

Article

Short-Term Electrical Load Forecasting Based on VMD and GRU-TCN Hybrid Network

Changchun Cai ^{1,2,*}, Yuanjia Li ^{1,2}, Zhenghua Su ³, Tianqi Zhu ^{1,2} and Yaoyao He ^{1,2}

¹ Jiangsu Key Laboratory of Power Transmission and Distribution Equipment Technology, Hohai University, Changzhou 213022, China; 18739086714@163.com (Y.L.); zhutq@163.com (T.Z.); heyy@hhu.edu.cn (Y.H.)

² College of The Internet of Things Engineering, Hohai University, Changzhou 213022, China

³ State Grid Changzhou Power Supply Company, Changzhou 210024, China; gerrysu@163.com

* Correspondence: 20031690@hhu.edu.cn

† Member, IEEE.

Abstract: With the continuous increase in user-side flexible controllable resources connected into a distribution system, the components of electrical load become too diverse and difficult to be accurately forecasted. A short-term load forecast method that integrates variational modal decomposition (VMD), gated recurrent unit (GRU) and time convolutional network (TCN) into a hybrid network is proposed in this paper. Firstly, original electrical load sequence data with noise are decomposed into intrinsic IMF components with different frequencies and amplitudes based on the VMD method. Secondly, a combined load forecasting method based on the GRU and TCN network is proposed for the high and low-frequency load subsequent signals, respectively. Finally, the high and low-frequency signals forecasting results of the GRU and TCN network are rebuilt for the final load forecasting. The experiment results based on actual operation data (data set 1) and simulation data (data set 2), which show that the proposed method can reduce the forecasting error by 36.20% and 10.8%, respectively, in comparison with VMD-GRU. The reliability and accuracy of the proposed method is verified through the comparison with other methods such as LSTM, Prophet and XG Boost.

Keywords: short-term load forecasting; variational modal decomposition (VMD); gated recurrent unit (GRU); time convolutional network (TCN); hybrid algorithm



Citation: Cai, C.; Li, Y.; Su, Z.; Zhu, T.; He, Y. Short-Term Electrical Load Forecasting Based on VMD and GRU-TCN Hybrid Network. *Appl. Sci.* **2022**, *12*, 6647. <https://doi.org/10.3390/app12136647>

Academic Editor: Luis Hernández-Calleja

Received: 1 June 2022

Accepted: 27 June 2022

Published: 30 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

With the presentation of the goal of “carbon peaking and carbon neutrality” in China, the construction of a new electrical power system based on new energy is imminent. Under the new electrical power system architecture, the interaction of renewable energy generations, energy storage and flexible load resources is becoming more frequent. The accuracy and reliable challenges posed by the higher penetration of renewable energy and flexible load resources should be considered significantly. Short-term load forecast is the basis work of power grid dispatching, and it is becoming more important in the new electrical power system [1–3].

Various short-term load forecasting models and methods have been proposed and applied during the past decades. Typically, load forecasting methods are divided into statistical-based and machine learning-based forecasting methods [4–7]. Statistical methods include multiple linear regression [8], autoregressive moving average [9], and autoregression [10,11]. These methods have a distinctive feature with faster calculation speed. However, they also have a weak expressive ability for nonlinear problems. With the development of artificial intelligence technology recently, machine learning methods such as gray system [12], recurrent neural network [13], and extreme learning machine [14] are focused and widely used in load forecasting areas. However, there are gradient extinction and gradient explosion problems in the back propagation of recurrent neural network with the layer increase in the deep structure. Extreme learning machine is a combined stochastic

prediction and multiple regression method; its performance reduces with the increasing complexity of load forecasting considering various meteorological and electrical consumption behaviors. The k-fold cross-validation estimator is widely used in classification and regression problems. A global support vector regression metamodeling approach based on the sorted k-fold method is proposed in [15], in which the k number of SVR metamodels are constructed based on a cross-validation technique. A k-fold cross-validation approach based on the DCNN algorithm is used for predicting the melanoma classes in [16], and the k-fold method can achieve a better performance on a small skin cancer data set.

Furthermore, load forecasting methods based on combined algorithms emerge continuously, which have achieved results with a faster calculation speed and higher precision. A deep learning bidirectional LSTM neural network is proposed in [17]; the method gives a deep learning structure of bidirectional LSTM to avoid the gradient explosion problem. A hybrid short-term load forecasting is proposed in [18], which uses the combination of correlation analysis and appropriate inputs to the individual Bayesian neural network, but the parallel training and its weight of Bayesian neural network introduce an extra computational burden. A novel hybrid with a new signal decomposition and correlation analysis technique method is proposed in [19]; T-copula is used to compensate the peak load to reduce the forecasting error. A combined method based on the maximum information coefficient, factor analysis, gray wolf optimization and generalized regression neural network is proposed in [20]; the proposed method has a strong global search ability, but it needs enough time for the training and forecasting.

The variation of electrical price will influence the consumption of electrical power, which depends on the electrical power curves. So, the electrical price is a key point for the load forecasting. A new approach combining parametric and nonparametric function autoregressive models for the prediction of electricity price is proposed in [21]; it provides insights into the scale and purchase mechanism with the price and demand formation process. The paper provides a new solution from the purchase curve side to improve the forecasting accuracy. A load forecasting model based on least square support vector machine and chaotic time series is proposed in [22], and the real-time price of power load and chaotic characteristics is considered. A short-term load forecasting method based on price-responsive load is proposed in [23], and the response of the aggregate load to price is modeled by a set of marginal utility curves and power consumption limitations. The price response factors are estimated using a generalized inverse optimization scheme.

Power load time-series data is a non-stationary random process, which has the characteristics of quasi-periodicity, non-stationarity and nonlinearity. In order to improve the stationarization of power load time-series data, empirical mode decomposition (EMD) [24] and the frequency domain decomposition algorithm [25] are used to decompose the original load time series. A data-driven deep learning framework to load forecasting is proposed in [26], Box-Cox transformation is used to process data, and deep belief network is used for the load forecasting. A user-side load prediction method that integrates empirical mode decomposition and deep belief network is proposed in [27]; by decomposing the original load data into several eigenmode functions with different frequencies and amplitudes, the deep belief network is used to feature extraction and time-series prediction for each modal function. A load forecasting method based on ensemble empirical mode decomposition (EMMD) and Elman network is proposed in [28], the EEMD sample entropy was used to decompose the original power load sequence into a series of subsequences. A fault feature extraction method based on VMD optimized with information entropy is proposed in [29], and several intrinsic model functions were obtained by signal decomposition. Variable mode decomposition is used to generate 2D time frequency spectrograms from the various fault conditions of bearing in [30], and deep convolutional neural network is applied for fault diagnosis.

With the development of deep learning theory in the field of time-series data processing, CNNs have an outstanding performance in comparison with traditional RNN models such as LSTM. It is believed that sequence tasks should be reconsidered with RNN and

the CNN method. The CNN neural network is seen as a new solution for the sequence modeling tasks [31]. Furthermore, a temporal convolutional network (TCN) is proposed, which accelerates the feedback and convergence of the deep network through the setting of the residual unit, and it solves the degradation phenomenon caused by the increase in the network level [32]. A short-term load forecasting model for industrial customers based on the temporal convolutional network (TCN) and Light Gradient Boosting Machine (LightGBM) is proposed in [33]; the proposed method can extract the hidden information and long-term temporal relationships in the electrical load. In [34], convolution neural networks (CNNs) and gated recurrent unit (GRU) is combined for probabilistic residential load forecasting, and a deep model based on the mixture density network is used to directly predict probability density functions. An improved stacked gated recurrent unit-recurrent neural network (GRU-RNN) for both univariate and multivariate scenarios is proposed in [35]; the modified algorithm establishes an accurate mapping between the selected variables and load through its self-feedback.

