ELSEVIER

Journal of Photochemistry and Photobiology A: Chemistry 149 (2002) 139–146

Journal of Photochemistry and Photobiology A:Chemistry

www.elsevier.com/locate/jphotochem

Artificial neural network modeling of a multiphase photodegradation system

V.K. Pareek [a], M.P. Brungs [a], A.A. Adesina [a,*], Raj Sharma [b]

[a] *Reactor Engineering and Technology Group, School of Chemical Engineering and Industrial Chemistry, University of New South Wales, Sydney, NSW 2052, Australia*

[b] *EPIN System Pvt. Ltd., 14, Uniara Garden, Moti Doongri Road, Jaipur 302004, India*

Received 3 September 2001; received in revised form 24 October 2001; accepted 14 November 2001

Abstract

Photodegradation of spent Bayer liquor was carried out in an 18 l pilot scale photoreactor. The experimental data indicated that the average reaction rate was a complex nonlinear function of various process variables, such as lamp power, catalyst loading, initial solution pH, liquid batch time, and total organic carbon (TOC) concentration. The experimental data were modeled using feed forward artificial neural networks (ANN). The networks were trained with 350 sets of input–output patterns using backpropagation algorithm. Out of several configurations, a three-layered network with eight-neurons in its hidden layer yielded optimal results with respect to data validation. The optimal model gave excellent predictions with a correlation coefficient of 0.9955.

Keywords: Photodegradation; Artificial neural networks (ANN); Bayer liquor; Backpropagation

1. Introduction

The last two decades have seen the emergence of a voluminous literature on the development of novel photocatalytic processes for the treatment of wastewater containing toxic and recalcitrant pollutants [1–3]. The advantages of photocatalysis over other wet aqueous removal routes are well espoused and include, complete mineralization of the organic species at relatively mild operating conditions. In many cases, titania-based compounds are used as the photocatalysts. The performance of a photoreactor is strongly influenced by several physicochemical factors, which are presumably independent variables [4]. However, the interaction effect of these variables suggests that conventional linear superposition techniques may not yield adequate model(s) for design purposes. For example, since photo-excitation of the semiconductor catalyst leads to the production of holes (positive vacancies) and electrons, which mediate in the substrate adsorption step(s), most photocatalytic destruction mechanisms are regarded as redox processes. Consequently, the pH of the solution impacts upon the elementary rates, suggesting a distinct interaction between pH and light intensity or lamp power (considered experimental input variable). By same token, the optimal catalyst loading is reportedly dependent on the pH of the reacting medium since particle aggregate size is influenced by prevalent electrostatic field within the solution [2]. This in turn attenuates the effective light distribution. These observations provide evidence of a latent linkage between pH, catalyst loading and light intensity. Although detailed mechanistic understanding to de-couple these variable interaction effects is unknown, effective process design may still be achieved by using modeling paradigms that account, and indeed take advantage of multiple interplay of factors to describe, predict and ultimately optimize process performance. It is in this context that the ANN modeling presents itself as a veritable approach [5].

In recent years, the concept of artificial neural networks (ANN) has gained wide popularity in many areas of chemical engineering [5]. The ability of ANN to recognize and reproduce cause–effect relationships through training, for multiple input–output systems, makes them efficient to represent even the most complex systems [6]. In general, the rate of reaction in heterogeneous photocatalytic reactors is a complex nonlinear function of catalyst loading, light intensity, initial solution pH, and reactant concentration.

In this paper, we have demonstrated the use of this powerful technique (ANN) in analyzing input–output pattern for a batch photoreactor. A set of 350 data points for the photodegradation of spent Bayer liquor in an 18 l pilot scale reactor were used to train the neural nets [7]. The input layer

* Corresponding author. Tel.: +61-2-9385-5268; fax: +61-2-9385-5966.
E-mail address: a.adesina@unsw.edu.au (A.A. Adesina).

1010-6030/02/$ – see front matter
PII: S1010-6030(01)00640-2

of the neural net consisted of catalyst loading, light intensity, initial solution pH and reactants concentration and the output layer contained one neuron—the average reaction rate.

2. Experimental details

2.1. Apparatus

A schematic of the 18 l reactor is shown in Fig. 1. The outer chamber was a stainless steel vessel (i.d. = 20 cm) inside which a commercial UV lamp was suspended from the top flange through double O-ring seals. The UV lamp was enclosed in a double-walled quartz hollow U-tube (o.d. = 4 cm) through which water was passed at 1.5 l min^{-1} as coolant to maintain reaction isothermality and to remove IR fraction of the incident radiation. Monitoring ports for pH, temperature and telescopic UV detector/radiometer were provided at indicated locations. Iwaki magnetic pumps were used for all liquid delivery (water and diluted Bayer liquor) while electronic mass flow controllers were used to meter all gas flow rates. Air from the flow controller was passed upwards through the stationary column of Bayer solution containing suspended titania particles via a 70 μm stainless mesh distributor. The effective height of reacting liquid column was 80 cm, and four sampling ports were provided at axial positions, Z = 10, 30, 50, and 70 cm. The gas flow rates ranged 1.0–10.0 l min^{-1} (τ_G = 2–18 min).

