Energy 179 (2019) 863–875

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Energy efficiency evaluation and energy saving based on DEA integrated affinity propagation clustering: Case study of complex petrochemical industries

Zhiqiang Geng [a, b], Rongfu Zeng [a, b], Yongming Han [a, b, *], Yanhua Zhong [c], Hua Fu [d]

[a] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, 100029, China
[b] Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing, 100029, China
[c] Jiangmen Polytechnic, Jiangmen, Guangdong, 529020, China
[d] Equipment Department, BGP INC., China National Petroleum Corporation, Hebei, 072750, China

ARTICLE INFO

Article history:
Received 24 July 2018
Received in revised form
1 March 2019
Accepted 5 May 2019
Available online 11 May 2019

Keywords:
Data envelopment analysis
Affinity propagation
Energy efficiency evaluation
Energy saving
Complex petrochemical industries

ABSTRACT

Data envelopment analysis (DEA) has been widely used in the energy efficiency analysis of industrial production processes. However, the traditional DEA model is not high in the division of the efficiency value of decision making units (DMUs), and produces a large number of DMUs with an efficiency value equal to 1, making it difficult to identify their merits and demerits. Therefore, a novel DEA model based on the affinity propagation (AP) clustering algorithm (AP-DEA) is proposed. Through the AP clustering algorithm, high influence input data of the energy efficiency can be obtained. The merits and demerits of DMUs can then be identified with a high degree of discrimination to obtain better efficiency groups. Finally, the proposed model is applied to evaluate the energy efficiency and optimize the energy configuration of the ethylene and pure terephthalic acid (PTA) production processes in complex petrochemical industries. The experimental results show that this proposed model can improve the efficiency value discrimination of efficiency values by effective DMUs better than the traditional DEA. Moreover, the energy saving potentials of ethylene and PTA production systems are approximately 0.49% and 24.74%, respectively, and the carbon emission reduction of the ethylene production system is approximately 10.04%.

1. Introduction

As a pillar industry closely related to the population's daily life, the petrochemical industry has promoted socioeconomic development and improved living standard [1,2]. However, according to the IPCC Fifth Assessment Report, the industrial sector contributed approximately 19% of energy utilization and 30% of greenhouse gas emissions in 2016 [3]. The main cause of global warming is greenhouse gas emissions, with carbon dioxide having the highest contribution rate [4]. It is essential to achieve lower carbon development in industrial production and find a more suitable method to reduce energy consumption and carbon emissions.

Ethylene production is a major proportion of the petrochemical industry. The energy consumption of the ethylene industry accounts for over half of industrial energy consumption [5], and there is a clear, strong relationship between energy efficiency and environmental impact as using fewer resources and releasing less pollution for generating the same products is associated with higher efficiency processes [6]. Therefore, clearly identifying the operating state of ethylene plants and the causes of energy consumption is a key method of improving the energy efficiency of the process [7]. As another major complex petrochemical and important raw material industry, the pure terephthalic acid (PTA) industry is closely related to the population's [8,9]. However, it is difficult to make breakthroughs in improving the industrial production efficiency of the PTA industry. From the analysis report of the IEA (International Energy Agency), the energy intensity improvement potential of the PTA industry is 43.2% [10]. Therefore, researching the energy efficiency evaluation of the PTA production industry has high economic benefits.

In practical efficiency evaluation, energy efficiency is usually analyzed following the mean and optimal index methods [11]. In

* Corresponding author. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, 100029, China.
E-mail address: hanym@mail.buct.edu.cn (Y. Han).

https://doi.org/10.1016/j.energy.2019.05.042
0360-5442/

addition, most ethylene plants use the special energy consumption (SEC) index to measure the energy consumption [12]. However, these methods cannot easily provide guidance for improving energy efficiency and suitably determine the objective connection of the input and output indicators of energy consumption. The index decomposition analysis (IDA) method can decompose energy consumption indicators that affect energy efficiency. Geng et al. used the IDA method to obtain three energy performance indicators (EPI) that affected the energy efficiency, including activity, structure and intensity [13]. Ang et al. proposed two criteria for the selection of the IDA method to understand the relative contribution of the overall activity level, activity structure, and energy intensity, and explain the difference in the total energy consumption between two countries or regions. Taking China's 30 regions as an example, their energy consumption has been compared and sorted [14]. The IDA method can analyze the hierarchical structure of energy efficiency indicators; however, due to the lack of energy-saving knowledge, it cannot play a guiding role in the optimization of the energy allocation by the decision-makers. The stochastic frontier analysis (SFA) method is a typical parameter method in the frontal analysis of energy efficiency. Lin et al. used the SFA method to study the average energy efficiency and energy saving potentials of the petrochemical industry based on the assumption of the *trans*-log production function [15]. Shui et al. used the SFA method to analyze the efficiency performance of a factory's “best practice” among peers, and assess the energy efficiency using factory production and utility consumption data in the automotive industry [16]. The SFA method is more complex for multi-output situations. It is necessary to merge multiple outputs or use distance functions, which is inconvenient for use in the complex petrochemical industry. The data envelopment analysis (DEA) model is a non-parametric programming model for evaluating the efficiency of multiple decision-making units (DMUs) with multiple inputs and outputs, and is widely used in energy and environmental efficiency assessment due to its excellent processing capability for multitudinous data and independence of the input-output dimension [17]. Chen et al. used the three-stage DEA model and a DEA model based on discriminant analysis to analyze the energy efficiency of the 30 provinces of China from the change trend in 2003 to the energy efficiency in 2011 [18]. Their results showed that, after using this method to eliminate the influence of environmental factors and random errors, the energy efficiency of the construction industry in most provinces had improved, and its effect has also been improved. Iftikhar et al. analyze the efficiency of energy and CO_2 emissions using the network DEA-SBM model and obtains guidance for resolving energy efficiency problems [19]. Yang et al. investigated the total-factor energy efficiency and analyses the trends of efficiency changes based on the DEA model in China's agricultural production. They also analyze the corresponding energy efficiency potential [20]. Kang et al. adopted a non-radial DEA model to examine the energy-environmental performance of China's manufacturing industry during 2006–2014 and obtained strategies for saving energy and reducing emission in the manufacturing industry [21].

However, the quantity and sample number of input-output indicators have a significant influence on the results of DEA analysis [22]. In addition, due to unsuitable weight distribution or the extreme input and output index, the traditional DEA model can easily show values of 1 for over one-third of the DMUs, which affects the decision-making process and makes it difficult to determine a direction for improving the production process. So there is a poor distinction between the results and the most efficient DMU cannot be identified. Therefore, the clustering algorithm is used to treat the input and the output of the DEA. Hashem et al. proposed an integrated fuzzy clustering cooperative game DEA model and uses a fuzzy C-means technique to cluster the DMUs and obtain efficient units [23]. Gong et al. proposed an improved energy efficiency evaluation method of ethylene production plants based on the DEA model integrated with comprehensive factor analysis [24]. Using the three key factors of raw material composition, cracking depth and load rate, the production conditions were determined using a k-means algorithm. Based on the multi condition model, the DEA model was used to evaluate the energy efficiency of ethylene production. The data of the highest energy consumption were screened by factor analysis, and supported the decisions made by operators to improve energy efficiency.

The above methods fully develop the decision-making role of the DEA model in the application process, and the method of combining the working condition with the k-means algorithm can be used as a preliminary treatment of the input index. However, the k-means algorithm is sensitive to outliers in the clustering process, and the clustering center K is set by human experience. Due to the uncertainty of the selection clustering of the k-means algorithm, the instability of the decision results in the actual application are unstable [25]. Therefore, there would be large deviations and instability when organizing complex industrial data. To solve this problem, the AP clustering algorithm without selecting the cluster center is chosen to cluster the input indices [26]. As the cluster center is not selected ahead of time, the AP clustering algorithm is considered as a deterministic algorithm, and is widely used in image recognition, image retrieval, image segmentation and text recognition [27–29].

To solve these problems, a novel DEA model based on the affinity propagation (AP) clustering algorithm (AP-DEA) is proposed. By discriminating the multiple input indices as cluster centers, the input indices are divided into low and high-influence indices based on the AP clustering algorithm. The high-influence indices are then taken as inputs for the DEA model to improve, deviation degree of all DMUs by using the DEA model. Finally, through calculating the relaxation coefficients of the relevant DMUs, the energy saving and emission reduction capabilities of ethylene and PTA production plants in complex petrochemical industries are obtained by the proposed model. The experimental results show that the proposed model can select the high-influence indices, which refines the input indices of the DEA model to a certain extent, and the result has a good discriminant ability. Compared with the traditional DEA model, the number of DMUs with low discrimination of PTA production data and ethylene production data decreased by 25% and 72%, respectively. Moreover, the proposed model can determine the improvement direction of the inefficient DMU to save energy and achieve effective production. Furthermore, the energy saving potentials of ethylene and PTA production systems are approximately 0.49% and 24.74%, respectively, and the carbon emission reduction of the ethylene production system is approximately 10.04%.

The rest of this paper is organized as follows: Section 2 introduces the AP-DEA model in detail. The improvement direction of the energy efficiency and energy saving and emission reduction potential of the PTA production plant and the ethylene production plant can be obtained by the proposed model in Section 3. Discussion and conclusion are given in Section 4 and Section 5, respectively.

