



OPEN Predicting mine water inflow volumes using a decomposition-optimization algorithm-machine learning approach

Jiaxin Bian^{1,2,3}, Tao Hou⁴, Dengjun Ren⁴, Chengsen Lin⁴, Xiaoying Qiao^{1,2,3}✉, Xiongde Ma^{1,2,3}✉, Ji Ma⁴, Yue Wang^{1,2,3}, Jingyu Wang⁴ & Xiaowei Liang⁴

Disasters caused by mine water inflows significantly threaten the safety of coal mining operations. Deep mining complicates the acquisition of hydrogeological parameters, the mechanics of water inrush, and the prediction of sudden changes in mine water inflow. Traditional models and singular machine learning approaches often fail to accurately forecast abrupt shifts in mine water inflows. This study introduces a novel coupled decomposition-optimization-deep learning model that integrates Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Northern Goshawk Optimization (NGO), and Long Short-Term Memory (LSTM) networks. We evaluate three types of mine water inflow forecasting methods: a singular time series prediction model, a decomposition-prediction coupled model, and a decomposition-optimization-prediction coupled model, assessing their ability to capture sudden changes in data trends and their prediction accuracy. Results show that the singular prediction model is optimal with a sliding input step of 3 and a maximum of 400 epochs. Compared to the CEEMDAN-LSTM model, the CEEMDAN-NGO-LSTM model demonstrates superior performance in predicting local extreme shifts in mine water inflow volumes. Specifically, the CEEMDAN-NGO-LSTM model achieves scores of 96.578 in MAE, 1.471% in MAPE, 122.143 in RMSE, and 0.958 in NSE, representing average performance improvements of 44.950% and 19.400% over the LSTM model and CEEMDAN-LSTM model, respectively. Additionally, this model provides the most accurate predictions of mine water inflow volumes over the next five days. Therefore, the decomposition-optimization-prediction coupled model presents a novel technical solution for the safety monitoring of smart mines, offering significant theoretical and practical value for ensuring safe mining operations.

Keywords Mine water inflow, Deep learning models, CEEMDAN, NGO, LSTM, Short-term prediction

Coal constitutes 56.2% of China's primary energy consumption structure¹. With ongoing advancements in coal production technology, most old mines have transitioned to deep mining phases, where mine water inflow has become a primary factor causing disasters in coal production². According to incomplete statistics, economic losses caused by mine water inflow over the past 30 years have reached 2.7 billion yuan, with an additional 15 billion yuan in reserves left undeveloped due to this threat³. Furthermore, accurately predicting mine water inflow volumes is crucial for ensuring the safety and ecological balance of mining operations^{4,5}. It is also essential for evaluating the rationality of coal mining schemes and the design of drainage systems, which directly affect mining efficiency and sustainable development⁶.

¹School of Water and Environment, Chang'an University, Xi'an 710064, China. ²Key Laboratory of Subsurface Hydrology and Ecological Effect in Arid Region of Ministry of Education, Chang'an University, Xi'an 710064, China. ³Key Laboratory of Eco-Hydrology and Water Security in Arid and Semi-Arid Regions of Ministry of Water Resources, Chang'an University, Xi'an 710064, China. ⁴Shaanxi Zhengtong Coal Industry Co, Ltd, Xianyang 713600, China. ✉email: qiaoxiaoy@163.com; hgmxd@chd.edu.cn

Methodologies for predicting mine water inflow typically classify models into dynamics-based and data-driven approaches^{7–9}. Dynamics-based models rely on a comprehensive understanding of the mechanisms behind mine water inflow and include mathematical representations and parameterizations of related physical processes, such as the water balance method, analytical method, and numerical methods^{10,11}. Despite their specific applicability and effectiveness, the continuous extraction of coal mines, leading to changes in some geological conditions, complicates accurate data collection and challenges the comprehensive and precise management of mine water inflow volumes¹². In contrast, data-driven models offer strong capabilities, low entry barriers, high adaptability, and fewer limitations, providing new avenues for addressing the variability and complexity of mine water inflow volumes. Correlation analysis and time series analysis models are widely used in this context. For instance, Miladinovic et al. utilized a linear correlation regression model to predict and calibrate mine water inflow volume in southwest Serbia, significantly reducing prediction errors due to unknown hydrogeological parameters¹³. Meanwhile, Yang et al.¹⁴ combined SARMA and ARIMA models to achieve a prediction accuracy of 94% for mine water inflow. However, these models primarily reveal linear relationships and fail to adequately address the nonlinear processes involved in mine water inflow, thus showing significant limitations in capturing abrupt changes in trends¹⁵.

Deep learning, an innovative data-driven predictive model, demonstrates remarkable resilience to changes in data dimensionality. Research consistently shows that these models can maintain robust predictive performance, even with single-dimensional inputs, provided the data sequence is rich in information¹⁶. Among these, Long Short-Term Memory (LSTM) networks, recognized for their effectiveness in time series prediction, excel at managing high-dimensional and nonlinear data relationships^{17,18}. These methods have been widely applied in predicting concentrations of air pollutants, runoff forecasts, and groundwater level predictions. M. Cheng et al. (2020) utilized Artificial Neural Networks (ANN) and LSTM models for daily and monthly runoff predictions in the Nan River Basin (NRB) and the Ping River Basin (PRB), demonstrating that LSTM's unique forget gate feature enhances performance in daily forecasts¹⁹. In 2018, Zhang and colleagues assessed the predictive performance of Multilayer Perceptrons, Wavelet Neural Networks, and LSTM, finding that LSTM provided more accurate multi-step time predictions²⁰. Zhang et al.²¹ developed a recurrent neural network prediction method using LSTM to forecast methane concentrations in coal mine production processes within the mining safety domain. Shu et al.²² integrated microseismic monitoring technology with LSTM, showing that LSTM-identified coal fracture microseismic events sequentially respond to gas concentrations, offering potential for early warning of gas accidents. In summary, while LSTM has been primarily focused on predicting methane gas eruptions and seismic activities, research on predictions related to water hazards remains limited²³.

Data preprocessing is crucial for enhancing predictive performance. Signal decomposition methods segregate fluctuations and variations at different scales from the original sequence while preserving the inherent characteristics of the data, facilitating better extraction of local features and detailed information²⁴. Researchers commonly use methods such as Singular Spectrum Analysis (SSA), Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to manage the non-stationarity in hydrological time series^{25,26}. Among these, SSA decomposition requires manual selection of windows, risking the loss of essential features or over-decomposition. EMD, which does not require manual intervention, performs better in handling non-stationary and nonlinear data²⁷. However, EMD faces the issue of mode mixing, leading to the development of improvements such as EEMD and CEEMDAN²⁸. Notably, Ji et al.²⁴ showed that CEEMDAN effectively avoids mode mixing and reduces noise in decomposed sequences, thus enabling more accurate trend predictions in models. Additionally, optimization algorithms play a vital role in enhancing model predictive capabilities by automatically adjusting model parameters to find global optima, thus minimizing manual intervention^{29,30}. Among the plethora of optimization algorithms, Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Grey Wolf Optimizer (GWO) have recently garnered significant attention and research^{31,32}. Jian Zhou et al.³³ introduced a hybrid model combining the Grey Wolf Optimizer with Random Forest (RF) for predicting tunnel water inflow volumes, demonstrating that this hybrid model predicts tunnel water inflow effectively and exhibits higher accuracy and stability compared to traditional methods.

Despite the significant insights provided by previous studies into the application potential of signal decomposition and optimization algorithms for time series data, researchers have not thoroughly explored critical issues regarding the generalizability of models and their ability to detect abrupt changes in the data. To address the volatility of water inflow data and enhance model precision, we developed robust data-driven models that integrate LSTM, Northern Goshawk Optimization (NGO), and CEEMDAN to predict mine water inflow. A comparative analysis of this model's performance against traditional data-driven models demonstrates its superior stability and accuracy. The primary contributions of this study are as follows:

1. Investigating the impact of sliding input steps on model prediction outcomes.
2. Developing a new “decomposition-optimization-prediction” coupled model framework for mine water forecasting, suitable for complex and highly variable mine water inflows. This framework integrates CEEMDAN, NGO, and LSTM, enhancing the capture of subtle changes and nonlinear characteristics in mine water inflow data and effectively predicting sudden trend shifts.
3. Analyzing the predictive performance of the coupled model across different forecasting periods (1 day, 3 days, 5 days, and 7 days) to determine the optimal forecasting period. This analysis provides robust support for short-term forecasting in mine water prevention and control.

