PERGAMON
www.elsevier.com/locate/watres

PII: S0043-1354(00)00217-7

Wat. Res. Vol. 34, No. 18, pp. 4371–4380, 2000

Printed in Great Britain
0043-1354/00/$ - see front matter

NEURAL NETWORK MODELS FOR ULTRAFILTRATION AND BACKWASHING

CARMEN TEODOSIU[1]*M, OCTAVIAN PASTRAVANU[2] and MATEI MACOVEANU[1]

[1]Department of Environmental Engineering, Faculty of Industrial Chemistry, Bd. D. Mangeron 71 A, 6600, Iasi, Romania and [2]Faculty of Automatic Control and Computer Engineering, Bd. D. Mangeron 53 A, 6600, Iasi, Romania

(First received 3 May 1999; accepted 15 March 2000)

Abstract—The possibility of predicting the time dependence of flux evolution for ultrafiltration systems is one of the most important factor that influences large-scale applications, optimization and automation, because flux affects productivity and thus all cost components. This paper presents the development and exploitation of mathematical models based on artificial neural networks, trained with experimental data obtained in a laboratory scale ultrafiltration system. Hollow-fibre ultrafiltration membranes and dead-end mode of operation were used for secondary refinery effluent treatment, at constant transmembrane pressure and fixed loss of the initial flux. For the training procedure of neural networks, ultrafiltration tests followed by backwashing with demineralized water, performed in the same operating conditions have been selected. Two neural network models were constructed to predict the flux at any time instant during ultrafiltration and after backwashing for arbitrary cycles, within the ultrafiltration-backwashing process. The trained networks are able to accurately capture the non-linear dynamics for initial fluxes in the range 80–145 l h^{-1} m^{-2} and for time horizons up to 2500 s.

Key words—ultrafiltration, fouling, wastewater recycling, modelling, neural networks, flux prediction

INTRODUCTION

Water and wastewater treatment using membrane processes found many pilot or full-scale applications due to a series of objective factors, in close connection either with the stringent standards for drinking water production and wastewater disposal, or with the need to use alternative sources and recycle industrial wastewater. All these factors and the development of membrane industry, including reduction of costs and energy consumption contributed to an increased number of applications in water and wastewater treatment processes (Mallevialle *et al.*, 1996).

There are clear advantages of membrane technologies, such as: physical separation of a broad range of contaminants at ambient temperature, easiness of operation and adaptation to existing units, continuous processing which can be automated, and low requirement of chemicals. The major drawbacks of these processes are concentration-polarization and fouling/scaling. These phenomena refer to the accumulation and deposition of submicron particles, solutes, microorganisms on the membrane surface and into the pores, thus contributing to *permeate flux decline* and decreasing *solute rejection*.

For ultrafiltration systems operated either in cross-flow or dead-end mode, using as influents natural waters or wastewater, fouling was found to be the most important factor influencing flux and productivity (Wetterau *et al.*, 1996; Teodosiu, 1996; Fratila, 1997), operational costs and membrane replacement (Erricson and Hallmans, 1994; Manth *et al.*, 1998). Membrane fouling can be defined from an *operational* point of view as *reversible* and *irreversible*, after the character of the deposit formed (temporary or permanent), and the possibilities of restoring initial flux by backwashing or chemical cleaning. Considering the *nature of the deposited material*, a distinction can be made between *organic, inorganic, particulate* and *biological* fouling, and in the case of wastewater, at least two or more of these foulants contribute to membrane fouling. The most important factors that influence fouling and its control are related to: the composition of feed water, membrane nature and surface properties, hydrodynamic and operating process conditions (Cheryan, 1998). These factors influence also fouling mechanisms, i.e. deposition on mem-

*Author to whom all correspondence should be addressed. Tel.: +40-32-237-594; fax: +40-32-237-594; e-mail: cteodosi@tuiasi.ro

brane surface or into the pores, adsorption, pore blocking and permeate flux variation.

Membrane productivity, operational costs and membrane replacement are also affected by fouling, due to the gradual decrease of ultrafiltration flux with time, even if the major operating parameters (pressure, flow rate, feed concentration, temperature) are kept constant. Frequent backwashing, chemical cleaning or an increased number of pretreatment techniques are required for its prevention, generating thus supplementary costs in comparison with other competing technologies.

Prediction of flux decline for ultrafiltration and generally all membrane applications is very important both for the assessment of overall productivity, optimization, automation and for scaling up these systems. Most of the mathematical models developed for ultrafiltration flux are based on the similarities that exist between pressure-driven membrane processes and normal filtration, or the "resistance in series" concept. Thus, the permeate flux can be calculated with the general relation (Jones *et al.*, 1993):

$$J = \frac{\Delta p}{\eta \cdot (R_m + R_c + R_i)} = \frac{\Delta p}{\eta \cdot R_t} \qquad (1)$$

where J is the membrane flux ($m^3\ m^{-2}\ s^{-1}$), Δp the transmembrane pressure (Pa), η the fluid viscosity (Nsm^{-2}), R_m the hydraulic resistance of the clean membrane (m^{-1}), R_c the resistance due to a cake layer or concentration-polarization (m^{-1}), R_i the resistance caused by irreversible fouling due to blocking filtration or adsorption (m^{-1}), and R_t the total hydraulic resistance (m^{-1}).

For long term flux decline, the irreversible and cake resistance become *time dependent*, the net effect being an increase of the total hydraulic resistance and subsequently a decrease of permeate flux during normal operation.

In order to use the "resistance in series" models for flux prediction it is absolutely necessary to calculate membrane resistances for a given set of operating conditions. These calculations can be made using either empirical equations or curve fitting estimations that have been obtained when water with a consistent quality (controlled number and concentration of pollutants) is used as influent for ultrafiltration. Many of these models are non-linear, due to the need to improve the accuracy in capturing the time dependence of flux (Fane and Fell, 1987; Jones *et al.*, 1993; Wetterau *et al.*, 1996). Although many studies of fouling on ultrafiltration membranes have been done, there exists no analytical or empirical model devised for the general case, that might be particularized for wastewater by taking into account influent properties and membrane properties.

