Journal of Environmental Management 205 (2018) 298–307

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Research article

Carbon emission analysis and evaluation of industrial departments in China: An improved environmental DEA cross model based on information entropy

Yongming Han [a, b], Chang Long [a, b], Zhiqiang Geng [a, b, *], Keyu Zhang [c, **]

[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] College of Economic, Capital University of Economic and Business, Beijing 100070, China

ARTICLE INFO

Article history:
Received 17 July 2017
Received in revised form
12 September 2017
Accepted 21 September 2017

Keywords:
Information entropy
Dynamic weights
Environmental DEA cross model
Monte Carlo simulation
Carbon emission reduction
Industrial departments

ABSTRACT

Environmental protection and carbon emission reduction play a crucial role in the sustainable development procedure. However, the environmental efficiency analysis and evaluation based on the traditional data envelopment analysis (DEA) cross model is subjective and inaccurate, because all elements in a column or a row of the cross evaluation matrix (CEM) in the traditional DEA cross model are given the same weight. Therefore, this paper proposes an improved environmental DEA cross model based on the information entropy to analyze and evaluate the carbon emission of industrial departments in China. The information entropy is applied to build the entropy distance based on the turbulence of the whole system, and calculate the weights in the CEM of the environmental DEA cross model in a dynamic way. The theoretical results show that the new weight constructed based on the information entropy is unique and optimal globally by using the Monte Carlo simulation. Finally, compared with the traditional environmental DEA and DEA cross model, the improved environmental DEA cross model has a better efficiency discrimination ability based on the data of industrial departments in China. Moreover, the proposed model can obtain the potential of carbon emission reduction of industrial departments to improve the energy efficiency.

1. Introduction

A large number of greenhouse gases of industrial departments caused by human activities have been emitted into the atmosphere, so the obviously global warming is much serious in the last decades. In all greenhouse gases, Carbon dioxide (CO_2) has the highest contribution rate and is the major factor that leads to the climate change and global warming in the longest survival period (Guo et al., 2011). The industrial departments are the important industries in the national production life. Meanwhile, the industrial departments are also the high pollution and high carbon emissions industries. Therefore, analyzing and evaluating the carbon emission of the industrial departments is imperative.

In many environment efficiency evaluation methods, Data Envelopment Analysis (DEA) method is widely used as a nonparametric programming method because of its significantly convenient in practice and economic background, especially in the field of environment performance evaluation of enterprise micro-level (Zhou et al., 2008a,b). Since Charnes et al. proposed the DEA method in 1978 (Charnes et al., 1978), the DEA method has been widely used in the evaluation of relative efficiency of industrial departments. The advantage of the DEA method is that the input and output variables can be of high dimension and it is unnecessary to know the function relationship between variables in the model construction. However, an important hypothesis in the traditional DEA model is that the output variables must be well and positive benefit index, and possess strongly disposable characteristic (Ramanathan, 2003), namely, the greater the input, the greater the output. Moreover, the evaluation of the environmental efficiency of the industry should not only take into account the expected output (positive utility) variables created in production activity but also the undesirable output (negative utility) variables produced with

* Corresponding author.College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China.
** Corresponding author.
E-mail addresses: gengzhiqiang@mail.buct.edu.cn (Z. Geng), zkyjesu@126.com (K. Zhang).

https://doi.org/10.1016/j.jenvman.2017.09.062

the expected output. Therefore, how to technically deal with undesirable output variables with negative externalities is the key to evaluate the environmental efficiency. The DEA model has been formally applied to evaluate the environmental efficiency of the industry after Chambers et al. proposed the assumption that pollutants and other unexpected outputs possess weakly disposable (Chambers and Fare, 1996), and the production possibility was concentrated disposal with the expected output (radial hypothesis). Hernandez et al. studied 42 large furniture manufacturers in Spain and found that that reducing pollutant emission could only be at the expense of expected outputs (Hernandez-Sancho et al., 2000). Zhou et al. studied and compared the environmental performance indices of several major countries in the world based on different production technology assumptions (Zhou et al., 2008a,b). Hua et al. used a three-stage DEA model to evaluate the CO_2 efficiency of each region in China (Huan et al., 2013). Hernández-Sancho et al. applied a non-radial DEA method to analyze energy efficiency indices for the sampling of wastewater treatment plants in Spain, and get the potential savings in economic terms and carbon emissions to provide the guidance for the efficiency improvement (Hernández-Sancho et al., 2011). Zhu et al. studied the DEA combined with life-cycle environmental impacts of products for eco-efficiency evaluation, examined the eco-efficiencies of ten comparable pesticides, and found out that the network DEA method could distinguish the differences in the eco-efficiency of products at the different stages (Zhu et al., 2014).

However, both the traditional DEA model and the environment DEA model are based on maximize their own Decision Making Unit (DMU) efficiency as the objective function, and lack of consideration of other DMUs, so the evaluation is too subjective. In addition, when there are multiple active units in the DMUs, the advantage and disadvantage cannot be further evaluated together. Therefore, Sexton proposed the DEA cross model, which evaluated its own efficiency values under the full consideration of other DMU situations (cross-evaluation) by calculating the cross-reciprocal matrix, and then took the arithmetic average to get comprehensive evaluation scores (Sexton and Silkman, 1986). Chen et al. compared the spatial eco-efficiency of 11 provinces in western China from 2005 to 2009 based on the DEA cross model (Chen and Xu, 2012). Xu et al. introduced the virtual unit to evaluate the urban efficiency of the prefecture-level cities in Gansu Province by using the DEA cross model (Xu and Xu, 2013).

Although the DEA cross model can take the inputs and outputs of all DMUs into account, the optimal weights of the DEA cross model are not unique (Despotis, 2002). For different assumptions on constraints (such as the strategy of confrontation or mutual assistance between DMUs), the optimal weight of some DMUs will have multiple solution sets, resulting in a completely different efficiency value for the DEA cross model. In order to solve above problems, Wang proposed the neural network method integrating the DEA to recalculate the cross evaluation matrix (Wang and Chin, 2010). Guo et al. proposed the fuzzy DEA method and got the comprehensive evaluation score by the fuzzy processing of DMUs (Guo and Wang, 2012; Hu and Xu, 2014). Johanshahloo et al. used symmetric weights to assign the cross evaluation matrix (CEM) (Jahanshahloo, 2011). These existing methods are mostly focused on improving the accuracy of the CEM or performing the fuzzy calculation on the efficiency values to eliminate the subjective problem caused by constructing the CEM.

Although there are many correction way to measure the weight of the cross-reciprocal index, the DEA cross model have introduced other parameters and man-made assignments, and recalculated the weights (Ramon and Ruiz, 2010). Moreover, the reason for the inaccurate cross-evaluation is that assigning the weight of each cross-evaluation index is equal to the self-evaluation index, and the information contained in the entire cross matrix is not taken into account. In addition, the fixed weights tend to lead to the lack of Pareto optimal (Wu et al., 2009). In order to solve this problem, this paper proposes the environmental DEA cross model based on the information entropy. The information entropy in the CEM is used to construct the entropy distance between the self-evaluation index and the cross-evaluation index. Meanwhile, the dynamic weight assignment of each cross-evaluation unit in the CEM is made by the distance of the entropy distance. Moreover, we use the Monte Carlo simulation to prove that the weight constructed based on the information entropy is globally optimal. Finally, the proposed method is applied to analyze and evaluate the carbon emission reduction of industrial departments in China by the data of Input-Output table. The results show that the proposed method can obtain the potential of carbon emission reduction of industrial departments to improve the energy efficiency and reduce carbon emissions.

