

## Research Article

# A Novel Clustering Model Based on Set Pair Analysis for the Energy Consumption Forecast in China

Mingwu Wang, Dongfang Wei, Jian Li, Hui Jiang, and Juliang Jin

Hefei University of Technology, 193 Tunxi Road, Hefei 230009, China

Correspondence should be addressed to Mingwu Wang; wanglab307@foxmail.com

Received 29 April 2014; Accepted 11 July 2014; Published 24 July 2014

Academic Editor: Guido Maione

Copyright © 2014 Mingwu Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The energy consumption forecast is important for the decision-making of national economic and energy policies. But it is a complex and uncertainty system problem affected by the outer environment and various uncertainty factors. Herein, a novel clustering model based on set pair analysis (SPA) was introduced to analyze and predict energy consumption. The annual dynamic relative indicator (DRI) of historical energy consumption was adopted to conduct a cluster analysis with Fisher's optimal partition method. Combined with indicator weights, group centroids of DRIs for influence factors were transferred into aggregating connection numbers in order to interpret uncertainty by identity-discrepancy-contrary (IDC) analysis. Moreover, a forecasting model based on similarity to group centroid was discussed to forecast energy consumption of a certain year on the basis of measured values of influence factors. Finally, a case study predicting China's future energy consumption as well as comparison with the grey method was conducted to confirm the reliability and validity of the model. The results indicate that the method presented here is more feasible and easier to use and can interpret certainty and uncertainty of development speed of energy consumption and influence factors as a whole.

## 1. Introduction

Nowadays China is in the middle term of industrialization and urbanization and is the world's second largest energy consumer. As we know, energy is an essential material base for economic development. Energy consumption skyrockets along with the rapid and steady economic growth, industrialization, and urbanization in China, which has resulted in a serious imbalance between supply and demand of energy [1, 2]. In China, the total energy consumption increased from 0.987 billion ton coal equivalents (tce) in 1990 to 3.25 billion tce in 2010 [3]. This has a significant influence on the current global energy profiles, and China also faces a big challenge of carbon emission decrease in the future. As we know, forecast of energy consumption is a significant precondition and basis for making energy policies of a developing country like China. Consequently, to keep sustainable and stable development, accurate forecast of energy consumption is essential, and the development of rational forecast model especially is urgent and necessary.

Many researchers have studied the relation between energy consumption and economic growth in national

or regional level and proposed some forecast models for countries such as Turkey, India, Iran, UK, Finland, New Zealand, and China [4–7]. Nonetheless, energy consumption is affected by outer environment and various elements. Thus, to obtain rational forecast results, various prediction models based on regression analysis [8, 9], artificial neural networks (ANN) theory [10–17], grey theory [18–20], multivariate statistical analysis theory, and time series theory [21, 22] were proposed. However, there is not a general model up to now because the previous models do not reveal general characteristics and have drawbacks themselves. Besides, traditional techniques, such as regression analysis, econometric model, and autoregressive integrated moving average (ARIMA) model, are poor in precision when data are few or exhibit nonlinear characteristics [20, 23–25]. Although the grey models (GMs) can overcome this problem induced by small data sets or data with limited information, the weakness of grey theory is uncontrollable on the accuracy of prediction since the effectiveness of the residual series of GM(1, 1) depends on the number of data points with the same sign, which is low when the observations are limited [26].

ANN techniques can avoid this disadvantage, but it shows inability to present an explicit relationship between energy consumption and impact factors, and its application is limited by knowledge acquisition. Other effective approaches such as support vector regression (SVR), adaptive particle swarm optimization (PSO), and genetic algorithm (GA) were also introduced to forecast electricity consumption [27–29]. But their procedures are complex and rather inconvenient for the decision-makers. Meantime, some integrated models [30, 31] were also proposed and the results exhibited superiority compared with the single optimization. To study energy savings in buildings, Gaitani et al. [32] defined five energy classes for 1100 schools from all the prefectures of Greece through clustering techniques and introduced the principal components analysis to identify the typical characteristics of school buildings belonging to a particular energy class. As mentioned above, much research has been performed to advance the techniques for energy consumption forecast, whereas energy consumption forecast is still not well resolved nowadays. And there are few studies focusing on energy consumption forecast by uncertainty analysis methods, and the previous researches cannot consider and depict quantitatively certainty and uncertainty relationship between energy consumption and influence factors as a whole. A newly proposed method of set pair analysis (SPA) can deal with the uncertainty problem from three aspects of identity, discrepancy, and contrary features and depict comprehensively essential characteristics of things [33–35]. So this method provides a fresh idea for energy consumption forecast.

