Ain Shams Engineering Journal (2016) xxx, xxx–xxx

Ain Shams University

Ain Shams Engineering Journal

www.elsevier.com/locate/asej
www.sciencedirect.com

CIVIL ENGINEERING

Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt

Mosaad Khadr *, **Mohamed Elshemy**

Irrigation and Hydraulics Engineering Department, Faculty of Engineering, Tanta University, 31734 Tanta, Egypt

Received 10 October 2015; revised 25 June 2016; accepted 11 August 2016

KEYWORDS

Data-driven modeling;
Water quality parameters;
Manzala Lake;
Egypt

Abstract Manzala Lake, the largest of the Egyptian lakes, is affected qualitatively and quantitatively by drainage water that flows into the lake. This study investigated the capabilities of adaptive neuro-fuzzy inference system (ANFIS) to predict water quality parameters of drains associated with Manzala Lake, with emphasis on total phosphorus and total nitrogen. A combination of data sets was considered as input data for ANFIS models, including discharge, pH, total suspended solids, electrical conductivity, total dissolved solids, water temperature, dissolved oxygen and turbidity. The models were calibrated and validated against the measured data for the period from year 2001 to 2010. The performance of the models was measured using various prediction skill criteria. Results show that ANFIS models are capable of simulating the water quality parameters and provided reliable prediction of total phosphorus and total nitrogen, thus suggesting the suitability of the proposed model as a tool for onsite water quality evaluation.

1. Introduction

The quality and quantity of water resource worldwide is a subject of ongoing concern [1]. Assessment and management of long-term water quality of water resources is also a challenging problem [2–4]. The determination of the water quality refers to the classification by considering the physical, chemical and biological characteristics according to the water usage range [5]. In water quality modeling, the mathematical modeling usually involves several parameters that cannot be measured or involve considerable expense [6,7]. A deterministic model may also have inevitably errors originated from model structures or other causes. Water quality models are still therefore simplified approximations of reality, and they inevitably contain certain kinds of errors that result in uncertainty in the model results [8]. Therefore the researchers tend to rely on conceptual or empirical models in practical applications to reduce this uncertainty. A new modeling paradigm such as data-driven modeling or data mining has recently been a considerable growth in the development and application of

* Corresponding author.
E-mail addresses: mosaad.khadr@f-eng.tanta.edu.eg (M. Khadr), m.elshemy@f-eng.tanta.edu.eg (M. Elshemy).

Peer review under responsibility of Ain Shams University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.asej.2016.08.004
2090-4479 © 2016 Ain Shams University. Production and hosting by Elsevier B.V.

computational intelligence and computer tools with respect to water-related problems [9,10].

These techniques are an approach to estimate the water quality parameters based on the field data sets and to map the relationship between the water quality parameter according to the temporal and spatial variation [11,12]. Data-driven models refer to a wide range of models that simulate a system by the data experienced in the real life of that system. Data-driven modeling (DDM) is based on analyzing the data characterizing the system under study; in particular, a model can be defined on the basis of finding connections between the system state variables (input, internal and output variables) without explicit knowledge of the physical behavior [13]. DDM includes different categories generally divided into statistical and artificial-intelligent models which include neural networks, fuzzy systems and evolutionary computing as well as other areas within artificial intelligence and machine learning [14–17].

The use of ANNs and fuzzy logic has many successful applications in hydrology; in modeling rainfall-runoff processes [18–21]; replicating the behavior of hydrodynamic/hydrological models of a river basin where ANNs are used to provide optimal control of a reservoir [22]; modeling stage-discharge relationships [23]; simulation of multipurpose reservoir operation [24–26]; and deriving a rule base for reservoir operation from observed data. The development and current progress in the integration of various artificial intelligence techniques (knowledge-based system, genetic algorithm, artificial neural network, and fuzzy inference system) in water quality modeling, sediment transportation and *DO* concentration have been studied by many researchers [27–29]. Egyptian northern lakes have been regarded highly as a fishery; therefore, monitoring of water quality of the drainage water input to the northern lakes is a major task for maintaining their ecology. In this study, adaptive neuro-fuzzy inference system (ANFIS) models were developed for prediction and simulation of water quality parameters in drains systems associated with Manzala Lake, the most important among all Egyptian Lakes, with emphasis on total phosphorus (*TP*) and total nitrogen (*TN*). *TN* and *TP* are considered as the most essential parameters to assess and control the water quality and trophic status of water bodies. In order to measure these two parameters, laboratory examinations should be done using water samples, which is costly and time consuming process. To the best of our knowledge, the issue on predicting of water quality in the study area using ANFIS model so far has not been addressed. It is hoped that the proposed approach and our findings obtained in this study are useful and valuable to assist in reporting the status of water quality in the study area.

