

DETECCIÓN Y DIAGNÓSTICO DE FALLAS EN UN BIORREACTOR BATCH
UTILIZANDO ANÁLISIS DE COMPONENTES PRINCIPALES MULTIVARIANTES.

MARÍA STEFANY ANGARITA GÓMEZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISCOQUÍMICAS
ESCUELA DE INGENIERÍA QUÍMICA
BUCARAMANGA

2016

DETECCIÓN Y DIAGNÓSTICO DE FALLAS EN UN BIORREACTOR BATCH
UTILIZANDO ANÁLISIS DE COMPONENTES PRINCIPALES MULTIVARIANTES.

MARÍA STEFANY ANGARITA GÓMEZ

Trabajo de grado presentado como requisito para optar al título de:

Ingeniera Química

Director

M. Nazmul Karim

Ingeniero Químico, Ph.D

Texas A&M University

Codirector

Liliana del Pilar Castro

Ingeniera Química, Ph.D

Universidad Industrial de Santander

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISCOQUÍMICAS
ESCUELA DE INGENIERÍA QUÍMICA
BUCARAMANGA

2016

A Dios, por siempre darme la fortaleza necesaria para seguir adelante, bendecirme inmensamente a lo largo de estos años y consentirme como la niña de sus ojos.

A mi mamá y a mi papí, Ena y Mario, por su sacrificio y esfuerzo para que yo alcance mis sueños, por siempre darme apoyo y ánimo, cuando las situación se ponía difícil, por las comidas, tintos, el amor y cariño en las noches largas que me llenaban de energía para seguir.

A mi familia, a mi abuelito Efraín, que desde el cielo me sigue cuidando, a mi abuelita Ena por todo su amor, a mis tíos, tías y primos, por todo su apoyo, sin los cuales esto no hubiese sido posible.

A Luis, por todo su amor, comprensión y compañía a lo largo de estos años, impulsándome a dar lo mejor de mí para seguir hasta el final, siempre con la meta de alcanzar nuestros sueños juntos.

AGRADECIMIENTOS

Expreso mis profundos agradecimientos a:

El Dr. Nazmul Karim, por la gran oportunidad brindada para trabajar en su grupo de investigación, por su apoyo a lo largo del proceso, y por su contante guía en el desarrollo de este proyecto.

Los miembros del grupo de investigación del Dr. Karim, especialmente a Tejasvi, Melanie, Jonathan, Xinghua y Chiranjivi por compartir sus conocimientos y experiencias, y siempre ayudar en cada uno de los pasos del proyecto.

El Departamento de Ingeniería Química de Texas A&M University y a la Universidad Industrial de Santander, por el convenio de intercambio que me permitió realizar esta pasantía de investigación.

Los profesores y a la Escuela de Ingeniería Química de la Universidad Industrial de Santander, los cuales me brindaron una muy buena formación académica, profesional e integral durante la realización de mis estudios.

Dra. Liliana Castro, por orientarme en la etapa final de este proyecto, por sus aportes para concluir este trabajo de la mejor manera, y por depositar su confianza en mí.

Mis amigos y compañeros, por las risas, las enseñanzas, las noches de estudio, la compañía y el apoyo, y todo aquello que ha contribuido a mi crecimiento como persona.

CONTENIDO

	Pág.
INTRODUCCIÓN	11
1.1. FUNDAMENTOS TEÓRICOS	13
1.1.1. Análisis De Componentes Principales.....	14
1.1.2. Análisis De Componentes Principales Multivariante.	16
2. METODOLOGÍA	19
2.1. FASE 1: PRODUCCIÓN DE B-CAROTENO A ESCALA LABORATORIO.	19
2.1.1. Microorganismo y Medio de Cultivo	19
2.1.2. Fermentación en el Biorreactor.....	20
2.2. FASE 2: MODELO Y GENERACIÓN DE LA MATRIX X.....	21
2.3. FASE 3: MODELO ESTADÍSTICO MPCA.	23
3. RESULTADOS Y DISCUSIÓN.....	25
3.1. PRUEBA DE CONTROL EN MATRIZ SIN FALLAS	25
3.2. PRUEBA DE CONTROL EN MATRIZ CON FALLA DE CONTAMINACIÓN...26	
4. CONCLUSIONES	33
CITAS BIBLIOGRÁFICAS.....	34
BIBLIOGRAFÍA.....	37

LISTA DE FIGURAS

	Pág.
Figura 1. Clasificación de los métodos para monitorear procesos.	13
Figura 2. (a) Arreglo tridimensional para un proceso batch. (b) Despliegue IXJK. (c) Despliegue JXKI.	16
Figura 3. Biorreactor de vidrio.	20
Figura 4. Datos típicos de la fermentación. Los marcadores representan los datos experimentales y las líneas continuas representan el ajuste del modelo.	23
Figura 5. Evaluación de la Estadística Q (a), y la estadística <i>Hotelling T2</i> (b), para una matriz con datos de 500 batch sin fallas.	26
Figura 6. Comparación de un Batch con operación normal y con contaminación. a) Glucosa. b) Etanol. c) Ácido Acético. d) Biomasa. e) Beta-Caroteno.	28
Figura 7. Falla de contaminación detectada en los batch número 12, 77 y 134.	29
Figura 8. Contribución a (a) <i>Hotelling T2</i> y (b) Estadística Q para un batch contaminado.	30
Figura 9. Gráficas de contribución. a) 7 hr. b) 11 hr. c) 54 hr.	31

RESUMEN

TÍTULO: DETECCIÓN Y DIAGNÓSTICO DE FALLAS EN UN BIORREACTOR BATCH UTILIZANDO ANÁLISIS DE COMPONENTES PRINCIPALES MULTIVARIANTES*

AUTOR: MARÍA STEFANY ANGARITA GÓMEZ**

PALABRAS CLAVE: procesos batch, lotes, control estadístico de procesos, PCA, componentes principales, detección de fallas.

DESCRIPCIÓN:

Los métodos estadísticos multivariantes son herramientas muy poderosas capaces de manejar una gran cantidad de datos, altamente correlacionados y con comportamiento variable además de poco predecible. El análisis de componentes principales es utilizado para detectar y diagnosticar fallas en un biorreactor *batch*. La detección de una falla consiste en saber que esta existe y el diagnóstico se basa en saber cuál es la variable que se encuentra por fuera del rango normal de operación. Debido a la naturaleza dinámica de los procesos *batch*, aproximaciones multivariantes tienen que ser utilizadas, en donde únicamente es necesario conocer datos históricos del proceso, dichos datos deben describir el comportamiento normal del *batch*. Los componentes principales son retenidos para mantener la mayoría de la varianza y reducir las dimensiones del conjunto de datos iniciales, esto permite detectar la falla fácilmente y facilita el análisis. Tan pronto la falla es detectada un análisis de gráficas de contribución permite el diagnóstico de la falla. Experimentos fueron realizados para validar los resultados del modelo, los datos experimentales se obtienen de la fermentación aeróbica para la producción de β -caroteno utilizando *Saccharomyces cerevisiae* SM14 cepa modificada genéticamente y como sustrato 20 g/L de glucosa inicial, entre los datos se analiza el consumo de glucosa, la formación y posterior consumo de los subproductos como el etanol y ácido acético, además de la formación de biomasa y carotenoides.

*Trabajo de Grado

** Facultad de ingenierías fisicoquímicas. Escuela de Ingeniería Química. Director: M. Nazmul Karim. Codirector: Liliana del Pilar Castro.

ABSTRACT

TITLE: FAULT DETECTION AND DIAGNOSIS IN A BATCH BIORREACTOR USING MULTIVARIATE PRINCIPAL COMPONENT ANALYSYS*

AUTHOR: MARÍA STEFANY ANGARITA GÓMEZ**

KEYWORDS: batch process, statistical process control, PCA, principal components, fault detection.

