Green efficiency performance analysis of the logistics industry in China: based on a kind of machine learning methods

Xin Yao[a], Yuanyuan Cheng[a], Li Zhou[b], Malin Song[c,*]

[a] *China Center for Energy Economics Research, School of Economics, Xiamen University, Xiamen, 361005, China*

[b] *Systems Management and Strategy Department, Business School, University of Greenwich, SE10 9LS, UK*

[c] *School of Statistics and Applied Mathematics, Anhui Finance and Economics University, Bengbu, 233030, China*

[*] Corresponding author. E-mail: songmartin@126.com

Abstract: This paper aims to analyze the green efficiency performance of the logistics industry in China's 30 provinces from 2008 to 2017. We first evaluate the green efficiency of the logistics industry through the non-directional distance function (NDDF) method. Then, we use the functional clustering method funHDDC, which is one of the popular machine learning methods, to divide 30 provinces into 4 clusters and analyze the similarities and differences in green efficiency performance patterns among different groups. Further, we explore the driving factors of dynamic changes in green efficiency through the decomposition method. The main conclusions of this paper are as follows: (1) In general, the level of green efficiency is closely related to the geographical location. From the clustering results, we can find that most of the eastern regions belong to the cluster with higher green efficiency, while most of the western regions belong to the cluster with lower green efficiency. However, the green efficiency performance in several regions with high economic levels, such as Beijing and Shanghai, is not satisfactory. (2) Based on the analysis of decomposition results, the innovation effect of China's logistics industry is the most obvious, but the efficiency change still needs to be improved, and technical leadership should be strengthened. Based on these conclusions, we further propose some policy recommendations for the green development of the logistics industry in China.

Keywords: logistics industry; green efficiency; non-directional distance function; functional clustering; machine learning

1. Introduction

1.1 Background

The logistics, which is a pillar industry, plays a pivotal role in the social production, but the huge energy consumption and environmental pollution it brings are also prominent (Oberhofer and Dieplinger 2013; Dai and Gao 2016; Sureeyatanapas et al. 2018). Specifically, according to the statistics of the International Energy Agency (IEA), the energy consumption of the transportation sector, which is an important part of the logistics, is 2808.15 Mtoe in 2017 and accounts for 28.90% of global energy consumption. And the CO_2 emissions of the transportation sector reach 8040 million tones in 2017, accounting for 24.48% of the world's total. As for China, the huge demand for logistics services has made the energy consumption and environmental pollution problems of the logistics industry increasingly serious. In 2017, the energy consumption of the transportation sector is 313 Mtoe in China, and the CO_2 emissions of the transportation sector are 889 million tones. Besides, the packaging in the Chinese logistics industry also leads to a lot of resource waste. For example, Double 11 shopping carnival of China generated 25 billion US dollars sales in 2017, but at the same time, it also brought 1.5 billion packaging parcels most of which were not recycle (Kuo et al. 2019).

With increasing attention to the energy-saving and environmental protection, many emerging conceptions in the logistics have been proposed to seek sustainable development patterns, such as reverse logistics (Zarbakhshnia et al. 2019; Chen et al. 2019; Suzanne et al. 2020; Dev et al. 2020), green coal logistics (Li et al. 2020), green maritime logistics (Davarzani et al. 2016), and green supply chain (Dekker et al. 2012; Khor and Udin 2013; Ameknassi et al. 2016; Ahmed et al. 2019). Green logistics aims to realize the economic profits as well as environmental benefits for the whole operation and management in the logistics industry (Sbihi and Eglese 2010; Chhabra et al. 2017). To better achieve the green and sustainable development of the logistics industry, the consumption of resources and the impact on the environment should be reduced as much as possible. Therefore, it is of great significance to analyze the green efficiency performance of the logistics industry.

1.2 Contribution

In this paper, we aim to comprehensively analyze the green efficiency performance of China's logistics industry. Therefore, three questions are explored: (1)

What is the specific green efficiency performance of the logistics industry in different provinces of China? (2) What are the differences in the green efficiency development patterns of the logistics industry in different provinces? (3) What are the driving factors that affect the dynamic changes in the green efficiency of the logistics industry? Based on the data from the logistics industry of 30 provinces in China from 2008 to 2017, we have done the following works for our analysis. For the first question, we calculate the green efficiency scores of the logistics industry in different regions based on the non-directional distance function (NDDF) method. For the second question, we use the functional clustering method (funHDDC) to divide the 30 provinces into 4 clusters and study the similarities of the green efficiency development patterns of each cluster and the differences among different groups. For the third question, we further obtain three different indexes through decomposing the dynamic changes in green efficiency. And we analyze the influencing factors of changes in green efficiency of the logistics industry from three perspectives: catch-up effect, innovation effect, and technology leadership.

1.2.1 Contribution for research

Firstly, this paper measures the green efficiency performance of the logistics industry in different provinces of China through specific values, which can reflect the green efficiency of China's logistics industry more clearly. Secondly, this paper applies cluster analysis to the green efficiency of the logistics industry, thereby more objectively summarizing the different development patterns of the green efficiency in the logistics industry. Thirdly, this paper studies the influencing factors of the dynamic changes in green efficiency through the decomposition method, which can better show the characteristics of green efficiency changes.

1.2.2 Contribution for practice

Based on the analysis of results in this paper, we put forward relevant policy recommendations from three aspects. The first is how to further improve the green efficiency performance of logistics in the regions with high economic levels, such as Beijing and Shanghai. The second is how to promote the coordinated development of the logistics industry in adjacent areas. The third is how to improve the green efficiency of the logistics industry through technological innovation.

The remainder of this paper is organized as follows. Section 2 is about the literature review. In Section 3, we explain the methods and data used in this paper.

Section 4 and 5 focus on specific results. In Section 6, we further discuss the related results. Section 7 is about conclusion and policy implications.

2. Literature review

2.1 Green development of the logistics

The emphasis of green development is the coordinated development of economy, environment, and resources (Lin and Zhu 2019a; Yi and Liu 2015). And it has received the concerns of many studies (Yi and Liu 2015; Bagheri et al. 2018; Jin et al. 2019; Lin and Zhu 2019b; Zhu et al. 2020). Among them, the green development of the industry is the hot topic that many researchers pay attention to (Berry et al. 2013; Yuan and Xiang 2018; Shen et al. 2018; Jiahuey et al. 2019; Hou et al. 2019; Li et al. 2019). Due to the outstanding problems of energy consumption and environmental pollution in the logistics, the green development of the logistics industry has received more and more attention. Some scholars have carried out related research on the overall green development of the logistics industry. For example, Yang et al. (2019) investigates the carbon emissions performance of the logistics in sixteen Chinese cities and finds that the economic development has a positive effect on improving carbon emissions performance in the logistics. Similarly, Aldakhil et al. (2018) analyzes the green development performance of the logistics industry in the BRICS countries, the findings indicate that there exists a positive correlation between the green logistics index and national per capita income. Based on the LMDI decomposition method, Dai and Gao (2016) calculates the energy consumption of the logistics industry in China and further analyzes the changes in the energy consumption structure as well as the ways to improve the energy efficiency. Through the data envelopment analysis, Rashidi and Cullinane (2019) evaluates the sustainable development performance of the logistics industry in OECD countries and further shows the comparative analysis results of this efficiency score and the Logistics Performance Index (LPI).

Some scholars are mainly concerned about the sustainable development of transportation. Yao et al. (2019) discusses the role of collaboration in city logistics and the simulation results verify that collaboration between the two carriers can increase profits while reducing carbon emissions during transportation. Similarly, the research of Demir et al. (2019) shows that the economic and environmental benefits of cargo transportation can achieve coordinated development. Based on the mathematical programming model constructed, Hong et al. (2019) analyzes the problem of transportation costs minimization and takes into account the carbon emissions as well as PM2.5. Goswami et al. (2020) finds that the environmental performance of freight transport can be improved by effective transport plans and distribution network strategy. Garza-Reyes et al. (2016) proposes a new methodology to improve the efficiency and environmental benefits in the transport and logistics sector. What's more, the last mile logistics is also one of the key points. Melkonyan et al. (2020) proposes three new distribution channel options and explores the sustainable development of the last mile logistics in food transportation. Ji et al. (2019) uses the integer programming model and the surrogate model to optimize the assignment problems of express cabinets and tries to reduce the energy consumption and costs of the last mile logistics in the cities.

Reverse logistics plays an important role in promoting the development of a circular economy. Based on the evolutionary game model and different reverse logistics strategies, Gu et al. (2019) analyzes the role of cooperation between firms in the energy-intensive industries on the improvement of environmental performance. In the context of carbon cap-and-trade emissions, Zhang et al. (2018) analyzes the effects of carbon policies on carbon emissions reduction in reverse logistics. Some other aspects have also been discussed by the researchers. For example, Graham et al. (2018) studies the positive impacts of downstream environmental logistics practices on both the environment and profits of firms, and the positive effects will increase when the firms cooperate with customers to solve environmental issues. Focusing on the inventory control systems, Tang et al. (2015) evaluates the potential of CO_2 emissions reduction and the influences on total cost in logistics by reducing shipment frequency.

2.2 Green efficiency evaluation

The data envelopment analysis (DEA) is a powerful methodology to estimate environmental and energy efficiency (Lin and Du 2015) and has been widely adopted in prior studies to measure green development performance. As a kind of nonparametric methods, the DEA method usually adopts the linear programming models to estimate the best-performance frontiers, and the distance from the frontiers is recognized as the inefficiency (Lin and Zhu 2019a). The conventional DEA methods usually base on the Shephard distance function with an unreasonable assumption that the input and outputs are changed in the same proportion, which leads to certain limitations (Zhang and Choi 2014). The directional distance function (DDF) method proposed by Chung et al. (1997) can solve this issue properly. However, the DDF method assumes that the inputs and outputs change proportionally. Specifically, it assumes that the reduction in inputs and undesirable outputs, as well as the growth of the desirable outputs, should change at the same rate. Therefore, the DDF method may underestimate the efficiency. Different from the DDF method, the NDDF method proposed by Zhou et al. (2012) has a relaxed assumption that the input and output factors can have disproportional adjustments. Therefore, the NDDF method exhibits higher discriminating power (Lin and Du 2015). In this paper, we adopt the NDDF method to evaluate the green development efficiency of the logistics industry.

