Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2019, 6(2):126-133**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Modeling and Analysis of Energy and Renewable Energy Profiles Based on SEDS Dataset

Wenwei Shen, Xinying Peng, Wei Chen, Wenbo Liu

School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China

Abstract Based on State Energy Data System (SEDS) Complete Dataset through 2009, energy and renewable energy profiles of four US states are analyzed through mathematical modeling. The four states are California (CA), Arizona (AZ), New Mexico (NM) and Texas (TX). First, we establish a composite index of energy to describe the energy profile for each states. Considering the integrity and redundancy of the data, we do data cleaning and filtering by manual and principal component analysis (PCA) selection. The 28 variables are screened from 605 ones, such as natural gas, biomass, hydroelectricity and so on. We focus on different types of energy use and supply, as well as energy consumption in different sectors. The basic energy profile is available through the datasheet. Then, to understand the similarities and difference between the four states, we develop a time evolution model of the energy profile. Through PCA attached other factors, the indicators with similar characteristics are combined to replace the large data. Finally, by the TOPSIS method, we find that, California uses the new renewable energy source most widely, while New Mexico relies heavily on non-renewable energy sources. Empirical research is performed on the analysis of energy development, which may help to exploit and utilize energy better, especially renewable energy.

Keywords Renewable energy, Data analysis, Principal component analysis, TOPSIS

1. Introduction

The production and use of energy are always the main part of any economy [1]. With the increasing pollution problem and the decrease of the non-renewable traditional energy sources, the clean and renewable new energy has gradually become the goal of energy development in all countries in the world [2]. In 1970, 12 western states in the U.S. formed the Western Interstate Energy Compact (WIEC) [3]. These states agree on specific policy issues and adopt a set of standards. California (CA), Arizona (AZ), New Mexico (NM) and Texas (TX) are along the U.S. border with Mexico. In the following, we take the four states to conduct empirical research on the analysis of energy development. This may help to understand the similarities and difference of energy and renewable energy profiles between them, and promote mutual cooperation on energy production and usage.

2. Data analysis of energy profile

The SEDS Complete Dataset through 2009 [4] 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. Data processing is needed before data analysis of energy profile, for it is a very large total of about 100,000 data.

2.1. Data sorting

First, we have the data in sheet seseds divided into four parts according to the four states, coupled with sheet msncodes data using excel "vlookup" function with the year as the abscissa and the consumption type as the ordinate. We have got four two-dimensional forms, and the nonexistent items are empty. Then, we carry out data filtering before data analysis by manual and PCA selection [5].

In order to extract the important data and simplify the data analysis, we reduce the dimension of this large amount of data set. The standard data cleanup model is to enter data into the data cleanup processor of "dimensionality reduction", clean up the data in a series of steps and then output the cleansed data in the desired format. Data cleaning use statistical methods to identify possible false values or outliers. After processing, the incomplete data in the table is deleted. Depending on the type of data and the relationship among the data according to the additional information, we have selected the most representative of the total energy consumption of various energy sources as a basis for data analysis and summarization.

Table 1: Composite classification of energy

unrenewable energy	Fossil fuelS	Petroleum products	Asphalt and road oil
			Aviation gasoline
			Distillate fuel oil
			Jet fuel
			Kerosene
			LPG
			Lubricants
			Motor gasoline
			Residual fuel oil
			Petroleum coke
		Coal	
		Natural gas	
		Supplemental gaseous fuels	
	Electricity (nuclear power)		
	Electricity		
renewable energy	Biomass		
	Geothermal energy		
	Hydroelectricity		
	Photovoltaic and solar thermal energy		
	Wood and waste		
	Electricity produced from wind		
	Fuel ethanol		

2.2. Data analysis of different types of energy use and supply

To observe each state's energy profile, the data is first divided into two parts of renewable and non-renewable energy from the years of 2000, 2005 and 2009.

