8-2018

Using multifactor inputs bp neural network to make power consumption prediction

Hao Song
Binghamton University--SUNY, hsong27@binghamton.edu

USING MULTIFACTOR INPUTS BP NEURAL NETWORK TO MAKE POWER CONSUMPTION PREDICTION

BY

HAO SONG

BS, Tianjin University, 2016

THESIS

Submitted in partial fulfillment of the requirements for
the degree of Master of Science in Electrical Engineering
in the Graduate School of
Binghamton University
State University of New York
2018

Accepted in partial fulfillment of the requirements for
the degree of Master of Science in Electrical Engineering
in the Graduate School of
Binghamton University
State University of New York
2018

June 1, 2018

Dr. Yu Chen, Chair & Faculty Advisor
Department of Electrical and Computer Engineering, Binghamton University

Dr. Ning Zhou
Department of Electrical and Computer Engineering, Binghamton University

Dr. Ziang Zhang
Department of Electrical and Computer Engineering, Binghamton University

Abstract

With the development of modern information and technology (IT), smart grids became one of the major components of smart cities, to take full advantage of the smart grid, the capability of intelligent scheduling and planning of electricity delivery is essential. For this purpose, researchers have investigated methodologies for power consumption prediction and demand side management (DSM). In addition, conducting a comprehensive analysis and obtaining an accurate evaluation of power consumption are the premise and basis for a more robust and efficient power grid design and transformation. Therefore, it is meaningful to explore forecasting models that are able to reflect the power consumption change effectively.

Making electricity consumption prediction based on neural network has been a popular research topic in recent years, and backpropagation neural network (BPNN) algorithm has been recognized as a mature and effective method. This thesis applies the BPN to predict the electricity consumption of Pecan Street, a community with a relatively large scale smart grid, and takes more factors into account, such as weather condition, weekend and holiday. The influences of each factor have been evaluated for a deeper insight. While what presented in this thesis is not mature, it may inspire more discussion and further study to guide the design of future smart grids.

Acknowledgement

First of all, I would like to express my sincerest appreciation to my advisor, Professor Yu Chen, for his great encouragement, patience and guidance throughout my graduate study in Binghamton University, not only in academic research, but also in my daily life. He is one of my greatest mentors. Also, I want to thank Professor Ning Zhou and Professor Ziang Zhang to serve on my committee.

In addition, I am grateful to my laboratory and group companions, Ronghua Xu, Qi Dong, Zekun Yang, Seyed Yahya Nikouei. They gave me a lot of constructive suggestions and advice.

The last but not the least, I deeply thank my girlfriend and my parents. They gave me mental and physical support and take care of my needs in daily life. Without their love, care, and constant support, I would not be able to finish this thesis work.

Table of Contents

List of Figures

List of Abbreviations

IT – Information Technology

DSM - Demand Side Management

ANN – Artificial Neural Network

BP - Back Propagation

BPNN – Back Propagation Neural Network

MATLAB - Matrix laboratory

ERT - Encoded Radio Transmission

NCDC – National Climatic Data Center

MSE – Mean Square Error

SVM – Support Vector Machine

PSO – Particle Swarm Optimization

GA – Genetic Algorithm

GM – Grey Model

Chapter 1. Introduction

Electricity consumption prediction has been considered an effective measure that helps the power grid designers and planners build robust, adaptive, efficient, and economic smart grids. It is aimed as modeling the electricity consumption under different constraints along with environmental factors and the rules. A pre-estimated and calculated electricity demand can be obtained based on the history data including dates, economic, climate and so on. Considering the dynamic pricing mechanism in today's market, an accurate power load forecasting is an effective tool for companies to optimize the scheduling load balance decisions to maximize their profit and minimize the probability of accidents like disturbance, overload, etc. Different prediction periods and precision are required for the large scale and complicated smart power grid systems.

This thesis is focused on the effects of different factors that may bring impacts on the prediction accuracy. For this purpose, leveraging a backpropagation neural network (BPNN) algorithm will allow multiple factors to be considered and their impacts studied. The accurate analysis and prediction of electrical consumption will help government agencies and the power industry make appropriate electricity utility policies and power scheduling plans. For individual households and the communities on the smart grid, the prediction will help people arrange their electricity usage with more intelligence.

1.1 Motivation

Electricity consumption forecasting is a work which is easy to iterate, but the amount of effort required to improve the quality for each incremental step is huge. The comprehensive consideration of various influencing factor, and the analysis and utilization of diverse types of data to the electricity consumption forecasting module are the requirements of the modern smart power grid. Mastering the way of electricity consumption forecasting with a decent prediction accuracy is a foundation for regional electric power planning, as well as the region's industrial layout, energy distribution, electric power dispatching and power grid investment as a reliable reference.

For better competition in the power market, participants need to accurately predict how much power they will need in a given cycle. On one hand, the underestimation of power demand will lead to higher operation cost [1], which cannot meet the development needs of local economy. On the other hand, overestimation of power demand results in the waste of power resources and investment costs. Therefore, electricity consumption forecasting is one of the most essential tasks in the power market. Electricity consumption prediction can be classified into super-short, short, medium and long term, based on the prediction cycles.

Currently, more and more researchers turn their attention to increase the accuracy of the electricity consumption prediction using multiple factors. Because of the limited access of real-world data set, most of the reported works are focused on a single factor or only historical data. There is not much reported research using multiple factors.

1.2 Contributions

This thesis considers the historical electricity consumption, the weather, and weekend/holiday information to get a better performance of prediction. The effects of different factors are analyzed for deeper understanding. More specifically, using the historical data of the Pecan Street smart community and the corresponding weather information, a BP neural network-based prediction has been conducted. The performances under different time resolutions, hourly, daily, weekly, and monthly, are investigated. In addition, the optimal setting of the BPN is explored according to the data set features by adjusting the parameters to reach the highest performance.

The rest of thesis is organized as follows. In Chapter 2, the research background and related work are introduced. Chapter 3 reviews the principles and design of the BP neural network model for power consumption prediction, The Pecan Street data set and its process are discussed in Chapter 4. Chapter 5 reports the experimental results in detail. Chapter 6 is the conclusion.

Chapter 2 Background and Related Work

2.1 Pecan Street Project

The Pecan Street Project is an advanced smart community project which located in Austin, Texas, USA. Technologies implemented in the participating homes include energy management systems, distributed solar photovoltaic energy, plug-in electric vehicles, smart meters, distributed energy storage, smart appliances, in-home displays, and programmable communicating thermostats [2].

The Pecan Street Smart Grid maintains over 1,000 households who shared their home or businesses' electricity consumption data with the project, through the methods such as green button protocols, smart meters, home energy monitoring system and so on. The households in the Pecan Street Project were just like pioneers, they have great interest in smart community products and services. They have relatively high education and income level in Texas State [2].

Through the Pecan Street Project, massive data can be obtained. Electricity utility data is available with 1-hour resolution, electric data at 1- minute resolution. The gas meter data and the water meter data are collected by ERT (encoded radio transmission). These data have some notable features for researching and modeling: large quantity, high resolution, sustainable access, high reliability and high integrity. It is an ideal candidate for modeling, forecasting and analyzing.

2.2 Energy Consumption Prediction

The relationship between the residential electricity consumption and influence factors is not a simple linear relationship. Quantity analysis-based electricity consumption prediction may get opposite results or unsatisfactory performance [3]. Researchers pay more attention to more intelligent algorithms and models. At first, the most popular method was linear regression [4]. Nowadays, new algorithms like the grey forecasting model, artificial neural network, support vector machine and their corresponded optimization and deformation algorithms becomes more and more popular and mature [5].

Currently, the research on residential electricity consumption is mainly based on household economic theory [6]. Every household purchases electricity corresponding to the household electrical appliance. If there are sufficient data, the model of residential electricity consumption based on household economic theory contains many significant factors, such as electricity price, household income, personal income, alternative energy price, household electrical appliances price, citizens population density, household size, and household area [7-11]. And there are many other factors which may have significant impacts on electricity preference, such as weather or climate condition, holiday, weekend and so on [12].

In some developed countries and areas, such as Europe, the United States, Japan, Hong Kong, there are many researches on multifactor residential electricity consumption models [13-16], due to the powerful data collection work system. But most of these researches only consider one or few of the factors mentioned above. Some researches even only focused on historical records of the electricity consumption and ignore all other factors.

Some researchers have done multiple indicators annual prediction. Most of them introduced the factors of installed power capacity, historical yearly electricity consumption, gross domestic product, popularity, imports, exports and so on [13], [14], [20]. Some of them only consider the historical consumption data [16]. According to the existing records and experimental results, the annual prediction is suitable for extra-large zone (e.g. a country). And when it is accessible to obtain data like gross domestic product and popularity, the artificial neural network (ANN) is mostly applied, if not, the grey model can take its advantages [16], [21].

For short-term and medium-term predictions, the ANN is also widely used.

Many researchers have done optimizations based on the neural network itself. GA, PSO and Elman Neural Network [15], [19], [22], [27], [29] are some examples that have good performance. After optimization, the training speed, the prediction accuracy become better at certain degree.

Some researchers have made changes on input factors or indicators, for example, introducing temperature, weekday or weekend, seasons [18], [23], [25]. According to some experimental results, changes of the number of historical consumption data will affect the prediction performance [26], [31], [32], [33], like the results of 1 hour before input and 24 hours input are different. Also, for the structure and parameters of neural network, there are more spaces to adjust, like the number of layers, the number of neurons in each hidden layer, the learning rules, the transfer function between each layer and so on [13], [20].

Besides ANN and GM (Grey Model), there are many excellent algorithms and models, such as SVM (Support Vector Machine), regression analysis, detail model simulation, statistical methods, and decision tree and so on. There are many comparisons among these methods [13], [18], [20], [23], [24], [30]. For different cases, different areas, different data set types, these methods have different performance and adaptation, and different methods have different model complexity, usability, running speed, input needs and accuracy. For example, in principle, the regression analysis does not have higher accuracy than ANN, but its model complexity is lower than ANN, sometimes the cost performance ratio is an important consideration basis.

2.3 BP Neural Network

Neural network is a complex nonlinear system that consists of numerous neurons. In this system, every neuron has a relatively simple function and construction. However, when they are merged together into the entire system, the behavior can be very complex. In artificial neural networks, strength and condition of every connection between nodes are adjustable, it has strong ability of self-learning and self-adaption. The artificial neural network can be applied to many aspects and research areas. Dividing the data samples into three parts: the training data set, the validation data set and the testing data set. The training data set and the validation data set are used in the process of "training", and the validation data set is randomly picked up from training data set in some proportion.

In an artificial neural network, the neurons can be classified into three types according to their position and the information they process: input units, hidden units and output units. The input units receive the input information of the system, which represent the outside signals or data. The output unit give the output after neural network processes, which represent the result. Hidden units form a layer between the input units and the output units. While they don't represent any information of the entire system, but they are significant to the entire neural network and have profound impact on the prediction results. The connection between each neuron mainly reflect the process of information from input to output. This process is repeated many epochs, it is the important part of artificial neural network learning and training.

The BP Neural Network is one of the artificial neural networks that are widely used in many research areas. This technique is also sometimes called backward propagation of errors are calculated because the error is calculated at the output layer and feedback through the network layers.

There are two main processes in BP Neural Network learning: The first one is propagation, which include the generation of the output from each layer and the error (the difference between actual output and target value); the second one is updating the weight.

For weight update, multiplying the error of weight's output and input activation, then finding the gradient of the weight. A ratio of the gradient of the weight is subtracted from the weight. This ratio is named as learning rate which can affect the training speed and performance.

If the learning rate is low, the training will become more reliable, but the optimization will take a long time, because each step of the minimum value of the orientation loss function is small.

If the learning rate is high, the training may not converge at all or even spread out. The change in weight can be so large that the optimization goes over the minimum, making the loss function worse.

The weights need to be updated in the opposite direction of the gradient, thus this method is called as gradient descent.

Chapter 3 BP Neural Network for Power Consumption Prediction

The BP neural network is an algorithm which calculates the errors and propagates the error in the opposite direction of network computation. The simplest BP neural network has three layers. There is no connection between neurons in a single layer, and no direct connection between the input layer and the output layer.

The BP neural network algorithm is one of the most widely used ANN models. It is a multi-layer feedforward network and its key feature is back propagating the error. It is applied to learn and memory huge amount of mapping relations of input-output models, and there is no need to disclose in advance the mathematical equation that describes these mapping relations. Its learning rule is to adopt the steepest descent method, where the back propagation is used to regulate the weight value and threshold value of the network to achieve the minimum error sum of square.

3.1 BP Neural Network Model

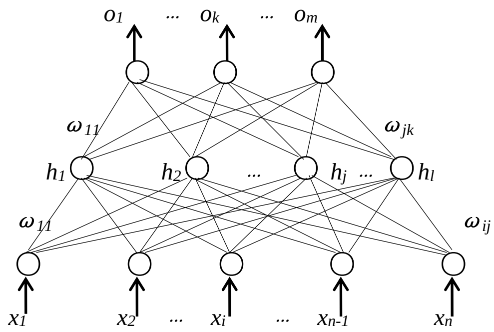

Figure 3.1 Structure of BP Neural Network.

In Figure 3.1, x_1, x_2, …, x_n are the neurons in the input layer, h_1, h_2, …, h_n are the neurons in the hidden layer, o_1, o_2, …, o_n are the neurons in the output layer. ω_{ij} is the weight from neuron *i* in the input layer to neuron *j* in the hidden layer. ω_{jk} is the weight from neuron *j* in the hidden layer to neuron *k* in the output layer.

As shown by Fig. 3.2, the BP learning process can be described as follows:

1. Forward propagation of operating signals: the input signals are propagated from the input layer via the hide layer to the output layer. During the forward propagation of operating signals, the weight values and offset values of the network are constant. The status of each layer of the neurons will only exert an effect on the next layer of the neurons. In case that the expected output cannot be achieved in the output layer, it can be switched into the back propagation of error signal.
2. Back propagation of error signals: the difference between the actual output and

the expected output of the network is defined as the error signals. In the back propagation, the error signals are propagated from the output end to the input layer in a layer-by-layer manner. During the back propagation of error signals, the weight values of network are regulated by the error feedback. The continuous modification of weight values and offset values is applied to make the obtained output of network be closer and closer to the expected one.

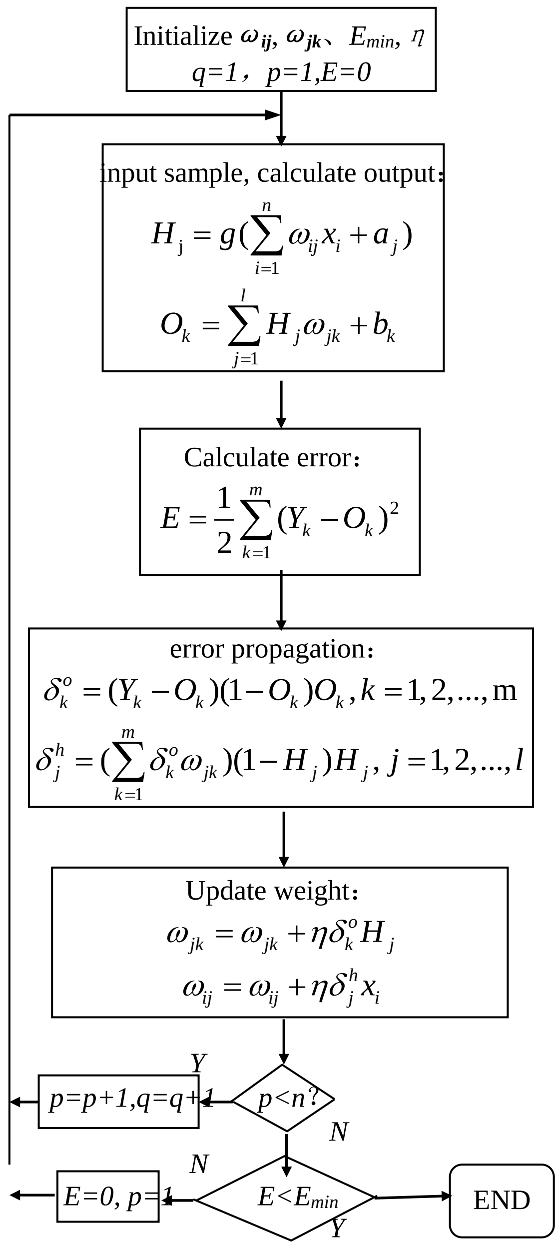

Figure 3.2 Flow Chart of Error Calculation.

