Title

Simulating thermal stratification and modeling outlet water temperature in reservoirs with a data mining method

Journal

AQUA - Water Infrastructure Ecosystems and Society, 68(1)

ISSN

2709-8028

Authors

Soleimani, Shima

Bozorg-Haddad, Omid

Saadatpour, Motahareh

et al.

Publication Date

2019-02-01

DOI

10.2166/aqua.2018.036

Peer reviewed

Simulating thermal stratification and modeling outlet water temperature in reservoirs with a data-mining method

Shima Soleimani, Omid Bozorg-Haddad, Motahareh Saadatpour and Hugo A. Loáiciga

ABSTRACT

This paper simulates the thermal stratification of the Karkhe Reservoir, Iran, with the CE-QUAL-W2 model for the period 1981–1995. The simulation of reservoir water quality requires meteorological, hydrological, chemical, and discharge time series to accurately predict the temperature of water releases from the reservoir. Outlet water temperature of the Karkhe Reservoir is calculated using the CE-QUAL-W2 model and the simulated outlet water temperature is thereafter modeled with the library for support vector machines (LIBSVM) data-mining model. Simulation results show thermal stratification in the Karkhe Reservoir occurs once a year. In addition, the data-mining model is a good surrogate model for the CE-QUAL-W2 model for estimating water temperature at different outlet levels in the reservoir. The root-mean square, mean absolute error and Nash-Sutcliffe criteria are used to assess the performance of the data-mining method. The LIBSVM model was found to be a suitable surrogate model for the main simulation model, and can be linked to optimization models with which to calculate reservoir operational rules for thermal control.

Key words | data-mining model, support vector machine, SVM, thermal stratification, two-dimensional simulation model, water temperature

Shima Soleimani
Omid Bozorg-Haddad (corresponding author)
Department of Irrigation and Reclamation, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, University of Tehran,
Karaj, Tehran 31587-77871,
Iran
E-mail: *obhaddad@ut.ac.ir*

Motahareh Saadatpour
School of Civil Engineering,
Iran University of Science and Technology,
16846-13114 Tehran,
Iran

Hugo A. Loáiciga
Department of Geography,
University of California,
Santa Barbara, California 93106,
USA

INTRODUCTION

The thermal stratification of lakes and reservoirs implies a change in the temperature at different depths in the lake or reservoir, and occurs because of the change in density with temperature. Thermal stratification develops two layers called epilimnion and hypolimnion separated by a layer of rapid temperature changes called metalimnion. The hypolimnion layer consists of water that is generally denser and colder than water in the epilimnion layer. Thermal stratification and heat budget significantly affect the water quality and ecological characteristic of lakes and reservoirs (Wetzel 1983; Wang *et al.* 2012), contaminant transport and hydrodynamic mixing in reservoir and lakes (Fischer *et al.* 1979; Kennedy *et al.* 1982), downstream irrigation area of reservoirs (Yang *et al.* 2012), and aquatic environment of lakes and reservoirs (Hanna *et al.* 1999). For instance, hypolimnion layers in stratified reservoirs contain anoxic water with poor quality which might contain dissolved iron, manganese, sulfide, ammonium, and phosphate (Dortch 1997). Rising water temperature in water reservoirs increases the speed of chemical and biological reactions, which have a strong influence on reservoir nutrient cycling and initial productivity (Sahoo & Schladow 2008).

Elci (2008) considered the mixing and thermal stratification effects in the Tahtali reservoir in Turkey using

doi: 10.2166/aqua.2018.036

multivariate analyses on qualitative data of the reservoir. Wei *et al.* (2011) developed a three-dimensional hydrodynamic simulation model and considered the effects of reservoir inflow temperature on thermal structure in a reservoir located in south China. Yang *et al.* (2012) simulated the thermal stratification effects on downstream irrigation area of the Xiahushan reservoir in China with a three-dimensional hydrodynamic model, environmental fluid dynamics code (EFDC). Wang *et al.* (2012) used a three-dimensional hydrodynamic simulation model (ELCOM) to consider the effects of local climate and hydrologic conditions on the thermal stratification in the Liuxihe Reservoir, China. Kerimoglu & Rinke (2013) estimated the response of the Bautzen reservoir, Germany, to different combinations of external factors such as the hydrological regime, water level fluctuation, dewatering depth and meteorological variables with the one-dimensional DYRESM model. Lugg & Copeland (2014) assessed the impacts of the location, quantity, and extent of cold-water pollution (CWP) in the downstream Murray-Darling basin in Australia. Bermudez *et al.* (2018) considered the impacts of a pumped-storage hydroelectric power plant on the reservoirs' thermal stratification. Gelda & Effler (2007) linked the CE-QUAL-W2 simulation model to an evolutionary optimization algorithm and showed that using selective withdrawal can decrease the epilimnion and metalimnion in the reservoir. All the models applied in the cited studies were solved based on differential equations and numerical methods called physically based models.

Data-mining methods can be applied as surrogates of physically based models to reduce the computational burden. Data mining is a process for selecting, identifying, and modeling based on large databases searched to discover relations among data providing useful results for database analysts (Giudici 2003). There are many data-mining methods such as linear and nonlinear regression, Kriging method, artificial neural network (*ANN*), genetic programming (GP), multilayer perceptron (MLP), and support vector machine (SVM). Xiang & Jiang (2009) applied least-squares support vector regression (LSSVR) method to predict water quality in the Liuxi River located in China. They predicted chemical oxygen demand (COD) and DO with an integrated algorithm consisting of the LSSVR method and the particle swarm optimization (PSO) algorithm. Raghavendra & Deka (2014) reviewed the application of support vector regression (SVR) in hydrology.

There have not been, to our knowledge, reported applications of the SVM method to modeling outlet water temperature in reservoirs even in recent investigations. This study simulates outlet temperature in a reservoir with the combined application of water-quality simulation with the CE-QUAL-W2 model and data mining with the LIBSVM, a variant of the SVM method. LIBSVM serves as the data-mining method for its capacity to model complex and non-linear water-quality dynamics present in the physically based model CE-QUAL-W2. The LIBSVM yields substantial reductions in the computational burden of water-temperature modeling while preserving predictive accuracy.

MATERIALS AND METHODS

The CE-QUAL-W2 model and the SVM method are briefly summarized in this section. The statistical criteria used to evaluate the LIBSVM method's performance are also defined in this section.

