



Article

# A Predictive Model for Voltage Transformer Ratio Error Considering Load Variations

Zhenhua Li <sup>1,2</sup>, Jiuxi Cui <sup>1,\*</sup>, Paulo R. F. Rocha <sup>3</sup>, Ahmed Abu-Siada <sup>4</sup>, Hongbin Li <sup>5</sup> and Li Qiu <sup>1,2</sup>

- <sup>1</sup> College of Electrical Engineering & New Energy, China Three Gorges University, Yichang 443002, China; lizhenhua@ctgu.edu.cn (Z.L.)
  - <sup>2</sup> Provincial Engineering Research Center of Intelligent Energy Technology, China Three Gorges University, Yichang 443002, China
  - <sup>3</sup> Centre for Functional Ecology-Science for People & the Planet, Associate Laboratory TERRA, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal; procha@uc.pt
  - <sup>4</sup> Department of Electrical and Computer Engineering, Curtin University, Perth, WA 6102, Australia; a.abusiada@curtin.edu.au
  - <sup>5</sup> School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
- \* Correspondence: cuiup@ctgu.edu.cn

**Abstract:** The accuracy of voltage transformer (VT) measurements is imperative for the security and reliability of power systems and the equitability of energy transactions. The integration of a substantial number of electric vehicles (EVs) and their charging infrastructures into the grid poses new challenges for VT measurement fidelity, including voltage instabilities and harmonic disruptions. This paper introduces an innovative transformer measurement error prediction model that synthesizes Multivariate Variational Mode Decomposition (MVMD) with a deep learning framework integrating Bidirectional Temporal Convolutional Network and Multi-Head Attention mechanism (BiTCN-MHA). The paper is aimed at enhancing VT measurement accuracy under fluctuating load conditions. Initially, the optimization of parameter selection within the MVMD algorithm enhances the accuracy and interpretability of bi-channel signal decomposition. Subsequently, the model applies the Spearman rank correlation coefficient to extract dominant modal components from both the decomposed load and original ratio error sequences to form the basis for input signal channels in the BiTCN-MHA model. By superimposing predictive components, an effective prediction of future VT measurement error trends can be achieved. This comprehensive approach, accounting for input load correlations and temporal dynamics, facilitates robust predictions of future VT measurement error trends. Computational example analysis of empirical operational VT data shows that, compared to before decomposition, the proposed method reduces the Root-Mean-Square Error (RMSE) by 17.9% and the Mean Absolute Error (MAE) by 23.2%, confirming the method's robustness and superiority in accurately forecasting VT measurement error trends.



**Citation:** 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. <https://doi.org/10.3390/wevj15060269>

Academic Editor: Joeri Van Mierlo

Received: 15 May 2024

Revised: 6 June 2024

Accepted: 14 June 2024

Published: 19 June 2024

**Keywords:** voltage transformer; measurement error prediction; dynamic load; MVMD; spearman rank correlation coefficient; BiTCN-MHA



**Copyright:** © 2024 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/>).

## 1. Introduction

Voltage transformers are key components in modern substations as they are used for measuring voltage signals that are required for power system protection and control [1]. To ensure the reliability of the measurement results of voltage transformers, their performance must meet the 0.2 class standard.

In recent years, with the global emphasis on environmental protection and sustainable development, electric vehicles (EVs) have emerged as an effective means to reduce greenhouse gas emissions and the consumption of mineral resources. Consequently, their market share is rapidly growing worldwide. According to the International Energy Agency

(IEA) report, the global ownership of electric vehicles exceeded 10 million in 2020, marking a 43% increase from the year 2019, with electric vehicles accounting for two-thirds of the total EV registrations for the year [2]. In China, the sales of purely electric vehicles reached approximately 5.365 million units in the year 2022. As the ownership of electric vehicles increases, the corresponding charging infrastructure is also rapidly expanding. By the end of 2021, the number of electric vehicle charging stations in China exceeded 1.4 million, with more than 10 million charging piles [3]. Moreover, related research indicates that, even at moderate development speeds, the proportion of electric vehicles in the US car market is expected to reach 35% by 2030 and 51% by 2050 [4]. To meet the growing EV charging demand, countries around the world are striving to build EV charging facilities on a large scale.

However, the rapid growth of electric vehicles and their charging demands pose significant challenges to grid stability and power quality. EVs act as nonlinear loads during the charging process, which may cause a wide range of problems when connected en masse to the active distribution grid, especially under uncoordinated charging conditions. This leads to large-scale grid integration effects, causing voltage fluctuations [5], increased load variance [6], heightened network losses in the power system [7], harmonic pollution [8], three-phase imbalance [9], and voltage distortion [10], among others. Consequently, the electrical signal parameters of the power system exhibit multimodal and complex time-varying characteristics. In this context, the challenge of measurement accuracy faced by voltage transformers becomes increasingly complex, especially since the fluctuation range of the primary side's for the random signals they sense may far exceed the 0.2% standard.

With the increasing integration of charging stations into the grid, these challenges are anticipated to significantly increase. In this context, distinguishing between gradual measurement errors caused by transformer degradation and the inherent fluctuations of the sensing signals becomes a significant challenge in evaluating the measurement accuracy of voltage transformers. Thus, there is an urgent necessity for a novel method to precisely assess the measurement errors of voltage transformers under variations under continuous load, ensuring trade fairness and grid security.

In recent years, with the development of artificial intelligence, algorithm-based models have been applied to the assessment of measurement errors in voltage transformers [11,12]. By making short-term predictions based on monitoring data, these models can provide state information for the future sampling moment, supporting risk warning, fault detection, and maintenance planning. For instance, reference [13] uses a BiLSTM network to directly predict the measurement errors of voltage transformers, while reference [14] employs GRU and MTL to predict the ratio error of voltage transformers. Reference [15] uses VMD to decompose the measurement error signals of voltage transformers and inputs features for prediction, but this approach overlooks the impact of decomposition residuals on model stability and lacks interpretability in empirical reconstruction. However, the direct application of all presented models so far does not take into account their generalizability and the impact of load changes on transformer measurement errors.

Currently, integrating theoretical knowledge from different disciplines has become an effective way to improve the generalizability and robustness of deep learning models. Techniques such as Fourier transform, Empirical Mode Decomposition (EMD), and Ensemble Empirical Mode Decomposition (EEMD) have been widely used in signal processing. However, the Fourier transform [16] is suitable for processing stationary signals and is commonly used to describe global oscillation characteristics. EMD [17] may encounter mode mixing issues during decomposition; EEMD [18], while mitigating the problem of mode mixing, suffers from low algorithm efficiency. The Variational Mode Decomposition (VMD) [19,20] method, aimed at nonlinear and non-stationary signals, achieves adaptive decomposition of signals through a variational framework. By iteratively updating the central frequencies and bandwidths of each oscillatory mode, it demonstrates good decomposition performance and noise resistance.

However, the aforementioned methods mainly target the identification of single-channel signals one by one, lacking the capability to process multi-channel signals. Multivariate Variational Mode Decomposition (MVMD) [21] constructs a variational optimization problem to extract multivariate signals, providing a theoretical foundation and solving issues related to endpoint effects and mode mixing.

This paper introduces a predictive model for assessing the measurement accuracy of voltage transformers by integrating MVMD with a Temporal Convolutional Network hybrid. This model is designed to evaluate and forecast the trends of measurement error deviations in voltage transformers influenced by load variations. Initially, the MVMD algorithm is enhanced through the implementation of a RIME optimization strategy, predicated on minimizing permutation entropy, which adaptively determines the optimal number of modal decompositions and penalty factors to refine decomposition accuracy and interpretability. Subsequently, the model decomposes load fluctuation data and the original ratio error of the transformer into Intrinsic Mode Functions (IMF) characterized by distinct frequency–amplitude profiles, utilizing the Spearman rank correlation coefficient to discern and isolate predominant modes. These principal modes are then configured as input signals for the hybrid Bidirectional Temporal Convolutional Network and Multi-Head Attention (BiTCN-MHA) model, facilitating the prediction of future ratio error fluctuations in voltage transformers. This model methodically accounts for the correlations among input loads and the dynamic evolution of time series. Through empirical analysis using real transformer data from substations, the efficacy of the proposed method in accurately predicting trends of 0.2-class ratio error fluctuations in voltage transformers has been verified as will be elaborated below.

