Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Nonlinear Model Predictive Control applied to E. Coli Cultures

G. Hafidi*, S. Tebbani*, D. Dumur*, A. Vande Wouwer**

** Supélec, Control Department, F 91192 Gif sur Yvette cedex, France (e-mail: ghizlane.hafidi@ supelec.fr, sihem.tebbani@supelec.fr, didier.dumur@supelec.fr).*
*** Faculté Polytechnique de Mons, Control Department, 7000 Mons, Belgium (e-mail: alain.vandewouwer@fpms.ac.be)*

Abstract: This paper proposes a nonlinear model predictive control scheme with particular application to the regulation of acetate concentration in order to maximize the biomass growth during fed-batch cultures of *E. Coli*. A reference feed rate which enables to maintain the acetate concentration at a specified level is first determined by means of optimal conditions. This feed rate is further used as the reference control trajectory within the NMPC algorithm. Finally, to avoid discretization problems during NMPC application, the on-line optimization is moved into a nonlinear programming strategy using the control vector parameterization approach (CVP). Some simulation results obtained on a fed-batch *E. Coli* bioreactor validate the efficiency of the proposed control strategy.

1. INTRODUCTION

The bacterium *Escherichia coli* represents the universal cell-factory for the fermentative production of bio-pharmaceuticals. Several valuable products such as recombinant proteins, antibiotics and amino acids are today commercially produced using fermentation techniques and there is an enormous economic incentive to optimize such processes. However, in industry, the control of bioreactors is most often limited to basic pH, temperature and partial pressure of dissolved oxygen and dissolved carbon dioxide regulations (Diaz *et al*., 1996). There is, however, no doubt that computer control of the biochemical state variables can help increasing the process performance significantly. In this way, (Jenzsch *et al*., 2005) proposed a generic model control of the specific growth rate in recombinant *E. Coli* cultivations. Another approach is concerned with the control of acetate concentration using model-based adaptive linearizing control by reducing the model (Bastin *et al*., 1990, Rocha, 2003). Other structures exist which apply linear predictive control to a linearized model, combining robustification features to deal with system nonlinearities and disturbances (Renard *et al*., 2006).

In this direction, this paper proposes a nonlinear model predictive control structure which aims at maximizing the biomass productivity of cultures of *E. Coli* through acetate concentration control, by manipulating the feed rate. Indeed, the presence of acetate affects the productivity; the growth of biomass being inhibited when acetate is present at high concentrations. The developed NMPC law includes on the one hand a reference to be followed by the acetate concentration and on the other hand a reference trajectory for the control signal, so called optimal reference feed rate. This original model reference NMPC appears to be a powerful strategy managing the difficulty of the related non-linear problems. Furthermore, complexity of on-line optimization related to NMPC for this kind of process is avoided through control vector parameterization (CVP).

The paper is organized as follows. Section 2 reminds the main steps leading to the elaboration of the dynamical model of *E. Coli* cultures. The theoretical background required to formulate the model reference NMPC structure is presented in Section 3, from the optimal reference feed rate design to the acetate concentration regulation constraining the feed rate to follow this optimal feed rate. This section also examines the CVP approach implemented in real time to decrease the complexity of the NMPC optimization. This control strategy is applied to fed-batch cultures of *E. Coli* in Section 4. Finally, some conclusions and perspectives are given in Section 5.

2. PROCESS MODELLING

The growth of the bacteria *E. Coli* follows the bottleneck theory (Rocha, 2003; Hollywood *et al*., 1976; Gray et al., 1966; Amarasingham *et al*., 1965; Xu *et al*., 1999; Luli *et al*., 1990). Indeed, *E. Coli* has a limited respiratory capacity. Acetate is produced when glucose exceeds the respiratory capacity, which corresponds to the oxido-fermentative regime. Acetate is consumed when glucose is less than the respiratory capacity; which is the oxidative regime. When the quantity of glucose exactly fills the respiratory capacity, the system operates in optimal conditions. This case corresponds to the edge between the two regimes, when acetate is not produced nor consumed (Fig. 1). It will be seen in further developments that the aim of the control is to force the *E. Coli* culture to remain at the edge between these regimes.

