ELSEVIER

Chemical Engineering Journal 140 (2008) 71–76

Chemical Engineering Journal

www.elsevier.com/locate/cej

Discoloration process modeling by neural network

Oswaldo Luiz Cobra Guimarães *, Marta Heloisa dos Reis Chagas, Darcy Nunes Villela Filho, Adriano Francisco Siqueira, Hélcio José Izário Filho, Henrique Otávio Queiroz de Aquino, Messias Borges Silva

School of Engineering of Guaratinguetá Sate University, UNESP, Brazil

Received 12 November 2005; received in revised form 22 August 2007; accepted 7 September 2007

Abstract

The photo-oxidation of acid orange 52 dye was performed in the presence of H_2O_2, utilizing UV light, aiming the discoloration process modeling and the process variable influence characterization. The discoloration process was modeled by the use of feedforward neural network. Each sample was characterized by five independent variables (dye concentration, pH, hydrogen peroxide volume, temperature and time of operation) and a dependent variable (absorbance). The neural model has also provided, through Garson Partition coefficients and the Pertubation method, the independent variable influence order determination. The results indicated that the time of operation was the predominant variable and reaction mean temperature was the lesser influent variable. The neural model obtained presented coefficients of correlation on the order 0.98, for sets of trainability, validation and testing, indicating the power of prediction of the model and its character of generalization.

Keywords: Neural modeling; Azo dye; UV/H_2O_2

1. Introduction

The most common dyes present in industrial wastewaters, particularly in the textile ones, are the azodyes (60–70% of the world production), which are defined as composites with one or more –N=N– bonds, known as chromophore structure, which are able to supply color through radiant energy absorption.

The presence of dyes in the discarded wastewater may be a hazard to public health and possibly to the ecosystem. The discard of colored compounds alters the water transparency, thus impeding the penetration of solar radiation and diminishing the photosynthetic activity of certain aquatic organisms. Therefore, these systems become unstable. The effects of the dye are substantial, in that even low concentrations of dye (less than 1 ppm) can have significant effects on aquatic environments.

With respect to the color, the treatment, via biodegradation with the action of microorganisms is revested of much complexity in function of the toxicity conditions supported by the microorganisms and in many situations the color does not change, due to the non-biodegradable nature of the dye. Shu et al. [1] point out that numerous dyes have complicated scent structures and are resistant to conventional biological treatment systems. Thus, the search for new processes and the optimization of processes already in use for discoloration and degradation of industrial effluents that contain synthetic dyes are necessary.

Muruganandham and Swaminathan [2] indicate the efficacy of advanced oxidation processes, particularly the binomial UV/H_2O_2 in the synthetic dye chromophore structure destruction.

Due to the large number of existing dyes it becomes very difficult to determine a single method which can reduce the dyeing grade and purify the waters from industrial processes, without taking in account the dye structure and its relation to the variables that influence particularly their discoloration process. The discoloration process modeling due to the dye complex nature and their dependence of many factors and variables revests the problem with a high difficulty level, characterizing it as a multiple analysis system [3]. In this sense, neural networks appear as a multiple variable processes modeling and according to Pareek et al. [4] it has been becoming very popular in the Chemical Engineering area. The authors emphasize the capacity of recognition and reproduction of the cause–effect relations for multiple output and input systems and, thus, it is possible to map a $A = f(c_0$, pH, t, V, $T)$ relation where A is the absorbance dependent vari-

* Corresponding author. Tel.: +55 1231595096.
E-mail address: oswaldocobra@debas.eel.usp.br (O.L.C. Guimarães).

1385-8947/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cej.2007.09.021

Nomenclature

A	predicted values of absorbance
[dye]	dye concentration
$f(s)$	transfer function
I	network input variable relative importance
MSE	mean square error
Purelin	linear function
R	Pearson correlation coefficients
S	weight pondered sum
tansig	sigmoidal tangent
T	photo-oxidizing process operation time
T	temperature
T	real values of absorbance
$V_{H_2O_2}$	hydrogen peroxide volume
w	synaptic connections
X	input vector
Y	output vector

Greek letters

ε	mean square error
λ	maximum wavelengths

able, c_0 the dye initial concentration, t the time of operation, V the hydrogen peroxide volume added and T is the reaction temperature.

