The 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)
May 18-20, 2018, Nanjing, China

Energy Efficiency Analysis of PTA Plants Based on PCA-DEACM

Zhiqiang Geng
Guizhou Provincial Key Laboratory of Public Big Data
Guizhou University
Guiyang, Guizhou 550025, China
College of Information Science and Technology
Beijing University of Chemical Technology
Beijing 100029, China
E-mail: gengzhiqiang@mail.buct.edu.cn

*Yongming Han**
Guizhou Provincial Key Laboratory of Public Big Data
Guizhou University
Guiyang, Guizhou 550025, China
College of Information Science and Technology
Beijing University of Chemical Technology
Beijing 100029, China
Corresponding author: hanym@mail.buct.edu.cn

Ning Chen
Guizhou Provincial Key Laboratory of Public Big Data
Guizhou University
Guiyang, Guizhou 550025, China
College of Information Science and Technology
Beijing University of Chemical Technology
Beijing 100029, China
E-mail: 13718030854@163.com

Yanhua Zhong
Jiangmen Polytechnic
Jiangmen, Guangdong 529020, China
E-mail: zhflowers@163.com

***Abstract*—The method of data envelopment analysis (DEA) is a nonparametric statistical method, which is widely used in energy efficiency analysis of complex industrial process. However, there are many factors for affecting the energy efficiency, so the input and output indexes of the DEA is too much to get the low resolution of analysis results. Therefore, this paper proposes an energy efficiency analysis method based on the DEA cross model (DEACM) integrating principal component analysis (PCA) (PCA-DEACM). After using the PCA to reduce the dimensions of input and output indexes, the DEACM is used to further improve the effective resolution of decision-making units (DMUs). And the analysis results are more objective and reasonable. Finally, the proposed PCA-DEACM method is applied in energy efficiency analysis of PTA plants. By comparing with the classical DEA method and the PCA-DEA method, the validity and applicability of the PCA-DEACM is verified, which provides accurate analysis results for optimizing the operation of PTA plants.**

Keywords—data envelopment analysis; energy efficiency analysis; principal component analysis; DEA cross model; PTA plants

I. INTRODUCTION

Terephthalic acid is one of the most important organic raw materials, and is widely used in various aspects of the national economy, such as chemical fiber, light industry, electronics, and construction. In recent years, with the continuous development of the polyester industry in China, the apparent consumption of the PTA has been steadily increasing. Energy saving and emission reduction is an effective way to reduce production costs and improve market competitiveness for PTA production enterprises. However, there is still a gap between the domestic PTA production technology and the foreign advanced technology in terms of energy consumption. Therefore, it is of certain economic value to analyze and evaluate the energy efficiency of PTA production plants and put forward some suggestions for energy efficiency improvement.

Researchers have put forward many methods in efficiency evaluation, including analytic hierarchy process(AHP), fuzzy comprehensive evaluation, artificial neural network (ANN), principal component analysis (PCA), data envelopment analysis (DEA) and so on [1-4]. The AHP involves too many subjective factors in the practical application. Fuzzy comprehensive evaluation method mainly uses the fuzzy matrix to process qualitative indicators. And the dependence of the neural network on the data is large, so it is often difficult to converge. The PCA is often used to extract data features and simplify data sets. The DEA is generally considered a more suitable analytical method for energy efficiency analysis than traditional data analysis methods [5]. DEA methods are applied to most energy efficiency assessment problems. In 1978, Charnes, Cooper and Rhodes first proposed the DEA method and established the CCR model [6]. Because the DEA method has no requirement for dimensional input and output variables, it can reflect the information and characteristics of the ideal evaluation object itself. Moreover, the DEA method has a good effect in dealing with the relative scale effectiveness and relative technical effectiveness of decision-making units (DMUs). Therefore, the applicability in various fields is relatively strong.