DESCRIPTION:

Multivariate statistical methods are powerful tools capable to handle huge, noise and highly correlated data sets with unpredictable behavior. Principal Component Analysis (PCA) will be used to detect and predict faults in a batch bioreactor system. The detection of a fault is to know that this exists and the diagnosis is based on knowing what the variable is outside the normal operating range. Due to inherent dynamics in batch processes, multiway methods have to be applied in which only historical data sets of past successful batches are needed, this historical data needs to be clear from any failure. Principal Components are kept to extract the majority of variance and reduce the dimensionality of the original data set, making easier the analysis and fault detection. As soon as the fault is detected, it is necessary to find, the variable that presents the major deviation from its expected value therefore, a contribution plot for diagnosis is required. Experiments are performed to validate the results obtained from models. Experimental data is obtained from batch fermentation to describe the glucose consumption, product formation and depletion such as ethanol and acetic acid, and β -carotene production of the *Saccharomyces cerevisiae* strain mutant SM14 with 20 g/L glucose as the carbon source.

*Degree work

** Physics-Chemistry Engineering Faculty. Chemical Engineering School. Director: M. Nazmul Karim. Codirector: Liliana del Pilar Castro.

INTRODUCCIÓN

La operación por lotes o *batch* es ampliamente utilizada en la industria, en la producción de productos altamente valiosos y de poco volumen, entre ellos se incluyen la industria farmacéutica, bioquímica, alimenticia y de polímeros [1].

Los procesos que se llevan a cabo en *batch*, presentan una naturaleza dinámica, por lo cual requieren control por medio de acciones correctivas, que minimice los efectos negativos sobre el proceso y sobre el producto final. Adicionalmente, la operación *batch* agrega nuevas restricciones, por ejemplo: punto de operación variante en el tiempo, perturbaciones inherentes internas y comportamiento irreversible, lo que dificulta el óptimo control del proceso [2].

Particularmente, uno de los bioprocesos que se desarrolla en operación por lotes es la producción de carotenoides. Los carotenoides son un grupo diverso de pigmentos amarillos-naranjas encontrados en sistemas biológicos. Estos pigmentos son producidos por diversos organismos como las plantas, hongos y bacterias [3-5]. Los carotenos como el β -caroteno tienen un impacto positivo en la salud humana como el precursor de la vitamina A, además de tener un efecto antioxidante y propiedades contra el cáncer [6-9].

Algunos carotenoides son producidos industrialmente por vía química sintética, sin embargo los subproductos tienen efectos indeseados al ser consumidos y la mayoría de los carotenoides como el β -caroteno tienen una estructura compleja que hace que la síntesis química sea una opción inviable. Por esta razón, la producción de carotenoides por rutas bioquímicas se ha posicionado como una alternativa para la producción de pigmentos con gran interés a nivel industrial [10-14].

Para la producción del β -caroteno se utilizó un cepa modificada de *S. cerevisiae* SM14, con 20 g/L D-glucosa como concentración de sustrato inicial, se tuvo que

utilizar una modificación de *S. cerevisiae*, ya que la cepa original no es productora de β -caroteno y su producto final es únicamente etanol.

La producción de carotenoides en un proceso por lotes se ve limitado por la complejidad que presenta medir las variables en línea, monitoreo y control del proceso. Lo anterior conlleva a lazos de control que contiene grandes tiempos de retardo y fallas en el producto final.

Una alternativa de mejora para el control de procesos es el monitoreo estadístico de procesos multivariantes (MSPM, *Multivariate Statistical Process Monitoring*). Este es una técnica emergente usada para monitorear procesos *batch* y *semi batch*. Como consecuencia al comportamiento no lineal que se presenta al operar los procesos por lotes, el MSPM está ganando más aceptación de la industria porque provee información asertiva, dando advertencias de desviaciones en el proceso tempranamente y basándose en estadísticas fáciles de computar e interpretar. [15-17].

Para monitorear estadísticamente el proceso de producción de carotenoides, se requiere una adecuada detección y el diagnóstico de fallas. Si las fallas son detectadas a tiempo, mientras el proceso se encuentra todavía operando en una región controlable, se puede evitar el progreso de la falla, manteniendo las variables del proceso en el valor deseado, lo que reduce la pérdida de productividad y mantiene las cualidades finales del producto.

Por lo anterior, el objetivo de este trabajo de investigación fue desarrollar un modelo estadístico para la detección y el diagnóstico de fallas durante el proceso de producción de carotenoides en operación *batch*.

1.1. FUNDAMENTOS TEÓRICOS

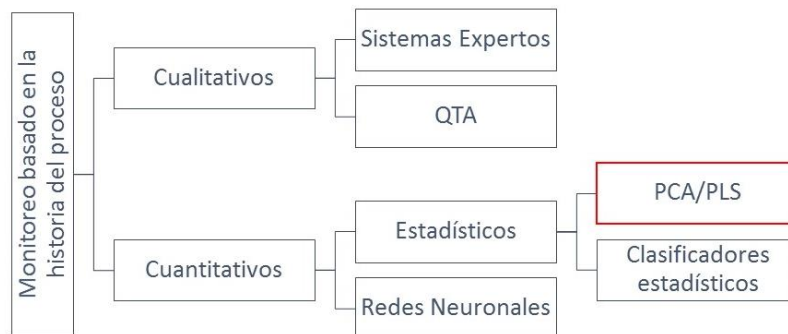
La literatura reporta diferentes métodos de monitoreo de procesos, los cuales se pueden resumir en tres grandes categorías que están basadas en: modelos cuantitativos, modelos cualitativos o en la historia del proceso [18-20].

El objetivo de los métodos es asegurar el éxito de las operaciones detectando las fallas de forma automática, lo que se traduce en mayor soporte e información para la toma de decisiones de cara a restaurar el comportamiento normal del proceso [21].

Los métodos basados en la historia del proceso son los más utilizados en los procesos por lotes, debido a que es muy complicado generar modelos matemáticos que correlacionen todas las variables presentes en un proceso biológico.

Particularmente las aproximaciones estadísticas como el análisis de componentes principales (Principal Component Analysis PCA Figura 1), es altamente utilizado debido a que extrae la mayor información acerca de las mayores tendencias dentro del conjunto de datos, usando un pequeño número de factores relevantes. Para aplicar este método es necesario tener a disponibilidad una gran cantidad de datos históricos del proceso. Dichas mediciones permiten la apropiada detección y diagnóstico de fallas.

Figura 1. Clasificación de los métodos para monitorear procesos.



Fuente: [18]

1.1.1. Análisis De Componentes Principales. El PCA es una técnica lineal muy conocida para la reducción de dimensiones en un conjunto de datos [22-24], capaz de capturar gran parte de la variabilidad de estos, determinando los componentes principales o también llamados vectores de carga, estos consisten en un conjunto de vectores ortogonales los cuales retienen la información más importante y esencial, lo que facilita el análisis [15, 21].

Si se considera un conjunto de datos de un proceso, arreglados en una matriz $X \in \mathcal{R}^{n \times m}$ en donde n son las muestras y m las variables del proceso:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \quad (1)$$

La matriz presentada en la ecuación (1) tiene un proceso de escalamiento, primero se debe centrar en la media, es decir a cada valor $(x_{11} \dots x_{1m}, x_{nm})$ se le resta el promedio de la correspondiente variable (columna), de tal manera que el nuevo promedio de dicha columna sea cero, además se puede recurrir a la varianza unitaria, igualmente esta consiste en dividir cada valor de la columna, en la desviación estándar de la correspondiente variable.

La matriz de covarianza R se puede descomponer utilizando la descomposición de valores singulares (singular value decomposition SVD)

$$R = \frac{1}{n-1} X^T X = V \Lambda V^T \quad (2)$$

$$R = V \Lambda V^T \quad (3)$$

De ahí se obtiene dos matrices importantes, V contiene los vectores propios o eigenvectores, y la matriz $\Lambda \in \mathcal{R}^{m \times m}$ contiene la diagonal de los valores propios de R o eigenvalores, en orden decreciente es decir donde $\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_m > 0$.

Para reducir la dimensionalidad original de la matriz solo se seleccionan los primeros a ($a < m$) (llamados también componentes principales) y también corresponde a la cantidad de vectores propios que se mantienen para posterior análisis $P \in R^{m \times a}$.

