Probabilistic Load Forecasting of Adaptive Multiple Polynomial Regression considering Temperature Scenario and Dummy variables

Jiang Li 1, Liyang Ren 1, Baocai Wang 1 and Guoqing Li 1

1Northeast Electric Power University, Jilin, Jilin, 132000, China
2China Electric Power Research Institute, Beijing, 100000, China
3Liyang Ren: 2024468066@qq.com

Abstract. The monthly or yearly low accurate history data always leads to the low prediction-accuracy for load forecasting. We use temperature data from Sydney, Australia and the New South Wales Natural Load Dataset. To improve the data-based forecasting accuracy and time related scenario, this paper builds an adaptive multiple polynomial regression model considering temperature scenario and dummy variables. These dummy variables are divided into three aspects: trend variables, date variables and temperature variables. Trend variables are used to predict the whole economic development and user habit. Date variables are introduced to deal with the characteristics of working days and holidays. Cubic function for temperature variables from Australia and the New South Wales electric load history data is constructed to describe the relationship between load and temperature scenario. A temperature scenario is generated by considering the different loads of different seasons and the probability search of different scenarios. The load forecasting interval under different scenarios is given and analyzed by using dummy variables. At last, the method is validated based on the history data in a certain area. The prediction result with high accuracy shows clear intuitive and powerful interpreting ability, which can provide reliable decision basis for long term load forecasting. After simulation analysis, the accuracy of load forecasting based on 3-year history increases by 3.8%.

1. Introduction
Long-term load forecasting is very important for the production, operation, planning and construction of power systems, which is the basis and also the value of history data mining [1, 2]. With load diversification and the accessing of large-scale distributed renewable energy sources, it is more and more difficult to give the load forecasting interval under complex scenarios.

In recent years, the load forecasting mainly focuses on short-term load forecasting, topic papers about long-term load forecasting are relatively fewer [3]. In practice, forecasting is essentially a stochastic problem. Thus, exact forecasting for the future is impossible, and it can be assumed that forecasting for long-term horizons can only be the reference for reducing the effect of uncertainty as few as possible [4]. One way to counter this assumption is the scenario analysis that looks into a selected scenario in the future. Due to the uncertainty in weather and economic forecasts, forecasting process is encouraged to provide explicit forecasting value based on different scenarios. The other load forecasting methods are predictive modeling, weather normalization, and probabilistic forecasting [5]. There are many
traditional short-term load prediction methods, such as regression prediction method and gray prediction method. There are also other intelligent prediction algorithms, such as support vector machine method and neural network method [6]. Gray prediction method requires less sample datum and is easy to achieve. However, the demand load has an exponential trend [7-9]. Neural network method has effective prediction results. The black-box model cannot explain the relationships between input and output variables, which makes the model less able to explain and is easily trapped in local optimal solution. Therefore, it is very difficult to initialize the model [10]. The regression analysis method is simple in calculating principle and has a clear solving algorithm. The prediction speed is fast and has a strong explanatory power of the model, and it is the earliest used in load forecasting. Literature [11] proposes a new approach to support the process of forecasting the hourly electric load values for the next day. The adopted methodology based on neural networks is only supported by detailed information related to consumers’ typical behavior and climatic information. The case study was tested in two real distribution substation outputs, demonstrating its effectiveness and practical applicability in [12]. Literature [13,14] provides new ideas for regression prediction. However, the method cannot reflect the inherent mechanism of load fluctuation, and just considers the quantitative factors such as gross national product and population, neglecting the meteorological temperature, periodic load characteristics, and the special nature of the holiday load, which affect the adaptability of proposed method under different scenarios.

With the increase of economic level, the proportion of temperature-sensitive load in the home is increasing, which makes the load more and more obvious with temperature. Due to the uncertainty of temperature, load forecasting is a random problem. The main methods are point forecasting and cannot determine the forecasting interval of load fluctuations in the future, So, it is unscientific to judge the long-term load forecasting by comparing the predicted and true values of the corresponding points [15-16]. The low accuracy features of the traditional prediction methods provide very limited information for the prediction model, their prediction errors are large and have poor interpretation ability, such as the monthly maximum or minimum temperature, and it cannot explain the specific moment when this temperature appears and the dynamic characteristics of the load with temperature [17, 18]. Therefore, this paper proposes a high-precision load forecasting method that adapts to different data quality to solve such problems.

