ELSEVIER

Ecological Modelling 75/76 (1994) 629–640

ECOLOGICAL MODELLING

Feature-based model identification of nonlinear biotechnological processes

L. Vermeersch [a,*], B. Kroes [b], P. Vanrolleghem [a]

[a] *Institute for Applied Mathematics and Biometrics, University of Ghent, Coupure Links 653, B-9000 Ghent, Belgium*

[b] *Delft University of Technology, Delft, Netherlands*

Abstract

Modelling ill-defined systems requires powerful tools to attain a quantitatively description of the studied systems. In this paper, a modelling concept is presented that tackles the problems inherent to processes for which the necessary a priori knowledge for deductive analysis is lacking. The approach can be summarised as follows. By means of relationship detectors, such as SAPS or fractional factorials, the existence of a causal structure can be deduced qualitatively. In a next step of the modelling task, the goal is to find the quantitative description of this relationship. This step can be subdivided in two phases, i.e. model structure characterisation and finally parameter estimation. In this paper new techniques are proposed (and validated on real-life experimental results) to achieve the latter steps. In order to separate the structure characterisation from the parameter estimation task, an approach is taken in which parameter-invariant features are extracted from the data. The properties of these features are chosen in such a way that a classifier can select a specific model description. The decomposition in Zernike features and the recurrent neural network classifier, introduced by Sudharsanan, constitute the implementation of this concept. To check the feasibility of this modelling approach, a biotechnological application is chosen as a test case. Due to the changing nature of the wastewater treatment process, reflected in a set of mathematical descriptions applicable at different time instances, the aforementioned methodology will give the possibility to develop more efficient adaptive controllers.

Key words: Biotechnological process

* Corresponding author.

SSDI 0304-3800(94)00025-D

1. Introduction

To compose the model of a real-life process, two types of information could be utilised (Karplus, 1976):

1. theoretically based knowledge and insight of the system,
2. experimentally obtained input/output data of the system.

The former kind leads to a deductive reasoning approach, in which specific knowledge is deduced from more general principles. However, when a plenitude of background knowledge is lacking, one has to rely on empiric observations, which force an inductive reasoning approach. In any case, the more a priori information en surplus to the input/output data is deployed, the more the chances are improved to find an adequate model for the system. The joint use of a priori knowledge and empiric observations leads to a semi-inductive approach.

The goal of this work is to develop a more general technique which is a realisation of this semi-inductive reasoning approach and to test it on non-linear ill-defined systems. In particular biotechnological processes require such a technique due to their inherent non-linear nature and the deficiency of a priori knowledge.

The main characteristic of the modelling approach, presented in the sequel, is the separation of the model characterisation phase and the parameter estimation phase. This is in contrast with the more traditional structure characterisation methods (e.g. AIC by Akaike, GIC, etc.) where structure characterisation and parameter estimation are intermingled. The reason for chosing an "a priori" modelling technique (where first the model structure is characterised and in a second phase the parameters are estimated) instead of an "a posteriori" technique (where the model structure characterisation and the parameter estimation is performed in one step) will be explained in greater detail in Section 6.1.

2. Relationship detection

Empirical data can be utilised to build an empirical model that will explain the behaviour of the dynamical variables over limited ranges (Box et al., 1978). Before achieving this, some important questions have to be answered during the data acquisition phase of the identification problem:

- which variables are the most important,
- over what ranges should the variables be studied,
- what metrics should be used for the variables,
- which multivariate transformations should be deployed to preprocess the data into a desirable form.

So, what class of data acquisition design should be taken? Different techniques have been proposed to give an answer to some or all of these questions.

A two-level factorial design, which has proven to be of great practical importance, could be taken. The main advantages of this approach are:

- only a few runs per factor studied are needed,

- it can give an indication of major trends and so determine promising directions for further research,
- 2^k-factorial designs are the basis for 2^k-fractional factorial designs.