This paper introduces a novel clustering model based on SPA for energy consumption prediction to deal with uncertainty relation between energy consumption and its influence factors. And the proposed model is used to forecast China's energy consumption, and its feasibility and effectiveness are also further discussed.

## 2. Theory

*2.1. Brief of Set Pair Analysis Theory.* The set pair analysis theory put forward by Zhao [33] is an uncertainty analysis method. Certainty and uncertainty can be taken into consideration as a whole and be treated dialectically during the process of SPA. The connection number, a basic idea of SPA theory, was described to depict the relation of a set pair made up by two sets [33–40]. And the corresponding function of the connection number can be written as

$$\mu_{(A,B)} = a + bi + cj, \quad (1)$$

where  $i$  is the discrepancy coefficient,  $i \in [-1, 1]$ , and the uncertainty number sometimes denotes only a mark of difference;  $j$  is the contrary coefficient,  $j \equiv -1$ , or a sign of the contrary;  $a$ ,  $b$ , and  $c$  are identical degree, different degree, and contrary degree, respectively, and  $a + b + c = 1$ .

In case application, SPA theory can describe certainty and uncertainty by connection numbers in one system. Meanwhile, IDC analysis of set pair is a dynamic process, in which ideality can transfer into discrepancy or contrary with condition changes. Therefore, SPA can be utilized to clearly interpret the uncertainty of energy consumption forecast.

*2.2. Optimal Partition Method.* Cluster analysis is widely used in pattern recognition, image analysis, information retrieval, and other fields. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals, or particular statistical distributions [41, 42]. The reported literatures show that most cluster methods cannot deal with the problem of ordered samples since the ordered samples cannot be scattered in the clusters. However, the optimal partition method presented by Fisher in 1958 can effectively overcome this problem [43, 44]. This method is of concise calculation process and efficient to find out reasonable groups and save computing cost. Suppose that  $l$  samples are partitioned into  $q$  clusters  $G(q, l)$ . The core of Fisher algorithm can be expressed as the following recursion formulas:

$$\begin{aligned} L[P(l, 2)] &= \min_{2 \leq q \leq l} \{D(1, q-1) + D(q, l)\}, \\ L[P(l, k)] &= \min_{k \leq m \leq l} \{D(q-1, k-1) + D(q, l)\}, \\ D(q, l) &= \sum_{t=q}^l (x_t - \bar{x})^2, \end{aligned} \quad (2)$$

where  $L[P(l, k)]$  denotes an objective function.  $D(q, l)$  is the diameter of a cluster  $G(q, l) = \{x_q, x_{q+1}, \dots, x_l\}$  ( $q < l$ ).  $x_t$  and  $\bar{x}$  are indicator value and mean value of ordinal samples in  $G(q, l)$ .

## 3. Development of Cluster Forecasting Model Based on Set Pair Analysis

*3.1. Basic Principle and Procedures.* Basic principle of SPA-based clustering model is depicted as follows. Based on ordered DRIs of historical energy consumption, first conduct cluster analysis with Fisher's partition method to obtain categories of development speed. Then analyze the uncertainty of influence factors through identity-discrepancy-contrary analysis. And forecast the energy consumption in a certain year according to the similarities of measured influence factors to each category. The corresponding flow chart is sketched as showed in Figure 1. And the detailed procedure is illustrated as below.

*Step 1.* Conduct cluster analysis with Fisher's optimal partition method for the ascending DRI of historical energy consumption. And calculate group centroid values of energy consumption and influence factors for each cluster.

*Step 2.* A criterion is constructed to conduct set pair analysis, and the corresponding formulas are described to transfer mean values of DRIs into connection numbers in order to calculate the identical degree, different degree, and contrary degree between influence factors and reference sets. Then, combined with weights of influence factors, obtain integrated connection numbers.

*Step 3.* Express influence factors of energy consumption to be forecasted in a certain year with connection numbers

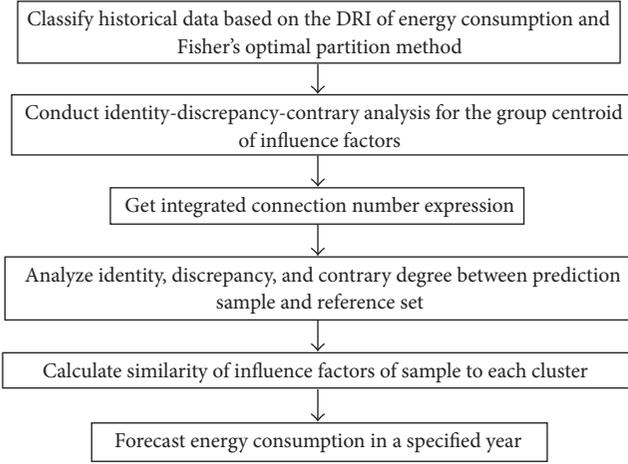


FIGURE 1: Flow chart of SPA-based cluster forecast model.

obtained from the IDC analysis, and calculate their similarities to each cluster.

