IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 5, 2021, accepted November 17, 2021, date of publication November 29, 2021, date of current version December 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3131237

A New Medium and Long-Term Power Load Forecasting Method Considering Policy Factors

BO ZHANG[1,2], XIAOHAN ZHAO[3], ZHENHAI DOU[1], AND LIANXIN LIU[1]
[1]School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, Shandong 255049, China
[2]State Grid Juancheng Power Supply Company, Heze, Shandong 274600, China
[3]School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China

Corresponding author: Zhenhai Dou (douzhenhai@sdut.edu.cn)

This work was supported in part by the Youth-Project of Shandong Natural Science Foundation, Research on optimal control algorithm of capacitor voltage balancing for hybrid MMC with a low DC-1ink voltage, under Grant ZR2020QE215; in part by the Funding-Project of Shandong Natural Science Foundation, under Grant ZR2017LEE022; and in part by the Key Research and Development Program of Zibo, Shandong, China, Development of multi-objective optimization for optimal dynamic controller of transformer, under Grant 2019ZBXC498.

ABSTRACT At present, China's power load development is facing a new situation in which policies such as the new economic norm, industrial structure adjustment, energy conservation and emission reduction, etc. are being deeply promoted, load growth in some areas begin to ease, and volatility of load gradually become prominent, which increases the difficulty of medium and long-term load forecasting. In this context, in view of multi-correlation, uncertainty of influence of policy factors on power load, in order to improve the accuracy of load forecasting under the influence of policy factors, and solve the problem that policy factors are ambiguous, difficult to be quantified, and difficult to be integrated into load forecasting model, a medium and long-term load forecasting model considering policy factors is proposed. First, by analyzing the influence of various policies on power load, a hierarchical policy influencing factor index system that combines macro and micro levels is constructed to systematically reflect the influence of economy and policies on load under the new situation. Then, in view of the traditional grey relational analysis model's insufficient consideration of the difference of historical data and future power development situation, by respectively weighting historical periods and factor indexes, a quantification analysis model of power load influencing factors based on two-way weighted grey relational analysis is proposed to quantify the influence of various policy factors on power load, achieve the combination of subjective weighting and objective weighting,and obtain quantification weights. Finally, the weighted fuzzy cluster analysis method combined with weights is used to predict load under the influence of policy factors. The proposed model can better solve the difficulty of medium and long-term load forecasting caused by the volatility of load under the influence of policy factors, and is suitable for medium and long-term load forecasting under the background of policy changes. The analysis of calculation examples shows that compared with traditional grey relational analysis model, the quantification results of proposed methods are more realistic, compared with traditional prediction methods such as time series extrapolation and elasticity coefficient, proposed method has better prediction accuracy and engineering application value.

INDEX TERMS Fuzzy cluster analysis, grey relational analysis, load forecasting, policy factors.

I. INTRODUCTION

As China's economy has entered the new norm, various industries and environmental protection policies have been further promoted, power load has gradually stepped out of the "unilateral rise" rapid growth model, and has begun to show a trend of slowing growth and fluctuation characteristics. The change in load growth model increases the difficulty of medium and long-term load forecasting, the factors that affect load change and the effects also changed in different degrees. The reason for this phenomenon, in the final analysis, is that policy changes in economy, industry, environmental protection and other aspects have a profound impact on power load [1]. However, policy has a strong uncertainty, affecting

The associate editor coordinating the review of this manuscript and approving it for publication was Payman Dehghanian.

many factors, the relationship between each other and its influence mechanism on power load is complex. Therefore, comprehensively and systematically sorting out relevant factors that affect power load under policy factors, analyzing its influence mechanism on power load, and establishing a correlation model for prediction are of great significance for guiding the power planning and load forecasting under the new situation [2].

In the past years when power transmission, distribution and sales were integrated, operators could obtain the change law of load characteristics by qualitative analysis of experience, and the prediction accuracy was feasible [3]. However, in the context of power load facing multiple policy influences, the law of load changes is becoming more and more complex [4]. In order to improve the accuracy of medium and long-term load forecasting, it is necessary to systematically sort out the relevant factors affecting power load under policy factors, combine the analysis of past data and research and judgment of future trend, and scientifically grasp the internal law of load changes [5]. In recent years, forecasting models with comprehensive factors and more refined modeling have gradually emerged [6], [7], but most of them are difficult to adapt to the forecasting accuracy required by power grid under the background of policy influence [8]. In the future, the requirements for safety, stability and economic operation of power grid will rise to a higher level [9], [10]. Therefore, it is necessary to propose a medium and long-term power load forecasting method that considers policy factors. It is difficult to ensure the accuracy of analysis and forecasting by relying solely on qualitative analysis methods. Quantitative analysis methods such as data analysis and mathematical modeling must be combined.

The medium and long-term load forecasting is developing rapidly, and the forecasting effect is being improved continuously [11], but there is still room for improvement in the following four aspects. In terms of analytical thinking, the time series extrapolation method [12], [13] is mostly purely using historical data to quantify and analyze the factors that influence load to achieve extrapolation forecasting. It is more suitable for load sequence with strong regularity and relatively stable, but for load sequence with strong volatility under policy factors, the forecasting effect is not good. It is difficult to well reveal the internal factors that influence load, and the differences of historical data and the future power development situation are not sufficiently considered. Single-factor correlation analysis methods [14], such as elastic coefficient method, output value unit consumption method, etc., predict by grasping the correlation between load and main influencing factors. Because of simple operation and good effect of this method, it is widely used in practical work. However, this type of method only considers a single factor, and it is gradually difficult to adapt to the diversified characteristics of factors that influence power load in the process of policy changes. When selecting load influencing factors, multi-factor correlation analysis methods [15], such as multiple linear regression, partial least squares regression, etc., mostly use micro indexes directly such as GDP, CPI, population, etc., and relatively lack the systematic and hierarchical index system from the macro perspective. Artificial intelligence methods [16]–[19] such as artificial neural networks and support vector machines establish the nonlinear relationship between input and output through the training of a large number of samples, which are more suitable for the prediction of sequences with strong volatility [20], [21], but because the sample size is often small in medium and long-term load forecasting, it is difficult to apply this type of method in practice. At present, some scholars have done research on the influence of policies on load. The Ref. [22] combines the product value of various industries, GDP and other economic indicators to predict, which reflects the impact of policy factors on load to a certain extent, but policy adjustment is a process of comprehensive changes in the economy, industry, and environmental protection. The policy contains many factors, and each factor has different effects on power load. Therefore, it is still a little one-sided to study the change rule of power load change only from the perspective of economy; The Ref. [23] uses AHP to realize the quantification of urbanization factors, but because it is a qualitative analysis method, the quantification results are too subjective and deviate too much from the actual situation; In Ref. [24], the grey relational analysis is weighted by analytic hierarchy process (AHP), and an improved grey relational analysis model is proposed to quantify policy factors. The subjective weighting of AHP and the objective weighting of grey relational analysis are combined to realize. Compared with qualitative analysis, quantitative analysis is more reliable and accurate. However, the opinions of many experts are not considered and only one expert's judgment is adopted. The reference opinions are too one-sided, and the quantification results are not used in load forecasting. Ref. [25] uses the method of grey relational analysis combined with fuzzy clustering, and the correlation coefficients obtained by grey relational analysis are used to optimize variables and parameters for data preprocessing. Because fuzzy clustering is not improve, The prediction effect of fuzzy clustering is not changed. In this article, fuzzy clustering is improved to a weighted fuzzy clustering model, which not only can consider multiple related factors at the same time, but also carries quantification weights of grey relational analysis, takes into account the different influences of different factors on load. Clustering effect and prediction accuracy are improved.

This paper explores the medium and long-term load forecasting technology under the influence of policy. There are two major research difficulties: ① quantification of policy factors; ② application of quantification results to load forecasting. In the existing research, there are few literatures on the influence of policy factors on load, and there are fewer literatures on load forecasting under the influence of policy factors. The focus of research is on one of above two points, and the comprehensive research on the two is relatively less. This article solves these two difficulties, and the method proposed has strong theoretical basis and engineering application value.

In order to quantify the influence of policy factors on load, and solve the problem of load forecasting under the influence of policy factors, this paper proposes a medium and long-term power load forecasting model based on improved grey relational analysis and fuzzy clustering. First, a hierarchical policy factor index system that combines macro and micro levels is constructed. Secondly, in view of the problem of insufficient consideration of difference of historical data and future development situation of traditional grey relational analysis model, a quantification analysis model of power load influencing factors based on two-way weighted grey relational analysis is proposed. On the one hand, analytic hierarchy process method is used to weight the correlation coefficients of each period in accordance with the principle of ''near large, far small'' in historical time; On the other hand, principal component analysis method is used to determine the importance weights of indexes, quantitatively represents the research and judgment of multiple experts and operators on the future situation, and the model is used to subjectively weight the influence of each index on power load. Finally, fuzzy cluster analysis method combined with weights is used to predict through clustering. This method systematically considers the influence of policy factors on load, quantifies it, and on this basis realizes medium and long-term load forecasting considering policy factors. Calculation examples show that this method proposed by this paper has high prediction accuracy and engineering application value.

