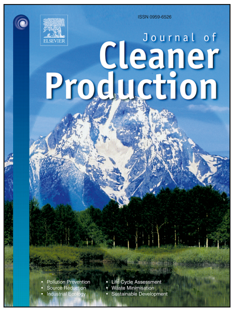
Journal Pre-proof



Sustainability efficiency assessment of wastewater treatment plants in China: A data envelopment analysis based on cluster benchmarking

Hui Jiang, Ming Hua, Jin Zhang, Panpan Cheng, Zhen Ye, Mai Huang, Qiang Jin

PII: S0959-6526(19)33599-1

DOI: https://doi.org/10.1016/j.jclepro.2019.118729

Reference: JCLP 118729

To appear in: *Journal of Cleaner Production*

Received Date: 4 July 2019

Revised Date: 25 September 2019

Accepted Date: 4 October 2019

Please cite this article as: Jiang H, Hua M, Zhang J, Cheng P, Ye Z, Huang M, Jin Q, Sustainability efficiency assessment of wastewater treatment plants in China: A data envelopment analysis based on cluster benchmarking, *Journal of Cleaner Production* (2019), doi: https://doi.org/10.1016/j.jclepro.2019.118729.

Sustainability efficiency assessment of wastewater treatment plants in China: A data envelopment analysis based on cluster benchmarking

Hui Jiang[1], Ming Hua[2], Jin Zhang[3], Panpan Cheng[1] , Zhen Ye[4],

Mai Huang[1], Qiang Jin[1]*

[1]School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

[2]School of the Environment, Nanjing University, Nanjing 210023, China

[3]School of Civil Engineering, Yantai University, Yantai 264005, China

[4]The Bartlett School of Construction and Project Management, University College London, London WC1E 6BT, UK

***Corresponding author:**

Qiang Jin, Ph.D.

School of Environmental Science and Engineering
Shanghai Jiao Tong University
Shanghai, 200240
China

Phone: +86-18621863638
Fax: +86-21-54740825
E-mail: jinqiang@sjtu.edu.cn

Word Count: 7375

Abstract

Quantitative evaluation on the efficiency of wastewater treatment plants (WWTPs) is a key issue that needs to be solved. For this purpose, data envelopment analysis (DEA) was employed to establish a comprehensive efficiency evaluation system on WWTPs, including three inputs of operating cost, electricity consumption and labor, three desirable outputs of chemical oxygen demand (COD) removal rate, ammonia nitrogen (NH_3-N) removal rate and reclaimed water yield, and one undesirable output of dry sludge yield. 861 WWTPs in China were assessed by a slacked-based DEA model based on cluster benchmarking. The technology gap ratio (TGR) confirmed that large WWTPs operated more efficiently than small ones. The WWTPs had an average efficiency score of 0.611. Among them, 170 samples were relatively efficient with a score of 1, which means these samples could be a benchmark for other inefficient samples. Different degrees of input excesses or output shortfalls existed in 691 inefficient samples and these samples should be the key objects to improve the operational efficiency. Furthermore, through the Kruskal-Wallis test, the influent COD concentration and capacity load rate showed significant effects on the WWTP performance. These findings, derived from a simple but effective framework, have potential value for managers to make decisions.

Keywords: Efficiency assessment; Data envelopment analysis; Cluster benchmarking; Sustainability; Wastewater treatment plant; Slack based measure

1. Introduction

Wastewater treatment plants (WWTPs) help protect the surrounding water environment by removing the contaminants from wastewater. In recent years, many countries, particularly quickly developing countries, have made great efforts to strengthen wastewater treatment. In China, for example, the number of municipal WWTPs has increased rapidly during the past 40 years. By 2016, 3552 WWTPs have been built and operated, with a total treatment capacity of 1.76×10^8 m^3/d (China

Urban Water Association, 2017). There is still a huge demand for future construction and technical upgrading of WWTPs. According to China's 13th Five-Year Plan (NDRC, 2016), by 2020, an additional treatment capacity of 50.22 million cubic meters per day is targeted; 42.2 million cubic meters per day of wastewater treatment facilities will be upgraded; 60,000 tons of extra daily sludge (with 80% water content) will be disposed of harmlessly and the newly-added reclaimed water utilization facilities will reach a scale of 15.05 million cubic meters per day.

To achieve effective treatment of urban wastewater and ensure sustainable development of the society, not only should the number of centralized wastewater treatment facilities be greatly increased, but also the operational efficiency of the treatment facilities should be improved. Therefore, how to scientifically evaluate and improve the operational efficiency of existing wastewater treatment facilities has attracted much attention. The efficiency evaluation can help the governmental departments formulate reasonable policies to promote the healthy development of WWTPs, and also provide targeted improvement recommendations for enterprises.

It is necessary to assess the WWTP efficiency from a comprehensive perspective. The operation of WWTPs is aimed at reducing wastewater pollution from human activities to minimize adverse impacts on the natural environment and human health (Wang et al., 2012). Proper treatment and further reuse of wastewater can help solve water shortage problems and save valuable resources. However, owing to chemical consumption, energy consumption, and various environmental emissions, WWTPs also bring many additional environmental impacts (Hospido et al., 2004). For instance, sludge is a by-product of wastewater treatment and contains lots of heavy metals, organic pollutants, pathogenic bacteria, and parasite eggs. As the amount of wastewater treatment increases, sludge production is also increasing rapidly, which easily leads to secondary pollution (Buonocore et al., 2018). Another challenge in wastewater treatment is to minimize the economic and manpower inputs involved in the operation. Therefore, in order to realize the goal of real sustainable development, the negative effects of a plant should not exceed the benefits of environmental remediation.

DEA is a mature nonparametric method to calculate the relative efficiency of similar entities with various input and output indicators (Charnes et al., 1978). Recently, DEA has been gradually applied in the efficiency evaluation of WWTPs. For example, Gao et al. (2006) applied a CCR model to evaluate five sewage treatment plants in Urumqi, and discussed the influencing factors of operational efficiency, but limited by the small number of samples. Hernandez-Sancho et al. (2011) used a non-radial DEA method to assess energy efficiency for WWTPs in Spain and studied the operating variables that lead to the difference among plants. Sala-Garrido et al. (2011) utilized the DEA meta-frontier model to evaluate four technologies involved in 99 Spanish WWTPs. Dong et al. (2017) innovatively combined a tolerance method with the DEA model to assess the WWTP eco-efficiency under uncertainty analysis. Generally speaking, the application of DEA in wastewater treatment industry is gradually diversified. These studies have demonstrated the adaptability of DEA and have achieved the evaluation of WWTPs in some regions and countries, but there are still some deficiencies. For instance, previous studies did not take into account such indicators as labor input, sludge production, and reclaimed water production. Most of them only considered operating costs and pollutant removal, so that the evaluation could not comprehensively reflect the sustainability of WWTPs. In addition, few studies did relevant cluster benchmarking and comparison of technology gap ratio (TGR) under the scale effect. What's more, quantifying potential improvements in each indicator is critical to supporting the decision-making process. Existing studies mainly focused on efficiency evaluation and failed to provide improvement suggestions for each WWTP, for lacking further analysis in input excesses and output shortfalls. Therefore, this paper aims to overcome these deficiencies and obtain more scientific and useful assessment results.

According to the difference in how to analyze the distance between units and measure the production frontier, DEA can generally be divided into radial and non-radial models (Castellet and Molinos-Senante, 2016). A notable disadvantage of radial models is that they set reduction/enlargement of input/output vectors in proportion, which is inconsistent with most actual situations. To solve this problem, Tone (2001) put forward a slack based measure (SBM) model, which is a non-radial DEA model designed to deal straightly with input excesses and output shortfalls. Hence, this study tends to adopt an input-oriented SBM-DEA model based on variable returns to scale (VRS) and cluster benchmarking to evaluate WWTPs.

