




# Determination of biochemical oxygen demand and dissolved oxygen for semi-arid river environment: application of soft computing models

Hai Tao<sup>1</sup> · Aiman M. Bobaker<sup>2</sup> · Majeed Mattar Ramal<sup>3</sup> · Zaher Mundher Yaseen<sup>4</sup>  · Md Shabbir Hossain<sup>5</sup> · Shamsuddin Shahid<sup>6</sup>

Received: 6 September 2018 / Accepted: 1 November 2018 / Published online: 12 November 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Surface and ground water resources are highly sensitive aquatic systems to contaminants due to their accessibility to multiple-point and non-point sources of pollutions. Determination of water quality variables using mathematical models instead of laboratory experiments can have venerable significance in term of the environmental prospective. In this research, application of a new developed hybrid response surface method (HRSM) which is a modified model of the existing response surface model (RSM) is proposed for the first time to predict biochemical oxygen demand (BOD) and dissolved oxygen (DO) in Euphrates River, Iraq. The model was constructed using various physical and chemical variables including water temperature (T), turbidity, power of hydrogen (pH), electrical conductivity (EC), alkalinity, calcium (Ca), chemical oxygen demand (COD), sulfate (SO<sub>4</sub>), total dissolved solids (TDS), and total suspended solids (TSS) as input attributes. The monthly water quality sampling data for the period 2004–2013 was considered for structuring the input-output pattern required for the development of the models. An advance analysis was conducted to comprehend the correlation between the predictors and predictand. The prediction performances of HRSM were compared with that of support vector regression (SVR) model which is one of the most predominate applied machine learning approaches of the state-of-the-art for water quality prediction. The results indicated a very optimistic modeling accuracy of the proposed HRSM model to predict BOD and DO. Furthermore, the results showed a robust alternative mathematical model for determining water quality particularly in a data scarce region like Iraq.

**Keywords** Water quality variables · Environmental prospects · Soft computing models · Euphrates River

---

Responsible editor: Marcus Schulz

---

✉ Zaher Mundher Yaseen  
yaseen@tdtu.edu.vn

- <sup>1</sup> Computer Science Department, Baoji University of Arts and Sciences, Baoji, Shaanxi, China
- <sup>2</sup> Chemistry Department, Faculty of Science, University of Benghazi, Benghazi, Libya
- <sup>3</sup> Dams and Water Resources Department, College of Engineering, University Of Anbar, Ramadi, Iraq
- <sup>4</sup> Sustainable Developments in Civil Engineering Research Group, Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam
- <sup>5</sup> Institute of Energy Infrastructure, Department of Civil Engineering, Universiti Tenaga Nasional, Kajang, Malaysia
- <sup>6</sup> Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia

## Introduction

Water is vital in many aspects of human life such as for drinking, personal hygiene, agricultural purposes, manufacturing and industrial processes, biotransformation, power generation, and contamination dissolution releasing (Zhang et al. 2012; Wu et al. 2018). Unsustainable anthropogenic activities often polluted water bodies and causes high stress on freshwater resources (Chau 2005). Because of the frequent episodes of water pollution in recent times, the prediction and assessment of water quality have gradually attracted the attention of the environmental management department of many countries (Gümrah et al. 2000; Page et al. 2017).

Iraq experienced a remarkable increase in water shortage in the last two decades due to intervention of water flow in the upstream of major rivers, changes in climate, and gradual declination of rainfall (Kadhem 2013; Zolnikov 2013). Water quality is another major problem that Iraq is facing for the past couple of decades (Zolnikov 2013). This is particularly becoming a major concern for Euphrates River where

water quality is drastically aggravated in recent years due to agricultural developments. The issue of river water quality has become critical for the country as it has gone beyond the standard required for industrial, domestic, and agricultural purposes (Rahi and Halihan 2010).

The quality of river water is defined by various physical, chemical, and biological properties of water. Among all the water quality parameters, dissolved oxygen (DO) is considered as the most important water quality parameter as it is essential for the survival of all aquatic organisms. Biochemical oxygen demand (BOD), on the other hand, is a measure of the amount of DO in river and thus defines the amount of organic matter available for oxygen-consuming bacteria. DO and BOD is a composite index that can be used to assess the favorable conditions for aquatic life and overall quality of water. The DO and BOD affects a large number of biological, chemical, and physical properties of water and thus considered as the most important index of water quality. The stream pollution control and management of river water quality and ecology activities are largely hinged on accurate determination of these two parameters. However, the analysis of these parameters is delicate and time-consuming compared to other water quality parameters. A great amount of cost, time, and energy can be saved if these water quality parameters can be predicted in a reasonable accuracy. This has inspired researchers to develop reliable models for prediction of BOD and DO from other easily available water quality data.

Modeling and forecasting of river water quality parameters is a challenging issue for a long time. The BOD and DO depend on many biotic and abiotic factors and their complex interactions. The knowledge of many of these interactions are still not clear, the data required for modeling such processes are difficult to acquire, and mathematical formulations of the processes are often very difficult. Therefore, physical-based models generally used for BOD and DO modeling simplify these complex physical processes and therefore often fail to predict BOD and DO with reasonable accuracy.

The BOD and DO in water bodies are found to change with time and follow stochastic behavior which encouraged developing stochastic prediction models. Regression models are most widely used for modeling stochastic behavior of BOD and DO. However, highly stochastic behavior of BOD and DO makes the reliable simulation of those parameters using conventional regression models a difficult task. It is expected that prediction models must have the high prescient ability in describing water quality. Therefore, simple statistical regression-based model cannot be used for operational management of river water quality.

Soft computing models such as artificial intelligence (AI) provide an excellent and reliable technique for modeling surface and underground water quality (Gümrah et al. 2000; Shi et al. 2018). Nevertheless, AI models exhibited robust and reliable modeling strategies for multiple hydrological,

climatological, and environmental applications (Wang et al. 2014; Olyaei et al. 2015; Chen and Chau 2016). The main advantage of the AI models is their capability of handling the highly complicated nonlinear inter-parameter relationship (Barzegar et al. 2016) on the contrary of the conventional statistical models that are based on the assumption of linear relationship. The applications of AI have been presented in several predictive model forms such as artificial neural network (Sudheer et al. 2006; Zou et al. 2007; May 2008; Palani et al. 2008; Singh et al. 2009; Song et al. 2010; Balabin et al. 2011; Khalil et al. 2011; Gazzaz et al. 2012; Klaslan et al. 2014; Wu et al. 2014), support vector machine (Xu et al. 2007; Bouamar and Ladjal 2008; Yunrong and Liangzhong 2009a; Jian et al. 2010; Singh et al. 2011; Liu and Lu 2014; Jadhav et al. 2015), adaptive neuro-inference system model (Sahu et al. 2011; Emamgholizadeh et al. 2014; Najah et al. 2014; Ahmed and Shah 2015), and genetic programming (Muttill and Chau 2006; Sreekanth and Datta 2010; Orouji and Haddad 2013; Olyaei et al. 2017). On the other hand, hybrid intelligence models revealed a good performance in modeling water quality parameters (Wang et al. 2010; Liu et al. 2013; Deng et al. 2015; Barzegar et al. 2016; Ravansalar et al. 2016). In spite of the enormous implementation of AI in water quality modeling, there are still several downsides of these AI models such as difficulty in tuning internal parameters, time-consuming algorithms, human modeling interaction, and lack of generalization. Therefore, exploring new and robust mathematical models that are featured by high flexibility in solving complicated environmental phenomena are on progress (Behmel et al. 2016).

Most recently, a new mathematical model called response surface method (RSM) has been well recognized for its ability to solve complex regression problem effectively (Cho 2007; Kewlani and Iagnemma 2008; Kim and Choi 2008; Wei et al. 2008; Acherjee et al. 2009; Roussouly et al. 2012). The main advantage of RSM is its employment of high-order polynomial function (Keshtegar et al. 2016). The precision of the RSM model relies upon the fundamental numerical capacity in light of the fact that the essential response surface function frames are given adaptability to model the targeted application (i.e., water quality variable). In the current research, a hybrid response surface method (HRSM) has been developed for the first time to predict water quality variables. The motivation of proposing this study was its successful implementation in the field of hydrology (Keshtegar et al. 2017).

The HRSM models have been developed in this study for the prediction of BOD and DO from other easily available water quality parameters including water temperature (T), turbidity, power of hydrogen (pH), electrical conductivity (EC), alkalinity, calcium (Ca), chemical oxygen demand (COD), sulfate (SO<sub>4</sub>), total dissolved solids (TDS), and total suspended solids (TSS). The modeling result is verified against the support vector regression (SVR) model. Various

AI models have been proposed for simulations of river water quality parameters as mentioned above. However, SVR has been reported in literature as the most predominate AI model in prediction of environmental phenomena (Fahimi et al. 2016; Fengxiang et al. 2010; Singh et al. 2011; Yunrong and Liangzhong 2009b). This study aims to develop a robust mathematical model for the prediction of BOD and DO in river water in order to aid river water quality management in a data scarce region such as Iraq. This type of model is extremely important for a developing country like Iraq where the amount of assigned budget for environmental quality monitoring and assessment is very limited, but the water pollution is very frequent and more disastrous. Hence, establishing the current research is highly significant for the sake of providing intelligent system to monitor the water quality variables of this most vital river of Iraq. To the best of the knowledge of the authors, there is no research previously conducted on such prospective and thus the novelty is presented at this point in addition to the proposed methodology.