California In fossil fuel use of California, natural gas has the largest proportion and kerosene accounts for the smallest proportion. From 2005 to 2009, its fossil fuel consumption decreased to varying degrees. California is clearly ahead of other states in the use of renewable energy. Biomess , geothermal energy, hydroeletricity , wood and waste all reached six digits (Billion Btu). Other minimal consumption also reached five digits (Billion Btu). In particular, hydroelectric power consumption reached about 390 million Billion Btu. Second, the total biomass consumption also exceeded 200,000 Billion Btu. From 2005 to 2009, Photovoltaic and solar thermal energy increased by 9.3% and Electricity produced from wind energy increased by 7.5%.

Arizona In Arizona, The consumption of non-renewable energy is dominated by coal consumption, it consumes about 420000 Billion Btu. While, it consumes a very small amount of kerosene. In 2009 it even had only 8.2 Billion Btu in the consumption of kerosene. In the consumption of renewable energy, the consumption of hydropower accounts for about 70%. In 2009, Arizona had wind power for the first time, consuming 288.4 Billion Btu. Ethanol fuel consumption increased rapidly between 2000 and 2005, increasing by 75.4%.

Texas Texas is the nation's largest producer of lignite coal. Texas is a major oil producer. Natural gas consumption has reached about 4000000 Billion Btu. However, between 2000 and 2009, natural gas consumption maintained a sustained and slow decline. Texas's consumption of wind power is up 70.4% from 2000 to 2005 and 46.6% from 2005 to 2009. Fuel ethanol also increased by 62.3% and 15.4% respectively. Texas maintains rapid growth in the development of both.

Table 2: Energy profile of California

California	energy year		2000	2005	2009	rate of change(%)	
						2000-2005	2005-2009
unrenewable energy	Fossil fuels	Coal	70011.7	67422.3	52403.1	-0.9	-6.1
		Distillate fuel oil	544379.1	564453.1	526665.1	0.9	-1.7
		Jet fuel	584018.1	593147.8	555575.4	0.4	-1.6
		Kerosene	2104.3	2291.2	1142.5	2.2	-16.0
		LPG	45297.1	44797.2	60934.4	-0.3	8.0
		Motor gasoline	1786458.2	1989629.7	1866541.0	2.7	-1.6
		Natural gas	2456438.2	2304462.9	2391376.8	-1.6	0.9
	Electricity (nuclear power)		366845.3	377312.5	332249.4	0.7	-3.1
	Electricity		832723.2	867499.3	885699.3	1.0	0.5
renewable energy	Biomess		164988.7	225495.2	224662.5	8.1	-0.1
	Geothermal energy		127599.7	132375.4	127461.1	0.9	-0.9
	Hydroeletricity		391043.0	396279.0	272187.2	0.3	-9.0
	Photovoltaic and solar thermal		23120.8	22035.1	31397.0	-1.2	9.3
	Wood and waste		159180.8	145587.5	140159.2	-2.2	-0.9
	Electricity produced from wind		35887.4	42618.0	56996.6	4.4	7.5
	Fuel ethanol		5513.8	79010.3	81718.3	94.6	0.8

Table 3: Energy profile of Arizona

Arizona	energy year		2000	2005	2009	rate of change(%)	
						2000-2005	2005-2009
unrenewable energy	Fossil fuels	Coal	432810.0	428444.8	413259.9	-0.3	-0.9
		Distillate fuel oil	116054.2	151044.2	143372.9	6.8	-1.3
		Jet fuel	59155.7	45463.3	32288.3	-6.4	-8.2
		Kerosene	31.1	51.5	8.2	13.4	-36.8
		LPG	5987.9	5048.6	7318.5	-4.2	9.7
		Motor gasoline	294006.1	352127.7	331839.1	4.6	-1.5
		Natural gas	208137.1	329301.7	376674.8	12.2	3.4
	Electricity (nuclear power)		316839.0	269326.5	320723.0	-4.0	4.5
	Electricity		208575.7	236761.0	250553.2	3.2	1.4
renewable energy	Biomess		13603.4	25121.7	35412.8	16.6	9.0
	Geothermal energy		286.2	282.9	329.1	-0.3	3.8
	Hydroeletricity		85221.4	64094.2	62730.9	-6.9	-0.5
	Photovoltaic and solar thermal		3515.0	2990.7	4732.1	-4.0	12.2
	Wood and waste		12150.5	11375.9	12867.3	-1.6	3.1
	Electricity produced from wind		0.0	0.0	288.4	0.0	/
	Fuel ethanol		1452.9	13745.8	19451.0	75.4	9.1

