Knowledge-based Modeling of Multi-factor Processes in Biotechnology and Microbial Ecology

SVETLA VASSILEVA
Department of Integrated Systems
Institute of System Engineering and Robotics – Bulgarian Academy of Sciences (ISER-BAS)
Akad. Georgi Bonchev str., bl.2, POB 79, 1113 Sofia,
BULGARIA
si27v06@gmail.com http://www.iser.bas.bg

Abstract: - Biotechnological and ecological processes are multi-factor nonlinear system, its dynamic could be considered as sequence of phases. Bacterial growth in batch culture can be modeled as a sequence of four integrated phases: lag phase, exponential or log-phase, stationary phase, and death phase. Ecological processes are connected with the seasonal changes for certain period of time – one season, one year, a decade or a century. Methodologies which can provide their adequate mathematical descriptions are based on the synthesis of local MIMO-models; the transition between phases is realized by using time or state conditions markers in form of IF-THEN rules, expressing complex relations between influential input-output variables. Obtaining of such relations is a nontrivial task. For this reason human expertise and learning capacity of modern AI-approaches is embedded.
Main purpose of the presented paper is to demonstrate these opportunities on some multi-factor and multiphase biotechnological processes. The application of knowledge-based system on the multiphase processes is presented in connection with monitoring and inferential measurements systems development.

Key-Words: - knowledge-based systems, intelligent industry, multi-factor nonlinear system, multiphase modeling, artificial intelligence, biotechnology, microbial ecology

1 Introduction

Today's explosive growth in innovations, new products and services can be attributed to earlier investments in developing knowledge in areas such as biology, chemistry, earth sciences, ecology, space and nuclear sciences, coupled with phenomenal advances in information and communication technology. Economic growth is accelerated with the investment in knowledge stocks and knowledge infrastructure. Such stocks of knowledge form part of the intellectual capital and can profitably be invested by both the public and private sectors. Transformation towards a knowledge-based economy will necessarily shift the proportion and growth of national income derived from knowledge-based industries, the percentage of the work force employed in knowledge-based jobs and the ratio of firms using technology to innovate.

Nowadays in the practice are known several categories of knowledge-based systems (KBS) as well as expert systems (ES) [45], linked systems [16], intelligent tutoring systems, case-based systems and intelligent user interface for databases [44,45]. Their objectives lies in providing a high intelligence level, assisting people in discovering new knowledge in different areas, developing unknown fields, aiding management, acquiring new perceptions by simulating unknown situations, improvement of software productivity and reduction of costs and time to develop computerized systems [23,36,45].

Modelling of real world systems is mostly carried out in the badly defined domain, uncertain or incomplete information. For this reason now AI approaches are widely implemented [2,18,20,26,27,28,30,32,34,35,36,37,38,43,47,55,56, 57,58,60,61,63]. In general AI-model-based reasoning refers to the inference method used in Expert Systems - computer systems that emulate the decision-making ability of a human expert [44].

With this approach, the main focus of application development is developing the model. Then at run-time, an "inference engine" combines this model knowledge with observed data to obtain conclusions such as a diagnosis or a prediction. Knowledge can be represented using causal rules in a model-based reasoning system.

The term "linked systems" is associated with a list of linked elements and usually chain method of connection between them [16]. Another method is to actually link each element of the list with a mental picture of an image that includes two elements in

the list that are next to each other. There are three limitations to the link system: first there is no numerical order imposed when memorizing, hence the practitioner cannot immediately determine the numerical position of an item; the second appears if any of the items is forgotten; the third is the potential for confusing repeated segments of the list, a common problem when memorizing binary digits.

The first problem can be solved by bundling numerical markers at set points in the chain. Solution of the last limitations is presented in [16, 23].

Intelligent Tutoring Systems (ITS) are computer-based tutors which act as a supplement to human teachers. The major advantage of an ITS is that it provides personalized instructions to the users according to their cognitive abilities. The classical model of ITS architecture has three main modules – domain, model, user model and teaching model.

There are many other forms of case-based models that may be used - quantitative models with mathematical equations or qualitative models, which use cause/effect models. They may include representation of uncertainty, behaviour over time, "normal" or abnormal behaviour. Model types and usage for model-based reasoning are discussed in [43].

Modern knowledge-based systems are defined as systems, its fundamental principle lies on the methods and techniques of artificial intelligence (AI).

Knowledge exist in the ordinary (in the human-expert memory, in the textbooks) and in the formalized form. In systems with artificial intelligence are submitted to external, logical and physical level [51,54-61]. Knowledge is classified on declarative (descriptive), which presents established facts with the nature of axioms in the formal logic and on procedural, with the algorithmic (formalized) presentation that creates conditions for a better flexibility of the knowledge base. The formalization of knowledge uses the principles of knowledge representation that is a part of the theoretical ideas of knowledge engineering.

Basic components of KBS are: knowledge-base (KB), which serves as a repository of domain knowledge and meta-knowledge; inference engine (IE) which is a software program that infers the knowledge available in the KB and enriches the KBS with self-learning capabilities, provides explanation and reasoning capabilities; friendly interface to users.

Advantages of recent KBS are permanent documentation of knowledge, cheaper solution and easy availability of knowledge, dual advantages of effectiveness and efficiency, consistency and reliability, justification for better understanding, self-learning and ease of updates [3-6,15,33,50,53].

The difficulties are due to the incompleteness of KB, identification of the characteristics of knowledge, large size of KB in most cases, acquisition of new knowledge and its verification, often slow learning and execution, development of the model and standards.

Our work is devoted to the KBS implementation in modeling and inferential measurements of key process parameters in biotechnology and microbial ecology.

2 State of the art in Biotechnology

Various engineering solutions were offered in the last 50 years with the introduction of knowledge-based and multi-agent systems [3-6,15, 24,25,36,37,38,40,43,44,50,53,54,59]. Nowadays by Internet-based technologies new sources of knowledge and data were created. The Worldwide Protein Data Bank (wwPDB) [5,6] consists of organizations that act as deposition, data processing and distribution centres for protein data. On importance of such databases shows fact the founding members are Protein Data Bank in Europe (PDBe) [62], Protein Data Bank in Japan (PDBJ), Research Collaboratory for Structural Bioinformatics Protein Database in USA(RCSB PDB), and Biological Magnetic Resonance Data Bank in USA (BMRB) [3,4,15,31]. Each member's site can accept structural data and process these data. The processed data is sent to the "archive keeper", which is RCSB PDB. This ensures that there is only one version of the data which is identical for all users. The modified database is available to the other wwPDB-members, each of whom makes the resulting structure files available through their websites to the public.

