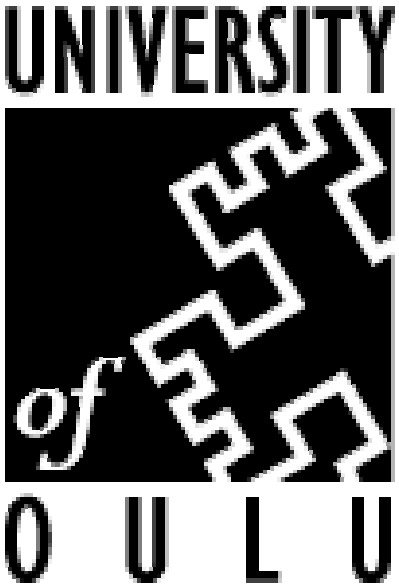


CONTROL ENGINEERING LABORATORY

Modelling of a Fed-Batch Fermentation Process

Ulla Saarela, Kauko Leiviskä and Esko Juuso

Report A No. 21, June 2003

University of Oulu
Control Engineering Laboratory
Report A No. 21, June 2003

MODELLING OF A FED-BATCH FERMENTATION PROCESS

Ulla Saarela, Kauko Leiviskä and Esko Juuso

Control Engineering Laboratory
Department of Process and Environmental Engineering
University of Oulu
P.O.Box 4300, FIN-90014 University of Oulu, Finland

Abstract: This report describes the building of a simulator for prediction of the dissolved oxygen concentration, the oxygen transfer rate and the concentration of carbon dioxide in a fermentation process. The steady state models were made using the linguistic equations method. The dynamic models were made using Simulink® toolbox in the Matlab®.

At the beginning, some basics about fermentation and microbiological reactions are stated. In the third chapter the modelling methods are presented. The modelling experiments are presented in chapter four and after that the results are stated. Chapter six includes discussion about the results and the conclusions. The simulation results were good.

Keywords: fermentation, modelling, linguistic equations

ISBN 951-42-7083-5
ISSN 1238-9390
ISBN 951-42-7514-4 (PDF)

University of Oulu
Control Engineering Laboratory
P.O.Box 4300
FIN-90014 University of Oulu

CONTENTS

1 INTRODUCTION

A process, which employs microorganisms, animal cells and/or plant cells for the production of materials, is a bioprocess. Most biotechnical products are produced by fermentation. In fermentation, the products are formed by catalysts that catalyse their own synthesis. Enzymes are biological catalysts and are produced as secondary metabolites of enzyme fermentation.

There are many aspects that complicate the modelling of the bioprocesses. A fermentation process has both nonlinear and dynamic properties. The metabolic processes of the microorganisms are very complicated and cannot be modelled precisely. Because of these reasons, traditional modelling methods fail to model bioprocesses accurately. The modelling is further complicated because the fermentation runs are usually quite short and large differences exist between different runs.

The purpose of this work was to create a model for prediction of dissolved oxygen concentration, oxygen transfer rate and carbon dioxide concentration. Earlier different modelling methods were compared and the method of linguistic equations was concluded to be the best method for this purpose /21/. Dynamic models were constructed based on these steady state models.

This work is a part of INTBIO – Intelligent Methods in the Analysis and Control of Bioprocesses research project, which is financially supported by Tekes, Genencor International and Hartwall. The goal of the project is to develop new measurements and soft sensors to aid the optimisation and control of the fed-batch fermentation process.

2 FERMENTATION

Fermentations can be operated in batch, fed-batch or continuous reactors. In batch reactor 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 process there is no input nor output flows. In fed-batch process, 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. There are both input and output flows in a continuous process, but the reaction volume is kept constant. /1/

2.1 Cells

Every cell in nature has a finite lifetime and in order to maintain the species the continuous growth of the organisms is needed. A bacterial cell is able to duplicate itself. The duplication process is quite complicated and includes as many as 2000 different chemical reactions. The generation time, that is the time needed for the cells to double the mass or the number of the cells, depends on the number of factors, both nutritional and genetic. For *Escherichia coli* in ideal conditions the doubling time can be as short as 20 min, but usually it takes a longer time. /2/

To be able to live, reproduce and make products, a cell must obtain nutrients from its surroundings. Heterotrophic microorganisms, which include most of the bacteria, require an organic compound as the carbon source. A cell can use either light or chemicals as its energy source. A chemotroph obtains energy by breaking high-energy bonds of chemicals. Most organisms that are used in industrial processes are chemoheterotrophs, i.e., organisms that use an organic carbon source and a chemical source of energy. /3/

A view of a cell as an open system is presented in Figure 1. A cell produces more cells, chemical products and heat from chemical substrates. A cell requires many different kinds of substrates to function. In most cases carbon is supplied as sugar or some other carbohydrate. Glucose is often used. In aerobic processes oxygen is a vital component. Oxygen can be fed into the process by continuous aeration. The most common source of nitrogen is ammonia or an ammonium salt. In some cases the growth rate of the organisms increases if amino acids are supplied. Required amounts of hydrogen can be derived from water and organic substrates. Other compounds that are needed for growth include P, S, K, Mg and trace elements, which are added in the growth media as inorganic salts. /1/

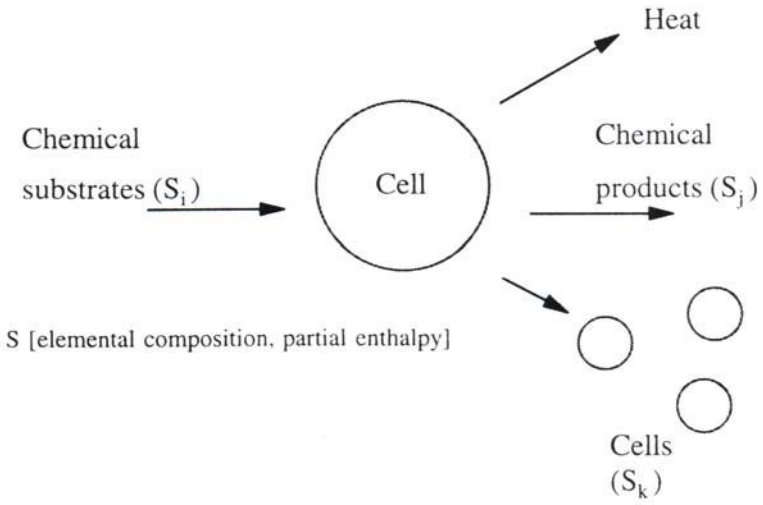


Figure 1. A view of a cell as an open system /3/.

When microorganisms are grown in a batch reactor certain phases of growth can be detected. A typical growth characteristic is shown in Figure 2. The appearance and the length of each phase depend on the type of organisms and the environmental conditions. /3/

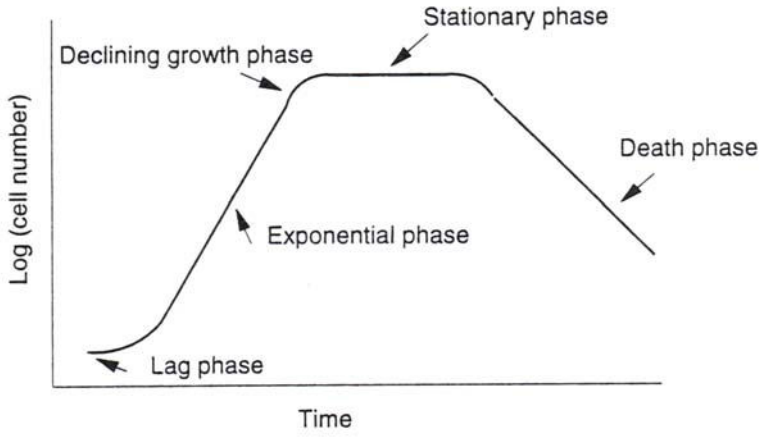


Figure 2. Growth phases in a batch process /3/.

