Modeling of Industrial Nylon-6,6 Polymerization Process in a Twin-Screw Extruder Reactor. II. Neural Networks and Hybrid Models

CLAUDIO AUGUSTO OLLER NASCIMENTO,[1] REINALDO GIUDICI,[1] NATÁLIA SCHERBAKOFF[2]

[1] Department of Chemical Engineering, Polytechnic School, University of São Paulo, P.O. Box 61548, 05424-970, São Paulo, SP, Brazil

[2] Rhodia S.A. (Rhône-Poulenc), Av. dos Estados 6144, cep 09210-900, Santo André, SP, Brazil

Received 9 December 1996; accepted 11 November 1997

ABSTRACT: This article describes the application of neural networks and hybrid models to the finishing stage of nylon-6,6 polycondensation in a twin-screw extruder reactor. A planned experiment in the industrial and in the pilot plant was employed to build the neural network and the hybrid model. The hybrid model combines information calculated from the phenomenological model with the neural network model. The comparison of experimental with calculated data shows good agreement. During two years, industrial data were collected. The comparisons of the models' prediction with these data were performed and reasonable results are achieved from the industrial point of view. These models help an increase of industrial production of about 20%. © 1999 John Wiley & Sons, Inc. J Appl Polym Sci 72: 905–912, 1999

Key words: nylon-6,6 polymerization; neural network model; twin-screw extruder; hybrid model

INTRODUCTION

Although the phenomenological modeling of the nylon-6,6 extruder process gives a reasonable prediction from the industrial point of view (Giudici et al.[1]), the nature of the process presents some difficulties, such as the complex nature of the flow in the extruder, the condensate removal (mass transfer limitations), and the kinetics of polycondensation and degradation reactions. The lack of internal or intermediate measurements along the industrial extruder reactor represents a limitation for a detailed model validation. In addition, the balance equation for the concentration of polymer end groups allows one to calculate only the number-average molecular weight. Thus, empirical correlations have to be included in the model to predict important polymer properties of interest to the final use, such as relative viscosity.

In light of these difficulties for developing complex phenomenological models, we propose alternative models of the nylon-6,6 process, using a pure neural network model as well as a hybrid model in which a phenomenological model is combined with neural networks. The neural network model consists of a set of processing units called neurons, connected to one another. The neural network used in this article is a feed-forward network with one hidden layer. By adjusting parameters in the coupling between neurons, the network is capable of learning from a set of numerical data corresponding to the input and the desired output.

Correspondence to: C. A. Oller Nascimento (oller@usp.br).

Journal of Applied Polymer Science, Vol. 72, 905–912 (1999)
 CCC 0021-8995/99/070905-08

REVIEW

Nylon-6,6 Polycondensation Process

Nylon-6,6 polymer is produced from hexamethylene diamine and adipic acid monomers. The design of the finishing stage reactors of this process requires special features since they usually operate with high-viscosity polymers under difficult conditions of condensate removal and heat transfer. Extruder reactors can be effective in handling these conditions, as described in reviews on reactive extrusion.[2–4]

The process under study is carried out in a self-wiping co-rotating twin-screw extruder. The extruder is fed with melt polymer. In the first section of the extruder, there is a vacuum vent port that promotes degassing conditions for water evaporation. It is followed by a second section, in which the main phenomenon is the polycondensation reaction with little or no evaporation taking place.

Neural Network Models

Neural networks have been attracting great interest as predictive models, as well as for pattern recognition.[5–6]

The potential for employing neural networks in the chemical industry is tremendous because nonlinearity in chemical processes constitutes the general rule. Neural networks possess the ability to learn what happens in the process without actually modeling the physical and chemical laws that govern the system. The success in obtaining a reliable and robust network depends strongly on the choice of process variables involved, as well as the available set of data and the domain used for training purposes.

The neural network employed is the feed-forward network (Fig. 1). In this network, information propagates in only one direction and is useful for steady-state modeling.

In general, the networks consist of processing neurons (represented by circles) and information flow channels between the neurons, usually denominated interconnects. The boxes represent neurons where the inputs to the network are stored. Each processing neuron calculates the weighted sum of all interconnected signals from the previous layer plus a bias term, eq. (1), and then generates an output through its activation function, eq. (2). The most widely used networks are made up of three layers, the input, hidden, and output layers. According to the literature,[7] this network is a universal approximator.

