

# Hybrid Adaptive Neuro-Fuzzy Models for Water Quality Index Estimation

Zaher Mundher Yaseen<sup>1</sup> · Majeed Mattar Ramal<sup>2</sup> ·  
Lamine Diop<sup>3,4</sup> · Othman Jaafar<sup>5</sup> · Vahdettin Demir<sup>6</sup> ·  
Ozgur Kisi<sup>7</sup>

Received: 4 June 2017 / Accepted: 19 January 2018 /  
Published online: 8 February 2018  
© Springer Science+Business Media B.V., part of Springer Nature 2018

**Abstract** Soft computing models are known as an efficient tool for modelling temporal and spatial variation of surface water quality variables and particularly in rivers. These model's performance relies on how effective their simulation processes are accomplished. Fuzzy logic approach is one of the authoritative intelligent model in solving complex problems that deal with uncertainty and vagueness data. River water quality nature is involved with high stochasticity and redundancy due to the its correlation with several hydrological and environmental aspects. Yet, the fuzzy logic theory can give robust solution for modelling river water quality problem. In addition, this approach likewise can be coordinated with an expert system framework for giving reliable and trustful information for decision makers in enhancing river system sustainability and factual strategies. In this research, different hybrid intelligence models based on adaptive neuro-fuzzy inference system (ANFIS) integrated with fuzzy c-means data clustering (FCM), grid partition (GP) and subtractive clustering (SC) models are used in modelling river water quality index (WQI). Monthly measurement records belong to Selangor River located in Malaysia were selected to build the predictive

---

✉ Zaher Mundher Yaseen  
yaseen@tdt.edu.vn

<sup>1</sup> Sustainable Developments in Civil Engineering Research Group, Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

<sup>2</sup> Dams and Water Resources Department, College of Engineering, University of Anbar, Ramadi, Iraq

<sup>3</sup> UFR S2ATA « Sciences Agronomiques, de l'Aquaculture et des Technologies Alimentaires », Université Gaston Berger (UGB) BP 234-Saint Louis, Saint-Louis, Sénégal

<sup>4</sup> Department of Food, Agricultural and Biological Engineering, The Ohio State University, 590 Woody Hayes Dr, Columbus, OH 43210, USA

<sup>5</sup> Civil and Structural Engineering Department, Faculty of Engineering and Built Environment, National University of Malaysia, UKM, 43600 Bangi, Selangor Darul Ehsan, Malaysia

<sup>6</sup> Engineering Faculty, Civil Eng. Dept., Karatay University, Konya, Turkey

<sup>7</sup> Faculty of Natural Sciences and Engineering, Ilia State University, Tbilisi, Georgia

models. The modelling process was included several water quality terms counting physical, chemical and biological variables whereas WQI was the target variable. At the first stage of the research, statistical analysis for each water quality parameter was analyzed toward the WQI. Whereas in the second stage, the predictive models were established. The finding of the current research provides an authorized soft computing model to determine WQI that can be used instead of the conventional procedure that consumes time, cost, efforts and sometimes computation errors.

**Keywords** Hybrid ANFIS models · Water quality index · Tropical environment · River sustainability

## 1 Introductory

Based on a steady stream of hydrological studies, a river is a stream that flows into channels in a natural way, where the quality can be affected by diverse endeavors, committed by nature and human (Das Gupta 2008; Grabowski and Gurnell 2016). Activities that concern with the geological, hydrological and climatic aspects can all determine river water quality. Having added uninvited materials into the water can lead to water pollution, as it can adversely affect the water quality. These water contaminants are classified into the point-source pollutants and the non-point pollutants; the latter of which serve as the major percentage of water pollutants (Lai et al. 2011). The impact of pollution is usually far from the contamination source, and yet it is proven to still be detrimental to human but also to other creatures and the water itself. The life and reproductive activities of human and aquatic entities can be harmed, as the pollution is naturally health-threatening, and it does not sit well with the water supply system. Surface water quality has become very essential and critical matter all around the nations, owing to the fact of fresh water trends to be scarce in the near future (Bharti and Katyal 2011). Hence monitoring and quantifying water quality is extremely momentous for fresh water protection.

Malaysia water demands was quantified 15,285 m<sup>3</sup>/day in 2010 and the predicted increment exceed the 60% (20,338 in 2020). Due to this, fresh water control, monitoring and maintenance are vastly required to ensure the safety of water availability (Fulazzaky et al. 2010). All over the world, various developmental projects have left a negative mark on the water quality and indirectly affected human life and the environment as well (Ouyang 2005). It is a fact that there are various sources of pollution in Malaysia (Shuhaimi-Othman et al. 2007; Azamuddin Arsad et al. 2012; Dada et al. 2012; Othman et al. 2012). It is also added by the fact that Malaysia is located in the tropical zone and due to this, the pollution sources are determined by the heavy rainfall, where the sources affect the catchment areas when various pollutants are carried away by the flood water. The proliferation of contaminants worsens the water quality, reducing the amount of the oxygen content as well as increasing harmful algae. To add, there are also other serious issues namely conditions that negatively impact the flow depth and river bed conditions. The WQI is created to evaluate the extent to which the water body is suitable for various water activities or uses (Tyagi et al. 2013). WQI is determined through highly complicated procedure of calculation that involve several individual water quality variables magnitude including dissolved oxygen (DO), total suspended solid (TSS), turbidity (TU), calcium (CA), biochemical oxygen demand (BOD), chemical oxygen demand (COD), temperature (TEMP), and pH. The procedure of the computation for WQI is traditionally world-wide known as very complex and might presents inaccurate determination. There are several countries use the same manual of calculation such as Korea, India, Portugal,

Brazil, USA (Cude 2001; Sargaonkar and Deshpande 2003; Bordalo et al. 2006; Abrahão et al. 2007; Song and Kim 2009). It is imperative to identify and recognize the quality of the water to enable effective supervision of water pollutants, and the sooner this is identified and addressed, the sooner for the WQI to be able to monitor and execute any vital improvement to restore the water quality (Avvannavar and Shrihari 2008).

The assessment of the water quality in developing countries such as Malaysia is important to enable us to comprehend the issue about river pollutants. It is also important to enable effective measures to be taken to treat the water, and further helps to improve the state of health of living beings. Several variables that can ascertain water quality are normally measured in the laboratory as the water quality is assessed (Barzegar et al. 2017). Because of the uneconomical and time-consuming process to analyses these parameters separately, this has led to the concerted effort to develop and explore soft computing models to accurately compute these parameters with less efforts (Maier et al. 2004). It is essential to devise a fast and reliable WQI computational approach, as it will help launch a more advanced monitoring of the contaminant levels and further raise the awareness of the surface water quality (Hameed et al. 2016). Thus, studying new methods to successfully predict the river water quality indexing is the mainstay of this study.

Artificial intelligence (AI) models including artificial neural network, adaptive neuro-fuzzy system (ANFIS), support vector machine, hybrid computing models are developed to address the non-linearity and non-stationarity of the water quality variables by several researchers (Palani et al. 2008; Khalil et al. 2011; Gazzaz et al. 2012; Niroobakhsh 2012; Wang et al. 2012; Nourani et al. 2013; Liu and Lu 2014). Between all these models, ANFIS model was exhibited an efficient and essential method in modelling water quality over the other models (Yan et al. 2010; Sahu et al. 2011; Orouji and Haddad 2013; Emamgholizadeh et al. 2014; Ahmed and Shah 2015; Wei-Bo and Wen-Cheng 2015). Fuzzy system is featured to be capable in formulating rules that are based on linguistic terms for learning processes purposes. As a fact, fuzzy system internal parameters are not simply can be tuned or optimized and particularly for prediction problem. The determination of the internal parameters of fuzzy system is still ongoing mission for the expertise researchers to be solved optimally. The integration of the fuzzy system model with different learning optimization algorithms can produce a robust hybrid intelligent model to solve this problem and improve the prediction performance. Fuzzy c-Means Clustering (FCM), Grid Partition (GP) and subtractive clustering (SC) algorithms have shown an effective optimization procedure for ANFIS model, several scholar in the last five years investigated the feasibility of the ANFIS-FCM, ANFIS-GP and ANFIS-SC models on numerous aspects in water resources engineering such as hydrology, environment, climate and agriculture problems including pan evaporation (Sanikhani et al. 2012), river flow estimation and forecasting (Sanikhani and Kisi 2012), suspended sediment modelling (Kisi and Zounemat-Kermani 2016), soil temperature modelling (Kisi et al. 2016), total dissolve oxygen prediction (Zaman Zad Ghavidel and Montaseri 2014), evapotranspiration simulation (Kisi et al. 2015). Up to date and for the best knowledge of the authors, there is no research has been conducted previously on the utilization of the hybrid ANFIS-FCM, ANFIS-GP and ANFIS-SC models for water quality index prediction.