Another approach towards relationship detection is taken in SAPS. It is an implementation of inductive reasoning, a technique which was invented in the 1970s by Klir (1985). The first implementation of SAPS (Uyttenhove, 1979) was not sufficiently flexible. The second edition was implemented by F. Cellier and his team (Cellier, 1991), it is called SAPS-II and is available as a CTRL-C library or MATLAB toolbox. Currently, an ameliorated version is developed by D. Van Welden (University of Ghent). SAPS-II, just as factorial designs, tries to give answers to the previous questions.

3. Feature extractor

After ascertaining the qualitative result that there is a relationship between some variables (i.e., input and output variables) by means of the relationship detectors, the purpose is to determine the kind of relationship. This second step in the modelling phase is to find a more quantitative description of the ongoing process, e.g. to enable a better control of the system.

An ideal situation would be one where the only input to a computer is the data measured from the phenomenon to be modelled, and from this data, the machine would produce a complete model. This is, however, utopian. A solution (Vansteenkiste et al., 1979) has been proposed where one more input is required: an input considered as a library of candidate models. The machine will search among the several candidate models in order to find a model whose structure is best adapted to the data. The situation that no reasonably good model can be retrieved from the library is also imaginable.

From a pragmatic point of view, a given structure results in a set of properties shared by the data generated by that structure. For a discriminating study it is sufficient to choose for each structure a property, which is sufficiently parameter invariant, and which is not shared, to the same extent, with the other candidate models. Thus, the feature of the model corresponding to the real phenomenon should be invariant to the real-world data and the features of the wrong models should vary as strongly as possible.

The main problem with parameter-invariant features corresponding to models of dynamical systems is that even in fairly simple testing examples, one is confronted with implicit or even transcendental functions when extracting analytically a real parameter-invariant feature from the candidate models. Moreover, it is a very problem-sensitive approach, which can hardly be automated.

Therefore, this parameter-invariant requirement is dropped from a practical viewpoint. The consequences for the identification of systems, based on a library of candidate models, are:

- a simplified feature extractor,
- a more selective but more complicated classifier.

The manner in which the more complex classifier will be built is explained in greater detail in Section 5. The simplified feature extractor will actually be a well-known data compression technique, as presented in Section 4. The next section deals with a new algorithm to find appropriate features.

4. Zernike features

The recognition of signals generated by real-life processes can be seen as a special case of the 2D-image recognition problem. In the field of image recognition it is also very important to recognise objects automatically and this independently with respect to the actual location, magnitude and orientation of these objects, cf. the recognition of objects in space, etc. Moments and functions of moments were used as feature extractors, containing enough information to reconstruct the image. The regular moments are by far the most popular type of moments and are defined as

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x, y)\, \mathrm{d}x\, \mathrm{d}y, \tag{1}$$

where m_{pq} is the $(p+q)$th order moment of the continuous image function $f(x, y)$. For digital images the integrals are replaced by summations and m_{pq} becomes

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y). \tag{2}$$

The definition of regular moments can be seen as the projection of the function $f(x, y)$ onto the monomial $x^p y^q$. Unfortunately, the basis constructed by these monomials is not orthogonal. Thus, the reconstruction of the image is more difficult and computationally intensive. Teague (1988) has suggested to use the theory of orthogonal polynomials as the basis for orthogonal moments. Zernike moments are a type of orthogonal moments. The Zernike features — i.e. the magnitude of the complex Zernike moments — possess the additional property of rotation invariancy. This means that the values of the Zernike features will not change when the image is rotated. Since the Zernike features are only rotation invariant, a preprocessing of the image, deploying regular moments, is necessary to obtain scale and translation invariancy.

