Author's Accepted Manuscript

www.elsevier.com/locate/omega

The quality of service: An overall performance assessment for water utilities

F.S. Pinto, A.S. Costa, J.R. Figueira, R.C. Marques

PII: S0305-0483(16)30491-1
DOI: http://dx.doi.org/10.1016/j.omega.2016.08.006
Reference: OME1700

To appear in: *Omega*

Received date: 8 March 2016
Revised date: 4 August 2016
Accepted date: 7 August 2016

Cite this article as: F.S. Pinto, A.S. Costa, J.R. Figueira and R.C. Marques, The quality of service: An overall performance assessment for water utilities, *Omega* http://dx.doi.org/10.1016/j.omega.2016.08.006

The quality of service: An overall performance assessment for water utilities☆

F. S. Pinto[a,*], A. S. Costa[b], J. R. Figueira[b], R. C. Marques[a]

[a]*CERIS, CESUR, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001, Lisboa, Portugal*
[b]*CEG, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal*

Abstract

Over the last decades, the delivery of water supply services has changed among regions and actors, in order to meet the changing needs. Those services can be balanced based on the relation between the obligations they must fulfill, the quality offered, and the associated price/value for money. Furthermore, due to the water services inherent characteristics, they are prone to poor quality of service (in general terms). Thus, global performance assessments are of paramount relevance under those constraints. In Portugal, the water regulator has developed a system of performance indicators (partial measures of performance) and, therefore, it is not possible to achieve the desired holistic performance evaluation (global measure). To enable such assessment, we propose an application based on the ELECTRE TRI-nC method to define quality of service categories and aggregate performance indicators. To obtain a coherent family of criteria (*i.e.*, exhaustive, cohesive, and non-redundant) we apply an iso-preference logic. The results obtained are presented through a geographical information system allowing for a clearer visualization of the overall performance of water utilities. The method proposed can be considered a suitable decision support system and useful regulatory tool, able to provide policy relevant outputs.

Keywords: Multicriteria, Decision support systems, Policy analysis, Quality of service, Water supply services, ELECTRE TRI-nC

1. Introduction

The delivery of water supply services has changed over time, distinctively across regions and actors, in order to meet the changing needs of customers, citizens and society [1, 2]. Due to their very nature, these services can be balanced based on the relation between the obligations they must fulfill, the quality offered, and the associated price/value for money, to achieve a social and economic cohesion [3]. This way even if their scope (*i.e.*, activities included in the delivery) and organization may vary according to history and cultures of government intervention, the management of those services should follow the principles of transparency, nondiscrimination and proportionality.

☆Case study: Portugal.

*Corresponding author. Tel.: +351 218 418 305.
Email addresses: frcsilvapinto@tecnico.ulisboa.pt (F. S. Pinto), anasaracosta@tecnico.ulisboa.pt (A. S. Costa), figueira@tecnico.ulisboa.pt (J. R. Figueira), rui.marques@tecnico.ulisboa.pt (R. C. Marques)

Since they also present an economic interest, they should be further ruled by the principles of quality, safety and affordability, equal treatment, the promotion of universal access and of user rights, independently of geographical, social or cultural situations [4]. However, generally, water supply services operate in a natural monopoly basis, requiring huge investments characterized by being long-lived and indivisible [5], and present asymmetric information among stakeholders [6]. Indeed, those features render the water services liable to poor quality of service (QS), namely to inappropriate use of resources (*e.g*., natural, human, and capital), insufficient cost recovery (heightened by high levels of non-revenue water), inadequate/ disrupted water supply, and poor water quality, among others [3]. Under those circumstances, benchmarking assumes a key role in the creation of incentives for efficiency and compliance with obligations [7]. This process involves a comparison of business processes and performance measures, which are usually based on the use of Performance Indicators (PI) to promote the desired virtual competition. A key point on this regulatory process based on PI is allowing for the public disclosure and discussion of results, promoting a 'name and shame' approach (sunshine regulation, see Marques and Simões [8]). This way, poor-performing water utilities will be motivated to improve their QS due to expected pressures on undesired performances (applied by stakeholders).

The use of PI has been widely fostered worldwide [9], either from an operational perspective [10, 11], requiring an exhaustive application, or from a regulatory point of view, compelling a condensed but still comprehensive approach [12, 13]. However, from a regulatory perspective the usage of several indicators, hampered by their differences in importance, does not allow to suitably assess the problem globally.

In Portugal, the water and waste services regulator (ERSAR) has successfully implemented a Performance Assessment (PA) system. However, even after a revision (reduced from 20 to 16 PI) the desired holistic QS evaluation remains to be achieved. A possible way to solve such hindrances is to aggregate PI by using some Multiple Criteria Decision Analysis (MCDA) methodologies, still, due to the nature of such methods, caution must be taken in their selection procedure, making sure that the chosen decision aiding method is adequate to the specific context [14, 15]. In related literature, there are studies that assess the suitability of particular PI to be included in the overall evaluation of water utilities [16–18]. Despite that, when considering the methods applied, such research cover applications that range from the Analytic Hierarchy Process (AHP) [19], to Goal Programming Synthetic Indicator (GPSI) or Principal Component Analysis (PCA) [20], as well as to the aggregation of criteria modelled through a 'fuzzy related' method [21]. The last one assumes an increased relevance as it covers the first generation of the ERSAR PA. However, there was a requirement to set the PA based on QS levels, built from characteristic reference profiles (*i.e.*, actions), while performing a suitable evaluation [22]. Such evaluation has to intuitively allow for judgements without converting the original scales into specific abstract ones (with an arbitrary imposed unit and range), while developing a suitable classification with predefined QS categories under a non-compensatory outranking preference model [23]. The mentioned method should also be able to take imperfect knowledge into account and avoid an extensive interplay [24], acknowledging that different methods will bear distinct results [25].