2. Method and materials

2.1. Study area and water quality data

Manzala Lake, which is located at the northern edge of the Nile Delta, is the largest of the Egyptian lakes along the Mediterranean coast (Fig. 1) [30,31]. The lake is bordered at the north by a sandy margin which separates the lake from the Mediterranean Sea except at three outlets where exchange of water occurs. These outlets are El-Gamil, El-Boughdady, and the new El-Gamil [31]. The eastern side of the lake is connected with the Suez Canal through El-Raswa Canal, a few kilometers to the south of Port Said City. To the west, the Damietta branch of the Nile River borders the lake and the southern side of the lake is bordered by cultivated land [30]. The Lake is exposed to high inputs of pollutants from industrial, domestic, and agricultural sources. The southern region of the lake is characterized by lower values of salinities and high concentration of nutrients and heavy metals as a result of receiving high volumes of low salinity drainage water through different drains. The Lake is enriched by drainage water transplanted by the drains which are connected to the Lake at the South and South Eastern Borders. Six major drains contribute by a flow rate of about 4170 million cubic meters annually [32]. The main two drains flow into Manzala Lake, which were considered in this study, are Bahr El-Baker Drain system and Bahr Hadous Drain system. Bahr El Baqar drain, which is heavily polluted and anoxic over its entire length, transports untreated and poorly treated wastewater to Lake Manzala over a distance of 170 km. Water quality data, that were used to develop the ANFIS models, are measured values at outfall measuring stations and have a record length of 10 years covering between 2001 and 2010. The data set includes discharge (*Q*), *pH*, total suspended solids (*TSS*), electrical conductivity (*EC*), total dissolved solids (*TDS*), water temperature, dissolved oxygen (*DO*) and turbidity (*TU*). The output of the model is two water quality parameters; total phosphorus (*TP*) and total nitrogen (*TN*). Both parameters were chosen to be the model objective output due to their main impacts on water quality status and control. Table 1 summarizes the statistical properties of input and output data used in the simulations.

2.2. Adaptive neuro-fuzzy inference system – ANFIS

An adaptive neuro-fuzzy inference system (ANFIS) is a fuzzy inference system formulated as a feed-forward neural network. Hence, the advantages of a fuzzy system can be combined with a learning algorithm [33,34]. ANFIS was introduced as an effective tool to represent simple and highly complex functions more powerfully than conventional statistical methods. Neuro-fuzzy modeling is a technique for describing the behavior of a system using fuzzy inference rules within a Neural Network (NN) structure. Using a given input/output data set, adaptive neuro-fuzzy inference system (ANFIS) constructs a FIS whose member ship function parameters are tuned using a back propagation algorithm [34]. So, the FIS could learn from the training data. In this study, the ANFIS models were developed in the MATLAB environment.

ANFIS was used to extract the relation of the total phosphorus (*TP*), total nitrogen (TN), discharge (*Q*), pH, total suspended solids (*TSS*), electrical conductivity (*EC*), total dissolved solids (*TDS*), water temperature, dissolved oxygen (*DO*) and turbidity (*TU*). The consequent part is total phosphorus (*TP*) or total nitrogen (*TN*). The structure of the ANFIS model consists of a Sugeno type fuzzy system with generalized bell input membership functions, which provided the best results in this study, and a linear output membership function. The Sugeno model makes use of "if then" rules to produce an output for each rule. It is similar to the Mamdani method in many respects. The first two parts of the fuzzy inference process, fuzzifying the inputs and applying the fuzzy operator, are exactly the same. The main difference between

Please cite this article in press as: Khadr M, Elshemy M, Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt, Ain Shams Eng J (2016), http://dx.doi.org/10.1016/j.asej.2016.08.004

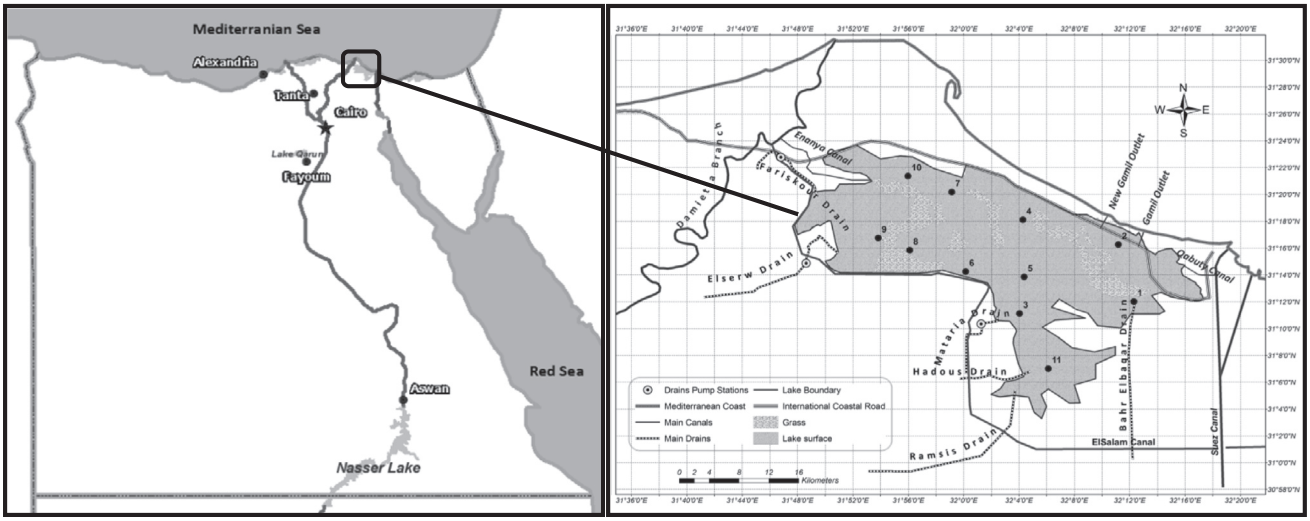

Figure 1 Layout of El-Manzala Lake and main canals and drainage system associated with it.

Table 1 Statistical measures of water quality parameters at the outlet of Bahr El Baqar and Bahr Hadous drains.

