Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2021, 8(3):145-151**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Evaluation and Prediction of Renewable Energy Development Based on Data Mining

Siyi Gu[1], Dingzhe Xie[1], Yu Zhang[2], Yanyu Ding[3], Dan Liu[4]

[1]School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, 201620, China
[2]School of Management Studies, Shanghai University of Engineering Science, Shanghai, 201620, China
[3]School of Chemistry and Chemical Engineering, Shanghai University of Engineering Science, Shanghai, 201620, China
[4]School of Urban Railway Transportation, Shanghai University of Engineering Science, Shanghai, 201620, China

Abstract Renewable energy is an important part of the energy system, and the development and utilization of energy are always the main part of any economy. To evaluate and predict the renewable energy development, we construct and solve the corresponding comprehensive measurement model and time evolution model, thus conducting horizontal and vertical comparative analysis. In the process of evaluation and prediction, we resort to these methods of data processing and statistical analysis, such as entropy weight method, grey prediction, multivariable linear regression method and PCA. The evaluation results show that California uses the new renewable energy source most widely, while Arizona and New Mexico rely heavily on non-renewable energy sources. The prediction results suggest continuing enhancing the usage rate of renewable energy although non-renewable resources will still have to play an irreplaceable role in production in the near future. Based on the numerical results and analysis, we give several suggestions for renewable energy development.

Keywords Renewable energy development, Comprehensive evaluation, GM(1,1), Data mining

1. Introduction

Renewable energy is an important part of the energy system and an energy resource that is conducive to the harmonious development of man and nature [1]. Energy development and utilization plays an important role in economic development. Renewable energy utilization can be divided into three categories: electricity, heating and cooling fuel, and transportation fuel [2]. At present, hydropower is the dominant source of renewable energy, followed by wind energy and solar power. Others such as biomass power generation, geothermal power generation, solar thermal power generation and ocean energy generation are only in a secondary position. Due to the clean and sustainable nature of renewable energy, renewable energy is being widely used to solve the most ideal energy for energy security, economic development, climate change and environmental pollution [2,3]. The development of renewable energy is entering the fast lane, and the production and use of renewable energy has made unprecedented progress.

In order to make better use of renewable energy, it is necessary to investigate the use and development of renewable energy [4-7]. Considering similarities and differences (e.g. geography, industry, population, and climate),we choose the four representative states in the United States, i.e., California (CA), Arizona (AZ), New Mexico (NM) and Texas (TX). The data source is the State Energy Data System (SEDS) Complete Dataset [8]. In the following, we do empirical research on evaluation and prediction of renewable energy development based on data mining.

2. Outline of the Approach

To evaluate and predict the renewable energy development of these four states, we will construct and solve the corresponding comprehensive measurement model and time evolution model, thus conducting horizontal and vertical comparative analysis. The following is a brief introduction to the data analysis and data estimation methods used in this paper.

2.1. Entropy method

In information theory, entropy is a measure of uncertainty. In the light of the features of entropy, the entropy method [4] can be used to evaluate the use of renewable energy. First of all, we will select 15 factors which can describe the use of cleaner, renewable energy in 2009, see Figure 1. Each factor is given an unique number s_{ij} which represents the jth factor of the ith state. In order to normalize these factors, we distinguish the positive and negative factors. For the characteristics of extremely large data and extremely small data, a normalized publicity is selected. Positive factors are normalized as follows.

$$s_{ij}^{'} = \frac{s_{ij} - \min\{s_{1j}, s_{2j}, s_{3j}, s_{4j}\}}{\max\{s_{1j}, s_{2j}, s_{3j}, s_{4j}\} - \min\{s_{1j}, s_{2j}, s_{3j}, s_{4j}\}} \tag{1}$$

Negative factors are normalized below.

$$s_{ij}^{'} = \frac{\max\{s_{1j}, s_{2j}, s_{3j}, s_{4j}\} - s_{ij}}{\max\{s_{1j}, s_{2j}, s_{3j}, s_{4j}\} - \min\{s_{1j}, s_{2j}, s_{3j}, s_{4j}\}} \tag{2}$$

The normalized data also recorded as s_{ij} .Then, calculate each state’s proportion of each factor and measure the entropy of them.

$$p_{ij} = \frac{s_{ij}}{\sum_{i=1}^{4} s_{ij}}, \qquad i = 1,2,3,4, \;\; j = 1,2...,15 \tag{3}$$

$$e_j = -k \sum_{i=1}^{4} p_{ij} \ln(p_{ij}) \tag{4}$$

Finally, calculate the weight of each factor and the overall score of the states by using MATLAB.

$$d_j = 1 - e_j \tag{5}$$

$$w_j = \frac{d_j}{\sum_{j=1}^{15} d_j}, \qquad j = 1,2,...,15 \tag{6}$$

$$s_i = \sum_{j=1}^{15} w_j \cdot p_{ij}, \qquad i = 1,2,3,4 \tag{7}$$

2.2. Gray prediction method

Gray models (GM) establish a gray differential prediction model through a small amount of data to make a fuzzy long-term description of the future growth. GM(1,1) is one of the gray prediction model [5,9].

