



# Pre- and post-dam river water temperature alteration prediction using advanced machine learning models

Dinesh Kumar Vishwakarma<sup>1</sup> · Rawshan Ali<sup>2</sup> · Shakeel Ahmad Bhat<sup>3</sup> · Ahmed Elbeltagi<sup>4</sup> · Nand Lal Kushwaha<sup>5</sup> · Rohitashw Kumar<sup>3</sup> · Jitendra Rajput<sup>5</sup> · Salim Heddad<sup>6</sup> · Alban Kuriqi<sup>7</sup>

Received: 5 April 2022 / Accepted: 16 June 2022 / Published online: 28 June 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Dams significantly impact river hydrology by changing the timing, size, and frequency of low and high flows, resulting in a hydrologic regime that differs significantly from the natural flow regime before the impoundment. For precise planning and judicious use of available water resources for agricultural operations and aquatic habitats, it is critical to assess the dam water's temperature accurately. The building of dams, particularly several dams in rivers, can significantly impact downstream water. In this study, we predict the daily water temperature of the Yangtze River at Cuntan. Thus, this work reveals the potential of machine learning models, namely, M5 Pruned (M5P), Random Forest (RF), Random Subspace (RSS), and Reduced Error Pruning Tree (REPTree). The best and effective input variables combinations were determined based on the correlation coefficient. The outputs of the various machine learning algorithm models were compared with recorded daily water temperature data using goodness-of-fit criteria and graphical analysis to arrive at a final comparison. Based on a number of criteria, numerical comparison between the models revealed that M5P model performed superior ( $R^2 = 0.9920$ ,  $0.9708$ ;  $PCC = 0.9960$ ,  $0.9853$ ;  $MAE = 0.2387$ ,  $0.4285$ ;  $RMSE = 0.3449$ ,  $0.4285$ ;  $RAE = 6.2573$ ,  $11.5439$ ;  $RRSE = 8.0288$ ,  $13.8282$ ) in pre-impact and post-impact spam, respectively. These findings suggest that a huge wave of dam construction in the previous century altered the hydrologic regimes of large and minor rivers. This study will be helpful for the ecologists and river experts in planning new reservoirs to maintain the flows and minimize the water temperature concerning spillway operation. Finally, our findings revealed that these algorithms could reliably estimate water temperature using a day lag time input in water level. They are cost-effective techniques for forecasting purposes.

**Keywords** Dams · Machine learning · Water temperature · M5P · RF · REPTree · RSS

## Introduction

Building dam reservoirs are one of the other oldest branches of engineering. Historically, human civilization developed on rivers. As humanity expanded and advanced worldwide, the number of constructed dams has increased, especially in nearly every water body region (Olden and Naiman 2010; Rheinheimer et al. 2015). Among the common and significant roles that dams play are water storage, water volume control, and flood protection which have not yet fully understood the ecologies of the global riverine system. Reservoirs

and dams and their operation can affect riverine ecology, including changing riverine thermal regimes and water temperature fluctuation alongside the rivers (Olden and Naiman 2010; Rheinheimer et al. 2015). Water temperatures can affect aquatic species' health, distribution, and functions (Jiang et al. 2018); therefore, as the number of constructed dams increases, the understanding of water temperature variation of fluctuation has become a priority for ecological researchers (Murchie et al. 2008; Olden and Naiman 2010). Indeed, it was shown that the dam's water release mechanism is the major and critical factor controlling the water temperature downstream of the dams (Tao et al. 2020). Generally, a high volume of cold water was passed down through "deep portals" beneath the thermocline, especially the hypolimnetic layer (Olden and Naiman 2010; Kushwaha and Bhardwaj 2016). Although this is a rare occurrence, water was passed down above the thermocline specifically

---

Responsible Editor: Marcus Schulz

---

✉ Dinesh Kumar Vishwakarma  
dinesh.vishwakarma4820@gmail.com

Extended author information available on the last page of the article

the hypolimnetic layer, causing an increase in the downstream water temperatures (Cheng et al. 2020).

Reservoirs impact the seasonal and annual thermal patterns of downstream water temperature. Indeed, it was demonstrated that, during the spring and summer seasons, water temperature fluctuation in large reservoirs was moved toward a decreased direction compared to the winter season, for which negligible fluctuation has been experienced. Compared to the well-informed natural rivers, a significant delay for the maxima values was exhibited (Olden and Naiman 2010). For comparison, it was shown that many dams constructed worldwide had encountered similar phenomena, among them are Hills Creek Dam in the USA (Angilletta Jr et al. 2008), the controlled dam in Scotland (Jackson et al. 2007), and the Burrendong Dams in Australia (Ryan et al. 2001; Preece 2004).

The impounded reservoirs behind the dams significantly influence the temperature regimes alongside the dam's river. Water is released through the dam at the upstream reservoir channels (Ali et al. 2019b). Temperature gradients observed over a long period are generally used as an alternative to assess a free-flowing river's natural thermodynamics within impounded water. They significantly affect the marine life upstream and downstream of the diversion of impoundment. Consequently, overall marine aquatic life is highly vulnerable to temperature fluctuation. All these marine organisms must adapt, relocate, or perish in response to the impacts of thermal regime modification.

The greatest production of power electricity in the world is guaranteed by the Three Gorges Dam (TGD). Also, it possesses the biggest stored water volume (Wu et al. 2012). Regarding its high hydraulic, hydrological, and ecological importance, a large number of investigations have been conducted over the TGD, i.e., hydrological alteration (Gao et al. 2012; Yu et al. 2017b; Wang et al. 2017), investigating the streamflow variation conducted by Gao et al. (2012), highlighting that the TGD has significantly contributed to the decreasing in the calculated downstream flow section. It has helped reduce the peak flows (Ali et al. 2019c).

Over the past two decades, artificial intelligence (AI) and machine learning techniques have been successfully developed and widely used for estimating and predicting (Citakoglu and Coşkun 2022), in particular, modeling non-linear hydrologic systems and agriculture field (Shukla et al. 2021), meteorological droughts and standardized precipitation index (SPI) (Malik et al. 2021; Xu et al. 2022), lake water level (Zhu et al. 2020), rainfall forecasting (Luk et al. 2001; Olsson et al. 2004; Abbot and Marohasy 2012; Lee et al. 2018; Mirabbasi et al. 2019; Adnan et al. 2020; Armin et al. 2021; Khosravi et al. 2022), streamflow forecasting (Yaseen et al. 2016; Shukla et al. 2021; Khodakhah et al. 2022), hydrological drought (Shamshirband et al. 2020; Aghelpour et al. 2021; Muhammad et al. 2021; Almikael et al. 2022), pan evaporation forecasting (Shiri

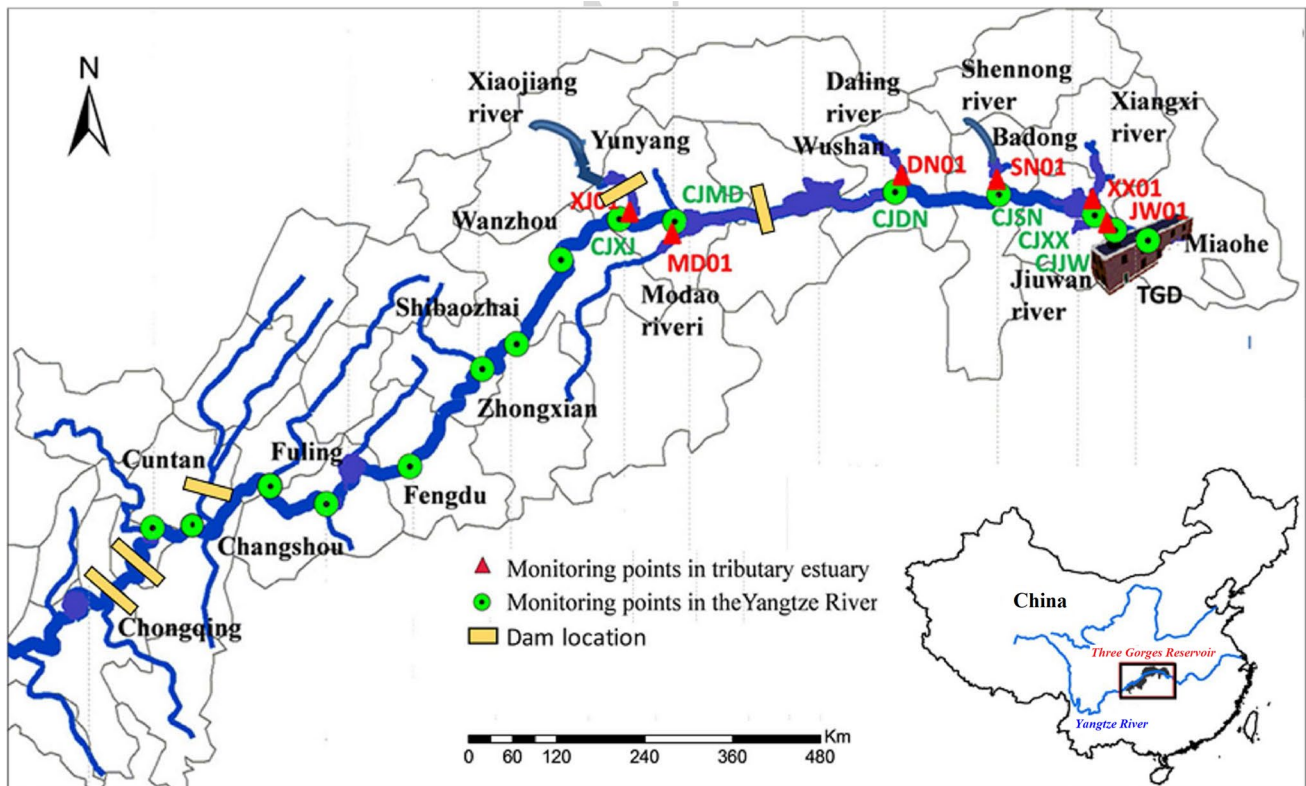
and Özgür 2011; Mohammad et al. 2019; Malik et al. 2020; Al-Mukhtar 2021; Kushwaha et al. 2021), evapotranspiration (Granata 2019; Wu et al. 2019; Tikhamarine et al. 2019, 2020; Chen et al. 2020; Chia et al. 2020; Ferreira and da Cunha 2020; Elbeltagi et al. 2022b), water level forecasting (Daliakopoulos et al. 2005; Nayak et al. 2006; Ali Ghorbani et al. 2010; Kisi et al. 2012; Buyukyildiz et al. 2014; Seo et al. 2015, 2017), velocity predictions in compound channels with vegetated floodplains (Harris et al. 2003), suspended sediment load prediction (Melesse et al. 2011; Rajaei et al. 2011; Azamathulla et al. 2013; Kakaei Lafdani et al. 2013; Gupta et al. 2021), soil temperature (Yang and Wang 2008; Bilgili 2010; Singh et al. 2018; Penghui et al. 2020), water quality (Singh et al. 2021b), groundwater quality variables (Esmailbeiki et al. 2020; Che Nordin et al. 2021; El Bilali et al. 2021; Singha et al. 2021; Shiri et al. 2021; Singh et al. 2022), soil permeability (Singh et al. 2020, 2021a; Özçoban et al. 2022), soil hydraulic conductivity (Allah et al. 2014; Sihag et al. 2019a; Singh et al. 2019; Araya and Ghezzehei 2019), runoff and suspended sediment simulation (Sharma et al. 2015; Kumar et al. 2019), soil infiltration (Kashi et al. 2014; Sihag et al. 2019b; Panahi et al. 2021; Sayari et al. 2021; Angelaki et al. 2021), global solar radiation (Hassan et al. 2017; Voyant et al. 2017; Cornejo-Bueno et al. 2019; Feng et al. 2019; Ağbulut et al. 2021), dew point temperature (Naganna et al. 2019; Qasem et al. 2019; Alizamir et al. 2020), chezy resistance coefficient in corrugated channels (Giustolisi 2004), manning's roughness coefficient in flows, (Bahramifar et al. 2013; Pradhan and Khatua 2017; Mohanta et al. 2018), and drought- and stress-tolerance (Kumar et al. 2022).

The main aim of this work is to provide an experimental evaluation of the effect of dams on river water temperature fluctuation. The study considered river water temperature over many years before and after selecting reservoirs (Kuriqi et al. 2020). The study findings are expected to allow users to establish a direct effect of the TDG on the river's thermal regime. The findings of this study provide an insight into future development projects; for instance, it can present valuable information and a priori view to support the engineers and practitioners to implement the structures to be constructed to cope with the floods and droughts when looking at prevailing climatic events. The finding of the study can be beneficial in planning and management of water resources at Yangtze River.

## Materials and methods

### Study area and climate characterization

China is blessed with an abundant number of rivers flowing from north to south, including the Yangtze River, among others. The Yangtze River is one of the longest rivers around



**Fig. 1** Locations of the hydrologic stations and the Yangtze River

the world, which collects water from several catchments. This paper uses the Yangtze River located in China with in the Coordinates latitude  $29.7204^{\circ}$  N, longitude  $112.6501^{\circ}$  E as a case study. It flows from Qinghai's southwest corner to Shanghai's north end. The river basin is approximately 1.8 million  $\text{km}^2$  in size. It provides approximately  $892 \text{ km}^3$  of water calculated as a river discharge for the period ranging from 1950 to 2010 (Yang and Lu 2012; Liu et al. 2018). The monsoon is a dominant component in this region. It is designed for the transportation of moist air, starting from the East and ending in the south China Sea, according to spatiotemporal data of rainfall alongside the river basin (Li et al. 2014; Wu et al. 2018), and there are numerous precipitation patterns over time (Zhao and Shepherd 2012). Summers receive a large amount of precipitation, leading to floods (Wu et al. 2012; Zhao and Shepherd 2012). The river is about 6400 km long and is Asia's longest river (Vezzoli et al. 2016; Ali et al. 2019b). Due to the river's length, nearly 50,000 reservoirs of various sizes have been built. The sources of nitrogen and phosphorus were highly influenced by the spatiotemporal fluctuation of the Yangtze River (Liu et al. 2018; Ali et al. 2019a). From year to year, it was shown from several conducted investigations that the natural aquatic habitat was significantly affected by the TGD project, whether at the upstream or the downstream locations of

the dam (Yu et al. 2017a). As a result, three stations on the dam's upstream sides were chosen for investigation in this study (Fig. 1) to depict the stations' positions. All stations were chosen according to their geographical situation and the availability of high-quality data. The Hydrologic Data Centre of China's Ministry of Water Resources provided the mean daily river stations and the Yangtze River afterdata.

### Mann-Kendall trend analysis

The Mann-Kendall statistical test for trend is used to assess whether a set of data values is increasing over time or decreasing over time and whether the trend in either direction is statistically significant. The Mann-Kendall test does *not* assess the magnitude of change. There are several trend assessments approaches available in the literature. However, the Mann-Kendall test is the most widely used test for assessing the trends in hydro-climatic studies. The Mann-Kendall test (Ahmed et al. 2017; Ali et al. 2019c), which is recommended by the World Meteorological Organization (WMO) often used as because it has several advantages: it does consider the data distribution, and it can cope with the outliers (Ali et al. 2019c). For a time-series data points  $Y = \{x_1, x_2, x_3, x_4, x_5, \dots, x_n\}$  with  $n > 10$ . The Mann-Kendall test

statistic,  $S$  is calculated as (Haktanir and Citakoglu 2014; Tefaruk and Hatice 2015; Citakoglu and Minarecioglu 2021)

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \operatorname{sgn}(x_j - x_k) \quad (1)$$

where  $n$  is the number of data points and  $\operatorname{sgn}(x_j - x_k)$  is calculated as

$$\operatorname{sgn}(x_j - x_k) = \begin{cases} 1 & \text{for } (x_j - x_k) > 0 \\ 0 & \text{for } (x_j - x_k) = 0 \\ -1 & \text{for } (x_j - x_k) < 0 \end{cases} \quad (2)$$

If we assume that selected data points are independent and randomly ordered, the mean of  $S = 0$  and the variance of M.K. statistics [ $\operatorname{Var}(S)$ ] is given by

$$\operatorname{Var}(S) = \frac{\left[ m(m-1)(2m+5) - \sum_{p=1}^q (t_p - 1)t_p(2t_p + 5) \right]}{18} \quad (3)$$

where  $q$  is the number of groups of tied rank, each with  $t_p$  tied observation. A tied group is a set of the same values in a selected dataset. The standard normal test statistic ( $Z$ ) is calculated as

$$Z = \frac{S - \operatorname{sgn}(S)}{\operatorname{Var}(S)^{1/2}} \quad (4)$$

The Sen's slope (S.S.) is represented by calculating the slope as a change in measurement per unit change in time:

$$S.S. = \operatorname{Median} \left[ \frac{w_j - w_i}{j - i} \right] \quad (5)$$

where  $w_j$  and  $w_i$  have represented the values of information at the time  $i$  and  $j$ , respectively, for all  $i < j$ .

