sustainability

MDPI

Article

An Improved Neural Network Algorithm for Energy Consumption Forecasting

Jing Bai [1], Jiahui Wang [1], Jin Ran [2], Xingyuan Li [2,*] and Chuang Tu [1]

[1] Economic and Management College, Yanshan University, Qinhuangdao 066004, China; bjr@ysu.edu.cn (J.B.)
[2] Xinjiang Key Laboratory of Green Construction and Smart Traffic Control of Transportation Infrastructure, Xinjiang University, Urumqi 830017, China
* Correspondence: lixy_sjp2023@xju.edu.cn

Abstract: Accurate and efficient forecasting of energy consumption is a crucial prerequisite for effective energy planning and policymaking. The BP neural network has been widely used in forecasting, machine learning, and various other fields due to its nonlinear fitting ability. In order to improve the prediction accuracy of the BP neural network, this paper introduces the concept of forecast lead time and establishes a mathematical model accordingly. Prior to training the neural network, the input layer data are preprocessed based on the forecast lead time model. The training and forecasting results of the BP neural network when and when not considering forecast lead time are compared and verified. The findings demonstrate that the forecast lead time model can significantly improve the prediction speed and accuracy, proving to be highly applicable for short-term energy consumption forecasting.

Keywords: neural network; energy consumption forecasting; forecast lead time; short-term forecasting

check for updates

Citation: Bai, J.; Wang, J.; Ran, J.; Li, X.; Tu, C. An Improved Neural Network Algorithm for Energy Consumption Forecasting. *Sustainability* **2024**, *16*, 9332. https://doi.org/10.3390/su16219332

Academic Editor: Sergio Nardini

Received: 6 September 2024
Revised: 19 October 2024
Accepted: 25 October 2024
Published: 27 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Energy serves as the foundation of human society's survival and a crucial driving force for the economic development of countries worldwide today. With the acceleration of economic globalization and the continuous advancement of industrialization in various nations, global energy consumption is gradually increasing. As the largest developing country, China's energy consumption has been increasing annually. Notably, since 2000, the growth rate of energy consumption in China has significantly accelerated, increasing at a rate of 190 million tons of standard coal each year. By 2009, China surpassed the United States to become the largest energy consumer globally. By 2022, China's total energy consumption reached 5.41 billion tons of standard coal [1].

Currently, Chinese society has entered a critical period of development, and the structure and operation of the energy system have undergone significant transformations compared to the past. The theme of China's energy development is to promote high-quality development, with the main line being to deepen the supply side structural reform, comprehensively promote the transformation of China's energy consumption mode, establish a diversified clean energy supply system, implement the strategy of innovation-driven development, continuously deepen the reform of China's energy system, and continuously promote international cooperation in China's energy field. In this context, scientific and precise energy consumption forecasting is of great significance for analyzing current policies and formulating correct energy development plans, promoting China's green economic development, and achieving grand strategic goals.

The BP neural network has been widely used in the field of prediction due to its excellent function approximation and nonlinear fitting ability [2–8]. And many scholars have improved neural network algorithms to study energy prediction problems [9–11]. For example, Zeng et al. [12] used an adaptive differential evolution algorithm to optimize a BP neural network for predicting energy consumption. Cellura et al. [13] used a joint

method of unsupervised learning and supervised learning of neural networks to predict the daily electricity load status of cities. More scholars have proposed hybrid models based on neural networks for short-term prediction [14–16].

Although the BP neural network algorithm has been widely used in energy consumption forecasting, most research methods are trained on the constructed neural network, using input and output layer data from the same period. The predicted value of input layer data indicators are obtained from self prediction or querying other scholars, institutions, etc. Then, these are applied to the trained network, and the calculated output layer value is the final predicted value, which is shown in Figure 1. However, there are often different lag times between the predicted objects and the influencing factors. For example, the population of the current year can directly affect the energy consumption level in the following year, but the energy processing and transformation technology needs a certain amount of time to promote and implement, so the factor of energy processing and transformation efficiency can have a certain impact on the energy consumption level after a longer period of time. In this way, it is necessary to adjust the interval period of the influencing factor data and the target prediction object in the time dimension to achieve a high correlation between the two. However, the current research ignores the time lag relationship between the data of the input layer and the output layer, which greatly reduces the prediction accuracy.

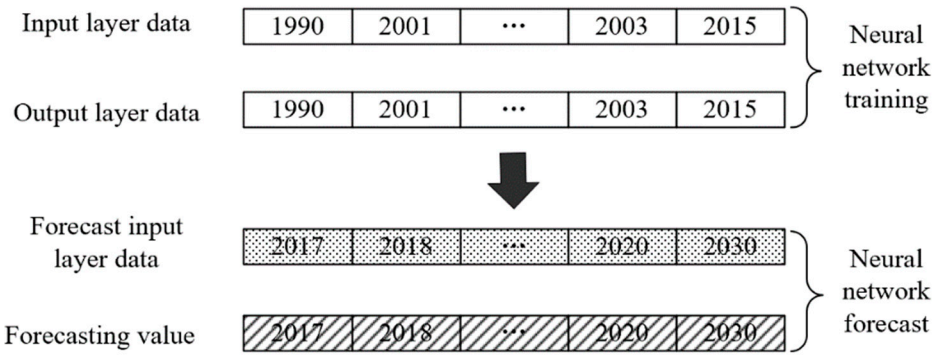


Figure 1. Explanation of traditional neural network prediction method.

Therefore, considering the different lag between the prediction object and the influencing factors, this paper proposes the concept of forecast lead time and constructs a model. Based on the model, the forecast lead time between the input layer nodes and the output layer data indicators is calculated, and the BP neural network, a typical algorithm for a neural network, is taken as an example to process the data index of its input layer. Then, the impact of considering and not considering the forecast lead time on the accuracy of energy consumption forecasting in China are compared and analyzed. The results show that applying the concept of forecast lead time to improve the BP neural network algorithm will be a good chance to improve tools for energy consumption forecasting in China.

The remainder of this paper is organized as follows. In Section 2, there are literature reviews about forecasting methods. In Section 3, the concept of forecasting lead time is proposed and a BP neural network solution model is constructed. In Section 4, we use data from 1990 to 2015 in China and the constructed model for forecasting, then we compare the experimental results of prediction algorithms that consider and do not consider the forecast lead time. Finally, Section 5 offers a comprehensive conclusion.

2. Literature Review

Energy consumption is influenced by various factors such as economic development level, population size and structure, policy orientation, and technological progress. These factors are usually nonlinear and non-stationary time series, and there may be some correlation between them. Many scholars have explored the issue of energy consumption forecasting from different perspectives and aspects, and have developed a large number of methods and models, as shown in Table 1. These methods and models can be classified

into the following categories: (1) statistical model prediction methods, among which regression prediction methods are the most widely used; (2) time series prediction methods; (3) grey prediction models; (4) artificial intelligence prediction models; and (5) hybrid prediction models.

Table 1. Summary of prediction methods.

Model	Input	Method Description	Advantages	Disadvantages
Regressive analysis	Processed historical data	Based on the changing patterns of historical data and factors affecting prediction target, search for the correlation between independent and dependent variables and their regression equations, determine model parameters, and infer future values based on this	The calculation principle and formal structure are simple, the prediction speed is fast, the extrapolation performance is good, and it shows good prediction for situations that have not occurred in history	Nonlinear regression models have non-unique expressions, relatively large workloads, and are prone to "spurious regression"
Time series method	Processed historical data	Based on historical data, establish a mathematical model of prediction target variation over time, and determine the prediction target expression based on this model	It can eliminate modelling problems such as multicollinearity, heteroscedasticity, and sequence correlation, and is computationally simple	The workload is relatively large, and the final form of the model is not unique. When there are significant changes in the external environment, there will be significant deviations
Grey prediction method	Processed historical data	Utilize the limited known information, seeking the inherent motion laws of the system, and then predict the future state of the system	Low data requirements, simple calculation, and wide applicability	Sensitive to data changes, a limited long-term predictive ability, and a fixed model structure
BP neural network	Processed historical data	A multi-layer feedforward neural network trained using an error backpropagation algorithm	A flexible network structure with a strong model generalization ability and nonlinear mapping ability	A slow learning speed, easy to fall into local optima, limited network generalization ability, lack of corresponding theoretical guidance for selecting network layers and the number of neurons
Recurrent neural network	Processed historical data	On the basis of the BP neural network, pre-sequential connections and post-sequential connections are provided for each node in the hidden layer to record pre-sequential information and apply it to post-sequential output calculation	A strong ability to process sequential data, weight sharing, flexibility, and short-term memory characteristics	Difficulty in ensuring the accuracy of information transmission for load sequences with large time spans; easy to encounter problems such as gradient disappearance and gradient explosion

