



Time Series Analysis of Fed-batch Fermentation Process for L-valine Production

Kalin Todorov¹, Tzanko Georgiev^{2*}, Velitchka Ivanova¹,
Julia Kristeva¹, Ignat Dimov¹, Alexander Ratkov¹

¹ Institute of Microbiology, Bulgarian Academy of Science
Acad. G. Bonchev str., bl. 26, 1113 Sofia, Bulgaria, phone /fax: +359 2 870 00 97
E-mail: ratkova@microbio.bas.bg

² Technical University, Faculty of Automatics, Chair Automation of Continuous Processes
1756 Sofia, Bulgaria, phone: +359 2 965 2408
E-mail: tzg@tu-sofia.bg

* Corresponding author

Received: February 1, 2006

Accepted: March 30, 2006

Published: April 26, 2006

Abstract: Fed-batch fermentation processes are some of the most efficient and widely applied types of cultivation for industrial production of most amino acids including L-valine. Time series analysis is an important tool for description of the experimental data. This article deals with statistical inference from the time series analysis of generalised stoichiometric equations as a hypothesis for modelling and optimisation. The aim of the article is to develop some time series models of generalized stoichiometric equations. The identification procedure includes the following steps: description of the process by generalized stoichiometric equations; preliminary data processing; model structure selection for each stoichiometric equation; estimation of the model's parameters; verification of the derived models.

Keywords: Modelling, Time series analysis, Fed-batch fermentation process, L-valine.

Introduction

Amino acids are some of the most important products of the industrial microbiology. L-lysine and L-valine are known as essential amino acids [3]. The use of the amino acids is based on their nutritional value, taste, physiological activities and chemical characteristics. The main fields of the application of the amino acids are [2, 3]:

- Human nutrition.
- Animal nutrition.
- Cosmetics.
- Medicine.
- Chemical Industry.

The distribution of world amino acids market is estimated as shown in Fig. 1.

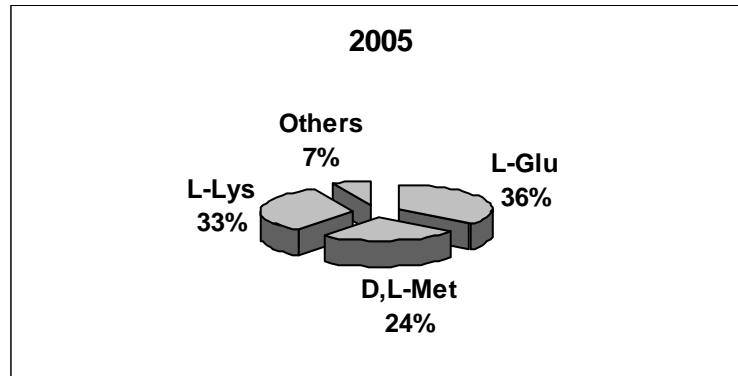


Fig. 1 Amino acids world supply

The identification problem of the biotechnological processes is threefold [1].

- Determination of the number of biological reactions.
- Identification of the underlying reaction network.
- Identification of the kinetics.

This article deals with confirmation of the chosen reaction network presented by generalized stoichiometric equations. The aim of the article is to develop some time series models of generalized stoichiometric equations. Suppose that if the stoichiometric hypothesis exists than the appropriate time series models could be established according to the experimental data. This is the purpose of the application of the time series models [5, 8]. The identification and simulation procedures are realised by MATLAB environment [4, 6].

Materials and methods

The variable volume fed-batch fermentation process is carried out at laboratory scale fermentors with 3 litres total volume. *Corynebacterium glutamicum sp. - B023* is used as a producer. Analytical methods used for the characterisation of the process are as follows: biomass is measured as dry cell mass [g/l]; sugar concentration – as reducible compounds [g/l]; L-valine – by chromatographic method [g/l] (Fig. 2). During the process on-line measurement of differed physical–chemical variables are done by proper sensors.

