ISSN 1 746-7233, England, UK
World Journal of Modelling and Simulation
Vol. 16 (2020) No. 4, pp. 273-287

Water-environment efficiency evaluation using data envelopment analysis with principal components analysis: a case study of Tuojiang River basin

Rui Chen[1] , Zhongwen Xu[2*]

Business School, Sichuan University, Chengdu, 610064, China

(*Received March 14 2020, Accepted October 10 2020*)

Abstract. Rapid economic development in China has aggravated water pollution problems, so improvements in water-environment efficiency, as a principal method of achieving water protection, are vitally important. Water-environment efficiency (WEE) is the ability to obtain the largest economic benefits with minimal water pollution. Therefore, quantitative water-environment efficiency analyses are not only essential for understanding the regional water resource protection situation, but also a prerequisite for designing and adjusting relevant policies. This paper develops a data envelopment analysis (DEA) model combined with principal components analysis (PCA) to evaluate a regional WEE and suggest developmental policies. The method consists of a two-stage analysis that begins with the PCA. In the second stage, the DEA model is used to rank the WEE in different regions. The WEE ranking for a sample area in the Tuojiang River basin is presented. Based on the research results, policy recommendations are then given on reducing water pollution and increasing China's water-environment efficiency.

Keywords: Water-environment efficiency, Principal components analysis, Data envelopment analysis, DEA-PCA model

1 Introduction

Water-environment issues have increasingly become an obstacle to economic and social development. Water pollution and other water-environment problems are growing, presenting a severe test to sustainable survival and development. In China, rapid economic growth has led to many water environmental problems, which not only affect the physical and mental health of the general population but also highlight the problems in the current economic growth mode. Hence, it is crucial to enhance the water environment and develop a series of effective water-environment efficiency evaluation methods. This paper is based on this background, in which an effective water environmental efficiency evaluation method is proposed, which could provide a valuable reference for decision-makers.

Various methods have been developed to evaluate efficiency, one of which is the DEA model, which has gained popularity as a methodology in evaluating bank performance [7, 13, 15, 17, 20], assessing university research efficiency [8], identifying excesses or deficits in production and examining buyer-supplier supply chains [25, 31, 32, 41]. Farrell [18] first proposed a non-parametric method of computing the relative efficiency of a decision making unit (DMU) on the basis of a set of DMUs. Two decades later, Charnes et al. [12] proposed a line programming model to evaluate technical efficiency and technological progress. After this time, the DEA has been widely used to measure energy and environmental efficiency at the macro-economic level.

In recent years, significant research has paid attention to environmental protection in China. Hu and Wang [23] proposed a measure for total factor productivity, which was used to measure 29 administrative regions in China using a CCR-DEA model from 1995 to 2002. Hu et al. [22] used the CCR model to analyze the technical

* Corresponding author. *E-mail address*: xuzhongwen1996@163.com

efficiency and productivity changes in 31 regions in China from 1997 to 2001. Chiu and Wu [14] used an undesirable measure DEA model to calculate the impact of undesirable output and input on energy efficiency. The research established two models to analyze 27 provinces and cities in China from 2000 to 2003. Yang and Wang [38] utilized a nonparametric DEA method to calculate the production frontier of environmental efficiency for provincial energy utilization. The results indicated that the energy utilization environmental efficiency in Chinese provinces still had significant room for improvement, and CO_2 emissions still need to be reduced. Zhou et al. [40] first proposed a new non-radial DEA approach by integrating the entropy weight and the SBM model to evaluate the environmental efficiency of the Chinese power industry at the provincial level from 2005-2010. Shi et al. [33] measured the total technological efficiency, pure technological efficiency, and scale efficiency of 28 administrative regions in China using an extensive DEA model.

Existing research, however, has tended to focus on environmental efficiency measurements. However water-environment efficiency (WEE) has not been mentioned science and technology academic research. Through a study of the concepts related to WEE [19, 21, 29], we define that WEE as the ability to obtain the best economic benefits at a minimal water pollution cost.

In this paper, an index system for the WEE evaluation, which has both water quality monitoring indicators and regional economic development evaluation indicators, is developed. Then, we build a DEA model based on the PCA to evaluate the WEE of the Tuojiang River basin. Then we rank the WEE values. The results and conclusions drawn from the DEA-PCA model are then discussed and summarized.

2 Methodology

In this section, water quality monitoring indicators serve as inputs, and regional economic development indicators serve as outputs. That is, the index system for the regional WEE evaluation is scientifically built based on sample data comparability and availability. Then an integrated DEA-PCA approach is introduced. Fig. 1 presents the detailed flow of the methodology. In summary, the detailed flow is as follows:

I. The concept of WEE is defined based on related research.

II. Appropriate water quality monitoring indicators and economic development indicators are selected as the DEA input and output indicators.

III. The PCA is applied to replace the original inputs with a smaller group of principal components (PCs).

IV. With the original input indicators and integrated output indicators, the WEE scores can be calculated using the DEA model.

V. The ranking results from the DEA are presented and analyzed. Different DMUs are marked with different colors.

VI. Suggestions for improving the regional water environment are proposed.

2.1 Indicators

From the WEE definition, indicators measuring water quality are chosen as input variables and indicators measuring the economic development level are chosen as the output variables. In this study, data about wastewater emissions, and wastewater into the river, Chemical Oxygen Demand (COD) and Ammonia Nitrogen ($NH_3 - N$) quantities in the river are the four water quality inputs. The economic development level refers to the size, speed, and attained level of economic development. There has not been any uniform standard for measuring the economic development level, so from previous research [4, 5, 10, 24, 36], we have divided it into two basic categories. One approach is to choose one of the most representative indicators, while the other is to establish an index system, with the latter method being increasingly utilized. Based on this and combined with the principles of data availability, an index system was established and is presented in Table 1. In the index system, the indicators selected were based on regional sustainable development research, then, using reliability and validity analysis, any unnecessary indicators were excluded. The final index system for the input and output indicators is listed in Table 2.

