Performance evaluation of problematic samples: a robust nonparametric approach for wastewater treatment plants

Alda A. Henriques[a], Milton Fontes[b], Ana S. Camanho[a1] , Giovanna D'Inverno[c], Pedro Amorim[a], Jaime Gabriel Silva[d],

[a] INESC TEC, Faculdade de Engenharia, Universidade do Porto, Portugal
[b] Águas do Centro Litoral (AdP Group)
[c] Faculty of Economics and Business, KU Leuven
[d] AdP Energias and Instituto Superior de Engenharia do Porto

Abstract: This paper explores robust unconditional and conditional nonparametric approaches to support performance evaluation in problematic samples. Real-world assessments often face critical problems regarding available data, as samples may be relatively small, with high variability in the magnitude of the observed indicators and contextual conditions. This paper explores the possibility of mitigating the impact of potential outlier observations and variability in small samples using a robust nonparametric approach. This approach has the advantage of avoiding unnecessary loss of relevant information, retaining all the decision-making units of the original sample. We devote particular attention to identifying peers and targets in the robust nonparametric approach to guide improvements for underperforming units. The results are compared with a traditional deterministic approach to highlight the proposed method's benefits for problematic samples. This framework's applicability in internal benchmarking studies is illustrated with a case study within the wastewater treatment industry in Portugal.

Keywords: Data Envelopment Analysis, Robust and Conditional Approach, Problematic Samples, Wastewater Treatment Plants

1. Introduction

In the context of nonparametric efficiency analysis, such as Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH), a set of individual entities is comparatively assessed to search for potential improvements in their performance. Each entity, or decision-making unit (DMU), conducts a given productive activity that can be seen as converting inputs into outputs. To achieve a fair efficiency assessment, the sample of DMUs used in the analysis must be homogeneous or, in other words, similar in several ways (Dyson et al., 2001). First, all DMUs must develop similar activities to obtain similar products or services (Pitfall 3.1, *ibidem*). However, although all DMUs may be converting the same type of inputs into the same type of outputs, in practice two additional issues can threaten the homogeneity assumption, namely the presence of atypical observations in the sample and different environmental conditions that the DMUs can face (Pitfall 3.2, *ibidem*). This issue often affects real-world samples, where high variability in input/output levels or exogenous conditions can be found, hence the name "problematic samples". The performance evaluation conducted on this type of samples might be strongly affected, resulting in biased efficiency measures. As an alternative to not conducting any efficiency analysis at all because some requirements are not perfectly met, or to implement a conventional framework with deficient results, the selection of a more refined method

[1] Corresponding author.
E-mail address: acamanho@fe.up.pt (A.S. Camanho). Faculdade de Engenharia da Universidade do Porto. Rua Dr. Roberto Frias, 4200-465 Porto, Portugal.

may allow to open the black box of the transformation process under analysis and to obtain more realistic results, extracting valuable information concerning the units under assessment.

This paper aims to show, from a novel perspective, how nonparametric efficiency evaluation tools can provide valuable information to assist companies or decision-makers in their management choices when evaluating DMUs in problematic samples. A problematic sample is defined as a sample with a small number of DMUs, and high variability in the indicators' values and/or environmental conditions. In relation to the first aspect, the meaning of "small" in the definition of problematic samples should be interpreted in relative terms by comparing the number of DMUs and the inputs, outputs, and exogenous factors considered in the performance assessment. The curse of dimensionality in nonparametric methods is a well-known problem in the literature (e.g., Lee and Cai, 2020), but no general guidelines are available to overcome this issue in real-world benchmarking studies.

Regarding the variability in the magnitude of the indicators' values, part of it may result from the presence of atypical observations. In this context, a DMU may differ to a large extent from the rest of the sample mainly for two reasons: there may be errors in the data, or the observations may be potentially correct but highly atypical. In a benchmarking setting, the DMUs located on the frontier should correspond to replicable DMU-level performance for the same set of circumstances, so that DMUs with exceptionally high relative performance may be removed from the sample for precautionary reasons.

The efficiency literature proposes several outlier detection procedures to guide the removal of suspected units, attempting to ensure that the frontier is well populated and not too distant from the bulk of DMUs in the analysis (e.g., Anderson and Petersen, 1993; Wilson, 1993; Wilson, 1995; Simar, 2003; De Witte and Marques, 2010b). However, for analyses involving DMUs in a problematic sample, especially due to the limited number of DMUs available, removing potential outliers cannot be considered a solution. Using robust methods (for example, bootstrapping techniques as proposed by Simar and Wilson, 1998, or robust optimization approaches as proposed by Sadjadi and Omrani, 2008, Omrani at al, 2021, Omrani at al., 2022) or partial frontiers instead of full frontier are very appealing alternatives for these cases (see Cazals et al, 2002; Aragon et al, 2005).

Furthermore, another key challenge in the performance measurement of DMUs in a problematic sample is accounting for different operating contexts. One-stage and two-stage approaches have been considered to address the impact of contextual conditions on DMUs' efficiency levels (see Avkiran and Rowlands, 2008 for a literature review). However, there is an unsettled debate in the literature regarding the most appropriate specification of these types of approaches. To overcome some of the limitations of the two-stage approaches, a conditional approach (Daraio and Simar, 2005 & 2007) has been proposed to circumvent the restrictive assumptions underlying efficiency assessments with heterogeneous environmental conditions (see De Witte and Marques, 2010a, for a critical discussion of different methodologies to incorporate heterogeneity).

In light of this, we found that the nonparametric order-*m* method (Cazals et al., 2002) fits all the requirements to attain the objectives of this paper. Firstly, alternatively to a full frontier method such as DEA, the partial frontier method is demonstrated to surmount the curse of dimensionality problem (Wheelock and Wilson, 2003), which is particularly useful when dealing with small samples. Additionally, the order-*m* method also allows to mitigate the effect of potential outliers, either occurring due to errors in the data or due to being real atypical observations, without the need to remove any DMU, which again is particularly useful when dealing with small samples. Moreover, regarding the variability that occurs due to DMUs facing different contextual conditions, a conditional order-*m* method (Daraio and Simar, 2005 & 2007) may provide additional insights into the role of those contextual conditions on the performance of the DMUs.

In this paper, we identify a suitable framework for internal benchmarking exercises and thus support real-world efficiency assessments. Specifically, we propose using the outlined robust order-*m* and conditional nonparametric estimation tools departing from their traditional applications. These approaches can be considered a starting point for extracting relevant information from internal benchmarking exercises, providing ad hoc suggestions, and ready-to-implement measures. The ultimate objective is to provide information to guide practical improvements, particularly in the case where the DMUs to be assessed are in a problematic sample, or, in other words, for samples that at first glance might discourage or invalidate the performance evaluation.

The suggested framework has a clear managerial application. It focuses on the evaluation of problematic samples and provides a clear identification of potential reductions in resource consumption or reallocation of resources, rather than on a more general analysis of relative efficiency levels that would be suitable for policy design at more aggregated levels. By applying this toolbox, we illustrate the relevance of identifying peers and targets for decision support, complementing the standard results provided by robust methods. We also exemplify how informative is the evidence provided by the conditional approach to guide more thoroughly comparative assessments. As a result, in real-world case studies, intervening at a very detailed level can improve the effectiveness of the performance assessment exercise, making processes evolve towards more sustainable and productive systems at social, economic, and environmental levels (Moldan et al., 2012).

Although the proposed efficiency analysis framework can be applied to several areas, we explore a particular case study within the wastewater treatment industry. The efficient use of resources in the productive activity of cleansing the wastewater is a topic that deserves increased attention by the scientific community and practitioners. As highlighted by Longo et al. (2018), the wastewater industry, and more specifically the case of wastewater treatment plants (WWTPs), constitutes an example of an application field where the homogeneity assumptions may be critical, although often disregarded in the literature. Our case study corresponds to a sample of 41 Portuguese WWTPs managed by the company *Águas do Centro Litoral* (AdCL). This sample presents remarkable variability in the input and output indicators and faces different operating characteristics. Therefore, it represents a good testbed to provide concrete examples of the advocated advantages of using robust and conditional efficiency method tools for performance measurement of DMUs in a problematic sample.

Two aspects are simultaneously considered to assist decision-making better: (1) The measurement of efficiency should be as fair as possible to provide an accurate picture of the relative performance behaviour of the sample under evaluation. When this aspect is taken into account, better prioritization of actions and eventual investments can be achieved. (2) The information on peers and targets for improvements should be provided. This allows plants to benefit from information sharing to guide performance improvements through adjustments of technical and operational aspects. The use of DEA to guide organizational improvement by identifying attainable targets, representing best practices that are aligned with management strategy, is a topic that is gaining momentum in the literature in recent years (e.g., see Ruiz and Sirvent, 2019).

DEA has also been widely used as the preferred methodology to assess the performance of wastewater treatment plants (WWTPs) (e.g., Hernández-Sancho and Sala-Garrido, 2009). Moreover, most studies within the context of the wastewater treatment industry recognize the relevance of the environmental conditions faced by WWTPs on their performance (e.g., Fuentes et al., 2015). Accordingly, for the practical implementation, we adopt a robust DEA estimator based on the concept of order-*m* frontiers introduced by Cazals et al. (2002) and further extended by Daraio and Simar (2007), together with its conditional version. Conducting a fair comparison of WWTPs and reporting their peers helps to look inside the black box of the transformation process. This directly allows the interested stakeholders, namely company and plants decision-makers, to receive tangible management solutions that can be practically implemented. Moreover, ensuring improvements in asset management is likely to have an impact beyond the direct benefits arising from organizing the WWTPs production process more efficiently and consequently saving resources at the company level.

The wastewater treatment industry plays a significant role in sustainable development worldwide. Improved functioning of the WWTPs would also support at least two goals of the Sustainable Development Agenda defined by the United Nations (United Nations, 2015): Goal 6 ("*By 2030, improve water quality by reducing pollution, eliminating dumping and minimizing release of hazardous chemicals and materials, halving the proportion of untreated wastewater and substantially increasing recycling and safe reuse globally*", 6.3) and Goal 7 ("*By 2030, double the global rate of improvement in energy efficiency*", 7.3).

The contribution of this paper to the Operations Research and Management Science literature is twofold. First, it contributes to the nonparametric efficiency literature mostly oriented to internal managerial practices (Camanho and Dyson, 1999; Barros, 2006 & 2008; Camanho et al., 2009; Vaz et al., 2010; Roháčová, 2015) by adapting the use of robust and conditional approaches from an innovative and operative point of view. A critical feature in internal benchmarking studies concerns the improvement of operational practices through learning from peers. In a recent paper by Lavigne et al. (2019), the authors emphasize the importance of identifying the peers underlying a robust conditional Benefit-of-the-doubt approach applied to waste management. Our paper extends this research line in two main directions. First, Lavigne et al. (2019) focused on constructing a composite indicator to aggregate outputs, while we look at the transformation of resources into outputs in efficiency assessments involving problematic samples. Furthermore, rather than just focusing on estimating peers, we also estimate the targets for the inefficient DMUs to provide guidelines for the amount of resources that can be potentially saved. We also critically discuss the different information that can be retrieved from robust and deterministic methods when assessing the performance of DMUs in a problematic sample. In addition to this, we point out some technical issues that must be considered during the implementation of order-*m* and conditional approaches to the case of such type of samples.

The second contribution of this paper is related to its empirical application, specifically to the literature on WWTPs efficiency evaluation. Despite the increasing prominence of this topic in recent years (see D'Inverno et al., 2018), to the best of our knowledge, only a few papers used a conditional approach to evaluate wastewater treatment plants efficiency (e.g., Fuentes et al., 2015; Guerrini et al., 2016). Furthermore, our study represents a development in relation to these approaches by proposing a novel WWTP output specification to fully capture the real output attained by each plant during the treatment process. This requires considering the rate of pollutant removal (which accounts for the concentration of the pollutant in the influent) and the total volume of wastewater treated. It is also the first study adopting an internal managerial perspective of WWTPs efficiency using Portuguese data.