2. DEA model based on AP clustering

2.1. Affinity propagation clustering algorithm

The AP clustering algorithm was first proposed by Frey and Dueck in 2007 [30] to find the optimal set of representative points to achieve the highest similarity of all data points to the nearest class representative points. As all the sample data points are used as the potential clustering center, the initial cluster center can be selected subjectively and effectively, which avoids the influence of outliers to the greatest extent, and effectively improves the reliability of the clustering results.

The AP clustering algorithm does not need to set the number of clusters in advance. The responsibility and availability values of each point during execution are updated until some high-quality clustering centers are generated. The remaining data points are then allocated to the corresponding cluster centers according to the similarity between the data objects based on the different criteria for different applications. This study selected Euclidean distance as a similarity criterion. First, the data is normalized based on the normalization formula shown in Eq. (1), where ϑ is the data before processing and ϑ^* is the data processed by normalization. These similarity data form a similarity matrix S, which is N^*N, to conduct the iterative computation process of the AP clustering algorithm.

$$\vartheta^* = \frac{\vartheta - min}{max - min} \tag{1}$$

The AP clustering algorithm uses the preference value P of the diagonal line in the similarity matrix S as the criterion for a data point to become a cluster center. In addition, the greater the value, the greater the likelihood of the point being the center of the cluster. The size of the preference P affects the number of cluster centers. When P is equal to the mean of all elements in the similarity matrix S, the number of cluster centers is a medium value. The reference value used in this paper is the mean value of all elements.

The AP algorithm passes two types of information in the process, the responsibility $R\,(i,k)$ and the availability $A\,(i,k)$. $R\,(i,k)$ describes the suitability of the data object X_k as the cluster center of the data object X_i, which represents the information transmitted from X_i to X_k. $A\,(i,i)$ describes the suitability of data object X_i to select data object X_k as its cluster center, which represents the information transmitted from X_k to X_i. The larger the values of $R\,(i,k)$ and $A\,(i,k)$, the more likely it is that the data object X_k will become the final cluster center. The responsibility and availability of each data object are updated until a specific number of times or products are iterated for some high quality cluster centers. Finally, the remaining data objects are assigned to the corresponding clustering results. The number of iterations is controlled by the set damping coefficient. When the damping coefficient is too small, the iteration process is easily concussed. The damping coefficient mentioned in this paper is set to 0.6 according to the linear relationship between the damping coefficient and the volatility [31]. When the number of iterations $t = 0$, the initial values of $R\,(i,k)$ and $A\,(i,k)$ are both 0, and the iterative process of the AP clustering algorithm are shown in Eqs (2)–(5):

The responsibility is calculated based on Eq. (2).

$$R_{t+1}(i,k) = \begin{cases} S(i,k) - \max\limits_{i' \neq k}\left\{A_t(i,k) + R_t\left(i,i'\right)\right\}, & when\ i \neq k \\ S(i,k) - \max\limits_{i' \neq k}\{S(i,k)\}, & when\ i = k \end{cases} \tag{2}$$

The iterative equation for calculating the responsibility is as follows:

$$R_{t+1}(i,k) = (1-\lambda)\cdot R_{t+1} + \lambda \cdot R_t(i,k) \tag{3}$$

The availability is calculated based on Eq. (4).

$$A_{t+1}(i,k) = \begin{cases} \min\left\{0, R_{t+1}(k,k) + \sum\limits_{i' \notin \{i,k\}} \max\{0, R_{t+1}(i',k)\}\right\}, when\ i \neq k \\ \sum\limits_{i' \neq k} \max\left\{0, R_{t+1}\left(i',k\right)\right\}, when\ i = k \end{cases} \tag{4}$$

The iterative equation for calculating the availability is as follows:

$$A_{t+1}(i,k) = (1-\lambda)\cdot A_{t+1} + \lambda \cdot A_t(i,k) \tag{5}$$

The responsibility value is iterated, and the availability value is then calculated by the result of the iteration. During each iteration, the values of $R\,(k,k)$ and $A\,(k,k)$ are greater than 0 along the diagonal as candidates for the cluster center. When the number of iterations exceeds the set threshold, or the cluster center does not change continuously for several iterations, the iteration process is stopped. The sample data are then divided into specific clusters according to the clustering results. This iterative algorithm considers the direct influence of the nearest neighbor point and the nearby area on the representation of clustering, and the clustering results of multiple independent operations have the stability to prevent them from falling into the local optimal predicament.

2.2. DEA model

In 1978, Charnes et al. proposed the DEA model [32], which is an ideal and effective model with "scale effectiveness" and "technical effectiveness". The main principle of the DEA model is to determine the relatively effective production frontier by using statistical data and mathematical programming. The input or output of the DMU is kept unchanged, and the projection of each DMU is obtained based on the production frontiers. The relative effectiveness of each DMU is evaluated by comparing their deviation degree at the production frontiers. As there is no need to make assumptions about the input and output, the DEA model excludes many subjective factors, and has good objectivity in practical applications. As an efficiency analysis model for non-parametric optimization, the DEA model is suitable for multiple input and output systems. For the multi-input and output systems in complex petrochemical processes, the DEA model can evaluate the influence of multiple factors on improving the energy efficiency well. Based on the classic CCR model (abbreviations of the names of three authors: A. Charnes & W. W. Cooper & E. Rhodes), suppose m DMUs. In addition, each DMU includes i inputs and j outputs. For the nth DMU, the input vector is $\boldsymbol{d}_r = (d_{1r}, d_{2r}, \ldots, d_{ir})^T$ and $\boldsymbol{d}_w > 0$, and the output vector is

$\boldsymbol{e}_r = (e_{1r}, e_{2r}, ..., e_{jr})^T$ and $\boldsymbol{e}_w > 0$, where $r = 1,2, \dots, m$; $k = 1,2, \dots, i$; and $l = 1,2, \dots j$. The DEA model is evaluated as follows (fractional form):

$$s.t.\begin{cases} \max \dfrac{\boldsymbol{u}^T \boldsymbol{e}_l}{\boldsymbol{v}^T \boldsymbol{d}_k} \\ \dfrac{\boldsymbol{u}^T \boldsymbol{e}_r}{\boldsymbol{v}^T \boldsymbol{d}_r} \le 1,\ r = 1, 2, \dots, m \\ u \ge 0, v \ge 0, u \ne 0, v \ne 0 \end{cases} \tag{6}$$

where $\boldsymbol{v} = (v_1, v_2, \dots, v_i)^T$ *and* $\boldsymbol{u} = (u_1, u_2, \dots, u_j)^T$ are the weight coefficients of the i inputs and j outputs, respectively.

The equivalent linear programming form of the DEA model can be obtained based on Eq. (7).

$$s.t.\begin{cases} \max \alpha^T \boldsymbol{e}_l = \boldsymbol{v}^l \\ \beta^T \boldsymbol{d}_r - \alpha^T \boldsymbol{e}_r \ge 0, r = 1, 2, \dots, m \\ \beta^T \boldsymbol{d}_k = 1, \beta \ge 0, \alpha \ge 0 \end{cases} \tag{7}$$

Assuming that the optimal target value in Eq. (7) is $v = 1$, the DMU is called the ineffective DEA (v is the efficiency index).

$$s.t.\begin{cases} \min \theta \\ \sum_{r=1}^{m} \boldsymbol{d}_r \lambda_r \le \theta \boldsymbol{d}_m, \sum_{r=1}^{m} \boldsymbol{e}_r \lambda_r \ge \boldsymbol{e}_m \\ \lambda_r \ge 0, r = 1, 2, \dots, m, \theta \in E^1 \end{cases} \tag{8}$$

Any optimal solution θ and λ of Eq. (8) can be solved by linear programming. The value of θ will not exceed 1. In the DEA model, if $\theta < 1$, the DMU is ineffective, and if $\theta = 1$, the DMU is effective. All effective DMUs constitute the effective production boundary in the DEA model. The smaller the value of θ, the farther the distance from the production boundary, and the lower the relative efficiency.

However, the DEA self-evaluation model mentioned above does not provide a specific direction for improvement. Therefore, the input-output relaxation coefficient and the non-Archimedes infinitesimal $\boldsymbol{\varepsilon}$ are introduced to Eq. (7), as shown in Eq. (9).

$$s.t.\begin{cases} \min\left[\theta - \varepsilon\left(e_1^T g^- + e_2^T g^+\right)\right] \\ \sum_{r=1}^{m} \lambda_r d_{kr} + g^- = \theta d_{kr}, k = 1, 2, \dots, i \\ \sum_{r=1}^{m} \lambda_r e_{lr} - g^+ = \theta e_{lr}, l = 1, 2, \dots, j \\ \lambda_r \ge 0, r = 1, 2, \dots, m \\ g^+ \ge 0, g^- \ge 0 \end{cases} \tag{9}$$

In practical applications, the non-Archimedes infinitesimal $\boldsymbol{\varepsilon} = 10^{-6}$, and g_t^-, g_t^+ is the relaxation coefficient. $g_t^- = (g_t^{1-}, g_t^{2-}, \dots, g_t^{i-})^T, g_t^+ = (g_t^{1+}, g_t^{2+}, \dots, g_t^{j+})^T$ represents the redundancy of the ith input and the deficiency of the jth output. In addition, $e_1^T = (1, 1, \dots, 1) \in R^i, e_2^T = (1, 1, \dots, 1)^T \in R^j$. θ represents the valid value of the input relative to the output.