The first phase in the growth, where the growth rate stays almost constant, is the lag phase. The lag phase is caused for many reasons. For example, when the cells are placed in fresh medium, they might have to adapt to it or adjust the medium before they can begin to use it for growth. Another reason for the lag phase might be that the inoculum is composed partly of dead or inactive cells /1/. If a medium consists of several carbon sources, several lag phases might appear. This phenomenon is called diauxic growth. Microorganisms usually use just one substrate at a time and a new lag phase really results when the cells adapt to use the new substrate. /3/

When a substrate begins to limit the growth rate the phase of the declining growth begins. The growth rate slows down until it reaches zero and the stationary phase begins. In the stationary phase the number of the cells remains practically constant, but the phase is important because many products are only produced during it. The last phase is called the death phase. During the death phase the cells begin to lyse and the growth rate decreases. /3/

The microorganisms can be divided into many groups depending of their need for oxygen. Although there are several groups, two main classes can be distinguished – aerobes and anaerobes. Organisms that cannot use oxygen are called anaerobes. They lack the respiratory system. Aerobes are capable of using oxygen and in many aerobic processes extensive aeration is required. /2/ The cells can usually use only water-dissolved substrates. Because of the limited solubility of oxygen into water, oxygen transfer can become a problem in the aerobic processes. The gas transfer from oxygen bubble into the cell includes many resistances, characterised by mass transfer constants. The most significant resistance in a well-stirred reactor is the diffusion through the stagnant liquid layer surrounding the air bubble.

Aeration is an important design parameter in the bioreactors and by its efficient control the overall productivity of the process can be increased. Product's requirements of oxygen depend on the energetics of the pathway leading to the product. Because the oxygen uptake is linked to the cellular metabolism, the oxygen dynamics reflect the changes in the environmental conditions. The rate of change of dissolved oxygen concentration is about 10 times faster than the cell mass or substrate concentrations. /4/

2.2 Enzyme production

Since 1980's a large increase has occurred in the range of commercial fermented products, particularly secondary metabolites and recombinant proteins. In the past, only the fermentation of extracellular enzymes, such as amylases and proteases, was industrially possible. The release of intracellular enzymes has become possible by large-scale mechanical techniques. Also chemical or physical methods can be used in the cell disintegration. /1/ Recombinant organisms will likely be used for producing a large proportion of enzymes in the future, because this approach enables the production of many different enzymes in substantial quantities and minimizes the production costs by using a small number of host/vector systems /5/. In many cases only low levels of protein can be produced by natural hosts. Systems, which have the gene of interest cloned and inserted in the expression vector, have been developed to achieve the abundant expression of the functional protein /6/.

The active form of an enzyme is a folded globular structure. If enzymes are subjected to stress, either *in vitro* or *in vivo*, they might unfold partially or completely. The stress can be provided by denaturants, high (or low) temperature or ionic composition of medium. When protein is overproduced in a recombinant microorganism, the local concentration of protein is raised and aggregation may occur. Denatured proteins may form bodies that cannot be recovered. /7/

2.3 Fed-batch fermentation

Fed-batch reactors are widely used in industrial applications because they combine the advantages from both batch and continuous processes. Figure 3 presents biomass concentration as the function of time in a typical fed-batch process. Process is at first started as a batch process, but it is exhibited from reaching the steady state by starting substrate feed once the initial glucose is consumed. The fermentation is continued at a certain growth rate until some practical limitation inhibits the cell growth. /1/

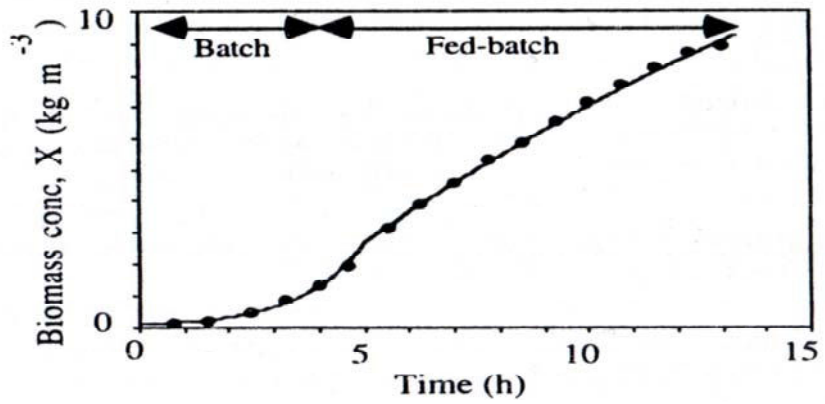


Figure 3. Biomass vs. time in a fed-batch process /1/.

The inlet substrate feed should be as concentrated as possible to minimize dilution and to avoid process limitation caused by the reactor size. In a fed-batch process the dilution rate means the components rate of dilution because of the volume increase caused by the inlet feed. The main advantages of the fed-batch operation are the possibilities to control both reaction rate and metabolic reactions by substrate feeding rate. The limitations caused by oxygen transfer and cooling can be avoided by controlling the reaction rate. /1/

In industrial fermentation systems, consistent operation is achieved by manual monitoring and control by process operators. The operators detect potential problems and make necessary modifications to the process based on their experience and knowledge of the process together with the information provided by supervisory control systems /8/. Because the models for model-based control are rare, fermentation processes are usually run with a predetermined feed profile /9,10/

A typical operation procedure is presented in /9/. The fermentation is started with a small amount of biomass and substrate in the fermenter. The substrate feed is started when most of the initially added substrate has been consumed. This procedure enables the maintaining of a low substrate concentration during fermentation, which is necessary for achieving a high product formation rate. The growth rate can be controlled by the substrate concentration to avoid catabolite repression and sugar-overflow metabolism /1/. The sugar-overflow metabolism, or glucose effect, occurs when glucose concentration exceeds a critical value and leads to excretion of partially oxidized products, such as acetic acid and ethanol. Most microorganisms exhibit some kind of overflow metabolism

and that is often detrimental to the process. Catabolite repression is a repression of the respiration on the enzyme synthesis level. It occurs during the long-term exposure of the cell to the high glucose concentration. /1/

Different types of substrate limitations can be used in the fed-batch processes. The repression of the growth rate can be achieved for example by sugar, nitrogen or phosphate sources. If no reaction rate control is used, and the cells are growing exponentially, the reaction will eventually be limited by oxygen or by heat. The metabolism control with the fed-batch process is useful also for the production of the secondary metabolites such as antibiotics, because the synthesis of them is repressed during the unrestricted growth. /1/

While in continuous fermentation the key variables are held constant, in fed-batch technique almost every key variable is changing as the process progresses. In order to give the best possible growing conditions the pH and temperature levels are usually kept constant /9/. The fermentation systems are very sensitive to abnormal changes in operating conditions. The performance of fermentation depends greatly on the ability to keep the system operating smoothly /11/. A smoothly operated process is likely to be more productive than one that is subjected to significant disturbances /8/.

2.3 Measurements

Instrumentation of the bioprocesses differs from that of a standard chemical reaction. Advantages of the bioprocesses are that they are quite stable and many variables change slowly over time. One of the challenges is that all the instruments inside the reactor must be absolutely sterile. The biggest problem in the instrumentation of a bioprocess is that there are no suitable sensors for on-line measurements of many important process parameters. For example, reliable measurement of the biomass or the glucose concentrations is not yet possible. /12,1/

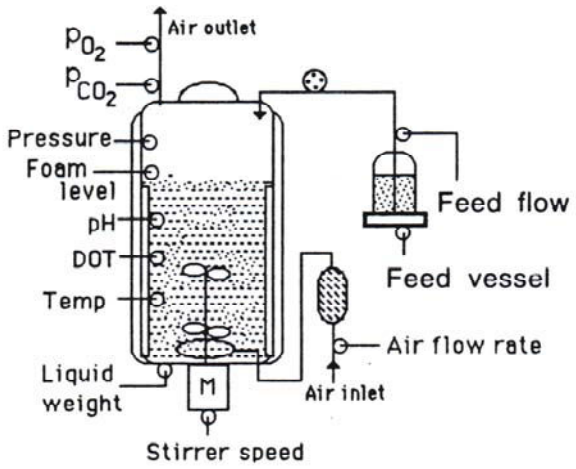


Figure 4. On-line measurements in bioreactor /1/.

Data acquisition of key fermentation variables is difficult due to the lack of reliable sensors for on-line measurements of biomass, substrate, and product concentrations. In recent years attention has been focused on the development of so-called “software sensors” /13/. A software sensor provides on-line estimates of unmeasurable variables, model parameters or helps to overcome measurement delays by using on-line measurements of some process variables and an estimation algorithm /14/.

3 MODELLING METHODS

Batch bioprocesses are difficult to model due to many aspects. The fermentation runs are short and large batch-to-batch differences exist in process conditions /9/. Modelling is further complicated because of bioprocesses' strong nonlinearity, dynamic behaviour, lack of complete understanding and unpredictable disturbances from their external environment. The data sets obtained from process are in practice specific sets obtained through different process performances because usually one or more substantial physical parameters, such as dissolved oxygen (DO), temperature or pH are maintained on the distinct level /15/. 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 /16/.

The models can be used for on-line fault diagnosis or for prediction of the product concentration /9/. The ability to control bioprocesses is of great interest, because it allows reduction of production costs and the increase of yield while maintaining the quality of the metabolic products /14/.

