A Setting System of Interstate Energy Cooperation Goals Based on Data Insight

Summary

After performing data analysis and modeling, we finally determine a set of development goals for the new four-state energy compact.

First, we preprocess the data provided, which includes default value processing, abnormal value process and data classification. For the sake of analysis, we divide various energy into two broad categories. One is cleaner renewable energy (CRE), the other is traditional fossil energy (TFE). After that, we select 11 important variables from the given data to create the energy profile for each of the four states. We call the 11 variables the basic variables

Next, we apply the decoupling theory to characterize the dynamic relationship between economic development and energy utilization, which can reflect the evolution of energy profile. We find that the four states differ in production and usage of various energy significantly. To determine the underlying factors that lead to the differences, we construct the simultaneous equations model. Combining natural environment information further, we find out the factors and know the respective strengths of the four states in CRE.

Then, we establish a multi-dimensional evaluation system to identify the state that has the "best" energy profile on the whole. We introduce the index, comprehensive utilization performance (*CUP*) to measure the energy profile. The *CUP* is composed of three parts, energy performance, economic performance and environment performance. And each of the three parts includes three indexes respectively, all of which are synthesized by the basic variables. We use the PCA method to integrate the nine indexes into an overall index, namely the *CUP*. Ranking *CUP*, we find that California is the "best".

Finally, we construct BP neural network to predict the energy profile. Analogous to Cobb-Douglas Production Function in economics, we define the *CUP* in a new way for predicting. Through setting various development scenarios, we get the predictions successfully. After that, we regard the four states as a whole to determine renewable energy usage targets for 2025 and 2050. In this process, we use the BP neutral network and previous models again. We collect real data from 2010 to 2015 to calculate the values of *CUP*. Compare them to the predicted value, we test our predicting system. The result shows that our predicting system works well.

Contents

MEMO

From: Team 73767, MCM 2018
To: The group of Governors
Date: February 13, 2018
Subject: Goals for the interstate energy compact

Dear governors, we are honored to inform you our achievement after performing data analysis and modeling.

First, we introduce the energy profiles of your states in 2009. We use comprehensive utilization performance (*CUP*) to reflect the energy profile on the whole. Ranking the *CUP* of your four states, California appears to be the best. In terms of the usage of cleaner renewable energy, California, Arizona, Texas and New Mexico has the largest consumption successively in the water power, nuclear power, wind power and solar power. It is decided by the natural environment of your four states, such as geography and climate.

Then, we provide you the goals for the interstate energy compact. Through predicting, your four states' average *CUP* will increase to 16.98 by 2025 and to 26.05 by 2050, if your four states can cooperate adequately. Based on the predictions, we set the development goals as follows:

- Building a community of energy utilization and economic development, so as to realize manual benefit and win-win result.

- Promoting the production and usage of new energy sources and reducing the dependence on traditional fossil fuels.

- Increasing the whole economy's comprehensive utilization performance to 16.98 by 2025 and to 26.05 by 2050, and achieving the harmony of economy, energy and environment ultimately.

According to the goals above and the characteristics of each of your four states, we offer you the following suggestions to meet the goals.

- Building an energy investment bank to integrate money for developing the clear renewable energy.

- Enhancing residuals' environmental protection consciousness.

- Adjusting the industrial structure properly.

We sincerely hope that you can achieve the goals above!

Please contact us if you have any problems.
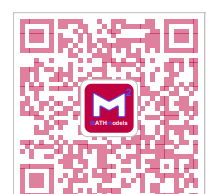
1 Introduction

1.1 Problem Statement

Energy is the main material base and driving force for human's daily production and life. The proper utilization of all kinds of energy are closely link to the sustainability of economic development. The excessive consumption of traditional fossil fuels not only restricts the economic development, but also causes a series of environmental problems, such as global warming. Therefore, in America, many states have been trying to improve the production and usage of energy. A successful practice is that some states with different strengths and weaknesses unite to form an interstate compact, for promoting the usage of cleaner, renewable energy sources, through cooperation and adherence to specific policies.

In the southwest of the U.S., there are four states – California (CA), Arizona (AZ), New Mexico (NM), and Texas (TX) – that hope to form a realistic new energy compact as well. Asked by the four governors of these sates, we determine a set of development goals for the energy compact.

1.2 Our Goals

Based on our understanding of the problem, we set the following goals:

- Use the given data to found an energy profile for each state.
- Develop a model system to show the dynamic relationships between various energy consumption and economic development of each of the four states, and explore the underlying factors that lead to these relationships.
- Define the "best" profile for use of cleaner, renewable energy, then set up a system of evaluation to determine the state that has the "best" profile in 2009.
- According to the analysis above, develop a model and set different scenarios to predict the energy profile of each state for 2025 and 2050.
- Based on the established models, decide the usage targets of cleaner and renewable energy for 2025 and 2050. Then provide three actions for the four states to achieve the goals.

1.3 Our Thinking

This is a typical big data problem, so we solve it from the point of view of statistical analysis. Here is our thinking.

First, we preprocess the data provided, which includes default value processing, abnormal value processing and data synthesis and classification. Based on our definitions, we select some major energy sources for analysis and divide them into two categories. One is called **cleaner renewable energy (CRE)**, the other is called **traditional fossil energy (TFE)**. The specific definitions will be given later.

Second, we select some important data from the dataset provided to construct the energy profile. Through statistical charts, we visualize them. In this way, we found an energy profile for each state. Comparing the energy files of the four states, we find differences in their utilization of the two kinds of energy sources. To further clarify the differences, we learn from the decoupling theory and use coefficients of elasticity to show the dynamic relationships between various energy consumption and economic development of each state. Then we use the simultaneous equations model to respectively analyze the four states' economy systems, in which way, we find out the reasons why the differences exist.

Third, we introduce the concept of the **comprehensive utilization performance (*CUP*)** to evaluate which of the four states appeared to have the "best" profile. Since different states have their own characteristics, therefore, we establish a multi-dimensional evaluation system to measure the comprehensive utilization performance in case of bias. Then we use principle component analysis (PCA) to integrate each evaluation index into an overall index, namely the comprehensive utilization performance.

Finally, we build the BP neural network to predict the energy profile of each state. By setting various change trajectory of independent variables, we get the predictions successfully. After that, we regard the four states as a whole to determine renewable energy usage targets for 2025 and 2050.

2 Assumptions and Notations

2.1 Assumptions

Due to lack of necessary data and limitation of our knowledge, we make the following assumptions to help us perform modeling. These assumptions are the premise for our subsequent analysis.

- For the four states, all kinds of energy they produce are consumed by themselves each year. Thus we can replace energy output with energy consumption in calculation.
- The policy of each state will not change in the future. This assumption may be not realistic, but it is essential for us when predicting the energy profile in 2025 and 2050.
- The natural environment of the each states will not change. So we can see it as constant. This assumption simplify our analysis, and it is reasonable, since the natural environment is usually stable.
- Once forming an interstate energy compact, the four states can develop and utilize resources together. Thus they can realize manual benefit and win-win result.

2.2   Notations

Here are the notations and their meanings in our paper:

Notation	Meaning
Epd	Electricity production
Tpd	Total energy production
Rce	Renewable energy expenditure
Gdp	GDP
Cpd	Renewable energy production
Pri	Price energy
Ind	Secondary industry consumption
Pop	Population
Wcs	Wood consumption
Tcs	TFE consumption
Enc	Total energy consumption
Cep	Carbon emission per capita
K	Energy expenditures as share of GDP

Table 1: notation

3   Data Preprocessing

For data-analysis problem, there are usually some incomplete and abnormal data in the large amount of raw data, which may seriously affect the efficiency of modeling and the accuracy of conclusions. So it is quite important to preprocess the data.

3.1   Default Value Processing

We use different methods to process variables with various degrees of data loss. (1) For variables with large amount of data missing, we just delete it. Because small data cannot provide enough and valuable information for our modeling. (2) For variables with a small amount of data missing, we use interpolation method to compensate the data. More specifically, we first use existing data points to establish an appropriate interpolation function, and then replace the missing value with the function value $f(x_i)$ at the corresponding point x_i.

3.2   Abnormal Value Processing

If a value in a set of data is more than twice the standard deviation of the average, we call it the abnormal value. Statistically we can use a box-plot to identify the abnormal values. For the abnormal value, we fix it with the average value of its two adjacent observations.

3.3 Data Synthesis and Classification

There are 605 variables in the dataset provided. Not all of them are used in our model. Thus, we sort the data we need into a new dataset. Besides, we divide the energy sources into two categories for the further analysis, One is called cleaner renewable energy, which includes the hydroenergy, the wind energy, the geothermal energy, the solar energy and the ethanol; the other is called traditional fossil energy, which includes the coal, the petroleum, the natural gas and other fuels.

Definition:

- **Cleaner renewable energy** is the energy that produces little pollution and can be used directly in the production and life.
- **Traditional fossil energy** is non−renewable energy and will cause air pollution after burning.

