CEEP-BIT WORKING PAPER SERIES

CEEP-BIT

Exploring the regional characteristics of inter-provincial CO_2 emissions in China: An improved fuzzy clustering analysis based on particle swarm optimization

Shiwei Yu
Yi-Ming Wei
Jing-Li Fan
Xian Zhang
Ke Wang

Working Paper 22
http://ceep.bit.edu.cn/english/publications/wp/index.htm

Center for Energy and Environmental Policy Research
Beijing Institute of Technology
No.5 Zhongguancun South Street, Haidian District
Beijing 100081
November 2011

This paper can be cited as: *Yu S, Wei Y-M, Fan J-L, Zhang X, Wang K. 2011. Exploring the regional characteristics of inter-provincial CO_2 emissions in China: An improved fuzzy clustering analysis based on particle swarm optimization. CEEP-BIT Working Paper.*

This study is supported by the National Natural Science Foundation of China under Grant Nos. 71103016 and 71020107026. And the Humanities and Social Sciences Fund, Ministry of Education of China under Grant No. 10YJC630356. The CAS Strategic Priority Research Program Grant No. XDA05150600. The views expressed herein are those of the authors and do not necessarily reflect the views of the Center for Energy and Environmental Policy Research.

The Center for Energy and Environmental Policy Research, Beijing Institute of Technology (CEEP-BIT), was established in 2009. CEEP-BIT conducts researches on energy economics, climate policy and environmental management to provide scientific basis for public and private decisions in strategy planning and management. CEEP-BIT serves as the platform for the international exchange in the area of energy and environmental policy.

Currently, CEEP-BIT Ranks 121, top10% institutions in the field of Energy Economics at IDEAS（http://ideas.repec.org/top/top.ene.htm), and Ranks 157, top10% institutions in the field of Environmental Economics at IDEAS (http://ideas.repec.org/ top/top.env.html).

Yi-Ming Wei
Director of Center for Energy and Environmental Policy Research, Beijing Institute of Technology

For more information, please contact the office:

Address:
Director of Center for Energy and Environmental Policy Research
Beijing Institute of Technology
No.5 Zhongguancun South Street
Haidian District, Beijing 100081, P.R. China

Access:
Tel: +86-10-6891-8551
Fax: +86-10-6891-8651
Email: ceeper@vip.163.com
Website: http://ceep.bit.edu.cn/english/index.htm

Exploring the Regional Characteristics of Inter-provincial CO_2 Emissions in China: An Improved Fuzzy Clustering Analysis Based on Particle Swarm Optimization

Shiwei Yu[1,2,3*] Yi-Ming Wei[2,3] Jingli Fan[2,3] Xian Zhang[2,3] Ke Wang[2,3]

(1. School of Economics and Management, China university of geosciences, Wuhan, 430074

2. Center for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing 100181

3. School of Management and Economics, Beijing Institute of Technology, Beijing 100181)

Abstract: The better to explore the regional characteristics of inter-provincial CO_2 emissions and the rational distribution of the reduction of emission intensity reduction in China, this paper proposes an improved PSO-FCM clustering algorithm. This method can obtain the optimal cluster number and membership grade values by utilizing the global capacity of Particle Swarm Optimization（PSO）on Fuzzy C-means（FCM）. The clustering results of CO_2 emissions indicate that the 30 provinces of China are divided into five clusters and each has its own significant characteristics. Compared with other clustering methods, the results of PSO-FCM are more explanatory. The most important indicators affecting regional emission characteristics are CO_2 emission intensity and per capita emissions, whereas CO_2 emission per unit of energy is not obvious in clustering. Furthermore, some policy recommendations on setting emission reduction targets according to the emission characteristics of different clusters are made.

Keywords: fuzzy C-means cluster; carbon emission; characteristics; mitigation policy

1. Introduction

Faced with the challenges of climate change, more and more countries are making unremitting efforts to mitigate global greenhouse gas (GHG) emissions [1, 2]. To demonstrate its commitment to reducing global emission, late in 2009, the Chinese government committed to cutting its carbon dioxide emission per unit of gross domestic product (GDP) by 40-45% of the 2005 levels by 2020, which had been declared in the Twelfth Five Year Plan of China's Economic Development to be a restrictive target. Although the national intensity reduction target is clearly set, how this target should be distributed among the various sectors, provinces and regions has not yet been determined. China covers a vast territory, but its energy resources are unevenly distributed. The considerable differences of socio-economic history and conditions among provinces have not only led to uneven levels of economic development, but also to regional differences in carbon emissions. The principle of "common but differentiated" is a reasonable base for the allocation of national intensity reduction obligations among provinces. On one hand, the inter-provincial differences in carbon emissions should be full considered. On other hand, provinces with the same carbon emission characteristics may be set a more uniform emission reduction obligation. Therefore, to provide a rational basis for allocation basis, all provinces should be divided into appropriate categories according to their

* Corresponding author. Tel: +86 010 68918551; fax: +86 010 68918551

E-mail address: ysw81993@sina.com (S.YU)

emission characteristics. On this basis, the present study attempts to answer the following questions: How many categories of carbon emission characteristics can be determined? What are the common characteristics of each category? What is the reason for these characteristics? How can the regional emission reduction targets be set according to these characteristics?

The studies on carbon emission characteristics at provincial level in China in existing literature can be divided into four categories: (1) studies on the differences in carbon emissions from the perspective of the regional spatial geography. Based on China's eastern, central, and western regional divisions in the Seventh Five-Year Plan of China's Economic Development ,scholars examined inter-provincial in equality in CO_2 emission within China using usual measures of inequality (coefficient of variation, Gini index, Theil index) to analyze provincial-level inequality [3-4]. (2) Discussions of the factors affecting carbon emissions using the *Kaya* equation (including the *IPAT* [5] and *STIRPAT* [6] models) and *LMDI* (Logarithmic Mean Division Index) decomposition analysis model [7-8]. (3) Descriptions of inter-provincial carbon emission characteristics based on statistical methods using indicators such as total CO_2 emission, emission intensity, per capita emissions and so forth [9-11]. (4)Analyses of the relationship between carbon emission and industries or energy [12]. Moreover, Zhang et al. argued that a drastic structural change in energy use in China made it difficult for any region to reduce CO_2 emissions in the short-term [13]. Furthermore, Wang et al. estimated the environmental efficiency, economic efficiency, economic environmental efficiency and two-stage efficiency of various provinces in China by considering carbon dioxide (CO_2) emissions [14].

It can therefore be concluded that previous studies on China's regional carbon emissions have mainly been divided into three regions, namely eastern, central and western, which is the traditional division by economic development level. However, there are some differences in the definitions of eastern, central and western, according to researchers' requirements. It is therefore worth examining how closely the characteristics of the carbon emissions of China's regions correspond to the traditional division by economic development level. Hence, a new approach to studying the issue of China's regional clustering of carbon emission is required.

Clustering analysis is an effective technique to divide a given set of data or objects into groups so that elements drawn from the same group are as similar to each other as possible, while those assigned to different groups are dissimilar. Thus, clustering analysis appears to be an appropriate choice for categorizing provinces with similar carbon emission characteristics and to facilitate the exploration of the specific characteristics of each category and its causes easily. There have been a few previous related studies. Yue and Zhu, using a statistical clustering method, divided the 30 provinces of China into four categories [15]. However, their statistical clustering, which included hierarchical clustering and systemic clustering, was based on crisp set, which divides the samples into one or the other category with a clear boundary. It does not reflect the intermediary of objective reality and is a hard partition. Conversely, the Fuzzy C-Means (FCM) algorithm, proposed based on fuzzy set theory, has been more widely used [16-19]. FCM detects clusters that have centroid prototypes of a roughly similar size and describes the intermediary through the use of membership grade values [20]. Concerning carbon emission, Zhang et al., based on a *Kaya* equation, categorized China's inter-provincial carbon emissions in terms of four regions [21]. However, their study applies a traditional fuzzy clustering method, which is simply a steepest descent algorithm with variable step length and often becomes stuck at suboptimal solutions based on the initial configuration [22]. In addition, the convergence and clustering results are easily influenced by initial value. This shortcoming is more prominent where there are a large number of clusters [23]. Therefore, to reduce the sensitivity of classification results, researchers have proposed the utilization of global intelligent optimization, for example, Genetic Algorithm (GA) [24-26] or Particle Swarm Optimization (PSO) [27-29] to replace the steepest descent optimization. The basic concept behind integrating intelligent optimization with FCM is the optimization of appropriate cluster centers at a previously-specified number of clusters. In most of the real-life situations the number of cluster in a data set is not known a priori. Furthermore, the object function or cluster validity index mainly applies to the traditional compactness measure function, which focuses only on the Euclidean distance-weighted sum between cluster centers and samples.

