Team Control Number

For office use only
T1 ________________
T2 ________________
T3 ________________
T4 ________________

82392

For office use only
F1 ________________
F2 ________________
F3 ________________
F4 ________________

Problem Chosen

C

2018
MCM/ICM
Summary Sheet

(Your team's summary should be included as the first page of your electronic submission.)

Type a summary of your results on this page. Do not include the name of your school, advisor, or team members on this page.

Energy production and usage are a major portion of any economy. In the United States, many aspects of energy policy are decentralized to state level. In our paper, we construct a series of models to analyze the energy profile for each four states in the U.S. and help their governors do some prediction, and succeed in setting up goals for this new four-state energy compact.

In the first part, we choose all the ‘total consumption’ data to construct most part of our model. We work out the ratios of the consumption of each kind of resource to the total energy consumption (energy ratio) as a standard of measuring energy profile. We also classify the ratios into two groups, one for renewable energy, and another for non-renewable energy.

In the second part, we build a Time Series Model to fit the reality and predict the future energy profile in each state by applying foregoing model based on our analysis of four states’ energy profiles. Moreover, our model has strong maneuverability and is reasonable enough to accommodate all kinds of energy consumption in each state.

After that, we use Multiple Linear Regression Method (MLRM) to find the similarities and differences among the four states. This model contains general information about how a state’s own characteristics, such as population, climate, impacts its energy ratio. Then, based on PCA and TOPSIS method, we analyze which of the four states appeared to have the “best” profile for use of cleaner, renewable energy in 2009.

In the third part, we make a key assumption that all the energy consumed are used to generate electricity. After that, we used the policy-decision function to determine renewable energy usage targets for future, and state them as goals for this new four-state energy compact. We use the function to help us to determine the policy to meet above goals. According to the above analysis, we come up the goals that four states should use more renewable energy to generate electricity.

Finally, we analyze the fitting and forecast values accuracy of Time Series Model by residuals and square residuals plot. The result shows our model is comparatively robust.

MEMORANDUM FOR: The governors for four states

FROM: Members form team 82392

SUBJECT: Summary and predict of energy profile, recommended energy compact

1. Energy profile from 1960 to 2009 in each of the four states

- From the prospective of proportion, the sort of application amount of renewable energy is $AZ > CA > NM \approx TX$.
- From the trend of change, all the four states have shown increasing trend. The use of renewable energy in CA and AZ shows a greatly fluctuation, whereas TX and NM are small and smooth.
- The sort of renewable energy profile is $AZ > CA > NM > TX$.

2. Prediction of energy profile in 2025 and 2050 in each of the four states

- If there is no policy change, only AZ will use renewable energy as the main source of energy. Other states will use more than half of the non-renewable energy.
- The proportion of using clean energy in four states is going up obviously. The amount of non-renewable energy usage will appear an upward trend in NM and TX without policy change.

3. Recommended goals for the energy compact

The goals of energy compact are measured by the target ratios of renewable energy generation to total electricity generation in four states.

- For California, the target ratio is 50% in 2025 and 64% in 2050.
- For Arizona, the target ratio is 24% in 2025 and 30% in 2050.
- For New Mexico, the target ratio is 10% in 2025 and 14% in 2050.
- For Texas, the target ratio is 20% in 2025and 38% in 2050.

4. Possible actions

- Give 6% tax discount in the following 8 years for all the firms related to renewable energy resources.
- Give 50% forever tax discount for the electricity generating from renewable and clean energy.
- Charge 20% more for net electricity overuse in a family.
- Invite firms to tender for a solar-electricity generating station project in Arizona.

Contents

1 Introduction

1.1 Background

If it were a decade ago, you cannot imagine that the homepage of "Twitter" can be opened more than millions in one minute; you cannot imagine that when you open "google maps", all the travel information is already in the palm of your hand; you cannot imagine that through data mining you can specify a country's future decision to guide its development. Nowadays big data has penetrated into our work and lives, and has brought such huge changes. In addition, it has become particularly important for us to find useful information from the mass of data to guide our work and life.

Nowadays, as the result of the repaid development of technology over the few years, we can definitely say that the Big Data Era has come. In the economic, administration and other areas, information will increasingly be made based on data mining and analysis, rather than based on experience or intuition. Also, government will use this method when making decision. Energy production and usage are a major portion of any economy, and governments has an obligation to control the production and consumption of various kinds of energy. By analyzing data, it could make decision more effective and efficient.

This article is about to formulate a realistic new energy compact for four states (California (CA), Arizona (AZ), New Mexico (NM), and Texas (TX)) focused on increased usage of cleaner, renewable energy sources. We aim to design a system of the trend on the usage of energy based on large quantities of data through data mining methods. To solve the problem, we will use the Time series Theory, TOPSIS method, Principal Component Analysis and some other methods to inform their development of a set of goals for their interstate energy compact.

1.2 Overview of Our Work

The flow chart of the whole model is presented below in Figure 1:

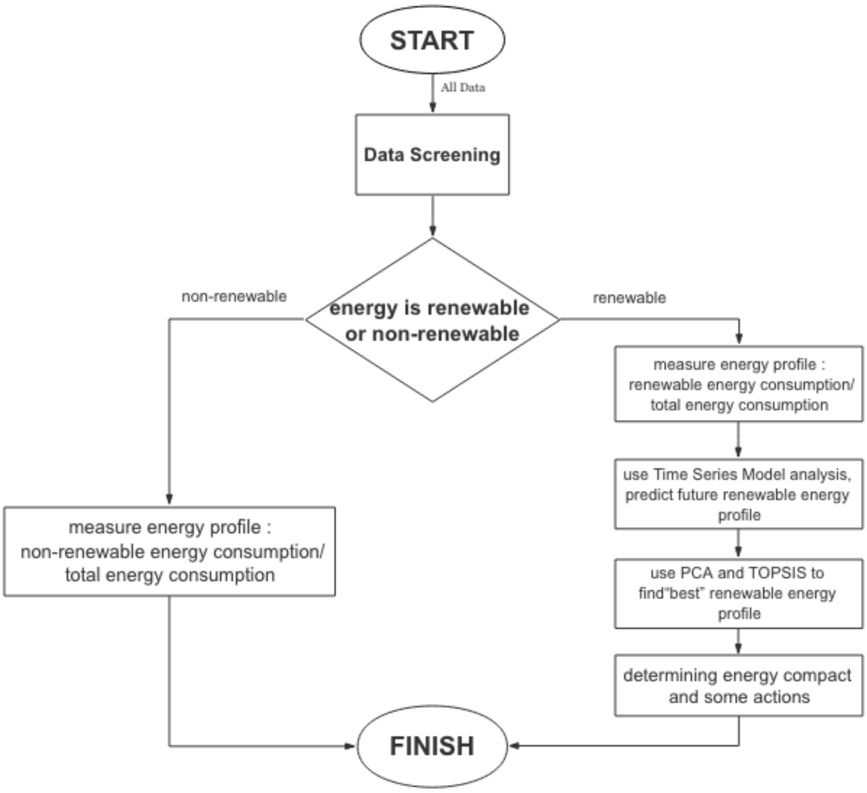


Figure 1 - Overview of our work

2 Data Screening

2.1 Assumptions

Due to limitation about the data, we make the following basic assumptions in order to simplify the problem. These simplified assumptions will be used through our paper and can be improved with more reliable data.

- The data collected are true and effective.
- The development of population, economy, society and environment are consistent to the planned goal.
- There is no change of policy when we making predictions.
- The impact of policy is deeply.
- Neglect the discount rate of the capital.
- All the energy is used to generate electricity.

2.2 Data Cleaning and Filtering

Any serious statistical analysis begins with substantial work on the data. Cleaning and screening data was prerequisite to model fitting and analysis.

First, data managing was performed in Python, where we reorganized variables into a more useful structure. For example, we deleted the same energy consumption measured by different units in the data. Because this situation will cause multicollinearity in our model.

Next, we also analyzed the missing data by Python. Python has many useful tools for dealing with missing data problems. We plot a picture to illustrate data loss situation. Figure 1 shows the initial situation.

In figure 1, the X axis represents all variables in the data base, the Y axis represents time. Black shadows represent missing data. It can be clearly seen from the figure that data from 1960 to 1970 are almost all missing, so we decided to remove the data in the first ten years, which did not affect the results of the model. Also, we will remove the variables that almost all the data are missing. For example, what we can see a very thick black line on the left of Figure 1, shows that for those variables, most of the data are missing.

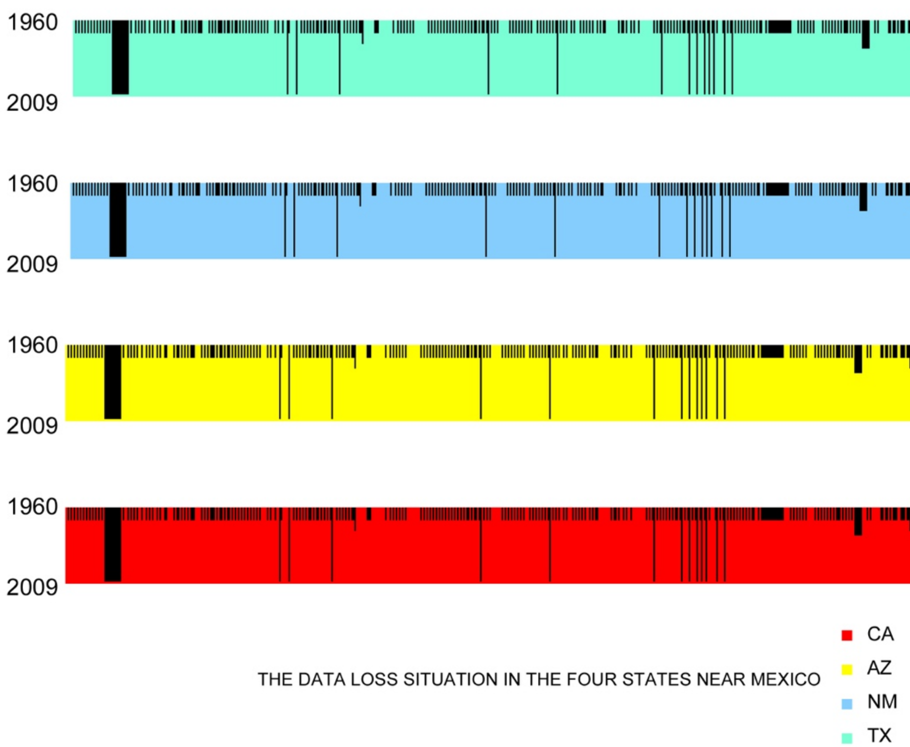


Figure 2 - Data Loss Situation

3 Model Construction

3.1 Energy Profile

By analyzing and screening our data, we roughly divide energy into two, renewable and non-renewable sources of energy. For example, coal, natural gas, and petroleum are non-renewable energy sources, while non-renewable energy sources have ethanol, solar energy, nuclear energy and so on.

In determining the energy profile of a state, the standard for measuring the energy profile is the ratio of the consumption of some energy to the total energy consumptio. The calculations were conducted as follows:

$$Total\ energy = \sum all\ kinds\ of\ energy$$

$$Ratio_i = \frac{energy_i}{total\ energy}$$

$$i = gas, coal;\ nuclear\ energy;\ etc.$$

The following five figures give a line graph of the renewable energy comparison in four states and the line graph for the comparison of renewable and non-renewable energy in four states.

