water

MDPI

Article

Water Quality Prediction of the Yamuna River in India Using Hybrid Neuro-Fuzzy Models

Ozgur Kisi [1,2,3], Kulwinder Singh Parmar [4,*], Amin Mahdavi-Meymand [5], Rana Muhammad Adnan [6,*], Shamsuddin Shahid [3] and Mohammad Zounemat-Kermani [7]

1 Civil Engineering Department, Ilia State University, Tbilisi 0101, Georgia
2 Department of Civil Engineering, Lübeck University of Applied Science, 23562 Lübeck, Germany
3 School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia
4 Department of Mathematical Sciences, IKG Punjab Technical University, Jalandhar 144103, India
5 Institute of Hydro-Engineering, Polish Academy of Sciences, 8-0328 Gdansk, Poland
6 School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China
7 Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman 76169, Iran
* Correspondence: kulmaths@gmail.com (K.S.P.); rana@gzhu.edu.cn (R.M.A.)

Abstract: The potential of four different neuro-fuzzy embedded meta-heuristic algorithms, particle swarm optimization, genetic algorithm, harmony search, and teaching–learning-based optimization algorithm, was investigated in this study in estimating the water quality of the Yamuna River in Delhi, India. A cross-validation approach was employed by splitting data into three equal parts, where the models were evaluated using each part. The main aim of this study was to find an accurate prediction model for estimating the water quality of the Yamuna River. It is worth noting that the hybrid neuro-fuzzy and LSSVM methods have not been previously compared for this issue. Monthly water quality parameters, total kjeldahl nitrogen, free ammonia, total coliform, water temperature, potential of hydrogen, and fecal coliform were considered as inputs to model chemical oxygen demand (COD). The performance of hybrid neuro-fuzzy models in predicting COD was compared with classical neuro-fuzzy and least square support vector machine (LSSVM) methods. The results showed higher accuracy in COD prediction when free ammonia, total kjeldahl nitrogen, and water temperature were used as inputs. Hybrid neuro-fuzzy models improved the root mean square error of the classical neuro-fuzzy model and LSSVM by 12% and 4%, respectively. The neuro-fuzzy models optimized with harmony search provided the best accuracy with the lowest root mean square error (13.659) and mean absolute error (11.272), while the particle swarm optimization and teaching–learning-based optimization showed the highest computational speed (21 and 24 min) compared to the other models.

Keywords: river water; pollution; chemical oxygen demand; neuro-fuzzy; meta-heuristic algorithms; harmony search

check for updates

Citation: Kisi, O.; Parmar, K.S.; Mahdavi-Meymand, A.; Adnan, R.M.; Shahid, S.; Zounemat-Kermani, M. Water Quality Prediction of the Yamuna River in India Using Hybrid Neuro-Fuzzy Models. *Water* **2023**, *15*, 1095. https://doi.org/10.3390/w15061095

Academic Editor: Domenico Cicchella

Received: 15 February 2023
Revised: 8 March 2023
Accepted: 10 March 2023
Published: 13 March 2023

1. Introduction

The industrialization of economics has caused serious environmental problems worldwide. This issue made the members of the United Nations agree to 17 sustainable development goals (SDGs) for growing economies and poverty reduction, while preserving the environment [1]. Conserving the oceans and seas is one of the fundamental goals of the SDGs. Rivers are one of the primary sources of water that discharge from the land to the sea, and can transfer pollution to the seas and oceans.

Water is vital for life, and the river is the major source of water for life. Therefore, river water quality (WQ) and maintaining river WQ are crucial for sustainable living on earth. They are also crucial for the sustainability of the global ecosystem. However, economic activities, industrialization, and urbanization have affected river WQ globally. This is more prominent in developing countries, due to rapid but often unplanned development. The

Yamuna River, the largest tributary of India's biggest river Ganges, is an example of such pollution. River water pollution continuously increased with increased transportation, urbanization, and industrialization. Industrial wastes, city sewerages, and agricultural runoff significantly reduced the river WQ [2–5] and disturbed the whole ecosystem, including animals and humans, especially children's health. Monitoring the WQ of the Yamuna River is urgent to adopt protective measures and ensure ecosystem stability [6,7]. However, precise WQ monitoring is challenging for the river Yamuna due to the effect of many points and non-point pollution sources. Robust models are required to estimate WQ changes accurately, with minimum environmental inputs [8].

Chemical oxygen demand (COD) indicates the amount of oxidizable organic material in the river water and, therefore, the dissolved oxygen (DO) levels and the anaerobic conditions. A higher COD indicates a lower DO level and insufficient conditions for aquatic life. Therefore, COD is widely used to measure river WQ [9–11]. Numerous models have been developed for predicting river WQ. Most of these models are statistical, based on multiple linear regression, moving average, and auto-regressive moving average. Such statistical methods cannot address the nonlinearity in data; thus, they often fail to predict WQ in complex situations [12–14]. Recent studies indicate that ordinary and advanced artificial intelligence (AI) models are robust tools in pattern recognition, and are gaining popularity [15]. Yilma et al. [16] recommended the application of an artificial neural network (ANN) for the prediction of the river WQ index. Ahmed et al. [17] compared the performance of an adaptive neuro-Fuzzy inference system (ANFIS) and two ANNs in the prediction of river WQ. The results demonstrated that the ANFIS was capable of providing greater accuracy. Abba et al. [18] developed three AI models for the prediction of WQ. The considered models included the ANFIS, ANN, and least square support vector machine (LSSVM). The obtained results indicated that the ANFIS outperformed the other methods. Lee and Kim [19] used an ANFIS structure for the simulation of biological oxygen demand (BOD) in the Dongjin River. The results confirmed the accuracy of the developed ANFIS. Wong et al. [20] used an ANN and square support vector machine (SVM) for monsoonal river classification based on water quality. The results approved the accuracy of both the ANN and LSSVM; however, the ANN was more accurate.

Hybrid AI models, e.g., LSSVM or ANFIS with meta-heuristic algorithms, have been introduced to address the drawbacks of statistical methods [21–24]. Fadaee et al. [25] used a butterfly optimization algorithm (BOA) for training the ANFIS to predict dissolved oxygen (DO) in rivers. The results showed that the BOA is stronger than other optimization algorithms in the literature. Song et al. [26] developed a model for the prediction of WQ based on the LSSVM and sparrow search algorithm (SSA). The capability of LSSVM–SSA was confirmed in the Yangtze River. Arya Azar et al. [27] developed two hybrid algorithms for estimating the longitudinal dispersion coefficient of river pollution. The models included a hybrid of the ANFIS and SVR, with Harris hawks optimization (HHO) meta-heuristic algorithm. The results demonstrated that the HHO may increase the performance of AI models.

Around 40% of India's populace relies on the Yamuna River for water supply. Therefore, the Yamuna River's WQ prediction using highly accurate models is directly related to national public health and a sustainable environment. In this study, AI-based models were used for accurate prediction of the Yamuna River's WQ. The least square support vector machine (LSSVM) model was developed using the strength of kernels, which can predict any phenomenon much more accurately than statistical models [28–30]. Kernel-based methods can handle the nonlinearity and non-stationary of time series and accurately predict the series [31–33]. The ANFIS has emerged as a powerful AI model for predicting environmental processes. It is more accurate than the classical AI models [34]. However, the ANFIS also requires tuning of its internal parameters for improved accuracy. The ANFIS uses derivative-based learning as the standard parameter learning process, which has a high probability of becoming trapped in local minima. The recent literature revealed that integrating AI models with optimization algorithms could improve their prediction performance by finding optimal control parameters. In the present study, the ANFIS was integrated with four meta-heuristic algorithms, particle swarm optimization (PSO), genetic algorithm (GA), harmony search (HS), and teaching–learning-based optimization (TLBO) to predict the Yamuna River, Delhi's long-term WQ. Most meta-heuristic algorithms need to be initialized before starting the iterations to calculate the best answer. TLBO was chosen since it is known as one of the optimization algorithms that needs the lowest number of initial parameters. PSO, GA, and HS are famous and powerful algorithms, and their performance has been confirmed in many disciplines. The performances of hybrid ANFISs were compared with the classical ANFIS method to show the efficiency of the TLBO algorithm compared to the classical method. The heuristic ANFIS methods were also compared with the LSSVM method, which was recently applied by Kisi and Parmar [21] to investigate the accuracy of proposed neuro-fuzzy methods in estimating COD. It is worth noting that the application of LSSVM, as well as TLBO, PSO, GA, and HS meta-heuristics algorithms together with the ANFIS to model WQ variables, is a novel comparison. Since the performance of meta-heuristics depends on the particular problem, the results of this research can determine the best candidates for practical applications with the Yamuna River.

A brief overview of the study area is provided in Section 2, whereas a description of the ANFIS and meta-heuristics algorithms are provided in Section 3. Section 4 discusses the results obtained through the application of the models, and finally, Section 5 provides the main conclusions derived from the study, including limitations and recommendations.