The remainder of the paper is organized as follows. Section 2 describes the methodological toolbox proposed for the efficiency analysis with problematic samples. Section 3 describes the case study used to demonstrate the usefulness of the framework proposed. Section 4 presents the results obtained and discusses their managerial implications. Finally, section 5 concludes, highlighting the major findings and limitations of this study.

2. The methodology

This section introduces the nonparametric efficiency tools commonly used in performance evaluations, which are adapted in the empirical application to deal with the challenges of problematic samples (i.e., small samples with high variability in input and output levels and/or exogenous conditions). Particular attention is devoted to the practical information on peers and targets that can be retrieved from their use.

One of the main advantages of nonparametric efficiency methods is that the efficient frontier is derived directly from observations. Therefore, there is no need for prior specification of its functional form. Depending on the assumptions characterizing the production technology, different nonparametric efficiency models can be considered, such as Data Envelopment Analysis, Free Disposal Hull, and Directional Distance Function.

To keep a concise exposition, we focus the discussion on Data Envelopment Analysis (DEA) models, with an input orientation setting under Variable Returns to Scale (VRS). However, the approach can be straightforwardly extended to other nonparametric techniques, such as the ones mentioned above, as well as to different orientations or returns to scale assumptions.

2.1 DEA: A baseline nonparametric efficiency model

Data Envelopment Analysis (DEA) is a non-parametric method that originated from the seminal work of Farrell (1957), who proposed the evaluation of relative efficiency as the radial distance to a production frontier estimated directly from empirical observations. Farrell (1957) concepts were operationalized for the first time using linear programming by Charnes et al. (1978).

DEA compares the efficiency of a relatively homogeneous set of DMUs in using multiple resources (inputs) to produce multiple outcomes (outputs). It derives a single summary measure of efficiency for each unit, which is based on a comparison with other units in the sample. DEA identifies two groups of units: radially efficient and inefficient. The DMUs on the frontier are considered examples of best practices (i.e., the benchmarks) and obtain an efficiency score equal to one. The efficient DMUs of the sample span the frontier. For the inefficient DMUs, the magnitude of their inefficiency is evaluated as the distance to the frontier. For an input-oriented assessment, it represents the proportional reduction of inputs that is required to reach the frontier. One of the advantages of DEA is to allow each DMU to select its own weighting system for the performance evaluation, recurring to optimization. This allows emphasizing the strengths of each DMU.

The formulation of the input-oriented DEA model with Variable Returns to Scale is shown in (1).

$$\min_{\theta_k, \lambda_j, s_i^-, s_r^+} \theta_k - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \qquad \textbf{(1)}$$

Subject to

$$\theta_k\, x_{ik} - \sum_{j=1}^{n} \lambda_j x_{ij} - s_i^- = 0 \qquad i = 1, \dots, m$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{rk} \qquad r = 1, \dots, s$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j, s_i^-, s_r^+ \geq 0, \qquad \forall j, i, r$$

In model (1), x_{ij} corresponds to the value of input i $(i = 1, \dots, m)$ and y_{rj} corresponds to the value of output r $(r = 1, \dots, s)$ observed for DMU j $(j = 1, \dots, n)$. The decision variables of the model (1) are θ_k , λ_j , s_i^- and s_r^+. θ_k corresponds to the radial efficiency score of the unit under assessment, represented by index k. λ_j is an intensity variable, which assumes a positive value whenever a unit j is used as peer for unit k under assessment. The VRS assumption is represented by the restriction imposing that the sum of λ_j, for $(j = 1, \dots, n)$, is equal to one. s_i^- and s_r^+ are slack variables, representing potential non-radial adjustments to the input and output levels of the unit under assessment, beyond the proportional reduction to the input levels represented by the radial efficiency score. ε is an infinitesimal used to ensure that slacks are not ignored in the efficiency assessment.

When analyzing the solution to model (1), a DMU k under evaluation is fully efficient if and only if θ_k =1 and all slacks are equal to zero ($s_{ik}^- = s_{rk}^+ = 0, \forall i, r$). In other words, a unit is fully efficient if it is not possible to decrease the consumption of any of its inputs without also decreasing the amount of at least one of the outputs or increasing the consumption of another input. As by-products of the assessment, for inefficient DMUs it is also possible to obtain peers and targets. The input and output targets for unit k are obtained as shown in expressions (2) and (3). The values of the decision variables obtained at the optimal solution to model (1) are signalled with an asterisk.

$$x_{ik}^T = \sum_{j=1}^{n} \lambda_{jk}^* x_{ij} = \theta_k{}^* x_{ik} - s_{ik}^{*-} \qquad i = 1, \dots, m \qquad \textbf{(2)}$$

$$y_{rk}^T = \sum_{j=1}^{n} \lambda_{jk}^* y_{rj} = y_{rk} + s_{rk}^{*+} \qquad r = 1, \dots, s \qquad \textbf{(3)}$$

2.2 Robust DEA: Mitigating the impact of atypical observations

The non-parametric nature of DEA is an important feature in situations where the production process is complex and the production function is unknown. However, its deterministic nature makes it very sensitive to atypical observations, as any unit can potentially contribute to shape the frontier, hence the name "full frontier". As a result, the presence of outliers and extreme values might lead to downwardly biased estimates (Fusco et al., 2020), as DMUs that differ to a large extent from the rest of the sample may lead to the shift of the frontier and have a significant impact on the evaluation of other DMUs.

Several methods can be employed to detect outliers that can potentially affect the position of the frontier, which are candidates to be removed from the sample (e.g., super-efficiency of Anderson and Petersen, 1993 or the approach proposed by Wilson, 1995). From an applied perspective of internal benchmarking, the outlying units might be of great interest to the stakeholders to the extent that they could potentially be the best or the worst-performing units. Especially in the context of small samples, each piece of information is crucial for a sound managerial decision, and the removal of observations should be avoided (De Witte and Marques, 2010b). To account for these issues, and at the same time to mitigate the impact of atypical observations, the literature has proposed the use of "partial" or "robust" frontiers, order-m (Cazals et al., 2002; Daraio and Simar, 2005 & 2007) or order-α methods (Aragon et al., 2005; Daouia and Simar, 2007). In compliance with this latter stream of literature, this paper adopts insights from the order-m method to evaluate performance. To keep the discussion focused on an operational perspective, we only present the basic idea and the main intuition of this robust approach. For an exhaustive exposition, we refer to Daraio and Simar (2007).

Computationally, in the full frontier estimation, the linear programming model (1) is solved once for each unit under assessment. Instead, in the partial frontier estimation the linear programming model (1) is computed B times for each unit according to a Monte-Carlo procedure, where B is a large number. For an input-oriented assessment, in each of these B iterations, m units are drawn at random with replacement among those producing at least the same level of output as the unit under evaluation. By doing this, the impact of the atypical observations is remarkably mitigated, and the unit under evaluation is compared with a less extreme benchmark. The robust efficiency score $\hat{\theta}_k^m$ is then obtained as the average of the efficiency scores $\theta_k^{b,m}$ computed at each b-th iteration ($b = 1, \dots, B$):

$$\hat{\theta}_k^m = \frac{1}{B}\sum_{b=1}^{B}\theta_k^{b,m} \tag{4}$$

Due to the subsampling, the evaluated unit might not be among the drawn m units. In this case, it does not constitute its own reference set, and it may be located above the frontier. An efficiency score greater than one identifies a 'super-efficient unit' that is more efficient than the average m units in its reference set. m can have a twofold interpretation (Daraio and Simar, 2005; De Witte and Schiltz, 2018). First, it can represent the number of competitors producing greater or the same output levels. Second, it can be seen as a trimming value to assert the robustness of the analysis. Among others, Cazals et al. (2002) and Rogge and De Jaeger (2013) note that the choice of m is not straightforward. m should not be set too low or too high, since "the obtained estimator will converge to the DEA frontier estimator if $m \to \infty$ but, for finite m, it will not envelop all the data points" (Cazals et al, 2002). Usually in practical applications $m < n$ and it should be set so to have a "sufficiently small decrease in the proportion of super-efficient observations" (Schiltz et al., 2020). A sensitivity analysis for different values of m helps to support the robustness of the findings.

The identification of peers and targets in evaluations with "partial" frontiers is also possible. When the focus is on small samples representing real-world settings, these outcomes can be easily investigated and practically explored. For example, the number of times a unit is considered a peer can be computed while keeping in mind that units are now drawn from a subset of the sample. Furthermore, the intensity values (λ_j^*) can be computed as an average value. Likewise, the target values can be obtained as the average target value across the B iterations, pointing at the average resource savings, once the units are evaluated against a "partial" frontier that envelops the data more tightly (representing a less demanding reference of best-practice). As a result, the robust efficiency score will always be higher than the one obtained from the full frontier estimation, and consequently the robust target values will be lower and more realistic to be achieved.

2.3 Conditional robust DEA: The role of contextual variables

Besides softening the impact of outlying observations in a problematic sample, the role of the operating context needs to be accounted for when it comes to giving operational guidelines to the units under assessment. This aspect enters both in the computation of the efficiency scores and in detecting the influence of these external factors on the production process itself. The literature on efficiency has acknowledged the relevance of these aspects and has developed mainly two approaches to address them, namely the 'one-stage' approach and the 'two-stage' approach (Daraio and Simar, 2007). In the former, environmental variables are directly included in the model estimation, but they first need to be classified a priori for the analysis either as an input or as an output. In the latter, the efficiency scores are parametrically regressed in a second stage on nondiscretionary variables. However, a few issues might arise. First, the efficiency scores are serially correlated and the first stage efficiency scores are biased. Second, even if these problems would be addressed by bootstrap techniques, this approach would need a priori specification of the parametric regression model and it would imply the "separability condition". Estimating first the efficiency scores and then regressing them on the environmental variables would assume that these variables do not influence the attainable set. In many practical applications, this assumption is difficult to defend (for a thorough explanation, we refer to Daraio and Simar (2007) and the references therein).

To circumvent the above-mentioned issues, conditional nonparametric frontier models have been introduced by Daraio and Simar (2005). Following insights from Daraio and Simar (2005, 2007) and De Witte and Kortelainen (2013), this paper integrates the robust DEA presented in the previous section with its conditional version.

Computationally, the conditional robust frontier estimation requires a slight adjustment concerning the (unconditional) frontier estimation as described above. The adjustment concerns the way m units are drawn among those producing at least the same output level as the evaluated unit (Rogge and De Jaeger, 2013). In the unconditional case, m units are drawn with replacement and with uniform probability, so that each unit is equally likely to be drawn. In the conditional case instead, m units are drawn with replacement but with a weight (probability) determined by an estimated kernel function. Accordingly, units that operate in a more similar environment will have a higher probability of being drawn and included in the reference set. Likewise, units that operate in a more dissimilar environment will have a lower probability of being drawn and considered in the reference set. In this sense, the efficiency score is 'conditional' upon the environmental variables z. The conditional order-*m* efficiency score $\hat{\theta}_k^{m,z}$ is obtained as the average of the conditional efficiency scores $\theta_k^{b,m,z}$ computed at each b-th iteration ($b = 1, \dots, B$):

$$\hat{\theta}_k^{m,z} = \frac{1}{B}\sum_{b=1}^{B} \theta_k^{b,m,z} \tag{5}$$

In addition to using the conditional approach to estimate efficiency scores that account for different operating contexts, the comparison between the conditional and the unconditional efficiency score helps to detect and to better understand the influence of the environmental variable on the efficiency. Practically, the ratio of the conditional and unconditional order-*m* efficiency estimates is computed as in (6) and non-parametrically regressed on the external variables z (Li and Racine, 2007).

$$Q_k = \frac{\hat{\theta}_k^{m,z}}{\hat{\theta}_k^{m}} \tag{6}$$

The framework for the non-parametric regression smoothing is the following:

$$Q_k = g(z_k) + \epsilon_k, \qquad k = 1, \dots, n \tag{7}$$

By making statistical inference, information about the sign and the statistical significance of the influence of environmental variables on the efficiency estimates can be retrieved (Rogge and De Jaeger, 2013). Moreover, it is possible to graphically visualize the ratio Q_k with a scatter plot when z is univariate or with a partial scatter plot when z is multivariate, so to help with the interpretation of the role of z on efficiency. In an input orientation, if Q_k is increasing, z plays a detrimental (unfavourable) influence on the efficiency. Vice versa, if Q_k is decreasing, z plays a conducive (favourable) influence on the efficiency (Daraio and Simar, 2005).