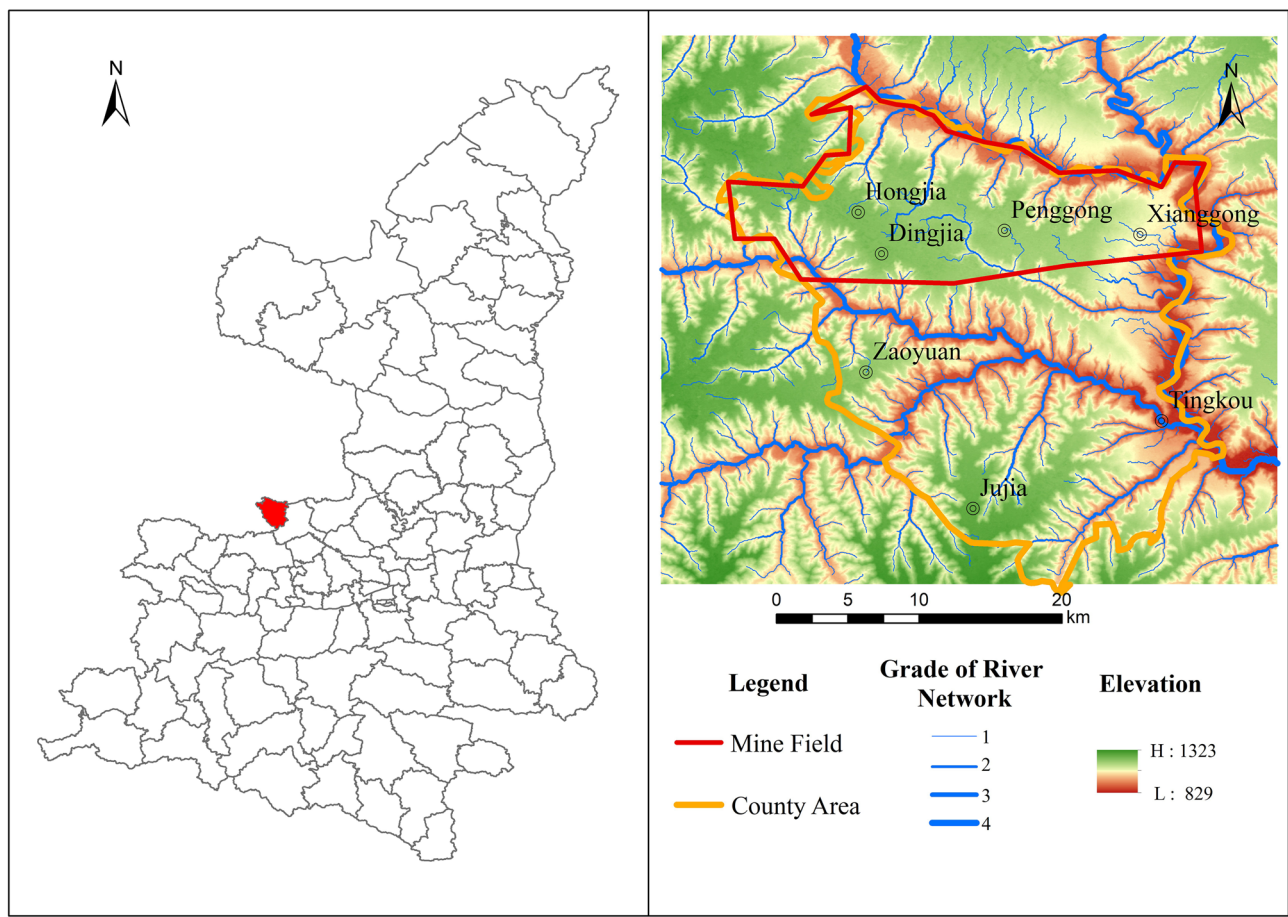


Figure 1. Location of the study area.

Study area and methodology

Experiment area and data collection

The study area, located in the Gaojiabao mining district of Xianyang, Shaanxi Province, extends approximately 25.7 km east to west and about 16.6 km north to south, covering an area of roughly 219 km² (Fig. 1). The coal seams within this district are buried at depths ranging from 600 to 1100 m, with a primary focus on exploiting the No. 4 coal seam. We estimate the extractable area for this seam to be 111.29 km², with an average extractable thickness of 9.33 m. Table 1 presents the hydraulic properties and seepage characteristics of high-pressure aquifers in the overlying strata, along with their distances from the coal seam roof boundary. The stratigraphy, from bottom to top, comprises the Yan'an Formation (J2y), Zhiluo Formation (J2z), Anding Formation (J2a), Luohui Formation (K1l), Huachi Formation (K1h), Neogene System (N), and Quaternary System (Q). The Jurassic Yan'an Formation within the mining area contains coal-bearing strata. The principal aquifers can be categorized as follows: the porous phreatic aquifer of the Quaternary system; the sandstone fissure aquifer of the Lower Cretaceous Huachi Formation; the sandstone porous-fissure confined aquifer of the Lower Cretaceous Luohe Formation; and the sandstone fissure confined aquifer of the Middle Jurassic Zhiluo Formation.

Methodology

CEEMDAN method

The CEEMDAN, an advanced decomposition method, builds on the principles of EMD³⁴ and EEMD³⁵. This method adaptively decomposes an original signal into various Intrinsic Mode Functions (IMFs), each incorporating a level-specific white noise to address the mode mixing issues found in EMD and improve upon EEMD's limitations, such as significant reconstruction errors and low computational efficiency³⁶. The main steps involved in the CEEMDAN process are:

1. Initial Decomposition: CEEMDAN utilizes Empirical Mode Decomposition (EMD) to perform n decompositions on the original signal $S(t)$ augmented with Gaussian white noise $w^i(t)$ that follows a standard normal distribution ($i = 1, 2, \dots, n$). The first modal component derived from this decomposition is expressed as:

$$IMF_1(t) = \frac{1}{M} \sum_{i=1}^M IMF_1^i(t) \quad (1)$$

	Stratum		Water-richness	Distance to the coal seam roof
Cretaceous	K ₁ h	Upper section	The aquifer is well-developed with weak water richness	800m
	K ₁ h	Lower section	Aquifer undeveloped, aquiclude	550m
	K ₁ l	Upper section	The aquifer is moderately developed with medium water richness	450m
		Upper section of the middle	The aquifer is highly developed and has strong water-rich properties.	350–400m
		Lower section of the middle	The aquifer is highly developed, with moderate to strong water-richness.	200–250m
		Upper section of the lower	The aquifer is moderately developed with medium water richness.	100–150m
		Lower section of the lower	The aquifer has a low level of development with weak water-rich properties.	100m
Jurassic	J ₂ a	Complete section	Aquifer undeveloped, aquiclude.	70m
	J ₂ z	Upper section	Aquifer undeveloped, aquiclude.	< 15m
		Lower section	The aquifer is underdeveloped, resulting in slow groundwater runoff.	< 15m
	J ₂ y	Upper section	The aquifer is underdeveloped, resulting in slow groundwater runoff.	<10m

Table 1. Hydraulic properties and seepage characteristics of high-pressure aquifers in overlying strata and their distance from the coal seam roof boundary.

The first decomposed component, $IMF_1(t)$, is then removed from the original time series $S(t)$, resulting in the first residual sequence $R_1(t)$:

$$R_1(t) = S(t) - IMF_1(t) \tag{2}$$

2. Subsequent Decompositions: This step replicates the initial procedure, where a series of adaptive Gaussian white noises $w^i(t)$ ($i = 1, 2, \dots, m$) are added to the first residual series. This involves the continued decomposition of the sequence $R_1(t) + \varepsilon_1 E_1[v_m(t)]$ using EMD, resulting in the extraction of the IMF_2 mode component:

$$IMF_2(t) = \frac{1}{M} \sum_1^M E_1[R_1(t)] + \varepsilon_1 E_1[v_m(t)] \tag{3}$$

3. Continued Decomposition: Repeating the aforementioned steps yields the k -th residual and the $(k + 1)$ -th order modal component. The CEEMDAN decomposition process terminates when the number of extrema in the residual does not exceed two, resulting in k -order modal components. The final residual is denoted as $r(t)$:

$$R_k(t) = R_{k-1}(t) - IMF_1(n) \tag{4}$$

$$IMF_{i+1}(t) = \frac{1}{M} \sum_1^M E_1(R_i(t) + \varepsilon_i E_i[v_m(t)]) \tag{5}$$

$$r(t) = S(t) - \sum_{K=1}^K IMF_1(t) \tag{6}$$

Therefore, the original time series $S(t)$, after undergoing CEEMDAN decomposition, can satisfy:

$$S(t) = r(t) + \sum_{K=1}^K IMF_K(t) \tag{7}$$

LSTM method

Introduced in 1997 by Hochreiter and Schmidhuber, LSTM^{37,38} is an enhancement over the existing Recurrent Neural Networks (RNNs) and addresses issues of vanishing and exploding gradients encountered during RNN training. This innovation presents an effective method for modeling complex time series with multiple variables over extended periods^{39,40}. The cell structure^{41,42} of the LSTM is illustrated in Fig. 2.

1. Forget Gate: The forget gate determines, based on a specific formula, which information is retained from the previous time step C_{t-1} to the current state C_t :

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{8}$$

Here, f_t denotes the output function value of the forget gate at time t . W_f and b_f represent the forget gate’s weight matrix and bias vector, respectively. The function σ indicates the sigmoid activation function, and h_{t-1} is the output value of the hidden layer at time $t-1$.

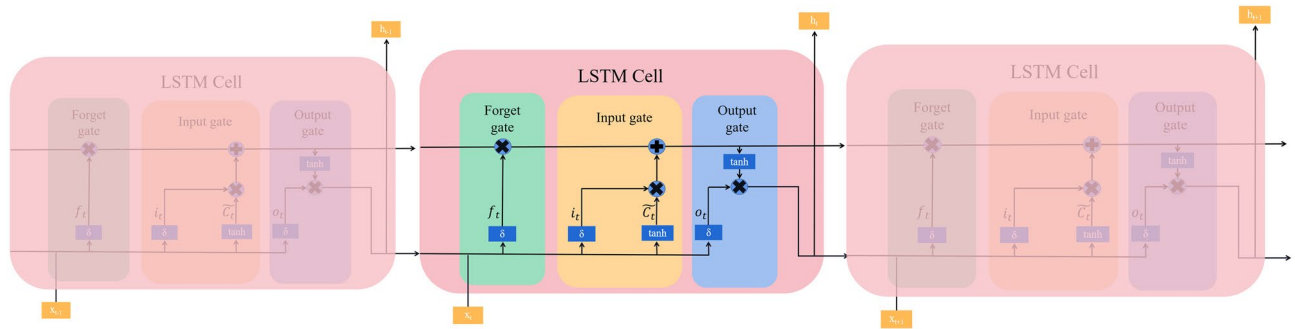


Figure 2. LSTM memory cell structure. C_t , h_t , and x_t is the cell state, hidden state, and input of the cell at time t .

- Input Gate: The input gate determines which new information should be updated to the cell state, which is formulated into the following formula:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (9)$$

In this formulation, it represents the activation function value of the input gate at time t . W_i and b_i respectively denote the input gate's weight matrix and bias vector.

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (10)$$

Furthermore, $\tilde{C}_t$ indicates the function value of the candidate cell state at time t , with $\tanh$ representing the hyperbolic tangent function. W_c and b_c respectively represent the candidate cell state's weight matrix and bias vector.

- Memory Update: The new cell state is formed by combining the old memory processed through the forget gate and the latest memory processed through the input gate. The cell state is updated from C_{t-1} to C_t using the following formula:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (11)$$

- Output Gate: The output gate determines which information will be transferred from the cell state to the hidden state, as calculated by the following equation:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (12)$$

In this context, o_t represents the function value of the output gate at time t , with W_o and b_o denoting the weight matrix and bias vector of the forget gate. The hidden state at time t , denoted as S_t , is determined by the output gate and the cell state.

$$S_t = \tanh(C_t) * o_t \quad (13)$$

NGO method

The Department introduced the NGO algorithm in 2021⁴³. This algorithm models the hunting behaviors of Northern Goshawks, including prey identification, attack, pursuit, and evasion strategies. The NGO algorithm effectively balances global and local searches, allowing it to tackle complex problems characterized by multiple local optima. Its robust adaptability simulates natural optimization processes, making it a powerful, flexible, and efficient tool for solving a variety of complex optimization challenges^{44,45}. The steps are outlined as follows:

- Algorithm Initialization Process: Mathematically, each goshawk in the population is represented as a vector within a population matrix. Initially, the members of the population are randomly initialized within the search space:

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N \times m} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix}_{N \times m} \quad (14)$$

Here, X denotes the matrix of the Northern Goshawk population; X_i is the initial solution of the i -th goshawk; N and m represent the number of individuals in the population and the dimensions of the problem space, respectively.

