



Development and evaluation of an innovative Enhanced River Pollution Index model for holistic monitoring and management of river water quality

Suyog Gupta¹ · Sunil Kumar Gupta¹

Received: 15 September 2020 / Accepted: 12 January 2021 / Published online: 27 January 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

The present study was conceptualized to develop the Enhanced River Pollution Index (ERPI) model. The ERPI model was used to evaluate the river water quality (RWQ) for its beneficial usage, i.e., drinking with (DCD) and without (DD) conventional treatment, outdoor-bathing (OB), wildlife and fisheries (WF), and industrial and irrigation (IIW). The adequacy of multiple linear regression (MLR) and support vector regression (SVR) models was also investigated to predict the ERPI for estimating the RWQ. The accuracy of the MLR and SVR models was tested by using the statistical parameters, i.e., root mean squared error (RMSE), coefficient of determination (R^2), and mean absolute error (MAE). The results revealed that the MLR models performed well (RMSE = 0.004 ± 0.0043 , $R^2 = 0.998 \pm 0.001$, and MAE = 0.002 ± 0.003) for the DD, DCD, and OB. However, the SVR models estimated the RWQ more accurately (RMSE = 0.041 ± 0.001 , $R^2 = 0.962 \pm 0.010$, and MAE = 0.026 ± 0.002) than the MLR models for WF and IIW. Moreover, this study disclosed that the RWQ was not excellent for DD, OB, and DCD. However, the RWQ was categorized from excellent to poor classes for WF, while it was suitable for IIW.

Keywords River water quality · Enhanced river pollution index · Multiple linear regression · Support vector regression · Kernel function · Monitoring · Modelling

Introduction

River water quality (RWQ) is an extremely delicate and fundamental issue in numerous nations. Likewise, it is a great need to evaluate and describe the expanded comprehension of the consequentiality of RWQ for the purport of drinking, bathing, wildlife, fisheries, irrigation, and industrial usage. The investigation includes water quality (WQ) data to demonstrate the absolute impact of natural factors on surface water, which gives the information of its quality (Ewaid et al. 2018; Sotomayor et al. 2018; Bhatti et al. 2019). The river water is as yet utilized for domestic

and industrial purposes (Fathi et al. 2018). The water nature of the river under mundane conditions is affected by sundry variables, i.e., geography, topography, atmosphere, populace, and anthropogenic elements. Other human impedances are the development of dams and repositories, channelization, industrialization, urban spread, and land use advancements all through the river basin (Wang et al. 2013; Zhang et al. 2019). Anthropogenic activities and natural procedures destruct river water and impede their utilization for agribusiness, drinking, regalement, and other different purposes (Mukate et al. 2019; Verma et al. 2019).

So, the WQ observing system is a fundamental requisite to water assets. The water quality index (WQI) approach has been applied for assessing the WQ of the surface as well as the ground water sources in the worldwide since the last few decades (Effendi 2016; Bhutiani et al. 2016; Bora and Goswami 2017; Ewaid et al. 2018; Verma et al. 2019). The principle motivation behind building up a WQI is to change an intricate arrangement of WQ data into clear and usable information by which a layperson can identify the status of the water source (Abbasi and Abbasi 2012; Şener et al. 2017; Ewaid et al. 2018; Verma et al. 2019). WQI aims

Responsible editor: Xianliang Yi

✉ Sunil Kumar Gupta
sunil@iitism.ac.in

Suyog Gupta
suyog.iitdh@gmail.com

¹ Department of Environmental Science & Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad 826004, India

to minimize the vast datasets to a great extent (Effendi 2016) and simplifies the interpretation of WQ for several purposes like drinking, irrigation, and aquaculture (Abbasi and Abbasi 2012).

The WQI is still utilizable to exhibit the quality of the river basin for low-cost water quality management (Wu et al. 2018; Tian et al. 2019; Tripathi and Singal 2019; Banda and Kumarasamy 2020). Several indices and modelling approaches were introduced to evaluate the status of RWQ in recent years (Sahoo and Jha 2013; Effendi et al. 2015; Effendi 2016; Bora and Goswami 2017; Şener et al. 2017; Ewaid et al. 2018; Kadam et al. 2019; Nayak et al. 2020). Effendi (2016) applied the pollution index and WQI to evaluate the WQ of the river Ciambulawung. Goher et al. (2014) used the weighted arithmetic method based WQI to evaluate the WQ of the Ismailia Canal for drinking, irrigation, and aquatic life. Şener et al. (2017) assessed the WQ of the river Aksu, Turkey, using the WQI and GIS methods. Verma et al. (2019) developed some simplified WQIs for the assessment of spatial and temporal variations in WQ of the river Damodar, India. Chen et al. (2019) employed the monthly river pollution index distributions on the highly urbanized Danshui River basin for sustainable and recreational management.

Due to some difficulties in dealing with the complexities involved in the WQI approach, a strong need for a more straightforward and precise modelling procedure for predicting the RWQ (Li et al. 2016; Rajaei et al. 2018; Leong et al. 2019). Therefore, some researchers used modest and upfront statistical and soft computing approaches, i.e., regression models for establishing a strong relationship between the dependent and independent variables. In recent years, regression models were effectively employed in the domain of water resources for modelling a wide range of hydrological processes, e.g., ground water level (Sahoo et al. 2015; Kommineni et al. 2020), water temperature (Rehana 2019); stream flows (Adnan et al. 2017, 2020), evapotranspiration (Tabari et al. 2012; Kundu et al. 2017), flood prediction (Mosavi et al. 2018; Bafitlhile and Li 2019), and rainfall-runoff (Granata et al. 2016; Sedighi et al. 2016).

The anterior studies were mainly focused on the development of the WQI for drinking purpose only (Şener et al. 2017; Barakat et al. 2018; Ewaid et al. 2018; Tang et al. 2019). These studies enumerated the suitability of the river water without considering the aptness of the river stretches for other beneficial purposes, i.e., OB, WF and IIW. Moreover, the river pollution indices proposed in previous studies (Wang et al. 2013; Hoseinzadeh et al. 2015; Alphayo and Sharma 2018) considered very few river water quality parameters (RWQPs), i.e., ammonia nitrogen ($\text{NH}_3\text{-N}$), dissolved oxygen (DO), biochemical oxygen demand (BOD_5), and suspended solids (SS). But looking into the wide spectrum of its utilization, there was a need to reconsider the RWQ classification in the light of other essential WQPs correlating their potential

cause of river pollution complications and concerns. The high concentration of fluoride (F^-), nitrate (NO_3^-), sulfate (SO_4^{2-}), total coliform (TC), and heavy metals are harmful to both humans and wildlife (Tchounwou et al. 2012). While chloride (Cl^-), electrical conductivity (EC), total dissolved solids (TDS), sodium absorption ratio (SAR), and pH are the critical parameters for irrigation and industrial usage (Zahedi 2017) and high concentrations corrode metals and affect the taste of food products. For wildlife and fisheries; pH, EC, free ammonia (FA), and DO also play a very significant role. High concentration of pH, EC, and FA kills fishes and decreases the species diversity. However, the high concentration of DO is desirable for healthy survival of aquatic life and indicates a good health of the river (EPA 2012). In view of a specific parameter importance and the limitations of the previous studies, the present study incorporated the above mentioned critical RWQPs while deciding the suitability of RWQ for different purposes, i.e., DD, OB, DCD, WF, and IIW, because the criteria of the RWQPs vary for different practices (Leong et al. 2019). Moreover, a strong need of a straightforward and precise modelling procedure was observed for predicting the RWQ (Li et al. 2016; Rajaei et al. 2018; Leong et al. 2019). Therefore, the MLR and SVR modelling techniques were also employed to simplify the complex calculations involved in the ERPI models.

The purpose of this study was to develop cost-effective rapid models to evaluate RWQ by considering the specific RWQPs. The principal targets of this study were (i) to develop and evaluate the ERPI model to investigate the RWQ, (ii) to classify the RWQ for different usage, (iii) to develop MLR and SVR models against the ERPI models as reference for the RWQ modelling, and (iv) to compare the performance of the MLR and SVR models.

