METHODS FOR A TRANSPARENT DEVELOPMENT AND OPTIMIZATION OF BIOTECHNOLOGICAL PROCESSES

Peter Marenbach* Kurt Dirk Bettenhausen* Henning Tolle*

* *Darmstadt University of Technology , Institute of Control Engineering, Department of Control Systems Theory & Robotics, Landgraf-Georg-Strasse 4, D-64283 Darmstadt, Germany. E-Mail:* mali@rt.e-technik.th-darmstadt.de, *URL:* http://www.rt.e-technik.th-darmstadt.de

Abstract. The article at hand describes an integrating system for the intelligent control of complex biotechnological processes including automatic modelling and model based control strategy generation. Starting with a summary of previously achieved results, some new approaches that provide better transparency to process engineers and operators are discussed. This includes aspects of self-organizing generation of structured dynamic nonlinear process models based upon the ideas of genetic programming as well as the transparent generation of fuzzy rules in a particular NeuroFuzzy approach. The latter is used for the classification of physiological states during batch and fed-batch fermentations and for the long time strategy generation to optimize the achievable product yield.

Keywords. Biotechnology, Process control, Fuzzy systems, Learning control, Genetic algorithms

1. INTRODUCTION

The control and optimization of biotechnological processes is a complex task of industrial relevance, due to the growing importance attached to biotechnology. Therefore the number of modern intelligent approaches of computer and control engineering applied for development and optimization of bioprocesses increases.

In biotechnological productions microorganisms are cultivated which have the special property to excrete or accumulate a desired product as part of their metabolism. The lack of a complete mathematical description that arises from the incomplete knowledge on the dominant biological pathways as well as the low availability of sensor information about the current physiological state are characteristic problems. Therefore an automatic control and optimization of biotechnological processes often appears to be very difficult. In industrial practice the development of biotechnological production processes is characterized by a big number of empirical test series that are expensive and time consuming. Organisms and substrate composition have to be selected or modified by microbiologists. Furthermore a large number of experimental runs is needed to find appropriate environmental process parameters (e.g. temperature or feed-rate). During batch or fed-batch fermentations often significant alternations in the cell metabolism due to changes in extracellular conditions can be observed. Therefore it is obvious that the environmental parameters have to be changed during a fermentation in order to achieve optimal product yield. However, in industrial practice usually constant set-points are applied to the whole fermentation which are chosen because they provided the best results during test series in laboratory scale (see e.g.

Original publication in IFAC Proceedings Volumes (now IFAC-Papers online), https://doi.org/10.1016/S1474-6670(17)58770-4

Bailey and Ollis, 1986). That is due to the fact that often appropriate methods for an analysis of experimental data are not available. Neither an analytical way exists in most cases to evaluate optimal temporal sequences for the environmental parameters by biological or physical reflections. For such processes computer based learning control approaches are an attractive way for automatic control and optimization.

The following section describes the basic ideas of learning control and a successful application of an approach for learning control and optimization of biotechnological processes. Afterwards, some extensions and improvements of this approach are introduced which increase the learning systems' transparency and provide enhanced possibilities for the use of knowledge and experience of process engineers and controllers. The experimental results achieved with the improved system under real-world conditions are discussed in sect. 6. A short summary and a preview concerning the future practical and theoretical work are given in the last section.

2. LEARNING PROCESS CONTROL AND OPTIMIZATION

The basic conception of learning control loops can briefly be explained by describing the system LERNAS (Tolle and Ersü, 1992), shown in fig. 1. Basically there are

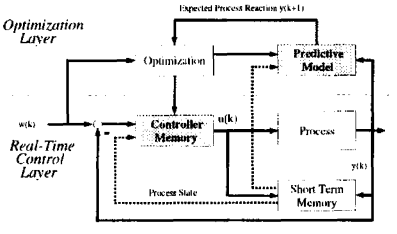

Fig. 1. Scheme of the learning control loop LERNAS (Tolle and Ersü, 1992).

two mechanisms working at the same time: First an associative memory – this could be a neural network – which is connected in parallel to the process learns the input/output mapping of a predictive process model. In order to enable a pseudo dynamic mapping several history values taken from a short term memory are used beside the actual process values as inputs of the associative memory. Second by applying different control inputs to the predictive model and assessing the predicted outputs with respect to a predefined optimization criterion advantageous control strategies are generated. Finally these control strategies are stored into another associative memory. Both, modelling and optimization, can be operated off-line based on stored process data as well as on-line.