Table 1. *Cont.*

Model	Input	Method Description	Advantages	Disadvantages
Long short-term memory neural network	Processed historical data	The internal recurrent unit structure of traditional recurrent neural networks cannot transmit the functional relationship between the preceding and following feature signals. Therefore, an improved long short-term memory neural network based on recurrent neural networks is proposed	Effectively solves the problems of gradient vanishing and exploding in traditional RNNs during long-term training, and performs well in handling large datasets with long time series	A high computational complexity, requiring a large amount of data and reverse training, difficult to explain the decision-making process of the network
Gated recurrent unit	Processed historical data	A simplified variant of the long short-term memory neural network unit, which combines the forget gate and input gate inside the LSTM loop body into an update gate, and replaces the output gate with a reset gate	It can simultaneously consider the temporal and nonlinear nature of power load sequences, greatly reducing the number of parameters and lowering the difficulty of network training	Difficulty in differentiating sequence features and the problem of information loss in dealing with non-continuous long time series
Convolutional neural network	Processed historical data	A feedforward neural network with a deep structure and convolutional computation	Has a strong nonlinear mapping ability, image feature extraction ability, self-learning, and an ability to adapt	When dealing with long time series, there are often limitations in the field of view, difficulties in extracting all temporal features, and a lack of memory function
Random forest algorithm	Processed historical data	An ensemble learning method, belonging to a type of supervised learning algorithm, consisting of a classifier or regressor composed of multiple decision trees	A high accuracy, strong robustness, ability to handle high-dimensional data, easy implementation, and parameter tuning	Possible overfitting, a high computational cost, difficulty in explaining individual predictions, and shows poor performance for datasets with complex interactions
Support vector machine	Processed historical data	A generalized linear classifier for the binary classification of data using supervised learning, whose decision boundary is the maximum margin hyperplane for solving the learning sample.	The efficient processing of high-dimensional data, with a good generalization ability and strong robustness	A high computational complexity, sensitivity to noise, and difficulty in parameter selection
Hybrid prediction model	Processed historical data	Using two or more different prediction methods for the same problem	Being able to comprehensively utilize the information provided by multiple prediction methods to improve prediction accuracy	A lack of uniformity in the criteria for determining combinations

2.1. Statistical Model Prediction Method

Due to their simplicity and ease of use, statistical methods have become the main method in the field of prediction. In many statistical models, regression prediction models have become the most commonly used method due to their high interpretability and

simplicity. Ding et al. [17] established a sampling regression model based on mixed data of different dynamic factors to predict China's sewage discharge. Tsekouras et al. [18] used a nonlinear multiple regression model to conduct mid-term energy forecasting for the Greek power system. Wang et al. [19] proposed an adaptive robust multi-kernel regression model for wind power prediction. Fan et al. [20] combined the phase space reconstruction PSR algorithm with the double square kernel BSK regression model and applied it to short-term power load forecasting. Wu et al. [21] proposed a hybrid model that combines seasonal index adjustment (SEAM) with regression methods, where SEAM and regression models are employed to forecast seasonal and trend items, respectively.

2.2. Time Series Prediction Method

However, regression prediction models have their own limitations. Identifying the influencing factors associated with the predicted object when utilizing regression models for prediction can be a relatively intricate and labour-intensive process. Consequently, scholars attempted to use time series methods for prediction. The predicted object has its own development trend, regarding stability and extrapolation. The time series method utilizes the properties of the predicted object itself for prediction, without considering the influence of external factors, and has a high prediction accuracy; thus, it has been widely applied. Yolcu et al. [22] combined cascaded feedforward neural networks with intuitionistic fuzzy time series for short-term load forecasting. The results show that the model has good prediction accuracy. Gupta et al. [23] proposed a hybrid method combining ARIMA and Ensemble models to predict daily and monthly inflows of reservoirs, and predicted the flow of three reservoirs in India. The results showed a high prediction accuracy and reference significance for hydropower generation. Jamil [24] used the autoregressive integrated moving average (ARIMA) model to predict hydropower consumption in Pakistan, and the results showed good fitting and minimal deviation. Sen et al. [25] applied the autoregressive integrated moving average (ARIMA) model to reveal that ARIMA (1,0,0) $\times$ (0,1,1) is the best fitting model for energy consumption prediction. Hussain et al. [26] applied Holt Winter and autoregressive integrated moving average (ARIMA) models to predict the total electricity consumption and individual component electricity consumption of Pakistan using time series quadratic data from 1980 to 2011. Martínez et al. [27] first used *k*NN regression to process data with seasonal characteristics, and, based on this, constructed a new time series prediction model. Lang et al. [28] analyzed the power load and corresponding temperature data as related time series and constructed them into a multivariate phase space, using a neural network with random weights and kernels as the prediction model.

2.3. Grey Prediction Model

The time series method relies on historical data and requires a high accuracy and the completeness of the data. Some simple time series models can make predictions with limited data, but more complex models typically require more data to support model building. If the quantity of data is insufficient, it will affect the accuracy of the prediction, and time series prediction methods are not suitable in such a case. The grey prediction model is a prediction method based on incomplete information, which can achieve a high prediction accuracy with less data and information. Therefore, in situations where information is limited, grey prediction models have been widely applied. Sang Bing Tsai et al. [29] predicted the growth trend in renewable energy consumption in China using grey models and nonlinear grey Bernoulli equations. Li et al. [30] introduced a constant free term based on the classical grey GM (1,1) model and proposed a modified grey GM (1,1) model, which was applied to predict per capita electricity consumption. Lu S [31] combined heuristic fuzzy time series with an improved grey prediction model HFEGM (1,1) to predict renewable energy in Taiwan. Xie N et al. [32] used an optimized univariate discrete grey prediction model to predict the total energy production and consumption in China. Huiping et al. [33] proposed a grey model with Consistent Fractional Reverse

Accumulation (CFOA) for predicting renewable energy consumption in Australia. The results show that the model can better capture the changing trend in renewable energy consumption and has a high prediction accuracy. Wu et al. [34] proposed a multivariate grey prediction model, considering the total population for predicting electricity consumption in Shandong Province. The results indicate that, compared with traditional grey prediction models, this model has significant predictive performance. Wang et al. [35] proposed a seasonal grey model (SGM (1,1) model) based on cumulative operators generated by seasonal factors to accurately predict the seasonal fluctuations in the electricity consumption of major economic sectors.

2.4. Artificial Intelligence Prediction Model

Compared with the above prediction methods, artificial intelligence technology has excellent performance in the field of prediction due to its strong nonlinear fitting ability, has an improving prediction accuracy, and is becoming the mainstream model in today's prediction field. Wangbin et al. [36] proposed an XGBoost model with a window mechanism and random grid search for short-term regional power load forecasting. The results indicate that the model has good predictive and generalization abilities. Grzegorz [37] uses the random forest algorithm for short-term load forecasting. LiLing et al. [38] proposed a power load forecasting method based on wavelet transform and random forest, which removes noise through wavelet transform and uses the random forest algorithm for prediction. Hosein et al. [39] used convolutional neural networks (CNN) to fully extract multidimensional features and used them as inputs for long short-term memory neural networks (LSTM) and gated recurrent unit (GRU) networks to predict each component, effectively improving the accuracy of load forecasting. Jinghua et al. [40] proposed an improved sparrow search algorithm (ISSA) to solve the hyperparameter selection problem of support vector machine (SVM) models, and used the ISSA-SVM model for long-term load forecasting. Omer et al. [41] predicted the carbon dioxide emissions of Bangladesh using several different algorithm models and found that dense neural networks (DNNs) have excellent performance. Ameyaw et al. [42] used long short-term memory neural networks (LSTM) to predict the carbon dioxide emissions from fossil fuel combustion. Mohamed et al. [43] integrated convolutional layers into gated recurrent unit (GRU) structures to effectively extract position and time features from photovoltaic power sequences. This model is used to predict photovoltaic power generation. Mehdi et al. [44] proposed a Conditional Generative Adversarial Network (cGAN) that uses four exogenous variables (highest and lowest temperatures, day of the week, and month) to predict daily loads. Abdelhak et al. [45] used artificial neural networks (ANN) and regression models to predict the power output of photovoltaic modules, and studied the effects of climate conditions and operating temperature on the estimated output.