In 1934, Zernike introduced a set of complex polynomials which form a complete orthogonal set over the interior of the unit circle, i.e. $x^2 + y^2 = 1$. Let the set of these polynomials be denoted by $\{V_{nm}(x, y)\}$. The form of these polynomials is

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm\theta), \tag{3}$$

where n is a positive integer or zero; m are positive and negative integers subject to constraints $n - |m|$ even and $|m| \leqslant n$; ρ is the length of vector from origin to

(x, y) pixel; θ is the angle between vector ρ and x axis in counterclockwise direction; $R_{nm}(\rho)$ is radial polynomial defined as

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \cdot \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!} \cdot \rho^{n-2s}. \quad (4)$$

Note that $R_{nm}(\rho) = R_{n,-m}(\rho)$.

These polynomials are orthogonal and satisfy

$$\iint_{x^2+y^2\leq 1} [V_{nm}(x, y)]^* V_{pq}(x, y)\, dx\, dy = \frac{\pi}{n+1}\delta_{np}\delta_{mq}.$$

Zernike moments are the projection of the image function onto these orthogonal basis functions. The Zernike moment A_{nm} of order n with repetition m for a continuous image function $f(x, y)$ that vanishes outside the unit circle is

$$A_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2\leq 1} f(x, y) V_{nm}^*(\rho, \theta)\, dx\, dy. \quad (5)$$

For a digital image, the integrals are replaced by summations to get

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V_{nm}^*(\rho, \theta), \quad x^2+y^2 \leq 1. \quad (6)$$

To compute the Zernike moments of a given image, the center of the image is taken as the origin and pixel coordinates are mapped to the range of the unit circle, i.e. $x^2+y^2 \leq 1$. Those pixels falling outside the unit circle are not used in the computation. Also note that $A_{nm}^* = A_{n,-m}$. As previously mentioned, the Zernike features are the magnitudes of the Zernike moments, and are calculated as

$$A_{nm}^f = |A_{nm}| = \sqrt{A_{nm} A_{nm}^*}. \quad (7)$$

It has to be noted that the Zernike features for their efficient data compression and reconstruction capabilities, the rotation invariancy characteristic and the applicability for more general two-dimensional figures are promising coding scheme for both signals and images. The data compression characteristic is very useful later on to keep the size of the classifying neural network within reasonable boundaries.

5. Neural classification

The final step of the model identification procedure, which has gained in importance due to the simplified feature extractor, is to determine the class the unknown process belongs to. To accomplish this, neural networks can be used. Artificial neural networks are biology-based replicas of the human brain, which give interesting perspectives on the research of human behaviour. On the other

hand, they are far from exact — biologically correct — replicas of the brain structure and dynamics. The main reasons are the lack of precise knowledge about the internals of the brain and the currently known complexity of cell, axon and synapse interactions, which is hard to simulate in finite time on currently available computers. Although artificial neural networks are much simpler compared to their biological counterparts, the theoretical research on them can be more founded. For some artificial neural networks for example, it is possible to prove the convergence of the learning algorithm, to give a lower limit of the learning speed, to propose synthesis methods instead of learning methods, etc.

The specific class of dynamical neural networks — alternatively called recurrent neural networks or Hopfield neural networks — that will be used as model classifier is described by

$$\dot{\vec{u}} = -A \cdot \vec{u} + W \cdot \vec{g}(\vec{u}) + B \cdot \vec{z}, \tag{8}$$

$$\vec{y} = H^T \cdot \vec{u}\,^*.$$

The different components are of the following type: $A = I \in \mathbf{R}^{n \times n}$, $\vec{u} \in \mathbf{R}^n$, $W \in \mathbf{R}^{n \times n}$, $B \in \mathbf{R}^{n \times m}$, $H \in \mathbf{R}^{n \times p}$, $\vec{y} \in \mathbf{R}^p$, $\vec{z} \in \mathbf{R}^m$ and $\vec{g}(\cdot): \mathbf{R}^n \rightarrow \mathbf{R}^n$. Elements of $\vec{g}$ are bounded functions (usually sigmoid) of the states $\vec{u}$. Model 8 has a discrete-time counterpart as well. The discrete-time neural model has been well discussed in the literature after Hopfield (1982) and can be described by