Methodology

Study area

In this study area encompasses the river Damodar, situated in the Damodar river basin (DRB), India (Fig. 1). The river originates from the Khamerpet hill and flows from Jharkhand and meets with the river Hoogli in West Bengal. This is a shallow, wide, and flashy rain-fed river. The full stretch and the catchment area of this river are approximately 541 km and 23,170 sq. km, respectively. It traverses through the steep slope of the pat region in its upper reaches to descend on the gneissic flat plain of Chandwa, and flow of the river becomes sluggish over the flattop surface. The mean discharge and annual runoff were observed as 296 m^3/s and 486 mm/year at Rhondia station. The physiography of upper catchment of DRB is quite different from the lower part as the different rock types, i.e.,

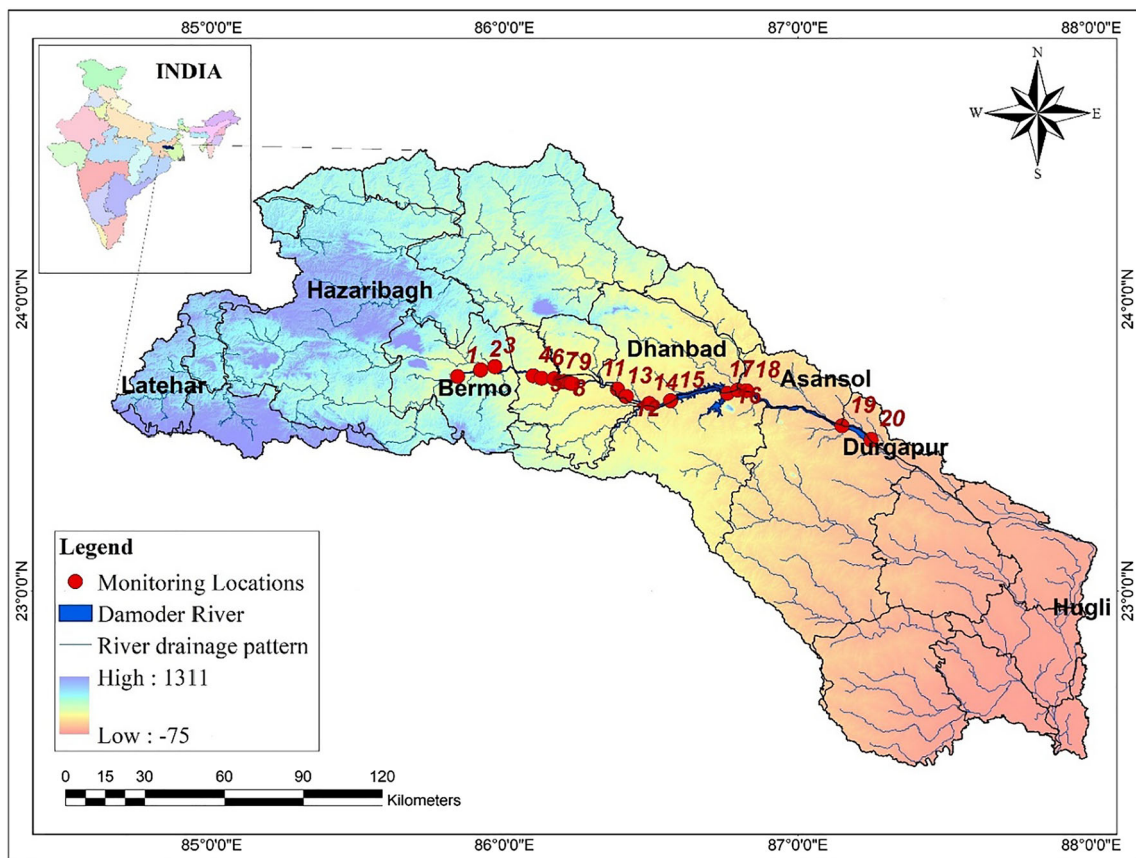


Fig. 1 Geographical location of the study area

igneous, sedimentary, and metamorphic rocks were found in different geological time scale. DRB is gifted with mineral resources of coal. It falls within dry and subhumid climatic zones and usually experiences a very hot and dry summer. The average temperature is of 30 °C, and it rises to 48 °C during the months of May–July. Winter is cold with temperature as low as 2 °C. The average annual precipitation of around 1350 mm. More than 80% of the total rainfall ensues during monsoon season between June and September months.

This river is not only the source of drinking water but also accomplishes the water necessity of irrigation and industrial activities at the region. The industrial activities consist of six steel power stations, four thermal power plants, and three hydroelectric power stations (Kumar et al. 2019; Verma et al. 2019). These industries influence the hydrological regime of the river by withdrawing a lot of water for their accompanying activities. These industries also discharge the substantial amount of effluent containing pollutants, e.g., heavy metals, fly ash, coal dust, and suspended solids, directly into the river, which deteriorate the RWQ. Besides industrial activities, urbanization and heavy encroachment at the bank of the river affect the RWQ and quantity (Mukherjee et al. 2012; Haldar et al. 2014; Verma et al. 2019).

River water sampling

The sampling was performed for the period of 2017–2019 during premonsoon, monsoon, and postmonsoon seasons at the selected monitoring locations on the river stretch. Three numbers of samples were collected from a single location thrice in a season, and the average of the same was reported. Twenty monitoring locations in the river Damodar stretch were carefully selected with consideration of the guidelines for water quality monitoring given by Central Pollution Control Board (CPCB), India (CPCB 2007). All water samples were stored in an insulated cool box together with cold packs and sent to the laboratory. At laboratory, water samples were immediately transferred to the refrigerator for further analysis.

Analytical methods

Fourteen RWQPs, i.e., pH, EC, BOD₅, DO, TDS, Cl⁻, TC, SAR, NO₃⁻, SO₄²⁻, F⁻, FA, Fe, and Pb were analyzed by considering the standard methods prescribed in the guidelines, published by American Public Health Association (Baird et al. 2017). pH and EC were recorded in-situ using a pH meter (Hanna® HI98107) and conductivity meter (HACH®

HQ40D multiparameter), respectively. BOD₅ was assayed by using the 5-day BOD test. DO was determined using the Winkler method. Estimation of TDS was done by gravimetric analysis. Cl⁻ was measured using the argentometric method. TC was examined by the multiple tube fermentation method. Sodium and calcium were estimated using the flame photometer (Systronics Flame Photometer 128), while magnesium was evaluated by EDTA titrimetric method to calculate the SAR. NO₃⁻, SO₄²⁻, F⁻, FA, Fe, and Pb were estimated by using an ultraviolet spectrophotometer (MOTRAS Scientific UV-Visible spectrophotometer).

Analytical quality assurance and quality control

As the natural variability is a fundamental feature of a river and cannot be controlled, to quantify this variability triplicate river water samples were collected during the sampling. The analytical data quality and accuracy were ensured through careful standardization by preparing and analyzing the reference water sample for determining the presence of any interference. For the precision of measurement, analysis of the river water sample was performed in triplicate and considered the average as the final value. The instrument was recalibrated when the relative percent difference (RPD) between the two river water samples transcend to $\pm 5\%$. Moreover, the analytical grade chemical reagents were used in the whole analysis procedure of RWQPs. The representativeness of the samples was controlled by selecting the appropriate locations and time for river water sampling.

Data processing

For regression models, the dataset of dependent and independent variables was normalized within a fixed range between 0 and 1, to transform all variables on a uniform scale. Moreover, the dataset was split into the training and testing set as, 70% for the training phase, and 30% for the testing phase (Bozorg-Haddad et al., 2017). The models were developed using training set and then validated by the testing set. The performance of the models was evaluated using the statistical metrics, i.e., RMSE, R^2 , and MAE. The program codes were written in R language using RStudio Desktop version 1.3 software.

Enhanced river pollution index model

The enhanced river pollution index model (ERPI) model was developed and evaluated for the monitoring and management of the RWQ for the specific usage of the river water, i.e., DD, OB, DCD, WF, and IIW, as categorized and described by the CPCB, India (CPCB 1979, 2007; BIS 1982). The ERPI model included the four essential steps. The first step was the selection of crucial RWQPs according to the particular use of the river water. The second step was to determine the relative

weights for the selected RWQPs (Olasoji et al. 2019). The third step was to calculate the subindex for each selected RWQPs. In the fourth step, all of the subindices were aggregated to evaluate the final value of the ERPI model. The ERPI model was described in Eq. (1).

$$ERPI = \sum_{j=1}^m SI_j / \sum_{j=1}^m W_j \quad (1)$$

where, SI_j is the subindex and W_j is the relative weight for j^{th} (1, 2, 3, ..., m) parameter of the river water. The calculation involved in the ERPI model was described in Eqs. (2–5)

$$SI_j = Q_j \times W_j \quad (2)$$

$$Q_j = [(|EV_j - IV_j|) / (SPV_j - IV_j)] \times 100 \quad (3)$$

$$W_j = k / SPV_j \quad (4)$$

$$k = 1 / \sum_{j=1}^m (1 / SPV_j) \quad (5)$$

where, Q_j is the quality rating, EV_j is the estimated value of parameter in river water sample, IV_j is the ideal value of parameter in pure water, SPV_j is the standard permissible value for RWQPs, and k is the constant of proportionality. The different categories for the values of ERPI model with respective RWQ were classified in Table 1 (Tyagi et al. 2013; Bora and Goswami 2017; Hussein and Ali 2017; Trikoilidou and Samiotis 2017; Ustaoğlu et al. 2020).

Multiple linear regression

Multiple linear regression (MLR) is a quantitative tool used for modelling by establishing linear relationship between two or more independent variables and a dependent variable (Tabari et al. 2012; Kadam et al. 2019) and is expressed in the form of Eq. (6).

