buildings

MDPI

Article

Temperature Prediction of Mass Concrete During the Construction with a Deeply Optimized Intelligent Model

Fuwen Zheng [1], Shiyu Xia [1], Jin Chen [2], Dijia Li [1], Qinfeng Lu [1], Lijin Hu [1], Xianshan Liu [3,*], Yulin Song [3] and Yuhang Dai [3]

[1] Construction Branch of Chongqing State Grid Power Company, Chongqing 401121, China; bbzfw@cq.sgcc.com.cn (F.Z.); pipige@cq.sgcc.com.cn (S.X.); lidijia@cq.sgcc.com.cn (D.L.); luqinfeng@cq.sgcc.com.cn (Q.L.); hu-lj@cq.sgcc.com.cn (L.H.)
[2] Chongqing State Grid Power Company, Chongqing 400015, China; xscj@cq.sgcc.com.cn
[3] School of Civil Engineering, Chongqing University, Chongqing 400045, China; 202316131352@stu.cqu.edu.cn (Y.S.); 202416131176t@stu.cqu.edu.cn (Y.D.)
* Correspondence: lzmoumou@163.com

Abstract

In the construction of ultra-high voltage (UHV) transformation substations, mass concrete is highly susceptible to temperature-induced cracking due to thermal gradients arising from the disparity between internal hydration heat and external environmental conditions. Such cracks can severely compromise the structural integrity and load-bearing capacity of foundations, making accurate temperature prediction and effective thermal control critical challenges in engineering practice. To address these challenges and enable real-time monitoring and dynamic regulation of temperature evolution, this study proposes a novel hybrid forecasting model named CPO-VMD-SSA-Transformer-GRU for predicting temperature behavior in mass concrete. First, sine wave simulations with varying sample sizes were conducted using three models: Transformer-GRU, VMD-Transformer-GRU, and CPO-VMD-SSA-Transformer-GRU. The results demonstrate that the proposed CPO-VMD-SSA-Transformer-GRU model achieves superior predictive accuracy and exhibits faster convergence toward theoretical values. Subsequently, four performance metrics were evaluated: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). The model was then applied to predict temperature variations in mass concrete under laboratory conditions. For the univariate time series at Checkpoint 1, the evaluation metrics were MAE: 0.033736, MSE: 0.0018812, RMSE: 0.036127, and R^2: 0.98832; at Checkpoint 2, the values were MAE: 0.016725, MSE: 0.00091304, RMSE: 0.019114, and R^2: 0.96773. In addition, the proposed model was used to predict the temperature in the rising stage, indicating high reliability in capturing nonlinear and high-dimensional thermal dynamics in the whole construction process. Furthermore, the model was extended to multivariate time series to enhance its practical applicability in real-world concrete construction. At Checkpoint 1, the corresponding metrics were MAE: 0.56293, MSE: 0.34035, RMSE: 0.58339, and R^2: 0.95414; at Checkpoint 2, they were MAE: 0.85052, MSE: 0.78779, RMSE: 0.88757, and R^2: 0.91385. These results indicate significantly improved predictive performance compared to the univariate configuration, thereby further validating the accuracy, stability, and robustness of the multivariate CPO-VMD-SSA-Transformer-GRU framework. The model effectively captures complex temperature fluctuation patterns under dynamic environmental and operational conditions, enabling precise, reliable, and adaptive temperature forecasting. This comprehensive analysis establishes a robust methodological foundation for advanced temperature prediction and optimized thermal management strategies in real-world civil engineering applications.

check for updates

Academic Editor: Humberto Varum

Received: 17 October 2025
Revised: 24 November 2025
Accepted: 25 November 2025
Published: 4 December 2025

Citation: Zheng, F.; Xia, S.; Chen, J.; Li, D.; Lu, Q.; Hu, L.; Liu, X.; Song, Y.; Dai, Y. Temperature Prediction of Mass Concrete During the Construction with a Deeply Optimized Intelligent Model. *Buildings* **2025**, *15*, 4392. https://doi.org/10.3390/buildings15234392

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Keywords: ultra-high voltage (UHV) transformation substations; mass concrete; construction phase; temperature series; CPO-VMD-SSA-Transformer-GRU model; temperature prediction

1. Introduction

In recent years, the State Grid Corporation of China (SGCC) has been actively implementing the national energy security strategy of "Four Revolutions and One Cooperation", promoting the high-quality development of ultra-high voltage (UHV) transmission infrastructure and accelerating the construction of power systems. The Sichuan–Chongqing 1000 kV UHV transmission project [1], commissioned in 2024, represents a landmark initiative in establishing a modern energy architecture in Southwest China and contributes substantially to goals of "carbon peak and carbon neutrality". Four new UHV substations along the transmission corridor all involve large-scale concrete raft foundations, such as single-pour concrete volumes of considerable size, and a great amount of high cementitious material content [2], resulting in significant heat of hydration release within a short period, generating considerable thermal gradients between the core and surface regions of the concrete. The thermal stress induced extensive cracking, critically compromising the structural integrity and load-bearing capacity of the foundations. Accurate temperature prediction and effective thermal control are therefore essential to mitigate excessive early-age temperature rise, enhance cracking resistance, and improve construction efficiency in mass concrete projects [3,4]. Consequently, precise temperature monitoring and proactive thermal management during the construction of mass concrete are of considerable engineering significance.

Current research encompasses theoretical analysis, experimental investigation, and numerical simulation. Liu et al. [5] examined the mechanisms and effectiveness of multi-admixtures in suppressing temperature rise in mass concrete; Yang et al. [6] optimized cooling pipe systems based on the hydration heat distribution characteristics of ultra-high-strength concrete for improved thermal control; Klemczak et al. [7] proposed a machine learning-based (ML) approach to predicting temperature rise and thermal gradients in mass concrete based on the dataset generated using the analytical CIRIA C766 method and compared it with linear regression, a decision tree, and XGBoost to confirm the suitability of ML models for reliable thermal response prediction, providing a usable alternative to conventional methods of temperature prediction in early age mass concrete. Hossein Kabir et al. [8] introduced a custom computer vision model trained on 6234 images, consisting of 4000 real and 2234 synthetic images, which automatically detected the water level in prismatic samples absorbing water, and an FPN-based water level detection model compared with other networks was utilized to present its superior computational efficiency to ensure accurate water level detection. Peng et al. [9] proposed numerical methods to simulate concrete temperature evolution after construction, thereby deriving patterns of temperature and stress development during the concrete curing process. It can be observed that although numerical simulation achieves relatively high computational accuracy, it requires high-quality, fine meshes that demand substantial computational resources, leading to low computational efficiency and an inability to enable real-time, rapid prediction. Given the limitations of conventional analytical methods, such as high costs, time consumption, and restricted applicability, an efficient and accurate temperature prediction approach is urgently needed for mass concrete in UHV substations during construction to address these critical challenges.

With the rapid advancement of big data and artificial intelligence technologies, data-driven machine learning approaches have garnered increasing attention for temperature prediction in mass concrete, offering a promising avenue for precise thermal regulation and crack propagation prevention in the construction of UHV substation foundations.

Xu et al. [10] pioneered a neural network model to predict peak temperatures in concrete pouring bays, utilizing a multi-parametric input vector comprising pouring temperature, cooling water flow rate, cooling water temperature, and ambient air temperature, with the measured maximum temperature serving as the output response. Zheng et al. [11] proposed an RBF-PSO hybrid framework for efficient inversion analysis of hydration heat parameters in mass concrete temperature fields, reducing reliance on costly finite element simulations while maintaining global search capabilities based on validation through multi-sensor field data, revealing computational intelligence applications in large-scale concrete structures and multi-physics coupling problems. Zhou et al. [12] employed the random forest algorithm to develop a high-fidelity predictive model for the internal peak temperature of concrete pouring bays, subsequently validating its effectiveness through application in a large-scale concrete dam project. Wang et al. [13] and Guo and Zhou [14] integrated numerical simulations with neural networks to establish a hybrid temperature prediction model for mass concrete, demonstrating strong agreement between predicted trends and empirical observations. These studies collectively provide substantial support for the development of deep learning-based temperature prediction models in mass concrete construction [15,16]. Given the inherent non-stationarity and stochastic fluctuations of temperature time series during the construction phase, advanced preprocessing techniques can be leveraged to decompose the high-dimensional and nonlinear characteristics embedded within raw time series data [17]. This decomposition strategy effectively overcomes the limitations of conventional methods, such as inadequate feature extraction and suboptimal predictive accuracy, when training and testing models on unprocessed data. Furthermore, hybrid deep learning architectures that explicitly capture temporal dependencies can be designed to address the shortcomings of traditional neural networks, including sensitivity to weight initialization and susceptibility to convergence at local optima. Accordingly, in view of the characteristics of temperature time series in typical mass concrete applications, Variational Mode Decomposition (VMD) [18–20] is incorporated due to its effectiveness in handling non-stationary signals and noise. Xing et al. [19] have proposed a combined prediction model based on Variational Mode Decomposition (VMD) and Gradient Boosting Regression Tree (GBRT) to exhibit superior short-term wind power prediction performance, with a Mean Squared Error, Mean Absolute Error, and Coefficient of Determination of 0.0244, 0.1185, and 0.9821. Xu et al. [20] has developed a hybrid VMD-BiTCN-Psformer model and compared it with four additional models, demonstrating the greatest level of forecast accuracy among all the models, with R^2 increased by 22.27%, 12.38%, 8.93%, and 2.59%, respectively. In addition, the self-attention mechanism of the Transformer model is introduced to achieve rapid parallel processing of sequential data [21–25]. Zheng et al. [23] addressed an innovative, cutting-edge hybrid model that integrates Gated Recirculation Unit (GRU) and Transformer technologies to demonstrate a significant advantage in predicting soil moisture content forecasts with enhanced precision and reliability when compared with models such as SVR, KNN, GBDT, XGBoost, and Random Forest, as well as DNN, CNN, and standalone GRU-branch and Transformer-branch models. When coupled with the Gated Recurrent Unit (GRU) [26–29], a hybrid temperature prediction model (VMD-Transformer-GRU) was established, enabling accurate modeling of complex, nonlinear time series dynamics. Hao et al. [26] constructed a new deep learning model called Attention-GRU to predict lake surface water temperature in Qinghai Lake in China. Compared with the seven baseline models, the proposed model achieved the most accurate prediction results and significantly outperformed the Air-Water model. Dai et al. [29] established an accurate Transformer loss prediction model based on the PCA-SSA-BP, compared with the Transformer loss prediction model based on the BP network, GA-BP, PSO-BP, and SSA-BP; the former model improves accuracy by using actual data and aids low-voltage recovery

in the distribution network. Based on above technologies, the Transformer self-attention mechanism may be applied to model long-range dependencies in the VMD-decomposed subsequences, while the GRU captures temporal patterns and long-term dependencies within each component. This combination can effectively mitigate issues such as gradient vanishing and gradient explosion. To further enhance the predictive performance of the VMD-Transformer-GRU (VMD-Tr-GRU) model, meta-heuristic optimization algorithms are employed. Specifically, the Crested Porcupine Optimizer (CPO) [30] is utilized to optimize the key parameters of VMD, thereby better addressing the nonlinear and multimodal nature of the temperature time series. Additionally, the Sparrow Search Algorithm (SSA) [29] is applied to optimize the model hyperparameters, significantly improving convergence speed, global search capability, and prediction accuracy. This dual optimization strategy effectively resolves the challenges associated with parameter tuning in complex hybrid models, ensuring both stability and generalization in practical engineering applications.