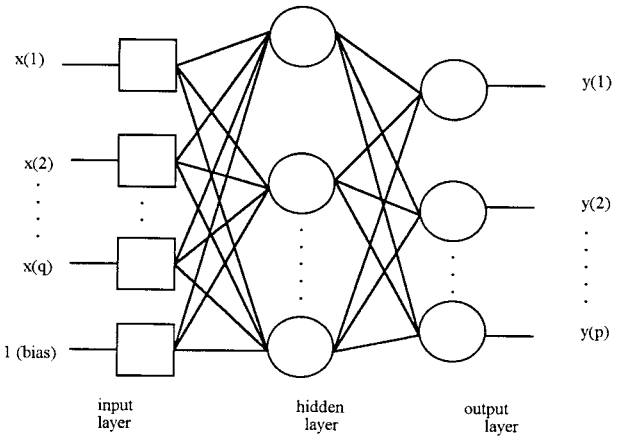

Figure 1 Multilayer feed-forward neural network.

$$S_j = \sum_{i=1}^{n} w_{i,j} X_i + w_{n+1,j} \tag{1}$$

$$f(S_j) = \frac{1}{1 + e^{-S_j}} \tag{2}$$

The system learns by making changes in the weights ($w_{i,j}$). The input and output variables chosen for the network training are normalized. At present, the most extensively adopted algorithm for the learning phase is the back propagation algorithm, which is a generalization of the steepest descent method. It consists of minimizing the mean square error (E), defined as

$$\min(E) = \min\left(\sum_{m=1}^{r} \sum_{k=1}^{p} (y_k^{(m)} - O_k^{(m)})^2 \right) \tag{3}$$

where y_k comes from the r input–output pairs of data (x, y) available for training the network, and O_k is obtained from the output layer signal, calculated by the following expression:

$$O_j = f(S_j) \tag{4}$$

In the traditional gradient approach for minimizing the mean square error E with respect to the weights $w_{i,j}$, one calculates the derivatives $dE/dw_{i,j}$ and then moves in the direction of steepest descendent. This technique requires using all the input–output pairs to determine the gradient.

The back propagation algorithm also uses gradient information to change the weights; however, it is calculated with respect to only one input–output pair at a time.[5]

Neural networks are characterized by the large number of parameters involved (weights) due to the high connectivity among the neurons. Normally, the data are split into two sets. One set is used to train the network, and the other is used to test the prediction capability. In training a network, the objective is to find an optimum set of weights. When the number of weights is higher than the number of available data, the error in fitting the nontrained data initially decreases but then increases as the network becomes overtrained. In contrast, when the number of weights is smaller than the number of data, the overfitting problem is not crucial.

Neural network computing is one of the fastest growing fields of artificial intelligence due to its ability to learn nonlinear relationships. Neural networks represent a promising alternate approach for modeling polymeric systems. Chan and Nascimento[8] have used neural networks to model industrial high-pressure olefin polymerization and comparison between the neural network and mechanistic models were made.

NEURAL NETWORK FOR THE EXTRUDER REACTOR

The finishing stage of nylon-6,6 is carried out in a twin-screw extruder. Modeling this process by a neural network model was carried out by Nascimento et al.[9] This technique requires reliable data to train the network. The following process industrial data were used to train the network.

- As the extruder input variables:

  (1) temperature;
  (2) pressure of the vacuum system;
  (3) flow rate;
  (4) pressure in the extruder head;
  (5) pressure after the pump on the extruder head;
  (6) screw rotation speed.

- As the output variables:

  (1) the relative viscosity (RV);
  (2) the amine end-groups (CA);
  (3) the carboxyl end groups (CC).

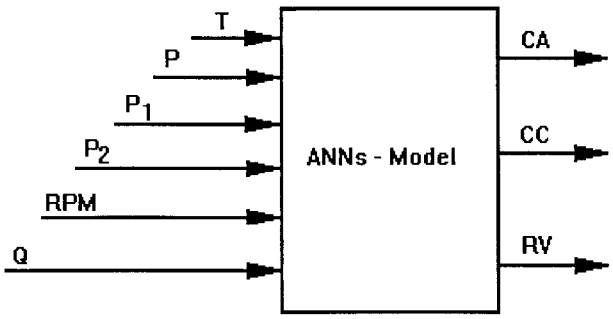

Figure 2 Neural network model (input and output variables).

Figure 2 shows schematically the input and output information fed to the neural network.