Table 4: Energy profile of Texas

Texas	energy year		2000	2005	2009	rate of change(%)	
						2000-2005	2005-2009
unrenewable energy	Fossil fuels	Coal	1548184.3	1627863.3	1497854.3	1.3	-2.1
		Distillate fuel oil	651512.1	744862.2	769735.1	3.4	0.8
		Jet fuel	582403.4	455765.4	350450.6	-5.9	-6.4
		Kerosene	1832.8	1690.0	615.5	-2.0	-22.3
		LPG	1466385.9	1496823.6	1489198.5	0.5	-0.1
		Motor gasoline	1301556.3	1452431.1	1510370.3	2.8	1.0
		Natural gas	4550105.7	3625116.4	3462210.4	-5.5	-1.1
	Electricity (nuclear power)		391669.5	398994.3	434065.1	0.5	2.1
	Electricity		1085911.7	1140489.2	1178148.5	1.2	0.8
renewable energy	Biomess		87162.1	81539.6	148263.6	-1.7	16.1
	Geothermal energy		562.0	1161.0	2057.0	19.9	15.4
	Hydroeletricity		8456.3	13324.3	10039.7	12.0	-6.8
	Photovoltaic and solar thermal		567.5	563.6	819.8	-0.2	9.8
	Wood and waste		81740.0	80147.9	72108.5	-0.5	-2.6
	Electricity produced from wind		5020.4	42367.9	195454.8	70.4	46.6
	Fuel ethanol		5422.2	37590.0	66731.0	62.3	15.4

Table 5: Energy profile of New Mexico

New mexico	energy / year		2000	2005	2009	rate of change(%) 2000-2005	rate of change(%) 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

2.3. Data analysis of energy consumption in different sectors

We extracted the total energy consumption of each state's industrial sector, the commercial sector, the residential sector, the transportation sector, and the electric power sector to making a line chart [6]. We can visually see the energy consumption of one state's different sector. The figure below shows the energy profile for California.

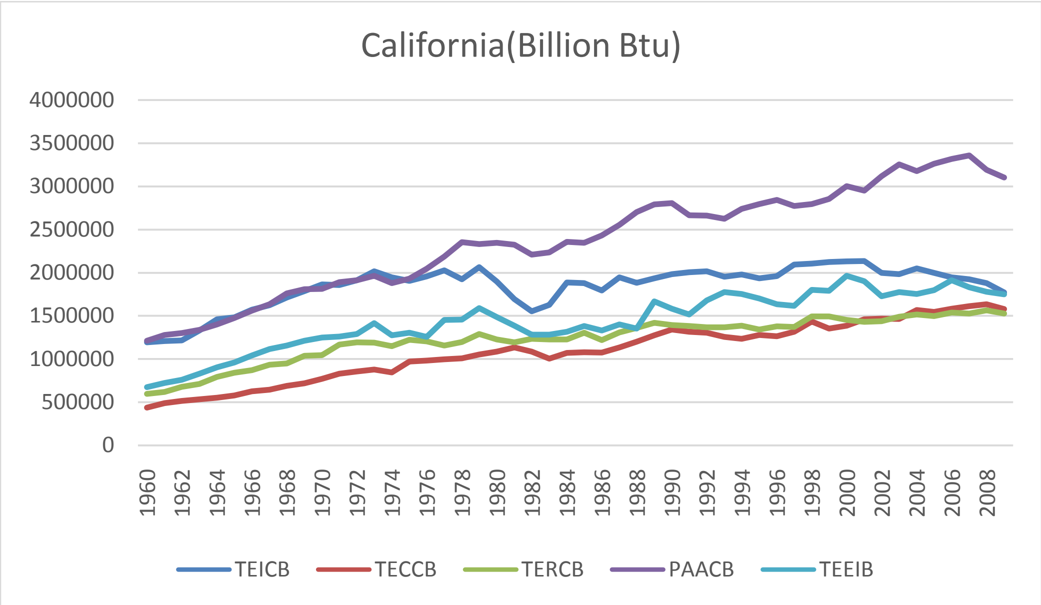

Figure 1: Energy growth of California

California The total energy consumption of the five selected parameters in 2009 is 9728241.747 Billion Btu. The industrial sector accounted for 18.19%. The commercial sector accounted for 16.23%. The residential sector accounted for 15.70%. The transportation sector accounted for 31.87%. The electric power sector accounted for 17.98%.