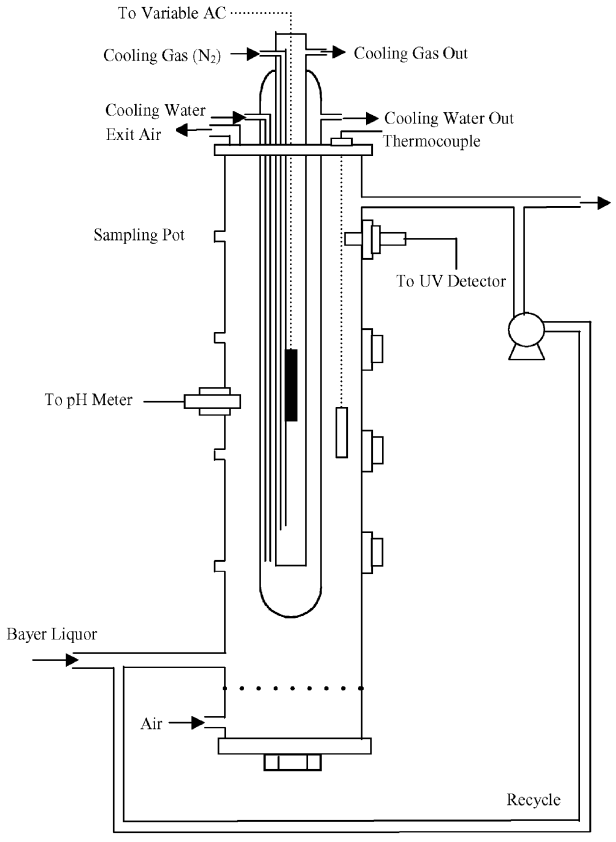

Fig. 1. A Schematic of pilot plant reactor.

2.2. Materials and methods

Titania (>99% anatase) was obtained from Aldrich Chemicals and used as supplied in all runs. Ordinary domestic water was used to dilute the industrial Bayer liquor. HCl was used to adjust the pH of the slurry prior to reaction. Ultra-pure N_2 gas (flowing over the external surface of lamp was used to remove ozonized air) was obtained from Linde, Sydney. Ambient compressed air was used as the feed gas for the reactor. A TPS Digital pH meter was used for continuous pH monitoring. Light intensity was measured by an IL1400 radiometer/detector (International Light, MA, USA) calibrated for 265–332 nm. Steady state aliquots were taken for different liquid and gas holdup times and analyzed for total carbon (TC) and total organic carbon (TOC) on a Shimadzu TOC Analyzer 5000.

3. Bayer liquor

Bayer liquor is a dark and viscous solution produced in the alumina refineries during caustic digestion of bauxite ore. In this process, alumina (Al_2O_3) as well as other high molecular weight organic compounds (essentially humic acids) associated with the ore are also converted to sodium salts (mostly aluminate, carbonate and oxalate). This results in substantial loss (>10%) in caustic solution which should be recycled to the digestion unit. Although sodium carbonate may be readily treated with lime to yield NaOH and $CaCO_3$, the removal of sodium oxalate is difficult. Owing to its low solubility in Bayer liquor, sodium oxalate easily crystallizes during alumina precipitation. This leads to increased generation of alumina trihydrate fines, and hence, poor hydrate classification efficiency. Additionally, sodium oxalate is a pollutant and its disposal must satisfy strict statutory environmental regulations [8].

In this study, the photodegradation of sodium oxalate in spent Bayer liquor was carried out to recover sodium hydroxide. Since the original Bayer liquor is a thick and viscous solution, the present pump capabilities did not allow the processing of pure Bayer liquor. Therefore, the feed to the reactor was diluted using water. Different values of dilution ratio were used, where the dilution ratio is defined as the ratio of volume of water added to the volume of pure Bayer liquor.

4. Artificial neural networks (ANN)

ANN are direct inspiration from the biology of human brain, where billions of neurons are interconnected to process a variety of complex information. Accordingly, a computational neural network consists of simple processing units called neurons (cf. Fig. 2). In general, a neural net, as

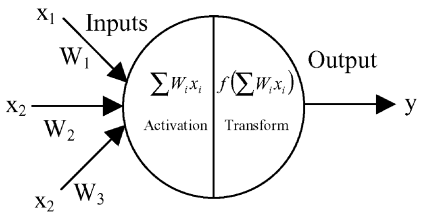

Fig. 2. An artificial neuron.

shown in Fig. 3, is parallel interconnected structure consisting of: (1) input layer of neuron (independent variables), (2) a number of hidden layers, (3) and output layer (dependent variables). The number of input and output neurons is determined by the nature of the problem. The hidden layers act like feature detectors and in theory, there can be more than one hidden layer. Universal approximation theory, however, suggests that a network with a single hidden layer with a sufficiently large number of neurons can interpret any input–output structure [9,10]. Therefore, in this study, we have used only one layer of hidden-neurons. The number of neurons in the hidden layer is determined by the desired accuracy in the neural predictions. Hence, it may be considered as a parameter for the neural net design.