The UniProt Knowledgebase (UniProtKB) is an expertly accurate database, a central access point for integrated protein information with cross-references to multiple sources. The UniProt Archive (UniParc) is a comprehensive sequence repository, reflecting the history of all protein sequences. UniProt [3] Reference Clusters (UniRef) merge closely related sequences based on sequence identity to speed up searches. While the UniProt Metagenomic [3] and Environmental Sequences database (UniMES) were created to respond to the expanding area of meta-genomic data [5,6].

Certainly, there are still many other databases available on the Internet - Protein Database in the National Center for Biotechnology Information –

Swiss, ENZYME of Enzyme Commission, Protein Geometry Database (PGD), Structural Classification of Proteins (SCP), Protein Research Foundation (PRF), etc.

A wide variety of software tools and systems were developed to solve requirements for high quality of life taking under consideration quality of produced industrial goods and saving environment [1,4,9,10,11,17,18,26,30,31,32,35,49,52,57,61,63, 64].

Traditional engineering interest and importance in biotechnology are: biomass feedstock pre-treatment, microbial strain selection, fermentation process optimal control and integration, process engineering and life cycle analysis, identification and characterization of reaction products and intermediates, kinetic of biomass transformation into active bio-substances, food safety and health, plant protection research, etc.

The inherent complexity of biotechnological processes makes the measurement problem significant. Because the measured variables are often interrelated, the measurements are inexact and uncertain. Mutual dependence among variables have to be taken into consideration when measure concentration of oxygen, carbon dioxide and other gaseous species in the exit gas, which frequently contains volatile substrates/products [7,8,10,57]. In addition, the gases concentration depends on the cultivation temperature; microbial respiratory processes are connected with viability of cultivated culture, specific morphological characteristics of cells and cultivation conditions in general.

The last years have shown that lack of accurate mathematical models limits the more realistic point of view and model accuracy, both considered in sense of the complex characteristics of the mathematical expression of biotransformation processes [7,8,9,12,13,14,20,22,26,29,30,33,40,42]. Some solutions of this problem offered in 70^{-th} years of the 20^{-th} century implementation of the flexible soft-computing techniques with "black-box" modeling, model-based predictive control [8,20,26, 27,28,30,37,38,47,52,55,56,60,62] and inferential control schemes by using inferential estimators'[1,7,8,12-14,21,27,28,54], adaptive control systems [1,12,42,48,62], etc.

The potential information about multi-factor processes regarding factors affecting plant operation might be obscured by the complete volume of data collected [18,51,55]. In addition, the process of data mining can be difficult because of high dimensionality, noise and low accuracy, redundant and incorrect values, non-uniformity in sampling and recording policies. Careful investigation of available data is required in order to detect either missing data or outliers, due to faults of measuring or transmission devices or to unusual disturbance, which can have unwanted effects on model quality. Skills and knowledge of plant experts should be considered a precious support to any numerical data processing approach.

In [31,32], independent component analysis (ICA) is used to process data in comparison with principal component analysis (PCA) relative to biological waste water treatment and best predictors selection [17,55,59].

Today, the introduction of intelligent systems with the realization of non-linear data mappings algorithms, owing effective computing properties like "low order interpolation" and "universal function approximation", parallel processing and learning capabilities, have been recognized as an attractive alternative to the on-line soft-sensors and indirect measurement systems design [1,7, 8,9,10,12,13,19,20,26,27,28,30,34-39,40,41-44,47,49, 55-63].

The robustness of the intelligent model-based control systems, implemented in soft-sensors and inferential measurement systems ensures stability of the overall system in the presence of external disturbances and uncertain information.

From the commercial point of view there are two groups of innovative solutions for monitoring the derived parameters, which characterize the trade biotechnological product rating: index of performance (quantitative measure) which presents relation between obtained bioactive substances volume and substrate and/or power costs and final product composition and quality (qualitative measure) - product purity, potency, stability, safety, specific organoleptic particularities, etc. [17,35,55]. With the focus on green biotechnology some tailor-made solutions, which are soon of great benefit for the brewery due to their low investment and operating costs, are offered by GEA Brewery Systems GmbH [25].

For example, bio-ethanol is produced by the biological fermentation of carbohydrates derived from plant material and from the food industry, beer and wine brewing. Theoretical Ethanol Yield Calculator of ethanol from feedstock calculates the theoretical ethanol yield of a potential feedstock for biomass ethanol production when the dry mass percentage of the material that is sugar components is known [24].

In the [63], a hybrid model of the differential catalytic hydrogenation reactor of carbon dioxide to methanol is proposed. The model consists of two parts: a mechanistic model and a neural one. The

mechanistic model calculates the effluent temperature of the reactor by taking outlet mole fractions for a neural model. The authors show that the hybrid model outperforms both a first principles model and a neural network model using the available experimental data.

An effort to include prior knowledge of a process into neural models in such a way that the interactions between the process variables are represented by the network's connections by means of regression networks is presented in [52]. A regression network is a framework by which a model structure can be represented using a number of feed-forward interconnected nodes, where each of them is characterized by its own transfer function. Black-box regression techniques are compared to the regression network and the latter is shown to give better performances.

Artificial neural networks (ANN) are implemented in solving various ecological processes as well as: toxicity prediction of aqueous effluents from specialised organic chemicals processes [55], biodegradation of naphthalene - a poorly water-soluble polycyclic aromatic compound (PAHs) [61], inhibitory effect of furfural on the bio-productivity of lactose-assimilating strains [49].

ANN are implemented in beer brewing for predicting hard-to-analyse factors of aging, taste and flavour of beer, oriented to the Total Antioxidant Capacity (TAC) measurement by FRAP (Ferric Reducing/Antioxidant Power), content of glutathione and total phenols analysis [35]. In wine manufacturing neuro-fuzzy systems are used for morphological parameters of yeast strain and quality rating prediction [55].

The fuzzy logic method improved by adaptive learning of a fuzzy inference system is used to demonstrate a sufficient solution of software analyzer-based measurement of hard-to-analyze variables – metal ions uptake and percent of metal ions removal [26]. A fuzzy clustering algorithm is applied to find the rule base of a soft sensor designed to infer the top composition of a distillation column [36].

