electronics

MDPI

Article

Metering-Error Prediction for Electronic Current Transformers Using Improved VMD and Two-Stage RA-BiGRU

Yedong Mao [1,2], Yanfeng Xiao [1,2], Yijun Xiao [1,2], Haoyu Chen [3,*] and Zhenhua Li [3]

[1] Hubei Technology Innovation Center for Smart Hydropower, Wuhan 430019, China; mao_yedong@ctg.com.cn (Y.M.); xiao_yanfeng@ctg.com.cn (Y.X.); xiao_yijun@ctg.com.cn (Y.X.)
[2] China Yangtze Power Co., Ltd., Wuhan 430014, China
[3] College of Electrical Engineering & New Energy, China Three Gorges University, Yichang 443002, China; lizhenhua1993@163.com
* Correspondence: 15168467713@163.com

Abstract

Forecasting electronic current transformer (ECT) metering errors is challenging because long-term operation produces nonstationary and multiscale variations under changing load and environmental conditions. This study proposes an improved variational mode decomposition (VMD) and two-stage residual-attention bidirectional gated recurrent unit (RA-BiGRU) framework. A dual-constraint objective evaluates reconstruction fidelity, intermode independence, permutation entropy, and decomposition complexity, while hard constraints prevent low-mode under-decomposition. An improved whale optimization algorithm (IWOA) adaptively determines the VMD parameters. Stage 1 predicts the future increments of the final two trend-dominant modes and reconstructs preliminary ratio-error and phase-displacement forecasts. Stage 2 estimates the remaining forecast residuals, and horizon- and error-specific weights selected on the validation set limit over-correction. Experiments using 15,812 field samples at horizons of 1, 3, 6, and 12 consistently outperform the Persistence baseline. At h = 3, the proposed method reduces the ratio-error and phase-displacement RMSE values by 54.27% and 50.37%, respectively. These results demonstrate the complementary effects of trend-increment prediction and controlled residual correction.

Keywords: electronic current transformer; metering-error prediction; variational mode decomposition; improved whale optimization algorithm; bidirectional gated recurrent unit; temporal attention; residual correction

Check for updates

Academic Editor: Herminio Martínez-García

Received: 21 June 2026
Revised: 14 July 2026
Accepted: 20 July 2026
Published: 1 August 2026

Copyright: © 2026 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.

1. Introduction

Electronic current transformers (ECTs) are key digital metering devices in modern power systems. Their measurement accuracy directly affects protection, energy accounting, condition monitoring, and the reliable integration of intelligent secondary equipment. During long-term outdoor operation, however, sensing-element nonidealities, electronic-component drift, secondary-load variations, environmental electric fields, temperature, humidity, vibration, magnetic fields, and aging may gradually degrade measurement performance [1–5]. Periodic calibration is therefore required to ensure that the ratio error and phase displacement remain within the prescribed limits.

Conventional periodic calibration generally requires a reference standard and the device under test to be connected to the same circuit. For high-voltage outdoor installations, offline calibration can involve service interruptions, transportation of bulky reference equipment, and substantial labor and safety costs. Although energized on-site calibration

alleviates some of these difficulties, its results can be affected by environmental disturbances and operational constraints [6,7]. Forecasting the future metering-error state provides a complementary approach: it can support risk-based calibration scheduling, identify devices whose accuracy is likely to deteriorate, and reduce unnecessary outages.

Existing forecasting approaches can be grouped into physics-based, statistical, and data-driven methods. Physics-based models are interpretable, but their parameters may vary with operating and environmental conditions. Classical statistical methods have limited ability to describe nonlinear, multivariable, and multiscale dependencies. Deep learning models have therefore been introduced into power-system forecasting [8–17]. In instrument-transformer applications, neural-network, optimization-assisted, BiLSTM, and GRU-based models have been used to learn temporal relationships between operating factors and measurement errors [18–22]. However, direct end-to-end prediction remains sensitive to nonstationarity and local fluctuations in the error sequence.

Signal-decomposition-assisted forecasting is an effective way to address nonstationary time series. CEEMDAN-based methods have been used to construct multiscale load predictors [23,24], whereas VMD-based models have been applied to photovoltaic-power and interference forecasting [25–27]. VMD provides a mathematically defined decomposition and good time–frequency localization [28,29]. Nevertheless, its performance depends strongly on the number of modes and the quadratic penalty factor. Empirical parameter selection can cause under-decomposition, over-decomposition, or mode mixing. Moreover, a single entropy criterion does not simultaneously characterize reconstruction fidelity, intermode redundancy, and decomposition complexity.

On the basis of these studies, existing methods can be further compared from the perspective of forecasting target, decomposition strategy, and error-correction mechanism. Neural-network-based metering-error predictors usually learn the mapping from operating factors to future errors directly, which is simple but may not sufficiently separate rapid fluctuations from slow drift. Decomposition-assisted methods improve the description of nonstationary sequences, but many of them still rely on empirically selected decomposition parameters and directly predict all decomposed components. More specifically, conventional VMD-based frameworks commonly use empirical or single-index parameter selection, build predictors for the original sequence or every mode, and reconstruct the result without controlling the contribution of the remaining residual. In contrast, the proposed framework uses a constrained multi-criterion objective to select K and α, predicts only the increments of the final two trend-dominant modes, and applies horizon- and error-specific residual weights selected on the validation set.

This study develops an improved VMD and a two-stage RA-BiGRU framework for jointly forecasting the ratio error and phase displacement of an ECT. The main contributions are as follows:

1. To solve unreliable VMD parameter selection and low-mode under-decomposition, a dual-constraint IWOA-VMD method is proposed. The composite objective evaluates reconstruction fidelity, intermode independence, permutation entropy, and decomposition complexity, while hard constraints exclude infeasible low-mode solutions.
2. To reduce the redundant learning burden associated with predicting the original sequence or every mode, Stage 1 predicts only the future increments of the final two trend-dominant modes. The remaining modes and reconstruction residual are propagated from the current time, allowing the network to focus on medium- and low-frequency trend evolution.
3. To recover systematic errors left by the preliminary reconstruction without amplifying uncertain residual predictions, Stage 2 learns the remaining residuals and applies horizon- and error-specific correction weights selected on the validation set.

The remainder of this paper is organized as follows. Section 2 formulates the ECT metering-error forecasting problem. Section 3 presents the dual-constraint IWOA-VMD method. Section 4 describes the two-stage RA-BiGRU. Sections 5 and 6 report the experimental setup and results, respectively. Section 7 concludes the paper.

2. ECT Metering Errors and Forecasting Formulation

An ECT converts the primary current into a digital secondary output through sensing, signal-conditioning, and acquisition units. Let $I_1(t)$ be the primary reference current, $I_2(t)$ be the measured secondary current, and K_r be the rated transformation ratio. The secondary current referred to the primary side is $K_\mathrm{r} I_2(t)$. The ratio error is defined as

$$e_\mathrm{r}(t) = \frac{K_\mathrm{r} I_2(t) - I_1(t)}{I_1(t)} \times 100\%. \tag{1}$$

Let $\theta_1(t)$ and $\theta_2(t)$ denote the phases of the primary reference signal and the referred secondary signal, respectively. The phase displacement is

$$e_\phi(t) = \theta_2(t) - \theta_1(t). \tag{2}$$

In this study, the external operating-factor vector is

$$\mathbf{q}(t) = [L(t), T(t), H(t), V(t), B(t)]^\mathrm{T}, \tag{3}$$

where $L(t)$, $T(t)$, $H(t)$, $V(t)$, and $B(t)$ denote load, temperature, humidity, vibration, and magnetic field, respectively. Defining $\mathbf{e}(t) = \left[e_\mathrm{r}(t), e_\phi(t)\right]^\mathrm{T}$, the multivariable forecasting task is expressed as

$$\begin{bmatrix} \hat{e}_\mathrm{r}(t+h) \\ \hat{e}_\phi(t+h) \end{bmatrix} = \mathcal{F}(\mathbf{e}(t - L_\mathrm{w} + 1 : t), \mathbf{q}(t - L_\mathrm{w} + 1 : t)), \tag{4}$$

where L_w is the sliding-window length, h is the prediction horizon, and $\mathcal{F}(\cdot)$ is the nonlinear mapping to be learned. Because the original error sequences contain high-frequency disturbances, local fluctuations, and medium- and low-frequency trends, direct forecasting can be unstable. The proposed framework first decomposes the two error sequences and then performs trend-increment prediction and residual correction.

3. Dual-Constraint IWOA-VMD

3.1. Variational Mode Decomposition

VMD was originally proposed as a nonrecursive variational signal-decomposition method [24]. Recent MDPI studies have further used optimized or improved VMD to reduce the volatility of nonstationary forecasting sequences before recurrent or attention-based prediction [30].

