




Original Paper

A Novel Multiple-Kernel Support Vector Regression Algorithm for Estimation of Water Quality Parameters

Mohammad Najafzadeh^{1,2} and Saeid Niazmardi^{1,2,3} 

Received 15 January 2021; accepted 30 May 2021
Published online: 17 June 2021

The quality of surface waters plays a key role in the sustainability of ecological systems. Measuring water quality parameters (WQPs) is of high importance in the management of surface water resources. In this paper, contemporary-developed regression analysis was proposed to estimate the hard-to-measure parameters from those that can be measured easily. To this end, we proposed a novel modification of support vector regression (SVR), known as multiple-kernel support vector regression (MKSVR) algorithm. The MKSVR learns an optimal data representation for regression analysis by either linear or nonlinear combination of some precomputed kernels. For solving the optimization problem of the MKSVR, the particle swarm optimization (PSO) algorithm was used. The proposed algorithm was assessed using WQPs taken from Karun River, Iran. MKSVR was used to estimate chemical oxygen demand (COD) and biochemical oxygen demand (BOD) using nine WQPs as the input variables, namely electrical conductivity, sodium, calcium, magnesium, phosphate, nitrite, nitrate nitrogen, turbidity, and pH. The results of the proposed MKSVR were compared with those obtained using the SVR and Random Forest regression (RFR). The results showed that the MKSVR algorithm (correlation coefficient $[R] = 0.8$ and root mean squared error $[RMSE] = 4.76$ mg/l) increased the accuracy level of BOD prediction when compared with SVR ($R = 0.68$ and $RMSE = 5.15$ mg/l) and RFR ($R = 0.77$ and $RMSE = 5.15$ mg/l). In the case of COD estimation, the performance of a developed support vector machine (SVM) technique was satisfying. Overall, the use of MKSVR along with the PSO algorithm could demonstrate the superiority of the newly developed SVM technique for the WQPs estimation in the natural streams.

KEY WORDS: Multi-kernel learning, Support vector regression, Water quality parameters, Natural water resources, Regression analysis, Kernel learning.

INTRODUCTION

Measuring water quality parameters (WQPs) plays a significant role in environmental monitoring. As water quality stands at a poor state, it has a negative influence on the life of living organisms in various aquatic ecosystems and water bodies such as natural rivers, lakes, dam reservoirs, confined and unconfined aquifers. Several parameters that characterize physical, chemical, and biological (or biochemical) characteristics of water are considered as

¹Department of Water Engineering, Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, P.O. Box 7631885356 Kerman, Iran.

²Department of Surveying Engineering, Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, P.O. Box 7631885356 Kerman, Iran.

³To whom correspondence should be addressed; e-mail: s.niazmardi@kgut.ac.ir

WQP in the literature, among which the most used ones are chemical oxygen demand (COD), and biochemical oxygen demand (BOD), phosphate, turbidity, and electrical conductivity.

The accurate measurement of WQPs is at the mercy of taking a lot of time and difficult-to-carry out procedures. Measuring some of these WQPs such as BOD and COD is more complicated than the others. To obtain a reasonable estimation of BOD concentration, two types of measurement should be considered: first, the amount of oxygen, required to perform oxidization process for all the organic elements in a specific water volume, and second, the amount of oxygen, absorbed by other living creatures. Thus, the performance of fields and experimental investigations may result in inaccuracies because the volume of absorbed oxygen is not taken into account. Similarly, in accordance with COD measurement, the results of experimental studies are distorted in the presence of different inorganic radicals (Verma & Singh, 2013).

To eradicate these difficulties, many researchers proposed to estimate the WQPs (for example, chemical oxygen demand, biochemical oxygen demand, dissolved oxygen demand) as a function of other common WQPs (for instance, nitrate, turbidity, electrical conductivity, pH) instead of measuring them. To this end, different regression models have been considered. From previous investigations, artificial neural networks (ANNs) (Ay & Kisi, 2011; Emamgholizadeh et al., 2014; Singh et al., 2009), adaptive neuro-fuzzy inference system (ANFIS) (Emamgholizadeh et al., 2014; Soltani et al., 2010), and support vector machine (SVM) (Bozorg-Haddad et al., 2017; Li et al., 2017) have been employed frequently to estimate the WQPs in various water bodies through the world. Furthermore, along with the use of these techniques, some other techniques such as gene-expression programming (GEP), evolutionary polynomial regression (EPR), M5 model tree (Najafzadeh et al., 2018), multivariate adaptive regression spline (MARS) (Heddami & Kisi, 2018; Najafzadeh & Ghaemi, 2019), and linear genetic programming (LGP) proved to be efficient for estimation of WQPs. In addition, wavelet decomposition techniques, locally weighted linear regression model and multigene genetic programming, have been applied recently to estimate accurately different WQPs in natural streams (Ahmadianfar et al., 2020; Jamei et al., 2020).

One of the most successful ones of these regression algorithms is the support vector regres-

sion (SVR), which is a support vector machine algorithm (SVM) developed for regression analyses (Cortes & Vapnik, 1995; Smola & Schölkopf, 2004). This algorithm has been used widely and proved efficient in different fields for parameter estimation and time-series analysis and forecasting (Mukherjee et al., 1997; Niazmardi et al., 2013; Tuia et al., 2011; Wu et al., 2004; Yu et al., 2006). However, the SVR suffers from the same shortcomings as the other members of the SVM family, i.e., its performance is highly tied with the proper selection of its kernel function (Abbasnejad et al., 2012). Selection (or construction of) an optimal kernel function for a learning problem is not a straightforward task. Thus, in the last decades, several kernel learning approaches have been proposed (Abbasnejad et al., 2012). Multiple-kernel learning (MKL) algorithms are one of the categories of kernel learning approach which, due to their sound theoretical background and outstanding results, gained the attention of many researchers (e.g., Bucak et al., 2014; Gönen & Alpaydm, 2011; Niazmardi et al., 2016, 2018).

One of the main issues of the SVM algorithm is that its performance depends highly on the choice of kernel function and fine-tuning the parameters of the selected kernel. During the last decade, several techniques have been proposed to assist this choice, among which the MKL framework was the most promising one (i.e., Niazmardi et al., 2018; Qiu & Lane, 2005, 2009; Yeh et al., 2011). In another word, the MKL algorithms can solve the kernel selection problem in the kernel-based learning algorithms. Different kernel functions are combined using a combination function, which can be either linear or nonlinear. The proposed algorithm is among the MKL algorithms that can handle both types of combination functions. The parameters of the combination function are estimated by solving the existing optimization problem.

The MKL algorithms address the problem associated with selecting the proper kernel function, by learning an optimal task-specific kernel through either linear or nonlinear combination of some precomputed kernels (Bucak et al., 2014). Most of the MKL algorithms have been proposed for classification purposes (Niazmardi et al., 2018), and there are only a few algorithms proposed for the regression analysis (Gonen & Alpaydm, 2010; Qiu & Lane, 2005, 2009; Yeh et al., 2011). These few algorithms cannot learn the nonlinear combination of kernels and usually use complex optimization strategies. To address these issues, we proposed

multiple-kernel support vector regression (MKSVR) algorithm for accurate estimation of WQPs. The MKSVR benefits from a flexible structure of MKL algorithm which can learn both linear and nonlinear combinations of kernels for regression analysis. Besides, this algorithm uses the particle swarm optimization algorithm to optimize the combination of the kernels, which makes its implementation very easy.

