

Article

Point-Interval Forecasting for Electricity Load Based on Regular Fluctuation Component Extraction

Bilin Shao, Zixuan Yao *  and Yifan Qiang

School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China

* Correspondence: yzx1999@xauat.edu.cn

Abstract: The fluctuation and uncertainty of the electricity load bring challenges to load forecasting. Traditional point forecasting struggles to avoid errors, and pure interval forecasting may cause the problem of too wide an interval. In this paper, we combine point forecasting and interval forecasting and propose a point-interval forecasting model for electricity load based on regular fluctuation component extraction. Firstly, the variational modal decomposition is combined with the sample entropy to decompose the original load series into a strong regular fluctuation component and a weak regular fluctuation component. Then, the gate recurrent unit neural network is used for point forecasting of the strong regular fluctuation component, and the support vector quantile regression model is used for interval forecasting of the weak regular fluctuation component, and the results are accumulated to obtain the final forecasting intervals. Finally, experiments were conducted using electricity load data from two regional electricity grids in Shaanxi Province, China. The results show that combining the idea of point interval, point forecasting, and interval forecasting for components with different fluctuation regularity can effectively reduce the forecasting interval width while having high accuracy. The proposed model has higher forecasting accuracy and smaller mean interval width at various confidence levels compared to the commonly used models.

Keywords: electricity load forecasting; point-interval forecasting; regular fluctuation component; variational modal decomposition; sample entropy; gate recurrent unit; support vector quantile regression



Citation: Shao, B.; Yao, Z.; Qiang, Y. Point-Interval Forecasting for Electricity Load Based on Regular Fluctuation Component Extraction. *Energies* **2023**, *16*, 1988. <https://doi.org/10.3390/en16041988>

Academic Editors: Antonio Gabaldón, María Carmen Ruiz-Abellón and Luis Alfredo Fernández-Jiménez

Received: 11 January 2023
Revised: 10 February 2023
Accepted: 14 February 2023
Published: 17 February 2023



Copyright: © 2023 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 emergence of various new electricity-consuming products, people's daily electricity consumption is increasing, and the electricity load is growing, which makes the complexity and uncertainty of the electricity system grow [1]. Therefore, accurate electricity load forecasting is crucial to help decision-makers reasonably arrange the start and stop of engine groups within the electricity grids, maintain the safety and stability of the grid operation, effectively reduce the cost of electricity generation, and improve the efficiency of the electricity system [2]. However, electricity load forecasting is not easy, especially because the fluctuation and uncertainty of electricity load bring great challenges to load forecasting [3].

Currently, the research on electricity load forecasting is mainly divided into deterministic point forecasting and uncertainty forecasting. There are three main types of models for deterministic point forecasting: conventional statistical models [4], machine learning models [5], and deep learning models [6]. Conventional statistical models, such as exponential smoothing [7], autoregressive moving average [8], and autoregressive integrated moving average [9], have long been used in load forecasting, but the accuracy of the models decreases as the forecasting time increases, and the ability to fit volatility datasets and highly nonlinear data is weak. Machine learning methods can capture complex nonlinear relationships and support vector regression [10], random forest [11], and other machine learning methods have been successfully applied to the field of load forecasting. However,

the forecasting results obtained by machine learning are often less accurate than deep learning. Deep learning conforms to the trend of big data and has a stronger learning ability for massive data. Lin et al. [12] used graph neural networks for short-term load forecasting. Rafi et al. [13] proposed a combination model of convolutional neural network (CNN) and long short-term memory network (LSTM) for electricity load forecasting. Fang et al. [14] proposed a combination model of CNN and bidirectional gated recurrent unit (BIGRU) based on attention mechanism (ATT) for short-term electricity load forecasting. Although scholars have constantly improved the common models in load forecasting and increased their forecasting accuracy, the error that can occur in point forecasting can never be avoided by model optimization. Compared with deterministic point forecasting, uncertainty forecasting can adequately consider the fluctuation of electricity load [15], reasonably quantify the potential fluctuation, and obtain a possible fluctuation range of electricity load, thus providing a more comprehensive reference for the planning, operation, and scheduling of electricity systems.

Uncertainty forecasting is mainly divided into probability density forecasting and interval forecasting. He et al. [16] proposed a probability density estimation method based on the copula theory for short-term electricity load forecasting. Ding et al. [17] proposed a kernel density estimation method based on bootstrap for natural gas demand forecasting. Although these probability density estimation methods have good forecasting results, they require data to obey a certain hypothetical distribution, which may not hold in reality, resulting in unreliable forecasting intervals obtained [18]. Gan et al. [19] performed interval forecasting with the temporal convolutional network. Saeed et al. [20] used a bidirectional LSTM model for interval forecasting. However, when using the neural network method for interval forecasting, if the training data fluctuate a lot, interval forecasting is not effective [21]. Quantile regression (QR) does not depend on the initial assumption distribution of the data and is not affected by extreme values, so many scholars have started to combine quantile regression with other models for interval forecasting [22]. Zhao et al. [23] combined the LS-SVM model with quantile regression for wind power forecasting and obtained better forecasting results. Ma et al. [24] combined Gaussian distribution correction with the CNN model and used the quantile regression for electricity load-interval forecasting, which improved the forecasting accuracy and flexibility of the model. Hu et al. [25] used quantile regression combined with a time convolution network (TCN) for wind power interval forecasting, and their results proved that the model has significant advantages in interval accuracy and width. The above researchers all perform interval forecasting for the original load series directly without decomposing the fluctuation of the original load series, extracting the irregular fluctuation from it, and performing interval forecasting for the irregular fluctuation only. The direct interval forecasting for the original load series causes the forecasting interval to be too wide, and when the interval width is too wide, the obtained forecasting results lose their value.

The key to interval forecasting is to discover the regularity of fluctuation and to forecast the possible fluctuation. Cartagena et al. [26] proposed that the forecasting interval is composed of a data-generating function and additive noise. Thus, it is necessary to separate the component of strong and weak regularity fluctuation from the original load series before performing interval forecasting and modeling for different fluctuation components separately to obtain more accurate forecasting results. Ding et al. [27] combined empirical modal decomposition (EMD) with singular spectrum analysis (SSA) to extract linear and nonlinear components and then used an autoregressive integrated moving average model (ARIMA) for point forecasting of linear components and combined BP neural network with improved first-order Markov chain (IFOMC) model for interval forecasting of the nonlinear component; however, the EMD decomposition has modal mixing phenomenon, which will affect the final forecasting results. Wang et al. [28] proposed EEMD-SE-RVM combined model for interval forecasting; firstly, the original load series was decomposed into multiple components using the ensemble empirical modal decomposition (EEMD) method, then the components were reconstructed using the sample entropy (SE), finally, using the relevance

vector machine (RVM) to forecast each component and accumulate to obtain the forecasting intervals. The experiment results showed that using sample entropy for reconstruction reduced the complexity of the model and ensured the accuracy and reliability of the forecasting intervals. Although EEMD solves the modal mixing phenomenon in EMD, the white noise it adds to the original signal contaminates the fluctuation trend of the original signal [29]. In contrast, variational modal decomposition (VMD) solves this problem well. Liu et al. [30] used the VMD technique to decompose the original series and used the ATT-GRU model to forecast the individual components, and accumulated the forecasting results to obtain the forecasting intervals. Zhang et al. [31] combined VMD with phase space reconstruction (PSR) for decomposition and reconstruction of the original series, then used the GRUQR model to perform interval forecasting for the reconstructed components and accumulate them to obtain the forecasting intervals. The model reconstructed the decomposed components of the original series, integrated the components with similar features, and solved the problem of model complexity caused by too many components. The above researchers decomposed the original load series and built models to forecast for different components and obtained better forecasting intervals, but still performed interval forecasting for all components, which led to wide interval width.

Based on the above research, this paper proposes a point-interval forecasting model for electricity load based on regular fluctuation component extraction. Firstly, VMD is used to decompose the original load series, and SE is used to reconstruct the decomposed components into strong regular fluctuation components and weak regular fluctuation components. Then, we build the GRU model for point forecasting of the strong regular fluctuation component and the support vector quantile regression (SVQR) model for interval forecasting of the weak regular fluctuation component. Finally, the fluctuation intervals obtained from interval forecasting and the point forecasting results are accumulated to obtain the forecasting intervals of electricity load. In addition, by setting up a variety of comparison experiments, the model forecasting results are analyzed and evaluated using root mean square error, mean absolute error, prediction interval coverage probability, prediction interval normalized average width, and other indicators. The innovations and contributions of this paper include the following:

1. Reasonable interval forecasting needs to dig out the regularity of load fluctuation and forecast the possible fluctuation. This paper combines point forecasting and interval forecasting, interval forecasting for weak regular fluctuation components, and point forecasting for strong regular fluctuation components, which can ensure the accuracy of the forecasting results and effectively reduce the forecasting interval width.
2. Reconstructing the decomposed components and merging the components with similar fluctuation regularity to avoid an overly complex model.
3. Using high accuracy deep learning model for point forecasting and machine learning model with strong anti-fluctuation and short training time for interval forecasting. It avoids the bad results of interval forecasting with deep learning when the training datasets fluctuate a lot and reduces the training time of the whole model.

The rest of this paper is presented as follows: Section 2 introduces the theoretical basis, model structure, and forecasting process, Section 3 describes the datasets and model evaluation metrics used in this paper and experiments and results are analyzed in Section 4, and Section 5 gives the conclusion of this research.

2. Materials and Methods

2.1. Variational Modal Decomposition

VMD is a non-recursive signal processing method that reduces the non-smoothness of time series with high complexity and strong nonlinearity. Using an iterative search for the optimal solution of the variational modals, the time series data are decomposed into multiple components with limited bandwidth, which are called intrinsic mode functions (IMF). Compared with EMD, VMD has better noise immunity and overcomes the problem of modal mixing [32]. The decomposition steps of VMD are as follows.