3.2 Theoretical Analysis

In BP neural networks, the activation function of neurons is a simulation of mathematic processes between each layer of neurons. The mathematic functions reflect the relationship between each layer. In the traditional 3-layer BP neural network, the most commonly used activation function is a standard sigmoid function. The standard sigmoid function's mathematic expression is shown below:

$$g(x) = \frac{1}{1+e^{-x}} \tag{1}$$

The first work is network initialization. Let's define the following parameters:

- n: the number of input layer nodes;
- l: the number of hidden layer nodes;
- m: the number of output layer nodes;
- ω_{ij}: the weight from neuron i of the input layer to neuron j of the hidden layers;
- ω_{jk}: the weight from neuron j of the hidden layer to neuron k of the output layer;
- a_j: the bias from the input layer to the hidden layer;
- b_k: the bias from the hidden layer to the output layer;
- η: the learning rate; and
- $g(x)$: the activation function.

In this case, the $g(x)$ is the sigmoid function. Thus, the output of the hidden layer is

$$H_j = g(\sum_{i=1}^{n} \omega_{ij} x_i + a_j) \tag{2}$$

The output of the output layer is

$$O_k = \sum_{j=1}^{l} H_j \omega_{jk} + b_k \quad (3)$$

The error calculation is

$$E = \frac{1}{2}\sum_{k=1}^{m}(Y_k - O_k)^2 \quad (4)$$

In Eq. (4), Y_k is the expectation output, and O_k is the actual value of output. To simplify the process, the following equation was introduced:

$$Y_k - O_k = e_k \quad (5)$$

This leads us to:

$$E = \frac{1}{2}\sum_{k=1}^{m} {e_k}^2 \quad (6)$$

where *i*, *j*, and *k* are integers, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, l$, and $k = 1, 2, \ldots, m$.

Thus, the weight updating equations are shown below. Eq. (7) shows the weight updating from the input layer to the hidden layer. Eq. (8) shows the weight updating from the hidden layer to the output layer.

$$\omega_{ij} = \omega_{ij} + \eta H_j (1 - H_j) x_i \sum_{k=1}^{m} \omega_{jk} e_k \quad (7)$$

$$\omega_{jk} = \omega_{jk} + \eta e_k H_j \quad (8)$$

The process is explained as follows. During the process of errors back propagation, the target makes the error function maintain the minimum value. In this case, the gradient descent algorithm is adopted, which means the weight update should be in direct proportion to the descent of error gradient.

The updated weight from the hidden layer to the output layer is:

$$\frac{\partial E}{\partial \omega_{jk}} = \sum_{k=1}^{m}(Y_k - O_k)(-\frac{\partial O_k}{\partial \omega_{jk}}) = (Y_k - O_k)(-H_j) = -e_k H_j \tag{9}$$

Thus, the updating formula is:

$$\omega_{jk} = \omega_{jk} + \eta e_k H_j \tag{10}$$

The updated weight from the input layer to the hidden layer is:

$$\frac{\partial E}{\partial \omega_{ij}} = \frac{\partial E}{\partial H_j} \cdot \frac{\partial H_j}{\partial \omega_{ij}} \tag{11}$$

The above equation leads to following process:

$$\begin{aligned} \frac{\partial E}{\partial H_j} &= (Y_1 - O_1)(-\frac{\partial O_1}{\partial H_j}) + ... + (Y_m - O_m)(-\frac{\partial O_m}{\partial H_j}) \\ &= -(Y_1 - O_1)\omega_{j1} - ... - (Y_m - O_m)\omega_{jm} \\ &= -\sum_{k=1}^{m}(Y_k - O_k)\omega_{jk} \\ &= -\sum_{k=1}^{m} e_k \omega_{jk} \end{aligned} \tag{12}$$

Another equation process is:

$$\begin{aligned} \frac{\partial H_j}{\partial \omega_{ij}} &= \frac{\partial g(\sum_{i=1}^{n}\omega_{ij}x_i + a_j)}{\partial \omega_{ij}} \\ &= g(\sum_{i=1}^{n}\omega_{ij}x_i + a_j)\cdot[1 - g(\sum_{i=1}^{n}\omega_{ij}x_i + a_j)]\cdot\frac{\partial(\sum_{i=1}^{n}\omega_{ij}x_i + a_j)}{\partial \omega_{ij}} \\ &= H_j(1 - H_j)x_i \end{aligned} \tag{13}$$

Thus, the updating formula of weight is:

$$\omega_{ij} = \omega_{ij} + \eta H_j(1 - H_j)x_i \sum_{k=1}^{m}\omega_{jk}e_k \tag{14}$$

The bias updating equations are shown below. Eq. (15) shows the bias updating from the input layer to the hidden layer. Eq. (16) shows the bias updating from the hidden layer to the output layer.

$$a_j = a_j + \eta H_j (1 - H_j) \sum_{k=1}^{m} \omega_{jk} e_k \quad (15)$$

$$b_k = b_k + \eta e_k \quad (16)$$

The process of calculation for the bias updating equation is as below, which follows a process similar to the weight updating formula. It also uses the gradient descent algorithm.

The update of bias from the hidden layer to the output layer:

$$\frac{\partial E}{\partial b_k} = (Y_k - O_k)(-\frac{\partial O_k}{\partial b_k}) = -e_k \quad (17)$$

Thus, the bias updating formula is:

$$b_k = b_k + \eta e_k \quad (18)$$

The update of bias from hidden layer to output layer

$$\frac{\partial E}{\partial a_j} = \frac{\partial E}{\partial H_j} \cdot \frac{\partial H_j}{\partial a_j} \quad (19)$$

More specifically:

$$\frac{\partial E}{\partial a_j} = \frac{\partial g(\sum_{i=1}^{n} \omega_{ij} x_i + a_j)}{\partial a_j} \quad (20)$$

$$= g(\sum_{i=1}^{n} \omega_{ij} x_i + a_j) \cdot [1 - g(\sum_{i=1}^{n} \omega_{ij} x_i + a_j)] \cdot \frac{\partial(\sum_{i=1}^{n} \omega_{ij} x_i + a_j)}{\partial a_j}$$

$$= H_j (1 - H_j)$$

and

$$\frac{\partial E}{\partial H_j} = (Y_1 - O_1)(-\frac{\partial O_1}{\partial H_j}) + \ldots + (Y_m - O_m)(-\frac{\partial O_m}{\partial H_j}) \tag{21}$$

$$= (Y_1 - O_1)\omega_{j1} + \ldots + (Y_m - O_m)\omega_{jm}$$

$$= -\sum_{k=1}^{m}(Y_k - O_k)\omega_{jk}$$

$$= -\sum_{k=1}^{m} e_k \omega_{jk}$$

Thus, the updating formula of bias is

$$a_j = a_j + \eta H_j (1 - H_j) \sum_{k=1}^{m} \omega_{jk} e_k \tag{22}$$

3.3 Selection of Parameters

In this project, the initial weight and bias of each layer are randomly picked by the MATLAB neural network tool box. The larger the learning rate, the quicker the learning processes; however, quicker learning processes have lower accuracy. To get a balance between learning speed and accuracy, the learning rate is selected as *0.01*.

There is one node in the output layer and it reflects the predicted electricity consumption.

The number of input layer nodes will vary depending on various factors. For example, if the historical electricity consumption is the only factor considered, there will be fewer nodes examined than when other factors are taken into account, such as the weather and weekend/holiday.

For the number of nodes in the hidden layer, two necessary conditions are considered:

1. The number of nodes in the hidden layer must be fewer than *N-1*. Here *N* is the number of training samples. Otherwise, the error of the network model will have no relationship with the training samples' features and it will race to zero. Thus, there is not a generalization ability of network model; and
2. The number of training samples should be larger than the connection weight of the network model.

According to these two conditions and previous research [34], there are four empirical formulas to quantify the hidden layers.

$$\sum_{i=0}^{n} C_m^i > k \tag{23}$$

where k is the number of training samples, m is the number of nodes in hidden layer, n is the number of nodes in input layer, and i is a constant value locate in [0, n].

The second equation is as follows:

$$n_1 = \sqrt{n+m} + a \tag{24}$$

where n_1 is the number of nodes in the hidden layer, n is the number of nodes in the input layer, m is the number of nodes in the output layer, and a is a constant value locate in [1, 10].

The third equation is as follows:

$$n_1 = \log_2 n \tag{25}$$

where n_1 is the number of nodes in the hidden layer and n is the number of nodes in the input layer.

The fourth equation is as follows:

$$n_1 = 2n+1 \tag{26}$$

where again n_1 is the number of nodes in the hidden layer and n is the number of nodes in the input layer.

In the actual implementation of experimental study, the trial and error method will be used to test these empirical formulas in order to find the optimal number of nodes in the hidden layer. This method will be discussed in Chapter 5.

Chapter 4 Data Processing

4.1 Data acquiring

he historical electricity consumption data is downloaded from the Pecan Street Project, as collected by the Pecan Street Inc. (www.pecanstreet.org). It provides the electricity consumption data per hour in its smart grid community. For this thesis, ten households' electricity consumption were collected. This ten households were randomly picked up from the community. These households have been offering their power consumption information for many years. As such, it is a stable and reliable source of data, which makes it convenient for future research and validation. Considering the size of the data set, two years (2016 and 2017) data are collected, which shows the households' hourly electricity consumption. The data has more than 17,000 record points which meets the requirement for the BP neural network prediction model.

The weather condition data of Austin, Texas was downloaded from the National Climatic Data Center (NCDC). In this thesis, two factors were taken into consideration: temperature and humidity.

Weekend information was gathered through observing general calendar trends. Because different states have different holidays, holiday information was taken from the Office Holidays of Texas State (www.officeholidays.com).

Thus, the raw data set has three main parts: historical electricity consumption, weather information and weekend/holiday information.

4.2 Limitations

The raw data from real-world is not perfect, there are two issues need to be addressed, missing data and data errors.

There are a few data errors where the information was recorded incorrectly. For example, Fig. 4.1 shows the temperature or humidity was recorded as -999.99 for a given hour.

4541	2016/7/9 20:00	59.50357	7	-999.9	-9999

Figure 4.1 Error Data Example.

To deal with this kind of error, the average value of two hours before the recorded error and two hours after the recorded error is used. If x_i is a wrong data, then the refilling value is as below:

$$x_{i\,new} = \frac{x_{i-2} + x_{i-1} + x_{i+1} + x_{i+2}}{4} \qquad (23)$$

In addition to data errors, there are many data points missing. While some missing data consists of single hour, in some cases, there are several continuous hours missing for the entire day. Theoretically speaking, there should be 17544 records for hourly electricity consumption for 2016 and 2017. However, the raw data only contains 17427 records. Even if the missing part is less than 1% of the entire data set, it will bring negative effects on the final prediction accuracy. Since the power consumption information is sequential, any missing data will break the consecutiveness of information. For example, if data is missing from 1 pm to 7 pm, the MATLAB program will automatically read 7 pm data to fill in 2 pm data slot and so on.

When an individual hour record is missing, the error is resolved in a manner similar to correcting data errors, using the average value of two hours before the missing value and two hours after the missing value.

If the missing data covers several continuous hours, a different method is applied. For example, the power consumption records for 2016/07/08 and 2016/07/09 are incomplete because there are many hourly data records missing. Figure 4.2 demonstrates this situation.

4534	2016/7/8 08:00	25.15785
4535	2016/7/8 09:00	28.54575
4536	2016/7/8 10:00	33.12165
4537	2016/7/8 11:00	40.7465
4538	2016/7/8 12:00	43.84068
4539	2016/7/8 13:00	57.37928
4540	2016/7/8 14:00	57.52302
4541	2016/7/9 20:00	59.50357
4542	2016/7/9 21:00	56.26608
4543	2016/7/9 22:00	57.23182
4544	2016/7/9 23:00	49.28663

Figure 4.2 Missing Data Example 1.

In this case, it is not feasible to simply calculate the average value of two hours before and two hours after because so many hours are missing. Instead, the data of two days before the missing value range and two days after the missing value range that cover the missing data range are considered. After identifying these four days, the values of the missing data range are averaged to refill the missing data.

To illustrate this process using the example described in Fig. 4.2, four sets of data are slected: 2016/07/06 15:00 to 2016/07/07 19:00, 2016/07/05 15:00 to 2016/07/06 19:00, 2016/07/10 15:00 to 2016/07/11 19:00, 2016/07/11 15:00 to 2016/07/12 19:00. Then the average value of the data range within the four sets of data is calculated. Then this average value is used to refill the missing data.

There is one case where this resolution did not work. On 2017/12/31, there were only records from 00:00 to 17:00. Many sequential hours were missing. As it was the last day of data set, no following records were available to calculate the average value. To maintain accuracy and performance, the 2017/12/31 record was deleted from entire data set. This occurrence is demonstrated in Fig. 4.3.

2017/12/31 13:00	23.3313	12
2017/12/31 14:00	25.8144	12
2017/12/31 15:00	29.15345	12
2017/12/31 16:00	30.28535	12
2017/12/31 17:00	33.84138	12

Figure 4.3 Missing Data Example 2.

4.3 Data Categories

In order to achieve higher prediction accuracy, the data sets have been adjusted accordingly depending on defferent prediction scales: hourly, daily, weekly and monthly.

For hourly prediction, the data set has 17520 rows and eight columns. Rows represent the hourly time (for example: 6:00, 7:00 and so on). The eight columns follow the given format: electricity consumption (in kWh), month, temperature (in Fahrenheit), humidity (in %), hour (which hour in a day), day (which day in a week), whether or not it is a weekend, and whether or not it is a holiday.

For daily prediction, the data set has 730 rows and nine columns. Each row corresponds to each day. The nine columns follow the given format: electricity consumption (in kWh), highest temperature of the day (in Fahrenheit), lowest temperature of the day (in Fahrenheit), average temperature of the day (in Fahrenheit), highest humidity of the day (in %), lowest humidity of the day (in %), average humidity of the day (in %), whether or not it is a weekend, and whether or not it is a holiday.

For weekly prediction, the data set has 104 rows and eight columns. Each row represents each week. The eight columns follow the given format: electricity consumption (in kWh), highest temperature of the week (in Fahrenheit), lowest temperature of the week (in Fahrenheit), average temperature of the week (in Fahrenheit), the highest humidity of the week (in %), lowest humidity of the week (in %), average humidity of the week (in %), and the number of holidays in the week.

For monthly prediction, the data set has 48 rows and nine columns. Rows represent each month. The nine columns follow the given format: electricity consumption (in kWh), highest temperature of the month (in Fahrenheit), lowest temperature of the month (in Fahrenheit), average temperature of the month (in Fahrenheit), highest humidity of the month (in %), lowest humidity of the month (in %), average humidity of the month (in %), the number of weekends in the month, and the number of holidays in the month.

The daily, weekly and monthly data are actually statistics of the hourly data. For example, to calculate the daily data set, the electricity consumption data is obtained by adding the total hourly electricity consumption on that day. The highest and lowest temperatures are found among the hourly reports for the day. The average temperature is the mean value of the 24-hour period. Calculating daily humidity follows the same process as calculating daily temperature. The weekend and holiday information can be checked using the calendar.

To calculate the weekly data set, the electricity consumption data is obtained by adding the total hourly electricity consumption for that week. The highest and lowest temperatures are found among the hourly reports for the week. The average temperature is the mean value of the 168-hour period. Calculating weekly humidity follows the same process as calculating weekly temperature. Because there are always two weekend days in a week, thus weekend information is meaningless in weekly data set. The holiday information can be checked using the calendar.