The main contribution of this paper is to generate temperature scenario and applied into probabilistic load forecasting problem by using dummy variables. The long-term load forecasting accuracy is improved and both upper boundary and lower boundary are given with probabilistic forecasting. [19]

Based on the hourly history data, we first establish a regression model, dummy variables are used to quantify the year, week and day of the dummy variables. When the weekly history data are classified, it should take into account the special nature of holidays; the temperature effect is considered for periodic load forecasting under working days and holidays. Compared with different term scale scenario for forecasting error, the optimal scenario with high accuracy is generated. The probability is used to optimize the load parameters and the forecasting interval is used to define the load change. We will explain the concept of the temperature scene in 3.3.4 and introduce the construction of the temperature scene in the form of simulation verification in Section 4.2.

The remainder of the paper is organized as follows: Firstly, the generalized multivariate linear regression model for load forecasting is established based on per hour history data and regression constant in section2. Then, the detailed model for probabilistic load forecasting is described by using trend variables, date dummy variables and temperature scenario in Section 3. Finally, the performance of the proposed method is verified in Section 4. Section 5 concludes the paper and proposes future work.

2. Generalized Multiple Linear Regression
In this section, a multivariate linear regression model is firstly given and then the polynomial regression model is proposed to solve uncertainties from working days and holidays.

The general form of a multivariate linear regression model:

\[ Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + e_i (i = 1,2,...n) \]  

(1)
Therefore, $\beta_0$ stands for the regression constant, $\beta_0, ..., \beta_p$ stands for the partial regression coefficient. Y is called the explained variable (dependent variable), $X_1, X_2, ..., X_p$ is called the explanatory variable (independent variable), $e_i$ is the random error [20-22]. Compared with other load forecasting methods, the proposed load forecasting method based on principal component regression effectively retains most information of the original variables and reduces the correlation among the data, finally improve the accuracy of load forecasting [23].

In practical problems, the relationship between the explained variable Y and the explanatory variable X is not linear in many models and they can be transformed into a linear relationship through the functional relationship of independent variables or dependent variables. Linear regression could be used to solve unknown parameters and make regression diagnosis [24].

In polynomial regression, the influencing variable may be a polynomial, or they are the two independent variables that have an interaction effect, the regression equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + e_i$$

(2)

The polynomial regression is transformed into a linear regression of four variables:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{12} + \beta_3 X_{13} + \beta_4 X_{14} + e_i$$

(3)

In the regression analysis, we first quantify the qualitative variables by quantifying some independent variable and then introduce dummy variables that take only two values of 0 and 1. When an attribute appears, the dummy variable takes 1, and otherwise 0. If a qualitative variable has K categories, it is necessary to introduce K-1 0-1 virtual arguments, taking working days and holidays as an example.

$$\begin{cases} 
X_1 = 1 & \text{working days} \\
X_1 = 1 & \text{holidays}
\end{cases}$$

Then a regression equation with load characteristics for the working days is described as follows:

$$Y = \beta_0 + \beta_1 X_1$$

(4)

When describing the working days, $X_1=1$, the regression equation is: $E(Y)=\beta_0+\beta_1$. When describing holidays, $X_1=0$ the regression equation is $E(Y)=\beta_0$ in [25]. The resulting daily load characteristics are described by regression constants.

3. Building the forecasting model

In this section, dummy variables, such as trend variables and data variables, are firstly introduced. Then, interaction among different variables is modeled in linear regression expression. Finally, two methods for generating temperature scenes, such as moving day temperature method and probabilistic temperature scene creation method, are proposed, and the probability prediction errors under different time scenarios are analyzed.

3.1. Trend variables

Data are sourced from Sydney's temperature in Australia and the Natural Load Dataset in New South Wales.

Figure 1 plots the hourly load and temperature scatter plots for a region from 2006 to 2013, and Table 1 shows the annual load table, and it is relatively stable. There is no annual increase or decrease trend. This may be caused by social-economic development and population growth resulting in increased electricity consumption.
In order to actually describe the trend of increasing load, we introduce the trend variables $T_r$ in the regression model and define the rise of a series of natural numbers per hour to quantify the load growth trend. For example, in the first hour of 2006, the trend variable was 1, the second hour was 2, and then the analogy. This trend variable is a linear approximation of the load growth sequence. The trend expression of economic growth is expressed as:

$$Load_t = \beta_0 + \beta_1 T_t + e$$

(5)

3.2. Date variable

Power consumption behavior is one of the main factors that affect load fluctuation. This section describes the load characteristics of periodic daily, weekly and yearly variables by date. As can be seen from Fig. 1, the annual load has a periodic pattern of load fluctuation. The yearly component of the load is closely related to seasonal climate characteristics. The peak loads of summer and winter reach maximum, while loads of spring and autumn are minimum. This paper introduces the virtual independent variable $M$ for 12 categories, the treatment is as follows.
The regression equation described the monthly load characteristics is:

\[
Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{12} X_{12}
\]

\[
\text{Load}_t = \beta_0 + \beta_1 M_t + \beta_2 D_t + \beta_3 H_t + \beta_4 D_t H_t + e
\]  

(6)

Where \(\beta\) is the regression coefficient and \(M_t, H_t, D_t\) is the dummy variable. \(H_t D_t\) represents the interaction between the dummy variables \(D\) (day) and \(H\) (hour). \(e\) indicates random error. When the load is described in January, the variables are \(X_1 = 1, X_2 = X_3 = \ldots = X_{11} = X_{12} = 0\) in the regression equation.