2. Artificial neural networks

The neural networks can be trained for complex mappings, for the hidden layer elements learn to respond to characteristics found in the input, which refer to correlations of activities among different input spots, allowing an input information abstract representation in the hidden layers. They have, the abstraction ability, the generalization ability, classifying a complex pattern correctly, even if it does not belong to the network training set. It is also vigorous and immune to little fails or noises present in the inputs [5].

In the feedforward network (Fig. 1), the neurons are connected to all posterior layer neurons, with the information from an anterior layersuffering a ponderation by a (w_{ij}) weight that is sent to all neurons of the next layer.

The networks with backpropagation training model refer to the way the weights are adjusted. This way is also known as General Delta Rule, based on the descending gradient optimization and has been used in the majority of the works applied to chemical processes [6].

Processing elements of the same layer act in parallel and layer-to-layer processing is sequential. The equations that administrate the feedforward processing are

$$s_j^{(k)} = w_{0_j}^{(k)} + \sum_{i=1}^{N_k} w_i^{(k)} x_i^{(k-1)} \quad (1)$$

$$x_j^{(k)} = f(s_j^{(k)}) \quad (2)$$

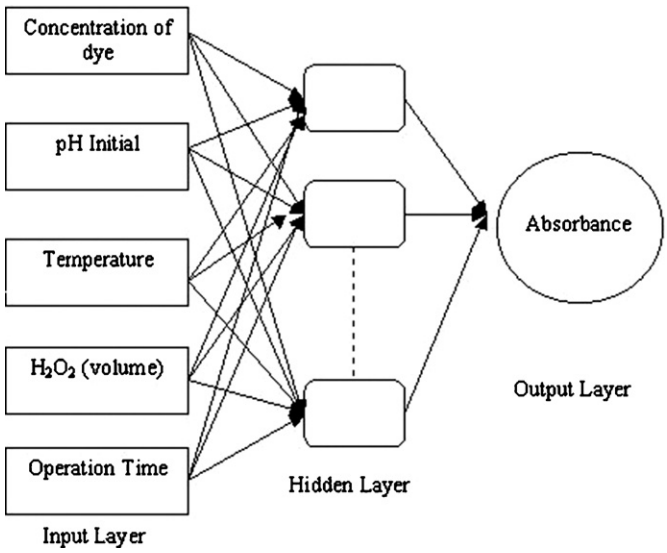

Fig. 1. Feedforward neural network model.

In this relation, $x_i^{(k)}$ refers to the activating function input of k layer i element, $s_j^{(k)}$ the pondered sum of the weights through the inputs and $w_{ij}^{(k)}$ refers to the synaptic connection weights at the k layer j element input, where i is the connection index and N_k is the k layer processing element number.

In the input, $x_i = x_i^{(0)}$ are X input vector components and, in the output, $y_i = x_i^{(m)}$ are the y output vector components. The neuron input and output may be related by $f(s_j) = 1/(1 + e^{-s_j})$ type transference function [7].

The purpose of training a network is the adjustment of its weights in such a way that the application of a pattern produces an output value and in this sense the General Delta Rule aims to reduce the network mean square error indicated by Eq. (3)

$$\varepsilon = \left(\frac{1}{2}\right) \sum_{j=1}^{m} (d_j - y_j)^2 \quad (3)$$

In Eq. (3), d_j is the desired (real) value and y_j is the value obtained by the network [5].

3. Materials

The acid orange 52 dye structure is presented in Fig. 2.

Hydrogen peroxide (30% by weight) was used in all the photo-oxidizing procedures. NaOH and H_2SO_4 (0.5N) were utilized to obtain the reaction mean initial pH. Distilled water was used to compose all the processes.

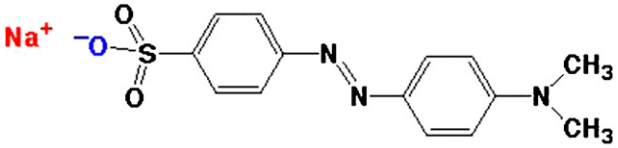

Fig. 2. Acid orange 52 structural formula.

The discoloration was evaluated as a function of the absorbance, measured every 5 min by removing 2 ml of sample, through Femto 600 spectrophotometer, at the maximum wavelength $\lambda = 463$ nm. Each sample was analyzed three times and the average value was used as the value for each experiment. Temperature and pH were monitored via pHmeter pG2000 Gehaka.

The photo-oxidizing process was performed in a Germetec GPJ 463-1 plug-flow reactor, emitting at 254 nm, with 21 W low-pressure radiation source (mercury vapor lamp). At the end of each experiment the system (for washing purposes) was filled with a slight acid solution and then recirculated. After discard and recirculation with distilled water, the system was disassembled and the reactor filled with 10% nitric acid solution for cleaning purposes.

