Article

A Hybrid System Based on LSTM for Short-Term Power Load Forecasting

Yu Jin, Honggang Guo *, Jianzhou Wang and Aiyi Song

School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China;
jinyudc@dufe.edu.cn (Y.J.); wangjz@dufe.edu.cn (J.W.); aiyi971208@163.com (A.S.)

* Correspondence: ghg970612@163.com

Received: 21 October 2020; Accepted: 23 November 2020; Published: 26 November 2020

Abstract: As the basic guarantee for the reliability and economic operations of state grid corporations, power load prediction plays a vital role in power system management. To achieve the highest possible prediction accuracy, many scholars have been committed to building reliable load forecasting models. However, most studies ignore the necessity and importance of data preprocessing strategies, which may lead to poor prediction performance. Thus, to overcome the limitations in previous studies and further strengthen prediction performance, a novel short-term power load prediction system, VMD-BEGA-LSTM (VLG), integrating a data pretreatment strategy, advanced optimization technique, and deep learning structure, is developed in this paper. The prediction capability of the new system is evaluated through simulation experiments that employ the real power data of Queensland, New South Wales, and South Australia. The experimental results indicate that the developed system is significantly better than other comparative systems and shows excellent application potential.

Keywords: power load forecasting; hybrid analysis-forecast system; data preprocessing strategy; deep learning structure; optimization algorithm

1. Introduction

Along with fast-economic development, power enterprises continue to expand their construction scales, and the corresponding power grid structures and operation modes are gradually becoming diversified [1]. Because electricity is a special type of energy, people cannot store much electricity. This requires the power generation capacity of power generation enterprises and the power supply capacity of power supply companies to maintain a state of dynamic balance; otherwise, the lives of residents and the production of enterprises will be affected, potentially endangering the security and availability of the whole electrified wire netting system. Accurate power load prediction provides an important guarantee to ensure that the power supply and demand remain in a stable state. An accurate power load will yield significant economic benefits. It is estimated that every 1% increase in the accuracy of power consumption forecasting will save millions of dollars in operating costs [2]. Improving the prediction performance of electrical load prediction will not only provide a solid foundation for the smooth operation of the power grid but also provide theoretical support for power supply and dispatching plans.

The power load forecast is closely related to the dispatch and normal operations of national electricity consumption, which affects the lives and production activities of residents and the country overall [3]. The current main research work focuses on ultra-short-term and short-term power load prediction, which are hotspots that will allow academia and power companies to dynamically adjust their power generation plans and trading plans in the market environment [4]. This article mainly studies short-term power load forecasting.

Based on the principles and structures of the research methods, power load research can be distinguished into three categories: (i) Physical prediction methods, (ii) statistical prediction methods, and (iii) machine learning methods. The physical prediction model is a prediction model established by combining some physical characteristics and historical power load data. Most of the model assumes that the relationship between power load data and related physical information is still valid in future predictions, and the power load is predicted through intuitive analysis [5]. The main methods involve the unit consumption method, the load density method, and the elastic coefficient method [6]. For example, Yang et al. [7] predicted the power load distribution by calculating the load density and combining it with the characteristics of power consumption in the area. However, physical prediction methods require a great deal of observation data, which inevitably consumes exorbitant and difficult-to-obtain computing resources. Moreover, these physical methods are more appropriate for long-term power load predictions than short-term power load predictions. Statistical methods are more fit for short-term power load predictions. In recent years, traditional learning methods and traditional statistical systems have been broadly used for predicting short-term power loads. Considering previous time-series data of power loads, Cui et al. [8] developed a new load forecasting model by combining the grey linear regression model with the Markov chain, which overcomes the shortcomings of the traditional grey model that ignores the linear factors. Commonly used traditional statistical systems include autoregressive (AR) and autoregressive moving average (ARMA) models [9]. Song et al. [10] proposed a non-parametric hybrid model. The results show that the hybrid model relying on non-parameters is generally better than other models. However, this method also has some shortcomings. For example, in multi-step predictions, this model's prediction accuracy is low, and the time series prediction method currently widely used cannot take into account the influence of meteorological factors [11]. Although some researchers have proposed some methods to increase the adaptability of time series forecasting methods to meteorological changes, such methods still use the ARIMA model, which has insufficient explanatory power and cannot fundamentally improve the adaptability of the time series prediction method to meteorological changes [12].

In the 1990s, computers gradually entered all walks of life, and power load prediction technology was created in large quantities. Artificial neural network models are becoming increasingly popular for power load prediction. The artificial neural network (ANN) model has very strong nonlinear modeling capabilities and is a data-driven nonlinear adaptive method [13]. At the present time, the main algorithms for artificial neural networks are the back propagation neural network (abbreviated as BP) and Elman neural network (abbreviated as Elman), which can approximate any nonlinear function without knowing the relationship between the predictive model and the data [14]. Moreover, the support vector machine (SVM) originally employed by Vapnik and others of Bell Labs is extensively used in the domain of power load prediction and has been continuously improved by researchers. Some of the artificial intelligence technologies have been applied as follows. Guo et al. [15] suggested a generalized neural network based on using the local mean decomposition to make predictions. Xu et al. [16] proposed an improved RBF prediction model using a weighted fuzzy clustering algorithm to determine the center of the benchmark function. Li et al. [17] established a hybrid neural network model using an improved gray wolf bionic algorithm. The experimental results show that in power system prediction, the hybrid model significantly reduces the prediction errors compared to other comparative models. The effectiveness of the neural network method in power load forecasting has been extensively verified. To date, deep learning (DL) technology has been applied in the automation of many industries, such as image and audio detection [18]. Compared with the neural network, deep learning has a deeper hidden layer. This layer can make the computer perceive problems like human beings by simulating the connection of the human brain grid. At present, DL has become one of the most attractive technologies in short-term electrified power load prediction as a result of its excellent end-learning capacity and offers the most advanced forecasting performance [19]. Li et al. [20] combined the convolution neural network (CNN), long short-term memory neural network (LSTM), and gated recurrent unit (GRU) algorithm and proposed a prediction model based on deep learning

for power load forecasting in Beijing. Massaoudi et al. [21] combined the savitzky Golay filter with the bi-directional long-term memory neural network (BiLSTM) to predict the short-term power load. Experimental results show that the proposed model is highly effective.

Although the DL algorithm improves the prediction accuracy of traditional prediction models, a single traditional prediction system often has some prominent and obvious shortcomings, which leads to unsatisfactory prediction results. Therefore, our study focuses on the combined model [22]. However, modeling of the classical hybrid model still needs to be improved. The high volatility and randomness of the power load data will affect the learning ability of the prediction model and lead to poor prediction performance when using raw data without processing directly [23]. Hence, to reduce the random interference of the data sequence and improve the prediction performance, data preprocessing methods such as empirical mode decomposition (EMD) and the deep learning image noise reduction algorithm (DA) are applied to time series prediction [24,25]. For example, Ribeiro et al. [26] proposed an adaptive, decomposition, heterogeneous, and integrated learning model. In the data preprocessing stage, the super parameters of complementary ensemble empirical mode decomposition were optimized, and three machine learning models were calculated to predict the short-term electricity price in Brazil market. Stefenon et al. [27] proposed a method of feature extraction using wavelet energy coefficient and combining with LSTM to predict power insulator fault. The experimental results show that the method has good prediction accuracy. Although these methods improve the prediction accuracy to a certain extent, they still have some shortcomings. For example, the problem of mode aliasing often occurs in EMD, and the residual noise in DA cannot be processed [28,29]. On the other hand, the final result of the deep learning network depends to some extent on the initial random hidden layer nodes, input unit length, and model hyperparameter optimization method, which will affect the instability of the prediction. However, there is a contradiction between the number of hidden layer nodes in the network and training ability. Universally speaking, when the hidden layers are few, the prediction accuracy is also poor. To some extent, as the hidden layer nodes increase in number, the forecasting accuracy also improves. Nonetheless, this correlativity is limited. When reaching the apex, the predictive power decreases as the number of hidden layer nodes increases. Therefore, it is very important to determine the number of hidden layers and the length of the input cells [30]. By reviewing the previous literature, we found that the above prediction method still has some inherent defects [31]. The shortcomings of these systems are summarized in Table 1.

According to the above analysis, this paper proposes a new hybrid system that combines data preprocessing technology, the advanced deep learning prediction method, and the bionic optimization algorithm to further improve the short-term power load forecasting accuracy. More specifically, based on variational mode decomposition (VMD), the original power load series are decomposed and reconstructed in this paper to effectively remove the noise of the original load data and extract the data features effectively. Then, we apply the deep learning prediction method using the LSTM to predict the processed power load data. Finally, A calculation technique employing the binary encoding genetic optimization algorithm (BEGA) based on swarm intelligent evolution and a bionic strategy is proposed to find the optimal LSTM model's hidden layer nodes and input unit length. The main contributions and innovations of this research are as follows:

1. The original power load data is decomposed and reconstructed using VMD technology to extract the effective features of the data. This reduces the adverse effects of the instability and irregularity of the original load data on the forecasting model.
2. The long-term and short-term memory neural network is applied to forecast the power load data. This solves the problem where the time series depends on previous data and overcomes the low accuracy and poor stability of traditional models.
3. A binary encoding genetic algorithm is proposed to adaptively decide the hidden layer nodes and the length of the input data unit of the LSTM. This algorithm abandons the traditional decimal coding method and uses binary coding for integer optimization.

4. The adaptive moment estimation (Adam) algorithm is employed for optimizing the model's hyperparameters, instead of the traditional gradient descent algorithm and stochastic gradient descent algorithm (SGD). This improves the convergence speed and prediction stability of the model.
5. The prediction model optimized by the hybrid optimization method has high prediction accuracy and good stability, thereby effectively improving the accuracy of power load prediction.

Table 1. Comparison of the advantages and disadvantages of existing power load forecasting systems.

Category	Advantage	Disadvantage	Method Sample	Sample Advantage	Sample Disadvantage
Physical arithmetic	Simple model ideal, wide parameters range	Need a lot of observation data, consumes a lot of computing resources, More suitable for long-term power load forecast	Single consumption method	Separate analysis of different types of electricity consumption	When there are many influencing factors, the prediction accuracy is not high
			Elastic coefficient method	Reflects the relationship between economic growth rate and power consumption growth rate	The calculation is complex and requires accurate statistics on economic growth
Statistical strategies	Wide application, higher prediction accuracy	In multi-step prediction, the prediction accuracy of the model is bad	Time series (autoregressive moving average (ARMA), ARIMA)	Simple model assumptions, good self-fitness	The extrapolation effect is poor, reducing the prediction range
			Grey prediction system	Less modeling information and convenient operation	Low accuracy lack of systematizations
Machine learning	Wide application, Strong generalization and robustness	High complexity, high requirements of knowledge	Neural network (back propagation neural network (BP), support vector machine (SVM), RBF)	Excellent fitting effect nonlinear property	High the degree of data dependence
			Data denoising (empirical mode decomposition (EMD), deep learning image noise reduction algorithm (DA))	Compared with other methods, it is easy to understand	EMD has mode aliasing and DA has insufficient noise reduction
			Deep learning (convolution neural network (CNN))	Strong fault tolerance, simple human-computer interaction	Long CPU operation time

Note: With the advance of forecasting technology, few researchers use a single method to forecast the power load. In this table, different power load forecasting methods used in previous studies are detailed introduced.