A recent study (Delgrange *et al.*, 1998) approaches the non-linearity of membrane resistance prediction only at the *beginning* and at the *end* of ultrafiltration and backwashing cycles using neural network models. The use of such a model allows the incorporation of water quality parameters (turbidity, temperature) and operating parameters (permeate flow) in cross flow operation of a large-scale ultrafiltration system for surface water. However, the prediction of hydraulic resistances (and indirectly of flux), is defined with regard to a certain number of cycles, with a fixed time duration per cycle, no information being available in between the sampling instants.

Our study is focused on the development of two neural network models able to predict the flux *at any time instant* during ultrafiltration and after backwashing, for arbitrary cycles, regardless the duration of ultrafiltration and backwashing, considering wastewater as influent and a hollow fibre membrane. This "black-box" type of model is developed on the basis of an experimental system that can produce reliable data, without needing any other theoretical knowledge of process mechanisms or the relationships that describe it. This modelling approach is motivated by the optimization problems, which require a thorough prediction of the *continuous-time evolution of flux*. The advantage of such a model results from the fact that, using a well-established set of experimental conditions, and information about time and initial fluxes, an accurate prediction of flux evolution for ultrafiltration and backwashing can be formulated. Thus, the major drawbacks and limits of the linear piecewise model developed in our previous work for blocking and cake filtration regions (Teodosiu, 1998; Teodosiu and Macoveanu, 1998) can be avoided. Actually, none of these models were accurate enough, because the phenomena representing the objective of model construction are essentially non-linear.

NEURAL NETWORK USAGE IN MODEL CONSTRUCTION

Multilayer feedforward neural networks offer a generous framework for modelling *non-linear* phenomena whenever (due to various reasons) the physical insight fails in providing relevant information for the construction of a parameterized model, whose parameters play precise roles in capturing the essentials of studied behaviour. The neural network operates as a non-linear mapping:

$$\mathbf{f}: \mathscr{R}^m \rightarrow \mathscr{R}^q \qquad (2)$$

parameterized by the weights and biases of its layers, which can be adjusted so as to fit experimental data, but without any physical meaning for the identified parameters. In other words, such a model is exploitable as a global entity, operating in input-output terms and any effort is pointless to reveal separate interpretation for its parameters.

The neural network architecture which is most frequently used in data fitting and non-linear approximation consists of three layers depicted as a block diagram in Fig. 1. Such a graphical representation is analogue to those in Freeman and Skapura (1991) and Hornik *et al.* (1989), but present the advantage of a simplified matrix–vector notation.

Consequently, the input–output transfer can be described as a composite mapping:

$$\mathbf{y} = f(\mathbf{x}) = \boldsymbol{F}_3(\boldsymbol{W}_3^*\boldsymbol{F}_2(\boldsymbol{W}_2^*\boldsymbol{F}_1(\boldsymbol{W}_1^*\mathbf{x}+\boldsymbol{b}_1)+\boldsymbol{b}_2)+\boldsymbol{b}_3) \quad (3)$$

where $\mathbf{W}_j$, $\mathbf{b}_j$ and $\mathbf{F}_j$ denote the weight matrix, bias vector and vector activation function corresponding to layer j ($j = 1, 2, 3$). Within a given layer j, the scalar functions $\mathbf{F}_j^i(\boldsymbol{u}_j^i)$ defining the vector activation function $\boldsymbol{F}_j$ are identical.

Our further discussion refers to the case of linear functions in the third layer, i.e. $\boldsymbol{F}_3^i(\boldsymbol{u}_3^i) = \boldsymbol{u}_3^i$, and tansigmoid functions in the first and second layer, i.e.:

$$\boldsymbol{F}_j^i(\mathbf{u}_j^i) = (\exp(\boldsymbol{u}_j^i) - \exp(-\boldsymbol{u}_j^i))/(\exp(\boldsymbol{u}_j^i) + \exp(-\boldsymbol{u}_j^i)),$$

$$j = 1, 2.$$

All the points of this discussion preserve their validity for logsigmoid functions instead of the tansigmoid ones. The number of neurons in the input and hidden layer, i.e. n and p are chosen by the designer, whereas m and q are constrained by the data to be fitted.

Such a network possesses an extremely useful property of approximation, as initially proven in (Cybenko, 1989) for the more restrictive case of two layer networks with sigmoid functions in the first layer and linear functions in the second layer. In our situation, if $\varphi:\mathscr{R}^m \to \mathscr{R}^q$ is a smooth function on a compact set $S \subset \mathscr{R}^m$, the approximation property can be stated as follows:

$$\forall \epsilon > 0,\ \exists\ n, p, \boldsymbol{W}_j, \boldsymbol{b}_j, j = 1, 2, 3;$$
$$\forall \mathbf{x} \in S \|\boldsymbol{f}(\mathbf{x}) - \varphi(\mathbf{x})\| < \epsilon \quad (4)$$

where $\| \ \|$ denotes any vector norm in $\mathscr{R}^p$. If the function φ is known in a finite number of points $(\mathbf{x}_i, \mathbf{z}_i)$, $i = 1, \ldots, N$, $\mathbf{x}_i \in \mathscr{R}^m$, $\mathbf{z}_i \in \mathscr{R}^q$ (assumption corresponding to the knowledge of φ by means of a set of experimental data), then the computation of network parameters (entries of matrices $\boldsymbol{W}_j$ and vectors $\boldsymbol{b}_j$, $j = 1, 2, 3$) is addressed as a minimization problem for the objective function:

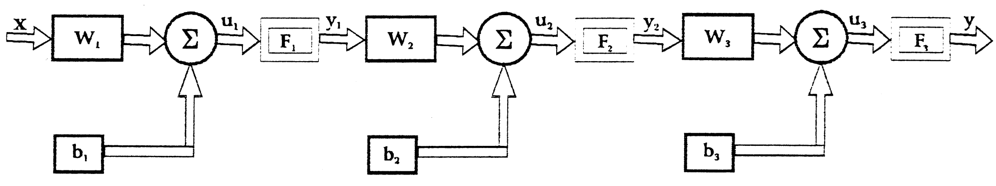
$$\mathrm{E}(\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{W}_3, \boldsymbol{b}_1, \boldsymbol{b}_2, \mathbf{b}_3) = \frac{1}{2}\sum_{i=1}^{N}\|\mathbf{f}(\mathbf{x}_i) - \mathbf{z}_i\|_2^2 \quad (5)$$

where each $\boldsymbol{f}(\boldsymbol{x}_i)$ for $i = 1, \ldots, N$, depends on $\boldsymbol{W}_j$, $\boldsymbol{b}_j$, $j = 1, 2, 3$, according to equation (3).

The particular expression of $\boldsymbol{f}$ in equation (3) induces matrix forms for the optimization algorithms based on equation (5), which, in the specific language of neural networks are referred to as "training (learning) procedures" (Haykin, 1994). Also in this language, vectors $\boldsymbol{x}_i$, $\boldsymbol{z}_i$, $i = 1, \ldots, N$, are called patterns, and targets, respectively.

MATERIALS AND METHODS

The experimental equipment has been thoroughly described in a previous paper (Teodosiu *et al.*, 1999). The experimental set-up consists of a system for creating, control and monitoring of pressure, storage tanks for wastewater and demineralized water for backwashing, ultrafiltration module, and a flowmeter connected with a computer for automatic data registering (flow, time and cumulative permeate volume were determined each 30 s for all the experiments). Two hollow fibre membrane modules (named A and A-LF), with a molecular weight cut-off of 150,000 Dalton, having 50 fibres, internal fibre diameter of 1.5×10^{-3} m, and membrane area 0.1 m^2. These membranes are made of *polyethersulphone* and *polyvinylpirrolidone*, the only difference between them being a special polymer coating (LF), for the low fouling module. In order to obtain a database for training the neural network and for testing the validity of this model, only the data



Input (first) layer

$\boldsymbol{W}_1 \in \Re^{n \times m}, \boldsymbol{b}_1 \in \Re^n,$
$\boldsymbol{x} \in \Re^m, \boldsymbol{u}_1 \in \Re^n, \boldsymbol{y}_1 \in \Re^n$
$\boldsymbol{y}_1^i = \boldsymbol{F}_1^i(\boldsymbol{u}_1^i), i = 1, \ldots, n$

Hidden (second) layer

$\boldsymbol{W}_2 \in \Re^{p \times n}, \boldsymbol{b}_2 \in \Re^p,$
$\boldsymbol{u}_2 \in \Re^p, \boldsymbol{y}_2 \in \Re^p$
$\boldsymbol{y}_2^i = \boldsymbol{F}_2^i(\boldsymbol{u}_2^i), i = 1, \ldots, p$

Output (third) layer

$\boldsymbol{W}_3 \in \Re^{q \times p}, \boldsymbol{b}_3 \in \Re^q,$
$\boldsymbol{u}_3 \in \Re^q, \boldsymbol{y} \in \Re^q$
$\boldsymbol{y}^i = \boldsymbol{F}_3^i(\boldsymbol{u}_3^i), i = 1, \ldots, q$

Fig. 1. Architecture of a three layer feed forward neural network with m inputs, q outputs, n nodes (neurons) in the input layer and p nodes (neurons) in the hidden layer.

obtained with membrane A-LF were used, since this membrane had better performances in terms of flux recovery, operating time and removal efficiencies (Teodosiu *et al.*, 1999). The wastewater originates from a refinery, being subjected to a treatment scheme involving primary, intermediate treatment, secondary treatment (activated sludge) and further tested by ultrafiltration in order to evaluate the possibilities to recycle it as cooling water make-up (Teodosiu, 1996).

Ultrafiltration (fouling) experiments were performed using the dead-end flow mode of operation, with a transmembrane pressure of 1 bar, and a fixed loss of the initial flux (20%) for membrane A-LF, while backwashing with demineralized water under 2 bar was made for short intervals of 1 min (database for training and testing the neural network model). Dead-end mode is recommended for wastewater operation, due to lower capital and operating costs (Blume *et al.*, 1995).

For membrane A, short ultrafiltration experiments (20% loss of the initial flux), at a transmembrane pressure of 0.3 bar (that gave similar fluxes with those obtained for the other module) were achieved. Backwashing was made with demineralized water at 0.6 bar during 1 min (data for testing the model).

All experiments that have been used in this study involve only cycles of ultrafiltration and backwashing, without any chemical cleaning experiments. In this text, experiments are going to be referred to as "tests", preserving the original labelling considered in the previous work (Teodosiu, 1996).

Experiments were performed with the same batch of wastewater, with characteristics presented in a previous study (Teodosiu *et al.*, 1999), at temperatures varying in the interval (15–18)°C. In order to eliminate the influence of temperature variation, permeate fluxes were calculated and corrected for 20°C using equation (6), and viscosities were determined with equation (7) (Koprowski, 1995):

$$J = \frac{\eta_T}{\eta_{20}} \cdot \frac{Q_{avg}}{A} \tag{6}$$

$$\eta_T = \frac{\eta_0}{(1 + 0.0337 \cdot T + 0.000221 \cdot T^2)} \tag{7}$$

where J is the flux ($lh^{-1}\ m^{-2}$), η_T and η_{20} the water viscosity at temperature T and 20°C (Nsm^{-2}), Q_{avg} the average flow registered by the flowmeter (lh^{-1}), T the water temperature (°C), and A the membrane area (m^2).

