



Reparametrización iterativa heurística a un modelo dinámico de crecimiento de microalgas en un reactor *fed-batch*

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

Daniela Soto¹, Alumno de sexto año

Benjamín Sánchez¹, Alumno de sexto año

Katherine Rojas¹, Alumno de sexto año

Francisco Jeria¹, Alumno de sexto año

Claudio Gelmi¹, Profesor Asistente

Ricardo Pérez-Correa¹, Profesor Titular

¹DEPARTAMENTO DE INGENIERÍA QUÍMICA Y BIOPROCESOS

INTRODUCCIÓN

El calentamiento global es uno de los problemas más preocupantes de nuestra época [1], por lo tanto, diversas fuentes de energía alternativas a los combustibles fósiles están siendo investigadas. Una fuente prometedora son los biocombustibles de segunda generación [2, 3], en particular el biodiésel producido por microalgas [4, 5]. Sin embargo, la aún baja eficiencia en la producción le impide competir con los combustibles convencionales [6].

Predecir el comportamiento de estos cultivos a través de modelos matemáticos es una herramienta útil para aumentar la eficiencia del proceso. Para que dichos modelos tengan capacidad predictiva se

requiere que sean robustos y confiables. Esto implica que el modelo debe tener pocos parámetros, sin problemas de identificabilidad, sensibilidad o significancia. Este problema se ha abordado en la literatura [7, 8] con técnicas como análisis de sensibilidad (estudiar el impacto en el modelo de variar el valor de cada parámetro), análisis de identificabilidad (detectar parámetros que estén correlacionados entre sí), y análisis de significancia (estudiar si cada parámetro es significativamente distinto de cero). Dichas técnicas han sido denominadas análisis pre/post regresión [9], y sirven para detectar parámetros del modelo que debiesen fijarse y estimarse. Sin embargo, dichas herramientas por sí solas no entregan información acerca de cómo encontrar una combinación óptima de parámetros a fijar, por lo que se hace necesario un procedimiento genérico para determinar el grupo de parámetros fijos que entregue un modelo que no solo ajuste bien, sino que sea confiable.

En este trabajo se presenta un procedimiento heurístico iterativo para disminuir el número de parámetros en un modelo dinámico, reportado previamente en la literatura [10], de crecimiento de microalgas para producción de biodiésel en un reactor *fed-batch*. En cada paso del procedimiento se realizó una estimación de parámetros mediante optimización global meta-heurística por búsqueda dispersa, y luego se emplearon técnicas de análisis pre/post regresión para decidir qué parámetro fijar. Como en cada iteración varios parámetros pueden ser fijados, se exploraron diversas ramas considerando reglas heurísticas para reducir el espacio de búsqueda.

METODOLOGÍA

El modelo de crecimiento heterotrófico de microalgas en un reactor *fed-batch* presentado por Surisetty et al. [10] consta de seis variables de estado (Tabla 1) y 12 parámetros (Tabla 2). Además, el modelo incluye seis ecuaciones diferenciales ordinarias y tres ecuaciones algebraicas (más información en referencia [10]).

Tabla 1. Variables de estado del modelo.

x	Concentración de biomasa activa funcional [g/mL]
S_1	Concentración de la fuente de nitrógeno en medio de cultivo [g/mL]
S_2	Concentración de la fuente de carbono en medio de cultivo [g/mL]
Q	Concentración celular total de nitrógeno [g/mL]
I_p	Concentración celular total de aceite [g/mL]
V	Volumen del cultivo [mL]

Tabla 2. Parámetros del modelo.

μ_m	Tasa máxima de crecimiento [1/h]
q_m	Fracción celular de nitrógeno mínima de crecimiento [-]
K_q	Constante de saturación de la fracción celular de nitrógeno [g/mL]
ρ_m	Tasa máxima de consumo de nitrógeno [1/h]
K_s	Constante de saturación de la fuente de carbono en el crecimiento [g/mL]
s_o	Concentración límite de la fuente de nitrógeno [g/mL]
Y_{xs}	Rendimiento de fuente de carbono en biomasa [g/g]
k_m	Constante de mantenimiento [1/h]
Y_{ps}	Rendimiento de fuente de carbono en producto [g/g]
π_m	Tasa de producción máxima de aceite [1/h]
K_π	Constante de saturación de la producción de aceite [g/mL]
Y_{xq}	Rendimiento de fracción de nitrógeno en biomasa [g/g]

Para llegar a un modelo dinámico de crecimiento de microalgas en un reactor *fed-batch* con un set significativo de parámetros, se realizó un procedimiento heurístico iterativo que involucra estimación de parámetros y análisis de pre/post regresión, el cual se muestra en la Figura 1. Inicialmente se estiman los valores de cada parámetro del modelo para ajustarse a los datos experimentales. Luego, en cada iteración se fija el valor de un parámetro en el valor del paso de estimación previo. Un parámetro se decide fijar si: (i) presenta valor absoluto de su coeficiente de correlación de Pearson con otro parámetro mayor a 0,95 (falta de identificabilidad), (ii) no incide en ninguna de las variables del modelo (falta de sensibilidad), o (iii) no es significativamente distinto de cero, es decir, tiene un coeficiente de confianza menor a 2 (falta de significancia).