2.3. Energy-efficiency analysis process based on the AP-DEA model

The main process of the AP-DEA model is shown in Fig. 1 and the following explanation is described.

Step 1 Data preprocessing. To better use the AP clustering algorithm, the data should be normalized according to the characteristics of the data and obtain the initial similarity matrix S.

Step 2 AP clustering process. The cluster center is obtained by iteratively solving the responsibility and the availability. The final clustering center is then obtained, and the sample data are divided into the specific clustering categories. Using the sample data corresponding to specific clustering categories, the efficiency values of each DMU are calculated by Eqs (6)–(8).

Step 3 Analyze the improvement direction of the corresponding DMUs. By solving the corresponding relaxation coefficient based on Eq. (9), the redundancy of input and the deficiency of the output can be obtained. Then the input-output of ineffective DMUs can then be improved and guided.

The DEA model has a certain required number of samples. The number of samples and indicators follow a certain principle - the number of samples (the number of DMUs) should not be lower than the product of the number of input and output indicators [33]. This principle is based on the minimum requirements of the data simulation results and efficiency discriminations [22,34,35]. In actual research, although the above requirements are satisfied, the discrimination may not satisfy the actual analysis needs.

3. Case study: energy efficiency evaluation and energy saving in complex petrochemical processes

3.1. Analysis of PTA plants

The main index for measuring the effectiveness and progressiveness of the PTA production plant is the consumption of acetic acid. If the goal of improving the overall production economic benefit is achieved, the consumption of acetic acid should be reduced as much as possible. In the actual PTA production process, the content of acetic acid is not easily measured, so it cannot be directly used as the input index of the experimental model [36]. The conductivity of the top tower corresponds to the acetic acid content at the top of the tower. Therefore, the top tower's conductivity is often indirectly used to indicate the changes in the acetic acid content.

In the PTA production process, the solvent dehydrating tower is the core device for separating the acetic acid solvent and water, which is a key component of the process as it removes the oxidation reaction to generate water, maintaining the water content of the whole oxidation unit and reducing the core function of acetic acid consumption. The solvent-dehydrating tower involves heat exchange. Therefore, it is a major component of the energy consumption analysis of the PTA plant [37]. A solvent-dehydrating tower of PTA plant is shown in Fig. 2.

There are seventeen main factors affecting the conductivity of the acetic acid. After the specific analysis, the seventeen factors are as follows: feed amount FC1501, temperature TI1504, water reflux FC1502, FC1503, and FC1504, feed temperature TI15010, steam flow FC1507, and temperatures inside the tower TI1511~TI1519, and TC1501. Specific input indicators and corresponding conductivity are shown in Fig. 3.

These factors are clustered by the AP clustering algorithm, and the results are shown in Table 1.

The water reflux FC1502, FC1503, steam flow FC1507, and temperatures inside the tower TI1516, TI1517 and TC1501 are obtained, which have a higher influence on the PTA production process. The six input indices are used as the input indices of the DEA model. The highest conductivity of the tower is the main output index of the DEA model. The improved DEA efficiency value of the PTA production plant is obtained based on the AP-DEA model,

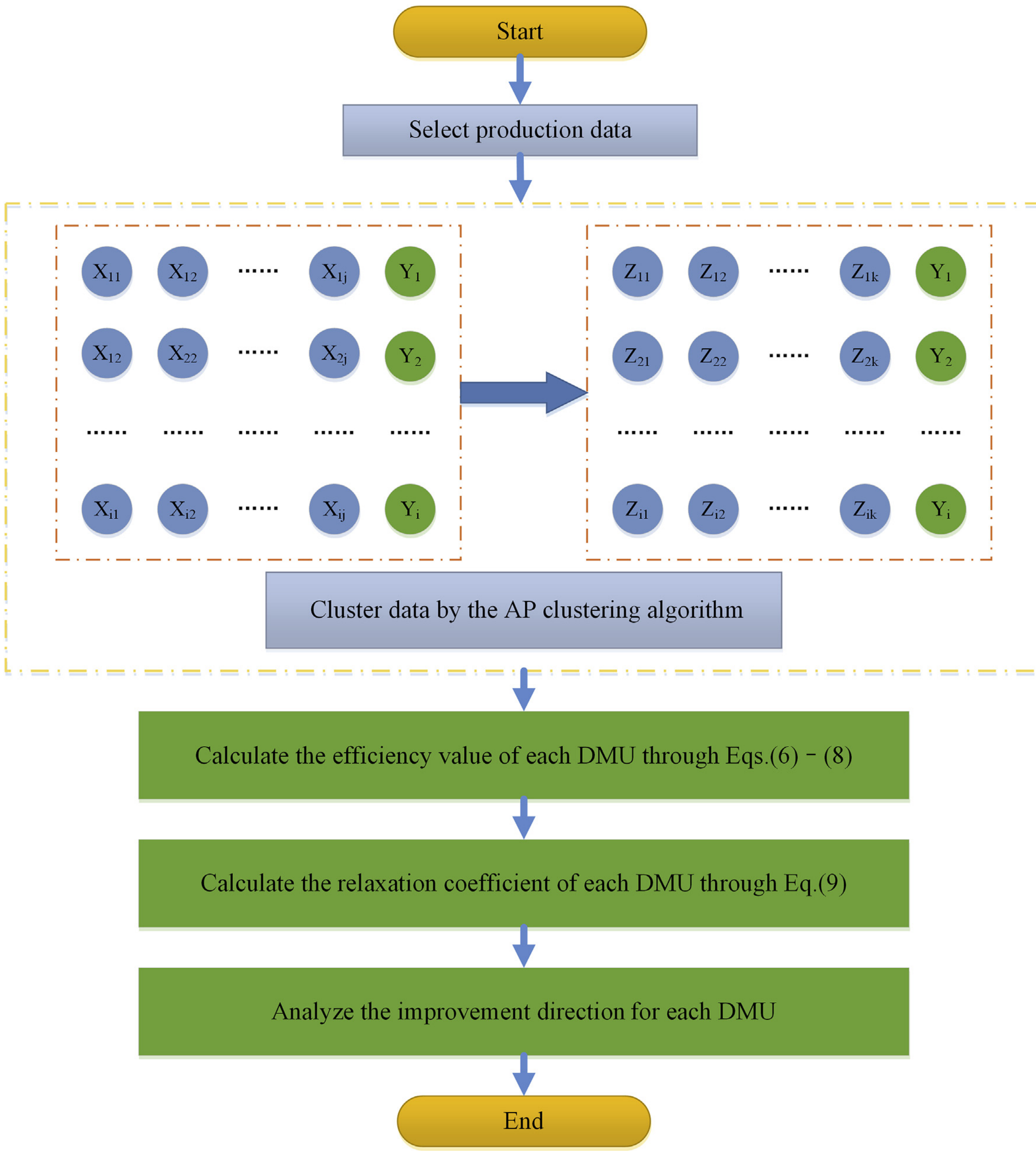

Fig. 1. The process of the AP-DEA model.

while the seventeen factors affecting the consumption of acetic acid are directly used as the input index of the traditional DEA model, and the conductivity (acetic acid content) is used as the output index.

There are 20 groups of data from a PTA production plant. The comparisons of the efficiency values of PTA production plants based on the traditional DEA and the improved AP-DEA are shown in Fig. 4.

As shown in Fig. 4, eight DMUs with an efficiency value of 1 were obtained using the traditional DEA model, exceeding one-third of the total number of samples, while the AP-DEA model has six DMUs with the efficiency value in 1. The number of ineffective units of the PTA production plant is decreased by 25%. The discrimination of prediction results is usually judged by using the variance, the standard deviation and the average deviation [38,39]. When the standard deviation is large, the discrimination of the DMUs is high [38,39]. Therefore, the standard deviation is used as the discrimination evaluation method in this study. Let the efficiency value θ_i ($i = 1,2, \ldots,20$) be the variable. The standard deviation of the data is obtained from Eq. (10) [40].

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i - \theta_{Mean})^2} \tag{10}$$

where θ_{Mean} is the mean efficiency value. According to Eq. (10), the standard deviations of the traditional DEA and AP-DEA models are 0.0097 and 0.0142, respectively. Therefore, the AP-DEA model is more accurate than the traditional DEA model, and the efficiency discrimination is higher, which is more suitable for determining the direction improving the energy efficiency.