Hydraulic resistance was calculated using clean water flux for the new membrane, at an operating pressure of 1 bar, with correction of flux and viscosity for a temperature of 20°C. For membrane A-LF, $R_m = 2.27 \times 10^{12}\ m^{-1}$.

RESULTS AND DISCUSSION

Two neural network models were constructed to predict the flux at *any time instant* during ultrafiltration and after backwashing for arbitrary cycles, regardless the duration of ultrafiltration and backwashing. Prediction of flux is based on sets of data obtained with membrane A-LF, for complete cycles ultrafiltration- backwashing. The following aspects are relevant for model development relying on artificial neural networks.

Ultrafiltration

- The initial permeate flux, calculated at 20°C can be considered for computation and representation of the total membrane resistance at the beginning of ultrafiltration tests. Such an approximation was made considering results of the analysis of clean water flux values (CWF—calculated after backwashing, before starting an ultrafiltration tests) and of initial permeate flux values, and is presented in Table 1; moreover, in practice, CWF measurement can be used to evaluate membrane permeability after chemical cleaning, applied after a certain number of ultrafiltration backwashing cycles has been already performed.
- It is considered that for short ultrafiltration tests, both mechanisms of blocking filtration and cake filtration are possible, previous experiments showing that the exact point of separation between these two mechanisms is practically impossible (Teodosiu, 1996).
- Influent quality was constant during all experiments.

Backwashing

- It is considered that each ultrafiltration test is followed by backwashing, which is achieved with the same pressure and an arbitrary time duration.
- For each backwashing procedure, an association can be made between the final permeate flux at the end of test $k-1$ and the initial permeate flux at the beginning of test k (after backwashing). Backwashing affects global resistance and subsequently the values of flux, due to the removal of the reversible fouled layer, but the influence of backwashing is observed only from the values of flux.
- For experiments in which no backwashing is done, the final and the initial permeate fluxes for two consecutive tests can be considered equal.

The design of the concrete network architecture used to develop separate models for ultrafiltration and backwashing takes into consideration the function ***f***, defined by expressions (2) and (3), for the particular case of $m = 2$ and $q = 1$, corresponding to the following significance of input (x) and output (y) variables, respectively.

- Ultrafiltration model: $x = [tJ_{ini}]^T$ and $y = J$, where T denotes transposition and t the time instant, measured with respect to the beginning of ultrafiltration (s), J_{ini} the initial flux value standardized at 20°C ($lh^{-1}m^{-2}$), and J the current flux value ($lh^{-1}\ m^{-2}$) corresponding to time instant t.
- Backwashing model: $x = [t_b J_{prev}]^T$ and $y = J_{next}$, where T denotes transposition and t_b the time duration for backwashing between two consequent ultrafiltration tests (s), J_{prev} the final flux value for the previous ultrafiltration test, per-

Table 1. Analysis of fluxes and hydraulic resistances—membrane A-LF, based on a set of tests selected from the experimental studies (Teodosiu, 1996)

CWF (lh^{-1} m^{-2})	R_t (1) (m^{-1})	Permeate flux (lh^{-1} m^{-2})	R_t (2) (m^{-1})	Relative error for calculated R_t (%)
Long ultrafiltration tests (50% loss of the initial flux)				
157.5	$2.263 \times 10^{+12}$	155.9	$2.286 \times 10^{+12}$	1.01
159.3	$2.237 \times 10^{+12}$	160.89	$2.215 \times 10^{+12}$	0.99
123.8	$2.879 \times 10^{+12}$	122.74	$2.904 \times 10^{+12}$	0.86
128.1	$2.782 \times 10^{+12}$	122.74	$2.904 \times 10^{+12}$	4.20
149.6	$2.382 \times 10^{+12}$	143.01	$2.492 \times 10^{+12}$	4.41
Short ultrafiltration tests (20% loss of the initial flux)				
157.5	$2.260 \times 10^{+12}$	152.01	$2.344 \times 10^{+12}$	3.58
141.9	$2.511 \times 10^{+12}$	143.97	$2.475 \times 10^{+12}$	1.45
140.7	$2.533 \times 10^{+12}$	139.40	$2.556 \times 10^{+12}$	0.90
127.9	$2.786 \quad 10^{+12}$	131.26	$2.715 \times 10^{+12}$	2.62
141.9	$2.511 \times 10^{+12}$	137.72	$2.588 \times 10^{+12}$	2.98
150.0	$2.376 \times 10^{+12}$	143.97	$2.475 \times 10^{+12}$	4.00
144.3	$2.470 \times 10^{+12}$	138.67	$2.570 \times 10^{+12}$	3.89
161.5	$2.207 \times 10^{+12}$	165.71	$2.151 \times 10^{+12}$	2.60

formed just before backwashing (lh^{-1} m^{-2}), and J_{next} the initial flux value for the next ultrafiltration test, to be performed after backwashing (lh^{-1} m^{-2}).

It should be mentioned that such a choice of the variables allows *sequencing* ultrafiltration–backwashing cycles, since, for the k-th cycle, one may write the following equalities: J_{ini} for cycle k (in ultrafiltration) = J_{next} for cycle $k-1$ (in backwashing); J_{prev} for cycle k (in backwashing) = J for final time in cycle k (in ultrafiltration).

The exploitation of neural networks for model development requires *four distinct stages*.