In this paper, three different learning algorithms namely FCM, GP and SC were integrated with ANFIS model for water quality index prediction. To achieve this, a highly stochastic river flow water quality case study located in Selangor state, Malaysia was selected as a case study. The modelling was carried out to predict the water quality index based on several water quality parameters including biological, physical and chemical. The attained modeling results are

compared with each other. The motivation of implementing these hybrid models to provide an authoritative predictive that can be applied for the tropical environment of Malaysian region.

## 2 Methods and Materials

In this research, three hybrid models namely ANFIS-FCM, ANFIS-GA and ANFIS-SC were developed and compared with each other for the ability to predict the WQI in tropical environment located in peninsular Malaysia for Selangor River. The MATLAB software environment was used for the application of each predictive model.

### 2.1 Adaptive Neuro Fuzzy Inference System (ANFIS)

ANFIS is a combination of fuzzy system and the learning capacity of neural networks (Tabari et al. 2012). There are three principal types of ANFIS, Mamdani, Sugeno, and Tsumoto; However, the Sugeon's system is the most used (Sanikhani et al. 2012). In Fuzzy logic, inputs data is converted into fuzzy values by employing membership functions. Fuzzy values are comprised between 0 and 1. Nodes which working as membership functions (MFs) and rules permitting to model the relationship between input and output, form the structure of the ANFIS model. Several types of Membership Function exist including triangular, trapezoidal, Gaussian and sigmoid. The Gaussian function (Eq. 1) is the most commonly used MF (Karimi et al. 2013).

$$V_{Ni} = \frac{1}{1 + \left[ \frac{(x-\alpha_i)}{\beta_i} \right]^{2\mu_i}} \quad (1)$$

where  $x$  represents the input at  $i$  node,  $V_{Ni}$  the membership function, and  $\alpha_i$ ,  $\beta_i$ , and  $\mu_i$  are the conditional parameters of the function.

In ANFIS, rules are defined based on their antecedents (If part) and consequents (Then part) and these rules are stocked in a fuzzy based rule system ('the IF-THEN' rules).

The two Eqs. (2 and 3) display the rules for a Sugeno ANFIS model with two inputs ( $x$  and  $y$ ) and one output  $f$  (Jang 1993).

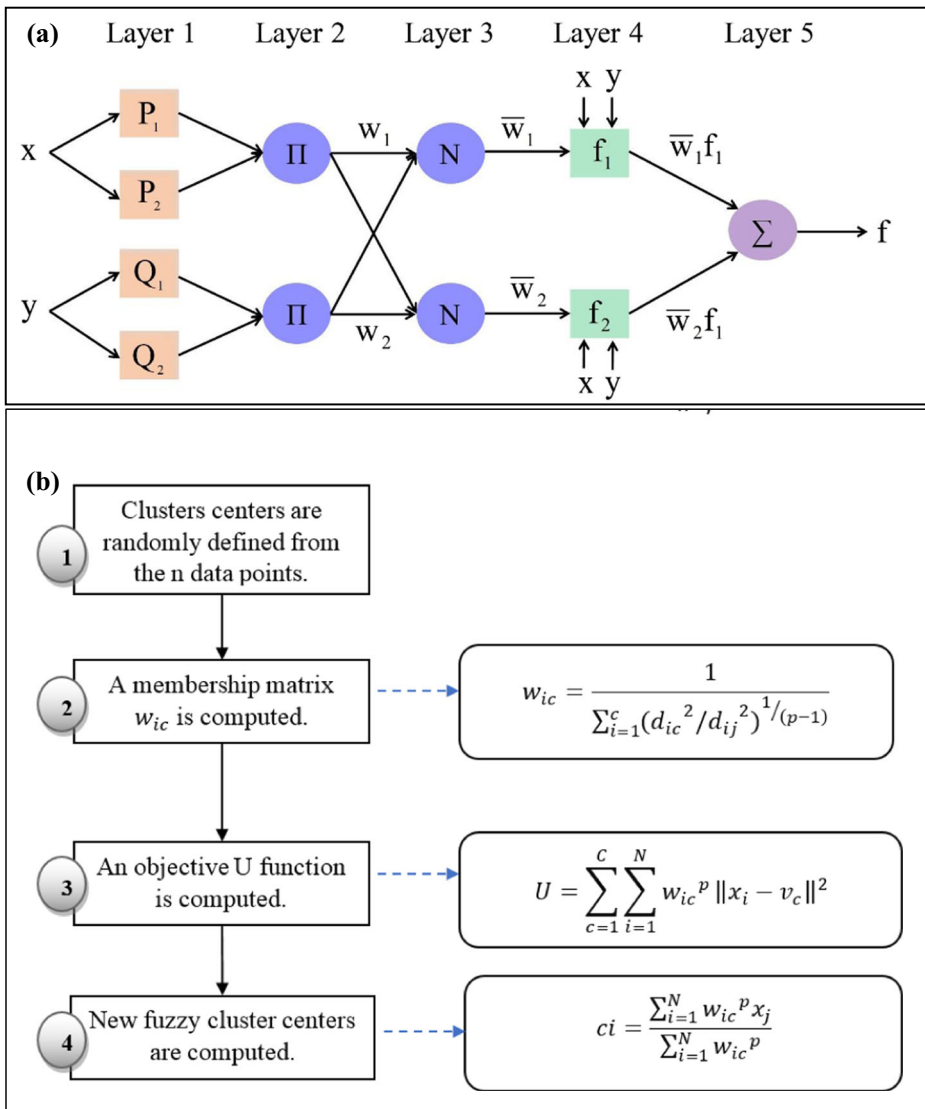
$$\text{Rule 1 : IF } x \text{ is } P_1 \text{ and } y \text{ is } Q_1, \text{ then } f_1 = p_1x + q_1y + r_1 \quad (2)$$

$$\text{Rule 2 : IF } x \text{ is } P_2 \text{ and } y \text{ is } Q_2, \text{ then } f_2 = p_2x + q_2y + r_2 \quad (3)$$

$P_i$  and  $Q_i$  are fuzzy sets,  $f_i$  represents output within the fuzzy region, and  $p_i$ ,  $q_i$ , and  $r_i$  are the design parameters determined during the training process. The Fig. 1a showed the architecture of an ANFIS with two inputs ( $x$  and  $y$ ) and one output ( $f$ ). More comprehensive details about the ANFIS approach can be found in (Jang 1993).

### 2.2 Fuzzy c-Means Clustering (FCM)

In data clustering, the data set is classified into groups and data sets with the same characteristics belong to the same clusters and non-similar data sets to different clusters. The FCM (Bezdek 1973), an improvement and modification of K-means clustering, uses a dataset of  $x_i$  data points to define  $C$  clusters by minimizing the objective function  $U$  defined in Fig. 1b. The algorithm of the FCM is presented in Fig. 1b.



**Fig. 1** a ANFIS model architecture with 2 inputs and 5 layers, b FCM algorithm procedure, c Grid partitioning with 2 inputs with  $K = 3$

FCM is an unsupervised algorithm and the objective function ( $U$  in Fig. 1b) is minimized before computing new fuzzy clusters.  $p$  (comprise entre 0 and 1) is the fuzzifier exponent,  $C$  is the number of clusters;  $N$  is the number of data points;  $c_i$  is the cluster’s centers.

### 2.3 Grid Partitioning (GP)

ANFIS-GP is a combination of ANFIS and grid partition. In Grid partition method, data is divided into grid data, based on the type and number of member function (MFs) in each dimension. The Fig. 1c presents a chart description of the ANFIS-GP with 3 partitions ( $k = 3$ ).

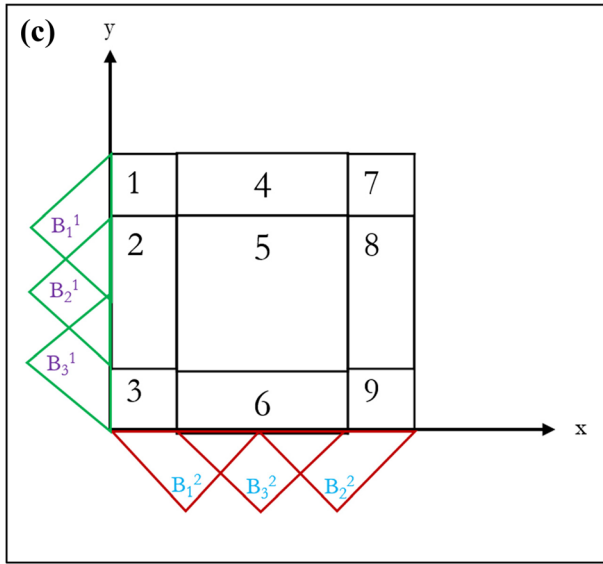


Fig. 1 (continued)

ANFIS-GP follows a process and begins with zero output and progressively learns the different fuzzy set rules and functions through the training procedure, (Cobaner 2011; Kisi and Zounemat-Kermani 2014). The least square estimate method is used for determining initial fuzzy sets and parameters depending on the partition and MF types (Kisi and Zounemat-Kermani 2014).