In this paper, a hybrid load forecasting method based on variational modal decomposition (VMD), temporal convolutional network (TCN) and gated recurrent unit (GRU) is proposed, and the contributions of this paper are as follows: (1) Variation modal decomposition is used for the original electrical load decomposition with different frequencies. (2) A combined load forecasting method based on TCN and GRU neural network is proposed in this paper, the two neural networks are used for the forecasting of different IMF components based on its frequency characteristics, and it can extract the characteristics of electrical load in different frequency modes. (3) The simulation based on the actual electrical load is implemented in this paper; the proposed method can improve the forecast result both in accuracy and speed.

The structure of the paper is arranged as follows: The variational modal decomposition method and its application in the power load are discussed in Section 2. In Section 3, the hybrid neural network method based on GRU and TCN is introduced, and the reasons for the application of each neural network for the low and high-frequency load decomposition signal are discussed. In Section 4, two experiment data sets are used to verify the proposed method; the results show that the proposed method can catch the dynamic of electrical load in different frequency bandwidths and improve the accuracy of the electrical load forecasting.

2. Variational Modal Decomposition of Electrical Load

With the deepening of the re-electrification process of the energy system, more and more renewable generations will connect into power system. The power consumption characteristics in the user side are changing significantly. It is difficult to forecast the load profile in the power station with large-scale renewable generations. On the one hand, the output power of renewable energy is volatile, which causes fluctuation of the power load in the power station. On the other hand, the power load data contain a lot of random noise from users' individual electricity habits. For example, the yearly and daily periodicities can be seen in Figure 1.

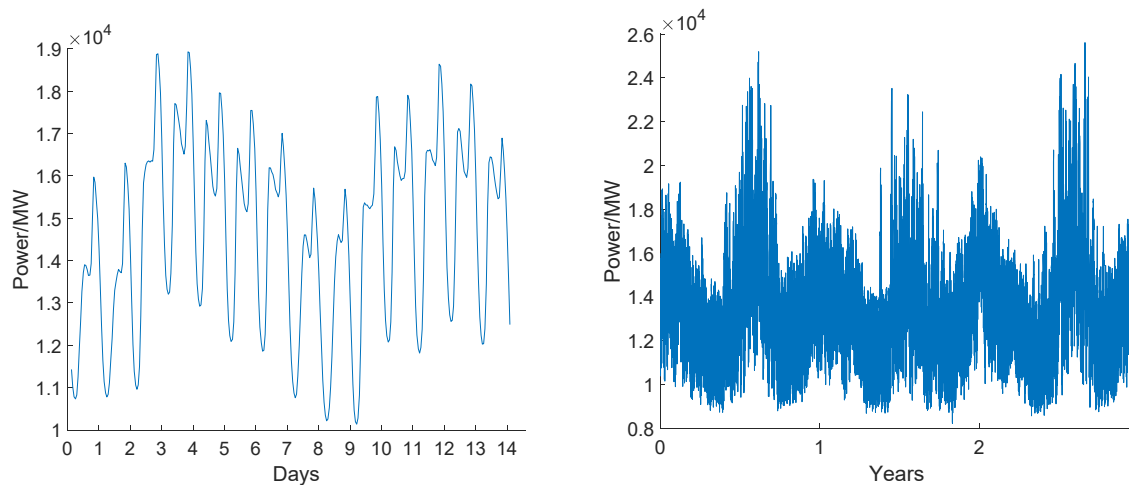


Figure 1. Yearly seasonality load profiles.

Load profile is essentially a comprehensive reflection of the dynamic response of various electrical resources in the load-side system. The principle output power of different components is similar and can be distinguished easily. It is mature to forecast a single specific component based on its physical mechanism. Therefore, the load forecast accuracy can be improved by decomposing the total load profile into a series of component levels with certain characteristics. VMD is an adaptive, non-recursive method for modal variation and signal processing. The number of modal decompositions of the signal sequence is determined by the change of the actual signal sequence. The optimal frequency center and limited bandwidth of each mode are adaptive matched and iterated to separate the intrinsic modal components (IMFs) and residuals. As a result, the frequency domain division of the power load time-series signal is obtained.

It assumes that all signal sub-components are narrowband signals which concentrated around their respective center frequencies in the VMD method. As a result, a constrained optimization problem based on the components' narrowband condition is established. The center frequencies of the different components and reconstructing their respective components is the core of VMD. It can effectively reduce the complexity of decomposition for the non-stationary random sequences compared with the EMD, and the subsequences have different frequency scales and stationary characteristics. The establishment and optimization of the Variational Problem (VP) are the two cores of the VMD method.

2.1. Variational Problem Establishment

In this paper, the alternating direction multiplier method is used to continuously update each mode and its center frequency. The corresponding fundamental frequency band of each mode is gradually demodulated to obtain the center frequency. We assume that the original signal sequence is $f(t)$, K is the number of center frequencies, and $\{u_k(t)\}$ represent the components of IMF in the limited bandwidth. Based on the Hilbert transform, the one-sided spectral signal of envelope signal $u_k(t)$ can be rewritten as $\left[\delta(t) + \frac{j}{\pi t}\right] * u_k(t)$. The frequency of each IMF is modeled to the baseband based on the center frequency ω_k . It can be written as $\left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)\right] e^{-j\omega_k t}$.

The Variational Problem (VP) based on each IMF component signal bandwidth can be written as

$$\min_{\{v_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes v_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (1)$$

$$\text{s.t.} \sum_{k=1}^K u_k = f \quad (2)$$

where $\{u_k\}$ is the IMF component, $\{\omega_k\}$ is the center frequency of each IMF component, \otimes is the convolution operator, $\delta(t)$ is the Dirac distribution, and ∂_t is the partial derivative function.

Lagrange Multiplier $\lambda(t)$ and bandwidth parameters α are introduced to turn the constrained Variational Problem into an unconstrained Variational Problem, and the extended Lagrangian function for the Variational Problem can be written as:

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \quad (3)$$

2.2. Variational Problem Solution

The IMF component $u_k(t)$ and center frequency $\omega_k(t)$ of each mode are updated based on the alternating direction multiplier method, and the new modal components can be written as:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + (\hat{\lambda}(\omega)/2)}{1 + 2\alpha(\omega - \omega_k)^2} \quad (4)$$

where, $\hat{f}(\omega)$, $\hat{u}_i(\omega)$, $\hat{\lambda}(\omega)$ and \hat{u}_k^{n+1} are the estimated value of $f(t)$, $u_t(t)$, $\lambda(t)$ and u_k^{n+1} after Fourier transform, n is the number of iterations, and ω is the frequency. The VMD calculation process is shown in Figure 2, and parameter estimation can be written as:

$$\hat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (5)$$

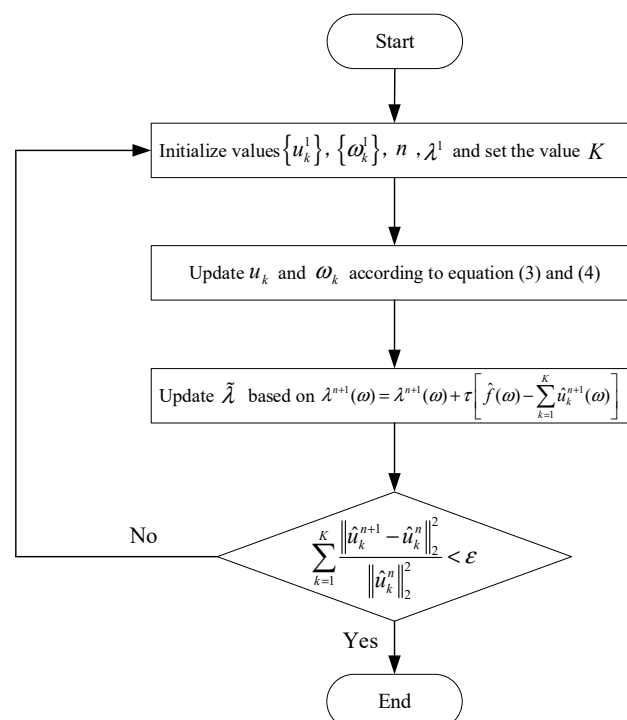


Figure 2. VMD calculation process.

3. Load Forecasting Based on GRU and TCN Hybrid Network

3.1. Gate Recurrence Unit (GRU)

A long short-term memory (LSTM) network can solve the gradient explosion and gradient disappearance problem in RNN, and LSTM has been widely used in the area of data processing in recent years. GRU is a variant of LSTM with fewer internal units; it

combines the input gate and forget gate into one update gate. It is easier to perform the training process and save computing time compared with LSTM. The substructure of GRU is shown in Figure 3; there are only two gates: namely, the update gate and the reset gate. Its mathematical description is shown in Formula (6).

$$\begin{cases} r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \\ z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \\ \tilde{h}_t = \phi(W_{\tilde{h}} \cdot [r_t \times h_{t-1}, x_t]) \\ h_t = (I - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \\ y_t = \sigma(W_o \cdot h_t) \end{cases} \quad (6)$$

where z_t and r_t are the update gate and reset gate, σ is the Sigmoid function, ϕ is the hyperbolic tangent function, x_t , h_{t-1} , h_t and y_t are the vector of input data, state memory variable in time $t - 1$ and t , and output data, respectively. W_r , W_z , $W_{\tilde{h}}$ and W_o are the weight factor of the connection matrix between x_t and h_{t-1} in the update gate, reset gate, memory unit, and output gate, respectively. I is the identity matrix. The update gate is used to control the information percentage of the previous moment when it is brought into the next moment. The reset gate controls how much information from the previous state is reserved to the current candidate set \tilde{h}_t .

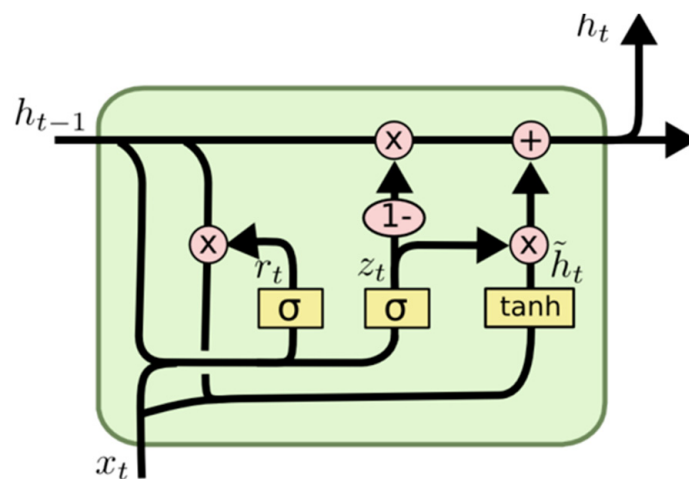


Figure 3. Basic structure of GRU.

3.2. Temporal Convolutional Network

Temporal convolutional network (TCN) is a variant of convolutional neural network (CNN). The core of TCN is atrous convolution and residual block. TCN has a larger causal convolution than traditional causal convolution with the same network depth, and it can flexibly adjust the size of the receptive field. The TCN model solves the gradient explosion and gradient disappearance problem by initializing weight parameters and regularization layers in time-series data prediction. The residual module is used to make the network information directly and deeply transmitted through the cross-layer form in TCN.

The convolution process of standard convolution and atrous convolution is shown in Figure 4. By injecting holes into the convolution, atrous convolution allows interval sampling in the input data, which increases the receptive field without going through the pooling layer. The convolution output contains a large range of information to ensure the authenticity of the information.

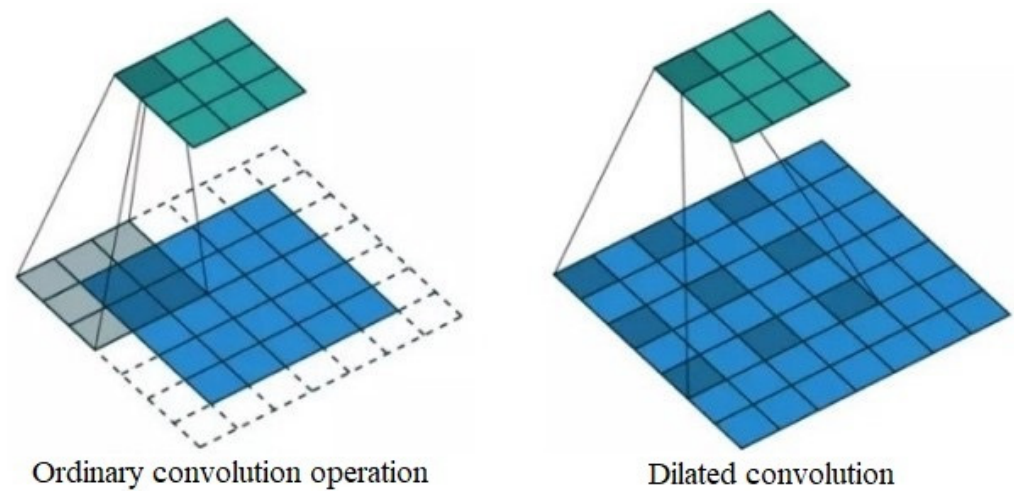


Figure 4. Comparison of CNN and TCN convolution.

The receptive field can be written as:

$$R = k_s d_{\max} s \quad (7)$$

For 1D sequence input X and filter f , atrous convolution operation F on the input sequence can be written as:

$$F(s) = \sum_{i=0}^{k-1} f(i) X_{s-d \cdot i} \quad (8)$$

where R is the size of the receptive field, k_s is the size of the convolution kernel, d_{\max} is the maximum number of hollows, s is the number of convolutional layers, d is the hollow coefficient, k is the filter size, and $s - d \cdot i$ is the direction of past data. It can be seen that the receptive field is positively correlated with the size of the convolution kernel, the maximum number of holes and the number of convolution layers.

The residual connection module uses the previous information and converted them as the current input to ensure the training efficiency; the residual connection module can be written as:

$$o = \text{Activation}(x + \chi(x)) \quad (9)$$

where o is the output of residual block, *Activation* is the activation function, x is previous information, and $\chi(x)$ is converted information.

