Application of Artificial Neural Network for Modeling Wastewater Treatment Process

A. Sebti[1,2(✉)], B. Boutra[2], M. Trari[1,2], L. Aoudjit[2], and S. Igoud[2]

[1] Chemical Engineering Department, National Polytechnic School, Algiers, Algeria
sebti.aicha@udes.dz, aicha_seb@yahoo.fr
[2] Unité de Développement des Equipements Solaires, UDES, Centre de Développement des Energies Renouvelables, CDER, 42415 Tipaza, Algeria

Abstract. Development of reliable model in chemical engineering facilitates all subsequent steps in process optimization and monitoring operation. Many chemical process intended to wastewater treatment can exhibit complex non-linear behaviour. In this paper, three layered feed forward neural network is used to predict the photo-catalytic degradation yield of solophenyl red, an azo dye widely used in textile industry. The approach adopted to find the optimal topology of the network is based on finding the architectural parameters (the hidden nodes number, the activation function and the training algorithm) that minimize the prediction error. Experimental data required for the development of the network are extracted from the study performed by [1]. In this study a new photo-catalyst have been used to eliminate under solar light the solophenyl red. The result show that the predicted data from the designed optimal neural network architecture was in good agreement with the experimental data. The excellent value of the correlation coefficient attested the accuracy of the model and proved its ability to fit this complex system.

Keywords: Photo-catalysis · Wastewater treatment · Artificial intelligence · Solar energy

1 Introduction

Development of reliable model in chemical engineering facilitates all subsequent steps in process optimization and monitoring operation. Many chemical process intended to wastewater treatment can exhibit complex nonlinear behavior. The use of analytical approach to model such system frequently have their limitations. Various papers have been proved that Artificial Neural Networks (ANNs) can overcome these limitations. ANNs are computational models that mimic the way in which the human brain deal with enormous amounts of sensory information. These empirical models are able to establish the unknown relationship between the input and output data without considering the complex physical and chemical laws governing the process under study [2–4]. They offer fast and accurate solutions to various problems in wide range of disciplines, particularly areas involving prediction, classification and data filtering [5]. An ANN always consists of at least three hierarchical layers of neurons fully

M. Hatti (Ed.): ICAIRES 2019, LNNS 102, pp. 143–154, 2020.
https://doi.org/10.1007/978-3-030-37207-1_15

interconnected and arranged in parallel structure: one input layer, one output layer and one or more hidden layers. Each neuron is linked to certain of its neighbors with weights that represent the strengths of these connections. According to the propagation direction of information, neural networks are categorized into: feed-forward (FFNN) networks and recurrent networks (RNN). In FFNN information flow only in one direction towards the output layer and there is no feedback between neurons. In RNN information pass both forward and backward by introducing loops between certain neurons in the network [3, 6, 7].

Neuron is the basic block in a network structure. It is a processing element which has usually one or more inputs and a single output (Fig. 1). Each neuron has an associated activation function (f) and weight (w_i). The purpose of the processing element is to perform simple computations (summation and multiplication) in order to determine the output signal (y) using the input values received from many other neurons (x_i) as a linear combination [8, 9].

$$y = f(\upsilon) \qquad with \qquad \upsilon = b_1 + \sum_{i=1}^{p-1} x_i * w_i$$

In fact, it applies the activation function to the sum of the weights and sends this signal in a forward direction layer by layer until last layer is reached [2]. The output delivered by the last layer is a prediction of the neural network that is compared to the target.

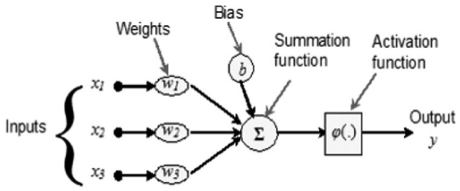

Fig. 1. Model of an artificial neuron [10].

When the predicted result is different of the target then the weight associated to each neuron is updated in an iterative procedure called training or learning. The most commonly learning method used to adjust weights is back propagation (BP). BP consists on calculating the error between the observed and the predicted responses and propagates this error signal in backward direction from the output layer to the input layer. This step is repeated until a convergence criterion is reached.