The purpose of this study is to assess the sustainable performance efficiency of 861 WWTPs in China, with a view to providing effective and targeted countermeasures. The main contents are as follows: (i) evaluate the efficiency of WWTPs with a comprehensive index system and cluster analysis; (ii) determine the best practices from samples and quantify the potential improvement for each WWTP; (iii) identify underlying factors that affect the plant performance. The assessment takes into account multiple dimensions and statistical analysis of large samples, making it sufficient to reveal the reasons for the efficiency gap. As this study can provide reliable scientific benchmarking data and control information, the final results are expected to have great practical significance for the construction, operation and management of wastewater treatment plants.

2. Methodology

Different WWTPs run at different efficiency levels. Choosing an appropriate DEA model and evaluation index is the key to obtain useful information, such as efficiency scores and projections. Generally, the choice of model needs to consider the undesirable output, orientation, benchmark and so on. To ensure the simplicity and effectiveness of the model, only a certain number of indicators are needed. However, the efficiency of WWTPs may be affected by various factors that have not been chosen as input or output indicators. Thus, after assessing the efficiency, the next phase is to find out the explanatory factors.

2.1. DEA model

DEA, developed by Charnes, Cooper and Rhodes, is a method to evaluate the relative efficiency of decision making units (DMUs) with various input and output indicators (Charnes et al., 1978). Efficiency is relatively assessed, which means that the efficiency of DMUs should be analyzed mutually. After the assessment, DMUs can be divided into two groups: efficient and inefficient. The envelope formed by efficient DMUs is called an efficient frontier, which covers inefficient DMUs like an envelopment, and the name of DEA comes from it.

Conventional DEA models suppose that every output index should be maximized under the current input level (Bi et al., 2014), such as CCR-DEA model. However, in the actual production process, desirable outputs are often accompanied by the production of undesirable outputs, such as solid waste. As undesirable outputs should be minimized as much as possible, the traditional DEA model is not well suited to assess the sustainability efficiency, which requires certain special treatment to achieve a more accurate assessment (Zhang et al., 2016).

The indirect method is to convert the numeric values related to the undesirable output into an input or a desirable output (Reinhard et al., 2000; Scheel, 2001). However, this method has certain limitations. For instance, considering an undesirable output as an input is impractical to a degree, since the input-output structure defining the production process has been lost (Seiford and Zhu, 2002). Performing a reciprocal or negative conversion process on the undesirable output may bias the final efficiency value (Fare and Grosskopf, 2004).

By comparison, the direct method does not change the undesirable output, but integrates it with constraints into the DEA model (Xiao et al., 2018; Yang et al., 2018). It has been proved that dealing with the undesirable output in its original form conforms to the standard axioms and physical laws of production theory (Fare and Grosskopf, 2004). Therefore, this study adopted the direct method to deal with the undesirable output.

In this paper, slack based measurement (SBM) was applied to the DEA model involved (Tone, 2001), which can modify the constraints to add the slacks of undesirable outputs into the target function. Zhou et al. (2008) have proved that the SBM model is especially proper for handling undesirable outputs, and it has very high recognition.

As stated above, an SBM-DEA model is constructed for this study. Suppose there are n DMUs, each DMU has k inputs, l desirable outputs and m undesirable outputs, which are denoted as $x \in R^k$, $y \in R^l$ and $u \in R^m$, respectively. Define three matrices X, Y, U as $X=[x_1,\ldots,x_n] \in R^{k\times n}$, $Y=[y_1,\ldots,y_n] \in R^{l\times n}$, $U=[u_1,\ldots,u_n] \in R^{m\times n}$. The mathematical expression of the input-oriented SBM based on VRS is shown as follows:

$$
\begin{aligned}
&\text{Min } \rho^* = \frac{1-\frac{1}{k}\sum_{i=1}^{k} {s_i}^- / s_{in}}{1+\frac{1}{l+m}\left(\sum_{r=1}^{l} {s_r}^+ / u_{rn} + \sum_{t=1}^{m} {s_t}^- / u_{tn}\right)} \\
&\text{s.t.} \sum_{j=1}^{n} \lambda_j x_{ij} + {s_i}^- = x_{in}, i = 1,2,\ldots,k \\
&\sum_{j=1}^{n} \lambda_j y_{rj} - {s_r}^+ = y_{rn}, r = 1,2,\ldots,l \\
&\sum_{j=1}^{n} \lambda_j u_{tj} + {s_t}^- = u_{tn}, t = 1,2,\ldots,m \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j \geq 0, {s_i}^- \geq 0, {s_r}^+ \geq 0, {s_t}^- \geq 0, j = 1,2,\ldots,n
\end{aligned}
\tag{1}
$$

where ${s_i}^-$, ${s_r}^+$ and ${s_t}^-$ represent the input excesses, desirable output shortfalls and undesirable output excesses, respectively. λ_j is a non-negative weight vector. In general, the term "efficiency" refers to the optimal use of resources to satisfy human desires and needs under given the inputs and technologies. $\text{Min}\,\rho$ is the objective function that defines the efficiency for each DMU, and the objective value ρ^* in the range of 0 and 1 denotes the efficiency score. The higher the value of ρ^* is, the better the efficiency of the DMU will be. Let the best solution for the above program be $(\lambda^*, {s_i}^{-*}, {s_r}^{+*}, {s_t}^{-*})$.

The DMU is efficient only when $\rho^*=1$ and ${s_i}^{-*} = {s_r}^{+*} = {s_t}^{-*} = 0$. If $\rho^* < 1$, the relevant DMU is inefficient and its input and output indices should be improved. Generally, the following SBM-projection can be employed to improve the efficiency of an inefficient DMU_0 (x_0, y_0, u_0) by reducing input excesses and undesirable output excesses, as well as making up for desirable output shortfalls:

$$x_0 - {s_i}^{-*} \to x_0', \; y_0 + {s_r}^{+*} \to y_0', \; u_0 - {s_t}^{-*} \to u_0' \tag{2}$$

Nevertheless, the DEA method considers that samples have similar properties when evaluating the efficiency of DMUs as a whole. The robustness of DEA results is determined by the homogeneity of samples (Corton and Berg, 2009), therefore traditional DEA models cannot be used to compare the DMUs with diverse characteristics (Lozano-Vivas et al., 2002).

Cluster benchmarking is a method of dividing a set of DMUs into groups (i.e. clusters) according to certain attributes, so DMUs in the same cluster are more similar to each other than to DMUs in other clusters. The major purpose of grouping is to maximize the homogeneity of DMUs in the same cluster and the heterogeneity of DMUs in different clusters (Patra et al., 2011). To achieve the efficiency evaluation of objects relative to the optimal practice under the respective clusters, it is necessary to construct the production frontier separately (Zhu, 2013).

Self-benchmarking refers to each cluster taking itself as a "reference set" and conducting "self-assessment" separately. By comparing the results of the non-grouped DEA model (traditional method, all DMUs as a reference set) and the results of self-benchmarking after grouping, the technology gap ratio (TGR) can be obtained. According to Rao et al. (2003), the formula for calculating TGR is as follows:

$$\mathrm{TGR} = \frac{\rho^*_{\mathrm{meta}}}{\rho^*_{\mathrm{cluster}}} \tag{3}$$

where ρ^*_{meta} refers to the efficiency value based on the meta-frontier, $\rho^*_{\mathrm{cluster}}$ refers to the efficiency value based on the cluster-frontier.

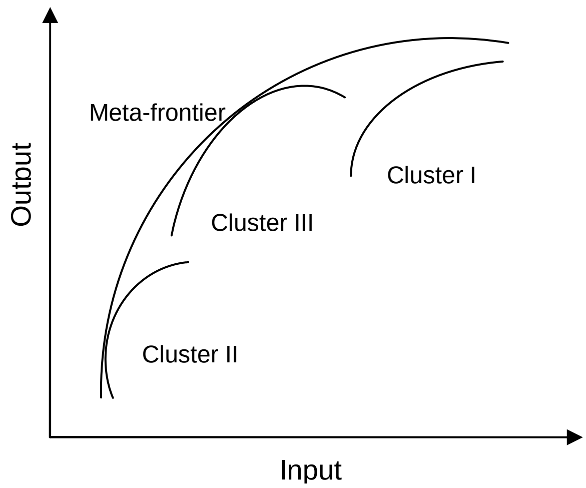


Fig. 1. Meta-frontier and cluster-frontier.