	Bahr El-Baqar drain				Bahr Hadous drain			
	Min.	Max.	Mean	Std.	Min.	Max.	Mean	Std.
Discharge (m^3/s)	42	59.9	50.38	4.93	2.34	11.65	5.44	2.77
PH	6.6	8.45	7.48	0.28	6.82	8.41	7.52	0.27
TSS (μg/L)	0	197	77.26	42.09	0.00	499.00	38.54	51.81
EC (μg/L)	1.15	6.57	4.36	0.94	0.50	2.28	1.41	0.27
TDS (μg/L)	291	4468	2779.1	645.21	342.0	1420.00	967.11	183.36
Temperature (°C)	13	31	23.05	5.31	11.00	32.00	21.90	5.35
DO (μg/L)	0.36	5.6	2.27	1.08	0.08	7.80	1.58	1.52
TU	39	200	108.73	44.7	14.00	73.00	39.59	15.41
TP (μg/L)	0.14	1.87	0.92	0.3	0.05	2.08	0.79	0.36
TN (μg/L)	0.7	55.6	15.41	13.44	0.31	80.77	10.29	13.17

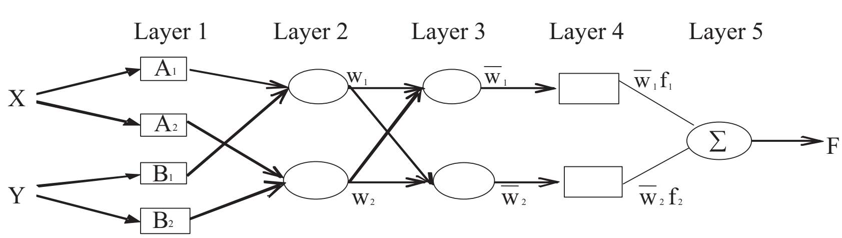

Figure 2 An ANFIS architecture for a two rule Sugeno system.

Mamdani and Sugeno is that in the Sugeno type rule outputs consist of the linear combination of the input variables plus a constant term; the final output is the weighted average of each rule's output. Adaptive neuro-fuzzy inference system mimics the operation of a Takagi–Sugeno–Kang (TSK) fuzzy system. Fig. 2 presents the typical architecture of ANFIS with a multilayer feed-forward network, which is linked with a fuzzy system for two inputs (x and y). Fuzzy inference systems are composed of five functional blocks and the ANFIS model contains the following [35]:

1. A rule base containing a number of if-then rules.
2. A database which defines the membership function.
3. A decision making interface that operates the given rules.
4. A fuzzification interface that converts the crisp inputs into "degree of match" with the linguistic values such as high or low.
5. A defuzzification interface that reconverts to a crisp output.

The rule base in the Sugeno model has the following form:

$$\text{If } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \text{ then } f_1 = p_1^* x + q_1^* y + r_1 \quad (1)$$

Please cite this article in press as: Khadr M, Elshemy M, Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt, Ain Shams Eng J (2016), http://dx.doi.org/10.1016/j.asej.2016.08.004

$$\text{If } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \text{ then } f_2 = p_2^* x + q_2^* y + r_2 \tag{2}$$

where x and y are predefined membership functions, A_i and B_i are membership values, p_i, q_i, and r_i are the consequent parameters that are updated in the forward pass in the learning algorithm, and f_i is the output within the fuzzy region specified by the fuzzy rule.

Let the membership functions of fuzzy sets A_i and B_j, be μ_{A_i} and μ_{B_i} respectively. The five layers that integrate ANFIS are as follows:

Let the output of the ith node in layer l is denoted as $O_{1,i}$, then,

Layer 1: Every node i in this layer is an adaptive node with node function;

$$Q_{1,i} = \mu_{A_i}(x) \quad \text{for } i = 1,2, \quad \text{or} \quad Q_{1,i} = \mu_{B_{i-2}}(y) \quad \text{for i} = 3,4 \tag{3}$$

where x (or y) is the input to the ith node and A_i (or B_i–2) is linguistic labels.

Layer 2: This layer consists of the nodes labeled which multiply incoming signals and send the product out. Each node output represents the firing strength of a rule;

$$O_{2,i} = w_i = \mu_{A_i}(x)\ \mu_{B_i}(y) \quad \text{for } i = 1,2 \tag{4}$$

Layer 3: In this layer, the nodes labeled N act to scale the firing strengths to provide normalized firing strengths;

$$O_{3i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \ i = 1,2 \tag{5}$$

Layer 4: The output of layer 4 is comprised of linear combination of inputs multiplied by normalized firing strengths. This layer's nodes are adaptive with node functions;

$$O_{4i} = w_i f_i = w_i(p_i x + q_i y + r_i) \tag{6}$$

where, w_i is the output of layer 3, and $\{p_i, q_i, r_i\}$ are the parameter set. Parameters of this layer are referred to as consequent parameters.

Layer 5: This layer consists of a single node and computes the final output as the summation of all incoming signals;

$$O_{5i} = \sum_{i=1} \bar{w}_i f_i = \frac{\sum_{i=1} w_i f_i}{\sum_{i=1} w_i} \tag{7}$$

Layers represented by squares are adaptive and their values are adjusted when carrying out the system training. Layers represented by circles remain invariable before, during and after the training [36]. Fig. 3 illustrates the network used in this paper and consists of eight inputs, and one output membership function (TP or TN).

2.3. Performance measures

Several measures of goodness of fit were used to evaluate the prediction performance of all the aforementioned ANFIS models. The measures that were used include Mean Absolute Deviations (MAD), the coefficient of determination (R^2), Root Mean Square Error ($RMSE$), correlation coefficient (C_r) and Nash–Sutcliffe coefficient (E). To investigate whether there is a significant difference between the mean from the observed and predicted data, the two-sample t-test for the means was employed. The MAD measures the average magnitude of the errors in a set of prediction, without considering their direction. It measures accuracy for continuous variables. Expressed MAD is calculated as follows:

$$MAD = \frac{\sum_{i=1}^{n} |WQ_{o_i} - WQ_{fi}|}{n} \tag{8}$$

where WQ_o is the observed value, WQ_f is the predicted value and n is the number of data points.