Based on Eq. (9), the corresponding relaxation coefficient of the

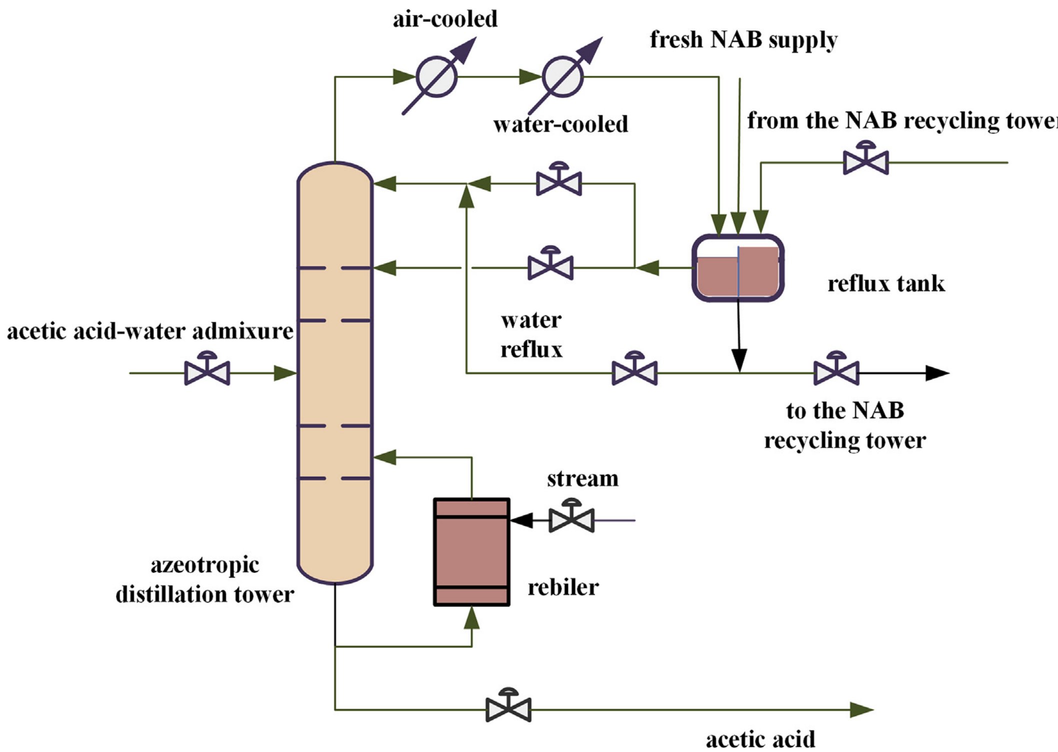

Fig. 2. A solvent dehydrating tower of the PTA production plant.

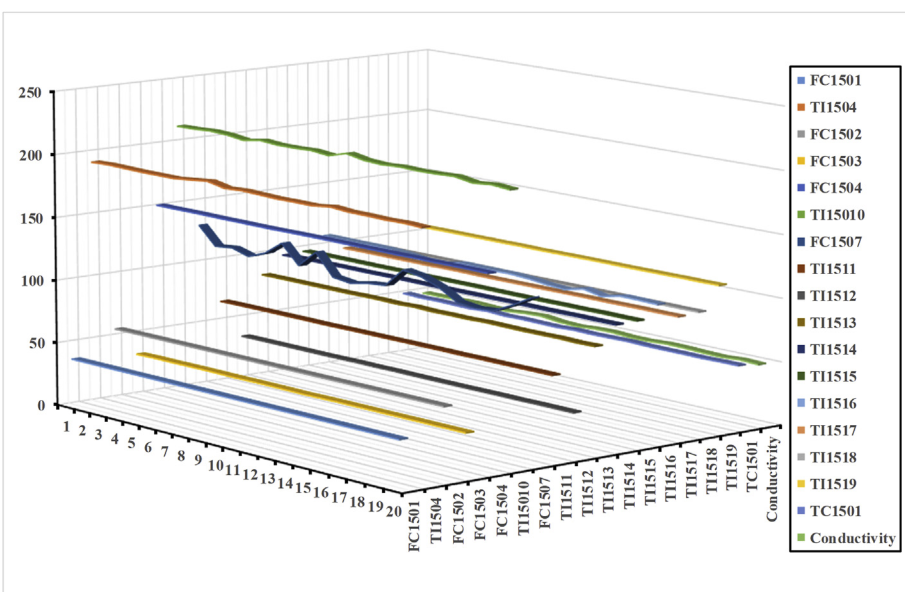

Fig. 3. Input index of PTA production plant.

inputs and outputs of the ineffective DMUs can be obtained. The redundancy of the input g_i^- ($i = 1, 2, \ldots, 6$, which represents the redundancy of the clustering results with the water reflux FC1502, FC1503, steam flow FC1507 and temperatures inside the tower TI1516, TI1517 and TC1501) and the deficiency of the output g^+ (which represents the deficiency of the conductivity.) of the PTA production plant are shown in Table 2.

Table 2 and Fig. 4 show that the efficiency values of samples 7, 8, 10, 11, 16 and 17 are all 1. According to the redundancy of the input and the deficiency of the output, the direction for improving the production configuration of the ineffective production samples with efficiency values lower than 1 can be obtained. For example, the efficiency value of the sixth sample is 0.96. If the water reflux FC1503 is unchanged, the water reflux FC1503, steam flow FC1507, and temperatures inside the tower TI1516 and TC1501, are reduced by 0.02, 0.17, 1.75, and 0.65, respectively, and the temperature inside the tower TI1517 and output are unchanged. This sample can then achieve effective production. Other ineffective samples can

Table 1
The input indexes of the PTA plant based on the AP clustering algorithm.

No.	Input Variable	Clustering results
1	Feed amount FC1501	
2	Feed temperature TI1504	
3	Feed temperature TI15010	
4	Water reflux FC1502	Water reflux FC1502
5	Water reflux FC1503	Water reflux FC1503
6	Water reflux FC1504	
7	Steam flow FC1507	Steam flow FC1507
8	Temperature inside the tower TI1511	
9	Temperature inside the tower TI1512	
10	Temperature inside the tower TI1513	
11	Temperature inside the tower TI1514	
12	Temperature inside the tower TI1515	
13	Temperature inside the tower TI1516	Temperature inside the tower TI1516
14	Temperature inside the tower TI1517	Temperature inside the tower TI1517
15	Temperature inside the tower TI1518	
16	Temperature inside the tower TI1519	
17	Temperature inside the tower TC1501	Temperature inside the tower TC1501

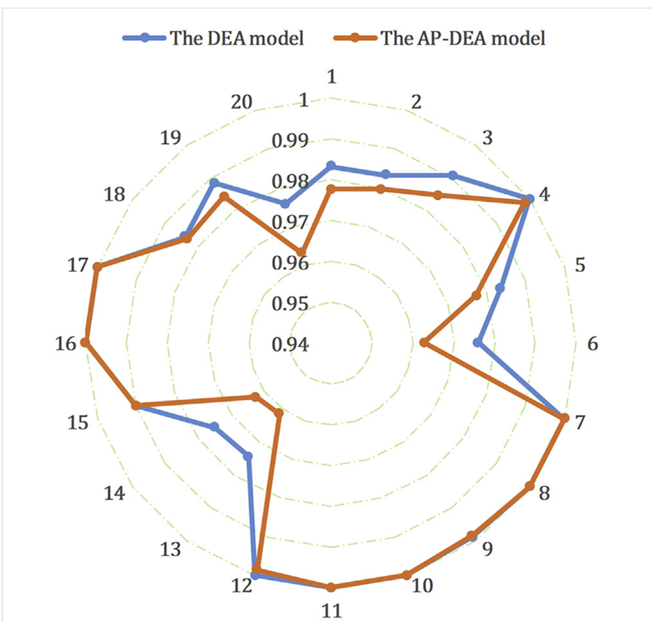

Fig. 4. Comparison of the efficiency values in PTA production plants.

also undergo similar analysis to achieve effective production. As the specific improvement direction is provided, it also ensures more effective energy utilization. As shown in Fig. 3, the acetic acid consumption is approximately about 968.31, but considering the improved efficiency values of all DMUs, the acetic acid consumption can be reduced by 3.89 through calculating the relaxation coefficient of the AP-DEA model. Also, the energy-saving potential of the PTA production plant can be improved by approximately 0.49%.

3.2. Energy efficiency analysis and carbon emissions reduction of ethylene production plants

Ethylene production plants are a key component of fossil fuel energy consumption, and the production process is complex, as shown in Fig. 5. The ethylene production process can be divided into two parts: cracking and separation. Separation mainly consists of three parts: rapid cooling, compression and separation. The cracking stage is the core component of the whole production process, and it also consumes the largest amount of energy throughout production. During pyrolysis, the cracking furnace requires a large amount of fuel to provide the heat required for production [41].

According to the ethylene production process mentioned above, feed, fuel, steam, electricity and water are used as input indicators for ethylene production plants. In the production process, these five input indicators refer specifically to naphtha, raffinate, hydrog. oil, LHydr, C345, and others, which belong to feed; fuel gas, which belongs to fuel; high, middle and low press, which belong to steam, circle, industrial and boiler water, which belong to water and electricity in the production process. The specific input indicators are shown in Fig. 6.

Most ethylene plants usually use the SEC as an indicator for analyzing the energy consumption in the production process. As shown in Fig. 7, the overall SEC presents the energy consumption of an ethylene plant from 2012 to 2013, but it cannot show how to reduce the energy consumption and calculate the energy-saving potential of the ethylene plant.

Table 2
The ineffective efficiency values and the relaxation coefficient of the PTA production plant.