The rest of this paper is organized as follows. First, brief reviews of the theorem associated with the SVR and MKL algorithms are presented. After that, the structure of the proposed MKSVR algorithm and its optimization strategy is described. Next, water quality data and experimental setups are presented, followed by the presentation of the results in terms of qualitative and quantitative performance. The conclusions are drawn in the final step of this research.

METHODOLOGY

Regression and Support Vector Regression Algorithms

Suppose we are provided with a set $T = \{x_i, y_i\}_{i=1}^n$ of n training samples $x_i \in \mathbb{R}^p$, each of which is assigned with a real-valued target $y_i \in \mathbb{R}$. Assume that these samples are obtained through sampling an unknown function $g: \mathbb{R}^p \rightarrow \mathbb{R}$. The main purpose of regression is to estimate a function $f: \mathbb{R}^p \rightarrow \mathbb{R}$ that approximates the unknown function g using the set of training samples (Mukherjee et al., 1997); accordingly, regression is also known as function approximation in some literature.

Several mathematical methods were proposed for solving regression problems. Among these models, the support vector regression attracted great attention (Camps-Vails et al., 2006; Gunn, 1998; Niazmardi et al., 2013; Qiu & Lane, 2009; Smola & Schölkopf, 2004). The SVR aims at approximating a linear function $f(x) = w^T \varphi(x) + b$ in which w and b are regression parameters that should be estimated using the training data. In this function, also known as the prediction function or regressor, $\varphi(\cdot)$ is a mapping function from the original space of data into the kernel space. The SVR algorithm estimates w and b through optimizing a loss function and a regularization term (Mukherjee et al., 1997). The loss function minimizes the error of estimated

function, while the regularization term controls its flatness (Rojo-Álvarez et al., 2018).

One of the most used loss functions in the structure of SVR is ϵ -insensitive loss function, proposed by Vapnik (2013). The value of this function L_ϵ for an error value e is calculated as:

$$L_\epsilon(e) = \max(0, |e| - \epsilon) \tag{1}$$

The flatness of the SVR regressor can be estimated through the calculation of its norm $\|w\|^2$. Accordingly, the SVR optimization is written as:

$$\min_{w,b} \left\{ C \sum_{i=1}^n L_\epsilon(y_i - w^T \varphi(x_i) - b) + \frac{1}{2} \|w\|^2 \right\} \tag{2}$$

where C is a positive real number (known as trade-off parameter) that controls the trade-off between the flatness and the error of the estimated function. It can be shown that the minimization of Eq. 2 is equivalent to the following constraint optimization problem (Scholkopf & Smola, 2001):

$$\min_{w,b,\xi_i,\xi_i^*} \left\{ C \sum_{i=1}^N (\xi_i + \xi_i^*) + \frac{1}{2} \|w\|^2 \right\} \tag{3}$$

Constrained to:

$$y_i - \langle w, \varphi(x_i) \rangle - b \leq \epsilon + \xi_i \quad \forall i = 1, \dots, N \tag{4}$$

$$\langle w, \varphi(x_i) \rangle + b - y_i \leq \epsilon + \xi_i^* \quad \forall i = 1, \dots, N \tag{5}$$

$$\xi_i, \xi_i^* \geq 0 \quad \forall i = 1, \dots, N. \tag{6}$$

In Eq. (3), ξ_i and ξ_i^* are the positive slack variables that are used to cope with the infeasible constraints (Smola & Schölkopf, 2004). With the aid of the Lagrange multiplier technique, Eq. (3) is re-written by the dual problem as:

$$\begin{aligned} \max_{\alpha, \alpha^*} & \left\{ \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) - \epsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \right\} \\ \text{subject to:} & \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ & \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \tag{7}$$

where α and α^* are the dual variables associated with inequality constraint of Eqs. (4) and (5), respectively. This is a convex optimization problem that is solved conveniently. After solving this optimization problem and obtaining optimum values for dual variables, b is estimated considering the conditions and training samples, and w is calculated as:

$$w = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \varphi(x_i) \tag{8}$$

After computing b and w , the target value y is estimated as:

$$y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{9}$$

where K is known as the kernel function with only two open parameters ε and C .

Multiple-Kernel Learning

Choosing a kernel function and fine-tuning its parameters have a profound effect on the performance of kernel-based learning algorithms such as support vector regression (Yeh et al., 2011). There are several different kernel functions available for a specific learning task, from which the user should choose the best performing one without any prior knowledge about their performances. There is a wide range of kernel learning methods that are employed to either assist this choice or to estimate a valid kernel function from the available training data (Abbasnejad et al., 2012). The most promising category of these algorithms is the MKL algorithms.

MKL algorithms estimate a (sub)-optimal kernel function, known as the composite kernel, for a specific learning task by combining a group of pre-computed basis kernels (Gönen & Alpaydın, 2011). The basis kernels are combined by the use of a parametric combination function into the composite kernel. Thus, the main goal of MKL algorithms is to estimate the optimal values for the parameters of the combination function (Niazmardi et al., 2018). MKL algorithms yield this goal by optimizing a target function with respect to these parameters. Although most of the MKL algorithms have been proposed for the classification problems, the optimization techniques and the combination functions associated with these algorithms can be also used for the regression problems (Bucak et al., 2014; Gönen & Alpaydın, 2011; Kloft et al., 2011; Niazmardi et al., 2016). In the MKL literature, it is a common practice to replace the kernel function of the dual problem related to the kernel-based learning task (Eq. 7 in the case of the SVR algorithm) with the composite kernel. This issue was considered as the target function of the MKL algorithm (Bucak et al.,

2014). Following this strategy, the target function of the MKSVR can be written as the following min-max problem:

$$\begin{aligned} \min_{\eta} \max_{z, z^*} & \left\{ -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K_c(x_i, x_j) - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \right\} \\ \text{Subject to: } & \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ & \alpha_i, \alpha_i^* \in [0, C] \\ & \eta \in \Delta \end{aligned} \tag{10}$$

where η and Δ , respectively, denote the parameter of the combination function and their feasible set; K_c shows the composite kernel. An alternative optimization strategy is used occasionally to solve this optimization because optimizing Eq. (10) with respect to η is not a convex problem (Gönen & Alpaydın, 2011).

As mentioned previously, one of the most important characteristics of the MKL algorithm is how the basis kernels are combined into the composite kernel which is controlled by the combination function. Considering the linearity of this function, the MKL algorithms can be categorized into two groups of linear and nonlinear algorithms (Niazmardi et al., 2016). The linear MKL algorithms apply the following function to n available basis kernels $K_i, i = 1, \dots, n$ to construct the composite kernel:

$$K_c = \sum_{i=1}^n d_i K_i \tag{11}$$

where $d_i, i = 1, \dots, n$ are non-negative weights associated with the basis kernels, which should be optimized by using the MKL algorithm.

