European Symposium on Computer-Aided Process Engineering –13
A. Kraslawski (Editor)

Comparison of Methods for Training Grey-Box Neural Network Models

Gonzalo Acuña[1], Francisco Cubillos[2], Jules Thibault[3] and Eric Latrille[4]

[1] Departamento de Ingeniería Informática; [2] Departamento de Ingeniería Química, Universidad de Santiago de Chile, Avda. Ecuador 3659, Casilla 10233, Santiago, Chile. E-mail: gacuna@lauca.usach.cl
[3] Department of Chemical Engineering, Laval University, Sainte-Foy (Quebec) Canada G1K 7P4
[4] INRA-LGMPA, CBAI INA-PG, 78850 Thiverval-Grignon, France.

Abstract
Due to its inherent plasticity, neural network models are well suited to represent complex functions such as those encountered in bioprocesses. In this paper, neural networks were used to model kinetic rate expressions that form an integral part of a grey-box model. A grey-box model normally consists of a phenomenological part (differential equations of heat and/or mass balances) and an empirical part (a neural network in this paper). The objective of this investigation is to compare three different methods to come up with the same neural network to represent two kinetic rate expressions that are used directly in the grey-box model. In one method (the direct method), the model is fitted directly on data obtained from the derivative of smoothed state variables whereas the other two methods (indirect methods) use a non-linear regression algorithm to fit the complete grey-box model in order to derive specific kinetic rate expressions that minimise an objective function made of some measured state variables. Results clearly show that indirect methods are superior to direct methods for the prediction of state variables. However, the derived kinetic rate models are not unique.

Keywords: Neural network, Grey-Box model, kinetic rate expression, fermentation

Introduction
The use of feedforward neural networks (FNN) for modelling dynamic non-linear systems, like biochemical processes, has been the subject of a great deal of recent research work. Some important features of neural network models used for biotechnological applications are their capability to act as universal approximators of non-linear functions, their applicability to multivariate modelling and their ability to assess process dynamics (Bishop, 1994).
Most of the efforts have been put on developing non-parametric or black-box models that simply map a set of input variables into the corresponding outputs.
However, in many cases it seems to be convenient to take advantage of the *a priori* knowledge of the non-linear system frequently expressed as a set of ordinary or partial differential equations representing mass or energy balances. In biotechnological processes, the most difficult modelling task remains the determination of some key parameters, like the specific kinetic rate expressions, which are usually complex functions of the various state variables.
To cope with this problem, grey-box models have been suggested (Psichogios and Ungar, 1992) which combine prior knowledge included in the phenomenological model with neural networks. This kind of models has proven to be successful for dynamic systems, with better generalisation features. In addition they can be identified with a more reduced set of data than the black-box equivalent model (Psichogios and Ungar, 1992).
Thompson and Kramer (1994) classified these grey-box models into two main types: FNN bringing intermediate values (parameters or variables) to be used by the phenomenological model (*series grey-box models*) or FNN compensating the bias of the parametric model in parallel with the dynamic model (*parallel grey-box models*). Van Can *et al*. (1996) demonstrated that the series approach results in more accurate models, with better dimensional and range extrapolation properties than the other approaches.
Focusing on the way the FNN is identified, two main approaches can be found: the most currently used (*direct approach*) consists in minimising a quadratic error function between the outputs of the neural network and the ideal values of the parameters computed from the smoothed state variables, using the available experimental data (see Figure 1).

European Symposium on Computer-Aided Process Engineering –13
A. Kraslawski (Editor)

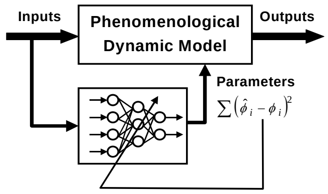

Figure 1: Series Grey-Box Model: Direct Approach

An alternative approach (*indirect approach*) to identify a suitable neural network of some key parameters of the process consists of using the values of the measured state variables rather than the errors of the output of the neural network. This procedure allows to determine indirectly the unknown dynamic model parameters (e.g. the kinetic rate coefficients) without having to perform additional data processing of the measured state variables (see Figure 2).
In this paper we propose a comparison between three methods for developing series grey-box models. Two of them can be classified as indirect approaches. The first consists of minimising an objective function of the state variables (outputs) by means of a non-linear optimisation technique (Thibault *et al.*, 1997) while the second performs the backpropagation of the output error into the neural network weights going through the phenomenological model equations (Piron *et al.*, 1997). Finally, the third method corresponds to the direct approach (Cubillos *et al.*, 1996).