2. Model theory and construction

2.1. Environmental DEA cross model

As mentioned in the introduction, the traditional DEA model does not consider the undesirable output, which is contrary to the actual production situation, such as the coal power generation will inevitably produce carbon dioxide, and the paper industry is bound to produce sewage. In production activities, it is desirable that the undesirable output is as little as possible, which is contrary to the general DEA model assuming that the output has strongly disposable. Therefore, if the output does not distinguish between the undesirable output and the expected output, the whole efficiency evaluation system will be distorted and invalid.

There are N objects in the overall environmental performance evaluation system, each object i is a decision unit (DMU_i). Its input factors, expected outputs and undesirable outputs are:

$X = (X_1, X_2, \ldots, X_M), X \in R_+^M$; That is, a total of M kinds of elements input factors.
$Y = (Y_1, Y_2, \ldots, Y_S), Y \in R_+^S$; That is, a total of S kinds of expected outputs.
$Z = (Z_1, Z_2, \ldots, Z_K), Z \in R_+^K$; That is, a total of K kinds of undesirable outputs.

Then the production technology can be expressed by production possibility set P, $P(X) = \{(Y, Z)$: input X can produce Y and $Z\}$.

According to setting of the environmental DEA by Tyteca (Tyteca, 1996), the expected output Y and the undesirable output Z are defined as possessing weakly disposable:

If $(Y, Z) \in P$, then $(\alpha Y, \alpha Z) \in P$, $\forall \alpha \in (0, 1]$. In other words, the only way to reduce the undesirable output is to reduce the number of expected outputs with same percentage. If the production technology P presents the characteristic of constant returns to scale (CRS), the environment DEA technology can be defined as

$$\begin{aligned} P(x) = \Bigg\{ (y, z) : & \sum_{i=1}^{N} \lambda_i x_{im} \le x_m, \ m = 1, 2, \ldots, M \\ & \sum_{i=1}^{N} \lambda_i y_{is} \ge y_s, \ s = 1, \ 2, \ldots, S \\ & \sum_{i=1}^{N} \lambda_i z_{ki} = z_k, \ k = 1, \ 2, \ldots, K \\ & \lambda_i \ge 0, \ i = 1, \ 2, \ldots, N \Bigg\} \end{aligned} \tag{1}$$

where, M kinds of elements are inputs, which can produce S kinds of expected outputs and K kinds of undesirable outputs. Wherein λ_i is the combined ratio of the i-th decision unit in the reconstructed effective decision unit combination. Then, according to the methodology proposed by Fare et al. (Fare and Grosskopf, 2004), the environment DEA technique in Eq. (1) can be transformed into

$$\begin{aligned}
&\min \theta \\
&\text{s.t} \sum_{i=1}^{N} \lambda_i x_{im} \le x_{m0} \quad m = 1, 2, M. \\
&\sum_{i=1}^{N} \lambda_i y_{is} \ge y_{s0} \quad s = 1, 2, S. \\
&\sum_{i=1}^{N} \lambda_i u_{ik} = \theta u_{k0} \quad k = 1, 2, \ldots, K \\
&\lambda_k \ge 0, \quad i = 1, 2, \ldots, N
\end{aligned} \tag{2}$$

From Eq. (2), the above-mentioned environmental DEA model focuses on the adjustment of undesirable output Z. Under the assumption of production technology CRS, the less the undesirable yield Z_i of DMU i, the higher production efficiency of the unit. Therefore, the environment DEA model only take the environmental performance of each decision-making unit into consideration, that is, the efficiency value θ represents the environmental performance index (EPI) of industry. If the value of θ is larger, indicating that the industry has no compressible space in reducing the undesirable output, the environment performance of industry will be determined better, and vice versa.

Eq. (2) is converted into a dual problem as the following.

$$\begin{aligned}
&\text{Max } E = \sum_{k=1}^{K} \alpha_k Y_k - \sum_{m=1}^{M} \beta_m X_m \\
&\text{s.t.} \sum_{k=1}^{K} \alpha_{ki} Y_{ki} - \sum_{m=1}^{M} \beta_{mi} X_{mi} - \sum_{s=1}^{S} \gamma_{si} Z_{si} \le 0 \\
&\sum_{s=1}^{S} \gamma_{si} Z_{si} = 1, \quad i = 1, 2, \ldots, N \\
&\alpha_k,\ \beta_m; \quad \gamma_z \text{ is a free variable}
\end{aligned} \tag{3}$$

The result of the target E in Eq. (3) should be the same as the result of θ in Eq. (2), which represents the environmental performance efficiency of the DMU. We can get the optimal weight $W^* = (\alpha^*, \beta^*, \gamma^*)$ of input and output vectors X, Y and Z by Eq. (3), wherein W^* is the vector of $m + k + s$ dimension.

Definition 1: Based on the definition of the cross-evaluation of the general DEA model by Sexton (Sexton and Silkman, 1986), we can define $E_{ij}^{E-cross}$ as environmental cross-evaluation index by Eq. (4).

$$E_{ij}^{E-cross} = \frac{Y_k^T \alpha_k^*}{X_m^T \beta_m^* - Z_s^T \gamma_s^*} \tag{4}$$

If the EPI is calculated by the department j through its own optimal weight ω_j^*, which is called environmental self-evaluation index, recorded as E_{jj}. The EPI is calculated by the department j through the optimal weight ω_i^* of the department i, which is called the environmental cross-evaluation index, record as E_{ij}. The generalized CEM is shown in the following:

$$CEM = \begin{array}{c} DMU_1 \\ DMU_2 \\ \ldots \\ DMU_n \end{array} \begin{pmatrix} E_{11} & E_{12} & \ldots & E_{1n} \\ E_{21} & E_{22} & \ldots & E_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ E_{n1} & E_{n2} & \ldots & E_{nn} \end{pmatrix}$$

Finally, the cross-evaluation indices can form an Environmental CEM. As shown in Table S1, diagonal elements are general environmental evaluation indices, and non-diagonal elements are environmental cross-evaluation indices. Each DMU_i can calculate its arithmetic mean value by the column direction, which is

$$\overline{E}_i = \frac{\sum_{i=1}^{N} E_{ij}}{N} \tag{5}$$

wherein, $\overline{E}_i$ is cross-evaluation value of this department under the traditional definition.

2.2. Cross evaluation method based on the information entropy

The concept of Entropy is derived from statistical physics and thermodynamics, and Shannon introduced its concept into information theory and named it as Information Entropy (Shannon, 1948), which aims to describe a system of disorder and uncertainty by the theory of probability and statistics. The information theory is that the smaller the degree of disorder of a system, the lower the information entropy, and the more chaotic of a system, the higher the information entropy. Therefore, based on the theory of the information entropy, this paper constructs the distance of the entropy to measure the degree of the self-evaluation index and the cross-evaluation index, and then give different weights of each index, and recalculate the new environmental performance of each unit fraction.