Several options available can be used as the combination function of the nonlinear MKL algorithms. Among these functions, the polynomial function of degree in the state of $d \geq 1$ is the most common (Cortes et al., 2009), and it is expressed as:

$$K_c = \sum_{q \in Q} \mu_{q_1 \dots q_n} K_1^{q_1} K_2^{q_2} \dots K_n^{q_n} \tag{12}$$

where Q is known as $\{q : q \in \mathbb{Z}_+^n, \sum_{i=1}^n q_i \leq d\}$ and $\mu_{q_1 \dots q_n}$ is equal to zero at least. Due to many open parameters, adopting Eq. (12) as the combination function leads to create an optimization problem with a high degree of complexity. To reduce the complexity of this problem, the following combination function is considered occasionally (Cortes et al., 2009):

$$K_c = \sum_{q \in R} \mu_1^{q_1} \mu_2^{q_2} \dots \mu_n^{q_n} K_1^{q_1} K_2^{q_2} \dots K_n^{q_n} \quad (13)$$

where $R = \{q : q \in \mathbb{Z}_+^n, \sum_{i=1}^n q_i = d\}$ and $\mu_{q_1, \dots, q_n} \geq 0$.

According to the literature, quite a few linear and nonlinear MKL algorithms have been employed for data classification by adopting various target functions and optimization strategies. Additionally, most of these algorithms are applied for regression problems. To the best of our knowledge, there is no general guideline for selecting fixed values for the SVR parameters. Thus, these parameters are estimated occasionally using an n -fold cross-validation technique. Although SVR demonstrated acceptable performance, this technique is highly time-consuming. Besides, the user should provide a set of candidate values from which the optimum values for the SVR parameters are selected. These issues may yield sub-optimal values for the parameters. However, these problems can be avoided through simultaneous optimization of the SVR parameters and the kernel parameters in the MKSVR algorithm. We refer the readers to Gönen and Alpaydm (2011), Bucak et al. (2014) and Niazmardi et al. (2018) for details of MKL algorithms.

Proposed MKSVR

The MKSVR algorithm should optimize Eq. (10) jointly with respect to the dual variables and parameters of the considered combination function. This is achieved occasionally using an alternative optimization (AO) strategy. The AO, introduced as a two-stage optimization strategy, optimizes either dual variables or the parameters of the combination functions at each stage while assigning fixed values to the other one. The AO algorithm iterates until a termination criterion is met.

Optimizing the target function of the MKSVR with respect to the dual variable is a convex problem that can be solved easily. However, optimization with respect to the parameters of the combination function can be highly challenging due to the non-convexity of the optimization problem. Occasionally, the gradient descendant method has been employed widely to solve this optimization problem

(e.g., Rakotomamonjy et al., 2008; Varma & Babu, 2009). In contrast, the gradient descendant method is an iterative optimization method that needs several evaluations of the gradient of the target function at each step. Accordingly, adopting this method increases the computational complexity of the MKL algorithm.

To address the above issue, we proposed an AO framework that replaces the gradient descendant method with the particle swarm optimization (PSO) algorithm. In the proposed optimization strategy, the parameters of the combination function are considered as the particles, whose search space is defined as the feasible set of the parameters of the combination function. To evaluate the fitness of each particle, one composite kernel is constructed by considering the values of each particle as the parameters of the combination function. Equation (7) can be solved easily by employing a convex optimization method, once this composite kernel as the main kernel function of Eq. (7) was adopted. After solving this optimization, the prediction accuracy of this function is evaluated by employing a five-fold cross-validation method. In fact, the fitness value of each particle is considered as the precision level of the models' performance.

The proposed optimization strategy of MKSVR has several advantages over the gradient descendant method. Firstly, the proposed strategy is very flexible and able to perform with any combination function. Secondly, this is a cost-effective method owing to the fact that there is no need to estimate or evaluate the gradient. Finally, in addition to the parameters of the combination function, the proposed strategy is capable of optimizing the SVR parameters (i.e., C and ε). In fact, SVR parameters are estimated occasionally using n -fold cross-validation, which is a very time-consuming technique. In this case, each particle consists of two separate parts for parameters of combination function and SVR parameters (C and ε). Table 1 shows the optimization strategy of the MKSVR algorithm. The flowchart of the processing steps of the MKSVR algorithm is illustrated in Fig. 1. It is noteworthy that the PSO algorithm is adopted due to the continuity of its search space and its satisfying performance (Sengupta et al., 2019). However, the PSO algorithm can be replaced by other meta-heuristics optimization techniques.

Table 1. Optimization strategy of MKSVR algorithm

Inputs:	Basis kernels, PSO parameters, type of combination function
Outputs:	Parameters of combination function, SVR parameters (C and ε)
1,	Generate randomly the initial swarm of PSO (the first part consists of combination function parameters and the second part consists of the SVR parameter)
2,	while (the PSO termination criteria is not met)
3,	for $i = 1$ to number of particles in the swarm
	1, Extract the parameters of the combination function and SVR parameters from the particle i .
	2, Construct the composite kernel using the extracted parameters of the combination function.
	3, Train the SVR algorithms with the obtained composite kernel and the extracted SVR parameters
	4, Fitness function of particle $i \leftarrow$ 5-fold cross-validation error of the trained SVR
4,	end for
5,	Update position of all particles using PSO
6,	end while
7,	Return the position of the best particle for parameters of combination function, C and ε

STUDY AREA, DATASET DESCRIPTION, AND EXPERIMENTAL SETUP

Study Area

We evaluated the performance of the proposed strategy for estimating WQPs of Karun River in Khuzestan Province, Iran. This river drains from the Bakhtiari area in the central Zagros Mountain and follows a tortuous course on the Khuzestan plain, and joins the Shatt al-Arab in Bousher before its final discharge into the Persian Gulf. Karun River (Fig. 2), with 829 km long and a watershed area of 65,230 km², is the longest and the only navigable waterway of Iran. Quite a few dams have been constructed on the Karun River, whose main aims are not only hydro-power generation but also flood control. There is no denying the fact that dams on the Karun River play a key role in the evolution of some riverine issues such as land use, sediment transport, and management of water quality. Karun River is also the main source of water for several cities, among which the largest is Ahvaz with just above 1.3 million residents. Thus, assessment of the water quality of this river is of high practical importance.

Dataset Description

In this paper, 11 different WQPs were considered, namely BOD, COD, electrical conductivity (EC), sodium (Na⁺), calcium (Ca²⁺), magnesium (Mg²⁺), phosphate (PO₄³⁻), nitrite (NO₂⁻), nitrate nitrogen (NO₃⁻), turbidity, and pH. The WQPs were measured monthly from eight hydrometric stations along the Karun River between the years 1995 and 2011; the location of these stations is shown in Fig. 3 and is also listed in Table 2. As mentioned, COD and BOD are harder to measure than the other WQPs. Thus, these parameters were considered as the main variables to be estimated using the other nine different parameters. In the case of Karun River, Emamgholizadeh et al. (2014) were the first researchers to use these WQPs data for estimation of BOD and COD by ANN and ANFIS models. Additionally, Najafzadeh et al. (2018) applied several explicit formulations for prediction of BOD and COD by using EPR, MT, and GEP models. Najafzadeh and Ghaemi (2019) applied MARS and SVM techniques recently to estimate BOD and COD; they used an improved simple version of SVM. In contrast, this present research aimed to employ a newly developed version of SVM on the basis of kernel learning. The main statistics of the WQP used in this study are given in Table 3.

