Background：

图片

On 2010, U.S. State Clean Energy Data Book reported that "In the United States, renewable energy has been capturing a growing percent of new capacity additions during the past few years. In 2009, renewable energy accounted for more than 55% of all new electrical capacity installations in the United States—a large contrast from 2004 when all renewable energy captured only 2% of new capacity additions. "

With increasing usage of cleaner, renewable energy, four state governors – California, Arizona, New Mexico , and Texas – wish to form a realistic new energy compact. Texas leads the country in total (non-hydro) installed renewable energy capacity, almost all of which comes from the state's 9,410 MW of wind capacity. California is the leader in solar energy installed capacity, both for photovoltaic technology (738 MW) and concentrating solar power (364 MW).

Restatement of problem:

We are required to provide 4 energy profiles of states above and model how these profiles have evolved from 1960 to 2009. Besides that, making our results understandable with discussion is necessary, including similarities and differences between 4 states. Then, we need to establish a criteria model and choose the "best" profile use of cleaner, renewable energy in 2009. Based on the results , we are asked for prediction of the development trend of energy profile for 2025 and 2050 without any policy changes by each governor's office. In addition, we should consider characteristic of each state, our prediction and criteria to set renewable energy usage targets. Finally, we need to give our suggestions about exact actions taken to achieve the goal we predict.

Related work:

energy profile is a general

cleaner, renewable energy (defination from SEDS and):
conventional hydroelectric power,
solar thermal direct use energy and photovoltaic electricity net generation,
electricity produced by wind,
wood and wood-derived fuels, biomass waste,
fuel ethanol minus denaturant.

Notation:

Abbreviation&Description

n — the number of samples
p — the number of original indicators
k — the number of principal component
X = {xij} — original data matrix (i = 1, 2,…, n ; j = 1,2,…,p)
Z = {zij} — standardization matrix with z-scores (i = 1, 2,…, n ; j = 1,2,…,p)
U = {uij} — partial covariance matrix (i = 1, 2,…, n ; j = 1,2,…,p)
C = {cjj'} — covariance matrix (j = 1,2,…,p)
R = {rjj'} — correlation matrix (j = 1,2,…,p)
F ={fjj'} — The eigenvectors matrix of Z (j = 1,2,…,p)
λ_j — The eigenvalues of Z (j=1,2,…,p)

yi — Principal component
kmo — the value of KMO test

Assumptions and Justifications:

1.忽略除煤电外发电引起的外部成本。
the mean damage is 3.2 cents per kWh
2005年美国由于燃煤排放的外部成本为3.2美分/kwh，天然气发电的外部成本为0.16美分/kwh，由数据分析知实际上几乎全部是由煤电引起，Life-cycle CO2 emissions from nuclear, wind, biomass, and solar appear so small as to be negligible compared with those from fossil fuels.

—HIDDEN COSTS OF ENERGY UNPRICED CONSEQUENCES OF ENERGY PRODUCTION AND USE

2.发电量和消耗量之间的转化率为100%，利用消耗量来表示发电量从而评估
energy conservation law

Model overview:

In order to solve those problems, we will proceed as follows:

流程框图

Firstly, considering energy production and consumption, environmental impact, technology, economy, we extract 12 variables from 605 variables to get general energy profile of each state.
In order to show the evolution of energy production, we construct a comprehensive evaluation index system for the level of cleaner, renewable energy development. We use the **principal component analysis method** to carry out the correlation cluster analysis of the index, To make the similarities and differences between 4 states understandable, we use **TOPSIS** method to show our results. Then, we use **AHP** to assign the weight of the extracted principal components choose the "best" profile from comprehensive evaluation results ranking.
After determining the "best" profile, we use **ARIMA method** to predict the energy profile of each state, particularly, we evaluate the cleaner, renewable energy development in the future. Based on our results,

Model theory:

. ***Principal component analysis theory***

The purpose of PCA is to use the idea of dimensionality reduction to convert multiple indicators into a few composite indicators. When using statistical methods to study multivariable problems, too many variables will increase the computational complexity and increase the complexity of analyzing the problems. We hope that in the process of quantitative analysis, fewer comprehensive variables are involved to make the results more Intuitive.

The method can be summarized as follows :

(1) Organize the data set and normalization

Supposing that there are n samples, and each sample has P indicators, the original variable data X are arranged as a set of n data vectors with each representing a single grouped observation of the p variables, which as shown in formula(1)

$$X = \begin{vmatrix} x_1^T \\ x_2^T \\ \cdots \\ x_n^T \end{vmatrix} = \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{vmatrix} \qquad (1)$$

We choose zero-mean normalization to standardize the original data matrix.The standard score of a raw score xij is shown as formula(2)

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \qquad (2)$$

where:
μj is the mean of the row j .
σj is the standard deviation of the row j.
Then we get the standardization matrix Z as shown in formula(3).

$$Z = \begin{vmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{vmatrix} \tag{3}$$

(2) Determine the correlation between the indicators

Correlation test is a statistical test of how the variables are related and how relevant they are. Thus, it's also the basis of principal component analysis. The Kaiser-Meyer-Olkin (KMO) test method is used in this model. Correlation coefficient is a statistical indicator of the close relationship between the response variables, whose range of values between 1 to -1 .

The formula(4) for KMO test is :

$$kmo_j = \frac{\sum_{i\neq j} r_{ij}^2}{\sum_{i\neq j} r_{ij}^2 + \sum_{i\neq j} u_{ij}^2} \tag{4}$$

where :
//R = [rij] is the correlation matrix.
U = [uij] is the partial covariance matrix.

And for reference, Kaiser put the following values on the results :

value range	result
0.00 to 0.49	unacceptable
0.60 to 0.69	miserable
0.70 to 0.79	mediocre
0.80 to 0.89	meritorious
0.90 to 1.00	marvelous

(3) Calculate the number of principal component

The correlation coefficient matrix of standardization matrix is calculated by formula (5).

$$C = \frac{Z^T Z}{n-1} \tag{5}$$

Note:The reasoning behind using N − 1 instead of N to calculate the covariance is Bessel's correction.
The eigenvalues of G is calculated by formula (6)

$$|Z - \lambda E| = 0 \tag{6}$$

where E is unit matrix.

According formula [7], arrange the eigenvalues in descending order as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$.
The cumulative contribution rate of top k factors is shown in formula (7).

$$\alpha_{(k)} = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{p} \lambda_j} \qquad (7)$$

The goal is to choose a value of k as small as possible while achieving a reasonably high value of g on a percentage basis. Generally, the value of $\alpha_{(k)}$ should be more than 80%.

(4) Determine the scores of principal components

The eigenvectors corresponding to eigenvalues are calculated by formula (8).

$$(Z - \lambda_j E)F_j = 0 \qquad (8)$$

where F_j is the eigenvectors, which is also the component consisting of $x_1, x_2, \ldots, x_p$.

The scores of principal components are the product of the eigenvectors of correlation coefficient matrix C and the standardization matrix Z .The calculate method is shown as formula (10) .

$$y_m = F_m \times B(m = 1,2,...,k) \qquad (10)$$