Most simple mathematical models are unable to describe the behaviour of the bioprocess well /10/. The predictive ability of conventional fermentation process models is quite limited /17/.

3.1 The method of linguistic equations

The linguistic equation models are made up of two parts. The linear equations handle the interactions and the membership definitions take the nonlinearities into account. An example of membership definitions is presented in Figure 5.

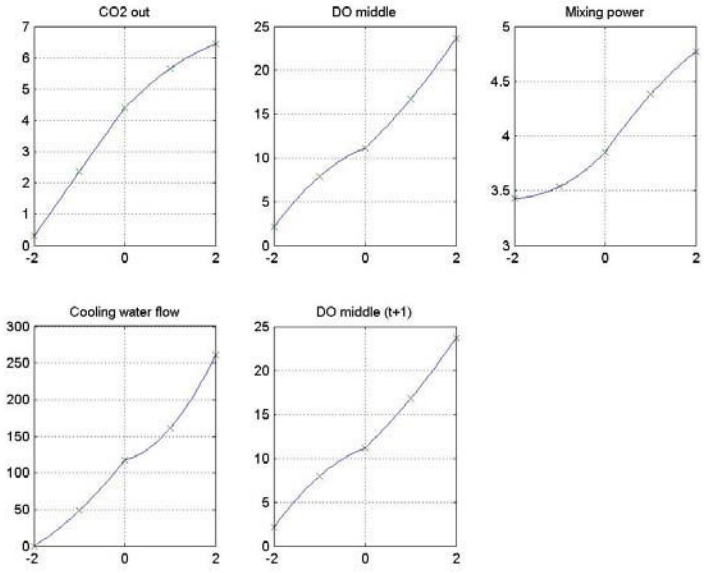


Figure 5. Membership definitions.

In the beginning of the modelling, the membership definitions and the feasible ranges must first be defined. They can be generated directly from the data or be defined manually. Expert knowledge can be used when defining the feasible ranges of the variables. The feasible range of a variable is defined as a membership function. The range of values a variable has is called the support area, and the main area of operation is called the core area. The corner parameters can be extracted from data or they can be defined by using expert knowledge. These parameters are made to equal linguistic values –2, -1, 1 and 2. The centre point is defined and it is given a linguistic value of 0. The centre point can be defined by some defuzzifying method or by using expert knowledge. /18/

The LE models can be used in any direction because the system is linearised by the nonlinear membership definition (NLMD) /19/. The NLMD consists of two monotonously increasing second-order polynomials, which are connected at the zero of linguistic variable. The NLMD transforms the real value of the input variable into a linguistic value in the range of [-2 +2]. The conversion (linguistification) is made by the following equation:

$$lv\left|x_{ij}(k)\right| = \begin{cases} 2\_if\_x_{ij}(k) \geq x_{ij}^{hl}(k) \\ \dfrac{-b_{ij} + \sqrt{b_{ij}^2 - 4 \times a_{ij} \times (c_{ij} - x_{ij}(k))}}{2 \times a_{ij}} \\ -2\_if\_x_{ij}(k) \leq x_{ij}^{ll} \end{cases} \quad (1)$$

where a_{ij}, b_{ij} are constants obtained from polynomials,
c_{ij} is the real value of the variable, which corresponds to the LV 0, and
x^{ll}, x^{hl} are the real values of the variable that correspond to the linguistic values of –2 and 2.

The linguistic value of the model output can be transformed (delinguistificated) to the real value by equation:

$$\text{rv} = y_i(k) = a_i \times lv\left|y_i(k)\right|^2 + b_i \times lv\left|y_i(k)\right| + c_i \quad (2)$$

The linguistic relations can be displayed as an equation /19/:

$$\sum_{j=1}^{m} A_{ij} X_j = 0 \quad (3)$$

where X_j is the linguistic level for the variable j, j=1…m, and
A_{ij} is the direction of interaction, $A_{ij} \in \{-1, 0, 1\}$.

When converting linguistic relations into equation form, linguistic values *very low, low, normal, high,* and *very high* are replaced by numbers –2, -1, 0, 1 and 2 indicating the linguistic level X.

3.2 Dynamic simulation

Dynamic fuzzy modelling can be performed based on state-space modelling, input-output modelling or semi-mechanistic modelling. Input-output models are often used when models are built from data. The most common structure for input-output models is the NARX (Nonlinear AutoRegressive with eXogenous input) model, which establishes a relation between the collection of the past input-output data and the predicted output /20/:

$$y(k+1) = F(y(k),\ldots, y(k-n+1), u(k), \ldots, u(k-m+1)) \quad (4)$$

where, k is discrete time sample, and
n, m are integers.

The basic form of the linguistic equation is a static mapping in the same way as the fuzzy systems and the neural networks, and therefore dynamic models include several inputs and outputs originating from a single variable. External models provide the dynamic properties. Since nonlinearities are taken into account by membership definitions, rather simple input-output models can be used. In these models the old value of the simulated variable and the current value of the control variable are used as inputs and the new value of the simulated variable as an output. In dynamic modelling of the linguistic equations, either single model or multimodel approach can be used, depending on the process. /19/ A multimodel approach is presented in Appendix 1 and a dynamic model for dissolved oxygen concentration (DO) in Appendix 2. The dynamic model in Appendix 2 can be presented by equation 5.

$$p_k = \int \left(\left[\sum rv \left(\sum A \cdot lv \left| x_{ij}(k) \right| \cdot \left(-\frac{1}{a_{ij}} \right) \right) - p_{k-1} \right] \cdot w \right) dt \quad (5)$$

where, p_k is the prediction
rv is the real (delinguistificated) value ,and
w is the weighting factor of a submodel.

Ode 45 (ordinary differential equation) solver was used in the integration. It is based on Runge-Kutta formula. Ode 45 is a one step solver – it needs only the solution at $y(t_{n-1})$ to calculate $y(t_n)$. /21/

A single model approach can be used in dynamic simulation if one set of membership definitions is able to describe the whole process. In small models, all the interactions are in a single equation. For larger models, a set of equations is needed, where each equation describes an interaction between two to four variables.

When one set of membership definitions cannot describe the system sufficiently, because of very strong nonlinearities, a multimodel approach can be used. This approach is able to combine specialized fuzzy LE submodels, which can have different equations and delays. A separate working point model defines the working area. If n working areas and m subareas has been defined, n×m submodels can be included in the model. The outputs of the submodels are aggregated by taking a weighted average of them. The working point model defines the degree of membership of each model, which equals the weight of the submodel. /19/

4 MODELLING EXPERIMENTS

The purpose of the experiments was to model the key parameters of fed-batch enzyme fermentation. A dynamic model for the prediction of the dissolved oxygen concentration, the concentration of carbon dioxide in the exhaust gas and the oxygen transfer rate was constructed. Different modelling methods were compared earlier and it was concluded that dynamic models were successful only when they were based on the linguistic equations models /22/. Other methods tested were the fuzzy modelling and artificial neural networks. Five different neural network types were tested. These types were perceptron, linear, feedforward, radial basis function and self-organizing networks. Also Takagi-Sugeno type fuzzy models created by using subtractive clustering were tested. Dynamic models can be used for control design and control of a process /23/. The data for modelling was obtained from an industrial fed-batch fermenter at the Genencor International plant in Hanko.

A part of the measurements were ignored from the modelling data because they were not suitable for it. Some variables remained constant during the whole process and some did not affect the course of the process. The number of variables for modelling was reduced to 51. After modifications required by the modelling program, the final size of the training data was around [438x59]. The number of rows in each data set varied according to the length of the fermentation. The training data set included data from seven different fermentations. The models were tested using a number of different test data, not included in the training data set.

Pre-processing of data was performed by the FuzzEqu Toolbox. By taking moving averages of the measured values the noise in the data was filtered when necessary. The variables for each model were mainly chosen based on correlation analysis performed using Microsoft Excel® 2000. The correlation value measures the linear dependence between two variables. The dependence is significant when the correlation is near 1 /23/. Variables that could be used for control were preferred when choosing the input variables of the model. These variables include mixing, aeration, substrate feed rate etc.

Different phases can be distinguished from the process and during these phases different variables affect the output variable. Because of this, it is reasonable to create different submodels for each phase in the fermentation process. The first phase, lag phase, starts at the beginning of the fermentation and lasts until the substrate feeding has begun. After the lag phase the model switches to the exponential growth phase. The phase of the exponential growth lasts until the substrate feeding is made constant. The last phase is called steady state and during it, the substrate feed is constant. The product is mainly produced during the steady state phase. In the models developed during this work, the typical number of data points is around 45 for the lag phase, 150 for the exponential phase and 200 for the steady state. The exact number of data points varies between different fermentations. A fuzzy decision system chooses the submodel, which suits best for each situation. The decision system chooses the submodel based on measurements from the process.

