Modelling biological oxygen demand in wastewater of Arak City, Iran

Mohamad Sakizadeh, Eshagh Shayegan and Mojtaba Sheykh Lar

ABSTRACT

Due to the fact that the laboratory analysis of biological oxygen demand (BOD) is time-consuming and uncertain because of some interferences during analysis, automatic estimation of BOD by modeling methods would be of great importance to researchers. The main aim of this study was to compare the performance of three models for the prediction of BOD in the wastewater of Arak City, Iran. The methods were artificial neural network (ANN with early stopping and ANN with Bayesian regularization), an Ensemble of ANN (EANN), and partial least squares regression. The models trained and were validated on a data set containing 18 parameters sampled periodically from the wastewater of Arak City. The performance of these models was assessed by mean squared error (MSE) and R^2 of the test data set besides correlation coefficient of observed and predicted values of BOD for each model. Concerning the Bayesian regularization algorithm, the perfect fitting of the model to both training and test data with correlation coefficients of 0.999 and 0.945 confirmed that this method outperformed the results of other models. Moreover, the results of sensitivity analysis indicated that chemical oxygen demand (COD), next to sulfate and fecal coliforms are the most important parameters in the prediction of BOD by ANN modeling.

Key words | artificial neural network, biological oxygen demand, partial least squares regression

Mohamad Sakizadeh (corresponding author)
Eshagh Shayegan
Mojtaba Sheykh Lar
Department of Environmental Sciences, Faculty of Sciences,
Shahid Rajaee Teacher Training University,
Shahid Shabanloo Avenue, Lavizan,
Tehran,
Iran
E-mail: *msakizadeh@gmail.com*

INTRODUCTION

The proper treatment of wastewater is one of the ongoing challenges in developing countries. In Iran, during the last few years, an enormous amount of money has been spent on the establishment of new wastewater treatment systems especially in metropolitan areas. Arak City is the capital of Markazi Province with a population of about 600,000 inhabitants. Municipal and industrial wastewaters of this city are directed to a wastewater treatment plant through sewers and the basic wastewater treatment process in Arak is a stabilization pond (Naddafi *et al.* 2009). Information about the quality of wastewater before discharging to surface and groundwater resources is of great importance. The Iranian Department of Environment monitors the quality of the physicochemical and biological parameters of Arak's wastewater. One of the most important parameters of wastewater is biological oxygen demand (BOD) that measures the approximate amount of biodegradable organic matter present in wastewater and serves as an indicator parameter for the extent of wastewater treatment. The oxygen demand resulting from the respiration of algae in the sample and the possible oxidation of ammonia in addition to the presence of toxic substances in the sample are some of the interferences that make the laboratory analysis of this method prone to measurement error and uncertainty (Basant *et al.* 2010). Moreover, analysis of this parameter takes a long time (five consecutive days) making it one of the most labor-intensive methods as well. These problems encourage researchers to develop methods for automatic prediction of this parameter through the available information associated with the other

doi: 10.2166/washdev.2015.169

physicochemical and biological parameters. In this field, the partial least squares (PLS) regression and artificial neural networks (ANNs) modeling approaches have the ability to relate the input and output variables (Basant *et al*. 2010). Awareness of neural modeling in the environmental field is becoming increasingly evident given the emerging number of published studies and reference works cataloged in the literature (e.g. Panda *et al*. 2004; Nour *et al*. 2006; Ozkaya *et al*. 2007; Chen *et al*. 2008; Boszke & Astel 2009; May & Sivakumar 2009; Hanrahan 2011; Wieland et al. 2012; Kisi *et al*. 2013). An artificial neural network (ANN), or simply neural network, is a type of artificial intelligence (computer system) that attempts to mimic the way the human brain processes and stores information. It works by creating connections between mathematical processing elements, called neurons. Knowledge is encoded into the network through the strength of the connections between different neurons, called weights, and by creating groups, or layers, of neurons that work in parallel. The system learns through a process of determining the number of neurons or nodes and adjusting the weights for the connections based upon training data. Details of the neural network, including different algorithms for network training, can be found in detail in the extensive published literatures in this field (e.g. Dreyfus 2005; Taylor 2006; Hsieh 2009). In theory, a neural network with one hidden layer with a sufficient number of hidden neurons is capable of approximating any continuous function (Cybenko 1989; Hornik *et al*. 1989). On the contrary, partial least squares (PLS) regression is a powerful linear regression technique for working with noisy and highly correlated data and in cases in which there is a limited number of observations (Cote *et al*. 1995).

Prediction of biological oxygen demand has been implemented earlier through artificial neural networks (e.g. Dogan *et al*. 2009; Singh *et al*. 2009; Emamgholizadeh *et al*. 2014) and partial least squares regression (PLSR) both in natural ecosystems and wastewater treatment systems (e.g. Oliveria-Esquerre *et al*. 2003, 2004; Lee *et al*. 2006; Basant *et al*. 2010). Considering the foregoing discussion, the main aim of this study was to compare the performance of three models for the prediction of BOD in the wastewater of Arak City, Iran. The first method is based on an artificial neural network (ANN with early stopping and ANN with Bayesian regularization), the second method is based on an Ensemble ANN (EANN), and the third method is partial least squares regression (PLSR). The first two methods are nonlinear methods whereas the third method is a linear method.

