2018 Mathematical Contest in Modeling (MCM) Summary Sheet
(Attach a copy of this page to each copy of your solution paper.)

Energy Data Analysis and Interstate Agreement Goals for Four States

Summary

There exists no doubt in the significance of energy to the development of economy and society. In order to better cooperate with other states on energy issues, our team was tasked with setting a series of goals for the interstate new energy compact.

In the first part, we preprocessed the data. We did data screening according to the integrity and redundancy of the information, selecting nine energy consumption indicators as main focuses for analysises. For a more intuitive sense, we plotted a graph that illustrates the changes of energy consumption and proportion over time, from which we examined general trends of energy profile.

Second, since both the overall average growth trend and the differences between individual development trend are needed, we built a double-layer model to achieve a longitudinal analysis of data. As a result, we obtained the characteristics of cleaner and renewable energy usage of each state.

Third, we constructed a cleaner, renewable energy evaluation index, which is the weighted sum of three indicators. The three are GDP, the proportion and the average growth of cleaner and renewable energy. The higher the value of this index, the higher the performance of renewable energy usage. The results showed that Arizona's index was the highest in 2009.

Next, we established a linear regression to predict the energy profile in 2025 and 2050. According to the value of the coefficient of determination, the degree of fitting of the last 50 years and the forecasts are credible without any policy changes by each governor's office.

Finally, we sorted and analyzed the result of the fourth problem, obtaining the goal of renewable energy through the best criterion.

Contents

1 Introduction

1.1 Problem Background

Energy is the significant material basis of economic growth and social development. As the major energy production and consumption power in the world, the U.S. state government attaches great importance to energy policy. The establishment of a interstate new energy contract not only helps to strengthen the cooperation between the states and jointly promote the sustainable, multivariate and clean development of energy but also satisfies the energy demand of the economic development of various industries.

California (CA), Arizona (AZ), New Mexico (NM) and Texas (Texas) plan to create a new realistic energy compact. We are tasked with conducting data analysis to model on the basis of what is known in all four states, and setting a series of goals for new energy contract which focused on cleaner and renewable energy.

Faced with the following difficulties, we try to find some solutions:

1. Analyze the overall situation of energy in four states according to the data in the table.

2. Develop a model to describe the energy profiles and analyze the similarities and differences of the four states.

3. Establish standards to evaluate the use of cleaner and renewable energy.

4. Build a forecasting model and Making predictions about the future energy situation.

5. set goals for interstate energy contract based on the results of the four questions above.

6. Identify effective actions to achieve the goal.

7. Summarize a memo to Governors.

1.2 Analysis and Approach Overview

Any serious statistical analysis begins with substantial work on the data. The data provided had a panoply of issues ranging from energy price, energy production,energy consumption to GDP, factor, and price.

For problem 1, due to the hugeness of data, we only select several important and representative data for analysis. We mainly choose the energy consumption.

For problem 2, in order to take both the overall and individual characteristics into account which is precisely the feature of longitudinal data analysis, we develop the energy profile model based on hierarchical linear. Then by solving and analyzing the result, we can get the similarities and differences in four states. The reason is analyzed in combination with the actual situation, such as geography, industry, population, and climate.

For problem 3, to fully assess the use of cleaner and renewable energy of four states, we combine the three variables in the table, eliminate the influence of dimensionality through normalization and then calculate the weighted sum of these three indexes as assessment criteria. To solve forecasting problem 4, we consider regression models and make

linear fitting for historical data of all kinds of energy.

For problem 5.6.7, the solutions are based on the results of the first four questions.

2   Assumptions

In order to simplify the course of modeling and draw some reasonable conclusions from our model, we make assumptions as follows

1. We assume that the data given by the title is true and reliable, the impact of human factors on the data is not considered.

2. We assume that the data given by the title is true and reliable, the impact of human factors on the data is not considered.