2.2.1. Definition of discrete entropy

According to Table S1, the environmental CEM $E = (E_{ij}) \in R_+^{n \times n}$, whose elements E_{ij} represent the environmental performance efficiency of DMU_j obtained with optimal weight ϖ_i of DMU_i. If we consider E as an indeterminate system, the discrete random variable E_{ij} represents a certain state, we can obtain the following definition:

Definition 2: According to the definition of Shannon, the information entropy of the whole environment cross-evaluation system can be expressed as

$$I = -\kappa \cdot \sum_{j=1}^{N} \sum_{i=1}^{N} D_{ij} \ln D_{ij} \quad i, j = 1, 2, \ldots, N \tag{6}$$

wherein κ represents the Bozeman coefficient, which is set to 1 in this paper; $D_{ij} = E_{ij} / \sum_{i=1}^{N} E_{ij}$ is normalized standard values, and $\sum_{i=1}^{N} D_{ij} = 1$, $0 \le D_{ij} \le 1$.

Thus, the information entropy of each cross-evaluation index E_{ij} is $I_{ij} = -D_{ij} \ln D_{ij}$, the information entropy of each unit j is $I_j = -\sum_{i=1}^{N} D_{ij} \ln D_{ij}$. It is easy to prove that definition of Eq. (6) satisfies three basic properties of the information entropy: non-negativity, additivity and extremality (when $D_{1j} = D_{2j} = \ldots = D_{nj} = 1/N$, there is the maximum value of the entropy of $I_j = \ln Nj = 1, 2, \ldots, N$). Thus, the definition of the information entropy of the cross-evaluation E is established.

2.2.2. Constructing an entropy model

Based on the self-evaluation index E_{jj} of the unit j, the higher consistency of the cross-evaluation index E_{ij} of any other unit i for unit j and E_{jj}, the index is more useful for evaluating the DMU. On the other hand, if the consistency of two indices is lower, then the cross-evaluation index is less effective for evaluating the own DMU. We continue to construct the distance of the entropy by the definition of the information entropy in the last section to measure the coherence of the self-evaluation index and the cross-evaluation index in the CEM. If the distance of the entropy is smaller, the uncertainty between two indices is also smaller, the consistency between the cross-evaluation index and the self-evaluation index is higher, and the accurate evaluation system will be more, and vice versa. By summing up the information contained in the total distance of the entropy, we can get a more reasonable weight for each evaluation index, rather than a simple equal weight (1/N) assignment.

Definition 3: Under the environmental DEA cross model, based on the idea of Wald statistics (Wald, 1945), the equation of the distance of the entropy of the self-evaluation index and the cross-evaluation index of any department j can be defined as

$$d_{ij} = (I_{ij} - I_{ii})^T \sum\nolimits_{ij}^{-1} (I_{ij} - I_{ii}),\ j = 1, ..., N \tag{7}$$

Wherein I_{ij} and I_{ii} are the entropy of cross-evaluation indices and self-evaluation indices, respectively. $\sum_{ij} = COV(I_{ij} - I_{ii})$ is the covariance matrix of the corresponding variable. When the value of d_{ij} is larger, the higher the degree of disorder in the system, the lower the consistency between the self-evaluation index of DMU_j and the cross-evaluation index measured by other department i, the less effect of this indicator on decision, and vice versa.

Definition 4: By the idea of Generalized Least Square (GLS), this paper defines that the most reasonable evaluation of its efficiency value for any department j should be the minimum square of sum of the distance of the entropy of the weighted information

$$\begin{aligned} &\min M_j = \sum_{i=1}^{n} (d_{ij}\lambda_i)^2 \\ &\text{s.t.} \sum_{i=1}^{n} \lambda_i = 1 \\ &\lambda_i > 0, \quad i = 1, 2, ..., n \end{aligned} \tag{8}$$

wherein λ_i is the weight corresponding to the different distance of the entropy.

Based on the same idea, for all N units in CEM, the optimal weight λ_i^* should satisfy the objective function of minimizing the square distance for each department at the same time. That is to say, the multi-objective programming problem is solved under the same restriction condition:

$$\begin{aligned} &\begin{cases} \min M_1 = \sum_{i=1}^{N} (d_{i1}\lambda_i)^2 \\ \min M_2 = \sum_{i=1}^{N} (d_{i2}\lambda_i)^2 \\ \quad \cdots \\ \min M_n = \sum_{i=1}^{N} (d_{in}\lambda_i)^2 \end{cases} \\ &\text{s.t.} \sum_{i=1}^{N} \lambda_i = 1 \\ &\lambda_i > 0, \quad i = 1, 2, ..., N \end{aligned} \tag{9}$$

By linearizing weights, the multi-objective problem can be transformed into a single programming problem, which is

$$\begin{aligned} &\min M = \sum_{j=1}^{N} \sum_{i=1}^{N} d_{ij}^2 \lambda_i^2 \\ &\text{s.t.} \sum_{i=1}^{N} \lambda_i = 1 \\ &\lambda_i > 0, \quad i = 1, 2, ..., N \end{aligned} \tag{10}$$

Eq. (10) is a typical quadratic optimization problem. By introducing the Lagrangian operator, we can calculate the analytic expression of weight λ_i.

$$\lambda_i^* = \left\{ \left[\sum_{i=1}^{N} \left(\sum_{j=1}^{N} d_{ij}^2 \right)^{-1} \right] \left[\sum_{j=1}^{N} d_{ij}^2 \right] \right\}^{-1} \tag{11}$$

Theorem 1: When the following two conditions are satisfied: The objective function in Eq. (10) is a convex function, and the restricted constraint set S is linear and nonempty set, if the optimal solution λ_i^* exists, it is unique and is the global optimal.

Prove: The optimization problem shown in Eq. (10), the objective function $M = \sum_{j=1}^{n} \sum_{i=1}^{n} d_{ij}^2 \lambda_i^2$ is a quadratic continuous function, so there is a second order partial derivative, and the Hessian matrix H is

$$H = \begin{bmatrix} \frac{\partial^2 M}{\partial \lambda_1^2} & \frac{\partial^2 M}{\partial \lambda_1 \lambda_2} \cdots\cdots\cdots & \frac{\partial^2 M}{\partial \lambda_1 \lambda_n} \\ \frac{\partial^2 M}{\partial \lambda_2 \lambda_1} \cdots\cdots\cdots\cdots & & \frac{\partial^2 M}{\partial \lambda_2 \lambda_n} \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots & \cdots\cdots \\ \frac{\partial^2 M}{\partial \lambda_n \lambda_1} & \frac{\partial^2 M}{\partial \lambda_n \lambda_2} \cdots\cdots\cdots & \frac{\partial^2 M}{\partial \lambda_n^2} \end{bmatrix}$$

wherein the non-diagonal elements $\frac{\partial^2 M}{\partial \lambda_i \lambda_j} = 0,\ \forall i \neq j$; While the diagonal elements $\frac{\partial^2 M}{\partial \lambda_i^2} = 2\sum_{j=1}^{n} d_{ij}^2 > 0,\ i = 1, 2, ..., n$, According to the definition of the positive definite matrix, H is a positive definite matrix. Thus, the objective function M is a convex function. In addition, the constraint set S is linear and must satisfy the condition: $y_1 \in S$, $y_2 \in S$, then $(1-\alpha)y_1 + \alpha y_2 \in S$, for $\forall \alpha \in [0, 1]$, and S is clearly a nonempty set. Thus, the set S is a convex set. In summary, the whole optimal problem is a typical convex programming, and as long as the solution exists, the solution is the global optimal solution. In addition, in the general DEA cross model, the equal weight assignment $(\lambda_1 = \lambda_2 = ... = \lambda_n)$ is only a special case of optimal weight λ_i^*.

