Journal of Industrial Microbiology, 13 (1994) 83–89

Published by The Macmillan Press Ltd

Use of an Extended Kalman Filter and development of an automated system for xylose fermentation by a recombinant *Escherichia coli*

A. K. Hilaly, M. N. Karim and J. C. Linden

Department of Agricultural and Chemical Engineering, Colorado State University, Fort Collins, CO 80523, USA

(Received 3 May 1993; revision received 18 October 1993; accepted 4 November 1993)

Key words: Extended Kalman Filter; Xylose fermentation; State estimation; Data acquisition; Automation and control

SUMMARY

An automated system was developed for on-line monitoring and control of xylose fermentation by a recombinant *Escherichia coli*. A 7-L fermenter was interfaced with a personal computer. Control circuits were constructed and a software was developed to estimate the states of the fermentation using an Extended Kalman Filter. The automated system combined with the Extended Kalman Filter provided a satisfactory way to obtain on-line information regarding estimation of fermentation parameters.

INTRODUCTION

Ingram et al. [2] constructed a recombinant *Escherichia coli* to ferment xylose to ethanol. Optimal operational schemes can be developed by analyzing the dynamics of the system and applying mathematical optimization techniques. There are two methods of implementing an optimal strategy: (i) open loop operation and (ii) closed loop operation. In an open loop operation, a predetermined policy is applied without any information about the current states of the system. It is presumed that the pattern of the behavior of the system remains the same as that in past experiments. The effectiveness of an open loop operation is, therefore, not insured. In a closed loop operation, an optimal decision is taken based on the knowledge of the current states of the system. Though relatively complicated, this method of operation is obviously better. In this paper, we describe the construction of an automated system for on-line estimation of the states of a recombinant fermentation. This automated system was used to implement optimization schemes in a feed-back mode. The results of the optimization study will be reported in a separate article.

This work was divided into two parts: (i) design and construction of the necessary hardware and (ii) development of a software for estimating the states of the fermentation. A 7-L fermenter (New Brunswick Scientific Inc., Edison, NJ, USA, model 110) was interfaced with a personal computer via a data acquisition card. This fermenter did not have any sophisticated instrumentation. Appropriate electrical circuits were designed to transduce the signals from various sensors (with different ranges not compatible with the data acquisition card) to input to the computer and to convert signals sent from the computer to activate different control elements. A computer program was written with a user-friendly interface. The on-line measurements were displayed graphically on the screen. Based upon the on-line measurements, the program estimated the states of the system using the Extended Kalman Filtering theory [1,4,5]. The estimated states were also displayed graphically on the screen.

Correspondence to: M.N. Karim, Department of Agricultural and Chemical Engineering, Colorado State University, Fort Collins, CO 80523, USA.

EXPERIMENTAL METHODOLOGY

Strain

Escherichia coli ATCC 11303 containing the plasmid pLOI297 was used in this study. This plasmid was a derivative of pUC18. The genes coding for pyruvate decarboxylase and alcohol dehydrogenase from *Zymomonas mobilis* were cloned in the plasmid. Ingram et al. [2] reported the details of the plasmid in the literature.

Medium and cultivation

Cultures were grown in Luria Broth containing xylose. The composition of Luria Broth was: 10 g L^{-1} bactotryptone, 5 g L^{-1} yeast extract and 5 g L^{-1} NaCl. Tetracycline (10 mg L^{-1}) was added to the medium to prevent the growth of plasmid-free cells. The temperature and pH were controlled at 34 °C and 6.5 respectively.

Cell concentration

Growth was monitored by measuring the optical density of the broth at 550 nm. A Bausch and Lomb (Rochester, NY, USA) model 20 spectrophotometer was used for measuring optical density. Cell dry weight was determined by centrifuging a fixed volume of liquid and drying the cell pellets in aluminum pans at 80 °C for 12 h.