A 1000 kV UHV power transmission and transformation was used as a case study, with temperature time series selected from the construction phase of mass concrete. Based on CPO-optimized VMD technology, the data was decomposed into subsequences. The multi-head attention mechanism of the Transformer was employed to extract time series, and the SSA was introduced to optimize corresponding hyperparameters, forming a CPO-VMD-SSA-Transformer-GRU (CPO-VMD-SSA-Tr-GRU) prediction model. The temperature predictions by the above methods were compared for sinusoidal time series functions by the Transformer-GRU (Tr-GRU), VMD-Tr-GRU, and CPO-VMD-SSA-Tr-GRU models, which demonstrated the feasibility and superiority of the CPO-VMD-SSA-Tr-GRU model for long-term temperature prediction. The model was then applied to a single time series temperature prediction for mass concrete pouring tests, with predicted values closely matching actual monitoring data, further validating its reliability. In addition, multi-variable time series temperature prediction showed higher accuracy compared to single time series predictions, confirming the reliability of the CPO-VMD-SSA-Tr-GRU model for deep learning of complex temperature time series. The above investigations provide a new approach for temperature prediction during mass concrete construction and offer technical support for the optimization of temperature control in different construction environments.

2. Construction of the Deeply Optimized CPO-VMD-SSA-Tr-GRU Model

2.1. CPO-Optimized VMD for Time Series Data Processing

The Variational Mode Decomposition (VMD) method enables adaptive decomposition of nonlinear and non-stationary signals by formulating a variational optimization problem to estimate the center frequency of each intrinsic mode function (IMF). Compared with the Empirical Mode Decomposition (EMD) method, VMD effectively alleviates the issue of mode mixing, thereby enhancing signal decomposition reliability. In practical applications involving real time series, the number of decomposition modes (K) directly determines the number of resulting IMFs; an excessively large K may lead to mode mixing, while an insufficiently small K may result in the loss of critical signal components. Simultaneously, the quadratic penalty factor α controls the spectral bandwidth of the decomposed modes. Therefore, the selection of these two parameters plays a crucial role in determining the quality and accuracy of signal decomposition. To overcome the limitations associated with empirical and manual parameter tuning, the Crested Porcupine Optimizer (CPO), a meta-heuristic algorithm inspired by the defensive behaviors of the crested porcupine, including visual, auditory, olfactory, and physical attack strategies, is employed to optimize the key parameters of VMD. The CPO algorithm exhibits high computational efficiency, rapid convergence, and strong robustness, making it particularly effective in addressing the

nonlinearity and multimodality inherent in temperature time series and enabling accurate extraction of temperature variation patterns.

The procedure for decomposing temperature time series based on CPO-optimized VMD (CPO-VMD) is as follows:

(1) Construct the variational constraint model:

$$\begin{cases} \min\limits_{\{g_k\},\{w_k\}} \sum\limits_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \dfrac{j}{\pi t} \right) * g_k(t) \cdot e^{-jw_k t} \right] \right\|_2^2 \\ s.t. \sum\limits_{k=1}^{K} g_k(t) = f(t) \end{cases} \tag{1}$$

In which $f(t)$ denotes the first high-frequency component obtained from the initial decomposition; $\delta(t)$ represents the impulse signal; ω_k indicates the central frequency of each component; $g_k(t)$ refers to the modal component; and * denotes the convolution operator.

(2) Initialize the parameters and search range of the Crested Porcupine Optimizer (CPO) and set the initial number of IMF components K, the quadratic penalty factor α, and the maximum number of iterations.

(3) Update the population positions based on the four defense strategies in the Crested Porcupine Optimizer (CPO) to effectively explore the search space.

(4) Perform iterative optimization until the optimal number of IMF components K and the optimal quadratic penalty factor α are obtained, thereby establishing the CPO optimized VMD modal.

2.2. Transformer Network Structure

The Transformer network encoder is part of an encoder–decoder architecture built upon a self-attention mechanism. The encoder consists of multiple stacked, identical encoding layers, each comprising two primary sub-layers: a multi-head self-attention mechanism and a fully connected feedforward neural network. Residual connections are applied around both sub-layers, followed by layer normalization to stabilize and accelerate training. The self-attention mechanism reduces dependence on sequential processing and external contextual assumptions, enabling the model to effectively capture long-range dependencies and intrinsic correlations within the input data. Specifically, it computes attention weights dynamically across the input sequence, allowing the model to adaptively assign importance to different time steps while incorporating relevant influencing factors before and after each point in the time series.

Assuming each input vector maps to three different spaces to generate the query vector, the key vector, and the value vector, these vectors are linearly mapped to three matrices, respectively, to obtain the sample generation W^Q, W^K, and W^V:

$$\begin{Bmatrix} Q = X \cdot W^Q \\ K = X \cdot W^K \\ V = X \cdot W^V \end{Bmatrix} \tag{2}$$

The scaled dot product is used as the attention scoring function:

$$Attention(Q, K, V) = Soft\max\left(\frac{Q_i K_i^T}{\sqrt{D_k}} \right) \tag{3}$$

In which Q_i, K_i^T are the score of the sample in each vector, $\sqrt{D_k}$ is the optimization of the training gradient, and $Soft\max$ is the function normalized by column.

Then, the calculations are normalized through the residual network, and corresponding results are given as inputs to the feedforward neural network.

The multi-head attention mechanism is constructed by concatenating multiple parallel self-attention modules, each referred to as a self-attention head. Each head is trained to focus on different representational subspaces. For instance, some heads may capture local temporal patterns, while others attend to global sequence structures. This design enables the model to simultaneously extract diverse features from various perspectives without interference among heads, thereby enhancing its capacity to handle complex sequential tasks. Specifically, the input is first linearly projected into lower-dimensional spaces to generate query, key, and value vectors for each head. These are then processed through a scaled dot-product attention mechanism to compute similarity patterns across different positions in the sequence. The outputs from all heads are concatenated and linearly transformed back to the original dimension, allowing the model to effectively integrate both local and global contextual information and extract comprehensive, high-level features.

$$Multi(Q, K, V) = Contact(head_1, \cdots, head_n) W^0 \tag{4}$$

In which $head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right)$.

The output vector can be calculated from two linear variations of the fully connected feedforward network:

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{5}$$

2.3. GRU Network

The GRU (Gated Recurrent Unit) network uses reset gates and update gates to control information transmission, selectively memorizes state information, and incor-porates input information to train the model through internal recurrent units, better capturing relationships in long time series and reducing problems such as vanishing gradients. A typical GRU network propagates in the direction of sequence transport, and the hidden layer units at the moment t are

$$R_t = \sigma(x_t W_{xr} + H_{t-1} W_{hr}) Z_t = \sigma(x_t W_{xz} + H_{t-1} W_{hz}) \tag{6}$$

$$\begin{aligned} \widetilde{H}_t &= \tan \mathrm{h}[x_t W_{xh} + (R_t \odot H_{t-1}) W_{h\mathrm{h}}] \\ H_t &= Z_t \odot H_{t-1} + (1 - Z_t) \odot \widetilde{H}_t \end{aligned} \tag{7}$$

In which σ is the sigmoid function, which changes the data to [0, 1]; R_t, Z_t are the reset gate and the update gate, respectively; W_{xr}, W_{hr}, W_{xz}, W_{hz}, W_{xh}, $W_{h\mathrm{h}}$ are the weights matrix $\widetilde{H}_t$, H_t are the candidate hidden layer state information and output result, respectively; and $\odot$ is the Hadamar product, representing the product of the elements at the corresponding positions in the matrix.

2.4. Transformer-GRU Model Structure

The temperature time series is used as the input sequence, and feature extraction and spatial relationship modeling are performed using the Transformer network. The resulting feature sequence is input into the GRU network for time series modeling and prediction to establish the Transformer-GRU model for temperature prediction, shown in Figure 1.

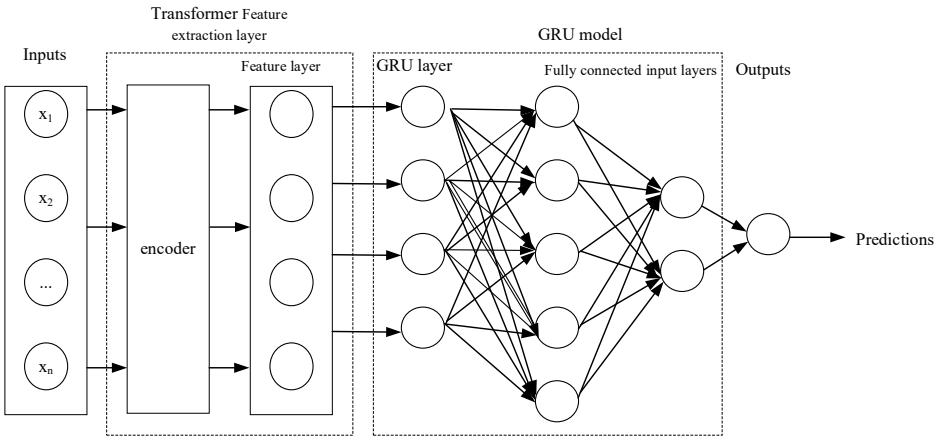


Figure 1. Structure of the Transformer-GRU model.

(1) Input layer: Normalize the temperature time series and apply it in the model inputs. Assume the length of the data is N, and describe as $X = [x_1, x_2, \cdots, x_N]$.

(2) Transformer feature extraction layer: Consists of position encoding, multi-head attention mechanism, and feedforward neural network. Position information is labeled for each normalized data, representing different semantic information:

$$\begin{aligned} PE_{(pos,2i)} &= \sin\left(\frac{pos}{10000^{\frac{2i}{d_{moder}}}}\right) \\ PE_{(pos,2i+1)} &= \cos\left(\frac{pos}{10000^{\frac{2i}{d_{moder}}}}\right) \end{aligned} \tag{8}$$

In which pos is the position of the input sequence, i is the dimension, and d_{moder} is the size of the vector dimension.

Use residual connections and layer normalization based on the multi-head attention mechanism and feedforward neural network:

$$\begin{aligned} S_{out} &= LN(x + S_{out}(x)) \\ LN(x_i) &= a \times \frac{x_i - u_L}{\sqrt{\varepsilon + \sigma_L^2}} + \beta \end{aligned} \tag{9}$$

In which LN is the normalization of the layer and u_L, σ_L^2 are the mean and variance, respectively.

