

Artículo Original

## **Análisis de componentes principales aplicado a la fermentación alcohólica**

### **Application of principal component analysis to alcoholic fermentation**

**López de la Maza, L.E.; Zumalacárregui de Cárdenas, L.; Pérez Ones, O.\***

<sup>1</sup>Universidad Tecnológica de La Habana "José Antonio Echeverría", Facultad de Ingeniería Química, Departamento de Ingeniería Química. La Habana, Cuba

#### **RESUMEN**

La etapa de fermentación es la más crítica en la fabricación de bebidas alcohólicas pues determina la cantidad de etanol que se va a producir. La naturaleza biológica de este proceso implica que incidan sobre su desarrollo una gran cantidad de variables, cuyos efectos combinados resultan de difícil interpretación. Las técnicas de análisis multivariante como el análisis de componentes principales permiten obtener una mejor comprensión de procesos complejos. La investigación tuvo como objetivo identificar la influencia de las condiciones iniciales de fermentación en el rendimiento de esta etapa. Para el estudio se recopilaron 249 instancias de las variables temperatura, concentración de células vivas en el inóculo, pH, sólidos aerométricos disueltos y grado alcohólico. Las variables originales fueron centradas y posteriormente se realizó en análisis de componentes principales utilizando el software Unscrambler v8.0. Con un único componente principal se explica el 94,9% de la varianza del conjunto inicial de datos. Se comprobó que el intervalo en que normalmente se trabajan las condiciones iniciales de las variables temperatura, pH, concentración de células vivas y concentración de sólidos aerométricos disueltos dan lugar a fermentaciones con rendimiento medio y bajo. Se identificó la potencialidad de incrementar los rendimientos de la etapa de fermentación a valores de grado alcohólico mayores que 5,9.

**Palabras clave:** Fermentación, análisis de componentes principales.

#### **ABSTRACT**

Fermentation is the most critical stage in the manufacture of alcoholic beverages. A large number of variables influences the process development due to its biological nature. Multivariate analysis techniques such as principal component analysis allow a better understanding of complex processes. The objective of the research was to identify the influence of the initial conditions of fermentation on the performance of this stage. For the research, 249 instances of temperature variables, concentration of living cells in the inoculum, pH, aerometric solids and alcoholic grade were collected. The original variables were centred before the principal component analysis were done using the Unscrambler v8.0. With a single main component, 94.9% of the variance of the initial data set was explained. It has been proven that the interval in which the initial conditions are usually seen results in medium and low yield fermentations. The potential to increase the yields of the fermentation stage to values of alcoholic grade greater than 5.9 was identified.

**Key words:** Fermentation, principal component analysis.

**\*Autor Correspondiente: Osney Pérez Ones.** Universidad Tecnológica de La Habana "José Antonio Echeverría", Facultad de Ingeniería Química, Departamento de Ingeniería Química. La Habana, Cuba. Correo electrónico: osney@quimica.cujae.edu.cu  
Fecha de recepción: 18/12/2018 Fecha de aceptación: 30/03/2019



## **INTRODUCCIÓN**

La fermentación alcohólica es la transformación anaeróbica de azúcares, principalmente hexosas como la glucosa y la fructosa, en etanol y dióxido de carbono, así como la generación de un gran número de subproductos (Buratti & Benedetti, 2016; Epifanio, 2005; Zamora, 2009). La obtención de etanol por vía fermentativa constituye un paso intermedio, pero crítico, en la producción de bebidas alcohólicas.

En las destilerías cubanas se utiliza la fermentación discontinua tradicional, tipo Jackemine y las mieles finales del proceso de fabricación de azúcar constituyen un medio de cultivo de alta calidad por la cantidad de nutrientes presentes en su composición (Menéndez, 2010).