Firstly, the original series S is decomposed into k subseries u , which makes the sum of the bandwidths of the decomposed components minimum and the sum of all components equal to the original series, and the constrained variation expression is obtained as follows:

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \sum_{k=1}^K \|\partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega_k t} \|_2^2 \\ \text{s.t.} \sum_{k=1}^K u_k(t) = f(t) \end{cases} \quad (1)$$

where $\{u_k\}$ represents the modal component of the decomposition, $\{\omega_k\}$ denotes the central frequency of each component, $f(t)$ describes the original series, ∂_t is the obtained gradient, t is the moment, $(\delta(t) + j/\pi t) \times u_k(t)$ is the Hilbert–Huang transform, and $e^{-j\omega_k t}$ is the forecasted central frequency.

Then, we add the penalty parameter, Lagrange multiplication operator, and transforming the constrained variational problem into the unconstrained variational problem to obtain the augmented Lagrange expression as follows:

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

where α is the penalty parameter and λ is the Lagrange multiplication operator.

Finally, using the alternating direction method of multipliers to continuously update each component and its central frequency, we calculate the decomposed sub-series, the central frequency of each series, and Lagrange multiplier according to Equations (3)–(5).

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (3)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (4)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau_v \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right) \quad (5)$$

where $\hat{u}_k^{n+1}(\omega)$ represents the Fourier transform of $u_k^{n+1}(\omega)$, other Fourier transform symbols are the same, τ_v is the tolerance of signal noise.

When the series obtained after $n + 1$ cycles is not much different from the one obtained in the previous cycle, the cycle is stopped, and the stopping condition is as in Equation (6).

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \varepsilon \quad (6)$$

where ε is the positive number that infinitely tends to zero.

2.2. Sample Entropy

Multiple components are obtained after decomposition using VMD; building a model for each component for interval forecasting not only greatly increases the complexity of the model but also ignores the similar characteristics among the components. Therefore, this paper uses sample entropy to evaluate the regularity of each component fluctuation.

Sample entropy is an improvement of approximate entropy, which is a proposed method to measure the complexity of time series based on approximate entropy [33]. The

lower the sample entropy, the higher the serial self-similarity and the stronger the regularity; conversely, the more complex the sample series, the weaker the regularity [34]. Currently, the sample entropy has an excellent effect in measuring the complexity of time series. The sample entropy is calculated as follows.

First, the length of the time series is noted as l . The time series is reconstructed by combining the time points of every m adjacent one into a vector to obtain a set of vector series with dimension, $X_m(1), \dots, X_m(l - m + 1)$, where $X_m(i) = \{x(i), x(i + 1), \dots, x(i + m - 1)\}$.

Then calculate the absolute value of the maximum difference in the corresponding elements of vectors $X_m(i)$ and $X_m(j)$, according to Equation (7), where j is the moment in the series that is not equal to i .

$$d[X_m(i), X_m(j)] = \max_{k \in [0, m-1]} |x(i+k) - x(j+k)| \quad (7)$$

For each node satisfying $\{i : 1 \leq i \leq l - m + 1\}$, give a tolerance deviation r , count the number of $d[X_m(i), X_m(j)] < r$, note as B_i , and the ratio of B_i to the total number of vectors is calculated by Equation (8).

$$B_i^m(r) = \frac{B_i}{l - m - 1} \quad (8)$$

Next, the average value of $B_i^m(r)$ is calculated as in Equation (9).

$$B^m(r) = \frac{1}{l - m} \sum_{i=1}^{N-m} B_i^m(r) \quad (9)$$

Finally, increase the dimension to $m + 1$, repeat the above steps, and calculate $B^{m+1}(r)$ and the sample entropy of finite length series by Equation (10).

$$\text{SampEn}(m, r, l) = -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right] \quad (10)$$

2.3. Gate Recurrent Unit

The most important feature of recurrent neural networks (RNN) is the ability to use the output of the previous moment as the input of the next moment, which makes the data with the time series feature form a connection in the time dimension. However, ordinary RNN cannot solve the long-term time series problem, so LSTM and GRU neural networks successively appeared. Cho et al. [35] proposed the GRU model based on the LSTM neural network model, which makes improvements on the basis of the LSTM neural network. The network structure is simplified by replacing the three-gate structure of forgetting, input, and output gate with an update and reset gate. Many researchers show that GRU possesses similar performance to LSTM and requires less computation [36–38]. The GRU network structure diagram is given in Figure 1.

The parametric relationships of the GRU network are as follows:

$$r_t = \sigma(W_r \cdot [s_{t-1}, x_t]) \quad (11)$$

$$z_t = \sigma(W_z \cdot [s_{t-1}, x_t]) \quad (12)$$

$$\tilde{s}_t = \tanh(W \cdot [r_t \times s_{t-1}, x_t]) \quad (13)$$

$$s_t = (1 - z_t) \times s_{t-1} + z_t \times \tilde{s}_t \quad (14)$$

where σ represents the sigmoid function, r_t stands for the reset gate, \tilde{s}_t means the new hidden state, s_{t-1} denotes the output state at the previous moment, s_t shows the current state, z_t is the update gate, x_t is the input at the current moment, and W is the weight matrix.

The reset gate determines how much information from the previous moment is retained, and the update gate determines how the retained information is combined with the current information. Compared with the LSTM model, GRU neural network has a more simplified network structure, fewer model parameters, and faster model training speed.

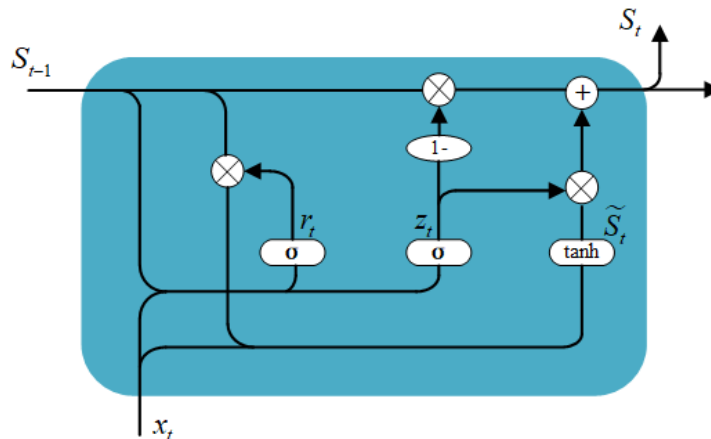


Figure 1. GRU network structure diagram.

2.4. Support Vector Quantile Regression

Support vector machine (SVM) is a supervised learning model that is widely used in classification and regression analysis problems [39]. When applied in time series regression problems, it is called support vector regression (SVR). SVR is a supervised learning algorithm for forecasting discrete values, its computational complexity does not depend on the dimensionality of the input space, and it has strong generalization ability and high forecasting accuracy [40].

For a given set of time series data sample A, their relationship is as follows:

$$f(x_t) = w\phi(x_t) + b \tag{15}$$

where x_i represents the input vector, y_i means the output vector, w denotes the weight vector, b stands for the deviation vector, and $\phi(x_t)$ is the nonlinear function that maps the input vector to the higher space.

w, b is calculated by minimizing the objective function, which is as follows:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k |y_t - f(x_t)| \tag{16}$$

where C represents the penalty parameter, y_t means the true value, and $f(x)$ is the forecast value.

Because of the multiple and complex factors affecting electric load, the traditional mean regression is less effective. Koenker et al. [41] proposed quantile regression, which not only retains the statistical information of explanatory and response variables but also reduces the effect of heteroscedasticity. They also proposed to optimize the quantile regression by using the test function and obtaining the optimal parameters by minimizing the test function, which is as follows:

$$\rho_\tau(\mu) = \mu(\tau - I(\mu)) \tag{17}$$

$$I(\mu) = \begin{cases} 1, & \mu < 0 \\ 0, & \mu \geq 0 \end{cases} \tag{18}$$

where τ is the quantile point.

Introduce QR into SVR [42], the penalty function part of the SVR is replaced by quantile regression, and obtain the support vector quantile regression (SVQR) model as follows:

$$\min_{w_\tau, b_\tau} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k \rho_\tau \left(y_t - b_\tau - \beta^T x_t - w^T \phi(x_t) \right) \quad (19)$$

where β represents the quantile matrix and ρ_τ denotes the value of the test function in Equation (17).

Rewrite Equation (19) into the quadratic programming as follows:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k (\tau \zeta_t + (1 - \tau) \bar{\zeta}_t^*) \quad (20)$$

$$\text{s.t.} \begin{cases} y_t - b - \beta^T u_t - w^T \phi(x_t) \leq \zeta_t \\ -y_t + b + \beta^T u_t + w^T \phi(x_t) \leq \bar{\zeta}_t^* \\ \zeta_t, \bar{\zeta}_t^* \geq 0 \end{cases} \quad (21)$$

To solve the above optimization problem, introduce the slack variables to construct the Lagrange function, and the results obtained from solving the quadratic programming problems of Equations (20) and (21) are as follows:

$$\begin{cases} w_\tau = \sum_{t=1}^k (\alpha_t - \alpha_t^*) \phi(x_t) \\ (b_\tau, \beta_\tau^T)^T = (U^T U)^{-1} U^T (y - K_t (\alpha - \alpha^*)) \\ Q_{y_t}(\tau | u_t, x_t) = b_\tau + \beta_\tau^T u_t + K_t (\alpha - \alpha^*) \end{cases} \quad (22)$$

where U is the matrix composed of $(1, u_t^T)$, K_t means the kernel matrix obtained from the kernel function, α, α^* denotes the slack variables, the parameters subscripted by τ all represent the parameter values at the τ quantile, and the parameters subscripted by t all represent the parameter values at the moment of t .

