A systematic methodology for the robust quantification of energy efficiency at wastewater treatment plants featuring Data Envelopment Analysis

S Longo*, J M Lema, A Hospido, M Mauricio-Iglesias

Department of Chemical Engineering, Institute of Technology, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain

*Corresponding author (E-mail: stefano.longo@usc.es)

Abstract

This article examines the potential benefits of using Data Envelopment Analysis (DEA) for conducting energy-efficiency assessment of wastewater treatment plants (WWTPs). WWTPs are characteristically heterogeneous (in size, technology, climate, function…) which limits the correct application of DEA. This paper proposes and describes the Robust Energy Efficiency DEA (REED) in its various stages, a systematic state-of-the-art methodology aimed at including exogenous variables in nonparametric frontier models and especially designed for WWTP operation. In particular, the methodology systematizes the modelling process by presenting an integrated framework for selecting the correct variables and appropriate models, possibly tackling the effect of exogenous factors. As a result, the application of REED improves the quality of the efficiency estimates and hence the significance of benchmarking. For the reader's convenience, this article is presented as a step-by-step guideline to guide the user in the determination of WWTPs energy efficiency from beginning to end. The application and benefits of the developed methodology are demonstrated by a case study related to the comparison of the energy efficiency of a set of 399 WWTPs operating in different countries and under heterogeneous environmental conditions.

Keywords: WWTP; Energy efficiency; Benchmarking; Two-stage DEA; Exogenous factors.

1. Introduction

Growing economic, social and administration pressures for improving energy efficiency has increased the interest of wastewater agencies, utilities and operators in the application of benchmarking procedures (Longo et al., 2016), which is considered a crucial approach to reduce operational costs (Doherty et al., 2017) and mitigate global warming (Wang et al., 2016). The European Union (EU) Energy Efficiency Directive (Directive 2012/27/EU) launched in 2012, outlines the actions deemed necessary to address the objective of "increasing energy efficiency in the EU". This has resulted in several measures, including the establishment of EU wide and national energy utilisation targets and the obligation to carry out energy audits periodically (Bertoldi et al., 2015). An example of such growing awareness also in the wastewater sector is ENERWATER[1], a project funded under the European Commission that aims at the development of a standard methodology for evaluation and improvement of energy performance in wastewater treatment plants (WWTPs).

The management tools should address the WWTP’s main goals, i.e. the compliance with the water requirements using energy, water and chemical resources in a cost-effective and sustainable way (Silva et al., 2014). This requirement is not trivial since WWTPs can perform different functions, e.g. removing chemical oxygen demand (COD), nutrients such as nitrogen (N) and/or phosphorus (P), or producing an effluent free of pathogens among others (Rodriguez-Garcia et al., 2011). Furthermore, wastewater is increasingly valued as a source of renewable resources (Fang et al., 2016), therefore a sound assessment of WWTPs performance must be capable to take into account the production of multiple outputs besides clean water (e.g. energy, fertilizers, biopolymers). In a water-resource efficiency context, Life Cycle Assessment (LCA) is highly relevant for environmental authorities, regulators, and utility managers aiming to comply with the requirement for sustainable water management (Corominas et al., 2013). However, given the centrality of the water-energy nexus, the present paper will focus on energy efficiency as one of the priority areas of European Commission, whose need for transparency will be one of the main elements addressed in the next Water Directive (European Commission, 2018).

From the aforementioned discussion, it seems clear that the usual measures of energy efficiency based on relative simple performance indicators and ratios of single input and output, such as energy use per volume

[1] ENERWATER - Standard method and online tool for assessing and improving the energy efficiency of wastewater treatment plants. More information: www.enerwater.eu.

of treated wastewater, are inadequate for evaluating the energy efficiency of WWTPs. Thanks to its ability to i) handle multiple inputs and outputs, ii) identify efficient input-output relations, and iii) identify sources and quantify inefficiency in each of the compared units, Data Envelopment Analysis (DEA) represents an attractive tool for performance assessment (Cook and Seiford, 2009) and focusing on the last 10 years, the application of DEA in energy efficiency analysis has increased. It currently represents the most widely used approach in published studies on WWTPs benchmarking (Guerrini et al., 2016).

The results of DEA applied to WWTPs have highlighted that exogenous factors (any factor that is not under the direct control of the management is exogenous to the WWTP system) need to be included in the analysis to obtain well-grounded comparisons of WWTPs sets (Picazo-Tadeo et al., 2009; Carvalho and Marques, 2011; Guerrini et al., 2016; Fuentes et al., 2017). The reason is that without controlling for exogenous factors, the efficiency estimates generated by DEA will be potentially biased as inefficiency in DEA is assumed to be fully attributable to managerial decisions, while exogenous factors are not under control of the management. A large part of the works that introduce exogenous factors in DEA efficiency analysis focuses on two-stage approaches (Liu et al., 2016). The method proposed by Simar and Wilson (2007) is a recognized statistical model of general applicability that led to valid, accurate inference in DEA framework (Bădin et al., 2014). The basic idea is to estimate efficiency scores in the first stage considering only the space of inputs and outputs, ignoring the exogenous factors. Then in the second stage, a bootstrap-based algorithm is used to assess the impact of the exogenous factors and obtain valid and accurate inference for bias correction of the efficiency estimates. However, the complexity of the aforementioned methods and the considerable number of open choices, may lead to non-comparable results depending on the user and the rigour in the application of DEA and regression analysis, with the risk of biasing the evidences on which decisions and energy policies are made.

Systematic procedures have been recognizes as the best manner to address complex procedures in several fields (Lazzaretto and Tsatsaronis, 2006; Fernández-Arévalo et al., 2014; Gurevitch et al., 2018) for their ability to be transparent, reproducible and address well-defined questions in a robust way. Therefore, the main contribution of the present paper is to bring the ideas together in the context of DEA applied to WWTPs and formulate them more clearly, to offer some clarification and direction on these matters, and to present a good case study. In order to do so, a new general methodology is introduced for carrying out energy efficiency quantification at WWTPs in a systematic and rigorous way featuring DEA, thereby increasing the quality of the efficiency estimates and hence the effectiveness of benchmarking.

2. Context and previous work

DEA is a technique that essentially quantifies the efficiency of entities of interest, called decision-making units (DMUs)[2] (Charnes et al., 1978), which eventually allows identifying the best performers in the use of resources, pointing out where the potential gains may be made from possible improvements in efficiency, and helping the non-performers to achieve their potential. A DEA model estimates the efficiency of a DMU relative to the other DMUs identifying a best practice frontier with a simple restriction: all DMUs lie on or below the efficiency frontier (Cooper et al., 2011). Using linear combinations of inputs and outputs, DEA determines how efficient a DMU is at producing an output and/or utilising an input, compared to similar DMUs.

Efficiency for a set of DMUs can be estimated by the CCR[3] DEA (Charnes et al., 1978). For p inputs, q outputs and n DMUs, we can determine the input oriented efficiency of the data matrix of input and output vectors (X, Y), by solving for each observation the following constrained linear programming problem:

$$\min_{\theta,\lambda} \quad \theta$$

subject to

$$\begin{aligned} &\theta x_k \geq X\,\lambda \\ &Y\lambda \geq y_k \\ &\lambda \geq 0. \end{aligned} \tag{1}$$

where the index k represents a given observation, X is the matrix of inputs, Y is the matrix of outputs, and λ is vector of weights given to each observation. Problem (1) can be interpreted as combining plants (by weights λ) to produce an output level at least equal to plant k ($Y\lambda \geq y_k$) and then selecting the combination with the minimum input level ($\theta x_k \geq X\,\lambda$ for minimum θ). Solving the linear programming problem (1) k times generates the efficiency indices θ_{k}, one for each DMU. WWTPs with efficiency scores $\theta_k < 1$ are inefficient, since they are capable of reducing their input(s) without affecting the amount of output(s). On the other hand, efficient WWTPs receive efficiency score $\theta_k = 1$. Output oriented efficiency

---

[2] In the field of wastewater treatment a DMU is a WWTP and its evaluation of performances is defined as the ability of the plant in converting at least one input (i.e. energy) to outputs (i.e. the kg of COD removed).
[3] From the initials of authors Charnes, Cooper and Rhodes.

can be estimated by solving a similar linear programming problem (1) with a different set of restrictions (Cooper et al., 2011).

