KSCE Journal of Civil Engineering
DOI 10.1007/s12205-024-2613-z
pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Environmental Engineering

Comparing Artificial Neural Networks and Regression-based Methods for Modeling Daily Dissolved Oxygen Concentration: A Study Based on Long-term Monitored Data

Sinan Nacar[a], **Betul Mete**[b], **and Adem Bayram**[b]

[a]Dept. of Civil Engineering, Faculty of Engineering and Architecture, Tokat Gaziosmanpaşa University, Tokat 60150, Turkey
[b]Dept. of Civil Engineering, Faculty of Engineering, Karadeniz Technical University, Trabzon 61080, Turkey

ARTICLE HISTORY

Received 13 December 2023
Revised 23 April 2024
Accepted 4 June 2024
Published Online 05 August 2024

KEYWORDS

Artificial neural networks
Clackamas river
Dissolved oxygen concentration
Modeling
Multivariate adaptive regression splines
Water quality

ABSTRACT

In this study, the ability of regression-based methods, namely conventional regression analysis (CRA) and multivariate adaptive regression splines (MARS), and artificial neural networks (ANNs) method was investigated to model the river dissolved oxygen (DO) concentration. Daily average data for discharge and water-quality (WQ) indicators, which include DO concentration, temperature, specific conductance, and pH, were provided for the monitoring stations USGS 14210000 (upstream) and USGS 14211010 (downstream) in the Clackamas River, Oregon, USA. Eight models were established using different combinations of the input parameters and tested to determine the contribution of each parameter used in the modeling to the performance of the models. The results of the models and methods were compared with each other using several performance statistics. Although the performances of the methods were quite close to each other, the highest estimation performance was obtained from the ANNs method in the testing data sets. According to the performance statistics, Model 8, in which all WQ indicators were included as input parameters, was selected as the optimal model to estimate DO concentration of different periods of the same stations. However, when estimating the DO concentration from one station to another, the highest performance statistics were obtained from Model 8 for upstream and Model 1 for downstream station using the CRA method. For the ANNs method, Model 1 having the single input for both stations was the best model.

1. Introduction

Monitoring and controlling of water-quality (WQ) indicators are of vital importance. Dissolved oxygen (DO) is a significant indicator of the health and productivity of aquatic ecosystems. It is essential for the survival of fish and other aquatic organisms and plays a key role in the decomposition of organic matter. Low DO concentrations can lead to fish kills and the proliferation of harmful algal blooms, while high DO concentrations can indicate an over-enriched ecosystem. Therefore, measuring DO concentration can provide valuable information about the overall health and functioning of an aquatic ecosystem. Additionally, it can help identify potential pollution sources and aid in the conservation and management of aquatic resources (Nas et al., 2008; Csabragi et al., 2017). DO concentration changes depending on various factors such as water temperature (WT), pH, conductivity, and salinity, as well as atmospheric pressure. The solubility of oxygen is inversely proportional to temperature. In other words, the solubility of oxygen decreases as WT increases. Because WT and pH have an inverse proportion, pH values could be used to estimate DO concentrations. Lower conductivity and salinity and higher atmospheric pressure are the factors that increase DO concentration in water (Kalff, 2002). On the other hand, hydraulic structures impact the amount of DO in a river system (Bayram and Kankal, 2015).

The impacts of anaerobic conditions caused by low DO concentrations in a river system generate unstable ecosystem with fish reproductive problems and mortalities, pollution, odor, and other unaesthetic situations (Cox, 2003). In addition, DO

CORRESPONDENCE Sinan Nacar sinannacar@hotmail.com Dept. of Civil Engineering, Faculty of Engineering and Architecture, Tokat Gaziosmanpaşa University, Tokat 60150, Turkey

concentration reveal more about aquatic systems than any other single parameter. Therefore, monitoring DO concentrations is important in most surface waters and near-shore coastal systems. There is a need for cost-efficient and effective methods that can estimate DO concentration using limited monitoring data due to the reasons, i.e., resource deficiencies, budget constraints, and difficulties of constantly monitoring DO changes (Money et al., 2009). Thus, modeling DO concentration using readily measurable and available variables has become an essential scientific advantage.

Recently, modeling river DO concentration has attracted considerable attention in the literature. Many researchers have tried to estimate DO concentration in surface waters using different methods such as artificial neural networks (ANNs) (Rajaee et al., 2020; Yang et al., 2021; Bhardwaj and Singh, 2022), fuzzy logic and neuro-fuzzy (Ay and Kisi, 2017; Arora and Keshari, 2021), evolutionary models (Sedighkia et al., 2021; Li et al., 2023), and support vector machine (Song et al., 2021; Nong et al., 2023). In addition, the use of combinations of different methods has been suggested in the literature (Tiyasha et al., 2021; Li et al., 2023). Until now, the multivariate adaptive regression splines (MARS) method has been successfully used by researchers in the modeling of various engineering problems such as predicting pile drivability (Zhang and Goh, 2016), estimating pan evaporation (Kisi, 2015), modeling air pollutants (Kisi et al., 2017), and estimating suspended sediment load (Yilmaz et al., 2018). The MARS method was firstly used by Heddam and Kisi (2018) for modeling daily DO concentrations together with a version of the support vector machine and Model 5 tree methods.

Regression-based methods, especially linear regression commonly used in DO modeling studies, can provide a strong starting point when model is sufficiently simple, whereas ANNs method can more flexible and suitable for modeling complex data structures (Ghahramani, 2015). The present study focuses on constructing different regression-based methods, namely conventional regression analysis (CRA) and MARS, and ANNs to model DO concentration in the Clackamas River, Oregon. The main reason for selecting the Clackamas River Basin as the study area is the lack of a comprehensive WQ modeling study previously conducted in this basin. The methods were selected to ensure the reliability and comparability of the analysis results. Furthermore, comparing the performance of different modeling methods on DO estimation, one of the main indicators of WQ, increases the value of this study. The literature review reveals that no such study has been conducted so far. This highlights the uniqueness of this modeling study on the WQ of the Clackamas River. The objectives of this study are to (i) compare the ability of ANNs and regression-based methods method to estimate the river DO concentration, (ii) investigate the role of WQ variables in DO estimation, and (iii) investigate whether the model established for a particular station can be used for another station in the basin. The methods used are important for monitoring and management of WQ in the basin, and it is thought that a comparison study, which has not yet been carried out in this field, will contribute to practice.

2. Material and Methods

2.1 Study Area

The Clackamas River Basin is located in the Pacific Northwest Region, one of the 22 hydrological regions in the USA. The river has approximately 134 km length main branch and 2,434 km^2 drainage area. The river is home to various fish species and provides hydroelectric power and drinking water to Portland metropolitan area (Edrington, 1993). Precipitation and soil erosion cause water-

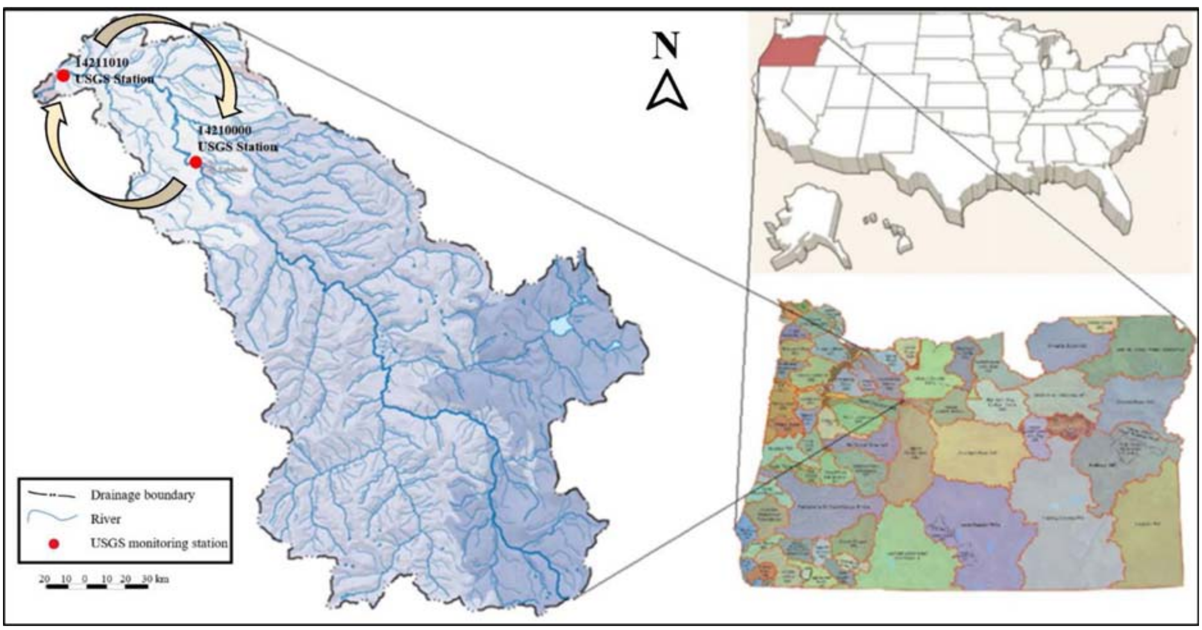

Fig. 1. The Locations of the Water-Quality Monitoring Stations Operated by USGS in the Clackamas River Basin, Oregon, USA (Lee, 2011)