Based on the Mann-Kendall test, the M-K significance of monthly, yearly, and seasonal dam temperature trends is assessed and tabulated in Table 1. Table 1 indicates that, during the January, February, April, June, October, November, and December months, the increasing temperature trend and the rest of the month were found to decrease but were statistically not significant in both cases. In Fig. 2, we depicted water temperature fluctuation at three different trends: monthly, yearly, and seasonal. Indeed, it is clear that the statistical test (Table 1) confirmed that a statistically positive trend could be highlighted. In addition, yearly and monthly fluctuation of water temperatures follows a rapidly ascending curve during the period of record, which is statistically significant after the dam project's realization. Taking into account the water temperature anomalies, it is clear that fitting the mean yearly fluctuation of the water temperature using a linear fit led to detect a non-significant

**Table 1** Statistic and change percentage (2010–2015)

Temperature trend (2010–2015)			
Time series	Mann-Kendall trend Test $S$ value	Significance	Sen's slope
January	7	-	0.429
February	5	-	0.148
March	-1	-	-0.029
April	7	-	0.262
May	-5	-	-0.238
June	5	-	0.063
July	-1	-	-0.019
August	-5	-	-0.294
September	-5	-	-0.083
October	7	-	0.354
November	9	-	0.323
December	9	-	0.335
Annual	5	-	0.072
Annual	5	-	0.072
Spring (April to May)	-1	-	-0.030
Summer (June to August)	-1	-	-0.015
Autumn (Sep to Oct)	5	-	0.165
Winters (Nov to Mar)	5	-	0.206

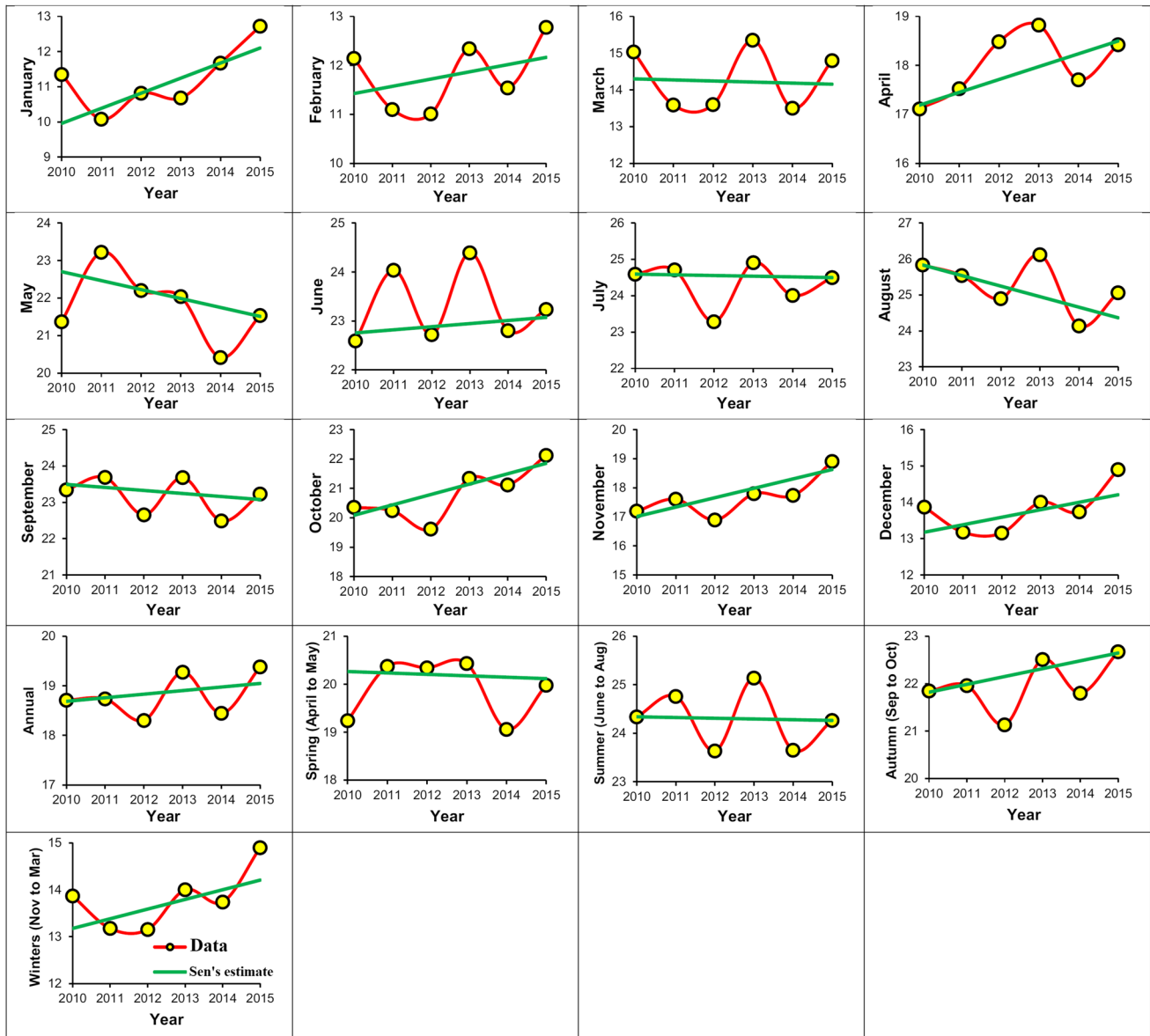
– values indicate statistical not significance as per the Mann-Kendall test (+ for increasing and – for decreasing trend).

and high trend of approximately  $\approx 0.072^\circ\text{C}$  for each year for the average water temperature and the seasonal water temperature was increased by approximately  $0.082^\circ\text{C}$  during the period ranging from 2010 to 2015. During the autumn (September to October) and winters (November to March) seasons, positive trends were detected by 0.165 and 0.206  $0.082^\circ\text{C}$ , respectively, and spring (April to May) and summer (June to August) seasons were detected negatively by  $-0.030$ ,  $-0.015^\circ\text{C}$ , respectively.

Factory sites around the city, a rise in human activity, and a lack of green spaces and parks all contribute to the city's warming. Furthermore, the mountains surrounding the city act as natural windshields, impeding smooth air circulation and contributing to the city's heat. It satisfies the accuracy requirement that the temperature simulation in the reservoirs essentially agrees with the recorded data, allowing the developed model to accurately simulate the trends and evolution of water temperature structure over space and time at the Xiangjiaba and Xiluodu reservoirs.

Seasonal variation exists in the stratification of water temperature in the Xiangjiaba Reservoir. It was almost visible from April to August and then vanished in other months. The surface water temperature rises increasingly rapidly in spring, and the stratification steadily intensifies. Due to the reservoir's flood, the bottom water temperature rose





**Fig. 2** Fluctuation and trend of monthly, yearly, and seasonally temperature at Yangtze River in China (2010–2015)

rapidly in the summer. The thickness of the isothermal layer increased, and the treatment decreased due to the strong vertical turbulent diffusion. The storage leads to conclude that the stratification dissolved in the autumn. The study found that the water temperature distribution in Xiangjiaba Reservoir is affected by the inflow temperature, meteorological elements, and intake elevation. The inflow temperature only affects the size of the water temperature in the Xiangjiaba Reservoir. However, the influence of the elevation and discharge ways on the vertical water structure in front of the dam was more notable. Meteorological elements control the surface water temperature within 10 m.

The lagging heating process was visible in spring after the impoundment of Xiluodu Reservoir. The water temperature

lowering process was relatively smooth in the fall and winter. The daily variable amplitude of the water temperature was reduced daily. The inflow temperature is nearly identical to the water temperature in front of the dam. The velocity in the Xiluodu Reservoir has grown greatly due to the increased inflow. The seasonal stratification of vertical water temperature in the Xiluodu Reservoir was noticeable. It could be divided into epilimnion, thermocline, and hypolimnion. The epilimnion depth increased, the thermocline thickness reduced, and the water temperature stratification structure strengthened as input and water temperature increased in spring. Due to the deep hole spillway, the thermocline moves slowly down during the flood season. The hypolimnion range shrank gradually, while the water temperature

remained stable at 14–15°. The inflow temperature had little influence on the vertical water structure in the front of the dam; the intake elevation significantly affected the thermocline depth in the Xiluodu Reservoir.

In Xiangjiaba Reservoir, located at the end of the cascade reservoirs in the Jinsha River, affected by the impoundment of Xiluodu Reservoir, the water temperature change process was lagged, and heating and cooling processes were smoother. The congestion time and the accumulative effect of water temperature on space were more significant. The impact of cascade reservoirs on downstream river water temperature processes can be summarized in two ways: the first is the homogenization effect, which refers to the amplitude of annual variation in water temperature decrease, and the second is the lagging effect, which refers to the apparent delay in water temperature change.

According to the cumulative impacts of the water temperature in the Xiangjiaba Reservoir and Xiluodu Reservoir, the temperature of the discharged water in the Xiangjiaba Reservoir was below the lower limit of demand from March to May. The control measures are as follows: in Xiangjiaba Reservoir, the left power station will be tested to operate from March to May. The right power station will be tested to operate from August to February. In Xiluodu Reservoir, the stop-log gate will be enabled in March. Then, the level deterioration should be started from January, as far as possible, down to 540 m before May 1.

## Dataset

In the present study, we examine the variation of water temperature in the upper and middle streams of the Yangtze River at Cuntan from 2010 to 2015. However, two scenarios were deeply analyzed: the pre-impact and post-impact. The period of 2010–2012 was considered pre-impact, while 2013–2015 was considered post-impact for Cuntan station. The descriptive statistics of the data selected for the two

scenarios are reported in Table 2. The statistical summary for the two scenarios during the training and testing period is given in Table 2, and the inter-co-relation among input variables is shown in Tables 3 and 4, respectively.

## Machine learning models

### Random Subspace (RSS)

The RSS generates several representations that can create a wide diversity of decision agents (Li et al. 2011; Pham et al. 2018). RSS, like bagging, modifies the training set; more precisely, the change is made for the future and not, for example, space. For a given  $p$ -dimensional vector ( $Z_j$ ) from the calibration dataset, i.e., ( $z_{j1}, z_{j2}, \dots, z_{jp}$ ), a ( $P$ ) features were randomly chosen. Hence, a Random Subspace of the first  $p$ -dimensional vector is presented due to this subspace selection. A new calibration dataset is designated as ( $Z = z_1, z_2, z_3, \dots, z_n$ ) of the initially  $p$ -dimensional training instances. Consequently, first base-level classifiers are constructed, and a voting mechanism is used to get a final prediction.

This technique is adopted to boost the accuracy achieved using poor classifiers performance (Plumpton et al. 2012). Following that, the RSS introduces randomness into the issue formulation by selecting certain variables to be substituted at random (Li et al. 2011). The RSS algorithm is a robust ensemble with several different classifiers (Plumpton et al. 2012). Integrating these weak classifiers becomes a robust model (Al-rimy et al. 2019).

Furthermore, stochastic discrimination theory is similar to the bagging method in that randomly selecting for the presented calibration dataset was adopted (Garca-Pedrajas and Ortiz-Boyer 2008); nevertheless, the RSS is selected using the fixed method calibration subset of attributes (Hong et al. 2017).  $M$  patterns were randomly chosen when building an RSS model to several aggregate classifiers for cataloging.

**Table 2** Statistics of measured daily water temperature at study stations

Statistics parameters	Pre-impact		Post-impact	
	Training period	Testing period	Training period	Testing period
Mean	18.10	20.02	18.38	20.93
Standard error	0.19	0.23	0.178	0.20
Median	18.00	21.30	19.00	22.00
Mode	11.50	23.10	11.90	23.20
Standard deviation	5.44	3.85	5.01	3.45
Sample variance	29.58	14.82	25.05	11.91
Kurtosis	−1.44	−0.87	−1.43	−0.80
Skewness	−0.03	−0.55	−0.06	−0.56
Minimum	9.10	11.50	10.10	13.10
Maximum	28.00	26.20	27.10	26.30
Count	800	297	800	297

They had L size without the need for any replacement. Each candidate example combines several single subsets representing an R subspace. Subsequently, a classifier is then calibrated using a sole subset of the all training set (Pham et al. 2018). The parameters selected for modeling pre- and post-impact in the RSS algorithm are presented in Table 5.

### Reduced Error Pruning Tree (REPTree)

The REPTree is one of the *ensembles* learning algorithms. It is used for building a decision tree (DT) model using an ensemble of dataset by decreasing the variance.

The information can be obtained using a splitting criterion, and decreasing the error pruning is the critical goal of the training process. Based on the division of the available instance, the REPTree can successfully handle missing data. For building a REPTree model, four pieces of information are necessary to be provided: (i) for each leaf of the threes, the minimum number of instances should be provided, (ii) the maximal value of the tree depth, (iii) for the split, the minimum ratio of the training set, and (iv) how many numbers of folds should be provided for better pruning (Srinivasan and Mekala 2014; Witten et al. 2016).

**Table 3** Correlation matrix and multicollinearity statistics analysis result from pre-impact intercomparison input combination (variables) characteristics

Correlation matrix									
	X <sub>(t-1)</sub>	X <sub>(t-2)</sub>	X <sub>(t-3)</sub>	X <sub>(t-4)</sub>	X <sub>(t-5)</sub>	X <sub>(t-6)</sub>	X <sub>(t-7)</sub>	X <sub>(t-8)</sub>	Pre-Impact
X <sub>(t-1)</sub>	1	0.997	0.994	0.990	0.986	0.982	0.978	0.973	0.997
X <sub>(t-2)</sub>	0.997	1	0.997	0.994	0.990	0.986	0.982	0.978	0.994
X <sub>(t-3)</sub>	0.994	0.997	1	0.997	0.994	0.990	0.986	0.982	0.990
X <sub>(t-4)</sub>	0.990	0.994	0.997	1	0.997	0.994	0.990	0.986	0.986
X <sub>(t-5)</sub>	0.986	0.990	0.994	0.997	1	0.997	0.994	0.990	0.982
X <sub>(t-6)</sub>	0.982	0.986	0.990	0.994	0.997	1	0.997	0.994	0.978
X <sub>(t-7)</sub>	0.978	0.982	0.986	0.990	0.994	0.997	1	0.997	0.973
X <sub>(t-8)</sub>	0.973	0.978	0.982	0.986	0.990	0.994	0.997	1	0.968
Pre-Impact	0.997	0.994	0.990	0.986	0.982	0.978	0.973	0.968	1

**Table 4** Correlation matrix and multicollinearity statistics analysis result from post-impact intercomparison input combination (variables) characteristics

Correlation matrix									
	X <sub>(t-1)</sub>	X <sub>(t-2)</sub>	X <sub>(t-3)</sub>	X <sub>(t-4)</sub>	X <sub>(t-5)</sub>	X <sub>(t-6)</sub>	X <sub>(t-7)</sub>	X <sub>(t-8)</sub>	Post-Impact
X <sub>(t-1)</sub>	1	0.996	0.992	0.989	0.985	0.982	0.978	0.974	0.996
X <sub>(t-2)</sub>	0.996	1	0.996	0.992	0.989	0.985	0.982	0.978	0.992
X <sub>(t-3)</sub>	0.992	0.996	1	0.996	0.992	0.989	0.985	0.982	0.989
X <sub>(t-4)</sub>	0.989	0.992	0.996	1	0.996	0.992	0.989	0.985	0.985
X <sub>(t-5)</sub>	0.985	0.989	0.992	0.996	1	0.996	0.992	0.989	0.982
X <sub>(t-6)</sub>	0.982	0.985	0.989	0.992	0.996	1	0.996	0.992	0.978
X <sub>(t-7)</sub>	0.978	0.982	0.985	0.989	0.992	0.996	1	0.996	0.974
X <sub>(t-8)</sub>	0.974	0.978	0.982	0.985	0.989	0.992	0.996	1	0.971
Post-Impact	0.996	0.992	0.989	0.985	0.982	0.978	0.974	0.971	1

**Table 5** The machine learning algorithm parameters are used for pre- and post-dam construction water temperature modeling

Model name	Description of parameters
Random Subspace (RSS)	Batch size = 100, classifier = REPTree, random seed = 1, subspace size = 0.5, numbers of executions slots = 1, number of iterations = 10
Reduced Error Pruning Tree (REPTree)	Batch size = 100, initial count = 0, number of folds = 3, random seed = 1, minimum proportion of the variance = 0.001, minimum number = 2, max depth = 1
Random Forest (RF)	Batch size = 100, bag size percent = 100, max depth = 0, numbers of executions slots = 1, number of iterations = 100, random seed = 1
M5 Pruned (M5P)	Batch size = 100, minimum number of instances = 4

It employs regression tree logic to generate iteratively after that successfully; it only chooses one which is considered the best (Rajesh and Karthikeyan 2017). Several authors used the REPTree model to predict air pollution concentration (Oprea et al. 2016; Vitkar 2017): Furthermore, the REPTree employs the validation dataset to accurately anticipate generalization errors (Nhu et al. 2020; Pham et al. 2021). From a computational point of view, backward overfitting is the first and sole responsibility of the pruning process achieved using the REPTree model. The essential benefit of the REPT technique is that it reduces the model complexities, escapes the over-fitting during the learning phase, and maintains accuracy (Khosravi et al. 2018). The parameters selected for modeling pre- and post-impact in the REPTree algorithm are presented in Table 5.