2.5. Hybrid Prediction Model

Due to its own limitations, a single-model algorithm has a limited prediction accuracy, and it is difficult to reduce errors solely by adjusting internal parameters. In order to improve the accuracy of predictions, scholars have combined the advantages of various single-prediction models to form a hybrid prediction model. Guo et al. [46] proposed a monthly electricity consumption forecasting model based on a Vector Error Correction Model (VECM) and Adaptive Screening (SAS) method, and verified the effectiveness of this model through empirical research. Before predicting energy consumption in Anhui Province, He et al. [47] first used a reasonable feature selection method of stepwise regression to extract important variables, and then proposed two probability density prediction methods based on the Box Cox transform quantile regression of the normal distribution (N-BCQR) and gamma distribution (G-BCQR). Manel et al. [48] proposed a new parallel hybrid prediction model for predicting climate data, which combines multi-resolution analysis wavelet transform (MRA-WT) and time convolutional network (TCN) models. Abedinia et al. [49] proposed a hybrid prediction method based on a combination of neural

networks and metaheuristic algorithms for solar photovoltaic power generation prediction. Wang et al. [50] proposed two combined simulation methods, ARIMA-BPNN and BPNN-ARIMA, based on the autoregressive integrated moving average (ARIMA) method and BP neural network, for predicting the carbon emissions of China, India, the United States, and the European Union in a pandemic-free scenario.

3. Improved BP Neural Network Model

3.1. The Concept of Forecast Lead Time

The so-called forecast lead time refers to how when predicting an object, most of the influencing factors or decision variable data select the leading data indicators. Due to the different lag between the prediction object and the influencing factors, the interval between the influencing factor data and the target prediction object needs to be adjusted in the time dimension to achieve a high correlation between them. This interval is called the forecast lead time in this paper. As shown in Figure 2, it is proposed to predict the energy consumption in 2004, assuming that the forecast lead time of x_1 is 1 and that of x_2 is 2. When training the neural network, the influencing factor x_1 should select the data from 2000 to 2002, and x_2 should select the data from 1999 to 2001 as the input layer data. The energy consumption data from 2001 to 2003 are mostly output layer data. After completing the training of the neural network, the prediction of energy consumption for the year 2004 involves using the input layer with the relevant data. Specifically, the values for x_1 from the year 2003 and x_2 from the year 2002, along with any other relevant input features, should be input into the trained neural network. The output layer of the network will then provide the predicted energy consumption for the year 2004.

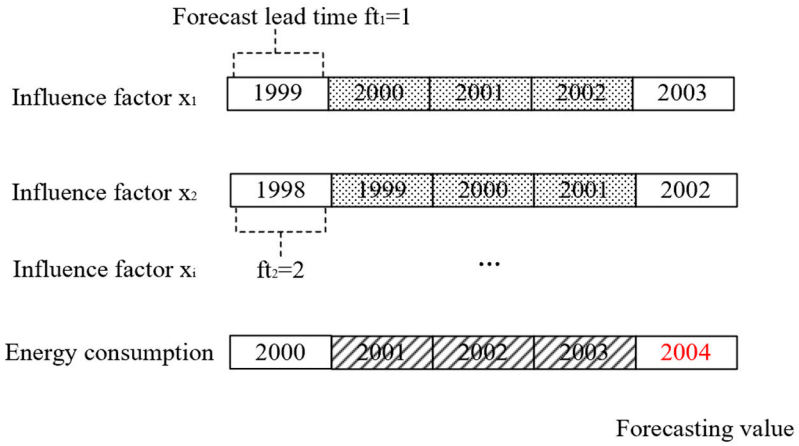


Figure 2. Concept of forecast lead time.

3.2. Overall Algorithm Flow

According to the concept of forecast lead time, an improved BP neural network algorithm is proposed, and its calculation process is shown in Figure 3. There are three core steps, which are as follows: (1) the preliminary determination of the influencing factors of energy consumption: mainly screening and determining based on the research results of existing scholars on the influencing factors of energy consumption. (2) Dealing with the influencing factors: first, the forecast lead time is determined according to the forecast lead time solution model, and then the principal component analysis is carried out to determine the training data. (3) Neural network training: based on the traditional BP neural network algorithm, the momentum term is added when adjusting the weight error. After the training is completed, the input layer data are adjusted and processed according to the forecast lead time solution model and principal component analysis, and brought into the network after the training is completed. The output layer data comprise the actual predicted values after inverse normalization processing.

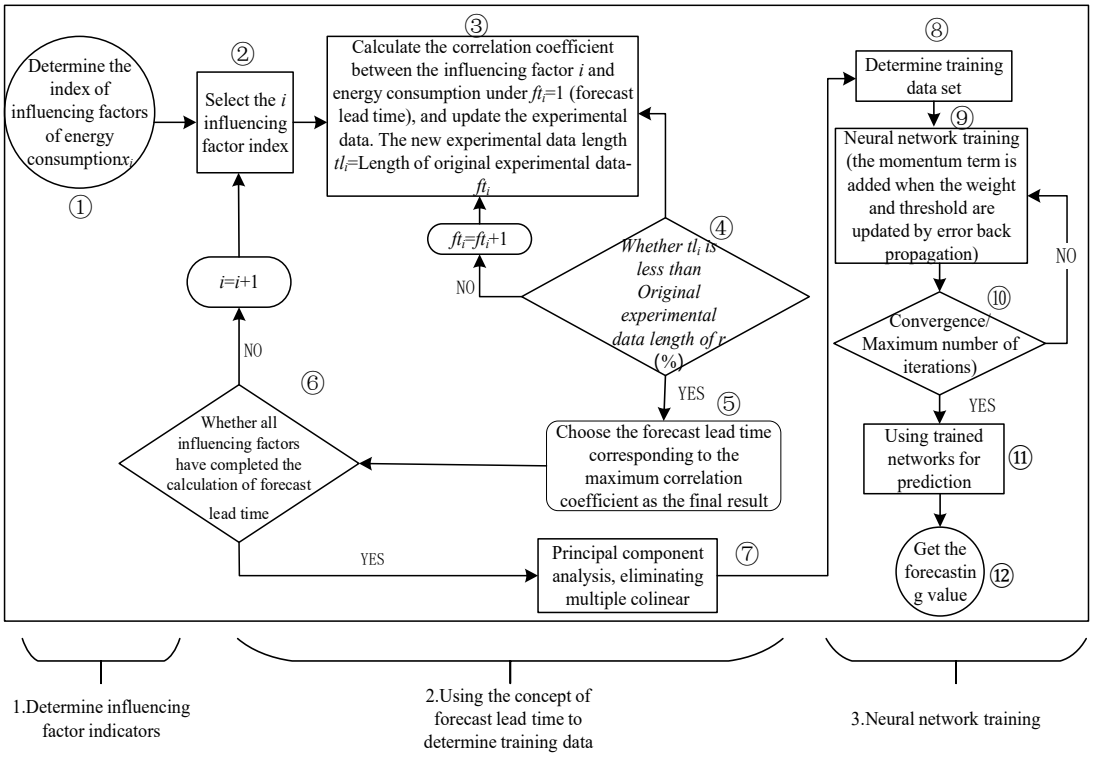


Figure 3. Overall flow of algorithm.

3.3. Construction of Forecast Lead Time Solving Model

A mathematical model has been constructed based on the forecast lead time concept, where Equation (1) represents the objective function, and Equations (2)–(6) denote the constraints. The goal of the forecast lead time solution model is to seek the maximum sum of the absolute values of correlation coefficients between various influencing factor indicators and energy consumption, as shown in Equation (1). The inclusion of absolute value constraints is necessary because there might be negative correlations between influencing factors and energy consumption. Without absolute value constraints, positive and negative correlations would offset each other during summation. Here, $R_{R_{TEC}}^{ft_i}$ represents the correlation coefficient between the influencing factor data indicator i and the energy consumption TEC under the forecast lead time ft_i. TEC denotes the time series dataset of energy consumption adjusted for forecast lead time, while EC represents the experimental dataset of original energy consumption.

$$\max\left(R_{R_{TEC}}^{ft_i}\right) \tag{1}$$

In order to ensure the accuracy of neural network training, it is necessary to restrict the length of input experimental data, as shown in Equation (2). r is the length constraint ratio of the experimental data; that is, the data length adjusted by the forecast lead time is not less than the r of the original experimental data length. l_i is the length of the experimental data of the input variable i, that is, the number of data. If the data of the influencing factor i is the annual data from 1995 to 2000, the value of l_i is 6.

$$l_i - ft_i \geq \lceil r \cdot l_i \rceil \tag{2}$$

$$\begin{cases} ti^n = i^{n+ft_i} \\ tEC^n = EC^{n+\max(ft_i)} \end{cases} \tag{3}$$

Equation (3) is the constraint that the original data value is equal to the experimental data value after the forecast lead time adjustment. n is the serial number of the time series

data, i^n is the actual data value n corresponding to the original data i of the influencing factors, ti^n is the data value n of the input variable i adjusted by the forecast lead time, EC^n is the value n of the original experimental dataset of energy consumption, and tEC^n is the data value n of the output variable energy consumption adjusted by the forecast lead time and meets $tEC^n \in EC$. As shown in Figure 4, the influencing factor i of the experimental data is the time series set from 2001 to 2008. If the forecast lead time is 1 and n is 1, then it means the corresponding value of the i^1 input variable I in 2001, and the value of the second piece of data in the original data of $\max(n) = l_i$, are equal to the first value after the adjustment of the forecast lead time.