The architect of neural nets may vary widely [5]. The net shown in Fig. 3 is called a feed forward neural net. In the feed forward neural net, all the neurons of a particular layer are connected to all the neurons of the layer next to it. Other types of nets that have been used in chemical engineering applications include recurrent networks. They are similar to feed forward neural nets, but also include time-delayed feedback or recycle [5].

The input layer of neurons acts like a distributor and the input to this layer is directly transmitted to the hidden layer. The inputs to hidden and output layers are calculated by performing a weighted summation of all the inputs received from the preceding layer. Generally, the output from the hidden layer is calculated by using a transfer function. Most widely used transfer function is the sigmoid transfer function given by

$$f(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

Details of mathematical expressions used in this study are reported in Appendix A.

5. Model development

The development of a neural net model generally consists of three steps. The first step is to generate the data for the training. The second step is the training of the neural net with the selected data. Here, the net is exposed to a certain number of patterns and an objective function is used

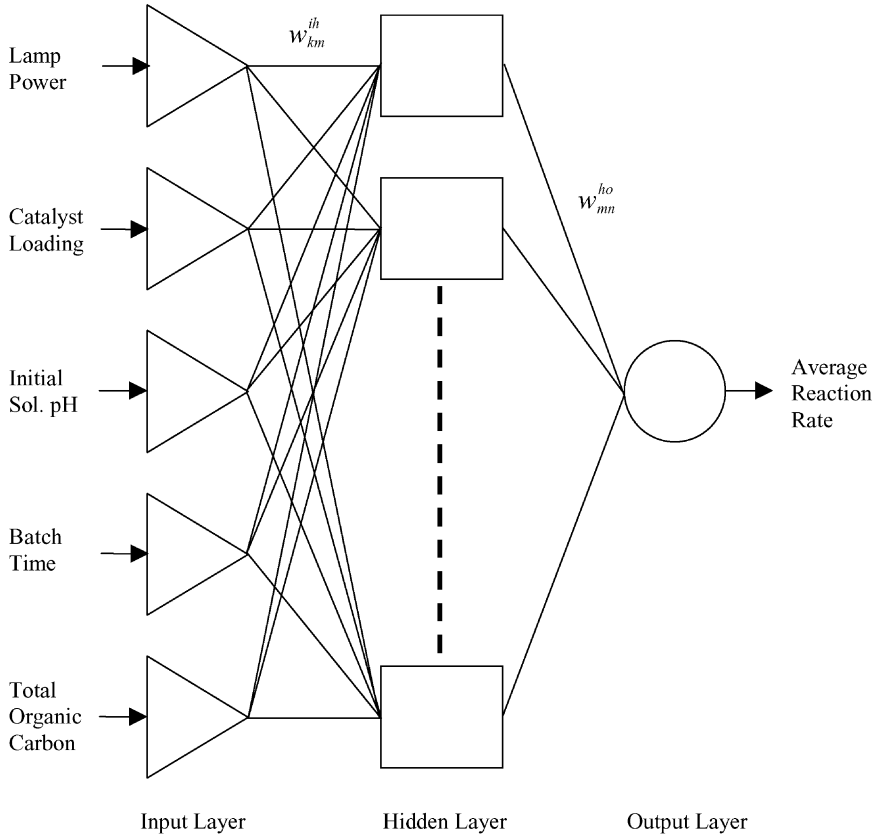

Fig. 3. Neural net model.

to minimize the errors in predicted and target values. The third step (testing) involves exposure of the trained network to unfamiliar dataset and consequently the accuracy in the predicted pattern is adjudged.

5.1. Data collection

Since an ANN learns by examples, exposure to multiple datasets improves the performance. Therefore, it is essential that one should have a sufficiently large number of datasets to enable a net training of valid generalization. A good training dataset should include the effect/variation of all the input variables on the output variables. Since it is not possible to collect data for the complete domain of a process, a subset of all the possible input–output patterns is usually used. However, in order to achieve a valid generalization, the training set must be representative of the domain of interest. A wrongly chosen training set may give very poor predictions with datasets unknown to the net. However, for some processes, it may be impossible to conceive all possible input–output patterns, the extrapolation of the input data using ANN may result in unacceptable predictions even with presumably right set of training data. Therefore, a careful monitoring of the ANN predictions is warranted while dealing with novel inputs.