**2.4 Subtractive Clustering (SC)**

ANFIS subtractive clustering (ANFIS-SC) model is a combination of combining ANFIS and subtractive clustering (SC) method. In SC every single data point set is considered as a potential cluster center. Therefore, a point with several neighboring points presents a high potential value. Thus, in order to determine the first cluster center, a measure of density ( $d_i$ ) is defined (Eq. 4) and the data set point with the highest density or potential value is elected as the first cluster center (Kisi et al. 2015).

$$d_i = \sum_{k=1}^N \exp\left(-\left(\frac{2}{r_0}\right)^2 \cdot \|x_i - x_k\|^2\right) \tag{4}$$

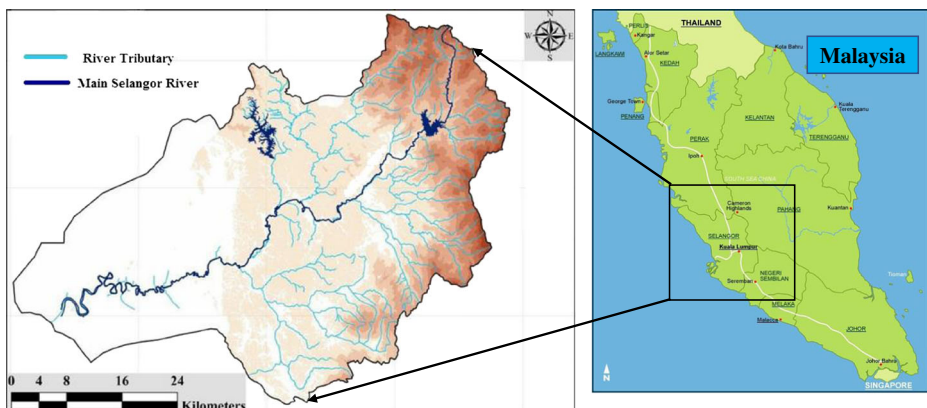
The  $d_i$  represents the density measurement,  $x_i$ : the considered cluster center,  $x_k$  is the remaining data points and  $r_0$  the influence radius. After selection of the first cluster center  $x_1$  with a density  $d_1$ , a new density ( $d_{i_{new}}$ ) is determined by excluding the impact of the first cluster center. The Eq. 5 presents the calculation of the new density measurement (Kisi et al. 2015).

$$d_{i_{new}} = d_i - d_1 \cdot \exp\left(-\frac{\|x_i - x_1\|^2}{\left(\frac{r_b}{2}\right)^2}\right) \tag{5}$$

$r_b$  is constant describing the neighborhood which will have measurable decreasing in potential. The closest points to the first cluster center will have significantly reduced potential and have less chance to be selected as next cluster center. The choice of the influential radius is very important for defining the number of clusters. Optimal  $r_a$  has to be between 0.1 and 2 (Kisi and Zounemat-Kermani 2014) and  $r_b$  equals to  $1.25r_a$  (Sanikhani and Kisi 2012). Many clusters lead to more rules, therefore, it is primordial to avoid very small radius.

## 2.5 Description of Selangor River

For this particular research, the choice was done based on the importance of the river basin (Selangor river basin). It covers approximately 25% of the whole Malaysia (see Fig. 2). Therefore, particular attention needs to be under taken for this catchment, principally in this context of water scarcity. As background knowledge of the case study, Selangor river basin is located in the state of Selangor, with the catchment area of about 2200 km<sup>2</sup>- covering approximately 25% of the whole state. At an elevation of 1700 m, the river begins at the border between Selangor and Pahang. The river flows southwesterly, covering a total distance of 110 km before discharging into the Straits of Malacca in Kuala Selangor and several other rivers serve as its main tributaries. The basin area is half-covered by natural forest and a small percentage is dominated by various agricultural activities. Both the point sources and the non-point sources constitute the river pollutants, and yet, the readily-available data would not be enough to estimate quantitatively the pollutant loads accurately. Due to the fact that the data of the water quality may have to be studied separately through the experiences and knowledge of experts, to know the impact of the element content in the water on both the environment and human, the results of the water quality analysis have become ambiguous. The assessment of the river water quality can be done in three ways- firstly the water quality evaluation, which looks into the physicochemical and biological qualities of the water, secondly the physical quality evaluation system that delves into the level of man-made change on the main channel,



**Fig. 2** Selangor river basin located in western part of peninsular Malaysia



channel margins and river banks and thirdly, the biological quality evaluation system which gauges the state of the biosciences of the aquatic environment.

### 3 Results and Discussion

As a preceding analysis for the actual data set, Table 1 gives the monthly statistical parameters of the Water Quality Index at the Selangor station for the time-period 2000 to 2011 including the monthly mean value ( $X_{\text{mean}}$ ), standard deviation ( $S_x$ ), coefficient of skewness ( $C_{\text{sx}}$ ), minimum ( $X_{\text{min}}$ ), and maximum value ( $X_{\text{max}}$ ) of the WQI, respectively. The recent data are not taken into account because those data are not available or not accessible. However, we used twelve years of monthly data that concern both dry and wet seasons. In water quality studies, having twelve years of data is very remarkable and it is very rare to continually monitor quality data in the context of the study area. This fact corroborates the importance of using data driven model to estimate the water quality index. Based on Table 1, results exhibit high variation of WQI with a standard deviation varying from 83.2 to 86.1 and a coefficient of skewness greater than 6.90. The period 2009–2011 shows the smallest amplitude compared to the other three periodicity scales. There is not an obvious temporal pattern of the WQI which confirm the complexity and non-linearity pattern of WQI.

The correlation matrix between WQI and each of the input variables used in the modelling displayed in Table 2. In this study the main variables that can affect the water quality index had been used and the stepwise approach was used to determine the main variables participating to the determination of water quality index. Indeed, this table permits to clearly identify two groups of inputs accordantly to the direction of the relationship between inputs and WQI: (i) inputs with a positive correlation and (ii) inputs with negative correlation. Here, the value of the correlation is very essential for the developed machine learning predictive models. The sign (or direction) of the correlation only shows the proportionality of independent and dependent variables. Only two water quality variables (DO and pH) out of the 8 inputs are positively correlated to the WQI with a value of 0.83 and 0.467 for DO and pH, respectively. Among the inputs with a negative correlation coefficient with WQI, total solids and turbidity give the highest correlation with a value of  $-0.769$  and  $-0.651$ , respectively. The temperature of water is lesser correlated to the WQI. This phenomenon best can be elucidated owing to the river environment which is tropical environment with heavy monsoon rainfall events. The percentages of DO and pH can be highly affected by the total surface water runoff in addition to its total suspended solid.

Based on the different inputs presented in Table 2, eight input combinations were defined: i to viii (Table 3). The stepwise approach was applied to determine each of the combination. For the first combination only the DO, presenting the highest positive correlation with the WQI was considered and for each of the remaining combination (from ii to viii combination), a

**Table 1** The monthly statistical parameters of WQI data set of Selangor station

Periods	$S_x$	$C_{\text{sx}}$	$x_{\text{min}}$	$x_{\text{max}}$	$x_{\text{mean}}$
2000–2002	84.262	8.353	−0.897	60.796	0.048
2003–2005	86.067	7.416	−0.735	69.541	−0.169
2006–2008	84.419	6.924	−0.703	66.451	−0.138
2009–2011	83.286	8.811	−0.174	64.766	−0.061



**Table 2** Correlation matrix between WQI and each input attribute of water quality variables

	DO mg/l	TS mg/l	TUR Neph.U	CA mg/l	BOD mg/l	COD mg/l	TEMP °C	pH Unit	WQI
DO mg/l	1.000								
TS mg/l	-0.549	1.000							
TUR N.U	-0.456	0.856	1.000						
CA mg/l	-0.503	0.405	0.210	1.000					
BOD mg/l	-0.341	0.250	0.225	0.321	1.000				
COD mg/l	-0.228	0.432	0.401	0.154	0.329	1.000			
TEMP °C	-0.508	0.318	0.203	0.321	0.261	0.151	1.000		
pH	0.375	-0.414	-0.236	-0.484	-0.106	-0.098	-0.256	1.000	
<b>WQI</b>	<b>0.830</b>	<b>-0.769</b>	<b>-0.651</b>	<b>-0.542</b>	<b>-0.537</b>	<b>-0.511</b>	<b>-0.471</b>	<b>0.464</b>	<b>1.000</b>

different water quality parameter considered as another input's. The water quality index may be affected also by the management practices and other river basin characteristics; therefore a model may integrate this information in their processes and we expect that with this approach we may get very good results. However, in the context of data scarcity and the complexity of developing these kind of models. It will be more practical and optimal to develop model that use less data and produce accurate results like Artificial Intelligence models.