The objective function for the problem is evaluated based on the proposed solutions by each member of the group, and the values of the objective function can be expressed as a vector:

$$(X) = \begin{bmatrix} F_1 = F(X_1) \\ \vdots \\ F_i = F(X_i) \\ \vdots \\ F_N = F(X_N) \end{bmatrix} \quad (15)$$

Here, F represents the objective function vector of the Northern Goshawk population, with F_i denoting the value of the objective function corresponding to the i -th solution.

2. Prey Identification and Attack: Initially, the goshawk selects prey randomly and rapidly attacks, representing the global search of the search space to identify optimal regions. This random prey selection enhances the exploratory capacity of the algorithm. The mathematical model is:

$$P_i = X_k, i = 1, 2, \dots, N, k = 1, 2, \dots, i - 1, i + 1, \dots, N \quad (16)$$

$$x_{i,j}^{new,P1} = \begin{cases} x_{i,j} + r(p_{i,j} - Ix_{i,j}), & F_{Pi} < F_i \\ x_{i,j} + r(x_{i,j} - P_{i,j}), & F_{Pi} \geq F_i \end{cases} \quad (17)$$

$$X_i = \begin{cases} X_i^{new,P1}, & F_i^{new,P1} < F_i \\ X_i, & F_i^{new,P1} \geq F_i \end{cases} \quad (18)$$

where, P_i represents the position of the i -th Northern Goshawk's prey, and F_{Pi} is its corresponding objective function value. The variable k is a randomly chosen integer within the interval $[1, N]$, distinct from i . The term $x_{i,j}^{new,P1}$ denotes the new solution value of the i -th goshawk in the j -th dimension. $x_{i,j}$ signifies the position of the i -th Northern Goshawk in the j -th dimension, while $p_{i,j}$ refers to the j -th dimensional position of the i -th Northern Goshawk's prey. The parameters r and I are random numbers used for searching and iteration updates; r ranges within $[0, 1]$, and I can be either 1 or 2. $X_i^{new,P1}$ represents the updated position of the i -th Northern Goshawk after the first phase, and $F_i^{new,P1}$ is the objective function value after the first phase.

3. Chase and Escape Operation: When prey attempts to flee following an attack, the Northern Goshawk, known for its keen responsiveness and exceptional speed, can almost continuously pursue and ultimately capture it under any circumstances. This simulation enhances the algorithm's ability to utilize local search within the search space. Assuming a hunting radius of R , the mathematical model for the chase between the Northern Goshawk and its prey is as follows:

$$x_{i,j}^{new,P2} = x_{i,j} + R2r - 1x_{i,j} \quad (19)$$

$$R = 0.021 - \frac{t}{T} \quad (20)$$

$$X_i = \begin{cases} X_i^{new,P2}, & F_i^{new,P2} < F_i \\ X_i, & F_i^{new,P2} \geq F_i \end{cases} \quad (21)$$

where, $x_{i,j}^{new,P2}$ represents the value of the i -th goshawk's solution at the new position in the j -th dimension. Here, t and T denote the current iteration and maximum number of iterations, respectively. $X_i^{new,P2}$ is the updated position of the i -th Northern Goshawk in the second phase, while $F_i^{new,P2}$ denotes the objective function value for the i -th Northern Goshawk after the update in the second phase.

Coupling prediction method

Figure 3 displays the schematic flowchart of the CEEMDAN-NGO-LSTM model. This study introduces a prediction model for mine water inflow utilizing an LSTM network, enhanced by data decomposition and optimization techniques. Initially, the CEEMDAN algorithm decomposes the raw data into multiple IMFs, which capture signal characteristics across various scales and frequencies. These IMFs provide a multi-layered input for the LSTM model. Subsequently, the NGO algorithm optimizes the LSTM parameters for each element. The LSTM utilizes its unique gating mechanism (input gate, forget gate, and output gate) to effectively capture long-term dependencies and filter out irrelevant or noisy information, thereby maintaining efficient memory and state updates during the prediction of mine water inflow. The final step involves using the optimized model to make predictions and aggregating the results from each element into a comprehensive forecast.

All models were executed using Matlab R2022b on a computer running Windows 10, equipped with an Intel(R) Core(TM) i5-9300H CPU @ 2.40 GHz and 16 GB of RAM.

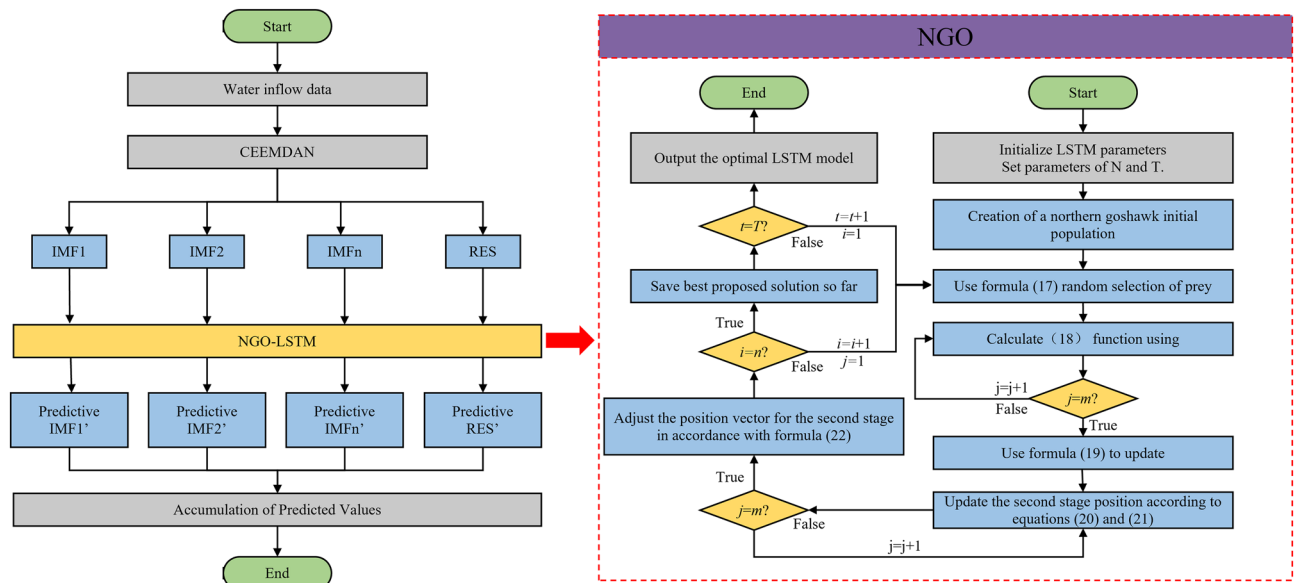


Figure 3. Flow prediction chart based on ICEEMDAN-NGO-LSTM methods.

Model performance evaluation indicators

To assess the predictive accuracy of the model, we selected the following metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the Nash–Sutcliffe Efficiency coefficient (NSE)^{46–49}. RMSE, which is the square root of the average of squared forecasting errors, measures deviations between observed and predicted values. MAE quantifies the absolute differences between observed and predicted values, providing a direct measure of disparity. MAPE, expressed as a percentage, indicates the proportionate error between observed and actual values and is primarily used to assess data anomalies. Smaller values of RMSE, MAE, and MAPE indicate higher accuracy of the prediction model. NSE, a widely used efficiency metric in hydrology, measures the consistency between model predictions and actual observations. NSE values range from $-\infty$ to 1, with values closer to 1 indicating a higher correspondence between predictions and observations. The theoretical formulas for RMSE, MAE, MAPE, and NSE satisfy:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_i^*)^2} \quad (22)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - y_i^*| \quad (23)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - y_i^*}{y_i} \right| \times 100 \% \quad (24)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (y_i - y_i^*)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (25)$$

where, y_i represents the actual mine water inflow data, $\bar{y}$ denotes the average value of the actual data, and y_i^* is the predicted mine water inflow.

Result and discussion

Change characteristics of water inflow

Figure 4a presents the measured data for water inflow volumes at the Gaojiabao mine from 2020 to 2023. As shown in Fig. 4, there are localized abrupt changes in water inflow volumes, exhibiting a gradually increasing trend. Specifically, the minimum inflow volume recorded on June 4, 2020, was 3506 m³/h, while the maximum volume reached 7956 m³/h on September 1, 2022, with an average volume of 5390.83 m³/h over the period. The extreme values differ by a factor of two, and significant fluctuations are observed at specific locations.

Box plots were used to analyze the distribution characteristics of mine water inflow volumes on a monthly scale. As depicted in Fig. 4b, the variations in mine water inflow volumes exceeded 1000 m³/h in March and April of 2020, in March, April, September, and October of 2022, and in January and February of 2023. The most significant disparity was observed in September 2022, reaching 2000 m³/h. This analysis highlights numerous abrupt changes and turning points in mine water inflow volumes over short periods. From March 2022 to March 2023, mine water inflow displayed a trough-like pattern with a skewness of 0.376. The right-skewed distribution with a long tail suggests a higher likelihood of exceptionally high inflows as mining progresses. This pattern indicates an increased probability of exceptional mine water inflow due to ongoing mining activities. The underlying