In order to comply with all the previous demands while assuring a requisite model (in the sense of a model 'whose form and content are sufficient to solve a particular problem' [26, p. 29]) we established an innovative QS Index

(QSI). The QSI has both methodological and practical novelties as we propose an application based on the ELECTRE TRI-nC method, where there is a previous modelling in order to aggregate PI according to their nature (accounting for preference information from the decision makers) and then avoid systematic compensation among criteria with different fundamental nature (*e.g.,* financial and environmental). The practical novelties are related to being the first application that is able to comply with the demands highlighted in the previous paragraph, bearing in mind the possibility of enabling an improved regulation of water supply services. Indeed, this method allows to evaluate and classify the Portuguese water utilities into different categories, i.e., distinct levels of QS, taking into account multiple criteria. The main features of applying this method are: (1) the absence of systematic compensation between 'good performances' and 'poor performances'; (2) the use of discriminant thresholds to consider the imperfect knowledge of data and arbitrariness when constructing the set of criteria; (3) the consideration of positive and negative reasons in the modelling of preferences, and also the possibility of using veto thresholds, which reinforces the non-compensatory character of the method (*cf.* Figueira et al. [27]).

This paper is organized as follows. Section 2 is devoted to the description of the method used (ELECTRE TRI-nC). Section 3 presents the case-study's decision context, highlighting the model's structure. Section 4 deals with the focus group and the developed interaction. Later in Section 5, the results achieved are assessed through sensitivity and robustness concerns. To conclude, Section 6 outlines the managerial implications drawn and Section 7 presents some concluding remarks.

2. The ELECTRE TRI-nC method

In this section, we present the ELECTRE TRI-nC method [22], a multiple criteria sorting based method of the ELECTRE family methods [28], which were designed into two main phases: (1) construction of one or several outranking relations, aimed at comparing each pair of actions; and (2) exploitation of these outranking relations (for more details, see [29]). This method follows a decision aiding constructive approach, through the interaction between the analyst(s) and the decision-maker(s) (DMs).

2.1. Concepts, definitions and notation

In what follows, let $A = \{a_1, ..., a_i, ...\}$ denote a set of *potential actions,* and g a given criterion. A *coherent family* of criteria, $F = \{g_1, ..., g_j, ..., g_m\}$, is defined in order to evaluate the potential actions to be assigned to a certain category [29]. A set of *completely ordered categories,* $C = \{C_1, ..., C_h, ..., C_q\}$, is defined, in which C_1 is the worst category and C_q is the best one, with $q \geqslant 2$. Let $B = \{B_1, ..., B_h, ..., B_q\}$ denote a set of *characteristic reference actions* that define the categories, and $B_h = \{b_h^r, r = 1, \ldots, m_h\}$ the subset of the reference actions that define the category C_h, such that $m_h \geqslant 1$ and $h = 1, \ldots, q$ (at least one m_h is greater than one). Note that if $m_h = 1$, for all h, then the considered method is the ELECTRE TRI-C [30]. Generally, the ELECTRE TRI-nC method takes into account more than one reference action to characterize each category. The two particular subsets of reference actions, denoted $B_0 = \{b_0^1\}$ and $B_q = \{b_{q+1}^1\}$, contain a reference action, such that $g_j(b_0^1)$ is the worst possible performance on criterion g_j, and $g_j(b_{q+1}^1)$ is the best possible performance on the same criterion g_j, for all $g_j \in F$, respectively. The worst and the best possible performances must be chosen such that, for any action a, one has $g_j(b_0^1) < g_j(a) < g_j(b_{q+1}^1)$, for all $g_j \in F$.

The imperfect character of the data from the computation of the performances $g_j(a)$, $\forall a \in A$ and $\forall g_j \in F$, and the arbitrariness that can affect the definition of the criteria, are taken into account by introducing two thresholds, the preference threshold (*cf.* Definition 1) and the indifference threshold (*cf.* Definition 2). Even though these thresholds can vary according to the performances $g_j(a)$ and $g_j(a')$, in this paper, we solely consider constant thresholds (for more details, see [31]). Let us consider that each criterion g_j will be taken as a pseudo-criterion (*cf.* Definition 3).

Definition 1 (Preference threshold). The preference threshold between two performances, denoted by p, is the smallest performance difference that when exceeded is judged significant of a strict preference in favour of the action having the best performance.

Definition 2 (Indifference threshold). The indifference threshold between two performances, denoted by q, is the largest performance difference that is judged compatible with an indifference situation between two actions having different performances.

Definition 3 (Pseudo-criterion with constant thresholds). A criterion g_j is considered as a pseudo-criterion when two thresholds are associated to g_j: an *indifference threshold*, q_j, and a *preference threshold*, p_j, such that $p_j \geqslant q_j \geqslant 0$.

Let us consider two actions a and a', where $g_j(a) \geqslant g_j(a')$, for a given criterion g_j to be maximized. The following binary relations can be derived for each criterion:

1. $|g_j(a) - g_j(a')| \leqslant q_j$ represents a non-significant advantage of one of the two actions over the other, meaning that a is *indifferent* to a' according to g_j, denoted aI_ja';
2. $g_j(a) - g_j(a') > p_j$ represents a significant advantage of a over a', meaning that a is *strictly preferred* to a' according to g_j, denoted aP_ja';
3. $q_j < g_j(a) - g_j(a') \leqslant p_j$ represents an ambiguous zone. There is a hesitation between indifference and strict preference, meaning that a is *weakly preferred* to a', denoted aQ_ja'.

In what follows, it is needed the following notation:

- $C(aIa')$ denotes the subset of criteria such that aI_ja';
- $C(aPa')$ denotes the subset of criteria such that aP_ja';
- $C(aQa')$ denotes the subset of criteria such that aQ_ja'.