VMD decomposes a signal $f(t)$ into K finite-bandwidth modes $u_k(t)$ centered at angular frequencies ω_k, together with a residual $r(t)$:

$$f(t) = \sum_{k=1}^{K} u_k(t) + r(t). \tag{5}$$

The constrained variational problem minimizes the sum of the estimated mode bandwidths:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \partial_t \left[\left(\delta(t) + \frac{\mathrm{j}}{\pi t} \right) \times u_k(t) \right] e_2^{-\mathrm{j}\omega_k t 2} \right\}, \tag{6}$$

subject to

$$\sum_{k=1}^{K} u_k(t) = f(t). \tag{7}$$

After introducing the quadratic penalty factor α and the Lagrange multiplier $\lambda(t)$, the augmented Lagrangian is

$$\begin{aligned}\mathcal{L}(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_{k=1}^{K} \partial_t \left[\left(\delta(t) + \frac{\mathrm{j}}{\pi t}\right) \times u_k(t)\right] e_2^{-\mathrm{j}\omega_k t 2} \\ + f(t) - \sum_{k=1}^{K} u_k(t)_2^2 + \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t).\end{aligned} \tag{8}$$

The mode number K controls decomposition granularity, whereas α controls mode bandwidth and separation. A small K can merge physically distinct scales, while a large K can create redundant modes. Likewise, an excessively large α can over-smooth local dynamics, whereas an excessively small value can cause mode mixing.

3.2. Dual-Constraint Objective Function

For a candidate pair (K, α), the normalized reconstruction error is defined as

$$J_{\mathrm{rec}} = \frac{\mathrm{RMSE}(f, \hat{f})}{\max(f) - \min(f)}, \ \hat{f}(t) = \sum_{k=1}^{K} u_k(t). \tag{9}$$

Intermode independence is quantified using the average absolute Pearson correlation coefficient:

$$J_{\mathrm{ind}} = \frac{2}{K(K-1)} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} |\rho(u_i, u_j)|. \tag{10}$$

The mean permutation entropy is

$$J_{\mathrm{pe}} = \frac{1}{K} \sum_{k=1}^{K} \mathrm{PE}(u_k), \tag{11}$$

and the normalized mode-number penalty is

$$J_K = \frac{K - K_{\min}}{K_{\max} - K_{\min}}. \tag{12}$$

The composite objective is

$$J(K, \alpha) = w_1 J_{\mathrm{rec}} + w_2 J_{\mathrm{ind}} + w_3 J_{\mathrm{pe}} + w_4 J_K, \tag{13}$$

where w_1–w_4 are weighting coefficients. A preliminary grid search showed that unconstrained minimization of Equation (13) may favor a small number of modes, even when the corresponding decomposition is under-resolved. Figure 1 illustrates this behavior.

To prevent low-mode under-decomposition, the first constraint requires

$$K \geq 9, \ \mathrm{NRMSE}_{\mathrm{r}} < 0.1\%, \ \mathrm{NRMSE}_{\phi} < 0.1\%. \tag{14}$$

The second constraint selects the candidate with the minimum value of Equation (13) among the feasible solutions. Table 1 lists the five best grid-search candidates after screening. The feasible optimum moves to the higher-mode region, indicating that the hard constraints effectively suppress under-decomposed solutions.

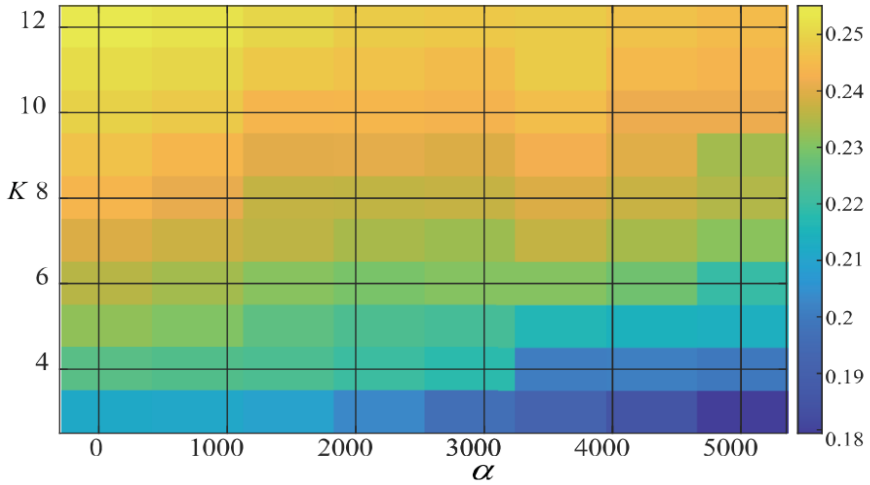


Figure 1. Grid-search distribution of the composite fitness for different values of K and α before applying the hard constraints.

Table 1. Candidate parameters after dual-constraint screening.

Rank	K	α	Fitness
1	11	50	0.198452991
2	12	50	0.218438037
3	10	50	0.283301996
4	9	50	0.363684024
5	12	100	0.464362050

3.3. IWOA-Based Parameter Optimization

The optimization vector and search domain are

$$\mathbf{s} = [K, \alpha],\ K \in [3, 12],\ \alpha \in [50, 5000], \tag{15}$$

where K is rounded to an integer and α is continuous. A logistic chaotic map is used to diversify the initial population:

$$z_{n+1} = \mu z_n(1 - z_n),\ \mu = 4. \tag{16}$$

To strengthen early global exploration and late local exploitation, the linear convergence factor of the standard whale optimization algorithm is replaced by

$$a(t) = 2\left[1 - \left(\frac{t}{T_{\max}}\right)^2\right], \tag{17}$$

where t and $T_{\max}$ denote the current and maximum iteration indices, respectively. With coefficient vectors $\mathbf{A} = 2a\mathbf{r} - a$ and $\mathbf{C} = 2\mathbf{r}$, the position update is

$$\mathbf{X}(t+1) = \mathbf{X}^*(t) - \mathbf{A}\cdot|\mathbf{C}\cdot\mathbf{X}^*(t) - \mathbf{X}(t)|,\ p < 0.5, |\mathbf{A}| < 1, \tag{18}$$

$$\mathbf{X}(t+1) = \mathbf{X}_{\text{rand}}(t) - \mathbf{A}\cdot|\mathbf{C}\cdot\mathbf{X}_{\text{rand}}(t) - \mathbf{X}(t)|,\ p < 0.5, |\mathbf{A}| \geq 1, \tag{19}$$

and

$$\mathbf{X}(t+1) = \left|\mathbf{X}^*(t) - \mathbf{X}(t)\right| e^{b\updownarrow}\cos(2\pi\updownarrow) + \mathbf{X}^*(t),\ p \geq 0.5, \tag{20}$$

where $\mathbf{X}^*(t)$ is the current best solution, $\mathbf{X}_{\text{rand}}(t)$ is a randomly selected individual, b is the spiral constant, and $\updownarrow \in [-1, 1]$. Infeasible solutions are penalized, whereas feasible solutions are ranked using Equation (13).

Figure 2 compares the convergence histories of WOA and IWOA. The standard WOA stagnates at approximately 0.2833 after about 52 iterations. IWOA escapes intermediate local optima and converges to approximately 0.1984 after about 65 iterations.

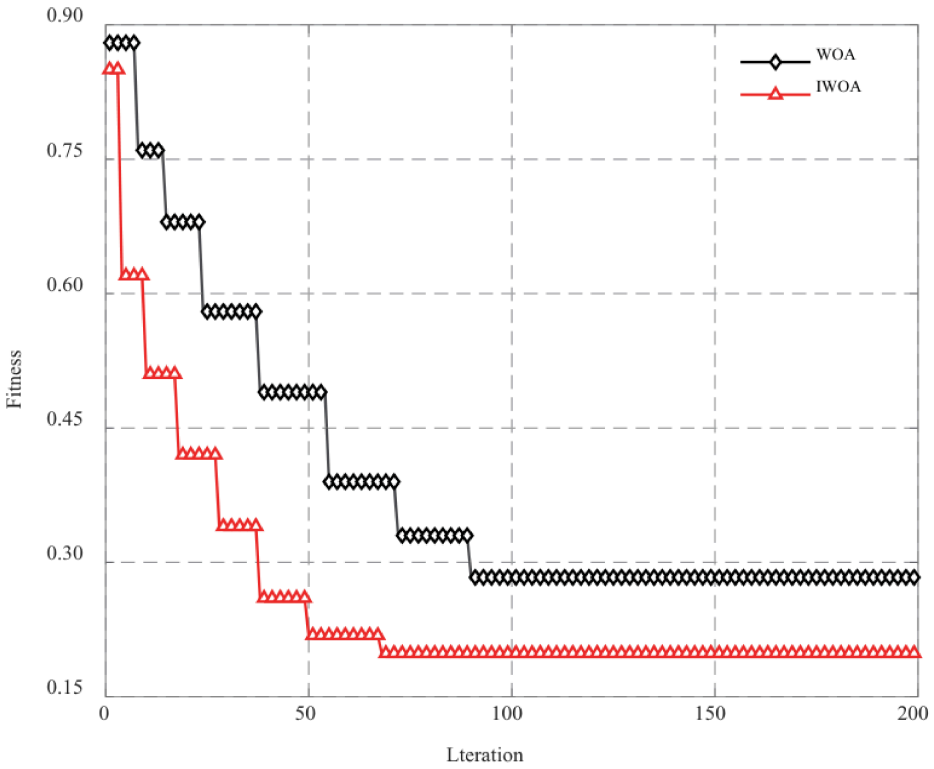


Figure 2. Convergence histories of WOA and IWOA under the dual-constraint VMD objective.

