Prediction and Analysis of Carbon Emissions of Chinese Provinces

Qing-yu ZHANG¹, Kang WANG¹, Chun-ling ZHANG¹ and Cui LI¹

¹Research Institute of Business Analytics & Supply Chain Management, College of Management, Shenzhen University, Shenzhen, 518060, China.

*Corresponding author: q.yu.zhang@gmail.com

Keywords: Carbon Emission, K-means Clustering, Logistic Regression Analysis

Abstract. The paper makes forecasts of the carbon emissions of Chinese provinces to provide references for the government's emission reduction policy. The methods to analyze carbon emissions include K-means clustering and logistic regression analysis. The data about carbon emissions for each province in China during 1995~2011 were collected from the statistical database of China's economic and social development (CNKI). The results show that the forecast value of carbon emissions in 28 provinces and regions in China in 2011~2020 is 488,232~6,618,514 million tons. The conclusion of the paper is that the accuracy of Logistic model is influenced by the value of K. As long as the value of K is determined in a reasonable range, it can give a relatively high fitting degree and low prediction error.

Introduction

The population of China as a developing country increases very quickly in recent years. The economic growth and the quality of life are also increasing but at the same time the consumption of household appliances, cars, and personal digital products are increasing, inevitably causing higher carbon dioxide emissions. Energy saving and emission reduction has become the consensus of the world, and as a major power, China plan to reduce carbon emissions per unit of gross domestic product (GDP) by 2020 to 40%~45% of that in 2005[1]. During the 12th Five-Year planning period, China accepted the increase in emissions, but its growth rate should be reduced. The National Development and Reform Commission predicted that China's greenhouse gas emissions will reach its peak in 2025, five years earlier compared with the previous forecast. To reduce carbon dioxide emissions, it is difficult to rely on the excellent emission reduction technology only, and there should be effective and rational policy as guidance. To make better policies, we should accurately predict the carbon emissions first. In fact, the local government plays a vital role in reducing carbon emissions. Therefore, through the analysis of carbon emissions in each province can we effectively predict the carbon emissions in China.
Literature Review

Due to the increasing attention to the greenhouse effects, therefore, there have been a lot of researches on carbon emissions [3, 4, 5]. Li [3] studied the energy consumption based system, analyzing the history of CO2 emissions in China, and predicting the future Chinese CO2 emissions. The results showed that, by 2030, emissions of CO2 in China will be 12.4 billion tons, per capita emissions of about 8.5t, CO2 emission intensity is 256t / million dollars. Based on the public data, Qu [7] analyzed carbon emissions in 30 provinces and cities in China in recent years. The STIRPAT model is used to predict that China will usher in carbon emissions peak in 2020-2045, and he proposed to increase the emission reduction technology investment and promote the use of clean energy, an important means to complete the carbon emission reduction targets in China the future. Nie [8] undertook a structural decomposition analysis to decompose the changes in CO2 emissions from 1997 to 2010 into six driving forces, and the results showed that declines in energy intensity had a decrease impact on CO2 emissions during the studied period. Yue [9] made an analysis of carbon emissions, per capital carbon emissions and carbon emission intensity in 1995 - 2007 for the three regions of East, middle and West. Based on the previous researches, this study selects the Logistic function which can describe the growth curve of S, and constructs the Logistic prediction model of carbon emissions growth. Then the data of each province are analyzed.

Method

Data Sources, Carbon Emissions and Method of Calculation

The carbon emission data selected in this study were collected from the statistical database of China's economic and social development (CNKI). This paper sets up 1995 as the starting point of time. In the history of national development, the time from 1995 to 2010 is the industrial development period of a large-scale consumption of fossil energy and is also the most important cumulative period of the fossil fuel carbon emissions. In this study, the historical cumulative emissions and per capita cumulative emissions of 28 provinces and autonomous regions of China were calculated. And calculated formulas are as follows:

Historical cumulative emissions by province:

\[ C_K = \sum_{1995}^{2010} C_i \]  

(1)

Emissions from various provinces over the years:

\[ C_K = (N_{K_i} / N_i) \times C_i \]  

(2)

Per capita cumulative carbon emissions:
\[
RC_i = \sum_{i=1995}^{2010} \frac{C_{ki}}{P_{ki}}
\]  

(3)

C presents carbon emission; N presents energy consumption; K presents provinces; P presents population at the end of the year; i presents particular year.

