

Article

A Grounding Current Prediction Method Based on Frequency-Enhanced Transformer

Na Zhang ¹, Gang Yang ^{1,*}, Zilong Fu ² and Junsheng Hou ²

¹ Electric Power Research Institute, State Grid Shanxi Electric Power Company, Taiyuan 030001, China; 19834502446@163.com

² Substation Operation and Maintenance Center, State Grid Shanxi Electric Power Company Jincheng Power Supply Company, Jincheng 048000, China; longsir523@163.com (Z.F.); hjs666888@163.com (J.H.)

* Correspondence: xxyfei1231@163.com

Abstract: Concerning the problem that the coupling relationship in substation scenarios is complex and the Transformer model makes it difficult to capture the correlation between multiple variables of grounding current, resulting in low accuracy of grounding current prediction, a ground current prediction method based on frequency-enhanced Transformer is proposed. Firstly, in the data preprocessing stage, the best frequency domain decomposition algorithm is designed to obtain the high-frequency and low-frequency component data containing different component features so as to enhance the initial features that the model focuses on. Secondly, the data slicing and embedding module is designed to replace the original embedding module of the Transformer to realize the enhanced extraction of local features of the data. Finally, in the feature extraction stage, an enhanced attention mechanism is introduced to replace the standard attention mechanism to capture the intrinsic features of the sequence time dimension and the variable dimension in parallel so as to improve the extraction ability of Transformer multivariate features. Experimental results on the self-built grounding current dataset and the public dataset show that the proposed method outperforms existing advanced methods, verifying the effectiveness of the proposed method.

Keywords: time series; attention mechanisms; grounding current prediction; transformer



Academic Editor: Junfeng Liu

Received: 4 November 2024

Revised: 9 December 2024

Accepted: 23 December 2024

Published: 25 December 2024

Citation: Zhang, N.; Yang, G.; Fu, Z.; Hou, J. A Grounding Current Prediction Method Based on Frequency-Enhanced Transformer. *Energies* **2025**, *18*, 32. <https://doi.org/10.3390/en18010032>

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

As important data in the quality of electrical energy, the ground current at the grounding site of the substation can reflect the operating status of the power station equipment in real-time [1]. Mastering its dynamic change law [2] is essential for the safe and stable operation of the substation and the fault prevention of the equipment [3]. The grounding current predictive task refers to the historical ground current data collected by the different location sensors. It is a prediction of the ground current data of the ground locations by establishing a predictive model to find its potential laws [4].

Grounding current data generally exhibit a strong periodicity, with brief disturbance fluctuations, and is a form of time-series information [2]. Therefore, this prediction task can be viewed as the application of time-series forecasting in the power industry. In actual forecasting work, the accurate prediction of grounding current faces two main challenges: (1) Grounding current is influenced by external factors such as temperature, region, and electromagnetic interference [5], showing high-frequency and nonlinear characteristics. This not only increases the difficulty of feature extraction for the model but also affects

prediction accuracy. (2) Due to the interconnection characteristics of the grounding network [6], there exists a dynamic correlation between the grounding currents at different grounding points, which further complicates the prediction task.

In response to this task, Early traditional methods, such as manual analysis and prediction [3] as well as machine learning methods [4], face challenges in handling the high-frequency nonlinear features and multivariate dynamic correlations in large and complex datasets when tasked with predicting grounding currents. This is primarily due to their heavy reliance on human expertise. In contrast, deep learning models can deeply explore the underlying relationships within the data, effectively addressing this issue. Currently, in the power scenario, mainstream deep learning methods for time series prediction can be roughly divided into two categories: one is based on recurrent neural network (RNN) architectures, such as RNNs, LSTMs, and their variants; the other is based on encoder-decoder architectures like Transformer models and their variants [7]. Although methods based on RNNs and their variants [8–10] can effectively extract potential patterns in the time dimension of the data during time series prediction, they face issues like gradient explosion and vanishing when dealing with long input–output sequences, such as ground current prediction tasks. Transformer models, however, can address these issues more effectively. This paper will focus on predicting ground current based on Transformer models.

The Transformer series models effectively capture the dependencies between time series through their attention mechanism, making them widely used in various time prediction tasks. Mainstream neural network models based on the attention mechanism, such as DSFormer [11], Informer [12], Autoformer [13,14], Reformer [15], Fedformer [16], and TimesNet [17], primarily focus on modeling temporal dependencies in data. However, they lack analysis of the underlying interdependencies between data points, making them unsuitable for scenarios like ground current prediction, where nonlinear dependencies need to be considered. In recent years, with the development of Graph Convolutional Networks (GCNs), some researchers have utilized graph structures to capture spatial features between variables, offering new solutions to the aforementioned problems. References [18,19] addressed the spatial relationships between samples and features by constructing graph structures, capturing interdependencies between features to some extent. However, the graphs used in these methods are usually static and based on prior human knowledge, while the spatial dependencies between nodes are dynamic. Therefore, these methods have significant limitations in tasks like ground current prediction. References [20–22] have dynamically captured both temporal and spatial features of each data time-stamp using spatial attention mechanisms, greatly improving prediction accuracy in related scenarios. However, they fail to fully consider the contextual semantic features between adjacent time-stamped variables. Moreover, the combination of these attention mechanisms with LSTM networks also limits their application in long time series prediction tasks. Therefore, effectively learning the dependencies of each time-stamp variable in long time series prediction remains a challenging task.

Moreover, the aforementioned single prediction models are unable to fit subsequences with varying volatility, which affects their predictive performance. Decomposing time series into subsequences with different characteristics for separate modeling has become a common approach. References [23,24] utilized empirical mode decomposition (EMD) and variational mode decomposition (VMD), respectively, to decompose time series data and then combine these with neural networks for prediction. This approach is more suitable for forecasting tasks involving complex feature time series than other methods. However, these decomposition techniques may encounter modal mixing issues during data processing, which can affect the results of the component data.

To address the aforementioned issues, this paper proposes a frequency-enhanced transformer for ground current prediction (Fdeformer) aimed at improving the model's ability to extract features from multivariable time series data. In response to the high-frequency characteristics of ground current, a best frequency domain decomposition algorithm (BFDD) is introduced to obtain the high and low-frequency component data of ground current. Compared to traditional time series decomposition methods, this module can more accurately decompose the ground current time series data, reducing the interference of modal stacking issues on the component data. To tackle the existing challenge where attention mechanisms struggle to adequately model the dependencies between variables, we first introduce a data slicing and embedding module (DSAE) to perform regional partitioning of the time series data. This results in time series segments that contain more contextual information. Then, enhanced attention is used to simultaneously capture the relationships between the time dimension and variable dimension of these segments, thereby improving the model's ability to extract local features and contextual information from the time series data. The proposed model is validated across different datasets, with results indicating that its predictive accuracy surpasses that of several existing Transformer-based baseline models across various evaluation metrics.

2. Research Methodology and Fdeformer

2.1. Transformer and Its Problems

The Transformer model is a deep learning model based on the self-attention mechanism, primarily used for sequence-to-sequence tasks. This model abandons the frameworks commonly used in traditional sequence-to-sequence tasks, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), and instead relies entirely on self-attention mechanisms and positional encoding strategies to analyze and model long-term complex dependencies in the input sequences. The model is mainly composed of stacked encoders and decoders, and its architecture is illustrated in Figure 1.

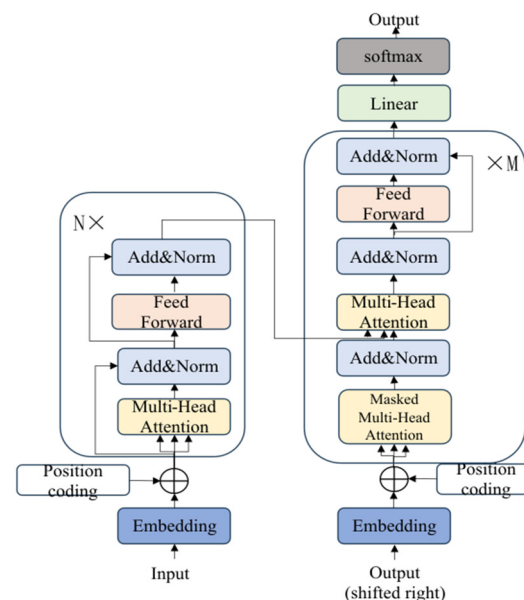


Figure 1. Transformer model.

2.1.1. Encoder Layer

The encoder is composed of N identical sub-layers stacked together, each containing two modules: one is a multi-head attention layer, which captures long-term dependencies in the temporal dimension of the sequence, and the other is a feedforward layer, which applies

a nonlinear transformation to each position. The encoder can capture global dependency information from the input sequence for subsequent decoder calculations. Additionally, the self-attention mechanism allows the model to process all positions in the sequence simultaneously, significantly improving the model's efficiency.

2.1.2. Decoder Layer

The decoder is composed of M identical sub-layers stacked together, each containing three modules. The masked multi-head attention is similar to the encoder's attention but employs a masking mechanism to ensure that when generating each target token, it can only attend to all positions before the current one. The cross multi-head attention is primarily used to aggregate the global dependency features captured during the encoder phase and the features learned through masked attention. Finally, the output from the encoder is processed through a feedforward layer.