CE-QUAL-W2 model

A physically based hydrologic model consists of a mathematical description of surface and subsurface processes, external and internal boundary conditions, and initial conditions (Furman 2008). Physically based models have been applied in numerous studies to address a wide range of water-quality questions (McCuen 1973; Rango & Martinec 1995; Montanari & Grossi 2008; Mendoza *et al.* 2015). The CE-QUAL-W2 model is a two-dimensional water quality and hydrodynamic simulation model developed by the U.S. Army Corps of Engineers' Waterways Experiment Station (WES). The temporal and spatial changes of water surface level and water temperature are modeled using CE-QUAL-W2 version 3.71 (Cole & Wells 2008). The required model data consist of reservoir geometric data, initial temperature conditions, boundary conditions (time series of meteorological, hydrological, water quality, released reservoir discharge, and hydraulic and kinetic parameters). The reasons for choosing the CE-QUAL-W2

version 3.71 are: (1) the model simulates flow, water level, horizontal and vertical velocities, water temperature, ice cover, and the concentrations of constituents such as ammonia, nitrate, phosphate; (2) in case of thermally stratified reservoirs the model allows reservoir operators to define water release from multiple floating or fixed elevational outlets that access waters at several depths with distinct temperature; and (3) it allows the operators to set priority allocations for each outlet to optimize water releases.

Support vector machine

SVM is a machine-learning system based on constrained optimization theory (Vapnik *et al.* 1996; Vapnik 1998). The SVM method has been applied in many fields of inquiry successfully. SVM regression determines a relational function between dependent or output variables (y) and independent or input variables (x). The relational function, $f(x)$, is obtained by the SVM model trained on a data set.

Structural risk minimization principle

The structural risk minimization principle guides the SVM search for an optimal function, $f(x)$, by minimizing the norm of differences, $L(y)$ and $f(x)$ between predicted values (y, the output or dependent variables) and observed ones (x, the input or independent variables). The minimization relies on the risk function R, defined as follows:

$$R = \int L(y, f(x) dP(x, y) \tag{1}$$

in which R is risk function, and $P(x,y)$ is the probability distribution function, which is unknown. Therefore, the expected risk function R written below is used instead:

$$R_{emp} = \frac{1}{l}\sum_{i=1}^{l} L(y, f(x)) \tag{2}$$

where R_{emp} is the empirical risk and l is the number of training data. Vapnik (1998) proposed a structural risk minimization inductive principle to minimize R_{emp} in the SVM method, which is explained below.

Application of the SVM method in function estimation

Vapnik (Vapnik *et al.* 1996; Vapnik 1998) proposed a risk function called *ε-insensitive* to solve regression problems. The function is shown in Equation (3):

$$L(y, f(x)) = |y - f(x)|_{\varepsilon} = \begin{cases} 0 & if |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon & \text{otherwise} \end{cases} \tag{3}$$

This risk function ignores risks that are smaller than ε and considers risk values according to $\xi = |y - f(x)| - \varepsilon$ or data in which the difference between observed values and estimated values exceeds the threshold ε. Bold font expresses vector values in the adopted notation.

The SVM method assumes the relation between input and output variables is nonlinear. A space called feature space is developed by a nonlinear mapping function that linearizes the relation between variables. The linear relation between input and output variables in the feature space is defined by Equation (4):

$$f(x) = \langle w, \phi(x) \rangle + b \tag{4}$$

in which x and $f(x)$ are input and output of training data, respectively, w is the weighting vector, $\phi(x)$ is a nonlinear function that maps the data from main space to feature space, and b is a bias factor. The norm $\langle ., . \rangle$ denotes the inner product of vectors.

The structural risk minimization is written in terms of the optimization problem (5):

$$\text{Minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$
$$\text{subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{5}$$

in which $\|w\|^2$ is the Euclidean norm value, i is the input variable counter, ξ and ξ^* represent the penalties applied to the objective function considering the value of ε. C is a constant parameter that determines the $\|w\|^2$ value in terms of the complexity of the risk function.

The procedure for solving problem (5) relies on the Lagrange form of the objective function, which is shown in Equation (6):

$$L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \sum_{i-1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*) - \sum_{i=1}^{l}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i\rangle + b) - \sum_{i=1}^{l}\alpha_i^*(\varepsilon + \xi_i^* - y_i + \langle w, x_i\rangle + b) \quad \text{in which} \quad \eta^{(*)}, \alpha^{(*)} \geq 0 \tag{6}$$

in which L is Lagrange function, $\eta_i^{(*)}$ and $\alpha_i^{(*)}$ are the Lagrange multipliers for the ith input ($\eta_i^{(*)}$ represents η_i or η_i^* (with or without star), analogous notational symbolism is used for $\alpha_i^{(*)}$). Partial derivatives of L are taken with respect to w, b, ξ, and ξ^*, the resulting expressions are set to equal to zero and solved for the unknown Lagrange parameters:

$$\partial_w L = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0 \tag{7}$$

$$\partial_b L = w - \sum_{i=1}^{i}(\alpha_i - \alpha_i^*)x_i = 0 \tag{8}$$

$$\partial_{\xi^{(*)}} L = C - \alpha_i^{(*)} - \eta^{(*)} = 0 \tag{9}$$

There is lack of information about the choice of nonlinear mapping function $\phi(\boldsymbol{x})$. Accordingly, a kernel function is defined that expresses the feature space linearly mapped from the main space. A kernel function is given by Equation (10):

$$K(x_i, x) = \langle \phi(x_i)', \phi(x)\rangle \qquad x \in X \tag{10}$$

where K is a kernel function. The kernel function in Equation (10) is used to construct the relational function:

$$f(x) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)K(x_i, x) + b \tag{11}$$

The kernel function can be selected from linear, polynomial, sigmoid, and radial basis (RBF), and MLP functions. This study applies the RBF kernel, which is given by:

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) \tag{12}$$

in which σ is the RBF kernel function coefficient. Figure 1 depicts the SVM method's flowchart. The LIBSVM model is a variant of the SVM method. It is implemented in this study as a data-mining model (Chang & Lin 2011).

Determination of the LIBSVM model parameters

The LIBSVM model parameters are ε, which expresses the *ε-insensitive* risk function, C (expresses functional capability), and the kernel function coefficient σ that appears in the RBF kernel function, Equation (12). These parameters can be obtained by sensitivity analysis resorting to the trial and error method which is commonly applied based on prior knowledge of the analyst (Vapnik 1998).