### 4.2 K-Means Cluster Analysis

In accordance with the historical cumulative emissions and per capita cumulative emissions, we use K-means clustering analysis to classify those provinces and autonomous regions. The steps of K-means clustering analysis as follow:

1. To determine how many categories will be divided, namely K. In this study, it is divided into 4 categories, K=4.
2. To determine the initial cluster centers of 4 categories.
3. According to the 4 initial cluster centers, the Euclidean distance of each sample to the 4 cluster centers will be calculated.

Euclidean distance formula:

\[
d_{kQ} = \sum_{\lambda=1}^{n} (h_{\lambda} - Q_{\lambda})^2
\]  

(4)

In the above formula: h and Q present samples; \(\lambda\) presents sample dimensions. According to the principle of the closest distance, all samples are divided into 4 categories.

4. According to the 4 categories, the average of each variable in each category is calculated, and the mean points are regarded as the new centers of the 4 categories. According to the new center location, the distance between each sample to the new center is recalculated and reclassified.

5. Repeat step (4), until the conditions for terminating the cluster are met. Set the termination clustering condition: ①The iteration number is 10 times; ②The maximum offset of the center point of the new clustering center and the last iteration is less than 0.02.

The figure 1 are showed that the first category includes Shandong, Hebei; The second category includes Anhui, Fujian, Guangxi, Heilongjiang, Hubei, Hunan, Jilin, Jiangxi, Shandong, Shanxi, Sichuan, Xinjiang, Yunnan, Zhejiang; The third category includes Guangdong, Henan, Jiangsu, Inner Mongolia; the fourth category includes Beijing, Gansu, Hainan, Ningxia, Qinghai, Tianjin. In order to test whether the above 4 categories are reasonable, we ran the variance analysis and the results are shown in Table 1. There are significant differences between two different clustering methods.

Table 1. Variance Analysis and Inspection of Per Capita Emissions.
Logistic Regression Model and Regression Analysis

The logistic regression model can describe the bounded growth of S-shaped growth curve well, and it has the simplicity of mathematical calculation and obvious economic meaning[10]. Köne and Aylin[11] argued that the figure of China’s CO2 emission from fossil energy combustion is S-shaped. Hence, China is selected as a sample to confirm the suitability of the logistic equation. The logistic regression model for the growth of carbon emissions was established as equation 5.

\[
\frac{dx}{dt} = rx(1 - \frac{x}{k})
\]

(5)

Where x is the amount of increased carbon emissions, k is the maximum capacity of carbon emissions, r is an indefinite constant value, t represents the year. The initial condition is \(X_0, t = 0\). From the equation (5), the following equation is obtained.

\[
x = \frac{k}{1 + \left(\frac{k}{X_0} - 1\right)e^{-rt}}
\]

(6)
Set up \[ Y = In \frac{k-1}{x_0}, \quad k_{\infty} - 1 = e^a \], we can obtain equation (7) from equation (6).

\[ Y = a - rt \]  

where \( a \) is a constant value.

**Regression Analysis**

This paper took Qinghai Province as an example and used nonlinear regression method. We determined the approximate value of the future range of \( K \) was from 6000 to 8000, and the results of regression analysis were obtained in Table 2.

As can be seen from Table 2, when \( K \) was from 6400 to 7000, the value of goodness of fit was the highest. This paper took \( K=7000 \), the \( R^2=0.984 \), then we could obtain the equation

\[ Y=0.215t+1.012 \]  

Substituting the time series into equation (9), we could obtain the predicted value of carbon emissions of Qinghai Province from 2011 to 2020. Based on principles of the above, we could obtain the predicted value of carbon emissions of other regions of China from the year 2011 to 2020. The results were showed in Table 3.

