



# Journal of Applied Sciences

ISSN 1812-5654

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>

## Genetic Algorithms Using for a Batch Fermentation Process Identification

Kouamé Kan Benjamin, Assidjo N. Emmanuel, Akaki David and Yao K. Benjamin  
Groupe Modélisation de l'Information et des Processus, Laboratoire des Procédés Industriels,  
de Synthèse et Environnement, Institut National Polytechnique Houphouët-Boigny,  
BP 1313 Yamoussoukro, Côte d'Ivoire

---

**Abstract:** This study presents the use of genetic algorithm methodology for identification of a micro-organism mixture batch fermentation process. This process is described by complex and nonlinear differential equations according the mass balance. The parameters of the model are estimated using a multi-objective genetic algorithm. Simulations of the model obtained show accuracy in prediction behaviour of the model.

**Key words:** Genetic algorithm, multi-objective optimisation, batch fermentation process, parameters identification, modelling

---

### INTRODUCTION

Fermentation processes are very complex and their modelling is rather a challenging task (Assidjo *et al.*, 2006; Coleman *et al.*, 2003; Garcia-Ruiz *et al.*, 2008). Indeed, dynamic models for industrial fermentations processes are difficult to identify because of a wide varieties of reasons e.g., micro-organisms complex dynamics, variables and ill-defined raw materials, varying inocula quality (Assidjo *et al.*, 2006; Coleman *et al.*, 2003; Garcia-Ruiz *et al.*, 2008). However, their study was intensively investigated and different models proposed for control (Moore *et al.*, 2001; Marin, 1999). A good model must take into account the effects of substrate limitation, substrate and products inhibition as well as maintenance energy and cells death on the cell growth and metabolism (Moore *et al.*, 2001).

However, is neither necessary nor desirable to construct comprehensive mechanistic process models that can describe the systems in all possible situations with a high accuracy.

In this intention, many mathematical models have been proposed to predict the influence of fermentation operating parameters on cell growth rate, cell concentration (biomass), substrate utilisation rate (Sarkar and Modak, 2005; Trelea *et al.*, 2004). The use of these models may lead to the development of better strategies for the optimisation of the fermentation process to ensure its economic viability. In fact, few fermentation models have been used for industrial scale fermentation optimisation (Nandasana and Kumar, 2008).

Genetic algorithms have been widely used for optimisation and system identification. They are robust, global and generally more straightforward to apply in situation where there is little or no a priori knowledge about the process to be controlled (Nandasana and Kumar, 2008; Sarkar and Modak, 2005; Goldberg, 1989). They are stochastic search technique for approximating optimal solution within complex search spaces. They are based upon the analogy with biological evolution, in which the fitness of individual determines its ability to survive and reproduce. Their mechanism, as shown in Fig. 1, starts by encoding the problem to produce a list of genes (Yao *et al.*, 2007; Davis, 1991). The genes are represented by either numeric or alphanumeric characters. The genes are randomly combined to give a population of chromosomes, each of which represents a possible solution. Genetic operations are performed on chromosomes that are randomly selected from the population producing therefore offspring (Yao *et al.*, 2007; Polifke *et al.*, 1998). The fitness of these chromosomes is measured and the probability of their survival is determined. In this study, a genetic algorithm has been used for a batch fermentation parameters identification.

### MATERIALS AND METHODS

**Data collection and treatment:** The brewing fermentations were performed using a Brunswick microferm fermentor (New Brunswick Scientific Co. Inc., New Jersey, USA), in batch mode. This micro-fermentor is characterised by a volume of about 15 L. Its vessel internal diameter is 10 cm and its height is 50 cm.