2.3 Clustering analysis

Clustering is a well-known unsupervised machine learning method that can group a large number of observations into several clusters. The features of different clusters have discrepancies, while the individuals in each subset are similar as much as possible. Clustering analysis can be applied in many fields such as biomedical, marketing, behavioral science, and social sciences (Marinakis et al. 2011; Rajabi et al. 2020). So far, many clustering methods have been proposed due to different merits, such as K-means, hierarchical classification, and model-based clustering method (Hartigan and Wong 1979; Celeux and Govaert 1995; Jacques and Preda 2014). In this paper, we try to identify the heterogeneity and similarity of green efficiency patterns of the logistics industry among different Chinese provinces with the clustering method. The clustering analysis in this paper is conducted with a recently popular method, namely the functional high dimensional data clustering (funHDDC) proposed by Bouveyron and Jacques (2011). The funHDDC method, which is based on the model-based clustering, can cluster the time series data (or more generally functional data) in group-specific subspaces of low dimensionality. The comparative results in the research of Bouveyron and Jacques (2011) show that the funHDDC method performs better than the functional clustering method fclust proposed by James and Sugar (2003), and appears to be more stable than the two-step clustering methods such as HDDC (Bouveyron et al. 2007) and MixtPPCA (Tipping and Bishop 1999). The funHDDC method has already been applied in some previous studies. For example, to forecast electricity demand accurately, Martínez-Álvarez et al. (2019) adopts the funHDDC method to find the patterns of the historical data, and then predict the future electricity demand based on the assumption of pattern sequence similarity. Based on the funHDDC method, Leroy et al. (2018) focuses on the improvement patterns of promising young swimmers, and the results show that the fastest progress of young swimmers usually occurs before the age of 16. And Bouveyron et al. (2015) tries to figure out the operating patterns in the bike-sharing systems based on several functional clustering methods such as funFEM, fclust, and funHDDC.

2.4 Innovation of this study

Evaluating the changes in the green efficiency of the logistics industry is crucial, and most of the literature does not discuss this topic from the aspect of heterogeneity and similarity of green efficiency patterns in different regions. Therefore, this study attempts to fill the literature gap. Specifically, based on the NDDF method, we first evaluate the green efficiency performance in the logistics industry of different Chinese provinces, then we adopt the functional clustering method funHDDC to reveal the regional heterogeneity and similarity of the green efficiency patterns among different provinces. We further analyze the driving factors of dynamic changes in green efficiency of the logistics industry.

3. Methods and data

3.1 Non-directional distance function

Suppose that in the production process, each province is regarded as a Decision-Making Unit (DMU). The provinces use the inputs include capital (*K*), labor (*L*) and energy (*E*) to produce desirable output *(Y)* and undesirable output CO_2 emissions (*C*). Thus, the production technology can be specified as:

$$T = \{(K, L, E, Y, C): K, L, E \text{ } can\ produce\ Y\ and\ C\}$$

$$T = \left\{ \begin{array}{c} (K,L,E,Y,C): \sum_{i=1}^{N}\sum_{t=1}^{T}\tau_{it}L_{it} \le L\,, \sum_{i=1}^{N}\sum_{t=1}^{T}\tau_{it}K_{it} \le K\,, \\ \sum_{i=1}^{N}\sum_{t=1}^{T}\tau_{it}E_{it} \le E\,, \sum_{i=1}^{N}\sum_{t=1}^{T}\tau_{it}Y_{it} \ge Y\,, \sum_{i=1}^{N}\sum_{t=1}^{T}\tau_{it}C_{it} = C. \\ \tau_{it} \ge 0;\ i = 1,2,\dots,N;\ \ t = 1,2,\dots,T. \end{array} \right\} \tag{1}$$

T is assumed as a close and bounded technology set in the production theory, which means that only finite outputs can be generated by finite inputs. Besides, the following two standard axioms are also assumed to be satisfied: (1) Inputs and desirable outputs supposed to be strongly or freely disposable; (2) There exists the possibility of inactivity. Furthermore, the weak-disposability and null-jointness assumptions proposed by Färe et al. (1989) need to be imposed. To be specific, the weak-disposability assumption implies that the reduction in undesirable outputs will lead to the opportunity cost calculated by the proportional decrease in desirable outputs. That is to say, for any $\eta \in [0,1]$, if $(K, L, E, Y, C) \in T$, then $(K, L, E, \eta Y, \eta C) \in T$. The null-jointness assumption implies that some undesirable outputs will be generated during the process of producing some desirable outputs. That is, if $C = 0$, and $(K, L, E, Y, C) \in T$, then $Y = 0$. Following Zhou et al. (2012), Lin and Du (2015), and Lin and Zhu (2019a), the NDDF can be defined as:

$$\overrightarrow{ND}(K, L, E, Y, C; g) = sup\{w^T\beta : [(K, L, E, Y, C) + diag(\beta) \cdot g] \in T\} \tag{2}$$

The slack vector $\beta = (\beta_K, \beta_L, \beta_E, \beta_Y, \beta_C)^T \geq 0$ measures the individual inefficiency for inputs and outputs. $diag$ is the diagonal matrix. The normalized weight vector $w = (w_K, w_L, w_E, w_Y, w_C)^T$ is assigned to input and output variables respectively. The directional vector $g = (g_K, g_L, g_E, g_Y, g_C)^T$ determines the scaled directions for the input and output variables. Therefore, Eq. (2) can be specified as the total-factors NDDF function:

$$\begin{aligned}
\overrightarrow{ND}_T(K, L, E, Y, C; g) = \max w_k\beta_k + w_L\beta_L + w_E\beta_E + w_Y\beta_Y + w_C\beta_C \\
\text{s.t. } \textstyle\sum_{i=1}^{N} \tau_i K_i \leq K - \beta_K g_K \\
\textstyle\sum_{i=1}^{N} \tau_i L_i \leq L - \beta_L g_L \\
\textstyle\sum_{i=1}^{N} \tau_i E_i \leq E - \beta_E g_E \\
\textstyle\sum_{i=1}^{N} \tau_i Y_i \geq Y + \beta_Y g_Y \\
\textstyle\sum_{i=1}^{N} \tau_i C_i = C - \beta_C g_C \\
\tau_{it} \geq 0;\ i = 1, 2, \ldots, N. \\
\beta_K, \beta_L, \beta_E, \beta_Y, \beta_C \geq 0
\end{aligned} \tag{3}$$

The vector w and g can be set differently according to requirements. To calculate the total-factor unified efficiency (TFUE) of the logistics industry, we assume an equal weight to inputs, desirable outputs, and undesirable outputs, respectively. Thus, the weight vector is defined as $w = \left(\frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{3}, \frac{1}{3}\right)^T$. Besides, to decrease inputs and undesirable output as well as increase desirable output simultaneously, we define the directional vector as $g = (-K, -L, -E, Y, -C\ )^T$. Then the TFUE can be formulated as:

$$TFUE = \frac{\frac{1}{4}[(1-\beta_K^*)+(1-\beta_L^*)+(1-\beta_E^*)+(1-\beta_C^*)]}{1+\beta_Y^*} \quad (4)$$

Using global environmental DEA to optimize Eq. (3) can obtain the optimal solutions $\beta^* = (\beta_K^*, \beta_L^*, \beta_E^*, \beta_Y^*, \beta_C^*)^T$. The TFUE considers all the changes in production efficiency, energy utilization and environmental improvement. Therefore, it is a good indicator to reflect the green performance of the logistics industry (Lin and Zhu 2019b). The TFUE lies between zero and unity and a higher TFUE corresponds to a better performance in the logistics industry.

3.2 Functional clustering approach

In this paper, we try to make clustering analysis on different Chinese provinces based on the time series of efficiency scores in the logistics industry. The clustering approach funHDDC proposed by Bouveyron and Jacques (2011) can group functional data into different clusters based on the model-based clustering method. By employing this approach, the discrete observations can be expanded by the basis functions, and the mixture of Gaussian distribution assumption can be applied to the associated coefficient vectors to get the final clustering results.

Step 1: Basis expansion.

Suppose that the observed curves $\{x_1, \dots, x_n\}$ belong to $L_2[0, T]$, and the functional forms of them are difficult to know. To construct the functional forms based on the discrete observations, the observed curves are usually expanded by some basis functions under the assumption that they belong to finite-dimensional space:

$$X(t) = \sum_{j=1}^{p} \gamma_j(X)\psi_j(t) \quad (5)$$

Where, $x_{ij} = x_i(t_{ij})$ is the finite discrete observation on original curves. $\{\psi_1, \dots, \psi_j, \dots, \psi_p\}$ is a basis of functions and $\{\gamma_1, \dots, \gamma_j, \dots, \gamma_p\}$ are random coefficient vectors.

Step 2: Coefficient vectors modeling.

Coefficient vectors are modeling by a latent mixture model. Suppose that a set of n_k observed curves are combined into the kth cluster, and the coefficient vectors $\{\gamma_1, \dots, \gamma_{nk}\}$ are independent realizations of vector Γ. Besides, the actual stochastic process related to the kth cluster is reflected on the low-dimensional functional latent subspace $\mathrm{E}_k[0,T]$ of $L_2[0,T]$, which has d_k dimensions. Let $\mathrm{E}_k[0,T]$ be spanned by the first d_k elements of the kth basis of functions $\{\varphi_{k1},\dots,\varphi_{kd_k}\}$, and $\varphi_{kj} = \sum_{l=1}^{p} q_{k,jl}\psi_l$. $Q_k = (q_{k,jl}) = [U_k, V_k]$ is an orthogonal $p \times p$ matrix. Let $\{\lambda_1,\dots,\lambda_{nk}\}$ be the latent expansion curve coefficients with basis $\{\varphi_{k1},\dots,\varphi_{kd_k}\}$, which are independent realizations of the latent vector Λ. And we can find that:

$$\Gamma = U_k\Lambda + \varepsilon \tag{6}$$

Where ε is the independent random noise. Suppose that Λ and ε both follow the multivariate Gaussian distribution. Therefore, we have:

$$\Gamma \sim \mathcal{N}(\mu_k, \Sigma_k) \tag{7}$$

Let the unobserved random variable $Z = (Z_1,\dots,Z_k)$ be the indicator variable of the group membership. Z_k equals to 1 if X is in the kth cluster and 0 otherwise. And suppose $\pi_k = P(Z_k = 1)$.

The coefficient vectors γ follow the mixture of Gaussian distribution and the density function is:

$$p(\gamma) = \sum_{k=1}^{K} \pi_k \phi(\gamma; \mu_k, \Sigma_k) \tag{8}$$

Where ϕ is the density function of standard Gaussian distribution.

Step 3: Model parameters estimation.

Bouveyron and Jacques (2011) estimates parameters θ by MLE through the EM algorithm. The log-likelihood of the data under the functional latent mixture (FLM) model is:

$$\begin{aligned} l_c(\theta; \gamma_1,\dots,\gamma_n; z_1,\dots,z_n) \\ = -\frac{1}{2}\sum_{k=1}^{K}\eta_k\Bigg[\sum_{j=1}^{d_k}\left(\log(a_{kj}) + \frac{q_{kj}^t C_k q_{kj}}{a_{kj}}\right) \\ + \sum_{j=d_k+1}^{p}\left(\log(b_k) + \frac{q_{kj}^t C_k q_{kj}}{b_k}\right) - 2\log(\pi_k)\Bigg] + \xi \end{aligned} \tag{9}$$

The hyper-parameters d_k and K cannot be estimated by MLE. Actually, the class specific dimension d_k is estimated by the scree-test (Cattell 1966) and the number of clusters K is chosen by BIC.