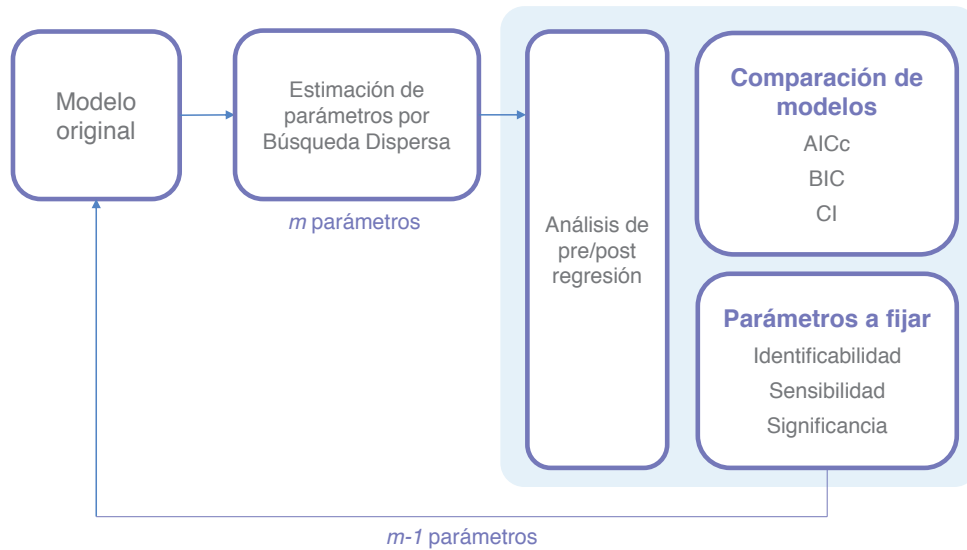


Figura 1. Ciclo iterativo del procedimiento de reparametrización.

Como en la mayoría de los casos más de un parámetro arroja al menos uno de los tres problemas mencionados, se deben explorar distintas ramas de reparametrizaciones. De esta manera, se genera un árbol exploratorio. Dicho árbol se recorre heurísticamente de manera de ahorrar tiempo de cómputo. De esta manera, una rama cualquiera se deja de explorar si después del análisis pre/post regresión: a) El número de parámetros correlacionados aumenta significativamente o b) un parámetro que a priori es conocido por ser relevante en el modelo se torna insensible. Además, si no se detectan problemas de significancia, sensibilidad e identificabilidad, se guarda la correspondiente reparametrización como exitosa.

Para poder comparar entre qué tan bueno es el ajuste del modelo a los datos experimentales de las reparametrizaciones exitosas obtenidas, se utilizan los criterios de Akaike corregidos (AICc) [11] y de información bayesiana (BIC)[12], los que indican qué tan bueno es un ajuste en relación con el número de parámetros estimados (no fijos) que involucra. Además, se consideraron los valores de los intervalos de confianza (CI) al 95% en los parámetros obtenidos. Estos valores fueron guardados en cada iteración para comparar posteriormente las reparametrizaciones obtenidas.

Para resolver la estimación de parámetros, se utilizó un algoritmo global de optimización metaheurística implementado en MATLAB [13] llamado búsqueda dispersa (SSm), el cual es óptimo para problemas complejos de optimización dinámica en la industria de bioprocesos [14]. Para cada estimación de parámetros de dicho procedimiento, se usó como valor inicial y rango permitido de búsqueda para los parámetros los reportados en el trabajo original [10]. Los datos para la calibración fueron facilitados por Surisetty et al. [10].