II. CONSTRUCTION OF POWER LOAD INFLUENCING FACTOR INDEX SYSTEM UNDER THE BACKGROUND OF POLICY INFLUENCE

Policy factors have a great impact on load forecasting, especially medium and long-term load forecasting. The three policy factors of economy, industry and environmental protection make load factors diversified, and the volatility of overall load sequence increase. However, at present, when considering policy factors, the empirical value is often used as the weight for adjustment, there is no specific quantitative mechanism, and the contribution to prediction accuracy is also very small. The research on this problem is as follows.

A. ANALYSIS OF THE INFLUENCE OF POLICY FACTORS ON POWER LOAD

This paper analyzes the influence of policy factors on power load from three aspects: economic policy, industrial policy and environmental protection policy [26].

1) ECONOMIC POLICY

The level of economic development largely determines the level of electricity consumption in the whole society. Economic development is the internal driving force of electric power development. The operating conditions of the national economy and other economic and non-economic factors directly or indirectly influence the power industry, and have a great impact on production and sales of power industry. Load level is an important index that reflects electric power development and economic growth. Therefore, economic development will drive the development of load level, and economic fluctuations will also lead to the change of load level.

The influence of new economic norm on power load is mainly manifested in two aspects: First, the slowdown of economic growth directly leads to the slowdown of electricity consumption demand growth; Second, the optimization and upgrading of economic structure brings new opportunities and challenges to the development of electric power.

2) INDUSTRIAL POLICY

The adjustment of industrial structure is an important factor affecting load changes. For a long time, in the central and western regions of China, the industrial system led by energy-intensive industries has been the source of regional economic development. The electricity consumption proportion of primary and tertiary industries and urban and rural residents in the grid load is relatively low, and load level mainly depends on the electricity consumption of secondary industry.

With the optimization and adjustment of industrial structure, the power consumption structure is gradually moving from the development stage of heavy and chemical industry to the development stage of ''the new norm''. The electricity consumption of tertiary industry and the electricity consumption of urban and rural residents are gradually becoming the main driving force for the growth of regional electricity consumption. Meanwhile, the load characteristics of power grid will be greatly changed, and the annual load changes will gradually fluctuate.

3) ENVIRONMENTAL PROTECTION POLICY

In recent years, as government has intensified the implementation of energy conservation and emission reduction policies, energy-intensive industrial structure ''based on coal'' and ''based on electricity'' has not only brought rapid economic growth, but also caused great pressure on energy conservation and emission reduction. The country's low-carbon economic policy requires to focus on the development of environmentally friendly, green and efficient industries, actively promote electric energy substitution, clean substitution, limit and phase out industries with backward production capacity and low value-added, and such industries are generally energy-intensive industries. Low-carbon economic policies will force these companies to undertake large-scale technological transformation and capacity upgrading, which will reduce their electricity consumption and slow down growth rate of electricity consumption.

B. CONSTRUCTION OF POLICY FACTOR INDEX SYSTEM

There are many factors that affect the level of power load, and the time span of medium and long-term load forecasting is long, which is a complex nonlinear problem [27], and

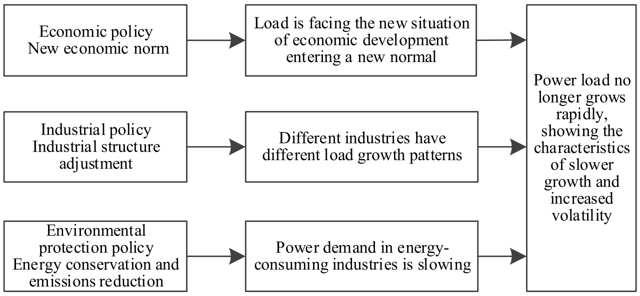

FIGURE 1. The influence of policy factors on load.

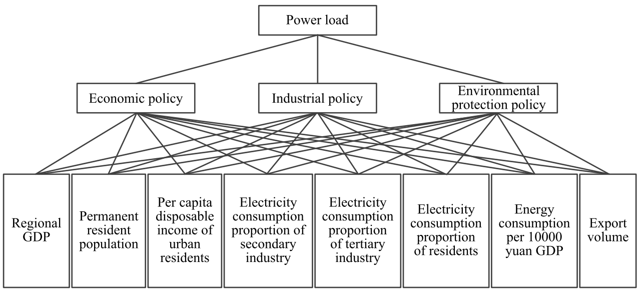

FIGURE 2. Ladder hierarchy structure of power load and various policy influencing factors.

the influence of economic changes, industrial adjustments, energy conservation and emission reduction policies on load is difficult to measure. In order to determine the quantitative relationship between medium and long-term load and policy factors, it is necessary to determine quantification indexes of load influencing factors, and conduct quantitative research on the relationship between power load and policy factors.

According to the ladder hierarchy structure of analytic hierarchy process (AHP) [28], this paper summarizes policy factors that influence load into three macro levels: economic policy, industrial policy, and environmental protection policy. 8 micro indexes, including regional GDP, permanent resident population, per capita disposable income of urban residents, electricity consumption proportion of secondary industry, electricity consumption proportion of tertiary industry, electricity consumption proportion of residents, energy consumption per 10000 yuan GDP, and export volume, are selected to establish an policy influencing factors index system. This index system covers three aspects: the new economic norm, industrial structure adjustment, and energy conservation and emission reduction, which can more comprehensively reflect the main aspects of impact of policy on power load. The ladder hierarchy model constructed is shown in Fig. 2.

III. ESTABLISHMENT OF IMPROVED GREY RELATIONAL ANALYSIS MODEL

Grey relational analysis is a quantitative description and comparison method for the development and change of a system. Its basic idea is to judge whether the relationship is close by determining the geometrical similarity between reference data column and several comparison data columns, which reflects the degree of correlation between curves. This method can usually be used to analyze the degree of influence of various factors on the results. The core is to establish the mother sequence that changes with time according to certain rules, take the change of each evaluation object with time as the sub-sequence, calculate the degree of correlation between each sub-sequence and mother sequence, and draw a conclusion according to the correlation size [29].

A. TRADITIONAL GREY RELATIONAL ANALYSIS MODEL

1) FORM ANALYSIS MATRIX

After selecting the influencing factor index of load, the analysis matrix can be formed:

$$(\boldsymbol{Y},\boldsymbol{X})=\begin{bmatrix} y(1) & x_1(1) & \cdots & x_i(1) & \cdots & x_n(1)\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\ y(t) & x_1(t) & \cdots & x_i(t) & \cdots & x_n(t)\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\ y(t) & x_1(T) & \cdots & x_i(T) & \cdots & x_n(T)\end{bmatrix} \tag{1}$$

where $t(t = 1, 2, \ldots, T)$ is the dimension of load sequence or each influencing factor index sequence; $\boldsymbol{Y}$ is load sequence, $y(t)$ is load data at time t; $\boldsymbol{X}$ is influencing factor index sequence, $x_i(t)$ is data of the i-th $(i = 1, 2, \ldots, n)$ influencing factor index at time t; n is the total number of influencing factor indexes.

2) GENERATE INITIAL VALUE MATRIX

Sequences are normalized according to (2)-(3), and the initial value mirror matrix is formed according to (4):

$$x_i' = \frac{x_i(t)}{x_i(1)} \quad (i=1,2,\cdots,n;\ t=1,2,\cdots,T) \tag{2}$$

$$y' = \frac{y(t)}{y(1)} \quad (t=1,2,\cdots,T) \tag{3}$$

$$(\boldsymbol{Y}',\boldsymbol{X}')=\begin{bmatrix} y'(1) & x_1'(1) & \cdots & x_i'(1) & \cdots & x_n'(1)\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\ y'(t) & x_1'(t) & \cdots & x_i'(t) & \cdots & x_n'(t)\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\ y'(t) & x_1'(T) & \cdots & x_i'(T) & \cdots & x_n'(T)\end{bmatrix} \tag{4}$$

3) GENERATE DIFFERENCE VALUE MATRIX

According to (5), the difference value calculation is performed on the elements in the initial mirror matrix, and the difference value between load and influencing factor index i is calculated, and the sequence matrix $\boldsymbol{\Delta}_i(t)$ as shown below is obtained.