3.3 Data description

3.3.1 Input variables

(1) Labor (*L*). In this paper, the labor factor is represented by the total employees of the logistics industry. The data is collected from the China National Bureau of Statistics.

(2) Capital (*K*). Usually, there are two indicators to measure the capital factors, namely fixed asset investment in each year and capital stock calculated by the perpetual inventory method. Following Yang et al. (2019), this paper adopts the fixed asset investment of the logistics industry to represent the capital input. The data is collected from the China National Bureau of Statistics.

(3) Energy (*E*). The different types of energy consumption in the logistics industry are collected from the CEIC.

3.3.2 Output variables

(1) Desirable output (Y). This paper uses the added value of the logistics industry to represent the desirable output. Relevant data is collected from the China National Bureau of Statistics.

(2) Undesirable output (C). Taking into account the environmental issue, this paper uses the CO_2 emissions of the logistics industry to represent the undesirable output. The total CO_2 emissions are calculated based on the energy consumption of the logistics industry and the emission coefficients provided by IPCC (2006).

Considering that the major production of the logistics industry is concentrated in transportation, storage, and postal industries, this paper uses the data of these three sub-industries to represent the logistics industry (Yang et al. 2019). The dataset only includes 30 provinces of China over the period of 2008 to 2017 because of data limitation, and the raw data is deflated in the 2008 constant price in the analysis.

4. Results

4.1 Total-factor unified efficiency

The specific TFUE results of the logistics industry in different Chinese provinces are shown in Table 1. According to the definition of the TFUE, the closer the value is to 1, the better the performance of green efficiency. Moreover, when the value of TFUE equals to 1, the unit is suggested to display the best performance of green efficiency and locate on the frontier. Overall, we can find that the TFUE scores fluctuated in the study period and all the provinces obtain relatively high green efficiency values, with values large than 0.65. The worst performing region is Yunnan, which gets the lowest efficiency scores among 30 regions, with values less than 0.70. By calculating the average value of the TFUE for all provinces in different years, we can observe that the mean values of the TFUE are relatively stable, and seem to be close to 0.85 in these ten years. We also find that the average value in 2017 is only slightly higher than the value in 2016 and even lower than the mean in 2008, which shows that the TFUE scores have not been effectively improved from the perspective of the whole of China. Therefore, how to better improve the green efficiency of China's logistics industry is an issue worthy of attention.

[Insert Table 1 here]

To more clearly show the changes in the green efficiency performance of different provinces over the past decade, we display the TFUE scores in 2008 and 2017 respectively in Fig.1. Fig.1 indicates that half of the provinces have witnessed an increase in the TFUE scores. In specific, Hebei appears to be the best performing province with the efficiency score to be 1 in 2008 and 2017 respectively, suggesting that Hebei lies in the frontier in these two years. And Yunnan appears to be the worst performing province among these regions. Besides, we also find that in some provinces, the green efficiency value has dropped significantly, such as Gansu and Qinghai. Therefore, it is very meaningful to study the similarities and differences of the green efficiency patterns in the logistics industry of different provinces.

[Insert Fig. 1 here]

4.2 Functional clustering results

Firstly, we should choose the suitable cluster number K and the appropriate FLM model. The funHDDC algorithm is allowed to test one partition or multiple partitions at the same time, and we initially let K equal to 2, 3, and 4 respectively. Besides, there are six FLM models ($A_{kj}B_kQ_kD_k$, $A_{kj}BQ_kD_k$, $A_kB_kQ_kD_k$, $AB_kQ_kD_k$, $A_kBQ_kD_k$, ABQ_kD_k) can be chosen, and we can also test multiple models simultaneously. So we have 18 options and we use the kmeans to initialize the E-M algorithm. According to the results of BIC, we finally select the model ABQ_kD_k with 4 clusters, and the specific result of each combination can be seen in Table 2.

[Insert Table 2 here]

Fig.2 shows that the green efficiency TFUE scores in the logistics industry of 30 provinces are eventually divided into 4 groups due to the method of funHDDC. In general, the TFUE of group 1 (black solid curves) shows a rapid downward trend, the TFUE of group 2 (red dashed curves) is at a medium level, the TFUE of group 3 (green dotted curves) is at a low level, and the TFUE of group 4 (blue dot-dashed curves) is at a high level. Compared with the original discretized time series data, the curves in Fig.2 have become smoother after clustering. This is because the discrete time series data of each province is approximated as a smooth functional curve through the FLM model.

[Insert Fig. 2 here]

To analyze the different clusters' change characteristics of green efficiency values more intuitively, we further obtain the estimated mean function of each cluster which can be seen in Fig.3. From Fig.3 we can find that the average green efficiency values in the logistics industry of these four groups show obvious differences. The mean function of group 1 (black solid curves) decreases rapidly, from about 0.95 at the beginning to less than 0.8. However, the mean function of group 2 (red dashed curves) fluctuates in the range of 0.85-0.95 and shows a rapid growth trend recently. Similar to group 2, the mean function of group 3 (green dotted curves) are also steadily increasing. But the mean function of group 3 is much lower than that of group 2, which fluctuates in the range of 0.75-0.85. Besides, the mean function of group 4 (blue dot-dashed curves) is relatively stable, with mean values change in the range of 0.95-1.

[Insert Fig. 3 and Table 3 here]

From Table 3, we can know the specific clustering results of 30 provinces in China. For the three provinces (Anhui, Gansu, and Qinghai) in group 1, the decline in the green efficiency scores of the logistics industry is very large from 2008 to 2017. In 2008, the green efficiency value of these three provinces played a leading role. But the green efficiency of these three provinces is at a relatively low level in 2017 from the perspective of the whole of China. As shown in Fig.4, the green efficiency of the logistics industry in the 13 regions of group 2 is steadily increasing, and among them, the green efficiency of three regions is growing particularly fast. Besides, the increase in the green efficiency of these 13 regions tends to converge. There are 11 regions in group 3. Compared with the other three groups, the green efficiency of group 3 is at a lower level, but there is still a slight upward trend recently. Group 4 consists of two eastern provinces (Tianjin and Hebei) and one western province (Ningxia). It can be seen from Fig.4 that the green efficiency of these three provinces is at a relatively high level compared to other regions in China, although the green efficiency values fluctuate greatly.

From the results of clustering, we can find that the patterns of green efficiency change in the logistics industry are closely related to the geographical location of the region. Generally speaking, the green efficiency values of the second and fourth groups are relatively high, and most of these regions are concentrated in eastern China. However, the regions with lower green efficiency values of group 3 are mainly distributed in western China. Therefore, we can conclude that the average green efficiency performance of the eastern region is better than that of other regions, while the green efficiency of the western region performs poorly. But it can also be found that some of the eastern regions like Beijing and Shanghai belong to the third group, although the green efficiency of these regions plays a leading role in group 3. In the following section, we are going to analyze the dynamic changes in green efficiency performance and the influencing factors after decomposition.

[Insert Fig. 4 here]

5. Further analysis: dynamic changes in green efficiency performance

5.1 Metafrontier non-radial Malmquist CO_2 emission performance index

The TFUE constructed in Section 3.1 is a static index that aims to describe the green efficiency performance in the logistics industry. Zhang and Choi (2013) further proposes the construction and decomposition of the Metafrontier non-radial Malmquist CO_2 emission performance index (MNMCPI) to study the dynamic changes in efficiency performance over time. Following Zhang and Choi (2013), we make further analysis of the dynamic changes in the green efficiency performance of the logistics industry.

To construct and decompose MNMCPI, three production technology sets (the contemporaneous production technology T_h^C, the intertemporal production technology T_h^I, and the global production technology T_h^G) are first to be defined. T_h^C represents the production technology for a particular group h at a specific period t.

$$T_h^C = \{(K^t, L^t, E^t, Y^t, C^t): K^t, L^t, E^t \; can \; produce \; Y^t \; and \; C^t; t = 1, \ldots, T\} \quad (10)$$

Correspondingly, the NDDF described in Eq. (2) can be rewritten as:

$$\overrightarrow{ND}^C(K, L, E, Y, C; g) = sup\{w^T \beta^c: [(K, L, E, Y, C) + diag(\beta^c) \cdot g] \in T_h^C\} \quad (11)$$

What's more, the intertemporal production technology T_h^I are defined as $T_h^I = T_h^1 \cup T_h^2 \cup \ldots \cup T_h^T$, which means the technology constructed from group h in the whole period. The global production technology T_h^G represents the technology constructed from all groups in the whole period, and $T^G = T_1^I \cup T_2^I \cup \ldots \cup T_H^I$.

And we can solve the NDDFs under different production technologies with the following model:

$$\begin{aligned} \overrightarrow{ND}_T^{\,d}(K^s, L^s, E^s, Y^s, C^s; g) &= \max w_k\beta_k^d + w_L\beta_L^d + w_E\beta_E^d + w_Y\beta_Y^d + w_C\beta_C^d \\ \text{s.t. } & \textstyle\sum_{set} \tau_i^s K_i^s \le K - \beta_k^d g_K \\ & \textstyle\sum_{set} \tau_i^s L_i^s \le L - \beta_L^d g_L \\ & \textstyle\sum_{set} \tau_i^s E_i^s \le E - \beta_E^d g_E \\ & \textstyle\sum_{set} \tau_i^s Y_i^s \ge Y + \beta_Y^d g_Y \\ & \textstyle\sum_{set} \tau_i^s C_i^s = C - \beta_C^d g_C \\ & \tau_i^s \ge 0 \\ & \beta_k^d, \beta_L^d, \beta_E^d, \beta_Y^d, \beta_C^d \ge 0 \end{aligned} \quad (12)$$

Where, according to different production technologies, we have:

$$\begin{aligned} & T_h^C: d \equiv C, set \equiv \{i \in h\} \\ & T_h^I: d \equiv I, set \equiv \{i \in h, s \in [1,2,..,T]\} \\ & T_h^G: d \equiv G, set \equiv \{i \in [1,2,\ldots,H], s \in [1,2,..,T]\} \end{aligned} \quad (13)$$

Therefore, the definition of MNMCPI is shown as follows:

$$MNMCPI(K^s, L^s, E^s, Y^s, C^s) = \frac{TFUE^G(K^{t+1}, L^{t+1}, E^{t+1}, Y^{t+1}, C^{t+1})}{TFUE^G(K^t, L^t, E^t, Y^t, C^t)} \quad (14)$$

Therefore, the changes of the TFUE from time t to t+1 can be captured by the MNMCPI index. Furthermore, to analyze the driving factors of changes in green efficiency, the MNMCPI index is decomposed into three parts. And the decomposition of MNMCPI is shown as follows:

$$
\begin{aligned}
& MNMCPI(K^s, L^s, E^s, Y^s, C^s) = \frac{TFUE^G(.^{t+1})}{TFUE^G(.^t)} \\
& = \left[\frac{TFUE^C(.^{t+1})}{TFUE^C(.^t)}\right] \times \left[\frac{TFUE^I(.^{t+1})/TFUE^C(.^{t+1})}{TFUE^I(.^t)/TFUE^C(.^t)}\right] \times \left[\frac{TFUE^G(.^{t+1})/TFUE^I(.^{t+1})}{TFUE^G(.^t)/TFUE^I(.^t)}\right] \\
& = \left[\frac{TE^{t+1}}{TE^t}\right] \times \left[\frac{BPR^{t+1}}{BPR^t}\right] \times \left[\frac{TGR^{t+1}}{TGR^t}\right] \\
& = EC \times BPC \times TGC
\end{aligned} \tag{15}
$$

Where the efficiency change (EC) term shows the catch-up effect under the contemporaneous production technology for group h from t to t+1. The best-practice gap change (BPC) term reflects the innovation effect between the contemporaneous and the intertemporal production technology during two periods. And the technology leadership change has been illustrated in the technology gap change (TGC) term because it measures the ratio change of the technology gap between the intertemporal and the global frontier from t to t+1.

