**RESEARCH ARTICLE**

# The development of dissolved oxygen forecast model using hybrid machine learning algorithm with hydro-meteorological variables

Abul Abrar Masrur Ahmed[1,2] · S. Janifer Jabin Jui[2] · Mohammad Aktarul Islam Chowdhury[3] · Oli Ahmed[4] · Ambica Sutradha[4]

## Abstract

Dissolved oxygen (DO) forecasting is essential for aquatic managers responsible for maintaining ecosystem health and the management of water bodies affected by water quality parameters. This paper aims to forecast dissolved oxygen (DO) concentration using a multivariate adaptive regression spline (MARS) hybrid model coupled with maximum overlap discrete wavelet transformation (MODWT) as a feature decomposition approach for Surma River water using a set of water quality hydro-meteorological variables. The proposed hybrid model is compared with numerous machine learning methods, namely Bayesian ridge regression (BNR), k-nearest neighbourhood (KNN), kernel ridge regression (KRR), random forest (RF), and support vector regression (SVR). The investigational results show that the proposed model of MODWT-MARS has a better prediction than the comparing benchmark models and individual standalone counter parts. The result shows that the hybrid algorithms (i.e. MODWT-MARS) outperformed the other models ($r = 0.981$, WI $= 0.990$, RMAE $= 2.47\%$, and MAE $= 0.089$). This hybrid method may serve to forecast water quality variables with fewer predictor variables.

## Nomenclature

| | | | |
|---|---|---|---|
| ACF | Autocorrelation function | CEEMDAN-BNR | Hybrid model integrating the CEEMDAN algorithm with BNR |
| ANN | Artificial neural network | CEEMDAN-KNN | Hybrid model integrating the CEEMDAN algorithm with KNN |
| BF | Basis functions | | |
| BNR | Bayesian ridge regression | CEEMDAN-KRR | Hybrid model integrating the CEEMDAN algorithm with KRR |
| BOD | Biological oxygen demand | | |
| COD | Chemical oxygen demand | CEEMDAN-MARS | Hybrid model integrating the CEEMDAN algorithm with MARS |
| CCF | Cross-correlation function | | |
| CEEMDAN | Complete ensemble empirical mode decomposition with adaptive noise | CEEMDAN-RF | Hybrid model integrating the CEEMDAN algorithm with RF |
| | | CEEMDAN-SVR | Hybrid model integrating the CEEMDAN algorithm with SVR |

Responsible editor: Marcus Schulz

✉ Abul Abrar Masrur Ahmed
  AbulMasrur.Ahmed@unimelb.edu.au;
  masrur@outlook.com.au

  S. Janifer Jabin Jui
  SJanifer.Jui@usq.edu.au

  Mohammad Aktarul Islam Chowdhury
  maiccee@gmail.com

  Oli Ahmed
  oliahmed3034@gmail.com

  Ambica Sutradha
  ambicasutradhar@gmail.com

[1]  Department of Infrastructure Engineering, University of Melbourne, Parkville, VIC 3010, Australia

[2]  School of Mathematics Physics and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia

[3]  Department of Civil and Environmental Engineering, Shahjalal University of Science and Technology, 3114 Sylhet, Bangladesh

[4]  School of Modern Sciences, Leading University, Sylhet 3112, Bangladesh

| EEMD | Ensemble empirical mode decomposition |
|---|---|
| EEMD-BNR | Hybrid model integrating the EEMD algorithm with BNR |
| EEMD-KNN | Hybrid model integrating the EEMD algorithm with KNN |
| EEMD-KRR | Hybrid Model integrating the EEMD algorithm with KRR |
| EEMD-MARS | Hybrid Model integrating the EEMD algorithm with MARS |
| EEMD-RF | Hybrid model integrating the EEMD algorithm with RF |
| EEMD-SVR | Hybrid model integrating the EEMD algorithm with SVR |
| EMD | Empirical mode Decomposition |
| EMD-BNR | Hybrid model integrating the EMD algorithm with BNR |
| EMD-KNN | Hybrid model integrating the EMD algorithm with KNN |
| EMD-KRR | Hybrid model integrating the EMD algorithm with KRR |
| EMD-MARS | Hybrid model integrating the EMD algorithm with MARS |
| EMD-RF | Hybrid model integrating the EMD algorithm with RF |
| EMD-SVR | Hybrid model integrating the EMD algorithm with SVR |
| DWT | Discrete wavelet Transformation |
| DWT-BNR | Hybrid model integrating the DWT algorithm with BNR |
| DWT-KNN | Hybrid model integrating the DWT algorithm with KNN |
| DWT-KRR | Hybrid model integrating the DWT algorithm with KRR |
| DWT-MARS | Hybrid model integrating the DWT algorithm with MARS |
| DWT-RF | Hybrid model integrating the DWT algorithm with RF |
| DWT-SVR | Hybrid model integrating the DWT algorithm with SVR |
| ECDF | Empirical cumulative distribution function |
| ELM | Extreme learning machine |
| FE | Forecasting error |
| GCV | Generalised cross-validation |
| IMF | Intrinsic mode functions |
| KNN | K-nearest neighbourhood |
| KRR | Kernel ridge regression |
| LM | Legates-McCabe's Index |
| LSSVM | Least square support vector machine |
| MAE | Mean absolute error |
| MAPE | Mean absolute percentage error |
| MARS | Multivariate adaptive regression splines |
| MLP | Multi-layer perceptron |
| MODWT | Maximum overlap discrete wavelet transformation |
| MODWT -BNR | Hybrid Model integrating the MODWT algorithm with BNR |
| MODWT -KNN | Hybrid Model integrating the MODWT algorithm with KNN |
| MODWT -KRR | Hybrid Model integrating the MODDWT algorithm with KRR |
| MODWT –MARS | Hybrid Model integrating the MODWT algorithm with MARS |
| MODWT -RF | Hybrid Model integrating the MODWT algorithm with RF |
| MODWT -SVR | Hybrid Model integrating the MODWT algorithm with SVR |
| MRA | Multi-resolution analysis |
| MSE | Mean squared error |
| NCA | Neighbourhood component analysis |
| NSE | Nash–Sutcliffe efficiency |
| PACF | Partial auto–correlation function |
| r | Correlation coefficient |
| RBF | Radial basis function |
| RF | Random forest |
| RMSE | Root-mean-square-error |
| RRMSE | Relative root-mean-square error |
| SVR | Support vector regression |
| TDS | Total dissolved solids |
| WQ | Water quality |

# Introduction

The deterioration of the quality of water sources throughout the world is considered a wide-reaching issue of importance. Because of the rapid rise of communities and the diversity of their activities, this deterioration is speeding up, and it could constitute a severe threat to the aquatic environment and human health (Henderson et al. 2009; Hur and Cho 2012; Mouri et al. 2011; Su et al. 2011).

The dissolved oxygen (DO) in water is a critical water quality variable that is crucial for the proper functioning of the aquatic ecosystem (Ranković et al. 2010). DO demonstrate the water pollution in rivers (Heddam and Kisi 2018; Mohan and Kumar 2016) and the state of the river's ecosystems (Mellios et al. 2015; Ranković et al. 2010). The concentration of dissolved oxygen (DO) in aquatic systems refers to the metabolism of the aquatic systems, and it reflects the transient balance between the oxygen system and the metabolic activity. The concentration of DO is affected by a variety of parameters, including salinity, temperatures,

and pressure (US-Geological-Survey 2016). Researchers investigated the concentration and change of DO over the last decade since the dynamics of DO are nonlinear (Kisi et al. 2020). It is very desirable for water resource managers to develop a DO model for rivers that can reliably quantify and predict DO concentrations based on hydro-meteorological variables.

There are various methods available for estimating the DO concentration, but most of them are time-consuming and expensive to use since they require numerous parameters that are not readily available in most cases (Suen and Eheart 2003). More to the point, conventional data processing techniques are no longer appropriate for water quality modelling, which may be linked to the explanation that many parameters affecting water quality have a complicated nonlinear interaction with one another (Ahmed 2017; Xiang et al. 2006). There are specific issues in developing a water quality model for tiny streams or rivers due to the lack of available data, investment, and many different inputs to consider. As a result, certain well-known water quality analysis models, such as the United States Environmental Protection Agency (USEPA): QUAL2E and QUAL2K, WASP6, require a great deal of information that is not always readily available (Ahmed 2017). Moreover, these models are complex and sensitive and, therefore, tough to recognise.

Machine learning-based data-driven algorithms have become potentially widespread in the field of water quality modelling (Ahmed 2017; Ahmed and Shah 2017a; Forough et al. 2019; Kuo et al. 2004; Tomic et al. 2018) and hydrological modelling (Ahmed et al. 2021b, c; Ahmed and Shah 2017b; Yaseen et al. 2016). In addition, a number of artificial intelligence (AI)-based models for predicting and estimating DO concentrations have been developed such as soft computing (Tao et al. 2019), artificial neural networks (ANNs), and hybrid ANN (Keshtegar et al. 2019; Zounemat-Kermani et al. 2019), fuzzy-based models (Heddam 2017; Raheli et al. 2017), multivariate adaptive regression spline (MARS) (Rezaie-Balf et al. 2019), support vector machine (SVM) (Heddam and Kisi 2018; Li et al. 2017b), extreme learning machine (ELM) (Heddam 2016; Heddam and Kisi 2017; Zhu and Heddam 2020), quantile regression (Ahmed and Lin 2021), and other potential approaches were applied for dissolved oxygen concentration modelling.

This study investigates the utilisation of multivariate adaptive regression splines (MARS) (Friedman 1991) to describe DO dynamics' intrinsic nonlinear and multidisciplinary relationship. Like neural networks, no prior information on the numerical function is required for MARS. The benefit of the MARS model is that it can accomplish complex data by grouping related data collected, permitting it to understand easily (Zhang and Goh 2016). Considering the positive attribute, the MARS model has been used in hydrology (Deo et al. 2017b; Heddam and Kisi 2018; Kisi

and Parmar 2016; Yin et al. 2018) and the energy sector (Al-Musaylh et al. 2019). Heddam and Kisi (2018) applied the least-square support vector machine (LSSVM), multivariate adaptive regression splines, and M5 model tree (M5T) for daily dissolved oxygen forecasting. The authors found the MARS model a substantial forecasting approach with a limited number of predictor variables. Therefore, incorporating the hybrid approaches and a potential feature selection algorithm may boost the result of forecasting. Nevertheless, the hybrid MARS models are yet to be executed in the study sites of Bangladesh.

Using multi-resolution analysis (MRA), a technique for extracting data features, the prediction performance can be enhanced significantly. Using the EMD, you can decompose a signal following the spirit of the Fourier series into a specific number of components. A coefficient representing Gaussian white noise with a unit variance is introduced sequentially to the time series in CEEMDAN-based decomposition to reduce the complexity and avoid the intricacy of the time series (Prasad et al. 2018). A coefficient denoting Gaussian white noise with covariance matrices is introduced sequentially to the time series in CEEMDAN-based decomposition to reduce the complexity and prevent the intricacy of the time series (Di et al. 2014). Previous studies have used CEEMDAN in forecasting soil moisture (Ahmed et al. 2021a; Prasad et al. 2018, 2019) with an earlier version (i.e. EEMD) used in forecasting stream-flow (Seo and Kim 2016) and rainfall (Beltrán-Castro et al. 2013; Jiao et al. 2016; Ouyang et al. 2016). Discrete wavelets transform (DWT) has been employed (Deo and Sahin 2016; Deo et al. 2016; Nourani et al. 2014, 2009) in different fields of hydrology. On the other hand, DWT has a limitation that prevents it from extracting all the features of the predictors in its entirety. An enhanced discrete wavelet transforms (DWT), such as the MODWT, can solve these problems (Cornish et al. 2006; Prasad et al. 2017; Rathinasamy et al. 2014). Al-Musaylh et al. (2020) successfully used MODWT to decompose the short-term electricity demand of Australia. The study incorporated the MODWT by separately splitting the data to training, testing, and validation to calculate the detailed approximation, as Quilty and Adamowski (2018) prescribed. The potential application of MODWT is further approved by Prasad et al. (2017), where MODWT was used to forecast stream-flow. However, neither the MODWT nor the DWT decomposition model has incorporated the MARS model in DO forecasting, as attempted in this study.

The feature selection technique, namely neighbourhood component analysis (NCA) for regression, was used in this investigation. As a result of the algorithm being slowed down by the extraneous and redundant features, the prediction model is less accurate (Arhami et al. 2013) different feature selection methods have been utilised in predictive

models (Ahmed et al. 2021a; Prasad et al. 2017, 2019). The NCA method has been successfully applied by Ahmed et al. (2021a) to forecast surface soil moisture. The study demonstrates that the feature weight calculated by NCA was found successful in forecasting soil moisture and to the study by Ghimire et al. (2019b), where they applied NCA for solar radiation forecasting. Forecasting DO concentration with a machine learning method incorporated with the NCA feature selection method and feature decomposition methods would substantially increase forecasting performance.

To the author's knowledge, there has been no systematic comparison of various feature decomposition strategies in improving MARS performance for daily DO estimates. The fundamental contribution of this study is the selection of an appropriate feature decomposition algorithm (i.e. MODWT, DWT, CEEMDAN, EEMD, and EMD) tailored MARS model for DO prediction. While effective adjustment of MARS parameters via feature decomposition algorithms can increase prediction accuracy, the incorporation of feature selection and feature decomposition theories can aid decision-makers in making the optimal choice for the best prediction model. Because attempting all available optimisation techniques is practically impossible, the scope of the current study has been reduced to a few potential algorithms to be merged with the MARS. As a result, the goal of this study is to (1) use 5 feature decomposition techniques to modify MARS ability, (2) compare the performances of hybridised MARS models, and (3) rank the hybridised MARS models using hydro-meteorological variables. The findings of this work will be a helpful tool that can provide valuable information for better water management.