In the following chapters, Section 2 analyzes transformer measurement errors. Section 3 explains and improves the Multivariate Variational Mode Decomposition (MVMD). Section 4 details the BiTCN-MHA model for predicting VT measurement errors. Section 5 presents a case study demonstrating the application and validation of the model using empirical data. Section 6 summarizes the research findings.

## 2. Transformer Measurement Error Analysis

VTs convert the primary side voltage of the power grid to a signal for secondary systems. Typically, there is a discrepancy between the actual voltage and the VT-derived secondary measurement data, commonly quantified as the ratio error  $f$  [22,23]:

$$f = \frac{K_r U_2 - U_1}{U_1} \times 100\% \quad (1)$$

where  $K_r$  represents the transformation ratio of the VT,  $U_1$  is the actual value of the primary side voltage signal, and  $U_2$  is the measured data of the secondary side output voltage.

To evaluate the influence of load fluctuations on the gradual error of operational transformers, the proposed method considers load fluctuations at any given time affecting the operational site of the transformer as  $X$ , with its time series matrix being  $X = (x_1, x_2, \dots, x_T)$ . Additionally, the transformer's historical ratio error is denoted by  $Z$ , and the predicted next moment's ratio error is expressed as:

$$\hat{f}_{T+1} = F_f(z_1, z_2, \dots, z_T, x_1, x_2, \dots, x_T) \quad (2)$$

Based on the above foundation, a predictive model is constructed based on load fluctuations and historical ratio error information.

## 3. Multivariate Variational Mode Decomposition

### 3.1. Principle of MVMD

MVMD extends Variational Mode Decomposition (VMD) to accommodate multivariate measurement data by transforming single-channel data into multiple channels using

the Frobenius norm. The essence of MVMD is to extract  $K$  oscillatory modes  $u_k^+(t)$  from the input data containing  $N$  data channels  $x(t)$ , as given by (3)

$$x(t) = \sum_{k=1}^K u_k(t) \tag{3}$$

The objective is to derive multivariate oscillatory modes  $\{u_k(t)\}$  from the input data, minimizing the bandwidth of these modes while ensuring accurate reconstruction of  $x(t)$ . Thus, the bandwidth of  $u_k(t)$  is estimated by employing the square  $L^2$  norm of the gradient of the demodulated signal  $u_k^+(t)$ :

$$g = \sum_{k=1}^K \|\partial_t(e^{-j\omega_k t} u_k^+(t))\|_2^2 \tag{4}$$

To identify multivariate oscillations with a singular common frequency component  $\omega_k$  in multi-channels, it is necessary to estimate the bandwidth of the modulated multivariate oscillatory signal, by shifting the one-sided spectrum of each channel of  $u_k^+(t)$  with center frequency  $\omega_k$ , utilizing the Frobenius norm to convert single-channel data into multivariate formats.

The Frobenius norm, which structures space topologically, is defined for a matrix  $\mathbf{W}$  as the square root of the sum of squares of all elements, expressed as follows:

$$\|\mathbf{W}\|_F = \sqrt{\sum_i \sum_j w_{i,j}^2} \tag{5}$$

Within the context of the MVMD algorithm, this is represented as follows:

$$g' = \sum_k \sum_h \|\partial_t(e^{-j\omega_k t} u_{k,n}^+(t))\|_2^2 \tag{6}$$

where  $u_{k,n}^+(t)$  is the analytic signal for channel number  $n$  and mode number  $k$ . The MVMD variational constraint model is constructed as follows:

$$\min_{\{u_{k,n}\}, \{\omega_k\}} \left\{ \sum_k \sum_n \|\partial_t(e^{-j\omega_k t} u_{k,n}^+(t))\|_2^2 \right\} \tag{7}$$

$$\text{s.t. } X_N(t) = \sum_k u_{k,n}^+(t), n = 1, 2, \dots, N \tag{8}$$

With multiple linear constraints, the corresponding augmented Lagrangian function is derived as follows:

$$L(\{u_{k,n}\}, \{\omega_k\}, \lambda_n) = \alpha \sum_{k=1}^K \sum_{n=1}^N \|\partial_t(e^{-j\omega_k t} u_{k,n}^+(t))\|_2^2 + \sum_{n=1}^N \left\| x_n(t) - \sum_{k=1}^K u_{k,n}(t) \right\|_2^2 + \sum_{n=1}^N \left\langle \lambda_n(t), x_n(t) - \sum_{k=1}^K u_{k,n}(t) \right\rangle \tag{9}$$

By iteratively updating  $u_k(t)$ ,  $\omega_k$ , and the Lagrangian multipliers  $\lambda$  through the Alternating Direction Method of Multipliers (ADMM), the optimal solution of the variational model, which represents all estimated frequency domain modes, can be obtained as follows:

$$\tilde{u}_{k,n}^{l+1}(\omega) = \frac{\hat{x}_n(\omega) - \sum_{k=1}^{k-1} \hat{u}_{i,n}^{l+1}(\omega) + \frac{\hat{\lambda}^l(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^l)^2} \tag{10}$$

where  $\alpha$  is the penalty factor;  $\hat{x}_n(\omega)$ ,  $\hat{\lambda}^l(\omega)$ ,  $\hat{u}_{k,n}^{l+1}(\omega)$ ,  $\hat{u}_{i,n}^{l+1}(\omega)$ , and  $\hat{u}_{i,n}^{l+1}(\omega)$  are the Fourier transforms of  $x_n(\omega)$ ,  $\lambda^l(\omega)$ ,  $u_{k,n}^{l+1}(\omega)$ , and  $u_{i,n}^{l+1}(\omega)$ , respectively; and  $l$  is the iteration number. The estimated central frequency of the modes is determined by

$$\omega_k^{l+1} = \frac{\sum \int_0^\infty \omega \left| \hat{u}_{k,n}^{l+1}(\omega) \right|^2 d\omega}{\sum \int_0^\infty \left| \hat{u}_{k,n}^{l+1}(\omega) \right|^2 d\omega} \quad (11)$$

### 3.2. Decomposition and Reconstruction

The effectiveness of MVMD outcomes primarily hinges on the parameters  $k$  and  $\alpha$ . The magnitude of  $k$  is pivotal in determining the precision of the decomposition results. A suboptimal  $k$ , being too low, may result in incomplete modal decomposition, whereas an excessively high  $k$  may lead to over-decomposition. The influence of  $\alpha$  on MVMD's performance is intertwined with the characteristics of both signal and noise. Therefore, the configuration of  $k$  and  $\alpha$  is vital for the MVMD process.

This paper formulates an objective function based on the minimal Permutation Entropy (PE), iteratively determines  $k$  and  $\alpha$  utilizing the RIME optimization algorithm, and selects the primary modal components of the signal channels through the Spearman rank correlation coefficient.

#### 3.2.1. Permutation Entropy

PE proposed by Bandt et al. [24] is a method used for assessing the complexity and dynamical alterations within time series data. This approach is noted for its simplicity in computation, robustness, and high computational efficiency.

$$H_{pe}(m) = - \sum_{j=1}^k P_{ej} \log P_{ej} \quad (12)$$

where  $P_{ej}$  represents the probability of the  $j$ th symbol occurrence;  $H_{pe}$  signifies the complexity and randomness level of the time series. A higher value indicates greater complexity and randomness of the series, whereas a lower value denotes a more regular series, implying enhanced predictability.

#### 3.2.2. RIME

The RIME [25] constitutes a novel metaheuristic framework, mirroring the frost crystallization phenomenon. Through comprehensive solution space examination coupled with the implementation of adaptive search paradigms, it proficiently amplifies the likelihood of ascertaining global optima, concurrently diminishing the susceptibility to suboptimal local minima entrapments.