The models were tested with data. The fitness of a model can be estimated by examining the correlation, R, relative error, fuzziness distribution, fuzziness and the model surfaces. The FuzzEqu program also draws the acquired model in the same chart with data where they can be visually compared. The value of correlation varies between 0 and 1, where 1 means that the model fits the data perfectly. Model was assumed to be good if the correlation was near 1. Fuzziness shows how well the equations represent the data. If there are large deviations from zero, it indicates that other variables affect the process than is included in the model. The fuzziness of the equations should be close to zero. Model surfaces are presented in Figure 6. They are important for examining the directions of interactions. The model surfaces should be quite smoothly changing.

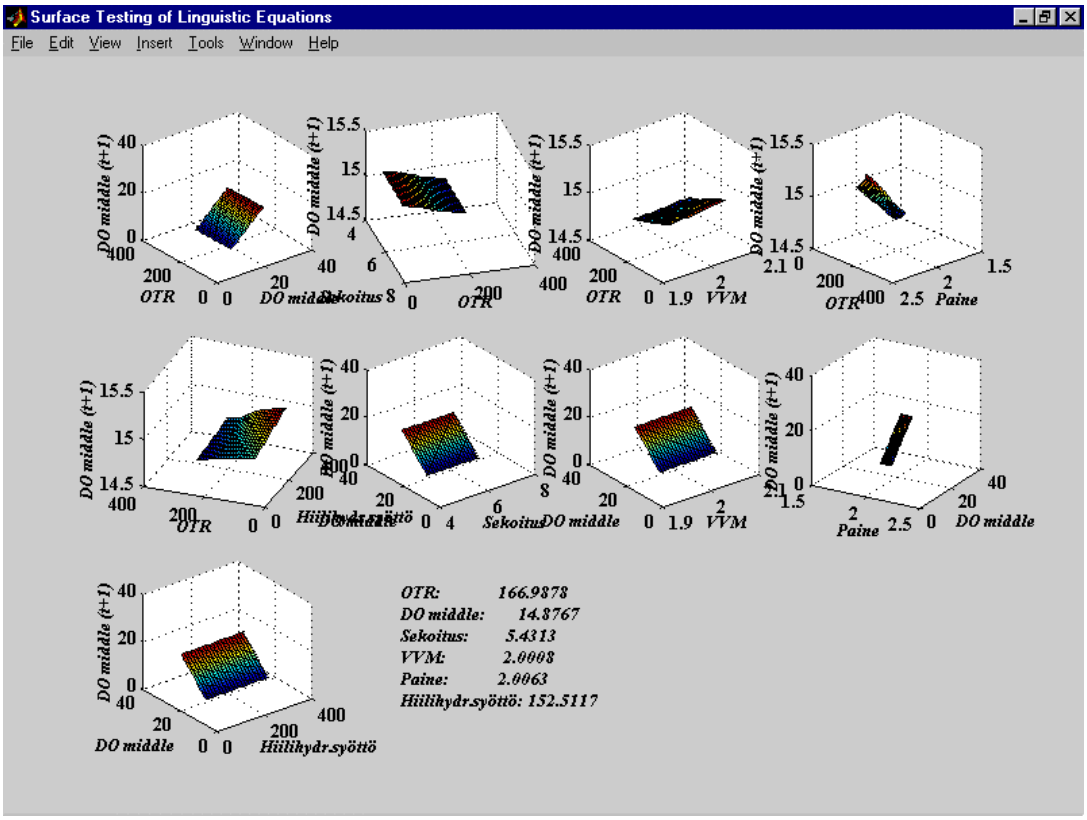


Figure 6. Model surfaces.

Dynamic modelling was performed starting from simulating steady-state models with the Matlab-Simulink® program. Steady-state models had a NARX (Nonlinear AutoRegressive with eXogenous input) structure; such that the output of the model was one-step ahead the inputs. The quality of modelling can be tested by simulation /24/.

The model for the prediction of the dissolved oxygen concentration (DO) is presented in APPENDIX 1. Predicted values of carbon dioxide and the oxygen transfer rate (OTR) are used in the dissolved oxygen model. Also the value of K_La (the volumetric oxygen transfer coefficient) is calculated based on the prediction of the oxygen transfer rate and the predicted value of dissolved oxygen.

In the dynamic model (APPENDIX 2), input data is converted to linguistic values in the subsystem of the model. The linguistic values of the inputs and the calculated value of output are weighted according to the parameters of the model and summed. The sum is multiplied by a parameter and delinguistified. The calculated variable is reduced from the result as seen in Figure 7. A new value for the output variable is obtained by integration. The integration term handles the dynamic effects of the model. The calculated new value and the training data are shown in the same display. The results can be examined visually from the display, or the correlation and the error of the dynamic model can be calculated using the FuzzEqu Toolbox.

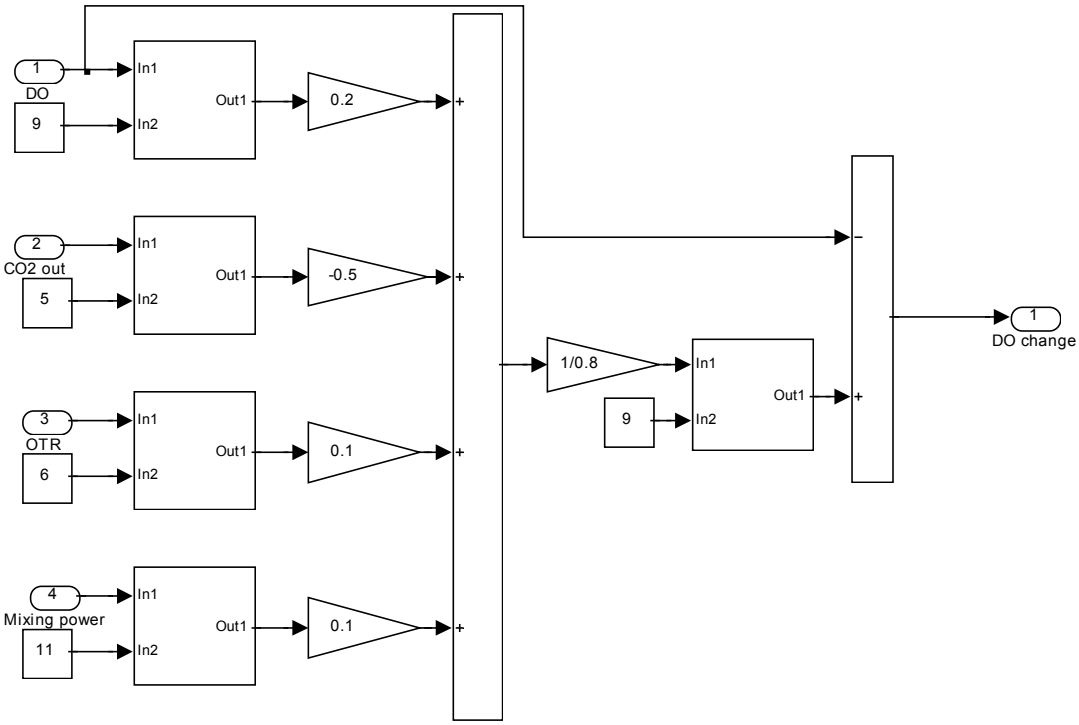


Figure 7. The linguistic equations approach in the dynamic model.

The fuzzy decision system that chooses the submodel to be used is presented in Figure 8. The model chooses the submodel based on the measurements made of time, the oxygen transfer rate and the glucose feed rate. The system gives a weighting factor (w) for each submodel, which is used to decide in which level its result is used. 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.