## Materials and methods

### Theoretical frameworks of proposed models

#### Multivariate adaptive regression spline

According to Friedman (1991), a non-parametric and nonlinear regression technique, the multivariate adaptive regression spline (MARS), was utilised in this investigation. MARS uses numerous splines to build nodes between these lines (Friedman 1991). The underlying functional link between inputs and outputs is not assumed in the MARS model. The data in each spline is assigned using basis functions (BF) in MARS models. It is possible to express the BF as a single equation between two knots. Two adjacent data domains converge at a knot, and the output is continuous. An adaptive regression algorithm is used (Heddam and Kisi 2018). The MARS model depicts the piecewise relationship between the input and output variables using numerous lines. The over-fitting of training

data is avoided by setting a predefined minimum number of observations between knots (Heddam and Kisi 2018).

Let y be the target output, and a matrix of n input variables be the vector $x = (x_1, \ldots, x_n)$. The data are then presumed to be created from an undisclosed 'true' model. In the case of a straight answer, this will be as follows:

$$y = f(x_1, \ldots, x_n) + \xi = f(x) + \xi \tag{1}$$

In which $\xi$ is the distribution of the model error, and n is the number of training data points. By adding sufficient BFs, MARS approximates the f(.). For linear functions piecewise: max $(0, x\text{-}t)$ where a knot exists at position $t$ (Zhang and Goh 2016). The max (.) equation implies that only the positive portion of (.) is used; otherwise, a zero value will be given corresponding to:

$$\max(0, x - t) = \begin{cases} x - t, & if x \geq t \\ 0, & otherwise \end{cases} \tag{2}$$

Thus, $f(x)$ is constructed as a linear BF(x) combination:

$$f(x) = \beta_0 + \sum_{i=1}^{n} \beta_i BF(x) \tag{3}$$

The coefficients $\beta$ are constants, calculated using the form of least squares. Initially, $f(x)$ is applied to input data in a forward–backward stepwise process to determine the knot's position where the feature value varies (Deo et al. 2017b). A broad model is built at the end of the forwards' stage to over-fit the qualified input data. According to the generalised cross-validation, the model is optimised by deleting one last basis function from the model (GCV). GCV for a model is computed as follows for the training data with $n$ observations:

$$GCV = \frac{\frac{1}{n} \sum_{i=1}^{n} \left[ y_i - f(x_i) \right]^2}{\left[ 1 - \frac{M + d \times (M-1)/2}{n} \right]^2} \tag{4}$$

where $M$ is the number of BF, $d$ is the penalising parameter, $n$ is the number of measurements, and $f(x_i)$ denotes the MARS model's expected values.

MARS is a non-parametric regression modelling technique that is flexible and does not make any assumptions about the relationships between the variables (Stull et al. 2014). The model is simple to understand and interpret (Kuhn and Johnson 2013). MARS models typically exhibit a favourable bias-variance trade-off. While the models are sufficiently flexible to account for nonlinearity and variable interactions (and so have a relatively low bias), the limited nature of the MARS basis functions precludes excessive flexibility (thus, MARS models have relatively low variance).

## Maximal overlap discrete wavelet transforms

Distinctive wavelet transforms (DWTs) are modified by the maximal overlap discrete wavelet transform (MODWT) (Li et al. 2017a). Ideally, time series analysis can be done using the MODWT's appealing qualities, which prevent missing data without subsampling. MODWT's ability to extract additional information is enhanced because the coefficients of decomposed components in each layer are identical to the original time series. Time-series data are broken down into high-pass and low-pass filters using MODWT, which handles two feature sets. Further, high-pass filters can be broken down into several information levels depending on the suitable time frame (He et al. 2017). Low-pass filters reflect the real-time-series signal pattern called an approximation. The signal $\varkappa_m$ is decomposed through wavelet low-pass $\pi_m$ and high-pass detail filters $h_m$ and reconstructed by digital reconstruction filters complementing decomposition filters. This principle is described in the equations below:

$$\mathcal{X}_{m+1}(K) = \sum_p h_{p-2k} \mathcal{X}_m(P) \tag{5}$$

$$d_{m+1}(K) = \sum_p l_{p-2k} \mathcal{X}_m(P) \tag{6}$$

$$\mathcal{X}_m(K) = \sum_p h'_{p-2k} \mathcal{X}_{m+1}(P) + \sum_p l'_{p-2k} \mathcal{X}_{m+1}(P) \tag{7}$$

## Comparing models

In this study, we proposed a MODWT-MARS model to predict the dissolved oxygen of a running river. To find a practical approach to machine learning methods and feature decomposition methods, a pool of six machine learning models and five feature decomposition methods were also incorporated. The theoretical description of the proposed algorithms (i.e. MODWT and MARS) was explained in the previous section, and this section provides a short overview of the comparing algorithms.

Breiman (2001) proposed an algorithm based on a random forest (RF), which included methods for regression and classification. The bootstrap resampling procedure generates a new set of training data from the initial training sample set $N$, and then bootstrap-set random forests are built using $K$ decision trees. The RF model's full specifications may be read here (Ali et al. 2020a). The random forests approach has become a prominent tool for classification, prediction, investigating variable relevance, selection, and outlier identification. RF comprises a group (ensemble) of basic tree predictors. Each tree may generate a response given a collection of predictor values (Jui et al. 2022; Yu et al. 2017).

With regularisation and the kernel technique, it is possible to reduce over-fitting using the KRR (Kernel Ridge Regression) regression model (Saunders et al. 1998). The "kernel technique" can be used to generate a nonlinear form of ridge regression. Extending the general framework, kernel ridge regression allows nonlinear prediction. Linear, polynomial and Gaussian kernels are only some of the many options available for enhancing overall performance (You et al. 2018). The suggested KRR technique has the fundamental advantage of learning a global function and predicting any target variable using a regularised variation of least squares.

The Bayesian modelling approach uses hierarchical data (Huang and Abdel-Aty 2010). Bayesian regression uses this regularisation parameter, easily tailored to the data. The Gaussian maximum posterior estimate is discovered before the coefficient $w$ and, with an accuracy of $\lambda$ (-1), is treated as a random variable instead of a lambda. In contrast, most decision-making analyses based on maximum likelihood estimation entail determining the values of parameters that may significantly impact the analysis outcome and for which there is considerable uncertainty. The capacity to include previous information is one of the primary advantages of the Bayesian technique (Saqib 2021).

A machine learning kernel method known as SVR (Support Vector Regression) can be used for various purposes, including forecasting time series. SVRs that use kernels can also learn the nonlinear trend of the training data. There are three SVR models to pick from, each with a different kernel (RBF, poly, and linear) (Yang et al. 2017). It should also be noted that the proposed KRR model in its generic sense has been used in many research including the forecasting of precipitation (Ali et al. 2020b), drought (Ali et al. 2019), wind speed (Alalami et al. 2019; Douak et al. 2013; Mishra et al. 2019; Naik et al. 2018; Zhang et al. 2019), and solar power (Dash et al. 2020).

K-nearest neighbours (KNN) algorithm is implemented using instance-based learning, which serves two purposes: (1) estimating the test data density function and (2) categorising the test data obtained from the test patterns (Shabani et al. 2020). Choosing the number of neighbours (k) is a crucial stage. This method's efficiency depends on selecting samples from the nearest reference database (or most similar). If $k$ is significant, other points from other classes can be placed inside the desired range of possibilities (Wu et al. 2008). The KNN method has been successfully applied previously (Ghiassi et al. 2017; Liu et al. 2020).

This study incorporated five decomposition methods (i.e. DWT, EMD, EEMD, MODWT and CEEMDAN) and six machine learning methods (i.e. MARS, RF, BNR, SVR, KNN and KRR) to address the prediction problem of dissolved oxygen concentration. Hyperspectral feature decomposition is DWT-assisted, and the features are evaluated

for their efficacy in discriminating between subtly different ground covers (Bruce et al. 2002). The theoretical explanation of the method is explained by other researchers (Agbinya 1996; Fowler 2005; Shensa 1992). Most recently, Huang et al. (Huang et al. 1998) developed an empirical mode decomposition (EMD) method for analysing the information contained in data derived from non-stationary and nonlinear systems. This algorithm decomposes the signal into a series of oscillatory functions that are 'well-behaved,' which are referred to as the intrinsic mode functions in this context (IMFs). When used with the powerful adaptive EMD tool, it behaves as a dyadic filter bank (Flandrin et al. 2004). It is handy for filtering out noise in the measurement domains (Khaldi et al. 2008). Torres et al. (2011) implemented the CEEMDAN process to reduce the computational cost and retain the ability to eliminate mode mixing. The readers are requested to go through the previous studies (Ahmed et al. 2021a; Zhang et al. 2017; Zhou et al. 2019) for getting further information on CEEMDAN.

## Study area and data

The Surma River, Bangladesh, provided daily water quality factors. Figure 1 depicts the Surma River monitoring stations. This river drains one of the heaviest runoffs in the Surma-Meghna Basin system (Chowdhury and Ali 2006). The Surma River originates in Assam's Cachar district, flows through Bangladesh's Sylhet and Sunamganj districts, joins the Meghna River near Bhairab Bazar Kishoreganj, and empties into the Bay of Bengal. Many studies are found regarding water quality analysis (Ahmed 2017; Ahmed and Shah 2017a, b), riverbank erosion (Islam and Hoque 2014), stream flows (Ahmed and Shah 2017b), and water level modelling (Biswas et al. 2009). The Surma River's Keane Bridge station provided the study's water quality variables between January 2017 and December 2019 obtained 15 cm to 20 cm below the surface.

The selection of prospective predictive factors is critical for predictive modelling. Various studies reveal that some variables predict DO better than the others (Ahmed 2017; Tomic et al. 2018). Ahmed (2017) used Biological Oxygen Demand (BOD) and Chemical oxygen demand (COD) for predicting the dissolved oxygen of the Surma River. Kisi and Ay (2012) observed that the temperature, pH, and electrical conductivity are highly influential over Fountain Creek, Colorado. However, Ranković et al. (2010) claimed that pH and water temperature have a practical relation in DO prediction, whereas nitrates, chloride, and total phosphate have poor connections. It is found that pH is a standard variable for predicting DO values using ANN, followed by temperature. However, along with pH and temperature, some authors used oxygen-containing ($PO_4^{3-}$, $NO_3$-N) variables or oxygen demanding variables ($NH^{-4}$ N, COD, and BOD)

(Wen et al. 2013). Turbidity (Iglesias et al. 2014) and total solid can be considered essential water quality parameters, as their high value indicates typically high values of other parameters associated with water quality. The missing values were interpolated from two adjacent values. The fundamental statistics of the input variables are tabulated in Table 1.

## Development of MODWT-MARS model

The multi-phase MODWT-MARS model and other benchmark models were created in Python using the sci-kit-learn machine learning platform (Pedregosa et al. 2011b). All simulations were performed on a machine with an Intel i7 processor running at 3.6 GHz and 16 GB of RAM. Furthermore, a software platform such as 'MATLAB2020' is employed for feature selection using neighbourhood component analysis (NCA). However, tools such as *matplotlib* (Barrett et al. 2004) and *seaborn* (Waskom et al. 2020) are employed to visualise the forecasted DO. Figure 2 depicts the workflow of the proposed MODWT-MARS model.

The wavelet transformation using MODWT was combined with the predictor variables filtered by the NCA approach to create the MODWT-MARS model. Identifying the wavelet-scaling filter types and decomposition level is vital in creating a substantial wavelet transformation model. Because there is no one approach to choose the optimal filter, Al-Musaylh et al. (2020) used a trial and error strategy. Quilty and Adamowski (2018) discovered an issue in the forecast model inputs due to erroneous wavelet decomposition during the wavelet-based forecasting model. The inaccuracy can be traced back to the decomposition process's boundary conditions. They identified three problems: (1) improper use of future data, (2) unsuitable selection of decomposition levels and filters, and (3) incorrect division of validation and calibration data. The readers are encouraged to look up more information about the findings of Quilty and Adamowski (2018). The authors' concern about the development of MODWT and DWT decomposition were addressed in this study. After separating the DO variables to resolve more comprehensive information to create the MODWT-MARS model, Fig. 3 displays the time-series of the intrinsic mode functions (IMFs) and the residual components and decomposed components of MODWT.