The second part introduces the specific methods used for the proposed model, including data preprocessing technology, the LSTM prediction model, and the binary encoding genetic algorithm; we describe the framework of the VLG model in detail in Section 3. In Section 4, we conduct three different experiments from different angles and probe the experimental conclusions of the combined system and other systems. To further certify the accuracy and effectiveness of the new hybrid system, Section 5 offers a specific discussion. Finally, the results and conclusions are given in Section 6.

2. Related Theory

In this part, the methods adopted in the employed hybrid system are explained.

2.1. Variational Mode Decomposition (VMD)

VMD is a completely non-recursive adaptive signal processing technology employed by Dragomiretskiy et al. [32], which can effectively achieve the adaptive separation of signals in the frequency domain.

Step 1: For each mode, the analytical signal related to each mode is calculated by the Hilbert transition, and its single-sided spectrum can then be acquired [33]:

$$\text{Hilbert}f(t) = \frac{1}{\pi} \text{p.v.} \int_{\mathbb{R}} \frac{f(v)}{t-v} dv \quad (1)$$

Step 2: Adjust the estimated center frequency $e^{-j\omega_k t}$ by adding an exponential term to each mode [34]:

$$\left[\left(\delta(t) + \frac{j}{\pi t} \right) \mu_k(t) \right] e^{-j\omega_k t} \quad (2)$$

Step 3: Each mode can be closely centered around the center pulse frequency. Gaussian smoothness is used to estimate the bandwidth of the signal above so that a constrained variation problem can be obtained:

$$\text{Min} \left\{ \sum_K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * \mu_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (3)$$

$$\text{s.t.} \sum_K \mu_k = f_0 \quad (4)$$

where f is the original signal, μ_k is the modal function, and $\delta(t)$ is the Dirac distribution [35].

Step 4: The second penalty factor ensures that the signal still offers better reconstruction accuracy under high noise conditions. The Lagrange multiplier maintains strict constraints, and the augmented Lagrange formula is [36–38]

$$L(\{\mu_k\}, \{\varphi_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\psi(t) + \frac{j}{\pi t} \right) * \mu_k \right] e^{-j\varphi_k t} \right\|_2^2 + \|f(t) - \sum_k m_k(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_k m_k(t) \rangle \quad (5)$$

where α is the penalty factor, λ is the Lagrange factor, and L is an augmented Lagrangian multiplier. The solution to the original minimization problem (3) is transformed as a saddle point of the augmented Lagrangian L .

2.2. Long Short Term Memory Neural Network (LSTM)

The LSTM, as a special RNN, has strong processing ability for time series data and effectively overcomes the defects of gradient disappearance and gradient explosion in RNNs in machine learning. Figure 1 shows the schematic structure of the LSTM.

The three gates included in the LSTM all use the sigmoid function; the sigmoid function σ is [39]

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

Step 1: The LSTM must decide what information is invalid and discard that information from the unit state. The Sigmoid layer named “Forgotten Gate” f_t will do this part of the work. The expression of f_t is

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (7)$$

where h_{t-1} is the hidden state, and x_t is the input vector.

Step 2: LSTM determines what information remains in the cell state. First, the “input gate layer” i_t determines which values will be updated. Next, the candidate information $\tilde{C}_t$ is created through a

$\tanh$ neural network layer, and the input gate also reads h_{t-1} and x_t . Next, multiply i_t and $\tilde{C}_t$ to obtain the new information needed to remember the cell state. The i_t and $\tilde{C}_t$ expressions are as follows [40]:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (8)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (9)$$

Figure 1. Flowchart of the long short-term memory neural network (LSTM) algorithm.

Step 3: According to the current input variables x_t and h_{t-1} , the cell C can be used to remove outdated old information and add the new information needed and thereby update the cell state and obtain the new cell state [41,42]:

$$C_t : C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (10)$$

Step 4: Multiply the latest cell state C_t result by the output gate O_t vector to obtain the LSTM's final output state vector h_t :

$$\tanh h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (11)$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (12)$$

$$h_t = O_t * \tanh h(C_t) \quad (13)$$

where W_f , W_i , W_C , and W_o are the coefficient matrix, and b_f , b_i , b_C , and b_o are the bias vector.

2.3. Binary Encoding Genetic Algorithm (BEGA)

The BEGA algorithm is an improved genetic algorithm with binary coding. The genetic algorithm is an evolutionary principle that simulates “the survival of the fittest” phenomenon in nature [43]. The basic operation of the binary encoding genetic algorithm is as follows:

Step 1: *Binary encoding*. The four bases in the human chromosome are simulated with two numbers, 0 and 1, and the variable is described with a string of a certain length containing 0 s and 1 s.

Step 2: *Select the operation.* The selection operation of the genetic algorithm is a form of roulette. In accordance with the selection strategy of the fitness ratio, the selected probability P_i of each individual i is

$$f_i = k/F_i \quad (14)$$

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j} \quad (15)$$

where f_i is the fitness value, which is the root mean square error value between the power load forecast value and the true value under the LSTM neural network in this study.

Step 3: Cross operation. A pair of genetic sequences are cut individually via two-point crossover, and then the cut sequences are randomly combined, which makes the binary string sequence more likely to be transcoded.

Step 4: Mutation operation. Due to the particularity of binary coding, the method of bit-flip mutation is used in the mutation operation. For each gene value in the individual, the opposite value is taken according to the given mutation probability P .

Step 5: Decode. assuming that the binary code is converted into the decimal range of $[-a, a]$ and that b -bit precision is required, we need to discretize the interval into $(|-a| + a) \times 10^b$ numbers. Then, we convert the binary code x^{bin} into a decimal real number x^{dec} . Next, through formula decoding using

$$x = -a + x^{dec} * \frac{2a}{2a^b - 1} \quad (16)$$

we can obtain a real number in the $[-a, a]$ interval.

2.4. Adaptive Moment Estimation Optimizer (Adam)

Adam is a random objective function optimization algorithm based on a one-step degree and also an adaptive estimation algorithm based on a low-order matrix. This method has high calculation efficiency and small memory requirements and is very suitable for optimization problems with a large data volume or many parameters [44]. The algorithm is as follows:

Step 1: Given an objective function $J(\theta)$ with some parameters θ , calculate the gradient at time t :

$$g_t = \nabla_{\theta} J(\theta_{t-1}) \quad (17)$$

Step 2: Comprehensively, consider the gradient momentum of the previous time step and calculate and update the gradient mean m_t and exponential moving average v_t of the square of the gradient:

$$m_t = \varphi_1 m_{t-1} + (1 - \varphi_1) g_t \quad (18)$$

$$v_t = \varphi_2 v_{t-1} + (1 - \varphi_2) g_t^2 \quad (19)$$

where φ_1 is the exponential decay rate, which controls weight distribution, and φ_2 is the exponential decay rate. This rate impacts the process of the previous squared gradient.

Step 3: Since the initial setting of m_0 and v_0 is 0, here we need to adjust the deviation of the gradient mean m_0 and the exponential moving average v_0 of the square of the gradient to alleviate the impact of the deviation:

$$\hat{m}_t = m_t / (1 - \varphi_1^t) \quad (20)$$

$$\hat{v}_t = v_t / (1 - \varphi_2^t) \quad (21)$$

Step 4: Update parameters θ :

$$\theta_t = \theta_{t-1} - \alpha * \hat{m}_t / (\sqrt{\hat{v}_t} + \varepsilon) \quad (22)$$

where α represents the step size, which is updated according to $\alpha_t = \frac{\alpha \sqrt{1-\varphi_2^t}}{1-\varphi_1^t}$, and the built-in parameters of the Adam algorithm are set to $\varphi_1 = 0.9$; $\varphi_2 = 0.999$; $\varepsilon = 10^{-8}$.

3. The Formation of the Combined Forecasting System

In this study, a VLG hybrid system for short-term load forecasting is proposed to improve the forecasting accuracy. This system mainly uses VMD technology to remove the noise of the original sequence to obtain a stable time series, which is used as the input feature of the model training and provides a high-quality training set for the model. Here, BEGA is used to adaptively optimize the length L of the input data sample unit of the LSTM and the number N of the cell units, allowing the LSTM to achieve optimal performance. The analysis framework of the prediction system is shown in Figure 2. The 1-step is VMD data preprocessing module, the 2-step is BEGA and Adam optimization algorithm module, and the 3-step is LSTM prediction module. The specific instructions are as follows.

Figure 2. The flowchart of variational mode decomposition-binary encoding genetic optimization algorithm-LSTM (VLG) system.

3.1. Data Preprocessing Module

As the original load data without data processing for feature extraction have strong volatility and randomness, if these data are input into the model input directly, the prediction performance will decline [45]. Therefore, the advanced decomposition ensemble method is used to eliminate the high frequency noise in the power load series and extract valuable information components from the original power load data, which can effectively improve the prediction accuracy. In this paper, we use

VMD to preprocess the data, decompose the original time series into several intrinsic mode functions (IMFs), and then reconstruct those functions as the input series of the prediction model.

3.2. Optimization Algorithm Module

The module is mainly classified into two portions. The first part is used to optimize the selection of the input unit length L and the number of the number of hidden layer nodes N of the LSTM through a BEGA algorithm. In this part of the optimization, the binary genetic sequence is initialized. Through the genetic algorithm, L and N are changed within a certain range. Then, pass the changed L and N to the LSTM for training to calculate the loss function as

$$loss = \frac{1}{M} \sum_{m=1}^M (y_m - \hat{y}_m)^2 \quad (23)$$

where y_m is the true value, and $\hat{y}_m$ is the predicted value. The smaller the value of the loss function, the smaller the model prediction error will be. Next, based on the loss function value, update the population, and obtain L and N values that minimize the value of the loss function. Finally, the length L of the input unit and the number N of the number of hidden layer nodes that optimize the performance of the model are determined.

The second part is to optimize the hyperparameters in LSTM. Adam is an adaptive objective function optimization algorithm that adaptively calculates the learning rates of different parameters on the basis of the first and second-moment estimates of the gradient to improve the model convergence speed and provide better prediction accuracy.

3.3. Forecast Module

To make the prediction performance more accurate and stable, the VLG forecasting system is built to forecast the power load data. By constructing the index system, the prediction errors between the prediction results and the original power load data are evaluated.

4. Experiment and Evaluation

The accurate prediction of short-term power load data has considerable practical significance and is very important for ameliorating the prediction performance of the model. Therefore, in this paper, we propose a hybrid system to enhance the accuracy of short-term power load prediction and apply other traditional nonlinear forecasting models for comparison in this section to assess the effectiveness of the employed hybrid system. Taking the power load data of New South Wales, South Australia, and Queensland (30 min power load data for each season in 2013) as examples, the performance of the hybrid model is evaluated. This part mainly outlines the procedure and conclusions of the experiment.

