Team Control Number

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

74030

Problem Chosen

C

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

2018 Mathematical Contest in Modeling (MCM/ICM) Summary Sheet

Our Energy Will Go on

Summary

The production and usage of energy has always been a topic of great concern, especially the development and utilization of renewable energy. In this paper, we try to find a solution to the problem by establishing feasible mathematic models based on the mining results of the energy data of four states. The core steps are as follow:

Firstly, the 583 variables in the energy data provided are classified into seven categories using the K-means clustering method. Seven indicators are then extracted from the seven categories, respectively. The state energy profile is established using the seven indicators as well as one additional indicator featuring at renewable energy.

Secondly, analytic hierarchy process (AHP) is adopted to assess energy profiles about the usages of clean, renewable energy in the four states. The AHP model contains two criteria and sets six kinds of clean and renewable energy, based on which we calculate the weights of the eight indicators and make a ranking list of the four states. Arizona is found to have the "best" profile. The rationale of our ranking list is further confirmed by comparison with those reported in *Renewable Energy data book*.

Thirdly, autoregressive integrated moving average (ARIMA) models are established to describe and predict the development of the energy profile of each of the four states. With the eight indicators we obtained, we establish eight models for each state, resulting in a total of 32 models. Based on the criteria in the AHP model and the predictions made by ARIMA models, we set the renewable energy usage targets for each state in 2025 and 2050. After taking the geographical, climatic and resources features into consideration, we also put forward some suggestions for the four states to meet their energy compact goals.

In the last part of this report, we state the strengths and weaknesses of our work. Data-driven is the most important feature of our modeling procedure. The usage of reliable data mining techniques, including K-means, AHP, and ARIMA, make our models reliable and objective. Nevertheless, some external influence variables might be ignored in the models.

MEMORANDUM

To: Mr. Doug Ducey, Mr. Jerry Brown, Ms. Susana Martinez and Mr. Greg Abbott

From: Team #74030

Subject: Information about interstate energy compact

Date: Feb. 12th, 2018

After analyzing the dataset provided, we are delighted to inform you of our findings. In general, with the largest amount of residual fuel oil consumption, California has the highest energy consumption in the transportation sector among these 4 states from 1960 to 2009, while Arizona and New Mexico have few of them. Also, in Texas, the annual changes of both industrial sector's energy consumption and kerosene consumption fluctuate wildly, with the biggest numerical values compared with the other states. But in Arizona and New Mexico, the situations are contrary. With regard to Arizona, both the ratio of clean and renewable energy consumption to total energy consumption and the ratio of clean and renewable energy's electricity generation to total electricity generation are the largest than those of any other state among which New Mexico and Texas have quite small ratios during this period. Anyway, the annual changes of both the ratio of energy expenditures to GPD and energy average price is very similar in all the states.

As to predictions, we forecast that in Arizona, the ratio of renewable energy consumption to total energy consumption [REC/TEC] will reach at a very low level in 2025 and 2050, but the ratio of renewable energy's electricity generation to total electricity generation [REEG/TEG] will keep a stable growth and reach 0.527 in 2050. In California, the REEG/TEG will be better in 2025 and 2050, but is still at a very low level. Besides, the REEG/TEG is going to be the lowest in four states. In New Mexico, the progress is visible in early 2025 and is very obvious in 2050. The REC/TEC always keeps the highest than those of four states. Besides, the ratio of total energy expenditures to current-dollar gross domestic product [EXP/GDP] has a rapid download tendency and will reach at 7% and 5.4% in 2025 and 2050.

We offer two types of indicators containing four goals for your references. The indicators include the ratio of renewable energy consumption to total energy consumption (REC/TEC) and the ratio of renewable energy's electricity generation to total electricity generation (REEG/TEG). Each year corresponds to two main goals for each state. Specifically, first, as for Arizona, the REC/TEC in 2025 and 2050 should separately reach 0.25 and 0.45, while the REEG/TEG in 2025 and 2050 should separately reach 0.50 and 0.70. Second, as for California, the indexes are 0.20, 0.30, 0.40 and 1, whose sequence follows the order of the former statement. Third, similarly, as for New Mexico, the indexes are 0.20, 0.50, 0.50 and 0.90. Last, the indexes of Texas are 0.16, 0.45, 0.60 and 1.

Please remember that any question is welcome. We are looking forward to receive any good news that you will take the goals we set up into considerations when making the interstate energy compact. At last, good luck to your plans about clean and renewable energy!

Contents

Our Energy Will Go on

1 Introduction

1.1 Statement of the Problem

Since the very past till now, fossil fuels have always been the main energy resources for people. But as the fact that fossil fuels are unrenewable, we have to face the possibility of energy depletion if we continue our pattern of energy-using. So that's why a policy about developing and increasing the usage of renewable energy is very important.

Governments definitely play a crucial role in formulating and executing any policy. In USA, state governments commonly delegate some authority to local units and channel policy decisions down to them for implementation. Besides, two or more states can enter into interstate compacts with one another. In order to create a more energy sustainable development, several states in USA, including California, Arizona, New Mexico, and Texas intend to build up a contractual agreement with a set of goals.

In accordance with these four governors' request, this paper utilizes a typical method to analyze the exact energy profiles, which is beneficial to provide a feasible arrangement of energy-using targets and deal with the tangled and absent time-series data. What's more, the strategy contains the assessment and predictions of their energy profiles.

1.2 Overview of Our Work

First of all, the problem is decomposed a few key points:

- Energy profiles should contain some typical energy-related features, so we need to extract them from the data provided.
- With hundreds of different variables, the amount of data is large. How can we extract key information?
- Also, the data provided include many missing data, which must be considered when we build up model.
- How can we develop a time-related model about energy profile evolvement which can predict the future?
- Renewable energy has many components. But different components reflect different aspects, so how can we judge their importance?

Then, on the basis of above questions, to determine the optimal investment strategy, we can fix them by following several steps:

- First, we utilize the K-means clustering algorithm to characterize these data into different types of features after filling up the missing data and normalizing all the data. Among these features, we select the typical variables as our indicators. With these indicators, we can easily create an energy profile for each state.

- Second, we adopt the analytic hierarchy process (AHP) to choose the state who has the best energy profile.
- Third, we build up ARIMA models to describe and predict the development of the energy profile of each of the four states.

1.3 Assumptions

- Any data provided in excel is correctly reported.
- Clean and renewable energy resources include biomass, wind energy, nuclear power, geothermal energy, hydroelectricity, and photovoltaic and solar thermal energy.
- The influences of other indicators which are not provided in the dataset should be ignored.
- Electricity produced by clean and renewable energy is taken charge by electric power sector.
- All the consumptions of clean and renewable energy focus on the production of electricity.
- All the electricity produced by clean and renewable energy is consumed.
- Annual gross electricity productions equal annual gross electrical system energy losses and annual gross electricity consumptions.

2 Data Processing

2.1 Data Processing

We sort out the data in the worksheet ("seeds") at first and find that it contains 583 variables. Then we created a 583-by-50 data matrix for each of the four states. Rows of the matrix correspond to variables (also known as features), while columns of the matrix correspond to years, from 1960 to 2009. This dataset suffers from missing values because some kinds of energy have not been recorded in 1960~1979. To cope with this problem, first, for each specific variable with missing values, we calculate the mean over each state. The missing values are then replaced by the mean value correspondingly.