$$v_i(k+1) = sgn\left(w_{ij} v_j(k) + b_i\right), \quad i = 1,2,\ldots,n, \tag{9}$$

where w_{ij} represents the (i, j)th element of the matrix W, v_i is the state of the ith neuron, b_i is an external input element and the $sgn(\cdot)$ function takes the values of one multiplied by the sign of its argument. Note that Eq. 9 is a synchronous update model and various asynchronous update models are also available. The stability of the model is shown by an energy function argument which requires that matrix W be symmetric and its diagonal are all zero. Another important aspect is the number of patterns that can be stored in the model 9. Hopfield showed that it should not be more than $0.15 \cdot n$ for efficient information retrieval. Additional extensions and comparisons to other techniques of the model described by Eq. 9 are numerous, e.g., the retrieval efficiency will improve if the memory vectors are orthogonal. Hopfield showed, also based on energy function considerations, that the stability of model 8 required symmetry conditions on the interconnection weight matrix W. [1] Hopfield and Tank (1985) used this neural model as a useful optimisation network by demonstrating the solution of a complex optimisation problem, namely the travelling salesman problem. This was carried out by relating the objective function to be minimised to the energy function of the neural network. Sudharsanan (1990) introduced a synthesis algorithm for the neural model 8 to solve quadratic minimisation problems. The main advantage of this neural approach is the realizability of this kind of neural networks by analogue active RC networks or even by

[1] The matrix A only has to be diagonal and not necessarily equal to I.

optical circuits. These implementations are intrinsicly highly parallel, which leads to important speedup of the minimisation calculations.

To implement the classifier, the neural net model 8 (see Fig. 1) introduced by Sudharsanan (1990) is applied. Let $\vec{z}=[z_1, z_2, \ldots, z_m]^T$ denote the input vector and let $\vec{y}$ denote the corresponding neural net output. These form the input and output layers respectively. The wanted output is represented by $\vec{y}^d$. So, the objective is to learn the dynamical neural network, described by:

$$\dot{\vec{u}} = -\vec{u} = +W \cdot \vec{g}(\vec{u}) + B \cdot \vec{z}, \tag{10}$$

$$\vec{y} = H^T \cdot \vec{u}^*, \tag{11}$$

to classify an unknown — noise-corrupted — signal. $\vec{u}^*$ in Eq. 11 is a stable equilibrium of Eq. 10 for the input pattern $\vec{z}$ presented.

It is clear from Fig. 1 and the above description that the underlying network has a three-layer architecture. The input layer comprises noninterconnected identity