From the above analysis, it can be concluded that there is a need for further study of the following: (1) as there are few descriptive indicators of the characteristics of carbon emission in the existing literature, more indicators reflecting the structure of emission should be considered, such as the proportion of industry emission and carbon emission per unit of energy. These will assist in the investigation of inter-provincial emission characteristics and the reasons for them. (2) In terms of carbon emission, the regions of China have traditionally been divided according to economic development level into eastern, central and western. Moreover, there are some differences in the definitions of the three regions definition according to researchers' requirements. Therefore, further research on how to divide the regions of carbon emission is called for. (3) Cluster analysis on carbon emissions has mainly been carried out by traditional K-means clustering or FCM clustering. These methods require further improvement. The existing improved FCM based on GA or PSO focuses on optimizing the cluster center vectors at a given number of clusters and does not consider the separation of the categories; hence, they are still unable to obtain the optimal number of clusters. At the same time, studies on carbon emission clustering analysis mainly utilize the traditional K-means or fuzzy clustering. Thus the methods require further improvement.

Therefore, the better to explore the regional characteristics of China's inter-provincial carbon emission, this paper proposes an improved PSO-FCM clustering algorithm. Its main feature is the addition of the distances between centers to the objection function and the optimization of the number of clusters by using discrete PSO. In order to divide the 30 provinces of China, have selected five indicators of carbon emission, namely, total CO_2 emission; per capita CO_2 emissions; emission intensity of CO_2; CO_2 emission per unit of energy, and; the proportion of industrial CO_2 emissons. The optimal cluster number and membership grade values can be obtained through the use of the proposed PSO-FCM. In addition, the characteristics and causes of each category have been thoroughly investigated. Furthermore, some policy recommendations on setting emission reduction targets have been proposed based on the emission characteristics of different clusters.

2. Methodology and data

2.1 The calculation of China's CO2 emission

According to IPCC (2007) Fourth Assessment Report, the main source of greenhouse gases is the combustion of fossil fuels (in 2004, around 95% of the world's carbon emissions was ascribed to fossil fuel combustion) [30]. We calculate the CO_2 emission of 30 provinces from fossil energy consumption data.

$$Tco_{2ij} = \sum_{i=1}^{n} \left\{ (A_{ij} - S_{ij}) e_i c_i \times 10^{-3} \right\} \cdot O_i \times \frac{44}{12} \qquad (1)$$

Where A_{ij} is the j th province's real consumption of the i th fuel. The sum of the energy consumption conservation sectors (the heating supply industry and the thermal power industry) and the final consumption sectors (farming, forestry and animal husbandry; fisheries & water conservancy; industry; construction; transport, storage and post; the wholesale and retail trades; hotels and restaurants; residential consumption, and; other); S_{ij} is the j th province's non- combustion consumption of i th fuel. e_i is the net calorific value of the i th fuel; c_i is the carbon content of i th fuel; O_i is the carbon oxidation rate of i th fuel. The coefficients of various types of energy can be found in reference [30].

2.2 PSO-FCM cluster method

FCM can be summarized as follows. Given $X = \{X_1, X_2, \cdots, X_N\} \subset R^n$, where R^n denotes a vector space with real n dimensions. For $\forall k,\ 1 \le k \le N$, $X_k = (x_{k1}, x_{k2}, \cdots, x_{kn})^T \in R^n$, where $x_{kj}(j = 1,2,\cdots,n)$ is the j th membership grade of sample $X_k(k = 1,2,\cdots,N)$. If $Z^T = (Z_1, Z_2, \cdots, Z_c)$ $(Z_i \in R^n, i = 1,2,\cdots,c)$ is the central vector of clustering, the fuzzy c-clustering for X can be expressed by the following mathematical programming:

$$\min\ J_m(U,Z,c) = \sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^h d_{ik}^2 = \sum_{k=1}^{N}\sum_{i=1}^{C}\mu_{ik}^h \|X_k - Z_i\|^2, \quad 1 \le h \le \infty \tag{1}$$

$$\text{s.t}\quad \begin{cases} \sum_{i=1}^{c}\mu_{ik} = 1,\ 1 \le k \le N \\ 0 \le \mu_{ik} \le 1,\ 1 \le i \le c;\ 1 \le k \le N \\ 0 < \sum_{k=1}^{N}\mu_{ik} < N \quad 1 \le i \le c. \end{cases} \tag{2}$$

where $d_{jk}^2 = \|X_k - V_j\|^2 = (X_k - V_j)^T (X_k - V_j)$.

The so-called Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA) proposed by Bezdek is used to conduct the iterative operation, after initialization of matrix $U^0 = ({\mu^0}_{ij})_{c\times N}$, for

$$Z_i = \frac{\sum_{k=1}^{N}(\mu_{ik})^h X_k}{\sum_{k=1}^{N}(\mu_{ik})^h} \quad i = 1,2,\cdots,c \tag{3}$$

and

$$\mu_{ik} = \left(\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{h-1}}\right)^{-1} \tag{4}$$

It can be demonstrated that the operation is convergent in the case of $h > 1$. Where $\max(\|\mu_{ik}(t+1) - \mu_{ik}(t)\|) < \varepsilon$. [31]. $\mu_{ik}(t)$ is a membership grade belonging to the k th category of the t iteration of the i th sample, ε is prior given threshold of convergence, h is the weight, usually set to 2. The traditional FCM algorithm has not yet discussed the optimal number of clusters, but only given a general range, namely $2 \le c \le \sqrt{N}$ for selecting the clusters. To solve this problem, Zhu et al revised the

Eq. (1) into Eq. (5) [32].

$$\min V_{zs} = \sum_{j=1}^{c}\sum_{i=1}^{n}\mu_{ij}^{m}d_{ij} + \frac{c}{\sum_{i,j=1}^{c}\left\|Z_i - Z_j\right\|^2} \tag{5}$$

Where $\left\|Z_i - Z_j\right\|^2$ is the distance between the i th centers and the j th centers. The Eq.(5) not only considers the compactness of samples but also measures the separation of different categories. However, the basic GA is applied to optimize the number of clusters in the paper and the drawbacks of GA, such as prematurity and redundancy, have been clearly recognized [33]. Thus, to improve the efficiency of FCM, the present paper utilizes a discrete particle swarm optimization (DPSO) to optimize c.

As a new global intelligent optimal algorithm, PSO mainly evolves via the comparison of the self position, and the surrounding positions [34]. Compared to GA, PSO has a much more profound and intelligent background and can be performed more easily without selection, crossover and mutation, as in GA [19, 35, 36].

This paper proposes an improved PSO-FCM clustering algorithm. The proposed algorithm flow is shown in Fig.1.In Fig.1, n is the population size of particles; d is the length of each particle, ε is the threshold of convergence, and $\max\_gen$ is the max iterations of algorithms. In the velocity updating equation, $v_{id}^{k+1} = \chi \cdot \left[v_{id}^{k} + c_1 \times rand_1 \times (P_{id}^{k} - x_{id}^{k}) + c_2 \times rand_2 \times (P_{gd}^{k} - x_{id}^{k})\right]$, where χ is a restriction factor. c_1 and c_2 denote the cognitive and the social parameters, respectively. $rand_1$ and $rand_2$ are random numbers distributed uniformly in the interval $[0,1]$. $\chi = \frac{2}{\left|2-\varphi-\sqrt{\varphi^2-4\varphi}\right|}$, For $\varphi > 4$, where $\varphi = c_1 + c_2$, usually $c_1 = c_2 = 2.05$. P_i^k represents the best previous position particle i has obtained until iteration k; P_g^t represents the best position obtained from P_i^k in the swarm or local neighborhood at iteration k.

