

Modelado borroso de una fermentación alcohólica a partir de la *Zymomonas mobilis**

Fuzzy modelling of an alcoholic fermentation with Zymomonas mobilis

M.Sc. Boris Luis Martínez Jiménez, Dr.C. Francisco Herrera Fernández

Dpto. de Automática y Sistemas Computacionales. Facultad de Ingeniería Eléctrica. UCLV, boris@uclv.edu.cu

Resumen — Este artículo presenta el método seguido para la obtención de un modelo borroso dinámico de un proceso de fermentación bacteriana de producción de bioetanol. Específicamente, este proceso trabaja con la bacteria *Zymomonas mobilis*. La fermentación alcohólica con esta bacteria presenta altos rendimientos de etanol, pero con comportamientos altamente complejos. Desde el punto de vista de la identificación y el control esto representa una gran dificultad, por lo que obtener un modelo preciso y adecuado es de gran importancia.

La metodología de identificación propuesta combina métodos de agrupamiento y de reducción de reglas para aproximar un sistema complejo por medio de un conjunto de modelos lineales locales, ya que se utiliza el modelado borroso basado en el modelo de Takagi-Sugeno (TS) donde los consecuentes son precisamente las funciones lineales.

Con el modelo del proceso presentado en este trabajo es posible capturar la dinámica del proceso de obtención de bioetanol mediante *Zymomonas mobilis*, siendo además compacto y con una estructura deseable para su aplicación futura en varias estructuras de control basado en modelos, en aras de optimizar la productividad del mismo y lograr una disminución en el costo de producción.

Palabras Claves — Modelo borroso Takagi-Sugeno, agrupamiento, reducción de reglas, fermentación alcohólica, bioetanol, *Zymomonas mobilis*.

Abstract — This article presents the followed method for the obtaining of a dynamic fuzzy model that represents a bacterial fermentation process of bioethanol production. Specifically, this process works with *Zymomonas mobilis* bacterium. The alcoholic fermentation with this bacterium presents high yields of ethanol, but with highly complex behaviors. From the identification and control point of view, this represents a great difficulty, and to obtain a precise and suitable model is very important.

The proposed identification methodology combines clustering and rules reduction methods to approximate a complex system by means of a set of local linear models, since the fuzzy modeling used is based on the Takagi-Sugeno (TS) model where the consequent ones are in fact the lineal functions.

With the presented process model in this work is possible to capture the dynamic of the process of bioethanol obtaining by means of *Zymomonas mobilis*, and in addition, the model is compact and with a desirable structure for its future application in several model based control structures, for the sake of optimizing the process productivity and to achieve a diminution in the production cost.

Keywords — Takagi-Sugeno fuzzy model, clustering, rules reduction, alcoholic fermentation, bioethanol, *Zymomonas mobilis*

INTRODUCCIÓN

La fabricación de alcohol como combustible alternativo para sortear los altos precios del petróleo, despierta actualmente un gran interés en muchas naciones de la región, ya que este es más barato y compatible con el medio ambiente, unido a los excelentes resultados obtenidos en varios países con Brasil a la cabeza. Además, son numerosísimas las demás aplicaciones del etanol. Entre ellas se destacan las bebidas, desinfectantes, disolventes y como materia prima para la obtención de gran variedad de compuestos. Por otro lado, el desarrollo ambientalmente compatible de la industria azucarera como productora no sólo de azúcar, sino también de materias primas y portadores energéticos como bioetanol y electricidad, es un imperativo de estos tiempos. La posibilidad de lograr,

* Parcialmente financiado por el proyecto MES 6.111“Aplicación de técnicas inteligentes en procesos de biotecnología”

en determinadas instalaciones de la industria azucarera, niveles de generación de etanol que satisfagan varias necesidades, tiene una relevancia económica estratégica. El bioetanol, además de ser una apuesta fuerte para el futuro de los combustibles, se puede considerar un tema económico, científico, medio ambiental y social, es decir, una primera plana en el conocimiento actual.