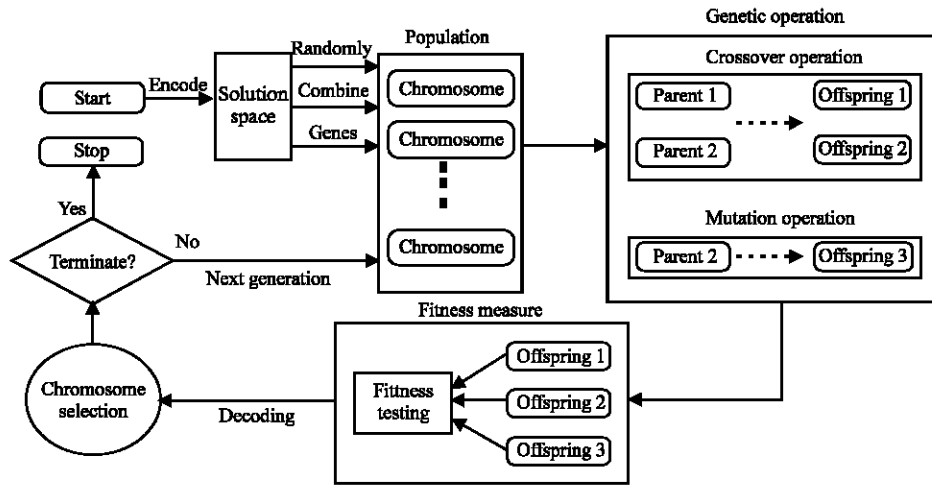


Fig. 1: Flowchart of genetic algorithm process

The wort was produced by crushing the malt into coarse flour, which was then mixed with water. The resulting porridge-like mash was heated to a selected temperature that permitted the malt enzymes to partially solubilize the ground malt. The resulting sugar-rich aqueous extract (wort), was then separated from the solids and boiled. The wort was then clarified, cooled and poured in the vessel of the micro-fermentor for inoculation. Inoculated ferments were traditionally produced by female brewers that use it for a well-known local beer (tchapalo or dolo) making. This ferment is in fact a mixture of micro-organisms containing different species (e.g., *Saccharomyces cerevisiae* and *Candida*). During the fermentation process ( $t = 0$  to 18 h), different parameters (pH, biomass, substrate, ethanol, carbon dioxide) were measured. The responses concerned in this study (biomass, substrate and alcohol (ethanol)) were determined by gravimetric and refractometric methods and refractometric method after a distillation, respectively (Thonart, 2001).

For the purpose of this study, 30 batches were performed.

**Statistical analysis:** In order to eliminate eventual outlier fermentations and because of the trilinear form of data (batches×time×responses), a PARAFAC analysis was performed.

PARAFAC is a multi-way decomposition method originating from psychometrics (Rutledge and Bouveresse, 2007; Khayamian, 2007). It is gaining more interest in chemometrics and associated areas for many reasons: simply increased awareness of the method and its possibilities, the increased complexity of the data dealt with in science and industry and increased computational power (Geladi, 1989; Rutledge and Bouveresse, 2007).

PARAFAC decomposes the array into sets of scores and loadings, which describe the data in a more condensed form than the original data array. It conceptually can be compared to bilinear PCA, or rather it is one generalization of bilinear PCA (Smilde, 1992; Bouveresse *et al.*, 2007). This technique is designed to decompose higher order data tables (e.g., cubes), again to reveal the underlying, latent phenomena for the purpose of data analysis and predictions.

The decomposition of the data is made into trilinear components, but instead of one score vector and one loading vector as in bilinear PCA, each component consists of one score vector and two loading vectors (Bouveresse *et al.*, 2007). It is common three-way practice not to distinguish between scores and loadings as these are treated equally numerically.

Therefore, a PARAFAC model of a three-way array is given by three loading matrices, A, B and C with elements  $a_{if}$ ,  $b_{jf}$  and  $c_{kf}$ . The trilinear model is found to minimize the sum of squares of the residuals,  $e_{ijk}$  in the model:

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk} \quad (1)$$

The advantage of the PARAFAC model is the uniqueness of the solution, meaning that no restrictions are necessary to identify estimate the model apart from trivial variations of scale and column order (Bouveresse *et al.*, 2007). Therefore the true and estimated models must coincide when the right number of components is chosen.