3.3. Load Forecasting Based on VMD-GRU-TCN

After VMD decomposition, different IMF components have different frequency characteristics. In this paper, different neural networks are selected for the data training and load forecasting based on the different frequency characteristics of the load modal components. As mentioned above, the GRU network can better capture the time-series characteristics and variation rules of stationary signals, and it can obtain higher prediction accuracy when the model parameters do not change much. It is suitable for the forecasting of the low-frequency load components. The convolutional layer of TCN has both atrous convolution and causal convolution structures. Atrous convolution allows the volume to obtain different receptive fields by controlling the sampling rate; it covers a suitable range of sequence values. In addition, through the parallel performance of hole convolution, the time-series tracking of high-frequency fluctuation signals can be improved significantly. The forecasting results of the high and low-frequency load components are rebuilt for the result of the load forecast; the process of the proposed method is shown in Figure 5. The specific forecast steps are as follows:

- (1) Data preprocessing. The original load data contain abnormal data and missing data. First, the data are cleaned by DBSCAN clustering, the abnormal data are eliminated and the missing data are supplemented.
- (2) Decomposition of time-series load data. The load time series data are decomposed by VMD to obtain the different $f(t)$, $u_k(t)$, and $\lambda(t)$, corresponding to the characteristic data with different frequencies.
- (3) Load forecasting. We construct a load forecasting model based on the GRU network and TCN network, select the load data on similar days to determine the training date and carry out model training to obtain the load forecasting model; then, the load data of different components in the future day are forecasted.
- (4) Load reconstruction. The forecast data of different components are superimposed and reconstructed to obtain the forecast daily load data.

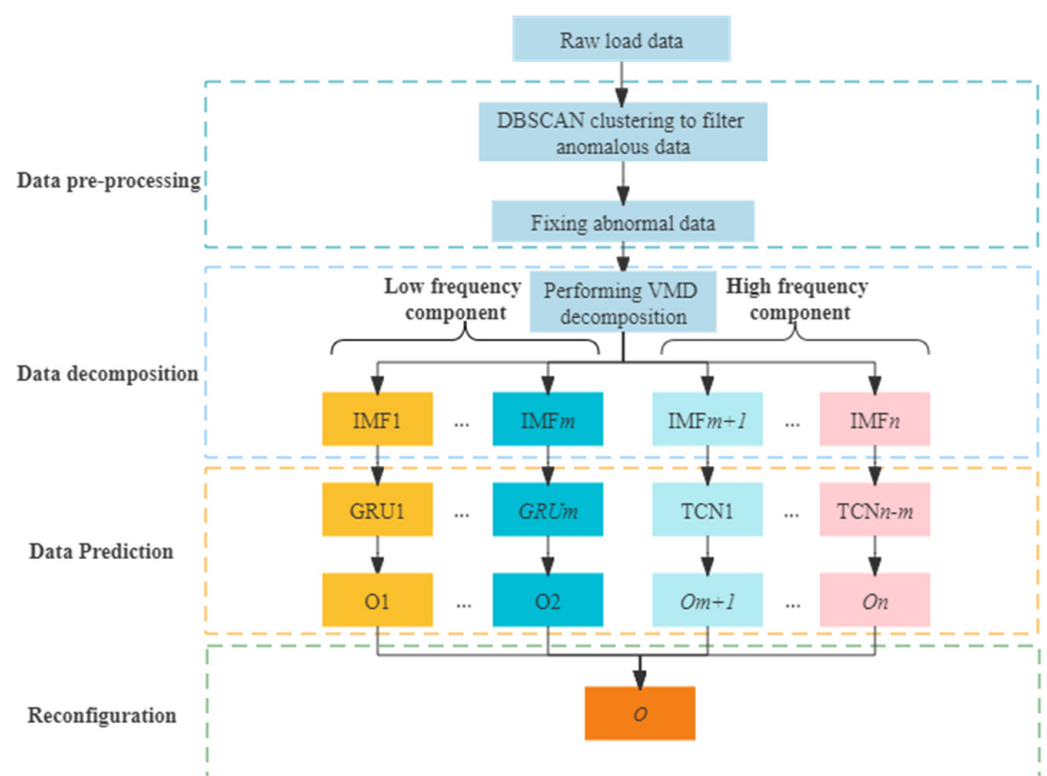


Figure 5. VMD-GRU-TCN based forecasting process.

4. Experiment Simulation and Data Analysis

4.1. Experimental Environment and Load Data

The experiment was completed using Python 3.8 n and the open source library vmdpy to write the VMD program. The neural network was implemented in Keras 2.6. The data set contains two parts in this experiment: the GEFCom2012 [36] load data and the actual load data of the power station in the southwest province in China during 1 January 2016 to 31 December 2018. The sampling time of the data set is 15 min.

Firstly, the load data are classified by the density-based clustering algorithm (DBSCAN). The clustering result is shown in Figure 6. Normal data are divided into one cluster, and the remaining clusters can be regarded as abnormal data. For anomalous data, it will be replaced by the mean of before and after data.

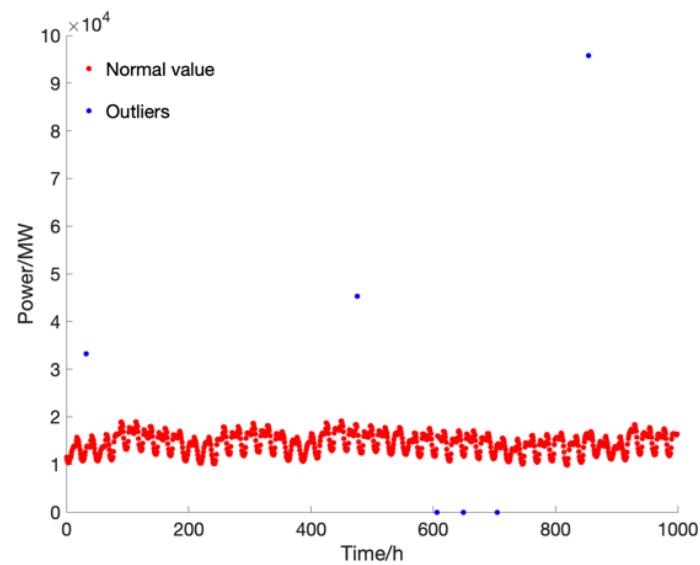


Figure 6. DBSCAN clustering diagram.

The root mean square error (RMSE) and mean absolute percentage error (MAPE) are used as the evaluation indicators in this paper. The smaller the value of evaluation indicators, the higher the forecast accuracy.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

where y_i represents the actual data, and \hat{y}_i represents the forecast data.

It is necessary to determine the value of the decomposition quantity to perform VMD decomposition. In general, the value is between 5 and 20 based on the literature. In this experiment, the center frequency value of each component is calculated with the decomposition quantity starting from 5. It can be seen that some important information in the original signal will be filtered when the decomposition quantity is too small, which will result in insufficient precision for the load forecast. However, the center frequencies of adjacent modal components will be too close, adding noise when the decomposition quantity increases.

Table 1 shows the center frequency of each IMF component when the decomposition quantity is between 5 and 10. It can be seen that the center frequency decreases when $K < 8$, which indicates that the IMF mode may be under-decomposed. When $K = 9$, IMF6 and IMF7 have similar modes, so the modal decomposition number $K = 8$ is suitable in the experiment in this paper. The dynamic profile of different component is shown in Figure 7.