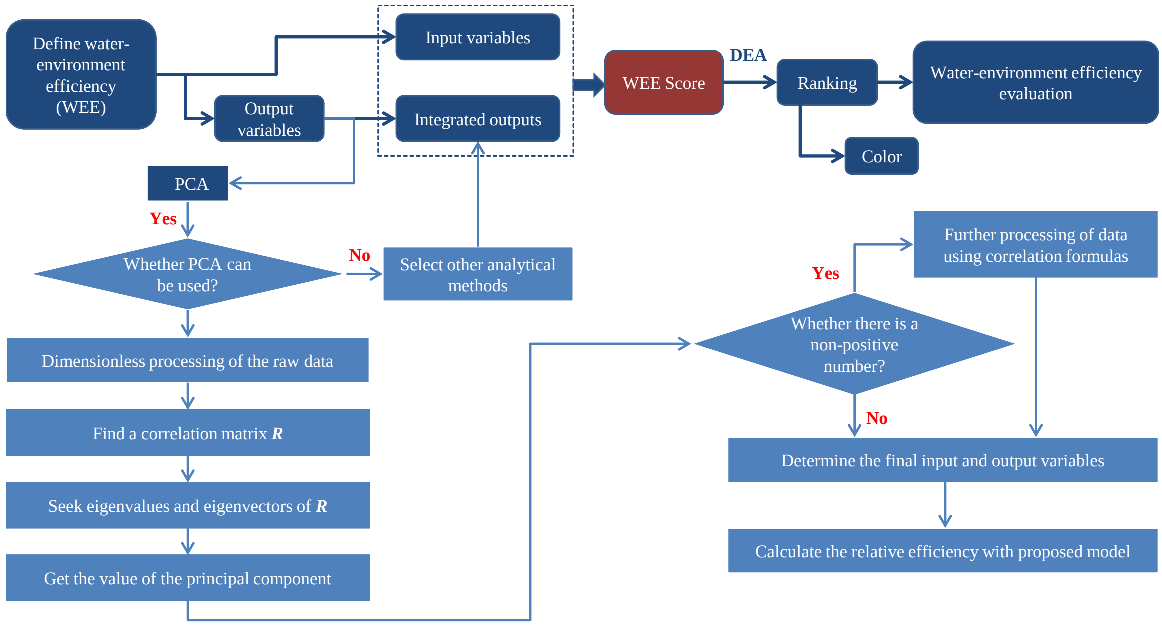

Fig. 1: The framework of methodology

Table 1: Index system for measuring the level of economic development in the region

First-class Targets	Second-class Targets
The economic scale	Administrative region land area
	Area of built-up area
	Household population at the end of the year
	Regional Gross Domestic Product (Regional GDP)
	Local general budget revenue
	Tax revenue
	Total investment in fixed assets
	Total retail sales of social consumer goods
	Primary industry added value
	Secondary industry added value
	Industrial added value
	Tertiary industry added value
	Private economy added value
The quality of economic development	Added value of the primary industry to the Regional GDP ratio
	Added value of the secondary industry to the Regional GDP ratio
	Added value of the tertiary industry to the Regional GDP ratio
	Regional GDP growth rate
	Total output of agriculture, forestry, animal husbandry and fishery
	Industrialization rate
	Total above-scale industrial profits
	Urban and rural residents' savings deposits
	Town residents' per capita disposable income
	Farmers' per capita net income

2.2 DEA-PCA model

The DEA, as a new assessment method, is based on relative efficiency. It enhances the assessed object, and provides an input-output economic significance, thereby assisting decision-makers when making decisions and proposing policy changes. The DEA's projection theorem provides advice on improving the assessed object. In recent years, the DEA has become a central technique in economics and management science to analyze productivity and efficiency, as it measures the technical efficiency relative to a deterministic best practice frontier which has been empirically built from observed inputs and outputs using linear programming techniques. Its main advantage is that it allows several inputs and outputs to be considered at the same time.

Table 2: Input and output variables

	Variables	Symbol	Unit
Input variables	Wastewater emissions	x_1	m^3
	Wastewater into the river	x_2	m^3
	COD in the river	x_3	t/a
	$NH_3 - N$ in the river	x_4	t/a
Output variables	Regional GDP	y_1	Yuan
	Local general budget revenue	y_2	Yuan
	Tax revenue	y_3	Yuan
	Total investment in fixed assets	y_4	Yuan
	Tertiary industry added value	y_5	Yuan
	Urban and rural residents' savings deposits	y_6	Yuan
	Town residents' per capita disposable income	y_7	Yuan

While there are many advantages to the DEA method, there are also some drawbacks. The rule of thumb in DEA theory requires that the number of DMUs be at least twice the numbers of evaluation units. If the number of indicators is excessive, which is contrary to the rule of thumb, it leads to unnecessary effective DMUs and the discrimination ability of the DEA model is greatly reduced [26]. The key limitation of the DEA is that it assumes data to be free of measurement error and, therefore, it is more sensitive to the presence of measurement errors than other parametric techniques. When using real-life data for the DEA analysis, the results are inevitably affected to a greater or lesser degree by statistical noise (measurement errors), so reducing the impact of these measurement errors (statistical noise) on the DEA analysis results increase the confidence of the analysis. Therefore, it is necessary to reduce the dimensionality or the number of variables in the DEA structure.

To avoid these shortcomings, the PCA is useful. The use of the PCA makes it possible to solve two problems simultaneously because; a) The principal components are sensitive to from measurement errors (statistical noise) from reallife data; and b) using the PCA reduces the dimensionality or the number of variables in the DEA structure. To solve these problems, Adler and Golany [1, 2] and Adler et al. [3] suggested using the PCA to produce uncorrelated linear combinations of the original variables or sub-indicators (inputs and outputs). Essentially, the PCA is designed to reduce the number of variables to a smaller number of indicators, called principal components, which are linear combinations of the original variables. It is commonly accepted that if most of the variance in the data can be attributed to the first few PCs, the original variables can be replaced by those components or indicators with a minimal loss of information and the measurement error (statistical noise) impact of measurement errors (statistical noise is modest. Hence, the principal component scores can be used instead of the original input and output variables, and they can be used to replace all of the inputs and/or outputs simultaneously or to replace certain groups of variables.