Table 1. Descriptive Statistics for the River Water-Quality Indicators Monitored at the USGS 14210000 and USGS 14211010 Stations on the Clackamas River, for the Training, Validating, Testing, and All Data Sets

Data Sets	Water-quality indicators	USGS 14210000						USGS 14211010					
		X_{min}	X_{ave}	X_{max}	S_x	C_{sx}	C_{kx}	X_{min}	X_{ave}	X_{max}	S_x	C_{sx}	C_{kx}
Training	DO (mg/L)	8.09	11.12	14.36	1.54	-0.14	-1.27	8.16	11.13	14.82	1.41	-0.12	-1.08
	WT (°C)	0.36	10.14	20.59	4.90	0.30	-1.15	0.23	11.48	23.61	5.59	0.36	-1.12
	EC (µS/cm)	18.04	37.69	64.02	13.13	0.47	-1.31	20.22	40.41	69.22	13.72	0.54	-1.24
	Q (ft^3/s)	650.57	2,694.47	23,875.78	2,430.22	2.82	12.11	624.36	3,363.32	27,515.38	3,183.89	2.53	9.17
	pH	6.97	7.45	8.36	0.13	0.56	3.79	6.77	7.48	8.09	0.23	-0.46	-0.48
Validating	DO (mg/L)	8.69	11.23	13.83	1.29	-0.08	-1.24	8.58	11.27	13.66	1.36	-0.36	-1.21
	WT (°C)	2.17	9.69	18.99	4.63	0.43	-1.15	3.14	10.95	22.09	5.25	0.50	-1.10
	EC (µS/cm)	18.62	39.18	60.48	11.75	0.47	-1.12	20.63	41.75	66.58	12.57	0.58	-0.98
	Q (ft^3/s)	548.68	1,997.62	15,159.08	1,666.46	3.31	15.72	709.54	2,547.67	20,303.74	2,262.37	3.31	16.40
	pH	7.10	7.48	7.79	0.13	0.18	-0.89	7.20	7.60	8.11	0.18	-0.30	-0.77
Testing	DO (mg/L)	8.32	11.26	14.72	1.57	-0.30	-1.26	8.20	11.08	13.67	1.42	-0.44	-1.12
	WT (°C)	2.48	10.25	20.32	5.05	0.58	-1.02	3.44	11.55	23.92	5.66	0.62	-0.98
	EC (µS/cm)	19.74	38.93	64.50	12.83	0.64	-1.16	22.48	42.15	71.38	13.77	0.67	-1.12
	Q (ft^3/s)	617.07	2,665.85	23,506.91	2,468.21	3.06	13.73	700.61	3,435.58	29,801.75	3,208.05	2.88	12.88
	pH	7.04	7.48	7.65	0.09	-0.75	0.82	7.01	7.51	8.01	0.21	-0.09	-0.60
All data	DO (mg/L)	8.09	11.17	14.72	1.50	-0.17	-1.24	8.16	11.15	14.82	1.41	-0.23	-1.11
	WT (°C)	0.36	10.08	20.59	4.89	0.39	-1.10	0.23	11.39	23.92	5.54	0.44	-1.07
	EC (µS/cm)	18.04	38.23	64.50	12.83	0.50	-1.24	20.22	41.02	71.38	13.54	0.57	-1.16
	Q (ft^3/s)	548.68	2,555.15	23,875.78	2,327.22	2.99	13.51	624.36	3,221.90	29,801.75	3,051.47	2.74	11.08
	pH	6.97	7.46	8.36	0.13	0.31	2.56	6.77	7.51	8.11	0.22	-0.46	-0.34

X_{min}: minimum, X_{ave}: mean, X_{max}: maximum, S_x: standard deviation, C_{sx}: skewness coefficient, and C_{kx}: kurtosis coefficient

quality problems such as high turbidity levels in the upper basin with steep topography and geologic instability. Moreover, the contaminants that emerged from the results of human activities (fishing, hiking, camping, hunting, etc.) have impacts on the river WQ in the lower basin (Carpenter, 2003).

The data used in this study are daily average values of measured and recorded data at thirty-minute intervals for each monitoring station operated by United States Geological Survey (USGS), namely USGS 14210000 (latitude 45°18'00" N and longitude 122°21'10" E) and USGS 14211010 (latitude 45°22'46" N and longitude 122°34'34" E) in the Clackamas River, Oregon, USA. The upstream station USGS 14210000 is in a relatively undisturbed and sparsely populated area, while the downstream station USGS 14211010 near the city of Oregon is located in a more densely populated area. The locations of the monitoring stations are illustrated in Fig. 1.

2.2 Data Sets and Modeling Applications

To decide input parameters is very important in modeling studies (Nacar et al., 2020a). Various water-quality indicators have been considered for modeling river DO in the literature (Table S1). In this study, monitoring stations USGS 14210000 and USGS 14211010 located in the Clackamas River Basin were taken into account. Considering the hydro-chemical indicators measured in the monitoring stations and the input parameters frequently preferred in the literature, it was decided to consider WT, pH, SC, and discharge (Q) for this study. The hydro-chemical data were downloaded from the USGS website. Electrical conductivity (EC) values corresponding to SC values were computed.

There were 3586 daily average data (from October 2012 to September 2022) for each indicator. The data were divided into three sets, training (six years from October 2012 to September 2018), validating (two years from October 2018 to September 2020), and testing (two years from October 2020 to September 2022). Thus, it was tested whether the function that gave good estimation or not for the training data set gave satisfactory estimations for another data set as well. Descriptive statistics of the daily monitored indicators for training, validating, testing, and all data are given in Table 1. The expressions X_{min}, X_{ave}, X_{max}, S_x, C_{sx}, and C_{kx} indicate the minimum, average, maximum, standard deviation, skewness coefficient, and kurtosis coefficient, respectively. Many statistical tests require normally distributed data. Neglecting to test for normality can affect the results of their findings (Matore and Khairani, 2020). Skewness and kurtosis quantitatively indicate non-normal variation in a statistical series. Skewness refers to the asymmetry of the curve, while kurtosis refers to the height or flatness of the curve (Blanca et al., 2013). Skewness and kurtosis values between -3 and 3 are considered normal by Peat and Barton (2005). When the C_{sx}, and C_{kx} values are considered, it is seen that the Q data do not fit the normal distribution. For this

Table 2. The Input Combinations of Different Dissolved Oxygen Models

Models	Independent variables			
M1	WT			
M2	WT	EC		
M3	WT	Q		
M4	WT	pH		
M5	WT	EC	Q	
M6	WT	EC	pH	
M7	WT	Q	pH	
M8	WT	EC	Q	pH

reason, log transformation was applied to Q data. Additionally, correlations between water-quality data were also examined. The Spearman correlation is a non-parametric measure used to assess the strength and direction of association between two variables. It does not rely on assumptions about the distribution of the variables, making it applicable to a wide range of data types. The Spearman correlation coefficient ranges from -1 to 1. A value of 1 (high) indicates a perfect positive relationship, -1 (high) indicates a perfect negative relationship, and 0 (low) indicates no relationship between the variables (Hauke and Kossowski, 2011).

The Spearman correlations between water-quality data are given in Table S2. The hydro-chemical indicators with the highest and lowest correlations with DO concentration for both stations are WT and pH, respectively. Eight models were established considering the correlations to determine the contribution of the parameters to be used in the modeling of DO concentration to the model performance. The univariate model was established using the river WT data, which are highly correlated with the DO concentration data. Other models were established by including different combinations of the river EC, Q, and pH in this univariate model. In this way, all models having one to four variables were tested. Those models are given in Table 2.