3. Case study

To demonstrate the practical usefulness of the methodological approach from a managerial and operational perspective, we use a case study of Portuguese wastewater treatment plants (WWTPs). It provides a practical example of the potential problems that may occur when attempting to conduct a fair efficiency assessment of DMUs in a problematic sample.

The Portuguese case study consists of a set of 41 WWTPs that are managed by *Águas do Centro Litoral* (AdCL), a public-owned company. All WWTPs are activated sludge systems, meaning that all WWTPs are converting inputs into outputs using the same secondary treatment technology.

The set of plants selected treated 80% of total volume of wastewater and were responsible for 91% of the electricity consumption of all WWTPs managed by AdCL in 2015. All data used in this study was provided by AdCL and refers to the year 2015. The plants are quite heterogeneous in terms of dimension: 20 WWTPs serve less than 5000 inhabitants (measured in Population Equivalent, PE), 16 WWTPs serve between 5000 and 50000 inhabitants, and 5 WWTPs serve more than 50000 inhabitants. The smallest of these WWTPs was designed to serve 79 PE and the largest to serve 230020 PE. Only 8 WWTPs conduct a disinfection process. Only 5 WWTPs conduct anaerobic digestion with biogas production for energy recovery.

3.1. Modelling the WWTPs activity: Input and output specification

The literature reports several applications of DEA in the context of WWTPs performance assessment (e.g., Hernández-Sancho and Sala-Garrido, 2009). A critical phase of any DEA assessment concerns selecting of the most appropriate indicators to use as inputs and outputs of the productive process. Accordingly, input and output variables have been selected in compliance with the existing literature and the data availability. Moreover, the company managers and decision-makers of the plants under assessment have ultimately validated them to ensure a proper estimation of efficiency.

In terms of the input indicators, the majority of studies reviewed used costs (operation, maintenance and other costs) either expressed in €/m^3 (e.g., Fuentes et al., 2015; Sala-Garrido et al., 2011; Sala-Garrido et al., 2012a; Molinos-Senante et al., 2014a,b; Hernández-Sancho et al., 2011) or in €/year (e.g., Castellet & Molinos-Senante, 2016; D'Inverno et al., 2018; Guerrini et al., 2016; Hernández-Sancho et al., 2009; Molinos-Senante et al., 2016; Sala-Garrido et al., 2012b). The study by Lorenzo-Toja et al. (2015) considered the physical quantities of resources used per year (i.e., electricity and chemicals).

In this study, we selected as inputs the physical resources (energy and labour) used by each plant per year. Due to data unavailability, we could not consider other resources in the efficiency evaluation, such as maintenance actions or materials consumed. Nonetheless, energy and labour represent the largest share of the operational expenditures for a typical WWTP (Silva and Rosa, 2015). The energy indicator corresponds to the energy balance at each WWTP, obtained as the difference between the sum of electricity and natural gas consumed annually by the plant and the total amount of electricity produced (measured in kWh/year). The labour indicator corresponds to the number of full-time equivalent workers assigned to each plant, obtained as the ratio between the number of weekly person-hours spent by workers at each plant and the number of weekly hours corresponding to a full-time worker. Note that our input specification also considers the scale of operation of the WWTP, as we do not normalise the resources used by the volume of wastewater treated (m^3). This ensures that no information inherent to the size of the plant is lost.

In terms of the output indicators used in WWTP assessments, the majority of studies reviewed use the amount of pollutants removed from the wastewater, either expressed in kg/year (e.g., Castellet & Molinos-Senante, 2016; Hernández-Sancho et al., 2009; Molinos-Senante et al., 2016; Sala-Garrido et al., 2011; Sala-Garrido et al., 2012b), or in g/m^3 (e.g., Hernández-Sancho et al., 2011; Molinos-Senante et al., 2014a,b; Sala-Garrido et al., 2012a; Fuentes et al., 2015), which corresponds to the difference between the pollutant in the influent and effluent. Dong et al. (2017) introduced a novel approach by using the rate of pollutants removal. This is interpreted as the ratio between the amount of pollutant removed and the amount of pollutant present in the raw wastewater (i.e., in the influent). Therefore, this indicator accounts for the concentration of each pollutant in the influent sewage. Indeed, considering two WWTPs (A and B) that treat the same volume of wastewater, it is not the same to remove 200 mg/l out of 600 mg/l on entry in plant A, or to remove 200 mg/l out of 1000 mg/l on entry in plant B. Although the absolute amount of pollutant removed by both plants is identical, the rate of pollutant removal is higher in A than in B (e.g., A:200/600 > B:200/1000). Therefore, considering only the absolute amount of pollutant removed could imply an unfavourable evaluation of plant A since it demands a higher level of resource consumption than B. For this reason, the use of the rate of pollutant removal instead of the absolute amount of pollutant removed as the output specification allows a fairer efficiency comparison among plants.

However, since the scale is a factor that affects the performance of WWTPs (e.g., Hernández-Sancho and Sala-Garrido, 2009), the rate of pollutant removal still does not correspond to the real amount of pollutant removed from the influent wastewater. This varies in pollutant concentration, which will also impact the required effort in terms of the total amount of resource consumption. In this study, we contribute to the literature by proposing a novel WWTP output specification, namely the "pollutants actually removed", which is defined as the rate of pollutant removal multiplied by the total volume of wastewater treated (m^3 per year). In this study, we used as outputs the "pollutants actually removed", both in terms of suspended solids (SS) and chemical oxygen demand (COD), which is consistent with the input specification in terms of the total amount of resources consumed (per year).

Table 1 shows the descriptive statistics of the inputs and outputs used in the DEA model. The variables display an important variability that emerges especially looking at the quartiles. This variability is a consequence of the characteristics of the plants that constitute the sample, both in terms of size and contextual factors affecting WWTP activity. Note that the data was collected from company records, and the values were confirmed with managers and engineers working at the WWTP. Therefore, existing outliers are not likely to be erroneous or corrupted data, but real observations of the operational activity. As pointed out in the previous sections, the performance evaluation aims to provide a full picture of the WWTP system to enable the design of sound asset management policies by company managers, as well as improvements in operational procedures to be implemented by local managers of all plants.

Table 1. Descriptive statistics for inputs and outputs of the DEA model (41 WWTPs, year of 2015).

		Mean	SD	Min	Max	Q1	Median	Q3
Inputs	Energetic balance [MWh/year]	376.90	781.48	21.50	4476.00	51.74	135.30	201.40
	Labour [number of full time equivalent workers]	1.10	1.89	0.20	9.00	0.30	0.45	0.61
Outputs	$(COD_{removed}/COD_{entry})\times$Vol. wastewater [thousand m^3 per year]	987.10	2468.53	2.01	12461	62.00	140.50	323.50
	$(SS_{removed}/SS_{entry})\times$Vol. wastewater [thousand m^3 per year]	1022.00	2557	2.01	12875	57.29	141.90	331.00

3.2. Modelling the WWTPs activity: The role of contextual variables

The threat to the homogeneity of the units under analysis emerges not only from the observed variability on the magnitude of the inputs and the outputs, but also from the environmental conditions faced by the WWTPs. The majority of studies recognize the relevance of the environmental conditions faced by WWTPs on their performance. In some of these studies, two-stage approaches have been used to explain the variability of the DEA efficiency scores (e.g., Hernández-Sancho et al., 2009; Dong et al., 2017; D'Inverno et al., 2018; Molinos-Senante et al., 2016; Hernández-Sancho et al., 2011; Molinos-Senante et al., 2014a,b; Sala-Garrido et al., 2012b). Alternatively, conditional frontier estimators have also been used within the context of WWTPs performance assessment, accounting for contextual conditions directly in the computation of the efficiency scores (e.g., Fuentes et al., 2015; Guerrini et al., 2016).

Table 2 provides a review of the contextual factors deemed to play a role in the WWTP activity, listing the most frequently investigated contextual variables, classified into structural factors, operational factors and other factors.

Table 2. Overview of the contextual variables used in DEA related studies of WWTPs.

	Structural factors						Operational factors					Other factors	
Study	Plant capacity (or plant size)	Plant age	Secondary treatment technology	Technology used for sludge treatment	Nutrient removal and Tertiary treatment	Type of aeration in the bioreactor	Characteristics of influent wastewater	Capacity utilization	Energy consumption	Sludge generated	Compliance with regulatory effluent standards	Climate	Seasonality
D'Inverno et al. (2018)	x	x	x		x		x				x		x
Dong et al. (2017)	x		x				x	x				x	
Fuentes et al. (2015)	x	x				x	x						
Gómez et al. (2017)	x	x	x					x	x				
Hernández-Sancho et al. (2009)	x												
Hernández-Sancho et al. (2011)	x	x				x	x						
Guerrini et al. (2016)	x	x	x			x	x	x			x		
Lorenzo-Toja et al. (2015)	x				x		x	x				x	
Molinos-Senante et al. (2016a)	x	x	x	x				x					
Molinos-Senante et al. (2014a)	x	x		x	x								
Molinos-Senante et al. (2014b)	x	x	x	x					x	x			
Sala-Garrido et al. (2012b)													x
No. times used	**11**	**8**	**6**	**3**	**2**	**3**	**6**	**6**	**2**	**1**	**1**	**2**	**1**

Among all variables considered in the literature, plant capacity (or plant size), plant age, secondary treatment technology, capacity utilization, and characteristics of the influent wastewater were the variables chosen more often to analyse their potential influence on the efficiency of the WWTPs.

As the contextual factors may strongly influence the performance of the WWTPs, we also discussed with AdCL managers what would be the most relevant set of contextual factors to consider in the analysis of their plants.

As managers were very concerned with asset management, they demonstrated particular interest in the age of the plants. Also, they were concerned with five additional aspects that could be relevant to explain differences in relative performance. The first of these aspects is the plant dimension. There are very big and very small plants in the sample, and their size affects the operation. The second aspect is capacity utilization. This is a particularly critical aspect in at least one plant that was built to accommodate seasonal events (religious). The third aspect is the existence of disinfection treatment in some of the plants. This occurs due to legal requirements concerning the effluent quality. In the plants considered in our sample, the disinfection treatment is made through UV radiation, which implies additional energy consumption at these plants. The fourth aspect is the existence of sludge dehydration in some of the plants, which requires additional labour and electricity consumption. The fifth aspect is the existence of pumping facilities inside some of the plants. This is required by the geographical features of the place where the plant is located, and causes additional energy consumption.

In summary, we selected the following set of contextual factors: (i) age of the plant (expressed in terms of the number of years since the construction of the WWTP or since the last rehabilitation), (ii) installed capacity (expressed in terms of the volume of wastewater that can be treated at each plant daily, in m^3/day), (iii) capacity utilization (expressed as the ratio between the average volume of wastewater treated per day and the installed capacity), (iv) disinfection treatment (binary variable), (v) sludge dehydration (binary variable), and (vi) pumping facilities inside the WWTP (binary variable).

All the contextual factors listed above represent structural conditions at the plants, except the plant's capacity utilization that corresponds to an uncontrollable operational aspect.