METHODS

Study area and data description

Arak City has a population of about 600,000 inhabitants. It is located in the southwest of Tehran between mountain ranges on the south, west, and southwest (Figure 1). With an elevation of 1,700 m above sea level, it has a cold and dry (continental) climate. The maximum temperature may rise up to 35 degrees Celsius in summer and may fall to below –25 degrees Celsius in winter. The average rainfall is around 350 mm and the annual relative humidity is 46%. The data used in this study included 19 wastewater parameters (BOD, Chemical oxygen demand [COD], NH3, Fluoride, pH, Total suspended solids [TSS], Total dissolved solids [TDS], Total solids [TS], Electrical conductivity [EC], Alkalinity, Phosphate, Nitrate, Total hardness, Chloride, Sulfate, Ca, Mg, Total coliform and Fecal coliform) which were sampled periodically from the wastewater of Arak City since 2009 by the Department of Environment of Markazi Province. The descriptive statistics associated with each parameter have been given in Table 1.

Partial least squares regression

In this study, partial least squares regression (PLSR) using the SIMPLS Algorithm was performed to predict the biological oxygen demand based on the 19 wastewater treatment parameters. Variables were standardized firstly to overcome the problem of their large scale variance, so the modeling process was actually based on the correlation matrix of the original data (Vinzi *et al*. 2010). PLSR works by finding latent variables called X-scores which are orthogonal, while capturing most of the variance in the original data. The latent variables are good predictors of Y and summaries of X. The latter matrices (e.g. X and Y) are modeled by these latent variables. The X-scores are denoted

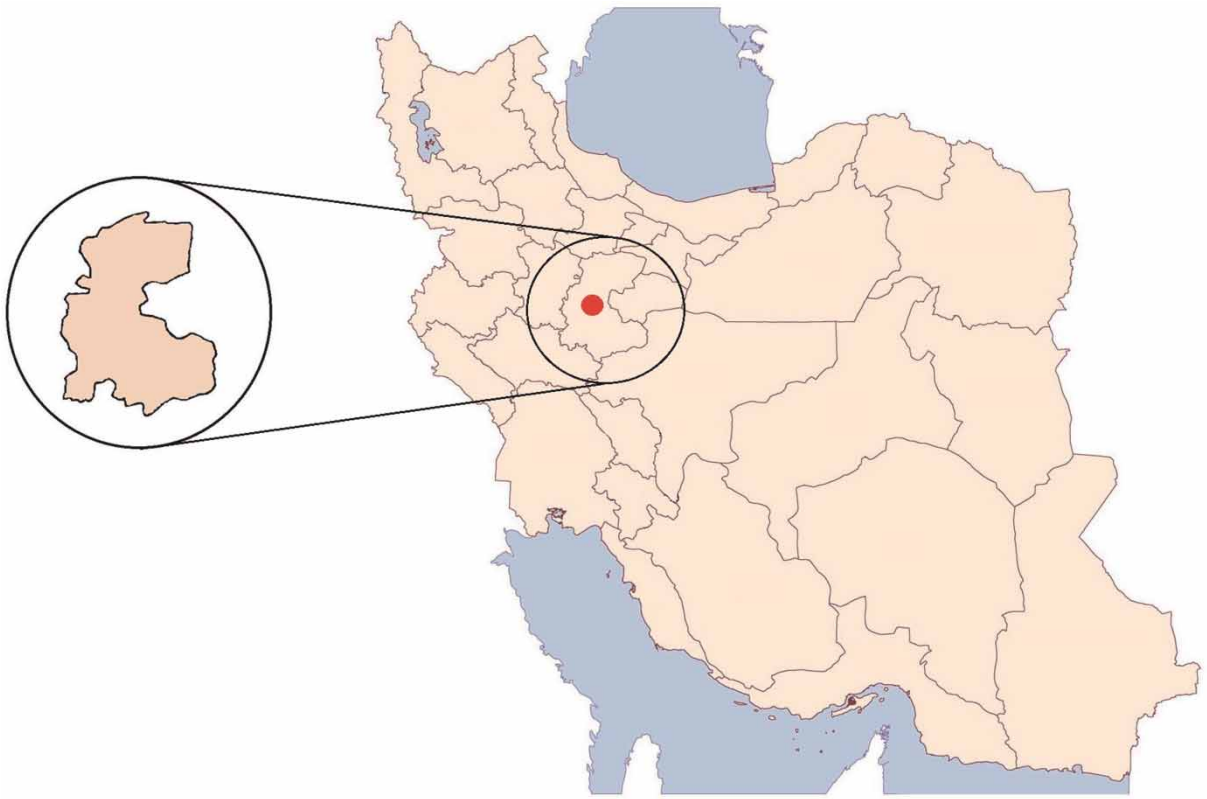

Figure 1 | A sketch map of the study area including Arak City located in Markazi Province.

