Ain Shams Engineering Journal 16 (2025) 103206

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Enhancing hydropower generation Predictions: A comprehensive study of XGBoost and Support Vector Regression models with advanced optimization techniques

Zhenya Qi *, Yudong Feng, Shoufeng Wang, Chao Li

Shandong Electric Power Engineering Consulting Institute Corr. LTD. Jinan, Shandong 250014 China

ARTICLE INFO

Keywords:
Hydropower Generation prediction
Machine learning
Support vector regression
Extreme Gradient Boosting
Optimization Methods

ABSTRACT

Hydropower plays a crucial role in electricity generation, contributing over 60% of total renewable energy output. Its ability to stabilize energy fluctuations makes it essential in green energy initiatives. Accurate prediction of hydropower production is vital, considering its dependence on various factors like weather, water storage, and electricity generation. Traditional methods struggle with the complexities involved. This study utilized Support Vector Regression (SVR) and eXtreme Gradient Boosting (XGBoost) algorithms, both individually and in hybrid models enhanced by optimization techniques like Slime Mould Algorithm (SMA), Aquila Optimizer (AO), and Grey Wolf Optimization (GWO). XGBoost outperformed SVR in single model predictions with an R^2 value of 0.8632 and RMSE of 40.90, and when optimized, the hybrid XGBoost models showed superior performance, with XGBoost-SMA achieving the highest accuracy. The results revealed that the XGBoost-SMA model achieved the most desired accuracy with an R^2 value of 0.9713 and a root mean square error of 18.73 for the test dataset. This research highlights machine learning's applicability in hydropower prediction and suggests hybrid models as a promising approach for better accuracy, emphasizing XGBoost's potential in efficient hydropower forecasting to meet global electricity demands.

1. Introduction

The growing global population, coupled with advancements in technology, has significantly boosted the demand for electricity [1]. Historically, our energy systems have been predominantly fueled by fossil fuels, contributing to notable environmental challenges [2–4]. This surge in electricity demand has propelled a rapid shift towards renewable energy sources, aiming to reduce carbon-intensive and nuclear power generation, fostering both environmental sustainability and economic progress [5,6]. Among these renewable sources, hydropower stands out for its cost-effectiveness, minimal environmental impact, and ability to quickly adapt to peak electricity demands. Renewable energy is vital for sustaining global energy demands sustainably, reducing greenhouse gas emissions, and tackling climate change. Sources like wind, solar, and hydropower are environmentally friendly and help decrease dependence on fossil fuels. It is also elevating economic growth, and forms green jobs and a crucial option for a sustainable and low-carbon future [7–10].

Hydropower contributes to a significant 75 % of the renewable energy mix worldwide [11]. Consequently, enhancing the efficiency of hydropower generation is paramount for both economic and environmental reasons, two pressing concerns in the contemporary era. An efficiently managed hydropower system offers numerous advantages, such as improved water resource utilization, augmented renewable energy output, a solution to escalating energy needs, minimized equipment wear, and prolonged machinery lifespan. Nonetheless, optimizing hydropower operations presents challenges. Thorough monitoring and a deep understanding of the entire energy conversion process within a hydropower facility are essential to achieving desired outcomes [12–14].

Hydro-power plants harness electricity from the energy generated by falling or swiftly flowing water, typically resulting from rainfall or the melting of snowpacks. Consequently, areas with abundant hydro-power facilities tend to be located on mountain slopes and downstream regions. Countries leading in hydroelectric production often feature prominent mountain ranges [15–17]. For instance, Norway generates over 95 % of its energy needs through hydropower. Other nations like China, the United States, Brazil, and Canada are also increasingly relying on hydropower to fulfill their energy requirements. This shift is largely driven

* Corresponding author.
E-mail address: sdyswlw@163.com (Z. Qi).

https://doi.org/10.1016/j.asej.2024.103206
Received 20 July 2024; Received in revised form 2 November 2024; Accepted 22 November 2024
Available online 7 December 2024

Nomenclature

A10	A10 Index
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Networks
AO	Aquila Optimizer
ARMA	Autoregressive Moving Average Model
C	The penalty coefficient
DL	Deep Learning
DRV	Deviation of Runoff Volume
$F_{t-1}(x_i)$	The cumulative output of the prior t-1 trees
$F(x_n^g)$	The fitness function
$F(x_w)$	The least favorable objective function
GA	Genetic algorithm
GPR	Gaussian Process Regression
GWO	Grey Wolf Optimization
HPG	Hydro-power generation
K(X, xi)	The kernel function
l(•)	The error function
Levy(D)	The Levy flight distribution function
MAE	Mean Absolute Error
MBE	Mean Bias Error
ML	Machine Learning
MLR	Multiple Linear Regression
QF	The primary function employed for the equilibrium search strategy
RMSE	Root Mean Square Error
R^2	Coefficient of Determination
SHP	Small hydropower
SMA	Slime Mould Algorithm
SVM	Support vector machine
T	Transpose operation
SVR	Support Vector Regression
XGBoost	eXtreme Gradient Boosting
x_b	The best solution
$X_{best}(t)$	The optimal solution
Xp	The position vectors of the prey
$X_R(t)$	A random solution
X_α, X_β, and X_δ	The present positions of the first three wolves
x_n^g	The position of the nth slime mould
y_i	Observed output value
$\Omega(f_t)$	Regularization component
ξ	The penalty assigned to prediction errors
α	The Alpha wolf
β	The Beta wolf
δ	Delta wolf
ω	The Omega wolf
ρ	A random value

by their self-set objectives to promote sustainable energy production [18,19].

The functioning of a hydropower plant can be examined either individually within a watercourse or as part of a cascading system. The operation can be categorized as run-of-river [20] or storage reservoir [21]. The process of generating energy encompasses three phases: pre-operation, real-time, and post-operation. Within the pre-operation phase, there are distinct stages such as long-term, medium-term, short-term, and real-time scheduling [22].

Predicting water availability in reservoirs is a complex issue addressed using various computational techniques. Some of these include time series processing algorithms [23], empirical orthogonal functions [24], error correction-based forecasting [25], multivariate strategies [26], and ensemble-based algorithms [27]. Moreover, Machine Learning (ML) methodologies have been effectively incorporated into solutions for these challenges [28].

Artificial Neural Networks (ANN) have been employed widely for tackling complex and non-linear data in hydropower generation prediction. While it is proven that the ANN is effective in modeling intricate patterns, concentrations on the more advanced methods combined with optimizers can beat ANN in accuracy and error decrement. Nevertheless, ANN remains a key tool in predictive modeling and contributes to progress in energy prediction [29–31].

Besides, ML techniques have emerged as valuable tools for harnessing accumulated data, gaining increasing prominence across diverse fields like science, finance, and industry [32–34]. While there are numerous uncertainties involved in hydroelectricity generation, ML models have exhibited remarkable adaptability in making predictions [35]. Traditional methods often struggled with capturing the intricate non-linear relationships between dependent and independent variables [36]. For instance, multi-linear regressions and autoregressive models face challenges due to their limited capacity to account for non-linearity and non-stationarity inherent in environmental and hydrological data [37]. These limitations are being addressed with advanced computational techniques [38]. As ML models find their way into the energy domain, a growing body of literature underscores the extensive research efforts to uncover novel applications in the energy landscape.

Numerous machine learning techniques have proven to be effective in predicting hydro-power generation (HPG). Valentina Sessa et al. (2020) delved into the integration of climate variability in sustainable power systems, particularly focusing on run-of-river hydropower generation influenced by weather variables. Translating meteorological data into hydropower outputs is complex, given the intricate relationships between water availability and electricity generation, with water flow influenced by nonlinear weather variables and basin characteristics. This study compared various machine learning regression algorithms for predicting hydropower generation in France, Portugal, and Spain, finding ensemble trees and neural networks to be the most effective [39]. Dehghani et al. (2019) examined how hydropower generation, influenced by dam reservoir inflow, can be predicted using the Grey Wolf Optimization (GWO) method combined with an Adaptive Neuro-Fuzzy Inference System (ANFIS). Utilizing various input combinations, including rainfall and prior power data, the coupled model demonstrated promise in forecasting. Notably, while GWO-ANFIS showed satisfactory results, standalone ANFIS failed in several scenarios [40]. Alrayees et al. (2018) emphasized the challenges in forecasting renewable energy integration, notably influenced by environmental factors. Hydropower's increasing role necessitates precise production predictions to optimize renewable energy marketing and system integration. This research utilized Machine Learning, including ANN, SVM, and DL, to predict short-term energy outputs of the Almus Dam and Hydroelectric Power Plant in Tokat, Turkey, analyzing data from 1993 to 2013 [41]. Li et al. (2019) highlighted the underexplored potential of deep learning in industrial applications due to data constraints and complexity. Introducing a novel deep neural network model that combines residual and recurrent networks, they innovatively tackled hydroelectric power generation prediction, segmenting data into recent, daily, weekly, and time-series levels. By fusing predictions from these components with varying weights, they achieved significantly superior results over traditional methods, suggesting a fresh avenue for enhanced data utilization in the energy sector [42]. Li et al. (2014) emphasized the challenges of forecasting small hydropower (SHP) energy generation due to limited historical records. They proposed a support vector machine (SVM) model, optimized by the genetic algorithm (GA) for short-term prediction, showcasing SVM's adaptability for small sample forecasting. Testing in Yunlong and Maguan Counties in

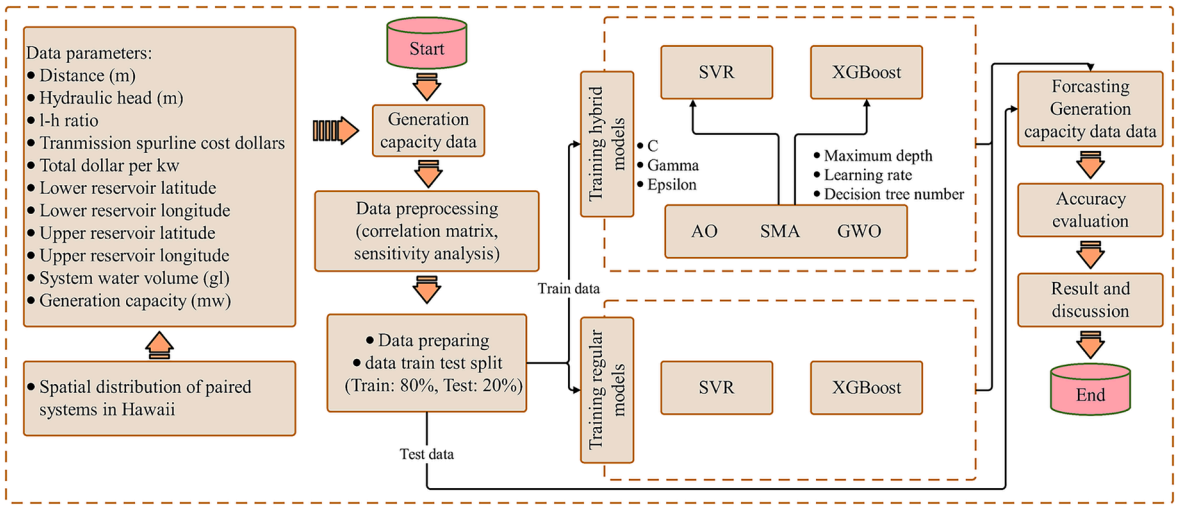


Fig. 1. Flowchart diagram of the current investigation.

Yunnan showcased the GA-SVM model's superiority over traditional ARMA models in predicting SHP energy outputs [43]. Hammid et al. (2016) addressed power challenges in developing nations and utilized Artificial Neural Networks (ANNs) to predict the performance of the Himreen Lake dam-Diyala's hydropower plant. Drawing on a decade of data, their ANN model effectively captured the plant's nonlinear behavior with a correlation coefficient (R) exceeding 0.96 [44]. Sattar Hanoon et al. (2023) leveraged machine learning models to predict power production from a Chinese reservoir, using data spanning 1979 to 2016. They evaluated algorithms like ANN, ARIMA, and SVM across daily, monthly, and seasonal forecasting scenarios. After thorough statistical preprocessing, the models' performances were assessed using five metrics, revealing their potential utility for energy decision-making. Uncertainty analysis further substantiated the robustness and reliability of the ANN and SVM models [45]. Ekanayake et al. (2021) developed predictive models for power generation at the Samanalawewa hydropower plant in Sri Lanka using regression-based machine learning and statistical techniques. Analyzing rainfall data from 1993 to 2019 and other meteorological variables, they evaluated the impact of collinearities between factors. Among the models, including GPR, SVR, MLR, and PR, machine learning methods, particularly GPR, demonstrated superior accuracy in forecasting hydropower based on rainfall predictions [18].