4.1. Model Evaluation Indicators

To judge whether the performance of one model is better than that of another model, the model evaluation criteria play a vital role. However, there is currently no unified standard to evaluate prediction performance, nor can a single model evaluation index fully reveal the excellent performance of a given model [46]. Therefore, we use the method of constructing an index evaluation system to evaluate the present model. This index system mainly includes the mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). The smaller the values of these indicators are, the more accurate the prediction and the better the model performance will be. In contrast, the higher the value of R^2 , the better the prediction model performance. The particular content and calculation formula for each index is introduced in Table 2.

Table 2. The evaluation metrics.

Metric	Definition	Equation
MAPE	Average absolute percentage errors	$MAPE = \frac{1}{N} \sum_{n=1}^N \left \frac{\hat{y}_i - y_i}{y_i} \right \times 100\%$
MAE	Average absolute error	$MAE = \frac{1}{N} \sum_{n=1}^N \hat{y}_i - y_i $
RMSE	Root Mean Square Error	$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (\hat{y}_i - y_i)^2}$
R ²	Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^N (F_i - \bar{F})^2}{\sum_{i=1}^N (F_i - \bar{F})^2}$

Note: In this table, $\hat{y}_i$ represents the predicted value of the system, y_i is the true value of the power load data, and $\bar{y}$ is the average value of the sequence. The calculation equation is $\bar{y} = 1/N(\sum_{i=1}^N y_i)$.

4.2. Experimental Setup

To estimate the performance of the VLG hybrid system for short-term power load predictions, the hybrid system is compared with the traditional models for each component to explore the impact of the combination model on short-term power load forecasting. We divide the experiment into three parts: Experiment 1, Experiment 2, and Experiment 3. In Experiment 1, The prediction performance of the model using VMD and other data preprocessing techniques is evaluated. Experiment 2 compares the performance of the VLG hybrid system with the performance of different neural networks and nonlinear prediction models in predicting the short-term power load. Experiment 3 discusses the impact of the adaptive optimization algorithm BEGA on model performance.

4.3. Data Description

In the experiment, since the power load data will change with the seasons, we selected the 2013 power load data of New South Wales, South Australia, and Queensland in Australia as the experimental verification data. We used the data from January 1 to January 31 as the summer dataset, the data from March 1 to March 31 as the autumn data set, the data from June 1 to June 30 as the winter dataset, and the data from September 1 until September 30th as the spring data set. In addition, the power load data set was given a time interval of 30 min. Among the data points included in the data set of each season, the first 80% were selected as the training set and the last 20% as the test set. Specifically, the lengths of the training set and test sets in the summer and autumn data sets of the three states were 1190 and 298, respectively. Correspondingly, the lengths of the training set and test set for the spring data set and winter data were 1152 and 288, respectively. When constructing the model input vector, we adopted a rolling acquisition mechanism. In other words, $\{x(1), x(2), \dots, x(t-1), x(t)\}$ was used as the basis of the latter data $x(t+1)$. Table 3 presents the descriptive statistics of the original power load data set, training data set, and test data set for the NSW seasonal data.

Table 3. Statistical description of the NSW seasonal power load data (MW).

Season	Data	Number	Min	Max	Mean	Standard
Summer	All samples	1488	5622.05	13787.85	8351.85	1519.63
	Training set	1190	5622.05	13787.85	8409.83	1584.77
	Testing set	298	5909.89	10928.47	8120.35	1200.07
Autumn	All samples	1488	5449.59	10724.86	7909.03	1166.83
	Training set	1190	5689.53	10080.21	7959.44	1104.36
	Testing set	298	5449.59	10724.86	7707.75	1372.34
Winter	All samples	1440	5997.81	11553.75	8602.26	1208.75
	Training set	1152	5997.81	11553.75	8562.20	1213.52
	Testing set	288	6191.79	11537.78	8762.52	1177.98
Spring	All samples	1440	5661.39	9916.19	7520.16	877.59
	Training set	1152	5661.39	9916.19	7543.92	868.22
	Testing set	288	5699.65	9081.59	7425.11	909.46

4.4. Model Parameter Setting

This section outlines the parameter settings of each component of the proposed hybrid prediction model VLG, including the length of the input unit, the initial learning rate, and other parameter settings of the LSTM, BEGA optimization algorithm, and Adam optimizer.

4.4.1. Parameter Setting of the ANN

Before the experiment, the parameters in the model and optimization algorithm must be defined. For the deep learning algorithm LSTM, the length of the input unit L and the number of hidden layer nodes N are optimized by the BEGA algorithm. The input unit length of BP is 5, and the number of hidden layer nodes is 11. See Table 4 for the other parameter settings.

Table 4. Parameter settings of the artificial neural network model.

Parameters	LSTM	BP
Length of input units	Based on BEGA algorithm / 5	5
Number of hidden layer nodes	Based on BEGA algorithm / 8	11
Objective function	MSE	MSE
Activation function	Sigmoid	PURELIN
Epochs	200	200
Initial learning rate	0.001	0.0001

Note: The input unit length of the LSTM model without the BEGA optimization algorithm is 5, and the number of hidden layer nodes is 8.

4.4.2. Parameter Settings of the Optimization Algorithm

Setting the parameters of the BEGA optimization algorithm and Adam optimizer is very important for the short-term power load forecasting accuracy. The specific parameter settings we studied are shown in Table 5.

Table 5. Model parameters.

Model	Parameters	Default Value
BEGA	Maximum number of iterations	30
	Binary code length	15
	Population number	10
	Fitness function	MSE
	Select operation	roulette wheel selection
Adam	Initial learning rate	0.001
	α	0.001
	β_1	0.9
	β_2	0.999

Figure 3 shows the power load data for South Australia, Queensland, and New South Wales. It is not difficult to see that the data fluctuations in New South Wales and Queensland are relatively stable, with South Australia fluctuating the most frequently. In the Figure 3, (a) is the situation of the three study regions. (b) is the feature selection part, which shows the decomposition results of power load series, and IMFs is arranged in descending order of frequency. Then, the high frequency IMF is removed and the remaining IMFs are reconstructed to obtain the optimal input sequence. The characteristics of the data series are obviously improved than the original ones.

Figure 3. Site data and data preprocessing results. (a) the situation of the three study regions. (b) Data preprocessing process.

4.5. Experiment I

In the previous section, we noted that the noise reduction technology of time series plays a crucial role in the performance and prediction accuracy of the model. Based on the research of previous scholars, noise reduction technology can mainly be divided into three major categories: empirical mode decomposition technology (EMD), denoising autoencoder technology (DA), and VMD, which is used in our proposed hybrid system. The principle of VMD was discussed in the previous section. EMD technology involves decomposing the internal model function from the original signal to

obtain a series of different intrinsic mode functions (IMFs). This method can decompose non-stationary and non-linear signals into stationary signals with different time scales. The denoising autoencoder (DA) is a type of unsupervised learning that uses an encoder and a decoder. White noise or Gaussian noise is added to the original sequence, and the neural network is continuously iterated to obtain a dimensionality reduction feature expression of the data to achieve the effect of feature extraction.

Experiment 1 aims to verify the influence of the noise reduction data set on the model and the performance of the VMD noise reduction method. In this way, the prediction accuracy of VMD-LSTM and other neural network models based on EMD and DA (i.e., EMD-BP, EMD-LSTM, DA-BP, and DA-LSTM) is compared. We performed three experiments on each model to assess the prediction accuracy of each model, and the specific prediction results based on NSW spring dataset are shown in Table 6.

Table 6. Comparison of the prediction performance of models using different data preprocessing systems.

Model	MAE			RMSE			MAPE(100%)			R ²		
	NSW	SA	QLD	NSW	SA	QLD	NSW	SA	QLD	NSW	SA	QLD
LSTM	83.91	32.65	51.08	110.57	50.15	67.69	0.8478	2.3571	0.8475	0.9919	0.9744	0.9920
BP	100.49	35.30	86.59	125.95	52.63	74.80	1.0305	2.5735	0.9509	0.9887	0.9743	0.9896
EMD-LSTM	51.95	13.25	30.99	63.05	17.21	37.89	0.5476	1.0236	0.5231	0.9963	0.9898	0.9949
EMD-BP	65.89	22.31	38.01	75.11	29.13	46.31	0.6558	1.7704	0.6846	0.9932	0.9823	0.9921
DA-LSTM	61.09	20.69	38.66	77.72	29.62	42.86	0.7646	1.6430	0.7733	0.9925	0.9836	0.9934
DA-BP	68.30	19.79	41.31	88.57	32.38	54.61	0.8553	1.5923	0.8062	0.9906	0.9841	0.9924
VMD-LSTM	49.42	11.69	28.75	62.17	15.60	34.47	0.4922	0.9352	0.4859	0.9971	0.9910	0.9967
VMD-BP	52.04	19.77	35.54	94.89	25.48	42.78	0.8251	1.5808	0.6421	0.9921	0.9802	0.9925

The details of the experiment are as follows:

- The comparison results of the evaluation index system of VMD-LSTM and other hybrid systems are introduced in this table. The VMD-LSTM shows better prediction accuracy and prediction performance on most evaluation indicators. For the comparison between the single LSTM model and the VMD-LSTM system, it is obvious that the evaluation index of the VMD-LSTM system is superior to that of the LSTM model in all cases. At the same time, the prediction accuracy of all noise reduction models is higher than that of the model based on original data, which indicates that data preprocessing is indispensable for power load prediction.
- In the case, we compare the VMD-LSTM model and several other hybrid model methods. Among the performance test index values of the experiments in various regions, the VMD-LSTM model offers the best MAPE results, with 0.4859%, 0.9352%, and 0.4922%. Secondly, the models based on prediction accuracy are VMD-LSTM, EMD-LSTM, EMD-BP, VMD-BP, DA-LSTM, and DA-BP, in order from high to low. Among the six models, VMD-LSTM has the best prediction accuracy. The coefficient of determination (R^2) reflects the difference in the performance of the prediction model from the fit. In this experiment, the R^2 of VMD-LSTM is the best with 0.9971, 0.9910, and 0.9967 in the three states. We also certify the effectiveness of the noise reduction model VMD employed in this paper.
- The previous time-series data denoising technology is also applied to the power load, short-term wind speed, and stock prediction models. Most of these models only discuss the improvement of model accuracy and performance via noise reduction technology but do not discuss the new sequence obtained after using the noise reduction method correlation with the original time series. Therefore, through the gray correlation method (GC) and the method of calculating the Pearson correlation coefficient (PE) and Spearman correlation coefficient (SP), the differences between different noise reduction methods are discussed from the perspective of the correlation between the new sequence and the original sequence. Detailed calculation results are given in Table 7.

Table 7 shows the correlation between the new sequence obtained by different noise reduction methods and the original power load data, in which the new sequence obtained using the VMD method has the highest correlation with the original sequence. In summary, using the VMD noise reduction method to process data not only performs well in improving the prediction accuracy of the model but also maintains more original information in the sequence, making it a more suitable method for the data preprocessing of short-term power load data.