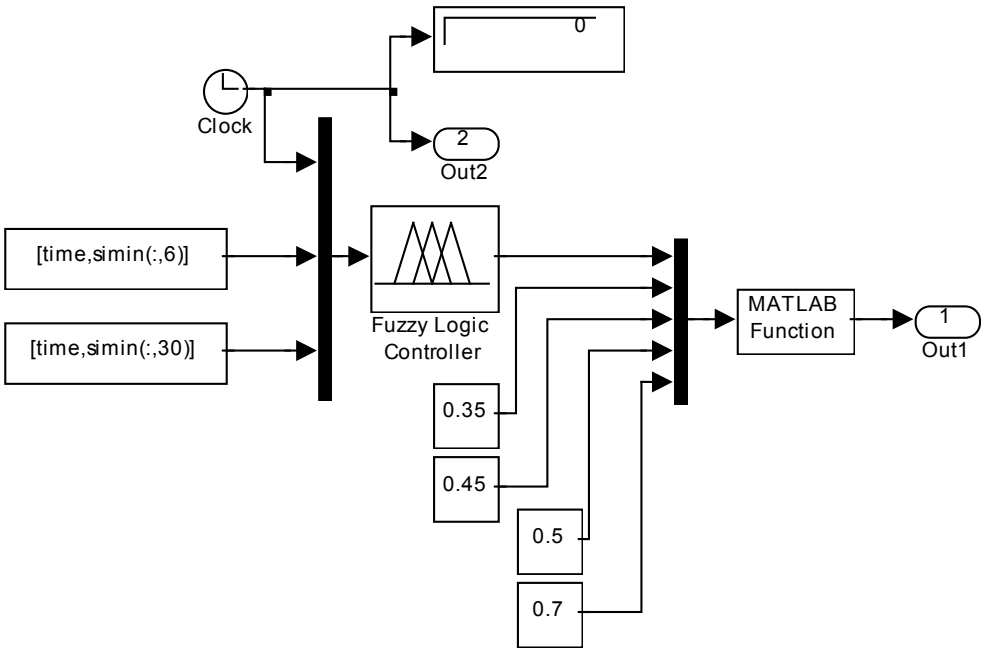


Figure 8. Fuzzy decision system for the selection of the submodel.

5 RESULTS

Models for the dissolved oxygen concentration, the oxygen transfer rate and the concentration of carbon dioxide in the exhaust gas were constructed. All the variables required specialized submodels for each growth phase. The variables used as inputs to the models include the mixing power, the VVM (volumes of air per volume of liquid per minute), the glucose feed rate, backpressure, and the k_La (the volumetric oxygen transfer coefficient).

First, steady state models for all three variables were made using the linguistic equations approach. An example of the testing of the models is presented in Figure 9. Correlation of the model is 0.98 and the relative error 0.07. With another set of testdata the correlation was 0.98 and the relative error 0.06. Similar results were obtained with all the steady state models used in the simulation model.

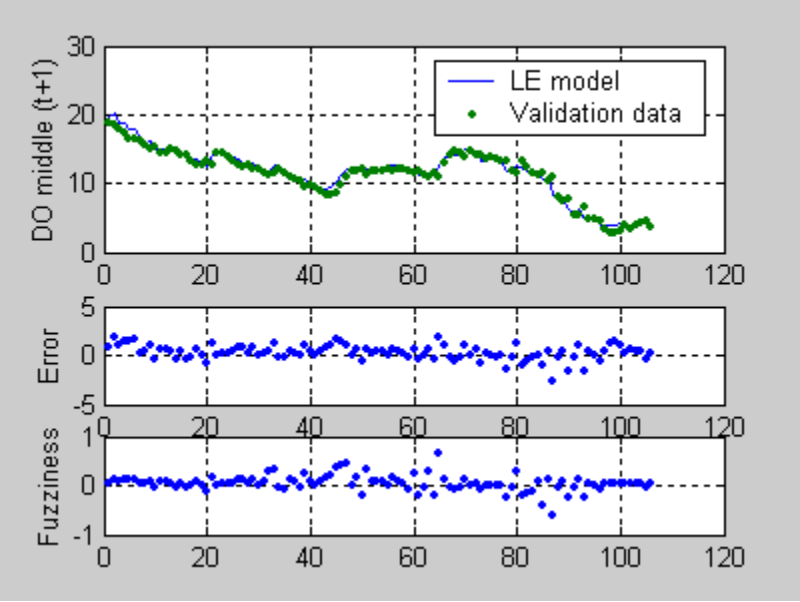


Figure 9. Testing, error and fuzziness of dissolved oxygen concentration model of exponential growth phase.

Figure 10 presents the weights of the submodels obtained from the fuzzy decision system. The first submodel, lag phase, is presented by the yellow line. The second phase is presented by the purple, and the third phase by blue line. The change from one phase to another is quite fast.

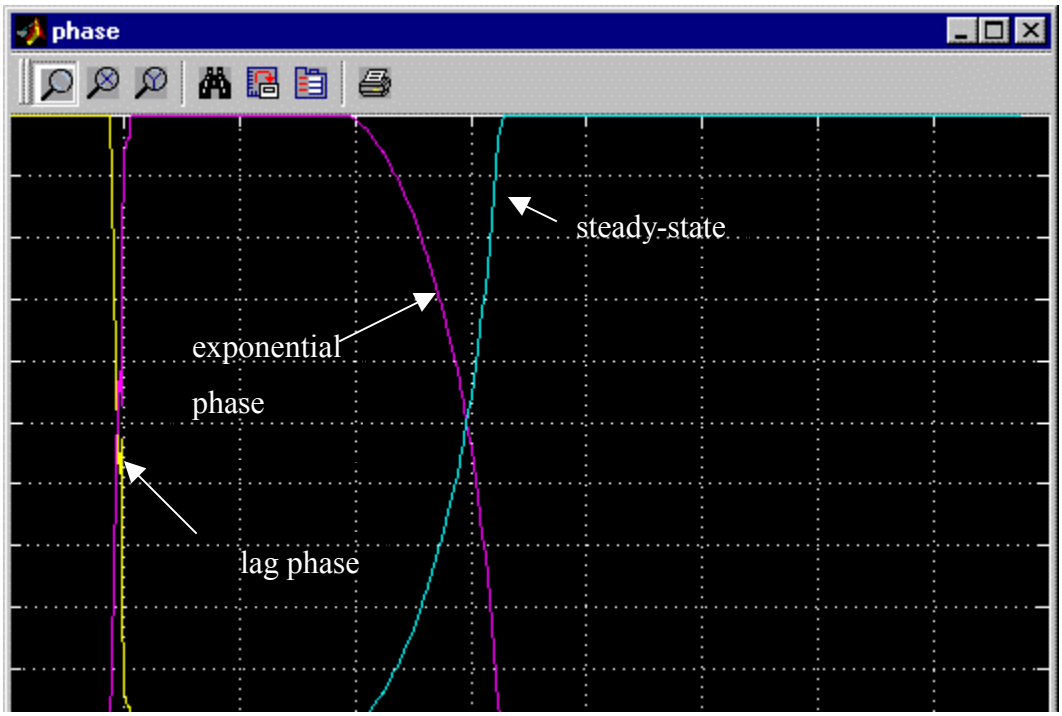


Figure 10. The weighting factors of different submodels.

In Figure 11, the estimation of the dissolved oxygen concentration is presented. In this model, the estimates of the oxygen transfer rate and the concentration of the carbon dioxide are used as inputs. The same timescale is used in all the Figures 10-13.

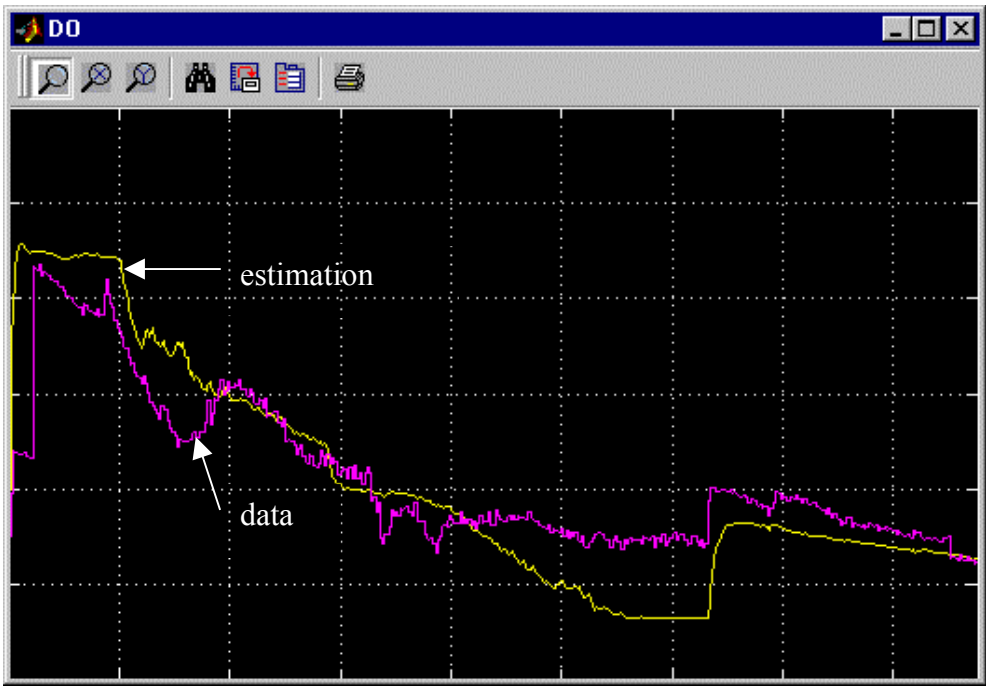


Figure 11. The estimation of dissolved oxygen concentration. The model is displayed by the yellow line and the training data by the purple line.