According to the criterion g_j, "a outranks a'", denoted aS_ja', if "a is at least as good as a'" on such a criterion g_j. In order to construct a fuzzy outranking relation, it is necessary to introduce three concepts: concordance, non-discordance, and a degree of credibility [32].

Concordance index. A set of intrinsic weights is associated with the criteria, each one denoted by w_j, such that $w_j > 0$, for $j = 1, ..., m$, and $\sum_{j=1}^{m} w_j = 1$. The overall concordance in favour of the assertion of "a outranks a'" (denoted aSa') is modelled through a *comprehensive concordance index*, denoted $c(a, a')$, that can be defined as follows:

$$c(a, a') = \sum_{j \in C(aPa')} w_j + \sum_{j \in C(aQa')} w_j + \sum_{j \in C(aIa')} w_j + \sum_{j \in C(a'Qa)} w_j \varphi_j, \tag{1}$$

where

$$\varphi_j = \frac{p_j - (g_j(a') - g_j(a))}{p_j - q_j} \in [0, 1[. \tag{2}$$

The fraction φ_j represents the way in which the power of vote decreases according to the criteria $g_j \in C(aQa')$. It should be mentioned that there is a difference in φ_j when we use a continuous scale and when we use a discrete scale (*cf.* Roy et al. [31]).

An additional threshold, called veto threshold, can also be assigned to certain criteria. This is a different threshold in comparison to the indifference and the preference ones. The veto threshold allows to increase the power of certain criteria. When a criterion opposes strongly to the assertion "a outranks a'", such a criterion puts its veto to this assertion. For instance, when "a outranks a'" on a large majority of the criteria, the extent to which some criteria are out of this majority (or the extent to which they are strongly discordant with the assertion "a outranks a'") should be evaluated, before definitely drawing a final conclusion. Veto thresholds can be introduced to this end.

Definition 4 (Veto threshold). The veto threshold, denoted v_j, such that $v_j \geqslant p_j$, is the minimal advantage of a over a', on the criterion g_j, to exceed for being incompatible with an overall outranking, indifference, or preference of a over a'.

Discordance index. For each criterion, the veto power is taken into account by a partial discordance index, denoted by $d_j(a, a')$, $j = 1, ..., m$, which is defined as follows:

$$d_j(a,a') = \begin{cases} 1 & if\ g_j\left(a'\right) - g_j\left(a\right) > v_j, \\ \dfrac{(g_j(a') - g_j\left(a\right)) - p_j}{v_j - p_j} & if\ p_j < g_j(a') - g_j\left(a\right) \leqslant v_j, \\ 0 & if\ g_j\left(a'\right) - g_j\left(a\right) \leqslant p_j. \end{cases} \tag{3}$$

Credibility index. The credibility index reflects the way the assertion "a outranks a'" is more or less well justified when taking into account all the criteria from F, and is defined as follows:

$$\sigma(a,a') = c(a,a') \prod_{j=1}^{m} T_j(a,a'), \tag{4}$$

where

$$T_j(a,a') = \begin{cases} \dfrac{1 - d_j(a,a')}{1 - c(a,a')} & if\ d_j(a,a') > c(a,a'), \\ 1 & otherwise. \end{cases} \tag{5}$$

Categorical credibility indices. The categorical credibility indices, $\sigma(a, B_h)$ and $\sigma(B_h, a)$, can be interpreted as the *categorical outranking degrees* of action a over the subset of reference actions B_h, and the *categorical outranked degrees* of action a over the subset of reference actions B_h, respectively, and are defined as follows:

1. $\sigma(a, B_h) = \max\limits_{r=1,\ldots,m_h} \{\sigma(a, b_h^r)\}$.
2. $\sigma(B_h, a) = \max\limits_{s=1,\ldots,m_h} \{\sigma(b_h^s, a)\}$.

2.2. The assignment procedure

Let λ denote the *credibility level*, which represents the minimum degree of credibility, $\sigma(a, a')$, considered needed by the DM(s) to validate (or not) the assertion "a outranks a'" taking all the criteria from F into account (λ usually takes a value within the range $[0.5, 1]$). In order to preserve the role of the characteristic reference actions, ELECTRE TRI-nC uses a selecting function, $\rho(a, B_h)$, which allows to choose between the two consecutive categories where an action a can be assigned to. This function can be defined in several ways. In this paper, we will make use of the following one:

$$\rho(a, B_h) = min\left\{\sigma(a, B_h), \sigma(B_h, a)\right\}. \tag{6}$$

The ELECTRE TRI-nC assignment procedure is composed of two joint rules, the descending rule (*cf.* Definition 4) and the ascending rule (*cf.* Definition 5), which must be conjointly used. These two rules allow to assign a category or a set of possible categories to an action a, which is compared with the reference actions B_h, considering the chosen credibility level λ.

Definition 5 (Descending rule). Choose a credibility level, λ $(0.5 \leqslant \lambda \leqslant 1)$ and decrease h from $(q+1)$ until the first value, t, such that $\sigma(a, B_t) \geqslant \lambda$:

1. For $t = q$, select C_q as a possible category to assign action a;
2. For $0 < t < q$, if $\rho(a, B_t) > \rho(a, B_{t+1})$, then select C_t as a possible category to assign a; otherwise, select C_{t+1};
3. For $t = 0$, select C_1 as a possible category to assign a.

Definition 6 (Ascending rule). Choose a credibility level, λ $(0.5 \leqslant \lambda \leqslant 1)$ and increase h from zero until the first value, k, such that $\sigma(B_k, a) \geqslant \lambda$:

1. For $k = 1$, select C_1 as a possible category to assign action a;
2. For $1 < k < (q+1)$, if $\rho(a, B_k) > \rho(a, B_{k-1})$, then select C_k as a possible category to assign a; otherwise, select C_{k-1};
3. For $k = (q+1)$, select C_q as a possible category to assign a.