Table 3 shows the descriptive statistics of the contextual variables used in the conditional efficiency assessment.

Table 3 – Descriptive statistics for the contextual variables (41 WWTPs, year of 2015).

Contextual variable	Description	Mean	SD	Min	Max
Plant age	Years since construction or last intervention	11	5	4	23
Installed capacity	m^3/day	5400	12015.49	90	49000
Percentage of utilization of the installed capacity	Average volume of wastewater treated (m^3/day)/Installed capacity (m^3/day)	54.32%	32.33%	6.59%	155%
Disinfection treatment	Dummy = 1 if present	19.51%			
Dehydration of sludge	Dummy = 1 if present	26.83%			
Pumping facilities	Dummy = 1 if present	58.54%			

To conclude, Figure 1 depicts a visual summary of the input and output indicators, as well as the environmental variables considered in this study.

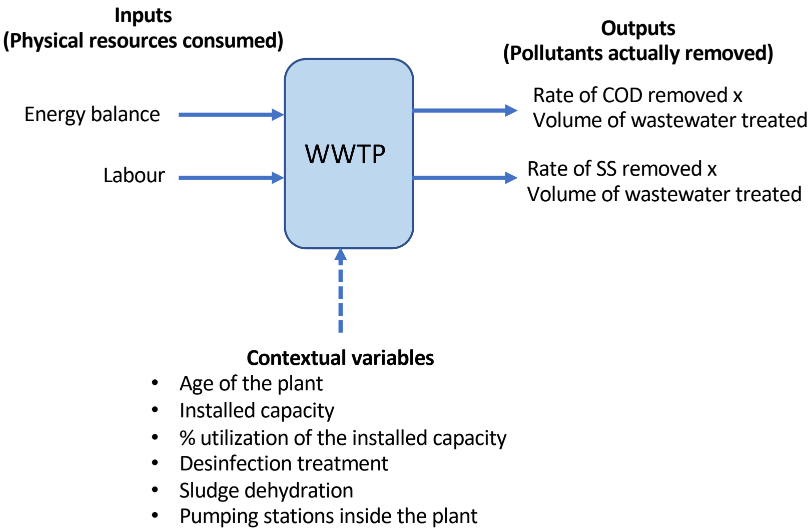

Figure 1 – Schematic representation of the conditional DEA model used to assess WWTPs performance in this study.

4. Results and Discussion

4.1 Critical comparison of results obtained from robust and deterministic DEA.

In this section, we compare the results obtained from a traditional (deterministic) DEA with those obtained using a robust unconditional approach to the problematic sample of WWTPs. We also explore the practical information to be delivered to WWTPs managers to guide performance improvements. Specifically, we used a VRS specification of the DEA model, to comply with existing literature demonstrating that these utilities exhibit VRS technology (e.g., Hernández-Sancho et al., 2009).

The first issue that must be resolved in assessments using the order-*m* method is choosing the value of *m* (number of units drawn at random with replacement among those producing at least the same output as the unit under evaluation). As explained in Section 2.2, the value of *m* can be selected by plotting the percentage of super-efficient units against the values of *m*, and then determining the elbow value after which the percentage of super-efficient units stabilizes. Figure 2 shows the sensitivity analysis for different values of *m*, both in terms of % of super-efficient units and average robust efficiency scores. The detailed data at the DMU level underlying the construction of the graphs shown in Figure 2 is shown in Appendix A.

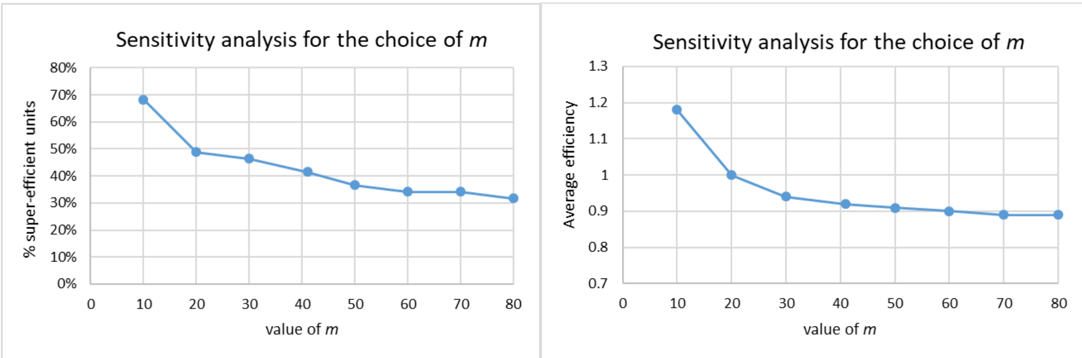

Figure 2 – Sensitivity analysis for selection of the *m*-value.

The analysis of these graphs led us to choose a value of *m* equal to 41, which corresponds to the number of DMUs in the sample. A value of *m* equal to the number of DMUs in the sample might be regarded as a recommended option for problematic samples due to the small dimension.

Table 4 shows the descriptive statistics of the robust unconditional efficiency scores for different values of *m,* together with the DEA scores of the deterministic DEA (full frontier without resampling) and deterministic DEA with peers restricted to the WWTPs with at least the same output level as the DMU under assessment.

Table 4. Summary information of efficiency results for different model specifications

Summary Statistics	Deterministic DEA		Robust DEA							
	conventional	Peers restricted	(*m* = 10)	(*m* = 20)	(*m* = 30)	(*m* = 41)	(*m* = 50)	(*m* = 60)	(*m* = 70)	(*m* = 80)
Mean	0.654	0.880	1.180	1.000	0.940	0.920	0.910	0.900	0.890	0.890
StDev	0.242	0.160	0.363	0.239	0.197	0.180	0.170	0.165	0.162	0.160
Min	0.245	0.530	0.647	0.573	0.547	0.537	0.532	0.529	0.528	0.527
Max	1	1	2.338	1.807	1.525	1.367	1.258	1.186	1.126	1.099
# of (super-)efficient units	8	21	29	23	23	22	21	21	21	21

The magnitude of the efficiency scores obtained for the deterministic DEA approaches and the robust DEA approaches are quite different. The efficiency scores obtained under the DEA approach (mean value of 0.654) are much lower than in the robust approaches, regardless of the value of *m* selected. This highlights the importance of using robust approaches to analyse problematic samples, as a full-frontier may lead to results with limited face validity from a practical perspective, compromising the design of policies for continuous improvement.

As expected, we notice that the average value of deterministic DEA with peers restricted is practically the same as that obtained in the robust DEA for large values of *m* compared to the sample size (*m* =

70 and *m* = 80). Furthermore, for the DMUs considered inefficient, the magnitude of the difference between the estimates obtained for individual efficiency scores in the Deterministic DEA with peers restricted model and the Robust DEA (m=80) was under 0,5% for 17 DMUs. The largest difference (2.6%) was observed for DMU P38 (see Table A.1 in Appendix). Also, the number of (super-) efficient DMUs observed in the deterministic DEA with peers restricted is similar to that obtained in the robust approach.

We highlight the finding that the input-oriented deterministic DEA with peers restricted is very similar to the input-oriented robust DEA for values of *m* close to the sample size. This could be anticipated given that both methods are using the same conditional assumption, i.e., each DMU can only be compared with those that produce at least the same amount of output. From a deterministic point of view, the deterministic DEA with peers restricted can be seen as a dynamic clustering (Golany and Thore, 1997), since the peer restriction used is softening the variability of the sample in terms of the output indicator, such that each DMU is fairly evaluated within a group of comparable peers. See in Appendix B the mathematical programming model formulation used for the deterministic DEA model with peers restricted, as well as the results obtained from its application to the sample of WWTPs from AdCL (Table B.2).

Using the robust DEA with *m* equal to 41, we found 19 plants performing inefficiently. Since this paper aims to get into the black box of the efficiency assessment, Table 5 presents the information obtained regarding efficiency scores and targets for the detected inefficient DMUs.

Table 5 – Results of the efficiency analysis for the inefficient units, by using the robust approach.

Inefficient WWTPs	# units with ≥ output level than the evaluated unit	Robust efficiency score	Target Energetic Balance (MWh/year)	Potential Energy savings (MWh/year)	Target Labour (# of FTE workers)	Potential Labour savings (# of FTE workers)
P5	4	0.924	1485.2	122.9	5.2	3.80
P9	7	0.857	489.6	421.0	1.2	0.20
P12	13	0.605	121.7	79.7	0.4	0.25
P14	18	0.572	126.7	94.5	0.3	0.26
P15	19	0.537	98.4	84.6	0.5	0.42
P16	13	0.670	115.2	56.3	0.4	0.21
P17	20	0.677	114.4	54.4	0.4	0.19
P19	11	0.870	153.1	22.8	0.2	0.20
P22	23	0.659	117.9	61.2	0.4	0.21
P23	22	0.769	103.9	31.4	0.5	0.14
P24	15	0.907	91.6	9.4	0.5	0.10
P26	40	0.696	24.9	11.0	0.3	0.15
P28	25	0.928	105.3	8.0	0.4	0.03
P29	29	0.678	105.3	50.4	0.3	0.15
P34	16	0.922	120.4	10.1	0.4	0.03
P35	21	0.674	145.1	72.2	0.3	0.13
P36	36	0.985	32.6	0.65	0.3	0.07
P38	34	0.820	34.4	7.6	0.3	0.13
P40	14	0.819	146.1	32.5	0.2	0.06
Mean		**0.766**	**196.4**	**64.8**	**0.7**	**0.35**
St.Dev		**0.136**	**317.7**	**93.0**	**1.1**	**0.84**
Min		**0.537**	**24.9**	**0.65**	**0.2**	**0.03**
Max		**0.985**	**1485.2**	**421.0**	**5.2**	**3.80**

From the analysis of Table 5, we can see that even despite the features underlying the robust efficiency assessment (mitigation of the impact of outliers in the comparative evaluation process through resampling and construction of a partial frontier based only on units with the same or higher level of output than the unit under evaluation) there is still room for improvement. The average inefficiency is 23.4% (1-76.6%), which means there is room for reducing the overall level of resource usage by 23.4%. For the energy input, the total energy wasted due to inefficiency is 1230.8 MWh/year, corresponding to 8% of the total energy consumed by the 41 WWTPs. In the case of labour, the total number of full-time equivalent workers that can be potentially reduced is 6.7 (or equivalently 268 hours per week, assuming 1 FTE corresponds to 40 hours per week). This labour time could be applied to different functions within the company to improve the overall functioning of the WWTPs system.

Table 6 complements the insights gained by looking into the black box of the efficiency assessment exercise with the description of the peers identified. Table 6 presents the vector of intensities λ_j for each DMU, which indicates how relevant other (observed) WWTPs (j) are for constructing the benchmark against which the efficiency of plant k is assessed. The values of lambda presented in the intensity matrix are the average values of the lambdas calculated for each plant in B iterations. Recall that the efficiency results are robustified by creating B (=2000) efficiency scores per plant given the subsample of *m* (=41) plants (the order-*m* approach) generated in each iteration.

Table 6 – List of peers and intensity values for the units considered inefficient under the robust approach.