Fermentation product analysis

Fermentation broth samples (10 ml) were centrifuged and the cell-free broths collected. Each clarified broth was then filtered through a 0.45-μm membrane filter (Millipore, Milford, MA, USA). The filtered broth was analyzed by HPLC (Waters Associates, Inc., Milford, MA, USA, model 501) with a differential refractive index detector (model Waters 410). An organic acid column (Showa Denko K.K., model SHODEX 1011 marketed by Waters) was used. The column was operated at 43 °C. The mobile phase was a filtered and degassed solution of 0.01 N H_2SO_4. The flowrate of the mobile phase was 1 ml min^{-1}. A sample of 25 μl was injected by a Hamilton injector. A Data Module (model Waters 746) was used to quantify the chromatographs based on a calibration table. The calibration table was constructed by injecting known concentrations of each component of the fermentation broth.

THE DATA ACQUISITION SYSTEM

A New Brunswick Microferm (model 110) fermenter was used in this study. Control circuits were designed to control the temperature, pH and feed flowrate by an IBM compatible computer based on a 286 microprocessor. A data acquisition card (Qua Tech Inc., Akron, OH, USA, SAC-12) was used to interface the fermenter with the computer.

There were three control circuits: (i) temperature control circuit, (ii) pH control circuit and (iii) feed flowrate control circuit. Circuits were also developed for processing the signals from an off-gas CO_2 analyzer (Infrared Industries Inc., Santa Barbara, CA, USA, model 703D), a gas flowmeter (Omega, Stamford, CT, USA, model FMA-219) and an electronic balance (A&D Engineering, Malipitas, CA, USA, model EK-1200A) which measured the amount of alkali added. Figure 1 shows a schematic diagram of the data acquisition system.

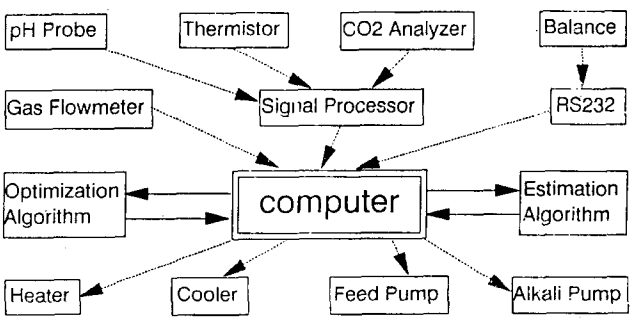

Fig. 1. A schematic diagram of the computer-controlled fermenter.

TEMPERATURE CONTROL

Temperature was controlled by a sequential operation of a heater and a cooling water assembly. A thermistor probe was used to measure the temperature. The resistance of a thermistor decreases with an increase in temperature. A circuit was designed to maintain a constant current through the thermistor. The change in voltage drop across the thermistor was, therefore, proportional to the change in the resistance. The voltage output (0–5 V) from the circuit was fed into the computer through the SAC-12 data acquisition card. The software program filtered the signal noise. The output from the control program was converted back into the analog voltage to activate several relays which sequentially turned on and off the heater and the solenoid valve for cooling water. A cascade relay system was used for the heater since it operated at a relatively high current of 6 A.

pH CONTROL

The pH was monitored by a pH electrode. The voltage output of the pH electrode was in the mV range. The output from the pH electrode was amplified in the range of 0–5 V and fed to the computer through the SAC-12 card. The output from the control program sent a digital signal through the SAC-12 card to turn on or off a Watson Marlow (Cornwall, UK) pump (model 501U) for adding alkali to the fermenter. The fermentation studied in this work produced acid, and hence the control of pH required alkali addition only.

FEED RATE CONTROL

The output from the digital to analog converter of the SAC-12 card was in the range of 0–5 V. This signal was sent to a Watson Marlow pump (model 501U) set in the remote control mode. A correlation between the feed flowrate and the signal voltage was obtained from experimental data. The control program used this correlation to maintain the desired feed flowrates.

MONITORING ALKALI ADDITION

A flask containing an alkali solution (4 N sodium hydroxide) for pH control was placed on an electronic balance (A&D Engineering, model EK-1200A). The signal of the balance was made available to the computer through the RS-232 serial port. A null modem, with pin 2 and pin 3 interchanged, was used to implement the hardware handshaking. A string processing subroutine was written to filter the strings sent by the balance for extracting the required numerals.