Network architecture design. The characteristics of the artificial neural networks that have been developed for flux prediction considering the cycles ultrafiltration-backwashing are in accordance with the notations and feedforward topology presented in Fig. 1. Thus, three layers of neurons have been used, with 15 neurons in the first layer, 5 neurons in the second layer (i.e. $n = 15$, $p = 5$) and 1 neuron in the final layer ($q = 1$). Tansigmoid activation functions were considered for all layers (Teodosiu, 1998). The dimensions of the input and hidden layers were chosen in correlation with the performances achieved during the training procedure for

Table 2. Characteristics of the neural network training procedures implemented in MATLAB, based on a set of tests selected from the experimental studies (Teodosiu, 1996)

Inputs	Ultrafiltration	Backwashing
Experimental data (x_i, z_i) $i = 1, \ldots, N$	*Patterns*: time instants (t) and initial flux value (J_{ini}) *Targets*: current flux values (J) tests 3.14–3.22 ($N = 541$)	*Patterns*: backwashing time (t_b) and final flux value for previous ultrafiltration test (J_{prev}) *Targets*: initial flux value for the next ultrafiltration test (J_{next}) tests 2.1–2.4; 3.2–3.7; 3.14–3.22 ($N = 16$)
Parameters for training control:		
maximum number of iterations (epochs)	5000	10,000
error goal	0.0001	0.00002
Computational aspects:		
total number of flops	2376×10^6	144×10^6
sum squared error achieved after training	0.00206	0.00046

Table 3. Analysis of relative errors for the evaluation of model validity—based on a set of tests selected from the experimental studies (Teodosiu, 1996)

Case	Test no.	Maximum relative error per test (%)	Mean relative error per test (%)
(1)	3.14	2.89	1.37
	3.17	1.02	0.26
	3.18	0.88	0.32
(2)	3.2	2.75	1.59
	3.3	3.18	1.92
	3.5	1.47	0.69
	3.8	1.42	0.80
	3.10	1.87	1.07
	3.12	1.77	0.76
	5.1	3.37	1.81
	5.2	7.70	3.80

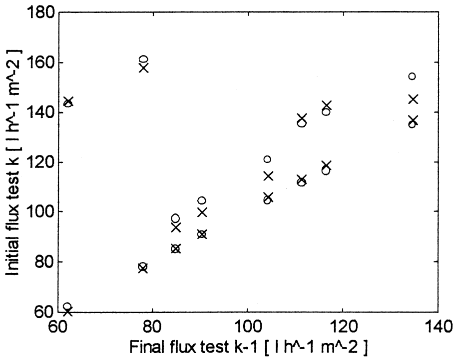


Fig. 3. Results of neural network training to approximate backwashing flux—membrane A-LF.

several architectures tested by modifying the number of nodes in the first and second layer, in order to accurately capture the non-linear time dependence of flux evolution. The proposed architecture represents a balance between quality of fitting and the complexity of network topology.

Data preparation and pre-processing. The pairs pattern-target $(\boldsymbol{x}_i, z_i)$, $i = 1, \ldots, N$, considered for minimization of objective function (5), i.e. for developing the neural-net based model, were constructed from the experimental data of some representative tests, as presented in Table 2. In these pairs, $\mathbf{x}_i \in \mathscr{R}^2$, $i = 1, \ldots, N$, specify numerical values corresponding to the generic vector $\boldsymbol{x}$ defined above for both ultrafiltration and backwashing models, while $z_i \in \mathscr{R}$, $i = 1, \ldots, N$, specify numerical values corresponding, as targets, to the network output y, in the two cases. Not all-available data were considered, because part of tests was left for further model validation. For numerical reasons, data were scaled before starting the training procedure, and afterwards, brought back in the initial range. This scaling allows the usage in the output layer of a tansigmoid activation function (instead of the linear one), in order to increase the quality of data fitting.

Neural network training. For both ultrafiltration and backwashing, MATLAB programs have been developed to implement the training procedure, taking advantage of the existing facilities available in Neural Network Toolbox (MATLAB, 1995). For both models, the initial values of the entries in weight matrices $\boldsymbol{W}_j$ and bias vectors $\boldsymbol{b}_j$ were random numbers in the interval $[-1, 1]$. As learning rule, *backpropagation with adaptive learning rate and momentum* was used (Freeman and Skapura, 1991). This procedure represents the translation in neural network terms of the standard steepest descent algorithm, applied for the objective function (5) and exploiting the particular form of input–output transfer (3) achieved by the neural network. Numerical values considered for parameters controlling the convergence of backpropagation, i.e. maximum number of iterations (epochs) and the threshold for the sum squared errors (error goal) are given in Table 2.

The training procedures implemented in MATLAB return the numerical values for the matrices $\boldsymbol{W}_{\mathrm{f}j}$, $\boldsymbol{W}_{\mathrm{b}j}$ and vectors $\boldsymbol{b}_{\mathrm{f}j}$, $\boldsymbol{b}_{\mathrm{b}j}$ $(j = 1, 2, 3)$, which ensure data fitting for ultrafiltration and backwashing, respectively. Inherent noise resulting from data acquisition is smoothed during training by the filtering characteristics of the network itself.

The neural network models have the following forms, derived from the general expression (3), using subscripts f and b for ultrafiltration and backwashing, respectively:

- model for predicting flux values in ultrafiltration:

$$
\begin{aligned}
J &= \boldsymbol{f}_{\mathrm{f}}\left(\begin{bmatrix} t \\ J_{\mathrm{ini}} \end{bmatrix}\right) \\
&= \boldsymbol{F}_3\left(\boldsymbol{W}^*_{\mathrm{f}3}\boldsymbol{F}_2\left(\boldsymbol{W}^*_{\mathrm{f}2}\boldsymbol{F}_1\left(\boldsymbol{W}^*_{\mathrm{f}1}\begin{bmatrix} t \\ J_{\mathrm{ini}} \end{bmatrix} + \boldsymbol{b}_{\mathrm{f}1}\right) + \boldsymbol{b}_{\mathrm{f}2}\right) \right. \\
&\quad \left. + \boldsymbol{b}_{\mathrm{f}3}\right) \qquad (8)
\end{aligned}
$$