The PCA is a vital statistical method for studying how to transform the multi-index into a less comprehensive index. The PCA can transform high-dimensional space issues into a low-dimensional space to deal with them, which makes the problem much simpler. The objective of the PCA is to identify a new set of variables, with each new variable (PC) being a linear combination of the original variables. The first new variable, y_1, accounts for the maximum variance in the sample data. In the CCR-DEA model, the PCA replaces the original m inputs with a smaller group of PCs explaining the variance structure of the data matrix through linear variable combinations. The new variables (PCs) are uncorrelated, so the PCA is performed by identifying the Eigen structure of the covariance or the singular value decomposition of the original data. This eventually leads to the DMUs scoring and rankings. Clearly, if we use less than full information, we lose some of the explanatory power of the data, but improve the discriminatory power of the model. The PCA used here is based on covariance rather than correlation, as the variables used in the DEA are often quantified in different units of measurement. In general, the inputs and outputs of a DEA need to be strictly positive, but PCs are able to contain negative values. Consequently, all the PC input data used in the DEA are processed to ensure they are positive.

During the last decade, efficiency measurement approaches have evolved very quickly. Among these, the advantages of non-parametric techniques have been widely recognized [27, 28]. We employ in this section a leading non-parametric efficiency measurement technique to evaluate a regional WEE. This DEA method was invented by Charnes et al. [12]. They conducted a mathematical programming model to measure the efficiency frontier using Pareto optimality. This approach was created to determine the organization's relative efficiency using multiple input and output measures [6]. The DEA is based on mathematical programming, in that the DEA evaluates the efficiencies of the DMUs through the building of linear programming model. In summary, in this paper, we selects a constant return to scale, output-oriented DEA model to measure the WEE values in the surrounding areas of the Tuojiang River Basin. The CCR-DEA model is then combined with PCA to develop an advanced DEA-PCA model.

The evaluation model is shown in Fig. 2. X expresses the input values and Y expresses the output values. As mentioned in the previous section, waste water emissions (x_1), and wastewater into the river(x_2), COD in the river (x_3), and $NH_3 - N$ in the river (x_4) are the input variables, regional GDP (y_1), local general budget revenue (y_2), tax revenue (y_3), total investment in fixed assets (y_4), tertiary industry added value (y_5), urban and rural residents savings deposits of urban and rural (y_6), and town residents' per capita disposable income (y_7) are the output variables. Using the PCA, the seven output variables are converted into integrated variables (PCs). F is the set of PCs, and the WEE is a value determined by X and Y. Then, X serves as the input, and F serves as the output, and by calculating the DEA model, the WEE values are obtained.

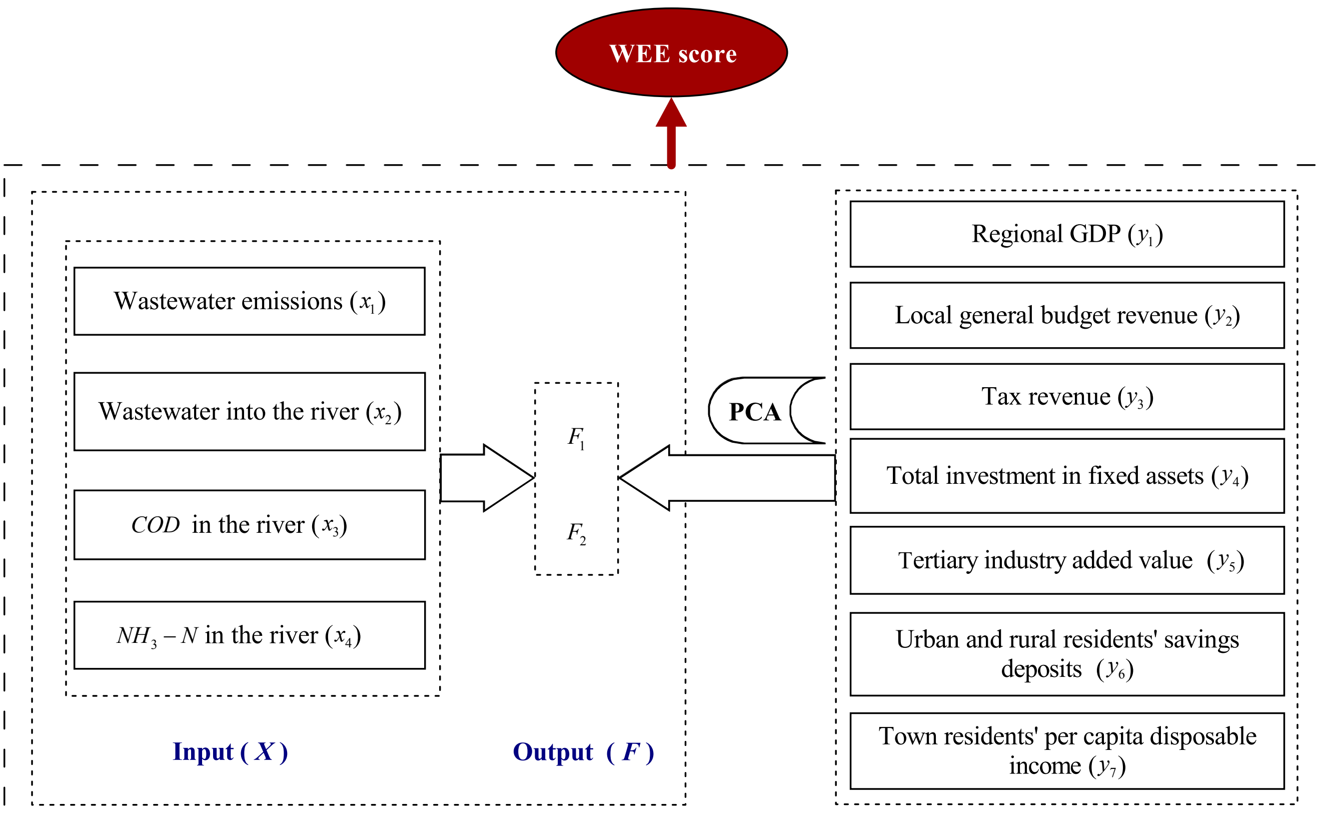

Fig. 2: Framework of the PCA-DEA model

3 Case study

In this section, we first describe the regions and areas in China as well as the associated data for the integrated energy and WEE measurements. Then, the PCA-DEA model is applied and the indicators for the WEE values in the Tuojiang River basin (Fig. 3) are calculated and analyzed.