These two rules are applied conjointly, and each one of these rules selects a possible category that an action a should be assigned, the highest and the lowest possible categories. The result of the assignment of an action a can be one of the following situations: one category, if the extreme selected categories are the same; or a range of categories, delimited by the extreme categories.

3. Case study: decision context

In this section, the decision context will be described to perceive the MCDA process at hand. This perception is acquired by understanding the objectives of both the set of players responsible for the decision and those that are likely to be affected. Thus, after a brief overview of the Portuguese regulatory framework, the Portuguese water market structure is characterized to contextualize the units being evaluated, and the ERSAR PA system is outlined.

3.1. Overview

The Portuguese Government has promoted several reforms in the water sector to counter raising challenges (e.g., access to appropriate services at affordable prices, increasing quality and environmental standards). Those reforms included a reorganization of the water supply services delivery, and the creation of a sector specific regulator (IRAR) to monitor and supervise such services [33]. IRAR, set in 1998, was first designated to regulate municipal and regional (multiple municipalities) concessions. Later, in 2009, through a reform in both its statutes and regulatory model, it became a stronger regulator. This reformed agency (ERSAR), has its independence and organizational framework regulated by a Portuguese framework law (Law 67/2013 of August 28). Since its creation (2009), although only effective after 2011, ERSAR has broad regulatory competences covering all delivery models. Those regulatory activities include a wide intervention in the sector structural regulation, the regulation of operators' behaviour (*e.g.*, economic), and other additional regulatory activities (*e.g.*, public disclosure of sound information and technical support). However, as highlighted elsewhere [34], those activities require an increased maturity, as improved compliance and standardization in reporting is needed (mainly by municipal direct delivery) to enable prompt and efficient analyses. In fact, several utilities (as of 2013) did not report the demanded information by ERSAR for the annual PA, disabling the possibility to develop their overall assessment (see Section 5).

The different regulatory activities involve diversified powers of regulation, some include a coercive power (*e.g.*, water quality), others are based on advice and guidelines (*e.g.*, tariff setting). Still, the monitoring of services coupled with public display of results can greatly improve the 'payoff' of those activities [8]. The supervision of quality of service, through the application of PI, their comparison and public display, brought undoubtedly relevant outcomes [33, 35]. The public display is usually published in an yearly report series named 'Annual report on water and waste services in Portugal', with 4 volumes. The first one, presents information about the organization of the sector and the implemented regulatory model, of quantified data about the utilities, and an overview of end-users' complaints and tariffs. The remaining volumes develop a market analysis of the sectors' main economic and financial indicators, of the quality of service provided to users and of drinking water quality.

In fact, although, a positive trend should be outlined, the national situation encompasses a diversified set of realities, at different development stages and with very distinct levels of service [36].

3.2. Water supply market structure

In Portugal, the market structure is quite complex as a result of several, perhaps contradictory, reforms. Accordingly, the wholesale (bulk) and retail (end-user) activities may be devised into a unitary system or, respectively, into regional and municipal systems [34]. In this research, since we want to perform an overall PA to allow dissemination of sound information for improved stakeholder engagement, we will focus on those that perform, at least, the retail segment. The main key institutional features enclose the role of the water regulatory authority, ERSAR, and of the central government, as a relevant operator.

The Portuguese water supply utilities vary in accordance with numerous characteristics, such as the management model and scale of operation. The main characteristics of the water supply retail providers are highlighted in Table 1.

The number of retail providers is indeed very large. However, the water supply services provided by Parishes and associations has a clear marginal role which is continuously decreasing [35]. Besides, this type of management model is mostly found providing water supply in municipalities jointly with other suppliers (see in Table 1, as in Portugal mainland there are only 278 municipalities for the 299 instances covered), which may highlight in some cases the need for their existence (*e.g.*, in secluded locations), or in other cases a clear redundancy (at best). Still, those entities

Table 1: Market structure of the Portuguese water retail sector (adapted from [35])

Management model	Entities (nr.) [%]	Municipalities covered (nr.)	Area (km^2)	Pop. covered^a (1000 inhab.) [%]	Pop. density^a (inhab./km^2)
Regional concessionaire	1 [0.3]	1	75	11 [0.1]	142
Municipal concessionaire	28 [8.1]	33	7609	1971 [19.5]	259
State delegation	1 [0.3]	1	85	548 [5.4]	6446
State/municipalities partnership	1 [0.3]	10	1476	340 [3.4]	230
(Inter)Municipal companies	23 [6.6]	28	8896	1793 [17.8]	202
Parishes/User associations	85 [24.5]	16	1 362	56 [0.6]	41
(Inter)Municipal service with autonomy	20 [5.8]	22	7690	2316 [23]	301
Municipal department	188 [54.2]	188	62391	3048 [30.2]	49
Total	347	299	–	10083	–

^{a}- The values specify the potentially covered population.

are not targeted by the ERSAR PA system and, thus, we did not consider them; pragmatically, their context is quite different and the lack of information is daunting.

The water supply is significantly provided directly by the municipalities (*i.e.*, through municipal departments or municipal services with autonomy) covering 70% of them and 53% of the potential population [35]. The remaining models are mostly found in coastal areas or wide urban centers. The corporatization of those services is an increasing trend, as since 2000, the population supplied by corporate models of service delivery grew from 20% to more than 45% in 2014 [35].

From a general point of view, the municipalities opt to provide the services by themselves (either directly or through externalization of the service), as 91% of municipalities covering 82% of the population is not integrated with other municipalities. In terms of process integration or verticalization, most municipalities have the service vertically unbundled (65%) and 11% of the municipalities have the service partially integrated. Furthermore, in terms of scope, nearly all municipalities provide the retail water supply service jointly with the retail wastewater service (95%) [35]. To a far less extent, the municipalities decide to provide urban solid waste services and, even less frequently, other services of general interest, such as urban transportation.

