JOURNAL OF CHEMOMETRICS
J. Chemometrics 2003; **17**: 453–462
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.791

Kinetic modelling of acetic fermentation in an industrial process by genetic algorithms with a desirability function

Consuelo Pizarro[1]*, José M. González-Sáiz[2] and Diego Garrido-Vidal[2]

[1]Analytical Chemistry, Department of Chemistry, University of La Rioja, E-26004 Logroño, La Rioja, Spain
[2]Chemical Engineering, Department of Chemistry, University of La Rioja, E-26004 Logroño, La Rioja, Spain

Received 15 June 2002; Revised 11 October 2002; Accepted 17 January 2003

The basic tool to simulate the evolution of a bioprocess, such as the industrial process of acetic fermentation, is the kinetic model. The model must be simple and have a high predictive ability to give results capable of explaining the real behaviour. The difficulties in the kinetic modelling of biological processes are mainly related to the description of the bacterial growth. In this paper a genetic algorithm is designed to obtain a set of kinetic parameters for the specific growth rate that enables the model to explain the industrial fermentation. Only acetic concentration data from industrial fermentators are required. A four-factor desirability function works properly as the response to maximize. The model optimized can explain not only the behaviour of the industrial batch process studied, but also the experimental results obtained with a pilot fermentator working continuously with and without cell recycling. Copyright © 2003 John Wiley & Sons, Ltd.

KEYWORDS: simulation process; acetic fermentation; kinetic modelling; genetic algorithms; desirability function

1. INTRODUCTION

A predictive simulation environment is a high-potential tool. The simulation tools allow the engineers to optimize the industrial processes or to develop new ones reducing the costs, since the experimental charge must be reduced owing to the high cost of an experiment in industrial plant. They are also useful to control the processes by applying techniques such as on-line state estimation [1,2]. The simulation tools are also some of the most important tools for the scientists, since they allow one to shorten the time required to study the experimental domain and to compare process alternatives and analyses of a large number of process conditions reducing the experimental costs. Applied to bioprocesses, the advantages of process simulation have been realised for a long time now [3–5].

The basic tool to simulate the evolution of a bioprocess, such as the industrial process of acetic fermentation, is the kinetic model. The model must be simple and have a high predictive ability to provide results capable of explaining the real behaviour. Micro-organisms are very complex systems, and therefore unstructured (the internal composition and structure of the cell are not considered) and unsegregated (all cells are considered identical) models have been used extensively [6].

The data obtained from industrial fermentators are used in this work to identify the set of kinetic parameters that provides a better prediction ability for the model proposed. The concentration data have been taken from the industrial plant of the most important vinegar company in Spain, Vinagrerías Riojanas SA.

Genetic algorithms have been applied for a long time to solve various problems [7–12], including kinetic modelling. A genetic algorithm has been developed in Matlab 6.1. to optimize the kinetic model proposed. One critical point in the building of the algorithm is to find an evaluation function with a high capacity to fit the simulation to the real data. In this paper the problem is solved with a four-factor desirability function that works as the response to maximize.

The model optimized has been tested in a simulation environment different from the process conditions of the industrial plant. The results of the simulations are compared with the experimental ones obtained with the pilot fermentation system.

The nomenclature used in the mathematical expressions of this paper is reported in the Appendix. The symbols have been organized by their affiliation to the evaluation function, the kinetic model or the pilot fermentation system.

*Correspondence to: C. Pizarro, Analytical Chemistry, Department of Chemistry, University of La Rioja, C/ Madre de Dios 51, E-26004 Logroño, La Rioja, Spain.
E-mail: consuelo.pizarro@dq.unirioja.es
Contract/grant sponsor: Spanish Ministry of Agriculture and Food; Contract/grant number: INIA CAL01-053.
Contract/grant sponsor: Autonomous Government of La Rioja; Contract/grant number: FPI-1999.

2. BACKGROUND AND THEORY

2.1. Genetic algorithms

Genetic algorithms are numerical optimization methods that try to simulate biological evolution, i.e. the process of optimization of the characteristics of individuals (chromosomes) to improve their fitness to the environment. The basic idea of evolution theory is that the individuals with best ability to fit to the environment have more chances of surviving and reproducing.

Many authors have applied genetic algorithms to various optimization problems since Holland published the first works in 1975. Genetic algorithms are well known for their ability to perform robust optimizations in a multidimensional space. The applications of genetic algorithms can be roughly divided into three main categories [7]: numerical problems [8], sequencing problems [9] and subset selection problems [10,11]. Several tutorials have been published, such as those by Lucasius and Kateman [12,13] and Wehrens and Buydens [14].

It is also possible to find works where genetic algorithms have been successfully applied to the empirical modelling of fermentation processes [15]. In our work the genetic algorithm is used to optimize the parameters of a mechanistic model proposed for industrial acetification. We have adapted the basic algorithm [12,13] to our problem and have introduced a desirability function as evaluation function.

2.2. Desirability functions

Multicriteria decision making (MCDM) is applied when decisions based on multiple criteria covering several responses must be made. There are a number of methods to take multicriteria decisions, and the use of desirability functions is one of them [7,16,17]. It is possible to find recent works where desirability functions are used in multiresponse optimization [18].

A desirability function is defined for each criterion considered. The function can adopt different mathematical structures such as linear, exponential or logarithmic, but the response must be always scaled between 0 (unacceptable) and 1 (maximum desirability):

$$d_{cn} = f_c(y_{cn}), \qquad 0 \leqslant d_{cn} \leqslant 1 \tag{1}$$

where y_{cn} is the response of sample n for criterion c, f_c is the kind of single desirability function chosen for the selected criterion c, and d_{cn} is the value of the single desirability function for response y_{cn}.