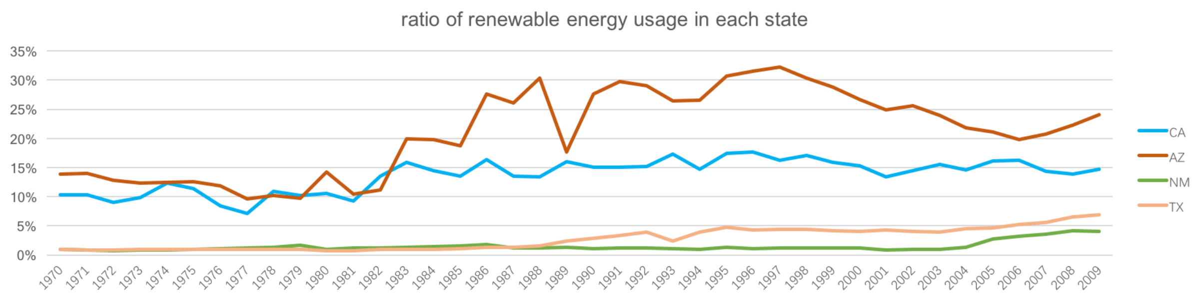


Figure 3 - energy ratio for renewable energy

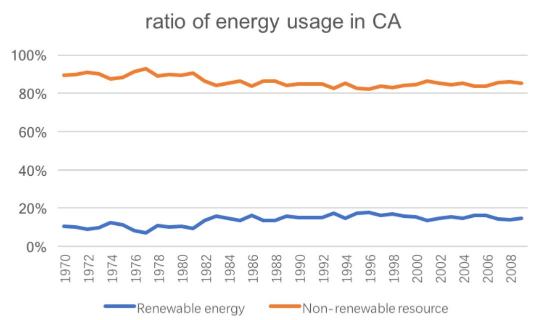
From figure 3, obviously, Arizona has the largest application amount among four states. But there was a rapid decline in 1989. Based on the data, we find that U.S. was established the Biosphere 2, which is a micro - artificial ecological cycle system, in Arizona in 1987.

Thus, we speculate that this system may take up some Arizona's renewable energy resources. The experiment failed in 1993, so the application amount of Arizona's renewable resources was back to the normal level.

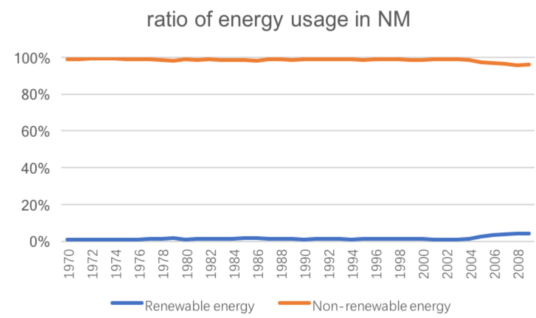
The second largest state of renewable use application amount is California. The line graph shows that the amount of renewable energy usage for California was slightly fluctuation for a long time, but the overall trend is still rising gradually.

The third one Texas. Texas renewable energy has been a dominant force politically and economically within the state. In order to encourage the development of using more renewable energy, Texas signed a series of initiatives to encourage the use of renewable energy at the end of 20th Century. Therefore, the amount of renewable energy has been rising since 21st Century.

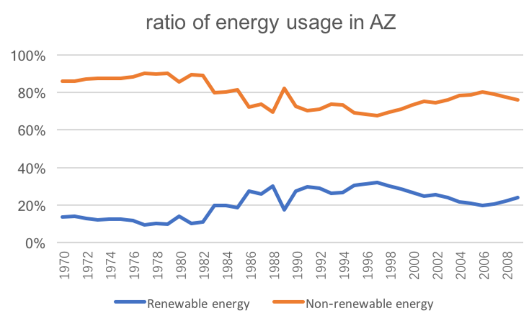
The lowest amount of usage among four is New Mexico. As the port for the U.S. of rocket and missile, the amount of using renewable energy has been very small. Until 2005, the U.S. signed the energy bill. After that, the amount of using renewable energy has only a slight increase.



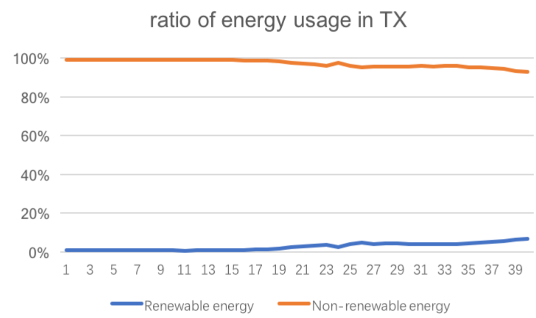
a) CA



c) NM



b) AZ



d) TX

Figure 4 - energy ratio of renewable energy and non-renewable energy

We can clearly see from the above four figures the comparison of the ratio of renewable energy and the ratio of non-renewable energy in four states.

3.2 Time Series Model

Based on the above analysis, we set up a model to describe the evolution of the energy profile over the period 1970-2009. Since the data structure is panel data, we build a Time Series Model from a traditional time series perspective.

After analyze the figure of the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) of renewable energy ratio, we can clearly assert form ACF and PACF patterns that ACF is tail off and PACF is first-order cutoff. So we use ARMA (1,0), that is, AR (1).

In selecting variables, we choose ten most representative indicators of all energy ratios. Continue to use the energy ratio defined above as the model's dependent variable.