Through the optimal weight λ_i^*, we can define the environmental cross-evaluation index based on the information entropy, that is, the comprehensive environmental evaluation index

$$E_j^{enttophy} = \sum_{i=1}^{N} \lambda_i E_{ij} \quad j = 1, 2, ..., N \tag{12}$$

The flowchart of the improved environmental DEA cross model based on the information entropy can be shown as Fig. 1.

3. Case study: carbon emission reduction analysis and evaluation of industrial departments in China

The greenhouse effect problem caused by carbon dioxide emissions is a hot topic in academia in recent decades. In general, there are three indicators for measuring carbon emissions of industry (department): carbon intensity (CO2/GDP), energy intensity (energy/GDP) and carbon factor (CO2/Energy). But above indicators, either only energy consumption problem is considered, or only economic output problem is considered.

Therefore, the EPI (Zhou et al., 2008a,b) of industrial departments will be measured through the environmental DEA cross model in this paper, and the results will be compared.

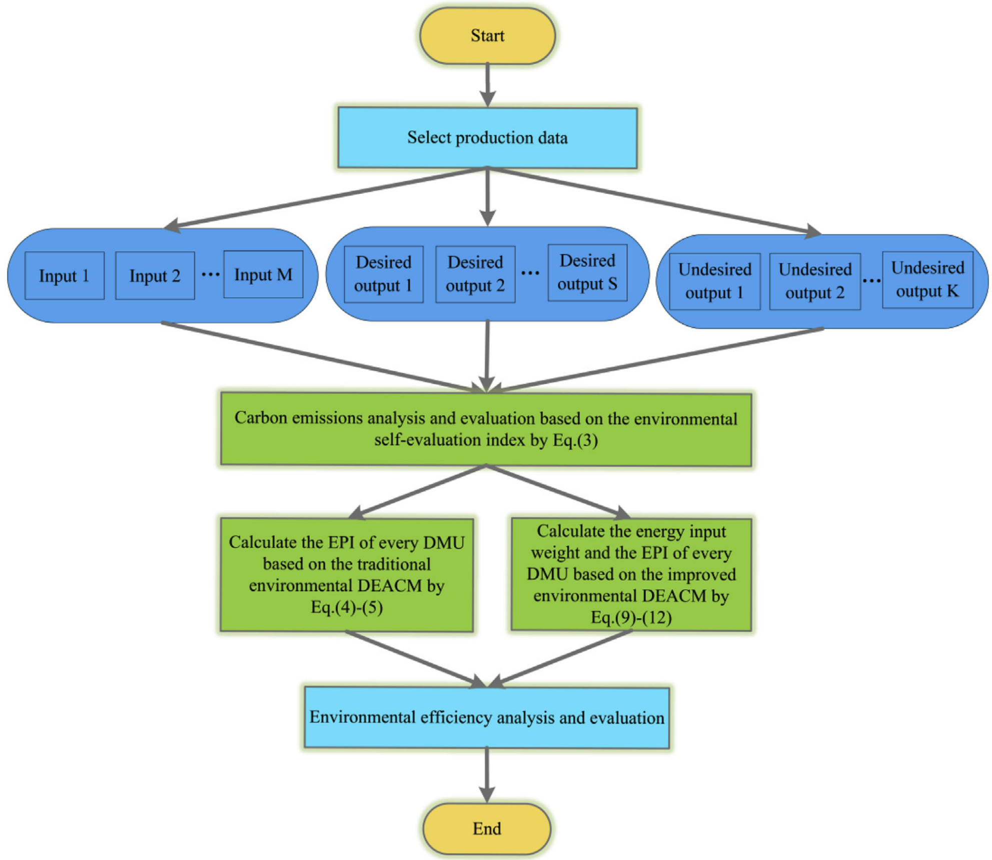

Fig. 1. The flowchart of the improved environmental DEA cross model based on the information entropy.

3.1. Data selection

The data are mainly from the Beijing Municipal Bureau of Statistics newly released "Beijing input and output table 2012" (Beijing Municipal Bureau of Statistics, 2015) and "China Energy Statistics Yearbook 2013" (National bureau of Statistics of China, Beijing, 2014). The number of departments studied is 42, namely $n = 42$. Which includes the number of the first industry department is 1, the number of secondary and tertiary departments respectively are 27, 14 (Table S1).

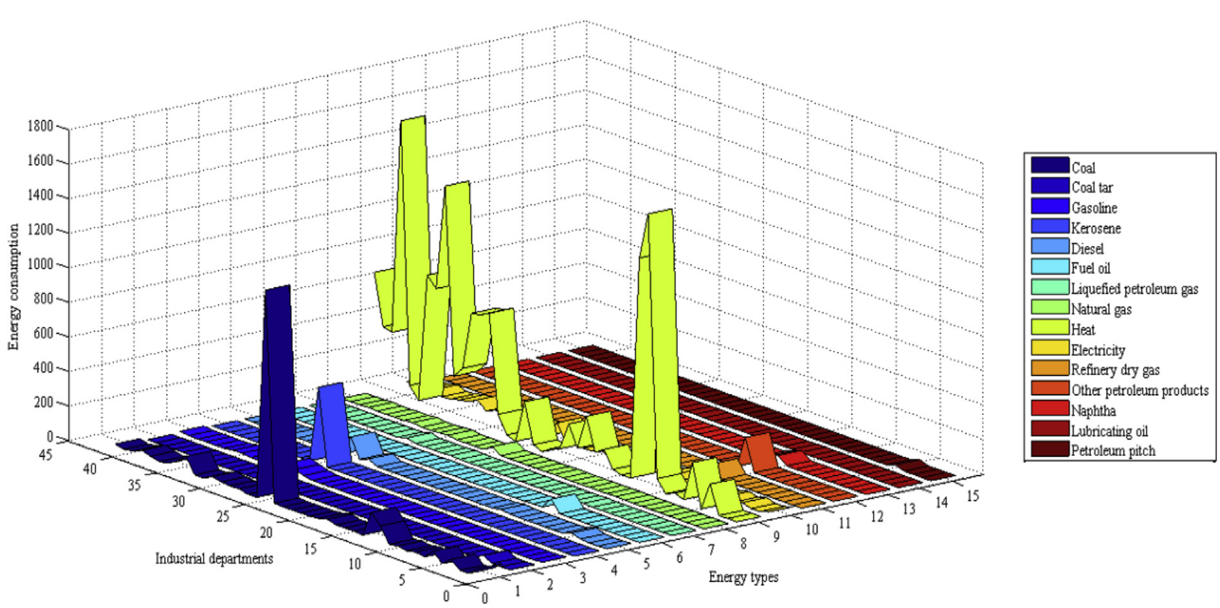

Fig. 2. 15 kinds of major energy use.

Table 1
Energy emission coefficient (Unit: 10,000 tons CO2/tonne).

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

Table 2
Input and output data.

	Variable name	Sample size	unit
Input vector (X)	Energy consumption (including 15 kinds of energy such as coal and oil)	42	Million tons, billion cubic meters and so on
Expected output vector (Y)	Total output	42	Million Yuan
Undesirable output vector (Z)	Carbon dioxide emissions	42	Million tons

Among them, the input factor X is set as all energy consumption of each department in this paper. Here, we only consider the production of fossil energy in the production process, such as coal, coke, gasoline, kerosene, etc. There are 15 kinds of energy consumption of 42 departments as shown in Fig. 2.