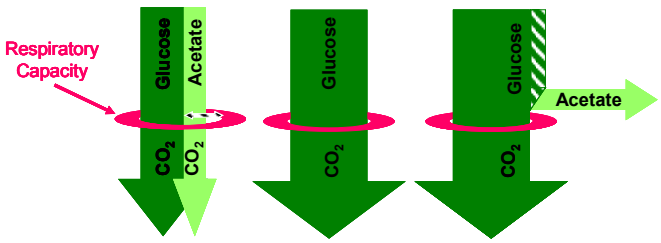

Fig. 1. Schematic representation of the bottleneck theory.

Following the previous scheme, the metabolism of *E. Coli* is described through three macroscopic reactions (Rocha, 2003; Galvanauskas *et al*., 1998; Cockshott *et al*., 1999):

- Glucose oxidation:

$$k_1 S + k_5 O \xrightarrow{r_1} X + k_8 C \quad (1)$$

- Glucose fermentation:

$$k_2 S + k_6 O \xrightarrow{r_2} X + k_9 C + k_3 A \quad (2)$$

- Acetate oxidation:

$$k_4 A + k_7 O \xrightarrow{r_3} X + k_{10} C \quad (3)$$

where S, O, X, C and A represent glucose (substrate), oxygen, biomass, carbon dioxide and acetate respectively (in the sequel S, O, X, C and A will represent concentrations). k_i are the stoichiometric coefficients. r_j are the growth rates. Reactions (1) and (3) describe the oxidative regime, reactions (1) and (2) the oxido-fermentative regime. The relationship between the growth rates and the specific growth rate μ_j, which depends on the operating regimes (Table 1), is as follows (Bernard, 2002):

$$r_j = \mu_j X \quad (4)$$

Table 1. Specific growth rates depending on the regimes

Specific growth rate	Oxido-fermentative regime	Oxidative regime
μ_1	$\frac{q_{S,crit}}{k_1}$	$\frac{q_S}{k_1}$
μ_2	$\frac{(q_S - q_{S,crit})}{k_2}$	0
μ_3	0	$\frac{q_{AC}}{k_4}$

Based on the previous reaction scheme, mass balances yield the following dynamic model (Bastin *et al*., 1990):

$$\frac{d}{dt}\begin{bmatrix} X \\ S \\ A \\ O \\ C \end{bmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ -k_1 & -k_2 & 0 \\ 0 & k_3 & -k_4 \\ -k_5 & -k_6 & -k_7 \\ k_8 & k_9 & k_{10} \end{pmatrix} \begin{bmatrix} \mu_1 X \\ \mu_2 X \\ \mu_3 X \end{bmatrix} - \frac{F_{in}}{W} \begin{bmatrix} X \\ S \\ A \\ O \\ C \end{bmatrix} + \begin{bmatrix} 0 \\ S_{in}\frac{F_{in}}{W} \\ 0 \\ OTR \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ CTR \end{bmatrix} \quad (5)$$

$$\frac{dW}{dt} = F_{in}$$

where W is the culture medium weight (kg), F_{in} is the influent flow-rate (kg/h), S_{in} is the influent glucose concentration (g/kg), OTR is the oxygen transfer rate from gas to liquid phase, CTR the carbon dioxide transfer rate from liquid to gas phase.

Considering the bottleneck assumption, the kinetic terms associated with the glucose consumption q_S, the critical glucose specific uptake rate $q_{S,crit}$ and the specific acetate uptake rate q_{AC} are expressed by the relations:

$$q_S = q_{S\max}\frac{S}{K_S + S} \quad (6)$$

$$q_{S\,crit} = \frac{q_{O\max}}{k_{OS}}\frac{K_{i,O}}{K_{i,O} + A} \quad (7)$$

$$q_{AC} = q_{AC\max}\frac{A}{K_A + A}\frac{K_{i,A}}{K_{i,A} + A} \quad (8)$$

where $q_{S\max}$, $q_{O\max}$ and $q_{AC\max}$ are the maximum specific growth rates, K_S and K_A are the saturation constants of substrate (glucose) and acetate respectively, k_{OS} is the oxygen yield related to glucose, $K_{i,O}$ and $K_{i,A}$ are the inhibition constants related to oxygen uptake and acetate uptake respectively. q_S follows a Monod law and q_{AC} a Monod law multiplied by an inhibition factor, which are both classically used by biologists to characterize the behaviour of uptake rates.