3 Symbols and Notations

Symbol or Notation	Specification
β_{0i}	the intercept of the equation
β_{1i}	the regression coefficient
ε	the residual
G	GDP
P	Average growth of clean and renewable energy within five years.
S	the proportion of clean and renewable energy
V	Average growth of clean and renewable energy within five years.
SST	the total sum of squares
SSR	the regression sum of squares
SSE	the sum of squares of residual
R^2	the coefficient of determination

4 Energy profile of four states

There are a lot of variables in the dataset. And we select nine representative indicators to analyze the energy profile of four states.We list them as follows:

MSN	Description
CLTCB	Coal total consumption
BMTCB	Biomass total consumption
GETCB	Geothermal energy total consumption
NGTCB	Natural gas total consumption
PMTCB	All petroleum products total consumption excluding fuel ethanol.
SOTCB	Photovoltaic and solar thermal energy total consumption
NUETB	Electricity produced from nuclear power
HYEGB	Hydroelectricity produced by the electric power sector
WYEGB	Electricity produced from wind energy by the electric power sector

Table1: the selected nine indicators and the specific meaning.

From the data, we can easily see that energy consumption in the four continents shows a slight fluctuation and a slow rise in 1960-2009.

Depicted below are graphs of the evolution of Arizonaąŕs energy consumption structure.

As can be seen from table 1, the overall trend in Arizona (AZ) shows that the

usage of all kinds of energy is on the rise. During the 1960s and the 1970s, oil and natural gas were the main components, and the proportion of both of them accounted for more than 95% of the total energy consumption. After the 1970są́ oil crisis, the usage of coal was constantly improving. Then the proportion reached 43.19% in the whole energy structure in 1982. By 1985 nuclear power begins to be adopted. However, the momentum of nuclear power plantą́ development was blocked because of Three Mile Island and Chernobyl accident in 1986, but then quickly restored to the original state. After 2000, the four main energy consumption was basically stable.

The chart below shows a variation of the California energy consumption structure.

As we can see from Figure 2, the two fossil fuels, oil and gas, are the main sources of energy in California (CA). In 1960 the proportion of oil and natural gas in the energy consumption structure has already up to 56.42% and 39.94%. Even in 2009 they were still 52.62% and 35.86% , far more than other sources of energy and took the absolute advantage. Over the past 50 years their use has been increasing year by year. .At the same time, it can be seen that in the mid-1980s, the usage of nuclear energy increased slowly, and remained around 5% in the later period. In addition, other types of energy consumption are very small.

New Mexico energy consumption structure changes as below:

It can be seen from figure 3 that at first New Mexico (NM) mainly used gas and oil. The proportion of coal increased in the energy consumption structure, which has risen from 1.25% in 1960 to 37.40% in 2009. Ultimately, the three are relatively stable in the overall energy ratio. The consumption of oil and coal increases over time, while natural gas fluctuates in a large range, and overall consumption does not change much. Other sources of energy account for very little of the whole energy.

The chart below shows the change of Texas energy structure.

Texas initially used two kinds of fossil fuels, natural gas and oil, and their proportion in the whole energy consumption structure was 63.54% and 35.03% respectively in 1960.

In the middle of 70s, the usage of coal began to increase, and the proportion in the whole energy consumption structure also increased gradually. It reached
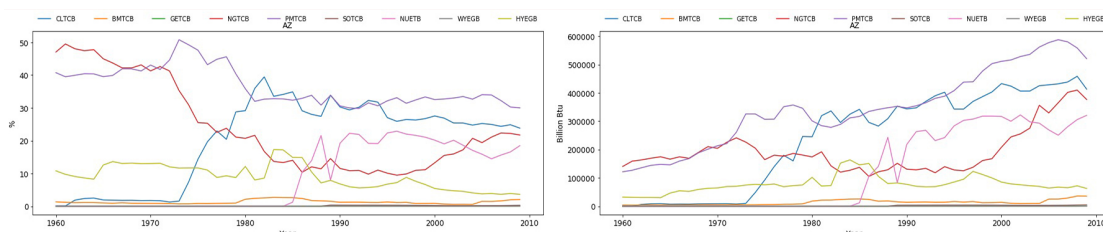