Table 1. The center frequency of each IMF component under different K .

K	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10
5	0.02	83.15	41.51	288.77	165.29					
6	0.02	83.14	41.51	208.46	290.39	164.28				
7	0.01	41.51	83.14	208.44	290.29	164.29	374.00			
8	0.01	41.51	83.14	164.31	208.44	290.29	373.77	457.30		
9	0.01	41.53	83.11	164.32	208.40	290.19	327.33	374.47	457.41	
10	0.01	41.52	83.14	164.34	208.04	248.93	291.26	329.93	374.67	457.53

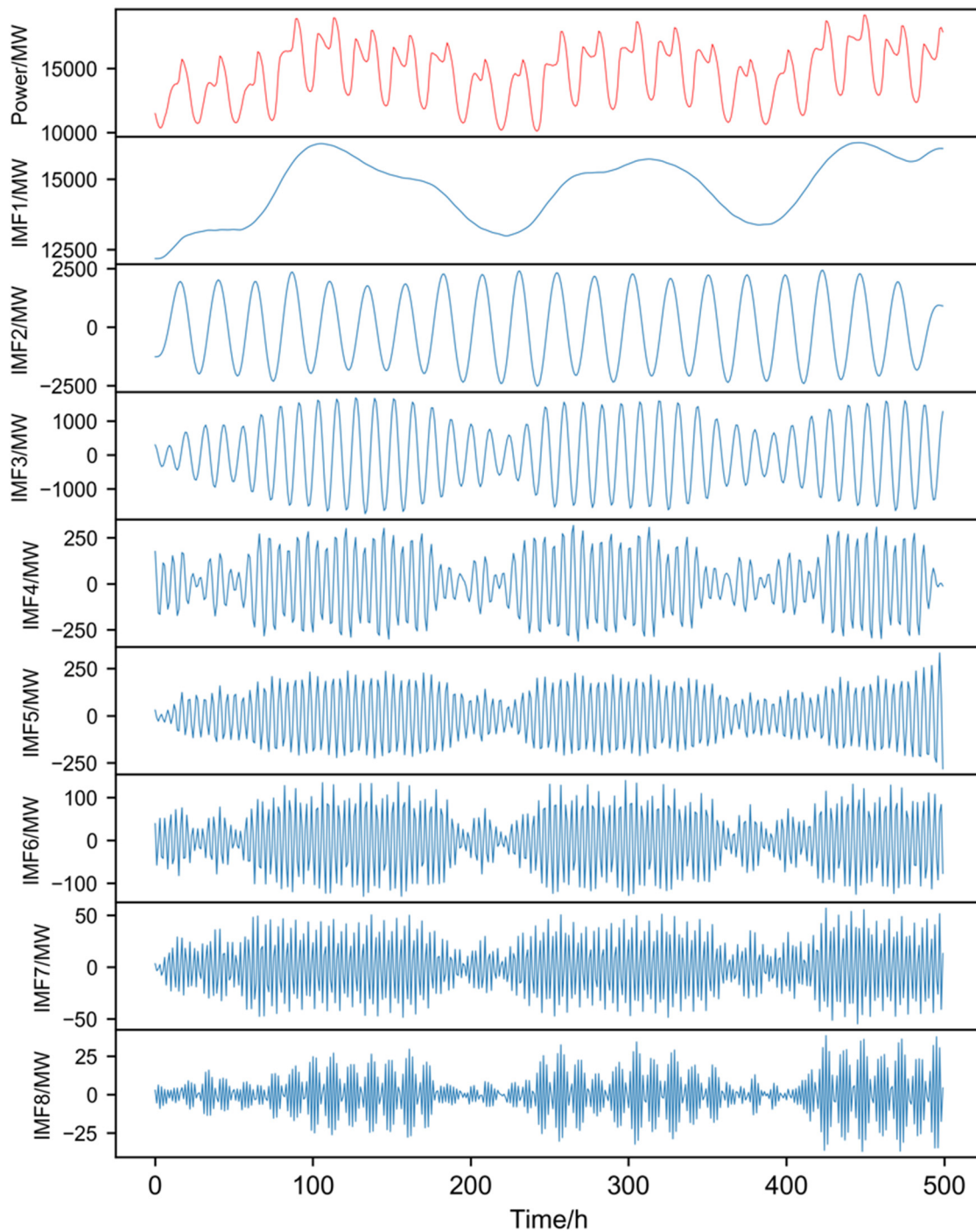


Figure 7. Modal component of VMD.

The penalty parameter α is generally set to be about twice the amount of data, and it will be 2000 in this paper. The other parameters will be set as follows: initial center frequency $\omega = 0$, and the convergence parameter is $r = 1 \times 10^{-7}$. The first 1000 pieces of data in this data set are decomposed, and the modal components diagram is shown in Figure 6. It can be seen that the modal components decomposed by VMD are smoother than the original data, which is useful and helpful for subsequent neural network training. Among them, the component IMF1 is close to the trend of the original load sequence profile, and it reflects the overall load trend.

4.2. Analysis of Complementary Characteristics of Hybrid Algorithms

In order to verify the reliability of the proposed model, the actual operation load data of a certain province are used as data set 1, the data from 0:00 on 1 January 2016 to 18:00 on 26 May 2018 are used as the training data set and a total of 50 h of data from 19:00 on 26 May to 20:00 on 28 May 2018 is used as the test data set; the sampling frequency is 1 h, and the iterative prediction is performed 6 h in advance.

The GEFCom2012 data are used as data set 2. The load data from 0:00 on 1 January 2009 to 23:45 on 25 January 2009 are used as the training data set and the load data at 23:45 on 7 February 2009 are used as the test set; the sampling frequency is 15 min. The iterative prediction is carried out 2 h in advance. The GRU network and TCN network parameters used, respectively, in the experiment are determined by training and testing. The parameters of the two neural networks are shown in Table 2.

Table 2. Parameters of neural networks.

Parameter	GRU	TCN
hidden layer	3	3
batch size	128	128
epoch	100	100
learning rate	0.01	0.01
dropout	0.1/0.2	0.1/0.2
activation function	relu	relu
dilation rate	-	1/2/4

Figure 8 shows the load forecasting results of the eight IMF components in data set 1 using the GRU model and the TCN model, respectively. For the low-frequency load components, the error between actual and forecasting data based on the GRU network is smaller than that of the TCN forecasting model. However, the prediction accuracy of GRU will decrease with the frequency increase such as IMF5 to IMF 8.