Finally, since oxygen is always regulated to induce no influence on the growth of bacteria, the dynamic model (5) can be formulated in a compact form, reduced to four differential equations, considered as the control model used during the NMPC synthesis:

$$\frac{dX}{dt} = r_X X - \frac{F_{in}}{W}X \quad (9)$$

$$\frac{dS}{dt} = -q_S X - \frac{F_{in}}{W}(S_{in} - S) \quad (10)$$

$$\frac{dA}{dt} = r_A X - \frac{F_{in}}{W}A \quad (11)$$

$$\frac{dW}{dt} = F_{in} \quad (12)$$

denoted as $\dot{x}(t) = f(x(t), F_{in}(t))$ in further developments with $x = [X \quad S \quad A \quad W]^{\mathrm{T}}$. r_X and r_A in (9) and (11) depend on the operating regime and are given in Table 2.

Table 2. Expressions of r_X and r_A

Growth rates	Oxido-fermentative regime	Oxidative regime
r_X	$\frac{(q_S - q_{S,crit})}{k_2} + \frac{q_{S,crit}}{k_1}$	$\frac{q_S}{k_1} + \frac{q_{AC}}{k_4}$
r_A	$k_3\frac{(q_S - q_{S,crit})}{k_2}$	$-q_{AC}$

3. NMPC APPLIED TO E. COLI BIOREACTOR

As mentioned in Section 2, the aim is to regulate the acetate concentration at zero, when the acetate is not produced nor consumed, that is at the border of the two regimes describing the optimal conditions. However, due to sensitivity problems of sensors in a too close neighbourhood of zero, it appears to

be much more realistic to regulate the acetate concentration at a reference value close to zero, but not zero. Based on the simplified model (9), (10), (11), (12) developed in Section 2, the following approach considers model reference NMPC formulation for acetate concentration regulation.

3.1. Problem formulation

The objective is to regulate acetate concentration A to a reference value A_{set} close to zero while constraining the substrate feed rate F_{in} to track a specified feed rate profile F_{ref} which will be determined later. The formulation of the model reference optimization problem without terminal constraints is as follows (Mayne *et al*., 2000):

$$\min_{\chi} \sum_{j=1}^{N} (A_{set_{k+j}} - \hat{A}_{k+j})^2 + \lambda \sum_{j=1}^{N} (F_{in_{k+j-1}} - F_{ref_{k+j-1}})^2 \quad (13)$$

$$\text{s.t.} \begin{cases} \hat{A}_{k+1} = H\, f(\hat{x}_k, F_{in_k}) \\ \vdots \\ \hat{A}_{k+N} = H\, f(\hat{x}_{k+N-1}, F_{in_{k+N-1}}) \\ F_{in_k} \geq 0, \quad \forall\, k \in IN \\ x_k \geq 0, \quad \forall\, k \in IN \end{cases} \quad (14)$$

where $\chi = \{A_{k+1}, ..., A_{k+N}, F_{ink}, ..., F_{ink+N-1}\}$ is the optimization vector, N is the prediction horizon (which is here assumed to equal the control horizon), $\hat{A}$ is the predicted output, $\hat{x}$ is the predicted state vector, λ is the control weighting factor and $H = [0 \quad 0 \quad 1 \quad 0]$.

Among the whole sequence resulting from the on-line optimization of (13) under nonlinear constraints (14), only the first optimal control is applied as input to the system. At the next sampling time, the current state is obtained and the optimization problem (13)-(14) is solved again with the new initial state value, according to the well-known receding horizon principle.

However, two major problems result when solving this optimization problem (13)-(14): This formulation requires discretization of the system with a small sampling time, so that the discretized model remains significant compared to the continuous one. This leads to a sampling time much too short compared to the time response of the system and the nonlinear constraints increase the on-line computation time of the optimization phase. However the most important drawback is that it may also induce unfeasibility problems.

