NEURAL IDENTIFICATION OF SUPERCRITICAL EXTRACTION PROCESS WITH FEW EXPERIMENTAL DATA

Rosana P.O. Soares*, Roberto Limão de Oliveira*, Vladimiro Miranda and José A. L. Barreiros***

** Electrical and Computer Engineering Department, Federal University of Pará -UFPA*
Campus Universitário do Guamá, CP 8619, CEP: 66075-900, Belém - PA, Brazil
Phone Number: +55 91 3183 1680 Fax: +55 91 3183 1634
***INESC Porto, Institute of Engineering in Systems and Computers of Porto, and FEUP, Faculty of Engineering of the University of Porto, Portugal*
Rua Dr. Roberto Frias, 378 4200-465 Porto, Portugal.
E-mails: rsoares@ufpa.br, limao@ufpa.br, vmiranda@inescporto.pt, barreiro@ufpa.br,

Abstract: In this paper, an Artificial Neural Network (ANN) is designed to model an extraction process that uses a supercritical fluid as solvent. Two approaches are used in the ANN training. They differ in the strategy used to complement the experimental data collected during extraction procedures of useful compositions for the pharmaceutical industry. While the first method involves fitting of data obtained during an operation of extraction, the second one uses pseudo experiments generated from real data and incorporating process qualitative characteristics.

Keywords: Training, Neural Network, Identification, Modeling, Supercritical Extraction Process

1. INTRODUCTION

The extraction processes that use supercritical fluids as solvent stand out among the techniques of extraction of active components of natural substratum due to their capacity to produce products free of solvent. The modeling of such a process used in the extraction of natural products of a great variety of plants has been receiving a lot of attention from researchers. Most of the time, simple numerical models of mass transfer are proposed and used in the project of supercritical extraction installations and in the identification of their best operation conditions. These types of models are based on phenomenological and simplified description of the process, using rigid restrictions (Reverchon, 1997). In Fonseca et al. (1999), the results of a hybrid neural model of the curves of the supercritical extraction of two Brazilian vegetal matrices were presented. In this approach, an artificial neural net is used to identify parameters for a phenomenological model, minimizing some restrictions of neural nets. Fonseca, et al. (2000) had also used neural nets to derive the mass transfer coefficients of a supercritical extraction. To avoid the difficulties associated with a reduced amount of experimental supercritical extraction system data, they used a particular technique to generate new semi-empiric data. This technique combines the experimental coefficients of mass transfer with those obtained from the available correlation in the literature, producing an amount of data enough for efficient neural identification.

The models considered in the present paper can predict the extraction curves in relation to the profile of solvent mass used for an operation point. The models are based on the application of Artificial Neural Networks (ANN) and inexperimental data collected in laboratory during extraction procedures of useful compositions for the pharmaceutical industry using residues (RAN–Residue Black Agglomerate) originating from cork as raw material. To correlate solvent mass of and raw material with the extracted mass of product, we used multiplayer

feedforward ANN designs. The training of ANN was achieved using MATLAB with the optimization algorithm of Levenberg-Marquardt and Bayesian regularization, where the values of weights and bias of the ANN are updated to minimize a linear combination of the square of the sum of the deviations between the output training pattern and the output of the ANN in training.
Due to the small amount of experimental data and the high cost involved in the execution of experiments, two different strategies were tested in order to create a larger set of data enabling an efficient ANN training. The first strategy consisted in fitting a curve to the data sampled from each extraction operation to generate a larger number of samples. It was assumed that even though the information had not been registered due to the characteristics of the extraction process, it existed in fact and, therefore, it could be estimated using a curve fitting method.

The second strategy consisted of adding pseudo experiments to the set of training, which had been generated from the real experiments using qualitative characteristics of the process. Its use, besides generating enough data for the training, also allows a priori knowledge inclusion in the ANN through the qualitative information of the process incorporated in the training patterns (Thompson et al., 1994; Niyogi et al., 1998).

2. PROCESS DESCRIPTION

The supercritical extraction processes have the capacity to produce products free from solvents and to extract or to fraction mixtures in soft conditions. In a general way, they are constituted by stages of:

- compression;
- heating or condensation;
- extraction;
- separation;
- and regeneration of the solvent used .