(3) GRU model: The temperature series after extraction by the Transformer is applied as the input. The layer consists of fully connected input layers and fully connected GRU level output layers. The input fully connected layer is:

$$H = \mathrm{Re}LU(LN + d) \tag{10}$$

Input the combined data as a newly generated feature into the GRU cell layer to obtain the GRU layer output value h.

2.5. The Sparrow Search Algorithm

The Sparrow Search Algorithm (SSA) is a swarm intelligence optimization algorithm proposed in 2020, mainly inspired by the foraging and anti-predation behaviors of sparrows. This method is not limited by the differentiability, derivability, and continuity of the objective function, with advantages of strong global generalization, good stability, and fast convergence. The sparrow population in SSA is divided into discoverers and followers. In each iteration, two ways are applied by the discoverer to update its position.

When the discoverer does not receive a warning signal from the scout, the updated position of discoverer is as shown in Equation (11):

$$P_{i,j}^{t+1} = P_{i,j}^{t} \cdot \exp\left(\frac{-i}{a \cdot epoch}\right), R_2 < ST \tag{11}$$

In which t is the current iteration, $epoch$ represents the maximum number of iterations, $a = [0, 1]$ is the random number, $P_{i,j}^{t}$ is the value of the jth dimension of the ith individual during the iteration, t, d represents the dimension for solving the problem, and $ST = [0.5, 1]$ and $R_2 = [0, 1]$ represent the threshold and the warning values, respectively.

When the discoverer receives the warning signal, knowing that the predator is approaching, the discoverer guides all sparrows to the safe area and updates the position to Equation (12):

$$P_{i,j}^{t+1} = P_{i,j}^{t} + r \cdot D, R_2 \geq ST \tag{12}$$

In which r is a random number that follows a normal distribution and D is a matrix where all elements are 1 describing as $1 \times D$. Predators with better fitness head toward the best discoverer, whose position is updated to Equation (13):

$$P_{i,j}^{t+1} = P_{best}^{t+1} + \left|P_{i,j}^{t} - P_{best}^{t+1}\right| \cdot A^{+} \cdot D, i \leq \frac{n}{2} \tag{13}$$

In which P_{best} represents the best position for the entire population, A is a matrix in which all elements are 1 describing as $1 \times d$, and all elements in the matrix are 1 or -1, $A^{+} = A^{T}\left(AA^{T}\right)^{-1}$. The remaining discoverers with less fitness who have not found food continue to look for food near the next location, and these discoverers update their positions to Equation (14):

$$P_{i,j}^{t+1} = P_{best}^{t+1} + K \cdot \left|\frac{P_{worst}^{t} - P_{i,j}^{t}}{i^2}\right|, i > \frac{n}{2} \tag{14}$$

If the finder in the best position detects danger, it becomes the scourer, and the scourer's position is updated to Equation (15):

$$P_{i,j}^{t+1} = P_{best}^{t+1} + K \cdot \frac{\left|P_{worst}^{t} - P_{i,j}^{t}\right|}{(f_i - f_w) + \varepsilon}, f_i = f_g \tag{15}$$

In which $K = [-1, 1]$ are the random numbers; f_i, f_g, f_w represent the current fitness of the sparrow, the optimal fitness of the population, and the worst fitness, respectively; and $\varepsilon \geq 0$ is a constant. If the scout is not in the best position, the scout moves to the best position to reduce the probability of being preyed upon, and the position is updated to Equation (16):

$$P_{i,j}^{t+1} = P_{i,j}^{t} + \beta \cdot \left|P_{i,j}^{t} - P_{best}^{t}\right|, f_i > f_g \tag{16}$$

2.6. Construction of Deeply Optimized VMD-Transformer-GRU Model

The temperature time series observed during the construction of mass concrete is significantly influenced by climatic conditions, admixtures, and other external factors, exhibiting highly nonlinear and high-dimensional characteristics. To accurately predict temperature variations during the construction phase, the CPO-optimized Variational Mode Decomposition (VMD) method is employed for time series preprocessing, decomposing the original signal into a set of intrinsic mode function (IMF) sub-series. Subsequently, the Transformer architecture, equipped with a multi-head self-attention mechanism, is utilized to extract deep multidimensional features from these decomposed components.

Furthermore, the Gated Recurrent Unit (GRU) model, which demonstrates strong selective memory capability and robustness against gradient explosion and vanishing issues, is integrated to enhance the temporal modeling performance of the network. This synergistic integration gives rise to the VMD-Transformer-GRU hybrid deep learning framework for temperature time series prediction. Given that this hybrid model involves a large number of hyperparameters, empirical parameter setting often leads to suboptimal performance and reduced generalization ability. Therefore, systematic hyperparameter optimization is essential to improving both computational efficiency and predictive accuracy. To this end, the Sparrow Search Algorithm (SSA) is introduced to optimize the key hyperparameters of the VMD-Transformer-GRU (VMD-Tr-GRU) model, resulting in an enhanced architecture termed the CPO-VMD-SSA-Transformer-GRU (CPO-VMD-SSA-Tr-GRU) model. This optimized framework demonstrates superior computational efficiency and prediction accuracy, enabling more precise and reliable forecasting of temperature behavior in mass concrete. The computational workflow of the CPO-VMD-SSA-Tr-GRU model is illustrated in Figure 2.

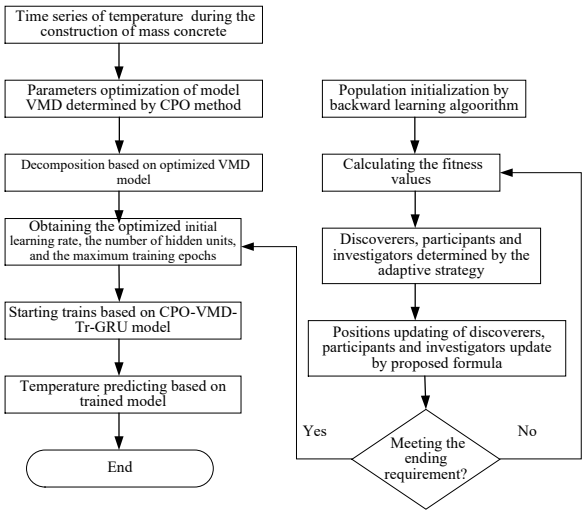


Figure 2. Analysis diagram of the CPO-VMD-SSA-Tr-GRU model.

Step 1: For measuring temperature time series of mass concrete, the Variational Mode Decomposition (VMD) process is optimized using the CPO algorithm by tuning key parameters, such as the IMF component numbers K and the secondary penalty factor α. The time series with these optimized parameters is decomposed into corresponding subsequences, which are subsequently normalized. In the optimization process, set populations as 5, max iterations as 20, and tolerance as 1×10^{-7}.

Step 2: The Transformer architecture is employed to extract features from the preprocessed subsequences, thereby capturing the nonlinear relationships between observed and predicted values. The Gated Recurrent Unit (GRU) model, which possesses selective memory capabilities, is utilized to capture temporal dependencies and dynamically adjust weights, thereby enhancing the visibility of subtle autocorrelations within the time series. The preprocessed data is then applied to the GRU model to construct a composite neural network model named VMD-Transformer-GRU (VMD-Tr-GRU) that integrates both nonlinear and temporal characteristics. Then, the time series is reconstructed to a matrix with self-defined independent variables and dependent variables; thus, the number of input and output layers is also configured accordingly.

Step 3: The Sparrow Search Algorithm (SSA) is introduced to optimize the hyperparameters of the VMD-Tr-GRU model. Key SSA parameters, including the maximum number of iterations (*epoch*), dimension (d), threshold (ST), and warning value R_2, are

initialized. For example, set waning index ST = 0.6, detector ratio PD = 0.7, and proportion of individuals aware of the danger of sparrows SD = 0.2. The algorithm optimizes critical hyperparameters, such as the initial learning rate, the number of hidden units, and the maximum training epochs. The fitness of each sparrow individual is evaluated, and the best position is updated iteratively. The corresponding hyperparameters are imported into the VMD-Tr-GRU model to compute fitness. Once the optimal position is reached, the algorithm terminates; otherwise, the new position is updated as the best position for the next iteration.

Step 4: The hyperparameters obtained through SSA optimization are input into the VMD-Tr-GRU model. The performance of the resulting CPO-VMD-SSA-Tr-GRU model is evaluated using standard evaluation metrics, including the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2), as defined in Equations (17)–(21).

Step 5: The optimized CPO-VMD-SSA-Tr-GRU model is applied to predict each time subsequence. The predicted values of all subsequences are then aggregated to reconstruct the predicted output for overall time series.

$$MAE = (\frac{1}{n})\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{17}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{18}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{19}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{20}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\overline{y}_i - y_i)^2} \tag{21}$$

In which $\hat{y}_i$ is the predicted temperature of the mass concrete during the construction and y_i is the corresponding monitoring temperature of the mass concrete.

3. Verification and Assessment of the Deeply Optimized VMD-Transformer-GRU Model

A sine function $f_{\sin} = \sin(\pi * t + \frac{\pi}{3}) + 0.2$ was selected for analysis, where t represents time, and the time interval was set to 0.01 s to construct the time series. The time series was then proportionally divided into input-output training and testing sample sequences, such as by allocating 90% of the data for training and 10% for testing. An optimized VMD was applied to the time series using an appropriate number of intrinsic mode functions (IMFs) K= 5, and a quadratic penalty factor α= 877. The optimized subsequences are illustrated in Figure 3

To evaluate the validity and predictive accuracy of the CPO-VMD-SSA-Tr-GRU model, several widely used performance metrics were employed, including the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). Specifically, MAE measures the average magnitude of absolute errors, while MAPE quantifies the average relative percentage error, thereby providing greater robustness in error assessment across different scales. MSE and RMSE are particularly effective for evaluating models sensitive to large error magnitudes due to their squared weighting of residuals. R^2, as a standard

metric for assessing the goodness-of-fit in deep learning models, reflects the proportion of variance explained by the model and thus provides insight into predictive accuracy. The optimal number of hidden units, maximum training epochs, and initial learning rate for the standalone Tr-GRU, VMD-Tr-GRU, and CPO-VMD-SSA-Tr-GRU models were systematically compared and analyzed to assess the relative predictive performance of the proposed framework. These results are summarized in Tables 1 and 2, as well as in Figure 4. Additionally, under a Sparrow Search Algorithm (SSA) configuration with a population size of 3 and 5 iterations, a comparative simulation study was conducted between the CPO-VMD-SSA-Tr-GRU model and other benchmark models, with results presented in Table 2 and Figure 5.