### Dataset and description of the study area

The water quality parameters of Euphrates River measured at Ramadi City, Anbar, Iraq (latitude 33°26'15"N; longitude 43°16'52"E) was used in the study (Fig. 1). The water quality of the Euphrates River has become a serious issue in recent years. The return flows from agricultural land and dumping of untreated sewage into the river and its tributaries for a long time have caused gradual deteriorations of the quality of river water. Therefore, forecasting water quality of Euphrates River is very important for environmental quality monitoring and management. The water sample at the intake of a large drinking water treatment plant in Ramadi City was collected for laboratory measurement of water quality parameters. The sampling process was based on monthly scale over the period of 2004–2013. Long-term reliable water quality data is a major problem in Iraq. Water quality data was available only for those 10 years when the study was conducted, and the available data was fully utilized in the present study. The main sources of water contamination of the Euphrates River are agricultural and domestic wastes. The salinity of river water is very high which increases along the course of the stream. In addition, discharging of the untreated sewage water in the river and its tributaries adds a serious hazard associated with different types of water contaminants. The analysis of water quality parameters was done upon ten physical and chemical water properties including *T*, turbidity, pH, EC, slkalinity, Ca, COD, SO<sub>4</sub>, TDS, TSS, DO, and BOD. The BOD and DO in river water system are affected by these water quality parameters, and therefore, those are selected for the development of prediction models.

## Theoretical review of the predictive models

### Hybrid response surface method

The RSM can be generally described as a set of approximation polynomials limited to quadratic order for experimental calibration. A set of polynomial functions for modeling of water quality variables (BOD and DO) at monthly time scale was derived in this study. The regression of several data points was used to obtain the polynomial set coefficients. The general method for a quadratic order approximating polynomial using RSM that was widely applied in previous researches (e.g., Afan et al. 2017; Keshtegar et al. 2016; Yeniay 2014) involves the following second-order function:

$$Y = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i=1}^n \sum_{j=i+1}^n a_{ij} x_i x_j \tag{1}$$

The conceptual idea of prediction “multivariate regression problem” using RSM hinges on the polynomial functions in high-order. Traditional RSM typically uses low-order polynomials to approximate highly nonlinear functions and thus suffers from limited predictive capability. Thus, selecting the appropriate polynomials is essential to attain a high level of accuracy. Different approaches such as multi-layer regression, Taguchi optimization, etc. have been adapted to improve the prediction characteristics RSM compared to those uses typical polynomial-based RSM. The main modification performed in the latest equation consisted of combining the polynomial and exponential functions to obtain a hybrid function. It is expected that if the polynomial expression is able to approximate the highly nonlinear relationship like the exponential relationship over an extended range, it will able to provide better prediction.

For data having small standard deviation, the data points usually have narrow bounds with normal distribution function near the mean. Hence, the input data to this function are usually normalized via the following equation:

$$N(X_n) = \frac{x_n^i - x_n^{mean}}{2\sigma_{xn}} \tag{2}$$

where  $x_n^{mean}$  and  $2\sigma_{xn}$  are the average and standard deviation of a normally distributed function in which  $2\sigma_{xn}$  can be replaced with  $\sigma_{xn}$  because of the narrow bounds with normal distribution function near the mean value. With respect to the following normalized exponential function (Eq. 3), the calibration of target values (BOD and DO in this case) considering normalized input water quality parameters can be done as

$$Y_n = a_{1n} + a_{2n} \exp[N(X_n)] \tag{3}$$

To approximate the unknown coefficients  $a_{1n}$  and  $a_{2n}$ , Eq. (3) can be incorporated with Eq. (1) to obtain the HRS function:

**Fig. 1** The case study location Ramadi water plant station located on the Euphrates River in Iraq



$$Y = a_0 + \sum_{i=1}^n a_i Y_{ni} + \sum_{i=1}^n a_{ii} Y_{ni}^2 + \sum_{i=1}^n \sum_{j=i+1}^n a_{ij} Y_{ni} Y_{nj} \quad (4)$$

where  $Y_{ni}$  is the same as given in Eq. (3). The unknown coefficients in the latest equation are usually estimated using the following error function:

$$e = [E - Y]^T [E - Y] \quad (5)$$

where  $Y = P(Y_n)^T a$ ,  $E$  are the observed target values (BOD and DO). The basic polynomial function that depends on Eq. (3) can be computed as

$$P(Y_n) = [1, y_{n1}, y_{n2}, \dots, y_{nn}, y_{n1}^2, y_{n2}^2, \dots, y_{nn}^2, y_{n1}y_{n2}, y_{n1}y_{n3}, \dots, y_{n_{n-1}}y_{nn}] \quad (6)$$

Following minimization of the error function given in Eq. (5), the unknown coefficients can be estimated as follows:

$$a = [P(Y_n)^T P(Y_n)]^{-1} [P(Y_n)^T E] \quad (7)$$

By combining the latest equation with  $Y$  given above, the prediction of BOD and DO can be fulfilled as

$$Y = P(Y_n)^T [P(Y_n)^T P(Y_n)]^{-1} [P(Y_n)^T E] \quad (8)$$

For illustration purpose, Fig. 2 shows the proposed hybrid response surface method. The figure displays four layers involved in the structure of the HRSM model that can be used for the prediction using exponential functions and hybrid polynomial.

The input data is presented in the first layer. The second layer defines the normalization process for the supplied data. This is followed by the calibration of the targeted BOD and DO variables in accordance with the normalized input attributes using Eq. (3), reported earlier. At the last layer, the regression problem is solved using the second-order polynomial function given in Eq. (4). More information of the development of HRSM can be found in other published researches (Keshtegar and Heddam 2017; Keshtegar et al. 2017).

### Support vector regression

As a distinguished intelligence predictive model, SVR had been applied successfully in environmental studies (Singh et al. 2011; Fahimi et al. 2016). Therefore, in this study, it was selected to verify the prediction proficiency of HRSM model. Recently, the SVR has been applied in several areas such as soft computing, environmental studies, and engineering as a learning algorithm. It has demonstrated a better prediction and forecasting accuracies when compared with other forecasting methods like neural network (Liu and Lu 2014). The process and theory of SVR development is available in

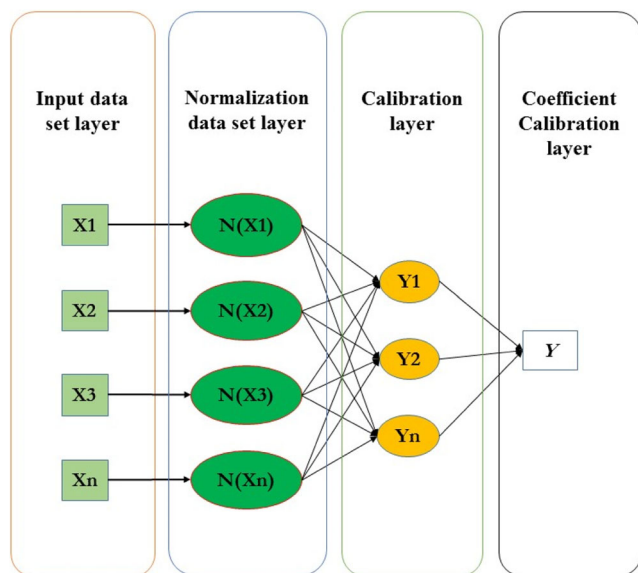


Fig. 2 The structure of the proposed HRSM predictive model

the literature (Vapnik 1995). A statistical way of machine learning and minimization of structural risk form the basis for the development of SVR, aimed at reducing upper bound error compared to the commonly experienced local training error in other machine learning methods. Based on recent evaluations, there are several improvements in the SVR compared to other soft computing learning algorithms such as the implementation of a set of kernel equations that is highly dimensionally spaced but does not involve nonlinear transformations, thereby making data to be indispensable and linearly separable since there is no room for assumption during the functional transformation. Besides, the method is unique in its solution due to the convex nature of the optimal problem. Different algorithms have been proposed for optimization of the internal parameters of SVR including bat algorithm, firefly algorithm, particle swarm optimization, and univariate

marginal distribution algorithm. Among those, the nature-inspired algorithm, firefly, has been highly recommended in recent studies (Ch et al. 2014; Moghaddam et al. 2016; Shamshirband et al. 2016; Ebtehaj et al. 2017; Ghorbani et al. 2017a, b, c; Tao et al. 2018). Yang (2010) developed firefly (FFA) algorithm as a biologically inspired metaheuristic optimization algorithm which depends on certain biological behaviors such as the characteristic flashing of light by Fireflies. Fireflies attract preys or mates through bioluminescence. When compared to other conventional metaheuristic algorithms, the FFA has shown promising, efficient, interesting, and robust potentials in achieving global optimization.