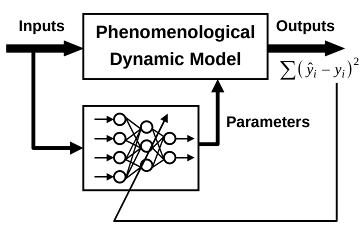

Figure 2: Series Grey-Box Model: Indirect Approach

The three methods are compared by modelling the fermentation of glucose to gluconic acid by the micro-organism *Pseudomonas ovalis* in a batch stirred tank reactor (Johansen and Foss, 1995).

Process Description
In this paper, the fermentation of glucose to gluconic acid by the micro-organism *Pseudomonas ovalis* in a batch stirred tank reactor is considered. The overall mechanism could be expressed as follows:

$$\text{Cells + Glucose + Oxygen} \longrightarrow \text{More cells}$$

$$\text{Glucose + Oxygen} \xrightarrow{Cells} \text{Gluconolactone}$$

$$\text{Gluconolactone + Water} \longrightarrow \text{Gluconic Acid}$$

This process has been the subject of considerable study and the following state space model has been derived to represent respectively the concentration of the cell (X), gluconic acid (p), gluconolactone (l), glucose (s, the substrate) and dissolved oxygen (c).

$$\dot{X} = \mu_m \frac{sc}{k_s c + k_0 s + sc} X = f_1 \quad X \qquad (1)$$

$$\dot{p} = k_p l \qquad (2)$$

$$\dot{l} = v_l \frac{s}{k_l + s} X - 0.91 k_p l = \qquad (3a)$$

$$= f_2 X - 0.91 \; k_p l \qquad (3b)$$

$$\dot{s} = -\frac{1}{Y_s} \mu_m \frac{sc}{k_s c + k_0 s + sc} X - 1.011 \, v_l \frac{s}{k_l + s} X = \qquad (4a)$$

$$= -\frac{1}{Y_s} f_1 \quad X - 1.011 \, f_2 \quad X \qquad (4b)$$

$$\dot{c} = K_L a \left(c^* - c\right) - \frac{1}{Y_0} \mu_m \frac{sc}{k_s c + k_0 s + sc} X - 0.09 \, v_l \frac{s}{k_l + s} X = \qquad (5a)$$

$$= K_L a \left(c^* - c\right) - \frac{1}{Y_0} f_1 X - 0.09 \, f_2 X \qquad (5b)$$

In this investigation, the values of the coefficients used to perform the simulation were identical to those used by Johansen and Foss (1995). For the sake of completeness, these values are reproduced in Table 1. To generate the batch fermentation data, a full two-level factorial design with two duplicates at the centre point was performed by varying three variables: the initial cell concentration (0.1, 0.4 and 0.7 UOD/mL), the initial substrate concentration (30, 40 and 50 g/L) and K_La (150, 175 and 200 h^{-1}). A total of 10 experiments of duration of 10 h with a sampling period of 0.5 h were simulated for a total of 200 values of each variable. It was assumed that all state variables were measured at each sampling instant. A 5% noise level was added to the simulated state variables in order to obtain more realistic data.

Table 1 - Constants in the state space simulation model.

μ_m	k_s	k_0	k_p	v_l
0.39	2.50	0.00055	0.645	8.30
h^{-1}	g/L	g/L	h^{-1}	mg/UOD h
k_l	K_La	Y_s	Y_0	C*
12.80	150-200	0.375	0.890	0.00685
g/L	h^{-1}	UOD/mg	UOD/mg	g/L

In this system of equations, the ability to represent adequately each of the five state variables lies in the

European Symposium on Computer-Aided Process Engineering –13
A. Kraslawski (Editor)

determination of functions f_1 and f_2. The other coefficients such as the yield coefficients and the oxygen mass transfer coefficient can usually be obtained more easily. The values of f_1 and f_2 will be calculated using a neural network.

Neural Networks

All the neural networks used in this work were multilayer perceptrons with sigmoidal transfer functions in the hidden and the output layer neurons. For FNN training, second-order minimisation routines like Levenberg-Marquardt and BFGS Q-Newton were used. The data set was splitted into a training set including 150 patterns and a test set with the remaining 50 patterns. To assure a good convergence, the training task was performed 20 times with different randomly chosen initial weights.

Identification procedures

Direct Approach: One way to identify a proper neural model of f_1 (a highly non-linear expression) and f_2 (a Monod model) is to compute them from the first derivatives of the involved state variables.
In this work, Euler discretization of equations (4b) and (5b) allowed to compute f_1X and f_2X at each sampling rate. These values were used to identify two static neural network models as shown in Figure 3.

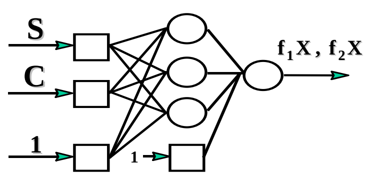

Figure 3: Static neural models for f_1X and f_2X.