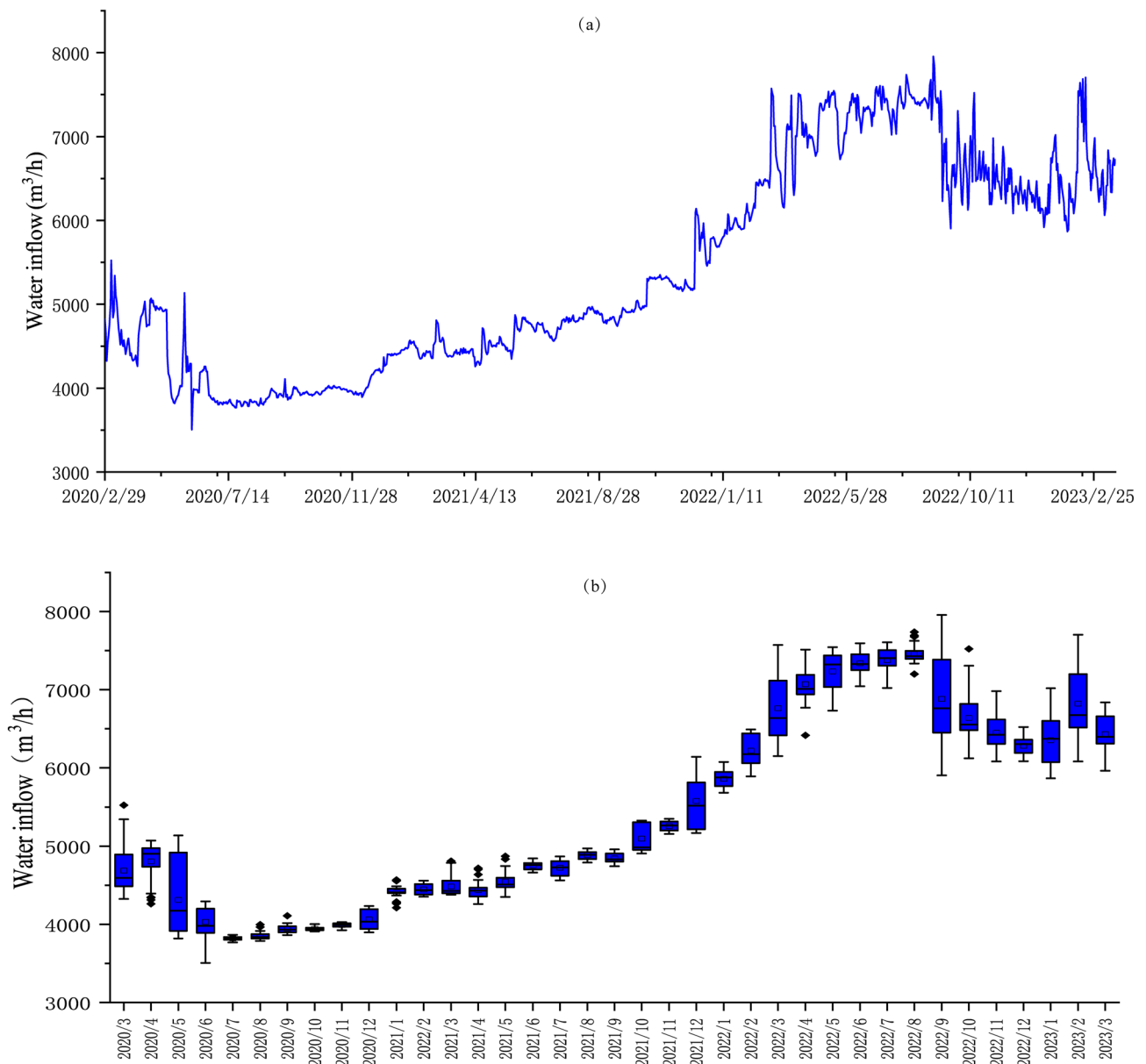


Figure 4. Time series diagram of water inflow (a) and Box plot for each month (b).

causes are likely related to pervasive joint fractures within the Luohe formation, significant static reserves, and the formation of conductive pathways when mining-induced fractures intersect with pre-existing joints.

Comparison of single prediction models

For the time series prediction, 80% of the historical data on water inflow volume from Fig. 4a was allocated to the training set, while the remaining 20% was used as the validation set⁵⁰. It is important to note that similar results were observed with other models; however, this study focuses primarily on the LSTM model as a representative

Lag(day)	1	3	5	7	10	15	20	25	30	35
MAE	187.492	186.468	189.754	188.424	193.453	192.962	204.101	201.930	213.504	220.818
MAPE	2.825%	2.811%	2.866%	2.841%	2.924%	2.920%	3.0981%	3.067%	3.258%	3.377%
RMSE	256.459	252.381	255.364	254.289	259.491	259.038	266.827	264.621	276.842	285.704
NSE	0.719	0.724	0.716	0.718	0.705	0.705	0.693	0.700	0.686	0.676

Table 2. The Impact of LSTM Time Lag Length on Predicting Water Inrush.

Epochs	100	200	300	400	500
MAE	186.712	185.462	186.093	186.468	186.341
MAPE	2.787%	2.784%	2.805%	2.811%	2.808%
RMSE	253.552	252.553	252.727	252.381	253.036
NSE	0.7213	0.7212	0.7238	0.7241	0.7231

Table 3. Impact of the Maximum Number of Epochs in LSTM on Water Inrush Prediction.

example. The detailed results of the evaluation metrics for the LSTM model, including sliding input steps and maximum epoch numbers, are presented in Tables 2 and 3. These tables show that the NSE coefficient initially increases and then decreases as the sliding input steps extends from 1 to 10 days, peaking at 3 days. Conversely, when the sliding input steps exceeds 10 days, the NSE values gradually decline. Other metrics, such as RMSE and MAE, display an opposite trend to NSE, decreasing initially and then increasing. This indicates that increasing the sliding input steps does not improve prediction results; optimal performance is achieved at a 3-day lag. Beyond this, the model may learn unnecessary and potentially noisy patterns, distorting the prediction outcomes. The outcomes of backward-in-time iterations show no direct correlation between increased time step value and enhanced forecast accuracy. Extending the time steps excessively compels the model to process irrelevant information, reducing predictive precision. Therefore, selecting a time step that aligns with the model's capabilities is crucial for achieving accurate predictions. The focus on mine water inflow prediction on short-term fluctuations, rather than on long-term trends or the cumulative impact of historical data, underscores the importance of choosing an appropriate time step.

In contrast, increasing the maximum number of epochs has only a minor effect on the prediction outcomes. The NSE coefficient initially increases and then decreases, while RMSE displays the opposite trend. The best results are achieved at 400 epochs. Furthermore, MAE achieves its optimum at 200 epochs, and MAPE reaches its best at 300 epochs. Considering all factors, the performance metrics of the parameters are optimal when the maximum number of training epochs is set to 400. Therefore, to enhance prediction accuracy for predicting the volume of mine water inflows, it is recommended to select a sliding input steps of 3 and a maximum of 400 epochs.

Figure 5 displays the predictive performance of six single-models: Autoregressive Integrated Moving Average (ARIMA), Back Propagation (BP), Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), and LSTM. The results show that the computational efficiency of each model is similar. However, except for LSTM, the other models generate relatively conservative predictions and do not achieve the predictive accuracy of LSTM. Specifically, LSTM achieves the following results in key performance metrics: a MAE of 186.468, a MAPE of 2.811%, a RMSE of 252.378, and a NSE coefficient of 0.724.

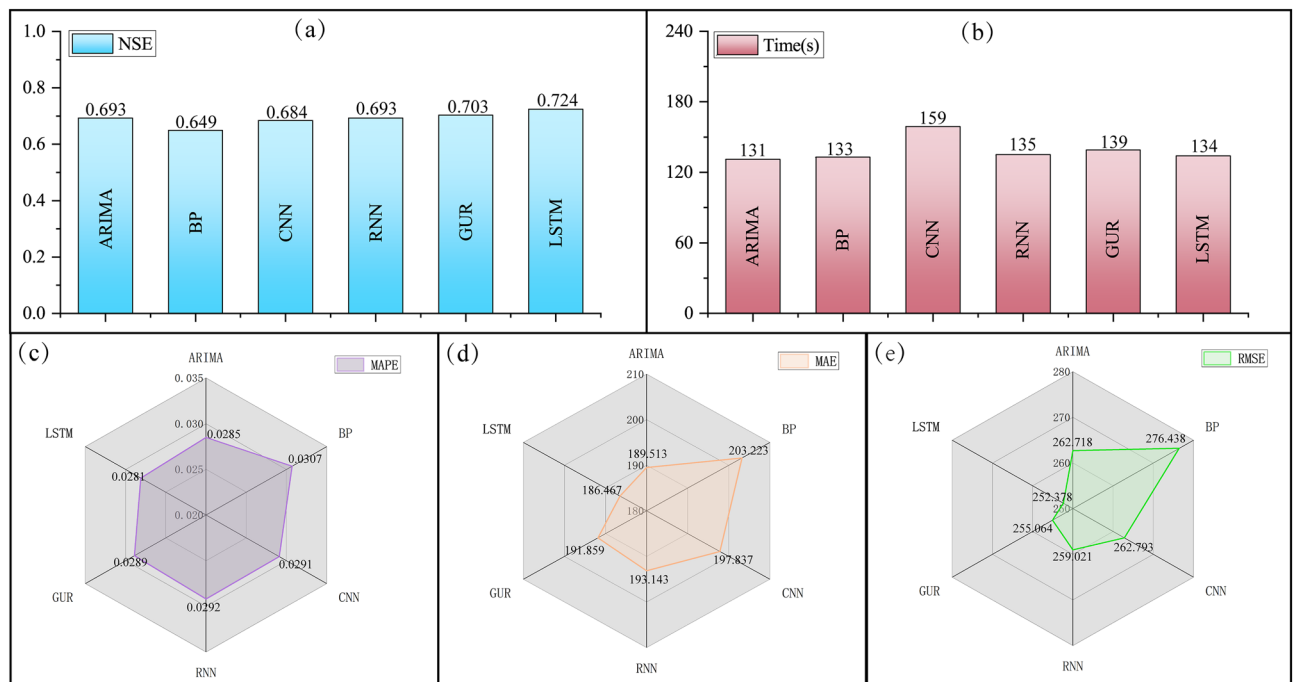


Figure 5. Comparison of 4 evaluation indicators for each model prediction (a) NSE (b) TIME (c) MAPE (d) MAE (e) RMSE.

These outcomes confirm LSTM's ability to uncover complex nonlinear relationships within time series data, demonstrating its superior predictive performance. However, LSTM's robustness in sequences with frequent abrupt changes can be compromised by vanishing gradients, leading to slight lags as LSTM struggles to fully capture the peaks of sudden changes, as observed in Fig. 6. This phenomenon reduces the predictive value for early warning of inrush events, indicating that achieving accurate predictions of water inflow volumes without comprehensive data preprocessing remains a significant challenge.

Decomposition-prediction coupled model

Time series decomposition efficiently extracts local characteristics and hidden information from data, segregating fluctuations and variations in mine water inflow data from the baseline dataset. In this study, we utilized three decomposition methods to analyze mine water inflow time series data: EMD, EEMD, and CEEMDAN as shown in Fig. 7. Due to the adaptive nature of these data-driven methods, the EMD decomposition identified 7 IMFs and one residual component R; EEMD decomposition revealed 8 IMFs and one residual component R; by contrast, CEEMDAN decomposition produced 9 IMFs and one residual component R. From IMF1 to the residual component, there is a progressive decrease in the amplitude of oscillations and an increase in wavelength, reflecting a transition from high to low frequencies and depicting the periodic variation characteristics of each component's time scale. These IMFs, along with the residual component R, enable the complete reconstruction of the original signal, ensuring no loss of information during frequency decomposition.