by $t_a (a = 1, 2, \ldots, A)$ and defined as follows:

$$t_{ia} = \sum_k W^*{}_{ka} X_{ik} \tag{1}$$

In which W^*_{ka} is the weight(s) and X_{ik} is calculated with the following equation:

$$X_{ik} = \sum_a t_{ia} p_{ak} + e_{ik} \tag{2}$$

In the later equation, p_{ak} and e_{ik} are the associated loading and X-residuals respectively. Similarly for the multivariate Y, we will have:

$$y_{im} = \sum_a c_{ma} t_{ia} + f_{im} \tag{3}$$

where c_{ma} and f_{im} are the weight and Y-residuals. If we combine Equations (1) and (3), a multiple regression model is produced as follows:

$$y_{im} = \sum_a c_{ma} \sum_k W^*{}_{ka} X_{ik} + f_{im} = \sum_k b_{mk} x_{ik} + f_{im} \tag{4}$$

In Equation (4), b_{mk} are PLS-regression coefficients. The PLS model performed in two stages, uses a set of calibration (training) samples to construct the model, which is employed to compute a set of regression coefficients. These coefficients are then used to make predictions of the dependent variables in new (test) experimental data. The optimum number of PLS components was selected on the basis of the percent of variance explained in X and captured by the model (Y) (Basant *et al.* 2010).

Development of the ANN model

To keep within the scope of this paper, we limited our survey of neural network (NN) models to the feed-forward

Table 1 | Descriptive statistics of wastewater quality parameters in Arak City

Wastewater quality parameters	Mean	Standard deviation	Min	Max
BOD (mg/l)	25.05	16.81	6.00	78.36
COD (mg/l)	50.12	26.29	19.50	162.10
Fluoride (mg/l)	0.87	0.55	0.10	2.50
NH3 (mg/l)	28.45	15.17	0.23	50.20
pH	7.83	0.30	7.40	8.46
TSS (mg/l)	76.35	61.15	4.00	316.67
TDS (mg/l)	751.42	103.07	588.40	1,012.90
TS (mg/l)	845.34	110.92	673.90	1,076.70
EC (μs/cm)	1,431.30	306.26	1,000.00	2,299.00
Alkalinity (mg/l)	323.37	112.60	137.20	466.40
Phosphate (mg/l)	17.38	11.09	1.00	57.00
Nitrate (mg/l)	13.37	9.22	1.60	30.00
Total hardness (mg/l)	318.86	50.11	233.90	434.00
Chloride (mg/l)	201.85	61.01	11.77	306.20
Sulfate (mg/l)	118.44	51.83	40.00	230.00
Ca (mg/l)	83.24	13.56	64.00	115.00
Mg (mg/l)	35.47	43.25	5.46	243.76
Total coliform (MPN/100 ml)	437.23	429.85	10.00	1,483.30
Fecal coliform (MPN/100 ml)	127.68	169.02	2.00	536.70

neural network with one hidden layer. The linear transfer function (e.g. $yi = xi$) and the following transfer function were used for the output and hidden layers, respectively:

$$y_j = \tan h\left(\sum_{i=1}^{d} w_{ij}x_i + b_j\right) \quad (5)$$

where w_{ij} and b_j are the weight and bias parameters in which 'i' and 'j' subscripts refer to the input and neuron respectively. In addition, the Levenberg-Marquardt algorithm was used to update the weight and bias of the network according to this formula:

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e \quad (6)$$

where J is the Jacobian matrix, which contains first derivatives of the network errors with respect to the weights and biases, e, is a vector of network errors, I, is the identity matrix, x, is a vector containing weights and biases, and μ is a scalar value. Prior to the data introduction to the neural network, standardization of the data (i.e. the data have zero mean and unit standard deviation) was carried out according to the following equation:

$$Z_i = \frac{x_i - \bar{x}_i}{s_i} \quad (7)$$

In which, $\bar{x}_i$ and s_i are the mean and standard deviation of the observed variables, respectively, whereas Z_i is the standardized value. One of the problems during neural network modeling is determination of the best network structure. The number of input predictors and output variables impose the number of input and output units (neurons) however a problem arises when we intend to select the optimum number of hidden nodes which is actually a trial and error process as there is no universally applicable rule concerning the optimal number of hidden nodes (Anctil *et al.* 2004; Palani *et al.* 2008; Singh *et al.* 2009; Basant *et al.* 2010). Increasing the number of hidden neurons incurs the danger of over-fitting leading to poor out-of-sample generalization error, while decreasing it to less than the optimum structure results in under-fitting. Over-fitting occurs when a forecasting model has few degrees of freedom. In other words, it has relatively few observations in relation to its parameters and therefore it is able to memorize individual points rather than learn the general patterns (Baum & Haussler 1989), so it has poor out-of-sample generalization error. In this study, different hidden node sizes ranging from 5 to 30 were applied, and given the optimum number of hidden nodes (based on the minimum mean squared error [MSE]), the best performed ANN was used for out-of-sample BOD prediction.

Early stopping

When the objective function evaluated over the validation data reaches a minimum, it gives a useful signal that this is the appropriate time to stop the training, as additional training epochs only contribute to over-fitting. This very common approach is called early stopping (Piotrowski & Napiorkowski 2013).