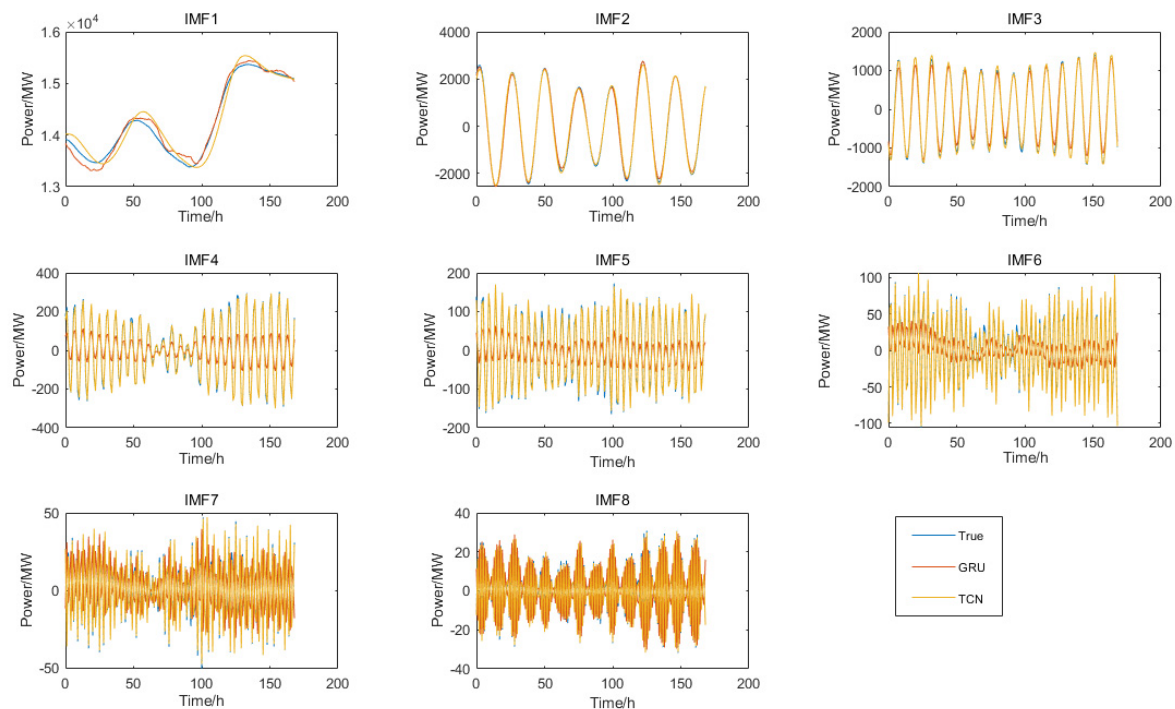


Figure 8. The forecasting results of GRU and TCN, respectively, in the high-frequency component.

Furthermore, the GRU model and the TCN model are used for the load forecasting of high-frequency components and low-frequency components in data set 2. The results are shown in Figures 9 and 10, respectively. It can be seen that in the low-frequency component, the forecasting result of the GRU model is more accurate than that of the TCN model. Figure 10 shows that the GRU model has a larger prediction error when the time-series data fluctuate too fast.

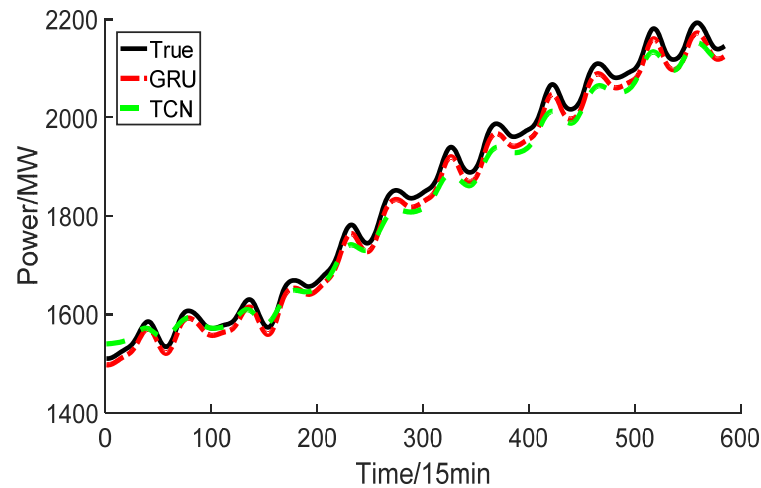


Figure 9. The forecasting results of GRU and TCN with low-frequency component.

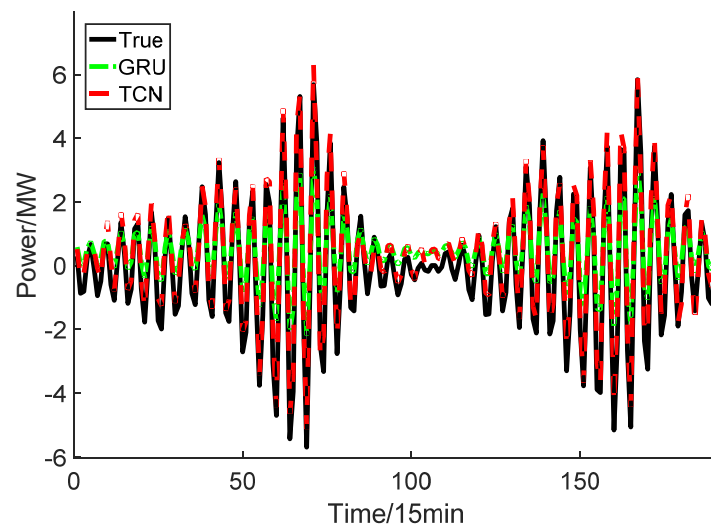


Figure 10. The forecasting results of GRU and TCN with high-frequency component.

4.3. The Comparison of Different Forecasting Models

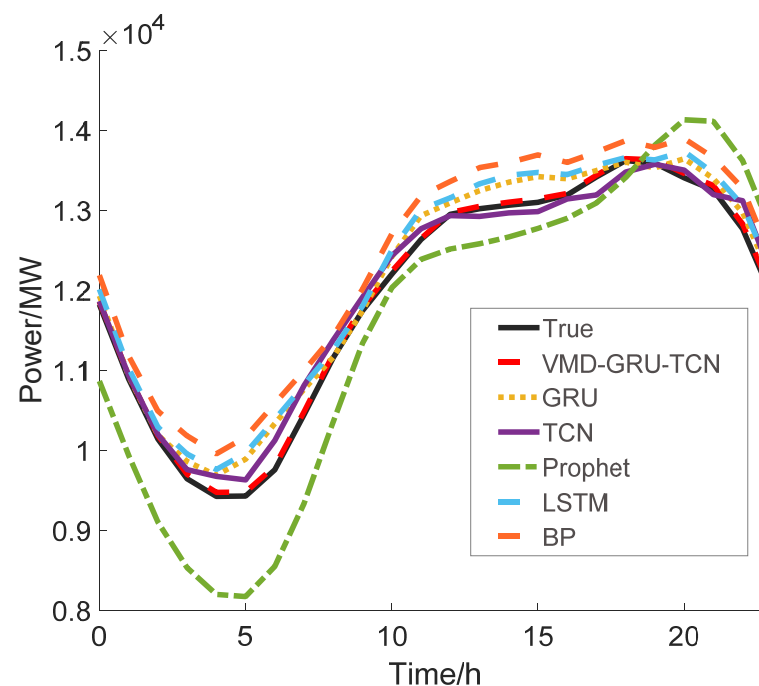
In order to verify the validity of the proposed method in this paper, GRU, TCN, LSTM, Prophet, VMD-GRU, and VMD-TCN were selected for experimental comparison, and the forecasting results are shown in Table 3. It can be seen that after the VMD decomposition, the forecasting errors of the GRU and TCN neural networks are lower than that without VMD decomposition. The performance of the forecasting is significantly improved, which indicates the feasibility of using VMD to decompose the original load sequence. Among the individual algorithms, the forecasting error of the GRU model is the smallest with the RMSE of 50.38 MW and MAPE of 1.12% in data set 1.