It may be challenging to model extreme values in data sets. Therefore, all data were normalized to minimize the effects of the size differences of the parameters in the data set on the modeling ability and to increase the model performances. Despite different normalization formulas in the literature (Dawson and Wilby, 1998; Bayram et al., 2012; Fetene et al., 2018), there are no definitive rules as to which procedure will be applied under what conditions. The most used transfer function for neurons in the hidden and output layers of an ANN has a limited output range between 0 and 1. For this reason, the data must be normalized so as to stay within the limited output range (Rajurkar et al., 2002). In this study, Eq. (1), in which X_n, X_i, X_{min}, and X_{max} stand for the normalized, raw, minimum, and maximum values, respectively, was used to normalize the data belonging to the river water-quality indicators. Normalized values close to the limit values cause the training of the network to slow down and become ineffective (Van Ooyen and Nichhuis, 1992). Therefore, the coefficients a and b were selected as 0.8 and 0.1, respectively. Thus, all data were normalized between 0.1 and 0.9.

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} a + b , \tag{1}$$

2.3 Conventional Regression Analysis (CRA)

Regression analysis includes determining whether there is a significant relationship between two or more variables, if there is a significant relationship, expressing this relationship mathematically, and determining the confidence intervals of predictions using equations (Bayazıt, 1981). In this study, four types of regression functions, which are linear (LF), power (PF), exponential (EF), and quadratic (QF), were used to model the river DO concentrations. The regression coefficients of the functions were obtained by using the least-squares method, which is a standard optimization method used to estimate the coefficients of the regression model, with IBM SPSS statistics 26 software. The least-squares method tries to minimize the sum of squared errors between the observations in the data set and the values estimated by the function.

2.4 Multivariate Adaptive Regression Splines (MARS)

MARS, introduced by Friedman (1991), is a form of nonparametric, flexible, and rapid regression method that does not make assumptions about the functional relationship between dependent and independent variables. This method has included basis functions and coefficients based on the available data. It divides the values of the independent variables into regions and explains each region with a regression equation. In addition, the independent variable is estimated by the contributions of both dependent variables and basis functions. In the MARS method, there is a two-stage process that continues until the best model is obtained. In the first stage, all possible basis functions are created. The model is developed with the basis functions that are added until the complexity of the model reaches its maximum level. In the second stage, the basis functions are eliminated one by one from the maximum model to achieve the optimum model. A general MARS model can be defined by Eq. (2), in which the n, N, X, a_n, β_0, $\beta_n(X_t)$, and ε_i stand for the number of knots, number of basis functions, argument, nst coefficient of the basis function, constant term in the model, n for the argument of tst basis function, and disturbance, respectively.

$$Y = \beta_0 + \sum_{n=1}^{N} a_n \beta_n (X_t) + \varepsilon_i , \tag{2}$$

2.5 Artificial Neural Networks (ANNs)

An ANN is a type of machine learning model that is inspired by the structure and function of the human brain. The ANNs are widely used to model complex and non-linear relationships, especially when the explicit form of the relationship between variables is unknown (Kohonen, 1988). An ANN architecture typically consists of three main layers, an input layer, a set of hidden layers and an output layer, which are connected to each other by neurons (Singh et al., 2009). The advantages of ANNs are that they make no assumptions about the structure of the data set and the models are developed based on the training algorithm (Rezaei et al., 2009).

2.6 Model Performance Statistics

To determine the most capable method used in the study, root mean square error (RMSE), Willmott's index of agreement (d), and Nash-Sutcliffe efficiency coefficient (NSEC) were used as performance statistics. Lower RMSE values indicate higher model performances (Moriasi et al., 2007). A statistical measure used to evaluate the degree of agreement between two data sets, d is commonly used in the fields of hydrology, meteorology, and environmental science. It ranges between 0 and 1, with 1 indicating perfect agreement between the two data sets, and 0 indicating no agreement at all (Willmott, 1981). A NSEC less than 0 occurs when the monitored mean is a better predictor than the model. As NSEC approaches 1, the model accuracy increases (Singh et al., 2005). While the NSEC primarily focuses on predictive performance relative to the mean of the monitored data, d focuses on both mean bias and scatter. The performance statistics are calculated by Eqs. (3) – (5) given below:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(t_i - td_i\right)^2}\ , \tag{3}$$

$$\mathrm{d} = 1 - \frac{\sum_{i=1}^{N}(t_i - td_i)^2}{\sum_{i=1}^{N}\left(\left|td_i - \bar{t}\right| + \left|t_i - \bar{t}\right|\right)^2}\ , \tag{4}$$

$$\mathrm{NSEC} = 1 - \frac{\sum_{i=1}^{N}(t_i - td_i)^2}{\sum_{i=1}^{N}(t_i - \bar{t})^2}\ , \tag{5}$$

where N, t_i, td_i, and $\bar{t}$, stand for data number, monitored and estimated values, and average of measured values, respectively.

3. Results

In this section, the performances of the estimation methods were investigated in modeling the river DO concentration. For each monitoring station, eight models were constituted using particular combinations of the hydro-chemical indicators, WT, pH, EC, and Q, as input parameter for modeling the river DO concentration. The performance statistics of the models and methods were

Table 3. The Performance Statistics of the CRA, MARS, and ANNs Methods Concerning the Training, Validating, and Testing Phases for the USGS 14210000 Station on the Clackamas River

Models	Statistics	Training						Validating						Testing					
		LF	PF	EF	QF	MARS	ANNs	LF	PF	EF	QF	MARS	ANNs	LF	PF	EF	QF	MARS	ANNs
M1	RMSE	0.357	0.702	0.353	0.353	0.327	0.347	0.447	0.494	0.440	0.439	0.444	0.427	0.414	0.689	0.402	0.400	0.392	0.374
	NSEC	0.946	0.792	0.947	0.947	0.955	0.949	0.880	0.854	0.884	0.884	0.882	0.891	0.930	0.806	0.934	0.935	0.941	0.946
	d	0.986	0.933	0.986	0.986	0.988	0.986	0.972	0.956	0.973	0.973	0.973	0.974	0.982	0.929	0.983	0.983	0.985	0.985
M2	RMSE	0.309	0.405	0.267	0.257	0.240	0.265	0.414	0.450	0.374	0.340	0.348	0.331	0.392	0.569	0.363	0.348	0.387	0.328
	NSEC	0.960	0.931	0.970	0.972	0.976	0.970	0.897	0.878	0.916	0.931	0.927	0.934	0.937	0.868	0.946	0.950	0.943	0.959
	d	0.990	0.981	0.992	0.993	0.994	0.992	0.975	0.966	0.980	0.983	0.982	0.984	0.984	0.958	0.986	0.987	0.984	0.989
M3	RMSE	0.254	0.347	0.214	0.203	0.178	0.264	0.370	0.503	0.328	0.329	0.304	0.264	0.303	0.406	0.247	0.256	0.284	0.214
	NSEC	0.973	0.949	0.981	0.983	0.987	0.971	0.918	0.848	0.935	0.935	0.945	0.958	0.963	0.933	0.975	0.973	0.969	0.982
	d	0.993	0.986	0.995	0.996	0.997	0.992	0.980	0.960	0.985	0.985	0.987	0.989	0.991	0.981	0.994	0.993	0.992	0.995
M4	RMSE	0.355	0.687	0.350	0.342	0.318	0.324	0.450	0.443	0.439	0.402	0.395	0.389	0.414	0.672	0.397	0.383	0.370	0.365
	NSEC	0.947	0.800	0.948	0.951	0.957	0.956	0.879	0.882	0.884	0.903	0.907	0.909	0.930	0.816	0.936	0.940	0.948	0.949
	d	0.986	0.937	0.986	0.987	0.989	0.988	0.972	0.965	0.973	0.977	0.978	0.978	0.982	0.934	0.983	0.984	0.986	0.986
M5	RMSE	0.251	0.339	0.213	0.198	0.168	0.247	0.363	0.491	0.331	0.328	0.348	0.281	0.290	0.429	0.254	0.272	0.313	0.202
	NSEC	0.973	0.951	0.981	0.984	0.988	0.974	0.921	0.855	0.934	0.935	0.927	0.953	0.966	0.925	0.974	0.970	0.962	0.984
	d	0.993	0.987	0.995	0.996	0.997	0.993	0.981	0.962	0.984	0.985	0.983	0.988	0.991	0.978	0.993	0.992	0.990	0.996
M6	RMSE	0.308	0.398	0.267	0.251	0.234	0.275	0.408	0.451	0.373	0.316	0.331	0.302	0.393	0.569	0.363	0.340	0.395	0.308
	NSEC	0.960	0.933	0.970	0.973	0.977	0.968	0.900	0.878	0.916	0.940	0.934	0.945	0.937	0.868	0.946	0.953	0.940	0.964
	d	0.990	0.982	0.992	0.993	0.994	0.992	0.976	0.966	0.980	0.985	0.984	0.986	0.984	0.957	0.985	0.987	0.984	0.990
M7	RMSE	0.250	0.346	0.213	0.196	0.173	0.243	0.353	0.502	0.323	0.302	0.276	0.266	0.298	0.405	0.245	0.249	0.278	0.182
	NSEC	0.974	0.949	0.981	0.984	0.987	0.975	0.925	0.849	0.937	***0.945***	***0.954***	***0.958***	0.964	0.933	***0.975***	0.975	***0.970***	***0.987***
	d	0.993	0.987	0.995	0.996	0.997	0.994	0.982	0.960	0.985	***0.987***	***0.989***	***0.990***	0.991	0.981	***0.994***	0.993	***0.992***	***0.997***
M8	RMSE	0.652	0.336	0.212	0.189	0.162	0.218	0.596	0.488	0.326	0.303	0.305	0.276	0.763	0.430	0.253	0.265	0.311	0.216
	NSEC	0.820	0.952	0.981	***0.985***	***0.989***	***0.980***	0.787	0.857	0.936	0.945	0.944	0.954	0.762	0.924	0.974	0.971	0.963	0.982
	d	0.943	0.987	0.995	***0.996***	***0.997***	***0.995***	0.939	0.962	0.985	0.987	0.986	0.989	0.927	0.978	0.993	0.992	0.990	0.995