The transportation sector accounted for the largest proportion in California's total energy consumption. The commercial sector usually accounted for the smallest proportion. But it seemed to surpass The energy consumption of the residential sector after 2000. The energy consumption of the industrial sector remained stable from 1984 to 2009. It is about 2000000 Billion Btu. All data showed a downward trend after 2006

Arizona The total energy consumption of the five selected parameters in 2009 is 2491456.496 Billion Btu. The industrial sector accounted for 8.34%. The commercial sector accounted for 14.12%. The residential sector accounted for 16.09%. The transportation sector accounted for 18.19%. The electric power sector accounted for 42.25%.

In Arizona, the consumption of the electric power sector had the fastest growth in the past. Its value is more than twice the value of the second(the transportation sector).The maximum reached about 1100000 Billion Btu. The consumption of the industrial sector has been stable at around 200000 Billion Btu.At the same time, it became the minimum after 1998.

New Mexico The total energy consumption of the five selected parameters in 2009 is 1055839.718 Billion Btu. The industrial sector accounted for 21.53%. The commercial sector accounted for 11.59%. The residential sector accounted for 11.17%. The transportation sector accounted for 18.09%. The electric power sector accounted for 37.42%.

The consumption of the industrial sector fluctuated around 200000 Billion Btu. The electric power sector has consumed more and more energy very fast. It reached 400000 Billion Btu after 2009. In 2009 the industrial sector was the second largest part. The commercial sector and the residential sector maintain a synchronous growth. They reached about 120000 Billion Btu in 2009.

Texas The total energy consumption of the five selected parameters in 2009 is 14757336.96 Billion Btu. The industrial sector accounted for 37.28%. The commercial sector accounted for 9.89%. The residential sector accounted for 10.92%. The transportation sector accounted for 17.79%. The electric power sector accounted for 24.20%.

The consumption of the commercial sector, the residential sector, the transportation sector, the electric power sector maintain very stable growth throughout historical data. The industrial sector is still the largest part of energy consumption but it is showing a downward trend after 1998 when it almost reached 7000000 Billion Btu. Now The industrial sector has dropped to 5500000 Billion Btu.

3. Modeling and analysis of energy and renewable energy profiles

3.1. Outline of the approach

Principal component analysis method The principal component analysis method (PCA) is a statistical method. Through this method's orthogonal transformation, a set of variables that may have correlation is transformed into a set of linearly uncorrelated variables.

Import raw data matrix X , standardized the matrix into R .

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{1p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (1)$$

$$r_{ij} = Cov(x_i, x_j) = \frac{\sum_{k=1}^{n}(x_i - \overline{x}_i)(x_j - \overline{x}_j)}{n-1} \qquad n > 1 \quad (2)$$

Calculate the eigenvalues of the matrix R and the corresponding eigenvectors. Then reach the contribution rate.

$$\text{cr} = \frac{\lambda_i}{\sum_{i=1}^{p}\lambda_i} \quad (3)$$

Finally, calculate the score of the main component.

$$F = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1k} \\ f_{21} & f_{22} & \cdots & f_{1k} \\ \vdots & \vdots & & \vdots \\ f_{n1} & r_{n2} & \cdots & f_{nk} \end{bmatrix} \quad (4)$$

$$f_{ij} = a_{j1}x_{i1} + a_{j2}x_{i2} + \cdots + a_{jp}x_{ip} \qquad i = 1, 2, \cdots, n \qquad j = 1, 2, \cdots, k \quad (5)$$

Continue modeling on the base of these statistics.