Leurgans *et al.* (1993) have shown that unique solutions can be expected if the loading vectors are linear independent in two of the modes and furthermore in the third mode the less restrictive condition is that no two

loading vectors are linearly dependent. In PARAFAC, one does not deflate the array, because the trilinear model calculated simultaneously for all components can be shown to fit the array better, than if the components were calculated successively as is possible in PCA (Mortensen and Bro, 2006; Leurgans *et al.*, 1993). As a consequence, extracting too many components does not only mean that noise is being increasingly modeled, but also that the true factors are being modeled by more (correlated) components.

In this study, calculations were performed using Matlab R2007b (MathWorks Inc., Massachusetts, USA) software.

**Process description:** Batch fermentation refers to a partially closed system in which most of required material for micro-organisms growth and maintenance are loaded before process starts. Conditions during fermentation are continuously changing with time leading the fermentor being an unsteady state system (Roeva *et al.*, 2004).

In alcoholic brewing fermentation, the micro-organisms (biomass) concentration is the central feature affecting the rate of growth, substrate consumption and product formation (Trelea *et al.*, 2004). Growth and alcohol formation rates vary with time due to a dependence on the present state of the batch which is characterised by biomass, substrate and product concentrations, oxygen tension and culture conditions (Roeva *et al.*, 2004; Thonart, 2001).

The model developed herein is based on the assumptions that:

- The bioreactor is completely mixed.
- The substrate is consumed mainly oxidatively and its consumption can be described by Monod kinetics.
- The ethanol production is assumed to be directly linked to the biomass formation.
- Variations in the growth rate, ethanol production and substrate consumption do not significantly change the elemental composition of biomass.

The rates of cells growth, substrate consumption, ethanol formation as well as carbon dioxide concentration in batch fermentation are commonly described according to the mass balance (Thonart, 2001):

$$\mu = \mu_{max} \cdot \frac{S}{S + K_s} \quad (2)$$

$$\frac{dX}{dt} = \mu \cdot X \quad (3)$$

$$\frac{dP}{dt} = \frac{\mu}{Y_{P/X}} \cdot X \quad (4)$$

$$\frac{dS}{dt} = -\frac{\mu}{Y_{X/S}} \cdot X + m_s \cdot X \quad (5)$$

The specific growth rate  $\mu$  is generally found to be function of three factors: the limiting concentration of substrate, the maximum growth rate  $\mu_{max}$  and the substrate specific constant  $K_s$ .

If taking account realistic aspects of the process:

- Substrate limitation
- Ethanol and substrate inhibition
- Lag phase...

The model drawn is as follows:

Lag phase

$$\frac{dx_{act}}{dt} = \mu_{lag} (0.48x_i - x_{act}) \quad (6)$$

$$\frac{dx_{lag}}{dt} = -\mu_{lag} x_{lag} \quad (7)$$

$$\frac{dx_{bot}}{dt} = K_s \cdot x_{act} - \mu_d \cdot x_{bot} \quad (8)$$

Fermentation phase

$$\frac{dx_{act}}{dt} = \mu_x \cdot x_{act} - K_m \cdot x_{act} + \mu_{lag} \cdot x_{lag} \quad (9)$$

$$\frac{dx_{lag}}{dt} = -\mu_{lag} x_{lag} \quad (10)$$

$$\frac{dx_{bot}}{dt} = K_m \cdot x_{act} - \mu_d \cdot x_{bot} \quad (11)$$

$$\frac{dS}{dt} = -\mu_s \cdot x_{act} \quad (12)$$

$$\frac{de}{dt} = \mu_a \cdot f \cdot x_{act} \quad (13)$$

The parameters

$$\mu_x = \mu_{x0} \cdot \frac{S}{0.5S_i + e} \quad (14)$$

$$\mu_s = \mu_{s0} \cdot \frac{S}{K_s + S} \quad (15)$$

$$\mu_a = \mu_{a0} \cdot \frac{S}{K_a + S} \tag{16}$$

$$\mu_d = 0.5 \cdot \mu_{d0} \cdot S_i \tag{17}$$

$$f = 1 - \frac{e}{0.5S_i} \tag{18}$$

This model was developed by Andres-Toro *et al.* (2004) who have adjusted it to laboratory scale experimental data. It takes account three components of the biomass: lag, active and dead cells and consider the active cells as the only fermentation agent. The model also includes sugar and ethanol concentrations and two important by-products of the fermentation that degrade beer quality: ethyl acetate and diacetyl.