1.1. Main Novelties and contributions

Previous research has developed diverse regression models, including SVR and XGBoost, to elevate the accuracy of the predictions, especially in complex implementations like hydropower generation.

In this research, the power of artificial intelligence is employed through the application of Support Vector Regression (SVR) and eXtreme Gradient Boosting (XGBoost) models to predict hydropower generation. To guarantee the utmost accuracy and reliability, SVR and XGBoost are primarily employed as individual predictive models. Advanced optimization methods, namely Aquila Optimizer (AO), Grey Wolf Optimization (GWO), and Slime Mould Algorithm (SMA) are introduced to calibrate and optimize these models, thereby improving their precision and effectiveness. The study's fundamental contributions revolve around tackling the intricate challenges related to hydropower prediction by defining a cutting-edge hybrid approach that integrates SVR and XGBoost algorithms with the power of AO, GWO, and SMA methods. Furthermore, this study conducts a detailed sensitivity examination by use of the Delta Moment Independent Index, which identifies 'system water volume' as the most impactful parameter. A comprehensive examination of diverse machine learning models is performed, which discloses the superior performance of the hybrid XGBoost-SMA model. This study addresses a gap in the current literature, and highlights the potential of optimized hybrid models, especially XGBoost, in improving the efficiency of hydropower forecasting to better meet global electricity demands.

1.2. Paper structure

Section 2 describes the utilized methodology, which focuses on the data collection procedure, machine learning models, and the optimization methods employed for hydropower generation prediction. Section 3 outlines the outputs and a detailed discussion, including correlation and sensitivity analyses, model performance evaluation, and comparisons of single and hybrid models. Finally, Section 4 concludes the findings and highlights the key outputs and contributions to the field of renewable energy predictions.

2. Methodology

In this study, machine learning algorithms have been employed to predict hydropower. The procedure for utilizing these algorithms is presented in the form of a flowchart within this section. A pivotal part of the study is the acquisition of data from a credible source, ensuring both the accuracy and sufficiency of the data. Based on Fig. 1, the utilized input variables include parameters of system water volume, transmission spurline cost, hydraulic head, and total dollar per kW. These inputs outline diverse factors that affect hydropower generation. The input variables were chosen based on their potential impact on the target output. The output report in this evaluation is the generation capacity, which depicts the predicted hydropower output. The relations between these input variables and the output are thoroughly examined through sensitivity investigation and correlation matrices to determine their impact on hydropower capacity prediction.

According to the depicted flowchart in Fig. 1, the obtained data were scrutinized prior to the study using specific methods. Interrelationships, as well as their impact on the output parameter, were investigated. The employed methodologies include sensitivity indices and correlation matrices. Through these methods, influential input parameters on the target parameter were studied. Subsequent to this step, data were divided into training and testing sets. The study employed Support Vector Regression (SVR) and XGBoost algorithms for training. The investigation was conducted initially using single algorithms, followed by hybrid models to enhance accuracy and precision and to enable a more comprehensive comparison. For the construction of hybrid models, optimizers such as SMA, AO, and GWO were utilized. These optimizers refined the hyperparameters of the single models, resulting in

Table 1
Statistical evaluation indexes [48].

Statistics	Criteria	Equation
RMSE	Root Mean Square Error	$\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{T}}$
MAE	Mean Absolute Error	$\frac{\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert}{n}$
MBE	Mean Bias Error	$\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)}{n}$
DRV	Deviation of Runoff Volume	$\frac{\sum_{i=0}^{N-1}y_i}{\sum_{i=0}^{N-1}\hat{y}_i}$
R^2	Coefficient of Determination	$1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$
A10	A10 Index	$\frac{1}{n}\sum_{i=1}^{n}\begin{cases}1 & if\ \frac{\lvert\hat{y}_i-y_i\rvert}{y_i}\leqslant 0.1\\ 0, & otherwise\end{cases}$

the creation of three distinct hybrid models.

The choice of utilized machine learning methods, specifically SVR and XGBoost, was justified by their proven capability to overcome presented complex, non-linear relations in hydropower generation prediction. SVR is identified for its effectiveness in generating regression models from data that may include noise and outliers, while XGBoost is a scalable algorithm that excels in predictive performance, especially when dealing with tabular data and intricate variable interactions [46,47].

The inclusion of the mentioned metaheuristic optimization methods was motivated by their capacity to enhance model accuracy through optimal hyperparameter tuning. These metaheuristics were selected for their distinctive abilities to explore and exploit the search zone effectively, leading to improvements in the predictive performance of the machine learning models. Their integration into hybrid models guarantees refined predictions, which makes them a favorable choice for addressing the complexities and uncertainties associated with hydropower predictions.

These models were then employed for prediction, and their results were validated and compared using various statistical indices including RMSE, MAE, MBE, DRV, A10, and R^2. Table 1 offers a comprehensive breakdown of the mathematical formulas related to these primary evaluation indicators.

2.1. Data collection

The data integral to our study was sourced from reference [49]. In this section, we elaborate on the resource assessment model crafted specifically for our primary area of investigation: Hawaii. Despite its relatively compact size, Hawaii's mountainous terrain leads to the identification of numerous potential reservoirs. Interestingly, Hawaii manifests a larger proportion of reservoirs disqualified based on the critical habitat area criteria compared to other regions. Although existing river systems serve as a primary reason for exclusion, the percentage of exclusions in Hawaii is notably less than in regions like Alaska. This can be partially attributed to the drier leeward regions of the islands, which exhibit a reduced density of rivers and streams. To sum it up, Hawaii witnesses a considerably lower number of reservoirs being excluded in comparison to Alaska.

The majority of the identified systems are situated on the Island of Hawaii, commonly referred to as the "Big Island". This island boasts significant elevation variations and has fewer rivers and streams, especially on its leeward side. Other islands like Maui, Molokai, and Lanai house a modest number of these systems, while Kauai is home to just one. Notably, Oahu, the island with the bulk of the state's population, lacks any of these systems. Cost-wise, the most affordable systems in Hawaii bear considerably steeper expenses than their counterparts in other regions, with costs oscillating between $1,542/kW and $5,485/kW. Such a trend is predominantly rooted in Hawaii's shortage of high-head systems, aligning with the limited elevation disparities present on the islands.

2.2. Machine learning methods

In this section of the study, algorithms employed for modeling purposes are introduced. These algorithms include SVR and XGBoost. Moreover, hybrid models of the mentioned algorithms have been optimized using optimizers such as GWO, AO, and SMA. This section offers a detailed exploration, presenting a succinct overview of the mathematical concepts and fundamental principles behind each of the previously mentioned techniques.

2.2.1. eXtreme Gradient Boosting (XGBoost)

XGBoost is a gradient-boosting algorithm that utilizes multiple decision trees. Each tree aims to correct the errors or residuals left by its predecessors. By combining the outputs of all these trees, the final prediction is obtained. One of the strengths of XGBoost is its ability to effectively process tabular data and its transparency in model interpretation. Additionally, XGBoost incorporates regularization into its objective function, reducing its complexity and mitigating the risk of overfitting[50]. Given a dataset with N instances and a model with Tr trees, the loss for the t-th tree is described as follows:

$$L_t = \sum_i l(y_i, F_{t-1}(x_i) + f_t(x_i)) + \Omega(f_t) \tag{1}$$

The error function for the i-th data point is denoted by $l(\bullet)$. The actual value of this i-th data point is represented by y_i. The prediction for the i-th data point, derived from the cumulative output of the prior t-1 trees, is $F_{t\text{-}1}(x_i)$. The current tree provides a prediction for the i-th data point, which is given by $f_t(x_i)$. To regulate the model's complexity and ward off overfitting, a regularization component, $\Omega(f_t)$, is incorporated.

2.2.2. Support Vector Regression (SVR)

Support Vector Regression (SVR) determines a linear relationship between dependent (or target) variables and independent (or predictor) variables. It does this by aiming to minimize a specific error function. SVR is especially apt for generating regression models from data that might be noisy or contain outliers [51]. When it comes to forecasting, the primary equation of SVR represents a linear function, which can be described as:

$$f(x) = w^{Tr}.x + b \tag{2}$$

In this equation, w is the weight vector associated with the inputs, and b acts as the bias for the product $w^{Tr}\cdot x$ compared to f(x), with T indicating the transpose operation. The output variable, as approximated by SVR, is f(x). To address the risk of overfitting in modeling, Vapnik [52] introduced an error metric in 1998 termed the epsilon-insensitive function.

$$|y - f(x)| = \begin{cases} 0 if |y - f(x)| \le k \\ |y - f(x)| - k = \zeta otherwise \end{cases} \tag{3}$$

In the framework of SVR, y is the observed output. The threshold k is an established error limit, meaning any prediction error below this value is dismissed. Meanwhile, ξ represents the penalty assigned to prediction errors that venture outside the interval (−k, +k).

The primary objective of SVR optimization revolves around minimizing the epsilon-insensitive function as well as reducing the magnitude of the vector w.

$$\text{Minimize}\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\zeta_i^- + \zeta_i^+) \tag{4}$$

$$\text{subject to } (w_i.x_i + b) - y_i < k + \zeta_i^+,\ i = 1, 2, \ldots, m$$

$$y_i - (w_i.x_i + b) \le k + \zeta_i^-,\ i = 1, 2, \ldots, m$$

$$\zeta_i^-, \zeta_i^+ \geq 0 \tag{5}$$

In the SVR optimization framework, C is the penalty coefficient, m refers to the number of training or calibration data points which indicates the sample size, and for the i-th training data point, ξ_i^- and ξ_i^+ depict the deviations that lie below and above the range (−k, +k) respectively. Meanwhile, for each point in the training dataset, w_i corresponds to the weight, x_i is the input variable's value, and y_i is the observed output value. The central goal of SVR is adjusting these parameters for optimal data fitting while minimizing predictive inaccuracies.

The decision variables for Eqs. (4) and (5) are denoted as w and b. To put it another way, these variables w and b are determined after the SVR training phase is finished. These calculated values of w and b are then integrated into Eq. (2) to derive predictions f(x) based on the given input variables x. While Eq. (2) showcases a linear regression approach, SVR can be extended to accommodate nonlinear relationships using various kernel functions. The nonlinear regression representation of SVR, when using the kernel, can be described as:

$$f(x) = w^{Tr}.k(X, x_i) + b \tag{6}$$

$$k(X, x_i) = \exp\left(-\frac{|X - x_i|^2}{2\gamma^2}\right) i = 1, 2, ..., m \tag{7}$$

In the context of the nonlinear SVR, K(X, xi) represents the kernel function. The main objective of SVR is to accurately determine the parameter values for k, C, and the kernel function's parameter, γ.

2.2.3. Grey Wolf Optimizer (GWO)

The Grey Wolf Optimizer (GWO) is an optimization algorithm rooted in nature, introduced by Mirjalili et al. in 2014[53]. Drawing inspiration from the social structure and predatory behavior of wolves, the algorithm simulates how an alpha wolf guides its pack during the hunting stages of tracking, encircling, and pursuing prey up to the point of attack. What makes GWO stand out is its adeptness at harmonizing localized searching with overarching optimization, a feat attributed to its dynamically adjusting convergence factor. This particular quality underscores its robust convergence capabilities. Furthermore, as [54], GWO boasts a straightforward design, requires fewer tuning parameters, and possesses a potent search prowess.

In the given mathematical framework, four distinct tiers are symbolized by diverse wolves, each pursuing its unique solution and having a specific role. At the pinnacle of this structure is the Alpha wolf (α), embodying the best global solution and overseeing decisions. Next in line is the Beta wolf (β), which signifies a less-than-perfect solution. The third position is held by the Delta wolf (δ), taking directives from both Alpha and Beta wolves and being in charge of tracking and pursuit. At the base is the Omega wolf (ω), assisting the aforementioned wolves, crucial for ensuring equilibrium within the group. The detailed procedures are outlined as follows.