TOPSIS method Technique for Order Preference by Similarity to an Ideal Solution（TOPSIS）method [7] is a multi-criteria decision analysis method. Its main idea is transform the solution to distance. The chosen solution should have the shortest geometric distance from the positive ideal solution (PIS) and the longest geometric distance from the negative ideal solution (NIS). In this task, the PIS represents the using of clear, renewable energy while the NIS represents unrenewable energy. Through the use of PCA, we can obtain essential data. The data processing method is described below.

Firstly, create the decision matrix $A=(a_{ij})_{m\times n}$. In order to remove the dimension effect, standardized processing is essential.

$$b_{ij}=\frac{a_{ij}}{\sqrt{\sum_{i=1}^{m}a_{ij}^2}},\quad i=1,\cdots,m,\quad j=1,\cdots,n \tag{6}$$

Compare the contribution of each factor, form different weights. The matrix after calculation is $C=(c_{ij})_{m\times n}$. Then, determine the PIS and NIS.

d_i^1 represents the distance between the factor and the PIS while d_i^0 represents the factor and the NIS.

$$d_i^1=\sqrt{\sum_{j=1}^{n}(c_{ij}-c_j^1)^2}\quad i=1,2,\cdots,m \qquad d_i^0=\sqrt{\sum_{j=1}^{n}(c_{ij}-c_j^0)^2}\quad i=1,2.\cdots,m \tag{7}$$

Calculate the distance d_i^1, d_i^0 and the evaluation reference value f_i .

$$f_i=\frac{d_i^0}{d_i^0+d_i^1}\qquad i=1,2,\cdots,m \tag{8}$$

Sort f_i from large to small, the largest f_i represents the most use of clear and renewable energy. The value shows the evolution of the usage of clear, renewable energy from 1960 to 2009.

3.2. Time evolution model of energy and renewable energy profiles

In order to find out how the energy profile of each of the four states has from 1960–2009, The main components need to be extracted from the cleaned data, the relevant data should be replaced by fewer groups, the principal component analysis can be used in the process. Analyzed data can be divided into renewable energy and non-renewable energy.

We acquire the results of PCA, the picture is displayed below.

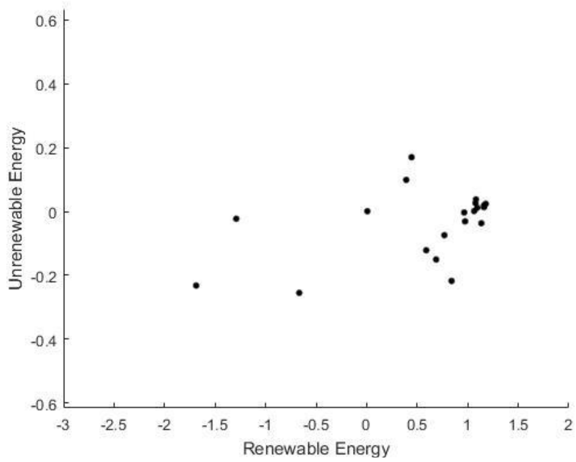

Figure 2: Result of principal component analysis method

According to the analyzing results, most of the factors tend to show greater impact on the usage of clear, renewable energy. Only few factors show significant impact on unrenewable energy consumption.

Table 6: Scores of the TOPSIS analysis

	1965	1969	1973	1977	1981	1985	1989	1993	1997	2001	2005	2009
AZ	0.0923	0.0728	0.0809	0.1015	0.0923	0.1247	0.0808	0.0746	0.0853	0.093	0.0864	0.0902
CA	0.7181	0.8184	0.7699	0.7766	0.7181	0.7182	0.728	0.6982	0.6642	0.6948	0.703	0.6981
NM	0.0893	0.0892	0.0886	0.0904	0.0893	0.0888	0.0871	0.0871	0.0864	0.0871	0.0879	0.0907
TX	0.6483	0.5809	0.6037	0.6427	0.6483	0.6398	0.6108	0.6212	0.6346	0.6296	0.6296	0.7249

We calculated the data from 1963 to 2009 by using the model. In order to list the table so we selected half of the data, however, the line chart is based on the complete data.