The recent concept of intelligent industry lies on the network of interacting intelligent systems for data acquisition, assessment, interpretation, decision support and control.

3 KBS for modeling of multi-factor processes – examples and discussion

Multi-factor processes are characterized by the following features:

- A large number of: control effects; controllable and uncontrollable environmental factors; qualitative and quantitative composition of some input material flows; parameters of the state of the technological environment and the parameters of the final output product;
- Non-linearity, non-stacionarity and/or quality uncertainty; presence of recirculating material flows; significant transportation delays between the input and output variables;
- Limited ability or inability for active impact on the identification of the object.

Principle of incompatibility between increasing complexity and high accuracy of the model is valid for multi-factor processes. For this reason new solutions are designed, in which the object model and control law are replaced by the knowledge-base. At present there are not absolutely comprehensive knowledge bases, because of the limited capacity of existing models for the representation of knowledge; incomplete knowledge of specific subject areas; the imperfection of methods of acquiring knowledge, etc.

In the kknowledge engineering "model for knowledge representation" is a term which shows a way to describe the subject area (notions, relations). Four types of models are implemented: logic, production, semantic and frame networks. Logical models are in the form of formulas that contain constants, variables, functions, predicates, logical links and quantors. Production models are sets of mutual connections and complementary production rules of the type: IF (condition, situation), THEN (the conclusion, action).

Methods for extraction, processing and representations of knowledge are conventional and Data Mining. Conventional approach involves complex research of experts in subject area. Data Mining is a modern approach is equivalent to the knowledge discovery in databases. Knowledge-discovery methods are pattern recognition, classification, clustering, decision tree, graphs, regression analysis, neural networks, induction of rules, etc.

The theory of fuzzy logic lies in the core of qualitative concept, where variables are natural language terms or sentences, known as linguistic variables its value expresses the quality particularities, defined by a fuzzy set. Object description is realized by the relations "IF Ai, THEN Bi, where Ai and Bi are fuzzy sets of the universal multitudes U, V. This relation is known as logical linguistic model.

Ability of ANN to distinguish objects from an object area on the basis of input data is an approach

for knowledge extraction. Objects can be visual, acoustic (sounds), electromagnetic, thermal, radiation signals or objects. For representation of situation of complex systems are established sets of characteristic symptoms (indicators). Objects with the similar symptoms form a class of objects each of them has its limits in the space of indicators. These limits are presented by disjunctive functions. Arguments in these functions are indicators and values of functions – percept class of patterns.

Main problems of aartificial intelligence refer to merge and joint different approaches ((hybrid approach) to retrieval and represent knowledge; to generate logical conclusions in relation to the area of application. For the purposes of such tasks at the beginning are used symbolic approaches (production rules, frames, semantic networks). Their advantages are good readability, clarity, easy handling of complex data structures (lists, trees, columns, frames) for the representation of human knowledge; the clear separation of knowledge from mechanisms for knowledge management; modularity, enabling easy expansion of knowledge bases; easy application of formal, analytical approach for logical conclusions. Despite these advantages, there are a number of restrictions which impede their effective implementation in practice: difficult adaptation to changes in the environment and the incoming data, greater sensitivity to the knowledge base incompleteness, limited opportunities for training. For this reason connectionist systems are preferred, because they use mainly various ANN-structures and numerical methods. Hybrid systems make it possible to model the two primary cognitive aspects of human expertise – logical reasoning and intuition (associative processing of a priori knowledge), such as associative aspect, which is implemented through the ANN and the logical aspect, which is implemented through symbolic methods.

In general, the knowledge engineering techniques are based on the knowledge transfer from domain expert directly to systems, which requires data collection and filtering; variables and model structure selection; model identification and model validation.

3.1. Example 1

Basic work in KBS design is knowledge-base creation. This problem is illustrated with an example from biotechnology – growth hormones obtaining. Two methods of knowledge extraction are demonstrated: expert knowledge acquisition and automated extraction of new knowledge from experimental data and its evaluation.

3.1.1 Expert knowledge acquisition

In this part the tacit knowledge and rules-of-thumb acquisition will be demonstrated on the experimental research of a multi-factor biotechnological process.

Growth hormones gibberellins (GA) are a widely group of plant hormones, which help to regulate growth and development of plants. GA can affect many mechanisms of plant growing, flowering and fruit development. The well established and widely employed major commercial applications are in malting agriculture, horticulture and viticulture. It is of significant interest to study fermentation techniques for production of Gas by selecting morphological mutants of strain *Fusarium moniliforme*. The fungal cultures are able to give higher yields of Gas on a variety of liquid media. The morphological mutants of *Fusarium moniliforme* led to lower viscosity in fermentation broth resulted in increased production of Gas.

To examine a variety of nutritional and physical factors such as temperature can be useful for commercial fermentation process.

The mutant strain *Fusarium moniliforme* 3211 obtained by γ-ray irradiation was studied in the present investigation. The mutant strain was characterized by enhanced biosynthesis of GA_3 and GA_4+GA_7. *F. moniliforme* was cultivated in 3 l fermentor at pH 6.1. For studying the effect of temperature (T): 29.5^{o}C; 30.5^{o}C; 31.5^{o}C; 32.5^{o}C on the gibberellin acids production, fungus cultures were investigated for 10 days. Samples were taken every 24 hours of cultivation. All fermentations were carried out in triplicate and the data for GAs analysis were averaged. Experimental data are shown in Fig. 1.

The observed fermentations started at equal initial conditions excepting the cultivation temperature: inoculum Xo=0.02 [g/l], initial substrate concentration So=98.78[g/l], pH=6.1[-].

After gathering experience with the technology of studied fermentation process, the tacit knowledge is formulated by experts as rules for each case as shown in Table 1.

The most important expert's tacit knowledge is expressed in rule 5, which expresses optimal conditions for higher yield of GA_4+GA_7 and rule 7, which presents conditions for higher production of GA_3.

Formulated rules regarding the growth hormone production from *Fusarium moniliforme* 3211 were

many times proved in the practice, which serves for its validation.