The purpose of this study is to demonstrate the ability of a three layered feed-forward neural network to model the complex behavior of a photo-catalytic wastewater treatment process.

2 Artificial Neural Network Modeling of Solophenyl Red Photo-Catalytic Degradation

2.1 Experimental Data

In this study, a three layered feed-forward neural network was used to model the photo-catalytic degradation of solophenyl red, an azo dye widely used in textile industry [1]. This organic pollutant is not biodegraded because of the complexity of the chemical structure and its presence in water is harmful to human health. The degradation experiments have been performed under solar light at different experimental conditions using a new catalyst (ZnO/Bentonite). According to the obtained results, the elimination of the SR, using the composite material, was successfully achieved within 160 min and the degradation yield of the azo dye was considered a function of: initial RS concentration, pH, ZnO/Bentonite dosage and irradiation time. The series of experiments conducted in batch reactor by [1] were collected and used for the development of the neural network. In fact, the 116 data were extracted from the experimental study. The four experimental variables (initial RS concentration, pH, ZnO/Bentonite dosage and irradiation time) were selected as inputs of the network whereas the degradation yield of the azo dye was chosen as output. The model variables and their ranges are summarized in Table 1.

Table 1. Neural network model variables and their ranges

Variables	Corresponding range
Initial RS concentration (mg/L)	5–75
pH	2,5–9
ZnO/Bentonite dosage (g/L)	0,25–1
Irradiation time (min)	0–210

In a neural network, the input layer accepts only independent variables. That's why, it's recommended to check whether the four variables are independent or not. In fact, this test reduces the size of the data set and feeds the network with only the most significant inputs by discarding those that are highly correlated. We use the Pearson's correlation coefficient that ranges from -1 (strong negative correlation) and 1 (strong positive correlation) with an insignificant correlation when the coefficient is close to 0 [11, 12].

The correlation coefficients for the SR photo-catalysis input variables are presented in the Fig. 2. The plot figure shows the Pearson coefficient for each pair of variables. The obtained values are close to 0 which indicate quite small correlation that can be neglected. Hence, all experimental variables are retained.

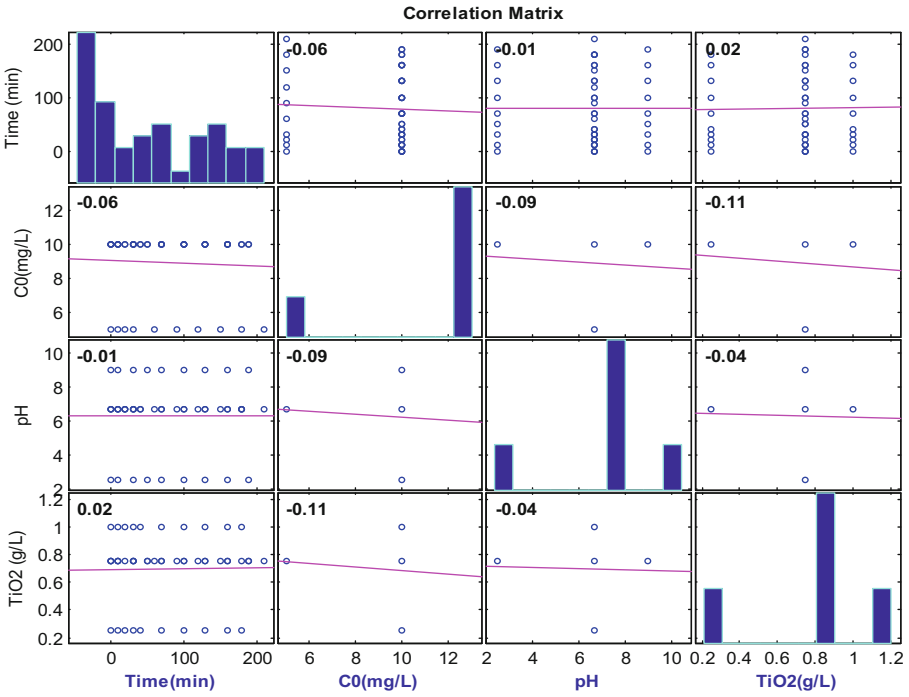

Fig. 2. Correlation coefficients for the solophenyl red photo-catalysis input variables.