Table 1
Model variables and their ranges

Variable	Range
Input layer lamp power	200–400 W
Catalyst loading	0.5–4.0 g l^{-1}
Initial solution pH	8.0–12.0
Batch time (liquid)	90–120 min
TOC concentration	700–1400 ppm (60–120 mmol l^{-1})
Output layer reaction rate	0.008–0.05 mmol l^{-1} min^{-1}

In this case, the set of training and testing data was taken from the photodegradation study of spent Bayer liquor in an 18 l pilot scale reactor [7]. Typical data patterns are shown in Fig. 4 and the range of the variables studied is summarized in Table 1. Fig. 4(a) shows the effect of catalyst

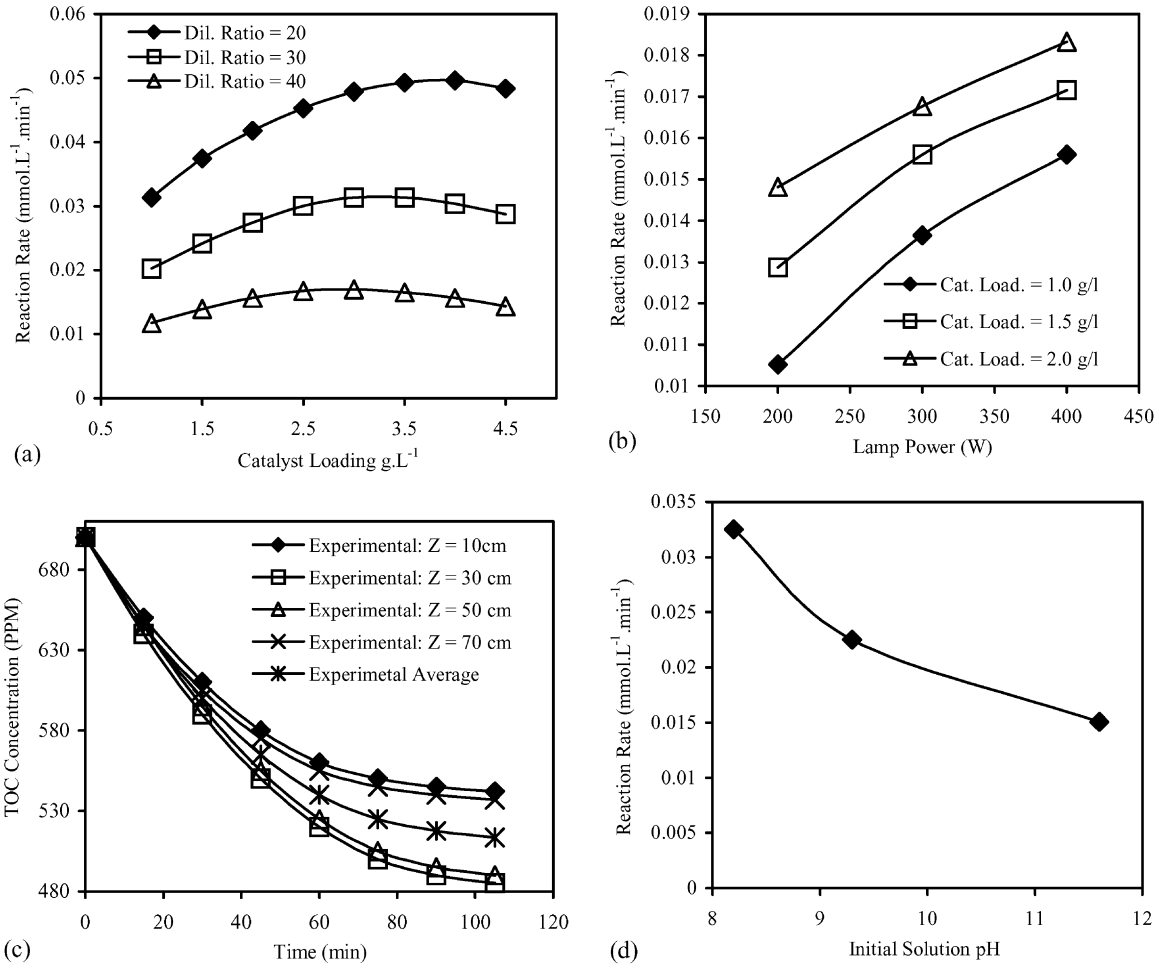

Fig. 4. Typical results from photodegradation of Bayer liquor. (a) Effect of catalyst loading, TOC concentration = 700, 1050 and 1400 ppm, respectively for dilution ratio 40, 30 and 20, lamp power = 200 W, initial solution pH = 11.75. (b) Effect of lamp power, TOC concentration = 700 ppm, initial solution pH = 11.75. (c) Variation in TOC concentration with batch time, initial solution pH = 11.75, lamp power = 200 W, 'Z' is axial position of sampling ports. (d) Effect initial solution pH, TOC concentration = 700 ppm, lamp power = 200 W, catalyst loading = 1.0 g l^{-1}.

loading on reaction rate. Based on the dilution ratio, an optimum catalyst loading between 1.5 and 3.5 g l^{-1} was obtained. The reaction rate increased progressively with lamp power (Fig. 4(b)). The TOC, in Fig. 4(c), decreased from an initial value of about 700–500 ppm in about 100 min duration. A slight axial gradient in TOC concentration (Z = 10, 30, 50 and 70 cm) could be attributed to insufficient mixing and spatial variation of light intensity. An exponential decrease in the reaction rate with initial solution pH (Fig. 4(d)) was consistent with the results obtained with pure sodium oxalate [8]. Out of the several data points generated, 400 data points were selected for neural net training and testing (350 points were chosen for training the neural nets and the remaining 50 data points were used in the validation).