Neural Network Training

A set of 21 planned experiments obtained in the industrial unit was used to train the neural network (15 runs) and another set to check the trained neural network (6 runs). The neural network has several parameters that can be chosen, as follows: the number of iterations and the number of neurons in the hidden layer. In order to select the optimum number of neurons to be used, a sensitivity analysis of the mean square error as a function of the number of neurons was carried out for 20,000 presentations, as an example. The results are shown in Figure 3 concerning the learning set and the test set. The number of neurons that provides the smallest mean square error for the training set is $NH = 6$. In contrast, the number of neurons that provides the smallest mean square error for the test set is $NH = 10$. Figure 4 shows the experimental versus predicted relative viscosity for the learning and the test set. The achieved agreement is within the experimental error.

Usually, the error in the learning set is much smaller than in the test set. However, the test set gives an evaluation of the interpolation–extrapolation potential of the neural network.[10]

This neural network model, as presented, is a black box of all process. Although the neural network absorbs most of the nonlinearity of the process, it is highly dependent on the quality and the range of the data fed into the network. A promising alternative is to apply neural networks to the parts of the problem where it is not possible to use phenomenological models. In such cases, a combined model is employed, the hybrid model.

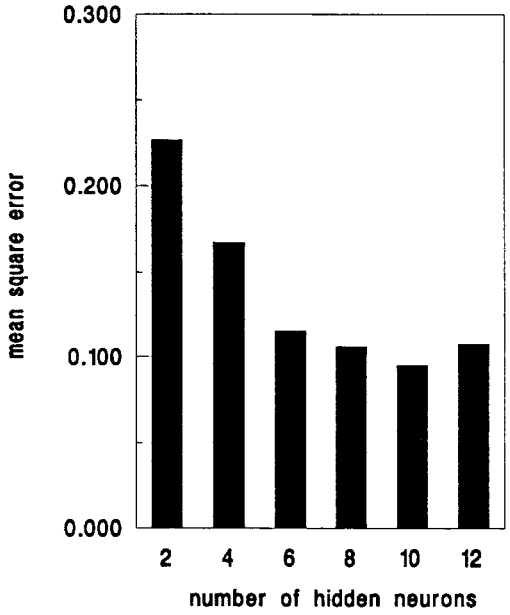

Figure 3 Residual sensitivity analysis after 20,000 iterations (test-set).

HYBRID MODEL

The steady-state phenomenological model (Giudici et al.[1]) using the plug flow assumption proved to be robust for predicting the *CA* and *CC* values for the nylon 6,6 processes compared with the experimental data. However, the prediction of *RV* by an empirical polynomial equation as a function

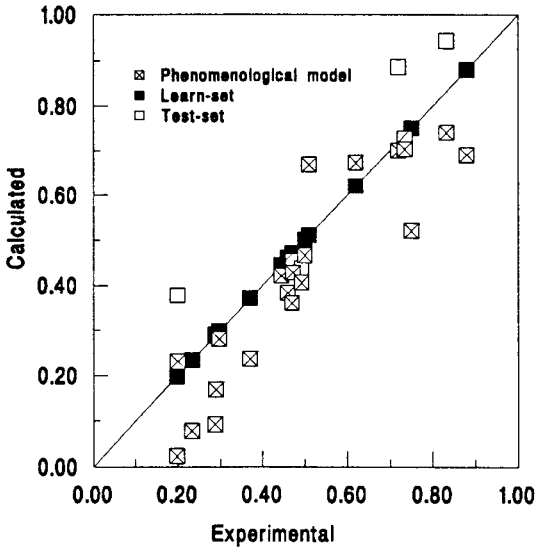

Figure 4 *RV* calculated by neural network (*NH* = 10) and phenomenological models versus experimental data.

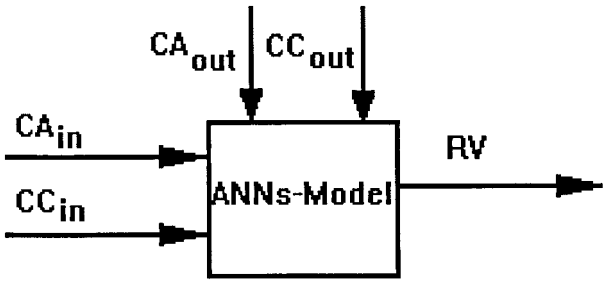

Figure 5 *RV* neural network model.

of M_n gave values lower than the expected industrial data (mainly at low values of *RV*). This may be explained by the fact that the number-average molecular weight M_n for linear polycondensation is calculated from values of *CA* and *CC* by eq. (5), as follows:

$$M_n = 2 \cdot 10^6/(CA + CC) \tag{5}$$

This equation does not reflect the true M_n if some degradation of the polymer occurs.