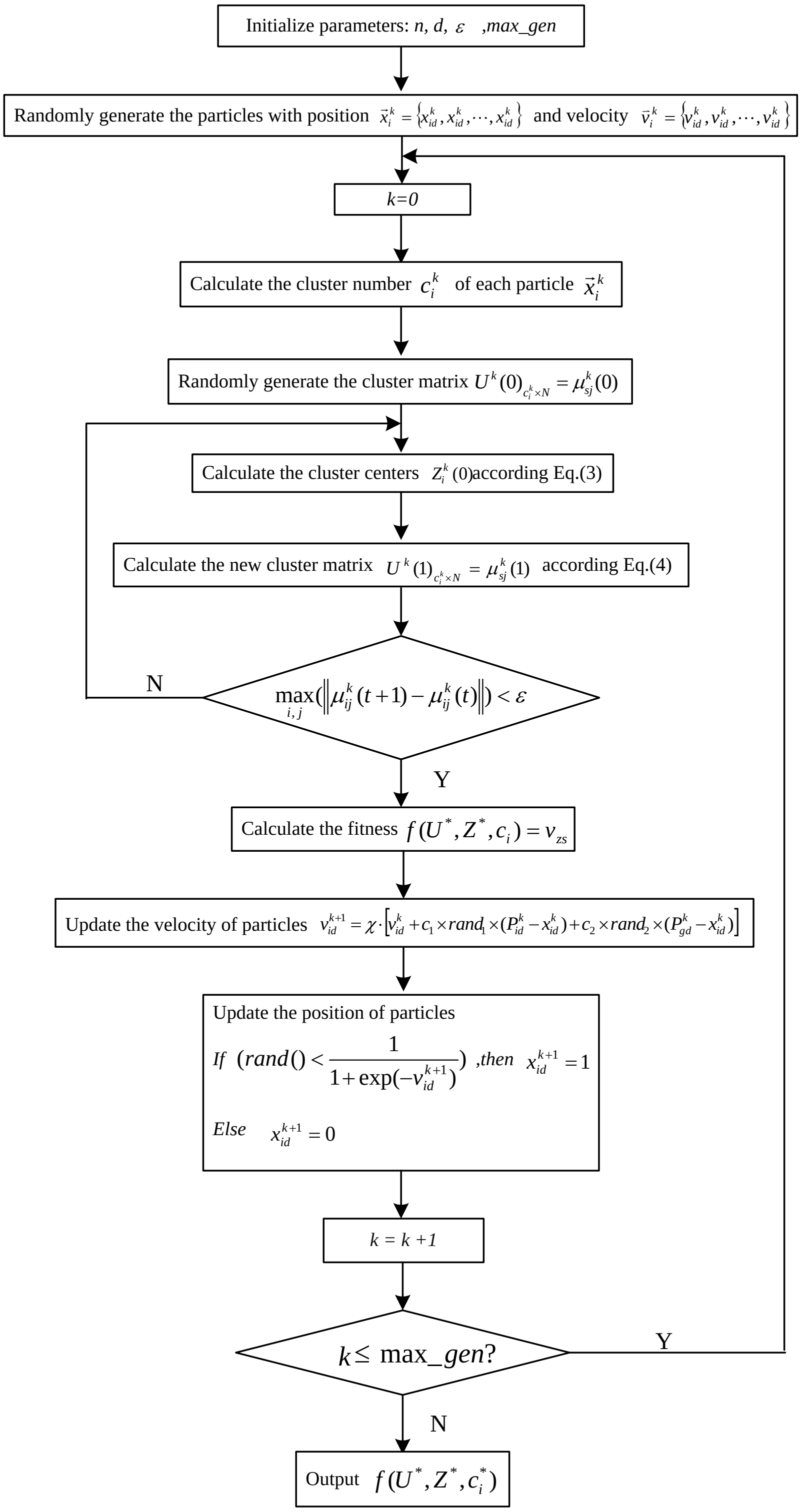

Fig.1: Flow chart for PSO-FCM

2.3 Data source

The characteristics of CO_2 emission can be depicted by five indicators: total CO_2 emission, per capita emissions, CO_2 emission per unit of GDP (intensity of CO_2 emission), CO_2 emissions per unit of energy and the proportion of industrial CO_2 emissions, representing respectively the emission scale, average interpersonal emission efficiency, resource endowment and the structure of emissions. The present study selected the five indicators from 2000 to 2009 as the clustering index to explore the regional characteristics of China's inter-provincial carbon emissions.

Fossil energy consumption data on China's 30 provinces were taken from the *China Energy Statistical Yearbook (CESY) 2001-2010* (Hong Kong, Macao, Taiwan and Tibet are not included for lack of data and the data on Ningxia is from the year 2002); the data on GDP and population were taken from the 2010 edition of the *China Statistical Yearbook*, and GDP is converted into 2000 constant price. According to equation 1, the average annual CO_2 emissions of 30 provinces during the period of 2000-2009 are shown in Fig 2.

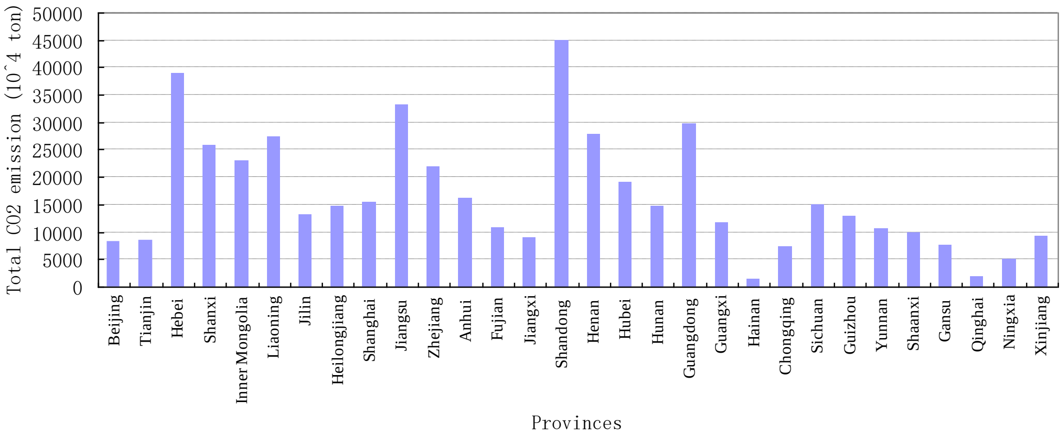

Fig.2: The average CO_2 emissions of each province from 2000 to 2009.

3. Clustering results and cause analysis

3.1 Clustering results

In the proposed PSO-FCM method of the present study, the size of particles is $n = 15$ and the length of each particle is $d = 5$ (because there are only 30 provinces in the study, $2^5 > 30$). The convergent threshold of ISODATA is $\varepsilon = 10^{-4}$; the fitness objective function is Eq.(1); the iterations of particles is $\max\_gen = 100$. The fitness value changes are shown in Fig.3.

After the selected characteristics of carbon emission data normalization and PSO-FCM implementation, the fitness value changes are shown in Fig.3. It can be seen that the PSO-FCM obtained the optimal cluster number $c = 5$ at around 80 iterations, and the carbon emission characteristics of China's provinces were divided into five categories. Furthermore, Fig. 3 also indicates that the convergence speed of the proposed PSO-FCM is faster than GA-FCM. The best fitness value is 3.5088. To validate the number of clusters, the clustering objective function value vs. the number of categories is shown in Fig. 4, depicting the least value in 5 categories. The optimal membership grades are given in appendix A Table and the clustering results are shown in Table 1, based on the principle of the largest membership grade. Compared with K-means clustering and hierarchical clustering, the results of the proposed method are quite different, except in the first category. As in the third category, it is difficult to discover the common characteristics between Hainan, in which each index of emission is the least of the 30 provinces, and the provinces with higher indicators, such as Jiangsu, Zhejiang, Guangdong and so forth. On the contrary, they have quite different characteristics. For example, the total CO_2 emission ranks third of the 30 provinces, with Hainan in penultimate place. The per capita emission of Zhejiang is 2.89 times that of Hainan, which is only 1.54 tons/ person. Furthermore, the CO_2 emission intensity of Shaanxi is 3.30 tons/10 thousand Yuan, which is considerably higher than that of Hainan, which is 1.48 tons/10 thousand Yuan. Although there are slight discrepancies in the units of energy emission and the proportion of industrial final energy-related emission, these cannot exclude more provinces from falling into this category. It is clear that the clustering results of K-means have little explanatory power. In terms of hierarchical clustering, Hainan has been placed in a separate category, but another twenty-two provinces, with significantly different emission characteristics have been clustered the same category, which does not appear logical.