The temperature for each experiment was kept constant through Optherm DC1 thermostat, in $T_i \pm 2\,^{\circ}C$, where T_i is the temperature of each experiment, from 22 °C to 45 °C.

4. Methodology

Initially a high dye concentration of 170 mg/l and a lesser amount of hydrogen peroxide (1 ml) was established for a model experiment. This model experiment was performed up to the point where the absorbance came close to zero value, providing a time of 150 min, that was set as this variable amplitude range maximum value, being characterized a process inspection model. Table 1 presents the levels for which the proposed neural network input variable dominium set was established.

The performance of the method indicated irrelevant results in the reduction of color at absence of peroxide or radiation in isolated processes. The input variable matrixes presented to the neural model are generically shown by

$$X = \begin{bmatrix} c_1 & \mathrm{pH}_1 & t_1 & V_1 & T_1 \\ c_2 & \mathrm{pH}_2 & t_2 & V_2 & T_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{218} & \mathrm{pH}_{218} & t_{218} & V_{218} & T_{218} \end{bmatrix} \quad (4)$$

Aiming to verify the existence of this matrix outliers, or solitary points of experiment, and in order to check the homogeneity of the data, each sample "leverage" (Fig. 3) was estimated, which is a measurement of how the sample influences the totality of data, and a small value identifies little sample influence over the model building.

Ferreira et al. [8] indicate that a critical value, or practical rule for the identification of anomalous points, namely, considered points with "leverage" higher than $3k/n$, where n is the number of samples (218) and k the number of main components or latent variables, five of them (analysis of components in Matlab environment) for the current work, resulting in a critical value of 0.068807 and, therefore, some samples were discarded from the set to be tested. Matlab prepcap (pn, 0.02) code transforms the input set data matrix already normalized (pn), retaining only the components that contribute with more than 2% in the input data set variance.

Table 1
Variables level

Variable level	Min	Max
H_2O_2 (ml)	2	15
[dye] (mg/l)	3	170
pH	2	12
Temperature (°C)	21	45
Operating time (min)	15	150

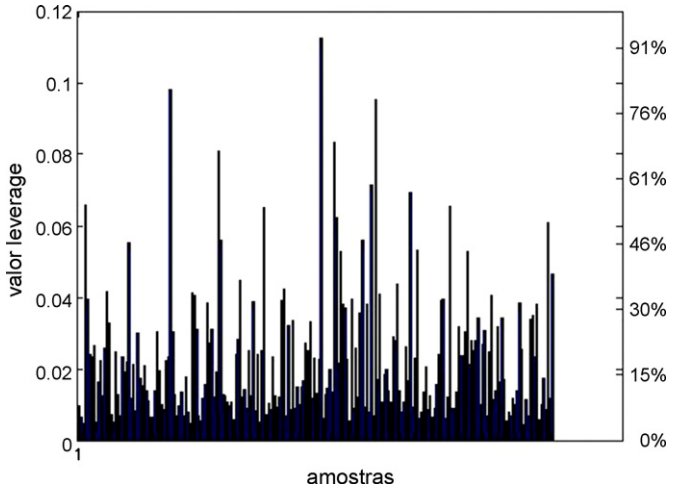

Fig. 3. Distribution of training, validation and test sets.

4.1. Training, validation and test set selection

There are several methods for picking out the sets to be used as training, validation and test sets. Kanduc et al. [9] establish the random selection, Kennard-Stone and Kohonen maps as some of the possibilities to be employed.

In the present work, the data were worked by following the basic algorithm given by:

1. a clustering was established using K-Means algorithm. The procedure followed the algorithm given by Martinez and Martinez [10];
2. after having determined the groups, a statistic test was used to set the training validation and test groups, in such a way that the training, validation and test sets pattern deviation and mean value be equal to less than a value tending to zero.
3. The input variables (in number of 5) and the output variables were processed in such a way that the mean value for each vector containing the dependent and independent variables be zero and the pattern deviation equal to 1, through the pn = (p − meanp)/stdp Matlab environment algorithm, where p is the input or output process matrix or data vector. In Matlab environment, this normalization and the generated set recording were performed by the command:

```
%NORMALIZED SET GENERATION
[pn, meanp, stdp, tn, meant, stdt] = prestd(p′, t′);
```

Table 2
Cluster distribution

Cluster	Cluster samples number
1	57
2	58
3	51
4	52

Table 3
Correlation coefficients

Hidden layer neuron number	R_1 (training set)	i_2 (validation set)	R_3 (test set)
8	0.988	0.982	0.979
12	0.976	0.971	0.963
15	0.990	0.980	0.979
16	0.991	0.986	0.981
20	0.990	0.984	0.977

The implementation of algorithm K-Means identified 4 clusters, herein named as clusters 1–4 (Table 2).