**Evaluation of model performance**

The simulation was done by using a new set of input variables, and the results obtained using HRSM and SVR models were compared with the observed BOD and DO, using different performance measuring indicators including scatter index (SI), mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute errors (MAE), root mean square relative error (RMSRE), mean relative error (MRE), BIAS, and correlation coefficient (R<sup>2</sup>), following most of machine learning researches evaluation (Moeeni et al. 2017):

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \tag{10}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| \times 100 \tag{11}$$

**Table 1** Basic statistics of the measured water quality variables

| Variable        | Unit  | Min  | Max  | Median | Mean   | SD    | CV%  |
|-----------------|-------|------|------|--------|--------|-------|------|
| Temperature     | °C    | 9    | 38   | 22     | 21.8   | 8.2   | 0.37 |
| Turbidity       | NTU   | 7.3  | 73.4 | 12.4   | 18.9   | 15    | 0.79 |
| pH              | –     | 7.3  | 8.1  | 7.8    | 7.7    | 0.1   | 0.02 |
| EC              | µs/cm | 1324 | 1896 | 1449.5 | 1484.4 | 125.2 | 0.08 |
| Alkalinity      | Mg/l  | 104  | 162  | 124    | 125.4  | 12.4  | 0.09 |
| Ca              | Mg/l  | 77   | 105  | 85     | 86.3   | 6.3   | 0.07 |
| COD             | Mg/l  | 8.3  | 108  | 11.1   | 12.2   | 9     | 0.73 |
| SO <sub>4</sub> | Mg/l  | 212  | 449  | 385    | 383.5  | 32.9  | 0.08 |
| TDS             | Mg/l  | 863  | 1289 | 1113.5 | 1102.7 | 93.8  | 0.08 |
| TSS             | Mg/l  | 12   | 196  | 35     | 56.7   | 49.5  | 0.87 |
| BOD             | Mg/l  | 2.9  | 5    | 3.7    | 3.8    | 0.5   | 0.15 |
| DO              | Mg/l  | 5.9  | 7.9  | 7.2    | 7      | 0.5   | 0.08 |

**Table 2** The correlation statistic between each inspected water quality (predictor attribute) and the targeted water quality BOD and DO

| Input variables attributes | Targeted variables |       |
|----------------------------|--------------------|-------|
|                            | BOD                | DO    |
| Temperature                | 0.678              | 0.746 |
| Turbidity                  | 0.250              | 0.332 |
| pH                         | 0.361              | 0.316 |
| EC                         | 0.330              | 0.360 |
| Alkalinity                 | 0.281              | 0.331 |
| Ca                         | 0.126              | 0.180 |
| COD                        | 0.442              | 0.304 |
| SO <sub>4</sub>            | 0.040              | 0.032 |
| TDS                        | 0.200              | 0.159 |
| TSS                        | 0.273              | 0.351 |

**Table 3** The investigated input combinations to predict BOD and DO water quality variables

| Models | Temperature | Turbidity | pH | EC | Alk | CA | COD | SO4 | TDS | TSS |
|--------|-------------|-----------|----|----|-----|----|-----|-----|-----|-----|
| M1     | √           |           |    |    |     |    |     |     |     |     |
| M2     | √           | √         |    |    |     |    |     |     |     |     |
| M3     | √           | √         | √  |    |     |    |     |     |     |     |
| M4     | √           | √         | √  | √  |     |    |     |     |     |     |
| M5     | √           | √         | √  | √  | √   |    |     |     |     |     |
| M6     | √           | √         | √  | √  | √   | √  |     |     |     |     |
| M7     | √           | √         | √  | √  | √   | √  | √   |     |     |     |
| M8     | √           | √         | √  | √  | √   | √  | √   | √   |     |     |
| M9     | √           | √         | √  | √  | √   | √  | √   | √   | √   |     |
| M10    | √           | √         | √  | √  | √   | √  | √   | √   | √   | √   |

$$SI = \frac{RMSE}{\bar{x}} \tag{12}$$

$$R^2 = \frac{\left(\sum_{i=1}^n (x_i - \bar{x}) - (y_i - \bar{y})\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \times 100 \tag{13}$$

$$RMSRE = \sqrt{\frac{\sum_{i=1}^n \left(\frac{x_i - y_i}{x_i}\right)^2}{n}} \tag{14}$$

$$BIAS = \frac{\sum_{i=1}^n (x_i - y_i)}{n} \tag{15}$$

**Table 4** The statistical performance indicators for the testing period of BOD variable prediction using HRSM and SVR models

| Models   | SI    | MAPE   | RMSE  | MAE   | RMSRE | MRE    | BIAS   | R <sup>2</sup> |
|----------|-------|--------|-------|-------|-------|--------|--------|----------------|
| HRSM     |       |        |       |       |       |        |        |                |
| Model 1  | 0.073 | 5.831  | 0.278 | 0.214 | 0.076 | 0.006  | -0.002 | 0.67           |
| Model 2* | 0.035 | 2.764  | 0.136 | 0.103 | 0.035 | -0.002 | 0.016  | 0.92           |
| Model 3  | 0.079 | 5.588  | 0.301 | 0.217 | 0.075 | -0.034 | 0.147  | 0.70           |
| Model 4  | 0.117 | 9.722  | 0.442 | 0.374 | 0.112 | -0.054 | 0.236  | 0.40           |
| Model 5  | 0.122 | 10.05  | 0.462 | 0.389 | 0.115 | -0.042 | 0.201  | 0.26           |
| Model 6  | 0.112 | 6.905  | 0.426 | 0.273 | 0.102 | -0.048 | 0.206  | 0.44           |
| Model 7  | 0.149 | 10.14  | 0.566 | 0.403 | 0.137 | -0.074 | 0.314  | 0.21           |
| Model 8  | 0.120 | 7.818  | 0.456 | 0.310 | 0.109 | -0.054 | 0.231  | 0.37           |
| Model 9  | 0.126 | 8.000  | 0.477 | 0.320 | 0.115 | -0.042 | 0.185  | 0.31           |
| Model 10 | 0.096 | 7.014  | 0.363 | 0.263 | 0.098 | -0.013 | 0.072  | 0.49           |
| SVR      |       |        |       |       |       |        |        |                |
| Model 1  | 0.097 | 8.332  | 0.369 | 0.304 | 0.102 | 0.005  | 0.014  | 0.41           |
| Model 2  | 0.119 | 9.279  | 0.452 | 0.337 | 0.126 | 0.028  | -0.076 | 0.31           |
| Model 3  | 0.066 | 5.645  | 0.252 | 0.209 | 0.068 | -0.021 | 0.097  | 0.76           |
| Model 4  | 0.122 | 8.578  | 0.464 | 0.339 | 0.113 | -0.077 | 0.312  | 0.51           |
| Model 5* | 0.066 | 5.046  | 0.251 | 0.194 | 0.063 | -0.043 | 0.173  | 0.85           |
| Model 6  | 0.106 | 8.535  | 0.403 | 0.329 | 0.102 | 0.0424 | -0.151 | 0.57           |
| Model 7  | 0.154 | 12.459 | 0.585 | 0.480 | 0.146 | -0.092 | 0.359  | 0.39           |
| Model 8  | 0.174 | 12.200 | 0.658 | 0.447 | 0.184 | 0.0943 | -0.349 | 0.40           |
| Model 9  | 0.142 | 11.177 | 0.537 | 0.415 | 0.145 | 0.0010 | -0.012 | 0.53           |
| Model 10 | 0.125 | 10.043 | 0.474 | 0.374 | 0.130 | 0.0398 | -0.130 | 0.37           |