(1) The wolves strategically position themselves to close in on their target from different angles, and they adjust the location of their pack periodically. The subsequent equations showcase this encircling mechanism within the grey wolf optimization process.

$$X(t+1) = X_p(t) - A.D \tag{8}$$

$$D = |C.X_P(t) - X(t)| \tag{9}$$

$$A = 2a.r_1 - a \tag{10}$$

$$C = 2.r_2 \tag{11}$$

Where Xp represents the position vectors of the prey, and X symbolizes the position vectors of the wolf pack. D stands for the interaction between them. C and A are coefficient vectors. r_1 and r_2 are vectors generated randomly within the range [0, 1]. To promote convergence, a is systematically reduced as the number of iterations increases.

(2) The mathematical representation illustrating how the wolves (ω) update their positions in relation to the top three optimal solutions, simulating the tracking of prey, is detailed below:

$$D_\alpha = |C_1.X_\alpha(t) - X| \tag{12}$$

$$D_\beta = |C_2.X_\beta(t) - X| \tag{13}$$

$$D_\delta = |C_3.X_\delta(t) - X| \tag{14}$$

Where D_α, D_β, and D_δ represent the distances between the first three wolves and the rest of the pack. X_α, X_β, and X_δ denote the present positions of the first three wolves, respectively. signifies the current position of the entire pack. C_1, C_2, and C_3 are randomly generated vectors.

(3) Subsequently, each wolf's position is adjusted dynamically by tracking the movements of the leading wolves, highlighting the evolution of the wolf pack's position. Additionally, as the wolves draw nearer to the prey, the value of a systematically drops to 0. Concurrently, the spectrum of A is confined within the interval [-a, a], aiding the wolf in ascertaining the precise location of the prey for a targeted strike.

$$X_1 = X_\alpha - A_1.D_\alpha \tag{15}$$

$$X_2 = X_\beta - A_2.D_\beta \tag{16}$$

$$X_3 = X_\delta - A_3.D_\delta \tag{17}$$

$$X_{t+1} = \frac{X_1 + X_2 + X_3}{3} \tag{18}$$

2.2.4. Slime Mould Algorithm (SMA)

A recent addition to the suite of nature-inspired metaheuristics is the SMA [55]. This method draws inspiration from the perturbation behavior exhibited by slime moulds. More details about this algorithm will be elaborated upon in the subsequent subsections.

Approach Food.

The initial phase of the SMA involves an inventive method to approach food. This step is typified by the slime mould being drawn towards food due to the attraction of its scent in the ambiance. This behavior can be described as:

$$X_n^g = \begin{cases} x_b + F_b.(\omega.x_{a1} - x_{a2}) \rho < pr \\ F_c.x_n^g \ \rho \geq pr \end{cases} \tag{19}$$

Where x_b represents the best solution obtained so far and F_b is a parameter that lies in the interval [−A, A]. The term pr stands for a probability rate, and F_c diminishes linearly from 1 down to 0. The symbol ρ embodies a random value selected within the range (0, 1). The variable g refers to the current generation, while x_n^g marks the position of the nth slime mould during the gth generation. Both x_{a1} and x_{a2} represent two randomly chosen solutions from within the current population. Lastly, ω is indicative of the slime mould's weight. The formulation for pr is provided below:

$$pr = \tanh|F(x_n^g) - F(x_b)| \tag{20}$$

where n takes values from the set {1, 2, …, N}. $F(x_n^g)$ represents the fitness function evaluated at (x_n^g), and $F(x_b)$ denotes the fitness function evaluated at x_b. The expression for μ_b is detailed next:

$$F_b = [-A, A] \tag{21}$$

$$A = \arctanh\left(-\left(\frac{g}{Maxg}\right)+1\right) \tag{22}$$

$$\omega(SlnD(n)) = \begin{cases} 1+\rho.\log\left(\frac{F(x_b)-F(n)}{F(x_b)-F(x_\omega)}+1\right) ConD \\ 1-\rho.\log\left(\frac{F(x_b)-F(n)}{F(x_b)-F(x_\omega)}+1\right) OS \end{cases} \tag{23}$$

$$SlnD = sort(F) \tag{24}$$

Where ConD signifies that F(k) falls within the top half of the population in terms of ranking. The term $F(x_w)$ designates the least favorable objective function value obtained during the entire optimization process. Meanwhile, SInD represents the ordered sequence of objective function values.

Wrap Food.

The subsequent mechanism pertains to the method by which the food is enveloped. This segment mathematically emulates the relationship between the width of the vein and the concentration of food. When concentrations are on the lower side, other regions are explored. This strategy is employed to refresh the position of the slime mould and is articulated as:

$$X = \begin{cases} rand.(U-L)+Lrand < S \\ x_b + F_b.(\omega.x_{a1} - x_{a2})\rho < pr \\ F_c.x_n^g\rho \geq pr \end{cases} \tag{25}$$

Where L and U denote the lower and upper bounds of the solution space, respectively. Both rand and S represent a random number within the range [0,1].

Grabble Food.

To simulate the variability in the vein width of the slime mould, the factors ω, F_b, and F_c are employed. The variable ω gauges the quality of food, emulating the oscillation frequency of the slime mould. This allows the modulation of the pace at which food is processed, enabling the slime mould to opt for a suitable food source. Throughout the experiment, F_b and F_c undergo random fluctuations within specific bounds. F_b oscillates within the interval [−A, A], and F_c does so within the interval [−1, 1]. As the optimization procedure progresses, both converge towards a value of zero. The variations in F_b also mirror the behavior of the slime mould when deciding whether to move closer to a current food source or seek out new ones upon the discovery of fresh food[56].

2.2.5. Aquila Optimizer (AO)

The AO optimization algorithm draws inspiration from the predatory behaviors exhibited by the Aquila, a majestic bird of prey. This optimization technique is segmented into four phases: first, it delineates the search space via vertical bending; next, it delves into a broad search space through short glides; it then narrows its focus in the search space by gradually descending; and finally, it executes a raid by walking followed by the act of predation. Importantly, the AO algorithm is adept at defining an effective search area during the optimization, whether operating within a vast or a limited view. The Aquila's quartet of hunting strategies is detailed in reference [57].

Extended Exploration (X_1): In this initial hunting approach, the Aquila employs a combination of high-altitude and vertical flight techniques to pinpoint the vicinity of its prey and to cherry-pick the most favorable hunting grounds. This specific behavior is mathematically represented by Eq. (26).

$$X_1 = X_{best}(t) \times \left(1-\frac{t}{T}\right) + (X_M(t) - X_{best}(t) \times rand) \tag{26}$$

In the given context, $X_{best}(t)$ represents the optimal solution before the t^{th} iteration, serving as an indicator of the estimated location of the prey. The variable Rand corresponds to a random value within the range of 0 to 1. The iteration count is denoted by t, representing the current step in the process, and T signifies the maximum allowable iterations.

Narrow Exploration (X_2): In this secondary hunting method, once the prey is detected from a lofty elevation, the Aquila lingers overhead, closely scrutinizing the designated vicinity of the targeted prey prior to initiating an assault. This strategy is termed the contour flight short glide attack, and its mathematical representation is given by Eq. (27).

$$X_2 = X_{best}(t) \times Levy(D) + X_R(t) + (y - x) \times rand \tag{27}$$

Within this framework, D signifies the dimensional space, and Levy (D) encapsulates the Levy flight distribution function, defined by Eq. (28). During the Rth iteration, $X_R(t)$ denotes a solution randomly selected from the range [1, N].

$$Levy(D) = s \times \frac{u \times \sigma}{|\nu|^{\frac{1}{\beta}}} \tag{28}$$

In Eq. (28), y and x are employed to depict the spiral shape during the search process, and they are determined as follows:

$$y = r \times \cos(\theta), x = r \times \sin(\theta) \tag{29}$$

$$r = r_1 + U \times D_1, \theta = -\omega \times D_1 + \frac{3\pi}{2} \tag{30}$$

Expanded Development (X_3): In this tertiary hunting approach, once the prey's location is pinpointed with precision and the Aquila is poised to swoop down and strike, it descends vertically, executing an attack maneuver to gauge the prey's reaction. Termed the low-flying slow-descent attack, this tactic involves the eagle strategically navigating toward the prey by exploiting the designated target area before executing its assault. The mathematical articulation of this behavior is precisely captured in Eq. (31).

$$X_3 = (X_{best}(t) - X_M(t)) \times \alpha - rand + ((UB - LB) \times UB + LB) \times \delta \tag{31}$$

In this context, LB stands for the lower boundary of the problem, delineating the minimum allowable value, while UB refers to the upper boundary, indicating the maximum permissible value within the defined problem parameters.

Reduced Exploitation (X4): In this fourth hunting methodology, as the Aquila nears its target, it engages and attacks the prey while walking on the ground, adapting to the prey's random movements. The mathematical depiction of this specific behavior can be found in Eq. (32).

$$X_4 = QF \times X_{best} - (G_1 \times X(t) \times rand) - G_2 \times Levy(D) + rand \times G_1 \tag{32}$$

In the given scenario, QF serves as the key function integral to the equilibrium search strategy, and its formulation is precisely conveyed as follows:

$$QF(t) = t^{\frac{2\times rand-1}{(1-T)^2}} \tag{33}$$

G_1 signifies the diverse maneuvers the Aquila employs while tracking prey that is attempting to escape. It is represented as:

$$G_1 = 2 \times rand - 1 \tag{34}$$

The diminishment of the G2 value from 2 to 0 signifies the gradient of the Aquila's flight as it traverses from its initial position (1) to its ultimate position (t) during the pursuit of the prey. This dynamic progression is succinctly represented as:

$$G_2 = 2 \times \left(1 - \frac{t}{T}\right) \tag{35}$$

2.2.6. Delta Moment independent index method

In recent years, global sensitivity analysis has become increasingly

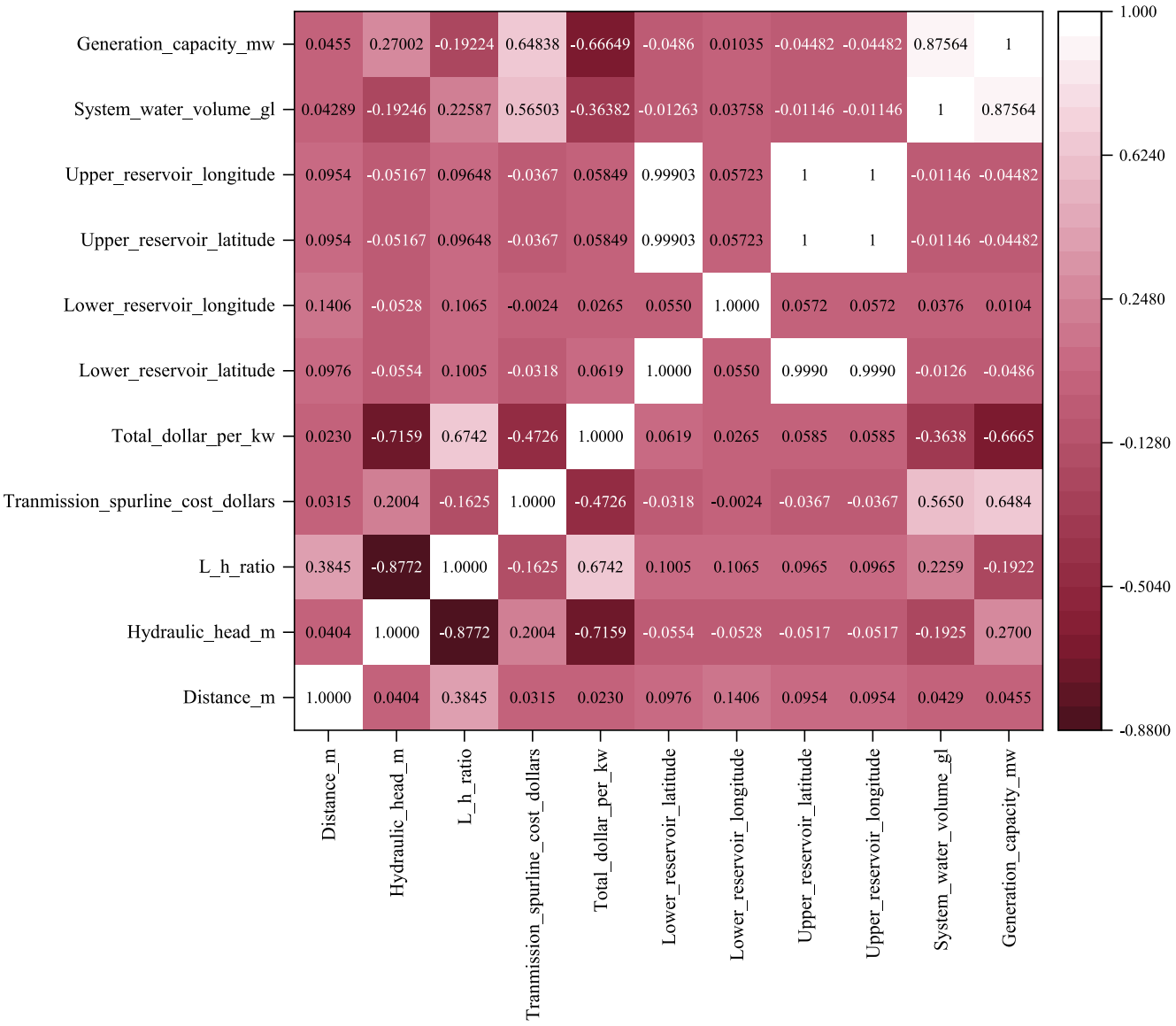


Fig. 2. The correlation matrix of features.

important, especially in environmental science, for assessing scientific models. Traditional global sensitivity analysis methods primarily focus on assessing the average contributions of input variables to the uncertainty of a model's output. However, these methods often overlook the specific effects of input variable ranges. To address this limitation, the Delta Moment-Independent method has gained attention.