Evaluation of the LIBSVM model performance

The skill of the LIBSVM model to predict the outlet water temperature in reservoirs with variable level is herein measured with the root mean square (*RMSE*), the mean absolute error (*MAE*), and the Nash-Sutcliffe coefficient (*NS*). The criteria are defined in Equations (13) to (15)

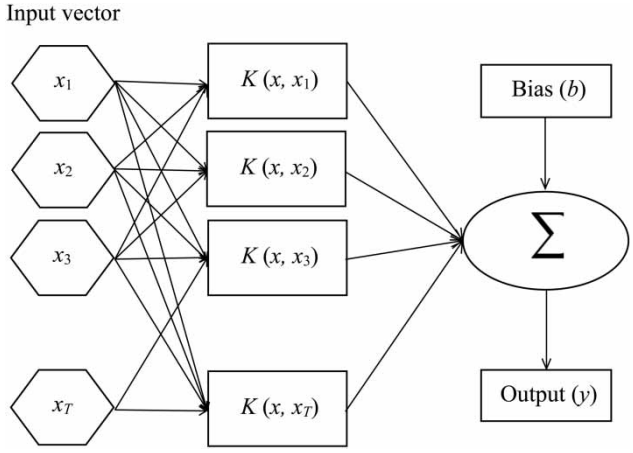

Figure 1 | The SVM method's flowchart.

(Wang *et al.* 2009; Orouji *et al.* 2013):

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t^{CE} - y_t^{LIB})^2} \quad (13)$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t^{CE} - y_t^{LIB}| \quad (14)$$

$$NS = 1 - \frac{\sum_{t=1}^{T}(y_t^{CB} - y_t^{LIB})^2}{\sum_{t=1}^{T}(\bar{y}_t^{CE} - \bar{y}_t^{LIB})^2} \quad (15)$$

in which y_t^{CE} and y_t^{LIB} are outlet water temperature calculated with the CE-QUAL-W2 and LIBSVM models, respectively. $\bar{y}_t^{CE}$ and $\bar{y}_t^{LIB}$ are the average outlet water temperature from the CE-QUAL-W2 and LIBSVM models, respectively, T is the total number of time steps, and t is the time step counter. The flowchart of the SVM approach herein developed is shown in Figure 2.

CASE STUDY

The Karkhe Reservoir is the sixth largest earthen reservoir in the world and the largest earthen reservoir in Iran. The catchment area of the Karkhe River is approximately about 44,000 km^2. It is located between 46°57′–49°10′ eastern longitudes and 31°48′–34°58′ northern latitudes. The reservoir volume equals 5×10^9 m^3, its length equals 64 km, and its surface area equals 162 km^2 at normal water pool (220 m above sea level). The average and maximum depth of the reservoir are 61.8 and 117 m, respectively, and the maximum and minimum elevation of the reservoir are 230 and 113 m above sea level, respectively. The reservoir has three outlets. One is a longitudinal

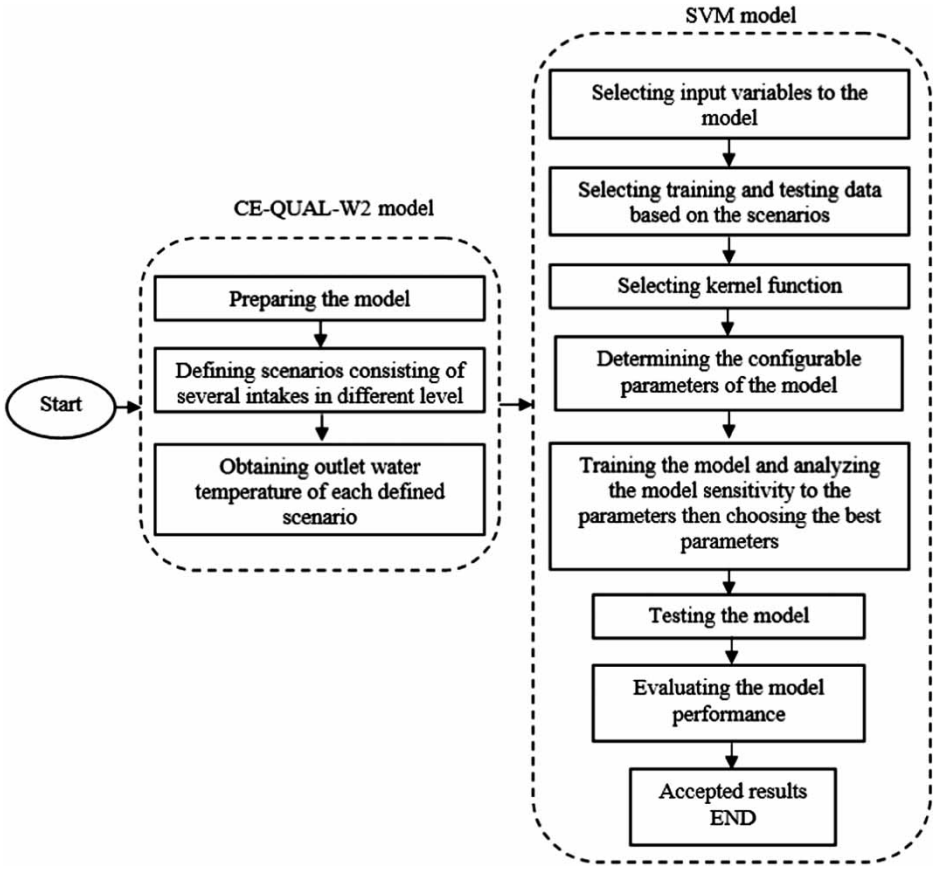

Figure 2 | The methodology's flowchart.

outlet whose outlet water temperature is not considered in this study, one agricultural outlet and one hydropower outlet located at 163 and 181 m above sea level, respectively. The reservoir supplies water for 180,000 hectares of agricultural land and domestic water for cities located in Khozestan province. Therefore, the reservoir is strategic for supplying reliable and high-quality water for irrigation and municipal consumption. According to field observations the reservoir is affected by eutrophication and is susceptible to thermal stratification. In recent years, the reservoir has exhibited algal blooms and water quality degradation. The reservoir average retention time is estimated about 0.74 years. The relative long retention time causes increasing nutrient load and water quality degradation, which call for detailed modeling, monitoring, and assessment of thermal stratification formation in the reservoir (Afshar & Saadatpour 2009; Saadatpour *et al.* 2017).

CE-QUAL-W2 set up

The set up of the CE-QUAL-W2 model involves: (1) specifying reservoir geometry; (2) adding component structures in the reservoir; (3) adding meteorological and hydrological data; (4) defining the simulation time period and time step; and (5) calibrating and verifying the model.