However, the Transformer series models face some issues during prediction. For example, the model directly fits the dependencies between variables using a single linear layer, which is then added to the embedded information in the temporal domain to serve as the input for the self-attention module. This approach makes it difficult for the model to account for the semantic relationships among multiple variables, affecting the allocation of weights within the self-attention module and resulting in degraded performance when applying Transformers to multivariable datasets such as grounding current predictions.

2.2. Fdeformer Model

To address the issues present in traditional Transformer models, this paper proposes the Fdeformer model. This model builds upon the Transformer framework by improving the original embedding module, thereby enhancing the model's ability to extract local features. Additionally, this paper introduces an enhanced attention mechanism to replace the traditional point attention mechanism, improving the model's expressive power in time series analysis and multivariable feature extraction. Furthermore, to strengthen the model's ability to capture the original characteristics of the data, a best-frequency domain decomposition module is introduced, which decomposes the input data into high-frequency and low-frequency components that reflect different features. Its specific structure is shown in Figure 2.

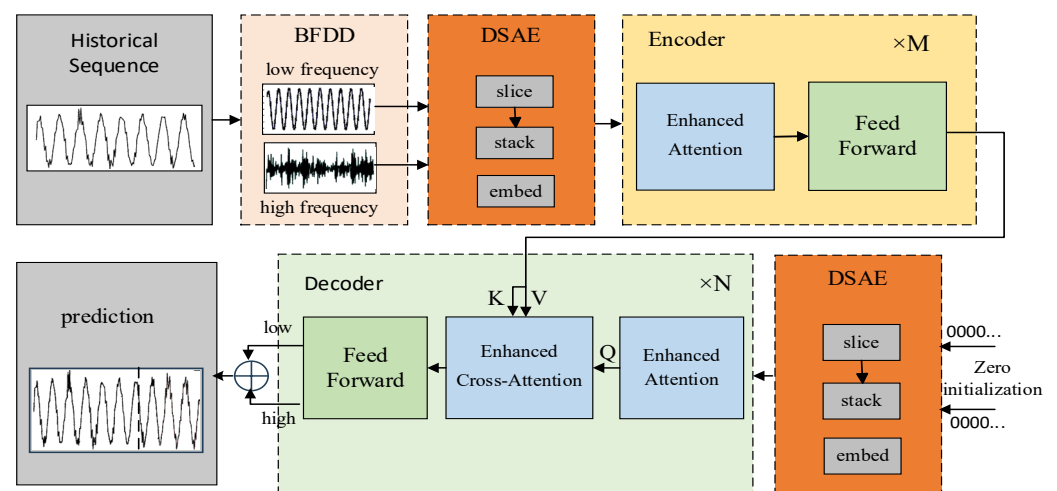


Figure 2. Model structure of Fdeformer.

First, the input grounding current data are passed through a best-frequency domain decomposition module to obtain high- and low-frequency component data that reflect the

different characteristics of the original data. Then, in the slicing and embedding module, the high- and low-frequency component data are sliced, stacked, and embedded to obtain embedding vectors of subsequences for different time segments. Afterward, the high- and low-frequency embedding vectors are fed into M serial encoders and N serial decoders for iterative training to obtain the prediction results for the high- and low-frequency components. Finally, the results are combined to obtain the final prediction.

The encoder parallelly captures the relationships between subsequence blocks in both the time and variable dimensions through an enhanced attention module, generating a key matrix K and value matrix V that contain the source sequence information. The decoder's input consists of zero-initialized vectors that have been enhanced through the data slicing and embedding module. The enhanced attention is then used to extract the query matrix Q for the target sequence, and cross-attention is applied to aggregate the results from the encoder. Finally, the prediction results are generated through a forward propagation mapping process.

2.2.1. Best Frequency Domain Decomposition (BFDD)

Through Fast Fourier Transform (FFT), the grounding current data can be decomposed into low-frequency components and high-frequency components, as shown in Figure 3. It can be observed that the low-frequency component data of the grounding current is relatively stable and regular, effectively representing the periodic and trend components of the historical data, whereas the high-frequency component data are significantly affected by noise, representing the random components within the historical data.

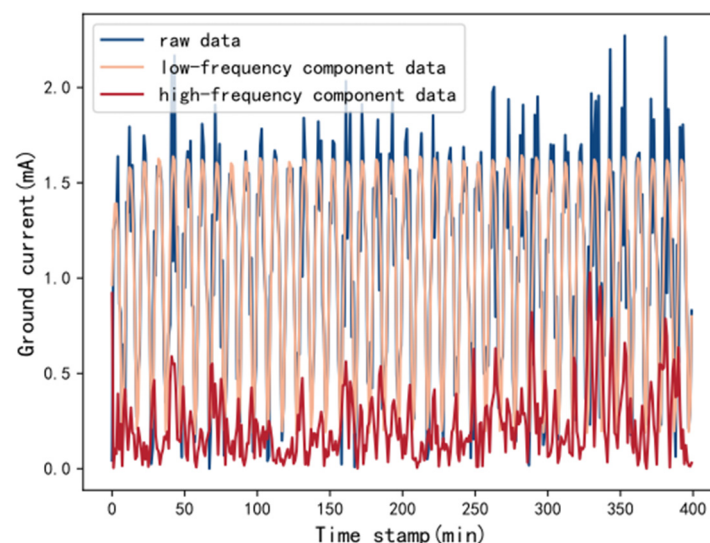


Figure 3. Schematic diagram of the data of the high- and low-frequency components at the random cut-off point.

The difference in the cut-off point between high and low frequencies directly affects the similarity between the resulting high and low-frequency component data and the original data. This paper uses the distance correlation coefficient [25] to calculate the correlation between the component data and the original data. This correlation coefficient improves upon the Pearson coefficient, as it can simultaneously represent both linear and nonlinear relationships between variables. The calculation method is as follows:

$$\rho^2(X, Z) = \frac{D^2(X, Z)}{\sqrt{D^2(X, X)D^2(Z, Z)}} \quad (1)$$

$$D^2(\mathbf{X}, \mathbf{Z}) = \frac{1}{n^2} \sum_{i,j=1}^n A_{i,j} B_{i,j} \quad (2)$$

$$A_{i,j} = \|X_i - X_j\|_2 - \frac{1}{n} \sum_{k=1}^n \|X_k - X_j\|_2 - \frac{1}{n} \sum_{l=1}^n \|X_i - X_l\|_2 + \frac{1}{n^2} \sum_{k,l=1}^n \|X_k - X_l\|_2 \quad (3)$$

$$B_{i,j} = \|Z_i - Z_j\|_2 - \frac{1}{n} \sum_{k=1}^n \|Z_k - Z_j\|_2 - \frac{1}{n} \sum_{l=1}^n \|Z_i - Z_l\|_2 + \frac{1}{n^2} \sum_{k,l=1}^n \|Z_k - Z_l\|_2 \quad (4)$$

Among them, $\mathbf{X}$ and $\mathbf{Z}$ are two vectors, which are used here to represent component data vectors of the grounding current and raw data vectors of the grounding current. n denotes the length of vectors $\mathbf{X}$ and $\mathbf{Z}$. $X_u, u \in \{i, j, k, l\}$ represents the u -th element in the vector $\mathbf{X}$. $Z_u, u \in \{i, j, k, l\}$ represents the u -th element in the vector $\mathbf{Z}$. $\rho^2(\mathbf{X}, \mathbf{Z})$ is the distance correlation coefficient between the two vectors; both $D^2(\mathbf{X}, \mathbf{Z})$ and $D^2(\mathbf{Z}, \mathbf{Z})$ can be calculated using the above formula.

In order to more accurately decompose the low-frequency and high-frequency component data, this paper selects the optimal frequency boundary point from the perspective of the correlation between the component data and the original data. As shown in Figure 4, when performing frequency domain decomposition of the grounding current, a larger selected frequency leads to a stronger correlation between the low-frequency component and the original data, while the opposite is true for the high-frequency component. The low-frequency component represents the fundamental trends and periodic changes in the data, offering higher predictability. In contrast, the high-frequency component mainly contains rapid changes, noise, and non-periodic fluctuations, making it more difficult to predict. If the proportion of low-frequency components in the data can be increased, accurately predicting the low-frequency components to balance the prediction errors of the high-frequency components will enhance overall prediction accuracy. Therefore, the optimal frequency cut-off selection approach proposed in this paper is to choose the highest possible frequency cut-off for the low-frequency component while ensuring that it occupies a significant proportion. However, the frequency cut-off cannot be too high, as this would render the high-frequency component unpredictable. Otherwise, it would be difficult to achieve a certain level of balance in the predictions.