The Delta Moment-Independent method is designed to determine how a specific range of an individual input variable influences the uncertainty in a model's results. It achieves this by comparing the cumulative distribution functions (CDF) of the model's output when all variables are considered versus when a particular input variable, denoted as xi, remains unchanged [58,59]. The sensitivity index (δ) for the Delta Moment-Independent approach can be calculated as follows:

$$S_i = \delta_i = \frac{1}{2} E_{X_i} \left| f_y(y) - f_{y|x_i}(y) \right| dy \tag{36}$$

Where:

$f_y(y)$ represents the probability density function for the complete model output (y).

$f_{y|x}(y)$ indicates the conditional density of the output (y) when a specific input variable x_i is set to a definite value.

Essentially, the δ_i sensitivity index quantifies the standardized expected alteration in the output (y) distribution caused by variations in the input variable x_i. These moment-independent techniques are particularly useful when the objective is to understand the full spectrum of effects arising from input fluctuations. They are valuable in scenarios where uncertain parameters can lead to rare yet impactful events within a given system [60].

The study implemented machine learning models and metaheuristic algorithms employing Python software, a versatile programming language widely used for data analysis and machine learning tasks. Key Python packages utilized in this research included Scikit-learn for implementing the SVR and XGBoost algorithms, and XGBoost library for model optimization. Metaheuristic optimizations, such as AO, GWO, and SMA, were coded employing custom implementations or imported from specialized Python packages that support nature-inspired optimization methods.

In this study, key parameters of SVR and XGBoost were optimized to improve model performance. For SVR, the main parameters fine-tuned

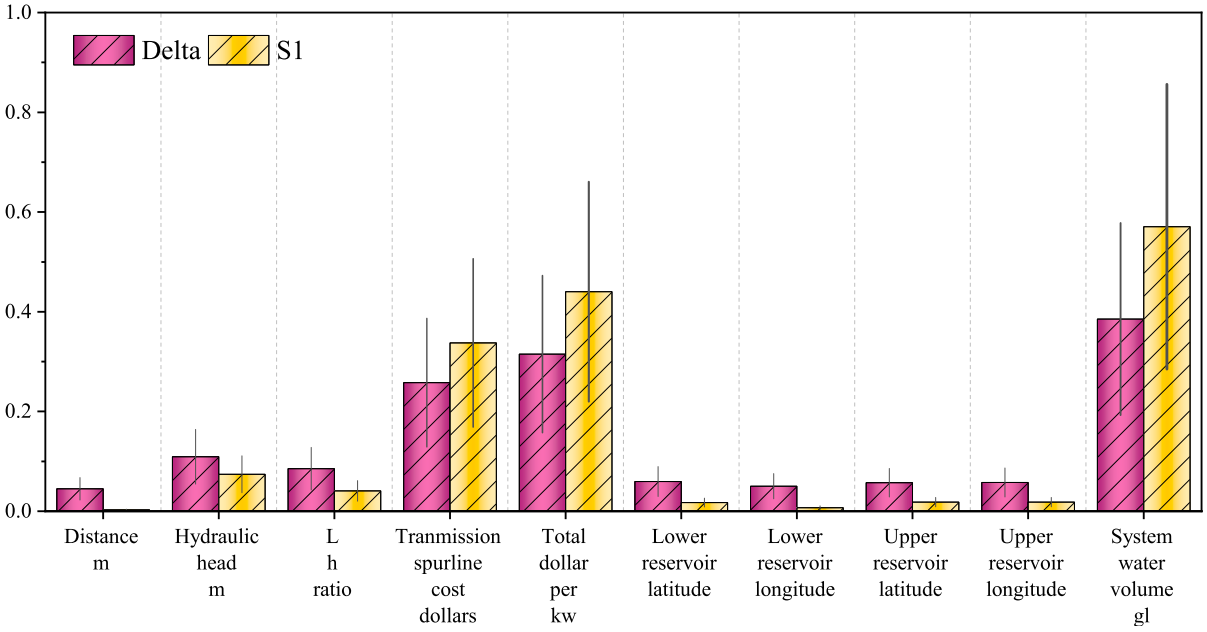


Fig. 3. Sensitivity analysis of variables based on the DMIM method.

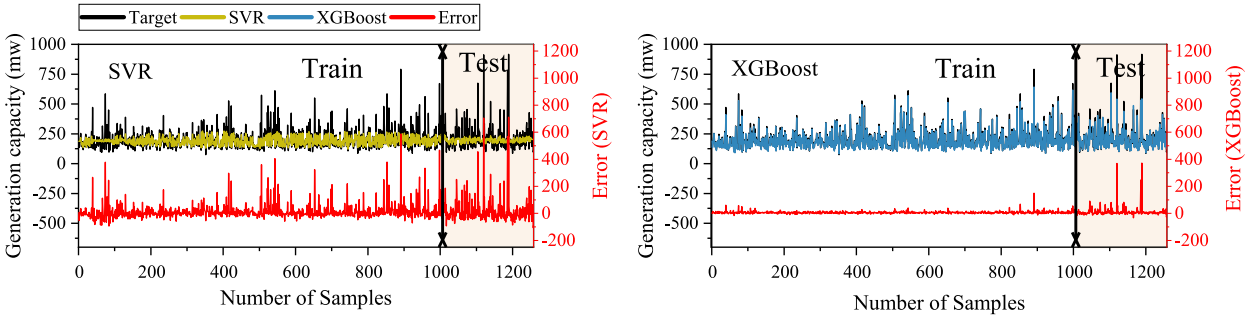


Fig. 4. Time series of actual and predicted based on XGBoost and SVR models.

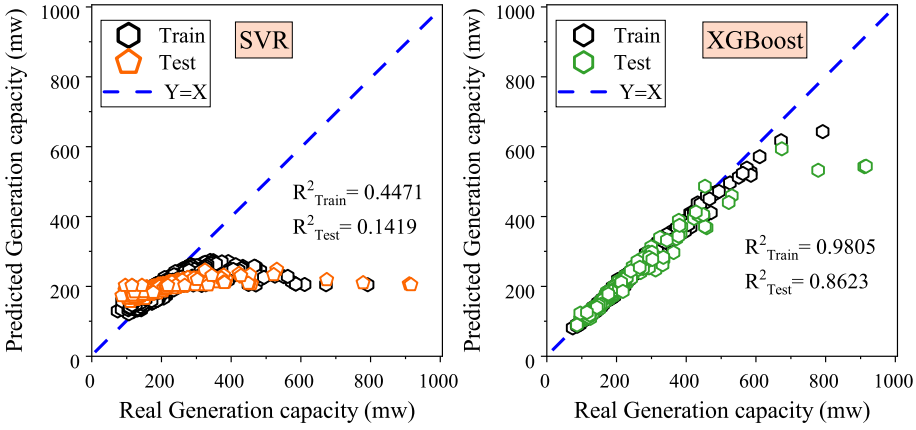


Fig. 5. Scatter plot of the observation-prediction for XGBoost and SVR models.

included the penalty coefficient, which balances minimizing training error and model complexity, with a search range of [0.1, 1000]; epsilon, which defines the error tolerance margin, with a range of [0.001, 1]; and the kernel function parameters, with a range of [0.0001, 1]. For XGBoost, the parameters adjusted were the learning rate, which determines the step size to minimize the loss function and ranges in [0.01, 0.3]; the maximum depth of the trees, set within [3,10]; the number of estimators, which outlines the boosting rounds, with a range of [50, 500]; and subsample, which refers to the proportion of samples employed for training each tree, set in the spectrum of [0.5, 1].

3. Results and discussion

In this section, a comprehensive analysis of methods used to predict

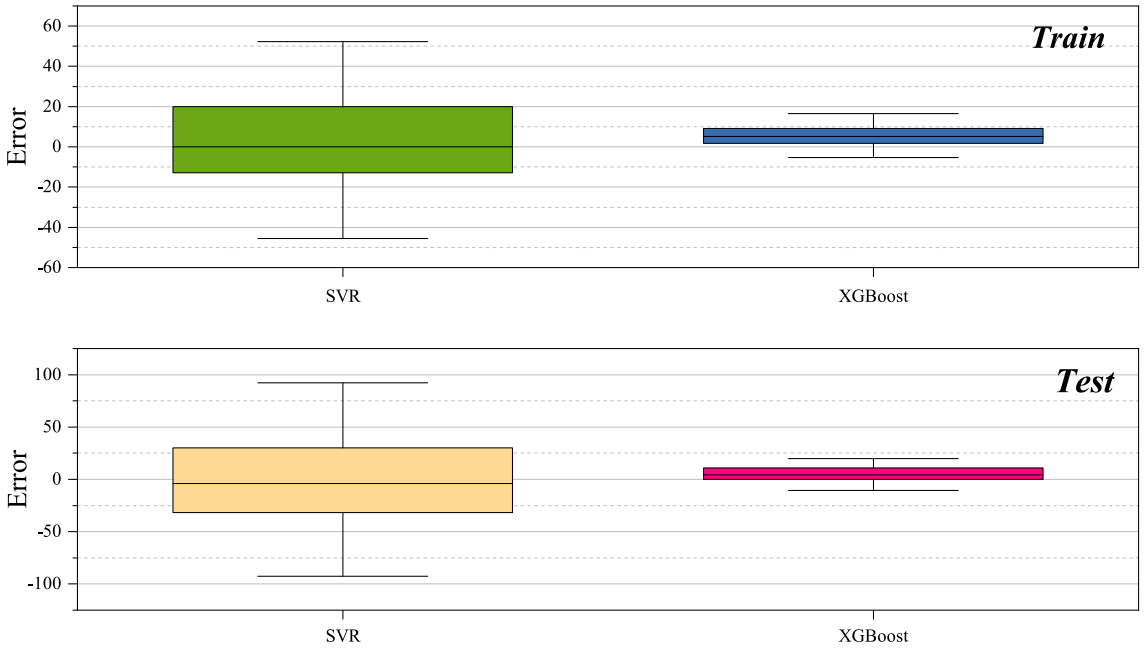


Fig. 6. Box plots of error measurements for XGBoost and SVR models during the testing and training phases.