TGR can reflect the gap between the cluster-frontier and the meta-frontier. It is employed to measure the technical efficiency gap of the same DMU under different frontiers. At the same time, TGR can also reflect the necessity of dividing different groups (Rao et al., 2003). The smaller the TGR value is, the greater the necessity of grouping will be, and vice versa. As illustrated in Fig. 1, the meta-frontier is an envelope curve higher than the frontier of each cluster, so $\rho^*_{cluster} \geq \rho^*_{meta}$, and the value of TGR ranges from 0 to 1. The closer the numerical value is to 1, the smaller the gap between the meta-frontier and the cluster-frontier is. The meta-frontier represents the latent technical level of the whole evaluated individuals, and the cluster-frontier represents the actual technical level of each cluster. The smaller the TGR is, the farther the actual technical level of the cluster deviates from the potential technical frontier, which means that the technology is relatively backward (Zhu, 2013).

2.2. Inputs and Outputs

Reasonable selection of input and output variables can improve the accuracy of DEA. Several dimensions, such as environment, economy, society, resource and energy, should be comprehensively considered to fully reflect the operation of WWTPs. In general, the consumption of labor, capital, resources or energy represent inputs, while products or services are outputs (Hu et al., 2019). It is important to avoid introducing a large number of variables, because the more variables there are in the model, the more difficult it is to differentiate the DMUs (Morita and Avkiran, 2009). Therefore, it is vital to minimize the number of variables while preserving necessary production factors.

This study referred to the variables selected in the previous efficiency evaluation of wastewater treatment industry (Caldas et al., 2019; Lorenzo-Toja et al., 2015; Molinos-Senante et al., 2014; Sala-Garrido et al., 2011; Zeng et al., 2017), and took into account the difficulty of obtaining the indicator data and the applicability of indicators in the selected model. To comprehensively evaluate the efficiency of WWTPs, this study regarded the wastewater treatment as a production process, involving three inputs, three desirable outputs, and one undesirable output. Among them, capital input was expressed by operating cost (x_1, CNY/m^3), energy input was expressed by electricity consumption (x_2, kWh/m^3), and labor input was expressed by the number of labors (x_3, person/$10^4 m^3$). The desirable outputs included chemical oxygen demand (COD) removal rate (y_1, %), ammonia nitrogen (NH_3-N) removal rate (y_2, %) and reclaimed water yield (y_3, 10^4 m^3/day), while dry sludge yield (u_1, t/10^4 m^3) was chosen as the undesirable output.

2.3. Explanatory factors

The efficiency of WWTPs may also be affected by various factors except for the variables selected above. Thus, it is essential to identify the potential factors affecting the WWTP efficiency. According to previous researches (Hu et al., 2019; Longo et al., 2018; Molinos-Senante et al., 2014) and considering the availability of statistics, the following factors were supposed likely to affect efficiency scores: (i) technology used to deal with the wastewater, (ii) capacity load rate of wastewater treatment plants; (iii) influent COD concentration of the wastewater; (iv) discharge standard of pollutants and (v) geographical location of plants.

Considering that the samples may not satisfy the assumptions of normalcy and homoscedasticity, the Kruskal–Wallis test (K-W test) was adopted in this paper. The K-W test is a non-parametric test of one-way variance analysis that can be used to test the consistency hypothesis of the overall function distribution and its alternative hypotheses. As an extension of the Mann-Whitney U test, the K-W test can verify the statistical significance of differences among multiple groups (Zeng et al., 2017). If the significance is greater than 0.05 (p-value > 0.05), there is no significant difference among the tested samples. Conversely, if the p-value is less than or equal to 0.05, the samples are significantly different.

3. Results and discussion

3.1. Characteristics of sample systems

This study investigated 1,456 WWTPs under 37 indicators on the Urban

Drainage Statistics Yearbook (China Urban Water Association, 2017). 861 valid samples were screened out from more than 50,000 data in 2016, which is the latest and most complete inventory of wastewater treatment plants in China.

Chen et al. (2006) found that the correlation between wastewater treatment scale and operating cost of WWTPs was significant: the operating cost of ton water decreased with the increase in treatment scale, and became relatively stable until the treatment scale was greater than 10^5 m^3/d. Therefore, in this study, WWTPs were divided into four clusters according to the designed treatment scale: Cluster I (micro WWTPs, ≤2×10^4 m^3/d), Cluster II (small WWTPs, 2~5×10^4 m^3/d), Cluster III (medium WWTPs, 5~10×10^4 m^3/d) and Cluster IV (large WWTPs, >10×10^4 m^3/d).

Table 1 listed the descriptive statistical results of sample WWTPs. With the increase in the treatment scale, the average values of three inputs gradually decreased, while desirable output variables gradually increased. To all WWTPs, the average operating cost, electricity consumption, and labor were 0.86 CNY/m^3, 0.33 kWh/m^3, and 9.96 persons/10^4 m^3 respectively. Average removal rates of COD and NH_3-N were 88.74% and 92.47%, while the average reclaimed water and dry sludge yield was 0.47×10^4 m^3/day and 1.47 t/10^4 m^3 respectively.

Table 1 Average statistics (standard deviation in parentheses) of variables for four clusters.

	Cluster I (≤2×10^4 m^3/d)	Cluster II (2~5×10^4 m^3/d)	Cluster III (5~10×10^4 m^3/d)	Cluster IV (>10×10^4 m^3/d)	Total
Inputs					
Operating cost (CNY/m^3)	1.06(±0.73)	0.78(±0.56)	0.75(±0.42)	0.70(±0.41)	0.86(±0.61)
Electricity (kWh/m^3)	0.35(±0.27)	0.33(±0.31)	0.29(±0.98)	0.29(±0.12)	0.33(±0.25)
Labor (person/10^4m^3)	16.51(±12.26)	7.55(±3.81)	5.52(±3.16)	3.62(±2.66)	9.96(±9.33)
Desirable outputs					
COD removal rate (%)	87.29(±6.26)	89.23(±4.11)	89.59(±4.12)	90.44(±4.21)	88.74(±5.12)
NH_3-N removal rate (%)	90.88(±9.42)	93.26(±5.97)	93.23(±6.75)	93.89(±6.62)	92.47(±7.65)
Reclaimed water (10^4m^3/d)	0.09(±0.32)	0.45(±1.63)	0.74(±1.93)	1.31(±4.31)	0.47(±1.98)
Undesirable output					
Dry sludge yield (t/10^4 m^3)	1.13(±1.11)	1.70(±1.92)	1.56(±1.33)	1.70(±1.43)	1.47(±1.53)

3.2. Efficiency analysis

3.2.1. Efficiency scores

In this study, the input-oriented SBM model based on VRS and cluster benchmarking was solved by the software MaxDEA Ultra 8 (No 812-182). Detailed data can be found in Table S1 in Appendix.

The efficiency scores of the WWTPs from four clusters were compared with those based on the meta-frontier, as illustrated in Fig.2. It is obvious that meta-frontier efficiency scores were lower than those calculated by cluster-frontier, just as theoretically derived. As shown in Table 2, the average TGRs of the four clusters were between 0.578 and 0.938. With the increase in the size of the WWTPs, the average efficiency and TGR were basically increasing, indicating that the gap between the two frontiers was smaller. In other words, the larger the plant is, the more efficient it may be. Moreover, the number of efficient WWTPs increased when using cluster benchmarking. Some WWTPs were inefficient in the meta-frontier analysis but efficient in the cluster analysis. Since cluster analysis is a method to maximize the homogeneity of DMUs in the same cluster, its assessment result is considered to be closer to reality, so the efficiency scores in the following discussion are scores based on the cluster-frontier.