5.2 Clustering results of the three decomposition components

5.2.1 Catch-up effect: clustering analysis of the EC term

When the value of the EC is bigger than 1, it indicates a fact of green efficiency improvement. And conversely, efficiency decreases. From Fig.5, we can find that the values of the EC term fluctuate greatly in the whole period, which means that the green efficiency performance of the logistics industry is not always in a better state. The specific clustering results are shown in Fig.6. According to the clustering results, 30 provinces can be divided into two groups. The first group (black solid curves) has a wavy shape, and the second group (red dashed curves) performs much more stable than the first group, with values going above or below 1. We can further divide the changing trend of the EC term in group 1 into four stages based on whether the average EC value is greater than 1. It shows that in the first and third stages, the seven provinces in group 1 move toward the contemporaneously green efficiency technology frontier, which suggests the existence of the positive catch-up effect in group 1. However, in the second and the fourth stage, these provinces perform poorly. In addition, we find that Beijing and Shanghai belong to group 1, which shows that the green efficiency improvement of the logistics industry in Beijing and Shanghai shows fluctuations, and has a downward trend in recent years. This may be the reason why the green efficiency values of the logistics industry in Beijing and Shanghai are not high.

[Insert Fig. 5, Fig. 6 and Table 4 here]

5.2.2 Innovation effect: clustering analysis of the BPC term

When the value of the BPC term is larger than 1, the contemporaneous technology frontier is closer to the intertemporal technology frontier, and the innovation effect exists. Instead, the contemporaneous technology frontier is farther away from the intertemporal technology frontier. From Fig.7 we can know that, similar to the EC term, the value of the BPC term also varies greatly. And it shows an overall upward trend which reflects the technological innovation improvement in the green efficiency performance of the logistics industry. Fig.8 shows that there are four clusters based on the results of funHDDC. Among them, both the BPC terms of group 1 (black solid curves) and group 2 (red dashed curves) show a clear upward trend. The average value of the BPC term in group 1 rises rapidly, and then gradually rises after a period of decline. And the average value of the BPC term in group 2 starts to rise rapidly after a period of fluctuations. From Table 5 we can know that Beijing and Shanghai belong to the second group, which indicates that these two cities have made more efforts to promote technological innovation in the logistics industry in recent years. Besides, the average values of the BPC term in group 3 (green dotted curves) and group 4 (blue dot-dashed curves) are slowly rising as well as fluctuating, and the changes in these two groups always show opposite trends.

[Insert Fig. 7, Fig. 8 and Table 5 here]

5.2.3 Technology leadership change: clustering analysis of the TGC term

If the value of the TGC term is larger than 1, the technology gap between intertemporal technology and global technology will be narrowing. Otherwise, the technology gap will widen. From Fig.9 we can find that the value of TGC does not change much compared to the variations of EC and BPC. Besides, from the results of clustering, we can see that the value changes of the TGC term in 30 provinces can be divided into two groups. There exists obvious volatility in the average value of the TGC term in group 1 (black solid curves). However, the average value of the TGC term in group 2 (red dashed curves) has smaller changes, which almost equals to 1. Furthermore, from Table 6 we can know that Beijing and Shanghai belong to the second group, which suggests that these two regions don’t achieve the target of becoming the technical leadership in the logistics industry, even though their economic levels are high. From 2008 to 2017, the mean values of the TGC term have increased slightly, which means that the gap between the intertemporal and the global frontier still exists.

[Insert Fig. 9, Fig. 10 and Table 6 here]

6. Discussion

From 2008 to 2017, the green efficiency performance of the logistics industry in most Chinese provinces has been steadily improving. Based on the green efficiency scores calculated by the NDDF method and the clustering results of the funHDDC method, we can conclude that there exists a certain correlation between the green efficiency of the logistics industry and the geographical location. On the whole, the green efficiency values of most provinces in the eastern, northeastern and central regions are at the middle-upper level in China, while the green efficiency values of most provinces in the western regions are at the lower level. Moreover, the clustering results show that the green efficiency performance of the logistics industry varies greatly in different regions. The 30 provinces can be divided into 4 clusters, namely a high level of green efficiency, a medium level of green efficiency, a low level of green efficiency, and a downward trend of green efficiency. Specifically, the green efficiency performance of the logistics industry in Tianjin, Hebei, and Ningxia plays a leading role in China, while the green efficiency performance of Anhui, Gansu, and Qinghai shows a significant downward trend comparing to other regions.

The development of the logistics industry needs the support of economic development. On the one hand, the regions with high economic levels usually show a higher demand for logistics services, which can stimulate the growth and expansion of the logistics industry. On the other hand, the expansion of the logistics industry can effectively promote the output growth of other industries. However, the green efficiency calculated by the NDDF method in this paper is a comprehensive consideration of desirable and undesirable output, which indicates that the impact on the environment should also be considered. Therefore, we cannot completely conclude that the higher the regional economic level, the higher the efficiency of the logistics industry. Similar results can be found in other industries, such as the construction industry (Zhou et al. 2019) and the iron and steel industry (Lin and Wu 2020). In the clustering results of this paper, we find that the green efficiency performance of the logistics industry in Beijing and Shanghai belongs to the third group, which is at a low level comparing to the green efficiency performance of the logistics industry in other regions. The poor green performance of the logistics industry in Beijing and Shanghai may be caused by various problems in city logistics development. For example, large demands for the logistics services may have a negative impact on efficiency due to the high population density and congested traffic (Shao et al. 2019; Firdausiyah et al. 2019). Therefore, queuing and congestion will lead to increased energy consumption and carbon dioxide emissions, and reduce the green efficiency in the logistics. Moreover, based on the further analysis of the decomposition of dynamic changes in green efficiency, we find that the efficiency improvements of the logistics industry in Beijing and Shanghai are not obvious, and these two cities lack technical leadership even though the technological innovation has been promoted in recent years. Therefore, the green efficiency performance of the logistics industry in Beijing and Shanghai is not satisfying.

7. Conclusion and policy recommendations

The logistics industry is an indispensable part of the social economy. At the same time, it also leads to large amounts of resource consumption and environmental pollution. Therefore, the high-quality development of the logistics industry needs to achieve the goal of green transformation. This article attempts to analyze the changing patterns and influencing factors of green efficiency in the Chinese logistics industry. We first use the NDDF method to calculate the green efficiency scores (TFUE) of the logistics industry in China's 30 provinces from 2008 to 2017. Then, based on the functional clustering method funHDDC, we conduct the clustering analysis on the green efficiency performance of the 30 provinces. Finally, the dynamic changes of green efficiency performance have been divided into three driving factors: the EC term, the BPC term, and the TGC term. The clustering results show that the green efficiency of China's logistics industry can be divided into four clusters, which are high level, medium level, low level, and downward trend respectively. By analyzing the similarities and differences of different clusters, we have proposed the following policy recommendations.

(1) The green efficiency evaluation of the logistics industry is not only related to the level of economic development, but also to environmental performance. Although the green efficiency of the logistics industry in eastern China is generally better than that in other regions, we still find that the green efficiency performance of some developed regions is not satisfactory, such as Beijing and Shanghai. Therefore, to achieve the goal of sustainable development in the regions with high economic level, the government should make more efforts to promote the level of management in the logistics and improve green efficiency. From the perspective of transportation, transportation vehicles with high pollution emissions should be phased out gradually, and transportation routes should be rationally planned. And promoting the development of sharing economy in the logistics can improve the efficiency of transportation. From the perspective of warehousing, the construction of urban and rural logistics warehousing centers should be accelerated to reduce the logistics pressure in the central cities.

(2) According to the clustering results, it can be seen that the logistics industry's green efficiency of most regions in adjacent geographical locations has a similar development pattern. Therefore, when the government formulates the policy for the logistics industry, it is important to carry out unified strategic planning for these adjacent areas. Considering the overall development requirements of these regions can not only stimulate the development potential of the logistics industry in each province, but also promote the coordinated development level among different provinces.

(3) From the results of the decomposition, we can see that technology innovation can be promising to stimulate the green efficiency performance of China's logistics industry. Therefore, the government should focus on promoting the transformation of logistics to the high-tech industry and effectively improving the innovation level of the logistics. For example, the government should stimulate the R&D activities in the logistics industry, strengthen the application of artificial intelligence and big data in the logistics. Moreover, the optimal operation management and allocation of resources should also be stressed.

Due to the data limitation, this paper analyzes the green efficiency performance of the logistics industry from the provincial level in China. And the study of city logistics can better reflect the development characteristics of the logistics industry in different areas. Therefore, how to analyze the green efficiency performance of the logistics industry and its dynamic changes from the perspective of city logistics is an issue worthy of attention in future research.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant numbers 71874149, 71934001, 71471001, 41771568, 71533004); the National Social Science Fund of China (grant number 20ZDA084); and the National Key Research and Development Program of China (grant number 2016YFA0602500).

Declaration of interest: None.

References

Ahmed, W., Najmi, A., Khan, F., 2019. Examining the impact of institutional pressures and green supply chain management practices on firm performance. Management of Environmental Quality. https://doi.org/10.1108/MEQ-06-2019-0115

Aldakhil, A.M., Nassani, A.A., Awan, U., Abro, M.M.Q., Zaman, K., 2018. Determinants of green logistics in BRICS countries: An integrated supply chain model for green business. Journal of Cleaner Production. 195, 861-868.

Ameknassi, L., Aït-Kadi, D., Rezg, N., 2016. Integration of logistics outsourcing decisions in a green supply chain design: A stochastic multi-objective multi-period multi-product programming model. International Journal of Production Economics. 182, 165-184.