The global desirability function for sample n, D_n, is then defined as the geometric mean of the single desirability functions:

$$D_n = \left(\prod^{c} d_{cn}\right)^{1/c} \tag{2}$$

2.3. Structure of the kinetic model

In this work we have assumed the general scheme for the reaction mechanisms in fermentation processes proposed by Sinclair and Kristiansen [19] (referred to throughout as Sinclair's model). This model is shown in Figure 1. The various rates in the figure are defined in Appendix A2.

2.3.1. *Growing cell kinetics*

We have applied some typical approximations to our fermentation model. These approximations are very necessary and have been used for a long time to obtain simple and predictive models.

- *Average cell approximation*. The cell-to-cell heterogeneity does not influence the model, and the average cellular properties are considered. The model is unsegregated.
- *Balance growth approximation*. The representation of the cell is a single component, and the biomass is considered as a component in solution, X_t. From this point of view the model is unstructured.

In the model proposed, the total biomass X_t is considered to be made up of viable, X_v, and dead, X_d, cells [20]. Considering the scheme in Figure 1 only for the cell growing, we have the following expressions for r_g, the overall cell growth rate, r_d, the overall cell death rate and r, the observed cell growth rate:

$$X_v \xrightarrow{r_g} X_v \xrightarrow{r_d} X_d \tag{3}$$

$$X_t = X_v + X_d \tag{4}$$

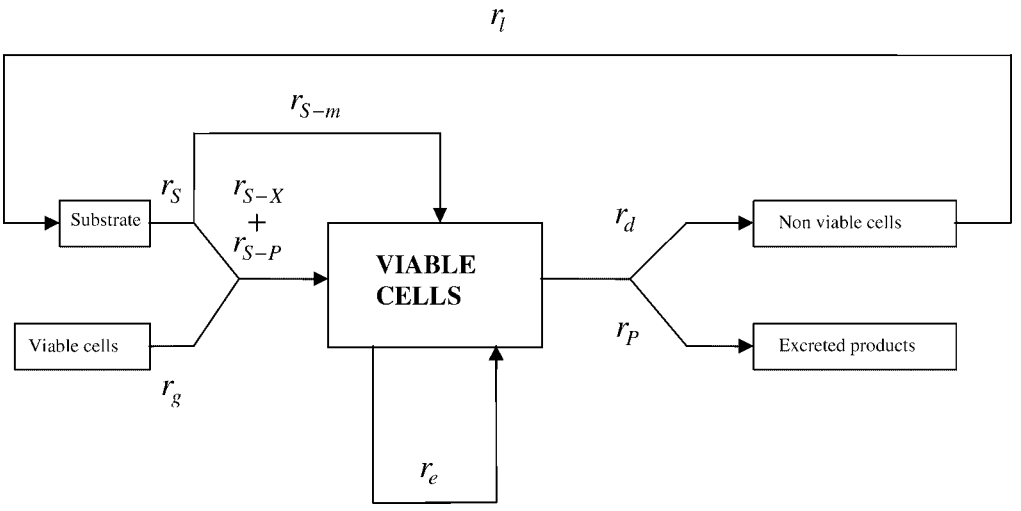

Figure 1. Sinclair's general model for fermentation processes.

$$r_g = \frac{dX_t}{dt} \tag{5}$$

$$r_d = \frac{dX_d}{dt} \tag{6}$$

$$r = r_g - r_d = \frac{dX_v}{dt} \tag{7}$$

The corresponding specific rates are defined by the overall rates and the viable biomass concentration:

$$\mu_g = \frac{1}{X_v}\left(\frac{dX_t}{dt}\right) \tag{8}$$

$$\mu_d = \frac{1}{X_v}\left(\frac{dX_d}{dt}\right) \tag{9}$$

$$\mu = \frac{1}{X_v}\frac{dX_v}{dt} \tag{10}$$

$$\mu = \mu_g - \mu_d \tag{11}$$

2.3.2. *Substrate and product kinetics*

Applying the general model of Sinclair to the substrate and the product, the following expressions are obtained in the acetic fermentation for the overall rate of ethanol consumption, r_E, the overall rate of oxygen consumption, r_O, and the overall acetic acid production rate r_A:

$$r_E = r_{E-X} + r_{E-m} + r_{E-P} + r_{E-AcEt} \tag{12}$$

$$r_O = r_{O-X} + r_{O-m} + r_{O-P} \tag{13}$$

$$r_A = r_{A-X} - r_{A-AcEt} \tag{14}$$

The various rates on the right-hand sides of Equations (12)–(14) are defined in Appendix A2. The mean contribution is due to the assimilative way, related to the rates of ethanol and oxygen consumption for the production of biomass, r_{E-X} and r_{O-X} respectively, and the rate of acetic acid excretion by biomass, r_{A-X}. It is reasonable to considerer that in the acetic fermentation the energetic requirements of the cells are basically due to the multiplication process. Therefore these expressions can be simplified:

$$r_E = r_{E-X} \tag{15}$$

$$r_O = r_{O-X} \tag{16}$$

$$r_A = r_{A-X} \tag{17}$$

It is very usual to make a link between the product formation–substrate consumption and the cell growth. Such substrate and product kinetics is usually called *growth associated* [6]:

$$\frac{dE}{dt} = \left(-\frac{1}{Y'_{X/E}}\right)\mu_g X_v \tag{18}$$

$$\frac{dO}{dt} = \left(-\frac{1}{Y'_{X/O}}\right)\mu_g X_v \tag{19}$$

$$\frac{dA}{dt} = Y_{A/E}\frac{1}{Y'_{X/E}}\mu_g X_v \tag{20}$$

$$Y'_{X/O} = Y_{E/O}Y'_{X/E} \tag{21}$$

In Equations (18)–(21) the relationship between the cell growth and the product formation–substrate consumption is established by $Y'_{X/E}$ and $Y'_{X/O}$, the yield factors of biomass/ethanol and biomass/oxygen respectively, and by $Y_{A/E}$ and $Y_{E/O}$, the stoichiometric coefficients of acetic acid/ethanol and ethanol/oxygen respectively.