5.2. Training of artificial neural networks

A number of neural net models were trained and tested using the 400 input–output patterns. As shown in Fig. 3, all the models developed in this study consisted of three layers. The first layer was input layer, possessing five-neurons (lamp power, catalyst loading, initial solution pH, liquid residence time (batch time), and concentration of organic species (TOC). The second layer was hidden layer, and it contained variable number of neurons (2–20). The third layer was the output layer and it contained one neuron (the average reaction rate).

The training of ANN is essentially an optimization process, where an error function is minimized by changing the neural net weights. The most widely used error function is the total sum of squared-error defined as [11]

$$E = \frac{1}{2}\sum_{j=1}^{N_s}\sum_{n=1}^{N^o}(y^o_{n,j} - T^o_{n,j})^2 \quad (2)$$

where the indices j and n refer to pattern and output neurons, respectively, $y^o_{n,j}$ the neural net prediction from the nth output neuron for jth input–output pattern, $T^o_{n,j}$ the corresponding target or actual value for the same neuron, N_s the number of datasets, and N^o is the number of output neurons (variables). Whenever the net is subjected to a new training pattern, it calculates the output using the input variables and the error given by Eq. (2) is estimated. In the present case, backpropagation algorithm was used to update the weights [11]. In backpropagation algorithm (see Appendix A) for every input–output pattern, a forward-pass is used to calculate the output and a backward-pass is used to adjust the weight.

5.3. Selection of optimal net configuration

Several net configurations with one hidden layer were trained using the input–output patterns. The trained nets were used to predict the photodegradation rate for 50 input data not previously known to the nets. The following criteria were used to assess the model predictions:

$$C_p = \frac{\sum_{j=1}^{N_s}(R_{p,j} - \bar{R}_{p,j})(R_{a,j} - \bar{R}_{a,j})}{\sqrt{\sum_{j=1}^{N_s}(R_{p,j} - \bar{R}_{p,j})^2 \sum_{j=1}^{N_s}(R_{a,j} - \bar{R}_{a,j})^2}} \quad (3)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{N_s}(R_{p,j} - R_{a,j})^2}{N_s}} \quad (4)$$

$$\mathrm{MAPE} = \frac{1}{N_s}\sum_{j=1}^{N_s}\frac{|R_{p,j} - R_{a,j}|}{R_{a,j}} \times 100 \quad (5)$$

where C_p is the Pearson correlation coefficient, RMSE the root mean square error, MAPE the mean absolute percentage error, R_a the actual value of reaction rate, R_p the predicted value of reaction rate, the subscript 'j' refers to validation pattern number, and N_s is the number of validation datasets.

Table 2
Prediction errors and correlation coefficients

Number of neurons in hidden layer	Correlation coefficient	RMSE[a]	MAPE (%)[b]
2	0.9123	0.0222	7.75
4	0.9938	0.0145	4.61
6	0.9939	0.0265	9.16
8	0.9955	0.0090	3.16
10	0.9930	0.0129	5.24
12	0.9948	0.0098	3.45
14	0.9937	0.0220	8.06
16	0.9955	0.0093	3.18
18	0.9955	0.0094	3.20
20	0.9929	0.0197	7.10

[a] RMSE: root mean square error.
[b] MAPE: mean absolute percentage error.

Table 2 shows the response of various nets to the evaluative criteria. With exception of the neural network with two neurons in the hidden layer, the Pearson correlation coefficient is generally greater than 0.99 suggesting a reasonable fit of the ANN model to the testing data. However, further model discrimination based on the RMSE and MAPE indicates the superiority of ANN models with 8, 16 and 18 neurons in the hidden layer over all other nets. These ANN configurations had not only the highest C_p-values, but also the lowest RMSE and MAPE. Indeed, the configuration with eight-neurons in the hidden layer was clearly the most optimal for the present data, and was therefore, chosen for subsequent investigation. Figs. 5 and 6 illustrate the agreement between predicted and actual values of reaction rate.