DEA, as originally proposed, is a methodology for evaluating the relative (in)efficiencies of a set of homogeneous DMUs (Charnes et al., 1978). From this assumption, we can derive the following three requirements for the correct application of DEA at WWTPs:

1. The plants under consideration perform the same function(s).
2. The factors (both inputs and outputs) characterizing the performance of all plants in the group are identical, except for differences in intensity or magnitude.
3. All the plants perform under the same set of environmental conditions.

Requirements 1 and 2 may be easily not met when comparing WWTPs since i) plants can provide the same function (e.g. removing P) using different inputs (e.g. electricity, chemicals) or ii) use the same input (e.g. electricity) to provide different services (e.g. removing COD or nutrients). Examples of such misspecifications are the inclusion of P removal rate as DEA output without including as an input the resource consumed for its removal (e.g. chemicals for P precipitation) (Dong et al., 2017) or the exclusion of the removed N when (at least part of) the plants in the analysed set carry out also N removal on top of COD removal (Guerrini et al., 2017). In such cases, unless the heterogeneity among inputs and/or outputs are properly taken into account, users are likely to have a misleading picture of the true energy efficiency of WWTPs and might make misguided decisions when investing on energy efficiency measures.

The last fundamental requirement of DEA is that DMUs operate within a homogenous environment. However, this assumption seldom holds in the wastewater sector in which the efficiency is influenced by several factors beyond managerial control. The inclusion of exogenous factors when estimating WWTPs efficiency has recently been tackled (Gómez et al., 2017; Guerrini et al., 2017). Although bias-corrected efficiency estimates (i.e. obtained from the two-stage DEA) are commonly perceived to be of better quality than efficiency estimates obtained with a single-stage DEA, the inference of the impact of the exogenous factors on the efficiency measures has to be carefully conducted because otherwise the results of the analysis may not be accurate. For example, earlier studies (Gómez et al., 2017; Guerrini et al., 2017) did not consider several regression model building issues such as the minimum required sample and multicollinearity. Furthermore, effective detection of outliers is critical for achieving useful results in benchmarking exercise. While outlier detection has been carried out by Gómez et al. (2017) by identifying observations that are “too good” relative to the DEA frontier (hereinafter referred to as “frontier outliers”), when two-stage DEA is considered, outliers that represents extreme observations with respect to the explanatory variables (i.e. exogenous factors) included in the regression model (hereinafter referred to as "regression outliers") might also distort the second stage results and cause misleading conclusions (Johnson and McGinnis, 2008).

Therefore, in light of the above considerations, a rigorous and systematic methodology for carrying out energy efficiency quantification using DEA is demanded. The Robust Energy Efficiency DEA (REED) methodology here presented overcomes these limitations by considering composite indicators to reduce heterogeneity and allowing comparability among the reference data set of WWTPs, using a systematic approach to select relevant input/output variables, and taking up a number of refined diagnostics for checking the adequacy of the second-stage regression model. Thorough examination of these properties is vital for properly capturing the effect of the exogenous factors on the WWTP efficiency as well as obtaining robust DEA efficiency scores. The usefulness of the presented REED methodology is demonstrated step-by-step on a comprehensive set of 399 plants. First, the user is guided through the data collection step, including the selection of inputs/outputs and exogenous factors, outlier detection and other validity checks, etc. Then, an appropriate DEA formulation (or model) is selected, possibly tackling the effect of the exogenous factors. Finally, the model results are refined and validated.

3. Robust energy efficiency DEA (REED) methodology

The REED methodology is based on decomposing the process of efficiency determination in a logical sequence of interconnected tasks (Fig. 1). This strategy involves four phases defined below: i) data collection and preparation, ii) model selection, iii) efficiency estimation, and iv) model refinement and validation. Clarifying comments to each of the steps are included in the methodology description as "remarks".

INSERT FIGURE 1 ABOUT HERE

3.1. Data collection and preparation

3.1.1. Data collection

Data collection involves obtaining data on the operation of a set of plants (e.g. influent and effluent characteristics) and the related energy consumption. Furthermore, for WWTP analysis, other types of variables reflecting WWTP characteristics must be included to account for known or potential influence on energy efficiency (see exogenous factors selection in section 3.3.1).

3.1.2. Inputs and outputs selection

DEA searches for units that minimize inputs and/or maximizes outputs to define the efficient performance. In other words, the resources used or required are usually the inputs and the outcomes are the outputs. In a WWTP, the outcomes are the quantities of pollutants removed from the water, e.g. COD, nutrients, pathogens, etc. depending on the function of the plant, while the inputs are the resources used for their removal (e.g. electricity and chemicals).

As the choice of variables is an area likely to suffer from user subjective preferences it is important to complement engineering knowledge with the use of a systematic method for selection of relevant inputs and outputs. This purpose makes the work proposed by Ruggiero (2005) to be very suitable framework for selecting DEA variables. This method is based on the fact that if a potential output (input) is omitted from the DEA model, then that output (input) will be positively correlated with the measured efficiency. This rule can be implemented using the regression model:

$$EE = \alpha + \beta_2 y_2 + \beta_3 y_3 + \cdots + \beta_m y_m + \varepsilon, \tag{2}$$

Where EE is the efficiency as given by DEA including only x and y_1, and y_2 through y_m are the potential outputs that could have been included in the model. Only if the parameters β_i are greater than zero, statistically significant at given level of significance (i.e. $\alpha = 0.10$) and have the proper signs (i.e. negative for outputs) is y_1 added to the model. The procedure is repeated, identifying one variable at a time and stops when there are no further variables with significant and properly signed coefficients.

Remark

For WWTPs, potential inputs commonly include electricity, energy carriers (e.g. gas, fuel) and chemicals. Potential outputs include the removal of COD, N and P (in kg removed per day), pathogens (in $m^3 \cdot log_{reduction}$), etc. In case of heterogeneity of input/output variables (i.e. requirements 1 and/or 2 in section 2 are not fulfilled) a composite indicator can be used in order to allow comparability. Requirement 1 can be overcome by joining the removal of COD, N and P in a single output expressed as total pollution equivalent (TPE) according to Benedetti et al. (2008).[4] *It is also possible to lump the different energy sources into a single input when they all refer to the same function (e.g. pollutant removal) by using the Cumulative Energy*

[4] A list of possible weights for the calculation of the TPE is reported in Longo et al. (2016).

Demand (CED) (Huijbregts et al., 2006) *to obtain the equivalent of primary energy consumption of a product over its entire lifecycle, as proposed in a publicly available deliverable of H2020 ENERWATER Project* (ENERWATER, 2018). *Using the CED the chemicals used for P removal can be converted into primary energy in order to be directly comparable with other sources of energy, e.g. electricity.*

Moreover, since the paradigm of wastewater treatment is changing towards the recovery of resources in addition to the treatment of wastewater, it may be convenient to consider the production of biogas, struvite or reclaimed water as outputs. Although we have limited the application of REED to energy efficiency, such a methodology would be readily extendable to other criteria such as capital and operational costs, space and environmental impacts; most of these criteria would be identified as inputs.

3.1.3. Preliminary checks

Size of the data sample. Two conflicting considerations are found when trying to define the right sample size. On the one hand, there is a tendency to increase the size of the dataset given that it is more likely that a large sample will contain high performance plants that would determine the efficiency frontier. On the other hand, a large set of plants has a lower probability of homogeneity within the set, and the results may be affected by some exogenous factors that are not of interest. Besides, the size of the WWTPs sample also depends on the number of inputs and outputs previously selected. A suggested rule of thumb is that to achieve a reasonable level of discrimination the number of units needs to be at least $2\,p \times q$ where $p \times q$ is the product of the number of inputs and outputs (Dyson et al., 2001). In general, a higher number of observations is required for the two-stage approach (see section 3.3.1).

Detection of frontier outliers. The accuracy of process data in WWTPs can be a significant barrier to benchmarking. Many data accuracy detection methods based on advanced statistical analysis can be used in the wastewater sector, such as mass balances, artificial neural network and principal component analysis (Doherty et al., 2017). However these methods are often unfeasible in WWTP benchmarking due to their high data requirements. As any deterministic frontier method, DEA is sensitive to extreme values and outliers. The super-efficiency test (Andersen and Petersen, 1993) can be used to individuate possible outcomes of recording or measurement errors, which is an approach widely used in non-parametric analysis. Based on this test if an efficient observation is an outlier that has been contaminated with noise then it is more likely to have an output (input) level much greater (lower) than other observations. Those observations with higher than a pre-selected screen super-efficiency scores should be eliminated.