3.4. Decomposition Results

The optimized parameters are

$$K = 11, \ \alpha = 50.026679. \tag{21}$$

The choice of K = 11 is also supported by the dual-constraint screening results in Table 1. Compared with smaller K values, K = 11 avoids merging components with different time scales; compared with K = 12, it gives a lower feasible fitness while avoiding one additional mode. Thus, K = 11 provides a balanced decomposition in terms of reconstruction accuracy, mode separation, and forecasting-model complexity.

The reconstruction errors for the ratio-error and phase-displacement sequences are

$$\begin{aligned} \text{RMSE}_{\text{r}} &= 1.51954 \times 10^{-5}, \ \text{NRMSE}_{\text{r}} = 0.012663\%, \\ \text{RMSE}_{\phi} &= 8.82670 \times 10^{-4}, \ \text{NRMSE}_{\phi} = 0.004601\%. \end{aligned} \tag{22}$$

Figure 3 shows the IWOA-VMD of the ratio-error sequence. The early modes primarily contain local high-frequency fluctuations, whereas the final modes evolve more smoothly. Figure 4 shows IMF10 and IMF11 for both error types. These modes retain the medium- and low-frequency trend information and are therefore selected as the Stage-1 prediction targets.

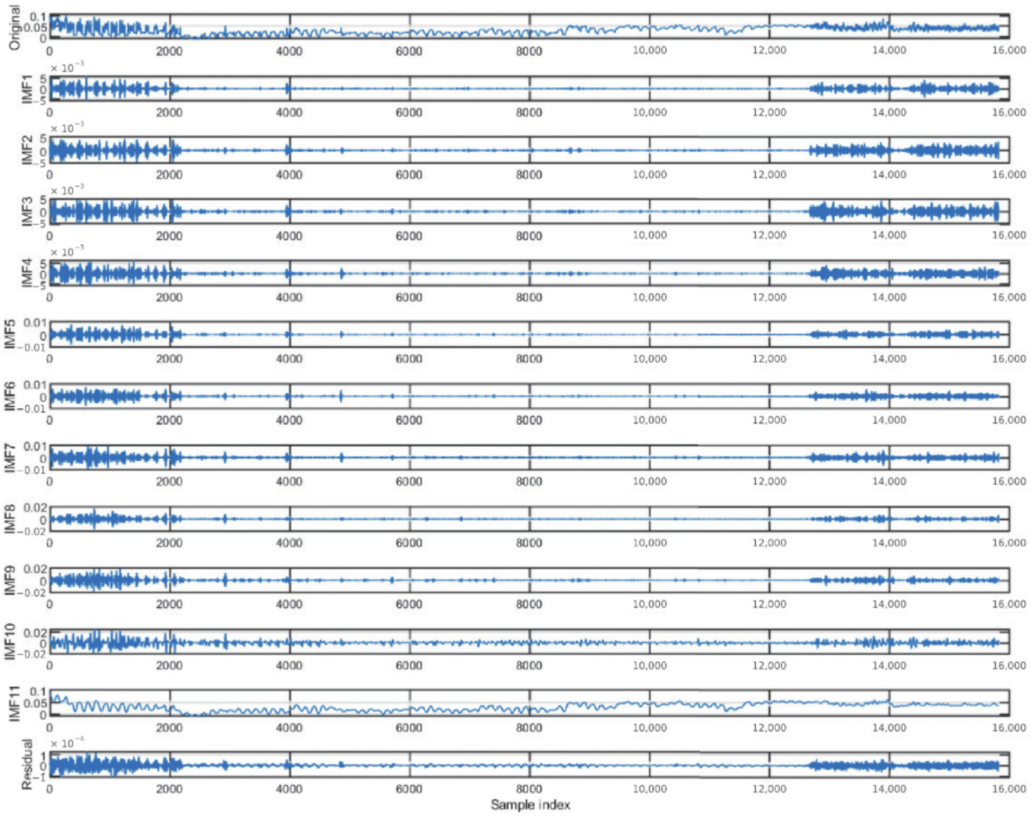


Figure 3. IWOA-VMD of the ratio-error sequence.

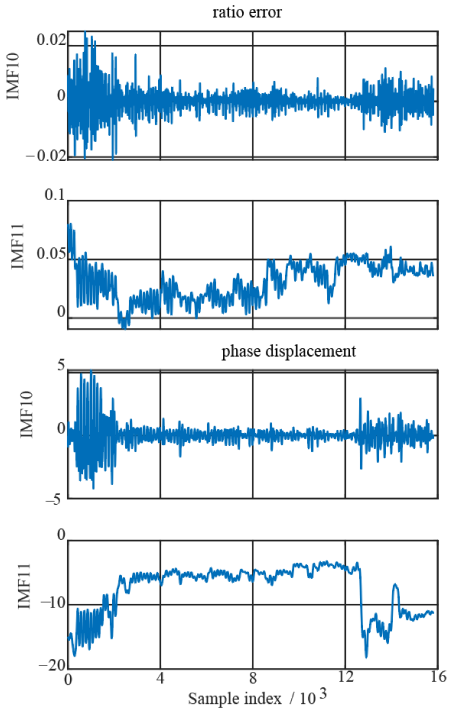


Figure 4. The last two medium- and low-frequency modes of the ratio-error and phase-displacement sequences.

As shown in Figure 3, the decomposed modes present different fluctuation scales. The earlier IMFs mainly describe high-frequency local fluctuations and measurement disturbances, whereas the later IMFs become smoother and contain more medium- and

low-frequency trend information. This decomposition helps separate rapid variations from slow drift in the original ratio-error sequence.

Figure 4 further compares the last two modes of the ratio error and phase displacement. IMF10 and IMF11 vary more smoothly than the earlier modes and retain the main medium- and low-frequency trend information. Therefore, they are selected as the Stage-1 prediction targets. Predicting their future increments emphasizes trend evolution instead of directly repeating the current metering-error value.

4. Two-Stage RA-BiGRU Forecasting Model

4.1. Input Construction

The ratio-error and phase-displacement sequences are decomposed as

$$\begin{aligned} e_{\mathrm{r}}(t) &= \sum_{k=1}^{K} u_{\mathrm{r},k}(t) + r_{\mathrm{r}}(t), \\ e_{\phi}(t) &= \sum_{k=1}^{K} u_{\phi,k}(t) + r_{\phi}(t), \end{aligned} \tag{23}$$

where $u_{\mathrm{r},k}$ and $u_{\phi,k}$ are the kth modes, and r_{r} and r_{ϕ} are the corresponding residuals. The input vector is

$$\mathbf{x}(t) = \left[u_{\mathrm{r},1}(t), \ldots, u_{\mathrm{r},K}(t), r_{\mathrm{r}}(t), u_{\phi,1}(t), \ldots, u_{\phi,K}(t), r_{\phi}(t), \mathbf{q}^{\mathrm{T}}(t)\right]^{\mathrm{T}}. \tag{24}$$

For $K = 11$, the 11 ratio-error modes, one ratio residual, 11 phase-displacement modes, one phase residual, and five operating factors yield 29 input features, as summarized in Table 2. A sliding window is constructed as

$$\mathbf{X}_t = [\mathbf{x}(t - L_{\mathrm{w}} + 1), \mathbf{x}(t - L_{\mathrm{w}} + 2), \ldots, \mathbf{x}(t)]. \tag{25}$$

Table 2. Input features of the two-stage forecasting model.

Feature Category	Variables	Number
Ratio-error VMD modes	$u_{\mathrm{r},1}, \ldots, u_{\mathrm{r},11}$	11
Ratio-error residual	r_{r}	1
Phase-displacement VMD modes	$u_{\phi,1}, \ldots, u_{\phi,11}$	11
Phase-displacement residual	r_{ϕ}	1
Operating factors	Load, temperature, humidity, vibration, magnetic field	5
Total	—	29

4.2. RA-BiGRU Prediction Unit

The RA-BiGRU unit in Figure 5 is used in both forecasting stages. It consists of a bidirectional GRU layer, a temporal attention layer, a residual mapping layer, and a fully connected output layer. For each input window, the unit receives the 29-dimensional feature sequence constructed in Section 4.1. In Stage 1, the output is the four mode increments defined in Equation (32), whereas in Stage 2, the output is the two residual terms defined in Equation (36).