In Figure 12 the estimation of the concentration of the carbon dioxide is presented.

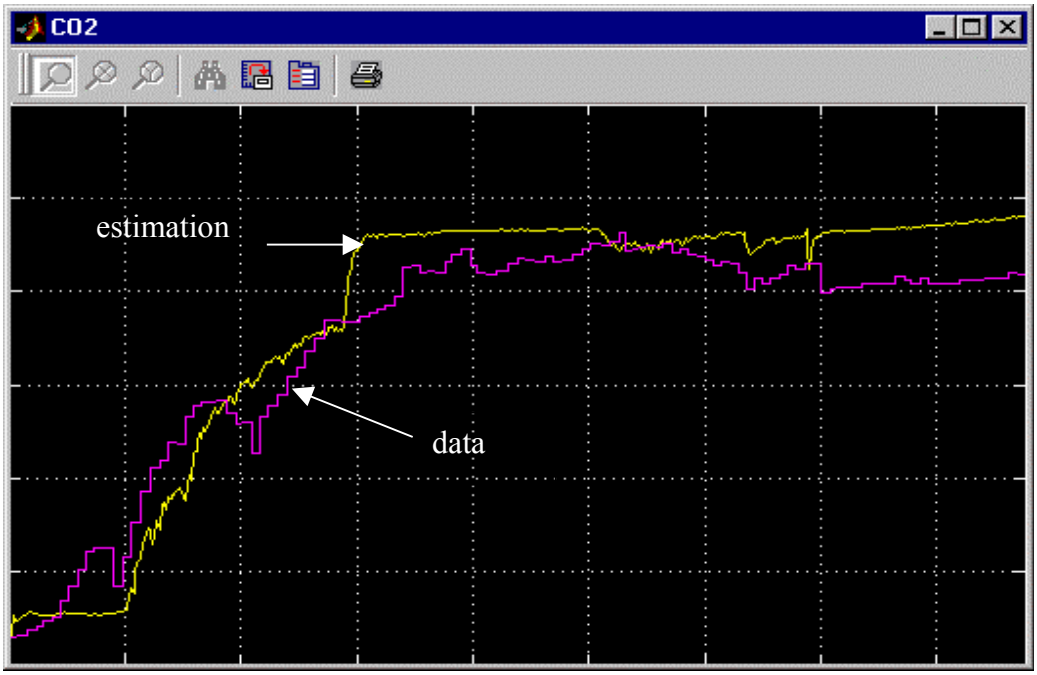


Figure 12. The estimation of the concentration of carbon dioxide in the exhaust gas. The model is displayed by the yellow line and the training data by the purple line.

The estimation of the oxygen transfer rate can be seen in the Figure 13. The estimate of the carbon dioxide concentration is used as an input of the model.

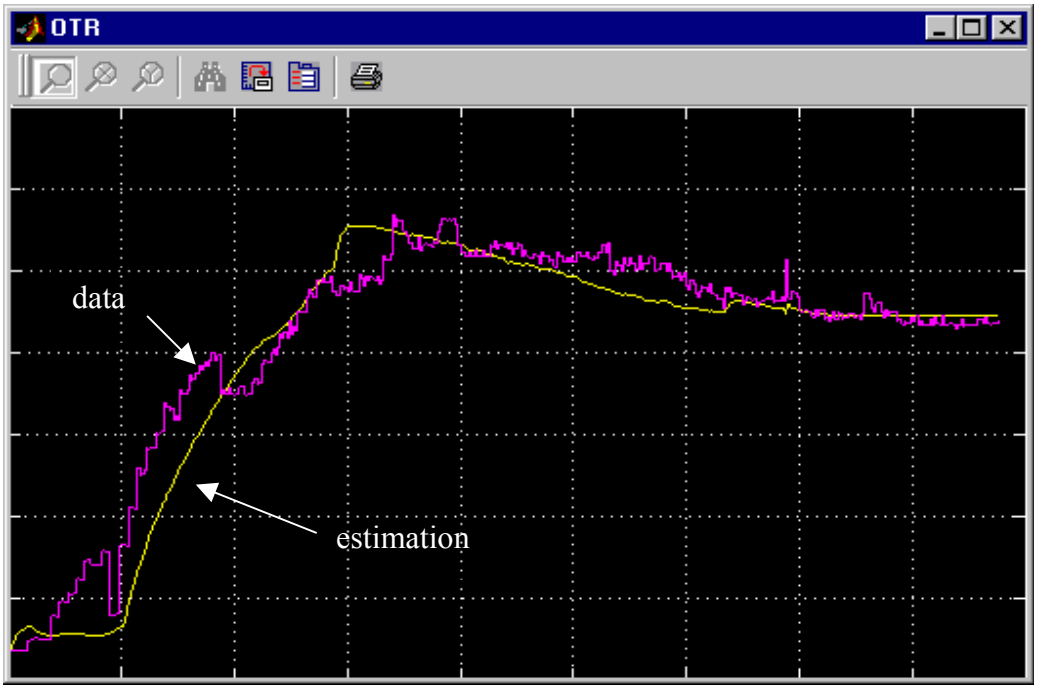


Figure 13. The estimation of the oxygen transfer rate. The model is displayed by the yellow line and the training data by the purple line.

6 DISCUSSION AND CONCLUSIONS

The results of the modelling were quite expected. The dynamic modelling proved to be a hard test for the performance of the model. The simulation results of the dynamic models for dissolved oxygen concentration, oxygen transfer rate and carbon dioxide concentration were good. The lag phase was most difficult to model. However, during the lag phase the concentration of dissolved oxygen in the fermentation broth is usually high and the prediction of it is not critical information. The linguistic equations method appears to be a suitable method for modelling of fermentation processes. These processes have been found too complicated for physical modelling /25/. Also linear models (Multiple Linear Regression (MLR), Principal Component Regression (PCR), Partial Least Squares (PLS) and Auto-Regressive Moving Average with eXogenous inputs (ARMAX)) have been applied to modelling of industrial fermentation process /26/ but their performance was not adequate enough. Artificial neural networks and NARMAX (Non-linear ARMAX) showed better performance.

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 contain enough data so that it can represent different batches. The results of the modelling can improve with the number of data runs employed for training. /20/ 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 every fermentation, the 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 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 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 amount of carbon dioxide in the exhaust gas, the mixing power, the glucose feed rate, the oxygen transfer rate, the dissolved oxygen concentration, the volumetric oxygen transfer coefficient, the position of the pressure valve and the VVM. The choice of the variables was quite similar to the choice of the modelling variables in the literature.

The concentration of the carbon dioxide in the exhaust gas is an important variable in fermentation process because the production of carbon dioxide is in proportion to the amount of consumed sugar /27/. The variations in the agitation speed can cause changes in oxygen transfer rate and an increase in it can cause an increase on production and yield of lipase enzyme /27/. In /29/ it is stated that the tension of dissolved oxygen was 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_La, is also an important process variable because it can be used to find the relationship between

OTR and enzyme production /28/ and it can be used in the control of dissolved oxygen tension /30/. The oxygen requirements of the bacteria differ at different fermentation stages /31/. By choosing a proper dissolved oxygen tension a product formation can be achieved without wasting the energy source.

The dynamic models presented in this work are used to predict the dissolved oxygen concentration, oxygen transfer rate and carbon dioxide concentration. The models are now in on-line testing. The predictions enable better operation of the process because necessary control operations can be made earlier. In the future, fault-diagnosis system is going to be developed based on the models. The models can be updated when new data is available.

REFERENCES

1. Enfors, S. - O. & Häggström, L., Bioprocess Technology Fundamentals and Applications. Royal Institute of Technology, Stockholm, 2000, 356p.

2. Madigan, M. T., Martinko, J. M. & Parker, J., Brock Biology of Microorganisms. Prentice-Hall Inc, 2000, 991 p.

3. Blanch, H. W. & Clark, D. S., Biochemical Engineering. Marcel Dekker, 1997, 702 p.

4. Gomes, J. & Menawat, A. S., Precise control of dissolved oxygen in bioreactors – a model based geometric algorithm. Chemical Engineering Science 55, 2000, pp. 67-76.