In previous research, the Karkhe Reservoir was described with 66 longitudinal segments each 1,000 m long (Figure 3), and up to 55 vertical layers depending on water depth, layers' thicknesses ranging from 2 to 5 m in each segment (Figure 4) (Afshar & Saadatpour 2009; Saadatpour *et al.* 2017). The same description is applied in this study. Reservoir components include a reservoir spillway, one longitudinal outlet, one agricultural oulet, and one hydropower outlet located at 163 and 181 m above sea level; two proposed outlets located at 120 and 140 m above sea level used to simulate water releases blending waters from different levels. Figure 4 specifies the reservoir's minimum and maximum elevations and the outlet elevations. Meteorological, hydrological, and water quality data were obtained from previous studies of the Karkhe system project by the Mahab Qods Consulting Engineers Company and the Meteorological Organization (Iran). Thermal stratification and outlet water temperature were simulated for the 15-year period 1981–1995. The first

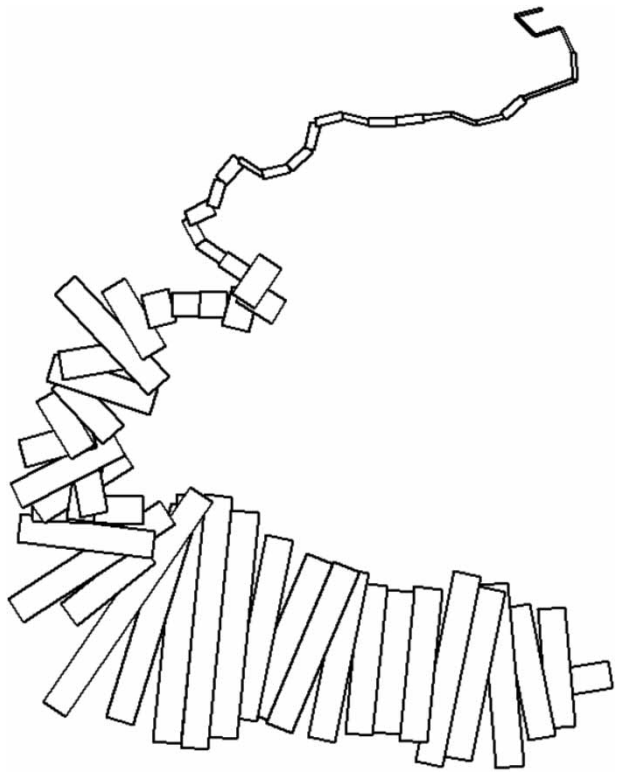

Figure 3 | Plan view of Karkhe Reservoir model segmentation (scale 1:15,000).

simulation day is September 19th, 1981, and the last day is September 15th 1995. The minimum and maximum simulation time steps are equal to one second and one hour, respectively. The CE-QUAL-W2 model was calibrated and verified by Afshar *et al.* (2011).

Implementation of the LIBSVM model

The implementation steps of the LIBSVM model are depicted in Figure 2. Using the LIBSVM model as a surrogate model for CE-QUAL-W2 model for water-quality simulations requires careful choice of the input variables. The input variables having the greatest impact on system response must be chosen so that the LIBSVM model's predictive skill is accurate.

The key variables governing surface heat exchange and reservoir thermal stratification in a reservoir are air temperature, dew point temperature, cloud cover, wind velocity, thermal inflow and outflow fluxes, water depth, and water volume (Edinger 2002; Afshar & Saadatpour 2009). An

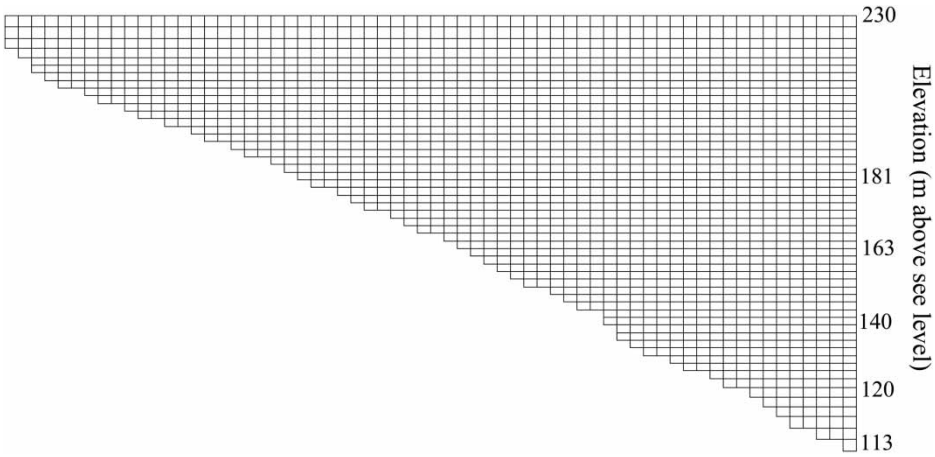

Figure 4 | Vertical view of Karkhe Reservoir model with depth layers.

information-theoretic approach based on the mutual information (MI) concept proposed by Tourassi *et al.* (2001) is applied herein to select the most significant variables among the cited governing variables to simulate Karkhe reservoir outlet water temperature and reduce the computational burden of the LIBSVM method (Saadatpour *et al.* 2017). The MI quantifies the 'amount of information' gained about outlet water temperature through other variables such as air temperature, cloud cover, and so on. Complex cause-and-effect relations govern feedbacks and time delays on large reservoir thermal responses. Therefore, appropriate time delays must be added to the input data. Moreover, the application of fractional withdrawals at various outlets must be taken into account in the LIBSVM input data to model outlet water temperature due to the withdrawals from different elevations in a lake.

The training (calibration) and testing input data of the LIBSVM model are obtained from inputs and outputs of 18 defined operational scenarios based on the authors' knowledge from simulations with the CE-QUAL-W2 model. This led to choosing the results of 12 and 6 operational scenarios among the 18 scenarios as training and testing data, respectively. Training and testing scenarios were chosen based on random selection (Liu & Motoda 2012). The withdrawal fractions at each outlet for different operational scenarios are listed in Table 1. Moreover, the LIBSVM model's performance is evaluated for two different structural states defined as follows: (1) the LIBSVM method is trained by operating scenarios 1 through 12, and tested with operating scenarios 13 through 18 for all the simulated years (1981–1995); (2) the LIBSVM method is trained with operating scenarios 1 through 12, and tested with operating scenarios 13 through 18 for similar months (12 months) during the simulation period (1981–1995).

Table 1 | Scenarios of combinations of individual withdrawal fractions at several outlets

	Outlet water level (m above sea level)			
Scenario number	**120**	**140**	**163**	**181**
1	1.00	0.00	0.00	0.00
2	0.00	1.00	0.00	0.00
3	0.00	0.00	1.00	0.00
4	0.00	0.00	0.00	1.00
5	0.33	0.33	0.33	0.00
6	0.50	0.00	0.00	0.50
7	0.50	0.50	0.00	0.00
8	0.00	0.15	0.30	0.55
9	0.50	0.00	0.50	0.00
10	0.00	0.25	0.00	0.75
11	0.00	0.00	0.50	0.50
12	0.33	0.00	0.33	0.33
13	0.00	0.55	0.30	0.15
14	0.00	0.33	0.33	0.33
15	0.00	0.75	0.00	0.25
16	0.00	0.50	0.00	0.50
17	0.15	0.30	0.00	0.55
18	0.00	0.50	0.50	0.00

RESULTS

The present study has the two main goals of: (1) simulating reservoir water temperature (both outlet water temperature and internal reservoir water temperature); and (2) implementing the LIBSVM model to estimate outlet water temperature at several outlet levels of the reservoir.