Figure 1: The change of the proportion of energy consumption, (left) vs. the change of the value of energy consumption (right) in Arizona, and the abscissa is the year.
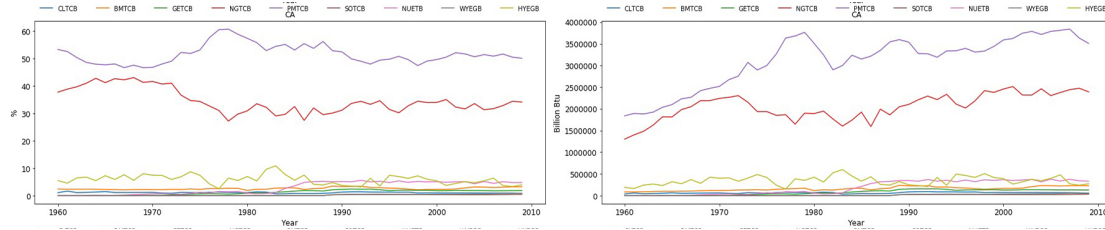
Figure 2: The change of the proportion of energy consumption, (left) vs. the change of the value of energy consumption (right) in California, and the abscissa is the year.
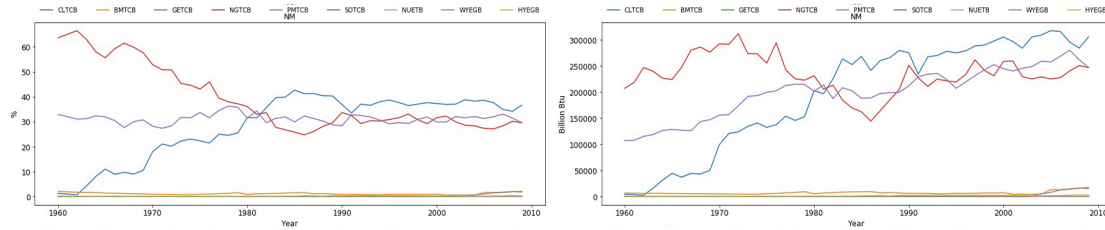

Figure 3: The change of the proportion of energy consumption, (left) vs. the change of the value of energy consumption (right) in New Mexico, and the abscissa is the year.
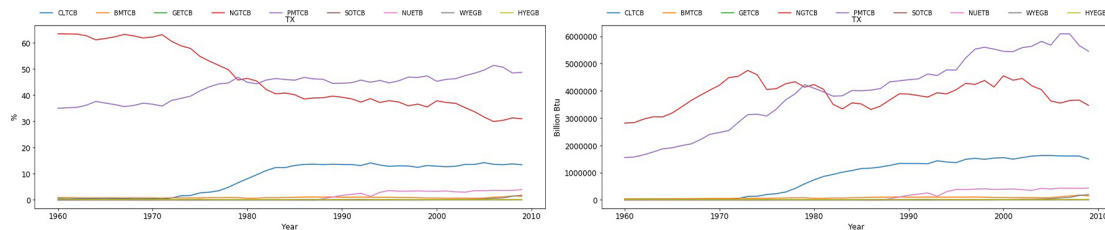
Figure 4: The change of the proportion of energy consumption, (left) vs. the change of the value of energy consumption (right) in Texas, and the abscissa is the year.

13.53% in 1986, and after that remained stable. The consumption of oil is increasing year by year, eventually replacing natural gas and becoming the most used energy in Dirk Seth. Nuclear energy began to be used in the late 90s, but the scale is small, and the proportion is only about 4%.

5   Energy Profile Model

Our task is to build a model that describes the energy profiles of four states from 1960 to 2009.

The data provided in the dataset is a longitudinal data, which is a repeated measure of the four states that can reflect the energy profile. In the analysis of longitudinal data, we mainly focus on two problems, one is to describe the overall average growth trend, and the other is to describe the differences in the growth trend among different individuals.

For the characteristics of longitudinal data., we adopt the energy profile model based on hierarchical linear. The hierarchical linear model can not only analyze

the overall average growth trend, but also pay attention to the differences between individual development trend and explain the reasons for the difference. That is to say, our model can describe the four statesąŕ development general situation of the energy, from 1960 to 2009, and also provide some parameters to illustrate the similarities and differences between states.

5.1　The establishment of a double-layer model

In order to better reflect the energy profile of the four states, as well as the use of cleaner and renewable energy, we develop a double - layer linear model. The first layer shows the overall situation of energy, and the second layer can express the consumption and change trend of cleaner and renewable energy. Here is a brief introduction to the principles of the double-layer linear model.

