

Article

Robust Wavelet Transform Neural-Network-Based Short-Term Load Forecasting for Power Distribution Networks

Yijun Wang ¹, Peiqian Guo ^{2,*} , Nan Ma ¹ and Guowei Liu ^{1,*}

¹ Shenzhen Power Supply Co., Ltd., Shenzhen 518020, China

² Department of Electrical Engineering, Tsinghua University, Beijing 100083, China

* Correspondence: guopeiqian@tsinghua.edu.cn (P.G.); liuguowei@sz.csg.cn (G.L.)

Abstract: A precise short-term load-forecasting model is vital for energy companies to create accurate supply plans to reduce carbon dioxide production, causing our lives to be more environmentally friendly. A variety of high-voltage-level load-forecasting approaches, such as linear regression (LR), autoregressive integrated moving average (ARIMA), and artificial neural network (ANN) models, have been proposed in recent decades. However, unlike load forecasting in high-voltage transmission systems, load forecasting at the distribution network level is more challenging since distribution networks are more variable and nonstationary. Moreover, existing load-forecasting models only consider the features of the time domain, while the demand load is highly correlated to the frequency-domain information. This paper introduces a robust wavelet transform neural network load-forecasting model. The proposed model utilizes both time- and frequency-domain information to improve the model's prediction accuracy. Firstly, three wavelet transform methods, variational mode decomposition (VMD), empirical mode decomposition (EMD), and empirical wavelet transformation (EWT), were introduced to transform the time-domain demand load data into frequency-domain data. Then, neural network models were trained to predict all components simultaneously. Finally, all the predicted data were aggregated to form the predicted demand load. Three cases were simulated in the case study stage to evaluate the prediction accuracy under different layer numbers, weather information, and neural network types. The simulation results showed that the proposed robust time–frequency load-forecasting model performed better than the traditional time-domain forecasting models based on the comparison of the performance metrics, including the mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE).

Keywords: distributed network; load forecasting; wavelet transform; neural network; smart grid



Citation: Wang, Y.; Guo, P.; Ma, N.; Liu, G. Robust Wavelet Transform Neural-Network-Based Short-Term Load Forecasting for Power Distribution Networks. *Sustainability* **2023**, *15*, 296. <https://doi.org/10.3390/su15010296>

Academic Editors: Pei Zhang, Zhao Liu and Zhenhuan Ding

Received: 30 November 2022

Revised: 20 December 2022

Accepted: 20 December 2022

Published: 24 December 2022



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

To deal with the severe climate change crisis, 193 countries agreed to adopt a set of global goals to end poverty, protect the planet, and ensure prosperity for all. Of particular interest is Goal 7, which aims to provide access to clean and affordable energy. The distribution network in a power system plays the role of an intermediate that links urban areas to the transmission/sub-transmission network. The distribution network is influenced by many factors, such as the power quality; the peak load, resistance, and reactance of the distribution lines; and the distance between the transformer and the consumers (including residential, commercial, and industrial consumers). Precise distribution-level short-term load forecasting (STLF) is a vital process that helps grid operators pursue this goal since a reliable STLF system provides critical input information for demand-side management (DSM), state estimation, maintenance scheduling, voltage support, etc. Moreover, providing the precise and rapid prediction of future demand is the foundation of hourly based applications, such as electricity market-clearing mechanisms and regulation bids. However, unlike high-voltage (HV) transmission networks, which are stable and remain unchanged for long periods, STLF for distribution networks have become more

challenging in recent years. The reasons for this can be summarized as follows: (1) The original distribution networks are designed as passive systems, as the electricity can only be transmitted from power plants to consumers. However, the networks have turned from passive into active bidirectional systems since more and more distribution generators (DGs) have penetrated the grid. (2) High uncertainty surrounds the low-capacity load inside distribution networks and the diversity of users' characteristics also influences load prediction accuracy. Compared to the well-developed load-forecasting procedures at the HV network level, distribution-level load-forecasting methods are still in the exploratory stage. Fortunately, a large amount of high-resolution electricity data has been collected with the widespread installation of smart meters. These aggregated smart meter data represent a useful data resource on the demand side to more accurately forecast the demand at the distribution level.

Load-forecasting methods have been well-discussed in the literature over the last few decades. Depending on the prediction duration, load forecasting can be further divided into STLTF, mid-term load forecasting, and long-term forecasting. Of all the load-forecasting categories, STLTF plays the most vital role in power systems due to its importance for power stability and economic performance. STLTF methods can be classified into conventional statistical methods and machine-learning-based methods. Traditional statistical methods try to establish linear equations to predict the demand load, with linear regression (LR) and autoregressive integrated moving averages (ARIMA) being the most common statistical methods. Distribution-level LR-based STLTF approaches are presented in [1–4]. As a naive forecasting approach, LR (or its modification, multivariable LR) aims to identify the linear correlations between the input features and the future demand load; the loss function employed by most LR methods is the least squares method, which minimizes the Euclidean distance between the predicted and true values. ARIMA is another statistical analysis model that uses time-series data to predict the future trends of datasets. It first applies lagged moving averages to smooth the demand load data. Then, it predicts the future demand load based on the assumption that the future trend will resemble the past trends [5]. The ARIMA algorithm is combined with other statistical methods to improve prediction accuracy. The methods that have been employed in the literature to enhance the accuracy of the ARIMA model include XGBoost [6], support vector machine (SVM) [7], decision trees (DTs) [8], and the hidden Markov model (HMM) [9]. However, the characteristics of the distribution-level demand load are nonlinear and nonstationary. Additionally, due to the high level of the incorporation of distributed renewable energy sources (such as solar panels and small wind turbines) and electric vehicles (EVs) into distribution networks in recent decades, the fluctuation of load curves has become increasingly extreme and unpredictable; hence, it is difficult to use a linear function to exactly estimate the demand load.

Machine-learning-based STLTF methods have shown higher prediction accuracy than traditional statistical methods, including artificial neural networks (ANNs), random forests, and SVM. ANN-based methods have achieved huge success in power system applications, such as STLTF [10], solar energy forecasting [11], energy management [12], and abnormal data detection [13]. The authors of [14] proposed a three-layer ANN STLTF predictive model. However, the proposed STLTF model was just a naive network that could only consider a single data point at a time and it could not apply the experiences learned from the historical training process. The improved ANN-based STLTF methods can be further divided into feedforward neural networks [10,15,16], backpropagation neural networks [17–19], recurrent neural networks [10,16], the restricted Boltzmann machine [20], and convolutional neural networks [21–23]. RNNs have achieved particularly high accuracy in recent decades. RNN models contain memory units that can study present input information and data from the past. This characteristic is very useful for time-series tasks such as forecasting. RNN models can be further classified as long short-term memory (LSTM) and gated recurrent unit (GRU) models. In [24], a LSTM-based STLTF model is utilized to predict short-term residential load. Firstly, Density-Based Spatial Clustering of Application with Noise (DBSCAN) is employed to cluster the load profiles into several groups and an LSTM

model is developed to forecast the load of each group. Four input vectors—the sequence of energy consumptions, the series of the time day indices, the corresponding day of week indices, and the corresponding binary holiday marks—are fed into the model to improve the accuracy further. Although RNN can map nonlinear features such as conventional approaches, this approach also has drawbacks: (1) RNN can only analyze information at the time domain and is insensitive to the frequency information, while the demand load is a combination of electricity appliances operating with different frequencies. (2) Moreover, the relevant weather information should also be fed as an important input feature.