After filling all the missing values, the data are standardized to facilitate further data analysis. Since data of four states will be used simultaneously in the following variable classification and modeling process, we merge the four data matrix so as to get a 583-by-200 matrix. Then we compute the z-scores of raw data in the new matrix.

2.2 Feature Clustering Analysis

The 583 variables might contain redundant and/or duplicated information for analyzing the state energy profile. Therefore, we adopt K-means clustering[1] to classify the variables based on their value distributions over the 50 years in the four states. The above procedure, often known as feature clustering, is a typical dimension reduction technique in data mining and can help improve the efficiency and accuracy of modeling.

- K-means Clustering

Given a set of observations

$$(x_1, x_2, \ldots, x_n)$$

where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k (≤ n) sets

$$\boldsymbol{S} = \{S_1, S_2, \ldots, S_k\}$$

so as to minimize the within-cluster sum of distance function of each point in the cluster to K center. That is to say, its objective is to find:

$$\boldsymbol{min} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \sum_{x \in \boldsymbol{Si}} (\boldsymbol{x} - \boldsymbol{\mu_i})^{\boldsymbol{2}}$$

where μ_i is the centroid of observations in S_i.

We take the variables as observations and k as the number of indicators. The bigger k is, the more indicators we get. Considering that it will be unnecessary and meaningless if we divide those variables into too many categories, we set the range of k as [2, 10]. The results of clustering analysis across different settings of k are evaluated using the silhouette score[15], as shown in Figure 1. We find that the silhouette score peaks at k = 7, suggesting that it is suitable to cluster variables into 7 categories.

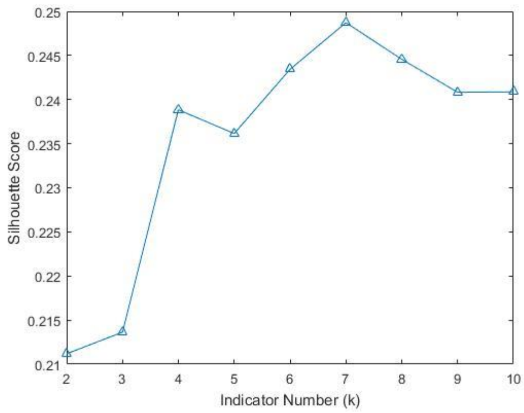


Figure 1. Silhouette Score figure

According to the 'Majority rule', we separately extract 11 variables as indicators to represent these 7 categories. Among these categories, there are 2 special categories. As for each of them, we use 2 variables to represent it. After proportion processing 4 variables into two indicators separately representing these two categories and adding up an extra indicator we are interested with, the whole indicators we use are as below:

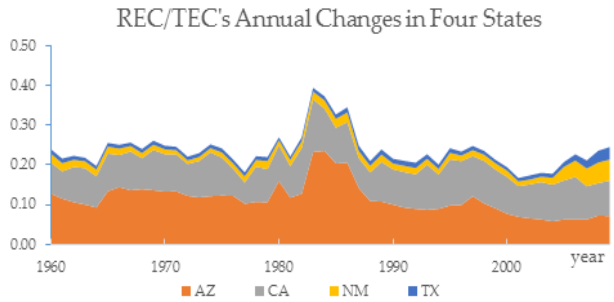
Table 1. The Information of Chosen Indicators.

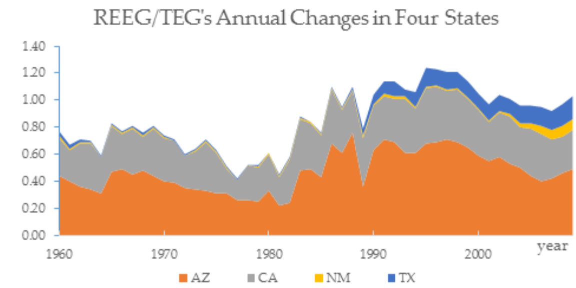
MSM	Description
TEACB	**Total energy consumed by the transportation sector.**
TEICB	**Total energy consumed by the industrial sector.**
TETCD	**Total energy average price.**
KSTCB	**Kerosene total consumption.**
RFTCB	**Residual fuel oil total consumption.**
EXP/GPD	**The ratio of total energy expenditures to current-dollar gross domestic product**
REC/TEC	**The ratio of clean, renewable energy consumption to total energy consumption**
REEG/TEG*	**The ratio of clean, renewable's electricity generation to total electricity generation**

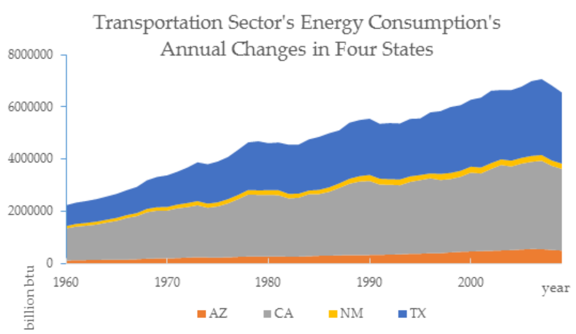
Note: the MSM with "*" is the added one.

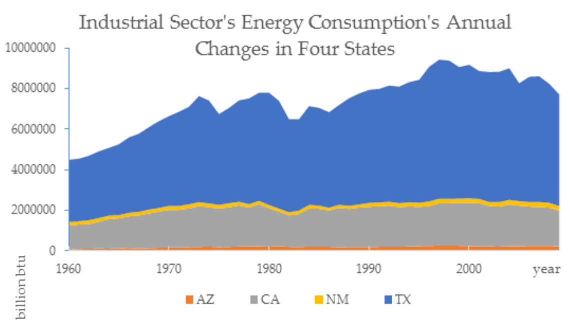
3 Energy Profile

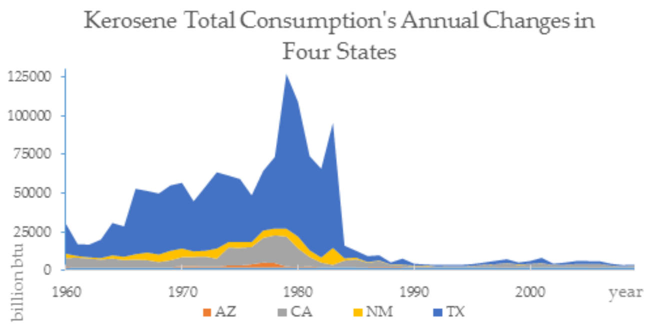
On the basis of above indicators extracted from the dataset, we intend to establish an energy profile for each state. The corresponding annual changes' patterns are shown in figures below. Also, the following energy profiles' descriptions will be based on this information. It is worth noting that AZ, CA NM and TX separately represent Arizona, California, New Mexico, and Texas.