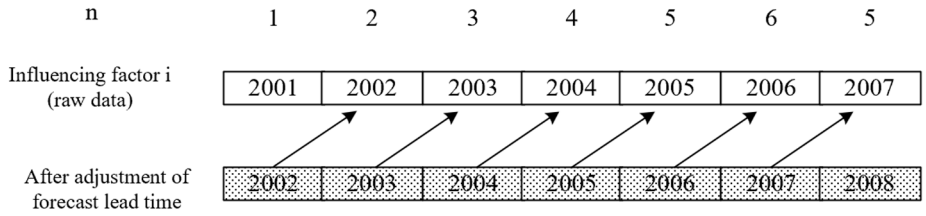


Figure 4. Corresponding description of experimental data values.

In order to meet the training requirements of the neural network, the data length of influencing factors should be equal to the original experimental data length of energy consumption. Equation (4) is the data length equality constraint. The data length adjusted by the forecast lead time is equal to its original data length minus the largest forecast lead time of all influencing factors, so as to ensure that the data length of each influencing factor is equal to the experimental data length of energy consumption. tl_i indic ates the data length of the influencing factor i after being adjusted in advance through prediction, l_{EC} indicates the original data length of energy consumption, that is, the number of elements in the set EC, and tl_{TEC} indicates the data length of energy consumption data after being adjusted in advance through prediction, that is, the number of elements in the set TEC.

$$\begin{cases} l_i = l_{EC} \\ tl_i = l_i - \max(ft_i) \\ tl_{TEC} = l_{EC} - \max(ft_i) \end{cases} \tag{4}$$

$$R_{TEC}^{ft_i} = \frac{\sum_{n=1}^{tl_i} \left(tx_i^n - \overline{TX_i}\right)\left(tEC^n - \overline{TEC}\right)}{\sqrt{\sum_{n=1}^{tl_i} \left(tx_i^n - \overline{TX_i}\right)^2} \cdot \sqrt{\sum_{n=1}^{tl_i} \left(tEC^n - \overline{TEC}\right)^2}} \tag{5}$$

$$\begin{cases} \mathrm{n} = 1, 2, 3 \cdots tl_i \\ ft_i, tl_i, tl_{TEC} \in N^+ \end{cases} \tag{6}$$

Equation (5) is the correlation coefficient formula solved by the Pearson correlation coefficient method, where TX_i is the dataset after the adjustment of the forecast lead time of the influencing factor i, and $\overline{TX_i}$ is the mean value of time series data after the adjustment of the influencing factor i. Equation (6) is the positive integer constraint of the variable.

3.4. Construction of BP Neural Network Model

The BP neural network belongs to the forward neural network, which emphasizes the learning algorithm of error backpropagation. It consists of an input layer, several hidden layers, and an output layer. The core idea is to continuously modify the weights and thresholds of the neural network through the sample training set, and gradually approach the expected output value. At the beginning of training, it propagates forward along the network, and then adjusts the weight and threshold according to the error between the network output value and the expected output value. The network training is completed

by repeatedly updating the network weights and thresholds to achieve the minimum error. Figure 5 shows the structure of a three-layer neural network.

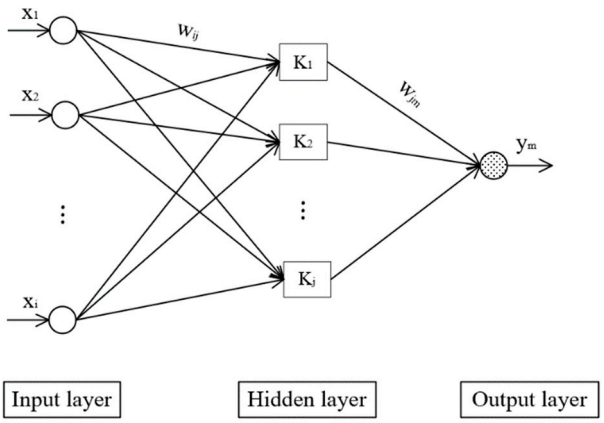


Figure 5. Three-layer neural network structure.

3.4.1. BP Neural Network Forward Propagation

(1) The input value of the j neuron in the hidden layer is as follows: $hid_{input_j} = \sum_{i=1}^{I} \omega_{ij} \cdot x_i$ Here, ω_{ij} is the weight between the input layer and the hidden layer, and x_i comprises the input layer data.

(2) The output value of the j neuron in the hidden layer is as follows: $hid_{output_j} = F\left(hid_{input_j}\right)$. $F(x)$ is the transfer function. Since the data in this paper are greater than 0, the transfer function is selected as follows: $F(x) = \frac{1}{1+e^{-x}}$;

(3) The output layer input calculation formula is as follows:

$out_{input_m} = \sum_{j=1}^{J} \omega_{jm} \cdot hid_{output_j}$. Here, ω_{jm} is the weight between the hidden layer and the output layer.

(4) The output layer calculation formula is as follows: $out_{output_m} = F\left(out_{input_m}\right)$.

3.4.2. Calculate Training Error

(1) The error of generation n is as follows: $e_m(n) = out_{output_m} - y_m$; here, y_m is the expected output.

(2) The total error of the network is as follows: $e(n) = \frac{1}{2}\sum_m^1 e_n^m(n)$.

3.4.3. Error Backpropagation

(1) The weight correction between the output layer and the hidden layer is as follows: $\triangle\omega_{jm}(n) = -\eta\frac{\partial e(n)}{\partial\omega_{jm}(n)}$; then, the revised weight is as follows: $\omega_{jm}(n+1) = \omega_{jm}(n) + \triangle\omega_{jm}(n)$. Here, η is the learning rate.

(2) The weight adjustment between the input layer and the hidden layer is as follows: $\omega_{ij}(n+1) = -\eta\frac{\partial e(n)}{\partial\omega_{ij}(n)} + \omega_{jm}(n)$.

3.4.4. Momentum BP Method

Because the BP neural network has the disadvantage of a slow convergence speed, this paper uses the momentum BP method to update the weight; that is, the momentum factor $\alpha(0 < \alpha < 1)$ is introduced when updating the weight, and the weight correction formula is as follows: $\Delta\omega(n) = -\eta(1-\alpha)\nabla e(n) + \alpha\Delta\omega(n-1)$.

4. Case Analysis

4.1. Index Selection and Standardization of Influencing Factors

Energy consumption is influenced by various factors such as economic growth, population, social development, industrial structure, energy consumption structure, supply, and technological progress. Based on the existing literature and considering the availability of data and other factors, seven influencing factors were selected, including China's economy (GDP), population (total population), social development (urban population ratio), industrial structure (the contribution rate of the secondary industry), energy consumption structure (proportion of raw coal to energy consumption), supply (total energy production), and technological progress (energy processing and conversion efficiency) from 1990 to 2015, as the input variables of the energy consumption forecast (data source: National Statistical Yearbook). In order to eliminate the dimensional impact between data indicators, the selected data are normalized to the [0, 1] interval by the way of deviation standardization according to Equation (7). Figure 6 shows the standardized experimental data.

$$x_i^* = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \tag{7}$$

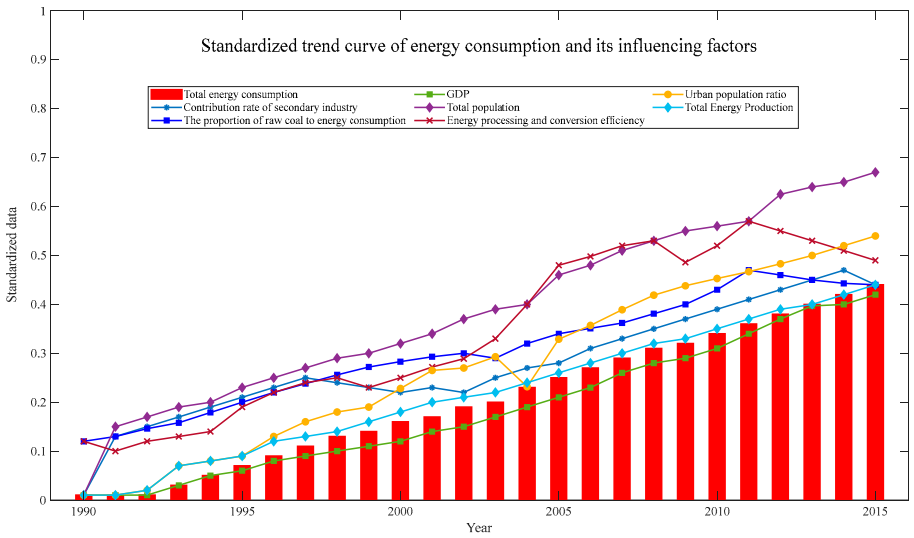


Figure 6. Data of energy consumption and influencing factors after standardization.