4.2. Neural network training

Table 3 presents the best results with a single hidden layer topology, with the respective linear (R) correlation coefficients. Neural networks with a hidden layer and a sufficiently large number of neurons can interpret any input–output structure and that the hidden layer neuron number is determined in function of the required accuracy.

All the configurations worked with the same 0.01 learning tax and the training performed in 22 epochs.

The functions used in the network-training algorithm were tansig and purelin (Matlab language) and the network weight actualization function was the Levenberg-Marquardt backpropagation (trainlm in Matlab language).

The function of error performance was MSE, or mean square error, and the performance learning function utilized was the descending gradient (learngdm).

Some of the parameters can be visualized in the sequence of commands given by

```
net = newff(minmax(pn),[co1{'tansig', 'purelin'}, 'trainml');
net.trainParam.epochs = 100; net.trainParam.goal = 0;
net.trainParam.1r = 0.01; % Learning tax
net.trainParam.show = 25; net.trainParam.mc = 0.9;
net.trainParam.1r_inc = 1.05; net.trainParam.1r_dec = 0.7;
  net.trainParam.max_perf_inc = 1.04;
net.performFcn='MSE';
```

The diagram of the network implemented may be seen in Fig. 4, where five input layer neurons related to the five network input variables, the 16 layer hidden layer and the input layer with a neuron corresponding to the absorbance output variable.

The linear activating function for the output layer is adequate for continuous phenomena, as for instance the oxygen biochemical demand or the absorbance degree in discoloration

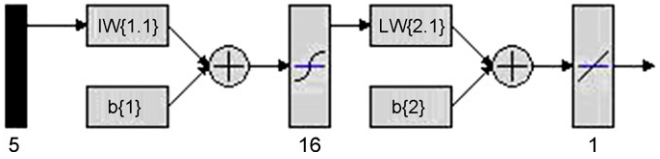

Fig. 4. Diagram of implemented neural model.

Table 4
Input variable classification

Variable	Importance (%)
H_2O_2	19.15
[dye]	21.44
pH	21.49
Temperature	16.97
Operating time	29.95

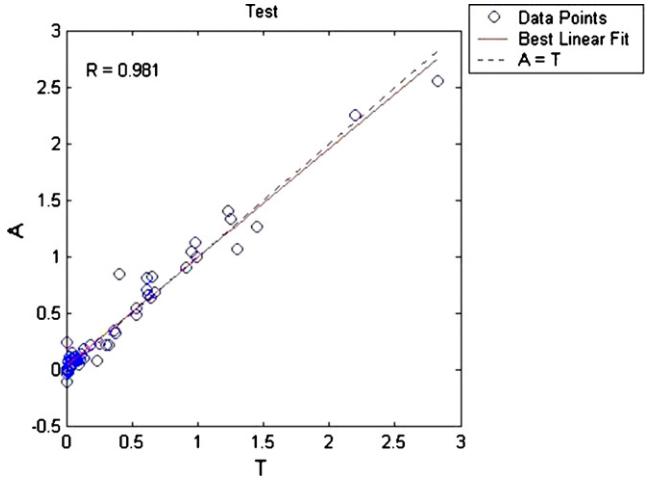

Fig. 5. Linear regression test set.

process. The sigmoidal type transference functions are necessary to introduce non-linearities in the network [11].

The IW {1,1} input neural weight matrix neural for the hidden layer may be visualized in Table 4 "input" lines and the hidden layer output neural weights in the "output" lines in this same table. The notations b{1} and b{2} refer to the bias vectors.