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m + \varepsilon \quad (6)$$

where, y is the dependent variable, α_0 is the intercept, $\alpha_1 - \alpha_m$ is the regression coefficients, $x_1 - x_m$ is the independent variables, m is the number of independent variables, and ε is the random error. In this study, outcomes of the ERPI models and their corresponding RWQPs were used as dependent and independent variables, respectively, to determine the RWQ by estimating the ERPI for different purposes.

Table 1 Categories for the values of ERPI model

Range	RWQ
0–25	Excellent
26–50	Good
51–75	Fair
76–100	Poor
> 100	Unfit

Support vector regression

Support vector machine (SVM) is a method educed from statistical learning theory and can be used both for classification and regression problems (Tabari et al. 2012; Ji et al. 2017). SVR imprints a linear model to separate the sample dataset from the input vectors through some nonlinear mapping techniques. In SVR, a nonlinear function is erudite a kernel induced feature space by a linear learning machine. The SVR model is trained on dataset $d = \{x_i, y_i; i = 1, 2, \dots, n\}$ with n -dimensional input vectors x_i and associated target y_i . SVR aims to discover a function $f(x)$ with at most error tolerance ε deviation from the target y for all the training datasets (Liu and Lu 2014; Raghavendra and Deka 2014). SVR deliberates the following estimation function, as shown in Eq. (7), to fulfil the aim.

$$f(x) = \omega \varphi(x) + b \tag{7}$$

where $\omega \in d$, d is the input space and b is the bias and $\varphi(x)$ is the high dimensional feature space. These coefficients can be estimated by the regularized risk function (R(f)) minimizing technique using Eqs. (8–9).

$$R(f) = C \sum_{i=1}^n L_\varepsilon(y_i - f(x_i)) + \|\omega\|^2 / 2 \tag{8}$$

$$L_\varepsilon(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon & |y_i - f(x_i)| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where, C is the cost function measuring empirical risk, $L_\varepsilon(f(x_i) - y_i)$ is the ε -insensitive loss function, $\|\omega\|^2 / 2$ is the Euclidean norm, ε is the difference between actual values, and n is the number of variables. Hence, the regression problem can be defined in the form of convex optimization problem and solved using Lagrange function (Raghavendra and Deka 2014). Hence, the regression function is shown in Eq. (10).

$$f(x) = \sum_{i,j=1}^n (\delta_i - \delta_i^*) (x_i, x_j) + b \tag{10}$$

where, $(\delta_i - \delta_i^*)$ is the Lagrange multiplier. The kernel function is involved to solve the nonlinear problems in the SVR models. This function maps the data into higher dimension feature space. The SVR model in the feature space can be expressed using $K(x_i, x_j)$ instead of (x_i, x_j) , then the SVR model can be expressed as Eq. (11).

$$f(x) = \sum_{i,j=1}^m (\delta_i - \delta_i^*) K(x_i, x_j) + b \tag{11}$$

where, $K(x_i, x_j)$ is the kernel function. From Eq. (11), the nonzero Lagrange multiplier data (support vector) is involved in the final SVR model. Finally, the SVR model can be expressed as the regression function given in Eq. (12).

$$f(x) = \sum_{k=1}^m (\delta_k - \delta_k^*) K(x_k, x) + b \tag{12}$$

where, x_k is the support vector and m : number of support vectors. The SVR model can be epitomized as a two-layer network architecture (Fig. 2) in which the weights are nonlinear in the first layer and linear in the second layer.

Performance analysis

In this study, root mean square error (RMSE), coefficient of determination (R^2), and mean absolute error (MAE) were used to compare the performance of the developed models in the estimation of RWQ. These were calculated using Eqs. (13–15), respectively.

$$RMSE = \sqrt{\sum_{i=1}^n (Y_{p_i} - Y_{o_i})^2 / n} \tag{13}$$

$$R^2 = 1 - \left[\frac{\sum_{i=1}^n (Y_{p_i} - Y_{o_i})^2}{\sum_{i=1}^n (Y_{o_i} - \bar{Y})^2} \right] \tag{14}$$

$$MAE = (1/n) \sum_{i=1}^n |(Y_{p_i} - Y_{o_i})| \tag{15}$$

where, n is the number of observations, Y_o is the observed value, Y_p is the predicted value, Y_o is the mean observed value, and Y_p is the mean predicted value of respective ERPI model.

Results and discussion

Characteristics of the river

The overall analytical results of RWQPs were epitomized in Table 2. The results deliberated the alkaline nature of the river Damodar. The increment in the pH was due to industrial effluent and agricultural runoff (Haladar et al. 2014; Verma et al. 2019). The concentration level of DO was found to be sufficient except for a few locations for various physiological activities because of the geological conditions, which increased the level due to high aeration (Mukherjee et al. 2012). TDS value was high in the sites near the small and large industries. The locations had a high concentration of BOD₅ and Cl⁻,

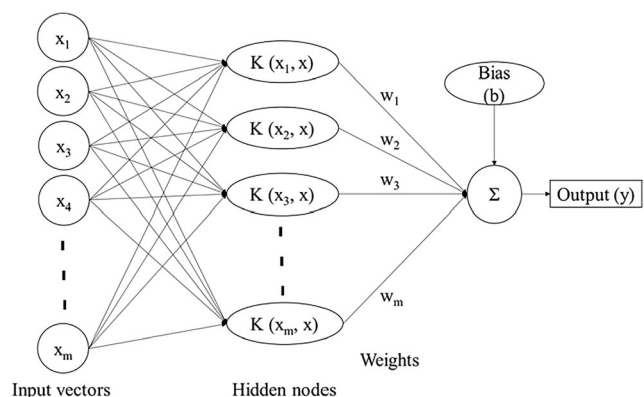


Fig. 2 Network architecture of SVR model

Table 2 Statistical summary of the RWQPs

RWQPs	Minimum	Maximum	Mean	Standard deviation
pH	6.10	9.10	7.67	0.59
BOD ₅ (mg/L)	1.10	17.50	4.82	3.02
DO (mg/L)	3.80	8.80	6.86	0.99
TDS (mg/L)	108	1445	312.28	259.24
Cl ⁻ (mg/L)	7.32	58.90	27.16	10.43
NO ₃ ⁻ (mg/L)	0.22	26.60	11.98	6.33
SO ₄ ²⁻ (mg/L)	10.20	144.20	56.21	36.92
TC (MPN/100 ml)	35	1112	381.24	303.28
F ⁻ (mg/L)	0.23	1.14	0.90	0.07
Fe (mg/L)	0.14	1.10	0.54	0.21
Pb (mg/L)	0.001	0.07	0.02	0.01
EC (μS/cm)	180	2390	497.28	415.68
FA (mg/L)	0.10	1.00	0.31	0.18
SAR	2.05	7.82	5.03	1.30

where the river received the urban waste. The higher BOD₅ level determined the presence of a greater amount of organic matter for the microorganisms due to wastewater discharge on the river stretch (Mukherjee et al. 2012; Tripathi and Singal 2019; Verma et al. 2019). The main contributors of sulfates were mine wastes, sewage treatment plants, industrial discharges, and runoff from agricultural lands (Verma et al. 2019). The maximum concentration of Pb and Fe was found at the locations near the discharge point of the thermal power station and metal industries. The value of TC was found higher than the prescribed limit (BIS 1982) at most of the sites.

Table 3 Spearman's correlation matrix of RWQPs

	pH	BOD	DO	TDS	Cl ⁻	NO ₃ ⁻	SO ₄ ²⁻	TC	F ⁻	Fe	Pb	EC	SAR	FA
pH	1													
BOD ₅	0.094	1												
DO	0.218**	-0.371**	1											
TDS	-0.163**	0.250**	-0.161**	1										
Cl ⁻	0.192**	0.270**	-0.159**	-0.218**	1									
NO ₃ ⁻	0.028	0.317**	-0.334**	0.139*	0.473**	1								
SO ₄ ²⁻	-0.157**	0.292**	-0.513**	0.411**	0.291**	0.429**	1							
TC	-0.022	0.379**	-0.082	0.273**	0.343**	0.148*	0.249**	1						
F ⁻	0.265**	0.223**	0.048	0.090	0.194**	0.151**	-0.028	0.307**	1					
Fe	-0.191**	-0.067	0.079	0.147*	-0.087	-0.167**	-0.158**	-0.085	0.002	1				
Pb	0.206**	0.240**	-0.207**	-0.057	0.354**	0.241**	0.323**	-0.045	0.095	0.050	1			
EC	-0.142*	0.261**	-0.176**	0.983**	-0.171**	0.204**	0.449**	0.279**	0.102	0.130*	-0.010	1		
SAR	0.044	0.058	-0.154**	0.284**	-0.131*	0.021	-0.049	0.090	0.099	0.094	-0.121*	0.261**	1	
FA	-0.030	0.200**	-0.311**	0.539**	-0.115*	0.199**	0.365**	0.151**	0.133*	-0.091	-0.025	0.533**	0.199**	1

*Correlation is significant at the 0.05 level (2-tailed)

**Correlation is significant at the 0.01 level (2-tailed)

As most of the RWQPs were not normally distributed, Spearman's correlation matrix was used to analyze the correlation between the RWQPs as shown in Table 3. The results obtained from Table 3 represented a very strong positive correlation between EC and TDS which indicated the presence of high level of inorganic salts and organic substances in the river water which may be attributed to the domestic, industrial, and agricultural pollutions. Moreover, a negative correlation between DO and BOD₅ was found, as high concentration of BOD₅ depleted the DO level of the river water. A positive correlation between BOD₅ and TC indicated the presence of domestic sources of pollution. This phenomenon designated that the overall RWQ was strongly affected by the domestic wastewater sources and effluents of the coal industries, steel plants, and thermal power stations situated at the Damodar River basin (Mukherjee et al. 2012; Verma et al. 2019).