Decomposition methods include empirical mode decomposition (EMD) [25], variational mode decomposition (VMD) [26,27], seasonal and trend decomposition using Loess decomposition (STL decomposition) [28], and empirical wavelet transforms (EWT) [29,30]. EMD-based STLF methods are introduced in [25]. As an adaptive nonlinear decomposition method, EMD decomposes the original signal into a series of intrinsic mode functions (IMFs) using Hilbert–Huang transform and each IMF is an amplitude modulation–frequency modulation (AM-FM) signal [31]. However, as a purely data-driven method, EMD lacks a mathematical definition, so it is difficult to understand the decomposition results; secondly, the decomposed signals will diverge at the endpoints and are highly sensitive to noise [12]. VMD-based STLF methods are presented in [27,32]. As an alternative algorithm to EMD, VMD is a non-recursive, adaptive decomposition estimation method to decompose the original signal into several mode functions with specific bandwidths in the frequency domain [33]. The latest decomposition algorithm, EWT, combines the strength of the wavelet’s mathematical definition with the flexibility of EMD [29].

However, as discussed above, it is vital to create a precise prediction for the distribution network. The high uncertainty of the distribution network and the strict limitation of the traditional time-domain STLF model are the main barriers for researchers to improve the accuracy further. Hence, this paper proposes a hybrid STLF method that can extract both time-domain and frequency-domain features with high adaptivity. Time–frequency transformation approaches, such as the Fourier transform, wavelet transform (WT), and least-squares wavelet (LSW) [34], can effectively convert the data from the time domain into the frequency domain. The literature proposes a hybrid STLF approach that combines the WT and ANN [25,27,29–31]. In these works, the original time-series demand load is transferred into the frequency-domain data, then a series of neural networks are trained simultaneously to forecast each frequency component. Finally, the predicted values of all the components are combined to create the overall prediction of the demand load. However, there is a lack of work to take advantage of all the WT algorithms, which can improve the prediction accuracy further. As illustrated, a wealth of work is available in the literature and the existing STLF models still have some knowledge gaps that can be filled.

- (1) Although STLF has been fully investigated in transmission networks and at the household-level, distribution-level STLF is a relatively weak segment in current power systems.
- (2) A new hybrid STLF that takes advantage of Variational Mode Decomposition (VMD), Empirical Mode Decomposition (EMD), and Empirical Wavelet Transform (EWT) should be proposed.

The remaining paper is organized as follows: The relevant knowledge regarding datasets, wavelet transform, recurrent neural networks, and the model structure is introduced in Section 2. In Section 3, three case studies are implemented, which compare the proposed load forecasting algorithm and other methods and evaluate the parameters that achieve the best performance. Section 4 discusses the result obtained from Section 3. The conclusion and final discussion are provided in the last section.

2. Materials and Methods

2.1. Data Description

The dataset employed in this paper includes distribution-level electricity data, which is constructed by combining household-level smart meter data and weather and temporal data.

2.1.1. Distribution-Level Electricity Data

In this paper, the distribution network level data are obtained from the physical/informatic aggregator and the individual household-level smart meter data from Pecan Street Dataport (Dataport) are added up to match the capacity of the feeder model. The geographical location of the elasticity data is N 30°15'59.9976", W 7°43'59.9880" (Austin, TX, USA). The feeder models used for this research are selected from standard feeder models provided by GridLAB-D 4.3. In this work, 976 houses are aggregated to match the R5-12.47-2 feeder model, indicating a moderate suburban area (demand capacity is 4500 kW). To reach the defined aggregation size, household smart meter data are picked up randomly from the Dataport dataset; an example of the demand load is shown in Figure 1.

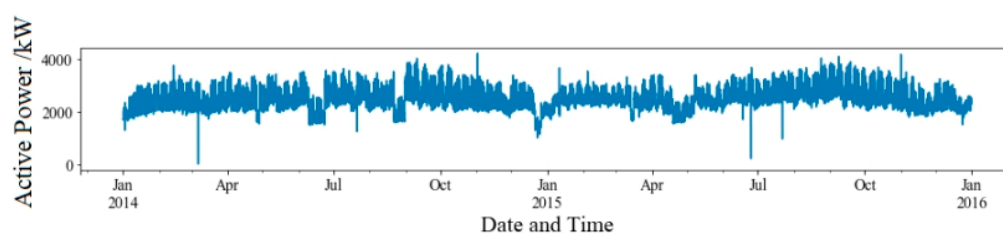


Figure 1. Active power of the distribution-level electricity data.

2.1.2. Weather and Temporal Information

The corresponding weather and temporal information at the same location are obtained from the National Solar Radiation Database (NSRDB). An example of the dataset is shown in Table 1 and Figure 2; the weather parameters include Dew point (°C), Temperature (°C), Pressure (Pa), and Relative Humidity (%RH). As for the temporal information, four variables are introduced, which are: Holiday (1 for holiday days and 0 for non-holiday days), Hour of the Day (HOD) (index range from 0 to 23), Day of the Week (DOW) (index range from 0 to 6), and Month of the Year (MOY) (index ranges from 1 to 12). As categorical variables, DOY, HOD, DOW, and MOY should be pre-processed by one-hot encoding.

Table 1. Shows the weather and temporal dataset.

Timestamp	Holiday	HOD	DOW	MOY	Dew. Point (°C)	Temperature (°C)	Pressure (Pa)	Relative Humidity (%RH)
1 January 2014 00:00	1	0	3	1	−1.25931	1.801934814	1001.035	80.1306
1 January 2014 01:00	1	1	3	1	−1.25199	1.385064697	1000.496	82.60188
1 January 2014 02:00	1	2	3	1	−1.25886	1.022241211	999.9987	84.73964
1 January 2014 03:00	1	3	3	1	−1.26002	0.723382568	999.3622	86.57533
1 January 2014 04:00	1	4	3	1	−1.23658	0.513696289	998.8751	88.04627
1 January 2014 05:00	1	5	3	1	−1.19007	0.407220459	998.4365	89.02894
1 January 2014 06:00	1	6	3	1	−1.09016	0.479547119	998.138	89.215
1 January 2014 07:00	1	7	3	1	−0.80314	1.532342529	998.006	84.46244