El número apropiado de componentes principales a se escoge por medio de la técnica de varianza acumulada (%VC) en la cual el porcentaje acumulado debe ser superior al 90% [25].

$$\%VC = \frac{\lambda_1 + \lambda_2 + \lambda_3 \dots + \lambda_a}{\lambda_1 + \lambda_2 + \lambda_3 \dots + \lambda_m} * 100\% \quad (4)$$

Las proyecciones de la matriz X , con un número menor de dimensiones se mantienen en la matriz T (scores), cada vector ortogonal en esta matriz contiene la información acerca de la variabilidad de los datos originales.

$$T = XP \quad (5)$$

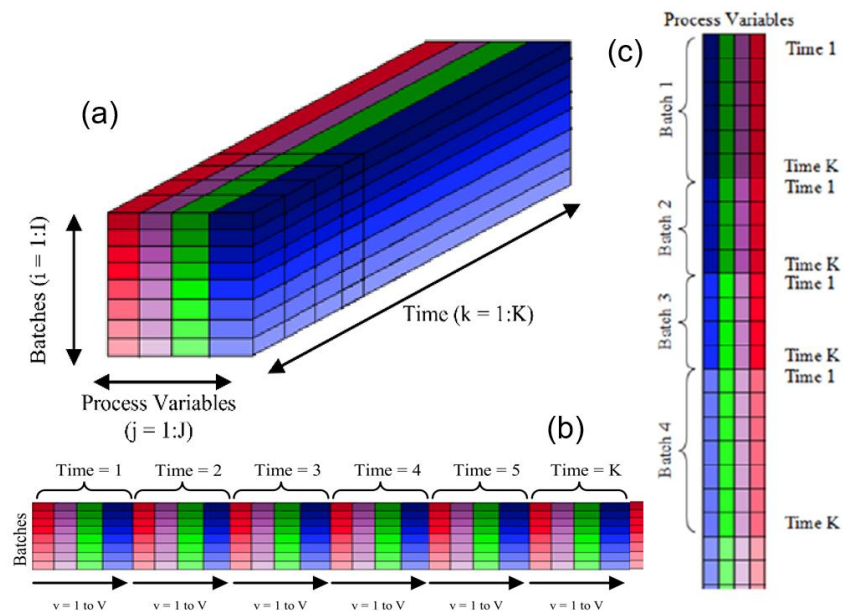
El nuevo modelo mantiene la mayoría de la varianza pero solo con a variables latentes, el resto de la varianza de las $n - a$ variables se encuentra en una matriz de error E .

$$X = \hat{X} + E = t_1 p_1^T + t_2 p_2^T + \dots + t_a p_a^T + E \quad (6)$$

$$X = \sum_{i=1}^a t_i p_i^T + E = TP^T + E \quad (7)$$

1.1.2. Análisis De Componentes Principales Multivariante. Aunque el PCA es muy usado para procesos continuos [26], aproximaciones como el MPCA por sus siglas en inglés (Multiway Principal Component Analysis MPCA) son las utilizadas para el monitoreo de procesos batch [27]. La principal característica de los conjuntos de datos de un proceso batch es el arreglo tridimensional de los mismos como se puede ver en la Figura 2 (a) en la cual (J) son las variables del proceso registradas por cada tiempo (K) y por cada batch (I). La relación entre MPCA y PCA, es que el MPCA es equivalente a aplicar el PCA en una matriz de dos dimensiones compuesta por uno de los seis posibles despliegues de la matriz tridimensional [27-29].

Figura 2. (a) Arreglo tridimensional para un proceso batch. (b) Despliegue IXJK. (c) Despliegue JXKI.



Fuente: [30]

La matriz tridimensional puede organizarse de seis formas diferentes $I \times JK$, $J \times KI$, $K \times IJ$ y sus equivalentes ($I \times KJ$, $J \times IK$, $K \times JI$). Los despliegues más utilizados son $I \times JK$ y $J \times KI$, el primero permite la comparación respecto a la trayectoria media de los *batch* (Figura 2 (b)), y por lo tanto se podría identificar los lotes que posiblemente tengan un comportamiento cuestionable. El segundo permite la comparación entre las variables Figura 2 (c), y por lo tanto también el monitoreo en línea [27, 30].

Para monitorear una nueva medición, se pueden utilizar dos distribuciones estadísticas diferentes. La primera es la distribución *Hotelling* T^2 , si solo esta distribución se encuentra por fuera de los límites de confianza, significa que el modelo del proceso es aun válido, pero la distancia entre el *batch* analizado y el centro del modelo es grande.

$$T^2 = x^T P(\Lambda_\alpha)^{-1} P^T x \quad (8)$$

La segunda distribución es la estadística Q, si esta se encuentra por encima de los límites de confianza, un nuevo evento se encuentra en los datos, que no es descrito por la matriz original del proceso.

$$Q = x^T (I - P P^T) x \quad (9)$$

Cada una de las distribuciones tiene un límite de confianza, por encima de los cuales las medidas serán descritas como una falla encontrada en el proceso. La distribución F es utilizada para calcular el límite de *Hotelling* T^2 en el cual α es el nivel de significancia [31].

$$T_\alpha^2 = \frac{a(n-1)}{n-a} F_{a,n-a,\alpha} \quad (10)$$

Y el límite para la distribución Q, en donde c_α es el valor de la distribución normal con α como nivel de significancia. Para calcular Q_α es necesario calcular primero

los parámetros θ_i y h_0 que dependen únicamente de los valores propios, y que se pueden hallar como se muestra en las ecuaciones 12 y 13 respectivamente.

$$Q_\alpha = \theta_1 \left(\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right)^{\frac{1}{h_0}} \quad (11)$$

$$\theta_i = \sum_{j=a+1}^m \lambda_j^i \quad (12)$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \quad (13)$$

2. METODOLOGÍA

El desarrollo de este trabajo se realizó en tres fases: producción de β -caroteno a escala laboratorio, modelo y generación de la matriz X y modelo estadístico MPCA.

2.1. FASE 1: PRODUCCIÓN DE B-CAROTENO A ESCALA LABORATORIO

La fase 1 se realizó principalmente para validar el modelo matemático que describe la cinética de la fermentación, con el fin de saber si el modelo previamente desarrollado en el grupo de investigación se ajustaba correctamente a los datos experimentales. Debido a que se requiere una gran cantidad de datos históricos del proceso para llevar a cabo el MPCA, para efectos de este trabajo esos datos fueron simulados con dicho modelo matemático, por lo que es importante confirmar el correcto ajuste de este.

A continuación se describe la metodología utilizada en el laboratorio para realizar la producción de β -caroteno.

2.1.1. Microorganismo y Medio de Cultivo. Se utilizó un cepa modificada de *S. cerevisiae* SM14, modificada para la producción de β -caroteno [32]. La cepa de levadura se almacena en viales congelados a -80°C y en placas a 4°C , que se subcultivaron cada tres semanas para el mantenimiento. Todos los experimentos se realizaron en medio de base nitrogenada (Yeast Nitrogen Base YNB) con 20 g/L D-glucosa.

El inóculo para el fermentador se preparó a partir de colonias individuales, inoculando 50 mL de medio YNB (20 g/L de glucosa) y fue incubada a 30°C por 72 horas con agitación constante 200 rpm.

2.1.2. Fermentación en el Biorreactor. La evaluación de la fermentación se llevó a cabo en un biorreactor de vidrio de 4 L, (Applikon®, Foster City, CA) Figura 3. Los experimentos se realizaron por duplicado en operación *batch*, a una temperatura de $30^{\circ}\text{C}\pm 2$, con una agitación constante de 800 rpm y un flujo de aire constante de 6 L/min, el pH se mantuvo en 4 mediante la adición de 2 M de HCl o 2 M de NaOH según como corresponda con el fin de contrarrestar el aumento o disminución del pH. El reactor fue alimentado con una solución de glucosa de 20 g/L y 50 mL de inóculo. La fermentación se realizó durante 72 horas (tiempo en el que el ácido acético producido es totalmente consumido).