3.2. Optimization through Control Vector Parameterization

In order to avoid these two problems, the formulation (13)-(14) is moved into a nonlinear programming problem (NLP) by time-discretizing the control actions $F_{in}(t)$ over the prediction horizon, choosing an adequate sampling time much larger than the one which could have resulted from the classical discretization. Furthermore, a piecewise constant approximation of these control actions is considered for the continuous-time computation of the predicted state vector, thus without discretizing the state variables. This approach called Control Vector Parameterization (CVP) (Vassiliadis, 1993) is classically considered also for chemical and biochemical processes (Schlegel *et al*., 2006; Balsa-Canto, 2001). Using this approach, the formulation of the NMPC problem becomes:

$$\min_{\chi'} \sum_{j=1}^{N} (A_{set_{k+j}} - \hat{A}_{k+j})^2 + \lambda \sum_{j=1}^{N} (F_{in_{k+j-1}} - F_{ref_{k+j-1}})^2 \quad (15)$$

$$\text{s.t.} \quad F_{in} \geq 0 \quad (16)$$

where $\chi' = \{F_{in_k}, F_{in_{k+1}}, \ldots, F_{in_{k+N-1}}\}$ is the new optimization vector. The number of constraints has been drastically decreased since the CVP technique implicitly considers model constraints when performing prediction of the state vector. Going one step further to avoid constraint (16), the variable in the optimization vector is moved to:

$$F_{in} = \exp(v) \quad (17)$$

leading to the following minimization problem without constraint anymore:

$$\min_{\chi''} \sum_{j=1}^{N} (A_{set_{k+j}} - \hat{A}_{k+j})^2 + \lambda \sum_{j=1}^{N} (\exp(v_{k+j-1}) - F_{ref_{k+j-1}})^2 \quad (18)$$

where $\chi'' = \{v_k \; v_{k+1}, \ldots, v_{k+N-1}\}$ is the final optimization vector. The overall structure of the developed NMPC strategy is summarized in Fig. 2.

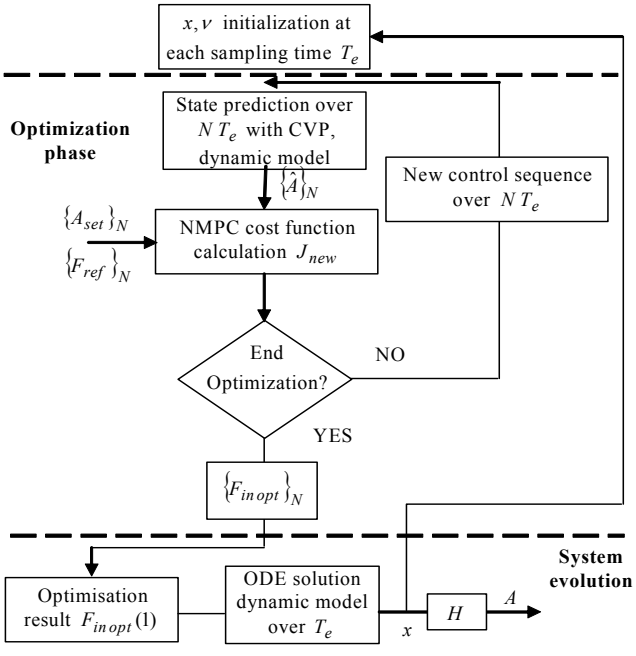

Fig. 2. Overall structure of the NMPC&CVP strategy.