For the monthly data set, the electricity consumption data is calculated by adding the total hourly electricity consumption for that month. The data cannot be pulled from weekly reports, because a given month does not always have the same number of weeks as another month. For example, the first part of a week could belong to June and the rest of the week could belong to July. The highest and lowest temperatures are found among the hourly reports for the month. The average temperature is the mean value of the hourly records for the month. Calculating monthly humidity follows the same process as calculating monthly temperature. The weekend and holiday information can be checked using the calendar.

Following these data processing steps, the possible errors in the obtained four data sets are minimized.

4.4 Data Normalization

In machine learning area, different evaluation indicators have different measurement units and order magnitude, these indicators refer to each column in the data set talked about. The original data with original measurement units and magnitude will make it hard to get satisfactory analysis result and training performance. To reduce this effect, standardization process is necessary and important. Among many standardization methods, the normalization process is one of the most typical approaches.

In this thesis, the normalization is processing input matric by map row minimum and maximum values to [-1,1]. Using processing instruction mapminmax in MATLAB tool box, this instruction can normalize the data set row by row. The computational formula is:

$$y = (ymax-ymin)*(x-xmin)/(xmax-xmin) + ymin$$

Chapter 5 Experimental Results

5.1 Experimental environment

The prediction scheme is tested using MATLAB. MATLAB software has both a powerful computing ability and good visualization ability. The program contains a lot of toolboxes, with the neural network toolbox being one of them. In this thesis, most computation, simulation and output results were finished by MATLAB.

5.2 Experimental Settings

There are two methods for obtaining hourly prediction: with and without the weather and weekend/holiday factors in the data set. When omitting the weather and weekend/holiday factors, the inputs only contain the historical electricity consumption data.

To find how many historical data inputs should be taken into consideration to achieve the highest accuracy and lowest mean square error (MSE) for hourly prediction, several tests were conducted by considering:

- 0 hour (only considering weather and weekend/holiday information);
- 1 hour (with & without weather and weekend/holiday information);

- 2 hours (with & without weather and weekend/holiday information);
- 4 hours (with & without weather and weekend/holiday information);
- 6 hours (with & without weather and weekend/holiday information);
- 12 hours (with & without weather and weekend/holiday information);
- 24 hours (with & without weather and weekend/holiday information).

For daily prediction, several tests were conducted following a pattern similar to hourly prediction by considering:

- 0 day (only considering weather and weekend/holiday information);
- 1 day (with & without weather and weekend/holiday information);
- 3 days (with & without weather and weekend/holiday information);
- 5 days (with & without weather and weekend/holiday information);
- 7days (with & without weather and weekend/holiday information);
- 9 days (with & without weather and weekend/holiday information);
- 11 days (with & without weather and weekend/holiday information);
- 13 days (with & without weather and weekend/holiday information).

For weekly prediction, several tests were conducted by considering:

- 0 week (only considering weather and weekend/holiday information);
- 1 week (with & without weather and weekend/holiday information);

- 2 weeks (with & without g weather and weekend/holiday information);
- 3 weeks (with & without weather and weekend/holiday information);
- 4 weeks (with & without weather and weekend/holiday information);
- 5 weeks (with & without weather and weekend/holiday information).

For monthly prediction, several tests were conducted by considering:

- 0 month (only considering weather and weekend/holiday information);
- 1 month (with & without weather and weekend/holiday information);
- 2 months (with & without weather and weekend/holiday information);
- 3 months (with & without weather and weekend/holiday information);
- 4 months (with & without weather and weekend/holiday information).

For the different tests above, the number of neurons in the input layer and the hidden layer must be adjusted correspondingly.

5.3 Experimental Results

5.3.1 Hourly prediction

a. With weather and weekend/holiday factors

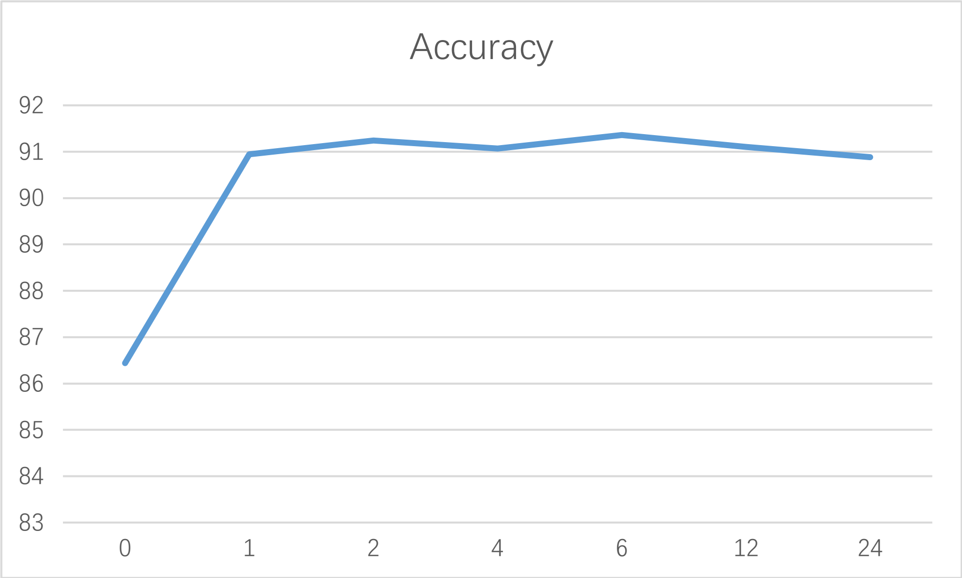

Figure 5.1 Accuracy Variation Diagram of Different Hours Input (with weather).

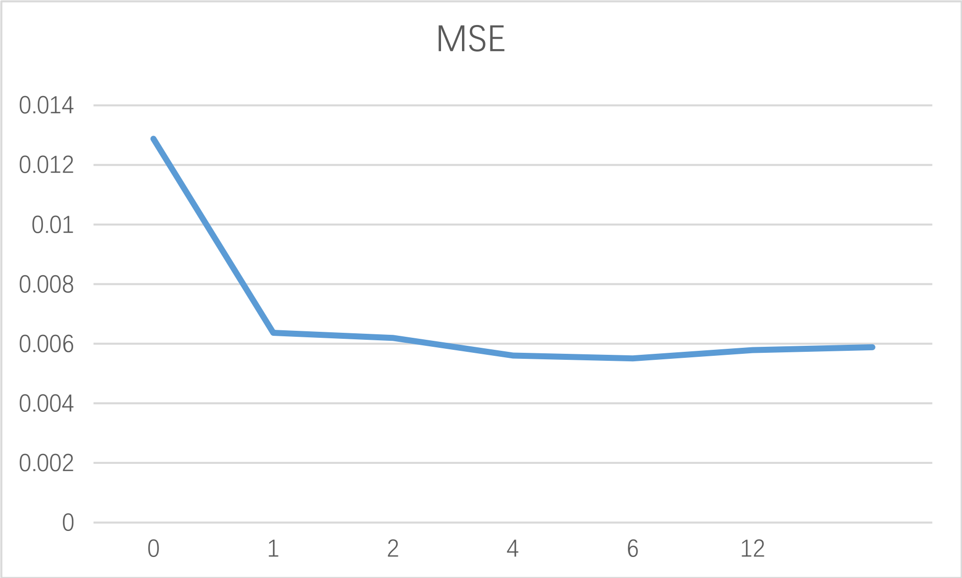

Figure 5.2 MSE Variation Diagram of Different Hours Input (with weather).

Figures 5.1 and 5.2 demonstrate that with the recorded historical electricity consumption data considered, the higher accuracy and lower MSE are obtained. Obviously, the historical electricity consumption data tells more than only considering the weather and weekend/holiday can do. However, there is not much benefit by applying longer history record.

b. Without weather and weekend/holiday factors included (date information)

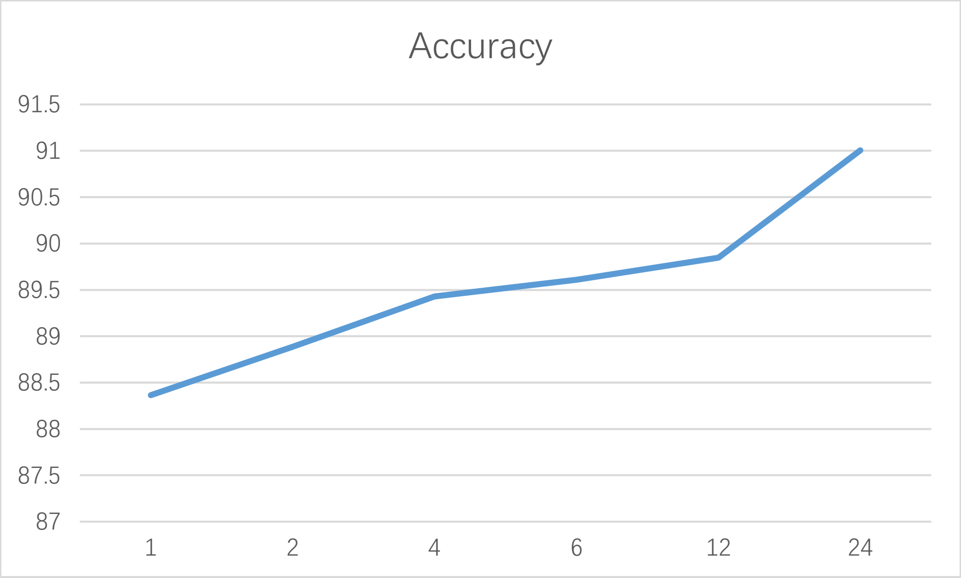

Figure 5.3 Accuracy Variation Diagram of Different Hours Input (without weather).

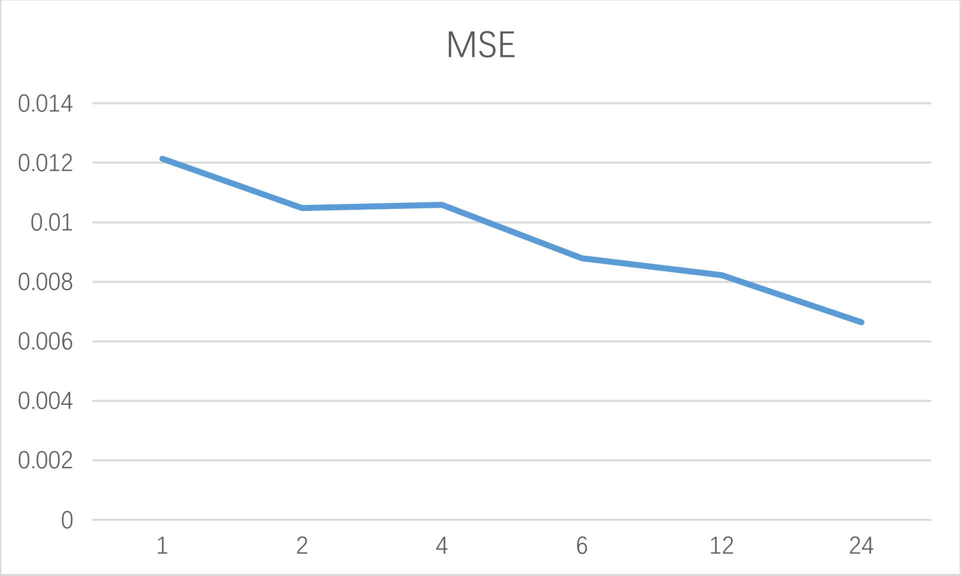

Figure 5.4 MSE Variation Diagram of Different Hours Input (without weather).

Figures 5.3 and 5.4 demonstrate that with more historical electricity consumption records, higher accuracy and lower MSE are obtained. It is really interesting that when the weather information and weekend/holiday factors are not taken into account, the historical electricity consumption record plays a more significant role and the longer a history is counted, the better performance is achieved. The trend does not stop even 24 hours history has been counted.

c. Comparison between cases a) and b)

Figures 5.5 and 5.6 demonstrate that for hourly prediction, considering weather and weekend/holiday factors has higher accuracy and smaller MSE than omitting these factors. While in general, the history information is useful to improve the prediction accuracy, much less historical records is needed to achieve the decent level when the weather and weekend/holiday factors are available. This actually implies lower computing and transmission overhead.

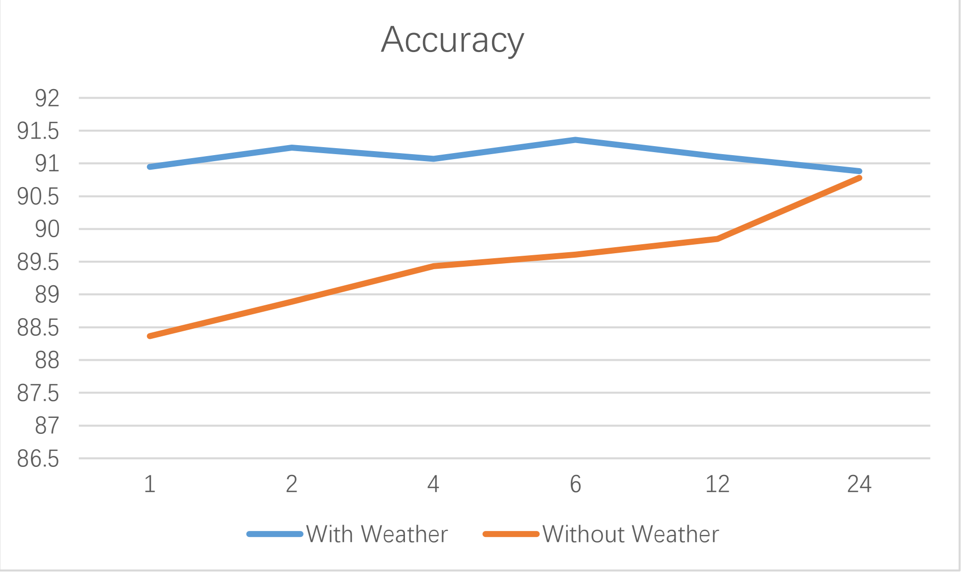

Figure 5.5 Comparison of Two Types Hourly Predictions' Accuracy Variation.

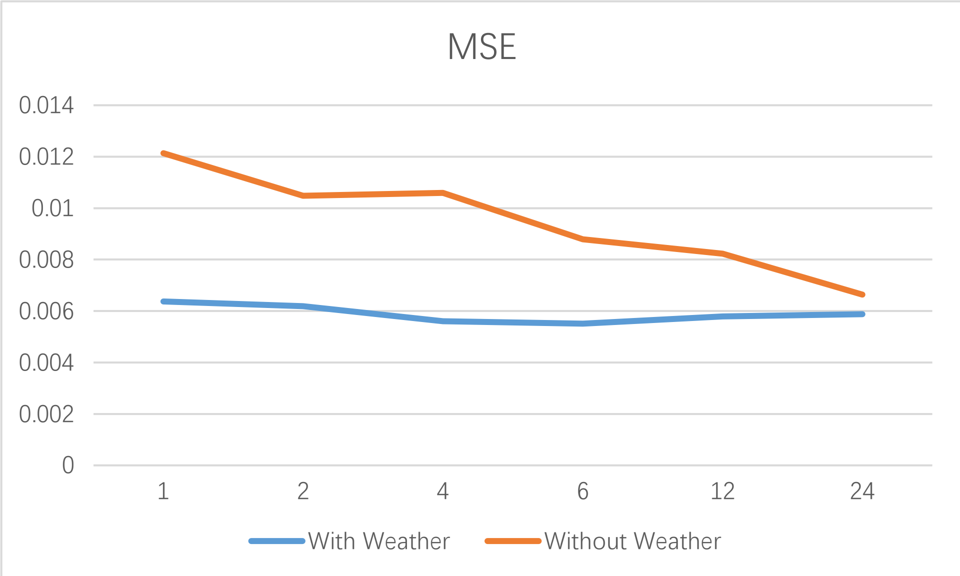

Figure 5.6 Comparison of Two Types Hourly Predictions' MSE Variation.