As is known from literature, the key to acquiring tacit knowledge is experience. Tacit knowledge has been described as "know-how"; it involves skill in a way that can not be written down. On this account knowing-how or embodied knowledge is characteristic of the expert, who acts, makes judgments, and so forth without explicitly reflecting on the principles or rules involved. The expert works without having a work theory or expert just performs skilfully without deliberation or focused attention. The tacit aspects of knowledge are those that cannot be codified, but can only be transmitted via training or gained through personal experience.

Polanyi [39] introduced the useful distinction between tacit and explicit knowledge, explained that tacit knowledge is difficult to formalize and communicate, whereas explicit knowledge is transmittable in formal language. For example, researchers attempting to construct expert systems are constantly challenged to quantify how decisions are made, and embrace any tools or methods that convert tacit into explicit knowledge. Connected with the multi-factor processes, problems of expert knowledge obtaining, relational method for creating complex production rules, formalizing logical inference and rules reliability assessment are considered in details in [51].

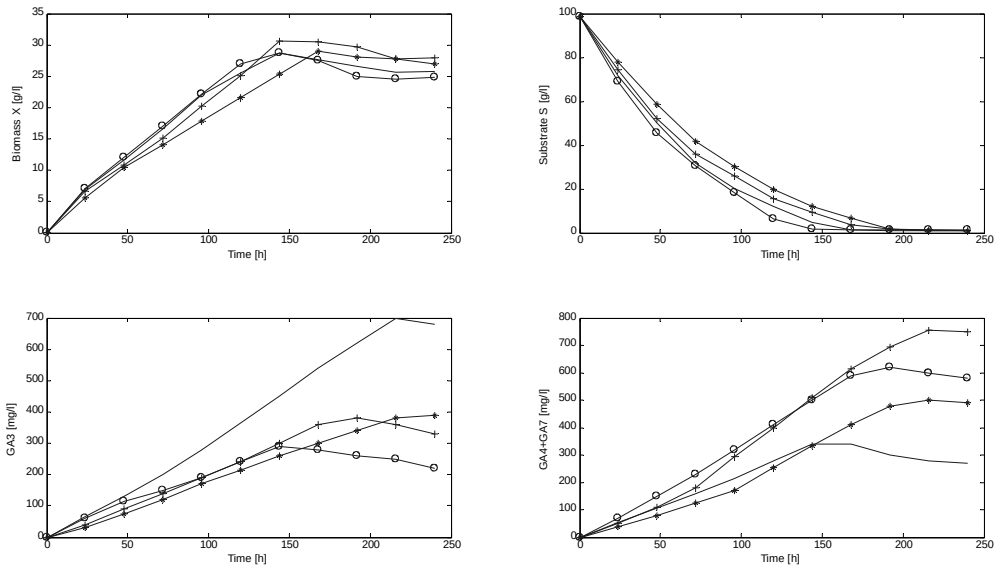

Fig.1. Experimental data for studying the effect of temperature (T): 29.5 ºC (denoted *-); 30.5 ºC (denoted +-); 31.5 ºC (denoted –); 32.5 ºC (denoted o-) on the biomass (X) growth, substrate (S) utilization and gibberellin acids (GA_3 and GA_4+GA_7) production.

3.1.2. Extraction of explicit knowledge from experimental data

Analysis of experimental data is a process of inspecting, cleaning, transforming, and modelling data with the goal of highlighting useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facts and approaches (statistics, artificial intelligence, machine learning, etc.) encompassing diverse techniques under a variety of names, in the different business, science, and social science domains. Modern trends in analysing experimental data – data mining method is a particular data analysis technique that focuses on modelling and knowledge discovery for predictive rather than purely descriptive purposes.

Related to the studied process for growth hormones obtaining, in this part are demonstrated results of the experimental data analysis by methods from statistics and artificial intelligence.

Table 1. Expert tacit knowledge

№	Temperature conditions of fermentation	Expert tacit rules
1	Fermentation at 29.5ºC	1. The final biomass concentration of *F. moniliforme* and final utilized substrate are close to these of fermentation 2. 2. Cultivation temperature is non optimal for gibberellines production. 3. The final biomass concentration of *F. moniliforme* is higher in comparison with the fermentations at lower or higher temperature (as well as fermentations 1, 3 and 4).
2	Fermentation at 30.5ºC	4. Final concentration of the utilised substrate is close to the fermentation 3. 5. Production GA_4+GA_7 is higher in comparison with fermentations at lower or higher temperature (as well as fermentations 1, 3 and 4), which shows that cultivation temperature and the used liquid media are optimal. 6. Substrate utilisation is better in comparison with the fermentations at lower or higher temperature (as well as fermentations 1, 2 and 4).
3	Fermentation at 31.5ºC	7. Yield of GA_3 is higher than in the fermentations at lower or higher temperature (as well as fermentations 1, 2 and 4), or this cultivation temperature and substrate content are optimal for GA_3 obtaining. 8. Yield of GA_4+GA_7 is lower in comparison with fermentations at lower or higher temperature (as well as fermentations 1, 2 and 4). 7. The fungal growth is less and substrate utilization is slower than in the other fermentations.
4	Fermentation at 32.5ºC	8. Production of GA_3 is lowest in comparison with the other processes. 9. Production of GA_4+GA_7 is higher than in fermentations 1 and 3, but lower than in fermentation 2.

The implementation of regression analysis in field of multifactor processes searches for a relationship between a variable of interest (dependant variable) and other variables (independent variables). Main goal of such relationship building is to forecast the dependant variable in the future based on past values of the dependant and independent variables. In Fig.2 are depicted examples of regression models of gibberellins dependence on biomass X formation and substrate S utilization with the corresponding residuals. These models are extracted from experimental data for purposes of best predictor of the final product selection.