Figura 3. Biorreactor de vidrio.



El consumo de sustrato, formación y consumo de subproductos fueron las variables de respuesta del proceso. Estas variables se cuantificaron usando el HPLC Agilent Technologies 1290 Infinity con el uso de una columna Aminex HPX-87H.

2.2. FASE 2: MODELO Y GENERACIÓN DE LA MATRIX X

La aplicación del MPCA requiere considerar datos históricos del proceso. Estos datos comúnmente se obtienen de registros de datos en la industria o el modelamiento del proceso. Para el desarrollo de este trabajo se utilizó como base de datos el modelamiento del proceso de producción de carotenoides (previamente desarrollado en el grupo de investigación del departamento de ingeniería química de la universidad de Texas A&M dirigido por el Dr. Karim), para posteriormente simular los datos del proceso.

La Figura 4 muestra los perfiles de fermentación para la producción de carotenoides, realizado con un cultivo de *S. cerevisiae* SM14 en un fermentador de tanque agitado con 20 g/L de glucosa inicial. En las primeras etapas de la fermentación el etanol y ácido acético es producido, debido a un metabolismo de exceso de sustrato. Cuando la glucosa es consumida, el etanol presente en el medio puede ser utilizado en el metabolismo aerobio para ser convertido a Acetyl-CoA, una vez el etanol se acaba, el acetato puede ser consumido [33]. El Acetyl-CoA se convierte en β -caroteno por medio del genoma introducido en la célula por medio de la recombinación [32, 34, 35].

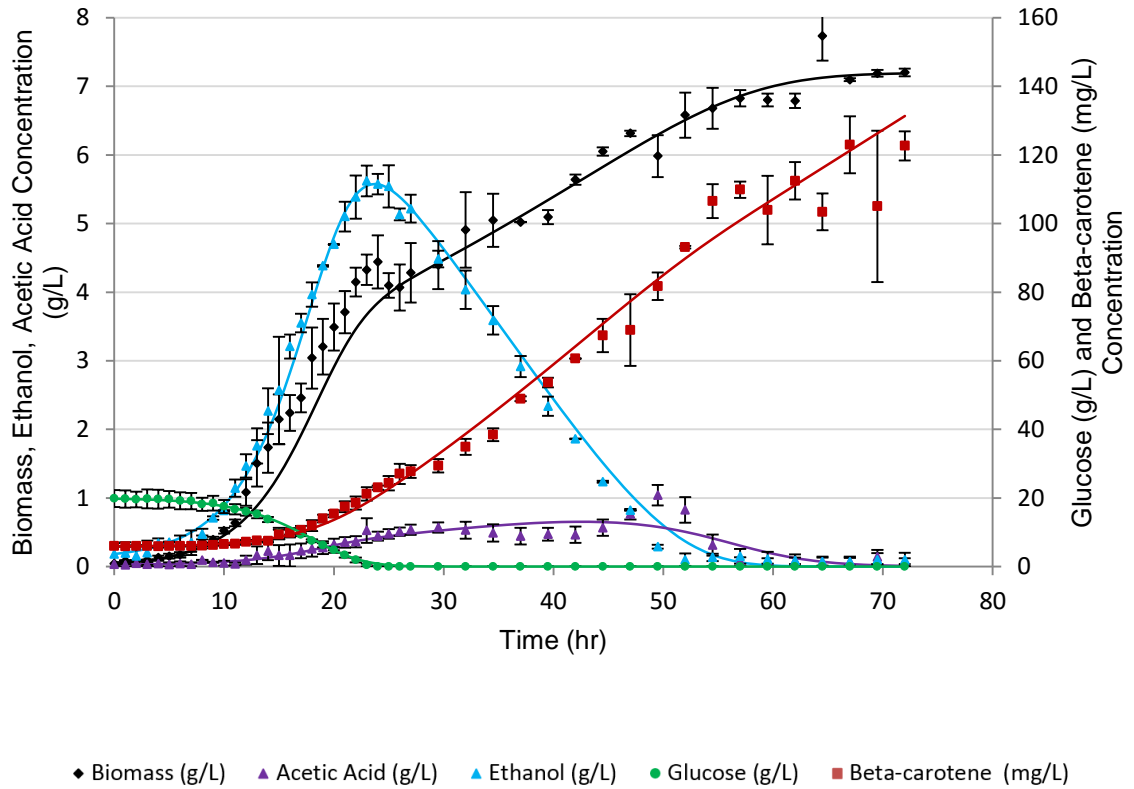
De acuerdo a la Figura 4, la levadura muestra un periodo de crecimiento exponencial en las primeras 24 horas, hasta que la glucosa se agota y el etanol se encuentra en su máxima concentración alrededor de 5.42 g/L. Después de que la glucosa se agota, el etanol producido es utilizado como fuente de carbono, lo que resulta en un agotamiento del etanol y al mismo tiempo el ácido acético alcanza su pico de concentración 1.19 g/L a las 50 horas, posteriormente se consume el ácido

acético. La biomasa alcanza un máximo de alrededor de los 7.9 g/L, la producción de Beta-caroteno comienza en etapas tempranas del crecimiento exponencial y continua a través del periodo de la fermentación, esta producción aumenta significativamente cuando se consumen los subproductos, alcanzando una concentración de 130 mg/L al final de las 72 horas.

Como se puede ver en la **Figura 4** el modelo matemático (representado por las líneas continuas), tiene un muy buen ajuste con respecto a los datos experimentales de la fermentación (representado por los puntos y sus respectivas desviaciones) y por lo tanto se puede utilizar para la simulación de los datos históricos.

Se simuló 500 *batch* utilizando el software MATLAB a partir del modelo matemático de las variables del proceso, para generar una matriz sin fallas (training data). Con el objeto de crear diferentes conjuntos de datos con un comportamiento real del proceso, durante la simulación se introdujeron variaciones típicas y aleatorias en las condiciones iniciales del proceso, por ejemplo: concentraciones iniciales de glucosa, biomasa y etanol.

Figura 4. Datos típicos de la fermentación. Los marcadores representan los datos experimentales y las líneas continuas representan el ajuste del modelo.



2.3. FASE 3: MODELO ESTADÍSTICO MPCA.

La primera parte para el desarrollo del modelo fue organización tridimensional de la matriz X ($I \times J \times K$) en donde I corresponde a los *batch* que fueron simulados, J son las 13 variables reportadas en la Tabla 1 en la cual las primeras 5 variables son registradas fuera de línea y el resto son variables registradas en línea, K corresponde a los 73 puntos en el tiempo (cada variable es registrada cada hora por 72 horas).

Posteriormente se desarrolló un programa en el software MATLAB, con el fin de realizar el tratamiento matemático descrito en la simulación, que consiste en el proceso de escalamiento, calcular la matriz de covarianzas, retener los

componentes principales y finalmente expresar la matriz X de la forma presentada en la ecuación 7.

$$X = \sum_{i=1}^a t_i p_i^T + E = TP^T + E$$

Por otra parte, se calculan los límites de confiabilidad de la estadística Q y de *Hotelling T²*, para que estos puedan ser utilizados en la detección de fallas. Un análisis adicional se puede realizar por medio de las gráficas de contribución, las cuales reportan para un momento determinado y un *batch* específico, la contribución de cada una de las variables al error conjunto.

Tabla 1. Variables usadas para monitorear la producción de carotenoides.