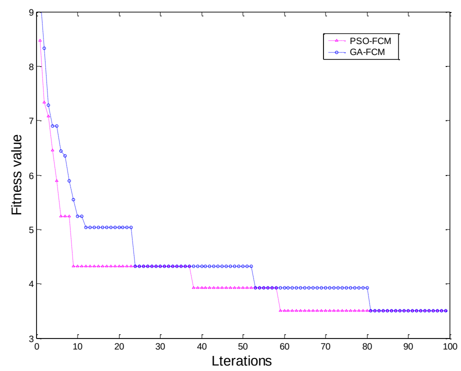

Fig.3 Fitness evolutions of the PSO-FCM and GA-FCM

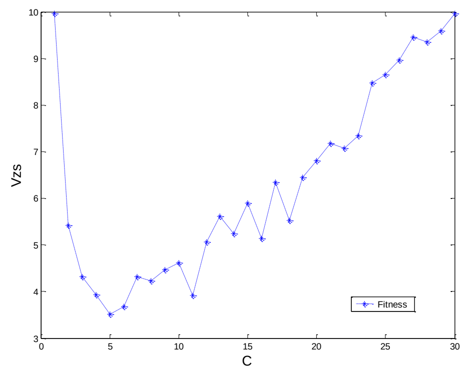

Fig.4 Cluster numbers VS. Fitness values

Table 1: The clustering results of PSO-FCM and comparisons with other methods

Classes	PSO-FCM Cluster	K-means Cluster	Hierarchical Cluster
Class 1	Shanxi Inner Mongolia Ningxia	Shanxi Inner Mongolia Ningxia Guizhou	Shanxi Inner Mongolia Ningxia Guizhou
Class 2	Tianjin Shanghai	Beijing Tianjin Jilin Shanghai Gansu Qinghai Xinjiang	Hebei Liaoning Shandong
Class 3	Jiangsu Zhejiang Shandong Guangdong	Heilongjiang Jiangsu Zhejiang Guangdong Hainan Shaanxi	Hainan
Class 4	Hebei Liaoning Anhui Fujian Jiangxi Hubei Henan Guangxi Chongqing Sichuan Yunnan	Anhui Fujian Jiangxi Hubei Hunan Guangxi Chongqing Sichuan Yunnan	Beijing Tianjin Jilin Heilongjiang Shanghai Jiangsu Zhejiang Anhui Fujian Jiangxi Henan Hubei Hunan Guangdong Guangxi Chongqing Sichuan Yunnan Shaanxi Gansu Qinghai Xinjiang
Class 5	Beijing Jilin Heilongjiang Hainan Guizhou Shaanxi Gansu Qinghai Xinjiang	Hebei Liaoning Shandong Henan	

*The clustering number of PSO-FCM is obtained by global optimization; K-means is set by the man-made 5; and the hierarchical cluster is decided by the maximum distance in categories in a tree diagram.

3.2 Characteristics and cause analysis

During the period between 2000 and 2009, the characteristics of China's inter-provinces carbon emission can be divided into five categories using the PSO-FCM method. The characteristics and their causes can be analyzed as follows.

(1) The first category is "double high", with a high intensity of CO_2 emission and a high per capita emission. This category is represented by Shanxi and Inner Mongolia provinces, which rank in the first class with a grade of more than 0.9 and which have significant characteristics. Looking at the data, it can be clearly seen that the emission intensity of CO_2 in these provinces (more than 0.73 tons/ thousand Yuan) is much higher than in other provinces. It is almost twice the national average level (0.37 tons/thousand Yuan). At the same time, the emission intensity of CO_2 in these provinces is also much higher, at 1.76 times as the

national average (4.31 tons/person), as shown in Fig. 5.

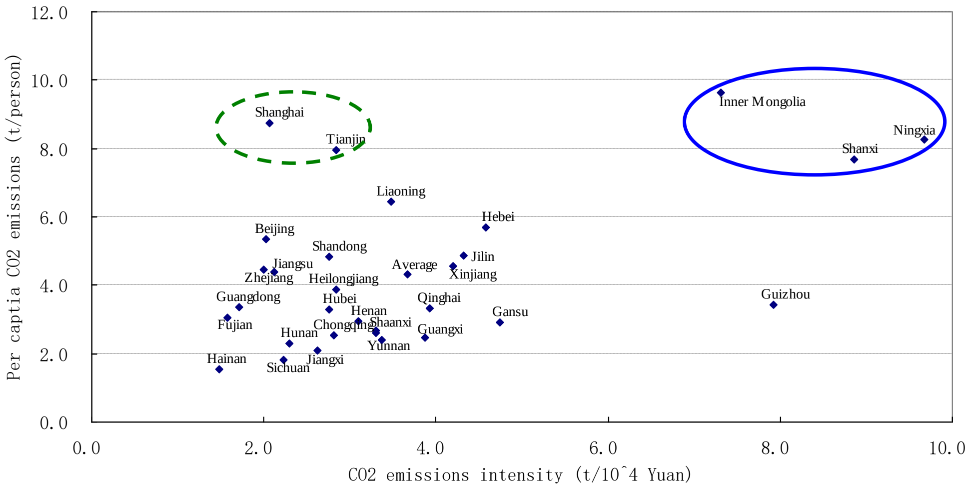

Fig.5 Scatter diagram of tCO_2 emission intensity and per capita emission

The main reasons for the appearance of this characteristic can be summarised as follows:

Firstly, there is the impact of resource endowment. China's main coal-rich provinces are Shanxi, Inner Mongolia, Shaanxi, Guizhou and Xinjiang. As shown in Fig. 6, the CO_2 emission intensity in these provinces is higher than others. More specifically, in the period 2000 to 2009, the coal resource reserves of Shanxi and Inner Mongolia comprised 31.90% and 21.41% respectively of China's entire coal resource reserves, , although their total GDPs were only 1.75% and 1.74% respectively of China's national GDP. Similarly, although the coal resource reserves of Ningxia comprised 1.94% of the country's total coal resource reserves, ranking No. 12 in the thirty province of China, Ningxia's total GDP accounted for only 0.26% of national GDP, giving it penultimate ranking in the country.

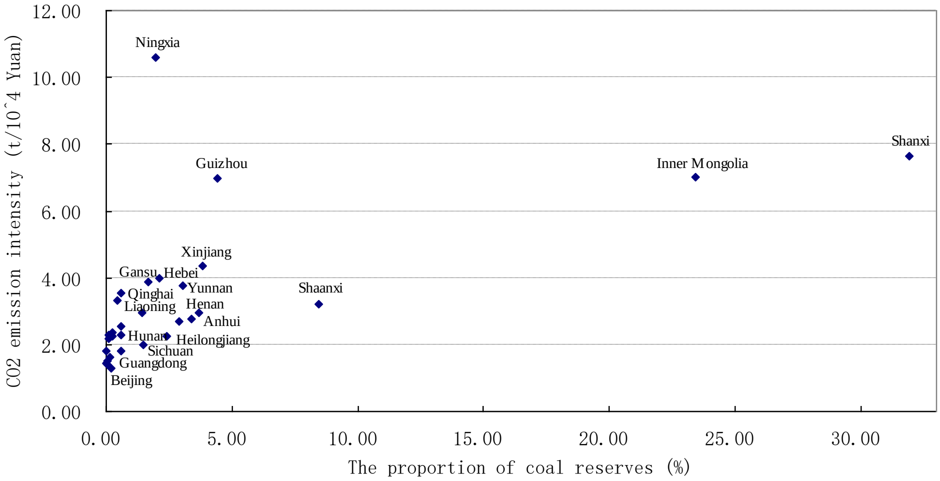

Fig. 6: Scatter diagram of CO_2 emission intensity and coal resource reserves (2000-2009)

Secondly, there is the impact of industry structure. The provinces of the first category are the major regions of energy production and processing. In addition, the proportion of energy-intensive industries, which are dominated by coal mining and processing industries, as well as the low technology of energy-consuming equipment, lead to a high emission intensity of carbon dioxide. From 2000 to 2009, the proportion of secondary industries of Shanxi, Inner Mongolia and Ningxia were 56.16%, 47.01% and 48.61% respectively, higher than the national average level of 46.40%. In particular, Shanxi has the largest proportion of secondary industry in all provinces

Finally, there is the impact of the national energy strategy layout. In 2011, the China land space development plan, "The Main Function of National Planning", stated that China would focus on constructing Shanxi, the Ordos Basin, eastern Inner Mongolia, southwestern of China, and Xinjiang as the five integrated national energy bases. It is well known that Shanxi, Inner Mongolia and Ningxia provinces have all been included in the comprehensive national energy bases. Shanxi, the Ordos Basin and eastern Inner Mongolia are focused on constructing the energy industries such as thermal power, coal mining and processing industries. From 2000 to 2009, the annual coal consumption of Shanxi, Inner Mongolia, and Ningxia was 4296.61 tce, 4866.72 tce, and 846.10 tce, respectively. However, the total CO_2 emissions from thermal power of these provinces were 42.18%, 54.80% and 54.54%, respectively shown in Fig. 7. These three provinces have undertaken a large CO_2 emission transference to other provinces as part of their economic development.