Sample	1	2	3	4	5	6	9	12	13	14	15	18	19	20
g_1^-	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
g_2^-	0.00	0.12	0.03	0.14	0.26	0.02	0.00	0.14	0.11	0.08	0.32	0.14	0.20	0.14
g_3^-	0.34	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00
g_4^-	3.76	3.05	3.00	1.71	2.94	1.75	0.32	1.71	1.88	0.18	0.00	0.79	3.00	2.44
g_5^-	0.65	0.00	0.00	0.00	0.24	0.00	0.04	0.00	0.00	0.00	0.60	0.50	0.00	0.17
g_6^-	1.01	0.26	0.50	0.00	0.00	0.65	0.00	0.00	0.10	0.40	0.31	0.00	0.11	0.00
g^+	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
θ	0.98	0.98	0.98	0.99	0.98	0.96	0.99	0.99	0.96	0.96	0.99	0.98	0.98	0.96

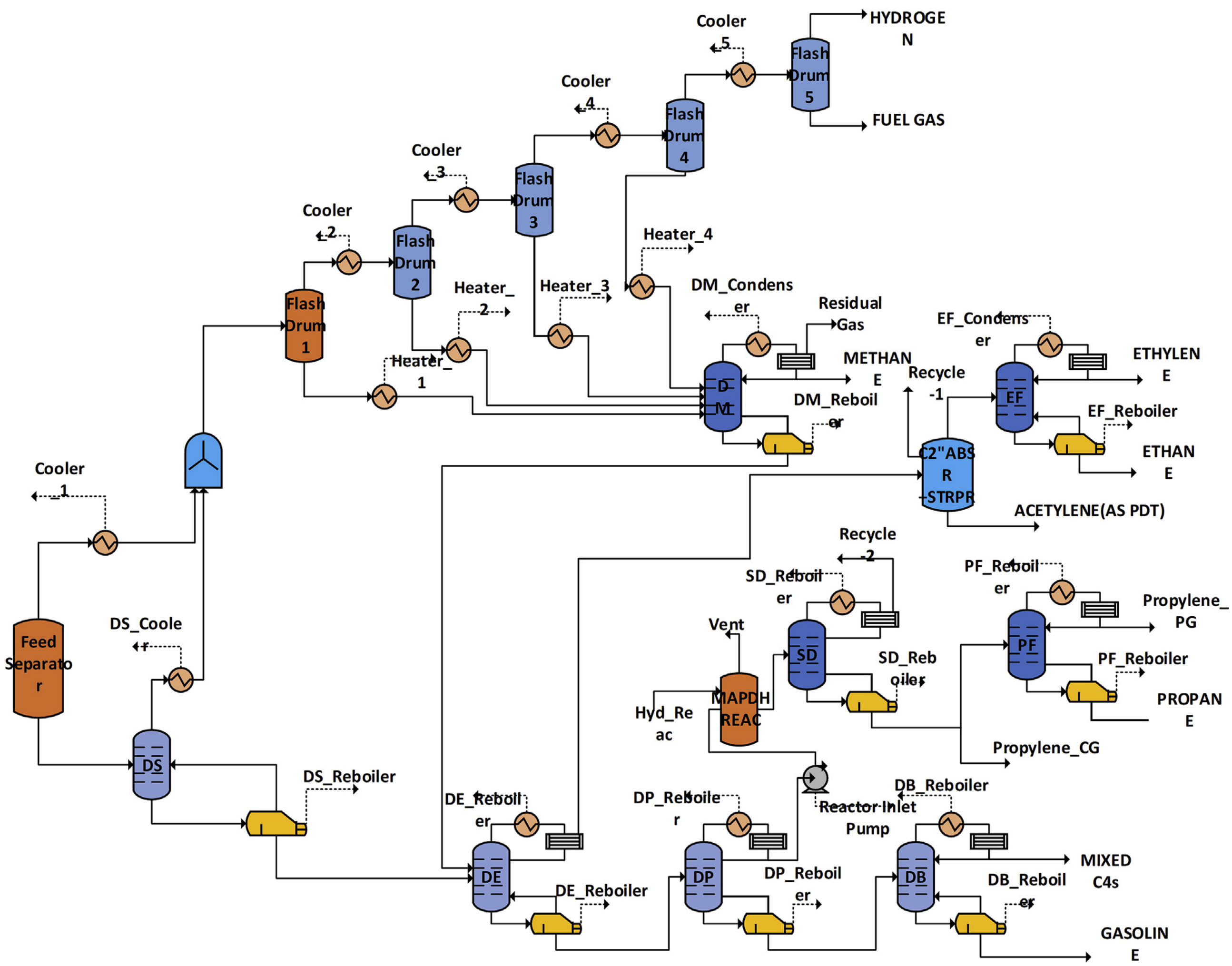

Fig. 5. The production process of a typical ethylene plant.

Based on the AP-DEA model, the corresponding energy-saving potential of the plant can be further analyzed after obtaining the corresponding efficiency value and relaxation coefficient. In addition, the input indices can be divided into energy and non-energy input indices according to the consumption of production and energy resources, such as feed, fuel, steam, water, and electricity, which are categorized as energy inputs that are linearly related to the degree of carbon emissions from the production process [42]. These carbon emissions cannot be accurately measured directly, but studying the carbon emissions in industrial production is a prerequisite for resolving energy efficiency problems [43,44]. Therefore, the carbon emissions of the production process can be estimated according to the linear relationship between energy input and carbon emissions. The formula of this calculation is shown in Eq. (11):

$$C_l = \sum_{k=1}^{P} \left(D_{kl} \times E_k \right) \tag{11}$$

where C_l and D_{kl} represent the carbon dioxide emission value consumption value of the kth energy input in the lth DMU, respectively. E_k represents the carbon emission coefficient of the kth energy input mentioned above. P is the number of energy inputs, and the common carbon dioxide emission coefficient is shown in Table 3.

The input-output relaxation coefficient of the corresponding DMU can be calculated using Eq. (9). The improvement direction of the DMU can be further clarified by analyzing the relaxation coefficient. The energy saving potential of each energy input unit and the overall carbon emission reduction potential in each DMU can also be obtained and analyzed, as shown in Eq. (12), where Q_{pk} represents the energy saving potential of the kth energy input of DMU, and C_r represents the potential for reducing carbon emissions.

$$Q_{pk} = 1 - \theta_{pk} D_{pk} + s_{pk}^{-}$$

$$C_r = \left(1 - \theta_{ck} \right) C_k \tag{12}$$

Taking these factors as the input index the clustering results can be obtained based on the AP clustering algorithm as shown in Table 4.

The naphtha, hydrofoil, industry water and electricity have a great influence on the ethylene production process. These four input indicators are used as input indices for the AP-DEA model, while the ethylene yield is used as the output index. The improved DEA efficiency value of the ethylene production plant is obtained from the AP-DEA model. Further the fourteen indicators of the

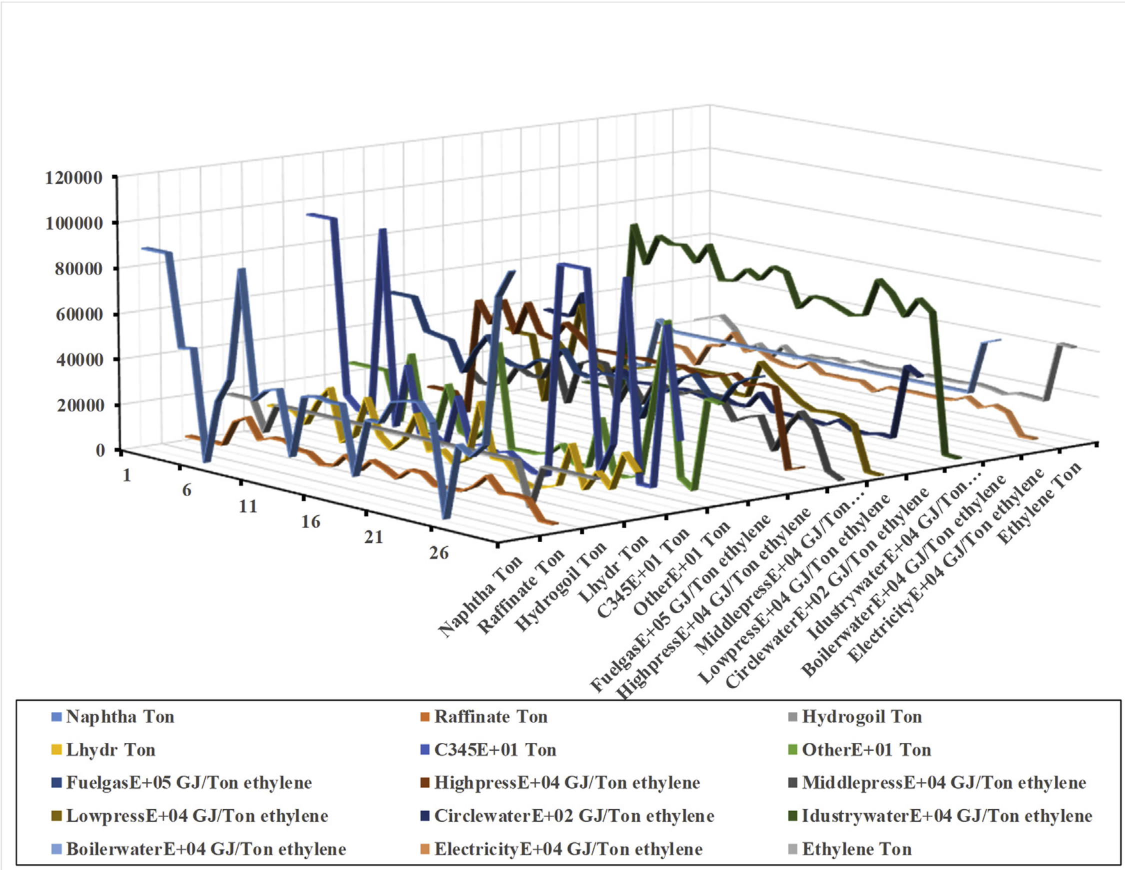

Fig. 6. Input indexes and output of the ethylene production plant.