RESULTADOS Y DISCUSIÓN

Realizando el procedimiento heurístico descrito se recorrió parte del árbol de reparametrizaciones del modelo de crecimiento de microalgas a través de 37 iteraciones. Cada iteración alcanzó tiempos de convergencia entre 30 y 60 minutos. Esto permitió obtener siete reparametrizaciones exitosas, es decir, con un set de parámetros sin problemas de identificabilidad, sensibilidad y significancia. Dentro de dicho set, la mejor solución (según los cálculos de AICc, BIC y CI) resultó tener seis parámetros fijos y seis parámetros libres. Los parámetros fijados correspondieron a q_m , K_q , ρ_m , Y_{xs} , k_m y π_m , siendo estimados μ_m , K_s , s_ϕ , Y_{ps} , K_π y Y_{xq} .

La solución se muestra gráficamente junto con la solución del trabajo original en la Figura 2. En ésta es posible observar que el ajuste obtenido por medio del procedimiento heurístico propuesto es muy similar al del modelo original, pero contiene tan sólo 6 parámetros estimados en vez de los 12 originales.

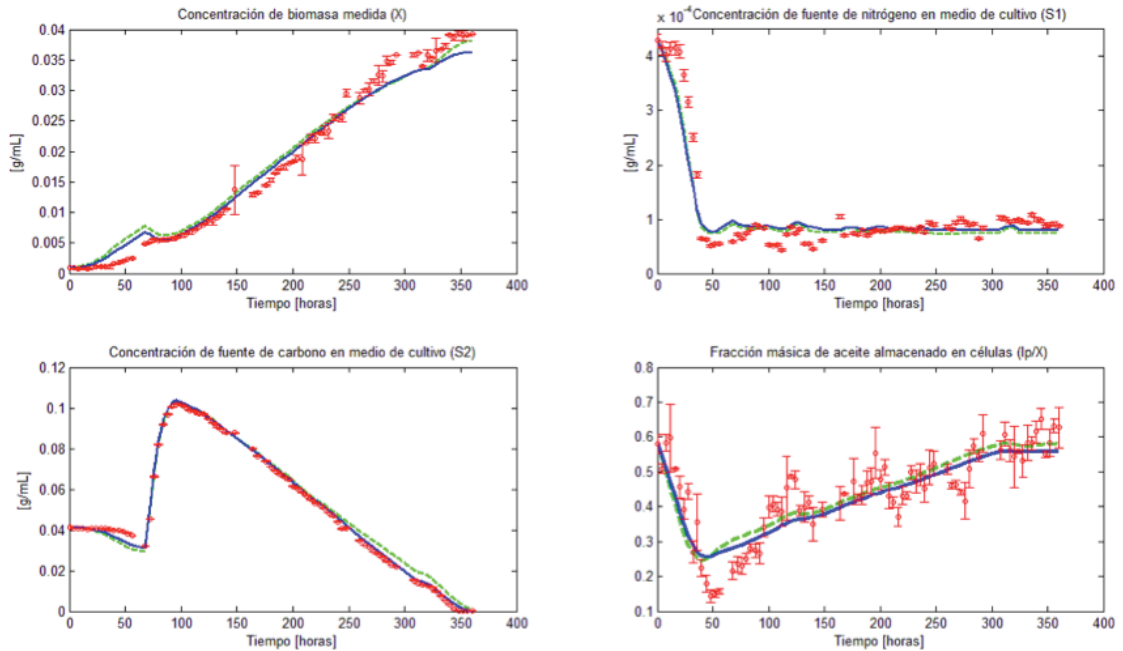


Figura 2. Solución original con 12 parámetros (línea discontinua verde), la reparametrización lograda con 6 parámetros (línea continua azul) y los datos experimentales (círculos rojos).

La solución encontrada presenta, además de la disminución del número de parámetros, características positivas respecto del modelo original. En primer lugar, la solución original contenía correlaciones absolutas mayores a 0,95 entre k_m e Y_{ps} , y entre K_π e Y_{xq} . La reparametrización propuesta, en cambio, no presenta ningún problema de correlación, es decir, todos sus parámetros son identificables, como se observa en la Figura 3. En segundo lugar, los análisis de sensibilidad muestran que todos los parámetros encontrados luego de la reparametrización influyen en al menos una de las variables de estado, a diferencia de la solución original (información no mostrada). Finalmente, la solución cumple con tener parámetros más significativos (y por lo tanto CI más angostos) que los del modelo original, siendo el promedio de los CC del modelo de seis parámetros tres veces menor.