One possible reason for the large spread of the empirical polynomial equation is that degradation is not taken into account. The inclusion of the end groups produced by the degradation reactions could improve the results. However, as already

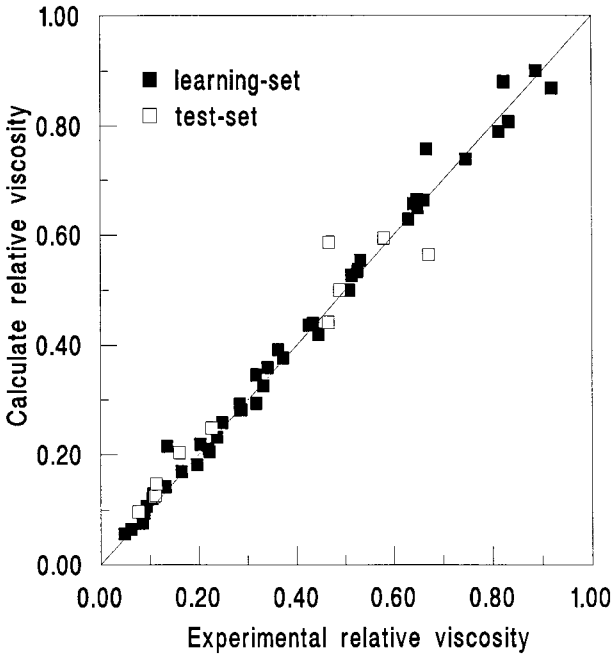

Figure 6 Comparison of *RV* correlation by neural network with experimental data.

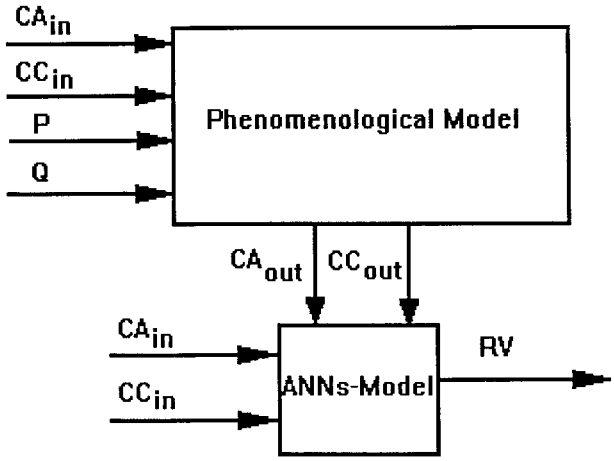

Figure 7 Overview of the hybrid model.

pointed out, this is not feasible due to the lack of these measurements.

If no degradation occured, the consumption of amine would be equal to the consumption of carboxyl end groups. One way to quantify the degradation is to compare the *CA* variation with the *CC* variation along the extruder. Giudici et al.[1] presented the ratio of differences of the input and output values of *CA* and *CC*. If no degradation occurs, the ratio should be 1. However, degradation always happens as a result of thermal shear, and other effects. Values smaller than one are an indication that degradation has occurred in some point in the process. Therefore, a feasible alternative to account for the degradation is

$$RV = f(CA, CC, CA_0, CC_0) \tag{6}$$

where CA_0 and CC_0 are the concentrations of amine and carboxyl end groups in the feed stream of the extruder, and *CA* and *CC* are at the extruder outflow. The neural network model presented in Figure 5 was developed to correlate *RV* according to eq. (6).

Two sets of experiments were used to feed the neural network. One consists of the 21 planned experiments made in the industrial plant, discussed in Giudici et al.[1] An additional 44 planned experiments were run in a pilot screw extruder reactor. The main reason for using the data from the pilot plant was that the values of *CA* and *CC* can be changed at the extruder entrance. In the industrial plant, it is not possible to change the values of *CA* and *CC* in the extruder inflow because this would provoke a change in the process beyond the extruder. The only permissible variation of *CA* and *CC* of the extruder inflow was that which resulted from process variability, representing in a narrow range. All the data were previously treated statistically to identify possible gross error. Four data were considered to possibly contain gross error and were excluded. The remaining data (21 from the industrial plant and 40 from the pilot plant) were randomly split into two groups for the learning set (80%) and test set (20%) of the neural network. These data are presented in Figure 6. The neural network model adjusted the data very well over the entire range of relative viscosity tested. The agreement of the test set was compatible with the experimental *RV* error analysis.