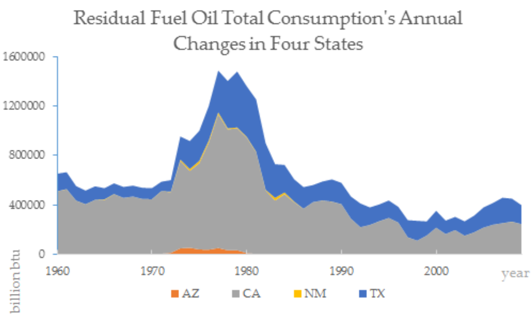


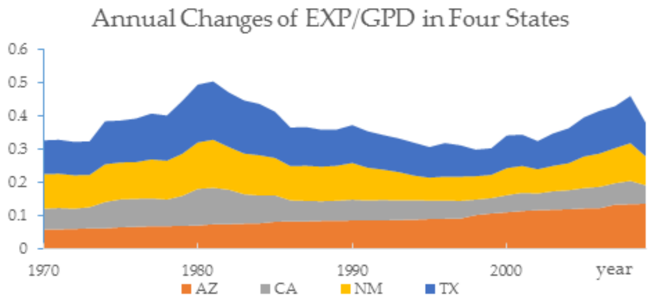


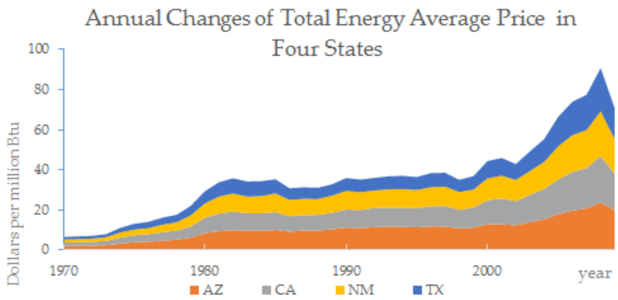












3.1 AZ's Energy Profile

As known for its Grand Canyon and Saguaro desert, Arizona's gross energy consumptions in both the transportation sector and industrial sector increase very slightly with time. What's more, because of the particular geographical and topographic features, this state has few fossil fuel resources. The shortage of fossil fuel resources may explain why the consumptions of some of petroleum products, including kerosene and residual fuel oil, are so low that in some periods, people can even consume none of them. As for the annual change of the ratio of total energy expenditures to current-dollar gross domestic product [EXP/GPD], there are two little peaks in around 1980 and 2007. The gradual increase of energy average price with years may indicate the currency inflation or energy shortage. Also, the distribution curve of the ratio of renewable energy consumption to total energy consumption [REC/TEC] looks like a unimodal curve whose top is in around 1994. However, we can be told from the figures that the ratio of renewable energy's electricity generation to total electricity generation [REEG/TEG] fluctuates wildly in that period, with numerical values ranging from 0.22 to 0.76.

3.2 CA's Energy Profile

As the most populated state in the nation[7]. California has a rich traffic network to meet the needs of public transportation. The data also reflects that California's energy consumption in the transportation sector was rather high in each specific year and became higher and higher with years. However, the energy consumed in the industrial sector remained rather steady in different years, which is lower than that in the transportation sector. The shape of the distribution of kerosene consumption from 1960 to 2009 also looks like a unimodal curve, whose top is in around 1978. The top kerosene consumption is less than 25,000 billion btu. Nevertheless, the residential fuel oil consumption fluctuates a bit during the period, with rather high numerical values in each year. The annual change patterns and the size of each year's numerical value of 'EXP/GPD' and 'energy average price' are very similar to those of Arizona, so they can be referred from the energy profile of Arizona. The annual changes of 'REEG/TEG' and 'REC/TEC' are quite stable with a little increase and decrease. Altogether, the numerical values of them are a bit smaller than those in energy profile of Arizona.

3.3 NM's Energy Profile

Similar to Arizona, New Mexico also has lots of natural landscapes including Rocky Mountains, desert canyons and so on[8]. With the sixth-least population[8], the energy consumptions in both the transportation sector and industrial sector change very slightly and remain a quite small numerical value with years. As a possible result, the consumptions of some of petroleum products, including kerosene and residual fuel oil were generally low. Also, the annual change patterns and the size of each year's numerical value of 'EXP/GPD' and 'energy average price' are very similar to those of Arizona, so they can be referred from the energy profile of Arizona. Both 'REEG/TEG' and 'REC/TEC' increase slightly with years. And basically, the numerical values of them are much smaller than those in energy profiles of the former two states.

3.4 TX's Energy Profile

Among the states, Texas has the second-largest population and the second-largest economy after California[9,10,11]. Like California, Texas's energy consumption in the transportation sector was also rather high in each specific year and became higher and higher with years. The state has many energy-intensive industries and the industrial sector accounts for the largest share of state energy use[12, 13]. That may help to explain why Texas consumed so much energy in the industrial sector from 1960 to 2009. However, the corresponding numerical values still fluctuate mildly. As a part of the energy supplies, kerosene plays a key role in the productions of the industrial sector. So the consumption of kerosene is the highest among all the four states. But the numerical values of different years are of super differences, ranging from 529.34 to 100188.90, among which the bigger numerical values mainly correspond to the earlier years. But the annual change pattern of residual fuel oil remains similar to that of California, with relatively smaller numerical values. What's more, Texas's annual EXP/GPDs, energy average prices and REEG/TEGs are of mild bigger numerical values than those of New Mexico, while REC/TECs are of mild smaller numerical values than those of New Mexico. But the annual change patterns of these variables are very similar to those of New Mexico, so they can be referred from the energy profile of New Mexico.

4 Using AHP to Rank the Profiles

4.1 Establish a hierarchical model

Here, we try to determine which state have the “best” profile for the use of cleaner and renewable energy in 2009. With the combination of above analyses and data provided, we think the criteria for evaluating cleaner, renewable energy profile contain two aspects. One is the ratio of clean and renewable energy consumption to total energy consumption, and the other is the ratio of electric power generation by clean and renewable energy in total electric power generation. The higher the ratio is, the better the profile for use of clean and renewable energy.

We then develop a hierarchy structure considering the two criteria and six kinds of clean and renewable energy, and choose the Analytic Hierarchy Process (AHP)[2] as the way to conform the weighting coefficients of all the indicators in the evaluation system. The hierarchy structure is shown in Figure 2.

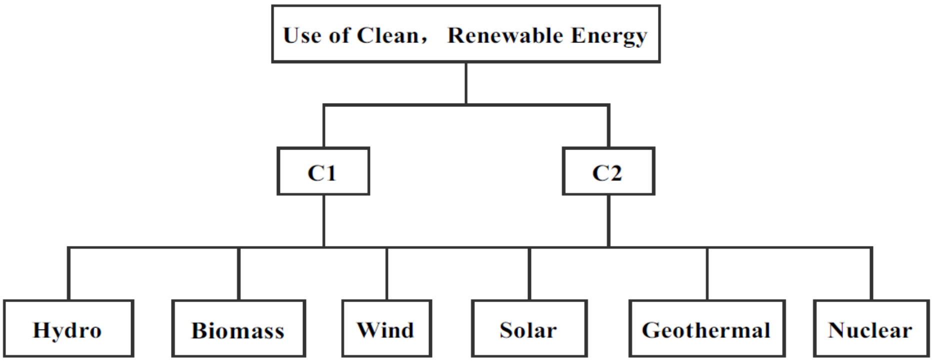


Figure 2. Hierarchy Model

In Figure 2, C1 represents the ratio of clean, renewable energy consumption to total energy consumption. C2 represents the ratio of electric power generation by clean, renewable energy to total electric power generation.

