Dynamic simulation of a fed-batch enzyme fermentation process

Esko K. Juuso
Control Engineering Laboratory, P.O.Box 4300, FIN-90014 Univesity of Oulu

Abstract

This article describes building of a dynamic simulation model for prediction of bioprocess variables. The simulator consists of three interacting dynamic model based on the method of linguistic equations. Each model has three versions, i.e. an own version for each phase of the fed-batch fermentation process. Steady state methods with dynamic structures were used in developing these linguistic equation models. With this simulator, it is possible to predict values of dissolved oxygen concentration, oxygen transfer rate and concentration of carbon dioxide in the exhaust gas through the whole process, using only the values of the control variables as inputs. The simulator can be used as a online forecasting tool in connection to the real process.

1 Introduction

Batch bioprocesses are difficult to model due to strong nonlinearity, dynamic behaviour, lack of complete understanding and unpredictable disturbances from their external environment [1]. As every cell in nature has a finite lifetime, a continuous growth of the organisms is needed to maintain the species. The generation time depends on both nutritional and genetic factors. To be able to live, reproduce and make products, a cell must obtain nutrients from its surroundings.

In batch reactors all components, except gaseous substrates such as oxygen, pH-controlling substances and antifoaming agents, are placed in the reactor in the beginning of the fermentation. During three process there is no input nor output flows. In fed-batch processes, nothing is removed from the reactor during the process but one substrate component is added in order to control the reaction rate by its concentration. The process is started as a batch process, and the substrate feed is started when the initial glucose is consumed. The fermentation continues at a certain growth rate until some practical limitation inhibits the cell growth. [2]

The data sets obtained from process are in practice distinct sets obtained through different process performances because usually one or more substantial physical parameters, such as dissolved oxygen, temperature or pH are maintained on distinct level [3]. The optimal values of parameters, such as pH, temperature and DO might not be the same for the growth phase and metabolite production phase in secondary metabolite production [4]. Large differences exist between different fermentation runs because of the variations in the feeding strategy, metabolic state of the cells and the amount of oxygen available. Even if the process conditions were kept same in each fermentation, the micro-organisms would behave differently every time. Detection of fluctuations in operating conditions is essential for making correct actions in time.

The concentration of carbon dioxide in the exhaust gas is an important variable in a fermentation process since the production of carbon dioxide is correlated to the amount of consumed sugar [5]. The variations in the agitation speed can cause changes in oxygen transfer rate, and an increase in it can cause an increase in production and yield [6]. In [7] it is stated that the tension of dissolved oxygen is an important variable in secondary metabolite production and remarkable impacts in production yields can be achieved by affecting this parameter by changes in aeration, agitation system and stirrer speed. The volumetric mass transfer coefficient, $k_L a$, is also an important process variable because it can be used to find the relationship between OTR and enzyme production [6] and it can be used in the control of dissolved oxygen tension [8].

The oxygen requirements of the bacteria differ at different fermentation stages [9]. By choosing a proper dissolved oxygen tension a product formation can be achieved without wasting the energy source. As the changes are slow, early forecasting of the process operation is needed. A smoothly operated process is likely to be more productive than one that is subjected to significant disturbances.

This paper describes a dynamic simulation model for prediction of the operation in a fed-batch fermentation process.

2 Modelling with linguistic equations

A nonlinear mapping has been developed to extract meanings of variables from measurement signals. The scaling function scale the real values of variables to the range of [-2, +2] which combines normal operation [-1, +1] with handling of warnings and alarms. The scaling function contains two monotonously increasing functions: one for the values between -2 and 0, and one for the values between 0 and 2.

2.1 Nonlinear scaling

The mapping functions can for example consist of two second-order polynomials,

$$\begin{aligned} f_j^- &= a_j^- X_j^2 + b_j^- X_j + c_j, \quad X_j \in [-2,0), \\ f_j^+ &= a_j^+ X_j^2 + b_j^+ X_j + c_j, \quad X_j \in [0,2], \end{aligned} \tag{1}$$

where a_j^-, b_j^-, a_j^+, and b_j^+ are coefficients of the polynomials, c_j is real value corresponding to the linguistic value 0.
The scaled values are obtained from the real values x_j by

$$X_j = \begin{cases} 2 & \text{with } x_j \geq \max(x_j) \\ \frac{-b_j^+ + \sqrt{{b_j^+}^2 - 4a_j^+ (c_j - x_j)}}{2a_j^+} & \text{with } c_j \leq x_j \leq \max(x_j) \\ \frac{-b_j^- + \sqrt{{b_j^-}^2 - 4a_j^- (c_j - x_j)}}{2a_j^-} & \text{with } \min(x_j) \leq x_j \leq c_j \\ -2 & \text{with } x_j \leq \min(x_j). \end{cases} \tag{2}$$