Remark

In rare occasions, extreme observations can also represent the best practices, making the WWTP(s) a reference for the others. Furthermore, given the presence of heterogeneity in the reference set, extremes values may be the results of the effect of some exogenous factor (i.e. plant operating in a particular favourable environment may appear much more efficient), and hence, worthy of further investigation.

3.2. DEA model selection

3.2.1. Model orientation

As efficiency can be thought as output/input ratio, there are two ways to increase the efficiency: input minimization or output maximization (Cooper et al., 2011). The model orientation is selected according to the objective of the analysis. For instance, efficient N elimination is achieved when the lowest amount energy is used to remove a given mass of N and comply with effluent regulations. Hence, the goal is to minimise an input and DEA would be input-oriented. In contrast, maximising an output such as the production of biogas (or other resource recovery process) would lead to the output-oriented DEA.

Remark

Despite the advent of resource recovery facilities, WWTPs must comply with effluent requirements and therefore it is recommended to use an input-oriented DEA unless the goal of the assessment is exclusively focused on resource recovery.

3.2.2. Return to scale

The return to scale (RTS) concept (Banker et al., 2011) refers to the rate by which output changes if all inputs are changed by the same factor. If input and output increase proportionally by factor α and β (i.e. $I_2 = \alpha\, I_1$ and $O_2 = \beta\, O_1$), constant returns to scale (CRS) applies if $\beta = \alpha$, increasing returns to scale (IRS) if $\beta > \alpha$, and decreasing returns to scale (DRS) if $\beta < \alpha$.

Remark

Prior studies indicate that increasing the plant size positively affects efficiency (Longo et al., 2016) *and therefore IRS is the recommended alternative for wastewater applications featuring the use of single-stage DEA. In the case of two-stage analysis, CRS DEA may be applied and the scale (in)efficiency may be taken into account in the second-stage regression by including a proxy of the size (e.g. flowrate, person equivalent) as exogenous factor.*

3.3. Efficiency estimation

If requirement 3 is not fulfilled, i.e. some exogenous factor may affect the efficiency estimation, WWTPs comparison can be done using the two-stage DEA, as described below. The problem that arise here in that the possible effect of the exogenous factors is not know a priori, hence unless the user considers that the set of plants is homogenous, it is suggested to apply the two-stage approach in the first instance and to test for the presence of heterogeneity depending on the significance of the coefficients of the second-stage regression. If the coefficients are not significant one may deduce that the homogeneity requirement is respected, and can apply the basic DEA model (1) and obtaining the final DEA efficiency estimates.

Remark

Regarding requirement 3, approaches based on one-stage DEA have attempted to reduce heterogeneity by breaking the set of DMUs into multiple groups, and then doing a separate DEA analysis for each group. As an example, Lorenzo-Toja et al. (2015) *divided in two blocks their set of plants depending on whether or not tertiary treatment was performed on top of conventional secondary treatment. This approach is not applicable, though, for several factors or sources of heterogeneity as it leads to a combinatory explosion of ever-smaller subsets. The greater the number of splits required, the more difficult it is to estimate meaningful efficiency as efficiency scores would be artificially inflated* (Cook et al., 2013).

3.3.1. Exogenous factors selection

The type and the number of exogenous factors to include in the analysis depend on the characteristics of the dataset. Furthermore, depending on the objective of the analysis, the user may be interested in selecting only some exogenous factors, e.g. to assess the impact of regulatory constraints upon treatment efficiency. The user in this phase should select all the factors whose effect on energy consumption is beyond the control of the management and for this reason whose inefficiencies are impossible to eliminate. The number of factors is, however, limited in order to provide adequate statistical power to detect meaningful effect of these factors. A common rule-of-thumb suggests that 10 observations per exogenous variable is the minimum required sample size for regression model to ensure correct estimation of regression coefficients and standard errors that display minimal bias (Harell, 2001).

Remark

In WWTPs, exogenous *factors may reflect differences in technology choice (i.e. membrane bioreactors are known to be more energy intensive in comparison with conventional activated sludge processes), regulatory constraints (i.e. areas where further treatment is necessary to comply with national and /or international directives), urban infrastructure (i.e. combined or separate sewer), climate (i.e. rain intensity and* *temperature) and so on. Whether a variable is considered as exogenous is context dependent, depending on the objective of the study and the stakeholder(s) involved. For example, the WWTP size might be exogenous for a water utility running a WWTP but not for a water regulation board considering merging of small WWTPs into larger ones; effluent limits are exogenous to most stakeholders in wastewater sector but not to environmental regulatory bodies which may wonder e.g. to what extent lowering the nitrogen requirement will impact the WWTP energy consumption. REED is therefore conceived as a flexible methodology that can accommodate different user's objectives while being robust and repeatable, provided that the goals of the REED analysis are clearly stated.*

3.3.2. Bias-correction of DEA efficiency estimates

To evaluate the impact of exogenous variables, we propose a modification of the method reported by Simar and Wilson (2007). As efficiency is, by definition, bounded between zero and one we use the inverse of the first-stage DEA estimates of efficiency: $\left(\frac{1}{\theta_k}\right) = \delta_k$. This variable is left-bounded to one and can be regressed using a left-truncated regression in the second stage. Overall, the two-stage DEA is done as follows:

1. Compute the efficiency scores θ_k , $k = 1, \ldots, n$ by solving the (first-stage) DEA linear programming problem (1).
2. Transform the efficiency scores according to $\left(\frac{1}{\theta_k}\right) = \delta_k$.
3. Regress $\hat{\delta}_k$ with respect to the exogenous factors Z using only the subset of inefficient observations, i.e. observations with an inverse efficiency (δ) greater than one: $\hat{\delta}_k = Z_k\beta + \varepsilon_k$. Note that step (3) is the (second-stage) truncated regression where β is a vector of parameters to be estimated and $\varepsilon \in N\ (0, \sigma_\varepsilon^2)$ describes the random term. Obtain estimates of β and σ_ε, namely $\hat{\beta}$ and $\hat{\sigma}_\varepsilon$.
4. Loop over steps (4.1) to (4.3) L_1 times (i.e. 200) to obtain a set of bootstrap estimates for δ:
   4.1. For each WWTP $k = 1, \ldots, n$, draw ε_k from a normal distribution $N\ (0, \hat{\sigma}_\varepsilon^2)$ with left-truncation at $\left(1 - Z_k\hat{\beta}\right)$.
   4.2. Compute $\delta_k^* = Z_k\hat{\beta} + \varepsilon_k$.
   4.3. For input-oriented DEA, set for all WWTPs $x_k^* = x_k \frac{\delta_k^*}{\hat{\delta}_k}$, $y_k^* = y_k$ and compute $\hat{\delta}_k^*$ by solving the linear programming problem (1) replacing x_k with x_k^*.
5. For each WWTP $k = 1, \ldots, n$, compute the bias-corrected efficiency estimator $\hat{\hat{\delta}}_k = \hat{\delta}_k - \widehat{bias}_k$, where $\widehat{bias}_k = \frac{1}{L}\sum_{l=1}^{L} \hat{\delta}_{l,k}^* + Z_k\hat{\beta}$.

6. Regress $\hat{\hat{\delta}}_k$ with respect to Z to yield estimates of $\hat{\hat{\beta}}$ and $\hat{\hat{\sigma}}_\varepsilon$.
7. Loop over steps (7.1) to (7.3) L_2 times (i.e. 2000) to obtain a set of bootstrap estimates $\hat{\beta}^*$ and $\hat{\sigma}_\varepsilon^*$:
   7.1. For each WWTP $k = 1, \dots, n$, draw ε_k from a normal distribution $N\left(0, \hat{\hat{\sigma}}\right)$ with left-truncation at $\left(1 - Z_k\hat{\hat{\beta}}\right)$.
   7.2. Compute $\delta_k^{**} = Z_k\hat{\hat{\beta}} + \varepsilon_k$.
   7.3. Regress δ_k^{**} with respect to Z to yield estimates of $\hat{\hat{\beta}}^*$ and $\hat{\hat{\sigma}}_\varepsilon^*$.
8. Finally, using the bootstrap values from step 7 and the estimates of $\hat{\hat{\beta}}$ and $\hat{\hat{\sigma}}_\varepsilon$ construct confidence intervals for β.