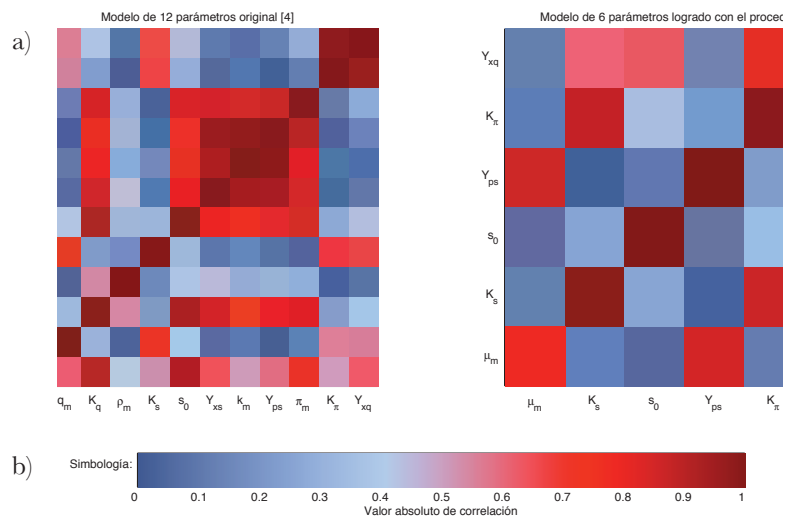


Figura 3. Mapa de colores de la correlación presentada entre los parámetros del modelo original (a) y del modelo reparametrizado (b). Los parámetros con correlaciones sobre 0,95 presentan problemas de identificabilidad.

CONCLUSIONES

En este trabajo hemos presentado una reparametrización a un modelo de crecimiento heterotrófico de microalgas para producción de biodiésel en un reactor *fed-batch* previamente reportado en la literatura. A partir de este procedimiento se logró reducir el número de parámetros del modelo original a la mitad, pasando de doce parámetros libres a sólo seis. El nuevo modelo logra un buen ajuste a los datos experimentales de acuerdo al criterio de Akaike corregido, al criterio de información bayesiana, y los intervalos de confianza. Esto implica el descenso en órdenes de magnitud del esfuerzo computacional de los algoritmos de optimización para realizar el ajuste, pudiendo llegar en menores tiempos de procesamiento a soluciones mejores.

Además de lo anterior, para que dicho modelo tenga capacidad predictiva se requiere que sea robusto y con parámetros sin problemas de identificabilidad, sensibilidad o significancia. La reparametrización fue realizada de tal forma que los seis parámetros estimados no presentan ninguno de estos problemas, aumentando la precisión de la matriz.

Recorrer todas las posibles opciones de reparametrización significa un gran esfuerzo computacional y temporal. El procedimiento heurístico iterativo empleado logró, a través de sucesivas reparametrizaciones y análisis pre/post regresión, encontrar una buena solución recorriendo de manera sistemática e inteligente un espacio acotado de todas las opciones posibles.

La reparametrización propuesta puede ser usada para predecir con mayor exactitud el crecimiento de estos cultivos. Con mejores representaciones predictivas en esta área se podrán bajar los costos de producción, permitiendo con ello que el biodiésel sea una alternativa viable a los combustibles fósiles.

PRINCIPIO CIENTÍFICO UTILIZADO

La estimación de parámetros correspondió a un problema de optimización dinámica, donde la función a minimizar fue la suma de los errores cuadráticos entre los valores de cada variable predichos por el modelo y los obtenidos experimentalmente, ponderados por la desviación estándar promedio de cada variable:

$$(1) \text{Min} \sum_{l=1}^N \sum_{i=1}^n \left(\frac{x_{il}^{\text{model}} - x_{il}^{\text{exp}}}{\bar{\sigma}_i} \right)^2$$

Donde N es el número de mediciones, n es el número de variables medidas experimentalmente (no necesariamente igual al número de variables de estado), y $\bar{\sigma}_i$ es la desviación estándar promedio para cada variable medida.

El análisis de pre/post regresión se involucró análisis de sensibilidad, identificabilidad y significancia. La sensibilidad fue calculada como:

$$(2) \mathbf{G}_{kj}(t) = \frac{\theta_j}{X_k} \frac{dX_k}{d\theta_j}$$

Donde θ_j corresponde al parámetro j y X_k a la variable k . Dicha sensibilidad se almacena para todo parámetro y para toda variable en una matriz de sensibilidad \mathbf{G}_t para todo tiempo t de la simulación.

Además, con las matrices \mathbf{G}_t se calculan matrices de correlación en cada tiempo, para identificar parámetros que presentan problemas de identificabilidad, es decir $|corr_{ij}| > 0.95$.