*Step 4.* Construct a forecasting model and predict the energy consumption in a specified year.

### 3.2. Expressions of Influence Factors in terms of Connection Numbers

*3.2.1. Definition of IDC Criteria.* To combine the concept of IDC with the development speed, suppose that the reference set is  $x_n^k = 1$  ( $n = 1, 2, \dots, N$ ;  $k = 1, 2, \dots, K$ ) because the original state of annual DRI is a relative certainty variable. Then, the IDC analysis between the annual DRI and the reference set can be conducted for the discussed year on a basis of IDC criteria. The IDC criteria are defined as follows. If the annual DRI value is greater than 1, it is defined as identity according to the SPA theory, and the bigger the amount of the dynamic relative number appears, the higher the possibility of the growth situation is. When the annual dynamic relative number is less than 1, this is defined as contrary, and the smaller the amount of dynamic relative number exhibits the higher possibility of attenuation. Besides, while the annual DRI is 1, it is called discrepancy, which means that the discussed indicator shows uncertainty characteristic against the initial state of the influence factor.

*3.2.2. Expressions of Group Centroids.* Let  $k$  ( $k = 1, 2, \dots, K$ ) represent the number of clusters or categories. The corresponding group centroid of influence factors for each category is defined as

$$\bar{x}_n^{(k)} = \frac{1}{M} \sum_{m=1}^M x_{nm}^{(k)}, \quad (3)$$

where  $\bar{x}_n^{(k)}$  is the mean value of  $n$ th influence factor on category  $k$ .  $x_{nm}^{(k)}$  is the corresponding DRI value of  $n$ th influence factor on sample  $m$  ( $m = 1, 2, \dots, M$ ). Then, based on the set pair analysis theory [40], integrated connection

number  $\mu_k$ , used to depict growth and decline features, can be written as

$$\mu_k = a_k + b_k i + c_k j = \sum_{n=1}^N a_{nk} w_n + i \sum_{n=1}^N b_{nk} w_n + j \sum_{n=1}^N c_{nk} w_n, \quad (4)$$

$$a_{nk} = \lambda \bar{x}_n^{(k)}, \quad (5)$$

$$c_{nk} = \frac{\eta}{\bar{x}_n^{(k)}}, \quad (6)$$

$$b_{nk} = 1 - a_{nk} - c_{nk}, \quad (7)$$

where  $a_k$ ,  $b_k$ , and  $c_k$  are integrated identical degree, different degree, and contrary degree, respectively;  $w_n$  is an indicator weight;  $a_{nk}$ ,  $b_{nk}$ , and  $c_{nk}$  are identical degree, discrepancy degree, and contrary degree of  $n$ th indicator relative to reference set on category  $k$ , respectively;  $\lambda$  and  $\eta$  are coefficients; if  $\bar{x}_n^{(k)} \in [x^L, x^U]$ ,  $\lambda$  and  $\eta$  will satisfy

$$0 < \lambda \eta \leq 0.25, \quad (8)$$

$$(x^U)^2 \lambda + \eta - x^U \geq 0, \quad (9)$$

$$(x^L)^2 \lambda + \eta - x^L \geq 0, \quad (10)$$

where  $x^L$  and  $x^U$  are the lower and upper values of DRI for historical indicators. The value ranges of  $\lambda$  and  $\eta$  are presented in Figure 2.

The intersection point of two linear functions (9) and (10) is always used to specify  $\lambda$  and  $\eta$ . Namely,

$$\lambda = \frac{1}{(x^U + x^L)}, \quad (11)$$

$$\eta = \frac{x^U x^L}{(x^U + x^L)}.$$

In a similar way, the connection degree between the measured indicator value of a forecasting year and reference set ( $x_n^k = 1$ ) can also be achieved.

*3.2.3. The Forecast Model Based on Similarity.* Let set  $A$  be DRIs of influence factors of energy consumption, set  $B$  is Reference set in which all the DRI values are 1, and set  $C$  represents measured DRIs of influence factors of a forecast object. According to the definition of set pair, two set pairs,  $H_1(A, B)$  and  $H_2(C, B)$ , can be established. And corresponding connection numbers are written as follows:

$$\mu_{(A,B)} = a_k + b_k i + c_k j, \quad (12)$$

$$\mu_{(C,B)} = a_f + b_f i + c_f j,$$

where  $a_f$ ,  $b_f$ , and  $c_f$  are identical degree, different degree, and contrary degree, respectively.