2.3.3. *Optimization parameters*

Assuming that the industrial data show a decrease in the acetic production only due to the ethanol consumption, we propose a function for μ_g based exclusively on the ethanol concentration. There is no oxygen concentration factor, because in the industrial fermentation processes the oxygen demand is always satisfied by a constant aeration and agitation:

$$\mu_g = K_1\frac{1}{K_2 + (1/E)^n} \tag{22}$$

We consider μ_d to be constant, since it is logical to evaluate an average vital cycle from the average cell approximation applied to the model.

The initial viable biomass/total biomass ratio R_{Xv_i} is an initial parameter in the simulation. For this reason, it must be optimized together with the kinetic parameters.

All in all, the problem to be solved was the optimization of five parameters of the model: K_1, K_2, n, μ_d and R_{Xv_i}.

3. GENETIC ALGORITHM APPLIED TO KINETIC MODELLING

The algorithm designed in this work is an evolution of the basic genetic algorithm (GA) adapted to the features of the problem to be solved. The basic GA can nowadays be easily consulted in the literature [12,13]. Many authors have adapted the basic algorithm to very different problems; for example, Pizarro Millán *et al*. [10] applied it to feature selection and Potocnik and Grabec [15] applied it to empirical kinetic modelling. This proves the flexibility of the algorithm and its applicability potential in the future.

3.1. Description of the genetic algorithm developed

The most important technical features of the algorithm are as follows.

1. Each chromosome represents a possible combination of values of the five parameters to optimize, in binary code. There is an allowed range of values for each parameter to adopt, as many values as allowed by the binary codification and the number of significant figures. The initial population is composed of randomly selected values for the parameters within the allowed ranges and codified into binary code.
2. The evaluation program decodes the values of the parameters for each chromosome and then uses them

to simulate a batch process with each sequence of parameters. The simulation algorithm solves the system of differential equations formed by Equations (8)–(10) and (18)–(20) by the Runge–Kutta algorithm [21]. The initial concentrations are those of the representative sequence of the process, and the initial viable biomass/total biomass ratio is that of the parameters of the chromosome. There is an oxygen control in the simulation, because in the real process the oxygenation conditions are enough to satisfy the oxygen demand.

The simulation algorithm has two important stop conditions.

(a) The simulation is stopped when no real positive values for one concentration are obtained.
(b) The simulation is stopped when the process time in the simulation has reached the final process time of the representative sequence.

In Section 3.2 it is shown how these stop conditions are important to compare the simulated acetic concentrations with the concentrations of the representative sequence by means of a desirability function.

3. A new generation with the same number of chromosomes is formed by applying reproduction, crossover and mutation operators. The chromosomes with the best fitting ability obtain the best value in the desirability function, i.e. closer to 1, and have more chances of being selected and copied into the next generation. This probability is expressed for chromosome i of n chromosomes as

$$\text{prob}(i) = \frac{\text{response}(i)}{\sum_{i=1}^{i=n} \text{response}(i)} \tag{23}$$

Uniform crossover is used and the five best chromosomes of each generation pass unchanged to the next generation. These chromosomes are called elitist chromosomes.

4. Twins and out-of-range chromosomes are disallowed by using a 'while' loop with filters. When some of these chromosomes are discovered after crossover, they are substituted by chromosomes also obtained by crossover, and if they are discovered after mutation, they are replaced by the original chromosomes in the same positions but mutated again with the same chances of mutation. With this process the mutability is not increased and the number of chromosomes remains constant.
5. The process stops after five generations without changes higher than a fixed percentage of the mean response of the elitist chromosomes.
6. The algorithm is completed five times each time the program is run. A final run where the initial population is composed of the best chromosomes found in each of the previous runs is performed.
7. All the programs have been written in Matlab 6.1.0.450 (The MathWorks, Inc.).

In Table I the configuration parameters of the algorithm are shown. The configuration has been developed to solve the critical points, as will be explained in the following subsection, since it is difficult to find a standard methodology to obtain the optimal settings for a particular problem, and therefore experience and knowledge of the influence of the different parameters are usually applied [7].

Table I. Configuration parameters of the genetic algorithm developed

Population	30 chromosomes
Mutation	1.5%
Elitism	Five elitist chromosomes in each population
Stop condition	Five generations without changes higher than 5% in the mean response of the elitist chromosomes

3.2. Critical points

In this subsection the critical points in the algorithm are exposed. These critical points have influenced very much the architecture of the program.

3.2.1. First critical point: exploitation ability of the algorithm

One of the most important goals in building a genetic algorithm is to obtain the highest exploitation; in our case, always within the ranges of values allowed for the parameters. We have achieved this by the following.

- The highest possible elitism.
- A quite high probability of mutation.
- A limited population size.
- Filtering out-of-range parameters. It is very important to concentrate the efforts on the most important regions. This is particularly critical in the last runs, where a global exploration has just been made and it is necessary to study the response in a short range.
- Filtering twins. In the last runs of the program it is also very important to avoid twins. Running the program without this filter caused the algorithm to provide us with final populations with a large number of repeated chromosomes, particularly with short allowed ranges of parameters, thus wasting energy owing to a useless time-consuming process.
- A final run with the best chromosomes, as reported before, is a refinement of the best solutions found.

By applying these considerations, we have avoided the premature meeting of a relative maximum in the response.

3.2.2. Second critical point: convergence

The algorithm is less time-consuming and convergence is achieved by the following.