4.2. Forecast Lead Time Solution

For the time series calculation data from 1990 to 2015 studied in this paper, according to the data in Figure 6, 20% (1990–1994), 30% (1990–1997), 40% (1990–1999), 50% (1990–2002), 60% (1990–2005), 70% (1990–2007), 80% (1990–2010), and 90% (1990–2012) of the length of the experimental data are selected as the training data, and the data from 2013 to 2015 are selected as the experimental data. A three-layer BP neural network is constructed, consisting of 7 nodes in the input layer (seven influencing factors), 11 nodes in the hidden layer, and 1 node in the output layer (energy consumption). The number of iterations is set to 2000 generations. The neural network is trained and the mean square deviation (the average value of the sum of squares of the difference between the predicted value and the actual value) is calculated, as shown in Figure 7. It can be intuitively observed from Figure 7 that when the length of the data is greater than or equal to 70% of the experimental data, the rate of decline of the mean square error is large and the difference between the mean square error with the length of the experimental data of 80% and 90% is small. Therefore, in Equation (2), the length constraint ratio r of experimental data is taken as 70%.

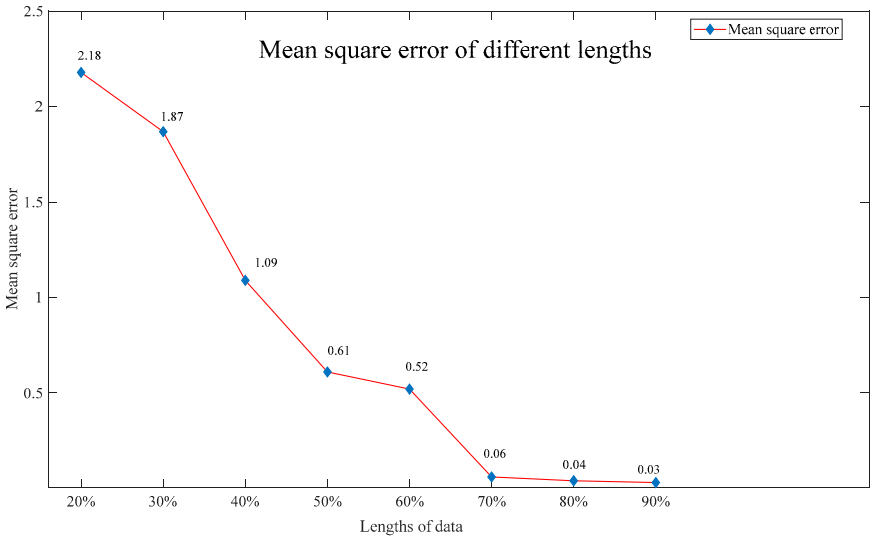


Figure 7. Mean square deviation of different data lengths.

According to the forecast lead time solution model (3.3) presented above, we use MATLAB to write a programme to solve the forecast lead time of various influencing factors, and the results are shown in Table 2. After adjusting the forecast lead time, the corresponding time dimension relationship between the influencing factors and the energy consumption experimental data is shown in Table 3, where the year in the table is the year corresponding to the energy consumption data.

Table 2. Calculation results of forecast lead time of various influencing factors.

Influence Factor	GDP	Urban Population Ratio	Total Population	Contribution Rate of Secondary Industry	Total Energy Production	Proportion of Raw Coal to Energy Consumption	Energy Processing and Conversion Efficiency
Forecast lead time	4	1	1	1	1	1	6

Table 3. Experimental data adjusted by forecast lead time.

GDP	Urban Population Ratio	Total Population	Contribution Rate of Secondary Industry	Total Energy Production	Proportion of Raw Coal to Energy Consumption	Energy Processing and Conversion Efficiency	Total Energy Consumption	Year
0.01	0.09	0.29	0.87	0.10	0.56	0.15	0.11	1996
0.03	0.14	0.35	0.85	0.11	0.51	0.08	0.11	1997
0.05	0.19	0.40	0.73	0.11	0.39	0.10	0.11	1998
0.06	0.23	0.45	0.75	0.10	0.21	0.26	0.13	1999
0.08	0.28	0.50	0.65	0.11	0.32	0.00	0.15	2000
0.09	0.33	0.54	0.75	0.13	0.14	0.71	0.17	2001
0.10	0.38	0.56	0.25	0.17	0.09	0.60	0.21	2002
0.11	0.43	0.61	0.36	0.20	0.18	0.55	0.30	2003
0.12	0.48	0.65	0.68	0.29	0.63	0.49	0.40	2004
0.14	0.52	0.68	0.45	0.40	0.81	0.49	0.49	2005
0.15	0.56	0.71	0.40	0.49	0.93	0.50	0.58	2006
0.18	0.60	0.74	0.37	0.55	0.95	0.54	0.64	2007
0.21	0.66	0.77	0.39	0.62	1.00	0.46	0.67	2008
0.25	0.69	0.80	0.33	0.67	0.83	0.50	0.72	2009

Table 3. *Cont.*

GDP	Urban Population Ratio	Total Population	Contribution Rate of Secondary Industry	Total Energy Production	Proportion of Raw Coal to Energy Consumption	Energy Processing and Conversion Efficiency	Total Energy Consumption	Year
0.30	0.74	0.83	0.47	0.71	0.30	0.65	0.79	2010
0.38	0.79	0.85	0.66	0.81	0.72	0.71	0.87	2011
0.45	0.84	0.88	0.46	0.92	1.00	0.68	0.92	2012
0.50	0.88	0.91	0.38	0.96	0.72	0.73	0.96	2013
0.59	0.920	0.940	0.33	0.99	0.58	0.75	0.99	2014
0.71	0.955	0.971	0.30	1.00	0.26	0.87	1.00	2015

4.3. Principal Component Analysis

Multiple collinearity leads to a problem where a minor change in one variable can lead to a significant alteration of the coefficient, thereby affecting the reliability of the results. Furthermore, for neural networks, if the variables in the input data exhibit collinearity, the finite number of input nodes will be "wasted"; that is, when there is only a certain number of input nodes, collinear variables contain repeated information, so all these collinear variables are included in the input node, which does not increase the amount of information, and these occupied nodes can be allocated to other factors, so the number of nodes is wasted. In order to eliminate the influence of multiple collinearity, it is essential to reduce the data dimensionality through the composite data index. Data dimensionality reduction can improve data quality, applicability, and accuracy, while also minimizing the influence of erroneous data. This paper will employ a principal component analysis (PCA) method to reduce the dimensionality of data and synthesize a new composite index. The principal component analysis is performed based on the experimental data that have been predicted and adjusted in advance. And 'Energy processing and transformation efficiency', 'Urban population ratio', and 'GDP' are selected as the principal components, which are represented by P1, P2, and P3 respectively. New experimental data are obtained, as shown in Figure 8.