The load on different date types is also very different within a week, but shows a clear periodic pattern. As shown in Figure 2. In normal days, there was a significant difference between weekdays and weekends, the total load of the weekends was significantly lower than the daily cyclical changes.

![Figure 2. Weekly load (2006/3/25—2006/4/2)](image-url)
In order to describe the load characteristics, we introduce the independent dummy variable D to describe the load difference between different date types. One week can be divided into 7 categories. 6 dummy arguments and processing methods are introduced into the monthly variable M [23]. Because it reduces industrial electricity consumption, in the daily cycle, the load characteristics were significantly different at different times of the day, and the nighttime electricity consumption was significantly lower than that during the daytime. The virtual independent variable H is introduced to describe the load characteristics, which is divided into 24 categories and introduced 23 dummy variables.

Working day morning shift load is significantly higher than the holiday morning, this is due to the fact that people do not have to get up early to work on a day off, and reduce the load on electricity, and we introduce interaction H and D in the model. Due to the different load components, each holiday generally occurs in a fixed period of time every year. During the holidays, a large number of factories, enterprises and institutions to withdraw from the electricity load, they are mainly including residential load, commercial load and non-stop industrial load, this made the load significantly reduced from the normal day [24]. According to the flexible adjustment policy, current holidays will be converted into lasted holidays or working days, and it will raise the overall forecast level. In summary, the regression equation can be expressed as follows,

\[
\text{Load}_t = \beta_0 + \beta_1 T_t + \beta_2 T^2_t + \beta_3 T^3_t + \beta_4 H_t + \beta_5 D_t H_t + e
\]

(7)

3.3. Temperature Scenarios Generation

1) Analysis of Temperature Variables

In this section, the load-temperature function of the cubic function is introduced. Figure 3 gives an hourly temperature-load scatter plot for a region from 2006 to 2011, and its section linear, quadratic, and cubic fitting functions are plotted. The temperature-load relationship is asymmetric, while the quadratic function can only describe the symmetrical function. Thus, the cubic function is better than the quadratic function for load forecasting.

2) Interaction of Temperature Variables

The temperature of summer is higher than that in winter, the temperature is distinguishing in different months. M*T should be considered in the interaction between month variable M and temperature variable T. During the day, the temperature in different time periods also changes regularly. The daytime temperature is higher than that of night, and the interaction between variables H and T need to be considered. The temperature function in the regression model is:

\[
\text{Load}_t = \beta_0 + \beta_1 T_t + \beta_2 T^2_t + \beta_3 T^3_t + \beta_4 H_t + \beta_5 T^2_t H_t + \beta_6 T^3_t H_t + \beta_7 T_t M_t + \beta_8 T^2_t M_t + \beta_9 T^3_t M_t + e
\]

(8)
Where $\beta$ is the regression coefficient and $M, H, D, T$ is the dummy variable. $H \times T, H \times T^2, H \times T^3$ is the interaction between the variables $H$ and $T$. $M \times T, M \times T^2, M \times T^3$ is the interaction between variables $M$ and $T$. $e$ indicates random error.

3) Proximity of Temperature

Proximity is a phenomenon in psychology, referring to the phenomenon that when people recognize a series of things, the memory effect of the last part of the items is better than that of the middle part. The same phenomenon exists between the load and the temperature, that is, the current time before the temperature will also affect the load changes. We add temperature variables into the model, introducing the same form of variables as $T$. $T_t$ refers to the temperature of the first i hours ($i = 1, 2, 3$), as, $T_{s,1}^i, T_{s,2}^i, T_{s,3}^i$. $T_t$ refers to the temperature average of the first 24 hours of the current time. Proximity variables $T_{s,1}^i, T_{s,2}^i, T_{s,3}^i$. $T_{s,1}^i, T_{s,2}^i, T_{s,3}^i$.