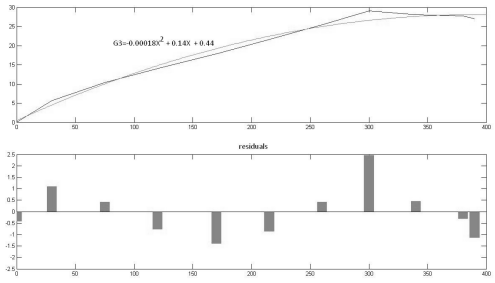

Fig.2a). Regression analysis with residuals of dependence between gibberelline GA_3 and biomass X at the cultivation temperature 29.5ºC

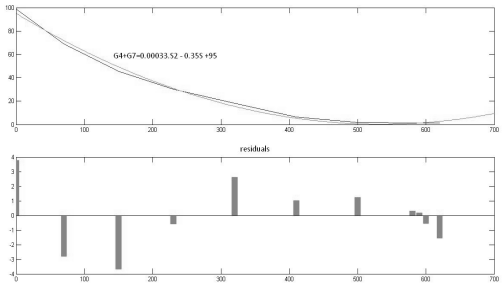

Fig.2b) Regression analysis with residuals of dependence between gibberelline GA_4+GA_7 and substrate S at the cultivation temperature 32.5ºC

A fuzzy regression model is used in evaluating the functional relationship between the dependent and independent variables in a fuzzy environment. Most fuzzy regression models are considered to be fuzzy outputs and parameters but non-fuzzy (crisp) inputs. In general, there are two approaches in the analysis of fuzzy regression models: linear-programming-based methods and fuzzy least-squares methods. In [34] are considered fuzzy linear regression models with fuzzy outputs, fuzzy parameters and also fuzzy inputs. A multi-objective programming method is formulated for the model estimation along with a linear-programming-based approach.

Fuzzy regression analysis is an efficient method capable of solving many modelling problems, connected with uncertain and incomplete data processing from one hand, and with nonlinear-system presentation when the relationship between input-output variables is not known from the other hand. In [56,60] a software system for fuzzy regression analysis is used for two ecological problems solution.

The first one is the development of fuzzy regression model for predicting bio-sorption of cuprum ion's depending on some input variables as well as microbial cells concentration, limiting substrate content, acidity of the cultural medium and other cultivation conditions. The problem is oriented to the waste water ecological treatment.

The second task solved is modelling of the influence of formaldehyde concentration on microbial population of strain *Candida didensii* 74-10. The mathematical investigations carried out concerning the negative influence of formaldehyde on the microbial growth and protein synthesizing ability. The problem is of great importance for minimization of wastes influence on the nature []
In the biotechnology and ecology are very important tasks of predicting key variables – final bioactive substance, ecological contamination, toxicity product, etc. In [58] it was found that both measured variables as well as biomass content X and carbohydrate source (utilizing substrate) S can be implemented with similar success for purposes of inferential measurements of final bio-products – gibberellines (Fig. 3)

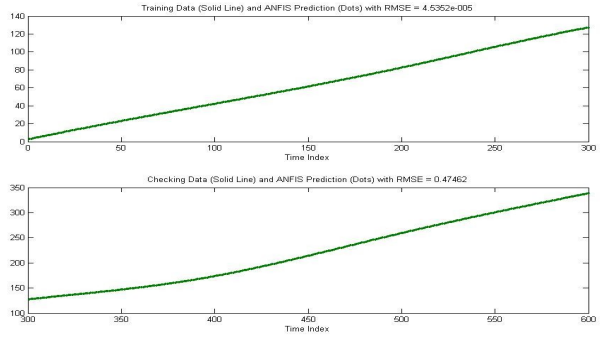

Fig. 3 Prediction of gibberellines GA_3 by using real measurements of carbohydrate source S

The software sensor design is carried out in two basic steps: selection of the best predictor by sensitivity analysis and after that – selection of best numerical data - predicting the output in real time.
For time series prediction a hybrid neuro-fuzzy approach was implemented by anfis [58].

Growth hormones biosynthesis as highly nonlinear process with many input and output variables requires many experimental investigations to obtain rules-of-thumb or to extract knowledge from the experimental data to create efficient knowledge-bases of software analyzers, useful in predictive control of industrial bioprocesses. Our results show that suitable empirical models, or data-driven models, producing reliable real-time estimates of process variables on the basis of their correlation with other relevant system variables can be useful tools in industrial applications, due to the complexity of the plant dynamics, which can prevent the first principles approach from being used.

The new developed GAs-predictive approaches, implemented in the inferential measurement systems of multi-factorial processes, show a high accuracy, simple structure and ask for low-cost equipment. They can be efficient when a lack of skilled workers and measurement equipment exist.

3.2. Example 2

As it was mentioned above, bacterial growth can be modeled as a sequence of four integrated phases: lag phase, exponential or log phase, stationary phase, and death phase. Mostly used mathematical

description represents the Monod kinetic model as a system of three differential equations, expressing relations between several input-output variables, as follows:

$$\frac{dX}{dt} = \mu X$$
$$\frac{dS}{dt} = -Y_{S/X} X$$
$$\frac{dP}{dt} = \eta X \quad (1)$$

where X is the biomass concentration [g/l], S is the substrate concentration [g/l], P is the final product concentration [g/l] and kinetic parameters are: YS/X - the yield of biomass depending on substrate utilization, η - the yield of product related to the biomass formation. The specific growth rate of biomass μ, according Monod description – a global model, is:

$$\mu = \frac{\mu_{max} S}{k_S + S}, \quad (2)$$

where μ_{max} is the maximal specific growth rate [1/h], k_S is Michaelis-Menten constant [g/l].

A set of experimental data for a batch process of glucoamylase production as well as biomass X, substrate concentration S, and product P is applied in order to investigate the system dynamics in the presence of the designed estimator. There are two main approaches for on-line state and parameter estimation of biotechnological processes – exponential estimators design (based on Kalman filtering method) and asymptotic estimators design (based on the linear algebra results) [27].
This example describes a modelling technique of biotechnological process based on the Extended Kalman Filter (EKF) – single-phased and multi-phased.

As a next stage of the problem of estimation model parameters in the present work are defined a separate phases of the studying nonlinear process and is applied a Kalman filter for each of them. The purpose is to compare an accuracy of single-phase modelling and multi-model process presentation.

The transition between phases is based on previously defined rules, one of them connected with the fixed period of time, which is preliminary determined by experts, and the other – with the change of the first derivative of specific growth rate, which in lag-phase is zero and after that at the beginning of the phase of the exponential growth is increasing. Usually for the studied process are most important these both phases.

The EKF algorithm, a frequently used recursive estimator [1], is a two-step algorithm using a time-update and a measurement-update. Given the current estimate, $\hat{x}_k$, the time update predicts the state value at the next sample, $\bar{x}_{k+1}$. Here is executed a one-step-ahead prediction using Runge-Kutta integration. The prediction of the corresponding error covariance, $\bar{P}_{k+1}$ is calculated using the linearised model. The next step in the EKF algorithm, the measurements-update, consists of approximation of the error covariance, $\hat{P}_{k+1}$ and the correction of the state based on the new measurement, y_{k+1}. There is used a correction term, which is a function of the innovation, that is, the discrepancy between the measured and predicted values.