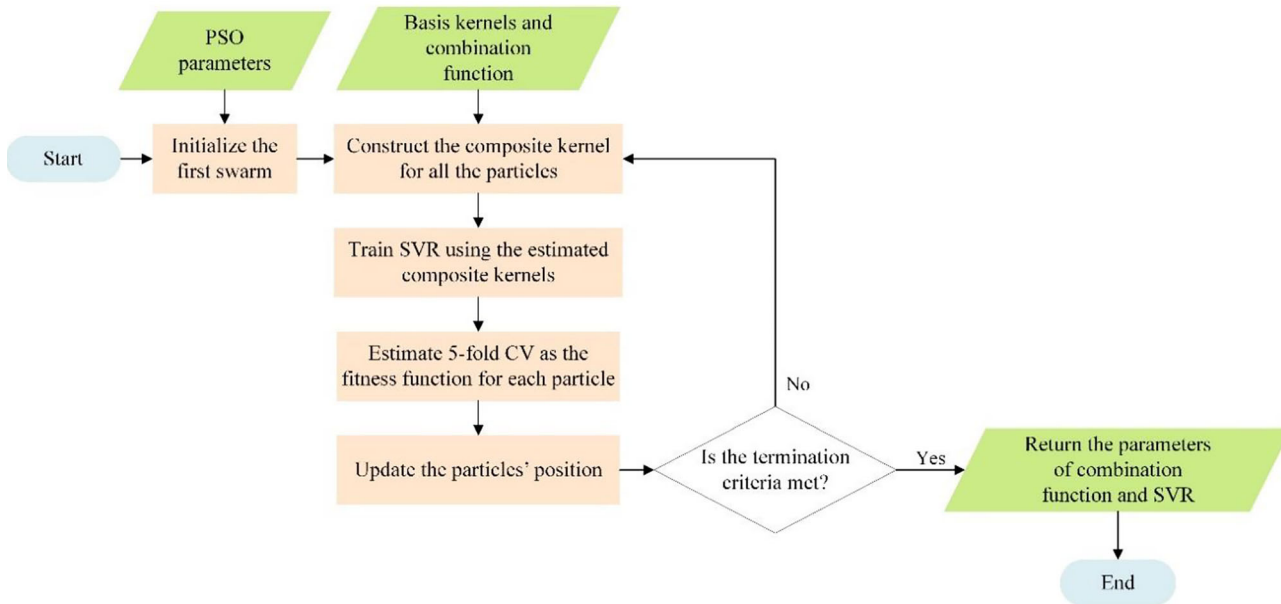


Figure 1. Flowchart of MKSVR.

In Table 3, magnesium and calcium with 60 mg/l and 58.4 mg/l, respectively, have the highest concentrations. These two parameters also had relatively large standard deviations, which can be interpreted as high dispersion in the concentration levels of these parameters. Nitrite and nitrate nitrogen have become almost stable during the measuring period; accordingly, they showed the smallest standard deviations among the parameters.

Experimental Setup

To assess the performance of the proposed MKSVR, we designed two different experiments. In the first experiment, the performance of the MKSVR was evaluated. The MKSVR algorithm with the proposed optimization strategy was implemented for both linear and nonlinear combination functions (2nd-degree polynomial). The values of the trade-off and the epsilon parameters were also optimized along with the parameters of the combination function. For this algorithm, 19 different functions were constructed as the basis kernels. The basis kernels consisted of nine different radial basis function (RBF) kernels and 10 polynomial kernels, whose parameters were selected, respectively, from the ranges $\{10^{-4}, 10^{-3}, \dots, 10^4\}$ and $\{1, 2, \dots, 10\}$. To run the MKSVR algorithm, control parameters

of the PSO algorithm, introduced as swarm size, the number of iterations, inertia weight, and the acceleration constants (shown usually by c_1 and c_2 in the literature), need to be set. Based on the suggestions by Shi and Eberhart (1998), Trelea (2003) and Bansal et al. (2011), the swarm size, the number of iterations, and inertia weight were fixed at 20, 300, and 0.72, respectively. Both acceleration parameters were set to 2.

In the second experiment, the results of the proposed method were compared with those obtained using other regression algorithms. In this step of the experiment, the Random Forest regression (RFR) and SVR were selected as benchmark for comparison. The SVR algorithm was implemented by adopting both the polynomial and the RBF as kernel functions. For this experiment, besides the value of the used kernel parameter (i.e., the spread of RBF kernel and the degree of polynomial), values of the SVR trade-off and the epsilon parameters should be set. Here, we used a fivefold cross-validation strategy to tune these parameters. The value of spread of the RBF kernel and the epsilon were both selected from the range $\{10^{-4}, 10^{-3}, \dots, 10^4\}$ and the used range for the value of trade-off parameters was $\{10^{-3}, 10^{-2}, \dots, 10^3\}$. The degree of polynomial kernel function was selected from the range $\{1, 2, \dots, 10\}$.



Figure 2. Drainage basin of Karun River (<https://en.wikipedia.org/wiki/Karun>).

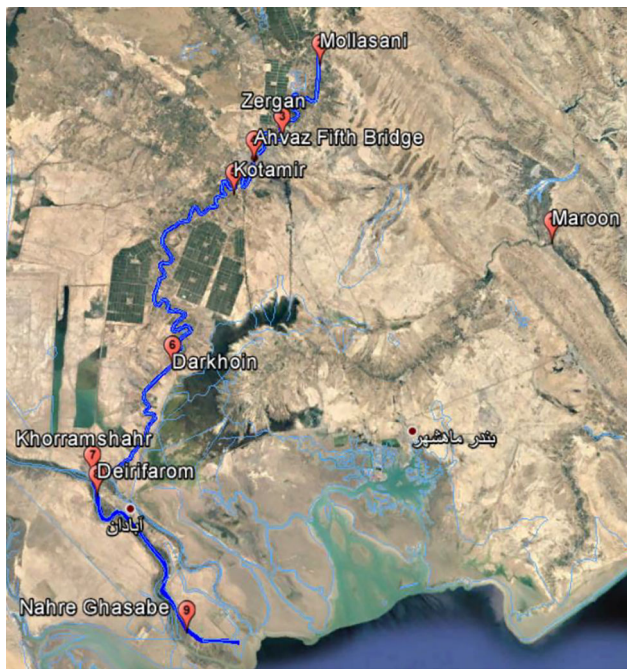


Figure 3. Location of hydrometry stations along the Karun River.