Table 7. Correlation coefficient between new sequence and original sequence.

Model	VMD			EMD			DA		
	NSW	SA	QLD	NSW	SA	QLD	NSW	SA	QLD
GC	0.8953	0.785	0.749	0.885	0.779	0.725	0.752	0.767	0.702
SP	0.999	0.990	0.999	0.998	0.990	0.999	0.996	0.986	0.997
PE	0.999	0.987	0.999	0.999	0.986	0.999	0.995	0.986	0.996

Remarks:

Based on the above experiments, the employed VMD-LSTM combined system has the highest prediction performance, and Queensland has the highest prediction accuracy, with a MAPE value of 0.4922%. The average prediction accuracy of the three regions is 0.6378%. This shows that the system has high prediction accuracy and excellent stability. Moreover, in the test index based on the correlation measure, VMD remains superior to other noise reduction models, further verifying the effectiveness of the system.

4.6. Experiment II

Power load data are very sensitive to natural factors such as the season and climate and is one of the dominant factors affecting the fluctuation characteristics of the power load. In this part, the 30 min power load data of New South Wales, Queensland, South Australia from March to April, June to July, September to October, and December to January in 2013 are used as seasonal data for this area. A comparison of the prediction performance differences between the different predictive methods based on the power load data of New South Wales is shown in Table 8. Moreover, Tables 9 and 10 provide a prediction performance comparison between the traditional predicted model based on the seasonal data of Queensland and South Australia and the combination model based on the LSTM employed in this study. In general, the VLG system employed in this research provides better performance than traditional prediction methods.

(a) For New South Wales, based on the annual average of the prediction results, BP, PSO-BP, LSTM, and our proposed hybrid model VLG obtained good accuracy results for all evaluation indicators. The ARIMA model also achieved a prediction effect with a MAPE value of less than 2%. However, the MAPE values of the other two models were higher than 10%, and the performance was poor. The ARIMA model has always been considered to provide superior performance in predicting power load data, and its MAPE value is lower than the values of the other traditional models discussed in this article, which is logical. However, the ARIMA model is not beneficial for long-term rolling forecasting and has certain disadvantages. The BP model and the optimization system based on the BP neural network design have always had good prediction performance in many experiments related to power load prediction. However, their MAPE values are still nearly three times higher than those of our proposed hybrid model VLG, which shows that our proposed hybrid model provides outstanding performance in short-term power load prediction.

For seasonal data, the proposed system obtained the best performance for each seasonal data set. For spring data, the MAPE value obtained by the VLG hybrid system was 0.3081%. Among these methods, the ARIMA, BP, PSO-BP, and LSTM models all perform better. The forecasting accuracy of the VLG prediction model was the highest, with a MAPE of 0.3081%, RMSE of 30.27, R^2 of 0.9979, and MAE of 24.88. In terms of the other seasonal feature data, the VLG model also obtained the best results,

with corresponding MAPE values of 0.4271%, 0.2724%, and 0.3717%. Among them, the forecasting accuracy of the autumn feature data set was the highest in each model, while the forecasting accuracy of the general system of the winter feature data set was worse than that of other seasons, which indicates indirectly that the forecasting of the power load is affected by regional and seasonal factors.

(b) For Queensland, from the perspective of annual average forecasting accuracy, the proposed combined system is still better than other classic models in prediction accuracy, with a MAPE value of 0.3486%. Among the remaining eight models, the forecasting accuracy of the VMD-LSTM system ranks second. The performance of the PSO-BP optimization model and the original LSTM model is similar, and the prediction accuracy is excellent. However, the MAPE of the hybrid system's VLG proposed in this paper was 0.1624%, 0.5052%, and 0.5499% lower than the values of the above model, respectively. Compared with the feature data of South Wales, the prediction accuracy of Elman and RBF models is improved but is still not ideal compared to the other models. For the seasonal data in Queensland, the VLG model still achieved the best results for the seasonal data, with corresponding MAPE values of 0.2602%, 0.3718%, 0.3101%, and 0.4524%, respectively. From the perspective of the coefficient of determination, the goodness of fit of the VLG combination forecasting model was the most significant, and R^2 values of the four seasonal data were determined as 0.9983, 0.9953, 0.9979, and 0.9942, respectively. Here, the general prediction accuracy for the winter feature data set was worse than that of other seasons. The performance rankings of each model differed little from the rankings based on the South Wales feature dataset.

(c) For South Australia, from the perspective of the annual average prediction accuracy, in all prediction models, the performance evaluation index values calculated by the VLG hybrid model were significantly better than the performance index values calculated by the other prediction model processing methods, and its MAPE value was 0.9800%. The prediction results of the characteristic data of South Australia were slightly lower than those of the models of South Wales and Queensland, but the prediction accuracy of the VLG, VMD-LSTM, LSTM, and PSO-BP models was still excellent. In comparison with the other three systems, the MAPE value of the VLG hybrid model system decreased by 0.4066%, 1.2675%, and 1.1927%, respectively. In terms of seasonal feature data, like with the prediction results of Queensland and South Wales, the VLG hybrid model provided the smallest prediction errors and the greatest point prediction accuracy in four-season feature data prediction compared with other traditional models. Figure 4 shows the performance comparison between the VLG prediction system and the seven comparison models based on the spring data set of South Australia. In addition, Figure 5 shows the fit between the predicted value of VLG and the actual value of power load based on the four seasonal data sets in South Australia.

Moreover, based on the seasonal feature data of the three states, the seasonal temperature and regional climate were some of the most important factors affecting the performance of short-term power load predictions. Among them, autumn was often the season with the highest power load forecast accuracy. On the contrary, the prediction performance of the three states based on the winter feature data was lower than that of the other three seasons. In terms of regional differences, the annual average forecast performance gap between New South Wales and Queensland was not large, with MAPE values of 0.3717% and 0.3486%. South Australia's annual average forecast performance was relatively poor, with a MAPE value of 0.9800%, which may be related to the South Australia power load data set.

Remark:

Based on Experiment 2, the performance evaluation value calculated by the hybrid model VLG is superior to the performance evaluation value calculated by any classic single-item model and hybrid model. Therefore, the experimental conclusions indicate that the employed VLG hybrid system performs well in short-term power load predictions. Simultaneously, the seasonal climate and other factors have a certain impact on power load forecasting.

Table 8. The performance indicators of the VLG model based on seasonal data from New South Wales compared with other traditional power load forecasting methods.

Model	Spring				Summer				Autumn				Winter				Annual Mean			
	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE
Elman	19.1725	0.8506	1691.73	1457.86	8.1867	0.9198	906.39	714.14	6.7683	0.9337	671.54	508.20	8.3994	0.9131	612.39	763.87	10.6317	0.9043	970.51	901.85
RBF	19.4860	0.8644	1707.73	1477.86	8.3402	0.9145	922.56	725.44	6.9717	0.9288	682.04	520.26	8.3686	0.9177	612.06	766.47	10.7916	0.9064	981.09	872.51
ARIMA	1.7767	0.9812	121.86	117.85	1.4676	0.9817	150.69	119.91	1.5563	0.9812	144.49	114.45	1.9512	0.9810	139.23	126.23	1.6880	0.9813	139.06	119.61
SVM	2.0421	0.9792	176.34	158.17	1.7475	0.9751	119.33	112.19	1.8620	0.9807	130.54	121.22	2.3122	0.9759	196.92	179.83	1.9910	0.9777	155.78	142.85
BP	1.2493	0.9836	127.40	101.96	1.2646	0.9833	138.28	103.55	1.2846	0.9830	126.88	98.95	1.1503	0.9839	115.42	89.81	1.2372	0.9835	126.99	98.56
PSO-BP	1.1766	0.9835	132.91	90.41	0.7660	0.9942	86.28	65.77	1.0928	0.9854	113.07	80.16	1.0047	0.9903	161.11	67.09	1.0100	0.9887	123.34	75.85
LSTM	0.8942	0.9944	90.83	71.91	0.9548	0.9934	112.82	87.33	1.0360	0.9877	103.10	78.48	1.0091	0.9852	107.30	72.56	0.9735	0.9902	103.51	77.57
VMD-LSTM	0.5032	0.9951	49.16	40.84	0.6901	0.9947	72.96	60.83	0.4762	0.9952	46.33	28.45	0.5423	0.9943	52.14	34.69	0.5530	0.9949	55.15	41.20
VLG	0.3081	0.9979	30.27	24.88	0.4271	0.9972	48.05	38.53	0.2724	0.9981	26.66	20.76	0.4792	0.9970	44.73	35.48	0.3717	0.9976	37.43	29.91

Note: The above tables report the details of different forecasting methods. Three indicators were selected—MAE, RMSE, and MAPE—to verify the performance of each model. The specific equations of the three indicators are $MAE = 1/N(\sum_{i=1}^N |\hat{y}_i - y_i|)$, $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$, $MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right|$, and $R^2 = 1 - \frac{\sum_{i=1}^N (F_i - A_i)^2}{\sum_{i=1}^N (F_i - \bar{F})^2}$. Here, bold numbers indicate that the index values of this system are superior to those of the other system.

Table 9. The performance indicators of the VLG model based on seasonal data from Queensland compared with other traditional power load forecasting methods.

Model	Spring				Summer				Autumn				Winter				Annual Mean			
	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE
Elman	7.4343	0.9212	531.83	430.17	6.8212	0.9314	514.17	392.03	7.1909	0.9117	538.69	404.58	4.3644	0.9504	300.81	252.10	6.4527	0.9287	471.37	369.72
RBF	7.4613	0.9209	534.82	431.19	6.9253	0.9286	523.02	398.09	7.2187	0.9106	545.30	406.02	4.3591	0.9510	298.00	251.76	6.4911	0.9278	475.28	371.76
ARIMA	1.9497	0.9801	137.21	107.65	2.1738	0.9664	133.49	122.18	2.1513	0.9658	136.75	123.22	1.8488	0.9809	112.16	127.64	2.0309	0.9733	129.90	120.17
SVM	1.4228	0.9885	113.49	101.34	1.6231	0.9790	124.26	112.32	1.2141	0.9865	101.93	89.44	1.3952	0.9860	106.44	94.76	1.4138	0.9850	111.53	99.47
BP	0.9847	0.9896	74.80	59.45	1.1686	0.9862	85.89	67.20	0.9670	0.9891	73.76	53.29	0.927	0.9885	68.63	53.65	1.0118	0.9884	75.77	58.39
PSO-BP	0.8657	0.9910	73.07	48.37	0.7973	0.9921	61.26	45.14	0.8817	0.9902	69.80	49.31	0.8704	0.9907	99.11	48.36	0.8538	0.9910	75.81	47.79
LSTM	0.8475	0.9919	67.69	51.08	0.9781	0.9897	70.98	56.83	0.8740	0.9913	65.07	49.20	0.8944	0.9905	64.60	50.46	0.8985	0.9909	67.08	51.89
VMD-LSTM	0.4859	0.9946	34.47	28.75	0.4453	0.9938	32.47	25.81	0.5228	0.9948	35.36	29.19	0.5912	0.9941	53.21	48.64	0.5110	0.9943	38.87	33.09
VLG	0.2602	0.9983	20.26	15.53	0.3718	0.9953	28.46	21.67	0.3101	0.9979	21.60	17.30	0.4524	0.9942	30.96	25.25	0.3486	0.9964	25.32	19.93

Note: The above tables report the details of different forecasting methods. Three indicators were selected—MAE, RMSE, and MAPE—to verify the performance of each model. The specific equations of the three indicators are $MAE = 1/N(\sum_{i=1}^N |\hat{y}_i - y_i|)$, $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$, $MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right|$, and $R^2 = 1 - \frac{\sum_{i=1}^N (F_i - A_i)^2}{\sum_{i=1}^N (F_i - \bar{F})^2}$. Here, bold numbers indicate that the index values of this system are superior to those of the other system.