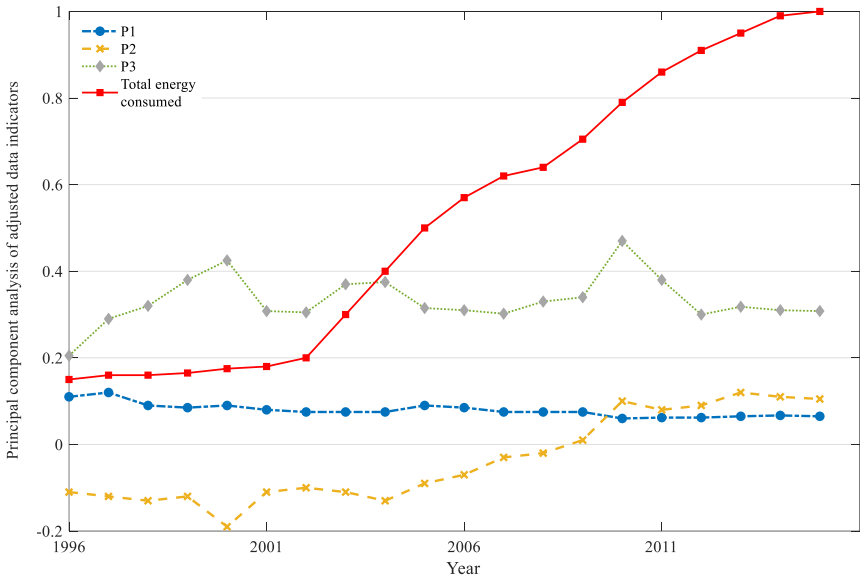
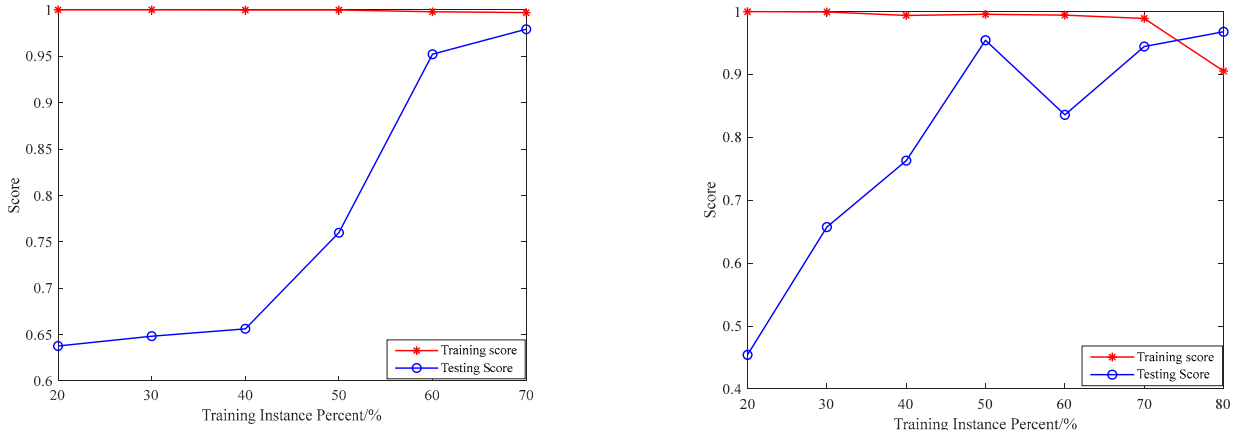


Figure 8. Adjusted data indicators of principal component analysis.

4.4. BP Neural Network Solving Model

The experimental data in Figure 9a,b are trained with the neural network, respectively, considering the forecast lead time and not considering the forecast lead time. A total of 70% of the experimental data were selected as training data and 30% as test data.



(**a**) After adjustment (**b**) Before adjustment

Figure 9. Learning curves.

A three-layer neural network model is established, in which the number of neural nodes in the input layer is the number of influencing factors in the two tables. The number of output neurons, that is, the value for energy consumption, is 1. The number of neurons in the hidden layer is set to 0.5, 1, and 1.5 times the number of neurons in the input layer, respectively. When the number of neurons in the hidden layer is set to 1.5 times the number of neurons in the input layer, both the error and convergence effect are good (0.5 times: the number of convergence iterations is 1893 generations, and the average error of the standard error is 0.75; when the number of neurons in the hidden layer is set to a factor of 1, the number of convergence iterations is 961 generations, and the average error of the standard error is 0.28; when the number of neurons in the hidden layer is set to a factor of 1.5, the number of convergence iterations is 374, and the average error is 0.01). Therefore, set the number of neurons in the hidden layer to 1.5 times the number of neurons in the input layer. The error capacity is set to 0.01, the learning rate is 0.6, the number of iterations is set to 2000, and the momentum factor is 0.8. Set the maximum number of iterations to less than the error capacity to achieve convergence.

Comparing the situations with and without consideration of the forecast lead time, from the experimental results in Table 4, it can be seen that the neural network algorithm when considering the forecast lead time is better than the neural network algorithm without considering the forecast lead time in terms of convergence speed, evaluation error, relative error, and other indicators.

Figure 10 shows the error iteration curves of the two types of experiments. The horizontal axis represents the number of training iterations, and the vertical axis represents the sum of the squares of the errors of each generation. It can be intuitively observed from Figure 10 that the error convergence speed of the neural network training speed considering the forecast lead time is significantly faster than that when not considering the forecast lead time. If the forecast lead time is considered, it will converge at 374 generations, but if the forecast lead time is not considered, it will converge when the maximum number of iterations is reached, and it will tend to be stable at 1600 generations, and it will achieve near convergence when the maximum number of iterations is set at 2000 generations. The results demonstrate that the training speed of a neural network when considering the forecast lead time is markedly superior than when not considering the forecast lead time.

Table 4. Experimental calculation results.

	Actual Value of Experimental Data	Experimental Results		Relative Error	
		Considering Forecast Lead Time	Not Considering Forecast Lead Time	Considering Forecast Lead Time	Not Considering Forecast Lead Time
Error convergence Iterations	-	374	2000	-	-
Training time	-	13s	21s	-	-
Average error (Average after sum of squares)	-	0.01	0.05	-	-
Test data prediction value	0.67	-	0.6	-	8.59%
	0.72	-	0.53	-	26.42%
	0.79	0.83	0.60	5.00%	23.26%
	0.87	0.81	0.70	6.95%	19.19%
	0.92	0.81	0.70	11.33%	23.13%
	0.96	0.85	0.71	12.01%	25.99%
	0.99	0.84	0.68	14.56%	30.13%
	1.00	0.84	0.69	15.58%	30.91%

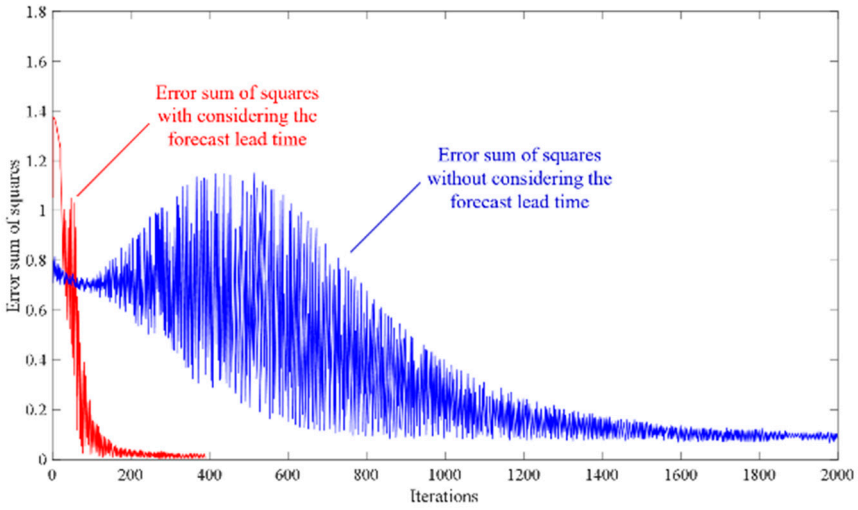


Figure 10. Comparison of error convergence curves.

Figure 11 shows the comparison between the measured data of the experimental prediction results and the original data. The solid line represents the predicted results when considering the forecast lead time, and the star point represents the actual value of the corresponding year after the standardization of energy consumption. The dotted line indicates the predicted results when not considering the forecast lead time, and the dot indicates the actual value corresponding to the corresponding year. In Figure 11, the error between the predicted value and the actual value when considering the forecast lead time is significantly less than that when not considering the forecast lead time, which shows that the neural network's prediction accuracy when considering the forecast lead time is higher than that when not considering the forecast lead time.

We compared the forecast value without considering the forecast lead time by SVM [51], the forecast value without considering the forecast lead time by RF [52], and the forecast value considering the forecast lead time by the BP network with actual values. The result is shown in Figure 12. It can be seen that the prediction accuracy of the BP neural network when considering the forecast lead time is higher than when not.

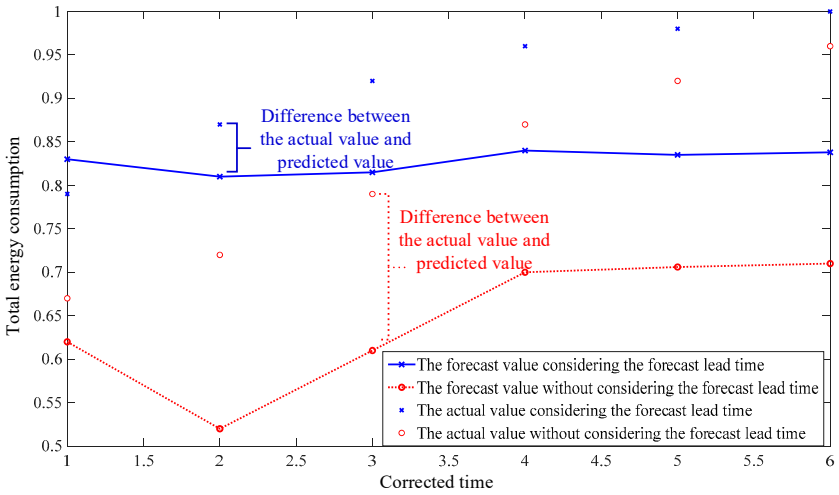


Figure 11. Comparison of prediction experiment results.