- model for predicting the flux value after back-

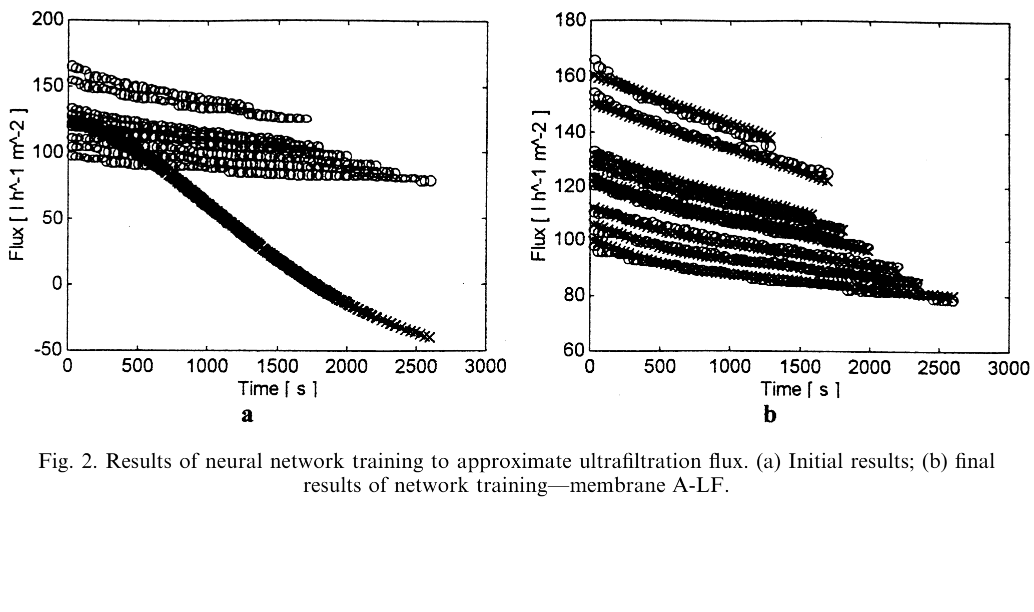


Fig. 2. Results of neural network training to approximate ultrafiltration flux. (a) Initial results; (b) final results of network training—membrane A-LF.

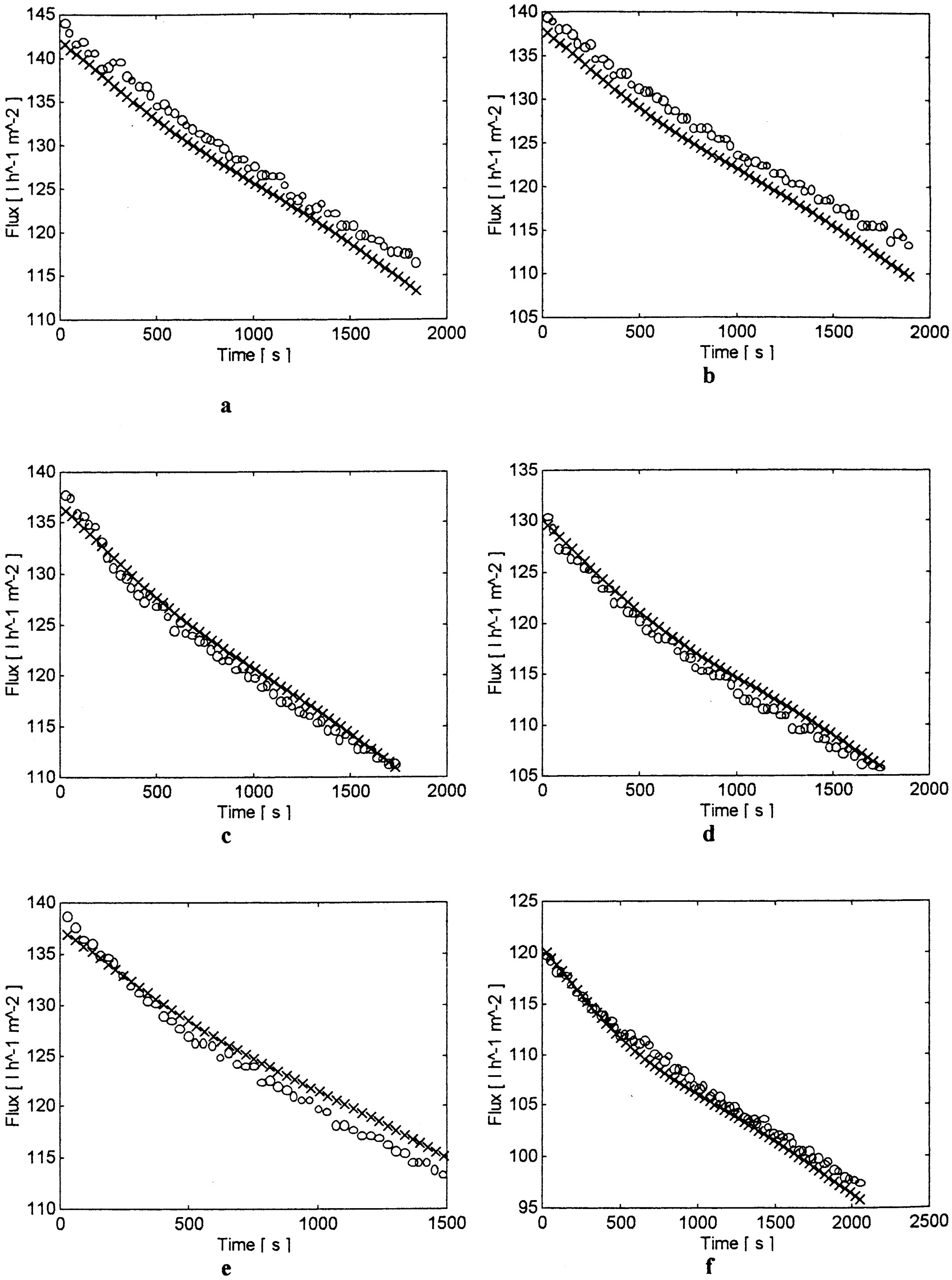


Fig. 4. Neural network approximation of experimental data for ultrafiltration—membrane A-LF, short tests: (a) test 3.2; (b) test 3.3; (c) test 3.5; (d) test 3.8; (e) test 3.10; (f) test 3.12.