Integrating EMD, EEMD, and CEEMDAN with LSTM led to the development of EMD-LSTM, EEMD-LSTM, and CEEMDAN-LSTM coupled predictive models. We determined the model with the highest accuracy by comparing the performance of these three coupled models, as detailed in Fig. 8 and Table 4. Specifically, the EMD-LSTM model exhibited a MAE of 143.306, a MAPE of 2.143%, a RMSE of 199.948, and an NSE coefficient of 0.821. Relative to EMD-LSTM, EEMD-LSTM showed a reduction of 8.185% in MAE, 6.337% in MAPE, 12.273% in RMSE, and an improvement of 4.580% in NSE. However, the CEEMDAN-LSTM model surpassed EEMD-LSTM in prediction accuracy, with further reductions of 2.082% in MAE, 2.880% in MAPE, 10.107% in RMSE, and an increase of 5.903% in NSE. The CEEMDAN model demonstrated the most significant performance improvements compared to EMD-LSTM and EEMD-LSTM, while the computation times for the models were similar. Although the CEEMDAN-LSTM model addressed the lag issue associated with LSTM, it was less effective in predicting abrupt changes in trends, showing a notable discrepancy between expected and actual peak values.

Decomposition-optimization-prediction coupled model

The NGO algorithm offers several significant advantages in addressing complex optimization challenges. Empirical evaluations across various benchmark tests and real-world engineering design problems have demonstrated that NGO outperforms other established algorithms, such as PSO, GA, and GWO, in terms of convergence speed and optimization accuracy. These outcomes highlight NGO's capability in handling complex and high-dimensional optimization tasks. Moreover, the NGO algorithm exhibits robust adaptability and stability across a range of problem environments, affirming its efficiency and reliability as an optimization tool⁴³. Therefore, as illustrated in Fig. 9 and Table 4, integrating the NGO algorithm into the CEEMDAN-LSTM model significantly enhanced the training performance of the coupled model. It accurately captured and predicted abrupt changes in water inflow volumes, thereby increasing the model's overall reliability in real-world scenarios. The optimized model demonstrated reductions of 25.039% in MAE, 24.525% in MAPE, and 22.536% in RMSE, along with a 5.415% improvement in the NSE coefficient. These key metrics substantiate the feasibility and significant practical application value of the CEEMDAN-NGO-LSTM in enhancing prediction accuracy.

To evaluate the model's applicability in various scenarios, the CEEMDAN-NGO-LSTM was used to predict the water inflow of a single working face in a mining area, utilizing data from November 4, 2021, to July 5, 2023. The division of the training and test sets is depicted in Figure S1, with the results presented in Fig. 10 and Table 4. The model achieved an NSE of 0.906, a MAPE of 4.060%, an MAE of 87.760, and an RMSE of 117.410. The accuracy

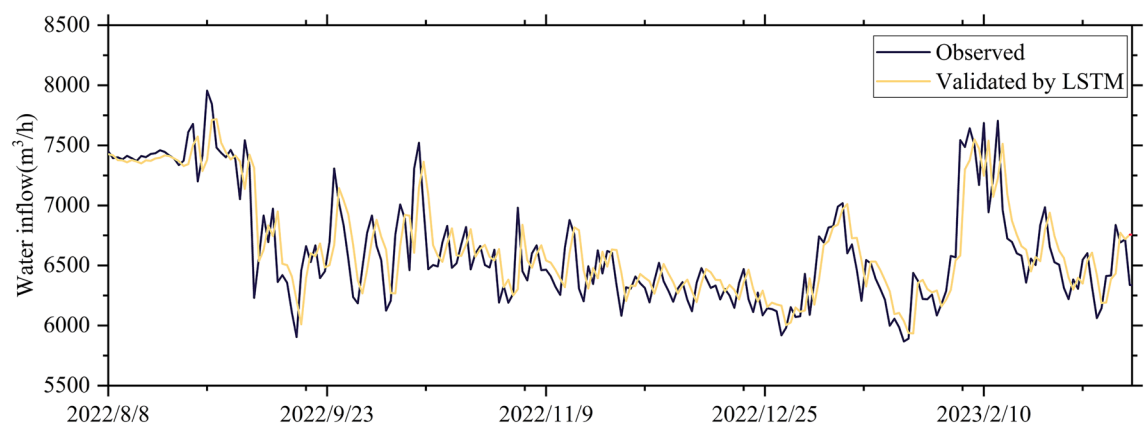


Figure 6. The validation results of the mine water inrush by LSTM.

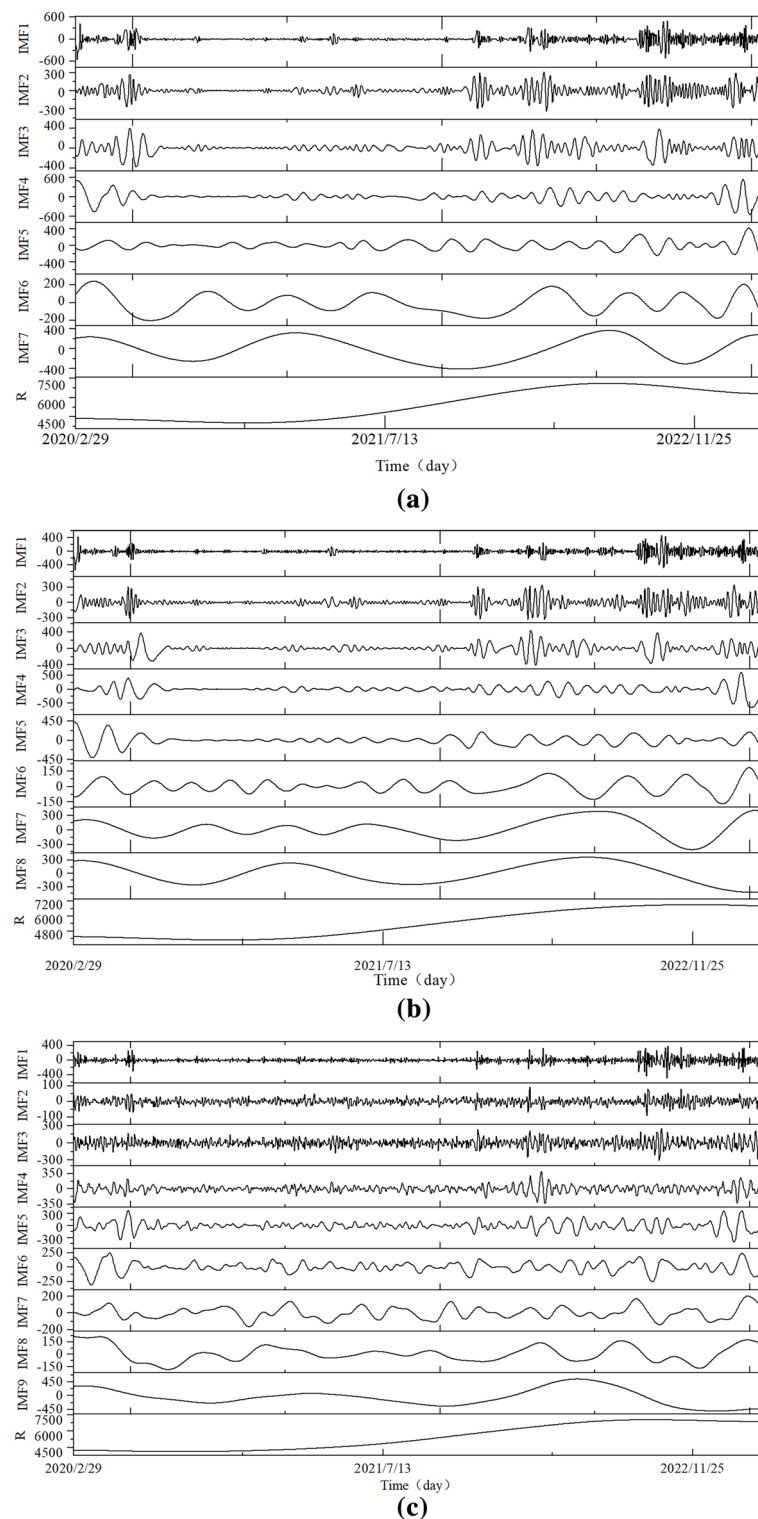


Figure 7. Results of (a) EMD (b) EEMD (c) CEEMDAN for the mine water inflow.

and capability of the CEEMDAN-NGO-LSTM to capture abrupt changes were also validated, confirming its effectiveness in practical applications.

Short-term forecasting

In our analysis, we selected the three top-performing models: the single prediction model, the decomposition prediction model, and the decomposition optimization prediction model. We applied linear fitting to the