- The stop conditions in the simulation, which have been reported before.
- Filtering out-of-range parameters and twins.
- Limiting the number of runs without changes in the response.
- Introducing a minimum percentage to consider that there have been changes in the mean response of the elitist chromosomes.

3.2.3. Third critical point: response

The response must be able to select the best set of kinetic parameters, which provides the highest predictive ability to the model, obtaining simulated acetic concentration sequences similar to the industrial one.

The mere use of a root mean square is not enough to compare the concentration sequences, since it does not provide information on the differences in the shape of the data sequence, the final concentration reached and the process time.

To solve the problem, we propose a four-factor desirability function as the response to maximize:

$$R = \left[\left(\frac{1}{1+\mathrm{RMS}}\right)\left(\frac{t_s}{t_r}\right)\left(\frac{1}{1+\mathrm{abs}(A_s - A_r)}\right) \times \left(\frac{1}{1+\mathrm{abs}[a\tan(B_s) - a\tan(B_r)]}\right)\right]^{\frac{1}{4}} \quad (24)$$

where RMS is the root mean square of the differences in the acetic acid concentrations between the simulated and the representative sequence, t_s is the process time in the simulation (h), t_r is the process time in the real sequence (h), A_s is the final acetic acid concentration reached in the simulation ($g\,l^{-1}$), A_r is the final acetic acid concentration reached in the real sequence ($g\,l^{-1}$), B_s is the slope of data in the simulated sequence with a process time range of 0–20 h and B_r is the slope of data with a linear behaviour in the real sequence (0–20 h).

Three main features of the desirability functions make them particularly useful as a tool to be applied to our problem.

1. The desirability functions are very restrictive, because if only one single function gives a zero value, the global value of the function is zero. This behaviour has been very useful to penalize seriously the sets of parameters providing simulations with concentrations lacking any analytical sense, since they are stopped without reaching the final process time and the final acetic concentration of the representative sequence, even if the RMS values are not so bad considering the scarcity of points to compare. In a similar way the sets of parameters that provide extremely fast simulated processes are also penalized. The sets of parameters providing very slow simulated processes are penalized, since the simulation is always stopped when the final process time of the representative sequence is reached, and thus the RMS is very high and the acetic concentration reached is very low.
2. On the other hand, all the single functions must give maximum values to obtain a maximum global value, conferring a more selective power on the evaluation function, particularly with simulated sequences with the same RMS but with very different shapes.
3. The desirability functions are scaled between 0 and 1, allowing us to compare objectively the fitting of each simulation to the representative sequence.

In conclusion, the main advantages of working with the desirability function instead of using only the root mean square are as follows.

- Better discrimination abilities compared with the RMS method.
- Evaluation function scaled between 0 and 1.
- The sets of kinetic parameters providing extremely long or short process times in the simulation are penalized.
- The sets of kinetic parameters reaching a final acetic concentration far from the real sequence are penalized.
- The sets of kinetic parameters providing sequence data with a linear segment in the simulation similar to the data with a linear behaviour in the real data sequence are rewarded.

4. EXPERIMENTAL DATA SETS

4.1. Industrial data set

The acetic concentrations in the fermentators of the industrial plant of Vinagrerías Riojanas SA, obtained by NIR, were studied. The data were obtained for a period of 4 months without changes in the industrial parameters of the process, i.e. oxygenation conditions and temperature. The average temperature was 29.5°C and the oxygenation conditions were enough to satisfy the oxygen demand, and thus the oxygen became a non-limiting substrate.

Nowadays the fermentators of the industrial plant work discontinuously with charges. The batch bioreactors studied were always fed with white wine of the same origin. The process time was about 30–31 h and 218 complete sequences were obtained.

An average concentration sequence was calculated by analysing the data. This sequence is representative of the process to be modelled. The variability in the concentrations among the sequences is due to analytical errors and to factors that cannot be controlled in an industrial process, i.e. differences in the ethanol concentration of the wine among batch processes. Therefore the model obtained with this sequence does not model this variance. Figure 2 shows the average concentration sequence.

4.2. Pilot processes

Two different continuous processes were developed in the pilot system to test the predictive ability of the model: one without cell recycling and another with cell recycling. All the experiments were carried out in a BioFlo IV fermentator of 10 l capacity. The fermentator was fed with white wine of the same origin as that used in the industrial plant; the wine was provided by the company. The parameters of the two processes were: temperature, 29.5°C; agitation, 400 rpm; aeration, 2.20 SLPM; overpressure, 0.5 atm. The oxygenation conditions were improved to meet the oxygen demand, even when there was an increase in cell population due to the effects of the recycling.

Figure 3 shows the configuration of the experimental system. In the steady state, where all the concentrations within the vessel are independent of time, it is possible to apply the following balance to any component i of the system:

$$F(C_{if} - C_i) + Vr_{fi} = 0 \quad (25)$$

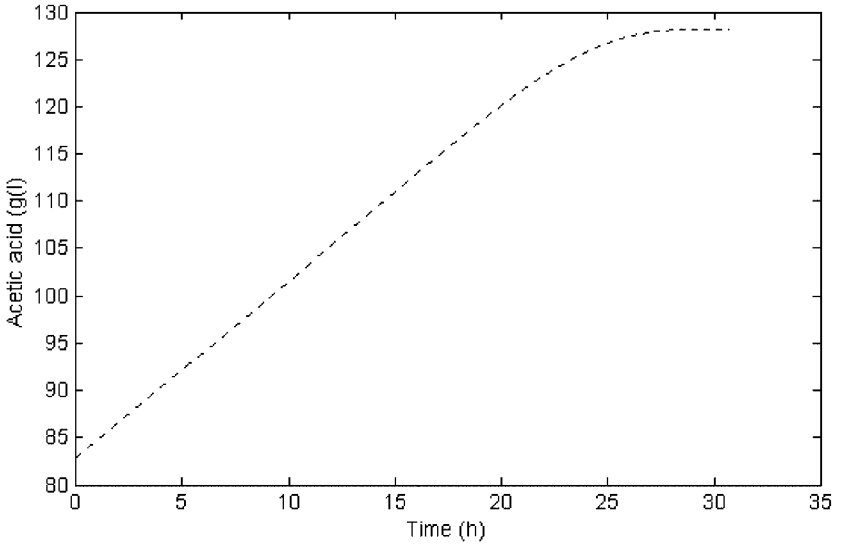

Figure 2. Average data sequence calculated from the industrial concentrations.