Establish GM(1,1) model for data $x^{(0)} = \left\{x^0(1), x^0(2), \cdots, x^0(n)\right\}$.

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \tag{8}$$

The result is

$$x^{(1)}(t) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-a(t-1)} + \frac{b}{a} \tag{9}$$

The predicted data is got by the equations below.

$$x^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{b}{a}\right]e^{-ak}+\frac{b}{a} \qquad k=1,2,\cdots,(n-1) \tag{10}$$

$$x^{(0)}(k+1)=x^{(1)}(k+1)-x^{(1)}(k) \qquad k=1,2,\cdots,(n-1) \tag{11}$$

2.3. Multivariable linear regression method

Multivariable linear regression (MLR) analysis [10] can accurately measure the degree of correlation between the various factors and the level of regression fit, improve the prediction of the effect of the equation. In many cases, one variable is affected by multiple variables. To ensure the model's excellent explanatory and predictive effect, the independent variables must have a significant impact on dependent variables. What's more, the selection of independent variables should have a certain degree of mutual exclusion.

$$y=B_0+B_1t_1+\cdots+B_kt_k+\varepsilon \tag{12}$$

B_k represents the parameter of the model while ε represents deviation part.

As for other two methods, that is, principal component analysis (PCA) and TOPSIS, we only give a brief explanation, please refer to the literatures [4,9-11] for details. The PCA method eliminates the influence among the evaluation indexes and reduces the workload of the index selection. There are a large number of indexes in this data, so in this way we can retain most of the information and use a few comprehensive indexes to replace the original Indicators for analysis. It is more convenient in the computer to achieve. On the other hand, the TOPSIS method can fully extract the characteristics of the original data and reflect the relationship between energy production and consumption to a large extent, and has the characteristics of reliability, intuition and high credibility.

3. Data processing

The data file of State Energy Data System (SEDS) Complete Dataset through 2009 provides 50 years of data in 605 variables on each of these four states' energy production and consumption, along with some demographic and economic information. It is a very large dataset. In order to extract important data and simplify data analysis to obtain representative conclusions, we reduce the dimension by manual and PCA selection. The 28 variables are screened from 605 ones, such as natural gas, biomass, hydroelectricity and so on. Depending on the type of data and the relationship among the data according to the additional information, we have selected the most representative ones of the total energy consumption of various energy sources as a basis for data analysis. Here, New Mexico is used as an example to show the results of data analysis of energy profiles, including the different types of energy use and supply and the evolution of related data over time, see Table 1.

Table 1: Energy profile of New Mexico

New mexico	energy / year		2000	2005	2009	rate of change(%)	
						2000-2005	2005-2009
unrenewable energy	Fossil fuels	Coal	305512.9	317916.1	306161.5	1.0	-0.9
		Distillate fuel oil	69533.8	83709.5	74505.9	4.7	-2.9
		Jet fuel	17109.0	12941.8	7587.3	-6.7	-12.5
		Kerosene	164.5	49.1	7.7	-26.1	-37.1
		LPG	10301.0	10286.8	22673.0	0.0	21.8
		Motor gasoline	110697.9	120085.3	120780.5	2.1	0.1
		Natural gas	259018.0	225352.0	247119.6	-3.4	2.3
	Electricity (nuclear power)		0.0	0.0	0.0	/	/
	Electricity		64147.9	70420.1	73860.0	2.4	1.2
renewable energy	Biomess		7323.2	13020.1	17295.2	15.5	7.4
	Geothermal energy		687.9	690.1	317.1	0.1	-17.7
	Hydroeletricity		2256.0	1649.8	2644.6	-7.5	12.5
	Photovoltaic and solar thermal		452.4	208.8	282.5	-17.6	7.8
	Wood and waste		4515.9	10807.8	11633.2	24.4	1.9
	Electricity produced from wind		0.0	7945.5	15096.0	/	17.4
	Fuel ethanol		2212.5	1044.8	4114.8	-17.1	40.9

The state of New Mexico is the only state in the four states that has no nuclear power. Non-renewable energy consumption is mainly consumed by coal and natural gas. The consumption of both has remained basically stable in the past ten years. The use of renewable energy is much lower than that of California and Arizona. New Mexico's biomass production can grow rapidly in a decade. Geothermal energy consumption decreased by 17.7% from 2005 to 2009. In the 21st century, New Mexico developed wind power and production and consumption grew rapidly.