To start with, in the first layer of data structure, taking tracking observation results as dependent variables and observing time as independent variables, we establish the first layer regression equation:

$$Y_{ij} = \beta_{0i} + \beta_{1i}X_{ij} + \varepsilon_{ij} \tag{1}$$

In the equation, the subscript $0, 1, i$ and j means the intercept, the slope, the sequence number of the observed object and the number of times of observation.β_{0i} is the intercept of the equation, whose meaning is the average of this observation. β_{1i} is the regression coefficient, which represents the rate of change of the observation. X represents the value of the independent variables. ε is on behalf of the residual, meaning the part that the measured value Y cannot be explained by the independent variable X.

In the second layer, we build two second-floor regression equations with the intercept and slope of the equation (1) as the independent variable, the individual characteristics of the observed objects or the experiments accepted as dependent variables. It is as follows:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}W_{1i} + \mu_{0i} \tag{2}$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11}W_{li} + \mu_{1i} \tag{3}$$

Some of the variables in the equation can be listed as below:

γ_{00}	the intercept of the equation (2)
γ_{01}	the regression coefficient of the independent variable W_1 in equation (2)
μ_{0i}	the residual of the equation (2)
γ_{10}	the intercept of the equation (3)
γ_{11}	the regression coefficient of the independent variable W_1 in equation (3)
μ_{1i}	the residual of the equation (3)

Table 2: variables and explanations

5.2   The solution of model

Taking California for example, we use the nine variablesąŕs corresponding energy consumption and proportion every five years to create two data files (by SPSS statistic software). The contents of the two files are as follows.

File 1		File 2	
energy types	measuring time	energy types	proportion in 1960
	energy consumption		Whether it is a renewable energy
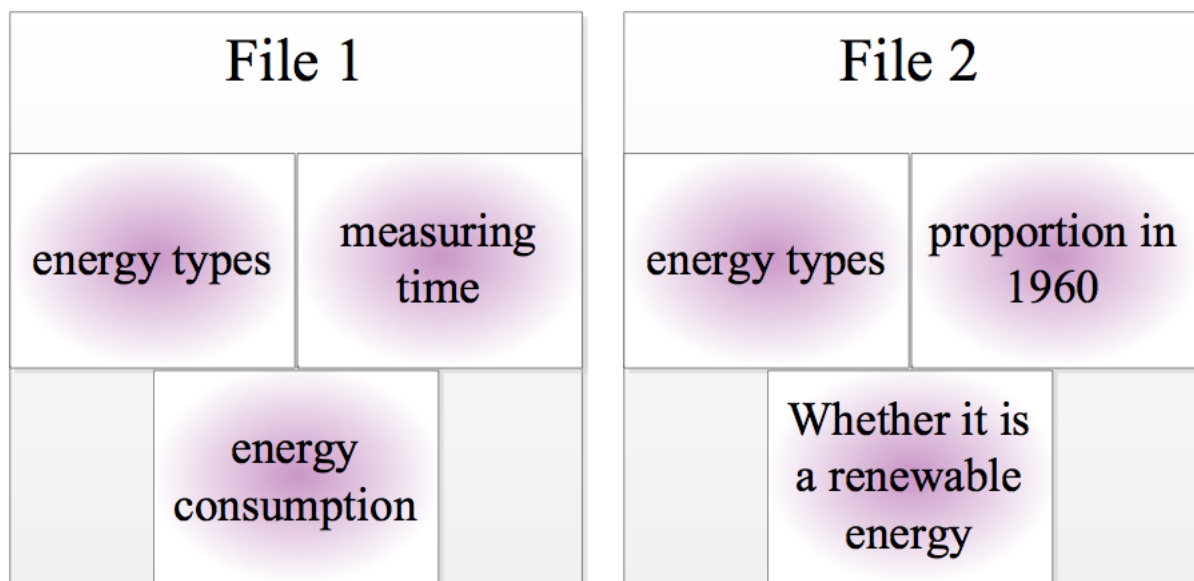
Figure 5: the contents of the two files

In the double-layer model, the equation of the first layer is.

$$Y = \beta_0 + \beta_1 \times t + \varepsilon \tag{4}$$

the equation of the second layer is

$$\beta_0 = \gamma_{00} + A\gamma_{01} + B\gamma_{02} + \mu_0 \tag{5}$$
$$\beta_1 = \gamma_{10} + A\gamma_{11} + B\gamma_{12} + \mu_1 \tag{6}$$