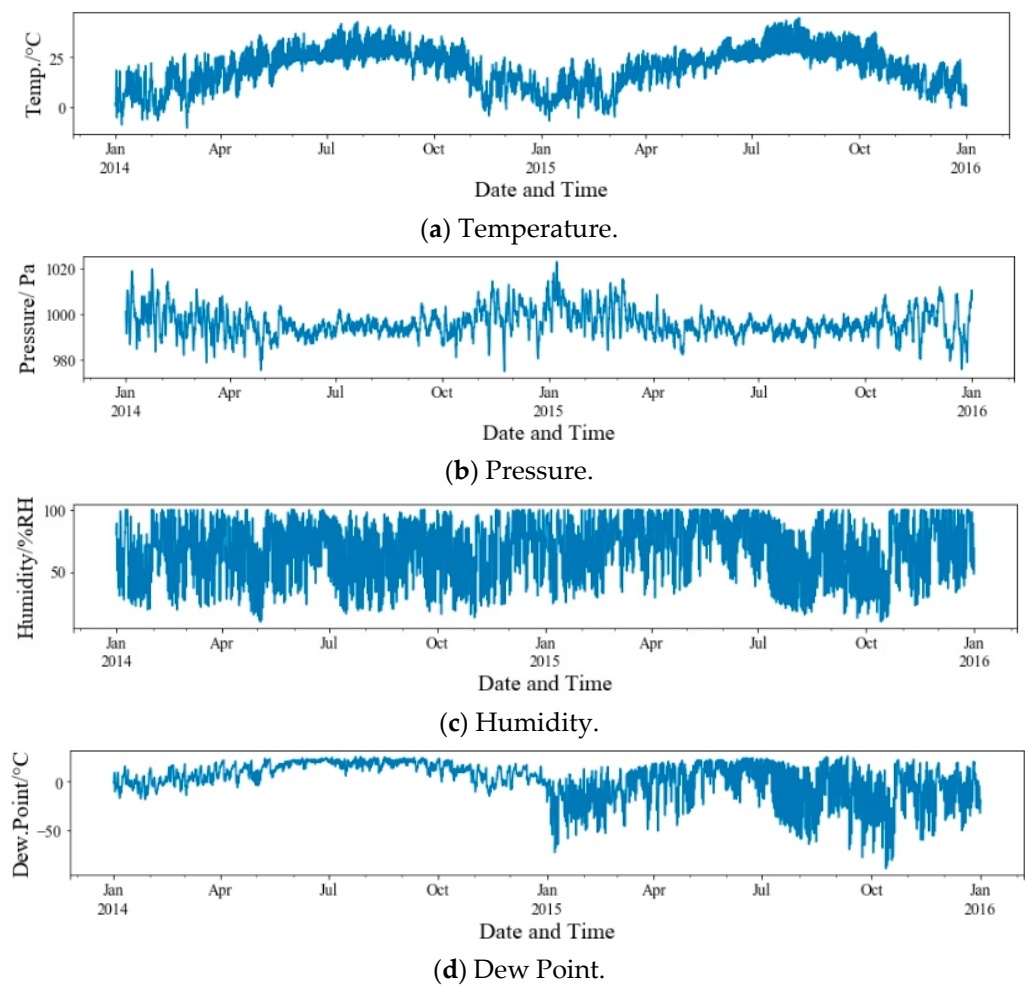


Figure 2. Visualization of the weather variables.

2.2. Methods

2.2.1. Empirical Mode Decomposition

EMD is a self-adaptive mode decomposition method and it can decompose the signal in temporal space directly without transferring it into frequency space (see Figure 3). The characteristic of EMD is that it does not rely on any mathematical functions but adopts the signal $f(t)$ accordingly; $f(t)$ is decomposed into $N + 1$ Intrinsic Mode Functions (IMFs) $f_k(t)$ and a residuum $r(t)$, see (1).

$$f(t) = \sum_{k=0}^N f_k(t) + r(t) \quad (1)$$

An IMF is an AM—FM function and can be expressed as follows:

$$f_k(t) = F_k(t) \cos(\varphi_k(t)) \text{ where } F_k(t), \varphi'_k(t) > 0 \forall t \quad (2)$$

The main assumption is that $F_k(t)$ and $\varphi'_k(t)$ varies much slower than $\varphi_k(t)$. The detailed process is presented in Algorithm 1. An IMF should satisfy two conditions: (1) the number of extrema and the number of zero crossings must be the same or differ at most by one; (2) at any point, the mean value of upper (defined by local maxima) and lower envelope (defined by local minima) is zero. The detailed process of EMD is demonstrated in Algorithm 1.

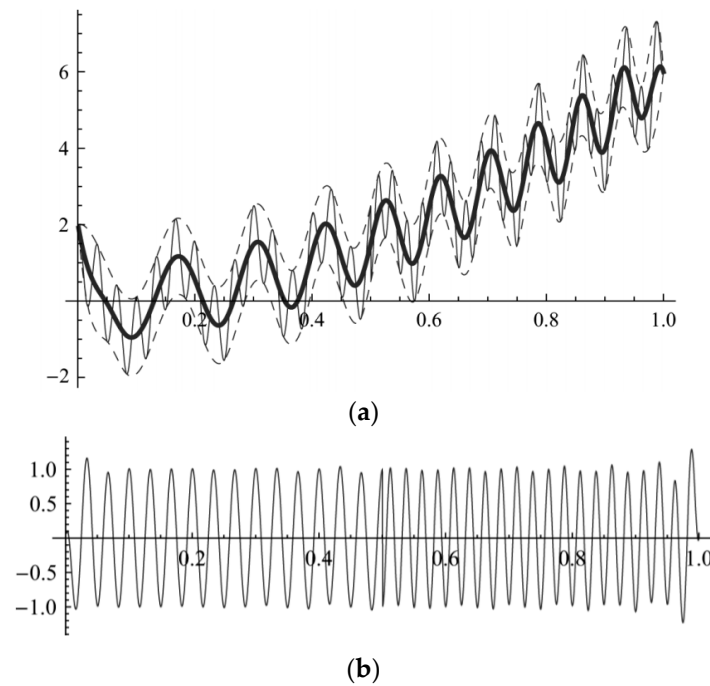


Figure 3. (a) EMD: basic IMF detection; (b) the first IMF candidate.

Algorithm 1: Empirical Mode Decomposition (EMD).

Input: Real-world signal $f(t)$.

Output: IMFs f_k , where $k = 1, 2, \dots, N + 1$.

Initialization: $n := 1, r_0(t) = f(t)$.

Step 1: Extract the n th IMF as follows:

(a): Initialize $h_0(t) := r_{n-1}(t)$ and $k := 1$.

(b): Detect the maxima and minima of $h_{k-1}(t)$.

(c): Compute the upper and lower envelope, $U_{k-1}(t)$ and $L_{k-1}(t)$ by a cubic spline interpolation from the maxima and minima (See Figure 3a).

(d): Compute the mean envelope: $m_{k-1}(t) = \frac{U_{k-1}(t) + L_{k-1}(t)}{2}$.

(e): Obtain the candidate component: $h_k(t) := h_{k-1}(t) - m_{k-1}(t)$ (See Figure 3b).

(f): If $r_1(t)$ satisfies conditions of an IMF:

(i): $x_n(t) := h_k(t)$ and $r_n(t) := r_{n-1}(t) - x_n(t)$.

(g): Else:

(i): $k := k + 1$.

(ii): Repeat steps b)-g) until $h_k(t)$ is an IMF.

Step 2: If $r_n(t)$ is a residuum, stop the process.

Else $n := n + 1$ and start from Step 1.

2.2.2. Variational Mode Decomposition

VMD was proposed by K. Ragomiretskiy in 2014 [33]; it is a non-recursive, adaptive decomposition estimation method to decompose the original signal into K mode functions $u_k(t)$ with specific bandwidth in the frequency domain. Additionally, each $u_k(t)$ is concentrated near the central frequency ω_k . The nature of the VMD method is an optimal process to look for K modes that cause the overall bandwidth to be the smallest, as shown in Algorithm 1:

Already modified by Figure 3a,b.

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|^2 \right\} \text{ s.t. } \sum_{k=1}^K u_k = f(t) \quad (3)$$

where $f(t)$ is the original signal, $\delta(t)$ represents the Dirac distribution function; $\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)$ is the corresponding unilateral spectrum of $u_k(t)$ by implementing the Hilbert transformation; u_k and ω_k represents the k th mode and corresponding central frequency; and $e^{-j\omega_k t}$ is the exponent term to adjust the frequency spectrum to the corresponding base frequency band.

Then, by introducing a quadratic penalty α and Lagrange multiplier operator $\lambda(t)$, the constrained problem mentioned in (3) is transformed into a non-constrained problem, the augmented Lagrangian expression is expressed as:

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \|f(t) - \sum_{k=1}^K u_k(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \rangle \quad (4)$$

where α is adopted to ensure the accuracy of the reconstruction; $\lambda(t)$ is employed to tighten the constraint; and $\|f(t) - \sum_{k=1}^K u_k(t)\|_2^2$ is a quadratic penalty term to speed up the convergence. The expression (4) can be solved by employing the alternate direction method of multipliers (ADMM) to compute the saddle point of the equation. According to the ADMM optimization method, u_k and are updated as follows:

$$u_k^{n+1} = \underset{u_k}{\operatorname{argmin}} L(\{u_i^{n+1}\}, \{\omega_i^n\}, \{\omega_i^n\}, \{\lambda^n\}) \quad (5)$$

$$\omega_k^{n+1} = \underset{\omega_k}{\operatorname{argmin}} L(\{u_i^{n+1}\}, \{\omega_i^{n+1}\}, \{\omega_i^n\}, \{\lambda^n\}) \quad (6)$$

$$\lambda^{n+1} = \lambda^n + \tau(f(t) - \sum_{k=1}^K u_k^{n+1}) \quad (7)$$

$$\sum_{k=1}^K \|u_k^{n+1} - u_k^n\|_2^2 / u_k^n^2 < \varepsilon \quad (8)$$

Finally, u_k^{n+1} and ω_k^{n+1} are solved as:

$$u_k^{n+1}(\omega) = \frac{f(\omega) - \sum_{i \neq k} u_i(\omega) + \frac{\lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (9)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |u_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |u_k^{n+1}(\omega)|^2 d\omega} \quad (10)$$

where $f(\omega)$, $\lambda(\omega)$, $u_i(\omega)$, and $u_k^{n+1}(\omega)$ represent the Fourier transform of $f(t)$, $\lambda(t)$, $u_i(t)$, and $u_k^{n+1}(t)$.