\* indicates the best modeling input combination

**Table 5** The statistical performance indicators for the testing period of DO variable prediction using HRSM and SVR models

| Models   | SI    | MAPE  | RMSE  | MAE   | RMSRE | MRE    | BIAS   | R <sup>2</sup> |
|----------|-------|-------|-------|-------|-------|--------|--------|----------------|
| HRSM     |       |       |       |       |       |        |        |                |
| Model 1  | 0.052 | 3.689 | 0.374 | 0.266 | 0.051 | −0.008 | 0.075  | 0.51           |
| Model 2* | 0.023 | 1.620 | 0.168 | 0.117 | 0.022 | −0.003 | 0.032  | 0.90           |
| Model 3  | 0.047 | 3.253 | 0.336 | 0.226 | 0.049 | 0.003  | −0.010 | 0.60           |
| Model 4  | 0.051 | 3.707 | 0.365 | 0.258 | 0.053 | 0.016  | −0.099 | 0.55           |
| Model 5  | 0.058 | 4.705 | 0.417 | 0.323 | 0.063 | 0.019  | −0.113 | 0.41           |
| Model 6  | 0.079 | 5.152 | 0.563 | 0.344 | 0.087 | 0.000  | 0.007  | 0.40           |
| Model 7  | 0.065 | 4.338 | 0.463 | 0.296 | 0.071 | 0.006  | −0.024 | 0.34           |
| Model 8  | 0.055 | 3.919 | 0.397 | 0.268 | 0.060 | 0.015  | −0.093 | 0.47           |
| Model 9  | 0.065 | 4.816 | 0.466 | 0.339 | 0.067 | −0.006 | 0.054  | 0.45           |
| Model 10 | 0.074 | 5.224 | 0.532 | 0.359 | 0.079 | −0.011 | 0.084  | 0.41           |
| SVR      |       |       |       |       |       |        |        |                |
| Model 1  | 0.050 | 3.896 | 0.358 | 0.278 | 0.049 | −0.007 | 0.066  | 0.54           |
| Model 2  | 0.045 | 3.595 | 0.322 | 0.256 | 0.045 | −0.002 | 0.030  | 0.62           |
| Model 3  | 0.043 | 3.608 | 0.310 | 0.255 | 0.043 | 0.002  | 0.000  | 0.65           |
| Model 4* | 0.031 | 2.688 | 0.223 | 0.188 | 0.032 | 0.009  | −0.056 | 0.83           |
| Model 5  | 0.063 | 4.698 | 0.452 | 0.334 | 0.063 | −0.005 | 0.055  | 0.40           |
| Model 6  | 0.070 | 4.995 | 0.502 | 0.339 | 0.077 | 0.040  | −0.266 | 0.37           |
| Model 7  | 0.062 | 4.754 | 0.446 | 0.339 | 0.061 | −0.014 | 0.118  | 0.40           |
| Model 8  | 0.066 | 5.495 | 0.472 | 0.390 | 0.066 | −0.008 | 0.067  | 0.46           |
| Model 9  | 0.062 | 4.678 | 0.444 | 0.318 | 0.068 | 0.020  | −0.126 | 0.37           |
| Model 10 | 0.073 | 5.425 | 0.525 | 0.388 | 0.072 | −0.009 | 0.083  | 0.30           |

\* indicates the best modeling input combination

$$MRE = \frac{\sum_{i=1}^n \left( \frac{x_i - y_i}{x_i} \right)}{n} \tag{16}$$

where  $x_i, y_i, \bar{x}$  and  $\bar{y}$  are observed, predicted mean value of observation, and mean value of predictions, respectively. The criteria perform in different ways; for instance, bias is a measure of the systematic tendency of a model to underestimate or overestimate the target values. A positive bias, for example, implies that observed values of BOD and DO, on average, are higher than that of predicted values and vice versa. The eight statistical metrics mentioned above can be used for the assessment of all forms of errors in model output as well as to assess the association and similarity of model output with observed data. Therefore, it is expected that the use of those eight criteria together would help in selection of the best model in an unbiased way.

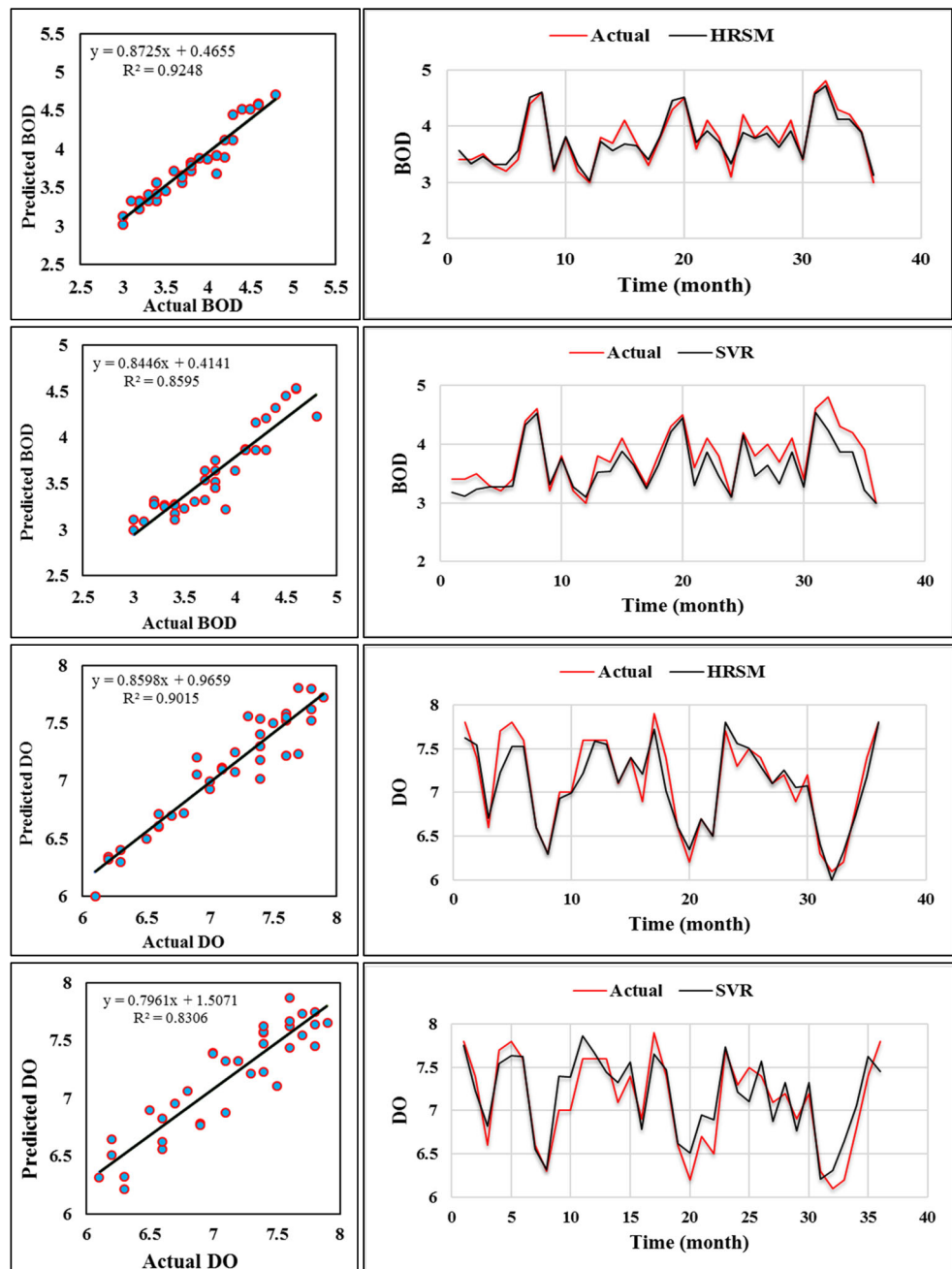
### Application results

Deterioration trend of water quality can be inspected via water quality prediction models. As described in the earlier parts, this

study mainly focused on the prediction of two important chemical parameters (i.e., BOD and DO). Both parameters have been classically used for decades as indicators of water quality, and undoubtedly accurate prediction, in this case, is essential to ease the protective initiatives. In this work, a new predictive HRSM model was introduced and the performance of new model is compared with very well-known AI model (e.g., SVR). HRSM is relatively a new approach that can predict complex patterns using the approximating tool. The superiority of the proposed model is checked by analyzing different forms of errors in model simulation.

Exploratory analysis of Euphrates River water quality parameters are given in Table 1. For a better understanding of the influence of each predictor on the targeted variables, correlation coefficients of each input variable with BOD and DO were computed (Table 2). It was found that the correlation coefficients of all water quality parameters except temperature were low and insignificant. A total of ten parameters were used to predict the BOD and DO. In this respect, ten different models were constructed with the combination of different input parameters, which were labeled as (M1, M2, M3, ..., M10). According to Table 3, model 1 (M1) consists of only one water quality parameter (i.e., temperature), M2 consists of two parameters, and likewise, M10 consists of all (ten) parameters to be fed as input attributes to the predictive models. As the number of input

**Fig. 3** Scatter plots and time series presentation for the actual and predictive models

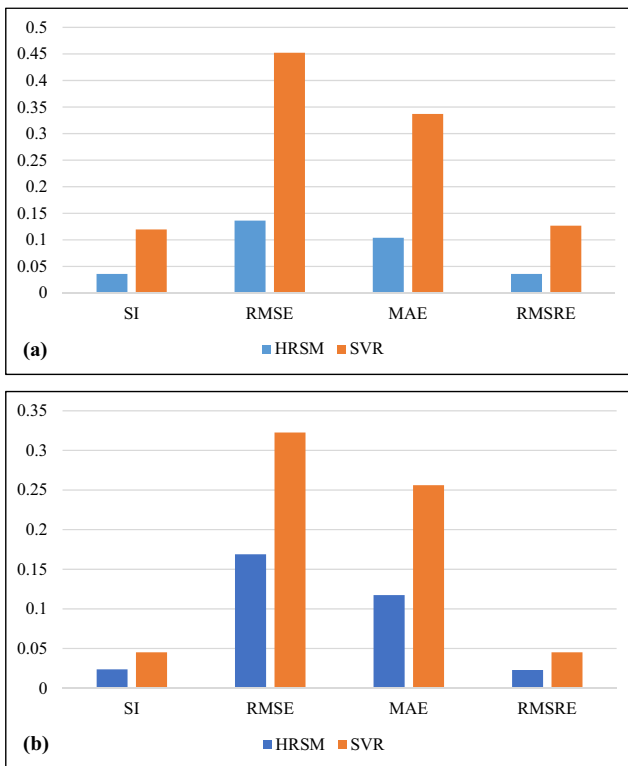


parameter gradually increased from M1 to M10, the changes in model performance provides the influence of each input parameter. Thus, ten models were constructed in this study in order to provide information regarding the sensitivity of the ten input parameters considered in this study in prediction of HRSM and SVR.