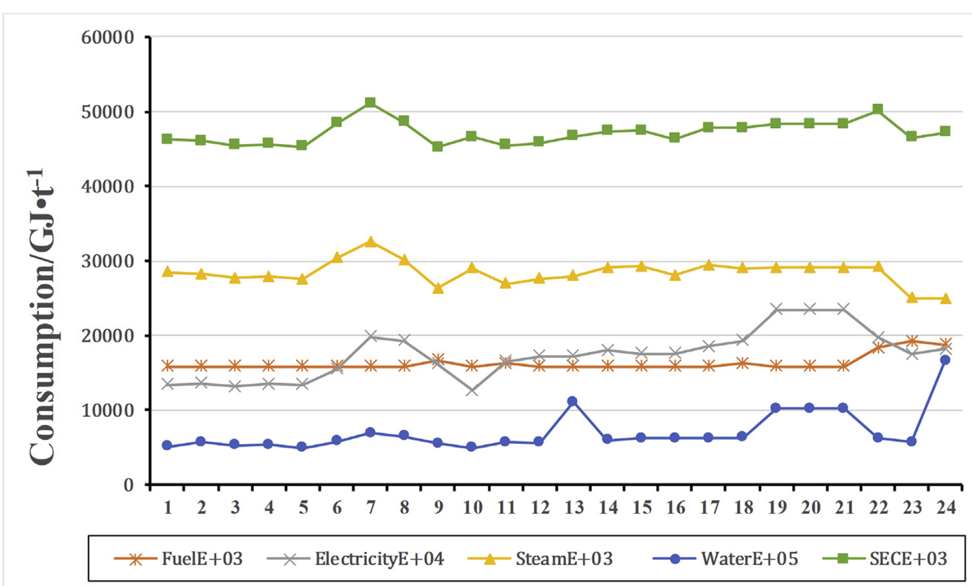

Fig. 7. Energy consumption of the ethylene production plant.

ethylene production process are used as the input index of the traditional DEA model, and the ethylene yield is used as the output index.

There are 30 groups of ethylene production data from a Chinese ethylene plant. The results of the efficiency values in ethylene production plants based on the traditional DEA and improved AP-DEA are compared in Fig. 8.

From analyzing Fig. 8, there are 27 DMUs with an efficiency value of 1 according to the traditional DEA model, exceeding one-third of the total number of samples, while the AP-DEA model

Table 3
Common carbon dioxide emission. (Kg-CO_2/Kg).

Energy	Carbon dioxide emission	Energy	Carbon dioxide emission
Raw coal	2.07	Kerosene	3.08
Cleaned coal	2.49	Diesel	3.16
Slime, Middlings	0.89	Refinery dry gas	2.65
Briquette	2.02	Liquefied petroleum gas	3.17
Coal Briquette total	2.23	Other coke products	3.04
Coke	3.04	Other petroleum products	2.95
Crude Oil	3.07	Natural gas	21.84
Fuel oil	3.24	Coke oven gas	7.71
Gasoline	3	Other coal gas	5.92
Electricity	3.56		

Table 4
The clustering results of ethylene production plants.

No.	Input Variable	Clustering results
1	Naphtha	Naphtha
2	Raffinate	
3	Hydrogoil	Hydrogoil
4	Lhydr	
5	C345	
6	Other	
7	High press	
8	Middle press	
9	Low press	
10	Circle water	
11	Industry water	Industry water
12	Boiler water	
13	Electricity	Electricity
14	Fuel gas	

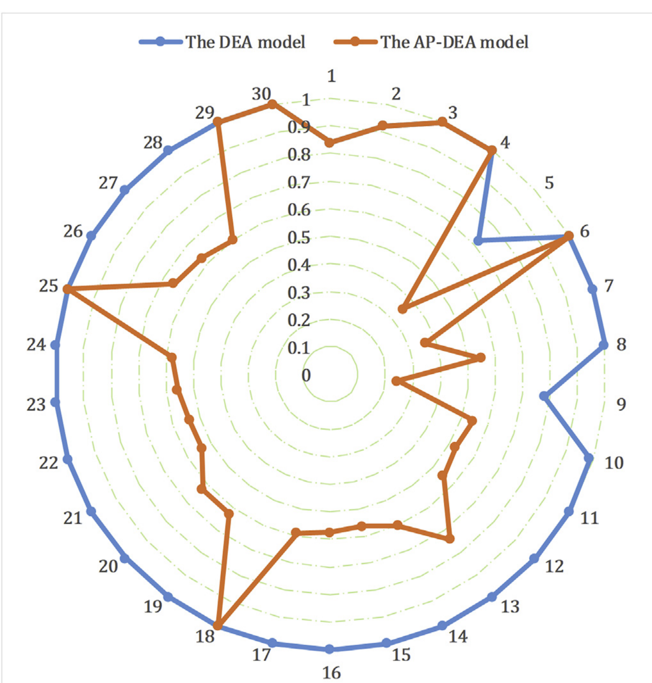

Fig. 8. Comparison of the efficiency values in ethylene production plants.

has nine DMUs with an efficiency value of 1. The number of ineffective units in the ethylene plant is decreased by 78%. According to Eq. (10), the standard deviations of the traditional DEA and AP-DEA models are 0.0679 and 0.2158, respectively. Therefore, the AP-DEA model is more accurate than the traditional DEA model, and the discrimination of the efficiency value is higher, which is more suitable for determining a direction for improvement.

The corresponding relaxation coefficients of the inputs and outputs of the ineffective DMUs can be obtained using Eq. (9). The redundancy of the input $\boldsymbol{g_i^-}$ (i = 1, 2, 3, 4, which represents the redundancy of the clustering results with the naphtha, hydrofoil, industry water and electricity.) and deficiency of the output $\boldsymbol{g}^+$ (which represents the deficiency of the ethylene yield.) of the ethylene production plant are shown in Table 5.

Table 5 and Fig. 8 show that the efficiency values of samples 2, 3, 4, 6, 7, 10, 12, 13, 18, 19, 20, 22, 25, 27, 29 and 30 are all 1. The other, inefficient DMUs can be improved by the relaxation coefficient. For example, the efficiency value of the eighth sample is 0.55. If the industry water is reduced by 0.16, the naphtha, hydrog. oil, electricity, and output are unchanged, and this sample can achieve effective production. Other ineffective samples can also undergo similar analysis to achieve effective production. As the improvement direction is specific, it also ensures more effective energy utilization. As shown in Fig. 6, the ethylene yield is approximately 564153.65 tons, but considering the improvement of the efficiency values of all DMUs, the ethylene yield can be increased by 118783.95 tons through calculating the relaxation coefficient of the AP-DEA model. Further, the energy-saving potential of the ethylene production plant can be improved by approximately 24.74%.

In complex ethylene production processes, carbon emissions are an indicator of concern. In addition, carbon emissions are related to the conservation of resources and environmental pollution from the ethylene production process. Therefore, determining methods to reduce carbon emissions from petrochemical industries is an important topic for all countries to study. Based on the production data from the same ethylene production plant in 2012 and 2013, the energy saving potential of each energy input and the carbon emission reduction potential are shown in Figs. 9–12, respectively. The feed, fuel and electricity consumptions of this plant in 2012 were 603034 tons, 191.56 GJ and 18.42 GJ as shown in Figs. 9–11, and the corresponding energy-saving potential can be calculated to be 124331 tons, 18.8 GJ, and 4.04 GJ, respectively, according to Eqs. (11) and (12). The feed, fuel, and electricity consumptions of this plant in 2013 were 531740 tons, 199.48 GJ and 23.41 GJ, and the corresponding energy-saving potentials were 129208 tons, 55.1 GJ and 11.32 GJ, respectively. The energy-saving potentials of the three main energy sources were increased by 17.81%, 3.68%, and 36.43% in 2013, respectively.

As shown in Fig. 12, the carbon emissions of this plant in 2012 and 2013 were 1852001 and 1633171 tons, respectively. The total carbon emission reduction of the ethylene plant was 10.04%. Moreover, the carbon emission reductions were 71293 and 247814 tons, respectively, and the potential for carbon emissions increased by 11.32% from 2012 to 2013. The carbon emission potential from 2012 to 2013 exhibited an increasing trend, indicating that the production efficiency in 2013 is not as good as that in 2012.

Table 5
The ineffective efficiency values and the relaxation coefficient of the ethylene production plant.