Considering the proximity effect, the function of load forecasting is:

$$Load_i = \beta_0 + \beta_1 T_i + \beta_2 T_i^2 + \beta_3 T_i^3 + \beta_4 T_{s,1}^i + \beta_5 T_{s,2}^i + \beta_6 T_{s,3}^i + \beta_7 T_i H_i + \beta_8 T_i^2 H_i + \beta_9 T_i^3 H_i + \beta_{10} T_{s,1}^i M_i + \beta_{11} T_{s,2}^i M_i + \beta_{12} T_{s,3}^i M_i + \beta_{13} T_i^2 M_i + \beta_{14} T_i^3 M_i + \beta_{15} T_{s,1}^i H_i + \beta_{16} T_{s,2}^i H_i + \beta_{17} T_{s,3}^i H_i + \beta_{18} T_i H_i + \beta_{19} T_i^2 H_i + \beta_{20} T_i^3 H_i + \beta_{21} T_{s,1}^i M_i + \beta_{22} T_{s,2}^i M_i + \beta_{23} T_{s,3}^i M_i + \beta_{24} T_i^2 M_i + \beta_{25} T_i^3 M_i + \beta_{26} T_{s,1}^i H_i + \beta_{27} T_{s,2}^i H_i + \beta_{28} T_{s,3}^i H_i + \beta_{29} T_i H_i + \beta_{30} T_i^2 H_i + \beta_{31} T_i^3 H_i + \beta_{32} T_{s,1}^i M_i + \beta_{33} T_{s,2}^i M_i + \beta_{34} T_{s,3}^i M_i + \beta_{35} T_i^2 M_i + \beta_{36} T_i^3 M_i + e$$  (9)

In summary, the multiple linear regression load forecasting model based on time-temperature near-effect is:

$$Load_i = \beta_0 + \beta_1 T_i + \beta_2 M_i + \beta_3 D_i + \beta_4 H_i + \beta_5 D_i H_i + \beta_6 T_i + \beta_7 T_i^2 + \beta_8 T_i^3 + \beta_9 T_i H_i + \beta_{10} T_i^2 H_i + \beta_{11} T_i^3 H_i + \beta_{12} T_{s,1}^i M_i + \beta_{13} T_{s,2}^i M_i + \beta_{14} T_{s,3}^i M_i + \beta_{15} T_i^2 M_i + \beta_{16} T_i^3 M_i + \beta_{17} T_{s,1}^i H_i + \beta_{18} T_{s,2}^i H_i + \beta_{19} T_{s,3}^i H_i + \beta_{20} T_i H_i + \beta_{21} T_i^2 H_i + \beta_{22} T_i^3 H_i + \beta_{23} T_{s,1}^i M_i + \beta_{24} T_{s,2}^i M_i + \beta_{25} T_{s,3}^i M_i + \beta_{26} T_i^2 M_i + \beta_{27} T_i^3 M_i + e$$  (10)

4) Probabilistic Temperature Scenario Generation

The meteorological characteristics of the same period of each year are similar. A type of meteorology may arrive a few days earlier, or it may arrive a few days later. For example, the temperature in May 2006 may be similar to the temperature in April or June 2009. There is a strong correlation between load and temperature, and the load will follow the temperature change. It is expressed that if the high temperature lasts for a long time in the summer, people will continue to use the air conditioner to cool down, resulting in an increase in load. This temperature change phenomenon will lead to large load differences, so this temperature change characteristics should be considered in long-term load forecasting, and power planning and scheduling should be done to fully cope with this uncertain meteorological variation.

In this section, a probabilistic temperature scenario generation method based on a moving temperature scenario is proposed, which is compared with the fixed temperature scenario generation method. If history year is k, then we will generate kth probability temperature scenario. The moving day temperature method is based on the change characteristic of temperature, and the historical temperature is moved forward or backward by n days to create more equal-probability historical temperature scenes, taking table.3 forward and backward one day as an example. If k history moves forward and backward by n days, then (2n+1) k temperature scenes are generated.

| Table 2. Schematic diagram of shifted-data |
|-------------------------------------------|
| Base year | $T_{(a+365-i)}$ | $T_{(a+1)}$ | $T_{(a+366)}$ |
| Move forward for 1 day | $T_{(a+1)}$ | $T_{(a+2)}$ | $T_{(a+3)}$ |
| Moved 1 day later | $T_{(a+364-i)}$ | $T_{(a+365-i)}$ | $T_{(a+366)}$ |

The history year and moving days of temperature are the key indicators that influence the prediction accuracy. We use the error parameter optimization method, and the formula is:
\begin{equation}
S(y_{t,q}, y_{t}, q) = \begin{cases} 
(1 - \frac{q}{100}) (y_{t,q} - y_{t}) & y_{t} < y_{t,q} \\
\frac{q}{100} (y_{t} - y_{t,q}) & y_{t} \geq y_{t,q}
\end{cases}
\end{equation}

Where, \( q \) is the given value (\( q=1,2,\ldots,99 \)). \( y_t \) is the actual load at time \( t \), \( y_{t,q} \) is the \( q \)-digit load at time \( t \). The smaller the value is, the smaller its error.