Evaluation of the RWQ using ERPI model

In this study, WQ of the river Damodar was evaluated for DD, OB, DCD, WF, and IIW purposes. The analysis results of all twenty sampling locations were used for RWQ estimation. To assess RWQ for purposes as mentioned earlier, five ERPI models, i.e., ERPI_{DD}, ERPI_{OB}, ERPI_{DCD}, ERPI_{WF}, and ERPI_{IIW}, were developed with different combinations (CPCB 1979; BIS 1982) of the RWQPs (Table 4).

To calculate the ERPI model values at each sampling location, the relative weights (W_i) for each RWQP, were computed according to their relative importance in the overall RWQ for different purposes (Table 5).

Table 4 Description of categorical ERPI with respective parameters

ERPI	Purpose	Mathematical expression	Input RWQPs
ERPI _{DD}	Drinking after disinfection	$\sum_{j=1}^{11} SI_j / \sum_{j=1}^{11} W_j$	DO, BOD ₅ , Cl ⁻ , F ⁻ , SO ₄ ⁻² , NO ₃ ⁻ , pH, TDS, TC, Fe, Pb
ERPI _{OB}	Outdoor bathing	$\sum_{j=1}^5 SI_j / \sum_{j=1}^5 W_j$	DO, BOD ₅ , F ⁻ , pH, TC
ERPI _{DCD}	Drinking after conventional treatment and disinfection	$\sum_{j=1}^{11} SI_j / \sum_{j=1}^{11} W_j$	DO, BOD ₅ , Cl ⁻ , F ⁻ , SO ₄ ⁻² , NO ₃ ⁻ , pH, TDS, TC, Fe, Pb
ERPI _{WF}	Wildlife & Fisheries	$\sum_{j=1}^4 SI_j / \sum_{j=1}^4 W_j$	DO, FA, pH, EC
ERPI _{IW}	Industrial & irrigation	$\sum_{j=1}^6 SI_j / \sum_{j=1}^6 W_j$	SAR, Cl ⁻ , SO ₄ ⁻² , pH, EC, TDS

The ERPI_{DD} model included eleven RWQPs to estimate the suitability of the river water for DD. The values of the ERPI_{DD} model lied between 55.252 and 122.590. The ERPI_{DCD} model also encompassed eleven RWQPs. The results determined that the values of ERPI_{DCD} model varied from 28.583 to 87.711. The ERPI_{OB} model contained five RWQPs to evaluate the fitness of the river water for OB. The results deliberated that the values of the ERPI_{OB} model were found between 51.814 and 111.047. The ERPI_{WF} model comprised four RWQPs to evaluate the aptness of the RWQ for WF. Values of the ERPI_{WF} model ranged between 17.038 and 82.014. The ERPI_{IW} model involved five RWQPs in estimating the suitability of the river water for IIW. Outcomes of the ERPI_{IW} were estimated between 10.010 and 68.956. Figure 3 showed the RWQ classification, based on evaluation of the ERPI models for a particular use of the river water.

Table 5 Relative weight of each input parameter for different ERPI models

RWQPs ^a	W _j				
	ERPI _{DD}	ERPI _{OB}	ERPI _{DCD}	ERPI _{WF}	ERPI _{IW}
DO	0.01121	0.15156	0.01867	0.20799	-
BOD ₅	0.03364	0.25259	0.02489	-	-
TC	0.00135	0.00152	0.00001	-	-
TDS	0.00013	-	0.00005	-	0.00298
Cl ⁻	0.00027	-	0.00012	-	0.01044
F ⁻	0.04485	0.50519	0.04978	-	-
SO ₄ ²⁻	0.00017	-	0.00019	-	0.00626
NO ₃ ⁻	0.00336	-	0.00149	-	-
pH	0.00792	0.08915	0.00878	0.09788	0.73669
Fe	0.22427	-	0.14934	-	-
Pb	0.67282	-	0.74668	-	-
EC	-	-	-	0.00083	0.00278
FA	-	-	-	0.69330	-
SAR	-	-	-	-	0.24084

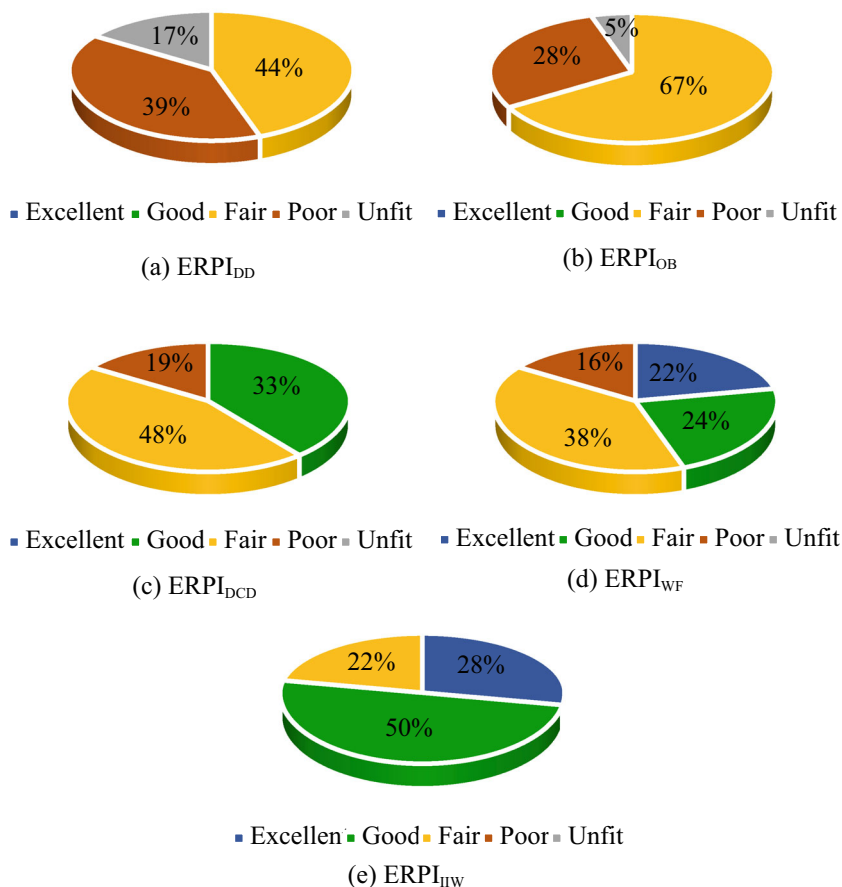
^a RWQPs, not contributed as input parameters were marked “-“ for particular ERPI

Overall, the ERPI was the most delicate model for different usage of the river water over the previously reported river pollution indices (Sahoo et al. 2015; Alphayo and Sharma 2018), which were focused only on drinking purpose by considering fewer RWQPs. ERPI method reflected the capability to overcome the limitations (e.g., parameter restriction and redundancy, lack of portability, and the inability to represent specific uses) of the hitherto approaches. The ERPI model assimilated only necessary RWQPs, i.e., eleven for ERPI_{DD}, five for ERPI_{OB}, eleven for ERPI_{DCD}, four for ERPI_{WF}, and five for ERPI_{IW}. The approach of selecting the more specific and essential RWQPs reduced the time and cost involved in the analytical procedures of the RWQPs. The developed models integrated the composite influence of different RWQPs for explicit purposes for communicating the global RWQ information to the general person as well as the decision-makers. According to the critical observation of the results for ERPI models shown in Fig. 3, ERPI_{DD} revealed that no stretch of the river had fallen under the excellent or good categories (Mukherjee et al. 2012; Singh et al. 2019) and fibbed only between fair to unfit classes for DD. But it could be improved after performing conventional treatments conferring to ERPI_{DCD}. ERPI_{OB} deliberated that most of the river stretch was falling under the fair class for OB, while 33% of the river stretch lied between poor to unfit. In the view of WF, ERPI_{WF} classified 16% of the river stretch in poor class. It was due to the discharge of municipal wastes and effluent from coal washeries, steel plants, and thermal power stations, situated near to the river bank (Mukherjee et al. 2012; Singh et al. 2019; Verma et al. 2019). However, ERPI_{IW} categorized the whole stretch of the river within excellent to good classes for IIW purposes.