Table 10. The performance indicators of the VLG model based on seasonal data from South Australia compared with other traditional power load forecasting methods.

Model	Spring				Summer				Autumn				Winter				Annual Mean			
	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE	MAPE (%)	R ²	RMSE	MAE
Elman	24.1390	0.7494	404.81	331.16	11.4670	0.8821	216.55	171.91	15.2161	0.8346	210.36	166.59	12.4362	0.8744	222.15	165.58	15.8146	0.8351	263.46	208.81
RBF	24.6208	0.7501	407.97	336.96	11.6931	0.8796	218.64	174.94	15.5422	0.8339	210.86	170.14	12.5620	0.8706	220.51	166.75	16.1045	0.8336	264.49	212.19
ARIMA	2.2543	0.9784	28.02	35.12	1.5754	0.9802	29.11	21.20	2.7415	0.9737	47.22	39.03	2.9042	0.9721	55.12	39.93	2.3687	0.9761	39.86	33.82
SVM	2.7534	0.9728	60.31	47.98	2.1322	0.9745	41.33	29.86	2.6262	0.9749	55.91	42.64	2.7440	0.9787	52.35	34.22	2.5639	0.9752	52.475	38.68
BP	2.5735	0.9736	52.63	35.30	2.2731	0.9703	48.26	29.43	2.5139	0.9778	50.11	36.12	2.606	0.9793	57.42	37.84	2.4916	0.9755	52.11	34.67
PSO-BP	2.1429	0.9781	51.53	29.82	1.8272	0.9820	45.49	27.61	2.4783	0.9789	46.91	31.36	2.2424	0.9801	49.50	28.20	2.1727	0.9798	48.35	29.24
LSTM	2.3571	0.9769	50.15	32.65	2.1480	0.9712	53.52	34.57	2.3904	0.9795	45.24	30.27	2.0946	0.9814	45.57	27.27	2.2475	0.9773	48.62	31.19
VMD-LSTM	1.2135	0.9885	19.74	16.22	0.9312	0.9907	19.23	14.11	1.1885	0.9894	18.88	14.52	1.3132	0.9862	44.21	27.54	1.1616	0.9887	25.51	18.09
VLG	0.9816	0.9912	17.62	13.50	0.8991	0.9919	18.13	13.98	0.8212	0.9923	13.93	10.19	1.2181	0.9896	25.41	20.21	0.9800	0.9913	18.77	14.47

Note: The above tables report the details of different forecasting methods. Three indicators were selected—MAE, RMSE, and MAPE—to verify the performance of each model. The specific equations of the three indicators are $MAE = 1/N(\sum_{n=1}^N |\hat{y}_i - y_i|)$, $RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (\hat{y}_i - y_i)^2}$, $MAPE = \frac{1}{N} \sum_{n=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right|$, and $R^2 = 1 - \frac{\sum_{i=1}^N (F_i - A_i)^2}{\sum_{i=1}^N (F_i - \bar{F})^2}$. Here, bold numbers indicate that the index values of this system are superior to those of the other system.

In Figure 6, (a) is the radar map of the comparison of annual average MAPE of different prediction models in three regions. (b) gives the prediction performance comparison of different data preprocessing methods. (c) shows the outstanding advantages of the BEGA optimization algorithm and the prediction results of VLG prediction system. (d) shows the three indicators of the classic prediction model increase by percentage compared with the VLG prediction system, in New South Wales.

Figure 4. The fitting situation and error description of the VLG prediction system and seven comparative prediction models based on spring dataset of South Australia.

Figure 5. The fit of the VLG prediction system under the training set and test set based on the four seasonal data sets of South Australia.

Figure 6. Comparison results of the hybrid forecasting models. (a) Comparison of annual average MAPE of different prediction models. (b) Compared results with different decomposed approaches. (c) Compared results of the BEGA algorithm. (d) performance demonstration of different prediction models based on NSW.

4.7. Experiment III

The results of Experiment 2 show that the hybrid model based on LSTM has excellent performance in short-term power load prediction. However, LSTM needs one to manually configure the input unit length L and the number of hidden layer nodes N . These configured parameters also largely determine the model's ability to engage in short-term power load forecasting. There is no fixed rule for determining appropriate parameters when the LSTM predicts a time series. Therefore, common solutions include using repeated trials or enumeration methods to obtain appropriate parameters for accurate prediction accuracy. However, methods such as enumeration consume considerable time and may not necessarily select the best parameters. On the other hand, for the LSTM network, the final values of the hyperparameters of each neural unit and gate structure also have a significant impact on the prediction performance. Usually, the gradient descent method

(GD) or stochastic gradient descent method (SGD) is used in an experiment to find and select these hyperparameters. The gradient descent method is prone to problems such as a local optimal solution and slow convergence. Although the prediction results of the VMD-LSTM were satisfactorily evaluated, considering the deficiencies of the LSTM and SGD algorithms, we do not consider the prediction performance to have reached its optimal value. In response, the VLG hybrid system uses a binary encoding genetic algorithm (BEGA) and the Adam optimizer to solve these problems.

In this part of the experiment, we focus on the hybrid model VMD-LSTM and the original LSTM neural network and use the Adam optimizer and BEGA for parameter optimization. Here, the performance of the feature data sets of each season in New South Wales under different configurations, with different numbers of input units L and numbers of hidden layer nodes N , was studied. That is, the experience of previous scholars was used to determine the two different sets of L and N values to verify the effects of the BEGA on model accuracy. The specific experimental steps are shown in Table 11. At the same time, using the spring data of New South Wales as the data set, the Adam optimizer is compared with the stochastic gradient descent method to discuss the differences in prediction performance produced by different hyperparameter optimization methods. The specific experimental steps are shown in Table 12.

Table 11. Comparison of the results for different input unit lengths and numbers of cell units based on the NSW seasonal dataset.

Model	Evaluation Parameters	MAPE	MAE	RMSE	R ²
Season: Spring		The optimal parameters: 12–8			
VMD-LSTM	BEGA	0.2602	15.53	20.26	0.9986
	5–8	0.3612	22.01	28.12	0.9973
	16–12	0.5194	30.50	35.13	0.9949
Season: Summer		The optimal parameters: 1–8			
VMD-LSTM	BEGA	0.3718	21.67	28.46	0.9971
	5–8	0.6480	37.35	45.24	0.9944
	16–12	0.5034	28.95	35.03	0.9952
Season: Autumn		The optimal parameters: 25–12			
VMD-LSTM	BEGA	0.3101	17.30	21.60	0.9978
	5–8	0.5785	30.68	37.96	0.9940
	16–12	0.3554	19.22	23.75	0.9969
Season: Winter		The optimal parameters: 4–6			
VMD-LSTM	BEGA	0.3758	20.90	25.77	0.9965
	5–8	0.5828	32.69	40.98	0.9938
	16–12	0.5521	30.63	37.55	0.9942

As can be determined from Table 11, the VLG model using the binary encoding genetic algorithm provides better prediction results than the model selected via empirical summary in each season data set. Based on the seasonal feature data set, the BEGA optimal input unit length L and the number of hidden layer nodes N are 12–8, 12–8, 25–12, and 4–6, respectively, and the prediction accuracy MAPE values are 0.2602%, 0.3718%, 0.3101%, and 0.3758%, respectively. It can be seen from the MAPE values that using the BEGA algorithm to select the optimal unit length L and the number of cell units N can increase the model accuracy by up to 30%. Based on a comparison of the RMSE, MAE, and R^2 values, the performance improvement of the VLG system in the spring data set was the highest, with MAE, RMSE, and R^2 values of 15.53, 20.46, and 0.9986, respectively. Compared with the experience summary method, the MAE and RMSE average decline was 10.72, 15.37. In this four-season feature data set, our proposed VLG model using the BEGA algorithm has very good prediction performance and can accurately predict future short-term power load changes.

Table 12. Comparison of the performance results of different optimizers based on the NSW spring dataset.

Forecasting Model	Optimizer	Metric	Parameters			
			BEGA (4–6)	12–8	8–10	16–12
LSTM	Adam	MAPE	0.8744	1.1920	0.9580	0.8939
		MAE	66.22	89.74	74.80	68.01
		RMSE	86.79	114.19	92.12	84.74
		R ²	0.9926	0.9874	0.9901	0.9913
LSTM	SGD	MAPE	1.071	1.2219	1.2685	1.2840
		MAE	81.34	91.59	96.27	97.95
		RMSE	109.94	118.64	123.39	126.43
		R ²	0.9885	0.9870	0.9868	0.9866
VMD-LSTM	Adam	MAPE	0.2281	0.2947	0.3671	0.3609
		MAE	19.58	22.35	28.53	19.21
		RMSE	22.89	27.02	33.65	35.50
		R ²	0.9987	0.9981	0.9971	0.9972
VMD-LSTM	SGD	MAPE	0.6253	0.7022	0.8376	0.9021
		MAE	60.04	65.98	61.64	66.66
		RMSE	48.52	51.61	77.12	82.78
		R ²	0.9935	0.9928	0.9916	0.9903

Table 12 shows the performance advantages and disadvantages of different optimizers based on LSTM and VMD-LSTM systems. The results show that the prediction values of LSTM networks with different structures are quite different. Through a longitudinal comparison, it can be seen that using the Adam optimizer is much better than using the random gradient descent for prediction. Under the same data set, the MAPE values of the LSTM and VMD-LSTM models using the Adam optimizer were 0.8744% and 0.2281%, while the MAPE values of the LSTM and VMD-LSTM models using the SGD algorithm were 1.071% and 0.6253%. Thus, compared to using the Adam optimizer, employing the SGD algorithm reduced the performance by 22% and 174%, respectively. This result is enough to show the importance of using the Adam optimizer for model prediction accuracy and that this optimizer significantly improves the stability and accuracy of the predictions. Moreover, through a horizontal comparison, it can be further shown that using the optimal number of input units L and the number of cell units N (4–6) determined by the BEGA algorithm can give the model the best performance in the field of measurement precision and stableness, thereby significantly improving model performance.

Remarks:

For prediction performance and prediction accuracy, using the proposed BEGA algorithm and Adam optimizer is significantly superior to applying the evaluation index value determined by summarizing experience and using the stochastic gradient descent method. This further confirms that the VLG hybrid system has outstanding performance in future short-term power load predictions.