The system identification is achieved by determining parameters like  $K_m, K_s, \mu_{x0}, \mu_{d0} \dots$

**RESULTS AND DISCUSSION**

**Outlier determination:** As described earlier, a PARAFAC decomposition was made; scores and loadings, for time, batches and responses, were drawn. Loadings (B) concerning batches are shown in Fig. 2 with the 95% confidence ellipse.

The analysis of Fig. 2 shows that data were regrouped in a set of batches inside the ellipse contour. Unfortunately, 2 batches are isolated from the others and out of this ellipse. These outlier (Johnson and Wichern, 1992) batches were removed from data set prior process identification.

**Parameter identification:** Identifying parameters represents a very difficult task to elucidate. In general, conventional methods, as simplex, are local optimisation methods based on gradient determination; supposing that functions must be derivable. Moreover, conventional search techniques are often incapable of optimising non-linear multi-modal functions. In such cases, a random and global search method might be required. Genetic algorithms do not use much knowledge about the problem to be optimised. They work with codes, which represent the problem parameters (i.e.,  $\mu_{x0}, \mu_{s0}, K_s, \mu_{a0}, K_a, \mu_{lag}, K_m, \mu_{d0}$ ).

The differential (6-13) and parameters (14-18) equations and initial values were implemented in a script under Matlab R2007b (MathWorks Inc., Massachusetts, USA) environment. Differential equations were resolved using Runge-Kutta 4th order algorithm at specified time (e.g., 0, 1, 2, 3, 4, 5 ... h).

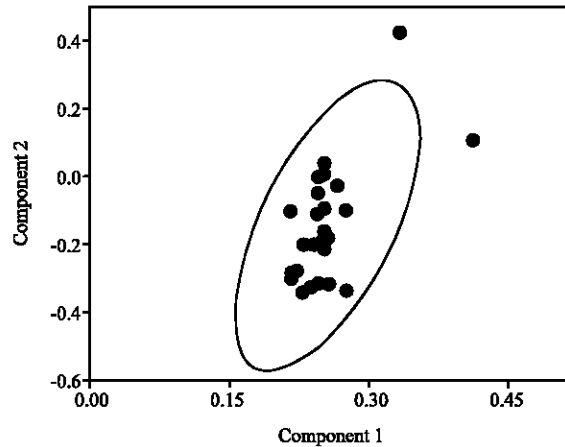


Fig. 2: Batches projection on PARAFAC components

Table 1: Initial conditions of genetic algorithms

Population size	100
Elite count	2
No. of generations	150
Fitness scaling function	@fitscalingrank
Selection function	@selectionstochunif
Crossover function	@crossoverscattered
Mutation function	@mutationuniform
Mutation probability	0.05

In order to optimise the batch fermentation modelled, another script containing the necessary instructions for the genetic algorithm toolbox has been developed. In this script, some initial parameters needed in the genetic algorithm toolbox were defined, e.g., individual number, generation maximum number, crossover and mutation rates, selection function...

In order to implement the genetic algorithm, the model parameters (e.g.,  $K_s, K_m, \mu_{lag}, \mu_{d0} \dots$ ) have to be presented as chromosomes. Decimal numbers ranging from 0 to 20 for the parameters values have been used to represent this principle. Each chromosome corresponds to a possible solution of the objective function. Generally, this function is expressed as the modelling error i.e., the mean square deviation between the model output and the corresponding data obtained during the fermentation. Therefore, the optimisation criteria are as follows:

$$J_f = \sum (Y_c - Y_o)^2 \Rightarrow \min \tag{19}$$

with  $Y_c$  the calculated value and  $Y_o$  its corresponding observed one, respectively for biomass, substrate and alcohol.

Initial tests were performed using conditions shown in Table 1.