Sample	1	5	8	9	11	14	15	16	17	21	23	24	26	28
g_1^-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
g_2^-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
g_3^-	0.00	0.49	0.16	0.22	0.81	0.58	0.32	0.77	0.60	0.09	0.12	0.84	0.16	0.32
g_4^-	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
g^+	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
θ	0.84	0.35	0.55	0.25	0.53	0.60	0.56	0.57	0.59	0.54	0.56	0.58	0.66	0.60

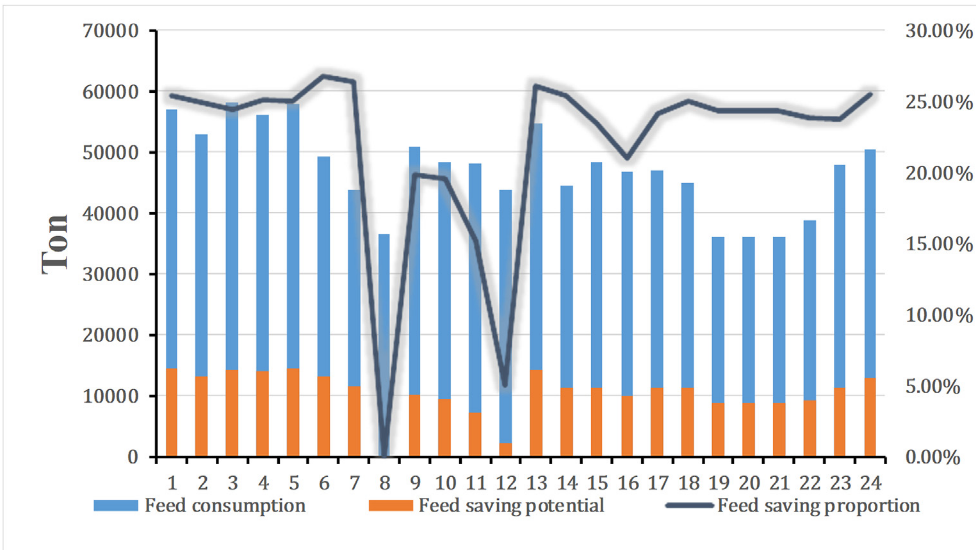

Fig. 9. Feed saving potential.

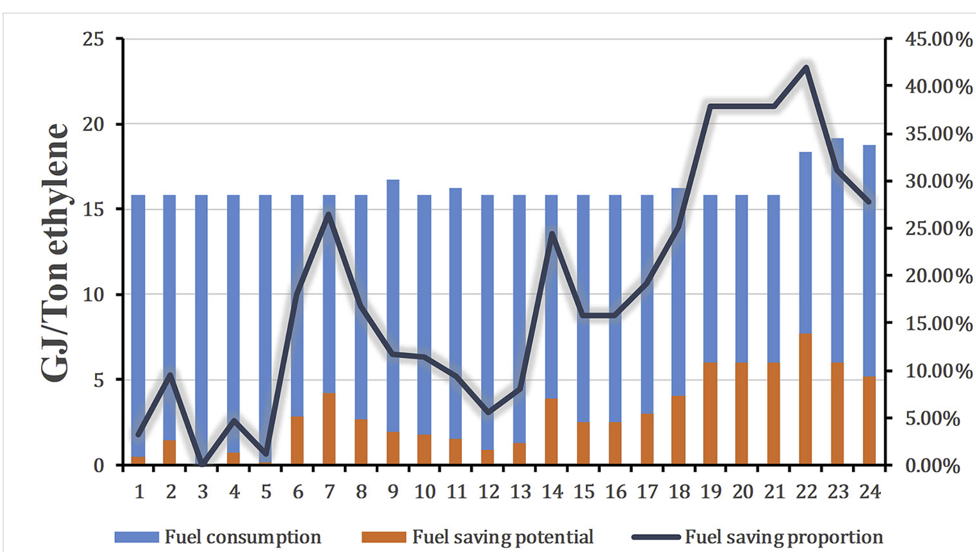

Fig. 10. Fuel saving potential.

Therefore, it is necessary to fulfill the carbon emission potential and make further suggestions for improving the efficiency [45]. The AP-DEA model mentioned above can produce DMU with greater diversity, and calculating the input-output relaxation coefficient can obtain a clearer improvement direction, which is beneficial for conserving energy and reducing emission in the environmental protection and production process.

4. Discussion

First, the AP-DEA model is proposed. The AP clustering algorithm can reduce a high number of input indicators of the DEA

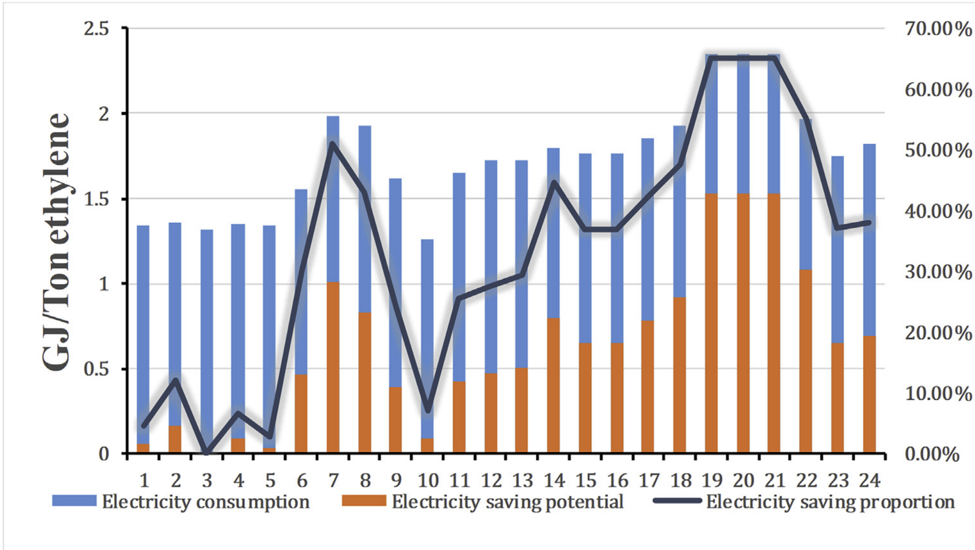

Fig. 11. Electricity saving potential.

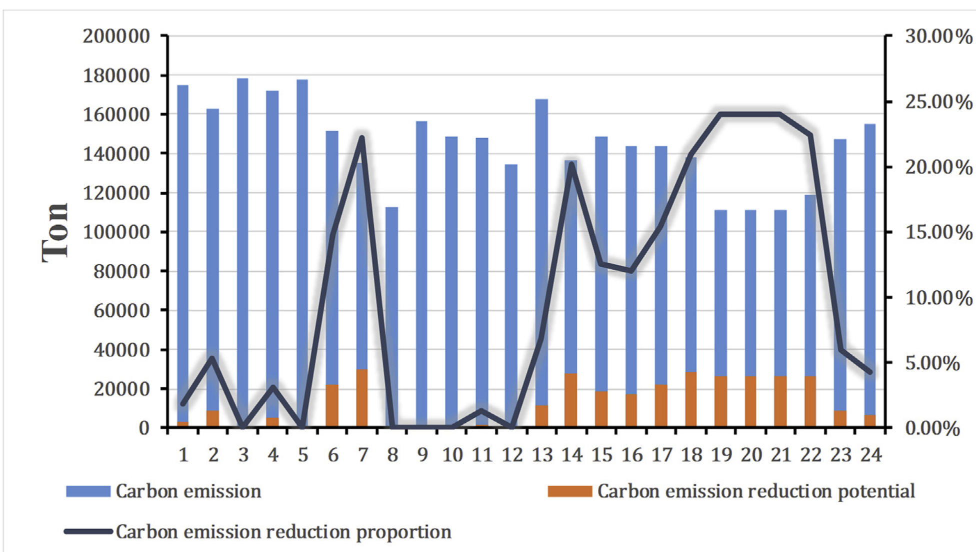

Fig. 12. Carbon emission reduction potential.

model. Moreover, by calculating and comparing the standard deviations of the DMUs, the AP-DEA model can achieve higher discrimination and better efficiency calculation results than the traditional DEA model.

Second, the proposed model is applied to analyze the energy efficiency and energy saving potential of complex petrochemical industries. By discriminating the multiple input indexes as cluster centers, the highly influential indices affecting the energy efficiency of complex petrochemical industries can be obtained based on the AP clustering algorithm, which are set as the inputs of the DEA model. Compared with the traditional DEA model, the number of DMUs with low discrimination of PTA production data and ethylene production data decreased by 25% and 72%, respectively. Therefore, the energy consumption of complex petrochemical production industry can be better improved. The energy saving and emission reduction capabilities of the ethylene and PTA production plants in complex petrochemical industries are then obtained by the proposed model. Furthermore, the energy savings of the PTA production and ethylene plant are improved by approximately 0.49% and 24.74%, and the carbon emissions reduction of the ethylene plant is 10.04%.

Third, the proposed model also has some deficiencies. The AP clustering process obtains clustering centers by continuous iteration, and the damping coefficient λ affects the AP algorithm. Therefore, particle swarm optimization (PSO) can be added to find the optimal damping coefficient λ and reduce the number of iterations, and reduce the computational complexity of the clustering process.

5. Conclusion

In this paper, an improved AP-DEA model is proposed. The input indicators are clustered by the AP clustering algorithm to screen high-influence factors. The clustering results are then used as the input of the DEA model to obtain the results of energy efficiency analysis and the related relaxation coefficient. Moreover, adjustments that can be made to the production configuration are obtained. Finally, the improved AP-DEA model is applied to analyze the energy efficiency, and reduce the energy consumption and carbon emissions of complex petrochemical industries. The seventeen input indices of the PTA production plant and the data of feed, fuel, steam, water and electricity of the ethylene production plant are clustered. The high-influence index and corresponding output indicators are then processed by the AP-DEA model. Compared with the traditional DEA model, the discrimination ability of the AP-DEA model is high. By using the AP-DEA model, the numbers of ineffective units in the PTA and ethylene production plants are reduced by 25% and 78%, respectively. According to the improved guidance given by the AP-DEA model, theoretically, the acetic acid consumption of the PTA production plant can be reduced by 3.89, and the ethylene yield of the ethylene plant can be increased by 118783.95 tons. Also, the carbon emissions of the ethylene plant in 2012 and 2013 are reduced by 71293 and 247814 tons, respectively.