3.4. System flow chart

This section predicts the date entered. First, it is judged whether it is a normal working day. If it is an abnormal working day, it is corrected according to the temperature scene and then enters the normal cycle. The normal cycle is predicted once per hour to 24 hours, and \( H \) returns to zero to start a new day forecast.

The load forecasting system flow chart is shown in Figure 4.
4. Case study
In this Section, we use temperature data from Sydney, Australia and the New South Wales Natural Load Dataset. the hourly data from 2006 to 2010 in regression model are used to predict the load in 2011 and
analyze error. The mean absolute percentage error (MAPE) is used to evaluate the prediction accuracy of the model.

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\%$$

Where $y_t$ is the real load, and $\hat{y}_t$ is the predictive value.

### Table 3 Deviation for different models

| Model | $R^2$  | MAPE (%) | Standard deviation (MW) |
|-------|--------|----------|-------------------------|
| 1     | 0.867  | 8.94     | 318.30                  |
| 2     | 0.944  | 4.27     | 186.57                  |
| 3     | 0.952  | 3.33     | 143.40                  |
| 4     | 0.967  | 2.95     | 128.59                  |

In model 1 single-factor variables ($T, M, D, H, T, T^2$ and $T^3$) are considered, the accuracy was also low because of interaction between variables. Model 2 added the coupled variable with temperature and date ($M\cdot T, M\cdot T^2, M\cdot T^3, D\cdot T, D\cdot T^2, D\cdot T^3, D\cdot H, H\cdot T, H\cdot T^2$ and $H\cdot T^3$) into the regression model for improving forecasting accuracy. Model 3 has been revised on the coupled variables of Model 2. Model 4 adds the short-term effect of temperature scenario, the corrected $R^2$ reaches 0.967, and MAPE is reduced to 2.95%, which verifies the validity of short-term effect to improve prediction accuracy for temperature.

The prediction model 4 considering temperature is

$$L_t = \beta_0 T_t + \beta_1 M_t + \beta_2 D_t + \beta_3 H_t + \beta_4 D_t H_t + f(T_t)$$

Where $f(T_t)$ is the temperature model and the expression is

$$f(T_t) = \beta_{11} T_{t}^3 + \beta_{12} M_{t} T_{t} + \beta_{13} D_{t} T_{t} + \beta_{14} H_{t} + \beta_{15}MT_{t}^2 + \beta_{16}MT_{t}^3 + \beta_{17}HT_{t} + \beta_{18}HT_{t}^2 + \beta_{19}MH_{t} + \beta_{110}MHT_{t}^2 + \beta_{111}MHT_{t}^3$$

During the year, summer is higher than winter temperature, that is, there are differences in temperature in different months, and the interaction between variables $M$ and $T$ should be considered. $M\cdot T, M\cdot T^2, M\cdot T^3$; During the day, the temperature of different time periods also changes periodically. The temperature at noon is higher than that at night.

#### 4.1. Data length for estimating the regression parameters

The length of the history data in the regression model is a key factor affecting the forecasting accuracy. Table 4 lists error for different data length. In the table 4, the second line is based on last 2-years history data to forecast the load power. It can be seen from the average value that the minimum error 3.17% of parameter estimation can be obtained by using 3-years history data. This paper selects the data from last 3-years to estimate the regression parameters.

### Table 4. Error for different data length

| DATA LENGTH (years) | 2009 (%) | 2010 (%) | 2011 (%) | 2012 (%) | Average (%) |
|---------------------|----------|----------|----------|----------|-------------|
| 1                   | 4.09     | 3.23     | 3.19     | 2.89     | 3.35        |
| 2                   | 4.30     | 2.94     | 2.94     | 2.67     | 3.21        |
| 3                   | 4.26     | 2.89     | 2.96     | 2.58     | 3.17        |
| 4                   | 4.37     | 2.64     | 3.20     | 2.70     | 3.23        |
4.2. Temperature scenario generation

Probabilistic load forecasting method flow is as follows. First, the probabilistic parameter optimization is performed. The k-n parameter with the highest accuracy is selected, and (2n+1)k temperature scenes are created as the input of the prediction model for each temperature scene. Forecasting separately to simulate the predicted annual temperature, and obtaining (2n+1)k prediction results, these results can be used to find the median or interval division, which is of great significance for guiding medium and long-term power grid planning and scheduling.