2.2.3. Empirical Wavelet Transforms (EWT)

After the data are denoised via DWT, the denoised data $f(t)$ are decomposed into N sub-layers via EWT. The EWT aims to extract multiple sub-layers by constructing adaptive wavelets. The EWT decomposition process is performed in the following steps. In EWT decomposition, the number of sub-layers N is defined at the beginning.

Step 1: Apply fast Fourier transform (FFT) to the denoised data $f(t)$ to obtain the frequency spectrum $F(\omega)$.

Step 2: Search the $F(\omega)$ to find the N local maxima $\partial = \{\partial_n\}_{n=1,2,\dots,N}$ and the corresponding frequencies $\omega = \{\omega_n\}_{n=1,2,\dots,N}$ by using the magnitude threshold α and frequency distance thresholds δ . α is set as 3% of the fundamental magnitude to detect the significant frequencies and δ is set as 8 Hz to avoid overestimation.

Step 3: Segment the frequency spectrum $[0, f_{sample}/2]$ into N segments and the boundaries Ω_n is the center line between two neighboring local maxima (see Figure 4), which can be calculated as:

$$\Omega_n = \frac{\omega_n + \omega_{n+1}}{2} \quad (11)$$

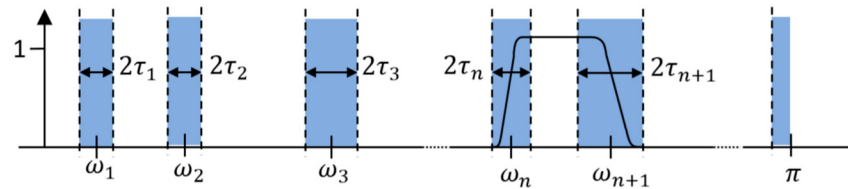


Figure 4. Segmenting Fourier spectrum into N contiguous segments (Adopted from [35]).

Step 4: Build N wavelet filters, including one low-pass filter and $N - 1$ band-pass filters based on the defined boundaries. The scaling and wavelet functions are defined in (12) and (13), respectively. A comparison between the demand load data at the time- and frequency-domain is shown in Figure 5; from the figure, it is observed that the power spectrums of the load demand not only contain time information but also indicate the frequency components.

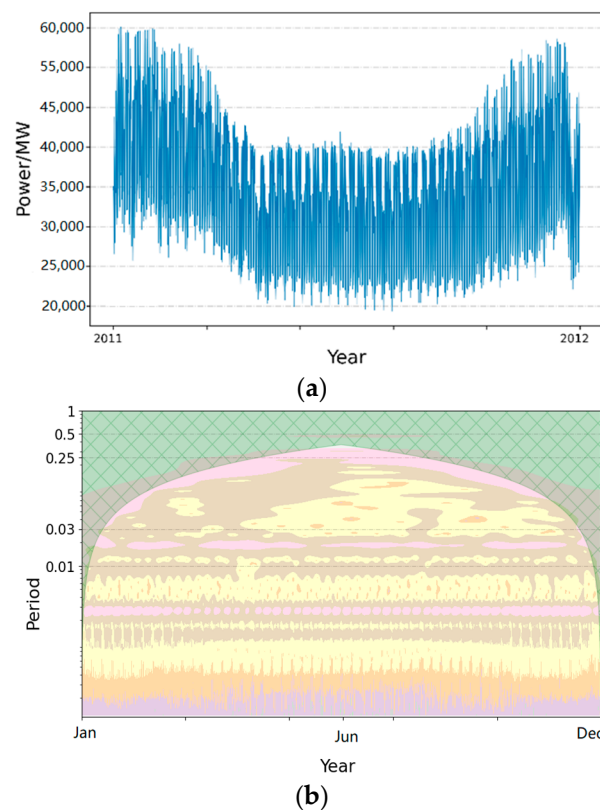


Figure 5. (a) Load demand curve; (b) Power spectrums of the load demand.

2.2.4. Recurrent Neural Network

The LSTM model was first proposed in 1997. As shown in Figure 6, in LSTM, the hidden state in traditional RNN is replaced by the memory cell C_t and three gates, i.e., the input gate I_t , the forget gate F_t , and the output gate O_t . The output of the previous time step h_{t-1} and the input sequence of the current time step X_t are adopted as the input of the gates. The sigmoid activation function σ controls these gates ($\cdot$): the information is preserved when the activation output is close to 1 and the information is eliminated when the activation output approaches 0. As for the memory cell C_t , a candidate memory cell $\tilde{C}_t$

is computed at first. The only difference between $\tilde{C}_t$ and the gates are that $\tilde{C}_t$ utilizes a Tanh activation function $\tanh(\cdot)$ ranging from -1 to 1 . Finally, the memory cell C_t is generated by combining $\tilde{C}_t$ and I_t and then the previous memory cell C_{t-1} with I_t and F_t , where I_t decides how many data from $\tilde{C}_t$ are useful and F_t decides how much information from the old memory cell is retained. The detailed formulas are presented as follows:

$$I_t = \sigma(W_{xi}X_t + W_{hi}h_{t-1} + b_i) \quad (12)$$

$$F_t = \sigma(W_{xf}X_t + W_{hf}h_{t-1} + b_f) \quad (13)$$

$$O_t = \sigma(W_{xo}X_t + W_{ho}h_{t-1} + b_o) \quad (14)$$

$$\tilde{C}_t = \tanh(W_{xc}X_t + W_{hc}h_{t-1} + b_c) \quad (15)$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \quad (16)$$

$$h_t = O_t \odot \tanh(C_t) \quad (17)$$

where $X_t, \odot$ represents element-wise multiplication; $W_{xi}, W_{xf}, W_{xo}, W_{xc}; W_{hi}, W_{hf}, W_{ho}, W_{hc}$ are the weight matrices; and b_i, b_f, b_o, b_c are the bias parameters.

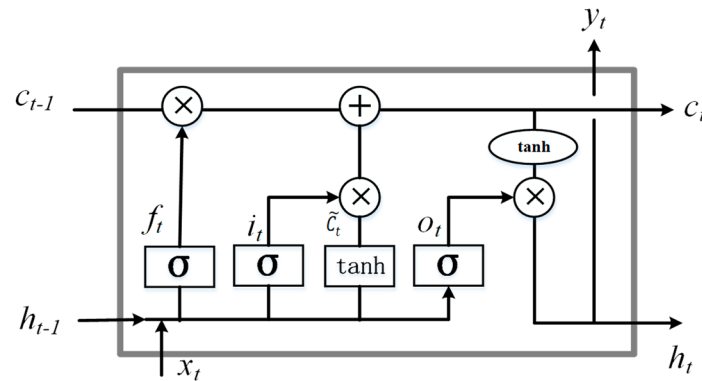


Figure 6. Block diagram of LSTM model.

2.2.5. Proposed Wavelet Transform-Based Forecasting System

As presented in Figure 7, the proposed method is divided into five steps and described as follows.