From this learning technique certain advantages arise compared to adaptive approaches, as they were proposed e.g. by Bastin and Dochain (1990) for bioprocess control. Adaptive approaches use simplified models in which the complex nonlinear dependence on environmental process parameters are not explicitly considered. Instead they are interpreted by time-varying parameters. That means changes – e.g. of temperature – lead to a new adaption of the model's parameters. Therefore, based on such models, there is no chance to find optimal set-point sequences due to the fact that the influence of variations of the environmental parameters cannot be predicted.

Gehlen *et al.* (1992) introduced an extension of LERNAS with respect to the specific properties of biotechnological processes. The system *BioX*, shown in fig. 2, includes a number of special solutions for an integrated knowledge-based and learning control of bioprocesses:

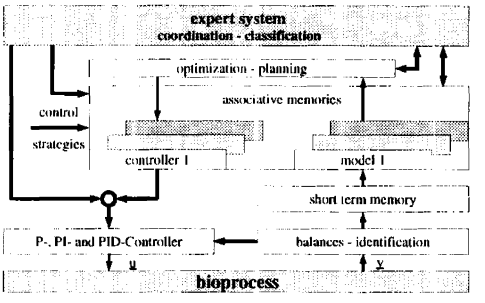

Fig. 2. Scheme of the system *BioX*.

(1) For the control of fermentations the major task is not to establish given set-points for environmental parameters at the reactor but to choose these set-points. Therefore the optimization module is used to generate optimal inputs for underlying conventional control loops.
(2) As already mentioned in sect. 1, fermentation processes are characterized by a temporal sequence of process phases, in which process behaviour can be very different. By a classification of the current physiological state combined with the use of phase specific models and control strategies, easier generation and better quality of the predictive model can be achieved (Halme, 1989; Gehlen and Bettenhausen, 1990). For this reason a rule based phase classifier was realized.
(3) Finally, a rule-based fault detection and a plausibility check for the generated control action was supplied to the process control system.

BioX was successfully applied to the process control of an α-amylase production with *B. subtilis*. By optimized dynamic variation of the environmental process parameters the product yield was increased by more than 100%.

In spite of these good results achieved at a real-world process with an industrial production strain there still remain some arguments against a wide application of this method in industrial practice:

(1) The black box structure of associative memories used for the control strategies does neither provide an insight into the optimized strategies to an operator nor offers him a simple possibility to introduce his personal experience.
(2) The configuration of a classical expert system for a robust and faultless detection of changes of the current process phase is very difficult.
(3) The predictive process model, which is realized as a black box too, enables the numerical optimization of set-points based on short-term predictions. However, long-term predictions, which are in particular useful for bioprocesses, based on this type of models are difficult in most cases. Furthermore, for the process engineer a trained associative memory does not provide an insight into how a measured value influences the process behaviour. This information would be of great interest to draw conclusions for the process development.

3. AUTONOMOUS GENERATION OF TRANSPARENT CONTROL STRATEGIES

During the last few years the mentioned points were systematically investigated. The extended concept *BioX*$^{++}$ was first presented by Bettenhausen and Tolle (1993). A homogeneous object oriented implementation concept was chosen to overcome the explicit separation between knowledge based and learning layer. Since that time a number of new techniques were developed to provide a better transparency and to make the system more user-friendly (see also Bettenhausen *et al.*, 1995*b*).