La fermentación alcohólica es un proceso de largos tiempos de operación, pues debido a su naturaleza biológica inciden sobre su desarrollo una gran variedad de variables y parámetros operacionales tales como: concentración de azúcares, temperatura, pH, concentración de células vivas, cepa utilizada, entre otros. Esta razón provoca que, desde el punto de vista económico, varios de esos parámetros operacionales deban encontrarse dentro de un intervalo restringido para garantizar la mayor eficiencia posible en un tiempo de operación razonable. Todos los estudios que permitan una mejor comprensión de esta etapa contribuirán al perfeccionamiento de los modos de operación y a la eficiencia del proceso.

El análisis de componentes principales (ACP) es una técnica de análisis multivariante que permite la reducción de la dimensionalidad de la base de datos y mejor interpretación de la misma, con una mínima pérdida de información. Desde el punto de vista matemático los componentes principales son combinaciones lineales de las variables originales de manera que sean ortogonales entre sí. La búsqueda de los componentes principales se reduce a la solución de un problema de autovectores y autovalores (Abdi & Williams, 2010; García & Londoño, 2015; Jolliffe & Cadima, 2016; Tharwat, 2016; Zayas, Lorenzo, Hernández, & Blanco, 2014).

Varios autores (Aceves-Lara, Latrille, Buffiere, Bernet, & Steyer, 2008; Gunther, Seborg, & Baclaski, 2006; Lee, Yoo, & Lee, 2004; Mesa & Gozá, 2018; Valles, 2015) han aplicado el análisis de componentes principales en datos referentes a procesos de fermentación para la determinación del mecanismo de reacción, la detección de fallas y la mejor comprensión de las fuentes de variabilidad.

La tecnología existente en la destilería "Héctor Molina" permite contar con grandes volúmenes de información de las variables y parámetros del proceso de fermentación alcohólica, pero su procesamiento no se explota con la finalidad de optimizar los rendimientos. El conocimiento de la influencia de las condiciones iniciales permitiría encontrar intervalos óptimos de variables de operación o en otro sentido poner mayor atención en otros aspectos operacionales para mejorar los rendimientos.

En este trabajo se realizó el ACP a las condiciones iniciales de fermentación en la destilería "Héctor Molina" a partir de la información almacenada en las hojas de fermentación con el objetivo de identificar la influencia de las condiciones iniciales de fermentación en el rendimiento de esta etapa.

## MATERIALES Y MÉTODOS

La realización de la investigación se dividió en 3 etapas:

1. Recopilación de los datos.
2. Transformación de los datos.
3. Análisis de componentes principales.

### Recopilación de datos

Los valores de las variables temperatura (T), concentración de células vivas en el inóculo (Xv), pH, sólidos aerométricos disueltos (<sup>0</sup>Brix) y grado alcohólico (cEt) utilizados en el análisis se extrajeron de las hojas de control en la etapa de fermentación que se registraron en la destilería "Héctor Molina". Los valores de T, Xv, pH y <sup>0</sup>Brix corresponden a condiciones iniciales de la etapa de fermentación y la cEt se determinó una vez concluida esta etapa. Se pudieron recopilar datos de 249 fermentaciones. Los intervalos en los que se mueven las variables estudiadas se muestran en la Tabla 1.

Durante la fermentación se tomaron muestras en los fermentadores y se realizaron mediciones de las variables pH, Xv y <sup>0</sup>Brix de acuerdo a las normas cubanas aprobadas (NC 709, 2009; NC 711, 2009, Vázquez et al.,2012). La temperatura se midió mediante un termómetro Alla France de acero inoxidable ( $\pm 1^{\circ}\text{C}$ ) directamente en el fermentador. El valor de cEt solo se midió cuando la fermentación se detuvo (NC 290; 2007).