$\min(x_j)$ and $\max(x_j)$ are minimum and maximum values of the real data corresponding to the linguistic values -2 and 2. Examples of scaling functions are presented in Figure 1.

2.2 Interactions

The nonlinear scaling is an essential part of the linguistic equation (LE) approach [10]. The LE models consist of two parts: interactions are handled with linear equations, and nonlinearity is taken into account by membership definitions. The general LE model can be presented by

$$AX + B = 0, \tag{3}$$

where vector X defines linguistic levels for variables. The direction and strength of interaction is presented by matrix A. Bias term B can be used to shift the model from the origin. Vector X is obtained by nonlinear scaling from vector x which can include direct measurement, features generated from them, principal components, etc.

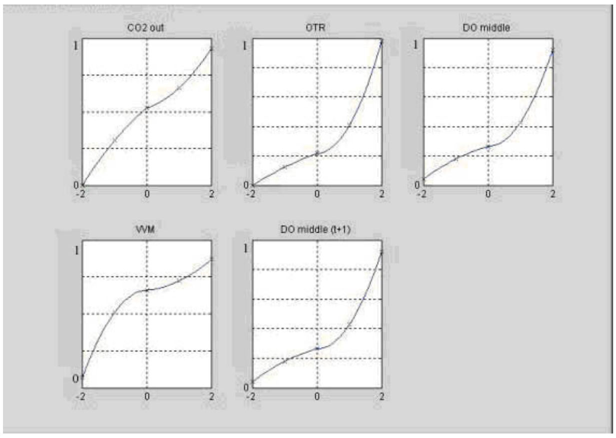

Figure 1: A set of membership definitions. The linguistic values of variables [-2 2] are on the x-axis and the real values of the variables (scaled from 0 to 1) are on the y-axis.

After the linguistic level of the model output, X_{out}, is calculated with linguistic equation model, it is converted to real value of the output, x_{out}, using the following equation:

$$x_{out} = \begin{cases} a_{out}^- {X_{out}}^2 + b_{out}^- X_{out} + c_{out} & \text{with } X_{out} < 0 \\ a_{out}^+ {X_{out}}^2 + b_{out}^+ X_{out} + c_{out} & \text{with } X_{out} \geq 0 \end{cases} \tag{4}$$

where a_{out}^-, b_{out}^-, a_{out}^+ and b_{out}^+ are coefficients of the polynomials, and c_{out} is the real value corresponding to the linguistic value 0.
Multimodel approach can realised as the linguistic equation models are affine in linguistic range. The main idea is to extend the operating range of each individual model by nonlinear scaling methods. Comparing different modelling techniques provides information on how well the nonlinear scaling has been done.

2.3 Dynamic models

Dynamic fuzzy models can be constructed on the basis of state-space models, input-output models or semi-mechanistic models [11]. In the state-space models, fuzzy antecedent propositions are combined with a deterministic mathematical presentation of the consequent. The most common structure for the input-output models is the NARX /Nonlinear AutoRegressive with eXogenous input) model

$$y(k+1) = F(y(k), \ldots, y(k-n+1), u(k), \ldots, u(k-m+1)), \tag{5}$$

where k denotes discrete time samples, n and m are integers related to the systems' order. This structure is directly used for multiple input, single output (MISO)

systems. Multiple input, multiple output (MIMO) sy
tems can be built as a set of coupled MISO models.
Delays can be taken into account by moving the valu
of input variables correspondingly. The basic form
the linguistic equation (LE) model is a static mappi
in the same way as fuzzy set systems and neural n
works, and therefore dynamic models will include se
eral inputs and outputs originating from a single va
able. External dynamic models provide the dynam
behaviour:

- Rather simple input-output LE models, where t
old value of the simulated variable and the cu
rent value of the control variable as inputs and t
new value of the simulated variable as an output,
can be used since nonlinearities are taken into account by membership definitions.