Results of MLR model

In the MLR models for DD (MLR_{DD}), OB (MLR_{OB}), DCD (MLR_{DCD}), WF (MLR_{WF}), and IIW (MLR_{IW}), the ERPI_{DD}, ERPI_{OB}, ERPI_{DCD}, ERPI_{WF}, and ERPI_{IW} variables were defined as the dependent variables. Moreover, the corresponding RWQPs, as described in Table 4, were assumed as independent variables for respective models. The reduction in the number of independent variables on the premise of less

Fig. 3 River classification based on ERPI model. (a) ERPI_{DD}, (b) ERPI_{OB}, (c) ERPI_{DCD}, (d) ERPI_{WF}, (e) ERPI_{IW}



influence on the accuracy of the MLR models was comprised to minimize the workload and the information overlapping. The best suited model was derived by testing the numerous blends of independent variables with respective dependent variable. The MLR models were developed using training dataset and then validated by the testing dataset using ‘lm’ function with ‘qr’ method, available in RStudio software. The statistical summary of the derived MLR models was abridged in Table 6.

As the results showed, the MLRDD, MLROB, and MLRDCD models had the best performance. In contrast, the MLR_{WF} and MLRIW models gave the poor RWQ estimates having low R² with high RMSE and MAE. The RWQ for the respective water usage, estimated by the MLR models and RWQ computed using the ERPI models as benchmark were

shown in Fig. 4. The comparative analysis between ERPI and the respective MLR models suggested that the multiple regression techniques can be an excellent way to deal with foresees ERPI for RWQ. Moreover, this sort of approach requires long haul physicochemical information to determine the parameters of the regression model, which are site and season reliant.

Results of SVR models

In the present study, four SVR models with different kernel functions, i.e., linear (LK), polynomial (PK), radial basis (RK), and sigmoid (SK), were developed for each usage of the river water. Overall, twenty SVR models were established as; LK-SVR_{DD}, PK-SVR_{DD}, RK-SVR_{DD}, and SK-SVR_{DD} for DD, LK-SVR_{OB}, PK-SVR_{OB}, RK-SVR_{OB}, and SK-SVR_{OB}

Table 6 Statistical summary of MLR models

Model	RMSE	R ²	MAE	Equation
MLR _{DD}	0.001	0.999	0.001	ERPI _{DD} = - 0.134 + 0.011pH + 0.297BOD ₅ - 0.006DO + 0.002TDS + 0.005NO ₃ ⁻ + 0.029TC + 0.027F ⁻ + 0.629Fe + 0.514Pb
MLR _{OB}	0.009	0.997	0.006	ERPI _{OB} = - 0.120 + 0.078pH + 0.927BOD ₅ - 0.049DO - 0.006TC + 0.190F ⁻
MLR _{DCD}	0.002	0.999	0.001	ERPI _{DCD} = - 0.146 + 0.017pH + 0.199BOD ₅ - 0.012DO + 0.002TDS + 0.001NO ₃ ⁻ - 0.002TC + 0.041F ⁻ + 0.340 Fe + 0.774Pb
MLR _{WF}	0.092	0.757	0.061	ERPI _{WF} = 0.279 + 0.713pH - 0.496DO + 0.079FA
MLR _{IW}	0.127	0.695	0.085	ERPI _{IW} = - 0.110 + 1.028pH - 0.077SO ₄ ²⁻ + 0.168EC - 0.175SAR

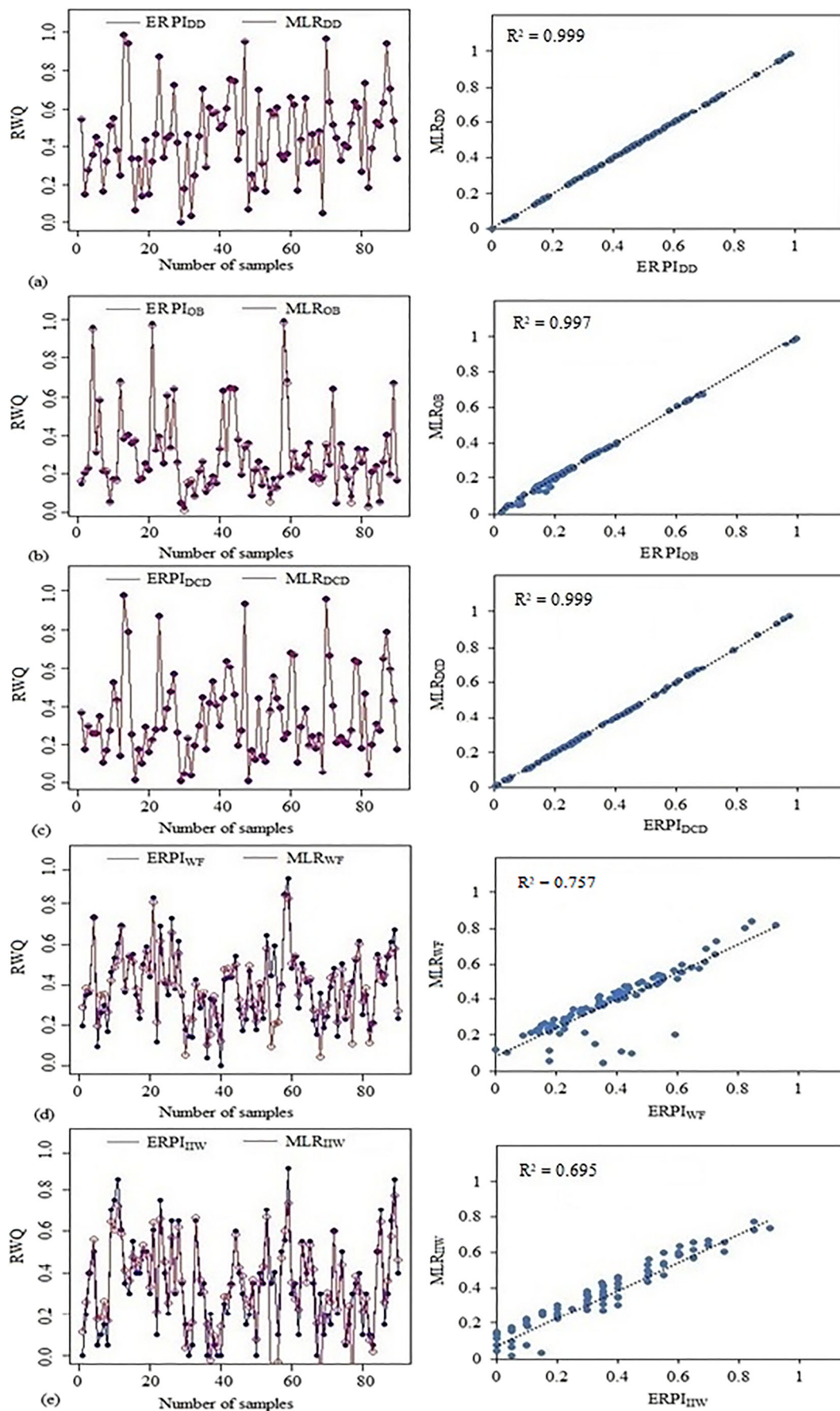


Fig. 4 RWQ estimated by the ERPI models and the MLR models in the testing phase for a DD; b OB; c DCD; d WF, and e IIW

Table 7 Performance of the SVR models for the testing phase

Model	Kernel	RMSE	R^2	MAE
LK-SVR _{DD}	LK	0.012	0.999	0.010
PK-SVR _{DD}	PK	0.058	0.932	0.042
RK-SVR _{DD}	RK	0.029	0.985	0.021
SK-SVR _{DD}	SK	1.030	0.071	0.472
LK-SVR _{OB}	LK	0.010	0.997	0.008
PK-SVR _{OB}	PK	0.102	0.853	0.061
RK-SVR _{OB}	RK	0.072	0.912	0.030
SK-SVR _{OB}	SK	0.804	0.312	0.457
LK-SVR _{DCD}	LK	0.013	0.999	0.011
PK-SVR _{DCD}	PK	0.053	0.944	0.039
RK-SVR _{DCD}	RK	0.033	0.981	0.023
SK-SVR _{DCD}	SK	1.010	0.064	0.455
LK-SVR _{WF}	LK	0.116	0.720	0.047
PK-SVR _{WF}	PK	0.129	0.529	0.098
RK-SVR _{WF}	RK	0.032	0.977	0.021
SK-SVR _{WF}	SK	0.861	0.061	0.504
LK-SVR _{IIW}	LK	0.157	0.688	0.067
PK-SVR _{IIW}	PK	0.145	0.599	0.108
RK-SVR _{IIW}	RK	0.029	0.984	0.020
SK-SVR _{IIW}	SK	1.301	0.042	0.682

for OB, LK-SVR_{DCD}, PK-SVR_{DCD}, RK-SVR_{DCD}, and SK-SVR_{DCD} for DCD, LK-SVR_{WF}, PK-SVR_{WF}, RK-SVR_{WF}, and SK-SVR_{WF} for WF and LK-SVR_{IIW}, PK-SVR_{IIW}, RK-SVR_{IIW}, and SK-SVR_{IIW} for IIW. The input combinations of variables with respective outputs for the SVR models were the same as that of used for the MLR models. Table 7 represented the results of the SVR models for RWQ estimation. It merely elaborated the performance of the developed SVR models with different kernel functions.