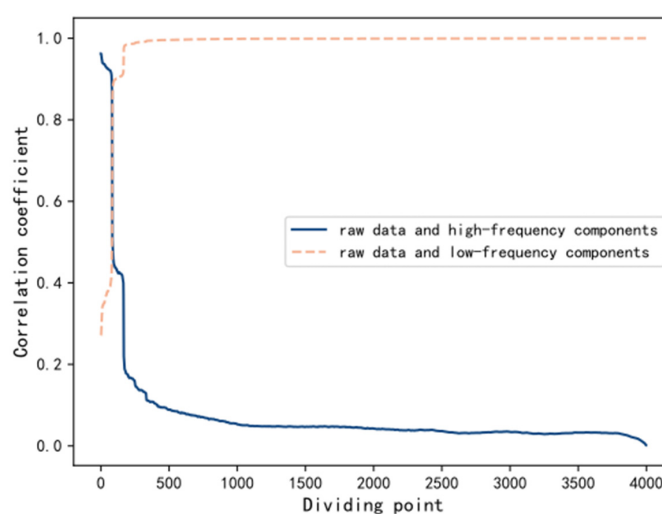


Figure 4. The correlation coefficient between the high and low-frequency components of different cut-off points and the original data.

In addition, as shown in Figure 4, when the frequency reaches a certain value, the correlation between the low-frequency component of the grounding current and the original data does not increase significantly as the frequency cut-off point increases, while the correlation between the high-frequency component and the original data continues to decrease. Therefore, this paper seeks to determine the optimal cut-off point for high and low frequencies by minimizing the difference in correlation coefficients between the high and low-frequency components and the original data. This can be regarded as a minimization optimization problem. The objective function is as follows:

$$O = \operatorname{argmin} |L_f - H_f|, f \in \left[0, \frac{N}{2}\right] \quad (5)$$

In the above equation, O represents the objective function. When the objective function reaches its minimum value, the correlation between the low-frequency component and the original data, as well as the correlation between the high-frequency component and the original data, achieves optimal balance. L_f and H_f represent the correlation coefficients of the low-frequency component and high-frequency component with the original data at frequency f , respectively. N denotes the length of the input samples.

The frequency component corresponding to the minimum value m of the objective function is selected as the optimal frequency cut-off point. Then, the inverse Fourier transform is applied to convert the frequency domain data into time domain data, thereby obtaining the time domain data of the grounding current for both the high-frequency and low-frequency components. The specific implementation process is shown in Equations (6)–(8).

$$F = \text{FFT}(X) \quad (6)$$

$$Z_h = \text{iFFT}(F[:m]) \quad (7)$$

$$Z_l = \text{iFFT}(F[m+1:]) \quad (8)$$

In the above equation, F represents the frequency domain component of the original data, FFT denotes the Fourier transform, Z_h represents the high-frequency time domain component of the grounding current, iFFT denotes the inverse Fourier transform, and Z_l represents the low-frequency time domain component of the grounding current.

2.2.2. Data Slicing and Embedding (DSAE)

For multivariate time series forecasting, the information provided by a single time point is limited. A period of time contains more local feature information, and extracting and utilizing these local features can effectively enhance the model's predictive accuracy. Additionally, when calculating the dependencies among multiple variables, considering only the associations between variables at a single time-stamp lacks contextual information, which can impact the effectiveness of multivariate predictions. This paper introduces a data slicing and embedding block that not only improves the temporal attention module's ability to extract features from local temporal information in sequences but also enables the model to better capture the relationships among multiple variables in the data.

After obtaining the high-frequency and low-frequency component data of the grounding current, they are input into DSAE, respectively, to obtain the feature-encoded vectors. The specific formulas are as follows:

$$[z_1^d, z_2^d, \dots, z_i^d] = \text{DSAE}(Z^d), i \in \left[1, \frac{L}{L_s}\right], d \in [1, D] \quad (9)$$

In the equation, Z^d represents the vector of the d -th dimension of the component data, z_i^d denotes the encoded result of the i -th sequence block in the dimension of d , L represents the length of the input sequence, L_s denotes the segment length of the sequence, and D represents the total number of variable dimensions in the input sequence.

First, the time series of the same dimension in the component data is extended using boundary value replication to ensure that the lengths of the segmented sequences remain consistent. Then, the sequences are sliced, which can be expressed by the following formula:

$$q_i^d = \text{unfold}(\text{padding}(Z^d)), i \in \left[1, \frac{L}{L_s}\right], d \in [1, D] \quad (10)$$

In the equation, q_i^d represents the slicing result of the i -th sequence block in the d -th dimension, unfold denotes the slicing operation on the sequence, and padding indicates the extension of the sequence.

Finally, the segmented sequences undergo value embedding and position embedding [14], resulting in the stacked embedded vectors e_{out} . The specific computation is shown in Equation (11).

$$\begin{aligned} e_{out} &= \text{valueEmbedding}(q_i^d) \\ &\quad + \text{positionEmbedding}(q_i^d) \\ e_{out} &= [z_1^d, z_2^d, \dots, z_i^d], i \in \left[1, \frac{L}{L_s}\right], d \in [1, D] \end{aligned} \quad (11)$$

2.2.3. Enhanced Attention Mechanism

Unlike univariate prediction tasks, in multivariate prediction tasks, variables from different dimensions influence each other. Therefore, while considering temporal dependencies, it is also necessary to account for the interrelationships among the time series data variables. To address this issue, this paper proposes an enhanced attention module that captures spatiotemporal features. This module primarily utilizes temporal attention and variable attention to capture the relationships between the segmented sequences' time dimensions and variable dimensions output by the DSAE module. A gating unit is then employed to merge these two dimensions' features, thereby enhancing the model's feature extraction capabilities. Its specific structure is shown in Figure 5.

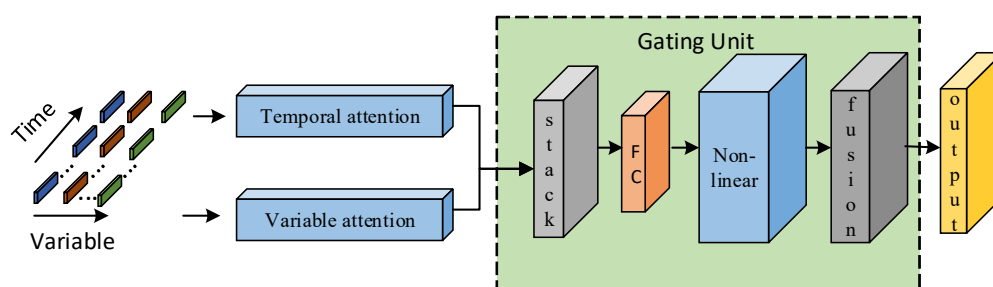


Figure 5. Enhanced attention mechanism.

Temporal attention primarily relies on an autocorrelation mechanism [14] to extract features across the time dimensions of the input sequence. Compared to the point attention mechanism used in Transformer, the autocorrelation mechanism offers higher computational efficiency and is better at capturing the periodic dependencies of time series. Its specific structure is shown in Figure 6.

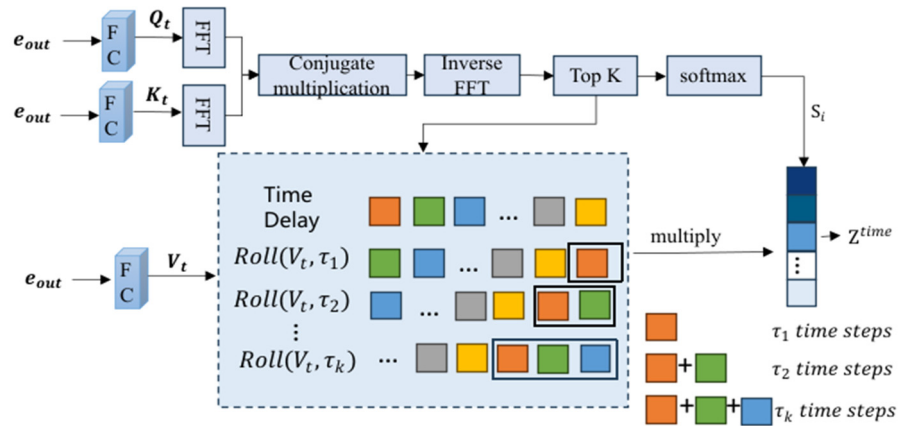


Figure 6. Autocorrelation mechanism.

First, the result e_{out} output by the DSEB module is mapped through the time dimension multi-head to obtain the query matrix Q_t , the key matrix K_t , and the value matrix V_t . The specific formulas are shown in Equations (12)–(14).

$$Q_t = e_{out} \times W_{Q_t}, W_{Q_t} \in \mathbb{R}^{d_k \times \frac{d_k}{s}} \quad (12)$$

$$K_t = e_{out} \times W_{K_t}, W_{K_t} \in \mathbb{R}^{d_k \times \frac{d_k}{s}} \quad (13)$$

$$V_t = e_{out} \times W_{V_t}, W_{V_t} \in \mathbb{R}^{d_k \times \frac{d_k}{s}} \quad (14)$$

In the above formula, W_{Q_t} , W_{K_t} and W_{V_t} are the weight matrices that can be learned by the model, d_k represents the model depth and s represents the number of attention heads.