In our model, some synthesized variables are used, for example, the average price of cleaner renewable energy. It can be calculated according to formula (1). Other synthesized variables will be explained later.

$$P = \frac{\sum\limits_{i=1} p_i \cdot q_i}{\sum\limits_{i=1} q_i} \qquad i = 1, 2, \cdots \tag{1}$$

4 Energy Profile

After data preprocessing, we choose 11 important variables to create the energy profile. we call them the basic variables

Notation	Abbreviation in the original data set
GDP	GDPRX
Total population	TPOPP
TFE consumption	FFTCB
Total consumption	TETCB
Wood consumption	WWTCB
Electricity production	ESTCB
Total energy production	TEPRB
Price of renewable energy	AVACD
Renewable energy production	REPRB
Renewable energy expenditure	RFEIV
Secondary industry consumption	TEICB

Table 2: basic variable

Descriptive statistics is usually the first step in statistical analysis, which can shows the important features of the data visually through graphs and tables. In view of this, we adopt the descriptive statistics method to show the energy profile of each state. Here is the energy profile of California, those of other states are attached to the appendix.
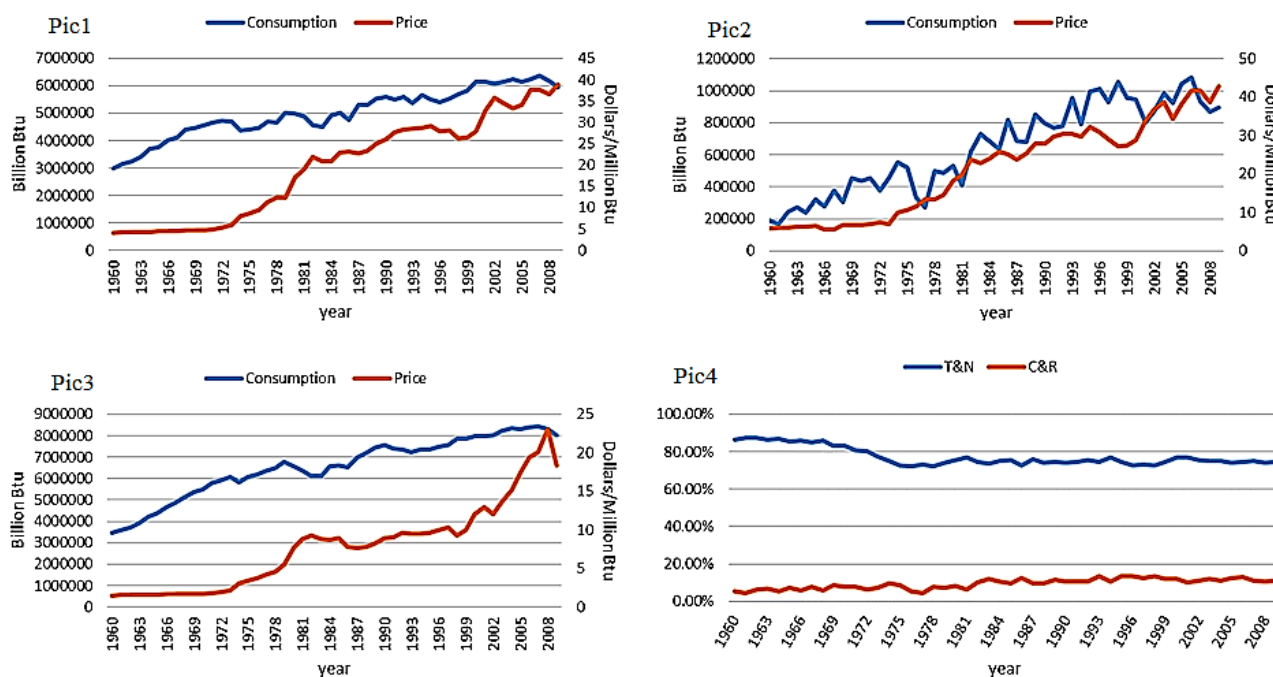
Figure 1: change trend of price and energy consumption in California
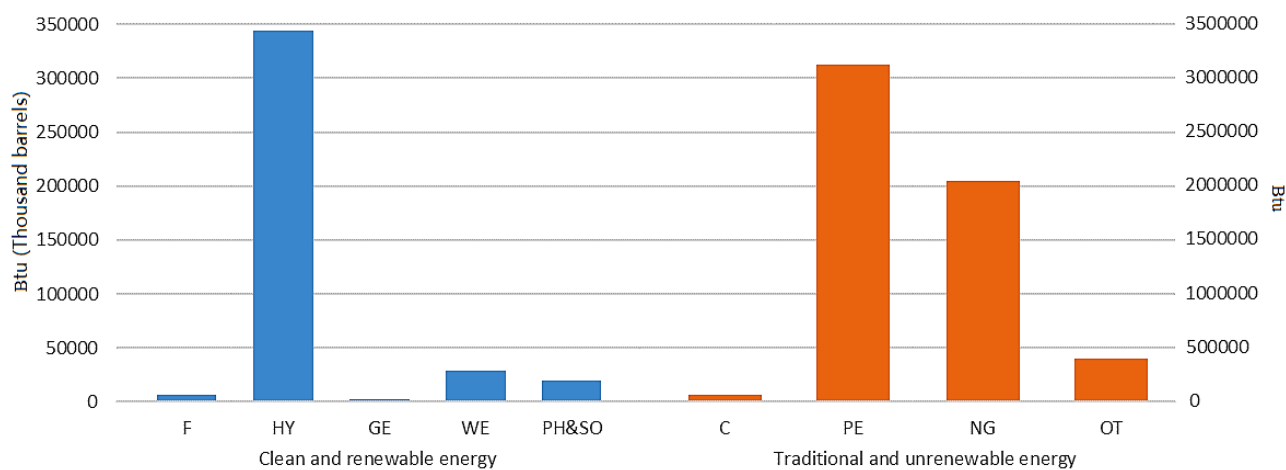

Figure 2: energy consumption of different items in California

Figure 2 :			
Pic1	Trend graph of TFE and its price		
Pic2	Trend graph of CRE and its price		
Pic3	Trend graph of total consumption(TC) and its price		
Pic4	Trend graph of the ratio of TFE to TC and CRE to TC		
Figure 3 :			
F	Fuel enthanol	C	Coal
HY	Hydroelectricity	PE	Petroleum products
GE	Geothermal energy	NG	Natural gas
WE	Wind energy	OT	Other
PH&SO	Photovoltaic and solar thermal energy		
(Only the unit of "F"is Thausand Barrels)			

Table 3: the cutline in Figure 2 and Figure 3
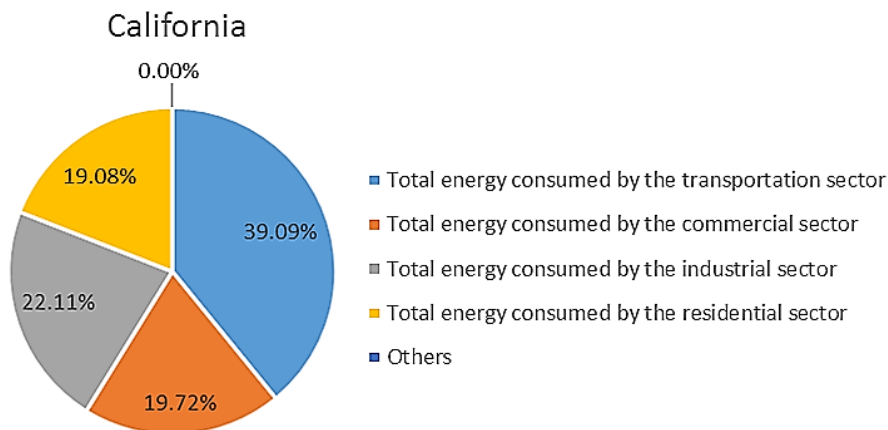
Figure 3: constitution of total energy consumption in California in 2009

Variable	Quantitative Change	Rate of Change
Energy total production	510310.93(Billion Btu)	156.56%
Renewable energy total production	28262.33(Billion Btu)	383.32%
Fossil fuels, total production	482048.60(Billion Btu)	151.32%
Electricity production	16995.11(Billion Btu)	2279.84%
Wood consumption	4244.64(Billion Btu)	64.04%
Secondary industry consumption	414753.42(Billion Btu)	199.64%
Renewable energy total end-use expenditures	750.17(Million dollars)	873.31%
GDP	59090(Million dollars)	593.15%

Table 4: quantitative change and rate of change of typical variables from 1960 to 2009

By comparing the energy profiles of the four states, we preliminarily reached the following conclusions:

- Overall, the four states differ in their utilization and consumption of the two types of energy (CRE&TFE).
- Each of the four states has a growing share of CRE consumption, while the share of TFE is still high.
- California ranks the first in terms of the amount of CRE consumption; Arizona ranks the first in terms of the share of CRE consumption. However, New Mexico lags behind in the usage and consumption of CRE.

5   Model Construction

5.1   Decoupling Analysis

Since having had a general idea of the energy profiles of the four states, we now intend to calculate the coefficients of elasticity to deeply analyze the dependency between economic development and the various energy consumption of each state. According to the decoupling theory, the higher the absolute value of the elasticity, the stronger the dependency.[1] The formula of coefficient of elasticity is as follow.

$$ela = \left| \frac{\Delta Enc}{\Delta Gdp} \cdot \frac{Gdp}{Enc} \right| \tag{2}$$

The calculation results are shown in figure 4. From the figure, we can intuitively know that generally the dependency between economic development and CRE consumption is on the rise for each state, but the intensity of the dependency differs among the four states.