MONITORING CO_2 IN THE EFFLUENT GAS

A carbon dioxide analyzer (Infrared Inc., model 703D) was used for monitoring the % CO_2 in the effluent gas from the fermenter. The CO_2 analyzer provided a voltage output

in the mV range. This signal was proportional to the % CO_2 in the effluent gas. An amplifier was designed to amplify the voltage signal from the analyzer in the range of 0–5 V. The amplified signal was then sent to the computer via the SAC-12 card. A filtering subroutine eliminated the noise from the amplifier. Nitrogen was used as a carrier gas at a flowrate of 2 L min^{-1}. The flowrate of the effluent gas was measured by a gas flowmeter (Omega, model FMA-219), which provided a voltage output proportional to the gas flowrate. This voltage signal was sent to the computer via the SAC-12 card.

THE SOFTWARE

The software program executed four main functions: (i) acquired the signals from various sensors; (ii) estimated the states of the fermentation; (iii) computed the decision variables and sent control signals accordingly and (iv) displayed graphically the on-line measurements and estimated states.

SIGNAL ACQUISITION

The analog signals from the CO_2 analyzer circuit, temperature control circuit and pH control circuit were input to the computer via the SAC-12 card. In order to filter the noise of the signals, it was assumed that the noises were white and Gaussian in nature and therefore, their mean should be zero. The noisy signal $X(j)$ from a circuit may be expressed as

$$X(j) = Y(j) + n(j) \tag{1}$$

where $n(j)$ denotes a noise and $Y(j)$ is the true signal at any time interval j. If the signal is accessed for N times, we may write,

$$\sum X(j) = \sum Y(j) + \sum n(j), \quad j = 1,2,...N \tag{2}$$

Now $\Sigma\, n(j)/N = 0$ when N is large, and one may write

$$\sum X(j)/N = \sum Y(j)/N \tag{3}$$

The above equation shows that the noise term is eliminated and hence the mean of a noisy signal over a finite time represented the mean of the noise-free signal.

As mentioned before, an electronic balance was connected to the computer through the RS-232 serial port. The balance was made by the manufacturer for one-way data transmission; it could only send signals out and was incapable of receiving any signal. This caused difficulty in running different control algorithms sequentially. The serial port input buffer filled up very quickly by the incessant signal flow from the balance. The problem was circumvented by clearing the input buffer frequently (approximately every 10 s).

THE CONTROLLER

The temperature controlling program compared the setpoint and measured temperature to decide the control action. The fermenter was equipped with a constant-current heating element. There was no mechanism to vary the input current in the heating element. Therefore, an on-off controller was designed. However, with a simple on-off algorithm, the controller performance was poor (about ±1 °C fluctuations). The control algorithm was modified to turn the heater or the cooler on or off at the borders of a 'window' surrounding the setpoint. The upper and lower borders of the window was +0.1 °C and −0.1 °C respectively relative to the setpoint. With this technique, the controller performance improved significantly (±0.2 °C).

The pH controller was also an on-off type. Through trial and error, the duration of the 'on' cycle was adjusted (1.0 s) to get a satisfactory controller performance (pH ±0.01). The feed flowrate was manipulated to regulate the substrate concentration in fed-batch fermentations. The program sent a signal to the feed pump in the range of 0–5 V to deliver the desired flowrates.

THE USER INTERFACE

A computer program was written to provide an effective user interface for monitoring the measurements and states at any time. The interface also allowed the user to change operating conditions whenever necessary. The graphical interface consisted of three windows. The first window plotted all the signals from the measurement devices, namely, temperature, pH, weight of alkali in the alkali-flask, flowrate of effluent gas and % CO_2. The second window plotted the time course of the estimated states, such as cell mass concentration, xylose concentration, ethanol concentration and total acids concentration. The third window provided several options to the user for changing the operating conditions, such as temperature set-point, pH set-point and feedrate. There was also a provision for updating the estimated states by actual HPLC analysis data. The program also saved the data on the time course of states, measurements and operating conditions.

THE STATE ESTIMATOR

The states of the fermentation were estimated by the algorithm reported by Yatawara et al. [5]. The algorithm was derived from the Extended Kalman filtering theory.