5.3.2 Daily Prediction

a. With weather and weekend/holiday factors included (date information)

Figures 5.7 and 5.8 demonstrate that there is not a clear relationship between prediction performance and the amount of historical electricity consumption data applied. However, it may be an experience that using seven days of historical record has the highest accuracy with a fair MSE.

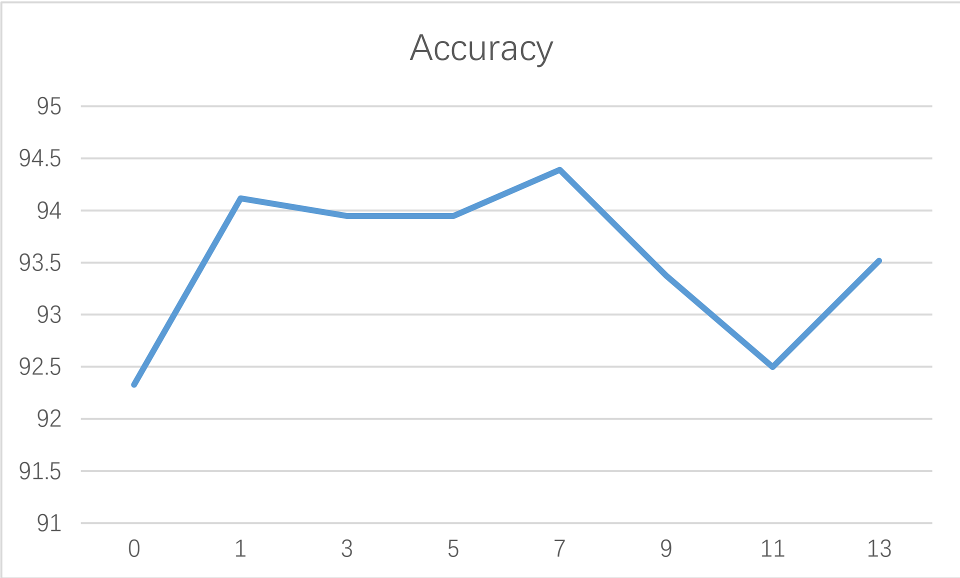

Figure 5.7 Accuracy Variation Diagram of Different Days Input (with weather).

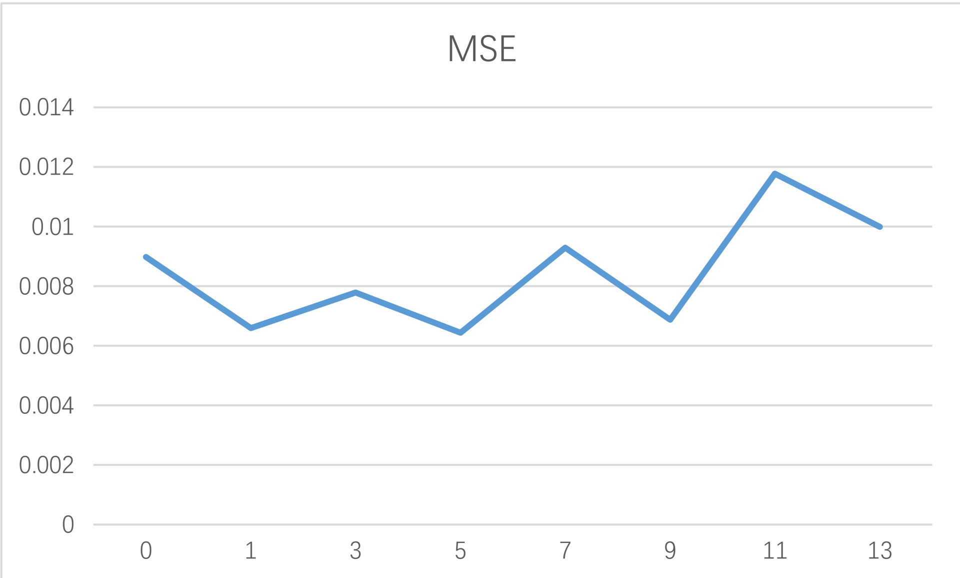

Figure 5.8 MSE Variation Diagram of Different Days Input (with weather).

b. Without weather and weekend/holiday factors included (date information)

When the weather and weekend/holiday factors are not considered, the results shown by Fig. 5.9 and Fig. 5.10 are really interesting. It looks the accuracy becomes worse when longer history record is applied, and the MSE is not related to the length of history records.

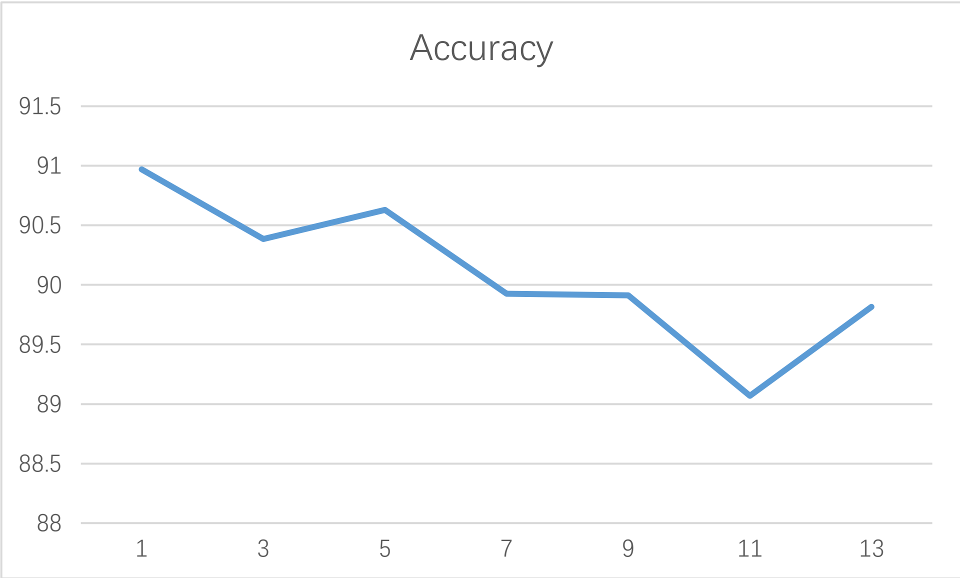

Figure 5.9 Accuracy Variation Diagram of Different Days Input (without weather).

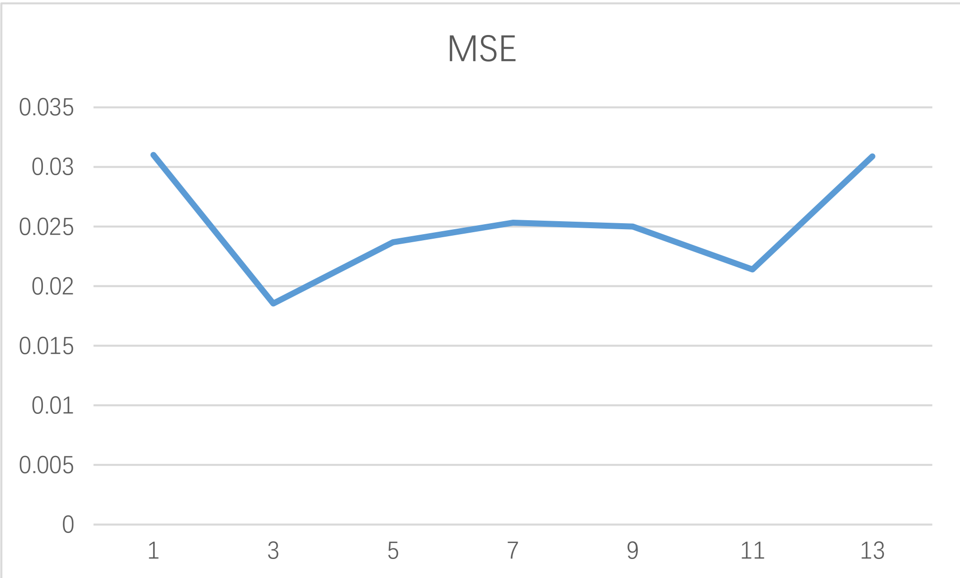

Figure 5.10 MSE Variation Diagram of Different Days Input (without weather).

c. Comparison between a) and b)

Putting the results from cases a) and b) together, Fig. 5.11 and Fig. 5.12 demonstrate that, for daily prediction, considering weather and weekend/holiday factors yields much higher accuracy and lower MSE. Meanwhile, the historical record does not help the prediction.

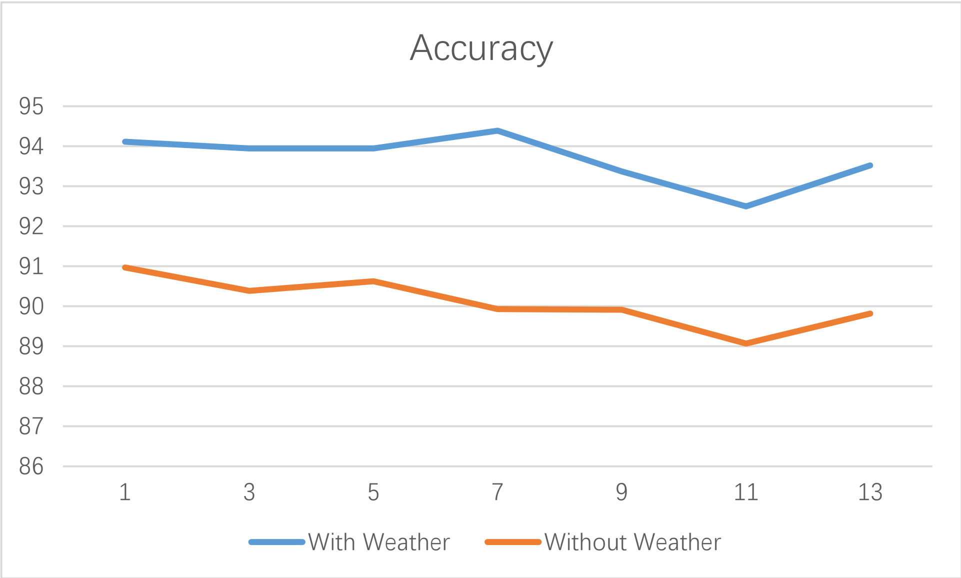

Figure 5.11 Comparison of Two Types Daily Predictions' Accuracy Variation.

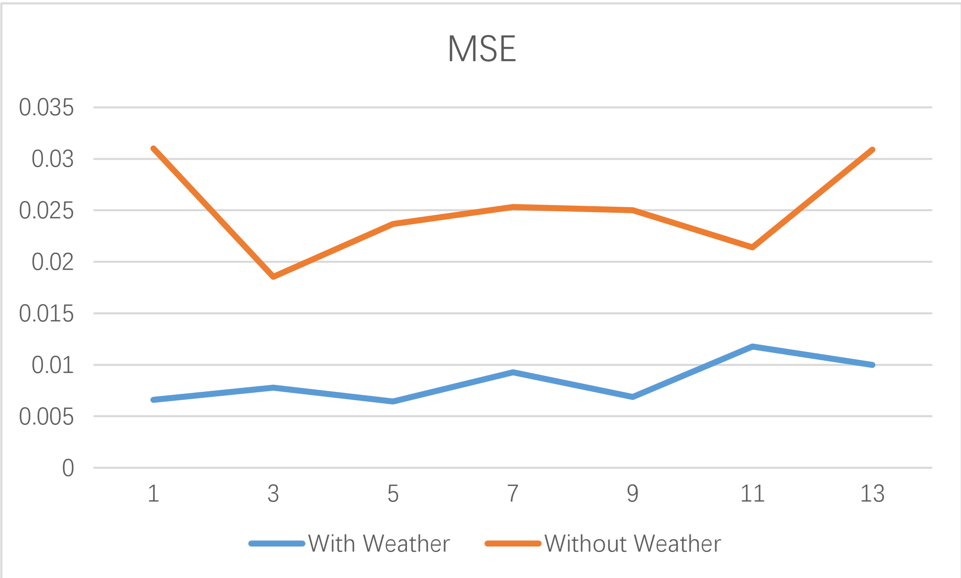

Figure 5.12 Comparison of Two Types Daily Predictions' MSE Variation.

5.3.3 Weekly prediction

a. With weather and weekend/holiday factors included (date information)

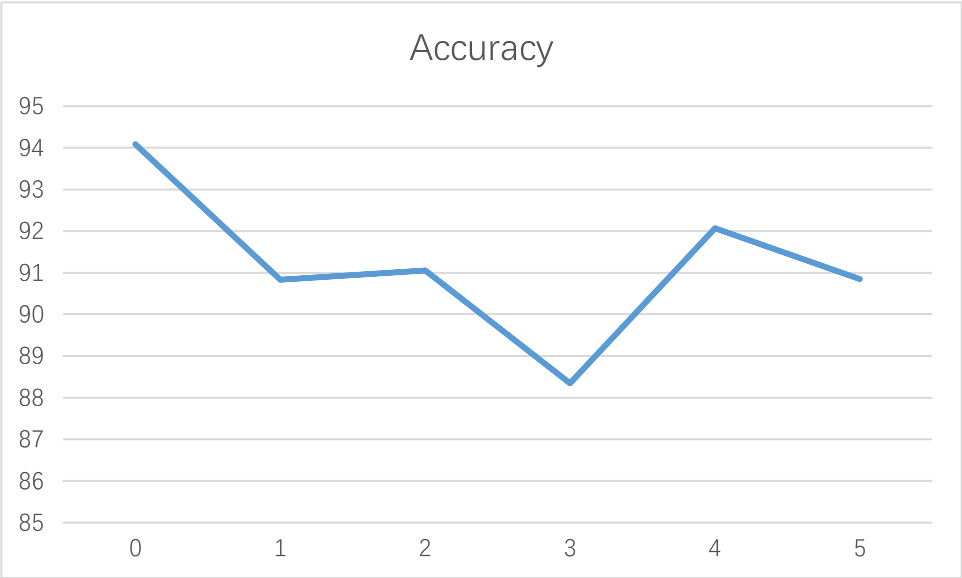

Figure 5.13 Accuracy Variation Diagram of Different Weeks Input (with weather).

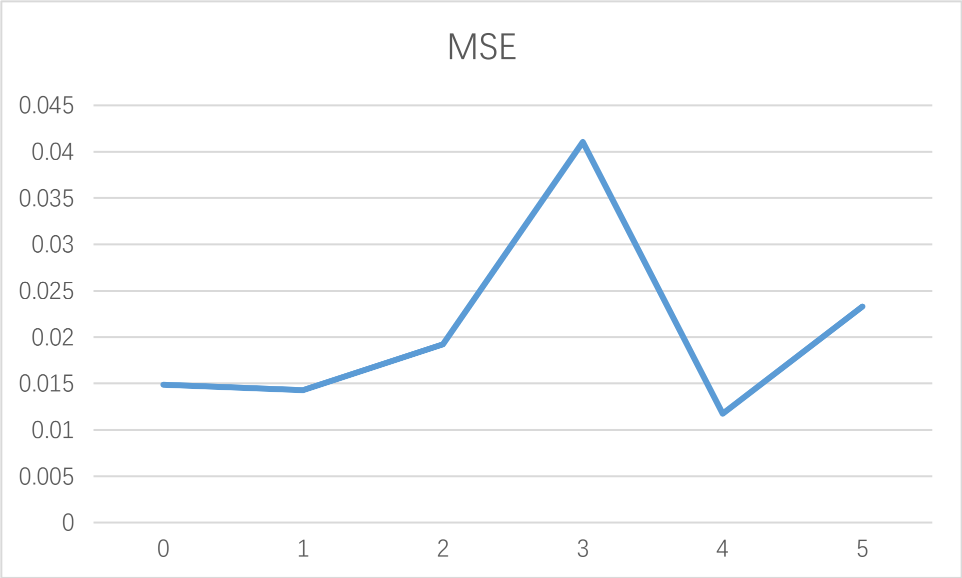

Figure 5.14 MSE Variation Diagram of Different Weeks Input (with weather).