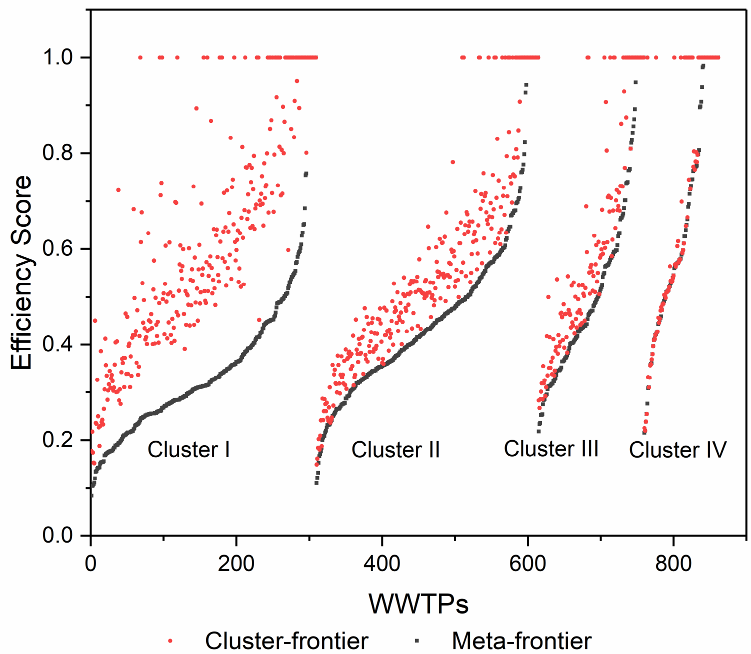


Fig. 2. Efficiency scores of 861 WWTPs based on cluster-frontier and meta-frontier respectively.
Note：Cluster I (≤ 2×10^4 m^3/d), Cluster II ($2\sim5\times10^4$ m^3/d), Cluster III ($5\sim10\times10^4$ m^3/d) and Cluster IV ($>10\times10^4$ m^3/d).

Table 2 Comparison of operating efficiency of WWTPs from different treatment scale clusters.

	Cluster I (≤2×10^4 m^3/d)	Cluster II (2~5×10^4 m^3/d)	Cluster III (5~10×10^4 m^3/d)	Cluster IV (>10×10^4 m^3/d)	Total
Number of plants	309	305	145	102	861
Number of Efficient plants	59	40	31	40	170
% Efficient plants	19.09	13.11	21.38	39.22	19.74
Average efficiency score	0.622	0.563	0.623	0.710	0.611
Technology gap ratio (TGR)	0.578	0.828	0.850	0.938	0.755

As shown in Table 2, 170 plants got a full efficiency score, meaning that 19.74% of plants were efficient and the rest were inefficient. The allocation of capital, labor, and energy in these efficient WWTPs was relatively good during the operation, and these plants were located on the best practice frontier among 861 samples. The WWTPs had an average efficiency score of 0.611. Considering that a WWTP with a score of 1 is the benchmark for best practice, the improvement potential of the sample plants is about 38.9%. As plotted in Fig. 3, 543 inefficient plants scored between 0.3 and 0.7. Therefore, inefficient WWTPs had a lot of room to improve their efficiencies.

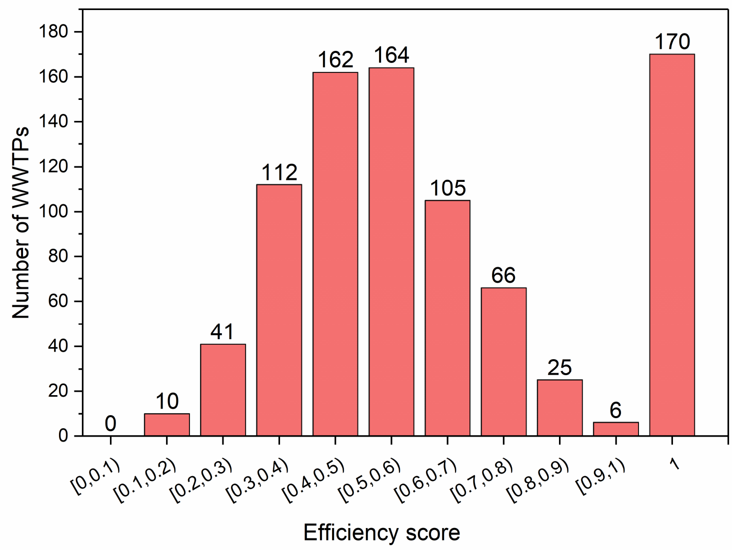


Fig. 3. Interval distribution of WWTP efficiencies.

3.2.2. Improvement potential of the WWTPs

The statistical comparison of efficient and inefficient WWTPs is shown in Fig. 4. It was found that the average values of three inputs and the undesirable output of efficient plants were significantly lower than those of inefficient plants. For instance, the average operating cost of efficient plants was 0.63 CNY/m^3, while that of inefficient plants was 0.92 CNY/m^3. In contrast, the reclaimed water yield of efficient plants was 1.41×10^4 m^3/d, far higher than that of inefficient plants, which was 0.24×10^4 m^3/d. Moreover, there was no significant difference in pollutant removal variables, for the removal rates of COD and NH_3-N could reach 90% in most WWTPs.

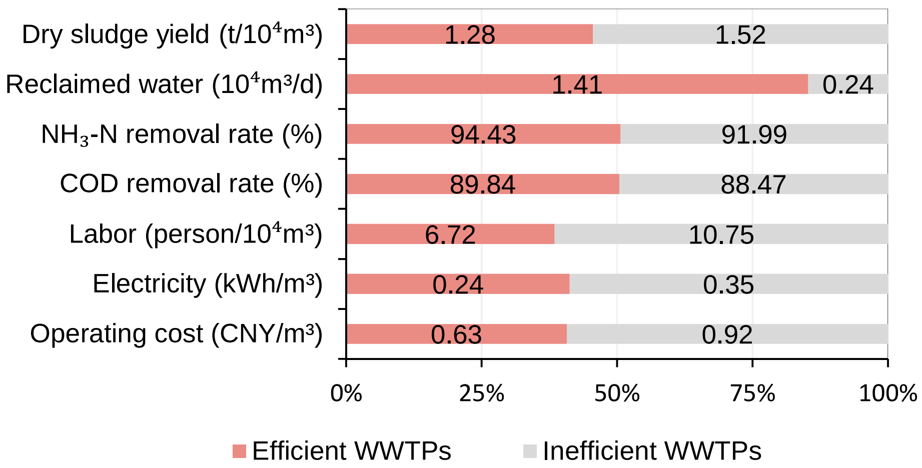


Fig.4. Comparison of seven variables in efficient and inefficient WWTPs.

In addition to providing overall efficiency scores, software MaxDEA can also provide target improvement for each plant, including all inputs and outputs. WWTPs can improve performance and become efficient by reducing input and undesirable output excesses, as well as making up for desirable output shortfalls. The results of improvement potential are shown in Table 3 and Fig. 5.

Table 3 The improvement potential of all WWTPs in this study.

	Origin	Projection	Improvement	Improvement ratio
Operating cost (10^4 CNY/d)	3776.05	1961.84	-1814.21	-48.05%
Electricity (10^4 kWh/d)	1498.70	986.51	-512.19	-34.18%
Labor (person)	28012	15832	-12179	-43.48%
COD removal rate (%)	88.74	90.67	1.93	2.17%
NH_3-N removal rate (%)	92.47	94.20	1.73	1.87%
Reclaimed water (10^4 m^3/d)	406.95	425.21	18.26	4.49%
Dry sludge yield (t/d)	8092.27	5704.07	-2388.20	-29.51%

Note：Negative values in improvement represent input excesses or undesirable output excesses that should be reduced, positive values represent the desirable output shortfalls that need to be made up for.