There is no formula for verifying whether or not a model's valid predictors are present (Tiwari and Adamowski 2013). Although the research describes three input selection strategies for picking the time series of lagged memories of DO and predictors for an optimum model, the literature does not specify which method should be used. The autocorrelation function (ACF), partial autocorrelation function (PACF), and cross-correlation function (CCF) approaches are the three types of approaches to consider. A substantial antecedent behaviour in terms of the lag of DO from the
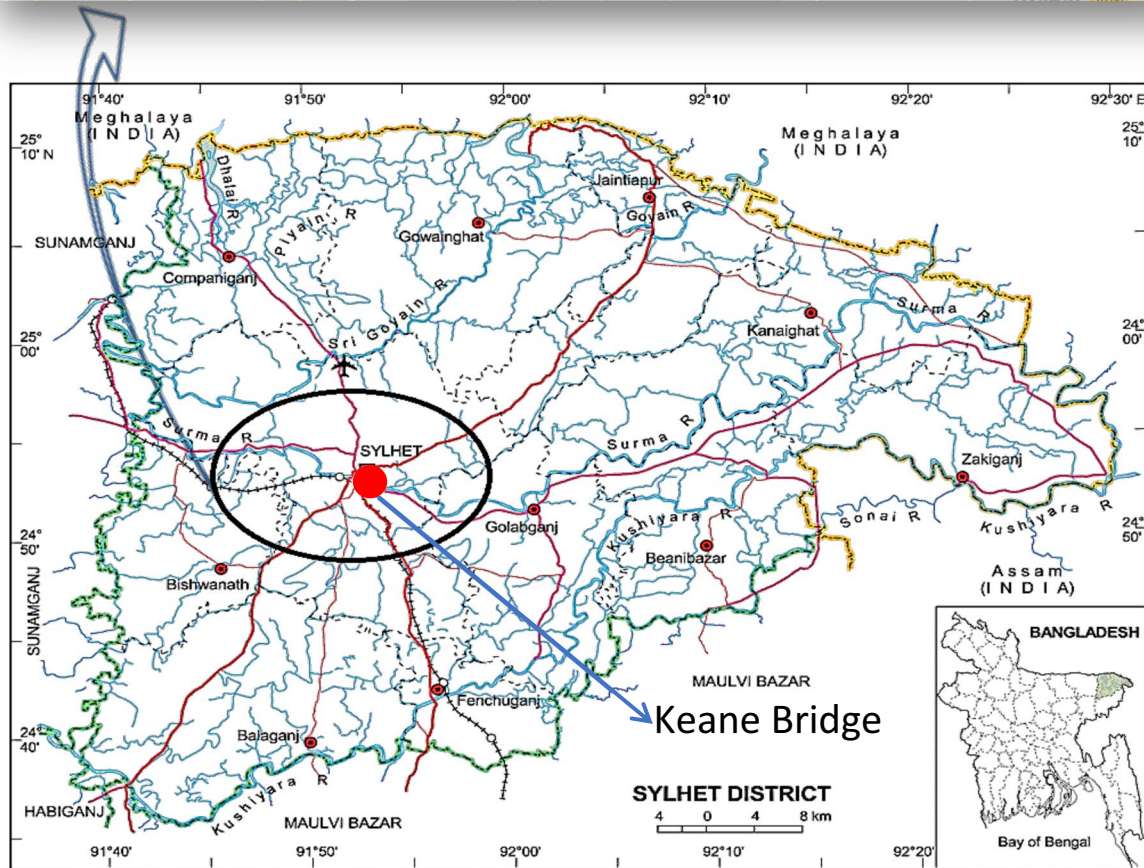
**Fig. 1** The study region showing the Keane Bridge station of Surma River, Sylhet, Bangladesh

**Table 1** Basic statistics i.e. minimum (min), maximum (max), mean (M), standard deviation (SD), and coefficient of variation (CV) of the water quality variables in Surma River, Sylhet, Bangladesh

| Variable | Acronyms | Unit | Min | Max | Mean | SD | CV (%) |
|---|---|---|---|---|---|---|---|
| Humidity | h | % | 0.01 | 3.79 | 0.53 | 0.70 | 132 |
| Water temperature | w | $^0$C | 0.18 | 4.0 | 1.53 | 1.05 | 69 |
| Rainfall | r | mm | 8.00 | 127 | 32.66 | 20.99 | 64 |
| TDS | td | Mg/l | 10.0 | 522 | 142.3 | 102.15 | 72 |
| pH | p | - | 5.70 | 8.25 | 6.92 | 0.55 | 8 |
| Turbidity | tr | (NTU) | 4.18 | 42.62 | 11.84 | 7.37 | 62 |
| Air temperature | a | $^0$C | 12.30 | 33.30 | 27.10 | 4.93 | 20.00 |
| DO | d | (mg/l) | 1.90 | 17.30 | 5.40 | 2.45 | 45 |

predictors was found in this study, utilising PACF as the predictor (Tiwari and Adamowski 2013; Tiwari and Chatterjee 2011). Figure 4 demonstrates the PACF for DO time series showing the antecedent behaviour in terms of the lag of DO and decomposed components of DO using MODWT. It is clear from the figure that antecedent monthly delays are found significant.

The cross-correlation function determines which predictor's antecedent lag selects the input signal pattern and which

pattern the predictor selects (Adamowski et al. 2012). The cross-correlation function is used to establish the statistical similarity between the predictors and the target variable. The cross-correlation function between the predictors and the DO for the River Surma is depicted in Fig. 5a. Afterwards, a set of significant input combinations were determined by assessing $r_{cross}$ of each predictor with DO. In this plot, a 95% confidence level of the statistically significant $r_{cross}$ is shown in the blue line. It is found from the Fig. 5a the correlation

**Fig. 2** The study's workflow details the steps in the model designing phase and the proposed hybrid CEEMDAN-MARS predictive models. Note: IMF = Intrinsic Mode Function, CCF = Cross-Correlation Functions, PACF = partial autocorrelation function, CEEMDAN = complete ensemble empirical mode decomposition with adaptive noise and DO = Dissolved Oxygen (mg/l)
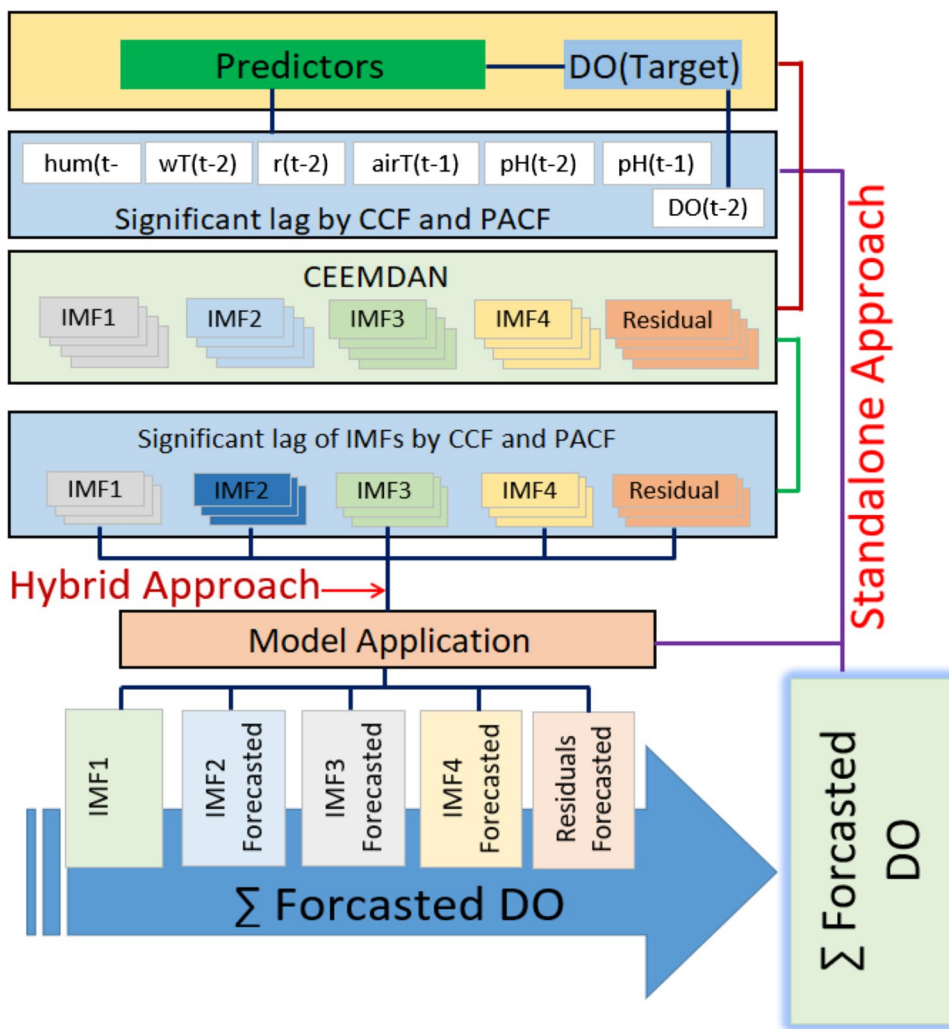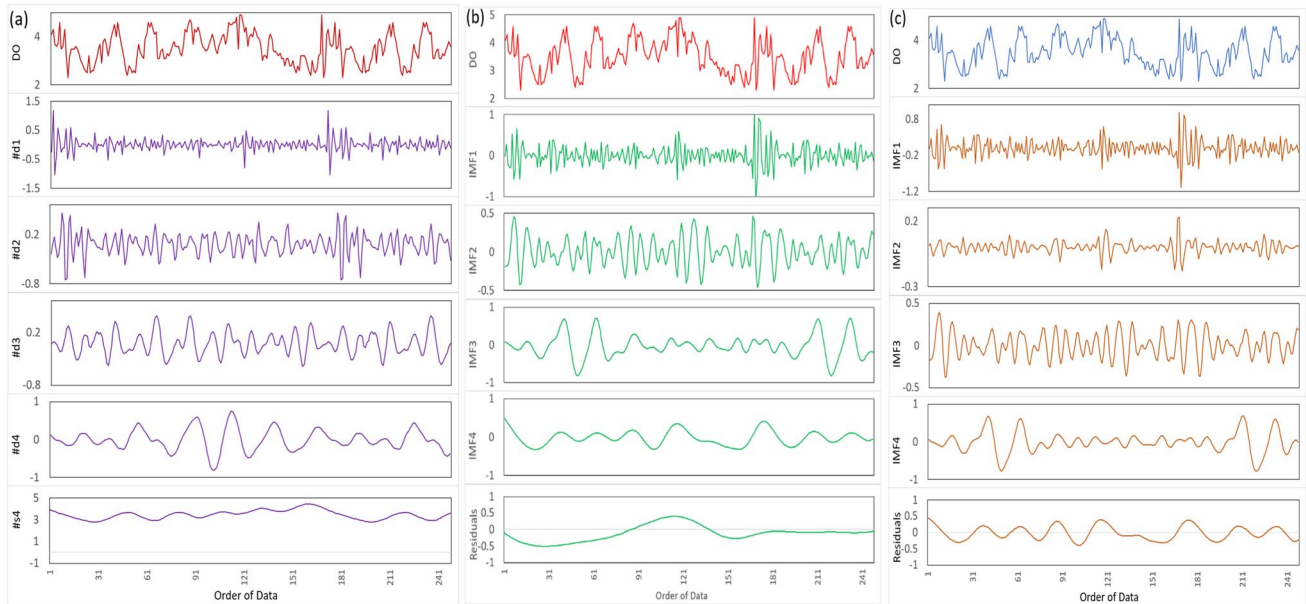
**Fig. 3** Time series of the **a** maximum overlap discrete wavelet coefficient (MODWC) of Dissolved Oxygen using MODWT, and intrinsic mode functions (IMFs) and the residual components after decompos-ing the DO in the training period using **b** CEEMDAN and **c** EEMD. The time series of the actual DO is plotted at the top of the figure

of respective data with DO was found as the highest for all stations at lag zero ($r_{cross} \approx 0.25$–0.45). A similar procedure is maintained for the decomposed predictor variables. Figure 5b–f demonstrate the $r_{cross}$ value between #$d_1$ (DO) and #$d_n$ (Predictors) and their respective residuals ($n = 1$ to 4). Figure 5 shows that the $r_{cross}$ value was ranged between 0.25 and 0.50 found more than 95% confidence level. The predictor data sets are normalised (Ahmed 2017; Ali et al. 2019) between 0 and 1 to minimise one variable's overestimation.

$$DO_{norm} = \frac{DO - DO_{min}}{DO_{max} - DO_{min}} \qquad (6)$$

Python-based *Scikit-learn* (Pedregosa et al. 2011a) was used to build this study's SVR, RF, KRR, BNR, and KNN model. For SVR, the RBF (Radial Basis Function) was employed in developing the SVR model (Suykens et al. 2002). The RBF uses a faster function during training to examine nonlinearities between the objective and predictor variables (Goyal et al. 2014; Lin 2003; Maity et al. 2010). The tricky process of creating an accurate SVR model required identifying the 3D parameters (C, σ, and ε) (Hoang et al. 2014). This is why the NCA algorithm was used to select the parameters with the smallest weight value. (Pedregosa et al. 2011a).