In real time experiments, signals from various sensors are subject to disturbances. Bad measurements due to intermittent sensor problems may bias the estimation of a Kalman Filter. The method of Yatawara et al. [5] can be used to design a filter resistant to bad measurements. The probability of an observation to be a bad measurement was calculated. Two parallel filters, having a lower and a higher gain, were designed. When the probability was high, the lower gain filter dominated the estimation process. The states and covariances of error of estimation were computed

as the weighted average of the two filters. The details of the method can be found in the literature [5]. The summary of the algorithm is described below.

Suppose the behavior of the system is described by nonlinear continuous equations

$$dx/dt = f(x(t), u(t)) + w(t) \tag{4}$$

and the equations for discrete measurements are

$$z(k) = h(x(k)) + v(k) \tag{5}$$

where x denotes the state variables; f represents a function; w and v are noise vectors; h represents a function and z denotes the measurements.

(i) Predict states $x_{k+1/k}$ at the time interval $k + 1$ using the values of the state variables at interval k

$$x_{k+1/k} = x_{k/k} + \int_k^{k+1} [f(x(t), u(t)]dt \tag{6}$$

(ii) Predict the covariance matrix of estimation error $P_{k+1/k}$ at the time interval $k + 1$ using the values at interval k

$$P_{k+1/k} = P_{k/k} + \int_k^{k+1} [J(k) \cdot P_{k/k} + P_{k/k} \cdot J^T(k) + Q(k)]dt \tag{7}$$

where J denotes the Jacobian matrix of the model equations and Q is the covariance matrix of the noise vector w. The Jacobian matrix is defined as

$$J(k) = \frac{\delta f(x(t), u(t))}{\delta x} \tag{8}$$

(iii) For $i = 0$ and 1, calculate the following

$$x_{k+1/k+1,i} = x_{k+1/k} + R_i^{-1} \cdot P_{k+1/k} \cdot H_{k+1}^T \cdot [z_{k+1} - h(x_{k+1/k})] \tag{9}$$

$$P_{k+1/k+1,i} = P_{k+1/k} - R_i^{-1} \cdot [P_{k+1/k} \cdot H_{k+1}^T \cdot H_{k+1} \cdot P_{k+1/k}] \tag{10}$$

$$R_i = \phi_i^2 + H_{k+1} \cdot P_{k+1/k} \cdot H_{k+1}^T \tag{11}$$

$$\phi_i^2 = (1-i)\,\phi_\epsilon^2 + i \cdot \phi_\gamma^2 \tag{12}$$

where ϕ_ϵ^2 is the variance of observation noise; ϕ_γ^2 is the variance of a Gaussian random variable and H represents the observation matrix. When $i = 1$, it is a case of bad measurement. The observation matrix is defined as

$$H(k) = \frac{\delta h(x)}{\delta x} \tag{13}$$

(iv) Calculate the posterior probabilities

$$\kappa_i = (\sqrt{2\pi} \cdot \phi_i)^{-1} \cdot K_i \cdot \sqrt{P_{k+1/k+1,i}} \cdot \exp[(-1/2) \{z_{k+1} - h(x_{k+1/k})\}^T \cdot R_i^{-1} \{z_{k+1} - h(x_{k+1})\}]/\sqrt{P_{k+1/k}} \tag{14}$$

where K_i is a constant.

(v) Update the state vector and the covariance matrix

$$x_{k+1/k+1} = \sum \kappa_i \cdot x_{k+1/k+1,i} \quad \text{for } i = 0,1\,. \tag{15}$$

$$P_{k+1/k+1} = \sum \kappa_i \cdot P_{k+1/k+1,i} \quad \text{for } i = 0,1\,. \tag{16}$$

(vi) Go back to step (i).