Hoy en día se conoce una gran variedad de especies capaces de producir bioetanol, principalmente levaduras y bacterias. Entre las levaduras ha adquirido gran importancia la *Saccharomyces cerevisiae*. Entre las bacterias se destaca la denominada *Zymomonas mobilis*, la cual tiene mayor productividad y tolerancia a etanol que la *Saccharomyces cerevisiae*. Sin embargo, unido a que los bioprocesos por lo general son altamente no lineales y complejos, la fermentación con *Zymomonas mobilis* presenta comportamientos altamente oscilatorios en las variables del proceso [1]. Desde el punto de vista de la identificación y el control esto representa una gran dificultad, por lo que la obtención de un modelo preciso y adecuado adquiere gran importancia. A esto se une que no existen muchos modelos matemáticos de obtención de bioetanol a partir de bacterias ya que la mayor parte se produce a partir de levaduras, y por tanto, la mayoría de los modelos matemáticos de fermentación alcohólica existentes se basan en datos experimentales obtenidos con este último tipo de organismos.

Dentro de los métodos de identificación, la modelación borrosa es uno que ofrece una forma muy estructurada de formulación del conocimiento, capaz de proporcionarlo a partir de datos e incorporarlo a partir de expertos. Otra propiedad importante de los modelos borrosos es su capacidad de representar sistemas no lineales [2, 3]. De esto se deriva que es provechosa su aplicación a plantas no lineales y difíciles de modelar como es el proceso que nos ocupa.

Para modelar, generalmente se utiliza el razonamiento, las leyes físicas, químicas, etc (modelado de caja blanca) o la experimentación y el análisis de datos (modelado de caja negra); sin embargo, en los últimos años se vienen aplicando una combinación de ambas técnicas, conocida como modelado de caja gris o modelado híbrido [3, 4].

El presente trabajo enfrenta el reto de modelar un proceso de fermentación alcohólica en el que se usa un microorganismo con altas complejidades cinéticas, lo que hace que en el sistema estén presentes dinámicas altamente no lineales de características difíciles de modelar. Se propone para ello utilizar sistemas borrosos Takagi-Sugeno dentro de los dos últimos tipos de modelado mencionados en el párrafo anterior. A continuación se presentarán las características de la fermentación alcohólica con *Zymomonas mobilis*, luego se expone la metodología de identificación borrosa utilizada. Finalmente, se realiza el análisis de los resultados y se presentan las conclusiones.

DESARROLLO

Fermentación alcohólica con Zymomonas mobilis

La *Zymomonas mobilis* se ha convertido en objeto creciente de estudio en estos años debido a un conjunto de ventajas potenciales las cuales se resumen a continuación:

1. tolerancia a concentraciones de azúcar y de etanol superiores a las de las levaduras
2. mayor velocidad de crecimiento,
3. reducción significativa de la cantidad de sustrato que se convierte en biomasa en lugar de etanol (sólo del 2-2,6 %) si se compara con las levaduras,
4. alcanza mayor productividad máxima

Además, la *Zymomonas* es la única especie de bacteria capaz de utilizar la vía de Entner-Doudoroff anaeróbicamente y por tanto sólo produce un mol de ATP por mol de sustrato utilizado. Pero acompañado a estos beneficios, exhiben el ya mencionado comportamiento no lineal y oscilatorio [1].

Teniendo en cuenta dicho comportamiento y sus altos rendimientos de etanol comparados con los de otros microorganismos, es de vital importancia entonces encontrar un modelo que capture la dinámica del proceso de obtención de bioetanol mediante *Zymomonas mobilis*, y que además, sea compacto y pueda ser aplicado en varias estructuras de control. De esta forma se busca maximizar el rendimiento de la bacteria y por ende, potenciar su utilización como productora de etanol.

Modelo matemático de la fermentación con *Zymomonas mobilis*

Un modelo satisfactorio debe cumplir dos requerimientos contradictorios: debe ser lo suficientemente detallado como para representar al sistema real con exactitud y a su vez, lo bastante sencillo como para permitir un análisis matemático práctico.