2. Case Study

The Yamuna River is the longest and largest tributary of the Ganga, the largest river in India. After originating from the Yamunotri Glacier in the Garhwal Himalayas in northern India, it travels 1376 km before merging with the River Ganga at Allahabad. The Yamuna River contributes 40.2% of the total water of the Ganga. Nearly 70% or 57 million inhabitants of the Indian capital Delhi depend on the Yamuna River for water. The river mixes with the drainage system many times during its long travel from its origin, which causes severe pollution of its water.

The sampling site at Nizamuddin in Delhi is used to monitor the WQ of the Yamuna. The industrial waste and sewerage of the states of Haryana and Delhi affect the WQ at the sample site (Figure 1). This study used 10-year monthly average COD data (January 1999–April 2009) collected by the Central Pollution Control Board (www.cpcb.nic.in, accessed on 1 July 2020). A basic statistical summary of the data is provided in Table 1. WQ parameters of free ammonia (AMM), total kjeldahl nitrogen (TKN), water temperature (WT), total coliform (TC), fecal coliform (FC), and potential of hydrogen (PH) were recorded at the sample site. Table 2 provides the Pearson's correlations between the WQ parameters and COD for all of the data sets. It is clear from the table that the COD is highly positively correlated with the river water parameters AMM, TKN, TC, and FC, while it has negative correlations with the pH and WT parameters. The mean values of the river water parameters for the studied period are 7.47225 mg/L, 65.05833 mg/L, 15.42467 mg/L, 20.498 mg/L, 25.68517 mg/L, 39,941,063 mg/L, and 5,084,043 mg/L for the pH, COD, AMM, TKM, WT, TC, and FC, respectively.

The WQ parameters were used as inputs to develop the COD prediction in different scenarios. The effect of the parameters was analyzed in these scenarios. Cross-validation was used to assess the model performance, where the available data (120 monthly values) were split into three equal parts, M1, M2, and M3, as shown in Table 3. Thus, the models were evaluated for three different data sets. The period of each training and test is provided in Table 3, where M1 indicates model 1, and vice versa. The data of the clusters were divided into two main parts, including the training and the testing data sets. The models were optimized using the training part, and the testing data sets were used for evaluating the accuracy of the predictions. About 15% of the training data were randomly separated during the optimization process to prevent overfitting. In cluster M1, the models were trained using data from 1999 January to 2005 August (80 monthly values), and were tested using data from 2005 September to 2009 December (40 monthly values). The other periods can be observed in Table 3.

Figure 1. Sampling site at Nizamuddin in Delhi.

Table 1. The monthly statistics of COD at the sampling site in Delhi during different periods (Kisi and Parmar, [21]).

Data Set	x_{mean}	S_x	C_{sx}	x_{min}	x_{max}
January 1999 to April 2002	56.8	22.6	−0.08	18	104
April 2002 to September 2005	70.4	25.4	−0.64	13	116
September 2005 to December 2009	68.0	31.9	−0.24	9	127

Note: x_{mean}, S_x, C_{sx}, x_{min}, and x_{max} indicate the overall mean, standard deviation, and skewness, respectively.

Table 2. Pearson's correlations between water quality parameters and COD.

		pH	AMM	TKN	WT	TC	FC
COD	Pearson Correlation	−0.048	0.823 **	0.741 **	−0.273 **	0.211 *	0.164
	Sig. (2-tailed)	0.603	0.000	0.000	0.003	0.021	0.074
	N	120	120	120	120	120	120

Notes: ** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Table 3. The training and test data sets used in the study (Kisi and Parmar, [21]).

Cross-Validation	Training	Testing
M1	Jan1999 to August 2005	September 2005 to December 2009
M2	January 1999 to April 2002 & September 2005 to December 2009	May 2002 to August 2005
M3	May 2002 to December 2009	January 1999 to August 2002

3. Methods

3.1. Least Square Support Vector Machine

The SVMs were constructed based on the statistical learning theory and the structural risk minimization principle. These make the SVMs sufficiently capable of not becoming trapped in local minima. However, reaching out to an accurate SVM model was challenging due to its requirement of solving a set of nonlinear quadratic equations. In this respect, Suykens et al. [35] introduced a simpler form of the SVM known as the least square support vector machine (LSSVM). LSSVM employs a set of linear equations to train models. Similarly to the SVMs, the LSSVMs models are based on kernel methods, which can accurately estimate hydrological phenomena during training and testing [21].

3.2. Adaptive Neuro-Fuzzy Inference System

The adaptive neuro-Fuzzy inference system (ANFIS) is a robust data-driven model that integrates a feed-forward artificial neural network (ANN) and fuzzy inference system (FIS) to simulate complex problems. In the ANFIS, the Sugeno-type FIS part is utilized to process the input information using different numbers and membership functions (MFs). For the adjustment of the fuzzy logic parameters, an adaptive learning algorithm that integrates the least square and ANN training algorithm (gradient descent) is utilized. Information about the theoretical and practical usage of the ANFIS can be found in several pertinent sources [36,37].

3.3. The Hybrid Procedure of ANFIS and Meta-Heuristic Algorithms

In the ANFIS, MF parameters, such as the center and the width in Gaussian MFs, should be optimized. In the standard version of the ANFIS, the amalgamation of gradient descent (CD) and least square estimator (LSE) optimizes the parameters. Instead of using the CD-LSE algorithm, the ANFIS structure can be merged with meta-heuristic algorithms. It has been reported in some previous studies that merging meta-heuristic algorithms improves the model accuracy in solving complex hydrological problems [38–40].

This study assessed the skill of the ANFIS model merged with four meta-heuristic algorithms, particle swarm optimization (PSO), genetic algorithm (GA), harmony search (HS), and teaching–learning-based optimization (TLBO), and compared their performance with the standalone ANFIS model. The performance was compared in optimizing the Gaussian MF parameters for the inputs and linear MF parameters for the output of the ANFIS. Figure 2 shows a flowchart of the developed integration of the ANFIS with meta-heuristic algorithms. A brief description of the algorithms is as follows.

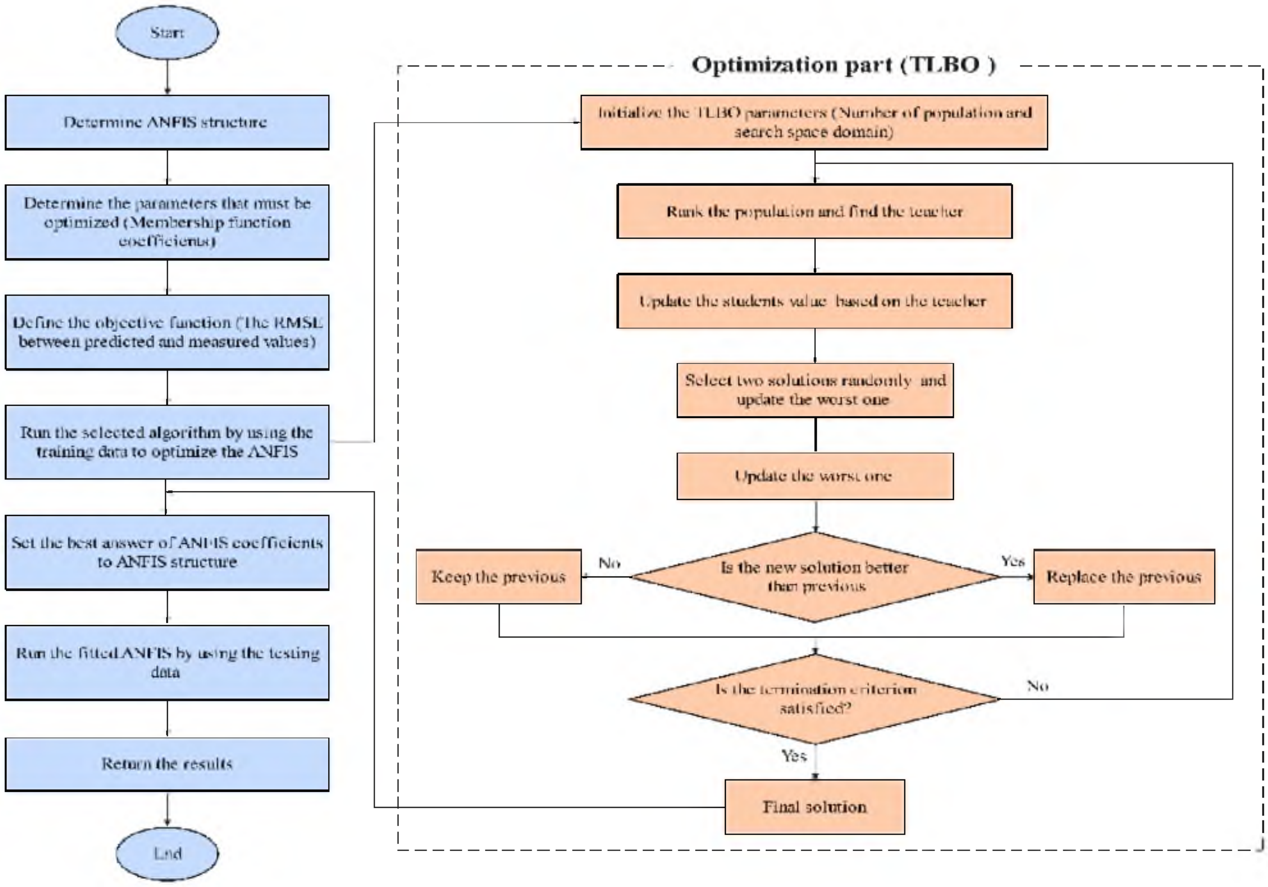

Figure 2. Flowchart of integrated ANFIS with TLBO.