Simulating thermal stratification

The thermal stratification of Karkhe Reservoir was simulated with the CE-QUAL-W2 model for the period 1981–1995. Figure 5 shows vertical water temperature profiles in seasons of 1982, 1985, 1988, 1991, and 1994, which displays that thermal stratification begins in spring in every year of simulation (1981–1995). The surface of the reservoir warms up with the onset of spring and the density of the upper layer decreases. This prevents vertical mixing in the reservoir. The calculated water temperature on May 17th of all the simulation years in the lowest and highest layers averaged 14.18 °C and 26.66 °C, respectively. This implies a temperature difference of 12.48 °C over 88 m of depth. With the onset of summer, the water surface continues to warm up and the vertical temperature gradient increases. The surface water temperature on August 15th of all the simulation years in the lowest and highest layers averaged 14.18 °C and 33.58 °C, respectively, producing a temperature difference equal to 19.39 °C over 88 m of depth. With the onset of autumn, the epilimnion, metalimnion and hypolimnion layers become clearly delineated by November 13th of all years. The depths of the epilimnion and hypolimnion layers in this season are approximately 17 and 35 m with temperatures equal to 25.50 °C and 15 °C, respectively. Autumn exhibits the steepest thermal slope of the metalimnion layer among all seasons. The surface layer cools with advancing autumn. This lowers the temperature difference between the epilimnion and metalimnion layers causing vertical mixing. Complete vertical mixing develops through February 16th of all years. In the winter season of all simulated years, the difference between the lowest and highest layer is almost 1.2 °C over 88 m of depth. Thermal stratification does not occur during winter.

Figure 5 establishes that Karkhe Reservoir is subjected to strong thermal stratification which prevents vertical mixing in the reservoir. Therefore, applying selective withdrawal to counter thermal stratification is necessary.

Application of the LIBSVM model

The input data to the LIBSVM model are a ten-day average of air temperature, wind speed, heat flux input (obtained by multiplying reservoir inflow discharge by inflow water temperature), water surface level of the reservoir with an average three-month lag from the current condition (average lag of 1, 2 and 3 months), reservoir outflow, and reservoir withdrawal fractions (at elevations equal to 120, 140, 163, and 181 m). These variables were selected as the most important input variables to the LIBSVM model based on MI criteria results reported by Saadatpour *et al.* (2017). Selection of the variables, averaging time intervals, and time lags are based on prior knowledge and sensitivity analyses (Saadatpour 2012; Saadatpour & Afshar 2013; Saadatpour *et al.* 2017).

Structural state 1

The ε, C and σ LIBSVM model parameters were obtained based on sensitivity analysis shown in Table 2. Accordingly, the *RMSE* and number of iterations were captured for different sets of LIBSVM parameters. According to Table 2 the *RMSE* decreases and the number of iterations increases with decreasing σ, which leads to an increase of the model run time. In addition, the *RMSE* value decreases and the number of iterations increases with increasing C. The best value of ε equals 0.075. The optimal ε, C, and σ which minimize the *RMSE* and the number of iterations were equal to 0.075, 500, and 0.0006, respectively. Figure 6 portrays the calculated outlet water temperature with LIBSVM and the CE-QUAL-W2 models for corresponding to scenarios 9 and 14 as training and testing scenarios, respectively.

Structural state 2

The LIBSVM model parameters were obtained from sensitivity analysis and are listed in Table 2. For all months except September these are ε, C, and σequal to 0.075, 5000, and 0.006, respectively, and for September they are equal to 0.75, 5000,