$$\Delta_i(t) = \left|y'(t) - x_i'(t)\right| \ (i=1,2,\cdots,n) \tag{5}$$

$$\boldsymbol{\Delta}_i(t) = \begin{bmatrix} \Delta_1(1) & \cdots & \Delta_i(1) & \cdots & \Delta_n(1) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \Delta_1(t) & \cdots & \Delta_i(t) & \cdots & \Delta_n(t) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \Delta_1(T) & \cdots & \Delta_i(T) & \cdots & \Delta_n(T) \end{bmatrix} \quad (6)$$

The maximum $\Delta_{\max}$ and minimum $\Delta_{\min}$ of the two-level range are selected.

$$\Delta_{\max} = \max(\max \Delta_i(t)), \quad \Delta_{\min} = \min(\min \Delta_i(t)) \quad (7)$$

4) CALCULATE GREY RELATIONAL DEGREE

The correlation coefficient $\lambda_i(t)$ of load and the influencing factor index i at time t is calculated by (8), and their correlation coefficient matrix $\boldsymbol{\lambda}_i(t)$ is formed, as shown in (9) $(i = 1, 2, \ldots, n; t = 1, 2, \ldots, T)$.

$$\lambda_i(t) = \frac{\Delta_{\min} + \rho \Delta_{\max}}{\Delta_i(t) + \rho \Delta_{\max}} \quad (8)$$

$$\boldsymbol{\lambda}_i(t) = \begin{bmatrix} \lambda_1(1) & \cdots & \lambda_i(1) & \cdots & \lambda_n(1) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \lambda_1(t) & \cdots & \lambda_i(t) & \cdots & \lambda_n(t) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \lambda_1(T) & \cdots & \lambda_i(T) & \cdots & \lambda_n(T) \end{bmatrix} \quad (9)$$

where ρ is the identification coefficient, usually take 0.5.

By calculating the mean value of each column in $\boldsymbol{\lambda}_i(t)$ matrix, grey relational degree r_i between influencing factor index i and load can be obtained.

$$r_i = \frac{1}{T} \sum_{t=1}^{T} \lambda_i(t) \quad (i = 1, 2, \cdots, n; t = 1, 2, \cdots, T) \quad (10)$$

B. PROBLEMS OF TRADITIONAL GREY RELATIONAL ANALYSIS MODEL

In the traditional grey relational analysis model, grey relational degree is obtained by averaging the correlation coefficients of each historical period. In fact, the degree of influence of historical data on current situation is inconsistent. Generally, under the same conditions, the closer the historical time of the data, the higher the impact on the current situation. For example, as China's economy has entered the new norm in recent years, the data in recent years is obviously more important than the previous data.

In addition, traditional calculation formula of relational degree adopts equal weight treatment for each influencing factor, which has two problems: First, the average value tends to conceal the personality of each influencing factor, fails to consider the experience or opinions of experts, and is difficult to deal with changes of future situation; Second, when correlation coefficient is relatively discrete, overall relational degree will be determined by the point with a large correlation coefficient, resulting in local correlation tendency and deviation of analysis results. Therefore, in order to make up for the shortcomings of traditional grey relational analysis model, longitudinal AHP weighting method and transverse PCA weighting method are proposed to carry out two-way weighting for grey relational analysis in this paper.

TABLE 1. Reference table for a_{ij} value.

a_{ij}	Qualitative description	a_{ji}
1	x_i and x_j are equally important	1
3	x_i is slightly more important than x_j	1/3
5	x_i is obviously more important than x_j	1/5
7	x_i is more important than x_j	1/7
9	x_i is extremely important than x_j	1/9

C. TWO-WAY WEIGHTED GREY RELATIONAL ANALYSIS MODEL

In this paper, grey relational analysis will be two-way weighted from two dimensions of historical period and index importance, so as to make up for shortcomings of grey relational analysis.

1) LONGITUDINAL AHP WEIGHTING METHOD

AHP (Analytical Hierarchy Process) is a subjective weighting method. It judges the relative importance between two different factors by using people's experience, and finally calculates the combination weight to realize the conversion from qualitative to quantitative. In this section, AHP is used to combine the experience of experts to weight the importance of each historical period [30].

Formula (10) is improved to formula (11) by longitudinal weighting. The calculation formula of longitudinal weighted relational degree r_i' between influencing factor index i and load at different time t is

$$r_i' = \sum_{i=1}^{T} \omega(t) \lambda_i(t) \quad (i=1, 2, \cdots, n; t=1, 2, \cdots, T) \quad (11)$$

where $\omega(t)$ is longitudinal weight, and represents the weight of correlation coefficient $\lambda_i(t)$ between influencing factor index i and load at time t.

AHP is used to calculate longitudinal weight $\omega(t)$, the steps are as follows:

(1) Determine program objectives and establish evaluation model.

(2) Construct pairwise comparison judgment matrix. Judge the relative importance of each factor at the same level. a_{ij} in the judgment matrix $\boldsymbol{A} = (a_{ij})_{T \times T}$ represents the importance of the i-th factor x_i relative to the j-th factor x_j in the level. Refer to Table 1 for the value of a_{ij}.

When value of a_{ij} is between the above values, the corresponding meaning of a_{ij} is also between different qualitative levels, so judgment matrix $\boldsymbol{A} = (a_{ij})_{T \times T}$ can be obtained. In order to facilitate the determination of judgment matrix of the principle of "near big, far small" in historical time, the empirical formula of importance of each historical period is obtained through a lot of experiments:

$$f(t) = 2t + T - 3 \quad (t = 1, 2, \cdots, T) \tag{12}$$

where T is total historical time, t is historical time. Experiments show that the historical weights obtained by this formula have strong reliability, which greatly facilitates the research on determining historical weights of "near large, far small".

(3) Calculate the maximum eigenvalue λ_{max} of judgment matrix $\boldsymbol{A}$ and its corresponding eigenvector ω, and ω' after normalizing ω is the weight of the level factor.

(4) Consistency check. First, the consistency index CI is calculated, CI $= (\lambda_{max}\text{-}\mu)/(\mu - 1)$; Secondly, the mean random consistency index RI is determined according to the order μ. Finally, the consistency ratio CR = CI/RI is calculated. If CR < 0.10, the consistency of judgment matrix is considered acceptable, otherwise judgment matrix should be appropriately modified.

In this way, the closer the historical data is to the present, the greater the longitudinal weight will be assigned, so as to highlight its importance.

2) TRANSVERSE PCA WEIGHTING METHOD

PCA (principal component analysis) transforms related variables into a number of unrelated comprehensive index variables through variable transformation methods, so as to reduce the dimensionality of data and simplify problem. At present, most of the applied research on PCA focuses on data simplification or comprehensive evaluation. This paper uses principal component analysis method combined with experts' experience to subjectively weight the influence of policy factors on power load [31].

On the basis of (11), formula (10) is further improved to formula (13) by transverse weighting. The calculation formula of two-way weighted relational degree R_i between influencing factor index i and load is

$$R_i = \sigma_i \sum_{i=1}^{T} \omega(t)\lambda_i(t) \quad (i = 1, 2, \cdots, n; t = 1, 2, \cdots, T) \tag{13}$$

where σ_i is transverse weight, indicating the importance of influencing factor index i relative to load.

PCA is used to calculate transverse weight σ_i, and the steps are as follows:

a: PROPOSE HYPOTHESIS

Supposing that the number of indexes that need to be weighted is n. Now v experts are consulted respectively to obtain n groups of weight scores, among which there are v elements in each group. The specific form can be shown in Table 2.

Due to the different research directions of experts, their scoring also has a certain bias, which brings certain fuzziness to determination of weight. The study found that the more experts there are, the more scientific the weights obtained are, and at the same time, the more fuzzy determination of weights becomes. On this basis, the following hypothesis is proposed, that is, under the condition that the number of experts is constant, the linear relationship between the scores of each expert is used to similarly simplify the number of experts actually participating in the scoring, so as to achieve the accuracy of weight evaluation. After analysis, the idea accords with the basic principle of PCA, so the method of PCA can be tried to determine weights.

TABLE 2. Experts scoring table.