#### 3.2.3. Spearman Rank Correlation Coefficient

The Spearman rank correlation coefficient [26] is a non-parametric measure of the dependency between two variables, which does not require the data to follow a normal distribution. In this paper, the Spearman rank correlation coefficient is employed to analyze the correlation among the decomposed signals:

$$\rho_{(X,Y)} = \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (x_i - \bar{y})^2}} \quad (13)$$

where  $\rho_{(x,y)}$  denotes the Spearman rank correlation between signals  $x$  and  $y$ . The value of  $\rho_{(x,y)}$  ranges between  $[-1, +1]$ , with its absolute value nearing 1 indicating a stronger correlation between the two signals, while a value close to 0 suggests an almost non-existent correlation relationship. The specific process is shown in Figure 1.

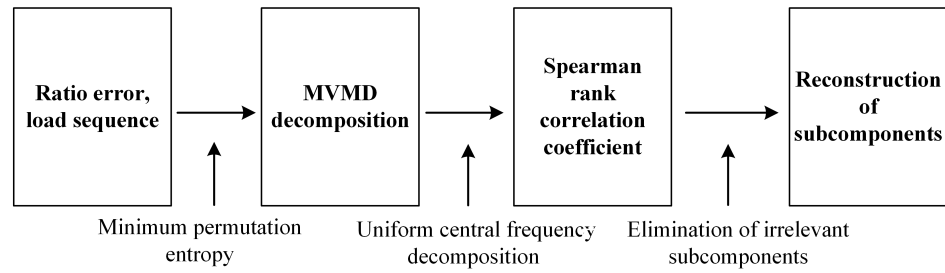


Figure 1. MVMD decomposition and reconstruction.

#### 4. BiTCN-MHA Model

##### 4.1. BiTCN

The Temporal Convolutional Network (TCN) consists of causal convolution, dilated convolution, and residual connections. As illustrated in Figure 2, the TCN employs a multi-layer architecture design, where each layer is stacked with residual blocks of varying dilation factors to enhance the model’s capability to process long sequence data while preserving essential original information.

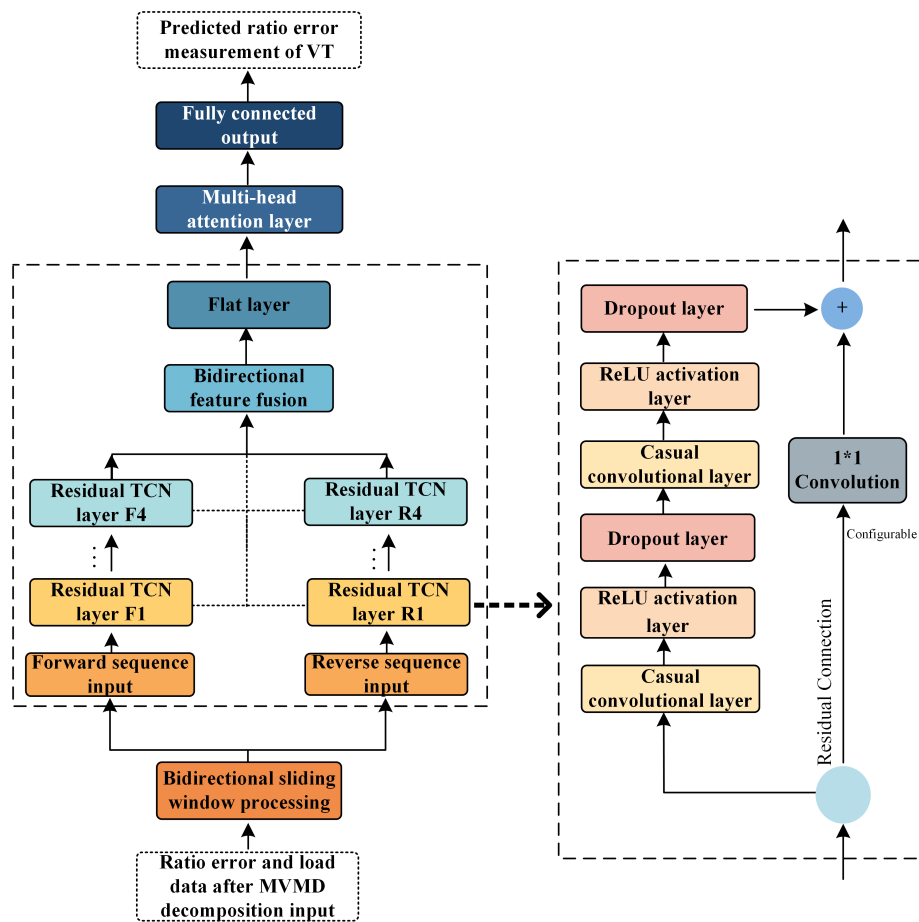


Figure 2. Based on the BiTCN-MHA deep learning model.

Causal convolution improves upon traditional one-dimensional convolution. Utilizing dilated causal convolution effectively prevents the leakage of future information, ensuring the posterior causality of the data. The model’s receptive field is expanded by adjusting the convolution kernel size, the number of layers, and the dilation factor, thereby capturing the dependency relationship between future moments and extended historical periods.

This design enables the TCN to utilize only data available up to the current moment at each step, thus avoiding interference from future data. The output at the  $t$ -th moment,  $z_t$  is given as follows:

$$z_t = g(u_0, u_1, u_2, \dots, u_t) \quad (14)$$

where  $g$  represents the one-dimensional convolution kernel.

Dilated convolution allows for an exponential expansion of the receptive field by introducing fixed intervals through the dilation factor, which reduces the computational demands for processing long sequence data. Moreover, employing a progressively increasing dilation factor strategy facilitates deep feature extraction from local details to comprehensive global insights, enriching the feature representation at the output layer. For the unidimensional data  $[u_0, u_1, \dots, u_t, u_{t+1}]$ , the expression for the output of dilated convolution at a given time  $t$  is

$$G(t) = (u \odot g)(t) = \sum_{i=0}^{k_{er}-1} v_{s.(i)} U_{t-m \cdot i} \quad (15)$$

where  $G(t)$  represents the output of the dilated convolution at time  $t$ ;  $v(i)$  is the weight of the convolution kernel at position  $i$ ;  $k_{er}$  represents the convolution kernel size;  $U_{t-m \cdot i}$  corresponds to the input sequence values after interval sampling; and  $m$  is the dilation factor.

Residual connections sum the model's extracted input features with the feature extraction results, assisting the model in preserving essential original information that might be lost during deep feature extraction. This enhances the model's stability and prevents the issue of vanishing gradients. The formula is as follows:

$$z = G(u, V) + u \quad (16)$$

where  $z$  is the output of the residual connection;  $u$  is the input; and  $G(u, V)$  is a residual network.

BiTCN enhances the capture of data dependencies through a bidirectional processing architecture, thereby efficiently extracting deep features of time series. Assuming a TCN consists of  $k$  stacked residual blocks, each residual block contains multiple layers of convolution, thus the output after passing through a residual block can be described as

$$y^{(j,k)} = [y_0^{(j,k)}, \dots, y_T^{(j,k)}] \quad (17)$$

$$y_t^{(j,k)} = \sum_{i=0}^{k_{er}-1} (f(i) \cdot y_{t-n \cdot i}^{(j-1,k)}) + y_t^{(1,k)} \quad (18)$$

where  $k$  is the index of the residual block,  $j$  is the layer index,  $y_t^{(j,k)}$  represents the convolutional kernel weight in the  $j$ th layer of the  $k$ th residual block, and  $T$  represents the total length of the sequence. To merge the features generated by the bidirectional processing paths, an additive fusion strategy is adopted to obtain the final output, as follows:

$$L_{BiTCN} = L_{for} \oplus L_{rev} \quad (19)$$

where  $L_{BiTCN}$  is the composite feature output of BiTCN, and  $L_{for}$  and  $L_{rev}$  are the feature sets outputted by the forward and reverse processing paths, respectively. By parallel processing of the forward and backward temporal information of the sequence, the capability to handle bidirectional dependencies is enhanced. This paper achieves multi-level feature mining from local to global through the flexible receptive field of BiTCN, identifying hidden features and improving the processing efficiency for long time series data.