Variable	Nombre
1	Glucosa g/L
2	Biomasa g/L
3	Carotenoides mg/L
4	Etanol g/L
5	Ácido Acético g/L
6	pH
7	Adición de Ácido ml
8	Temperatura C
9	Adición de Base ml
10	dO2 Oxígeno disuelto
11	Agitación rpm
12	CO2 off gas
13	O2 off gas

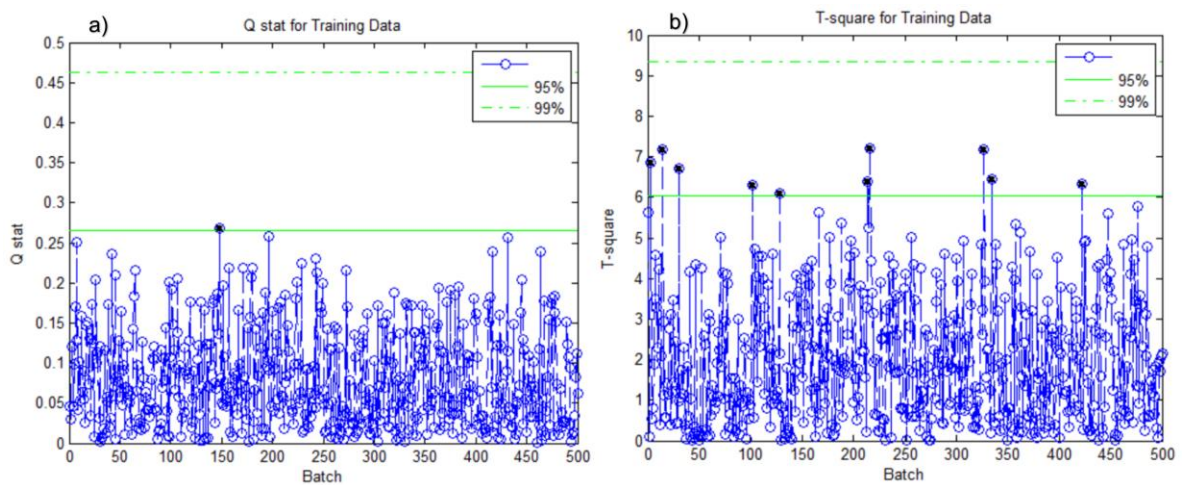
3. RESULTADOS Y DISCUSIÓN

En la siguiente sección se muestran los resultados obtenidos para el MPCA (Multiway Principal Component Analysis), aplicado al proceso de producción de carotenoides. En la primera parte se prueba el modelo con un conjunto de datos sin falla, con el fin de demostrar que los puntos se encuentran en la región de control. Posteriormente se introducen fallas individuales en distintas variables, en donde el arreglo $I \times JK$ permite detectar cual *batch* presenta la falla, después se procede a seleccionar el *batch* en cuestión y el arreglo bidimensional $KI \times J$ permite visualizar el momento en el que se presenta la falla. Por último por medio del análisis de graficas de contribución se evidencia la variable que se encuentra fuera del rango normal de operación y conseguir diagnosticar la falla.

3.1. PRUEBA DE CONTROL EN MATRIZ SIN FALLAS

La Figura 5 (a) Y 5 (b) muestran el análisis estadístico para 500 *batch* sin anomalías y operación normal. Los resultados estadísticos muestran que las pruebas están por debajo del 99% del límite de confiabilidad esto indica que la operación no presenta fallas y se corrobora que el modelo no muestra falsos positivos, ni arroja resultados imprecisos. Esta prueba fue realizada como control para la detección de fallas.

Figura 5. Evaluación de la Estadística Q (a), y la estadística *Hotelling T*² (b), para una matriz con datos de 500 batch sin fallas.



Aunque en las gráficas existen algunos puntos que superan el límite del 95% (representados por los puntos negros), para efectos de este trabajo la falla se rectifica si también supera el 99%. Cabe resaltar que solo se mantienen 2 componentes principales, los cuales retienen el 99.95% de la varianza total (existen 949 componentes).

En síntesis, la Figura 5 muestra que cuando el modelo de MPCA realizado se aplica a una matriz que no presente fallas ni alteraciones, su respuesta frente al análisis de los datos es de manera congruente, lo cual es un indicador de que el modelo es correcto, y que la cantidad de componentes principales elegidos, representan adecuadamente el comportamiento general de la matriz.

3.2. PRUEBA DE CONTROL EN MATRIZ CON FALLA DE CONTAMINACIÓN

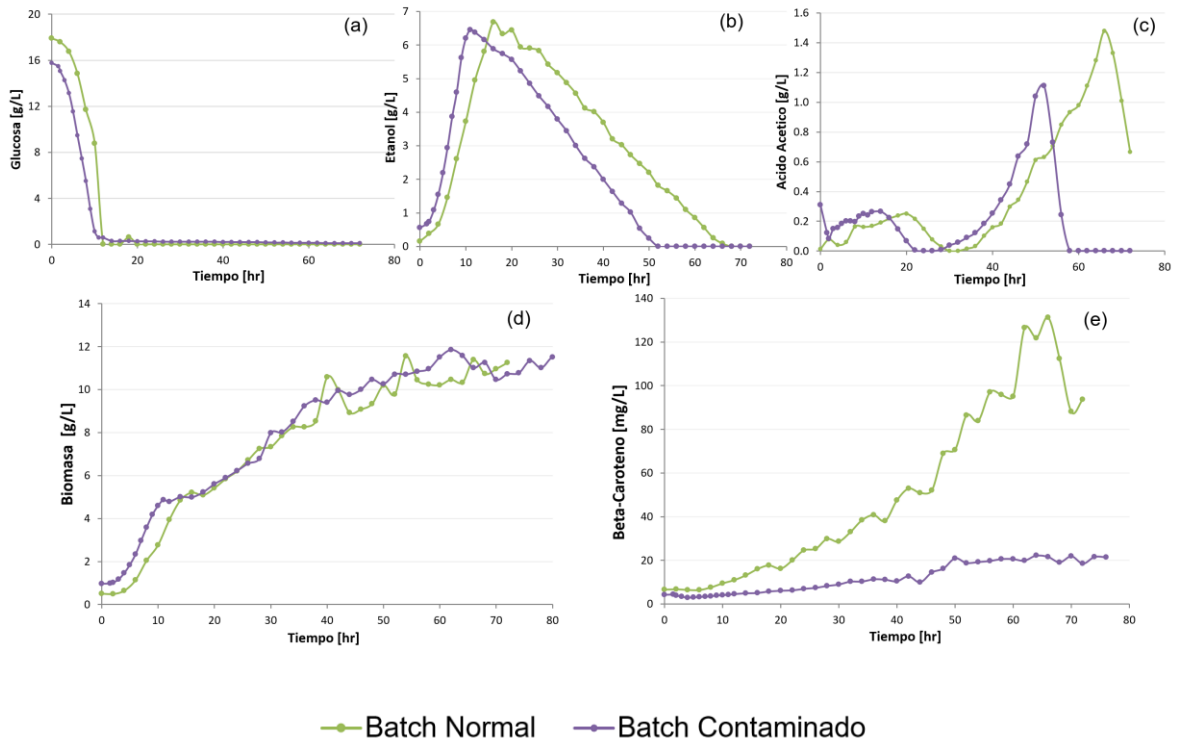
Una de las fallas estudiadas en el grupo de investigación, fue la contaminación del proceso, esto se debe principalmente a que la contaminación puede afectar la

calidad y la cantidad del producto final, lo que se ve reflejado en una pérdida de dinero y de tiempo en la industria de procesos.

La contaminación se realizó con *Saccharomyces cerevisiae*, sin ninguna modificación (wild strain FY2), cabe resaltar que esta no es productora de β -caroteno, por lo contrario su producto es el etanol, el cual también es fuente de carbono para la cepa SM14, motivo por el cual también se quiso saber sus efectos sobre el proceso.

En la Figura 6 presenta los perfiles de biomasa, sustrato y producto para las fermentaciones con y sin contaminación. Los resultados muestran que el consumo de glucosa no presenta una desviación importante. Por lo contrario la producción y consumo del etanol y del ácido acético muestran aproximadamente la misma tendencia, pero con variaciones de los tiempos, por ejemplo para el caso del etanol este es consumido 10 horas antes que el tiempo promedio de un batch sin contaminación. Se pudo observar que la variable más afectada por la contaminación del sistema fue la producción de Carotenoides, en donde esta se redujo en aproximadamente un 80%, lo que implica que únicamente hasta que se termina el proceso y se extrae el producto final de la célula, se conocería que hubo una falla en el proceso.