3.3. Determination of the reference feed rate

The objective is to maintain operating conditions to the optimal biological behaviour. In fact, as mentioned before

(Fig. 1), these conditions are satisfied if the system works at the edge between the two operating regimes; in this case, the acetate is not produced, leading to $\mu_2 = 0$ and thus:

$$q_S = q_{Scrit} \tag{19}$$

and the acetate is not consumed, leading to $\mu_3 = 0$ and thus:

$$q_{AC} = 0 \tag{20}$$

The optimal feed rate is thus defined for a unique pair of acetate and glucose concentrations, satisfying (19) and (20):

$$A(t) = 0, \quad S(t) = S_{crit} \tag{21}$$

Condition (21) in (10) leads to $\dot{S} \equiv 0$ which provides the expression of the reference feed rate:

$$F_{ref}(t) = \left. \frac{q_S X(t) W(t)}{(S_{in} - S)} \right|_{S=S_{crit}} \tag{22}$$

From the values in Table 2 and condition (21), it appears that r_X in equation (9) is constant in the two operating regimes. In this case, solving the differential equations (9) and (12) gives the final expression of the reference feed rate:

$$F_{ref}(t) = \left. \frac{q_S X(0) W(0)}{(S_{in} - S)} e^{r_X t} \right|_{S=S_{crit},\, A=0} \tag{23}$$

The same approach can be applied to determine the optimal feed rate when the acetate concentration is assumed to be maintained at a value A_{set} close to zero. Indeed, in this general case, acetate must be produced leading the system to work in the oxido-fermentative regime. In order to maintain A to A_{set}, the operating regime cannot change, which implies, according to the bottleneck theory, that the substrate concentration will also be maintained at a constant value S_{crit}. Similarly to the previous case, the optimal feed rate is thus defined for a unique pair of acetate and glucose concentrations $A = A_{set}$ and $S = S_{crit}$. Combining (10) and (11) when the derivatives are equal to zero, the expression of S_{crit} results from the following equation:

$$(S_{in} - S) \left. \frac{r_A}{A} \right|_{A=A_{ref}} - q_S = 0 \tag{24}$$

This yields to the general expression of the optimal feed rate:

$$F_{ref}(t) = \left. \frac{q_S X(0) W(0)}{(S_{in} - S)} e^{r_X t} \right|_{S=S_{crit},\, A=A_{set}} \tag{25}$$

4. SIMULATION RESULTS

The proposed strategy (model reference NMPC including CVP and change of variable) is now implemented to the acetate concentration $A(t)$ regulation of the *E. Coli* bioreactor. Parameters and initial values considered for simulations are reported in Appendix A and B. The sample time T_e is chosen equal to 2 min. The prediction horizon and weighting factor are respectively equal to $N = 10$ and $\lambda = 1$. The proposed approach has been implemented in the Matlab™ environment, using the optimization LSQNONLIN routine under Matlab R2006a on a 3 Ghz PC with 512 Mo of ram. The chosen setpoint is $A_{set} = 0.5\,\text{g/kg}$. Mean computation time for each control value calculation is 2 *s*.

Figures 3 to 8 show the evolution of the state and control variables over 20 hours (required time to complete the culture). The acetate regulation works well, which was the objective of the control strategy. Before $t = 15\,\text{min}$, the system operates in the oxidative regime as shown in Fig. 8, trying to track the optimal feed rate. After $t = 15\,\text{min}$, the feed rate recovers the optimal trajectory, the operating regime is maintained in the oxido-fermentative regime.

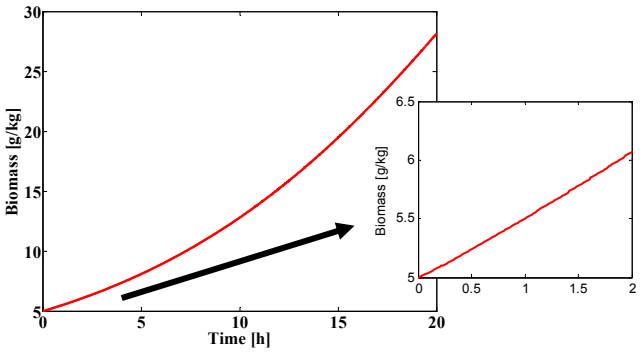

Fig. 3. Biomass concentration evolution over 20 hours.