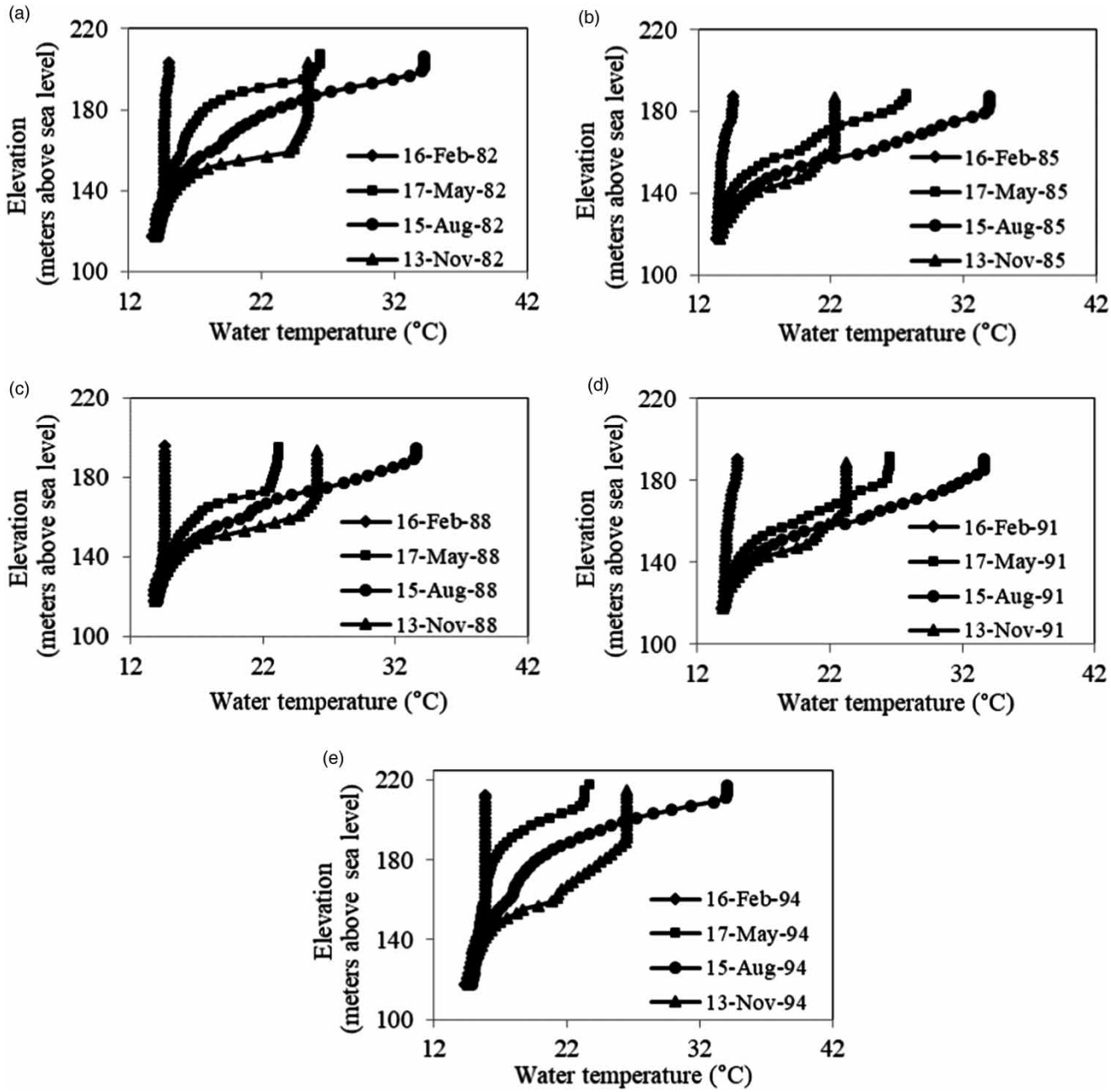

Figure 5 | Vertical water temperature profiles in several seasons of years: (a) 1982, (b) 1985, (c) 1988, (d) 1991, (e) 1994.

and 0.006, respectively. Figure 7 depicts the outlet water temperature obtained from CE-QUAL-W2 and LIBSVM models for April and September obtained from Scenario 9 and 14 as the training and testing scenarios, respectively.

Figures 6 and 7 demonstrate the LIBSVM performance in modeling outlet water temperature was excellent for the two structural states of the model. Moreover, the accuracy of the model in approximating maximum and minimum outlet water temperature is also good for each of two modeling states.

Figure 7(b) shows that in 1978 through 1990 the difference between outlet temperature predicted with the

Table 2 | Sensitivity analysis of LIBSVM parameters for structural state 1

ε	C	σ	RMSE (°C)	Number of iterations
0.075	50,000	0.06	0.2739	10,000,000
0.0750	5,000	0.06000	0.283	2,753,779
0.0750	5,000	0.00600	0.420	1,276,493
0.0750	50,000	0.00060	0.484	2,959,564
0.0750	5,000	0.00060	0.490	192,499
0.0750	500	0.00060	0.494	31,644
0.0075	5,000	0.00060	0.515	346,312
0.7500	5,000	0.00060	0.568	27,912
0.0750	500	0.00006	0.573	52,883
0.0750	50	0.00060	0.938	10,798

CE-QUAL-W2 and LIBSVM models is larger than in other years. The input variables to the LIBSVM model consist of time series data that vary substantially within a specified month through different years. Therefore, the input data for September of the years 1987 through 1990 were different from the input data of this same month in other years because of natural or man-made changes in data series. The LIBSVM identified fitting functions for each particular scenario. The performance of the LIBSVM to model outlet temperature in some years was not as good as in other years. The LIBSVM model predictions exhibit some error, but, generally, the predictive performance of the model for outlet water temperature is excellent overall.

The LIBSVM model performance based on statistical criteria is listed in Table 3, in which *Max Error* was the maximum difference between obtained the CE-QUAL-W2 and the LIBSVM model outputs. Table 3 lists the maximum values of the *RMSE*, *MAE*, and *Max Error* being equal to 0.76 °C, 0.65 °C, and 4.80 °C, respectively, and the minimum value of the *NS* is 0.92 for both structural states. It shows that both structural states had excellent performance modeling outlet water temperature. The averaged value of *RMSE*, *MAE*, *Max Error*, and *NS* for structural state 2 are 0.41 °C, 0.30 °C, 1.88 °C, and 0.96, respectively. Comparing the averaged statistical results of structural state 2 to the statistical results of structural state 1 established the LIBSVM model's excellent performance in predicting outlet water temperature for each of two structural states.

The run times for calculating outlet water temperature with CE-QUAL-W2 and LIBSVM model were about 20 minutes and less than two minutes, respectively. The LIBSVM model calculations were about 10 times faster than those for the CE-QUAl-W2 model. Evidently, the LIBSVM model exhibited accurate and relatively rapid performance for extracting optimal selective withdrawal rules for thermal control of the environmental demand.

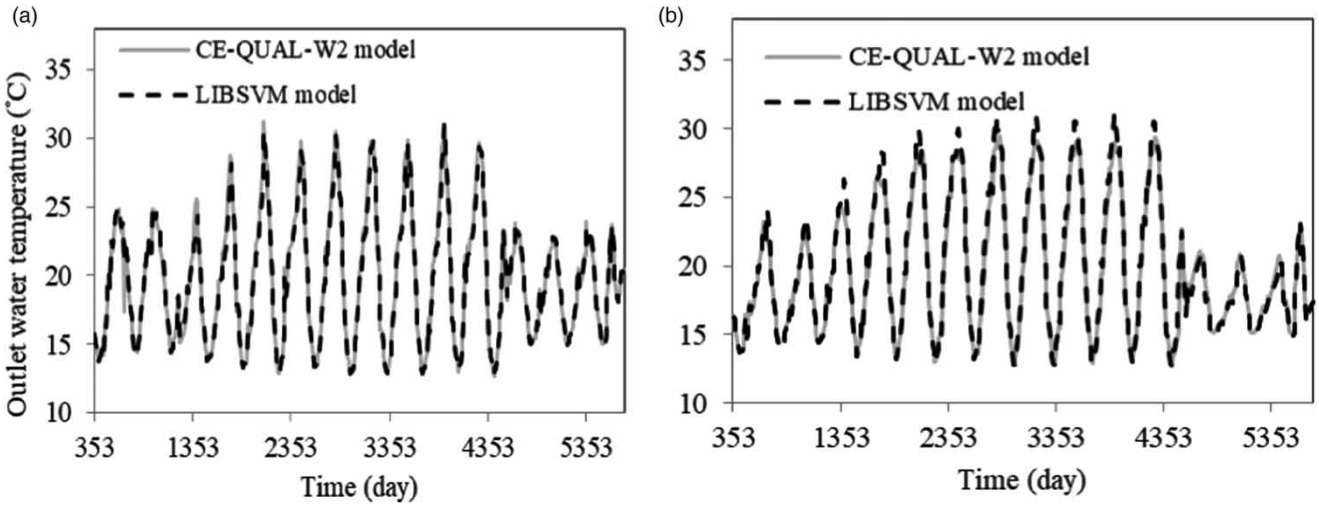

Figure 6 | Calculated outlet water temperature with CE-QAUL-W2 and the LIBSVM models during all the simulated years (1981–1995) corresponding to Scenarios (a) 9 and (b) 14 serving as training and testing cases, respectively.