The estimation algorithm required a model of the fermentation. The model equations of the fermentation are described below.

$$\frac{dX}{dt} = \mu \cdot X \cdot (1 - X/X_{max})^{n1} - F \cdot X/V \tag{17}$$

$$\frac{dS}{dt} = -\alpha_1 \cdot dX/dt - \beta_1 \cdot X \cdot (1 - E/E_{max})^{n2} + F \cdot (S_f - S)/V \tag{18}$$

$$\frac{dE}{dt} = \alpha_2 \cdot dX/dt + \beta_2 \cdot X \cdot (1 - E/E_{max})^{n3} - F \cdot E/V \tag{19}$$

$$\frac{dA}{dt} = \alpha_3 \cdot dX/dt + \beta_3 \cdot X \cdot (1 - E/E_{max})^{n3} - F \cdot A/V \tag{20}$$

where X is the cell mass concentration; S is the xylose concentration; E is the ethanol concentration; A is the sum of organic acid (succinic, lactic and acetic acid) concentrations; μ is the specific cell growth rate; α_i is the specific production (or consumption) rate of component i in the growth phase and β_i is the specific production (or consumption) rate of component i in the non-growth phase; n1, n2 and n3 are constants; X_{max} and E_{max} are the maximum cell concentration and ethanol concentration in the logistic model; F is the feed flowrate.

The specific growth rate was expressed as a function of cell mass concentration and substrate concentration

$$\mu = \mu_{max} \frac{S}{f(S)} \tag{21}$$

where μ_{max} is the maximum specific growth rate; $f(S)$ is a polynomial of substrate concentration. The polynomial $f(S)$ was expressed as

$$f(S) = a_0 + a_1 \cdot S + a_2 \cdot S^2 + a_3 \cdot S^3 \tag{22}$$

The constants a_i were found by a nonlinear regression of experimental data. The parameters α_i and β_i were determined by the method reported by Ollis [3]. The values of various parameters are listed in Table 1.

Two on-line measurements, namely the % CO_2 in the effluent gas and the amount of alkali added, were used for estimating the states. These measurements were related with ethanol and organic acid production respectively. The functional relationship between the measurements and the states are developed below.

From the metabolic pathway of xylose fermentation, one may write

TABLE 1

Model parameters for Eqns 17–22

Parameter	Value
a_0	−8.06
a_1	2.39
a_2	−4.95e − 02
a_3	4.78e − 04
α_1	3.0
β_1	1.0
α_2	3.1
β_2	0.4
α_3	0.5
β_3	0.1
n1	1/3
n2	1/4
n3	1/2
μ_{max}	0.138
X_{max}	3.1
E_{max}	38.0

$$C_3H_4O_5 \longrightarrow CO_2 + CH_3CHO$$
(pyruvate) (acetaldehyde)

$$CH_3CHO + NADH + H^+ \longrightarrow C_2H_5OH + NAD^+$$
(acetaldehyde) (Ethanol)

When the pH of the culture medium is kept constant, the amount of CO_2 dissolved in the aqueous phase remains unaltered. Under the conditions of constant pH, therefore, CO_2 evolution is proportional to ethanol formation. If σ_j denotes the volume of CO_2 produced at time interval j, one may write

$$\sigma_j = 0.5\,(PCO_{j-1} + PCO_j) \cdot G_j \cdot \delta t_j/100 \tag{23}$$

where PCO represents the % CO_2 measured by the CO_2 analyzer and G denotes the total gas flowrate. The sum of σ_j is related with ethanol production, and one may write

$$\sum \sigma_j = \omega\,(E)$$
$$= \theta_0 + \theta_1 \cdot E + \theta_2 \cdot E^2 \tag{24}$$

where E denotes ethanol concentration and θ_i are the coefficients of a polynomial. The values of θ_i were determined by a nonlinear regression of experimental data.

A considerable amount of succinic acid, acetic acid and lactic acids were also produced during the course of the fermentation. If $[HA]_j$ denotes moles of undissociated acid, one may write

$$[HA]_j = [H^+]_j + [A^-]_j \tag{25}$$

The dissociation constant K_j is

$$K_j = \frac{[H^+]_j[A^-]_j}{[HA]_j} \tag{26}$$

After several steps of mathematical simplification, one can write

$$[H^+] = 1/2[-K_j + (K_j^2 + 4K_jC_j)^{1/2}] \tag{27}$$

where $C_j = [HA]_j + [H^+]_j$.