In a typical operation of supercritical extraction, the raw material is initially placed inside the extractor, being continuously in contact with the supercritical fluid. This fluid can dissolve chemical substances similarly to a liquid and penetrate in porous matrices similarly to a gas. Figure 1 and Table 1, show the main components of a typical pilot installation of extraction supercritical.

The solvent most commonly used is CO_2. In the extraction equipment, it is submitted to a high pressure, under the action of a compressor that leads it to operate in supercritical phase. In the extractor, when the carbon dioxide passes through the raw material, substances are dissolved and extracted until a level of balance solubility. The gaseous solution, when leaving the extractor, passes through an expansion valve and is submitted to a lower pressure, causing the precipitation of the components in the separator. In this way, the extracted substances of the raw material are separated from the carbon dioxide, which is recycled by the compressor. Such recurrent process is repeated until all components are extracted and collected in the separator. The characteristics of the recycling such as: temperature, flow, pressure and duration are calculated and adjusted in a way to maximize the extraction. These adjustments, in general, depend on the product and the components desired to be extracted.

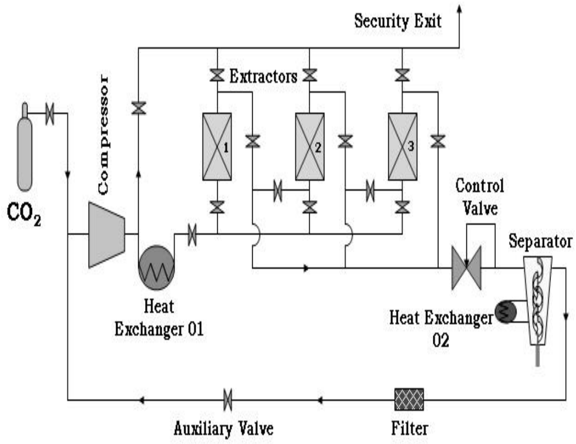

Fig. 1. Supercritical extraction installation scheme.

Table 1. Basic Components of the Pilot Installation

EQUIPMENT	TASK
COMPRESSOR	To elevate the pressure so that CO_2 operate in the supercritical state.
HEAT EXCHANGER 01	To stabilize the temperature on the side of high pressure (extraction stage).
HEAT EXCHANGER 02	To guarantee the stabilized temperature in the cyclone (separator) avoiding that the extreme cooling happens.
EXTRACTOR	Place where the raw material is submitted to the continuous contact with supercritical fluid
EXPANSION VALVE (CONTROL VALVE)	To reduce the pressure of the CO2+ Extract to control the rate of flow of solvent mass and to cause the precipitation of the components in the cyclone.
SEPARATOR (CYCLONE)	Place where the precipitation of the extracted components occurs

3. NEURAL NETWORK MODELS

Typically, ANN are constituted by artificial neurons connected so that the information along the network may be processed in a simultaneous and parallel way. A diagram that represents the model for an artificial neuron is shown in the Figure 2, where: $X_1, X_2. . . , X_n$ represent the inputs of the neurons; $w_1, w_2. . . , w_n$ represent the weights for each input and b represents the bias for each neuron. In a static neuron the output S is given by the linear combination of its inputs as shown in (1). The weighted sum S is an input to f (.), denominated activation function.

$$S = \omega_1.x_1 + \omega_2.x_2 + \cdots + \omega_n.x_n + b \qquad (1)$$

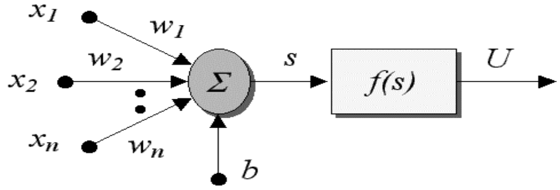

Fig. 2. Artificial Neuron

Multilayer Feedforward Neural Networks are well known since the early 90's to be successful in identification and control systems applications (Hunt, 1992). Their architecture is typically composed of a set of input patterns, one or more hidden layers and one output layer.