Bagheri, M., Guevara, Z., Alikarami, M., Kennedy, C.A., Doluweera, G., 2018. Green growth planning: A multi-factor energy input-output analysis of the Canadian economy. Energy Economics. 74, 708-720.

Berry, S., Davidson, K., Saman, W., 2013. The impact of niche green developments in transforming the building sector: The case study of Lochiel Park. Energy Policy. 62, 646-655.

Bouveyron, C., Côme, E., Jacques, J., 2015. THE DISCRIMINATIVE FUNCTIONAL MIXTURE MODEL FOR A COMPARATIVE ANALYSIS OF BIKE SHARING SYSTEMS. The Annals of Applied Statistics. 9(4), 1726-1760.

Bouveyron, C., Girard, S., Schmid, C., 2007. High dimensional data clustering. Computational Stats & Data Analysis. 52, 502-519.

Bouveyron, C., Jacques, J., 2011. Model-based clustering of time series in group-specific functional subspaces. Advances in Data Analysis and Classification. 5, 281-300.

Cattell, R.B., 1966. The scree test for the number of factors. Multivariate behavioral research. 1(2), 245-276.

Celeux, G., Govaert, G., 1995. Gaussian parsimonious clustering models. Pattern recognition. 28(5), 781-793.

Chen, D., Ignatius, J., Sun, D., Zhan, S., Zhou, C., Marra, M., Demirbag, M., 2019. Reverse logistics pricing strategy for a green supply chain: A view of customers' environmental awareness. International Journal of Production Economics. 217, 197-210.

Chhabra, D., Garg, S.K., Singh, R.K., 2017. Analyzing alternatives for green logistics in an Indian automotive organization: A case study. Journal of Cleaner Production. 167, 962-969.

Chung, Y.H., Färe, R., Grosskopf, S., 1997. Productivity and undesirable outputs: a directional distance function approach. Journal of Environmental Management. 51(3), 229-240.

Dai, Y., Gao, H.O., 2016. Energy consumption in China's logistics industry: A decomposition analysis using the LMDI approach. Transportation Research Part D: Transport and Environment. 46, 69-80.

Davarzani, H., Fahimnia, B., Bell, M., Sarkis, J., 2016. Greening ports and maritime logistics: A review. Transportation Research Part D. 48, 473-487.

Dekker, R., Bloemhof, J., Mallidis, I., 2012. Operations Research for green logistics - An overview of aspects, issues, contributions and challenges. European Journal of Operational Research. 219, 671-679.

Demir, E., Hrušovský, M., Jammernegg, W., Van Woensel, T., 2019. Green intermodal freight transportation: bi-objective modelling and analysis. International Journal of Production Research. 57(19), 6162-6180.

Dev, N.K., Shankar, R., Swami, S., 2020. Diffusion of green products in industry 4.0: Reverse logistics issues during design of inventory and production planning system. International Journal of Production Economics. 223, 107519.

Färe, R., Grosskopf, S., Lovell, C.K., Pasurka, C., 1989. Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. The Review of Economics and Statistics. 90-98.

Firdausiyah, N., Taniguchi, E., Qureshi, A.G., 2019. Modeling city logistics using adaptive dynamic programming based multi-agent simulation. Transportation Research Part E. 125, 74-96.

Garza-Reyes, J.A., Villarreal, B., Kumar, V., Ruiz, P.M., 2016. Lean and green in the transport and logistics sector - a case study of simultaneous deployment. Production Planning & Control. 27(15), 1221-1232.

Goswami, M., De, A., Habibi, M.K.K., Daultani, Y., 2020. Examining freight performance of third-party logistics providers within the automotive industry in India: an environmental sustainability perspective. International Journal of Production Research. https://doi.org/10.1080/00207543.2020.1756504

Graham, S., Graham, B., Holt, D., 2018. The relationship between downstream environmental logistics practices and performance. International Journal of Production Economics. 196, 356-365.

Gu, W., Wei, L., Zhang, W., Yan, X., 2019. Evolutionary game analysis of cooperation between natural resource- and energy-intensive companies in reverse logistics operations. International Journal of Production Economics. 218, 159-169.

Hartigan, J. A., Wong, M. A., 1979. Algorithm AS 136: A k-means clustering algorithm. Applied Statistics. 28(1), 100-108.

Hong, J., Alzaman, C., Diabat, A., Bulgak, A., 2019. Sustainability dimensions and PM2.5 in supply chain logistics. Annals of Operations Research. 275, 339-366.

Hou, D., Li, G., Chen, D., Zhu, B., Hu, S., 2019. Evaluation and analysis on the green development of China's industrial parks using the long-tail effect model. Journal of Environmental Management. 248, 109288.

IPCC, 2006. Guidelines for National Greenhouse Gas Inventories. Intergovernmental Panel on Climate Change. Available at: https://www.ipcc-nggip.iges.or.jp/public/2006gl/index.html.

Jacques, J., Preda, C., 2014. Functional data clustering: a survey. Advances in Data Analysis and Classification. 8(3), 231-255.

James, G.M., Sugar, C.A., 2003. Clustering for sparsely sampled functional data. Journal of the American Statistical Association. 98(462), 397-408.

Ji, S.F., Luo, R.J., Peng, X.S., 2019. A probability guided evolutionary algorithm for multi-objective green express cabinet assignment in urban last-mile logistics. International Journal of Production Research. 57(11), 3382-3404.

Jiahuey, Y., Liu, Y., Yu, Y., 2019. Measuring green growth performance of China's chemical industry. Resources, Conservation and Recycling. 149, 160-167.

Jin, P., Peng, C., Song, M., 2019. Macroeconomic uncertainty, high-level innovation, and urban green development performance in China. China Economic Review. 55, 1-18.

Khor, K.S., Udin, Z.M., 2013. Reverse logistics in Malaysia: Investigating the effect of green product design and resource commitment. Resources, Conservation and Recycling. 81, 71-80.

Kuo, T.C., Chiu, M.C., Chung, W.H., Yang, T.I., 2019. The circular economy of LCD panel shipping in a packaging logistics system. Resources, Conservation & Recycling. 149, 435-444.

Leroy, A., Marc, A., Dupas, O., Rey, J.L., Gey, S., 2018. Functional Data Analysis in Sport Science: Example of Swimmers' Progression Curves Clustering. Applied Sciences. 8(10), 1766.

Li, A., Chen, Y., Wang, D., 2020. An empirical study of the factors influencing the willingness to implement green coal logistics in China. Journal of Cleaner Production. 245, 118932.

Li, W., Wang, J., Chen, R., Xi, Y., Liu, S.Q., Wu, F., Masoud, M., Wu, X., 2019. Innovation-driven industrial green development: The moderating role of regional factors. Journal of Cleaner Production. 222, 344-354.

Lin, B., Du, K., 2015. Energy and CO2 emissions performance in China's regional economies: do market-oriented reforms matter? Energy Policy. 78, 113-124.

Lin, B., Wu, R., 2020. Designing energy policy based on dynamic change in energy and carbon dioxide emission performance of China's iron and steel industry. Journal of

Cleaner Production. 256, 120412.

Lin, B., Zhu, J., 2019a. Impact of energy saving and emission reduction policy on urban sustainable development: Empirical evidence from China. Applied energy. 239, 12-22.

Lin, B., Zhu, J., 2019b. Fiscal spending and green economic growth: Evidence from China. Energy Economics. 83, 264-271.

Marinakis, Y., Marinaki, M., Doumpos, M., Matsatsinis, N., Zopounidis, C., 2011. A hybrid ACO-GRASP algorithm for clustering analysis. Annals of Operations Research. 188, 343-358.

Martínez-Álvarez, F., Schmutz, A., Asencio-Cortés, G., Jacques, J., 2019. A Novel Hybrid Algorithm to Forecast Functional Time Series Based on Pattern Sequence Similarity with Application to Electricity Demand. Energies. 12(1), 94.

Melkonyan, A., Gruchmann, T., Lohmar, F., Kamath, V., Spinler, S., 2020. Sustainability assessment of last-mile logistics and distribution strategies: The case of local food networks. International Journal of Production Economics. 228, 107746.

Oberhofer, P., Dieplinger, M., 2013. Sustainability in the Transport and Logistics Sector: Lacking Environmental Measures. Business Strategy and the Environment. 23(4), 236-253.

Rajabi, A., Eskandari, M., Ghadi, M.J., Li, L., Zhang, J., Siano, P., 2020. A comparative study of clustering techniques for electrical load pattern segmentation. Renewable and Sustainable Energy Reviews. 120, 109628.

Rashidi, K., Cullinane, K., 2019. Evaluating the sustainability of national logistics performance using Data Envelopment Analysis. Transport Policy. 74, 35-46.

Sbihi, A., Eglese, R.W., 2010. Combinatorial optimization and Green Logistics. Annals of Operations Research. 175, 159-175.

Shao, S., Xu, G., Li, M., Huang, G.Q., 2019. Synchronizing e-commerce city logistics with sliding time windows. Transportation Research Part E. 123, 17-28.

Shen, Z., Baležentis, T., Chen, X., Valdmanis, V., 2018. Green growth and structural change in Chinese agricultural sector during 1997–2014. China Economic Review. 51, 83-96.

Sureeyatanapas, P., Poophiukhok, P., Pathumnakul, S., 2018. Green initiatives for logistics service providers: An investigation of antecedent factors and the contributions to corporate goals. Journal of Cleaner Production. 191, 1-14.

Suzanne, E., Absi, N., Borodin, V., 2020. Towards circular economy in production planning: Challenges and opportunities. European Journal of Operational Research. 287, 168-190.

Tang, S., Wang, W., Yan, H., Hao, G., 2015. Low carbon logistics: Reducing shipment frequency to cut carbon emissions. International Journal of Production Economics. 164, 339-350.

Tipping, M.E., Bishop, C.M., 1999. Mixtures of probabilistic principal component analyzers. Neural Computation. 11(2), 443-482.

Yang, J., Tang, L., Mi, Z., Liu, S., Li, L., Zheng, J., 2019. Carbon emissions performance in logistics at the city level. Journal of Cleaner Production. 231, 1258-1266.

Yao, X., Cheng, Y., Song, M., 2019. Assessment of collaboration in city logistics: From the aspects of profit and CO2 emissions. International Journal of Logistics Research and Applications. 22(6), 576-591.

Yi, H., Liu, Y., 2015. Green economy in China: regional variations and policy drivers. Global Environmental Change. 31, 11–19.

Yuan, B.L., Xiang, Q.L., 2018. Environmental regulation, industrial innovation and green development of Chinese manufacturing: Based on an extended CDM model. Journal of Cleaner Production. 176, 895-908.

Zarbakhshnia, N., Soleimani, H., Goh, M., Razavi, S.S., 2019. A novel multi-objective model for green forward and reverse logistics network design. Journal of Cleaner Production. 208, 1304-1316.