In the training phase, the importance of each input variable is related to the magnitude of its range. Therefore, it is indispensable to normalize the raw data in the same range prior to feeding network. We normalize data in order to: equalize the importance of the input variables, minimize the bias and weight values within the network and reduce the training time.

Many techniques of normalization are proposed in the literature. The min-max method is adopted in this study and we normalize the data to fall in the range [−1, 1] [12].

2.2 ANN Model Development

The 116 experimental data were divided into three subsets training (60%), validation (20%) and test (20%). The training subset allowed the network to learn the linear and nonlinear relationships between the inputs and output vectors. The validation subset was employed to evaluate the generalization capability of the network and avoid overfitting. The test data were fed to the network to assessing its prediction performance.

The number of hidden nodes, the activation function and learning algorithm are architectural parameters that affect the performance of the network model. The strategy adopted in this paper to optimize these parameters and determine the best feed forward network topology consists on:

- Varying the number of hidden nodes between 1 and 10.
- Testing four activation functions (logsig, tansig, hardlim and purelin) and ten training algorithms (trainlm, trainscg, trainbr, trainbfg, trainrp, traincgb, traincgf, traincgp, trainoss and traingdx).

This approach has allowed to select the optimal hidden nodes number and the suitable activation function and training algorithm that minimizes the prediction error. The script developed in Matlab has generated 400 topologies. Their performance to simulate the photo-catalysis process under study was evaluated by calculating the validation Root Mean Squared Error (RMSE) and the Perason's correlation coefficient between experimental and predicted data.

$$RMSE = \sqrt{\sum_{p=1}^{P}\sum_{j=1}^{n_L}\left(y_{jp}^{L} - t_{jp}\right)^2}$$

Where:

t_{jp}: Expected target at the jth output neuron for the input p.
y_{jp}^{L}: The output of the jth neuron at layer L

2.3 Results

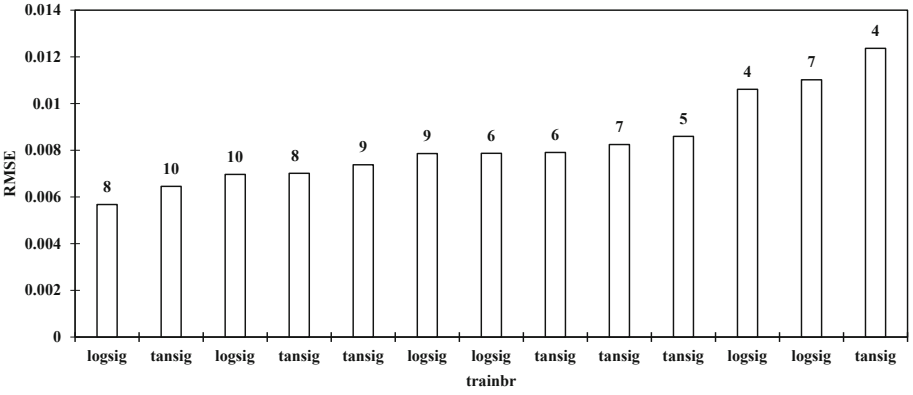

Fig. 3. Variation of RMSE as function of activation function for Bayesian regularization algorithm.

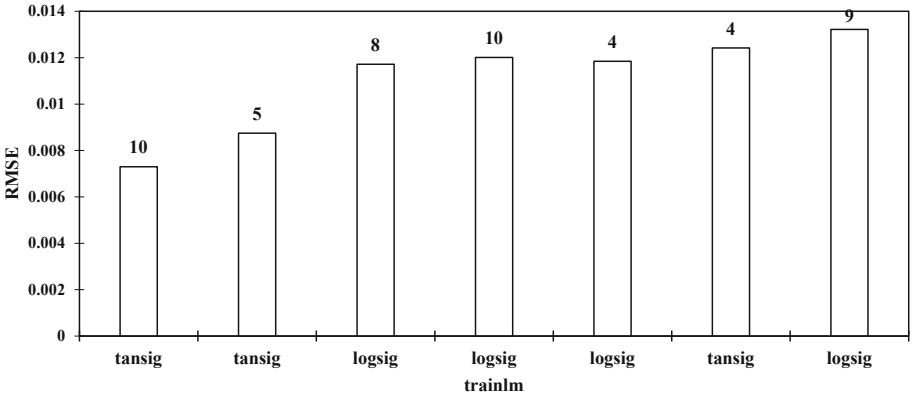

Fig. 4. Variation of RMSE as function of activation function for Levenberg-Marquardt algorithm.