Below is the Time Series Model was constructed as follows:

$$\begin{cases} Ratio_t = \alpha_t + \beta \times time + \varepsilon_t \\ \varepsilon_t = \dfrac{\Theta(l)}{\Phi(l)} \nu_t \end{cases}$$

Where $Ratio_t$ is the Ratio in different year, t=1970, 1971, …,2009

ε_t is residual, both $\Theta(l)$ and $\Phi(l)$ are lag operator polynomial, ν_t is White Noise. We choose $l = 1$.

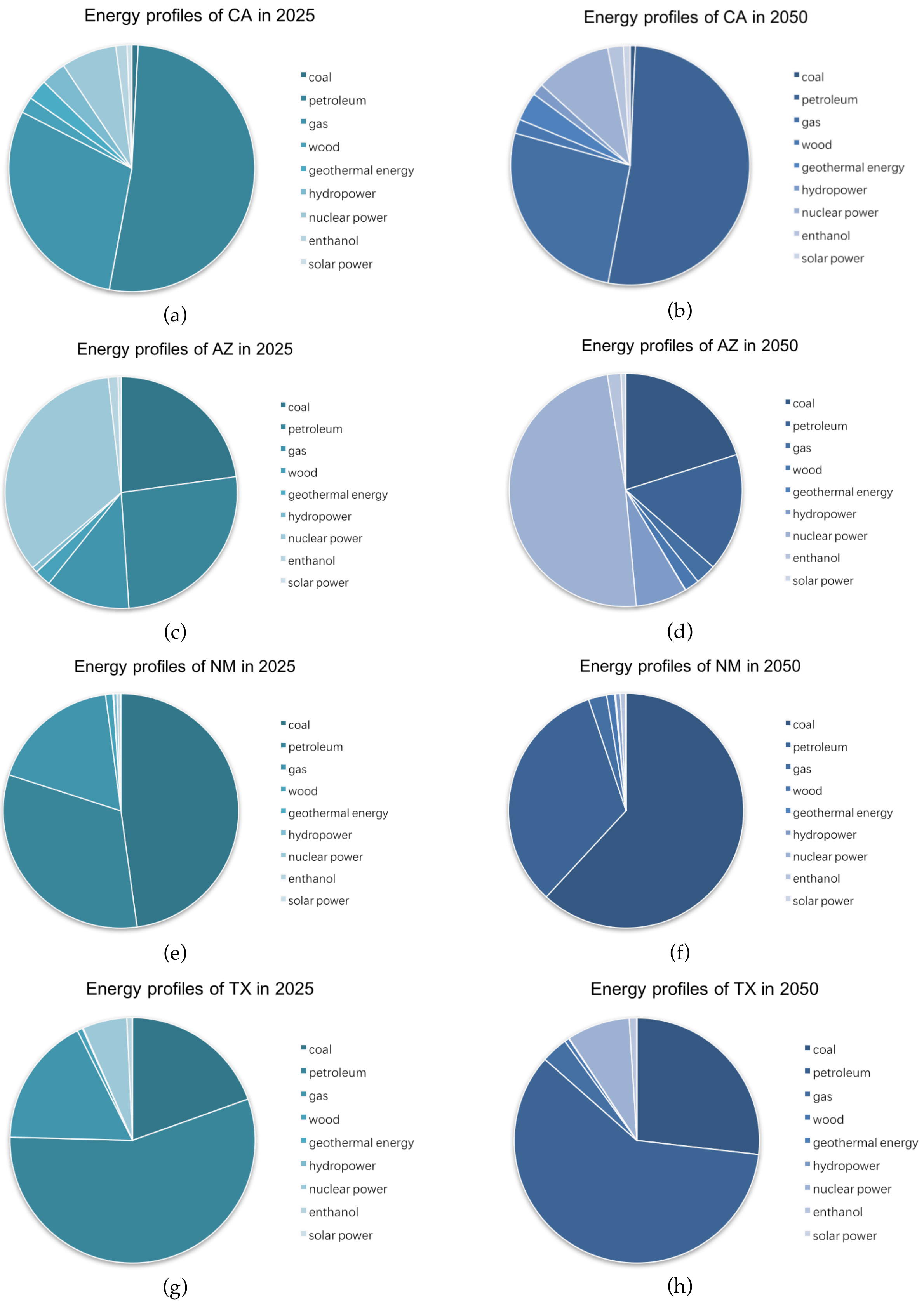


Figure 5 - Proportion of energy consumption in 2025 and 2050

These figures explain the energy structure in each state in both 2025 and 2050

3.3 Result of Time Series Model

Practically each energy profile feature prominent trend and cyclical effects. Based on the assumptions and the requirements of the title mentioned above, we used a Time Series Model constructed before to predict the energy profile of each states in 2025 and 2050. And the prediction is based on the time dummy variables and the real or predicted data of the last year. The Calculation of Time Series Model is completed by R.

Figure 5 shows the proportion of energy consumption of four states in 2025 and 2050 intuitively.

In each pie chart, we used deeper colors to represent relatively unclean energy, and lighter colors to represent relatively clean energy. As a result, we can see the overview of the energy consumption in different states. Besides, we can find out the differences of clean energy consumption between four states.

Some preliminary conclusions:

- According to the eight charts, we predicted that Arizona still has the highest percentage of clean energy consumption ratio, at about 15%. While New Mexico has the lowest percentage, roughly 65%. This trend was basically consistent with ten years ago.
- Nuclear power makes up the largest part of renewable energy, while coal and petroleum comprise the more than 50% of non-renewable energy. Petroleum and coal occupied a major position in the past decades. If there is no policy intervention, these two kinds of energy will still be the main resources used by four states in the future.

In particular, California had a nuclear accident in 2012. Therefore, the government of California had come up with some polices to close the nuclear power station. Although it is not in conformity with the facts, we respect the background of subject and still using data before 2009.