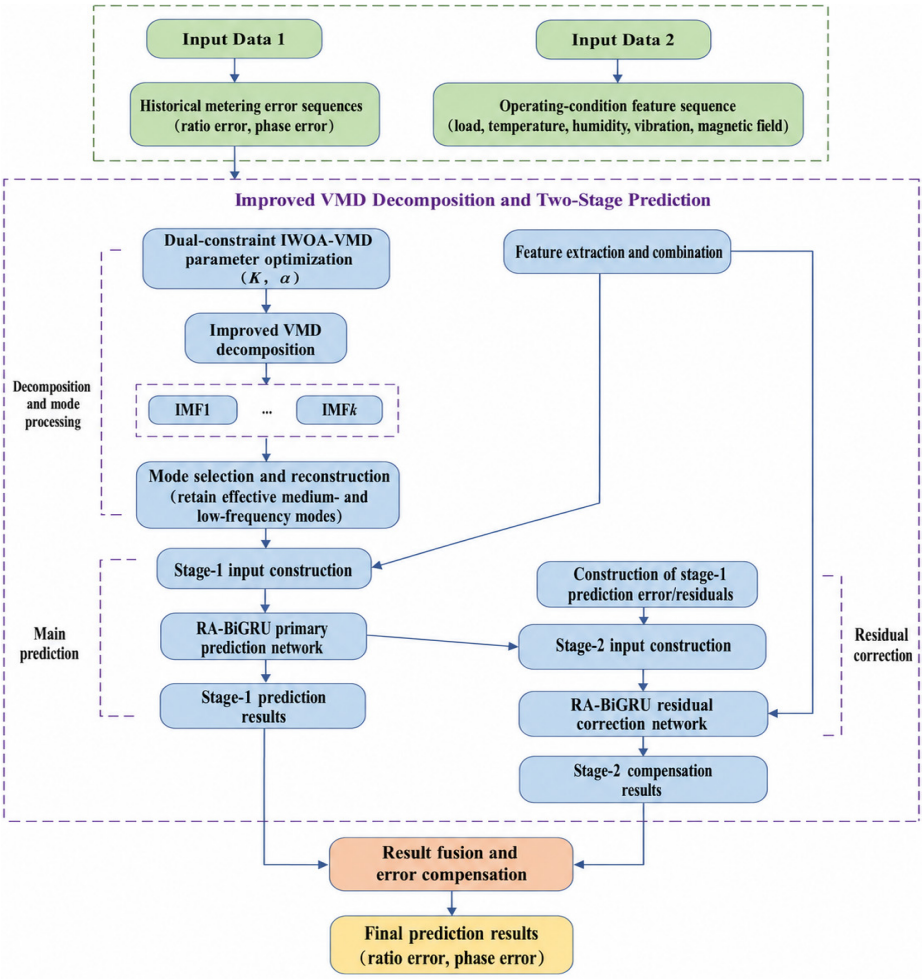


Figure 5. Architecture and signal flow of the proposed improved-VMD two-stage RA-BiGRU framework.

The use of BiGRU and attention follows the common recurrent-attention forecasting idea, in which bidirectional recurrent structures extract contextual temporal information and attention emphasizes informative time steps [31].

For time step i in an input window, the forward and backward GRU states are:

In Equation (26), the forward GRU describes the temporal evolution from past samples to the current sample, whereas the backward GRU extracts reverse contextual information within the same window. The concatenated hidden state therefore provides more complete temporal features than a single-direction GRU.

$$\overrightarrow{\mathbf{h}}_i = \mathrm{GRU}_{\mathrm{f}}(\mathbf{x}_i, \overrightarrow{\mathbf{h}}_{i-1}),\ \overleftarrow{\mathbf{h}}_i = \mathrm{GRU}_{\mathrm{b}}(\mathbf{x}_i, \overleftarrow{\mathbf{h}}_{i+1}),\ \mathbf{h}_i = [\overrightarrow{\mathbf{h}}_i; \overleftarrow{\mathbf{h}}_i]. \tag{26}$$

A temporal attention layer assigns larger weights to informative time steps:

In Equation (27), the attention weight represents the relative importance of each hidden state. A larger weight means that the corresponding time step contributes more to the prediction. This is useful because the samples in the sliding window do not have the same influence on future metering-error changes.

$$s_i = \mathbf{v}^{\mathrm{T}} \tanh(\mathbf{W}_a \mathbf{h}_i + \mathbf{b}_a),\ \beta_i = \frac{\exp(s_i)}{\sum_{j=1}^{L_{\mathrm{w}}} \exp(s_j)},\ \mathbf{c}_t = \sum_{i=1}^{L_{\mathrm{w}}} \beta_i \mathbf{h}_i. \tag{27}$$

To preserve the local inertia at the end of the window, a residual mapping is added:

The residual mapping in Equation (28) introduces the latest input into the attention feature to retain short-term inertia.

$$\mathbf{z}_t = \mathbf{c}_t + \mathbf{W}_s \mathbf{x}(t) + \mathbf{b}_s, \tag{28}$$

followed by

$$\widetilde{\mathbf{y}} = \mathbf{W}_o \sigma(\mathbf{z}_t) + \mathbf{b}_o, \tag{29}$$

where $\sigma(\cdot)$ is a nonlinear activation function.

Finally, the fully connected output layer maps the residual-attention feature to the target variables of the corresponding stage. In this way, the same RA-BiGRU structure can be used for both medium- and low-frequency mode-increment prediction and residual prediction.

4.3. Stage 1: Medium- and Low-Frequency Mode-Increment Prediction

The future increments of the last two ratio-error modes are

$$\begin{aligned} \Delta u_{\mathrm{r},K-1}(t+h) &= u_{\mathrm{r},K-1}(t+h) - u_{\mathrm{r},K-1}(t), \\ \Delta u_{\mathrm{r},K}(t+h) &= u_{\mathrm{r},K}(t+h) - u_{\mathrm{r},K}(t), \end{aligned} \tag{30}$$

and the corresponding phase-displacement-mode increments are

$$\begin{aligned} \Delta u_{\phi,K-1}(t+h) &= u_{\phi,K-1}(t+h) - u_{\phi,K-1}(t), \\ \Delta u_{\phi,K}(t+h) &= u_{\phi,K}(t+h) - u_{\phi,K}(t). \end{aligned} \tag{31}$$

The Stage-1 output is

$$\hat{\mathbf{y}}^{(1)}(t+h) = \left[\hat{\Delta u}_{\mathrm{r},K-1}, \hat{\Delta u}_{\mathrm{r},K}, \hat{\Delta u}_{\phi,K-1}, \hat{\Delta u}_{\phi,K}\right]^{\mathrm{T}}. \tag{32}$$

The preliminary forecasts are reconstructed as

$$\hat{e}_{\mathrm{r}}^{(1)}(t+h) = \sum_{k=1}^{K-2} u_{\mathrm{r},k}(t) + \sum_{k=K-1}^{K} \left[u_{\mathrm{r},k}(t) + \hat{\Delta u}_{\mathrm{r},k}(t+h)\right] + r_{\mathrm{r}}(t), \tag{33}$$

$$\hat{e}_{\phi}^{(1)}(t+h) = \sum_{k=1}^{K-2} u_{\phi,k}(t) + \sum_{k=K-1}^{K} \left[u_{\phi,k}(t) + \hat{\Delta u}_{\phi,k}(t+h)\right] + r_{\phi}(t). \tag{34}$$

Thus, the first $K-2$ modes and the residual are propagated using their current values, whereas the final two trend modes are corrected by their predicted increments.

4.4. Stage 2: Residual Correction

The Stage-1 residual targets are

$$\varepsilon_{\mathrm{r}}(t+h) = e_{\mathrm{r}}(t+h) - \hat{e}_{\mathrm{r}}^{(1)}(t+h), \ \varepsilon_{\phi}(t+h) = e_{\phi}(t+h) - \hat{e}_{\phi}^{(1)}(t+h). \tag{35}$$

The second RA-BiGRU predicts

$$\hat{\varepsilon}(t+h) = \left[\hat{\varepsilon}_{\mathrm{r}}(t+h), \hat{\varepsilon}_{\phi}(t+h)\right]^{\mathrm{T}}. \tag{36}$$

To avoid direct and potentially excessive residual superposition, an error-specific correction matrix is introduced:

$$\Gamma_h = \mathrm{diag}(\gamma_{\mathrm{r},h}, \gamma_{\phi,h}), \ \gamma_{\mathrm{r},h}, \gamma_{\phi,h} \in \{0, 0.05, 0.10, \ldots, 1.00\}. \tag{37}$$

The weights are selected on the validation set separately for each prediction horizon and error type. For a given prediction horizon and error type, each candidate weight in {0, 0.05, 0.10, …, 1.00} is applied to the Stage-2 residual prediction on the validation set. The candidate that gives the minimum validation RMSE is selected as the final correction weight.