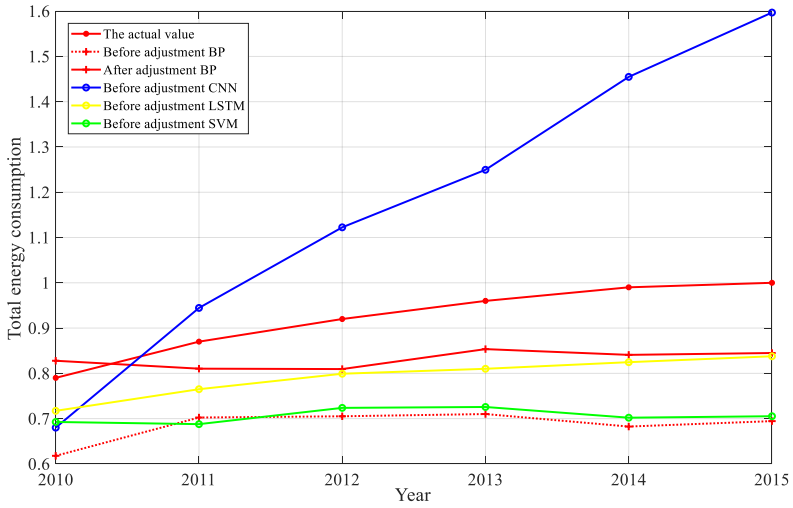


Figure 12. Comparison between improved BP and other algorithms.

We used three different neural networks, CNN [53], BP, and LSTM [54], for prediction. The results are shown in Figure 13. The error between the forecast value and the actual value considering the forecast lead time is less than when not considering the forecast lead time.

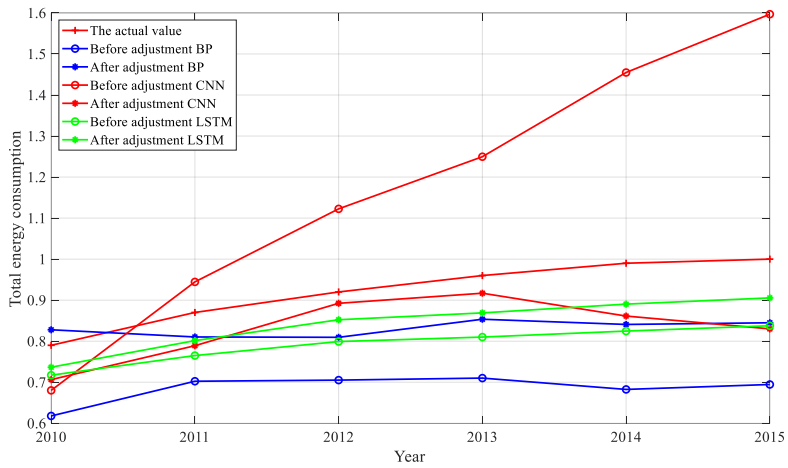


Figure 13. Comparison between considering forecast lead time and not considering forecast lead time of BP, CNN, and LSTM.

Additionally, we compared the prediction results of the three neural network algorithms considering and not considering the forecast lead time, using the Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) as our evaluation standards. The results are shown in Table 5.

Table 5. Comparison of experimental results of different algorithms.

	Experimental Results					
	Before Adjustment BP	**After Adjustment BP**	**Before Adjustment CNN**	**After Adjustment CNN**	**Before Adjustment LSTM**	**After Adjustment LSTM**
Mean Squared Error	0.059047	0.012479	0.11914	0.010275	0.017851	0.006546
Mean Absolute Percentage Error	25.3439%	10.8894%	30.2157%	9.5658%	13.8295%	8.4989%
Mean Absolute Error	0.23627	0.10314	0.2897	0.089017	0.12938	0.079135

From the comparison results in Table 5, we can find that regardless of the type of neural network algorithm, the prediction results when considering the forecast lead time are more effective than those when not considering the forecast lead time, which shows that forecast lead time can significantly improve prediction accuracy, and this concept also applies to other neural network algorithms.

5. Conclusions

On the basis of using BP neural network to predict China's energy consumption, this paper introduces the concept of forecast lead time, sets the influencing factor index and energy consumption index of the neural network input layer through the designed forecast lead time solution model, and then improves the neural network prediction process. At the same time, through the analysis of practical examples, the training and prediction results of the neural network when not considering the forecast lead time were compared with those considering the forecast lead time. By adjusting the influencing factor indicators according to the forecast lead time, the convergence speed could be increased by 38% on average, and the relative error could be reduced by 15%, which can effectively improve the prediction effect of the neural network, and it also has obvious advantages in the prediction accuracy and network training speed. However, this model is more applicable to short-term energy consumption forecasts. Due to the small forecast lead time, for long-term energy consumption forecast, future research needs to improve the forecast lead time of influencing factor indicators through time series and other methods, and explore the adjustment of the forecast lead time in the process of neural network backpropagation.

Author Contributions: Conceptualization, J.B.; methodology, J.W.; software, C.T.; validation, writing—review and editing, J.W.; supervision, J.B.; project administration, X.L.; funding acquisition, J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the "Tianchi Talent" Introduction Plan Leading Innovative Talents Project of Xinjiang, National Social Science Foundation of China, under Grant 21CJY051.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study, and written consent was obtained from the patients to publish this paper.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. National Bureau of Statistics. *2023 China Statistical Yearbook*; China Statistical Publishing House: Beijing, China, 2023.
2. Geem, Z.W. Transport energy demand modeling of South Korea using artificial neural network. *Energy Policy* **2011**, *39*, 4644–4650. [CrossRef]
3. Yu, L.A.; Zhao, Y.Q.; Tang, L.; Yang, Z.B. Online big data-driven oil consumption forecasting with Google trends. *Int. J. Forecast.* **2019**, *35*, 213–223. [CrossRef]
4. Sözen, A.; Gülseven, Z.; Arcaklioğlu, E. Forecasting based on sectoral energy consumption of GHGs in Turkey and mitigation policies. *Energy Policy* **2007**, *35*, 6491–6505. [CrossRef]
5. Kankal, M.; Akpınar, A.; Kömürcü, M.İ.; Özşahin, T.Ş. Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. *Appl. Energy* **2011**, *88*, 1927–1939. [CrossRef]
6. Ju, Y.; Sun, G.; Chen, Q.; Zhang, M.; Zhu, H.; Rehman, M.U. A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting. *IEEE Access* **2019**, *7*, 28309–28318. [CrossRef]
7. Song, X.Y.; Liu, Y.; Xue, L.; Wang, J.; Zhang, J.; Wang, J.; Jiang, L.; Cheng, Z. Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network mode. *J. Pet. Sci. Eng.* **2020**, *186*, 106682. [CrossRef]
8. Alsina, E.F.; Bortolini, M.; Gamberi, M.; Regattieri, A. Artificial neural network optimisation for monthly average daily global solar radiation prediction. *Energy Convers. Manag.* **2016**, *120*, 320–329. [CrossRef]
9. Yu, F.; Xu, X.Z. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Appl. Energy* **2014**, *134*, 102–113. [CrossRef]
10. Sun, W.; Huang, C.C. A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. *J. Clean. Prod.* **2020**, *243*, 118671. [CrossRef]
11. Wang, D.Y.; Luo, H.Y.; Grunder, O.; Lin, Y.B.; Guo, H.X. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. *Appl. Energy* **2017**, *190*, 390–407. [CrossRef]
12. Zeng, Y.R.; Zeng, Y.; Choi, B.; Wang, L. Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. *Energy* **2017**, *127*, 381–396. [CrossRef]
13. Cellura, M.; Brano, V.L.; Marvuglia, A. Forecasting daily urban electric load profiles using artificial neural networks. *Energy Convers. Manag.* **2004**, *45*, 2879–2900.
14. Wang, S.X.; Zhang, N.; Wu, L.; Wang, Y.M. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [CrossRef]
15. Ren, C.; An, N.; Wang, J.Z.; Li, L.; Hu, B.; Shang, D. Optimal parameters selection for BP neural network based on particle swarm optimization: A case study of wind speed forecasting. *Knowl. Based Syst.* **2014**, *56*, 226–239. [CrossRef]
16. Zhang, Y.G.; Chen, B.; Pan, G.F.; Zhao, Y. A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting. *Energy Convers. Manag.* **2019**, *195*, 180–197. [CrossRef]
17. Ding, L.; Lv, Z.; Han, M.; Zhao, X.; Wang, W. Forecasting China's wastewater discharge using dynamic factors and mixed-frequency data. *Environ. Pollut.* **2019**, *255*, 113148. [CrossRef]
18. Tsekouras, G.; Dialynas, E.; Hatziargyriou, N.; Kavatza, S. A non-linear multivariable regression model for midterm energy forecasting of power systems. *Electr. Power Syst. Res.* **2007**, *77*, 1560–1568. [CrossRef]
19. Wang, Y.; Hu, Q.; Meng, D.; Zhu, P. Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model. *Appl. Energy* **2017**, *208*, 1097–1112. [CrossRef]
20. Fan, G.; Peng, L.; Hong, W. Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model. *Appl. Energy* **2018**, *224*, 13–33. [CrossRef]
21. Wu, J.; Wang, J.; Lu, H.; Dong, Y.; Lu, X. Short term load forecasting technique based on the seasonal exponential adjustment method and the regression model. *Energy Convers. Manag.* **2013**, *70*, 1–9. [CrossRef]
22. Yolcu, C.O.; Lam, K.H.; Yolcu, U. Short-term load forecasting: Cascade intuitionistic fuzzy time series—Univariate and bivariate models. *Neural Comput. Appl.* **2024**, *36*, 20167–20192. [CrossRef]
23. Gupta, A.; Kumar, A.; Khatod, K.D. Optimized scheduling of hydropower with increase in solar and wind installations. *Energy* **2019**, *183*, 716–732. [CrossRef]
24. Jamil, R. Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030. *Renew. Energy* **2020**, *154*, 1–10. [CrossRef]
25. Sen, P.; Roy, M.; Pal, P. Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. *Energy* **2016**, *116*, 1031–1038. [CrossRef]
26. Hussain, A.; Rahman, M.; Memon, A.J. Forecasting electricity consumption in Pakistan: The way forward. *Energy Policy* **2016**, *90*, 73–80. [CrossRef]
27. Martínez, F.; Frías, P.M.; Pérez-Godoy, D.M.; Rivera, A.J. Dealing with seasonality by narrowing the training set in time series forecasting with kNN. *Expert Syst. Appl.* **2018**, *103*, 38–48. [CrossRef]
28. Lang, K.; Zhang, M.; Yuan, Y.; Yue, X. Short-term load forecasting based on multivariate time series prediction and weighted neural network with random weights and kernels. *Clust. Comput.* **2019**, *22*, 12589–12597. [CrossRef]
29. Tsai, S.-B. Using grey models for forecasting China's growth trends in renewable energy consumption. *Clean Technol. Environ. Policy* **2016**, *18*, 112–135. [CrossRef]