3.1 Data collection

The Tuojiang River is one of the most important tributaries on the left bank upstream of the Yangtze River. As a key area for the construction of an ecological barrier for the upstream Yangtze River area and Sichuan Provinces economic development, the ecological security of the Tuojiang River is of great importance to further on economic development in Sichuan Province and the ecological security of Yangtze River basin. The forest cover rate in the Tuojiang river basin is only 6.1%, the lowest of all rivers in Sichuan.

Further, the Tuojiang river basin has much of sichuan province's industrial concentration development, so consequently has a significantly higher population density than other river basins in the province. There are 5 large and medium-sized cities in the basin; Chengdu, Deyang, Neijiang, Zigong, Luzhou; and thousands of large and medium-sized factories.

Accidents have already occurred which have had a dramatic influence on the economy and the lives of the residents. For example, in February 2004, the Tuojiang River water pollution accident resulted in nearly 1 million peoples having no access to drinking water for 25 days, 1 million kilograms of dead fish 1 million, and direct economic losses of 219 million yuan, in Jianyang, Ziyang, Zizhong, and Neijiang. Therefore, considering the important position of the Tuojiang river basin and the present less optimistic situation, studies on Tuojiang River basin water efficiency are imperative.

In this study, because of an absence of reliable data, we were only able to examine 8 regions in the Tuojiang River basin; Jintang County in Chengdu City (D1), Jingyang District in Deyang City (D2), Zhongjiang County in Deyang City (D3), Mianzhu County in Deyang City (D4), Yanjiang District in Ziyang City (D5), Jianyang County in Ziyang City (D6), Shizhong District in Neijiang City (D7), and Fushun County in Zigong City (D8). As mentioned before, the DEA-PCA approach was used to measure the efficiencies in the eight regions of the Tuojiang River basin in 2010. Each region was classified as a separate DMU. Therefore, a set of 8 DMUs was observed. In Fig. 3, the 8 sample areas are marked with triangles.

Fig. 3: Location of the study area

Based on the Chengdu 2010 statistical yearbook and other prefecture-level cities in Sichuan Province, we conducted a detailed analysis of the regional economic development levels in these 8 regions. Data for the indicators used in this case study are presented in Table 3.

Table 3: DMUs and raw data in DEA analysis

	x_1 (10^4)	x_2 (10^4)	x_3	x_4	y_1 (10^8)	y_2 (10^4)	y_3 (10^4)	y_4 (10^8)	y_5 (10^8)	y_6 (10^8)	y_7
D1	86.89	83.25	130.25	38.38	134.48	60175.00	43,028.00	110.54	50.95	90.97	15,290.00
D2	571.54	491.45	402.38	95.59	284.87	106018.00	46,978.00	147.49	77.61	226.92	18,072.00
D3	1452.56	1280.61	1300.05	525.51	161.91	24756.00	17,445.00	72.16	45.97	107.84	14,550.00
D4	394.00	374.08	310.30	10.60	118.10	56889.00	47,515.00	107.30	28.37	88.30	16,001.00
D5	1345.00	1105.00	2144.00	441.00	214.82	45201.00	25,956.00	130.25	45.98	123.35	15,248.00
D6	472.50	390.10	657.45	112.79	203.64	75025.00	51,984.00	132.07	51.05	149.63	16,493.00
D7	1655.9	1353.40	2311.25	406.35	126.85	15168.00	10,028.00	74.81	31.32	121.67	14,285.00
D8	739.00	603.80	1044.50	221.10	121.94	36218.00	21,649.00	64.00	36.36	104.24	13,846.00

3.2 Data envelopment analysis with pca

The input variable correlations are shown in Table 4. From Table 4 we can see that there is a strong correlation between some input variables. Further, the Kaiser-Meyer-Olkin (KMO) and Bartlett's Test were conducted, the results for which are given in Table 5.

Table 4: Correlations

		y_1	y_2	y_3	y_4	y_5	y_6	y_7
y_1	Pearson Correlation	1						
	Sig. (2-tailed)							
y_2	Pearson Correlation	0.721*	1					
	Sig. (2-tailed)	0.044						
y_3	Pearson Correlation	0.372	0.866**	1				
	Sig. (2-tailed)	0.364	0.005					
y_4	Pearson Correlation	0.781*	0.854**	0.761*	1			
	Sig. (2-tailed)	0.022	0.007	0.028				
y_5	Pearson Correlation	0.881**	0.786*	0.447	0.68	1		
	Sig. (2-tailed)	0.004	0.021	0.266	0.063			
y_6	Pearson Correlation	0.902**	0.712*	0.317	0.627	0.837**	1	
	Sig. (2-tailed)	0.002	0.048	0.445	0.096	0.01		
y_7	Pearson Correlation	0.775*	0.938**	.800*	0.885**	0.747*	0.767*	1
	Sig. (2-tailed)	0.024	0.001	0.017	0.003	0.033	0.026	

$*$. Correlation is significant at the 0.05 level (2-tailed).

$**$. Correlation is significant at the 0.01 level (2-tailed).