### Random Forest (RF)

Random Forest (RF) is a strong artificial intelligence technique developed by Breiman (2001) for measuring the considerable level of predictive parameters and producing accurate results without any of the overfitting fitting issues (Devasena 2014). It is a classifier composed of a collection of classification trees mainly related to the variables. Every tree produces a unique class, and all classes are then aggregated. The overwhelming vote predicts the outcomes (Pavey et al. 2017). It is used in classification and regression situations. The algorithm can be used for learning a complicated large dataset.

In contrast, a forest grows from numerous regression trees, putting them together and building an ensemble (Breiman 2001). Equal bias values characterize all trees; however, variances minimization can be achieved by lowering the link between the coefficients (Hastie et al. 2009). The results are numerical values, and the training sample is expected to be statistically independent.

The main advantages of the RF technique can be summarized as follows: (i) high generalization capacity, (ii) slightly sensitive to the attribute values, and (iii) can be easily calibrated using cross-validation. The ability of the R.F. methodology in simulating long-term monthly air temperature was studied, and its accuracy was examined by Mohsenzadeh Karimi et al. (2020); application examples showed advantageous characteristics of the R.F., which has higher accuracy. Several other researchers favor RF model machine learning techniques to judge relevance, namely, for climatological, hydrological, and environmental studies (Rahman and Islam 2019; Salam et al. 2021; Saha et al. 2020). Islam et al. (2020) employed the RF model to investigate whether variables impact COVID-19 mortality in Bangladesh cities. The architecture and parameters selected for modeling pre- and post-impact in the RF algorithm are presented in Table 5.

### M5 Pruned (M5P)

The model trees were developed by Quinlan (1992). The M5P is the most well-known reported algorithm for regression problems among the developed model's trees. Linear functions are used instead of discrete class labels at the leaves; M5P predicts that functional reliance is not constant across the domain but could be considered in smaller sub-domains (Demir 2022). M5P is an upgraded model of the M5 technique. Its major feature is efficiently handling large datasets with high dimensionality. If the training set is limited, the classification error rate may be large compared to the number of classes. The M5P method does not require parameter configuration. As a result, this algorithm does not require knowledge discovery. The M5P model can also be used in hydrology to model the stage-discharge connection (Ajmera and Goyal 2012), long-term streamflow forecasting (Yaseen et al. 2016), lake level forecasting (Demir 2022) and simulate the rainfall-runoff process (Solomatine and Xue 2004).

It is quick, straightforward, and accurate throughout the procedure. M5P uses a multivariate linear regression model to generate classification and regression trees. As a result, it can reduce variation within a specific subspace. These model trees are reminiscent of piecewise linear functions. The M5P algorithm is named the robust algorithm when dealing with missing data. The parameters selected for modeling pre- and post-impact in the M5P algorithm are presented in Table 5.

### Statistical performance assessment

There are numerous applications for performance evaluation in the real world. When a consumer wants to buy a computer, for example, he must compare costs, CPU speed, RAM, pre-installed software, and other factors among several options before deciding which one to purchase. We may ask which search engine will return the most relevant information for the given searches when retrieving information on the Internet. In performance evaluation, hypotheses are selected or ranked based on performance comparison of hypotheses on sample data (Leighton and Srivastava 1999). Hypotheses' performance measurements are numerical numbers that must be derived from sample data and may contain noise. Furthermore, in real-world applications, evaluating all hypotheses is typically impractical or impossible due to time and resource restrictions. As a result, statistical measures are utilized to efficiently evaluate the performance of hypotheses using a small quantity of sample data. There are a variety of statistical metrics available, and their conclusions are dependent on a number of criteria, including the size of the sample data and the distribution of hypotheses performance measurements. It's difficult to choose the best acceptable statistical measurements.



Models evaluation and comparison of actual and forecasted data of water temperature were achieved based on several performances metrics, namely, (i) Pearson correlation coefficient (PCC), (ii) the mean absolute error (MAE), (iii) the root mean square error (RMSE), (iv) the relative absolute error (RAE), (v) the coefficient of determination  $R^2$ , and (vi) the root-relative square error (RRSE), calculated as follows (Shukla et al. 2021; Vishwakarma et al. 2022):

$$PCC = \frac{\sum_{i=1}^N \left( (Tw)_{(Obs)_i} - (\overline{Tw})_{(Obs)_i} \right) \left( (Tw)_{(Est)_i} - (\overline{Tw})_{(Est)_i} \right)}{\sqrt{\sum_{i=1}^N \left( \left( (Tw)_{(Obs)_i} - (\overline{Tw})_{(Obs)_i} \right) \right)^2 \left( \left( (Tw)_{(Est)_i} - (\overline{Tw})_{(Est)_i} \right) \right)^2}} \quad (6)$$

$$R^2 = \left[ \frac{\sum_{i=1}^N \left( (Tw)_{(Obs)_i} - (\overline{Tw})_{(Obs)_i} \right) \left( (Tw)_{(Est)_i} - (\overline{Tw})_{(Est)_i} \right)}{\sqrt{\sum_{i=1}^N \left( \left( (Tw)_{(Obs)_i} - (\overline{Tw})_{(Obs)_i} \right) \right)^2 \left( \left( (Tw)_{(Est)_i} - (\overline{Tw})_{(Est)_i} \right) \right)^2}} \right]^2 \quad (7)$$

$$RMSE = \sqrt{\sum_{i=1}^N \frac{1}{N} \left[ \left( (Tw)_{(Obs)_i} - (\overline{Tw})_{(Est)_i} \right) \right]^2} \quad (8)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N \left| \left( (Tw)_{(Obs)_i} - (\overline{Tw})_{(Obs)_i} \right) \right| \quad (9)$$

$$RAE = \left| \frac{(Tw)_{(Obs)_i} - (\overline{Tw})_{(Est)_i}}{(\overline{Tw})_{(Est)_i}} \right| \times 100 \quad (10)$$

$$RRSE = \frac{\sqrt{\sum_{i=1}^N \left[ \left( (Tw)_{(Obs)_i} - (\overline{Tw})_{(Est)_i} \right) \right]^2}}{\sqrt{\sum_{i=1}^N \left[ \left( (Tw)_{(Obs)_i} - (\overline{Tw})_{(Obs)_i} \right) \right]^2}} \quad (11)$$

where  $(Tw)_{(Obs)_i}$  is the  $i^{th}$  measured/actual values,  $(\overline{Tw})_{(Obs)_i}$  is the average of the observed/actual values,  $(\overline{Tw})_{(Est)_i}$  is the  $i^{th}$  calculated values,  $(\overline{Tw})_{(Est)_i}$  is the average of the estimated/predicted values, and  $N$  is the total number of observations. The values of RMSE range from 0 to  $\infty$ , PCC  $-1$  to  $1$ ,  $R^2$  0 to 1, MAE 0 to  $\infty$ , and RAE and RRSE 0 to 1. Good forecasting accuracy corresponds to a value of PCC and  $R^2$  nearly equal to 1, while for the other metrics, their values should be close to zero (Yaseen et al. 2016, 2018; Ayele et al. 2017; Shukla et al. 2021; Demir 2022; Pham et al. 2022; Vishwakarma et al. 2022). Each one of these measures’ descriptive performances is as follows:

- The lower the value of MAE and RMSE and near to zero MBE, the better the model performance (Vishwakarma et al. 2022).
- For  $R^2$ :  
 Very good ( $0.7 < R^2 \leq 1$ ); good ( $0.6 < R^2 \leq 0.7$ ); satisfactory ( $0.5 < R^2 \leq 0.6$ ); and unsatisfactory ( $R^2 \leq 0.5$ ) (Ayele et al. 2017).

In RAE, total absolute error is normalized by dividing it by the total absolute error of the basic indicator in the RAE, whereas in RRSE, the total squared error is normalized by dividing it by the total squared error of the basic indicator in the RSE. The error is reduced to the same dimensions as the quantity being predicted by taking the square root of the relative squared error. Taylor diagrams, radar charts, and box plots were also investigated to visually compare model performance (Taylor 2001; Citakoglu 2021; Başakın et al. 2022; Görkemli et al. 2022). More details about models evaluation and comparison can be found in Kushwaha et al. (2021), Elbeltagi et al. (2022a), and Vishwakarma et al. (2022).

**Table 6** The summary of best subset regression variables in (pre-impact)

S. No.	Number of variables and their combination	MSE	$R^2$	Adjusted $R^2$	Mallows’ Cp	Akaike’s IC	Schwarz’s BC	Amemiya’s PC
1	$X_{(t-1)}$	0.148	0.994	0.994	22.181	-2093.580	-2083.582	0.006
2	$X_{(t-1)}/X_{(t-3)}$	0.145	0.994	0.994	4.307	-2111.307	-2096.309	0.006
<b>3</b>	<b><math>X_{(t-1)}/X_{(t-3)}/X_{(t-7)}</math></b>	<b>0.145</b>	<b>0.994</b>	<b>0.994</b>	<b>4.311</b>	<b>-2111.308</b>	<b>-2091.310</b>	<b>0.006</b>
4	$X_{(t-1)}/X_{(t-3)}/X_{(t-6)}/X_{(t-7)}$	0.145	0.995	0.994	1.764	-2113.880	-2088.883	0.006
5	$X_{(t-1)}/X_{(t-3)}/X_{(t-4)}/X_{(t-6)}/X_{(t-7)}$	0.145	0.995	0.994	3.251	-2112.397	-2082.400	0.006
6	$X_{(t-1)}/X_{(t-2)}/X_{(t-3)}/X_{(t-4)}/X_{(t-6)}/X_{(t-7)}$	0.145	0.995	0.994	5.157	-2110.491	-2075.495	0.006
7	$X_{(t-1)}/X_{(t-2)}/X_{(t-3)}/X_{(t-4)}/X_{(t-5)}/X_{(t-6)}/X_{(t-7)}$	0.145	0.995	0.994	7.070	-2108.579	-2068.584	0.006
8	$X_{(t-1)}/X_{(t-2)}/X_{(t-3)}/X_{(t-4)}/X_{(t-5)}/X_{(t-6)}/X_{(t-7)}/X_{(t-8)}$	0.145	0.995	0.994	9.000	-2106.650	-2061.655	0.006

The best model for the selected selection criterion is displayed in bold.

**Table 7** The summary of best subset regression variables in (post-impact)

S. No.	Number of variables and their combination	MSE	R <sup>2</sup>	Adjusted R <sup>2</sup>	Mallows' Cp	Akaike's IC	Schwarz's BC	Amemiya's PC
1	X <sub>(t-1)</sub>	0.193	0.992	0.992	4.975	-1803.824	-1793.825	0.008
2	X <sub>(t-1)</sub> /X <sub>(t-2)</sub>	0.191	0.992	0.992	-0.941	-1809.761	-1794.763	0.008
3	X <sub>(t-1)</sub> /X <sub>(t-2)</sub> /X <sub>(t-4)</sub>	0.191	0.992	0.992	0.648	-1808.175	-1788.177	0.008
4	X <sub>(t-1)</sub> /X <sub>(t-2)</sub> /X <sub>(t-4)</sub> /X <sub>(t-6)</sub>	0.191	0.992	0.992	1.162	-1807.672	-1782.675	0.008
5	X <sub>(t-1)</sub> /X <sub>(t-2)</sub> /X <sub>(t-3)</sub> /X <sub>(t-4)</sub> /X <sub>(t-6)</sub>	0.191	0.992	0.992	3.032	-1805.803	-1775.807	0.008
6	X <sub>(t-1)</sub> /X <sub>(t-2)</sub> /X <sub>(t-3)</sub> /X <sub>(t-4)</sub> /X <sub>(t-6)</sub> /X <sub>(t-7)</sub>	0.192	0.992	0.992	5.027	-1803.808	-1768.812	0.008
<b>7</b>	<b>X<sub>(t-1)</sub>/X<sub>(t-2)</sub>/X<sub>(t-3)</sub>/X<sub>(t-4)</sub>/X<sub>(t-6)</sub>/X<sub>(t-7)</sub>/X<sub>(t-8)</sub></b>	<b>0.192</b>	<b>0.992</b>	<b>0.992</b>	<b>7.002</b>	<b>-1801.833</b>	<b>-1761.838</b>	<b>0.008</b>
8	X <sub>(t-1)</sub> /X <sub>(t-2)</sub> /X <sub>(t-3)</sub> /X <sub>(t-4)</sub> /X <sub>(t-5)</sub> /X <sub>(t-6)</sub> /X <sub>(t-7)</sub> /X <sub>(t-8)</sub>	0.192	0.992	0.992	9.000	-1799.835	-1754.841	0.008

The best model for the selected selection criterion is displayed in bold.

**Table 8** Standardized coefficients and sensitivity analysis of linear regression of different input combinations in pre-impact

Source	Standardized coefficients				Sensitivity analysis	
	Value	Standard error	t	Pr >  t	Lower bound (95%)	Upper bound (95%)
X <sub>(t-1)</sub>	1.084	0.020	53.012	<0.0001	1.044	1.124
X <sub>(t-2)</sub>	0.000	0.000	-	-	-	-
X <sub>(t-3)</sub>	-0.068	0.026	-2.658	0.008	-0.119	-0.018
X <sub>(t-4)</sub>	0.000	0.000	-	-	-	-
X <sub>(t-5)</sub>	0.000	0.000	-	-	-	-
X <sub>(t-6)</sub>	0.000	0.000	-	-	-	-
X <sub>(t-7)</sub>	-0.019	0.013	-1.413	0.158	-0.046	0.007
X <sub>(t-8)</sub>	0.000	0.000	-	-	-	-

**Table 9** Standardized coefficients and sensitivity analysis of linear regression of different input combinations post-impact

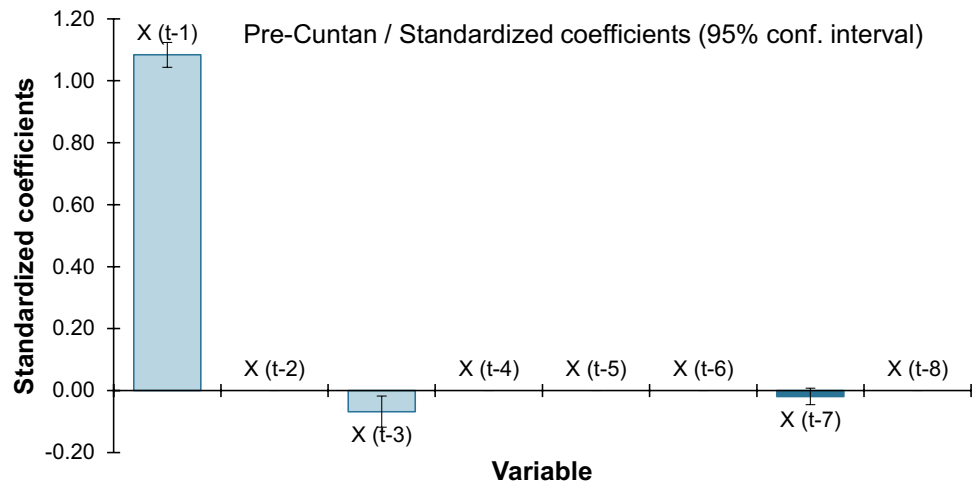
Source	Standardized coefficients				Sensitivity analysis	
	Value	Standard error	t	Pr >  t	Lower bound (95%)	Upper bound (95%)
X <sub>(t-1)</sub>	0.911	0.030	30.054	< 0.0001	0.851	0.970
X <sub>(t-2)</sub>	0.080	0.041	1.959	0.050	0.000	0.161
X <sub>(t-3)</sub>	-0.015	0.041	-0.357	0.721	-0.095	0.066
X <sub>(t-4)</sub>	0.047	0.037	1.285	0.199	-0.025	0.119
X <sub>(t-5)</sub>	0.000	0.000	-	-	-	-
X <sub>(t-6)</sub>	-0.025	0.037	-0.695	0.487	-0.097	0.046
X <sub>(t-7)</sub>	-0.006	0.041	-0.157	0.875	-0.087	0.074
X <sub>(t-8)</sub>	0.005	0.030	0.156	0.876	-0.055	0.064