In statistics, the coefficient of determination, R^2 is used in the context of statistical models whose main purpose is the prediction of future outcomes on the basis of other related information. The absolute fraction of variance, R^2, is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\mathrm{WQo_i} - \mathrm{WQ_{fi}})^2}{\sum_{i=1}^{n} (\mathrm{WQ}o_i)^2} \tag{9}$$

with the variables already having been defined. The $RMSE$ is the square root of the variance of the residuals. It indicates the absolute fit of the model to the data–how close the observed data points are to the model's predicted values. Whereas R-squared is a relative measure of fit, $RMSE$ is an absolute measure of fit and lower values of $RMSE$ indicate better fit. $RMSE$ is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (WQ_{o_i} - WQ_{fi})^2}{n}} \tag{10}$$

The correlation coefficient is a concept from statistics, and it is a measure of how well trends in the predicted values follow trends in past actual values (historical releases). The correlation coefficient is calculated as follows:

$$C_r = \frac{\sum_{i=1}^{n} WQ_{o_i} WQ_{fi} - \frac{(\sum WQ_{o_i})(WQ_{fi})}{n}}{\sqrt{\left[\left(\sum_{i=1}^{n} WQ_{o_i}^2 - \frac{(\sum_{i=1}^{n} WQ_{o_i})^2}{n}\right)\left(\sum_{i=1}^{n} WQ_{fi}^2 - \frac{(\sum_{i=1}^{n} \mathrm{WQ}_{fi})^2}{n}\right)\right]}} \tag{11}$$

The efficiency E proposed by Nash and Sutcliffe [37] is defined as one minus the sum of the absolute squared differences between the predicted and observed values normalized by the variance of the observed values during the period under investigation. It is calculated as follows:

$$E = 1 - \left[\frac{\sum_{i=1}^{n} (WQ_{o_i} - WQ_{fi})^2}{\sum_{i=1}^{n} (WQ_{o_i} - \overline{WQ}_o)^2}\right] \tag{12}$$

where $\overline{WQ}_o$ is the average of the considered parameter.

3. Results and discussion

The adaptive neuro-fuzzy inference system (ANFIS) was used to derive and to develop models for prediction of water quality parameters in Bahr El-Baker Drain system and Bahr Hadous Drain system. To simulate and predict the behavior of water quality parameters in the two drains systems, a time series of the ten previously noted parameters in a 10-year (120-month) period was used. For ANFIS models construction, monthly data set has been randomly partitioned into two parts for the training and testing processes by considering 70% and 30% respectively, which are common divisional percentages in

Please cite this article in press as: Khadr M, Elshemy M, Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt, Ain Shams Eng J (2016), http://dx.doi.org/10.1016/j.asej.2016.08.004

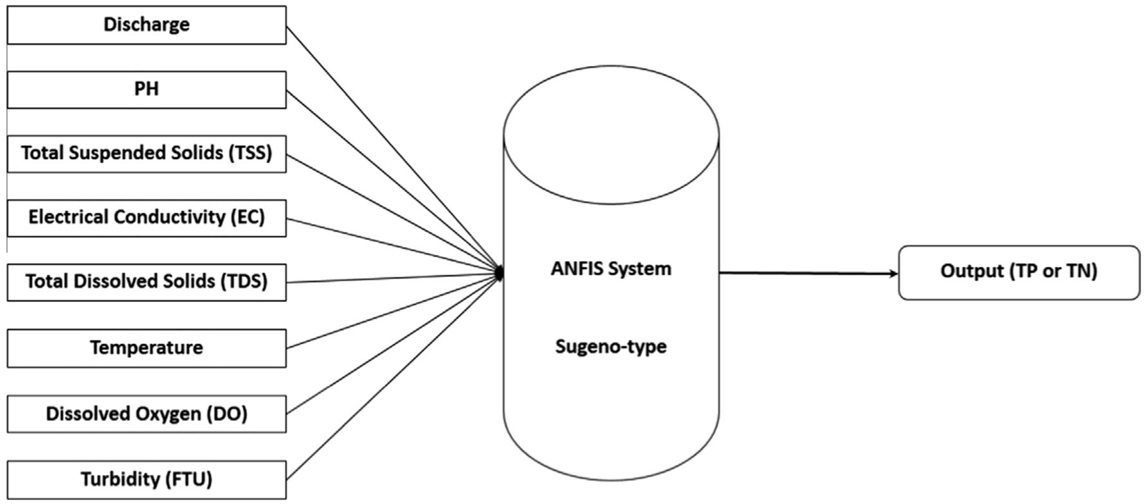

Figure 3 An ANFIS architecture for TP and TN prediction.

Table 2 Performance of particular ANFIS models according to the *RMSE*, with respect to the selection of input–output membership functions.

		Bahr El-Baqar drain					Bahr Hadous drain				
		gaussmf	*dsigmf*	*trapmf*	*gbellmf*	*trimf*	*gaussmf*	*dsigmf*	*trapmf*	*gbellmf*	*trimf*
Training	TP	0.073	0.092	0.068	0.029	0.061	0.151	0.182	0.142	0.102	0.132
	TN	1.352	1.295	1.086	0.756	0.954	0.092	0.087	0.093	0.018	0.076
Testing	TP	0.084	0.094	0.073	0.023	0.062	0.334	0.297	0.161	0.122	0.143
	TN	1.272	1.383	1.317	1.109	1.253	0.705	0.833	0.917	0.478	0.624

data-driven models. Accordingly, data set was divided into used 7 and 3-year periods, respectively, for the training and testing data sets. Fuzzy inference system structure of a Sugeno-type was then generated using subtractive clustering and the separate sets of input and output data as input arguments and this was applied to *TP* and *TN* as well. The aim of this step was to determine the number of rules and antecedent membership functions and then used linear least squares estimation to determine each rule's consequent equations to cover the feature space.