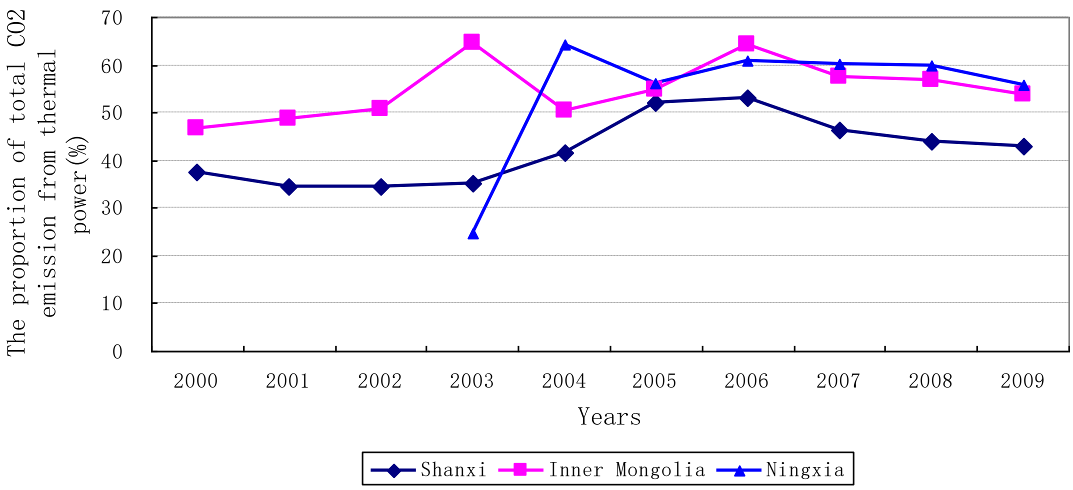

Fig.7: The proportion of thermal power CO_2 emissions of the first category provinces in China

（2）Tianjin and Shanghai, directly governed by the central government, belong to the second category. The grade of this class can reach up to 0.99. This type of carbon emission has the characteristics of high per capita emission and low intensity. These two cities have high per capita carbon dioxide emissions (over 7.90 tons per person), 1.80 times of the average national level (4.30 tons per person). However, their carbon intensity (less than 2.90 tons / 10 thousand Yuan), which is less than the 79.00% average national level (3.70 tons /10 thousand Yuan), is relatively low, as illustrated in Fig 5.

The main causes of these characteristics are as follows. First, these two cities are located in eastern coastal areas, where there is a high level of economic development and have a great demand for fossil energy. As a result, they have high carbon dioxide emissions. However, according to National Residence Register, the population is small; therefore, these cities have high per capita carbon dioxide emissions. From 2000 to 2009, the carbon dioxide emissions of Tianjin and Shanghai represented 1.84% and 3.70% of the national total emission respectively, while their populations only comprised 0.83% and 1.37% of the country's total population. Accordingly, these two cities had the highest per capita carbon dioxide emissions. (However, it should be noted here that both Tianjin and Shanghai have large floating population, while only the official Residence Register figures are taken into account. If the floating population was taken into consideration, the per capita carbon dioxide emission could fall to some extent.)

Secondly, Shanghai, at the head of the Yangtze River Delta, has begun to make changes in its economic structure and is gradually becoming the national center of overseas trade, finance and R & D. The proportion of secondary industry, which produces high carbon dioxide emissions, is in decline, with its proportion falling from 47.30% in 2000 to 39.90% in 2009. Tianjin, the northern economic center, has seen a rapid increase in GDP during the last ten years, but the proportion of its secondary industry has remained at around 50.00%. At the same time, due to their lack of fossil fuel energy, most of their electric power is provided by other provinces. From 2000 to 2009, the annual transferred proportion of the electricity consumed in Tianjin and Shanghai rose from around 11.98% and 21.07% to as much as 27.33% and 33.80%, respectively, as illustrated in Fig 8. In view of electricity in China being generated mainly by thermal power, Tianjin and Shanghai have already diverted a certain large amount of carbon dioxide emissions, so that their emission intensities have become lower than the national average level.

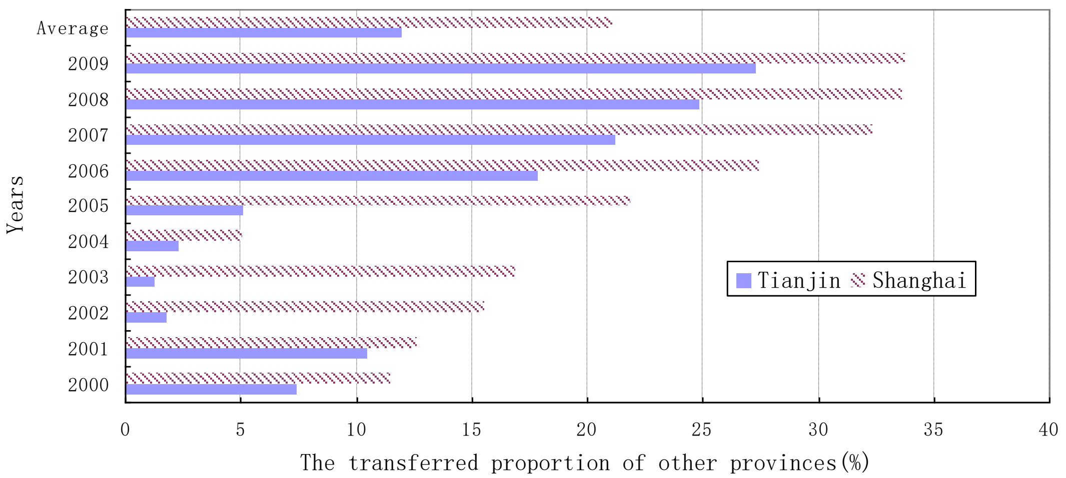

Fig.8: The proportion of imported electricity consumption amount in the total consumption of the second category provinces

(3) The CO_2 emissions of the third category provinces amount to more than 210.00 million tons, but per capita emissions are less than 5.00 tons. Thus, the characteristic of these provinces can be described as high total emission and low per capita emission, as illustrated in Fig.9. The provinces in this category are mainly coastal regions such as Jiangsu, Zhejiang, Shandong, and Guangdong, and a large province in central China, Henan. The total CO_2 emission of these provinces is much higher than the national average level (about 165.29 million tons), comprising around 31.79% of CO_2 emission of China.

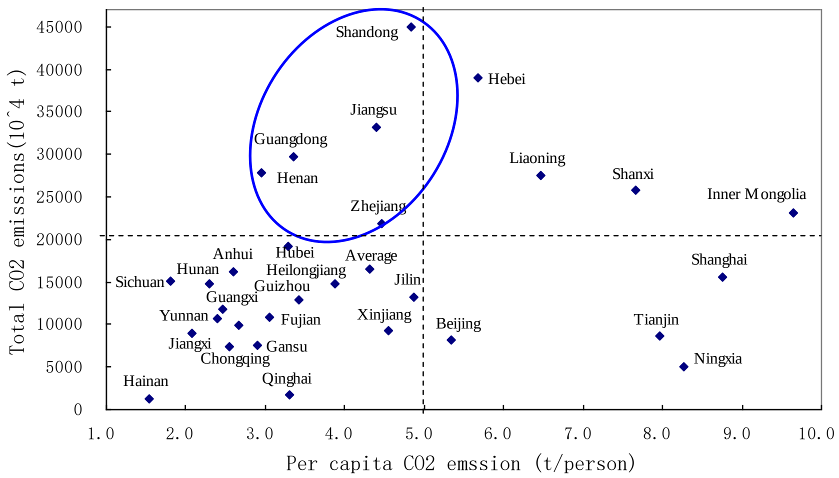

Fig. 9: Scatter diagram of the total CO_2 emissions and per capita emissions of each province （2000-2009）

The main reasons for the appearance of these characteristics can be summarized as follows:

According to the Kaya equation, factors contributing to CO_2 emissions are population, economic development, energy intensity and energy structure. The first two factors are the main driving factors. The presence of such characteristics in the five provinces are mainly due to large economic size and population size, leading to high total emission, although per capita emission is not high.