**Tabla 1:** Intervalos de las variables de estudio

<b>Variables</b>	<b>Máximo</b>	<b>Mínimo</b>
Brix (g sólidos solubles/g miel diluida)	14,3	11,2
pH	4,1	3,5
T (°C)	34	30
Xv (millones de células/mL)	168	90
c (Et) (mL etanol/mL vino de fermentación)	5,9	4,9

Se decidió añadir a la matriz inicial de datos la variable categórica "rendimiento" con los niveles medio y bajo asociados a la cEt y a indicadores productivos de procesos de fermentación alcohólica en destilerías que utilizan miel final como sustrato sin sistema de enfriamiento (Pérez, 2011). La adición de esta variable permite la identificación de condiciones iniciales favorables para alcanzar altos rendimientos en la etapa de fermentación.

### Transformación de los datos

La transformación previa de los datos consistió en el centrado y el escalado de los datos ya que las variables tienen diferentes magnitudes, con el propósito de que todas las columnas tengan media cero (centrado) y la misma varianza (escalado). Para el centrado, el programa utiliza la Ecuación 1 y para el escalado, la Ecuación 2. Con el centrado la información sobre el origen se pierde, pero las distancias entre los puntos permanecen inalteradas. El autoescalado elimina cualquier peso indeseado debido a la diferencia de magnitud entre las variables o de las muestras dándoles a todas las mismas posibilidades de competir; además garantiza que el ACP se realice con la matriz de correlaciones (Jolliffe & Cadima, 2016).

$$x_{(i,k)c} = x_{i,k} - |\bar{x}_i| \quad (1)$$

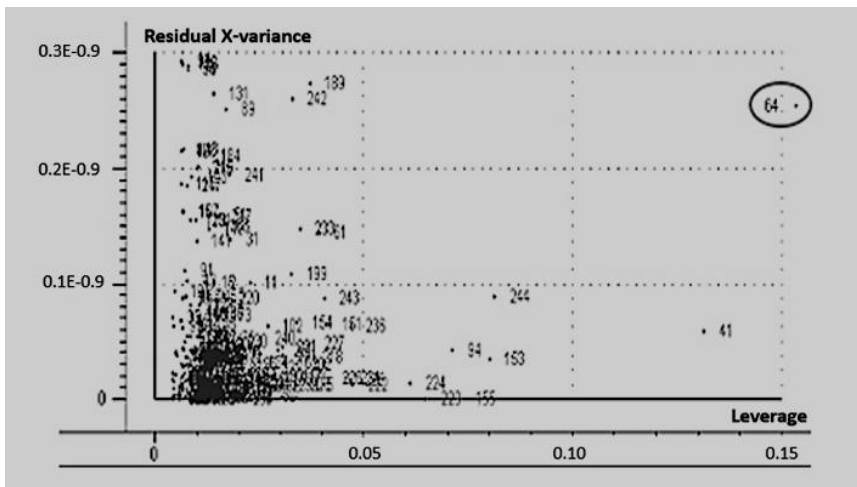
$$x_{(i,k)s} = \frac{x_{i,k}}{\sigma} \quad (2)$$

**Análisis de componentes principales**

Se utilizó el ACP para el conjunto de variables Xv, pH, T y °Brix para la descripción del proceso fermentativo. El ACP se realizó con el software Unscrambler v8.0 por sus facilidades en el uso y excelentes capacidades de representación gráfica (Hernández & Rodríguez, 2013).

**RESULTADOS**

El ACP se realizó partiendo de un fichero de datos para 249 muestras, con las variables temperatura (T), concentración de iones hidrógeno (pH), concentración de células vivas (Xv) y °Brix (Brix). El gráfico de la influencia se muestra en la Figura 1.