Table 3. Comparison of different forecasting methods.

Method	Data Set 1			Data Set 2		
	RMSE/MW	MAPE	Time/s	RMSE/MW	MAPE	Time/s
GRU	58.43	1.19	120	25.21	1.83	169
TCN	77.08	1.32	131	44.24	2.50	146
LSTM [37]	64.05	1.27	157	33.54	2.09	190
Prophet [38]	144.58	4.71	172	87.25	4.41	231
XG Boost [38]	125.23	1.68	185	36.25	2.35	199
VMD-GRU	50.38	1.12	982	11.43	1.21	1292
VMD-TCN	56.64	1.14	374	24.72	1.75	495
VMD-GRU-TCN	32.14	0.42	731	10.19	0.97	1007

Due to the high randomness and non-stationarity of the original load profile, the forecasting model constructed by a separate algorithm cannot fit the actual load situation very well. The results show that the hybrid model combined with the GRU and TCN network is the smallest both in data set 1 and data set 2. The RMSE of the forecasting result of VMD-GRU is reduced by 36.20% and 10.8%, respectively, while the MAPE of the forecasting result of VMD-GRU is reduced by 62.5% and 19.8%.

At the same time, the training and operation time of different prediction models is shown in Table 3. It can be seen that the operation time of the model proposed in this paper is slightly longer than that of the other non-decomposition prediction. The time-series load data are decomposed into eight components by VMD, and the process is serial, which increase the calculation time. However, if the different components are calculated in parallel, then the operation time will be greatly reduced. From Table 3, the TCN network itself has the ability of parallel calculation. Therefore, the computation time of the VMD-GRU-TCN model is reduced compared with the VMD-GRU model. Figures 11 and 12 show the comparison of different models based on data set 1 with 24 sample points and data set 2 with 96 sample points on the next day. It can be seen from Figures 11 and 12 that the load forecasting result has a high similarity with the original load profiles.

**Figure 11.** The forecasting result of data set 1.

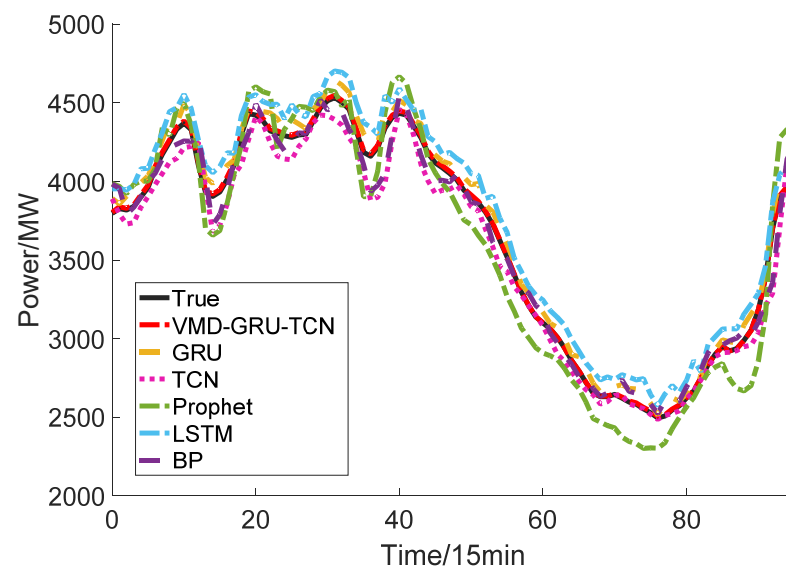


Figure 12. The forecasting result of data set 2.

5. Conclusions

This paper proposes a short-term load forecasting method based on a VMD and GRU-TCN combined network, which effectively solves the problem of low accuracy of traditional forecasting models due to the high randomness and non-stationarity of original load data. The variational modal decomposition is used to decompose the noisy original load sequence data into intrinsic IMF modal components with different frequencies and amplitudes, which can effectively reduce the complexity of the original time-series data. The gated recurrent unit (GRU) and temporal convolutional network (TCN) are used to extract, train and predict the IMF modal components. The simulation results show that the combined method has a certain degree of improvement in the prediction accuracy compared with the single prediction model.

However, there are some limitations of the proposed method in this paper. (1) The future work will focus on the coordination mechanism of the combined neural network. The forecasting results of different IMF models based on TCN and GRU will be linearly combined together for the result of electrical load forecasting, and the adaptive combined method will be a more suitable strategy. (2) From Table 3, which shows a comparison with the VMD-TCN method, the proposed model needs more time to obtain a better result. Thus, we will build a parallel computing framework for the proposed method in future work.

Author Contributions: Data curation, Y.L.; Formal analysis, T.Z.; Resources, Z.S.; Software, Y.H.; Writing—original draft, C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by “National Natural Science Foundation of China, grant number 51607057”, “The Fundamental Research Funds for the Central Universities, grant number 2020B22514” and “The open funding of Jiangsu Key Laboratory of Power Transmission & Distribution Equipment Technology, grant number 2021JSSPD07”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [\[CrossRef\]](#)
2. Taieb, S.B.; Huser, R.; Hyndman, R.J.; Genton, M.G. Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. *IEEE Trans. Smart Grid* **2016**, *7*, 2448–2455. [\[CrossRef\]](#)