	Intensity values of WWTPs considered peers																					
Inefficient WWTP	P1	P2	P3	P6	P10	P18	P19	P20	P21	P24	P25	P26	P28	P30	P32	P33	P34	P36	P37	P38	P40	P41
P5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
P12	0	0	0	0.46	0	0	0.03	0	0	0	0.51	0	0	0	0	0	0	0	0	0	0	0
P14	0	0	0	0.5	0	0	0	0	0	0.03	0.38	0	0	0	0	0	0.07	0	0	0	0.02	0
P15	0	0	0	0.08	0	0	0	0	0	0.08	0.8	0	0	0	0	0	0.03	0	0	0	0	0
P16	0	0	0	0.37	0	0	0.02	0	0	0	0.62	0	0	0	0	0	0	0	0	0	0	0
P17	0	0	0	0.3	0	0	0	0	0	0.06	0.54	0	0	0	0	0	0.09	0	0	0	0	0
P19	0	0	0	0.98	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P22	0	0	0	0.32	0	0	0	0	0	0.07	0.46	0	0	0	0	0	0.14	0	0	0	0.01	0
P23	0	0	0	0.15	0	0	0	0	0	0.09	0.69	0	0	0	0	0	0.06	0	0	0	0	0
P24	0	0	0	0	0	0	0	0	0	0.05	0.94	0	0	0	0	0	0	0	0	0	0	0
P26	0	0.01	0	0	0	0	0	0	0	0	0	0.01	0	0.08	0.23	0.63	0	0.04	0	0	0	0
P28	0	0	0	0.14	0	0	0	0	0	0	0.12	0	0.02	0	0	0	0.02	0	0.67	0	0.02	0
P29	0	0	0	0.1	0	0.1	0	0.03	0	0	0	0	0	0	0	0	0.01	0	0	0	0.03	0.72
P34	0	0	0	0.42	0	0	0	0	0	0.03	0.47	0	0	0	0	0	0.07	0	0	0	0	0
P35	0	0	0	0.69	0	0	0.02	0	0	0.02	0.14	0	0	0	0	0	0.02	0	0	0	0.11	0
P36	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0.69	0	0	0	0.21	0	0.08	0	0
P38	0	0	0.06	0	0.02	0	0	0	0	0	0	0	0	0.72	0	0	0	0	0	0.21	0	0
P40	0	0	0	0.82	0	0	0	0	0	0	0.13	0	0	0	0	0	0	0	0	0	0.05	0

From an individual perspective, we can take the example of WWTP P15 that was assigned the lowest efficiency value to give an in-depth understanding of the practicality inherent to the robust approach. This plant was considered comparable to 19 other WWTPs of AdCL, meaning that these other 19 plants produce the same or more output that WWTP P15 (see Table B1 in Appendix B). P15 uses a very high percentage of the installed capacity (in fact, on average, it was working with overload in the year analysed). It is a relatively new facility (with 9 years in 2015). Table 6 shows that plant P25 was considered the most relevant peer for plant P15, with an intensity value of 0.8. This suggests that P15 should look for the best practices implemented in plant P25 to foster performance improvements. In fact, the similarities between P15 and P25 are noticeable. Both were overloaded in the period analysed, and P25 is only 6 years older than P15, indicating that both are in useful age. Note that using the deterministic DEA approach, the efficiency score of WWTP P15 was only 0.334 (a difference of 20.3% in relation to the robust approach).

Another example is WWTP P13, with the biggest difference between the results obtained in the DEA and robust approaches (64.7%). In the robust approach, this plant is considered comparable to 7 other plants (see Table B1 in Appendix B), and was considered slightly super-efficient. This suggests that plant P13 was being unfairly evaluated when using the DEA (efficiency score equal to 0.353) due to comparisons with very different WWTPs.

Finally, we consider the case of WWTP P9, which was identified as the most inefficient plant in the DEA approach (efficiency score equal to 0.245). This plant was assigned an efficiency score of 0.857 in the robust approach. Looking at the characteristics of this particular WWTP, we see that it had a very low average percentage of utilization of the installed capacity in the year of 2015 (37.9%). This plant is significantly affected by seasonality due to religious events that attract many people to the city of Fátima for a few days per year in the area served by this WWTP. In addition, this WWTP also undertakes a disinfection treatment. Although it may be challenging to improve the efficiency of this plant due to the adverse contextual conditions, the robust approach allows comparisons with 7 other plants in the sample and found WWTP P21 as its peer. This plant also had a very low average percentage of utilization of the installed capacity (18.1%) in 2015. It can be used as a benchmark to share best practices and guide the implementation of enhanced management and operational procedures at plant P9.

As expected, all units considered efficient in the deterministic DEA approach continue to be considered efficient in the robust approach. However, the number of efficient units increases significantly in the robust approach by mitigating the effect of outliers in the assessment and imposing a more restricted assumption regarding the comparability among DMUs. Accordingly, a total of 22 WWTPs were identified as (super-)efficient in the robust approach, against only 8 efficient WWTPs under the deterministic approach. Consequently, the robust efficiency assessment results match more closely stakeholders perceptions concerning WWTPs' performance than the results of the deterministic DEA model. The targets obtained using DEA would be very demanding and perceived as impractical to accomplish with the technology available at the WWTPs.

Other important conclusions can be drawn by comparing the number of times each WWTP is identified as peer, and the average intensity value (λ_j) obtained in deterministic DEA versus robust DEA approaches. In the robust approach, the number of times a plant is identified as a peer is obtained by counting the number of times the average results of the intensity values obtained in the B iterations were greater than zero. This means that even a WWTP considered inefficient (on average) may have been considered efficient at least once in the B iterations and therefore be counted as a peer of itself. Tables 7 and 8 summarise this information concisely.

Table 7 –Number of times each WWTP is identified as peer and average intensity values for robust DEA.

WWTP	P1	P2	P3	P4	P6	P7	P8	P10	P11	P13	P16	P18	P19	P20	P21	P23	P24
# of times as peer	2	5	9	1	27	2	2	9	5	1	1	7	14	9	4	3	13
Average intensity	0.049	0.019	0.035	0.024	0.165	0.024	0.024	0.033	0.002	0.024	0.000	0.028	0.002	0.028	0.049	0.000	0.012
WWTP	**P25**	**P26**	**P27**	**P28**	**P30**	**P31**	**P32**	**P33**	**P34**	**P35**	**P36**	**P37**	**P38**	**P39**	**P40**	**P41**	
# of times as peer	21	2	1	5	12	4	4	3	13	2	6	8	6	9	19	6	
Average intensity	0.171	0.000	0.015	0.000	0.084	0.024	0.029	0.036	0.012	0.000	0.008	0.041	0.007	0.006	0.007	0.038	

Table 8 – Number of times each WWTP is identified as peer and average intensity values for deterministic DEA.

WWTP	P4	P6	P7	P27
# of times as peer	29	29	4	35
Average intensity	0.058	0.267	0.025	0.649

There are three main aspects to retain from Table 7 and Table 8. First, WWTP P6 is considered the preferable peer both in the DEA and the robust approaches. This signals outstanding relative performance at this plant in the year 2015.

Secondly, WWTP P4 was considered 29 times as peer for the underperforming units in the DEA benchmarking exercise (although with an average intensity value of only 0.06), whereas P4 is only peer to itself in the robust approach. In fact, WWTP P4 is the largest of the sample in terms of output levels, so although all other plants could have used this DMU as a peer, they have selected other plants as benchmarks in the robust analysis (it was only used once as a peer).

Conversely, the smallest plant of the sample is P27. It has been considered 35 times as a peer in the DEA approach, with the highest intensity value (0.65). However, in the robust approach, this WWTP could not be included in the comparator set of any other plant (see Table B1 in Appendix B). Therefore, given the assumptions underlying the robust approach, this plant could not be a peer to any other plant except itself (and indeed, it is a self-comparator, despite having a small lambda value in its own efficiency evaluation).

Thirdly, WWTP P25 is considered as a preferable peer in the robust approach (from Table 7 it can be observed that 21 plants have used it as a benchmark in the efficiency evaluation), whereas it is considered inefficient in the DEA approach (efficiency score of 0.739). In fact, P25 is comparable only to 12 plants in the overall sample (see the line corresponding to P25 in Table B1 of Appendix B). This shows that by mitigating the effect of outliers, some plants that are considered inefficient in the DEA approach can stand out as examples of best practices in the robust assessment.

The information retrieved from this analysis can be summarised in three main takeaways. The first aspect is that valuable managerial information can be retrieved by exploring the by-products of a robust non-parametric model. Identifying peers paves the way for information sharing among managers, operators, and technical staff. It thus constitutes a starting point for the planning and action-taken that leads to performance improvements. Moreover, the computation of targets gives both managers and operational staff a direct indication of necessary improvements. It also triggers the gathering of new insights into both current practices and best practices observed in identified peers. In addition, the possibility of conducting internal benchmarking to identify best practices is vital to motivate continuous improvement, taking as inspiration examples known for the company. Otherwise, company's operational staff may perceive goal-setting based on external benchmarking or without empirical support as unrealistic. The second one concerns the comparison of the results corresponding to robust and deterministic approaches. Best-practice sharing can be best promoted using robust approaches, as the peers identified are more similar to the unit under assessment. This is actually an essential ingredient for fair benchmarking evaluations. Finally, even if only deterministic models are used, their implementation should consider peers' restrictions. This will enable obtaining results that are more aligned with the robust approaches (especially for large values of m).

4.2 Results of the conditional robust DEA

As demonstrated in the previous section, the robust approach allows a trustworthy comparative evaluation of units, which serves as a baseline reference for estimating the efficiency levels of the WWTPs. To complement this analysis, we also explored the relationship of contextual factors with plants' performance. This exploratory study aims to grasp additional insights concerning the evaluation of small samples with variability in the values of the input and output indicators and facing heterogeneous environmental conditions.

The statistical significance of the models is an issue that must be explored with caution. Also, for small samples, it is not possible to include all potentially relevant contextual factors in the same model, which prevents exploring interactions among variables. Nevertheless, conditional models can still guide decision-makers concerning the significance and direction of influence of each contextual variable on the performance of the units under assessment.

Table 9 shows the summary results obtained for the conditional models tested. For the factors with a statistically significant impact on efficiency, this table also indicates the direction of influence of the conditional factor on WWTP efficiency (favorable or unfavorable). Appendix C (Table C.1) provides the complete information on the efficiency scores obtained for each WWTP in the models considered.

Table 9 – Results of the robust conditional efficiency analysis and statistical inference.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Variables	Age	Installed capacity	% of utilization of installed capacity	Disinfection treatment	Sludge dehydration	Pumping facilities
Mean	0.945	0.904	0.940	0.916	0.922	0.910
St.Dev.	0.111	0.168	0.129	0.168	0.146	0.170
Min.	0.639	0.528	0.526	0.535	0.535	0.536
Max.	1.058	1.195	1.166	1.323	1.155	1.249
Direction of influence		Unfavorable		Unfavorable	Unfavorable	Unfavorable
p-value	0.067	0.049*	0.407	0.001*	0.044*	0.018*

*Statiscally significant at a 5% confidence level

We can conclude that two of the factors considered (age and percentage of utilization of installed capacity) do not have a statistically significant relationship with efficiency, and thus the direction of influence is not reported on Table 9. Regarding the variable age, similar results were found, for example, by Hernández-Sancho et al. (2011) and D'Inverno et al. (2018).

Regarding the percentage of utilization of the installed capacity, Guerrini et al. (2016) found that high use of the installed capacity positively influences the performance levels. Although the analysis of our sample did not confirm this result at a statistically significant level, the analysis of the scatter plot for this conditional model (see Figure 3) shows a decreasing slope for a percentage of the utilization of the installed capacity between 50% and 110%, signaling that in this range of values the effect of increasing the use of the installed capacity may have a favourable influence on efficiency. This result is in accordance with what is described by Lorenzo-Toja et al. (2015) and by Dong et al. (2017).