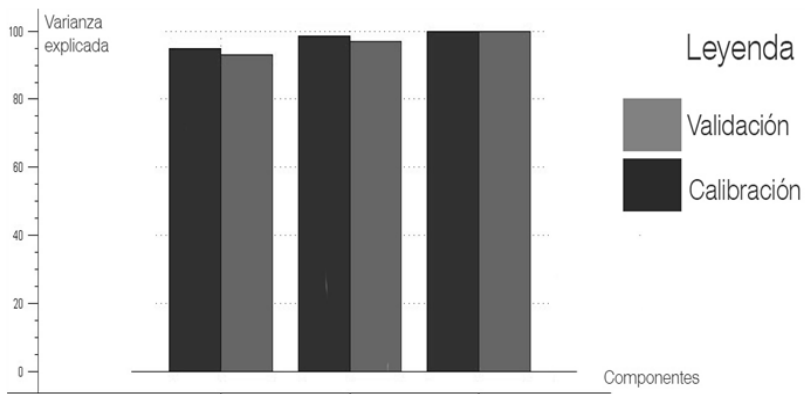


**Figura 1:** Gráfico de la influencia.

En el gráfico de la influencia, los puntos con elevado error residual se consideran discrepantes, pues quedarían pobremente descritos por el modelo resultante. Por otra parte, muestras con altos valores de distanciamiento tienen gran influencia en el modelo y se deben analizar profundamente pues pudieran ser discrepantes o representar información importante y valiosa que no sería recomendable obviar. El tipo de dato discrepante más peligroso es el que tiene altos de valores de distanciamiento y error residual y se debe eliminar del análisis.

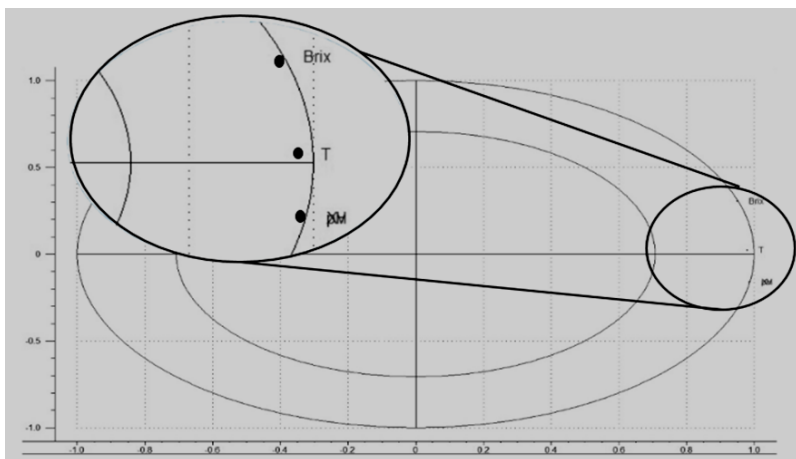
La Figura 2 muestra el gráfico de la varianza explicada. La varianza de calibración se basa en adaptar los datos de calibración al modelo. La varianza de validación, en este caso utilizando validación cruzada, se computa probando el modelo con datos que no se usaron para construir el modelo. Si se comparan las dos curvas y difieren significativamente, hay buena razón para dudar de la

calidad de los datos. Esto quiere decir que, aunque el modelo se ajuste bien a los datos de calibración, no describe bien nuevos datos.



**Figura 2:** Gráfico de la varianza explicada.

El gráfico de los pesos correlacionados, Figura 3, muestra las relaciones entre las variables y la influencia sobre cada componente. Este gráfico contiene dos elipses que indican cuánta varianza se explica. La elipse exterior indica el 100% de la varianza explicada y la elipse interior indica el 50 % de la varianza explicada. La importancia de cada variable en cada componente principal se visualiza con mayor claridad en este gráfico.



**Figura 3:** Gráfico de los pesos correlacionados.

En la Tabla 2 se encuentran los pesos de cada una de las variables en los dos primeros componentes principales.