In contrast to the traditional techniques, the Kalman filter algorithm takes the measurement error characteristics into account. The EKF can be implemented online since it is based on the measurements up to and including the moment of estimation.

The results are presented in four figures Fig.4-Fig.7 and Table2.

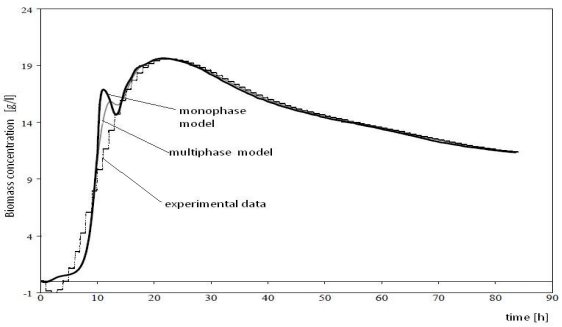

Fig.4 Biomass concentration – a comparison between real data, mono-phase and multiphase model

The results show that the estimated values (with using a single and multiphase model) of biomass and product concentrations tend to experimental data with a satisfactory accuracy (Fig.4 and 5). Modelling with multi-model approach observed a high accuracy in the time interval after the 15-th hour and a higher accuracy in the transition between two phases – from lag to exponential, which is due to the chosen model structure. The results are sufficiently good and satisfy the

requirements for model accuracy (Fig.6 and 7) and gives less residual error (Table 2).

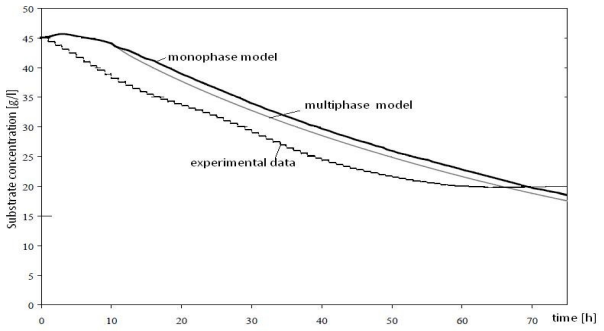

Fig. 5 Substrate concentration – a comparison between real data, mono-phase and multiphase model

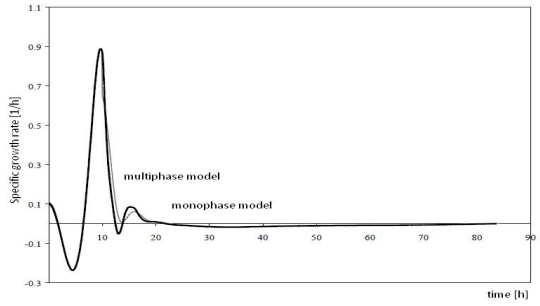

Fig. 6 Specific growth rate [1/h] - a comparison between mono-phase and multiphase model

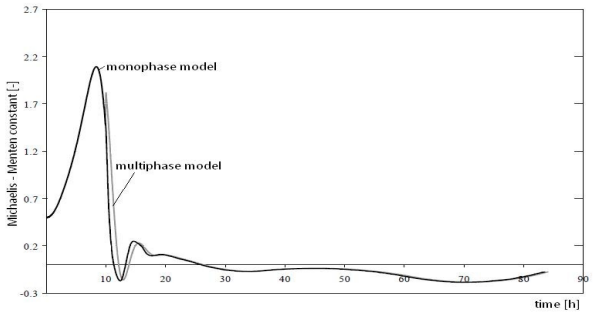

Fig.7 Mihaelis-Menten constant – a comparison between mono-phase and multiphase model

Our expectations are, that at more detailed analysis of model structure, taking into account more growth factors, for example pH, which have influence on biomass synthesis, will be obtained a better accuracy of assessment. A higher accuracy of the obtained model would be achieved and with dividing a process to more than two phases, as well as if are used other modelling techniques – fuzzy logic, NN, under which is expected to achieve more smoothly transition between the process phases.

Table 2. Accuracy of compared models

Process phases	Average residuals (SSR criterion)	
	Mono-phase model	Multi-phase model
Lag-phase	0.202819	0.142642
Exponential phase	0.010124	0.007471
Global model	0.04344	0.030604

The discussion shows opportunity to implement these different models in a multi-model system for multiphase process monitoring and control.

4 Conclusion

The concept of intelligent industry lies on the network of interacting intelligent systems for data acquisition, assessment, interpretation, decision support and control, to provide efficiency savings and to assure consistent quality of product – safety products and environment.

Microbiological industry owns a wide range of same technological processes, used in different plants in the whole world, which could be contacted through the knowledge bases in specialized technological centres, offering immediate consulting to improve the technology and to guarantee a production of high quality.

Creation of knowledge-based systems for monitoring and control of ecology and industrial processes is a powerful approach, which expands its own efficiency by implementing the recent communications and multi-agent technologies, for inferential control of the distributed plants.

References:

[1] Acha, V., Meurens, M., Naveau, H., Dochain, D., Agathos, S.N. Detoxification of a mixture of aliphatic chlorinated hydrocarbons in a fixed-bed bioreactor: continuous on-line monitoring via an ATR-FTIR sensor and model-based estimation of state variables. *Proc. 5th Latin-American Workshop-Seminar Wastewater Anaerobic Treatment*, Vina del Mar, Chile, Vol.1, 1998.

[2] Angelova M., Pedro Melo-Pinto, Pencheva T. Modified Simple Genetic Algorithms

Improving Convergence Time. *WSEAS Trans. on Systems, Special issue "Modeling and Control of the Integrated Bio-systems"*, 11(7), 2012, pp.256-267.

[3] Apweiler, R., A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, Vol. 32, No Database Issue, 2004, pp. D115-D119.

[4] Bairoch, A., R. Apweiler, C. H. Wu, W. C. Barker, B. Boeckmann, et al. The universal protein resource (UniProt). *Nucleic Acids Research*, Vol.33, No. Database Issue, 2005, pp. D154-D159.

[5] Berman H., K. Henrick, and H. Nakamura, Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol*., Vol. 10, No. 12, pp. 980, 2003.