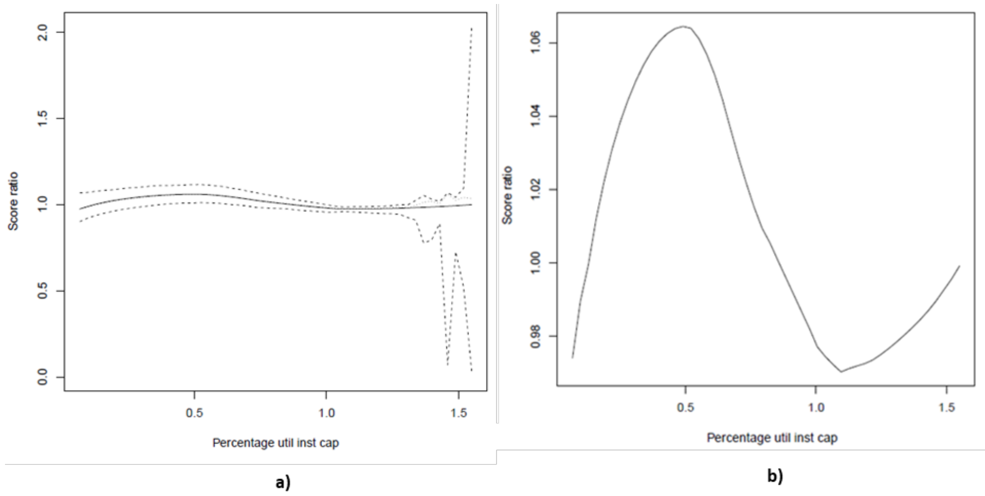

Figure 3 – Univariate scatter plot for the influence of the percentage of utilization of the installed capacity a) with confidence interval; b) without confidence interval.

According to our sample, only severe over-sizing (below 50% of utilization) and under-sizing (more than 120% of utilization) are detrimental to good performance. In our sample, 18 WWTPs are oversized and 4 WWTPs are undersized (totalizing more than half of the plants in the sample), which justifies the observed unfavorable trend of this contextual factor on performance. Furthermore, as the average percentage of utilization of the installed capacity is around 50%, there is considerable room for increasing the capacity utilization in the long run. In the future, increasing the volume of wastewater treated at the 18 currently oversized plants might enhance efficiency. Accordingly, to face a potential increase in wastewater treatment needs, it could be wise to analyse the feasibility of draining the additional volumes to these existing WWTPs, as an alternative, if possible, to the construction of new facilities.

One of the contextual factors that is commonly studied in the WWTPs performance assessment literature is the plant dimension (see Table 2). This factor is proxied in our study by the installed capacity. Table 9 shows that the installed capacity has an unfavourable effect on WWTP efficiency for the AdCL sample, meaning that higher installed capacity is associated with lower levels of efficiency. Although the study by Fuentes et al. (2015), also using a robust conditional approach, reached a similar conclusion regarding the unfavourable effect of dimension on efficiency, this is a result contrary to what is commonly described in the literature (e.g., Guerrini et al., 2016). Although this effect was found to be statistically significant at a 5% level for our sample, the p-value is quite high (0.049), so the results should be viewed with caution. Note also that in our sample, the majority of plants (36 out of 41) have a small installed capacity (i.e., lower than 6500 m^3/day), while only 5 plants have a big installed capacity (i.e., greater than 20000 m^3/day).

Concerning the disinfection treatment, sludge dehydration, and the existence of pumping facilities, all these factors have a statistically significant association with the efficiency of WWTPs, and the direction of their influence is unfavorable. Thus, some of the inefficiency identified in the robust unconditional model can be attributed to the influence of these contextual conditions. This result is as expected, as all these factors imply electricity consumption and require additional workforce.

It has been argued in the literature (e.g., Lorenzo-Toja et al., 2015) that the plants with a tertiary treatment, which includes disinfection, should not be comparatively assessed with plants that don't conduct this kind of treatment. However, in the presence of small samples, it is not possible to separate the plants into groups according to the combination of contextual conditions that they face. This is precisely what motivated this paper. We provide an alternative approach to evaluate small samples, consisting of using a robust unconditional DEA model to estimate a baseline efficiency level for all plants under evaluation, followed by a robust conditional model, such that the decision maker becomes aware of the particular issues (contextual conditions) that may be correlated with the performance of the plants. Consequently, the design of policy measures to improve efficiency can be done in light of the additional information provided by the conditional analysis.

It should also be noted that although the conditional analysis revealed that some of the factors have a statistically significant relationship with the efficiency levels, the correlation between the efficiency scores obtained with robust conditional and unconditional models is very high (see Table 10).

Table 10 - Spearman correlation matrix for different model specifications

	Robust Conditional DEA					
	Age	Installed capacity	% utilization of installed capacity	Disinfection treatment	Sludge dehydration	Pumping facilities
Robust DEA	0.7673	0.9606	0.8112	0.9770	0.9737	0.9922

In summary, our results demonstrate that four contextual variables have a statistically significant association with the efficiency levels. Therefore, the results revealed by the conditional models should be taken into account, especially when considering the identification of peers to guide performance improvements at the plants.

5. Conclusion

In order to guide the design of policies for performance improvement, efficiency studies at macro level are often directed to a sector or area to understand trends and reveal opportunities and threats. In these cases, the samples to analyse are carefully selected, such that it is possible to reach unbiased conclusions. At a micro level, companies may also be interested in conducting efficiency studies to tackle better internal challenges for the improvement of their utilities. In these internal benchmarking situations, the samples available may be problematic, undermining the reliability of the evaluation process. As defined in this paper, problematic samples are small samples whose units present high variability in terms of input-output indicators and/or environmental conditions.

Considering the importance of performance assessment studies in the search for improvements within organizations, the variability within the samples should be treated without compromising the quality of the information obtained. This must be done without disregarding any unit, especially in companies or entities that manage a small number of units.

In this paper, we showed the potential of an efficiency analysis framework based on the combination of robust unconditional and conditional DEA models. This procedure was implemented using a real-world problematic sample of a Portuguese wastewater company, and was validated by company managers. It was demonstrated that a robust efficiency approach could successfully tackle the challenge of identifying best practices and uncovering the real potential for improvement within a company. This approach can identify a more realistic set of targets than the deterministic DEA formulation.

For example, in the case study consisting of 41 WWTP from the AdCL company, the baseline value for reducing the total energy consumed was 8%. This is more conservative and realistic than the value proposed by the traditional DEA approach (28%). Furthermore, using a robust conditional approach, it was possible to derive additional information about the effect of the environmental factors that were not previously considered directly in the unconditional efficiency assessment. More precisely, it was concluded that plants' installed capacity, the existence of disinfection treatment, sludge dehydration and pumping facilities inside the plants have a statistically significant association with plants' efficiency levels. Therefore, these factors must be taken into account in the identification of the peers where best practices should be observed to foster performance improvements at underperforming WWTPs.

The information gathered can help identify priorities of investments, determine the type of technical measures to implement at inefficient units based on information sharing, and assist the decision-making process.

The practical application of the efficiency assessment framework proposed in this paper also constitutes an example of a formative evaluation procedure that other water companies may adopt in the pursuit of improved sustainability of their assets. Although the performance assessment framework proposed in this paper was applied to the specific case of the wastewater industry, it can also be adopted in other industry contexts, especially in internal benchmarking exercises where the samples available are small, and the operating characteristics or environmental conditions differ significantly among DMUs.

Our study only covers information from one year, so future research should explore the evolution of performance over time. This could mitigate the variability in input and output levels that may occur over the years and systematically identify the plants that consistently maintain high or low levels of performance over the years. It would also allow monitoring the effect of the environmental conditions on DMUs performance and investigate the changes in productivity levels over time.

Acknowledgements

The authors wish to acknowledge the financial support of Project "More Sustainable WWTPs”. This project is financed by the Portuguese water company Águas do Centro Litoral (AdCL). The authors are especially grateful to all the support provided by Tiago Braga and Paulo Leitão from AdCL. The authors also wish to gratefully acknowledge Flávia Barbosa for the computational assistance in building the DEA model with peers restricted, and to Kristof De Witte and Laura Carosi for their valuable comments and suggestions on a previous version of the paper. The first author also wishes to acknowledge the financial support from FCT - Fundação para a Ciência e a Tecnologia (Portuguese national funding agency for science, research and technology), given through the Grant PD/BD/142815/2018 respecting the Doctoral Program on Sustainable Energy Systems by MIT-Portugal. The fourth author also gratefully acknowledges financial support from Research Foundation-Flanders, FWO (Postdoctoral Fellowship 12U0219N).

References

Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management science*, *39*(10), 1261-1264.

Aragon, Y., Daouia, A., & Thomas-Agnan, C. (2005). Nonparametric frontier estimation: a conditional quantile-based approach. *Econometric Theory*, *21*(2), 358-389.

Avkiran, N. K., & Rowlands, T. (2008). How to better identify the true managerial performance: State of the art using DEA. Omega, 36(2), 317-324.

Barros, C. P. (2008). Efficiency analysis of hydroelectric generating plants: A case study for Portugal. *Energy* Economics, 30(1), 59-75.

Barros, C. P. (2006). Productivity Growth in the Lisbon Police Force. Public Organization Review, 6(1), 21-35.

Camanho, A. S., & Dyson, R. G. (1999). Efficiency, size, benchmarks and targets for bank branches: an application of data envelopment analysis. Journal of the Operational Research Society, 50(9), 903-915.

Camanho, A. S., Portela, M. C., & Vaz, C. B. (2009). Efficiency analysis accounting for internal and external non-discretionary factors. Computers & Operations Research, 36(5), 1591-1601.

Castellet, L., & Molinos-Senante, M. (2016). Efficiency assessment of wastewater treatment plants: A data envelopment analysis approach integrating technical, economic, and environmental issues. Journal of environmental management, 167, 160-166.

Cazals, C., Florens, J. P., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. Journal of econometrics, 106(1), 1-25.

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

D'Inverno, G., Carosi, L., Romano, G., & Guerrini, A. (2018). Water pollution in wastewater treatment plants: An efficiency analysis with undesirable output. European Journal of Operational Research, 269(1), 24-34.

Daouia, A., & Simar, L. (2007). Nonparametric efficiency analysis: a multivariate conditional quantile approach. Journal of Econometrics, 140(2), 375-400.

Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach. Journal of Productivity Analysis, 24(1), 93-121.

Daraio, C., & Simar, L. (2007). Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. Journal of Productivity Analysis, 28(1), 13-32.

De Witte, K., & Kortelainen, M. (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. Applied Economics, 45(17), 2401-2412.

De Witte, K., & Marques, R. C. (2010a). Incorporating heterogeneity in non-parametric models: a methodological comparison. International Journal of Operational Research, 9(2), 188-204.

De Witte, K., & Marques, R. C. (2010b). Influential observations in frontier models, a robust non-oriented approach to the water sector. Annals of Operations Research, 181(1), 377-392.

De Witte, K., & Schiltz, F. (2018). Measuring and explaining organizational effectiveness of school districts: Evidence from a robust and conditional Benefit-of-the-Doubt approach. European Journal of Operational Research, 267(3), 1172-1181.

Dong, X., Zhang, X., & Zeng, S. (2017). Measuring and explaining eco-efficiencies of wastewater treatment plants in China: an uncertainty analysis perspective. Water Research, 112, 195-207.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. European Journal of Operational Research, 132(2), 245-259.

Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society: Series A (General), 120(3), 253-281.

Fuentes, R., Torregrosa, T., & Ballenilla, E. (2015). Conditional order-m efficiency of wastewater treatment plants: The role of environmental factors. Water, 7(10), 5503-5524.

Fusco, E., Vidoli, F., & Rogge, N. (2020). Spatial directional robust Benefit of the Doubt approach in presence of undesirable output: An application to Italian waste sector. Omega, 94, 102053.

Golany, B., & Thore, S. (1997). Restricted best practice selection in DEA: An overview with a case study evaluating the socio-economic performance of nations. Annals of Operations Research, 73, 117-140.

Gómez, T., Gémar, G., Molinos-Senante, M., Sala-Garrido, R., & Caballero, R. (2017). Assessing the efficiency of wastewater treatment plants: A double-bootstrap approach. Journal of Cleaner Production, 164, 315-324.

Guerrini, A., Romano, G., Mancuso, F., & Carosi, L. (2016). Identifying the performance drivers of wastewater treatment plants through conditional order-m efficiency analysis. Utilities Policy, 42, 20-31.