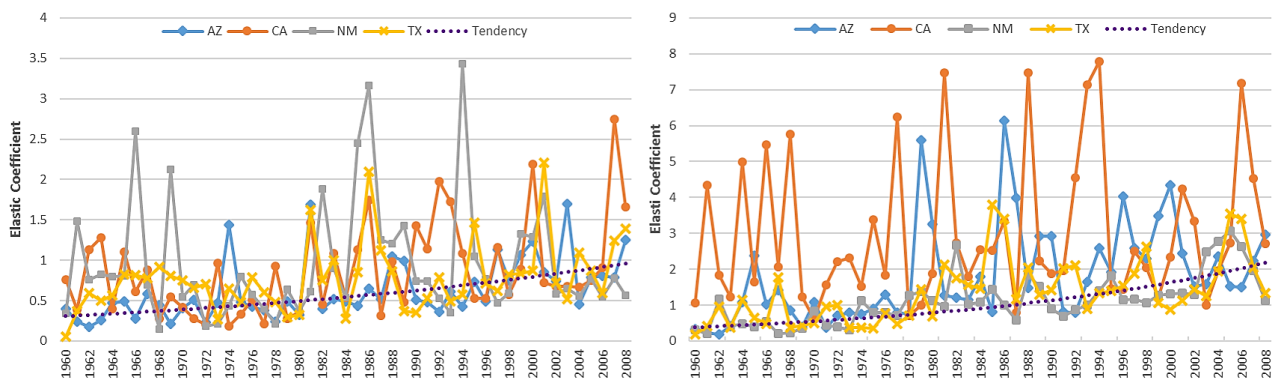Figure 4: elasticity coefficient of FTE(left) and CRE(right)

5.2   The Simultaneous Equations Model

To explore the underlying factors that leads to the differences of energy files among the four states, we develop a simultaneous equations model, as shown below. Different

from single equation regression, simultaneous equations model can explain the complex economic system more comprehensively and accurately. We introduce three equations into the model, since there are interactions between each two of the three elements, economic development, energy consumption, and environmental pollution.[2]

$$\begin{cases} \ln(Enc) = \beta_0 + \beta_1\ln(Gdp) + \beta_2\ln(Cep) + \beta_3\ln(Pr) + \beta_4\ln(pop) + \beta_5\ln(Ind) + \beta_6\ln(K) + X + \varepsilon_3 \\ \ln(Gdp) = \alpha_0 + \alpha_1\ln(Enc) + \alpha_2\ln(Cep) + \alpha_3\ln(K) + \alpha_4\ln(Pop) + \varepsilon_1 \\ \ln(Cep) = \gamma_0 + \gamma_1\ln(Gdp) + \gamma_2\ln(Enc) + \gamma_3\ln(Pop) + \gamma_4\ln(Ind) + \varepsilon_2 \end{cases} \quad (3)$$

where X is a set of controlling variables on the natural environment

Lack of variable data on the natural environment, we neglect the set of controlling variables X to solve the model. The regression results of equation $\ln(Enc)$ are shown in the following table. The regression coefficients reflect the influence of explanatory variables on dependent variables, from which we can determine the impact of different factors on the energy variable.

Equation 1	TFE				CRE			
	AZ	CA	NM	TX	AZ	CA	NM	TX
$\ln(Gdp)$	0.2189***	0.5122**	0.4896**	1.2368***	0.1904***	0.4915**	0.2167*	0.4532**
$\ln(Cep)$	0.0012	0.0104**	0.1326**	0.1725**	0.0031	0.0113*	0.0682**	0.1651***
$\ln(K)$	0.2169**	0.3481**	0.2018*	0.3018**	0.2267**	0.4162**	0.2481***	0.3156*
$\ln(Pop)$	0.0104	0.0341	0.0361*	-0.2214	0.0133	0.0265	0.0421*	-0.2153
$\ln(Pri)$	0.1152*	0.1421	0.1102**	0.2451	0.1842*	0.1821*	0.2012**	0.2201
$\ln(Ind)$	0.1421*	0.4142***	0.2269*	0.2684**	0.1723*	0.3841***	0.2362*	0.2758*
Constant	17.6421*	19.2364*	13.2631*	50.3641*	16.6372*	37.6298*	13.2156*	61.2571*

Table 5: the regression results of Equation 1

*，**，*** respectively represent the confidence of 10%,5%,1%

From the Table 5, we can see that population has no significant influence on the energy consumption. While the the economic development and the share of secondary industy have significant influence on the energy consumption.

To determine the influence of geography and climate, we collect information about the natural environment of the four states, which is shown in Figure 5. It shows the resource distribution of solar power, wind power and hydroelectric power.

From Figure 5, we know that the four states differ in the potential of CRE production. The differences are caused by the various natural environment of the four states, such as climate and geography.
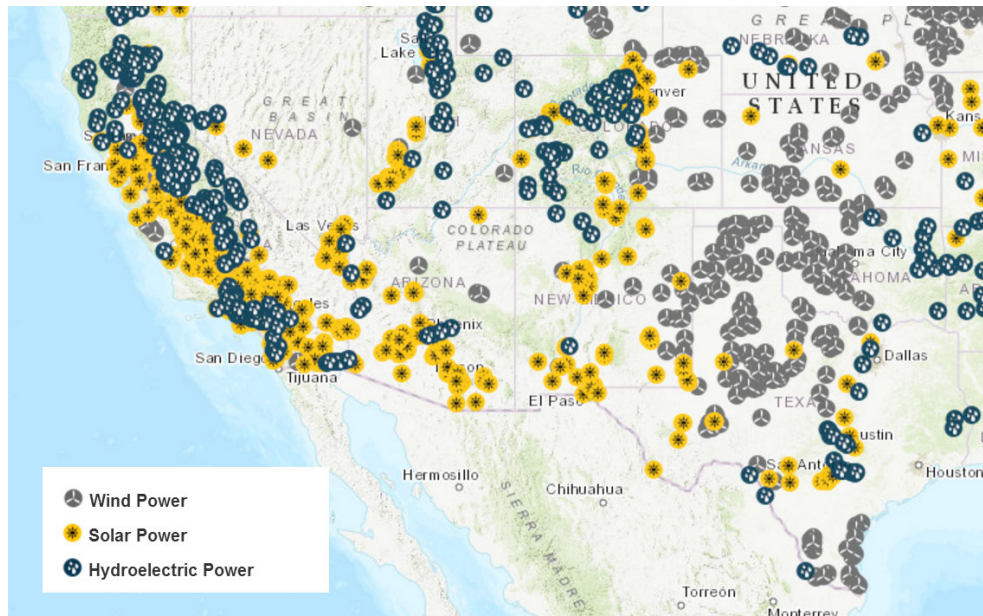
Figure 5: resource distribution of CRE

5.3   The Multi-dimensional Evaluation System

In this part, we will determine which of the four states has the "best" energy profile in 2009. Through the above analysis, we know that the four states differ in strengths and weaknesses congenitally. Thus, for avoiding bias, we establish a multi-dimensional evaluation system to measure their energy profile.

5.3.1   Construction of Indexes

The core indexes of our evaluation system is called comprehensive utilization performance (CUP), which reflects the development and utilization level of CRE. The CUP is composed of three parts, energy performance, economic performance and environment performance. And each of the three parts includes three indexes as well. These indexes and their formulas are detailed in the Table 6.
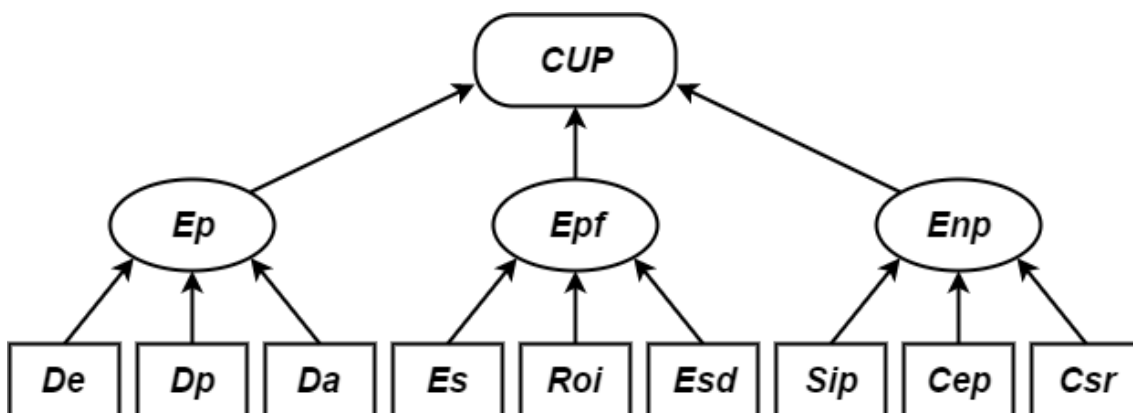

Figure 6: evaluation index systerm

	Name of Index	Meaning	Formula
Energy performance (Ep)	Development Efficiency (De)	Reflects the status of the CRE development and utilization in the electricity industry.	$De = \dfrac{Cpd}{Epd}$
	Development Potential (Dp)	Reflects the sustainability of CRE development and utilization.	$Dp = \dfrac{(Cpd_t - Cpd_{t-1})}{Cpd_{t-1}}$
	Development Achievements (Dpa)	Reflects the contribution of CRE development and utilization to the state energy structure.	$Dpa = \dfrac{Cpd}{Tpd}$
Economic proformance (Epf)	Economic support (Es)	Reflect the state's support for renewable energy at the investment level	$Es = \dfrac{Rce}{Gdp}$
	Rate of return on investment (Roi)	Reflect the efficiency and profitability of investment on renewable energy	$Roi = \dfrac{Cpd}{Rce}$
	Equilibrium of supply and demand (Esd)	Reflect supply's satisfaction with demand	$Esd = \dfrac{1}{Pc}$
Environment performance (Enp)	Secondary industry proportion (Sip)	Reflects the dependence on polluting industries in each state.	$Sip = \dfrac{Scs}{Tocs}$
	Carbon emission per capita (Cep)	Reflectis the situation of the greenhouse gas emission per capita.	$Cep = \dfrac{Tcs}{Tpo}$
	Consumption of natural resources per capita (Csr)	Reflects the situation of natural resources consumption per capita.	$Cnr = \dfrac{Wcs}{Tpo}$

Table 6: indexes and their formulas

5.3.2 Principle Component Analysis

We use the PCA method to integrate each evaluation index into an overall index, namely the comprehensive utilization performance. Then through the comparison of the *CUP* values of the four states, we determine the "best" state.