- Linear state-space models of different operating conditions are transformed into a linguistic equation model. Since the LE model can handle nonlinearities, at least some of the rules can be combined.

- In semi-mechanistic models, approximation rules are combined into linguistic equations. The approximation rules can be based on qualitative knowledge, and the mechanistic models take care of the basic level dynamic simulation.

Need for higher order dynamic models can be tested by applying classical identification to the data after nonlinear scaling. As parametric dynamic models, e.g. autoregressive moving average (ARMAX), autoregressive with exogenous inputs (ARX), Box-Jenkins and Output-Error (OE), are based on linear techniques, nonlinear scaling reduces the number of input and output signals needed for modelling. Normally, this analysis confirms the applicability of a compact input-output linear model, which needs only the old value of the simulated variable and the current value of the control variable as inputs and the new value of the simulated variable as an output [12].

3 Dynamic modelling

The dynamic modelling was based on the process data obtained from an industrial fed-batch fermenter. The models were tested using a number of different testing data, which were not included in the training data set. When necessary, the noise in the data was filtered by taking moving averages of the measured values. The variables for each model were chosen mainly based on correlation analysis. Variables that could be used for control were preferred when choosing the input variables of the model. These variables include mixing rate, aeration, substrate feed rate etc.

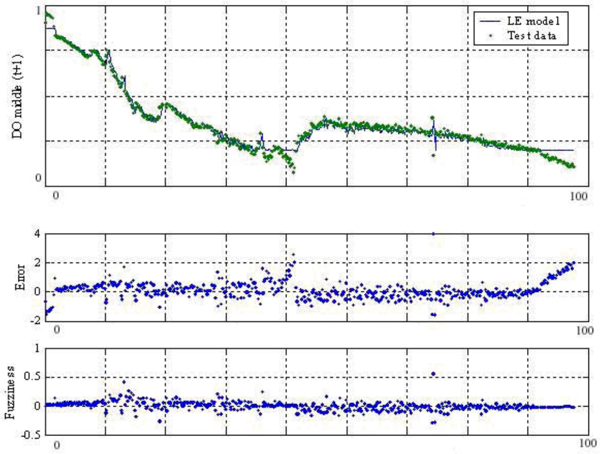

Figure 2: Results from the testing of steady-state models. Time from 0 to 100 is shown on the x-axis and the values of dissolved oxygen concentration, error and fuzziness on the y-axis.

The models have a NARX (Nonlinear AutoRegressive with eXogenous input) structure. A multimodel approach was applied as different growth phases need different models. As the prediction of the future values required three interacting models, which each produce prediction of a different variable, the overall system consists of nine models. Various modelling methodologies have been compared. The compact implementation of the linguistic equation models made such a complex structure possible to use. Smooth transitions between the phase models are based on fuzzy logic.

The controllable variables were preferred as inputs and these include mixing, aeration, feed rate, pressure, temperature and cooling power. The variables used in the models include the concentration of carbon dioxide in the exhaust gas, mixing power, feed rate, oxygen transfer rate, dissolved oxygen concentration, volumetric oxygen transfer coefficient, position of the pressure valve and VVM (volumes of air per volume of liquid per minute). The choice of the variables was quite similar to the normal choice in the literature.

An example of data-driven modelling results for the prediction of the dissolved oxygen is presented in Figure 2. The new measure, fuzziness, is used for detecting areas where the models should be considerably different. Fuzziness can also be considered as an additional unknown variable. In this case the fuzziness is very low.

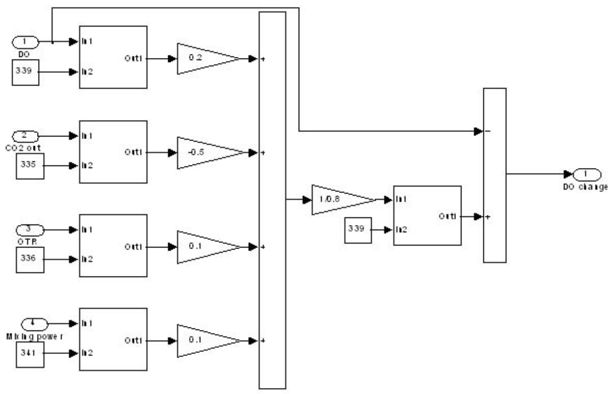

Figure 3: Dynamic model for dissolved oxygen concentration.