The structure of the hybrid model, in which the phenomenological model and the neural network model are coupled, is presented in Figure 7.

A hybrid model was developed in such a way that the information about *CA* and *CC* are computed by the phenomenological model, and *CA* and *CC* are used to calculate *RV* by the neural network model. The neural network model for *RV* is fed with the following input information: the value of *CA* and *CC* in the feed stream and *CA* and *CC* predicted at the outflow of the extruder. The output information is the *RV* value.

At this point, it is interesting to compare the empirical relation[1] and the proposal neural network model for *RV* prediction. The neural network uses more information (*CA*, *CC*, CA_0, CC_0) than the empirical relation (*CA*, *CC*). Although the accuracy of the empirical relation might be improved by adding CA_0 and CC_0 information, one would have is to identify *a priori* the relationship between these variables. The advantage of

Table I Mechanistic Model (MM) × Hybrid Model (HM) × Plant Data (PD)

Run	Extruder	*RV* : MM	*RV* : HM	*RV* : PD
Run 1B	Extruder 3	0.450	0.483	0.517
Run 2B	Extruder 3	0.833	0.817	0.817
Run 3B	Extruder 8	0.433	0.483	0.517

Table II Mechanistic Model (MM) × Hybrid Model (HM) × Plant Data (PD)

Run	Extruder	RV : MM	RV : HM	RV : PD
Run 1B	Extruder 3	0.450	0.483	0.517
Run 2B	Extruder 3	0.833	0.817	0.817
Run 4B	Extruder 3	0.450	0.483	0.500
Run 5B	Extruder 3	0.800	0.767	0.783

the neural network approach is that this relationship is automatically given by the neural network structure.

Comparison Between Industrial Plant Data and the Hybrid Model

An increase of production was planned after the annual plant shutdown. One month after the start up, new plant data was collected in extruder 3 (runs 1B and 2B) for two different polymer grades, without a production increase. The developed models were checked, and the predicted RV values are shown in Table I. The mechanistic model (MM = phenomenological model with RV predicted by polynomial correlation[1]) has predicted the plant data very well for high RV, but failed for process situations where the resulting polymer had a lower RV. This may be explained by the fact that although the mechanistic model predicts CA and CC well, the empirical relation of $RV = f(CA, CC)$ is the weak point. Tests were then carried out in other similar industrial extruders (run 3B), and the models were applied in the same way. The hybrid model shows good agreement for polymer of low and high RV.

The comparison of MM and HM compared with the plant data showed the possibility of increasing the production by changing process variables. A production increase of 15% of extruder 3 was carried out and the model prediction (MM and HM) was checked with the plant data (runs 4B and 5B) for polymer of different grades (Table II). The hybrid model appeared to be more robust for extrapolation purposes than the MM model.

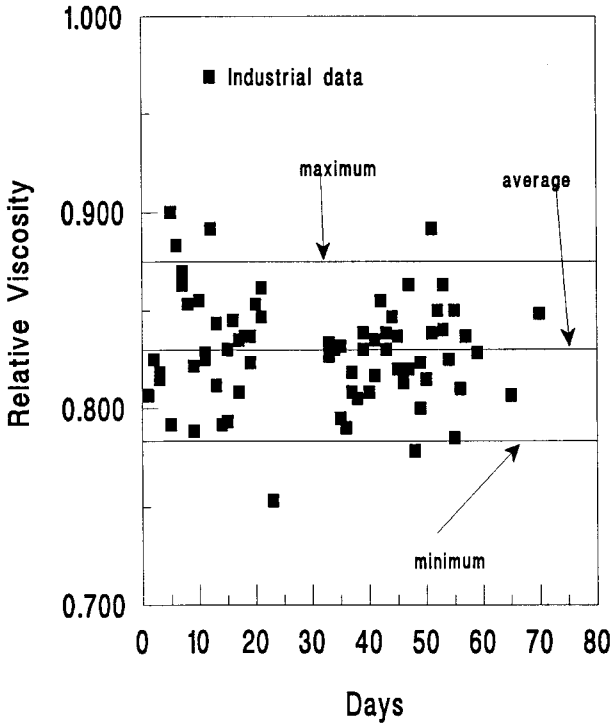

Figure 8 Confidence intervals (95%) of the relative viscosity (high relative viscosity).