Expected output Y is total output of each department.

The undesirable output Z is the total amount of carbon dioxide gas emitted directly by each department when 15 kinds of fossil fuels are consumed in production activities. Where the carbon conversion coefficient of each energy source is mainly based on the combustion calorific value method provided by (Intergovernmental Panel on Climate Change, 2006) IPCC as shown in Table 1 (2006 IPCC Guidelines for National Greenhouse Gas Inventories, 2006).

As the input of the energy of some departments is zero, which does not meet the basic requirements for the data of the DEA model, so the pre-data conversion is necessary.

Let x_{ij} denote the i-th element of j-th variable. The converted data becomes

$$x_{ij}^{*} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \tag{13}$$

Among them, $x_j^{\max} = \max\{x_{1j}, x_{2j}, \ldots, x_{nj}\}$, $x_j^{\min} = \min\{x_{1j}, x_{2j}, \ldots, x_{nj}\}$, $j = 1, 2, \ldots, 15$.

The input factor X, the expected output Y and the undesirable output Z are shown in Table 2.

3.2. Monte Carlo simulation

In order to avoid the randomness of the results of an observed data, Monte Carlo simulation (Pereira et al., 2014; Silva et al., 2015) will be used in this section to test whether the results of the three DEA models will be significantly different.

Based on the data of "Beijing input and output table 2012" as reference (Beijing Municipal Bureau of Statistics, 2015), the stochastic perturbation factor ε, $\varepsilon \sim N(0,1)$ is introduced into the intermediate consumption matrix A, that is, ε follows the standard normal distribution with average of 0 and variance of 1. Thus, each element a_{ij} of matrix A will become the a_{ij}^{*} of the newly generated matrix A^{*} by using the Eq. (14).

$$a_{ij}^{*} = a_{ij} + \varepsilon, \ \mathrm{i, \ j} = 1, \ 2, \ldots, \ 42 \tag{14}$$

Based on the newly generated matrix A^{*}, the new input and output vectors (X^{*}, Y^{*}, Z^{*}) can be obtained by the simple calculation. Then, we calculate the EPI, the environment cross index and the comprehensive environmental index (based on the information entropy) of each department, respectively. The whole process should be simulated 1000 times based on Monte Carlo approach (Pereira et al., 2014; Silva et al., 2015) as shown in Fig. 3.

Finally, the Wilcoxon test (Wilcoxon, 1945) is used to test the significant difference of two scenarios as shown in Table 3. And compared the order of the 1000 sets of simulated data in each environment DEA model to the EPI ranking of original data, the stability of the EPI are checked after the introduction of the error term.

Note: The data in the table represents the significant number of Z statistic ratios.

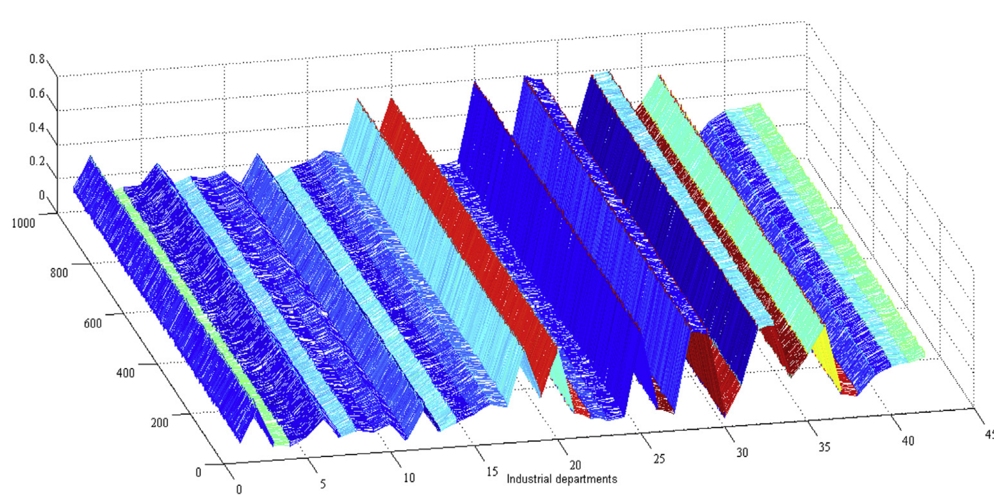

Fig. 3. The simulated data with 1000 times based on Monte Carlo simulation.

Table 3
The results by Monte Carlo simulation.

	Environment DEA	DEA cross model	DEA cross model based on the information entropy
intra-group comparison	4.3%	17.3%	1.1%
comparison among groups	94.2%	98.9%	

First, from the comparison between groups, the results of the general environmental DEA model show that there are 43 times Z statistics significant by Wilcoxon test in 1000 simulation, that is, there is 4.3% chance that the industry ranking is affected by the random noise, and significant changes will occur, and the performance is relatively stable. However, the results of the environment DEA cross model show that there are 173 times Z statistics significant, which is much worse than the general environmental DEA model. Finally, there are only 11 times significantly differences of the industry ranking based on the DEA cross model based on the information entropy, which is the least affected by the noise in the three models, and the conclusion is the most stable. Moreover, compared with the EPI of the environment DEA model, there are 942 times significant differences of the industry ranking in 1000 times simulations of the environment DEA cross model based on the information entropy, and compared with the DEA cross model, there are 989 times significant differences in 1000 times simulations.

From the simulation results, the EPI calculated by the environment DEA cross model based on the information entropy possess a strong stability for interference of data randomness. Meanwhile, from the results of the industry efficiency ranking, there are significant differences in results of the DEA cross model based on the information entropy compared with the environment DEA model a DEA cross model.

Therefore, the improved environmental DEA cross model based on the information entropy is best for dealing with the efficiency.

3.3. *Carbon emission reduction analysis and evaluation of industrial departments*

First, the self-evaluation of environmental performance indices of the 42 departments included in the "Beijing Input-Output Table 2012" is conducted based on the general environmental DEA model (based on the production technology CRS hypothesis). And then the environment DEA cross model was used to evaluate the 42 departments. The results of the two models are shown in Fig. 4.

It can be seen from Fig. 4 that the blue thin line represents industry efficiency scores of the general environmental DEA, there are three departments of the EPI reached the DEA effective (efficiency value reached 1), which are production and supply of electricity and heat, construction industry and financial industry, respectively (department codes are 25, 28 and 33, respectively), which is consistent with the original intention of the cross-evaluation. When there are many departments to achieve the DEA effective, we cannot further distinguish its merits. The red dot-like thick line represents the results of the cross-evaluation, we can see that the efficiency value of industry self-evaluation dropped a lot after an average of cross-evaluation indices and self-evaluation indices, where production and supply of electricity and heat, construction industry and financial industry are still top three, but it has been able to distinguish among their ranking (from high to low are 33, 25 and 28, respectively).

Next, the information entropy value of each element in the system is solved through the environment cross evaluation matrix in Eq. (6). The values of diagonal elements of the CEM are self-evaluation efficiency values obtained by the general environmental DEA model as shown in Fig. 5.

According to Eq. (7), the distance of entropy between all elements in the CEM with their own self-evaluation index is obtained. By solving the optimization problem defined by the distance of entropy in Eq. (11), we can work out the optimal weight of all 42 departments in the CEM based on the information entropy. The results are shown in Table 4.