Figure 5.13 and Figure 5.14 demonstrate that, for weekly prediction, when the weather and weekend/holiday factors are considered, including more weekly historical electricity consumption data actually decreases the prediction accuracy. When we consider the historical data input for three weeks, the performance is the worst.

b. Without weather and weekend/holiday factors included (date information)

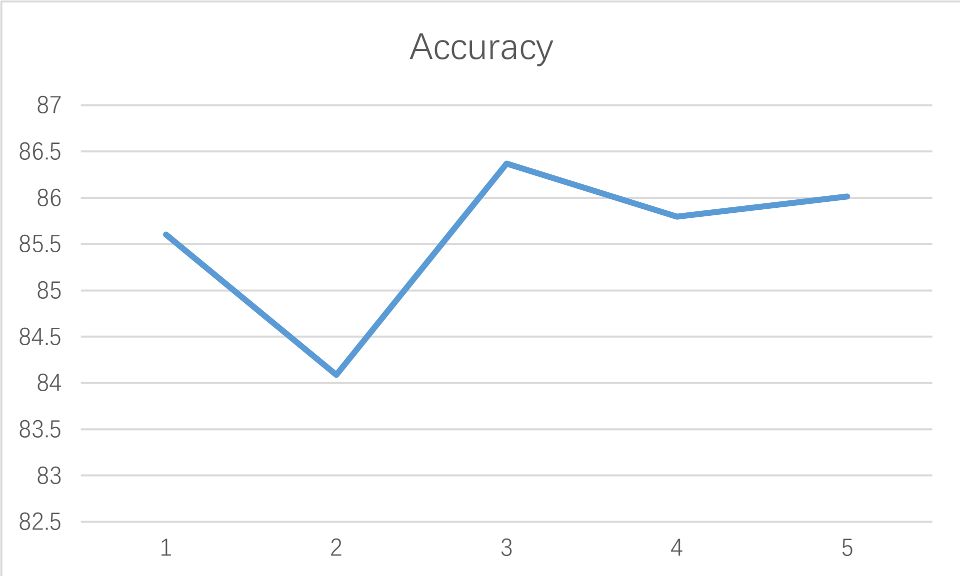

Figure 5.15 Accuracy Variation Diagram of Different Weeks Input (without weather).

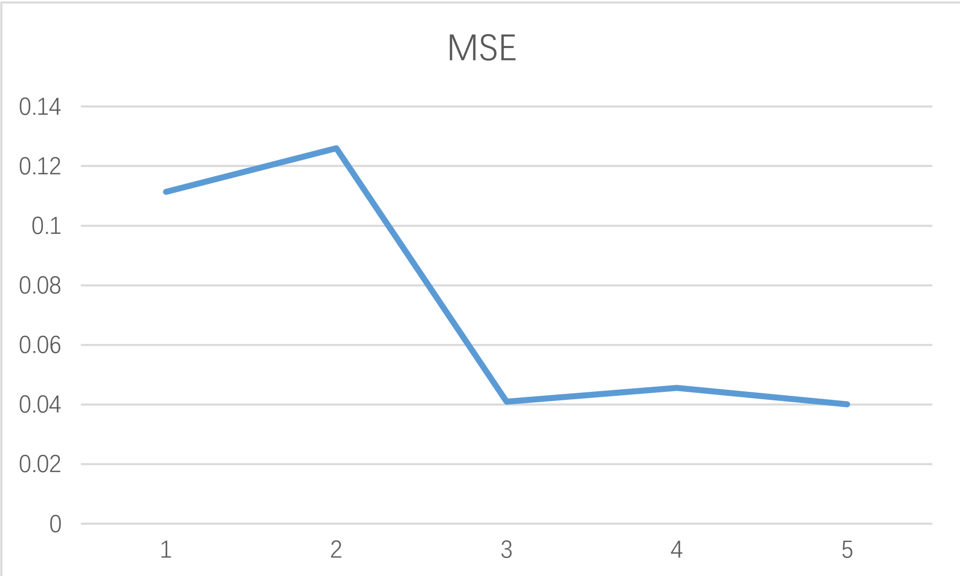

Figure 5.16 MSE Variation Diagram of Different Weeks Input (without weather).

As shown by Fig. 5.15 and Fig. 5.16, for weekly prediction, when the weather and weekend/holiday factors are not considered, more weekly historical electricity consumption data does not always contribute to the predictions. Using three weeks historical data got the best prediction with the highest accuracy and the lowest MSE.

But two weeks of the historical data leads to the worst result.

c. Comparison between a) and b)

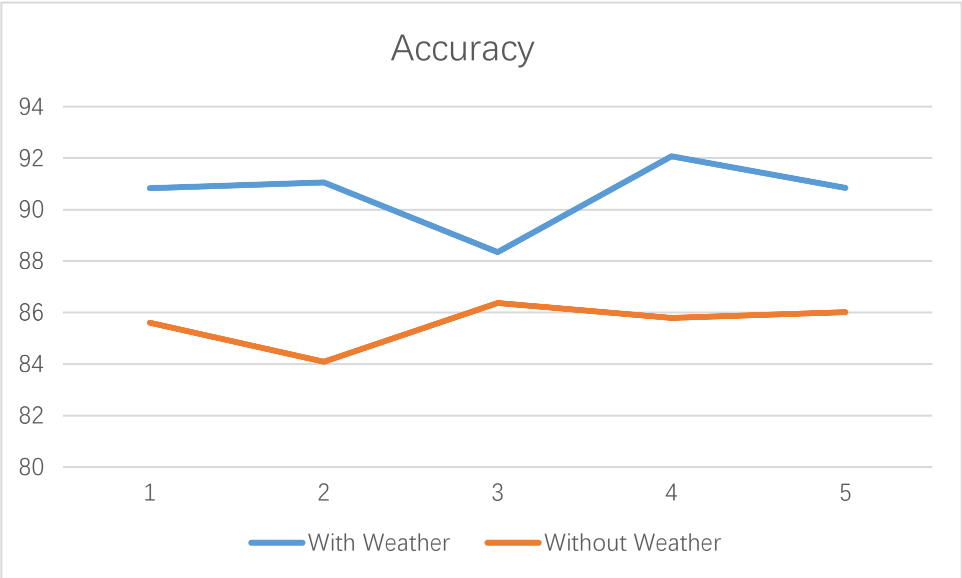

Figure 5.17 Comparison of Two Types of Weekly Predictions' Accuracy Variation.

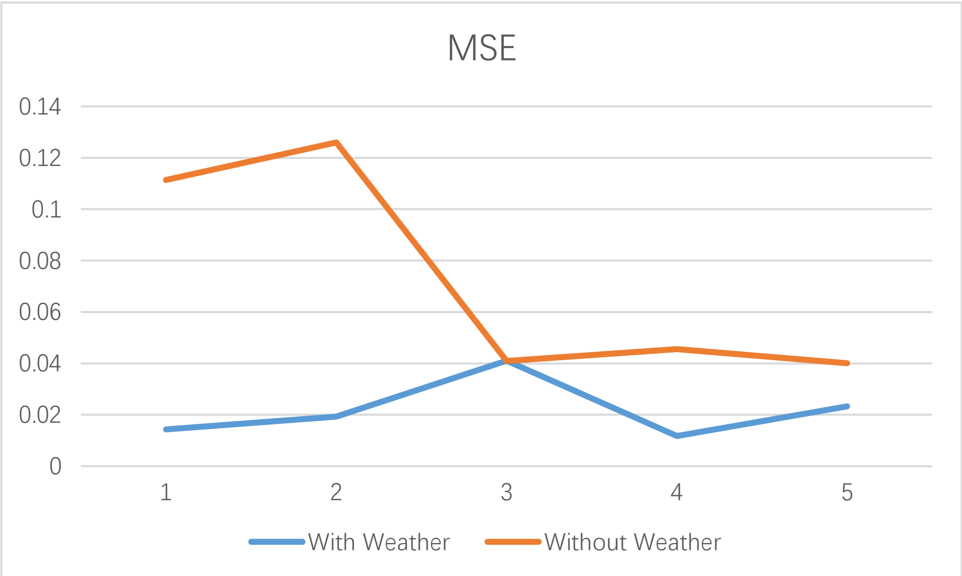

Figure 5.18 Comparison of Two Types of Weekly Predictions' MSE Variation.

Figure 5.17 and Figure 5.18 demonstrate that, for weekly prediction, including weather information and weekend/holiday factors yields more accurate prediction performance than omitting these factors. However, the amount of historical record does not bring much difference.

5.3.4 Monthly prediction

a. With weather and weekend/holiday factors included (date information)

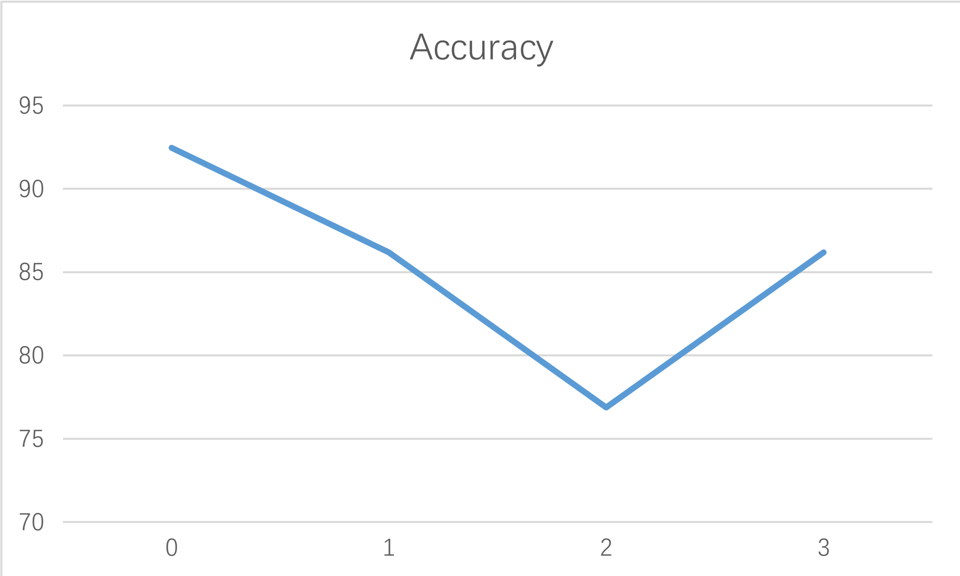

Figure 5.19 Accuracy Variation Diagram of Different Months Input (with weather).

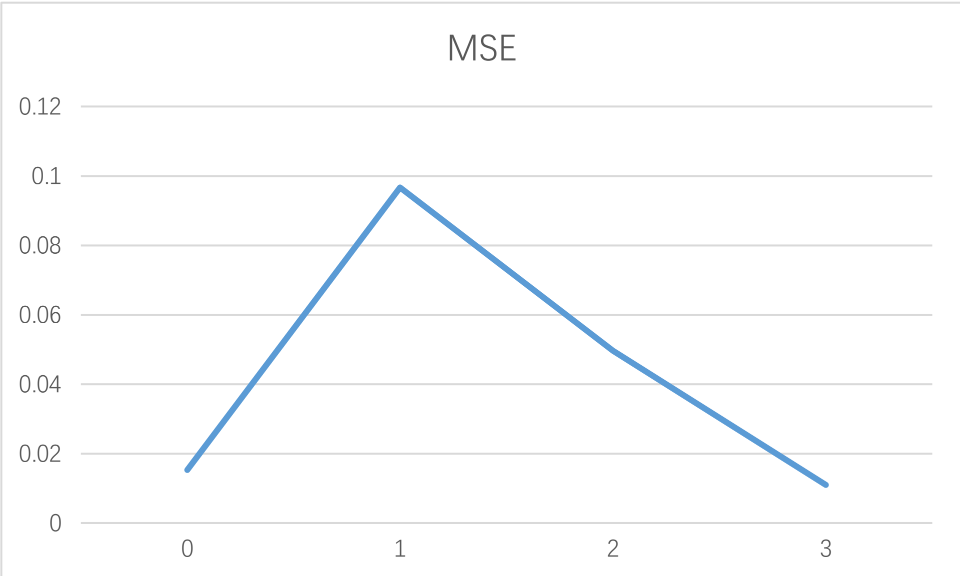

Figure 5.20 MSE Variation Diagram of Different Months Input (with weather).

The curves shown in Fig. 5.19 and Fig. 5.20 demonstrate the change of prediction performance for monthly prediction with weather information and weekend/holiday factors included. The accuracy decreases when more monthly historical electricity consumption data is used. When two months of historical data is used, the accuracy is the lowest. The performance recovers after that, however, due to the limited data set, this work could not try more.

b. Without weather and weekend/holiday factors included (date information)

It is very interesting as shown by Fig. 5.21 and Fig. 5.22, for monthly prediction, when the weather information and weekend/holiday factors are not considered, the prediction performance will improve as more historical electricity consumption data inputs are added.

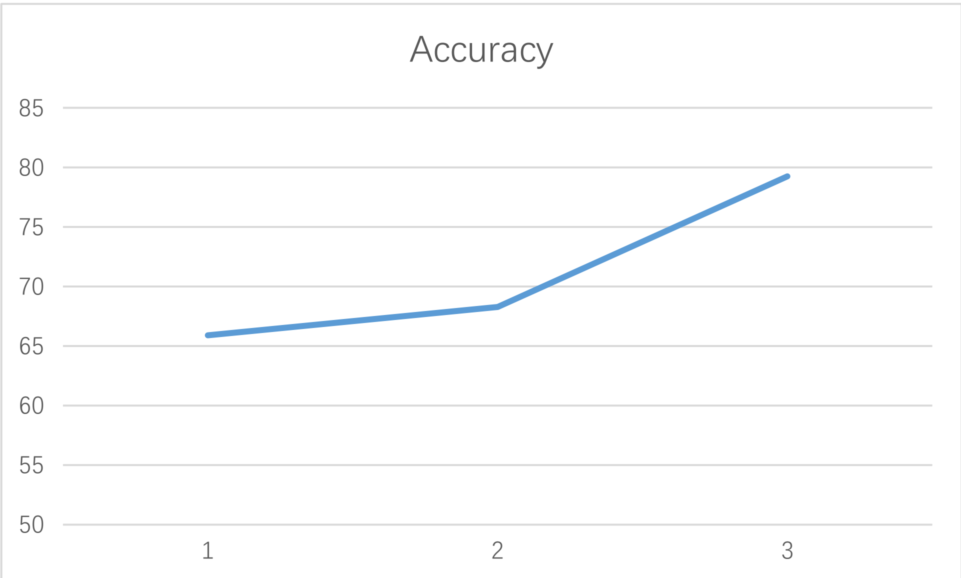

Figure 5.21 Accuracy Variation Diagram of Different Months Input (without weather).