Table 2. Names and coordinates of the hydrometry stations

Hydrometry Station	Region	Longitude	Latitude	Height
Mollahsani	Karun	48°, 53', 00"	31°, 35', 00"	18
Zergan	Karun- Northern Ahvaz	48°, 45', 41"	31°, 22', 27"	20
Ahvaz Fifth Bridge	Karun- Southern Ahvaz	48°, 40', 09"	31°, 18', 26"	21
Kot Amir	Karun- Southern Ahvaz	48°, 36', 17"	31°, 12', 55"	2
Maroun	Maroun-Shadegan	49°, 36', 13"	31°, 04', 31"	64
Darkhoin	Karun	48°, 25', 00"	30°, 44', 00"	1700
Khorramshar	Karun	48°, 10', 00"	30°, 26', 00"	1560
Deirifarom	Arvand-Khoramshahr	48°, 11', 09"	30°, 22', 53"	4
Nahre Ghasebe	Nahre	48°, 29', 00"	30°, 01', 00"	1000

In both experiments, 75% of the datasets were used for training the algorithms and the remaining 25% was applied for validation. The performance of the algorithms was measured by means of different validity measures including correlation coefficient (R), root mean squared error (RMSE), and mean absolute error (MAE). These error criteria are used frequently in the literature for evaluation of environmental processes (e.g., Ahmadianfar et al., 2020; Jamei & Ahmadianfar, 2020; Jamei et al., 2020; Najafzadeh & Ghaemi, 2019; Pourrajab et al., 2020). In addition to these validity measures, some recent statistical measures were used, namely uncertainty at 95% confidence level (denoted as U95), reliability, and resilience (Zhou et al., 2017).

Framework of Random Forest Regression

The RFR algorithm acts on the basis of assembling the tree-like structure. It is capable of establishing congruous formulation among a set of input–output variables (Jamei et al., 2021). Generally, it can be noted that RFR model increases different decision trees (DTs), which are learned by means of a sample of input datasets that are bootstrappable. The final output vector of RFR model is calculated by taking the average of these prediction trees. According to Svetnik et al. (2003) research, RFR is a triple step model. At first, X matrix is defined as the training datasets which has N samples. Afterward, k samples are selected randomly by using the bootstrap resampling approach in order to generate k regression trees. In this stage, probability values pertaining to those samples that were excluded (P) are calculated as (Jamei et al., 2021):

$$P = \left(1 - \frac{1}{N}\right)^N \quad (14)$$

Based on Eq. (14), if N has infinite value, the probability will become approximately 37% of the main training datasets that are not taken into account to be drawn, being introduced as out-of-bag datasets and considered for the performance of testing stages. In the second stage of RFR development, regression trees (RTs) which were not pruned, k bootstrapped data samples are generated. As a tree is grown structurally, in each node, an input variable (attribute) is selected randomly from all input variables (A), introduced as internal nodes. Thus, a minimum Gini index is applied to measure how each attribute has a contribution to evaluating elements of tree structures (i.e., nodes and leaves). In this way, the optimum input variable is defined by a splitting variable in order to generate the branches hierarchy. Through the last phase of RFR development, the final model is composed of extracted k regression trees. Fundamentally, there are two statistical measures, introduced as the coefficient of determination and mean squared error, to assess the accuracy level of RFR.

RESULTS AND DISCUSSION

Results of the First Experiment

Table 4 presents the quantitative performance of the MKSVR algorithm developed by linear and nonlinear combination functions for the estimation of BOD and COD.

Table 3. Statistical properties of water quality parameters in the Karun River

Parameter	Unit	Max	Min	Average	Std. Dev	Skewness
Ca ²⁺	mg/l	58.4	1	12.47	9.20	1.92
Na ⁺	mg/l	40	1.42	18.48	7.50	0.144
Mg ²⁺	mg/l	60	2.1	13.21	11.74	2.038
NO ₂ ⁻	mg/l	2.1	0.08	0.41	0.31	1.88
NO ₃ ⁻	mg/l	2.7	0.34	1.01	0.38	1.074
PO ₄ ³⁻	mg/l	3.21	0.13	1.21	0.78	0.701
EC	Decisiemens per meter	9.26	1.7	4.72	1.62	0.743
pH	-	8.71	5.1	7.15	0.77	-0.294
Turbidity	Nephelometric turbidity units	25	1	7.10	5.30	1.051
BOD	mg/l	40.6	3.7	19.21	10.28	0.421
COD	mg/l	34.2	1.06	15.86	9.30	0.381

Table 4. Accuracy level of the MKSVR algorithms in the estimation of the WQP parameters

WQP	Combination function	Validity measures					
		R	RMSE	MAE	U95	Reliability (%)	Resilience (%)
BOD	Linear	0.79	4.87	3.80	22.49	60	55
	Nonlinear	0.80	4.76	3.73	22.40	64	55
COD	Linear	0.71	4.75	3.35	19.14	56	54.54
	Nonlinear	0.72	4.59	3.57	18.98	54	52.17

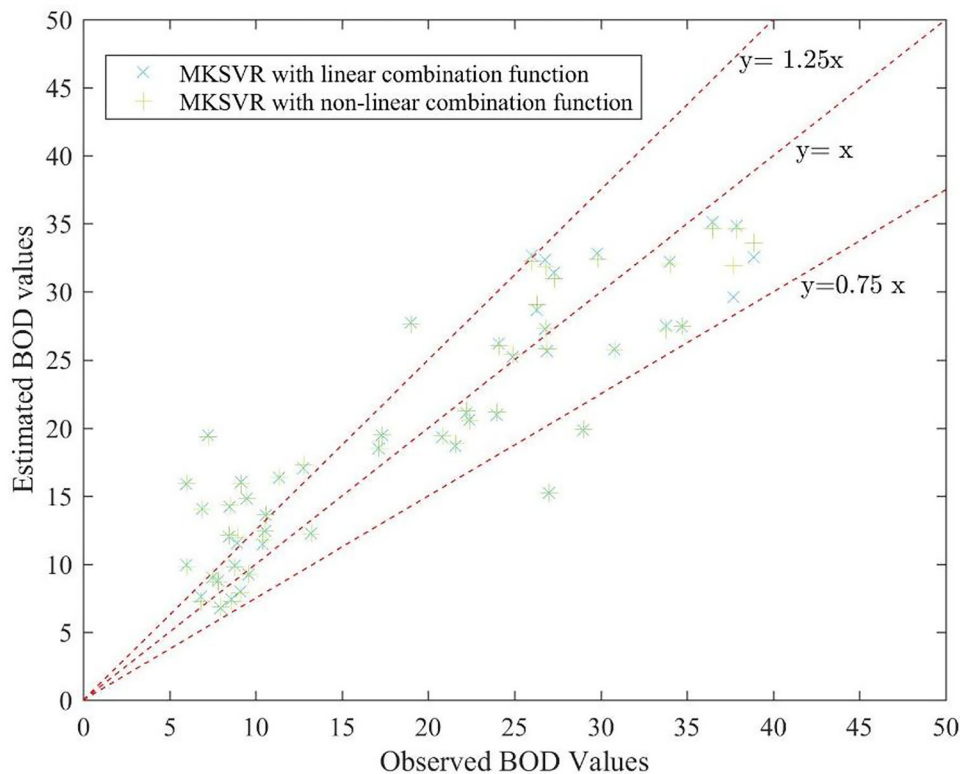


Figure 4. Scatter plot of observed BOD values versus estimated ones by MKSVR.