In this study, we randomly retained 38.5 percent of the original data as the test set to ensure that a completely independent set was used for out-of-sample error evaluation. Following division of the data to training (70 percent), validation (15 percent), and test (15 percent) set, the above-mentioned algorithm was implemented on the rest of the data. The developed neural network trained 20 times and the generalization error of the model, based on MSE, was considered on the test set. Since during the training the weights and biases of the ANN are reinitialized each time and the data are divided randomly to training, validation and test set, so data that is a test set for one network will likely be used for training or validation by other neural networks. That is why the original data were divided into two parts to ensure that a completely independent test set was preserved (Beale *et al*. 2013).

Ensemble averaging in conjunction with early stopping

As described by Khalil *et al*. (2011), the ANN ensemble is a group of ANNs that are trained for the same problem, followed by combination of ANN's results to produce the ANN ensemble output. Several methods can be used to generate an ensemble of neural networks as suggested by Shu & Ouarda (2007). In this study we manipulated the initial random weights to construct an ensemble of neural networks. In this field, the previous studies proved that, if the ensemble techniques are appropriately selected, it will improve the performance of a single ANN (e.g. Shu & Burn 2004; Zaier *et al*. 2010; Cheng *et al*. 2012).

The optimum ensemble size is a calibration parameter however, as suggested by Hansen & Salamon (1990), Agrafiotis *et al*. (2002) and Shu & Burn (2004), since beyond the neural network size of 10, no improvement is observed in generalization ability, an ensemble size of 10 was used in the present paper.

In this research, a completely independent data set, accounting for 38.5 percent of the original data, was selected randomly. Then, the same network architecture was trained 10 times on the remaining part of the data using early stopping, with different data divisions for training, validation and test and different weight and bias initialization. Finally, the performance for the average output was considered on the second independent data set.

Bayesian regularization

One of the methods to improve out-of-sample error (generalization) is regularization. This involves modifying the performance function, which is normally chosen to be the sum of squares of the network errors on the training set as follows:

$$\mathrm{mse} = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(t_i - a_i)^2 \tag{8}$$

where t_i and a_i are the target and predicted values of the water quality parameter (BOD in this case). In this method, the improvement of generalization is implemented through modifying the performance function by adding a term that consists of the mean of the sum of squares of the network weights and biases:

$$\mathrm{msereg} = \gamma\mathrm{mse} + (1-\gamma)\mathrm{msw} \tag{9}$$

In which γ is the performance ratio and

$$\mathrm{msw} = \frac{1}{n}\sum_{j=1}^{n} w_j^2 \tag{10}$$

where w_j is the associated weight of the neural network. Using this performance function will cause the network to have smaller weights and biases, and this will force the network response to be smoother and less likely to over-fit.

Sensitivity analysis

The relative importance of each wastewater quality parameter on prediction of BOD was considered using a sensitivity analysis. In model building studies, sensitivity analysis generally refers to assessment of the importance of predictors in the fitted models. During this process, the variables are usually ranked according to the deterioration of the model performance criterion (e.g. MSE in this case) if a variable is removed from the model. This analysis is helpful for the identification of less important variables to be removed or ignored in subsequent studies in addition to the most essential variables (Gazzaz *et al*. 2012). The leave-

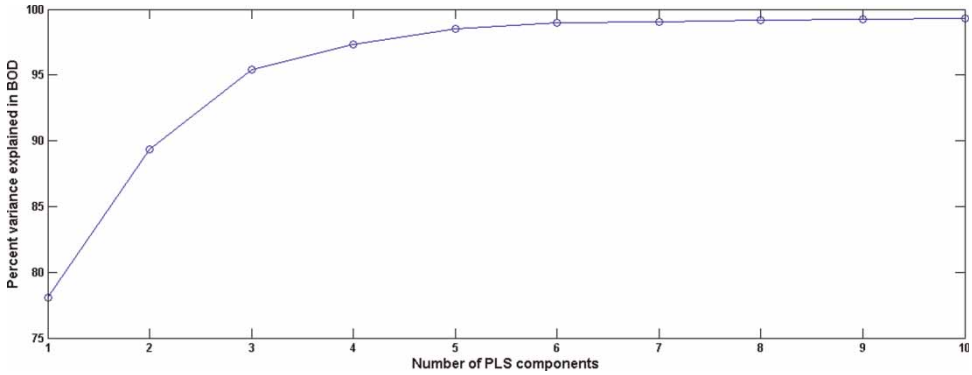

Figure 2 | Percent of variance explained in BOD against the number of PLS components.

one-out method was implemented with ANN modeling which corresponds to assessing changes in the network error that will be obtained if each input variable is removed one at a time (e.g. Sahoo *et al*. 2006). All the computations in this study were performed using MATLAB(R2013b) (MathWorks, Inc., Natwick, MA).