Sexton et al. proposed a cross efficiency evaluation method based on the classic DEA model [7]. Geng et al. expounded an improved DEA cross model (DEACM) and its application in complex chemical process [8]. Zhang et al. described the environment DEACM based on the information entropy and its application analysis [9]. Fan et al. introduced the group cross efficiency evaluation method based on the heterogeneity of the DMU [10]. Liu et al. introduced a DEA weighted cross efficiency evaluation method based on the reliability and transfer matrix [11]. Ruiz et al. described the fuzzy cross-efficiency evaluation with a possibility approach [12]. The evaluation and analysis method based on the DEA integrating the PCA (PCA-DEA) is widely used in ecological environment, electricity, economy and other fields [13-15]. Zhu et al. introduced an energy efficiency analysis method based on the PCA-DEA for ethylene plants [16], the PCA was used to reduce input and output dimensions, and to alleviate or eliminate the problem of too many decision units. However, when using the traditional DEA model to analyze the energy efficiency, only effective and ineffective units can be analyzed, and the ineffective units cannot be sorted. For two or more effective units, it is impossible to compare.

According to the characteristics of multi input and output indicators of PTA plants in the energy efficiency, using the PCA method for reducing dimensions of input and output indicators, avoid the excessive influence effect analysis. And then the DEACM is used to sort of all DMUs fully for avoiding input-output weights in the extreme optimization process. Therefore, this paper introduces a DEACM based on the PCA to analyze the energy efficiency index of PTA plants.

II. The Introduction of Method

A. The PCA

In 1901, Pearson proposed the concept of principal component [17]. In 1933, Hotelling improved it on this basis and extended the concept of principal component to the application of random variables [18]. The PCA is a new way to reduce dimensions by forming new variables through linear combination of original variables. In order to reduce lost information, a method of analyzing larger and larger independent components is used to analyze things. The PCA can transform a number of interrelated variables into a few independent variables, and is not objectivity influenced by the criterion weight and subjective preferences. Meanwhile, it reserves the correlation structure between process variables, which is suitable for dimensionality reduction of the high-dimensional data. The following is a brief introduction to the principles of the PCA.

Assuming that there are n samples, each sample has m index variables, so it can be constructed as a n*m order data of the matrix X.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} = \left(X_1, X_2, \cdots, X_m\right) \tag{1}$$

- Standardized raw data matrix, the original data x_{ij} is normalized to $y_{ij} = (x_{ij} - \bar{x}_j)/S_j$, in which $\bar{x}_j$ and S_j are the mean and standard deviation of x_j, respectively.
- Solving the eigenvalue λ_j and the eigenvector l_j of the correlation coefficient matrix $R = Y^T Y/(n-1)$ of Y, in which $j = 1,2,\cdots,m$.
- From small to large, if the cumulative variance contribution rate of the former t eigenvectors is:

$$\eta_t = \sum_{j=1}^{t} V_{\lambda_j} \Big/ \sum_{j=1}^{m} V_{\lambda_j} \geq 85\% \tag{2}$$

in which V_{λ_j} is the variance of l_j, the principal component of the original data is $f_i = Xl_i$, in which $i = 1,2,\cdots,t$.

B. The DEA

The DEA is a nonparametric statistical method. The concept of the relative efficiency of the DMU is introduced, and the relative technical effectiveness and relative scale effectiveness of the DMU are calculated. the DEA is generally considered to be a more suitable analytical method for energy efficiency analysis than traditional data analysis methods. Most DEA methods are applied to analyze energy efficiency assessment problems. The following is an introduction to the CCR model of the classic DEA model.