Las matrices de sensibilidad \mathbf{G}_t también permiten calcular la Matriz de Información de Fisher (FIM) (15) y extraer así la varianza (σ_j) de cada parámetro (16). Esta última fue utilizada para obtener los valores de los intervalos de confianza de los parámetros obtenidos, los que fueron calculados al 95% de confianza:

$$(3) \text{CI}_j = [\hat{\theta}_j \pm 1,96 \sigma_j]$$

Donde θ_j es el valor del parámetro j . La significancia de un parámetro se expresó a través del coeficiente de confianza (CC), calculado como:

$$(4) \text{CC}_j = \frac{\Delta(\text{CI}_j)}{\hat{\theta}_j}$$

Donde $\Delta(\text{CI}_j)$ corresponde al largo del CI. Un parámetro se define significativo si su respectivo CI no contiene el valor cero; luego se debe cumplir que $\frac{\Delta(\text{CI}_j)}{2} \leq \hat{\theta}_j$, es decir $\text{CC}_j \leq 2$, de lo contrario el parámetro debe fijarse en cero.

GLOSARIO

Identificabilidad: Un parámetro es identificable si puede ser estimado unívocamente al ajustar el modelo a los datos experimentales.

Sensibilidad: Un parámetro es sensible si tiene una fuerte influencia en al menos una variable de estado del modelo.

Significancia: Un parámetro es estadísticamente significativo si, a pesar del ruido experimental, la probabilidad de que la magnitud de dicho parámetro sea distinta de cero excede el umbral de confianza.

AGRADECIMIENTOS

Agradecemos al grupo del profesor Amos Ben-Zvi (Department of Chemical and Materials Engineering, University of Alberta), que generosamente nos facilitó todas las mediciones experimentales de este trabajo.

REFERENCIAS

1. Monastersky, R. A burden beyond bearing. *Nature* 458 (7242): 1091-1094, 2009.
2. Naik, S. N.; Goud, V. V., Rout, P. K. and Dalai, A. K. Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews* 14(2): 578-597, 2010.
3. Antizar-Ladislao, B.; Turrión-Gómez, J. Second-generation biofuels and local bioenergy systems. *Biofpr: Biofuels, Bioproducts and Biorefining* 2 (5): 455-469, 2008.
4. Chisti, Y. Biodiesel from microalgae. *Biotechnology Advances* 25 (3): 294-306, 2007.
5. Chisti, Y. Biodiesel from microalgae beats bioethanol. *Trends in Biotechnology* 26 (3): 126-131, 2008.
6. Amaro, H. M.; Guedes, A. C.; Malcata, F. X. Advances and perspectives in using microalgae to produce biodiesel. *Applied Energy* 88 (10): 3402-3410, 2011.
7. Chu, Y.; Hahn, J. Integrating parameter selection with experimental design under uncertainty for nonlinear dynamic systems. *AIChE Journal* 54 (9): 2310-2320, 2008.
8. Sacher, J.; Saa, P.; Cárcamo, M.; López, J.; Gelmi, C.; Pérez Correa, R. Improved calibration of a solid substrate fermentation model. *Electronic Journal of Biotechnology* 14 (5), 2011.
9. Jaquaman, K.; Danuser, G. Linking data to models: data regression. *Nature Reviews. Molecular Cell Biology* 7 (11): 813-819, 2006.
10. Surisetty, K.; Hoz Siegler, H. D. I.; McCaffrey, W. C.; Ben-Zvi, A. Robust modeling of a microalgal heterotrophic fed-batch bioreactor. *Chemical Engineering Science* 65 (19): 5402-5410, 2010.
11. Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6): 716-723, 1974.
12. Hurvich, C. M.; Tsai, C-L. Regression and time series model selection in small samples. *Biometrika* 76 (2): 297-307, 1989.
13. MATLAB. Version 7.10.0 (R2010a). Natick, Massachusetts: The MathWorks Inc., 2010.
14. Egea, J.; Rodríguez-Fernández, M.; Banga, J.; Martí, R.. Scatter search for chemical and bio-process optimization. *Journal of Global Optimization* 37 (3): 481-503, 2006.
15. Petersen, B.; Gernaey, K.; Vanrolleghem, P. A.; Practical identifiability of model parameters by combined respirometric-titrimetric measurements. *Water Science and Technology* 43 (7): 347-355, 2001.
16. Landaw, E. M.; Distefano, J.J. Multiexponential, multicompartmental, and noncompartmental modeling. II. Data analysis and statistical considerations. *American Journal of Physiology* 246 (5): R665-R677, 1984.

EQUIPO DE INVESTIGADORES

- 1 Daniela Soto
- 2 Benjamín Sánchez
- 3 Katherine Rojas
- 4 Francisco Jeria
- 5 Profesor Claudio Gelmi
- 6 Profesor Ricardo Pérez-Correa