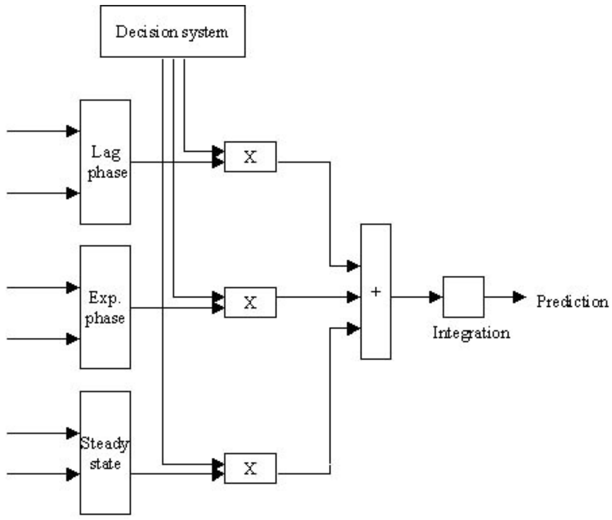

Figure 4: The dynamic model structure of one predicted variable.

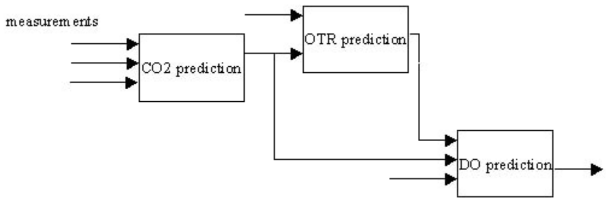

Figure 5: The overall structure of the model.

Three modelling techniques with several variants were compared including the methods of linguistic equations, neural networks and fuzzy modelling. Static modelling of the fermentation variables was not difficult for these intelligent modelling methods: linguistic equation method, linear neural network, feedforward neural networks and Takagi-Sugeno fuzzy models created by subtractive clustering appeared to be the best[13]. However, dynamic simulation turned out to be too demanding for most of these methodologies.

The membership definitions and coefficients A_{ij} from steady state linguistic models are transferred into the dynamic model. Membership definitions are used in converting real valued measurement into linguistic values. The four blocks on the left hand side of Figure 3 perform this operation. The delinguistification block on the right hand side converts the linguistic value back to a real value. The triangle shaped blocks contain the coefficients of the steady state models. The new prediction is calculated using previous values of the predicted value and the previous values of control variables.

Different growth phases can be distinguished from the fermentation process and during these phases different variables affect the output variables. Because of this, three submodels for each predicted variable were created corresponding to each phase in the fermentation process (Figure 4). The first phase at the beginning of the fermentation is called the lag phase. The second phase is the exponential growth phase. During this phase the growth is exponential. The last phase is called the steady state phase. The secondary metabolic products, such as enzymes, are produced mainly during the steady state phase.

The same structure is used for all the predicted variables. New predictions are obtained by integrating the calculated changes to the previous value with an ordinary differential equation solver based on an explicit Runge-Kutta (4,5) formula, the Dormand-Prince pair, with variable step.

The overall model consists of three models, which each produce prediction of a different variable. Inputs to the models include measurements from the process, such as mixing power, aeration rate, pressure, and substance concentrations. The inputs to the models of oxygen transfer rate prediction and dissolved oxygen concentration prediction include also predicted values from other models (Figure 5). Each model contains three submodels and a decision system as shown in Figure 4.

The overall dynamic model shown in Figure 6 contains an additional model for calculating the volumetric mass transfer coefficient, $k_L a$.