A temperature scenario based on from history 2005-2012 was generated, and the load forecasting was based on 2013 actual date type (k = 1, 2, ..., 8). When k = 1, the scenario is generated from the 2012 temperature data, and when k = 2, the temperature scenario is based on 2011-2012, and so on. This section will search for the optimal k-n parameter by moving the history year temperature data under k-n year(s) and n day(s), and Table 4 shows the probabilistic error of different temperature scenarios.

As can be seen in the Table 5, the optimal k-n parameter is 8-13 days, which is based on the 8-year and 13 days, probabilistic error is 59.35%, which has higher accuracy than 63.05%for 8-year fixed day temperature scenario. Probabilistic error for 8-year and the fixed-day temperature scenario was 4.69%, and the median error of MAPE was 5.01%. Figure 5 shows the probability error curves of temperature scenes based on different history years (k = 1, 2, ..., 8). It can be seen that with moving-days increasing, the probability error fluctuating. In the initial stage, the forecasting accuracy can be significantly improved by increasing the number of moving days. When k = 1, it means the temperature data moving forward three days in 2012, the probability error can be reduced from 78.37% to 66.27%, and the median MAPE dropped from 5.67 to 5.01, the accuracy increased to 11.6%.

| k-n | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  |
|-----|----|----|----|----|----|----|----|----|----|
| 0   | 78.37 | 77.23 | 69.63 | 67.86 | 64.48 | 63.49 | 63.05 | 63.05 |
| 1   | 76.42 | 76.41 | 63.36 | 64.45 | 61.83 | 61.08 | 60.99 | 63.58 |
| 2   | 70.43 | 68.82 | 61.63 | 62.41 | 60.61 | 60.12 | 60.13 | 61.57 |
| 3   | 66.27 | 64.34 | 60.79 | 61.21 | 59.99 | 59.94 | 59.86 | 60.02 |
| 4   | 65.45 | 64.17 | 60.49 | 61.00 | 59.78 | 59.87 | 59.63 | 60.09 |
| 5   | 64.77 | 63.71 | 60.40 | 60.32 | 59.77 | 59.77 | 59.63 | 59.92 |
| 6   | 63.57 | 62.46 | 60.38 | 60.83 | 59.83 | 59.76 | 59.68 | 59.55 |
| 7   | 63.62 | 63.50 | 60.31 | 60.63 | 59.88 | 59.76 | 59.70 | 59.70 |
| 8   | 62.61 | 62.8 | 60.26 | 60.55 | 59.87 | 59.75 | 59.73 | 59.69 |
| 9   | 62.80 | 62.67 | 60.24 | 60.30 | 59.82 | 59.65 | 59.78 | 59.48 |
| 10  | 62.56 | 62.64 | 60.22 | 60.28 | 59.80 | 59.75 | 59.84 | 59.42 |
| 11  | 61.91 | 62.21 | 60.20 | 60.13 | 59.73 | 59.74 | 59.87 | 59.38 |
| 12  | 61.03 | 61.90 | 60.17 | 60.01 | 59.68 | 59.73 | 59.86 | 59.35 |
| 13  | 61.72 | 61.85 | 60.20 | 60.04 | 59.70 | 59.72 | 59.93 | 59.43 |
| 15  | 61.23 | 62.16 | 60.32 | 60.4 | 59.91 | 59.87 | 60.12 | 59.75 |
| 20  | 62.03 | 62.57 | 60.73 | 61.08 | 60.07 | 60.20 | 60.30 | 60.29 |
| 30  | 62.85 | 62.71 | 61.5 | 61.87 | 60.78 | 60.94 | 61.15 | 61.07 |
The regression model parameters are estimated by using temperature scenario and the 2014 load power are forecasted. The dashed line in Figure 6 is the forecasting curve based on the temperature scenario. Black is the actual load and the red line is the median load. It can be seen that compared with the forecasting accuracy of fixed-day temperature, the forecasting accuracy of moving-day temperature scenario is obviously improved [26].

Table 6 summarizes the forecast results. The results show that the probability error at the moving day temperature of July 16 to July 22, 2014 is 1.82, which is larger than the 2.62 from the fixed day temperature. In the median, the MAPE was 4.07, representing an increase of 40.32% under the fixed-day temperature of 6.82. The 2014 full-year median load MAPE dropped from 4.90 to 4.76. It can be seen that the predictive accuracy of the probabilistic load based on the moving day temperature scenario is significantly higher than that of fixed day forecast, especially in the case of historical temperature.
data is limited, and it can also reach higher forecast level by moving daily temperature.