In the SVQR model, the selection of the kernel function plays an important role in the forecasting results. Since the electricity load data are very complex and not linearly separable, they need to be processed through ascending dimensionality. In this paper, using the Gaussian kernel as the kernel function of the SVQR model, the Gaussian function is as follows:

$$K(x, \hat{x}) = \exp \left(-\frac{(x - \hat{x})^2}{2\sigma_k^2} \right) \quad (23)$$

where \hat{x} is the kernel function center and σ_k is the width parameter of the kernel function.

2.5. Point-Interval Forecasting Model

Based on the above methods, this paper proposes a point-interval forecasting model based on regular fluctuation component extraction. Firstly, the maximum–minimum method is used to normalize the electricity load data to eliminate the influence of the dimension difference between different variables on the experiment results. Next, aiming at the fluctuation and uncertainty of the electricity load, the original electricity load series is decomposed into multiple components with different fluctuation regularity using VMD. Since the sum of the components after VMD decomposition is approximately equal to the original load series, the accumulation method is used in the component reconstruction. By calculating and comparing the sample entropy of the original series and each component, the components larger than the sample entropy of the original series are combined into weak regular fluctuation components, and those smaller ones are combined into strong regular fluctuation components by accumulation. This simplifies the model and enables the separation of the components with different strengths and weaknesses of the fluctuation regularity and the use of appropriate models to forecast for the components with

different features. The GRU neural network is used for point forecasting of strong regular fluctuation components by taking advantage of the high accuracy of deep learning, and the SVQR model is obtained by combining the quantile regression model and support vector regression model for interval forecasting of weak regular fluctuation component by taking advantage of the strong anti-fluctuation, simple model, and short training time of machine learning. Finally, the fluctuation intervals obtained from the interval forecasting are accumulated with the point forecasting results to get the electricity load forecasting intervals for the future period. Figure 2 shows the forecasting process of the proposed model.

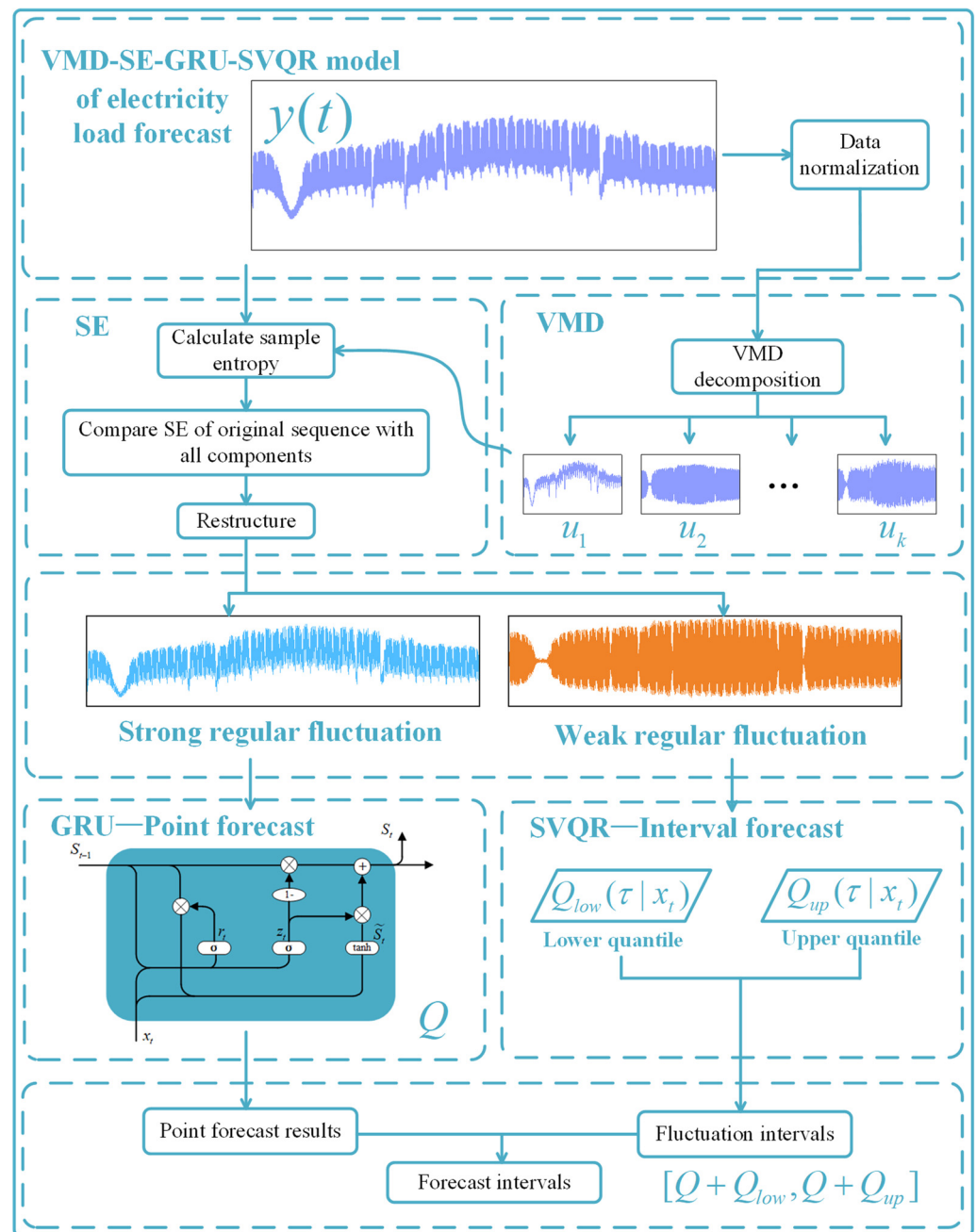


Figure 2. Process diagram of the proposed model.

The specific steps are as follows:

1. Use the maximum–minimum method to normalize the original data and obtain the datasets $y(t)$.

2. Apply VMD technique to decompose the electricity load series into multiple subseries with different fluctuation regularity u_k .
3. Calculate the sample entropy of each component, the components larger than the sample entropy of the original series are combined into weak regular fluctuation components, and the components smaller than the sample entropy of the original series are combined into strong regular fluctuation components by accumulation.
4. Perform point forecasting for the strong regular fluctuation component by GRU neural network and obtain the forecasting results Q .
5. Perform interval forecasting for the weak regular fluctuation component by SVQR model and obtain the fluctuation intervals $[Q_{low}, Q_{up}]$.
6. Accumulate the point forecasting results with the forecasting intervals to get the forecasting intervals $[Q + Q_{low}, Q + Q_{up}]$.

3. Data Description and Evaluation Metrics

3.1. Data Description

Electricity load data from the electricity grids in central and southern China's Shaanxi Province are used as the experimental datasets, noted as dataset A and dataset B. The datasets contain meteorological factors, week types, and electricity load values at the 15-min interval from 1 January to 31 December 2017, for a total of 35,040 time points. The meteorological factors include average temperature, maximum temperature, minimum temperature, relative humidity, and precipitation. The division of the dataset is shown in Figure 3, and the statistical descriptive information of the electricity load dataset is given in Table 1, including mean, maximum, minimum, standard deviation, skewness, and kurtosis.

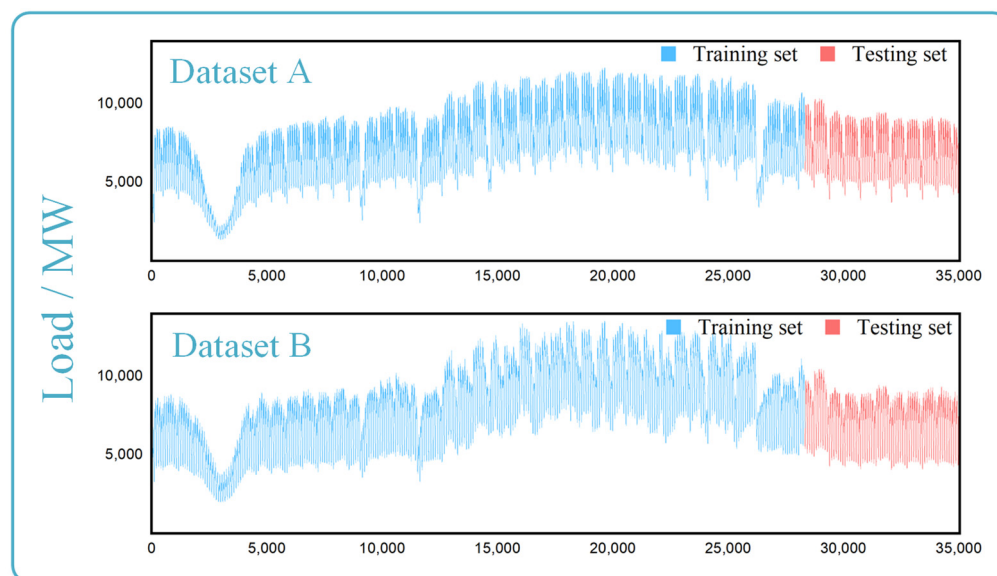


Figure 3. Division of electricity load datasets.

Table 1. Statistical information of the datasets.

Dataset	Min	Max	Mean	Std	Skewness	Kurtosis
A	1351.87	12,296.85	7274.30	2222.29	−0.11	−0.41
B	2000.13	13,536.75	7855.88	2347.06	0.03	−0.58

To reflect the applicability of the proposed model, the datasets are divided into training and test sets in a 4:1 ratio, and one week from each quarter is taken as the validation of the model, with 11 to 17 April in spring, 26 July to 1 August in summer, 17 September to 25 September in fall, and 25 December to 31 December in winter.