RESULTS AND DISCUSSION

To determine the optimum number of PLS components, the percent of variance explained in BOD was plotted against the number of PLS components (Figure 2). Considering this figure, it can be concluded that with 5 PLS components, 98.5 percent of variance in the biological oxygen demand has been covered, while thereafter the variance explained by each PLS components was negligible. Therefore, the parameters of PLSR were recalculated with 5 PLS components. The regression coefficients and performance of PLS for 10 and 5 PLS components have been presented in Table 2. Those variables with significant contribution to the BOD prediction are COD, Total hardness and Fecal coliform considering the coefficients for 5 PLS components whereas the least significant variables (i.e. regression coefficient larger than 0.05) as suggested by Kovoor and Nandagiri (2007) are TSS, TDS, EC, Chloride, Ca and Mg. The model yielded R^2 values of 0.98 and 0.82 for the training and test data with 5 PLS components whereas they were 0.99 and 0.70 in the case of 10 PLS components indicating a better out-of-sample prediction with 5 PLS components (Table 2). The measured and predicted values of BOD for training and test data have been illustrated in Figure 3. The high correlation coefficient ($r = 0.99$) of the training data suggested a good fit to the training set using the PLS model. On the contrary, with respect to the test data set, for the low values of BOD the model is linear whereas for medium and high values, the

Table 2 | The regression coefficients and performance of PLS for 10 and 5 PLS components

Variables	Coefficients for 10 PLS components	Coefficients for 5 PLS components
COD	0.64	0.71
Fluoride	0.07	0.08
NH3	0.13	0.19
pH	–0.14	–0.11
TSS	0.04	0.03
TDS	–0.02	–0.03
TS	–0.17	–0.09
EC	0.05	–0.02
Alkalinity	0.13	0.12
Phosphate	0.10	0.12
Nitrate	0.07	0.10
Total hardness	–0.35	–0.23
Chloride	0.06	0.01
Sulfate	0.30	0.18
Ca	0.00	0.02
Mg	0.13	0.03
Total coliform	0.10	0.08
Fecal coliform	–0.44	–0.33
Intercept	0	0
test R^2	0.702	0.820
training R^2	0.993	0.985

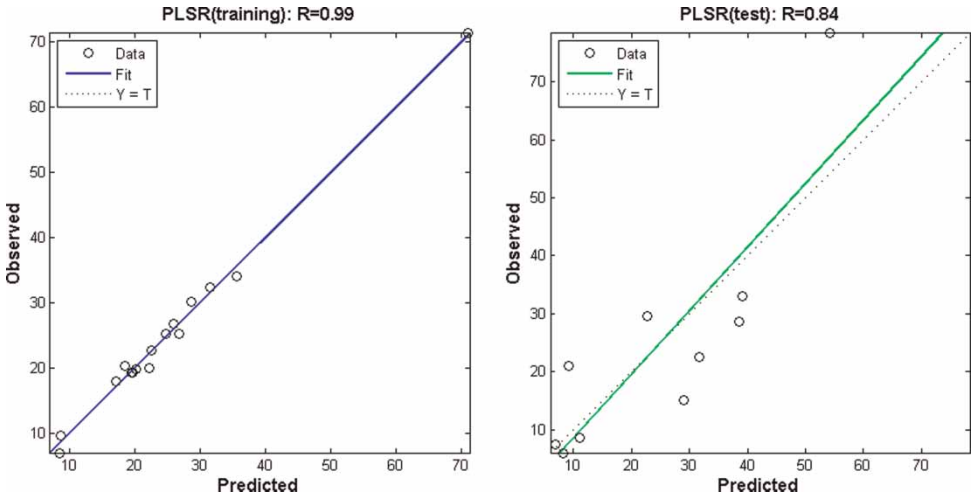

Figure 3 | The correlation coefficient between observed and predicted BOD by PLSR modeling.

data are scattered indicating that a nonlinear model may have a better out-of-sample generalization error. The low correlation coefficient ($r = 0.84$) confirms this to some extent.

On the other hand, the results of other modeling methods namely ensemble averaging, early stopping and Bayesian regularization for different numbers of hidden nodes have been given in Table 3. Considering ensemble averaging, the minimum MSE was obtained for an ensemble size containing 15 hidden neurons. The associated MSE for the training and test data were 32.38 and 30.63, whereas beyond this size the MSE increased accordingly. This shows that ensemble averaging enabled the ANNs to generalize and thereby the accuracy of out-of-sample generalization of this method (for the test data) was similar to the accuracies of the training data.

The optimum number of hidden nodes for ANN with the Bayesian regularization algorithm was the same (15 nodes). However, in this case since the MSE of the training set for lower numbers of hidden nodes was considerably higher than that of 15 hidden nodes, it might have been a sign of under-fitting of the model to the training set, so a neural network structure with 15 hidden nodes size was deemed as the optimum size. This result proved that both too few and more than enough hidden nodes may contribute to under-fitting and over-fitting, respectively. On the other hand, the best structure for early stopping was a 25 hidden nodes size while with respect to this method, the training and test MSEs were 11.79 and 36.20, respectively.

Given the optimum structure for each modeling method, we plotted the model's predicted values against the observed BOD for the training and test parts separately (Figure 4). Regarding early stopping, the correlation coefficients between the observed and predicted values of BOD for the training and test set were 0.985 and 0.790, respectively.