In our work, we used nets with one or two hidden layers. We have confirmed that a model of a supercritical extraction plant can be represented by a ANN if a correct training is accomplished. The algorithm used by us for ANN training is based on the method of Levenberg-Marquardt (Hagan, Menhaj, 1994), with Bayesian regularization (Dan, Hagan, 1997), whose main characteristics are the reduced computational effort during the training and producing ANN with good generalization qualities. In the ANN training, a typical fitting function is obtained by the average of the quadratic sum of the ANN errors, according to Equation 2.

$$F = mse = \frac{1}{N}.\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}.\sum_{i=1}^{N}(t_i - a_i)^2 \quad (2)$$

The Levenberg-Marquardt algorithm was created to obtain a solution that minimizes the function F without the need of calculating the Hessian matrix. In particular, when the fitting function has the form of the Equation 2, the Hessian matrix can be approximate to:

$$H = J^T.J \quad (3)$$

and the gradient can be calculated for:

$$g = J^T.e \quad (4)$$

where

J – Jacobian matrix that contains the first derivative of the errors of the ANN in relation to the weights and bias, and

e – error vector of the ANN output

The Jacobian Matrix can be calculated using the error backpropagation method, which initially calculates the errors derived (delta vector) for the output layer, which are then backpropagated through the ANN until a delta vector for all the layers of the net has been calculated . The variables (weights and bias) are continuously updated by the following equation:

$$X_{K+1} = X_K - [J^T.J + \mu.I]J^T.e \quad (5)$$

To improve the generalization capacity of the ANN, a procedure denominated regularization is used. Such procedure involves the modification of the fitting function, according to the equation 6.

$$msereg = \gamma.mse + (1-\gamma).msw \quad (6)$$

where γ is the fitting rate and msw is given by:

$$msw = \frac{1}{n}\sum_{j=1}^{n} w^2{}_j \quad (7)$$

The difficulty in regularization is to determine the optimal value of γ. However, using Bayesian Regularization (Dan and Hagan, 1997), it is possible to determine this parameter automatically using statistical techniques.

4. NEURAL MODELS FOR THE EXTRACTION CURVE

4.1 General Considerations

This section presents two methodologies to build neural models that may predict extraction curves, for a supercritical extraction process operation condition, in relation to the solvent mass profile used.

The models have been obtained with experimental data collected during experiments performed in a pilot installation at the Institute of Experimental and Technological Biology - IBET -located in Oeiras (Lisbon, Portugal), with the extraction of useful substances for pharmaceutical industry from residues originating from cork production (RAN - Residue Black Agglomerate). However, because the experimental data collected were scarce, due to the nature of the process, the protocol involved and the high cost involved in experiments execution, there was an insufficient amount of data for ANN training. Therefore, two strategies for data complementation have been tested.

The general operation conditions of the pilot installation, for the extraction of the pharmaceutical product whose composition cannot be disclosed for confidentiality reasons, had been previously identified in research accomplished at IBET and the results of those research efforts suggested the use of an extraction autoclave of 2,0 liters, initial mass of RAN around 400g, and temperature and pressure references of 40°C and 250bar. Analyzing the pilot installation operation during the extraction, it was observed that the CO_2 temperature that is controlled along the process did not present significant variations during the experiments. The ON-OFF action of the expansion valve, combined with the compressor action, controlled the pressure in the supercritical cycle. However, the CO_2 solvent mass during the extraction process depends strongly on the heating of CO_2 reservoirs (especially if they are subject to solar radiation) and on the amount of CO_2 accumulated in them, influencing the amount of product extracted. Trusting the experience of IBET researchers, we opted to neglect the perturbations caused by temperature and pressure and to use the structure of the model shown in the Figure 3.