3.4. Regression model refinement and validation

In this section we take up a number of standard refined diagnostics for checking the adequacy of the regression model and the final validation. These include methods for identifying problem of multicollinearity, outliers and influential observations (i.e. regression outliers).

3.4.1. Regression diagnostics

If the exogenous factors are correlated (i.e. multicollinearity among explanatory variable exists), the regression coefficients cannot be reliably estimated even though the model may reproduce the sampled data. Variance inflation factors (VIF) (Kutner et al., 2005) is used in this framework to detect multicollinearity. A value of VIF higher than 10 is taken as an indication that multicollinearity may be significantly influencing the regression estimates. If highly correlated exogenous factors are detected, they should be removed from the model.

A second source of spurious influence on the regression coefficients is the presence of outlying or extreme observations. When the two-stage DEA is used, outliers that represent particularly bad performance as well as bad monitoring/reporting in the explanatory variables (exogenous factors) may distort the second stage results (Johnson and McGinnis, 2008). As a consequence it is recommended that also in the second-stage (regression) analysis outlier detection to be carried out. The studentized residual, DFFITS, and Hat Matrix are three widely used methods to assess the robustness of the fit (Kutner et al., 2005).

Remark

In case of multicollinearity, a regression coefficient does not reflect any inherent effect of a particular variable but only a marginal or partial effect. For instance, correlating overall energy consumption w.r.t.

both WWTP size and flowrate as exogenous factors may result in finding that only size is relevant while flowrate appears as non-significant. As both are highly correlated, the effect of the flowrate is "shadowed" by the WWTP size.

Regarding regression outliers, if it is obvious that the outlier is due to incorrectly entered or measured data it should be dropped from the dataset. Otherwise, it remains ultimately to the user's judgement to decide whether an observation should be taken out of a data set. Removing outliers may provide more representative regression coefficients but it can dramatically narrow down the range of validity of the analysis and eliminate the actual best practices.

3.4.2. Model validation

The final step of the analysis consists in the model validation, also called sanity test/check, which refers to the evaluation of the reasonableness of the regression coefficients, the plausibility of the regression function, and the ability to generalize inferences drawn from the regression analysis. In this phase the model needs to be checked in detail for the effect from exogenous factors, what its direction might be, and only finally, what the magnitude of the effect might be. When possible, theory or previous empirical results may be useful in determining whether the selected model is reasonable.

4. Application of REED methodology for WWTPs energy performance assessment

The usefulness of the REED methodology (Fig. 1) presented in section 3 is demonstrated step-by-step by the estimation of energy efficiency of a set of WWTPs so as to i) estimate the effect of the exogenous factors on WWTP energy efficiency, ii) evaluate the energy efficiency loss or gain caused by the exogenous factors, and iii) rank a set of WWTPs according to their energy efficiency.

4.1. Data collection and preparation

4.1.1. Data collection

Data collection was carried out in the context of the H2020 ENERWATER coordination and support action, to provide an energy database for benchmarking energy efficiency (ENERWATER, 2015). The dataset used in this study was gathered i) by web-search engines; ii) by collecting energy data from regional water agencies (in particular from Germany, Spain and Switzerland); by private communications. Those WWTPs with insufficient information were omitted from the analysis, so the final dataset consisted of 399 WWTPs receiving municipal wastewater. Descriptive statistics for all variables used in the analysis are given in Table 1. Both the database and the computer code used in this case study are available upon request from the authors.

INSERT TABLE 1 ABOUT HERE

Energy consumption was gathered together with data related to the operation, namely: population equivalent (PE) load basis, both the designed value and the actually served value; average flow rate; influent and effluent wastewater characteristics, e.g. COD, total N and P.

Moreover, since energy consumption depends heavily on the technology (Krampe, 2013), WWTPs were classified according and the type of secondary treatment. The sample ranges from a few dozen PE to more than 500000 PE, and cover a wide range of technologies, e.g. biological nutrient removal (*BNR*), oxidation ditch (*OD*), membrane biological reactor (*MBR*), trickling filter (*TF*), mixed tricking filter and activated sludge processes (*TFAS*) and medium/high loading rate activated sludge (*MHLOAD*). Furthermore plants were classified based on the presence or absence of tertiary treatment (i.e. whether the plant carried out final filtration or ultraviolet disinfection). This sample covers most common layouts (up to 80%) of WWTPs in Europe in terms of treatment intensity, i.e. WWTPs including secondary or both secondary and tertiary treatment (EEA, 2013).

From the analysis of the collected data, two WWTP operational indices, dilution factor (*DF*) and load factor (*LF*), were defined based on Longo et al. (2016). *DF* is mainly function of the sewer network design, age and materials, while *LF* represents the capacity utilization of the plant compared to the design capacity, showing then if a plant is under- or over-loaded. In addition, the annual average outdoor temperature (*TEMP*) was included as a proxy of the WWTP climate.

4.1.2. Input and output selection

The efficiency of the WWTPs was analysed for the following functions: removal of COD and nutrients, e.g. N and P. The candidates to output variables were the average mass of pollutants (in kg) removed per day, which were estimated as the product of the average flowrate (in m^3/day) times the effluent/influent difference in pollutant concentration (in kg pollutant/m^3). The input variable was the overall electricity consumption (expressed as kWh/day).

The result of the regression-based test described in section 3.1.2 confirms that the inclusion of the *COD* and *N* in the output set is correlated with inefficiency differences among the WWTPs sample. In contrast, *P* was identified as not relevant and as consequence omitted, as none of the treatment technologies in our dataset is intended to carry out biological P removal. An assessment of energy efficiency including P removal would require estimating the embedded energy of chemicals for P removal (i.e. using the CED method); however, as data on the consumption of chemicals were not available, it was decided to limit the scope of the analysis to the assessment of the energy efficiency for the removal of COD and N, hence excluding P.

4.1.3. Preliminary checks

Size of the data sample. In our empirical example with one input and two outputs the minimum number of WWTPs in the dataset is 4 (i.e. $2(1 \times 2)$), which is largely exceeded.

Detection of outliers in frontier estimation. The test of super-efficiency was applied to individuate possible outcomes of recording/measurement errors using a pre-selected screen super-efficiency scores equal to 2.5. None of the WWTPs falls into this category. Therefore, all plants initially included in the dataset were considered for the analysis in this phase.

4.2. DEA model selection

4.2.1. Model orientation

The input oriented model was selected since all the outputs are bounded by the effluent regulation. As a consequence, the goal of the efficiency estimation is to identify plants that are over-utilizing resources to remove COD and N.

4.2.2. Return to scale

The CRS DEA model is selected and the difference in scale was accounted for at the second-stage (regression) analysis by including a proxy of scale (*SIZE*).

4.3. Efficiency estimation

WWTPs in the comparison set use different technologies as secondary and/or tertiary treatment (i.e. a different function that requires extra energy supply). Moreover, the WWTPs are operated under very different process conditions (e.g. large range of influent dilution and load factor), located in different countries with different climates, thus, the two-stage approach is selected to determine and correct the efficiency estimates based on a set of exogenous factors.

4.3.1. Exogenous factors selection

Four factors that may influence the energy consumption at WWTPs were selected: secondary treatment technology, plant size, influent dilution and load factor. Furthermore, the outdoor temperature was included as an additional exogenous factor. In this phase, variables that are proxies of the same factors were excluded (i.e. volume of treated wastewater, in order to avoid multicollinearity with *PE*). Then, since some of the WWTPs carry out also tertiary beside secondary treatment, the dummy variable *TERTIARY* was included to control for plants that have additional tertiary beside secondary treatment. Finally, we included a dummy variable to represent the geographical location of each plant as differences may be expected due the environmental regulations and technical progress. The resulting DEA model of WWTP energy performance has one input (*E*), two outputs (*COD*, *N*) and seven possible exogenous factors (*COUNTRY*, *SECONDARY*, *TERTIARY*, *SIZE*, *LF*, *DF*, *TEMP*). Considering our dataset composed by 399 observations, the rule-of-thumb of 10 observations for each exogenous variable is largely satisfied.