Then, frequency domain conversion is performed through FFT, and the autocorrelation coefficient between Q_t and the lag sequence K_t which is shifted backward by τ time points, is calculated. The specific calculation method is shown in Equation (15).

$$R_{Q_t, K_t}(\tau) = \text{iFFT}(\text{FFT}(Q_t) \cdot \text{FFT}^*(K_t)), \tau \in \{1, \dots, L_s\} \quad (15)$$

In the above formula, τ represents the translation length and FFT^* represents performing FFT on the input vector first and then finding its conjugate vector. Next, obtain the K translation lengths with the largest autocorrelation coefficients, and normalize them through softmax to obtain the attention weight matrix S_i between sequences under different translation lengths. This process can be described as follows:

$$\tau_1, \dots, \tau_k = \text{argTopk}(R_{Q_t, K_t}(\tau)) \quad (16)$$

$$S_1, \dots, S_k = \text{softmax}(R_{Q_t, K_t}(\tau_1), \dots, R_{Q_t, K_t}(\tau_k)) \quad (17)$$

In the above formula, argTopk represents obtaining the K largest translation lengths. Finally, based on the obtained K translation lengths, the vector V_t is shifted backward and then multiplied by its attention weight to obtain the data characteristics after time-delay aggregation. The specific calculation process is as follows:

$$\text{AutoCorrelation}(Q_t, K_t, V_t) = \sum_i^m \text{Roll}(V_t, \tau_i) S_i \quad (18)$$

Among them, Roll represents the operation of shifting the input sequence backward. Based on the above process, the dependency relationship Z^{time} of the input sequence in the time dimension can be obtained.

$$Z^{time} = \text{AutoCorrelation}(Z_{:,d}, Z_{:,d}, Z_{:,d}) \quad (19)$$

In the equation, AutoCorrelation represents the autocorrelation mechanism and $Z_{:,d}$ denotes all the time series vectors in the variable dimension of d .

This paper employs a variable attention mechanism to learn the interrelationships between grounding currents at different connection points, and its specific structure is shown in Figure 7.

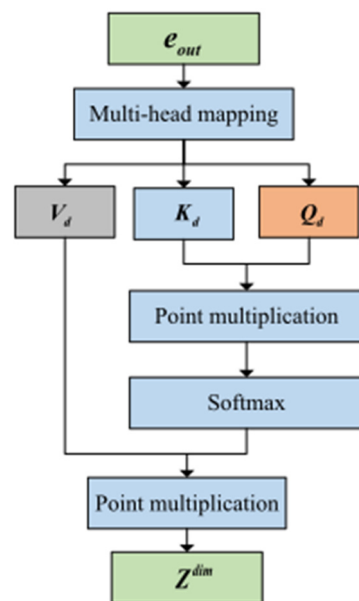


Figure 7. Variable attention mechanisms.

First, the output from the DSAE block is processed through variable multi-head mapping to obtain the query matrix Q_d , key matrix K_d , and value matrix V_d corresponding to the input vector. Then, the dot product of Q_d and K_d is computed to obtain the attention weight scores A between variables. After normalization using an activation function, this is multiplied by the value matrix V_d to obtain the high-dimensional features of the variable dimension. The calculation formula for this attention module is as follows:

$$\text{MSA}(Q_d, K_d, V_d) = \text{softmax}\left(\frac{Q_d \times K_d^T}{\sqrt{d}}\right) \times V_d \quad (20)$$

Based on the aforementioned self-attention module, the dependencies Z^{dim} among the variable dimensions of the input sequence can be obtained.

$$Z^{dim} = \text{MSA}(Z_{i,:}, Z_{i,:}, Z_{i,:}) \quad (21)$$

In the equation, $Z_{i,:}$ represents the vectors of all dimensions at the time step i .

To combine the feature information from the two attention modules, a gating unit is used to allow each attention module to obtain a learnable weight. First, the dimension of the feature-extracted vector Z^{dim} is transformed to align with the dimension of vector Z^{time} . Then, they are concatenated directly, followed by a linear projection layer and softmax function to obtain the weight distributions h_1 and h_2 corresponding to the two

attention modules. Based on this distribution, the output of the enhanced attention module is obtained through concatenation. The specific computation process is as follows:

$$G = W \cdot \text{Concat}(Z^{dim}, Z^{time}) + b \quad (22)$$

$$h_1, h_2 = \text{softmax}(G) \quad (23)$$

$$X_{att} = \text{Concat}(h_1 \times Z^{dim}, h_2 \times Z^{time}) \quad (24)$$

In the equation, W and b are the weight matrices that can be learned by the model, G is the feature matrix obtained from linear mapping, and X_{att} represents the output result of the enhanced attention module.

2.3. Encoder Layer

As shown in Figure 2, the encoder consists of an enhanced attention module and a forward propagation module. The input to the encoder is the high- and low-frequency component data obtained after best frequency domain decomposition. The component data are first segmented and embedded through data slicing and embedding, and then features are extracted by the enhanced attention module to produce the output of the different component data encoders. This process can be described in detail as follows:

$$e_{out}^c = \text{DSAE}(Z_c) \quad (25)$$

$$X_{en}^c = \text{ENAttn}(e_{out}^c, e_{out}^c, e_{out}^c) \quad (26)$$

In the above equation, $Z_c, c \in \{l, h\}$ represents the high-frequency time-domain component and the low-frequency time-domain component while X_{en}^c denotes the output of the encoder and ENAttn represents the enhanced attention mechanism.

2.4. Decoder Layer

The decoder is responsible for interpreting the high-dimensional features extracted by the encoder in both the time and variable dimensions. It aggregates the initialized sequences into predicted values and then merges the predicted values from different components into the final prediction result.

The input to the decoder is a zero matrix with the same shape as the prediction target. Similar to the encoder, it first passes through the data slicing and embedding module to obtain sequence blocks de_c for different time regions. The specific process can be described as follows:

$$de_c = \text{DSAE}(p_c) \quad (27)$$

In the equation, $p_c, c \in \{\text{low}, \text{high}\}$ represent the initial values for the low-frequency and high-frequency components during the decoding phase, which are zero vectors. Then, the enhanced attention module is used to extract time-dimensional features from the input data and learn the relationships between variables, resulting in the features X_{de}^c after attention-enhanced extraction. This can be expressed as follows:

$$X_{de}^c = \text{ENAttn}(de_c, de_c, de_c) \quad (28)$$

Subsequently, the enhanced cross-attention module aggregates the source sequence features outputted by the encoder and the features extracted by the enhanced attention in the decoder. Finally, a fully connected layer produces the predicted sequences for different

components, resulting in the final prediction outcome. The specific calculation process is as follows:

$$Y_{de}^c = \text{ENAttn}(X_{de}^c, X_{en}^c, X_{en}^c) \quad (29)$$

$$r_c = \text{FeedForWard}(Y_{de}^c) \quad (30)$$

$$r = r_{low} + r_{high} \quad (31)$$

In the equation, Y_{de}^c represents the output of the enhanced cross-attention, FeedForWard denotes the forward propagation, $r_c, c \in \{\text{low}, \text{high}\}$ represent the predicted results for the low-frequency and high-frequency components, respectively, and r is the final prediction result.

3. Experimental Analysis

3.1. Experimental Datasets

To validate the method presented in this paper, a simulation experiment was conducted based on the actual equipment conditions of a 220 kV substation in northern China. A simulation model of the grounding current collection system for the 220 kV substation, as shown in Figure 8, was constructed in MATLAB(R2021a)/Simulink. This simulation model includes five types of equipment from left to right: circuit breakers, lightning arresters, current and voltage transformers, and transformers. The grounding network model is composed of multiple horizontal and vertical grounding electrodes, where each pair of parallel (or vertical) horizontal grounding sections forms a π -type circuit. The grounding currents (current) at six grounding points labeled a, b, c, d, e, and f, as shown in Figure 8, are collected and denoted as I_a, I_b, I_c, I_d, I_e , and I_f , respectively. Additionally, all equipment used in the model is consistent with the parameters of actual substation equipment, ensuring the rigor of the experimental data.

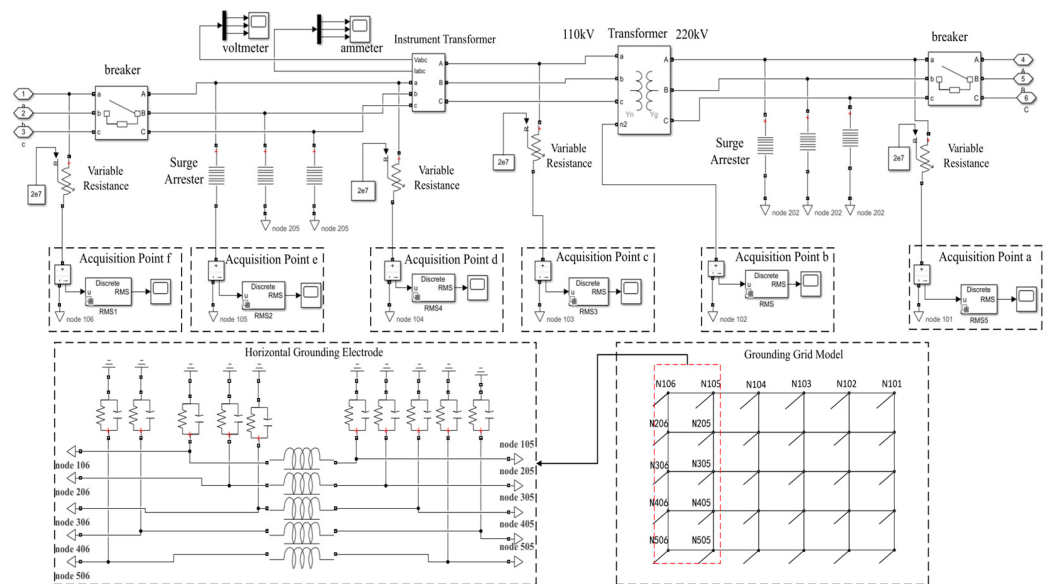


Figure 8. Simulation model of the grounding current collection system for the 220 kV substation.