Altogether, the overall model contains 9 different submodels: three for each predicted variable. The three submodels (lag phase, exponential phase, and steady state) shown in Figure 4 form subsystems of the prediction models. The same fuzzy decision system weights the outputs of each of these submodels. In Figure 3, it can be seen how the linguistic equation approach calculates the output value of the dissolved

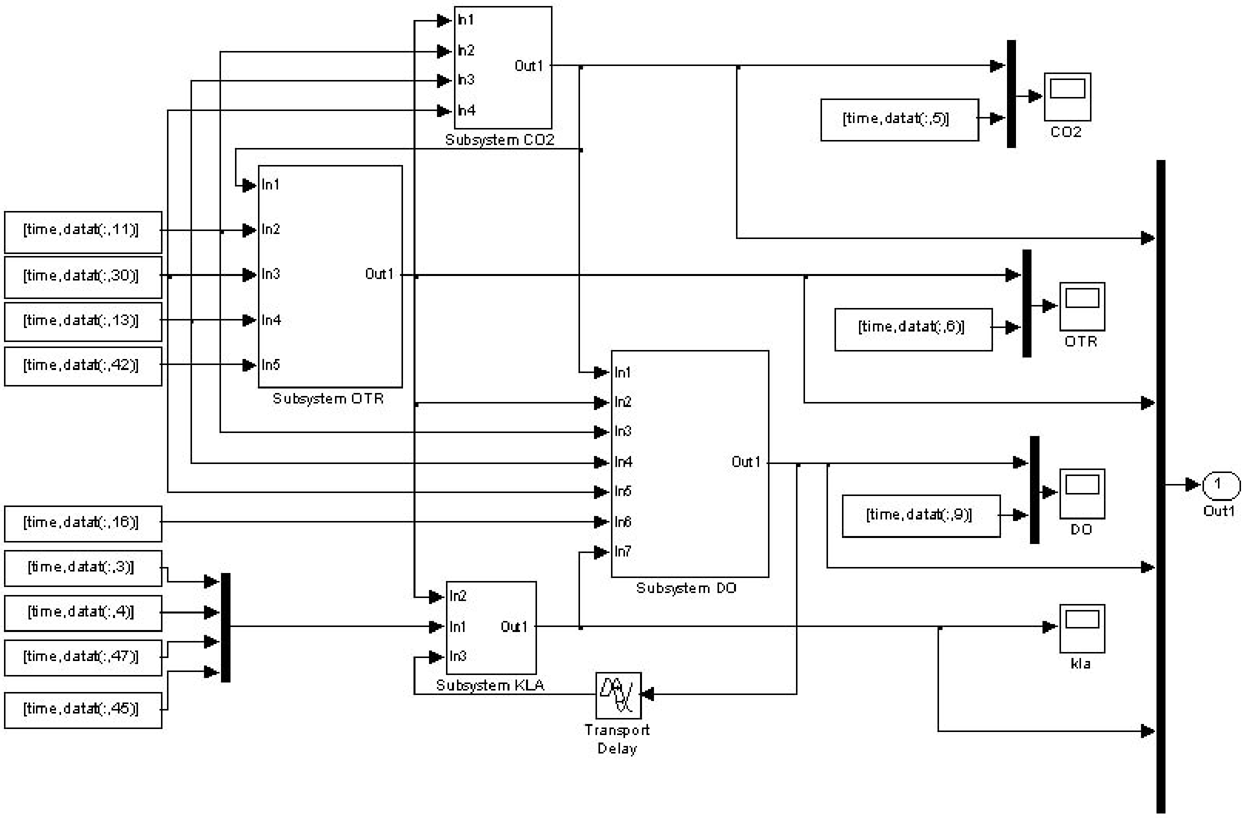

Figure 6: The overall structure of the model.

oxygen model. The other submodels have similar structures: lingustification, linear model and delinguistification.

The fuzzy decision system chooses the right submodel phase of the process by using the measurements of time, oxygen transfer rate and substrate feed rate. The inference system presented in Figure 7 includes membership functions for these three variables and for the output variable, and a set of eight fuzzy rules for deduction. The system gives a weighting factor from 0 to 1 for each submodel according to which level its results are used. The system was constructed using the FIS Editor of Fuzzy Logic Toolbox in the Matlab. For example in the beginning of the fermentation the first submodel, lag phase, is given a weight of one, and the other two submodels have the weight of zero. This means that only the output of first submodel is used in calculating the prediction. The transition from one phase to another happens smoothly, thus during the transition phase two outputs of the submodels can be used simultaneously (Figure 8).