4.3.2. Bias-correction of DEA efficiency estimates

A modification of the Algorithm II of Simar and Wilson (2007) is applied to estimate bias-corrected efficiency estimates following the procedure in section 3.3.2. Two freely available toolboxes were used: the linear programming problem was solved using the Data Envelopment Analysis Toolbox for MATLAB (Álvarez et al., 2016), while for the truncated regression was employed the James Lesage Econometrics Toolbox (LeSage, 1999). The procedure to obtain the bias-corrected DEA efficiency scores was implemented in MATLAB.

4.4. Regression model refinement and validation

4.4.1. Regression diagnostics

Multicollinearity was studied by calculating the VIF. The VIF values of *COUNTRY* and *TEMP* greatly exceeded 10, which indicate that country and temperature are correlated variables. However, for their relevance these two variables are interesting to study, therefore we decided to develop two different models, one using the categorical variable *COUNTRY* and another using *TEMP* as continuous variable (Model 1 and 2, respectively in Table 2).

Outlier diagnostic methods suggested possible evidence of regression outliers at observations 167 and 204, which may affect the regressions residuals as well as the fit. To decide whether they should be removed, we proceeded by removing them from the sample and repeating the estimation procedure. Their omission was not found to have a large effect on the statistical interference. Moreover, no indication of incorrectly entered or measured data was encountered. Thus, we proceeded to maintain all the observations in the dataset.

4.4.2. Model validation

The final step of the analysis consisted in the model validation, i.e. evaluation of the reasonableness of the regression coefficients, the plausibility of the regression function, and the ability to generalize inferences drawn from the regression analysis. This step is discussed in next section 5.1 together with the presentation of the estimated energy efficiency estimation results by comparing when possible our results with the theory, previous empirical results and engineering considerations.

5. Discussion

5.1. Empirical findings

The results of the two-stage DEA are given in Table 2. Preliminary data analysis showed that energy consumption at WWTPs has a nonlinear dependency with respect to the operational variables (Longo et al., 2017). Therefore, all the continuous variables are log-transformed. Moreover, since we used the reciprocals of the efficiency scores as dependent variable in the second-stage regression a negative sign means efficiency enhancing and vice versa. The results prove that it is important to account for the characteristics and the heterogeneity of WWTPs.

INSERT TABLE 2 ABOUT HERE

Size. We first observe the expected positive relationship between the plant size and energy efficiency in the two specifications. This is consistent with previous studies (Longo et al., 2016).

Load factor. *LF* also shows a positive and highly significant relationship, however the available literature is quite conflicting on this factor. Using also a two-stage DEA approach, Gómez et al. (2017) found that the over- or under-loaded conditions does not significantly affect the WWTP efficiency, while Guerrini et al. (2017) reported increasing efficiency while increasing the ratio of used capacity (*LF* in this study). Our results confirm that plants receiving lower loads than design value present a significantly worse energy performance, and energy efficiency increases when approaching values of *LF* to 100% or higher. Interestingly, energy efficiency keeps increasing for over-loaded plants (in the range under assessment). Note, however, that malfunctions are likely to occur in severely over-loaded plants, leading to effluent quality deterioration and non-compliance with effluent requirements. A possible explanation is that in, general, design guidelines propose over-dimensioned WWTP designs. For example, Corominas et al. (2010) calculated that the aerobic volume could be reduced by 35% compared to the design of Metcalf and Eddy (2003) without affecting the design effluent requirements, and in Benedetti et al. (2010) the volumes obtained with the German Standard ATV design guidelines were reduced up to 60% of its original volume.

Dilution. A factor that negatively affects energy efficiency is the influent dilution (*DF*) for example deriving from rainwaters and/or infiltrations; this effect is highly statistically significant in the two models. It strongly supports the hypothesis that plants receiving more diluted wastewater require more energy per mass of pollutant removed, even at equal pollutant loadings, caused by, e.g. pumping greater volumes of wastewater.

Technology. The type of secondary treatment can impact on the energy efficiency (Fig. 2).

INSERT FIGURE 2 ABOUT HERE

TF is the less energy intensive technology in comparison with BNR. Tricking filter's low energy consumption is the result of a simpler operation not requiring mixed liquor inventory control and sludge wasting. As a drawback, the produced effluent has higher turbidity than activated-sludge treatment (Metcalf and Eddy, 2003). For that reason, TF are also used in combined processes with activated sludge to exploit the benefits of both processes. However, based on our results this configuration (*TFAS*) in not significantly different from BNR in terms of efficiency. It is interesting to note that BNR systems show extremely various results, including some very efficient WWTPs (red crosses in Fig. 2). This could be due to the fact that BNR category includes different configurations such as plug flow, step feed, LE, MLE, etc. Among all the technologies, MBR has the lowest energy efficiency due to intensive membrane aeration rates required to manage the fouling and clogging (Verrecht et al., 2008). Finally, a statistically significant and positive coefficient (i.e. negative effect on energy efficiency) was found for those plants that besides secondary carry out also tertiary treatment (an additional function) due to the additional energy consumption due to filtration or UV disinfection.

Geographical location. After controlling for the plant-specific heterogeneity (e.g. size, influent dilution and load factor, as well as the technology), it results interesting to investigate whether additional differences exist among countries. Our results suggest that these differences are present and are highly statistically significant. A plant located in Spain or Italy is on average less efficient than a plant located in Switzerland, which resulted as the most efficient country in our sample. This result is in accordance with the findings of Wett et al. (2007) who reported a 38% energy consumption reduction as a result of the effort carried out in Switzerland for the development of detailed energy management manuals. Additionally, it supports the hypothesis that policies for energy efficiency and benchmark initiatives are excellent measures to improve energy performance of WWTPs. However, testing adequately this hypothesis would require a representative and randomly selected subset of the WWTPs' population in the different countries.

It is worth nothing that Switzerland is the only country where all the trickling filter plants are located. As a result, in M1 the variable *COUNTRY* partially captured the effect of *TF*. *TF* in M1 has the correct sign (negative as in M2) but is not significant because its positive effect (e.g. lower energy use) is already controlled by *COUNTRY*.

Temperature. We finally found a negative and highly significant relationship of *TEMP* with the energy efficiency (Model 2). On the one hand increasing the temperature increases the biological activity, both the substrate uptake rate as the endogenous respiration. On the other hand, oxygen solubility decreases sharply when increasing temperature, leading to a higher energy demand for aeration. It is difficult to conclude which of these effects prevail. The results suggest that, in the analysed range, the higher aeration energy demand may be more significant. Although the decreasing efficiency with temperature would partially explain the lowest energy efficiency of Spanish or Italian WWTPs, since this correlation does not imply causation future studies are needed to investigate these differences among countries.

5.2. Impact of exogenous factors on estimated energy efficiency level

Fig. 3 represents the energy efficiency estimates for the WWTPs under analysis resulting from the bias-correction procedure.

INSERT FIGURE 3 ABOUT HERE

Keeping the notation used previously let Z be the vector of exogenous factors that impact the WWTP energy efficiency. In an input oriented framework (like in this study), a favourable Z means that the exogenous variable operates as a sort of an 'extra' output freely available. For this reason the exogenous factors may be considered as 'favourable' to the WWTP. Controlling for the exogenous factors will decrease the efficiency of plants operating under favourable conditions (e.g. bigger plants, operating under high values of *LF,* and low values of *DIL)* such as WWTP 180 or 387 (Fig. 3). On the contrary, an unfavourable Z means that the exogenous variable acts as a 'compulsory' or unavoidable output to be produced as a result of the 'negative' environmental condition. In other words, Z penalizes the removal of pollutants during wastewater process by increasing the amount of energy needed. In this situation, controlling for the exogenous factors will increase the efficiency of plants operating under unfavourable conditions, such as the WWTP 17 or 21 (Fig. 3). Finally, the exogenous factors can have no impact on the efficiency or favourable and unfavourable conditions can exist at the same time, cancelling out positive and negative impacts. In this case the efficiency will not change after controlling for the exogenous factors. This is the case for example of WWTP 5 or 113.