After several and different runs, the conditions retained are as follows:

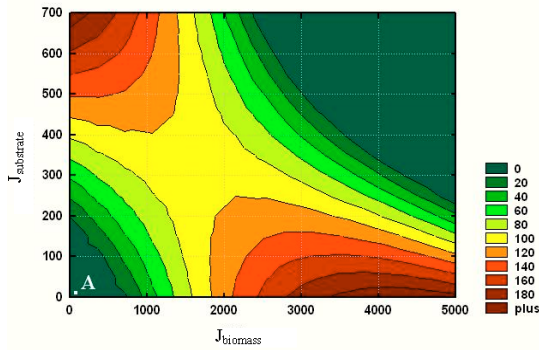


Fig. 3: Contour plot of Pareto solutions

Table 2: Values of parameters after identification

Parameters	Values
$\mu_{x0}$	1.2500
$\mu_{a0}$	0.0686
$\mu_{a0}$	1.1167
$\mu_{lag}$	0.5770
$\mu_{d0}$	0.3366
$K_s$	0.4515
$K_a$	0.9347
$K_m$	0.8392

- Population size = 50
- No. of generations = 100

The release R2007b of Matlab integrates in genetic toolbox the possibility of optimising multi-objective function (Mathworks, 2007). The solutions called Pareto solutions for the three functions ( $J_{Biomass}$ ,  $J_{Substrate}$  and  $J_{Alcohol}$ ) are shown in Fig. 3 in contour plot.

Glancing at Fig. 3, it appears that two optimal regions are represented concerning the  $J_{Alcohol}$  criterion. But if taking account the other criteria (i.e.,  $J_{Biomass}$  and  $J_{Substrate}$ ), the optimal region to consider is the lower green one ranging from 0 to 800 for  $J_{Biomass}$  and from 0 to 200 for  $J_{Substrate}$ . The retained point (i.e., point A in the Fig. 3) that is a good compromise of different functions minimum is situated in this zone. This point corresponds to parameters shown in Table 2.

In this case, the values of the criteria are:

$$J_{Biomass} = 44.982, J_{Substrate} = 9.724 \text{ and } J_{Alcohol} = 7.604.$$

Figure 4 shows the experimental and predicted values of, respectively biomass, substrate (i.e., glucose) and ethanol (alcohol).

These different phases were already extensively described in literature: the lag-phase during where micro-organisms adapt themselves to the culture medium; the active phase or exponential phase during which the yeasts multiply themselves exponentially, consuming sugar and producing alcohol. The decrease of yeast

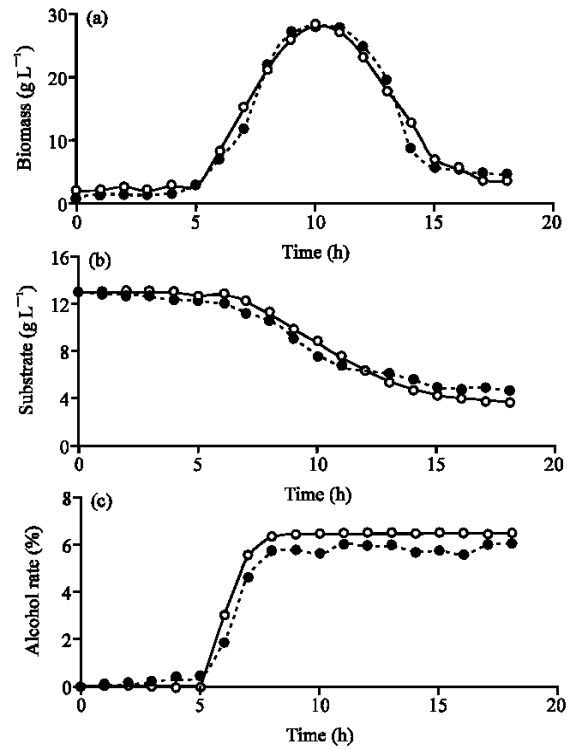


Fig. 4: Model (○) and experimental (●) data plot, (a) Biomass, (b) Substrate and (c) Alcohol

number in suspension is due to their fall down in fermentation tank confirming therefore, the lower fermentation type.