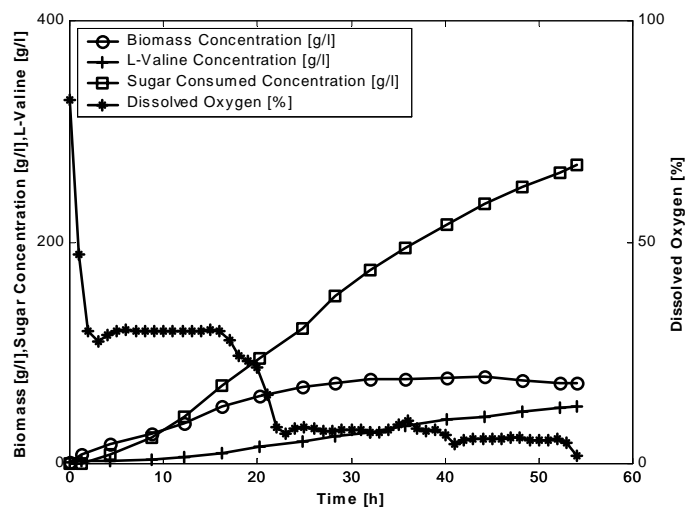


Fig. 2 Time course of experimental data from L-valine fed-batch fermentation

Primary Data Processing

Calculation of the specific rates is a final aim of this procedure. The stages of the primary processing procedure are [2]:

- Transformation the different measurable units of the concentration to unit [g/l].
- Equalization of the fed-batch process to batch one.
- Calculation of the specific rates: growth rate (μ), substrate utilisation rate (v), L-lysine or L-valine production rate (ρ_v).

Transformation the different measurable units

For expressing the biomass concentration the measured optical density units are transformed to the [g_{gry biomass}/l] using the coefficient $m_1 = 0,5887$ [g_{gry biomass}/for one unit of optical density].

Equalization of the fed-batch process to batch one

The specific rates are calculated using kinetics variables (X – biomass concentration [g/l], S_R – sugar remain concentration [g/l], S_c – sugar consumed concentration [g/l], L_V – L-valine concentration [g/l]) without any dilution effect from the feeding solution. To eliminate the dilution effect from the kinetic variables the appropriate correction is done. The correction is obtained by equations:

$$\begin{aligned} X(k)_{\text{corrected}} &= X(k)_{\text{experimental}} \exp\left(\frac{F(k)}{V(k)} h\right) \\ L_V(k)_{\text{corrected}} &= L_V(k)_{\text{experimental}} \exp\left(\frac{F(k)}{V(k)} h\right) \\ S_R(k)_{\text{corrected}} &= S_R(k)_{\text{experimental}} \exp\left(-\frac{F(k)}{V(k)} h\right) - \frac{F(k)}{V(k)} S(0)h \end{aligned} \quad (1)$$

where:

h is time step $h = t_k - t_{k-1}$, [h];

F – feeding rate, [l/h];

V – volume, [l];

$S(0)$ – concentration of the solution, [g/l];

L_V – L-valine concentration, [g/l].

Calculation of the specific rates

The specific growth rate, the rate of substrate utilisation, the specific production rates are calculated by equations:

$$\mu = \frac{\dot{X}}{X}, \quad v = \frac{\dot{S}}{X}, \quad \rho_v = \frac{\dot{L}_v}{X}. \quad (2)$$

Results and discussions

The fermentation process is described by the following schemes of generalized stoichiometric equations:



where: $\varphi_X, \varphi_g, \varphi_S, \varphi_L, \varphi_F$ are rates of the reactions, [g/l/h];
 V_0 – initial volume, V_f – final volume of the culture broth, [l];
 X – biomass concentration, [g/l];
 S – substrate concentration as a sugar remain concentration – S_R or sugar consumed concentration – S_C , [g/l];
 L_V – L-valine concentration, [g/l];
 C – dissolved oxygen tension, [%].

The validation of the above stoichiometric equations based on the experimental data, used in this article, is confirmed by an extended autoregressive (**ARX**), (4) and autoregressive moving average (**ARMAX**), (5) models written as:

$$A(q^{-1})y(t) = B(q^{-1})u(t - nk) + e(t) \tag{4}$$

$$A(q^{-1})y(t) = B(q^{-1})u(t - nk) + C(q^{-1})e(t), \tag{5}$$

where: $A(q^{-1}), B(q^{-1})$ and $C(q^{-1})$ – polynomials in the delay operator q^{-1} ; $y(t)$ – output variable; $u(t)$ – input variables; $e(t)$ – residuals of the model.