**Tabla 2:** Pesos de las variables en los dos primeros componentes principales

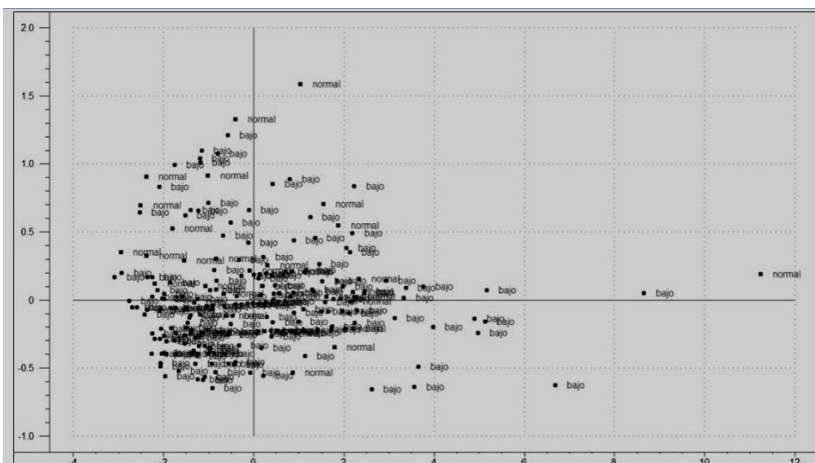
Variables	CP1	CP2
Brix	0,487	0,803
T	0,502	0,065
pH	0,506	-0,419
Xv	0,506	-0,419

El primer componente que permite explicar un 94,92% de la varianza total puede representarse por la Ecuación 3. Todos los pesos del primer componente tienen el mismo signo y por tanto representa un promedio ponderado de todas las variables.

$$CP1 = 0,487Brix + 0,502T + 0,506pH + 0,506Xv \quad (3)$$

La ecuación se pudiera utilizar para la estimación de la concentración de etanol dados los valores iniciales de las variables Brix, T, pH y Xv mediante modelos de regresión simple o técnicas de inteligencia artificial. La aplicación de este modelo se restringe a los intervalos de las variables estudiadas que se presentaron en la Tabla 1.

Las proyecciones de las muestras en el nuevo espacio representado por los componentes principales se pueden observar en el gráfico de las puntuaciones. Si los dos primeros componentes principales explican la mayor parte de la varianza, el gráfico de las puntuaciones resulta una proyección confiable en dos dimensiones de la distribución espacial de la muestra inicial multidimensional. En la Figura 4 se muestra el gráfico de las puntuaciones. Examinar el gráfico de las puntuaciones permite identificar agrupamientos que no se detectan a simple vista.



**Figura 4:** Gráfico de las puntuaciones.

## DISCUSIÓN

La Figura 1 permite apreciar que la muestra 64 tiene elevados valores de error residual y distanciamiento y debe por tanto ser eliminada del análisis. En esta muestra la concentración de células vivas ( $X_v$ ) presenta un valor bajo. La  $X_v$  es una variable que se determina mediante análisis de laboratorio y este comportamiento puede estar relacionado con un error en su medición o con el pase del pie de fermentación al fermentador antes de tiempo. En cualquiera de los dos casos anteriores no resulta interesante el análisis de esta muestra en particular pues este tipo de situaciones no son habituales en la destilería. Por tanto, la condición inicial representada por la muestra 64 no debe reflejarse en el modelo de ACP. Una vez eliminada la muestra 64 se repitió el ACP.

Dos componentes principales logran explicar el 98,55 % de la varianza de los datos iniciales (ver Figura 2). El primer componente explica el 94,92 % de la varianza mientras que el segundo explica 3,63 % de la varianza.

Las barras que representan la varianza explicada de calibración y validación tienen un comportamiento muy parecido por lo que este modelo no solo describe bien los datos de calibración, sino que también describe bien nuevos datos.

Como se puede observar en la Figura 3 todas las variables se encuentran entre las dos elipses, muy cercanas a la elipse exterior, por lo que todas son significativas ya que tienen un elevado aporte al primer componente que logra explicar el 94,92 % de la varianza total inicial. Por tanto, las cuatro variables (pH, T, °Brix y  $X_v$ ) se deben tomar en cuenta en cualquier descripción del proceso que se pretenda realizar.

El hecho de que con dos componentes se explique el 98,55 % de la varianza demuestra que la proyección en dos dimensiones de los datos iniciales (gráfico de las puntuaciones) refleja en gran medida la distribución real de los datos.