After being selected, the correction weight is fixed and directly used on the test set. Thus, the test samples are not involved in determining the correction weight. This validation-based selection balances insufficient correction and over-correction. The final forecast is

$$\hat{\mathbf{e}}(t+h) = \hat{\mathbf{e}}^{(1)}(t+h) + \boldsymbol{\Gamma}_h \hat{\boldsymbol{\varepsilon}}(t+h), \tag{38}$$

that is,

$$\hat{\varepsilon}_{\mathrm{r}}(t+h) = \hat{\varepsilon}_{\mathrm{r}}^{(1)}(t+h) + \gamma_{\mathrm{r},h}\hat{\varepsilon}_{\mathrm{r}}(t+h),\ \hat{\varepsilon}_{\phi}(t+h) = \hat{\varepsilon}_{\phi}^{(1)}(t+h) + \gamma_{\phi,h}\hat{\varepsilon}_{\phi}(t+h). \tag{39}$$

5. Experimental Setup

5.1. Dataset and Data Partitioning

The experiments use continuous monitoring data from an operating ECT. The variables comprise ratio error, phase displacement, load, temperature, humidity, vibration, and magnetic field. After validity screening, 15,812 samples were retained. To preserve temporal order, the data were divided chronologically into 11,068 training samples, 2371 validation samples, and 2373 test samples.

The VMD parameters were set to $K = 11$ and $\alpha = 50.026679$. The model input therefore contains 29 features. The sliding-window length was $L_{\mathrm{w}} = 24$, and the prediction horizons were $h \in \{1, 3, 6, 12\}$.

Computational cost was evaluated under the same software and hardware environment. Training time was recorded from network initialization to completion, while testing time was measured for all 2373 test samples. The average inference time per sample was obtained from the total testing time, and each timing experiment was repeated five times.

5.2. Evaluation Metrics

For N test samples, the evaluation metrics are

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{40}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \tag{41}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%, \tag{42}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\max(y) - \min(y)} \times 100\%, \tag{43}$$

and

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2}. \tag{44}$$

Lower RMSE, MAE, MAPE, and NRMSE values indicate smaller prediction deviations, whereas an R^2 value closer to one indicates a better fit. Because metering-error values may

approach zero, MAPE can become unstable; consequently, the principal comparisons in this study use RMSE, MAE, and R^2.

5.3. Compared Models

The main comparison includes Persistence, Stage-1 RA-BiGRU, and the complete two-stage RA-BiGRU. This comparison is designed primarily as an ablation analysis to distinguish the contributions of trend-mode increment prediction and residual correction.

6. Results and Discussion

6.1. Forecasting Performance at $h = 1$

Figure 6 compares the measured values with the predictions at $h = 1$. Persistence follows the overall trajectory but exhibits visible deviations during local fluctuations. Stage 1 improves trend tracking, and the two-stage model follows the measured sequence more closely after residual correction.

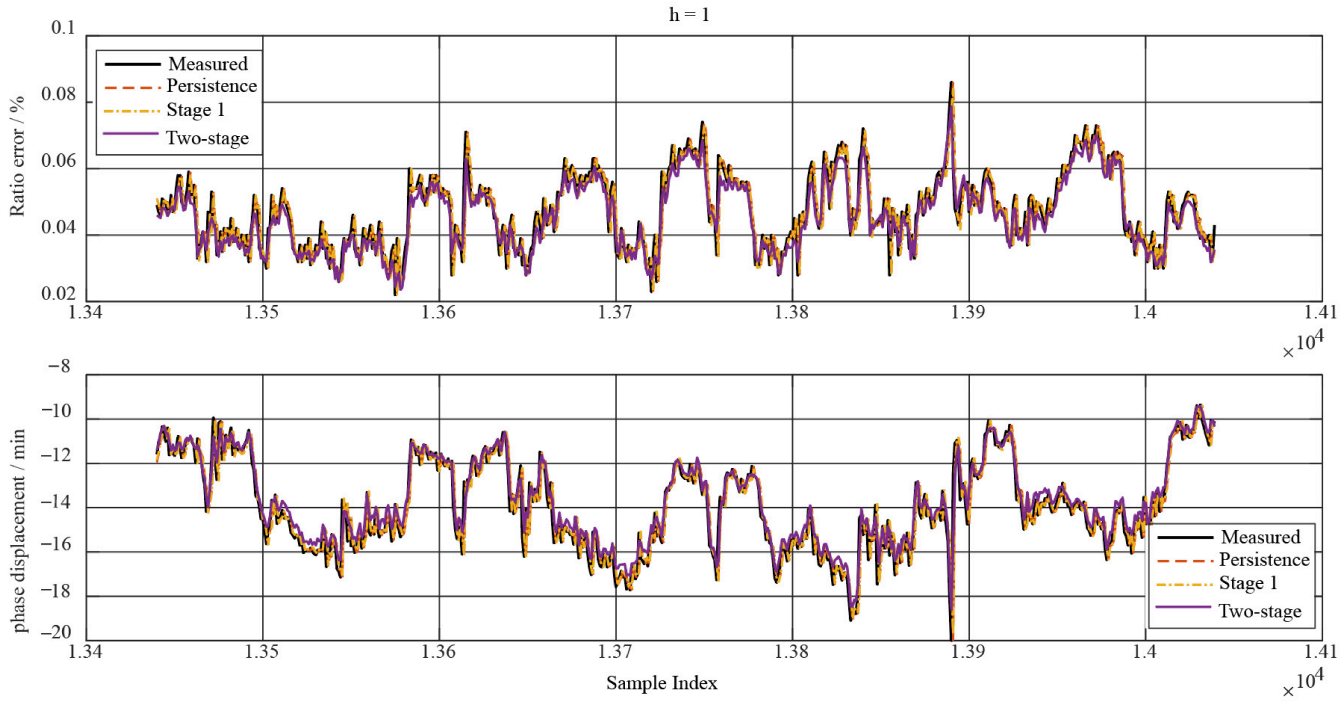


Figure 6. Prediction results on the test set at $h = 1$.

Table 3 reports the quantitative results. For ratio error, Stage 1 provides a modest improvement in RMSE and R^2, but its MAE remains higher than that of Persistence, indicating that trend-increment prediction alone does not consistently reduce all error measures. Stage 2 reduces RMSE to 0.003081 and increases R^2 to 0.8871. For phase displacement, the two-stage model reduces RMSE to 0.4942 min and increases R^2 to 0.9564.

Table 3. Prediction performance at $h = 1$.

Error Type	Model	RMSE	MAE	R^2
Ratio error	Persistence	0.005773	0.004358	0.6038
	Stage-1 RA-BiGRU	0.005719	0.005330	0.6112
	Two-stage RA-BiGRU	0.003081	0.002438	0.8871
Phase displacement	Persistence	0.8960	0.6577	0.8567
	Stage-1 RA-BiGRU	0.7514	0.6524	0.8588
	Two-stage RA-BiGRU	0.4942	0.3831	0.9564

The quantitative progression also clarifies the roles of the two forecasting stages. At h = 1, Stage 1 reduces the ratio-error and phase-displacement RMSE values by 0.94% and 16.14% relative to Persistence. After Stage 2, the total reductions reach 46.63% and 44.84%, showing that residual correction provides the main additional improvement, especially for ratio error.

6.2. Ablation Analysis and Correction Weights

The progression from Persistence to Stage 1 and then to the full two-stage model isolates the contributions of trend-increment prediction and residual correction. The final model outperforms the Persistence baseline at every horizon, as shown in Table 4. The largest improvement occurs at $h = 3$, where the RMSE reductions reach 54.27% for ratio error and 50.37% for phase displacement. At $h = 12$, the corresponding reductions remain 30.40% and 42.62%, indicating that the residual-correction mechanism remains effective for longer-horizon forecasting.

Table 4. Improvement of the two-stage RA-BiGRU over Persistence.

Horizon	Ratio RMSE/%	Ratio MAE/%	Phase RMSE/%	Phase MAE/%
1	46.63	44.05	44.84	41.74
3	54.27	52.14	50.37	45.81
6	23.75	21.47	33.73	27.48
12	30.40	29.52	42.62	40.85

The improvement is not monotonic across horizons, indicating that forecasting performance depends not only on horizon length but also on the predictability of the remaining residual. This observation is consistent with the horizon-dependent correction weights in Table 5.

Table 5. Validation-selected residual-correction weights.

Horizon	$\gamma_{r,h}$	$\gamma_{\phi,h}$
1	1.00	0.85
3	0.85	0.70
6	0.55	0.85
12	0.45	1.00

The validation-selected correction weights are listed in Table 5. Their variation across horizons and error types confirms that a fixed residual-correction strength would be inappropriate. The validation search balances under-correction against the amplification of Stage-2 prediction errors.

The Stage-1-to-final RMSE reductions quantify the horizon-dependent contribution of residual correction. For ratio error, the additional reductions are 46.12%, 52.51%, 18.51%, and 10.81% at h = 1, 3, 6, and 12, respectively; for phase displacement, the corresponding reductions are 44.44%, 49.24%, 30.83%, and 36.91%. The weaker long-horizon contribution for ratio error is consistent with the smaller validation-selected weights at h = 6 and 12, whereas the phase-displacement residual remains sufficiently predictable to provide clear improvement at all horizons. These results support horizon- and error-specific weights instead of a fixed correction strength.

Figure 7 compares the absolute ratio-error prediction errors before and after residual correction at $h = 12$. The corrected curve is generally lower, particularly around several high-error regions, although individual samples may still show limited over-correction.

This observation supports the use of validation-selected weights rather than unrestricted residual addition.