4.2 Results & Analysis

Starting from the second level of the above hierarchy model, we create a judgement matrix by Comparison Method of 1-9. Then we get the weights of C1 and C2 as shown in Table 2.

Table 2. Judgement matrix of hierarchy Ⅰ-Ⅱ

Criteria	C1	C2	Weight
C1	1	2	0.6667
C2	1/2	1	0.3333

By calculating the weights of two factors in hierarchy Ⅱ, the maximum eigenvalue we get is 2, index of consistency is 0 < 0.1.

We create matrices of hierarchy Ⅱ - Ⅲ according to the U.S. energy consumption and electricity generation data in 2009, which can be found on the Internet[3, 4]. Then we get the weights of those six kinds of energy as shown in Table 3 and Table 4.

Table 3. Judgement matrix of hierarchy Ⅱ-Ⅲ between C1 and six energy resources

C1	Hydro	Biomass	Wind	Solar	Geothermal	Nuclear	Weight
Hydro	1	4	5	8	6	1/2	0.3073
Biomass	1/4	1	2	5	3	1/4	0.1202
Wind	1/5	1/2	1	4	3	1/6	0.0850
Solar	1/8	1/5	1/4	1	1/3	1/8	0.0284
Geothermal	1/6	1/3	1/3	3	1	1/7	0.0495
Nuclear	2	4	6	8	7	1	0.4096

After calculating the weights of six energy sources of C1 in hierarchy Ⅲ and further analysis, we obtain two values including the maximum eigenvalue and the consistency ratio. The former one is 6.3217 and the latter one is 0.0519 which is smaller than 0.1.

Table 4. Judgement matrix of hierarchy Ⅱ-Ⅲ between C2 and six energy resources

C2	Hydro	Biomass	Wind	Solar	Geothermal	Nuclear	Weight
Hydro	1	7	4	8	7	1/2	0.3182
Biomass	1/7	1	1/2	3	1	1/7	0.0577
Wind	1/4	2	1	5	3	1/5	0.1116
Solar	1/8	1/3	1/5	1	1/3	1/8	0.0291
Geothermal	1/7	1	1/3	3	1	1/8	0.0535
Nuclear	2	7	5	8	8	1	0.4298

Similarly, after calculating the weights of six energy resources of C2 in hierarchy Ⅲ and further analysis, the maximum eigenvalue we get is 6.2572 and the consistency ratio we obtain is 0.0415 which is also smaller than 0.1.

From above tables, we know that those judgement matrices pass all consistency tests (consistency ratio < 0.1), which means that the weights determined by AHP are very reasonable. Then we use weight coefficients shown above to evaluate four states' clean and renewable energy profiles in 2009.

We set:

- the ratio of any type of clean and renewable energy consumption to total energy consumption of a state in 2009 as r_{ij}^{C1},
- the ratio of electric power generation by clean and renewable energy to total electric

power generation of a state in 2009 as r_{ij}^{C2},

- the weight coefficient between C1 and one of six energy sources as w_i^{C1},
- the weight coefficient between C2 and one of six energy sources as w_i^{C2},
- the weight coefficient of C1 as c_1, and the weight coefficient of C2 as c_2.

We use the weighted average method[14] to calculate the score of each state's profile for use of clean and renewable energy in 2009. The expression is shown as follow :

$$s_j = \sum_{i=1}^{n} r_{ij}^{C1} * w_i^{C1}, * c_1 + \sum_{i=1}^{n} r_{ij}^{C2} * w_i^{C2} * c_2$$

where j = 1, 2, 3, 4, refers to the four states; n = 1, 2, …, 6, represents six kinds of clean renewable energy.

We obtain the four states' scores and rank them in descending order. The result is shown in Table 4.

Table 4. Ranking of states

Rank	State
1	Arizona
2	California
3	Texas
4	New Mexico

From Table 4, we can see that Arizona have the 'best' profile for use of clean and renewable energy in 2009. This result is a little bit different from the report in *2009 Renewable Energy data book*[5], whose ranking is California, Texas, Arizona, and New Mexico.

Since we take nuclear energy, which is a kind of clean energy but not a renewable energy, into consideration and think it more highly than any other energy resources, it is no wonder that nuclear power has a larger weight. Thus, a state with a higher ratio of electric power generation by nuclear power to total electric power generation will be more likely to have the 'best' profile for use of cleaner, renewable energy. To test our criteria, we delete nuclear power from our model and run the program again. This time we get the same ranking as reported in 2009[5], so the criteria we set is feasible.

Finally,after considering the strength and scale of use of nuclear power, we believe that the weight on nuclear power is reasonable, so is our result.

5 ARIAM Model Construction

5.1 Concept of ARIAM Model

The ARIMA model (Autoregressive Integrated Moving Average Model) is widely used for predictions in the fields of energy, climate and so on[6]. The basic idea of the ARIMA model is that a sequence formed by the changes of the projective objects over time will be considered as a random sequence. Based on the autocorrelation analysis of time series we can approximate the sequence by applying a mathematical model.

The ARIMA model is the combination of differential methods and ARMA model (Auto-Regressive and Moving Average Model). Where the ARMA model use the following objective function:

$$\Phi(B)\, x_t = \psi(B) a_t$$

where:

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p$$
$$\psi(B) = 1 - \psi_1 B - \psi_2 B^2 - \cdots - \psi_q B^q$$

This model can be transformed as:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots + \psi_q a_{t-q}$$

Where a_t is the t-th random error, Φ_i and ψ_j is coefficients. This model is marked as ARMA (p, q) model. The parameters p and q in the model can be determined by observing the autocorrelation and partial correlation of time series. The autocorrelation correlation takes the general form as:

$$\overline{\rho_k} = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

And the partial correlation:

$$\Phi_{kk} = \frac{D_k}{D}$$

Where D is the matrix of autocorrelation coefficients and D_k is the k-order autoregressive fitting.

Besides, because only a smooth time series can be directly used to establish the ARIMA (p, d, q) model, we first using a differential method to get a smooth time series and then establish it. It can be expressed as:

$$\Phi(B)\,(1-B)^d\, x_t = \psi(B) a_t$$

Where d is the order of differences that we can achieve a smooth time series.

In this paper we use the ARIMA model to describe and predict the time series of energy development.

5.2 Example of AIMAM Model

The procedures of analyzing all models are not fully reported here due to the page constraint. We take the ratio of renewable energy consumption to total energy consumption in Arizona as an example to introduce the detailed process of ARIMA model analysis.

First of all, the autocorrelation (fig 3. left) and partial autocorrelations (fig 3. right) are all tailed, indicating that this time series is non-stationary. So we can do further analyses by a differential method.