A hybrid learning algorithm was used to identify parameters of Sugeno-type fuzzy inference systems by applying a combination of the least-squares method and the back propagation gradient descent method for training FIS membership function parameters to emulate a given training data set. The network was trained to obtain the nearest output to the target. In order to find the best model, an optimization model was developed to find the ANFIS parameters that give best performance. The performance function that was used for feed-forward is the root of mean square error between the network outputs and the target output. More than 119 models were tested to select the best model which fits data space with best performance. According to the least *RMSE* from Table 2 it is very obvious that the best ANFIS model, used for training and testing, is the one using generalized bell-shaped membership functions per each input and a linear output membership function. Once the best working model was selected through ANFIS training, the TP and TN predicted values would be worked out and compared with actual measured values to validate and examine the reliability of the developed ANFIS model as shown in Fig. 4.

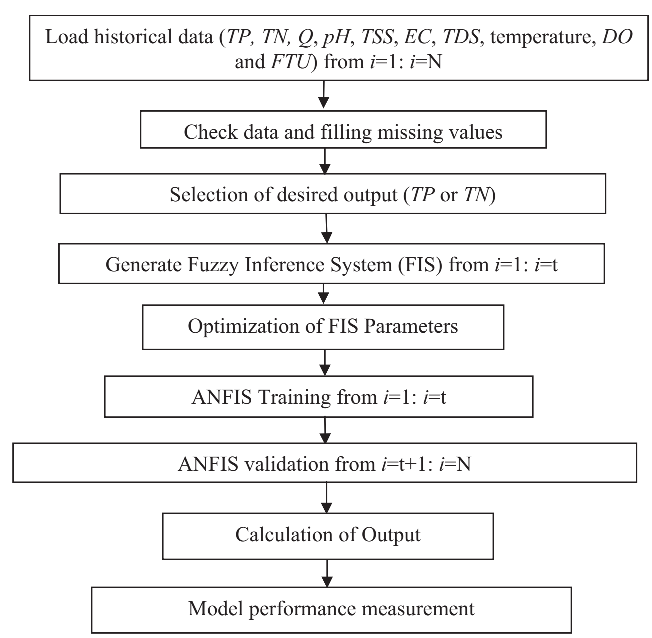

Figure 4 Flowchart of the water quality parameters simulation using ANFIS.