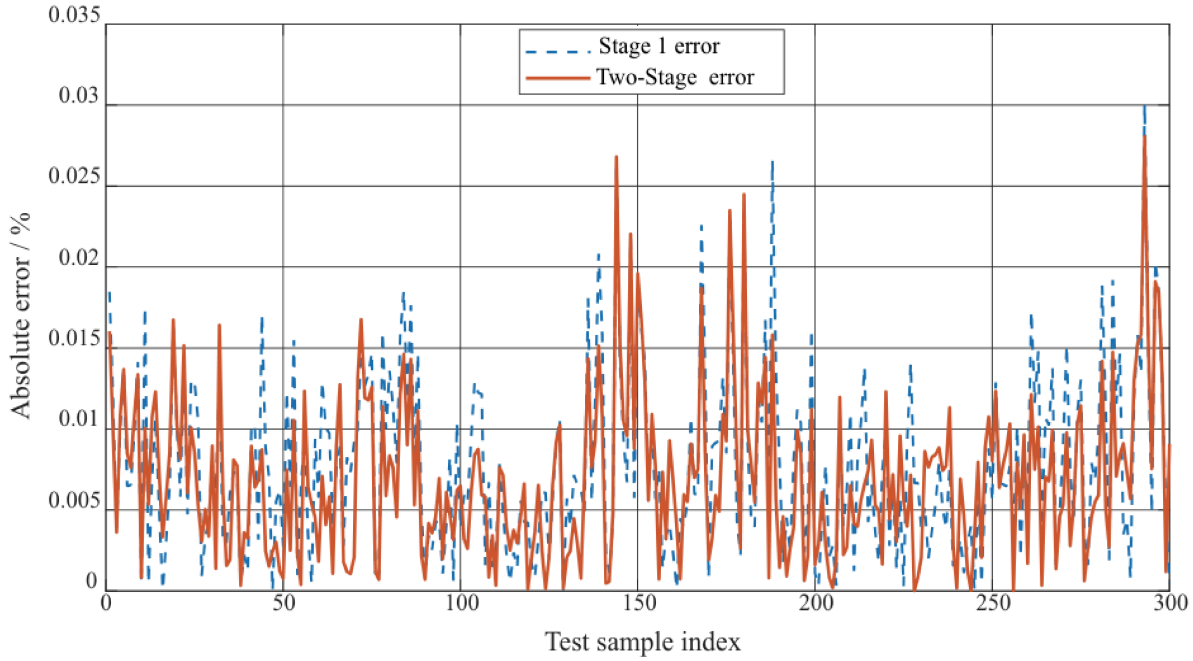


Figure 7. Absolute-error comparison before and after residual correction at $h = 12$.

6.3. Multi-Horizon Forecasting

Figure 8 summarizes the ratio-error RMSE values at the four horizons. Prediction errors increase overall as the horizon grows, as expected, but the two-stage model maintains a lower RMSE than Persistence in all cases. The full prediction traces for $h = 3$, $h = 6$, and $h = 12$ are provided in Appendix A. This indicates that the proposed trend-increment prediction and residual correction remain effective under different forecasting horizons.

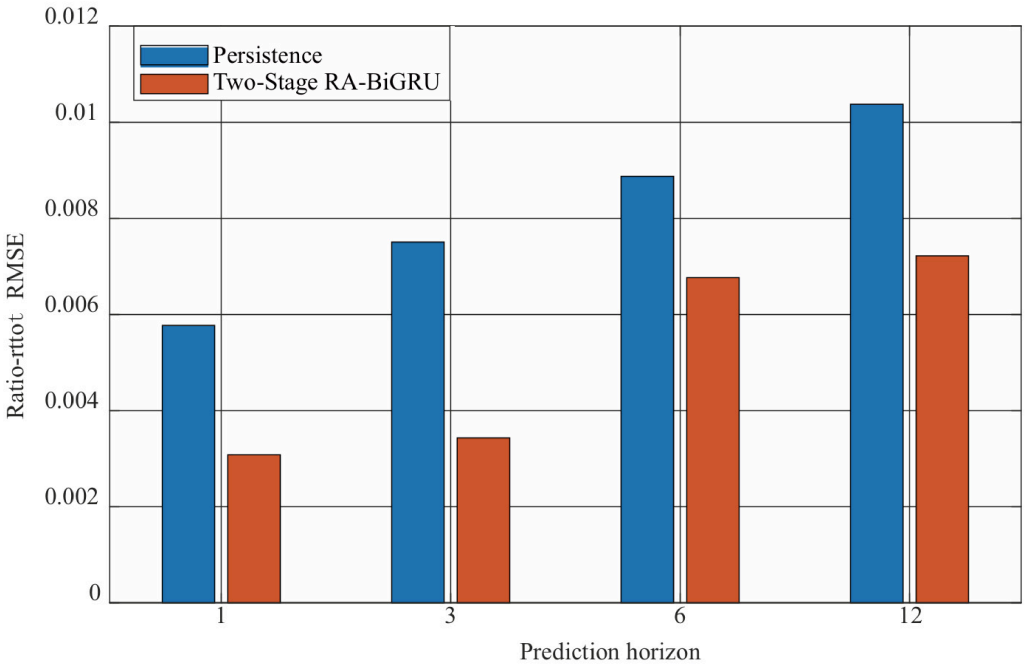


Figure 8. Ratio-error RMSE at different prediction horizons.

The supplementary recurrent-model comparison further shows that the improvement is not caused by the recurrent unit alone. At h = 12, the ratio-error RMSE values of GRU, BiLSTM, and single-stage RA-BiGRU are 0.009587, 0.009398, and 0.009377, respectively, whereas the complete two-stage framework achieves 0.007222. For phase displacement, the corresponding RMSE values are 1.4136, 1.5108, and 1.3970 min, compared with 0.9515 min for the proposed framework. This confirms that the main gain is produced by the coordinated use of adaptive decomposition, mode-increment prediction, and residual correction.

To further verify the influence of the VMD parameters on forecasting performance, K and α were varied while the data partition, forecasting model, and other settings were kept unchanged. The validation RMSE was calculated for each parameter combination.

As shown in Figure 9, the minimum validation RMSE values for ratio error and phase displacement are obtained at K = 11 and α = 50, reaching 0.0054 and 0.718, respectively. Within the tested range, the RMSE generally increases when K deviates from 11 or α increases above 50. The selected α is also close to the IWOA-refined value of 50.026679, which further supports the final parameter setting.

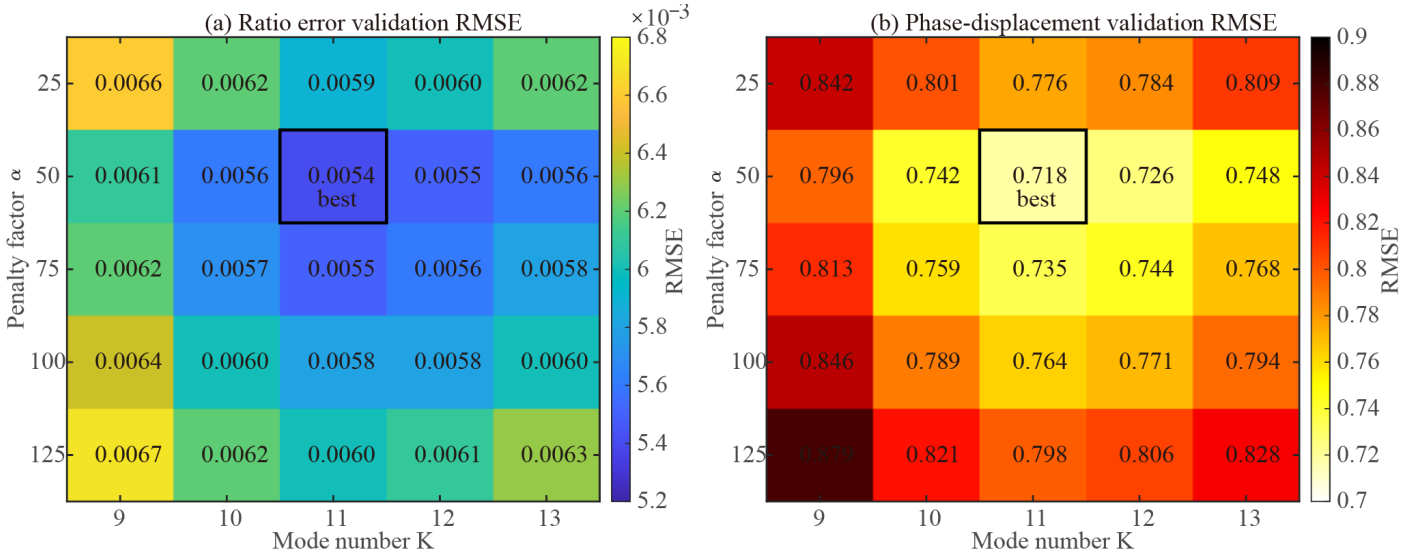


Figure 9. Validation RMSE under different values of K and α: (**a**) ratio error; (**b**) phase displacement.