In the simulation model, the grounding resistance is set to $3 \times 10^7 \Omega$ by default to simulate the grounding conditions of the substation equipment during normal operation. Additionally, considering the actual operating conditions of the substation, variable resistors are added at each equipment grounding point to dynamically adjust the grounding resistance, with a variation range of $[3 \times 10^7 \Omega, 1 \times 10^7 \Omega]$. An RMS module is included at the current output terminal to record the effective value of the current every minute.

To assess the correlation between the grounding currents at different grounding points, this study employs the Pearson correlation coefficient for analysis, as shown in Figure 9. It is clear that there is a significant correlation between the grounding currents at different points, which supports the validation of the proposed model's performance in multivariable prediction tasks.

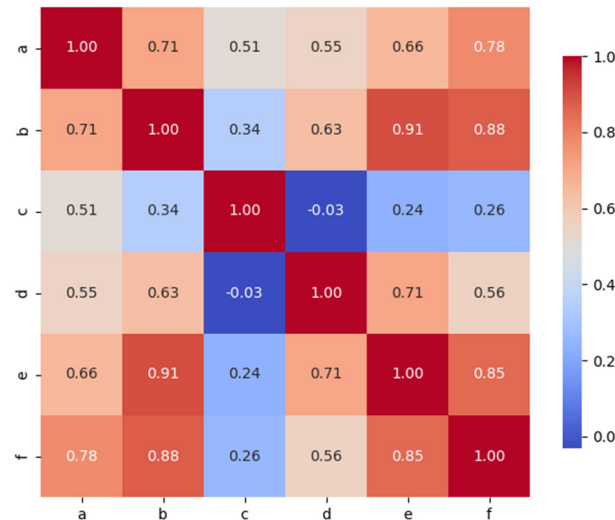


Figure 9. Pearson correlation coefficient between grounding currents at different grounding points.

In addition, to verify the generalization ability of Fdeformer, this paper also selects the transformer temperature (ETT) [14] and exchange rate (Exchange) [14] datasets to analyze the effect of the proposed method in this paper in the multivariate prediction task. The transformer temperature dataset consists of six different types of power load characteristics and the predicted target value “oil temperature”, and the exchange rate dataset consists of the exchange rates of 8 countries. All datasets are divided into the training set, validation set, and test set according to a ratio of 7:1:2. The information on different datasets is shown in Table 1.

Table 1. Description of the experimental dataset.

Dataset	Time Frame	Time Step	Sampling Interval
current	15 December 2023–5 January 2024	30,000	1 min
ETT	1 July 2016–26 June 2018	69,680	15 min
Exchange	1 January 1990–10 October 2010	7588	24 h

3.2. Data Preprocessing

During the collection of grounding current information for substation equipment, the data can be affected by factors such as sensor malfunctions, transmission equipment issues, and external environmental interference, leading to problems like outliers and abrupt changes. This paper employs the robust quartile method [26] to detect and remove outliers that are minimally influenced by extreme values. Given the strong correlation between adjacent data points in the time series, missing values are filled using interpolation, replacing the missing values with the average of the nearest preceding and succeeding data points at equal intervals. Some of the processed monitoring data are shown in Table 2.

Table 2. Some monitoring data after exception handling.

Time	Ia/mA	Ib/mA	Ic/mA	Id/mA	Ie/mA	If/mA
24 December 2023 T09:30:00	0.74	3.24	0.65	0.18	2.55	0.35
24 December 2023 T09:31:00	0.51	3.59	0.20	0.58	3.56	0.64
24 December 2023 T09:32:00	0.66	2.60	0.32	0.56	5.57	0.31
24 December 2023 T09:33:00	1.01	2.42	0.43	0.33	3.49	0.12
24 December 2023 T09:34:00	0.92	1.59	0.12	0.43	2.33	0.48
24 December 2023 T09:35:00	0.58	3.12	0.56	0.22	1.74	0.51
24 December 2023 T09:36:00	0.47	2.27	0.45	0.36	2.79	0.20

3.3. Experimental Environment and Evaluation Metrics

The model is implemented using the PyTorch 2.0.1 framework and Python 3.8. The hardware environment consists of an NVIDIA GeForce RTX 3060 Ti GPU and an Intel Core i7-12700F CPU. The prediction results in this paper are evaluated using three metrics: Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). A smaller value for these evaluation metrics indicates better predictive performance of the model.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (32)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (33)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (34)$$

In the equation, y_i represents the predicted result of the model, $\hat{y}_i$ represents the true value of the data.

3.4. Model Parameter Settings and Benchmark Model Selection

Before training the Fdeformer model, it is necessary to set the basic parameters for each part of the model. The number of encoder layers, decoder layers, attention heads, and batch size are set to 2, 1, 8, and 32, respectively. The Adam optimizer is chosen, and the GELU activation function is used. During training, the learning rate is set to 0.0001, with a maximum of 10 training epochs. After each epoch, the validation set is tested, and training is stopped if the validation loss increases for three consecutive epochs. The weights corresponding to the best performance on the validation set are used for subsequent predictions. For grounding current prediction, the hyperparameter L_s is set to 20, while for predicting the public datasets, it is set to 24.

In order to objectively evaluate our method, we used five popular transformer-based time series prediction models as the benchmark: Reformer (Kitaev N et al., 2020) [15], Informer (Zhou et al., 2021) [12], Autoformer (Wu et al., 2021) [14], Fedformer (Zhou et al., 2022) [16] and TimesNet (Wu et al., 2023) [17]. In order to enhance the competitiveness of benchmark models, their hyperparameters have been adjusted to ensure that they are in their best-fit state.

3.5. Analysis of Experimental Results

3.5.1. Best Frequency Domain Decomposition Algorithm

In Section 2.1.1, this paper constructs an objective function with the goal of minimizing the difference in the correlation coefficients between the high-frequency and low-frequency components and the original data. The frequency spectrum range of the input grounding

current data is searched to find the point where the objective function reaches its minimum value, which is taken as the high–low frequency division point. The variation function of the objective value is shown in Figure 10. When the objective function reaches its minimum, the correlation between the high-frequency and low-frequency components and the original data is optimally balanced. The determined decomposition point can effectively separate steady-state and transient features, enhancing the component features contained in the data.

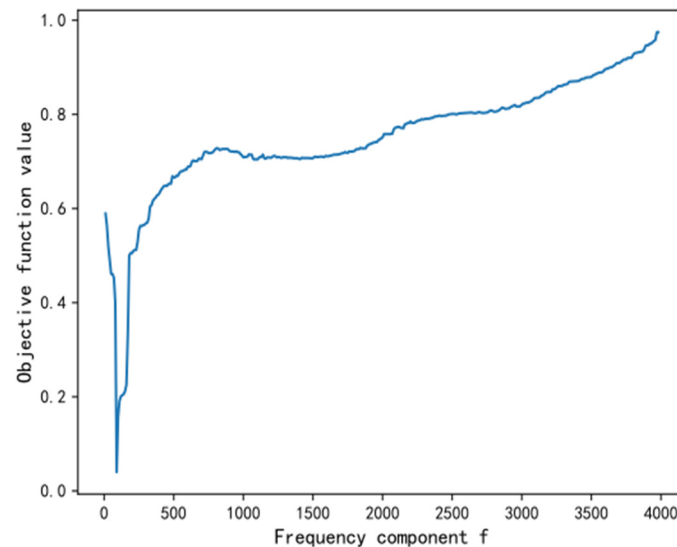


Figure 10. Curve of change in target value.

To verify the effectiveness of the proposed time-series decomposition algorithm, the original data are processed using four strategies: no decomposition, decomposition using the EMD algorithm [23], decomposition using the VMD algorithm [24], and decomposition using the best frequency domain decomposition algorithm to obtain high and low-frequency component data. The proposed model is then used to predict the decomposed component data. The comparison results of the prediction models for different decomposition algorithms are shown in Table 3.

Table 3. Comparison of the prediction performance of grounding current in different decomposition models.