Finally, according to Eq. (12), we can work out comprehensive EPIs based on the information entropy of 42 departments as shown

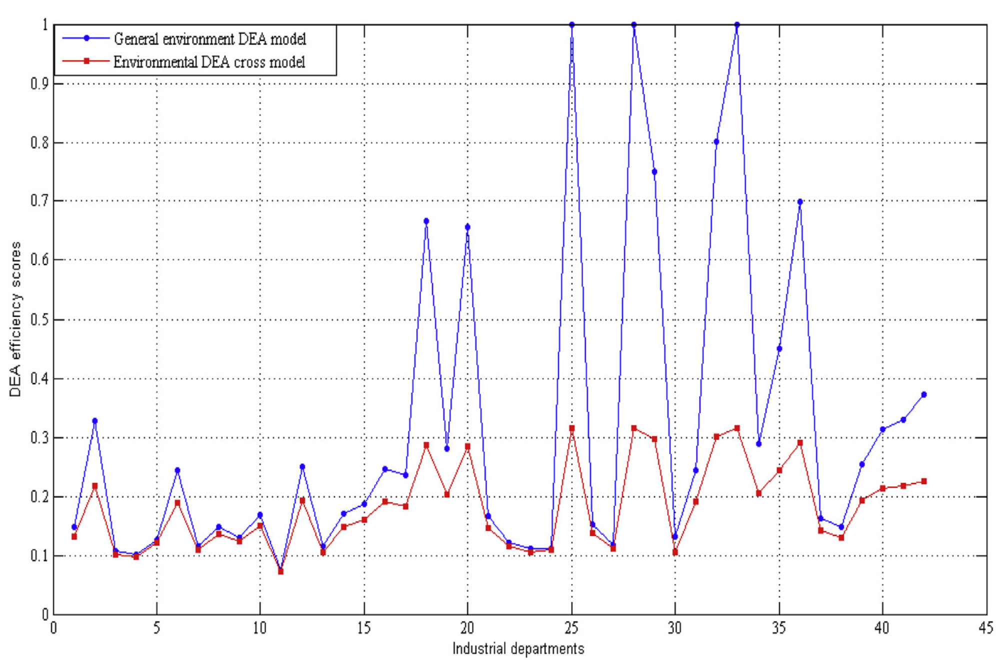

Fig. 4. Comparison of self-evaluation results and cross-evaluation results.

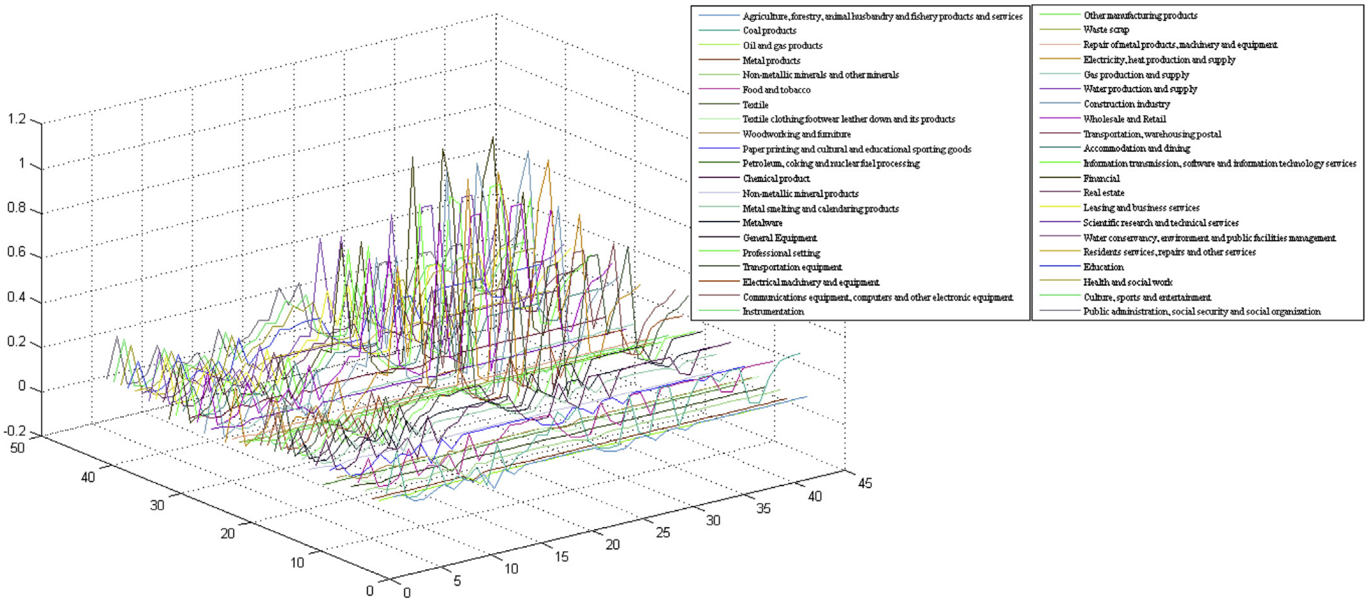

Fig. 5. The CEM of 42 departments.

Table 4
Entropy weights (unit: 10^{-3}).

$\lambda_1 = 6.66$	$\lambda_2 = 18$	$\lambda_3 = 5$	$\lambda_4 = 4.7$	$\lambda_5 = 5.7$	$\lambda_6 = 11.6$	$\lambda_7 = 0.53$
$\lambda_8 = 6.6$	$\lambda_9 = 5.9$	$\lambda_{10} = 7.5$	$\lambda_{11} = 3.8$	$\lambda_{12} = 12$	$\lambda_{13} = 5.3$	$\lambda_{14} = 7.6$
$\lambda_{15} = 8.4$	$\lambda_{16} = 11.8$	$\lambda_{17} = 11.1$	$\lambda_{18} = 11.5$	$\lambda_{19} = 14.1$	$\lambda_{20} = 11.2$	$\lambda_{21} = 7.4$
$\lambda_{22} = 5.6$	$\lambda_{23} = 5.1$	$\lambda_{24} = 5.2$	$\lambda_{25} = 29.6$	$\lambda_{26} = 6.8$	$\lambda_{27} = 5.4$	$\lambda_{28} = 29.6$
$\lambda_{29} = 12.2$	$\lambda_{30} = 5.9$	$\lambda_{31} = 11.7$	$\lambda_{32} = 10$	$\lambda_{33} = 29.6$	$\lambda_{34} = 14.8$	$\lambda_{35} = 34.2$
$\lambda_{36} = 12.5$	$\lambda_{37} = 7.2$	$\lambda_{38} = 6.6$	$\lambda_{39} = 12.3$	$\lambda_{40} = 16.8$	$\lambda_{41} = 18.2$	$\lambda_{42} = 2.3$

in Fig. 6.

It can be seen from the efficiency values that the EPI based on the distance of the entropy of the dynamic evaluation is similar to the general DEA cross model, which shows this proposed method retains a good distinguish ability of the efficiency. Among them, there are eight departments' EPI values are above 0.5, whose department codes are 25, 33, 28, 32, 29, 36, 18 and 20, respectively. However, only five departments' EPI values based on the general DEA cross model above 0.3. Since the efficiency values are only relative values, what is more important is that the order of the EPI among departments. The ranked EPI of all industry departments in China based on the three models of the general environment DEA model (Model I), the general environment DEA cross model (Model II) and the environment DEA cross model (Model III) based on information entropy are shown in Fig. 7.