**Fig. 4** Partial autocorrelation function (PACF) plot of the DO time series exploring the antecedent behaviour in terms of the lag of daily DO. The blue line in the figures indicates the ±95% confidence level
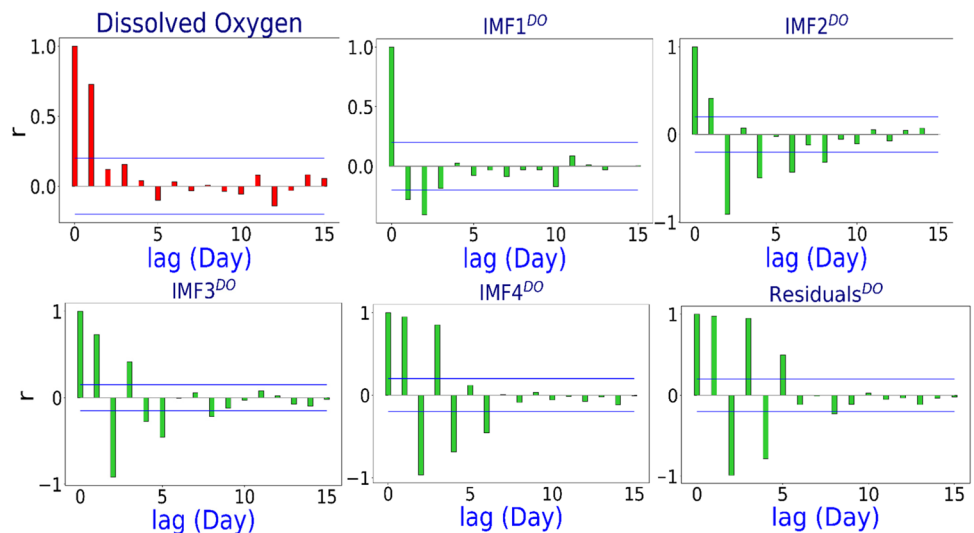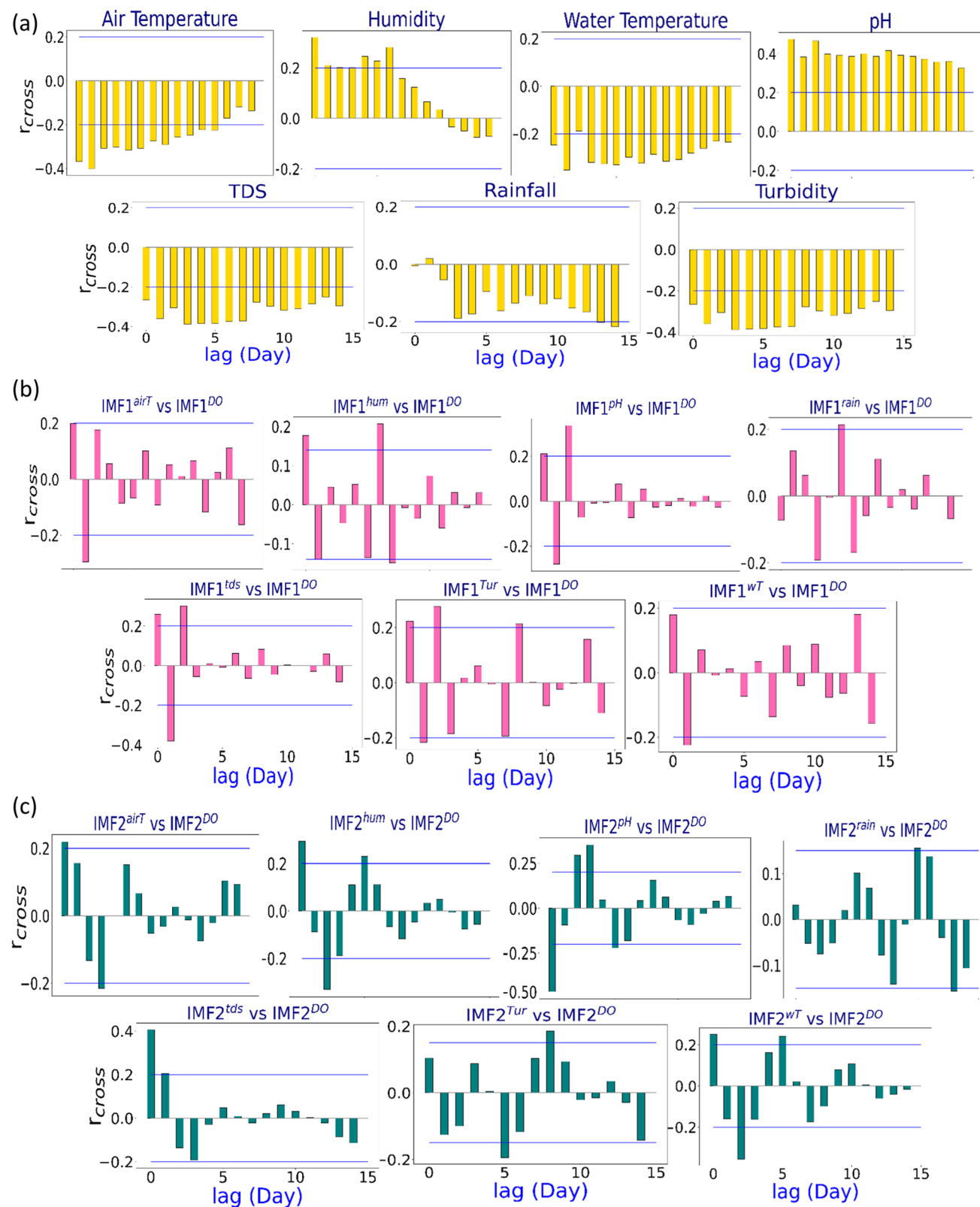
**Fig. 5** An analysis of the statistically significant cross-correlation function plots of **a** actual variables vs DO, **b** IMF1 of all variables vs IMF1 of DO, **c** IMF2 of all variables vs IMF2 of DO, **d** IMF3 of all variables vs IMF3 of DO, **e** IMF4 of all variables vs IMF4 of DO, f) residuals of all variables vs residuals of DO

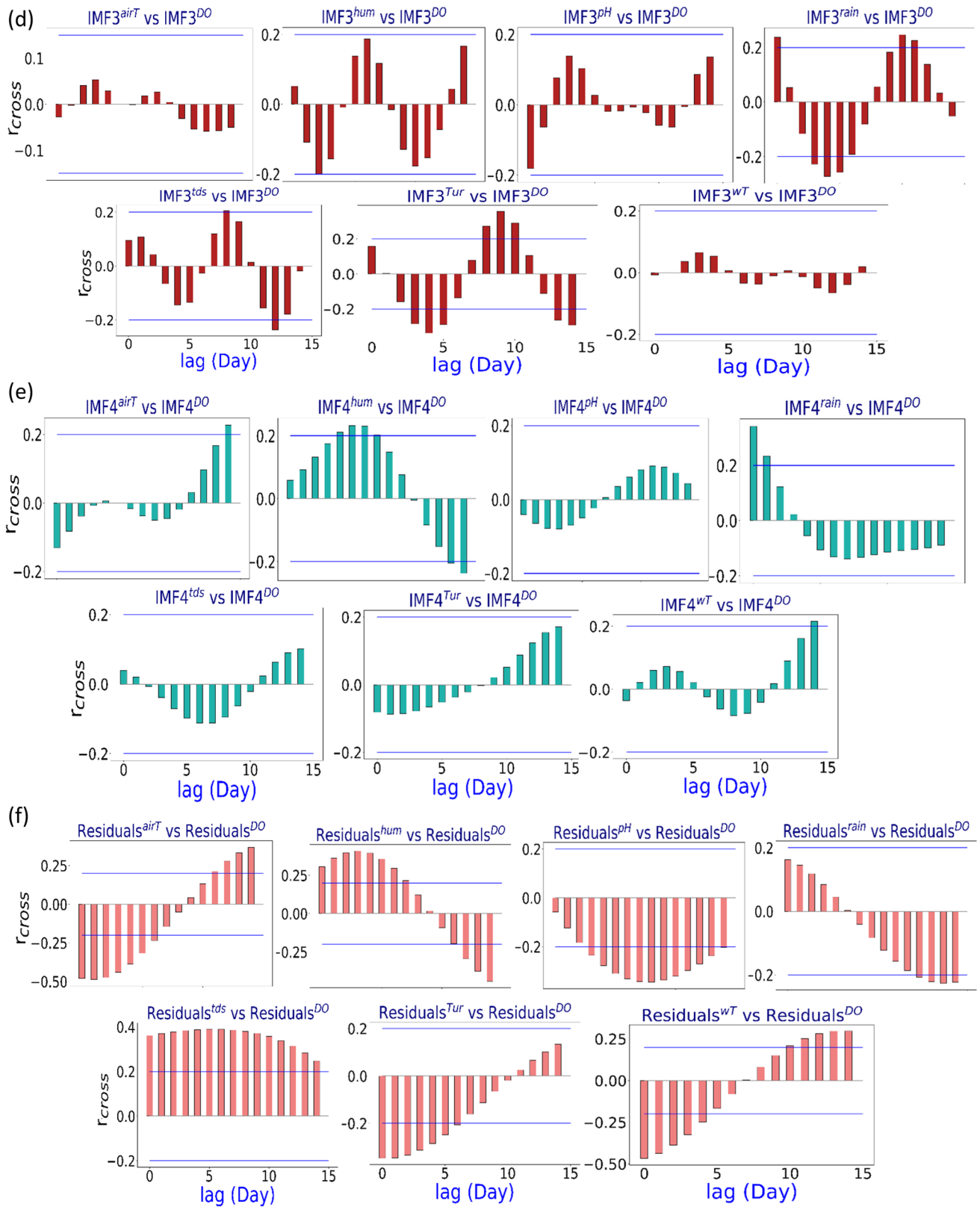**Fig. 5** (continued)

Alternatively, the MARS model adopted the Python-based Py-earth package (Rudy and Cherti 2017). The two MARS models used are cubic or linear piecewise functions. This study used a piecewise cubic model because it provided a smoother response. Also, the generalised recursive partitioning regression was adopted since it can handle multiple preconditioners. A forward and backward selection was used for optimisation. Initially, the algorithm ran with a "naïve" model that only contained the intercept term. The training MSE was reduced by iteratively adding the reflected pairs of basis functions (Table 2).

The accuracy of the hybrid MARS and other comparing models was constructed using piecewise cubic and linear regression functions, respectively. The best MARS model was selected using the lowest Generalised Cross-Validation (GCV) (Lin 2003); the MODWT-MARS model yielded the lowest RMSE and the highest LM, demonstrating the most accurate predictions. The optimum tuning parameters of various machine learning methods are tabulated in Table 3.

## Model evaluation benchmarks

Several statistical score metrics were considered in the rigorous evaluation of the proposed model (i.e. MODWT-MARS) compared with the counterpart models. The commonly adopted model score metrics such as Pearson's correlation coefficient ($r$), root-mean square error (RMSE; mg/l), mean absolute error (MAE; mg/l), Nash–Sutcliffe efficiency (NSE), Absolute Percentage Bias (APB; %), and Willmott's Index agreement (WI) (Krause et al. 2005; Legates and McCabe 1999; Nash and Sutcliffe 1970; Willmott et al. 2012) were used as the popular metrics used elsewhere (Ahmed et al. 2021a; Ghimire et al. 2019c). Due to the stations' geographic alterations, the percentage error measures relative error values such as RRMSE, RMAE, and MAPE were considered. Owing to the inherent merits and weaknesses of the metrics, combining them is prudent (Sharma et al. 2019). Different sets of model evaluation metrics such as RMSE, MAE, and $r^2$ (coefficient of determination) (Chu et al. 2020); NSE, RMSE, MAE, and PERS (persistence index)

**Table 2** Different input combinations prepared by using the NCA feature selection algorithm. Numerical values after the variable indicate respective lag memories of the datasets

| No | Different input combinations |
|---|---|
| 1 | $h_3$ |
| 2 | $h_3, tr_5$ |
| 3 | $h_3, tr_5, h_5$ |
| 4 | $h_3, tr_5, h_5, h_6$ |
| 5 | $h_3, tr_5, h_5, h_6, tr_4$ |
| 6 | $h_3, tr_5, h_5, h_6, tr_4, tr_7$ |
| 7 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3$ |
| 8 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4$ |
| 9 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6$ |
| 10 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4$ |
| 11 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2$ |
| 12 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1$ |
| 13 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8$ |
| 14 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4$ |
| 15 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2$ |
| 16 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3$ |
| 17 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3$ |
| 18 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5$ |
| 19 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4$ |
| 20 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1$ |
| 21 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1, w_3$ |
| 22 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1, w_3, a_3$ |
| 23 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1, w_3, a_3, w_2$ |
| 24 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1, w_3, a_3, w_2, r_{13}$ |
| 25 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1, w_3, a_3, w_2, r_{13}, w_4$ |
| 26 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1, w_3, a_3, w_2, r_{13}, w_4, r_1$ |
| 27 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1, w_3, a_3, w_2, r_{13}, w_4, r_1, w_1$ |
| 28 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1, w_3, a_3, w_2, r_{13}, w_4, r_1, w_1, d_1$ |
| 29 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1, w_3, a_3, w_2, r_{13}, w_4, r_1, w_1, d_1, w_5$ |
| 30 | $h_3, tr_5, h_5, h_6, tr_4, tr_7, tr_3, h_4, tr_6, td_4, td_2, td_1, td_8, p_4, a_2, td_3, p_3, p_5, a_4, a_1, w_3, a_3, w_2, r_{13}, w_4, r_1, w_1, d_1, w_5, p_1$ |

(Tiwari and Chatterjee 2010); Legates-McCabe's Index (LM), Willmott's Index (WI), RRMSE, and RMAE (Ali et al. 2019; Ghimire et al. 2019b; Yaseen et al. 2019) were selected for evaluating the model with numerous sets of variables. The correlation coefficient ($r$) provides information about the linear association between forecasted and observed DO data; therefore, it is limited in its capacity. However, $r$ is considered oversensitive to extreme values (Willmott et al. 1985). Moreover, RMSE and MAE can provide appropriate information regarding the forecasting skill, whereby RMSE evaluates the robustness of the model related to high values but focuses on the deviation of the forecasted value from the observed (Deo et al. 2017a). Alternatively, MAEs are not a perfect replacement for RMSEs (Chai and Draxler 2014). The Nash–Sutcliffe efficiency (NSE) is a widely used model evaluation criteria for the hydrological models. NSE is a dimensionless metric and a scaled version of MSE, offering a better physical interpretation (Legates and McCabe 2013). However, the NSE over-emphasises the higher values of outliers, and lower values are neglected (Legates and McCabe 1999). Due to the standardisation of the observed and predicted means and variance, the robustness of $r$ is limited. Willmott's Index (WI) was utilised to address this issue by considering the mean squared error ratio instead of the differences. The mathematical notations of the statistical metrics are as follows:

$$\text{MAE(mg/L)} = \frac{1}{N} \sum\nolimits_{i=1}^{N} \left| DO_{for}^i - DO_{obs}^i \right| \tag{7}$$

$$\text{RMSE(mg/L)} = \sqrt{\frac{1}{N} \sum\nolimits_{i=1}^{N} (DO_{for}^i - DO_{obs}^i)^2} \tag{8}$$

$$\text{NSE} = 1 - \left[ 1 - \frac{\sum_{i=1}^{N} DO_{for}^i{}^2}{\sum_{i=1}^{N} \left( DO_{obs}^i - \overline{DO}_{for}^i \right)^2} \right] \tag{9}$$

$$r = \left\{ \frac{\sum_{i=1}^{N} \left( DO_{obs}^i - \overline{DO}_{obs}^i \right)\left( DO_{for}^i - \overline{DO}_{for}^i \right)}{\sqrt{\sum_{i=1}^{N} \left( DO_{obs}^i - \overline{DO}_{obs}^i \right)^2 \sum_{i=1}^{N} \left( DO_{for}^i - \overline{DO}_{for}^i \right)^2}} \right\}^2 \tag{10}$$

$$\text{RRMSE(\%)} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (DO_{for}^i - DO_{obs}^i)^2}}{\frac{1}{N} \sum_{i=1}^{N} (DO_{obs}^i)} \times 100 \tag{11}$$

$$\text{RMAE(\%)} = \frac{\frac{1}{N} \sum_{i=1}^{N} \left| DO_{for}^i - DO_{obs}^i \right|}{\frac{1}{N} \sum_{i=1}^{N} DO_{obs}^i)} \times 100 \tag{12}$$

$$\text{MAPE(\%)} = \frac{1}{N} \left( \sum\nolimits_{i=1}^{N} \left| \frac{DO_{for}^i - DO_{obs}^i}{DO_{obs}^i} \right| \right) * 100 \tag{13}$$

$$\text{WI} = 1 - \left[ \frac{\sum_{i=1}^{N} (DO_{for}^i - DO_{obs}^i)^2}{\sum_{i=1}^{N} \left( \left| DO_{for}^i - \overline{DO}_{obs}^i \right| + \left| DO_{obs}^i - \overline{DO}_{obs}^i \right| \right)^2} \right] \tag{14}$$

$$\text{APB(mg/l)} = \left[ \frac{\sum_{i=1}^{N} \left( \left| DO_{for}^i - DO_{obs}^i \right| \right) * 100}{\sum_{i=1}^{N} \left| DO_{obs}^i \right|} \right] \tag{15}$$

$$\text{LM} = 1 - \left[ \frac{\sum_{i=1}^{N} \left| DO_{for}^i - DO_{obs}^i \right|}{\sum_{i=1}^{N} \left| DO_{obs}^i - \overline{DO}_{obs}^i \right|} \right] \tag{16}$$

where $DO_{obs}^i$ and $DO_{for}^i$ denote the observed and model-forecasted values from the $i$th element; $\overline{DO}_{obs}^i$ and $\overline{DO}_{for}^i$ denote their average, respectively, and N represents the observation's number of the DO.