For each input combination, the four sub time periods (M1, M2, M3, and M4) and their optimal models were tabulated in Table 4. The choice of these optimal models was based on the number of cluster and number of iteration for ANFIS-FCM, on the membership function type, number of membership functions, and number of iterations for the ANFIS-GP; on the radii value and number of iterations for the ANFIS-SC.

The hybrid integrative ANFIS predictive models were evaluated and assessed using various statistical indicators to outline the prediction accuracy. In fact, soft computing methodologies usually react and behave differently from one case to another, this is owing the degree of the complexity of the problem. Hence, different evaluation metrics including root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination R<sup>2</sup> were inspected.

RMSE satisfies the triangle inequality that is required for a distance function metric for model evaluation and it is preferred in data assimilation field where the sum of squared errors is often defined as the cost function to be minimized by adjusting model parameters. MAE indicates the average magnitude of the model error and the fact of taking the absolute value avoids error compensation. R<sup>2</sup> describes the degree of linear association between observed and predicted values directly such that it lies between 0 and 1 (with 1 being a perfect model).

**Table 3** The input combinations investigated in the current research to predict the WQI

Input combination	DO mg/l	TS mg/l	TUR Neph.U	CA mg/l	BOD mg/l	COD mg/l	TEMP °C	pH
(i)	√							
(ii)	√	√						
(iii)	√	√	√					
(iv)	√	√	√	√				
(v)	√	√	√	√	√			
(vi)	√	√	√	√	√	√		
(vii)	√	√	√	√	√	√	√	
(viii)	√	√	√	√	√	√	√	√

**Table 4** Optimal models for each combination

Statistics	Cross validation		Input combinations						
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	
ANFIS-FCM	M1	(3,80)*	(7,70)	(3,50)	(3,30)	(3,10)	(4,10)	(3,30)	(3,30)
	M2	(3,100)	(9,90)	(9,4,100)	(8,90)	(8,100)	(8,100)	(8,40)	(8,90)
	M3	(6,50)	(7,90)	(7,2,90)	(8,100)	(8,70)	(8,80)	(8,80)	(8,100)
	M4	(8,100)	(7,70)	(9,3,80)	(8,100)	(8,60)	(8,5,100)	(8,5,100)	(7,80)
ANFIS-GP	M1	(4,2,30)**	(4,3,2,100)	(6,2,2,2,100)	(4,2,2,3,2,30)	(1,2 × 5,100)	(8,2 × 6,100)	(2,2 × 7,100)	(4,2 × 8,100)
	M2	(5,3,100)	(4,3,3,90)	(4,3,3,100)	(1,3,3,2,100)	(4,2 × 5,100)	(4,2 × 6,100)	(4,2 × 7,100)	(2,2 × 8,100)
	M3	(8,2,70)	(8,3,3,100)	(2,3,3,3,100)	(4,3,3,2,3,100)	(4,2 × 5,100)	(2,2 × 6,100)	(8,2 × 7,100)	(8,2 × 8,100)
	M4	(5,3,90)	(4,3,3,100)	(4,3,3,3,100)	(4,3,3,3,100)	(8,2 × 5,100)	(8,2 × 6,100)	(8,2 × 7,100)	(2,2 × 8,100)
ANFIS-SC	M1	(0,2,10)***	(0,4,10)	(0,4,10)	(0,6,20)	(0,5,20)	(0,5,10)	(0,7,20)	(0,7,20)
	M2	(0,4,100)	(0,2,100)	(0,2,90)	(0,2,100)	(0,2,100)	(0,2,100)	(0,2,100)	(0,3,100)
	M3	(0,3,30)	(0,2,10)	(0,2,100)	(0,2,100)	(0,2,100)	(0,2,100)	(0,2,100)	(0,2,100)
	M4	(0,2,100)	(0,2,90)	(0,2,100)	(0,2,100)	(0,2,100)	(0,2,50)	(0,2,10)	(0,2,20)

\* (3,80) indicates 3 clusters and 80 iterations

\*\* (4,2,30) indicate 4 membership function type (1 = dsgimf, 2 = gaussmf, 3 = gauss2mf, 4 = gbellmf, 5 = pimf, 6 = psigmf, 7 = trapmf, 8 = trimf) given in Tables 4, 2 number of membership functions and 30 iterations. It should be noted that in the ANFIS-GP models comprising 5,6,7 and 8 inputs, 2 membership functions were used. 2 × 5 indicates 2 mf for 5 inputs and vice versa

\*\*\* (0,2,10) indicate 0.2 radii value and 10 iterations

However, this is a linear metric and is based on the covariance in the observed and predicted data, so only the dispersion of the predicted and observed data is quantified (one of the main disadvantage).

The combination of these three metrics leads to a valuable conclusion of the accuracy of a model. The prediction performance of the ANFIS-FCM, ANFIS-GP, and ANFIS-SC models were tabulated in Tables 5, 6, and 7, respectively. It is shown that in general that ANFIS-SC gave the best performance compared to the other models. For all three models, the combination with the full inputs give the best results for M2, M3, and M4 while for M1 there is not a consistency of the results. The RMSE varied from 1.04 to 3.68, between 1.82 and 4.26, and from 0.56 to 0.95 for the ANFIS-FCM, ANFIS-GP, and ANFIS- SC, respectively; the MAE ranged from 0.70 to 2.86, from 1.22 to 2.92, and between 0.40 and 2.86 for the ANFIS-FCM, ANFIS-GP, and ANFIS- SC, respectively. The coefficient of determination is varying between 0.77 and 0.97 for the ANFIS-FCM, 0.77 to 0.88 for ANFIS-GP, and 0.78 and 0.98 for the ANFIS-SC. When considering the different input combinations and the optimal models together, the mean RMSE (MAE) is 1.86 (1.39), 2.84 (1.91), and 1.18 (1.19) for the ANFIS-FCM, ANFIS-GP, and ANFIS-SC, respectively; the coefficient of determination is 0.93, 0.87, and 0.93, respectively. This confirms again the superiority of the ANFIS-SC compared to the other two models. The ANFIS-FCM which is also cluster based model has also better accuracy than the ANFIS-GP.

For further analysis, each of the best combination of optimal models and input variable combinations are considered for each of the ANFIS-FCM, ANFIS-GP, and ANFIS-SC model. Figure 3a–d give the graphical comparison of optimal models in prediction of WQI. For optimal model M1, the input combination vi, iv, and v gave the best results for ANFIS-FCM, ANFIS-GP, and ANFIS-SC models, respectively; while for M2 model, the full input variables (combination viii) exhibit the best performance for all models. For M3 and M4, the full variable combination displays the best results for ANFIS-FCM and ANFIS-GP whereas, for M3 and M4, the vi and v combination demonstrate the accurate results for ANFIS-SC. Across all input combinations, the model M1 presents the poorest results for all the three ANFIS. The scatterplot and the hydrographs of calculated versus predicted WQI presented in Figs. 2 and 3

**Table 5** Test statistics of ANFIS-FCM models in prediction of WQI

Statistics	Cross validation	Input combinations								Mean
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	
RMSE	M1	4.7047	2.8679	3.1531	2.6476	2.3275	2.2828	2.4801	2.8689	<b>2.9166</b>
	M2	3.0508	2.0065	1.8296	1.8063	1.6683	0.8970	0.5710	0.4329	<b>1.5328</b>
	M3	3.7212	2.1794	1.4965	1.4965	1.3145	0.7875	0.6236	0.4345	<b>1.5067</b>
	M4	3.2347	2.1246	1.7289	1.7862	1.1551	0.8238	0.5155	0.4309	<b>1.4750</b>
	<b>Mean</b>	<b>3.6778</b>	<b>2.2946</b>	<b>2.0520</b>	<b>1.9341</b>	<b>1.6163</b>	<b>1.1978</b>	<b>1.0476</b>	<b>1.0418</b>	<b>1.8578</b>
MAE	M1	3.3821	2.1447	2.2865	2.0368	1.6894	1.7537	1.7747	1.9775	<b>2.1307</b>
	M2	2.5269	1.6214	1.5134	1.2610	1.2284	0.6809	0.3959	0.2696	<b>1.1872</b>
	M3	3.0816	1.6289	1.0715	1.0715	0.8896	0.5768	0.4160	0.2645	<b>1.1250</b>
	M4	2.4689	1.6550	1.3112	1.4714	0.8438	0.5787	0.3563	0.2932	<b>1.1223</b>
	<b>Mean</b>	<b>2.8648</b>	<b>1.7625</b>	<b>1.5456</b>	<b>1.4602</b>	<b>1.1628</b>	<b>0.8975</b>	<b>0.7357</b>	<b>0.7012</b>	<b>1.3913</b>
R <sup>2</sup>	M1	0.6801	0.8955	0.8691	0.9131	0.9206	0.9271	0.9187	0.8864	<b>0.8763</b>
	M2	0.8347	0.9312	0.9383	0.9396	0.9490	0.9861	0.9941	0.9967	<b>0.9462</b>
	M3	0.7083	0.9046	0.9529	0.9529	0.9632	0.9870	0.9922	0.9960	<b>0.9321</b>
	M4	0.8684	0.9409	0.9609	0.9586	0.9838	0.9910	0.9966	0.9977	<b>0.9623</b>
	<b>Mean</b>	<b>0.7729</b>	<b>0.9181</b>	<b>0.9303</b>	<b>0.9411</b>	<b>0.9542</b>	<b>0.9728</b>	<b>0.9754</b>	<b>0.9692</b>	<b>0.9292</b>