Method	MAE	MSE
No Decomposition	0.421	0.346
EMD	0.403	0.339
VMD	0.382	0.322
BFDD	0.360	0.313

From Table 3, the following conclusions can be drawn:

- (1) Compared to predicting directly with the original data without decomposition, using the EMD algorithm, VMD algorithm, and optimal frequency domain decomposition algorithm yields better results. The MAE accuracy decreased by 4.5%, 10.2%, and 16.9%, respectively, while the MSE accuracy decreased by 2.1%, 7.5%, and 10.5%. It can be observed that decomposition before prediction improves prediction accuracy;
- (2) Compared to the EMD and VMD algorithms, the proposed optimal frequency domain decomposition algorithm shows a reduction in MAE accuracy of 11.9% and 6.1%, respectively, and a reduction in MSE accuracy of 8.3% and 2.9% on the self-constructed

grounding current dataset, validating the effectiveness of the proposed decomposition algorithm.

3.5.2. Low-Frequency Component Prediction

To validate the effectiveness of the predictions made by Fdeformer on the low-frequency components obtained after optimal frequency domain decomposition, we conducted experimental comparisons with classical time series prediction models: Autoformer, Informer, Reformer, Fedformer, and TimesNet.

Figure 11 shows a comparison of the prediction results for the low-frequency components of the grounding current at the grounding point across different models. It is evident that Fdeformer outperforms the other time series prediction models in predicting the low-frequency components of the grounding current. At most time points, Fdeformer's predicted values are nearly consistent with the actual values. In contrast, Fedformer, Autoformer and TimesNet exhibit significant prediction errors at certain time points. For instance, at the 43rd minute, the predicted values of Fedformer, Autoformer and TimesNet differ from the actual values by 0.11 mA, 0.13 mA and 0.13 mA, respectively, while Fdeformer's prediction deviates by only 0.08 mA. Informer and Reformer, compared to the other models, can only provide a rough trend of the actual values and fail to achieve accurate predictions.

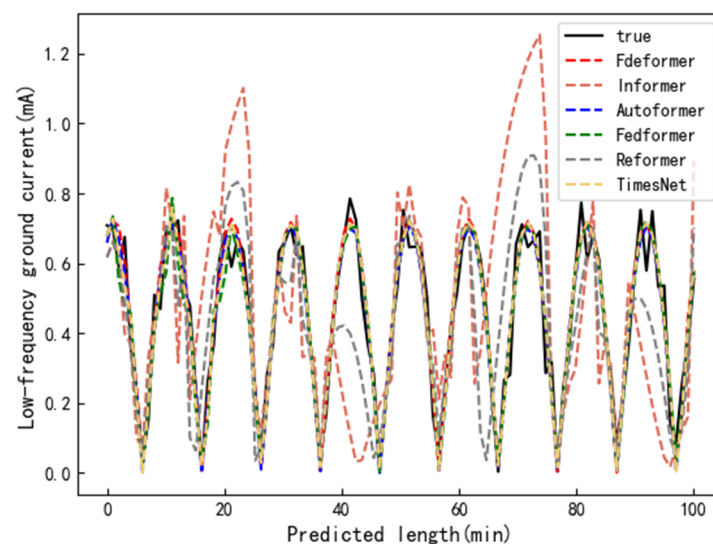


Figure 11. The prediction results of the low-frequency components of each model.

Table 4 presents the prediction errors of each model for the low-frequency components of the grounding current. As shown in the table, Fdeformer demonstrates lower evaluation metrics compared to the other models when predicting the low-frequency components of the grounding current. Specifically, the MAE is reduced by 8.1%, 66.2%, 46.6% and 1.3% compared to Autoformer, Informer, Reformer, and Fedformer, respectively. Additionally, the MSE is reduced by 16.1%, 61.3%, 56.2%, 8.5% and 1.3% compared to Autoformer, Informer, Reformer, Fedformer and TimesNet, respectively.

In summary, when predicting the low-frequency component of grounding current, using the Fdeformer proposed in this paper can achieve higher prediction accuracy.

Table 4. Comparison of prediction performance of low-frequency components of grounding current in different models.

Model	MAE	MSE
Informer	0.389	0.350
Autoformer	0.253	0.252
Fedformer	0.237	0.229
Reformer	0.343	0.339
TimesNet	0.234	0.220
Fdeformer	0.234	0.217

3.5.3. High-Frequency Component Prediction

To validate the effectiveness of Fdeformer in predicting the high-frequency component obtained after optimal frequency domain decomposition, experimental comparisons were conducted with classical time-series forecasting models such as Autoformer, Informer, Reformer, Fedformer, and TimesNet.

Figure 12 shows the comparison of the prediction results of the high-frequency component of the grounding current at grounding point a for each model. Although the aforementioned models struggle to achieve accurate predictions when dealing with the clearly nonlinear characteristics of the high-frequency component, Fdeformer's predicted values are closest to the actual values at the most extreme points, demonstrating the best-fitting performance.

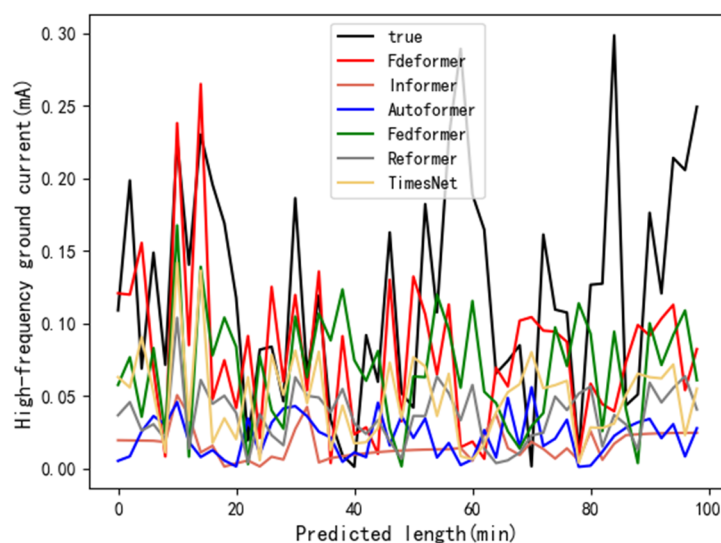
**Figure 12.** The prediction results of the high-frequency components of each model.

Table 5 presents the prediction errors of each model for the high-frequency component of the grounding current data. It can be seen from the table that Fdeformer outperforms other models in predicting the low-frequency component of the grounding current, with all evaluation metrics being lower than those of the other models. Specifically, compared to Fedformer, Informer, Reformer, Autoformer and TimesNet, the MAE is reduced by 8.3%, 59.2%, 60.4%, 16% and 1.5%, respectively, while the MSE is reduced by 5.7%, 61.7%, 62.5%, 8.8% and 1.7%.

Table 5. Comparison of prediction performance of high-frequency components of grounding current in different models.

Model	MAE	MSE
Informer	1.568	1.163
Autoformer	0.762	0.488
Fedformer	0.697	0.472
Reformer	1.612	1.187
TimesNet	0.642	0.453
Fdeforner	0.639	0.445

In summary, when predicting the high-frequency component of grounding current, choosing the Fdeforner, which considers the dependencies between variables, achieves higher prediction accuracy compared to other models.

3.5.4. Comparison and Analysis of the Proposed Model with Other Models

To validate the overall predictive performance of Fdeforner, predictions were made using both the aforementioned single models and the proposed method on both a self-constructed grounding current dataset and a public dataset. The input data length for the self-constructed dataset was 300 min, with predictions made for the grounding current for the next 100, 200, and 300 min. The error comparisons of the models on the self-constructed dataset are shown in Table 6. For the public dataset, the input length was 96, and the output lengths for predictions were 96, 192, 336, and 720. The error comparisons of the models on the public dataset are shown in Table 7, with bold black indicating the best prediction performance.

Table 6. Comparison experiment of self-built dataset models.

Dataset	Model	100 min			200 min			300 min		
		MAE	MSE	MAPE	MAE	MSE	MAPE	MAE	MSE	MAPE
current	Reformer	0.875	1.077	25.90%	0.899	1.120	26.58%	0.906	1.132	26.91%
	Informer	0.672	0.813	22.73%	0.662	0.749	23.31%	0.873	0.743	24.35%
	Autoformer	0.398	0.348	15.42%	0.406	0.357	15.89%	0.409	0.362	16.47%
	Fedformer	0.384	0.338	13.41%	0.395	0.354	14.17%	0.404	0.357	14.73%
	TimesNet	0.365	0.317	12.89%	0.379	0.350	13.84%	0.390	0.352	14.01%
	Fdeforner	0.360	0.313	12.37%	0.372	0.343	13.32%	0.381	0.346	13.89%

Table 7. Comparison of experimental results of public dataset models.