Se utilizó el gráfico de las puntuaciones con el objetivo de identificar qué condiciones iniciales de fermentación garantizaban obtener una concentración de etanol al final de dicha etapa acorde a indicadores operacionales propuestos por Pérez (Pérez, 2011). Para destilerías sin sistema de enfriamiento en los fermentadores el grado alcohólico al final de la etapa de fermentación debe estar en el intervalo de 5,5 a 6,5% (v/v). Como puede observarse en la Figura 4 se obtienen rendimientos catalogados como normales y bajos para las condiciones iniciales en que se opera la destilería, y en muy pocos casos rendimientos evaluados como altos.

Mesa (Mesa & Gozá, 2018) y Valles (Valles, 2015) han aplicado ACP en proceso fermentativos con células de mamíferos para mejorar la comprensión del proceso y explicar las fuentes principales de variabilidad. En ambos casos el análisis partía de un modo de operación de los reactores (continuo con recirculación de células) diferente al del presente estudio y; se tuvieron en cuenta dada la alta complejidad que implica el trabajo con células modificadas genéticamente, además de temperatura y pH, otras variables como: el flujo de aire, la velocidad de agitación, la velocidad específica de crecimiento, la presión, el flujo de medio de cultivo, viabilidad celular y tiempo del cultivo. Los resultados evidencian la alta importancia que tienen las variables temperatura, pH, flujo de medio de cultivo y viabilidad celular en los dos estudios para la descripción de la fermentación en función de la variabilidad del proceso.

Analizando los resultados de estos dos trabajos y la investigación en cuestión es evidente la alta significación de factores como la temperatura y el pH para los dos modos de operación (discontinuo y continuo con recirculación de células) y, de factores relacionados con el sustrato (flujo de medio y

concentración de total de sólidos disueltos), y factores relacionados con la cantidad y estado fisiológico de las células (concentración de células vivas y viabilidad celular).

Las variaciones de las variables durante el transcurso del proceso de fermentación reflejan los procedimientos operacionales como: el momento de realizar los incrementos de melaza al fermentador y la adición de nitrógeno, factores de gran importancia. En las hojas de fermentación no se registraba la variación de las variables durante el tiempo. Resultaría muy útil complementar el análisis anterior con la incorporación del estudio del procedimiento de operación que actualmente se desarrolla en la destilería y cómo influye este en el rendimiento de la etapa de fermentación.

## CONCLUSIONES

Mediante el análisis de componentes principales se realizó la descripción de la etapa de fermentación alcohólica en función de las variables operacionales que se miden en el proceso. La construcción de dos componentes principales que explican el 98,551% de la variabilidad total contribuyó a la proyección, con gran similitud, en sólo dos dimensiones del comportamiento de este proceso multidimensional atendiendo al gran número de variables que intervienen.

La proyección del proceso en dos dimensiones no evidenció agrupamiento en las condiciones iniciales de las variables estudiadas. Se comprobó que actualmente todas las combinaciones posibles de las variables T, pH, Brix y Xv pueden resultar en fermentaciones con rendimiento normal y bajo. Por tanto, no pueden establecerse intervalos idóneos para las condiciones iniciales en el fermentador dentro de todas las estudiadas en este trabajo.

En ninguna de las fermentaciones estudiadas se obtuvo un valor de  $c(Et)$  mayor que 5,9. Las condiciones actuales de la destilería permiten potencialmente la obtención de valores de  $c(Et)$  de hasta 6,5 según la bibliografía consultada (Pérez, 2011). La ampliación en futuros estudios de los intervalos de las variables analizadas y la inclusión de los procedimientos operacionales se presentan como alternativas atractivas para incrementar el rendimiento de la fermentación alcohólica en la destilería Héctor Molina.