The complete two-stage model requires training an additional residual-prediction RA-BiGRU compared with Stage 1, so its offline training cost is higher. However, IWOA-VMD optimization and network training are completed offline; online prediction only requires feature updating and forward inference through the trained networks. The computational cost is summarized in Table 6.

Table 6. Computational cost of the compared models.

Model	Training Time/s	Test Time for 2373 Samples/s	Inference Time/ms per Sample
Persistence	—	0.003	0.001
Stage-1 RA-BiGRU	78.4	0.156	0.066
Two-stage RA-BiGRU	116.8	0.0254	0.107

The reported values are averaged over the four prediction horizons. The one-time IWOA-VMD parameter optimization is excluded because it is performed offline before network training. The training time of the two-stage model includes the training of both the mode-increment prediction network and the residual-prediction network.

6.4. Discussion and Limitations

The results indicate that the proposed framework benefits from two complementary mechanisms. First, VMD separates rapid local fluctuations from smoother trend components, and Stage 1 predicts only the increments of the final two trend-dominant modes. This targeted prediction reduces the tendency of the network to reproduce the current state. Second, Stage 2 learns the systematic residuals that remain after the preliminary reconstruction. The validation-selected weights control the contribution of the residual prediction because its reliability varies with the prediction horizon and error type. Therefore, the observed improvement should be attributed to the coordinated use of constrained decomposition, selective trend-mode increment prediction, and controlled residual correction, rather than to VMD or RA-BiGRU alone.

For practical application, the framework has the potential to operate in an offline-training and online-prediction manner. IWOA-VMD parameter optimization, network training, and correction-weight selection can be completed offline using historical monitoring data. During online prediction, the optimized parameters and trained networks remain fixed, and the latest sliding-window data are used for feature updating and forward inference. However, a rolling-window or causal decomposition strategy is required to ensure that future information is not used during strictly online forecasting.

Several limitations should be addressed in future work. The present evaluation is based on one field dataset and a limited set of baselines. A controlled comparison with GRU, LSTM, BiGRU, and their decomposition-assisted variants has not yet been completed under identical data partitions, input structures, hyperparameter-tuning budgets, and repeated training runs. Future studies will therefore include these comparisons, together with alternative decomposition methods and VMD optimization strategies. Cross-device, cross-site, and cross-season validation, repeated experiments with different random seeds, confidence intervals, statistical significance tests, strict data-leakage checks, and end-to-end computational-cost evaluation including online decomposition are also required for a more comprehensive assessment.

7. Conclusions

This study proposed an improved-VMD two-stage RA-BiGRU framework for forecasting the ratio error and phase displacement of electronic current transformers under multivariable operating conditions. The dual-constraint IWOA-VMD method combines a composite objective with reconstruction-error and mode-number constraints to avoid low-mode under-decomposition. The optimized parameters were $K = 11$ and $\alpha = 50.026679$, providing accurate reconstruction of both metering-error sequences.

In Stage 1, the future increments of the final two trend-dominant modes are predicted to obtain preliminary forecasts. In Stage 2, the remaining residuals are estimated and combined with validation-selected correction weights to reduce over-correction. Experiments at prediction horizons of 1, 3, 6, and 12 showed that the complete framework consistently outperformed Persistence. At $h = 3$, the ratio-error and phase-displacement RMSE values were reduced by 54.27% and 50.37%, respectively; at $h = 12$, the reductions remained 30.40% and 42.62%.

These results confirm the complementary roles of constrained decomposition, trend-mode increment prediction, and controlled residual correction. Future work will focus on multi-device validation, comparisons with representative recurrent models, and rolling-window or causal decomposition for online implementation.

Author Contributions: Conceptualization, Y.M., H.C. and Y.X. (Yanfeng Xiao); methodology, Y.X. (Yijun Xiao); software, Y.M. and H.C.; validation, Y.M., H.C. and Z.L.; formal analysis, Y.X. (Yijun

Xiao); investigation, Y.M.; resources, H.C.; data curation, Z.L.; writing—original draft preparation, Y.X. (Yijun Xiao) and Y.M.; writing—review and editing, Y.M. and H.C.; visualization, Z.L.; supervision, Y.X. (Yanfeng Xiao); project administration, Y.M.; funding acquisition, Y.M., Y.X. (Yanfeng Xiao) and Y.X. (Yijun Xiao). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Open Research Fund of the Hubei Technology Innovation Center for Smart Hydropower (HBCXZX-JJ-202407).

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request, subject to the data-sharing restrictions of the operating organization.

Conflicts of Interest: Author Yedong Mao, Yanfeng Xiao and Yijun Xiao were employed by the company Hubei Technology Innovation Center for Smart Hydropower and China Yangtze Power Co., Ltd. Author Haoyu Chen and Zhenhua Li were employed by College of Electrical Engineering & New Energy, China Three Gorges University. The authors declare that this study received funding from the Open Research Fund of the Hubei Technology Innovation Center for Smart Hydropower (HBCXZX-JJ-202407). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Abbreviations

The following abbreviations are used in this manuscript:

ECT	Electronic current transformer
VMD	Variational mode decomposition
WOA	Whale optimization algorithm
IWOA	Improved whale optimization algorithm
GRU	Gated recurrent unit
BiGRU	Bidirectional gated recurrent unit
RA-BiGRU	Residual-attention bidirectional gated recurrent unit
IMF	Intrinsic mode function
RMSE	Root-mean-square error
MAE	Mean absolute error
MAPE	Mean absolute percentage error
NRMSE	Normalized root-mean-square error

Appendix A. Prediction Results at Additional Horizons

Figures A1–A3 show the test-set predictions at horizons of 3, 6, and 12, respectively.

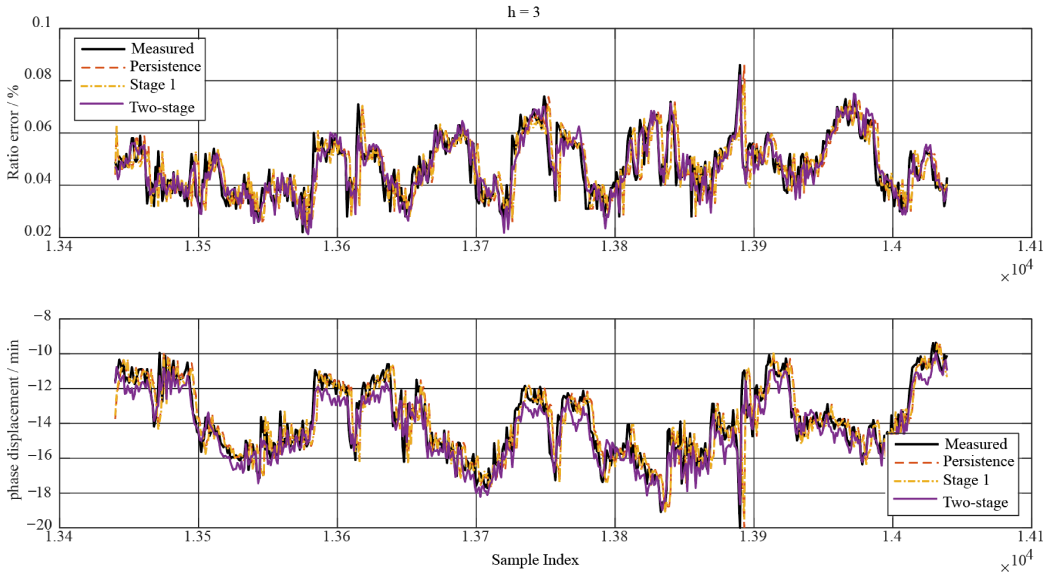


Figure A1. Prediction results at $h = 3$.

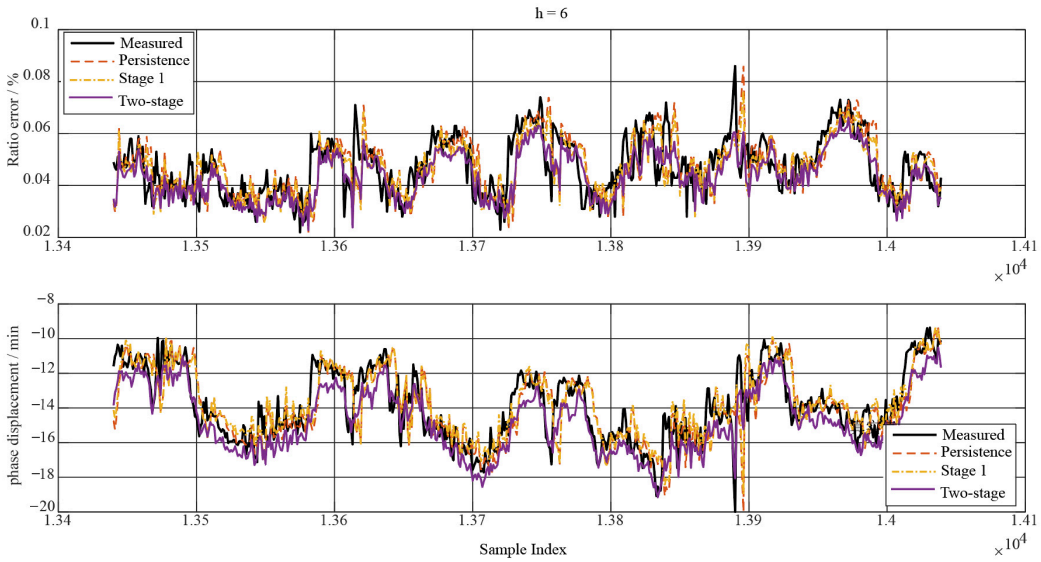


Figure A2. Prediction results at $h = 6$.