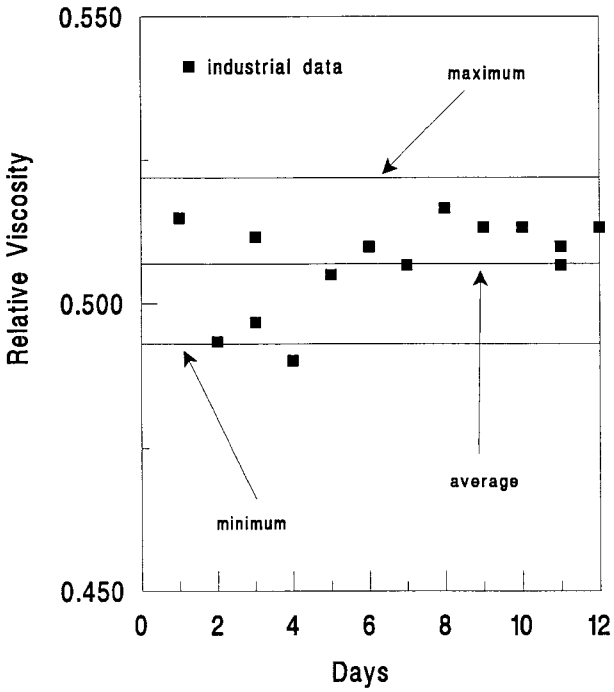

Figure 9 Confidence intervals (95%) of the relative viscosity (low relative viscosity).

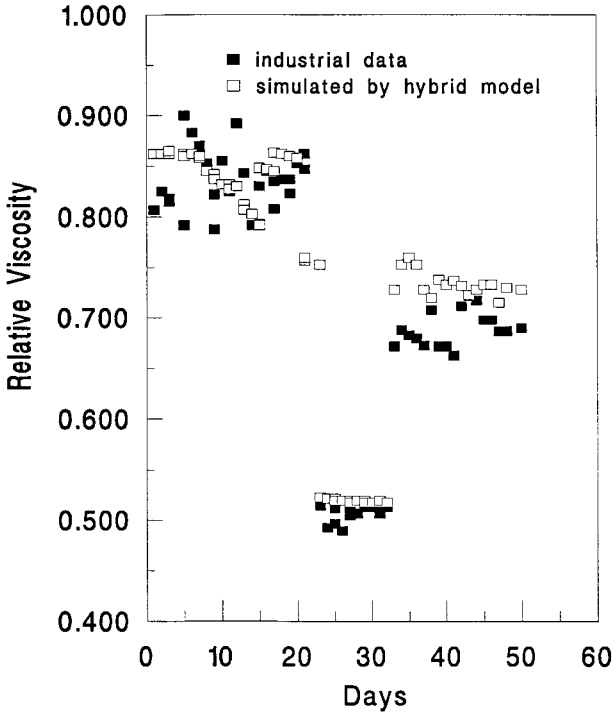

Figure 10 Experimental versus calculated data (year of 1995; extruder 3).

To show the consistency between experiment and the calculated data, an analysis of operation data must be done. Data covering a period of about 1 year were analyzed. During this period, different polymer grades were produced. The data were treated and separated into groups according the relative viscosity (*RV*) level. Figures 8 and 9 show the variation of the relative viscosity caused by process variability. The range of acceptable data was within the confidence interval of 95%. Comparing the acceptable range of the process variability with the model predictions, we may conclude that the predictions of both models are acceptable at high relative viscosity. At low relative viscosity, the predictions of the hybrid model are acceptable, but those of the MM model not.

We have not rejected the data outside the region of the limits shown in Figures 8 and 9 as an outliers since we do not have enough information to eliminate the possibility of abnormal problems that might occur in the process during these periods. However, the elimination of the possible outliers would not substantially alter the analysis.

The comparison of the hybrid model predictions with the industrial data collected during the year of 1995 is shown in Figures 10 and 11. The model predictions showed good agreement with the industrial data collected during 1995 in different extruders.

CONCLUSIONS

In this article, we have successfully modeled the nylon-6,6 polycondensation process in a twin-screw extruder reactor via a neural network and by a hybrid model in which phenomenological with neural network models are combined. These models were shown to be an important tool for increasing the plant production. Some plant bottlenecks were achieved after an increase of 20% of plant production.