Analysis of accuracy level obtained using the MKSVR algorithm shows that this algorithm yielded acceptable performances in estimating both BOD and COD. However, marginally better performances were obtained in the case of using the non-linear combination functions. This is because, due to their flexibility, the nonlinear combination functions can better model the underlying structure of data. These results were also in line with the results obtained by Cortes et al. (2009), where nonlinear combination of kernels was used for classification algorithms. However, it should be noted that linear combination functions are less computationally complex than the nonlinear ones, so their adoption will reduce the computational complexity of the algorithm.

The performance of the MKL algorithms depends highly on their optimization strategy by which they find the optimal combination of the kernels. The good performance of the MKSVR algorithm using the proposed optimization strategy can substantiate the effectiveness of PSO as the optimization algorithm. Figures 4 and 5 show the qualitative

comparison of the MKSVR algorithm for BOD and COD, respectively. For BOD measured values between 5 and 10 mg/l, Fig. 4 indicates that some estimated values of BOD were placed out of the $\pm 25\%$ allowable error range. As shown in Fig. 5, both MKSVR techniques had over-predicted slightly COD values between 2 and 5 mg/l; for COD = 5–7 mg/l, all the models have shown a remarkable over-estimation for some measured COD values.

Comparison to Other Regression Algorithms

Table 5 summarizes the obtained accuracies of the regression algorithms used as comparison benchmarks.

As observed from the results, the RMSE values of the SVR algorithm for estimation of COD and BOD in the case of adopting polynomial kernel, respectively, were 5.79 and 6.32, respectively. However, these values reduced to 4.85 and 5.97 when the RBF was adopted as the kernel function. The higher performance of the RBF kernel in comparison with

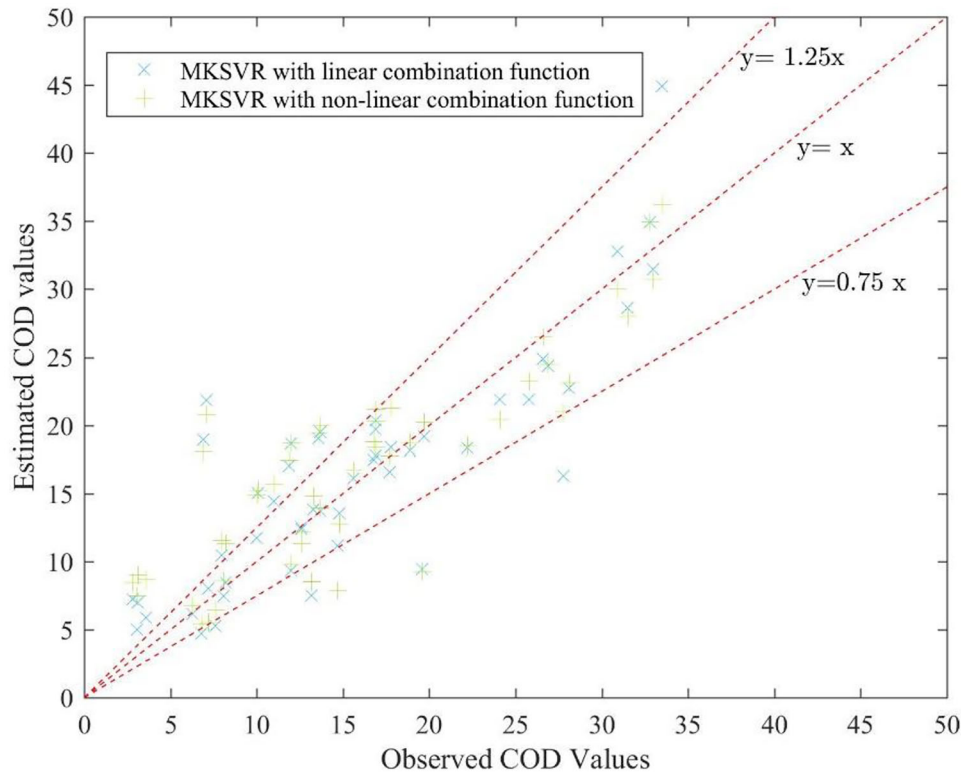


Figure 5. Scatter plot of observed COD values versus estimated ones by MKSVR.

the other kernel is due to its higher ability to characterize the data. Besides, the RBF kernel function also has fewer numerical difficulties, so it is a more appropriate choice than the polynomial kernel for the SVR algorithm. Comparison of the performances of the MKSVR, RFR, and SVR algorithms shows that, in most cases, the MKSVR algorithm outperformed the other algorithms. As an example, the RMSE of the MKSVR considering the nonlinear combination function for estimating the BOD was 4.76 mg/l, while for SVR using RBF kernel function

and for the RFR algorithm the RMSEs were 5.97 mg/l and 5.15 mg/l, respectively. The worst performance of SVR compared to the MKSVR is due to the fact that using a single kernel cannot guarantee to obtain the best model for data under consideration. However, different kernel functions can provide different modeling of the data, and their combination can lead to the best possible data model.

Figures 6 and 7 illustrate the qualitative performance of SVR and RFR models for both BOD

Table 5. Performance of benchmark regression algorithms

WQI	Algorithm	validity measures					
		R	RMSE	MAE	U95	Reliability (%)	Resilience (%)
BOD	SVR (RBF kernel)	0.68	5.97	4.55	23.49	46	48.15
	SVR (Polynomial kernel)	0.63	6.32	4.95	23.84	46	40.74
	RFR	0.77	5.15	4.28	22.74	54	43.48
COD	SVR (RBF kernel)	0.70	4.85	3.43	19.23	58	47.62
	SVR (Polynomial kernel)	0.65	5.79	4.71	20.21	36	34.37
	RFR	0.72	4.63	3.59	19.02	56	40.9