## Result and discussion

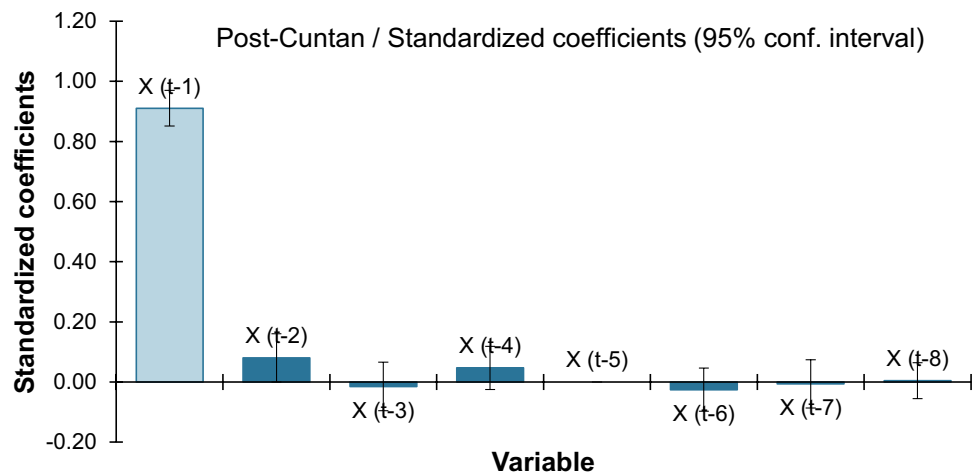
Temperature is one of the most significant parameters for evaluating the water environment since water temperature fluctuation mainly governs several freshwater

processes. The reservoir impoundment is responsible for water temperature fluctuation and distribution, as is the annual water temperature change in the downstream river. At the same time, because the reservoir space architecture is relatively dense in the cascade development mode, a single reservoir's influence on water temperature is bound

**Fig. 3** The standardized coefficients of input variable for sensitivity analysis (pre-impact)



**Fig. 4** The standardized coefficients of input variable for sensitivity analysis (post-impact)



**Table 10** Statistical measurements of the proposed methods to forecast water temperature in pre-impact spam

Scenario	Model	Performance indicators					
		R <sup>2</sup>	PCC	MAE	RMSE	RAE (%)	RRSE (%)
Training	Random Subspace	0.9952	0.9976	0.2683	0.3748	5.5112	6.8969
	REPTree	0.9952	0.9976	0.2648	0.3754	5.4393	6.9076
	Random Forest	0.9988	0.9994	0.1444	0.1862	2.9654	3.4253
	M5P	0.9952	0.9974	0.2566	0.3928	5.2701	7.2283
Testing	Random Subspace	0.9862	0.9931	0.3398	0.455	8.9076	10.592
	REPTree	0.9872	0.9936	0.3232	0.4365	8.4729	10.163
	Random Forest	0.9872	0.9936	0.3206	0.4356	8.404	10.1421
	M5P	0.9920	0.9960	0.2387	0.3449	6.2573	8.0288

to be in some form, resulting in a cumulative effect on water temperature. In this paper, two two-dimensional models modeled on two-dimensional averages were laterally averaged for Xiangjiaba Reservoir and Xiluodu Reservoir downstream of Jinsha River to simulate the water temperature. The model parameters were calibrated using 2014 temperature data and then confirmed using

2015 data. These models are capable and most suited to simulating the two libraries' hydrodynamic processes and geographical water temperature distributions.

Using simulation findings, the present research examined water temperature fluctuation over space and time in Xiangjiaba and Xiluodu reservoirs.

**Table 11** Statistical measurements of the proposed methods to forecast water temperature in post-impact spam

Scenario	Model	Performance indicators					
		$R^2$	PCC	MAE	RMSE	RAE (%)	RRSE (%)
Training	Random Subspace	0.9956	0.9978	0.2482	0.3350	5.5173	6.6966
	REPTree	0.9950	0.9975	0.2612	5.8081	5.8081	7.0346
	Random Forest	0.9988	0.9994	0.1327	0.1675	2.9498	3.3496
	M5P	0.9946	0.9973	0.2589	0.3662	5.7572	7.3211
Testing	Random Subspace	0.9704	0.9851	0.4212	0.5969	11.3469	13.9353
	REPTree	0.9661	0.9829	0.464	0.6442	12.5022	15.0396
	Random Forest	0.9704	0.9851	0.439	0.6006	11.8284	14.0229
	M5P	0.9708	0.9853	0.4285	0.4285	11.5439	13.8282

Furthermore, a cumulative effect evaluation method was built. The characteristics of cumulative effects of water temperature over space and time in the Xiangjiaba and Xiluodu reservoirs were evaluated. As a result, the downstream control approach for cumulative effects was developed.

### Input variables selection for modeling of pre- and post-impact on water temperature

The success of machine learning models is mainly governed by a good selection of the best predictors, i.e., the best input variables (Malik et al. 2019; Shukla et al. 2021; Kushwaha et al. 2021; Elbeltagi et al. 2022b, a; Kumar et al. 2022). From a general point of view, based on the available input variables, we believe that testing several input combinations is the more suitable procedure for obtaining the best final model; in addition, testing several input combinations can help provide a multitude of alternatives with different structures. As reported in Tables 6 (for pre-impact) and 7 (for post-impact), eight scenarios were analyzed in the present study having different input variables. The best input combinations are reported in bold. However, all combinations were selected based on several indices, namely, Amemiya's PC (A-PC), Schwarz's BC, Akaike's IC, the MSE, and Mallows' Cp (M-Cp), the  $R^2$ , and the adjusted  $R^2$  (A- $R^2$ ). According to Table 6, for the pre-impact scenario, it is clear that the best model corresponds to the third input combination using the first, second, and seventh lag times, i.e., ( $t-1$ ), ( $t-2$ ), and ( $t-7$ ), respectively, and exhibiting the most significant statistical indices with MSE,  $R^2$ , A- $R^2$ , M-Cp, AIC, SBC, and APC values of approximately 0.145, 0.994, 0.994, 4.311, -2111.308, -2091.310, and 0.006, respectively. Similarly, for the post-impact, as reported in Table 7, the best model was obtained when the input variables were selected as the first eight successive lag times, excluding the fifth lag time, i.e., ( $t-1$ ) to ( $t-4$ ) in addition to ( $t-6$ ) to ( $t-8$ ), for which the statistical MSE,  $R^2$ , A- $R^2$ , M-Cp, AIC, SBC, and APC values were approximately 0.192, 0.992, 0.992, 7.002, -1801.833, -1761.838, and 0.008, respectively.

### Sensitivity analysis

From the input variables selection reported above, it is clear that the variables' contribution varies from one to another, and the best input selection highly influenced the model's performance. Tables 8 and 9 and Figs. 3 and 4 depict the obtained standard coefficients of the linear regression (SC-LR). According to Table 8, for the pre-impact scenario, the input variables corresponding to the three lags times, i.e., ( $t-1$ ), ( $t-2$ ), and ( $t-7$ ), exhibited the highest absolute standard coefficients, i.e., 1.084, 0.068, and -0.019, respectively. Similarly, for the post-impact simulation, the values of the SC-LR were 0.911, 0.080, 0.015, 0.047, 0.025, 0.006 and 0.005, respectively (Table 9).

### Modeling of pre- and post-impact on water temperature

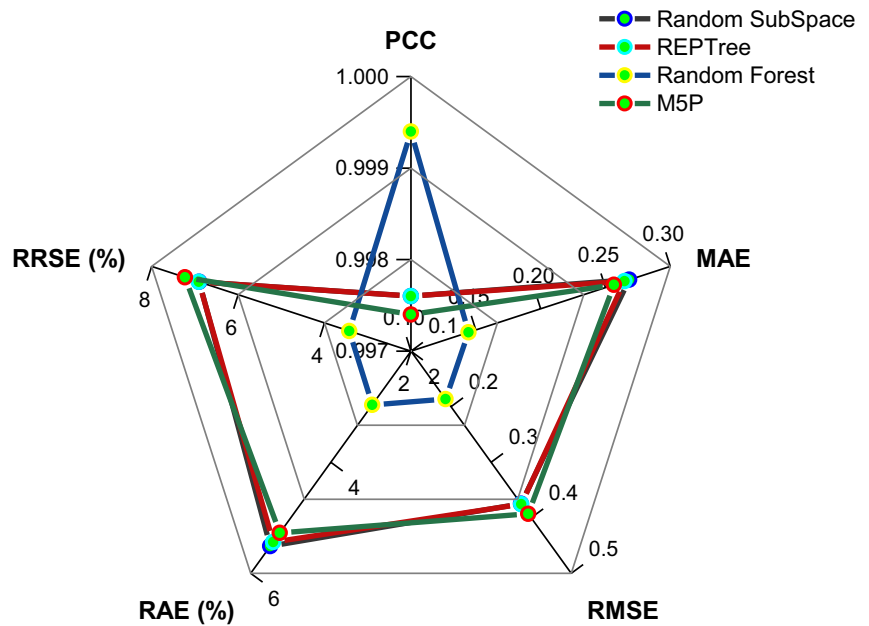
The hybrid models, i.e., RS, REPTree, RF, and M5P, were calibrated according to the best input variables selected based on the finding reported in Tables 6 and 7. Selection models were calibrated beyond the input variables using 75% of daily observed data and validated using the remaining 25%. Both goodness-of-fit measurements and graphical presentations were used to assess the models' performance. Tables 10 and 11 describe the overall performance of all AI-based models throughout the calibration and testing stages for the estimate of daily observed water temperature at all stations using five statistical indicators. In Figs. 5 and 6, the statistical measures are also shown using a radar chart.

### Evaluation developed models in pre-impact water temperature forecasting

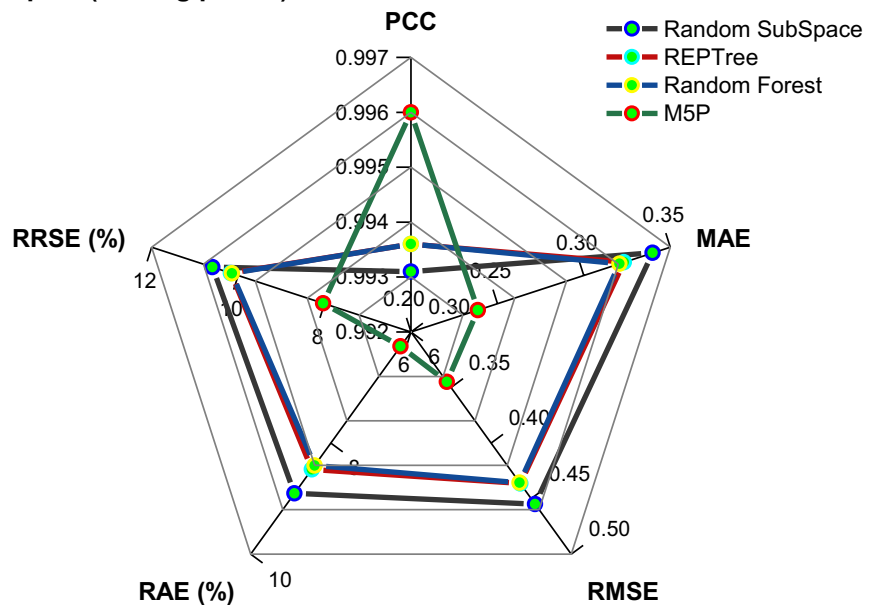
Using various assessment criteria, we examined the robustness of the proposed models during the calibration and testing stages (Table 10). In addition, all soft computing models use identical statistical techniques to train and evaluate datasets. It can be seen that overfitting does not occur in any of the models. The M5P model has the highest accuracy

**Fig. 5** Radar charts display the goodness-of-fit measures of Random Subspace, REPTree, Random Forest, and M5P models during **a** training and **b** testing period in pre-impact water temperature

**(a) Pre-impact (Training period)**



**(b) Pre-impact (Testing period)**



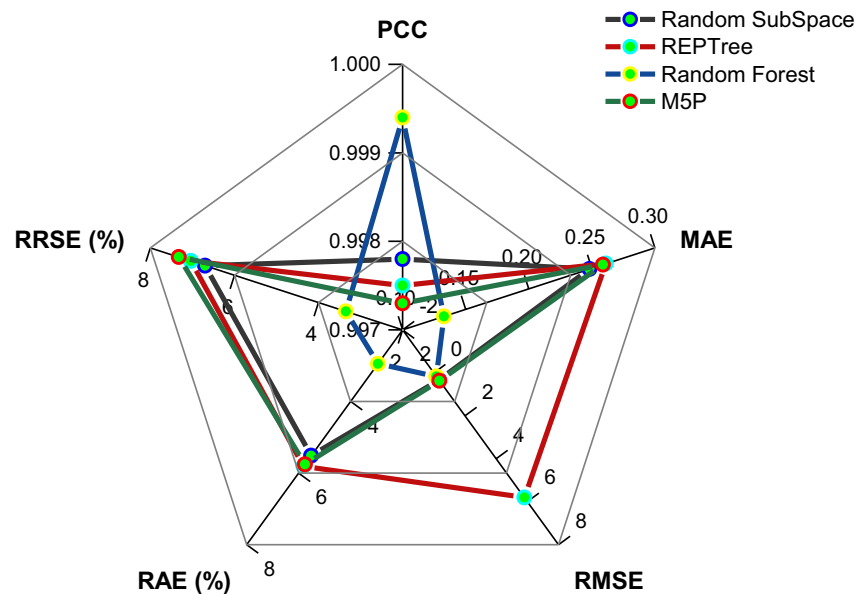
during the calibration and testing stages of training compared to the other suggested models, as shown in Table 10. Based on examining the numerical performances reported in Table 10, extremely strong prediction performance ( $R^2 > 0.9$ ) was achieved using all models. Our  $R^2$  result revealed highly reasonable model performances. However, the highest numerical performances were obtained using the M5P when performance measurements were taken into account, exhibiting the largest  $R^2$  value (0.9920); the RF (0.9872) came

second, REPTree (0.9872) came third, and RS (0.9862) was ranked fourth in the list of models during the validation stage. By referencing the RMSE values, it is clear that the M5P model obtained the poorest RMSE values corresponding to the highest predictive accuracy ( $RMSE \approx 0.3349$ ); the RF (0.4356) came second, REPTree (0.4365) came third, and RS (0.455) was ranked fourth in the list of models during the validation stage. Similarly, based on the MAE criteria, the M5P model ( $RMSE \approx 0.2384$ ) worked best, the RF

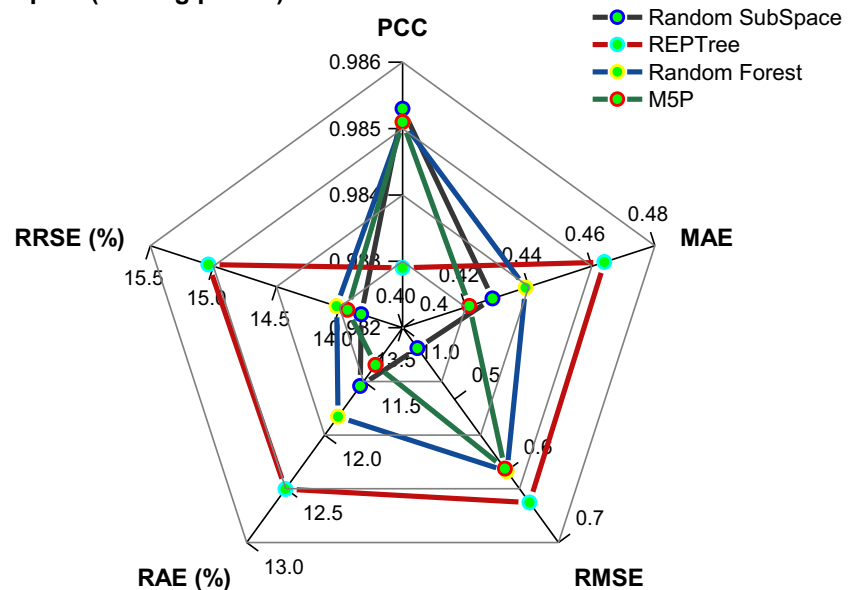


**Fig. 6** Radar charts display the goodness-of-fit measures of Random Subspace, REPTree, Random Forest, and M5P models during **a** training and **b** testing period in post-impact water temperature

**(a) Post-impact (Training period)**



**(b) Post-impact (Testing period)**



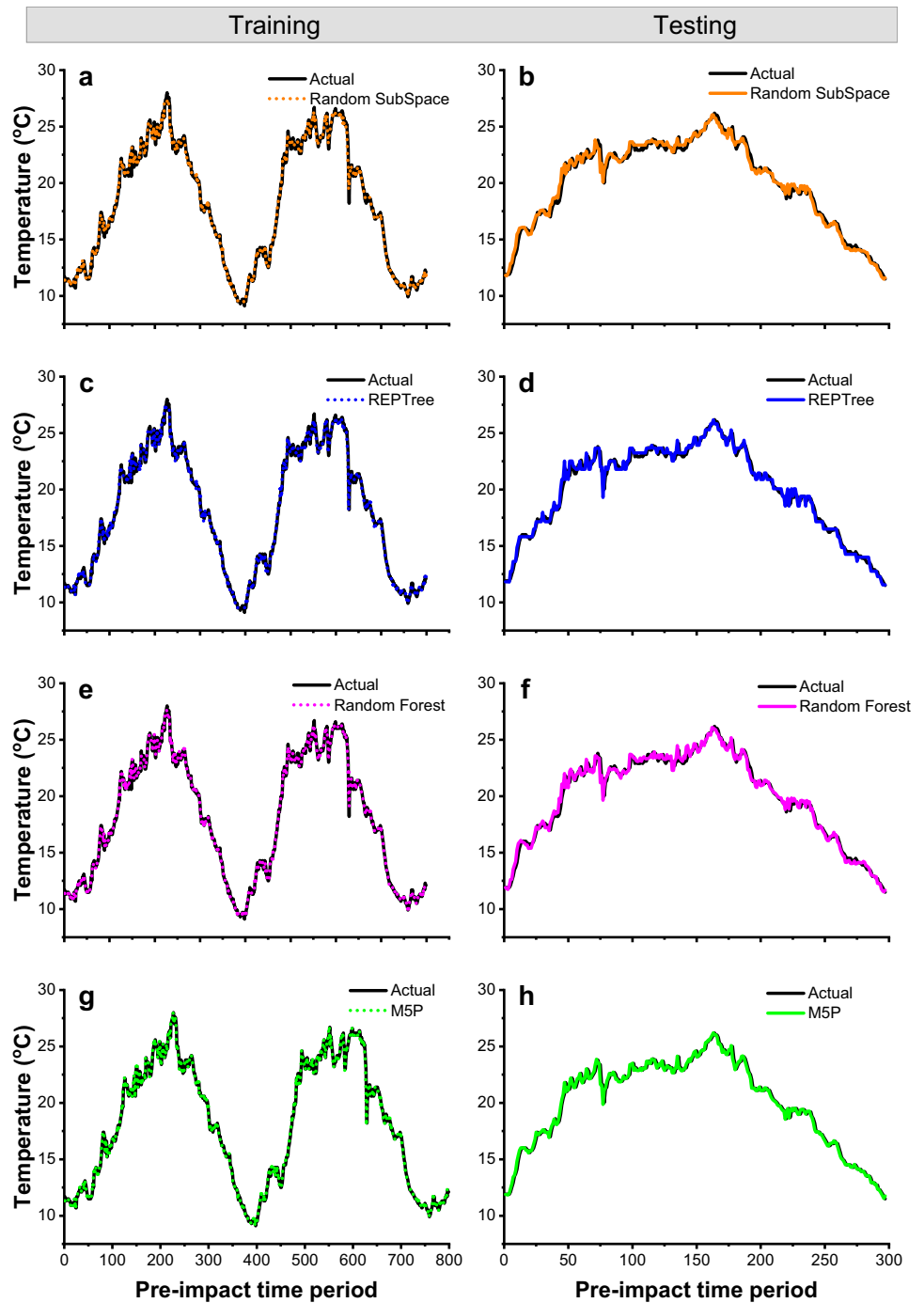
(0.3206) came second, REPTree (0.3232) came third, and RS (0.9938) was ranked fourth during the validation stage.