3.3.1. Particle Swarm Optimization

Particle swarm optimization (PSO) is considered a population-based evolutionary optimization algorithm that can be applied to decision-making functions. Its creation was inspired by the sociological and biological behavior of animals in groups (e.g., flocks of birds). In PSO, each potential solution (swarm) represents the particle of a population. Particles follow the optimal particle (global best; Gbest) through a multi-dimensional search space with keeping the memory of their own previous best personal solution (Pbest). In this regard, each particle updates its position and velocity vector according to the values of Pbest and Gbest [41,42].

3.3.2. Genetic Algorithm

Genetic algorithm (GA) is a search technique that is widely employed to solve optimization issues. It is a particular kind of evolutionary algorithm that makes use of concepts from evolutionary biology including natural selection and genetic drift. GAs use Darwinian principles of natural selection to arrive at the best possible formula for making a prediction or modifying a pattern. They work well with regression-based forecasting methods. It mimics the way natural selection works to solve problems. Some of the inputs derive solutions via genetic selection, which are then evaluated as candidates using the fitness function. The process is iterated until the termination condition is fulfilled. In general, GA is an iteration-based algorithm which finds the solution through a random process [43–45].

3.3.3. Harmony Search

Harmony search (HS) is one of the newest and simplest meta-heuristic methods that mimics an orchestra's harmonic behavior of searching for the optimal feasible solution. In other words, finding an optimal solution for a complex problem resembles playing music.

The HS has recently become a popular optimization algorithm due to its applicability for discrete and continuous optimization problems, its few mathematical calculations, simple concept, few parameters, and easy running. Furthermore, compared to other meta-heuristic methods, it has fewer mathematical requirements, and has been widely adapted for solving different engineering issues through simply changing the parameters and operators. Another advantage of this method over the GA is that it uses all the available solutions in its memory, which yields higher flexibility in searching the solution spaces [46–48].

3.3.4. Teaching–Learning-Based Optimization Algorithm

The teaching–learning-based optimization algorithm (TLBO), proposed by Rao [49], was designed based on principles of learning and teaching, where the teacher plays an essential role in the class, and can raise students' levels and the average level of the class through a good speech and communication style. Generally, an individual with a better value and higher level compared to others is determined to be a teacher who shares his/her knowledge with others. The TLBO algorithm comprises two optimization phases, the teacher and learning phases.

In the teacher phase, the average class level is raised to the teacher level; thus, the student's level changes in this phase. The teacher phase is followed by the learning phase, where the students can learn from and influence each other to improve students' levels [50–52].

4. Application of the Methods

Four meta-heuristic algorithms, PSO, GA, HS, and TLBO, were applied to improve the skill of the classic ANFIS in estimating river water chemical oxygen demand (COD) from six water quality (WQ) parameters, free ammonia (AMM), total kjeldahl nitrogen (TKN), water temperature (WT), total coliform (TC), fecal coliform (FC), and potential of hydrogen (PH). Meta-heuristic algorithms were integrated with theANFIS to improve its performance. The improvement was measured by comparing the hybrid ANFIS model with the classical ANFIS and least square support vector machine (LSSVM) models. The following input combinations were attempted, following the previous study of Kisi and Parmar [13,15]:

i. AMM, TKN, and WT;
ii. AMM, TKN, WT, and TC;
iii. AMM, TKN, WT, TC, and FC;
iv. AMM, TKN, WT, TC, FC, and PH.

The parameter values of the meta-heuristic algorithms are provided in Table 4. These values were selected based on recommendations from the literature [53,54], and trial and error. The models' performance was evaluated using root mean square error (RMSE), correlation coefficient (R^2), mean absolute error (MAE), and peak percent threshold statistics (PPTS), as described in Equations (1)–(4), following the study of [15]:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(COD_{i,o} - COD_{i,e})^2} \tag{1}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|COD_{i,o} - COD_{i,e}| \tag{2}$$

$$PPTS_{(l,u)} = \frac{1}{(k_l - k_u + 1)}\sum_{i=k_l}^{k_u}|E_i| \tag{3}$$

$$E_i = \frac{COD_{i,o} - COD_{i,e}}{COD_{i,o}} \times 100 \tag{4}$$

where N is the sample size; $COD_{i,o}$ and $COD_{i,e}$ are the measured and modelled COD, respectively; $k_l = \frac{lxN}{100}$ and $k_u = \frac{uxN}{100}$ in which u and l are higher and lower bounds in %, respectively; E_i denotes the relative error of the ith data. $PPTS_{(l,u)}$ indicates the mean absolute relative error in modeling COD varying between the top u% and l% data.

Table 4. The parameter values of the four meta-heuristic algorithms used in this study.

Optimization Method	Parameters
PSO	Population Size = 500 Maximum Iteration = 2000 Iteration Weight = 1 Inertia Weight Damping Ratio = 0.95 Personal Learning Coefficient = 1 Global Learning Coefficient = 2
GA	Population Size = 500 Maximum Iteration = 2000 Crossover Percentage = 0.7 Mutation Rate = 0.01
HS	Harmony Memory Size = 500 Maximum Iteration = 2000 Pitch Adjustment Rate = 0.1 Harmony Memory Consideration Rate = 0.9
TLBO	Population Size = 500 Maximum Iteration = 2000

5. Results and Discussion

Table 5 presents the performance of the applied methods in modeling Chemical Oxygen Demand (COD) using three inputs, free ammonia (AMM), total kjeldahl nitrogen (TKN), and water temperature (WT). Average statistics of the test results are also provided in Table 5. It shows that the ANFIS with meta-heuristic algorithms performed better than the least square support vector machine (LSSVM). As expected, merging the neuro-fuzzy method and the new algorithms remarkably enhanced the predictivity of the classical ANFIS method. Among the hybrid ANFIS methods, the ANFIS–HS provided the best RMSE (14.024 mg/L), MAE (11.033 mg/L), and the best PPTS criterion estimates. The ANFIS–HS decreased the RMSE of the classical ANFIS and LSSVM from 16.261 mg/L and 15.093 mg/L to 14.024 mg/L, with percentages of 13.75 and 7.08, respectively.

The RMSE, R^2, MAE, and PPTS statistics of the different neuro-fuzzy methods and LSSVM are shown in Table 6 for input combination (ii) (AMM, TKN, WT, and total coliform, TC). Here also, the methods showed the best and worst predictivity for the M2 and M1 data sets. Training the ANFIS with meta-heuristic algorithms improved its accuracy, similarly to the previous input combination. Hybrid ANFIS methods, except ANFIS–PSO, outperformed the LSSVM, while ANFIS–PSO showed similar performance. Among the hybrid methods, ANFIS–TLBO and ANFIS–HS showed the best performance. However, the PPTS 5%, PPTS 10%, and PPTS 20% values of ANFIS–TLBO were lower than ANFIS–HS, which indicates that TLBO acted slightly better than HS. The RMSE of the ANFIS and LSSVM methods reduced from 16.722 mg/L and 15.177 mg/L to 14.565 mg/L, or by 13% and 4% using the ANFIS–TLBO and ANFIS–HS methods, respectively. The addition of TC as input could not improve the accuracy of the applied models.

Tables 7 and 8 show the test results of the applied methods for input combinations (iii) and (iv), respectively. The results showed that all hybrid ANFIS methods provided a higher skill than the classical ANFIS and LSSVM algorithms in modelling COD. For input combination (iii), the ANFIS–HS provided the best accuracy in terms of various comparison statistics. This method increased the RMSE of the classical ANFIS and LSSVM methods by 7% and 6%, respectively. For input combination (iv), the ANFIS–TLBO showed the best performance in RMSE, while the ANFIS–HS showed slightly lower PPTS than the

ANFIS–TLBO. The RMSE of the ANFIS and LSSVM methods reduced from 16.451 mg/L and 15.987 mg/L to 15.158 mg/L, or by 8% and 5%, respectively, using the ANFIS–TLBO. The results indicate that COD estimation accuracy did not increase by including fecal coliform (FC) and potential of hydrogen (PH) as inputs.

Table 5. Comparison of models' performance with AMM, TKN, and WT as inputs.