The similarity between set pairs  $H_1$  and  $H_2$  can be defined as

$$d_k = \left[ (a_k - a_f)^2 + (b_k - b_f)^2 + (c_k - c_f)^2 \right]^{1/2}, \quad (13)$$

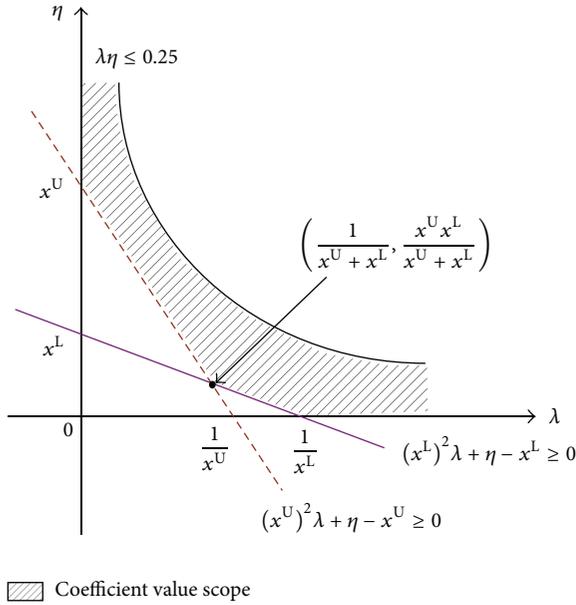


FIGURE 2: Coefficient value scope of  $\lambda$  and  $\eta$ .

where  $d_k$  is the similarity to category  $k$ . And the forecast model for the annual development speed is given as

$$y = \frac{\sum_{k=1}^K (\bar{y}_k / d_k)}{\sum_{k=1}^K (1/d_k)}, \quad (14)$$

where  $\bar{y}_k$  is the centre of energy consumption on category  $k$  and  $y$  is the forecasted value.

#### 4. Case Study

The model discussed is applied to analyze and predict China's energy consumption. Data from China statistics yearbook were used to confirm its validity and effectiveness [3]. Many reports have shown that energy consumption has a deep association with factors such as GDP, proportion of secondary industry, urbanization level, and price index. Thus, in this model, GDP, proportion of secondary industry, urbanization level, and price index were taken as major influence factors for energy consumption. Detailed materials of sample data and annual DRI values are listed in Tables 1 and 2.

According to Table 2, from 1990 to 2010, energy consumption increased significantly from 0.987 to 3.25 billion tce with an average annual development speed of 1.07. Chained development speed of Chinese energy consumption varied from 1.002 to 1.162. Based on Fisher's optimal partition method, the objective function varied with the number of clusters as shown in Figure 3. The knee point of the objective function was observed when the number of clusters is equal to 4. So chained speed of development for historical energy consumption should be divided into 4 categories. Then, the group centroids of influence factors as listed in Table 3 were obtained according to (3). Thereby, we substituted mean values of influence factors into (5), (6), and (7), which

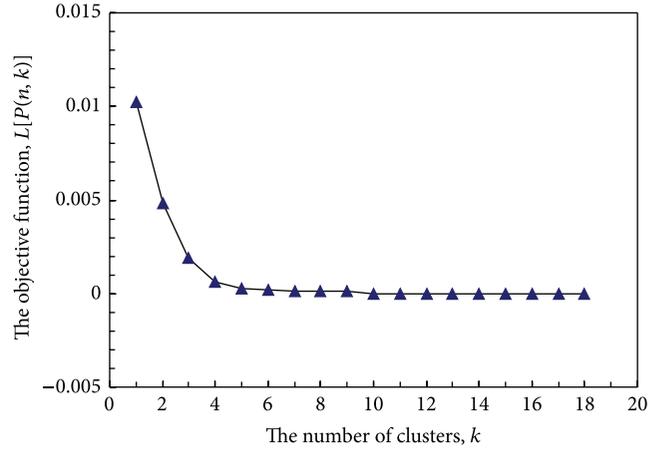


FIGURE 3: Relationship between the number of clusters and the objective function.

got corresponding connection numbers for each category. Herein,  $\lambda$  and  $\eta$  were 0.5 and 0.2, respectively, obtained on the basis of formulas (11). Combined with the same weights of influence factors, through (4), the integrated connection degree was obtained. The results were presented in Table 4.