Bold italic: the highest values,
The unit of RMSE: mg/L

Table 4. The Performance Statistics of the CRA, MARS, and ANNs Methods Concerning the Training, Validating, and Testing Phases for the USGS 14211010 Station on the Clackamas River

Models	Statistics	Training						Validating						Testing					
		LF	PF	EF	QF	MARS	ANNs	LF	PF	EF	QF	MARS	ANNs	LF	PF	EF	QF	MARS	ANNs
M1	RMSE	0.266	0.495	0.240	0.241	0.234	0.245	0.217	0.428	0.206	0.206	0.209	0.201	0.213	0.490	0.200	0.200	0.213	0.205
	NSEC	0.965	0.877	0.971	0.971	0.973	0.970	0.974	0.901	0.977	0.977	0.976	0.978	0.978	0.881	0.980	0.980	0.979	0.981
	d	0.991	0.964	0.993	0.993	0.993	0.992	0.993	0.969	0.994	0.994	0.994	0.994	0.994	0.960	0.995	0.995	0.995	0.995
M2	RMSE	0.265	0.387	0.233	0.225	0.214	0.227	0.215	0.479	0.208	0.199	0.240	0.194	0.211	0.452	0.198	0.206	0.232	0.209
	NSEC	0.965	0.925	0.973	0.975	0.977	0.974	0.975	0.876	0.977	0.979	0.969	0.980	0.978	0.899	0.981	0.979	0.975	0.980
	d	0.991	0.980	0.993	0.994	0.994	0.993	0.993	0.962	0.994	0.994	0.992	0.995	0.994	0.968	0.995	0.994	0.993	0.995
M3	RMSE	0.266	0.388	0.239	0.234	0.224	0.229	0.217	0.488	0.207	0.194	0.195	0.193	0.213	0.419	0.201	0.206	0.230	0.204
	NSEC	0.965	0.925	0.971	0.973	0.975	0.974	0.975	0.871	0.977	0.980	0.980	0.980	0.978	0.913	0.980	0.979	0.976	0.981
	d	0.991	0.979	0.993	0.993	0.994	0.993	0.993	0.960	0.994	0.995	0.995	0.995	0.994	0.974	0.995	0.994	0.994	0.995
M4	RMSE	0.266	0.492	0.238	0.236	0.222	0.240	0.214	0.449	0.197	0.191	0.203	0.190	0.211	0.501	0.195	0.193	0.205	0.193
	NSEC	0.965	0.879	0.972	0.972	0.975	0.971	0.975	0.891	0.979	0.980	0.978	0.980	0.978	0.876	0.981	0.982	0.981	0.983
	d	0.991	0.965	0.993	0.993	0.994	0.993	0.994	0.965	0.995	0.995	0.994	0.995	0.994	0.958	0.995	0.995	0.995	0.996
M5	RMSE	0.263	0.379	0.228	0.218	0.210	0.218	0.207	0.487	0.196	0.192	0.221	0.187	0.208	0.426	0.195	0.195	0.223	0.199
	NSEC	0.965	0.928	0.974	0.976	0.978	0.976	0.977	0.872	0.979	0.980	0.974	0.981	0.979	0.910	0.981	0.981	0.977	0.982
	d	0.991	0.980	0.993	0.994	0.994	0.994	0.994	0.961	0.995	0.995	0.993	0.995	0.995	0.972	0.995	0.995	0.994	0.995
M6	RMSE	0.265	0.377	0.226	0.212	0.197	0.210	0.208	0.437	0.176	0.168	0.196	0.164	0.207	0.426	0.180	0.179	0.188	0.169
	NSEC	0.965	0.929	0.974	0.977	0.981	***0.978***	0.977	0.897	0.983	***0.985***	0.979	0.985	0.979	0.910	0.984	0.984	***0.984***	***0.987***
	d	0.991	0.981	0.993	0.994	0.995	***0.994***	0.994	0.969	0.996	***0.996***	0.994	0.996	0.995	0.972	0.996	0.996	***0.996***	***0.997***
M7	RMSE	0.266	0.380	0.236	0.224	0.207	0.222	0.214	0.451	0.191	0.175	0.194	0.169	0.212	0.391	0.192	0.185	0.205	0.187
	NSEC	0.965	0.928	0.972	0.975	0.979	0.975	0.975	0.890	0.980	0.984	0.980	0.985	0.978	0.925	0.982	0.983	0.981	0.984
	d	0.991	0.980	0.993	0.994	0.995	0.994	0.994	0.967	0.995	0.996	0.995	0.996	0.994	0.978	0.995	0.996	0.995	0.996
M8	RMSE	0.312	0.368	0.224	0.207	0.192	0.216	0.237	0.443	0.175	0.170	0.184	0.158	0.272	0.393	0.181	0.172	0.192	0.170
	NSEC	0.951	0.932	0.975	***0.979***	***0.982***	0.977	0.970	0.894	0.983	0.984	***0.982***	***0.986***	0.963	0.924	0.984	***0.985***	0.983	0.987
	d	0.988	0.982	0.994	***0.995***	***0.995***	0.994	0.992	0.968	0.996	0.996	***0.995***	***0.996***	0.991	0.977	0.996	***0.996***	0.996	0.997

Bold italic: the highest values,
The unit of RMSE: mg/L

compared with each other to determine the model and method with the highest performance. Firstly, the CRA method was applied to the river water-quality data, and the optimum values for the coefficients of different regression forms were determined. The optimum coefficients obtained for each model and function as a result of the analysis are given in Tables S3 and S4 for upstream and downstream monitoring stations, respectively. In the next step, the MARS and ANNs methods were implemented. For the MARS method, the equations were determined using the basis functions. The basis functions and equations of the MARS method are presented in Tables S5 – S7 for the same monitoring stations, respectively.

The estimation methods were compared with each other based on their performance statistics in training, validating, and testing sets. It was determined that there were differences between the performances of the models for all data sets. However, it was observed that each model had acceptable performance statistics, which are given in Tables 3 and 4 for the monitoring stations, respectively. For both stations, the most appropriate performance statistics were obtained from QF for training and validating data sets, and EF and QF for the testing data sets in the CRA method. Since the performance statistics values of EF and QF were very close to each other in the testing data sets, QF was selected as the best function of the CRA method and its results were considered in the evaluation of the models.

For upstream station, the RMSE range was 0.189 – 0.439 mg/L for CRA, 0.162 – 0.444 mg/L for MARS, and 0.182 – 0.427 mg/L for ANNs. The NSEC values changed 0.884 – 0.985, 0.882 – 0.989, and 0.891 – 0.987 for the same methods, respectively. The d values changed 0.973 – 0.996, 0.973 – 0.997, and 0.974 – 0.997 for the same methods, respectively. The lowest RMSE and highest NSEC and d values were obtained from Model 8 for training phase by the MARS method, and Model 7 for validating and testing phases by the ANNs method.