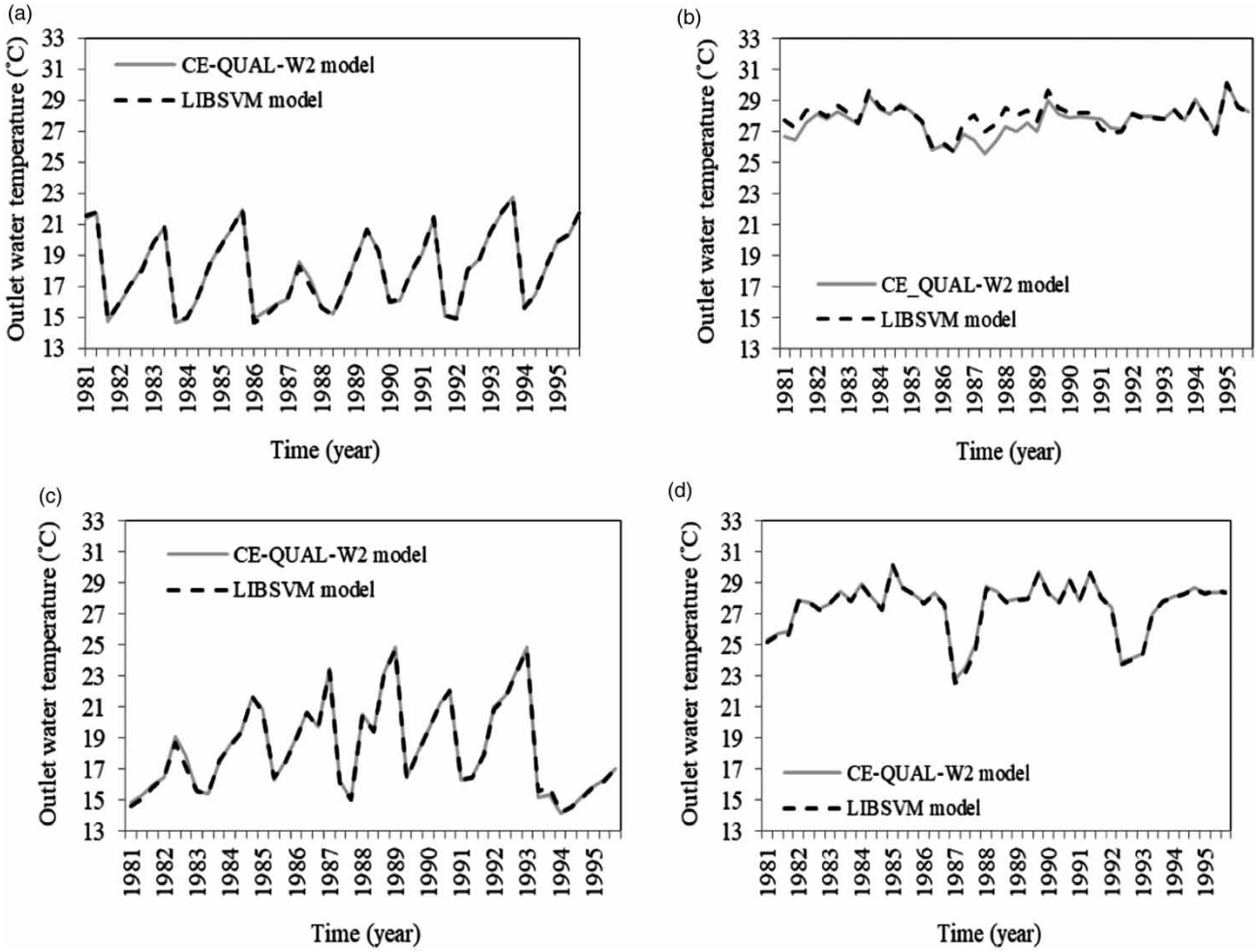

Figure 7 | Calculated outlet water temperature obtained from CE-QUAL-W2 and LIBSVM models corresponding to (a) April months of Scenario 9, (b) September months of Scenario 9, (c) April months of Scenario 14, and (d) September months of Scenario 14 during simulated years. The April and September calculations represent the training and testing results, respectively, in (a), (b), (c), and (d).

CONCLUDING REMARKS

This research simulated the thermal stratification in the Karkhe Reservoir in the period 1981–1995 employing the CE-QUAL-W2 model. The LIBSVM model was applied as a surrogate of the CE-QUAL-W2 model to approximate outlet water temperature of the reservoir. The application of LIBSVM reduces the computational time, and the number of input data needed for simulation purposes. Results demonstrated the run time of the LIBSVM model was approximately 10 times shorter than the run time of CE-QUAL-W2 model in this study. The simulation of reservoir thermal stratification showed a stratification cycle during each simulated year, in which the thermal stratification began in spring and complete vertical mixing occurred in the mid-winter of all the simulated years. The statistical results of the data-mining model performance measured the capacity of the LIBSVM model to approximate outlet water temperature, with the minimum and maximum values of the *NS* criteria equaling 0.92 and 0.99, respectively. These encouraging results indicate the LIBSVM model can be used instead of the more complex CE-QUAL-W2 model and be coupled with an optimization model to calculate operating policies at reservoir outlets to control outflow temperature from the reservoir. The trained algorithm performed well in predictions for Karkhe

Table 3 | LIBSVM model performance based on statistical criteria for two different structural states for testing data

Structural state	Months of simulated years	*RMSE* (°C)	*MAE* (°C)	*Max Error* (°C)	*NS*
1	All months	0.49	0.25	4.80	0.99
2	January	0.18	0.13	1.03	0.97
2	February	0.09	0.07	0.41	0.99
2	March	0.09	0.07	0.50	0.98
2	April	0.35	0.22	2.10	0.95
2	May	0.37	0.27	1.33	0.98
2	June	0.44	0.33	2.55	0.98
2	July	0.40	0.30	1.77	0.99
2	August	0.65	0.45	2.80	0.98
2	September	0.74	0.65	3.07	0.99
2	October	0.76	0.47	3.71	0.93
2	November	0.52	0.35	2.36	0.92
2	December	0.30	0.34	0.97	0.96

Reservoir. Such performance would have to be evaluated in other reservoirs.

ACKNOWLEDGEMENT

The authors thank Iran's National Science Foundation (INSF) for its financial support of this research.

REFERENCES

Afshar, A. & Saadatpour, M. 2009 Reservoir eutrophication modeling, sensitivity analysis, and assessment; application to Karkheh reservoir, Iran. *Journal of Environmental Engineering Science* **26**, 1227–1238.