The results are shown in Fig. 7 that the gap between the three models is not significant based on the included department code: compared with Model I and Model III, the first six contain the same department codes, and there is only one different department in last six (transportation, warehousing and postal services). Meanwhile, compared with Model II and Model III, there is only one different department in first six (financial industry). In addition, the department codes of last six are the same. We continue to use the Wilcoxon rank sum test to determine whether there is a significant difference in the rankings of the three models, where the original hypothesis H_0 is not significantly different with the two models to be tested. Since samples size of three sequences used are all larger ($n = 42$), the default Wilcoxon statistic follows the standard normal distribution under the original hypothesis.

From the results of Table 5, we can see that if 10% significance level is selected, compared with the environmental DEA cross model, there is no significant difference in the ranking of environmental performance indices for each department based on the cross-evaluation of the improved environmental DEA cross model (the value of p is 0.3, and the original hypothesis cannot be rejected). Meanwhile, compared with the general DEA cross model, there is a significant difference in the EPI ranking of departments (the value of p is less than 0.1, rejecting the original hypothesis).

Therefore, although there is no significant different in the ranking order between Model I and Model III, which shows that the conclusion is more accurate, and the results of the general environment DEA model are more subjective (only self-rated scores). However, the result of the general DEA cross model is not accurate enough instead. The main reason is that it is too subjective to give each unit equal weight, which causes the distortion of the calculation result.

Based on Figs. 6 and 7, we can obtain the EPI of the electricity and heat production and supply department is highest. Compared with other high energy consumption departments, the total output of the electricity and heat production and supply department is very high as about 346,945 million Yuan, whose produces 15 times outputs than that of the Gas production and supply department and 17 times than that of the Metal products department. Meanwhile, compared with other low energy consumption departments, the total output of the electricity and heat production and supply department is 8.8 times than that of the Agriculture, forest, livestock and fishing services department, and 3 times than that of the Education department. Moreover, the input-output weight comparison of other departments based on the self-evaluation and the cross evaluation of the proposed method from the environmental

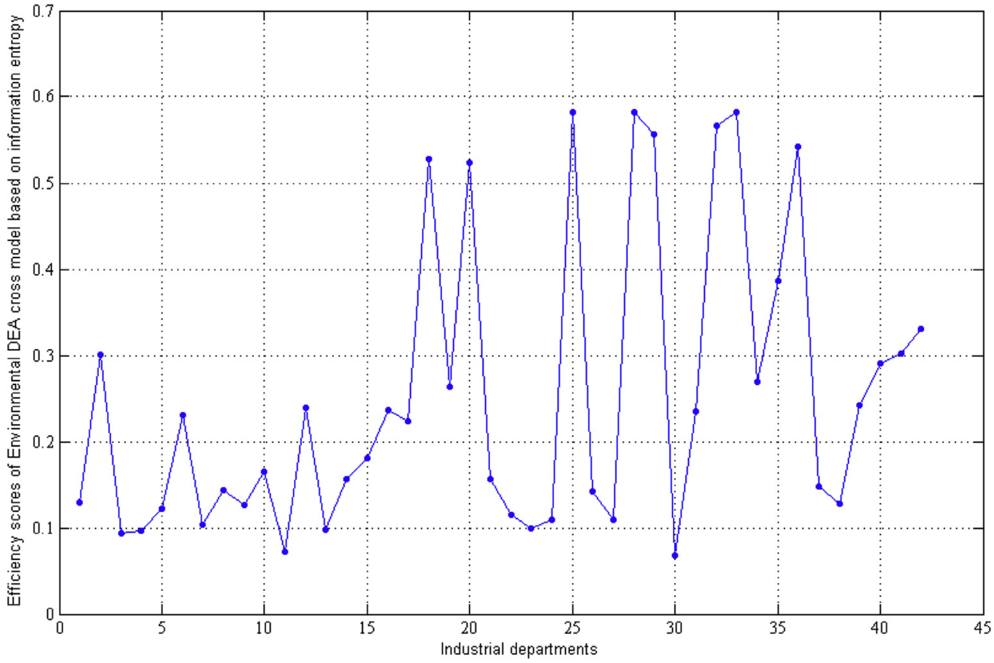

Fig. 6. EPIs based on the information entropy.

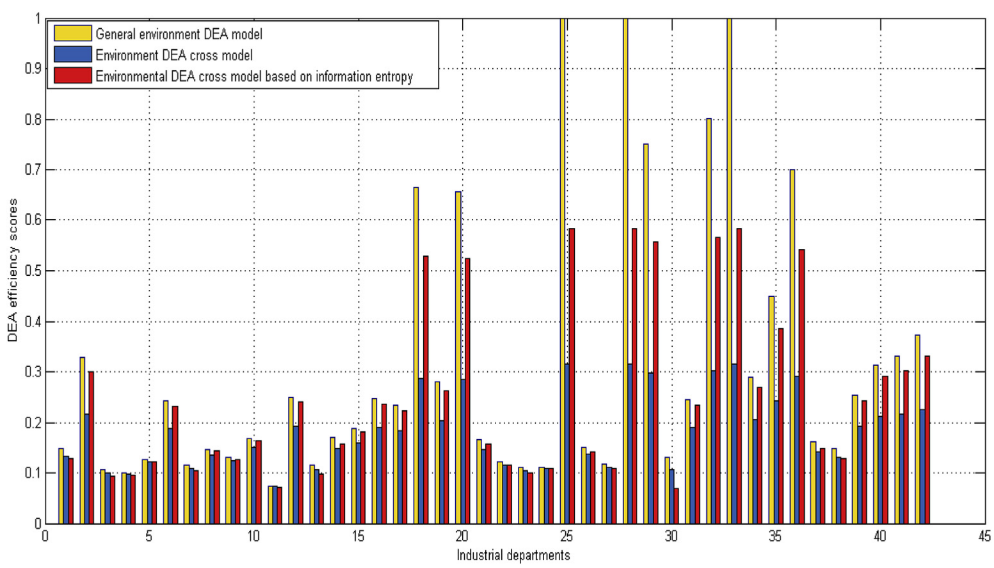

Fig. 7. Comparison of the three models.

Table 5
Wilcoxon test results.

Statistics	Comparison of Model I with Model III	Comparison of Model II and Model III
Z	1.006 (p = 0.32)	-1.95^{*} ($p = 0.093$)

Note: * Indicates that the statistic is significant at a level of $\alpha = 10\%$. And p is the probability value of rejecting the original hypothesis.

DEA model are similar, so the weight and the EPI of the electricity and heat production and supply department is relatively high. However, the $C0_2$ emission of the electricity and heat production and supply department is also high as 3,473,000 Tons, because the consumption of coal, natural gas and electricity is 1,213,000 Tons, 3400 million M^3 and 9200 million KWH, respectively. Meanwhile, the nature gas produces less CO_2 and the fossil fuels produce more CO_2 (2006 IPCC Guidelines for National Greenhouse Gas Inventories, 2006). Therefore, we should use more the natural gas and less the coal to improve energy efficiency and reduce the $C0_2$ emission. The similar analysis result of other departments can be obtained based the improved environment DEA cross model.

4. Discussion

Firstly, based on the traditional DEA cross model, this paper proposed a novel environment DEA cross model under the assumption of production technology CRS.