Table 3 | The results of ensemble averaging, early stopping and Bayesian regularization for different numbers of hidden nodes

		Number of hidden nodes					
Type of NN models	**Performance base on MSE**	**5**	**10**	**15**	**20**	**25**	**30**
Ensemble averaging	Test	88.79	82.55	30.63	120.74	121.65	190.28
	Training	103.86	47.21	32.38	49.44	67.12	57.27
Early stopping	Test	92.31	73.18	65.68	67.22	36.20	96.57
	Training	4.41	7.03	11.63	21.89	11.79	8.20
Bayesian regularization	Test	64.01	42.52	80.21	81.81	324.82	87.99
	Training	12.95	14.99	0.11	0.13	0.63	10.52

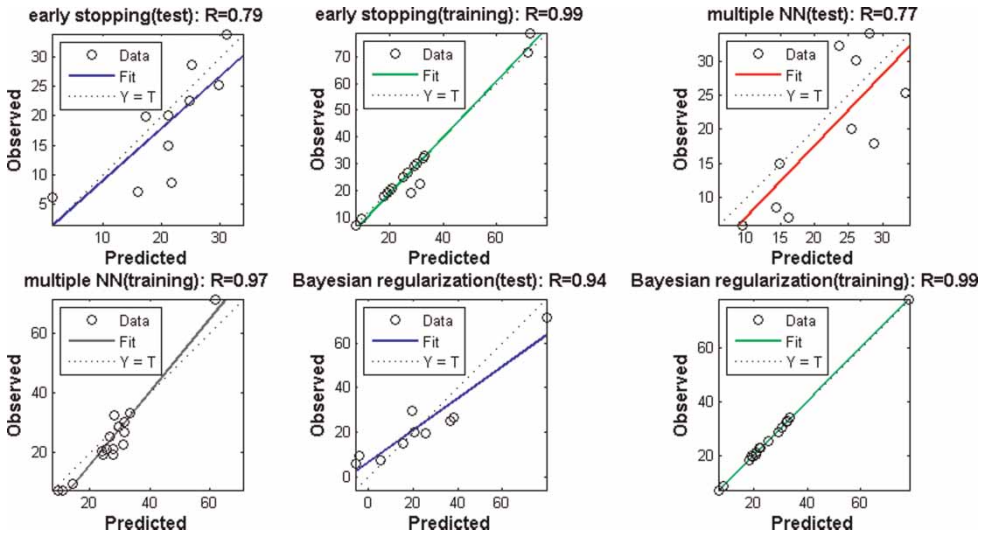

Figure 4 | Correlation coefficient between predicted and observed BOD by ANN and EANN modelling.

On the other hand, the condition for ensemble averaging was nearly the same in which the training and test correlation coefficients were in turn 0.967 and 0.768 showing a poor out-of-sample prediction and suggesting that these two latter models have memorized the training set and probably have over-fitted the training data. Concerning the Bayesian regularization algorithm, the perfect fitting of the model to both training and test data with correlation coefficients of 0.999 and 0.945 confirmed that this method outperformed the results of other models. The results of our study were consistent with that of Daliakopoulos *et al.* (2005) and Anctil *et al.* (2004) who concluded that the advantage of the Bayesian regularization algorithm is that whatever the size of the network, the function will not be over-fitted. However, these results are not in agreement with some previous studies (e.g. Zaier *et al.* 2010; Khalil *et al.* 2011) who believed that the performance of a single ANN can be improved by using an ensemble of artificial neural networks, although as mentioned earlier there are various algorithms for construction of an ensemble of neural networks other than manipulation of initial random weight, which was used in this study, such as using different network topology, different training algorithms and manipulating the training set (Shu & Ouarda 2007). Application of any of these algorithms may improve the out-of-sample generalization of ANN accordingly. One of the main reasons for poor generalization of ANN in conjunction with early stopping is the way that the available data are divided into training, validation and test set (Sahoo *et al.* 2006). In addition, Ray & Klindworth (2000) reported that the random initial weights for a given architecture appear to have some impact on back propagation neural network (BPNN) prediction accuracy. An optimal data set for training would be the one that fully represents the modeling domain (Sahoo *et al.* 2006). The random division of the data which was implemented in this study may be one of the reasons contributing to this poor out-of-sample generalization.

The results of sensitivity analysis indicated that COD, next to sulfate and fecal coliform are the most important parameters in prediction of BOD by ANN modeling (Figure 5). The relationship between COD and BOD is evident as high levels of the organic matter consume large amounts of oxygen that contribute to biological oxygen demand as well. Moreover, these results confirm that a high level of organic matter content in Arak's wastewater can be attributed to contamination with fecal wastes from sewage which exerted a high biological oxygen demand for its treatment. On the other hand, Arak is one of the key industrial cities of Iran, the main industries ranging from Arak (Shazand) Oil Refinery Company to Arak Petrochemical Company being present in the area. Wastewater