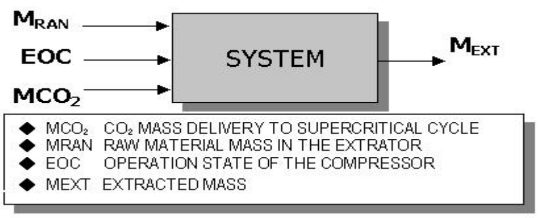

Fig. 3. Model Structure

4.2 Fitting Data

The strategy of fitting data consists of fitting a curve to the information collected during each extraction operation, with the purpose of having a larger number of samples. Samples were registered in the instants when the product was collected in the output of the separator (in intervals of approximately 60 minutes). From curve fitting, new points were derived for 10-minute intervals. Therefore, from experiments with just 3 samples we built a set of 18 samples. To fit the data we used the *interp1* function of MATLAB. With the interpolated data, the best training results were obtained with an ANN architecture of 3 hidden layers, with 11, 50 and 03 neurons, and sigmoid activation functions. In the output layer we used the linear activation function usual in Neural Identification applications. The training patterns were linearly normalized in the interval [0,1] and the internal structure of the supercritical extraction model shown in Figure 4 was adopted to characterize the inputs and outputs of the ANN model.

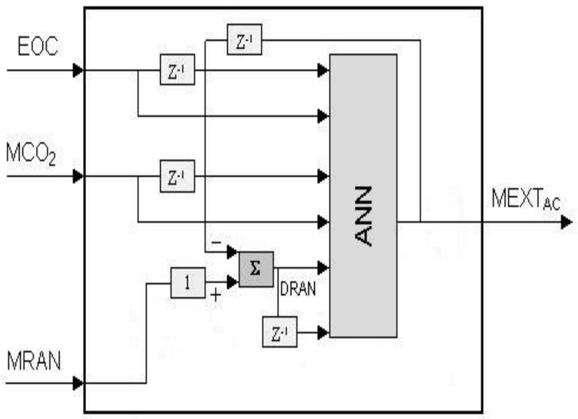

Fig. 4. Structure of the Supercritical Extraction Model

4.3 Building Pseudo Data

The pseudo data strategy consisted of generating additional data that could incorporate in the training patterns some qualitative information of the process. This was made taking in account an intrinsic characteristic of the process of extracting product; in fact, some product may be retained in the extraction ducts, during a partial extraction operation, and be collected in a subsequent stage together with the mass accumulated in the corresponding period. So, thinking about a training model for the ANN that could also take in consideration that characteristic, pseudo experiments were added to the training set. The approach adopted to generate the new patterns was based on fictitious measures of extracted mass obtained from the real measures with increments of ±DM%. For each experiment a random value DM was used, in the band suggested by the operator of the installation (from 5 to 15).

The pseudo experiments associated with each real experiment were generated maintaining the inputs and combining the new measures of extracted mass to compose new experiments. For instance, for an experiment, we could obtain 27 (3^3) possible combinations that corresponded to 26 pseudo experiments and 1 real experiment. For a total of 37 real experiments, 999 experiments were obtained (27x37) where 962 were pseudo experiments and 37 real experiments. Some possible combinations are shown in the Table 2.

With the pseudo data, the best training results were obtained with 3 hidden layers, with 7, 15 and 5 neurons, and sigmoid activation functions. In the output layer we used linear activation. The training patterns were linearly normalized in the interval of [0,1] and the internal structure of the supercritical extraction neural model, shown in the Figure 5, was adopted to characterize the inputs and outputs of the ANN.

Table 2. Combinations for Composition of Pseudo experiments

TYPE EXPERIMENT	MEASURE 01	MEASURE 02	MEASURE 03
REAL	MR01	MR02	MR03
PSEUDO	MR01	MR02	MR03+
PSEUDO	MR01	MR02	MR03-
PSEUDO	MR01	MR02+	MR03
PSEUDO	MR01	MR02+	MR03+
PSEUDO	MR01	MR02+	MR03-
PSEUDO	MR01	MR02-	MR03
PSEUDO	MR01	MR02-	MR03+
PSEUDO	MR01	MR02-	MR03-