5. Discussion

This section provides an in-depth exploration of the above experimental results—that is, the availability of the hybrid system, the performance differences between the optimization algorithm hybrid model used and other optimization hybrid prediction systems, and the improvements of the evaluation index of the proposed hybrid system.

5.1. Effectiveness of the Proposed System

First, the predicted error is a vital indicator to estimate the performance of the prediction system. In this section, the validity of the hybrid model developed by the Diebold-Mariano test is tested, and the significance level of the prediction error of the different models is demonstrated via the hypothesis test method. The VLG system is then compared with other models. The Diebold-Mariano test evaluates

the differences between different prediction systems according to the error of the system prediction performance [47]. The null hypothesis H_1 and the alternative hypothesis H_0 are presented in Equation (25) and (26):

$$H_1 : E_a[L(err_i^1)] \neq E_b[L(err_i^2)] \quad (24)$$

$$H_0 : E_a[L(err_i^1)] = E_b[L(err_i^2)] \quad (25)$$

$$DM = \frac{dmean(L(err_i^1) - L(err_i^2))}{\sqrt{S^2/n}} S^2 \quad (26)$$

The DM test consequences of the employed system and those of the other comparative models are shown in Table 13.

Table 13. The Diebold-Mariano test results for each model.

Site	Model	VLG (Adam)	Model	VLG (Adam)
NSW	Elman	9.8311 *	EMD-LSTM	2.1136 **
	RBF	9.9634 *	DA-BP	3.0329 *
	SVM	4.1134 *	DA-LSTM	3.9364 *
	BP	3.1654 *	VMD-BP	3.8029 *
	PSO-BP	2.9253 *	VMD-LSTM (SGD)	3.8395 *
	LSTM	3.8639 *	VMD-LSTM (Adam)	2.0104 **
	EMD-BP	3.8217 *	VLG (SGD)	3.7156 *
Site	Model	VLG (Adam)	Model	VLG (Adam)
QLD	Elman	8.2174 *	EMD-LSTM	2.3522 **
	RBF	8.2316 *	DA-BP	2.9871 *
	SVM	4.3108 *	DA-LSTM	2.8346 *
	BP	3.5936 *	VMD-BP	3.2135 *
	PSO-BP	3.1732 *	VMD-LSTM (SGD)	2.8724 *
	LSTM	3.1420 *	VMD-LSTM (Adam)	2.3412 **
	EMD-BP	2.4718 **	VLG (SGD)	2.3439 **
Site	Model	VLG (Adam)	Model	VLG (Adam)
SA	Elman	7.9347 *	EMD-LSTM	2.1261 **
	RBF	7.9931 *	DA-BP	2.8469 *
	SVM	3.3416 *	DA-LSTM	2.9336 *
	BP	3.3153 *	VMD-BP	2.1743 **
	PSO-BP	3.1732 *	VMD-LSTM (SGD)	2.0214 **
	LSTM	3.3509 *	VMD-LSTM (Adam)	1.9542 ***
	EMD-BP	2.1798 **	VLG (SGD)	2.0147 **

* represents the DM value at a confidence level of 1%. ** represents the DM value at a confidence level of 5%.

*** represents the DM value at a confidence level of 10%.

On account of the above results, the following conclusions can be drawn:

- (1) By contrasting and dissecting the forecasting errors of different hybrid systems, the DM test consequences of different prediction models are all at the upper limit at a confidence level of 1%;
- (2) The Diebold-Mariano test was performed on the prediction errors of four different traditional single models, and the test results of the VLG were all higher than the upper limit at a confidence level of 1%;
- (3) The minimum value of the comparison between the VLG combination system and the LSTM model using other data preprocessing technologies is 2.0104, which is also far beyond the 5% significance level threshold.

Therefore, according to the Diebold-Mariano test results, it can be legitimately summarized that the employed predicted model not only has greater prediction capacity than other systems but also

indicates crucial distinctions in the level of prediction accuracy and superiority in short-term power load forecasting.

5.2. Model Stability Study

This section starts with the stability of the model and proposes two different sets of multiple experiments to estimate the prediction ability of the VLG system. By comparing the prediction stability values of the three systems of LSTM, BP, and PSO-BP, the prediction stability is verified for the core part of the hybrid system, LSTM, employed in this paper. All three models use raw data and can be modeled without denoising. Based on the Spring feature data set in Queensland, Australia, 20 repeated experiments were conducted on the three models to explore the volatility of the prediction accuracy of the models. As is well-known, variance can reflect the robustness and volatility of the forecasting system. The smaller the standard deviation of the prediction error is, the more robust the prediction system and the weaker the volatility will be. Therefore, the standard deviation of the predicted error is used to appraise the robustness of the employed combination prediction system and other contradistinctive systems.

Figure 7a illustrates the differences in forecasting stability between the three systems over 20 tests. The results show that the BP has the largest volatility: The maximum MAPE is 1.5648%, the minimum is 0.9186%, and the standard deviation is 0.2022. In contrast, the prediction capacity of the LSTM model is better and more stable. The maximum value of MAPE is 0.9126%, the minimum value is 0.8241%, and the standard deviation is only 0.0367. Among these ANNs, the LSTM model has the best effectiveness in prediction accuracy and stability, as well as the least standard deviation, which reflects the stability advantages of the LSTM model.

Figure 7. Prediction performance and stability of the STM neural network and Adam optimizer with the Queensland datasets. (a) Experimental results stability of different prediction models. (b) Experimental results of stability of different optimizers.

Moreover, the excellent prediction performance of these models depends to a great extent on the optimization method of the model parameters. Different parameter optimization methods have certain effects on the forecasting accuracy and convergence speed of the system. To ensure a single variable, the LSTM model was selected to apply the Adam optimizer and the SGD stochastic gradient descent method for 100 trials, and the differences between the two optimization methods were compared. Figure 7 (b) shows the results. The average value of MAPE of LSTM model obtained by Adam optimization method is 0.8457%, which is significantly improved than that of SGD. In the scatter plot of MAPE, the MAPE obtained by the Adam optimizer is between 0.8% and 0.9%, and the standard deviation is 0.0341. Moreover, the bandwidth of the scatter plot is narrower, indicating that the model prediction is more stable and has smaller prediction volatility. On the contrary, the MAPE value of the model using the stochastic gradient descent method is between 0.9% and 1.3%, with a standard deviation of 0.2451. Here, the prediction error of the model is increased, the stability is worse than that of the Adam optimizer, the prediction volatility is large, and the prediction performance is unstable.

5.3. Multi-Step Prediction and Result Analysis

In the experiments in Section 4, the forecast model was used to make the next forecast of the power load data—that is, a single-step forecast. This section compares the three different hybrid models proposed to appraise the prediction performance of the employed hybrid system VLG in a multi-step prediction test.

Unlike the comparison model proposed in the previous section, this experiment aims to verify whether the new VLG model is comparable with the other two VMD-based hybrid models (i.e., VMD-GWO-SVM and VMD-PSO-BP). SVM and BP models perform well in dealing with time series forecasting problems and are also generally employed in short-term power load data predictions. In the model's hyperparameter optimization algorithm, the two comparative experiments, respectively, used the PSO optimization algorithm and the gray wolf optimization algorithm. In the spring data of Queensland, BP and SVM are optimized to certify the accuracy of the VLG hybrid system based on a comparison with excellent classic hybrid models (see Table 14 and Figure 8 for comparison results). In Figure 8, (a) is the forecasting and actual values of the proposed model and the classical hybrid prediction model in the multi-step prediction. (b) shows the index results of the proposed prediction system and the classical hybrid prediction model.).

Table 14. Comparison of the forecasting performance of the combined model and that of other VMD-based models.

Model	MAE			RMSE			MAPE (100%)			Percentage of the MAPE		
	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step
VMD-GWO-SVM	39.22	146.36	164.35	99.87	183.45	198.24	0.8846	1.9159	2.0213	72.33%	60.05%	32.88%
VMD-PSO-BP	19.90	119.96	133.26	48.52	161.51	172.97	0.2821	1.6333	1.6722	13.22%	53.14%	18.87%
VLG	18.13	48.77	103.22	23.54	56.73	133.86	0.2448	0.7653	1.3565	-	-	-

For the spring data of Queensland, in a one-step forecast, the proposed combined model VLG and comparative model VMD-PSO-BP had no significant differences in the various evaluation indicators but still obtained the best MAPE, MAE and RMSE, respectively, with 0.2448%, 18.13, and 23.54. In the one-step prediction, although the employed system does not reflect the outstanding advantages of the classic combined model VMD-PSO-BP in prediction accuracy and performance, the performance remains no worse than that of any other models. When the prediction of the model involves a two-step prediction, the VLG achieves obvious advantages, and the MAPE, MAE, and RMSE are 0.7653%, 48.77, and 56.73, respectively. Instead, the VMD-PSO-BP model has no significant difference in prediction accuracy compared to the VMD-GWO-SVM in the two-step prediction. The prediction accuracy of both is far poorer than that of one-step prediction. The MAPE values are 1.6333% and 1.9159%, which are 0.868% and 1.1506% higher than the values of the VLG system. In the three-step prediction, the MAPEs of the three hybrid systems are all greater than 1%, but the MAPE of the proposed VLG

combined system is 1.35655%. This system still offers the greatest prediction ability among the three excellent VMD-based combined systems. From the perspective of the improved MAPE, the prediction capacity of the VLG system improved the greatest in the two-step prediction, with 60.05% and 53.14%, respectively. Figure 8 demonstrates a comparison of the forecasting performance of the experiment with the spring data of Queensland steps 1, 2, and 3. Among the three-step prediction models, the VLG hybrid system is still the most accurate and valid prediction system.

Figure 8. Comparison between the VLG model and the classical hybrid model in multi-step predictions. (a) Prediction model fitting. (b) Comparison of prediction performance evaluation indexes.

5.4. Improvement of the Evaluation Index

In previous index evaluation systems, the MAPE values of each prediction model were too small, and the RMSEs were different because of the differences in the data dimension, making it challenging to intuitively display the degree of differences in the model prediction accuracy [48]. In this study,

we use the percentage improvements of the MAPE and RMSE criteria. In this way, a comprehensive analysis of the proposed combined system can be carried out. The definition is as follows:

$$P_{MAPE} = \left| \frac{MAPE_1 - MAPE_2}{MAPE_1} \right| \times 100\% \quad (27)$$

$$P_{RMSE} = \left| \frac{RMSE_1 - RMSE_2}{RMSE_1} \right| \times 100\% \quad (28)$$

The improved MAPE and RMSE indicators are shown in Table 15. Considering the results in Table 15, the prediction capacity of the employed combined system is discussed and analyzed as follows:

- The predictive capacity of the system employed in this study is clearly commendable.
- The value of the percentage improvement of the evaluation index shows a clear decreasing trend. This indicates that the prediction veracity of the system is gradually improved due to the data preprocessing technology and simulation optimization algorithm playing a vital role in improving the prediction ability of the system.
- Notably, the VLG hybrid system presents obvious advantages over other systems.