Hernández-Sancho, F., & Sala-Garrido, R. (2009). Technical efficiency and cost analysis in wastewater treatment processes: A DEA approach. Desalination, 249(1), 230-234.

Hernández-Sancho, F., Molinos-Senante, M., & Sala-Garrido, R. (2011). Energy efficiency in Spanish wastewater treatment plants: A non-radial DEA approach. Science of the Total Environment, 409(14), 2693-2699.

Lavigne, C., De Jaeger, S., & Rogge, N. (2019). Identifying the most relevant peers for benchmarking waste management performance: A conditional directional distance Benefit-of-the-Doubt approach. Waste Management, 89, 418-429.

Lee, C. Y., & Cai, J. Y. (2020). LASSO variable selection in data envelopment analysis with small datasets. Omega, 91, 102019.

Li, Q., & Racine, J. S. (2007). Nonparametric econometrics: theory and practice. Princeton University Press.

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 research, 141, 317-328.

Lorenzo-Toja, Y., Vázquez-Rowe, I., Chenel, S., Marín-Navarro, D., Moreira, M. T., & Feijoo, G. (2015). Eco-efficiency analysis of Spanish WWTPs using the LCA+ DEA method. Water research, 68, 651-666.

Moldan, B., Janoušková, S., & Hák, T. (2012). How to understand and measure environmental sustainability: Indicators and targets. Ecological Indicators, 17, 4-13.

Molinos-Senante, M., Gémar, G., Gómez, T., Caballero, R., & Sala-Garrido, R. (2016). Eco-efficiency assessment of wastewater treatment plants using a weighted Russell directional distance model. Journal of Cleaner Production, 137, 1066-1075.

Molinos-Senante, M., Hernández-Sancho, F., Mocholí-Arce, M., & Sala-Garrido, R. (2014a). Economic and environmental performance of wastewater treatment plants: Potential reductions in greenhouse gases emissions. Resource and Energy Economics, 38, 125-140.

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

Omrani, H., Alizadeh, A., Emrouznejad, A., & Teplova, T. (2021). Data envelopment analysis model with decision makers' preferences: a robust credibility approach. Annals of Operations Research, 1-38.

Omrani, H., Alizadeh, A., Emrouznejad, A., & Teplova, T. (2022). A robust credibility DEA model with fuzzy perturbation degree: An application to hospitals performance. Expert Systems with Applications, 189, 116021.

Roháčová, V. (2015). A DEA based approach for optimization of urban public transport system. Central European Journal of Operations Research, 23(1), 215-233.

Rogge, N., & De Jaeger, S. (2013). Measuring and explaining the cost efficiency of municipal solid waste collection and processing services. Omega, 41(4), 653-664.

Ruiz, J. L., & Sirvent, I. (2019). Performance evaluation through DEA benchmarking adjusted to goals. Omega, 87, 150-157.

Sadjadi, S. J., & Omrani, H. (2008). Data envelopment analysis with uncertain data: An application for Iranian electricity distribution companies. Energy policy, 36(11), 4247-4254.

Sala-Garrido, R., Hernández-Sancho, F., & Molinos-Senante, M. (2012a). Assessing the efficiency of wastewater treatment plants in an uncertain context: a DEA with tolerances approach. Environmental Science & Policy, 18, 34-44.

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

Sala-Garrido, R., Molinos-Senante, M., & Hernández-Sancho, F. (2012b). How does seasonality affect water reuse possibilities? An efficiency and cost analysis. Resources, conservation and recycling, 58, 125-131.

Schiltz, F., De Witte, K., & Mazrekaj, D. (2020). Managerial efficiency and efficiency differentials in adult education: a conditional and bias-corrected efficiency analysis. Annals of Operations Research, 288(2), 529-546.

Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. Management science, 44(1), 49-61.

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. Journal of Productivity Analysis, 20(3), 391-424.

Silva, C., & Rosa, M. J. (2015). Energy performance indicators of wastewater treatment: a field study with 17 Portuguese plants. Water Science and Technology, 72(4), 510-519.

United Nations (2015). Transforming our world: the 2030 agenda for sustainable development. General Assembly Resolution A/RES/70/1. New York: United Nations.

Vaz, C. B., Camanho, A. S., & Guimarães, R. C. (2010). The assessment of retailing efficiency using network data envelopment analysis. Annals of Operations Research, 173(1), 5-24.

Wheelock, D. C., & Wilson, P. (2003). Robust nonparametric estimation of efficiency and technical change in US commercial banking. Federal Reserve Bank of St. Louis Working Paper Series, (2003-037).

Wilson, P. W. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. Journal of Business & Economic Statistics, 11(3), 319-323.

Wilson, P. W. (1995). Detecting influential observations in data envelopment analysis. Journal of Productivity Analysis, 6(1), 27-45.

Appendix A. Supplementary information on the choice of the *m*-value.

Table A.1. Robust efficiency scores of WWTPs according to each *m*-value tested.

WWTP	Deterministic DEA	Deterministic DEA (peers restricted)	Robust DEA (m = 10)	Robust DEA (m = 20)	Robust DEA (m = 30)	**Robust DEA (m = 40)**	Robust DEA (m = 50)	Robust DEA (m = 60)	Robust DEA (m = 70)	Robust DEA (m = 80)	Robust DEA (m = 90)
P1	0.815	1	1	1	1	**1**	1	1	1	1	1
P2	1	1	1.325	1.114	1.048	**1.024**	1.013	1.006	1.004	1.003	1.002
P3	1	1	1.467	1.205	1.1	**1.057**	1.032	1.02	1.013	1.01	1.005
P4	1	1	1	1	1	**1**	1	1	1	1	1
P5	0.433	0.924	0.928	0.924	0.924	**0.924**	0.924	0.924	0.924	0.924	0.924
P6	1	1	2.139	1.346	1.122	**1.04**	1.014	1	1	1.001	1
P7	1	1	1.017	1	1	**1**	1	1	1	1	1
P8	0.409	1	1.134	1.015	1	**1**	1	1	1	1	1
P9	0.245	0.857	0.952	0.866	0.859	**0.857**	0.857	0.857	0.857	0.857	0.857
P10	1	1	1.658	1.359	1.21	**1.14**	1.085	1.062	1.04	1.026	1.018
P11	1	1	1.247	1.077	1.03	**1.007**	1.003	1.001	1	1	1
P12	0.439	0.593	0.763	0.654	0.619	**0.605**	0.599	0.595	0.594	0.593	0.593
P13	0.353	1	1.132	1.012	1.002	**1**	1	1	1	1	1
P14	0.396	0.562	0.664	0.603	0.583	**0.572**	0.568	0.566	0.564	0.563	0.563
P15	0.334	0.526	0.647	0.573	0.547	**0.537**	0.532	0.529	0.528	0.527	0.527
P16	0.487	0.657	0.842	0.723	0.687	**0.671**	0.663	0.66	0.658	0.658	0.657
P17	0.400	0.663	0.787	0.713	0.688	**0.677**	0.672	0.669	0.666	0.665	0.664
P18	0.543	1	1.212	1.081	1.043	**1.025**	1.015	1.009	1.007	1.005	1.003
P19	0.619	0.867	1.068	0.905	0.879	**0.871**	0.869	0.868	0.868	0.868	0.867
P20	0.536	1	1.237	1.097	1.058	**1.036**	1.025	1.016	1.01	1.006	1.004
P21	0.647	1	1.386	1.065	1.009	**1**	1	1	1	1	1
P22	0.333	0.641	0.768	0.696	0.672	**0.658**	0.652	0.648	0.645	0.643	0.642
P23	0.381	0.749	0.902	0.817	0.787	**0.77**	0.762	0.756	0.754	0.751	0.751
P24	0.572	0.899	1.094	0.956	0.922	**0.906**	0.904	0.901	0.9	0.9	0.899
P25	0.739	1	1.373	1.123	1.051	**1.023**	1.011	1.003	1.002	1.001	1
P26	0.611	0.663	0.954	0.769	0.72	**0.695**	0.682	0.674	0.671	0.667	0.665
P27	1	1	1.625	1.295	1.2	**1.139**	1.107	1.086	1.063	1.05	1.035
P28	0.445	0.914	1.042	0.965	0.94	**0.928**	0.923	0.92	0.917	0.916	0.915
P29	0.444	0.659	0.797	0.72	0.692	**0.677**	0.67	0.667	0.663	0.662	0.661
P30	0.982	1	2.338	1.807	1.525	**1.367**	1.258	1.186	1.126	1.099	1.059
P31	0.681	1	1.588	1.182	1.051	**1.018**	1.002	1.002	1.002	1	1
P32	0.968	1	1.652	1.285	1.18	**1.12**	1.087	1.067	1.047	1.038	1.024
P33	0.969	1	1.48	1.194	1.092	**1.049**	1.033	1.017	1.012	1.007	1.003
P34	0.637	0.910	1.04	0.957	0.934	**0.922**	0.916	0.914	0.912	0.911	0.911
P35	0.5	0.647	0.848	0.737	0.693	**0.672**	0.663	0.656	0.653	0.65	0.649
P36	0.751	0.937	1.371	1.094	1.017	**0.982**	0.967	0.955	0.951	0.949	0.943
P37	0.486	1	1.176	1.061	1.026	**1.01**	1.006	1.004	1.002	1.001	1.001
P38	0.635	0.742	1.18	0.938	0.861	**0.824**	0.802	0.786	0.774	0.768	0.76
P39	0.696	0.951	1.371	1.139	1.054	**1.009**	0.986	0.973	0.966	0.958	0.957
P40	0.667	0.806	1.031	0.878	0.836	**0.818**	0.813	0.81	0.808	0.807	0.807
P41	0.667	1	1.216	1.096	1.052	**1.026**	1.019	1.01	1.006	1.004	1.003
# of (super-) efficient units	8	21	29	23	23	**22**	21	21	21	21	21
Mean	0.654	0.882	1.182	1.001	0.944	**0.918**	0.906	0.898	0.893	0.890	0.887
StDev	0.242	0.156	0.363	0.239	0.197	**0.180**	0.170	0.165	0.162	0.160	0.158
Min	0.245	0.526	0.647	0.573	0.547	**0.537**	0.532	0.529	0.528	0.527	0.527
Max	1	1	2.338	1.807	1.525	**1.367**	1.258	1.186	1.126	1.099	1.059

Appendix B. Suplementary information upon deterministic DEA with the peers in the order-m approach restricted.

Mathematical programming model for Deterministic DEA with peers restricted

$$\min_{\theta_{j0}^{VRS},\lambda_j,s_i^-,s_r^+} \theta_k^{VRS} - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \qquad \textbf{(1)}$$

Subject to

$$\theta_{j0}^{VRS} x_{ik} - \sum_{j\in C_k} \lambda_j x_{ij} - s_i^- = 0 \qquad i = 1,\dots,m$$

$$\sum_{j\in C_k} \lambda_j y_{rj} - s_r^+ = y_{rk} \qquad r = 1,\dots,s$$

$$\sum_{j\in C_k} \lambda_j = 1$$

$$\lambda_j, s_i^-, s_r^+ \geq 0, \qquad \forall j,i,r$$

In formulation (1), C_k is the group of peers allowed for the efficiency evaluation of DMU k under assessment.

In the case study of WWTPs from AdCL, the peers used for the evaluation of DMU k are those shown in the columns of Table B.1.