First, we use the given data and the formulas in Table 6 to calculate the evaluation indexes. And then, we use the equation below to standardize the values of these indexes.

$$\bar{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j}$$

$$\mu_j = \frac{1}{n} \sum_{i=1}^{n} a_{ij} \qquad (4)$$

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (a_{ij} - \mu_j)^2}$$

Next, we use the equation N to calculate the correlation coefficient matrix of the standardized data.

$$r_{ij} = \frac{\sum\limits_{k=1}^{n} \bar{a}_{ki} \cdot \bar{a}_{kj}}{n-1} \tag{5}$$

Then we compute the eigenvalue λ_i and eigenvector $\vec{u}_i$ of the correlation coefficient matrix. According to the eigenvectors, we can get the principal components y_i. We can get nine principal components in total, but we can only use three of them, which have the highest contribution rate.

$$\begin{aligned}
y_1 &= u_{11}\bar{x}_1 + u_{21}\bar{x}_2 + \cdots + u_{m1}\bar{x}_m, \\
y_2 &= u_{12}\bar{x}_1 + u_{22}\bar{x}_2 + \cdots + u_{m2}\bar{x}_m, \\
&\vdots \\
y_n &= u_{1n}\bar{x}_1 + u_{2n}\bar{x}_2 + \cdots + u_{mn}\bar{x}_m,
\end{aligned} \tag{6}$$

Finally, we calculate the contribution rate based on the formula 7, then use the formula 8 to determine an overall index, which is named comprehensive utilization performance by us. Sorting the *CUP*, we can find out which of the states is the best.

$$b_j = \frac{\lambda_j}{\sum\limits_{k=1}^{m} \lambda_k} \tag{7}$$

$$Z = \sum_{j=1}^{p} b_j y_j \tag{8}$$

The results of PCA is showed in table 7. from the table, we can see that **California has the "best" energy profile in 2009.**

Rank	State	PCA Score
1	California	0.9433
2	Texas	0.5561
3	Arizona	0.2466
4	New Mexico	-1.7459

Table 7: PCA scores

5.4 Energy Profile Predicting System

5.4.1 Determining the Predictors

In part 5.3, we measured the energy file in three dimensions, energy performance, economic performance and environment performance. For the following analysis, we

select an index from each of them to calculate the comprehensive utilization performance (*CUP*) in a different way.

The indexes we select are Dpa, Roi and Cep. Analogous to Cobb-Douglas Production Function in economics, we derive the following formula.

$$CUP = A \cdot (Dpa)^{\alpha} \cdot (Roi)^{\beta} \cdot Cep^{\gamma} \tag{9}$$
$$where:$$
$$A\ is\ unit\ correction\ factor;$$
$$\alpha,\beta,\gamma\ are\ weight\ coefficients\ of\ variables;$$
$$\alpha + \beta + \gamma = 1.$$

In the previous analysis, we pointed out that, there is an interaction between each tow of the three elements, economy, energy and environment.[3] Therefore, we choose *GDP per capital (Gdp)* , *Secondary industry proportion (Sip)* and *carbon emission per capita (Cep)* as the independent variables in prediction. Then we can drive the following equation.

$$CUP = \theta_0 + \theta_1 Gdp + \theta_2 Sip + \theta_3 Cep + \varepsilon \tag{10}$$

5.4.2 Constructing BP Neural Network

Since we have to predict the energy profiles of the four states respectively, it is a heavy task. Thus, we utilize the intelligent algorithm (BP neural network) with high prediction efficiency to finish this work. It has been proved theoretically that BP neural network can approach any nonlinear function with higher precision. The flow chart below shows the principle of BP neural network.
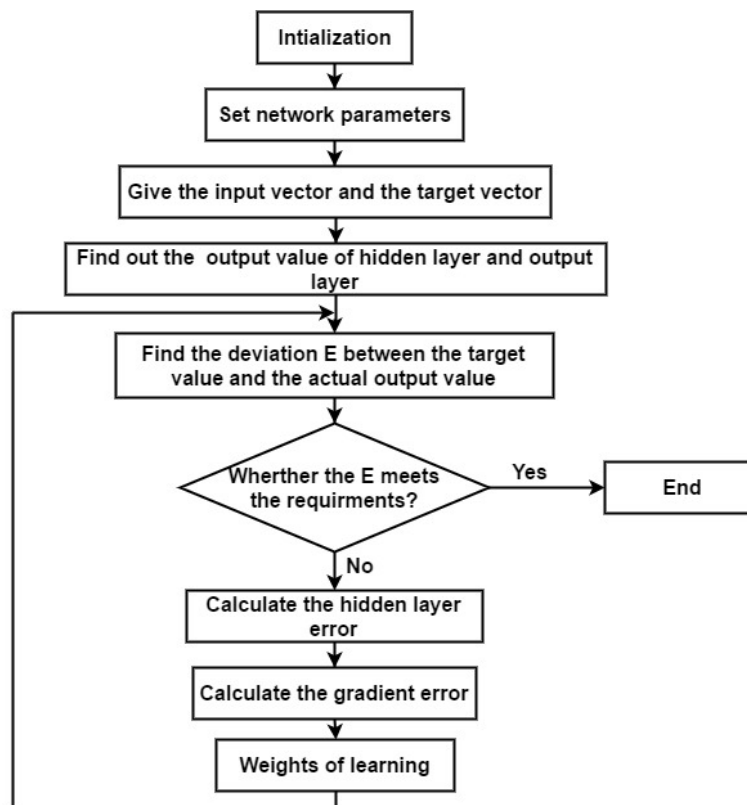

Figure 7: the principle of BP neural network

There are two steps in the process of constructing BP neural network. First, we should set the network parameter k that represents the numbers of neurons. Here we let $k = 9$. Second, we use existing data of input and output to train the network. We now drive the core formulas of the BP Network.

Set a pair of samples (X, Y), $X = [x_1, x_2, \cdots, x_m]'$, $Y = [y_1, y_2, \cdots, y_n]'$. The hidden neural unit is $O = [O_1, O_2, \cdots, O_l]$. The network weight matrix between the input layer and the hidden layer neurons W^1, and the network weight matrix between the hidden layer neurons and the output layer W^2 are shown below.

$$W^1 = \begin{bmatrix} \omega_{11}^1 & \omega_{12}^1 & \cdots & \omega_{1m}^1 \\ \omega_{21}^1 & \omega_{22}^1 & \cdots & \omega_{2m}^1 \\ \vdots & \vdots & & \vdots \\ \omega_{l1}^1 & \omega_{l2}^1 & \cdots & \omega_{lm}^1 \end{bmatrix}, \quad W^2 = \begin{bmatrix} \omega_{11}^2 & \omega_{12}^2 & \cdots & \omega_{1l}^2 \\ \omega_{21}^2 & \omega_{22}^2 & \cdots & \omega_{2l}^2 \\ \vdots & \vdots & & \vdots \\ \omega_{n1}^2 & \omega_{n2}^2 & \cdots & \omega_{nl}^2 \end{bmatrix} \tag{11}$$

The threshold of the hidden layer neurons θ^1 and the output layer neurons θ^2 are :

$$\theta^1 = [\theta_1^1, \theta_2^1, \cdots, \theta_l^1]', \quad \theta^2 = [\theta_1^2, \theta_2^2, \cdots, \theta_n^2]' \tag{12}$$

So the output of the hidden layer is:

$$O_j = f(\sum_{i=1}^{m} \omega_{ji}^1 x_i - \theta_j^1) = f(net_j) \tag{13}$$

where $f(\cdot)$ is the transfer function of the hidden layer

And the output of the hidden layer is:

$$z_k = g(\sum_{j=1}^{l} \omega_{kj}^2 O_j - \theta_k^2) = g(net_k) \tag{14}$$

where $g(\cdot)$ is the transfer function of the output layer

The error between the network's output and the expected output is:

$$\begin{aligned} E &= \frac{1}{2} \sum_{k=1}^{n} (y_k - z_k)^2 \\ &= \frac{1}{2} \sum_{k=1}^{n} [y_k - g(\sum_{j=1}^{l} \omega_{kj}^2 O_j - \theta_k^2)]^2 \\ &= \frac{1}{2} \sum_{k=1}^{n} \{y_k - g[\sum_{j=1}^{l} \omega_{ij}^2 f(\sum_{i=1}^{m} \omega_{ij}^1 x_i - \theta_j^1) - \theta_k^2]\}^2 \end{aligned} \tag{15}$$

ω_{kj}^2 is the weight of the neurons between hidden layer and output layer. The partial derivative of the error E to ω_{kj}^2 is:

$$\frac{\partial E}{\partial \omega_{kj}^2} = \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial \omega_{kj}^2} = -(y_k - z_k) g'(net_k) O_j = -\delta_k^2 O_j \tag{16}$$

ω_{ji}^1 is the weight of the neurons between hidden layer and output layer.The partial derivative of the error E to ω_{ji}^1 is:

$$\frac{\partial E}{\partial \omega_{ji}^1} = \sum_{k=1}^{n}\sum_{j=1}^{l} \frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial O_j}\frac{\partial O_j}{\partial \omega_{ji}^1} = -\sum_{k=1}^{n}(y_k - z_k)g'(net_k)\omega_{kj}^2 f'(net_j)x_i = -\delta_j^l x_i \qquad (17)$$

According to formula (16) and (17), we can drive the adjusting formula of weight：

$$\begin{cases} \omega_{ji}^1(t+1) = \omega_{ji}^1(t) + \Delta\omega_{ji}^1 = \omega_{ji}^1(t) - \eta^1 \dfrac{\partial E}{\partial \omega_{ji}^1} = \omega_{ji}^1(t) + \eta^1 \partial_j^1 x_i \\[3mm] \omega_{kj}^2(t+1) = \omega_{kj}^2(t) + \Delta\omega_{kj}^2 = \omega_{kj}^2(t) - \eta^2 \dfrac{\partial E}{\partial \omega_{kj}^2} = \omega_{kj}^2(t) + \eta^2 \partial_j^2 O_j \end{cases} \qquad (18)$$

where η^1，η^2 are the learning steps of hidden layer and output layer respectively.