In an analysis of the five provinces belonging to the third category, it can be seen from Fig. 10 that those whose average GDP comes in the top 5 out of the 30 provinces from 2000 to 2009 accounted for 41.00% of China's GDP. With the expansion of total economic output, energy-related CO_2 emissions are accelerating in an upward trend, which means that these provinces are developing economically at the cost of excessive energy consumption.

Furthermore, provinces in the third category also have large populations. These five provinces made up 31.00% of the population of China from 2008 to 2009. Among these provinces, Henan, Shandong and Guangdong take the first three places with populations of 94.86 million, 92.32 million and 87.67 million respectively, while Jiangsu and Zhejiang account for 5.84% and 3.78%, respectively, of the total population, as shown in Fig 11. Therefore, the "double high" of economic scale and population has led to the characteristics of these provinces, which are high total emissions but low per capita emissions.

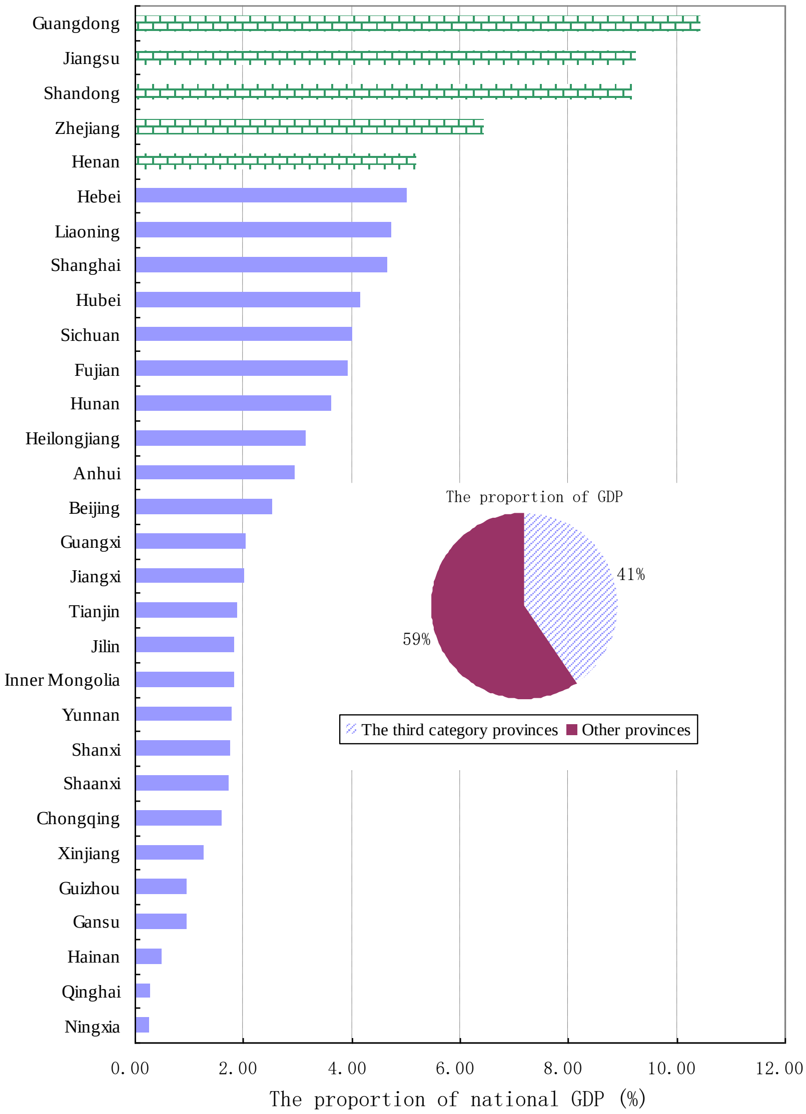

Fig. 10: The GDP rankings and proportions of the third category provinces (2000-2009)

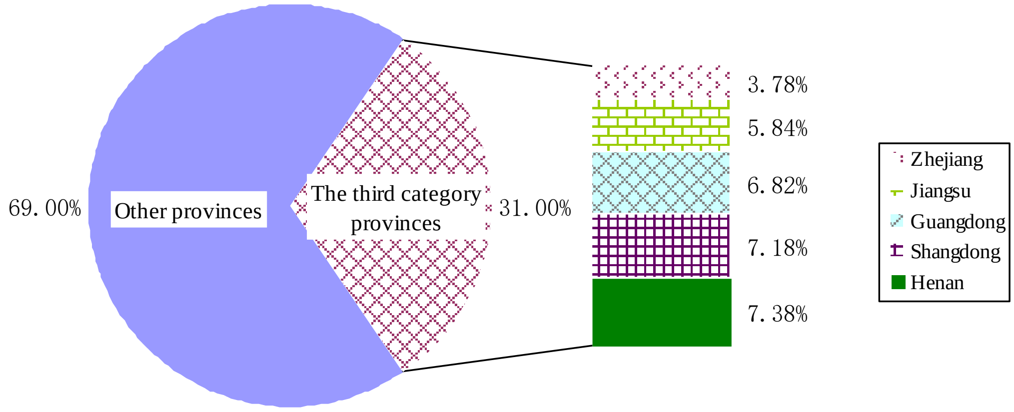

Fig.11: The proportion of the populations of the third category provinces in relation to the total population of China

(4) The fourth category consists of the provinces which have many final industries consuming a great deal of fossil fuel energy. The provinces whose proportion of the industrial carbon dioxide emission is more than 40.00%, but with an intensity of less than 5.00 tons/10 thousand Yuan, are the industrial provinces of Hebei, Liaoning, and Anhui. The characteristic of these provinces is that the proportion of industrial carbon emissions is high, while the intensity is low, as illustrated in Fig 12.

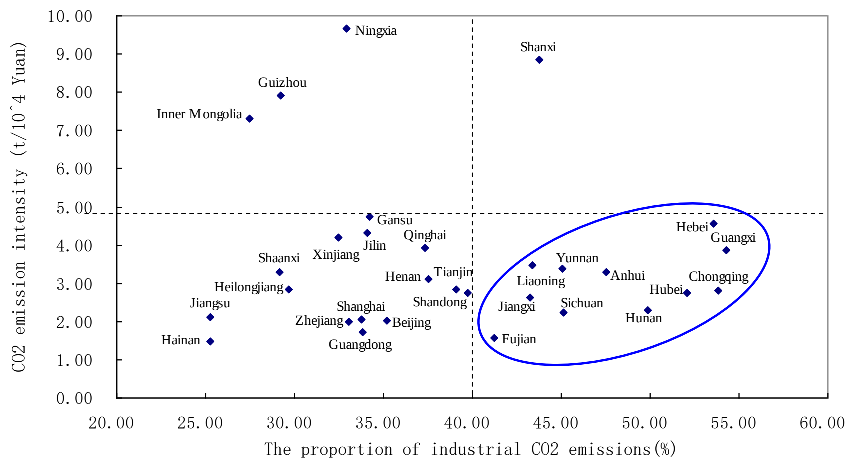

Fig. 12: Scatter diagram of the emission intensity and proportion of industrial emissions (2000-2009)

The main causes for this high proportion of industrial emissions and low intensity are as follows.

Firstly, the industrial GDP is higher than 50.00% among some of the fourth category provinces, such as Hebei and Liaoning, where average industrial GDP in industry accounted for 51.73% and 50.39% respectively from 2000 to 2009. In these two provinces, the metallurgy industry and the oil processing industry consume a great deal of fossil fuel energy, resulting in high proportions of industrial emissions.