The usage of existing ideas and experiences of process operators – which are expressed in a linguistic way in general – are the foundations of the popularity of fuzzy controllers designed and applied to complex nonlinear systems like chemical or biotechnological processes. The linguistically initialized fuzzy inferences as well as the trainable neural networks are based upon a function approximation by static nonlinear mappings with a weighted superposition of basis functions. The usage of this similarity – which can also be observed in the internal representation (cf. for example Brown and Harris, 1994) – leads to optimizable or trainable fuzzy systems, the so called *NeuroFuzzy* approaches.

A fuzzy approach offers several possibilities for automatic manipulation and optimization, i.e. the introduction of learning capabilities, examples are the use of parameterizable operators for union and intersection, the modification of number and shape of the membership functions and the weighting of existing rules. An extension of the last mentioned approach – that we call *completely defined rule space* – expands the well known rule matrix (fig. 3) by introducing weights for each term of a linguistic output variable (fig. 4). The value of 1 means

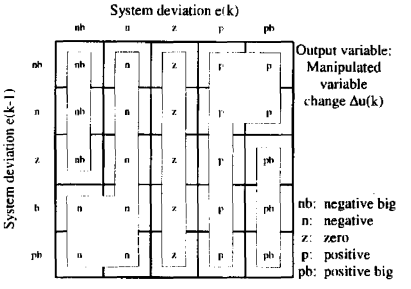

Fig. 3. Standard rule matrix of a fuzzy-controller with two inputs $(e(k-1), e(k))$ and one output $(\Delta u(k))$.

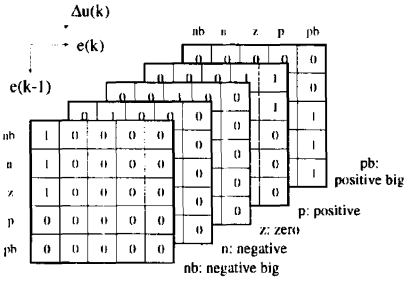

Fig. 4. Completely defined rule space of the same fuzzy controller as in fig. 3.

that the conjunction of all input terms fully activates the concerned output term. A rule base represented in this extended form can be initialized by the process operator using the conjunctive normal form of the existing or available rules and setting the corresponding weights to 1. Afterwards this initial set of weights can automatically be modified in order to optimize the fuzzy system's behaviour, i.e. wrong rules can be deleted by reducing their weight and new rules can be generated by increasing their weight.

Due to previously gained theoretical and practical experiences *algebraic product* is used as t-norm and *algebraic sum* as t-conorm operator in our actual implementation. The *center-of-singleton* defuzzyfication is the standard method for time-reducing implementations. For the efficient optimization of the rule weights, algorithms were developed integrating a local parameter search algorithm and stochastically stimulating components for reducing early optimization ending in local extrema.

Bettenhausen *et al.* (1995*c*) presented first experimental results using this approach for the generation and storage of strategies for the control of a – still simulated – biotechnological batch fermentation. These experiments have shown, that a significant increase of the product yield can be achieved – similar to the results gained by the usage of the interpolating associative memory described in sect. 2. The major advantage of the new approach is the increased transparency caused by the representation of control strategies as a readable set of fuzzy rules and the possibility of initializing the system with existing linguistic prior knowledge in a simple way.

4. SELF-ORGANIZING GENERATION OF STRUCTURED MODELS

The disadvantages that arise from the use of a pure input/output model of the biotechnological process for control purposes were already mentioned. The missing transparency of neural models, means that since the input/output behaviour is approximated by a black box approach no direct insight into the process and its underlying relationships can be gained. This usually leads to a low acceptance of these approaches in industrial applications. Another point is that a process engineer usually wants to have a structured mathematical description of his process, even if an associative memory which perfectly reproduces the observed process behaviour would be available. This is due the fact that information about what influences the different process variables have can be an important help for the optimization of the biological side of the process, too.