3.3. *The ERSAR PA system*

PI tools are recommended to measure/evaluate the utility's effectiveness and efficiency regarding the service provided. In this sense, the desired quality is obtained by reaching defined standards/levels. In Portugal, ERSAR decided to perform such assessment through a set of PI selected from a wider PI database [9]. Due to its regulatory function, such a PA system had to be as comprehensive as possible, and thus, twenty indicators were carefully chosen incorporating three main dimensions (*i.e.*, protection of user interests, operator sustainability, and environmental sustainability [37]). However, since 2005, this PA tool has been scrutinized and was later updated to its current version of sixteen PI [38].

In the ERSAR PA system each PI is qualitatively assessed through three different QS levels: the 'unsatisfactory', the 'acceptable', and the 'good'. This way, each utility has a basis to assess and compare its QS PI, their evolution over time and the improvements required to achieve the defined standards, or from a similar perspective, proposed targets. Likewise, as mentioned elsewhere, a possible integrated evaluation would allow to establish an overall assessment of utilities, which could 'stimulate a continuous improvement of performance, sustainability and quality of the water supply services' [21, p. 760].

The new hierarchical structure of the ERSAR PA system frames the sixteen PI into the mentioned three main dimensions/groups (which were kept from the previous twenty PI system), subdivided into seven sub-groups (*cf.* Table 2).

Each PI in Table 2 has its own set of required data with a computed relationship (*i.e.*, formulae), an inherent unit that defines the magnitude of measurement, as well as an objective function 'trend', that is, what is required to improve the result in a PI (*e.g.*, maximize, minimize, reach a specific target).

The complexity of the Portuguese water supply structure highlights the diversity between the utilities' scope (*cf.* Sub-section 3.2), this phenomenon renders the PA to hold specific PI that do not apply to some utilities, *i.e.*, there are specific activities that are not performed by all retail utilities (*e.g.*, treatment of raw water).

4. Working the focus group

In this study, we adopt a multiple criteria decision aid approach, based on the ELECTRE TRI-nC method, in order to achieve an overall PA, evaluating and classifying the Portuguese retail water utilities into different categories of QS. After achieving a requisite understanding of the problem (*cf.*, Section 3), the specified reasoning requires the intervention and interaction of a focus group able to address the subjectivity needed to attend the points of view of the different stakeholders [39]. Thus, the developed model was co-constructed through the interaction between the analysts and the focus group (interchangeably referred as DM). In this section, we provide the details about how such an interaction occurred. Also in Figure 1, we highlight the main tasks involved, as well as some important issues of such interplay.

Table 2: ERSAR PA System for retail water supply services

Code	Description	Units	IT[a]	Notes
Protection of user interests				
Accessibility of service				
AA01b	Coverage	%	max	Varies with the urban-rural typology
AA02b	Affordability of the service	%	min	–
Quality of service provided to users				
AA03b	Service interruptions	$\frac{nr}{1000c^b.year}$	min	–
AA04b	Safe water	%	max	–
AA05b	Reply to written suggestions and complaints	%	max	–
Operator sustainability				
Economic sustainability				
AA06b	Coverage of total costs	-	target	–
AA07b	Connection to the service	%	max	–
AA08b	Non-revenue water	%	min	–
Infrastructural sustainability				
AA09b	Adequacy of treatment capacity	%	max	–
AA10b	Mains replacement	$\frac{\%}{year}$	target	–
AA11b	Mains failures	$\frac{nr}{100km.year}$	min	–
Physical productivity of staff				
AA12b	Adequacy of staff	$\frac{nr}{1000c}$	target	Varies with the urban-rural typology
Environmental sustainability				
Efficient use of environmental resources				
AA13b	Water losses	$\frac{l}{c.day}$	min	–
AA14b	Fulfilment of the water intake licensing	%	max	–
AA15b	Standardized energy consumption	$\frac{Kwh}{m^3.100m}$	min	–
Efficiency in pollution prevention				
AA16b	Sludge disposal	%	max	–

[a]- IT: Improvement trend; [b]- c: Water supply service connections.

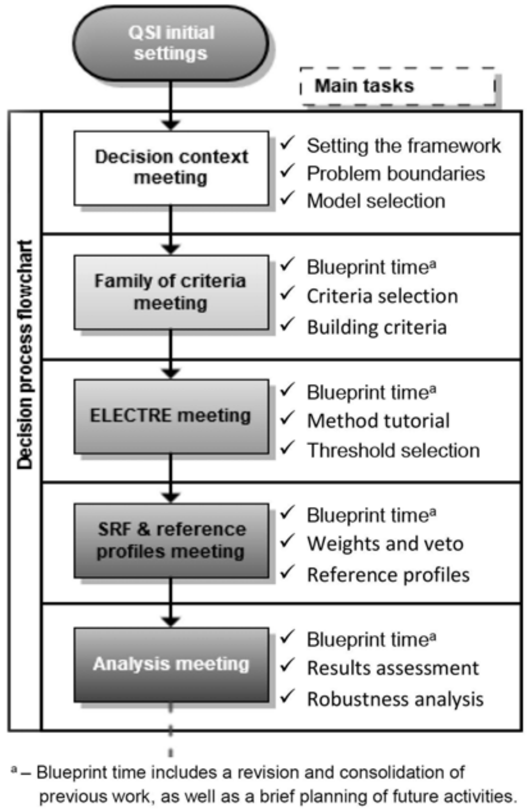


Figure 1: Flowchart of the decision aiding process

The focus group was made up by three individuals that, somehow, represented the stakeholders involved in the regulatory process, particularly, the municipalities (the owners of the infrastructure), the providers (which may or may not be the municipalities themselves), the regulator, and the customers. The first member is a senior consultant in the water sector with a vast experience in working together with the municipalities and the providers. The second individual is a previous director of ERSAR that has a wide knowledge on the institution and the PI system. Finally, the third is a young analyst that has been developing some research in governance and other social aspects of the water sector. The focus group intervened in all the meetings (decision conferences) highlighted in Figure 1, and had important contributions not only in providing preference information, but also in setting the requirements that the model had to meet. Indeed, regarding the mentioned requirements, we first proposed a value function approach, but as the systemic compensation among criteria with different fundamental nature was not accepted, we suggested an ELECTRE TRI-nC based approach.