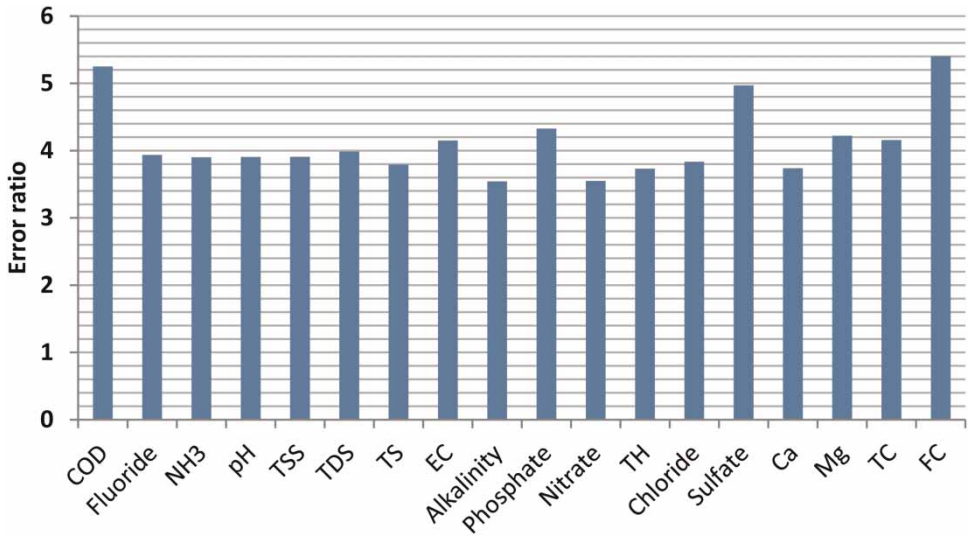

Figure 5 | Results of sensitivity analysis with 18 wastewater quality variables.

produced by many of these industrial processes often has both chemical oxygen demand and sulfate concentration (Zitomer & Shrout 2000) which affects the BOD level of wastewater. Beside these three parameters, the importance of all other parameters for BOD prediction is roughly the same.

CONCLUSION

The potential of three modeling methods including the artificial neural network technique (ANN with early stopping and ANN with Bayesian regularization), ensemble of artificial neural network (EANN) and partial least squares regression (PLSR) in the estimation of BOD in wastewater of Arak City in Iran was considered in this study. The models trained and validated on a data set containing 18 parameters sampled periodically from the wastewater of Arak City since 2009 by the Department of Environment of Markazi Province. The present study showed that the best performing ANN was the Bayesian regularization method which was capable of predicting the out-of-sample fluctuations of BOD with reasonable accuracy. Considering the low test error (based on MSE), ANN with Bayesian regularization can be considered as an alternative modeling method to laboratory analysis of BOD which has a high degree of uncertainty. Sensitivity analysis was also conducted to determine the most important variables in estimation of BOD indicating that COD and sulfate next to fecal coliform are the key parameters in this regard.

REFERENCES

Agrafiotis, D. K., Cedeno, W. & Lobanov, V. S. 2002 On the use of neural network ensembles in QSAR and QSPR. *J. Chem. Inf. Comput. Sci.* **42**, 903–911.

Anctil, F., Perrin, C. & Andreassian, V. 2004 Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models. *Environ. Modell. Software.* **19** (4), 357–368.

Basant, N., Gupta, S., Malik, A. & Singh, K. P. 2010 Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water – A case study. *Chemometr. Intell. Lab.* **104**, 172–180.

Baum, E. B. & Haussler, D. 1989 What size net gives valid generalization? *Neurocomputing* **6**, 151–160.

Beale, M. H., Hagan, M. T. & Demuth, H. B. 2013 *Neural Network Toolbox User Guide*. The Math Works, Inc., Natick, MA.

Boszke, L. & Astel, A. 2009 Application of neural-based modeling in an assessment of pollution with mercury in the middle part of the Warta River. *Environ. Monit. Assess.* **152**, 133–147.

Chen, S. H., Jakeman, A. J. & Norton, J. P. 2008 Artificial intelligence techniques: an introduction to their use for modelling environmental systems. *Math. Comput. Simulat.* **78**, 379–400.

Cheng, S., Li, L., Chen, D. & Li, J. 2012 A neural network based ensemble approach for improving the accuracy of meteorological fields used for regional air quality modeling. *J. Environ. Manage.* **112**, 404–414.

Cote, M., Grandjean, B. P. A., Lessard, P. & Thibault, J. 1995 Dynamic modelling of the activated sludge process: Improving prediction using neural networks. *Water Res.* **29**, 995–1004.

Cybenko, G. 1989 Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* **2**, 303–314.

Daliakopoulos, I. N., Coulibaly, P. & Tsanis, I. K. 2005 Groundwater level forecasting using artificial neural networks. *J. Hydrol.* **309**, 229–240.

Dogan, E., Sengorur, B. & Koklu, R. 2009 Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *J. Environ. Manage.* **90**, 1229–1235.

Dreyfus, G. 2005 *Neural Networks Methodology and Applications*. Springer, France, 1–493.

Emamgholizadeh, S., Kashi, H., Marofpoor, I. & Zalaghi, E. 2014 Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *Int. J. Environ. Sci. Technol.* **11**, 645–656.

Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H. & Ramli, M. F. 2012 Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Mar. Pollut. Bull.* **64**, 2409–2420.

Hanrahan, G. 2011 *Artificial Neural Network in Biological and Environmental Analysis*. Taylor and Francis Group, USA, pp. 119–147.

Hansen, L. & Salamon, P. 1990 Neural network ensembles. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 993–1001.

Hornik, K., Stinchcombe, M. & White, H. 1989 Multilayer feedforward networks are universal approximators. *Neural Networks* **2**, 359–366.