[6] Berman H., K. Henrick, H. Nakamura, and J. L. Markley, The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res*., Vol.35, No. Database issue, 2007, pp. D301-D303.

[7] Biryukov, V.V. Non-traditional tasks for control of microorganisms' cultivation processes by using computers. *Theory and practice of continuous micro-organisms' cultivation*, Moscow, Nauka (in Russian), 1980.

[8] Biryukov, V.V., Kantere, V.M. *Optimisation of microbial synthesis batch processes*, Moscow, Nauka, 1985.

[9] Blakebrough, N. and Moresi, M., Modelling of the process yields of whey fermentation, *Eur.J. Appl.Mikrobiol. Biotechnol*., Vol.13, No1, 1981, pp.1-9.

[10] Collins T. and J.Bandy. CO_2 generation and Harvesting. *MBAA Technical Quarterly*. Vol. 37, No2, pp.255-260.

[11] Dimitrova N., Krastanov M. On the Asymptotic Stabilization of an Anaerobic Digestion Model with Unknown Kinetics. *WSEAS Trans. on Systems, Special issue "Modeling and Control of the Integrated Bio-systems"*, 11(7), 2012, pp.244-255.

[12] Dochain, D. Design of on-line estimation and adaptive control algorithms for bioreactors. *Proc. 1st Int. Conf. Modelling and Control of Biotechnological, Ecological and Biomedical Systems*, Varna, Bulgaria, No2, 1990.

[13] Dochain, D. Software Sensors and Adaptive Control in Biotechnology. *Proc. ACoFoP III*, Paris, 1994.

[14] Dochain, D., Bastin, G., Pauss, A., Nyns, E..J. and H. Naveau. Software sensors: on-line estimation of the specific growth rate in microbial reactors. *Proc. Journée de la biotechnologie*, Belgian Society of Industrial Chemistry, 1985.

[15] Dutta, S., K. Burkhardt, J. Young, G. J. Swaminathan, T. Matsuura, et al. Data deposition and annotation at the worldwide protein data bank. *Molecular Biotechnology*, Vol. 42, No. 1, 2009, pp. 1-13.

[16] Einstein, G. and McDaniel, M. *Distinctiveness and the Mnemonic Benefits of Bizarre Imagery. Imagery and Related Mnemonic Processes: Theories, Individual Differences, and Applications*. New York: Springer-Verlag, 1987, pp.79-102.

[17] Eyben D. Brewery waste water and solid waste treatment. *Monograph XIX E.B.C. Symposium Waste reduction in brewery operations*, Rheinfelden, Switzerland, 1992, pp.121-134.

[18] Flynn D, Ritchie J, Cregan M. Data mining techniques applied to power plant performance monitoring. *Proc.16th IFAC World Congress*, Prague, Czech, 2005.

[19] Gaur S. and T Reed. *Thermal Data for Natural and Synthetic Fuels*. Marcel Dekker Inc ,1998

[20] Grancharova, A., Zaprianov, J. and Vassileva, S. Dynamic Modelling of Bioprocess with Artificial Neural Network. *Proc.Int. Sc. Conf. Computer Science*, Ostrava, Czech, 1995.

[21] Heijden, van der, Reinier, T.J.M., Hellinga, C., Luyben, K. and Honderd, G. State estimators (observers) for the on-line estimation of non-measurable process variables. *Trends in Biotechnology*, Vol.7, No 8, 1989, pp.205.

[22] Hinshelwood, S.N., *The chemical kinetics of the bacterial cell*, Oxford University Press, London, 1946.

[23] Hock, H.S., Romanski, L., Galie, A., Williams, C.S. Real-World Schemata and Scene Recognition in Adults and Children. *Memory and Cognition*, Vol.6, 1987, pp.423-431.

[24] http://www.eere.energy.gov

[25] http://www.geabrewery.com/].

[26] Ianis M., Tzekova K., Vasileva S.. Copper biosorption by Penicillium cyclopium: equilibrium and modellinig study. *Biotechnology and Biotechnol.Equipment*, №20/2006/1, 2006, pp.195-201.

[27] Jang, J.-S. R. Fuzzy Modeling Using Generalized Neural Networks and Kalman Filter Algorithm. *Proc. 9-th National Conf. on Artificial Intelligence* (AAAI'91), 1991.

[28] Karim, M. and Rivera, S. Comparison of feed-forward and recurrent neural networks for bioprocess state estimation, *Proc. ESCAPE-1*, Elsinore, Denmark, 1992.

[29] Kleinstreuer, C. and Poweigha, T., Modeling and Simulation of Bioreactor Process Dynamics. *Advances in Biochemical Engineering and Biotechnology*, Spinger-Verlag, Berlin, Heidelberg, N.Y., Tokio, Vol.30, 1984, pp. 91-146.

[30] Latrille, E., Teissier, P., Perret, B., Barille, J.M. and Corrieu, G. Neural Network Modelling and Predictive Control of Yeast Starter Production in Champagne. *CD-Proc. ECC'97*, 1997.

[31] Lee J.M., Yoo C., Lee I.B. Statistical process monitoring with independent component analysis. *Journal of Process Control*, Vol. 14, Issue 5, 2004, pp.467-485.

[32] Luo, R.F. Fuzzy-neural-net-based inferential control for a high-purity distillation column. *Control Engineering Practice*, Vol.3, No1, 1995, pp.31-40.

[33] Markley, J. L. E. L. Ulrich, H. M. Berman, K. Henrick, H. Nakamura, et al. BioMagResBank (BMRB) as a partner in the worldwide protein data bank (wwPDB): new policies affecting biomolecular NMR depositions. *Journal of Biomolecular NMR*, Vol. 40, No. 3, 2008, pp. 153-155.

[34] Masatoshi Sakawa, Hitoshi Yano. Multi-objective fuzzy linear regression analysis for fuzzy input-output data. *Fuzzy sets and Systems*, Vol.47, No2, 1992, pp.173-181.

[35] Mileva S., Vassileva S. ANN-based Prediction of Antioxidant Characterizations during the Brewery Fermentation. *Proc. International Scientific Conference Computer Science'2008*, Heron Press, 2008, pp.164-169.

[36] Montellano, R., Bernier, M.P., Cheruy, A. and Farza, M. Knowledge-based System in modeling and control for biotechnological processes. *Proc. 11th Triennial World Congress IFAC*, Pregamon Press Inc., Elmsford, NY, USA, Vol.4, 1991.