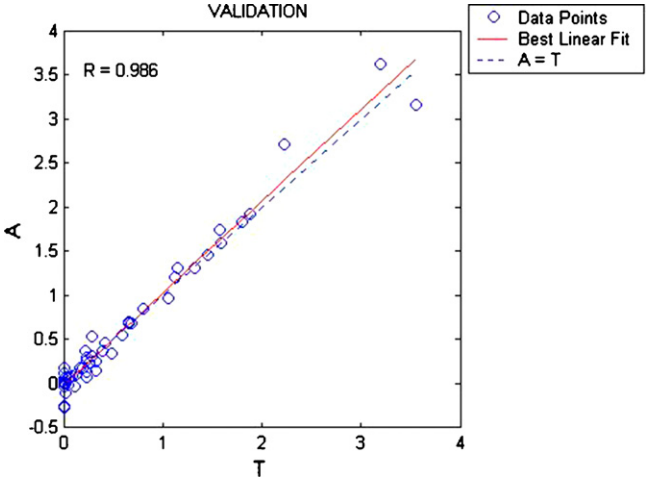

Fig. 6. Linear regression for validation set.

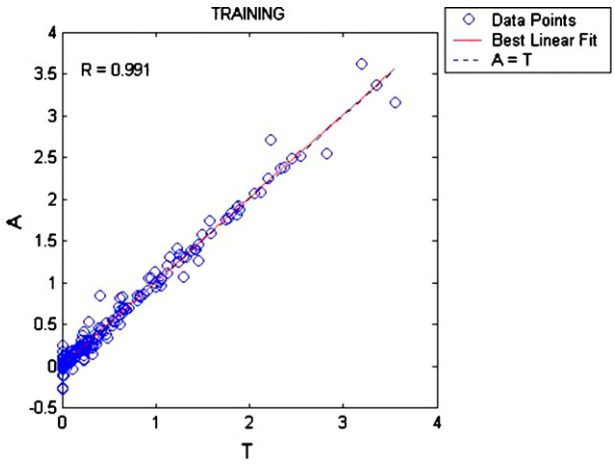

Fig. 7. Linear regression for training set.

In order to prevent overfitting problem, the training is interrupted if the error for the validation set becomes bigger than the training set error.

In function of the results obtained, hidden layer 16-neuron configuration was chosen. Graphically, the results may be visualized via Figs. 5–7.

5. Neural weights interpretation

The level of influence of each input variable concerning the modeling problem output variable may be obtained through the neural weight matrix. In order to reach this goal, Pareek et al. [4] an algorithm proposed by Garson [12], based on the neural weight partition.

$$I_j = \frac{\sum_{m=1}^{N^{\mathrm{h}}} \left(\left(\left| w_{jm}^{ih} \right| / \sum_{k=1}^{N^i} \left| w_{km}^{ih} \right| \right) \times \left| w_{mn}^{h_{\mathrm{o}}} \right| \right)}{\sum_{k=1}^{k=N^i} \left\{ \sum_{m=1}^{m=N^{\mathrm{h}}} \left(\left(\left| w_{km}^{ih} \right| \right) / \sum_{k=1}^{N^i} \left| w_{km}^{ih} \right| \right) \times \left| w_{mn}^{h_{\mathrm{o}}} \right| \right\}}$$

In the relation above mentioned, I_j is the j-esima input variable relative importance, N^{i} and N^{h} are the input and hidden neuron numbers, respectively and w stands for the neural weights, and i, h and o superscripts refer to the input, hidden and output layers. The subscripts k, m and n refer, respectively to the input, hidden and output layers.

As it can be seen in Table 4, all independent variables strongly influence the absorbances of the discoloration process.

In order to confirm the value importance order classification the pertubation method was applied. Gevrey et al. [13] indicates the perturbation method for input variable analysis. This consists of changes in the form $x_{\mathrm{i}} = x_{\mathrm{i}} + \delta$, where x_{i} is the selected input variable and δ is the variable change or noise. The method consists of attributing this noise and verifying the changes in the output y_{i} variable. In this work, the mean square error was used as comparison criterion.

The value $\delta = 10\%$ attributed in each variable, maintaining the other constants, produced the graph shown in Fig. 8, where the major importance of time of operation (t) is visualized, followed by the reaction mean (pH) and hydrogen peroxide volume (V).

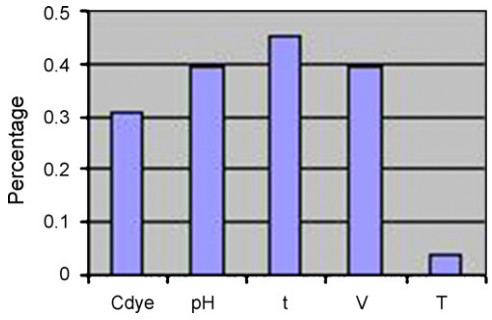

Fig. 8. MSE variation percentage.