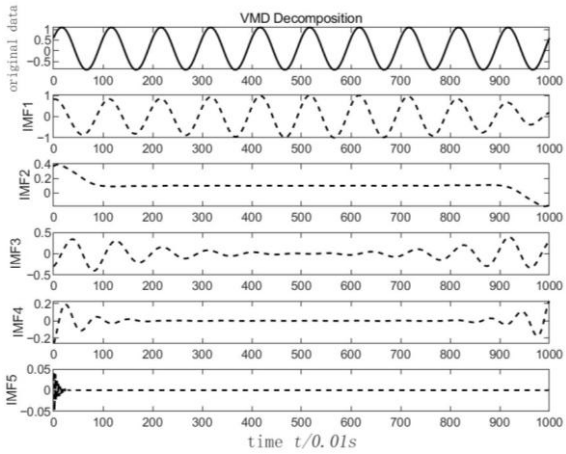


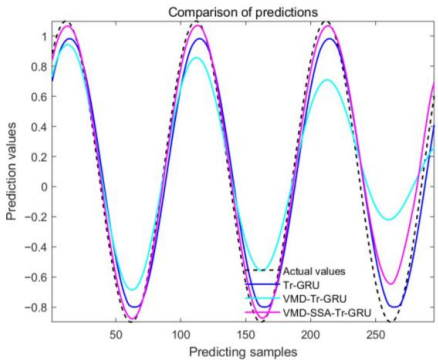
Figure 3. Time series for sin function based on CPO-VMD.

Table 1. Evaluation index for different models and corresponding optimized parameters.

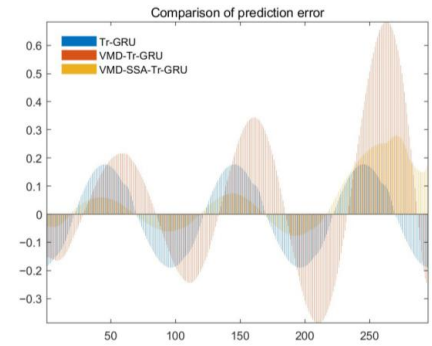
Models Type	Mean Absolute Error (MAE)	Mean Absolute Percentage Error (MAPE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Coefficient of Determination (R^2)	Optimized Number of Hidden Units	Optimized Maximum Training Epochs	Optimized Initial Learning Rate
Tr-GRU (65%)	0.098616	0.56336	0.012012	0.1096	0.9732			
VMD-Tr-GRU (65%)	0.10277	0.7167	0.013282	0.11525	0.97037			
CPO-VMD-SSA-Tr-GRU (65%)	0.06747	0.53725	0.005306	0.072839	0.98816	189	276	0.001
Tr-GRU (80%)	0.12107	0.45884	0.013775	0.11737	0.97154			
VMD-Tr-GRU (80%)	0.27802	0.71633	0.018572	0.13628	0.96163			
CPO-VMD-SSA-Tr-GRU (80%)	0.077425	0.50335	0.006795	0.08243	0.98596	81	72	0.01
Tr-GRU (90%)	0.071149	0.20958	0.006526	0.080781	0.98391			
VMD-Tr-GRU (90%)	0.11331	0.48885	0.01676	0.12946	0.95867			
CPO-VMD-SSA-Tr-GRU (90%)	0.032486	0.15236	0.001123	0.033505	0.99723	173	300	0.001

Table 2. Comparison of the predicted values in CPO-VMD-SSA-Tr-GRU model and other models.

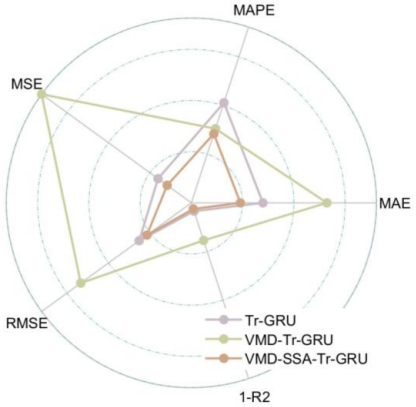
Theoretical Values	Tr-GRU	VMD-Tr-GRU	CPO-VMD-SSA-Tr-GRU
−0.74438	−0.63768	−0.54035	−0.71126
−0.75424	−0.64419	−0.55027	−0.72158
−0.76316	−0.64973	−0.55933	−0.73072
−0.77113	−0.65466	−0.56754	−0.7389
−0.77815	−0.65879	−0.57487	−0.74619
−0.70032	−0.65063	−0.52141	−0.68849
−0.6862	−0.64517	−0.50911	−0.67512
−0.67121	−0.63904	−0.49598	−0.66085
−0.65536	−0.63012	−0.48203	−0.6457
−0.63867	−0.61818	−0.46727	−0.62947
0.366769	0.254163	0.455529	0.380002
0.397657	0.282355	0.483981	0.410418
0.428351	0.31043	0.512218	0.440727
0.458819	0.33836	0.540204	0.471013
0.489032	0.366116	0.56791	0.50064
0.99653	0.845004	1.011617	0.96932
1.015128	0.863365	1.024764	0.983131
1.032921	0.881051	1.035954	0.995928
1.049893	0.898045	1.045545	1.007117
1.066025	0.914332	1.053257	1.016754



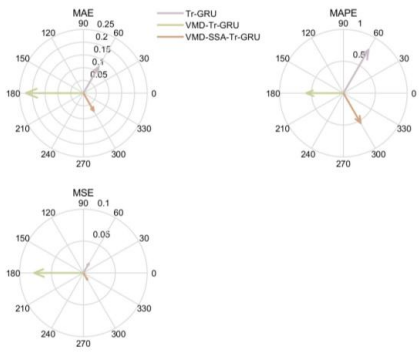
(**a**) Comparison of predictions



(**b**) Comparison of prediction error

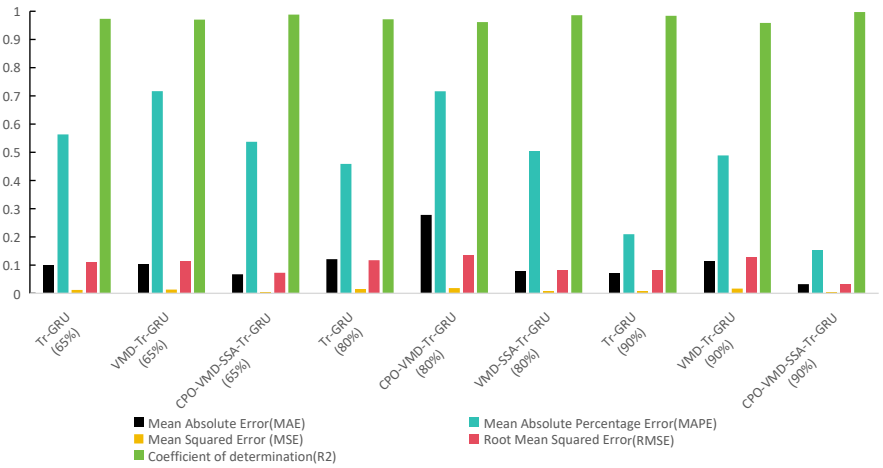


(**c**) Radar map of evaluation index

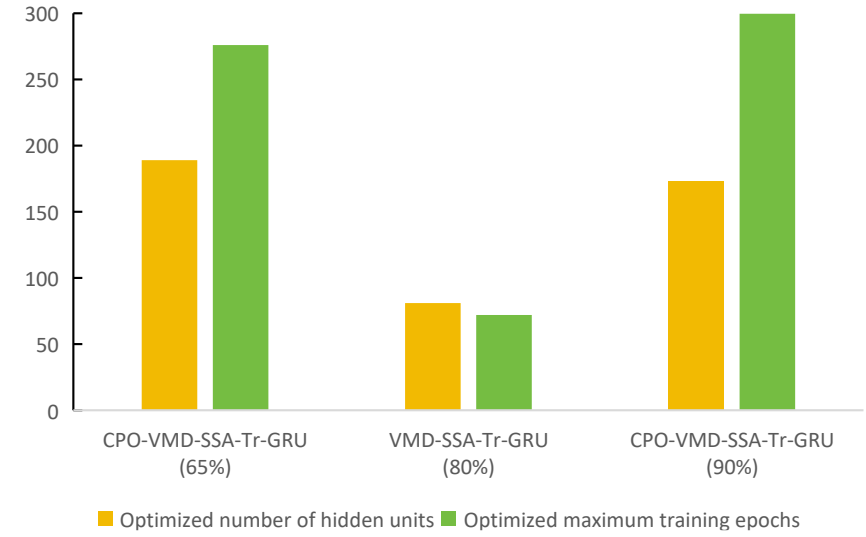


(**d**) Radar map of MAE, MAPE, MSE

Figure 4. Comparison of error and predicted values in CPO-VMD-SSA-Tr-GRU model and other models.



(**a**) Comparison of evaluation index by different models.



(**b**) Comparison of optimized parameters by different models.

Figure 5. Evaluation index for CPO-VMD-SSA-Tr-GRU model with different training samples and corresponding optimized parameters.

The results indicate that for the Tr-GRU model, the values of MAE, MAPE, MSE, RMSE, and R^2 are 0.098616, 0.56336, 0.012012, 0.1096, and 0.9732, respectively. For the VMD-Tr-GRU model, the corresponding values are 0.10277, 0.7167, 0.013282, 0.11525, and 0.97037. For the VMD-SSA-Tr-GRU model, the values are 0.06747, 0.53725, 0.0053055, 0.072839, and 0.98816, respectively. Notably, the CPO-VMD-SSA-Tr-GRU model achieves the lowest error values and the highest R^2 among all models. Furthermore, it demonstrates a faster convergence rate and enhanced global search capability. The model is capable of effectively capturing the nonlinear and dynamic characteristics inherent in long-term time series data and achieving a more precise global optimal solution. This indicates that the proposed model can rapidly converge to high-quality solutions for temperature prediction and is well-suited for simulation and forecasting tasks involving complex, high-dimensional, and large-scale datasets.

Furthermore, temperature predictions were conducted using different proportions of training and testing datasets, with 65%, 80%, and 90% of the total data allocated to the training set, respectively. The optimal number of hidden units and the maximum training epochs under these configurations are presented in Table 2 and illustrated in Figure 5. The results demonstrate that the CPO-VMD-SSA-Tr-GRU model exhibits strong adaptability across varying data partitioning scenarios and consistently achieves high training accuracy.

This underscores the model's robust generalization capability in predicting temperature time series characterized by strong nonlinearity and dynamic variability.

4. Study on Temperature Prediction of Mass Concrete Based on a Deeply Optimized VMD-SSA-GRU Model

4.1. Temperature Prediction with Single Time Series for Lab Construction

4.1.1. Test Design

The testing concrete had dimensions of 1 m × 1 m × 1 m. The mix design included Portland cement (P.O) with a compressive strength of not less than 42.5 MPa, coarse and medium sand with a fineness modulus of 2.5 or higher, continuously graded crushed stone with a particle size range of 5–31.5 mm, tap water, fly ash, and citric acid as a retarder. The detailed mix proportions are presented in Table 3, and the prepared concrete sample prior to pouring is shown in Figure 6.