5.4. Comparison with regression model

The data used for training the neural nets were also employed in the development of an independent nonlinear

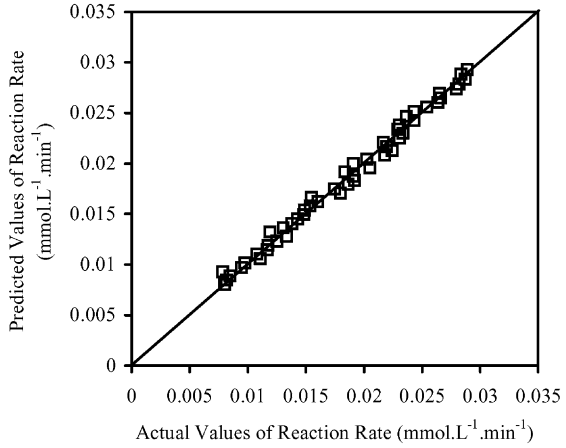

Fig. 5. ANN prediction vs. actual value.

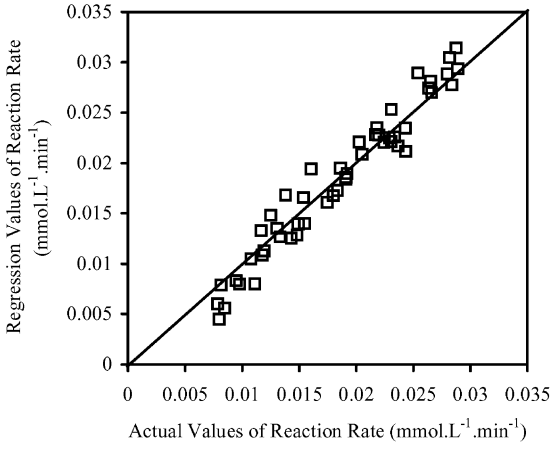

Fig. 7. Prediction from regression model vs. actual value.

regression model. This exercise provides:

$$-\hat{r}_{\mathrm{TOC}} = 0.29 \times W_{\mathrm{lamp}}^{0.75}\, \mathrm{e}^{-0.212\mathrm{pH}} \frac{C_{\mathrm{cat}}}{(a + C_{\mathrm{cat}})^2} \times \frac{(1 - \mathrm{e}^{-0.001\tau})}{\tau} C_{\mathrm{TOC}} \quad (6)$$

where $-\hat{r}_{\mathrm{TOC}}$ is the average reaction rate (mmol l^{-1} min^{-1}), W_{lamp} the lamp power (W), C_{cat} is the catalyst loading (g l^{-1}), τ the liquid residence time or batch time (min), C_{TOC} is the TOC concentration (mmol l^{-1}), and a is the constant with a value of 1.1 g l^{-1}.

A plot of predicted reaction rates using Eq. (6) and actual reaction rates (for the same input pattern as earlier described) is shown in Fig. 7. The comparison of Fig. 7 has a correlation coefficient of about 0.87, which is somewhat poorer than the values obtained with the neural net models. Additionally, the nonlinear regression fit gave higher values of RMSE (0.026) and MAPE (16.2%). On the strength of these findings, it would seem that neural network modeling is a better approach for the modeling of photodegradation kinetics.

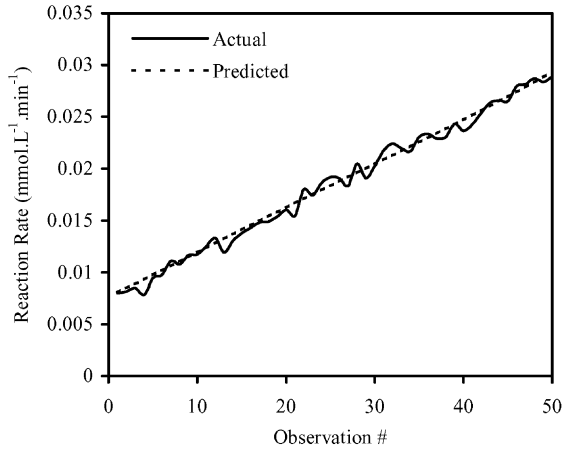

Fig. 6. Line plot of ANN prediction vs. actual value.

5.5. Interpreting neural network weights

The neural net weight matrix can be used to assess the relative importance of the various input variables on the output variables. Garson [12] proposed an equation based on partitioning of connection weights:

$$I_j = \frac{\sum_{m=1}^{m=N^{\mathrm{h}}} \left(\left(|w_{jm}^{\mathrm{ih}}| / \sum_{k=1}^{N^{\mathrm{i}}} |w_{km}^{\mathrm{ih}}| \right) \times |w_{mn}^{\mathrm{ho}}| \right)}{\sum_{k=1}^{k=N^{\mathrm{i}}} \left\{ \sum_{m=1}^{m=N^{\mathrm{h}}} \left(|w_{km}^{\mathrm{ih}}| / \sum_{k=1}^{N^{\mathrm{i}}} |w_{km}^{\mathrm{ih}}| \right) \times |w_{mn}^{\mathrm{ho}}| \right\}} \quad (7)$$

where I_j is the relative importance of the jth input variable on output variable, N^{i} and N^{h} are the number of input and hidden-neurons, respectively, w's are connection weights, the superscripts 'i', 'h' and 'o' refer to input, hidden and output layers, respectively, and subscripts 'k', 'm' and 'n' refer to input, hidden and output neurons, respectively.