It's important to report that among the 400 networks generated by the Matlab program, 21 topologies predict the photo-catalytic removal efficiency of the pollutant with high accuracy. In fact, the corresponding determination coefficients are over 0.97 and the validation RMSE are ranged between 0.005 and 0.013.

According to the values of the cross-validation mean squared errors and the determination coefficients between measured and predictable data, Bayesian regularization back-propagation (trainbr) and Levenberg-Marquardt (trainlm) show excellent ability to forecast the process efficiency compared to other algorithms (Figs. 3 and 4).

In addition, the two transfer functions namely hyperbolic log sigmoid (logsig) and tangent sigmoid (tansig) give the best fitting models.

The combined effect of the training algorithm and the activation function to improve the performance of the network is well confirmed and the best simulation of the complex system is performed by the network with Bayesian regularization back-propagation training algorithm and hyperbolic log sigmoid transfer function for the hidden layer.

After choosing the suitable combination of training algorithm and activation function, we discuss the neural network optimization with respect to the hidden neurons number.

The Fig. 5 shows the variation of the root mean squared error as function of the hidden neurons number and activation function for the best learning algorithm (trainbr). In fact, the Bayesian regularization algorithm (trainbr) has demonstrated an excellent ability to predict the degradation yield of the azo dye among the ten algorithms tested. In deed the logsig and tansig functions are the most suitable to model this complex and nonlinear system. For the logsig activation function, the best forecasting performance (least RMSE = 0.005 and high R^2 = 0.99) is achieved with eight neurons (4-8-1). Hence, the optimal topology of the feed forward network has four input neurons, eight hidden neurons and one output neuron.

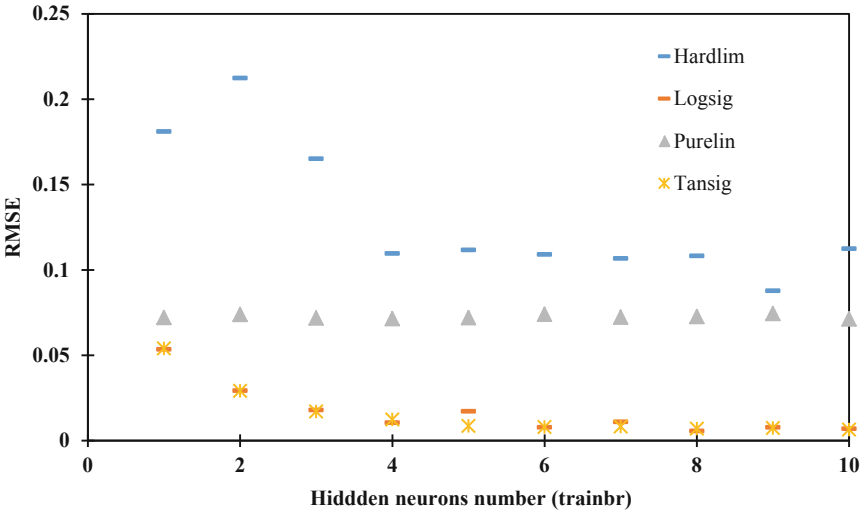

Fig. 5. RMSE as function of hidden nodes number for the trainbr learning algorithm.

In Fig. 6, the variation of the RMSE during the training step is reported as function of the epoch for the optimal topology (4-8-1). 100 training epoch was sufficient to reach the desired performance.