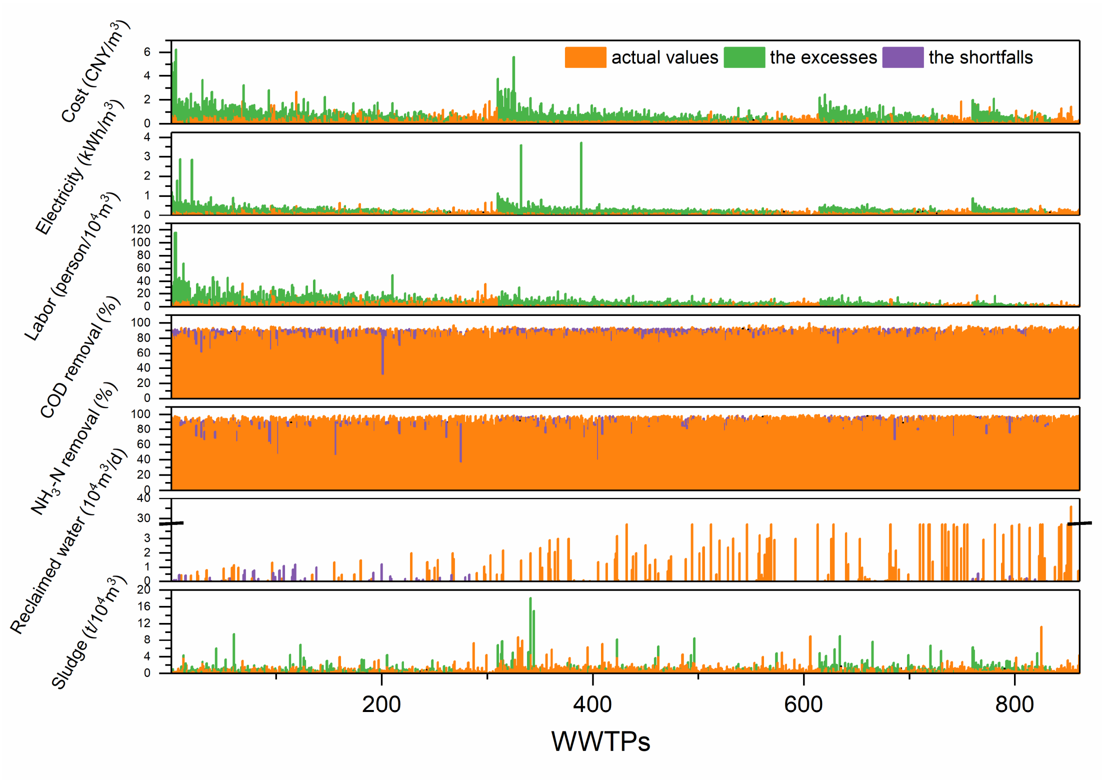


Fig. 5. The improvement potential of each WWTP.

Under the current output level, ineffective WWTPs had diverse levels of input excesses, i.e. inefficient allocations of capital, labor, and energy. For the existing 861 samples, the operating costs, electricity consumption and labor could be reduced by 1.81×10^7 CNY/d, 5.12×10^6 kWh/d, and 12,179 people respectively.

Under the current input level, the ineffective WWTPs also had some output shortages, in other words, there is room for improvement in COD removal rate, NH_3-N removal rate and reclaimed water yield. For the existing 861 samples, these three desirable output indicators could be increased by 1.93%, 1.73%, and 1.83×10^5 m^3/d, respectively. For undesirable output, dry sludge yield could be reduced by 2388.21 t/d.

Overall, according to the improvement ratio, the improvement space of input indicators is larger than that of output indicators. Through a further and more detailed analysis, it is possible to determine which items must be focused on for each plant to improve efficiency (see Table S2 in Appendix).

3.3. Explanatory factors

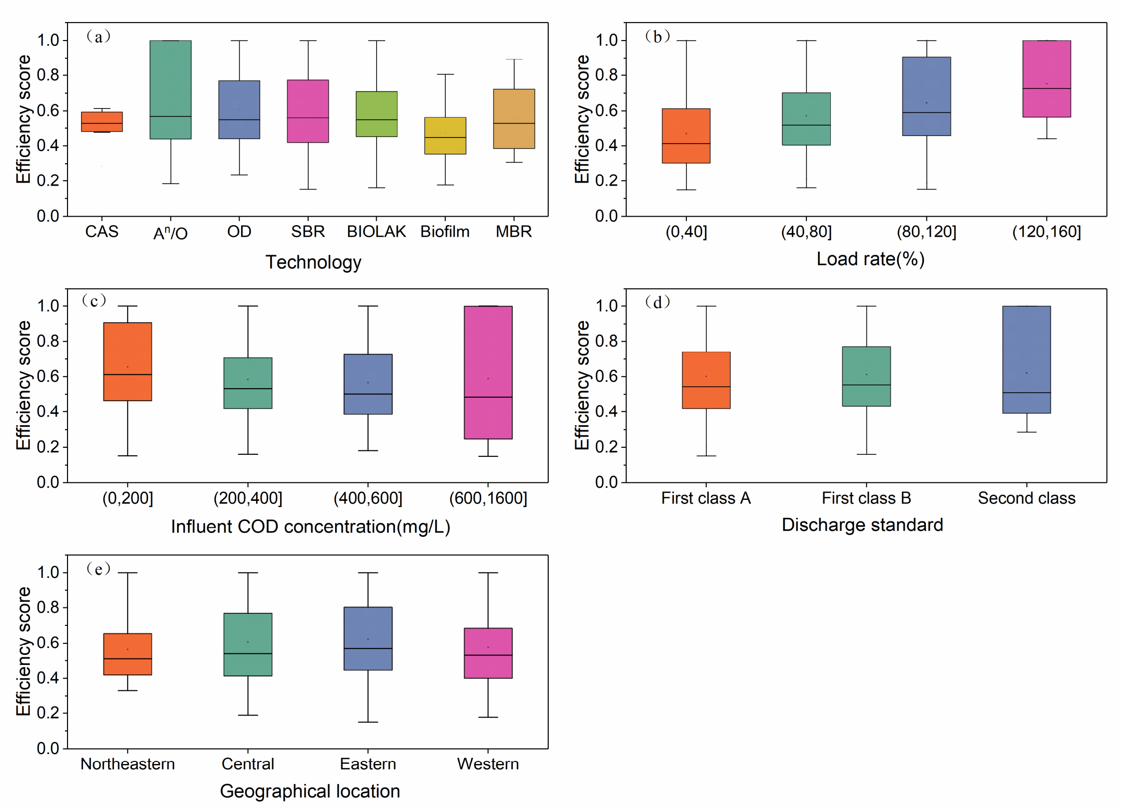


Fig. 6. Boxplots of the explanatory factors.

Efficiency scores of plants were grouped according to the explanatory factors in this study. Fig. 6 visually shows the characteristics of efficiency categorized by the five factors. To examine the differences among these groups, Kruskal-Wallis tests were conducted by SPSS 24.0. The results are listed in Table 4.

Table 4 Efficiency scores by explanatory factors and Kruskal-Wallis test results.

	Total WWTPs	Efficient WWTPs	% Eff.	Mean	Std.dev.	P-value	Chi-sq.
Technology							
CAS	8	0	0.0	0.513	0.106		
A^n/O	300	81	27.0	0.635	0.255		
OD	226	37	16.4	0.613	0.228		
SBR	164	30	18.3	0.612	0.238	0.076	11.437
BIOLAK	26	5	19.2	0.620	0.226		
Biofilm	25	0	0.0	0.470	0.158		
MBR	9	0	0.0	0.553	0.195		
Load rate							
(0,40]	38	3	7.9	0.469	0.224		
(40,80]	328	45	13.7	0.572	0.225	0.000	38.522
(80,120]	479	117	24.4	0.645	0.241		
(120,160]	16	5	31.3	0.753	0.213		
Influent COD							

concentration							
(0, 200]	345	84	24.3	0.655	0.242		
(200, 400]	440	69	15.7	0.584	0.226	0.000	20.688
(400, 600]	61	12	19.7	0.565	0.257		
(600, 1600]	15	5	33.3	0.588	0.335		
Discharge standard							
First class A	448	89	19.9	0.601	0.240		
First class B	308	57	18.5	0.612	0.235	0.769	0.525
Second class	30	8	26.7	0.620	0.276		
Geographical location							
Northeastern	44	5	11.4	0.564	0.200		
Central	136	28	20.6	0.607	0.245	0.114	5.959
Eastern	563	123	21.8	0.624	0.244		
Western	118	14	11.9	0.575	0.216		

3.3.1. Treatment technology

In China, there are many kinds of technologies applied in WWTPs, and different technologies often have diverse costs and treatment effects. Considering such critical conditions as influent characteristics and discharge standards, managers often select certain technologies for maximizing WWTP efficiency (Dong et al., 2017).