Table 15. Improvement percentage of the proposed model.

Model P_{MAPE} (100%)	Spring			Summer			Autumn			Winter			Annual Mean		
	NSW	QLD	SA	NSW	QLD	SA	NSW	QLD	SA	NSW	QLD	SA	NSW	QLD	SA
Elman	98.39	96.50	95.93	94.78	94.54	92.15	95.97	95.68	94.60	94.29	89.63	90.20	96.50	94.59	93.80
RBF	98.41	96.51	96.01	94.87	94.6	92.31	96.09	95.70	94.71	94.27	89.62	90.31	96.55	94.62	94.26
ARIMA	82.66	86.65	56.45	70.89	82.89	42.87	82.49	85.58	70.04	75.44	75.53	58.06	77.98	82.83	58.63
SVM	84.91	81.71	64.35	75.56	77.09	57.83	85.37	74.46	68.73	79.27	67.57	55.61	81.33	75.34	61.78
BP	75.33	73.57	61.85	66.22	68.18	60.44	78.79	67.83	67.33	58.34	51.19	53.25	69.95	65.54	62.56
PSO-BP	73.81	69.94	54.19	44.24	53.36	50.79	75.12	64.82	66.86	52.30	48.02	45.67	63.19	59.17	54.89
LSTM	65.22	67.70	58.3	55.26	61.98	58.14	73.70	64.51	66.64	52.51	49.41	41.84	61.81	54.99	56.39
VMD-LSTM	38.77	46.44	19.11	38.11	16.50	3.44	42.79	40.68	30.90	11.64	23.47	7.24	32.78	31.78	15.63
EMD-LSTM-GA	10.11	14.89	10.36	8.74	9.47	7.32	14.31	12.74	9.33	7.35	7.27	6.11	10.31	11.09	8.23
DA-LSTM-GA	35.79	39.46	27.33	28.22	32.12	27.22	38.12	36.48	29.87	0.13	19.52	17.84	30.83	31.89	28.86
VMD-PSO-BP	12.77	16.31	9.80	10.06	9.89	4.22	15.25	13.22	11.47	8.36	7.98	5.31	11.12	11.85	7.75

Model P_{RMSE} (100%)	Spring			Summer			Autumn			Winter			Annual Mean		
	NSW	QLD	SA	NSW	QLD	SA	NSW	QLD	SA	NSW	QLD	SA	NSW	QLD	SA
Elman	98.21	96.19	95.64	94.69	94.46	91.62	96.03	95.99	93.37	92.70	89.70	88.56	96.14	94.62	92.88
RBF	98.23	96.21	95.68	94.79	94.55	91.70	96.09	96.03	93.39	92.69	89.61	88.47	96.18	94.67	92.91
ARIMA	75.16	85.23	37.12	68.11	78.68	37.71	81.55	84.20	70.49	67.87	72.39	53.90	73.08	80.51	52.91
SVM	82.83	82.14	70.78	59.73	77.09	56.13	79.58	78.80	75.08	77.28	70.91	51.46	85.11	77.29	64.23
BP	76.24	72.91	66.52	65.25	66.86	62.43	78.99	70.71	72.20	61.25	54.88	55.75	70.53	66.583	63.98
PSO-BP	77.23	72.27	65.81	44.31	53.54	60.14	76.42	69.05	70.30	72.24	68.76	48.67	69.65	66.60	61.17
LSTM	66.67	70.06	64.87	57.41	59.90	66.12	74.04	66.80	69.21	58.31	52.07	44.24	63.84	62.25	61.39
VMD-LSTM	38.43	41.22	10.739	34.14	12.35	5.720	42.46	38.91	26.21	13.44	41.81	42.52	32.13	34.86	26.42
EMD-LSTM-GA	10.45	13.96	11.43	8.41	9.33	7.42	15.09	11.98	9.48	6.89	7.05	6.35	10.64	11.01	8.44
DA-LSTM-GA	32.48	37.97	26.71	30.15	34.01	26.17	37.38	35.84	28.85	0.24	18.89	16.97	32.85	32.22	27.26
VMD-PSO-BP	13.34	17.54	9.66	10.23	8.97	4.63	14.17	14.03	11.66	8.72	7.64	5.44	11.10	12.30	7.21

5.5. Future and Prospects

The hybrid prediction system based on deep learning proposed in this study overcomes the shortcomings of traditional prediction models. In the rapid development of the intelligent information age, accurate load forecasting has become an indispensable part of the power load field, which plays a vital role in the safe operation, daily distribution, and economy of the power system. From artificial neural networks, to machine learning, and finally to deep learning, the construction of load forecasting model is approaching ever closer to the actual situation. Given the existing prediction models, the prediction models constructed by deep learning methods such as LSTM and CNN are obviously better than those constructed by artificial neural networks and traditional machine learning algorithms. A hybrid prediction model combining the intelligent optimization algorithm and deep learning network has higher practical application value and stronger expansibility, making it able to more easily fit nonlinear time series with strong volatility.

In addition to traditional load forecasting, intermittent renewable energy, such as photovoltaic power generation and wind power generation, features stronger volatility, randomness, and instability. However, research on predicting intermittent renewable energy using deep learning methods is still very limited. The development of a comprehensive and effective deep learning method is expected to become a new direction of smart grid research. The application of artificial intelligence algorithms in the field of new energy is a new typical application scenario for artificial intelligence and also provides a new intelligent solution for building a global low-carbon energy future.

6. Conclusions

Short-term power load predictions play a crucial role in the safe operation and risk assessment of electrified wire netting, which has aroused attention and excitement among scholars. Because of the inherent uncertainty and randomness of power load sequences, determining how to efficiently and effectively predict the power load is still a challenging task. A string of research has also been developed to enhance the performance of power load prediction. Unfortunately, these methods are mostly restricted to using a onefold prediction model to predict power load sequences; however, any single model will have inherent shortcomings. Moreover, most previous studies have not considered the effect of data preprocessing and sequence noise on the model prediction accuracy. In response, this paper proposed a new predictive analysis system that overcomes the above shortcomings and provides an effective technical means for short-term power load analysis and monitoring. In the developed model, variational modal decomposition (VMD) was employed to divide the original sequence from high to low into a set of components. Through reconstruction, the global noise was effectively removed. Then, the long-term short-term memory neural network (LSTM) was used instead of the classical neural network to predict the power load data, which effectively improved the prediction accuracy. Finally, to further improve the modeling performance and robustness, an improved binary coding genetic algorithm was proposed based on the genetic algorithm; this algorithm achieved high accuracy and maintained strong stability. The effectiveness of the algorithm was verified by experiments. The predictive analysis system developed by the application was used to predict the seasonal data of New South Wales, Queensland, and South Australia and calculate multiple performance indicators (MAPE, RMSE, MSE, and PMAPE). The experimental results indicate that the minimum MAPE values of the employed VLG system are 0.3717%, 0.3486%, and 0.9800%, respectively, which are better than the values of the comparison model. Like the traditional excellent hybrid model, the multi-step prediction of this model also provides strong prediction performance. In general, the proposed predictive analysis system shows excellent performance in analyzing and monitoring short-term power loads. Specifically, this system does not only deeply analyze the information related to people's activities and lives but can also accurately and steadily approach the actual values. Therefore, future power plant decision-makers and grid investors could make reasonable decisions based on the system presented in this article to monitor and predict power loads.

Author Contributions: Conceptualization, Y.J. Methodology, J.W. Visualization, A.S. Writing—review and editing, H.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Foundation of Liaoning Social Science Federation [Grant No. 20201slktyb-031].

Acknowledgments: This research was supported by Foundation of Liaoning Social Science Federation [Grant No. 20201slktyb-031]. And all the authors do not have any possible conflicts of interest.

Conflicts of Interest: We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work. This manuscript is our own work and the content of this paper has not been copied from elsewhere. This manuscript has not been published before nor submitted to another journal for the consideration of publication and all data measurements are genuine results and have not been manipulated. In addition, none of the authors have any financial or scientific conflicts of interest with regard to the research described in this manuscript. And this manuscript was supported by Foundation of Liaoning Social Science Federation [Grant No. 20201slktyb-031]. And all the authors do not have any possible conflicts of interest.

Abbreviations

List of terminologies (method and indices)			
EMD	Empirical model decomposition	DA	Denoise autoencoder
VMD	Variational Modal Decomposition	BP	Back propagation neural network
LSTM	Long Short-Term Memory neural network	RBF	Radial basis function
Elman	Elman neural network	SVM	Support vector machine
ARMA	Autoregressive moving average model	ARIMA	Autoregressive interval moving average model
PSO	Partial swarm optimization algorithm	GWO	Grey wolf optimization algorithm
EMD-BP	BP after EMD technology	DA-BP	BP after DA technology
VMD-BP	BP after VMD technology	EMD-LSTM	LSTM after EMD technology
DA-LSTM	LSTM after DA technology	VMD-LSTM	LSTM after VMD technology
GC	Grey correlation method	IMFs	Intrinsic mode functions
SP	Spearman correlation coefficient	PE	Pearson correlation coefficient
MAPE	Mean absolute percentage error	MAE	Mean absolute error
RMSE	Root Mean Square Error	R ²	Coefficient of determination
AI	Artificial intelligence	PSO-BP	BP after PSO optimization algorithm
Adam	Adaptive moment estimation optimization	SGD	Stochastic Gradient Descent
BEGA	Binary encoding genetic algorithm	EMD-LSTM GA	LSTM after EMD and GA optimization
DA-LSTM-GA	LSTM after DA and GA optimization	VMD-PSO-BP	BP after PSO and VMD optimization
VMD-GWO-SVM	VMD after VMD and GWO optimization	DM	Diebold-Mariano test
DL	Deep learning	RNN	Recurrent neural networks
CNN	Convolution neural network	BiLSTM	Bi-directional long-term memory neural network
ANNs	Artificial neural networks	P _{MAPE}	The improvement in MAPE values
ADMM	Alternate Direction Method of Multipliers	VLG	The model combined VMD-BEGA -LSTM
List of terminologies (parameters and variables)			
ω_k	Center pulse frequency	NI	Population Size of the GA
f	Actual signal	α	Initial learning rate
λ	Lagrange factor	x_t	Corresponding input
$u_k(\omega)$	Modal function of VMD Technology	L1	Binary encoding length
f_t	Forget gate	i_t	Input gate
O_t	Output gate	C_t	Memory cell state
h_t	T-1 time input	<i>sigmoid</i>	Input gate layer
W_f	Coefficient matrix	b_f	Bias vector
g_t	T gradient of time step	m_t	Gradient mean
v_t	Exponential moving average	β	Exponential decay rate
p_i	Selection probability in genetic	x^{bin}	Binary code
x^{dec}	Decimal code	L	Input unit length
N	Number of cells in the hidden layer	loss	Loss function value of training model
y_m	True value of data	$\hat{y}_m$	The value of model forecast
iter	Model iterations	$\bar{y}$	Average of sequence data
ω_k	Center pulse frequency	NI	Population Size of the GA
f	Actual signal	x_t	Corresponding input

The main terminologies mentioned in this paper (including indices, methods, variables and parameters).