Method	Cross-Validation	Statistics					
		RMSE	R	MAE	PPTS 5%	PPTS 10%	PPTS 20%
ANFIS	M1	17.874	0.824	13.845	29.360	30.872	34.236
	M2	13.770	0.837	11.411	23.314	24.465	26.914
	M3	17.139	0.743	13.552	30.714	32.284	35.637
	Mean	16.261	0.801	12.936	27.796	29.207	32.262
ANFIS–PSO	M1	15.872	0.864	12.333	26.291	27.624	30.582
	M2	13.723	0.840	11.327	22.364	23.469	25.750
	M3	15.396	0.759	12.399	27.155	28.470	31.163
	Mean	14.997	0.821	12.020	25.270	26.521	29.165
ANFIS–GA	M1	15.646	0.870	11.970	24.669	25.931	28.535
	M2	13.802	0.837	11.315	23.415	24.597	27.082
	M3	15.372	0.739	12.298	28.318	29.796	32.933
	Mean	14.940	0.815	11.861	25.467	26.775	29.517
ANFIS–HS	M1	15.226	0.878	11.650	24.416	25.659	28.400
	M2	12.802	0.860	10.249	19.934	20.978	23.030
	M3	14.043	0.795	11.199	25.386	26.671	29.228
	Mean	14.024	0.844	11.033	23.245	24.436	26.886
ANFIS–TLBO	M1	15.470	0.874	11.946	25.600	26.902	29.749
	M2	13.280	0.850	11.051	22.889	23.966	26.371
	M3	15.479	0.747	12.523	28.691	30.174	33.405
	Mean	14.743	0.824	11.840	25.727	27.014	29.842
LSSVM *	M1	16.460	0.867	12.720	28.110	29.520	32.520
	M2	13.590	0.915	11.150	22.760	23.980	26.500
	M3	15.230	0.841	12.420	28.760	30.200	33.270
	Mean	15.093	0.874	12.097	26.543	27.900	30.763

Note: * Results were obtained from Kisi and Parmar [21].

Table 6. Comparison of the applied models with AMM, TKN, WT, and TC as inputs.

Method	Cross-Validation	Statistics					
		RMSE	R	MAE	PPTS 5%	PPTS 10%	PPTS 20%
ANFIS	M1	16.403	0.860	12.637	28.844	30.341	33.574
	M2	17.720	0.743	12.928	23.115	24.323	26.839
	M3	16.042	0.710	12.944	28.882	30.297	33.398
	Mean	16.722	0.771	12.836	26.947	28.320	31.270
ANFIS–PSO	M1	16.588	0.856	12.786	28.556	30.046	33.303
	M2	13.290	0.850	11.006	22.369	23.454	25.810
	M3	15.652	0.728	12.421	28.336	29.807	32.971
	Mean	15.177	0.811	12.071	26.420	27.769	30.695
ANFIS–GA	M1	16.827	0.850	12.990	29.630	31.141	34.486
	M2	13.830	0.835	11.213	22.845	24.036	26.378
	M3	15.757	0.720	12.601	28.231	29.685	32.898
	Mean	15.471	0.802	12.268	26.902	28.287	31.254
ANFIS–HS	M1	16.184	0.859	12.482	26.589	27.955	31.013
	M2	12.940	0.858	10.683	22.403	23.580	26.161
	M3	14.571	0.786	11.517	25.569	26.822	29.547
	Mean	14.565	0.834	11.561	24.854	26.119	28.907

Table 6. *Cont.*

Method	Cross-Validation	Statistics					
		RMSE	R	MAE	PPTS 5%	PPTS 10%	PPTS 20%
ANFIS–TLBO	M1	15.539	0.872	11.843	24.392	25.652	28.355
	M2	13.427	0.846	10.578	21.101	22.192	24.656
	M3	14.729	0.770	11.906	26.633	28.049	30.984
	Mean	14.565	0.829	11.442	24.042	25.298	27.998
LSSVM *	M1	16.540	0.865	12.830	28.130	29.520	32.490
	M2	13.760	0.837	11.250	22.840	24.020	26.580
	M3	15.230	0.749	12.420	28.760	30.200	33.270
	Mean	15.177	0.817	12.167	26.577	27.913	30.780

Note: * Results were obtained from Kisi and Parmar [21].

Table 7. Comparison of the applied models with AMM, TKN, WT, TC, and FC as inputs.

Method	Cross-Validation	Statistics					
		RMSE	r	MAE	PPTS 5%	PPTS 10%	PPTS 20%
ANFIS	M1	16.766	0.851	12.959	29.562	31.069	34.420
	M2	14.895	0.812	11.793	23.277	24.444	26.965
	M3	15.709	0.722	12.570	28.059	29.511	32.677
	Mean	15.790	0.795	12.441	26.966	28.341	31.354
ANFIS–PSO	M1	16.595	0.853	12.559	28.952	30.457	33.915
	M2	14.517	0.824	10.678	20.474	21.536	23.953
	M3	15.644	0.724	12.449	28.473	29.961	33.219
	Mean	15.585	0.800	11.895	25.966	27.318	30.362
ANFIS–GA	M1	16.764	0.851	12.959	29.560	31.066	34.416
	M2	14.823	0.816	11.822	23.061	24.225	26.661
	M3	14.985	0.749	12.128	27.780	29.207	32.282
	Mean	15.524	0.805	12.303	26.800	28.166	31.120
ANFIS–HS	M1	15.761	0.866	11.738	23.790	25.013	27.775
	M2	13.358	0.858	10.324	22.240	23.391	26.072
	M3	15.177	0.762	11.486	23.987	25.145	27.593
	Mean	14.765	0.829	11.183	23.339	24.516	27.147
ANFIS–TLBO	M1	16.243	0.858	11.936	25.587	26.867	29.685
	M2	12.889	0.862	10.489	22.358	23.483	25.978
	M3	15.522	0.726	12.419	27.342	28.685	31.691
	Mean	14.885	0.815	11.615	25.096	26.345	29.118
LSSVM *	M1	16.440	0.868	12.620	27.800	29.260	32.350
	M2	15.400	0.802	12.460	23.690	24.890	27.620
	M3	15.250	0.748	12.450	28.640	30.070	33.130
	Mean	15.697	0.806	12.510	26.710	28.073	31.033

Note: * Results were obtained from Kisi and Parmar [21].

The computational times of the applied hybrid methods are reported in Table 9 for comparison. The computer's properties were an Intel CPU, Core i7, 8 GB RAM. The total average computational time (in minutes) provided in the table shows that ANFIS–HS predicted the COD in the lowest time during calibration, while the ANFIS–GA was the slowest method. The ANFIS–PSO and ANFIS–TLBO also showed high computational speed compared to the ANFIS–GA method. This clearly indicates the superiority of the PSO and TLBO algorithms over GA.

Table 8. Comparison of the models' performance with all variables as inputs.

Method	Cross-Validation	Statistics					
		RMSE	r	MAE	PPTS 5%	PPTS 10%	PPTS 20%
ANFIS	M1	16.864	0.848	13.068	29.912	31.473	34.998
	M2	15.451	0.802	12.383	23.949	25.160	27.705
	M3	17.038	0.707	13.315	30.207	31.728	34.754
	Mean	16.451	0.786	12.922	28.023	29.454	32.486
ANFIS–PSO	M1	15.534	0.873	11.782	24.583	25.842	28.630
	M2	14.224	0.829	11.550	23.603	24.783	27.307
	M3	15.726	0.722	12.583	28.078	29.529	32.684
	Mean	15.161	0.808	11.972	25.421	26.718	29.540
ANFIS–GA	M1	16.634	0.860	12.679	27.777	29.227	32.420
	M2	15.335	0.804	12.276	23.917	25.103	27.664
	M3	15.250	0.748	12.174	27.890	29.352	32.517
	Mean	15.740	0.804	12.376	26.528	27.894	30.867
ANFIS–HS	M1	16.498	0.857	12.703	24.229	25.408	27.878
	M2	13.659	0.844	11.272	23.102	24.198	26.654
	M3	15.998	0.717	12.759	28.147	29.606	32.547
	Mean	15.385	0.806	12.245	25.159	26.404	29.026
ANFIS–TLBO	M1	16.746	0.851	12.842	28.011	29.456	32.690
	M2	12.850	0.860	10.480	21.734	22.821	25.178
	M3	15.878	0.725	12.663	27.699	29.131	32.144
	Mean	15.158	0.812	11.995	25.815	27.136	30.004
LSSVM *	M1	16.590	0.861	12.720	28.400	29.870	33.170
	M2	15.180	0.809	12.630	23.970	25.080	27.530
	M3	16.190	0.706	13.140	31.150	32.680	35.950
	Mean	15.987	0.792	12.830	27.840	29.210	32.217

Note: * Results were obtained from Kisi and Parmar [21].

Table 9. Computational time (min) in predicting COD by the applied hybrid methods.