Table 5: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.739
Bartlett's Test of Sphericity	Approx. Chi-Square	48.311
	df	21
	Sig.	0.001

Having carried out the PCA for the output variables, the synthetic outputs were derived. In Table 6, the factorial matrices and the variables from which we formed our synthetic input variables are shown. The first column is the names of the synthetic variables, the second column are the eigenvalues for the PCs; $\lambda_1 = 5.45 > 1$, $\lambda_2 = 1.047 > 1$, and all others are smaller than 1, with λ_1 and λ_2 capturing 92.816% of the variation in the output variables. The third column shows the relative weight of the original variable with respect to a specific principal component, and the fourth column shows the cumulative weight of the PCs. According to the eigenvectors, the first PC represents a weighted average of all input variables. Columns 5-7 show the correlation coefficients between the PCs and the input variables. As shown in Figure 4, there are two eigenvalues greater than 1. According to the principle $kM > 1$, we selected the first two PCs which, as stated, covered 92.816% of the original variable information. The equations for these first two PCs are:

$$\begin{cases} z_1 = 0.38y_1 + 0.41y_2 + 0.31y_3 + 0.39y_4 + 0.37y_5 + 0.41y_6 + 0.36y_7 \\ z_2 = -0.40y_1 + 0.22y_2 + 0.65y_3 + 0.19y_4 - 0.32y_5 + 0.15y_6 - 0.45y_7 \end{cases} \tag{1}$$

Table 6: Total variance explained

comp-	Initial Eigenvalues			Extraction Sums of Squared Loadings		
onent	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.45	77.861	77.861	5.45	77.861	77.861
2	1.047	14.954	92.816	1.047	14.954	92.816
3	0.243	3.472	96.288	0.243	3.472	96.288
4	0.18	2.567	98.855	0.180	2.567	98.855
5	0.04	0.569	99.424	0.040	0.569	99.424
6	0.028	0.405	99.829	0.028	0.405	100.000
7	0.012	0.171	100.000	0.012	0.171	100.000

The component plot for the first two PCs is shown in Fig. 4. From Eq. 1 and Fig. 4, we can see that:

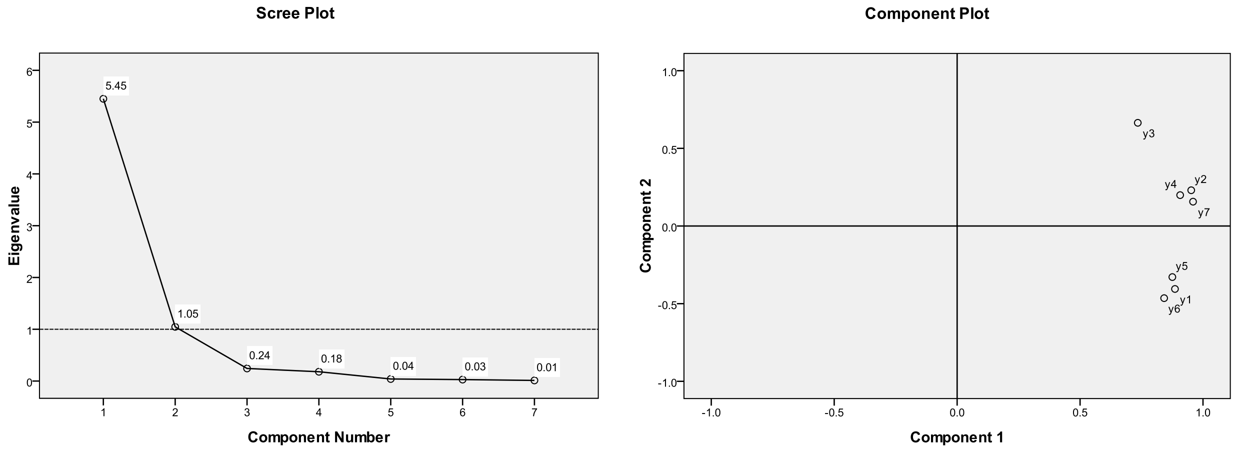

Fig. 4: Scree and component plots

(a) In the first principle component expression z_1, the coefficients for the local general budget revenue (y_2), the urban and rural residents savings deposits (y_6) and the total investment in fixed assets (y_4) play a main role. The coefficients for the local general budget revenue (y_2) and urban and rural residents' savings deposits (y_6) are the largest for the expression z_1. The coefficients for the total investment in fixed assets (y_4), regional Gross Domestic Product (regional GDP, y_1), tertiary industry added value (y_5), and the town residents' per capita disposable income (y_7) decrease in turn. The tax revenue coefficient (y_3) is the lowest. The integrated indicator z_1 can be seen as the urbanization level.

(b) In the expression for z_2, the tax revenue (y_3) is the largest. The coefficients for the regional Gross Domestic Product (regional GDP, y_1) and town residents' per capita disposable income (y_7) are larger than the remaining indicator coefficients. The tax revenue (y_3), Gross Domestic Product (GDP, y_1), and town residents' per capita disposable income (y_7) can be seen to play a main role in the second principle component z_2. We can regard the second principle component (z_2) as a composite indicator of economic profitability and growth ability, or the economic development level. The variance contribution rate of the second PC z_2 was 25.63%, which indicates that profitability and growth play an important part in a region's economic survival and development.

After the PCA process, our synthetic output variables were obtained, and the final input and output variables are shown in Table 7. There were no negative PCs.

Table 7: Final input and output variables of DEA modelp

	Input				Output	
	x_1	x_2	x_3	x_4	z_1	z_2
D1	86.89	83.25	130.25	38.38	43750.97	34465.41
D2	571.54	491.45	402.38	95.59	64829.72	46028.87
D3	1452.56	1280.61	1300.05	525.51	20986.53	10224.68
D4	394.00	374.08	310.30	10.60	44063.06	36328.11
D5	1345.00	1105.00	2144.00	441.00	32304.21	20018.81
D6	472.50	390.10	657.45	112.79	53119.05	43054.23
D7	1655.90	1353.40	2311.25	406.35	14633.79	3398.08
D8	739.00	603.80	1044.50	221.10	26706.89	15865.87

To calculate the efficiency value of D1, the following Eq. 2 was developed by inputting the values of the indices in Table 7.