Suppose there are n DMUs, each DMU has m entry and s output, then the input vector of the j^{th} DMU is recorded as $x_j = (x_{1j}, x_{2j}, \cdots, x_{mj})^T$, and the output vector is recorded as $y_j = (y_{1j}, y_{2j}, \cdots, y_{sj})^T$, $j = 1,2,\cdots,n$ .The input and output vector of DMU_{j_0} is (x_{j_0}, y_{j_0}) , the following abbreviated as (x_0, y_0) . So the CCR model is:

$$\begin{cases} \max u^T y_0 / v^T x_0 \\ \text{s.t. } u^T y_j / v^T x_j \leq 1 \\ u \geq 0, v \geq 0 \end{cases} \tag{3}$$

in which $v = (v_1, v_2, \cdots, v_m)^T$ and $u = (u_1, u_2, \cdots, u_s)^T$ represent the weight coefficients of the input and output variables, respectively.

The dual model of equivalent linear programming model is:

$$DUM_{CCR} = \begin{cases} \min \theta \\ \text{s.t.} \sum_{j=1}^{n} x_j \lambda_j + s^- = \theta x_0 \\ \sum_{j=1}^{n} y_j \lambda_j - s^+ = y_0 \\ \lambda_j \geq 0,\ j = 1,2,\cdots,n \end{cases} \quad (4)$$

in which θ represents the technical efficiency of DMU_{j_0}, and s^-, s^+ represent the relaxation variables.

If $\theta^0 = 1$, $s^- = 0$, $s^+ = 0$, DMU_{j_0} is effective. The greater the θ^0, the higher the relative efficiency of its decision making unit. The ineffective DMUs can be adjusted to the valid input and output according to the adjustment formula. The adjustment formula is:

$$\begin{cases} x_0' = x_0 \theta^0 - s^{0-} \\ y_0' = y_0 + s^{0+} \end{cases} \quad (5)$$

C. The DEACM

The cross efficiency evaluation method usually includes the following three stages:

1) DMUs self-evaluation: calculating the classic DEA model to determine the optimal weight.

2) Cross evaluation of DMUs: the cross efficiency evaluation value of the DMU_d evaluation DMU_j is θ_{dj}, and the cross efficiency matrix E is composed of all cross efficiency evaluation values.

$$\theta_{dj} = \sum_{r=1}^{s} u_{rj}^* y_{rd} \Big/ \sum_{i=1}^{m} v_{ij}^* x_{id}\ , d, j = 1,2,\cdots,n \quad (6)$$

$$E = \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1n} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \theta_{n1} & \theta_{n2} & \cdots & \theta_{nn} \end{bmatrix} \quad (7)$$

3) Cross efficiency value: calculating the final cross efficiency value, and the cross efficiency value of DMU_j is:

$$e = \overline{E}_j = \frac{1}{n} \sum_{d=1}^{n} E_{dj},\ j = 1,2,\cdots,n \quad (8)$$

III. ENERGY EFFICIENCY ANALYSIS OF THE PCA-DEACM

Choosing reasonable input and output indexes, and using the PCA to reduce the dimension of input and output data matrix. The input and output principal components will be used as input and output indexes of the DEACM, respectively. The energy efficiency analysis process of the PCA-DEACM is shown in Fig1.

The PCA-DEACM combines the PCA and the DEACM to analyze the energy efficiency of PTA plants. By using the advantages of nonparametric and objectivity in the analysis of the CCR model, and adding the concept of mutual evaluation of DMU, the distinction and accuracy of the model analysis results are greatly improved. At the same time, the PCA-DEACM can obtain fewer new indexes, without losing the data characteristic of the original indexes, and solve the problem that the resolution of the analysis result is too low because of too much input and output indexes.