where F is the volumetric flow rate of the feed and effluent liquid streams, C_{if} is the component i concentration in the feed stream, C_i is the component i concentration in the bioreactor and in the effluent stream, V is the reactor volume and r_{fi} is the rate of formation of component i within the reactor.

Rearranging the expression:

$$r_{fi} = \frac{F}{V}(C_i - C_{if}) = D(C_i - C_{if}) \tag{26}$$

The parameter D is called the dilution rate and characterizes the holding time or processing rate of the continuous reactor.

The feed stream was under proportional, integrative and derivative (PID) control to maintain the pH constant and therefore to maintain the concentrations of ethanol and acetic acid constant. When a constant stream feed was observed for more than 100 h and the oxygen dissolved in the medium was constant, it was considered that a steady state had been reached.

From Equation (26) we realize that when the steady state is reached, the rate of consumption or production of compound i may be easily evaluated. In our experiments the volume of the reactor was maintained constant at 7 l, and the acetic and ethanol concentrations in the wine stream were measured off-line and the concentrations inside the fermentator were measured on-line, using a FOSS NIRSystems 5000 liquid analyser previously calibrated. In the experiments the residual ethanol was maintained at about 20 g l^{-1}. The volumetric flow rate was easily calculated by measuring the wine consumption during a given period of time. In this way the acetification rate and the ethanol consumption rate were evaluated for the continuous process with and without cell recycling.

Working with cell recycling, we had to manage two streams of effluents, which we have called F_1 and F_2 in Figure 3; these are the filtered vinegar stream and the purged vinegar stream respectively. It is obvious that

$$F = F_1 + F_2 \tag{27}$$

and that the concentration of ethanol and acetic acid does not change between the two streams. Therefore the recycling rate R is calculated as

$$R = \frac{F_1}{F} \tag{28}$$

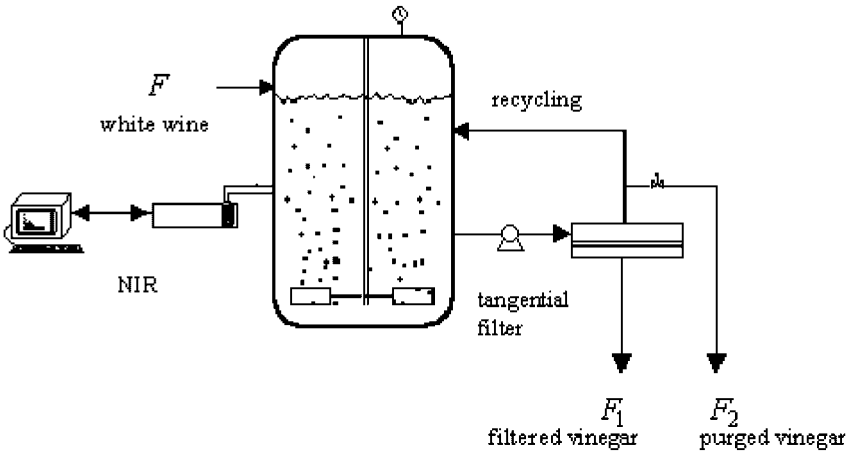

Figure 3. Pilot fermentation system.

Table II. Optimized set of kinetic parameters

Kinetic parameter	Optimized value
K_1	5.60×10^{-5}
K_2	6.30×10^{-4}
n	3.16
μ_d	$3.09 \times 10^{-2}\,h^{-1}$
R_{Xv_i}	0.63

The continuous process with cell recycling was developed with a recycling rate of 0.4.

5. OPTIMIZATION RESULTS

After running the program several times, an optimized set of parameters was obtained. The ranges allowed for n, μ_d and R_{Xv_i} were the logical ones from the point of view of the biological knowledge of the model proposed i.e. [1–10] for n, [0.01–0.1] for μ_d and [0–1] for R_{Xv_i}. The ranges for K_1 and K_2 were very large in the first runs, $[1 \times 10^{-6}$–$1 \times 10^{6}]$ for both parameters, to explore the whole experimental domain and to find the areas with the best predictive ability. Then we studied these areas to find the best zone, and gradually the allowed range was shortened until we obtained the ranges with which we performed the last runs, i.e. $[1 \times 10^{-5}$–$1 \times 10^{-4}]$ for K_1 and $[1 \times 10^{-4}$–$1 \times 10^{-3}]$ for K_2. The optimized set of parameters is shown in Table II.

6. PREDICTIVE ABILITY OF THE MODEL OBTAINED

The model obtained is able to explain the evolution of a batch fermentation under the temperature and oxygenation conditions of the industrial process modelled. Furthermore, if we want to test the real usefulness of the model from an engineering point of view, we should test the predictive ability of the model with different bioprocess designs. Therefore the model was introduced in a simulation environment that reproduces the conditions of those continuous processes with and without cell recycling that were developed in the pilot fermentator, as reported in Section 4.2.

The simulation programs were developed in Matlab 6.1.0.450 (The MathWorks, Inc.).