Figura 6. Comparación de un Batch con operación normal y con contaminación. a) Glucosa. b) Etanol. c) Ácido Acético. d) Biomasa. e) Beta-Caroteno.

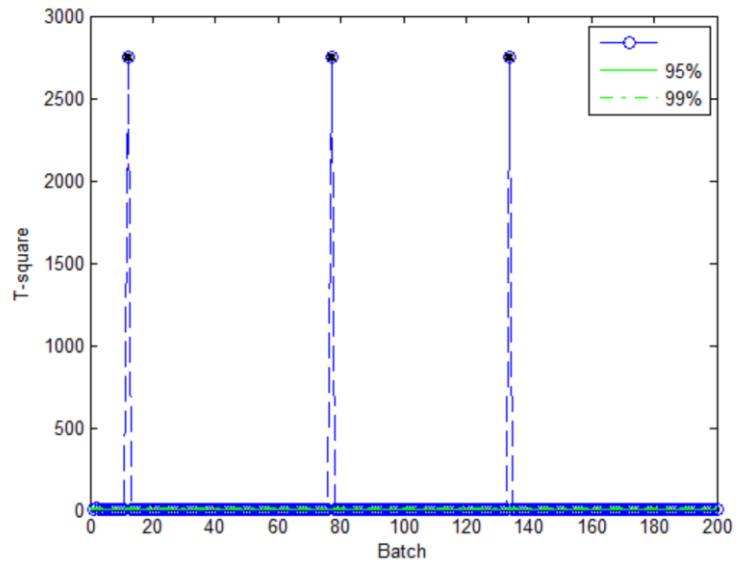


Aunque la contaminación del proceso se realizó en la hora cero, únicamente se comienzan a ver diferencias apreciables en el ácido acético después de 50 horas de fermentación y al ser el β -caroteno un producto intracelular que requiere un largo proceso de extracción, no se podría deducir que hubo una falla en el proceso hasta que este transcurriera en su totalidad.

Con el propósito de comprobar el resultado del MPCA cuando la matriz se somete a la presencia de fallas, se realiza el mismo procedimiento descrito en la sección 3.1, en esta oportunidad se introducen los datos correspondientes al comportamiento de un *batch* contaminado. La Figura 7 muestra 200 ensayos simulados con 3 contaminaciones en diferentes lotes. Las fallas fueron introducidas en los lotes número 12, 77, 134. Los resultados muestran que las fallas fueron

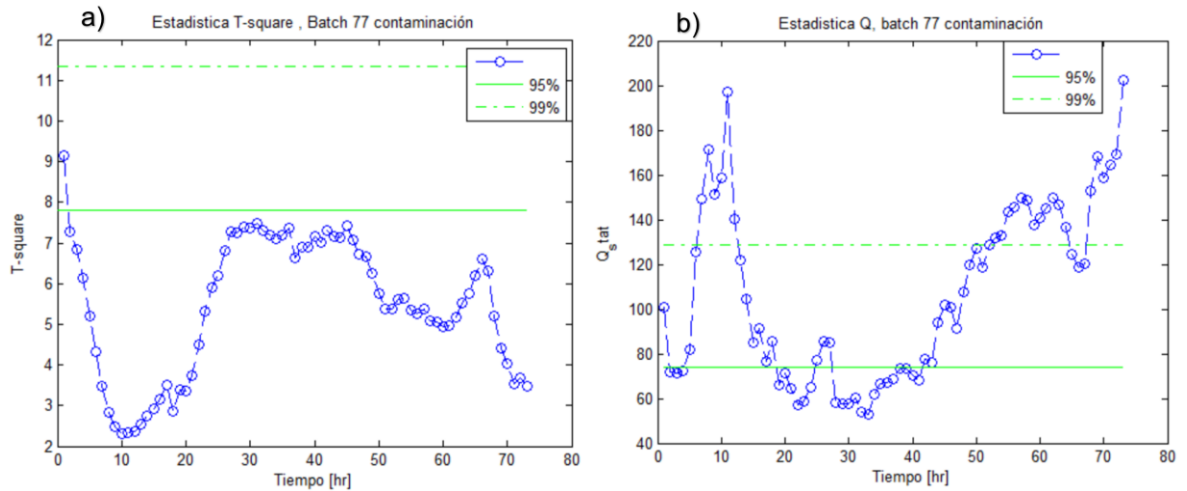
detectadas por el programa, los límites de confiabilidad se encuentran en la parte inferior (dado a que la escala estadística tiene valores altos, los límites no se observan).

Figura 7. Falla de contaminación detectada en los batch número 12, 77 y 134.



Si la matriz original se reorganiza en el arreglo $KI \times J$, esto permite la comparación dentro de un *batch*, por lo cual se procede a analizar el *batch* 77 (Figura 8). En este ensayo se introdujo una falla de contaminación, como resultado se puede observar el momento en el cual la falla es detectada. Este arreglo provee múltiples ventajas, como una detección temprana de cualquier falla y además un análisis de graficas de contribución.

Figura 8. Contribución a (a) *Hotelling T²* y (b) Estadística Q para un batch contaminado.

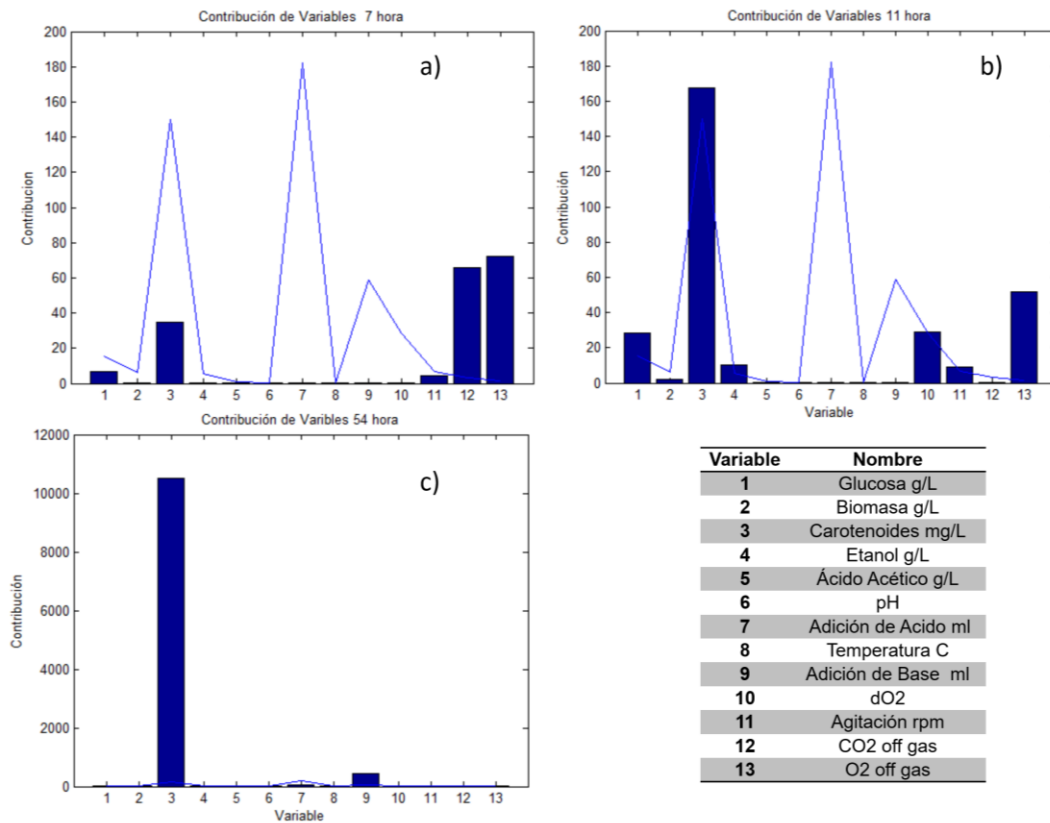


La Estadística *Hotelling T²* no muestra fallas Figura 8 (a), se concluye que el modelo es aun válido. Sin embargo para el *batch 77* únicamente la estadística Q muestra falla (Figura 8 b) , lo cual significa que se ha encontrado un nuevo evento en los datos y que este no es representado por el MPCA planteado para el proceso, presuntamente por que no se encuentra descrito en la matriz sin fallas (training data). Para este análisis solo se retienen 3 componentes principales, que mantienen el 99.66% de la varianza total.