Table B.1. Matrix describing the fulfilment of the comparability criterium for the WWTPs under evaluation

	WWTPs that produce at least the same quantity																																								
WWTPs under evaluation	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30	P31	P32	P33	P34	P35	P36	P37	P38	P39	P40	P41
P1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1
P3	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1	0	1	0	1	1	1
P4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P5	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P6	1	0	0	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P7	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P8	1	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P9	1	0	0	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P10	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1	0	1	0	0	1	1
P11	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1	0	1	1	1	1	1
P12	1	0	0	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P13	1	0	0	1	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P14	1	0	0	1	1	1	1	1	1	0	0	1	1	1	0	1	0	0	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0
P15	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0
P16	1	0	0	1	1	1	1	1	1	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P17	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0
P18	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	0	0	1	1	0	1	0	0	1	0
P19	1	0	0	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P20	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0	1	0	0	1	1	0	1	0	0	1	0
P21	1	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P22	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0
P23	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	0	1	0	1	1	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0
P24	1	0	0	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0
P25	1	0	0	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P26	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P28	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	0	1	0	0	1	0	0	1	1	0	1	0	0	1	0
P29	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	1	0	1	0	0	1	1
P30	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1	0	1	0	0	1	1
P31	1	0	0	1	1	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1
P33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P34	1	0	0	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0
P35	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0
P36	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1
P37	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	0	0	0	0	1	0	0	1	1	0	1	0	0	1	0
P38	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1	0	1	1	1	1	1
P39	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1	0	1	0	1	1	1
P40	1	0	0	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0
P41	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0	1	0	0	1	1	0	1	0	0	1	1

Table B.2. presents the comparison between the results of the deterministic DEA model with peers restricted and the robust DEA model (with *m*-value equal to 41).

Table B.2. Information for comparison of deterministic DEA with peers restricted and robust DEA.

WWTP	Efficiency score (deterministic DEA with peers restricted)	Efficiency score (robust DEA)	Difference (in %)	Peers Deterministic DEA with peers restricted	Peers Robust DEA (more than 30% of the DEA runs)
P1	1	1	0.0%	DMU 1 (lambda=1)	DMU 1
P2	1	1.0231	2.3%	DMU 2 (lambda=1)	DMU 2
P3	1	1.0528	5.3%	DMU 3 (lambda=1)	DMU 3
P4	1	1	0.0%	DMU 4 (lambda=1)	DMU 4
P5	0.9236	0.9236	0.0%	DMU 1 (lambda=1)	DMU 1
P6	1	1.0426	4.3%	DMU 6 (lambda=1)	DMU 6
P7	1	1	0.0%	DMU 7 (lambda=1)	DMU 7
P8	1	1	0.0%	DMU 8 (lambda=1)	DMU 8
P9	0.8571	0.8573	0.0%	DMU 21 (lambda=1)	DMU 21
P10	1	1.1382	13.8%	DMU 10 (lambda=1)	DMU 10
P11	1	1.0078	0.8%	DMU 3 (lambda=0.7300); DMU 6 (lambda=0.2699)	DMU 3; DMU 6
P12	0.5928	0.6034	1.1%	DMU 6 (lambda=0.4614); DMU 25 (lambda=0.5386)	DMU 6; DMU 25
P13	1	1.0001	0.0%	DMU 13 (lambda=1)	DMU 13
P14	0.5622	0.5724	1.0%	DMU 6 (lambda=0.5422); DMU 25 (lambda=0.4578)	DMU 6; DMU 25
P15	0.5263	0.5363	1.0%	DMU 6 (lambda=0.0879); DMU 25 (lambda=0.9121)	DMU 6; DMU 25
P16	0.6570	0.6705	1.3%	DMU 6 (lambda=0.3526); DMU 25 (lambda=0.6474)	DMU 6; DMU 25
P17	0.6631	0.6771	1.4%	DMU 6 (lambda=0.3404); DMU 25 (lambda=0.6596)	DMU 6; DMU 25
P18	1	1.0264	2.6%	DMU 18 (lambda=1)	DMU 18
P19	0.8674	0.8699	0.2%	DMU 6 (lambda=1)	DMU 6
P20	1	1.0325	3.3%	DMU 20 (lambda=1)	DMU 20
P21	1	1	0.0%	DMU 21 (lambda=1)	DMU 21
P22	0.6405	0.6589	1.8%	DMU 6 (lambda=0.3856); DMU 25 (lambda=0.6144)	DMU 6; DMU 25
P23	0.7489	0.7679	1.9%	DMU 6 (lambda=0.1690); DMU 25 (lambda=0.8310)	DMU 6; DMU 25
P24	0.8994	0.9060	0.7%	DMU 25 (lambda=1)	DMU 25
P25	1	1.0167	1.7%	DMU 25 (lambda=1)	DMU 25
P26	0.6634	0.6958	3.2%	DMU 2 (lambda=0.0149); DMU 33 (lambda=0.9851)	DMU 2; DMU 33
P27	1	1.1336	13.4%	DMU 27 (lambda=1)	DMU 27
P28	0.9137	0.9274	1.4%	DMU 6 (lambda=0.1553); DMU 37 (lambda=0.8447)	DMU 6; DMU 37
P29	0.6590	0.6766	1.8%	DMU 6 (lambda=0.0343); DMU 41 (lambda=0.9657)	DMU 6; DMU 41
P30	1	1.3898	39.0%	DMU 30 (lambda=1)	DMU 30
P31	1	1.0141	1.4%	DMU 31 (lambda=1)	DMU 31
P32	1	1.1237	12.4%	DMU 32 (lambda=1)	DMU 32
P33	1	1.0484	4.8%	DMU 33 (lambda=1)	DMU 33
P34	0.9103	0.9222	1.2%	DMU 6 (lambda=0.4530); DMU 25 (lambda=0.5470)	DMU 6; DMU 25
P35	0.6466	0.6727	2.6%	DMU 6 (lambda=0.8045); DMU 25 (lambda=0.1955)	DMU 6; DMU 25
P36	0.9367	0.9841	4.7%	DMU 30 (lambda=1)	DMU 30
P37	1	1.0098	1.0%	DMU 37 (lambda=1)	DMU 37
P38	0.7423	0.8230	8.1%	DMU 30 (lambda=1)	DMU 30
P39	0.9506	1.0017	5.1%	DMU 6 (lambda=0.1482); DMU 30 (lambda=0.8518)	DMU 6; DMU 30
P40	0.8063	0.8178	1.2%	DMU 6 (lambda=0.8604); DMU 25 (lambda=0.1396)	DMU 6; DM U 25
P41	1	1.0260	2.6%	DMU 41 (lambda=1)	DMU 41

From Table B.2. we observe that, concerning the identification of efficient DMUs, the two methods coincide. In both cases, 22 WWTPs are considered efficient. However, while in the robust conditional DEA, 17 of these were classified as super-efficient, in deterministic DEA with restricted peers, all efficient DMUs are assigned a score equal to one.

Also from Table B.2., if we focus only the WWTPs that were assigned an efficiency score lower than 1 in the deterministic DEA with peers restricted, we observe that the scores of the two approaches are very similar for all WWTPs, only with the exception of five WWTPs (P26, P35, P36, P38, and P39) that present a difference between both approaches higher than 2.6%. However, we didn't find a particular reason that could explain this difference. The higher values of efficiency are systematically corresponding to the robust approach. This can be easily explained by the stochastic nature of the robust approach, since in each sample replication m units are randomly drawn with replacement. In this process, there is a probability for the more efficient units being left out from the sample, resulting in less demanding frontiers in some of the DEA runs.

Concerning the by-products of the DEA assessment, the peers identified by deterministic DEA are exactly the same as the peers identified using the robust DEA method with the filter of reporting only those that were used more than in 30% of the 2000 replications. Both methods allow the determination of targets for improvement for the inefficient DMUs according to the equations (2) and (3) of the main text of the paper. The intensity values (lambdas) of the peers identified for each inefficient DMU are reported in Table B.2. for the case of the deterministic DEA model with peers restricted.

Appendix C. Detailed information resulting from deterministic, robust and robust conditional DEA.

Table C. 1. Complete table with efficiency scores

WWTP	Deterministic DEA	Robust DEA	Robust Conditional DEA (*Age*)	Robust Conditional DEA (*Installed capacity)*	Robust Conditional DEA (Percentage utilization of installed capacity)	Robust Conditional DEA (D*isinfection treatment)*	Robust Conditional DEA (Sludge *dehydration)*	Robust Conditional DEA (Pumping facilities)
P1	0.8149	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
P2	1.0000	1.0219	1.0000	1.0091	1.0347	1.0136	1.0083	1.0263
P3	1.0000	1.0514	1.0000	1.0252	1.0000	1.0322	1.0223	1.0332
P4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
P5	0.4333	0.9236	0.9998	0.9832	0.9236	0.9236	0.9236	0.9236
P6	1.0000	1.0275	1.0000	1.0000	1.0000	1.0124	1.0221	1.0171
P7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
P8	0.4089	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
P9	0.2454	0.8574	1.0000	1.0000	0.8579	0.9174	0.9461	0.8573
P10	1.0000	1.1291	1.0000	1.0715	1.0370	1.0950	1.0685	1.0902
P11	1.0000	1.0099	1.0000	1.0042	1.0000	1.0025	1.0023	1.0129
P12	0.4393	0.6052	1.0000	0.5929	0.9942	0.5985	0.6678	0.6003
P13	0.3532	1.0004	1.0000	1.0000	1.0000	1.0001	1.0001	1.0003
P14	0.3961	0.5719	0.6420	0.5652	0.6367	0.5707	0.6286	0.5704
P15	0.3343	0.5365	0.7510	0.5279	0.5263	0.5343	0.5345	0.5358
P16	0.4873	0.6701	0.9283	0.6580	0.9998	0.6635	0.7159	0.6661
P17	0.3995	0.6766	0.7684	0.6667	0.7388	0.6739	0.7012	0.6799
P18	0.5427	1.0244	1.0038	1.0091	1.0176	1.0001	1.0084	1.0187
P19	0.6191	0.8703	0.8886	0.8674	0.8694	0.8688	0.9254	0.8690
P20	0.5357	1.0353	1.0005	1.0149	1.0002	1.0000	1.0107	1.0284
P21	0.6467	1.0022	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
P22	0.3333	0.6592	0.7650	0.6468	0.7119	0.7609	0.7358	0.6553
P23	0.3809	0.7686	0.9484	0.7544	0.8579	0.7649	0.9088	0.7729
P24	0.5717	0.9070	1.0000	0.8998	1.0004	0.9038	0.8995	0.9043
P25	0.7393	1.0187	1.0000	1.0003	1.0000	1.0068	1.0000	1.0127
P26	0.6108	0.6963	0.7006	0.6784	0.6650	0.6825	0.6788	0.6909
P27	1.0000	1.1417	1.0267	1.0965	1.0012	1.1166	1.0923	1.1152
P28	0.4448	0.9277	1.0000	0.9192	0.9795	0.9250	0.9453	0.9308
P29	0.4444	0.6775	0.6984	0.6665	0.7279	0.7893	0.6780	0.6735
P30	0.9824	1.3476	1.0486	1.1909	1.1458	1.3637	1.1711	1.2490
P31	0.6811	1.0188	1.0000	1.0000	1.0020	1.0038	1.0000	1.0047
P32	0.9684	1.1183	1.0000	1.0813	1.0124	1.1000	1.0757	1.0955
P33	0.9691	1.0474	1.0115	1.0225	1.0003	1.0295	1.0209	1.0550
P34	0.6369	0.9216	0.9818	0.9115	1.0016	0.9187	0.9747	0.9184
P35	0.5000	0.6739	0.6753	0.6563	0.8414	0.6659	0.7643	0.6783
P36	0.7508	0.9853	0.9424	0.9641	0.9962	0.9928	0.9605	0.9975
P37	0.4861	1.0120	1.0000	1.0033	1.0023	1.0077	1.0027	1.0084
P38	0.6351	0.8198	1.0000	0.7913	0.9386	0.8013	0.7849	0.8079
P39	0.6961	1.0050	0.9557	0.9747	1.0045	0.9912	1.0000	0.9858
P40	0.6667	0.8187	1.0000	0.8071	1.0021	0.8147	0.9311	0.8177
P41	0.6667	1.0269	1.0000	1.0111	1.0055	1.0301	1.0148	1.0229