**Table 6** Test statistics of ANFIS-GP models in prediction of WQI

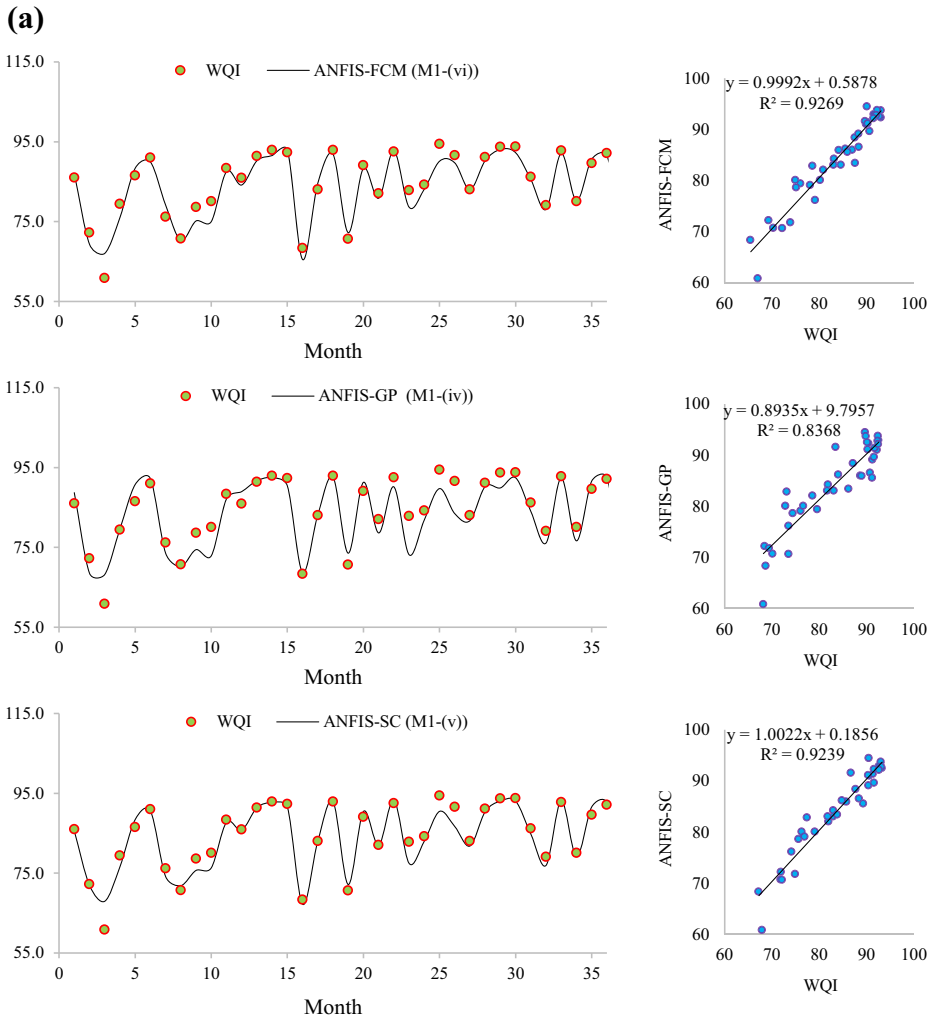
Statistics	Cross validation	Input combinations								Mean
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	
RMSE	M1	4.6134	2.9632	3.8633	3.5672	3.8372	4.3739	15.3906	11.3374	<b>6.2433</b>
	M2	3.1402	2.5283	2.3366	1.8344	1.6536	1.2603	0.7672	0.4125	<b>1.7416</b>
	M3	3.5711	2.4614	1.7629	1.4397	1.5652	1.0766	0.6033	0.1792	<b>1.5824</b>
	M4	3.5163	2.7618	1.8987	3.5672	1.4098	0.5613	0.2859	0.2089	<b>1.7762</b>
	<b>Mean</b>	<b>3.7103</b>	<b>2.6787</b>	<b>2.4654</b>	<b>2.6021</b>	<b>2.1165</b>	<b>1.8180</b>	<b>4.2618</b>	<b>3.0345</b>	<b>2.8359</b>
MAE	M1	3.4668	2.1301	2.9882	2.7394	2.7746	2.6786	6.9130	6.4745	<b>3.7706</b>
	M2	2.5737	2.0102	1.9396	1.4108	1.2014	0.9833	0.5292	0.2804	<b>1.3661</b>
	M3	2.9544	1.8642	1.2200	0.9933	1.1776	0.7927	0.3465	0.0980	<b>1.1808</b>
	M4	2.7011	2.1168	1.4221	2.7394	1.0606	0.4154	0.1932	0.1067	<b>1.3444</b>
	<b>Mean</b>	<b>2.9240</b>	<b>2.0303</b>	<b>1.8925</b>	<b>1.9707</b>	<b>1.5535</b>	<b>1.2175</b>	<b>1.9955</b>	<b>1.7399</b>	<b>1.9155</b>
R <sup>2</sup>	M1	0.6975	0.8836	0.7812	0.8368	0.7839	0.7279	0.1375	0.5603	<b>0.6761</b>
	M2	0.8252	0.8899	0.8995	0.9390	0.9532	0.9714	0.9904	0.9970	<b>0.9332</b>
	M3	0.7297	0.8815	0.9365	0.9573	0.9488	0.9765	0.9928	0.9993	<b>0.9278</b>
	M4	0.8456	0.8995	0.9533	0.8368	0.9759	0.9960	0.9989	0.9994	<b>0.9382</b>
	<b>Mean</b>	<b>0.7745</b>	<b>0.8886</b>	<b>0.8926</b>	<b>0.8925</b>	<b>0.9155</b>	<b>0.9179</b>	<b>0.7799</b>	<b>0.8890</b>	<b>0.8688</b>

showed that the best model to predict the WQI is ANFIS-SC (M4-vii) with a slope of 1, an intercept of - 0.043, and a perfect coefficient determination of 1.

Results from this study indicate generally that the performance of ANFIS-SC in modelling the water quality index is better than the ANFIS-FCM and ANFIS-GP. These results are in accord with other studies like Sanikhani and Kisi 2012 that proved the ANFIS-SC feasibility performed better than ANFIS-GP in streamflow forecasting at the Firat-Dicle Basin of Turkey in the Besiri Station and the Baykan Stations. Also, (Kisi and Zounemat-Kermani 2016) demonstrated the superiority of ANFIS-FCM over the ANFIS-GP to simulate the daily suspended sediment at Muddy Creek in the USA; however, these findings contrast with the output from the research of Kisi and Zounemat-Kermani 2014 whom concluded that ANFIS-GP gave better performance than ANFIS-SC in modelling daily reference evapotranspiration.

**Table 7** Test statistics of ANFIS-SC models in prediction of WQI

Statistics	Cross validation	Input combinations								Mean
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	
RMSE	M1	4.5034	3.1739	3.3453	3.3358	2.2995	2.3880	2.3858	2.2449	<b>2.9596</b>
	M2	3.0643	2.3025	2.2231	1.1125	0.0006	0.0005	0.0008	0.0004	<b>1.0881</b>
	M3	3.6298	2.2736	2.0636	1.3058	0.0005	0.0012	0.0008	0.0010	<b>1.1595</b>
	M4	3.2623	2.2671	2.2551	1.1483	0.0008	0.0007	0.0004	0.0006	<b>1.1169</b>
	<b>Mean</b>	<b>3.6149</b>	<b>2.5043</b>	<b>2.4718</b>	<b>1.7256</b>	<b>0.5754</b>	<b>0.5976</b>	<b>0.5969</b>	<b>0.5617</b>	<b>1.5810</b>
MAE	M1	3.3980	2.3177	2.3963	2.5580	1.6030	1.7830	1.7295	1.7022	<b>2.1859</b>
	M2	2.5278	1.7494	1.8283	0.7248	0.0003	0.0002	0.0005	0.0002	<b>0.8539</b>
	M3	2.9473	1.7824	1.5461	0.9535	0.0003	0.0005	0.0005	0.0005	<b>0.9039</b>
	M4	2.5636	1.6543	1.6460	0.8399	0.0004	0.0003	0.0003	0.0004	<b>0.8381</b>
	<b>Mean</b>	<b>2.8592</b>	<b>1.8759</b>	<b>1.8542</b>	<b>1.2691</b>	<b>0.4010</b>	<b>0.4460</b>	<b>0.4327</b>	<b>0.4258</b>	<b>1.1955</b>
R <sup>2</sup>	M1	0.7088	0.8982	0.8887	0.8708	0.9241	0.9280	0.9270	0.9434	<b>0.8861</b>
	M2	0.8332	0.9018	0.9191	0.9779	1.0000	1.0000	1.0000	1.0000	<b>0.9540</b>
	M3	0.7213	0.8952	0.9140	0.9672	1.0000	1.0000	1.0000	1.0000	<b>0.9372</b>
	M4	0.8645	0.9345	0.9332	0.9832	1.0000	1.0000	1.0000	1.0000	<b>0.9644</b>
	<b>Mean</b>	<b>0.7819</b>	<b>0.9074</b>	<b>0.9138</b>	<b>0.9498</b>	<b>0.9810</b>	<b>0.9820</b>	<b>0.9817</b>	<b>0.9859</b>	<b>0.9354</b>