The wastewater treatment process usually falls into three stages: primary, secondary and tertiary treatment. In order to make a deeper and centralized comparison, this study only considered secondary treatment technology. According to the classification of secondary treatment provided by Li et al. (2018), the WWTPs were divided into seven categories, including CAS (conventional activated sludge process), A^n/O (anoxic/oxic process), OD (oxidation ditch), SBR (sequencing batch reactor), BIOLAK, biofilm, and MBR (membrane bio-reactor).

The boxplot of efficiency scores by seven technologies is demonstrated in Fig. 6(a). The average scores of CAS, A^n/O, OD, SBR, BIOLAK, biofilm, and MBR were 0.513, 0.635, 0.613, 0.612, 0.620, 0.470 and 0.553, respectively. Among them, the average efficiencies of A^n/O, OD and SBR were high, especially A^n/O, while the biofilm method had the lowest average efficiency, only 0.470. Therefore, the activated sludge process was more effective than the biofilm process. This conclusion is consistent with the relevant technical mechanism, as the biofilm process has poorer operational flexibility and additional investment for biofilm carrier (Hu et al. 2019; Wanner et al., 1988)

As a high-energy technology (Tolkou and Zouboulis, 2016), MBR had the highest average operating cost and electricity consumption among all technologies, 1.13 CNY/m^3 and 0.53 kWh/m^3, respectively, so its relative efficiency was low. Unlike MBR plants, WWTPs using the A^n/O process had moderate operating costs and electricity consumption (0.84 CNY/m^3, 0.32 kWh/m^3 on average), so A^n/O plants generally had higher efficiency levels. It is suggested to adopt the A^n/O process in economically backward areas.

However, the selection of specific treatment technology did not seem to have a significant impact on the performance of WWTPs. A K-W test was carried out for seven technologies and the result indicated that there was no statistical difference ($p=0.076>0.05$), as shown in Table 4. This result agreed with the findings of Hu et al. (2019) and D'Inverno et al. (2018). It is worth noting that the uneven distribution of the sample may affect the analysis result. The proportion of seven technologies in sample WWTPs is illustrated in Fig. S1 in Appendix. A^n/O, OD, and SBR were the three most common wastewater treatment technologies. 690 out of 861 plants adopted these three technologies and the daily treatment capacity reached 3.75×10^7 m^3. On the other hand, fewer than 20 plants adopted CAS and MBR processes.

3.3.2. Load rate

According to the previous analysis, the performance of a plant is tightly bound to its scale. Nevertheless, the designed capacity cannot completely represent the operating condition. Actually, operating conditions of WWTPs are not always consistent with the expected conditions, so the operating load of WWTPs may be high or low, which will affect the operating performance of WWTPs to a certain extent (Teklehaimanot et al., 2015). Therefore, the operating load rate was chosen as a potential factor affecting the efficiency of the WWTPs. Four groups have been classified according to the capacity load rate: (i) less than 40%, (ii) between 40% and 80%, (iii) between 80% and 120%, and (iv) greater than 120%.

As shown in Fig. 6(b) and Table 4, the load rate had a positive and significant influence on plant efficiency. With the increase in load rate, the efficiency of WWTPs also increased. The average efficiency scores of the four groups were 0.469, 0.572, 0.645 and 0.753, respectively. Interestingly, the efficiency continued to increase for overload plants (within the scope of assessment). It should be noted that heavy overload plants may be out of order, resulting in deterioration of wastewater quality and non-compliance with discharge standards (Longo et al., 2018). Actually, the existing literature is quite contradictory in this factor. Gomez et al. (2017) discovered that overload or underload conditions had no significant influence on WWTP efficiency. Dong et al. (2017) considered that WWTPs with excessive capacity are often inefficient. Our result is consistent with Longo et al. (2018). The higher the capacity load rate, the more efficient WWTP tends to be, even if WWTP is overloaded.

3.3.3. Influent COD concentration

The influent component is a powerful driving force affecting wastewater treatment performance. For instance, aeration demand and sludge yield have close ties with influent COD concentration (Dong et al., 2017).

According to the influent COD concentration, the WWTPs were categorized into four groups: (i) below 200 mg/L, (ii) between 200 mg/L and 400 mg/L, (iii) between 400 mg/L and 600 mg/L, and (iv) above 600 mg/L. The efficiency scores of these four groups are illustrated in Fig. 6(c). The K-W test verified that the impact of the influent COD concentration was significant ($p=0.000<0.05$).

For the four groups of WWTPs, the median efficiency was monotonically decreasing (0.612, 0.532, 0.500, 0.484), but the average efficiency was not monotonically decreasing (0.655, 0.584, 0.565, 0.588). For the first three groups, the efficiency of plants decreased gradually as the influent COD concentration increased, meaning that reducing the low influent COD concentration is helpful to improve the efficiency of plants. Hu et al. (2019) drew a similar conclusion. When the concentration was more than 600 mg/L, the distribution of efficiency score became more discrete, that is to say, the standard deviation value was higher. The result implied that high influent COD concentration helped improve the WWTP efficiency to some extent. The reason might be that high COD concentration contributed to the anaerobic reaction.

3.3.4. Discharge standard

With the tightening of the national environmental protection policy, lots of WWTPs are facing the situation of upgrading. On the basis of the Discharge Standard of Pollutants for Municipal Wastewater Treatment Plant (GB18918-2002) in China (NEPA, 2002), three main groups were divided: (i) the first class A, (ii) the first class B, and (iii) the second class. Among the 861 WWTPs, about 52% of plants met the first class A discharge standard and 36% met the first class B.

In order to achieve more comprehensive pollution control objectives, WWTPs generally need to invest more energy and chemical consumption to meet higher discharge standards (Fine and Nadas, 2012). Compared with the plants meeting the second class standard, although the pollutant removal rate of the WWTPs meeting the first class A increased, at the same time, the operating cost, labor input and electricity consumption increased by 0.06 CNY/m^3, 0.07 kWh/m^3 and 2 persons/$10^4 m^3$ respectively. As shown in Fig. 6(d), when the standards became more stringent, the efficiency score of WWTPs demonstrated a downward trend, with an average score of 0.620, 0.612 and 0.601, respectively. However, the K-W test revealed that the discharge standard had no significant effect at a 5% significance level. It is well known that higher discharge standards may bring more reuse of reclaimed water. Therefore, under our assessment framework, it is believed that the benefits of improving effluent quality can make up for the shortage of high input. If WWTPs are all upgraded to the first class A, it is estimated that the electricity consumption will increase by 1.18×10^6 kWh/d and reclaimed water yield will increase by 2.00×10^6 m^3/d.

3.3.5. Geographical location

The sample WWTPs covers 155 cities in 23 provinces (municipalities, autonomous regions) from China. It is interesting to study whether there are additional differences in geographical location. Considering the average score of all plants in a province as an inter-provincial efficiency, the spatial discrepancy of plants can be observed. As shown in Fig. 7, different colors represent different efficiency levels. Overall, inter-provincial efficiencies were mostly at a medium level. High-efficiency provinces mainly included Shaanxi (0.972), Hainan (0.832), and Guangdong (0.799). Ningxia performed the worst in all the provinces evaluated, with an average efficiency of only 0.418. The results require further consideration to avoid biased conclusions, as the sample distribution was geographically uneven and some provinces were not evaluated for the lack of data.