washing:

$$\begin{aligned} J_{\text{next}} &= \boldsymbol{f}_{\text{b}}\left(\begin{bmatrix} t_{\text{b}} \\ J_{\text{prev}} \end{bmatrix}\right) \\ &= \boldsymbol{F}_3\left(\boldsymbol{W}^*_{\text{b}3}\boldsymbol{F}_2\left(\boldsymbol{W}^*_{\text{b}2}\boldsymbol{F}_1\left(\boldsymbol{W}^*_{\text{b}1}\begin{bmatrix} t_{\text{b}} \\ J_{\text{prev}} \end{bmatrix} + \boldsymbol{b}_{\text{b}1}\right)\right.\right. \\ &\quad \left.\left. + \boldsymbol{b}_{\text{b}2}\right) + \boldsymbol{b}_{\text{b}3}\right) \end{aligned} \tag{9}$$

It is worth mentioning that the *computational requirements* for constructing models (8) and (9) do not increase when applied to large-scale applications, since only a representative set of experimental data is necessary. This set of experimental tests has to cover the whole time range and the

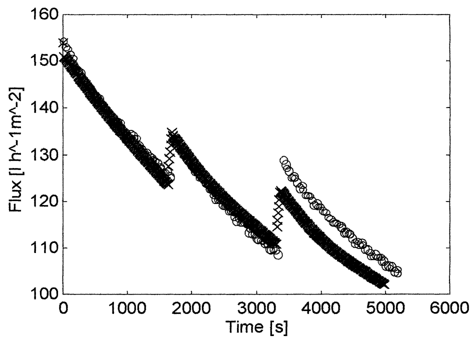


Fig. 5. Neural network approximation for ultrafiltration-backwashing cycles (tests 3.7, 3.8 and 3.9, membrane A-LF).

operating conditions envisaged for the practical usage of the model. Thus, once the network is appropriately trained, one needs only the *initial flux value* (which is currently measured in all ultrafiltration systems) and the *backwashing duration* in order to predict flux behaviour in ultrafiltration and backwashing.

Evaluation of model validity. The validity of the proposed model is evaluated with respect to the following criteria:

1. accuracy of the approximation for experimental data used in training (data "seen" during the learning phase);
2. accuracy of the approximation for other experimental data than those used in training (data "unseen" during the learning phase).

The above criteria are analytically expressed in terms of relative errors (referring to predicted vs measured values). Numerical details are relevant for the ultrafiltration model, since the construction of such models requires a large volume of experimental data. In case (1), Table 3 presents the *maximum* and *mean* value of relative error per test for those "seen" tests, which characterize the best and, respectively, the poorest fitting of experimental data. In the same table, there are also presented these values for case (2), referring to a number of tests "unseen" during training.

The discussed criteria are illustrated graphically, by simultaneous plots of time dependence for the neural network predicted flux and experimentally acquired values. In all graphical representations the following symbols have been used: o—experimental data, x—neural network approximation.

The plots in Figs 2 and 3 refer to the fulfilment of criterion (1). Figure 2 depicts the flux prediction for ultrafiltration given by the untrained network, corresponding to random values assigned to weights and biases—Fig. 2a, and by the trained network as given by equation (8)—Fig. 2b. Figure 3 presents the results of neural network training for the prediction of initial flux (test k) after backwashing, as a function of final permeate flux of the previous test ($k-1$) and backwashing duration, according to equation (9).

Regarding the fulfilment of criterion (2), the following graphical plots are considered relevant for model validation in case of both ultrafiltration and backwashing:

- short ultrafiltration tests performed with the same membrane, A-LF, in the same operating conditions as those used for training (results presented in Fig. 4a–f);
- a series of short ultrafiltration tests, each of them followed by backwashing in order to observe the quality of flux prediction when sequencing both models (results presented in Fig. 5);
- short ultrafiltration tests performed with the other membrane, A, in different operating conditions as those used for training, according to the Materials and Methods section (results presented in Fig. 6a–b);

As it can be observed for all the new examples illustrated by Fig. 4 (experiments "unseen" by the net-

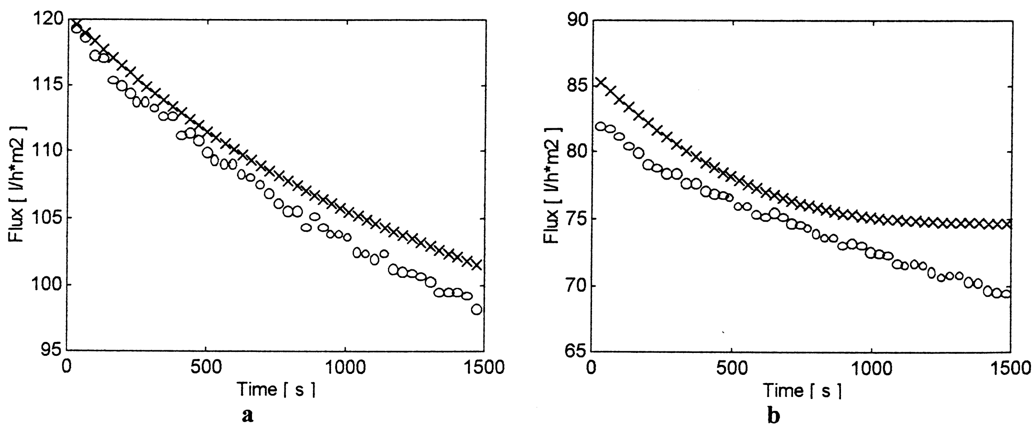


Fig. 6. a–b. Neural network approximation for ultrafiltration—membrane A, short tests: (a) test 5.1; (b) test 5.2.

works during the training procedure), the ultrafiltration model is able to accurately predict the non-linear evolution of flux at any time instant during ultrafiltration. Considering the errors for these predictions, the worst case is for test 3.3, with maximum and mean relative errors of 3.18% and 1.92%, respectively, according to Table 3.