## Results

In this study, MARS models optimised using a feature decomposition approach were utilised to forecast DO time series using hydro-meteorological variables. Several ways were employed to do this, including the conventional machine learning models (i.e. MARS, RF, SVR, KNN, and KRR), feature decomposition methods (i.e. MODWT, CEEMDAN, EEMD, EMD, and DWT), and the feature selection method (i.e. NCA) to screen the optimal model to forecast the DO. Though the mathematical metrics are so ambiguous that there is no way to evaluate the suitable alternative, it is reasonable to use multiple performance evaluation approaches. Compared to the other models, the hybrid, and standalone models of BNR, KNN, KRR, and RF outstripped all decomposition methods. The performance of MODWT-MARS has revealed that the NCA algorithm helped choose the relevant features to assist the MARS in better emulating the future DO concentration. MODWT found important performance matrices, such as r, NSE, WI, RMSE, and MAE. The MODWT-MARS model outperforms all the other tested models.

This study used the NCA algorithm to screen the appropriate predictor variables in the model. Table 2 provides the input combination for forecasting DO. The robustness of

**Table 3** The optimal hyperparameter of the proposed MARS model, including that of the other benchmark models' methods include machine learning (i.e. BNR, SVR, KRR, KNN, and RF)

| Name | Hyper-parameters | Acronym | Optimum |
|------|------------------|---------|---------|
| KRR | Regularisation strength | *alpha* | *1.5* |
| | Kernel mapping | *kernel* | *linear* |
| | Gamma parameter | *gamma* | *None* |
| | Degree of the polynomial kernel | *degree* | *3* |
| | Zero coefficient for polynomial and sigmoid kernels | *coef0* | *1.2* |
| BNR | Maximum number of iterations | *n_iter* | *200* |
| | Stop the algorithm if w has converged | *tol* | *0.0001* |
| | Shape parameter for Gamma distribution over alpha | *alpha_1* | *1e-05* |
| | Inverse scale parameter over alpha | *alpha_2* | *1e-05* |
| | Shape parameter for Gamma distribution over lambda | *lambda_1* | *1e-06* |
| | Inverse scale parameter for Gamma distribution over lambda | *lambda_2* | *1e-04* |
| | The initial value for alpha | *alpha_init* | *None* |
| KNN | Number of neighbours | *n_neighbors* | *5* |
| | Weights | *Weights* | *uniform* |
| | The algorithm used to compute the nearest neighbours | *algorithm* | *auto* |
| | Leaf-size passed | *leaf_size* | *30* |
| | Power parameter for the Minkowski metric | *p* | *2* |
| | The distance metric to use for the tree | *metric* | *minkowski* |
| | Additional keyword arguments for the metric | *metric_params* | *none* |
| | The number of parallel jobs | *n_jobs* | *int* |
| MARS | maximum degree of terms | *max_degree* | *1* |
| | Smoothing parameter used to calculate GCV | *penalty* | *3.0* |
| RF | Number of trees in the forest | *n_estimators* | *120* |
| | Maximum depth of the tree | *max_depth* | *2* |
| | Minimum number of samples for an internal node | *min_sample_split* | *2* |
| | Number of features for the best split | *max_features* | *auto* |

the NCA integrated BNR, KNN, KRR, MARS, and SVR model is provided in Tables A1–A6 in terms of statistical metrics would be found as supplementary materials. Tables show that each model's optimum standalone models were found between combinations from 19 to 29. For the case of the BNR model, the standalone model ($BNR_{28}$) shows poor performance ($r = 0.809$, WI $= 0.887$, RMAE $= 7.55\%$, and MAE $= 0.275$) comparing with the BNR-MODWT model ($r = 0.977$, WI $= 0.987$, RMAE $= 3.37\%$, and MAE $= 0.117$). Moreover, the hybrid models showed improved performance ranging from 0.888 to 0.977 and 7.17 to 3.37% for $r$ and RMAE accordingly. The $MARS_{29}$ model was found as the optimum model ($r = 0.824$, WI $= 0.895$, RMAE $= 7.97\%$, and MAE $= 0.277$) among all combinations of MARS model. The MODWT-MARS model was found as the highest performed model with substantial performance parameters ($r = 0.981$, WI $= 0.990$, RMAE $= 2.47\%$, and MAE $= 0.089$) which is followed by CEEMDAN-MARS model ($r = 0.949$, WI $= 0.971$, RMAE $= 4.65\%$, and MAE $= 0.156$). Mentionable that the highest model of SVR was found for CEEMDAN-SVR ($r = 0.971$, WI $= 0.983$, RMAE $= 3.36\%$) compared to the optimum standalone model ($SVR_{20}$).

Mentionable that KRR, KNN, and RF model provides poor performance comparatively.

Further analysis through a box plot showing the forecasted vs observed DO and absolute forecasting error of all hybrid models is illustrated in Fig. 6. The absolute forecasted error was determined as $|FE| = DO^{for} - DO^{obs}$. The box plot demonstrates the observed ($DO^{obs}$) data dispersion and forecasted ($DO^{for}$) DO from the proposed machine learning approaches and comparing models. Figure 6b, c, and e visualise the quartiles' data with distinctly larger outliers. The lower end of the plot lies between the lower quartile (25th percentile) and the upper quartile (75th percentile). The MODWT-MARS model shows an identical prediction compared with MODWT-SVR, with higher outliers for the SVR model. A more in-depth inspection of the absolute forecast error ($|FE|$) from the hybrid MODWT-MARS model further strengthens the suitability of the hybrid MARS approach in predicting the DO of the Surma River, which has the narrowest distribution compared with other models. The MODWT-MARS model has a significant percentage (98%) of the $|FE|$ in the first error brackets ($0 < |FE| < 0.25$), while the MODWT-SVR model has a percentage of 95%.

**Fig. 6** Box plots of hybrid models (MODWT-MARS) and their respective standalone counterparts (i.e. MARS, BNR, KRR, KNN, RNN, and SVR) in forecasting DO compare to the observed DO of Surma River
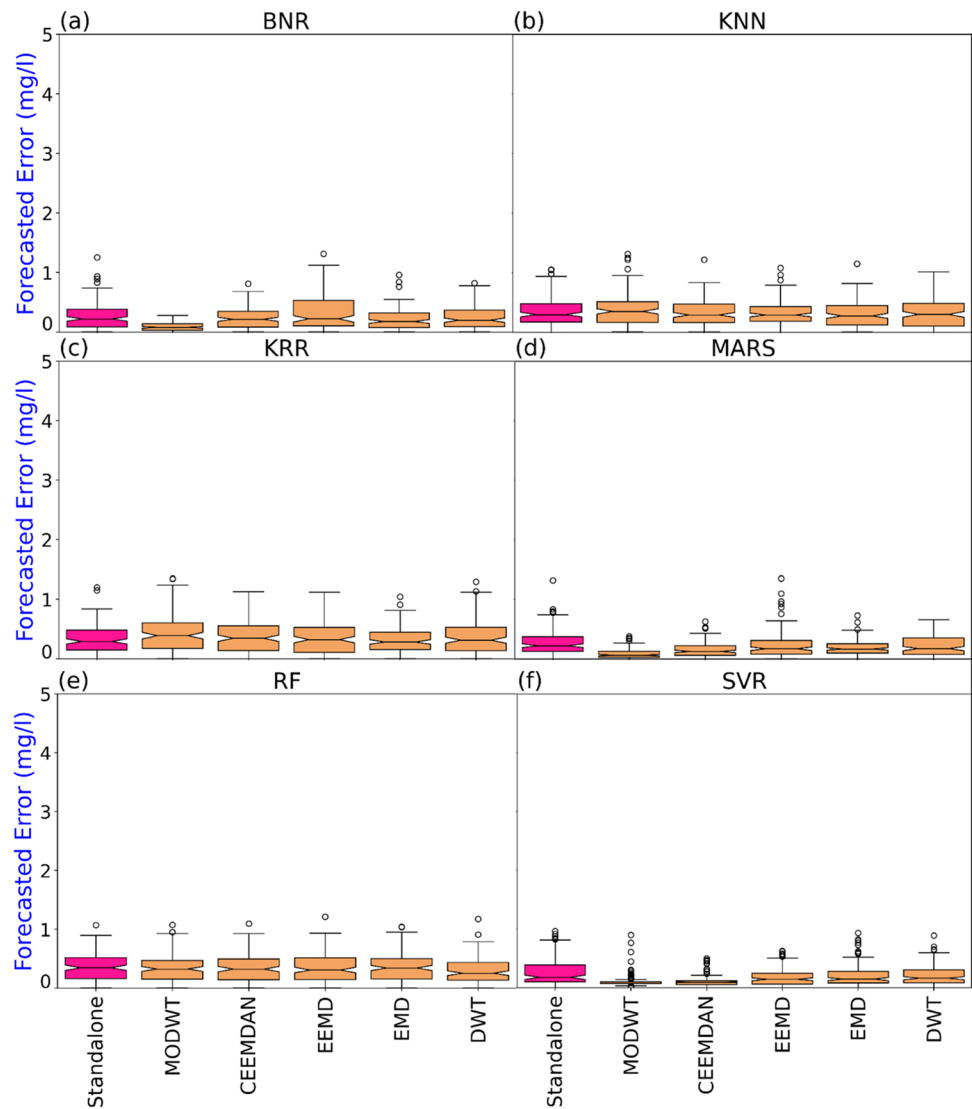


The empirical cumulative distribution function (ECDF) visualisation demonstrates the forecast error data's feature from the least to highest and perceives the full features circulated across the dataset. Figure 7 represents the empirical CDF of all six models for objective models and comparing models. The hybrid MODWT-MARS model was seen as reasonably sound against other models. The MODWT-MARS generated errors significantly lower from 0 to 0.25 mg/l. In the model-like KNN, KRR, and RF, the distribution of CDF was larger comparatively. The analysis also revealed that the standalone models showed a poor distribution, proving that MODWT-MARS was the most precise and responsive model.

To analyse the proposed MODWT-MARS model's further robustness, the models' forecasting performance was further assessed based on RRMSE and MAPE for all tested models, as shown in Fig. 8. From the figure,

the magnitude of RRMSE and MAPE for the objective model (MODWT-MARS) is significantly lower, which clarifies the potential merits of the proposed model. The best RRMSE (3.6%) and MAPE (2.2%) were found for the MODWT-MARS model, which was followed by the MODWT-BNR model with moderate RRMSE (4.0%) and MAPE (3%). Besides, KNN, KRR, and RF models with MODWT showed RRMSE (11.5% to 13.5%) and MAPE (9.5% to 12%) values, demonstrating poor performance.