As shown, the SVR models were sensitive to the choice of kernel functions (Tabari et al. 2012; Raghavendra and Deka 2014). An appropriate selection of the kernel function allowed the nonseparable RWQ data in the original input space to become separable in the new feature space (Raghavendra and Deka 2014). The comparison of the RWQ values calculated by ERPI models and RWQ values predicted by the four different SVR models was shown in Fig. 5. It was seen from Fig. 5 that the LK-SVR_{DD} model closely followed the ERPI_{DD} model values of RWQ. The same pattern was attained by LK-SVR_{OB} and LK-SVR_{DCD} models for ERPI_{OB} and ERPI_{DCD} values of RWQ, respectively. However, the RK-SVR_{WF} and RK-SVR_{IIW} models were evidenced as the best models for the RWQ estimation for WF and IIW, respectively. Moreover, the PK and RK functions were also found significant for SVR_{DD}, SVR_{OB}, and SVR_{DCD} models. It confirmed the RMSE and MAE statistics given in Table 7.

Comparison of the MLR and SVR models

Table 8 represented the comparative study between the performance of MLR and SVR models for the testing phase. The vigorous MLR and SVR models were selected and ranked according to the RMSE values, for estimating the RWQ for respective usage. The results described that the MLR_{DD} model with RMSE of 0.001 could be designated as the best model for RWQ valuation for DD in the study area. The LK-SVR_{DD} model attained the rank second with the RMSE of 0.012. The RK-SVR_{DD} model with RMSE of 0.028 could be considered as the next best model, successively followed by the PK-SVR_{DD} (RMSE = 0.058) and SK-SVR_{DD} (RMSE = 1.031). Same pattern was found for OB (MLR_{OB} > LK-SVR_{OB} > RK-SVR_{OB} > PK-SVR_{OB} > SK-SVR_{OB}) and DCD (MLR_{DCD} > LK-SVR_{DCD} > RK-SVR_{DCD} > PK-SVR_{DCD} > SK-SVR_{DCD}). The RK-SVR_{WF}, MLR_{WF}, LK-SVR_{WF}, PK-SVR_{WF} and SK-SVR_{WF} models ranked 1st place to 5th, respectively, for RWQ estimation for WF. For IIW, the models followed a similar ranking (RK-SVR_{IIW} > MLR_{IIW} > LK-SVR_{IIW} > PK-SVR_{IIW} > SK-SVR_{IIW}) as used for WF.

The MLR models were found beneficial to realize the association between dependent ERPI variables and respective independent RWQPs. However, the MLR models are not flexible enough to seize the complex associations and poorly perform with nonlinear relationships (Tabari et al. 2012; Rajaei et al. 2018; Keshtegar et al. 2019). In contrast, the SVR models have the adaptability and capacity to demonstrate the same. Moreover, the training process of the SVR models consistently looks for a globally optimized solution which avoids the problem of over-fitting. The SVR method can select the support vectors (key vectors) and expel the nonsupport vectors (nonkey vectors) consequently from the models. This approach increases the model flexibility into noisy conditions. The SVR models achieve high accuracy because of simultaneous minimization of prediction error and model complexity. The critical limitation of the SVR technique is that it is a black-box data-driven technique without any physical basis (Bozorg-Haddad et al., 2017; Liu and Lu 2014; Raghavendra and Deka 2014). Moreover, the SVR models can only be used when the training datasets are available (Ji et al. 2017). Overall, it was found that the MLR and SVR (except SK-based SVR) models afforded good agreement with the ERPI models to evaluate the RWQ. The comprehensive ranking of the developed models was shown in Table 9.

Fig. 5 RWQ values evaluated by the ERPI and SVR models with different kernel functions in the testing phase for **a** DD; **b** OB; **c** DCD; **d** WF, and **e** IIW

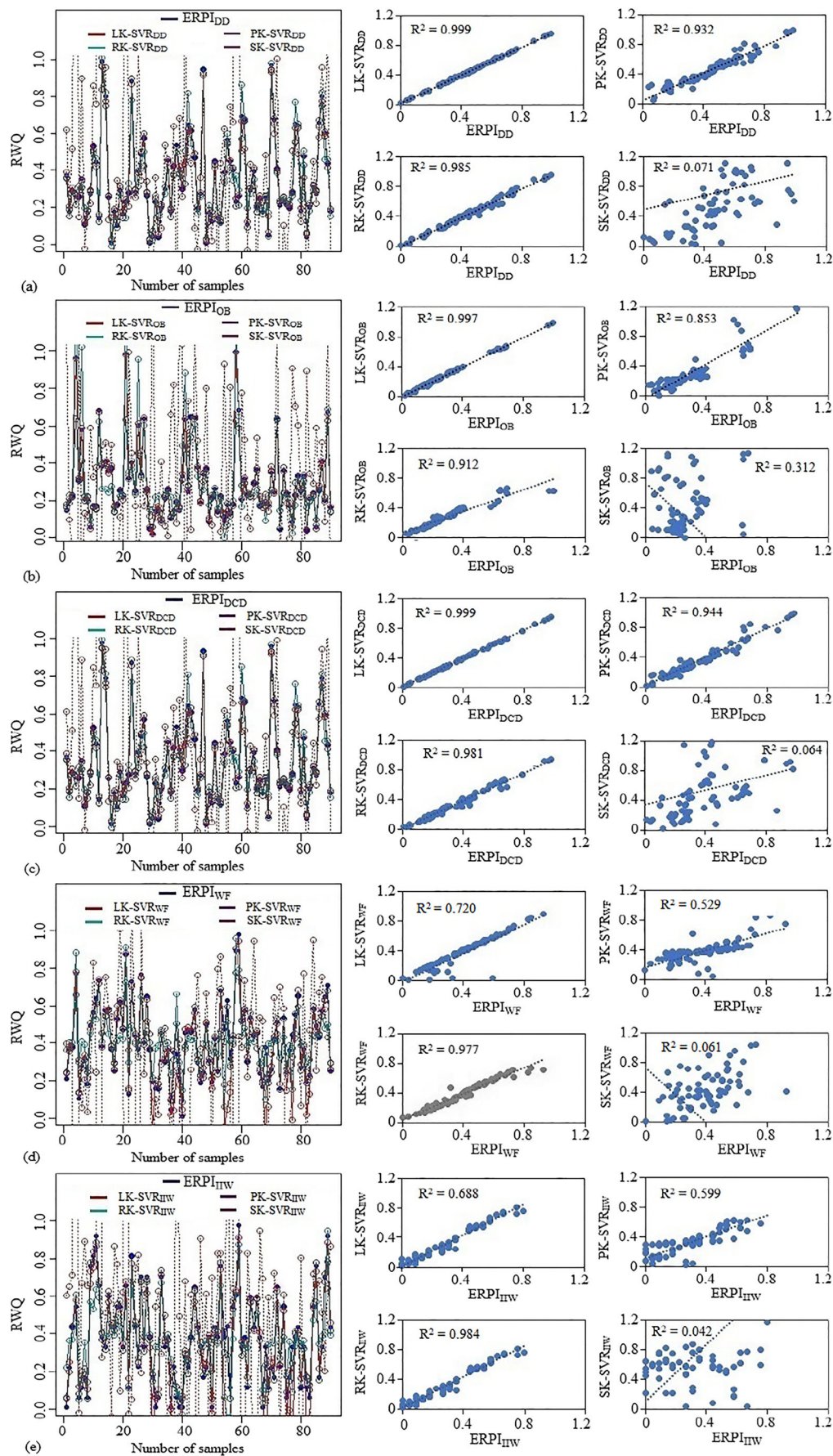


Table 8 Performance comparison of MLR and SVR models for the testing phase

MLR models	RMSE	R ²	MAE	SVR models	RMSE	R ²	MAE				
MLR _{DD}	0.001	0.999	0.001	LK-SVR _{DD}	0.012	0.999	0.010				
				PK-SVR _{DD}	0.058	0.932	0.042				
				RK-SVR _{DD}	0.029	0.985	0.021				
				SK-SVR _{DD}	1.030	0.071	0.472				
				MLR _{OB}	0.009	0.997	0.006	LK-SVR _{OB}	0.010	0.997	0.008
MLR _{OB}	0.009	0.997	0.006	PK-SVR _{OB}	0.102	0.853	0.061				
				RK-SVR _{OB}	0.072	0.912	0.030				
				SK-SVR _{OB}	0.804	0.312	0.457				
				MLR _{DCD}	0.002	0.999	0.001	LK-SVR _{DCD}	0.013	0.999	0.011
				MLR _{DCD}	0.002	0.999	0.001	PK-SVR _{DCD}	0.053	0.944	0.039
RK-SVR _{DCD}	0.033	0.981	0.023								
SK-SVR _{DCD}	1.010	0.064	0.455								
MLR _{WF}	0.092	0.757	0.061					LK-SVR _{WF}	0.116	0.720	0.047
MLR _{WF}	0.092	0.757	0.061					PK-SVR _{WF}	0.129	0.529	0.098
				RK-SVR _{WF}	0.032	0.977	0.021				
				SK-SVR _{WF}	0.861	0.061	0.504				
				MLR _{IIW}	0.127	0.695	0.085	LK-SVR _{IIW}	0.157	0.688	0.067
				MLR _{IIW}	0.127	0.695	0.085	PK-SVR _{IIW}	0.145	0.599	0.108
RK-SVR _{IIW}	0.029	0.984	0.020								
SK-SVR _{IIW}	1.301	0.042	0.682								