Optimization Method	Inputs			
	AMM, TKN and WT	AMM, TKN, WT and TC	AMM, TKN, WT, TC and FC	Total Average CPU Time (min)
ANFIS–PSO	20	21	23	21
ANFIS–GA	104	106	114	108
ANFIS–HS	12	13	13	13
ANFIS–TLBO	22	24	25	24

The observed and model estimated CODs for the input combinations (i), (ii), (iii), and (iv) are illustrated in Figures 2–5, respectively. The figures indicate less scattered estimates by the hybrid ANFIS methods than the ANFIS and LSSVM methods. It is worth noting that the R^2 values were compatible with the model accuracy in some cases, as it indicated a linear relationship between the observed and model estimations. However, $R^2 = 1$ does not indicate that the model exactly estimated all target values. This can also be observed in Tables 5–8, in which correlation was incompatible with the RMSE and/or MAE. In such cases, the RMSE and/or MAE statistics should be considered as the main criterion to determine the best model. The hybrid models were more successful in modelling peak values than the ANFIS and LSSVM, as confirmed by the PPTS statistics in Tables 5–8. It can also be observed that the ANFIS–HS model with input combination (i) and the M2 data set provided more precise results with smaller values of the RMSE (13.358 mg/L) and mean MAE (10.324 mg/L). It is also visible from Figures 3–6 that all of the models could not estimate the extreme COD values. The main reason for this was that limited data involving extreme values prevented the models from learning the extreme phenomena appropriately.

Figure 7 visually compares the RMSE and MAE of the best models using bar charts. This graph also shows the superior accuracy of hybrid ANFIS models over the single ANFIS and LSSVM. A Taylor's diagram of the models for the M2 data set and input combination (i) is illustrated in Figure 8. It shows that the hybrid ANFIS models were slightly more accurate than the ANFIS.

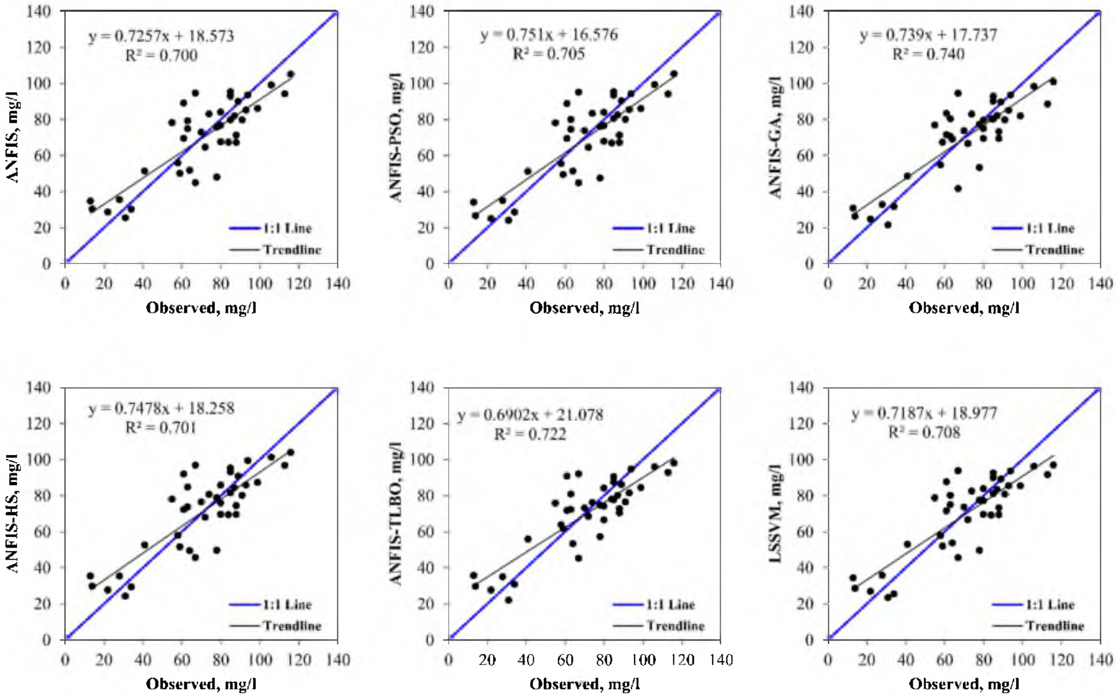

Figure 3. The observed and model-estimated CODs for the M2 data set with AMM, TKN, and WT as inputs.

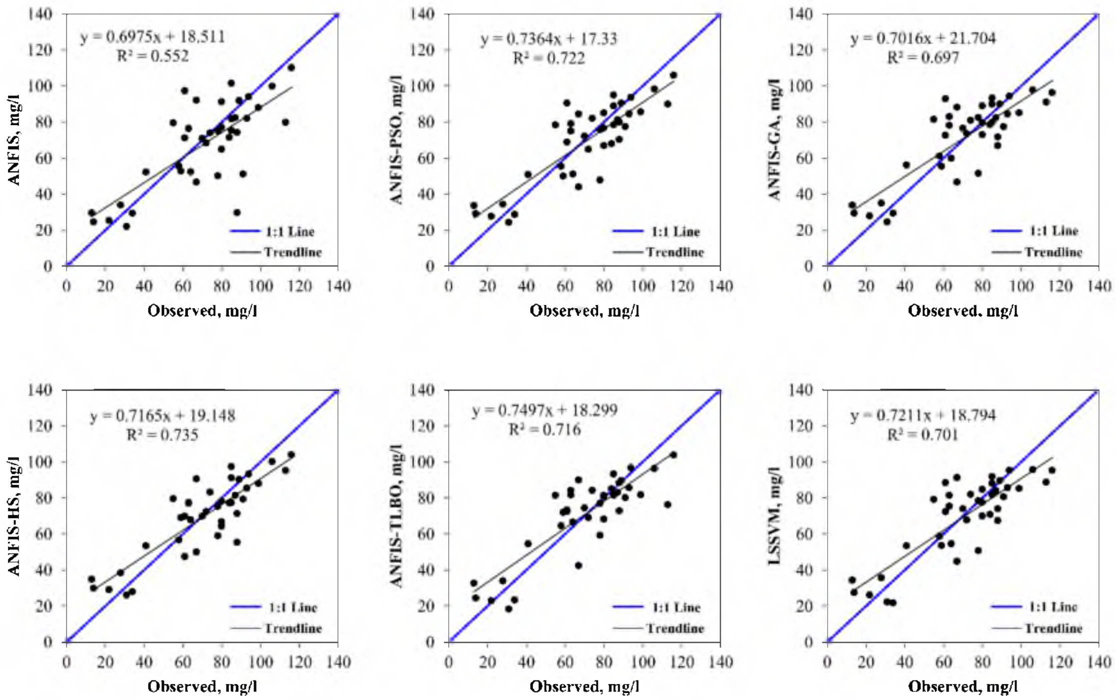

Figure 4. The observed and model-estimated CODs for the M2 data set with AMM, TKN, WT, and TC as inputs.

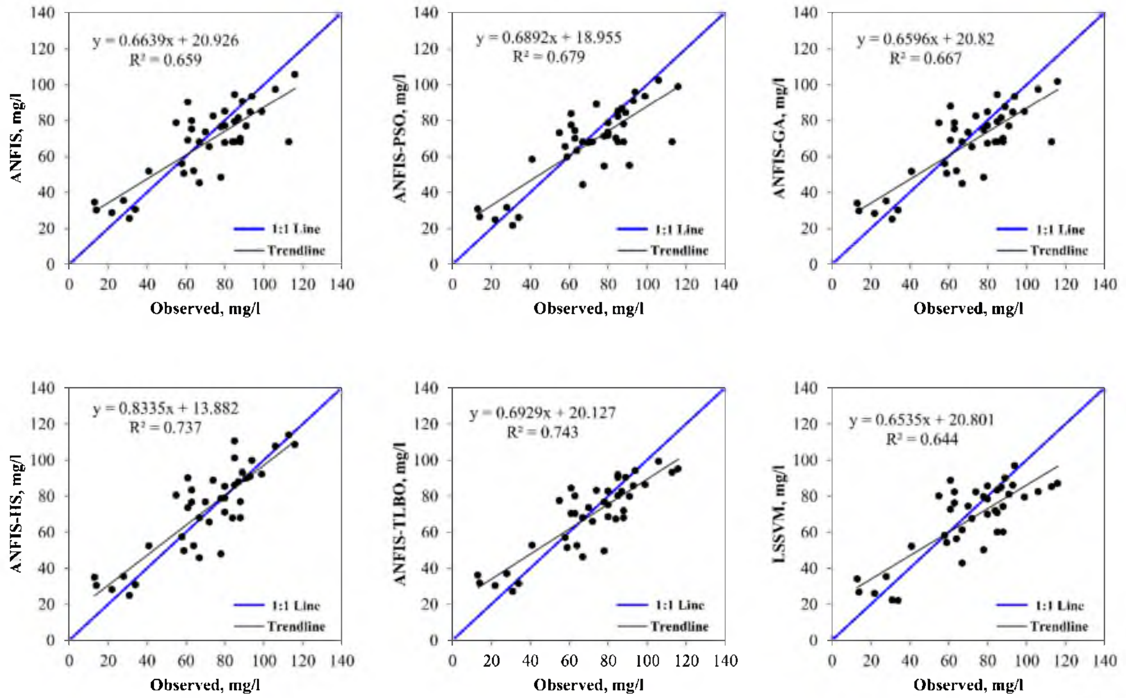

Figure 5. The observed and model-estimated CODs for the M2 data set with AMM, TKN, WT, TC, and as inputs.