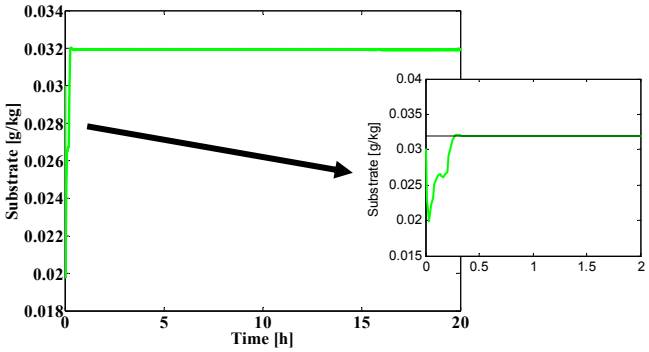

Fig. 4. Substrate concentration evolution over 20 hours.

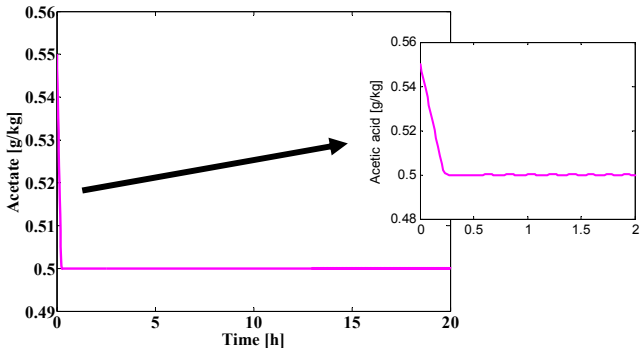

Fig. 5. Acetate concentration evolution over 20 hours.

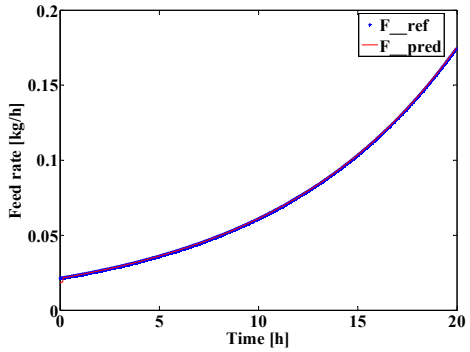

Fig. 6. Feed rate and reference feed rate evolutions over 20 hours.

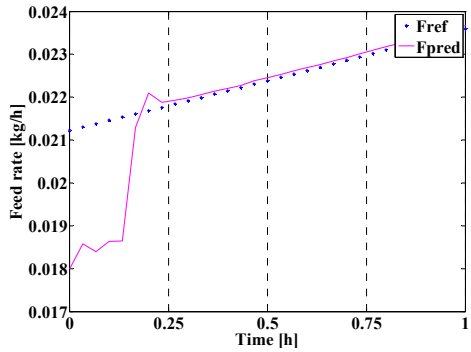

Fig. 7. Feed rate and reference feed rate, zoom over one hour.

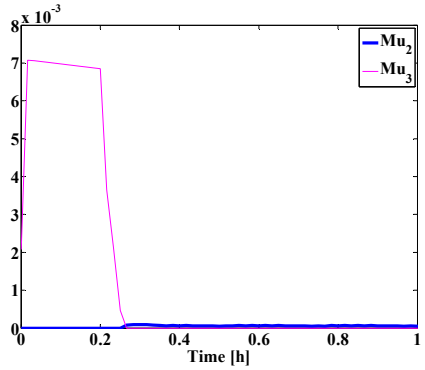

Fig. 8. Growth rates evolution over one hour.

Figures 9 to 12 show the impact of a change of acetate concentration setpoint, occurring at time $t = 10\,\mathrm{h}$, over 20 hours running with a zoom between 8 and 12 hours.

Even if the change of acetate reference value is not realistic since the aim of the regulation is to track a value very close to zero, in fact this scenario shows the adequate tracking behaviour of the developed control strategy. According to equations (24) and (25), Figure 12 illustrates the fact that a unique pair of acetate and substrate define the optimal feed rate. Indeed, after the change of the acetate setpoint, the glucose concentration stabilizes at a lower value, inducing another exponential profile for the feed rate.