It is clear from the results of this study that estimates of efficiency are conditional on the given exogenous factors and the technology used. A WWTP may appear inefficient for one technology, but it could be efficient with respect to a different technology. The implication for empirical analysis is that, when estimating the technical/operational inefficiencies of plants operated under different treatment technologies, it should be done with respect to the appropriate technology. For example, if we compare MBRs and BNRs together there might be unobserved or unknown differences in technology. In such circumstances, the differences in technology might be inappropriately labelled as inefficiency if such variations in technology are not taken into account, as done using the two-stage approach.

6. Conclusions

The growing number of applications of DEA in wastewater treatment must be accompanied by a rigorous approach in the selection of inputs and outputs according to the benchmarking objective and a sound treatment of the exogenous factors. The REED methodology described in this manuscript is meant to guide operators, plant managers, and engineers through all the steps required to correctly use DEA for comparison of energy efficiency of WWTPs.

The use of two-stage DEA to tackle the impact of the different characteristics and environmental conditions of WWTPs leads to a larger pool of open choices for the user, potentially leading to non-comparable results. By systematizing the selection criteria and offering guidance to the reader through the different choices, REED leads to robust energy efficiency quantification at WWTPs, thereby increasing the quality of the efficiency estimates and hence the effectiveness of benchmarking. Providing explicit details about the correct application of DEA for energy efficiency quantification in the REED methodology is therefore essential for clarity, transparency, and future reproducibility.

The case study demonstrates that adjusting for the effect of exogenous factors can lead to substantial changes in efficiency estimates since they can be altered up to ± 50% compared to a single-stage DEA depending on the adverse or favourable environmental conditions a WWTP is operating, hence suggesting that given the characteristics of wastewater treatment sector the inclusion of exogenous factors in the benchmarking process by the two-stage approach is required to obtain meaningful results.

Acknowledgments

The authors belong to the Galician Competitive Research Group GRC2013-032 and the CRETUS strategic partnership (AGRUP2015/02), co-funded by FEDER (EU). Besides, they are supported by ‘ENERWATER’ Coordination Support Action that has received founding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 649819.

References

Álvarez I, Barbero J, Zofio Prieto JL. A Data Envelopment Analysis Toolbox for MATLAB - Working Papers in Economic Theory 2016/03, Universidad Autónoma de Madrid; 2016.

Andersen P, Petersen NC. A procedure for ranking efficient units in data envelopment analysis. Manage Sci 1993;39(10):1261-1264.

Bădin L, Daraio C, Simar L. Explaining inefficiency in nonparametric production models: The state of the art. Ann Oper Res 2014:214(1):5-30.

Banker RD, Cooper WW, Seiford LM, Zhu J. Returns to scale in DEA. In: Cooper WW, Seiford LM, Zhu J, editors. Handbook on data envelopment analysis, NewYork: Springer Science & Business Media; 2011.

Benedetti L, De Keyser W, Nopens I, Vanrolleghem PA. Probabilistic modelling and evaluation of wastewater treatment plant upgrades in a water quality based evaluation context. J Hydroinf 2010;12(4):380-395.

Benedetti L, Dirckx G, Bixio D, Thoeye C, Vanrolleghem PA. Environmental and economic performance assessment of the integrated urban wastewater system. J Environ Manage 2008;88(4):1262-1272.

Bertoldi P, Castellazzi L, Oikonomou V, Fawcett T, Spyridaki NA, Renders N, Moorkens I. How is article 7 of the Energy Efficiency Directive being implemented? An analysis of national energy efficiency obligation schemes. ECEEE 2015 Summer Study proceedings 2015:455-465.

Carvalho P, Marques RC. The influence of the operational environment on the efficiency of water utilities. J Environ Manage 2011;92(10):2698-2707.

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

Cook WD, Harrison J, Imanirad R, Rouse P, Zhu J. Data envelopment analysis with nonhomogeneous DMUs. Oper Res 2013;61(3):666-676.

Cook WD, Seiford LM. Data envelopment analysis (DEA)–Thirty years on. Eur J Oper Res 2009;192(1):1-17.

Cooper WW, Seiford LM, Zhu J. Handbook on data envelopment analysis. NewYork: Springer Science & Business Media; 2011.

Corominas L, Flores-Alsina X, Muschalla D, Neumann MB, Vanrolleghem PA. Verification of WWTP design guidelines with activated sludge process models. Proceedings of the Water Environment Federation 2010;(18):137-146.

Corominas L, Foley J, Guest J, Hospido A, Larsen H, Morera S, Shaw A. Life cycle assessment applied to wastewater treatment: state of the art. Water Res 2013;47(15):5480-5492.

Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32. O J 2012;315:1-56.

Doherty E, McNamara G, Fitzsimons L, Clifford E. Design and implementation of a performance assessment methodology cognisant of data accuracy for irish wastewater treatment plants. J Clean Prod 2017;165:1529-1541.

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

Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS, Shale EA. Pitfalls and protocols in DEA. Eur J Oper Res 2001;132(2):245-259.

European Commission. Proposal for a directive of the European Parliament and of the Council on the quality of water intended for human consumption (recast). The European Commission 2018. http://ec.europa.eu/environment/water/water-drink/pdf/revised_drinking_water_directive.pdf

European Environment Agency. Urban Waste Water Treatment. https://www.eea.europa.eu/data-and-maps/indicators/urban-waste-water-treatment/urban-waste-water-treatment-assessment-3 [accessed 23 November 2017].

ENERWATER. Study of published energy data. ENERWATER Project, Deliverable D2.1; 2015. http://www.enerwater.eu/download-documentation/ [accessed 01 December 2017].

ENERWATER. ENERWATER methodology. ENERWATER Project, Deliverable D3.4; 2018. http://www.enerwater.eu/download-documentation/ [January 03 December 2018].

Fang LL, Valverde-Pérez B, Damgaard A, Plósz BG, Rygaard M. Life cycle assessment as development and decision support tool for wastewater resource recovery technology. Water Res 2016;88:538-549.

Fernández-Arévalo T, Lizarralde I, Grau P, Ayesa E. New systematic methodology for incorporating dynamic heat transfer modelling in multi-phase biochemical reactors. Water Res 2014;60:141-155.

Fuentes R, Torregrosa-Martí T, Hernández-Sancho F. Productivity of wastewater treatment plants in the Valencia Region of Spain. Util Policy 2017;46:58-70.

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

Guerrini A, Romano G, Indipendenza A. Energy Efficiency Drivers in Wastewater Treatment Plants: A Double Bootstrap DEA Analysis. Sustainability 2017;9(7):1126.

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

Gurevitch J, Koricheva J, Nakagawa S, Stewart, G. Meta-analysis and the science of research synthesis. Nature 2018 (In Press); https://doi.org/10.1038/nature25753

Harell F. Regression modeling strategies. Guildford: Springer; 2001.

Huijbregts MA, Rombouts LJ, Hellweg S, Frischknecht R, Hendriks AJ, van de Meent D, et al. Is cumulative fossil energy demand a useful indicator for the environmental performance of products? Environ Sci Technol 2006;40(3):641-648.

Johnson AL, McGinnis LF. Outlier detection in two-stage semiparametric DEA models. Eur J Oper Res 2008;187(2):629-635.

Krampe J. Energy benchmarking of South Australian WWTPs. Water Sci Technol 2013;67(9):2059-2066.

Kutner MH, Nachtsheim C, Neter J, Li W. Applied linear statistical models. New York: McGraw-Hill/Irwin; 2004.

Lazzaretto A, Tsatsaronis G. SPECO: A systematic and general methodology for calculating efficiencies and costs in thermal systems. Energy 2006;31(8-9):1257-1289.

LeSage JP. Applied Econometrics Using Matlab, Department of Economics, University of Toledo; 1999. https://www.spatial-econometrics.com/html/mbook.pdf [accessed 05 December 2017].

Liu JS, Lu LY, Lu W. Research fronts in data envelopment analysis. Omega 2016;58:33-45.

Longo S, Lema JM, Mauricio-Iglesias M, Hospido A. Measuring Energy Demand and Efficiency at WWTPs: An Econometric Approach. In: Mannina G, editor. Frontiers in Wastewater Treatment and Modelling. FICWTM 2017. Lecture Notes in Civil Engineering, 4; 2017.

Longo S, d'Antoni BM, Bongards M, Chaparro A, Cronrath A, Fatone F, et al. Monitoring and diagnosis of energy consumption in wastewater treatment plants. A state of the art and proposals for improvement. Appl Energy 2016;179:1251-1268.