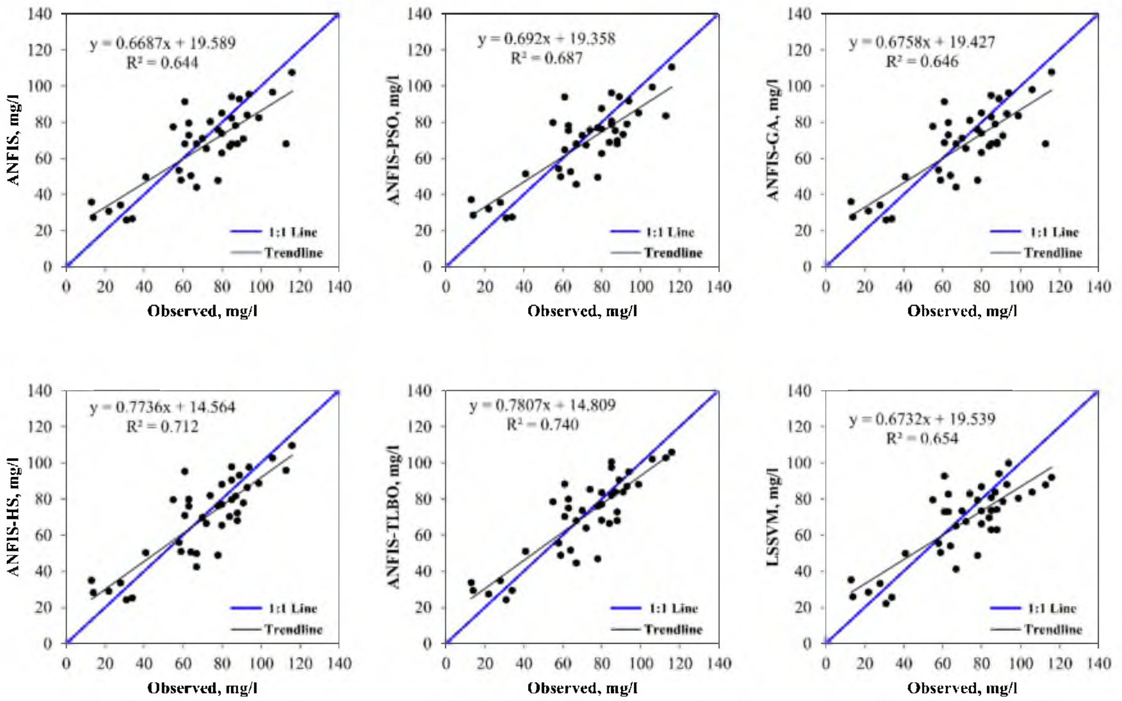

Figure 6. The observed and model-estimated CODs for the M2 data set with all variables as input.

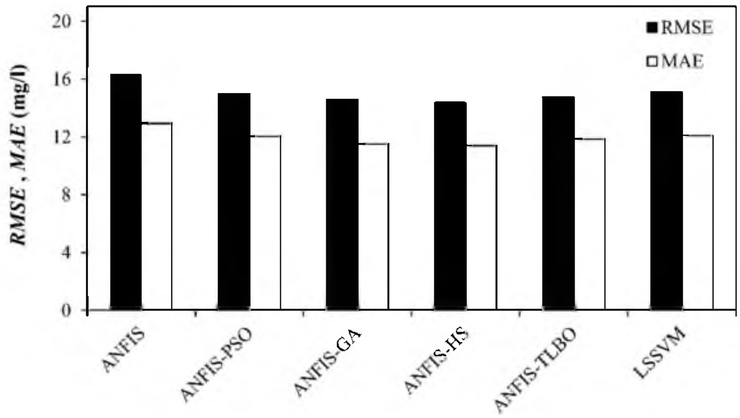

Figure 7. Visual comparison of model performance for the best input combination and data set.

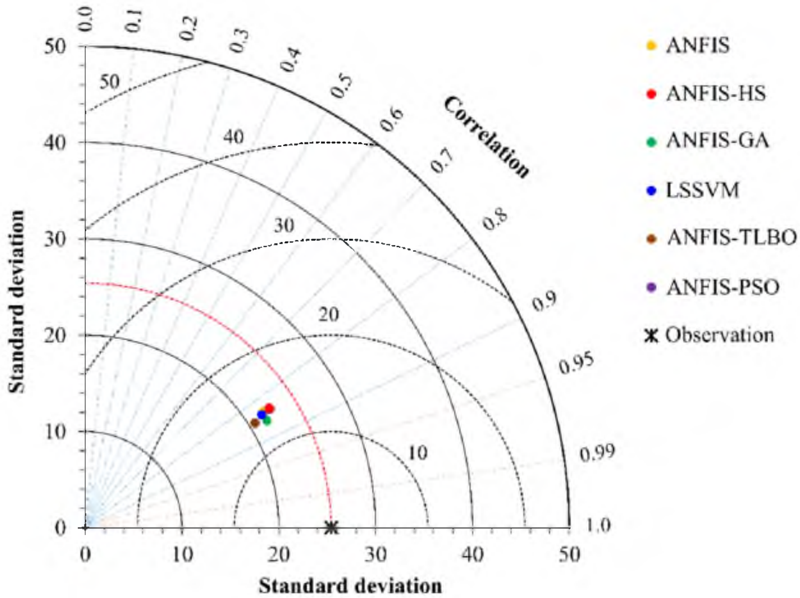

Figure 8. Taylor diagram of the predicted COD by ANFIS-based and LSSVM models using best input combination and data set during testing phase.

6. Discussion

The potential of four hybrid ANFIS methods was investigated in this study in estimating the chemical oxygen demand COD of the Yamuna River in Delhi, India, using monthly water quality parameters, total kjeldahl nitrogen, free ammonia, total coliform, water temperature, potential of hydrogen, and fecal coliform as inputs to the models. The outcomes of the implemented methods were compared with those of Kisi and Parmar [21].

The tables and figures revealed that the first input combination (AMM, TKN, and WT) provided the best accuracy in modelling COD, as reported in the previous study [21]. Bhardwaj and Parmar [55] reported that COD has high positive correlations with AMM (0.823) and TKN (0.741), and a negative correlation with WT (−0.273). Kora et al. [56] found no correlation between COD with TC and FC at Hussain Sagar Lake, Hyderabad, India. Kagalou et al. [57] studied the interrelationships between increased bacterial concentrations in near-bottom samples and an increase in TC and FC counts after precipitation. Evidence supports the idea that bacteria rely more on the source of pollution than the total organic load, indicating weak or negative relationships between bacteriological indices and BOD and COD levels.

It was observed from Table 4 that the best accuracy of the methods for the M2 data set, while the M1 data set resulted in the worst accuracy. This may be because of a different data range of M1 compared to M2 and M3 (see Table 1), which caused difficulties for the applied models in data extrapolation, as stated by Kisi and Parmar [21]. In addition, the training data were more skewed ($C_{sx} = -0.64$ and -0.24 for M2 and M3, respectively) than the test data ($C_{sx} = -0.08$) in this case.

The results showed that the accuracy of the models considerably fluctuated for these inputs. For example, the adaptive neuro-fuzzy inference system (ANFIS) showed an RMSE of 13.770 mg/L for M2, while it yielded 17.874 and 17.139 mg/L for M1 and M3, respectively. This indicates that the testing methods with only one data set may mislead the modeler about model performance. Therefore, cross-validation is very necessary for a robust evaluation of the methods.

Liu et al. [58] predicted COD using dynamic kernel extreme learning machine (DKELM) method, and compared it using partial least squares, ELM, dynamic ELM, and kernel ELM. The best model (DKELM) provided an R^2 of 0.7585 in the test stage. Sharafati et al. [59] used ada boost regression, gradient boost regression, and random forest regression for the prediction of COD; the highest correlation (R = 0.751) was found with the gradient boost regression. In the present study, the ANFIS–GA produced an $R^2 = 0.740$ (or R = 0.860), which is acceptable compared to that of previous studies.

The main limitation of the present study is the use of limited data. That data interval was monthly, and the available data period was very short. In order to justify the models' robustness and/or generalization capability, more data from different regions should be applied. It was clearly seen from the scatterplots that the hybrid methods could detect the extreme values well, and this can be explained by the limited number of training examples, especially for the COD extremes.