#### 4.2. Multi-Head Attention Mechanism

The multi-head attention mechanism, an enhancement of the self-attention mechanism, aims to dynamically focus on important information in the sequence through a weighted method. It enables the model to learn information across multiple representation subspaces, thereby enhancing the ability to extract global contextual features. The core lies in comparing the similarity between the output results of the previous layer and the current output results, calculating the weight factors, and finally generating the self-attention coefficients, as detailed below.

Firstly, a linear transformation is performed. The multi-head attention mechanism obtains the query matrix  $\mathbf{Q}$ , key matrix  $\mathbf{K}$ , and value matrix  $\mathbf{V}$  through linear transformations with different weight matrices, as shown in the following equations:

$$\begin{cases} \mathbf{Q} = \mathbf{W}_q \mathbf{x} \\ \mathbf{K} = \mathbf{W}_k \mathbf{x} \\ \mathbf{V} = \mathbf{W}_v \mathbf{x} \end{cases} \quad (20)$$

where  $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v \in R^{m \times n}$  are learnable weight matrices, with  $m$  and  $n$  being the dimensions of the input and output vectors of the attention mechanism, respectively, and  $\mathbf{x}$  being the input matrix.

Further, the attention scores for the input vectors are calculated and normalized, then input into the Softmax function for activation, and multiplied by the value matrix  $\mathbf{V}$  to obtain the weighted calculation result of self-attention, as shown in the following equation:

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Softmax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (21)$$

where  $d_k$  is the scaling factor, used to adjust the sensitivity of the attention mechanism. Through learnable weight matrices  $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v$ , the outputs of BiTCN are mapped to different subspaces, and each head independently calculates attention scores in its corresponding subspace, capturing distinct key feature dimensions in sequence data. That is,  $\{\mathbf{Q}_i\}_{i=1}^h, \{\mathbf{K}_i\}_{i=1}^h, \{\mathbf{V}_i\}_{i=1}^h$ , where  $i \in [1, h]$ . After iteration, the attention weights for the corresponding number of heads  $\mathbf{Head}_i$  are as shown in the following equation:

$$\mathbf{Head}_i = Attention(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (22)$$

where  $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v$  represent the parameter mapping matrices for  $\mathbf{Q}$ ,  $\mathbf{K}$ , and  $\mathbf{V}$  for the  $i$ th attention head, respectively. Finally, through a concatenation operation to form a comprehensive attention weight vector, and processed by the output transformation matrix  $\mathbf{W}_o$ , the final output is obtained as follows:

$$MHA(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Concat(\mathbf{Head}_1, \dots, \mathbf{Head}_h)\mathbf{W}_o \quad (23)$$

where Concat represents the concatenation operation of the output vectors, and  $h$  is the number of attention heads. This study utilizes the multi-head attention mechanism to apply weighted processing to the nodes output by BiTCN, enhancing the model's sensitivity to key sequence information points.

#### 4.3. Model Process

The specific steps of the RIME-MVMD-P-BiTCN-MHA prediction model proposed in this paper are as follows:

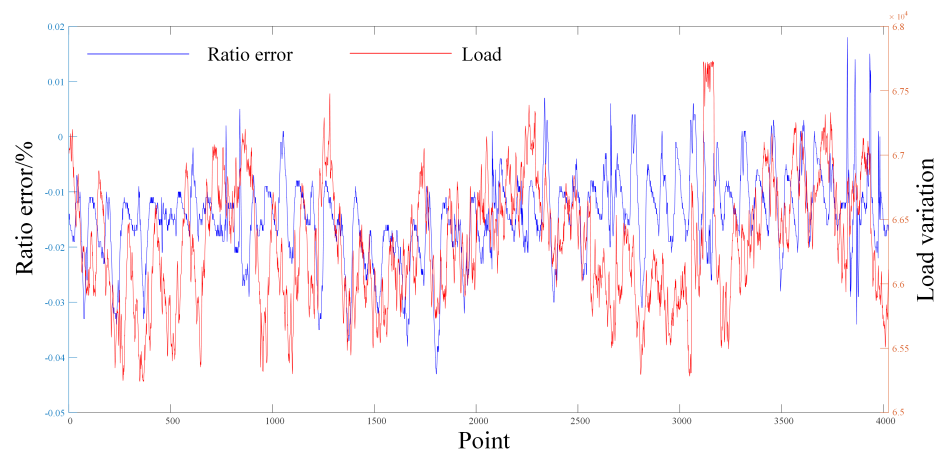
- **Data Collection and Preprocessing:** Collect historical ratio error sequences and load fluctuation data from VTs. The input step size is 24. After removing invalid data points and those beyond three standard deviations, the data are normalized. The final results are then denormalized.



- *Adaptive MVMD Decomposition*: Set the minimum permutation entropy as the objective function, use the RIME optimization algorithm to solve for the optimal decomposition parameters  $(K, \alpha)$  of MVMD, and decompose the collected load and ratio error sequences.
- *Feature Selection and Reconstruction*: Utilize the Spearman rank correlation coefficient to filter the decomposed subcomponents, eliminating irrelevant components.
- *Sequence Reconstruction*: Reconstruct the decomposed sequences and establish predictive sub-models for the reconstructed sequences.
- *BiTCN-MHA Model Training*: Predict each input signal component through BiTCN to extract deep features and generate a multidimensional feature matrix. After passing through a flattening layer to the MHA unit for marking important information, output the predicted subcomponents. The final prediction result is obtained by accumulation.

## 5. Case Study Analysis and Discussion

This case study was conducted in a Matlab 2023b environment, with data sourced from an operational 110 kV substation in Henan, China. Standard transformers connected online facilitated the real-time data collection via a 24-bit acquisition card for 0.2 class current transformers in operation. The data, transmitted through a merging unit to the verification platform, include secondary outputs from the transformer and standard output data. The device's sampling frequency is set at 10 min per sample, recording the intended output signals. After removing data points invalid data points and those beyond three standard deviations, a total of 4025 data values were recorded for September. Each dataset is divided into two parts: a training dataset and a testing dataset, with a ratio of 0.8:0.2, the last six days designated as the test set. The distribution of the original ratio error and load sample points is depicted in Figure 3.



**Figure 3.** Original ratio error and load sample points (left Y-axis for difference ratio error, right Y-axis for load).

### 5.1. Model Evaluation Metrics

To evaluate the performance of the model, three statistical metrics are employed: Mean Absolute Error (MAE), Root-Mean-Squared Error (RMSE), and Median Absolute Error (MedAE). The mathematical expressions for these metrics are given below:

$$E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (24)$$

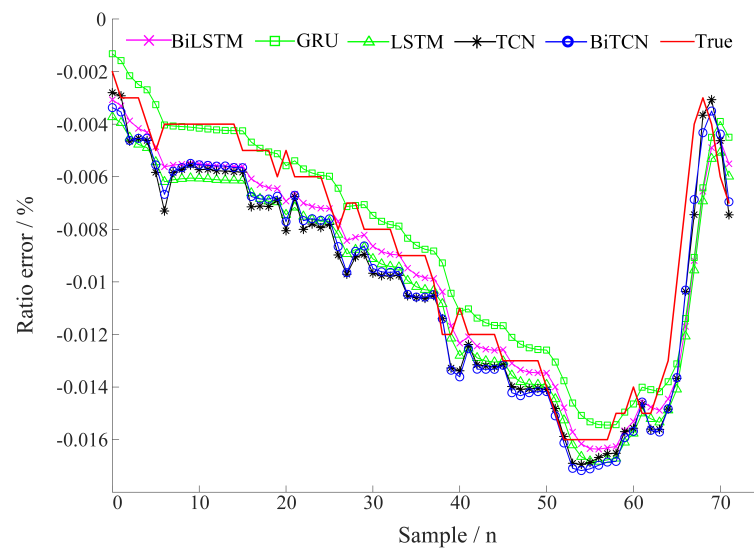
$$E_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (25)$$

$$E_{\text{MedAE}} = \text{median}(|\hat{y}_1 - y_1|, \dots, |\hat{y}_n - y_n|) \quad (26)$$