The specific meanings of each variable in the energy profile model are listed below:

Y	energy consumption	
t	measuring time	
A	0	renewable energy
	1	non-renewable energy
B	proportion in 1960	
γ_{00}	the intercept	
γ_{10}	the intercept's rate of change	
γ_{01}	renewable energy effect.	
γ_{11}	renewable energy effect's rate of change	
γ_{02}	energy initial proportion effect	
γ_{12}	energy initial proportion effect's rate of change	

Table3: specific meanings

In addition, considering the energy proportion's score is 0 to 10, in order to explain conveniently, we centralized energy proportion. After parameter estimation(by HLM6.0), the results are shown in table4:

	coefficient
γ_{00}	807764
γ_{01}	−4916770
γ_{10}	77479. 22

Table 4: double-layer model parameter estimation results of California

The chart has shown that the current energy consumption of clean energy is 807764 billion btu, which is 491,6770 billion btu lower than non-clean energy and the energy consumption trend of clean energy is increased by 77479 billion btu every five years, with an average growth rate of 16.02%.

Data from the remaining three states were shown in table 5:

	AZ	NM	TX
γ_{00}	222535	11831	273548
γ_{01}	-440221	-500009	-8274811
γ_{10}	38803	2826	74054

Table 5: double-layer model parameter estimation results of other three states

From the above table, we can know that the four states of energy is similar in the constant improvement of clean energy consumption and the possession of proportion in the whole energy structure also increases year by year. While the development speed is relatively slow, for the reason that the high cost of the use of renewable energy and people still tend to use fossil fuels.The entire energy structure is still dominated by non-renewable energy sources such as oil, gas and coal, which is expected to continue for some time.

The differences between the four states are that there are more people in California and Texas, so they consume more energy than the other two states.They develop different types of clean energy, such as California, which has plenty of light and is suitable for the development of solar energy. Arizona is rich in water resources and can promote hydropower generation.Texas has rich wind energy, which can boost its use of wind power. And New Mexico is rich in most energy, so the development of cleaner and renewable energy is slow compared to other states.

6　Performance evaluation model for cleaner, renewable energy

To determine which of the four statesʹŕ performance of using cleaner, renewable energy in 2009 is the best, we build an index to quantify their use of clean and renewable sources of energy, which is convenient for our choice.

On the basis of existing data in the dataset, we can easily find that there are

three indicators well reflecting the four statesąŕs use condition of cleaner and renewable energy in 2009. They are GDP, the ratio of cleaner and renewable energy to total energy consumption and the average growth of cleaner and renewable energy in the last five years of each state. And the weighted sum of these three indexes is what we want.

6.1 Reasons for choice

By referring to relevant information,we find that growth of GDP leads to increase in energy demand and consumptive endogeneity. Thus, the value of GDP reflects the energy consumption. Through cleaner and renewable energy use accounting for the proportion of total energy consumptionąŕs statistics, we can compare it visually. The average growth can help us learn about the potential for clean and renewable energy in the future.

6.2 The establishment of indicators

In the following equation, the variables and their meanings are as bellow:

G	GDP (Million dollars)
S	The proportion of clean and renewable energy (%)
P	Average growth of clean and renewable energy within five years.
V	Final evaluation index
k_1	The weight of G
k_2	The weight of S
k_3	The weight of V

Table 6: the variables and their meanings

$$V = k_1 G + k_2 S + k_3 P \tag{7}$$

The results of three indicators are calculated as follows:

	G (Million dollars)	S (%)	P (%)
AZ	230854	24.6%	30.08%
CA	1736901	7.6%	16.02%
NM	69052	1.1%	23.02%
TX	1066437	4.3%	40.62%

Table 7: the results of three indicators

In order to eliminate the influence of dimension and data value range on the result, we normalized the data:

$$x^* = \frac{x - min}{max - min} \tag{8}$$

We take $k_1 = 0.1, k_2 = 0.7, k_3 = 0.2$,calculating the states' V as below:

	AZ	CA	NM	TX
V (%)	82.33%	29.36%	5.69%	35.51%

Table 8: the results of V

The results show that Arizona used clean energy and renewable energy in 2009, showing the best effect.