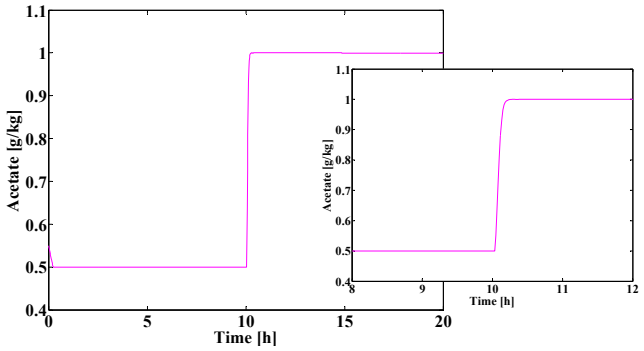

Fig. 9. Acetate concentration evolution over 20 hours.

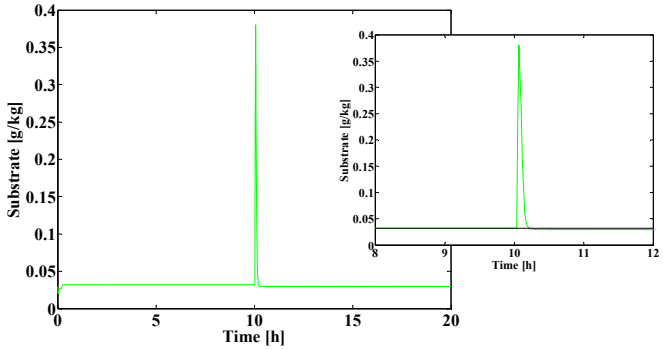

Fig. 10. Substrate concentration evolution over 20 hours.

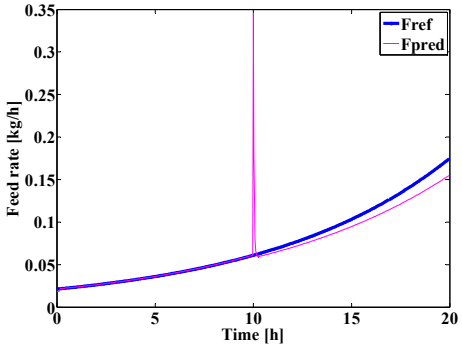

Fig. 11. Feed rate and reference feed rate evolutions over 20 hours.

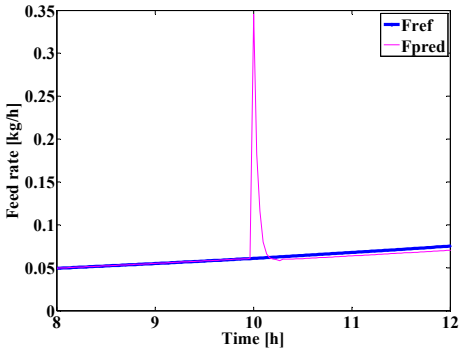

Fig. 12. Feed rate and reference feed rate, zoom between 8 h and 12 h.

5. CONCLUSIONS

This paper proposes a model reference NMPC strategy to control the fed-batch *E. Coli* bioreactor main state variables and achieve regulation of the acetate concentration in order to maximize the biomass productivity. This structure introduces an optimal reference feed rate to be tracked by the bioreactor feed rate, forcing the system to remain at the edge between the two operating regimes. To avoid problems linked with the application of NMPC algorithms, on-line optimization is moved into a nonlinear programming problem through CVP, with only discretizing the control signal. One major advantage of the developed structure is that it can be extended to other types of cultures modelled under the general macroscopic mass-balance model.

Further studies will consider robustness aspects of this control structure with respect to modelling errors or parameters uncertainties, and influence of NMPC tuning parameters. For that purpose, sensitivity functions will be analysed to highlight the most influential parameters.

ACKNOWLEDGMENT

Part of this research was undertaken within the Belgian Network DYSCO (Dynamical Systems, Control and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its author(s).

REFERENCES

Amarasingham, C.J. and B.D. Davis (1965). Regulation of α-ketoglutarate dehydrogenase formation in Escherichia Coli. *J. Biol. Chem*, **240**, pp. 3664-3668.

Balsa-Canto, E., J.R. Banga, A.A. Alonso and V.S. Vassiliadis (2001). Dynamic optimization of chemical and biochemical processes using restricted second-order information. *Computers and Chemical Enginee-ring*, **25**, pp. 539-546.