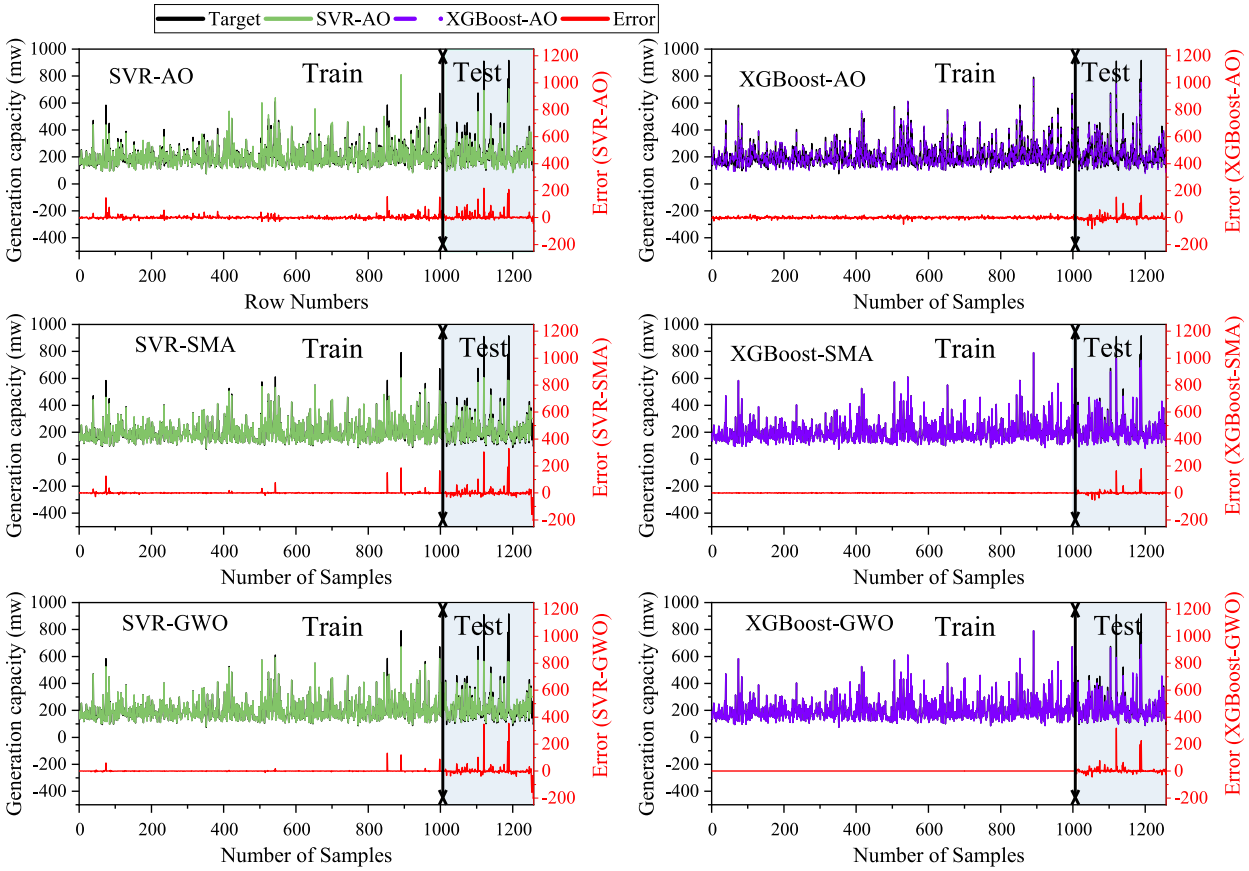


Fig. 7. Time series of actual and predicted based on XGBoost and SVR hybrid models.

the output of HP is presented. Specifically, the focus is on two algorithms: XGBoost and SVR. Both of these algorithms are evaluated individually and in a combined approach. The results of these tests are illustrated through various diagrams and tables. Based on this visual and tabulated data, a deep assessment of the efficiency of these algorithms is conducted, leading to the determination of the best-performing method for various locations.

Fig. 2 illustrates a correlation matrix created based on input and output parameters. According to this figure, the study encompasses ten inputs and one output, namely 'generation capacity.' This figure depicts

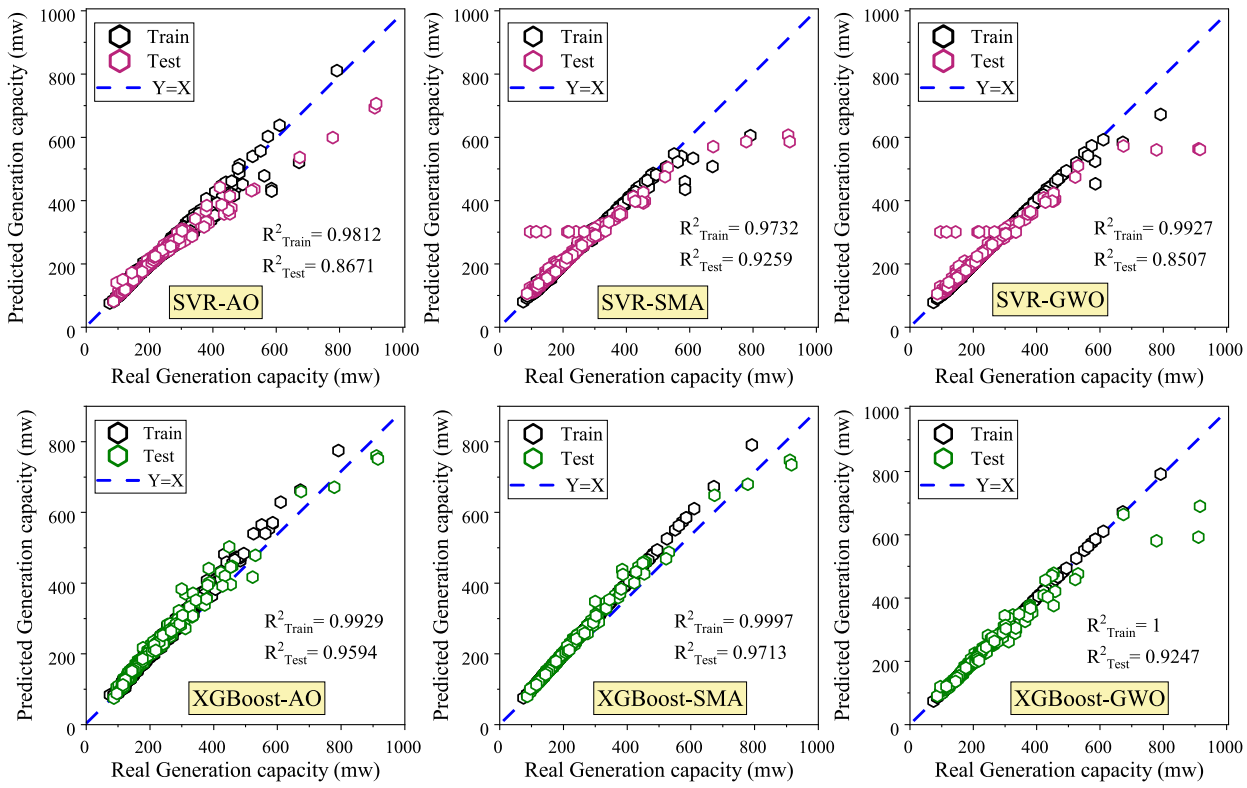


Fig. 8. Scatter plot of the observation-prediction for XGBoost and SVR hybrid models.

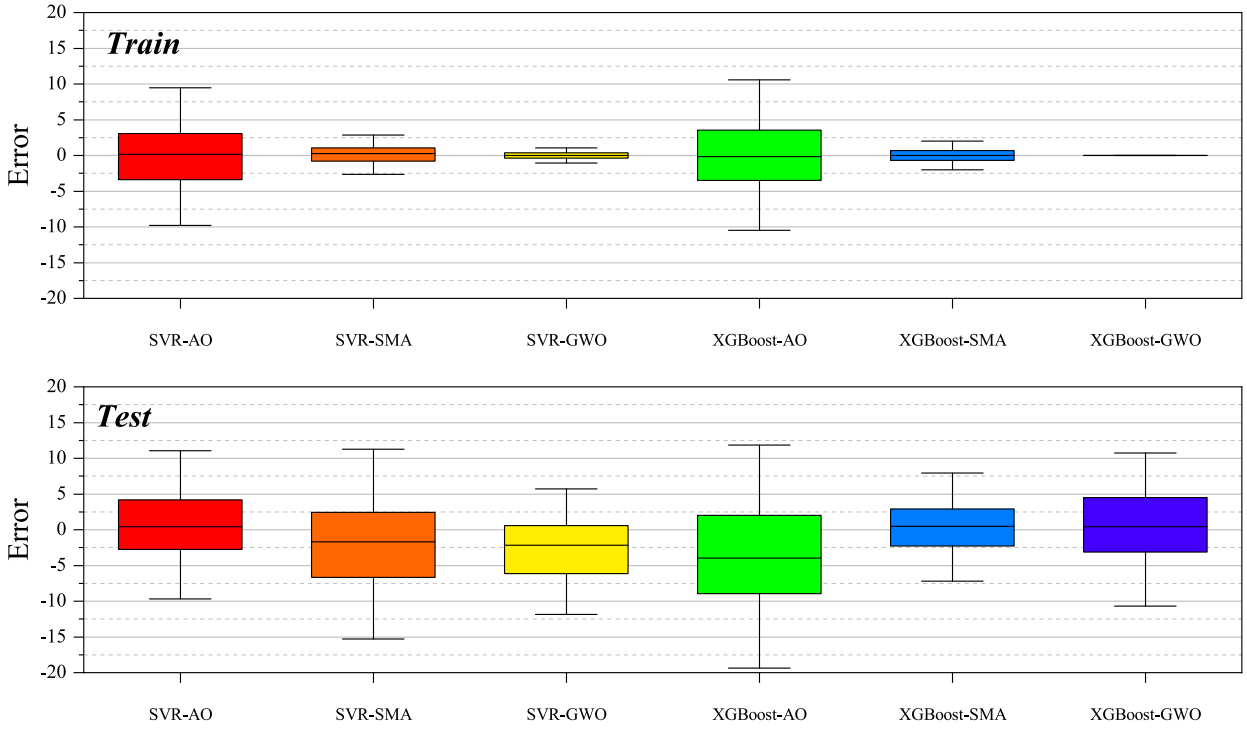


Fig. 9. Box plots of error measurements for XGBoost and SVR hybrid models during the testing and training phases.

the relationships between input parameters and their associations with the output. The matrix values range from −1 to + 1, where positive values indicate a positive correlation and a direct influence, while negative values signify a negative correlation and an inverse impact. According to Fig. 2, the parameter 'system water volume' exhibits the most significant positive influence and a direct relationship with the output. The next influential parameter is 'transmission_spurline_cost.' Even parameters like 'hydraulic head' display a slight positive correlation. However, the parameter 'total_dollar_per_kw' demonstrates the most substantial negative correlation. The remaining parameters have a