Index \ Expert	h_1	h_2	…	h_v
x_1	d_{11}	d_{12}	…	d_{1v}
x_2	d_{21}	d_{22}	…	d_{2v}
⋮	⋮	⋮	…	⋮
x_n	d_{n1}	d_{n2}	…	d_{nv}

b: THE PROCESS OF WEIGHT DETERMINATION

The average value and variance of each column elements in experts scoring table $\boldsymbol{D} = (d_{ij})_{n\times v}$ are calculated respectively, and the original data are transformed into the data that obey the standard normal distribution by (14).

$$\bar{d}_j = \frac{1}{n}\sum_{i=1}^{n} d_{ij}, var(d_j) = \frac{1}{n}\sum_{i=1}^{n}(d_{ij} - \bar{d}_j) \quad (j = 1, 2, \cdots, v)$$

$$z_{ij} = \frac{d_{ij} - \bar{d}_j}{\sqrt{var(d_j)}} \quad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, v) \tag{14}$$

where $\bar{d}_j$ is average value of elements in the j-th column of matrix $\boldsymbol{D}$; $var(d_j)$ is variance of elements in the j-th column of matrix $\boldsymbol{D}$. From (14), the normalized matrix $\boldsymbol{Z} = (z_{ij})_{n\times v}$ of matrix $\boldsymbol{D}$ can be obtained, and the correlation coefficient between each column and other columns in matrix $\boldsymbol{Z}$ is calculated by (15).

$$\varepsilon_{ij} = \frac{\sum_{k=1}^{n}(z_{ki} - \bar{z}_i)(z_{kj} - \bar{z}_j)}{\sqrt{\sum_{k=1}^{n}(z_{ki} - \bar{z}_i)^2 \sum_{k=1}^{n}(z_{kj} - \bar{z}_j)^2}} \tag{15}$$

where ε_{ij} is the correlation coefficient. The correlation coefficient matrix $[\varepsilon_{ij}]_{v\times v}$ can be calculated by (15), and the eigenvalues of correlation coefficient matrix can be calculated to obtain v eigenvalues $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_v \geq 0$. Then the cumulative contribution rate is calculated according to (16).

When $\eta \geq 70\%$, the first s principal components are retained.

$$\eta_i = \frac{\gamma_1 + \gamma_2 + \cdots + \gamma_i}{\sum_{j=1}^{v}\gamma_j} \tag{16}$$

According to above conditions, the process of determining weights is the process of obtaining comprehensive evaluation

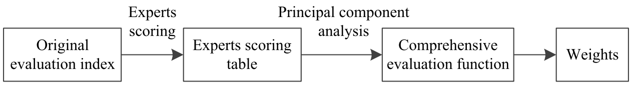

FIGURE 3. Flow chart of PCA weight determination.

function by PCA. In this process, the indexes in the original evaluation system become samples, and the existing indexes are experts. The specific process of determining weights can be shown in Fig. 3.

c: WEIGHT MODEL

The first determined primary weight model is the principal component model.

$$\begin{cases} F_1 = u_{11}h_1 + u_{21}h_2 + \cdots + u_{v1}h_v \\ F_2 = u_{12}h_1 + u_{22}h_2 + \cdots + u_{v2}h_v \\ \vdots \\ F_s = u_{1s}h_1 + u_{2s}h_2 + \cdots + u_{vs}h_v \end{cases} \tag{17}$$

where $F_1, F_2, \ldots, F_s$ are the s principal components obtained after analysis; u_{ij} is the coefficient in decision matrix.

On this basis, the comprehensive evaluation function is constructed:

$$F_z = \sum_{j=1}^{s} (\gamma_j/g)F_j = p_1h_1 + p_2h_2 + \cdots + p_vh_v$$
$$V_{zi} = \sum_{j=1}^{v} p_jd_{ij} \quad (i = 1, 2, \cdots, n) \tag{18}$$

where $p_1, p_2, \ldots, p_v$ are the comprehensive importance of indexes $h_1, h_2, \ldots, h_v$ in principal components. On this basis, combined with the actual scoring of experts, the comprehensive value of original index score can be calculated.

$$V_{zi} = \sum_{j=1}^{v} p_jd_{ij} \quad (i = 1, 2, \cdots, n) \tag{19}$$

The transverse weight of each index is

$$\sigma_i = \frac{V_{zi}}{\sum_{i=1}^{n} V_{zi}} \tag{20}$$

In this way, the research and judgment of experts and operators on future load influencing factors can be quantitatively reflected through transverse weights.

Finally, the final weight value of each influencing factor index relative to power load is calculated by (21).

$$w_i = \frac{R_i}{\sum_{i=1}^{n} R_i} \quad (i = 1, 2, \cdots, n) \tag{21}$$

IV. MEDIUM AND LONG-TERM LOAD FORECASTING BASED ON WEIGHTED FUZZY CLUSTER ANALYSIS

In order to solve the problem of load forecasting under the influence of policy factors, the established forecasting model must be able to consider multiple related factors at the same time, because the influence of policy on load has multi-correlation, and must consider the different influence of different factors on load, and be able to carry the quantification weight, which greatly increases the difficulty of load forecasting modeling.

Fuzzy cluster analysis is a clustering technology based on fuzzy theory. It realizes clustering by determining the similarity between different objects. It reflects the intermediary of samples and can better reflect the "is this and also is that" characteristics of objects in real world. The main idea of fuzzy cluster analysis, which is applied to medium and long-term load forecasting, is to cluster the known year samples together with the predicted year samples. In the same category, the load growth rate of known years is used as the load growth rate of predicted years, so as to realize load forecasting. Fuzzy cluster analysis method can consider many related factors at the same time, which is consistent with the characteristics of policy influencing many related factors [32]. However, fuzzy clustering cannot carry quantification weights, and the influence of different factors on load is not considered, so it is improved to weighted fuzzy clustering on the basis of fuzzy clustering.

(1) Standardize original data

Supposing that the number of original data samples is m and the number of influencing factors is n, then the original data matrix is $\boldsymbol{X} = (x_{ij})_{m\times n}$. Because the dimensions and sizes of different factors are different, it is necessary to standardize the data. First, the mean value and variance of each dimension factor are calculated according to (22):

$$\bar{x}_k = \frac{1}{m}\sum_{i=1}^{m} x_{ik}, \quad S_k^2 = \frac{1}{m}\sum_{i=1}^{m}(x_{ik} - \bar{x}_k) \tag{22}$$

Then, the data are preliminarily standardized according to (23):

$$x'_{ik} = \frac{x_{ik} - \bar{x}_k}{S_k} \tag{23}$$

In order to further compress the data into the interval [0,1], the extremum standardization formula is used:

$$x''_{ik} = \frac{x'_{ik} - x'_{k\,\min}}{x'_{k\,\max} - x'_{k\,\min}} \tag{24}$$

(2) Determine similarity and establish fuzzy similarity relation matrix.

Considering the different influences of different factors on load, the scalar product method of traditional fuzzy cluster analysis is weighted and improved into weighted fuzzy cluster analysis that is (25), combining weight of each main component, and calculating the similarity

q_{ij} between samples.

$$q_{ij} = \frac{\sum_{k=1}^{n} w_k x''_{ik} w_k x''_{jk}}{\sqrt{\sum_{k=1}^{n} (w_k x''_{ik})^2 \sum_{k=1}^{n} (w_k x''_{jk})^2}} \tag{25}$$

where w_k is the weight of the k-th influencing factor index, and $\sum_{k=1}^{n} w_k = 1$.

Based on this, the fuzzy similarity relation matrix is established as $\boldsymbol{Q}_m = (q_{ij})_{m\times m}$.

(3) Calculate the transitive closure $\tilde{\boldsymbol{Q}}_m$ of similarity relation matrix $\boldsymbol{Q}_m$ and obtain the dynamic clustering graph.

Transitive closure is the minimum transitive relation of binary relations defined on a set.

In order to obtain transitive closure of $\boldsymbol{Q}_m$, the composite operation is performed according to (26).

$$\boldsymbol{Q}_m^2 = \boldsymbol{Q}_m \circ \boldsymbol{Q}_m, \quad \boldsymbol{Q}_m^4 = \boldsymbol{Q}_m^2 \circ \boldsymbol{Q}_m^2 \tag{26}$$

The specific composite operation is performed according to (27):

$$\boldsymbol{E}_{m\times n} = \boldsymbol{B}_{m\times\varphi} \circ \boldsymbol{C}_{\varphi\times n}, \quad e_{ij} = \vee \left\{ b_{ik} \wedge c_{kj} | 1 \le k \le \varphi \right\} \tag{27}$$

where e_{ij} is element of $\boldsymbol{E}_{m\times n}$; $\vee$ and $\wedge$ denote the large operation and the small operation respectively. In this way, there must be a positive integer l, such that $\boldsymbol{Q}_m^{2l} = \boldsymbol{Q}_m^l \circ \boldsymbol{Q}_m^l$. In this case, $\tilde{\boldsymbol{Q}}_m = Q_m^l$ is the transitive closure of $\boldsymbol{Q}_m$, which is a fuzzy equivalence relation.

V. THE IDEAS AND FRAMEWORK OF MODELING

This article analyzes main influence factors of policy on power load from three aspects: economy, industry and environmental protection. Based on selected indexes, complex medium and long-term load forecasting work under influence of policy factors is carried out. The work is divided into two parts: quantification and forecasting.