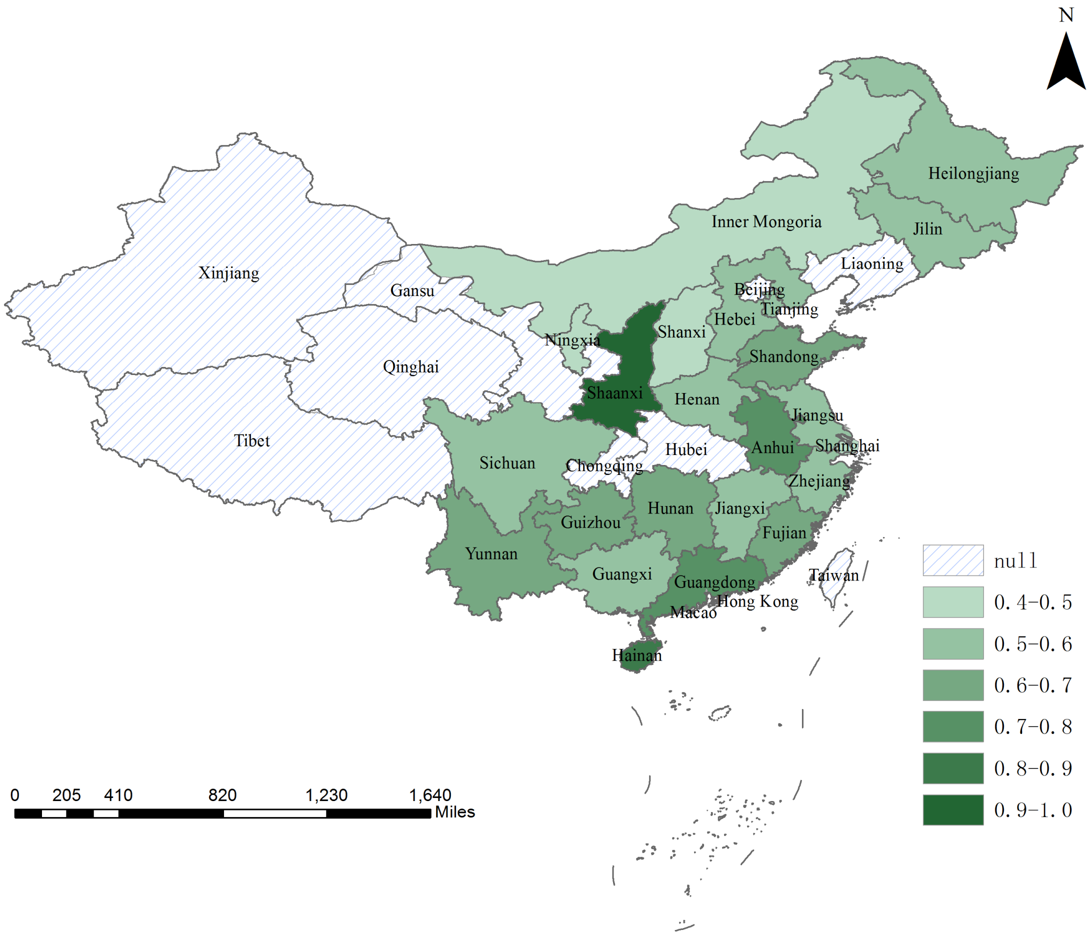


Fig. 7. Average efficiency scores in different provinces of China.

Since China is a country with a vast territory and uneven socio-economic development level (Zhang et al., 2016), in order to make the results more extensive, a K-W test was carried out on four independent samples from eastern, northeastern, central and western China. In this study, the eastern region included nine provinces (municipalities) with an average efficiency of 0.624, which was higher than the national average and ranked first in the four regions. In fact, the K-W test showed no significant difference in the distributions of efficiency scores among four regions ($p=0.114>0.05$). As the frontier of China's reform and opening up, the eastern region has great advantages in absorbing, introducing and utilizing advanced wastewater treatment technologies at home and abroad. Its management system is also in a leading position. Generally speaking, some advanced foreign technologies are often first absorbed by the eastern region, and then other regions begin to learn corresponding technologies from the eastern region. Therefore, the industry development of the eastern region basically represents the best state of China and is the benchmark of other regions.

4. Conclusions

With the requirement of sustainable development, a comprehensive and robust evaluation for WWTPs is receiving more and more attention. In this study, an SBM-DEA model based on cluster benchmarking was employed to assess 861 WWTPs in China. The evaluation index system considered multiple dimensions such as economic, environment and society. The result showed 170 plants obtained a full efficiency score. From a policy perspective, the assessment could help government agencies identify the best practices in China and set appropriate improvement targets for inefficient plants on the basis of projection values. According to the improvement ratio, there is still much room for saving in the three input indicators (operating cost, electricity consumption, and labor), especially for small plants ($\leq 5\times10^4$ m^3/d). Therefore, WWTP managers should focus on strengthening self-inspection of the plant, such as rational investment of funds and employees, and energy consumption analysis. Energy saving and consumption reduction is a comprehensive work, which requires attention to various aspects such as process, equipment, electrical and automatic control. The K-W test revealed that the influent COD concentration and capacity load rate affected the plant efficiency significantly. It is recommended to control the load rate at about 100% and the influent COD concentration below 200 mg/L. Moreover, let more wastewater be reused for urban miscellaneous water consumption or scenic environment use.

Overall, the DEA model applied to this paper can combine the efficiency of WWTPs with sustainability issues, and can be used as a benchmark model to provide recommendations for improving plants, thereby achieving the best utilization of existing resources. There is no doubt that the methodology and applications in this study are useful for government departments and enterprise managers. The evaluation system can help the government understand the relative operation of WWTPs and strengthen the supervision of third-party enterprises on the operation and management of WWTPs. It can also help enterprises to understand their own shortcomings and improvement potential. It is feasible to conduct incentive control on WWTPs, with rewards and punishments based on periodic evaluations or rankings of operational efficiency.

In addition, it is worth noting that DEA can effectively assess the relative efficiency, but cannot reflect the absolute efficiency. Due to the data uncertainty and quantity limitation of indices, further study can conduct sensitivity analysis and extend the index set, such as the use of chemical reagents and potential environmental impacts. Life cycle assessment (LCA) is a new tool for environmental management to quantitatively evaluate environmental impacts. Therefore, the combination of LCA and DEA may achieve a more accurate and comprehensive evaluation.

Acknowledgments

This work was supported by Major Science and Technology Program of China for Water Pollution Control and Treatment (2017ZX07204001) and Participation in Research Program (T160PRP35015) from Shanghai Jiao Tong University. The authors thank Grammarly (https://app.grammarly.com) for providing language help during the preparation of this paper.

Appendix A. Supplementary data

Supplementary data related to this article can be found at

References

China Urban Water Association, 2017. Urban Drainage Statistics Yearbook 2017. China Urban Water Association, Beijing (in Chineses).

Bi, G.B., Song, W., Wu, J., 2014. A clustering method for evaluating the environmental performance based on slacks-based measure. Comput Ind Eng. 72, 169-177.

Buonocore, E., Mellino, S., De Angelis, G., Liu, G.Y., Ulgiati, S., 2018. Life cycle assessment indicators of urban wastewater and sewage sludge treatment. Ecol Indic. 94, 13-23.

Caldas, P., Ferreira, D., Dollery, B., Marques, R., 2019. Are there scale economies in urban waste and wastewater municipal services? A non-radial input-oriented model applied to the Portuguese local government. J Clean Prod. 219, 531-539.

Castellet, L., Molinos-Senante, M., 2016. Efficiency assessment of wastewater treatment plants: A data envelopment analysis approach integrating technical, economic, and environmental issues. J Environ Manage. 167, 160-166.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. Eur J Oper Res. 2(6), 429-444.