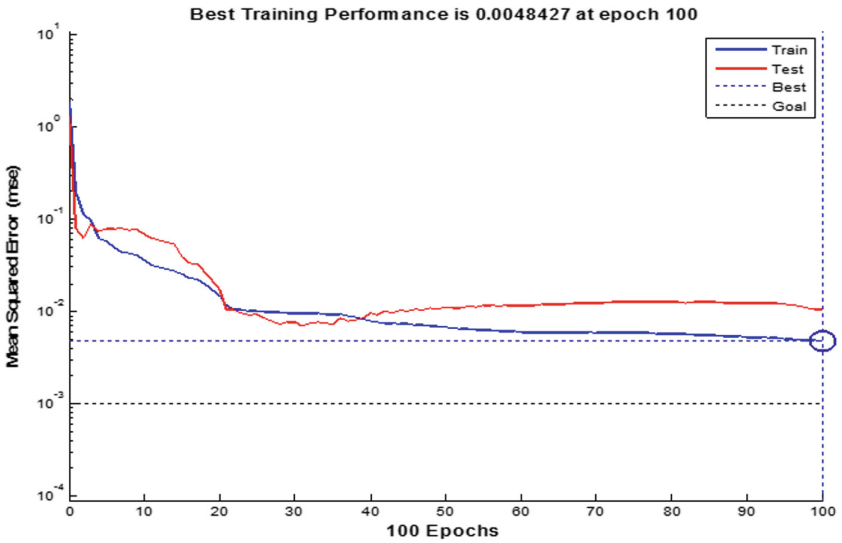

Fig. 6. RMSE as function of training epoch for the optimal network (4-8-1).

For further explanation, the prediction accuracy of the optimal neural network is investigated. Hence, Figs. 7, 8 and 9 illustrate the plots of the experimental results against the predicted ones for training, validation and test sets.

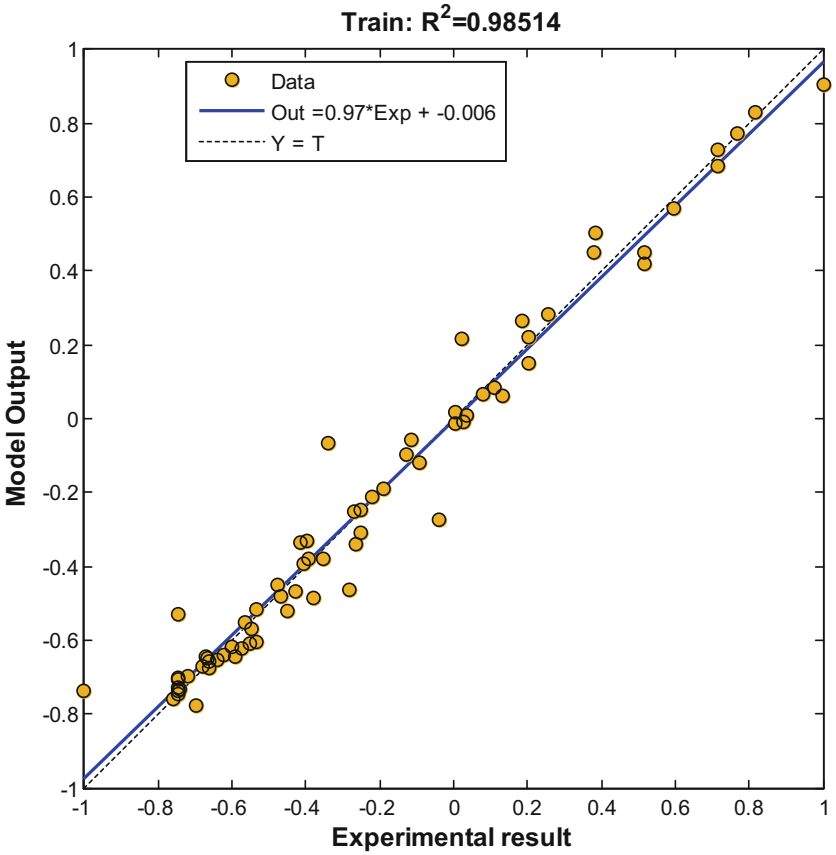

Fig. 7. Experiment results versus predicted ones for training set.