7   Linear regression prediction analysis

In order to make predictions about the energy profile of four states in 2025 and 2050 without any policy changes, we make linear fitting for historical data of all kinds of energy.

Regression analysis is a statistical analysis method to determine the quantitative relationship between two or more variables. If there is only one independent variable and one dependent variable included, and the relationship between the two can be approximated by a straight line, then we call it linear regression analysisy.

Now we define the symbol. We define X as the independent variable in which case X is the year and Y as the dependent variable, namely the value of the energy type that needs to be predicted. a and b is the correlation coefficient. Linear regression expressions are as follows:

$$Y = aX + b$$

To obtain a, b, we need to solve $\min Q(a,b)$:

$$\min Q(a, b) = \sum_{i=1}^{n}(Y_i - (aX_i + b))^2$$

Reduce it to get:

$$Q(a, b) = n\overline{Y^2} - 2an\overline{XY} - 2bn\overline{Y} + a^2 n\overline{X^2} + 2abn\overline{X} + nb^2$$

To get the minimum of $Q(a, b)$, we find the partial derivatives of Q for a and Q for b , then make it all zero.

$$\frac{\partial Q}{\partial a} = -2n\overline{XY} + 2an\overline{X^2} + 2bn\overline{X} = 0$$

$$\frac{\partial Q}{\partial b} = -2n\overline{Y} + 2an\overline{X} + 2nb = 0$$

Then we get the formula:

$$a = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{(\overline{X})^2 - \overline{X^2}}$$

$$b = \overline{Y} - a\overline{X}$$

We use the (the coefficient of determination) to determine the fitting degree of the regression equation.

The total sum of squares (proportional to the variance of the data):

$$SST = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

The regression sum of squares, also called the explained sum of squares

$$SSR = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2$$

The sum of squares of residuals, also called the residual sum of squares:

$$SSE = \sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 = \sum_{i=1}^{n}e_i^2$$

The most general definition of the coefficient of determination is

$$R^2 = 1 - \frac{SSE}{SST}$$

The partial fitting curve are as follows.

Finally, we have the following results and we listed the partial of them:

Year	State	MSN	Data (Billion Btu)	Year	State	MSN	Data (Billion Btu)
2025	AZ	WYEGB	33.24612	2025	CA	WYEGB	64042.66
2050	AZ	WYEGB	50.20843	2050	CA	WYEGB	93905.81
2025	AZ	NUEGB	472573.5	2025	CA	NUEGB	584283.2
2050	AZ	NUEGB	686576.9	2050	CA	NUEGB	828888.8
2025	AZ	SOEGB	73.62086	2025	CA	SOEGB	8533.295
2050	AZ	SOEGB	110.5323	2050	CA	SOEGB	12469.39
2025	AZ	HYEGB	107878.2	2025	CA	HYEGB	396039.7
2050	AZ	HYEGB	126652.1	2050	CA	HYEGB	428247.9

Year	State	MSN	Data (Billion Btu)	Year	State	MSN	Data (Billion Btu)
2025	NM	WYEGB	7772.633	2025	TX	WYEGB	70670.05
2050	NM	WYEGB	11675.86	2050	TX	WYEGB	106136.3
2025	NM	NUEGB	0	2025	TX	NUEGB	589664.3
2050	NM	NUEGB	0	2050	TX	NUEGB	864801
2025	NM	SOEGB	0	2025	TX	SOEGB	1.738693
2050	NM	SOEGB	0	2050	TX	SOEGB	2.406394
2025	NM	HYEGB	3585.202	2025	TX	HYEGB	16226.46
2050	NM	HYEGB	4923.45	2050	TX	HYEGB	18270.45