Compared to the standalone models using the Taylor diagram (Taylor 2001), the proposed model performance improves the interpretation presented in Fig. 9. The Taylor diagram demonstrates that the MODWT-MARS model with the NCA algorithm is closer to the observation than the comparing models. Again, the forecasted DO illuminates the proposed model's better pertinency than the standalone and benchmark models. The benchmark models' performance
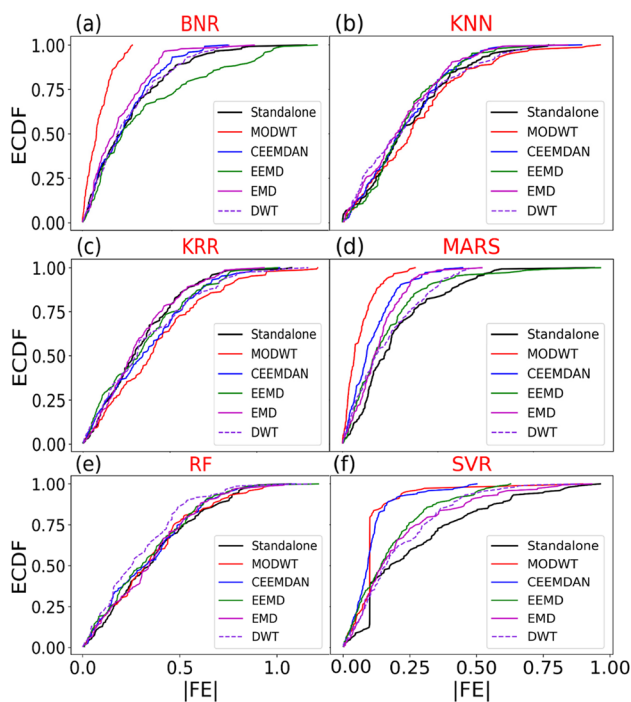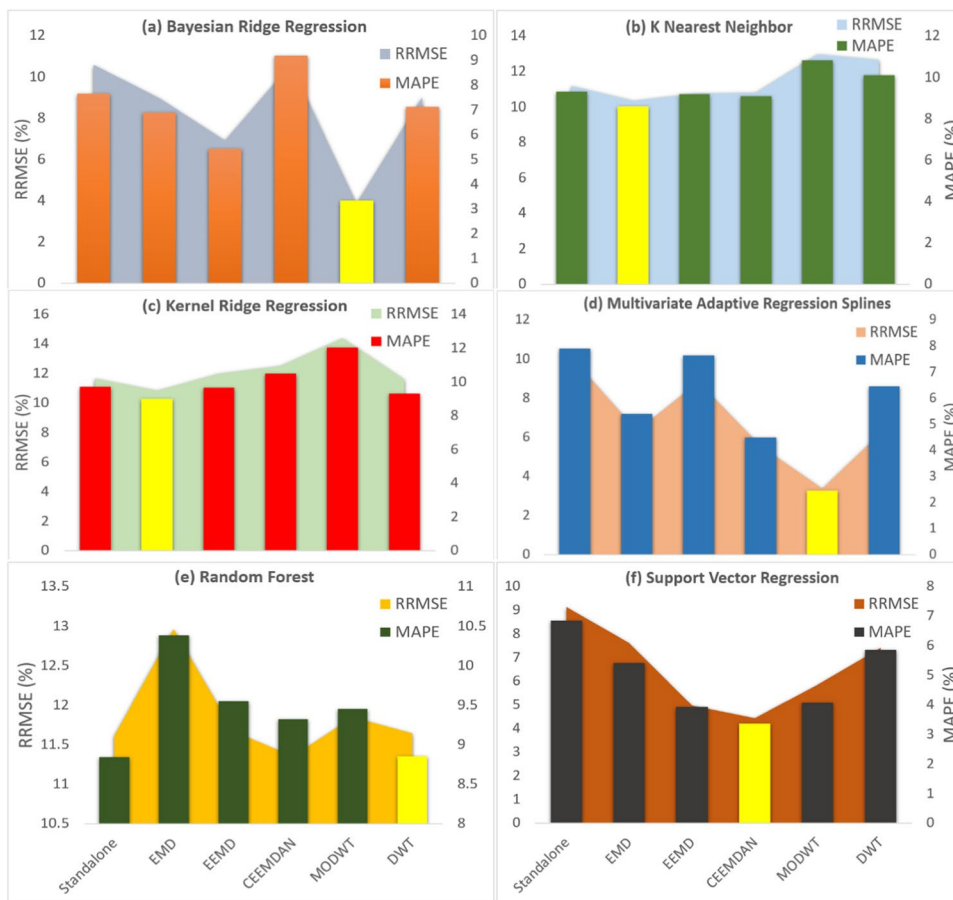
**Fig. 7** Empirical cumulative distribution function (CDF) of forecasted error |FE| of DO generated by the proposed MODWT-MARS and comparing models

with CEEMDAN and MODWT (i.e. MODWT-SVR, CEEMDAN-SVR, and CEEMDAN-MARS) achieved closer proximity to the observed values. However, the proximity of the observed DO for the MARS model with MODWT feature decomposition is the closest.

The scatter plot of the forecasted and observed DO for the proposed MODWT-MARS model portrayed a detailed comparison of DO forecasting (Fig. 10). The scatter plots comprise with the coefficient of determination ($R^2$) with goodness-of-fit between forecasted vs observed DO and a least-square fitting line and the corresponding equation; $DO_{for} = m \times DO_{obs} + C$, where, $m$ is referred to as the gradient, and $C$ is denoted as the y-intercept. Figure 10 reveals that the proposed model displays significant performance with a more considerable $R^2$ value. The DO forecasting using a hybrid machine learning model (i.e. MODWT-MARS) performed significantly better than the other models. The magnitudes registered from the hybrid MODWT-MARS model were the closest to unity, which, in pairs ($m|R^2$), are 0.978|0.976, followed by MODWT-SVR (0.939|0.965). Moreover, the CEEMDAN-SVR (0.699|0.795) and CEEMDAN-MARS (0.700|0.794) models provide a comparatively lower pair. Alternatively, y-intercepts [*ideal value* = 0] was found close to zero i.e. 0.084 for the proposed model.

**Fig. 8** Comparison of the forecasting skill of proposed models in terms of RRMSE (%) and MAPE (%) in the testing period

**Fig. 9** Tylor diagram representing correlation coefficient and the standard deviation difference for proposed hybrid models vs benchmark models
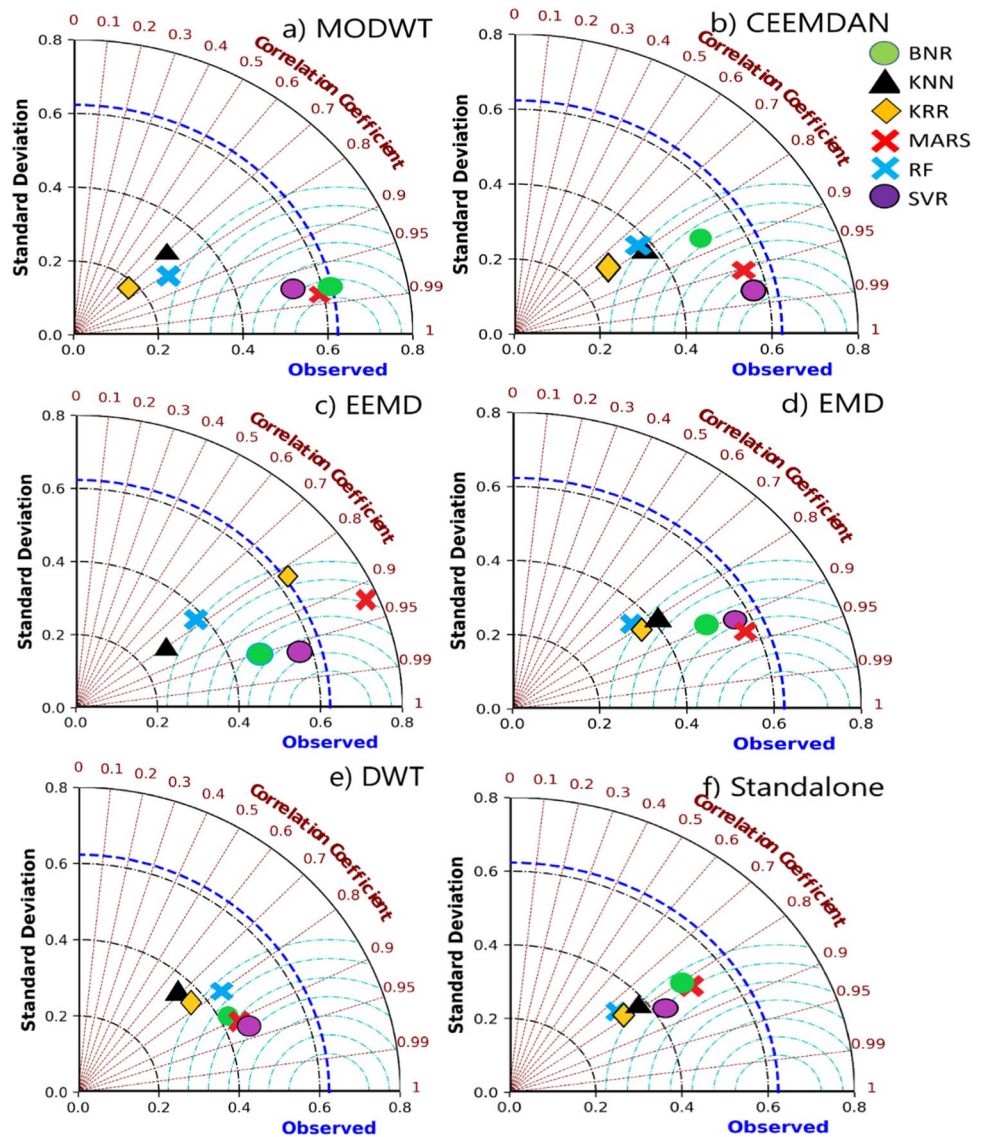


However, the y-intercept deviated from the ideal value with more outliers for the other models.

To attain a different interpretation of the proposed MODWT-MARS model's accuracy, the time series plot is used to comprehend the proposed model's forecasting ability. Figure 11 demonstrates the time series plot of forecasted and observed DO with MODWT-MARS compared to the standalone MARS model. Results show that the proposed MODWT-MARS model is found close to the observed DO revealed a high predictive accuracy. After applying the NCA algorithm as a feature selection approach and MODWT as a feature decomposition technique, the forecasted DO is enhanced.

Notably, five unique decomposition algorithms, EMD, EEMD, CEEMDAN, DWT, and MODWT, are incorporated to enhance the MARS-based predictive model. In terms of $r$, LM, and APB of DO forecasting, the MODWT effectively

forecasts improvement (Fig. 12). In the MARS model, $r$ and LM values using the MODWT model increased by ~19% and ~20% accordingly, and APB decreased by ~68%. Similarly, for the BNR model, MODWT feature decomposition skill increased $r$ and LM values up to ~21% and ~59% accordingly, and APB is decreased by ~57%. Additionally, $r$ and LM values for the MARS model with CEEMDAN are increased by ~15% and ~50%, respectively. Similarly, the inclusion of DWT, EMD, and EEMD also substantially improved the $r$, LM, and APB values.

## Discussion

According to the findings of this study, different input combinations have varying effects on the outcomes. Then, several input variables must be analysed, and the most
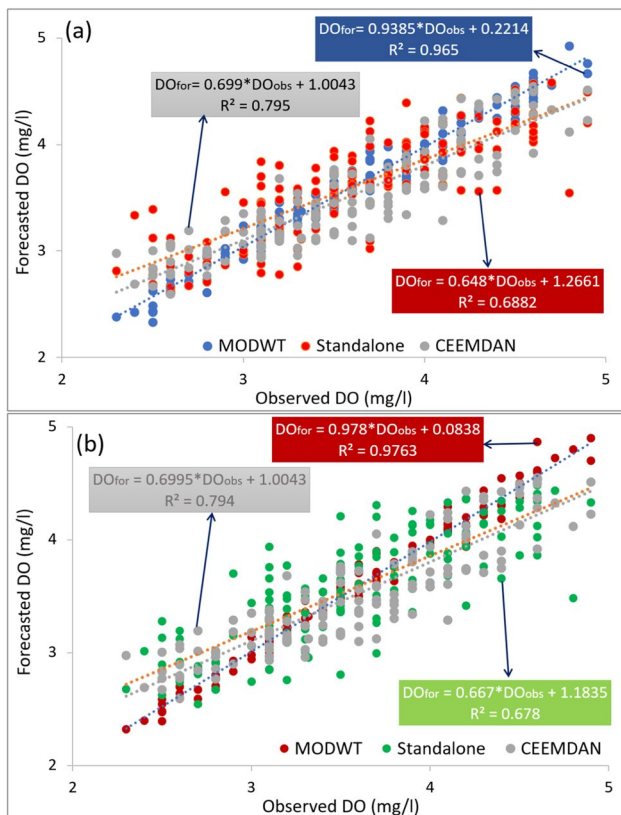
**Fig. 10** Scatter plot of forecasted vs observed DO, using **a** Bayesian ridge regression (BNR) and **b** multiple adaptive regression splines model using MODWT and CEEMDAN decomposition. A least square regression line and coefficient of determination ($R^2$) with a linear fit equation are shown in each sub-panel

appropriate collection of variables must be employed to optimise the products. Every model should have its ideal combination; yet the most effective combination is rare throughout the various models. Al-Musaylh et al. (2019) used the hybrid MARS model in forecasting electricity demand with a good performance. This study demonstrated profound forecasting of Dissolved Oxygen (DO)

concentration. Our findings have led to better forecasting than any algorithm evaluated in standalone and hybrid versions. We propose more studies to forecast DO using wet and dry season datasets and compare the results with the whole dataset's findings. Different pre-processing techniques could also enhance the projection accuracy of the MARS model. First, it is possible to implement a suitable feature selection approach such as NCA (Ahmed et al. 2021a; Ghimire et al. 2019a) algorithm to pick the input variables that significantly impact the model. The feature weight calculated using neighbourhood component analysis (NCA) respective to predictor variables was added one by one based on the highest to lowest feature weight to improve the model performance. The optimum combination of input parameters was found significantly in the proposed hybrid MARS model. By fitting piecewise linear regressions, MARS essentially creates flexible models by approximating the nonlinearity of a model using discrete linear regression slopes in various intervals of the independent variable space. An expansion in product spline basis functions of the predictors selected during a forwards and backwards recursive partitioning technique is how MARS best fits a model given a collection of predictor variables.