3.2. Evaluation Metrics

To better evaluate the accuracy and applicability of the model, this paper uses root mean square error (RMSE) and mean absolute error (MAE) as evaluation indicators for the point forecasting, and prediction interval coverage probability (PICP) and prediction interval normalized average width (PINAW) as evaluation metrics for the whole interval forecasting.

RMSE is used to measure the deviation between the forecasting values and the true values and is often used as a measure of the machine learning model's forecasting results. MAE is the mean of the absolute error, which can better reflect the actual situation of the forecasting results error. These indicators are defined as follows.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2} \quad (24)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |f(x_i) - y_i| \quad (25)$$

PICP and PINAW are commonly used indicators to evaluate the validity of the forecasting intervals. The larger the PICP and the smaller the PINAW of the forecasting intervals, the higher the correctness and reasonableness of the intervals, PICP and PINAW are calculated as follows.

$$\text{PICP} = \frac{1}{k} \sum_{i=1}^k C_i \quad (26)$$

$$C_i = \begin{cases} 1, & y_i \in [L_i, U_i] \\ 0, & y_i \notin [L_i, U_i] \end{cases} \quad (27)$$

$$\text{PINAW} = \frac{1}{kR} \sum_{i=1}^k (U_i - L_i) \quad (28)$$

where k is the total number of samples, y_i represents the true value, U_i denotes the upper limit of the interval, L_i denotes the lower limit of the interval, R means the sample range, which is the maximum value minus the minimum value.

While considering the PICP, it is also important to keep the PINAW as small as possible. When PINAW is too large, there must be a large PICP, but the obtained forecasting results are not helpful for decision. Therefore, MC is introduced as the evaluation indicator for interval forecasting [43]. The smaller the MC of the model, the smaller the PINAW, and the larger the PICP, the better the forecasting effect of the model. MC is defined as follows.

$$\text{MC} = \frac{\text{PINAW}}{\text{PICP}} \quad (29)$$

3.3. Data Standardization

In this paper, there are no missing values in the experimental data, but in order to eliminate the influence of the dimension difference between different variables on the experiment results, the maximum–minimum method is used to normalize the data. The method is as follows:

$$x = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (30)$$

where x represents the normalized load value, x_i denotes the original load value, x_{\max} is the maximum load value, and x_{\min} is the minimum load value.

4. Experiment and Results

4.1. Data Decomposition and Reconstruction

Before training the interval forecasting model, the VMD-SE method is used to decompose and reconstruct the original load series. However, in the decomposition process of

VMD, the decomposition results are mainly affected by the modal number K . When K takes too small a value, the original series decomposition is not complete and some important information will be filtered out; when K takes too large a value, it will lead to excessive decomposition, resulting in modal duplication or additional noise. The main difference between different modals is the difference in their center frequencies. When the value of K increases by 1, the center frequency of the new modal (the rightmost modal of each column in the table is the new modal) does not change much, the value of K at this time is the optimal number of decomposition. Tables 2 and 3 show the center frequencies of each modal after decomposition.

Table 2. Center frequency of each modal in dataset A.

K-Value	Center Frequency of Each Modal								
$K = 3$	3.40×10^{-5}	0.145	0.270						
$K = 4$	3.38×10^{-5}	0.083	0.218	0.353					
$K = 5$	3.37×10^{-5}	0.083	0.165	0.269	0.406				
$K = 6$	3.37×10^{-5}	0.093	0.165	0.219	0.270	0.395			
$K = 7$	3.09×10^{-5}	0.029	0.094	0.219	0.270	0.354	0.396		

Table 3. Center frequency of each modal in dataset B.

K-Value	Center Frequency of Each Modal								
$K = 3$	5.39×10^{-6}	0.010	0.093						
$K = 4$	5.29×10^{-6}	0.011	0.143	0.218					
$K = 5$	5.17×10^{-6}	0.011	0.041	0.165	0.228				
$K = 6$	5.20×10^{-6}	0.011	0.083	0.135	0.207	0.324			
$K = 7$	5.17×10^{-6}	0.011	0.081	0.134	0.207	0.238	0.353		
$K = 8$	5.17×10^{-6}	0.011	0.081	0.124	0.177	0.227	0.280	0.352	

The value of K is determined by the center frequency of the new modal after decomposition, which is shown in Tables 2 and 3. In dataset A, the center frequency of the new modal starts to stabilize after $K = 5$ in dataset B, the center frequency of the new modal starts to stabilize after $K = 6$. Therefore, the VMD method is used to decompose the electricity load series of dataset A into five modals, and the electricity load series of dataset B into six modals; each modal is considered a subseries. Figure 4 shows the decomposition results of VMD.

From Figure 4, it can be found that the electricity loads of the two datasets have some fluctuation and nonlinear characteristics, and there are some differences in the fluctuation degree and fluctuation pattern of each subseries. Therefore, we calculate the sample entropy of each subseries after the decomposition and compare it with the sample entropy of the original series. Table 4 shows the sample entropy of the subseries and original series.

Table 4. Sample entropy of subseries and original series.

IMF	Dataset A		Dataset B	
	Subseries	Original Series	Subseries	Original Series
1	0.123		0.031	
2	1.792		0.397	
3	2.344		2.222	
4	2.798	0.522	2.648	0.476
5	2.283		2.838	
6	—		2.708	

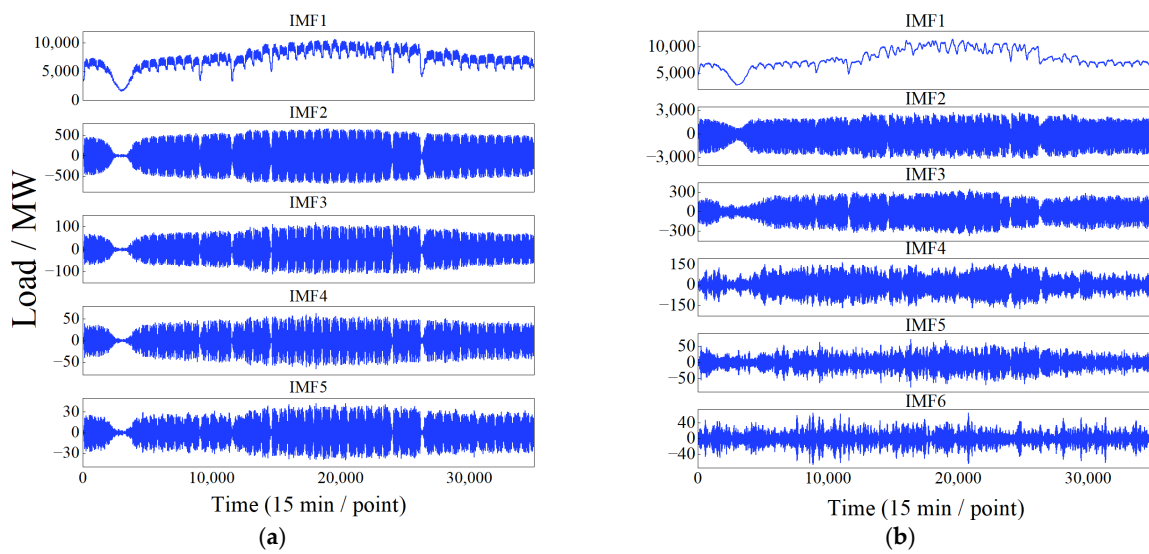


Figure 4. Decomposition results of VMD: (a) Dataset A; (b) Dataset B.

The sample entropy of the original load series in dataset A is larger than that in dataset B. And the sample entropy of subseries 1 in dataset A is 0.123, while it is only 0.031 in dataset B. Therefore, the fluctuation of dataset A is more irregular relative to dataset B. Only the sample entropy of subseries 1 in dataset A is smaller than the sample entropy of the original series, so subseries 1 is taken as the strong regular fluctuation component, and the rest subseries are reconstructed as the weak regular fluctuation component. The sample entropy of subseries 1 and 2 in dataset B is smaller than the sample entropy of the original series, so subseries 1 and subseries 2 are reconstructed as strong regular fluctuation components, and the rest subseries are reconstructed as weak regular fluctuation components. Figure 5 shows the reconstructed component series.