4. Evaluation model of renewable energy development

The current energy use ratio must be between the excellent clean energy use and the traditional non-renewable energy use. To provide an inductive analysis of the level of clean energy use in the four states, an entropy approach can be used to analyze the current level of clean energy use in order to find out which of the four states appeared to have the "best" profile for use of cleaner. In addition, this method can also be used to predict future renewable energy consumption.

Based on the above data processing, we select 15 factors which can describe the use of cleaner, renewable energy, and establish a comprehensive index system of renewable energy, see Figure 1.

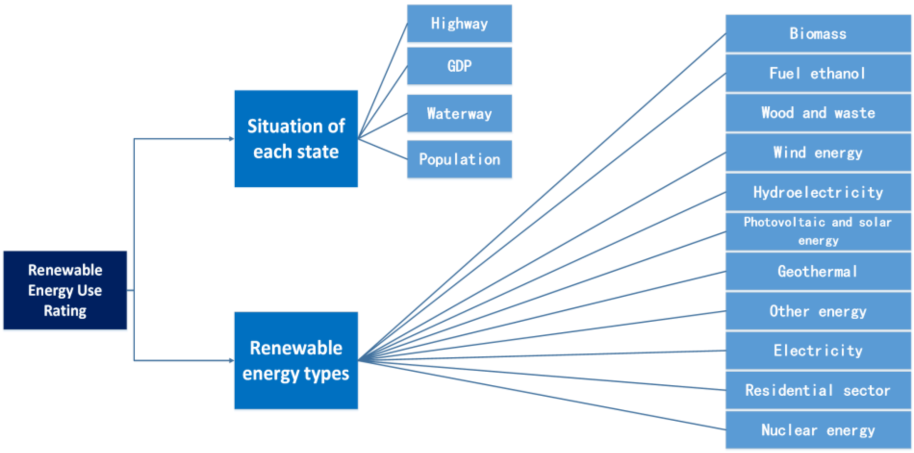

Figure 1: Comprehensive index system of renewable energy

After calculated by the entropy method, the weight of factors are described as Table 2 shows.

Table 2: Entropy weight

Factor	Weight	Factor	Weight
Biomass	0.0549	Nuclear energy	0.0295
Electricity	0.0639	Wood and waste	0.0718
Fuel ethanol	0.0442	Wind energy	0.0663
Geothermal	0.1295	Other energy	0.0575
Residential sector	0.076	Highway	0.0707
Hydroelectricity	0.0809	Waterway	0.0589
Photovoltaic and solar energy	0.0929	GDP	0.0533
		Population	0.0496

Table 3: Entropy scores

State	AZ	CA	NM	TX
Score	0.070	0.639	0.010	0.282

The overall scores are shown in Table 3. California holds the highest score among four states. So California has the "best" profile for use of cleaner, renewable energy in 2009 while the scores of Arizona and New Mexico are very poor, which means Arizona and New Mexico rely much on unrenewable energy. The result agrees with the conclusion in [12]. In particular, New Mexico has made energy exports an important industry. The scores of the TOPSIS analysis are also calculated from 1963 to 2009, see Figure 2. California presents the best performance in new energy usage, followed by Texas. This confirms the results of the entropy weight analysis.

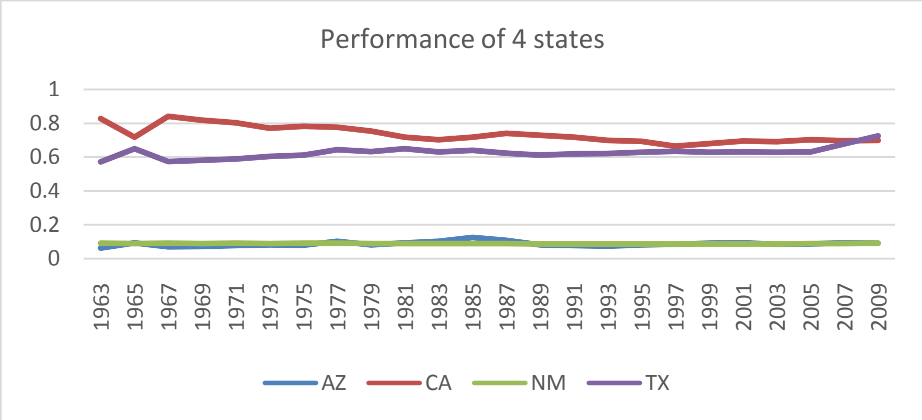

Figure 2: TOPSIS scores in using renewable energy

5. Time series prediction model of renewable energy development

There are many types of energy consumption in the table, which is equivalent to a large number of independent variables. The main independent variables can be found through MLR. It will occupy an important part in energy production and use. Long-term simulations of energy use are predicted by using grey predictions. According to the evolution of the usage of energy, each state shows different growth trend and potentiality, see Tables 4-7.