5. Buckland, B. C. & Lilly, M. D., Fermentation: An Overview. In Rehm, H. – J. & Reed, G (ed.). Biotechnology Volume 3, 2nd edition. Volume editor Stephanopoulos, G., Bioprocessing. Weinheim 1993, VCH Verlagsgesellschaft mbH. pp. 7-22.

6. Foster, K.A., Frackman, S. & Jolly, J.F., Production of Enzymes as Fine Chemicals. In Rehm, H. – J. & Reed, G (ed.). Biotechnology Volume 9, 2nd edition. Volume editor Reed, G. & Nagodawithana, T.W., Bioprocessing. Weinheim 1993, VCH Verlagsgesellschaft mbH. pp. 73-120.

7. Smith, G.M., The Nature of Enzymes. In Rehm, H. – J. & Reed, G (ed.). Biotechnology Volume 9, 2nd edition. Volume editor Reed, G. & Nagodawithana, T. W.,, Bioprocessing. Weinheim 1993, VCH Verlagsgesellschaft. pp. 5-72.

8. Lennox, B., Montague, G. A., Hiden, H. G., Kornfeld, G. & Goulding, P. R., Process Monitoring of an Industrial Fed-Batch Fermentation. Biotechnology and Bioengineering, Vol. 74, No. 2, July 20, 2001, pp. 125-135.

9. Gregersen, L. & Jorgensen, S. B., Supervision of fed-batch fermentation. Chemical Engineering Journal 75 (1999), pp. 69-76.

10. Zuo, K. & Wu, W. T., Semi-realtime optimization and control of a fed-batch fermentation system. Computers and Chemical Engineering 24 (2000), pp. 1105-1109.

11. Lennox, B., Hiden, H. G., Montague, G. A., Kornfeld, G. & Goulding, P. R., Applications of multivariate statistical process control to batch operations. Computers and Chemical Engineering, 24 (2000), pp. 291-296.

12 Lydersen, B. K., Bioprocess engineering systems, equipment and facilities. Edited by Bjorn K. Lydersen, Nancy A. D'Elia & Kim L. Nelson. John Wiley, 1994, 805 p.

13. Thibault, J., Van Breusegem, V. & Chéruy**,** A., On-Line Prediction of Fermentation Variables Using Neural Networks.. Biotechnology and Bioengineering Vol. 36 (1990) pp. 1041-1048.

14. José de Assis, A. & Maciel Filho, R., Soft sensors development for on-line bioreactor state estimation. Computers and Chemical Engineering 24, 2000, pp. 1099-1103.

15. Georgieva, O., Wagenknecht, M. & Hampel, R., Takagi-Sugeno fuzzy model development of batch biotechnological process. International Journal of Approximate Reasoning, 26, 2001, pp. 233-250.

16. Yegneswaran, P. K., Gray, M. R. & Thompson, B. G., Effect of Dissolved Oxygen Control on Growth and Antibiotic Production in Streptomyces clavuligerus Fermentation. Biotechnology Progress 7, 1991, pp. 246-250.

17. Feyo de Azevedo, S., Dahm, P. & Oliveira, F. R., Hybrid Modelling of Biochemical Processes: A comparison with the conventional approach. Computers and Chemical Engineering, 21, 1997, Suppl., pp. S751-756.

18. Juuso, E. K., Linguistic Equations for Data Analysis: FuzzEqu Toolbox. Proceedings of TOOLMET2000 Symposium. Oulu, April 13-14, 2000, pp. 212-226, Oulun yliopistopaino.

19. Juuso, E. K., Fuzzy Control in Process Industry: The Linguistic Equation Approach. In: Verbruggen, H. B., H.-J. Zimmermann and R. Babuska, editors, Fuzzy Algorithms for Control, International Series in Intelligent Technologies, pp. 243-300. 1999, Kluwer, Boston.

20. Babuska, R., Setnes, M., Kaymak, U. & Verbruggen, H. B., Fuzzy Modeling: a Universal and Transparent Tool. Proceedings of TOOLMET'97 - Tool Environments and Development Methods for Intelligent Systems. Editors: Yliniemi, L. & Juuso, E., Oulu, 1997, pp. 83-106.

21. Matlab Help. Matlab 6.1. The MathWorks Inc, 2001.

22. Saarela, U., Modelling of a fed-batch enzyme fermentation process. Master's Thesis, University of Oulu, 2002, 93p.

23. Leiviskä, K., Kokeellinen mallintaminen. Department of Process Engineering, University of Oulu, 1998, 49p.

24. Juuso, E. & Jäväjä, E., Intelligent simulation of batch cooking. In: Proceedings of TOOLMET'01 Symposium – Tool Environments and Development Methods for Intelligent Systems. Editors: Yliniemi, L. & Juuso, E. 19-20, April 2001, University of Oulu, pp. 118-129.

25. Soufian, M. & Soufian, M., Parallel Genetic Algorithms for Optimised Fuzzy Modelling with Application to a Fermentation Process. Genetic Algorithms in Engineering Systems: Innovations and Applications, 2-4 September 1997, Conference Publication No. 446, IEE, 1997.

26. Warnes, M. R., Glassey, J., Montague, G. A. & Kara, B., On Data-Based Modelling Techniques for Fermentation Processes. Process Biochemistry Vol. 31, No. 2, pp.147-155, 1996.

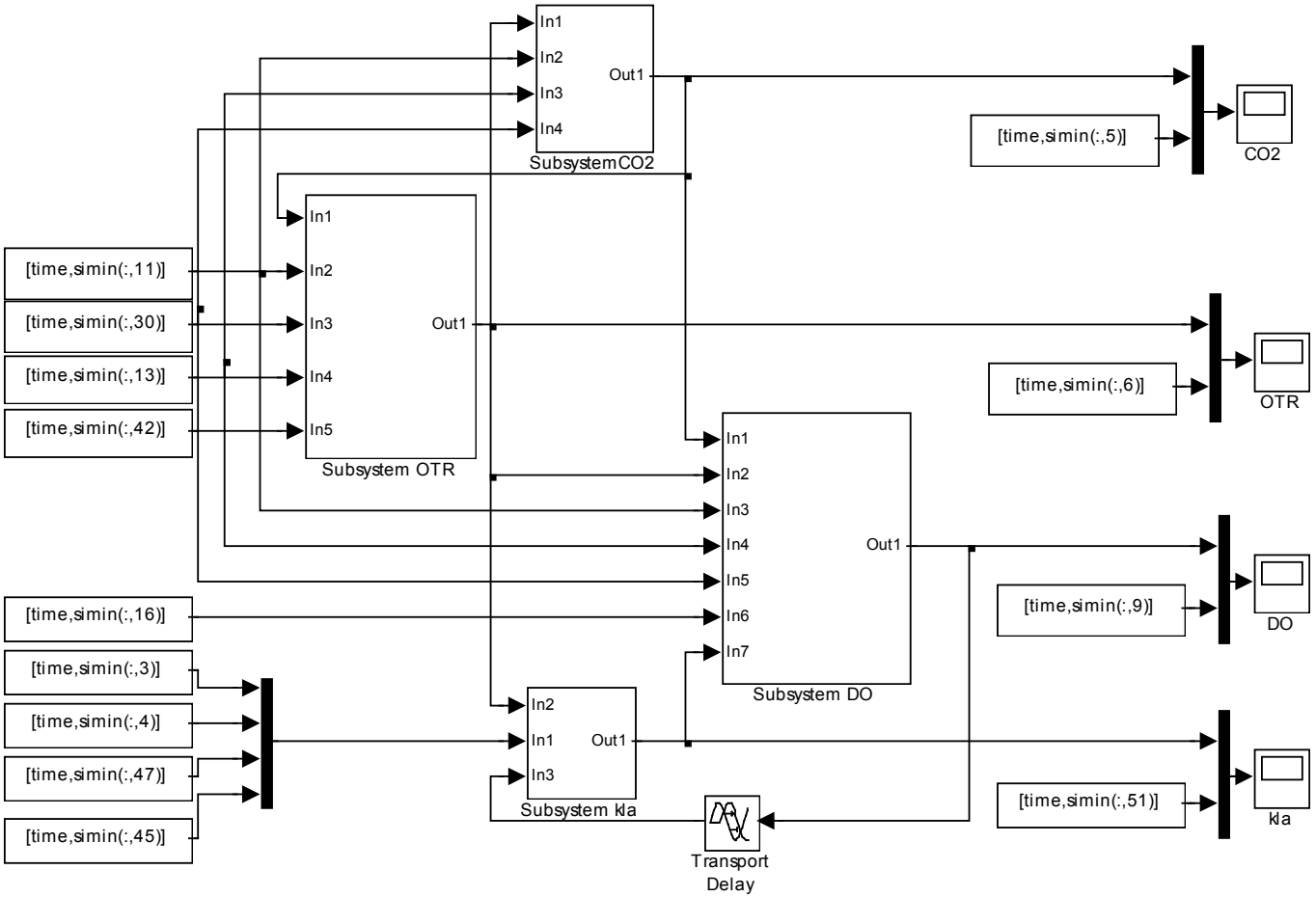
27. Martínez, G., López, A., Esnoz, A., Vírseda, P. & Ibarrola, J., A new fuzzy control system for white wine fermentation. Food Control 10 (1999), pp. 175-180.