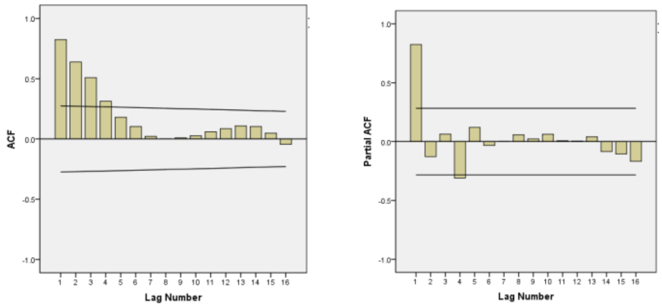


Figure 3.

After the second-order difference, the numbers in the differential sequence is basically evenly distributed on both the upper and lower sides of the zero (fig 4.), so the sequence can be considered as stable in a way. The numerical value of coefficient d is 2.

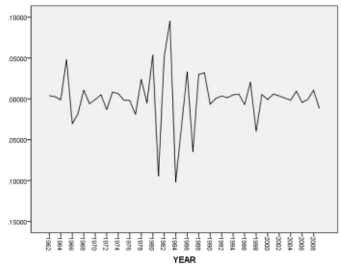


Figure 4.

Then look at the autocorrelation and partial autocorrelations graphs (fig 5.) for the differential sequence. The autocorrelation graph shows that coefficient p can be 1 and the partial autocorrelation graph shows that coefficient q can be 1 or 2.

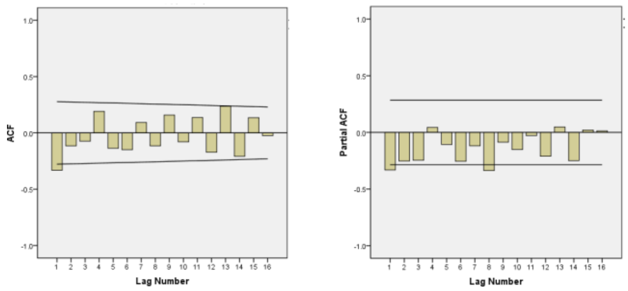


Figure 5.

After comparing the model of ARIMA(1,2,1) with ARIMA(1,2,2) model, the Akaike information criterion(AIC) index shows that the ARIMA(1,2,1) model fits the data better. This model explains the 44 percent variance of data (fig 6.).

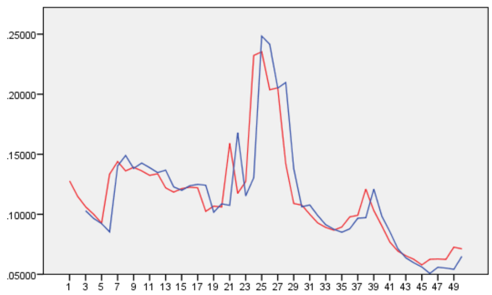


Figure 6. Model-Fit Graph

Note: the red line is observed data and the blue line is fitted data.

Besides, the autocorrelation and partial autocorrelations graphs of residuals (fig 7.) show the residual sequence is stable, therefore the ARIMA(1,2,1) model is feasible。

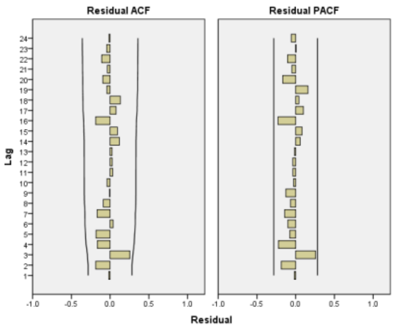


Figure 7.

Parameter estimation specified form of this model:

$$(1+0.034B)\ (1-B)^2\ x_t = (1+B)\ a_t$$

Here the numerical value of d indicates that it is not a stable time sequence. After combining the function and the fitted graph, it can be figured out that the model has a slow downward trend.

Finally, using the ARIMA(1,2,1) model we can predict the changes of the ratio in the next 41 years and estimate the 95% confidence interval. The prediction curve (fig 8.) shows a steady and slow downward trend.

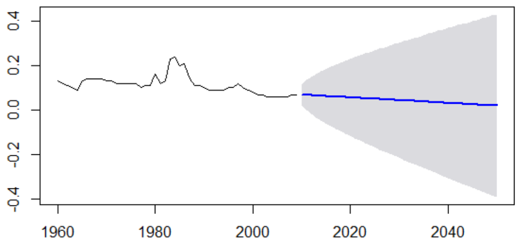


Figure 8.

Note: black line is the time sequence from 1960(/1970) to 2009; blue line is the predictive numerical values; grey area is the confidence interval of the predictive numerical valus.

As all of the other models use the same analytical procedures as above, here we only report the results.

5.3 Results and Interpretations

When it comes to the clean and renewable energy, we consider two indicators: the ratio of renewable energy consumption to total energy consumption (REC/TEC), the ratio of renewable energy's electricity generation to total electricity generation (REEG/TEG). We can see from the figure that the developments of clean and renewable energy in Texas and New Mexico are very similar, especially the REC/TEC (fig 9., left).

Taking account of the parameter estimation results, we find that although the REC/TEC in both states is low (about 0.02), it has been developing rapidly in recent years.

That's to say both the trends of New Mexico and Texas are positive, while the former one is more positive than the latter one. On the contrary, the other two states have negative trends. This difference may be due to the different geographical environment and location.

As for the ratio of renewable energy's electricity generation to total electricity generation (fig 9., right), we find some differences in all the four states. Texas and New Mexico still have lower baselines and faster development potentials than the other ones.

The pattern of change of the REEG/TEG in California is very volatile, but it still shows the similar growth tendency as Arizona's. But it is noteworthy that Arizona has shown a clear downward trend in the past decade. It may need some measures for Arizona's government to reverse the negative trend.

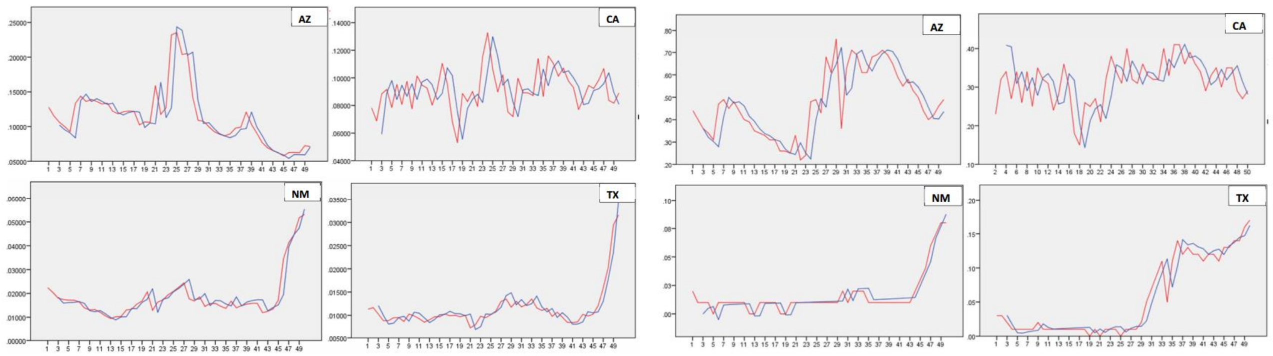


Figure 9. Model-Fitted Graphs of REC/TEC and REEG/TEG

Note: the red line is observed data and the blue line represents fit data.