[37] Nagai EY, Arruda LVR. Soft sensor based on fuzzy model identification. *Proc.16th IFAC World Congress*, Prague, 2005.

[38] Patnaik, P.R. Artificial intelligence as a tool for automatic state estimation and control of bioreactors. *Laboratory Robotics and Automation*, Vol.9, No6, 1997, pp.297-304.

[39] Polanyi M., *The Tacit Dimension*, Routedge and Kegan Paul, 1966.

[40] Rabinovich, S.G. Efficient calculation for indirect measurements and a new approach to the theory of indirect measurements. *Proc. Measurement Science Conference*, Newport Beach, Anaheim, CA, USA, 1996.

[41] Robenack K., Nicolas D. Observer Based Current Estimation for Coupled Neurons *WSEAS Trans. on Systems, Special issue "Modeling and Control of the Integrated Bio-systems"*, 11(7), 2012, pp.268-281.

[42] Rouzic, Y.Le, Rakotopara, D., Calvet, J.P. and Pgauthier, J. Soft Sensor for adaptive pH control, an industrial application, *Proc. European Control Conf ECC'97*, 1997.

[43] Russel, Stuart J., Norvig, Peter. *Artificial Intelligence: A Modern approach* (2nd ed.), Upper saddle River, New Jersey: Prentice Hall, 2003.

[44] Samad T. *Perspectives in Control Engineering: Technologies, Applications and New Directions*. New York: IEEE Press, 2001.

[45] San, K-Y. Expert System Based Intelligent Control Scheme for Space Bioreactors. *NASA Technical Memorandum*, Vol.1, 1988.

[46] Slavov T., Roeva O. Genetic Algorithm Tuning of PID Controller for Glucose Concentration Control using Software Sensor. *WSEAS Trans. on Systems, Special issue "Modeling and*

Control of the Integrated Bio-systems", 11(7), 2012, pp.223-233.

[47] Thibault, J, Breusegem, . V. van and Cheruy, A. On-line prediction of fermentation variables using neural nets, *Biotechnology and Bioengineering*, Vol.36, No10, 1990, pp.1041-1048.

[48] Thomas, J.P. and Wei, R.P. Standard error estimates for rates of change from indirect measurements. *Technometrics*, Vol.38, No.1, 1996, pp.59-68.

[49] Tzvetkova B., Vassileva S. Inhibitory effect of furfural on the bioproductivity of *Candida blankii* 35. *Proc. BIOPS'2003*, 2003.

[50] Ulrich, E. L. H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, et al. BioMagResBank. *Nucleic Acids Res*., Vol. 36, No. Database issue, 2008, pp. D402-D408.

[51] Vachova, B. *Derivation and reliability evaluation of knowledge for multi-factor technological processes*. PhD Thesis, Bulgarian Academy of Sciences, 2009, http://hsi.iccs.bas.bg /Staff/B.Vatchova/links /PhD_BV.pdf

[52] Van Deventer J.S.J., Kam K.M., van der Walt T.J., Dynamic modelling of a carbon-in-leach process with the regression network. *Chemical Engineering Science*, Vol. 59, 2004, pp.4575–4589.

[53] Vassileva S., F.Neri. An Introduction to the Special Issue: Modeling and Control of Integrated Bio-Systems. *WSEAS Transactions on Systems*, 7(11), 2012, pp. 221- 222.

[54] Vassileva S., Georgiev G., Mileva S. Knowledge-Based Control Systems via Internet Part I. Applications in Biotechnology *Bioautomation*, Vol.2, 2005, pp. 37-48.

[55] Vassileva S., Mileva S. Intelligent Software Analyzer Design for Biotechnological Processes. *Advanced topics on Neural Networks. Artificial Intelligence Series*, Publ. WSEAS Press, Hon.Ed. L.A.Zadeh and J.Kacprzyk, Ed.D.Dimitrov, S.Jordanova, V.Mladenov and N.Mastorakis, 2008, pp.48-53.

[56] Vassileva S., Trifonov T., Nikolov K. Ecological Variable Prediction By Fuzzy Regression. *In CD. Proc. Ecology, Scientific Articles of the Int. Conf. Ecology'2005*, Sunny Beach, Bulgaria, vol. III, part 2, 2005, pp.203-215.]

[57] Vassileva S., Tzvetkova B. ANN-based software analyzer design and implementation in processes of microbial ecology. *Advanced topics on Neural Networks. Artificial Intelligence Series* Publ. WSEAS Press, 2008, pp.142-147.

[58] Vassileva S.. Advanced Fuzzy Modeling of Integrated Bio-systems. *WSEAS Transactions on Systems*, 7(11), 2012, pp. 234-243.

[59] Vassileva, S. I., Mileva S.D., Andreeva P. Ts. Sensitivity Analysis for Environmental Sensory Models and Monitoring Networks. *Journal Ecology & Safety*, Vol 2, Part1, 2008, pp. 507-518.

[60] Vassileva, S., Trifonov, T., Nikolov, K. Fuzzy Regression Algorithm for the Multi-Modeling Approach. Advanced Studies on Contemporary Mathematics/ *Proceedings of the Jangjeon Mathematical Society (ASCM/PJMS)*, Seoul, S. Corea, Vol.9, № 1, 2006, pp. 83-89.

[61] Vassileva,S., B. Tuleva., N. Christova, P.Petrov. NN-based modeling of biodegradation of naphthalene by Bacillus subtilis 22BN. *Ecological Engineering and Environmental Protection*, No4, 2004, pp.48-56.

[62] Velankar, S., Y. Alhroub, A. Alili, C. Best, H. C. Boutselakis, S. Caboche, et al. PDBe: protein data bank in Europe. *Nucleic Acids Res*., Vol. 39, No. Database Issue, 2011, pp. D402-D410.

[63] Yu, J.J. and Zhou, C.-H. Soft-Sensing Techniques in process Control. Chinese *J. of Control Theory and Applications*, No13, 1996, pp.137.

[64] Zahedi G., Elkamel A., Lohi A., Jahanmiri A., Rahimpor M.R. Hybrid artificial neural network – First principle model formulation for the unsteady state simulation and analysis of a packed bed reactor for CO_2 hydrogenation to methanol. *Chemical Engineering Journal*, Vol.115, 2005, pp.113–120.