In the year of 2010 China's practical information about energy consumption was used to test and verify this model (see Table 5). Firstly, let set  $C$ , which consists of the statistical values of GDP, be proportion of secondary industry, urbanization level, and price index for the year of 2010. Based on SPA theory, conduct IDC analysis between influence factors value and reference set  $B = \{1, 1, 1, 1\}$ , and calculate the corresponding connection numbers. Meanwhile, according to the definition of similarity in expression (13) and a nearby rule, chained development speed for 2010 was specified as category 3. Finally, by (14), the forecasted chained development speed of energy consumption value was 1.085 in reference to 2009, which is of 2.37% relative error in comparison to the real value of 1.06. Besides, the prediction was conducted with the GM(1, 1) model, and its forecasted value was 1.024, whose relative error reached 3.40%. It indicated that the proposed model improves prediction performance significantly and is feasible and effective.

As noted above, the method proposed here overcomes drawbacks of conventional methods based on single type information and a static perspective. And it will enable us to provide a more comprehensive background for the characterization of energy consumption and to make appropriate energy policies of a developing country. However, the forecast of energy consumption involves various factors of incompatibility, complexity and diversity, combination, and dynamic uncertainty. Consequently, it would be important to clarify effects of factors on the energy forecast in various time frames. The same weights of influence factors used to calculate the integrated connection degree in the case study may neglect the importance of indicators and effect on the prediction. To provide more information about the most sensitive parameters and improve the forecast accuracy, considerable amount of work both on the sensitivity analysis

TABLE 1: Energy consumption and its influence factor from 1990 to 2010 in China.

Year	Energy consumption ( $10^8$ tce)	GDP (billion RMB)	Proportion of secondary industry (%)	Urbanization level (%)	Price index
1990	9.87	18.55	41.34	26.41	100
1991	10.38	20.25	41.79	26.94	112.97
1992	10.92	23.13	43.44	27.46	131.45
1993	11.6	26.36	46.57	27.99	179.71
1994	12.27	29.81	46.57	28.51	212.03
1995	13.12	33.07	47.18	29.04	230.51
1996	13.52	36.38	47.54	30.48	253.99
1997	13.59	39.76	47.54	31.91	277.61
1998	13.62	42.88	46.21	33.35	275.14
1999	14.06	46.14	45.76	34.78	277.61
2000	14.55	50.04	45.92	36.22	320.36
2001	15.04	54.19	45.05	37.66	321.01
2002	15.94	59.11	44.79	39.09	321.3
2003	18.38	65.04	45.97	40.53	345.07
2004	21.35	71.59	46.23	41.76	378.55
2005	23.6	79.69	47.37	42.99	435.36
2006	25.87	89.79	47.95	44.34	487.17
2007	28.05	102.51	47.34	45.89	508.12
2008	29.14	112.39	47.45	46.99	612.75
2009	30.66	122.74	46.3	48.34	546.59
2010	32.5	135.39	46.75	49.95	635.65

TABLE 2: Ordered relative numbers of energy consumption and corresponding DRIs of influence factors.

Year	DRI of energy consumption	DRI of GDP	DRI of proportion of secondary industry	DRI of urbanization level	DRI of price index
1998	1.0022	1.0785	0.9720	1.0451	0.9911
1997	1.0052	1.0929	1.0000	1.0469	1.0930
1996	1.0305	1.1001	1.0076	1.0496	1.1019
1999	1.0323	1.0760	0.9903	1.0429	1.0090
2001	1.0337	1.0829	0.9811	1.0398	1.0020
2000	1.0349	1.0845	1.0035	1.0414	1.1540
2008	1.0389	1.0964	1.0023	1.0240	1.2059
1991	1.0517	1.0916	1.0109	1.0201	1.1297
1992	1.0520	1.1422	1.0395	1.0193	1.1636
2009	1.0522	1.0921	0.9758	1.0287	0.8920
1994	1.0578	1.1309	1.0000	1.0186	1.1798
2002	1.0598	1.0908	0.9942	1.0380	1.0009
1993	1.0623	1.1396	1.0721	1.0193	1.3671
1995	1.0693	1.1094	1.0131	1.0186	1.0872
2007	1.0843	1.1417	0.9873	1.0350	1.0430
2006	1.0962	1.1267	1.0122	1.0314	1.1190
2005	1.1054	1.1131	1.0247	1.0295	1.1501
2003	1.1531	1.1003	1.0263	1.0368	1.0740
2004	1.1616	1.1007	1.0057	1.0303	1.0970

TABLE 3: Interval of energy consumption and mean values of DRIs of influenced factors for each cluster.