Experimental data of three consecutive tests have been compared with the predicted values given by neural network models of ultrafiltration and backwashing. As observed from Fig. 5, the backwashing model is able to give a good prediction of initial flux value for the next ultrafiltration test, thus creating a link between these operations and making possible a global description of flux evolution.

Even for membrane A (tests 5.1, 5.2), which exhibited a different experimental behaviour due to different material properties (Teodosiu *et al.*, 1999), the neural network approximation is rather good, as reflected by Fig. 6a–b. Considering the relative errors, the worst case is for test 5.2, with maximum and mean relative errors of 7.7% and 3.8%, respectively, according to Table 3. The role of these validations for membrane A is just to demonstrate some capabilities for the network to extrapolate the flux behaviour within a phenomenological context slightly different from the training conditions. The construction of a proper model for membrane A requires the application of the whole training procedure with a set of data pertaining to this membrane.

CONCLUSIONS

Based on a series of representative experimental data, the artificial neural networks were trained to accurately describe both flux evolution during ultrafiltration on hollow fibre membranes (as a function of time and initial permeate flux value), and to predict the initial permeate flux after backwashing (as a function of final permeate flux of previous ultrafiltration test and the backwashing duration). These models allow a unified approach that represents the background for optimization and computer-aided control of ultrafiltration and backwashing operations.

The advantage of these models, besides an accurate description of the global process, is that they may be adapted for very different operating or backwashing conditions, and influent qualities, provided that a sufficient number of relevant experimental tests are available for the training procedure. Another advantage of these models is that flux is a parameter constantly monitored in all ultrafiltration systems, operated in either cross-flow or dead-end flow, and consequently the application of such models can also be done for large-scale treatment plants, without other supplementary investments.

A very good prediction of flux at any time instant can be achieved using the artificial neural network and this aspect is proved by very small values of the relative errors obtained after training for both ultrafiltration and backwashing. Altogether, the validity of these models has been verified with very good results for "seen" and "unseen" tests, offering a fairly accurate description of experimental flux evolution and, thus, creating theoretical premises for further approach to the optimization of the process.

When variances are available for the measured experimental data, a statistical interpretation, based on Fisher's test, can analyse the conditions for which errors introduced by data processing in model construction do not exceed the inherent errors resulted from data acquisition.

REFERENCES

Blume I., Koenhen D. M. and Roesink H. D. W. (1995) Cross flow vs dead-end in large scale ultrafiltration processes. Euromembrane '95 Bath: 1–6.

Cheryan M. (1998) *Ultrafiltration and Microfiltration Handbook*. Technomic Publishing Co. Inc.

Cybenko G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* **2**(4), 303–314.

Delgrange N., Cabassud C., Durand-Bourlier L. and Laine J. M. (1998) Modelling of ultrafiltration fouling by neural network. *Desalination* **118**, 213–227.

Erricson B. and Hallmans B. (1994) Membrane applications in raw water treatment with and without reverse osmosis desalination. *Desalination* **98**, 3–16.

Fane A. G. and Fell C. J. D. (1987) A review of fouling and fouling control in ultrafiltration. *Desalination* **62**, 117–136.

Fratila L. (1997) Membrane filtration for combined refinery and petrochemical wastewater reuse. *M.Sc. thesis*, DEW 10, IHE Delft, 1–97.

Freeman J. A. and Skapura D. M. (1991) *Neural Networks: Algorithms, Applications and Programming Techniques*. Addison Wesley Publishing Company.

Haykin S. (1994) *Neural Networks*. IEEE Press and MacMillan.

Hornik K., Stinchombe M. and White H. (1989) Multilayer feedforward networks are universal approximations. *Neural Networks* **2**, 359–366.

Jones K. L., Odderstol E. S., Wetterau G. E. and Clark M. M. (1993) Using a hydraulic model to predict hollow-fiber UF performances. *Journal AWWA* **October**, 87–97.

Koprowski T. P. (1995) Development of the modified fouling index for the prediction of fouling of hyper-, nano-, and ultrafiltration membranes. *M.Sc. thesis*, IHE, Delft: 1–62.

Mallevialle J., Odendaal P. E. and Wiesner M. R. (1996) The emergence of membranes in water and wastewater treatment. In *Water Treatment Membrane Processes*, eds J. Mallevialle, P. E. Odendaal and M. R. Wiesner, pp. 1.1–1.10. AWWA Research Foundation, McGraw-Hill.

Manth T., Frenzel J. and Vlerken A. (1998) Large scale application of UF and RO in the production of demineralized water. *Desalination* **118**, 255–262.

MATLAB Neural Network Toolbox, version 2.0 (1995) The Mathworks Inc.

Teodosiu C. (1996) Feasibility of advanced treatment for recycling secondary refinery effluent as cooling water make-up. *M.Sc. thesis*, EE 334, IHE Delft, 1–122.

Teodosiu C. and Macoveanu M. (1998) Ultrafiltration mechanisms in the treatment of refinery wastewater. *Scientific Annals of University "Al.I.Cuza"*, **VI**, Chemistry, Iaşi, 167–176.

Teodosiu C., Kennedy M., van Straten H. and Schippers J. C. (1999) Evaluation of secondary refinery treatment using ultrafiltration membranes. *Water Research* **33**(9), 2172–2180.

Teodosiu C. (1998) Optimisation of an advanced wastewater treatment process for the removal of non-biodegradable organic matter and suspended solids removal. Ph.D thesis, Technical University of Iaşi.

Wetterau G. E., Clark M. M. and Anselme C. (1996) A dynamic model for predicting fouling effects during the ultrafiltration of a groundwater. *J. Membr. Sci.* **109**, 185–204.