Planned industrial and pilot plant experiments were used to correlate the neural network model. Although neural networks are relatively easy to use, prior process variables should be known, and the quality of information and the domains of the data are crucial to obtain reasonable results. Neural networks are especially useful for predict *RV* as a function of

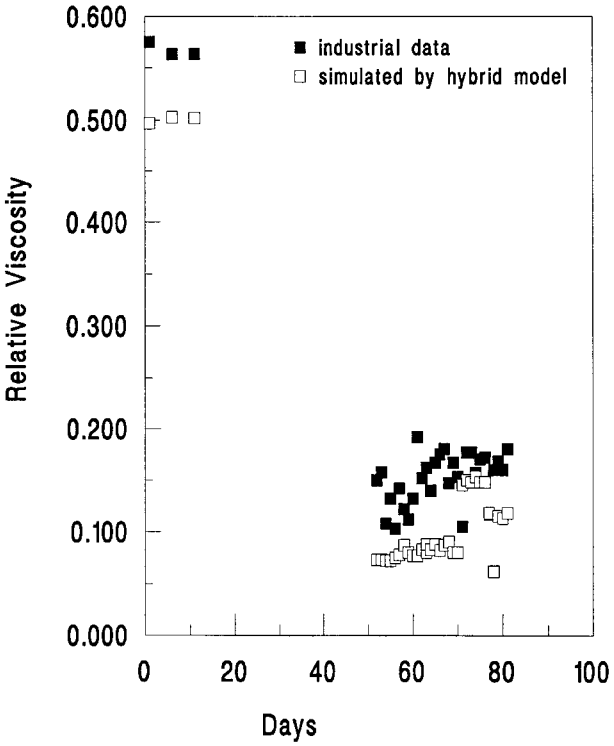

Figure 11 Experimental versus calculated data (year of 1995; extruder 7).

the end group concentrations via a relation, which is not known *a priori.*

The authors would like to express their gratitude to Rhodia S. A. (Rhône Poulenc Group) for supporting this work and to Prof. Frank Quina for revising the manuscript. Also, partial support from FAPESP and CNPq is gratefully appreciated.

NOTATION

CA concentration of amine end groups
CC concentration of carboxyl end groups
CA_0 concentration of amine end groups in the feed stream
CC_0 concentration of carboxyl end groups in the feed stream
E quadratic deviation
f Sigma function, in eq. (2)
M_n number-average molecular weight
n number of input variables in the neural network model
NH number of neurons in the hidden layer
O_j output from neuron j
O_k output from neuron k in the output layer
p number of output variables in the neural network model
P pressure imposed by vacuum system
P_1 pressure in the extruder head
P_2 pressure after the pump in the extruder head
Q flow rate
r number of input–output pairs in the learning set
RPM screw rotation speed
RV relative viscosity
S_j weighted sum of inputs to a neuron
T temperature
$w_{i,j}$ weight of variable i, in neuron j
X_i normalized input variable i in the neural network model
y_k normalized output variable k in the neural network model

Superscripts

m point in learning set

REFERENCES

1. Giudici, R.; Nascimento, C. A. O.; Beiler, I. C.; Scherbakoff, N. J Appl Polym Sci 1998, 67, 1573.
2. Tzoganakis, C. Adv Polym Technol 1989, 9, 321.
3. Berghauss, U.; Michaeli, W. Kunststoffe 1991, 81, 479.
4. Xanthos, M., Ed. Reactive Extrusion: Principles and Practice; Hansen Publishers: Munich, 1992.
5. Bhat, N. V.; McAvoy, T. Comput Chem Eng 1990, 14, 573.
6. Hoskins, J. C.; Himmelblau, D. M. Comput Chem Eng 1988, 12, 881.
7. Cybenko, G. Mathematics Control, Signals, Systems 1989, 2, 303.
8. Chan, W. M.; Nascimento, C. A. O. J Appl Polym Sci 1994, 53, 1277.
9. Nascimento, C. A. O.; Giudici, R.; Scherbakoff, N. in Proceedings 5th World Congress of Chemical Engineering, San Diego, CA, IV; 1996, p 515.
10. Pollard, J. F.; Broussard, M. R.; Garrison, D. B.; San, K. Y. Comput Chem Eng 1992, 16, 253.