Table 5
Correlation coefficients under noise in the input variables

	R_1	R_2	R_3
Noise (0%)	0.978	0.977	0.947
Noise (10%)	0.974	0.968	0.923

The factors that presented the minor MSE importance were the dye concentration (C_{dye}) and temperature (T), keeping this order of importance. It is noticed the coincidence in the three most important factors in the Garson Partition and Pertubation methods, namely, time of operation, pH and hydrogen peroxide volume.

In order to verify the stability of the values obtained through Garson Partition and Pertubation methods, the network trainings were repeated 10 times and the average contribution of each variable was calculated.

By comparing the results obtained through Garson Partition and Pertubation methods, an inversion is noticed concerning the dye concentration and temperature variables behavior, but equivalence was observed in the other variables, maintained the levels of importance.

A 10% noise value is attributed to the input data matrix aiming to verify the network capacity to self-adapt and prevent small failures or measurement errors, and Table 5 shows the network adaptation capacity to these noises, with the mean quadratic errors, and the linear correlation coefficients for training set (R_1), validation set (R_2) and test set (R_3).

6. Conclusions

An acid orange 52 dye discoloration neural model, with hydrogen peroxide, activated by UV radiation, was evaluated concerning five factors. The neural network was trained with 218 samples and utilized a configuration with a hidden layer and 16 neurons in this layer, presenting high correlation coefficients for training, validation and test sets (>0.98), verifying the network prediction capacity with high accuracy level. The input layer is formed by five variables: dye concentration, initial pH, time of operation, hydrogen peroxide volume at 30% and temperature. The study of the variable influence level determined that the input variables that influence the acid orange 52 discoloration process are time of operation, initial pH and hydrogen peroxide volume. However, temperature and concentration

of the dye should not be neglected, as they also appear to be significant factors.

References

[1] H.Y. Shu, M.C. Chang, H.J. Fan, Decolorization of azo dye black 1 by UV/H_2O_2 process and optmization of operating parameters, J. Hazard. Mater. B113 (2004) 201–208.

[2] M. Muruganandham, M. Swaminathan, Photochemical oxidation of reactive azo dye with UV–H_2O_2 process, Dyes Pigments 62 (2004) 269–275.

[3] Guimarães, O.L. Cobra, H.O.Q. Aquino, I.S. Oliveira, D.N. Villela, H.J. Izário, A.F. Siqueira, M.B. Silva, Prediction through neural networks of the residual of hydrogen peroxide used in photo-Fenton processes for effluent treatment, Chem. Eng. Technol. 30 (2007) 1134–1139.

[4] V.K. Pareek, M.P. Brungs, A.A. Adesina, R. Sharma, Artificial neural network modeling of a multiphase photodegradation system, J. Photochem. Photobiol. A: Chem. 149 (2002) 139–146.

[5] C. Loesch, S.T. Sari, Neural Networks Artificial Beddings and Models, Publishing Company of the Furb, 1996.

[6] A.F.M. Silvares, Mathematical Modeling of Photochemist Reactors Applied to the Treatment of Effluent, Polytechnical School of the University of São Paulo, 2001.

[7] A. Durán, J.M. Monteagudo, M. Mohedano, Neural networks simulation of photo-Fenton degradation of reactive blue 4, Appl. Catal. B: Environ. 65 (2006) 127–134.

[8] M.C.M. Ferreira, A.M. Antunes, M.S. Melgo, P.L. Volpe, Chemometrics I: Multivaried Calibration, a Tutorial, vol. 22, Química Nova, 1999.

[9] R.K. Kanduc, J. Zupan, N. Madcen, Separation of data on the training and test for modelling: a case study for modelling of five colour properties of a white pigment, Chemometr. Intell. Lab. Syst. 65 (2003) 221–229.

[10] W.L. Martinez, A.R. Martinez, Computational Statistics Handbook with MatLab, ChapMan & Hall/CRC, 2002.

[11] O.L.C. Guimarães, M.B. Silva, Hybrid neural model for descoloration by UV/H_2O_2 envolving process variables and structural parameters characteristics to azo dyes, Chem. Eng. Process. 46 (2007) 45–51.

[12] G.D. Garson, AI Expert, 1991, p. 46.

[13] M. Gevrey, I. Dimopoulos, S. Lek, Review and comparison of methods to study contribution of variables in artificial neural networks models, Ecol. Model. 160 (2003) 249–264.