- From the prospective of time series, California has a better prospect in renewable energy. Compare the above pie charts according to the time, this state used more renewable energy and less non-renewable energy in the future. In the contrary, the percentage of energy consumption of other two states have an opposite tendency. Besides, Arizona has no obvious tendency in using energy.

3.4 Multiple Linear Regression Model

In order to compare the similarities, differences and their influencing factors between the four states, we used a Multiple Linear Regression Model(MLRM) to analyze the main energy sources of the four states. We selected population, total biomass, annual air temperature (including mean, maximum and minimum temperature) and annual precipitation as independent variables, and the proportion of energy consumption as dependent variable.

The model was constructed as follows:

$$Ratio_{it} = \alpha_{it} + \sum_{j=1}^{6} \beta_{i,j} \times Characteristic_{it,j} + \varepsilon_{it}$$

Where $Characteristic_{it,j}$ is the factors influence the ratio of energy consumption in each state in each year such as total biomass.

The results obtained by linear fitting include the coefficients and the statistics of each regression. We find out that most of the results fitted surprisingly well no matter in statistic or graphic. Therefore, we believe that these regression coefficients prominently represent the influence of various factors on the proportion of energy consumption in different states.

3.5 Results of Multiple Linear Regression Model

According to the results of the MLRM, we can find out the differences among energy profiles of each state by comparing regression coefficients between each state. Based on the principle of linear regression model, the mathematical meaning of every regression coefficient is the degree of influence of the factors on the energy ratio. The analysis of regression results can be divided into two steps:

Firstly, the regression coefficients have positive and negative. This means the influence for four states of each factor is difference.

The fact represents possible influential factors of the similarities and differences of energy ratio between each state. For example, as the population increasing, the wood consumption of four states all present an upward trend. Whereas, when the average annual temperature increasing, the wood consumption of California and Arizona show a rising trend, while the wood consumption of New Mexico and Texas show a declining trend.

	POPULATION				BIOMASS				ANNUAL_AVE				JAN_AVE				JUL_AVE				PRECIPATION			
	CA	AZ	NM	TX	CA	AZ	NM	TX	CA	AZ	NM	TX	CA	AZ	NM	TX	CA	AZ	NM	TX	CA	AZ	NM	TX
COAL	-	-	+	+	+	+	-	+	+	+	-	-	-	-	+	+	-	-	+	+	+	+	+	+
ETHANOL	+	+	+	+	+	+	+	+	-	-	+	+	-	-	+	-	+	+	-	-	+	-	-	-
ELEC	+	+	+	+	-	-	-	+	+	+	+	-	+	-	-	-	-	-	-	+	-	-	+	-
GEOTHERM	+	+	+	+	+	-	-	+	+	+	+	+	+	-	+	+	-	-	-	-	-	+	-	-
HYDRO	-	-	+	-	-	+	-	+	-	-	+	-	+	+	-	+	+	-	-	-	+	-	-	+
NG	+	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+	+	+	-	-	-	-	-	-
NUCLEAR	+	+	-	+	+	-	-	+	+	+	-	+	+	+	-	+	-	-	-	+	-	+	-	+
PETRO	-	-	-	+	+	-	+	+	+	-	+	-	+	-	-	-	-	+	+	+	-	-	+	-
SOLAR	+	+	+	+	+	-	-	+	-	+	+	+	+	-	-	+	-	-	-	+	+	+	-	+
WOOD	-	-	-	-	+	+	+	+	+	+	-	-	+	+	-	+	-	-	+	-	-	+	+	-
WIND	+	+	+	+	+	+	+	+	-	-	-	-	-	+	+	-	+	+	-	-	-	-	-	-

Figure 6 - positive or negative influence

Secondly, the absolute values of the regression coefficients of the four states are ordinal in view of the factors affecting each kind of energy. This represents the similarities or differences between the degree of influence between each state.

The table below shows the result from Multiple Linear Regression Model. Each cell represents the most sensitive state to the factor.

	POPULATION	BIOMASS	ANN_AVE	JAN_AVE	JUL_AVE	PRECIP[1]
COAL	NM	AZ	AZ	NM	AZ	AZ
ETHANOL	NM	AZ	AZ	AZ	CA	AZ
ELECTRICITY	NM	NM	AZ	CA	CA	NM
GEOTHERM	NM	CA	CA	CA	CA	CA
HYDROELECTRICITY	AZ	AZ	AZ	AZ	AZ	CA
NATRUAL GAS	NM	AZ	AZ	CA	AZ	NM
NUCLEAR POWER	AZ	AZ	AZ	CA	AZ	AZ
PETROLEUM	AZ	AZ	AZ	TX	AZ	AZ
SOLAR POWER	NM	NM	AZ	AZ	AZ	AZ
WOOD	NM	NM	AZ	AZ	CA	AZ
WIND POWER	NM	NM	NM	NM	NM	NM

Figure 7 - Table: the most sensitive state to each factor

[1] ‘ann_ave’ means annual average temperature in each of the four states; ‘jan_ave’ and ‘jul_ave’ means January and July’s average temperature for each of the four states respectively; ‘parcip’ means annual precipitation in each of the four states.

4 Identifying the "best" Energy Profile

First, we use PCA that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. We use Python for this step.

After that, we predict the "best" Energy Profile via TOPSIS method. The TOPSIS is a multi-criteria decision analysis method, which is one of the most famous classic index methods. TOPSIS is based on two concepts: the positive ideal solution (PIS) and the negative ideal solution (NIS). The so-called the PIS is the optimal solution or scheme, and its attribute value reaches the best value of each alternative and vice versa. The rule of plan sorting is to compare each alternative with PIS. The best scheme satisfies that it's the nearest to PIS. This distance is measured by Euclidean geometric distance. For this step, we use Python too. An assumption here is that the criteria are monotonically increasing.