Dataset	Model	96		192		336		720	
		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
ETTm2	Reformer	0.619	0.658	0.827	1.078	0.972	1.549	1.242	2.631
	Informer	0.453	0.365	0.563	0.533	0.887	1.363	1.388	3.379
	Autoformer	0.339	0.255	0.340	0.281	0.372	0.339	0.419	0.422
	Fedformer	0.287	0.203	0.328	0.269	0.366	0.325	0.415	0.421
	TimesNet	0.283	0.201	0.324	0.265	0.351	0.321	0.413	0.418
	Fdeforner	0.281	0.201	0.324	0.269	0.359	0.318	0.408	0.417
Exchange	Reformer	0.829	1.065	0.906	1.188	0.976	1.350	1.016	1.510
	Informer	0.752	0.847	0.895	1.204	1.036	1.672	1.310	2.478
	Autoformer	0.323	0.197	0.369	0.300	0.524	0.509	0.941	1.447
	Fedformer	0.278	0.148	0.380	0.271	0.500	0.460	0.841	1.195
	TimesNet	0.254	0.137	0.348	0.226	0.468	0.389	0.832	1.148
	Fdeforner	0.255	0.132	0.348	0.225	0.462	0.397	0.822	1.144

Experiments were conducted on a dataset with interrelated variables related to grounding current. The results indicate that the proposed model achieved the lowest loss value. Compared with TimesNet, which performs best among the comparison models, the MSE index of Fdeformer decreased by an average of 1.68%, the MAE decreased by an average of 1.88%, and the MAPE decreased by an average of 2.98%. This demonstrates that Fdeformer effectively considers the dependencies between variables and fully leverages the advantages of the enhanced attention mechanism in handling multivariate data. To further validate the generalization capability of the proposed method, experiments were conducted on a public dataset. The experimental results showed that our proposed Fdeformer achieved 12 top-1 and 4 top-2 results out of a total of 16 cases, significantly outperforming the comparison methods. This result confirms the strong predictive capability of the Fdeformer model on multivariate public datasets.

Figure 13 shows the grounding current prediction results for the next 100 min at grounding point f for different models. It is evident that both Informer and Reformer fail to accurately predict the changes in grounding current. While Autoformer can predict the trend of grounding current changes accurately, it struggles to fit the data during abrupt changes. In contrast, Fedformer and TimesNet better capture the data at the points of abrupt changes; however, as the prediction horizon increases, the performance of Fedformer and TimesNet starts to decline. The prediction results of the proposed model show a closer alignment with the actual data over time, particularly after 60 min. By this point, Fedformer is unable to fit the fluctuations in the inflection points, while the proposed model continues to capture the changes in the data effectively.

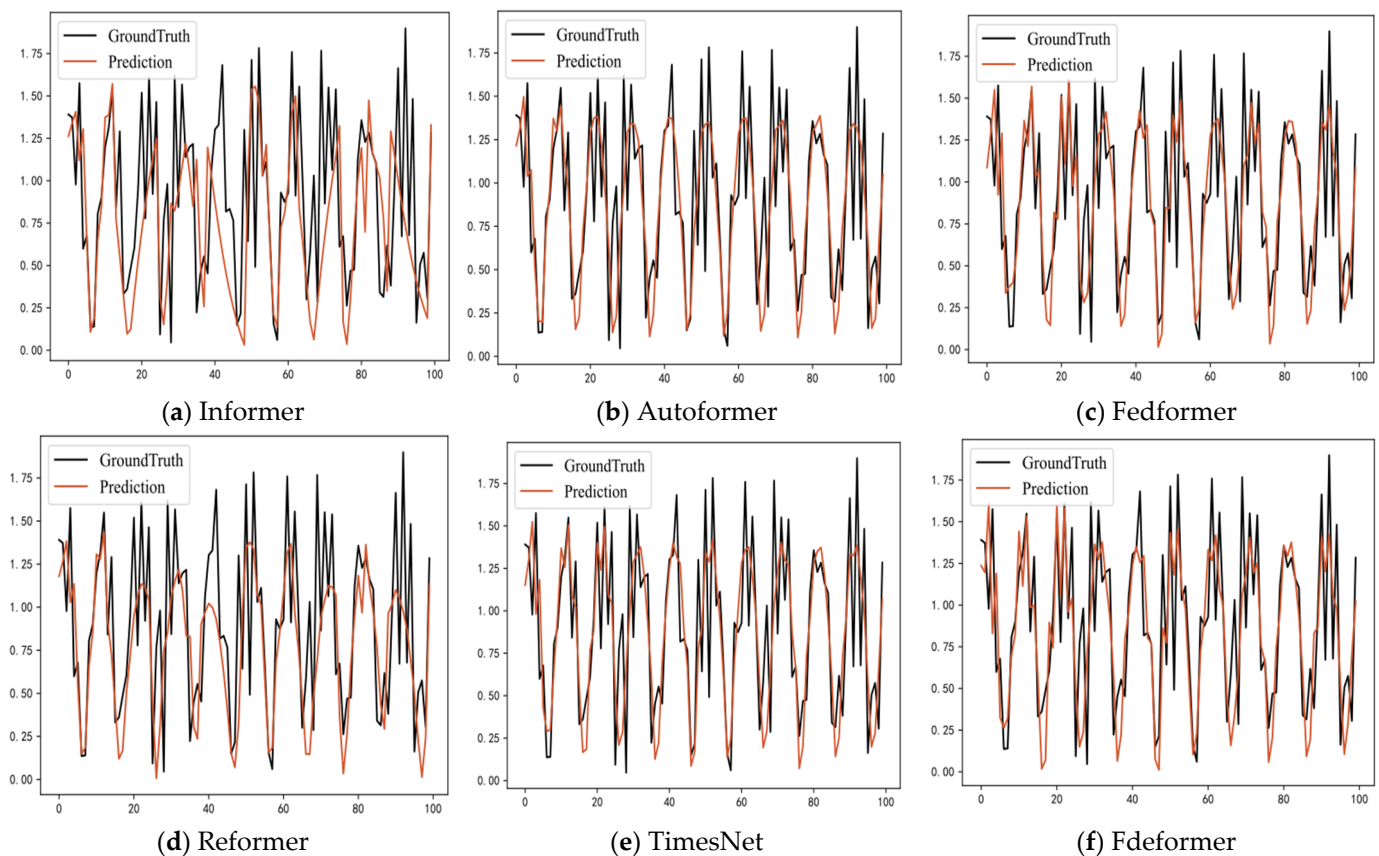


Figure 13. The prediction results of the ground current of each model.

3.5.5. Ablation Studies

To validate the effectiveness of the best frequency domain decomposition module in the Fdeformer model, as well as the data slicing and embedding module and the enhanced attention mechanism, this paper conducted ablation experiments. A 100 min prediction task was performed on a self-constructed dataset, using the Fdeformer model as the baseline. It was compared with variants of Fdeformer, where the optimal frequency domain decomposition module was removed, where the data slicing and embedding module was removed, and where the autocorrelation mechanism was used instead of the enhanced attention mechanism. The results are shown in Table 8. Here, the average increase in error refers to the average increase in MAE and MSE metrics for each model variant compared to Fdeformer.

Table 8. Comparison of the performance of Fdeformer with its various variants.

Model	MAE	MSE	Average Increase in Error
Fdeformer	0.360	0.313	----
Remove BFDD	0.421	0.346	13.7%
Remove DSAE	0.367	0.322	2.4%
Auto-Correlation	0.398	0.348	10.9%

It can be observed that (1) the optimal frequency domain decomposition module plays a positive role, as it effectively decomposes the data, allowing the model to better extract features from different components; (2) the data slicing and embedding module contributes to improved prediction accuracy, as it fully exploits and utilizes the local temporal information in the time series data; (3) the Fdeformer with the enhanced attention mechanism outperforms the Fdeformer using the autocorrelation mechanism, as the enhanced attention mechanism can better extract the implicit associations between the time dimension and variable dimension of the time series data. These experiments demonstrate that each module of the Fdeformer model can enhance the model's predictive performance.

3.5.6. Hyperparameter Sensitivity Analysis

Figure 14 verifies the impact of key hyperparameters L_s on model performance in the prediction task of 100, 200, 300 time steps of grounding current. The setting of these parameters affects the model's ability to extract localized features. As the value of L_s increases, the model's prediction error first increases and then decreases. This is because, initially, as the segment length increases, the model can capture more localized features in the sequence. However, a longer segment length L_s does not necessarily yield better results, as excessively long segments can lead to model overfitting, resulting in decreased performance. Experimental validation shows that when $L_s = 20$, a balance is achieved in capturing multi-scale local features in the grounding current prediction task. This setting neither overemphasizes the microscopic variations caused by high-frequency noise interference nor overlooks the key trend information in the low-frequency components. It effectively prevents the performance degradation caused by feature extraction bias in the complex and variable grounding current data. Therefore, we select the value of 20, where the error is at its lowest, as the segment length for grounding current prediction.

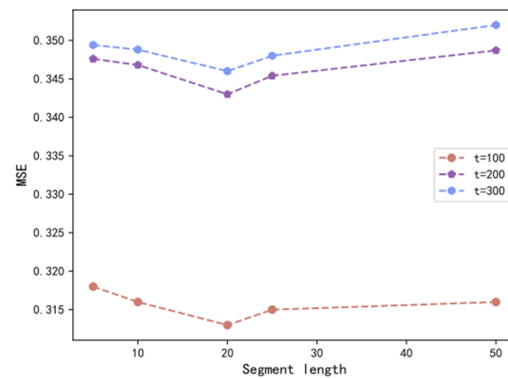


Figure 14. MSE changes with segment length.