Secondly, because of the particular regional situation, provinces in this area, apart from Anhui, are neither centralised heating supply provinces nor thermal power generation provinces. However, the proportion of energy transfer sector (providing heating supply and thermal power generation) consumption is low compared to the total energy consumption and the final energy is mostly consumed by the industrial sector. Taking the consumption of coal in 2009 as an example, for example, in the provinces belonging to the fourth category, the consumption of the thermal power generation and heating supply was lower than the national average level of 53.86%, while industry final consumption was higher than the national average level of 21.45%. Of these provinces, Sichuan, Chongqing, Guangxi, Hunan and Hubei had industry final consumptions exceeding 40%, around twice the national average level, as illustrated in Fig. 13.

Thirdly, the provinces in the fourth category do not have abundant energy resources. The total emission of carbon dioxide in most provinces is lower than the average level of the thirty provinces, while the scales of GDP are average, so that their intensity is not the highest.

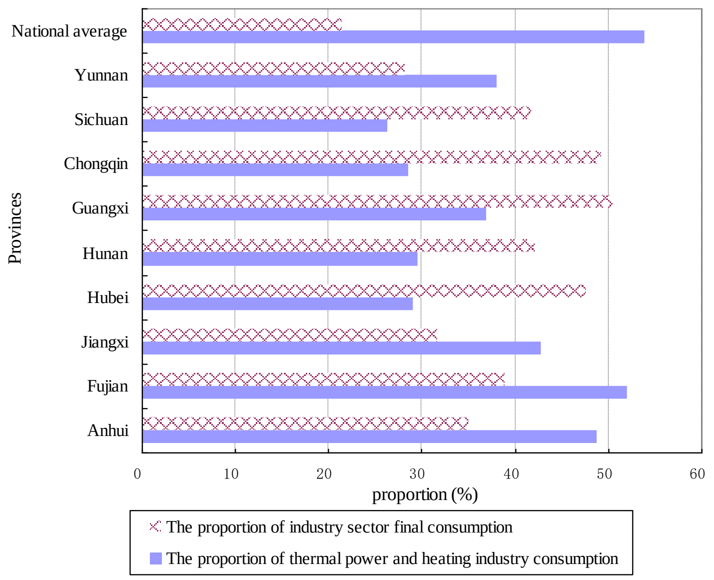

Fig. 13: The proportions of energy consumption of thermal power, heating supply and the industrial sector in the fourth category provinces in 2009

(5) The fifth category mainly comprises the remaining provinces, whose characteristics are not particularly significant. The grade of some of the provinces belonging to this category is not high. For example, Beijing, Hainan and Guizhou have only about 0.5. Through careful analysis, it can be concluded that the total CO_2 emission is less than 150 million tons, lower than the national average level of 165.29 million tons. Furthermore, the proportion of industrial discharge is less than 37.00%, which does not reach the national average level (38.74%), as illustrated in Fig.14. Although Ningxia would be classified into this group based on this characteristic, its grade in the first category is as high as 0.91. Therefore, the classification method of PSO-FCM does not place it in this group.

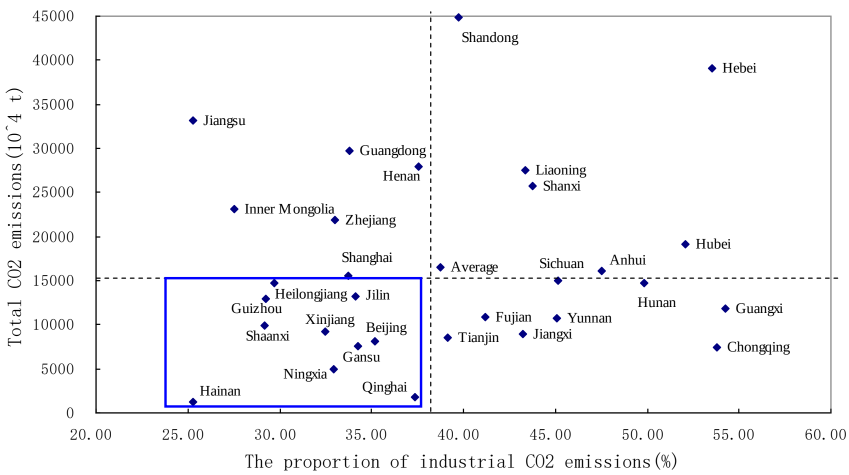

Fig. 14: Scatter diagram of the total emissions and the proportion of industrial emissions (2000-2009)

4 Main conclusions and implications for policy on setting emission reduction targets

4.1 Conclusions

(1) Taking into account inter-category and within-category distances, the proposed PSO-FCM clustering method can partition samples more effectively. Moreover, PSO-FCM can obtain the global optimal number of clusters and membership grade values. Compared to other traditional clustering methods, the results of PSO-FCM method are more explanatory and rational.

(2) During the period between 2000 and 2009, the characteristics of China's inter-provincial carbon emissions can be divided into five categories. These are the "double high" category, with a high intensity of CO_2 emission and high per capita emission, represented by Shanxi and Inner Mongolia provinces; the "high per capita-low intensity" category, with high per capita emission but low emission intensity, represented by Shanghai and Tianjin; "high emission-low per capita" category with high total emission but relatively low per capita emission, represented by Jiangsu and Zhejiang; "high industry-low intensity" category with a high ratio of industrial emission and low emission intensity, represented by Hebei and Liaoning; and the "double low" category, with low emission and low industrial emission, represented by Beijing and Jilin.

(3) The two most important indicators affecting the regional characteristics of China's carbon emission are CO_2 emission intensity and per capita emission, while the role of CO_2 emission per unit of energy in the classification is not apparent. The main reason for this situation is that energy consumption is based mainly on coal consumption in each province, with little difference among provinces. For example, from 2000 to 2009, the national average level of emission per unit of energy was 2.48 t / tce, while the variance between the 30 provinces was only 0.01; thus, its impact on the regional characteristics of carbon emission is not significant. In fact, there would be no significant change in classification results if this indicator was removed.

4.2 Implications for policy on setting emission reduction targets

Based on the analysis of clustering characteristics and their causes in the carbon emissions of China's provinces, the major policy recommendations on setting emission reduction targets for different regions are as follows. The targets can be set to vary from category to category, categories, and provinces belonging to the same category can have the same target.

(1)Although the provinces in the first group have the highest carbon emissions of the 30 provinces of China and therefore have the greatest potential for cutting their carbon dioxide emissions, the target for these provinces should not be set as the highest of all the provinces. The reason for this is that is that the national energy strategy and energy resource endowments must be taken into consideration, as these take a considerable amount of transferred carbon emissions of other provinces. In contrast, the provinces in the third category, whose emissions and GDP are the highest in China, can have their target set as the highest of all the provinces of China. On one hand, this could encourage them to take advantage of economics and technology to improve energy efficiency or to support clean energy construction. On the other hand, they could be pressed to accelerate the promotion of high discharge industry in order to promote the development of the low carbon service industry.

(2) Despite the carbon intensity of the fourth group being less than the national average level, the proportion of industrial sector emissions is relatively high and the energy efficiency of the industrial sector is very low. Thus, the target could be set to national average level, namely 40-45%. However, the provinces of fifth category, in which total emission and the proportion of industrial sector emissions are not high,

could be set a target lower than national average level.

(3) Based on the characteristics of carbon emissions and the reasons for them, the target of the provinces in the second group could be set to be similar to that of the third category provinces. In sum, the targets of the five categories can be arranged in the following order: third category >second category >first category (national level)>fourth category >fifth category.

References

[1] Akins MJ, Morrison AS, Walmsleym RW. Carbon Emission Pinch Analysis (CEPA) for emission reduction in the New Zealand electricity sector. Applied Energy 2010; 87(3): 982-987.

[2] Sharma SS. Determinants of carbon dioxide emission: Empirical evidence from 69 countries. Applied Energy 2011; 88(1):376-382.

[3] Clarke-sather A, Qu J, Wang Q, etc.. Carbon inequality at the sub-national scale: A case study of provincial-level inequality in CO_2 emission in China1997-2007. Energy Policy 2011;doi:10.1016/j.enpol.2011.05.021.

[4] Yue C, Hu X, He C et al. Provincial carbon emission and carbon intensity in China from 1995 to 2007(Carbon Emission and Social Development, III).Acta Scientiarum Naturalium Universitatis Pekinensis 2010;46(4):510-516.(In Chinese)

[5] Feng K, Hubacker K, Guan D. Lifestyles, technology and CO2 emission in China: A regional comparative analysis .Ecological Economics 2009;69(1):145-154.