For downstream station, the RMSE range was 0.168 – 0.241 mg/L for CRA, 0.184 – 0.240 mg/L for MARS, and 0.158 – 0.245 mg/L for ANNs. The NSEC values changed 0.971 – 0.985, 0.969 – 0.984, and 0.970 – 0.987 for the same methods, respectively. The d values changed 0.993 – 0.996, 0.992 – 0.996, and 0.992 – 0.997 for the same methods, respectively. The lowest

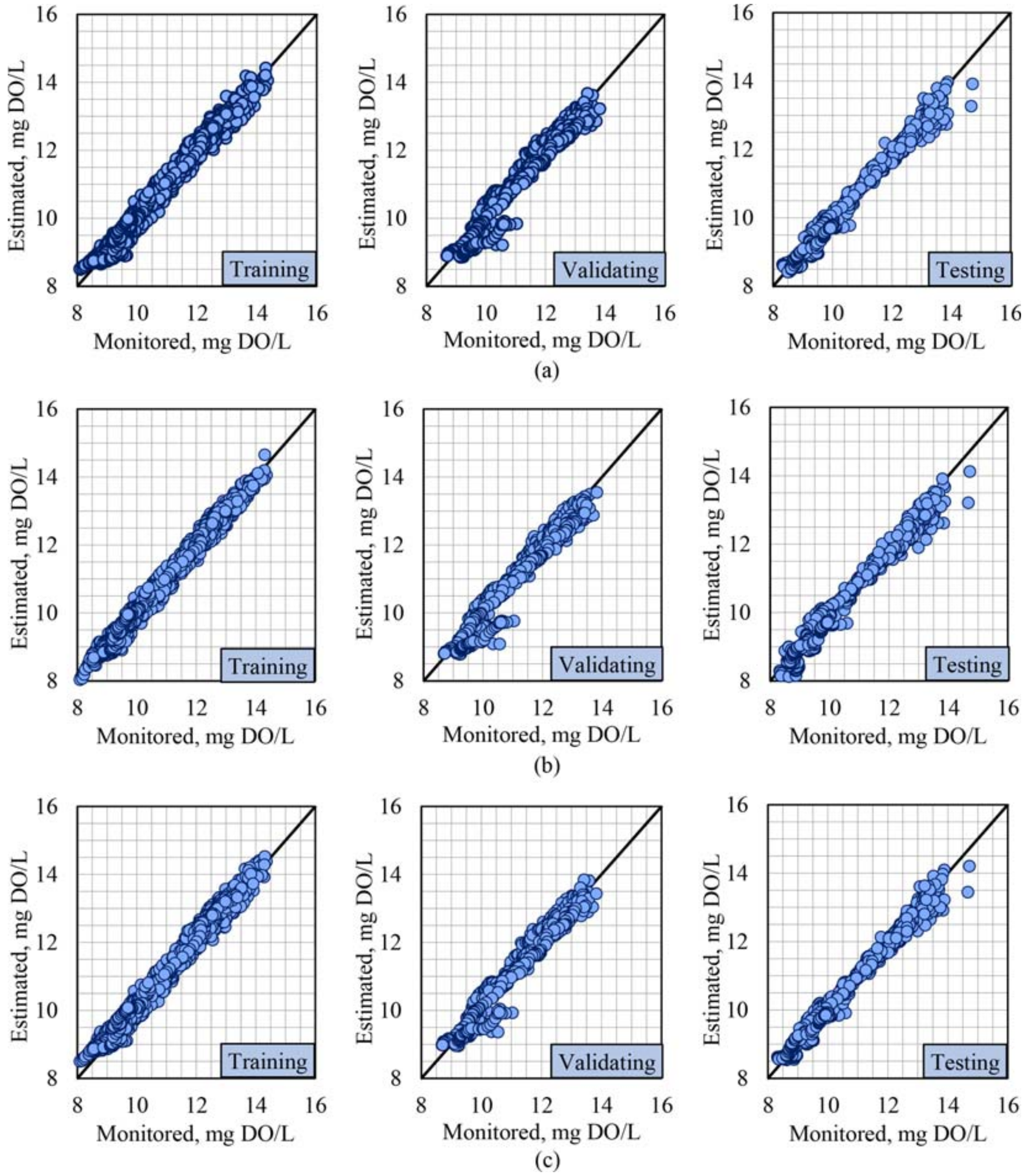

Fig. 2. The Scatter Plots of the River DO Concentrations Monitored versus Estimated by Model 8 Using the Methods: (a) CRA_QF, (b) MARS, (c) ANNs for USGS 14210000

RMSE and highest NSEC and d values were obtained from Model 8 for training phase by the MARS method, Model 8 for validating phase by the ANNs method, and Model 6 for testing phases by the ANNs method.

It is seen from Tables 3 and 4 that the performance statistics obtained from Models 7 and 8 for upstream station and Models 6 and 8 for downstream station are close to each other, and the river DO concentration can be successfully modeled for all methods. Based on the results, the optimal model for both stations was selected as Model 8, the combination using all hydro-chemical indicators as input. For the optimal model, the graphical results in the training, validating, and testing phases are given in the forms of scatter plots (Figs. 2 and 3). As the estimated and monitored values get closer to each other, the scattering becomes intense around the diagonal. The relative error on the diagonal is zero. It is seen from Figs. 2 and 3 that the estimated values are very close to the measured data.

Eight models established with data from October 2012 to September 2018 data for upstream station were tested with the same period data for downstream station. Similarly, the models built for downstream station were tested with the same period data for upstream station. The performance statistics calculated for the test results of the stations related to these studies are given in Table 5.

The most appropriate performance statistics of the CRA method were obtained from the LF and QF for upstream station

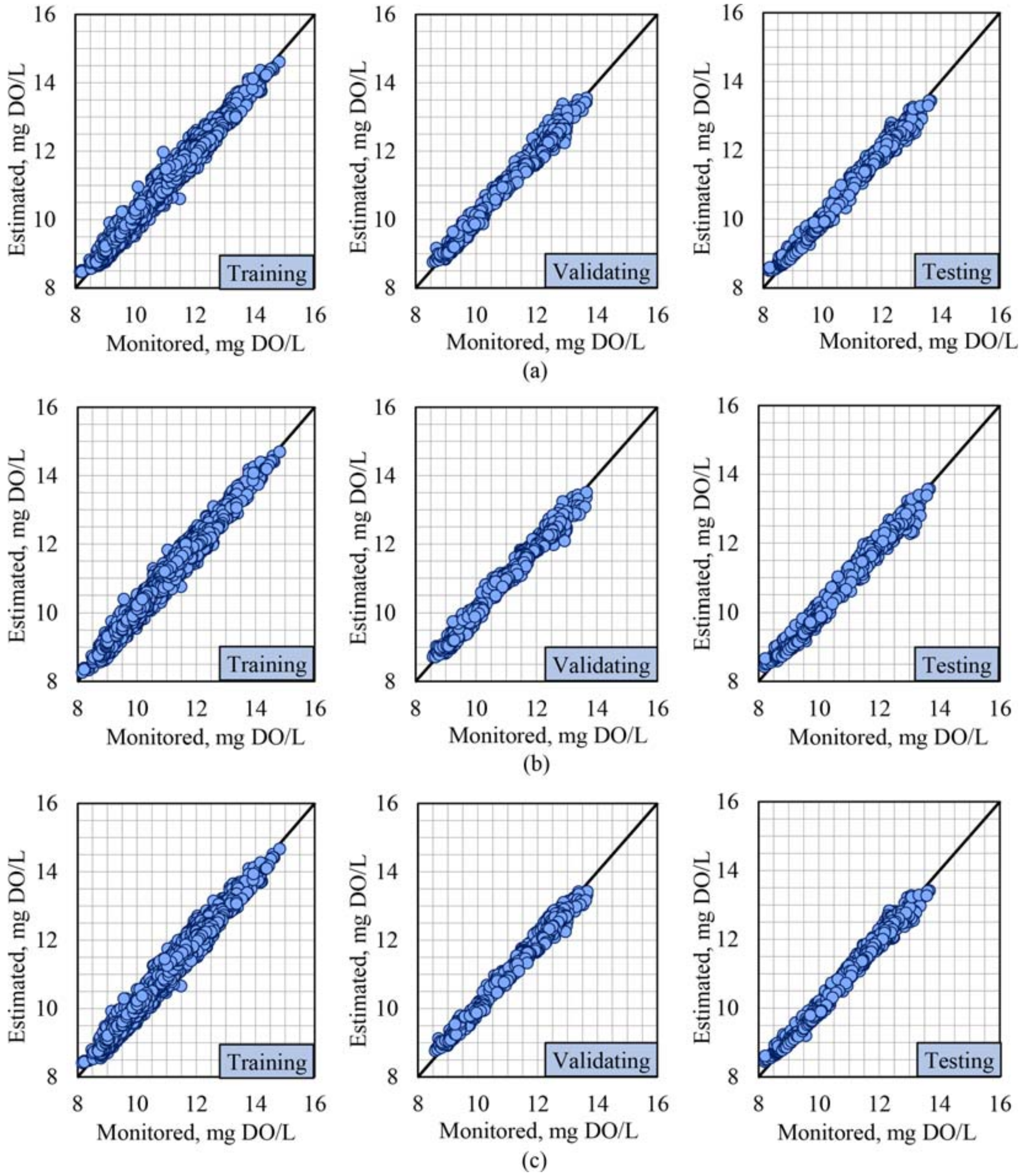

Fig. 3. The Scatter Plots of the River DO Concentrations Monitored versus Estimated by Model 8 Using the Methods: (a) CRA_QF, (b) MARS, (c) ANNs for USGS 14211010

and EF and QF for downstream station. The lowest RMSE values and highest NSEC and d values for the CRA, MARS, and ANNs methods were obtained from Models 8, 7, and 1, respectively, for upstream station and Models 4, 4 and 1, respectively, for downstream station. The highest NSEC values for the CRA, MARS and ANNs methods were 0.940, 0.926, and 0.910, respectively, for upstream station and 0.957, 0.887, and 0.948, respectively, for downstream station. According to the classification by Moriasi et al. (2007), these values are categorized as "very good".