7. Conclusions

In the present study, the potential of four meta-heuristic-algorithm-integrated adaptive neuro-fuzzy inference system (ANFIS) models in estimating river water chemical oxygen demand (COD) was explored. The ability of hybrid neuro-fuzzy methods was investigated for different combinations of water quality (WQ) parameters, free ammonia (AMM), total kjeldahl nitrogen (TKN), water temperature (WT), total coliform (TC), fecal coliform (FC), and potential of hydrogen (PH) as inputs. Various input combinations were used by applying a cross-validation method, and the results were compared with the classical ANFIS and least square support vector machine (LSSVM) methods. The ANFIS comprising AMM, TKN, and WT input parameters provided the best accuracy in estimating monthly COD. The analysis outcomes revealed that employing meta-heuristic algorithms improved the accuracy of the classical ANFIS, and generally outperformed the LSSVM method in modelling COD. The ANFIS with harmony search algorithm provided the best COD estimates in terms of accuracy and computational time. The applications produced considerable fluctuations in estimations for the implemented models for three different data sets; this suggested the necessity of using cross-validation for better assessment of the applied methods.

The outcomes of this study led us to recommend the use of the hybrid neuro-fuzzy model tuned with harmony search algorithm for estimating the water quality of the Yamuna River, India. The results can be helpful for authorities and decision makers in managing water pollution in this region. The hybrid model developed in this study can be used to model COD, a vital WQ index, from AMM, TKN, and WT. The case study selected in the current study is important for India, as the selected river is the water source for 40% of the country's population. Measuring COD requires sample preparation and chemical pre-treatment, which are time-consuming and labor-intensive. The models developed in this study can be employed for estimating COD amounts at critical points of the river, which can be helpful for monitoring and controlling industrial and sewerage effects.

The developed models could not be generalized because only data from one site were available to assess them; this can be carried out in future studies using more data from other regions. The implemented methods can be applicable for other sites, but they require enough data and training. The models implemented by this study can be compared with other advanced methods, such as hybrid artificial neural networks, extreme learning machine, and deep learning models, in future studies using daily or monthly water quality data for longer durations.

Author Contributions: Conceptualization: R.M.A., O.K. and K.S.P.; formal analysis: M.Z.-K. and A.M.-M.; validation: O.K. and R.M.A.; supervision: O.K. and R.M.A.; writing—original draft: M.Z.-K., O.K., A.M.-M., K.S.P., O.K. and R.M.A.; visualization: M.Z.-K. and A.M.-M.; investigation: M.Z.-K., O.K., A.M.-M., K.S.P., O.K. and R.M.A.; writing—review and editing: O.K. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study will be available upon an interesting request from the corresponding author.

Conflicts of Interest: There are no conflict of interest in this study.

References

1. *A/RES/70/1*; United Nations General Assembly. United Nations: New York, NY, USA, 2015. Available online: https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcopact/A_RES_70_1_E.pdf (accessed on 2 June 2020).
2. Shah, M.I.; Alaloul, W.S.; Alqahtani, A.; Aldrees, A.; Musarat, M.A.; Javed, M.F. Predictive Modeling Approach for Surface Water Quality: Development and Comparison of Machine Learning Models. *Sustainability* **2021**, *13*, 7515. [CrossRef]
3. Soni, K.; Parmar, K.S.; Agarwal, S. Modeling of Air Pollution in Residential and Industrial Sites by Integrating Statistical and Daubechies Wavelet (Level 5) Analysis. *Model. Earth Syst. Environ.* **2017**, *3*, 1187–1198. [CrossRef]
4. Akoto, O.; Adiyiah, J. Chemical analysis of drinking water from some communities in the Brong A hafo region. *Int. J. Environ. Sci. Technol.* **2007**, *4*, 211–214. [CrossRef]
5. Alam, M.J.B.; Muyen, Z.; Islam, M.R.; Islam, S.; Mamun, M. Water quality parameters along rivers. *Int. J. Environ. Sci. Technol.* **2007**, *4*, 159–167. [CrossRef]
6. APHA. *Standard Methods for Examination of Water and Waste Water*; American Public Health Association: Washington, DC, USA, 1995.
7. WHO. *International Standards for Drinking Water*; World Health Organization: Geneva, Switzerland, 1971.
8. Rodríguez, R.; Pastorini, M.; Etcheverry, L.; Chreties, C.; Fossati, M.; Castro, A.; Gorgoglione, A. Water-Quality Data Imputation with a High Percentage of Missing Values: A Machine Learning Approach. *Sustainability* **2021**, *13*, 6318. [CrossRef]
9. Dong, Q.; Wang, Y.; Li, P. Ultifractal behavior of an air pollutant time series and the relevance to the predictability. *Environ. Pollut.* **2017**, *222*, 444–457. [CrossRef]
10. Bhardwaj, R.; Parmar, K.S. Water quality index and fractal dimension analysis of water Parameters. *Int. J. Environ. Sci. Technol.* **2013**, *10*, 151–164.
11. Wong, Y.J.; Shimizu, Y.; He, K.; Nik Sulaiman, N.M. Comparison among different ASEAN water quality indices for the assessment of the spatial variation of surface water quality in the Selangor river basin, Malaysia. *Environ. Monit. Assess.* **2020**, *192*, 644. [CrossRef]
12. Singh, S.; Parmar, K.S.; Kumar, J. Soft computing model coupled with statistical models to estimate future of stock market. *Neural Comput. Appl.* **2021**, *33*, 7629–7647. [CrossRef]
13. Li, X.; Peng, L.; Yao, X.; Cui, S.; Hu, Y.; You, C.; Chi, T. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environ. Pollut.* **2017**, *231*, 997–1004. [CrossRef]
14. Li, Y.; Jiang, P.; She, Q.; Lin, G. Research on air pollutant concentration prediction method based on self-adaptive neuro-fuzzy weighted extreme learning machine. *Environ. Pollut.* **2018**, *241*, 1115–1127. [CrossRef] [PubMed]
15. Sharma, P.; Said, Z.; Kumar, A.; Nižetić, S.; Pandey, A.; Hoang, A.T.; Huang, Z.; Afzal, A.; Li, C.; Le, A.T.; et al. Recent Advances in Machine Learning Research for Nanofluid-Based Heat Transfer in Renewable Energy System. *Energy Fuels* **2022**, *36*, 6626–6658. [CrossRef]
16. Yilma, M.; Kiflie, Z.; Windsperger, A.; Gessese, N. Application of artificial neural network in water quality index prediction: A case study in Little Akaki River, Addis Ababa, Ethiopia. *Model. Earth Syst. Environ.* **2018**, *4*, 175–187. [CrossRef]
17. Ahmed, A.N.; Othman, F.B.; Afan, H.A.; Ibrahim, R.K.; Fai, C.M.; Hossain, M.S.; Ehteram, M.; Elshafie, A. Machine learning methods for better water quality prediction. *J. Hydrol.* **2019**, *578*, 124084. [CrossRef]