Table 4: Predictions of energy growth rate of California

CA	Year			Rate of change (%)	
Energy	2025	2035	2050	2025-2035	2035-2050
Fossil Energy	6005899.4	5919799.336	5813537.9	-0.16	-0.13
Electricity	1325681.12	1726372.791	2498666.5	2.98	2.68
Nuclear Energy	358271.688	362480.948	368794.56	0.13	0.12
New Energy	1066529.69	1231150.476	1526927.5	1.61	1.55
Total Energy	9727037.44	10401769.27	11478061	0.75	0.71

Table 5: Predictions of energy growth rate of Arizona

AZ	Year			Rate of change (%)	
Energy	2025	2035	2050	2025-2035	2035-2050
Fossil Energy	1977563.89	2428829.291	3030138.8	2.31	1.59
Electricity	460899.955	741532.2173	1442988	5.43	4.87
Nuclear Energy	337655.434	357959.3399	388405.31	0.65	0.58
New Energy	143997.899	160681.0617	189398.92	1.23	1.18
Total Energy	2451271.33	2835580.051	3500472.8	1.63	1.52

Table 6: Predictions of energy growth rate of New Mexico

NM	Year			Rate of change (%)	
Energy	2025	2035	2050	2025-2035	2035-2050
Fossil Energy	870587.038	910148.3611	969079.04	0.49	0.45
Electricity	130418.156	198390.0281	356922.84	4.77	4.28
Nuclear Energy					
New Energy	37718.7125	58912.36652	114995.51	5.08	4.89
Total Energy	823997.924	904847.6076	1024409.8	1.05	0.89

Table 7: Predictions of energy growth rate of Texas

TX	Year			Rate of change (%)	
Energy	2025	2035	2050	2025-2035	2035-2050
Fossil Energy	9362841.03	8548369.758	7309345.3	-1.01	-1.11
Electricity	2167492.62	3218269.577	5596971.9	4.49	4.03
Nuclear Energy	734268.713	913622.9524	1181718.2	2.46	1.85
New Energy	374339.357	537677.9773	925565.99	4.11	3.96
Total Energy	14040431.3	15309814.72	17167729	0.97	0.82

California will see the fastest growth in electricity consumption, followed by new energy resources. In the future, the consumption of fossil fuels will keep negative growth, dropping to 581,135,379.9 Billion Btu in 2050.As the state that uses the most renewable energy, the proportion of traditional energy use in California is declining. The use of nuclear energy in **New Mexico** remained at zero for the same energy policy. Electricity and new energy are the main components of increased energy consumption. It is estimated that the use of new energy sources will increase from 37718.7125 Billion Btu in 2025 to 114995.51Billion Btu in 2050.**Texas**, as a state that actively uses renewable energy, the use of fossil fuels is expected to decline by about 1% in the future. Texas is expected to reduce its fossil energy consumption to 7309245.35Billion Btu in 2050. As to **Arizona**, it is estimated that the average annual growth rate of electricity consumption in Arizona will be 5.43% between 2025 and 2045 and 4.87% between 2035 and 2050.

6. Analysis of renewable energy usage targets and actions

At the level of specific types of clear and renewable energy, the four states shows different characteristics. For example, the usage of geothermal energy in California rise at the rate of 4.24% from 2025 to 2035. California is located in the Pacific Rim, which is rich in geothermal energy. So developing and using therenewable energy can bring both economic and environmental benefits for the state.

- Balance four states energy consumption by transmitting electricity from one to another.

Table 8: Energy transmission

State	Rate of change (%)	
	2025-2035	2035-2050
AZ	0.94	0.84
CA	0.65	0.59
NM	0.28	0.25
TX	1.51	1.36

Because of the shortage of adequate power transmission facilities, the conversion of abundant renewable energy into electricity cannot be effectively transported, making it a major obstacle to the development of renewable energy [12]. Power loss is also a serious problem in energy transportation. Carrying out this action can help improve the level of energy infrastructure construction. Table 8 shows the preferred growth rate in transmitting electricity. Texas, as a longtime energy exporter, transmit more electricity can help promote other states' development and gain profits.