Table 3. The ratio of the concrete specimen mix.

Materials	Cement	Sand	Rock	Water	Fly Ash	Citric Acid
mix proportion/(kg/m^3)	255	792	1047	160	76.5	0.5



Figure 6. The sample of mass concrete during construction.

4.1.2. Layout of the Cooling Pipes

The cooling water pipes were arranged in a single layer, with each layer positioned at the center of the concrete thickness (0.5 m). The flow rate was maintained at 1 m/s. The piping layout is illustrated in Figure 7, the corresponding parameters are listed in Table 4, and the cooling water temperature was assumed to be equivalent to the environment temperature.

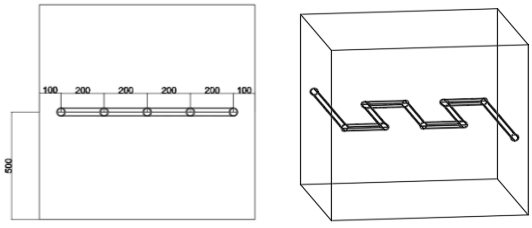


Figure 7. The arrangement of the single-layer pipes.

Table 4. The parameters of the pipes.

Parameters	Thermal Conductivity	Density	Thermal Expansion Coefficient	Outer Diameter	Wall Thickness
	W·(m^2·K)	g·cm^{-3}	(1/°C)	mm	mm
Values	0.15	1.35	7.0×10^{-5}	10	2

4.1.3. Placement of Temperature Measurement Points

The thermometer model used was SIN-RN3000, and the temperature sensor was RN3000. A total of six internal measurement points were symmetrically arranged within the concrete block, as illustrated in Figure 8. The temperature of the concrete during the water-cooling process was automatically collected and recorded.

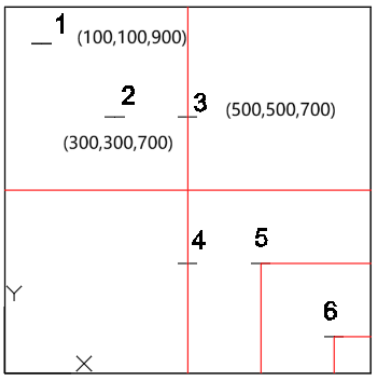
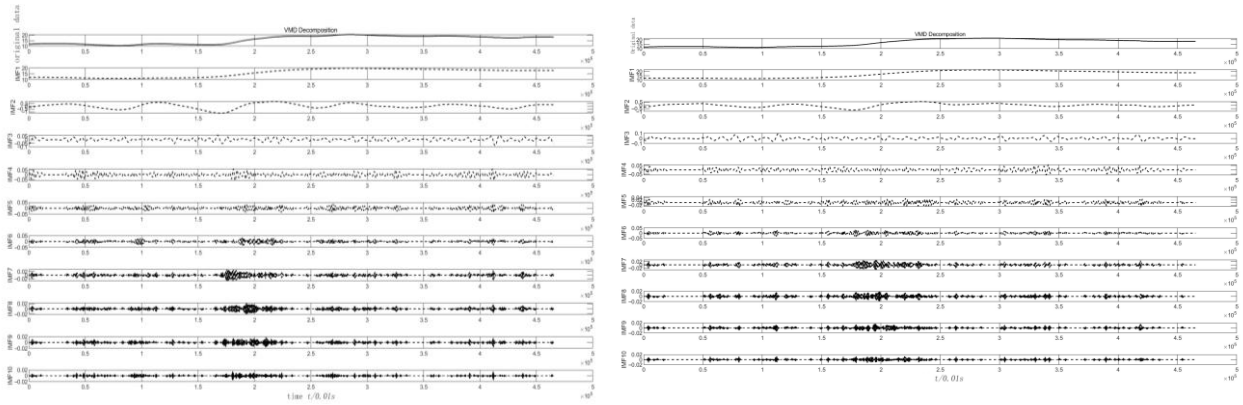


Figure 8. The arrangement of the checkpoints and corresponding coordinates (mm).

4.1.4. Temperature Time Series Prediction Based on Lab Tests

To validate the reliability and superior predictive performance of the CPO-VMD-SSA-Tr-GRU model in forecasting temperature time series for mass concrete, a comparative analysis was conducted against benchmark models, including the standalone Tr-GRU and VMD-Tr-GRU models. The temperature time series from Checkpoints 1 and 2 during laboratory-based concrete construction were used as input datasets. These selected time series were first preprocessed using the CPO-VMD technique (Figure 9) and then fed into the respective models for prediction. The computational results are summarized in Tables 5–9 and illustrated in Figures 10 and 11. Based on these outcomes, the predictive performance of each model was systematically evaluated and compared. The results presented in Tables 5–8 and Figures 10 and 11 indicate that the Tr-GRU and VMD-Tr-GRU models exhibited sensitivity to data fluctuations and had limited accuracy in capturing peak values and overall trend patterns, reflecting inherent constraints in their nonlinear fitting capabilities. In contrast, the VMD-SSA-Tr-GRU model achieved significantly higher prediction accuracy across the entire fluctuation range. This improvement is primarily attributed to the Sparrow Search Algorithm (SSA), which optimizes key hyperparameters, such as the number of hidden units, maximum training epochs, and initial learning rate, during the training process, thereby enhancing model convergence and overall predictive capability.



(**a**) Checkpoint 1 (**b**) Checkpoint 2

Figure 9. Time series for testing temperature in lab based on CPO-VMD.

Table 5. Error comparison of temperature prediction with different models (Checkpoint 1).

Model (90%)	Mean Absolute Error (MAE)	Mean Absolute Percentage Error (MAPE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Coefficient of Determination (R^2)
Tr-GRU	0.26985	0.014994	0.074717	0.27334	0.33159
VMD-Tr-GRU	0.18487	0.010216	0.041103	0.20274	0.6323
CPO-VMD-SSA-Tr-GRU	0.033736	0.0018812	0.001305	0.036127	0.98832

Table 6. Error comparison of temperature prediction with different models (Checkpoint 2).

Model (90%)	Mean Absolute Error (MAE)	Mean Absolute Percentage Error (MAPE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Coefficient of Determination (R^2)
Tr-GRU	0.11977	0.0065454	0.014632	0.12096	−0.29235
VMD-Tr-GRU	0.10629	0.0058106	0.012472	0.11168	−0.10157
CPO-VMD-SSA-Tr-GRU	0.016725	0.00091304	0.000365	0.019114	0.96773

Table 7. Comparison of the predicted values of CPO-VMD-SSA-Tr-GRU model and other models (Checkpoint 1).

Time	Temperature in Checkpoint 1	Tr-GRU	VMD-Tr-GRU	CPO-VMD-SSA-Tr-GRU
2 March 2025 08:37:29	17.7	17.509176	17.53813	17.702385
2 March 2025 08:39:29	17.7	17.509176	17.535011	17.700686
2 March 2025 08:41:29	17.7	17.509176	17.531834	17.703773
2 March 2025 08:43:29	17.7	17.509176	17.528616	17.69643
2 March 2025 08:45:29	17.6	17.509176	17.525337	17.640261
.....	.....			
2 March 2025 11:57:29	17.5	17.326246	17.415298	17.495182
2 March 2025 11:59:29	17.5	17.326246	17.417297	17.496754
2 March 2025 12:01:29	17.5	17.326246	17.419388	17.497347
2 March 2025 12:03:29	17.5	17.326246	17.421614	17.497026
2 March 2025 12:05:29	17.5	17.326246	17.424023	17.502207
2 March 2025 12:07:29	17.5	17.326246	17.426636	17.509899
.....	.....			
2 March 2025 14:27:29	18	17.600578	17.667452	17.918709
2 March 2025 14:29:29	18	17.637152	17.673679	17.94673
2 March 2025 14:31:29	18	17.673702	17.68005	17.960423
2 March 2025 14:33:29	18	17.710236	17.686457	17.953716
2 March 2025 14:35:29	18	17.74675	17.692862	17.963177
.....	.....			
2 March 2025 21:13:29	18.2	17.965757	17.969032	18.217176
2 March 2025 21:15:29	18.2	17.965757	17.968611	18.224201
2 March 2025 21:17:29	18.2	17.965757	17.968224	18.231836
2 March 2025 21:19:29	18.2	17.965757	17.967867	18.224049
2 March 2025 21:21:29	18.2	17.965757	17.967585	18.219481

As a result, the CPO-VMD-SSA-Tr-GRU model demonstrated the highest prediction accuracy among all tested models. A comparative analysis of the results confirms the effectiveness of the integrated hyperparameter optimization strategy, which substantially improves both convergence speed and computational accuracy, validating its feasibility and superiority in modeling complex thermal dynamics in mass concrete.

Table 8. Comparison of the predicted values of CPO-VMD-SSA-Tr-GRU model and other models (Checkpoint 2).

Time	Temperature in Checkpoint 2	Tr-GRU	VMD-Tr-GRU	CPO-VMD-SSA-Tr-GRU
2 March 2025 08:37:29	18.6	18.438955	18.591953	18.655531
2 March 2025 08:39:29	18.6	18.402805	18.585485	18.653751
2 March 2025 08:41:29	18.6	18.366646	18.579208	18.656033
2 March 2025 08:43:29	18.6	18.366646	18.573275	18.65411
2 March 2025 08:45:29	18.6	18.366646	18.567556	18.636728
		.....		
2 March 2025 18:07:29	18.2	18.095301	18.193825	18.254673
2 March 2025 18:09:29	18.2	18.077196	18.191532	18.242689
2 March 2025 18:11:29	18.2	18.059097	18.189159	18.235815
2 March 2025 18:13:29	18.2	18.040998	18.186775	18.23225
2 March 2025 18:15:29	18.2	18.0229	18.184435	18.2293
		.....		
2 March 2025 21:13:29	18.3	18.095301	18.187534	18.315294
2 March 2025 21:15:29	18.3	18.095301	18.187517	18.315332
2 March 2025 21:17:29	18.3	18.095301	18.187502	18.315245
2 March 2025 21:19:29	18.3	18.095301	18.18749	18.315292
2 March 2025 21:21:29	18.3	18.095301	18.187479	18.315359

Table 9. Error comparison of temperature prediction with different models in the rising stage (Checkpoint 1).