As for the non-renewable energy resources, we take the consumption of kerosene and residual fuel oil as examples and illustrate their changes.

We discovered an interesting phenomenon in the consumption of kerosene (fig 10., left): although every distribution curve of kerosene's consumption in each state has a peak in the evolvement from 1960 to 2009, California's and Arizona's curves peak before 1980s whereas the peaks of Texas's and New Mexico's curves appear after 1980s. 1980s is a period with so many possible incidents when the second industrial revolution happened. Then the changes have remained stable in recent years.

As for residual fuel oil (fig 10., right), the situation is different. The peaks' distributions of the right part of figure 10 are very similar to the corresponding ones of the left part. But the changing pattern is different. The consumption of residual fuel oil in Arizona and New Mexico reached a level that is close to zero, whereas the figures of Texas and California shows similar changing patterns which maintain at a rather high level. This phenomenon deserves the attention of the local governments.

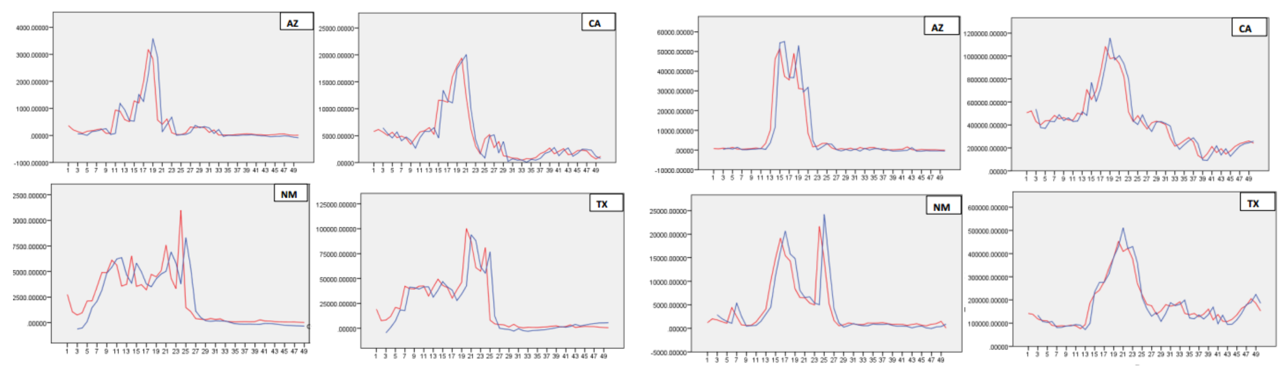


Figure 10. Model Fit Graph of the Consumption of Kerosene and Residual Fuel Oil

About the situations of energy consumptions in the transportation and industrial sectors, we find a nearly consistent increase in energy consumption in the transportation sector (fig 8., left) among the four states. Here Arizona has not only the highest consumption but also the most rapid energy-consuming growth in the transportation sector. Only New Mexico fluctuated a little bit wildly and started a continual decrease in the last decade.

But the situation is very different in the industrial sector (fig 11., right). Although all the sub-figures show growing tendencies in the patterns of these 4 states, Arizona has the lowest consumption, while Texas stands out flocks with the most rapid growth. Generally speaking, California has developed stably in the past decade, but New Mexico is on the contrary.

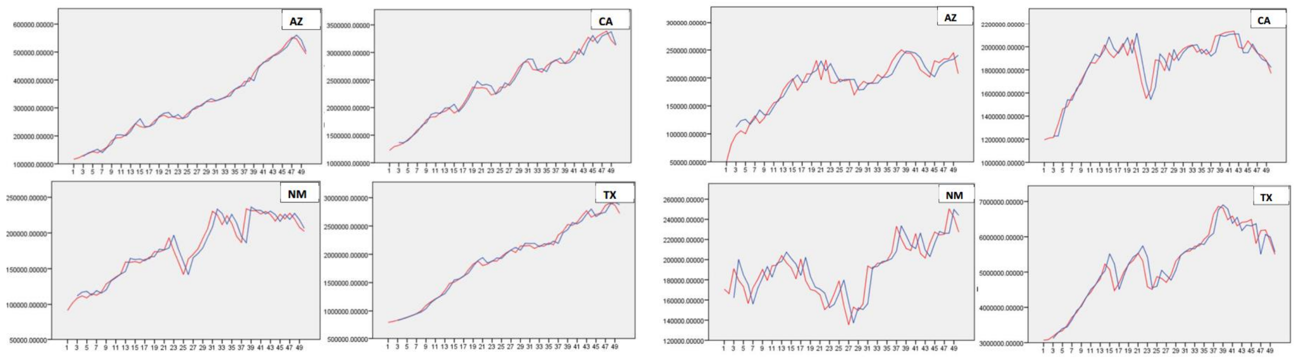


Figure 11. Model Fit Graph of Energy Consumptions in the Transport and Industrial Sectors

Considering the ratio of total energy expenditures to current-dollar gross domestic product (EXP/GDP) and energy average price, we found some similarities and differences that deserve attentions for four states' governors.

First of all, as the left 4 parts of figure 12. shows, the changing trends of EXP/GDP share some similarities in Texas and New Mexico. The EXP/GDP in these two states have become higher and higher for nearly a decade. But the trends in Arizona and California are quite stable.

Secondly, although all states show similar changing patterns in energy average price (fig 12., right), they still have clearly rankings in the average price: Arizona > California > New Mexico > Texas. It was also tested by one-sample ANOVA (F(3,196)=415.142, p=0). This can be possibly related to the degree of modernization in four states.

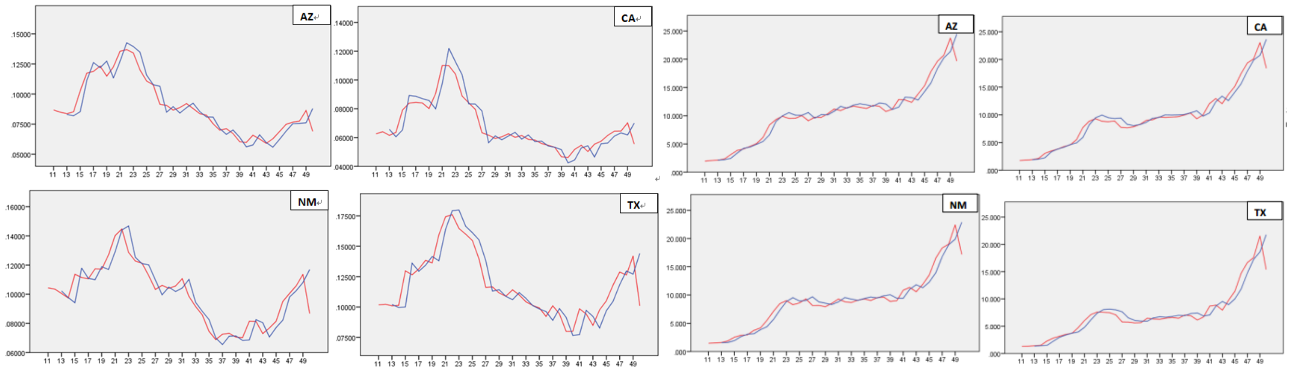


Figure 12. Model-Fitted Graphs of the EXP/GDP and energy average price

Generally speaking, from the results, we can see that there are many similarities and differences about the evolvement of energy profiles among these four states. We hope the results we obtain can make any contribution to formulate the newly interstate energy compact among these four states.