3. Kong, W.C.; Dong, Z.Y.; Hill, D.J.; Luo, F.J.; Xu, Y. Short term residential load forecasting based on resident behaviour learning. *IEEE Trans. Power Syst.* **2018**, *33*, 1087–1088. [\[CrossRef\]](#)
4. Wang, Y.; Chen, Q.X.; Zhang, N.; Wang, Y.S. Conditional residual modeling for probabilistic load forecasting. *IEEE Trans. Power Syst.* **2018**, *33*, 7327–7330. [\[CrossRef\]](#)
5. Cho, M.Y.; Hwang, J.C.; Chen, C.S. Customer short term load forecasting by using ARIMA transfer function model. *Proc. IEEE Int. Conf. Energy Manag. Power Del.* **1995**, *1*, 317–322.
6. Liu, B.; Nowotarski, J.; Hong, T.; Weron, R. Probabilistic load forecasting via quantile regression averaging on sister forecasts. *IEEE Trans. Smart Grid* **2017**, *8*, 730–737. [\[CrossRef\]](#)
7. Caprio, U.D.; Genesio, R.; Pozzi, S.; Vicino, A. Short term load forecasting in electric power systems: A comparison of ARMA models and extended wiener filtering. *J. Forecast.* **2010**, *2*, 59–76. [\[CrossRef\]](#)
8. Nury, A.H.; Hasan, K.; Alam, M.J.B. Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in Northeastern Bangladesh. *J. King Saud Univ. Sci.* **2017**, *29*, 47–61. [\[CrossRef\]](#)
9. Sajjad, M.; Khan, Z.A.; Ullah, A.; Hussain, T.; Ullah, W.; Lee, M.Y.; Baik, S.W. A novel CNN-GRU-based hybrid approach for short-term residential load forecasting. *IEEE Access* **2020**, *8*, 143759–143768. [\[CrossRef\]](#)
10. He, Y.; Liu, R.; Li, H.; Wang, S.; Lu, X. Short-term power load probability density forecasting method using Kernel-based support vector quantile regression and Copula theory. *Appl. Energy* **2017**, *185*, 254–266. [\[CrossRef\]](#)
11. Li, Y.J.; Yang, Y.; Zhu, K.; Zhang, J.L. Clothing sale forecasting by a composite GRU-Prophet model with an attention mechanism. *IEEE Trans. Ind. Inform.* **2021**, *17*, 8335–8344. [\[CrossRef\]](#)
12. El-Fouly, T.H.M.; El-Saadany, E.F.; Salama, M.M.A. Grey predictor for wind energy conversion systems output power prediction. *IEEE Trans. Power Syst.* **2006**, *21*, 1450–1452. [\[CrossRef\]](#)
13. Vermaak, J.; Botha, E.C. Recurrent neural networks for short-term load forecasting. *IEEE Trans. Power Syst.* **1998**, *13*, 126–132. [\[CrossRef\]](#)
14. Li, S.; Goel, L.; Wang, P. An ensemble approach for short-term load forecasting by extreme learning machine. *Appl. Energy* **2016**, *170*, 22–29. [\[CrossRef\]](#)
15. Meng, X.Z.; Zhou, Q.; Hu, J.X.; Shu, L.S.; Jiang, P. A global support vector regression based on sorted k-fold method. In Proceedings of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 10–13 December 2017; IEEE: New York, NY, USA, 2017.
16. Nie, Y.; Santis, L.D.; Carratu, M.; Nils, M.O.; Sommella, P.; Lundgren, J. Deep Melanoma classification with K-Fold Cross-Validation for Process optimization. In Proceedings of the 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Bari, Italy, 1 June–1 July 2020; IEEE: New York, NY, USA, 2020.
17. Cai, C.C.; Tao, Y.; Zhu, T.Q.; Deng, Z.X. Short-Term load forecasting based on deep learning bidirectional LSTM neural network. *Appl. Sci.* **2021**, *11*, 8129. [\[CrossRef\]](#)
18. Ghofrani, M.; Ghayekhloo, M.; Arabali, A.; Ghayekhloo, A. A hybrid short-term load forecasting with a new input selection framework. *Energy* **2015**, *81*, 777–786. [\[CrossRef\]](#)
19. Haq, M.R.; Ni, Z. A new hybrid model for short-term electricity load forecasting. *IEEE Access* **2019**, *7*, 125413–125423. [\[CrossRef\]](#)
20. Ge, L.J.; Xian, Y.M.; Wang, Z.G.; Gao, B.; Chi, F.; Sun, K. Short-term load forecasting of regional distribution network based on generalized regression neural network optimized by grey wolf optimization algorithm. *CSEE J. Power Energy Syst.* **2021**, *7*, 1093–1101.
21. Shah, I.; Lisi, F. Forecasting of electricity price through a functional prediction of sale and purchase curves. *J. Forecast.* **2020**, *39*, 242–259. [\[CrossRef\]](#)
22. Wang, S.Y.; Ji, T.Y.; Li, M.S. Power load forecasting of LSSVM based on real time price and chaotic characteristic of loads. In Proceedings of the 2020 IEEE Sustainable Power and Energy Conference (iSPEC), Chengdu, China, 23–25 November 2020; IEEE: New York, NY, USA, 2020.
23. Gallego, J.S.; Morales, J.M. Short term forecasting of price-responsive loads using inverse optimization. *IEEE Trans. Smart Grid* **2018**, *9*, 4805–4814. [\[CrossRef\]](#)
24. Liu, Q.Z.; Shen, Y.B.; Wu, L.; Li, J.; Zhuang, L.R.; Wang, S.F. A hybrid FCW-EMD and KF-BA-SVM based model for short-term load forecasting. *CSEE J. Power Energy Syst.* **2018**, *4*, 226–237. [\[CrossRef\]](#)
25. Dong, W.C.; Sun, H.X.; Li, Z.; Zhang, J.X.; Yang, H.F. Short-term wind-speed forecasting based on multiscale mathematical morphological decomposition, K-Means clustering, and stacked denoising autoencoders. *IEEE Access* **2020**, *8*, 146901–146914. [\[CrossRef\]](#)
26. Ouyang, T.H.; He, Y.S.; Li, H.J.; Sun, Z.Y.; Baek, S. Modeling and forecasting short-term power load with copula model and deep belief network. *IEEE Trans. Emerg. Top. Comput. Intell.* **2019**, *3*, 127–136. [\[CrossRef\]](#)
27. Yu, Y.; Ji, T.Y.; Li, M.S.; Wu, Q.H. Short-term load forecasting using deep belief network with empirical mode decomposition and local predictor. In Proceedings of the 2018 IEEE Power & Energy Society General Meeting (PESGM), Portland, OR, USA, 5–10 August 2018.
28. Song, L.Q.; Xie, Q.Y.; He, Y.K.; Dang, P.J. Ultra-short-term wind power combination forecasting model based on MEEMD-SAE-Elman. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020.

29. Yang, J.Z.; Zhou, C.J.; Li, X.F. Research on fault feature extraction method based on parameter optimized variational mode decomposition and robust independent component analysis. *Coatings* **2022**, *12*, 419. [[CrossRef](#)]
30. Akhenia, P.; Bhavsar, K.; Panchal, J.; Vakharia, V. Fault severity classification of ball bearing using SinGAN and deep convolutional neural network. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2021**, *236*, 3864–3877. [[CrossRef](#)]
31. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Proc. Conf. Adv. Neural Inform. Process. Syst.* **2015**, *28*, 91–99. [[CrossRef](#)]
32. Luo, F.; Poslad, S.; Bodanese, E. Temporal convolutional networks for multiperson activity recognition using a 2-D LIDAR. *IEEE Internet Things J.* **2020**, *7*, 7442–7462. [[CrossRef](#)]
33. Wang, Y.Y.; Chen, J.; Chen, X.Q.; Zeng, X.J.; Kong, Y.; Sun, S.F.; Guo, Y.S.; Liu, Y. Short-term load forecasting for industrial customers based on TCN-LightGBM. *IEEE Trans. Power Syst.* **2020**, *36*, 1984–1997. [[CrossRef](#)]
34. Afrasiabi, M.; Mohammadi, M.; Rastegar, M.; Stankovic, L.; Afrasiabi, S.; Khazaei, M. Deep-Based conditional probability density function forecasting of residential loads. *IEEE Trans. Smart Grid* **2020**, *11*, 3646–3657. [[CrossRef](#)]
35. Xia, M.; Shao, H.D.; Ma, X.D.; Silva, C.W.D. A Stacked GRU-RNN-Based approach for predicting renewable energy and electricity load for smart grid operation. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7050–7059. [[CrossRef](#)]
36. Hong, T.; Fan, S. Guest Editorial: Special section on analytics for energy forecasting with applications to smart grid. *IEEE Trans. Smart Grid* **2014**, *5*, 399–401. [[CrossRef](#)]
37. Zhao, J.T.; Huang, J.C. Application and research of short-term power load forecasting based on LSTM. In Proceedings of the 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Xi'an, China, 14–16 August 2020; IEEE: New York, NY, USA, 2020.
38. Gupta, R.; Yadav, A.K.; Jha, S.K.; Pathak, P.K. Time Series Forecasting of Solar Power Generation Using Facebook Prophet and XG Boost. In Proceedings of the 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 11–13 February 2022; IEEE: New York, NY, USA, 2022.