<table>
<thead>
<tr>
<th>K</th>
<th>6200</th>
<th>6400</th>
<th>6600</th>
<th>6800</th>
<th>7000</th>
<th>7200</th>
<th>7400</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^2 )</td>
<td>0.982</td>
<td>0.984</td>
<td>0.984</td>
<td>0.984</td>
<td>0.984</td>
<td>0.983</td>
<td>0.983</td>
</tr>
<tr>
<td>( r )</td>
<td>-0.258</td>
<td>0.244</td>
<td>-0.233</td>
<td>-0.223</td>
<td>-0.215</td>
<td>0.208</td>
<td>0.202</td>
</tr>
<tr>
<td>( a )</td>
<td>0.918</td>
<td>0.939</td>
<td>0.961</td>
<td>0.987</td>
<td>1.012</td>
<td>1.041</td>
<td>1.069</td>
</tr>
</tbody>
</table>

**Model Validation and Results Discussion**

In order to validate the model's fitting effect, we collected the carbon emission data from year 2011 to 2014, and compared this real data with the forecasting data of logistic regression. From the results, we can find that there are only the relative error of Ningxia Autonomous Region beyond 20%, the relative error between forecasting results of carbon emissions of other province and the true value are all below 10%. For the 28 provinces of China, the average relative error of carbon emissions from 2011 to 2014 were 6.94%, 6.18%, 5.77%, 4.70%, all lower 7%. There is to say, the logistic model can highly fit the carbon emission trends of various provinces and regions of China, and the performance of this Logistic model is very good.

Using the logistic regression model of carbon emissions growth, we obtained the predictive value of carbon emissions of regions of China from year 2011 to 2020. And then we summed up the predicted value of carbon emissions in China, the results were showed in table 3. The results indicated that the value of carbon emissions
continuously increases from year 2011 to 2020. In 2020, the total value of carbon emissions reached 940,239 thousand tons, the annual average growth rate was 9.99%.

The results were compared to the results by Energy Research Institute under National Development and Reform Commission Jiang[12] and results by Yue[9]. The forecast value of carbon emissions for 2020 is 399,590 thousand tons, which is 540,649 thousand tons lower than the result of this paper. The national assessment report predicts China's carbon emissions in 2050 will reach 2Pg, the average lever is 1.5t per person[13]. But the statistical results by the Carbon Dioxide Information Analysis Center of United States Oak Ridge National Laboratory[14] showed that China's carbon emissions had reached 2.25Pg in 2010. Therefore, the prediction by China's national assessment report was significantly lower.

Table 3. Total Value of Carbon Emission in the Regions of China from the Year 2011 to 2020 (10,000t).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>911,262</td>
<td>917,734</td>
<td>922,988</td>
<td>927,247</td>
<td>930,697</td>
<td>933,489</td>
<td>935,747</td>
<td>937,573</td>
<td>939,048</td>
<td>940,239</td>
</tr>
</tbody>
</table>

**Conclusion**

The accuracy of Logistic model is greatly affected by the value of K. As long as the value of K is determined in a reasonable range, it can give a relatively high fitting degree, and low prediction error. With the development of economy and the progress of energy saving and emission reduction technology, China's future carbon emissions will be similar to the "S" shaped growth of biological population. Based on the actual situation of China in recent years, the predicted values from the national assessment report[14] and the research of Yue[9] are lower than China's actual carbon emissions. Therefore, the results of this study are closer to the reality of China.

**Acknowledgments**

This work is partially supported by National Natural Science Foundation of China under Grant No. 71572115; Major Program of Social Science Foundation of Guangdong under Grant No. 2016WZDXM005; National Science Foundation of SZU under Grant No. 836.

**References**