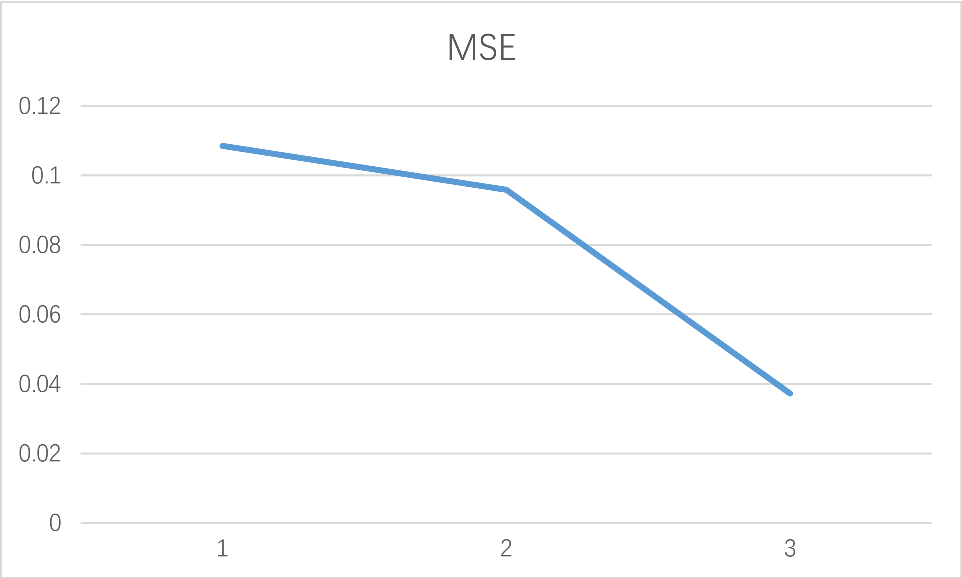

Figure 5.22 MSE Variation Diagram of Different Months Input (without weather).

c. Comparison between a) and b)

Putting the performance curves of these two scenarios together, Fig. 5.23 and Fig. 5.24 verified that for monthly prediction, including the weather information and weekend/holiday factors yields more accurate prediction performance than omitting these factors. But again, the influences of the length of historical data is not significant.

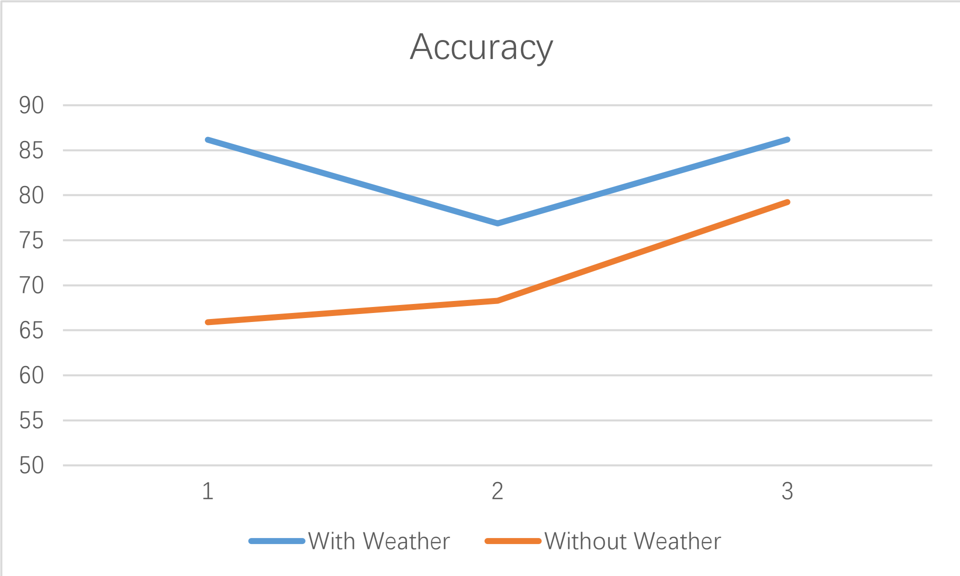

Figure 5.23 Comparison of Two Types of Monthly Predictions' Accuracy Variation.

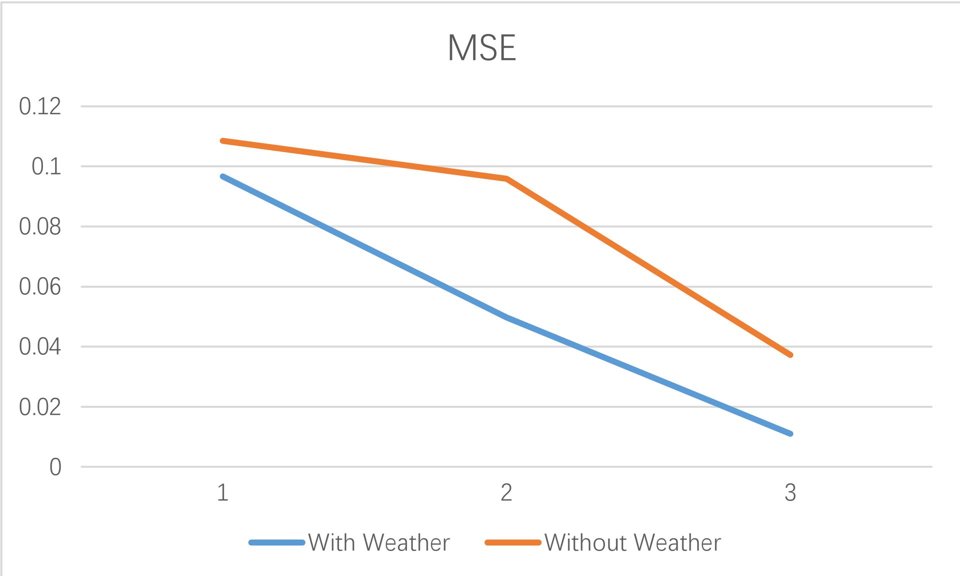

Figure 5.24 Comparison of Two Types of Monthly Predictions' MSE Variation.

5.3.5 Discussions

a. Influence of weather and weekend/holiday factors

As the experimental results presented above, it is clear that taking the weather and weekend/holiday factors into account yield better prediction performance than omitting these factors. However, it is not clear how much historical information should be leveraged to achieve the optimal prediction result.

Because of the randomness of computer training and the drawback of the BPNN, the initial weight and bias are generated randomly. The computer training results can easily run into the local optimization solution but not the global optimization solution. For each experiment, the program run ten times and the average value is adopted as the final result.

Figures 5.25 to 5.28 show the best prediction performance in four time scales with the weather condition factor considered.

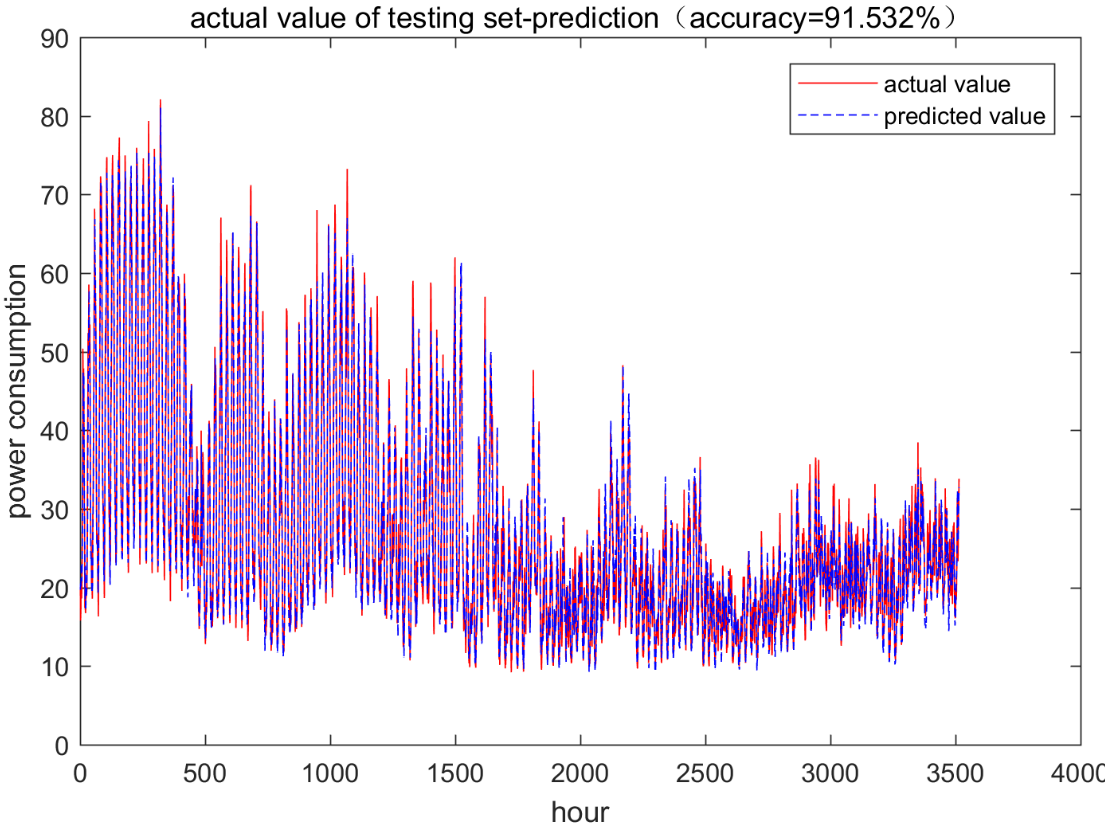

Figure 5.25 Best Performance of Hourly Prediction.

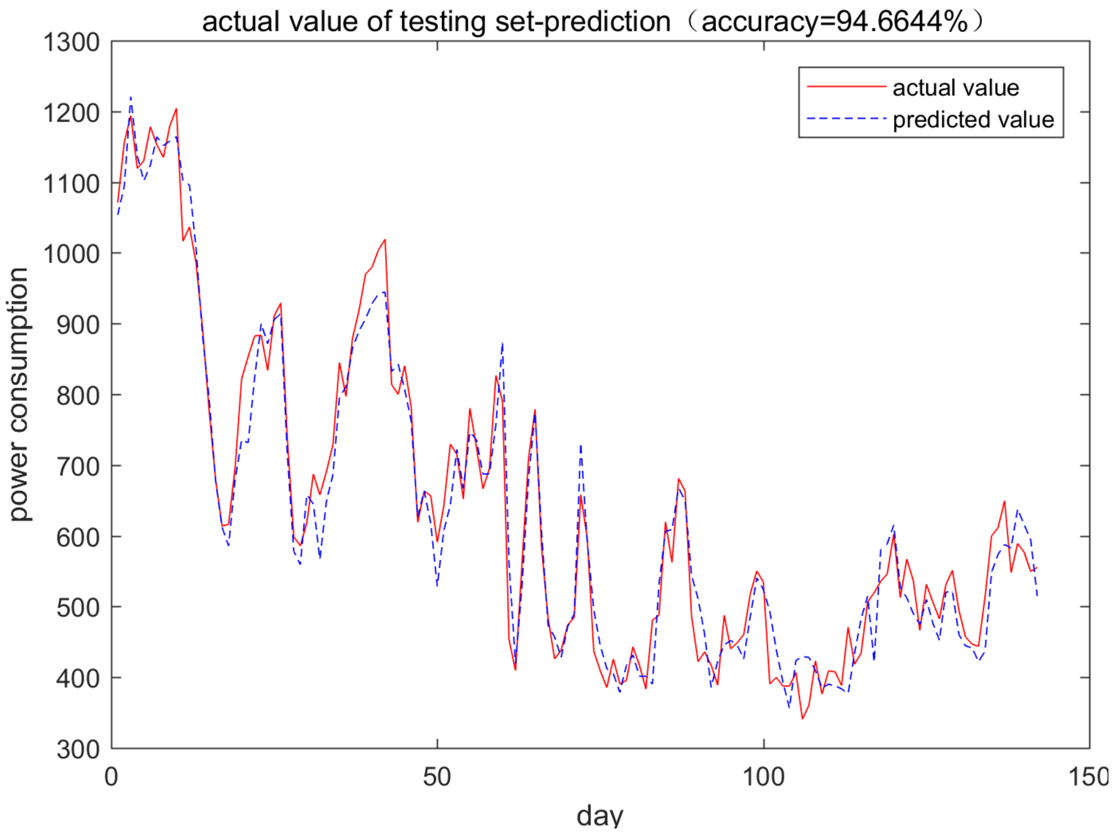

Figure 5.26 Best Performance of Daily Prediction.

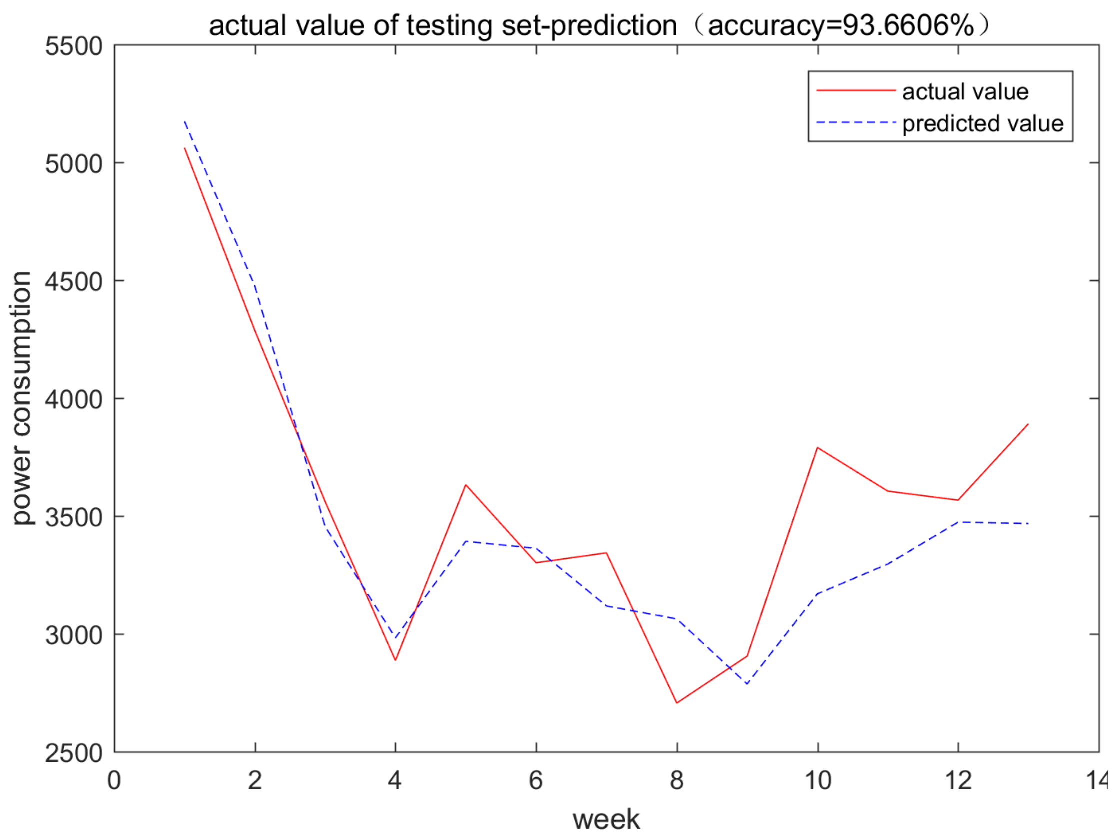

Figure 5.27 Best Performance of Weekly Prediction.