Zhang, N., Choi, Y., 2013. Total-factor carbon emission performance of fossil fuel power plants in China: A metafrontier non-radial Malmquist index analysis. Energy Economics. 40, 549-559.

Zhang, N., Choi, Y., 2014. A note on the evolution of directional distance function and its development in energy and environmental studies 1997-2013. Renewable and Sustainable Energy Reviews. 33, 50-59.

Zhang, Y., Alshraideh, H., Diabat, A., 2018. A stochastic reverse logistics production routing model with environmental considerations. Annals of Operations Research. 271, 1023-1044.

Zhou, P., Ang, B. W., Wang, H., 2012. Energy and CO2 emission performance in electricity generation: a non-radial directional distance function approach. European Journal of Operational Research. 221(3), 625-635.

Zhou, Y., Liu, W., Lv, X., Chen, X., Shen, M., 2019. Investigating interior driving factors and cross-industrial linkages of carbon emission efficiency in China's construction industry: Based on Super-SBM DEA and GVAR model. Journal of Cleaner Production. 241, 118322.

Zhu, B., Zhang, M., Huang, L., Wang, P., Su, B., Wei, Y. M., 2020. Exploring the effect of carbon trading mechanism on China's green development efficiency: A novel integrated approach. Energy Economics. 85, 104601.

List of tables:

Table 1: Total-factor unified efficiency of the logistics industry in different provinces

Table 2: Results of Bayesian Information Criterion

Table 3: Specific clustering results of 30 provinces

Table 4: Specific grouping results for EC

Table 5: Specific grouping results for BPC

Table 6: Specific grouping results for TGC

Table 1 Total-factor unified efficiency of the logistics industry in different provinces

Region	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Beijing	0.777	0.786	0.809	0.823	0.807	0.813	0.810	0.814	0.813	0.811
Tianjin	0.931	0.928	0.943	0.952	0.935	0.980	1.000	0.978	0.976	1.000
Hebei	1.000	0.985	0.970	0.979	0.946	0.982	1.000	1.000	0.974	1.000
Shanxi	0.946	0.830	0.844	0.849	0.803	0.840	0.852	0.871	0.874	1.000
Inner Mongolia	0.866	0.852	0.848	0.860	0.846	0.871	0.890	0.871	0.920	0.941
Liaoning	0.772	0.774	0.777	0.797	0.802	0.810	0.808	0.837	0.879	1.000
Jilin	0.808	0.809	0.805	0.815	0.823	0.786	0.783	0.785	0.786	0.788
Heilongjiang	0.839	0.820	0.828	0.799	0.803	0.800	0.811	0.826	0.835	0.851
Shanghai	0.768	0.745	0.787	0.798	0.828	0.817	0.874	0.804	0.796	0.800
Jiangsu	0.911	0.892	0.900	0.923	0.885	0.891	0.882	0.880	0.876	0.873
Zhejiang	0.849	0.845	0.854	0.851	0.848	0.856	0.858	0.852	0.855	0.857
Anhui	0.876	0.877	0.875	0.863	0.836	0.822	0.814	0.809	0.800	0.796
Fujian	0.920	0.903	0.893	0.886	0.892	0.889	0.894	0.905	0.904	0.905
Jiangxi	0.891	0.876	0.850	0.847	0.885	0.859	0.847	0.839	0.842	0.866
Shandong	0.912	0.875	0.869	0.882	0.876	0.866	0.879	0.884	0.888	0.898
Henan	0.888	0.880	0.860	0.856	0.872	0.887	0.896	0.893	0.898	0.899
Hubei	0.775	0.782	0.788	0.781	0.782	0.784	0.791	0.791	0.776	0.785
Hunan	0.881	0.853	0.845	0.846	0.873	0.859	0.854	0.846	0.851	0.857
Guangdong	0.848	0.829	0.832	0.845	0.861	0.845	0.853	0.853	0.852	0.853
Guangxi	0.770	0.769	0.782	0.799	0.772	0.821	0.807	0.818	0.819	0.830
Hainan	0.751	0.731	0.738	0.770	0.736	0.748	0.776	0.764	0.774	0.787
Chongqing	0.798	0.819	0.810	0.810	0.807	0.816	0.823	0.812	0.814	0.817
Sichuan	0.797	0.754	0.749	0.754	0.761	0.795	0.813	0.839	0.831	0.835
Guizhou	0.906	0.902	0.900	0.912	0.901	0.889	0.882	0.879	0.871	0.882
Yunnan	0.679	0.677	0.655	0.658	0.671	0.664	0.655	0.664	0.662	0.669
Shaanxi	0.801	0.767	0.783	0.789	0.805	0.802	0.812	0.819	0.843	0.836
Gansu	1.000	0.860	0.831	0.844	0.850	0.752	0.739	0.753	0.743	0.758
Qinghai	1.000	0.970	0.952	0.864	0.855	0.818	0.824	0.824	0.791	0.745
Ningxia	1.000	1.000	1.000	1.000	1.000	0.970	0.941	0.929	0.924	0.914
Xinjiang	0.741	0.758	0.732	0.738	0.780	0.754	0.778	0.787	0.805	0.793
Average	0.857	0.838	0.837	0.840	0.838	0.836	0.841	0.841	0.842	0.855

Table 2 Results of Bayesian Information Criterion

Order	Model	K	Threshold	Complexity	BIC
1	ABQ_kD_k	4	0.2	41	-2,949.97
2	$A_{kj}BQ_kD_k$	4	0.2	44	-2,957.95
3	$A_kBQ_kD_k$	3	0.2	36	-3,927.08
4	$A_kBQ_kD_k$	4	0.2	47	-3,984.82
5	$A_{kj}BQ_kD_k$	2	0.2	26	-5,048.50
6	ABQ_kD_k	2	0.2	24	-5,048.64
7	$A_kBQ_kD_k$	2	0.2	28	-7,349.12
8	$A_{kj}B_kQ_kD_k$	2	0.2	31	-7,664.58
9	$AB_kQ_kD_k$	2	0.2	28	-7,666.82
10	$A_kB_kQ_kD_k$	2	0.2	29	-7,670.21
11	$A_kB_kQ_kD_k$	3	0.2	38	-7,775.86
12	$A_{kj}B_kQ_kD_k$	4	0.2	55	-9,100.29
13	$A_{kj}B_kQ_kD_k$	3	0.2	\	\
14	$AB_kQ_kD_k$	3	0.2	\	\
15	$A_{kj}BQ_kD_k$	3	0.2	\	\
16	ABQ_kD_k	3	0.2	\	\
17	$A_kB_kQ_kD_k$	4	0.2	\	\
18	$AB_kQ_kD_k$	4	0.2	\	\

Table 3 Specific clustering results of 30 provinces

Region	Province	Group	Description
Northeast	Heilongjiang	2	Medium
	Jilin	3	Low
	Liaoning	2	Medium
Central	Shanxi	2	Medium
	Henan	2	Medium
	Hubei	3	Low
	Hunan	2	Medium
	Jiangxi	2	Medium
	Anhui	1	Decrease
East	Beijing	3	Low
	Tianjin	4	High
	Hebei	4	High
	Shandong	2	Medium
	Jiangsu	2	Medium
	Shanghai	3	Low
	Zhejiang	2	Medium
	Fujian	2	Medium
	Guangdong	2	Medium
	Hainan	3	Low
West	Chongqing	3	Low
	Sichuan	3	Low
	Guangxi	3	Low
	Guizhou	2	Medium
	Yunnan	3	Low
	Shaanxi	3	Low
	Gansu	1	Decrease
	Inner Mongolia	2	Medium
	Ningxia	4	High
	Xinjiang	3	Low
	Qinghai	1	Decrease

Table 4 Specific grouping results for EC

Group	Description	Province
1	Fluctuant	Beijing, Liaoning, Shanghai, Jiangsu, Hunan, Guangdong, Xinjiang
2	Stable	Tianjin, Hebei, Shanxi, Inner Mongolia, Jilin, Heilongjiang, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Guangxi, Hainan, Chongqing, Sichuan, Guizhou, Yunnan, Shaanxi, Gansu, Qinghai, Ningxia

Table 5 Specific grouping results for BPC

Group	Description	Province
1	Rise rapidly→Decline→Rise gradually	Shanxi, Inner Mongolia, Shandong, Gansu
2	Fluctuate→Rise rapidly	Beijing, Liaoning, Shanghai, Hunan, Xinjiang
3	Fluctuate	Heilongjiang, Anhui, Jiangxi, Henan, Hubei, Sichuan, Yunnan, Shaanxi, Qinghai, Ningxia
4	Fluctuate (opposite)	Tianjin, Hebei, Jilin, Jiangsu, Zhejiang, Fujian, Guangdong, Guangxi, Hainan, Chongqing, Guizhou

Table 6 Specific grouping results for TGC

Group	Description	Province
1	Fluctuant	Inner Mongolia, Anhui, Henan, Gansu
2	Stable	Beijing, Tianjin, Hebei, Shanxi, Liaoning, Jilin, Heilongjiang, Shanghai, Jiangsu, Zhejiang, Fujian, Jiangxi, Shandong, Hubei, Hunan, Guangdong, Guangxi, Hainan, Chongqing, Sichuan, Guizhou, Yunnan, Shaanxi, Qinghai, Ningxia, Xinjiang

List of figures:

Fig.1: Total-factor unified efficiency of different provinces in 2008 and 2017

Fig.2: Clustering results of 30 provinces’ green efficiency in the logistics industry