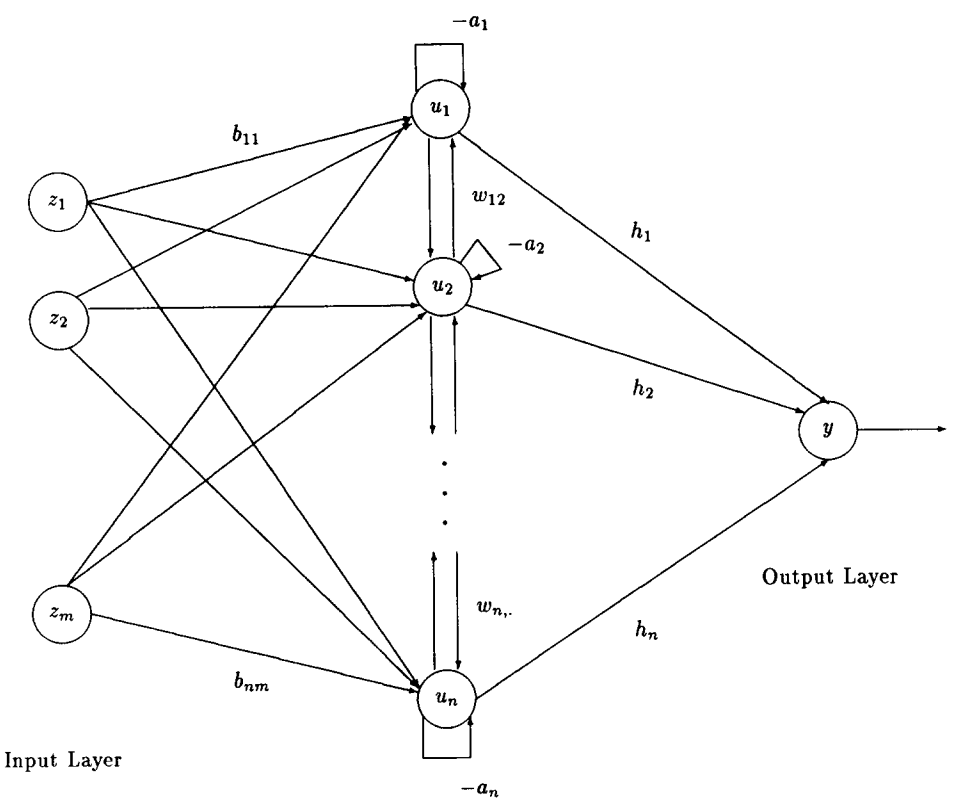

Fig. 1. Proposed process classifying network.

units, i.e., the activation output is identical to its input. Linear combination of these outputs are fed to the hidden layer, which is a true dynamical neural network with sigmoid processing units. The output units simply calculate a linear combination of the outputs of the hidden layer units. Sudharsanan (1990) proves that the standard three-layer feedforward network is a special case of this type of three-layer networks. Sudharsanan further claims that better approximation capabilities, reduced number of hidden nodes, are the advantage of this dynamic neural network.

6. Application in wastewater treatment

In this section, the proposed model classifier is used to determine which model is the most appropriate to describe the biological phenomena observed in an activated sludge wastewater treatment process. Although extensive research in the last few decades has provided us with relatively good insight in the process, control mainly relies on manual adjustment of the few control variables available. The main reason for this is the lack of appropriate instrumentation for on-line measurement of biologically meaningful parameters.

Recently, a new sensor was developed at the Lab of Microbial Ecology that aims at providing such data. It is a biosensor in the sense that it contains biological components that generate a signal for further analysis. The path from the raw biological signal to the final sensor output takes advantage of the classifier and this is illustrated below.

6.1. Principles

A detailed description of the biosensor can be found elsewhere (Vanrolleghem et al., 1990); the principles involved, however, are essential for a good understanding of the work. The central part of the sensor device is a small bioreactor in which activated sludge is maintained. Activated sludge is the name given to the consortium of organisms that degrade the waste contained in wastewater. In the sequel we will use "substrate" to describe the waste, because what we define as waste is essential feed for the growth of the activated sludge. The aim of the sensor is to provide a control system with information on the interaction between wastewater and activated sludge. In view of this, it is essential that the sensor's reactor contains activated sludge taken from the plant to control so that the biosensor's outputs are directly related to the system involved.

The degradation in activated sludge systems is an oxidative process. Therefore, the oxygen consumption (microbial respiration) is directly related to the amount of waste (substrate) that is eliminated from the water. In the respirographic biosensor, the oxygen uptake is recorded using an oxygen mass balance over the bioreactor. More information on how the oxygen uptake data are obtained can be found in Henze et al. (1987).

Essentially, the sensor interprets the response of the activated sludge biosystem on a pulse substrate (waste) injection in the bioreactor. Model-based interpretation

of this "impulse response" allows system identification. It has been shown elsewhere (Van Impe et al., 1991) that the models used to describe the biosystem in the monitoring device can be used to build a powerful control system. The direct connection between sensor and control system that is established in this way allows adaptivity of the controller.

However, the models that are required to adequately describe the bioprocesses are highly non-linear. As a result, parameter estimation is a computing-intensive task. To complicate the task of extracting the necessary information from the oxygen uptake data even more, the wastewater composition can change drastically and this is reflected not only in a change in parameters but also in a change of model structure. A number of candidate model structures are available and extensive experimental results have indicated their ability to describe most of the data obtained so far.