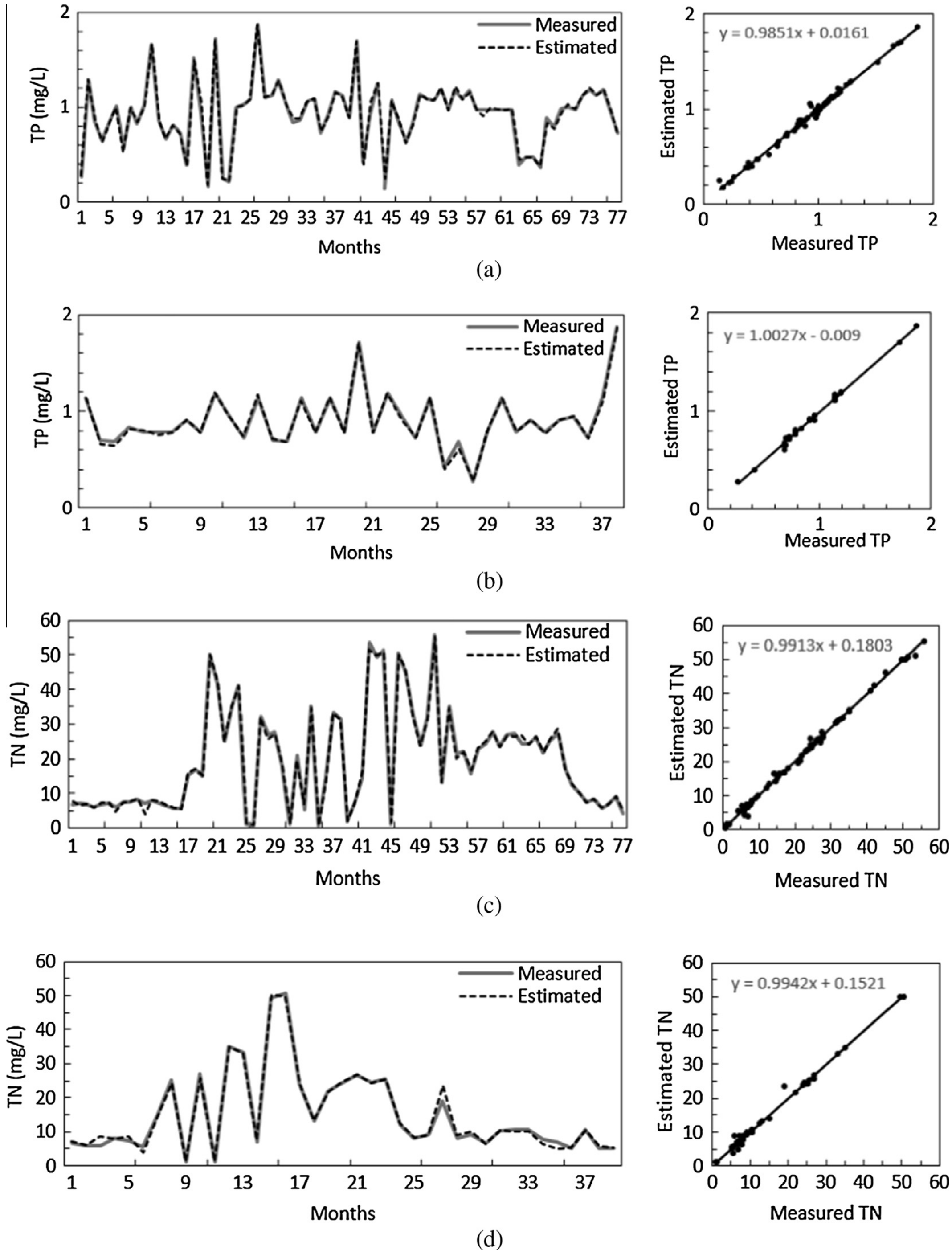

Figure 5 Observation and estimation of water quality parameters using ANFIS model in Bahr El-Baker Drain. (a) TP training period; (b) TP testing period; (c) TN training period; (d) TN testing period.

Figs. 5 and 6 present the monthly values of total phosphorus (TP) and total nitrogen (TN) estimated by ANFIS versus the corresponding measured values for the training and the test data set for Bahr El-Baker Drain system and Bahr Hadous Drain system respectively. It is shown in Figs. 5 and 6 that the two curves of observed and estimated data almost overlap each other and the trend between the measured and estimated values is similar except few records which are more deviated from actual measured values.

The training, testing and validation results for both the total phosphorus (TP) and total nitrogen (TN) prediction models are summarized in Table 3. It is clearly seen from Table 3 that the ANFIS performs satisfactory and the overall prediction results are fairly good from the $RMSE$, MAD and

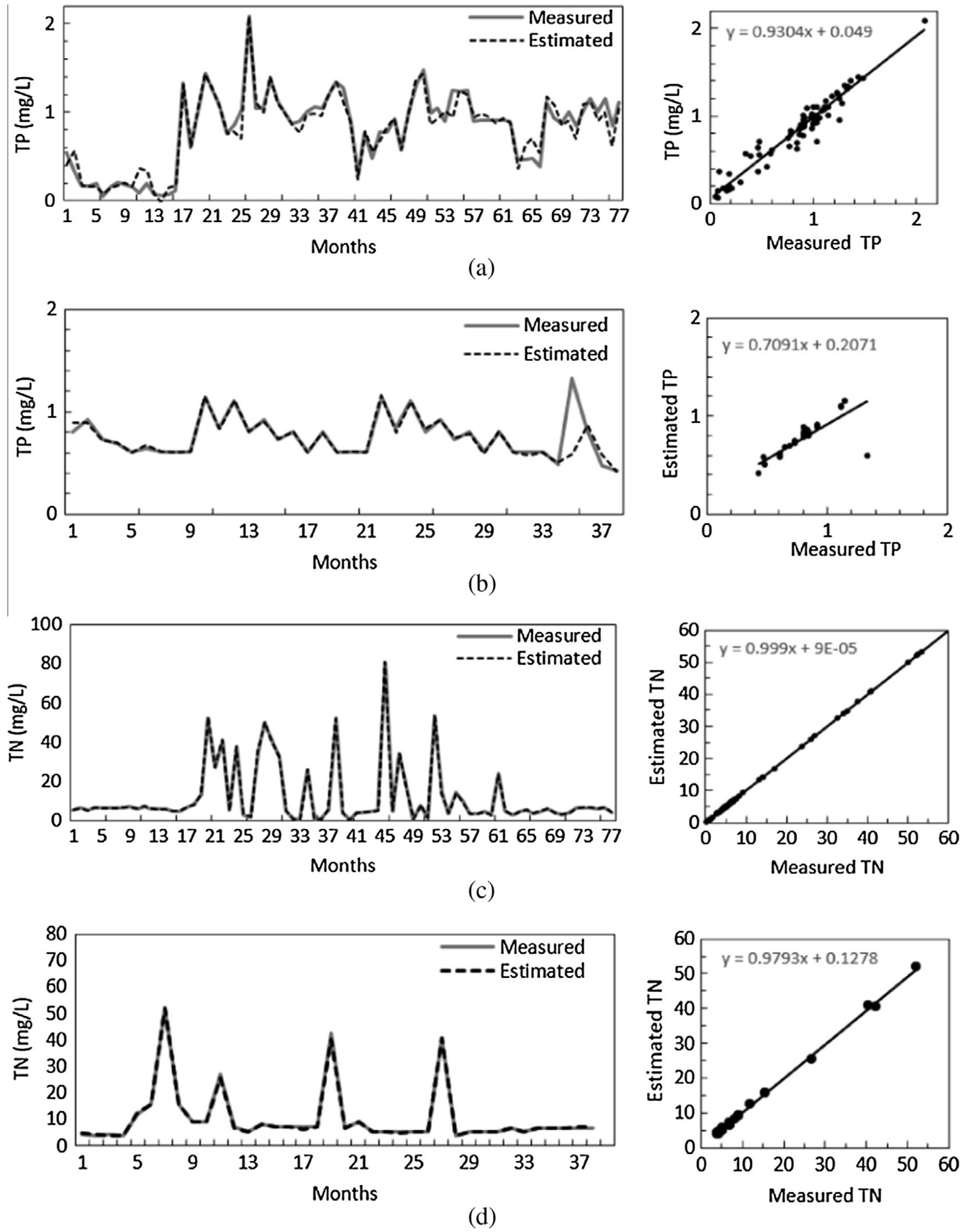

Figure 6 Observation and estimation of water quality parameters using ANFIS model in Bahr Hadous Drain. (a) TP training period; (b) TP testing period; (c) TN training period; (d) TN testing period.