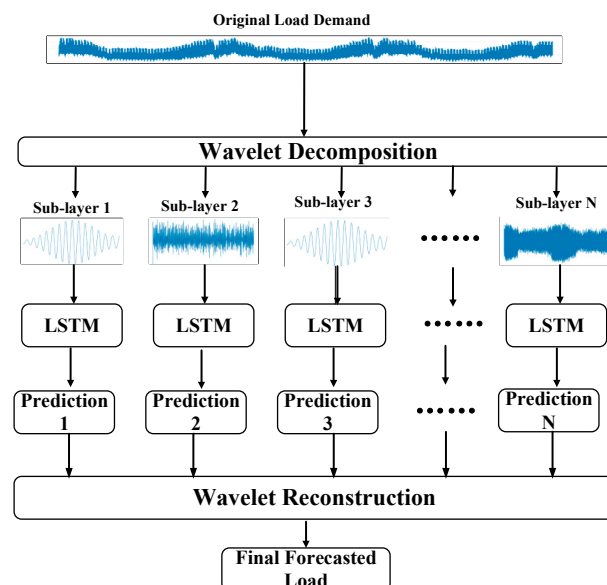


Figure 7. The overall process of the proposed spectral load forecasting model.

Step A: Data pre-processing is implemented in the first stage. The original input dataset, which includes data cleaning, is applied to the original dataset to populate the missing features. Then, a max–min scaling function is applied to the original dataset to limit the range of data between 0 and 1.

Step B: The denoised electric load is decomposed into N sub-layers via the WT decomposition algorithm. An example with nine sub-layers shows the decomposed components from the original load curve.

Step C: Then N LSTM prediction models are constructed and each BLSTM neural network model is trained for one sub-layer.

Step D: In the final step, the prediction results for all the sub-layers are reconstructed to present the final load forecasting results. Repeat Steps A–D until reaching the end of the testing dataset.

2.2.6. Performance Metrics

To assess the performance of the proposed predictor, four performance metrics are adopted, which are the mean absolute error (MAE), mean absolute percentage error (MAPE), rooted mean squared error (RMSE), and R^2 . The detailed formulas are shown as follows:

(1) MAE:

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (18)$$

(2) MAPE:

$$MAPE = \frac{\sum_{i=1}^N |(y_i - \hat{y}_i) / y_i|}{N} \times 100\% \quad (19)$$

(3) RMSE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N [y_i - \hat{y}_i]^2}{N}} \quad (20)$$

(4) R^2 :

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (21)$$

where SS_{RES} is sum squared regression error and SS_{TOT} is sum squared total error.

3. Results

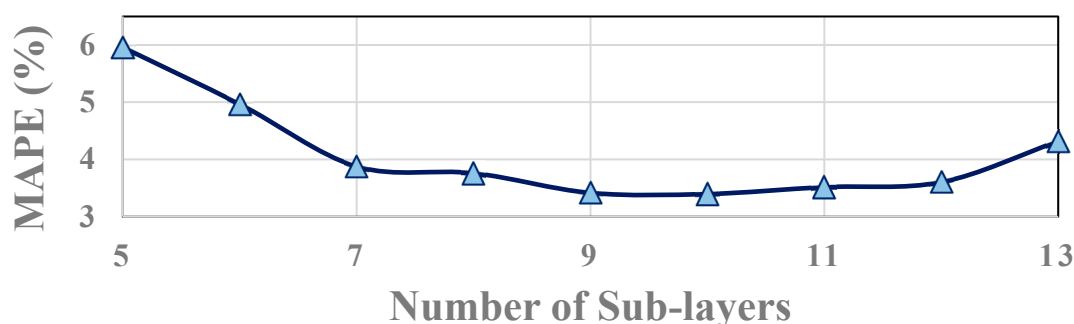
This section evaluates the proposed robust wavelet transform-based STL model by implementing three case studies. The first case study investigates the sub-layer number of the WT, the second case investigates the influence of weather information on forecasting accuracy, and the last compares different algorithms.

3.1. Sub-Layer Number

Referring to the EWT decomposition technique introduced in Section 2, the original time-varying load demand is decomposed into N sub-layers by the EWT, which is defined as S_1 – S_N in this study. The number of N has a significant impact on the final forecasting performance. In this study, the range of N increases from 5 to 13. The performance of the proposed model with different numbers of N is summarized in Table 2 and Figure 8. From these tables, it is observed that the MAE, MAPE, and RMSE are relatively large when N is too tiny (near 5) or too large (near 13) (see Figure 8). Among all N values, the dominant value is $N = 10$, followed by $N = 9$, where the RMSE values are 101.089 kW and 102.900 kW, respectively.

Table 2. Day-ahead prediction performance of the proposed model with different sublayer numbers.

N	MAE (kW)	MAPE (%)	RMSE (kW)	R ²
5	146.721	5.959	202.909	0.725
6	121.614	4.965	159.892	0.837
7	93.622	3.878	124.382	0.928
8	90.313	3.762	116.665	0.936
9	80.222	3.416	102.900	0.954
10	79.948	3.398	101.089	0.956
11	83.153	3.517	106.233	0.947
12	84.610	3.604	105.636	0.949
13	100.364	4.318	122.169	0.931

**Figure 8.** MAPEs of the proposed model with different sub-layer numbers.

Once the optimal number of the decomposition layers is determined, the load demand data are decomposed by EWT to obtain the sub-components. Then, N LSTM predictors are trained simultaneously to predict each sub-component. The predictions for the decomposed sublayers regarding the validation set are shown in Figure 9. The load demand is decomposed into N sub-layers by the EWT, which provides the best performance of the selected datasets ($N = 10$). Sub-layers (S_1 – S_3) capture the low-frequency oscillation of the baseline and the curves of these sub-layers vary smoothly and change steadily. The predicted curves of these four layers achieve higher accuracy from the prediction results. Because the Sub-layers (S_8 – S_{10}) capture high-frequency components with a high fluctuation range and include the most noise, most of the prediction errors come from these components' predictions.

3.2. Case Study 2: Influence of Weather/Temporal Information

The weather information, such as temperature, humidity, and pressure, has a significant impact on the prediction accuracy of the load forecasting model. The relevant work has demonstrated that the model with weather information as inputs achieves a higher accuracy than the model with such information. However, whether this conclusion still stands in the proposed robustness time—frequency model is unclear. In addition, the consumer's electricity activity is strongly correlated with temporal information. For instance, the electricity consumption is high during peak times (7–10 am and 4–10 pm) every day, while the electricity usage is significantly low during the off-peak times (such as midnight). Hence, in this case study, the model with weather/temporal information input is compared with the model without weather/temporal information. In addition, relevant weather variables are employed as the input variables of the proposed STLF model. A comparison is created among STLF without external data, STLF with weather information, and STLF with both weather and temporal information, see Table 3. From the table, it is observed that the model without external information achieves the lowest prediction accuracy. When the weather information, e.g., temperature, humidity, and the dew point, is added as input variables, the prediction accuracy improves RMSE by 6.34%, MAE by 7.94%, and MAPE by 7.89%, respectively. While introducing the temporal information, e.g.,

DOY, HOD, DOW, and MOY, the prediction performance improves further, demonstrating that the relevant variables can enhance the prediction accuracy.