Category	DRI interval for energy consumption	Number of samples	DRI of GDP	DRI of proportion of secondary industry	DRI of urbanization level	DRI of price index
$k = 1$	[1.00, 1.05)	7	1.0873	0.9938	1.0414	1.0796
$k = 2$	[1.05, 1.08)	7	1.1138	1.0151	1.0232	1.1172
$k = 3$	[1.08, 1.10)	3	1.1272	1.0081	1.0319	1.1040
$k = 4$	[1.10, 1.20]	2	1.1005	1.0160	1.0336	1.0855

TABLE 4: Connection numbers between influence factor and reference set.

Category	$k = 1$	$k = 2$	$k = 3$	$k = 4$
GDP	$0.544 + 0.272i + 0.184j$	$0.557 + 0.264i + 0.180j$	$0.564 + 0.259i + 0.177j$	$0.550 + 0.268i + 0.182j$
Proportion of secondary industry	$0.497 + 0.302i + 0.201j$	$0.508 + 0.295i + 0.197j$	$0.504 + 0.298i + 0.198j$	$0.508 + 0.295i + 0.197j$
Urbanization level	$0.521 + 0.287i + 0.192j$	$0.512 + 0.290i + 0.194j$	$0.516 + 0.290i + 0.194j$	$0.517 + 0.290i + 0.193j$
Price index	$0.540 + 0.275i + 0.185j$	$0.559 + 0.262i + 0.179j$	$0.552 + 0.267i + 0.181j$	$0.543 + 0.273i + 0.184j$
Integrated connection number	$0.525 + 0.284i + 0.191j$	$0.534 + 0.279i + 0.188j$	$0.534 + 0.278i + 0.188j$	$0.529 + 0.281i + 0.189j$

TABLE 5: Forecasted results and comparison with the gray method.

Year	Integrated connection number	Similarity $d_k$				Proposed model	GM(1, 1) model
		$k = 1$	$k = 2$	$k = 3$	$k = 4$		
2010	$0.539 + 0.275i + 0.186j$	0.0167	0.0062	0.0059	0.0115	1.085	1.024
2015	$0.528 + 0.282i + 0.190j$	0.0035	0.0070	0.0073	0.0017	1.105	1.051

and on the comparison with other methods is required to conduct with actual indicator weights in future.

Based on mean values of influence factors within 5 years from 2006 to 2010, the energy consumption in 2015 can be predicted with this discussed method as showed in Table 5. If the economy continues as usual, energy demand in China will continue to increase rapidly to 5.35 billion tce by 2015. However, there is a huge potential for reducing this projected level, since growth could be better by adjusting the energy and industrial structure and strengthening technology innovation. They may be especially helpful to frame suitable energy policy.

## 5. Conclusions

To provide reliable data for the decision-making of macroeconomic policy, a rational forecast model for energy consumption is of significance since well-targeted policies and reasonable measures are indispensable for rational energy consumption forecast. However, energy consumption forecast is a complex and uncertainty problem due to interactive factors. In this study, based on historical data of China's energy consumption and influence factors, a novel clustering forecast model based on SPA was presented to analyze energy consumption. Some conclusions can be drawn as follows.

- (1) The results indicate that this novel method used to forecast energy consumption is feasible and effective and convenient for practical applications. This cluster forecast model provides a potential method for other uncertainty problems.

- (2) The expressions in terms of connection number for group centroids of influence factors can depict the certainty and uncertainty of development speed as a whole.

- (3) Based on the similarities of DRIs of influence factors, interaction among the influence factors and similar information between historical samples and prediction object can be taken into account in the proposed model. Although our work has provided a useful clustering tool for making full use of similar information from historical samples for the energy consumption forecast and analyzing the certainty and uncertainty of evaluation indicators from three aspects embracing identity, discrepancy, and contrary, further investigations will still be in progress with sensitivity analysis to clarify effects of indicators on the prediction in various time frames.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

Financial support provided by the National Natural Sciences Foundations, China (no. 41172274 and no. 71273081), is gratefully acknowledged. The authors would also like to express their sincere thanks to the reviewers for their thorough reviews and useful suggestions.