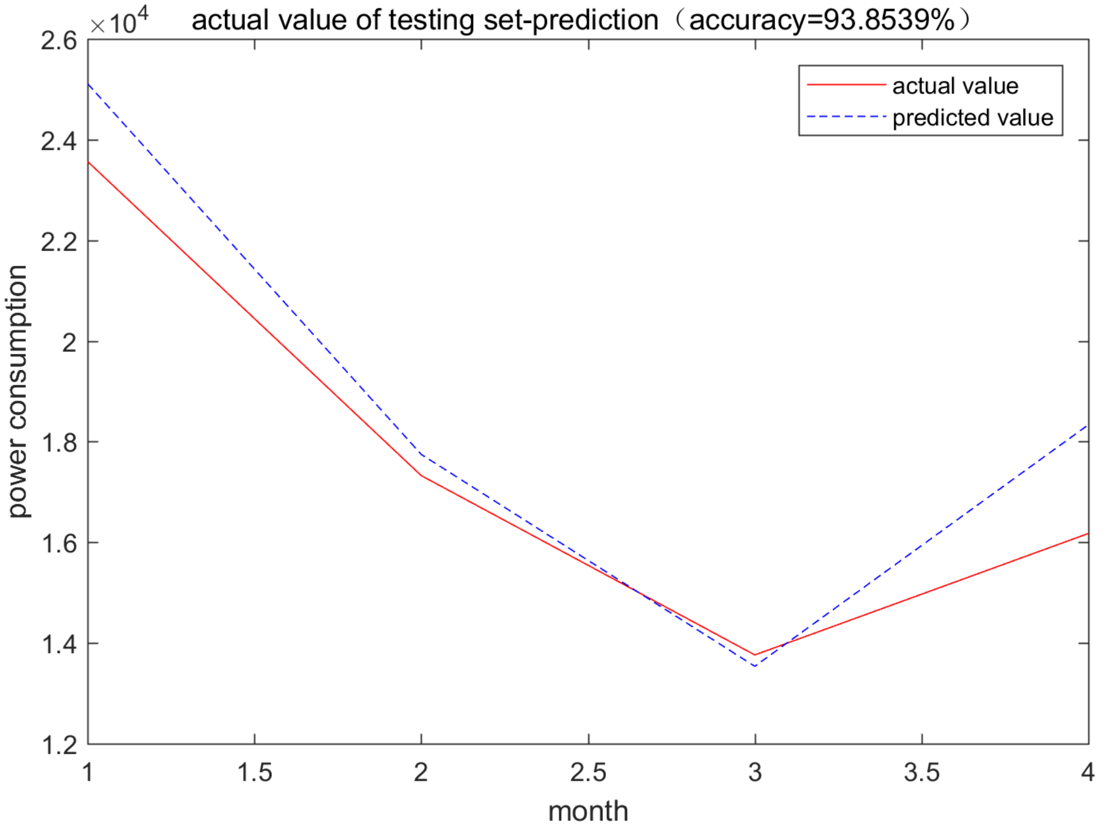

Figure 5.28 Best Performance of Monthly Prediction.

As shown in Figs. 5.25, 5.26, 5.27 and 5.28, among the four prediction time resolutions, the hourly prediciton and the daily prediction have achieved better fit than the weekly and the monthly predictions. However, the average accuracy of the monthly and the weekly predictions were slightly higher than the hourly prediction. The average MSE of the hourly and the daily predictions are lower than the weekly and the monthly predictions. However, the prediction performance of the monthly and the weekly predictions are unstable. For example, the accuracy of prediction varies from 69% to 93%. The performance of the daily and the hourly predictions is much more stable with the variation below 2%.

It is not a surprise that this researh confirms the data size is the dominant factor that brings impact to the variation in accuracy. The hourly prediction and the daily prediction have much larger training and testing data sets. Under the repeated experiments, the chance factor and local optimum are removed. Theoretically, the larger the data size, the better the prediction performance.

Sometimes, there are several data points that deviate from the norm range. But their influences can be mitigated when a sufficiently large data set is available. Meanwhile, if the data set is limited, the abnormal data points will make the prediction result fluctuate wildly, even make the misprediction where the predicted result has the opppsite changing trend with actual value.

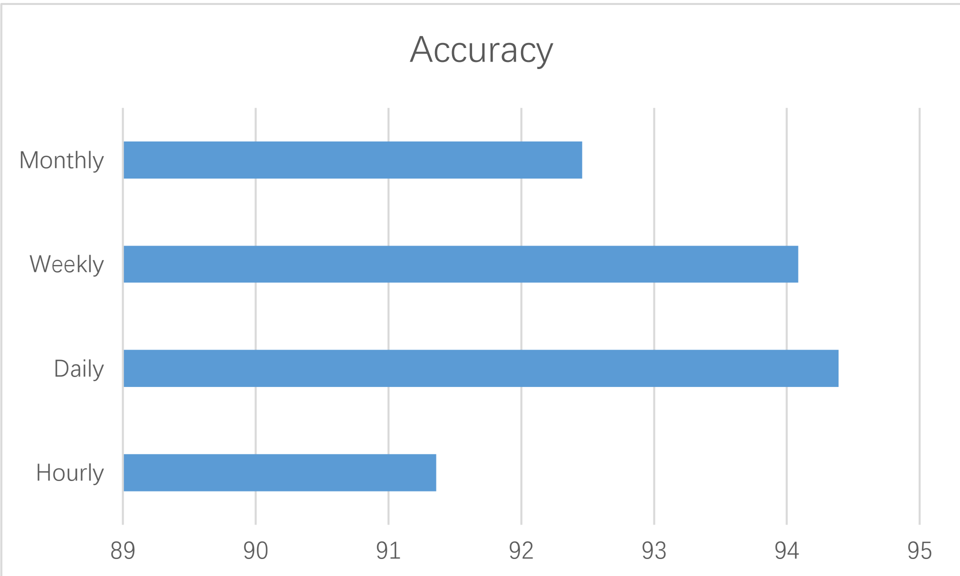

Figure 5.29 The accuracy of predictions on four time scales with weather factors considered.

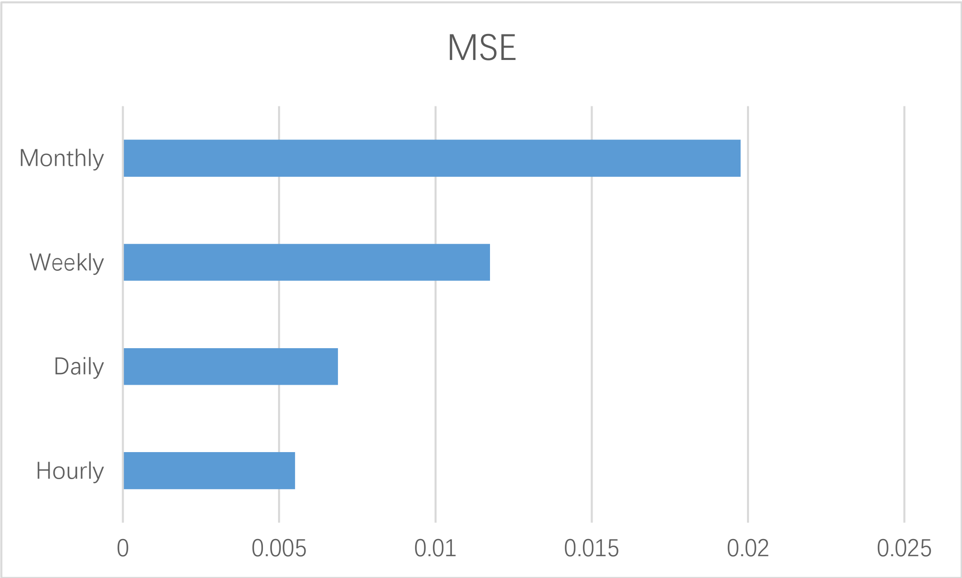

Figure 5.30 The MSE of predictions on four time scales with weather factors considered.

Without weather condition, best performance of four kinds of predictions

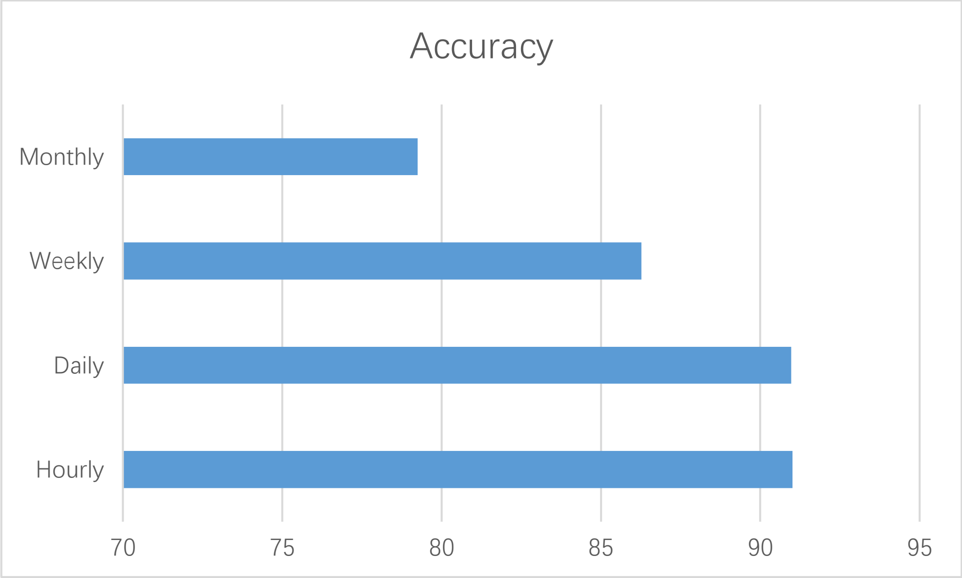

Figure 5.31 The Accuracy of predictions on four time scales without weather factors considered.

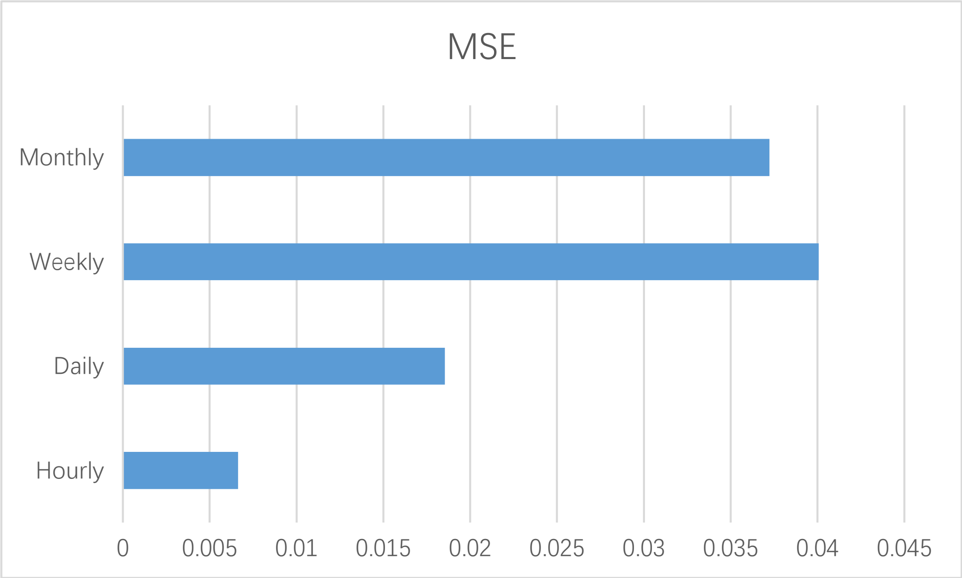

Figure 5.32 The MSE of predictions on four time scales without weather factors considered.

b. Influence of each individual factors

In this work, the impacts of each individual factors are studied, in all four prediction time scales. Figures 5.33, 5.34, 5.35, and 5.36 present the drop of prediction accuracy when one of the factors is ignored.

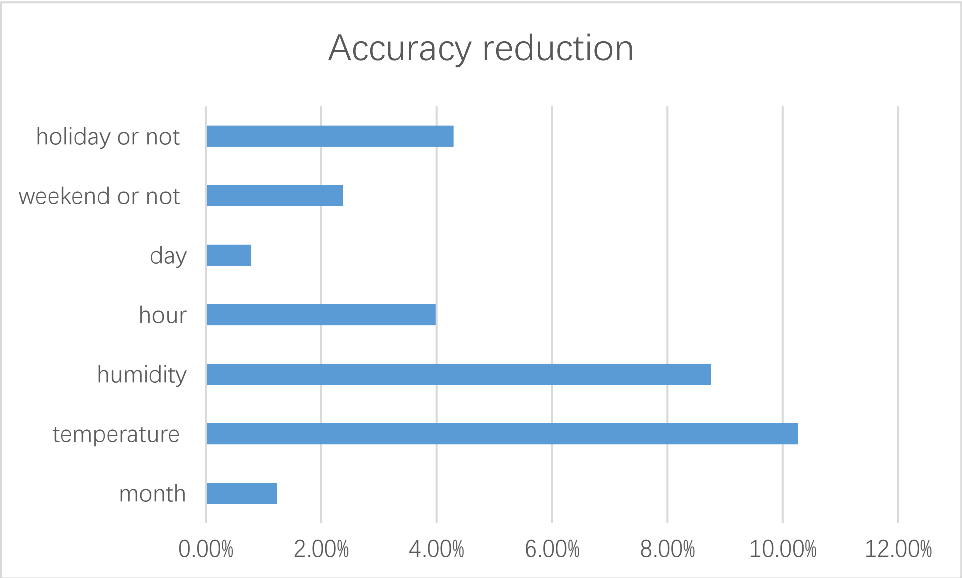

Figure 5.33 The influences of each individual factors on the prediction accuracy: hourly prediction.

Figure 5.33 illustrates the impacts on prediction accuracy for hourly prediction when one of the factors is ignored. The temperature and humidity have the largest influence on the prediction performance. However, omitting a given day in a week or a month in a year does not introduce a significant effect in prediction performance.

For daily prediction, the impacts of each factor are illustrated in Fig. 5.34. The highest and the lowest temperatures have the largest effect on prediction performance. However, the average humidity has the least effect on prediction performance.

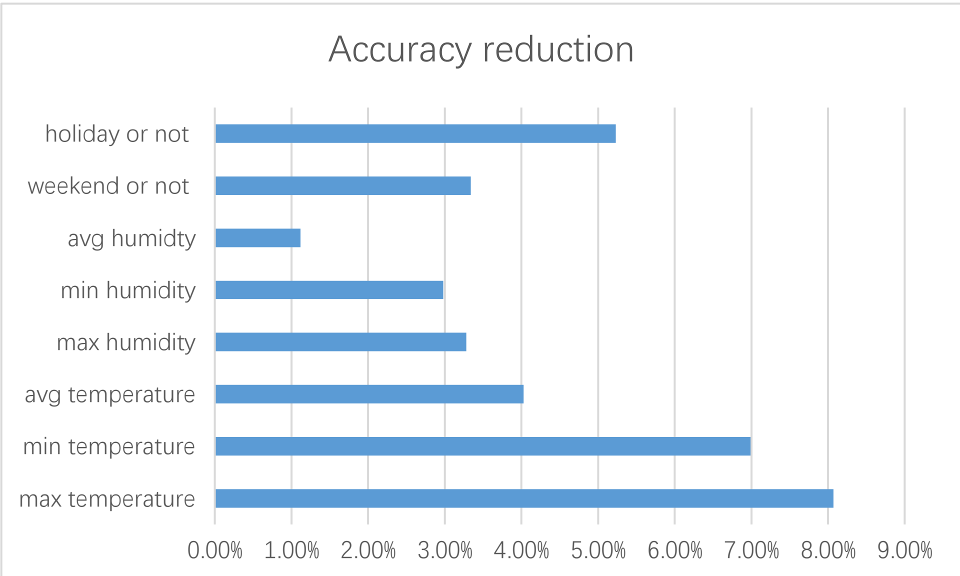

Figure 5.34 The influences of each individual factors on the prediction accuracy: daily prediction.

Figure 5.35 shows the case of the weekly prediction. The maximum and the minimum temperatures have the most significant influence on the prediction performance. However, the maximum humidity has the least effect on prediction performance. It is because the maximum humidity maintains above 90% for many weeks at a stable level, it did not greatly affect prediction accuracy.