A new approach that we call *self-organizing generation of structured models* is an attempt to overcome these disadvantages. The general idea is to automate the iterative methodology of empirical modelling used by a process engineer. Therefore, existing knowledge of biochemical experts on structural properties should be taken into account. Fig. 5 shows the basic scheme of self-organizing model generation, which is indeed very similar to the way models are developed by a process engineer. The algorithm distinguishes between two tasks: One is the optimization or identification of the structure's inherent set of parameters which is achieved by well known conventional methods. The other even more interesting task is the symbolic generation of an appropriate model structure which is done by means of genetic programming (for an introduction see Koza, 1992).

Block diagrams which are often employed in control theory are used for the symbolic representation of process models. In order to apply the methods of genetic programming for generation and modification of these models they are internally represented in a tree structure. Fig. 6 shows an example of a simple model depicted as a block diagram (left) and the equivalent genetic representation in a tree structure (right).

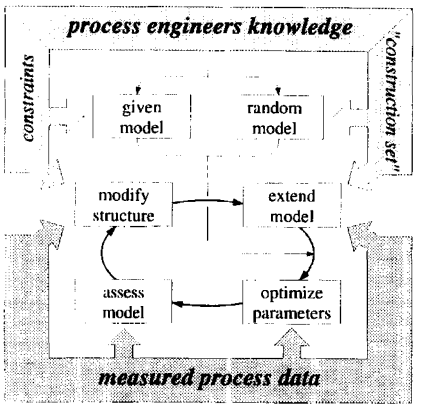

Fig. 5. Scheme of automatic generation of structured process models.

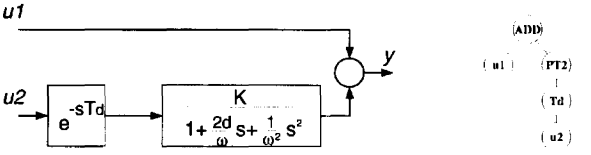

Fig. 6. Simple process model: depicted as a block diagram (left) and as a tree structure (right).

Starting with a collection of elementary transfer elements like time-delay or Monod kinetics placed in a so called "model construction set" a number of models is created. The degree of how much this is done by random depends on how much a priori knowledge is available. In an evolutionary process the following three steps are performed iteratively. First each model of a generation is adapted to measured process data by optimizing its internal parameters, using well known parameter search methods. After that a fitness value is evaluated for each model by assessing its accuracy and complexity. Directed by this fitness value new models are created by modifying and extending the actual model's structure. This iterative methodology finally leads to models that hopefully combine high accuracy and low complexity, which are needed for most kinds of control and design purposes. A priori knowledge on structural properties can be taken into account in this process by constraining the elements in the model construction set and by influencing their selection frequency. Furthermore, certain parts of the models, e.g. basic balance equations of fermentation processes, can be predefined and elements can be combined to "super-blocks" that are treated as if they where single elements. Therefore both, predefined

submodels and super-blocks, cannot be divided by genetic operators.

A detailed description of this approach as well as first experimental results taken from a cooperation with BASF AG, Germany, in which the self-organizing modelling technique was applied to a biotechnological fed-batch fermentation have been published by Bettenhausen *et al.* (1995*a*).

5. LEARNING DETECTION OF PHASE TRANSITIONS

Another property of biological processes, which causes difficulties in process control, is that even under identical measured experimental conditions different fermentations produce characteristic but nevertheless considerable differing time responses. Together with the difficulties that arise from alternations in the behaviour of the microorganism due to changes in cell metabolism this leads to a point where even data driven modelling becomes a very complex task.

An efficient way to overcome these difficulties is to divide the process behaviour into characteristic physiological states that correspond to the different phases (see Halme, 1989; Gehlen and Bettenhausen, 1990). If now phase specific models are used instead of trying to find an accurate model for the whole fermentation modelling becomes much easier, because only those effects have to be considered that are dominant in the actual phase.