[6] Ma J, Chen Z, You J. Spatial econometric analysis on China's provincial primary energy-related carbon dioxide emission. Technology Economics2010;29 (12):62-67.(In Chinese)

[7] Zhu C, Ma Z, Wang C, Liu Z. Analysis of difference features of energy-related CO_2 emission in China. Ecology and Environment 2006;15(5):1029-1034.(In Chinese)

[8] Wei Y-M, Liu L-C. Fang Y. China energy report (2008): carbon emission. Scientific Press, Beijing, 2008. (In Chinese)

[9] He J, Li S. Carbon Reduction and regional economy. Management review 2010; 22(6):9-16. (In Chinese)

[10] Wang R, Liu W., Xiao L, etc.. Path towards achieving of China's 2020 carbon emission reduction target-A discussion of low-carbon energy policies at province level. Energy Policy 2011; 39(5):2740-2747.

[11] Yang G, Tang Z, Shi H, et al.. Regional differences of carbon dioxide emission in China and emission-reduction policy recommendations. Advances in Climate Change Research 2011;7(1):35-40.(In Chinese)

[12] Gao W, Jiang W, Xie H. Effect of economic development on energy carbon emission distribution. Journal of Liaoning Technical University（Natural Science）2009;28(2):296-299. (In Chinese)

[13] Zhang L, Huang Y, Li Y, et al.. An investigation on spatial changing pattern of CO_2 emissions in China. Resources Science 2010; 32(2):211-217. (In Chinese)

[14] Wang Q, Zhou P, Zhou D. efficiency measurement with carbon dioxide emissions: The case of China. Applied Energy 2011; doi:10.1016/j.apenergy.2011.02.022.

[15] Yue R, Zhu Y. Provincial cluster analysis on china's energy carbon emission from 1990 to 2007. Technology Economic 2010; 29(3):40-45.

[16] Bezdek JC. Pattern Recognition with fuzzy objective function algorithms. Plenum, New York, 1981.

[17] Pan T-H, Shieh S-S, Jang S-S, et al.. Statistical multi-model approach for performance assessment of

cooling tower. Energy Conversion and Management 2011; 52(2): 1377-1385.

[18] Besheer AH, Emara HM, Abdel_Aziz MM. Wind energy conversion system regulation via LMI fuzzy pole cluster approach. Electric Power Systems Research 2009; 79(4):531-538.

[19] Gómez M, López A, Jurado F. Optimal placement and sizing from standpoint of the investor of photovoltaics grid-connected systems using binary particle swarm optimization. Applied Energy 2010; 87(6): 1911-1918.

[20] Rezaee B. A cluster validity index for fuzzy clustering .Fuzzy Sets and Systems 2010; 161 (23):3014-3025.

[21] Zhang B, Yao N. Liu X-M. Preliminary fuzzy clustering analysis on carbon emission in different provinces of China. China population, resources and environment 2011; 21(1):54-56. (In Chinese)

[22] Groll L, Jakel J. A new convergence proof of fuzzy C-means. IEEE Transactions Fuzzy Systems 2005; 13 (5):717-720.

[23] Juang C, Hsiao C-M, Hsu C-H. Hierarchical cluster-based multispecies particle-swarm optimization for fuzzy-system optimization. IEEE Transactions on fuzzy systems 2010; 18(1):14-26.

[24] Filho FM, Gomide F. Fuzzy clustering in fitness estimation models for genetic algorithms and applications. Proceedings of IEEE International Conference on Fuzzy Systems, 2006.pp. 1388 -1395.

[25] Saha S, Bandyopadhyay S. A new point symmetry based fuzzy genetic clustering technique for automatic evolution of clusters. Information Sciences 2009; 179(19): 3230-3246.

[26] Yoon J-W, Cho S-B. An efficient genetic algorithm with fuzzy c-means clustering for traveling salesman problem. Proceedings of IEEE Congress on Evolutionary Computation 2011,pp. 1452 -1456.

[27] Runkler TA, Katz C. Fuzzy clustering by particle swarm optimization. Proceedings of 2006 IEEE International Conference on Fuzzy Systems, Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada.

[28] Huang Y K. A hybrid particle swarm optimization approach for clustering and classification of datasets. Knowledge-Based Systems 2011; 24(3):420-426.

[29] Kuo RJ, Wang MJ Huang TW. An application of particle swarm optimization algorithm to clustering analysis. Soft Computing-A Fusion of Foundations, Methodologies and Applications 2010; 15(3): 533-542.

[30] IPCC. Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change, eds Solomon S, et al., eds. Cambridge University Press, Cambridge, UK), 2007.

[31] Bezdek JC, Pal NR. Some new indices of cluster validity. IEEE Transaction on Systems, Man, and Cybernetics 1998; 28(3): 301-315.

[32] Zhu K, Su S, Li J. Optimal number of clusters and the best partition in fuzzy C-mean. Systems Engineering-Theory &Practice 2005; 25(3):52-61.

[33] Hsieh S-T, Sun T, Liu C-C. Potential offspring production strategies: An improved genetic algorithm for global numerical optimization .Expert Systems with Applications 2009; 36(8):11088-11098.

[34] Kennedy J, Eberhart RC. Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks; 1995. p. 1942-1948.

[35] Yu S, Zhu K，Guo S. A hybrid MPSO-BP structure adaptive algorithm for RBFNs. Neural Computation& Applications , 2009(18):769-779.

[36] Unler A, Murat A.A discrete particle swarm optimization method for feature selection in binary classification problems. European Journal of Operational Research 2010; 206 (3): 528-539.

Appendix A

The membership grades of energy -related CO_2 emission characteristics from 2000-2009 according to province

Provinces	Class 1	Class 2	Class 3	Class 4	Class 5	Provinces	Class 1	Class 2	Class 3	Class 4	Class 5
Beijing	0.0198	0.3054	0.0899	0.1001	**0.4848**	Henan	0.0096	0.0242	**0.7244**	0.129	0.1128
Tianjin	0.0007	**0.9953**	0.001	0.0008	0.0022	Hubei	0.0033	0.0135	0.0256	**0.9361**	0.0214
Hebei	0.141	0.1271	0.239	**0.3992**	0.0937	Hunan	0.0001	0.0003	0.0006	**0.9981**	0.0008
Shanxi	**0.9181**	0.0245	0.0175	0.0207	0.0193	Guangdong	0.0018	0.0119	**0.9478**	0.0095	0.029
Inner Mongolia	**0.9394**	0.0294	0.0123	0.005	0.014	Guangxi	0.0015	0.003	0.0039	**0.9850**	0.0067
Liaoning	0.1493	0.2212	0.2466	**0.2643**	0.1186	Hainan	0.0275	0.1239	0.1776	0.0915	**0.5795**
Jilin	0.0305	0.0818	0.0701	0.0497	**0.7679**	Chongqing	0.0011	0.0032	0.0034	**0.9854**	0.0069
Heilongjiang	0.0045	0.0409	0.1249	0.0138	**0.8159**	Sichuan	0.0013	0.0055	0.0193	**0.9429**	0.031
Shanghai	0.0012	**0.9917**	0.0033	0.0008	0.003	Guizhou	0.2152	0.0617	0.092	0.0644	**0.5667**
Jiangsu	0.009	0.0302	**0.8896**	0.0129	0.0582	Yunnan	0.0017	0.0038	0.0064	**0.9688**	0.0194
Zhejiang	0.0032	0.0476	**0.8135**	0.0139	0.1218	Shaanxi	0.0008	0.0033	0.009	0.0037	**0.9831**
Anhui	0.0001	0.0002	0.0004	**0.9987**	0.0007	Gansu	0.0017	0.0037	0.0056	0.0093	**0.9797**
Fujian	0.0065	0.0661	0.1209	**0.4221**	0.3844	Qinghai	0.0048	0.0184	0.013	0.0559	**0.9079**
Jiangxi	0.005	0.0127	0.0217	**0.8936**	0.0669	Ningxia	**0.9055**	0.0372	0.0121	0.0109	0.0345
Shandong	0.0129	0.0346	**0.8892**	0.0334	0.0298	Xinjiang	0.0019	0.0141	0.0091	0.0038	**0.9711**