Stoichiometric identification with maximum likelihood principal component analysis

Johan Mailier · Marcel Remy · Alain Vande Wouwer

Received: 20 February 2012 / Revised: 7 June 2012 / Published online: 21 July 2012
© Springer-Verlag 2012

Abstract This study presents an effective procedure for the determination of a biologically inspired, black-box model of cultures of microorganisms (including yeasts, bacteria, plant and animal cells) in bioreactors. This procedure is based on sets of experimental data measuring the time-evolution of a few extracellular species concentrations, and makes use of maximum likelihood principal component analysis to determine, independently of the kinetics, an appropriate number of macroscopic reactions and their stoichiometry. In addition, this paper provides a discussion of the geometric interpretation of a stoichiometric matrix and the potential equivalent reaction schemes. The procedure is carefully evaluated within the stoichiometric identification framework of the growth of the yeast *Kluyveromyces marxianus* on cheese whey. Using Monte Carlo studies, it is also compared with two other previously published approaches.

Keywords Mathematical modeling · Parameter estimation · Reaction network · Bioprocess

Mathematics Subject Classification (2000) 92B08 · 92B01 · 92D08

List of symbols

\[ \mathbf{1}_N \] \( N \)-dimensional vector of ones
\[ a, b \] Indices of the state partition

J. Mailier (✉) · M. Remy · A. Vande Wouwer
Automatic Control Laboratory, University of Mons,
31 Boulevard Dolez, 7000 Mons, Belgium
e-mail: johan.mailier@umons.ac.be

A. Vande Wouwer
e-mail: Alain.VandeWouwer@umons.ac.be
$c$ Vector of the row offsets
$C$ Yield submatrix of a C-identifiable scheme
$d$ Vector of the column offsets
$D$ Input dilution rate
$\mathcal{E}$ Ethanol
$F$ Vector of external feed rates
$H$ Linear transformation matrix
$\hat{H}$ Estimate of the linear transformation matrix
$I_M$ $M$-dimensional identity matrix
$J_p$ ML criterion of a $p$-dimensional linear subspace
$J_*$ ML criterion of the true model
$k_i$ $i$th yield coefficient
$K$ Yield matrix
$K, K_0$ $M$-dimensional linear subspaces
$K_O$ Half-saturation constant related to oxygen
$K_E$ Half-saturation constant related to ethanol
$K_S$ Half-saturation constant related to lactose
$M$ Number of macroscopic reactions
$N$ Number of reacting species
$n_s$ Number of experimental samples
$O$ Dissolved oxygen
$p$ Pseudo-rank of the data matrix
$P_j$ Projection matrix of the $j$th measurement
$P_k$ Set of the products of the $k$th reaction
$R_j, R^\Delta_j$ Error covariance matrix of the $j$th measurement
$\mathcal{R}_k$ Set of the reactants of the $k$th reaction
$s_j$ $j$th singular value
$S$ Matrix of the singular values
$\hat{S}$ Truncated matrix of the singular values
$L$ Lactose
$T, T^*$ Invertible transformation matrix
$U$ Matrix of the left-singular vectors
$\hat{U}$ Truncated matrix of the left-singular vectors
$v$ Transport vector
$V$ Matrix of the right-singular vectors
$\hat{V}$ Truncated matrix of the right-singular vectors
$v_O$ Oxygen transport term
$x_j$ $j$th error-free measurement
$x_{j,M}$ Subvector of the $j$th error-free measurement
$\chi$ Biomass
$y_j, y^\Delta_j$ $j$th corrupted measurement
$\hat{y}_j$ Estimate of the $j$th corrupted measurement
$Y, Y^\Delta$ Data matrix
$z$ Transformed state vector
$\alpha$ Significance level