Lorenzo-Toja Y, Vázquez-Rowe I, Chenel S, Marín-Navarro D, Moreira MT, Feijoo G. Eco-efficiency analysis of Spanish WWTPs using the LCA + DEA method. Water Res 2015;68(0):651-666.

Metcalf L, Eddy HP. Wastewater engineering: treatment and reuse. New York: McGraw Hill; 2003.

Picazo-Tadeo AJ, Sáez-Fernández FJ, González-Gómez F. The role of environmental factors in water utilities' technical efficiency. Empirical evidence from Spanish companies. Appl Econ 2009;41(5):615-628.

Rodriguez-Garcia G, Molinos-Senante M, Hospido A, Hernández-Sancho F, Moreira MT, Feijoo G. Environmental and economic profile of six typologies of wastewater treatment plants. Water Res 2011;45(18):5997-6010.

Ruggiero J. Impact assessment of input omission on DEA. Int J Inf Technol Decis Mak 2005;4(03):359-368.

Silva C, Quadros S, Ramalho P, Alegre H, Rosa M. J. Translating removal efficiencies into operational performance indices of wastewater treatment plants. Water Res 2014;57:202-214.

Simar L, Wilson PW. Estimation and inference in two-stage, semi-parametric models of production processes. J Econ 2007;136(1):31-64.

Verrecht B, Judd S, Guglielmi G, Brepols C, Mulder J. An aeration energy model for an immersed membrane bioreactor. Water Res 2008;42(19):4761-4770.

Wang H, Yang Y, Keller AA, Li X, Feng S, Dong Y, et al. Comparative analysis of energy intensity and carbon emissions in wastewater treatment in USA, Germany, China and South Africa. Appl Energy 2016;184:873-881.

Wett B, Buchauer K, Fimml C. Energy self-sufficiency as a feasible concept for wastewater treatment systems. In: IWA Leading Edge Technology Conference, Singapore. London: IWA Publishing; 2007.

Table 1. Descriptive statistics for the dataset used. Notes: Variables are estimated on the daily basis. The reference for categorical variables is the most common value (mode).

Variable	Definition	Obs.	Mean	SD	Min	Max
Input						
E	Electricity consumption (kWh)	399	2271	4628	18.58	36653
Outputs						

COD	COD removed (kg)	399	2414	5659	2.694	58318
N	N removed (kg)	399	145.9	365.9	0.089	4098
P	P removed (kg)	399	27.17	65.18	0.003	704.5
Exogenous categorical variables						
COUNTRY (Ref = Switzerland)						
FRA	France	19	/	/	/	/
DEU	Germany	79	/	/	/	/
ITA	Italy	15	/	/	/	/
ESP	Spain	111	/	/	/	/
SECONDARY (Ref = Conventional activated sludge)						
EA	Extended aeration	150	/	/	/	/
MHLOAD	Medium/high rate activated sludge	25	/	/	/	/
MBR	Membrane bioreactor	9	/	/	/	/
OD	Oxidation ditch	18	/	/	/	/
TF	Tricking filter	20	/	/	/	/
TFAS	Tricking filter-activated sludge	5	/	/	/	/
TERTIARY (Ref = No tertiary treatment)						
YES	Filtration or UV disinfection	41				
Exogenous continuous variables						
SIZE	Actual plant size (PE)	399	21381	50164	23.91	507511
LF	Load factor (%)	399	71.80	59.26	4.192	782.5
DF	Dilution factor (L/PE·d)	399	380.0	380.4	61.70	3060
TEMP	Temperature (°C)	399	12.06	3.229	9.500	18.10

Table 2. Estimated WWTP energy efficiency function.

Variable	Model 1		Model 2	
	Coefficient	t-statistic	Coefficient	t-statistic
Constant	4.0826***	10.2811	5.0668***	21.2251
COUNTRY				
FRA	0.3287	0.5173	/	
DEU	0.6811	1.4258	/	
ITA	2.4044***	3.1881	/	
ESP	1.7731***	3.7307	/	
SECONDARY				
EA	0.6438	1.3879	0.1254	0.3146
MHLOAD	0.8645	1.3977	0.2626	0.4828

MBR	2.7999***	3.1453	2.5417***	2.8789
OD	-0.0288	-0.0404	0.3605	0.5395
TF	-0.8467	-1.2100	-1.4523**	-2.2565
TFAS	-0.9157	-0.7746	-1.5247	-1.3265
TERTIARY				
YES	1.3685***	2.7623	1.4736***	3.0299
SIZE	-1.8492***	-10.3355	-1.8376***	-10.3004
LF	-0.8301***	-5.1393	-0.7799***	-4.7580
DL	0.4077**	2.0584	0.4085**	2.0413
TEMP	/		0.6835**	3.6467
σ^2	6.2159		6.3093	
Log-Likelihood	-921.4066		-924.3264	

Note: FRA = France; DEU = Germany; ITA = Italy; ESP = Spain. MBR = membrane bio-reactors; EA = extended aeration; TFAS = tricking filter-activated sludge; MHLOAD = medium/high loading rate activated sludge; OD = oxidation ditch; TF = tricking filter.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

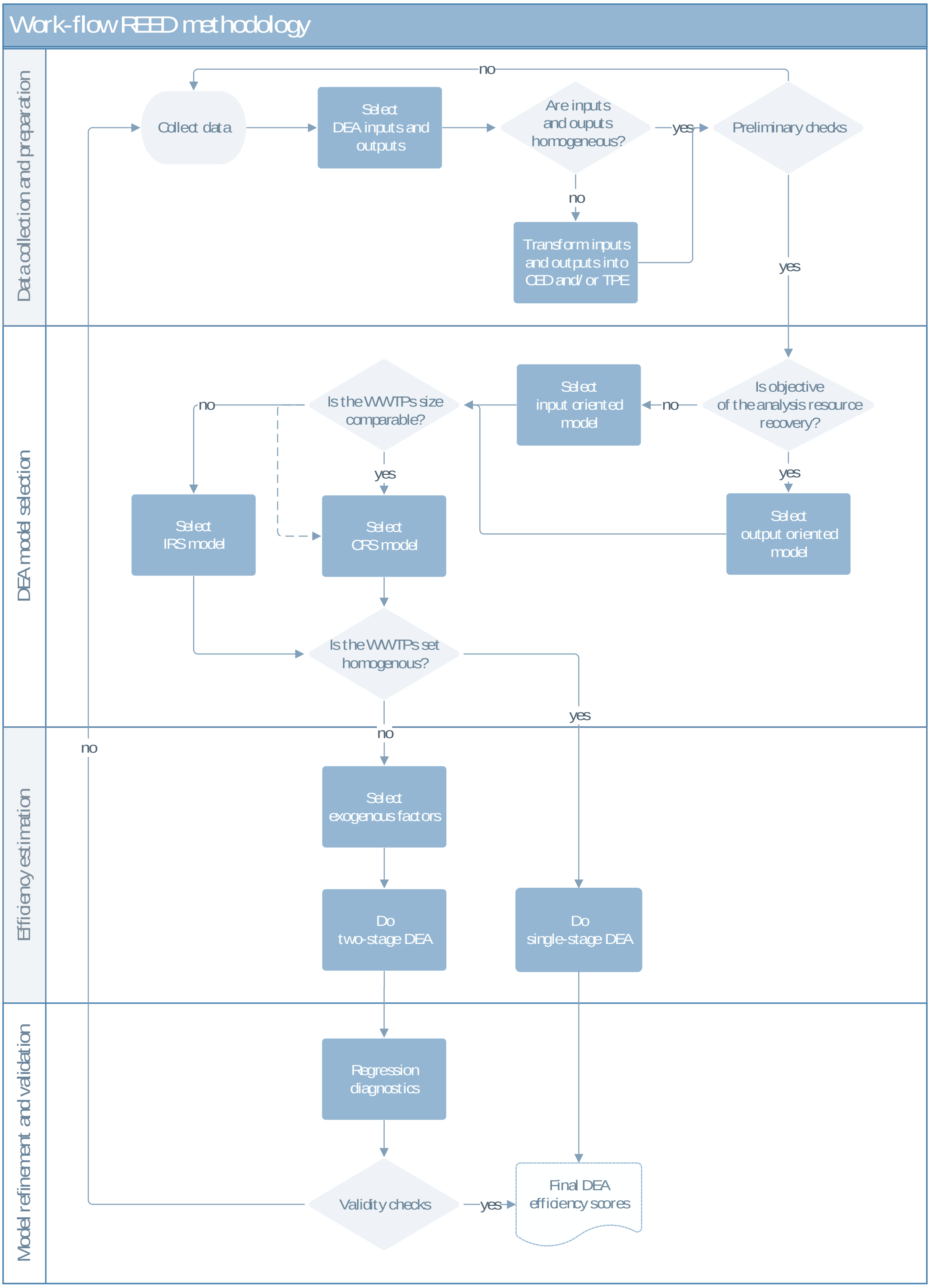

Fig. 1. REED methodology decision guidance flowchart for WWTP energy efficiency determination using DEA.