The M5P model produced the lowest MAE criteria (0.2384), followed by the Random Forest (0.3206), REPTree (0.3232), and Random Subspace (0.9938) model, using the RAE and RRSE statistical evaluation criteria which were least in M5P (6.2573 and 8.0288, respectively) and followed by the RF (8.404 and 10.1421, respectively), REPTree (8.4729 and 10.163, respectively), and RSS (8.9076 and 10.592, respectively) models. Statistical metrics are also presented using the radar graph in Fig. 5. All four

hybrid models performed excellently, but the M5P model worked largely better than the other models in estimating daily water temperature in all the six statistics at all study locations during pre-impact spam. The relative performance indicated that they performed similarly. The performance lines of all five models overlap on the radar map, showing that the models perform similarly to one another. However, a closer examination of the data indicated that the M5P largely exceeds the remaining models.

The scatterplot of the measured and estimated data of daily water temperature in the calibration and

**Fig. 7** Comparison of the results between the measured/ actual and predicted water temperature for the training and testing dataset in using **a–b** Random Subspace; **c–d** REPTree; **e–f** Random Forest; and **g–h** M5P for pre-impact time series

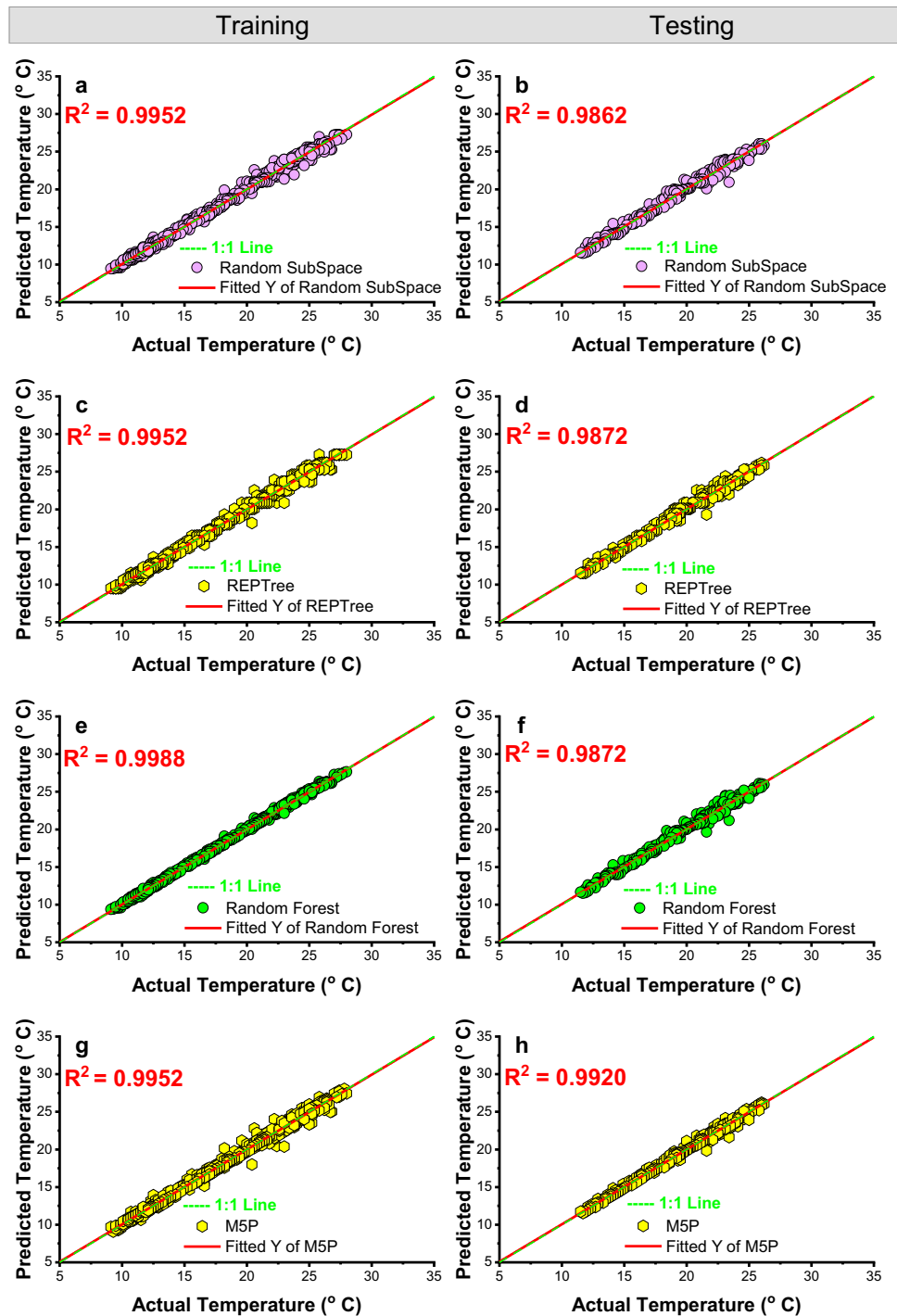


testing stages for all proposed models are depicted in Figs. 7 and 8, showing a good match during the two stages. All models with excellent levels guaranteed high predictive accuracy. At the same time, only M5P could perfectly predict the fluctuation of the water temperature of training and testing of pre-impact spam (Figs. 9 and 10).

We also further analyzed model efficiency using the Box and Whisker Plot of the models (Fig. 11a) and

Taylor diagrams (Fig. 12). The box and whisker plots for predicting the maximum and minimum data point using the M5P were approximately equal to the measured data. In contrast, RSS, REPTree, and RF slightly underestimated water temperature. The quartile, median, mean, and standard deviation of all models could closely predict water temperature values to the measured data having a significant predictive degree. Indeed, the M5P showed better accuracy.

**Fig. 8** Scatter plots of observed vs. predicted water temperature in the training and testing phase: **a–b** Random Sub-space; **c–d** REPTree; **e–f** Random Forest; and **g–h** M5P during the preimpact time

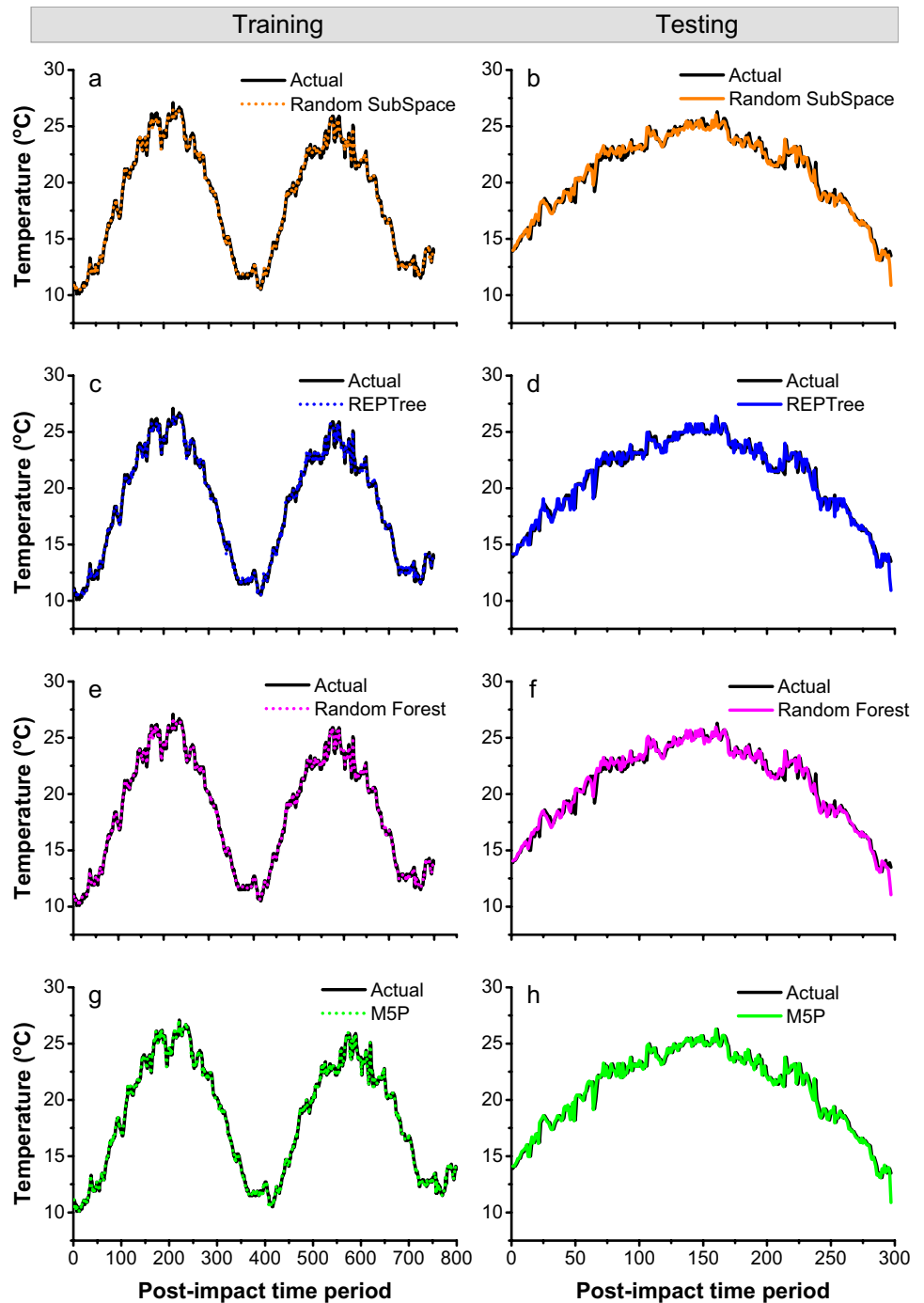


The better performance shown in Taylor diagrams (Fig. 12), the closer each produced model's point is to the observed position. The models had a strong predictive capacity in this case. However, the M5P approach provided the greatest R and poorest RMSE values. The SD of the M5P model was close to the actual SD-based values; however, the SD of the RS and RF models was lower, followed by the REPTree models.

### Evaluation developed models in post-impact water temperature forecasting

The model's performance during post-impact is summarized in Table 11 in terms of six statistical metrics. All four hybrid models performed significantly better in predicting water temperature in all six statistics in the post-impact phase than the baseline model. The comparison of the relative

**Fig. 9** Comparison of the results between the measured/ actual and predicted water temperature for the training and testing dataset in using: **a–b** Random Subspace; **c–d** REPTree; **e–f** Random Forest; and **g–h** M5P for post-impact time series

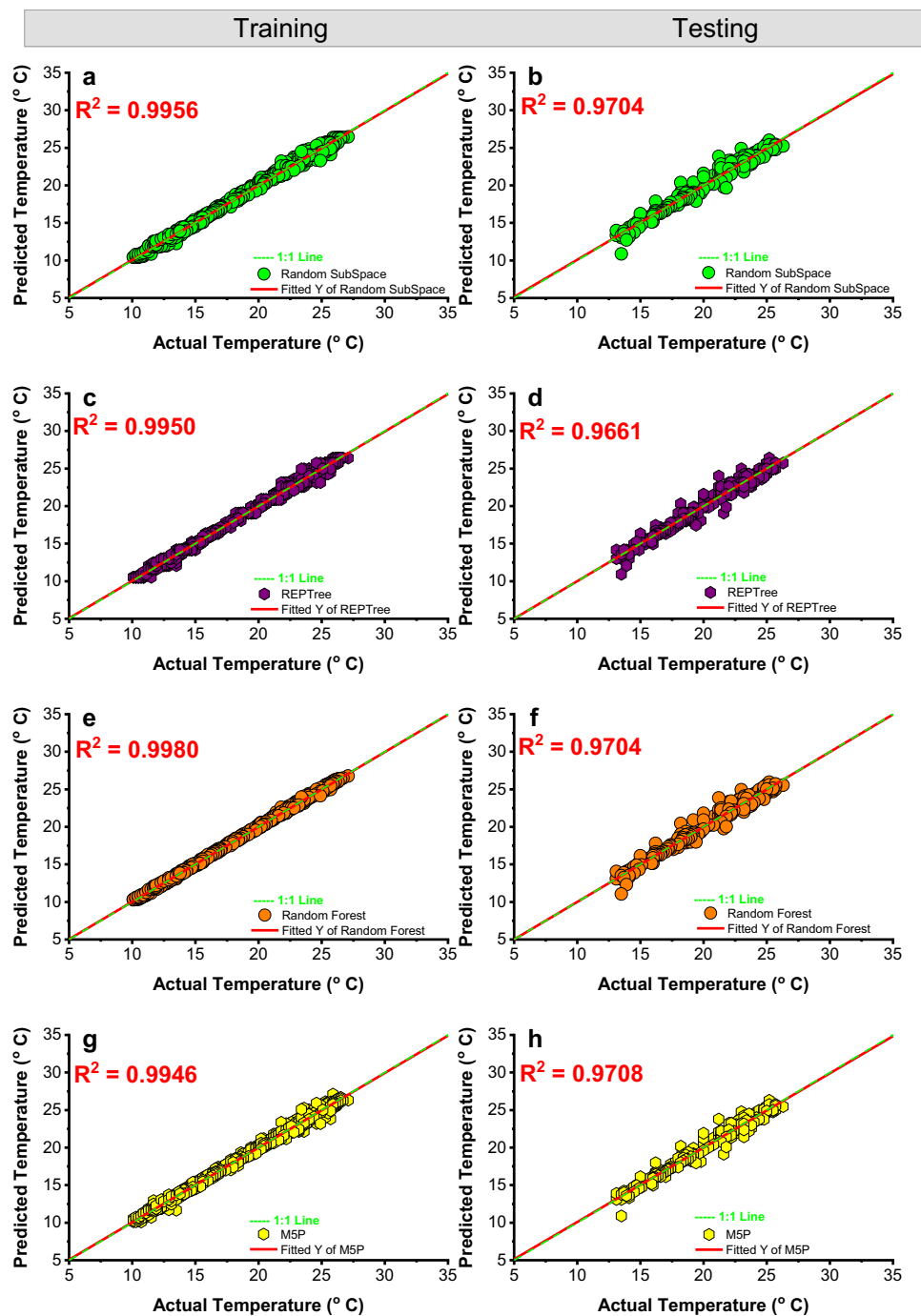


performances of the hybrid models found that they were quite close to the observed values. Our findings (Table 11) revealed that these models are acceptable and provide good results based on testing data. However, considering the  $R^2$  and PCC, the M5P was the most accurate and exhibited a value of approximately 0.9708 and 0.9853, followed by the RSS and RF, which are equal ( $R^2 = 9704$ ,  $PCC = 0.9851$ ) and REPTree (0.9661 and 0.9829). The MAE, RMSE, RAE, and RRSE were obtained as 0.4212, 0.5969, 11.3469, and

13.9353, respectively, for RSS; 0.439, 0.6006, 11.8284, and 14.0229 for RF; and 0.464, 0.6442, 12.5022, and 15.0396 for REPTree, respectively. The low MAE, RMSE, RAE, and RRSE and higher value or near-ideal  $R^2$  and PCC values designate a better model predictive performance. As indicated in Table 11, there is an excellent concert of the M5P model in estimating daily water temperature for post-impact.