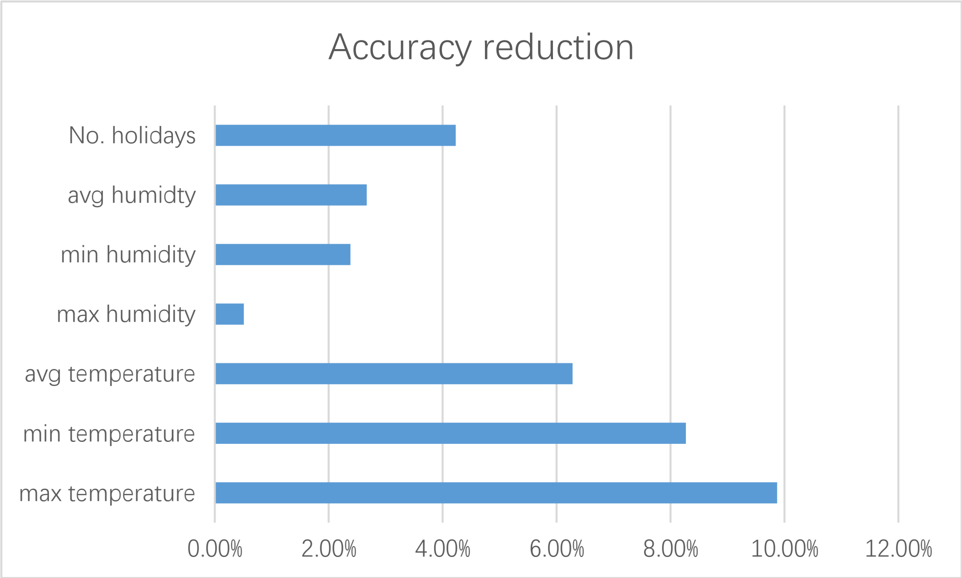

Figure 5.35. The influences of each individual factors on the prediction accuracy: weekly prediction.

For monthly prediction, delete one of the input factors, the accuracy drops down figure. Because nearly all the months had a max humidity of 99%, max humidity did not affect prediction accuracy.

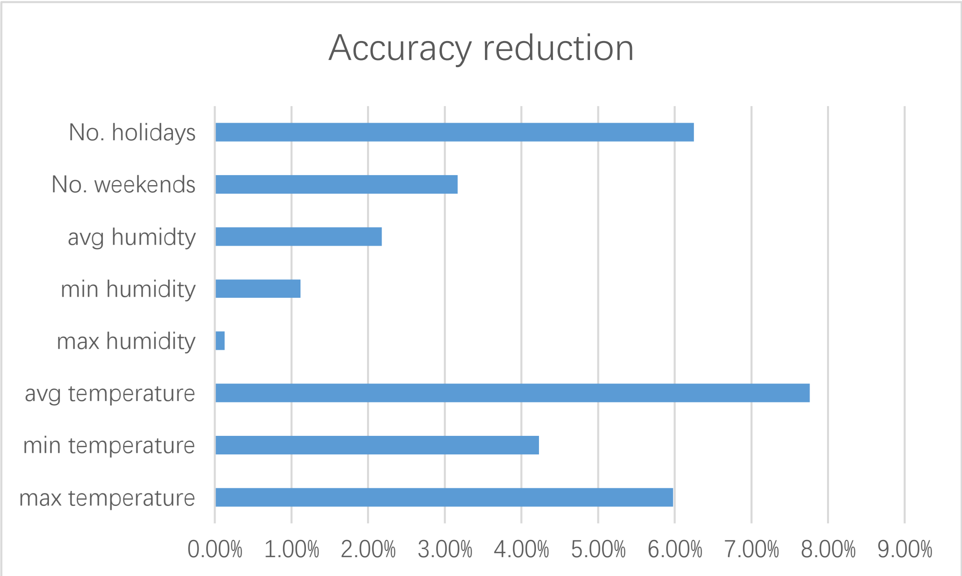

Figure 5.36 The influences of each individual factors on the prediction accuracy: monthly prediction.

c. Influence of BPNN hidden layer design

To find the most appropriate number of nodes in the hidden layer, this thesis validates multiple empirical formulas because there is not a universal formula for prediction models. Therefore, the method of trial and error is utilized. The test case is the daily prediction with the weather and weekend/holiday factors considered, using seven days' historical electricity consumption history. Because this model has relatively stable performance. Here the number of input nodes is 15 and the number of output nodes is 1.

For the four different empirical formulas, the number of nodes in the hidden layer should be as follows:

1. $\sum_{i=0}^{n} C_m^i > k$, in this case, because the number of samples is larger than 700, the number of nodes in the hidden layer should be larger than 10.
2. $n_1 = \sqrt{n+m} + a$, the number of nodes in the hidden layer should be around 5 to 14.
3. $n_1 = \log_2 n$, the number of nodes in the hidden layer should be around 4.
4. $n_1=2n+1$, the number of nodes in the hidden layer should be around 31.

The experimental result is shown in Figs. 5.37 and 5.38. When the number of nodes in the hidden layer increases, the prediction accuracy appears a constant flux wave, but follows a downward trend. In terms of MSE, the best prediction performance occurs when the number of nodes in the hidden layer is 15 or 16.

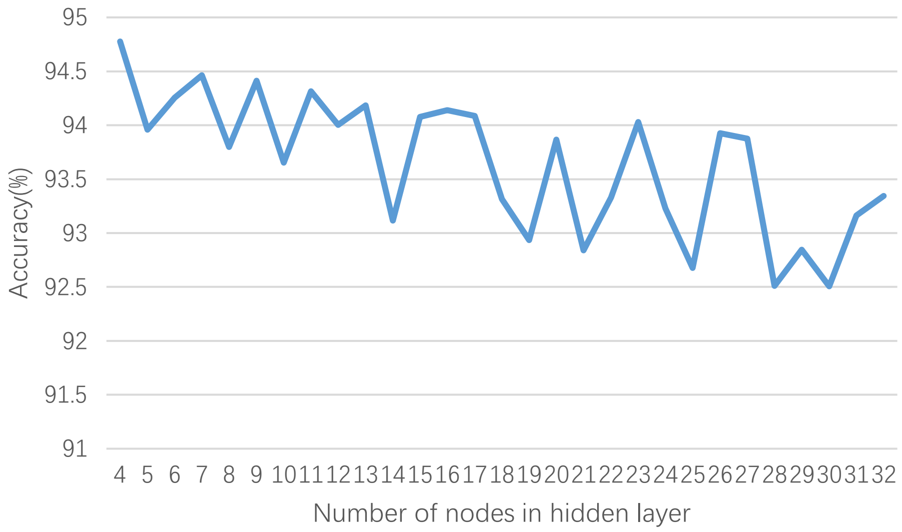

Figure 5.37 Prediction Accuracy with Different Number of Nodes in Hidden Layer.

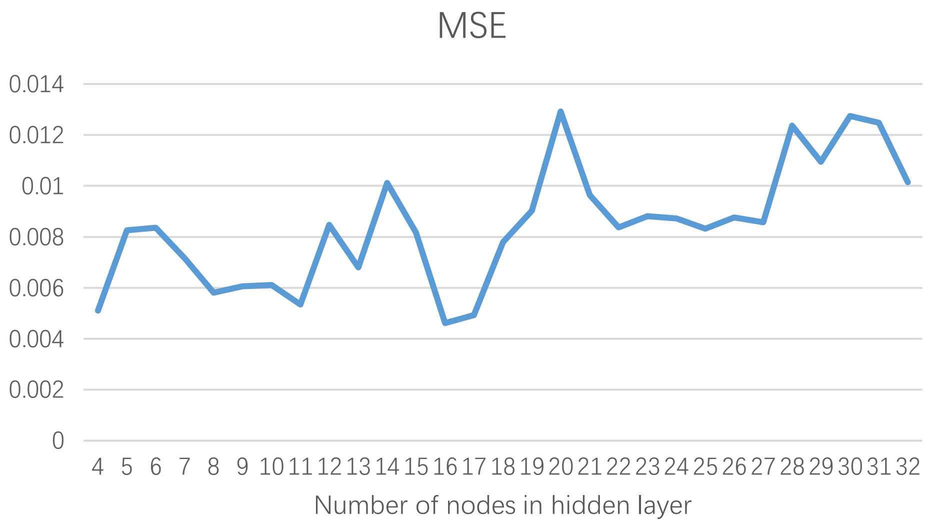

Figure 5.38 Prediction MSE n with Different Number of Nodes in Hidden Layer.

According to the experimental results, these four empirical formulas are not accurate. They are concluded based on some earlier experiences. But for different BPNN models, the number of nodes in the hidden layer has to be tried many different values based on empirical formula, there is not a clear rule.

Chapter 6 Conclusions and Discussions

This thesis has explored to apply BPNN in electricity consumption prediction. Leveraging the data set from the Pecan Street Project, the influences of different factor are experimentally investigated. According to the experimental results, it is clear that including weather and weekend/holiday factors will increase the accuracy and reduce the MSE for all four prediction scales: hourly, daily, weekly, and monthly.

For different prediction scales, the effects of including weather and weekend/holiday factors are different. For example, the maximum humidity has an insignificant effect on weekly and monthly predictions than for hourly and daily predictions.

Hourly and daily predictions have relatively better performance than weekly and monthly predictions. One of the most important reasons could be the size of the data set. Theoretically, the more data we train the computer to process, the higher the accuracy. Therefore, the experiments will be improved.

Regarding the BPNN, there are many empirical formulas or commonly used values, such as learning rate, epoch times, proportion of different data set (training, testing and validation), the number of nodes in hidden layer and so on. However, when conducting real experiments, we have to repeat experiments to find the best value because there are not well-defined guidelines for choosing the parameters.

Although this thesis has made relatively satisfactory prediction performance for hourly and daily prediction, there are multiple insufficiencies, which can be solved through future research.

The size of the data set is not large enough to conduct a decent weekly and monthly prediction, which causes fluctuations in prediction results. It is expected that better performance can be achieved if more data sets are available. If the data set is large enough, long-term prediction can be explored, such as quarterly prediction or even annual prediction.

If more data sets are acquired, the hourly and daily predictions will have numerous inputs samples. Due to the drawback of the BP Neural Network, the computer training time will be quite long. It will be easily trapped into the local optimum value but not the global optimum value. This case cannot be solved easily by repeating the experiments more times. Some optimization methods can be introduced to speed up the training process and avoid getting into the local optimum solution, such as PSO (particle swarm optimization), GA (genetical algorithm) and so on.

Because of the limitation of data set, in this thesis, there is not an obvious effect of GA optimized BPNN, the improvement of training speed and prediction accuracy is little.

Compared with other methods, the traditional BPNN is suitable for short-term, multifactor electricity consumption prediction of the Pecan Street Project. Before determining which method to use, many research works have been done, and some methods have been tried.

The grey model is aimed at information poor and data incomplete cases, although it has following advantages: light computation work, easily modeling, finding unobvious relationship and information from unregular raw data, relatively high accuracy. Its advantages cannot be fully used in this high resolution, rich data type's prediction case.

The regression analysis is a simple model but has good extrapolation. But it is mainly used in single factor modeling, such as only considering historical consumption data. It has following disadvantages: not very fit for complex forecasting case, cannot include multifactor effects.

The BPNN is a mature algorithm which has a long history. There are new neural network models that avoid the drawback of the BPNN, like the LSTM (long short-term memory) Neural Network, which is suitable for time series. In future work, these two kinds of neural networks can be compared in various aspects, including learning speed, accuracy, MSE and so on.

References

[1] K. Muralitharan, R. Sakthivel and R. Vishnuvarthan, "Neural network based optimization approach for energy demand prediction in smart grid", 2018.

[2] U. Obinna, P. Joore, L. Wauben and A. Reinders, "Comparison of two residential Smart Grid pilots in the Netherlands and in the USA, focusing on energy performance and user experiences", 2018.

[3] Advances in computer science and information engineering. Berlin: Springer, 2012.

[4] E. Almeshaiei and H. Soltan, "A methodology for Electric Power Load Forecasting", 2018.

[5] Al-Hamadi, H.M. and Soliman, S.A., 2005. Long-term/mid-term electric load forecasting based on short-term correlation and annual growth. Electric power systems research, 74(3), pp.353-361.

[6] Samarasinghe, S., 2016. Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. CRC Press.

[7] Hahn, H., Meyer-Nieberg, S. and Pickl, S., 2009. Electric load forecasting methods: Tools for decision making. European journal of operational research, 199(3), pp.902-907.

[8] Hong, W.C., 2011. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. Energy, 36(9), pp.5568-5578.

[9] Park, D.C., El-Sharkawi, M.A., Marks, R.J., Atlas, L.E. and Damborg, M.J.,

1991. Electric load forecasting using an artificial neural network. IEEE transactions on Power Systems, 6(2), pp.442-449.

[10] Papalexopoulos, A.D. and Hesterberg, T.C., 1990. A regression-based approach to short-term system load forecasting. IEEE Transactions on Power Systems, 5(4), pp.1535-1547.

[11] Papalexopoulos, A.D., Hao, S. and Peng, T.M., 1994. An implementation of a neural network based load forecasting model for the EMS. IEEE transactions on Power Systems, 9(4), pp.1956-1962.

[12] Hippert, H.S., Pedreira, C.E. and Souza, R.C., 2001. Neural networks for short-term load forecasting: A review and evaluation. IEEE Transactions on power systems, 16(1), pp.44-55.

[13] Ekonomou L. Greek long-term energy consumption prediction using artificial neural networks[J]. Energy, 2010, 35(2): 512-517.

[14] Kavaklioglu K, Ceylan H, Ozturk H K, et al. Modeling and prediction of Turkey's electricity consumption using artificial neural networks[J]. Energy Conversion and Management, 2009, 50(11): 2719-2727.

[15] Li K, Hu C, Liu G, et al. Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis[J]. Energy and Buildings, 2015, 108: 106-113.

[16] Akay D, Atak M. Grey prediction with rolling mechanism for electricity demand forecasting of Turkey[J]. Energy, 2007, 32(9): 1670-1675.

[17] Sarda P, Antoch J. Electricity consumption prediction with functional linear regression[J].

[18] Tso G K F, Yau K K W. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks[J]. Energy, 2007, 32(9): 1761-1768.

[19] Beccali M, Cellura M, Brano V L, et al. Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area[J]. Renewable and Sustainable Energy Reviews, 2008, 12(8): 2040-2065.

[20] Kaytez F, Taplamacioglu M C, Cam E, et al. Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines[J]. International Journal of Electrical Power & Energy Systems, 2015, 67: 431-438.

[21] Hamzacebi C, Es H A. Forecasting the annual electricity consumption of Turkey using an optimized grey model[J]. Energy, 2014, 70: 165-171.

[22] Azadeh A, Ghaderi S F, Tarverdian S, et al. Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption[J]. Applied Mathematics and Computation, 2007, 186(2): 1731-1741.

[23] Neto A H, Fiorelli F A S. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption[J]. Energy and buildings, 2008, 40(12): 2169-2176.

[24] Zhao H, Magoulès F. A review on the prediction of building energy consumption[J]. Renewable and Sustainable Energy Reviews, 2012, 16(6): 3586-3592.

[25] Gonzalez P A, Zamarreno J M. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network[J]. Energy and Buildings, 2005, 37(6): 595-601.

[26] Karatasou S, Santamouris M, Geros V. Modeling and predicting building's energy use with artificial neural networks: Methods and results[J]. Energy and buildings, 2006, 38(8): 949-958.

[27] Azadeh A, Ghaderi S F, Sohrabkhani S. A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran[J]. Energy Policy, 2008, 36(7): 2637-2644.

[28] Platon R, Dehkordi V R, Martel J. Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis[J]. Energy and Buildings, 2015, 92: 10-18.

[29] Escrivá-Escrivá G, Álvarez-Bel C, Roldán-Blay C, et al. New artificial neural network prediction method for electrical consumption forecasting based on building end-uses[J]. Energy and Buildings, 2011, 43(11): 3112-3119.

[30] Darbellay G A, Slama M. Forecasting the short-term demand for electricity: Do neural networks stand a better chance?[J]. International Journal of Forecasting, 2000, 16(1): 71-83.

[31] Park D C, El-Sharkawi M A, Marks R J, et al. Electric load forecasting using an artificial neural network[J]. IEEE transactions on Power Systems, 1991, 6(2): 442-449.

[32] Taylor J W, Buizza R. Neural network load forecasting with weather ensemble predictions[J]. IEEE Transactions on Power systems, 2002, 17(3): 626-632.

[33] Mandal P, Senjyu T, Urasaki N, et al. A neural network based several-hour-ahead electric load forecasting using similar days approach[J]. International Journal of Electrical Power & Energy Systems, 2006, 28(6): 367-373.

[34] Karsoliya S. Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture[J]. International Journal of Engineering Trends & Technology, 2012, 3(6).