Secondly, the information entropy is introduced into the DEA cross model, which relaxes the assumption of the general cross model for each index. Each cross-evaluation index is equal to the self-evaluation index, and the weight is $1/n$. By using the entropy distance, the optimal problem based on the complete information is constructed. Moreover, the Lagrangian operator is used to calculate the analytic expression for the new weight W^*, and it is theoretically proved that the new weight W^* is the optimal solution of the whole region. Finally, we use W^* to define the comprehensive EPI based on the information entropy.

Finally, based on the data of "Beijing Input-Output Table 2012" and the Monte Carlo simulation, compared with the general environmental DEA model, the DEA cross model and the DEA cross model based on the information entropy integrated EPIs of industry departments, the simulation results show that the DEA cross model based on the information entropy has the lowest sensitivity to data noise impact. In addition, from the actual effect, there are also significant differences in EPIs of the environment DEA cross model based on the information entropy. Therefore, the DEA cross model based on the information entropy is effective and accurate to measure the efficiency of the environmental performance.

5. Conclusion

This paper proposes a novel environmental DEA cross model based on the information entropy. The information entropy is used to build the entropy distance based on the turbulence of the whole system, and calculate the weight in the CEM of the environmental DEA cross model in a dynamic way. The theoretical results show that the new weights constructed based on the information entropy is unique and globally optimal. Moreover, compared with the environmental DEA model and the traditional environmental DEA cross model, the improved environmental DEA cross model has a better efficiency discrimination by using the Monte Carlo simulation based on the data of industrial departments in China. Finally, the proposed model analyze and evaluate the carbon emission reduction of industrial departments in China, and it can obtain the improvement direction of carbon emission-reduction of the electricity and heat production and supply department by comparing the EPI and the total output of other high and low energy consumption departments. Moreover, the carbon emission reduction potential of other inefficient industrial departments can be obtained to improve the energy efficiency.

Meanwhile, the proposed model can also be obtained based on the assumptions of production technology VRS or NIRS, and the same result can also be analyzed by the non-radial DEA model. Moreover, continuing to blur the weight of the information entropy may be more effective in eliminate subjective factors, which are subject to the further research.

Acknowledgments

This research was partly funded by National Natural Science Foundation of China (61603025, 61673046), the Fundamental Research Funds for the Central Universities (ZY1703, JD1608) and the Natural Science Foundation of Beijing, China (4162045).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jenvman.2017.09.062.

References

Beijing Input and Output Table 2012. Beijing Municipal Bureau of Statistics, 2015. Beijing.
Chambers, R., Fare, R., 1996. Productivity growth in APEC countries. Pacific Econ. Rev. 1 (12), 181–190.
Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. Eur. J. Oper. Res. 2 (6), 429–444.
Chen, X.P., Xu, X.Y., 2012. Spatial and temporal changes of ecological efficiency in western China based on DEA cross model. J. Lanzhou Univ. 7, 48–53.
Despotis, D.K., 2002. Improving the discriminating power of DEA: focus on significant efficient units. J. Oper. Res. Soc. 53 (3), 314–323.
Fare, R., Grosskopf, S., 2004. Modeling undesirable factors in efficiency evaluation: comment. Eur. J. Oper. Res. (157), 242–245.
Guo, Q.E., Wang, X.Q., 2012. Fuzzy comprehensive evaluation model based on DEA cross evaluation and its application. Control Decis. (4), 575–583.
Guo, X.D., Zhu, L., Fan, Y., Xie, B.C., 2011. Evaluation of potential reductions in carbon emissions in Chinese provinces based on environmental DEA. Energy Pol. 39 (5), 2352–2360.
Hernandez-Sancho, F., Picazo-Tadeo, A.J., Reig-Martinez, E., 2000. Efficiency and environmental regulation. Environ. Resour. Econ. 15, 365–378.
Hernández-Sancho, F., Molinos-Senante, M., Sala-Garrido, R., 2011. Energy efficiency in Spanish wastewater treatment plants: a non-radial DEA approach. Sci.Total Environ. 409 (14), 2693.
Hu, Q.H., Xu, Q., 2014. Application of DEA cross efficiency evaluation method based on fuzzy expectation value. J. Comput. Appl. 23 (9), 224–229.
Huan, J., Ren, J., Xu, M., 2013. Study on performance evaluation of CO2 emission reduction in China based on three-stage DEA. J. Res. Sci. 35 (7), 1447–1454.
Intergovernmental Panel on Climate Change. 2006 IPCC Guidelines for National Greenhouse Gas Inventories, 2006. Intergovernmental Panel on Climate Change.
Jahanshahloo, G.R., 2011. Selecting symmetric weights as a secondary goal in DEA cross-efficiency evaluation. App. Math. Model. 35 (1), 544–549.
China Energy Statistics Yearbook 2013, National Bureau of Statistics of China, 2014. China Statistics Press, Beijing.
Pereira, E.J.D.S., Pinho, J.T., Galhardo, M.A.B., 2014. Methodology of risk analysis by Monte Carlo Method applied to power generation with renewable energy. Renew. Energy 69 (3), 347–355.
Ramanathan, R., 2003. An Introduction to Data Envelopment Analysis: a Tool for Performance Measurement. Sage Publications, New Delhi.
Ramon, N., Ruiz, J.L., 2010. On the choice of weights profiles in cross-efficiency evaluations. Eur. J. Oper. Res. (207), 1564–1572.
Sexton, T.R., Silkman, R.H., 1986. Data Envelopment Analysis: Critique and Extensions, vol. 32. Jossey-Bass, San Francisco, pp. 73–105.
Shannon, C.E., 1948. The mathematical theory of communication. Bell Sys. Tech. J. (27), 379–423.
Silva, A.N.C., Amaral, R.S., Santos, J.A., Vieira, J.W., Menezes, R.S.C., 2015. Statistical analysis of discrepant radioecologicalradio ecological data using Monte Carlo bootstrap method. J. Radioanal. Nucl. Chem. 306, 571–577.
Tyteca, D., 1996. On the measurement of the environmental performance of firms. J. Environ. Manag. (46), 281–308.
Wald, A., 1945. Generalization of a theorem by v. Neumann concerning zero sum two person games. Ann. Math. 46 (2), 281–286.
Wang, Y.M., Chin, K.S., 2010. A neutral DEA models for cross-efficiency evaluation and its extension. Expert Syst. Appl. 37 (5), 3666–3675.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biomet. Bull. 1 (6), 80–83.
Wu, J., Liang, L., Yang, F., 2009. Determination of the weights for the ultimate cross efficiency using Shapley value in cooperative game. Expert Syst. Appl. 36 (1), 872–876.
Xu, J.W., Xu, X.Y., 2013. Evaluation of urban efficiency in Gansu province based on DEA crossover model. J. Nat. Resour. 4, 618–624.
Zhou, P., Ang, B.W., Poh, K.L., 2008a. A survey of data envelopment analysis in energy and environmental studies. Eur. J. Oper. Res. 189 (1), 1–18.
Zhou, P., Ang, B.W., Poh, K.L., 2008b. Measuring environmental performance under different environmental DEA technologies. Energy Econ. 30, 1–14.
Zhu, Z., Wang, K., Zhang, B., 2014. Applying a network data envelopment analysis model to quantify the eco-efficiency of products: a case study of pesticides. J. Clean. Prod. 69, 67–73.