First, on the one hand, AHP and the proposed empirical formula are used to weight correlation coefficients of each historical period to obtain longitudinal weights. On the other hand, PCA is used to determine the importance weight of each index, that is transverse weights, which quantitatively represents the research and judgment of experts and operators on the future situation. Through the above process, historical period and factor indexes are respectively weighted. Based on traditional grey relational analysis model, a quantification analysis model of power load influencing factors based on two-way weighted grey relational analysis is proposed to quantify the influence of policy factors on power load. Then, basis on fuzzy clustering model, weighted fuzzy clustering model is proposed, which not only can consider multiple related factors at the same time, but also carries quantification weights, takes into account the different influences of different factors on load. Known year and predicted year are clustered, and the predicted values are obtained by further processing the clustering results. The flow chart of modeling is shown in Fig. 4.

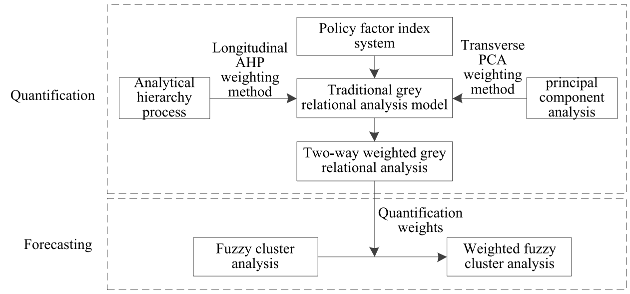

FIGURE 4. Flow chart of modeling.

1. The policy factors that affect load are summarized into three macro levels: economic development, industrial structure adjustment, and energy conservation and emission reduction. Eight micro policy factor indexes are selected to construct a hierarchical policy factor index system that combines macro and micro levels.

2. Longitudinal AHP weighting method combined with the proposed empirical formula is used to weight correlation coefficients of each historical period to obtain longitudinal weights in accordance with the principle of "near large, far small" in historical time.

3. Transverse PCA weighting method combined with many experts scoring is used to determine the importance weight of each index, that is transverse weights.

4. Formula (10) of traditional grey relational analysis is improved into (13) by longitudinal weighting and transverse weighting. Combined with longitudinal weights and transverse weights obtained in step 2 and step 3, the weight of influence of each influencing factor on power load is calculated, that is two-way weighted grey correlation degree, so as to realize the quantification of the influence of policy factors on power load.

5. Scalar product method of traditional fuzzy cluster analysis is weighted and improved into (25). Combined with final weights quantified in step 4, clustering prediction is made, so as to solve the load forecasting problem under the influence of policy factors.

VI. EXAMPLE ANALYSIS

Taking a province in China as an example, this paper collects electricity consumption data of whole society of the province and data of various load influencing factor indexes from 2008 to 2020, as shown in Table 3. Taking 2007 as the basic year, the load growth rate from 2018 to 2020 is predicted.

This article has two main research contents: quantification and forecasting. The original data in Table 3 is divided into four parts: red, yellow, blue, and green. The red part is the predicted data of each index in predicted years, the yellow part is the actual data of each index in known years, the green part is the actual annual electricity consumption in known years,

TABLE 3. Electricity consumption data of whole society and data of various load influencing factor indexes.

Year	y Electricity consumption of whole society/GWh	x_1 Regional GDP /billion RMB	x_2 Permanent resident population/ million people	x_3 Per capita disposable income of urban residents/RMB	x_4 Electricity consumption proportion of secondary industry/%	x_5 Electricity consumption proportion of tertiary industry/%	x_6 Electricity consumption proportion of residents/%	x_7 Energy consumption per 10000 yuan GDP/TCE	x_8 Export volume /billion USD
2008	37785	72.59	6.04	9177	91.6	2.8	2.9	3.806	0.9.45
2009	43978	91.911	6.10	10859	92.1	2.7	2.9	3.28	1.089
2010	43962	120.392	6.18	12932	91.2	3	3.3	2.644	1.259
2011	46296	135.331	6.25	14025	90.9	3.3	3.5	2.444	0.743
2012	54677	168.965	6.32	15344	91.6	3.6	2.7	2.18	1.17
2013	72453.9	210.221	6.39	17291	93.1	3.2	2	2.245	1.6
2014	74179.2	234.129	6.47	19507	92.7	3.5	2.3	2.091	1.641
2015	81118	257.757	6.54	21476	92.5	3.5	2.4	2.012	2.552
2016	84875	275.21	6.61	23285	92.1	3.6	2.6	1.933	4.303
2017	87833	291.177	6.68	25186	91.5	3.9	2.7	1.846	2.976
2018	88691	316.859	6.75	27153	90.79	4.36	2.87	1.767	2.485
2019	97830	345.393	6.82	29472	90.81	4.49	2.8	1.688	3.74
2020	101309	369.44	6.89	31872	90.62	4.49	2.91	1.604	3.06

and the blue part is the annual electricity consumption in predicted years. In the quantification part, two-way weighted grey relational analysis model is used to compare the columns of green part and yellow part vertically to get the weights of influence of policy on load; In the prediction part, weighted fuzzy cluster analysis model is used to compare the lines of yellow part and red part horizontally to realize clustering.

A. CALCULATION OF TRADITIONAL GREY RELATIONAL DEGREE

According to (1)-(9), the correlation coefficient matrix $\boldsymbol{\lambda}_i(t)$ obtained is

$$\begin{bmatrix} 1.000 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 \\ 0.919 & 0.882 & 0.983 & 0.879 & 0.853 & 0.876 & 0.792 & 0.990 \\ 0.700 & 0.892 & 0.824 & 0.875 & 0.882 & 0.978 & 0.711 & 0.872 \\ 0.644 & 0.858 & 0.792 & 0.832 & 0.811 & 0.984 & 0.664 & 0.724 \\ 0.567 & 0.742 & 0.837 & 0.721 & 0.646 & 0.691 & 0.569 & 0.847 \\ 0.541 & 0.573 & 0.972 & 0.561 & 0.584 & 0.484 & 0.465 & 0.837 \\ 0.478 & 0.564 & 0.877 & 0.548 & 0.563 & 0.496 & 0.449 & 0.836 \\ 0.451 & 0.520 & 0.856 & 0.504 & 0.537 & 0.467 & 0.416 & 0.676 \\ 0.427 & 0.500 & 0.799 & 0.482 & 0.526 & 0.461 & 0.399 & 0.333 \\ 0.406 & 0.486 & 0.733 & 0.465 & 0.453 & 0.453 & 0.385 & 0.583 \end{bmatrix}$$

B. CALCULATION OF TWO-WAY WEIGHTED GREY RELATIONAL DEGREE

1) CALCULATION OF LONGITUDINAL WEIGHTS

Longitudinal AHP weighting method is used. According to reference data in Table 1 or the empirical formula (12), judge relative importance of each factor in pairs, and establish the judgement matrix. This calculation example uses the empirical formula to obtain judgement matrix and its calculation results are shown in Table 4.

According to Table 4, the longitudinal weight $\omega(t)$ of correlation coefficient of each historical period is as follows: 0.0500, 0.0611, 0.0722, 0.0833, 0.0944, 0.1056, 0.1167, 0.1278, 0.1389, 0.1500 in order of time from far to near. longitudinal weight of each historical period is summarized and shown in Table 5.

2) CALCULATION OF TRANSVERSE WEIGHTS

Transverse PCA weighting method is used. Based on the experience of 6 experts, the influence of these 8 indexes on electricity consumption is scored subjectively according to Fig. 2 and transverse PCA weighting method, and the scoring table of each index obtained is shown in Table 6.

For Table 6, the correlation coefficient matrix is calculated and obtained according to (14)-(15).

$$\begin{bmatrix} 1.000 & 0.401 & 0.392 & 0.597 & 0.701 & 0.406 \\ 0.401 & 1.000 & 0.429 & 0.022 & 0.540 & 0.241 \\ 0.392 & 0.429 & 1.000 & 0.448 & 0.564 & 0.614 \\ 0.597 & 0.022 & 0.448 & 1.000 & 0.127 & 0.285 \\ 0.701 & 0.540 & 0.564 & 0.127 & 1.000 & 0.610 \\ 0.406 & 0.241 & 0.614 & 0.285 & 0.610 & 1.000 \end{bmatrix}$$

The eigenvalues of correlation coefficient matrix are: 0.06, 0.27, 0.58, 0.78, 1.11, and 3.19.

The eigenvectors corresponding to each eigenvalue are

$$\begin{bmatrix} 0.596 & 0.025 & 0.380 & 0.504 & -0.208 & 0.451 \\ -0.072 & -0.425 & -0.501 & 0.406 & 0.534 & 0.335 \\ 0.300 & 0.520 & -0.538 & -0.388 & -0.035 & 0.445 \\ -0.467 & -0.169 & -0.253 & 0.174 & -0.755 & 0.298 \\ -0.571 & 0.413 & 0.421 & 0.043 & 0.316 & 0.474 \\ 0.078 & -0.591 & 0.270 & -0.632 & 0.013 & 0.416 \end{bmatrix}$$

TABLE 4. Judgement matrix and calculation results.