Since we intend to propose a method able to suitably aggregate all the PI in the ERSAR's PA, aiming to use the data sets published by ERSAR as input, the structure of our problem adopts the same hierarchy (*i.e.*, dimensions/groups, sub-groups and PI). By using the sixteen PI we are able to build criteria that describe and represent the characteristics of this problem [40]. Indeed, in the construction of such criteria, the principles proposed by Keeney and Raiffa [41] were followed, such as: the required completeness, non-redundancy, minimum size, operationality, as well as decomposability.

A coherent family of criteria, selected through the interplay with the focus group (several structuring methods were used), was achieved with a set of six criteria set in accordance with their fundamental nature. Since all the criteria had specific particularities, they were modelled through iso-preference logic [42]. This way, the focus group had a considerable ease to relate iso-preference thresholds (known PI values to which there is a special comfort to assign a status of equal preference) and give them meaning. Naturally, any value in the between was determined by interpolation, in this case, linear interpolation was deemed suitable. However, the authors are aware that this situation was only possible because the focus group had some experience with MCDA methodologies and a deep knowledge on the ERSAR PA. In fact, the criteria that cover three PI had to be built upon different steps, first eliciting the PI iso-preferences in a pairwise fashion. The final figures (*e.g.*, Figure 2) with the iso-preference planes and the water utilities' scores, allowed to perform a consistency test.

Table 3 describes the criteria used in our work to evaluate the water utilities, as well as the scales of those criteria. All scales of the criteria are quantitative, and all criteria were considered to be maximized, *i.e.*, the preference increases when the criterion performance increases. The 'maximization' assumption led to the application of a negative unitary multiplier to 'minimization' cases, and to evaluate the distance to a specific point coupled with application of a negative unitary multiplier to the 'target' situations (see Table 2). The mathematical transformations performed was just to allow a swifter programming, as for the focus group interpretations they were not totally introduced, as seen in Figure 2.

Figure 2 outlines for two criteria their iso-preference modelling, bearing in mind that the criteria had to be bounded to values in which the participants could relate and infer iso-preference. An example is with the PI AA01b, where below a specific threshold the value was impossible to relate, as below specific levels of service accessibility the focus group considered that 'nothing else mattered'. See Appendix A for a pictorial description of all remaining criteria.

In this case-study, the overall evaluation should be granted to all water utilities properly rated in all PI. This universe of comparison implies that those PI which are not applicable to some utilities should not be considered in those utilities assessment. Thus, we had to introduce coefficients in some criteria as specified in Table 3. Those coefficients were modelled as a polynomial approximation able to consider the iso-preference elicitation.

If the utility did not provide all the required data they were removed from the overall assessment and classified as non-compliant. Following those rules, we were able to use data from 99 water utilities (161 utilities were rejected).

The water utilities were then evaluated according to the set of coherent criteria. The performances of our sample of water utilities were measured and then (*cf.*, Section 6) linked to a geographical information system (GIS) for an easier visualization/interpretation [43].

Table 4 presents both discriminating thresholds (indifference and preference thresholds) chosen for the criteria. Please note that the indifference and preference thresholds were introduced to take into account and modeling potential

Table 3: Details of QSI criteria

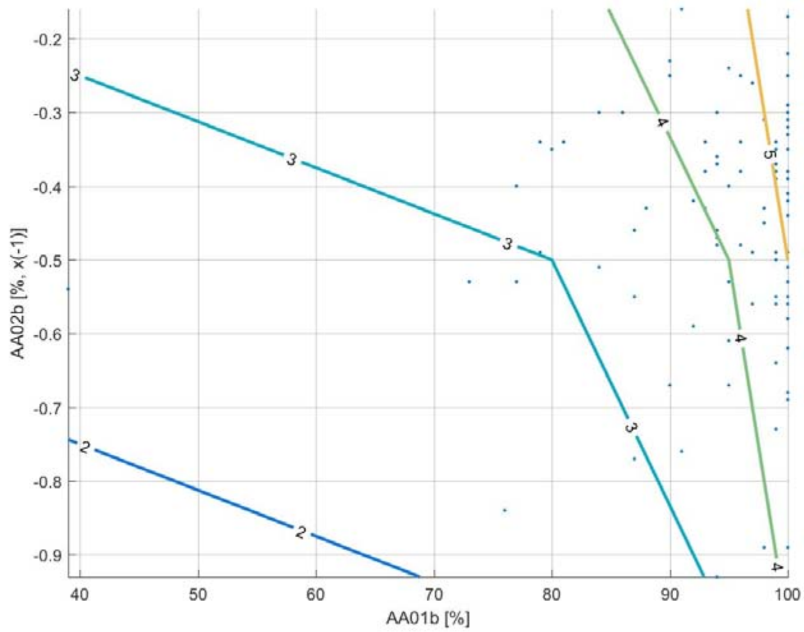
Code	Description	Range	PI covered	Notes
g_1	Accessibility of service	$[0, 6]$	AA01b, AA02b	The variations with the urban-rural typology (in AA01b) were considered.
g_2	Quality of service	$[0, 9]$	AA03b, AA04b, AA05b	–
g_3	Economic sustainability	$[0, 9]$	AA06b, AA07b, AA08b	–
g_4	Infrastructural sustainability	$[0, 6]$	AA09b, AA10b, AA11b	AA09b was considered as a polynomial coefficient.
g_5	Physical productivity of staff	$[0, 3]$	AA12b	The variations with the urban-rural typology (in AA12b) were considered.
g_6	Environmental sustainability	$[0, 6]$	AA13b, AA14b, AA15b, AA16b	AA14b and AA16b were considered polynomial coefficients.