Hsieh, W. M. 2009 *Machine Learning Methods in the Environmental Science, Neural Networks and Kernels*. Cambridge University Press, UK, 86–157.

Khalil, B., Ouarda, T. B. M. J. & St-Hilaire, A. 2011 Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *J. Hydrol.* **405**, 277–287.

Kisi, O., Akbari, N., Sanatipour, M., Hashemi, A., Teimourzadeh, K. & Shiri, J. 2013 Modeling of dissolved oxygen in river water using artificial intelligence techniques. *J. Environ. Inform.* **22** (2), 92–101.

Kovoor, G. M. & Nandagiri, L. 2007 Developing regression models for predicting pan evaporation from climatic data – a comparison of multiple least-squares, principal components, and partial least-squares approaches. *J. Irrig. Drain. E-SCE.* **133** (5), 444–454.

Lee, D. S., Lee, M. W., Woo, S. H., Kim, Y. J. & Park, J. M. 2006 Nonlinear dynamic partial least squares modeling of a full-scale biological wastewater treatment plant. *Process. Biochem.* **41**, 2050–2057.

May, D. B. & Sivakumar, M. 2009 Prediction of urban stormwater quality using artificial neural networks. *Environ. Modell. Softw.* **24**, 296–302.

Naddafi, K., Hassanvand, M. S., Dehghanifard, E., Faezi Razi, D., Mostofi, S., Kasaee, N., Nabizadeh, R. & Heidari, M. 2009 Performance evaluation of wastewater stabilization points in Arak-Iran. *Iran. J. Environ. Health. Sci. Eng.* **6** (1), 41–46.

Nour, M. H., Smith, D. W., Gamal El-Din, M. & Prepas, E. E. 2006 The application of artificial neural networks to flow and phosphorus dynamics in small streams on the Boreal Plain, with emphasis on the role of wetlands. *Ecol. Model.* **191**, 19–32.

Oliveria-Esquerre, K. P., Da Costa, A. C., Bruns, R. E. & Mori, M. 2003 Simulation of aerated lagoon using artificial neural networks and multivariate regression techniques. *Appl. Biochem. Biotech.* **105–108**, 437–449.

Oliveria-Esquerre, K. P., Seborg, D. E., Mori, M. & Bruns, R. E. 2004 Application of steady-state and dynamic modeling for the prediction of the BOD of an aerated lagoon at a pulp and paper mill. Part II. Nonlinear approaches. *Chem. Eng. J.* **105**, 61–69.

Ozkaya, B., Demir, A. & Bilgili, S. 2007 Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. *Environ. Modell. Softw.* **22**, 815–822.

Palani, S., Liong, S. Y. & Tkalich, P. 2008 An ANN application for water quality forecasting. *Mar. Pollut. Bull.* **56**, 1586–1597.

Panda, S. S., Garg, V. & Chaubey, I. 2004 Artificial neural networks application in lake water quality estimation using satellite imagery. *J. Environ. Inform.* **4** (2), 65–74.

Piotrowski, A. P. & Napiorkowski, J. J. 2013 A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modeling. *J. Hydrol.* **476**, 97–111.

Ray, C. & Klindworth, K. K. 2000 Neural networks for agrichemical vulnerability assessment of rural private wells. *J. Hydrol. Eng.* **5** (2), 162–171.

Sahoo, G. B., Ray, C., Mehnert, E. & Keefer, D. A. 2006 Application of artificial neural networks to assess pesticide contamination in shallow groundwater. *Sci. Total Environ.* **367**, 234–251.

Shu, C. & Burn, D. H. 2004 Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resour. Res.* **40**, 1–10.

Shu, C. & Ouarda, T. B. M. J. 2007 Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. *Water Resour. Res.* **43**, W07438.

Singh, K. P., Basant, A., Malik, A. & Jain, G. 2009 Artificial neural network modeling of the river water quality – A case study. *Ecol. Model.* **220**, 888–895.

Taylor, B. J. 2006 *Methods and Procedures for the Verification and Validation of Artificial Neural Networks*. Springer, USA, 1–275.

Vinzi, V. E., Chin, W. W., Henseler, J. & Wang, H. 2010 *Handbook of Partial Least Squares, Concepts, Methods and*

Applications. Springer Handbooks of Computational Statistics, Heidelberg, 791pp.

Wieland, R., Mirschel, W., Zbell, B., Groth, K., Pechenick, A. & Fukuda, K. 2012 A new library to combine artificial neural networks and support vector machines with statistics and a database engine for application in environmental modeling. *Environ. Modell. Softw.* **25**, 412–420.

Zaier, I., Shu, C., Ouarda, T. B. M. J., Seidou, O. & Chebana, F. 2010 Estimation of ice thickness on lakes using artificial neural networks ensembles. *J. Hydrol.* **380**, 330–340.

Zitomer, D. H. & Shrout, J. D. 2000 High-sulfate, high-chemical oxygen demand wastewater treatment using aerated methanogenic fluidized beds. *Water Environ. Res.* **72** (1), 90–97.

First received 15 October 2014; accepted in revised form 18 May 2015. Available online 13 July 2015