The global analysis of prevision shows in Fig. 4a-c) points out the good ability of the model to predict successfully the variation of substrate consumption, biomass concentration and alcohol production. Indeed, the model obtained gives values very close to the observed ones, whatever the response concerned during the fermentation studied.

### CONCLUSION

In this research, the model developed by Andrés-Toro *et al.* (2004) for beer fermentation was proposed to fit a batch fermentation by micro-organism mixture data. The study is accomplished through the formulation of the identification problem as an optimisation problem and the application of multi-objective genetic algorithm in order to estimate the unknown parameters from input-output space.

The simulations operated thereby, concerning the substrate consumption, biomass concentration and the alcohol production, validate the predictions of the genetic algorithm formalism with a good accuracy. Therefore, the

multi-objective genetic algorithm methodology presented in this paper offers a cost-effective and relatively simple alternative for process modelling and optimization.

### NOMENCLATURE

$\mu$  : Specific growth rate ( $\text{h}^{-1}$ )  
 $\mu_a$  : Alcohol production rate ( $\text{h}^{-1}$ )  
 $\mu_{a0}$  : Specific alcohol production rate ( $\text{h}^{-1}$ )  
 $\mu_d$  : Yeast settling down rate ( $\text{h}^{-1}$ )  
 $\mu_{d0}$  : Specific yeast settling down rate ( $\text{h}^{-1}$ )  
 $\mu_{lag}$  : Specific rate of latent formation ( $\text{h}^{-1}$ )  
 $\mu_{max}$  : Maximum growth rate ( $\text{h}^{-1}$ )  
 $\mu_s$  : Substrate consumption rate ( $\text{h}^{-1}$ )  
 $\mu_{s0}$  : Specific substrate consumption rate ( $\text{h}^{-1}$ )  
 $\mu_x$  : Yeast growth rate ( $\text{h}^{-1}$ )  
 $\mu_{x0}$  : Specific yeast growth rate ( $\text{h}^{-1}$ )  
 $e$  : Concentration of the alcohol ( $\text{g L}^{-1}$ )  
 $f$  : Fermentation inhibitor factor  
 $K_a$  : Alcohol inhibition parameter  
 $K_m$  : Yeast growth inhibition parameter  
 $K_s$  : Sugar inhibition parameter  
 $m_s$  : Maintenance constant ( $\text{g g}^{-1} \text{h}^{-1}$ )  
 $P$  : Concentration of the product ( $\text{g L}^{-1}$ )  
 $S$  : Concentration of the substrate ( $\text{g L}^{-1}$ )  
 $S_i$  : Initial concentration of substrate ( $\text{g L}^{-1}$ )  
 $X$  : Concentration of biomass ( $\text{g L}^{-1}$ )  
 $x_{act}$  : Concentration of active biomass ( $\text{g L}^{-1}$ )  
 $x_{bot}$  : Concentration of bottom biomass ( $\text{g L}^{-1}$ )  
 $x_i$  : Initial concentration of biomass ( $\text{g L}^{-1}$ )  
 $x_{lag}$  : Concentration of latent biomass ( $\text{g L}^{-1}$ )  
 $Y_{p/x}$  : Yield coefficient ( $\text{g g}^{-1}$ )  
 $Y_{x/s}$  : Yield coefficient ( $\text{g g}^{-1}$ )