En los cultivos microbianos, la monografía de Monod en 1942 marca un hito ya que el autor demostró que el crecimiento bacteriano puede formularse en términos de rendimiento de crecimiento, velocidad específica de crecimiento y concentración del sustrato. La teoría matemática fue tomada por analogía con los procesos químicos, pero estas ecuaciones no abarcan a los procesos biológicos en su totalidad. Las diferencias parten de que en las reacciones químicas todos los iones o moléculas se consideran similares, en cambio, en los sistemas biológicos las células tienen distintas “edades”: mientras unas están “naciendo”, otras están en “máxima actividad” o “muriendo”.

En tal sentido se han propuesto muchos modelos matemáticos para describir la producción de etanol. En los modelos cinéticos en los cuales las variables de estado son las concentraciones de biomasa (X), de sustrato (S) y del producto etanol (P), el modelado de la velocidad de crecimiento de la biomasa (μ) y de la velocidad de producción de etanol (r_p) son los pasos más importantes y aún no existe una solución aceptada por todos. En el presente trabajo se asume que μ responde a un modelo de inhibición que involucra al producto y al sustrato [5]. El modelo matemático está dado por:

$$\frac{dX}{dt} = r_x = \mu X \quad (1)$$

$$\frac{dS}{dt} = -r_s = -\left(\frac{\mu}{Y_{XS}} + m\right)X \quad (2)$$

$$\frac{dP}{dt} = r_p = (\alpha\mu + \beta)X \quad (3)$$

$$\mu(S, P) = \mu_{max} \frac{S}{K_S + S} \left(1 - \frac{P}{P_m}\right)^n \quad (4)$$

donde X , S y P son, respectivamente, las concentraciones de biomasa, de sustrato y del producto etanol (g/l); r_x , r_s y r_p son, respectivamente, las velocidades de crecimiento celular, de consumo de sustrato y de aparición del producto ((g/l)/h), μ es la velocidad específica de crecimiento (1/h), Y_{XS} es la constante de rendimiento celular (g_{bio}/g_{sus}), m es la constante de mantenimiento ($(g_{sus}/g_{bio})/h$), α expresa la porción del etanol producido asociado con el crecimiento de las células (g_{eta}/g_{bio}), β expresa la porción del etanol producido para el mantenimiento de las células ($(g_{eta}/g_{bio})/h$), μ_{max} es la razón de crecimiento máximo (1/h), K_S es la constante de saturación (g_{sus}/l), P_m es la máxima concentración de etanol (g_{eta}/l) y n es la constante de toxicidad del etanol. Ver los valores de los parámetros en [5].

Sistemas Borrosos Takagi-Sugeno (TS)

Un modelo borroso conveniente para la aproximación de una gran cantidad de sistemas dinámicos no lineales es el conocido sistema borroso Takagi-Sugeno [6]. Estos son populares en el área de la identificación debido a que pueden aproximar cualquier función con cualquier grado de exactitud y combinan la modelación lingüística y la matemática. En estos sistemas el antecedente describe regiones borrosas del espacio de entrada donde un consecuente funcional lineal es válido. Además, desde el punto de vista de la implementación, los modelos borrosos Takagi-Sugeno (TS) son simples y computacionalmente eficientes [2]. En este tipo de modelo TS con funciones lineales de las entradas como consecuentes, las reglas se presentan en la siguiente forma:

$$R_i: \text{ IF } x_1 \text{ is } A_{i1} \text{ and...and } x_p \text{ is } A_{ip} \text{ THEN } y_i = \mathbf{a}_i^T \mathbf{x} + b_i, \quad i = 1, 2, \dots, L \quad (5)$$

donde R_i denota la regla i , L es el número de reglas del modelo, $\mathbf{x} = [x_1, \dots, x_p]^T$ es el vector de entrada, A_{ij} son conjuntos borrosos caracterizados por las funciones de pertenencia $\mu_{A_{ij}}(x_i)$, y_i es la salida de la i -ésima regla, $\mathbf{a}_i = [a_{i1}, \dots, a_{ip}]$ es un vector de parámetros y b_i es un escalar. La salida de este sistema borroso es un promedio ponderado de los valores de los consecuentes de las reglas. La fórmula de inferencia es:

$$y = \left(\sum_{i=1}^L w_i(\mathbf{x}) y_i \right) / \sum_{i=1}^L w_i(\mathbf{x}) = \left(\sum_{i=1}^L w_i(\mathbf{x}) (\mathbf{a}_i^T \mathbf{x} + b_i) \right) / \sum_{i=1}^L w_i(\mathbf{x}) \quad (6)$$

donde y es la salida total del sistema y w_i es el nivel de impacto de la regla i , el cual se genera utilizando como operador AND el producto, de forma que:

$$w_i(\mathbf{x}) = \mu_{A_{i1}}(x_1) \times \mu_{A_{i2}}(x_2) \times \dots \times \mu_{A_{ip}}(x_p) \quad (7)$$

Definiendo el nivel de impacto normalizado como

$$w_i^*(\mathbf{x}) = w_i(\mathbf{x}) / \sum_{i=1}^L w_i(\mathbf{x}) \quad (8)$$

y sustituyéndolo en la ecuación 6, se obtiene que

$$y = \left(\sum_{i=1}^L w_i^*(\mathbf{x}) \mathbf{a}_i^T \right) \mathbf{x} + \sum_{i=1}^L w_i^*(\mathbf{x}) b_i = \mathbf{a}^T(\mathbf{x}) \mathbf{x} + b(\mathbf{x}) \quad (9)$$

En la ecuación 9 se aprecia que el modelo TS puede expresarse como un modelo lineal con parámetros dependientes de las entradas. Esta característica permite el análisis en un marco similar al de los sistemas lineales.

Metodología de identificación propuesta

Para formar particiones borrosas se utilizan dos procedimientos. El primero comienza con una partición y se van adicionando más particiones basados en algún criterio, como es el caso de los métodos de agrupamiento. En el modelado borroso, al aplicar estos algoritmos al conjunto de datos se determina la localización de las funciones de pertenencia de las reglas borrosas. En la realización de este trabajo se utilizan métodos de agrupamiento simples y efectivos propuestos para reducir el alto costo computacional presente en otras técnicas de este tipo, a saber, el agrupamiento sustractivo [7], el método de agrupamiento evolutivo ECM [8], y el ECM modificado [9]. Una vez determinado los parámetros del antecedente, los parámetros de los consecuentes se determinan por medio de los mínimos cuadrados.

La segunda proposición para formar particiones borrosas comienza con un número sobredimensionado de particiones borrosas y luego se eliminan las menos significativas como es el caso de los métodos de transformación ortogonal basados en la Descomposición en Valores Singulares (*SVD*). Aquí se utiliza el algoritmo SVD-QR con pivoteo de columna [10]. Este método se basa en que el antecedente de un modelo borroso resulta de la partición borrosa del espacio de entrada. De aquí que se construya una matriz de los niveles de impacto normalizados, donde cada columna se corresponde con una de las particiones. Si una columna es linealmente dependiente o todos sus elementos son ceros, entonces la partición asociada es redundante, y su eliminación no afectará significativamente el desempeño del modelo analizado. Matemáticamente, una columna cero o linealmente dependiente hará que la matriz tenga valores singulares con valor cero. Entonces, analizando los valores singulares de la matriz podremos determinar las particiones redundantes o irrelevantes [10].

En los métodos de agrupamiento, se aplican varios procedimientos para la determinación del número C de agrupamientos (*clusters*), como las medidas de validez y la combinación de agrupamientos semejantes [2]. Pero por lo general, la selección del número “correcto” de *clusters* se realiza de forma iterativa con distintos valores de C y/o con diferentes inicializaciones para igual número de *clusters*. Nótese entonces que estos procedimientos generalmente requieren elevados recursos de cómputo debido a su carácter iterativo.

Para paliar esta problemática, se propone una metodología que combina los procedimientos antes expuestos. Esta consiste en comenzar aplicando un algoritmo de agrupamiento con un número de *clusters* mayor que el número ‘óptimo’ para el problema en cuestión. Cada *cluster* encontrado se corresponde con una regla en el modelo. Luego, mediante el algoritmo de transformación ortogonal, se eliminan las particiones redundantes o menos significativas. Finalmente, se “reajusta” el modelo.