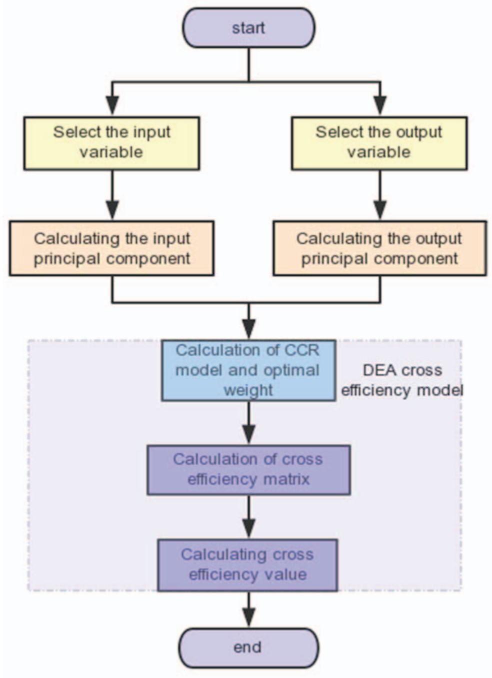

Fig. 1. The flowchart of the PCA-DEACM methodology.

IV. ENERGY EFFICIENCY ANALYSIS OF PTA PLANTS

In recent years, people have been unable to obtain economic benefits simply from PTA production process improvement and production equipment transformation, but to optimize operation of PTA production device, play the inherent potential of production device, and obtain at low energy consumption and low cost. In the process of PTA production, the solvent consumption of the acetic acid is an important production index that needs to be paid more attention.

A. Selection of variables

The process characteristics and mechanism of the PTA production process are more complex. The acetic acid in materials is highly corrosive. It causes great difficulty in measurement. Therefore, the change of the acetic acid content in the top of the tower is replaced by measuring the conductivity of acetic acid at the top of the tower. In the PTA solvent system, the main influencing factors of the acetic acid consumption are the following seventeen parts [19]: the feed volume FC1501, the temperature TI1504, the return flow FC1502~FC1504, the temperature TI15010, the reboiler steam flow rate FC1507, the tower temperature TI1511~TI1519, TC1501.

The number of decision making units in DEA model isn't less than the product of input and output variables. If the sample size is not large enough, and there is strong correlation between variables, the resolution of DEA model results will be relatively low. Therefore, this paper selects 20 sets of data from a PTA production plant as the analysis data, affecting the seventeen factors of acetic acid consumption as the input index of the production device, and the conductivity of acetic acid as the output index.

B. *Example analysis*

Seventeen factors affecting the consumption of acetic acid are used directly as the 17 input indexes of the DEA, which is shown in Table 1, and the results are shown in the DEA curve in Fig 2. It can be seen that the efficiency values of the 1st, 4th, 6th, 12th, 16th and 20th DMUs are all 1, and the efficiency values of other DMUs are also more than 0.96. Obviously, because of the excessive input indexes of the DEA method, all DMUs are highly effective, and the efficiency gap between each DMU is very little. The PCA-DEA method is adopted in the PCA-DEA curve in Fig 2. First, the 17 input indexes are processed by the PCA method, and the cumulative variance contribution rate of the first three principal components (principal component 1, principal component 2 and principal component 3) is greater than 85%. Then the three principal components are used as the new input indexes of the DEA model as shown in Table 1. The results are shown in PCA-DEA curve. It is not difficult to see that the efficiency value area diversity of the ineffective DMUs is obviously increased, which is a good phenomenon, but the efficiency values of the 1st, 4th, 5th, 6th, 10th DMUs are all 1, which cannot identify their advantages and disadvantages, and still need to be improved. The 3 new indexes after the PCA dimension reduction are used as the inputs of the DEACM, which is shown in the PCA-DEACM curve in Fig 2. The results show that the analysis result of the proposed PCA-DEACM method has the same trend as those of the DEA and the PCA-DEA, and the efficiency value of the 4th DMU is the highest in the 20 DMUs, and the efficiency values of the other DMUs are different. The energy efficiency of each DMU can be clearly determined. The PCA-DEACM maximizes information of original indexes to ensure the comprehensive energy efficiency analysis, improves the resolution of the analysis result, and solves the problem that so many DMUs are effective at the same time.