For the system *BioX* described in sect. 2 an expert system based on the programming language Lisp was used for recognition of the different process phases. The disadvantage of this approach was, that rules in the way they are formulated by an expert are not enough to achieve a robust and faultless recognition. In their fundamental paper Konstatinov and Yoshida (1989) already proposed the use of a learning fuzzy-approach – similar to the one described in sect. 3 – for the representation of control strategies.

This offers the possibility to first take the linguistic formulated rules for initialization and then optimize the rule base according to phase separations of measured process data which also should be done by an expert. A tool that provides the functionality to do this separation graphically on a computer screen and to optimize the fuzzy rules afterwards together with first experimental results were part of the oral presentation of Bettenhausen and Tolle (1993).

6. APPLICATION

The previous sections discussed a couple of methods and approaches which are part of the integrating system conception *BioX*$^{++}$. Due to several practical necessities, these methods were examined and applied to different processes.

In order to summarize this work and to give some general information about the flexibility of the developed tools, a recipe for the application and some benchmark values to make the tools comparable to other developed approaches, the actual examinations are concentrated on a complete application to a process examined together with the Institute of Biochemistry at Darmstadt University of Technology [1] . This process is the production process of α-amylase with *B. subtilis*. The complete

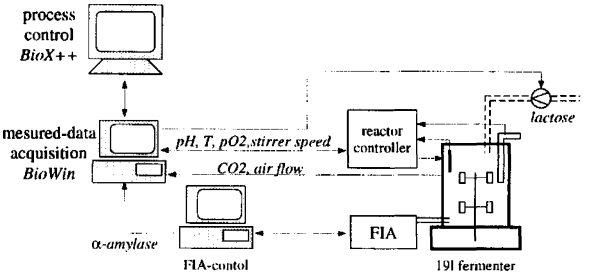

Fig. 7. Experimental set-up.

installation based upon a 19*l* fermenter with flow injection analyzer (FIA) for the on-line measurement of α-amylase concentration is shown in fig. 7. For the control of the FIA and the data acquisition conventional PCs are used. The additional components – integrated in the system *BioX*$^{++}$ – take place on a UNIX workstation.

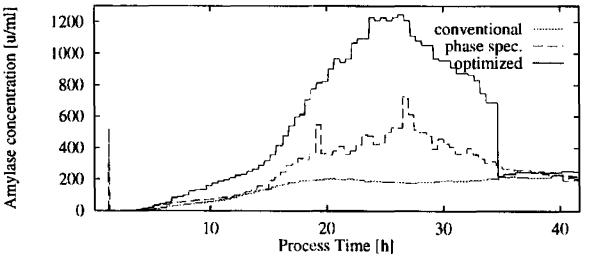

Fig. 8. Experimental results.

Fig. 8 shows actual results of three fermentations demonstrating the enormous capabilities and robustness of the integrating approach. A small number – only one is shown – of runs with constant set-points for temperature and pH was used as training data for the phase classification. Then phase specific constant set-points accord-

[1] These examinations are partially granted by the German Ministry of Education and Research (BMBF).

ing to a priori knowledge were applied (cf. fig. 9). The resulting data was used for training of an associative memory. The off-line generated short-time strategies increased the product yield considerably, particularly taking the previously gained results into account. However,

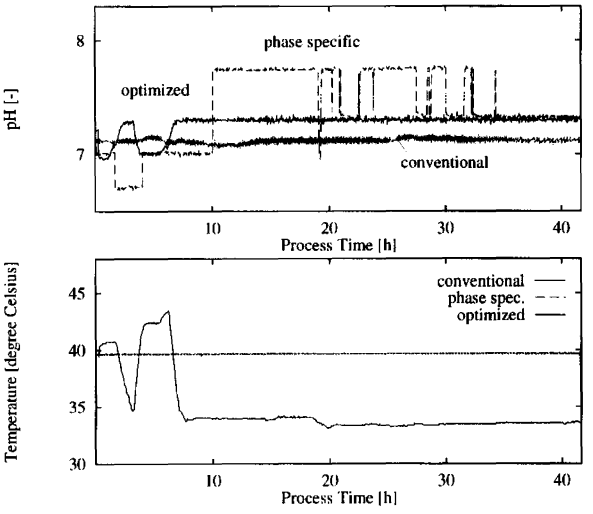

Fig. 9. Environmental parameters: pH an temperature.

the structured modelling approach has not been applied to *B. subtilis* up to now.