5.4.3   Analyzing the Error

Entering the selected variable data into the network, we get the fitting figures, shown in Figure 8. Then we use the following formula to calculate the average fitting error, the results are showed in table 8.

$$Afe = \frac{1}{n}\sum_{i=1}^{n}|e_i| = \frac{1}{n}\sum_{1}^{n}|y_i - \hat{y}_i| \qquad (19)$$

State	California	Arizona	New Mexico	Texas
Afe	6.47%	5.37%	8.82%	12.09%

Table 8: average fitting error

5.4.4   Predicting the Target Values

To predict the dependent variable, we have to know the independent variables first. Unfortunately, we lack the values of Gdp, Sip and Cep for 2025 and 2050. So we can only predict the dependent variable by setting the change trajectory of the three independent variables in advance. We set two scenarios in total.[4]
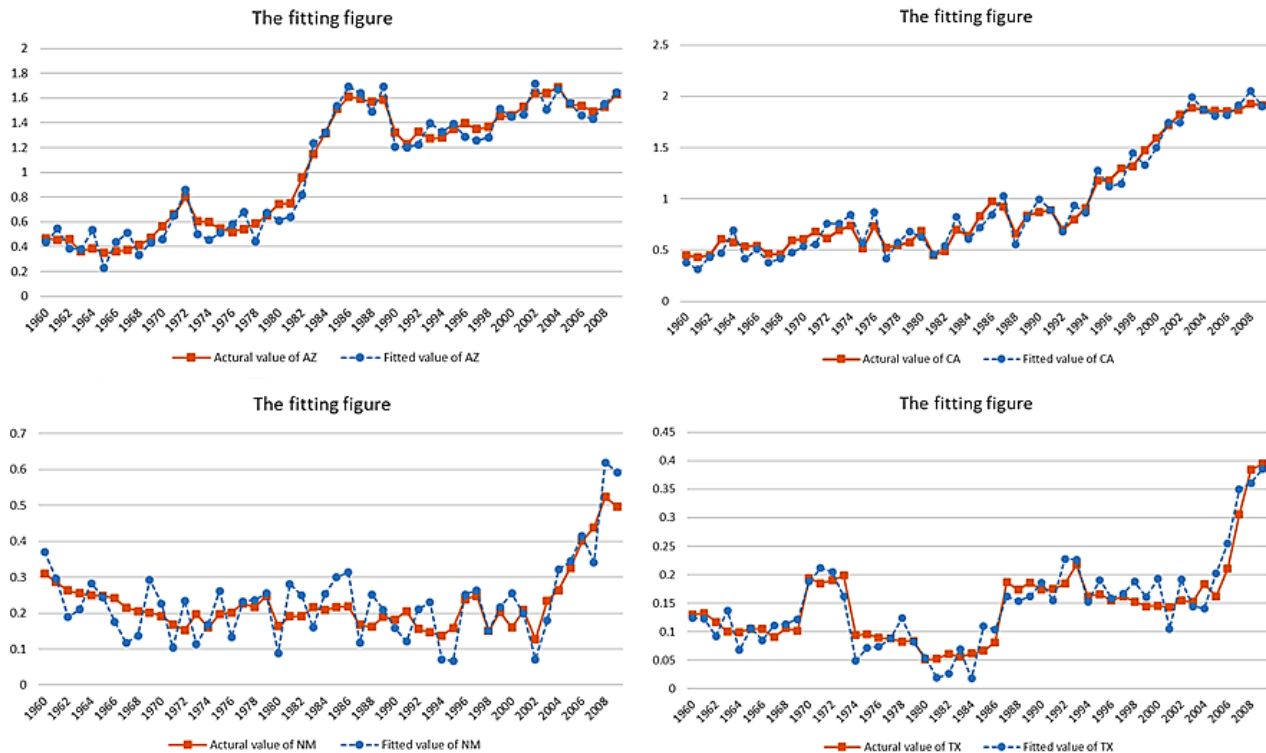
Figure 8: the fitting results
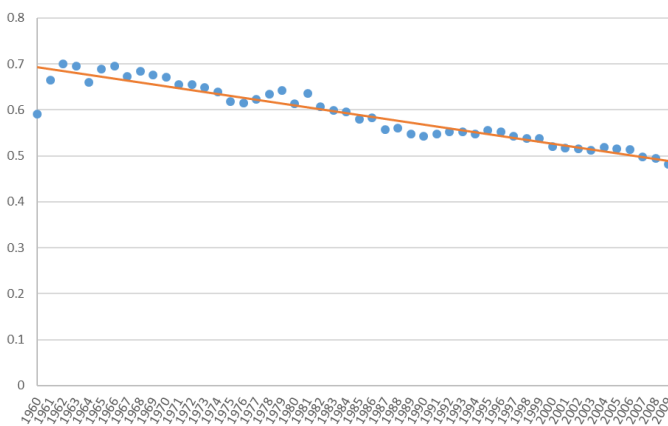

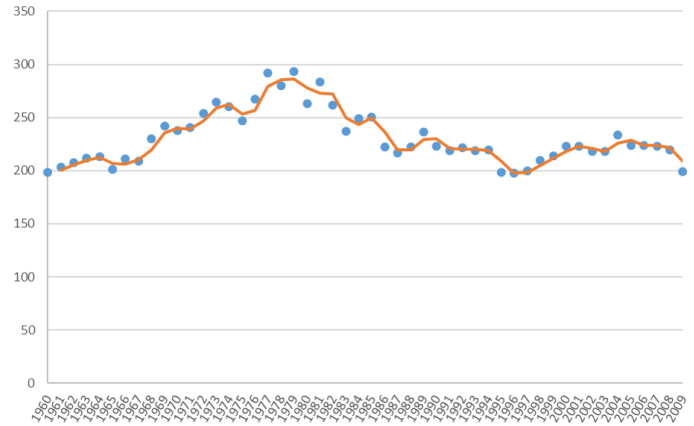Figure 9: Scenario one   Figure 10: Scenario two

- **Scenario one: Linear trend**
  As is shown in figure 9, if the change of the three independent variables follows the linear trend. We can use the formula 15 to characterize their trajectory.

$$x_t = x_{t-1} + a \tag{20}$$

- **Scenario two: Smooth fluctuation**
  As is shown in figure 10, if the change of the three independent variables follow the smooth fluctuation. We can use autoregressive moving average model, ARMA(p,q) to characterize their trajectory.[5]

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} - \theta_q \varepsilon_{t-q} + \mu \qquad (21)$$

Once determining the change trajectory of independent variables, we can estimate their values. Entering them to the BP networks, we get the predicted values of the *CUP* of each states. The results are shown in table 9.

		Scenario 1				Scenario 2			
	Year	DPA	ROI	CEC	CUP	DPA	ROI	CEC	CUP
AZ	2025	0.06	25.16	170.97	23.92	0.05	22.15	190.26	23.49
	2050	0.07	50.45	100.45	40.19	0.06	50.15	120.15	35.22
CA	2025	0.08	30.04	138.63	28.39	0.10	29.36	145.26	26.14
	2050	0.09	57.87	89.14	45.15	0.12	55.86	108.15	39.15
NM	2025	0.07	50.14	350.45	8.20	0.07	41.26	351.22	7.89
	2050	0.08	59.85	282.43	15.16	0.09	57.12	301.86	14.12
TX	2025	0.05	20.60	348.15	7.84	0.05	19.26	369.15	6.95
	2050	0.08	48.29	321.53	14.50	0.08	44.12	348.85	12.19

Table 9: the predicting results

6 Testing Our Model

6.1 Predicting Reliability Test

To test the predicting reliability of the BP network, we collect relevant data of each state from 2010 to 2015, and use them to predict the *CUP*. Then we use formula 15 again to compute the average error between the predicted values and the actual ones. Here are the results.

The predicting error				
State	California	Arizona	New Mexico	Texas
Error	5.89%	6.24%	7.98%	10.07%

Table 10: the predicting error

6.2 Sensitivity Analysis

In the process of constructing BP network, we let the network parameter $k = 9$. How does the change of k influence the predicting results? We analyze the average deviation of *CUP* caused by changing k slightly.

State	k-2	k-1	k+1	k+2
AZ	11.08%	4.43%	3.54%	8.51%
CA	8.19%	3.15%	4.66%	10.80%
NM	9.78%	5.39%	4.66%	12.28%
TX	6.66%	2.19%	2.28%	12.33%
Mean	8.93%	3.79%	3.79%	10.98%

Table 11: the influence by changing k

From table 11, we can find that the influence of k is not big, which we can bear.

7   Determining Goals for the Energy Compact

7.1   Idea one: Don't cooperate

If the four states don't cooperate, which means they develop independently, they can only utilize their own resources to enhance the level of renewable energy usage. In this condition, we can just use the predicted values in part 5.4 as their respective targets.