Fig.3: Estimated mean function of each cluster

Fig.4: Four clusters of 30 provinces’ green efficiency in the logistics industry

Fig.5: Trends in the EC

Fig.6: Clustering results of the EC

Fig.7: Trends in the BPC

Fig.8: Clustering results of the BPC

Fig.9: Trends in the TGC

Fig.10: Clustering results of the TGC

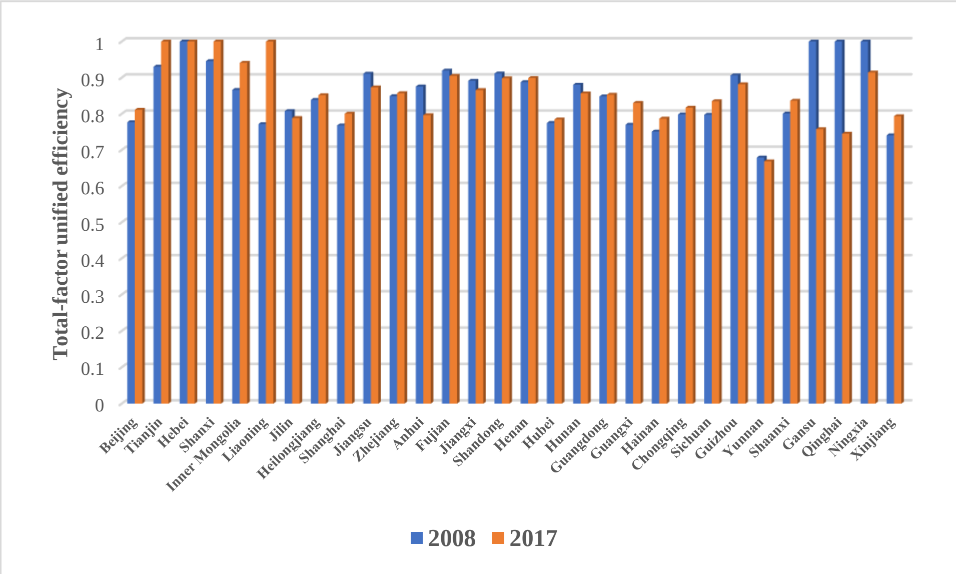

Fig.1 Total-factor unified efficiency of different provinces in 2008 and 2017

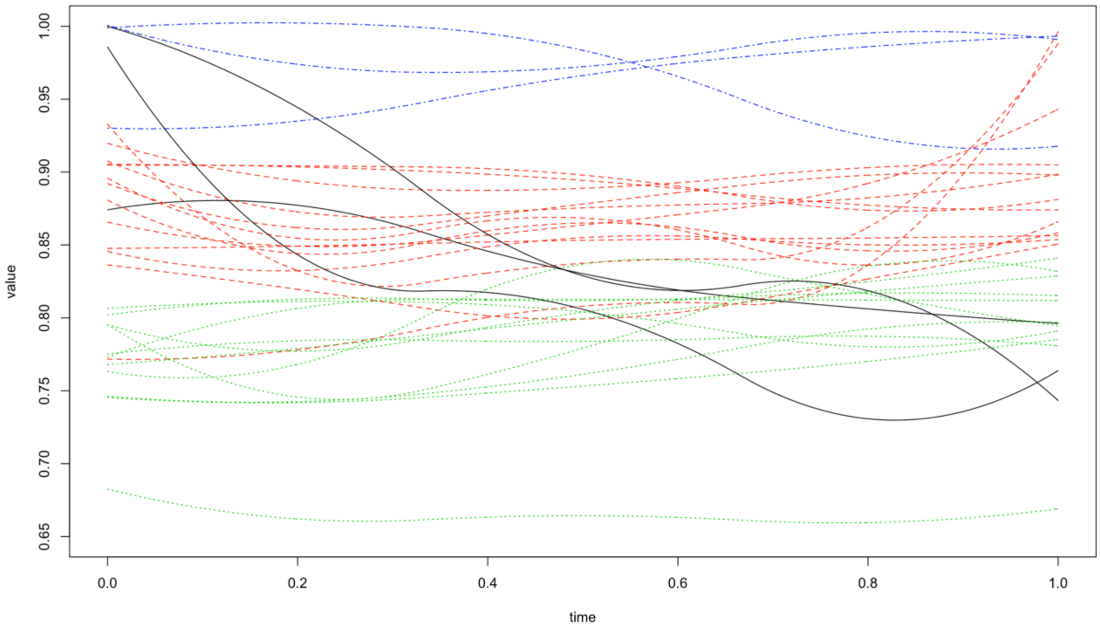

Fig.2 Clustering results of 30 provinces' green efficiency in the logistics industry