En resumen, el método propuesto consiste en:

Paso 1: Aplicar un algoritmo de agrupamiento con un número de *clusters* mayor que el número ‘óptimo’ para el problema en cuestión. En nuestro trabajo el modelo inicial se obtiene aplicando alguna de las tres técnicas de agrupamiento sencillas mencionadas anteriormente.

Paso 2: Aplicar el algoritmo SVD-QR para eliminar las particiones redundantes o no importantes.

Paso 3: Ajuste “fino” de los parámetros del modelo por alguna de las técnicas conocidas, como el ANFIS implementado en MATLAB.

Esta metodología busca reducir el alto costo computacional de varios procedimientos de cálculo del número de *clusters*, y por ende, del número de reglas. Además, mediante el análisis de los valores singulares se tiene también idea de cuáles son las particiones más importantes. En la próxima sección se aplica el método propuesto en el proceso fermentativo expuesto en este trabajo.

RESULTADOS Y DISCUSIÓN

En la introducción se describen las formas generales de modelar, dentro del modelado de caja blanca estaría el modelo descrito por las ecuaciones (1)-(4). En lo adelante se trabajarán los casos del modelado de caja negra y el modelado de caja gris o modelado híbrido.

Modelo global (modelo de caja negra)

A continuación se explica la obtención del modelo dinámico del etanol a partir de datos por ser esta la variable de mayor interés pero para el resto de las variables se trabaja de forma análoga.

Aquí se adopta un modelo TS con una estructura de primer orden y dinámico, ya que la salida del sistema borroso aparece como una de sus entradas. La forma de cada regla queda:

$$\text{IF } X(k) \text{ is } A_{i1} \text{ and } S(k) \text{ is } A_{i2} \text{ and } P(k) \text{ is } A_{i3} \text{ THEN } P_i(k+1) = a_{i1}X(k) + a_{i2}S(k) + a_{i3}P(k) + b_i \quad (10)$$

donde X (biomasa), S (sustrato) y P (etanol) constituyen las variables de estado. La salida del sistema borroso, $P(k+1)$, es un promedio ponderado de la salida de cada regla y se calcula por (6).

El modelo inicial con un número de *clusters* sobredimensionado se obtiene aplicando alguna de las técnicas de agrupamiento mencionadas en la sección anterior. Los resultados obtenidos con cada una fueron similares, aunque aquí se mostrarán los resultados del agrupamiento sustractivo. En este caso se obtiene un modelo inicial con nueve reglas. Dicho modelo es reducido mediante la aplicación del algoritmo SVD-QR, obteniéndose un modelo con cinco reglas y suficiente exactitud. Esta reducción implica menos parámetros a actualizar durante el entrenamiento, más rapidez en el cálculo y por tanto, un costo computacional menor. La figura 1 muestra los resultados de estimación obtenidos con el modelo reducido.

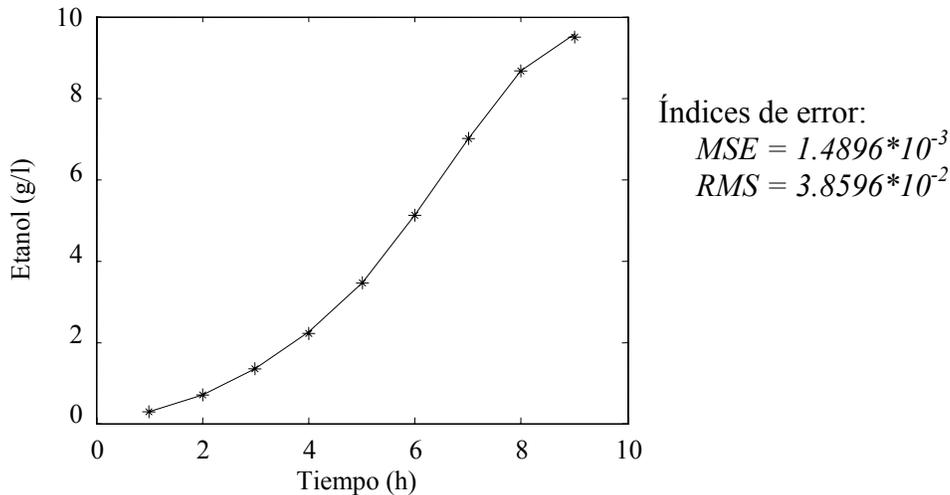


Fig. 1. Estimación del producto etanol. Datos experimentales (*), valores generados por el modelo (-). Las condiciones iniciales son: $X_0 = 0.03$ g/l, $S_0 = 20$ g/l y $P_0 = 0$ g/l