imperfections of the data used to compute the performances of the actions, and potential arbitrariness that may affect the criteria definition in accordance with the method proposed in Roy et al. [31]. The focus group contributed with their familiarity about the sources of data used in the computation of performances as well as the operational instruction used to define the criteria, whenever the values to be assigned to the indifference and preference thresholds were determined. However, the focus group had the instinct to consider the discriminating (indifference and preference) thresholds as preference parameters, and similarly, to assume that these thresholds are part of the scale definition, being consequently intrinsically linked. Such assumption led us to highlight that "it is the discriminating power of criteria that indifference and preference thresholds have as the object to take into account: the indifference and preferences which follow from it do not belong to the definition of the scale, but to the way the criterion applies actions of A to the scale" [31, p. 16].

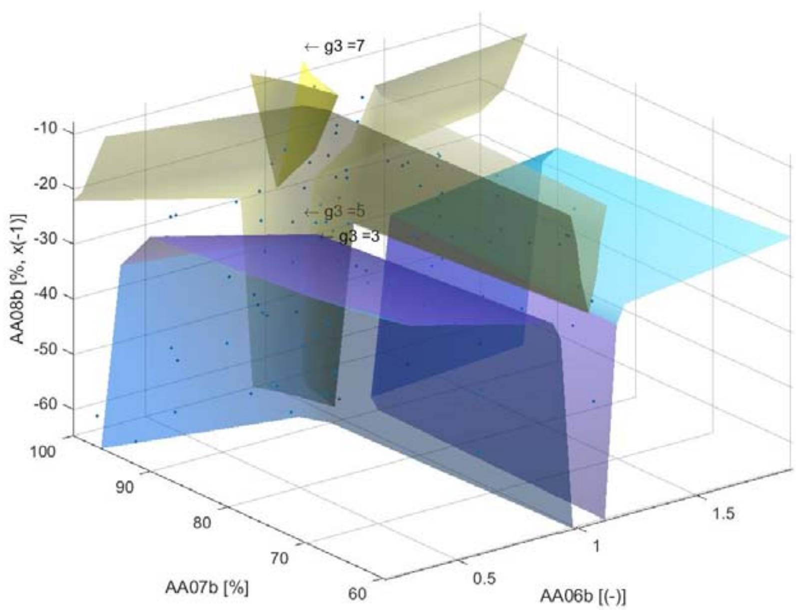
Table 4: Thresholds

Thresholds	g_1	g_2	g_3	g_4	g_5	g_6
q	0.5	1	0.5	1	0.2	0.5
p	1	2	1	2	0.4	1

In order to assign the weights to the six selected criteria, we applied the revised Simos' procedure, which follows an interactive process between the analysts and the focus group (for more details, see [44]). Initially, the focus group



(a) Criteria g_1, set of two attributes



(b) Criteria g_3, set of three attributes

Figure 2: Iso-preference constructed criteria

had a struggle in the definition of the Z value (*i.e.*, in understanding the ratio), which was countered by proposing the following question: "Bearing in mind that the least important criterion is assigned with one vote, how many votes do you assign to the most important?". Despite this drawback, the SRF methodology was very well accepted by the focus group since they are familiarized with decision adding methods (for a similar application of SRF, where the outcome was also considered to be very positive by the actors involved, see Bottero et al. [45]). From the application of such a procedure, we obtained the information presented in Table 5 $(Z = 8)$.

Table 5: Cards ranking given by the DM $(Z = 8)$

Criteria	Rank	No. of blank cards
g_6	1	
		0
g_5	2	
		2
g_4	3	
		1
g_3	4	
		0
g_2	5	
		1
g_1	6	

Subsequently, we used the SRF (Simos-Roy-Figueira) software to obtain the numerical values to the (normalized) weights of the criteria. Table 6 provides the weights of the criteria, which were validated by the focus group. As mentioned earlier, the ELECTRE TRI-nC method allows to associate a veto threshold with the criteria (all or some criteria), which reinforces their role. In this study, veto thresholds were defined for some criteria (*cf.* Table 6).

Table 6: Weights and veto

	g_1	g_2	g_3	g_4	g_5	g_6
Weights	29.6	23.9	21	15.2	6.6	3.7
Veto	3	6	3	6	–	3

ELECTRE TRI-nC allows to classify each water utility into a set of pre-established ordered categories, which are characterized by one or more characteristic reference actions, or say, utilities that are formed from the representative attribute values for each category [23]. In our study, five completely ordered categories were considered. The reference water utilities were defined in an interaction between the analysts and the focus group, while taking into consideration the focus group experience. The performances of those ‘reference water utilities’ are presented in Table 7.