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Eigenvector of the maximum eigenvalue
2008	1	9/11	9/13	9/15	9/17	9/19	9/21	9/23	9/25	9/27	0.0500
2009	11/9	1	11/13	11/15	11/17	11/19	11/21	11/23	22/25	11/27	0.0611
2010	13/9	13/11	1	13/15	13/17	13/19	13/21	13/23	13/25	13/27	0.0722
2011	15/9	15/11	15/13	1	15/17	15/19	15/21	15/23	15/25	15/27	0.0833
2012	17/9	17/11	17/13	17/15	1	17/19	17/21	17/23	17/25	17/27	0.0944
2013	19/9	19/11	19/13	19/15	19/17	1	19/21	19/23	19/25	19/27	0.1056
2014	21/9	21/11	21/13	21/15	21/17	21/19	1	21/23	21/25	21/27	0.1167
2015	23/9	23/11	23/13	23/15	23/17	23/19	23/21	1	23/25	23/27	0.1278
2016	25/9	25/11	25/13	25/15	25/17	25/19	25/21	25/23	1	25/27	0.1389
2017	27/9	27/11	27/13	27/15	27/17	27/19	27/21	27/23	27/25	1	0.1500

TABLE 5. Longitudinal weight of each historical period.

Year	2008	2009	2010	2011	2012
Longitudinal weight	0.0500	0.0611	0.0722	0.0833	0.0944
Year	2013	2014	2015	2016	2017
Longitudinal weight	0.1056	0.1167	0.1278	0.1389	0.1500

TABLE 6. Experts scoring table.

Index \ Expert	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6
x_1	5	3	5	4	5	5
x_2	3	4	3	3	3	4
x_3	3	3	2	3	3	2
x_4	5	4	4	3	4	3
x_5	4	2	3	5	2	3
x_6	2	3	4	3	3	3
x_7	1	2	3	2	2	3
x_8	3	2	1	2	3	3

In the table: 5-very important; 4- more important; 3- generally important; 2- less important; 1- not important

According to (16), the eigenvectors corresponding to eigenvalues 1.11 and 3.19 are retained, and the eigenvectors are substituted into the initial weight model (17), formula (28) can be obtained.

$$\begin{cases} F_1 = 0.451w_1 + 0.335w_2 + 0.445w_3 \\ \qquad +0.298w_4 + 0.474w_5 + 0.416w_6 \\ F_2 = 0.208w_1 - 0.534w_2 + 0.035w_3 \\ \qquad +0.755w_4 - 0.316w_5 - 0.013w_6 \end{cases} \tag{28}$$

Combining the above results and the scoring table, the secondary weight model and the transverse weight of each influencing factor are obtained according to (18)-(20), and the transverse weight of each factor is calculated as shown in Table 7.

TABLE 7. Transverse weight of each influencing factor.

Index	x_1	x_2	x_3	x_4
transverse weight	0.1842	0.1278	0.1048	0.1512
Index	x_5	x_6	x_7	x_8
transverse weight	0.1369	0.1177	0.0847	0.0927

$$F_z = 0.388w_1 + 0.111w_2 + 0.339w_3 \\ +0.416w_4 + 0.270w_5 + 0.305w_6$$

C. ANALYSIS OF QUANTIFICATION RESULTS

According to (10) and (13), traditional grey relational degree and two-way weighted grey relational degree between annual electricity consumption and each influencing factor can be obtained respectively, as shown in Table 8.

From Table 8, influencing factors of the province's annual electricity consumption are arranged according to traditional relational degree: per capita disposable income of urban residents > export volume > permanent resident population > electricity consumption proportion of residents > electricity consumption proportion of secondary industry > electricity consumption proportion of tertiary industry > regional GDP > energy consumption per 10000 yuan GDP. The conclusion drawn by traditional grey relational analysis model is that: Regional GDP, which is closely related to electricity consumption, ranks only seventh among load influencing factors; Electricity consumption of secondary industry accounts for more than 90% of total electricity consumption, and ranks only fifth, while uncontrollable population factor ranks third. It can be seen that although traditional grey relational analysis model can obtain the ranking and weight of influence degree of each load influencing factor through quantitative analysis, but it is somewhat inconsistent with actual situation.

Influencing factors of the province's annual electricity consumption are arranged according to two-way weighted grey

TABLE 8. Relational degree comparison.

Influencing factor	Regional GDP	Permanent resident population	Per capita disposable income of urban residents	Electricity consumption proportion of secondary industry
Traditional grey relational degree	0.613	0.702	0.867	0.687
Two-way weighted grey relational degree	0.102	0.082	0.089	0.095
Influencing factor	**Electricity consumption proportion of tertiary industry**	**Electricity consumption proportion of residents**	**Energy consumption per 10000 yuan GDP**	**Export volume**
Traditional grey relational degree	0.686	0.689	0.585	0.770
Two-way weighted grey relational degree	0.086	0.073	0.045	0.067

TABLE 9. Final weight of each influencing factor index on electricity consumption.

Index	Weight
x_1	0.1597
x_2	0.1289
x_3	0.1393
x_4	0.1485
x_5	0.1352
x_6	0.1145
x_7	0.0698
x_8	0.1043

relational degree: regional GDP > electricity consumption proportion of secondary industry > per capita disposable income of urban residents > electricity consumption proportion of tertiary industry > permanent resident population > electricity consumption proportion of residents > export volume > energy consumption per 10000 yuan GDP. Obviously, this ranking is more consistent with actual situation.

The final weight of each factor calculated by formula (21) is shown in Table 9.

D. LOAD FORECASTING

Taking 2007 as basic year, original data is treated as growth rate relative to the previous year and standardized. Combining the weighting results in Table 9, the correlation coefficient matrix and the transitive closure matrix of the fuzzy similarity relation are calculated as shown in Table 10 and Table 11 respectively.

The clustering level $\tau \in [0, 1]$ represents the confidence degree of the clustering results. The calculation formula of clustering level τ is shown in (29). The closer τ is to 1, the higher the similarity among the samples in the clustering result, and the better the clustering effect; On the contrary, the closer τ is to 0, the lower the similarity among the samples in the clustering result, and the worse the clustering effect. The clustering level τ is continuously changed from 1 to 0, and the two objects corresponding to elements larger than τ in fuzzy equivalence matrix are classified into one category. When $\tau = 1$, each object is divided into a category respectively. With the gradual decrease of τ, the classified categories gradually decrease, until all the predicted year samples are classified into the categories containing the known years, so that the dynamic clustering diagram is obtained, as shown in Fig. 5.

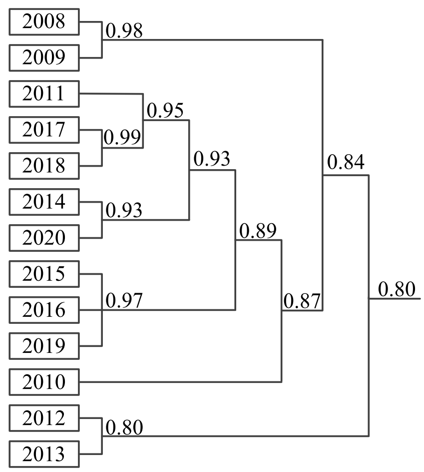

FIGURE 5. Dynamic clustering graph.

$$\tau_{i+1} = \tau_i - 0.01 \quad (\tau_1 = 1, \tau \in [0, 1]) \tag{29}$$

The numbers in Fig. 5 indicate that when the clustering level is reduced to this value, the corresponding objects on the left are clustered into one category. When a predicted year is first classified into the category containing known years, the load growth rate of these known years can be processed to obtain the load growth rate of the predicted year in the same category [33]. As can be seen from Fig. 5, when the clustering level is 0.97, the predicted year 2019 and the sample years 2015 and 2016 are classified into the same category. Therefore, the predicted value of 2019 is obtained by averaging the load growth rate in the two years of 2017 and 2016. Similarly, the load growth rate in 2018 is equal to that in 2017, and the predicted value of 2020 is obtained by

TABLE 10. Fuzzy correlation coefficient matrix.