- Continue enhancing renewable energy using rate [3,13].

According to the relationship between the energy usage and the environment, these four states should enhance the rate of renewable energy consumption. The goal is displayed in Figure 3.At the same time, promote and monitor the use of clear and renewable energy through subsidy and penalty policy [11-14].

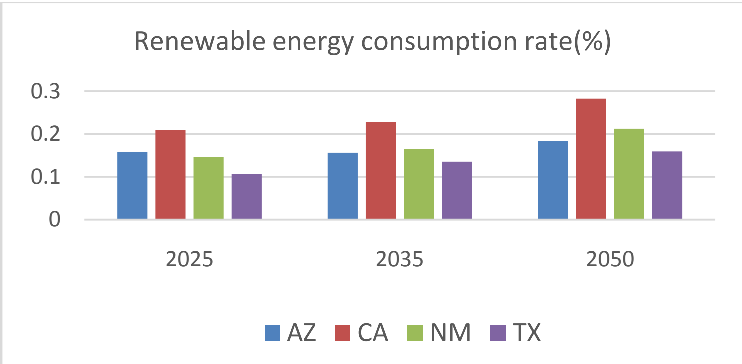

Figure 3: Predictions of renewable energy consumption rate

- Reduce the cost in using renewable energy through scientific research and technological innovation [1,2,8].

The main obstacle of the large-scale promotion of renewable energy is the high cost. A survey shows that in order to produce one kilowatt hour electricity, solar energy cost 30 cents while fossil fuel only cost 5.5 cents. So the government tends to use more unrenewable energy. With the development of science and technology, it is sure that the cost of renewable energy will be significantly reduced in the near future, and it will be more widely used.

7. Conclusions

We did case studies and empirical analysis on evaluation and prediction of renewable energy development. The evaluation results show that California presents the best performance in renewable energy usage among the four states. As we predicted, total energy consumption in 2050 will increase by 140% in Arizona, compared with 1.43 times in California than that in 2009, and about 150% in both New Mexico and Texas. The prediction results also suggest continuing enhancing renewable energy using rate. Through our research, we can help to understand the similarities and differences in the use of energy and renewable energy in the four states, and promote the cooperation of relevant departments in energy production and use in the future.

Acknowledgments

This work was supported by Municipal Undergraduate Innovation Training Programof SUES under Grant Nos. cs2021005, cs2021002. We would like to thank our instructor, Zhongtuan Zheng, for constructive comments and suggestions on this work.

References

[1]. BP, Statistical Review of World Energy. https://www.bp.com/zh_cn/china/reports-and-publications.html
[2]. REN21, Renewables 2017 Global Status Report http://www.ren21.net/status-of-renewables/global-status-report/
[3]. Jay Squalli. Renewable energy, coal as a baseload power source, and greenhouse gas emissions: Evidence from U.S. state-level data [J]. Energy, 2017, 127:479-488.
[4]. Long Zhang. Research on performance evaluation of renewable energy development and utilization [D]. China University of Geosciences, 2017.
[5]. Bing Wang. Renewable energy system risk assessment method and its application [D]. Beijing Institute of Technology, 2016.
[6]. Pei Zhang, Huang Hao.An overview of US new energy development [J]. Power SystemTechnology, 2011, 35(7): 48-53.
[7]. Huang Zhou. Status of renewable energy development and energy conservation in America[J]. China Foreign Energy, 2007, 12(1): 1-6.
[8]. U.S. Energy Information Administration (EIA), State Energy Data System (SEDS). https://www.eia.gov/state/seds/seds-data-complete.php?sid=US; Analysis and Projections. https://www.eia.gov/renewable/.
[9]. Yudong Guo. Research on the coordinated development of energy consumption and economic growth-ecological environment [D]. China University of Geosciences, 2015.
[10]. James G, Witten D, Hastie T, et al. An introduction to statistical learning: with Applications in R [M]. New York: Springer, 2013.
[11]. Fang Zhao. Research on energy development policy based on 3E coordination [D], Ocean University of China, 2008.
[12]. Xuxuan Xie, Tian Wang, Dongming Ren. The latest development of renewable energy quota system in the United States and its enlightenment to China [J]. China Energy, 2012, 34(3): 33-37.
[13]. Huan Le. US energy policy research [D]. Wuhan University, 2014.
[14]. Qiankun Wang, Qionghui Li, Guohui Xie. Renewable portfolio standard in California andits enlightenment for China [J]. Sino-Global Energy, 2012, 17(9): 25-31.