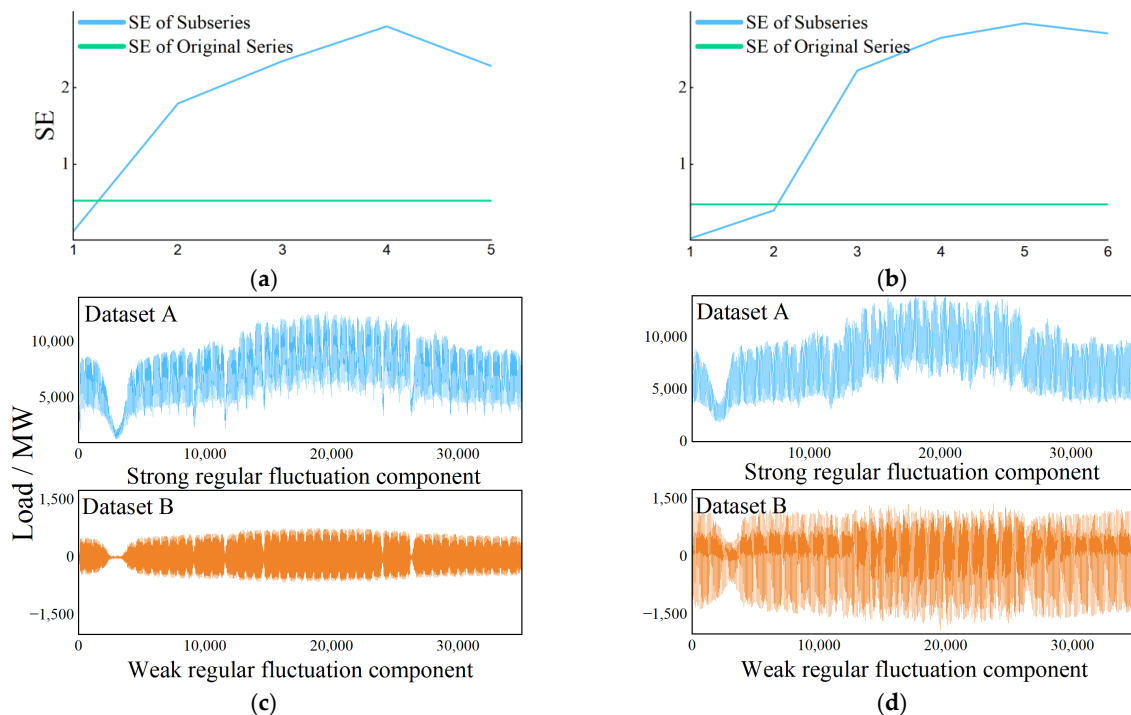


Figure 5. Subseries sample entropy and reconstruction: (a) subseries sample entropy of dataset A; (b) subseries sample entropy of dataset B; (c) reconstructed components of dataset A; (d) reconstructed components of dataset B.

4.2. Point Forecast

The strong regular fluctuation component obtained after decomposition and reconstruction is taken into the point forecasting model. In addition to the VMD-SE-GRU-SVQR model proposed in this paper, five different classes of representative models are compared, namely the SVQR and GRUQR models without decomposition and reconstruction, the SVQR and GRUQR models with VMD-SE decomposition and reconstruction, and the VMD-SE-SVR-GRUQR model. Among them, the SVQR and GRUQR models do not have the point forecasting process, so the VMD-SE decomposition reconstructed SVQR and GRUQR models, as well as the VMD-SE-GRU-SVQR model and the VMD-SE-SVR-GRUQR model were compared, and the four models, respectively, used the SVR and GRU models for point forecasting. The forecasting results of the SVR and GRU models for point forecasting of the strong regular fluctuation component are shown in Figures 6 and 7.

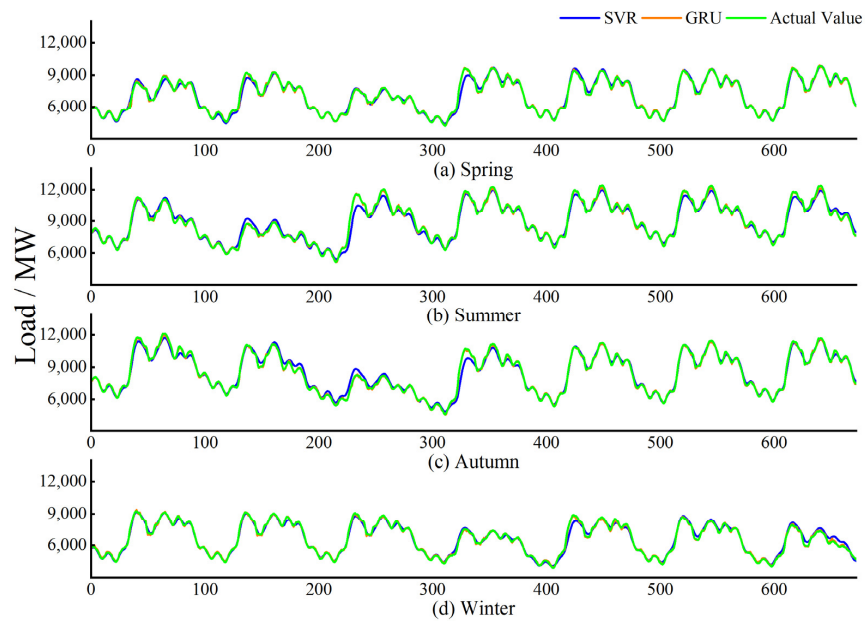


Figure 6. Point forecasting results of dataset A.

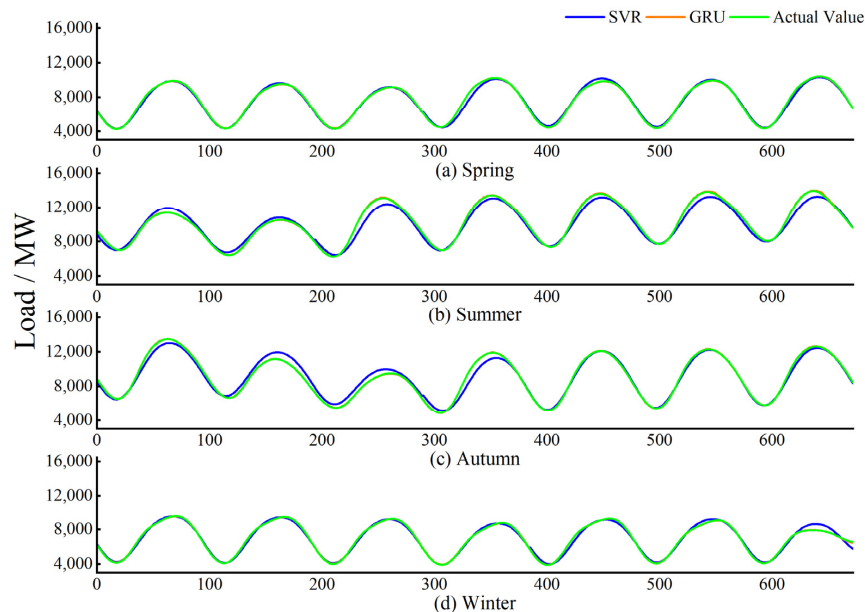


Figure 7. Point forecasting results of dataset B.

From Figures 6 and 7, it can be concluded that point forecasting using the GRU model is more accurate than point forecasting using the SVR model in both dataset A and dataset B. This is because the GRU neural network, which belongs to the category of deep learning, is better than the SVR model, which belongs to the category of machine learning, in terms of accuracy and precision. In order to show the forecasting errors of the two types of models more directly, Figure 8 shows the RMSE and MAE of the two types of models.

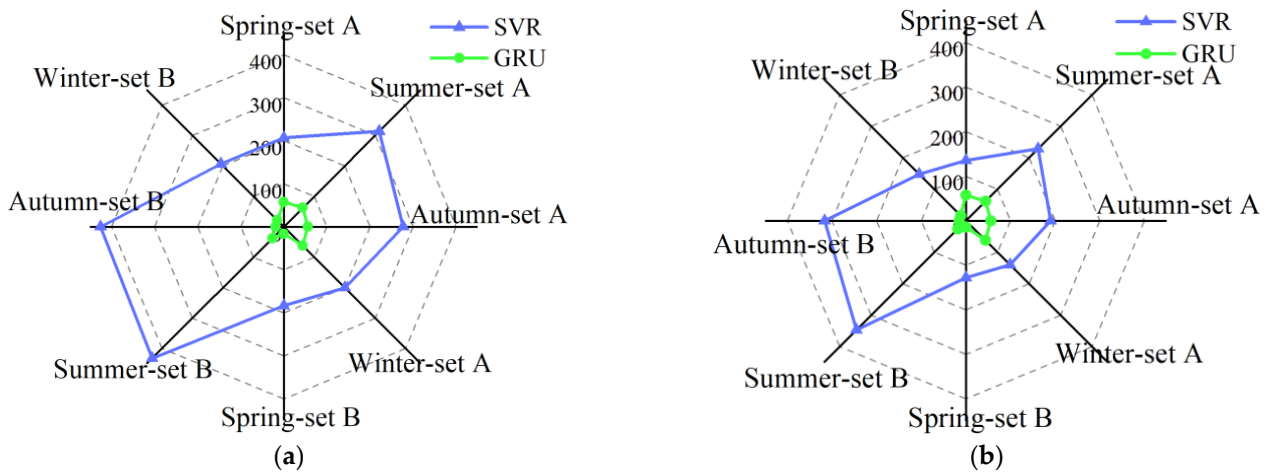


Figure 8. RMSE and MAE of model: (a) RMSE; (b) MAE.

From Figure 8, it can be concluded that the accuracy of the GRU model is higher compared to the SVR model, whatever the season, where the root mean square error is reduced by at least 100 and the mean absolute error is reduced by at least 50. The errors are always much smaller than the SVR model, further demonstrating that the accuracy of the GRU model is much better than the SVR model.

4.3. Interval Forecast

The weak regular fluctuation component obtained after decomposition and reconstruction is taken into the interval forecasting model, and the fluctuation intervals obtained from interval forecasting are cumulated with the point forecasting results to obtain the electricity load forecasting intervals. In order to better analyze and compare the forecasting results, electricity load-interval forecasting is performed at 80%, 85%, and 90% confidence levels, respectively. The forecasting results of dataset A and dataset B at a 90% confidence level are given in Figures 9 and 10. Meanwhile, in order to observe the details more clearly, partial enlargements are given on the lower right side of each figure.

The following conclusions can be drawn from Figures 9 and 10:

1. Compared to the pure interval forecasting model, the addition of the point-interval idea ensures both interval accuracy and reduces the interval width.
2. The accuracy of the intervals obtained by using SVR for point forecasting in the combination model is not as high as that by using the GRU model, which indicates that deep learning still has a great advantage over machine learning in the aspect of accuracy.
3. According to the interval width and interval coverage, the models can be generally classified into three levels; the first level is the model that uses deep learning GRU neural network for point forecasting, such as VMD-SE-GRUQR, VMD-SE-GRU-SVQR; the second level is the model that uses machine learning SVR for point forecasting, such as VMD-SE-SVQR, VMD-SE-SVR-GRUQR; and the third level is the forecasting model that does not use the point-interval idea, such as SVQR and GRUQR. The forecasting intervals obtained from the first level model have higher coverage, narrower interval width, and are closer to the actual electricity load values.