6 Predictions and Goals

6.1 Predictions

Using ARIMA model to predict the development of the next 41 years, we get the following results:

For REC/TEC (fig 13., left) and REEG/TEG (fig 13. right), we can find some similarities in the prediction of two indicators. Both REC/TEC and REEG/TEG seem to increase at a rapid pace in New Mexico and Texas, but remains stable in Arizona and California. The REC/TEC in Arizona even have a downward trend, which needs to be noticed by the governors.

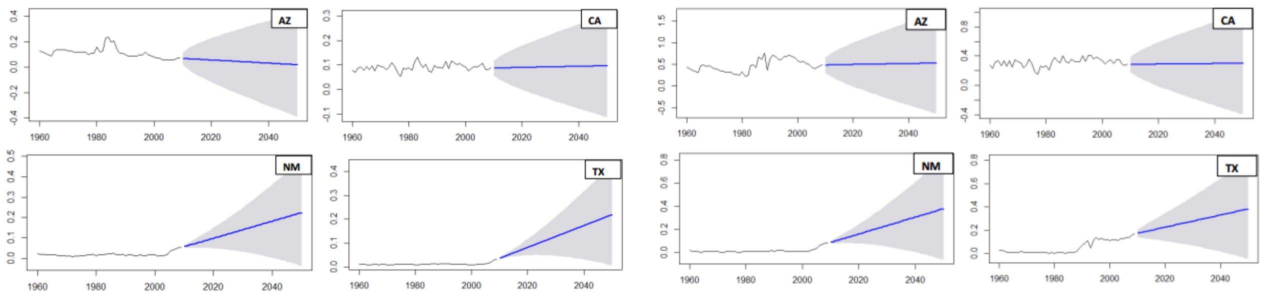


Figure 13 Predictive Graphs of REC/TEC and REEG/TEG

Note: black line is the time sequence from 1960 to 2009; blue line represents the predicted numerical values from 2010 to 2050; grey area shows the confidence interval of predicted numerical value.

From left part of Figure 11, we can see that the kerosene total consumption will remain stable in nearly all states has reached a very low level (close to zero), while residual fuel oil total consumption stays low only in Arizona and New Mexico (fig 11. right). Besides, the downward tendency of RFTCB in Texas is faster than that in California.

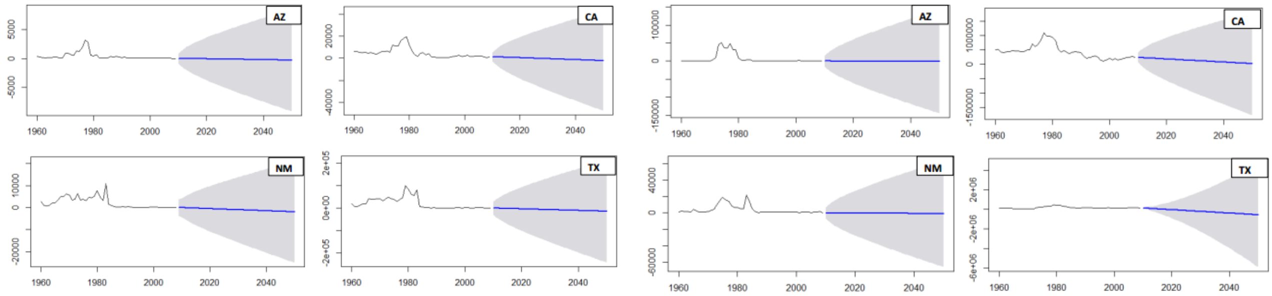


Figure 14 Predictive Fitted Graphs of Kerosene's and Residual Fuel Oil's Consumptions

As for the total energy consumed in the transportation sector (fig 12., left), we find that it has the tendency to increase fastest in all states. However, the total energy consumed in industrial sector is going to grow slowing in New Mexico and Texas and even has a slight downward trend in California and Texas.

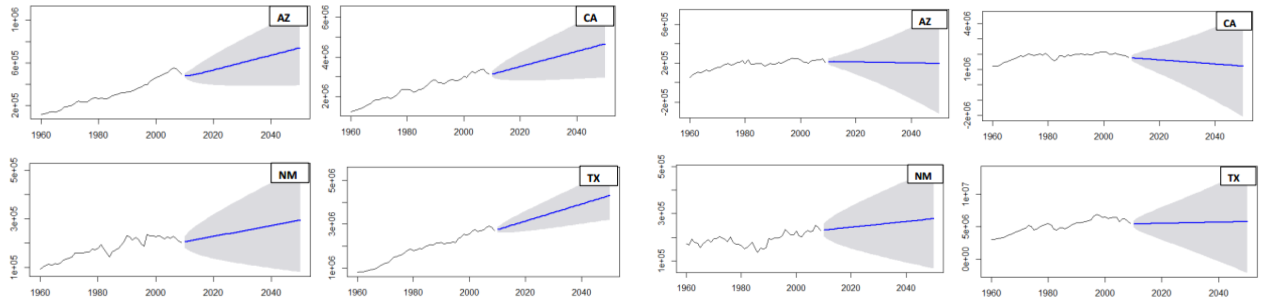


Figure 12 Predictive Graphs of Energy Consumptions in the Transportation and Industrial Sectors

In addition, the ratio of total energy expenditures to current-dollar gross domestic product in all states will decrease slowly in the next 41 years according to our prediction.

Based on the predicted numerical values (Table 5.), we can predict that in Arizona, the REC/TEC will reach a very low level in 2025 and 2050, but the REEG/TEG keeps a stable growth and attains 52.7% in 2050. The consumption of non-renewable energy, such as kerosene and residual fuel oil, will remains low as in 2009. And the average price of energy in Arizona seems to be still the highest among four states.

In California, the REC/TEC will be higher in future, but is still at a very low level. Moreover, the REEG/TEG in California is going to be the lowest among all states, while its consumption of residual fuel oil is the highest. These prediction results merit the attention of the governors in California.

We expect to see significant improvement on the use of clean, renewable energy in New Mexico. The REC/TEC will maintain the highest rank among four states. Besides, the EXP/GDP keeps a rapid downward trend and will decrease to 7% in 2025 and 5.4% in 2050.

For Texas, whose development of energy profile is similar to New Mexico, the usage of renewable energy will expand in the following years. The REEG/TEG in Texas will rise to

38.3% in 2050. The average energy price will be the lowest among all states as before, but the EXP/GDP will stay in a high level without a significant drop.