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
2008	1.0000	0.9830	0.8432	0.6608	0.7495	0.7108	0.7256	0.7335	0.6051	0.5488	0.5848	0.6849	0.6522
2009	0.9830	1.0000	0.8097	0.5603	0.7476	0.7970	0.6495	0.6901	0.5355	0.4215	0.4592	0.6292	0.5626
2010	0.8432	0.8097	1.0000	0.7664	0.7286	0.5367	0.8722	0.7189	0.6196	0.6245	0.6651	0.6233	0.5969
2011	0.6608	0.5603	0.7664	1.0000	0.7117	0.4040	0.9267	0.7793	0.7670	0.9514	0.9509	0.7908	0.9027
2012	0.7495	0.7476	0.7286	0.7117	1.0000	0.8018	0.6957	0.7670	0.6852	0.5772	0.6379	0.8149	0.6279
2013	0.7108	0.7970	0.5367	0.4040	0.8018	1.0000	0.4694	0.6219	0.4586	0.2421	0.2585	0.6103	0.4568
2014	0.7256	0.6495	0.8722	0.9267	0.6957	0.4694	1.0000	0.8720	0.8356	0.8811	0.8715	0.8162	0.8615
2015	0.7335	0.6901	0.7189	0.7793	0.7670	0.6219	0.8720	1.0000	0.9752	0.7828	0.7997	0.9704	0.8832
2016	0.6051	0.5355	0.6196	0.7670	0.6852	0.4586	0.8356	0.9752	1.0000	0.8202	0.8377	0.9664	0.8853
2017	0.5488	0.4215	0.6245	0.9514	0.5772	0.2421	0.8811	0.7828	0.8202	1.0000	0.9872	0.8109	0.9315
2018	0.5848	0.4592	0.6651	0.9509	0.6379	0.2585	0.8715	0.7997	0.8377	0.9872	1.0000	0.8324	0.9118
2019	0.6849	0.6292	0.6233	0.7908	0.8149	0.6103	0.8162	0.9704	0.9664	0.8109	0.8324	1.0000	0.8942
2020	0.6522	0.5626	0.5969	0.9027	0.6279	0.4568	0.8615	0.8832	0.8853	0.9315	0.9118	0.8942	1.0000

TABLE 11. Fuzzy equivalence relation matrix (transitive closure).

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
2008	1.0000	0.9830	0.8432	0.8432	0.8149	0.8018	0.8432	0.8432	0.8432	0.8432	0.8432	0.8432	0.8432
2009	0.9830	1.0000	0.8432	0.8432	0.8149	0.8018	0.8432	0.8432	0.8432	0.8432	0.8432	0.8432	0.8432
2010	0.8432	0.8432	1.0000	0.8722	0.8149	0.8018	0.8722	0.8722	0.8722	0.8722	0.8722	0.8722	0.8722
2011	0.8432	0.8432	0.8722	1.0000	0.8149	0.8018	0.9267	0.8942	0.8942	0.9514	0.9514	0.8942	0.9315
2012	0.8149	0.8149	0.8149	0.8149	1.0000	0.8018	0.8149	0.8149	0.8149	0.8149	0.8149	0.8149	0.8149
2013	0.8018	0.8018	0.8018	0.8018	0.8018	1.0000	0.8018	0.8018	0.8018	0.8018	0.8018	0.8018	0.8018
2014	0.8432	0.8432	0.8722	0.9267	0.8149	0.8018	1.0000	0.8942	0.8942	0.9267	0.9267	0.8942	0.9267
2015	0.8432	0.8432	0.8722	0.8942	0.8149	0.8018	0.8942	1.0000	0.9752	0.8942	0.8942	0.9704	0.8942
2016	0.8432	0.8432	0.8722	0.8942	0.8149	0.8018	0.8942	0.9752	1.0000	0.8942	0.8942	0.9704	0.8942
2017	0.8432	0.8432	0.8722	0.9514	0.8149	0.8018	0.9267	0.8942	0.8942	1.0000	0.9872	0.8942	0.9315
2018	0.8432	0.8432	0.8722	0.9514	0.8149	0.8018	0.9267	0.8942	0.8942	0.9872	1.0000	0.8942	0.9315
2019	0.8432	0.8432	0.8722	0.8942	0.8149	0.8018	0.8942	0.9704	0.9704	0.8942	0.8942	1.0000	0.8942
2020	0.8432	0.8432	0.8722	0.9315	0.8149	0.8018	0.9267	0.8942	0.8942	0.9315	0.9315	0.8942	1.0000

averaging the load growth rate in the four years of 2011, 2014, 2017, and 2018. (Note: The load growth rate in 2018 uses predicted value, that is, the load growth rate in 2017).

The actual load growth rate of each year is shown in Table 12.

TABLE 12. Actual load growth rate of the province from 2006 to 2018.

Year	load growth rate /%	Year	load growth rate /%
2008	10.2417	2015	9.3541
2009	16.3901	2016	7.6315
2010	-0.0364	2017	3.4851
2011	5.3091	2018	0.9769
2012	18.1031	2019	10.3043
2013	32.5126	2020	3.5564
2014	2.38124		

E. ANALYSIS OF FORECASTING RESULTS

Forecasting results of fuzzy clustering (FCA) without considering weights, grey relational analysis-fuzzy clustering (GRA-FCA) and two-way weighted grey relational analysis-fuzzy clustering (WGRA-FCA) are compared. At the same time, compared with traditional forecasting methods such as grey GM(1,1) model, GDP elastic coefficient method and multiple linear regression model, the results are shown in Table 13.

From forecasting results, it can be seen that forecasting accuracy of GM(1,1) model is the lowest, because under the influence of policy factors, volatility of load sequence is obviously strengthened, and the extrapolation prediction method is difficult to adapt. However, the method proposed in this paper makes prediction on the basis of clustering, so it will not be affected by the shape of load sequence, and it is compatible for the characteristics that fluctuation of load sequence is gradually prominent and the law is difficult to be grasped under the influence of policy factors; GDP elasticity coefficient method considers the correlation between load and GDP, but load factors influenced by policy adjustment are gradually diversified, and it is too simple to consider only from the perspective of GDP, so prediction accuracy is not high enough. However, the method proposed in this paper comprehensively considers multiple policy factors that influence power load, and on this basis, two-way weighted grey relational analysis method is used to weight each factor, which is suitable for the characteristics of numerous factors and different influence degrees under the background of policy influence; When selecting load influencing factors,

TABLE 13. Comparison of forecasting results and accuracy.

Year	Actual load growth rate/%	FCA /%	GRA-FCA /%	WGRA-FCA/%	GM(1,1)/%	GDP elasticity coefficient method/%	multiple linear regression model/%
2018	0.9769	5.3091	4.3971	3.4851	4.1932	3.8502	3.3766
2019	10.3043	8.4928	8.4928	8.4928	5.8396	6.1321	8.3048
2020	3.5564	5.3091	4.3971	3.6651	7.1260	5.6311	6.3758
Average forecasting error		2.6321	2.0241	1.4761	3.7502	3.0400	2.4062

the multiple linear regression model considers the correlation between load and multiple factors, but relatively lacks a systematic and hierarchical index system from a macro perspective.

Fuzzy cluster analysis method considers multiple factors at the same time and realizes prediction through similar clustering among samples. Therefore, it is less influenced by the fluctuation of load sequence and more suitable for load forecasting under the background of policy influence. On this basis, the weighted fuzzy clustering method takes into account the different influences of different factors on load, so as to improve the clustering effect and improve the prediction accuracy.

VII. CONCLUSION

This paper has the following two main tasks:

1. The influence of policy factors on power load is quantified.

This paper summarizes the policy factors that affect load into three macro levels: economic development, industrial structure adjustment, and energy conservation and emission reduction. Eight micro policy factor indexes are selected to construct a hierarchical policy factor index system that combines macro and micro levels.

Aiming at the problem of insufficient consideration of difference of historical data and future development situation of traditional grey relational analysis model, longitudinal AHP weighting method and transverse PCA weighting method are proposed, and a quantification analysis model of power load influencing factors based on two-way weighted grey relational analysis is proposed by weighting historical period and factor indexes respectively. The two-way weighted grey relational analysis method is used to quantify and weight the influence of various policy factors on load. The quantitative analysis method is mainly used to study the policy factors affecting power load, and research and judgment of many experts on future situation are integrated. Compared with qualitative analysis, quantitative analysis has better reliability and accuracy, and is more suitable for power load analysis under the influence of policy factors; Compared with traditional grey relational analysis, WGRA model proposed in this paper considers the difference of historical data and research and judgment of many experts on future situation are integrated, the quantification results obtained are more realistic.

2. The problem of power load forecasting under the influence of policy factors is solved.

Aiming at the problem that fuzzy cluster analysis method does not consider the influence of different factors on load is different, in order to improve clustering effect, traditional fuzzy clustering is weighted and improved to weighted fuzzy clustering. Quantification weights obtained by WGRA model are applied to fuzzy cluster analysis, and a new load forecasting model based on WGRA-FCA is established.

The WGRA-FCA forecasting model proposed is applied to medium and long-term power load forecasting. The results show that compared with GM(1,1) method, GDP elasticity coefficient method, multiple linear regression and fuzzy cluster analysis method without considering weights, the proposed method has certain advantages.

Although some achievements have been made in this paper, there are still some work to be done:

1. The "carbon peak, carbon neutralization" policy recently proposed will have a profound impact on the energy industry, and the impact of the policy on load will be taken into account subsequently.