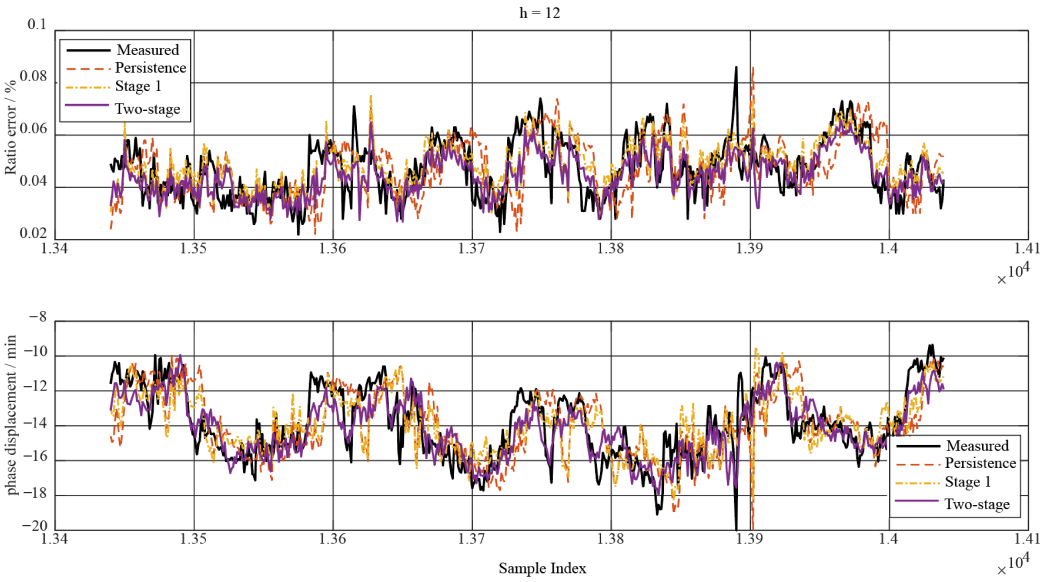


Figure A3. Prediction results at $h = 12$.

References

1. Cui, J.; Li, Z.; Lu, H. A data-driven hybrid framework for voltage transformer ratio error prediction: Addressing challenges in complex power systems. *Int. J. Electr. Power Energy Syst.* **2025**, *170*, 110825. [CrossRef]
2. Li, Z.; Cui, J.; Lu, H.; Zhou, F.; Diao, Y.; Li, Z. Prediction method for instrument transformer measurement error: Adaptive decomposition and hybrid deep learning models. *Measurement* **2025**, *253*, 117592. [CrossRef]
3. Chen, K.-L.; Wan, R.-S.; Guo, Y.; Chen, N.; Lee, W.-J. A redundancy mechanism design for Hall-based electronic current transformers. *Energies* **2017**, *10*, 312. [CrossRef]
4. Xiao, X.; Hu, H.; Xu, Y.; Lei, M.; Xiong, Q. Research on the error characteristics of a 110 kV optical voltage transformer under three conditions: In the laboratory, off-line in the field and during on-line operation. *Sensors* **2016**, *16*, 1303. [CrossRef] [PubMed]
5. Mingotti, A.; Bartolomei, L.; Peretto, L.; Tinarelli, R. On the long-period accuracy behavior of inductive and low-power instrument transformers. *Sensors* **2020**, *20*, 5810. [CrossRef] [PubMed]
6. Li, Z.; Li, Q.; Wu, Z.; Li, Z. Research into an online calibration system for the errors of voltage transformers based on open–closed capacitor. *Energies* **2018**, *11*, 1455. [CrossRef]
7. Mingotti, A.; Peretto, L.; Tinarelli, R. Calibration procedure to test the effects of multiple influence quantities on low-power voltage transformers. *Sensors* **2020**, *20*, 1172. [CrossRef] [PubMed]
8. Bianchi, F.M.; De Santis, E.; Rizzi, A.; Sadeghian, A. Short-term electric load forecasting using echo state networks and PCA decomposition. *IEEE Access* **2015**, *3*, 1931–1943. [CrossRef]

9. Zhu, Q.; Chen, J.; Zhu, L.; Duan, X.; Liu, Y. Wind speed prediction with spatio–temporal correlation: A deep learning approach. *Energies* **2018**, *11*, 705. [CrossRef]
10. Lim, B.; Arik, S.O.; Loeff, N.; Pfister, T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [CrossRef]
11. Qin, Y.; Song, D.; Chen, H.; Cheng, W.; Jiang, G.; Cottrell, G.W. A dual-stage attention-based recurrent neural network for time series prediction. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 2627–2633. [CrossRef] [PubMed]
12. Cho, K.; van Merri"enboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [CrossRef]
13. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
15. Sun, L.; Qin, H.; Przystupa, K.; Majka, M.; Kochan, O. Individualized short-term electric load forecasting using a data-driven meta-heuristic method based on an LSTM network. *Sensors* **2022**, *22*, 7900. [CrossRef] [PubMed]
16. Xu, X.; Hu, S.; Shao, H.; Shi, P.; Li, R.; Li, D. A spatio-temporal forecasting model using an optimally weighted graph convolutional network and gated recurrent unit for wind speeds at different sites distributed in an offshore wind farm. *Energy* **2023**, *284*, 128565. [CrossRef]
17. Niu, Z.; Zhong, G.; Yue, G.; Wang, L.-N.; Yu, H.; Ling, X.; Dong, J. Recurrent attention unit: A new gated recurrent unit for long-term memory of important parts in sequential data. *Neurocomputing* **2023**, *517*, 1–9. [CrossRef]
18. Xu, Y.; Li, Y.; Xiao, X.; Xu, Z.; Hu, H. Monitoring and analysis of electronic current transformer field operating errors. *Measurement* **2017**, *112*, 117–124. [CrossRef]
19. Zhang, Z.; Chen, Q.; Hu, C.; Li, H.; Chen, M. Evaluating the metering error of electronic transformers online based on VN-MWPCA. *Measurement* **2018**, *130*, 1–7. [CrossRef]
20. Li, Z.; Zheng, Y.; Abu-Siada, A.; Lu, M.; Li, H.; Xu, Y. Online evaluation for the accuracy of electronic voltage transformers based on recursive principal component analysis. *Energies* **2020**, *13*, 5576. [CrossRef]
21. Xiong, G.; Przystupa, K.; Teng, Y.; Xue, W.; Huan, W.; Feng, Z.; Qiong, X.; Wang, C.; Skowron, M.; Kochan, O.; et al. Online measurement error detection for the electronic transformer in a smart grid. *Energies* **2021**, *14*, 3551. [CrossRef]
22. Zhang, Y.; Zhang, C.; Li, H.; Chen, Q. An online detection method for capacitor voltage transformers with excessive measurement errors based on multi-source heterogeneous data fusion. *Measurement* **2022**, *187*, 110262. [CrossRef]
23. Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic, 22–27 May 2011; pp. 4144–4147. [CrossRef]
24. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
25. Rehman, N.U.; Aftab, H. Multivariate variational mode decomposition. *IEEE Trans. Signal Process.* **2019**, *67*, 6039–6052. [CrossRef]
26. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [CrossRef]
27. Li, Z.; Cui, J.; Rocha, P.R.F.; Abu-Siada, A.; Li, H.; Qiu, L. A predictive model for voltage transformer ratio error considering load variations. *World Electr. Veh. J.* **2024**, *15*, 269. [CrossRef]
28. Zhong, Y.; Li, T.; Przystupa, K.; Lin, C.; Yang, G.; Yang, S.; Kochan, O.; Sikora, J. Spatiotemporal correlation analysis for predicting current transformer errors in smart grids. *Energies* **2024**, *17*, 1608. [CrossRef]
29. Liu, J.; Hu, C.; Li, Z.; Cui, J. Multi-scale temporal learning with EEMD reconstruction for non-stationary error forecasting in current transformers. *Electronics* **2026**, *15*, 325. [CrossRef]
30. Fu, H.; Zhang, J.; Xie, S. A novel improved variational mode decomposition-temporal convolutional network-gated recurrent unit with multi-head attention mechanism for enhanced photovoltaic power forecasting. *Electronics* **2024**, *13*, 1837. [CrossRef]
31. Luo, B.; Zuo, P.; Zhu, L.; Hua, W. A wind power density forecasting model based on RF-DBO-VMD feature selection and BiGRU optimized by the attention mechanism. *Atmosphere* **2025**, *16*, 266. [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.