The model performance of HRSM and SVR is tabulated in Tables 4 and 5. According to the presented values, the HRSM was found to perform excellent in the prediction of both BOD and DO using the second input combination (M2) (temperature and turbidity). On the other hand, the benchmark model (SVR) attained the best results for fifth input combination

(M5) for prediction of BOD and fourth input combination (M4) for prediction of DO. This can be explained owing to the fact that mathematical models behave differently from one case to another following the explicitness of the internal mechanism between the predictors and the predictand. Figure 3 exhibits the model performance using the scatter plots and time series plots over the testing phase. The illustrated results of BOD belong to the best achievement of the HRSM for M2 and the best performance of SVR for M5. The HRSM prediction showed more accurate performance than the best SVR model. The highest correlation  $R^2$  was obtained using HRSM, 0.92 (M2), whereas it was 0.85 for SVR (M5). Figure 3 also shows





**Fig. 4** **a** Comparing errors in BOD prediction for M2. **b** Comparing errors in DO prediction for M2

the results for DO models. The highest values of correlation for HRSM and SVR DO models were ( $R^2 = 0.9$  (M2)) and ( $R^2 = 0.83$  (M4)), respectively.

In clearer appraisal for the various performance indicators, the results of the best input combinations are enlightened in Fig. 4. The SI, RMSE, MAE, and RMSRE were compared using a bar diagram in the figure. The results showed that the HRSM had significantly lesser error compared to the SVR for all cases. For instance, the scatter index (SI) of BOD prediction using HRSM and SVR was 0.035 and 0.119, respectively. The scatter index (SI) of DO prediction using HRSM and SVR was 0.023 and 0.031, respectively. It is evident that a remarkable augmentation was achieved using HRSM. Similarly, the RMSE, MAE, MAPE, and RMSRE indicators showed very promising results using HRSM for both the targeted variables.

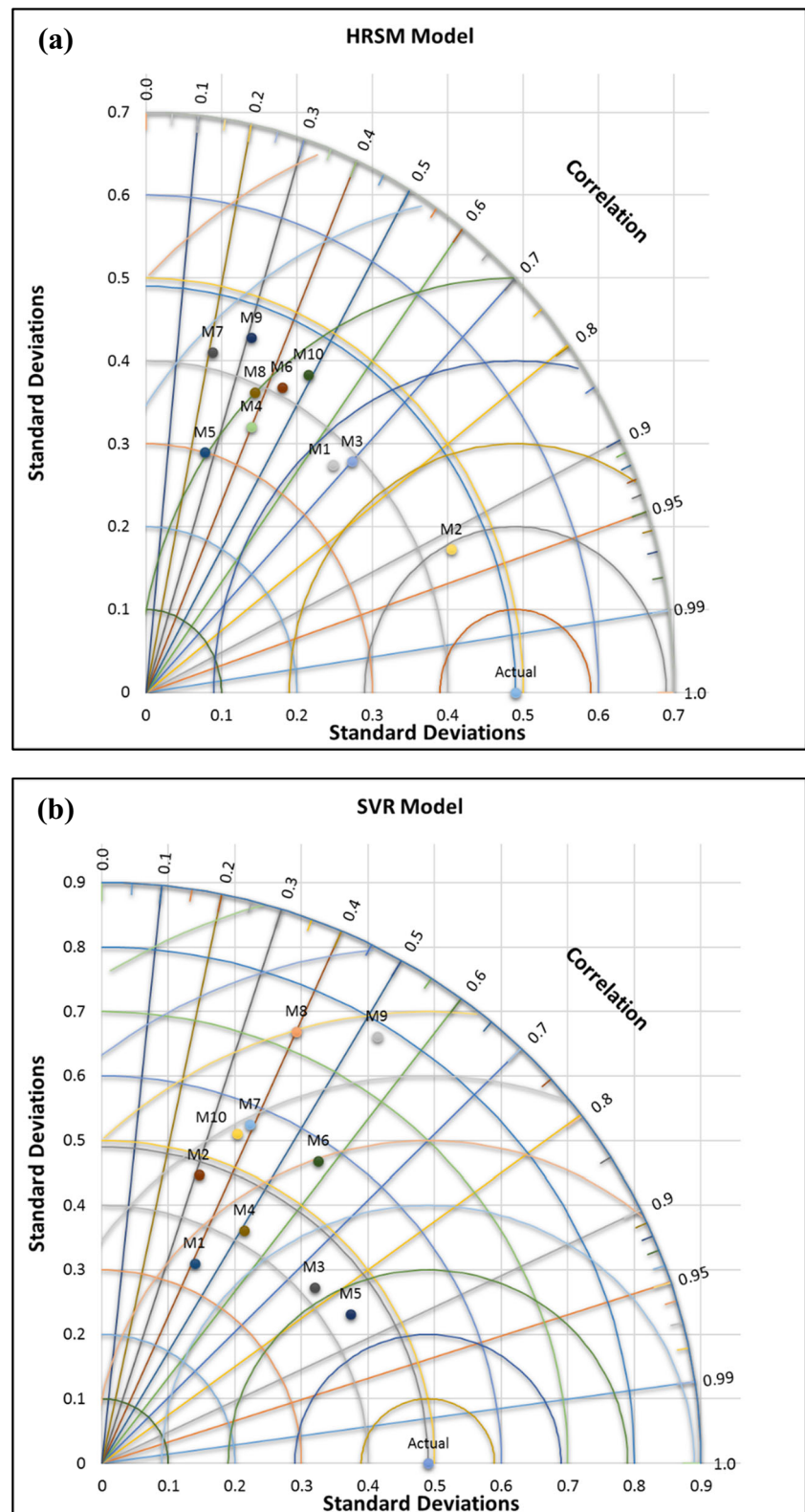
It was observed that HRSM showed better performance compared to SVR in predicting BOD and DO in most of the cases. It proves the robustness of the proposed model in comprehending the internal relationship between the predictors and predictand of the water quality parameters. The correlation coefficient achieved using HRSM was 0.9 (M2). On the other hand, SVR predicted DO with best input combination M4 with  $R^2$  value of 0.83. The details of the performance indicators of HRSM and SVR models in predicting DO are given in Tables 4 and 5, respectively. The results of other performance indicators such as SI, RMSE, MAE, and RMSRE for M2 are given in Fig. 4b.

In both BOD and DO prediction, it seems that consideration of more input variables is not always better for prediction. The prediction matrices demonstrated better prediction skill when fewer variables were used in constructing the predictive model. It was found that the HRSM performed better in the prediction of both BOD and DO when trained with only two parameters (i.e. temperature and turbidity). This observation matches well with the correlation values presented in Table 2 in which temperature and turbidity were found as the major attributes that affect the BOD and DO magnitudes. Indeed, the primary goal of a unique prediction model should be achieved closer approximation rather than including more parameters in the process. It is significantly important from the perspective of laboratory efforts. Also, this is highly valuable for the catchments that lack environmental information. The results indicate that it is important to focus on particular parameters only that have significant impact on prediction process and internal relations. Involving more parameters sometime may confuse the model and lead to inaccurate prediction or the astray. Here, the accuracy in the prediction of BOD and DO are prioritized than the number of parameter involvements. In this case study, M2 outperformed in almost every cases which consists of only two parameters (i.e., temperature and turbidity). Temperature was exclusively considered in the M1 model. Therefore, the result indicates that the turbidity is the key parameter that provides the best prediction. It should be noted that turbidity has a significantly high coefficient of variance (CV%) in Euphrates river along with TSS, which is also directly related to turbidity. When more parameters were included (i.e., M3, M4 and so on), both the models have to reform the relations among the parameters and get biased by the parameters which have less or no influence on BOD and DO variations.