Afshar, A., Kazemi, H. & Saadatpour, M. 2011 Particle swarm optimization for automatic calibration of large scale water quality model (CE-QUAL-W2); application to Karkheh Reservoir, Iran. *Journal of Water Resource Management* **25**, 2613–2632.

Bermudez, M., Cea, L., Puertas, J., Rodriguez, N. & Baztan, J. 2018 Numerical modeling of the impact of a pumped-storage hydroelectric power plant on the reservoirs' thermal stratification structure: a case study in NW Spain. *Environmental Modeling and Assessment* **23**, 71–85.

Chang, C.-C. & Lin, C.-J. 2011 LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* **2**, 1–27.

Cole, T. M. & Wells, S. A. 2008 *CE-QUAL-W2: A Two-Dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model*. Version 3.71, User Manual, Washington, DC, USA.

Dortch, M. S. 1997 *Water quality considerations in reservoir management*. Water Resources Update, Universities Council on Water Resources, Southern Illinois, University Carbondale, USA.

Edinger, J. E. 2002 *Waterbody Hydrodynamic and Water Quality Modeling: an Introductory Workbook and CD-ROM on Three-Dimensional Water Body Modeling, Ch. 11*. ASCE Press, Reston.

Elci, S. 2008 Effects of thermal stratification and mixing on reservoir water quality. *Limnology* **9**, 135–142.

Fischer, H. B., List, E. J., Koh, R., Imberger, J. & Brooks, N. H. 1979 *Mixing in Inland and Coastal Waters*. Academic Press, New York, USA.

Furman, A. 2008 Modeling coupled surface–subsurface flow processes: a review. *Vadose Zone Journal* **7**, 741–756.

Gelda, R. K. & Effler, S. 2007 Testing and application of a two-dimensional hydrothermal model for a water supply reservoir: implications of sedimentation. *Journal of Environmental Engineering and Science* **6**, 73–84.

Giudici, P. 2003 *Applied Data Mining: Statistical Methods for Business and Industry, No. 1*. Wiley, New York, USA.

Hanna, R. B., Saito, L., Bartholow, J. M. & Sandelin, J. 1999 Results of simulated temperature control device operations on in-reservoir and discharge water temperatures using CE-QUAL-W2. *Lake and Reservoir Management* **15**, 87–102.

Kennedy, R. H., Gunkel Jr, R. C. & Thornton, K. W. 1982 The establishment of water quality gradients in reservoirs. *Canadian Water Resources Journal* **7**, 71–87.

Kerimoglu, O. & Rinke, K. 2013 Stratification dynamics in a shallow reservoir under different hydro-meteorological scenarios and operational strategies. *Water Resources Research* **49**, 7518–7527.

Liu, H. & Motoda, H. 2012 *Feature Selection for Knowledge Discovery and Data Mining*, Vol. 454. Springer Science and Business Media, Berlin.

Lugg, A. & Copeland, C. 2014 Review of cold water pollution in the Murray–Darling basin and the impacts on fish communities. *Ecological Management and Restoration* **15**, 71–79.

McCuen, R. H. 1973 The role of sensitivity analysis in hydrologic modeling. *Journal of Hydrology* **18**, 37–53.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G. & Gupta, H. 2015 Are we unnecessarily constraining the agility of complex process-based models? *Water Resources Research* **51**, 716–728.

Montanari, A. & Grossi, G. 2008 Estimating the uncertainty of hydrological forecasts: a statistical approach. *Water*

Resources Research **44**, W00B08. doi: 10.1029/2008WR006897.

Orouji, H., Bozorg-Haddad, O., Fallah-Mehdipour, E. & Mariño, M. A. 2013 Modeling of water quality parameters using data-driven models. *Journal of Environmental Engineering* **139**, 947–957.

Raghavendra, N. S. & Deka, P. C. 2014 Support vector machine applications in the field of hydrology: a review. *Applied Soft Computing* **19**, 372–386.

Rango, A. & Martinec, J. 1995 Revisiting the degree-day method for snowmelt computations. *JAWRA Journal of the American Water Resources Association* **31**, 657–669.

Saadatpour, M. 2012 Deriving Optimal Reservoir Operational Strategy Considering Quality and Quantity Objectives. PhD Dissertation, Department of Civil Engineering, Iran University of Science and Technology, Tehran, Iran (in Persian).

Saadatpour, M. & Afshar, A. 2013 Multi objective simulation-optimization approach in pollution spill response management model in reservoirs. *Water Resources Management* **27**, 1851–1865.

Saadatpour, M., Afshar, A. & Edinger, J. E. 2017 Meta-model assisted 2D hydrodynamic and thermal simulation model (CE-QUAL-W2) in deriving optimal reservoir operational strategy in selective withdrawal scheme. *Water Resources Management* **31**, 2729–2744.

Sahoo, G. B. & Schladow, S. G. 2008 Impacts of climate change on lakes and reservoirs dynamics and restoration policies. *Sustainability Science* **3**, 189–199.

Tourassi, G. D., Frederick, E. D., Markey, M. K. & Floyd Jr, C. E. 2001 Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Am. Assoc. Phys. Med.* **28**, 2394–2402.

Vapnik, V. N. 1998 *Statistical Learning Theory*. John Wiley, New York, USA.

Vapnik, V., Golowich, S. E. & Smola, A. 1996 Support vector method for function approximation, regression estimation, and signal processing. *Proceedings of the 9th International Conference on Neural Information Processing Systems Advances*, Denver, USA, 3–5 December.

Wang, W. C., Chau, K. W., Cheng, C. T. & Qiu, L. 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* **374**, 294–306.

Wang, S., Qian, X., Han, B., Luo, L. & Hamilton, D. P. 2012 Effects of local climate and hydrological conditions on the thermal regime of a reservoir at tropic of cancer, in southern China. *Water Research* **46**, 2591–2604.

Wei, L., Dingguo, J. & Tao, C. 2011 Effects of flood on thermal structure of a stratified reservoir. *Procedia Environmental Sciences* **10**, 1811–1817.

Wetzel, R. G. 1983 *Limnology*. Saunders College Publishing, Philadelphia.

Xiang, Y. & Jiang, L. 2009 Water quality prediction using LS-SVM with particle swarm optimization. In: *Second International Workshop on Knowledge Discovery and Data Mining, IEEE 2009*, Moscow, Russia.

Yang, M., Li, L. & Li, J. 2012 Prediction of water temperature in stratified reservoir and effects on downstream irrigation area: a case study of Xiahushan reservoir. *Physics and Chemistry of the Earth Parts A/B/C* **53**, 38–42.

First received 21 March 2018; accepted in revised form 9 September 2018. Available online 7 November 2018