As seen in Fig. 6, the statistical measures are also provided using a radar map. Excellent accuracies were achieved

**Fig. 10** Scatter plots of observed vs. predicted water temperature in the training and testing phase: **a–b** Random Subspace; **c–d** REPTree; **e–f** Random Forest; and **g–h** M5P during the post-impact time

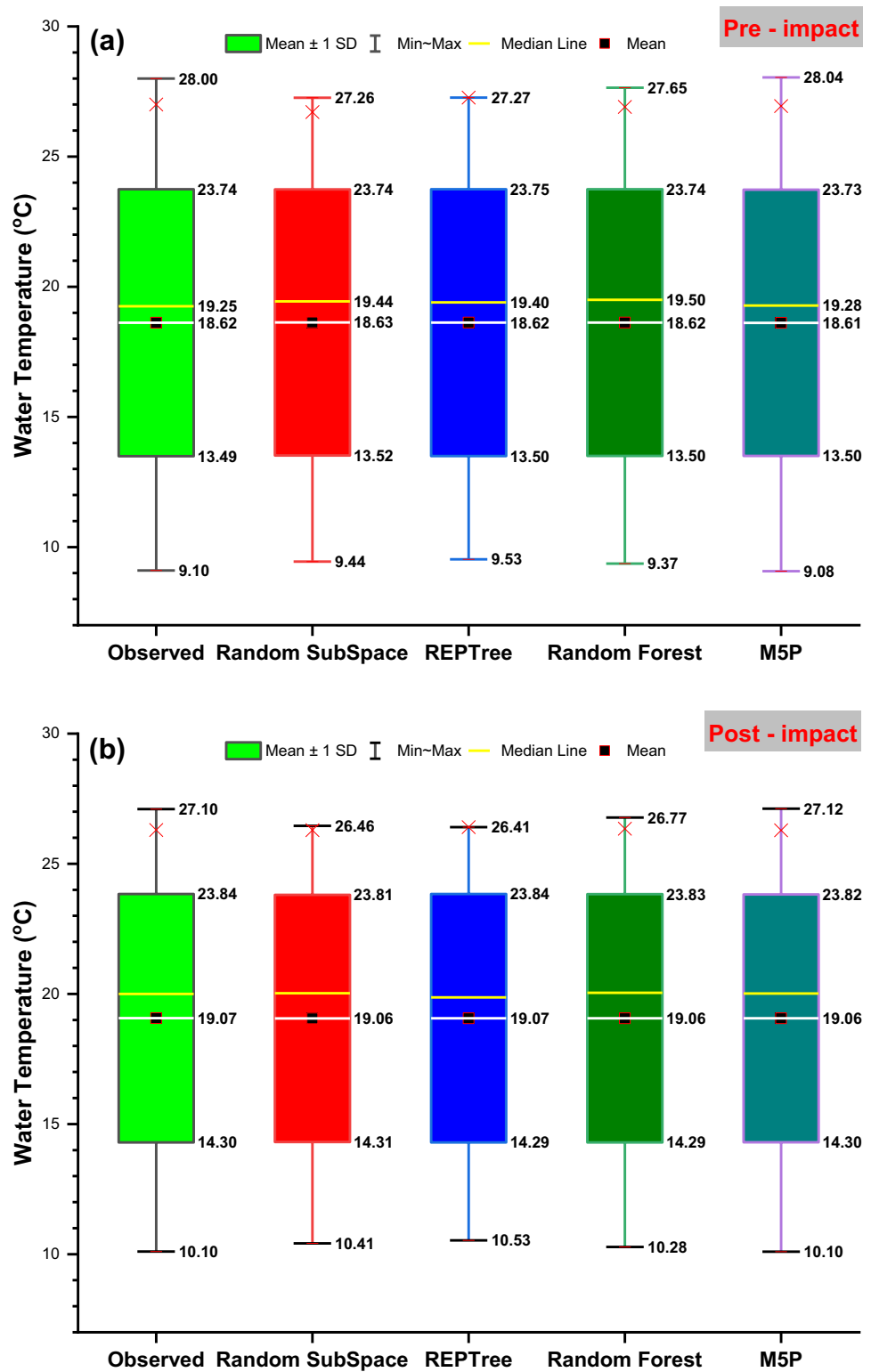


using the four proposed models. However, the superiority of the M5P model compared to the other models in projecting daily water temperature according to the six statistics metrics at all study sites during the pre-impact spam period is more obvious. The radar map demonstrates that the performance lines of all five models overlap, suggesting that the models perform similarly to one another in terms of overall performance. However, an in-depth examination of the data indicated slight superiority of the M5P model compared to the other.

Figures 9 and 10 indicate that the suggested soft computing algorithms predicted and observed values and scatter plots are consistent. This graph demonstrates that the proposed models can accurately predict water temperature. When employing the M5P model, the data points projected as measured versus predicted values were close, one on top of the other, indicating high fitting capabilities. We examine model efficiency using box and whisker plots (Fig. 11b) and Taylor diagrams (Fig. 13). Figure 11b shows the model's box and whisker plot results. Like the M5P model, the M5P box



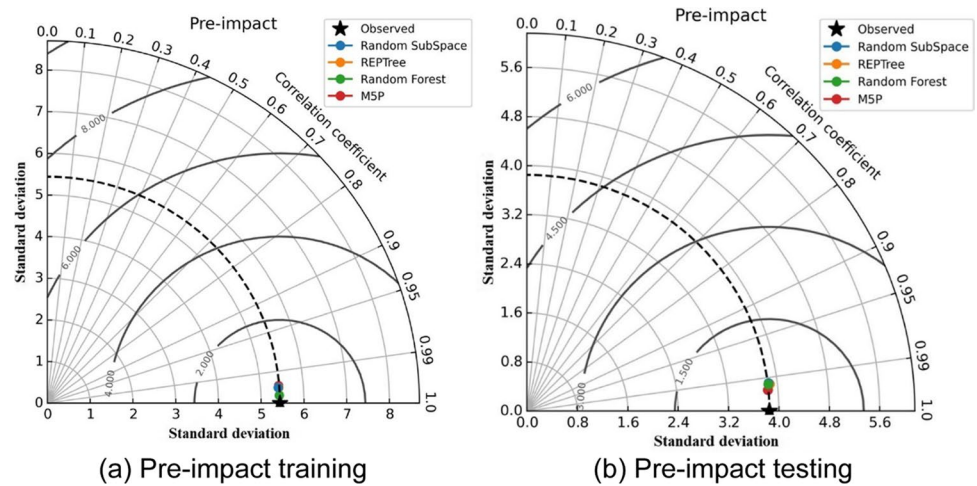
**Fig. 11** Box and Whisker plot of the models: **a** pre-impact and **b** post-impact



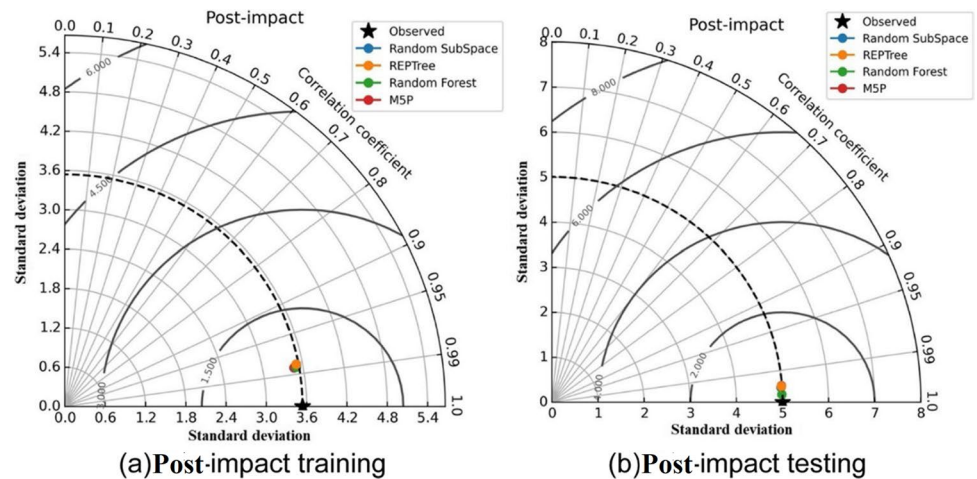
and whisker plot predicted maximum and minimum values very close to the actual values. However, Random Subspace, REPTree, and Random Forest slightly underestimated the water temperature. The M5P model outperformed the other quartile, median, mean, and standard deviation.

Figure 13 shows that, according to Taylor diagrams, the model should be considered better if it is near the observed point's position. Hence, it is clear that the M5P algorithm was the strong model in terms of forecasting capabilities and performances, which is reflected by its high PCC and lowest

**Fig. 12** Taylor diagram of the models during pre-impact spam. **a** Pre-impact training. **b** Pre-impact testing



**Fig. 13** Taylor diagram of the models during post-impact spam. **a** Post-impact training. **b** Post-impact testing



RMSE. In addition, from the Taylor diagrams, the M5P was also the sole model having an SD relatively equal to the measured data. However, the RT and RF models exhibited a small SD, while the REPTree model also had a small standard deviation.

## Discussion

Obtained results in the present study are very encouraging and promising. While the performances of all models for the pre-impact spam were more accurate compared to those of post-impact spam, in overall, numerical performances revealed the suitability of the proposed machine learning models as a robust tool for water temperature prediction. According to the obtained results and to what is discussed above, the mean PCC, RMSE, and MAE values were 0.994, 0.306°C, and 0.418°C for pre-impact spam and 0.985, 0.438, and 0.568 for post-impact spam, which are superior to the values reported by Heddami et al. (2020), i.e., 0.980, 1.413 °C and 1.085

°C, respectively, and it is clear that the superiority of the M5P, RSS, RF, and REPTree models was more obvious taking into account the error metrics, i.e., the RMSE and MAE values. In a recently published paper, Heddami et al. (2022) reported that river water temperature can be predicted with sufficient accuracy by hybrid machine learning combined with signal decomposition, and it was found that the high values of the PCC, RMSE, and MAE were 0.980, 1.304°C, and 1.018°C, respectively, which were significantly less than the obtained values in our present study. In another study, Yousefi and Toffolon (2022) compared between long short-term memory (LSTM), RF, ERT, *K*-nearest neighbor (KNN), decision tree (DT), adaptive neuro fuzzy inference system (ANFIS), multi-layer perceptron neural network (MLPNN), and support vector regression (SVR) for predicting river water temperature, and they reported that none of the reported models was able to reduce the RMSE below the level of 1.400°C, therefore highlighting the extent and the importance of the modelling framework reported in the present study.

## Conclusions

According to the numerical results obtained in this study, we can conclude that dam reservoirs contributed significantly to the alteration of the thermal water regimes. Especially, they are responsible for the continuous and progressive water heating in the downstream river environment. Consequently, building models for simulating dam's reservoir behaviors, continuous monitoring, and control of water temperature need to be continuously observed. Accurate forecasting of the water temperature variation in dams and lakes may help in the building and managing dams and lake's water utilization. To predict daily water temperature fluctuation of the Yangtze River in Cuntan, China, we tested and developed several artificial intelligence models, namely RSS, REPTree, RF, and M5P, according to several input variable combinations. The best input combination was found to be water temperature measured at three lags times, i.e., (t-1), (t-3), and (t-7) for pre-impact and (t-1) to (t-8) with the exclusion of (t-5) for post-impact. Our findings indicated that M5P outperformed all models exhibiting high performances and the best forecasting accuracy with the lowest MAE, RMSE, RAE, RRSE, and the greatest  $R^2$  and PCC. Furthermore, model validation based on graphical analysis revealed that plotting the data points using histograms and scatterplot demonstrate the superiority of the M5P for which data were less scattered than the other models indicating that it has potential for broader use in water temperature prediction.

**Acknowledgements** We also wish to thank the China Hydrological Bureau of the Yangtze River Water Resources Commission and the China Meteorological Administration for providing data for this study. Alban Kuriqi acknowledges the Portuguese Foundation for Science and Technology (FCT) support through PTDC/CTA-OHR/30561/2017 (WinTheface).

**Author contribution** Rawshan Ali and Ahmed Elbeltagi contributed conceptualization and framework of this research, processed data, and designed the experiment. Dinesh Kumar Vishwakarma, Shakeel Ahmad Bhat, Jitendra Rajput, and Nand Lal Kushwaha validated the processing results and wrote the manuscript. Rohitashw Kumar, Rawshan Ali, Alban Kuriqi, Ahmed Elbeltagi, and Salim Heddham edited the manuscript, reviewed it, and helped improve the English language, grammar, punctuation, spelling, and technical improvements. Dinesh Kumar Vishwakarma: writing—resources, figure plots, and map preparation. Alban Kuriqi, Salim Heddham, and Rohitashw Kumar: supervision.

**Data availability** The data observed from an extensive experimental study in a laboratory was used for analysis to develop models and performance assessments. Data supporting this study's findings are available from the corresponding author upon reasonable request.

## Declarations

**Ethics approval** Not applicable

**Consent to participate** Yes.

**Consent for publication** Yes

**Conflict of interest** The authors declare no competing interests.

## References

- Abbot J, Marohasy J (2012) Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Adv Atmos Sci* 29:717–730. <https://doi.org/10.1007/s00376-012-1259-9>
- Adnan RM, Liang Z, Heddham S et al (2020) Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *J Hydrol* 586:124371. <https://doi.org/10.1016/j.jhydrol.2019.124371>
- Ağbulut Ü, Gürel AE, Biçen Y (2021) Prediction of daily global solar radiation using different machine learning algorithms: evaluation and comparison. *Renew Sustain Energy Rev* 135:110114. <https://doi.org/10.1016/j.rser.2020.110114>
- Aghelpour P, Bahrami-Pichaghchi H, Varshavian V (2021) Hydrological drought forecasting using multi-scalar streamflow drought index, stochastic models and machine learning approaches, in northern Iran. *Stoch Environ Res Risk Assess* 35:1615–1635. <https://doi.org/10.1007/s00477-020-01949-z>
- Ahmed K, Shahid S, Chung E et al (2017) Spatial distribution of secular trends in annual and seasonal precipitation over Pakistan. *Clim Res* 74:95–107. <https://doi.org/10.3354/cr01489>
- Ajmera TK, Goyal MK (2012) Development of stage–discharge rating curve using model tree and neural networks: an application to Peachtree Creek in Atlanta. *Expert Syst Appl* 39:5702–5710. <https://doi.org/10.1016/j.eswa.2011.11.101>
- Ali Ghorbani M, Khatibi R, Aytak A et al (2010) Sea water level forecasting using genetic programming and comparing the performance with artificial neural networks. *Comput Geosci* 36:620–627. <https://doi.org/10.1016/j.cageo.2009.09.014>
- Ali R, Ismael A, Heryansyah A, Nawaz N (2019a) Long term historic changes in the flow of Lesser Zab River. *Iraq. Hydrology* 6:22. <https://doi.org/10.3390/hydrology6010022>
- Ali R, Kuriqi A, Abubaker S, Kisi O (2019b) Hydrologic alteration at the upper and middle part of the Yangtze River, China: towards sustainable water resource management under increasing water exploitation. *Sustainability* 11:5176
- Ali R, Kuriqi A, Abubaker S, Kisi O (2019c) Long-term trends and seasonality detection of the observed flow in Yangtze River using Mann-Kendall and Sen's innovative trend method. *Water* 11:1855
- Alizamir M, Kim S, Kisi O, Zounemat-Kermani M (2020) Deep echo state network: a novel machine learning approach to model dew point temperature using meteorological variables. *Hydrol Sci J* 65:1173–1190. <https://doi.org/10.1080/02626667.2020.1735639>
- Allah NA, Nima C, Tsai FT-C, Asghari MA (2014) Bayesian artificial intelligence model averaging for hydraulic conductivity estimation. *J Hydrol Eng* 19:520–532. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000824](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000824)
- Almikael W, Čubanová L, Šoltész A (2022) Hydrological drought forecasting using machine learning—Gidra River case study. *Water*, p 14
- Al-Mukhtar M (2021) Modeling the monthly pan evaporation rates using artificial intelligence methods: a case study in Iraq. *Environ Earth Sci* 80:39. <https://doi.org/10.1007/s12665-020-09337-0>
- Al-rimy BAS, Maarof MA, Shaid SZM (2019) Crypto-ransomware early detection model using novel incremental bagging with