Table 5. Forecasting Numerical values

State	Year	REC/TEC	REEG/TEG	KSTCB	RFTCB	TEACB	TEICB	EXP/GPD	TETCD
AZ	2025	0.050	0.499	0	0	567684	208993	0.050	26.9
	2050	0.019	0.527	0	0	740148	196340	0.033	38.2
CA	2025	0.091	0.291	0	153854	3706182	1546720	0.043	25.5
	2050	0.097	0.300	0	20476	4663912	1213627	0.034	36.3
NM	2025	0.120	0.197	0	0	238221	248201	0.070	24.5
	2050	0.224	0.379	0	0	294790	278217	0.054	34.9
TX	2025	0.105	0.253	0	0	3339869	5572220	0.095	22.2
	2050	0.219	0.383	0	0	4317028	5794551	0.092	31.8

6.2 Setting Renewable Energy Usage Targets

In view of the criteria presented in AHP model and information provided by the data, we think two of the most important indictors about renewable energy usage profile are:

- The ratio of renewable energy consumption to total energy consumption
- The ratio of renewable energy's generation to total electricity generation

So, the core goal of states' energy usage is to increase these two ratios. We then set targets to be reached according to our prediction model. For the purposes of urging states to promote renewable energy use, we set target ratios based on the upper numerical values of the prediction intervals. The goals we set are shown in Table 6. The targets to meet seem beyond the capacity of states now. However, we believe governments have the potentials to achieve them, if some effective actions are taken.

Table 6. Renewable energy usage targets for 2025 and 2050

Goals	Arizona	California	New Mexico	Texas
1.1 The ratio of renewable energy consumption to total energy consumption to be reached in 2025	25%	20%	20%	16%
1.2 The ratio of renewable energy consumption to total energy consumption to be reached in 2050	45%	30%	50%	45%
2.1 The ratio of renewable energy's generation to total electricity generation to be reached in 2025	50%	40%	50%	60%
2.2 The ratio of renewable energy's generation to total electricity generation to be reached in 2050	70%	100%	90%	100%

6.3 Suggestions for Energy Compact Goals

Due to the results of our analysis, we suggest that these four states might take the following actions to realize energy compact goals:

✧ Investment makes improvement. Increase the budgets that invest in renewable energy generation and high - technology industries. This action will attract lots of potential investors and companies, creating more environmental friendly workplaces.

- ✧ Learning quickens development. As the fact that almost all of renewable energy generating equipment is set in remote areas, governments should figure out how to decrease the losses of energy in energy transmission. The energy losses are related to two factors, including distances and conductive materials. Since we hardly shorten the distances based on the geographic restrictions, we need to optimize the conductive materials. Optimizing them is not a short-term thing, it'd be better to develop advanced technology or buy the best conductive materials from other countries after considering the time restriction.
- ✧ Education is the foundation. Besides directly increasing the budgets of renewable energy related industries, governments should also increase the financial supports to universities and research institutions to develop new technologies that can improve the efficiency of renewable energy transformation and transmission. That's because without any technological researches, the governments won't get the latest and the most advanced techniques at first, which may bring about a slower development than some country.
- ✧ Cooperation realizes fourfold wins. Located in the southwest of the United States, all of these four states have abundant wind, solar and geothermal energy potentials. But since the different population densities and economic power of these states, the energy needs and renewable energy developments are different among them. For example, California which has the most well-developed economy and largest population among these 4 states needs the largest amount of potential energy to supply the whole state. Also, because of the abundant financial supports, its renewable-energy related techniques may be the most advanced among all the four states. That's to say, California can be the main technical base area which provides well-designed technical guidance for other three states to tap their renewable energy. As a return, the other three states should share some of the newly-extracted renewable energy to California which is also beneficial to their renewable energy consumption patterns.

7 Conclusions

To help the governors of California, Arizona, New Mexico, and Texas form the energy compact, we develop an AHP evaluation system and ARIMA models to describe the energy profile and predict the development of energy use of the four states.

We use many theory and methods such as K-means clustering, AHP, ARIMA to complete our work, combine many possible influential factors and examine our results with those in other official reports. We believe that the work we have presented here is a significant and successful attempt to assist governors formulating the compact. But like any model, no model is correct, only useful model, our model also has its strengths and weaknesses. They are presented below:

7.1 Strengths

☺ Rather than choosing indicators subjectively, we cluster the 583 variables into 7 categories by using K-means clustering, which facilitate the following data analysis and make our energy profile more robust and reliable.

☺ By establishing an AHP model, we get the weight coefficients of all indicators of clean, renewable energy usage, making our analysis more objective.

☺ To further examine our analysis, we compare our ranking results of the energy profile with those in the reports from *2009 Renewable Energy data book*, which confirms the reliability and accuracy of the energy profile.

☺ Compared to less fitting results that are common in regression, ARIMA model makes the best use of the autoregressive characteristics of time series. Thus, it can represent the development of data more appropriately, and the prediction is more accurate.

☺ We give full consideration to the effect of the geographical, climatic and resource factors on the promotion of renewable energy use in the four states.

7.2 Weaknesses

☹ There are some vacancies in the data provided. And we fill the blanks with the mean value of each variable for each state, which might have an impact on the variables clustering results.

☹ Although we refer to the actual data of U.S. energy usage when we assign values to the judgment matrix in the AHP model, the influence of subjectivity exists inevitably.

☹ Since the ARIMA model needs to process the time series smoothly, the forecasting tendency is relatively stable. Besides, as the interval between prediction time and time in data grows longer, the standard error of prediction is getting bigger, so is the confidence interval.

Reference

[1]https://en.wikipedia.org/wiki/K-means clustering

[2] http://en.wikipedia.org/wiki/Analytic_hierarchy_process

[3]http://energyfuture.wikidot.com/us-energy-use

[4]http://rodhandeland.com/Energy%20Food%20Water%20Technology/Energy%20Challenges.htm

[5]Gelman, R. *2009 Renewable Energy Data Book, August 2010*. United States. doi:10.2172/989021.

[6]Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). Arima models to predict next-day electricity prices. *IEEE Transactions on Power Systems, 18*(3), 1014-1020.

[7]U.S. EIA, State Energy Consumption Estimates, 1960 through 2015, DOE/EIA-0214(2015) (June 2017), Table C10, Energy Consumption by End-Use Sector, Ranked by State, 2015.

[8]U.S. Census Bureau, Geography, State Area Measurements and Internal Point Coordinates, accessed December 20, 2017.

[9]U.S. Census Bureau, State Population Totals Tables: 2010-2016, Population Estimates, Population Change, and Components of Change, Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2016 (NST-EST2016-01).

[10]U.S. EIA, Texas, Profile Data, Energy Indicators, accessed December 4, 2017.

[11]U.S. EIA, California, Profile Data, Energy Indicators, accessed December 4, 2017.

[12] U.S. EIA, Energy Explained, Energy Use in Industry, accessed December 5, 2017.

[13]U.S. Bureau of Economic Analysis, Regional Data, GDP and Personal Income, Annual Gross Domestic Product (GDP) by State, GDP in current dollars, All industries, Texas, 2015.

[14]https://en.wikipedia.org/wiki/Weighted_arithmetic_mean

[15]https://en.wikipedia.org/wiki/Silhouette_(clustering)