$$\min e_1$$

$$s.t.\begin{cases} 86.89\omega_1+571.54\omega_2+1452.56\omega_3+394.00\omega_4+1345.00\omega_5+472.50\omega_6 \\ +1655.90\omega_7+739.00\omega_8\leq 86.89e_1 \\ \\ 83.25\omega_1+491.45\omega_2+1280.61\omega_3+374.08\omega_4+1105.00\omega_5+390.10\omega_6 \\ +1353.40\omega_7+603.80\omega_8\leq 83.25e_1 \\ \\ 130.25\omega_1+402.38\omega_2+1300.05\omega_3+310.30\omega_4+2144.00\omega_5+657.45\omega_6 \\ +2311.25\omega_7+1044.50\omega_8\leq 130.25e_1 \\ \\ 38.38\omega_1+95.59\omega_2+525.51\omega_3+10.60\omega_4+441.00\omega_5+112.79\omega_6 \\ +406.35\omega_7+221.10\omega_8\leq 38.38e_1 \\ \\ 43750.97\omega_1+64829.72\omega_2+20986.52\omega_3+44063.06\omega_4+32304.21\omega_5 \\ +53119.05\omega_6+14633.79\omega_7+26706.89\omega_8\geq 43750.97 \\ \\ 34465.41\omega_1+46028.87\omega_2+10224.68\omega_3+36328.11\omega_4+20018.81\omega_5 \\ +43054.23\omega_6+3398.08\omega_7+15865.87\omega_8\geq 34465.41 \\ \\ \omega_j\geq 0, j=1,\cdots,8 \end{cases} \tag{2}$$

4 Results and discussion

Through solving Eq. 2 using Lingo19.0, we determined the optimal solution to the DEA-PCA model for D1. Similarly, we can determine the optimal solutions for the other 7 DMUs. All optimal solutions are presented in Table 8. The value of e_i was a WEE score of ith DMU ($i = 1, \cdots, 8$).

4.1 Efficiency measurement results and findings

The first performance measure was the number of DEA efficient DMUs, of which only one was found. Jintang (x_1) was found to have the highest WEE, with a WEE score of 1. Because $e_1 = 1$, Jintang (D1) was deemed to be DEA efficient. Because $e_i < 1 (i = 2, \cdots, 8)$, Jingyang (D2), Zhongjiang (D3), Mianzhu (D4), Yanjiang (D5), Jianyang (D6), Shizhong (D7), and Fushun (D8) were all found to be DEA inefficient. Make those relative efficiency values in descending order, it can be seen from Fig. 5 that except for Jintang (D1), the overall WEE scores in the other seven regions were a little low at less than 0.5. Further, the difference between the regional WEE scores was relatively large reaching as high as 1 and as low as 0.02. Jingyang (D2) and Mianzhu (D4) were found to have similar WEE values approaching 0.5, and Zhongjiang (D3), Yanjiang (D5), Shizhong (D7), and Fushun (D8) also had similar WEE values, but approaching 0.

Table 8: The optimal solutions of the models

DMUs	e_i	$\omega^i \geq 0 (i = 1, \cdots, 8)$
D1(Jintang)	1.00	$[1.00, 0, 0, 0, 0, 0, 0, 0]^T$
D2(Jingyang)	0.48	$[1.48, 0, 0, 0, 0, 0, 0, 0]^T$
D3(Zhongjiang)	0.05	$[0.48, 0, 0, 0, 0, 0, 0, 0]^T$
D4(Mianzhu)	0.44	$[1.05, 0, 0, 0, 0, 0, 0, 0]^T$
D5(Yanjiang)	0.06	$[0.74, 0, 0, 0, 0, 0, 0, 0]^T$
D6(Jianyang)	0.27	$[1.25, 0, 0, 0, 0, 0, 0, 0]^T$
D7(Shizhong)	0.02	$[0.33, 0, 0, 0, 0, 0, 0, 0]^T$
D8(Fushun)	0.08	$[0.61, 0, 0, 0, 0, 0, 0, 0]^T$