Table 3 Performance measures for comparison of observed and predicted water quality parameters.

		Bahr El-Baqar drain					Bahr Hadous drain				
		MAD	R^2	*RMSE*	C_r	*E*	*MAD*	R^2	*RMSE*	C_r	*E*
Training	TP	0.016	0.98	0.029	0.976	0.992	0.071	0.98	0.102	0.971	0.942
	TN	0.457	0.99	0.756	0.945	0.983	0.009	0.97	0.018	0.981	0.932
Testing	TP	0.015	0.94	0.023	0.901	0.994	0.036	0.97	0.122	0.802	0.953
	TN	0.682	0.92	1.109	0.857	0.971	0.209	0.91	0.478	0.829	0.928

Please cite this article in press as: Khadr M, Elshemy M, Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt, Ain Shams Eng J (2016), http://dx.doi.org/10.1016/j.asej.2016.08.004

R^2 viewpoints. All values of $RMSE$ shown in Table 3 are very small compared to mean values of both TP and TN during training and testing periods, which is an indicator for the high efficiency of the data-driven model. Table 3 also indicates that the $RMSE$, for all models, is always closer to or equal to the MAD which indicates that all the errors are of the same magnitude. A significant positive correlation was obtained between the two groups of data for both TP and TN in case of Bahr El-Baker Drain with values of 0.901 and 0.857 respectively; however, in Bahr Hadous Drain C_r equals 0.802 and 0.929 respectively. The large values of R^2 are indicative of a perfect relationship between the observed and predicted values.

According to Eq. (12), the range of E lies between 1.0 (perfect fit) and $-\infty$; high value of E is indicative of a more efficient data-driven model. Values of E in the range $(-\infty, 0)$ occur when the mean observed value is a better estimation than the model prediction or simulation value, which indicates unacceptable performance. Values of E shown in Table 3 are greater than 0.9, which indicates that proposed models have perfect fit for all of the quality parameters in both training and testing data sets. The two-sample t-test failed to reject the null hypothesis at the 5% significance level ($h = 0$) that both predicted and measured data come from independent random samples from normal distributions with equal means.

4. Conclusions

In this study, the capabilities of data-driven models to predict water quality parameters were investigated. This study adopted ANFIS models to achieve easier and faster water quality parameter predictions with emphasis on total phosphorus (TP) and total nitrogen (TN) in drain systems associated with Manzala Lake, the most important among all Egyptian Lakes. Two main ANFIS models, for each drain system, were constructed for both TP and TN. The performance of the developed models was measured on a 10-year database of ten water quality parameters. Comparison between predicted and measured data, using several evaluation criteria, showed the efficiency of the applied models and confirmed the accuracy of the developed ANFIS models. Validation statistics also indicate that the correlation between predicted and actual measured values was fairly good. With reference to our findings, we propose the developed models as a simple tool for predicting water quality parameters and for onsite water quality parameters evaluation.

References

[1] Diamantopoulou MJ, Antonopoulos VZ, Papamichail DM. The use of a neural network technique for the prediction of water quality parameters of Axios River in Northern Greece. European Water 2005;11/12:55–62, 2005.
[2] Koschel R, Adams DD. An approach to under- standing a temperate oligotrophic Lowland Lake (Lake Stechlin, Germany). Archiv fur Hydrobiol, Spec Issues Adv Limnol 2003;58(2003):1–9.
[3] Parparov A, Hambright KD, Hakanson L, Ostapenia AP. Water quality quantification: basics and implementation. Hydrobiologia 2006;560(1):227–37, 2006.
[4] Parparov Arkadi, Gal Gideon, Hamilton David, Kasprzak Peter, Ostapenia Alexandr. Water quality assessment, trophic classification and water resources management. J Water Resour Prot 2010;2:907–15.
[5] Hambright KD, Parparov A, Berman T. Indices of water quality for sustainable management and conservation of an arid region lake, L. Kinneret (Sea of Gali-lee), Kinneret. Aquat Conservation, Mar Freshwater Ecosyst 2000;10(6):393–406.
[6] Rode Michael, Arhonditsis George, Balin Daniela, Kebede Tesfaye, Krysanova Valentina, Van Der Zee ATM. New challenges in integrated water quality modelling. Hydrol Process 2010;24:3447–61.
[7] Maia Celsemy Eleutério, Rodrigues Kelly Kaliane Rego da Paz. Proposal for an index to classify irrigation water quality: a case study in northeastern Brazil. R Bras Ci Solo 2012;36:823–30.
[8] Ghajarnia N, Haddad OB, Mariño MA. Performance of a novel hybrid algorithm in the design of water networks. Proc Inst Civ Eng Water Manage 2011;164(4):173–219.
[9] Orouji H, Orouji H, Bozorg Haddad O, Fallah-Mehdipour E. Modeling of water quality parameters using data-driven models. J Environ Eng © ASCE 2013:947.
[10] Daniel Edsel B, Camp Janey V, Le Boeuf Eugene J. Watershed modeling and its applications: a state-of-the-art review. Open Hydrol J 2011;5:26–50.
[11] Liu Z-J, Weller DE. A stream network model for integrated watershed modeling. Environ Modell Assess 2007;13(2):291–303.
[12] Wu CL, Chau KW. Data-driven models for monthly streamflow time series prediction. Eng Appl Artif Intell December 2010:1350–67.
[13] Abrahart R et al. Data-driven modelling: concepts, approaches and experiences, practical hydroinformatics. Water science and technology library. Berlin Heidelberg: Springer; 2008. p. 17–30.
[14] Kholghi M, Hosseini SM. Comparison of groundwater level estimation using neuro-fuzzy and ordinary kriging. Environ Model Assess 2009;14(6):729–37.
[15] Ghavidel SZZ, Montaseri M. Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud basin. Stoch Environ Res Risk Assess 2014;28 (8):2101–18.
[16] Sanikhani H, Kisi O. River flow estimation and forecasting by using two different adaptive neuro-fuzzy approaches. Water Resour Manage 2012;26(6):1715–29.
[17] Sanikhani H, Kisi O, Kiafar H, Ghavidel SZZ. Comparison of different data-driven approaches for modeling lake level fluctuations: the case of Manyas and Tuz Lakes (Turkey). Water Resour Manage 2015;29(5):1557–74.
[18] Dawson CW, W.R.. An artificial neural network approach to rainfall-runoff modelling. Hydrol Sci J 1998;43(1).
[19] Dibike YSD, Abbott MB. On the encapsulation of numerical-hydraulic models in artificial neural network. J Hydraul Res 1999;37(2).
[20] Abrahart RJ, S.L.. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecast in two contrasting catchments. Hydrol Process 2000;14.
[21] El-Shafie A, Jaafer O, Seyed A. Adaptive neuro-fuzzy inference system based model for rainfall forecasting in Klang River, Malaysia. Int J Phys Sci 2011;6(12):2875–88.
[22] Solomatine DP, TL. Neural network approximation of a hydrodynamic model in optimizing reservoir operation. In: Proc. 2nd int conference on hydroinformatics, Balkema, Rotterdam; 1996.
[23] Sudheer KP, J.S.. Radial basis function neural network for modeling rating curves. ASCE J Hydrol Eng 2003;8(3).
[24] Fontane DG, Gates TK, Moncada E. Planning reservoir operations with imprecise objectives. J Water Resour Plan Manage, ASCE 1997;123(3):154–62.
[25] Shrestha BP, Duckstein LE, Stokhin Z. Fuzzy rule-based modeling of reservoir operation. J Water Resour Plan Manage 1996;122(4):262–9.
[26] Khadr M, Schlenkhof A. Integration of data-driven modeling and stochastic modeling for multi-purpose reservoir simulation. In: 11th international conference on hydroscience & engineering