30. Li, R. Application of a Modified Grey Model Based on Least Squares in Energy Prediction. *Adv. Comput. Signals Syst.* **2023**, *7*, 112–145.
31. Lu, S. Integrating heuristic time series with modified grey forecasting for renewable energy in Taiwan. *Renew. Energy* **2019**, *133*, 1436–1444. [CrossRef]
32. Xie, N.; Yuan, C.; Yang, Y. Forecasting China's energy demand and self-sufficiency rate by grey forecasting model and Markov model. *Int. J. Electr. Power Energy Syst.* **2015**, *66*, 1–8. [CrossRef]
33. Huiping, W.; Zhun, Z. Forecasting the renewable energy consumption of Australia by a novel grey model with conformable fractional opposite-direction accumulation. *Environ. Sci. Pollut. Res. Int.* **2023**, *30*, 104415–104431.
34. Wu, L.; Gao, X.; Xiao, Y.; Yang, Y.; Chen, X. Using a novel multi-variable grey model to forecast the electricity consumption of Shandong Province in China. *Energy* **2018**, *157*, 327–335. [CrossRef]
35. Wang, Z.; Li, Q.; Pei, L. A seasonal GM(1,1) model for forecasting the electricity consumption of the primary economic sectors. *Energy* **2018**, *154*, 522–534. [CrossRef]
36. Cao, W.; Liu, Y.; Mei, H.; Shang, H.; Yu, Y. Short-term district power load self-prediction based on improved XGBoost model. *Eng. Appl. Artif. Intell.* **2023**, *126*, 1223–1245. [CrossRef]
37. Grzegorz, D. A Comprehensive Study of Random Forest for Short-Term Load Forecasting. *Energies* **2022**, *15*, 7547. [CrossRef]
38. Peng, L.L.; Fan, G.F.; Yu, M.; Chang, Y.C.; Hong, W.C. Electric Load Forecasting based on Wavelet Transform and Random Forest. *Adv. Theory Simul.* **2021**, *4*, 211–243. [CrossRef]
39. Hosein, E.; Maryam, I.; Parsa, M.M. Convolutional and recurrent neural network based model for short-term load forecasting. *Electr. Power Syst. Res.* **2021**, *195*, 107173.
40. Li, J.; Lei, Y.; Yang, S. Mid-long term load forecasting model based on support vector machine optimized by improved sparrow search algorithm. *Energy Rep.* **2022**, *8*, 491–497. [CrossRef]
41. Faruque, M.O.; Rabby, M.A.J.; Hossain, M.A.; Islam, M.R.; Rashid, M.M.U.; Muyeen, S.M. A comparative analysis to forecast carbon dioxide emissions. *Energy Rep.* **2022**, *8*, 88046–88060. [CrossRef]
42. Ameyaw, B.; Yao, L.; Oppong, A.; Agyeman, J.K. Investigating, forecasting and proposing emission mitigation pathways for CO_2 emissions from fossil fuel combustion only: A case study of selected countries. *Energy Policy* **2019**, *13*, 7–21. [CrossRef]
43. Abdel-Basset, M.; Hawash, H.; Chakrabortty, R.K.; Ryan, M. PV-Net: An innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production. *J. Clean. Prod.* **2021**, *303*, 127037. [CrossRef]
44. Bendaoud, N.M.M.; Nadir, F.; Samir, A.B. Comparing Generative Adversarial Networks architectures for electricity demand forecasting. *Energy Build.* **2021**, *247*, 111152. [CrossRef]
45. Keddouda, A.; Ihaddadene, R.; Boukhari, A.; Atia, A.; Arıcı, M.; Lebbihiat, N.; Ihaddadene, N. Solar photovoltaic power prediction using artificial neural network and multiple regression considering ambient and operating conditions. *Energy Convers. Manag.* **2023**, *288*, 117186. [CrossRef]
46. Guo, H.; Chen, Q.; Xia, Q.; Kang, C.; Zhang, X. A monthly electricity consumption forecasting method based on vector error correction model and self-adaptive screening method. *Int. J. Electr. Power Energy Syst.* **2018**, *95*, 427–439. [CrossRef]
47. He, Y.; Zheng, Y.; Xu, Q. Forecasting energy consumption in Anhui province of China through two Box-Cox transformation quantile regression probability density methods. *Measurement* **2019**, *136*, 579–593. [CrossRef]
48. Rhif, M.; Abbes, A.B.; Martínez, B.; Farah, I.R. Veg-W2TCN: A parallel hybrid forecasting framework for non-stationary time series using wavelet and temporal convolution network model. *Appl. Soft Comput. J.* **2023**, *137*, 110172. [CrossRef]
49. Abedinia, O.; Amjady, N.; Ghadimi, N. Solar energy forecasting based on hybrid neural network and improved metaheuristic algorithm. *Comput. Intell.* **2018**, *34*, 241–260. [CrossRef]
50. Wang, Q.; Li, S.Y.; Li, R.R.; Jiang, F. Underestimated impact of the COVID-19 on carbon emission reduction in developing countries-A novel assessment based on scenario analysis. *Environ. Res.* **2022**, *204*, 111990. [CrossRef]
51. Nijhawan, P.; Bhalla, K.V.; Singla, K.M.; Gupta, J. Electrical Load Forecasting using SVM Algorithm. *Int. J. Recent Technol. Eng. (IJRTE)* **2020**, *8*, 4811–4816. [CrossRef]
52. Wang, Y.; Sun, S.; Cai, Z. Daily Peak-Valley Electric-Load Forecasting Based on an SSA-LSTM-RF Algorithm. *Energies* **2023**, *16*, 7964. [CrossRef]
53. Wang, D.; Li, S.; Fu, X. Short-Term Power Load Forecasting Based on Secondary Cleaning and CNN-BILSTM-Attention. *Energies* **2024**, *17*, 4142. [CrossRef]
54. Qi, Y.; Luo, H.; Luo, Y.; Liao, R.; Ye, L. Adaptive Clustering Long Short-Term Memory Network for Short-Term Power Load Forecasting. *Energies* **2023**, *16*, 6230. [CrossRef]

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