Table 9 Summary of the ranking for the developed models

Purpose	Model	Rank
DD	MLR _{DD}	1
	LK-SVR _{DD}	2
	PK-SVR _{DD}	4
	RK-SVR _{DD}	3
	SK-SVR _{DD}	5
OB	MLR _{OB}	1
	LK-SVR _{OB}	2
	PK-SVR _{OB}	4
	RK-SVR _{OB}	3
	SK-SVR _{OB}	5
DCD	MLR _{DCD}	1
	LK-SVR _{DCD}	2
	PK-SVR _{DCD}	4
	RK-SVR _{DCD}	3
	SK-SVR _{DCD}	5
WF	MLR _{WF}	2
	LK-SVR _{WF}	3
	PK-SVR _{WF}	4
	RK-SVR _{WF}	1
	SK-SVR _{WF}	5
IIW	MLR _{IIW}	2
	LK-SVR _{IIW}	4
	PK-SVR _{IIW}	3
	RK-SVR _{IIW}	1
	SK-SVR _{IIW}	5

Conclusion

In this study, the ERPI model was proposed to determine the RWQ of different stretches of the river, not only for DD but also for OB, DCD, WF, and IIW. Assessment of the various monitoring locations determined that the river stretch was not excellent for DD, OB, and DCD. Less than 50% of the river stretch was classified in excellent and good classes for WF. However, the whole river stretch was found suitable for IIW. The ERPI_{MLR} and ERPI_{SVR} models developed here can be implanted for estimating the RWQ to simplify the interpretation of the ERPI models. MLR_{DD}, MLR_{OB}, MLR_{DCD}, RK-SVR_{WF}, and RK-SVR_{IIW} models performed well to evaluate the RWQ for respective usage. The verdicts of this case study offered a rudimentary direction to the water resource managers, irrigation engineers, aqua-culturists, and the general public. Further research is needed to test the developed models for more RWQPs to evaluate their potentiality in RWQ determination.

Acknowledgments The authors acknowledge the Department of Environmental Science & Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, India, for providing the support in carrying out the research work.

Authors' contributions Conceptualization: Suyog Gupta; methodology: Suyog Gupta; formal analysis and investigation: Suyog Gupta; writing - original draft preparation: Suyog Gupta; writing - review and editing: Suyog Gupta, Sunil Kumar Gupta; resources: Sunil Kumar Gupta; supervision: Sunil Kumar Gupta.

Data availability All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Compliance with ethical standards

Ethics approval and consent to participate Not applicable

Consent for publication Not applicable

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

References

- Abbasi T, Abbasi S (2012) Water quality indices. Elsevier, The Netherlands
- Adnan RM, Yuan X, Kisi O, Anam R (2017) Improving accuracy of river flow forecasting using LSSVR with gravitational search algorithm. *Adv Meteorol* 2017:1–23. <https://doi.org/10.1155/2017/2391621>
- Adnan RM, Liang Z, Heddami S, Zounemat-Kermani M, Kisi O, Li B (2020) Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *J Hydrol* 586: 124371. <https://doi.org/10.1016/j.jhydrol.2019.124371>
- Alphayo SM, Sharma MP (2018) Water quality assessment of Ruvi River in Tanzania using NSFQWI. *J Sci Res Rep* 20:1–9. <https://doi.org/10.9734/jsr/2018/44324>
- Bafitilhile TM, Li Z (2019) Applicability of ϵ -support vector machine and artificial neural network for flood forecasting in humid, semi-humid and semi-arid basins in China. *Water* 11(1):85. <https://doi.org/10.3390/w11010085>
- Baird RB, Eaton AD, Rice EW, Bridgewater L (eds) (2017) Standard methods for the examination of water and wastewater. American Public Health Association, Washington, DC
- Banda TD, Kumarasamy MV (2020) Development of water quality indices (WQIs): A review. *Pol J Environ Stud* 29:2011–2021. <https://doi.org/10.15244/pjoes/110526>
- Barakat A, Meddah R, Afdali M, Touhami F (2018) Physicochemical and microbial assessment of spring water quality for drinking supply in Piedmont of Béni-Mellal Atlas (Morocco). *Phys Chem Earth* 104: 39–46. <https://doi.org/10.1016/j.pce.2018.01.006>
- Bhatti EUH, Khan MM, Shah SAR, Raza SS, Shoaib M, Adnan M (2019) Dynamics of water quality: Impact assessment process for water resource management. *Processes* 7:1–14. <https://doi.org/10.3390/pr7020102>
- Bhutiani R, Khanna DR, Kulkarni DB, Ruhela M (2016) Assessment of Ganga river ecosystem at Haridwar, Uttarakhand, India with reference to water quality indices. *Appl Water Sci* 6(2):107–113. <https://doi.org/10.1007/s13201-014-0206-6>
- BIS (1982) Inland surface water standards, IS-2296:1982. Bur Indian Stand New Delhi:1–9
- Bora M, Goswami DC (2017) Water quality assessment in terms of water quality index (WQI): case study of the Kolong River, Assam, India. *Appl Water Sci* 7:3125–3135. <https://doi.org/10.1007/s13201-016-0451-y>
- Borzog-Haddad O, Soleimani S, Loáiciga HA (2017) Modeling water-quality parameters using genetic algorithm – least squares support vector regression and genetic programming. *J Environ Eng (N Y)* 143(7):04017021. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001217](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001217)
- Chen SK, Jang CS, Chou CY (2019) Assessment of spatiotemporal variations in river water quality for sustainable environmental and recreational management in the highly urbanized Danshui River basin. *Environ Monit Assess* 191(2):100. <https://doi.org/10.1007/s10661-019-7246-1>
- CPCB (1979) CPCB, 1979. <https://cpcb.nic.in/wqstandards/>. Accessed 26 Jul 2020
- CPCB (2007) Guidelines for water quality monitoring central pollution control board Parivesh Bhawan East Arjun Nagar, Delhi-32. Cent Pollut board
- Effendi H (2016) River water quality preliminary rapid assessment using pollution index. *Procedia Environ Sci* 33:562–567. <https://doi.org/10.1016/j.proenv.2016.03.108>
- Effendi H, Romanto, Wardiatno Y (2015) Water quality status of Ciambulung River, Banten Province, based on Pollution Index and NSF-WQI. *Procedia Environ Sci* 24:228–237. <https://doi.org/10.1016/j.proenv.2015.03.030>
- EPA (2012) Aquatic Life Ambient Water Quality Criteria for Carbaryl - 2012
- Ewaid SH, Abed SA, Kadhum SA (2018) Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis. *Environ Technol Innov* 11:390–398. <https://doi.org/10.1016/j.eti.2018.06.013>
- Fathi E, Zamani-Ahmadmohammadi R, Zare-Bidaki R (2018) Water quality evaluation using water quality index and multivariate methods, Beheshtabad River, Iran. *Appl Water Sci* 8(7):210. <https://doi.org/10.1007/s13201-018-0859-7>
- Goher ME, Hassan AM, Abdel-moniem IA et al (2014) Evaluation of surface water quality and heavy metal indices of Ismailia Canal, Nile River, Egypt. *Egypt J Aquat Res* 40:225–233. <https://doi.org/10.1016/j.ejar.2014.09.001>
- Granata F, Gargano R, De Marinis G (2016) Support vector regression for rainfall-runoff modeling in urban drainage: a comparison with the EPA's Storm Water Management Model. *Water* 8(3):69. <https://doi.org/10.3390/w8030069>
- Haldar D, Halder S, Das P, Halder G (2014) Assessment of water quality of Damodar River in South Bengal region of India by Canadian Council of Ministers of Environment (CCME) Water Quality Index: a case study desalination and water treatment assessment of water quality of Damodar River in South. *Desalin Water Treat* 57(8):3489–3502. <https://doi.org/10.1080/19443994.2014.987168>
- Hoseinzadeh E, Khorsandi H, Wei C, Alipour M (2015) Evaluation of Aydughmush River water quality using the National Sanitation Foundation Water Quality Index (NSFWQI), River Pollution Index (RPI), and Forestry Water Quality Index (FWQI). *Desalin Water Treat* 54:2994–3002. <https://doi.org/10.1080/19443994.2014.913206>
- Hussein S, Ali S (2017) Water quality index for Al-Gharraf River, Southern Iraq. *Egypt J Aquat Res* 43:117–122. <https://doi.org/10.1016/j.ejar.2017.03.001>
- Ji X, Randy A, Shang X et al (2017) Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang River, China. *Environ Sci Pollut Res* 24:16062–16076. <https://doi.org/10.1007/s11356-017-9243-7>
- Kadam AK, Wagh VM, Muley AA, Umrikar BN, Sankhua RN (2019) Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Model Earth Syst Environ* 5:951–962. <https://doi.org/10.1007/s40808-019-00581-3>
- Keshtegar B, Heddami S, Hosseinabadi H (2019) The employment of polynomial chaos expansion approach for modeling dissolved oxygen concentration in river. *Environ Earth Sci* 78:1–18. <https://doi.org/10.1007/s12665-018-8028-8>
- Kommineni M, Reddy KV, Jagathi K, Reddy BD, Roshini A, Bhavani V (2020). Groundwater level prediction using modified linear