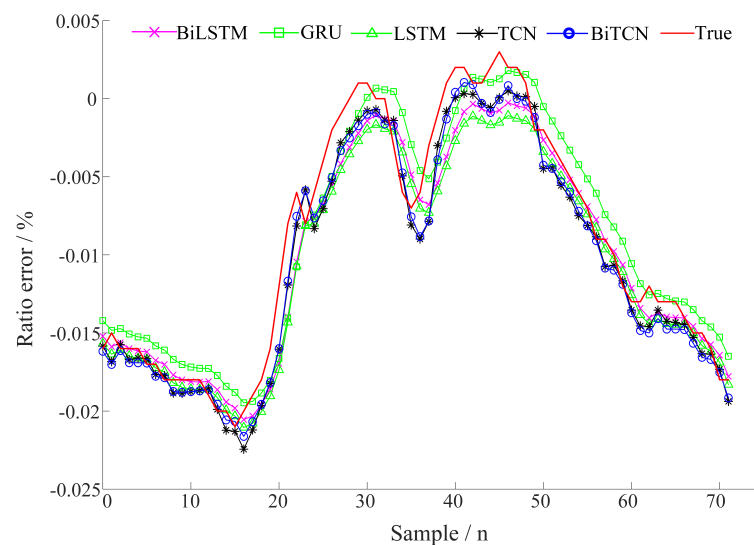
where  $y_i$  denotes the actual ratio error values,  $\hat{y}_i$  the predicted ratio error values, and  $n$  the total number of samples. Lower values of  $E_{MAE}$ ,  $E_{RMSE}$ , and  $E_{MedAE}$  signify minor deviations between the predicted and actual values, indicating higher accuracy of the model.

### 5.2. Model Evaluation Metrics

To verify model improvements through backtesting, original data are utilized as input, against which six benchmark models are compared: GRU [27], LSTM [28], BiLSTM [29], TCN [30], and the enhanced TCN, namely the bidirectional BiTCN model. For consistent lateral comparison, a uniform learning rate of 0.0025 is set, along with a batch size of 256 and a step size of 24. The L2 regularization parameter for all models is set to  $5 \times 10^{-4}$ . The hidden layers for GRU, LSTM, and BiLSTM models are configured to 100, while both TCN and BiTCN have four layers, with a dropout rate of 0.1. Comparative assessments of prediction performance are illustrated in Table 1. Randomly select two consecutive 72-points for display in Figures 4 and 5.



**Figure 4.** The evolution of the optimal solution over iterations: randomly select 72 consecutive points (12 h).



**Figure 5.** The evolution of the optimal solution over iterations: randomly select 72 consecutive points (12 h).

**Table 1.** Comparison of performance of five basic algorithms.

Models	$E_{RMSE}$	$E_{MAE}$	$E_{MedAE}$
GRU	$2.987 \times 10^{-3}$	$1.727 \times 10^{-3}$	$1.023 \times 10^{-3}$
LSTM	$2.988 \times 10^{-3}$	$1.807 \times 10^{-3}$	$1.054 \times 10^{-3}$
BiLSTM	$2.837 \times 10^{-3}$	$1.576 \times 10^{-3}$	$0.811 \times 10^{-3}$
TCN	$2.730 \times 10^{-3}$	$1.685 \times 10^{-3}$	$1.188 \times 10^{-3}$
BiTCN	$2.670 \times 10^{-3}$	$1.562 \times 10^{-3}$	$1.143 \times 10^{-3}$

The table and figures above demonstrate the performance differences among various models on this dataset. The BiTCN and BiLSTM, incorporating the capabilities of bidirectional networks, outperform their unidirectional equivalents, TCN and LSTM, respectively. Among them, BiLSTM shows the best performance on the MedAE metric with a score of  $0.811 \times 10^{-3}$ , indicating effective control over the median error. In contrast, BiTCN performs best in terms of MAE and RMSE, with scores of  $1.562 \times 10^{-3}$  and  $2.670 \times 10^{-3}$ , respectively. This suggests that BiTCN provides high accuracy in handling most data points but struggles to control the median prediction error in certain specific circumstances.

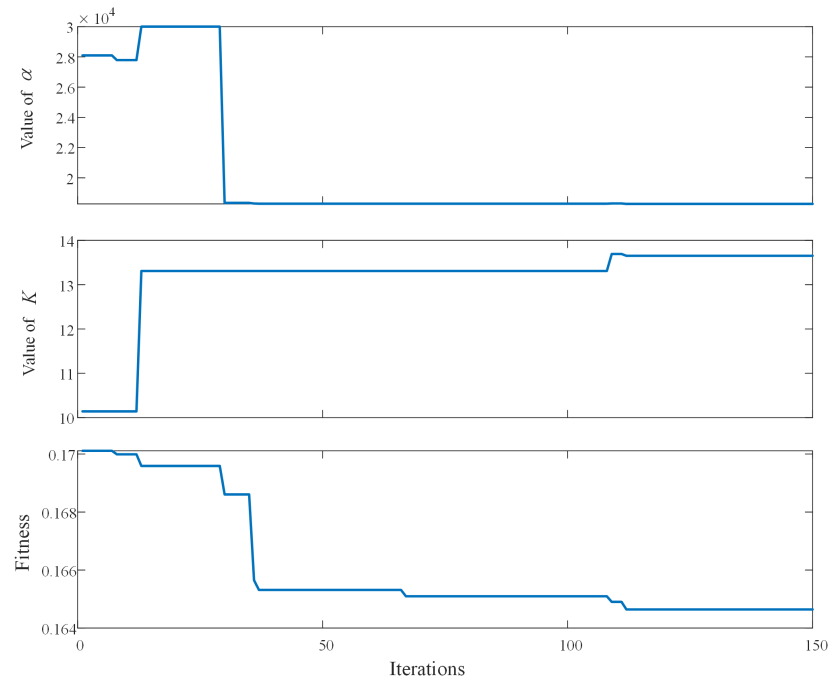
The reason lies in the design of the BiTCN model which, although effective at capturing long-term dependencies and cyclical patterns in time series, generates a skewed error distribution for certain atypical outlier points within this dataset, thereby resulting in substantial median errors. Although these errors are significant, their low frequency minimizes their impact on the RMSE. Additionally, the model's optimization strategy primarily focuses on reducing overall errors, which can sometimes compromise the predictive accuracy at critical points. Therefore, it is necessary to further optimize the BiTCN model to enhance its adaptability to data and improve its capability to handle outlier points.

### 5.3. MVMD Decomposition and Reconstruction

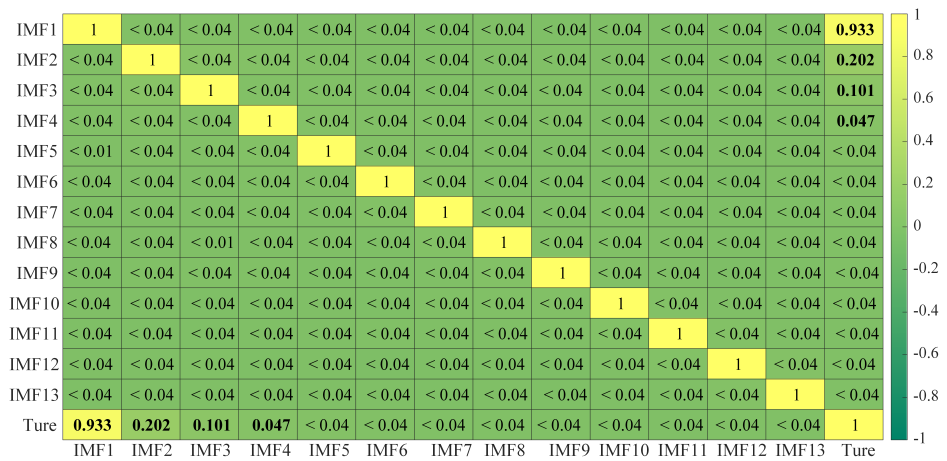
In MVMD decomposition, the penalty function and the number of modal components are formulated into a two-dimensional optimization problem. This problem is iteratively addressed through an optimization algorithm as depicted in Figure 6 to enhance the decomposition quality. By meticulously managing the decomposition details, common issues such as limited interpretability with empirical methods, incomplete decomposition, and inadvertent noise introduction are mitigated.