28. Elibol, M. & Ozer, D., Influence of oxygen transfer on lipase production by *Rhizopus arrhizus*. Process Biochemistry 36 (2000), pp. 325-329.

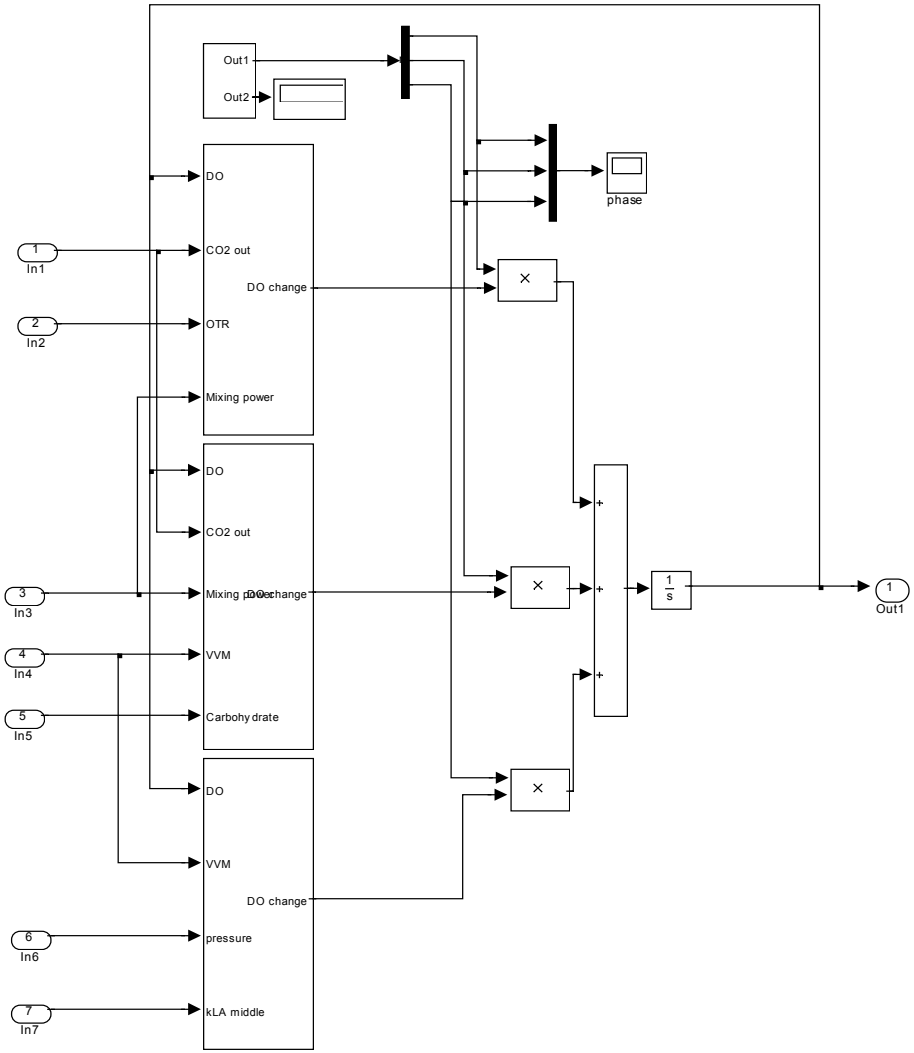
29. Pfefferle, C., Theobald, U., Gürtler, H. & Fiedler, H. -P., Improved secondary metabolite production in the genus *Streptosporangium* by optimisation of the fermentation condition. Journal of biotechnology 80 (2000), pp. 135-142.

30. Simon, L. & Nazmul Karim, M., Identification and Control of Dissolved Oxygen in Hybridoma Cell Culture in a Shear Sensitive Environment. Biotechnology Progress 17 (2001), pp. 634-642.

31. Yao, H. M., Tian, Y. C., Tadé, M. O. & Ang, H. M., Variations and modelling of oxygen demand in amino acid production. Chemical Engineering and Processing 40 (2001), pp. 401-409.



APPENDIX 1. A multimodel approach in dynamic modelling



APPENDIX 2. A dynamic model for dissolved oxygen concentration prediction

ISBN 951-42-7083-5
ISSN 1238-9390
University of Oulu
Control Engineering Laboratory – Series A
Editor: Leena Yliniemi

1. **Yliniemi L, Alaimo L & Koskinen J,** Development and tuning of a fuzzy controller for a rotary dryer. December 1995. ISBN 951-42-4324-2.
2. **Leiviskä K**, Simulation in pulp and paper industry. February 1996. ISBN 951-42-4374-9.
3. **Yliniemi L, Lindfors J & Leiviskä K,** Transfer of hypermedia material through computer networks. May 1996.ISBN 951-42-4394-3.
4. **Yliniemi L & Juuso E** (editors)**,** Proceedings of TOOLMET'96 – Tool environments and development methods for intelligent systems. May 1996. ISBN 951-42-4397-8.
5. **Lemmetti A, Leiviskä K & Sutinen R,** Kappa number prediction based on cooking liquor measurements. May 1998. ISBN 951-42-4964-X.
6. **Jaako J,** Aspects of process modelling. September 1998. ISBN 951-42-5035-4.
7. **Lemmetti A, Murtovaara S, Leiviskä K & Sutinen R,** Cooking variables affecting the craft pulp properties. June 1999. ISBN 951-42-5309-4.
8. **Donnini P A,** Linguistic equations and their hardware realisation in image analysis. June 1999. ISBN 951-42-5314-0.
9. **Murtovaara S, Juuso E, Sutinen R & Leiviskä K,** Modelling of pulp characteristics in kraft cooking. December 1999. ISBN 951-42-5480-5.
10. **Cammarata L & Yliniemi L,** Development of a self-tuning fuzzy logic controller (STFLC) for a rotary dryer. December 1999. ISBN 951-42-5493-7.
11. **Isokangas A & Juuso E,** Fuzzy modelling with linguistic equation methods. February 2000.33 p. ISBN 951-42-5546-1.
12. **Juuso E, Jokinen T, Ylikunnari J & Leiviskä K,** Quality forecasting tool for electronics manufacturing. March 2000. ISBN 951-42-5599-2.
13. **Gebus S,** Process Control Tool for a Production Line at Nokia. December 2000. 27 p. ISBN 951-42-5870-3.
14. **Juuso E & Kangas P,** Compacting Large Fuzzy Set Systems into a Set of Linguistic Equations. December 2000. 23 p. ISBN 951-42-5871-1.
15. **Juuso E & Alajärvi K,** Time Series Forecasting with Intelligent Methods. December 2000. 25 p. Not available.
16. **Koskinen J, Kortelainen J & Sutinen R**, Measurement of TMP properties based on NIR spectral analysis. February 2001. ISBN 951-42-5892-4.
17. **Pirrello L, Yliniemi L & Leiviskä K,** Development of a Fuzzy Logic Controller for a Rotary Dryer with Self-Tuning of Scaling Factor. 32 p. June 2001. ISBN 951-42-6424-X.
18. **Fratantonio D, Yliniemi L & Leiviskä K,** Fuzzy Modeling for a Rotary Dryer. 26 p. June 2001. ISBN 951-42-6433-9.
19. **Ruusunen M & Paavola M,** Quality Monitoring and Fault Detection in an Automated Manufacturing System - a Soft Computing Approach. 33 p. May 2002. ISBN 951-42-6726-5.

20. **Gebus S, Lorillard S & Juuso E,** Defect Localization on a PCB with Functional Testing. 44 p. May 2002. ISBN 951-42-6731-1.
21. **Saarela, U., Leiviskä, K. & Juuso, E.,** Modelling of a fed-batch fermentation process. 26 p. June 2003. ISBN 951-42-7083-5