El análisis de gráficas de contribución permite saber la contribución o aporte de cada variable en determinado tiempo, lo cual ayuda a establecer cuáles variables se encuentran fuera del rango de operación normal en ese momento. Al ver la Figura 8 (b) se puede observar que se presentan fallas desde la hora 7 hasta la hora 12 y nuevamente después de la hora 52 (si se considera únicamente el límite del 99%).

La Figura 9 presenta el análisis de las variables que se encuentran fuera del rango normal de operación, se analiza la hora 7 ,11 y 54.

Figura 9. Gráficas de contribución. a) 7 hr. b) 11 hr. c) 54 hr.



Las líneas continuas representan el límite por debajo del cual esa variable aún se encuentra en normalidad para el proceso, y las barras representan la contribución de esa variable a la estadística Q para un tiempo analizado.

Para este tipo de gráficas de contribución no se puede saber si la variable está por encima o por debajo del rango normal de operación, por lo cual únicamente se puede concluir acerca de que se encuentra en un valor que no es normal. En la Figura 9 (a) se puede observar fallas en el registro de la salida de gases del oxígeno y dióxido de carbono, lo cual es un indicador directo de que existe un problema con el crecimiento microbiano dentro del reactor, por ejemplo si existe diferentes tipos de células el consumo de oxígeno y la producción de dióxido de carbono puede variar considerablemente. A partir de ese momento se puede tomar la decisión de

terminar el proceso y desechar el *batch*, ahorrando tiempo y dinero, o se puede seguir analizando la contribución de las variables en tiempos posteriores.

En la Figura 9 (b) la falla se extiende a la glucosa y a la producción de Carotenoides, esto no implica que exista en el reactor más glucosa o menos carotenoides, solo refleja que las variables se encuentran por fuera de los límites normales de operación. Posteriormente en la Figura 9 (c) ya se puede comprobar el mayor error presente en la producción de β -caroteno al finalizar el proceso.

4. CONCLUSIONES

- El modelo estadístico desarrollado por medio del análisis de componentes principales multivariante, permite la correcta detección y diagnóstico de fallas, en el proceso de producción de β -caroteno en un reactor modo *batch*, manteniendo límites de confiabilidad del 99%.
- Las gráficas de contribución permiten analizar el aporte de cada una de las variables, para el caso del análisis de una falla de contaminación con *Saccharomyces cerevisiae* la detección temprana se presenta alrededor de la 7 hora, y muestra una desviación en la salida de gases, lo que significa cambios en la respiración celular, bien sea por mayor consumo de oxígeno y producción de dióxido de carbono o lo contrario, en cualquier caso indica que el crecimiento celular no está ocurriendo adecuadamente.
- El MPCA requiere un gran número de datos histórico del proceso, la simulación es un método efectivo para obtener los datos del proceso para es necesario una buena representación de la operación normal del proceso, y de las variaciones típicas de un proceso biológico, es por esto que se requiere la introducción de variaciones en la condiciones iniciales del proceso.

CITAS BIBLIOGRÁFICAS

1. Duchesne, C. and J.F. MacGregor, *Multivariate analysis and optimization of process variable trajectories for batch processes*. Chemometrics and Intelligent Laboratory Systems, 2000. **51**(1): p. 125-137.
2. Gómez, L.M., H.D. Álvarez, and H.A. Botero, *Limitaciones de la Controlabilidad de Estados en los Procesos por Lotes*. Limitations of the State Controllability of Batch Processes., 2012. **23**(5): p. 97-108.
3. Armstrong, G.A. and J.E. Hearst, *Carotenoids 2: Genetics and molecular biology of carotenoid pigment biosynthesis*. The FASEB Journal, 1996. **10**(2): p. 228-37.
4. Bender, D.A., *Nutritional Biochemistry of the Vitamins*. Vol. 2nd ed. 2003, Cambridge, UK: Cambridge University Press.
5. Hulshof, P.J.M., et al., *Quick screening of maize kernels for provitamin A content*. Journal of Food Composition and Analysis, 2007. **20**(8): p. 655-661.
6. Nelis, H.J. and A. De Leenheer, *Microbial sources of carotenoid pigments used in foods and feeds*. Journal of Applied Bacteriology, 1991. **70**(3): p. 181-191.
7. Frengova, G.I. and D.M. Beshkova, *Carotenoids from Rhodotorula and Phaffia: yeasts of biotechnological importance*. Journal of industrial microbiology & biotechnology, 2009. **36**(2): p. 163-180.
8. Vachali, P., P. Bhosale, and P.S. Bernstein, *Microbial carotenoids*. Microbial carotenoids from fungi: methods and protocols, 2012: p. 41-59.
9. Lee, P. and C. Schmidt-Dannert, *Metabolic engineering towards biotechnological production of carotenoids in microorganisms*. Applied Microbiology and Biotechnology, 2002. **60**(1-2): p. 1-11.
10. Kiokias, S. and M.H. Gordon, *Antioxidant Properties of Carotenoids In Vitro and In Vivo*. Food Reviews International, 2004. **20**(2): p. 99-121.
11. Johnson, E.A. and W.A. Schroeder, *Microbial carotenoids, in Downstream processing biosurfactants carotenoids*. 1995, Springer. p. 119-178.

12. Raja, R., S. Hemaiswarya, and R. Rengasamy, *Exploitation of Dunaliella for β -carotene production*. Applied Microbiology and Biotechnology, 2007. **74**(3): p. 517-523.
13. Liu, Y.S. and J.Y. Wu, *Modeling of Xanthophyllomyces dendrorhous growth on glucose and overflow metabolism in batch and fed-batch cultures for astaxanthin production*. Biotechnology and bioengineering, 2008. **101**(5): p. 996-1004.
14. Luna-Flores, C., et al., *Batch and fed-batch modeling of carotenoids production by Xanthophyllomyces dendrorhous using Yucca fillifera date juice as substrate*. Biochemical Engineering Journal, 2010. **53**(1): p. 131-136.
15. Louwarse, D.J. and A.K. Smilde, *Multivariate statistical process control of batch processes based on three-way models*. Chemical Engineering Science, 2000. **55**(7): p. 1225-1235.
16. MacGregor, J. and A. Cinar, *Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods*. Computers & Chemical Engineering, 2012. **47**(0): p. 111-120.
17. Nomikos, P. and J.F. MacGregor, *Multivariate SPC Charts for Monitoring Batch Processes*. Technometrics, 1995. **37**(1): p. 41-59.
18. Venkatasubramanian, V., et al., *A review of process fault detection and diagnosis: Part I: Quantitative model-based methods*. Computers & Chemical Engineering, 2003. **27**(3): p. 293-311.
19. Venkatasubramanian, V., R. Rengaswamy, and S.N. Kavuri, *A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies*. Computers & Chemical Engineering, 2003. **27**(3): p. 313-326.
20. Venkatasubramanian, V., et al., *A review of process fault detection and diagnosis: Part III: Process history based methods*. Computers & Chemical Engineering, 2003. **27**(3): p. 327-346.
21. Garcia-Alvarez, D. and M.J. Fuente, *Estudio comparativo de técnicas de detección de fallos basadas en el Análisis de Componentes Principales (PCA)*. Revista Iberoamericana de Automática e Informática Industrial RIAI, 2011. **8**(3): p. 182-195.
22. Ku, W., R.H. Storer, and C. Georgakis, *Disturbance detection and isolation by dynamic principal component analysis*. Chemometrics and Intelligent Laboratory Systems, 1995. **30**(1): p. 179-196.