2. In order to verify the accuracy and practicality of the method proposed in this paper, The calculation example in this paper adopts the method of comparing the predicted value of the known year with the actual value. This method will try to be applied to actual projects that forecast and plan the load of the 14th Five-Year Plan based on the data of the 12th Five-Year Plan and the 13th Five-Year Plan

3. The method proposed in this paper can not only be applied to quantify the impact of policies on load and then predict, but also can be used to quantify the impact of urbanization, economic and other uncertain factors on load and then predict. Therefore, this direction needs to be further studied.

REFERENCES

[1] S. Wang, Y. Zhang, and L. Ge, "Interpretative structural model of influencing factors for distribution network of new-type town," *Electr. Power Autom. Equip.*, vol. 35, no. 11, pp. 75–81, 2015.

[2] Y. Gao, Y. Fang, H. Dong, and Y. Kong, "A multifactorial framework for short-term load forecasting system as well as the Jinan's case study," *IEEE Access*, vol. 8, pp. 203086–203096, 2020.

[3] S. Khuntia, J. Rueda, and M. van Der Meijden, "Forecasting the load of electrical power systems in mid-and long-term horizons: A review," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 16, pp. 3971–3977, 2016.

[4] K. G. Boroojeni, M. H. Amini, S. Bahrami, S. S. Iyengar, A. I. Sarwat, and O. Karabasoglu, "A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon," *Electr. Power Syst. Res.*, vol. 142, pp. 58–73, Jan. 2017.

[5] H. Dong, Y. Gao, X. Meng, and Y. Fang, "A multifactorial short-term load forecasting model combined with periodic and non-periodic features–A case study of Qingdao, China," *IEEE Access*, vol. 8, pp. 67416–67425, 2020.

[6] B. Zhang, W. Liu, and S. Li, "Short-term load forecasting based on wavelet neural network with adaptive mutation bat optimization algorithm," *IEEJ Trans. Elect. Electron. Eng.*, vol. 14, no. 3, pp. 376–382, 2019.

[7] Z. Zhang, W.-C. Hong, and J. Li, "Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm," *IEEE Access*, vol. 8, pp. 14642–14658, 2020.

[8] T. Zhao, J. Wang, and Y. Zhang, "Day-ahead hierarchical probabilistic load forecasting with linear quantile regression and empirical copulas," *IEEE Access*, vol. 7, pp. 80969–80979, 2019.

[9] J. Zhang, Z. Tan, and Y. Wei, "An adaptive hybrid model for short term electricity price forecasting," *Appl. Energy*, vol. 258, Jan. 2020, Art. no. 114087.

[10] J. Zhang, Y. Wei, D. Li, Z. Tan, and J. Zhou, "Short term electricity load forecasting using a hybrid model," *Energy*, vol. 158, pp. 774–781, Sep. 2018.

[11] T. Yuan, J. Yuan, and Q. Chao, "Study on the comprehensive model of mid-long term load forecasting," *Power Syst. Protection Control*, vol. 40, no. 14, pp. 143–146, 2012.

[12] L. Yin, Z. Sun, F. Gao, and H. Liu, "Deep forest regression for short-term load forecasting of power systems," *IEEE Access*, vol. 8, pp. 49090–49099, 2020.

[13] D. Wang and B. Wang, "Medium-and long-term load forecasting based on variable weights buffer grey model," *Power Syst. Technol.*, vol. 37, no. 1, pp. 167–171, 2013.

[14] Y. He, C. Wen, and Q. Xu, "A method to predict probability density of medium-term power load considering temperature factor," *Power Syst. Technol.*, vol. 39, no. 1, pp. 176–181, 2015.

[15] Y. Wang and M. Wu, "Medium and long term load forecasting model based on partial least square regression under new normal economy," *Electr. Power Automat. Equip.*, vol. 38, no. 3, pp. 133–139, 2018.

[16] B. Li, J. Zhang, Y. He, and Y. Wang, "Short-term load-forecasting method based on wavelet decomposition with second-order gray neural network model combined with ADF test," *IEEE Access*, vol. 5, pp. 16324–16331, 2017.

[17] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sep. 2019.

[18] W. Li, Q. Shi, M. Sibtain, D. Li, and D. E. Mbanze, "A hybrid forecasting model for short-term power load based on sample entropy, two-phase decomposition and whale algorithm optimized support vector regression," *IEEE Access*, vol. 8, pp. 166907–166921, 2020.

[19] Y. Xuan, W. Si, J. Zhu, Z. Sun, J. Zhao, M. Xu, and S. Xu, "Multi-model fusion short-term load forecasting based on random forest feature selection and hybrid neural network," *IEEE Access*, vol. 9, pp. 69002–69009, 2021.

[20] L. Chen, H. Chiang, N. Dong, and R. Liu, "Roup-based chaos genetic algorithm and non-linear ensemble of neural networks for short-term load forecasting," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 6, pp. 1440–1447, 2016.

[21] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Sep. 2017.

[22] M. Cui and J. Gu, "Mid-long term load forecast of power system based on data-mining," *Electr. Power Automat. Equip.*, vol. 24, no. 6, pp. 18–21, 2004.

[23] Y. Li, Z. Yan, and D. Feng, "Mid/long-term load forecasting model considering urbanization characteristics," *Electr. Power Automat. Equip.*, vol. 36, no. 4, pp. 54–61, 2016.

[24] Y. Wang, M. Wu, Z. Zhou, and H. Ma, "Quantitative analysis model of power load influencing factors based on improved grey relational degree," *Power Syst. Technol.*, vol. 41, no. 6, pp. 1772–1778, 2017.

[25] L. Shunfu, X. Chao, and L. Dongdong, "Load preprocessing method based on grey relational analysis and fuzzy clustering," *Electr. Meas. Instrum.*, vol. 54, no. 11, pp. 36–42, 2017.

[26] G. Huebner, D. Shipworth, I. Hamilton, Z. Chalabi, and T. Oreszczyn, "Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes," *Appl. Energy*, vol. 177, pp. 692–702, Sep. 2016.

[27] B. Chen and Y. Wang, "Short-term electric load forecasting of integrated energy system considering nonlinear synergy between different loads," *IEEE Access*, vol. 9, pp. 43562–43573, 2021.

[28] C. Li, D. Niu, and L. Meng, "A comprehensive model for long-and medium-term load forecasting based on analytic hierarchy process and radial basis function neural network," *Power Syst. Technol.*, vol. 33, no. 2, pp. 99–104, 2009.

[29] J. Shi, Z. Ding, W.-J. Lee, Y. Yang, Y. Liu, and M. Zhang, "Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 521–526, Jan. 2014.

[30] Y. Zhang, H. Shi, W. Zhai, and C. He, "Line loss rate prediction based on combinatorial modeling by analytic hierarchy process-gray comprehensive relationship and multi grey models," *Power Syst. Technol.*, vol. 35, no. 6, pp. 71–76, 2011.

[31] F. M. Bianchi, E. De Santis, A. Rizzi, and A. Sadeghian, "Short-term electric load forecasting using echo state networks and PCA decomposition," *IEEE Access*, vol. 3, pp. 1931–1943, 2015.

[32] F. Liu, T. Dong, T. Hou, and Y. Liu, "A hybrid short-term load forecasting model based on improved fuzzy C-means clustering, random forest and deep neural networks," *IEEE Access*, vol. 9, pp. 59754–59765, 2021.

[33] R. Chen, Y. Zhang, and Z. Wu, "Application of improving fuzzy clustering algorithm to power load forecasting," *Proc. Electr. Power Syst. Automat.*, vol. 3, pp. 73–77, May 2005.

BO ZHANG was born in Shandong, China, in 1993. He received the B.S. and M.S. degrees in electrical engineering from the Shandong University of Technology, Shandong, in 2017 and 2020, respectively. He is currently working at State Grid Juancheng Power Supply Company. His research interests include power system short-term load forecasting and medium and long-term forecasting.

XIAOHAN ZHAO was born in Shandong, China, in 1995. She received the B.S. degree in electrical engineering from the Shandong University of Technology, Shandong, in 2018. She is currently pursuing the M.S. degree with the China University of Mining and Technology. Her current research interests include power electronics and motor drive.

ZHENHAI DOU was born in Beijing, China, in 1970. He received the M.S. degree in electrical engineering from the Inner Mongolia University of Science and Technology, Inner Mongolia, China, in 2005, and the Ph.D. degree in electrical engineering from China Agricultural University, Beijing, in 2014. He is currently an Associate Professor with the School of Electrical and Electronic Engineering, Shandong University of Technology. His research interests include power system load forecasting and power system automation.

LIANXIN LIU was born in Shandong, China, in 1965. He received the B.S. degree from Shandong University, Jinan, China, in 1987. He is currently an Associate Professor with the School of Electrical and Electronic Engineering, Shandong University of Technology. His research interest includes power system automation.