Once the neural models were identified, they were used to compute the remaining state variables, *X*, *p* and *l* by means of equations (1), (2) and (3b). Finally, f_1 and f_2 were computed simply dividing f_1X and f_2X by the estimated values of X.

Indirect Approach: Two ways for indirectly obtaining f_1 and f_2 were explored. In the first one (*Indirect 1*) the neural network model shown in Figure 4 was identified in order to minimise the following objective function using a non linear optimisation routine (Thibault *et al*., 1997).

$$J=\sum_{i=1}^{N}\frac{1}{3}\left(\hat{X}_n-X_n\right)^2+\frac{1}{15}\left(\hat{l}_n-l_n\right)^2+\frac{1}{50}\left(\hat{s}_n-s_n\right)^2 \quad (6)$$

The outputs of the FNN were not used to directly identify the neural weights but to feed the state space model of equations (1) – (5) (Figure 2).

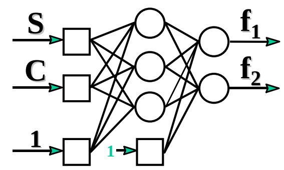

Figure 4: Static neural network for f_1 and f_2 in the Indirect 1 approach

The second way (*Indirect 2*) consists on backpropagating the output error of a neural network that exactly reproduces the state space model of equations (1) –(5) by using adequate fixed weights and activation and transfer functions, as schematically shown in Figure 5 (Piron *et al*., 1997). In this case the same static neural network of Figure 4 lies within the global neural network that reproduces the whole model, bringing the adequate f_1 and f_2 values. It should then be emphasised that the only variable weights of this global model to be identified are the same as those of the FNN shown in Figure 4. The quadratic output error function of the global neural network was modified in order to be the same as (6), hence, only *X*, *l* and *s* values were used for minimisation.

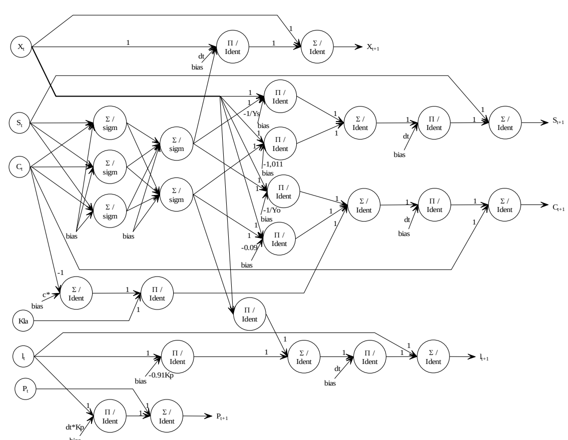

Figure 5: Neural network representation of the dynamic model of equations (1) – (5)

Results and Discussion

In order to compare the performance of the neural network models brought by each of the three described methods, the Residual Standard Error (RSE) index was used (Eq. 7)

$$RSD=\sqrt{\frac{\sum_{i=1}^{n}(o_i-p_i)^2}{N}} \qquad (7)$$

European Symposium on Computer-Aided Process Engineering –13
A. Kraslawski (Editor)

where o_i and p_i are the observed and predicted values respectively at time i, while N is the total number of data.
The RSD values obtained for the test set data are included in Table 2.

Table 2: RSD index for the state variables and the specific kinetics f_1 and f_2 when applying the Direct Method (DM), the Indirect 1method (I1), and the Indirect 2 method (I2) to the test set data.

	X UOD/ mL	l g/L	s g/L	f_1 1/h	f_2 1/h
DM	0.47	0.96	1.34	0.08	2.23
I1	0.12	0.35	1.00	0.03	0.47
I2	0.16	0.37	1.45	0.05	0.78

Table 2 shows close performances of the two indirect methods for almost all of the evaluated variables excepting *s*. In contrast, the Direct Method fails to give adequate predictions of some of the variables, specially the specific kinetic rate expressions. In this case, the noise strongly affects the evaluation of the derivatives of the state variables used to train the neural network. On the other hand, smoothing these variables also poses the problem of choosing an adequate smoothing algorithm that does not introduce a bias.
In contrast, the two Indirect Methods perform better because they do not have to evaluate the derivatives from the experimental data of any state variable. In fact they succeeded in finding good predictions of the state variables (see Figure 6A) included in their objective functions. However, the predictions of the specific kinetic rate functions were affected by the noise (see Figure 6B and 6C) thus showing that these kind of models are able to represent very well the dynamics of the state variables despite the presence of a large number of acceptable kinetic rate models.