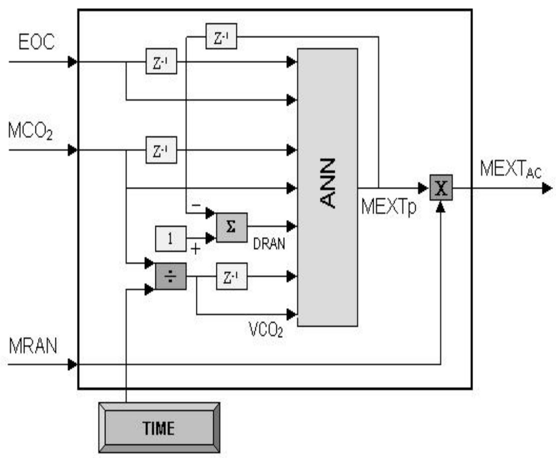

Fig. 5. Structure of the Supercritical Extraction Model

5. RESULTS

The raw data were obtained form 37 real experiments at IBET lab facility. In general, the results reached with the two strategies presented outputs with non-significant errors between the output of the ANN and the experimental data, namely in the experiments that were not part of the training set and that formed the test set. However, different problems were identified in each of the models, which will be mentioned in the following sections.

5.1 Neural Model with Fitted Data.

For this model, coherence loss was observed in some experiments, because the accumulated extracted mass estimated by the ANN assumed smaller values than the previous sample. Also, the first samples of the experiment also in some cases were very different from the real data – this is certainly due to tail effects of the fitting curve. However, the main advantage in using this strategy is to obtain a not too sensitive resulting model to deviations of the mass solvent profile during an extraction. The results of the simulations made with the neural model obtained from the interpolated data are shown in Figures 6-8.

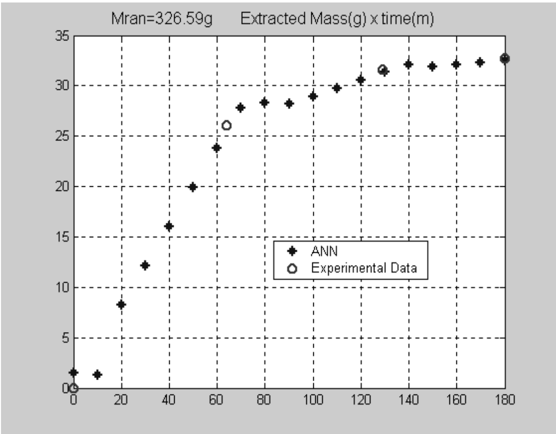

Fig. 6. Training Experiment Type

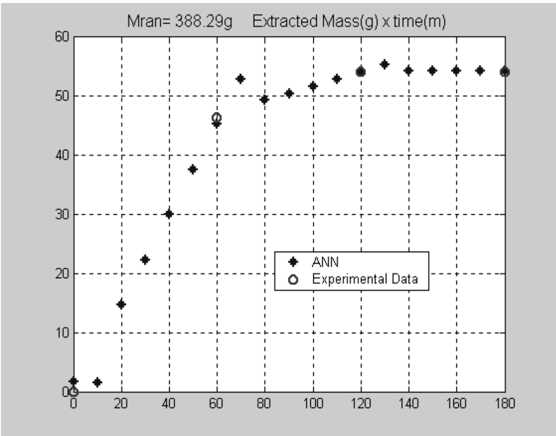

Fig. 7. Validation Experiment Type

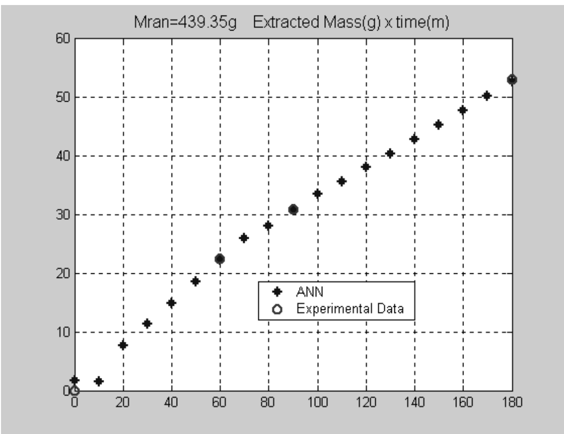

Fig. 8. Test Experiment Type

5.2 Neural Model with Pseudo Data.

For this model, the largest errors were around 2g to 3g. These errors were found in only a small number of experiments and in only one of the measurements taken. In spite of having presented better precision than the previous neural model, it showed as disadvantage a high sensitivity to deviations of the mass solvent profile during an extraction. Table 3 presents the numeric results of the simulations performed with the neural model obtained from the pseudo data.