An a posteriori approach is possible in which all candidate model structures are evaluated on the basis of their ability to fit the data. Numerous objective decision tools have been developed for structure characterisation. Criteria have been developed that take into account that fit is not the only guideline for inferring the functional relationships most adapted to the available experimental data. In the AIC-criterion, structure complexity (as reflected in the number of parameters) is used as a penalty with respect to improved fit. The major drawback of such a posteriori methods is that they rely on extensive parameter estimation which can take a considerable amount of computing time if the number of candidate models increases. Experience has shown that the time elapsing before a new impulse response becomes available allows the fitting of 2 to max. 3 different models (on a 486 PC). Real-time constraints therefore ask for another approach in which the parameter estimation problem is postponed until a choice is made on the structure that will optimally describe the data.

A priori model structure characterisation is proposed as a means of meeting the real-time constraints in this specific application. The choice between the different models is based on the Zernike moments that act as features specific for each mathematical structure. The neural network model classifier is being trained with simulated data obtained from the candidate models with different sets of realistic parameters.

The data flow in the proposed sensor is then the following: raw data is processed in the Zernike feature extractor; the Zernike moments are used as input to the Neural network Model Classifier that points to the most probable model structure; with this model structure, parameter estimation is performed; the output vector from the sensor contains both information on the chosen model and the associated parameters.

6.2. Results

The testbed candidate models are:

$$\frac{dS_1}{dt} = -K \cdot S_1 \cdot X, \tag{12}$$

$$\frac{dS_1}{dt} = -\mu_{max} \cdot \frac{S_1}{K_m + S_1} \cdot X, \tag{13}$$

$$\begin{cases} \dfrac{dS_1}{dt} = -\mu_{max1} \cdot \dfrac{S_1}{K_{m1} + S_1} \cdot X, \\ \dfrac{dS_2}{dt} = -\mu_{max2} \cdot \dfrac{S_2}{K_{m2} + S_2} \cdot X, \end{cases} \tag{14}$$

where S_1 and S_2 are the substrates; μ_{max}, μ_{max1} and μ_{max2} are the maximum growth rates on the respective substrates; K_m, K_{m1} and K_{m2} are the affinity constants that describe the dependency of the growth rate on the substrate concentration; K is a first order constant. Note that the biomass X is assumed constant within the short time interval of the experiments. The Oxygen Uptake Rate is then defined as:

$$OUR = -\sum_{\forall i} Y_{O_2,i} \cdot \frac{dS_i}{dt} \tag{15}$$

with $Y_{O_2,i}$ the amount of oxygen consumed per substrate degraded.

In order to obtain a neural network model classifier that is able to distinguish the models by looking at the Zernike features of the OUR-data, that neural network has to be trained by giving it examples of the models. Based on these training-patterns the network adapts itself until it "knows" how to classify all the training-patterns. A number of training-patterns were obtained from simulated data for each of the model classes with different realistic parameters, based on Henze et al. (1987). After the neural networks were trained, their performance was tested on real-life as well as simulated data to which random noise was added. These patterns were then classified. Apart from the described Sudharsanan neural network a backpropagation neural network was trained with the same training-patterns.

The performances of the trained neural networks and the ability of the neural networks to learn all the training-patterns correctly depended strongly on the choice of the neural network parameters and on the parameters for the feature extractor. In the end, all the resulting networks were able to classify the simulated data with added noise perfectly. The neural networks were, however, not very good at classifying the real-life data; both the Sudharsanan and backpropagation network correctly classified about 60% of the experimental results. However, the data that are to be described by the mathematically most complicated model were classified for only 30% to the proper model class. However, the two kinds of neural networks gave different results.