**Fig. 3** The graphical comparison of optimal model in prediction the WQI, **a** M1, **b** M2, **c** M3, and **d** M4

From these, it is clearly shown that the modelling is site specific and different models may yield different performance depending on the domain of research as well as the geographical area. It is clear that the results of modelling depend on several factors (data, geographical area, quality of data, type of model, etc.); therefore, one model can perform better in one context and give poor results in another situation. For this reason, models have to be test in the given situation before its use. Other studies like Kisi and Ay 2013, they pointed out the accuracy similarities of the ANFIS-SC, ANFIS-GP, and ANFIS-FCM to simulate chemical oxygen demand. In other cases, depending on the time scale, models perform differently. For example, Sanikhani et al. 2015 found that the optimal ANFIS-GP models performed better than the optimal ANFIS-SC in forecasting 1 and 3-month ahead of the Manyas and Tuz lake levels whereas the ANFIS-SC model showed better accuracy in 2-month ahead forecasting. Patki et al. (2013) applied fuzzy logic based models for modelling water quality index using physico-chemical properties of the water, such as pH, dissolved oxygen, total alkalinity, total

(b)

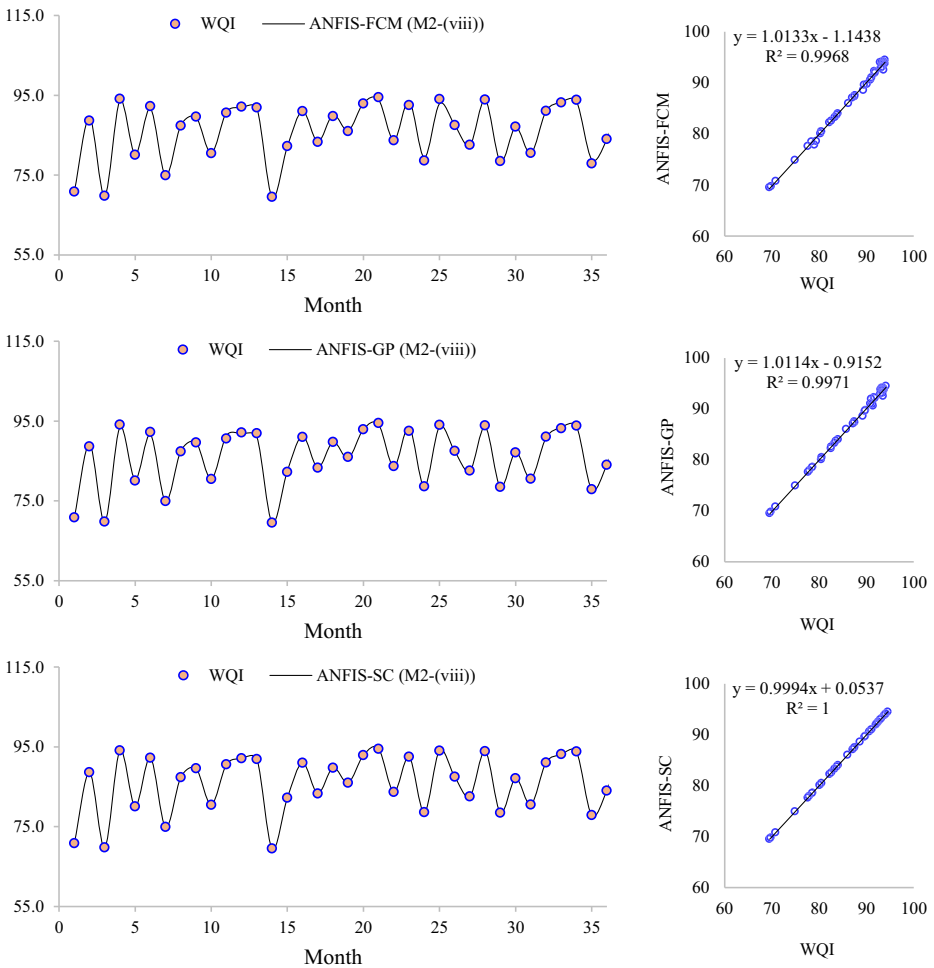


Fig. 3 (continued)

hardness, total solids and most probable number as inputs to the models. The best fuzzy model provided RMSE = 4.49, MAE = 3.53 and  $R^2 = 0.98$  in modelling WQI in the test period. They used ANFIS-GP model in their study and reported that the applied models captured the trend fully and there is no scatter in training period. During testing, however, the ANFIS-GP performed very poorly. According to the authors, the poor performance of ANFIS might be due to creating more rules, classifying limits for subsets and fixing overlapping pattern on its own by the ANFIS editor. Based on the current performance metrics (Tables 5, 6, and 7), it is clear that the applied ANFIS models provided accurate results in modelling WQI.

The results inform also that inputs variable play a key role in the accuracy of the modelling. The model with more input variables give the best results in general; however, in one hand, outputs from this research point out that the pH variable does not ameliorate the performance and the accuracy of the best model (ANFIS-FC (M4-vii)); in the other hand, the full input give

(c)