Figure 3 shows the dynamic model for the dissolved oxygen concentration (DO). Each submodel has been developed separately on the basis of selected training data. The combined model (Figure 6) has been

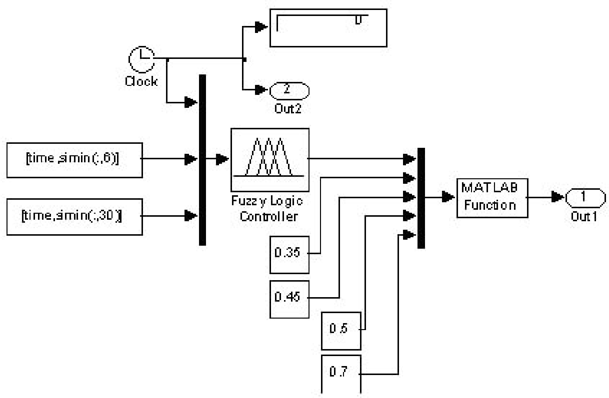

Figure 7: The fuzzy decision system.

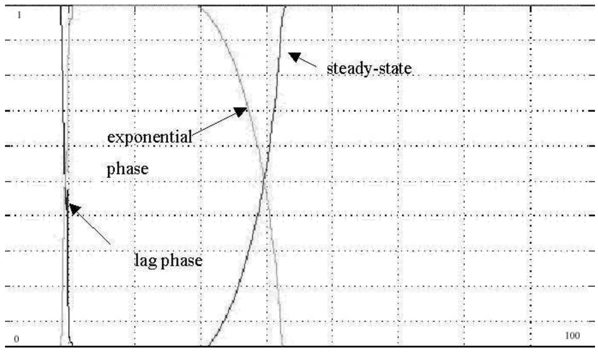

Figure 8: The weights from the decision system. The x-axis represents time and the y-axis weighting factors [0 1].

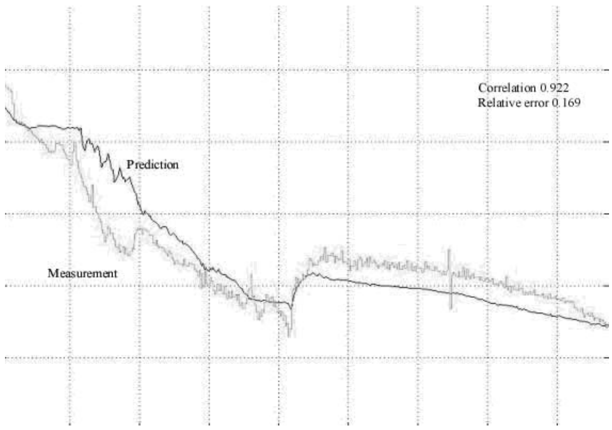

Figure 9: Prediction of dissolved oxygen concentration. Time is on the x-axis and the dissolved oxygen concentrations on the y-axis.

tested with data collected from various fermentation runs. In the simulation tests, the input values were taken from the previously collected data. During the on-line tests, the prediction system collects the data from the automation system and starts the simulation on chosen time intervals. The prediction results were written back to the data collection system.

The calculated prediction and the training data are printed in the same display (Figures 9 and 10). The results can be examined visually, or the correlation and the error of the dynamic model can be calculated using the FuzzEqu Toolbox [14].

4 Results

The models were tested with a set of test data. The fitness of a model can be estimated by examining the correlation, R, relative error, fuzziness and the model surfaces. The FuzzEqu program also draws the results of the predictions in the same chart with the test data

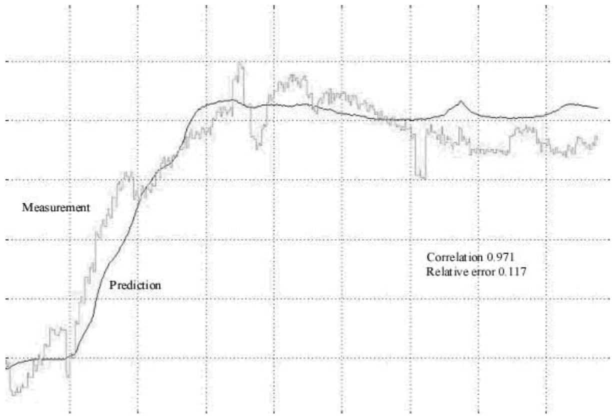

Figure 10: Prediction of oxygen transfer rate. Time is on the x-axis and the oxygen transfer rates on the y-axis.

where they can be compared visually. The fuzziness of the equations should be close to zero. It shows how well the equation represents the data [15].