Bastin, G. and D. Dochain (1990). *On-line estimation and adaptive control of bioreactors*. Elsevier Science Publishers, Amsterdam.

Bernard, O. (2002). Mass balance modelling of bioprocesses. (A. Agrachev. (Ed)). *Int. center for theoretical physics*, Trieste, pp. 769-812.

Cockshott A. and I. Bogle (1999). Modelling the effect of glucose feeding on a recombinant E. coli fermentation. *Biopr. Eng.,* **20**, pp. 83-90.

Diaz, C., P. Dieu, C. Feuillerat, P. Lelong and M. Salome (1996). Simultaneous adaptive predictive control of the partial pressuers of the dissolved oxygen (pO2) and dissolved carbon dioxide (pCO2) in a laboratory-scale bioreactor. *Journal of Biotechnology,* **52**, pp. 135-150.

Galvanauskas, V., R. Simutis, N. Volk and A. Lübert (1998). Model based design of a biochemical cultivation processes. *Bioprocess Engineering*, **18**, pp. 227-234.

Gray, C.T., J.W.T. Wimpenny and M.R. Mossman (1966). Regulation of metabolism in facultative bacteria. II. Effects of aerobiosis and nutrition on the formation of Krebs cycle enzymes in Escherichia Coli. *Biochim. Biophys*. *Acta* , **117**, pp. 33-41.

Hollywood, N. and H.W. Doelle (1976). Effect of specific growth rate and glucose concentration on growth and glucose metabolism of Escherichia Coli K-12. *Microbios* **17**, pp. 23-33.

Jenzsch, M., R. Simutis and A. Luebbert (2006). Generic model control of the specific growth rate in recombinant Escherichia Coli cultivations. *Journal of Biotechnology,* **122**, pp. 483-493.

Luli, G.W. and W.R. Strohl (1990). Comparison of growth, acetate production, and acetate inhibition of Escherichia Coli strains in batch and fed-batch fermentations. *Applied and Environmental Microbiol.,* **56**, pp. 1004-1011.

Mayne, D.Q., J.B. Rawlings, C.V. Rao and P.O. Scokaert (2000). Constrained model predictive control: stability and optimality. *Automatica*, **36**, pp. 789-814.

Renard, F., A. Vande Wouwer, S. Valentinotti and D. Dumur (2006). A practical robust control scheme for yeast fed-batch cultures – An experimental validation, *Journal of Process Control*, **16**, pp. 855-864.

Rocha, I. (2003). Model-based strategies for computer-aided operation of recombinant E. Coli fermentation. *PhD thesis,* Universidade do Minho.

Schlegel, M., A. Hartwich and W. Marquardt (2006). Adaptive control vector parameterization for nonlinear model-predictive control. *IFAC Workshop on NMPC for fast Systems*, Grenoble, France.

Vassiliadis, V.S. (1993). Computational solution of dynamic optimization problems with general differential-algebraic constraints. *PhD thesis*, Imperial College, University of London, London, UK.

Xu, B., M. Jahic and S.O. Enfors (1999). Modeling of overflow metabolism in batch and fed-batch cultures of Escherichia Coli. *Biotechnology Progress*, **15**, pp. 81-90.

Appendix A. PARAMETER VALUES

Parameters	Values	Units
k_1	3.164	g/kg
k_2	25.22	g/kg
k_3	10.9	g/kg
k_4	6.382	g/kg
$q_{S\max}$	1.832	g/(kg.h)
K_S	0.1428	g/(kg.h)
k_{OS}	2.02	-
$q_{O\max}$	0.7218	g/(kg.h)
$K_{i,O}$	6.952	g/(kg.h)
$q_{AC\max}$	0.0967	g/(kg.h)
K_A	0.5236	g/(kg.h)
$K_{i,A}$	5.85	g/(kg.h)

Appendix B. INITIAL STATE VARIABLES VALUES

Variables	X	S	A	W
Values	5	0.03	0.55	3.17
Units	g/kg	g/kg	g/kg	kg