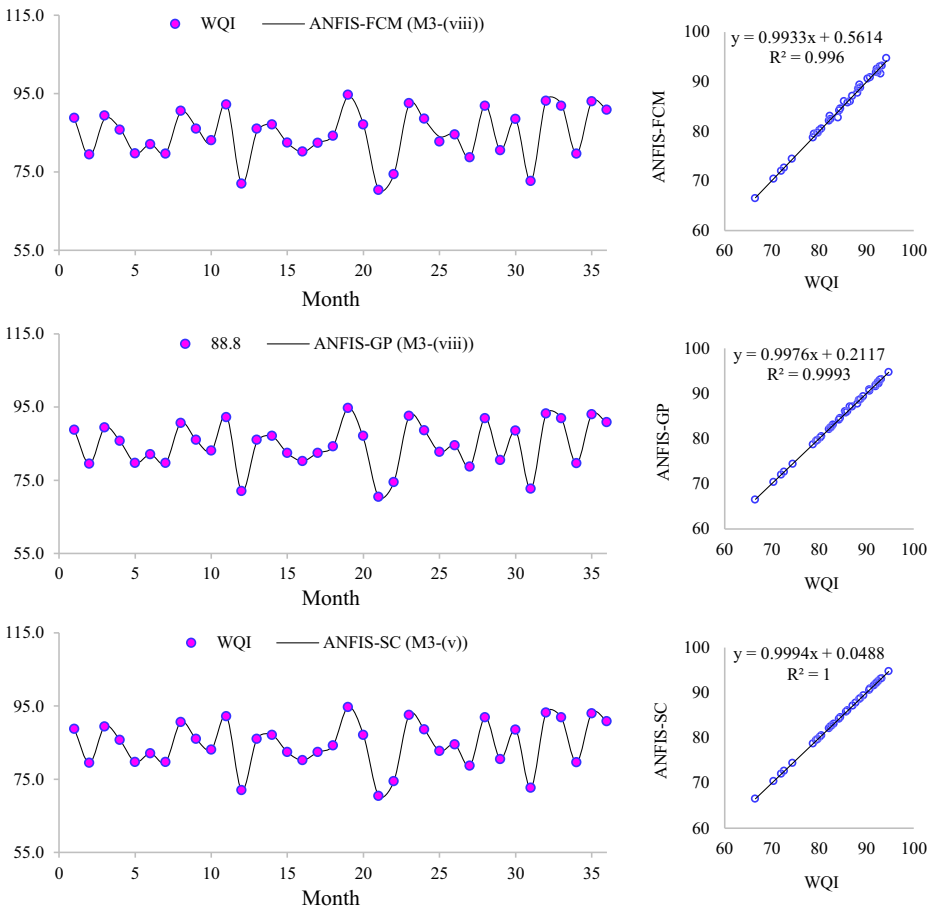
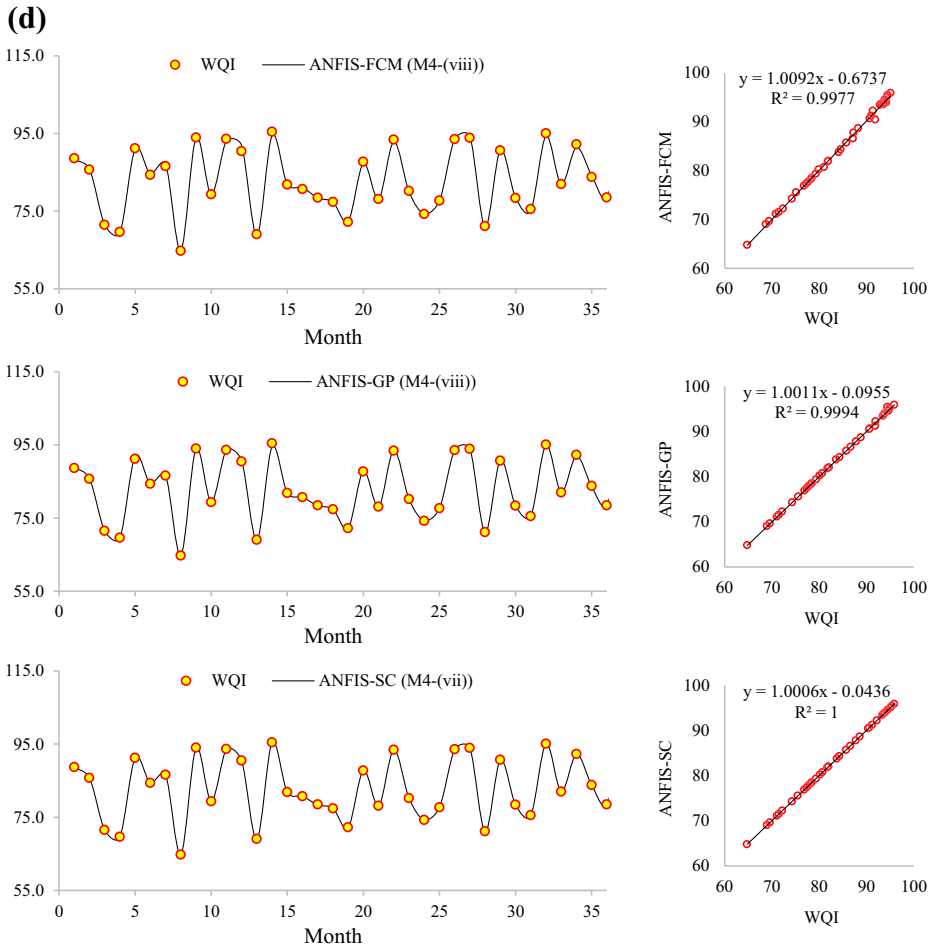


Fig. 3 (continued)

the best results for the optimal models M2, M3, and M4 for both ANFIS- GP and ANFIS-FCM.

In overall, the cluster based ANFIS-FCM and ANFIS-SC models perform superior to the ANFIS-GP in WQI modelling. The main advantage of these cluster based ANFIS models is their less complexity in comparison to the GP based models. ANFIS-FCM and ANFIS-SC models which comprised less number of parameters take less computational time than the ANFIS-GP model in modelling WQI in the current study. For the M4 data set and input combination (viii), the optimal ANFIS-GP model has 8 inputs and 2 Gumbel membership functions (gbellmf) for each input, so it has  $8 \times 2 \times 3 = 48$  premise parameters. Additionally, it has constant output and therefore  $2^8$  rules comprising 256 consequent parameters. Totally, the best ANFIS-GP model has  $48 + 256 = 304$  parameters. The ANFIS-FCM model has 3 clusters representing 3 Gaussian membership functions comprising 2 parameters for each input. It has  $8 \times 3 \times 2 = 48$  premise parameters. Additionally, it has linear output and 3 rules comprising  $3 \times (8 + 1) = 27$  consequent parameters. Totally, the best ANFIS-FCM model has  $48 + 27 = 75$





**Fig. 3** (continued)

parameters. Similar to the cluster based ANFIS-FCM, the ANFIS-SC model with 0/7 radii has 2 clusters representing 2 Gaussian membership functions comprising 2 parameters for each input. It has  $8 \times 2 \times 2 = 32$  premise parameters. Additionally, it has linear output and 2 rules comprising  $2 \times (8 + 1) = 18$  consequent parameters. Totally, the best ANFIS-SC model has  $32 + 18 = 50$  parameters. Hence, it should be noted that the cluster based ANFIS-FCM and ANFIS-SC models which are simpler than the ANFIS-GP model can be successfully applied for modelling WQI.

### 4 Conclusion

By the present study, the ability of three different hybrid neuro fuzzy methods, ANFIS-GP, ANFIS-FCM and ANFIS-SC, in modelling WQI by using monthly water quality parameters measured in Selangor river basin, Malaysia was investigated. For better evaluation of the soft

computing methods, cross validation was employed by dividing data into 4 equal parts. Eight different input combinations were decided via applying correlation analysis and each model was tested four times for each input combination. From the comparison results, the following results can be drawn:

- All three hybrid models provided good estimates for the WQI at studied river basin. According to the mean of the applied statistics, the cluster based ANFIS-SC and ANFIS-FCM performed superior accuracy to the ANFIS-GP.
- The best ANFIS-SC model (input combination viii) increased the RMSE, MAE and  $R^2$  accuracy of the best ANFIS-FCM (input combination viii) by 46, 39, 1.7% and the best ANFIS-GP (input combination vi) by 69, 65 and 7.4, respectively.
- It was found that the cluster based ANFIS-FCM and ANFIS-SC models have much less parameters resulting less computational time in comparison to the ANFIS-GP models.
- Among the applied models, the ANFIS-SC models were found to have better accuracy than the other hybrid models and should be preferred in modelling monthly WQI.
- Overall, the applied hybrid fuzzy intelligence model demonstrated a robust and authoritative model for WQI prediction and more for the environmental river engineering practice, particularly for Selangor River located in tropical environment, Malaysian region.

In this study, the ability of three different neuro fuzzy methods in modeling water quality parameters was examined for the Selangor river basin. In future studies, more data from different areas may be used for deriving more concrete conclusions. The applied methods may also be compared with other soft computing techniques such as support vector machine, extreme learning machine etc.

**Acknowledgements** Authors would like to acknowledge their gratitude and appreciate for the Department of Environment Malaysia, for providing the river water quality variables data set of the studied case study and their admirable cooperation. We thank all reviewers and the editor for their insightful comments that improved the clarity of the final paper.

## References

- Abrahão R, Carvalho M, Da Silva WR et al (2007) Use of index analysis to evaluate the water quality of a stream receiving industrial effluents. *Water SA* 33(4):459–465. <https://doi.org/10.4314/wsa.v33i4.52940>
- Ahmed AAM, Shah SMA (2015) Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River. *J King Saud Univ Eng Sci* 29(3):237–243. <https://doi.org/10.1016/j.jksues.2015.02.001>
- Arsad A, Abustan I, Rawi CSM, Syafalni S (2012) Integrating biological aspects into river water quality research in Malaysia: an opinion. *OIDA Int J Sustain Dev* 4:107–122
- Avvannavar SM, Shrihari S (2008) Evaluation of water quality index for drinking purposes for river Netravathi, Mangalore, South India. *Environ Monit Assess* 143(1-3):279–290. <https://doi.org/10.1007/s10661-007-9977-7>
- Barzegar R, Asghari Moghaddam A, Adamowski J, Ozga-Zielinski B (2017) Multi-step water quality forecasting using a boosting ensemble multi-w Stoch Env Res Risk A 1–15. <https://doi.org/10.1007/s00477-017-1394-z>
- Bezdek JC (1973) Cluster validity with fuzzy sets. *J Cybern* 3:58–73. <https://doi.org/10.1080/01969727308546047>
- Bharti N, Katyal D (2011) Water quality indices used for surface water vulnerability assessment. *J Environ Sci* 2: 154–173. <https://doi.org/10.6088/ijes.00202010017>
- Bordalo AA, Teixeira R, Wiebe WJ (2006) A water quality index applied to an international shared river basin: the case of the Douro River. *Environ Manag* 38(6):910–920. <https://doi.org/10.1007/s00267-004-0037-6>