Table 6. Error statistics

| error        | 2014 | 7/16-7/22,2014 |
|--------------|------|----------------|
|              | Fixed-day | Mobile-day | Fixed-day | Mobile-day |
| Probability  | 63.63    | 60.53       | 2.62      | 1.82        |
| error (MAPE) | 4.90     | 4.76        | 6.82      | 4.07        |

Based on the preferred parameter k-h, according to the 2014 actual data type, we forecast the 2014 monthly electricity consumption and monthly maximum load a 10% quantile, a median and a 90% quantile load value are taken at each of the forecast points.

Figure 7 and 8 show monthly maximum load and monthly electricity consumption respectively. The broken line in the figure represents the predicted value based on the historical temperature and the daily moving temperature data from 2005 to 2013, and the three solid lines from top to bottom are respectively 90% quantiles, median and 10% quantiles, black solid points represent the true value of the load.

As can be seen from the figure, most of the actual load points fall near the median, and individual points fall outside the range of 10% and 90% quantile lines. The 10% and 90% quantile loadings indicate the extreme cases with a lower probability of occurrence, but that does not mean it will never happen. The maximum load in February, October, November and December and the total electricity consumption in June and October are all in the 10% quantile line. In March and April, electricity consumptions all in the 90% quantile. The maximum load in May was below the 10% quantile, and the electricity consumption in January exceeded the 90% quantile. It can be seen that the defined prediction interval can reflect the real value of the load more accurately.

As can be seen from the figure, most of the actual load points fall near the median, and individual points fall outside the range of 10% and 90% quantile lines. The 10% and 90% quantile loadings indicate the extreme cases with a lower probability of occurrence, but that does not mean it will never happen. The maximum load in February, October, November and December and the total electricity consumption in June and October are all in the 10% quantile line. In March and April, electricity consumptions all in the 90% quantile. The maximum load in May was below the 10% quantile, and the electricity consumption in January exceeded the 90% quantile. It can be seen that the defined prediction interval can reflect the real value of the load more accurately.

Figures 9 and 10 show the temperature scenarios created with the data of 2005-2011 to forecast the monthly maximum load and monthly electricity consumption for 2012-2014. As can be seen from the figure, the annual maximum load in February-May and the electricity consumption in August-November is small, and it even below the 10% quantile load value, during which the relevant part should be done according to the 10% quantile. In January of each year, it is reasonable to arrange the power generation plan according to the median monthly maximum load. In July, the power generation plan in the light of 90% quantile load would be more reasonable.
In comparison with the point load forecasting, the proposed probability forecasting method provides a series of load changes, which can reflect the fluctuation range and trend of load fluctuation more accurately and define different quantile intervals as needed. In different time periods, the width of the forecast interval is different, this provides policy makers with more useful information, which is unmatched by point forecasts.

5. Conclusion
This paper extends the linear multiple linear regression model into the adaptive polynomial multiple regression model. Trend variables, date variables and temperature variables as dummy variables are used to describe the inherent characteristics of load changes in future. Economic development, utility consumption habit of working day and holidays, temperature effect and so on are viewed as linear, quadratic and even triple terms of the polynomial model. The proposed method quantifies 12 months, 7 days and 24 hours categories as the main factors for scenario generation. Temperature scenario optimization is applied to analyzing load forecasting median error and border, and load forecasting accuracy based on 3 years history is improved with 3.8%.

Case studies show that the proposed probability forecasting method can explain the trend of future load changes more accurately, and it can provide more useful information for the long-term load forecasting. It will help policy-makers estimate the possible uncertainties and risk factors of future loads. This will lay a solid foundation for load forecasting in complex operations.

Acknowledgment
Thanks to the National Natural Science Foundation of China for the project: Probability prediction and active smoothing theory of renewable energy slope events in AC and DC power grids (Project No. 51977030)

Author Biographies
Jiang Li In 2003, he obtained a bachelor's degree in electrical engineering and automation from Shanghai Electric Power College; In 2006, he obtained a master's degree in electrical engineering and automation from Northeastern Electric Power University; In 2010, he obtained a doctorate degree in electrical engineering and automation from North China Electric Power University; Visiting scholar at Cornell University in the United States in 2014; In 2015, he was a visiting scholar at the American Energy System Research Center.

Liyang Ren In 2017, he obtained a bachelor's degree in electrical engineering and automation from Northeastern Electric Power University; Master's degree in Power Systems and Automation from Northeastern Electric Power University from 2017 to the present;
References

[1] Kanggu Park; Seungwook Yoon; Euiseok Hwang. Hybrid Load Forecasting for Mixed-Use Complex Based on the Characteristic Load Decomposition by Pilot Signals. IEEE Access. December 2018; pp.12297-12306.