- enhanced semi-random subspace selection. *Futur Gener Comput Syst* 101:476–491. <https://doi.org/10.1016/j.future.2019.06.005>
- Angelaki A, Singh Nain S, Singh V, Sihag P (2021) Estimation of models for cumulative infiltration of soil using machine learning methods. *ISH J Hydraul Eng* 27:162–169. <https://doi.org/10.1080/09715010.2018.1531274>
- Angilletta MJ Jr, Ashley Steel E, Bartz KK et al (2008) Big dams and salmon evolution: changes in thermal regimes and their potential evolutionary consequences. *Evol Appl* 1:286–299. <https://doi.org/10.1111/j.1752-4571.2008.00032.x>
- Araya SN, Ghezzehei TA (2019) Using machine learning for prediction of saturated hydraulic conductivity and its sensitivity to soil structural perturbations. *Water Resour Res* 55:5715–5737. <https://doi.org/10.1029/2018WR024357>
- Armin A, Saeed F, Hadi S et al (2021) Approaches for optimizing the performance of adaptive neuro-fuzzy inference system and least-squares support vector machine in precipitation modeling. *J Hydrol Eng* 26:4021010. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0002069](https://doi.org/10.1061/(ASCE)HE.1943-5584.0002069)
- Ayele GT, Teshale EZ, Yu B et al (2017) Streamflow and sediment yield prediction for watershed prioritization in the upper blue Nile River basin, Ethiopia. *Water* 9:782
- Azamathulla HM, Cuan YC, Ghani AA, Chang CK (2013) Suspended sediment load prediction of river systems: GEP approach. *Arab J Geosci* 6:3469–3480. <https://doi.org/10.1007/s12517-012-0608-4>
- Bahramifar A, Shirkhani R, Mohammadi M (2013) An anfis-based approach for predicting the manning roughness coefficient in alluvial channels at the bank-full stage. *Int J Eng* 26:177–186
- Başakın EE, Ekmekcioğlu Ö, Çıtakoğlu H, Özger M (2022) A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment. *Neural Comput Appl* 34:783–812. <https://doi.org/10.1007/s00521-021-06424-6>
- Bilgili M (2010) Prediction of soil temperature using regression and artificial neural network models. *Meteorol Atmos Phys* 110:59–70. <https://doi.org/10.1007/s00703-010-0104-x>
- Breiman L (2001) Random Forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Buyukyildiz M, Tezel G, Yilmaz V (2014) Estimation of the change in lake water level by artificial intelligence methods. *Water Resour Manag* 28:4747–4763. <https://doi.org/10.1007/s11269-014-0773-1>
- Che Nordin NF, Mohd NS, Koting S et al (2021) Groundwater quality forecasting modelling using artificial intelligence: a review. *Groundw Sustain Dev* 14:100643. <https://doi.org/10.1016/j.gsd.2021.100643>
- Chen Z, Zhu Z, Jiang H, Sun S (2020) Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. *J Hydrol* 591:125286. <https://doi.org/10.1016/j.jhydrol.2020.125286>
- Cheng Y, Voisin N, Yearsley JR, Nijssen B (2020) Reservoirs modify river thermal regime sensitivity to climate change: a case study in the Southeastern United States. *Water Resour Res* 56:e2019WR025784. <https://doi.org/10.1029/2019WR025784>
- Chia MY, Huang YF, Koo CH, Fung KF (2020) Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—a review. *Agron*. 10
- Citakoglu H (2021) Comparison of multiple learning artificial intelligence models for estimation of long-term monthly temperatures in Turkey. *Arab J Geosci* 14:2131. <https://doi.org/10.1007/s12517-021-08484-3>
- Citakoglu H, Coşkun Ö (2022) Comparison of hybrid machine learning methods for the prediction of short-term meteorological droughts of Sakarya Meteorological Station in Turkey. *Environ Sci Pollut Res*. <https://doi.org/10.1007/s11356-022-21083-3>
- Citakoglu H, Minarecioglu N (2021) Trend analysis and change point determination for hydro-meteorological and groundwater data of Kizilirmak basin. *Theor Appl Climatol* 145:1275–1292. <https://doi.org/10.1007/s00704-021-03696-9>
- Cornejo-Bueno L, Casanova-Mateo C, Sanz-Justo J, Salcedo-Sanz S (2019) Machine learning regressors for solar radiation estimation from satellite data. *Sol Energy* 183:768–775. <https://doi.org/10.1016/j.solener.2019.03.079>
- Daliakopoulos IN, Coulibaly P, Tsanis IK (2005) Groundwater level forecasting using artificial neural networks. *J Hydrol* 309:229–240. <https://doi.org/10.1016/j.jhydrol.2004.12.001>
- Demir V (2022) Enhancing monthly lake levels forecasting using heuristic regression techniques with periodicity data component: application of Lake Michigan. *Theor Appl Climatol* 148:915–929. <https://doi.org/10.1007/s00704-022-03982-0>
- Devasena CL (2014) Comparative analysis of random forest, REP tree and J48 classifiers for credit risk prediction. *Int J Comput Appl*:975–8887
- El Bilali A, Taleb A, Brouziyne Y (2021) Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agric Water Manag* 245:106625. <https://doi.org/10.1016/j.agwat.2020.106625>
- Elbeltagi A, Kushwaha NL, Rajput J et al (2022a) Modelling daily reference evapotranspiration based on stacking hybridization of ANN with meta-heuristic algorithms under diverse agro-climatic conditions. *Stoch Environ Res Risk Assess*. <https://doi.org/10.1007/s00477-022-02196-0>
- Elbeltagi A, Raza A, Hu Y et al (2022b) Data intelligence and hybrid metaheuristic algorithms-based estimation of reference evapotranspiration. *Appl Water Sci* 12:152. <https://doi.org/10.1007/s13201-022-01667-7>
- Esmailbeiki F, Nikpour MR, Singh VK et al (2020) Exploring the application of soft computing techniques for spatial evaluation of groundwater quality variables. *J Clean Prod* 276:124206. <https://doi.org/10.1016/j.jclepro.2020.124206>
- Feng Y, Gong D, Zhang Q et al (2019) Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. *Energy Convers Manag* 198:111780. <https://doi.org/10.1016/j.enconman.2019.111780>
- Ferreira LB, da Cunha FF (2020) New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. *Agric Water Manag* 234:106113. <https://doi.org/10.1016/j.agwat.2020.106113>
- Gao B, Yang D, Zhao T, Yang H (2012) Changes in the eco-flow metrics of the Upper Yangtze River from 1961 to 2008. *J Hydrol* 448–449:30–38. <https://doi.org/10.1016/j.jhydrol.2012.03.045>
- Giustolisi O (2004) Using genetic programming to determine Chèzy resistance coefficient in corrugated channels. *J Hydroinformatics* 6:157–173. <https://doi.org/10.2166/hydro.2004.0013>
- Görkemli B, Citakoglu H, Haktanir T, Karaboga D (2022) A new method based on artificial bee colony programming for the regional standardized intensity–duration–frequency relationship. *Arab J Geosci* 15:272. <https://doi.org/10.1007/s12517-021-09377-1>
- Granata F (2019) Evapotranspiration evaluation models based on machine learning algorithms—a comparative study. *Agric Water Manag* 217:303–315. <https://doi.org/10.1016/j.agwat.2019.03.015>
- Gupta D, Hazarika BB, Berlin M et al (2021) Artificial intelligence for suspended sediment load prediction: a review. *Environ Earth Sci* 80:346. <https://doi.org/10.1007/s12665-021-09625-3>
- Haktanir T, Citakoglu H (2014) Trend, independence, stationarity, and homogeneity tests on maximum rainfall series of standard durations recorded in Turkey. *J Hydrol Eng* 19:05014009. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000973](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000973)



- Harris EL, Babovic V, Falconer RA (2003) Velocity predictions in compound channels with vegetated floodplains using genetic programming. *Int J River Basin Manag* 1:117–123. <https://doi.org/10.1080/15715124.2003.9635198>
- Hassan MA, Khalil A, Kaseb S, Kassem MA (2017) Potential of four different machine-learning algorithms in modeling daily global solar radiation. *Renew Energy* 111:52–62. <https://doi.org/10.1016/j.renene.2017.03.083>
- Hastie T, Tibshirani R, Friedman JH, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction. Springer, Springer Science+Business Media, New York
- Heddam S, Ptak M, Zhu S (2020) Modelling of daily lake surface water temperature from air temperature: extremely randomized trees (ERT) versus Air2Water, MARS, M5Tree. RF and MLPNN. *J Hydrol* 588:125130. <https://doi.org/10.1016/j.jhydrol.2020.125130>
- Heddam S, Ptak M, Sojka M et al (2022) Least square support vector machine-based variational mode decomposition: a new hybrid model for daily river water temperature modeling. *Environ Sci Pollut Res*. <https://doi.org/10.1007/s11356-022-20953-0>
- Hong H, Liu J, Zhu A-X et al (2017) A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China). *Environ Earth Sci* 76:652. <https://doi.org/10.1007/s12665-017-6981-2>
- Islam ARMT, Ahmed I, Rahman MS (2020) Trends in cooling and heating degree-days overtimes in Bangladesh? An investigation of the possible causes of changes. *Nat Hazards* 101:879–909. <https://doi.org/10.1007/s11069-020-03900-5>
- Jackson HM, Gibbins CN, Soulsby C (2007) Role of discharge and temperature variation in determining invertebrate community structure in a regulated river. *River Res Appl* 23:651–669. <https://doi.org/10.1002/rra.1006>
- Jiang B, Wang F, Ni G (2018) Heating impact of a tropical reservoir on downstream water temperature: a case study of the Jinghong Dam on the Lancang River. *Water* 10:951
- Kakaei Lafdani E, Moghaddam Nia A, Ahmadi A (2013) Daily suspended sediment load prediction using artificial neural networks and support vector machines. *J Hydrol* 478:50–62. <https://doi.org/10.1016/j.jhydrol.2012.11.048>
- Kashi H, Emamgholizadeh S, Ghorbani H (2014) Estimation of soil infiltration and cation exchange capacity based on multiple regression, ANN (RBF, MLP), and ANFIS Models. *Commun Soil Sci Plant Anal* 45:1195–1213. <https://doi.org/10.1080/00103624.2013.874029>
- Khodakhah H, Aghelpour P, Hamed Z (2022) Comparing linear and non-linear data-driven approaches in monthly river flow prediction, based on the models SARIMA, LSSVM, ANFIS, and GMDH. *Environ Sci Pollut Res* 29:21935–21954. <https://doi.org/10.1007/s11356-021-17443-0>
- Khosravi K, Pham BT, Chapi K et al (2018) A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci Total Environ* 627:744–755. <https://doi.org/10.1016/j.scitotenv.2018.01.266>
- Khosravi K, Golkarian A, Tiefenbacher JP (2022) Using optimized deep learning to predict daily streamflow: a comparison to common machine learning algorithms. *Water Resour Manag* 36:699–716. <https://doi.org/10.1007/s11269-021-03051-7>
- Kisi O, Shiri J, Nikoofar B (2012) Forecasting daily lake levels using artificial intelligence approaches. *Comput Geosci* 41:169–180. <https://doi.org/10.1016/j.cageo.2011.08.027>
- Kumar A, Kumar P, Singh VK (2019) Evaluating different machine learning models for runoff and suspended sediment simulation. *Water Resour Manag* 33:1217–1231. <https://doi.org/10.1007/s11269-018-2178-z>
- Kumar A, Singh VK, Saran B et al (2022) Development of novel hybrid models for prediction of drought- and stress-tolerance indices in teosinte introgressed maize lines using artificial intelligence techniques. *Sustainability* 14:2287. <https://doi.org/10.3390/su14042287>
- Kuriqi A, Ali R, Pham QB et al (2020) Seasonality shift and streamflow flow variability trends in central India. *Acta Geophys* 68:1461–1475. <https://doi.org/10.1007/s11600-020-00475-4>
- Kushwaha NL, Bhardwaj A (2016) Hydrologic response of Takarlaballowal watershed in Shivalik foot-hills based on morphometric analysis using Remote Sensing and GIS. *Indian Water Resour Soc* 36:17–25
- Kushwaha NL, Rajput J, Elbeltagi A et al (2021) Data intelligence model and meta-heuristic algorithms-based pan evaporation modelling in two different agro-climatic zones: a case study from Northern India. *Atmosphere (Basel)* 12:1654. <https://doi.org/10.3390/atmos12121654>
- Lee J, Kim C-G, Lee JE et al (2018) Application of artificial neural networks to rainfall forecasting in the Geum River basin. *Korea. Water* 10:1448
- Leighton HV, Srivastava J (1999) First 20 precision among World Wide Web search services (search engines). *J Am Soc Inf Sci* 50:870–881. [https://doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:10<870::AID-AS14>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-4571(1999)50:10<870::AID-AS14>3.0.CO;2-G)
- Li H, Lee Y-C, Zhou Y-C, Sun J (2011) The random subspace binary logit (RSBL) model for bankruptcy prediction. *Knowledge-Based Syst* 24:1380–1388. <https://doi.org/10.1016/j.knosys.2011.06.015>
- Li Z, Yang D, Hong Y et al (2014) Characterizing spatiotemporal variations of hourly rainfall by Gauge and Radar in the mountainous Three Gorges Region. *J Appl Meteorol Climatol* 53:873–889. <https://doi.org/10.1175/JAMC-D-13-0277.1>
- Liu X, Beusen AHW, Van Beek LPH et al (2018) Exploring spatiotemporal changes of the Yangtze River (Changjiang) nitrogen and phosphorus sources, retention and export to the East China Sea and Yellow Sea. *Water Res* 142:246–255. <https://doi.org/10.1016/j.watres.2018.06.006>
- Luk KC, Ball JE, Sharma A (2001) An application of artificial neural networks for rainfall forecasting. *Math Comput Model* 33:683–693. [https://doi.org/10.1016/S0895-7177\(00\)00272-7](https://doi.org/10.1016/S0895-7177(00)00272-7)
- Malik A, Kumar A, Kisi O, Shiri J (2019) Evaluating the performance of four different heuristic approaches with gamma test for daily suspended sediment concentration modeling. *Environ Sci Pollut Res* 26:22670–22687. <https://doi.org/10.1007/s11356-019-05553-9>
- Malik A, Kumar A, Kim S et al (2020) Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model. *Eng Appl Comput Fluid Mech* 14:323–338. <https://doi.org/10.1080/19942060.2020.1715845>
- Malik A, Tikhmarine Y, Sammen SS et al (2021) Prediction of meteorological drought by using hybrid support vector regression optimized with HHO versus PSO algorithms. *Environ Sci Pollut Res* 28:39139–39158. <https://doi.org/10.1007/s11356-021-13445-0>
- Melesse AM, Ahmad S, McClain ME et al (2011) Suspended sediment load prediction of river systems: an artificial neural network approach. *Agric Water Manag* 98:855–866. <https://doi.org/10.1016/j.agwat.2010.12.012>
- Mirabbasi R, Kisi O, Sanikhani H, Gajbhiye Meshram S (2019) Monthly long-term rainfall estimation in Central India using M5Tree, MARS, LSSVR, ANN and GEP models. *Neural Comput Appl* 31:6843–6862. <https://doi.org/10.1007/s00521-018-3519-9>
- Mohammad Z-K, Ozgur K, Jamshid P, Amin M-M (2019) Assessment of artificial intelligence-based models and metaheuristic