- Cobaner M (2011) Evapotranspiration estimation by two different neuro-fuzzy inference systems. *J Hydrol* 398(3–4):292–302
- Cude CG (2001) Oregon water quality index: a tool for evaluating water quality management effectiveness. *J Am Water Resour Assoc* 37(1):125–137. <https://doi.org/10.1111/j.1752-1688.2001.tb05480.x>
- Dada AC, Asmat A, Gires U et al (2012) Bacteriological monitoring and sustainable management of beach water quality in Malaysia: problems and prospects. *Global J Health Sci* 4(3):126–138. <https://doi.org/10.5539/gjhs.v4n3p126>
- Das Gupta A (2008) Implication of environmental flows in river basin management. *Phys Chem Earth* 33(5): 298–303. <https://doi.org/10.1016/j.pce.2008.02.004>
- 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(3):645–656. <https://doi.org/10.1007/s13762-013-0378-x>
- Fulazzaky MA, Seong TW, Masirin MIM (2010) Assessment of water quality status for the selangor river in Malaysia. *Water Air Soil Pollut* 205(1–4):63–77. <https://doi.org/10.1007/s11270-009-0056-2>
- Gazzaz NM, Yusoff MK, Aris AZ, Juahir H, Ramli MF (2012) Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Mar Pollut Bull* 64(11): 2409–2420. <https://doi.org/10.1016/j.marpolbul.2012.08.005>
- Grabowski R, Gurnell AM (2016) Hydrogeomorphology- ecology interactions in river systems. *River Res Appl* 22:1085–1095. <https://doi.org/10.1002/tra>
- Hameed M, Sharqi SS, Yaseen ZM, Afan HA, Hussain A, Elshafie A (2016) Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. *Neural Comput Applic* 28(S1):1–13. <https://doi.org/10.1007/s00521-016-2404-7>
- Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23(3): 665–685. <https://doi.org/10.1109/21.256541>
- Karimi S, Kisi O, Shiri J, Makarynsky O (2013) Neuro-fuzzy and neural network techniques for forecasting sea level in Darwin Harbor, Australia. *Comput Geosci* 52:50–59. <https://doi.org/10.1016/j.cageo.2012.09.015>
- Khalil B, Ouarda TBMJ, St-Hilaire a. (2011) Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *J Hydrol* 405(3–4):277–287. <https://doi.org/10.1016/j.jhydrol.2011.05.024>
- Kisi O, Ay M (2013) Modelling COD concentration by using three different ANFIS techniques. In: International Balkans Conference on Challenges of Civil Engineering, pp 23–25
- Kisi O, Zounemat-Kermani M (2014) Comparison of two different adaptive Neuro-fuzzy inference Systems in Modelling Daily Reference Evapotranspiration. *Water Resour Manag* 28(9):2655–2675. <https://doi.org/10.1007/s11269-014-0632-0>
- Kisi O, Zounemat-Kermani M (2016) Suspended sediment modeling using Neuro-fuzzy embedded fuzzy c-means clustering technique. *Water Resour Manag* 30(11):3979–3994. <https://doi.org/10.1007/s11269-016-1405-8>
- Kisi O, Sanikhani H, Zounemat-Kermani M, Niazi F (2015) Long-term monthly evapotranspiration modeling by several data-driven methods without climatic data. *Comput Electron Agric* 115:66–77. <https://doi.org/10.1016/j.compag.2015.04.015>
- Kisi O, Sanikhani H, Cobaner M (2016) Soil temperature modeling at different depths using neuro-fuzzy, neural network, and genetic programming techniques. *Theor Appl Climatol* 129(3–4):1–16. <https://doi.org/10.1007/s00704-016-1810-1>
- Lai YC, Yang CP, Hsieh CY, Wu CY, Kao CM (2011) Evaluation of non-point source pollution and river water quality using a multimedia two-model system. *J Hydrol* 409(3–4):583–595. <https://doi.org/10.1016/j.jhydrol.2011.08.040>
- Liu M, Lu J (2014) Support vector machine-an alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river? *Environ Sci Pollut Res* 21(18):11036–11053. <https://doi.org/10.1007/s11356-014-3046-x>
- Maier HR, Morgan N, Chow CWK (2004) Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. *Environ Model Softw* 19(5):485–494. [https://doi.org/10.1016/S1364-8152\(03\)00163-4](https://doi.org/10.1016/S1364-8152(03)00163-4)
- Niroobakhsh M (2012) Prediction of water quality parameter in Jajrood River basin: application of multi layer perceptron (MLP) perceptron and radial basis function networks of artificial neural networks (ANNs). *Afr J Agric Res* 7(29):4131–4139. <https://doi.org/10.5897/AJAR11.1645>
- Nourani V, Khanghah TR, Sayyadi M, et al. (2013) Application of the artificial neural network to monitor the quality of treated water. *Int J Manag Inf Technol* 2(2):38–45
- Orouji H, Haddad O (2013) Modeling of water quality parameters using data-driven models. *J Environ Eng* 139(7):947–957. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0000706](https://doi.org/10.1061/(ASCE)EE.1943-7870.0000706)
- Othman F, AE ME, Mohamed I (2012) Trend analysis of a tropical urban river water quality in Malaysia. *J Environ Monit* 14(12):3164–3173. <https://doi.org/10.1039/c2em30676j>

- Ouyang Y (2005) Evaluation of river water quality monitoring stations by principal component analysis. *Water Res* 39(12):2621–2635. <https://doi.org/10.1016/j.watres.2005.04.024>
- Palani S, Liong SY, Tkalich P (2008) An ANN application for water quality forecasting. *Mar Pollut Bull* 56(9): 1586–1597. <https://doi.org/10.1016/j.marpolbul.2008.05.021>
- Patki VK, Shrihari S, Manu B, Deka PC (2013) Fuzzy system modeling for forecasting water quality index in municipal distribution system. *Urban Water J* 12(2):89–110. <https://doi.org/10.1080/1573062X.2013.820333>
- Sahu M, Mahapatra SS, Sahu HB, Patel RK (2011) Prediction of water quality index using Neuro fuzzy inference system. *Water Qual Expo Health* 3(3–4):175–191. <https://doi.org/10.1007/s12403-011-0054-7>
- Sanikhani H, Kisi O (2012) River flow estimation and forecasting by using two different adaptive Neuro-fuzzy approaches. *Water Resour Manag* 26(6):1715–1729. <https://doi.org/10.1007/s11269-012-9982-7>
- Sanikhani H, Kisi O, Nikpour MR, Dinpashoh Y (2012) Estimation of daily pan evaporation using two different adaptive Neuro-fuzzy computing techniques. *Water Resour Manag* 26(15):4347–4365. <https://doi.org/10.1007/s11269-012-0148-4>
- Sanikhani H, Kisi O, Kiafar H, Ghavidel SZZ (2015) Comparison of different data-driven approaches for modeling Lake level fluctuations: the case of Manyas and Tuz Lakes (Turkey). *Water Resour Manag* 29(5):1557–1574. <https://doi.org/10.1007/s11269-014-0894-6>
- Sargaonkar A, Deshpande V (2003) Development of an overall index of pollution for surface water based on a general classification scheme in Indian context. *Environ Monit Assess* 89(1):43–67. <https://doi.org/10.1023/A:1025886025137>
- Shuhaimi-Othman M, Lim EC, Mushrifah I (2007) Water quality changes in Chini Lake, Pahang, West Malaysia. *Environ Monit Assess* 131(1–3):279–292. <https://doi.org/10.1007/s10661-006-9475-3>
- Song T, Kim K (2009) Development of a water quality loading index based on water quality modeling. *J Environ Manag* 90(3):1534–1543. <https://doi.org/10.1016/j.jenvman.2008.11.008>
- Tabari H, Kisi O, Ezani A, Hosseinzadeh Talaei P (2012) SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *J Hydrol* 444–445:78–89. <https://doi.org/10.1016/j.jhydrol.2012.04.007>
- Tyagi S, Sharma B, Singh P, Dobhal R (2013) Water quality assessment in terms of water quality index. *Am J Water Res* 1:34–38. <https://doi.org/10.12691/ajwr-1-3-3>
- Wang L, Li X, Cui W (2012) Fuzzy neural networks enhanced evaluation of wetland surface water quality. *Int J Comput Appl Technol* 44(3):235. <https://doi.org/10.1504/IJCAT.2012.049087>
- Wei-Bo C, Wen-Cheng L (2015) Water quality modeling in reservoirs using multivariate linear regression and two neural network models. *Adv Artif Neural Syst* 2015:1–12. <https://doi.org/10.1155/2015/521721>
- Yan H, Zou Z, Wang H (2010) Adaptive neuro fuzzy inference system for classification of water quality status. *J Environ Sci (China)* 22(12):1891–1896. [https://doi.org/10.1016/S1001-0742\(09\)60335-1](https://doi.org/10.1016/S1001-0742(09)60335-1)
- Zaman Zad Ghavidel S, Montaseri M (2014) Application of different data-driven methods for the prediction of total dissolved solids in the Zarinerohd basin. *Stoch Env Res Risk A* 28(8):2101–2118. <https://doi.org/10.1007/s00477-014-0899-y>

## Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

[onlineservice@springernature.com](mailto:onlineservice@springernature.com)