Modelo Híbrido (modelo de caja gris)

Si previamente se conoce la relación ‘correcta’ de la razón de crecimiento microbiano (μ), un modelo con suficiente exactitud puede obtenerse combinando dicha relación con un modelo macroscópico descrito por ecuaciones de balance. En la práctica, dicha relación ‘correcta’ generalmente no es conocida a priori y la identificación del parámetro se dificulta. En este trabajo se adopta un modelo borroso TS para la estimación del parámetro μ a partir de datos, el cual describe la dependencia de este parámetro de las variables de estado. Dicho modelo se incorpora después en el modelo descrito por las ecuaciones (1), (2) y (3).

La estructura del modelo TS en este caso también es de primer orden y cada regla está dada por:

$$\text{IF } X(k) \text{ is } A_{i1} \text{ and } S(k) \text{ is } A_{i2} \text{ and } P(k) \text{ is } A_{i3} \text{ THEN } \mu_i(k+1) = a_{i1}X(k) + a_{i2}S(k) + a_{i3}P(k) + b_i \quad (11)$$

Al aplicar el método propuesto se obtiene un modelo inicial con siete reglas y un modelo reducido con cuatro reglas. Los resultados de la estimación del parámetro μ se muestran en la figura 2 y en la figura 3 se grafican las estimaciones de las variables de estado. Se hace notar que para el caso del producto etanol, se obtienen los índices de error $MSE = 3.8821 * 10^{-3}$ y $RMS = 6.2306 * 10^{-2}$.

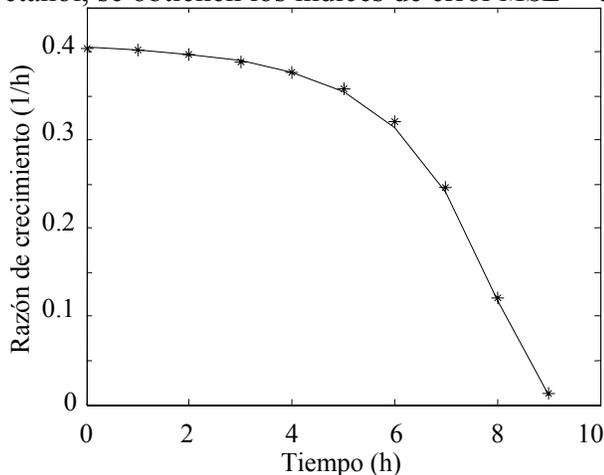


Fig. 2. Estimación de la razón de crecimiento μ . Datos experimentales (*), valores generados por el modelo (-). Las condiciones iniciales son: $X_0 = 0.03$ g/l, $S_0 = 20$ g/l y $P_0 = 0$ g/l