Tables 5 and 6 show the MC values of the forecasting intervals obtained by each model under different seasons. Table 7 shows the training time of each model.

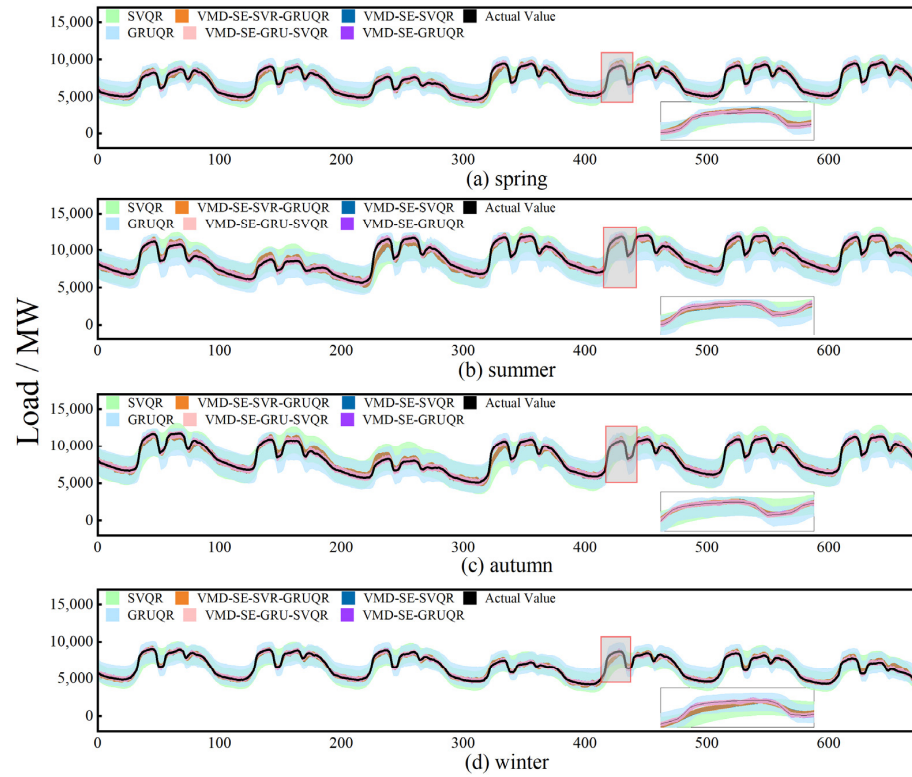


Figure 9. Forecasting results of dataset A at 90 confidence level.

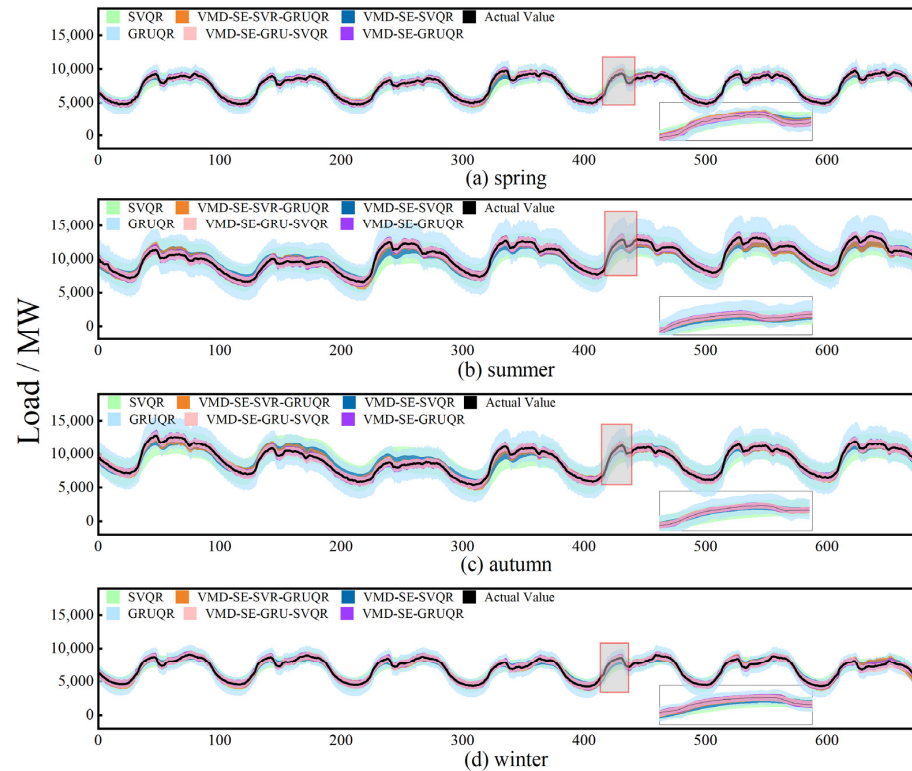


Figure 10. Forecasting results of dataset B at 90 confidence level.

Table 5. MC values in dataset A at three confidence levels.

Dataset A Models	Spring			Summer			Autumn			Winter		
	80%	85%	90%	80%	85%	90%	80%	85%	90%	80%	85%	90%
SVQR	0.445	0.486	0.527	0.450	0.502	0.538	0.433	0.469	0.553	0.447	0.500	0.545
GRUQR	0.478	0.528	1.105	0.494	0.577	1.169	0.494	0.521	1.055	0.485	0.539	1.110
VMD-SE-GRUQR	0.139	0.153	0.165	0.130	0.145	0.158	0.111	0.124	0.135	0.114	0.123	0.140
VMD-SE-SVQR	0.150	0.163	0.175	0.152	0.162	0.172	0.129	0.139	0.147	0.133	0.144	0.154
VMD-SE-SVR-GRUQR	0.148	0.160	0.172	0.150	0.159	0.171	0.128	0.137	0.146	0.133	0.140	0.153
VMD-SE-GRU-SVQR	0.141	0.156	0.169	0.132	0.146	0.160	0.113	0.127	0.135	0.114	0.127	0.139

Table 6. MC values in dataset B at three confidence levels.

Dataset B Models	Spring			Summer			Autumn			Winter		
	80%	85%	90%	80%	85%	90%	80%	85%	90%	80%	85%	90%
SVQR	0.337	0.366	0.401	0.387	0.466	0.524	0.449	0.475	0.510	0.281	0.326	0.410
GRUQR	0.365	0.432	0.482	0.593	0.670	0.771	0.478	0.543	0.620	0.410	0.487	0.548
VMD-SE-GRUQR	0.131	0.156	0.200	0.185	0.196	0.209	0.137	0.157	0.177	0.124	0.153	0.238
VMD-SE-SVQR	0.126	0.138	0.165	0.209	0.215	0.223	0.177	0.193	0.209	0.133	0.140	0.223
VMD-SE-SVR-GRUQR	0.135	0.157	0.200	0.194	0.202	0.215	0.142	0.160	0.178	0.132	0.161	0.241
VMD-SE-GRU-SVQR	0.120	0.139	0.173	0.170	0.183	0.191	0.131	0.151	0.176	0.116	0.141	0.233

Table 7. Training time of six models.

Models	SVQR	GRUQR	VMD-SE-GRUQR	VMD-SE-SVQR	VMD-SE-SVR-GRUQR	VMD-SE-GRU-SVQR
Time/s	100	7427	14,211	256	6926	7536

In order to show the results of each model more visually, Figure 11 shows the comparison of different models' MC values under each quarter. From Tables 5–7 and Figure 11, the following conclusions can be drawn:

1. VMD-SE-GRUQR model and VMD-SE-GRU-SVQR model possess better forecasting intervals compared with other models.
2. Comparing GRUQR, VMD-SE-GRUQR, and VMD-SE-GRU-SVQR models, the results show that decomposing and reconstructing the time series data before forecasting can effectively reduce the model MC values.
3. Comparing the VMD-SE-GRUQR, VMD-SE-SVQR, VMD-SE-SVR-GRUQR, and VMD-SE-GRU-SVQR models, it is found that the models with point forecasting by GRU possess smaller MC values.
4. VMD-SE-GRUQR is compared with the VMD-SE-GRU-SVQR model, and the results show that both models have low MC values, indicating that the forecasting intervals obtained by both models are superior, but the training time of the latter model required is shorter than that of the former.

To better evaluate the forecasting ability of each model under different datasets, Table 8 shows the average values of PICP, PINAW, and MC with different confidence levels. From Table 8, it can be found that all the models, except the VMD-SE-SVQR model, possess more than 90% interval coverage, which is caused by the lack of machine learning accuracy and is improved after using the deep learning model. In terms of the average interval width, the VMD-SE-SVQR and VMD-SE-GRU-SVQR models have narrower interval widths, while the VMD-SE-GRUQR and VMD-SE-SVR-GRUQR models have relatively wider interval width, which is due to the ineffectiveness of the neural network method for interval forecasting when the training data are highly fluctuating. In addition, the VMD-SE-GRU-SVQR model proposed in this paper possesses a smaller MC, which is only 0.13 and is lower than other models.