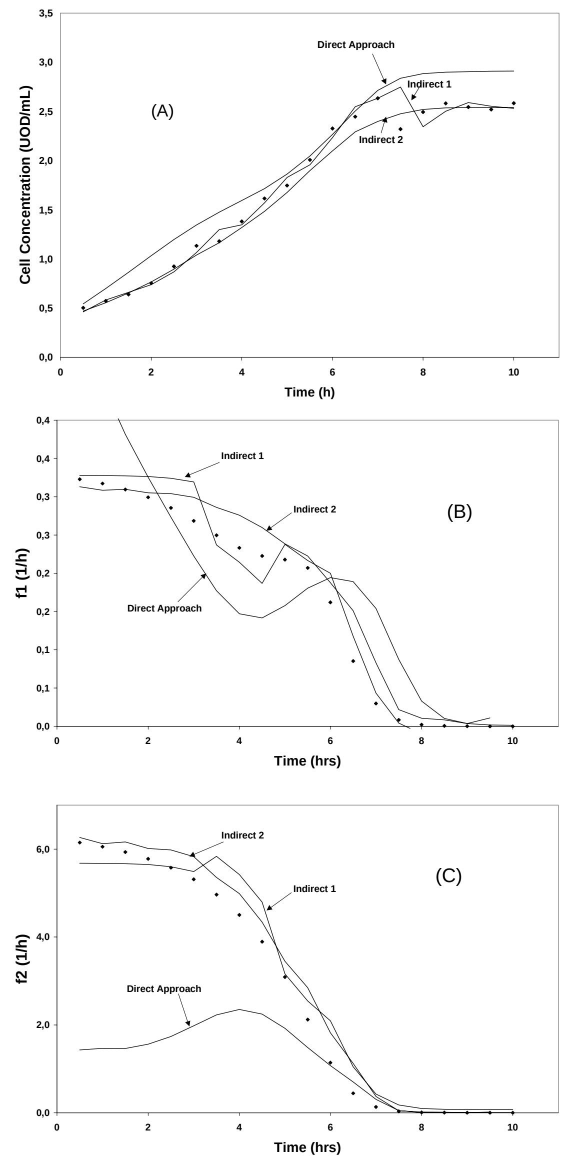

Figure 6: Experimental points and predicted values (solid lines) of X, f_1 and f_2 using the three analysed methods over one experiment included in the test set.

Conclusions

This paper has considered the use of neural networks in conjunction to a phenomenological model to form a grey-box model used to represent the dynamics of complex fermentation systems. The neural network *per se* was used to model kinetic rate expressions that are usually very difficult to obtain. Three different methods to fit the neural models were investigated. It was found that to obtain a good predictive model of state variables, the two indirect methods were superior to the direct method because they are less sensible to measurement noise than the direct approach. However, numerous neural network rate expression model lead to nearly identical overall predictive performance

European Symposium on Computer-Aided Process Engineering –13
A. Kraslawski (Editor)

References

Bishop C.M. (1994). Neural networks and their applications. *Rev. Sci. In*strum. **65**, 1803-1832.

Cubillos, FA, Alvarez, PI, Pinto, JC and Lima, EL (1996). Hybrid-Neural Modelling for Particulate Solid Drying Processes*. Powder Technology* **87**, 153-160.

Johansen, TA and Foss, BA (1995). Semi-Empirical Modeling of Non-Linear Dynamic Systems Through Identification of Operating Regimes and Locals Models. In: *Neural Network Engineering in Control Systems*, K Hunt, G Irwin and K Warwick, Eds., pp 105-126, Springer-Verlag,.

Piron, E, Latrille, E and René, F (1997). Application of artificial neural networks for crossflow microfiltration modelling: “black-box” and semi-physical approaches. *Computers Chem Engng.* **21**, 1021-1030.

Psichogios, D and Ungar, L. (1992). A hybrid neural network-first principles approach to process modeling. *AIChE J*. **38**, 1499-1511.

Thibault, J, Acuña, G and Grandjean, B. (1997), Identification of kinetic rate expressions in fermentation systems, Proceedings of the International *Conference on Engineering Applications of Neural Networks*; A. Bulsari, S. Kallio. Eds, Abo Akademis Tryckeri, Turku, Finland, pp. 123-126.

Thompson, M and Kramer, MA (1994). Modeling Chemical Processes Using Prior Knowledge and Neural Networks. *AIChE* **40**, 1328-1340.

Van Can, HJL, Hellinga, C, Luyben, KAM and Heijnen, JJ (1996). Strategy for Dynamic Process Modeling Based on Neural Networks in Macroscopic Balances. *AIChE* **42**, 3403-3418.

Acknowledgements

The authors would like to acknowledge grants from the Canadian Government (NSERC), from the Chilean Government (FONDECYT N^{o} 1970359, FONDECYT N^{o} 1980667), from USACH (DICYT grant 0698-19AL) and from joint Chilean and French Research agencies (ECOS C96-B01).