- regression. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) IEEE 1164–1168. <https://doi.org/10.1109/ICACCS48705.2020.9074313>
- Kumar R, Aaditya S, Khushbu C (2019) Water-quality assessment of Damodar River and its tributaries and subtributaries in Dhanbad Coal mining areas of India based on WQI. *Sustain Water Resour Manag* 5:381–386. <https://doi.org/10.1007/s40899-017-0159-7>
- Kundu S, Khare D, Mondal A (2017) Geoscience frontiers future changes in rainfall, temperature and reference evapotranspiration in the central India by least square support vector machine. *Geosci Front* 8: 583–596. <https://doi.org/10.1016/j.gsf.2016.06.002>
- Leong WC, Bahadori A, Zhang J, Ahmad Z (2019) Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). *Int J River Basin Manag*: 1–8. <https://doi.org/10.1080/15715124.2019.1628030>
- Li R, Zou Z, An Y (2016) ScienceDirect Water quality assessment in Qu River based on fuzzy water pollution index method. *J Environ Sci (China)* 50:87–92. <https://doi.org/10.1016/j.jes.2016.03.030>
- 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>
- Mosavi A, Ozturk P, Chau K-W (2018) Flood prediction using machine learning models: literature review. *Water* 10(11):15361. <https://doi.org/10.3390/w10111536>
- Mukate S, Wagh V, Panaskar D, Jacobs JA, Sawant A (2019) Development of new integrated water quality index (IWQI) model to evaluate the drinking suitability of water. *Ecol Indic* 101:348–354. <https://doi.org/10.1016/j.ecolind.2019.01.034>
- Mukherjee D, Dora SL, Tiwary RK (2012) Evaluation of Water Quality Index for drinking purposes in the case of Damodar River, Jharkhand and West Bengal Region, India. *J Bioremediation Biodegrad* 3(9):1–5. <https://doi.org/10.4172/2155-6199.1000161>
- Nayak JG, Patil LG, Patki VK (2020) Development of water quality index for Godavari River (India) based on fuzzy inference system. *Groundw Sustain Dev* 10:100350. <https://doi.org/10.1016/j.gsd.2020.100350>
- Olasoji SO, Oyewole NO, Abiola B, Edokpayi JN (2019) Water quality assessment of surface and groundwater sources using a water quality index method: a case study of a peri-urban town in southwest, Nigeria. *Environ* 6(2):23. <https://doi.org/10.3390/environments6020023>
- Raghavendra S, Deka PC (2014) Support vector machine applications in the field of hydrology: a review. *Appl Soft Comput J* 19:372–386. <https://doi.org/10.1016/j.asoc.2014.02.002>
- Rajae T, Ravansalar M, Adamowski JF, Deo RC (2018) A new approach to predict daily pH in rivers based on the “à trous” redundant wavelet transform algorithm. *Water Air Soil Pollut* 229(3):85. <https://doi.org/10.1007/s11270-018-3715-3>
- Rehana S (2019) River water temperature modelling under climate change using support vector regression. In: *Hydrology in a Changing World*. Springer, Cham, pp 171–183
- Sahoo S, Jha MK (2013) Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. *Hydrogeol J* 21(8):1865–1887. <https://doi.org/10.1007/s10040-013-1029-5>
- Sahoo MM, Patra KC, Khatua KK (2015) Inference of Water Quality Index using ANFIA and PCA. *Aquat Procedia* 4:1099–1106. <https://doi.org/10.1016/j.aqpro.2015.02.139>
- Sedighi F, Vafakhah M, Reza M (2016) Rainfall – runoff modeling using support vector machine in snow-affected watershed. *Arab J Sci Eng* 41(10):4065–4076. <https://doi.org/10.1007/s13369-016-2095-5>
- Şener Ş, Şener E, Davraz A (2017) Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). *Sci Total Environ* 584–585:131–144. <https://doi.org/10.1016/j.scitotenv.2017.01.102>
- Singh B, Sihag P, Deswal S (2019) Modelling of the impact of water quality on the infiltration rate of the soil. *Appl Water Sci* 9:1–9. <https://doi.org/10.1007/s13201-019-0892-1>
- Sotomayor G, Hampel H, Vázquez RF (2018) Water quality assessment with emphasis in parameter optimisation using pattern recognition methods and genetic algorithm. *Water Res* 130:353–362. <https://doi.org/10.1016/j.watres.2017.12.010>
- Tabari H, Kisi O, Ezani A, Talae PH (2012) SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *J Hydrol* 444–445:78–89. <https://doi.org/10.1016/j.jhydrol.2012.04.007>
- Tang T, Strokal M, van Vliet MTH, Seuntjens P, Burek P, Kroeze C, Langan S, Wada Y (2019) Bridging global, basin and local-scale water quality modeling towards enhancing water quality management worldwide. *Curr Opin Environ Sustain* 36:39–48. <https://doi.org/10.1016/j.cosust.2018.10.004>
- Tchounwou PB, Yedjou CG, Patlolla AK, Sutton DJ (2012) Molecular, clinical and environmental toxicology. *Mol Clin Environ Toxicol* 3:133–164. <https://doi.org/10.1007/978-3-7643-8340-4>
- Tian Y, Jiang Y, Liu Q, Dong M, Xu D, Liu Y, Xu X (2019) Science of the total environment using a water quality index to assess the water quality of the upper and middle streams of the Luanhe River, northern China. *Sci Total Environ* 667:142–151. <https://doi.org/10.1016/j.scitotenv.2019.02.356>
- Trikoilidou E, Samiotis G (2017) Evaluation of water quality indices adequacy in characterizing the physico-chemical water quality of lakes. *Environ Process* 4(1):35–46. <https://doi.org/10.1007/s40710-017-0218-y>
- Tripathi M, Singal SK (2019) Ecotoxicology and environmental safety allocation of weights using factor analysis for development of a novel water quality index. *Ecotoxicol Environ Saf* 183:109510. <https://doi.org/10.1016/j.ecoenv.2019.109510>
- Tyagi S, Sharma B, Singh P, Dobhal R (2013) Water quality assessment in terms of water quality index. *American J Water Res* 1(3):34–38. <https://doi.org/10.12691/ajwr-1-3-3>
- Ustaoğlu F, Tepe Y, Taş B (2020) Assessment of stream quality and health risk in a subtropical Turkey river system: a combined approach using statistical analysis and water quality index. *Ecol Indic* 113:113. <https://doi.org/10.1016/j.ecolind.2019.105815>
- Verma RK, Murthy S, Tiwary RK, Verma S (2019) Development of simplified WQIs for assessment of spatial and temporal variations of surface water quality in upper Damodar river basin, eastern India. *Appl Water Sci* 9:1–15. <https://doi.org/10.1007/s13201-019-0893-0>
- Wang Y, Wang P, Bai Y, Tian Z, Li J, Shao X, Mustavich LF, Li BL (2013) Assessment of surface water quality via multivariate statistical techniques: a case study of the Songhua River Harbin region, China. *J Hydro-Environ Res* 7:30–40. <https://doi.org/10.1016/j.jher.2012.10.003>
- Wu Z, Wang X, Chen Y, Cai Y, Deng J (2018) Assessing river water quality using water quality index in Lake Taihu. *Sci Total Environ* 612:914–922. <https://doi.org/10.1016/j.scitotenv.2017.08.293>
- Zahedi S (2017) Modification of expected conflicts between Drinking Water Quality Index and Irrigation Water Quality Index in water quality ranking of shared extraction wells using Multi Criteria Decision Making techniques. *Ecol Indic* 83:368–379. <https://doi.org/10.1016/j.ecolind.2017.08.017>
- Zhang Q, Yu R, Jin Y, Zhang Z, Liu X, Xue H, Hao Y, Wang L (2019) Temporal and spatial variation trends in water quality based on the WPI index in the shallow lake of an arid area: a case study of lake Ulansuhai, China. *Water* 11(7):1410. <https://doi.org/10.3390/w11071410>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.