CA	AZ	NM	TX
0.60319	0.69047	0.27879	0.27136

Figure 8 - the 'greatness' data (rounded to 5 valid digits)

So, Arizona appeared to have the "best" profile for use of cleaner, renewable energy in 2009. We can also clearly see in the figure 3 that Arizona performed better than other three states.

5 Suggestion of the WIEC

5.1 Policy Impact on energy resources usage

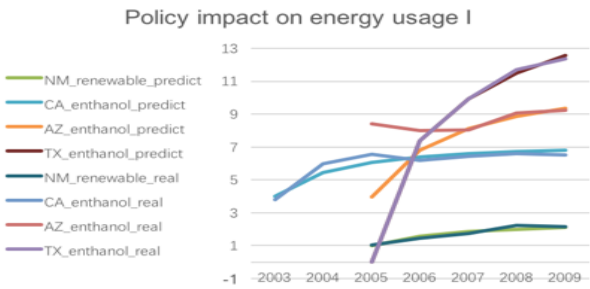
In each state, we made several models to explain how a specific policy infect the usage ratio of each kind of energy resource among all energy used in an annual timescale.

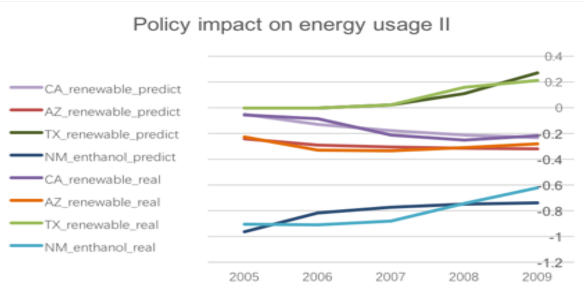
As for the policy, we chose EPACT2005, EISA2007[2] and California Diablo Canyon nuclear generator station, and analyzed the impact of their taking place.

EPACT2005 encouraged the development of clean energy resource and the electricity generating from renewable energy resource, discounting the tax paid by related firms or project group by up to 15%.

---

[2] Also named H.R.6, a law to encourage new energy use





[3]

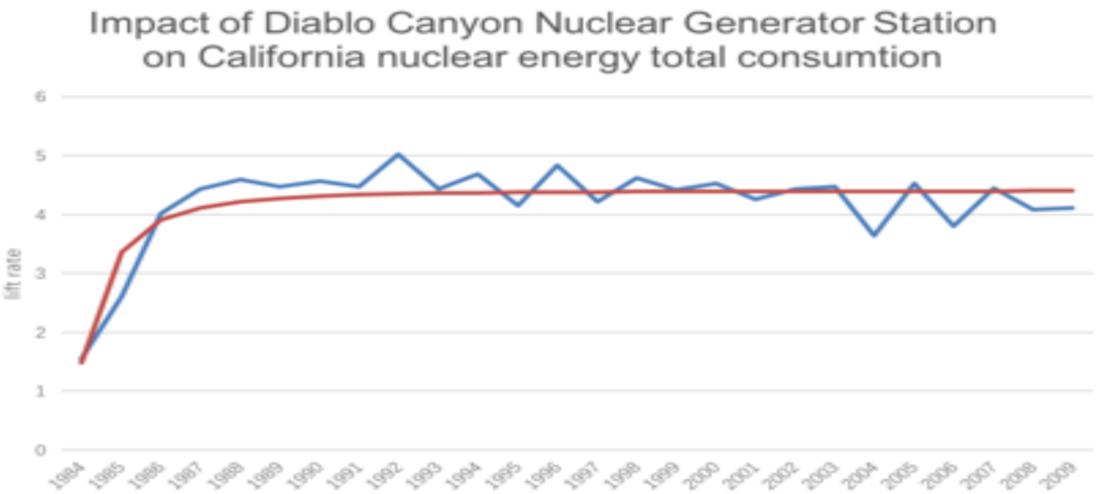


Figure 9 - regression result in picture; the result is almost perfect

One of the restrictions of EISA 2007 is that the car gasoline in 9 areas where ozone-index is over standard must used special fuel with 85% of MTBE, while other fuel contains at least 5.6%. This policy greatly encouraged the production and usage of fuel ethanol.

We describe the impact of a specific policy by lift rate, which is determined by the following functions:

$$lift = real\ value - prediction\ value_{model\ by\ former\ data}$$

$$lift\_rate = \frac{lift}{prediction\ value_{model\ by\ former\ data}}$$

---

[3] In this model, we changed $timedist$ to $timedist^{1.8}$ to meet faster reaction to new hardware.

As long as the policy impact is assumed to be grow over time, and finally reach a peak 'lift value' as the final influence on the field development. Specifically, a policy influences each power ratio stronger and stronger with time goes, and the ratio of real power ratio to the prediction value would finally be near a constant. According to this analysis, we chose the following policy-decision function as the pattern of Policy Impact:

$$lift\_rate = a \cdot e^{-\frac{b}{timedist}}$$

In the formula, $timedist$ is the number of years from the policy-publishing year, a and b are constants.

Now that the Time Series Model included all the data chosen by us, the influence of the policy taken place has already been considered in our previous prediction. Instead of adding further adjust based on our former model, we only need to extract patterns of different stylish of policies on each state.

Based on our chosen policies, we get the following pattern (rounded to three valid digits). In this modeling, we used mathematical manipulation and linear regression to calculate the final pattern, so we used linear correlation coefficient to describe the convenience of our model.