- algorithms in modeling evaporation. *J Hydrol Eng* 24:4019033. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001835](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001835)
- Mohanta A, Patra KC, Sahoo BB (2018) Anticipate Manning's coefficient in meandering compound channels. *Hydrol* 5
- Mohsenzadeh Karimi S, Kisi O, Porrajabali M et al (2020) Evaluation of the support vector machine, random forest and geo-statistical methodologies for predicting long-term air temperature. *ISH J Hydraul Eng* 26:376–386. <https://doi.org/10.1080/09715010.2018.1495583>
- Muhammad J, Muhammad BI, Dongkyun K, Tae-Woong K (2021) Comprehensive evaluation of machine learning techniques for hydrological drought forecasting. *J Irrig Drain Eng* 147:4021022. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0001575](https://doi.org/10.1061/(ASCE)IR.1943-4774.0001575)
- Murchie KJ, Hair KPE, Pullen CE et al (2008) Fish response to modified flow regimes in regulated rivers: research methods, effects and opportunities. *River Res Appl* 24:197–217. <https://doi.org/10.1002/rra.1058>
- Naganna SR, Deka PC, Ghorbani MA et al (2019) Dew point temperature estimation: application of artificial intelligence model integrated with nature-inspired optimization algorithms. *Water*:11
- Nayak PC, Rao YRS, Sudheer KP (2006) Groundwater level forecasting in a shallow aquifer using artificial neural network approach. *Water Resour Manag* 20:77–90. <https://doi.org/10.1007/s11269-006-4007-z>
- Nhu V-H, Shahabi H, Nohani E et al (2020) Daily water level prediction of Zrebar lake (Iran): a comparison between M5P, random forest, random tree and reduced error pruning trees algorithms. *ISPRS Int J Geo-Information* 9:497
- Olden JD, Naiman RJ (2010) Incorporating thermal regimes into environmental flows assessments: modifying dam operations to restore freshwater ecosystem integrity. *Freshw Biol* 55:86–107. <https://doi.org/10.1111/j.1365-2427.2009.02179.x>
- Olsson J, Uvo CB, Jinno K et al (2004) Neural networks for rainfall forecasting by atmospheric downscaling. *J Hydrol Eng* 9:1–12. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2004\)9:1\(1\)](https://doi.org/10.1061/(ASCE)1084-0699(2004)9:1(1))
- Oprea M, Dragomir EG, Popescu M, Mihalache SF (2016) Particulate matter air pollutants forecasting using inductive learning approach. *Rev Chim* 67:2075–2081
- Özçoban MŞ, Isenkul ME, Sevgen S, et al (2022) Modelling the effects of nanomaterial addition on the permeability of the compacted clay soil using machine learning-based flow resistance analysis. *Appl. Sci.* 12
- Panahi M, Khosravi K, Ahmad S et al (2021) Cumulative infiltration and infiltration rate prediction using optimized deep learning algorithms: a study in Western Iran. *J Hydrol Reg Stud* 35:100825. <https://doi.org/10.1016/j.ejrh.2021.100825>
- Pavey TG, Gilson ND, Gomersall SR et al (2017) Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *J Sci Med Sport* 20:75–80. <https://doi.org/10.1016/j.jsams.2016.06.003>
- Penghui L, Ewees AA, Beyaztas BH et al (2020) Metaheuristic optimization algorithms hybridized with artificial intelligence model for soil temperature prediction: novel model. *IEEE Access* 8:51884–51904. <https://doi.org/10.1109/ACCESS.2020.2979822>
- Pham BT, Prakash I, Tien Bui D (2018) Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* 303:256–270. <https://doi.org/10.1016/j.geomorph.2017.12.008>
- Pham BT, Jaafari A, Nguyen-Thoi T et al (2021) Ensemble machine learning models based on reduced error pruning tree for prediction of rainfall-induced landslides. *Int J Digit Earth* 14:575–596. <https://doi.org/10.1080/17538947.2020.1860145>
- Pham QB, Kumar M, Di Nunno F et al (2022) Groundwater level prediction using machine learning algorithms in a drought-prone area. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-022-07009-7>
- Plumpton CO, Kuncheva LI, Oosterhof NN, Johnston SJ (2012) Naive random subspace ensemble with linear classifiers for real-time classification of fMRI data. *Pattern Recognit* 45:2101–2108. <https://doi.org/10.1016/j.patcog.2011.04.023>
- Pradhan A, Khatua KK (2017) Gene expression programming to predict Manning's n in meandering flows. *Can J Civ Eng* 45:304–313. <https://doi.org/10.1139/cjce-2016-0569>
- Preece R (2004) Cold water pollution below dams in New South Wales: a desktop assessment. Water management Division, Department of Infrastructure, Planning and ...
- Qasem SN, Samadianfard S, Sadri Nahand H et al (2019) Estimating daily dew point temperature using machine learning algorithms. *Water* 11:582
- Quinlan JR (1992) Learning with continuous classes. In: 5th Australian joint conference on artificial intelligence. World Scientific, pp 343–348
- Rahman MS, Islam ARMT (2019) Are precipitation concentration and intensity changing in Bangladesh overtimes? Analysis of the possible causes of changes in precipitation systems. *Sci Total Environ* 690:370–387. <https://doi.org/10.1016/j.scitotenv.2019.06.529>
- Rajaei T, Nourani V, Zounemat-Kermani M, Kisi O (2011) River suspended sediment load prediction: application of ANN and wavelet conjunction model. *J Hydrol Eng* 16:613–627. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000347](https://doi.org/10.1061/(asce)he.1943-5584.0000347)
- Rajesh P, Karthikeyan M (2017) A comparative study of data mining algorithms for decision tree approaches using WEKA tool. *Adv Nat Appl Sci* 11:230+
- Rheinheimer DE, Null SE, Lund JR (2015) Optimizing selective withdrawal from reservoirs to manage downstream temperatures with climate warming. *J Water Resour Plan Manag* 141:4014063. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000447](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000447)
- Ryan T, Webb AA, Lennie R, Lyon JP (2001) Status of cold water releases from Victorian dams [Report produced for Catchment and Water, Department of Natural Resources and Environment]. Department of Natural Resources and Environment Arthur Rylah Institute 123 Brown Street Heidelberg, Victoria, 3084
- Saha S, Saha M, Mukherjee K et al (2020) Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPtree: a case study at the Gumani River basin. *India. Sci Total Environ* 730:139197. <https://doi.org/10.1016/j.scitotenv.2020.139197>
- Salam R, Towfiqul Islam ARM, Shill BK et al (2021) Nexus between vulnerability and adaptive capacity of drought-prone rural households in northern Bangladesh. *Nat Hazards* 106:509–527. <https://doi.org/10.1007/s11069-020-04473-z>
- Sayari S, Mahdavi-Meymand A, Zounemat-Kermani M (2021) Irrigation water infiltration modeling using machine learning. *Comput Electron Agric* 180:105921. <https://doi.org/10.1016/j.compag.2020.105921>
- Seo Y, Kim S, Kisi O, Singh VP (2015) Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J Hydrol* 520:224–243. <https://doi.org/10.1016/j.jhydrol.2014.11.050>
- Seo Y, Choi E, Yeo W (2017) Reservoir water level forecasting using machine learning models. *J Korean Soc Agric Eng* 59:97–110. <https://doi.org/10.5389/KSAE.2017.59.3.097>
- Shamshirband S, Hashemi S, Salimi H et al (2020) Predicting standardized streamflow index for hydrological drought using machine learning models. *Eng Appl Comput Fluid Mech* 14:339–350. <https://doi.org/10.1080/19942060.2020.1715844>
- Sharma N, Zakaullah M, Tiwari H, Kumar D (2015) Runoff and sediment yield modeling using ANN and support vector machines: a case study from Nepal watershed. *Model Earth Syst Environ* 1:23. <https://doi.org/10.1007/s40808-015-0027-0>
- Shiri J, Özgür K (2011) Application of artificial intelligence to estimate daily pan evaporation using available and estimated climatic data

- in the Khozestan Province (South Western Iran). *J Irrig Drain Eng* 137:412–425. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000315](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000315)
- Shiri N, Shiri J, Yaseen ZM et al (2021) Development of artificial intelligence models for well groundwater quality simulation: different modeling scenarios. *PLoS One* 16:e0251510
- Shukla R, Kumar P, Vishwakarma DK et al (2021) Modeling of stage-discharge using back propagation ANN-, ANFIS-, and WANN-based computing techniques. *Theor Appl Climatol*. <https://doi.org/10.1007/s00704-021-03863-y>
- Sihag P, Esmailbeiki F, Singh B et al (2019a) Modeling unsaturated hydraulic conductivity by hybrid soft computing techniques. *Soft Comput* 23:12897–12910. <https://doi.org/10.1007/s00500-019-03847-1>
- Sihag P, Singh VP, Angelaki A et al (2019b) Modelling of infiltration using artificial intelligence techniques in semi-arid Iran. *Hydrol Sci J* 64:1647–1658. <https://doi.org/10.1080/02626667.2019.1659965>
- Singh VK, Singh BP, Kisi O, Kushwaha DP (2018) Spatial and multi-depth temporal soil temperature assessment by assimilating satellite imagery, artificial intelligence and regression based models in arid area. *Comput Electron Agric* 150:205–219. <https://doi.org/10.1016/j.compag.2018.04.019>
- Singh VK, Kumar D, Kashyap PS, Singh PK (2019) Predicting unsaturated hydraulic conductivity of soil based on machine learning algorithms. In: Conference: Proceedings of International Conference Opportunities and Challenges in Engineering, Management and Science (OCEMS-2019)At: Bareilly, India
- Singh VK, Kumar D, Kashyap PS et al (2020) Modelling of soil permeability using different data driven algorithms based on physical properties of soil. *J Hydrol* 580:124223. <https://doi.org/10.1016/j.jhydrol.2019.124223>
- Singh B, Sihag P, Pandhiani SM et al (2021a) Estimation of permeability of soil using easy measured soil parameters: assessing the artificial intelligence-based models. *ISH J Hydraul Eng* 27:38–48. <https://doi.org/10.1080/09715010.2019.1574615>
- Singh VK, Kumar D, Singh SK et al (2021b) Development of fuzzy analytic hierarchy process based water quality model of Upper Ganga river basin. India. *J Environ Manage* 284:111985. <https://doi.org/10.1016/j.jenvman.2021.111985>
- Singh VK, Panda KC, Sagar A et al (2022) Novel genetic algorithm (GA) based hybrid machine learning-pedotransfer function (ML-PTF) for prediction of spatial pattern of saturated hydraulic conductivity. *Eng Appl Comput Fluid Mech* 16:1082–1099. <https://doi.org/10.1080/19942060.2022.2071994>
- Singha S, Pasupuleti S, Singha SS et al (2021) Prediction of groundwater quality using efficient machine learning technique. *Chemosphere* 276:130265. <https://doi.org/10.1016/j.chemosphere.2021.130265>
- Solomatine DP, Xue Y (2004) M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. *J Hydrol Eng* 9:491–501. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2004\)9:6\(491\)](https://doi.org/10.1061/(ASCE)1084-0699(2004)9:6(491))
- Srinivasan DB, Mekala P (2014) mining social networking data for classification using Reptree. *Int J Adv Res Comput Sci Manag Stud* 2:155–160
- Tao Y, Wang Y, Rhoads B et al (2020) Quantifying the impacts of the Three Gorges Reservoir on water temperature in the middle reach of the Yangtze River. *J Hydrol* 582:124476. <https://doi.org/10.1016/j.jhydrol.2019.124476>
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106:7183–7192. <https://doi.org/10.1029/2000JD900719>
- Tefaruk H, Hatice C (2015) Closure to “Trend, Independence, Stationarity, and Homogeneity Tests on Maximum Rainfall Series of Standard Durations Recorded in Turkey” by Tefaruk Haktanir and Hatice Citakoglu. *J Hydrol Eng* 20:7015017. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001246](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001246)
- Tikhmarine Y, Malik A, Kumar A et al (2019) Estimation of monthly reference evapotranspiration using novel hybrid machine learning approaches. *Hydrol Sci J* 64:1824–1842. <https://doi.org/10.1080/02626667.2019.1678750>
- Tikhmarine Y, Malik A, Souag-Gamane D, Kisi O (2020) Artificial intelligence models versus empirical equations for modeling monthly reference evapotranspiration. *Environ Sci Pollut Res* 27:30001–30019. <https://doi.org/10.1007/s11356-020-08792-3>
- Vezzoli G, Garzanti E, Limonta M et al (2016) Erosion patterns in the Changjiang (Yangtze River) catchment revealed by bulk-sample versus single-mineral provenance budgets. *Geomorphology* 261:177–192. <https://doi.org/10.1016/j.geomorph.2016.02.031>
- Vishwakarma DK, Pandey K, Kaur A et al (2022) Methods to estimate evapotranspiration in humid and subtropical climate conditions. *Agric Water Manag* 261:107378. <https://doi.org/10.1016/j.agwat.2021.107378>
- Vitkar S (2017) Comparative analysis of various data mining prediction algorithms demonstrated using air pollution data of Navi Mumbai. *Res J Chem Environ Sci* 5:79–85
- Voyant C, Notton G, Kalogirou S et al (2017) Machine learning methods for solar radiation forecasting: a review. *Renew Energy* 105:569–582. <https://doi.org/10.1016/j.renene.2016.12.095>
- Wang Y, Wang D, Lewis QW et al (2017) A framework to assess the cumulative impacts of dams on hydrological regime: a case study of the Yangtze River. *Hydrol Process* 31:3045–3055. <https://doi.org/10.1002/hyp.11239>
- Witten IH, Frank E, Hall MA, Pal CJ (2016) Practical machine learning tools and techniques, 4th edn. Elsevier, Amsterdam
- Wu J, Gao X, Giorgi F et al (2012) Climate effects of the Three Gorges Reservoir as simulated by a high resolution double nested regional climate model. *Quat Int* 282:27–36. <https://doi.org/10.1016/j.quaint.2012.04.028>
- Wu H, Li J, Song F et al (2018) Spatial and temporal patterns of stable water isotopes along the Yangtze River during two drought years. *Hydrol Process* 32:4–16. <https://doi.org/10.1002/hyp.11382>
- Wu L, Peng Y, Fan J, Wang Y (2019) Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data. *Hydrol Res* 50:1730–1750. <https://doi.org/10.2166/nh.2019.060>
- Xu D, Zhang Q, Ding Y, Zhang D (2022) Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. *Environ Sci Pollut Res* 29:4128–4144. <https://doi.org/10.1007/s11356-021-15325-z>
- Yang X, Lu XX (2012) Model of water regulation in the Yangtze River basin and its effects using remote sensing techniques. In: Proceedings of the IAHS International Commission on Continental Erosion Symposium, Chengdu, China (pp. 11–15) AISH Publication. Erosion and Sediment Yields in the Changing Environment (Proceedings of a symposium held at the Institute of Mountain Hazards and Environment, CAS-Chengdu, China, pp 235–243
- Yang K, Wang J (2008) A temperature prediction-correction method for estimating surface soil heat flux from soil temperature and moisture data. *Sci China Ser D Earth Sci* 51:721–729. <https://doi.org/10.1007/s11430-008-0036-1>
- Yaseen ZM, Kisi O, Demir V (2016) Enhancing long-term streamflow forecasting and predicting using periodicity data component: application of artificial intelligence. *Water Resour Manag* 30:4125–4151. <https://doi.org/10.1007/s11269-016-1408-5>
- Yaseen ZM, Ramal MM, Diop L et al (2018) Hybrid adaptive neuro-fuzzy models for water quality index estimation. *Water Resour Manag* 32:2227–2245. <https://doi.org/10.1007/s11269-018-1915-7>

- Yousefi A, Toffolon M (2022) Critical factors for the use of machine learning to predict lake surface water temperature. *J Hydrol* 606:127418. <https://doi.org/10.1016/j.jhydrol.2021.127418>
- Yu HC, Liu HD, Huang ZQ, Shi GC (2017a) Experimental study on time-dependent behavior of silty mudstone from the Three Gorges Reservoir area, China. *KSCE J Civ Eng* 21:715–724. <https://doi.org/10.1007/s12205-016-3645-9>
- Yu Y, Wang C, Wang P et al (2017b) Assessment of multi-objective reservoir operation in the middle and lower Yangtze River based on a flow regime influenced by the Three Gorges project. *Ecol Inform* 38:115–125. <https://doi.org/10.1016/j.ecoinf.2017.02.004>
- Zhao F, Shepherd M (2012) Precipitation changes near Three Gorges Dam, China. Part I: a spatiotemporal validation analysis. *J Hydrometeorol* 13:735–745. <https://doi.org/10.1175/JHM-D-11-061.1>
- Zhu S, Lu H, Ptak M et al (2020) Lake water-level fluctuation forecasting using machine learning models: a systematic review. *Environ Sci Pollut Res* 27:44807–44819. <https://doi.org/10.1007/s11356-020-10917-7>

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

## Authors and Affiliations

Dinesh Kumar Vishwakarma<sup>1</sup>  · Rawshan Ali<sup>2</sup> · Shakeel Ahmad Bhat<sup>3</sup> · Ahmed Elbeltagi<sup>4</sup>  · Nand Lal Kushwaha<sup>5</sup>  · Rohitashw Kumar<sup>3</sup>  · Jitendra Rajput<sup>5</sup> · Salim Heddami<sup>6</sup>  · Alban Kuriqi<sup>7</sup> 

Rawshan Ali  
rawshan.ali@epu.edu.iq

Shakeel Ahmad Bhat  
wakeelbhat@gmail.com

Ahmed Elbeltagi  
ahmedelbeltagi81@mans.edu.eg

Nand Lal Kushwaha  
nand.kushwaha@icar.gov.in

Rohitashw Kumar  
rohituhf@rediffmail.com

Jitendra Rajput  
jitu.caepht@gmail.com

Salim Heddami  
heddamisalim@yahoo.fr

Alban Kuriqi  
alban.kuriqi@tecnico.ulisboa.pt

<sup>2</sup> Department of Petroleum, Koya Technical Institute, Erbil Polytechnic University, Erbil 44001, Iraq

<sup>3</sup> College of Agricultural Engineering and Technology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar, Jammu and Kashmir 190025, India

<sup>4</sup> Agricultural Engineering Department, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt

<sup>5</sup> Division of Agricultural Engineering, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India

<sup>6</sup> Faculty of Science, Agronomy Department, Hydraulics Division, Laboratory of Research in Biodiversity Interaction Ecosystem and Biotechnology, University 20 Août 1955, Route El Hadaik, BP 26, Skikda, Algeria

<sup>7</sup> CERIS, Instituto Superior Técnico, University of Lisbon, 1649-004 Lisbon, Portugal

<sup>1</sup> Department of Irrigation and Drainage Engineering, G.B. Pant University of Agriculture and Technology, Pantnagar 263145, India