## REFERENCIAS BIBLIOGRÁFICAS

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *John Wiley & Sons, Inc.*, 2, 433-459.
- Aceves-Lara, C. A., Latrille, E., Buffiere, P., Bernet, N., & Steyer, J.-P. (2008). Experimental determination by principal component analysis of a reaction pathway of biohydrogen production by anaerobic fermentation. *Chemical Engineering and Processing*, 47, 1968-1975.
- Buratti, S., & Benedetti, S. (2016). *Alcoholic Fermentation Using Electronic Nose and Electronic Tongue Electronic Noses and Tongues in Food Science* (pp. 291-299): Elsevier.
- Epifanio, S. (2005). *Influencia de la tecnología de vinificación en la microbiología y el desarrollo de la fermentación alcohólica*. Tesis Doctoral. Universidad de la Rioja, España.
- García, N. G., & Londoño, A. T. (2015). *Análisis de Componentes Principales SPARSE. Formulación, algoritmos e implicaciones en el análisis de datos*. (Máster en Análisis avanzado de datos multivariantes Trabajo fin de Máster), Universidad de Salamanca.
- Gunther, J. C., Seborg, D. E., & Baclaski, J. (2006). *Fault Detection and Diagnosis in Industrial Fed-Batch Fermentation*. Paper presented at the 2006 American Control Conference, Minneapolis, Minnesota, USA.
- Jolliffe, I., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions*. doi: 10.1098/rsta.2015.0202



- Lee, J.-M., Yoo, C., & Lee, I.-B. (2004). Fault detection of batch processes using multiway kernel principal component analysis. *Computers & chemical engineering*, 28(9), 1837-1847.
- Menéndez, Z. (2010). *Desarrollo de módulos de cálculo para los procesos de fermentación alcohólica*. (Tesis de maestría en Análisis y Control de Procesos), Universidad Tecnológica de La Habana "José Antonio Echeverría" CUJAE.
- Mesa, L., & Gozá, O. (2018). Aplicación del Análisis de Componentes Principales en el proceso de fermentación de un anticuerpo monoclonal. *VacciMonitor*, 27(1), 8-15.
- NC 290: 2007: Norma Cubana Bebidas alcohólicas- Determinación del grado alcohólico en alcoholes, bebidas alcohólicas destiladas, vinos, licores, bebidas alcohólicas preparadas, cocteles y extractos hidroalcohólicos. Oficina Nacional de Normalización. 2007.
- NC 709: 2009: Norma Cubana MIEL FINAL — Determinación de sólidos aerométricos disueltos. Oficina Nacional de Normalización. 2009.
- NC 711: 2009: Determinación potenciométrica del pH. Oficina Nacional de Normalización. 2009.
- Pérez, O. (2011). *Modelación, simulación y análisis con fines energéticos de destilerías de etanol hidratado*. (Tesis de doctorado en Ingeniería Química), Universidad Tecnológica de La Habana "José Antonio Echeverría" CUJAE.
- Tharwat, A. (2016). Principal component analysis-a tutorial. *International Journal of Applied Pattern Recognition*, 3(3), 197-240.
- Valles, A. (2015). *Análisis del proceso de fermentación en la planta de EPOhr aplicando técnicas de Minería de Datos*. (Tesis de maestría en Análisis y Control de Procesos), Universidad Tecnológica de La Habana "José Antonio Echeverría" CUJAE.
- Vázquez, M., Cuellar, H., García, R., Díaz, M. E., Reyes, A., González, M. D., Ibáñez, M.-L. (2012). Técnicas microbiológicas. *Manual de técnicas analíticas para destilerías*. Editorial ICIDCA. (pp. 78-85).
- Zamora, F. (2009). Biochemistry of alcoholic fermentation *Wine chemistry and biochemisry*. New York: Springer.
- Zayas, E. P., Lorenzo, M., Hernández, D., & Blanco, I. (2014). Clasificación multivariante de ronnes añejos cubanos. *Revista Cubana de Ingeniería*, 5, 62-67.