[2] Mohamed Reda Nezzar; Nadir Farah; Tarek Khadir. Mid-long term Algerian electric load forecasting using regression approach. IEEE Transactions on Power Systems, July 2013; pp.121-126.

[3] Weicong Kong; Zhao Yang Dong; et al. Short-Term Residential Load Forecasting Based on Resident Behaviour Learning. 2018; pp.1087-1088

[4] T. Hong, J. Wilson, J. Xie.; Long term probabilistic load forecasting and normalization with hourly information; 2013, pp. 456-462

[5] Qingshan Xu; Yifan Ding; Qingguo Yan; et al. Day-Ahead Load Peak Shedding/Shifting Scheme Based on Potential Load Values Utilization: Theory and Practice of Policy-Driven Demand Response in China. IEEE Access August 2017; pp.22892-22901.

[6] Chen Y.; Kloft M.; Yang Y.; et al. Mixed kernel based extreme learning machine for electric load forecasting. Neurocomputing, 2018.

[7] Zhang X.; Wang R.; Zhang T.; et al. Short-Term load forecasting using a novel deep learning framework. Energies, 2018, 11, 1554.

[8] Shepero M.; Meer D. V. D.; Munkhammar J.; et al. Residential probabilistic load forecasting: A method using Gaussian process designed for electric load data. Applied Energy, 2018; pp.159-172.

[9] Bowen Li; Jing Zhang; Yu He; et al. Short-Term Load-Forecasting Method based on Wavelet Decomposition With Second-Order Gray Neural Network Model Combined With ADF Test. IEEE Access, May 2017; pp.16324-16331.

[10] Li Y.; Huang Y.; Zhang M.; et al. Short-Term load forecasting for electric vehicle charging station based on niche immunity lion algorithm and convolutional neural network. Energies, 2018.

[11] Wang Y.; Zhang N.; Chen Q.; et al. Data-driven probabilistic net load forecasting with high penetration of invisible PV. IEEE Transactions on Power Systems, 2017; pp.1-1.

[12] Fan G.F.; Peng L.L.; Hong W.C.; Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model. Applied Energy, 2018, 224, 13-33.

[13] Giuseppe fenza; Mariacristina Gallo; Vincenzo Loia. Drift-Aware Methodology for Anomaly Detection in Smart Grid, IEEE Access. December 2018; pp.9645-9657.

[14] Singh P.; Dwivedi P.; Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem. Applied Energy, 2018, 217, 537-549.

[15] Prakash A.; Xu S.; Rajagopal R.; et al. Robust building energy load forecasting using physically-based kernel models. Energies, 2018, 11, 862.

[16] Yang Y.; Li S.; Li W.; et al. Power load probability density forecasting using Gaussian process quantile regression. Applied Energy, 2018, 213.

[17] Barman M.; Choudhury N.B.D.; Sutradrhar S. A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India. Energy, 2018, 145.

[18] Karimi M.; Karami H.; Gholami M.; et al. Priority index considering temperature and date proximity for selection of similar days in knowledge-based short term load forecasting method. Energy, 2018, 144, 928-940.

[19] Simona Vasilica Oprea; Adela B’RA; Vlad Diaconta. Sliding Time Window Electricity Consumption Optimization Algorithm for Communities in the Context of Big Data Processing. IEEE Access December 2018, pp. 13050-13067.

[20] Yang Z.C.; Discrete cosine transform-based predictive model extended in the least-squares sense for hourly load forecasting. IET Generation Transmission & Distribution, 2016, 10, 3930-3939.

[21] Kaur A.; Nonnenmacher L.; Coimbra C.F.M.; Net load forecasting for high renewable energy penetration grids. Energy, 2016, 114, 1073-1084.
[22] Gu C.; Jirutitijaroen P. Dynamic state estimation under communication failure using kriging based bus load forecasting. IEEE Transactions on Power Systems, 2015, 30, 2831-2840.
[23] Park H.; Baldick R.; Morton D.P.; A stochastic transmission planning model with dependent load and wind forecasts. IEEE Transactions on Power Systems, 2015, 30, 3003-3011.
[24] Che J.X.; Wang J.Z.; Short-term load forecasting using a kernel-based support vector regression combination model. Applied Energy, 2014, 132, 602-609.
[25] Hernández L.; Baladrón C.; Aguiar J.M.; et al. Artificial neural networks for short-term load forecasting in microgrids environment. Energy, 2014, 75, 252-264.
[26] Vasudev Dehalwar; Akhtar Kalam; Mohan LalKolhe; et.al. Electricity load forecasting for urban area using weather forecast information. IEEE International Conference on Power and Renewable Energy, Oct2016; pp. 21-23.