The dissociation of sodium hydroxide and neutralization reaction may be represented as

$$[NaOH] \rightarrow [Na^+] + [OH^-]$$
$$[OH^-] + [H^+] \rightarrow H_2O$$

Each kind of acid produced a certain amount of $[H^+]$ depending upon the value of K_j. The only relevant data available was the total amount of sodium hydroxide consumed in the neutralization reaction. Therefore, it was not possible to quantify the amount of individual acids. Only the sum of the concentrations of the individual acids was quantified. The relation between the total amount of alkali added (ζ_j) and the sum of the concentrations of all acids (A) at instant j was expressed as

$$\zeta_j = Y(A)$$
$$= \xi_0 + \xi_1 \cdot A + \xi_2 \cdot A^2 + \xi_3 \cdot A^3 \tag{28}$$

The constants ξ_i were determined by a nonlinear regression of experimental data.

Equations 24 and 28 provided the functional relations to construct the observation matrix H. Equation 13 was used to determine the elements of the matrix H.

PERFORMANCE OF THE FILTER

The performance of the filter given by Eqn 15 is shown in Figs 2 and 3. Usually values of initial P and Q matrices are unknown, and they are chosen to be diagonal. The values of initial covariances P and Q are listed in Table 2.

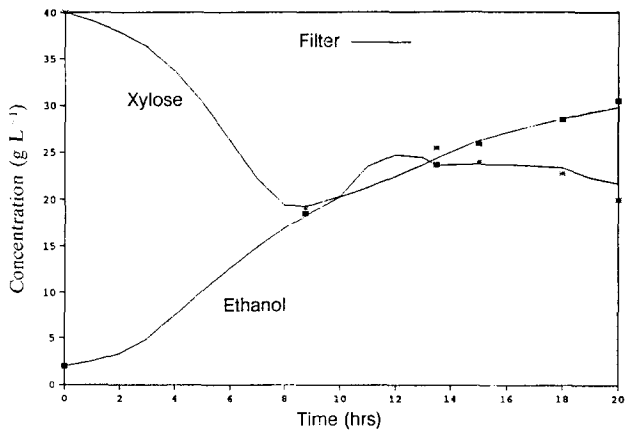

Fig. 2. The estimated concentration profiles of xylose and ethanol.

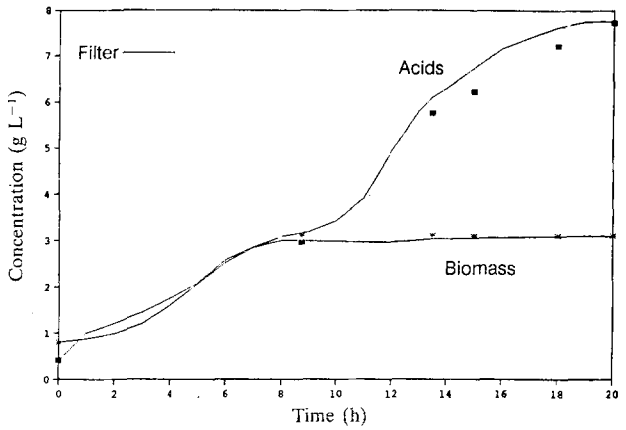

Fig. 3. The estimated concentration profiles of organic acids and biomass.

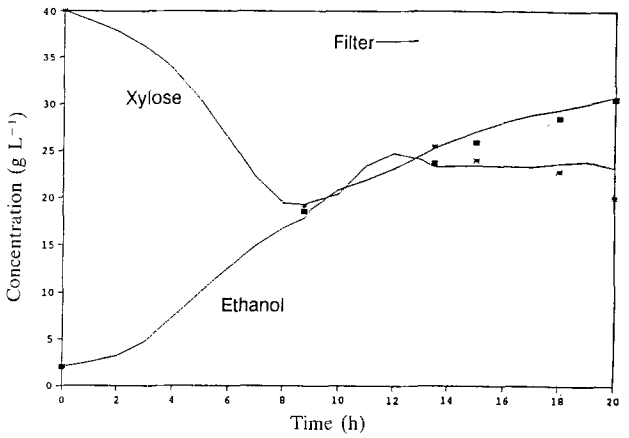

Fig. 4. The estimated concentration profiles of xylose and ethanol in the presence of sporadic faulty sensor signals.