As demonstrated in the figure above, adaptive decomposition effectively segments the original ratio error and load sequences into distinct subcomponents. The parameters  $(K, \alpha)$  are configured to round towards zero, where the optimal decomposition values are (13, 18,271). The decomposed IMF components are then screened using the Spearman rank correlation coefficient. Only those subcomponents that exhibit strong correlations with the original sequence are retained. The results of this correlation analysis are presented in Figure 7.

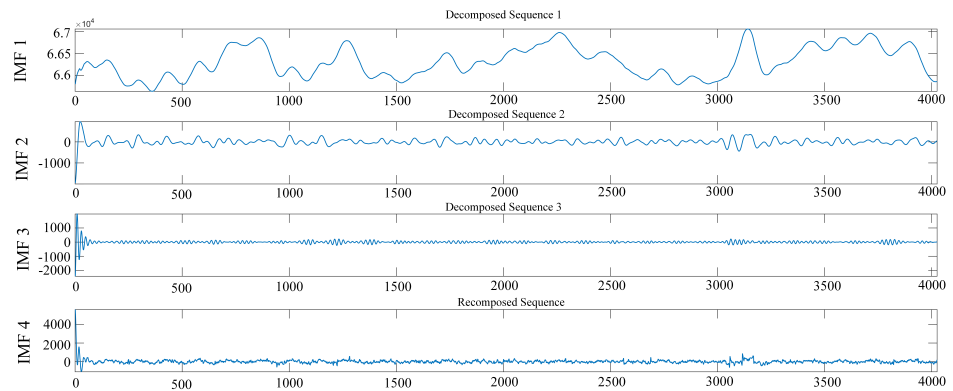
Based on the correlation coefficients calculated in Figure 7, a threshold value  $Q$  of 0.1 is set. Subcomponents with correlation coefficients greater than  $Q$  are retained as relevant components for building the prediction model. Subcomponents with values less than  $Q$ , IMF3 to IMF13, are combined to form a high-frequency component. The residual components from the decomposition are also added to this combination, resulting in a new reconstructed component referred to as IMF4. The reconstructed load and ratio error components are shown in Figures 8 and 9.



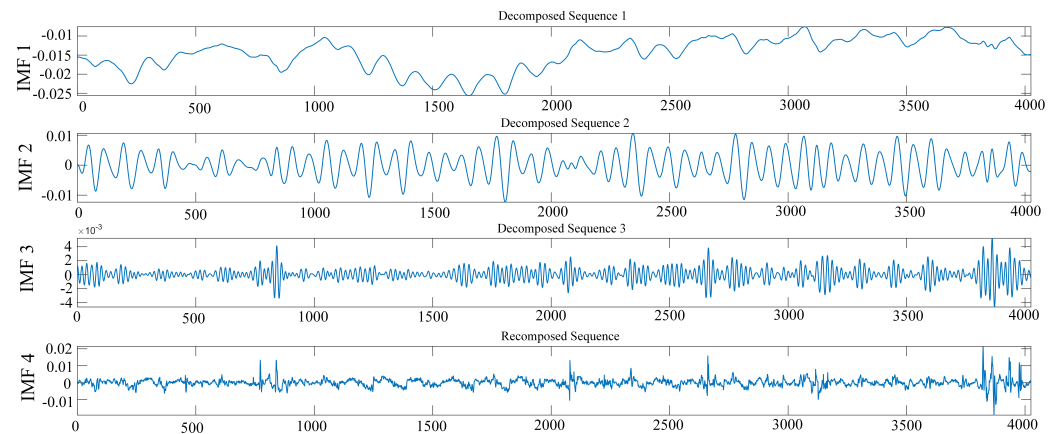
**Figure 6.** The evolution of the optimal solution over iterations, reflecting the optimization process of parameters  $K$  and  $\alpha$ , and corresponding changes in the objective function.



**Figure 7.** Spearman rank correlation coefficient analysis. To make the chart clearer, we have replaced the values within the range  $[-0.04, 0.04]$  with  $<0.04$  and bolded the values greater than 0.04.



**Figure 8.** Reconstructed load components.



**Figure 9.** Reconstructed ratio error components.

Figure 8 shows the reconstructed load component, and Figure 9 displays the reconstructed ratio error component. As seen in the figures, after the improved dual-channel decomposition of MVMD, decomposing the original sequence into simpler components with different frequencies and amplitudes, each of the IMF1, IMF2, and IMF3 exhibits distinct trends. It is also observable from the figures that IMF1 has the most significant correlation with the original sequence, as shown in Figure 3. Compared to the original sequence, IMF1 is more stable and orderly in trend, demonstrating the effectiveness of the MVMD decomposition and validating the correctness of the Spearman rank screening method. Decomposing non-stationary sequences into simpler trend series and replacing the original sequences as model inputs reduce the complexity of model training while increasing the training data, thereby improving the model's overall generalization and performance.

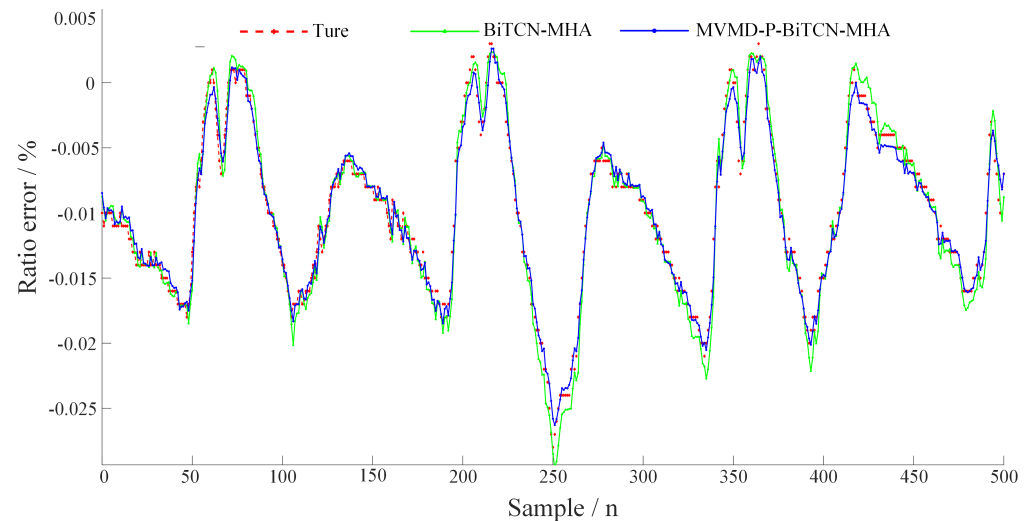
After the dual-channel decomposition using MVMD, each of the IMF1, IMF2, and IMF3 exhibits distinct trend components. It is also observable from the figures that IMF1 has the most significant correlation with the original series shown in Figure 3. Compared to the original series, IMF1 is more stable and orderly in trend, which demonstrates the effectiveness of the MVMD decomposition and validates the correctness of the Spearman rank screening method. Decomposing non-stationary sequences into simpler trend series can reduce the complexity of model training and enhance the generalization performance of the model, thereby improving the learning capabilities for complex models.