4. Conclusions

This paper presents a frequency-enhanced transformer method for grounding current prediction, which effectively captures the dependencies in both the time and variable dimensions of the data for accurate forecasting. The proposed method first determines the optimal boundary between high- and low-frequency components during the data preprocessing stage by minimizing the difference in correlation coefficients between these components and the original data, addressing the issue of mode stacking that can occur with traditional time series decomposition methods. Next, a data slicing and embedding module is designed to facilitate sequence-level analysis of the time series data. Finally, an enhanced attention mechanism is proposed to achieve synchronous extraction of features in both the time and variable dimensions of the time series data, improving the model's feature extraction capability. Extensive experiments conducted on both a self-built dataset and public datasets demonstrate that, compared to traditional attention models, this method achieves higher prediction accuracy in multivariate time series forecasting tasks. Ablation studies on the self-built dataset validate the effectiveness of the proposed modules.

Although Fdeformer has achieved good results, it still has limitations in extremely complex scenarios. For example, in substations affected by strong and fluctuating electromagnetic interference, frequent random electrical faults, and harsh natural environments, the grounding current data may exhibit rare, complex, and irregular fluctuation patterns. In such cases, Fdeformer may experience a decline in prediction accuracy due to the difficulty in covering all extreme operating conditions in the training data. This is particularly evident when there are sudden sharp changes or prolonged abnormal fluctuations in the current, leading to deviations between predicted and actual values. Future research will focus on improving the model's adaptability to extreme and complex conditions, expanding the diversity of training data, optimizing the model structure and parameters, and further enhancing the robustness and generalization capability of Fdeformer.

Author Contributions: Conceptualization, N.Z. and G.Y.; methodology, N.Z. and G.Y.; software, Z.F.; validation, N.Z., Z.F. and J.H.; formal analysis, N.Z.; investigation, Z.F.; resources, J.H. and G.Y.; data curation, Z.F.; writing—original draft preparation, N.Z.; writing—review and editing, N.Z. and G.Y.; visualization, Z.F.; supervision, J.H.; project administration, G.Y.; funding acquisition, G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of State Grid Shanxi Electric Power Company: Research and application of synchronous measurement of micropower consumption and anomaly identification technology of ground loop current in substation (520530230010).

Data Availability Statement: The original contributions presented in this study are included in this article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Na Zhang and Gang Yang were employed by the State Grid Shanxi Electric Power Company. Authors Zilong Fu and Junsheng Hou were employed by State Grid Shanxi Electric Power Company Jincheng Power Supply Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Miao, P.; Wang, Q.B.; Wu, W.J.; Zhan, N.; Lin, F. Analysis on harmonic characteristics of grounding current of converter transformer core clamp. *Guangdong Electr. Power* **2024**, *37*, 90–97.
2. Guo, W.; Pan, Z.H.; Ren, Z.G.; Zhou, S.Y.; Li, H.C.; Wang, M.X. Multidimensional characteristic analysis of grounding current harmonics for thermal aging evaluation of XLPE power cable. *Proc. CSU-EPSCA* **2023**, *35*, 134–139.
3. Shan, Y.Z.; Xie, J.C.; Wang, S.L.; Wang, C.T. Analysis of cause and treatment of a 220kV starter transformer core grounding current exceeding standard. *Transformer* **2023**, *60*, 69–71.
4. Meng, Y.K.; Pan, Z.H.; Liu, B.; Chen, P.; Zhao, J.Y. Defect recognition of distribution power cable through harmonic analysis of grounding current. *Proc. CSU-EPSCA* **2024**, *36*, 116–121.
5. Xing, C.; Zhao, J.; Liu, T.; Yang, Y.; Yang, S.B.; Wang, H. On-site detection method for resistive current of metaloxide surge arrester based on ground current of capacitive equipment. *Insul. Surge Arresters* **2022**, *1*, 113–117.
6. Mu, L.H.; Li, G.X.; Meng, Q.H. Effect of interconnection of protective earthing networks on high-voltage and low-voltage distribution systems. *J. China Coal Soc.* **2004**, *2*, 238–242.
7. Wen, Q.S.; Zhou, T.; Zhang, C.L.; Chen, W.Q.; Ma, Z.Q.; Yan, J.C.; Sun, L. Transformers in time series: A survey. *arXiv* **2022**, arXiv:2202.07125.
8. Campos, F.D.; Sousa, T.C.; Barbosa, R.S. Short-Term Forecast of Photovoltaic Solar Energy Production Using LSTM. *Energies* **2024**, *17*, 2582. [\[CrossRef\]](#)
9. Peng, J.; Yu, W. The algorithm of current prediction based on multi-dimensional Long Short Term Memory networks. *Energy Rep.* **2021**, *7*, 1114–1120. [\[CrossRef\]](#)
10. Qi, L.; Zhang, Q.; Xie, Y.; Zhang, J.; Ke, J. Research on Wind Turbine Fault Detection Based on CNN-LSTM. *Energies* **2024**, *17*, 4497. [\[CrossRef\]](#)
11. Zhang, J.L.; Li, Y.; Zhu, C.S.; Xue, H.L.; Ma, J.W.; Zhang, L.X.; Bi, S. Substation equipment malfunction alarm algorithm based on dual-domain sparse transformer. *Comput. Sci.* **2024**, *51*, 62–69.
12. Zhou, H.Y.; Zhang, S.H.; Peng, J.Q.; Zhang, S.; Li, J.X.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [\[CrossRef\]](#)
13. Zhong, J.; Li, H.; Chen, Y.; Huang, C.; Zhong, S.; Geng, H. Remaining Useful Life Prediction of Rolling Bearings Based on ECA-CAE and Autoformer. *Biomimetics* **2024**, *9*, 40. [\[CrossRef\]](#)
14. Wu, H.X.; Xu, J.H.; Wang, J.M.; Long, M.S. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.
15. Kitaev, N.; Kaiser, L.; Levskaya, A. Reformer: The efficient transformer. *arXiv* **2020**, arXiv:2001.04451.
16. Zhou, T.; Ma, Z.Q.; Wen, Q.S.; Wang, X.; Sun, L.; Jin, R. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. *Int. Conf. Mach. Learn.* **2022**, *162*, 27268–27286.
17. Wu, H.X.; Hu, T.G.; Liu, Y.; Zhou, H.; Wang, J.M.; Long, M.S. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv* **2022**, arXiv:2210.02186.
18. Wang, Y.; Xu, Z.; Zhao, S.; Fan, Y. Performance degradation prediction of rolling bearing based on temporal graph convolutional neural network. *J. Mech. Sci. Technol.* **2024**, *38*, 4019–4036. [\[CrossRef\]](#)
19. Chen, H.; Zhu, M.; Hu, X.; Wang, J.; Sun, Y.; Yang, J. Research on short-term load forecasting of new-type power system based on GCN-LSTM considering multiple influencing factors. *Energy Rep.* **2023**, *9*, 1022–1031. [\[CrossRef\]](#)
20. Yan, J.H.; Liu, Z.Y.; Wang, Z. Vehicle Trajectory Prediction based on Spatial-Temporal Attention Mechanisms and Multi-Scale Convolutional Social Pooling. *Comput. Eng.* **2024**, 1–10.
21. Guan, Y.; Li, N.; Chen, P. Vehicle trajectory prediction method integrating spatiotemporal relationships with hybrid time-step scene interaction. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2024**, 09544070241277412. [\[CrossRef\]](#)
22. Gao, L.; Gu, X.; Chen, F.; Wang, J. Sparse Transformer Network with Spatial-Temporal Graph for Pedestrian Trajectory Prediction. *IEEE Access* **2024**, *12*, 144725–144737. [\[CrossRef\]](#)
23. Zhang, Z.; Deng, A.; Wang, Z.; Li, J.; Zhao, H.; Yang, X. Wind Power Prediction Based on EMD-KPCA-BiLSTM-ATT Model. *Energies* **2024**, *17*, 2568. [\[CrossRef\]](#)
24. Li, W.; Fan, N.; Peng, X.; Zhang, C.; Li, M.; Yang, X.; Ma, L. Fault Diagnosis for Motor Bearings via an Intelligent Strategy Combined with Signal Reconstruction and Deep Learning. *Energies* **2024**, *17*, 4773. [\[CrossRef\]](#)

25. Li, C.; Zhu, D.; Hu, C.H.; Li, X.L.; Nan, S.Q.; Huang, H. ECDX: Energy consumption prediction model based on distance correlation and XGBoost for edge data center. *Inf. Sci.* **2023**, *643*, 119218. [[CrossRef](#)]
26. Guan, Q.; Geng, K. Research on Cleaning Method of Wind Power Data Based on Improved Quartile Method. In Proceedings of the 2024 2nd International Conference on Signal Processing and Intelligent Computing (SPIC), Guangzhou, China, 20–22 September 2024; IEEE: New York, NY, USA, 2024; pp. 478–482.

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.