The stated advantages of Sudharsanan were found to be correct, the number of hidden nodes was low and the approximation capabilities were good. The learning-time was long (usually around 30 h on an Apollo DN10000 uni-processor with 16 Mb internal memory) for both kinds of neural networks. The time that was needed for one learning-step for the backpropagation neural network was smaller

than that for the Sudharsanan neural network, but the number of needed learning-steps was smaller for the Sudharsanan network. All in all the total learning-time was equal for both kinds of neural networks.

7. Conclusion

A novel approach using Zernike features and a recurrent neural network as classifier has been presented. This semi-inductive model identification technique was found to be not as powerful as expected in the case of a highly non-linear biotechnological process. Presumably, the discriminative power of the Zernike features is too low. Therefore, further research will be oriented towards evaluation of other coding schemes as feature extractors, such as fast Fourier transform, Karhune-Loève transform, decimation of a signal, etc.

Acknowledgements

The authors are grateful to Willy Verstraete and Ghislain Vansteenkiste, Professors at the University of Ghent, Belgium, for their initiation and support on the developments.

References

Box, G.E.P., Hunter, W.G. and Hunter, J.S., 1978. Statistics for Experimenters. Wiley, New York.

Cellier, F.E., 1991. Continuous System Modeling, first edition. Springer, New York.

Henze, M., Grady Jr., C.P.L., Gujer, W., Marais, G.v.R. and Matsuo, T., 1987. Activated sludge model nr. 1. Scientific and Technical Reports 1, IAWPRC, London.

Hopfield, J.J., 19982. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA, 79: 2554–2559.

Hopfield, J.J. and Tank, D.W., 1985. "neural" computation of decisions in optimization problems. Biol. Cybern., 52: 141–152.

Karplus, W.J., 1976. The spectrum of mathematical modelling and systems simulation. In: L. Dekker (Editor), Simulation of Systems. Proceedings of 8th Imacs (formerly AICA) Congress. North Holland, Amsterdam, pp. 5–13.

Klir, G.J., 1985. Architecture of Systems Problem Solving. Plenum Press, New York.

Sudharsanan, S.I., 1990. Equilibrium characterization for a class of dynamical neural networks with applications to learning and synthesis. PhD thesis, University of Arizona, Department of Electrical and Computer Engineering, Tucson, AZ.

Teague, M., 1988. Image analysis via the general theory of moments. J. Opt. Soc. Am., 70: 920–930.

Uyttenhove, H.J., 1979. SAPS — System Approach Problem Solver. PhD thesis, SUNY Binghampton, New York.

Van Impe, J.F., Vanrolleghem, P.A., Verstraete, W., De Moor, B.L. and Vandewalle, J., 1991. Model based monitoring and control of activated sludge wastewater treatment processes. Part ii: Nonlinear control of the biotransformation and the sedimentation process. In: E.J.H Kerckhoffs, H. Koppelaar, H. Van Der Beken and G.C. Vansteenkiste (Editors), Modelling and Control of Water Resources Systems and Global Changes. SCS, San Diego, CA, pp. 221–226.

Vanrolleghem, P., Dries, D. and Verstraete, W., 1990. Rodtox: Biosensor for rapid determination of the biochemical oxygen demand and the on-line monitoring of the toxicity of wastewaters. In: Proceedings Fifth European Congress on Biotechnology, July 1990, Copenhagen, Denmark, pp. 161–164.

Vanrolleghem, P.A., Van Impe, J.F., Vandewalle, J. and Verstraete, W., 1992. Advanced monitoring and control of the activated sludge process: On-line estimation of crucial biological variables in a structured model with the rodtox biosensor. In: Proceedings IFAC-BIO2, Symposium on Modeling and Control of Biotechnical Processes, April 1992, Keystone Colorado, USA. IFAC.

Vansteenkiste, G.C., Spriet, J. and Bens, J., 1979. Structure characterization for system modeling in uncertain environments. In: Methodology in Systems Modelling and Simulation. North Holland, Amsterdam, pp. 289–303.