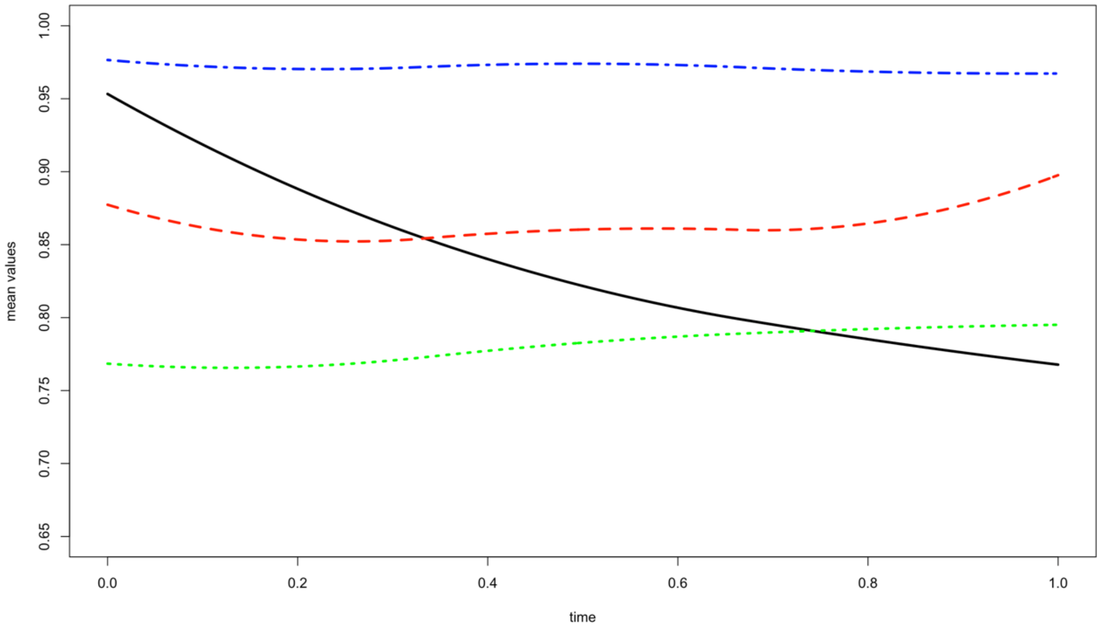

Fig.3 Estimated mean function of each cluster

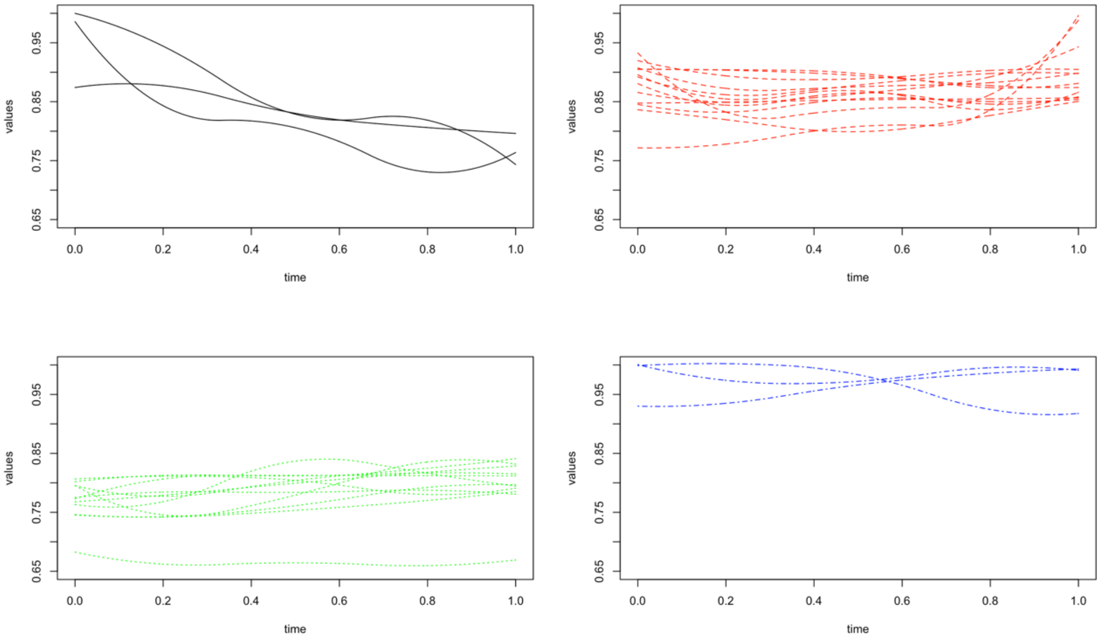

Fig.4 Four clusters of 30 provinces' green efficiency in the logistics industry

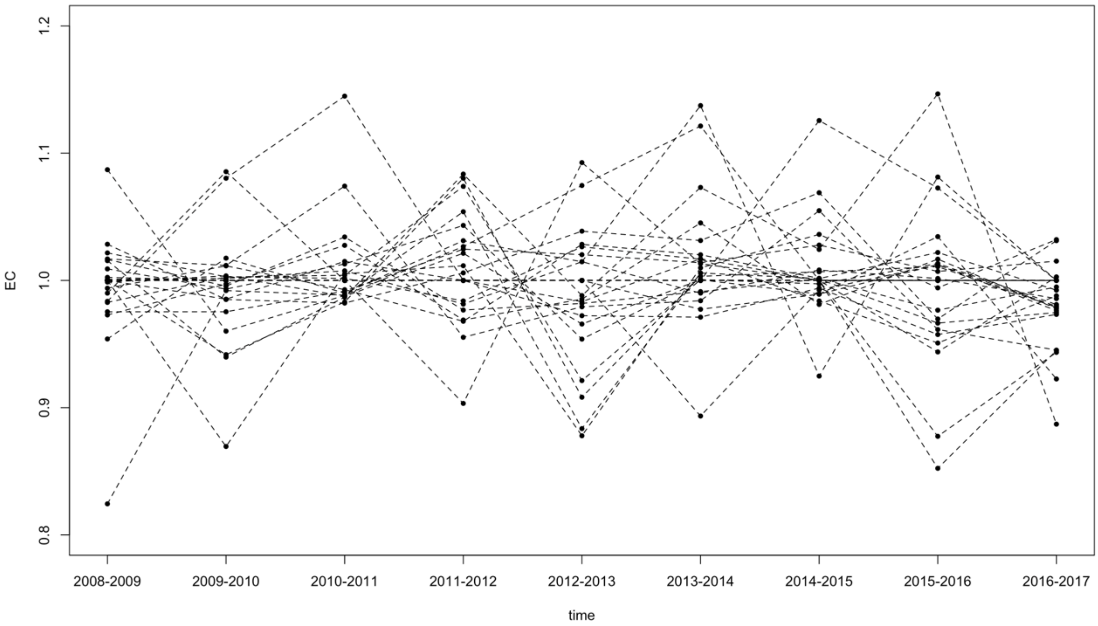

Fig.5 Trends in the EC

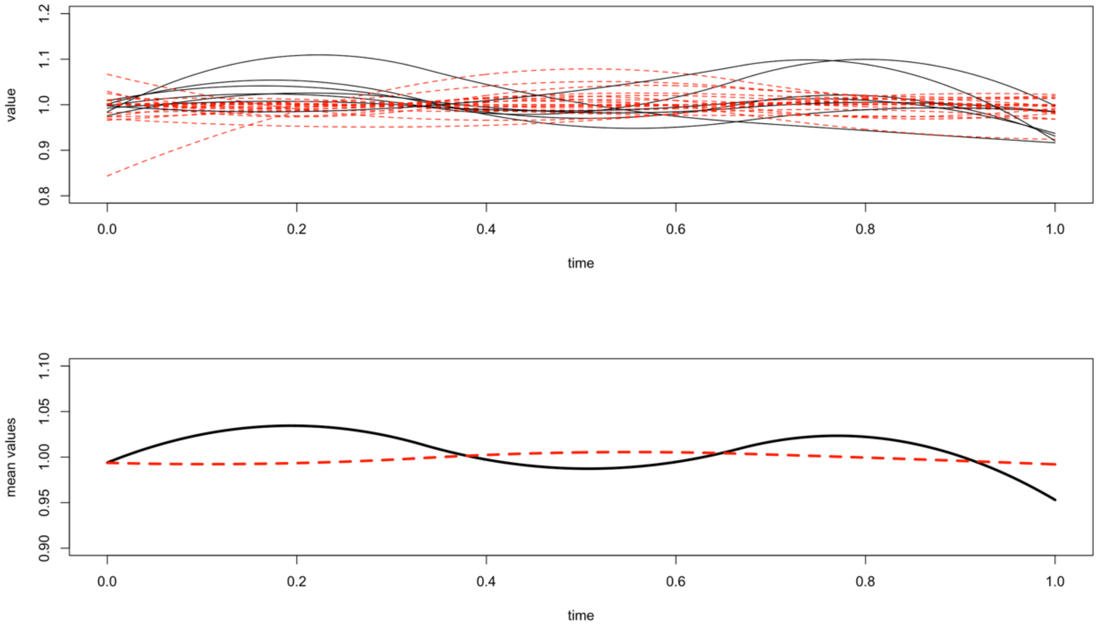

Fig.6 Clustering results of the EC

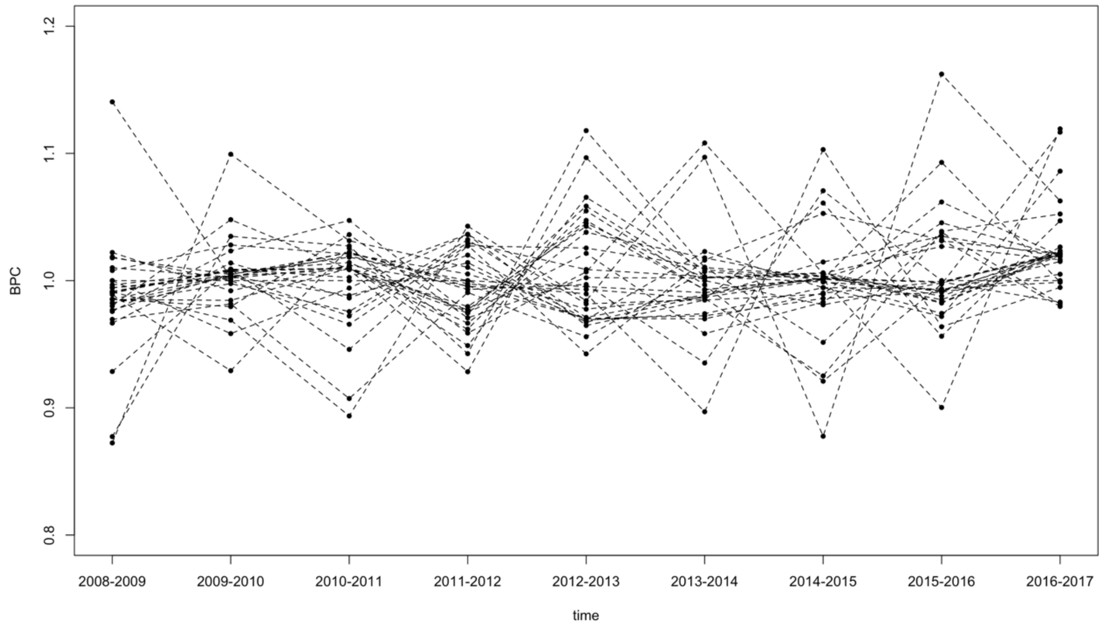

Fig.7 Trends in the BPC

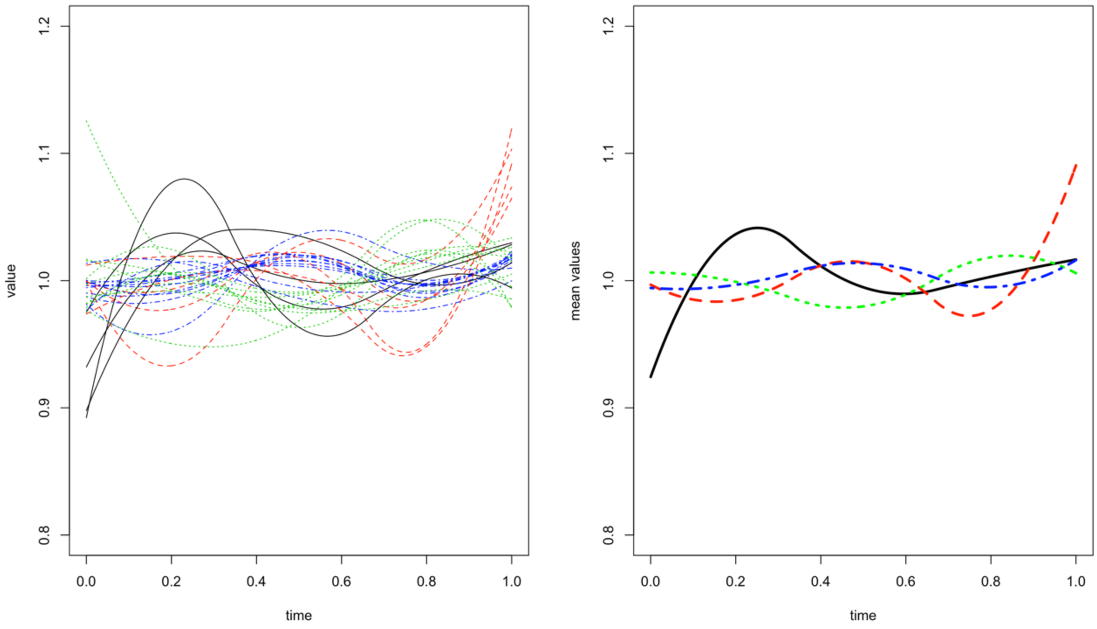

Fig.8 Clustering results of the BPC

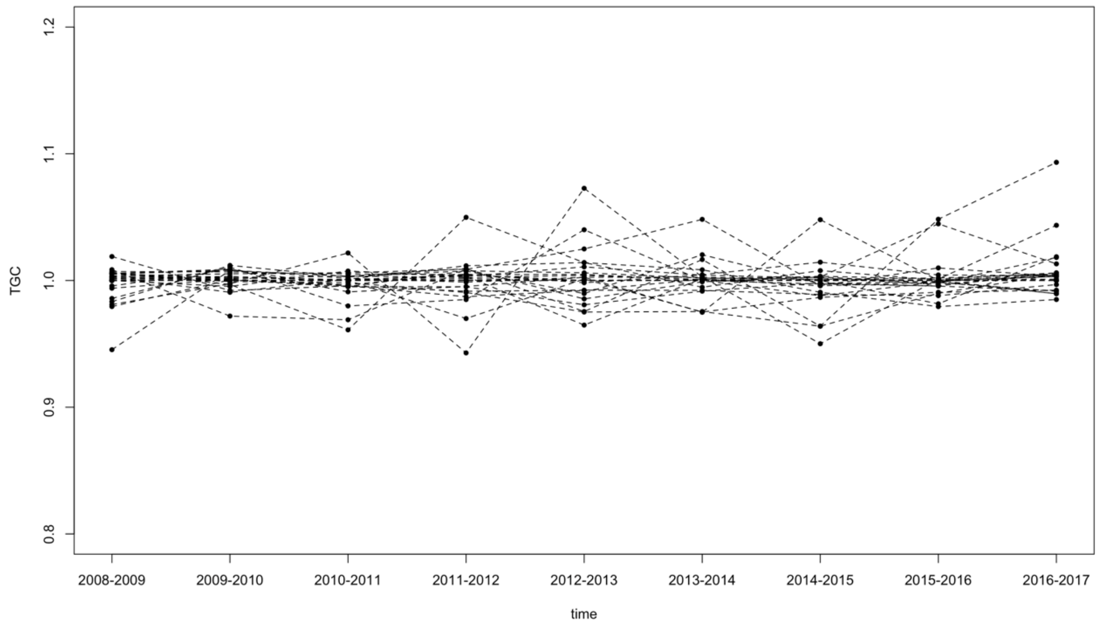

Fig.9 Trends in the TGC

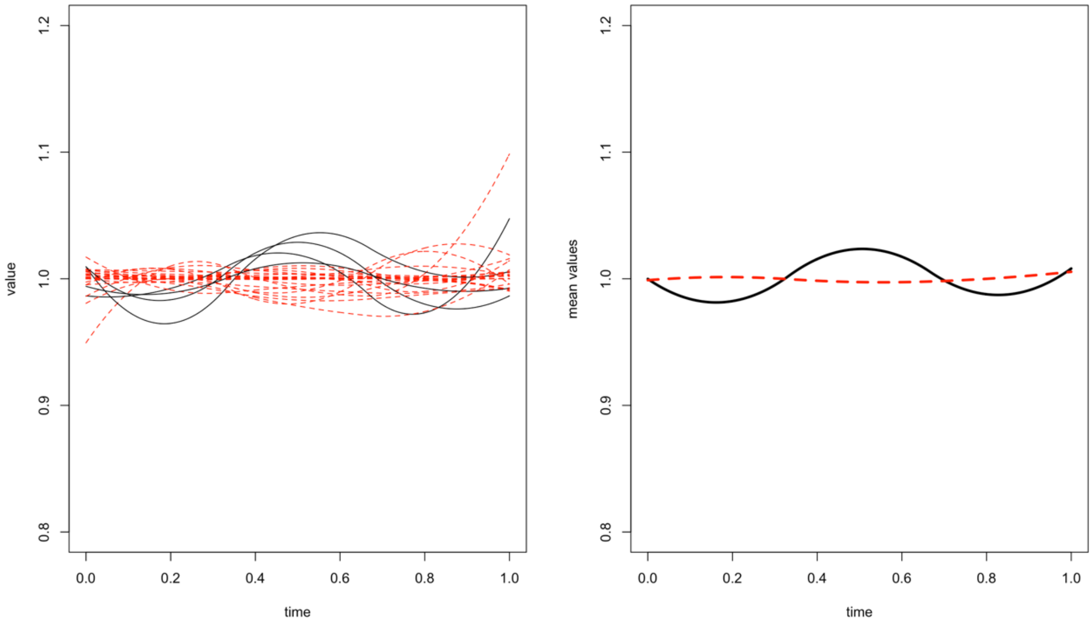

Fig.10 Clustering results of the TGC