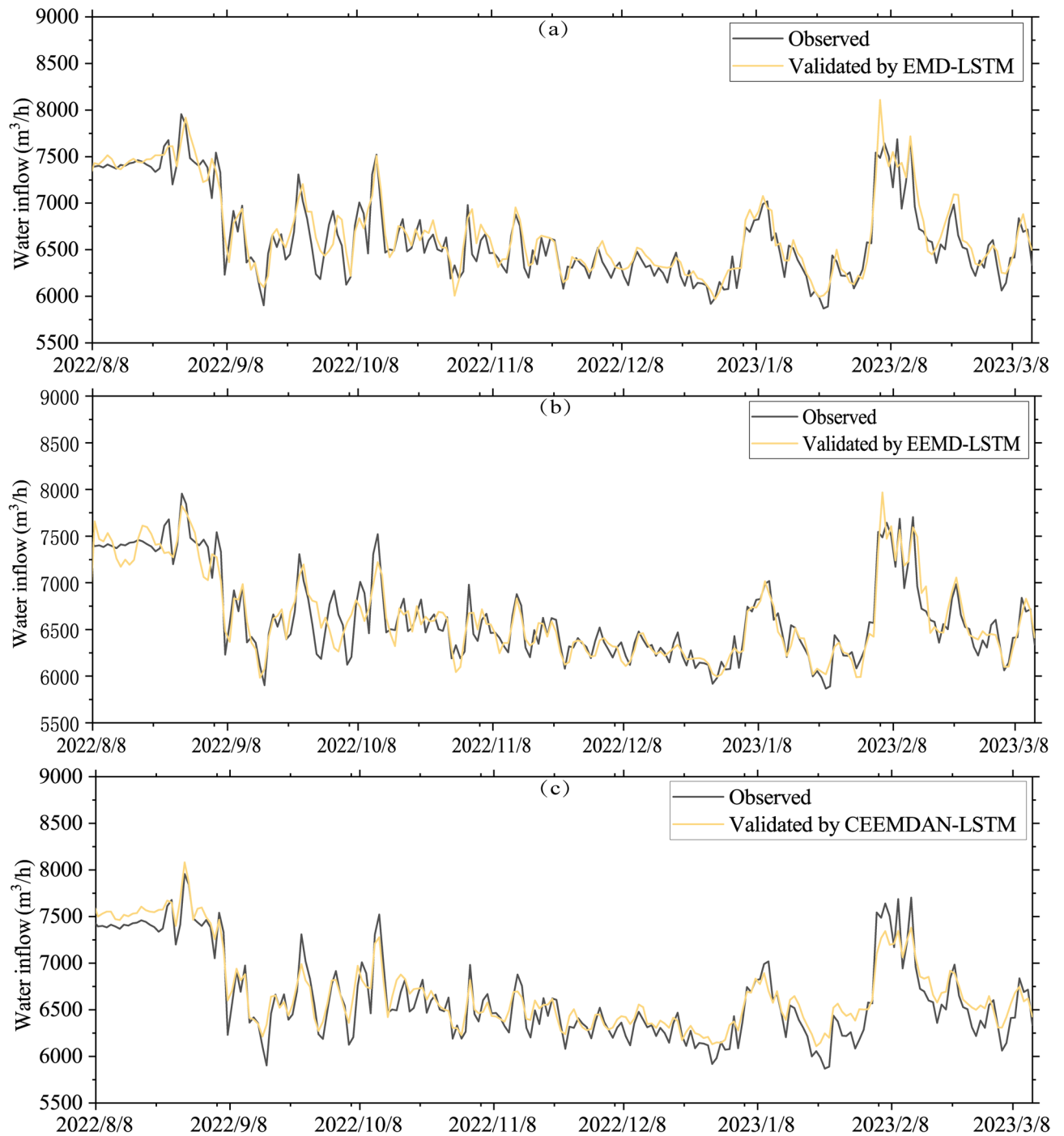


Figure 8. The validation results of the mine water inrush by EMD-LSTM (a), EEMD-LSTM (b) and CEEMDAN-LSTM (c).

predicted versus actual values during the training and validation phases, as illustrated in Fig. 11. The results of this approach are revealing.

The linear equations during the training phase are as follows:

$$\text{LSTM training phase : } Y = 1.002X - 10.913 (R^2 = 0.990)$$

$$\text{CEEMDAN - LSTM training phase : } Y = 0.990X - 60.955 (R^2 = 0.997)$$

$$\text{CEEMDAN - NGO - LSTM training phase : } Y = 0.998X + 22.376 (R^2 = 0.999)$$

	MAE	MAPE	RMSE	NSE	TIME
LSTM	186.468	2.811%	252.381	0.724	2min14s
EMD-LSTM	143.306	0.0214%	199.948	0.821	26min55s
EEMD-LSTM	131.576	0.0201%	175.408	0.858	30min13s
CEEMDAN-LSTM	128.836	0.0195%	157.678	0.909	33min29s
CEEMDAN-NGO-LSTM	96.577	0.0147%	122.143	0.958	1h40min32s
CEEMDAN-NGO-LSTM (working face)	87.760	4.060%	117.410	0.906	55min27s

Table 4. Metrics MAE, MAPE, RMSE, and NSE for Decomposition-Prediction and Decomposition-Optimization-Prediction models.

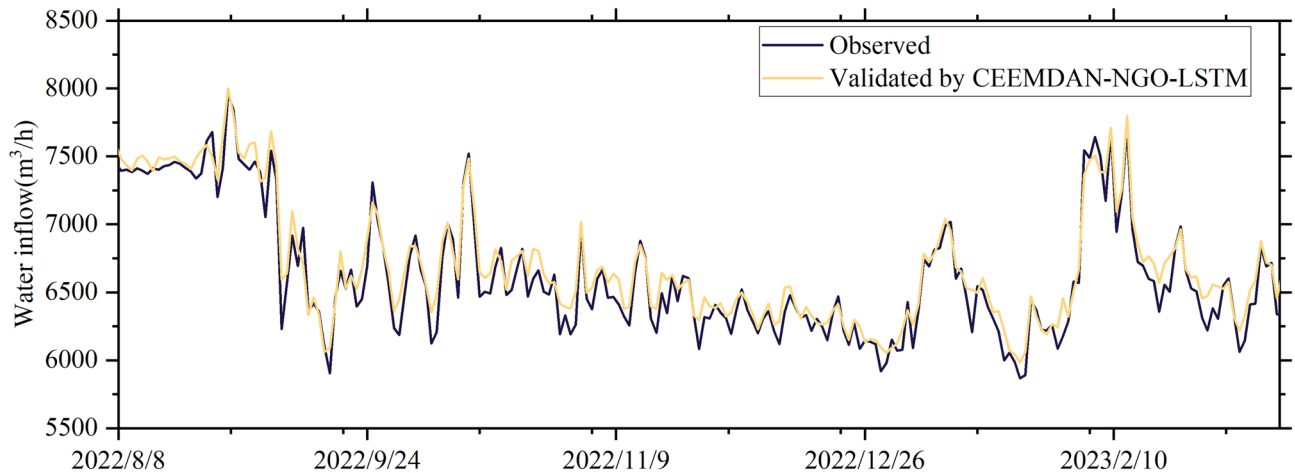


Figure 9. The validation results of the mine water inrush by CEEMDAN-NGO-LSTM.

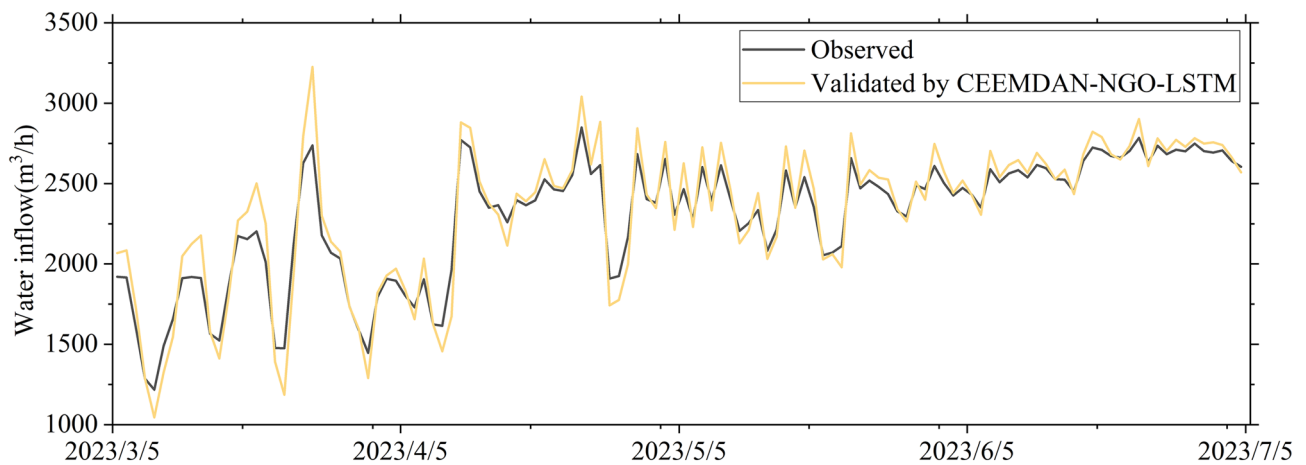


Figure 10. The validation results of the working face mine water inrush by CEEMDAN-NGO-LSTM.

During the validation phase, the equations shifted:

$$\text{LSTM : } Y = 0.935X + 405.915 (R^2 = 0.723)$$

$$\text{CEEMDAN - LSTM : } Y = 1.034X - 292.777 (R^2 = 0.909)$$

$$\text{CEEMDAN - NGO - LSTM : } Y = 1.020X - 206.682 (R^2 = 0.958)$$

These equations demonstrate a strong correlation between the predicted and actual values, with improvement noted in the order presented. However, the scatter plots for the first two models still show some outliers. One