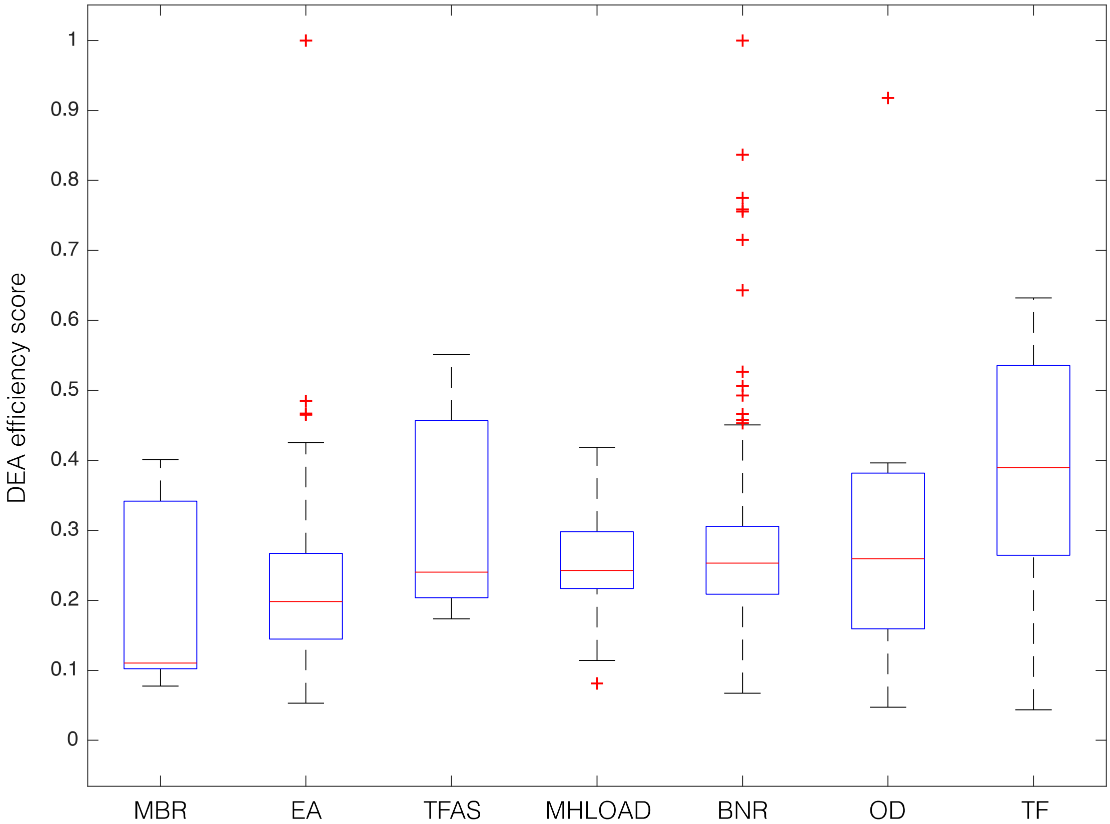

Figure 2. Energy efficiency for different treatment technologies. Note: MBR = membrane bio-reactors; EA = extended aeration; TFAS = tricking filter-activated sludge; MHLOAD = medium/high loading rate activated sludge; BNR = biological nutrient removal; OD = oxidation ditch; TF = tricking filter.

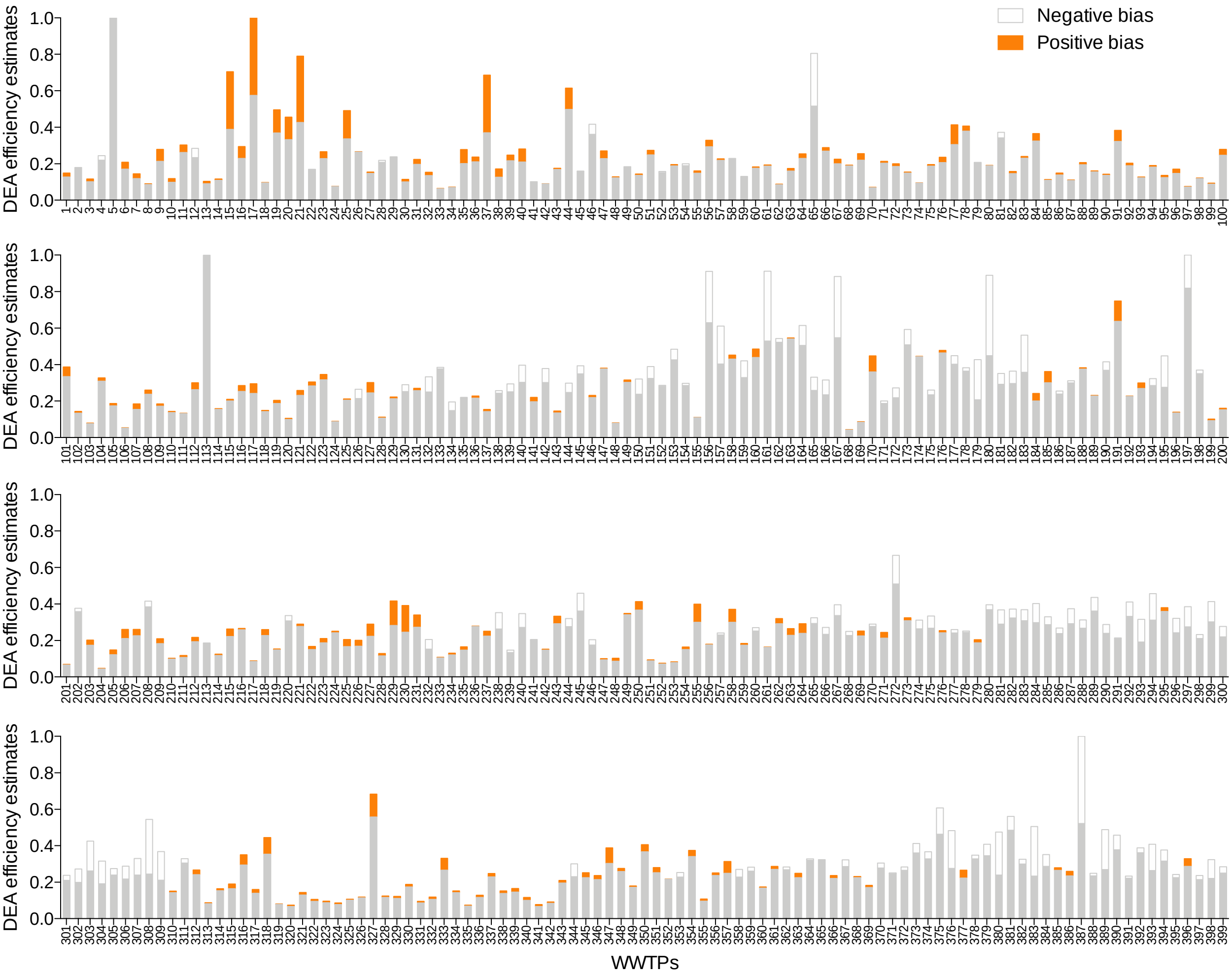

Figure 3. Bias-corrected efficiency estimates. Note: grey bars indicate original single-stage DEA scores; orange bars indicate positive bias (increase of the efficiency) and grey empty bars indicate negative bias (reduction of the efficiency). Full bars, independently of the colour, represent the bias-corrected final DEA efficiency.