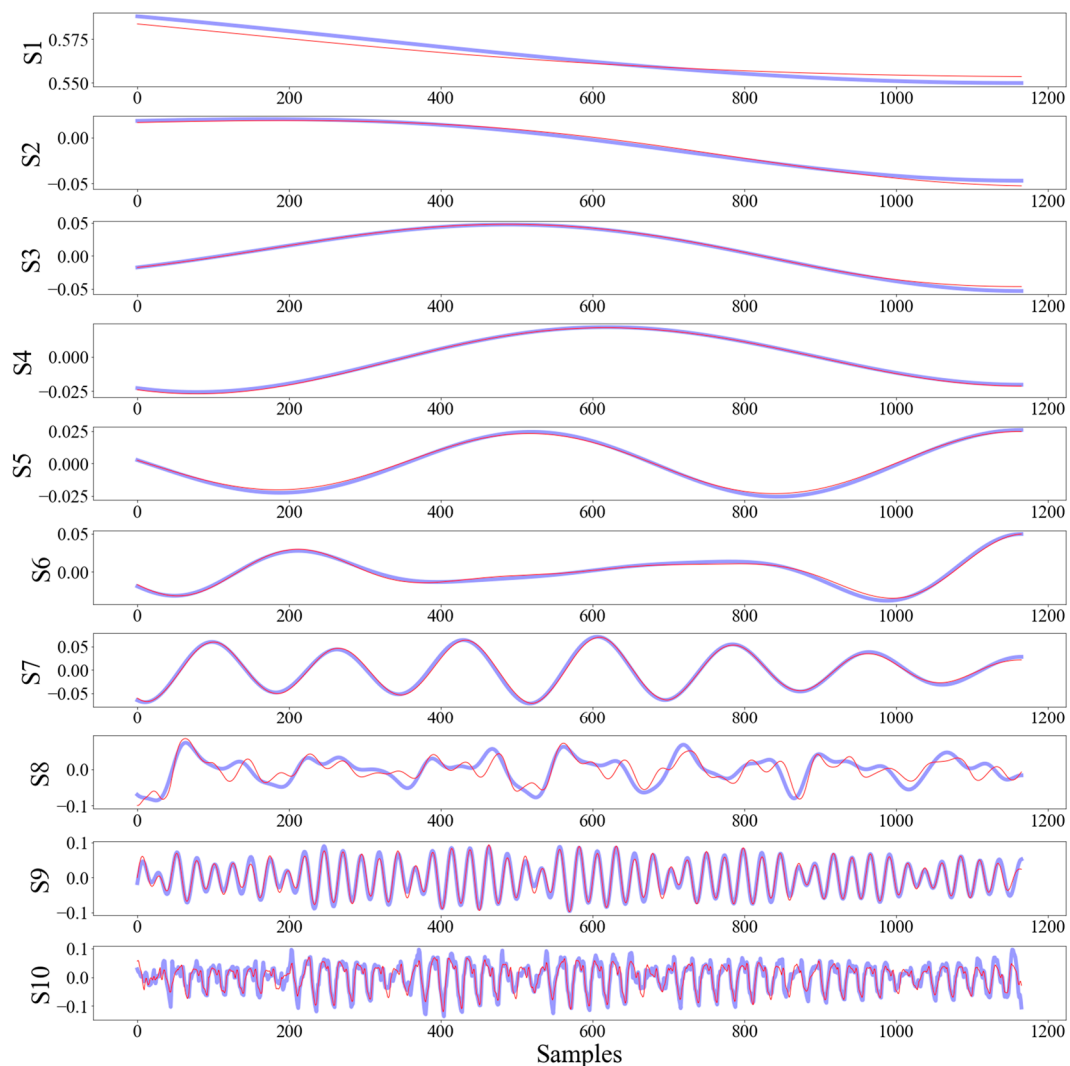


Figure 9. Validation for each sublayer in the validation set.

Table 3. Comparison of methods with/without weather information.

Method	MAE (kW)	MAPE (%)	RMSE (kW)	R^2
Model + Weather Information	73.602	3.130	94.676	0.962
Model + Weather Information + Temporal Information	70.666	3.004	91.084	0.966
Model without External Information	79.948	3.398	101.089	0.956

3.3. Case Study 3: Comparison of Different Algorithms

In this case, study, the one-step forecasting performance of the proposed method is compared with relevant forecasting approaches. A detailed description of the models adopted in this study is listed below: 1D CNN-LSTM STL model; 1D CNN-GRU STL model; EMD-LSTM STL model; VMD-LSTM STL model; and proposed model.

For models 1 and 2, the original time-varying load demand is adopted as the input of neural network models. However, for models 3, 4, and 5, the actual load demand data are decomposed via EMD/VMD/EWT, respectively, and then the neural network is trained for each sub-layer.

Table 4 shows the performance of five models considering the performance metrics, i.e., MAE, MAPE, RMSE, and R^2 , of the predicted load demand regarding the distribution-level dataset. As shown in the table, the proposed ND-EWT-LSTM-BHO outperforms other models. Moreover, the spectral load forecasting methods, including ND-EWT-LSTM-BHO, EMD-LSTM, and VMD-LSTM, have better prediction accuracy than conventional deep learning methods, including 1D CNN-LSTM and 1D CNN-GRU. 1D CNN-LSTM and 1D CNN-GRU models have the worst estimation performance, with the highest MAE, MAPE, and RMSE in all the experiment groups. The prediction performance of VMD-LSTM and EMD-LSTM are similar, just below the proposed method. Figure 10 compares the predicted values with the testing set using the proposed and benchmark models. The results predicted by the proposed model are the closest to the ground truth measurements. Moreover, the results estimated by the CNN-LSTM/CNN-GRU model are the farthest from the ground truth curve, showing that CNN-LSTM and CNN-GRU perform worst among all the algorithms.

Table 4. Prediction performance of the proposed model and related works.

Method	MAE (kW)	MAPE (%)	RMSE (kW)	R^2
1D CNN-LSTM	189.822	8.564	267.284	0.487
1D CNN-GRU	205.014	9.270	284.339	0.429
VMD-LSTM	122.899	5.010	171.473	0.803
EMD-LSTM	150.303	6.286	196.932	0.709
Proposed Method	70.666	3.004	91.084	0.966

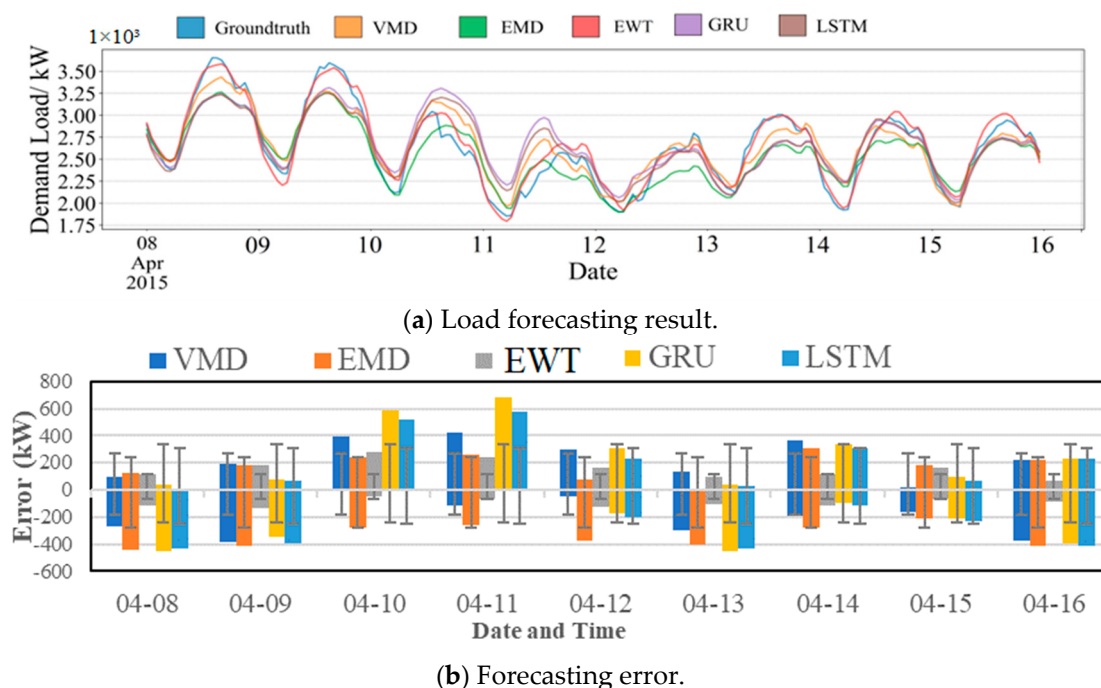


Figure 10. Day-ahead forecasting results on the distribution-level load. (a) Load demand profiles. (b) Load demand forecasting error.