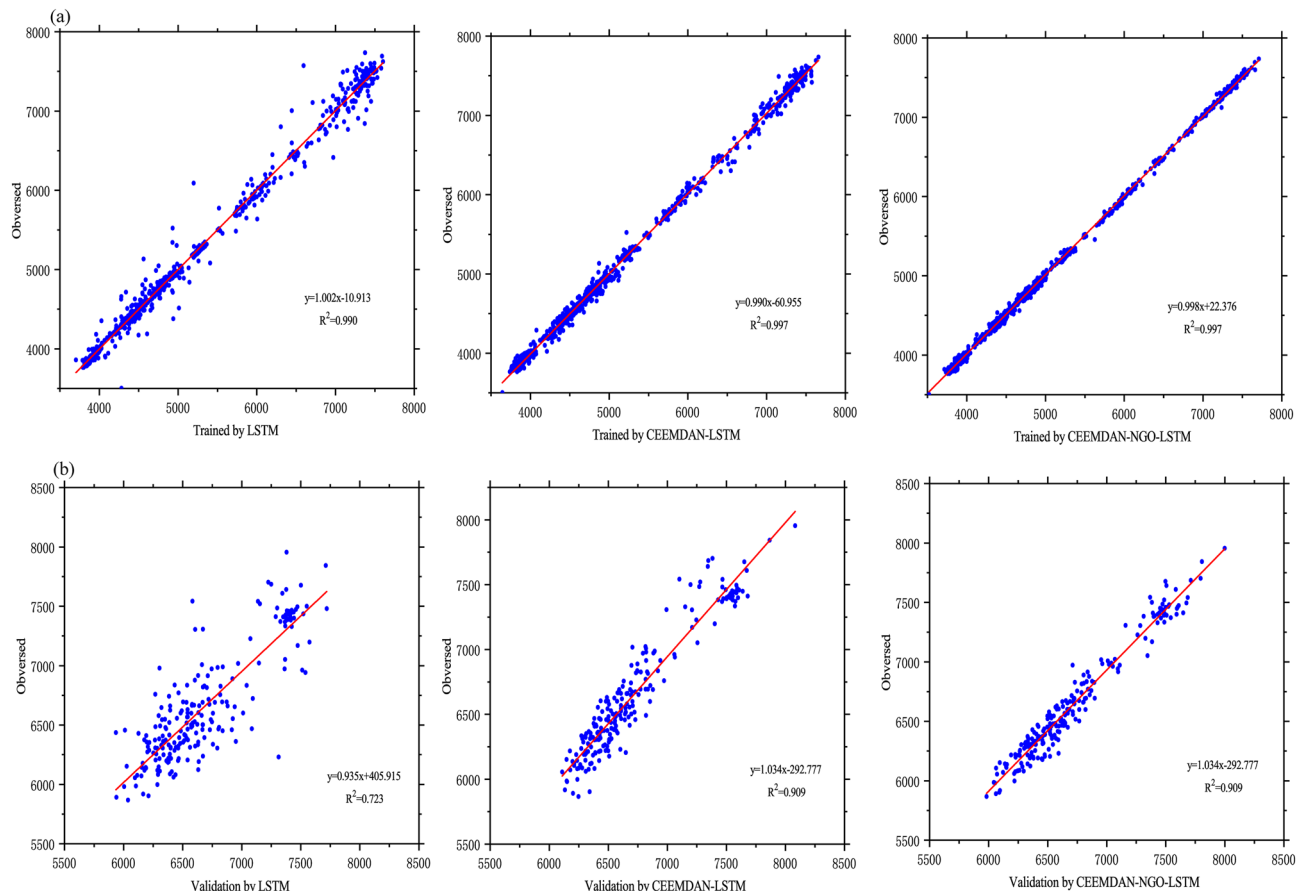


Figure 11. Scatter plot of the simulation vs the observation of the mine water inrush of (a) training and (b) validation.

potential explanation is that the abrupt changes are intrinsic rather than acquired characteristics, while another posits that the series exhibits a high degree of autocorrelation. These findings support the notion that optimization algorithms can significantly enhance feature extraction and learning within decomposition models, thereby reducing the impact of autocorrelation in the series.

Our study employed the LSTM, CEEMDAN-LSTM, and CEEMDAN-NGO-LSTM models to forecast mine water inflow over future intervals of 1, 3, 5, and 7 days as depicted in Fig. 12. Given the brief duration of forecasts, conventional indices like the NSE are not suitable for assessing accuracy. Instead, we used RMSE and MAPE for this purpose. In single-step forecasts, the MAPE values were 0.0650, 0.0460, and 0.0430, respectively, coupled with RMSE of 412.110, 288.230, and 271.430. The CEEMDAN-NGO-LSTM model yielded predictions closest to actual values. For three-step forecasts, MAPE values improved to 0.0290, 0.0160, and 0.0230, with RMSEs of 247.040, 128.170, and 155.437, indicating enhanced accuracy across all models. Still, only the CEEMDAN-NGO-LSTM model accurately captured the actual trend. In the five-step forecasts, the initial volumes of mine water inflow rose, followed by stable fluctuations. The LSTM model predicted a gentle ascending trend, culminating near the interval's maximum value. The CEEMDAN-LSTM model showed an ascending-descending-ascending pattern, but with notable fluctuations and some predictions not aligning with the actual trend. In contrast, the CEEMDAN-NGO-LSTM model closely matched the natural trend, with predictions slightly higher than actual values, providing a solid foundation for mine water hazard prevention. This model's MAPE values were 0.0270, 0.0290, and 0.0160, with RMSEs of 214.060, 198.020, and 122.790, respectively. For seven-step forecasts, the MAPE values were 0.0510, 0.0460, and 0.0460, with RMSEs of 410.100, 394.370, and 382.370, respectively. Again, the CEEMDAN-NGO-LSTM model outperformed the others. However, predictions for the last two days deviated significantly from the actual trend, impairing forecast accuracy. In conclusion, the CEEMDAN-NGO-LSTM model, utilizing the Northern Goshawk Optimization algorithm, achieves superior accuracy and trend prediction in five-step forecasts. This model effectively navigates the complexities of nonlinear, non-stationary signals such as mine water inflow. Additionally, the predicted water inflow trend for the next five days at the working face is highly consistent with the actual trend (Fig. 13), with RMSE and MAPE values of 236.580 and 9.836%, respectively. These results further substantiate that optimization algorithms can enhance the extraction and learning of features within decomposition models, mitigating the effects of series autocorrelation.

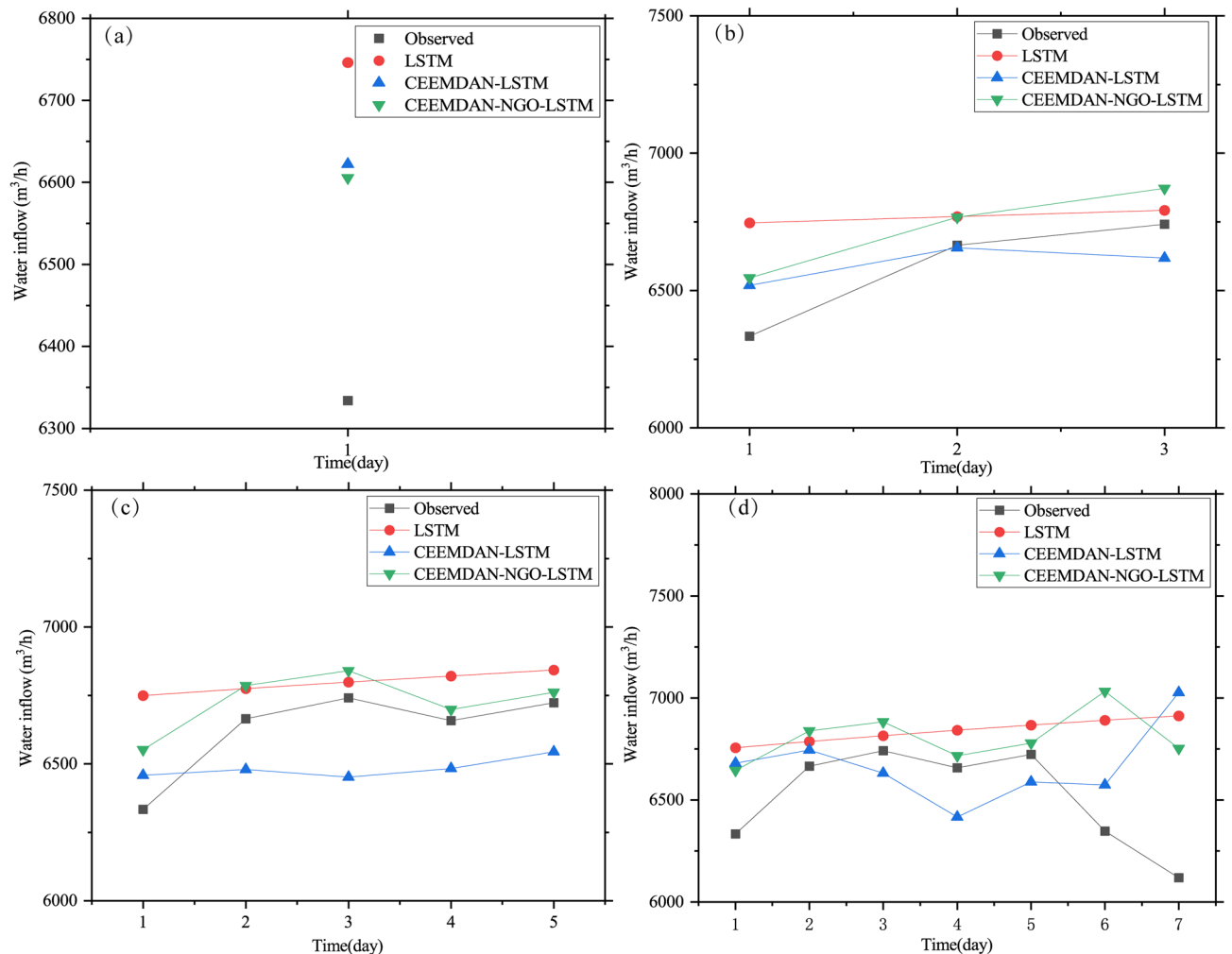


Figure 12. Comparison chart of predicted versus actual values of mine water inflow for the future (a) 1 day, (b) 3 days, (c) 5 days, (d) 7 days.

Conclusion

This study introduces and applies a novel "decomposition-optimization-prediction" model to address the complex, nonlinear, and non-stationary challenges inherent in predicting mine water inflow volumes. We employed the newly developed CEEMDAN-NGO-LSTM model, which features exceptional decomposition preprocessing capabilities and sophisticated memory functions. This model effectively tackles the challenges of predicting mine water inflow. The principal conclusions are as follows:

- (1) The single time series prediction model exhibits optimal performance when employing a sliding input step of 3 steps and a maximum of 400 epochs, achieving MAE, MARE, RMSE, and NSE values of 186.468, 2.825, 255.364, and 0.724, respectively.
- (2) In comparison to the LSTM and CEEMDAN-LSTM models, the CEEMDAN-NGO-LSTM model demonstrates average performance improvements of 44.950% and 19.400%, respectively. It excels in capturing abrupt changes in data and minimizing lags, underscoring the pivotal role of the CEEMDAN and NGO algorithms in enhancing the accuracy of predictions for mine water inflow.
- (3) The CEEMDAN-NGO-LSTM model forecasts mine water inflow for the next five days with high accuracy, closely mirroring actual trends and achieving the lowest RMSE value of 122.790. Thus, this model provides a solid foundation for the effective prevention and control of mine water hazards.