Table3. Numeric Results

TYPE EXPERIMENT	TRAIN	VALIDATION	TEST
MRAN (g)	326.59	388.29	439.35
TIME (min)	[64 129 180]	[60 120 180]	[60 90 180]
PROFILE MCO_2 (g)	6786 13727 19410	6881 13278 19801	4198 6349 13293
EXTRACTED MASS (ANN) (g)	26.00 31.57 32.60	46.29 53.94 54.04	20.83 29.25 51.48
EXTRACTED MASS (REAL DATA) (g)	26.02 31.59 32.72	46.30 53.94 53.94	22.29 30.71 52.94

6. CONCLUSIONS

The application of ANN in the identification of a supercritical extraction process is suggested as an efficient alternative considering that the phenomenological model and the industrial operative process are complex. The work was only possible because we had access to the installation pilot of IBET, which allowed us to get and register experimental data for ANN training.

The difficulty of ANN training with small amount of available experimental data, which is common in supercritical extraction processes, was minimized through two different data complementation strategies. Although the data curve fitting during each experiment has produced satisfactory results for ANN training and in the simulation of the process, its use does not have a formal justification and, therefore, it is not possible to conclude about its success in other situations or operation conditions.

The pseudo experiments were essential in the ANN training making available a larger number of data points, besides allowing to consider the characteristics of the product collection process in the data. If the imprecision in collecting product is not taken in account, this may cause error in the measuring of the extracted mass. The high degree of non linearity of the process and the limited spectrum of the collected data restricted the efficiency of the model obtained with the pseudo experiments turning it too sensitive to large deviations from the profile of mass solvent during an extraction. However, in the band of operation considered in this study, it presented good characteristics of precision and generalization capacity, becoming very important in the identification of strategies of manipulation of the CO_2 mass in the supercritical cycle to maximize the product to be extracted.

The work reported cannot be considered completed but the results so far obtained are promising and serve as inspiration to the development of more accurate ANN models.

An ANN model of a supercritical extraction plant will have both an academic interest, because it will serve to validate analytical models, and an industrial interest, because it may serve to help operator in predicting the outcome of operations. In further studies, ANN such as these will help in building up optimized control strategies and processes for industrial supercritical extraction plants.

ACKNOWLEDGEMENTS

The authors would like to thank the CAPES for the support in the development of this research and the Institute of Experimental and Technological Biology (IBET), in the person of Manoel Nunes Da Ponte, for allowing access to the laboratory and the use of experimental data in this work

REFERENCES

Dan Foresee, F.; Hagan, M.T. (1997). *Gauss - Newton approximation to Bayesian learning*. International Conference on Neural Networks vol 3, 9-12. pp. 1930 –1935.

Fonseca, A. P., Stuart, G., Oliveira, J. V. et al. (1999). *Using hybrid neural models to describe supercritical fluid extraction processes*. Braz. J. Chem. Eng., vol.16, no.3, p.267-278. ISSN 0104-6632.

Fonseca, A.P., Oliveira, J.V. e Lima, E.L. (2000). *Neural networks for predicting mass transfer parameters in supercritical extraction*. Braz. J. Chem. Eng., vol.17, no. 4-7, p.517-524. ISSN 0104-6632.

Hagan, M.T.; Menhaj, M.B. (1994). *Training feedforward networks with the Marquardt algorithm*. IEEE Transactions on Neural Networks, Volume: 5 Issue: 6, pp. 989 – 993.

Hunt, K. J.; Sbarbaro, D.; Gawthrop, P. J. (1992). *Neural Network for Control Systems – A survey in Automatica*, v.28, n.6, pp. 1083-1112.

Niyogi, P. F; Girosi, and Poggio, T. (1998). *Incorporating prior information in machine learning by creating virtual examples*. Proceedings of the IEEE., 86(11), pp.2196--2209.

Reverchon, E. (1997). *Supercritical Fluid Extraction and Fractionation of Essential Oils and Related Products*, J. Supercrit. Fluids, 10, 1

Thompson, M. L. and Kramer, M. A. (1994). *Modelling chemical processes using prior knowledge and neural networks*. AIChE Journal, 40(8): 1328-1340.