### REFERENCES

- Andrés-Toro, B., J.M. Girón-Sierra, P. Fernández-Blanco, J.A. López-Orozco and E. Besada-Portas, 2004. Multiobjective optimization and multivariable control of the beer fermentation process with the use of evolutionary algorithms. *J. Zhejiang Univ. Sci.*, 5 (4): 378-389.
- Assidjo, E., B. Yao, D. Amame, G. Ado, C. Azzaro-Pantel and A. Davin, 2006. Industrial brewery modelling by using artificial neural network. *J. Applied Sci.*, 6 (6): 1123-1127.
- Bouveresse, D.J.R., H. Benabid and D.N. Rutledge, 2007. Independent component analysis as a pretreatment method for parallel factor analysis to eliminate artefacts from multiway data. *Anal. Chem. Acta*, 589 (2): 216-224.
- Coleman, M.C., K.K. Buck and D.E. Block, 2003. An integrated approach to optimization of *Escherichia coli* fermentations using historical data. *Biotechnol. Bioeng.*, 84 (4): 274-285.
- Davis, L., 1991. *Handbook of Genetic Algorithms*: Van Nostrand Reinhold: New York, USA.
- García-Ruiz, A., B. Bartolomé, A.J. Martínez-Rodríguez, E. Pueyo, P.J. Martín-Álvarez and M.V. Moreno-Arribas, 2008. Potential of phenolic compounds for controlling lactic acid bacteria growth in wine. *Food Control*, 19 (9): 835-841.
- Geladi, P., 1989. Analysis of multi-way (multi-mode) data. *Chemom. Intel. Lab. Syst.*, 7 (1-2): 11-30.
- Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley, New York, USA.
- Johnson, R.A. and D.W. Wichern, 1992. *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, New Jersey, USA., pp: 1-36.
- Khayamian, T., 2007. Robustness of PARAFAC and N-PLS regression models in relation to homoscedastic and heteroscedastic noise. *Chemom. Intel. Lab. Syst.*, 88 (1): 35-40.
- Leurgans, S., R.T. Ross and R.B. Abel, 1993. A decomposition for three-way arrays. *SIAM J. Matrix Anal. Applied*, 14 (4): 1064-1083.
- Marin, M.R., 1999. Alcoholic fermentation Modelling: current state and perspectives. *Am. J. Enol. Vitic.*, 50 (2): 166-178.
- Mathworks, 2007. *Genetic Algorithm and Direct Search Toolbox for Use with MATLAB*; The Mathworks Inc. Natick, MA (USA).
- Moore, C.F., N. Kaistha and M.S. Johnson, 2001. Monitoring and analysis of batch profiles. *J. Proc. Anal. Chem.*, 2 (3): 65-80.
- Mortensen, P.P. and R. Bro, 2006. Real-time monitoring and chemical profiling of a cultivation process. *Chemom. Intel. Lab. Syst.*, 84 (1-2): 106-113.
- Nandasana, A.D. and S. Kumar, 2008. Kinetic modelling of lactic acid production from molasses using *Enterococcus faecalis* RKY1. *Biochem. Eng. J.*, 38 (3): 277-284.
- Polifke, W., W. Geng and K. Dobbeling, 1998. Optimization of rate coefficients for simplified reaction mechanisms with genetic algorithms. *Combust. Flame*, 113 (1-2): 119-134.
- Roeva, O., T. Pencheva, B. Hitzmann and S. Tzonkov, 2004. A genetic Algorithms Based Approach for identification of *Escherichia coli* Fed-batch fermentation. *Bioautomation*, 1 (2): 30-41.

- Rutledge, D.N. and D.J.R. Bouveresse, 2007. Multi-way analysis of outer product arrays using PARAFAC. *Chemom. Intel. Lab. Syst.*, 85 (2): 170-178.
- Sarkar, D. and J.M. Modak, 2005. Pareto-optimal solutions for multi-objective optimization of fed-batch bioreactors using nondominated sorting genetic algorithm. *Chem. Eng. Sci.*, 60 (2): 481-492.
- Smilde, A.K., 1992. Three-way analyses. Problems and prospects. *Chemom. Intel. Lab. Syst.*, 15 (2-3): 143-157.
- Thonart, P., 2001. Industries d'extraction et de fermentation: 1ère partie, Centre Wallon de Biologie Industrielle, Gembloux, Belgique.
- Trelea, I.C., M. Titica and G. Corrieu, 2004. Dynamic optimisation of the aroma production in brewing fermentation. *J. Proc. Control*, 14 (1): 1-16.
- Yao, K.B., N.E. Assidjo, G. Ado, Azzaro-Pantel C. and D. Andre, 2007. Modelling and optimization of M-cresol isopropylation for obtaining N-thymol: Combining a hybrid artificial neural network with a genetic algorithm. *Int. J. Chem. React. Eng.*, 5: A870.