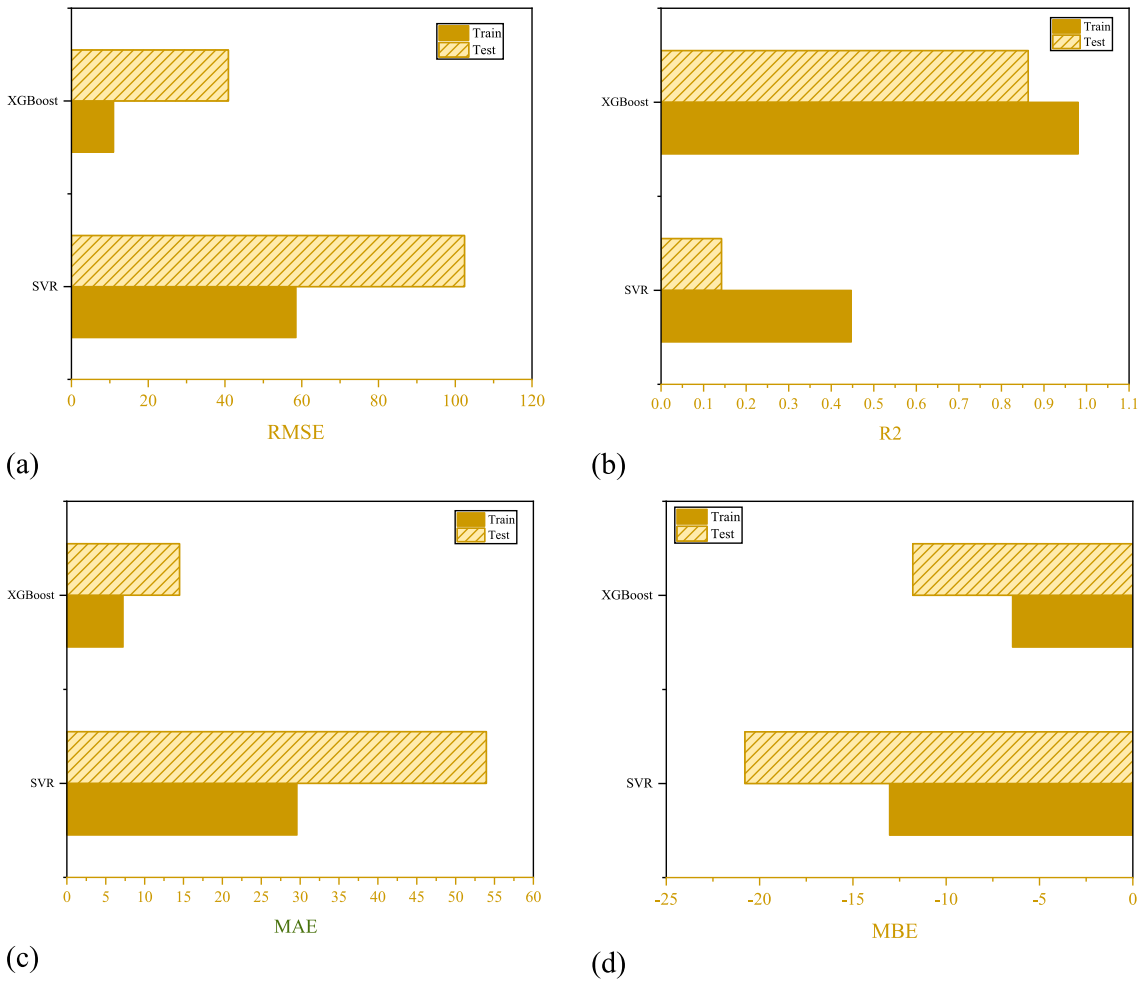


Fig. 10. Error metrics plots for proposed single models.

relatively minimal impact on the output and are predominantly neutral in their associations.

Fig. 3 presents a detailed sensitivity analysis of ten input variables, employing the Delta Moment independent index. In Fig. 3, S1 symbolizes one of the key input variables employed in predicting hydropower generation. The sensitivity analysis, carried out using the Delta Moment Independent Index method, explores how each input affects the output. In this visualization, all input factors, such as S1, are scaled between 0 and 1. A score of 1 for S1 means it has the greatest effect on the predicted outputs, which indicates that variations in this input significantly affect the model's accuracy. On the other hand, if S1 has a score closer to 0, it would suggest that this parameter has a relatively minor impact on the prediction of hydropower generation. The sensitivity index quantifies the reduction in output fluctuations, while the delta index denotes the magnitude of distributional changes. Based on Fig. 3, the parameter 'system water volume' has exhibited the highest sensitivity to the output. Therefore, this parameter, along with 'tranmission_spurline_cost' and 'total_dollar_per_kw,' has demonstrated the most sensitivity. The parameter 'hydraulic head' also shows some sensitivity to the output, albeit to a lesser extent. However, the remaining parameters have shown nearly similar levels of sensitivity to the output based on the delta index.

Fig. 4 depicts a time series chart of observed and computed data based on the SVR and XGBoost algorithms. The error rate in these charts is highlighted in red for each algorithm. According to this figure, the XGBoost algorithm has exhibited a lower error rate and performed better in prediction. If you focus on the peak points in the charts, it is evident that the SVR algorithm did not perform well in this regard, and, in contrast, the XGBoost algorithm had better coverage of the peaks.

To assess the suitability and precision of the algorithms more accurately, scatter plots have also been provided. According to these plots, the algorithm is more accurate when the data points cluster closer to the x = y axis. In Fig. 5, it is evident that the data clustering in the scatter plot for SVR is not satisfactory, whereas the XGBoost algorithm is much more accurate. In addition to the plot, the R^2 coefficient has been calculated for both algorithms. According to this metric, the XGBoost algorithm, with an R^2 value of 0.8623, has performed better in prediction.

Fig. 6 displays Box plots related to the single algorithms. These plots allow us to observe the spread of errors and data variability. These plots have been prepared for both the training and testing datasets. As expected, based on Fig. 6, the range of errors and their variability in the XGBoost algorithm is much lower compared to SVR, indicating the suitable performance of the XGBoost algorithm.

To improve accuracy and compare the algorithms, hybrid models have been developed. For this purpose, three optimizers, AO, SMA, and GWO, have been employed. They have specific parameters that affect their optimization capabilities. For AO, key parameters comprise the maximum iterations, which control the exploration and exploitation process, the search zone bounds to define the parameter search zone, and the convergence factor, which adjusts the gradient of Aquila's flight pattern for improved convergence. In GWO, the parameters comprise the number of wolves that dictate the population size, maximum iterations, coefficient vectors that simulate social behavior and guide the wolves' movement, and a convergence factor that refines the search from exploration to exploitation over time. SMA employs a weight

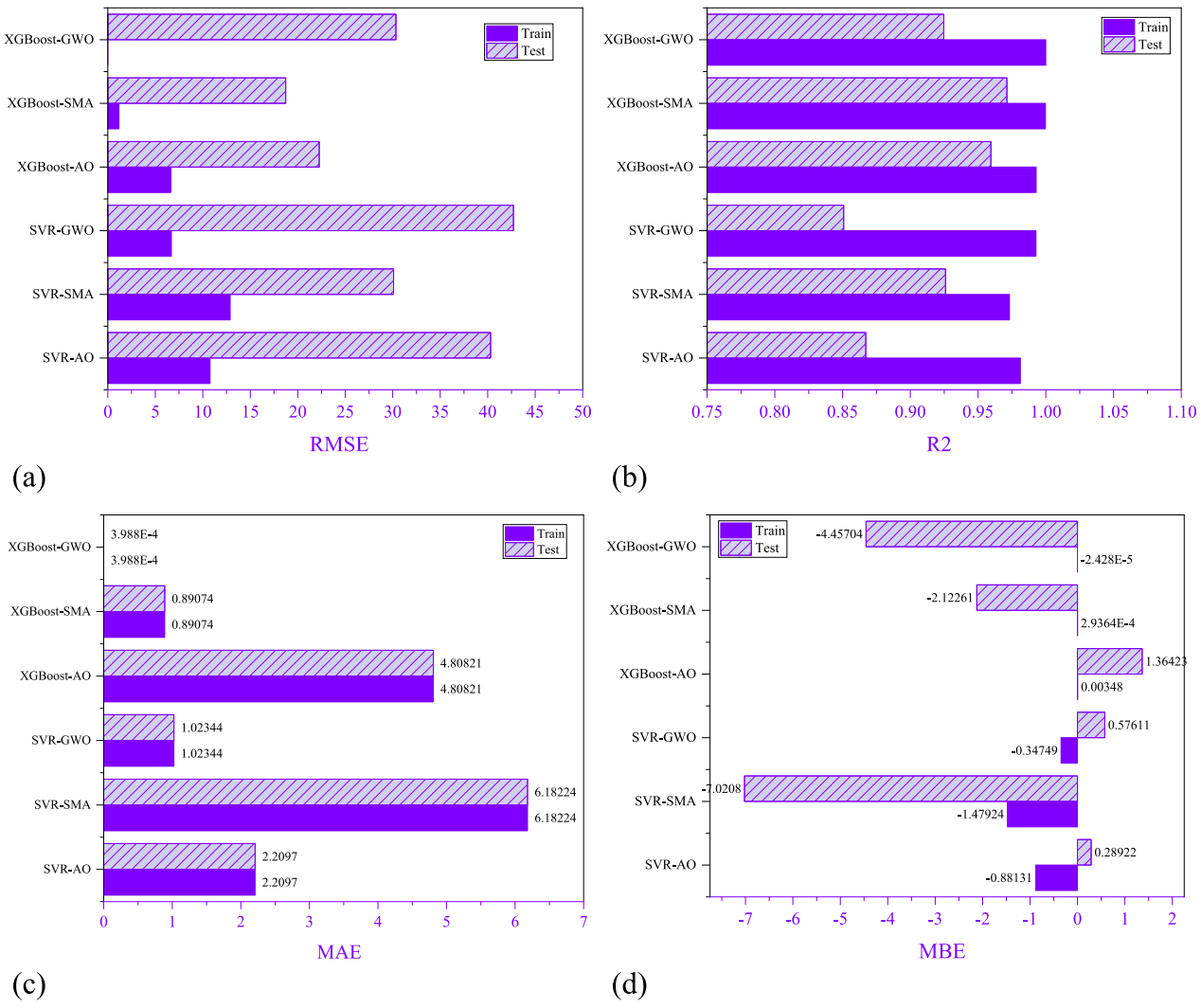


Fig. 11. Error metrics plots for proposed hybrid models.

Table 2
Error metrics for proposed XGBoost and SVR models.

Optimizer	SVR	XGBoost
MBE	−13.03609978	−6.4467
MAE	29.59052496	7.22564
RMSE	58.51083131	10.999
R^2	0.44712319	0.98046
DRV	0.93592913	0.96832
A10	0.56262425	0.97813
MBE	−20.79277809	−11.797
MAE	53.95582412	14.4936
RMSE	102.4276368	40.9032
R2	0.14199745	0.86317
DRV	0.90436762	0.94574
A10	0.31746032	0.89286

coefficient to imitate food quality, a probability rate that balances exploration and exploitation, bounds for defining the search zone, and a contraction–expansion coefficient to simulate the slime mould's movement treatment. These parameters collectively contribute to the optimization process, and improve the predictive accuracy of SVR and XGBoost models.

Fig. 7 illustrates a time series chart of observed and computed data for the hybrid models of XGBoost and SVR. Based on Fig. 7, the significant impact of the optimizers on the SVR models is evident. The accuracy of SVR models has greatly improved, underscoring the importance of optimizers. However, according to Fig. 7, the hybrid XGBoost models still perform better, with lower error rates, although the difference in results has significantly diminished.

Fig. 8 presents scatter plots for the hybrid XGBoost and SVR models. The R^2 coefficient has also been provided for both the test and training datasets. Based on Fig. 8, the data points cluster more closely to the XGBoost plots, indicating better performance. The R^2 coefficient further confirms this. According to this metric, the XGBoost-SMA model, with an R^2 value of 0.9713, has shown the best performance. Additionally, the SVR-GWO model, with an R^2 value of 0.8507, exhibited the weakest performance.

Fig. 9 displays Box plots for the hybrid models for both the training and testing datasets. Based on this figure, in the training dataset, the models optimized with SMA exhibit a higher degree of error variability and a wider error range, while the other models show less error variability. The situation is different in the testing dataset. The XGBoost-SMA algorithm has the widest error range, and the XGBoost-AO algorithm has the narrowest error range.

To enable a precise comparison and assessment of the algorithms mentioned, whether they have been used in hybrid or standalone models, a range of statistical metrics has been utilized. This section collectively presents these metrics to distinctly assess their performance. Fig. 10 presents error metrics for single XGBoost and SVR models. According to Fig. 10a, which shows model errors based on the RMSE index,

Table 3
Error metrics derived from the application of XGBoost and SVR hybrid models.