The time complexity of machine learning models is very important for the better application of the ML models. All the ML models used in our study shows less complex in terms of training time with less than 2 min for almost all the models. The incorporation of five feature decomposition approaches is vital to understanding the diverse implementation of the models in association with data pre-processing (i.e. feature selection and feature decomposition). The results showed that the inclusion of feature decomposition methods such as MODWT, CEEMDAN, EEMD, EMD, and DWT increased the performance of DO forecasting compared to the respective standalone methods. As MODWT can handle any sample size, the smooth and detail coefficients of MODWT filters and produces a more asymptotically efficient wavelet variance estimator than the

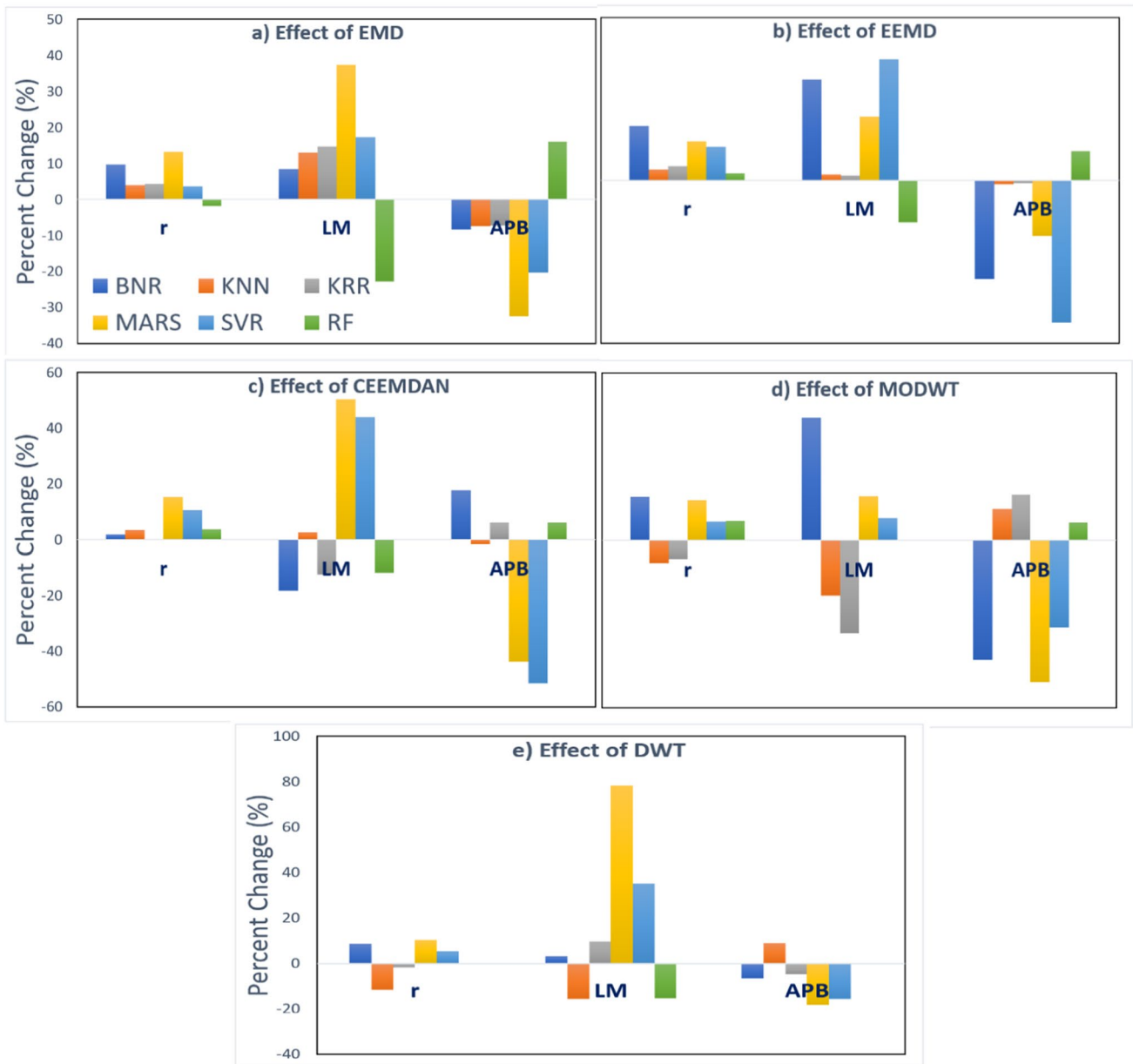**Fig. 11** Comparison between forecasted DO and observed DO during model testing using MODWT-MARS and Standalone MARS model
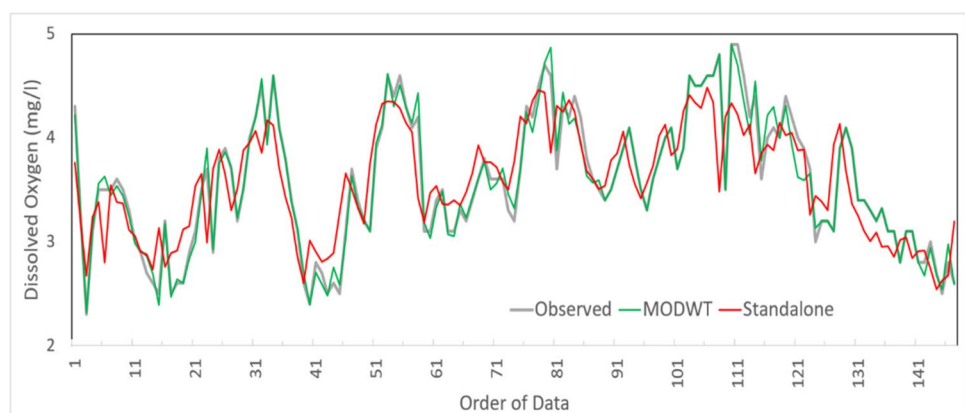
**Fig. 12** Effect of **a** EMD, **b** EEMD, **c** CEEMDAN, **d** MODWT, and **e** DWT of the performance of six models based on *r*, LM, and APB

DWT. However, the MODWT-MARS model was found as the optimum. Different researchers reported similar performance, where MODWT data decomposition is reported to improve performance (Li et al. 2017a; Prasad et al. 2017).

## Conclusions

This study developed hybrid machine learning models incorporating neighbourhood component analysis (NCA) as a feature selection method, multivariate adaptive regression splines (MARS) as a predictive model, and MODWT

as a feature decomposition method for DO forecast of the river Surma. The study used five distinct feature decomposition approaches (i.e. MODWT, CEEMDAN, EEMD, EMD, and DWT) and six machine learning models (i.e. BNR, KNN, KRR, MARS, RF, and SVR) for developing the optimum model. A new approach to the DO forecasting model was created using a decedent-lagged memory framework to explain the forecasting problem more appropriately and its consequences afterwards. The proposed MODWT-MARS approach provides the optimal performance among the benchmarked models. From this analysis, the following observations can be made.

1. The achieved results demonstrated that the NCA algorithm would be a helpful option for getting the predictor variables' substantial features. The model's performance metrics indicate that the NCA algorithm was a suitable tool for feature selection, as the NCA and MODWT optimised models showed better performance than the standalone models.

2. The proposed hybrid MODWT-MARS model outperformed all other models in forecasting the dissolved oxygen concentration of the Surma River. A low MAE (0.089) and a high NS (0.990) value substantiate the MODWT-MARS model's superiority. Correlation coefficient (r) values increased by 20%, and LM index values increased by 19% compared to their respective standalone models. To be more precise, the MODWT-MARS performed the best when r (0.981), WI (0.990), RMSE (0.121 mg/l), and MAE (0.089 mg/l) values were considered.

3. Based on the analysis, it is recognised that the hybrid MODWT-MARS model with the NCA feature algorithm shows superior forecasting of DO. The station's antecedent values of water quality parameters and hydro-meteorological variables embed the machine learning approach's future forecasting success. So, this type of forecast is used for better water quality management.

4. The current analysis strongly implies that MODWT and NCA methods with the MARS model can be used to forecast accurately. Their employment improved the accuracy of the MODWT-MARS model established in the current study and reduced the model's complexity by removing unnecessary input variables by incorporating NCA.

In addition to providing scientific benefits, the MARS low input need combined with their substantial predictive capability also provided significant practical benefits. They enable the development of a station-specific prudent predictive model of DO for monitoring river health at a minimal cost and the development of region-specific management plans across a range of land use and land cover gradients in a cost-effective manner.

## Declarations

## References

Adamowski J, Fung Chan H, Prasher SO, Ozga-Zielinski B, Sliusarieva A (2012) Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour Res 48(1)

Agbinya JI (1996) Discrete wavelet transform techniques in speech processing, Proceedings of Digital Processing Applications (TENCON'96). IEEE 2:514–519

Ahmed AAM (2017) Prediction of dissolved oxygen in Surma River by biochemical oxygen demand and chemical oxygen demand using the artificial neural networks (ANNs). J King Saud Univ Eng Sci 29(2):151–158

Ahmed MH, Lin L-S (2021) Dissolved oxygen concentration predictions for running waters with different land use land cover using a quantile regression forest machine learning technique. J Hydrol 597:126213

Ahmed AAM, Shah SMA (2017a) Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River. J King Saud Univ Eng Sci 29(3):237–243

Ahmed AAM, Shah SMA (2017b) Application of artificial neural networks to predict peak flow of Surma River in Sylhet Zone of Bangladesh. Int J Water 11(4):363–375

Ahmed AAM, Deo RC, Raj N, Ghahramani A, Feng Q, Yin Z, Yang L (2021a) Deep Learning forecasts of soil moisture: convolutional neural network and gated recurrent unit models coupled with satellite-derived MODIS, Observations and synoptic-scale climate index data. Remote Sens 13(4):554

Ahmed AAM, Deo RC, Ghahramani A, Raj N, Feng Q, Yin Z, Yang L (2021b) LSTM integrated with Boruta-random forest optimiser for soil moisture estimation under RCP4. 5 and RCP8. 5 global warming scenarios. Stoch Env Res Risk A 35(9):1851–1881

Ahmed AAM, Deo RC, Feng Q, Ghahramani A, Raj N, Yin Z, Yang L (2021c) Hybrid deep learning method for a week-ahead evapotranspiration forecasting. Stoch Env Res Risk A 36(3):831–849

Al-Musaylh MS, Deo RC, Adamowski JF, Li Y (2019) Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF Reanalysis atmospheric predictors in southeast Queensland. Renew Sust Energ Rev 113(2019):109293

Al-Musaylh MS, Deo RC, Li Y (2020) Electrical energy demand forecasting model development and evaluation with maximum overlap discrete wavelet transform-online sequential extreme learning machines algorithms. Energies 13(9):2307

Alalami MA, Maalouf M, EL-Fouly TH (2019) Wind Speed forecasting using kernel ridge regression with different time horizons, International Conference on Time Series and Forecasting, Springer, Cham, p 191–203

Ali M, Deo RC, Maraseni T, Downs NJ (2019) Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. J Hydrol 576:164–184

Ali M, Deo RC, Xiang Y, Li Y, Yaseen ZM (2020a) Forecasting long-term precipitation for water resource management: a new multi-step data-intelligent modelling approach. Hydrol Sci J 65(16):2693–2708

Ali M, Prasad R, Xiang Y, Yaseen ZM (2020b) Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. J Hydrol 584:124647

Arhami M, Kamali N, Rajabi MM (2013) Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. Environ Sci Pollut Res 20(7):4777–4789

Barrett P, Hunter J, Miller JT, Hsu JC, Greenfield P (2004) matplotlib – a portable python plotting package. In Proceedings of the Astronomical Data Analysis Software and Systems XIV, Pasadena, CA, p 91. [Google Scholar]

Beltrán-Castro J, Valencia-Aguirre J, Orozco-Alzate M, Castellanos-Domínguez G, Travieso-González CM (2013) Rainfall forecasting based on ensemble empirical mode decomposition and neural networks, International Work-Conference on Artificial Neural Networks. Springer, pp 471–480

Biswas R, Jayawardena A, Takeuchi K (2009) Prediction of water levels in the Surma River of Bangladesh by artificial neural network, Proceeding of 2009 Annual Conference

Breiman L (2001) Random forests. Mach Learn 45:5–32

Bruce LM, Koger CH, Li J (2002) Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. IEEE Trans Geosci Remote Sens 40(10):2331–2338

Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geosci Mod Dev 7(3):1247–1250

Chowdhury RK, Ali SIM (2006) Investigation of phosphate and ammonia-nitrogen concentrations at some selected locations of the Malnichara channel and the Surma river. ARPN Journal of Engineering and Applied Sciences 1(2)

Chu H, Wei J, Wu W (2020) Streamflow prediction using LASSO-FCM-DBN approach based on hydro-meteorological condition classification. J Hydrol 580(2020):124253

Cornish CR, Bretherton CS, Percival DB (2006) Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. Bound-Layer Meteorol 119(2):339–374

Dash P, Majumder I, Nayak N, Bisoi R (2020) Point and interval solar power forecasting using hybrid empirical wavelet transform and robust wavelet kernel ridge regression. Nat Resour Res 29(5):2813–2841

Deo RC, Sahin M (2016) An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. Environ Monit Assess 188(2):90

Deo RC, Wen X, Qi F (2016) A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Appl Energy 168:568–593

Deo RC, Downs N, Parisi AV, Adamowski JF, Quilty JM (2017a) Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. Environ Res 155:141–166

Deo RC, Kisi O, Singh VP (2017b) Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmos Res 184:149–175

Di C, Yang X, Wang X (2014) A four-stage hybrid model for hydrological time series forecasting. PLoS ONE 9(8):e104663

Douak F, Melgani F, Benoudjit N (2013) Kernel ridge regression with active learning for wind speed prediction. Appl Energy 103:328–340

Flandrin P, Rilling G, Goncalves P (2004) Empirical mode decomposition as a filter bank. IEEE Signal Process Lett 11(2):112–114

Forough K-T, Mousavi S-F, Khaledian M, Yousefi-Falakdehi O, Norouzi-Masir M (2019) Prediction of water quality index by support vector machine: a case study in the Sefidrud Basin, Northern Iran. Water Resour 46(1):112–116

Fowler JE (2005) The redundant discrete wavelet transform and additive noise. IEEE Signal Process Lett 12(9):629–632

Friedman JH (1991) Multivariate adaptive regression splines. The annals of statistics 19(1):1–67

Ghiassi M, Fa'al F, Abrishamchi A (2017) Large metropolitan water demand forecasting using DAN2, FTDNN, and KNN models: a case study of the city of Tehran, Iran. Urban Water J 14(6):655–659

Ghimire S, Deo RC, Downs NJ, Raj N (2019a) Deep learning neural networks trained with MODIS satellite-derived predictors for long-term global solar radiation prediction. Energies 12(12):2407

Ghimire S, Deo RC, Downs NJ, Raj N (2019b) Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia. J Clean Prod 216:288–310

Ghimire S, Deo RC, Raj N, Mi J (2019c) Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Applied Energy 253:113541

Goyal MK, Bharti B, Quilty J, Adamowski J, Pandey A (2014) Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert Syst Appl 41(11):5267–5276