Table 9: the results of partial predictions

AZ

- CLTCB: $R^2=0.90$ y=11057.79x+-21703542.22
- BMTCB: $R^2=0.49$ y=423.87x+-828076.22
- GETCB: $R^2=0.77$ y=8.12x+-15994.99
- NGTCB: $R^2=0.21$ y=2457.19x+-4682277.02
- PMTCB: $R^2=0.93$ y=9034.48x+-17583080.04
- SOTCB: $R^2=0.71$ y=105.95x+-208722.62
- NUETB: $R^2=0.80$ y=8560.13x+-16861698.94
- WYEGB: $R^2=0.06$ y=0.68x+-1340.70
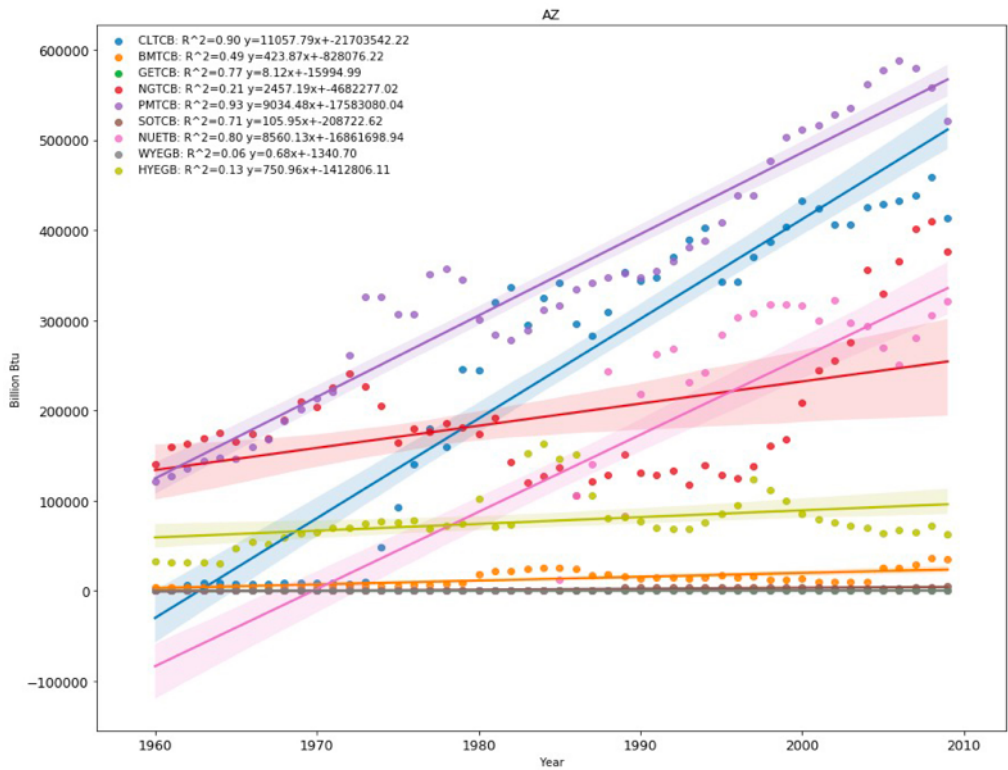- HYEGB: $R^2=0.13$ y=750.96x+-1412806.11
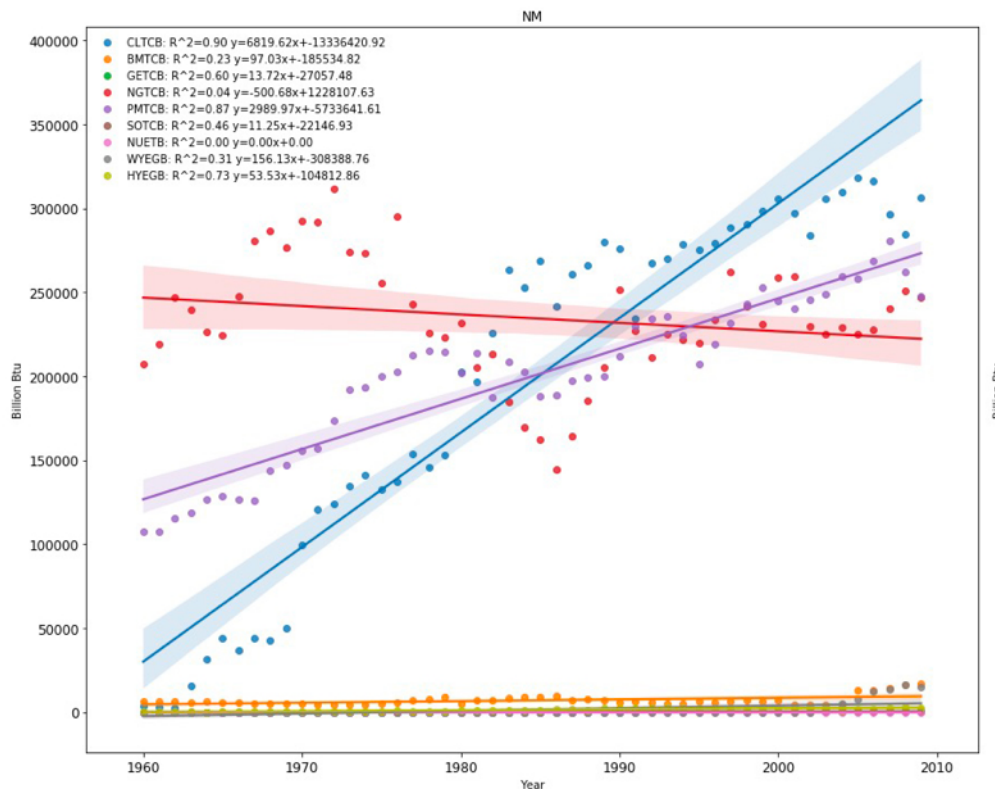
Figure 6: Fitting curve of Arizona