## References

- [1] C. Yuan, S. Liu, Z. Fang, and N. Xie, "The relation between Chinese economic development and energy consumption in the different periods," *Energy Policy*, vol. 38, no. 9, pp. 5189–5198, 2010.
- [2] Y. Fan and Y. Xia, "Exploring energy consumption and demand in China," *Energy*, vol. 40, no. 1, pp. 23–30, 2012.
- [3] National Bureau of Statistics of China, *China Statistical Yearbook 2012*, Statistics Press, Beijing, China, 2012.
- [4] K. Fatai, L. Oxley, and F. G. Scrimgeour, "Modeling and forecasting the demand for electricity in New Zealand: a comparison of alternative approaches," *Energy Journal*, vol. 24, no. 1, pp. 75–102, 2003.
- [5] C. Hamzaçebi, "Forecasting of Turkey's net electricity energy consumption on sector bases," *Energy Policy*, vol. 35, no. 3, pp. 2009–2016, 2007.
- [6] K. J. Baker and R. M. Rylatt, "Improving the prediction of UK domestic energy-demand using annual consumption-data," *Applied Energy*, vol. 85, no. 6, pp. 475–482, 2008.
- [7] S. Siitonen, M. Tuomaala, M. Suominen, and P. Ahtila, "Implications of process energy efficiency improvements for primary energy consumption and CO<sub>2</sub> emissions at the national level," *Applied Energy*, vol. 87, no. 9, pp. 2928–2937, 2010.
- [8] A. Azadeh, M. Saberi, and O. Seraj, "An integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data: a case study of Iran," *Energy*, vol. 35, no. 6, pp. 2351–2366, 2010.
- [9] L. Tang, L. Yu, S. Wang, and J. Li, "A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting," *Applied Energy*, vol. 93, pp. 432–443, 2012.
- [10] P. A. González and J. M. Zamarreño, "Prediction of hourly energy consumption in buildings based on a feedback artificial neural network," *Energy and Buildings*, vol. 37, no. 6, pp. 595–601, 2005.
- [11] A. Sözen, Z. Gülseven, and E. Arcaklıoğlu, "Forecasting based on sectoral energy consumption of GHGs in Turkey and mitigation policies," *Energy Policy*, vol. 35, no. 12, pp. 6491–6505, 2007.
- [12] P. Lauret, E. Fock, R. N. Randrianarivony, and J.-F. Manicom-Ramsamy, "Bayesian neural network approach to short time load forecasting," *Energy Conversion and Management*, vol. 49, no. 5, pp. 1156–1166, 2008.
- [13] A. Azadeh, S. F. Ghaderi, and S. Sohrabkhani, "Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors," *Energy Conversion and Management*, vol. 49, no. 8, pp. 2272–2278, 2008.
- [14] H. T. Pao, "Forecasting energy consumption in Taiwan using hybrid nonlinear models," *Energy*, vol. 34, no. 10, pp. 1438–1446, 2009.
- [15] L. Ekonomou, "Greek long-term energy consumption prediction using artificial neural networks," *Energy*, vol. 35, no. 2, pp. 512–517, 2010.
- [16] P. C. Chang, C. Y. Fan, and J. J. Lin, "Monthly electricity demand forecasting based on a weighted evolving fuzzy neural network approach," *International Journal of Electrical Power and Energy Systems*, vol. 33, no. 1, pp. 17–27, 2011.
- [17] A. Kialashaki and J. R. Reisel, "Modeling of the energy demand of the residential sector in the United States using regression models and artificial neural networks," *Applied Energy*, vol. 108, pp. 271–280, 2013.
- [18] C. Hsu and C. Chen, "Applications of improved grey prediction model for power demand forecasting," *Energy Conversion and Management*, vol. 44, no. 14, pp. 2241–2249, 2003.
- [19] P. Zhou, B. W. Ang, and K. L. Poh, "A trigonometric grey prediction approach to forecasting electricity demand," *Energy*, vol. 31, no. 14, pp. 2839–2847, 2006.
- [20] Y. Lee and L. Tong, "Forecasting energy consumption using a grey model improved by incorporating genetic programming," *Energy Conversion and Management*, vol. 52, no. 1, pp. 147–152, 2011.
- [21] U. Kumar and V. K. Jain, "Time series models (Grey-Markov, Grey model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India," *Energy*, vol. 35, no. 4, pp. 1709–1716, 2010.
- [22] Y. S. Lee and L. I. Tong, "Forecasting nonlinear time series of energy consumption using a hybrid dynamic model," *Applied Energy*, vol. 94, pp. 251–256, 2012.
- [23] F. B. Gorucu and F. Gumrah, "Evaluation and forecasting of gas consumption by statistical analysis," *Energy Sources*, vol. 26, no. 3, pp. 267–276, 2004.
- [24] E. Delarue and W. D'haeseleer, "Adaptive mixed-integer programming unit commitment strategy for determining the value of forecasting," *Applied Energy*, vol. 85, no. 4, pp. 171–181, 2008.
- [25] Z. Wu and J. Xu, "Predicting and optimization of energy consumption using system dynamics-fuzzy multiple objective programming in world heritage areas," *Energy*, vol. 49, no. 1, pp. 19–31, 2013.
- [26] S. F. Liu and L. M. Xie, *Grey System Theory and Its Application*, Science Press, Beijing, China, 4th edition, 2008, (Chinese).
- [27] J. Huang, G. Tzeng, and C. Ong, "Two-stage genetic programming (2SGP) for the credit scoring model," *Applied Mathematics and Computation*, vol. 174, no. 2, pp. 1039–1053, 2006.
- [28] J. Z. Wang, W. J. Zhu, W. Y. Zhang, and D. H. Sun, "A trend fixed on firstly and seasonal adjustment model combined with the E-SVR for short-term forecasting of electricity demand," *Energy Policy*, vol. 37, no. 11, pp. 4901–4909, 2009.
- [29] S. Zhu, J. Wang, W. Zhao, and J. Wang, "A seasonal hybrid procedure for electricity demand forecasting in China," *Applied Energy*, vol. 88, no. 11, pp. 3807–3815, 2011.
- [30] S. Yu and K. Zhu, "A hybrid procedure for energy demand forecasting in China," *Energy*, vol. 37, no. 1, pp. 396–404, 2012.
- [31] Y. Zhu, G. H. Huang, Y. P. Li, L. He, and X. X. Zhang, "An interval full-infinite mixed-integer programming method for planning municipal energy systems—a case study of Beijing," *Applied Energy*, vol. 88, no. 8, pp. 2846–2862, 2011.
- [32] N. Gaitani, C. Lehmann, M. Santamouris, G. Mihalakakou, and P. Patargias, "Using principal component and cluster analysis in the heating evaluation of the school building sector," *Applied Energy*, vol. 87, no. 6, pp. 2079–2086, 2010.
- [33] K. Q. Zhao, *Set Pair Analysis and Its Preliminary Application*, Zhejiang Science and Technology Press, Hangzhou, China, 2000, (Chinese).
- [34] M. Wang, L. Li, and J. Jin, "Set pair analysis-variable fuzzy set model for the evaluation of surrounding rock stability," *Chinese Journal of Geotechnical Engineering*, vol. 30, no. 6, pp. 941–944, 2008 (Chinese).
- [35] M. Wang and G. Chen, "A novel coupling model for risk analysis of swell and shrinkage of expansive soils," *Computers and Mathematics with Applications*, vol. 62, no. 7, pp. 2854–2861, 2011.