7. CONCLUSIONS

The article at hand describes the characteristic main parts of the system *BioX*$^{++}$ for the autonomous control of biotechnological batch and fed-batch fermentations. The system is based on two modelling approaches – neural modelling for continuous short-time optimization and data-driven self-organizing generation of structured models by the means of genetic programming as a basis for the generation of long-term strategies and for the increase of process understanding. Due to the well-known fact that different dominant effects can be observed in different process phases, the expenditure of modelling can easily be decreased by a stable classification and separation of these characteristic phases. A transparent approach based on existing operator knowledge and a data-driven extension and optimization is discussed and presented in detail. Current work is concentrated on an intense examination of the different techniques and the transition to fed-batch operating mode.

8. REFERENCES

Bailey, James E. and David F. Ollis (1986). *Biochemical engineering fundamentals.* 2 ed.. McGraw-Hill. New York.

Bastin, Georges and Denis Dochain (1990). *On-line Estimation and Adaptive Control of Bioreactors.* Elsevier Science. New York.

Bettenhausen, Kurt Dirk and Henning Tolle (1993). *BioX*$^{++}$ – extended learning control of biotechnological processes. In: *IFAC World Congress.* Vol. 7. Sydney, Australia.

Bettenhausen, Kurt Dirk, Peter Marenbach, Stephan Freyer, Hans Rettenmaier and Ulrich Nieken (1995*a*). Self-organizing structured modelling of a biotechnological fed-batch fermentation by means of genetic programming. In: *Proc. Int. Conf. on Genetic Algorithms in Engineering Systems: Innovations and Applications.* Vol. 414 of *IEE Conference Publication.* London, UK.

Bettenhausen, Kurt Dirk, Stefan Gehlen, Peter Marenbach and Henning Tolle (1995*b*). *BioX*$^{++}$ – new results and conceptions concerning the intelligent control of biotechnological processes. In: *Proc. 6th Int. Conf. on Computer Applications in Biotechnology.* Garmisch-Partenkirchen, Germany.

Bettenhausen, Kurt Dirk, Stefan Möller and Henning Tolle (1995*c*). Autonomous and transparent generation of control strategies. In: *Third European Control Conference ECC '95.* Vol. 2. Rome, Italy.

Brown, Martin and Christopher J. Harris (1994). *Neurofuzzy Adaptive Modelling and Control.* Prentice Hall.

Gehlen, Stefan and Kurt Dirk Bettenhausen (1990). Modelling of biotechnological processes with interpolating associative memories. In: *Int. Symposium on Mathematical and Intelligent Models in System Simulation.* Brüssel.

Gehlen, Stefan, Henning Tolle, Jürgen Kreuzig and Peter Friedl (1992). Integration of expert systems and neural networks for the control of fermentation processes. In: *IFAC Symposium on Modelling and Control of Biotechnological Processes.* Keystone, Colorado, USA.

Halme, Aarne (1989). Expert system approach to recognize the state of fermentation and to diagnose faults in bioreactors. In: *4th Int. Conf. Computer Applications in Fermentation Technology.*

Konstatinov, Konstatin and Toshiomi Yoshida (1989). Physiological state control of fermentation processes. *Biotechnology and Bioengineering* **33**.

Koza, John R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* The MIT Press. Cambridge, Massachusetts.

Tolle, Henning and Enis Ersü (1992). *Neurocontrol.* Vol. 172 of *Lecture Notes in Control and Information Sciences.* Springer-Verlag.