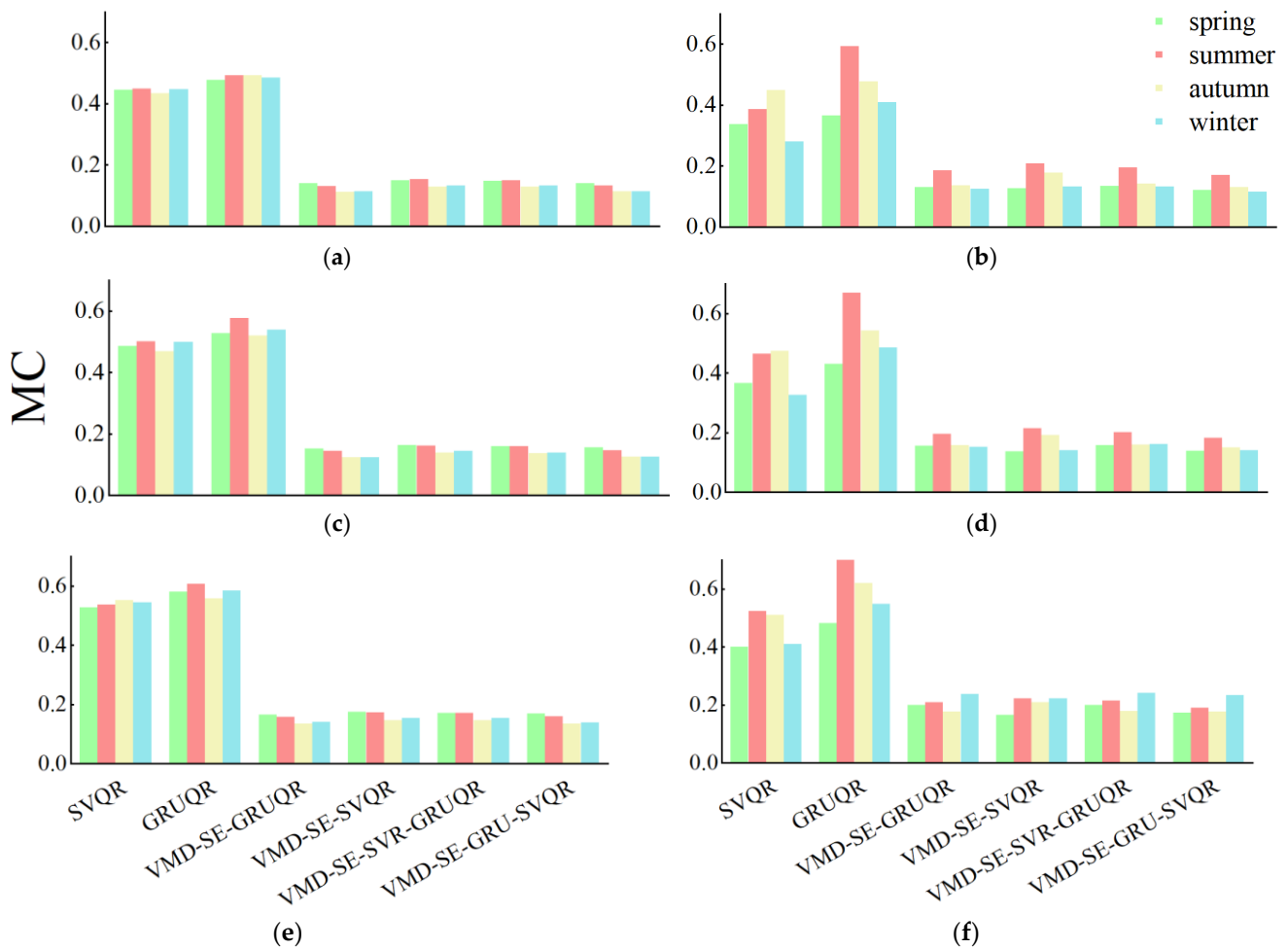


Figure 11. MC values of the model at three confidence levels: (a) MC in 80% confidence level of dataset A; (b) MC in 80% confidence level of dataset B; (c) MC in 85% confidence level of dataset A; (d) MC in 85% confidence level of dataset B; (e) MC in 90% confidence level of dataset A; (f) MC in 90% confidence level of dataset B.

Table 8. Averages of the evaluation metrics.

Models	80%			85%			90%		
	PICP	PINAW	MC	PICP	PINAW	MC	PICP	PINAW	MC
SVQR	0.944	0.419	0.446	0.964	0.498	0.516	0.979	0.564	0.576
GRUQR	1.000	0.435	0.436	1.000	0.471	0.471	1.000	0.520	0.520
VMD-SE-GRUQR	0.999	0.134	0.134	0.999	0.151	0.151	1.000	0.178	0.178
VMD-SE-SVQR	0.841	0.127	0.151	0.781	0.141	0.182	0.917	0.168	0.184
VMD-SE-SVR-GRUQR	0.919	0.134	0.145	0.942	0.151	0.160	0.959	0.178	0.185
VMD-SE-GRU-SVQR	0.996	0.129	0.130	0.998	0.146	0.146	1.000	0.172	0.172

Since the models that did not combine point-interval forecasting clearly have worse MC values than the forecasting results of the other models, the models that combined point-interval forecasting are compared again. Figure 12 shows the MC averages of the four combined point-interval forecasting models, in which the VMD-SE-GRU-SVQR model proposed in this paper all has the lowest MC averages at three confidence levels, the MC values of VMD-SE-GRUQR, VMD-SE-SVR-GRUQR, and VMD-SE-SVQR are slightly higher, while the SVQR and GRUQR models have excessive MC values, which fully demonstrates the superiority of the point-interval forecasting model proposed in this paper.

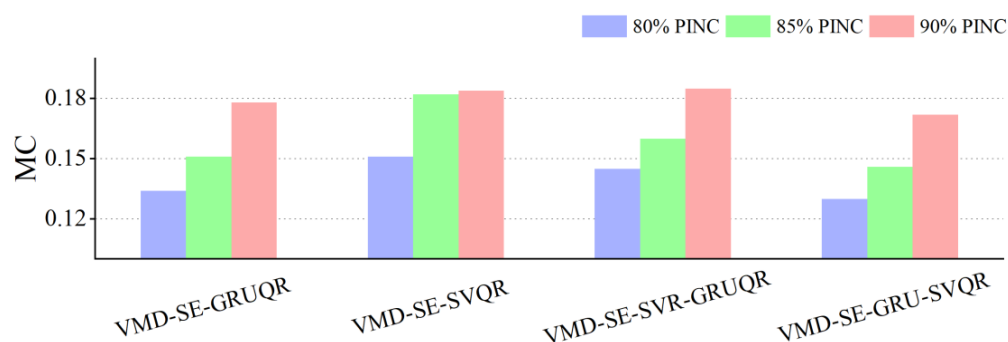


Figure 12. MC averages of some models at three confidence levels.

5. Conclusions

Due to the certain fluctuation and uncertainty of electricity load, the point forecasting models cannot avoid the error, and the direct interval forecasting may have too wide an interval width. Combining the ideas of point forecasting and interval forecasting, taking the respective advantages of deep learning and machine learning, this paper proposes a point-interval forecasting model for electricity load based on regular fluctuation component extraction (VMD-SE-GRU-SVQR model). According to the experimental analysis based on two datasets, the following conclusions are drawn:

1. The forecasting interval obtained by the VMD-SE-GRU-SVQR model achieves optimal results under a variety of confidence levels.
2. Separating components with different strengths and weaknesses of fluctuation regularity from the original load series and using the idea of point interval for interval forecasting can greatly reduce the interval width while ensuring interval accuracy.
3. Taking full advantage of the high accuracy of deep learning, the GRU neural network is used for point forecasting of strong regular fluctuation component, which improves the forecasting accuracy of the model; taking advantage of the strong anti-fluctuation and short training time of machine learning, the SVQR model is used for interval forecasting of weak regular fluctuation component, which reduces the interval width and training time.
4. The model proposed in this paper inherits the advantages of quantile regression, which can reduce the confidence level while satisfying the accuracy and further narrowing the interval width.

The limited influencing factors of the dataset used in this paper may lead to less accurate forecasting results of the model, and further research is needed later to incorporate more relevant factors affecting the electricity load.

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

Funding: This research was funded by the National Natural Science Foundation of China (grant number: No. 62072363).

Data Availability Statement: Not applicable.

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

Nomenclature

K	Total number of modes	w, W	Weight matrix
u_k	Modal component of the decomposition	b	Deviation vector
ω_k	Central frequency of each component	C	Penalty parameter in SVQR
$f(t)$	Original series	y_t	True value
∂_t	Gradient	$f(x)$	Forecast value
t	Moment	$\rho_\tau(\bullet)$	Test function in quantile regression
$e^{-j\omega_k t}$	Forecasted central frequency	ξ_t, ξ_t^*	Relaxation variables
α	Penalty parameter in VMD	τ	Quantile point
λ	Lagrange multiplication operator	β	Quantile matrix
$\hat{u}_k^{n+1}(\omega)$	Fourier transform of $u_k^{n+1}(\omega)$	ρ_τ	Test function for quantile regression
τ_0	Tolerance of signal noise	K_t	Kernel matrix
ε	Positive number that infinitely tends to zero	α, α^*	Slack variables
r	Tolerance deviation	\hat{x}	Kernel function center
σ	Sigmoid function	σ_k	Width parameter of the kernel function
\tilde{s}_t	New hidden state	Q	Forecasting results
s_t	Current state	Q_{low}	Lower limit of fluctuation intervals
s_{t-1}	Output state at the previous moment	Q_{up}	Upper limit of fluctuation intervals
x_t	Input at the current moment	n	Total number of samples
r_t	Reset gate	U_i	Upper limit of forecasting intervals
z_t	Update gate	L_i	Lower limit of forecasting intervals