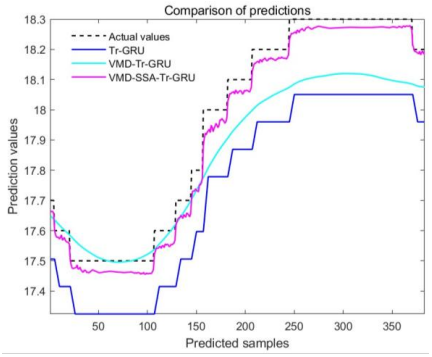
Model (90%)	Mean Absolute Error (MAE)	Mean Absolute Percentage Error (MAPE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Coefficient of Determination (R^2)
Tr-GRU	0.68044	0.034632	0.47444	0.6888	−0.25914
VMD-Tr-GRU	0.84392	0.042979	0.72543	0.85172	−0.92526
CPO-VMD-SSA-Tr-GRU	0.086023	0.004347	0.0087186	0.093374	0.97686

Based on the error analysis and predictions, the results for monitoring point 1 are as follows. For the standalone Tr-GRU model, the values of the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) are 0.26985, 0.014994, 0.074717, 0.27334, and 0.33159, respectively. For the VMD-Tr-GRU model, the corresponding values are 0.18487, 0.010216, 0.041103, 0.20274, and 0.6323. For the CPO-VMD-SSA-Tr-GRU model, the values are 0.033736, 0.0018812, 0.0013051, 0.036127, and 0.98832, respectively. For monitoring point 2, the results are as follows. For the standalone Tr-GRU model, the values are 0.11977, 0.0065454, 0.014632, 0.12096, and −0.29235. For the VMD-Tr-GRU model, the values are 0.10629, 0.0058106, 0.012472, 0.11168, and −0.10157. For the CPO-VMD-SSA-Tr-GRU model, the values are 0.016725, 0.00091304, 0.00036536, 0.019114, and 0.96773.

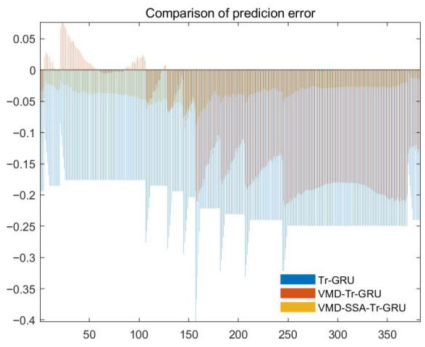
In addition, the model was applied to predict the temperature in the rising stage, as shown in Table 9 and Figure 12, with good accuracy for the proposed model. Thus, the proposed CPO-VMD-SSA-Tr-GRU model can capture the nonlinear and complex characteristics of the time series and give good predictions in all stages for the mass concrete construction.

The CPO-VMD-SSA-Tr-GRU model consistently outperformed the other models in terms of prediction accuracy, demonstrating that the integration of CPO for signal decomposition preprocessing and SSA for hyperparameter optimization significantly enhanced its predictive performance. Furthermore, to further improve temperature prediction accuracy in mass concrete construction, it is recommended to incorporate the correlation between dominant temperature-influencing factors and actual temperature measurements. A multivariate temperature series prediction model based on the CPO-VMD-SSA-Tr-GRU

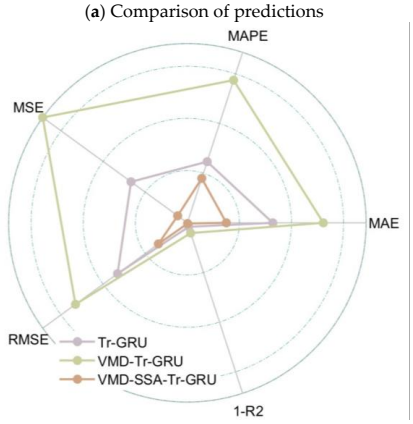
framework can be developed to enhance both the efficiency and accuracy of thermal forecasting in complex construction environments.



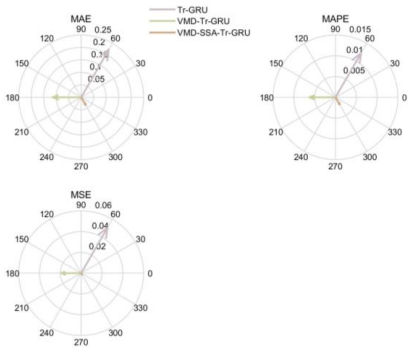
(**a**) Comparison of predictions



(**b**) Comparison of prediction error

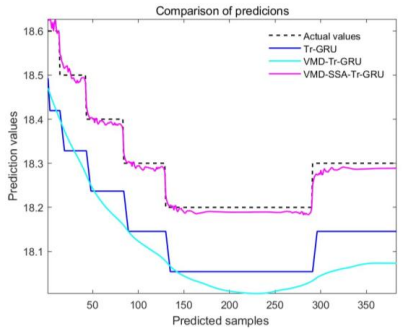


(**c**) Radar map of evaluation index

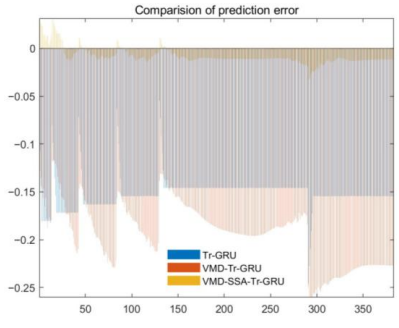


(**d**) Radar map of MAE, MAPE, MSE (Checkpoint 1)

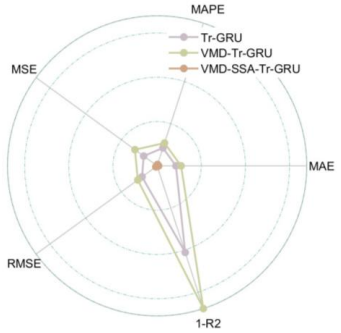
Figure 10. Comparison of error and predicted values of CPO-VMD-SSA-Tr-GRU model and other models (Checkpoint 1).



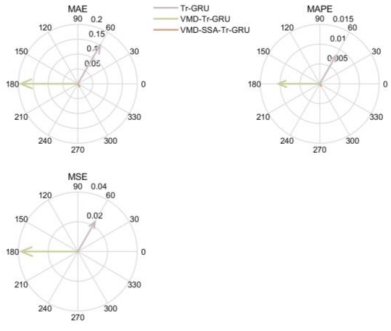
(**a**) Comparison of predictions



(**b**) Comparison of prediction error

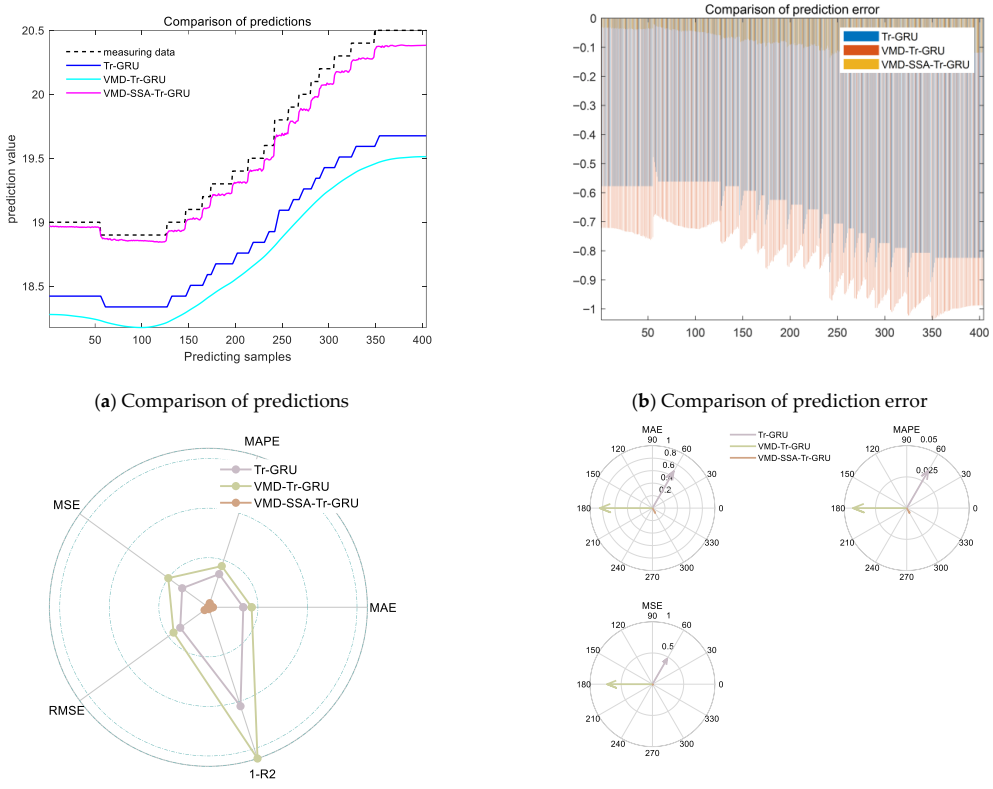


(**c**) Radar map of evaluation index



(**d**) Radar map of MAE, MAPE, MSE

Figure 11. Comparison of error and predicted values of CPO-VMD-SSA-Tr-GRU model and other models (Checkpoint 2).



(**a**) Comparison of predictions

(**b**) Comparison of prediction error

(**c**) Radar map of evaluation index

(**d**) Radar map of MAE, MAPE, MSE

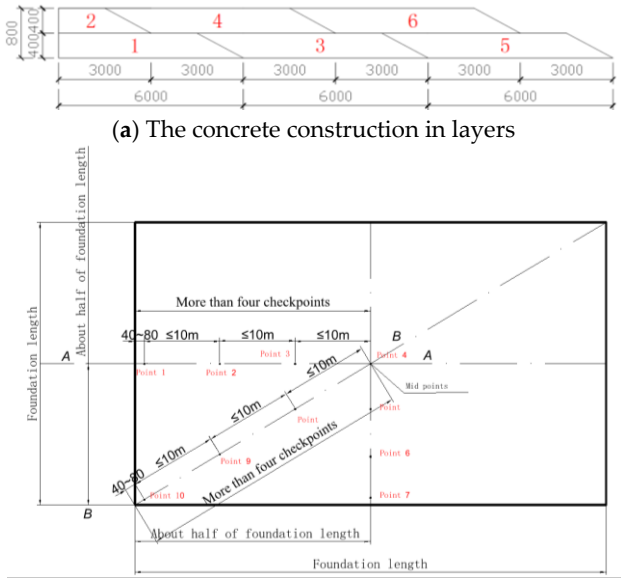
Figure 12. Comparison of error and predicted values of CPO-VMD-SSA-Tr-GRU model and other models (Checkpoint 2).

4.2. Field Temperature Prediction Based on Multivariate CPO-VMD-SSA-Tr-GRU Model

4.2.1. Project Overview

For a typical 1000 kV UHV substation, the pile raft foundation is 400 m long in the longitudinal direction, and concrete pouring involves 22 pouring bays. According to construction requirements, the hydration heat release rate and the variation law of heat release for various admixture combinations of cement-based materials were experimentally determined. Temperature sensors were installed to enable continuous temperature monitoring, as shown in Figure 13. Based on the proposed prediction method, the temperature evolution of mass concrete during construction was simulated to provide technical support for optimizing temperature control strategies.