23. Chiang, L.H., R.D. Braatz, and E.L. Russell, *Fault detection and diagnosis in industrial systems*. 2001: Springer Science & Business Media.
24. MacGregor, J.F. and T. Kourti, *Statistical process control of multivariate processes*. Control Engineering Practice, 1995. **3**(3): p. 403-414.
25. Afifi, A., S. May, and V.A. Clark, *Computer-aided multivariate analysis*. 2003: CRC Press.
26. Wold, S., K. Esbensen, and P. Geladi, *Principal component analysis. Chemometrics and intelligent laboratory systems*, 1987. **2**(1): p. 37-52.
27. Nomikos, P. and J.F. MacGregor, *Monitoring batch processes using multiway principal component analysis*. AIChE Journal, 1994. **40**(8): p. 1361-1375.
28. Flores-Cerrillo, J. and J.F. MacGregor, *Multivariate monitoring of batch processes using batch-to-batch information*. AIChE Journal, 2004. **50**(6): p. 1219-1228.
29. Yoo, C.K., et al., *On-line monitoring of batch processes using multiway independent component analysis*. Chemometrics and Intelligent Laboratory Systems, 2004. **71**(2): p. 151-163.
30. Molloy, M. and E.B. Martin. *Application of Multiway Principal Component Analysis for Identification of Process Improvements in Pharmaceutical Manufacture*. in *Dynamics and Control of Process Systems*. 2013.
31. Wang, B., X. Yan, and Q. Jiang, *Loading-Based Principal Component Selection for PCA Integrated with Support Vector Data Description*. Industrial & Engineering Chemistry Research, 2015. **54**(5): p. 1615-1627.
32. Reyes, L.H., J.M. Gomez, and K.C. Kao, *Improving carotenoids production in yeast via adaptive laboratory evolution*. Metabolic engineering, 2014. **21**: p. 26-33.
33. Kappeli, O., *Regulation of carbon metabolism in Saccharomyces cerevisiae and related yeasts*. Adv Microb Physiol, 1986. **28**: p. 181-209.
34. Postma, E., et al., *Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of Saccharomyces cerevisiae*. Applied and environmental microbiology, 1989. **55**(2): p. 468-477.
35. Jacobson, M.K. and C. Bernofsky, *Mitochondrial acetaldehyde dehydrogenase from Saccharomyces cerevisiae*. Biochimica et Biophysica Acta (BBA)-Enzymology, 1974. **350**(2): p. 277-291.

BIBLIOGRAFÍA

AFIFI, A., MAY, S., Y CLARK, V.A., Computer-aided multivariate analysis: CRC Press, 2003.

CHIANG, L.H., BRAATZ, R.D., Y RUSSELL, E.L., Fault detection and diagnosis in industrial systems: Springer Science & Business Media, 2001.

DUCHESNE, C. Y MACGREGOR, J.F., Multivariate analysis and optimization of process variable trajectories for batch processes. Chemometrics and Intelligent Laboratory Systems, 2000. **51**(1): p. 125-137.

FLORES-CERRILLO, J. Y MACGREGOR, J.F., Multivariate monitoring of batch processes using batch-to-batch information. AIChE Journal, 2004. **50**(6): p. 1219-1228.

GARCIA-ALVAREZ, D. Y FUENTE, M.J., Estudio comparativo de técnicas de detección de fallos basadas en el Análisis de Componentes Principales (PCA). Revista Iberoamericana de Automática e Informática Industrial RIAI, 2011. **8**(3): p. 182-195.

GARCÍA-MUÑOZ, S., KOURTI, T., Y MACGREGOR, J.F., Model Predictive Monitoring for Batch Processes. Industrial & Engineering Chemistry Research, 2004. **43**(18): p. 5929-5941.

GOLSHAN, M., et al. Latent Variable MPC for trajectory tracking in batch processes: Role of the model structure. In American Control Conference, 2009. ACC '09. 2009.

GOLSHAN, M., ET AL., Latent Variable Model Predictive Control (LV-MPC) for trajectory tracking in batch processes. Journal of Process Control, 2010. **20**(4): p. 538-550.

GOLSHAN, M., MACGREGOR, J.F., Y MHASKAR, P., Latent variable model predictive control for trajectory tracking in batch processes: Alternative modeling approaches. *Journal of Process Control*, 2011. **21**(9): p. 1345-1358.

GÓMEZ, L.M., ÁLVAREZ, H.D., Y BOTERO, H.A., Limitaciones de la Controlabilidad de Estados en los Procesos por Lotes. *Limitations of the State Controllability of Batch Processes.*, 2012. **23**(5): p. 97-108.

KASSIDAS, A., MACGREGOR, J.F., Y TAYLOR, P.A., Synchronization of batch trajectories using dynamic time warping. *American Institute of Chemical Engineers. AIChE Journal*, 1998. **44**(4): p. 864.

KOURTI, T. Y MACGREGOR, J.F., PROCESS ANALYSIS, Monitoring and diagnosis, using multivariate projection methods. *Chemometrics and Intelligent Laboratory Systems*, 1995. **28**(1): p. 3-21.

KOURTI, T., NOMIKOS, P., Y MACGREGOR, J.F., ANALYSIS, Monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *Journal of Process Control*, 1995. **5**(4): p. 277-284.

KU, W., STORER, R.H., Y GEORGAKIS, C., Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 1995. **30**(1): p. 179-196.

LOUWERSE, D.J. Y SMILDE, A.K., Multivariate statistical process control of batch processes based on three-way models. *Chemical Engineering Science*, 2000. **55**(7): p. 1225-1235.

MACGREGOR, J. Y CINAR, A., Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods. *Computers & Chemical Engineering*, 2012. **47**(0): p. 111-120.

MACGREGOR, J. Y KOURTI, T., Statistical process control of multivariate processes. *Control Engineering Practice*, 1995. **3**(3): p. 403-414.

MACGREGOR, J.F., et al., Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal*, 1994. **40**(5): p. 826-838.

MACGREGOR, J.F. Y KOURTI, T., Statistical process control of multivariate processes. *Control Engineering Practice*, 1995. **3**(3): p. 403-414.

MOLLOY, M. Y MARTIN, E.B. Application of Multiway Principal Component Analysis for Identification of Process Improvements in Pharmaceutical Manufacture. in *Dynamics and Control of Process Systems*. 2013.

NOMIKOS, P. Y MACGREGOR, J.F., Monitoring batch processes using multiway principal component analysis. *AIChE Journal*, 1994. **40**(8): p. 1361-1375.

NOMIKOS, P. Y MACGREGOR, J.F., Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics*, 1995. **37**(1): p. 41-59.

NOMIKOS, P. Y MACGREGOR, J.F., Multi-way partial least squares in monitoring batch processes. *Chemometrics and Intelligent Laboratory Systems*, 1995. **30**(1): p. 97-108.

VENKATASUBRAMANIAN, V. Abnormal events management in complex process plants: challenges and opportunities in intelligent supervisory control. in *Proceedings FOCAPO*. 2003.

VENKATASUBRAMANIAN, V., RENGASWAMY, R., Y KAVURI, S.N., A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. *Computers & Chemical Engineering*, 2003. **27**(3): p. 313-326.

VENKATASUBRAMANIAN, V., et al., A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering*, 2003. **27**(3): p. 327-346.

VENKATASUBRAMANIAN, V., et al., A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, 2003. **27**(3): p. 293-311.

WANG, B., YAN, X., Y JIANG, Q., Loading-Based Principal Component Selection for PCA Integrated with Support Vector Data Description. *Industrial & Engineering Chemistry Research*, 2015. **54**(5): p. 1615-1627.

WOLD, S., ESBENSEN, K., Y GELADI, P., Principal component analysis. *Chemometrics and intelligent laboratory systems*, 1987. **2**(1): p. 37-52.

WOLD, S., et al., The PLS model space revisited. *Journal of Chemometrics*, 2009. **23**(2): p. 67-68.

YOO, C.K., et al., On-line monitoring of batch processes using multiway independent component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2004. **71**(2): p. 151-163.

YOON, S. Y MACGREGOR, J.F., Principal-component analysis of multiscale data for process monitoring and fault diagnosis. *AIChE Journal*, 2004. **50**(11): p. 2891-2903.