For the two scenarios in part 5.4, we think that scenario two is more reasonable and realistic. Because the evolution of the macro-economy is usually stable, especially for advanced economies like the United States.[6] So we just take scenario two into consideration.

In scenario two, the average comprehensive utilization performance can increase to 16.12 and 25.17 successively in 2025 and 2050.

7.2   Idea two: Cooperate

Given that the four-state energy compact is an interstate compact, it is necessary for the tour states to cooperate. We presume that they can develop and utilize resources together. In this case, we regard the four states as a whole economy. Then we use previous method and models to predict the whole economy's comprehensive utilization performance for 2025 and 2050.

Year	2025	2050
The whole economy's CUP	16.98	26.05

Table 12: predictions of the whole economy

The results shows that the whole economy's comprehensive utilization performance can increase to 16.98 and 26.05 successively in 2025 and 2050, and each of them is higher than that in idea one. Maybe it is because cooperation enables the four states to give full play to their own advantages and promote the development and utilization of cleaner renewable energy.[7]

7.3   Goals and Actions

Based on the results in idea two and the previous analysis, we set the following goals for the four-state energy compact.

- Building a community of energy utilization and economic development, so as to realize manual benefit and win-win result.

- Promoting the production and usage of new energy sources and reducing then dependence on traditional fossil fuels.

- Increasing the whole economy's comprehensive utilization performance to 16.98 by 2025 and to 26.05 by 2050, and achieving the harmony of economy, energy and environment ultimately.

According to the goals above and the characteristics of each state, we propose the actions and measures for the four states.

- Focus on developing their own advantageous energy resource and share their energy and achievements with others.

- Make more investment in scientific research to make better use of renewable energy to improve the low utilization efficiency of new energy sources.

- Introduce a proper subsidy policy to award the enterprises who develop new energy source.

- Make certain quantitative index and do statistic periodically to make sure the direction of energy development is going well.

8   Conclusions

We are asked by the governors of the four states to set some goals for their interstate energy compact. After performing data analysis and modeling, we have finished the task successfully. First, Using decoupling theory and simultaneous equations model, we characterize the evolution of energy profile of each state from 1960 – 2009, and find out the influential factors. We have known that it is the differences of economic level, industrial structure and natural environment that lead to the distinct energy profiles of each state. In terms the production and usage of cleaner renewable energy, each of the four states has a growing share of CRE consumption, while the share of TFE consumption is still high.

Second, we construct a multi-dimensional Evaluation System and introduce the core concept in this paper, named comprehensive utilization performance (*CUP*), which reflects the development and utilization level of CRE. Through the PCA method, we determine that California has the "best" energy profile.

Finally, based on the previous analysis, we build BP neural network for predicting the energy profile of each state. By setting various change trajectory of independent variables, we get the target values successfully. After that, we regard the four states as a whole to determine renewable energy usage targets for 2025 and 2050. In this process, we use BP neutral network again.

9   Strengths and Weaknesses

9.1   Strengths

- **Data preprocessing.**When faced with big data problem, the data processing is very important. Through this step, we greatly improve the quality of the data. Thus, it is more efficient and convenient for us to solve the problem.

- **Accuracy and stability.**We use BP neutral network to make predictions. It is a powerful algorithm with great nonlinear approximation ability. The error test shows that our predicting results are more accurate. Besides, when changing the network parameter k, its influence is not large. So the BP network is more stable.

- **Good expansibility and flexibility.** There are three parameters α, β and γ in our *CUP* equation, which can be used to reflect the importance of corresponding indicators. In different situations, we can adjust them flexibly.
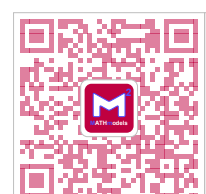
关注数学模型

获取更多资讯

9.2   Weaknesses

- **Subjectivity.** The calculation of some synthesized variables is subjective. It can cause extra error of our models.

- **Lack necessary data.** Lack of data on energy production, we can only use data on energy consumption to replace them.

- **Simplifying assumption.** For convenience of modeling, we neglect the elements about import and export, which play import roles in economy.

References

[1] Carley S.State renewable energy electricity policies: An empirical evaluation of Effectiveness [J]. Energy Policy, 2009, 37(8): 3071–3081.

[2] Verbruggen A, Fischedick M, Moomaw W, et al. Renewable energy costs, potentials, Barriers: Conceptual issues [J]. Energy Policy 2010, 38(2): 850–861.

[3] Vries B J M D, Vuuren D P V, Hoogwijk M M. Renewable energy sources: Their global potential for the first-half of the 21st century at a global level: An integrated Approach [J]. Energy Policy, 2007, 35(4):2590-2610. [36] Sliz-Szkliniarz B. Assessment of the renewable energy-mix and land use trade-off at a regional level: A case stud for the Kujawsko–Pomorskie Voivodship [J]. Land Use Policy, 2013, 35: 257–270.

[4] Al-Badi A H, Malik A, Gastli A. Assessment of renewable energy resources potential in Oman and identification of barrier to their significant utilization[J]. Renewable and Sustainable Energy Reviews, 2009, 13(9): 2734–2739.

[5] Akella A K, Saini R P, Sharma M P. Social, economic and environmental impacts of renewable energy systems [J]. Renewable Energy, 2009, 34 (2): 390–396

[6] Kaygusuz K. Environmental impacts of the solar energy systems [J]. Energy Sources Part A:Recovery Utilization and Environmental Effects, 2009, 31: 1366-1376.

[7] Wee H M, Yang W H, Chou C W, et al. Renewable energy supply chains, performance, application barriers, and strategies for further development [J]. Renewable and Sustainable Energy Reviews, 2012, 16 (8): 5451–5465.
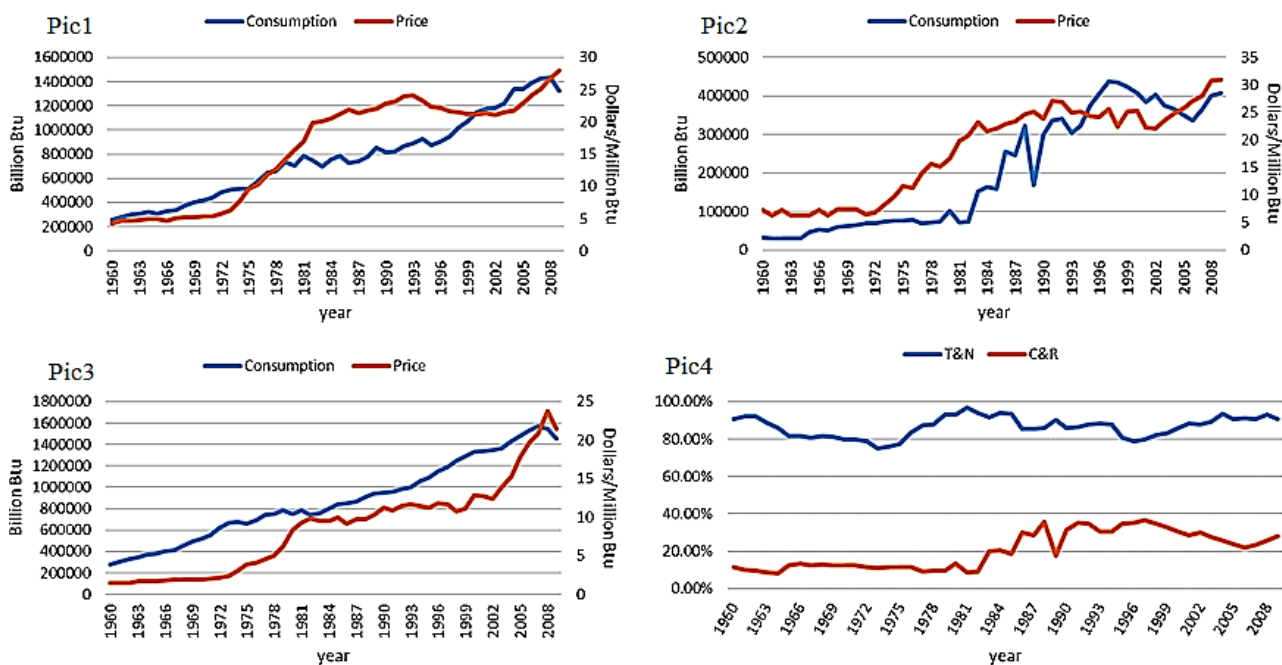
Appendices