6.1. Prediction of the pilot process: continuous process without cell recycling

The simulation output of the continuous process without cell recycling, using the optimized kinetic model, is shown in Figure 4. The simulation starts with a batch process with the initial concentrations of the representative sequence. When the ethanol concentration reaches $20\,g\,l^{-1}$, owing to the ethanol consumption, the program simulates the PID control of the feed stream, which is programmed with the same concentrations of ethanol and acetic acid as in the white wine used in the pilot processes. In Figure 4, when the ethanol concentration is controlled at $20\,g\,l^{-1}$, the concentrations of viable and total biomass and acetic acid are also maintained constant. If we study the simulated profiles of growth rate and acetification rate, they are stabilized after the simulated PID control starts. From these results it can be inferred that the steady state has been reached in the simulation, as it was reached with the pilot system described in Section 4.2. The feed stream profile it is not shown in the figure, but it is not difficult to deduce that it will be constant when the steady state is reached.

Before the simulated feed stream flow is started, the concentration profiles correspond with a batch process. The viable and total biomass concentrations and the acetic acid concentration increase and the ethanol concentration decreases owing to the logical evolution of the fermentation. The simulated acetification profile in this stage is adjusted to the acetification profile of the representative sequence in the first 15 h.

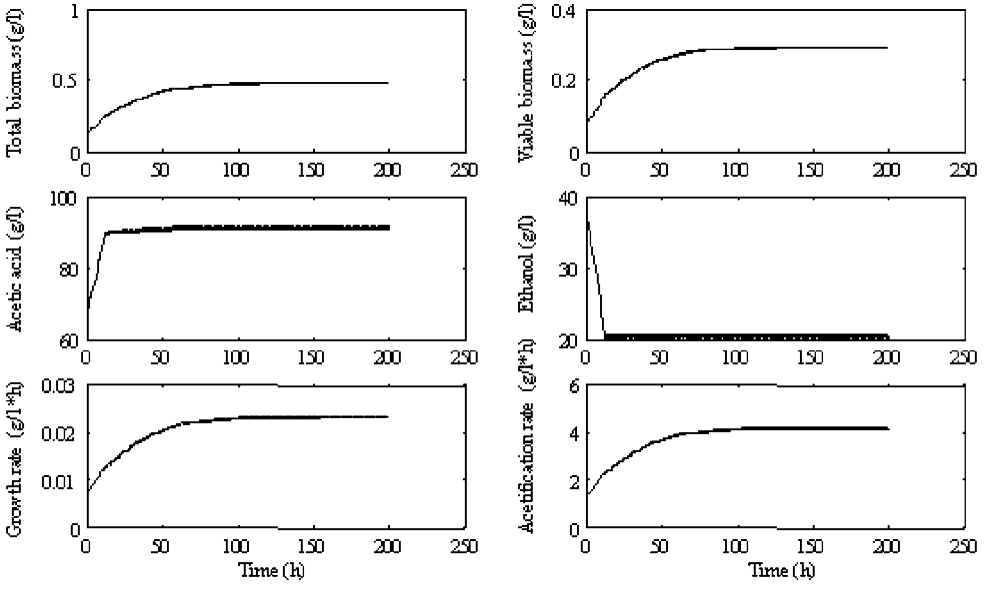

Figure 4. Simulation of a continuous process without cell recycling.

Table III. Simulation parameters of the continuous processes

	Without cell recycling	With cell recycling
Fermentator volume (l)	25000	25000
Residual ethanol ($g\,l^{-1}$)	20	20
Recycling ratio	0	0.4
Simulation time (h)	200	200

Table IV. Simulation results of the continuous processes

	Without cell recycling	With cell recycling
Average feed flow ($l\,h^{-1}$)	1131	1773
Average acetification rate ($g\,l^{-1}\,h^{-1}$)	3.91	6.13
Average feed flow in last 10 h ($l\,h^{-1}$)	1225	2000
Average acetification rate in last 10 h ($g\,l^{-1}\,h^{-1}$)	4.24	6.92
Average acetification rate in last 10 h obtained with pilot fermentator ($g\,l^{-1}\,h^{-1}$)	4.12 ± 0.21	6.90 ± 0.35

The simulation parameters are reported in Table III and the results obtained by simulation can be seen in Table IV.

The experimental results with the pilot system showed a steady state with an acetification rate of $4.12 \pm 0.21\,g\,l^{-1}\,h^{-1}$, similar to the value obtained in the simulation, i.e. $4.24\,g\,l^{-1}\,h^{-1}$. In conclusion, the simulation with the model optimized is able to explain the pilot process.

6.2. Prediction of the pilot process: continuous process with cell recycling

The simulation output of the continuous process with cell recycling, using the optimized kinetic model, is shown in Figure 5. The simulation program is similar to the program described in Section 6.1, but a cell recycling ratio is introduced when the feed flow starts. We can make similar considerations about the concentration profiles to those made with Figure 4. The simulation parameters and the results obtained by simulation are also reported in Tables III and IV respectively.

The experimental results with the pilot system showed a steady state with an acetification rate of $6.90 \pm 0.35\,g\,l^{-1}\,h^{-1}$, similar to the value obtained in the simulation, i.e. $6.92\,g\,l^{-1}\,h^{-1}$. We can also conclude that the simulation with the model optimized is able to explain the pilot process.

Furthermore, it was expected to observe an increase in the acetification rate due to the cell recycling, as was obtained with the real and the simulated process. Comparing Figures 4 and 5, it is also clear that the total and viable biomass concentrations reached in the simulated steady state with cell recycling are higher than those reached without cell recycling. This is shown in a similar way in Table IV, as the average feed flow is higher in the process with cell recycling, owing to the increase in the cell population and therefore in the ethanol requirements.