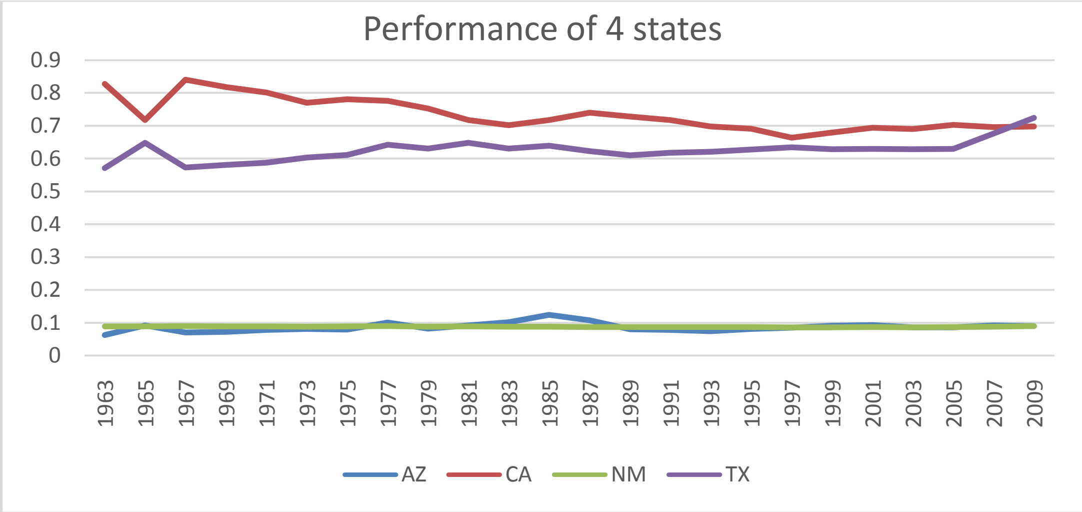

Figure 3: Performance in using renewable energy

From the line chart, we can find that California is the best performer in new energy usage, followed by Texas. Texas has exceeded California. We have found the reason by looking up the energy usage data that Texas electricity aspect has shifted from export to import.

4. Conclusions

Taking four states in the United States, we conducted empirical analysis of energy and renewable energy development. We made energy classification, and then performed data analysis on different types of energy use and supply, as well as energy consumption in different sectors. Finally, by the PCA and TOPSIS methods, we constructed and solved the time evolution model of energy and renewable energy profiles.

We have got that from 1960 to 2009, although California has almost not any coal reserves or production, it produces most electricity generation from renewable resources, while Arizona and New Mexico rely too much on the non-renewable resources, where solar and wind energy are not fully developed until 2003. Meanwhile, as one of the biggest energy state, the use of non-renewable energy in Texas dropped when the proportion of renewable energy increased sharply, which means it has already realized the huge meaning of renewable energy using.

The extracted categories are much smaller than the 605 categories given, and the data errors of doing so will be more obvious. The interpretation of PCA has somewhat fuzzy, unlike the original meaning of the variable so clear, exactly. This is the price that the variable dimension has to pay. Because each state has its own background in the TOPSIS method, the weight of information also has a priori law, so the result has some subjectivity. All of this is left for future improvements and for future work.

Acknowledgments

This work was supported by Municipal Undergraduate Innovation Training Program of SUES under Grant Nos. cs1821003 and cs1821002.

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]. The Western Interstate Energy Board, October-16-1970-Public-Law-91-461-Western-Interstate-Nuclear-Compact.pdf. https://westernenergyboard.org/wp-content/uploads/2015/07/October-16-1970-Public-Law-91-461-Western-Interstate-Nuclear-Compact.pdf

[4]. U.S. Energy Information Administration (EIA), State Energy Data System (SEDS). https://www.eia.gov/state/seds/seds-data-complete.php?sid=US

[5]. James G, Witten D, Hastie T, et al. An introduction to statistical learning: with Applications in R [M]. New York: Springer, 2013.

[6]. Yudong Guo. Research on the Coordinated Development of Energy Consumption and Economic Growth-Ecological Environment [D]. China University of Geosciences, 2015.

[7]. Long Zhang. Research on Performance Evaluation of Renewable Energy Development and Utilization [D]. China University of Geosciences, 2017.