References

1. Yang, W.; Wang, J.; Tong, N. A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting. *Appl. Energy* **2019**, *235*, 1205–1225. [\[CrossRef\]](#)
2. Guo, Z.; Zhou, K.; Zhang, X.; Yang, S. A deep learning model for short-term power load and probability density forecasting. *Energy* **2018**, *160*, 1186–1200. [\[CrossRef\]](#)
3. Zhang, X.; Wang, J. A novel decomposition-ensemble model for forecasting short-term load-time series with multiple seasonal patterns. *Appl. Soft Comput.* **2018**, *65*, 478–494. [\[CrossRef\]](#)
4. Wang, R.; Wang, J.; Xu, Y. A novel combined model based on hybrid optimization algorithm for electrical load forecasting. *Appl. Soft Comput.* **2019**, *82*, 105548. [\[CrossRef\]](#)
5. He, Q.; Wang, J.; Lu, H. A hybrid system for short-term wind speed forecasting. *Appl. Energy* **2018**, *226*, 756–771. [\[CrossRef\]](#)
6. Zhao, H.; Han, X.; Guo, S. DGM (1, 1) model optimized by MVO (multi-verse optimizer) for annual peak load forecasting. *Neural Comput. Appl.* **2018**, *30*, 1811–1825. [\[CrossRef\]](#)

7. Yang, Z.; Niu, H. Research on urban distribution network planning management system based on load density method. *Eng. Technol. Res.* **2018**, *8*, 76–77.
8. Cui, Q.; Shu, J.; Wu, Z.; Huang, L.; Yao, W.; Song, X. Medium- and long-term load forecasting based on glrm model and MC error correction. *New Energy Progress* **2017**, *5*, 472–477.
9. Jaihuni, M.; Basak, J.K.; Khan, F.; Okyere, F.G.; Arulmozhi, E.; Bhujel, A.; Park, J.; Hyun, L.D.; Kim, H.T. A Partially Amended Hybrid Bi-GRU—ARIMA Model (PAHM) for Predicting Solar Irradiance in Short and Very-Short Terms. *Energies* **2020**, *13*, 435. [\[CrossRef\]](#)
10. Song, J.; Wang, J.; Lu, H. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting. *Appl. Energy* **2018**, *215*, 643–658. [\[CrossRef\]](#)
11. Lydia, M.; Kumar, S.S.; Selvakumar, A.I.; Kumar, G.E.P. Linear and nonlinear auto-regressive models for short-term wind speed forecasting. *Energy Convers. Manag.* **2016**, *112*, 115–124. [\[CrossRef\]](#)
12. Kavasseri Rajesh, G.; Seetharaman, K. Day-ahead wind speed forecasting using f-ARIMA models. *Renew. Energy* **2009**, *34*, 1388–1393. [\[CrossRef\]](#)
13. Zhang, X.; Wang, J.; Gao, Y. A Hybrid Short-Term Electricity Price Forecasting Framework: Cuckoo Search-Based Feature Selection with Singular Spectrum Analysis and Svm. *Energy Econ.* **2019**, *81*, 899–913. [\[CrossRef\]](#)
14. Liu, J.; Liu, X.; Le, B.T. Rolling Force Prediction of Hot Rolling Based on GA-MELM. *Complexity* **2019**, 3476521. [\[CrossRef\]](#)
15. 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*. [\[CrossRef\]](#)
16. Xu, P. *Research on Load Forecasting Method Based on Fuzzy Clustering and RBF Neural Network*; Guangxi University: Nanning, China, 2012.
17. Xingjun, L.; Zhiwei, S.; Hongping, C.; Mohammed, B.O. A new fuzzy-based method for load balancing in the cloud-based Internet of things using a grey wolf optimization algorithm. *Int. J. Commun. Syst.* **2020**, *33*. [\[CrossRef\]](#)
18. Almalaq, A.; Edwards, G. A review of deep learning methods applied on load forecasting. In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, Cancun, Mexico, 18–21 December 2017; pp. 511–516.
19. Ryu, S.; Noh, J.; Kim, H. Deep neural network based demand side short term load forecasting. *Energies* **2016**, *10*, 3. [\[CrossRef\]](#)
20. Massaoudi, M.S.; Refaat, S.; Abu-Rub, H.; Chihi, I.; Oueslati, F.S. PLS-CNN-BiLSTM: An End-to-End Algorithm-Based Savitzky–Golay Smoothing and Evolution Strategy for Load Forecasting. *Energies* **2020**, *13*, 5464. [\[CrossRef\]](#)
21. Li, H.; Liu, H.; Ji, H.; Zhang, S.; Li, P. Ultra-Short-Term Load Demand Forecast Model Framework Based on Deep Learning. *Energies* **2020**, *13*, 4900. [\[CrossRef\]](#)
22. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* **2017**, *8*, 841–851. [\[CrossRef\]](#)
23. Zhang, W.; Qu, Z.; Zhang, K.; Mao, W.; Ma, Y.; Fan, X. A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. *Energy Convers. Manag.* **2017**, *136*, 439–451. [\[CrossRef\]](#)
24. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [\[CrossRef\]](#)
25. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *2009*, 1–41. [\[CrossRef\]](#)
26. Ribeiro, M.H.D.M.; Stefenon, S.F.; De Lima, J.D.; Nied, A.; Mariani, V.C.; Coelho, L.S. Electricity Price Forecasting Based on Self-Adaptive Decomposition and Heterogeneous Ensemble Learning. *Energies* **2020**, *13*, 5190. [\[CrossRef\]](#)
27. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Mariani, V.C.; Dos Santos Coelho, L.; Da Rocha, D.F.M.; Grebogi, R.B.; De Barros Ruano, A.E. Wavelet group method of data handling for fault prediction in electrical power insulators. *Int. J. Electr. Power Energy Syst.* **2020**, *123*. [\[CrossRef\]](#)

28. He, X.; Nie, Y.; Guo, H.; Wang, J. Research on a Novel Combination System on the Basis of Deep Learning and Swarm Intelligence Optimization Algorithm for Wind Speed Forecasting. *IEEE Access* **2020**, *8*, 51482–51499. [\[CrossRef\]](#)
29. Zhang, Y.; Pan, G.; Chen, B.; Han, J.; Zhao, Y.; Zhang, C. Short-term wind speed prediction model based on GA-ANN improved by VMD. *Renew. Energy* **2020**, *156*, 1373–1388. [\[CrossRef\]](#)
30. He, Z.; Chen, Y.; Shang, Z.; Li, C.; Li, L.; Xu, M. A novel wind speed forecasting model based on moving window and multi-objective particle swarm optimization algorithm. *Appl. Math. Model.* **2019**, *76*, 717–740. [\[CrossRef\]](#)
31. Zhu, C.; Teng, K. An early fault feature extraction method for rolling bearings based on variational mode decomposition and random decrement technique. *Vibroeng. Procedia* **2018**, *18*, 41–45.
32. Chen, X.; Yang, Y.; Cui, Z.; Shen, J. Wavelet Denoising for the Vibration Signals of Wind Turbines Based on Variational Mode Decomposition and Multiscale Permutation Entropy. *IEEE Access* **2020**, *8*, 40347–40356. [\[CrossRef\]](#)
33. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [\[CrossRef\]](#)
34. Zhao, N.; Mao, Z.; Wei, D.; Zhao, H.; Zhang, J.; Jiang, Z. Fault Diagnosis of Diesel Engine Valve Clearance Based on Variational Mode Decomposition and Random Forest. *Appl. Sci.* **2020**, *10*, 1124. [\[CrossRef\]](#)
35. Song, E.; Ke, Y.; Yao, C.; Dong, Q.; Yang, L. Fault Diagnosis Method for High-Pressure Common Rail Injector Based on IFOA-VMD and Hierarchical Dispersion Entropy. *Entropy* **2019**, *21*, 923. [\[CrossRef\]](#)
36. Wang, J.; Wang, S.; Yang, W. A novel non-linear combination system for short-term wind speed forecast. *Renew. Energy* **2019**, *143*, 1172–1192. [\[CrossRef\]](#)
37. Sun, H.; Fang, L.; Zhao, F. A fault feature extraction method for single-channel signal of rotary machinery based on VMD and KICA. *J. Vibroeng.* **2019**, *21*, 370–383. [\[CrossRef\]](#)
38. Lin, H.; Hua, Y.; Ma, L.; Chen, L. Application of ConvLSTM network in numerical temperature prediction interpretation. In Proceedings of the ICMLC '19—2019 11th International Conference on Machine Learning and Computing, Zhuhai, China, 22–24 February 2019; pp. 109–113. [\[CrossRef\]](#)
39. Wang, J.Q.; Du, Y.; Wang, J. LSTM based long-term energy consumption prediction with periodicity. *Energy* **2020**, *197*, 117197. [\[CrossRef\]](#)
40. Sakinah, N.; Tahir, M.; Badriyah, T.; Syarif, I. LSTM with adam optimization-powered high accuracy preeclampsia classification. In Proceedings of the 2019 International Electronics Symposium (IES), Surabaya, Indonesia, 27–28 September 2019; pp. 314–319. [\[CrossRef\]](#)
41. Li, C.; Xie, C.; Zhang, B.; Chen, C.; Han, J. Deep Fisher discriminant learning for mobile hand gesture recognition. *Pattern Recognit.* **2018**, *77*, 276–288. [\[CrossRef\]](#)
42. Qin, X.; Zhang, W.; Gao, S.; He, X.; Lu, J. Sensor fault diagnosis of autonomous underwater vehicle based on LSTM. In Proceedings of the 2018 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; pp. 6067–6072. [\[CrossRef\]](#)
43. Li, H.; Wang, J.; Li, R.; Lu, H. Novel analysis–forecast system based on multi-objective optimization for air quality index. *J. Clean. Prod.* **2019**, *208*, 1365–1383. [\[CrossRef\]](#)
44. Bera, S. Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification. *Int. J. Remote Sens.* **2020**, *41*. [\[CrossRef\]](#)
45. Yang, W.; Wang, J.; Wang, R. Research and Application of a Novel Hybrid Model Based on Data Selection and Artificial Intelligence Algorithm for Short Term Load Forecasting. *Entropy* **2017**, *19*, 52. [\[CrossRef\]](#)
46. Yang, W.; Wang, J.; Niu, T.; Du, P. A Novel System for Multi-Step Electricity Price Forecasting for Electricity Market Management. *Appl. Soft Comput.* **2020**, *88*. [\[CrossRef\]](#)
47. He, B.; Ying, N.; Jianzhou, W. Electric Load Forecasting Use a Novelty Hybrid Model on the Basic of Data Preprocessing Technique and Multi-Objective Optimization Algorithm. *IEEE Access* **2020**, *8*, 13858–13874.
48. Yechi, Z.; Jianzhou, W.; Haiyan, L. Research and Application of a Novel Combined Model Based on Multiobjective Optimization for Multistep-Ahead Electric Load Forecasting. *Energies* **2019**, *12*, 1931. [\[CrossRef\]](#)

© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).