Please cite this article in press as: Khadr M, Elshemy M, Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt, Ain Shams Eng J (2016), http://dx.doi.org/10.1016/j.asej.2016.08.004

(ICHE2014), Hamburg, Germany, 28 September to 2 October 2014.

[27] Broomhead D, Lowe D. Multivariable functional interpolation and adaptive networks. Complex Syst 1998;2(6):321–55.

[28] Dogan E, Sengorur B, Koklu R. Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. J Environ Manage 2009;90:1229–35.

[29] Ranković V, Radulović J, Radojević I, Ostojić A, Ćomić L. Neural network modeling of dissolved oxygen in the Gruza reservoir, Serbia. Ecol Model 2010;221:1239–44.

[30] Ali Mohamed HH. Assessment of some water quality characteristics and determination of some heavy metals in Lake Manzala, Egypt. J Aquat Biol Fish 2008;12(2):133–54.

[31] Abdel Mola Hesham R, Abd El – Rashid Mohamed. Effect of drains on the distribution of zooplankton at the southeastern part of Lake Manzala, Egypt. J Aquat Biol Fish 2012;16:57–68, ISSN 1110–1131.

[32] Laila Shakweer. Ecological and fisheries development of lake Manzala, Egypt. Egypt J Aquat Res 2005;31(1).

[33] Zadeh LA. Outline of a new approach to analysis of complex systems and decision processes. IEEE Trans Syst Man Cybernetics, Smc 1973;3(1):28–44.

[34] Labani MM, Salahshoor K. Estimation of NMR log parameters from conventional well log data using a committee machine with intelligent systems: a case study from the Iranian part of the South Pars gas field, Persian Gulf Basin. J Petrol Sci Eng 2010;72:175–85.

[35] Venugopal C, Devi SP, Rao KS. Predicting ERP user satisfaction-an adaptive neuro fuzzy inference system (ANFIS). Approach Intell Inf Manage 2010;2:422–30.

[36] Kablan A. Adaptive neuro-fuzzy inference system for financial trading using intraday seasonality observation model. World Acad Sci, Eng Technol 2009;58:479–88.

[37] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part 1 - A discussion of principles. J Hydrol 1970;10 (3):282–90.

Mosaad Khadr is an Assistant Professor in Irrigation and Hydraulics Engineering Department, Faculty of Engineering, Tanta University, Egypt. In 1997, he did B.Sc. in Civil Eng. (Structural Engineering), and in 2002, M.Sc. in Civil Engineering, Irrigation and Hydraulic Engineering, Faculty of Engineering, Tanta University, Egypt. In 2011, he did Ph.D. in Water Resources Management, University of Wuppertal, Germany.

Mohamed Elshemy is an Assistant Professor in Irrigation and Hydraulics Engineering Department, Faculty of Engineering, Tanta University, Egypt. In 1997, he did B.Sc. in Civil Eng. (Structural Engineering), and in 2002, M.Sc. in Civil Engineering, Irrigation and Hydraulic Engineering, Faculty of Engineering, Tanta University, Egypt. In 2010, he did Ph.D. in Civil Engineering (Environmental Engineering), University of Braunschweig, Germany.

Please cite this article in press as: Khadr M, Elshemy M, Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt, Ain Shams Eng J (2016), http://dx.doi.org/10.1016/j.asej.2016.08.004