The feasibility of the CPO-VMD-SSA-Tr-GRU model was demonstrated in Section 4.1 using a single temperature time series. However, during mass concrete construction, temperature fluctuations are closely influenced by multiple factors, including admixture composition, ambient temperature, and other environmental conditions. Relying solely on a univariate time series for temperature prediction may result in reduced accuracy due to the omission of critical external variables. To enhance predictive performance, a deep learning-based model incorporating multi-factor cooperative control mechanisms and taking into account the typical characteristics of monitoring temperature series is proposed (Table 10). This multivariate approach enables more accurate forecasting during actual construction, thereby providing scientific support for the effective formulation and optimization of temperature control strategies in mass concrete.



(**a**) The concrete construction in layers

(**b**) Arrangement of checkpoints

Figure 13. Jumping construction of mass concrete and corresponding checkpoints.

Table 10. Training and testing datasets (point 1).

No.	Air Temperature	Temperature in Point 1
1	27	61.5
2	29	61.2
3	35	60.8
4	30	59.7
5	25	59.1
6	24	58.6
7	33	58.3
8	26	58.1
9	24.5	57
10	23.5	56.9
11	33	56.5
12	24	56.2
	.....	
155	29	40.7
156	23.5	40.5
157	21	40.4
158	20.5	40.1
159	26	39.2
160	23	38.6
161	19.5	37.8
162	19	36.8

4.2.2. Temperature Prediction Analysis

The temperature series presented in Table 11 was decomposed into subsequences using the CPO-VMD technique, as illustrated in Figure 14. A total of 162 data points were obtained, with 5 consecutive data points selected as independent variables (input) and the 6th data point designated as the dependent variable (output). Based on this sliding window approach, the time series was reconstructed into a matrix consisting of 157 rows and 6 columns, where the output layer dimension was 1. Subsequently, the resulting subsequences were fed into the VMD-SSA-Tr-GRU model to generate prediction results for monitoring point 1 in Chamber 1, which are summarized in Table 12 and depicted in Figure 15.

Table 11. Error comparison of temperature prediction with different models (Checkpoint 1).

Model (90%)	Mean Absolute Error (MAE)	Mean Absolute Percentage Error (MAPE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Coefficient of Determination (R^2)
Tr-GRU	3.0803	0.074958	10.04	3.1686	−0.35282
VMD-Tr-GRU	1.9667	0.047363	4.0716	2.0178	0.45136
CPO-VMD-SSA-Tr-GRU	0.56293	0.013677	0.34035	0.58339	0.95414

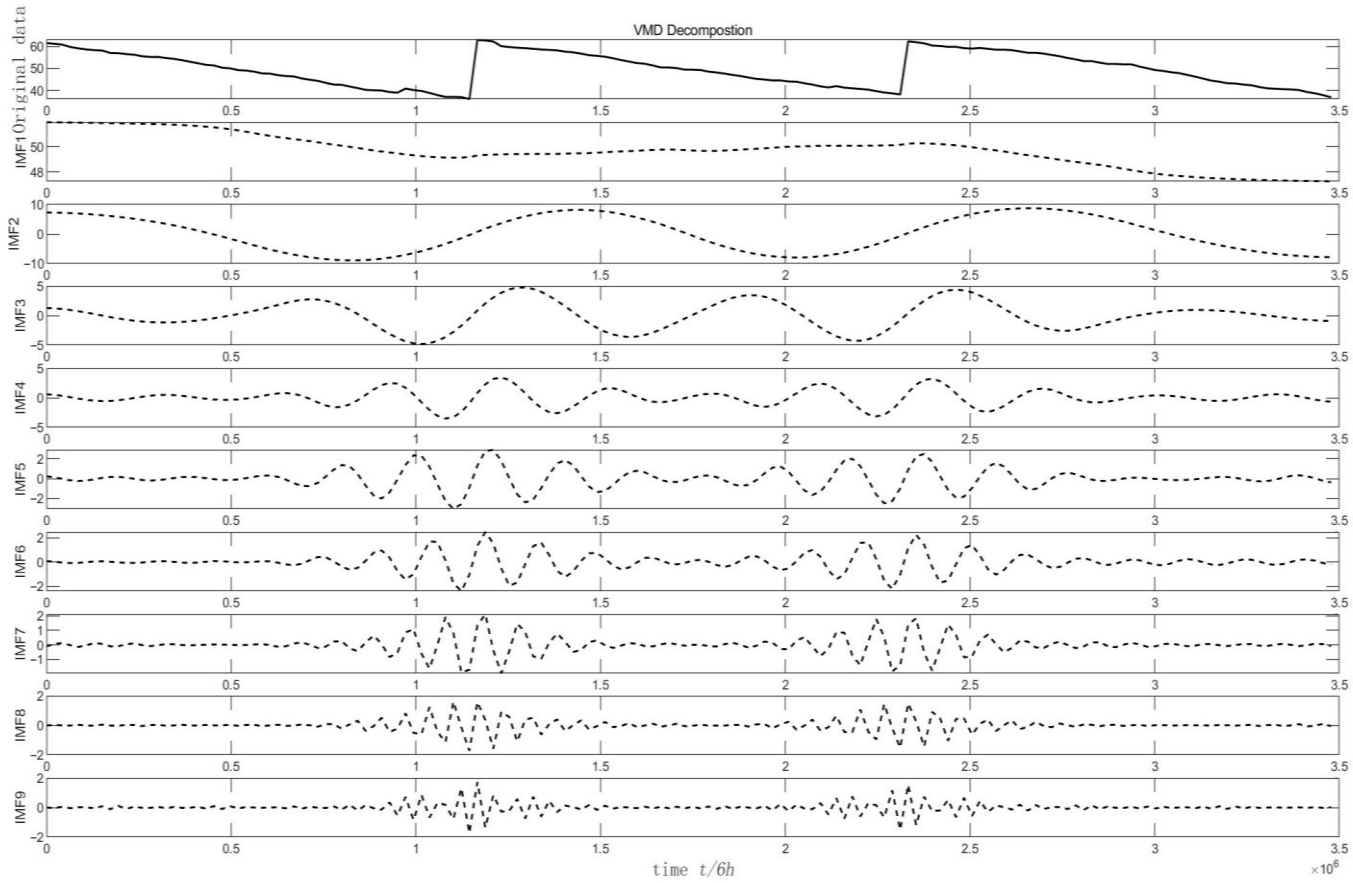
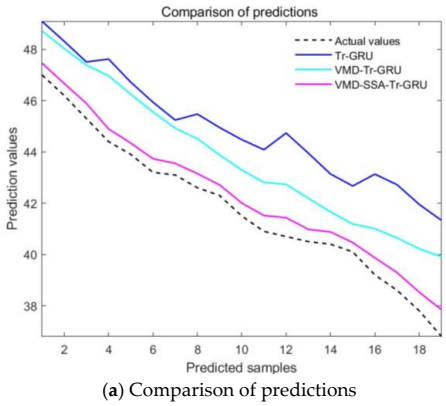


Figure 14. Time series for monitoring temperature in transmission infrastructure based on CPO-VMD.

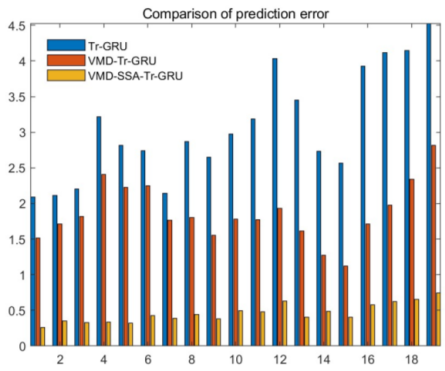
Table 12. Comparison of the predicted values of CPO-VMD-SSA-Tr-GRU model and other models (point 1).

Actual Temperature in Point 1	Tr-GRU	VMD-Tr-GRU	CPO-VMD-SSA-Tr-GRU
47	48.998734	48.496426	47.413113
46.2	48.129364	47.932735	46.654602
45.3	47.373768	47.168629	45.828285
44.4	47.313866	46.59811	44.816147
43.9	46.617676	46.036911	44.310432
43.2	45.771885	45.418903	43.722919
43.1	45.075943	44.689621	43.50808
42.6	45.1549	44.154358	43.091953
42.3	44.761082	43.633194	42.700665
41.5	44.24445	43.097313	42.002522
40.9	43.835213	42.524963	41.471081
40.7	44.286972	42.286591	41.33123
40.5	43.759117	41.922653	41.030804
40.4	42.908455	41.520748	40.916813
40.1	42.51244	40.93829	40.453896
39.2	42.76593	40.625797	39.809101
38.6	42.542824	40.3825	39.277225
37.8	41.927334	40.073124	38.56871
36.8	41.40247	39.710876	37.821724

The results show that for the CPO-VMD-SSA-Tr-GRU model, the values of the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) are 0.56293, 0.34035, 0.58339, and 0.95414, respectively. In comparison with single-variable prediction, the corresponding MAE, MSE, RMSE, and R^2 values were 3.0803, 10.04, 3.1686, and −0.35282 for one alternative model and 3.3609, 12.063, 3.4731, and −0.62542 for another. These results demonstrate that the CPO-VMD-SSA-Tr-GRU model achieves significantly higher prediction accuracy when applied to multivariate temperature series, highlighting its strong generalization capability and effectiveness in capturing complex temporal patterns and inter-variable dependencies within the temperature dynamics.

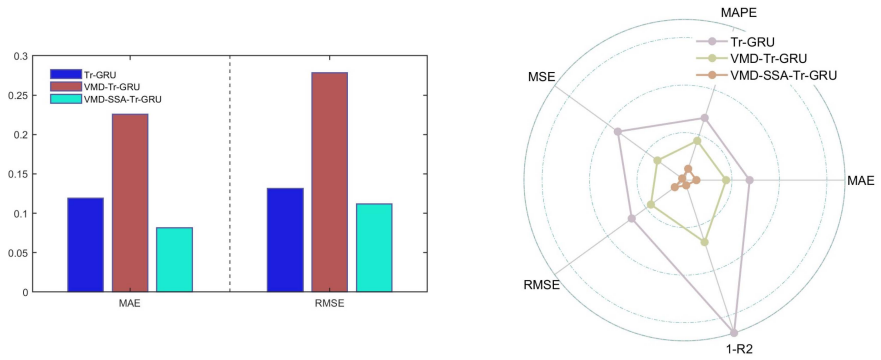


(**a**) Comparison of predictions

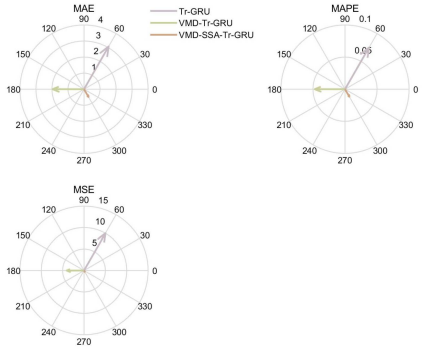


(**b**) Comparison of prediction error

Figure 15. *Cont.*



(**c**) Comparison of MAE and RMSE (**d**) Radar map of evaluation index



(**e**) Radar map of MAE, MAPE, MSE

Figure 15. Comparison of error and predicted values of CPO-VMD-SSA-Tr-GRU model and other models (point 1).