He F, Zhang Y, Liu D, Dong Y, Liu C, Wu C (2017) Mixed wavelet-based neural network model for cyber security situation prediction using MODWT and Hurst exponent analysis, In International Conference on Network and System Security Springer, Cham, p 99–111

Heddam S (2016) Use of optimally pruned extreme learning machine (OP-ELM) in forecasting dissolved oxygen concentration (DO) several hours in advance: a case study from the Klamath River, Oregon, USA. Environ Process 3(4):909–937

Heddam S (2017) Fuzzy neural network (EFuNN) for modelling dissolved oxygen concentration (DO). Intelligence Systems in Environmental Management: Theory and Applications. Springer, pp 231–253

Heddam S, Kisi O (2017) Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors. Environ Sci Pollut Res 24(20):16702–16724

Heddam S, Kisi O (2018) Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. J Hydrol 559:499–509

Henderson RK, Baker A, Murphy K, Hambly A, Stuetz R, Khan S (2009) Fluorescence as a potential monitoring tool for recycled water systems: a review. Water Res 43(4):863–881

Hoang N-D, Pham A-D, Cao M-T (2014) A novel time series prediction approach based on a hybridization of least squares support vector regression and swarm intelligence. Applied Computational Intelligence and Soft Computing 15(8). https://doi.org/10.1155/2014/754809

Huang H, Abdel-Aty M (2010) Multilevel data and Bayesian analysis in traffic safety. Accid Anal Prev 42(6):1556–1565

Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N-C, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond Ser A Math Phys Eng Sci 454(1971):903–995

Hur J, Cho J (2012) Prediction of BOD, COD, and total nitrogen concentrations in a typical urban river using a fluorescence excitation-emission matrix with PARAFAC and UV absorption indices. Sensors 12(1):972–986

Iglesias C, Torres JM, Nieto PG, Fernández JA, Muñiz CD, Piñeiro J, Taboada J (2014) Turbidity prediction in a river basin by using artificial neural networks: a case study in northern Spain. Water Resour Manag 28(2):319–331

Islam MS, Hoque F (2014) River bank erosion of the Surma River due to slope failure. Int J Res Innov Earth Sci 1(2):54–58

Jiao G, Guo T, Ding Y (2016) A new hybrid forecasting approach applied to hydrological data: a case study on precipitation in Northwestern China. Water 8(9):367

Jui SJJ, Ahmed AAM, Bose A, Raj N, Sharma E, Soar J, Chowdhury MWI (2022) Spatiotemporal hybrid random forest model for tea yield prediction using satellite-derived variables. Remote Sens 14(3):805

Keshtegar B, Heddam S, Hosseinabadi H (2019) The employment of polynomial chaos expansion approach for modeling dissolved oxygen concentration in river. Environ Earth Sci 78(1):1–18

Khaldi K, Alouane MT-H, Boudraa A-O (2008) A new EMD denoising approach dedicated to voiced speech signals, 2nd International Conference on Signals, Circuits and Systems. IEEE, pp 1–5

Kisi O, Ay M (2012) Comparison of ANN and ANFIS techniques in modeling dissolved oxygen. Proceedings of the Sixteenth International Water Technology Conference (IWTC 16), Istanbul, Turkey, pp 7–10

Kisi O, Parmar KS (2016) Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. J Hydrol 534:104–112

Kisi O, Alizamir M, Gorgij AD (2020) Dissolved oxygen prediction using a new ensemble method. Environ Sci Pollut Res 27(9):9589–9603

Krause P, Boyle D, Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment. Adv Geosci 5:89–97

Kuhn M, Johnson K (2013) Applied predictive modeling, 26. Springer

Kuo YM, Liu CW, Lin KH (2004) Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan. Water Res 38(1):148–158

Legates DR, McCabe GJ (1999) Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resour Res 35(1):233–241

Legates DR, McCabe GJ (2013) A refined index of model performance: a rejoinder. Int J Climatol 33(4):1053–1056

Li M, Chen W, Zhang T (2017a) Application of MODWT and log-normal distribution model for automatic epilepsy identification. Biocybern Biomed Eng 37(4):679–689

Li X, Sha J, Wang Z-L (2017b) A comparative study of multiple linear regression, artificial neural network and support vector machine for the prediction of dissolved oxygen. Hydrol Res 48(5):1214–1225

Lin H (2003) Hydropedology: Bridging disciplines, scales, and data. Vadose Zone J 2(1):1–11

Liu M, Huang Y, Li Z, Tong B, Liu Z, Sun M, Jiang F, Zhang H (2020) The Applicability of LSTM-KNN model for real-time flood forecasting in different climate zones in China. Water 12(2):440

Maity R, Bhagwat PP, Bhatnagar A (2010) Potential of support vector regression for prediction of monthly streamflow using endogenous property. Hydrol Process 24(7):917–923

Mellios N, Kofinas D, Laspidou C, Papadimitriou T (2015) Mathematical modeling of trophic state and nutrient flows of Lake Karla using the PCLake model. Environ Process 2(1):85–100

Mishra S, Dhar S, Dash P (2019) An effective battery management scheme for wind energy systems using multi Kernel Ridge regression algorithm. J Energy Storage 21:418–434

Mohan S, Kumar KP (2016) Waste load allocation using machine scheduling: model application. Environ Process 3(1):139–151

Mouri G, Takizawa S, Oki T (2011) Spatial and temporal variation in nutrient parameters in stream water in a rural-urban catchment, Shikoku, Japan: Effects of land cover and human impact. J Environ Manag 92(7):1837–1848

Naik J, Satapathy P, Dash P (2018) Short-term wind speed and wind power prediction using hybrid empirical mode decomposition and kernel ridge regression. Appl Soft Comput 70:1167–1188

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—a discussion of principles. J Hydrol 10(3):282–290

Nourani V, Komasi M, Mano A (2009) A multivariate ANN-wavelet approach for rainfall–runoff modeling. Water Resour Manag 23(14):2877–2894

Nourani V, Baghanam AH, Adamowski J, Kisi O (2014) Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. J Hydrol 514:358–377

Ouyang Q, Lu W, Xin X, Zhang Y, Cheng W, Yu T (2016) Monthly rainfall forecasting using EEMD-SVR based on phase-space reconstruction. Water Resour Manag 30(7):2311–2325

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011a) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011b) Scikit-learn: machine learning in Python. J Mach Learn Res 12(Oct):2825–2830

Prasad R, Deo RC, Li Y, Maraseni T (2017) Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. Atmos Res 197:42–63

Prasad R, Deo RC, Li Y, Maraseni T (2018) Soil moisture forecasting by a hybrid machine learning technique: ELM integrated with ensemble empirical mode decomposition. Geoderma 330:136–161

Prasad R, Deo RC, Li Y, Maraseni T (2019) Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. CATENA 177:149–166

Quilty J, Adamowski J (2018) Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. J Hydrol 563:336–353

Raheli B, Aalami MT, El-Shafie A, Ghorbani MA, Deo RC (2017) Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: a case study of Langat River. Environ Earth Sci 76(14):1–16

Ranković V, Radulović J, Radojević I, Ostojić A, Čomić L (2010) Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. Ecol Model 221(8):1239–1244

Rathinasamy M, Khosa R, Adamowski J, Ch S, Partheepan G, Anand J, Narsimlu B (2014) Wavelet-based multiscale performance analysis: an approach to assess and improve hydrological models. Water Resour Res 50(12):9721–9737

Rezaie-Balf M, Maleki N, Kim S, Ashrafian A, Babaie-Miri F, Kim NW, Chung I-M, Alaghmand S (2019) Forecasting daily solar radiation using CEEMDAN decomposition-based MARS model trained by crow search algorithm. Energies 12(8):1416

Rudy J, Cherti M (2017) Py-earth: a python implementation of multivariate adaptive regression splines. https://github.com/scikit-learn-contrib/py-earth

Saqib M (2021) Forecasting COVID-19 outbreak progression using hybrid polynomial-Bayesian ridge regression model. Appl Intell 51(5):2703–2713

Saunders C, Gammerman A, Vovk V (1998) Ridge regression learning algorithm in dual variables, Appears in Proceedings of the 15th International Conference on Machine Learning, ICML515-521

Seo Y, Kim S (2016) Hydrological forecasting using hybrid data-driven approach. Am J Appl Sci 13(8):891–899

Shabani S, Samadianfard S, Sattari MT, Mosavi A, Shamshirband S, Kmet T, Várkonyi-Kóczy AR (2020) Modeling pan evaporation using gaussian process regression k-nearest neighbors random forest and support vector machines. Comp Anal Atmos 11(1):66

Sharma E, Deoa RC, Prasadb R, Parisia AV (2019) A hybrid air quality early-warning framework: hourly forecasting model with online sequential extreme learning machine and empirical mode decomposition algorithm. Sci Total Environ 709:135934:1–23

Shensa MJ (1992) The discrete wavelet transform: wedding the a trous and Mallat algorithms. IEEE Trans Signal Process 40(10):2464–2482

Stull KE, L'Abbé EN, Ousley SD (2014) Using multivariate adaptive regression splines to estimate subadult age from diaphyseal dimensions. Am J Phys Anthropol 154(3):376–386

Su S, Li D, Zhang Q, Xiao R, Huang F, Wu J (2011) Temporal trend and source apportionment of water pollution in different functional zones of Qiantang River, China. Water Res 45(4):1781–1795

Suen J-P, Eheart JW (2003) Evaluation of neural networks for modeling nitrate concentrations in rivers. J Water Resour Plan Manag 129(6):505–510

Suykens JA, De Brabanter J, Lukas L, Vandewalle J (2002) Weighted least squares support vector machines: robustness and sparse approximation. Neurocomputing 48(1–4):85–105

Tao H, Bobaker AM, Ramal MM, Yaseen ZM, Hossain MS, Shahid S (2019) Determination of biochemical oxygen demand and dissolved oxygen for semi-arid river environment: application of soft computing models. Environ Sci Pollut Res 26(1):923–937

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106(D7):7183–7192

Tiwari MK, Adamowski J (2013) Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour Res 49(10):6486–6507

Tiwari MK, Chatterjee C (2010) Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. J Hydrol 394(3–4):458–470

Tiwari MK, Chatterjee C (2011) A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting. J Hydroinf 13(3):500–519

Tomic SA, Antanasijevic D, Ristic M, Peric-Grujic A, Pocajt V (2018) A linear and non-linear polynomial neural network modeling of dissolved oxygen content in surface water: Inter- and extrapolation performance with inputs' significance analysis. Sci Total Environ 610–611:1038–1046

Torres ME, Colominas MA, Schlotthauer G, Flandrin P (2011) A complete ensemble empirical mode decomposition with adaptive noise, 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 4144–4147

US-Geological-Survey (2016) The US Geological Survey, digital spectral reflectance library. https://usgs.gov

Waskom M, Botvinnik O, Ostblom J, Gelbart M, Lukauskas S, Hobson P, Gemperline DC, Augspurger T, Halchenko Y, Cole JB (2020) Mwaskom/Seaborn: v0.10.1 (April 2020). Zenodo. 2020. Available online: https://ui.adsabs.harvard.edu/abs/2020zndo...3767070W%2F/abstract. Accessed 25 Dec 2021

Wen X, Fang J, Diao M, Zhang C (2013) Artificial neural network modeling of dissolved oxygen in the Heihe River, Northwestern China. Environ Monit Assess 185(5):4361–4371

Willmott CJ, Ackleson SG, Davis RE, Feddema JJ, Klink KM, Legates DR, O'donnell J, Rowe CM (1985) Statistics for the evaluation and comparison of models. J Geophys Res Oceans 90(C5):8995–9005

Willmott CJ, Robeson SM, Matsuura K (2012) A refined index of model performance. Int J Climatol 32(13):2088–2094

Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Philip SY (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14(1):1–37

Xiang S, Liu Z, Ma L (2006) Study of multivariate linear regression analysis model for ground water quality prediction. Guizhou Sci 24(1):60–62

Yang P, Xia J, Zhang Y, Hong S (2017) Temporal and spatial variations of precipitation in Northwest China during 1960–2013. Atmos Res 183:283–295

Yaseen ZM, Jaafar O, Deo RC, Kisi O, Adamowski J, Quilty J, El-Shafie A (2016) Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. J Hydrol 542:603–614

Yaseen ZM, Sulaiman SO, Deo RC, Chau K-W (2019) An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. J Hydrol 569:387–408

Yin Z, Feng Q, Wen X, Deo RC, Yang L, Si J, He Z (2018) Design and evaluation of SVR, MARS and M5Tree models for 1, 2 and 3-day lead time forecasting of river flow data in a semiarid mountainous catchment. Stoch Environ Res Risk Assess 32(9):2457–2476

You Y, Demmel J, Hsieh C-J, Vuduc R (2018) Accurate, fast and scalable kernel ridge regression on parallel and distributed systems, Proceedings of the 2018 International Conference on Supercomputing, pp 307–317

Yu P-S, Yang T-C, Chen S-Y, Kuo C-M, Tseng H-W (2017) Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. J Hydrol 552:92–104

Zhang W, Goh AT (2016) Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geosci Front 7(1):45–52

Zhang W, Qu Z, Zhang K, Mao W, Ma Y, Fan X (2017) A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. Energy Convers Manag 136:439–451

Zhang S, Zhou T, Sun L, Liu C (2019) Kernel ridge regression model based on beta-noise and its application in short-term wind speed forecasting. Symmetry 11(2):282

Zhou Y, Li T, Shi J, Qian Z (2019) A CEEMDAN and XGBOOST-based approach to forecast crude oil prices. Complexity 2019:1–15

Zhu S, Heddam S (2020) Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN). Water Qual Res J 55(1):106–118

Zounemat-Kermani M, Seo Y, Kim S, Ghorbani MA, Samadianfard S, Naghshara S, Kim NW, Singh VP (2019) Can decomposition approaches always enhance soft computing models? Predicting the dissolved oxygen concentration in the St. Johns River, Florida. Appl Sci 9(12):2534