Appendix A   Profiles of Arizona

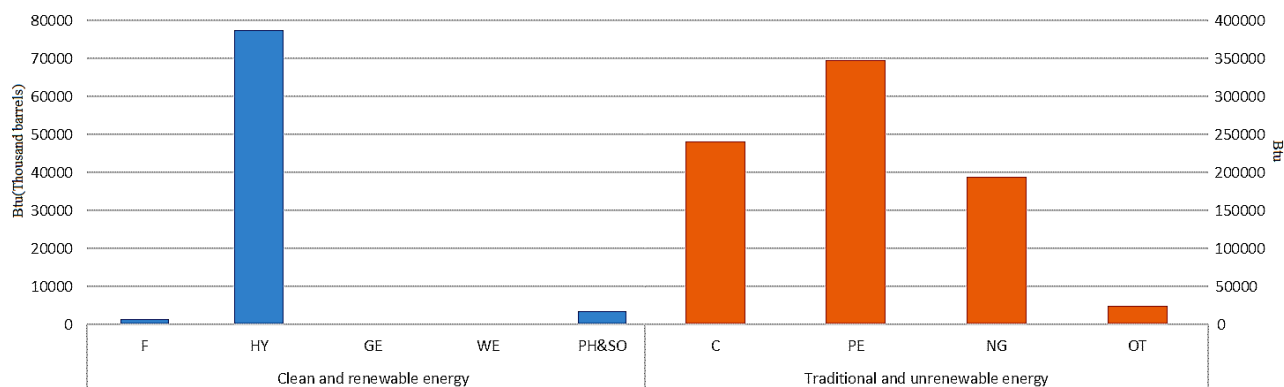Figure 11: Energy profile of Arizona

Figure 12: Means of energy consumption in Arizona

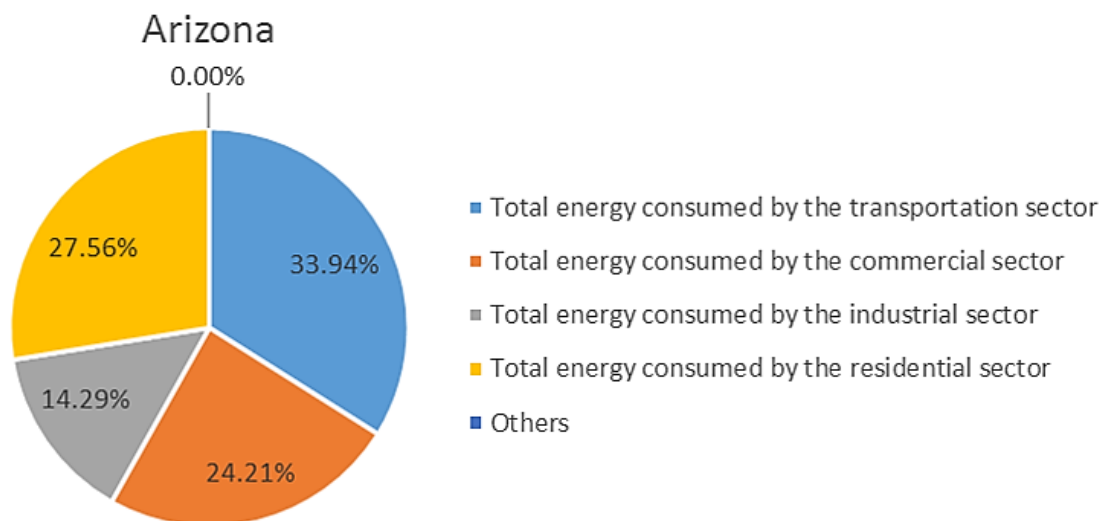| Figure 12 :			
Pic1	Trend graph of TFE and its price		
Pic2	Trend graph of CRE and its price		
Pic3	Trend graph of total consumption(TC) and its price		
Pic4	Trend graph of the ratio of TFE to TC and CRE to TC		
Figure 13 :			
F	Fuel enthanol	C	Coal
HY	Hydroelectricity	PE	Petroleum products
GE	Geothermal energy	NG	Natural gas
WE	Wind energy	OT	Other
PH&SO	Photovoltaic and solar thermal energy		
(Only the unit of "F"is Thausand Barrels)			

Table 13: the cutline in Figure 2 and Figure 3

Figure 13: The constitute of total energy consumption in Arizona in 2009

Data from 1960 to 2009		
Variable	Quantitative Change	Rate of Change
Energy total production	1116104.735(Billion Btu)	374.35%
Renewable energy total production	67311.832(Billion Btu)	186.04%
Fossil fuelS, total production	1048792.903(Billion Btu)	400.36%
Electricity production	672434.71972(Billion Btu)	2090.35%
Wood consumption	6547.67591(Billion Btu)	171.44%
Secondary industry consumption	1127324.01293(Billion Btu)	816.29%
Renewable energy total end-use expenditures	4885.009(Million dollars)	4728.95%
GDP	201016(Million dollars)	6.736912662

Table 14: The quantitative change and the rate of change of typical variables
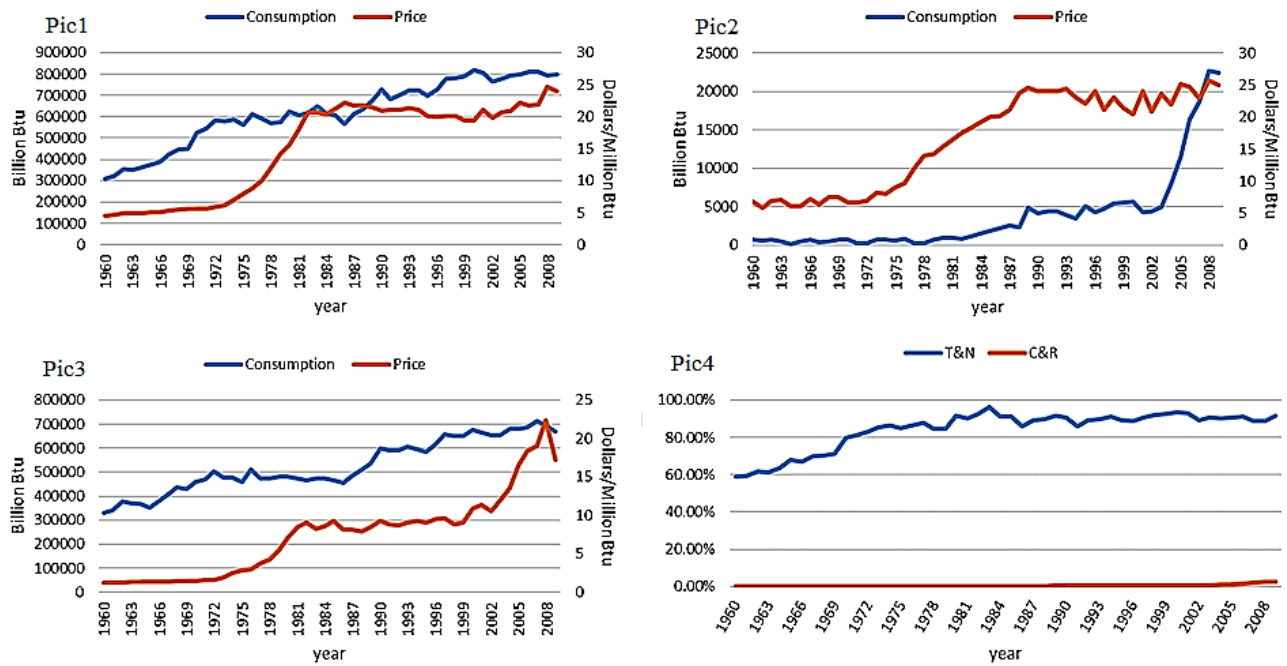
Appendix B   Profile of New Mexico
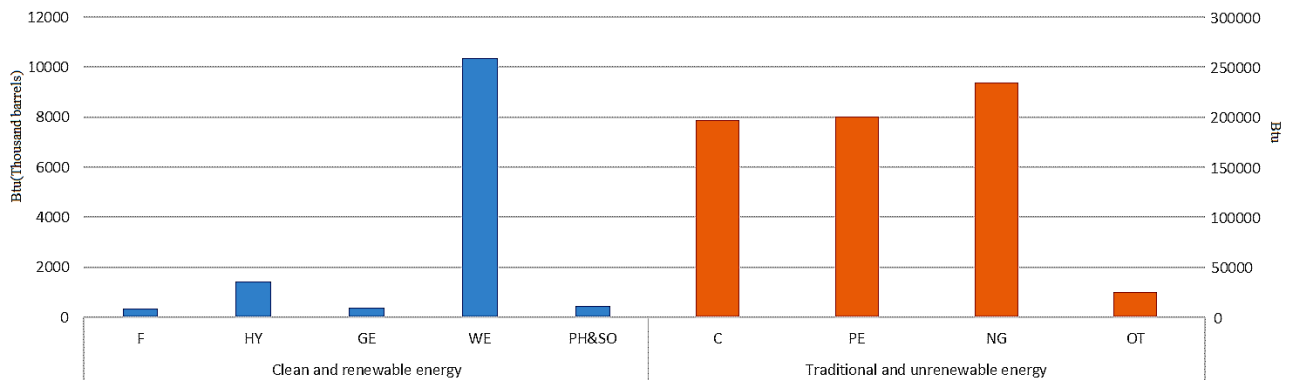

Figure 14: energy profile of New Mexico

Figure 15: means of energy consumption in New Mexico

Appendix C   Profiles of Texas

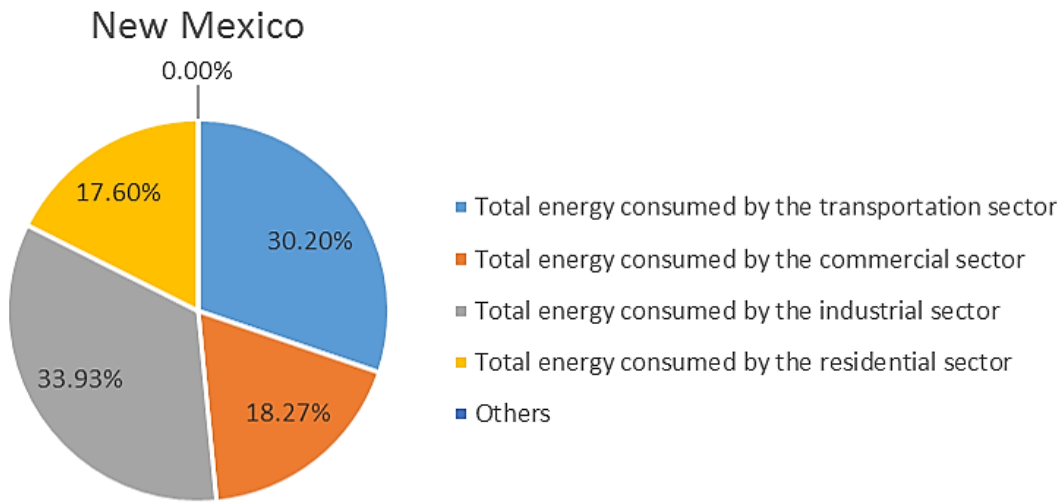| Figure 15 :			
Pic1	Trend graph of TFE and its price		
Pic2	Trend graph of CRE and its price		
Pic3	Trend graph of total consumption(TC) and its price		
Pic4	Trend graph of the ratio of TFE to TC and CRE to TC		
Figure 16 :			
F	Fuel enthanol	C	Coal
HY	Hydroelectricity	PE	Petroleum products
GE	Geothermal energy	NG	Natural gas
WE	Wind energy	OT	Other
PH&SO	Photovoltaic and so-lar thermal energy		
(Only the unit of "F"is Thausand Barrels)			