The Taylor diagram is another way to compare model performance by visualizing the errors (Taylor 2001). These diagrams graphically summarize the model efficiency and are quite new in the presentation of water quality model performance. The Taylor diagrams of HRSM and SVR for both BOD and DO are given in Fig. 5a–d. The position of each model on the diagram measures the accuracy of the model in simulating BOD and DO compare to observed data. Figure 5a shows the highest correlation for HRSM with M2 combination. The simulated BOD that approximated well compared to actual data lies nearer to the point marked “actual” on the x-axis. The RMSE between the simulated and the observed BOD is represented in the diagram as proportional to the distance from the observed point. The standard deviation of the simulated BOD is proportional to the radial distance from the origin of the diagram. The standard deviation for all cases was found to vary between 0.3 and 0.5.

On the other hand, the predicted BOD by SVR model, as given in Fig. 4b indicates that SVR with M5 is the closest to

**Fig. 5** Taylor diagram graphical presentation for BOD variable over the testing phase using **a** HRSM and **b** SVR predictive models. Taylor diagram graphical presentation for DO variable over the testing phase using **c** HRSM and **d** SVR predictive models



the observed data as it has the largest correlation and lowest value of RMSE. The M2 which was the best model of HRSM showed correlation value slightly more than the SVR model. The models M3, M4, and M6 were found to fall in the mid-

range category. Also note that, although models M4 and M9 showed the same correlation, M4 approximated the amplitude of the variations (standard deviation) better than M9 and yielded a lower RMSE.

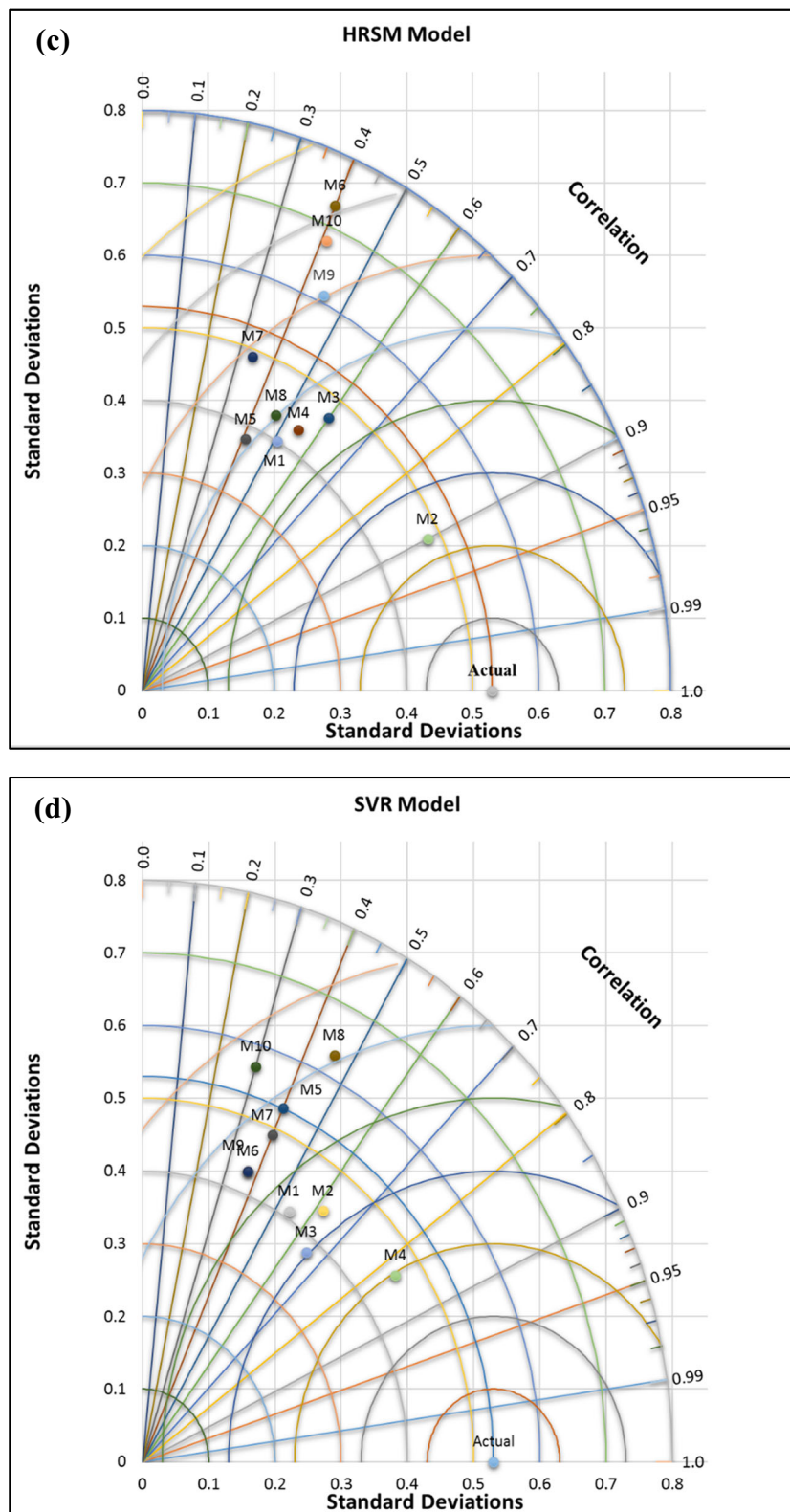


Fig. 5 (continued)

Figure 5c, d demonstrated the Taylor diagram for DO. Like BOD, M2 was found to perform exceptionally well among all the model input combinations by having the largest correlation

of 0.9. It also achieved the lowest standard deviation and RMSE. The M6, M10, and M5 correlated similarly with the observed values (0.4); yet, M5 showed better standard

deviation among the three models. Although the models M1, M3, and M4 performed moderately and had the correlation values between 0.5 and 0.6, M3 achieved the second position in approximating the DO using HRSM model. For SVR, it was found that the M4 performed very well as it showed the highest correlation (0.8) and the lowest RMSE. The M2 and M3 performed moderately well (correlation coefficients were 0.62 and 0.65, respectively) whereas M3 had slightly better standard deviation and RMSE in comparison to M2. Among all the models, M10 performed very poorly as it showed the lowest correlation (0.3) and the largest RMSE values.

From the above analysis and the description of the plotted Taylor diagrams, it can be argued that HRSM can simulate BOD and DO more accurately when trained with the temperature and the turbidity data of water. It is very clear, especially for this case study, the BOD and DO are regulated by the turbidity of the river. The SVR method came in the close prediction of BOD for M5 (correlation 0.85), but HRSM outperformed for M2 (correlation 0.92). In case of DO, again, HRSM was found to correlate better with the observed data with a correlation coefficient of 0.9 (M2) which is higher than SVR (0.83 for M4).

## Research findings discussion

RSM provides technique for mapping multi-dimensional pattern of responses of an outcome variable to the changes in controlling variables that govern physical processes of the system. The strength of the method lies in capturing accurate smooth approximations of responses through the selection of a set of polynomial functions which are able to capture the non-linearity in system behavior. The main advantage of RSM over other AI techniques is its ability to employ high-order polynomial functions for accurate approximation of responses (Keshtegar et al. 2016). Therefore, it has more explanatory potential compared to other AI-based regression analyses (Edwards 2007). The smooth nature of polynomial-based approximations eliminate numerical noises and allows efficient prediction of response variable (Yeniay 2014). The major improvement in the capability of RSM is obtained in this study by combining the polynomial and exponential functions to obtain a hybrid function which has increased the ability of RSM to approximate the highly nonlinear relationship and thus the predictive capability. This has made the HRSM model superior over the SVR model in simulating the BOD and DO from other water quality parameters.

The RSM method optimizes the response surface using the best set of predictors which are highly correlated with target variable considering that other less correlated variables can add uncertainty in model prediction. Uncertainty is one of the major limitations of using AI model predictions for water quality management. The AI models are usually data-driven,

and therefore, the major uncertainty in prediction arises from the uncertainty in inputs (Noori et al. 2013). The uncertainty of model inputs is propagated through the model towards the model outputs and acts as the major source of uncertainty in prediction (Beven 2006; Griensven and Meixner 2006). In order to reduce uncertainty, the RSM uses the best set of predictors to find a suitable approximation for the functional relationship between the targeted variables and the independent variables. In the present case study, the input parameters are gradually entered in the HRSM to identify the most appropriate input variables in order to reduce uncertainty in model prediction. The HRSM found only two easily measurable water quality parameters (temperature and turbidity) as most suitable for approximation of required functional relationship. Therefore, it can be remarked that the HRSM models developed in this study with only two input variables can be used for prediction of BOD and DO with less uncertainty.