All in all, not only the results of the simulation using the optimized kinetic model agree with the real ones, but both of them demonstrate clearly the increase in productivity that could be reached by the development of processes with cell recycling in the industrial plant.

7. CONCLUSIONS AND FURTHER RESEARCH

The present work shows a successful optimization of the kinetic parameters of the model proposed by applying the genetic algorithm developed.

Only a representative sequence of the acetic acid concentrations from the industrial fermentators was needed; no biomass concentrations were required. The desirability function used as evaluation function was able to select the best set of parameters for the model.

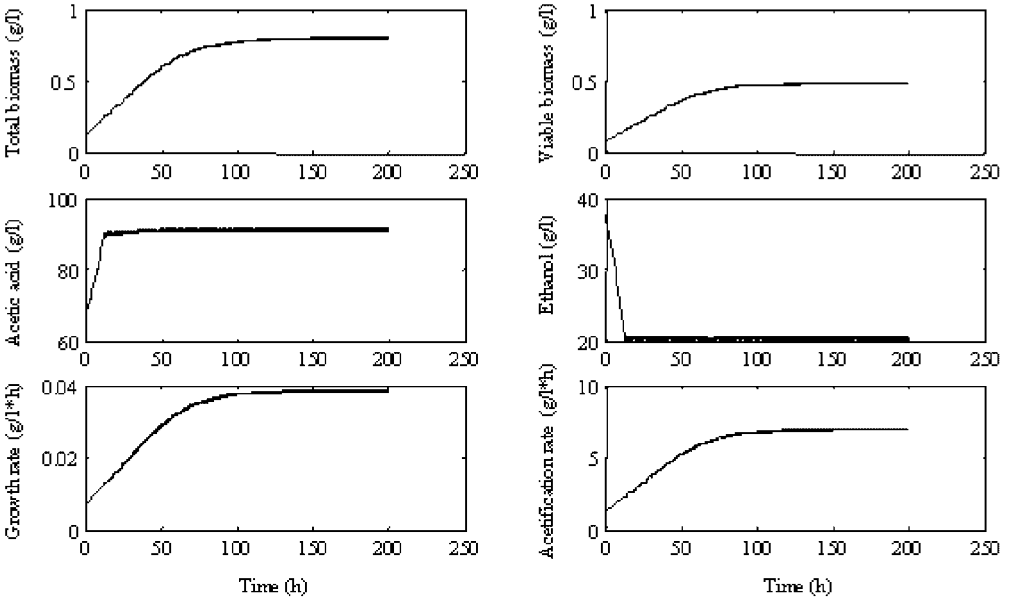

Figure 5. Simulation of a continuous process with cell recycling.

The model obtained predicts the behaviour of the industrial process of acetification. The model was used successfully in the simulation of a pilot continuous process with and without recycling. The results obtained in the simulation were a good prediction of the experimental ones.

Given the good results obtained with the model, it is currently being used to develop new industrial systems of fermentation. As a matter of fact, it has been demonstrated by simulations with the model optimized that the process with cell recycling increases the production of acetic acid, as was inferred from the pilot processes.

If the desirability function proposed is adapted, the algorithm can be used to obtain an economic optimization of the process. Depending on the goal, some factors to optimize other parameters, such as aeration, agitation, cooling or energy consumption, should be included in the response.

Acknowledgements

The authors acknowledge the Spanish Ministry of Agriculture and Food (INIA CAL01-053) and the Autonomous Government of La Rioja (Consejería de Educación, Juventud y Deportes, research grant FPI-1999) for their financial support.

APPENDIX. NOMENCLATURE

A1. Evaluation function

A_r	final acetic acid concentration reached in real sequence ($g l^{-1}$)
A_s	final acetic acid concentration reached in simulation ($g l^{-1}$)
B_r	slope of data with linear behaviour in real sequence (0–20 h)
B_s	slope of data in simulated sequence with process time range of 0–20 h
d_{cn}	value of single desirability function for response y_{cn}
D_n	value of global desirability function for sample n
f_c	kind of single desirability function chosen for selected criterion c
RMS	root mean square of differences in acetic acid concentrations between simulated and representative sequence
t_r	process time in real sequence (h)
t_s	process time in simulation (h)
y_{cn}	response of sample n for criterion c

A2. Kinetic model

A	acetic acid concentration ($g l^{-1}$)
E	ethanol concentration ($g l^{-1}$)
O	oxygen concentration ($g l^{-1}$)
r	observed cell growth rate ($gDW l^{-1} h^{-1}$)
r_A	overall acetic acid production rate ($g l^{-1} h^{-1}$)
r_{A-AcEt}	acetic acid consumption rate by ethyl acetate formation ($g l^{-1} h^{-1}$)
r_{A-X}	rate of acetic acid excretion by biomass ($g l^{-1} h^{-1}$)
r_d	overall cell death rate ($gDW l^{-1} h^{-1}$)
r_e	rate of consumption of organic cell matter in endogenous respiration ($gDW l^{-1} h^{-1}$)
r_E	overall rate of ethanol consumption ($g l^{-1} h^{-1}$)
r_{E-AcEt}	ethanol consumption rate by ethyl acetate formation ($g l^{-1} h^{-1}$)
r_{E-m}	ethanol consumption rate for production of maintenance energy ($g l^{-1} h^{-1}$)
r_{E-P}	ethanol consumption rate for formation of products ($g l^{-1} h^{-1}$)
r_{E-X}	ethanol consumption rate for production of biomass ($g l^{-1} h^{-1}$)
r_g	overall cell growth rate ($gDW l^{-1} h^{-1}$)
r_l	cell lysis rate ($gDW l^{-1} h^{-1}$)
r_O	overall rate of oxygen consumption ($g l^{-1} h^{-1}$)
r_{O-m}	oxygen consumption rate for production of maintenance energy ($g l^{-1} h^{-1}$)
r_{O-P}	oxygen consumption rate for formation of products ($g l^{-1} h^{-1}$)
r_{O-X}	oxygen consumption rate for production of biomass ($g l^{-1} h^{-1}$)
r_P	overall product formation rate ($g l^{-1} h^{-1}$)
r_S	overall rate of substrate consumption ($g l^{-1} h^{-1}$)
r_{S-m}	substrate consumption rate for production of maintenance energy ($g l^{-1} h^{-1}$)
r_{S-P}	substrate consumption rate for formation of products ($g l^{-1} h^{-1}$)
r_{S-X}	substrate consumption rate for production of biomass ($g l^{-1} h^{-1}$)
R_{Xv_i}	initial viable biomass/total biomass ratio
X_d	death biomass concentration ($gDW l^{-1}$)
X_t	total biomass concentration ($gDW l^{-1}$)
X_v	viable biomass concentration ($gDW l^{-1}$)
$Y_{A/E}$	acetic acid/ethanol stoichiometric coefficient
$Y_{E/O}$	ethanol/oxygen stoichiometric coefficient
$Y'_{X/E}$	biomass/ethanol yield factor
$Y'_{X/O}$	biomass/oxygen yield factor
μ	observed specific growth rate (h^{-1})
μ_d	overall specific death rate (h^{-1})
μ_g	overall specific growth rate (h^{-1})