TABLE 2

Initial values of covariance matrices P and Q

Matrix	Elements
P	diagonal (0.01,0.5,0.05,0.05)
Q	diagonal (0.002,0.5,0.05,0.05)

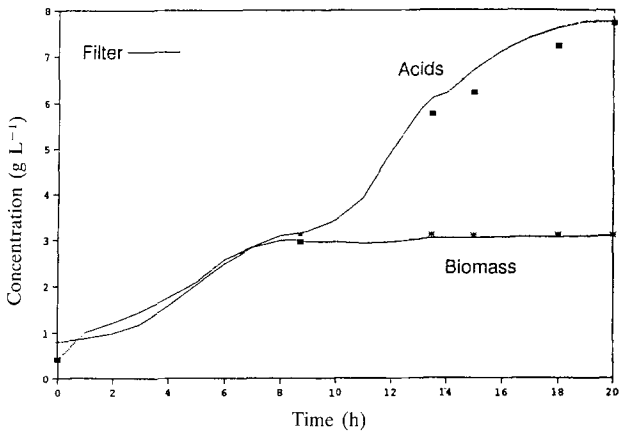

Fig. 5. The estimated concentration profiles of organic acids and biomass in the presence of sporadic faulty sensor signals.

The values were selected by trial and error. The ratio of variances of the two parallel filters were chosen to be 2.0. Yatawara et al. [5] theoretically showed that the filter designed by Eqn 15 is robust for different values of the ratio of variances of the two parallel filters.

The fermentation was initially carried out in the batch mode. After 9 h, feed was initiated. The feed flowrate was equal to the value of the expression $(\alpha_1 \cdot dX/dt + \beta_1 \cdot X \cdot)V/(S_f - S)$. This feedrate was chosen to maintain the xylose concentration in the range 25–30 g L^{-1}. The filter predictions for cell mass concentration, ethanol concentration and total acids concentration were in good agreement with experimental data. However, in the case of xylose, there was a little discrepancy between filter prediction and the experimental data.

In order to test the performance of the filter in the presence of faulty sensor readings, disturbances (200% of actual reading) were introduced in the CO_2 analyzer signal at 9, 12 and 15 h. Figures 4 and 5 illustrate the result. The filter predictions deteriorated only slightly. Therefore, the filter was sufficiently resistant against sporadic bad measurements.

The states estimated by the Extended Kalman Filter was used to implement optimization schemes. The accuracy and robustness of the estimator is very important for optimal performance of a fermentation. A filter with no outlier protection may totally spoil a fermentation run. The estimator designed in this study provided a means to carry out fermentations optimally.

CONCLUSIONS

In this paper, we described the development of an automated system to carry out computer controlled fermentations. The automated system controlled operating conditions (temperature and pH) and estimated the states of the fermentation. The regulation of the temperature and pH was satisfactory. The estimator performed reasonably well in the presence of sporadic bad sensor signals. It is very convenient to implement control and optimization schemes in a computer-controlled fermenter. In this study, we demonstrated that it is not difficult to convert a fermenter without any sophisticated instrumentation into a computer-controlled fermenter with the capability of estimating the states of the fermentation. By following our procedure, fermentation engineers at the industries may upgrade the existing fermenters.

REFERENCES

1 Bellgardt, K.H., W.I. Kuhlmann, H.D. Meyer, K. Schugerl and M. Thoma. 1986. Application of an extended Kalman filter for

state estimation of a yeast fermentation. IEEE Proc. 133(5): 226–234.

2 Ingram, L.O., T. Conway, D.P. Clark, G.W. Sewell and J.F. Preston. 1987. Genetic engineering of ethanol production in *Escherichia coli*. Appl. Environ. Microbiol. 53: 2420–2425.

3 Ollis, D.F. 1983. A simple batch fermentation model: theme and variation. Ann. New York Acad. Sci. 413: 144–157.

4 San, K.Y. and G. Stephanopoulos. 1984. Studies on on-line bioreactor identification. I. Theory. Biotechnol. Bioeng. 26: 1176–1188.

5 Yatawara, N., B. Abraham and J.F. MacGregor. 1991. A kalman filter in the presence of outliers. Commun. in Statist. – Theory Meth. 20(5): 1803–1820.