4. Discussion

This study, which estimated DO concentrations using daily average WQ data from two stations on the Clackamas River, has some limitations. Firstly, the use of WQ data in the study, although intended to assess general trends over a given time period, may be limited in reflecting changes in higher temporal resolution data. Furthermore, the study results may not be directly applicable to other rivers or different geographical regions due to the focus on specific stations on a particular river. The most frequently used independent variables in DO modeling studies are WT, Q, EC or SC, and pH, respectively (Nacar et al., 2020a). Therefore, these variables were considered in this study. Eight models were established to evaluate the effect of each parameter on model performance. Although the selected WQ variables represent various aspects of DO concentrations, they do not cover all factors affecting

Table 5. The Comparison of the Results from the CRA, MARS, and ANNs Methods for the USGS 14210000 and USGS 14211010 Stations, Clackamas River, Oregon (October 2012 to September 2018)

Models	Statistics	USGS14210000						USGS14211010					
		LF	PF	EF	QF	MARS	ANNs	LF	PF	EF	QF	MARS	ANNs
M1	RMSE	0.480	0.751	0.480	0.479	0.543	0.460	0.295	0.506	0.280	0.279	0.566	0.323
	NSEC	0.903	0.762	0.903	0.903	0.875	***0.910***	0.956	0.872	0.961	***0.961***	0.840	***0.948***
	d	0.972	0.923	0.972	0.972	0.963	***0.975***	0.990	0.961	0.991	***0.991***	0.967	***0.987***
M2	RMSE	0.475	0.568	0.462	0.460	0.494	0.463	0.339	0.450	0.327	0.326	0.598	0.370
	NSEC	0.904	0.863	0.910	0.910	0.897	0.910	0.943	0.899	0.947	0.947	0.821	0.931
	d	0.973	0.958	0.974	0.975	0.970	0.974	0.986	0.974	0.987	0.987	0.957	0.984
M3	RMSE	0.483	0.566	0.475	0.471	0.525	0.469	0.420	0.532	0.421	0.425	0.585	0.480
	NSEC	0.901	0.865	0.905	0.906	0.884	0.907	0.912	0.858	0.911	0.910	0.829	0.885
	d	0.972	0.959	0.973	0.974	0.966	0.973	0.980	0.966	0.980	0.979	0.965	0.972
M4	RMSE	0.489	0.727	0.501	0.496	0.508	0.501	0.301	0.506	0.292	0.353	0.475	0.551
	NSEC	0.899	0.776	0.894	0.896	0.891	0.894	0.955	0.872	0.957	0.938	***0.887***	0.848
	d	0.971	0.927	0.970	0.970	0.969	0.970	0.989	0.963	0.990	0.983	***0.975***	0.955
M5	RMSE	0.485	0.554	0.472	0.467	0.548	0.468	0.430	0.511	0.415	0.413	0.571	0.477
	NSEC	0.901	0.870	0.906	0.908	0.873	0.907	0.907	0.869	0.914	0.915	0.837	0.886
	d	0.972	0.961	0.973	0.974	0.964	0.974	0.979	0.969	0.980	0.980	0.964	0.973
M6	RMSE	0.489	0.607	0.502	0.505	0.427	0.580	0.348	0.465	0.328	0.327	0.508	0.530
	NSEC	0.899	0.844	0.894	0.892	0.923	0.858	0.939	0.892	0.946	0.947	0.871	0.860
	d	0.972	0.954	0.970	0.969	0.979	0.964	0.985	0.972	0.987	0.987	0.971	0.963
M7	RMSE	0.489	0.608	0.502	0.504	0.419	0.574	0.433	0.532	0.422	0.447	0.533	0.579
	NSEC	0.899	0.844	0.894	0.892	***0.926***	0.861	0.906	0.858	0.911	0.900	0.858	0.832
	d	0.971	0.954	0.970	0.969	***0.980***	0.964	0.978	0.966	0.979	0.975	0.970	0.957
M8	RMSE	0.378	0.600	0.500	0.514	0.465	0.500	0.486	0.511	0.417	0.450	0.519	0.582
	NSEC	***0.940***	0.848	0.895	0.889	0.908	0.894	0.882	0.870	0.913	0.899	0.865	0.831
	d	***0.984***	0.955	0.971	0.968	0.975	0.970	0.965	0.969	0.980	0.974	0.970	0.957

Bold italic: the highest values
The unit of RMSE: mg/L

DO dynamics. There are also studies in the literature (Nacar et al., 2020b; Abba et al., 2021; Garabaghi et al., 2023) to determine the effects of parameters on the performance of DO models.

In the estimation of DO concentrations for different periods of the same station, the model having all parameters as input was selected as the optimal model. For both stations, the highest estimation performance was obtained from the MARS method in the training data set and from the ANNs method in the validating and testing data sets. However, the best performance statistical values vary slightly from model to model. In the DO modeling study conducted by Yaseen et al. (2018) using WT, pH, SC, and Q parameters, the most successful predictions were obtained from the model having all parameters as input. It was stated that providing more information to the model by increasing the number of input parameters could improve the model performance. The performance of the models established by adding different combinations of EC, Q, and pH parameters to Model 1, which was created using WT, the WQ variable with the highest correlation with DO concentration for upstream station, increased between 1.32 and 4.33%. It was calculated that the performances of the two-input models established by adding EC, Q, and pH parameters, respectively, to Model 1 increased by 1.32, 3.81 and 0.28%, respectively, compared with Model 1. It was determined that Model 7 established with the Q and pH parameters added to Model 1 had the highest performance in DO estimation. According to the NSEC values, Model 7, in which EC parameter was excluded, outperformed the optimal model by 0.53%. In the study performed by Kisi et al. (2020), it was determined that SC parameter had a negligible effect on DO concentration compared to WT and pH parameters. For downstream station, the performance of the models established by adding different combinations of EC, Q and pH parameters to Model 1 changes between -0.07 and 0.63%. It was calculated that the performances of the two-input models established by adding EC, Q and pH parameters, respectively, to Model 1 increased by -0.07, 0.02 and 0.21%, respectively, compared with Model 1. It was determined that Model 6 established with the EC and pH parameters added to Model 1 had the highest performance. Although Model 6 had the highest performance, it performed only 0.02% above the optimal model according to NSEC values. Ay and Kisi (2012) stated that the accuracy of the model obtained by including Q in the model established with WT, EC, and pH parameters was lower for ANNs method. For downstream station, the effect of the parameters added to Model 1 is lower compared to upstream station. This is thought to be due to the fact that the factors affecting WQ may be more balanced and stable in upstream station, which is surrounded by relatively undisturbed lands, and may be also more diverse and variable in downstream station, where human effect is higher because it is closer to settlements.

The best model for estimating DO concentrations from one station to another varies according to the methods. The lowest RMSE and the highest d and NSEC values were obtained from Model 8 using LF in the CRA method, Model 7 in the MARS method and Model 1 in the ANNs method for upstream station, and from Model 1 using QF in the CRA method, Model 4 in the MARS method, and Model 1 in the ANNs method for downstream station. The highest performance statistics among the models determined as the best belong to the CRA method for both stations. When compared the NSEC values of the best models for each method, it was determined that the performance of the CRA method was higher than the MARS and ANNs methods as 1.52 and 3.27%, respectively, for upstream station and 8.32 and 1.40%, respectively, for downstream station. Ahmed and Lin (2021) stated that vegetation cover, agricultural and build-up areas in the watersheds may be effective in the relationship between SC and pH parameters with DO. While pH parameter had a great importance in DO estimation in some studies (Heddam and Kisi, 2018; Yang et al., 2021), its effect on model performance was negligible in some studies (Ay and Kisi, 2017; Nacar et al., 2020b). This is thought to be due to the fact that the models established are specific to the relevant stations.