The coefficients R^2 for training, validation and test are 0.985, 0.995 and 0.983 respectively. These high values indicate a strong linear correlation between observed and predicted data. This result confirm the reliability of the developed model to predict the photo-catalytic degradation yield of solophenyl red (Fig. 10).

Artificial neural network are more and more frequently applied in various fields essentially to resolve forecasting problems. Nevertheless, these empirical models are generally regarded to behave as black box systems, unable to clarify the contribution of the independent variables to dependent one. Several authors have thus focused on the analysis of the relative influence of the input variables on the neural network response in order to make ANNs more interpretable [13]. In this paper, the algorithm proposed by Garson is performed. It's a sensitive analysis method which uses combinations of

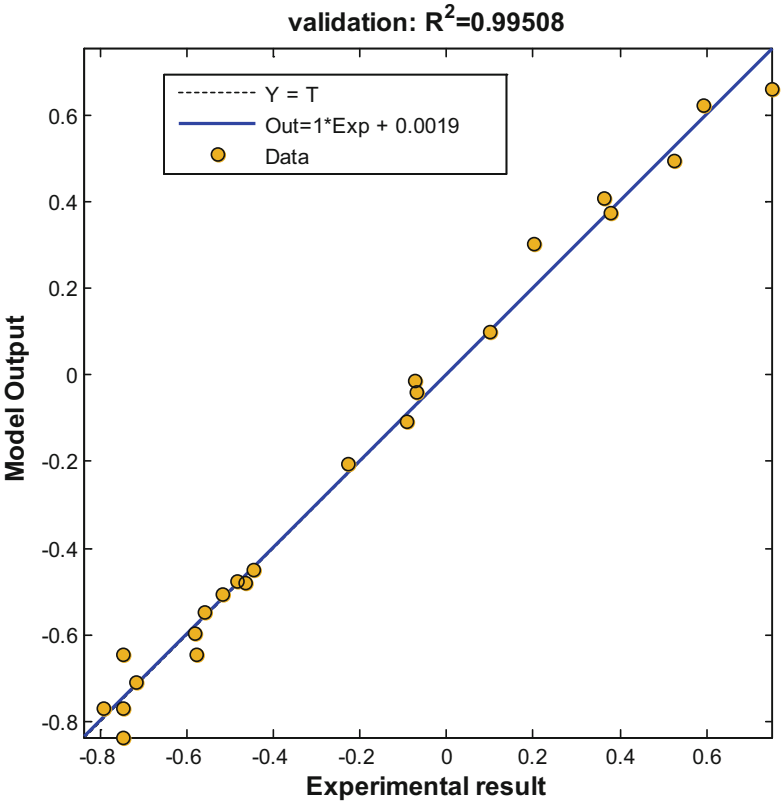

Fig. 8. Experiment results versus predicted ones for validation set.

the absolute values of the weights between layers of the neural network as shown in the formula (1) to obtain the relative importance of the input variable i with respect to the output variable k:

$$GRI_{ik} = \frac{\sum_{j=1}^{m}\left(\frac{|w_{ij}|}{\sum_{i=1}^{n}|w_{ij}|} * |v_{jk}|\right)}{\sum_{i=1}^{n}\left(\sum_{j=1}^{m}\left(\frac{|w_{ij}|}{\sum_{i=1}^{n}|w_{ij}|} * |v_{jk}|\right)\right)}$$

where i, j, k, respectively, refers to input layer, hidden layer and output layer neurons; w_{ij} is the connection weights between input layer and hidden layer neurons, v_{ij}, is the connection weights between hidden layer and output layer neurons, m is the total number of input neurons and n is the total number of hidden layer neurons.