The polynomials $A(q^{-1}), B(q^{-1})$ and $C(q^{-1})$ are written as:

$$A(q^{-1}) = 1 + a_1q^{-1} + a_2q^{-2} + \dots + a_{na}q^{-na}$$

$$B(q^{-1}) = B_0 + B_1q^{-1} + B_2q^{-2} + \dots + B_{nb}q^{-nb}$$

$$C(q^{-1}) = 1 + c_1q^{-1} + c_2q^{-2} + \dots + c_{nc}q^{-nc},$$

where: na, nb and nc – orders of the polynomials;

B_0, B_1, \dots, B_{nb} – an $(n_y \times n_u)$ matrices;

n_y and n_u – numbers of the outputs and inputs, respectively;

nk – number of delays from input to output.

The **ARX** and **ARMAX** models are derived to confirm the validation of the dependencies described by the above stoichiometric equations. The first **ARX** model expresses the relation between the consumed sugar and biomass concentration as an input and output variables, respectively. The second **ARX** model expresses the extended hypothesis about biomass formulation as a relation between two input variables: consumed sugar and dissolved oxygen, and output variable biomass concentration. The **ARMAX** model is used to express the relation between the residual sugar in culture broth and feeding solution (**F.S(0)**) as input variables and the consumed sugar as an output variable. The third **ARX** model includes three input variables: concentration of biomass, consumed sugar and dissolved oxygen and L-valine concentration as an output variable.

The structural and parametric identification of the presented **ARX** and **ARMAX** models is realised by System Identification Toolbox within **MATLAB** environment. Adequacy of the **ARX** and **ARMAX** is proved as follows:

- Initial estimation of the **ARX** and **ARMAX** model orders by using autocorrelation and partial autocorrelation functions.
- Model structure selection by using Akaike’s Information Criterion (**AIC**).
- Investigation of the residuals.
- Visualisations of the results.

The normal distribution of the residuals, not presented here, and independence between the residuals and past inputs are investigated within **MATLAB** environment. The following conclusions could be drawn from the validation of models:

- Suppose that if the stoichiometric hypothesis (3) exists than the **ARX** and **ARMAX** models could be established according to the experimental data. This is the purpose of the presented time series models.
- The presented models describe stationary time series according to the data. The analysis of the model approximation shows that the models correctly present the trends in the experimental data.
- The analysis of the normal probability plots of the residuals shows that the majority of them are normally distributed but outliers could be seen.
- Tools included in the **MATLAB** package for analysing correlation and cross-correlation functions prove the independence between the residuals and past inputs of the models.

The derived results are presented as follows:

$S_C \xrightarrow{\varphi_x} X$	$C + S_C \xrightarrow{\varphi_g} X$
$A(q) = 1 - 1,026 q^{-1} + 0,556 q^{-2}$	$A(q) = 1 - 0,5898 q^{-1}$
$B(q) = 0,8514 - 0,7337 q^{-1}$	$B_1(q) = 0,8583 - 0,7851 q^{-1}$
	$B_2(q) = 0,025 + 0,06101 q^{-1}$