TABLE I. INPUT INDICATSORS FOR A PTA PLANT

DEA Input Indexes			PCA-DEA and PCA-DEACM Input Indexes
FC1501	FC1502, FC1503, FC1504	FC1507	Principal Component 1
TI1504	TI1511, TI1512, TI1513, TI1514 TI1515, TI1516, TI1517, TI1518, TI1519	TI15010	Principal Component 2
TC1501			Principal Component 3

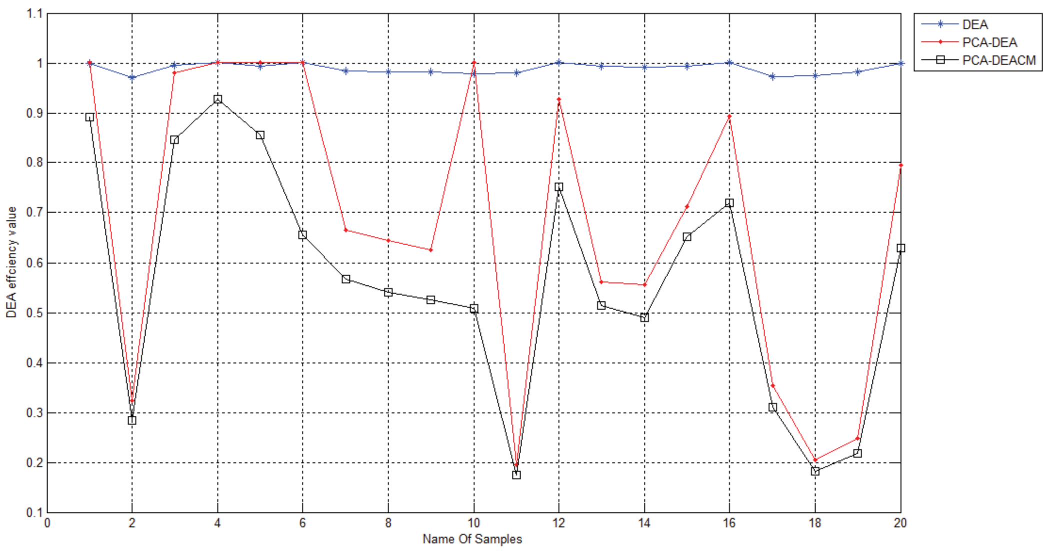

Fig. 2. Comparison analysis results of a specific PTA plant.

Though the comparison of efficiency value curves of three methods in Fig 2, it is easily known that the DEA method can only distinguish whether the DMUs are valid or not, but the efficiency values of the ineffective units are too high, which is inconsistent with the actual situation. And it cannot provide effective suggestions to improve the energy efficiency of PTA plants. The PCA-DEA method can increase the evaluation result resolution of DMUs. But there is still the phenomenon of multiple units effectively at the same time. So this part of the DMUs can't be distinguished. Based on the analysis of the

PCA-DEACM presented in this paper, its trend of efficiency value curve is in common with those of DEA method and PCA-DEA method, and the correctness of the PCA-DEACM is illustrated. The PCA-DEACM not only keeps the original input and output data characteristics, but also solves the problem of multi indexes at the same time. It clears and accurately estimates the energy efficiency of each DMU, and plays a more precise role in guiding the performance improvement of the PTA plants.

V. CONCLUSION

Aiming at the data of the actual production process of PTA plants, this paper presents an energy efficiency method based the PCA-DEACM to reduce the dimensions of actual production process data. And then the DEACM is applied to analyze and evaluate the new input and output data. The PCA-DEACM overcomes the problem of too many DMUs due to the excessive input and output indexes of the classical DEA model. Finally, the proposed PCA-DEACM is applied in the energy efficiency analysis of PTA plants. By comparing with the classical DEA method and the PCA-DEA method, the validity and applicability of the PCA-DEACM is verified. Furthermore, the accurate analysis results for optimizing the operation of PTA plants can be obtained.

REFERENCES

[1] Q. Wu, L. Cheng, H. Huang, S. Gao and Y.F. Guo. Comprehensive energy efficiency evaluation method of energy Internet based on AHP[J]. Electrical Application, 2017(17): 62-68.