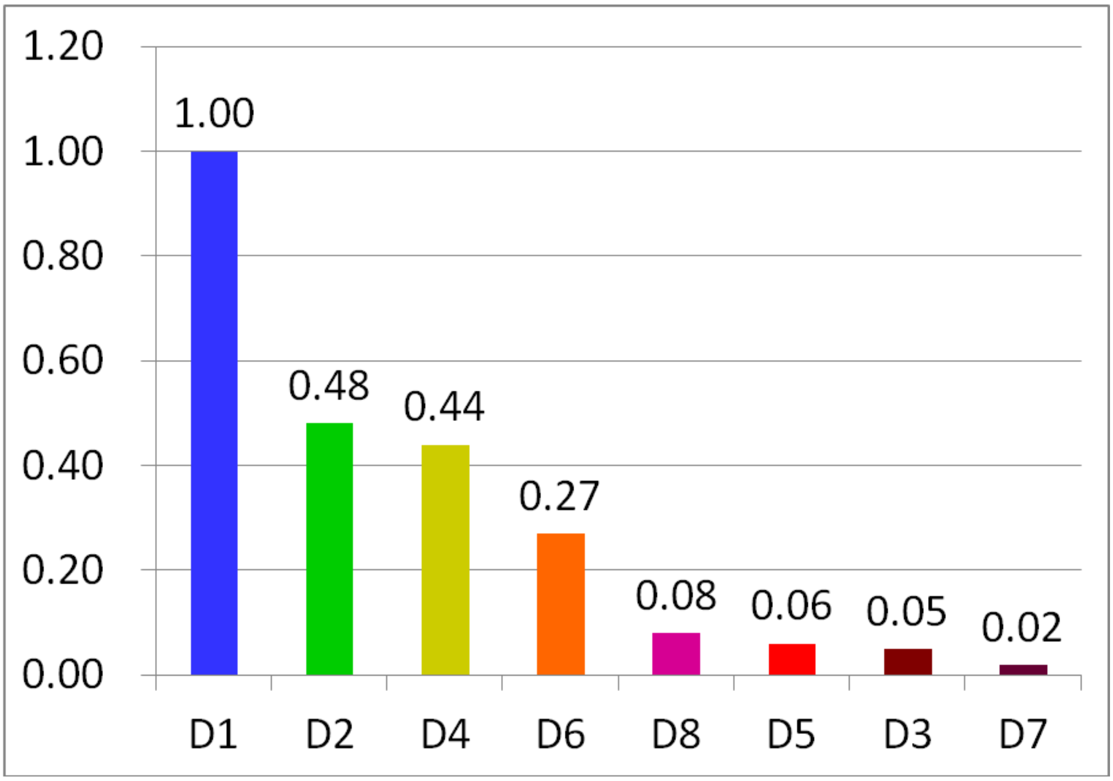

Fig. 5: The efficiency scores (WEE scores) and the rankings of the DMUs

Based on the values of e_i and $\omega^i \geq 0(i = 1, \cdots, 8)$ in Table 8, $\frac{1}{e_i}\sum_{j=1}^{8}\omega_j^i$ can be calculated to analyze a return to scale (RTS) for the DMUs, as shown in Table 9. $\frac{1}{e_1}\sum_{j=1}^{8}\omega_j^1 = 1.00$, so D1 (Jintang) can be seen to have constant returns to scale. $\frac{1}{e_i}\sum_{j=1}^{8}\omega_j^i > 1$, and D2 (Jingyang), D3 (Zhongjiang), D4 (Mianzhu), D5 (Yanjiang), D6 (Jianyang), D7 (Shizhong), and D8 (Fushun) have decreasing returns to scale. That is, the increase in output is unable to keep up with the increase in investment. This has important implications for providing important guidance for producers when they make their production plans using the RTS to analyze the relative resource use efficiency of the production units on the production frontier surface.

4.2 Further analysis

The major results of this study show that:

(i) There are significant differences in the water-environment efficiencies in each region. Jingtang, as a model in the forefront compared with other provinces, should lead the other regions in improving their industrial technologies and economic development.

(ii) From the perspective of a specific area, the WEE levels in most regions are lower than 50% of the ideal or target level. Therefore, except for Jingtang, most other regions have significant room for improvement in terms of water resources protection.

(iii) To order to maintain or increase the water-environment efficiency for the regions under a decreasing RTS, it is acceptable, but not recommended that they increase their economic development scales. Ideally, decision makers should reduce the region's water pollutant emissions and reduce the economic development to save water resources. For those regions with a constant return to scale, it is also acceptable but not recommended that factories maintain their economic development scale.

Table 9: The return to scale of the DMUs

DMU	$\frac{1}{e_i}\sum_{j=1}^{8}\omega_j^i$	RTS
D1(Jintang)	$\frac{1}{e_1}\sum_{j=1}^{8}\omega_j^1 = 1.00$	constant RTS
D2(Jingyang)	$\frac{1}{e_2}\sum_{j=1}^{8}\omega_j^2 = 3.08$	decreasing RTS
D3(Zhongjiang)	$\frac{1}{e_3}\sum_{j=1}^{8}\omega_j^3 = 9.60$	decreasing RTS
D4(Mianzhu)	$\frac{1}{e_4}\sum_{j=1}^{8}\omega_j^4 = 2.39$	decreasing RTS
D5(Yanjiang)	$\frac{1}{e_5}\sum_{j=1}^{8}\omega_j^5 = 12.33$	decreasing RTS
D6(Jianyang)	$\frac{1}{e_6}\sum_{j=1}^{8}\omega_j^6 = 4.63$	decreasing RTS
D7(Shizhong)	$\frac{1}{e_7}\sum_{j=1}^{8}\omega_j^7 = 16.5$	decreasing RTS
D8(Fushun)	$\frac{1}{e_8}\sum_{j=1}^{8}\omega_j^8 = 7.63$	decreasing RTS

(iv) These results imply that policies which enhance the performance of the WEE in the Tuojiang River basin should take these regional disparities into consideration.

Several suggestions can be proposed based on above suggestions. First, water policies should be tailored to areas with different levels of development. The regions which have low WEE values and high enhancement potential should focus on saving their water resources. The service industry needs to be better developed as this sector has the possibility of improving productivity and meeting the residents needs. The proportion of the tertiary industry in the regional economy should be increased, industrial structure optimized and environmentally friendly growth achieved.

5 Conclusion

Over the last several decades, Chinas economy has developed significantly. With such rapid economic development, many water pollution problems have occurred. Firstly, the concept of WEE was proposed and defined according to previous research. Secondly, to accurately evaluate regional economic development, the input and output index system for the WEE evaluation was developed using the PCA. Thirdly, based on a minimum output improvement strategy, we proposed a DEA-PCA model. Finally, the proposed original model was applied to the Tuojiang River basin to analyze the water pollutant emissions from regional economic development. These empirical results provided policy implications to alleviate the water concerns in the Tuojiang River, but also provide evidence for broader policy implications.

The proposed approach was developed based on a constant return to scale water-environment DEA technology. It can be easily extended to cases of variable returns to scale (VRS) technology and can be applied to other practical settings such as complete provinces and countries.

Problems to be addressed in future studies may include the following: a) China's environmental efficiency was compared only between the regions of the Tuojiang River basin. The efficiency scores would likely have been much worse if the data included other advanced regions; b) The data collected was only one year cross-sectional. To capture the more dynamic nature of the industry in China, panel data should be collected in the future. All these remain avenues for future research.

Acknowledgements

This research was supported by the Key Program of NSFC (Grant No. 70831005,70971091), the National Natural Science Foundation for Young Scholars of China (Grant No. 71301109), the Research Foundation of Ministry of Education for the Doctorpal Program of Higher Education of China (Grant No. 20130181110063), the Western and Frontier Region Project of Humanity and Social Sciences Research, Ministry of Education of China (Grant no. 13XJC630018), the Foundation for Fostering Young Scientists of Sichuan Province of China (Grant No. 2013DTPY0020), and the Initial Funding for Young Teachers of Sichuan University (Grant no. 2013SCU11014).

References

[1] Adler, N., Golany, B., 2001. Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe. European Journal of Operational Research, 132(2), 18-31.

[2] Adler, N., Golany, B., 2002. Including principal component weights to improve discrimination in data envelopment analysis. Journal of the Operational Research Society, 53(9), 985-991.

[3] Adler, N., Friedman, L., 2002. Sinuany-Stern Z. Review of ranking methods in the data envelopment analysis context. European Journal of Operational Research, 140(2), 249-65.