18. Abba, S.I.; Pham, Q.B.; Saini, G.; Linh, N.T.T.; Ahmed, A.N.; Mohajane, M.; Khaledian, M.; Abdulkadir, R.A.; Bach, Q.V. Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index. *Environ. Sci. Pollut. Res.* **2020**, *27*, 41524–41539. [CrossRef]
19. Lee, E.; Kim, T. Predicting BOD under Various Hydrological Conditions in the Dongjin River Basin Using Physics-Based and Data-Driven Models. *Water* **2021**, *13*, 1383. [CrossRef]
20. Wong, Y.J.; Shimizu, Y.; Kamiya, A.; Maneechot, L.; Bharambe, K.P.; Fong, C.S.; Sulaiman, N.M.N. Application of artificial intelligence methods for monsoonal river classification in Selangor river basin, Malaysia. *Environ. Monit. Assess.* **2021**, *193*, 438. [CrossRef]
21. Kisi, O.; Parmar, K.S. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *J. Hydrol.* **2016**, *534*, 104–112. [CrossRef]
22. Adnan, R.M.; Liang, Z.; Trajkovic, S.; Zounemat-Kermani, M.; Li, B.; Kisi, O. Daily streamflow prediction using optimally pruned extreme learning machine. *J. Hydrol.* **2019**, *577*, 123981. [CrossRef]
23. Cheng, M.Y.; Cao, M.T. Evolutionary multivariate adaptive regression splines for estimating shear strength in reinforced-concrete deep beams. *Eng. Appl. Artif. Intell.* **2014**, *28*, 86–96. [CrossRef]
24. Alizamir, M.; Kisi, O.; Adnan, R.M.; Kuriqi, A. Modelling reference evapotranspiration by combining neuro-fuzzy and evolutionary strategies. *Acta Geophys.* **2020**, *68*, 1113–1126. [CrossRef]
25. Fadaee, M.; Amin Mahdavi-Meymand, A.; Zounemat-Kermani, M. Seasonal Short-Term Prediction of Dissolved Oxygen in Rivers via Nature-Inspired Algorithms. *CLEAN—Soil Air Water* **2020**, *48*, 1900300. [CrossRef]
26. Song, C.; Yao, L.; Hua, C.; Ni, Q. A water quality prediction model based on variational mode decomposition and the least squares support vector machine optimized by the sparrow search algorithm (VMD-SSA-LSSVM) of the Yangtze River, China. *Environ. Monit. Assess.* **2021**, *193*, 363. [CrossRef]
27. Arya Azar, N.; Milan, S.G.; Kayhomayoon, Z. The prediction of longitudinal dispersion coefficient in natural streams using LS-SVM and ANFIS optimized by Harris hawk optimization algorithm. *J. Contam. Hydrol.* **2021**, *240*, 103781. [CrossRef]
28. Maheshwaran, R.; Khosa, R. Long term forecasting of groundwater levels with evidence of non-stationary and nonlinear characteristics. *Comput. Geosci.* **2013**, *52*, 422–436. [CrossRef]
29. Emadi, A.; Zamanzad-Ghavidel, S.; Fazeli, S.; Zarei, S.; Rashid-Niaghi, A. Multivariate modeling of pan evaporation in monthly temporal resolution using a hybrid evolutionary data-driven method (case study: Urmia Lake and Gavkhouni basins). *Environ. Monit. Assess.* **2021**, *193*, 355. [CrossRef]
30. Ikram, R.M.A.; Dai, H.L.; Ewees, A.A.; Shiri, J.; Kisi, O.; Zounemat-Kermani, M. Application of improved version of multi verse optimizer algorithm for modeling solar radiation. *Energy Rep.* **2022**, *8*, 12063–12080.
31. Adnan, R.M.; Mostafa, R.R.; Islam, A.R.M.T.; Gorgij, A.D.; Kuriqi, A.; Kisi, O. Improving Drought Modeling Using HybridRandom Vector Functional Link Methods. *Water* **2021**, *13*, 3379. [CrossRef]
32. Sapankevych, N.I.; Sankar, R. Time series prediction using support vector machines: A survey. *Comput. Intell. Mag. IEEE* **2009**, *4*, 24–38. [CrossRef]
33. Shamir, E.; Megdal, S.B.; Carrillo, C.; Castro, C.L.; Chang, H.; Chief, K.; Corkhill, F.E.; Georgakakos, S.E.K.P.; Nelson, K.M.; Prietto, J. Climate change and water resources management in the Upper Santa Cruz River, Arizona. *J. Hydrol.* **2015**, *521*, 18–33. [CrossRef]
34. Shoorehdeli, M.A.; Teshnehlab, M.; Sedigh, A.K.; Khanesar, M.A. Identification using ANFIS with intelligent hybrid stable learning algorithm approaches and stability analysis of training methods. *Appl. Soft Comput.* **2009**, *9*, 833–850. [CrossRef]
35. Suykens, J.A.; De Brabanter, J.; Lukas, L.; Vandewalle, J. Weighted least squares support vector machines: Robustness and sparse approximation. *Neurocomputing* **2002**, *48*, 85–105. [CrossRef]
36. Jang, J.S. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. [CrossRef]
37. Kisi, O.; Shiri, J.; Karimi, S.; Adnan, R.M. Three different adaptive neuro fuzzy computing techniques for forecasting long-period daily streamflows. In *Big Data in Engineering Applications*; Springer: Singapore, 2018; pp. 303–321.
38. Kumar, V.; Yadav, S.M. Optimization of Reservoir Operation with a New Approach in Evolutionary Computation Using TLBO Algorithm and Jaya Algorithm. *Water Resour. Manag.* **2018**, *32*, 4375–4391. [CrossRef]
39. Ikram, R.M.A.; Mostafa, R.R.; Chen, Z.; Islam, A.R.M.T.; Kisi, O.; Kuriqi, A.; Zounemat-Kermani, M. Advanced Hybrid Metaheuristic Machine Learning Models Application for Reference Crop Evapotranspiration Prediction. *Agronomy* **2023**, *13*, 98. [CrossRef]
40. Adnan, R.M.; Mostafa, R.R.; Elbeltagi, A.; Yaseen, Z.M.; Shahid, S.; Kisi, O. Development of new machine learning model for streamflow prediction: Case studies in Pakistan. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 999–1033. [CrossRef]
41. Poli, R. Analysis of the publications on the applications of particle swarm optimization. *J. Artif. Evol. Appl.* **2008**, *2008*, 685175.
42. Chaganti, R.; Mourade, A.; Ravi, V.; Vemprala, N.; Dua, A.; Bhushan, B. A Particle Swarm Optimization and Deep Learning Approach for Intrusion Detection System in Internet of Medical Things. *Sustainability* **2022**, *14*, 12828. [CrossRef]
43. Dai, L.; Lu, H.; Hua, D.; Liu, X.; Chen, H.; Glowacz, A.; Królczyk, G.; Li, Z. A Novel Production Scheduling Approach Based on Improved Hybrid Genetic Algorithm. *Sustainability* **2022**, *14*, 11747. [CrossRef]
44. Mirjalili, S. Genetic algorithm. In *Evolutionary Algorithms and Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 43–55.

45. Katoch, S.; Chauhan, S.S.; Kumar, V. A review on genetic algorithm: Past, present, and future. *Multimed. Tools Appl.* **2021**, *80*, 8091–8126. [CrossRef]
46. Manjarres, D.; Landa-Torres, I.; Gil-Lopez, S.; Del Ser, J.; Bilbao, M.N.; Salcedo-Sanz, S.; Geem, Z.W. A survey on applications of the harmony search algorithm. *Eng. Appl. Artif. Intell.* **2013**, *26*, 1818–1831. [CrossRef]
47. Ocak, A.; Nigdeli, S.M.; Bekdaş, G.; Kim, S.; Geem, Z.W. Optimization of Seismic Base Isolation System Using Adaptive Harmony Search Algorithm. *Sustainability* **2022**, *14*, 7456. [CrossRef]
48. Abualigah, L.; Diabat, A.; Geem, Z.W. A comprehensive survey of the harmony search algorithm in clustering applications. *Appl. Sci.* **2020**, *10*, 3827. [CrossRef]
49. Rao, R.V.; Savsani, V.J.; Vakharia, D.P. Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems. *Comput.-Aided Des.* **2011**, *43*, 303–315. [CrossRef]
50. Sahu, R.K.; Shaw, B.; Nayak, J.R. Short/medium term solar power forecasting of Chhattisgarh state of India using modified TLBO optimized ELM. *Eng. Sci. Technol. Int. J.* **2021**, *24*, 1180–1200. [CrossRef]
51. Almutairi, K.; Algarni, S.; Alqahtani, T.; Moayedi, H.; Mosavi, A. A TLBO-Tuned Neural Processor for Predicting Heating Load in Residential Buildings. *Sustainability* **2022**, *14*, 5924. [CrossRef]
52. Gao, N.; Zhang, Z.; Tang, L.; Hou, H.; Chen, K. Optimal design of broadband quasi-perfect sound absorption of composite hybrid porous metamaterial using TLBO algorithm. *Appl. Acoust.* **2021**, *183*, 108296. [CrossRef]
53. Adnan, R.M.; Yuan, X.; Kisi, O.; Adnan, M.; Mehmood, A. Stream Flow Forecasting of Poorly Gauged MountainousWatershed by Least Square Support Vector Machine, Fuzzy Genetic Algorithm and M5 Model Tree Using Climatic Data from Nearby Station. *Water Resour. Manag.* **2018**, *32*, 4469–4486. [CrossRef]
54. Ikram, R.M.A.; Hazarika, B.B.; Gupta, D.; Heddam, S.; Kisi, O. Streamflow prediction in mountainous region using new machine learning and data preprocessing methods: A case study. *Neural Comput. Appl.* **2022**, 1–18. [CrossRef]
55. Bhardwaj, R.; Parmar, K.S. Wavelet and statistical analysis of river water quality parameters. *Appl. Math. Comput.* **2013**, *219*, 10172–10182.
56. Kora, A.J.; Rastogi, L.; Kumar, S.J.; Jagatap, B.N. Physico-chemical and bacteriological screening of Hussain Sagar lake: An urban wetland. *Water Sci.* **2017**, *31*, 24–33. [CrossRef]
57. Kagalou, I.; Tsimarakis, G.; Bezirtzoglou, E. Inter-relationships between bacteriological and chemical variations in Lake Pamvotis-Greece. *Microb. Ecol. Health Dis.* **2002**, *14*, 37–41.
58. Sharafati, A.; Asadollah, S.B.H.S.; Hosseinzadeh, M. The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty. *Process Saf. Environ. Prot.* **2020**, *140*, 68–78. [CrossRef]
59. Liu, H.; Zhang, Y.; Zhang, H. Prediction of effluent quality in papermaking wastewater treatment processes using dynamic kernel-based extreme learning machine. *Process Biochem.* **2020**, *97*, 72–79. [CrossRef]