Optimizer	SVR SVR-AO	SVR-SMA	SVR-GWO	XGBoost XGBoost-AO	XGBoost- SMA	XGBoost-GWO
MBE	−0.88131	−1.47924	−0.34749	0.00348231	0.00029364	−0.00002428
MAE	2.209698	6.182242	1.023442	4.80820987	0.89073695	0.0003988
RMSE	10.76845	12.88072	6.703906	6.61372041	1.17388515	0.00057667
R^2	0.981273	0.973206	0.992742	0.99293605	0.99977746	1
DRV	0.995668	0.99273	0.998292	1.00001712	1.00000144	0.99999988
A10	0.992048	0.975149	0.996024	0.98707753	1	1
MBE	0.289222	−7.0208	0.576112	1.36423184	−2.1226139	−4.45703691
MAE	14.36013	11.97018	14.2013	11.49694462	6.18592276	9.77016551
RMSE	40.30475	30.08921	42.71388	22.26194985	18.72612019	30.33942213
R2	0.867148	0.925958	0.850792	0.95946962	0.97132193	0.92472186
DRV	1.00133	0.967709	1.00265	1.00627452	0.99023745	0.97950072
A10	0.888889	0.896825	0.884921	0.91666667	0.97619048	0.9484127

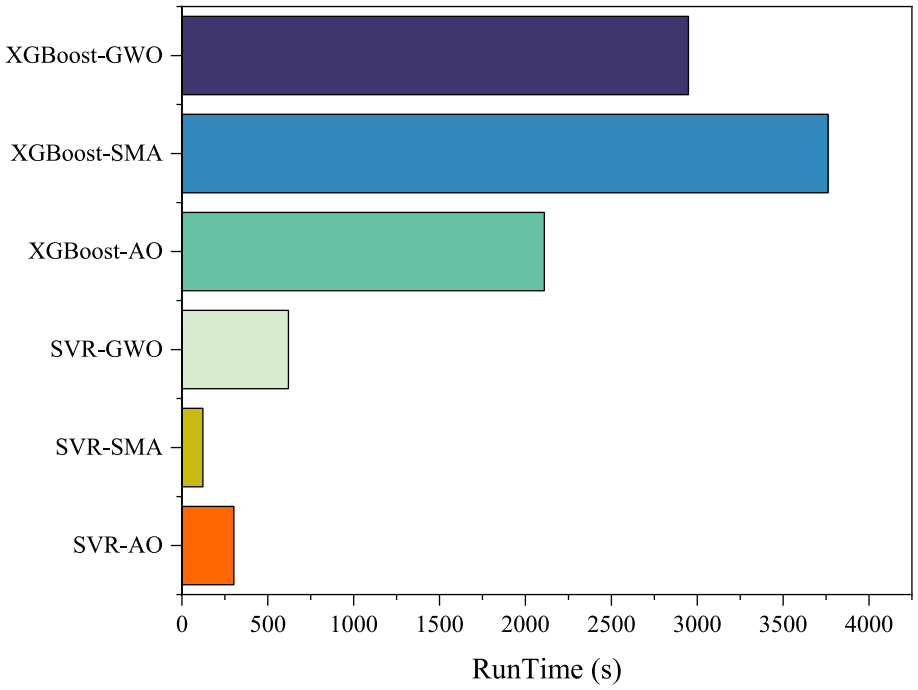


Fig. 12. Comparison of runtime for various hybrid models.

the XGBoost algorithm has lower errors and better performance. Additionally, Fig. 10b displays model errors based on the R^2 index, with XGBoost outperforming SVR in prediction, having a higher R^2 value. Similar results are seen for MAE and MBE indices, as shown in Fig. 10c and d.

Fig. 11 also illustrates error metrics for hybrid models. According to Fig. 11a, which is based on the RMSE index, the XGBoost-SMA hybrid model has the lowest error. This is consistent with Fig. 11-b, which relates to the R^2 index, where the XGBoost-SMA hybrid model exhibits the highest correlation. Fig. 11c and d, which correspond to MAE and MBE indices, show that the XGBoost-SMA hybrid model has the best predictive performance. In addition to the mentioned indices, the accuracy and performance of models have been evaluated based on other indices, including DRV and A10, the numerical results of which are provided in Tables 2 and 3 for both hybrid and single models.

Fig. 12 offers a graphical depiction of the execution times for the hybrid algorithms using bar charts. Based on the data presented in Fig. 12, it is clear that the SVR-SMA hybrid models typically exhibit shorter runtimes compared to all other hybrid models. The hybrid model XGBoost-SMA has had the longest runtime. In general, SVR models have exhibited shorter runtimes compared to the hybrid XGBoost models.

Fig. 13 showcases the convergence trajectories of the hybrid models throughout 500 iterations. Convergence is gauged using the Mean Squared Error (MSE) as the primary metric. Based on Fig. 13, the SVR-GWO model had the lowest Mean Squared Error (MSE), while the XGBoost-AO model had the highest MSE. Additionally, it's worth noting that most of the algorithms converged to a solution after a high number of iterations.

The presented study indicates plenty of advantages. It employs advanced machine learning methods, especially Support Vector Regression (SVR) and eXtreme Gradient Boosting (XGBoost), to address complex, nonlinear relations in hydropower prediction. The integration of hybrid models with optimization algorithms comprising Aquila Optimizer (AO), Grey Wolf Optimization (GWO), and Slime Mould Algorithm (SMA) significantly improves the accuracy of the predictions. Moreover, the detailed sensitivity evaluation utilizing the Delta Moment Independent Index identifies critical input parameters, which provides valuable insights for optimizing energy production. A comprehensive evaluation of diverse machine learning models also suggests a clear understanding of the most effective methodologies for hydropower prediction and underscores the potential of hybrid approaches, remarkably XGBoost-based models.

However, the study depicts some limitations. The models heavily depend on the quality and availability of input data, which makes them

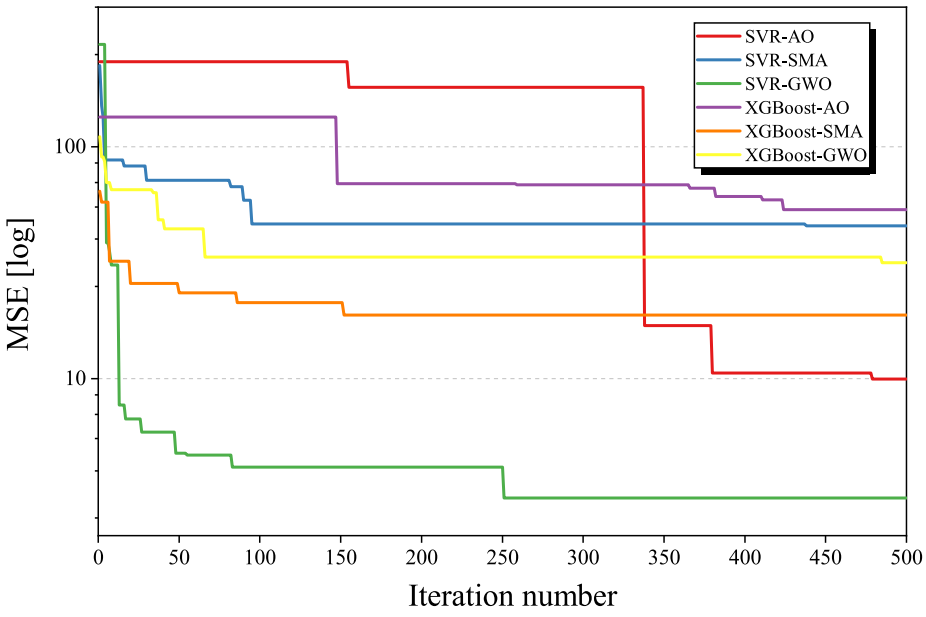


Fig. 13. The convergence plots of the XGBoost and SVR hybrid models.

sensitive to data inaccuracies. Although hybrid models and optimization algorithms were explored, the concentration on specific algorithms (SVR and XGBoost) may limit the exploration of other advanced methods, including deep learning, which could further improve predictive performance. The application of these models to a specific case study may also restrict their generalizability to diverse hydropower environments. Furthermore, the computational complexity and time required for running hybrid models, especially with extensive optimizations, may bring out challenges for real-time applications. Future research could address these limitations by incorporating additional datasets, exploring other machine learning algorithms, and developing real-time prediction systems for dynamic hydropower management.

4. Conclusion

In this study, a pioneering methodology is introduced for the prediction of hydropower capacity, leveraging advanced machine learning techniques. The research specifically delves into modeling the intricate relationships between meteorological, and climatic variables, and the production capacity of hydropower. Notably, the efficacy of the eXtreme Gradient Boosting (XGBoost) algorithm is explored, alongside the application of statistical learning methods such as Support Vector Regression (SVR). Moreover, the study investigates the performance of hybrid models, fine-tuned through optimization processes involving Grey Wolf Optimization (GWO), Slime Mould Algorithm (SMA), and the unique Aquila Optimizer (AO). The research employs a variety of machine learning techniques and optimization methods to enhance the accuracy of predictions, and it analyzes the performance of these models across different locations and scenarios. The findings of the study reveal several key insights:

- The XGBoost algorithm consistently outperforms the SVR algorithm in predicting hydropower generation. It exhibits lower errors, higher correlation coefficients (R^2), and better accuracy across various metrics.
- Hybrid models, particularly those utilizing optimization techniques like AO, SMA, and GWO, improve the accuracy of predictions. These models enhance the performance of both SVR and XGBoost, with the XGBoost-SMA hybrid model showing the best performance.
- Sensitivity analysis highlights the importance of certain input parameters, such as 'system water volume,' in influencing hydropower generation. Understanding these influential factors is crucial for optimizing energy production.
- The study offers valuable insights into the relationships between input parameters and their impact on hydropower generation, providing a foundation for better system management and decision-making.
- The article presents comprehensive data on the performance of machine learning models, including single algorithms and hybrid models, in predicting hydropower generation. These results can guide decision-makers in selecting the most suitable approach for their specific requirements. Overall, this research underscores the potential of machine learning techniques in improving the efficiency and reliability of hydropower generation predictions. It offers valuable contributions to the field of renewable energy and provides a roadmap for enhancing the utilization of hydropower resources in the transition to more sustainable and environmentally friendly energy production. As the demand for clean energy continues to grow, the application of machine learning in this domain becomes increasingly relevant and promising.

CRediT authorship contribution statement

Zhenya Qi: Writing – original draft, Project administration, Conceptualization. **Yudong Feng:** Funding acquisition, Data curation. **Shoufeng Wang:** Resources, Methodology. **Chao Li:** Validation, Software.

Funding

Not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Not applicable.

References

[1] Noorzadeh S, Yari M, Mahmoudi SMS. Assessing the thermoeconomic performance of a solar-powered trigeneration system with an upgraded transcritical carbon dioxide unit. Process Saf Environ Prot 2024. https://doi.org/10.1016/j.psep.2024.09.024.

[2] Landis DA, Gratton C, Jackson RD, Gross KL, Duncan DS, Liang C, et al. Biomass and biofuel crop effects on biodiversity and ecosystem services in the North Central US. Biomass Bioenergy 2018;114:18–29.

[3] Shaosen S, Chen D, Srinivasan K, Chen B, Meijuan X, Garg A, et al. Experimental and artificial intelligence for determination of stable criteria in cyclic voltammetric process of medicinal herbs for biofuel cells. Int J Energy Res 2019;43:5983–91.

[4] Arun M, Barik D, Chandran SSR, Govil N, Sharma P, Yunus Khan TM, et al. Twisted helical Tape's impact on heat transfer and friction in zinc oxide (ZnO) nanofluids for solar water heaters: Biomedical insight. Case Stud Therm Eng 2024;56:104204. https://doi.org/10.1016/j.csite.2024.104204.

[5] Dmitrieva K. Forecasting of a hydropower plant energy production 2015.

[6] Abd Aziz AU, Ammarullah MI, Ng BW, Gan H-S, Abdul Kadir MR, Ramlee MH. Unilateral external fixator and its biomechanical effects in treating different types of femoral fracture: A finite element study with experimental validated model. Heliyon 2024;10:e26660. https://doi.org/10.1016/j.heliyon.2024.e26660.

[7] Babu ER, Reddy NC, Babbar A, Chandrashekar A, Kumar R, Bains PS, et al. Characteristics of pulsating heat pipe with variation of tube diameter, filling ratio, and SiO2 nanoparticles: Biomedical and engineering implications. Case Stud Therm Eng 2024;55:104065. https://doi.org/10.1016/j.csite.2024.104065.

[8] Nagaraja S, Anand PB, Mohan Kumar K, Ammarullah MI. Synergistic advances in natural fibre composites: a comprehensive review of the eco-friendly bio-composite development, its characterization and diverse applications. RSC Adv 2024;14:17594–611. https://doi.org/10.1039/D4RA00149D.

[9] Muchammad M, Tauviqirrahman M, Ammarullah MI, Iqbal M, Setiyana B, Jamari J. Performance of textured dual mobility total hip prosthesis with a concave dimple during Muslim prayer movements. Sci Rep 2024;14:916. https://doi.org/10.1038/s41598-023-50887-7.

[10] Ammarullah MI, Hartono R, Supriyono T, Santoso G, Sugiharto S, Permana MS. Polycrystalline diamond as a potential material for the hard-on-hard bearing of total hip prosthesis: von mises stress analysis. Biomedicines 2023;11:951. https://doi.org/10.3390/biomedicines11030951.