A3. Pilot fermentation system

C_i	component i concentration in bioreactor and in effluent stream ($g l^{-1}$)
C_{if}	component i concentration in feed stream ($g l^{-1}$)
D	dilution rate (h^{-1})
F	volumetric flow rate of feed and effluent liquid streams ($l h^{-1}$)
F_1	filtered vinegar stream ($l h^{-1}$)
F_2	purged vinegar stream ($l h^{-1}$)
R	recycling rate
r_{fi}	rate of formation of component i within reactor ($g l^{-1} h^{-1}$)
V	reactor volume (l)

REFERENCES

1. Bogaerts Ph, Hanus R. On line state estimation of bioprocesses with full horizon observers. *Math. Comput. Simul.* 2001; **56**: 425–441.
2. Diaz C, Lelong Ph, Dieu P, Feuillerat C, Salomé M. On-line analysis and modelling of microbial growth using a hybrid system approach. *Process Biochem.* 1999; **34**: 39–47.

3. Petrides D. Biopro designer: an advanced computing environment for modelling and design of integrated biochemical processes. *Comput. Chem. Eng.* 1994; **18**: 621–625.
4. Cooney CL, Petrides D, Barrera M, Evans L. Computer-aided design of a biochemical process. *ACS Symp. Ser.* 1988; **39**: 39–61.
5. Evans LB, Field RP. Requirements of a biotechnology process simulator. *Proc. World Congr. II of Chemical Engineers*, 1986; 890–893.
6. Bailey JE, Ollis DF. *Biochemical Engineering Fundamentals* (2nd edn). McGraw-Hill: New York, 1986.
7. Massart DL, Vandeginste BGM, Buydens LMC, de Jong S, Lewi PJ, Smeyers-Verbeke J. *Handbook of Chemometrics and Qualimetrics*: *Part A*. Elsevier: Amsterdam, 1997; 783–802.
8. Reijmers TH, Wehrens R, Buydens LMC. Quality criteria of genetic algorithms for construction of phylogenetic trees. *J. Comput. Chem.* 1999; **20**: 867–876.
9. Huang M-W, Hsieh CC, Arora JS. A genetic algorithm for sequencing type problems in engineering design. *Int. J. Numer. Methods Eng.* 1997; **40**: 3105–3115.
10. Pizarro Millán C, Forina M, Casolino MC, Leardi R. Extraction of representative subsets by potential functions methods and genetic algorithms. *Chemometrics Intell. Lab. Syst.* 1998; **40**: 33–52.
11. Leardi R. Application of genetic algorithm–PLS for feature selection in spectral data sets. *J. Chemometrics* 2000; **14**: 643–655.
12. Lucasius CB, Kateman G. Understanding and using genetic algorithms. Part 1. Concepts, properties and context. *Chemometrics Intell. Lab. Syst.* 1993; **19**: 1–33.
13. Lucasius CB, Kateman G. Understanding and using genetic algorithms. Part 2. Representation, configuration and hybridization. *Chemometrics Intell. Lab. Syst.* 1994; **25**: 99–145.
14. Wehrens R, Buydens LMC. Evolutionary optimization: a tutorial. *Trends Anal. Chem.* 1998; **17**: 193–203.
15. Potocnik P, Grabec I. Empirical modelling of antibiotic fermentation process using neural networks and genetic algorithms. *Math. Comput. Simul.* 1999; **49**: 363–379.
16. Hendriks MMWB, de Boer JH, Boey P. Multicriteria decision making. *Chemometrics Intell. Lab. Syst.* 1992; **16**: 175–191.
17. Keller HR, Massart DL, Brans JP. Multicriteria decision making: a case study. *Chemometrics Intell. Lab. Syst.* 1991; **11**: 175–189.
18. Berget I, Næs T. Sorting of raw materials with focus on multiple end-product properties. *J. Chemometrics* 2002; **16**: 263–273.
19. Sinclair GC, Kristiansen B. *Fermentation Kinetics and Modelling* (1st edn). Open University Press: London, 1987.
20. Sinclair CG, Topiwala HH. Model for continuous culture which considers the viability concept. *Biotechnol. Bioeng.* 1970; **12**: 1069–1079.
21. Mathews JH, Kurtis DF. *Numerical Methods Using MATLAB* (3rd edn). Prentice-Hall: Englewood Cliffs, NJ, 1999.