#### 5.4. Model Ablation Study

To further analyze the effectiveness of the mixed deep learning model and quantify the impact of the MVMD framework on the model proposed in this paper, this section conducts detailed comparative experiments. The experiments compare three different model configurations: the standard BiTCN model, the BiTCN integrated with Multi-Head Attention mechanism (abbreviated as BiTCN-MHA), and the BiTCN-MHA using components decomposed of MVMD and screened by Spearman rank as inputs (abbreviated as MVMD-BiTCN-MHA). The specific results of these comparisons are detailed in Figure 10 and Table 2.

The results in Table 2 indicate that the method proposed in this study outperforms the standard BiTCN and the enhanced BiTCN-MHA model in various evaluation metrics. Specifically, improvements of 17.9%, 23.2%, and 43.1% in RMSE, MAE, and MedAE were noted compared to the standard BiTCN, and 15.9%, 17.1%, and 23.9% improvements compared to the BiTCN-MHA. This method not only shows superior performance in overall error management compared to BiTCN and BiTCN-MHA but also demonstrates higher robustness and effectiveness in managing median errors. The value of  $0.6507 \times 10^{-3}$  is significantly lower than  $1.143 \times 10^{-3}$  for BiTCN and  $0.856 \times 10^{-3}$  for BiTCN-MHA, clearly improving on the limitations of previous BiTCN models. This indicates that the proposed

method not only minimizes overall error management but also has significant advantages in handling outliers and extreme data points. From Figure 10, it is also observed that MVMD-BiTCN-MHA not only shows more precise trend following in tracking real values but also exhibits better adaptability and prediction accuracy when dealing with extreme data points.



**Figure 10.** Ablation study analysis.

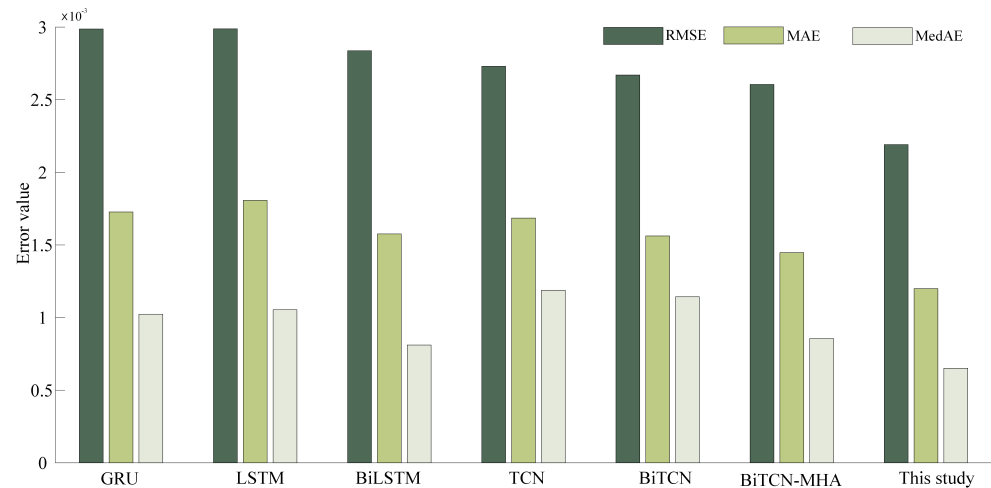
**Table 2.** Comparison of the evaluation effect of ablation study.

Models	$E_{RMSE}$	$E_{MAE}$	$E_{MedAE}$
BiTCN	$2.670 \times 10^{-3}$	$1.562 \times 10^{-3}$	$1.143 \times 10^{-3}$
BiTCN-MHA	$2.605 \times 10^{-3}$	$1.447 \times 10^{-3}$	$0.856 \times 10^{-3}$
This study	$2.191 \times 10^{-3}$	$1.199 \times 10^{-3}$	$0.651 \times 10^{-3}$

These results validate the optimal utility of MVMD processing, Spearman rank screening, and reconstruction, along with the multi-head self-attention mechanism in enhancing deep learning models for handling complex time series data, particularly in improving prediction accuracy and responsiveness to extreme changes.

To visually compare the performance of different models, data from Tables 1 and 2 are consolidated and presented in a bar chart, as shown in Figure 11. The scatter plot of absolute errors per point for each model is shown in Figure A1. The method introduced in this study successfully addresses the challenge of applying a uniform decomposition standard to heterogeneous data by incorporating the optimal use of MVMD. The precise decomposition via MVMD and subsequent reconstruction using the Spearman rank effectively mitigates the negative impact of the original complex non-stationary sequences on model performance.

Additionally, the strategy of integrating a hybrid model framework further enhances the prediction accuracy, enabling the proposed method to effectively predict the ratio error of transformers and also confirming the inferences made during the theoretical research phase.



**Figure 11.** Comparison of ratio error assessment effectiveness.

## 6. Conclusions

This paper considers the impact of load fluctuations on voltage transformers and proposes a voltage transformer measurement error prediction model that integrates multi-variate variational mode decomposition (MVMD) with a hybrid temporal convolutional network (TCN) to effectively predict the ratio error in transformers. The main conclusions drawn from this study can be summarized as follows:

- The enhanced MVMD algorithm improves the precision and interpretive power of dual-channel signal decomposition and utilizes the Spearman rank correlation coefficient to select dominant modes after decomposition, reconstructing the signal channels. This advances the model's generalization capabilities and its ability to learn from complex sequences.
- By integrating a bidirectional temporal convolutional network and a multi-head attention mechanism, the model takes into account both the correlation of the input load and the dynamic changes of the time series, enhancing its predictive stability for future trends.

In summary, the proposed MVMD-PE-P-TCN-MHA model provides a new method for assessing and predicting measurement errors in voltage transformers under load changes, contributing to the safe operation of power systems and the fairness of energy transactions. The proposed model is expected to play a vital role in improving the reliability of future grids in which dynamic loads of uncertain location and operational mode such as EVs will be significant.

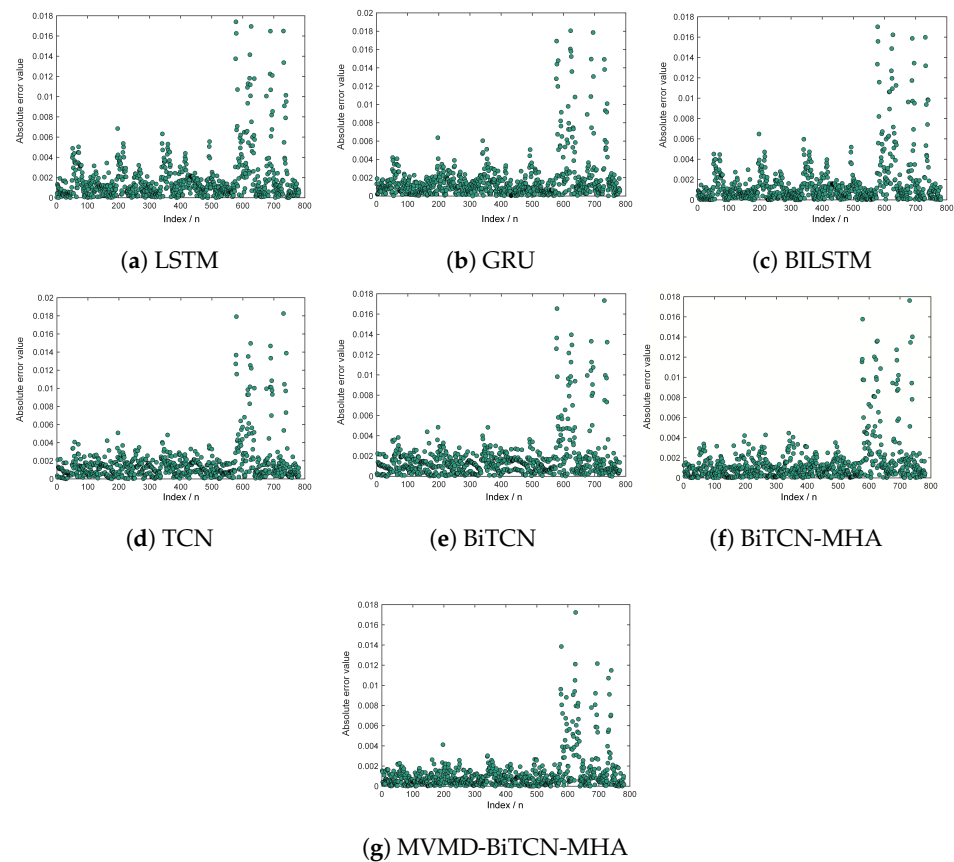
**Author Contributions:** Conceptualization, P.R.F.R. and A.A.-S.; data curation, L.Q.; writing—original draft preparation, J.C.; writing—review and editing, Z.L. project administration, H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under Grant 52311530337, in part by the China Scholarship Council, and in part by the Interdisciplinary program of Wuhan National High Magnetic Field Center, Huazhong University of Science and Technology under Grant WHMFC202202.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A