	CA-R	AZ-R	NM-R	TX-R	CA-E	AZ-E	NM-E	TX-E	CA-N[4]
$\boldsymbol{b}$	-1.94	-0.349	-0.918	-18.1	-0.622	-1.07	-0.334	-1.76	-1.09
$\boldsymbol{ln\|a\|}$	-1.07	-1.07	0.921	2.31	2.01	2.45	-0.372	2.88	1.49
$\boldsymbol{\gamma}$	-0.930	-0.724	-0.964	-0.967	-0.954	-0.869	0.654	-0.998	-0.927

Figure 10 - regression result for policy-decision model

The comparison of simulation curves and real data curve is shown in the figures. According to the result, the model better predicts the lift rate of power ratio in a mid-term timescale or in long run than the ratio in recent future. This meet our need perfectly.

[4] The Diablo Canyon Nuclear Generating Station puts a new machine group into electricity generating. The influence works so fast that we

We found California reacts fast to the new policy while New Mexico's power ratio hardly changes with the new policies' publishing, and the impact might appear with years of stalling. Texas, however, used to have very few renewable resource usage, so they get up to others with a high speed.

5.2 The goal of using renewable energy

By processing the model above, considering the potential, reaction pattern and altitude to compact rules, we worked out an appropriate policy-decision function for each state.

	CA-R	AZ-R	NM-R	TX-R
b	-1.94	-0.349	-0.918	-18.1
a	-1.07	-1.07	0.921	2.31

Figure 11 - simulate model of lift rate after the policy takes place

This goal is out of our policy in chapter 5.3.

In our linear regression result, we find that electricity demand grows steadily with time, and renewable resources are mainly used to generate electricity, except for Arizona. So, we think the tendency is to use clean and renewable resources for generating to meet the electricity demand eco-friendlier. (Fuel ethanol is mainly used as the additive of gasoline, more than 95%) We set the goal in the main factor for energy use to encourage the usage of renewable energy usage (geothermal, solar energy and wind energy, specifically). Our focus is, in two words, electricity generating.

unit: %

	CA	**AZ**	**NM**	**TX**
2025	50	24	10	20
2050	64	30	14	38

Figure 12 - Goals of electricity generating in 2025 and 2050

Figure 12 shows the target ratio of renewable energy generation to total electricity generation in four states in 2025 and 2050. Because the prediction ratios in four states are different without the support of new energy compact, and the factors that influence energy profiles are different, so the target ratios of each state are discrepant.

For California, cleaner renewable energy should exceed half of the total energy in 2025 and up to 64% in 2050. And the target ratio in Arizona is 24% and 30%. For the beginning of the implementation of new energy policy in New Mexico, the target ratio in 2025 and 2050 is 10% and 14%. Finally, the target ratio in Texas is 20% and 38%.

Also, we are glad to see if the residential sector can eliminate 50% of the ratio of electricity usage out of all their energy demand (like people in Arizona). This goal is mainly as an encouragement, as the insistent strategy to save energy as much as possible never really stopped the growing demand of electricity.

5.3 Possible Actions

The energy compact is based on state implementation in order to improve energy using situation. The efforts listed below give us a good start to the goal of achieving a clean energy program.

There are three parts of our policy to reach out our goal: providing better tax discounting policy for more stimulation on renewable energy resource generating, compose the residential sector to save more electricity by overrise the price of overuse, and establish brand new solar energy electricity generating station in Arizona.

Our first advice is to give 6% tax discount in the following 8 years for all the firms related to renewable energy resources, which is two times as the discount in 2005. What's more, we are going to give 50% forever tax discount for the electricity generating from renewable and clean energy.

To begin with, projects of hydro-electricity generating and nuclear-electricity generating are only done by the specialist firm, which does not need or able to be encouraged by this action. The first strategy is a maximized version of the similar one in EPACT2005, aimed to stimulate the field with a harder force.

The second one is that we would charge 20% more for net electricity overuse in a family. This strategy is widely used in various resource-saving project around the world, like Beijing tap water.

Finally, we would like to invite firms to tender for a solar-electricity generating station project in Arizona.

In Arizona, actually, we found only a little part of solar energy is used for generating, although Arizona enjoys almost most sunshine in the US. Checked the data, we speculate that Arizona people used to use solar energy in daily life. In this way, much of the solar energy would be wasted, although it does not cost much. Furthermore, Arizona, did not really generated much electricity, compared to the other states among the four.

6 Model Analysis

6.1 Residuals Analysis

In Time Series Model, we change the time dummy variable and analyze the residual of regression[5] and square residual [6]of prediction (SE), the results are as follows:

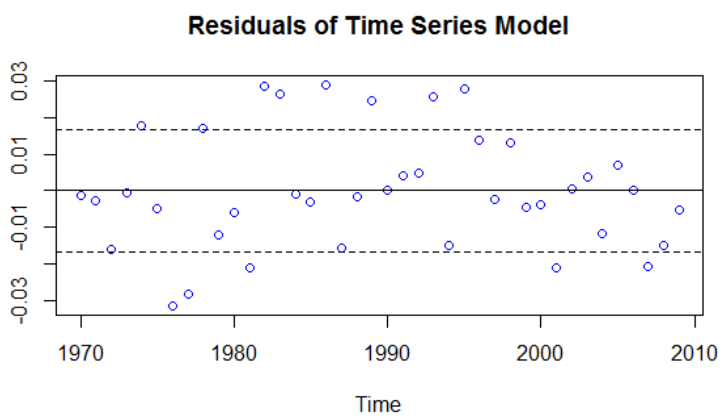


Figure 13 – regression residual of time series model

5 the residual of an observed value is the difference between the observed value and the estimated value of the quantity of interest

6 the sum of the squares of residuals (deviations predicted from actual empirical values of data). It is a measure of the discrepancy between the data and an estimation model. A small RSS indicates a tight fit of the model to the data. It is used as an optimality criterion in parameter selection and model selection.