In the further works, we will explore and integrate particle swarm optimization (PSO) and self-adaptive method to find the best damping coefficient λ and compare with the current works. Moreover, economic, environmental and human factors, as well as other pollutant data, will be taken into account to refine the input data and further improve the accuracy of the results. Furthermore, the proposed model can be widely used in energy efficiency evaluation and energy saving of other complex petrochemical industries.

Acknowledgement

This work is partly financial supported by National Key Research and Development Program of China (2017YFC1601800), National Natural Science Foundation of China (61603025 and 61533003) and Fundamental Research Funds for the Central Universities (XK1802-4).

References

[1] Aryanasl A, Ghodousi J, Arjmandi R, Mansouri N. Components of sustainability considerations in management of petrochemical industries. Environ Monit Assess 2017;189(6):266–74.
[2] Chen YX, Han YM, Zhu QX. Energy and environmental efficiency evaluation based on a novel Data Envelopment Analysis: an application in petrochemical industries. Appl Therm Eng 2017;119:156–64.
[3] Sueyoshi T, Yuan Y, Goto M. A literature study for DEA applied to energy and environment. Energy Econ 2017;62:104–24.
[4] Tesfamichael T, Jacobs B, Tabor N, Michel L, Currano E, Feseha M, Barclay R, Kappelman J, Schmitz M. Settling the issue of "decoupling" between atmospheric carbon dioxide and global temperature: $[CO2]_{atm}$ reconstructions across the warming Paleogene-Neogene divide. Geology 2017;45(11):999–1002.
[5] Tao R, Patel M, Blok K. Olefins from conventional and heavy feedstocks: energy use in steam cracking and alternative processes. Energy 2006;31(4):425–51.
[6] Dincer I, Rosen MA. A worldwide perspective on energy, environment and sustainable development. Int J Energy Res 2015;22(15):1305–21.
[7] Meng D, Shao C, Zhu L. Ethylene cracking furnace TOPSIS energy efficiency evaluation method based on dynamic energy efficiency baselines. Energy 2018;156:620–34.
[8] Song CX, Li MJ, Wen ZX, He YL, Tao WQ, Li YZ, Wei XY, Yin XL, Huang X. Research on energy efficiency evaluation based on indicators for industry sectors in China. Appl Energy 2014;134:550–62.
[9] Geng ZQ, Qin L, Han YM, Zhu QX. Energy saving and prediction modeling of petrochemical industries: a novel ELM based on FAHP. Energy 2017;122:350–62.
[10] Chan YL, Huang CF, Lin WC, Hong GB. Energy efficiency benchmarking of energy-intensive industries in Taiwan. Energy Convers Manag 2014;77(1):216–20.
[11] He Y, Liao N, Zhou Y. Analysis on provincial industrial energy efficiency and its influencing factors in China based on DEA-RS-FANN. Energy 2018;142:79–89.
[12] Han YM, Geng ZQ, Zhu QX, Lin XY. Energy consumption hierarchical analysis based on interpretative structural model for ethylene production. Chin J Chem Eng 2015;12.
[13] Geng ZQ, Gao HC, Wang YQ, Han YM. Energy saving analysis and management modeling based on index decomposition analysis integrated energy saving potential method: application to complex chemical processes. Energy Convers Manag 2017;145:41–52.
[14] Ang BW, Xu XY, Su B. Multi-country comparisons of energy performance: the index decomposition analysis approach. Energy Econ 2015;47:68–76.
[15] Lin BQ, Long HY. A stochastic frontier analysis of energy efficiency of China's chemical industry. J Clean Prod 2015;87(1):235–44.
[16] Shui HY, Jin XN, Ni J. Manufacturing productivity and energy efficiency: a stochastic efficiency frontier analysis. Int J Energy Res 2015;39(12):1649–63.
[17] Zhu QY, Wu J, Li XC, Xiong BB. China's regional natural resource allocation and utilization: a DEA-based approach in a big data environment. J Clean Prod 2016;142(2):809–18.
[18] Chen Y, Liu BS, Shen YH, Wang XQ. The energy efficiency of China's regional construction industry based on the three-stage DEA model and the DEA-DA model. KSCE J Civ Eng 2016;20(1):34–47.
[19] Iftikhar Y, Wang ZH, Zhang B, Wang B. Energy and CO_2 emissions efficiency of major economies: a network DEA approach. Energy 2018;147:197–207.
[20] Yang ZH, Wang D, Du TY, Zhang AL, Zhou YX. Total-factor energy efficiency in China's agricultural sector: trends, disparities and potentials. Energies 2018;11(4):853.
[21] Kang YQ, Xie BC, Wang J, Wang YN. Environmental assessment and investment strategy for China's manufacturing industry: a non-radial DEA based analysis. J Clean Prod 2018;175:501–11.
[22] Cook WD. Introduction to data envelopment analysis and its uses: with DEA solver software and references, by William W. Cooper; lawrence M. Seiford; Kaoru Tone. Interfaces 2006;36(5):474–5.
[23] Hashem O, Khatereh S, Ali E. An integrated fuzzy clustering cooperative game data envelopment analysis model with application in hospital efficiency. Expert Syst Appl 2018.
[24] Gong S, Shao C, Zhu L. Energy efficiency evaluation based on DEA integrated factor analysis in ethylene production. Chin J Chem Eng 2017;25(6):793–9.
[25] Li HY, He HZ, Wen YG. Dynamic particle swarm optimization and K-means clustering algorithm for image segmentation. Optik - Int J Light Electron Optic 2015;126(24):4817–22.
[26] Zhao JL, Qu H, Zhao JH, Luan ZR, Guo Y. Towards controller placement problem for software-defined network using affinity propagation. Electron Lett 2017;53(14):928–9.
[27] Zhang W, Wu XF, Zhu WP, Lu Y. Unsupervised image clustering with SIFT-based soft-matching affinity propagation. IEEE Signal Process Lett 2017;24(4):461–4.
[28] Liu J, Zhao XD, Xu ZH. Identification of rock discontinuity sets based on a modified affinity propagation algorithm. Int J Rock Mech Min Sci 2017;94:32–42.
[29] Zhao Y, Hong RC, Jiang JG. Visual summarization of image collections by fast RANSAC. Neurocomputing 2016;172:48–52.
[30] Frey BJ, Dueck D. Clustering by passing messages between data points. Science 2007;315(5814):972–6.
[31] Gui B, Yang XP. Research on parameters of affinity propagation clustering. Adv Technol Embed Multimed Hum Cent Comput 2014;260:637–44.
[32] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision-making units. Eur J Oper Res 1978;2(6):429–44.
[33] Jacobs R, Smith PC, Street A. Measuring efficiency in health care: analytic techniques and health policy. Cambridge University Press; 2006.
[34] Ruggiero J. Impact assessment of input omission on DEA. Int J Inf Technol Decis Mak 2005;4(3):359–68.
[35] Xu Y, Zhu QX. Research and implementation of decreasing the acetic acid consumption in purified terephthalic acid solvent system. Chin J Chem Eng 2008;16(4):650–5.
[36] Emrouznejad A, Amin GR. DEA models for ratio data: convexity consideration. Appl Math Model 2009;33(1):486–98.
[37] Siamak E, Mohammad SS. Introducing an application method for industries air pollutants emission control planning by preparing environmental flow diagram maps 2018;178:768–75.
[38] Han YM, Geng ZQ, Zhu QX, Wang Z, Cui YF. Energy consumption analysis and evaluation of petrochemical industries using an improved fuzzy analytic hierarchy process approach. J Intell Fuzzy Syst 2017;32(6):4183–95.
[39] Yousefi S, Soltani R, Saen RF, Pishvaee MS. A robust fuzzy possibilistic programming for a new network GP-DEA model to evaluate sustainable supply chains. J Clean Prod 2017;166:537–49.
[40] Han DQ, Han CZ, Deng Y, Yang Y. Weighted combination of conflicting evidence based on evidence variance. Acta Electron Sin 2011;39(3A):153–7.
[41] Han YM, Geng ZQ, Wang Z, Peng M. Performance analysis and optimal temperature selection of ethylene cracking furnaces: a data envelopment analysis cross model integrated analytic hierarchy process. J Anal Appl Pyrolysis 2016;122:35–44.
[42] Huang L, Krigsvoll G, Johansen F, Liu YP, Zhang XL. Carbon emission of global construction sector. Renew Sustain Energy Rev 2017;81:1906–16.
[43] Bie P, Fang X, Li Z, Wang Z, Hu J. Emissions estimates of carbon tetrachloride for 1992-2014 in China. Environ Pollut 2017;224:670–8.
[44] Chen WD, Wu FY, Geng WX, Yu GY. Carbon emissions in China's industrial sectors. Resour Conserv Recycl 2017;117:264–73.
[45] Lin B, Ahmad I. Analysis of energy related carbon dioxide emission and reduction potential in Pakistan. J Clean Prod 2017;143:278–87.