[4] Azadeh, A., et al., 2007. An integrated DEA PCA numerical taxonomy approach for energy efficiency assessment and consumption optimization in energy intensive manufacturing sectors. Energy Policy, 35, 3792-3806.

[5] Bai, Y.P., Hao, Y.P., 2012. Research on regional environmental performance and its influential factor in the north of China based on non-radial DEA model. ICEWE 2012, 7, 171-180.

[6] Banker, R.D., et al., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30(9), 1078-92.

[7] Barros, C.P., et al., 2012. The technical efficiency of the Japanese banks: non-radial directional performance measurement with undesirable output. Omega, 40(1), 1-8.

[8] Beasley, J., 1995. Determining teaching and research efficiencies. Journal of the Operational Research Society, 46(4), 441-452.

[9] Bian, Y.W. , Yang, F., 2010. Resource and environment efficiency analysis of provinces in China: a DEA approach based on Shannons entropy. Energy Policy, 38, 1909-1917.

[10] Cai, X.C., Xiao, X.A., Study on energy consumption efficiency of Chinas various regions based on super - efficiency DEA. Statistics & Information Forum, 25(4), 33-40.

[11] Słlvia, C., et al., 2014. Translating removal efficiencies into operational performance indices of wastewater treatment plants. Water Research, 57, 202-214.

[12] Charnes, A., et al., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research, 2, 429-444.

[13] Charnes, A., et al., 1990. Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. Journal of Econometrics, 46(1), 73-91.

[14] Chiu, Y., Wu, M.F., 2010. Environmental efficiency evaluation in China: application of undesirable data envelopment analysis. Polish Journal of Environmental Studies, 19(6), 1159-1169.

[15] Cook, W.D., et al., 2000. Multi-component efficiency measurement and shared inputs in data envelopment analysis: an application to sales and service performance in bank branches. Journal of Productivity Analysis, 14(3), 209-224.

[16] Trevisan, D., et al., 2012. POPEYE: A river-load oriented model to evaluate the efficiency of environmental policy measures for reducing phosphorus losses. Journal of Hydrology, 254-266.

[17] Drake, L., Hall, M., 2003. Efficiency in Japanese banking: an empirical analysis. Journal of Banking & Finance, 27(5), 891-917.

[18] Farrell, M.J., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society Series A (General), 120(3), 253-290.

[19] Freeman, A.M., et al., 1973. Economics of environmental policy. New York: John Wiley and Sons, Inc.

[20] Fukuyama, H., Weber, W.L., 2002. Estimating output allocative efficiency and productivity change: application to Japanese banks. European Journal of Operational Research, 137(1), 177-190.

[21] Kortelainen, M., 2008. Dynamic environmental performance analysis: a Malmquist index approach. Ecological Economics, 64(4), 701-715.

[22] Hu, J.L., et al., 2005. Under the shadow of Asian Brown Clouds: unbalanced regional productivities in China and environmental concerns. The International Journal of Sustainable Development & World Ecology, 12(4), 429-442.

[23] Hu, J.L., Wang, S.C., 2006. Total-factor energy efficiency of regions in China. Energy Policy, 34(17), 3206-3217.

[24] Jia, Y.P., Liu, R.Z., 2012. Study of the Energy and Environmental Efficiency of the Chinese economy based on a DEA Model. Procedia Environmental Sciences, 13, 2256-2263.

[25] Liang, L.F., et al., 2006. DEA models for supply chain efficiency evaluation. Annals of Operations Research, 145(1), 35-49.

[26] Luo, Y., 2012. DEA-based research on indicator selection and environmental performance measurement. PH.D Thesis, University of Science and Technology of China.

[27] Martins-Filho, C., Yao, F., 2007. Nonparametric frontier estimation via local linear regression. Journal of Econometrics, 141(1), 283-319.

[28] Martins-Filho, C., Yao, F., 2008. A smooth nonparametric conditional quantile frontier estimator. Journal of Econometrics, 143(2), 317-333.

[29] Reinhard, S., et al., 2000. Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA. European Journal of Operational Research, 121, 287-303.

[30] Roger, L.O., et al., 2012. Water quality sample collection, data treatment and results presentation for principal components analysis C literature review and Illinois River watershed case study. Water Research, 46(9), 3110-3122.

[31] Seiford, L.M., Zhu, J., 1998. Identifying excesses and deficits in Chinese industrial productivity (1953-1990): a weighted data envelopment analysis approach. Omega, 26(6), 279-296.

[32] Seiford, L.M., Zhu, J., 1999. Infeasibility of super-efficiency data envelopment analysis models. Infor, 37(2), 174-187.

[33] Shi, G.M., et al., 2010. Chines regional industrial energy efficiency evaluation based on a DEA model of fixing non-energy inputs. Energy, 38(10), 6172-6179.

[34] Stroosnijder, L., et al., 2012. Improving water use efficiency in drylands. Current Opinion in Environmental Sustainability, 4(5),497-506.

[35] Tong, X.J., et al., 2014 Ecosystem water use efficiency in a warm-temperate mixed plantation in the North China. Journal of Hydrology, 512, 221-228.

[36] Wang, K., et al., 2013. Chinas regional energy and environmental efficiency: a DEA window analysis based dynamic evaluation. Mathematical and Computer Modelling, 58, 1117-1127.

[37] Lorenzo-Toja, Y., et al., 2015. Eco-efficiency analysis of Spanish WWTPs using the LCA + DEA method. Water Research, 68, 651-666.

[38] Yang, L., Wang, K.L., 2013. Regional differences of environmental efficiency of Chinas energy utilization and environmental regulation cost based on provincial panel data and DEA method. Mathematical and Computer Modelling, 58(5), 1074-1083.

[39] Zhou, P., et al., 2008. A survey of data envelopment analysis in energy and environmental studies. European Journal of Operational Research, 189(1), 1-8.

[40] Zhou, Y., et al., 2013. Environmental efficiency analysis of power industry in China based on an entropy SBM model. Energy Policy, 57, 68-75.

[41] Zhu, J., 2014. Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets and DEA excel solver. Vol.213, Springer.