Table 15: the cutline in Figure 2 and Figure 3

Figure 16: the constitute of total energy consumption in New Mexico in 2009

Table 16: The quantitative change and the rate of change of typical variables

| Data from 1960 to 2009		
Variables	Quantitative Change	Rate of Change
Energy total production	510310.930(Billion Btu)	156.56%
Renewable energy total production	28262.327(Billion Btu)	383.32%
Fossil fuelS, total production	482048.602(Billion Btu)	151.32%
Electricity production	16995.11376(Billion Btu)	2279.84%
Wood consumption	4244.64258(Billion Btu)	64.04%
Secondary industry consumption	414753.41587(Billion Btu)	199.64%
Renewable energy total end-use expenditures	750.173(Million dollars)	873.31%
GDP	59090(Million dollars)	593.15%
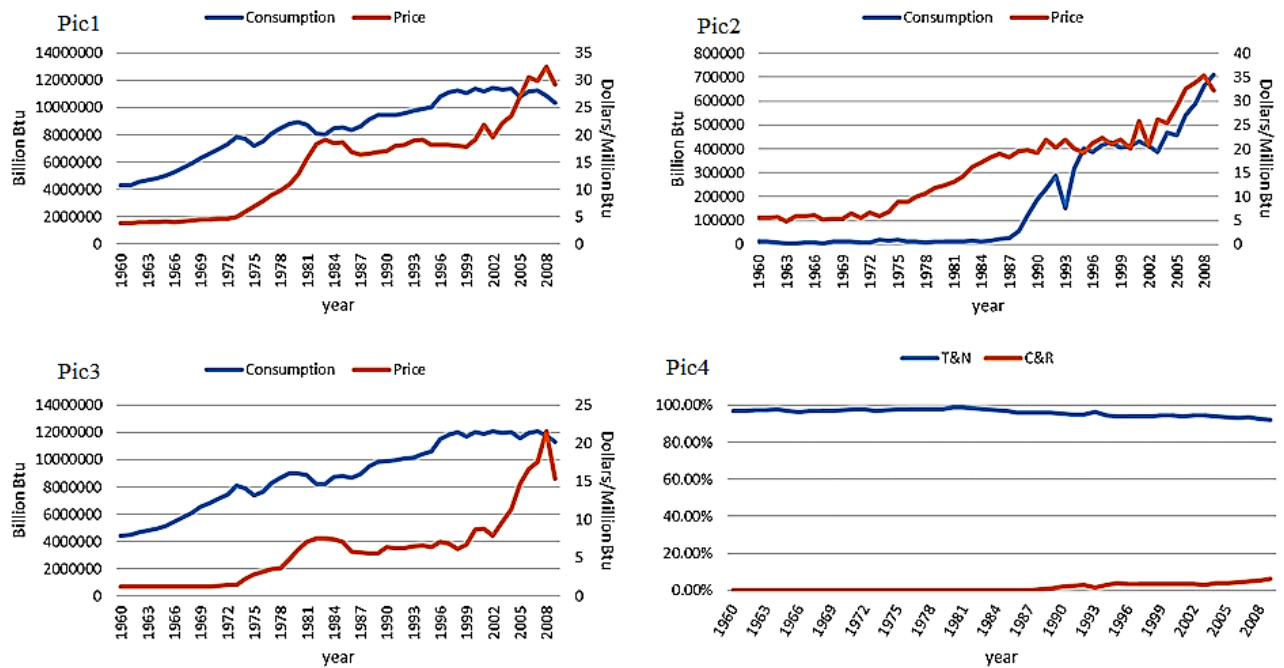
Figure 17: Energy profile of Texas
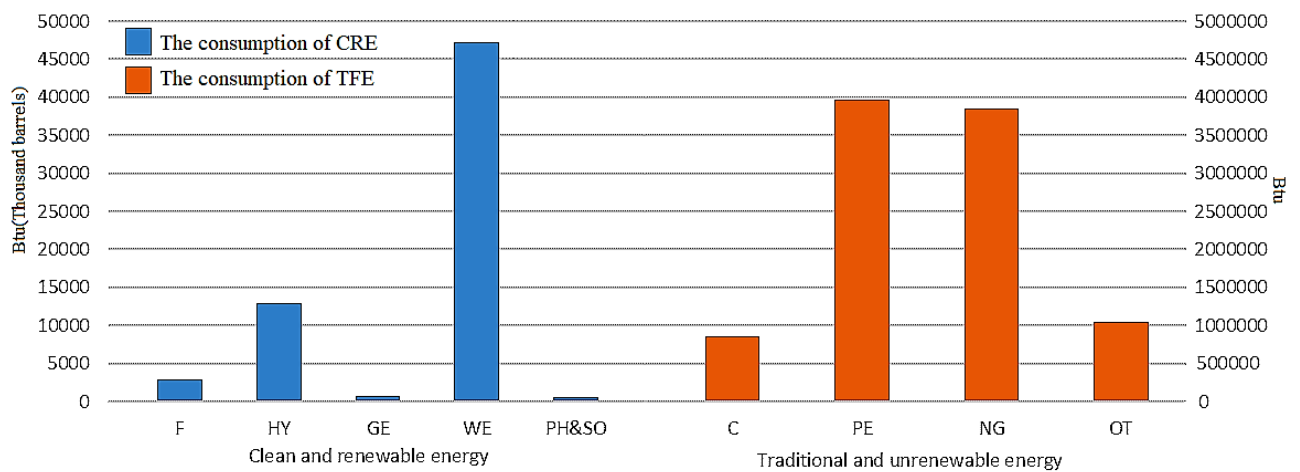
Figure 18: Means of energy consumption in Texas

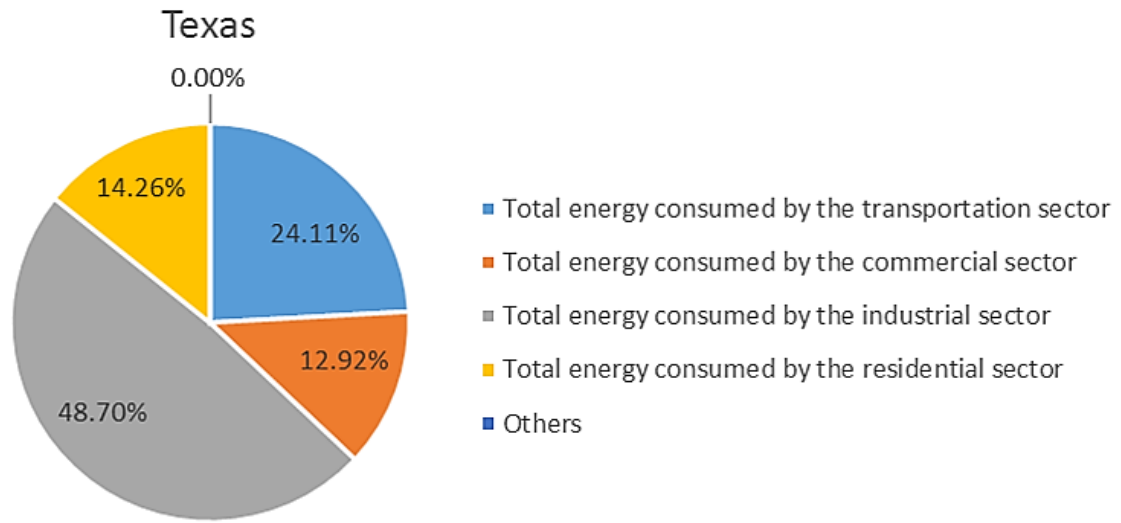| Figure 18 :			
Pic1	Trend graph of TFE and its price		
Pic2	Trend graph of CRE and its price		
Pic3	Trend graph of total consumption(TC) and its price		
Pic4	Trend graph of the ratio of TFE to TC and CRE to TC		
Figure 19 :			
F	Fuel enthanol	C	Coal
HY	Hydroelectricity	PE	Petroleum products
GE	Geothermal energy	NG	Natural gas
WE	Wind energy	OT	Other
PH&SO	Photovoltaic and so-lar thermal energy		
(Only the unit of "F"is Thausand Barrels)			

Table 17: the cutline in Figure 2 and Figure 3

Figure 19: The constitute of total energy consumption in Texas in 2009

Variable	Quantitative Change	Rate of Change
Energy total production	6319816.106(Billion Btu)	142.26%
Renewable energy total production	306479.838(Billion Btu)	611.07%
Fossil fuelS, total production	6013336.268(Billion Btu)	136.90%
Electricity production	1061765.94759(Billion Btu)	8953.51%
Wood consumption	23236.12241(Billion Btu)	60.67%
Secondary industry consumption	5556015.3619(Billion Btu)	158.61%
Renewable energy total end-use expenditures	24323.858(Million dollars)	5688.46%
GDP	876004(Million dollars)	460.01%

Table 18: The quantitative change and the rate of change of typical variables