Although the current CEEMDAN-NGO-LSTM model is proficient in predicting mine water inflow, it necessitates additional testing across diverse datasets to confirm its robustness. Future research will focus on the

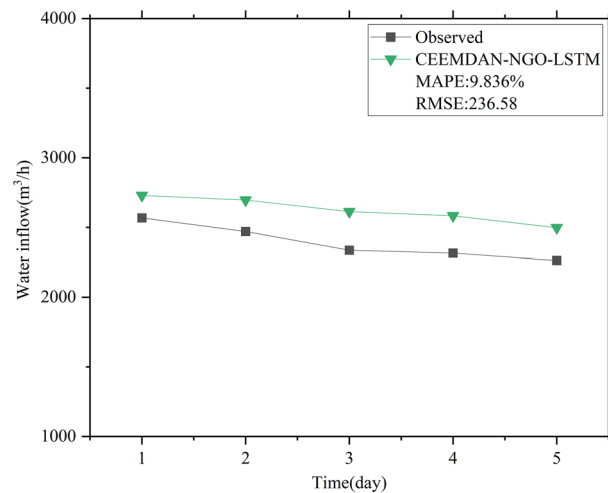


Figure 13. Comparison chart of predicted versus actual values of working face mine water inflow for the future 5 days.

potential for real-time implementation and integration of the model into existing systems. By developing interfaces and optimization schemes, the model can be seamlessly integrated with current mine monitoring and management systems. Furthermore, advanced predictive visualization tools will be developed by designing intuitive visualization interfaces. These tools will utilize various methods to display predictive data and trends, enabling users to customize views according to specific needs, thereby providing more targeted predictive insights. These research initiatives aim to further enhance the accuracy and stability of mine water inflow predictions, offering more scientific and precise tools for the field.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 22 April 2024; Accepted: 18 July 2024

Published online: 01 August 2024

References

- Dong, W. W., Jie, L. S. & Xi, H. J. Analysis of the main global coal resource countries' supply-demand structural trend and coal industry outlook. *Min. Mag.* **24**, 5–9 (2015).
- Qu, X., Yu, X., Qu, X., Qiu, M. & Gao, W. Gray evaluation of water inrush risk in deep mining floor. *ACS Omega*. **6**, 13970–13986 (2021).
- Qu, X., Shi, L. & Han, J. Preventing water-inrush from floor in coal working face with paste-like backfill technology. *Sci. Rep.* **13**, 15947 (2023).
- Du, Z., Wu, Q., Zhao, Y., Zhang, X. & Yao, Y. A multi-constraint and multi-objective optimization layout method for a mine water inrush monitoring network. *Sci. Rep.* **13**, 11817 (2023).
- Hu, W. & Zhao, C. Evolution of water hazard control technology in China's coal mines. *Mine Water Environ.* **40**, 334–344 (2021).
- Xu, J. & Xu, J. Y. Y. Prediction of the maximum water inflow in Pingdingshan No. 8 mine based on grey system theory. *J. Coal Sci. Eng. (China)* **18**, 55–59 (2012).
- Li, B., Li, T., Zhang, W., Liu, Z. & Yang, L. Multisource information risk evaluation technology of mine water inrush based on VWM: a case study of Weng'an coal mine. *Geofluids* **2021**, 1–12 (2021).
- Li, Z., Gao, T., Guo, C. & Li, H. A. A gated recurrent unit network model for predicting open channel flow in coal mines based on attention mechanisms. *IEEE Access*. **8**, 119819–119828 (2020).
- Zhai, H. *et al.* Prediction of the mine water inflow of coal-bearing rock series based on well group pumping. *Water*. **15**, 3680 (2023).
- Marinelli, F. & Niccoli, W. L. Simple analytical equations for estimating ground water inflow to a mine pit. *Ground Water*. **38**, 311–314 (2010).
- Press, P., Incorporated. Recommendations for the treatment of water inflows and outflows in operated underground structures. *Tunn. Undergr. Space Technol.* **4**, 343–407 (1989).
- Zhang, B., Hu, S. & Li, M. Comparative study of multiple machine learning algorithms for risk level prediction in goaf. *Heliyon*. **9**, e19092 (2023).
- Miladinović, B., Vakanjac, V., Bukumirović, D., Dragišić, V. & Vakanjac, B. Simulation of mine water inflow: case study of the Štavalj coal mine (southwestern Serbia). *Arch. Min. Sci.* **40**, 955–969 (2015).
- Yang, X. *et al.* Prediction of mine water flow based on singular spectrum analysis and multiple time-series coupled model. *Arabian J. Geosci.* **14**, 1–18 (2021).
- Al-Qaness, M. A. *et al.* Predicting CO₂ trapping in deep saline aquifers using optimized long short-term memory. *Environ. Sci. Pollut. Res.* **30**, 33780–33794 (2023).
- Fischer, T. & Krauss, C. Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.* **270**, 654–669 (2018).
- Ma, J. *et al.* Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Sci. Total Environ.* **705**, 135771 (2020).

18. Sahoo, B. B., Panigrahi, B., Nanda, T., Tiwari, M. K. & Sankalp, S. Multi-step ahead urban water demand forecasting using deep learning models. *SN Comput. Sci.* **4**, 752 (2023).
19. Cheng, M., Fang, F., Kinouchi, T., Navon, I. & Pain, C. Long lead-time daily and monthly streamflow forecasting using machine learning methods. *J. Hydrol.* **590**, 125376 (2020).
20. Zhang, D., Lindholm, G. & Ratnaweera, H. Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. *J. Hydrol.* **556**, 409–418 (2018).
21. Zhang, T. *et al.* Research on gas concentration prediction models based on LSTM multidimensional time series. *Energies*. **12**, 161 (2019).
22. Shu, L., Liu, Z., Wang, K., Zhu, N. & Yang, J. Characteristics and classification of microseismic signals in heading face of coal mine: implication for coal and gas outburst warning. *Rock Mech. Rock Eng.* **55**, 6905–6919 (2022).
23. Yin, H. *et al.* Predicting mine water inrush accidents based on water level anomalies of borehole groups using long short-term memory and isolation forest. *J. Hydrol.* **616**, 128813 (2023).
24. Ji, C. *et al.* A multi-scale evolutionary deep learning model based on CEEMDAN, improved whale optimization algorithm, regularized extreme learning machine and LSTM for AQI prediction. *Environ. Res.* **215**, 114228 (2022).
25. Liu, F., Liu, Y., Yang, C. & Lai, R. A new precipitation prediction method based on CEEMDAN-IWOA-BP coupling. *Water Resour. Manag.* **36**, 4785–4797 (2022).
26. Liu, H., Mi, X. & Li, Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. *Energy Convers. Manag.* **159**, 54–64 (2018).
27. An, L. *et al.* Simulation of karst spring discharge using a combination of time–frequency analysis methods and long short-term memory neural networks. *J. Hydrol.* **589**, 125320 (2020).
28. Zhou, F., Huang, Z. & Zhang, C. Carbon price forecasting based on CEEMDAN and LSTM. *Appl. Energy* **311**, 118601 (2022).
29. Li, H., Zhang, X., Sun, S., Wen, Y. & Yin, Q. Daily flow prediction of the Huayuankou hydrometeorological station based on the coupled CEEMDAN–SE–BiLSTM model. *Sci. Rep.* **13**, 18915 (2023).
30. Nuri Balov, M. & Altunkaynak, A. The impacts of climate change on the runoff volume of Melen and Munzur Rivers in Turkey based on calibration of WASMOD model with multiobjective genetic algorithm. *Meteorol. Atmos. Phys.* **132**, 85–98 (2020).
31. Chen, S. & Dong, S. A sequential structure for water inflow forecasting in coal mines integrating feature selection and multi-objective optimization. *IEEE Access* **8**, 183619–183632 (2020).
32. Wang, W., Tong, M. & Yu, M. Blood glucose prediction with VMD and LSTM optimized by improved particle swarm optimization. *IEEE Access* **8**, 217908–217916 (2020).
33. Zhou, J. *et al.* Enhancing the performance of tunnel water inflow prediction using Random Forest optimized by Grey Wolf Optimizer. *Earth Sci Inform.* **16**, 2405–2420 (2023).
34. Huang, N. E. *et al.* The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Sci.* **454**, 903–995 (1998).
35. Zhaohua, W. U. & Huang, N. E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **1**, 1–41 (2011).
36. Zhang, W. *et al.* A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. *Energy Convers. Manag.* **136**, 439–451 (2017).
37. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
38. Sahoo, B. B., Sankalp, S. & Kisi, O. A novel smoothing-based deep learning time-series approach for daily suspended sediment load prediction. *Water Resour. Manag.* **37**, 4271–4292 (2023).
39. Qiao, M., Yan, S., Tang, X. & Xu, C. Deep convolutional and LSTM recurrent neural networks for rolling bearing fault diagnosis under strong noises and variable loads. *IEEE Access* **8**, 66257–66269 (2020).
40. Sahoo, B. B., Jha, R., Singh, A. & Kumar, D. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophysica* **67**, 1471–1481 (2019).
41. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D: Non-linear Phenomena* **404**, 132306 (2020).
42. Swagatika, S., Paul, J. C., Sahoo, B. B., Gupta, S. K. & Singh, P. Improving the forecasting accuracy of monthly runoff time series of the Brahmani River in India using a hybrid deep learning model. *J. Water Climate Change* **15**, 139–156 (2024).
43. Dehghani, M., Hubalovsky, S. & Trojovsky, P. Northern Goshawk optimization: A new swarm-based algorithm for solving optimization problems. *IEEE Access* **9**, 162059–162080 (2021).
44. El-Dabah, M. A., El-Sehiemy, R. A., Hasanien, H. M. & Saad, B. Photovoltaic model parameters identification using Northern Goshawk Optimization algorithm. *Energy* **262**, 125522 (2023).
45. Liang, Y., Hu, X., Hu, G. & Dou, W. An enhanced Northern Goshawk optimization algorithm and its application in practical optimization problems. *Mathematics* **10**, 4383 (2022).
46. Wang, W. C., Xu, D. M., Chau, K. W. & Chen, S. Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. *J. Hydroinformatics*. **15**, 1377–1390 (2013).
47. Gao, S., Huang, Y., Zhang, S., Han, J. & Lin, Q. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *J. Hydrol.* **589**, 125188 (2020).
48. Swanson, D. A., Tayman, J. & Bryan, T. MAPE-R: a rescaled measure of accuracy for cross-sectional subnational population forecasts. *J. Popul. Res.* **28**, 225–243 (2011).
49. Meng, E. *et al.* A robust method for non-stationary streamflow prediction based on improved EMD-SVM model - ScienceDirect. *J. Hydrol.* **568**, 462–478 (2019).
50. Fang, L. & Shao, D. Application of long short-term memory(LSTM) on the prediction of rainfall-runoff in Karst area. *Front.* **9**, 685 (2022).

Acknowledgements

This work was financially supported by grants from the Provincial Key R&D Program of Shaanxi [grant number 2021ZDLSF05-03].

Author contributions

J.X.B.: Conceptualization, Methodology, Data curation, Writing—Original draft. X.Y.Q. and X.D.M.: Writing—review & editing, Supervision, Funding acquisition. T.H. and .D.J.R.: Writing—review & editing, Funding acquisition. C.S.L. and J.M.: Writing—review, Supervision. Y.W. and J.Y.W.: Data curation & editing. X.W.L.: Writing—review.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-67962-2>.

Correspondence and requests for materials should be addressed to X.Q. or X.M.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024