Figure 7: Fitting curve of New Mexico

8 The goals for renewable energy use in 2025 and 2050

In the first part, we have predicted the energy profile of eachstate in 2025 and 2050, as well as the proportion of renewable energy in the whole energy structure system. The results can be listed as bellow:

Year	State	The renewable energy（Billion Btu）	The total energy (Billion Btu)	proportion
2025	AZ	114173	2310769	0.0494091
	CA	725057	8816443	0.0822392
	NM	12719	3619707	0.0035138
	TX	89224	12814598	0.0069627
2050	AZ	135816	3120748	0.0435203
	CA	896725	10544201	0.0850444
	NM	18585	1273259	0.0145964
	TX	127841	19403560	0.0065885

Table 10: The prediction results

As can be seen from the above table 10, the predicted value of renewable energy is very small in the whole energy structure, which may be due to the small base of renewable energy and the insufficient growth of renewable energy in the initial stage. But for the development of the future energy structure, this circumstance is not science.According to the best standard, referring to a five-year

average growth, the proportion of renewable energy in the whole energy of four areas in 2025 and 2025 are as shown in table 11:

Year	State	proportion	The renewable energy （Billion Btu）
2025	AZ	54%	1247815
	CA	11.92%	1050920
	NM	2.05%	74203.99
	TX	11.80%	1512123
2050	AZ	83.12%	2593966
	CA	18.70%	1971766
	NM	3.81%	48511.17
	TX	32.38%	6282873

Table 11: The new prediction results

The new renewable energy sources will be added to the renewable energy targets in 2025 and 2050, and the goals we got are 3885061 billion btu and 10897115 billion btu.

9   Three actions

1. Increase the number of large-scale production equipment, which can aggrandize the capacity of renewable energy in a relatively short time.

2. Sign the purchase agreement of electricity for large orders and obtain renewable energy sources elsewhere.

3. Develop a diversified energy collection model and make full use of renewable energy resources.

10   The Conclusion

In our study, we have done lots of data processing work, putting forward the energy profile model. We also constructed a cleaner, renewable energy evaluation index which can be used to quantify the statesąś use of clean and renewable sources of energy.In addition, we made predictions about the energy profile and set goals for the interstate energy agreement.

11   Summary memorandum

Introduce the total energy consumption, the proportion of clean energy proportion in the whole energy structure and its growth rate from 1960 to 2009, let the audience can have a general knowledge of energy situation.

What is the development trend of different types of energy? Seize the tide of The Times.

With the absence of any policy changes, my forecast of the use of renewable energy will reach 941173 Billion Btu and 118967 Billion Btu in 2025 and 2050.

We suggest that the new energy target be set in a higher level, which will be 3885061 Billion Btu and 10897115 Billion Btu in 2025 and 2050.

References

[1] zhang xu. Analysis of multi-energy system model based on regional energy supply [D]. Shanghai jiaotong university,2015

[2] zhang long. Research on performance evaluation of renewable energy development [D]. China university of geosciences,2017

[3] `https://www.jianshu.com/p/fcd220697182`