The relative importance of various variables as calculated by Eq. (7) is shown in Table 3. As may be seen, all of the variables (lamp power, TOC concentration, catalyst loading, initial solution pH, and liquid residence time or batch time) have strong effects on the average photodegradation rate. Therefore, none of the variables studied in this work could have been neglected from the present analysis. However, as expected the lamp power, initial solution pH and TOC concentration, with relative importance 26.9, 26.1 and 25.7%, respectively, appeared to influence the reaction rate most

Table 3
Relative importance of input variables on the value of average reaction rate

Variable	Importance (%)
Lamp power	26.9
Catalyst loading	12.0
Initial solution pH	26.1
Batch time (liquid)	9.3
TOC concentration	25.7
Total	100

markedly. The light intensity distribution (lamp power) is of paramount importance in the analysis and design of photoreactors [13]. The results of this study further emphasize importance of lamp power. The pH of the reaction mixture in photocatalytic systems is important for reactions involving ionic species [14,15]. The photodegradation of sodium oxalate in Bayer liquor involves redox mechanism, therefore, the initial solution pH was found to be the second most important variable ($I_j = 26.1$).

6. Conclusion

Artificial neural network modeling has been used to investigate the cause effect relationship prevalent in a photodegradation process for the rejuvenation spent Bayer liquor. Neural nets were trained with a set of 350 input–output patterns using three-layered ANN configurations having one hidden layer. The input layer of neurons consisted of five variables—lamp power, catalyst loading, initial solution pH, liquid residence time (batch time), and concentration of organic species (TOC). Although all ANN models examined gave relatively high correlation coefficients (>0.97), a configuration with eight-neurons in the hidden layer was the most optimal considering all three evaluative criteria (correlation coefficient, RMSE and MAPE). An independent nonlinear regression model was also used to describe the data. However, a comparative study revealed that ANN configurations gave better predictions than the regression model indicating that in highly complex systems, neural network modeling may be a better option to data analysis. Analysis also confirmed that all input variables selected have significant effect on the reaction rate, although light intensity, TOC concentration, and initial solution pH had the most important influence.

Acknowledgements

The authors wish to acknowledge the financial support of the Australian Research Council (ARC) for this study. VKP appreciates the provision of a UNSW-International Post-graduate Scholarship as well as a study leave from MREC, Jaipur (India).

Appendix A. Numerical algorithm

1. The training and testing data is normalized using the formula:

$$\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{A.1}$$

where $x_{\min}$ and $x_{\max}$ are minimum and maximum value of the variable 'x'.

2. The output from the kth neuron of the input layer is equal to the input ($\hat{x}_k$) to this neuron:

$$\hat{y}_k^i = \hat{x}_k \tag{A.2}$$

where $k = 1, 2, \ldots, N^{\mathrm{i}}$, and N^{i} is number of input-neurons.

3. Initially, all the connection weights (w_{km}^{ih} and w_{mn}^{ho}) are randomly-generated numbers (having values between 1 and -1).
4. The inputs to the hidden layer are calculated by carrying out a weighted-summation of the all the outputs from the input layer (first layer). Therefore, the input to mth neuron of hidden layer is [16]:

$$x_m^{\mathrm{h}} = \sum_{k=1}^{N^{\mathrm{i}}} w_{km}^{\mathrm{ih}} \hat{y}_k^{\mathrm{i}} \tag{A.3}$$

where $m = 1, 2, \ldots, N^{\mathrm{h}}$, superscripts 'i' and 'h' refer to input and hidden layers, w_{km}^{ih} is the weight corresponding to the connection between kth input-neuron and mth hidden-neuron (Fig. 3), and N^{h} is the number of hidden-neurons.

5. Then the output from the neurons of the hidden layer is calculated using

$$y_m^{\mathrm{h}} = f(x_m^{\mathrm{h}}) = \frac{1}{1 + \exp(-x_m^{\mathrm{h}})} \tag{A.4}$$

where $f(x)$ is sigmoid transfer function defined in Eq. (1).

6. The inputs to the output layer are calculated from weighted-summation of the outputs from the hidden layer:

$$x_n^{\mathrm{o}} = \sum_{m=1}^{N^{\mathrm{h}}} w_{mn}^{\mathrm{ho}} y_m^{\mathrm{h}} \tag{A.5}$$

where $n = 1, 2, \ldots, N^{\mathrm{o}}$, superscript 'o' refers to output layer, w_{mn}^{ho} is the weight corresponding to the connection between mth hidden-neuron and nth output-neuron (Fig. 3), N^{o} the number of output neurons (in the present case, $N^{\mathrm{o}} = 1$).