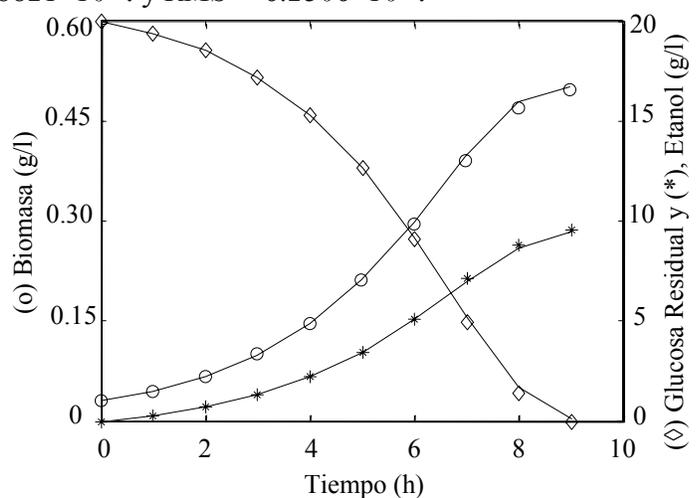


Fig. 3. Estimaciones de la biomasa (o), la glucosa residual (\diamond) y el producto (*). Los símbolos representan los datos experimentales y las líneas los valores generados por el modelo. Las condiciones iniciales son: $X_0 = 0.03$ g/l, $S_0 = 20.0$ g/l y $P_0 = 0.0$ g/l

CONCLUSIONES

En este trabajo se propone una estrategia que combina varias técnicas para la modelación borrosa de un bioproceso, con vistas a reducir el tiempo de obtención de los modelos y reducir la dependencia computacional. Específicamente, se utilizan métodos de agrupamiento sencillos y eficientes para obtener un modelo inicial y luego se aplica a este modelo el algoritmo SVD-QR en aras de reducir la complejidad de dicho modelo.

Aplicando esta metodología se obtuvieron dos modelos borrosos TS a partir de datos, uno para la predicción de las variables de estado de un modelo global del sistema y otro para la estimación de un parámetro que luego se combina con ecuaciones de balance en una estrategia de modelación híbrida. La exactitud numérica de las estimaciones en ambas estrategias de modelado es buena.

Los modelos borrosos obtenidos presentan una estructura matemática simple, lo que facilita el análisis así como un posterior tratamiento en el diseño y análisis de un sistema de control para el proceso en cuestión. Estos modelos podrían utilizarse para varios objetivos como el análisis del proceso, la predicción de las variables de estado, aplicarlo en estrategias de control basado en modelos, etc.

AGRADECIMIENTOS

Al doctor Oscar Alfredo Garro de la Facultad de Agroindustrias, Universidad Nacional del Nordeste, Argentina por su colaboración desinteresada.

REFERENCIAS

1. Echeverry, N., et al. *Control de un biorreactor para fermentación alcohólica en continuo*. in *XI Congreso Latinoamericano de Control Automático CLCA'04*. 2004. Cuba.
2. Babuska, R., *Fuzzy Modelling for Control*. 1998: Kluwer Academic Publishers.
3. Roubos, H., et al. *Hybrid modeling of bioprocesses*. in *Third Annual Symposium: Mastering the Molecules in Manufacturing*. 2000. Delft, The Netherlands.
4. Franco, E., *Framework for online modeling, optimization and monitoring of bioprocesses*. 2002, Martin-Luther-Universität: Halle -Wittenberg.
5. Garro, O.A., *Fermentación alcohólica con Zymomonas sp. Estudio de modelos matemáticos y su verificación mediante ensayos de fermentación.*, in *Facultad de Bioquímica, Química y Farmacia*. 1993, Universidad Nacional de Tucumán.
6. Takagi, T. and M. Sugeno, *Fuzzy identification of systems and its applications to modelling and control*. IEEE Trans. On Systems, Man, and Cybern, 1985. **15** (1): p. 116-132.
7. Chiu, S.L., *Fuzzy Model Identification Based on Cluster Estimation*. Journal of Intelligent & Fuzzy Systems, 1994. **2**(3).
8. Song, Q. and N. Kasabov. *ECM- A Novel On-Line Evolving Clustering Method and Its Applications*. in *Proceedings of the Fifth Biannual Conference on Artificial Neural Networks and Expert Systems*. 2001.
9. Fernández, J.A. and B.L. Martínez, *Algoritmos de clusterización en línea*, in *Dpto. de Automática y Sistemas Computacionales, Facultad de Ing. Eléctrica*. 2003, Universidad Central "Marta Abreu" de Las Villas: Santa Clara, Cuba.
10. Yen, J. and L. Wang, *Simplifying fuzzy rule-based models using orthogonal transformation methods*. IEEE Trans. On Systems, Man, and Cybern, Part B, 1999. **29**: p. 13-24.