First, steady-state models for all three variables were made by the linguistic equations approach. Correlations of the dissolved oxygen models for different testing data were between [0.88-0.98] and the relative errors between [0.03-0.18]. For models of oxygen transfer rate the correlations were between [0.72-0.99] and the relative errors between [0.02-0.33]. Similar results were obtained with all the static models used in the simulation model. The first part of the process was the most difficult to model, largely due to differences between fermentations. However, during the beginning of the process the concentration of the dissolved oxygen is usually quite high and its predicted value is not so critical information.

Dynamic modelling and simulation was performed by a Matlab Simulink program. Figure 8 presents the weights of the submodels obtained from the fuzzy decision system. The change from one phase to another is quite fast. The estimation of the dissolved oxygen concentration is presented in Figure 9. In this model, the estimations of the oxygen transfer rate and the concentration of carbon dioxide are used as inputs. In Figure 9, the estimation of the concentration of carbon dioxide is presented. The estimation of the oxygen transfer rate can be seen in the Figure 10. The estimate of the carbon dioxide concentration is used as an input of the model. The correlations and relative errors of these results are shown in the figures. With the exception of a few fermentations that largely differed from the others, the results were similar for other testdata. The estimation was easier for oxygen transfer rate and the carbon dioxide concentration than for dissolved oxygen concentration.

A multimodel approach was applied as different growth phases need different models. As the prediction of the future values required three interacting models, which each produce prediction of a different variable, the overall system consists of nine models. The compact implementation of the linguistic equation models made such a complex structure possible to use. Smooth transitions between the phase models are based on fuzzy logic.

The important factors in the success of the modelling were the choice of the input variables, the choice of the model type and structure, and the choice of training data. The training data should be sufficiently large so that it can represent different fermentations. The results of the modelling can improve with the number of data runs employed for training [16]. Large differences exist between different fermentation runs because the variations in the feeding strategy, metabolic state of the cells and the amount of oxygen available. Even if the process conditions were kept same in each fermentation, the micro-organisms would behave differently every time.

The choice of the input variables was difficult. Different variables affect the output variables in the different phases of the process. All the influences of the variables could not be examined because the data was obtained from an industrial fermenter and a part of the variables were controlled to remain constant. The data based modelling methods require changes in the data to be able to model it.

The dynamic simulator operates accurately throughout the fermentation even for more than 40 hours as a real simulation, i.e. the simulator uses in each time step only the previous simulated value and the values of the variables which control the process, according to the dynamic model. Differences between the calculated and measured are reasonable and provide a good basis for detecting fluctuations in operating conditions.

The simulator can be used as a on-line forecasting tool in connection to the real process. The simulator is started on chosen time intervals: the previous online measurements on a chosen horizon are used for constructing a starting point and the simulator predicts the operation on a chosen prediction horizon by using the planned control actions. In the on-line tests, the prediction horizon has been one hour and the time interval between predictions six minutes. The operation corresponds to the basic principle of model-based predictive control (Figure 11) as the future control actions should be used in simulation. Actually, generating a good

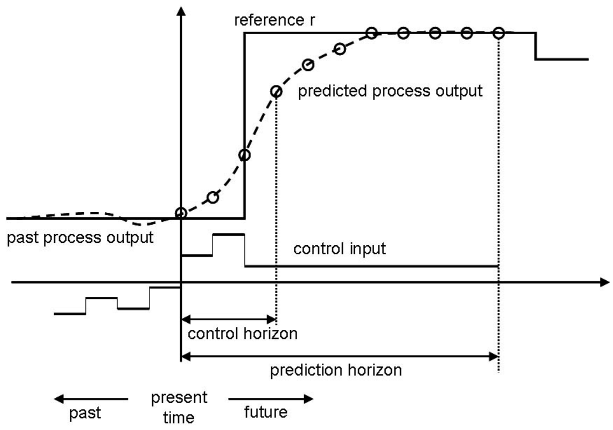

Figure 11: The basic principle of model-based predictive control.

starting point for simulation calculations was more demanding than the prediction part.