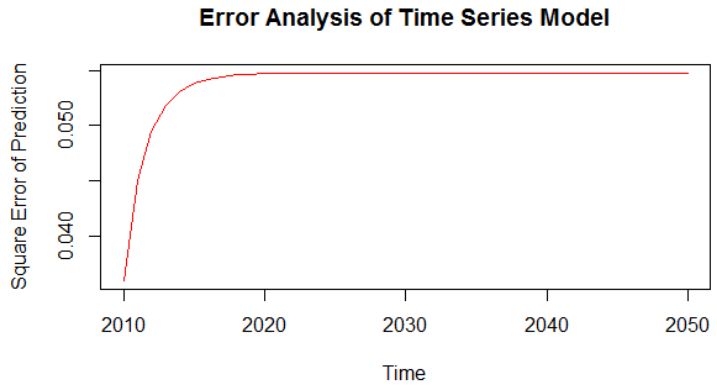


Figure 14 - square residual of prediction values

From the figure, we find that a majority of residuals fall within the mean error (ME) interval, which conform to our expectation. Moreover, the SE of prediction firstly rise to approximately 0.054, then it will not increase any more after 2020.

6.2 Strengths

- Maneuverability: While there are many steps to go from inputs and outputs, each step individually is easy to understand, that is to say, the model at each stage is interpretable and maneuverable.
- Technical Supporting: In the process of constructing the model, we use many statistical theories and methods to support our work. Each one is used reasonably and properly.
- Accuracy: Because we take Time Series and Linear Regression approach, we maintain full accountability for residuals. Therefore, our model has a good fitting of reality.

6.3 Weaknesses

- Simplifying Assumptions: Simplifying assumptions had to be made to create a solvable model. Thus, some valuable data and information are unable to use. The actual situation development is not consistent to the plan, which leads to the error of the result predicted
- Lack of data: Model's outcome greatly depends on given statistics, thus the outcome might not be robust if lots of information is lost. There are too many vacancies (NULL) in the data attached. And we simply abandon these data, which

is a main source of errors.

- Subjective: When we do classification for evaluation metrics, we choose variables by manual filtration instead of algorithm method. So, the accuracy of outputs is limited

6.4 Potential Improvements

We believe there are many improvements which could be made to this model. When we construct the Time Series Model, we use a linear trend and a first order lag to fit the data. That is, we have assumed that the growth rate of energy ratio is fixed, which could be not perfectly fit the reality enough.

Therefore, when fitting the dynamics of energy ratio, the accuracy of the model would potentially increase when using other kinds of trend, such as quadratic trend and exponential trend.

The model would also benefit from additional factors included in energy ratio. In particular, we can add dependent variables in Time Series Model to fit the dynamic relativity between energy ratio and factors (e.g. population and temperature).

7 Conclusion

We have been asked by governors of four states in United States to determine the goals of new energy compact and recommend possible actions to meet the goals.

Since it is a big data problem, we use data classification and data selection as pretreatment. We delete the first ten-year data and some variables which miss almost all the data, both are completed by Python. Then we select main factors of energy profiles by manual sort.

Then we constructed a series of model to fit the historical data. We built a Time Series Model to fit the tendency and cyclical dynamics, and a policy-decision model to determine the target ratios of four states. The results basically conform to our expectation and reality, in other words, each model is used reasonably and properly.

Finally, we offer possible actions for four state governors in order to achieve given goals. The first one is providing tax discount policy which can stimulate the development of renewable energy. Moreover, a policy that charging more for net electricity overuse families could be formulated to support resource-saving project. We also recommend Arizona to establish a new solar-electricity generating station because Arizona enjoys almost most sunshine in US.

Reference

[1] Wang, mou. Analysis on American Clean Energy and Security Act[J]. Advance in Climate Change Research, 2010, (4): 307-312

[2] Sun, Xiaoyu. Analysis of financial data based on linear time series model[J]. Times Fiance, 2016, (5): 264-265

[3] Cui, Cheng. Japan US clean energy agreement and its impact analysis[J]. Overseas Energy, 2011, (4): 31-33

[4] North America Clean Energy[EB/OL]. http://www.nacleanenergy.com/.

[5] Liu, Yiding. China and the United gas competition[N]. China Energy Newspaper, 2013.3.18(14).

[6] Technews. Renewable energy accounted for 33% of the total power. http://guangfu.bjx.com.cn/news/20170522/826706.shtml.

[7] Jin, youde. Nonlinear regression analysis (Common curves and equations) [EB/OL]. https://wenku.baidu.com/view/7c987575e2bd960590c677c9.html?qq-pf-to=pcqq.group.

[8] Thaws. TOPSIS method introduction and program implementation[EB/OL]. http://blog.csdn.net/Thaws/article/details/51955861?locationNum=2.

[9] Zhang, Yue. Evaluation of ideal solution by TOPSIS method[EB/OL]. http://blog.csdn.net/zhangyue_lala/article/details/70149388.

[10] Comparison of Title XIII OF H.R.6, The "Enhanced Energy Infrastructure and Technology Tax Act of 2005, "as passed by the House of Representatives and Title XV OF H.R.6, The "Energy Policy TaxIncentives Act of 2005, "as amended by the Senate[J], Prepared by the Staff of the Joint Committee on Taxation.

[11] Description and Technical Explanation of the Conference Agreement of H.R.6, TITLE XIII, the "Energy Tax Incentives Act of 2005"[J], Prepared bythe Staff of the Joint Committee on Taxation.

[12] NOAA. Climate at a Glance[EB/OL]. https://www.ncdc.noaa.gov/cag/time-series/us.

[13] Windpower. Capricorn Ridge (GE Energy) (USA)[EB/OL]. https://www.thewindpower.net/windfarm_ en_3164_capricorn_ridge_(ge_energy).php.