The models developed in this study can be used for prediction of BOD and DO at any location of Euphrates River through simple measurement of river water temperature and turbidity. It is well known that DO in water body has inverse relation while BOD has direct relation with temperature and turbidity. DO decreases and BOD increases as temperature or turbidity increases. Therefore, identification of temperature and turbidity as the most suitable input by HRSM for prediction of BOD and BO is justifiable.

River water temperature is related to air temperature, and turbidity is related to rainfall and different properties of catchment such as land use and soil. In the future, models can be developed to predict river water temperature from air temperature and turbidity from catchment rainfall-runoff model which can be integrated with the HRSM models developed in this study for prediction of BOD and DO from rainfall and temperature data. Such models can also be used for assessment of the impacts of climate and land use changes on BOD and DO in Euphrates River and planning management strategies for mitigation of the impacts of environmental changes on river water quality.

## Conclusion

In this research, the capability of a new model, HRSM, in prediction of environmental variable has been inspected. The HRSM is developed in this study to predict BOD and DO in river water. The motivation of implementing this model was to provide a robust approach to determine water quality variables using historical laboratory observations. The models were developed using various physical and chemical quality variables of water as input attributes. A period of one-decade (2004–2013) laboratory information data was used to construct the models. The HRSM model was validated against a very well-known regression data-intelligence model known as SVR.

The obtained results of HRSM showed higher performance in comparison with SVR model. In addition, the proposed model demonstrated less approximation in terms of the input attributes that is extremely important for prediction of BOD and DO in catchments having less environmental or ecological information. Overall, the results revealed that HRSM can be used as a robust predictive model for Euphrates River water quality variables. Future research can be conducted for the improvement of the performance of prediction models through incorporation of more informative input attributes such as microbiological, hydrological, or even climatological variables. In addition, feasibility of natural inspired algorithms can be explored to select the appropriate casual information between the predictors and predictand as reported by Cho and Hermsmeier (2002) and Muttill and Chau (2007) in order to select appropriate input variables. Furthermore, recent water quality data can be used in which it can provide more informative attributes to the predictive model.

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

### References

- Acherjee B, Misra D, Bose D, Venkadeshwaran K (2009) Prediction of weld strength and seam width for laser transmission welding of thermoplastic using response surface methodology. *Opt Laser Technol* 41:956–967. <https://doi.org/10.1016/j.optlastec.2009.04.007>
- Afan HA, Keshtegar B, Mohtar WHMW, El-Shafie A (2017) Harmonize input selection for sediment transport prediction. *J Hydrol* 552:366–375
- 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*. <https://doi.org/10.1016/j.jksues.2015.02.001>
- Balabin RM, Lomakina EI, Safieva RZ (2011) Neural network (ANN) approach to biodiesel analysis: analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. *Fuel* 90:2007–2015. <https://doi.org/10.1016/j.fuel.2010.11.038>
- Barzegar R, Adamowski J, Moghaddam AA (2016) Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran. *Stoch Env Res Risk A* 30:1797–1819. <https://doi.org/10.1007/s00477-016-1213-y>
- Behmel S, Damour M, Ludwig R, Rodriguez MJ (2016) Water quality monitoring strategies—a review and future perspectives. *Sci Total Environ* 571:1312–1329. <https://doi.org/10.1016/j.scitotenv.2016.06.235>
- Beven K (2006) On undermining the science? *Hydrol Process* 20:3141–3146. <https://doi.org/10.1002/hyp.6396>
- Bouamar M, Ladjal M (2008) A comparative study of RBF neural network and SVM classification techniques performed on real data for drinking water quality. 2008 5th International MultiConference on Systems Signals and Devices 1–5. <https://doi.org/10.1109/SSD.2008.4632856>
- Ch S, Sohani SK, Kumar D, Malik A, Chahar BR, Nema AK, Panigrahi BK, Dhiman RC (2014) A support vector machine-firefly algorithm based forecasting model to determine malaria transmission. *Neurocomputing* 129:279–288. <https://doi.org/10.1016/j.neucom.2013.09.030>
- Chau KW (2005) Characterization of transboundary POP contamination in aquatic ecosystems of Pearl River delta. *Mar Pollut Bull* 51:960–965
- Chen XY, Chau KW (2016) A hybrid double feedforward neural network for suspended sediment load estimation. *Water Resour Manag* 30: 2179–2194. <https://doi.org/10.1007/s11269-016-1281-2>
- Cho T (2007) Prediction of cyclic freeze-thaw damage in concrete structures based on response surface method. *Constr Build Mater* 21: 2031–2040. <https://doi.org/10.1016/j.conbuildmat.2007.04.018>
- Cho SJ, Hermsmeier MA (2002) Genetic algorithm guided selection: variable selection and subset selection. *J Chem Inf Comput Sci* 42: 927–936. <https://doi.org/10.1021/ci010247v>
- Deng W, Wang G, Zhang X (2015) A novel hybrid water quality time series prediction method based on cloud model and fuzzy forecasting. *Chemom Intell Lab Syst* 149:39–49. <https://doi.org/10.1016/j.chemolab.2015.09.017>
- Ebtehaj I, Bonakdari H, Shamshirband S, Ismail Z, Hashim R (2017) New approach to estimate velocity at limit of deposition in storm sewers using vector machine coupled with firefly algorithm. *J Pipeline Syst Eng Pract* 8:04016018. [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000252](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000252)
- Edwards JR (2007) Polynomial regression and response surface methodology. *Perspectives on organizational fit*:361–372
- 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. <https://doi.org/10.1007/s13762-013-0378-x>
- Fahimi F, Yaseen ZM, El-shafie A (2016) Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. *Theor Appl Climatol* 128:1–29. <https://doi.org/10.1007/s00704-016-1735-8>
- Fengxiang R, Xuanchi Z, Chao Z (2010) Application of rough set-SVM model in the evaluation on water quality. In: *Proceedings - 2010 International Conference on Communications and Intelligence Information Security, ICCIIS 2010*. pp 156–159
- 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* 64:2409–2420. <https://doi.org/10.1016/j.marpolbul.2012.08.005>
- Ghorbani MA, Deo RC, Karimi V, et al (2017a) Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey. *Stochastic Environmental Research and Risk Assessment* 1–15
- Ghorbani MA, Deo RC, Yaseen ZM, H. Kashani M, Mohammadi B (2017b) Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. *Theor Appl Climatol* 133:1119–1131. <https://doi.org/10.1007/s00704-017-2244-0>
- Ghorbani MA, Deo RC, Yaseen ZM, Kashani MH (2017c) Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. <https://doi.org/10.1007/s00704-017-2244-0>
- Griensven A, Meixner T (2006) Methods to quantify and identify the sources of uncertainty for river basin water quality models. *Water Sci Technol* 53:51–59. <https://doi.org/10.2166/wst.2006.007>
- Gümrah F, Öz B, Güler B, Evin S (2000) The application of artificial neural networks for the prediction of water quality of polluted aquifer. *Water Air Soil Pollut* 119:275–294. <https://doi.org/10.1023/A:1005165315197>