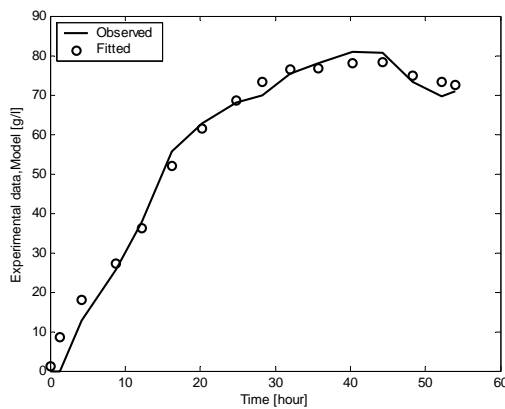


Fig. 3 Time course of the equation

$$S_C \xrightarrow{\varphi_x} X$$

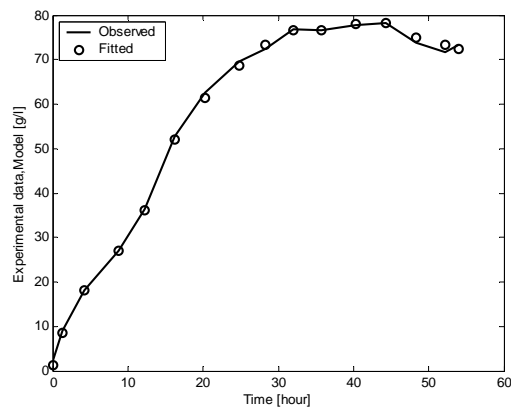


Fig. 4 Time course of the equation

$$C + S_C \xrightarrow{\varphi_g} X$$

$F.S(0) + S_R \xrightarrow{\varphi_S} S_C$	$C + S_C + X \xrightarrow{\varphi_L} L_V$
$A(q) = 1 - 2,108 q^{-1} + 1,121 q^{-2}$	$A(q) = 1 - 0,4575 q^{-1}$
$B1(q) = 0,01508 q^{-1} + 0,08072 q^{-2}$	$B_1(q) = -0,01665$
$B2(q) = -0,1035 q^{-1} + 0,07319 q^{-2}$	$B_2(q) = 0,1128$
$C(q) = 1 - 1,009 q^{-1}$	$B_3(q) = 0,01443$

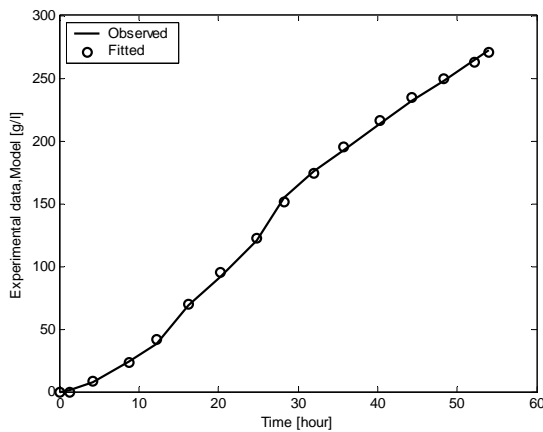


Fig. 5 Time course of the equation
 $F.S(0) + S_R \xrightarrow{\varphi_S} S_C$

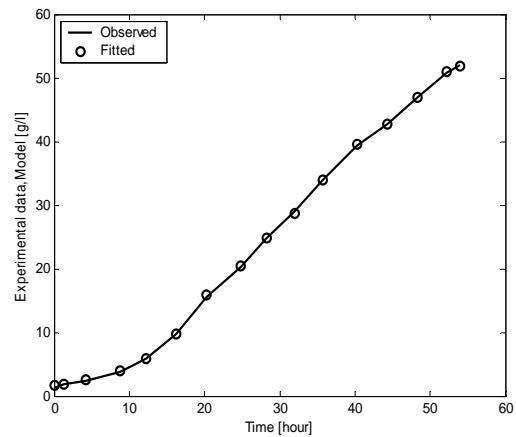


Fig. 6 Time course of the equation
 $C + S_C + X \xrightarrow{\varphi_L} L_V$

The obtained models are appropriated ones and describe the tendency of the basic kinetics variables. The hypothesis that the equation exist is revealed by the time series model based on the experimental data.

Conclusions

Regression and autoregressive (ARX) models prove the accepted number of the generalised stoichiometric reactions and underlying reaction network. The analysis of the model approximation shows that the models correctly present the trends in the experimental data. In additional their residuals are investigated.

References

1. Bastin G., L. Chen, V. Chotteau (1992). Can we Identify Biotechnological Processes?, Proceedings of IFAC, Modelling and Control of Biotechnological Processes, Colorado, USA, 83-88.
2. Georgiev Tz., Al. Ratkov, St. Tzonkov (1997). Mathematical Modelling of Fed-batch Fermentation Processes for Amino Acid Production, Mathematics and Computers in Simulation, 44, 171-285.
3. Leuchtenberger W. (1996). Amino Acids - Technical Production and Use in Biotechnology, Sc. Comp. Rev. Edition, H. J. Rehm and G. Reed Ed., 6, 465-503.
4. MathWorks Inc. (2003). MATLAB User's Guide.
5. MathWorks Inc. (2003). System Identification Toolbox, User's Guide.
6. MathWorks Inc. (2003). SIMULINK, Using SIMULINK.
7. MathWorks Inc. (2003). Optimization Toolbox, User's Guide.
8. STATGRAPHICS (2000). Version 5.1 Plus For Windows, User Manual, Magnugistics Inc. USA.