Figure 11 shows the scatter plot of different forecasting models' ground truth and forecasting values. The scatter plot shows the correlation relationship between the two variables. The higher the R^2 value, the stronger the correlation between the predictions and ground truth, representing better accuracy achieved by the forecasting model. For the proposed model, the scatter about the line is relatively small and most of the points are on the regression line, with only several data values far from other data values. For other spectral methods, the R^2 of VMD-LSTM and EMD-LSTM models also show a strong

correlation with the ground truth curve, with R^2 values over 0.70. CNN-GRU shows the worst correlation from the scatter plot, with R^2 values of 0.429.

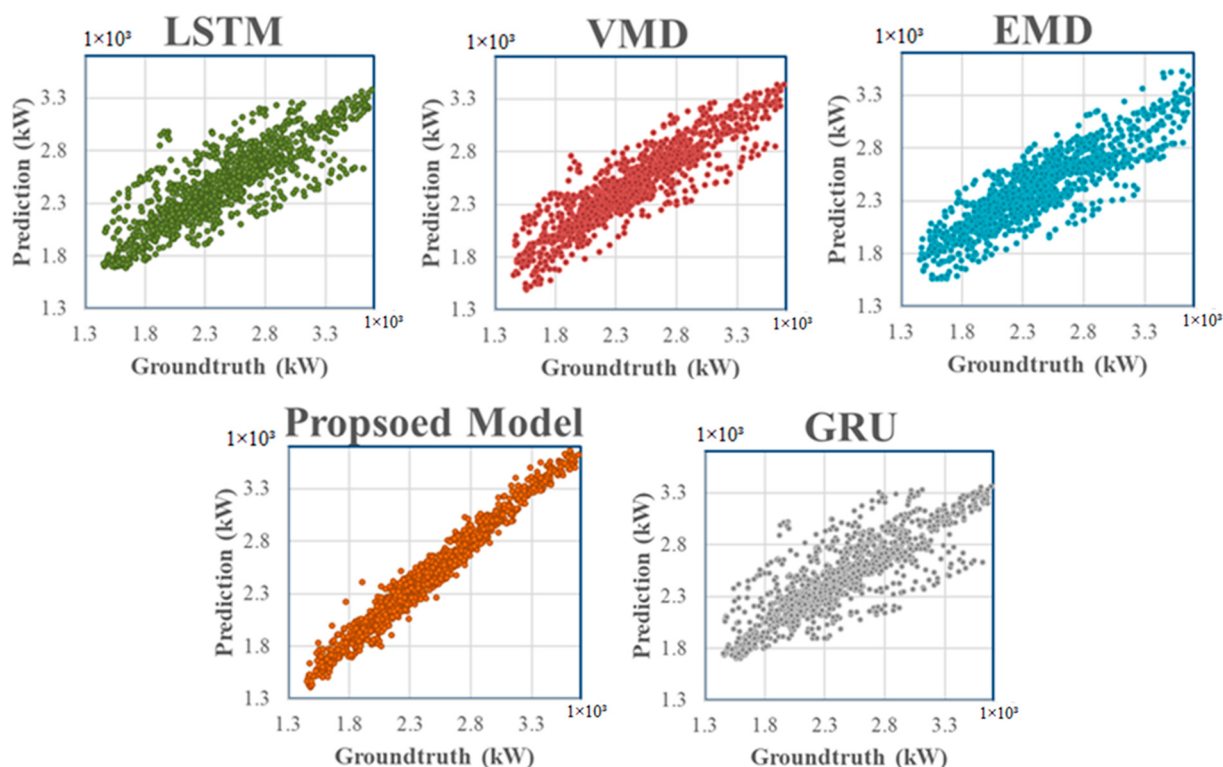


Figure 11. High-density scatter plot of ground truth and prediction values of day-ahead load forecasting models.

4. Discussion

In Section 3, three case studies are implemented. From the first case study, it is observed that the layer number N has a significant impact on the final forecasting result. A small or large value of N will reduce the prediction accuracy. When N is too small (such as 2), the EWT is inefficient in separating all the frequency components, different frequency components are overlapping, and the LSTM model causes it to be difficult to produce a precise prediction for each element. When N is too large, there are too many LSTM predictors trained and the errors from all the predictors are added up, increasing the error of the predicted active power. As for case study 2, the two important variables that strongly influence human activity, namely weather and temporal information, are investigated. The result is obtained by comparing the models with/without the weather and temporal information. From the result of the case study, it is observed that the weather information, especially temperature T and humidity H , can improve the prediction accuracy and reduce the MAE by 3.98%. The temporal data, including DOY, HOD, DOW, and MOY, can reduce the prediction error further. The results indicate that it is essential to consider the weather and temporal variables for distribution-level load forecasting. In the last case study, the robustness of the time—frequency-domain STLF models (VMD-LSTM, EMD-LSTM, and VMD-LSTM) are compared with the conventional time-domain STLF models (including LSTM and GRU). The simulation result shows that all three time-frequency domain STLF models have lower prediction errors than LSTM and GRU models, especially the EWT model. The explanation for this result is: (1) The robustness of the time—frequency-domain STLF models can extract frequency and time information from the original dataset; this information is vital to the analysis of the trend of the demand load. (2) EWT takes advantage of both EMD (high adaptation) and VMD (a strong mathematical expression). Hence, EWT achieves the highest accuracy among the five models.

5. Conclusions

In this paper, a robustness frequency—time-domain STLF model was proposed to improve the prediction accuracy of the distribution network. The proposed model utilized the three wavelet transform approaches (EWT, EMD, and VMD) to decompose the original data to extract the inherent frequency components of the load and then several LSTM predictors are trained to predict each frequency component. Besides, this paper investigated the important variables that influence the prediction results. From the case study analysis, it was found that the decomposition layer number, weather, and temporal information have a significant impact on the load forecasting results. The results verified that the model that utilizes both time and frequency data could better estimate the trend of the load. However, there are also limitations to this work. The training time of the model is longer than the normal LSTM model and it is time-consuming to tune the hyperparameters. In future work, a fast parameter-tuning algorithm will be proposed to reduce the training time and improve the model's efficiency.

Author Contributions: Conceptualization, P.G. and Y.W.; methodology, G.L.; validation and formal analysis, N.M.; supervision, P.G.; project administration, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Key Science and Technology Project of China Southern Power Grid, grant number SZKJXM20210136.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations and Notations

The following abbreviations are used in this manuscript:

STLF	Short-term load forecasting
FFT	Fast Fourier transform
WT	Wavelets transform
RNN	Recurrent neural network
EWT	Empirical wavelet transform
EMD	Empirical mode decomposition
VMD	Variational mode decomposition
ML	Machine learning
DL	Deep learning
LSTM	Long-short term memory
GRU	Gated recurrent unit
CNN	Convolutional neural network
IMF	Intrinsic mode functions
AM-FM	Amplitude modulation–frequency modulation
SVM	Support vector machine
φ_h	Low-pass filter
φ_g	High-pass filter
$D[k]$	Detail coefficients
$A[k]$	Approximation coefficients
ρ_T	Thresholding function
$\hat{f}$	Fourier spectrum
ω_n	Support boundaries
T_n	Transition area width
N	Number of sub-layers
f_t	Forgetting gate
i_t	Input gate
$\tilde{C}_t$	Candidate state value of the cell state

C_t	Cell state
C_{t-1}	Cell state of the previous step
O_t	Output gate
σ	Activation function
T_L	Look-back steps
T_F	Forecasting steps