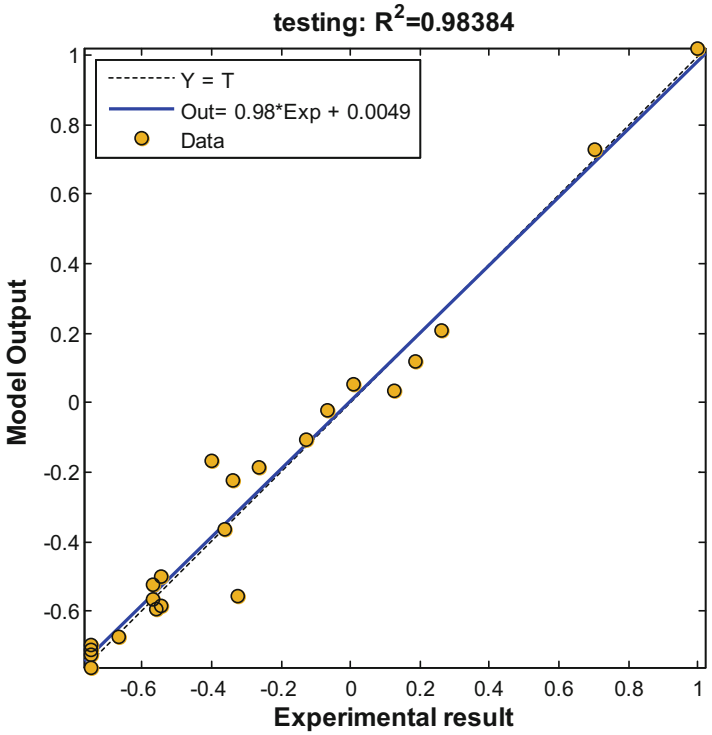

Fig. 9. Experiment results versus predicted ones for validation set.

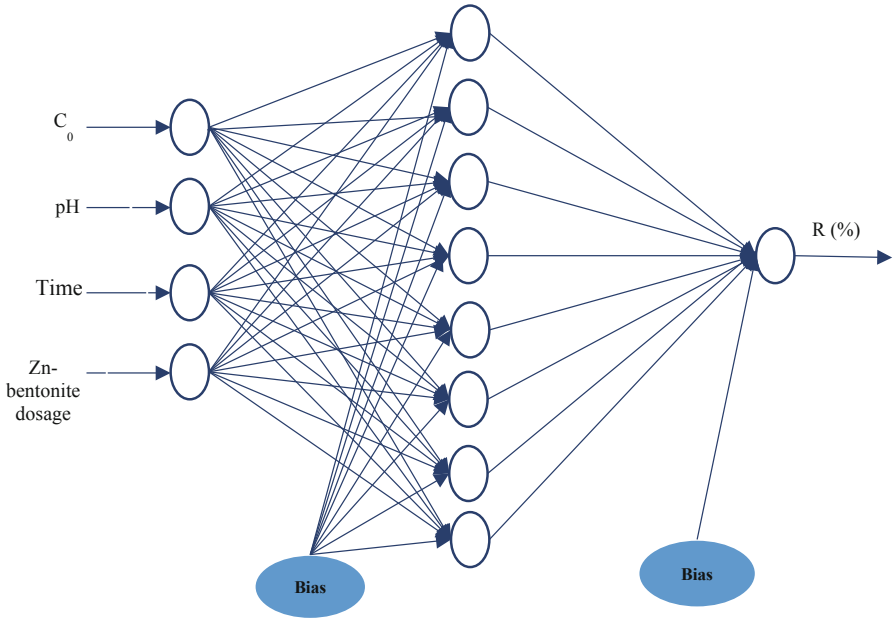

Fig. 10. Structure of the optimal neural network (4:8:1).

The weights matrix of the optimal neural network is given in Table 2.

Table 2. Weights matrix of the optimal neural network.

Weights and bias : Input layer – Hidden layer						Weights and bias: Hidden layer – Output layer		
Neuron	Inputs				Bias	Neuron	Weights	bias
	Irradiation time (h)	Initial solophenyl red concentration (mg/L)	Initial pH	Catalyst amount ZnO-Bentonite (g/L)				
1	−3,35	2,14	−1,31	2,07	−4,13	**1**	−3,46	−0,28
2	−0,44	−1,45	3,65	1,62	−0,76	**2**	3,99	
3	1,19	0,01	−1,78	−0,28	0,53	**3**	2,17	
4	3,86	−1,99	3,21	−1,22	4,02	**4**	−3,19	
5	1,54	1,34	0,40	0,05	3,20	**5**	2,54	
6	−0,28	−2,52	2,30	1,45	0,23	**6**	−2,53	
7	−1,64	1,57	1,64	3,94	1,26	**7**	−2,09	
8	−1,56	−0,31	−0,15	1,49	0,34	**8**	1,59	

The main objective of applying the Garson algorithm is to rank the four variables in order of the relative share of their contribution to the prediction of the solophenyl red removal efficiency by photo-catalytic under solar radiation. The obtained results are summarized in the Table 3.