Chen, H.B., Yu, F., Sun, B.Y., He, Q.B., Ming, L.L., 2006. Study on optimized scale of centralized municipal wastewater treatment system. China Water & Wastewater. (21), 26-30 (in Chinese).

Corton, M.L., Berg, S.V., 2009. Benchmarking Central American water utilities. Util Policy. 17(3-4), 267-275.

D'Inverno, G., Carosi, L., Romano, G., Guerrini, A., 2018. Water pollution in wastewater treatment plants: An efficiency analysis with undesirable output. Eur J Oper Res. 269(1), 24-34.

Dong, X., Zhang, X.Y., Zeng, S.Y., 2017. Measuring and explaining eco-efficiencies of wastewater treatment plants in China: An uncertainty analysis perspective. Water Res. 112, 195-207.

Fare, R., Grosskopf, S., 2004. Modeling undesirable factors in efficiency evaluation: Comment. Eur J Oper Res. 157(1), 242-245.

Fine, P., Nadas, E., 2012. Options to reduce greenhouse gas emissions during wastewater treatment for agricultural use. Sci Total Environ. 416, 289-299.

Gao, Q., Zhang, H.G., An, Z., Zhu, B., 2006. Research on the validity of scale and technology of urumqi's sewage disposal factories based on DEA method. Journal of Xinjiang University (Natural Science Edition). (02), 211-215 (in Chinese).

Gomez, T., Gemar, G., Molinos-Senante, M., Sala-Garrido, R., Caballero, R., 2017. Assessing the efficiency of wastewater treatment plants: A double bootstrap approach. J Clean Prod. 164, 315-324.

Hernandez-Chover, V., Bellver-Domingo, A., Hernandez-Sancho, F., 2018. Efficiency of wastewater treatment facilities: The influence of scale economies. J Environ Manage. 228, 77-84.

Hernandez-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), 2693-2699.

Hospido, A., Moreira, M.T., Fernandez-Couto, M., Feijoo, G., 2004. Environmental performance of a municipal wastewater treatment plant. Int J Life Cycle Ass. 9(4), 261-271.

Hu, W.Q., Guo, Y., Tian, J.P., Chen, L.J., 2019. Eco-efficiency of centralized wastewater treatment plants in industrial parks: A slack-based data envelopment analysis. Resour Conserv Recy. 141, 176-186.

Li, Z., Zhao, L.J., Zhu, H.F., Song, X.C., Wang, J., 2018. Analysis of construction and operation status and existing problems of municipal wastewater treatment plants in China. Water & Wastewater Engineering. 54(04), 52-57 (in Chinese).

Longo, S., Hospido, A., Lema, J.M., Mauricio-Iglesias, M., 2018. A systematic methodology for the robust quantification of energy efficiency at wastewater treatment plants featuring Data Envelopment Analysis. Water Res. 141, 317-328.

Lorenzo-Toja, Y., Vazquez-Rowe, I., Chenel, S., Marin-Navarro, D., Moreira, M.T., Feijoo, G., 2015. Eco-efficiency analysis of Spanish WWTPs using the LCA plus DEA method. Water Research. 68, 651-666.

Lozano-Vivas, A., Pastor, J.T., Pastor, J.M., 2002. An efficiency comparison of European banking systems operating under different environmental conditions. J. Prod. Anal. 18(1), 59-77.

Molinos-Senante, M., Hernandez-Sancho, F., Sala-Garrido, R., 2014. Benchmarking in wastewater treatment plants: a tool to save operational costs. Clean Technol Envir. 16(1), 149-161.

Morita, H., Avkiran, N.K., 2009. Selecting Inputs and Outputs in Data Envelopment Analysis by Designing Statistical Experiments. J Oper Res Soc Jpn. 52(2), 163-173.

NDRC, 2016. "13th Five-Year Plan" for Urban Wastewater Treatment and Recycling Facilities Construction. National Development and Reform Commission, Beijing (in Chinese). (accessed at 31 December 2016). http://www.ndrc.gov.cn/zcfb/zcfbtz/201701/t20170122_836015.html.

NEPA, 2002. Discharge Standard of Pollutants for Municipal Wastewater Treatment Plant. China's National Environmental Protection Agency, Beijing (in Chinese).

Patra, B.K., Nandi, S., Viswanath, P., 2011. A distance based clustering method for arbitrary shaped clusters in large datasets. Pattern Recogn. 44(12), 2862-2870.

Rao, D., O'donnell, C.J., Battese, G.E., 2003. Metafrontier functions for the study of inter-regional productivity differences. School of Economics, University of Queensland, Australil.

Reinhard, S., Lovell, C.A.K., Thijssen, G.J., 2000. Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA. Eur J Oper Res. 121(2), 287-303.

Sala-Garrido, R., Molinos-Senante, M., Hernández-Sancho, F., 2011. Comparing the efficiency of wastewater treatment technologies through a DEA metafrontier model. Chem Eng J. 173(3), 766-772.

Scheel, H., 2001. Undesirable outputs in efficiency valuations. Eur J Oper Res. 132(2), 400-410.

Seiford, L.M., Zhu, J., 2002. Modeling undesirable factors in efficiency evaluation. Eur J Oper Res. 142(1), 16-20.

Teklehaimanot, G.Z., Kamika, I., Coetzee, M.A.A., Momba, M.N.B., 2015. Population Growth and Its Impact on the Design Capacity and Performance of the Wastewater Treatment Plants in Sedibeng and Soshanguve, South Africa. Environ Manage. 56(4), 984-997.

Tolkou, A.K., Zouboulis, A.I., 2016. Effect of climate change in WWTPs with a focus on MBR infrastructure. Desalin Water Treat. 57(5), 2344-2354.

Tone, K., 2001. A slacks-based measure of efficiency in data envelopment analysis. Eur J Oper Res. 130(3), 498-509.

Wang, X., Liu, J.X., Ren, N.Q., Duan, Z.S., 2012. Environmental profile of typical anaerobic/anoxic/oxic wastewater treatment systems meeting increasingly stringent treatment standards from a life cycle perspective. Bioresource Technol. 126, 31-40.

Wanner, J., Kucman, K., Grau, P., 1988. Activated sludge process combined with biofilm cultivation. Water research (Oxford). 22(2), 207-215.

Xiao, C., Wang, Z., Shi, W., Deng, L., Wei, L., Wang, Y., Peng, S., 2018. Sectoral energy-environmental efficiency and its influencing factors in China: Based on S-U-SBM model and panel regression model. J Clean Prod. 182, 545-552.

Yang, T., Chen, W., Zhou, K.L., Ren, M.L., 2018. Regional energy efficiency evaluation in China: A super efficiency slack-based measure model with undesirable outputs. J Clean Prod. 198, 859-866.

Zeng, S.Y., Chen, X., Dong, X., Liu, Y., 2017. Efficiency assessment of urban wastewater treatment plants in China: Considering greenhouse gas emissions. Resour Conserv Recy. 120, 157-165.

Zhang, J.R., Zeng, W.H., Shi, H., 2016. Regional environmental efficiency in China: Analysis based on a regional slack-based measure with environmental undesirable outputs. Ecol Indic. 71, 218-228.

Zhou, J., Chang, V.W.C., Fane, A.G., 2011. Environmental life cycle assessment of reverse osmosis desalination: The influence of different life cycle impact assessment methods on the characterization results. Desalination. 283, 227-236.

Zhou, P., Ang, B.W., Poh, K.L., 2008. A survey of data envelopment analysis in energy and environmental studies. Eur J Oper Res. 189(1), 1-18.

Zhu, D.J., 2013. The performance of regional carbon emissions in China: an analysis based on the perspective of the technology gap. Shandong University, China (in Chinese).

Highlights

- 861 WWTPs were assessed by an SBM-DEA model based on cluster benchmarking.
- The evaluation index was extended to economic, environmental and social domains.
- 170 plants were regarded as best practices over the latest inventory in China.
- The improvement potential for sample plants was about 38.9%.
- Potential factors affecting the performance efficiency of WWTPs were discussed.