The simulator is aimed primarily on detection of fluctuations of the process control. However, model-based predictive control can be considered as a new option since the simulator is very compact.

5 Conclusions

The linguistic equation approach is a feasible platform for developing interacting dynamic models for bioprocesses with several process phases. A steady-state modelling an be used for developing dynamic models. The dynamic simulator can be used as an on-line forecasting tool in connection to the real process. The simulator is started on chosen time intervals

References

[1] L. Gregersen and S. B. Jorgensen. Supervision of fed-batch fermentation. *Chemical Engineering Journal*, 75:69–76, 1999.

[2] S.-O. Enfors and L. Häggström. *Bioprocess Technology Fundamentals and Applications*. Royal Institute of Technology, Stockholm, 2000. 356 pp.

[3] O. Georgieva, M. Wagenknecht, and R. Hampel. Takagi-sugeno fuzzy model development of batch biotechnological process. *International Journal of Approximate Reasoning*, 26:233–250, 2001.

[4] P. K. Yegneswaran, M. R. Gray, and B. G. Thompson. Effect of dissolved oxygen control on growth and antibiotic production in streptomyces clavuligerus fermentation. *Biotechnology Progress*, 10:246–250, 1991.

[5] G. Martínez, A. López, A. Esnoz, P. Vírseda, and J. Ibarrola. A new fuzzy control system for white wine fermentation. *Food Control*, 10:175–180, 1999.

[6] M. Elibol and D. Ozer. Influence of oxygen transfer on lipase production by rhizopus arrhizus. *Process Biochemistry*, 36:325–329, 2000.

[7] C. Pfefferle, U. Theobald, H. Gürtler, and H.-P. Fiedler. Improved secondary metabolite production in the genus streptosporangium by optimisation of the fermentation condition. *Journal of Biotechnology*, 80:135–142, 2000.

[8] L. Simon and M. Nazmul Karim. Identification and control of dissolved oxygen in hybridoma cell culture in a shear sensitive environment. *Biotechnology Progress*, 17:634–642, 2001.

[9] H. M. Yao, Y. C. Tian, M. O. Tadé, and H. M. Ang. Variations and modelling of oxygen demand in amino acid production. *Chemical Engineering and Processing*, 40:401–409, 2001.

[10] E. K. Juuso. Integration of intelligent systems in development of smart adaptive systems. *International Journal of Approximate Reasoning*, 35:307–337, 2004.

[11] R. Babuška, M. Setnes, U. Kaymak, and H. B. Verbruggen. Fuzzy Modelling: a Universal and Transparent Tool. In L. Yliniemi and E. Juuso, editors, *Proceedings of TOOLMET'97 - Tool Environments and Development Methods for Intelligent Systems, Oulu, April 17-18, 1997*, pages 1–27, Oulu, 1997. Oulun yliopistopaino.

[12] E. K. Juuso. Modelling and simulation with intelligent methods. In *White paper of the Virtual Institute for Simulation (Sim-Serv): www.sim-serv.com*. Sim-Serv, 2004. 17 pp.

[13] U. Saarela, K. Leiviskä, E. Juuso, and A. Kosola. Modelling of a fed-batch enzyme fermentation process. In *IFAC International Conference on Intelligent Control Systems and Signal Processing. Faro, Portugal, April 8-11, 2003*. IFAC, 2003.

[14] E. K. Juuso. Linguistic equations for data analysis: Fuzzequ toolbox. In L. Yliniemi and E. Juuso, editors, *Proceedings of TOOLMET 2000 Symposium - Tool Environments and Development Methods for Intelligent Systems, Oulu, April 13-14, 2000*, pages 212–226, Oulu, 2000. Oulun yliopistopaino.

[15] E. K. Juuso. Fuzzy control in process industry: The linguistic equation approach. In H. B. Verbruggen, H.-J. Zimmermann, and R. Babuska, editors, *Fuzzy Algorithms for Control, International Series in Intelligent Technologies*, pages 243–300. Kluwer, Boston, 1999.

[16] S. Feyo de Azevedo, P. Dahm, and R. R. Oliveira. Hybrid modelling of biochemical processes: A comparison with the conventional approach. *Computers and Chemical Engineering*, 21(Suppl.):S751–756, 1997.