**Figure A1.** Scatter plot of absolute errors per point: (a) Scatter plot of absolute errors per point (LSTM). (b) Scatter plot of absolute errors per point (GRU). (c) Scatter plot of absolute errors per point (BiLSTM). (d) Scatter plot of absolute errors per point (TCN). (e) Scatter plot of absolute errors per point (BiTCN). (f) Scatter plot of absolute errors per point (BiTCN-MHA). (g) Scatter plot of absolute errors per point (MVMD-BiTCN-MHA).

## References

- Long, Z.; Li, W.; Zhou, F.; Liu, S.; Fan, J.; Hu, K. Device of 1200 kV Wideband Capacitive Divider Based on High-voltage Standard Capacitor. *High Volt. Eng.* **2022**, *48*, 1826–1835. [[CrossRef](#)]
- Bibra, E.M.; Connelly, E.; Gerner, M.; Lowans, C.; Paoli, L.; Tattini, J.; Teter, J. *Global EV Outlook 2021: Accelerating Ambitions Despite the Pandemic*; National Academy of Sciences: Washington, DC, USA, 2021.
- General Metrological Terms and Definitions*; Chinese Quality Inspection Press: Beijing, China, 2012.
- Cusenza, M.A.; Bobba, S.; Ardente, F.; Cellura, M.; Di Persio, F. Energy and environmental assessment of a traction lithium-ion battery pack for plug-in hybrid electric vehicles. *J. Clean. Prod.* **2019**, *215*, 634–649. [[CrossRef](#)]
- Jones, C.B.; Lave, M.; Vining, W.; Garcia, B.M. Uncontrolled Electric Vehicle Charging Impacts on Distribution Electric Power Systems with Primarily Residential, Commercial or Industrial Loads. *Energies* **2021**, *14*, 1688. [[CrossRef](#)]
- Chadha, S.; Jain, V.; Singh, H.R. A review on Smart Charging impacts of Electric Vehicles on Grid. *Mater. Today Proc.* **2022**, *63*, 751–755. [[CrossRef](#)]
- Crozier, C.; Deakin, M.; Morstyn, T.; McCulloch, M. Coordinated electric vehicle charging to reduce losses without network impedances. *IET Smart Grid* **2020**, *3*, 677–685. [[CrossRef](#)]
- Rodríguez-Pajarón, P.; Hernández, A.; Milanović, J.V. Probabilistic assessment of the impact of electric vehicles and nonlinear loads on power quality in residential networks. *Int. J. Electr. Power Energy Syst.* **2021**, *129*, 106807. [[CrossRef](#)]
- Fu, Y.; Meng, X.; Su, X.; Mi, Y.; Tian, S. Coordinated charging control of PEV considering inverter's reactive power support and three phase switching in unbalanced active distribution networks. *Electr. Power Autom. Equip.* **2020**, *40*, 1–7.
- Baraniak, J.; Starzyński, J. Modeling the Impact of Electric Vehicle Charging Systems on Electric Power Quality. *Energies* **2020**, *13*, 3951. [[CrossRef](#)]



11. Hafeez, G.; Alimgeer, K.S.; Khan, I. Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. *Appl. Energy* **2020**, *269*, 114915. [[CrossRef](#)]
12. Zhang, C.; Ma, H.; Hua, L.; Sun, W.; Nazir, M.S.; Peng, T. An evolutionary deep learning model based on TVFEMD, improved sine cosine algorithm, CNN and BiLSTM for wind speed prediction. *Energy* **2022**, *254*, 124250. [[CrossRef](#)]
13. Zhou, F.; Zhao, P.; Lei, M.; Yue, C.; Yu, J.; Liang, S. Capacitive voltage transformer measurement error prediction by improved long short-term memory neural network. *Energy Rep.* **2022**, *8*, 1011–1021. [[CrossRef](#)]
14. Zhang, W.; Shi, Y.; Yu, J.; Yang, B.; Lin, C. Online measurement of capacitor voltage transformer metering errors based on GRU and MTL. *Electr. Power Syst. Res.* **2023**, *221*, 109473. [[CrossRef](#)]
15. Yang, X.; Li, Z.; Zhong, Y.; Li, H. Ultra-short term transformer error forecast based on variational mode decomposition and CNN-GRU-ED. *Dianli Xitong Baohu Yu Kongzhi/Power Syst. Prot. Control* **2023**, *51*, 68–77.
16. Brigham, E.O. *The Fast Fourier Transform and Its Applications*; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1988.
17. Mounir, N.; Ouadi, H.; Jrhilifa, I. Short-term electric load forecasting using an EMD-BI-LSTM approach for smart grid energy management system. *Energy Build.* **2023**, *288*, 113022. [[CrossRef](#)]
18. Li, D.; Jiang, M.-R.; Li, M.-W.; Hong, W.-C.; Xu, R.-Z. A floating offshore platform motion forecasting approach based on EEMD hybrid ConvLSTM and chaotic quantum ALO. *Appl. Soft Comput.* **2023**, *144*, 110487. [[CrossRef](#)]
19. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [[CrossRef](#)]
20. Xue, T.; Jiang, F.; Zhang, L.; Xu, X. Strategy of Power Allocation and Two-layer Energy Management in Hybrid Energy Storage. *J. China Three Gorges Univ. (Nat. Sci.)* **2023**, *45*, 80–87.
21. ur Rehman, N.; Aftab, H. Multivariate variational mode decomposition. *IEEE Trans. Signal Process.* **2019**, *67*, 6039–6052. [[CrossRef](#)]
22. Li, Z.; Lan, F.; Zhong, Y.; Qiu, L.; Cheng, L. Measurement-protection-integrated Current Sensor Based on Double-bobbin Co-winding Technology. *High Volt. Eng.* **2022**, *48*, 4427–4429.
23. Transformers—Part, I. In 5: *Additional Requirements for Capacitor Voltage Transformers*; Standard IEC: Geneva, Switzerland, 2011.
24. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [[CrossRef](#)]
25. Su, H.; Zhao, D.; Heidari, A.A.; Liu, L.; Zhang, X.; Mafarja, M.; Chen, H. RIME: A physics-based optimization. *Neurocomputing* **2023**, *532*, 183–214. [[CrossRef](#)]
26. Stephanou, M.; Varughese, M. Sequential estimation of Spearman rank correlation using Hermite series estimators. *J. Multivar. Anal.* **2021**, *186*, 104783. [[CrossRef](#)]
27. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
28. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
29. Peng, T.; Zhang, C.; Zhou, J.; Nazir, M.S. An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting. *Energy* **2021**, *221*, 119887. [[CrossRef](#)]
30. Zhu, J.; Su, L.; Li, Y. Wind power forecasting based on new hybrid model with TCN residual modification. *Energy AI* **2022**, *10*, 100199. [[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.