The hydro-chemical data for upstream station were also used by Keshtegar and Heddam (2018) with different variables, methods, and time intervals for DO modeling. They used two nonlinear mathematical modeling approaches, namely modified response surface method (MRSM) and multilayer perceptron neural network (MLPNN). As input parameters, Q, pH, SC, and turbidity were considered in their estimation models. The NSEC values were calculated as 0.802 and 0.804 for the training data set, 0.776 and 0.796 for the validating data set, and 0.782 and 0.791 for the testing data set for the MLPNN and MRSM methods, respectively. This study has higher NSEC values, for DO concentrations modeled using different parameters and methods in different time periods at the same station It is thought that the characteristics of the different models and data sets used are effective on these values. However, the calculated NSEC values support the efficiency and reliability of the methods used for modeling DO concentrations.

5. Conclusions

In this study, the applicability of three modeling methods, conventional regression analysis (CRA), multivariate adaptive regression splines (MARS), and artificial neural networks (ANNs), were investigated in modeling daily average dissolved oxygen (DO) concentration. The hydro-chemical data of two monitoring stations (USGS 14210000 and USGS 14211010) located on the Clackamas River, Oregon, were used for modeling. The daily average data of the discharge (Q), water temperature (WT), pH, and specific conductance (SC), as well as DO data from October 2012 to September 2022, were used for the modeling DO. Eight models having different input combinations have been compared using several performance statistics.

The function with the best performance in the CRA method varied from model to model and station to station. The highest performance values were generally obtained from QF in the CRA method. The performances of the methods are quite close to each other. For both stations, it was determined that the best methods were the MARS for the training data set and the ANNs for the validating and testing data sets.

In the estimation of DO concentrations for different periods of the same stations, the model performance increases for all methods as the number of inputs increases. On the other hand, for the same period from one station to another, the models with fewer input parameters performed better in the ANNs method.

The models and methods could estimate the river DO concentrations very close to monitored data both same periods for another station and another period for same station. In this way, it is thought that data deficiencies caused by the inability to perform measurements at the stations due to various reasons, such as adverse weather conditions, maintenance and repair works, and staff shortage, could be overcome.

Modeling complex and dynamic river systems requires simplifications and assumptions that no fully reflect the complexity of real-world conditions. In this study, daily averages of hydro-chemical parameters were considered, and temporal variations were neglected. In addition, climate change projections and land use changes were not included in the models. It is thought that investigating the accuracy of estimations by including these parameters in the models may be the subject of another study.

Acknowledgments

The authors sincerely thank the United States Geological Survey staff who made this work possible by ensuring the monitoring, processing, and management of the river water-quality data. They are also appreciative of the providers of the Salford Predictive Modeler 8 software, which was used to perform the analyses. The authors would also like to thank anonymous reviewers for their constructive comments and suggestions which helped to improve the article.

ORCID

Sinan Nacar https://orcid.org/0000-0003-2497-5032
Betul Mete https://orcid.org/0000-0002-3689-6430
Adem Bayram https://orcid.org/0000-0003-4359-9183

References

Abba SI, Abdulkadir RA, Sammen SS, Usman AG, Meshram SG, Malik A, Shahid S (2021) Comparative implementation between neuro-emotional genetic algorithm and novel ensemble computing techniques for modelling dissolved oxygen concentration. *Hydrological Sciences Journal* 66(10):1584-1596, DOI: 10.1080/02626667.2021.1937179

Abba SI, Linh NTT, Abdullahi J, Ali SIA, Pham QB, Abdulkadir RA, Costache R, Nam VT, Anh DT (2020) Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration. *IEEE Access* 8:157218-157237, DOI: 10.1109/ACCESS.2020.3017743

Ahmed MH, Lin LS (2021) Dissolved oxygen concentration predictions for running waters with different land use land cover using a quantile regression forest machine learning technique. *Journal of Hydrology* 597:126213, DOI: 10.1016/j.jhydrol.2021.126213

Arora S, Keshari AK (2021) Dissolved oxygen modelling of the Yamuna River using different ANFIS models. *Water Science and Technology* 84(10-11):3359-3371, DOI: 10.2166/wst.2021.466

Ay M, Kisi O (2012) Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado. *Journal of Environmental Engineering* 138(6):654-662, DOI: 10.1061/(ASCE)EE.1943-7870.0000511

Ay M, Kisi O (2017) Estimation of dissolved oxygen by using neural networks and neuro fuzzy computing techniques. *KSCE Journal of Civil Engineering* 21(5):1631-1639, DOI: 10.1007/s12205-016-0728-6

Bayazıt M (1981) Hidrolojide istatistik yöntemler, İstanbul Teknik Ünivesitesi Matbaası, İstanbul, Turkiye, 223 (in Turkish)

Bayram A, Kankal M (2015) Artificial neural network modeling of dissolved oxygen concentration in a Turkish watershed. *Polish Journal of Environmental Studies* 24(4):1507-1515

Bayram A, Kankal M, Onsoy H (2012) Estimation of suspended sediment concentration from turbidity measurements using artificial neural networks. *Environmental Monitoring and Assessment* 184(7):4355-4365, DOI: 10.1007/s10661-011-2269-2

Bhardwaj R, Singh RK (2022) Water quality modeling of the river ganga in the northern region of India using the artificial neural network technique. *Journal of Water Management Modeling* 30:C486, DOI: 10.14796/JWMM.C486

Blanca MJ, Arnau J, Lopez-Montiel D, Bono R, Bendayan R (2013) Skewness and kurtosis in real data samples. *Methodology* 9(2):78-84, DOI: 10.1027/1614-2241/a000057

Carpenter KD (2003) Water-quality and algal conditions in the Clackamas River Basin, Oregon, and their relations to land and water management. United States Geological Survey, Water-Resources Investigations Report 02-4189, Retrieved June 9, 2023, https://pubs.usgs.gov/wri/WRI02-4189/

Cox BA (2003) A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers. *Science of the Total Environment* 314:335-377, DOI: 10.1016/S0048-9697(03)00063-9

Csabragi A, Molnar S, Tanos P, Kovacs J (2017) Application of artificial neural networks to the forecasting of dissolved oxygen content in the Hungarian section of the river Danube. *Ecological Engineering* 100:63-72, DOI: 10.1016/j.ecoleng.2016.12.027

Dawson CW, Wilby R (1998) An artificial neural network approach to rainfall-runoff modelling. *Hydrological Sciences Journal* 43(1):47-66, DOI: 10.1080/02626669809492102

Dehghani R, Torabi Poudeh H, Izadi Z (2021) Dissolved oxygen concentration predictions for running waters with using hybrid machine learning techniques. *Modeling Earth Systems and Environment* 8(2):2599-2613, DOI: 10.1007/s40808-021-01253-x

Edrington MS (1993) Clackamas national wild and scenic river and state scenic waterway. United States Department of Agriculture - Environmental Assessment and Management Plan. Retrieved June 9, 2023, https://www.rivers.gov/rivers/rivers/sites/rivers/files/documents/plans/clackamas-plan-ea.pdf

Fetene BN, Shufen R, Dixit US (2018) FEM-based neural network modeling of laser-assisted bending. *Neural Computing and Applications* 29(6):69-82, DOI: 10.1007/s00521-016-2544-9

Friedman JH (1991) Multivariate adaptive regression splines. *The Annals of Statistics* 19(1):1-67, DOI: 10.1214/aos/1176347963

Garabaghi FH, Benzer S, Benzer R (2023) Modeling dissolved oxygen concentration using machine learning techniques with dimensionality reduction approach. *Environmental Monitoring and Assessment* 195(7):879, DOI: 10.1007/s10661-023-11492-3

Ghahramani Z (2015) Probabilistic machine learning and artificial intelligence. *Nature* 521(7553):452-459, DOI: 10.1038/nature14541

Hauke J, Kossowski T (2011) Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae* 30(2):87-93

Heddam S (2014) Modelling hourly dissolved oxygen concentration (DO) using dynamic evolving neural-fuzzy inference system (DENFIS)-based approach: Case study of Klamath River at Miller Island Boat Ramp, OR, USA. *Environmental Science and Pollution Research* 21(15):9212-9227, DOI: 10.1007/s11356-014-2842-7