[2] Q.K. Zhu, Q.C. Liang, H.Y. Zhan. Analysis of aero-engine performance and selection based on fuzzy comprehensive evaluation[J]. Procedia engineering, 2017, 174: 1202-1207.

[3] Y. Kim, H. Jang, J. Kim and J. Lee. Prediction of storage efficiency on CO2 sequestration in deep saline aquifers using artificial neural network[J]. Applied energy, 2017, 185: 916-928.

[4] A. Huang, L.A. Yang, T. Du, A.L. Wang, B. Zhang and G. Yuan. Comprehensive evaluation of soil nutrients based on principal component analysis[J]. Arid Zone Research, 2014, 31(5): 819-825.

[5] A. Mardani, E.K. Zavadskas, D. Streimikiene, A. Jusoha and M. Khoshnoudi. A comprehensive review of data envelopment analysis (DEA) approach in energy efficiency[J]. Renewable and Sustainable Energy Reviews, 2017, 70: 1298-1322.

[6] A. Charnes, W.W. Cooper, E. Rhodes. Measuring the efficiency of decision making units[J]. European Journal of Operational Research, 1978, 2(6): 429-444.

[7] T.R. Sexton, R.H. Silkman, A.J. Hogan. Data envelopment analysis: Critique and extensions[J]. New Directions for Evaluation, 1986, 1986(32): 73-105.

[8] Z.Q. Geng, J.G. Dong, Y.M. Han and Q.X. Zhu. Energy and environment efficiency analysis based on an improved environment DEA cross-model: Case study of complex chemical processes[J]. Applied Energy, 2017, 205: 465-476.

[9] K.Y. Zhang, J. Yang. DEA model of cross environment based on information entropy and its application[J]. Statistics & Information Forum, 2017, 32(7): 23-29.

[10] J.P. Fan, Y.Y. Zhao, M.Q. Wu. Group of cross-efficiency evaluation method based on non-homogeneous of decision making units[J]. Computer Engineering and Applications, 2017, 53(16): 241-248.

[11] W.L. Liu, Y.M. Wang, S.L. Lv. DEA weighted cross-efficiency evaluation based on the reliability level and the transition matrix[J]. Journal of Systems Engineering, 2017, 32(4): 570-576.

[12] J.L. Ruiz, I. Sirvent. Fuzzy cross-efficiency evaluation: a possibility approach[J]. Fuzzy Optimization and Decision Making, 2017, 16(1): 111-126.

[13] M. Wang, X.P. Zhang. Regional eco-efficiency evaluation of China based on PCA-DEA model[J]. Journal of University of Chinese Academy of Sciences, 2015, 32(4): 520-527.

[14] D.D. Xu. Study of power efficiency based on PCA-DEA[D]. He Fei: University of Science and Technology of China, 2015.

[15] J.P. Liu, W. Zhou, J.M. He and S. Meng. Bank efficiency and strength appraisal based on multi-dimensional statistical combination model[J]. Journal of Beijing Institute of Technology (SOCIAL SCIENCE), 2011, 13(3): 21-26.

[16] Q.X. Zhu, X. Chen, Y.L. He, X.Y. Lin and B.X. Gu. Energy efficiency analysis of ethylene plant based on PCA-DEA[J]. Journal of chemical industry, 2015, 66(1): 278-283.

[17] K. pearson. On lines and planes of closest fit to systems of points in space[J]. Philosophical Magazine, 1901, 2(11): 559-572.

[18] H. Hotelling. Analysis of a complex of statistical variables into principal components[J]. Journal of Educational Psychology, 1933, 24(6): 417-520.

[19] Z.Q. Geng, X. Yang, Y.M. Han, Q.X. Zhu. Energy optimization and analysis modeling based on extreme learning machine integrated index decomposition analysis: Application to complex chemical processes[J]. Energy, 2017, 120: 67-78.