- Jadhav MS, Khare KC, Warke AS (2015) Water quality prediction of Gangapur reservoir (India) using LS-SVM and genetic programming. *Lakes Reserv Res Manag* 20:275–284. <https://doi.org/10.1111/lre.12113>
- Jian C, Hongsheng H, Suxiang Q, Gongbiao Y (2010) Research on the water quality forecast method Based on SVM. In: ICMIT 2009: MECHATRONICS AND INFORMATION TECHNOLOGY
- Kadhem AJ (2013) Assessment of water quality in Tigris River-Iraq by using GIS mapping. *Nat Resour* 04:441–448. <https://doi.org/10.4236/nr.2013.46054>
- Keshtegar B, Heddami S (2017) Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: a comparative study. *Neural Comput & Applic*:1–12
- Keshtegar B, Allawi MF, Afan HA, El-Shafie A (2016) Optimized river stream-flow forecasting model utilizing high-order response surface method. *Water Resour Manag* 30:3899–3914. <https://doi.org/10.1007/s11269-016-1397-4>
- Keshtegar B, Kisi O, Asce M (2017) Modified response-surface method: new approach for modeling pan evaporation. 22:1–14. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001541](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001541)
- Kewlani G, Iagnemma K (2008) A stochastic response surface approach to statistical prediction of robotic mobility. *IEEE International Conference on Intelligent Robots and Systems (IROS)* 22–26. <https://doi.org/10.1109/IROS.2008.4651187>
- 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* 405:277–287. <https://doi.org/10.1016/j.jhydrol.2011.05.024>
- Kim C, Choi KK (2008) Reliability-based design optimization using response surface method with prediction interval estimation. *J Mech Des* 130:121401. <https://doi.org/10.1115/1.2988476>
- Klaslan Y, Tuna G, Gezer G et al (2014) ANN-based estimation of groundwater quality using a wireless water quality network. *Int J Distribut Sensor Netw* 10:458329. <https://doi.org/10.1155/2014/458329>
- Liu M, Lu J (2014) Support vector machine—an alternative to artificial neuron network for water quality forecasting in an agricultural non-point source polluted river? *Environ Sci Pollut Res* 21:11036–11053. <https://doi.org/10.1007/s11356-014-3046-x>
- Liu S, Tai H, Ding Q, Li D, Xu L, Wei Y (2013) A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. *Math Comput Model* 58:458–465. <https://doi.org/10.1016/j.mcm.2011.11.021>
- May RJ (2008) Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. *Environ Model Softw* 23:1289–1299. <https://doi.org/10.1016/j.envsoft.2008.03.008>
- Moeni H, Bonakdari H, Fatemi SE (2017) Stochastic model stationarization by eliminating the periodic term and its effect on time series prediction. *J Hydrol* 547:348–364. <https://doi.org/10.1016/j.jhydrol.2017.02.012>
- Moghaddam TB, Soltani M, Shahraki HS, Shamshirband S, Noor NBM, Karim MR (2016) The use of SVM-FFA in estimating fatigue life of polyethylene terephthalate modified asphalt mixtures. *Measurement* 90:526–533. <https://doi.org/10.1016/j.measurement.2016.05.004>
- Muttill N, Chau K-W (2006) Neural network and genetic programming for modelling coastal algal blooms. *Int J Environ Pollut* 28:223–238. <https://doi.org/10.1504/IJEP.2006.011208>
- Muttill N, Chau KW (2007) Machine-learning paradigms for selecting ecologically significant input variables. *Eng Appl Artif Intell* 20:735–744. <https://doi.org/10.1016/j.engappai.2006.11.016>
- Najah A, El-Shafie A, Karim OA, El-Shafie AH (2014) Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring. *Environ Sci Pollut Res* 21:1658–1670. <https://doi.org/10.1007/s11356-013-2048-4>
- Noori R, Karbassi A, Ashrafi K et al (2013) Development and application of reduced-order neural network model based on proper orthogonal decomposition for BOD5 monitoring in river systems: uncertainty analysis. *Environ Prog Sustain Energy* 32:120–127
- Olyaie E, Banejad H, Chau KW, Melesse AM (2015) Erratum to a comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States. *Environ Monit Assess* 187(4):4381. <https://doi.org/10.1007/s10661-015-4381-1>
- Olyaie E, Zare Abyaneh H, Danandeh Mehr A (2017) A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. *Geosci Front* 8:517–527. <https://doi.org/10.1016/j.gsf.2016.04.007>
- Orouji H, Haddad O (2013) Modeling of water quality parameters using data-driven models. *J Environ Eng* 139:947–957. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0000706](https://doi.org/10.1061/(ASCE)EE.1943-7870.0000706)
- Page RM, Besmer MD, Epting J, Sigrist JA, Hammes F, Huggenberger P (2017) Online analysis: deeper insights into water quality dynamics in spring water. *Sci Total Environ* 599–600:227–236. <https://doi.org/10.1016/j.scitotenv.2017.04.204>
- Palani S, Liong SY, Tklich P (2008) An ANN application for water quality forecasting. *Mar Pollut Bull* 56:1586–1597. <https://doi.org/10.1016/j.marpolbul.2008.05.021>
- Rahi KA, Halihan T (2010) Changes in the salinity of the Euphrates River system in Iraq. *Reg Environ Chang* 10:27–35. <https://doi.org/10.1007/s10113-009-0083-y>
- Ravansalar M, Rajaei T, Zounemat-Kermani M (2016) A wavelet-linear genetic programming model for sodium (Na<sup>+</sup>) concentration forecasting in rivers. *J Hydrol* 537:398–407. <https://doi.org/10.1016/j.jhydrol.2016.03.062>
- Roussouly N, Petitjean F, Salaun M (2012) A new adaptive response surface method for reliability analysis. *Probabilistic Engineering Mechanics*
- Sahu M, Mahapatra SS, Sahu HB, Patel RK (2011) Prediction of water quality index using neuro fuzzy inference system. *Water Qual Expo Health* 3:175–191. <https://doi.org/10.1007/s12403-011-0054-7>
- Shamshirband S, Mohammadi K, Tong CW, Zamani M, Motamedi S, Ch S (2016) A hybrid SVM-FFA method for prediction of monthly mean global solar radiation. *Theor Appl Climatol* 125:53–65. <https://doi.org/10.1007/s00704-015-1482-2>
- Shi B, Wang P, Jiang J, Liu R (2018) Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomalies. *Sci Total Environ* 610–611:1390–1399. <https://doi.org/10.1016/j.scitotenv.2017.08.232>
- Singh KP, Basant A, Malik A, Jain G (2016) Artificial neural network modeling of the river water quality—a case study. *Ecol Model* 220:888–895. <https://doi.org/10.1016/j.ecolmodel.2009.01.004>
- Singh KP, Basant N, Gupta S (2011) Support vector machines in water quality management. *Anal Chim Acta* 703:152–162. <https://doi.org/10.1016/j.aca.2011.07.027>
- Song S, Zheng X, Li F (2010) Surface water quality forecasting based on ANN and GIS for the Chanzhi reservoir, China. In: 2nd International Conference on Information Science and Engineering, ICISE2010 - Proceedings. pp 4094–4097
- Sreekanth J, Datta B (2010) Multi-objective management of saltwater intrusion in coastal aquifers using genetic programming and modular neural network based surrogate models. *J Hydrol* 393:245–256. <https://doi.org/10.1016/j.jhydrol.2010.08.023>
- Sudheer KP, Chaubey I, Garg V (2006) Lake water quality assessment from landsat thematic mapper data using neural network: an approach to optimal band combination selection. *J Am Water Resour Assoc* 42:1683–1695. <https://doi.org/10.1111/j.1752-1688.2006.tb06029.x>
- Tao H, Diop L, Bodian A, et al (2018) Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: regional case study in Burkina Faso

- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106:7183–7192. <https://doi.org/10.1029/2000JD900719>
- Vapnik V (1995) *The Nature of statistical Learning Theory*
- Wang Z, Huang K, Zhou PJ, Guo HC (2010) A hybrid neural network model for cyanobacteria bloom in Dianchi Lake. *Procedia Environ Sci* 2:67–75. <https://doi.org/10.1016/j.proenv.2010.10.010>
- Wang W, Xu D, Chau K (2014) Assessment of river water quality based on theory of variable fuzzy sets and fuzzy binary comparison method. *Water Resour Manag* 28:4183–4200. <https://doi.org/10.1007/s11269-014-0738-4>
- Wei D, Cui Z, Chen J (2008) Optimization and tolerance prediction of sheet metal forming process using response surface model. *Comput Mater Sci* 42:228–233. <https://doi.org/10.1016/j.commatsci.2007.07.014>
- Wu W, Dandy GC, Maier HR (2014) Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modeling. *Environ Model Softw* 54:108–127. <https://doi.org/10.1016/j.envsoft.2013.12.016>
- Wu Z, Wang X, Chen Y, Cai Y, Deng J (2018) Assessing river water quality using water quality index in Lake Taihu Basin, China. *Sci Total Environ* 612:914–922. <https://doi.org/10.1016/j.scitotenv.2017.08.293>
- Xu L, Wang J, Guan J, Huang F (2007) A support vector machine model for mapping of lake water quality from remote-sensed images. *IC- Med* 1:57–66. <https://doi.org/10.1080/1931308X.2007.10644137>
- Yang XS (2010) Firefly algorithm, stochastic test functions and design optimisation. arXiv preprint arXiv:1003–1409
- Yeniay O (2014) Comparative study of algorithms for response surface optimization. *Math Comput Appl* 19(1):93–104
- Yunrong X, Liangzhong J (2009a) Water quality prediction using LS-SVM with particle swarm optimization. In: WKDD: 2009 SECOND INTERNATIONAL WORKSHOP ON KNOWLEDGE DISCOVERY AND DATA MINING, PROCEEDINGS. pp 900–904
- Yunrong XYX, Liangzhong JLJ (2009b) Water quality prediction using LS-SVM and particle swarm optimization. 2009 Second International Workshop on Knowledge Discovery and Data Mining 900–904. <https://doi.org/10.1109/WKDD.2009.217>
- Zhang R, Qian X, Li H, Yuan X, Ye R (2012) Selection of optimal river water quality improvement programs using QUAL2K: a case study of Taihu Lake Basin, China. *Sci Total Environ* 431:278–285. <https://doi.org/10.1016/j.scitotenv.2012.05.063>
- Zolnikov TR (2013) The maladies of water and war: addressing poor water quality in Iraq. *Am J Public Health* 103:980–987. <https://doi.org/10.2105/AJPH.2012.301118>
- Zou R, Lung W-S, Wu J (2007) An adaptive neural network embedded genetic algorithm approach for inverse water quality modeling. *Water Resour Res* 43:1–13. <https://doi.org/10.1029/2006WR005158>