- [36] X. H. Yang, X. J. Zhang, X. X. Hu, Z. F. Yang, and J. Q. Li, "Nonlinear optimization set pair analysis model (NOSPAM) for assessing water resource renewability," *Nonlinear Processes in Geophysics*, vol. 18, no. 5, pp. 599–607, 2011.
- [37] J. L. Jin, Y. M. Wei, L. L. Zou, L. Liu, and J. Fu, "Risk evaluation of China's natural disaster systems: an approach based on triangular fuzzy numbers and stochastic simulation," *Natural Hazards*, vol. 62, no. 1, pp. 129–139, 2012.
- [38] W. Wensheng and L. Yueqing, "Hazard degree assessment of landslide using set pair analysis method," *Natural Hazards*, vol. 60, no. 2, pp. 367–379, 2012.
- [39] M. W. Wang, P. Xu, J. Li, and K. Y. Zhao, "A novel set pair analysis method based on variable weights for liquefaction evaluation," *Natural Hazards*, vol. 70, no. 2, pp. 1527–1534, 2014.
- [40] M. W. Wang, J. L. Jin, and Y. L. Zhou, *Set Pair Analysis Based Coupling Methods and Applications*, Science Press, Beijing, China, 2014, (Chinese).
- [41] G. Ragno, M. D. Luca, and G. Ioele, "An application of cluster analysis and multivariate classification methods to spring water monitoring data," *Microchemical Journal*, vol. 87, no. 2, pp. 119–127, 2007.
- [42] D. F. Levia Jr. and D. R. Page, "The use of cluster analysis in distinguishing farmland prone to residential development: a case study of Sterling, Massachusetts," *Environmental Management*, vol. 25, no. 5, pp. 541–548, 2000.
- [43] W. D. Fisher, "On grouping for maximum homogeneity," *Journal of the American Statistical Association*, vol. 53, pp. 789–798, 1958.
- [44] Y. F. Sun, Y. C. Liang, W. L. Zhang, H. P. Lee, W. Z. Lin, and L. J. Cao, "Optimal partition algorithm of the RBF neural network and its application to financial time series forecasting," *Neural Computing and Applications*, vol. 14, no. 1, pp. 36–44, 2005.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