Table 3. Relative importance of the process input variables

Input variable	Relative importance (%)	Rank
Irradiation time (h) (g/L)	7,10	4
Initial solophenyl red concentration (mg/L)	**58,78**	**1**
Initial pH	8,27	3
ZnO-Bentonite dosage (g/L)	25,84	2

Through the results of the sensitive analysis, it is possible to observe that the considered input variables possess strong effect on the pollutant removal. The initial concentration of solophenyl red is the most influential variable with relative importance of 58%. This parameter is followed by the catalyst loading (25%) and initial pH (8%).

3 Conclusion

The complexity of the photo-catalytic wastewater treatment process makes difficult the use of analytical approach for modelling purpose. Hence, in this work development of artificial neural network model was performed to predict the degradation yield of the solophenyl red azo dye. The strategy based on varying the three architectural parameters namely: hidden neurons number, activation function and learning algorithm has been adopted to optimize the network topology able to reproduce the experimental data. The best prediction performance was achieved with eight hidden neurons, logsig activation function and Bayesian regularization algorithm. This optimal architecture allowed to forecast the observed data with correlation coefficient of 0.995.

References

1. Boutra, B., Trari, M.: Solar photodegradation of a textile azo dye using synthesized ZnO/Bentonite. Water Sci. Technol. **75**(5), 1211–1220 (2017)
2. Dreyfus, G., Martinez, J.M., Samuelides, M., Gordon, M.B., Badran, F.: Réseaux de neurones, Méthodologie et applications, France (2004)
3. Bolanca, T., Ukic, S., Peternel, I., Kusic, H., Bozic, A.L., et al.: Artificial neural network models for advanced oxidation of organics in water matrix-comparison of applied methodologies. Indian J. Chem. Techn. **21**(1), 21–29 (2014)
4. Bagheri, M., Mirbagheri, S.A., Bagheri, Z., Kamarkhani, A.M.: Modeling and optimization of activated sludge bulking for a real wastewater treatment plant using hybrid artificial neural networks-genetic algorithm approach. Process Saf. Environ. Prot. **95**, 12–25 (2015)
5. Croall, I.F., Mason, J.P. (eds.): Industrial Applications of Neural Networks: Project ANNIE Handbook. Springer, Berlin; New York (1992)
6. Al Shamisi, M.H., Assi, A.H., Hejase, H.A.: Using MATLAB to Develop Artificial Neural Network Models for Predicting Global Solar Radiation in Al Ain City-UAE. INTECH Open Access Publisher (2011)
7. Burney, S.M.A., Jilani, T.A., Ardil, C.: A comparison of first and second order training algorithms for artificial neural networks. In: International Conference on Computational Intelligence, pp. 12–18 (2004)
8. Touzet, C.: les réseaux de neurones artificiels, introduction au connexionnisme. EC2 (1992)
9. Ammar, M.Y.: Mise en ø euvre de réseaux de neurones pour la modélisation de cinétiques réactionnelles en vue de la transposition batch/continu. Ph.D. thesis (2007)
10. Magoulas, G.D., Vrahatis, M.N.: Adaptive algorithms for neural network supervised learning: a deterministic optimization approach. Int. J. Bifurcat. Chaos **16**(7), 1929–1950 (2006)
11. Mjalli, F.S., Al-Asheh, S., Alfadala, H.E.: Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. J. Environ. Manag. **83**(3), 329–338 (2007)
12. Sebti, A., Souahi, F., Mohellebi, F., Igoud, S.: Experimental study and artificial neural network modeling of tartrazine removal by photocatalytic process under solar light. Water Sci. Technol. **76**(2), 311–322 (2017)
13. Olden, J.D., Joy, M.K., Death, R.G.: An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Model. **178**(3), 389–397 (2004)