Heddam S (2016) Simultaneous modelling and forecasting of hourly dissolved oxygen concentration (DO) using radial basis function neural network (RBFNN) based approach: A case study from the Klamath River, Oregon, USA. *Modeling Earth Systems and Environment* 2(3):135, DOI: 10.1007/s40808-016-0197-4

Heddam S, Kisi O (2018) Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology* 559:499-509, DOI: 10.1016/j.jhydrol.2018.02.061

Ji X, Shang X, Dahlgren RA, Zhang M (2017) Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: A case study of Wen-Rui Tang River, China. *Environmental Science and Pollution Research* 24(19):16062-16076, DOI: 10.1007/s11356-017-9243-7

Kalff J (2002) Limnology: Inland water ecosystems. Prentice-Hall, New Jersey, USA, 92

Kanda E, Kipkorir E, Kosgei J (2016) Dissolved oxygen modelling using artificial neural network: A case of River Nzoia, Lake Victoria basin, Kenya. *Journal of Water Security* 2:jws2016004, DOI: 10.15544/jws.2016.004

Keshtegar B, Heddam S (2018) Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: A comparative study. *Neural Computing and Applications* 30(10):2995-3006, DOI: 10.1007/s00521-017-2917-8

Keshtegar B, Heddam S, Hosseinabadi H (2019) The employment of polynomial chaos expansion approach for modeling dissolved oxygen concentration in river. *Environmental Earth Sciences* 78(1):34, DOI: 10.1007/s12665-018-8028-8

Khan UT, Valeo C (2016) Dissolved oxygen prediction using a possibility theory based fuzzy neural network. *Hydrology and Earth System Sciences* 20(6):2267-2293, DOI: 10.5194/hess-20-2267-2016

Kisi O (2015) Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology* 528:312-320, DOI: 10.1016/j.jhydrol.2015.06.052

Kisi O, Akbari N, Sanatipour M, Hashemi A, Teimourzadeh K, Shiri J (2013) Modeling of dissolved oxygen in river water using artificial intelligence techniques. *Journal of Environmental Informatics* 22(2):92-101, DOI: 10.3808/jei.201300248

Kisi O, Alizamir M, Gorgij AD (2020) Dissolved oxygen prediction using a new ensemble method. *Environmental Science and Pollution Research* 27(9):9589-9603, DOI: 10.1007/s11356-019-07574-w

Kisi O, Parmar KS, Soni K, Demir V (2017) Modeling of air pollutants using least square support vector regression, multivariate adaptive regression spline, and M5 model tree models. *Air Quality Atmosphere and Health* 10(7):873-883, DOI: 10.1007/s11869-017-0477-9

Kohonen T (1988) An introduction to neural computing. *Neural Networks* 1(1):3-16, DOI: 10.1016/0893-6080(88)90020-2

Lee KK (2011) Seepage investigations of the Clackamas River, Oregon: United States Geological Survey Scientific Investigations Report 2011-5191, 16p. Retrieved June 9, 2023, https://pubs.usgs.gov/sir/2011/5191/pdf/sir20115191.pdf

Li Y, Li X, Xu C, Tang X (2023) Dissolved oxygen prediction model for the Yangtze River Estuary Basin using IPSO-LSSVM. *Water* 15(12): 2206, DOI: 10.3390/w15122206

Matore EM, Khairani AZ (2020) The pattern of skewness and kurtosis using mean score and logit in measuring adversity quotient (AQ) for normality testing. *International Journal of Future Generation Communication and Networking* 13(1):688-702

Money E, Carter GP, Serre ML (2009) Using river distances in the space/time estimation of dissolved oxygen along two impaired river networks in New Jersey. *Water Research* 43:1948-1958, DOI: 10.1016/j.watres.2009.01.034

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* 50(3):885-900, DOI: 10.13031/2013.23153

Nacar S, Bayram A, Baki OT, Kankal M, Aras E (2020a) Spatial forecasting of dissolved oxygen concentration in the Eastern Black Sea Basin, Turkey. *Water* 12(4):1041, DOI: 10.3390/w12041041

Nacar S, Mete B, Bayram A (2020b) Estimation of daily dissolved oxygen concentration for river water quality using conventional regression analysis, multivariate adaptive regression splines, and TreeNet techniques. *Environmental Monitoring and Assessment* 192(12):752, DOI: 10.1007/s10661-020-08649-9

Najah A, El-Shafie A, Karim OA, El-Shafie AH (2014) Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring. *Environmental Science and Pollution Research* 21(3):1658-1670, DOI: 10.1007/s11356-013-2048-4

Nas SS, Bayram A, Nas E, Bulut VN (2008) Effects of some water quality parameters on the dissolved oxygen balance of streams. *Polish Journal of Environmental Studies* 17(4):531-538

Nong X, Lai C, Chen L, Shao D, Zhang C, Liang J (2023) Prediction modelling framework comparative analysis of dissolved oxygen concentration variations using support vector regression coupled with multiple feature engineering and optimization methods: A case study in China. *Ecological Indicators* 146:109845, DOI: 10.1016/j.ecolind.2022.109845

Olyaie E, Abyaneh HZ, Mehr AD (2017) A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. *Geoscience Frontiers* 8(3):517-527, DOI: 10.1016/j.gsf.2016.04.007

Peat J, Barton B (2008) Medical statistics: A guide to data analysis and critical appraisal. Blackwell Publishing, Massachusetts, USA

Rajaee T, Khani S, Ravansalar M (2020) Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. *Chemometrics and Intelligent Laboratory Systems* 200:103978, DOI: 10.1016/j.chemolab.2020.103978

Rajurkar MP, Kothyari UC, Chaube UC (2002) Artificial neural networks for daily rainfall-runoff modelling. *Hydrological Sciences Journal* 47(6):865-877, DOI: 10.1080/02626660209492996

Rezaei K, Guest B, Friedrich A, Fayazi F, Nakhaei M, Beitollahi A, Fatemi Aghda SM (2009) Feed forward neural network and interpolation function models to predict the soil and subsurface sediments distribution in Bam, Iran. *Acta Geophysica* 57:271-293, DOI: 10.2478/s11600-008-0073-3

Sarkar A, Pandey P (2015) River water quality modelling using artificial neural network technique. *Aquatic Procedia* 4:1070-1077, DOI: 10.1016/j.aqpro.2015.02.135

Sedighkia M, Datta B, Abdoli A, Moradian Z (2021) An ecohydraulic-based expert system for optimal management of environmental flow at the downstream of reservoirs. *Journal of Hydroinformatics* 23(6): 1343-1367, DOI: 10.2166/hydro.2021.112

Singh KP, Basant A, Malik A, Jain G (2009) Artificial neural network modeling of the river water quality - a case study. *Ecological Modelling* 220(6):888-895, DOI: 10.1016/j.ecolmodel.2009.01.004

Singh J, Knapp HV, Arnold JG, Demissie M (2005) Hydrological modeling of the Iroquois river watershed using HSPF and SWAT. *JAWRA Journal of the American Water Resources Association* 41(2):343-360, DOI: 10.1111/j.1752-1688.2005.tb03740.x

Song C, Yao L, Hua C, Ni Q (2021) 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. *Environmental Monitoring and Assessment* 193(6):363, DOI: 10.1007/s10661-021-09127-6

Tiyasha T, Tung TM, Bhagat SK, Tan ML, Jawad AH, Mohtar WHMW, Yaseen ZM (2021) Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models. *Marine Pollution Bulletin* 170:112639, DOI: 10.1016/j.marpolbul.2021.112639

Van Ooyen A, Nienhuis B (1992) Improving the convergence of the back-propagation algorithm *Neural Networks* 5(3):465-471, DOI: 10.1016/0893-6080(92)90008-7

Willmott CJ (1981) On the validation of models. *Physical Geography* 2(2):184-194, DOI: 10.1080/02723646.1981.10642213

Yang F, Moayedi H, Mosavi A (2021) Predicting the degree of dissolved oxygen using three types of multi-layer perceptron-based artificial neural networks. *Sustainability* 13(17):9898, DOI: 10.3390/su13179898

Yaseen ZM, Ehteram M, Sharafati A, Shahid S, Al-Ansari N, El-Shafie A (2018) The integration of nature-inspired algorithms with least square support vector regression models: Application to modeling river dissolved oxygen concentration. *Water* 10(9):1124, DOI: 10.3390/w10091124

Yilmaz B, Aras E, Nacar S, Kankal M (2018) Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models. *Science of the Total Environment* 639:826-840, DOI: 10.1016/j.scitotenv.2018.05.153

Zhang W, Goh AT (2016) Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geoscience Frontiers* 7(1):45-52, DOI: 10.1016/j.gsf.2014.10.003