References

- Serrano-Guerrero, X.; Briceño-León, M.; Clairand, J.-M.; Escrivá-Escrivá, G. A new interval prediction methodology for short-term electric load forecasting based on pattern recognition. *Appl. Energy* **2021**, *297*, 117173. [\[CrossRef\]](#)
- Hou, H.; Liu, C.; Wang, Q.; Wu, X.; Tang, J.; Shi, Y.; Xie, C. Review of load forecasting based on artificial intelligence methodologies, models, and challenges. *Electr. Power Syst. Res.* **2022**, *210*, 108067. [\[CrossRef\]](#)
- Wan, C.; Xu, Z.; Pinson, P.; Dong, Z.Y.; Wong, K.P. Optimal Prediction Intervals of Wind Power Generation. *IEEE Trans. Power Syst.* **2014**, *29*, 1166–1174. [\[CrossRef\]](#)
- Sim, S.-K.; Maass, P.; Lind, P.G. Wind Speed Modeling by Nested ARIMA Processes. *Energies* **2019**, *12*, 69. [\[CrossRef\]](#)
- Lu, H.; Azimi, M.; Iseley, T. Short-term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine. *Energy Rep.* **2019**, *5*, 666–677. [\[CrossRef\]](#)
- Pang, C.; Zhang, B.; Yu, J. Short-term power load forecasting based on LSTM recurrent neural network. *Electr. Power Eng. Technol.* **2021**, *1*, 175–180. [\[CrossRef\]](#)
- Baykal, T.M.; Colak, H.E.; Kilinc, C. Forecasting future climate boundary maps (2021–2060) using exponential smoothing method and GIS. *Sci. Total Environ.* **2022**, *848*, 157633. [\[CrossRef\]](#)
- Xin, Y.; Gao, J.; Yang, X.; Yang, J. Maximum likelihood estimation for uncertain autoregressive moving average model with application in financial market. *J. Comput. Appl. Math.* **2023**, *417*, 114604. [\[CrossRef\]](#)
- Xiang, Y. Using ARIMA-GARCH Model to Analyze Fluctuation Law of International Oil Price. *Math. Probl. Eng.* **2022**, *2022*, 3936414. [\[CrossRef\]](#)
- Gao, X.; Jia, B.; Li, G.; Ma, X. Calorific Value Forecasting of Coal Gangue with Hybrid Kernel Function–Support Vector Regression and Genetic Algorithm. *Energies* **2022**, *15*, 6718. [\[CrossRef\]](#)
- Dang, S.; Peng, L.; Zhao, J.; Li, J.; Kong, Z. A Quantile Regression Random Forest-Based Short-Term Load Probabilistic Forecasting Method. *Energies* **2022**, *15*, 663. [\[CrossRef\]](#)
- Lin, W.; Wu, D.; Boulet, B. Spatial-Temporal Residential Short-Term Load Forecasting via Graph Neural Networks. *IEEE Trans. Smart Grid* **2021**, *12*, 5373–5384. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Fang, N.; Yu, J.; Li, X.; Wang, C. Short-Term Power Load Forecasting Based on CNN-BIGRU-ATTENTION. *Comput. Simul.* **2022**, *2*, 40–44. [\[CrossRef\]](#)
- Khosravi, A.; Nahavandi, S.; Creighton, D.; Atiya, A.F. Comprehensive Review of Neural Network-Based Prediction Intervals and New Advances. *IEEE Trans. Neural Netw.* **2011**, *22*, 1341–1356. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Ding, L.; Zhao, Z.; Wang, L. Probability density forecasts for natural gas demand in China: Do mixed-frequency dynamic factors matter? *Appl. Energy* **2022**, *312*, 118756. [\[CrossRef\]](#)
- Wang, R.; Li, C.; Fu, W.; Tang, G. Deep Learning Method Based on Gated Recurrent Unit and Variational Mode Decomposition for Short-Term Wind Power Interval Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 3814–3827. [\[CrossRef\]](#)
- Gan, Z.; Li, C.; Zhou, J.; Tang, G. Temporal convolutional networks interval prediction model for wind speed forecasting. *Electr. Power Syst. Res.* **2021**, *191*, 106865. [\[CrossRef\]](#)

20. Saeed, A.; Li, C.; Danish, M.; Rubaiee, S.; Tang, G.; Gan, Z.; Ahmed, A. Hybrid Bidirectional LSTM Model for Short-Term Wind Speed Interval Prediction. *IEEE Access* **2020**, *8*, 182283–182294. [[CrossRef](#)]
21. Zhang, Y.; Zhao, Y.; Kong, C.; Chen, B. A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic. *Energy Convers. Manag.* **2020**, *203*, 112254. [[CrossRef](#)]
22. Alcántara, A.; Galván, I.M.; Aler, R. Deep neural networks for the quantile estimation of regional renewable energy production. *Appl. Intell.* **2022**. [[CrossRef](#)]
23. Zhao, X.; Ge, C.; Ji, F.; Liu, Y. Monte Carlo Method and Quantile Regression for Uncertainty Analysis of Wind Power Forecasting Based on Chaos-LS-SVM. *Int. J. Control Autom. Syst.* **2021**, *19*, 3731–3740. [[CrossRef](#)]
24. Ma, X.; Dong, Y. An estimating combination method for interval forecasting of electrical load time series. *Expert Syst. Appl.* **2020**, *158*, 113498. [[CrossRef](#)]
25. Hu, J.; Luo, Q.; Tang, J.; Heng, J.; Deng, Y. Conformalized temporal convolutional quantile regression networks for wind power interval forecasting. *Energy* **2022**, *248*, 123497. [[CrossRef](#)]
26. Cartagena, O.; Parra, S.; Muñoz-Carpintero, D.; Marín, L.G.; Sáez, D. Review on Fuzzy and Neural Prediction Interval Modelling for Nonlinear Dynamical Systems. *IEEE Access* **2021**, *9*, 23357–23384. [[CrossRef](#)]
27. Ding, W.; Meng, F. Point and interval forecasting for wind speed based on linear component extraction. *Appl. Soft Comput.* **2020**, *93*, 106350. [[CrossRef](#)]
28. Wang, S.; Sun, Y.; Zhou, Y.; Jamil Mahfoud, R.; Hou, D. A New Hybrid Short-Term Interval Forecasting of PV Output Power Based on EEMD-SE-RVM. *Energies* **2020**, *13*, 87. [[CrossRef](#)]
29. Shao, B.; Yan, Y.; Zeng, H. VMD-WLSLSTM Load Prediction Model Based on Shapley Values. *Energies* **2022**, *15*, 487. [[CrossRef](#)]
30. Liu, H.; Han, H.; Sun, Y.; Shi, G.; Su, M.; Liu, Z.; Wang, H.; Deng, X. Short-term wind power interval prediction method using VMD-RFG and Att-GRU. *Energy* **2022**, *251*, 123807. [[CrossRef](#)]
31. Zhang, C.; Ji, C.; Hua, L.; Ma, H.; Nazir, M.S.; Peng, T. Evolutionary quantile regression gated recurrent unit network based on variational mode decomposition, improved whale optimization algorithm for probabilistic short-term wind speed prediction. *Renew. Energy* **2022**, *197*, 668–682. [[CrossRef](#)]
32. Kumar, A.; Zhou, Y.; Xiang, J. Optimization of VMD using kernel-based mutual information for the extraction of weak features to detect bearing defects. *Measurement* **2021**, *168*, 108402. [[CrossRef](#)]
33. Jiang, Y.; Zhu, Y.; Yang, X.; Jiang, X. Hankel-SVD-CEEMDAN improved threshold partial discharge feature extraction method. *Power Syst. Technol.* **2022**, *11*, 4557–4567. [[CrossRef](#)]
34. Yang, D.; Gao, Z.; Li, Y. Interval prediction of wind power based on bivariate empirical mode decomposition and least squares support vector machine. *Electr. Power Constr.* **2019**, *5*, 118–127. [[CrossRef](#)]
35. Cho, K.; Merriënboer, B.v.; Gülgehre, Ç.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [[CrossRef](#)]
36. Veeramsetty, V.; Reddy, K.R.; Santhosh, M.; Mohnot, A.; Singal, G. Short-term electric power load forecasting using random forest and gated recurrent unit. *Electr. Eng.* **2022**, *104*, 307–329. [[CrossRef](#)]
37. Li, C.; Guo, Q.; Shao, L.; Li, J.; Wu, H. Research on Short-Term Load Forecasting Based on Optimized GRU Neural Network. *Electronics* **2022**, *11*, 3834. [[CrossRef](#)]
38. 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. [[CrossRef](#)]
39. Tian, P.; Yu, Y.; Dong, M.; Jiang, Z.; Bao, P.; Wu, G.; Zhang, T.; Hu, P. A CNN-SVM-based fault identification method for high-voltage transmission lines. *Power Syst. Prot. Control* **2022**, *13*, 119–125. [[CrossRef](#)]
40. Zhang, Y.; Li, W.; Dong, F. Medium and Long-Term Power Demand Forecasting Based on DE-GWO-SVR. *Electr. Power* **2021**, *9*, 83–88. [[CrossRef](#)]
41. Koenker, R.; Bassett, G.W. Regression quantiles. *Econometrica* **1978**, *1*, 33–50. [[CrossRef](#)]
42. Shim, J.; Kim, Y.; Lee, J.; Hwang, C. Estimating value at risk with semiparametric support vector quantile regression. *Comput. Stat.* **2012**, *27*, 685–700. [[CrossRef](#)]
43. Zhang, Z.; Qin, H.; Liu, Y.; Yao, L.; Yu, X.; Lu, J.; Jiang, Z.; Feng, Z. Wind speed forecasting based on Quantile Regression Minimal Gated Memory Network and Kernel Density Estimation. *Energy Convers. Manag.* **2019**, *196*, 1395–1409. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.