7. The final output is equal to the input to the output neurons:

$$y_n^{\mathrm{o}} = x_n^{\mathrm{o}} \tag{A.6}$$

8. Calculate the total error between predicted and 'actual target' values of the output variables [11]:

$$E = \frac{1}{2} \sum_{j=1}^{N_s} \sum_{n=1}^{N^{\mathrm{o}}} (y_{n,j}^{\mathrm{o}} - T_{n,j}^{\mathrm{o}})^2 \tag{A.7}$$

where index $j = 1, 2, \ldots N_{\mathrm{s}}$ refers to different data points (input–output pairs), and N_{s} is the number of data points.

If the error 'E' is less than a specified tolerance limit then stop training else go to next step.

9. Calculate the new value of the weights connecting hidden layer and output new layer using generalized delta rule [17]:

$$w_{mn}^{\mathrm{ho}}(p+1) = w_{mn}^{\mathrm{ho}}(p) + \Delta w_{mn}^{\mathrm{ho}}(p) \tag{A.8}$$

where

$$\Delta w_{mn}^{\mathrm{ho}}(p) = \beta \sum_{j=1}^{N_s} (y_{n,j}^{\mathrm{o}}(1 - y_{n,j}^{\mathrm{o}})(y_{n,j}^{\mathrm{o}} - T_{n,j}^{\mathrm{o}}) y_m^{\mathrm{h}}) + \alpha\, \Delta w_{mn}^{\mathrm{ho}}(p-1) \tag{A.9}$$

where α and β are momentum and learning rate parameters (generally, their values are set between 0.6 and 0.9), index 'p' refers to pattern number or iteration number.

10. Similarly, update the weights for the connection between input and hidden layer:

$$w_{km}^{\mathrm{ih}}(p+1) = w_{km}^{\mathrm{ih}}(p) + \Delta w_{km}^{\mathrm{ih}}(p) \tag{A.10}$$

where

$$\Delta w_{km}^{\mathrm{ih}}(p) = \beta \sum_{j=1}^{N_s} \left(y_{m,j}^{\mathrm{h}}(1 - y_{m,j}^{\mathrm{h}}) \sum_{n=1}^{N^{\mathrm{o}}} \delta_{n,j} w_{mn}^{\mathrm{ho}}(p) \right) + \alpha\, \Delta w_{km}^{\mathrm{ih}}(p-1) \tag{A.11}$$

where $\delta_{n,j} = y_{n,j}^{\mathrm{o}}(1 - y_{n,j}^{\mathrm{o}})(y_{n,j}^{\mathrm{o}} - T_{n,j}^{\mathrm{o}})$ is the error gradient for the output layer.

11. Go back to step 4.

References

[1] D.F. Ollis, E. Pelizzetti, N. Serpone, Environ. Sci. Technol. 25 (1991) 1522.
[2] N. Serpone, E. Pelizzetti, Photocatalysis: Fundamentals and Applications, Wiley, New York, 1989.
[3] A. Mills, S. Le Hunte, J. Photochem. Photobiol. A: Chem. 108 (1997) 1.
[4] M. Mehrvar, W.A. Anderson, M. Moo-Young, P.M. Reilly, Chem. Eng. Sci. 55 (2000) 4885.
[5] D.M. Himmelablau, Korean J. Chem. Eng. 17 (2000) 373.
[6] J. Michalopoulos, S. Papadokonstadakis, G. Arampatzis, A. Lygeros, Trans. IChemE 79 (2001) 137.
[7] V.K. Pareek, M.P. Brungs, A.A. Adesina, Ind. Eng. Chem. Res. 40 (2001) 5120.
[8] J. Lea, A.A. Adesina, Chem. Eng. Sci. 54 (1999) 2209.
[9] A. Elkamel, S. Abdul-Wahab, W. Bouhamra, E. Alper, Adv. Environ. Res. 5 (2001) 47.
[10] S.S. Tampe, B.D. Kulkarni, P.B. Deshpande, Elements of Artificial Networks with Selected Applications in Chemical Engineering, and Chemical and Biological Sciences, Simulations and Advanced Control Ltd., Louisville, KY, USA, 1996.
[11] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Nature 323 (1986) 533.
[12] G.D. Garson, AI Expert, April 1991, p. 46.
[13] A.E. Cassano, C.A. Martin, R.J. Brandi, O.M. Alfano, Ind. Eng. Chem. Res. 34 (1995) 2155.
[14] M. Trillas, J. Peral, X. Domenech, Appl. Catal. B: Environ. 5 (1995) 377.
[15] Y. Inel, A.N. Okte, J. Photochem. Photobiol. A: Chem. 96 (1996) 175.
[16] R.P. Lippmann, IEEE ASSP Magazine, April 1987, p. 4.
[17] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, in: D.E. Rumelhart, J.L. McClelland, (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, MIT Press, Massachusetts, 1986, pp. 318–362.