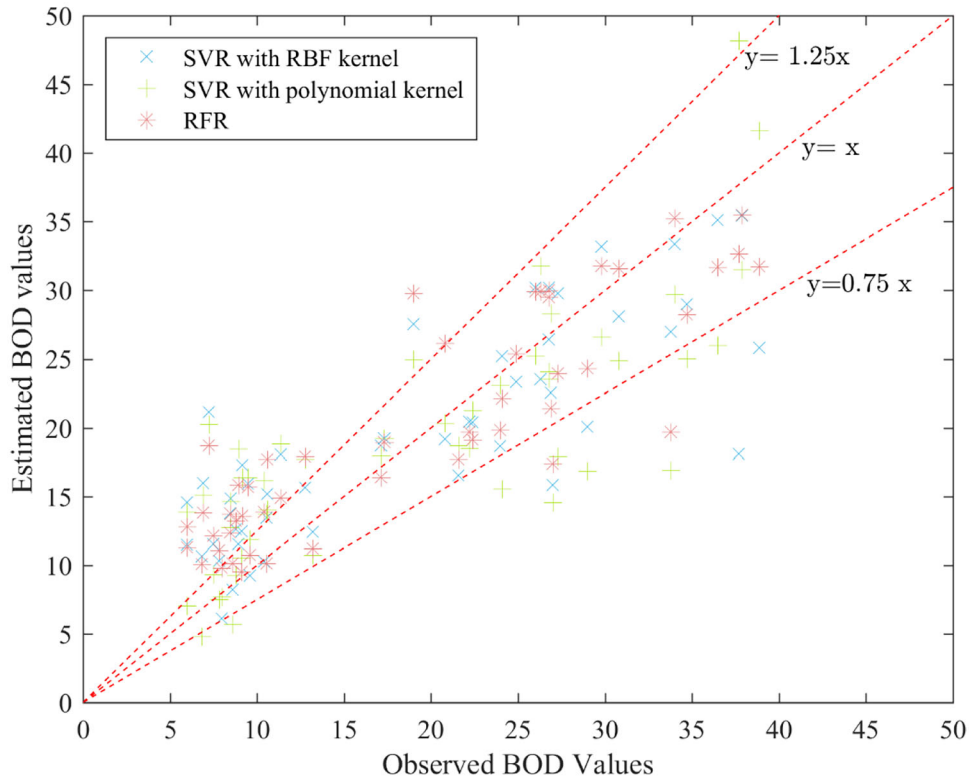


Figure 6. Scatter plot of observed BOD values versus estimated ones by SVR and RFR models.

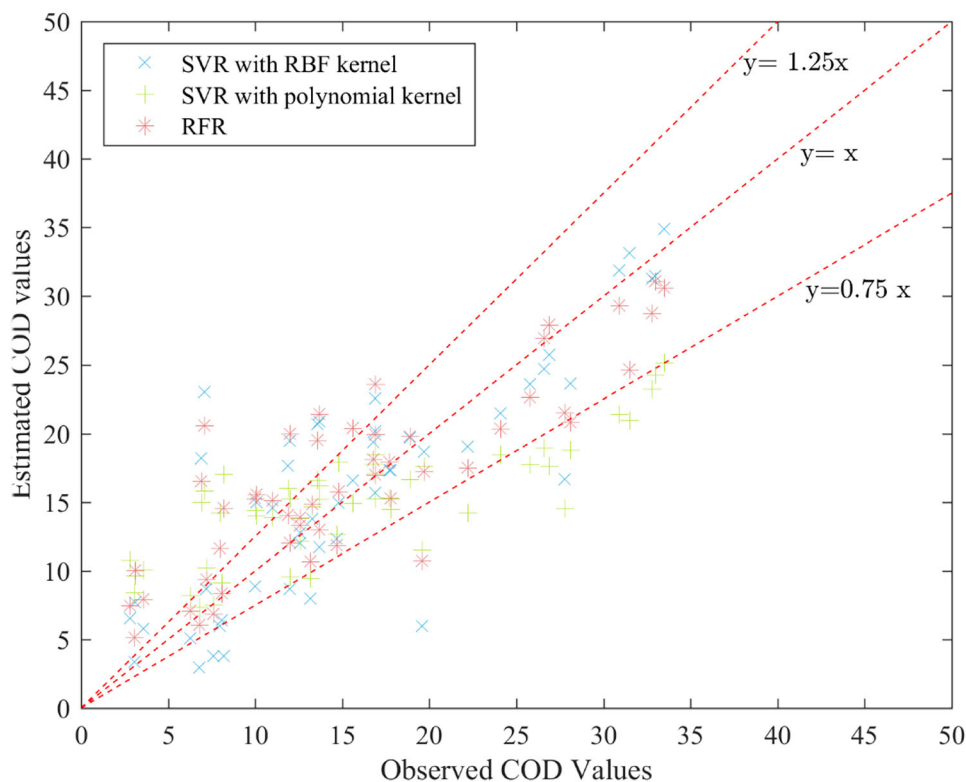


Figure 7. Scatter plot of observed COD values versus estimated ones by SVR and RFR models.

and COD indices, respectively. From Fig. 6, for BOD of 5–10 mg/l, the two SVR models and the RFR technique indicated relatively high over-estimation and consequently placing out-of-error bound. Additionally, for BOD of 25–40 mg/l, there was slight under-estimation of BOD. In Fig. 7, for COD of 2.5–15 mg/l, all the three models relatively over-estimated COD. SVR with polynomial kernel has under-estimated COD remarkably when COD was between 25 and 35 mg/l.

CONCLUSIONS

In this paper, regression analysis was used to obtain hard-to-estimate water quality parameters such as BOD and COD, as a function of other easy-to-measure field parameters. In this way, various regression algorithms in terms of a newly improved SVR model were considered. The performance of the SVR algorithm depended on the mathematical structures of the kernel function. To address issues associated with various types of kernel selection,

novel multiple-kernel support vector regression (MKSVR) algorithms were proposed. These algorithms are capable of learning an optimal kernel through linear or a nonlinear combination of some precomputed basis kernels. From this study, the following conclusions have been drawn.

- Using the SVR algorithm, both BOD and COD were estimated using other water quality parameters with acceptable accuracy.
- The performance of the SVR algorithm was highly dependent on the kernel function and on fine-tuning its parameters. Additionally, the RBF kernel applied in the SVR algorithm yielded better results than those obtained by the second-order polynomial kernel and RFR model.
- The MKSVR algorithm could increase the performance of the SVR algorithm for the estimation of BOD and COD. For this algorithm, the nonlinear combination functions yielded better performance than the linear ones.

For future studies, it is highly recommended to study the role of different setting parameters of SVR and MKSVR models on the accurate estimation of BOD and COD. Although the second-degree polynomial was used to construct the nonlinear combination of the kernels in this research, there is a strong need to employ a higher degree of polynomial kernel in the topology design of the MKSVR algorithm.