5. Conclusions

A deeply optimized VMD-Tr-GRU model was developed to address the nonlinear characteristics of temperature time series in mass concrete during the construction phase, with a focus on improving prediction accuracy under complex environmental conditions. The key findings are as follows:

(1) A CPO-VMD-SSA-Tr-GRU model was proposed for predicting temperature time series in mass concrete construction. Comparative experiments were conducted using different training dataset proportions to evaluate its adaptability and robustness across varying data availability scenarios.

(2) A simulation study based on the sine function was carried out using the CPO-VMD-SSA-Tr-GRU model. By comparing the performance and simulation results of the Tr-GRU, VMD-Tr-GRU, and CPO-VMD-SSA-Tr-GRU models, it was demonstrated that the CPO-VMD-SSA-Tr-GRU model effectively optimizes the decomposition parameters of VMD through the CPO algorithm and selects optimal hyperparameters via the SSA. This approach alleviates the challenge of manual hyperparameter tuning, thereby enhancing the model learning capacity and generalization ability for complex temperature dynamics.

(3) The CPO-VMD-SSA-Tr-GRU model was compared with the Tr-GRU and VMD-Tr-GRU models to validate its reliability and superiority in temperature prediction. The results show that the proposed model exhibits stronger adaptability and improved generalization capability, particularly in capturing long-term temporal patterns in extended time series.

(4) After hyperparameter optimization, the CPO-VMD-SSA-Tr-GRU model demonstrates strong predictive performance and notable generalization ability at real-world monitoring points. For multivariate temperature series prediction at monitoring point 1, the evaluation metrics are as follows: Mean Absolute Error (MAE): 0.56293; Mean Squared Error (MSE): 0.34035; Root Mean Squared Error (RMSE): 0.58339; Coefficient of Determination (R^2): 0.95414. These results are significantly more accurate than those obtained from univariate time series predictions, indicating that the multivariate framework captures

temperature variations more precisely and offers enhanced computational efficiency. This further confirms the model robustness in handling high-dimensional, strongly nonlinear temperature series.

In conclusion, the CPO-VMD-SSA-Tr-GRU model is a feasible and effective solution for temperature prediction during mass concrete construction. As for temperature prediction, the data acquisition frequency should be increased to enlarge the dataset, and predetermined parameters should not be used in the proposed model due to the given concrete materials and in order to reduce the computing cost. The model successfully captures the high-dimensional nonlinear characteristics of thermal behavior, while featuring a computationally efficient architecture, fast convergence, strong global optimization capability, and high prediction accuracy. This model provides scientific support for both accurate temperature forecasting and optimized control strategies in practical engineering applications.

Author Contributions: Conceptualization, X.L., Y.S. and Y.D.; methodology, X.L.; software, X.L., Y.S. and Y.D.; validation, F.Z. and S.X.; formal analysis, J.C.; investigation, D.L.; writing—original draft preparation, X.L., Y.S. and Y.D.; data curation, writing—review and editing, Q.L.; resources, writing—review and editing, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Project in the Construction Branch of Chongqing State Grid Power Company, grant number SGTYHT/24-JS-001.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy reasons.

Conflicts of Interest: Authors Fuwen Zheng, Shiyu Xia, Dijia Li, Qinfeng Lu and Lijin Hu were employed by the company Construction Branch of Chongqing State Grid Power Company. Author Jin Chen was employed by the company Chongqing State Grid Power Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. *China Electric Power Development Report 2024*; National Energy Administration Electric Power Planning and Design Institute: Beijing, China, 2024. Available online: https://www.nea.gov.cn/2024-07/19/c_1310782066.htm (accessed on 20 November 2025).
2. Li, L.; Wang, S.; Ma, M. Finite Element Analysis of Large-volume Concrete Temperature Field Under the Conditions of Sequence Method. *Constr. Technol.* **2019**, *48*, 89–92.
3. *GB50496-2018*; National Standard of the People's Republic of China. Code for Construction of Mass Concrete. Architecture & Building Press: Beijing, China, 2018.
4. *GBT51028-2015*; National Standard of the People's Republic of China. Technical Specification for Temperature Measurement and Control of Mass Concrete. Architecture & Building Press: Beijing, China, 2015.
5. Zheng, W.; Yuan, Z.; Qiao, H.; Song, P.; Zhu, Z. Analysis of Temperature Field of Mass Concrete Considering the Effect of Hydration Degree. *Mater. Rep.* **2024**, *38 (S1)*, 262–268.
6. Yang, D.; Chen, R.; Yang, R.; Zhang, Y. Analysis of Time-Varying Temperature Effect of Hydration Heat of Ultra-High Strength Mass Concrete. *Ind. Constr.* **2025**, *55*, 246–253. [CrossRef]
7. Klemczak, B.; Bąba, D.; Siddique, R. Machine Learning-Based Prediction of Heat Transfer and Hydration-Induced Temperature Rise in Mass Concrete. *Energies* **2025**, *18*, 4673. [CrossRef]
8. Kabir, H.; Wu, J.; Dahal, S.; Joo, T.; Garg, N. Automated estimation of cementitious sorptivity via computer vision. *Nat. Commun.* **2024**, *15*, 9935. [CrossRef]
9. Peng, W. Simulation Analysis of Temperature Stress During the Construction Period of Concrete Pouring Blocks on Rock Foundation. *Water Conserv. Plan. Des.* **2022**, *2*, 88–93.
10. Xu, B.; Huang, Y.; Fu, X.; Zhou, J.; Tang, T. Concrete Pouring Block's Highest Temperature Prediction Model andIt's Application Based on Uniform Design. *Water Resour. Power* **2014**, *32*, 90.
11. Zheng, S.; Lin, L.; Mao, W.; Wang, Y.; Liu, J.; Yuan, Y. A Hybrid RBF-PSO Framework for Real-Time Temperature Field Prediction and Hydration Heat Parameter Inversion in Mass Concrete Structures. *Buildings* **2025**, *15*, 2236. [CrossRef]
12. Zhou, J.; Fan, S.; Fang, C.; Huang, Y.; Liu, F. Internal Maximum Temperature Prediction of Concrete Pouring Block Based on Random Forest Algorithm. *Water Resour. Power* **2024**, *42*, 84–87.

13. Wang, K.; Zhang, Y.; Liu, J.; He, X.; Cai, C.; Huang, S. Prediction of Concrete Box-Girder Maximum Temperature Gradient Based on BP Neural Network. *J. Railw. Sci. Eng.* **2024**, *21*, 837–850.
14. Guo, S.; Zhou, S. Study on Temperature Evolution Rule and Prediction Method of Mass Concrete under Natural Condition. *Water Resour. Hydropower Eng.* **2018**, *49*, 188–196.
15. Shen, C.; Tan, C.; Wang, X.; Huang, J.; Wang, Y.; Zhai, C. Prediction of Concrete Temperature in Temperature Rising Stage for High Arch Dam Based on LSTM Neural Network. *Water Resour. Power* **2022**, *40*, 101–104.
16. Cheng, J.; Kong, C.; Zou, K. Temperature Field Prediction During Concrete Construction Period of Pump and Sluice Project Based on LSTM. *Adv. Sci. Technol. Water Resour.* **2023**, *43*, 76–81.
17. Li, Z.; Liu, H. Combining Global and Sequential Patterns for Multivariate Time Series Forecasting. *Chin. J. Comput.* **2023**, *46*, 70–84.
18. Wang, Z.; Wang, Q.; Wu, T. A Novel Hybrid Model for Water Quality Prediction Based on VMD and I-GOA Optimized for LSTM. *Front. Environ. Sci. Eng.* **2023**, *17*, 88. [CrossRef]
19. Xing, F.; Song, X.; Wang, Y.; Qin, C. A New Combined Prediction Model for Ultra-Short-Term Wind Power Based on Variational Mode Decomposition and Gradient Boosting Regression Tree. *Sustainability* **2023**, *15*, 11026. [CrossRef]
20. Xu, W.; Dai, W.; Li, D.; Wu, Q. Short-Term Wind Power Prediction Based on a Variational Mode Decomposition-BiTCN-Psformer Hybrid Model. *Energies* **2024**, *17*, 4089. [CrossRef]
21. Galassi, A.; Lippi, M.; Torroni, P. Attention in Natural Language Processing. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4291–4308. [CrossRef]
22. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition Transformers with Auto Correlation for Long Term Series Forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.
23. Zheng, W.; Zheng, K.; Gao, L.; Zhang, L.; Lan, R.; Xu, L.; Yu, J. GRU-Transformer: A Novel Hybrid Model for Predicting Soil Moisture Content in Root Zones. *Agronomy* **2024**, *14*, 432. [CrossRef]
24. Chen, H.; Jiang, D.; Sahli, H. Transformer Encoder with Multi-Modal Multi-Head Attention for Continuous Affect Recognition. *IEEE Trans. Multimed.* **2020**, *23*, 4171–4183. [CrossRef]
25. Huang, Y.; Liu, K.; Cheng, T.; Zhu, Y.; Xiong, M.; Zhao, C.; Peng, M. Intelligent Prediction Model for Deformation Induced by Excavation of Harbor Foundation Pit Based on Temporal Fusion Transformer. *J. Basic Sci. Eng.* **2025**, *33*, 1287–1296. [CrossRef]
26. Hao, Z.; Li, W.; Wu, J.; Zhang, S.; Hu, S. A Novel Deep Learning Model for Mining Nonlinear Dynamics in Lake Surface Water Temperature Prediction. *Remote Sens.* **2023**, *15*, 900. [CrossRef]
27. Li, H.; Gao, B. Short-Term PV Power Forecasting Based on Improved VMD and SAS-Attention-GRU. *Acta Energiae Solaris Sin.* **2023**, *44*, 292–300.
28. Zou, Z.; Wu, T.; Zhang, X.; Zhang, Z. Short-term Load Forecast Based on Bayesian Optimized CNN-BiGRU Hybrid Neural Networks. *High Volt. Eng.* **2022**, *48*, 3935–3945.
29. Dai, Z.; Shi, K.; Zhu, Y.; Zhang, X.; Luo, Y. Intelligent Prediction of Transformer Loss for Low Voltage Recovery in Distribution Network with Unbalanced Load. *Energies* **2023**, *16*, 4432. [CrossRef]
30. Abdel-Basset, M.; Mohamed, R.; Abouhawwash, M. Crested Porcupine Optimizer: A new nature-inspired metaheuristic. *Knowl.-Based Syst.* **2024**, *284*, 111257. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.