References

- Aslam, J.; Latif, W.; Wasif, M.; Hussain, I.; Javaid, S. Comparison of Regression and Neural Network Model for Short Term Load Forecasting: A Case Study. *Eng. Proc.* **2021**, *12*, 29.
- Sun, X.; Ouyang, Z.; Yue, D. Short-term load forecasting based on multivariate linear regression. In Proceedings of the 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, 26–28 November 2017; pp. 1–5.
- Dudek, G. Pattern-based local linear regression models for short-term load forecasting. *Electr. Power Syst. Res.* **2016**, *130*, 139–147. [\[CrossRef\]](#)
- Dhaval, B.; Deshpande, A. Short-term load forecasting with using multiple linear regression. *Int. J. Electr. Comput. Eng.* **2020**, *10*, 3911. [\[CrossRef\]](#)
- Siami-Namini, S.; Tavakoli, N.; Namin, A.S. A comparison of ARIMA and LSTM in forecasting time series. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 1394–1401.
- Zheng, H.; Yuan, J.; Chen, L. Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies* **2017**, *10*, 1168. [\[CrossRef\]](#)
- Al Amin, M.A.; Hoque, M.A. Comparison of ARIMA and SVM for short-term load forecasting. In Proceedings of the 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON), Jaipur, India, 13–15 March 2019; pp. 1–6.
- Liu, S.; Cui, Y.; Ma, Y.; Liu, P. Short-term load forecasting based on GBDT combinatorial optimization. In Proceedings of the 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, 20–22 October 2018; pp. 1–5.
- Hermias, J.P.; Teknomo, K.; Monje, J.C.N. Short-term stochastic load forecasting using autoregressive integrated moving average models and Hidden Markov Model. In Proceedings of the 2017 International Conference on Information and Communication Technologies (ICICT), Sanya, China, 1–2 January 2017; pp. 131–137.
- Wu, L.; Kong, C.; Hao, X.; Chen, W. A Short-Term Load Forecasting Method Based on GRU-CNN Hybrid Neural Network Model. *Math. Probl. Eng.* **2020**, *2020*, 1428104. [\[CrossRef\]](#)
- Zhang, X.-Y.; Watkins, C.; Kuenzel, S. Multi-quantile recurrent neural network for feeder-level probabilistic energy disaggregation considering roof-top solar energy. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104707. [\[CrossRef\]](#)
- Zhang, X.Y.; Córdoba-Pachón, J.R.; Guo, P.; Watkins, C.; Kuenzel, S. Privacy-Preserving Federated Learning for Value-Added Service Model in Advanced Metering Infrastructure. *IEEE Trans. Comput. Soc. Syst.* **2022**, 1–15. [\[CrossRef\]](#)
- Gao, H.X.; Kuenzel, S.; Zhang, X.Y. A Hybrid ConvLSTM-Based Anomaly Detection Approach for Combating Energy Theft. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–10. [\[CrossRef\]](#)
- Sharma, G. ANN created real time load pattern base frequency normalization studies of linked electric power system. *Electr. Power Compon. Syst.* **2020**, *48*, 1649–1659. [\[CrossRef\]](#)
- López, M.; Sans, C.; Valero, S.; Senabre, C. Empirical Comparison of Neural Network and Auto-Regressive Models in Short-Term Load Forecasting. *Energies* **2018**, *11*, 2080. [\[CrossRef\]](#)
- Kwon, B.-S.; Park, R.-J.; Song, K.-B. Short-Term Load Forecasting Based on Deep Neural Networks Using LSTM Layer. *J. Electr. Eng. Technol.* **2020**, *15*, 1501–1509. [\[CrossRef\]](#)
- Li, Z.; Qin, Y.; Hou, S.; Zhang, R.; Sun, H. Renewable energy system based on IFOA-BP neural network load forecast. *Energy Rep.* **2020**, *6*, 1585–1590. [\[CrossRef\]](#)
- Fan, G.-F.; Guo, Y.-H.; Zheng, J.-M.; Hong, W.-C. A generalized regression model based on hybrid empirical mode decomposition and support vector regression with back-propagation neural network for mid-short-term load forecasting. *J. Forecast.* **2020**, *39*, 737–756. [\[CrossRef\]](#)
- Zhuang, L.; Liu, H.; Zhu, J.; Wang, S.; Song, Y. Comparison of forecasting methods for power system short-term load forecasting based on neural networks. In Proceedings of the 2016 IEEE International Conference on Information and Automation (ICIA), Ningbo, China, 1–3 August 2016; pp. 114–119.
- Xu, A.; Tian, M.-W.; Firouzi, B.; Alattas, K.A.; Mohammadzadeh, A.; Ghaderpour, E. A New Deep Learning Restricted Boltzmann Machine for Energy Consumption Forecasting. *Sustainability* **2022**, *14*, 10081. [\[CrossRef\]](#)
- Rafi, S.H.; Nahid Al, M.; Deeba, S.R.; Hossain, E. A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network. *IEEE Access* **2021**, *9*, 32436–32448. [\[CrossRef\]](#)
- Imani, M. Electrical load-temperature CNN for residential load forecasting. *Energy* **2021**, *227*, 120480. [\[CrossRef\]](#)
- Alhussein, M.; Aurangzeb, K.; Haider, S.I. Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting. *IEEE Access* **2020**, *8*, 180544–180557. [\[CrossRef\]](#)

24. Han, Z.; Cheng, M.; Chen, F.; Wang, Y.; Deng, Z. A spatial load forecasting method based on DBSCAN clustering and NAR neural network. In Proceedings of the Journal of Physics: Conference Series, Kunming, China, 20–22 May 2020; p. 012032.
25. Mathew, J.; Behera, R.K. EMD-Att-LSTM: A Data-Driven Strategy Combined with Deep Learning for Short-Term Load Forecasting. *J. Mod. Power Syst. Clean Energy* **2021**, *10*, 1229–1240. [[CrossRef](#)]
26. Shi, X.; Lei, X.; Huang, Q.; Huang, S.; Ren, K.; Hu, Y. Hourly day-ahead wind power prediction using the hybrid model of variational model decomposition and long short-term memory. *Energies* **2018**, *11*, 3227. [[CrossRef](#)]
27. Kim, S.; Lee, G.; Kwon, G.-Y.; Kim, D.-I.; Shin, Y.-J. Deep Learning Based on Multi-Decomposition for Short-Term Load Forecasting. *Energies* **2018**, *11*, 3433. [[CrossRef](#)]
28. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2018.
29. Liu, H.; Long, Z. An improved deep learning model for predicting stock market price time series. *Digit. Signal Process.* **2020**, *102*, 102741. [[CrossRef](#)]
30. Zhang, X.; Kuenzel, S.; Colombo, N.; Watkins, C. Hybrid Short-term Load Forecasting Method Based on Empirical Wavelet Transform and Bidirectional Long Short-term Memory Neural Networks. *J. Mod. Power Syst. Clean Energy* **2022**, *10*, 1216–1228. [[CrossRef](#)]
31. Zhu, Z.; Sun, Y.; Li, H. Hybrid of EMD and SVMs for short-term load forecasting. In Proceedings of the 2007 IEEE International Conference on Control and Automation, Hyderabad, India, 5–7 January 1995; pp. 1044–1047.
32. Semero, Y.K.; Zhang, J.; Zheng, D. EMD-PSO-ANFIS-based hybrid approach for short-term load forecasting in microgrids. *IET Gener. Transm. Distrib.* **2019**, *14*, 470–475. [[CrossRef](#)]
33. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]
34. Ghaderpour, E. Least-squares wavelet and cross-wavelet analyses of VLBI baseline length and temperature time series: Fortaleza–Hartebeesthoek–Westford–Wettzell. *Publ. Astron. Soc. Pac.* **2021**, *133*, 014502. [[CrossRef](#)]
35. Liu, T.; Luo, Z.; Huang, J.; Yan, S. A comparative study of four kinds of adaptive decomposition algorithms and their applications. *Sensors* **2018**, *18*, 2120. [[CrossRef](#)] [[PubMed](#)]

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.