REFERENCES

- Abbasnejad, M. E., Ramachandram, D., & Mandava, R. (2012). A survey of the state of the art in learning the kernels. *Knowledge and Information Systems*, 31(2), 193–221.
- Ahmadianfar, I., Jamei, M., & Chu, X. (2020). A novel hybrid wavelet-locally weighted linear regression (W-LWLR) model for electrical conductivity (EC) prediction in surface water. *Journal of Contaminant Hydrology*, 232, 103641.
- Ay, M., & Kisi, O. (2011). Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado. *Journal of Environmental Engineering*, 138(6), 654–662.
- Bansal, J. C., Singh, P., Saraswat, M., Verma, A., Jadon, S. S., & Abraham, A. (2011). Inertia weight strategies in particle swarm optimization. In *11 third world congress on nature and biologically inspired computing (NaBIC)s* (pp. 633–640) IEEE.
- Bozorg-Haddad, O., Soleimani, S., & Loáiciga, H. A. (2017). Modeling water-quality parameters using genetic algorithm-least squares support vector regression and genetic programming. *Journal of Environmental Engineering*, 143(7), 04017021.
- Bucak, S. S., Jin, R., & Jain, A. K. (2014). Multiple kernel learning for visual object recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 1354–1369.
- Camps-Valls, G., Bruzzone, L., Rojo-Álvarez, J., & Melgani, F. (2006). Robust support vector regression for biophysical variable estimation from remotely sensed images. *IEEE Geoscience and Remote Sensing Letters*, 3(3), 339–344.
- Cortes, C., Mohri, M., & Rostamizadeh, A. (2009). Learning nonlinear combinations of kernels. In *Advances in neural information processing systems* (pp. 396–404).
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Emamgholizadeh, S., Kashi, H., Marofpoor, I., & Zalaghi, E. (2014). Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *International Journal of Environmental Science and Technology*, 11(3), 645–656.
- Gonen, M., & Alpaydin, E. (2010). Localized multiple kernel regression. In *2010 20th international conference on pattern recognition* (pp. 1425–1428) IEEE.
- Gönen, M., & Alpaydin, E. (2011). Multiple kernel learning algorithms. *Journal of Machine Learning Research*, 12, 2211–2268.
- Gunn, S. R. (1998). Support vector machines for classification and regression. *ISIS Technical Report*, 14(1), 5–16.
- Heddad, S., & Kisi, O. (2018). Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology*, 559, 499–509.
- Jamei, M., Ahmadianfar, I., Chu, X., & Yaseen, Z. M. (2020). Prediction of surface water total dissolved solids using hybrid wavelet-multigene genetic programming: New approach. *Journal of Hydrology*, 589, 125335.
- Jamei, M., & Ahmadianfar, I. (2020). A rigorous model for prediction of viscosity of oil-based hybrid nanofluids. *Physica A Statistical Mechanics and its Applications*, 556, 124827.
- Jamei, M., Ahmadianfar, I., Olumegbon, I. A., Karbasi, M., & Asadi, A. (2021). On the assessment of specific heat capacity of nanofluids for solar energy applications: Application of Gaussian process regression (GPR) approach. *Journal of Energy Storage*, 33, 102067.
- Kloft, M., Brefeld, U., Sonnenburg, S., & Zien, A. (2011). Lp-norm multiple kernel learning. *The Journal of Machine Learning Research*, 12, 953–997.
- Li, X., Sha, J., & Wang, Z.-L. (2017). A comparative study of multiple linear regression, artificial neural network and support vector machine for the prediction of dissolved oxygen. *Hydrology Research*, 48(5), 1214–1225.
- Mukherjee, S., Osuna, E., & Girosi, F. (1997). Nonlinear prediction of chaotic time series using support vector machines. In *Neural networks for signal processing [1997] VII. Proceedings of the 1997 IEEE workshop* (pp. 511–520) IEEE.
- Najafzadeh, M., & Ghaemi, A. (2019). Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods. *Environmental Monitoring and Assessment*, 191(6), 380.
- Najafzadeh, M., Ghaemi, A., & Emamgholizadeh, S. (2018). Prediction of water quality parameters using evolutionary computing-based formulations. *International Journal of Environmental Science and Technology*, 16(10), 6377–6396.
- Niazmardi, S., Demir, B., Bruzzone, L., Safari, A., & Homayouni, S. (2016). A comparative study on Multiple Kernel Learning for remote sensing image classification. In *2016 IEEE international geoscience and remote sensing symposium (IGARSS)* (pp. 1512–1515) IEEE.
- Niazmardi, S., Demir, B., Bruzzone, L., Safari, A., & Homayouni, S. (2018). Multiple kernel learning for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(3), 1425–1443. <https://doi.org/10.1109/TGRS.2017.2762597>.
- Niazmardi, S., Shang, J., McNairn, H., & Homayouni, S. (2013). A new classification method based on the support vector regression of NDVI time series for agricultural crop mapping. In *2013 second international conference on agro-geoinformatics (Agro-Geoinformatics)* (pp. 361–364) IEEE.
- Pourrajab, R., Ahmadianfar, I., Jamei, M., & Behbahani, M. (2020). A meticulous intelligent approach to predict thermal conductivity ratio of hybrid nanofluids for heat transfer applications. *Journal of Thermal Analysis and Calorimetry*. <https://doi.org/10.1007/s10973-020-10047-9>.
- Qiu, S., & Lane, T. (2005). Multiple kernel learning for support vector regression. *Computer Science Department, The University of New Mexico, Albuquerque, NM, USA, Technical Report* (p. 1).
- Qiu, S., & Lane, T. (2009). A framework for multiple kernel support vector regression and its applications to siRNA efficacy prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 6(2), 190–199.
- Rakotomamonjy, A., Bach, F., Canu, S., & Grandvalet, Y. (2008). SimpleMKL. *Journal of Machine Learning Research*, 9, 2491–2521.
- Rojo-Álvarez, J. L., Muñoz-Marí, J., Camps-Valls, G., & Martínez-Ramón, M. (2018). *Digital signal processing with Kernel methods*. Wiley.

- Scholkopf, B., & Smola, A. J. (2001). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press.
- Sengupta, S., Basak, S., & Peters, R. A. (2019). Particle Swarm optimization: A survey of historical and recent developments with hybridization perspectives. *Machine Learning and Knowledge Extraction, 1*(1), 157–191.
- Shi, Y., & Eberhart, R. C. (1998). Parameter selection in particle swarm optimization. In *International conference on evolutionary programming* (pp. 591–600).
- Singh, K. P., Basant, A., Malik, A., & Jain, G. (2009). Artificial neural network modeling of the river water quality—A case study. *Ecological Modelling, 220*(6), 888–895.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*(3), 199–222.
- Soltani, F., Kerachian, R., & Shirangi, E. (2010). Developing operating rules for reservoirs considering the water quality issues: Application of ANFIS-based surrogate models. *Expert Systems with Applications, 37*(9), 6639–6645.
- Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences, 43*(6), 1947–1958. <https://doi.org/10.1021/ci034160g>.
- Trelea, I. C. (2003). The particle swarm optimization algorithm: Convergence analysis and parameter selection. *Information Processing Letters, 85*(6), 317–325.
- Tuia, D., Verrelst, J., Alonso, L., Pérez-Cruz, F., & Camps-Valls, G. (2011). Multioutput support vector regression for remote sensing biophysical parameter estimation. *IEEE Geoscience and Remote Sensing Letters, 8*(4), 804–808.
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer.
- Varma, M., & Babu, B. R. (2009). More generality in efficient multiple kernel learning. In *Proceedings of the 26th annual international conference on machine learning* (pp. 1065–1072) ACM.
- Verma, A., & Singh, T. (2013). Prediction of water quality from simple field parameters. *Environmental Earth Sciences, 69*(3), 821–829.
- Wu, C.-H., Ho, J.-M., & Lee, D.-T. (2004). Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems, 5*(4), 276–281.
- Yeh, C.-Y., Huang, C.-W., & Lee, S.-J. (2011). A multiple-kernel support vector regression approach for stock market price forecasting. *Expert Systems with Applications, 38*(3), 2177–2186.
- Yu, P.-S., Chen, S.-T., & Chang, I.-F. (2006). Support vector regression for real-time flood stage forecasting. *Journal of Hydrology, 328*(3–4), 704–716.
- Zhou, Y., Chang, F.-J., Guo, S., Ba, H., & He, S. (2017). A robust recurrent anfis for modeling multi-step-ahead flood forecast of three gorges reservoir in the yangtze river. *Hydrology and Earth System Sciences Discuss, 5*, 1–29.