[11] Turgeon K, Trottier G, Turpin C, Bulle C, Margni M. Empirical characterization factors to be used in LCA and assessing the effects of hydropower on fish richness. Ecol Indic 2021;121:107047.

[12] Bernardes Jr J, Santos M, Abreu T, Prado Jr L, Miranda D, Julio R, et al. Hydropower operation optimization using machine learning: a systematic review. AI 2022;3:78–99.

[13] Farooq MU, Anwar S, Bhatti HA, Kumar MS, Ali MA, Ammarullah MI. Electric discharge machining of Ti6Al4V ELI in biomedical industry: Parametric analysis of surface functionalization and tribological characterization. Materials 2023;16:4458. https://doi.org/10.3390/ma16124458.

[14] Ammarullah MI, Hidayat T, Lamura MD, Jamari J. Relationship between deformation and running-in wear on hard-on-hard bearings from metal, ceramic, and diamond materials for total hip prosthesis. J Tribol 2023;38:69–81.

[15] Condemi C, Casillas-Perez D, Mastroeni L, Jiménez-Fernández S, Salcedo-Sanz S. Hydro-power production capacity prediction based on machine learning regression techniques. Knowl Based Syst 2021;222:107012.

[16] Santoso G, Ammarullah MI, Sugiharto S, Hidayat T, Khoeron S, Bayuseno AP, et al. TRIZ-based method for developing a conceptual laparoscopic surgeon's chair. Cogent Eng 2024;11. https://doi.org/10.1080/23311916.2023.2298786.

[17] Hidayat T, Ismail R, Tauviqirrahman M, Saputra E, Ammarullah MI, Lamura MDP, et al. Running-in behavior of dual-mobility cup during the gait cycle: A finite element analysis. Proc Inst Mech Eng H 2024;238:99–111. https://doi.org/10.1177/09544119231216023.

[18] Ekanayake P, Wickramasinghe L, Jayasinghe JMJW, Rathnayake U. Regression-based prediction of power generation at samanalawewa hydropower plant in Sri Lanka using machine learning. Math Probl Eng 2021;2021:1–12.

[19] Lestari WD, Adyono N, Faizin AK, Haqiyah A, Sanjaya KH, Nugroho A, et al. Optimization of the cutting process on machining time of ankle foot as transtibial prosthesis components using response surface methodology. Results Eng 2024;21:101736. https://doi.org/10.1016/j.rineng.2023.101736.

[20] Ahmed H, Airoboman EA, Ibrahim SO. Assessment of run off river hydropower potential within river kaduna. In: 2020 6th IEEE International Energy Conference (ENERGYCon). IEEE; 2020. p. 509–10.

[21] Barros MTL, Tsai FTC, Yang S, Lopes JEG, Yeh WWG. Optimization of large-scale hydropower system operations. J Water Resour Plan Manag 2003;129:178–88.

[22] Bordin C, Skjelbred HI, Kong J, Yang Z. Machine learning for hydropower scheduling: State of the art and future research directions. Procedia Comput Sci 2020;176:1659–68.

[23] Plucinski B, Sun Y, Wang S-YS, Gillies RR, Eklund J, Wang C-C. Feasibility of multi-year forecast for the Colorado river water supply: time series modeling. Water (basel) 2019;11:2433.

[24] Pan H, Lv X. Reconstruction of spatially continuous water levels in the Columbia river estuary: The method of empirical orthogonal function revisited. Estuar Coast Shelf Sci 2019;222:81–90.

[25] Zhang X, Liu P, Zhao Y, Deng C, Li Z, Xiong M. Error correction-based forecasting of reservoir water levels: Improving accuracy over multiple lead times. Environ Model Softw 2018;104:27–39.

[26] Goovaerts P. Geostatistical prediction of water lead levels in Flint, Michigan: A multivariate approach. Sci Total Environ 2019;647:1294–304.

[27] Karri RR, Wang X, Gerritsen H. Ensemble based prediction of water levels and residual currents in Singapore regional waters for operational forecasting. Environ Model Softw 2014;54:24–38.

[28] Karim MR, Ashiquzzaman Nipu SM, Hossain Shawon MS, Kumar R, Salman S, Verma A, et al. Machinability investigation of natural fibers reinforced polymer matrix composite under drilling: Leveraging machine learning in bioengineering applications. AIP Adv 2024;14. https://doi.org/10.1063/5.0200625.

[29] Sen B, Bhowmik A, Prakash C, Ammarullah MI. Prediction of specific cutting energy consumption in eco-benign lubricating environment for biomedical industry applications: Exploring efficacy of GEP, ANN, and RSM models. AIP Adv 2024;14. https://doi.org/10.1063/5.0217508.

[30] Lestari WD, Adyono N, Faizin AK, Haqiyah A, Sanjaya KH, Nugroho A, et al. Optimization of 3D printed parameters for socket prosthetic manufacturing using the taguchi method and response surface methodology. Results Eng 2024;21:101847. https://doi.org/10.1016/j.rineng.2024.101847.

[31] Rao PM, Dhoria SH, Patro SGK, Gopidesi RK, Alkahtani MQ, Islam S, et al. Artificial intelligence based modelling and hybrid optimization of linseed oil biodiesel with graphene nanoparticles to stringent biomedical safety and environmental standards. Case Stud Therm Eng 2023;51:103554. https://doi.org/10.1016/j.csite.2023.103554.

[32] Mehdinejadiani B, Fathi P, Khodaverdiloo H. An inverse model-based Bees algorithm for estimating ratio of hydraulic conductivity to drainable porosity. J Hydrol (amst) 2022;608:127673.

[33] Kodandappa R, Nagaraja S, Chowdappa MM, Krishnappa M, Poornima GS, Ammarullah MI. Application of the Taguchi method and RSM for process parameter optimization in AWSJ machining of CFRP composite-based orthopedic implants. Open Eng 2024;14. https://doi.org/10.1515/eng-2024-0057.

[34] Manola MS, Singh B, Singla MK, Chohan JS, Kumar R, Bisht YS, et al. Investigation of melt flow index and tensile properties of dual metal reinforced polymer composites for 3D printing using machine learning approach: Biomedical and engineering applications. AIP Adv 2024;14. https://doi.org/10.1063/5.0207551.

[35] Isa IGT, Ammarullah MI, Efendi A, Nugroho YS, Nasrullah H, Sari MP. Constructing an elderly health monitoring system using fuzzy rules and internet of things. AIP Adv 2024;14. https://doi.org/10.1063/5.0195107.

[36] Wang Y, Guo S, Yang G, Hong X, Hu T. Optimal early refill rules for Danjiangkou Reservoir. Water Sci Eng 2014;7:403–19.

[37] Hassan S, Amjad A, Farooq MU, Anwar S, Ammarullah MI. Applying lean production system philosophy to reduce patient waiting time in healthcare services: Simulation-based optimization and validations through experiment. AIP Adv 2024;14. https://doi.org/10.1063/5.0210721.

[38] Jain SK. Development of integrated sediment rating curves using ANNs. J Hydraul Eng 2001;127:30–7.

[39] Sessa V, Assoumou E, Bossy M. Modeling the climate dependency of the run-of-river based hydro power generation using machine learning techniques: an application to French. Portuguese and Spanish Cases 2020.

[40] Dehghani M, Riahi-Madvar H, Hooshyaripor F, Mosavi A, Shamshirband S, Zavadskas EK, et al. Prediction of hydropower generation using grey wolf optimization adaptive neuro-fuzzy inference system. Energies (basel) 2019;12:289.

[41] Alrayess H, Gharbia S, Beden N, Keskin AU. Using machine learning techniques and deep learning in forecasting the hydroelectric power generation in almus dam, Turkey. SAFETY 2018;72.

[42] Li L, Yao F, Huang Y, Zhou F. Hydropower generation forecasting via deep neural network. In: 2019 6th International Conference on Information Science and Control Engineering (ICISCE), IEEE; 2019. p. 324–8.

[43] Li G, Sun Y, He Y, Li X, Tu Q. Short-term power generation energy forecasting model for small hydropower stations using GA-SVM. Math Probl Eng 2014; 2014..

[44] Hammid AT, Bin SMH, Abdalla AN. Prediction of small hydropower plant power production in Himreen Lake dam (HLD) using artificial neural network. Alex Eng J 2018;57:211–21.

[45] Hanoon MS, Ahmed AN, Razzaq A, Oudah AY, Alkhayyat A, Huang YF, et al. Prediction of hydropower generation via machine learning algorithms at three Gorges Dam, China. Ain Shams Engineering Journal 2023;14:101919.

[46] Keerthiveettil Ramakrishnan S, Vijayananth K, Arivendan A, Ammarullah MI. Evaluating the effects of pineapple fiber, potato waste filler, surface treatment, and fiber length on the mechanical properties of polyethylene composites for biomedical applications. Results Eng 2024;24:102974. https://doi.org/10.1016/j.rineng.2024.102974.

[47] Balasubramainian NK, Kothandaraman L, Sathish T, Giri J, Ammarullah MI. Optimization of process parameters to minimize circularity error and surface roughness in fused deposition modelling (FDM) using Taguchi method for biomedical implant fabrication. Adv Manuf Polym Compos Sci 2024;10. https://doi.org/10.1080/20550340.2024.2406156.

[48] Khajavi H, Rastgoo A. Improving the prediction of heating energy consumed at residential buildings using a combination of support vector regression and meta-heuristic algorithms. Energy 2023;272:127069.

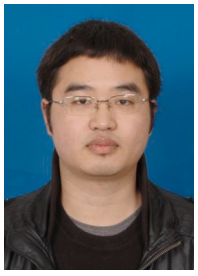
[49] Rosenlieb E. Closed loop pumped storage hydropower resource assessment of the United States. DOE Open Energy Data Initiative (OEDI); National Renewable Energy Lab.(NREL …; 2022.

[50] Zhang W, Wang H, Lin Y, Jin J, Liu W, An X. Reservoir inflow predicting model based on machine learning algorithm via multi-model fusion: A case study of Jinshuitan river basin. IET Cyber-systems and Robotics 2021;3:265–77.

[51] Ji C, Zhou T, Huang H. Operating rules derivation of jinsha reservoirs system with parameter calibrated support vector regression. Water Resour Manag 2014;28:2435–51.

[52] Vapnik VN. An overview of statistical learning theory. IEEE Trans Neural Netw 1999;10:988–99.

[53] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw 2014;69:46–61.
[54] Mirjalili S. How effective is the Grey Wolf optimizer in training multi-layer perceptrons. Appl Intell 2015;43:150–61.
[55] Li S, Chen H, Wang M, Heidari AA, Mirjalili S. Slime mould algorithm: A new method for stochastic optimization. Futur Gener Comput Syst 2020;111:300–23.
[56] Liu Y, Heidari AA, Ye X, Liang G, Chen H, He C. Boosting slime mould algorithm for parameter identification of photovoltaic models. Energy 2021;234:121164.
[57] Xu Y, Zheng Y, Du Y, Yang W, Peng X, Li C. Adaptive condition predictive-fuzzy PID optimal control of start-up process for pumped storage unit at low head area. Energy Convers Manag 2018;177:592–604.
[58] Borgonovo E. Measuring uncertainty importance: investigation and comparison of alternative approaches. Risk Anal 2006;26:1349–61.
[59] Borgonovo E. A new uncertainty importance measure. Reliab Eng Syst Saf 2007;92:771–84.
[60] Hadjimichael A, Quinn J, Reed P. Advancing diagnostic model evaluation to better understand water shortage mechanisms in institutionally complex river basins. Water Resour Res 2020;56:e2020WR028079.



Zhenya Qi 1988-, male, Han ethnicity, from Heze, Shandong, China, is a master's student with a research focus on watershed hydrological simulation and forecasting.

Yudong Feng 1995-, male, Han ethnicity, born in Nanyang, Henan Province, China. He is a master's student with a research focus on marine engineering environment.

Shoufeng Wang 1987-, male, Han ethnicity, born in Jinan, Shandong, with a bachelor's degree and research focus on hydrological and meteorological surveying.

Chao Li 1986-, male, Mongolian, from Tongliao, Inner Mongolia, with a bachelor's degree in hydrology and water resource engineering.