Table 7: Categories and reference water utilities

Category			g_1	g_2	g_3	g_4	g_5	g_6
	All min		0	0	0	0	0	0
C_1	Very Poor	b_1^1	3	3	1.5	1.75	0.75	1.25
		b_1^2	2	2	1	1	0.5	1
C_2	Poor	b_2^1	3.5	4.5	3	2.25	1.25	2.5
		b_2^2	3.25	3.5	2	2	1	1.5
C_3	Neutral	b_3^1	4	6.5	4	3.5	1.5	3.5
		b_3^2	4	5	4	3	1.5	3
C_4	Good	b_4^1	4.5	8	5	4.5	2	4
		b_4^2	5	7	6	4	2	3.75
C_5	Very Good	b_5^1	5.5	8.5	7	5.25	2.5	5
		b_5^2	5.5	8.25	8	5	2.5	5
	All max		6	9	9	6	3	6

5. Empirical analysis

5.1. Results

The QSI results were obtained with the MCDA Ulaval multi criteria decision aiding tool (software version: MCDA-Ulaval 0.6.1). The chosen credibility level was 0.65. As mentioned earlier, ELECTRE TRI-nC provides the lowest and highest possible categories a water utility should be assigned to. If they are the same, then the utility is, without ambiguity, assigned to such a category; and, if they are different, then the utility's category is ill-determined. In the last case, the DM must decide which category the utility should be assigned to.

A PA process should seek to consolidate a culture of concise, reliable and easily interpretable information for all. And since the current PA system does not provide an integrated evaluation leading to an overall assessment of utilities' quality of service, the use of a synthetic approach is of great relevance, which can be further improved by jointly applying a GIS related software [46]. Hence, after the modeling procedure, the ArcGIS © software was used to geographically represent the QS evaluation, assigning each utility to a QS category. Figure 3 highlights the results achieved. All the represented QS categories are outlined, and for a more comprehensive approach all the utilities that could possibly be assigned to two different categories were also depicted in that way.

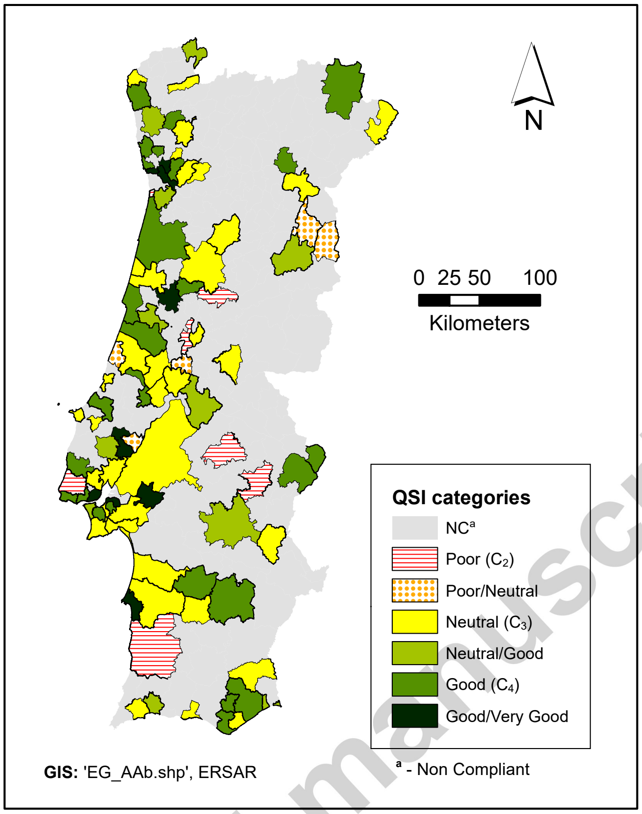


Figure 3: Spatial visualization of QSI results

5.2. Analysis and robustness

In order to assess the results achieved, we performed a robustness analysis, by varying the values of the weights and the credibility level. We used SRF software in order to obtain several sets of weights by varying the Z value from 6 to 10 (intervals of 0.25), while the rank defined by the focus group was maintained. Subsequently, the ELECTRE TRI-nC method was applied using those sets of weights and distinct credibility levels. The variations were combined in several possible ways: for each set of weights, three credibility levels were tested (0.6, 0.65 and 0.7). Therefore, several scenarios were analyzed. The water utilities' classification remains mostly unchanged (Appendix B).

According to the assignment results, the ELECTRE TRI-nC method provides a unique category in 73 of the 99 studied water utilities (73.7%) and two possible categories to assign the water utilities in 26 of them (26.3%). Considering all the scenarios, the method provides a unique category (being always the same) for 54.5% of the utilities. Therefore, results show that this method leads to robust conclusions suitable to develop recommendations regarding the assessment and classification of water utilities.

6. Managerial implications

The use of PA tools enable the comparison between the utilities, fostering the application of benchmarking. The proposed method allows to move from a system of operational PI (*i.e.*, ERSAR system), which are partial measures of performance, to a global assessment (measurement), and to integrate the results into a GIS spatial structure allowing to devise improved policies. Furthermore, the QSI enables to differentiate the 'importance' of those criteria (set of PI) and to deal with difficulties in handling large amounts of complex information in a consistent way [47]. From a regulatory point of view, such development is significant as it may allow to use 'carrot and stick' regulations. Indeed, the setting of yardstick competition with a possibility to include rewards in favor of good performances, and penalties punishing poor performances, is a major policy feature [48], as usually that type of regulations was restricted to other techniques. The previous methods are mostly based on input/output measurements, parametric, such as Stochastic Frontier Analysis (SFA), or non-parametric, as Data Envelopment Analysis (DEA). Indeed, an important consequence of the QSI application relies on the possibility to, depending on the evaluation (category assigned), apply different regulatory settings to regulated utilities [49]. Hence, it would allow to prioritize processes and easily perceive the comparative overall performance.

Regarding the QSI implementation, the objective was to take advantage of the procedures already set for the ERSAR system, such as the submission of data, their validation and the processing of results for each utility. Pragmatically, the added value would be significant, since it would require little marginal effort for a clear visualization and interpretation of each utility's overall QS performance, also it could be easily published with the ERSAR's annual report. Furthermore, each operator would be able to understand how its overall management evolves over time, and, due to the mentioned comparison between utilities, to settle appropriate targets.

The mentioned heterogeneity (*cf.* Sub-section 3.2) disclosed specific flaws, as highlighted in Figure 3. In brief, the integration with a GIS structure allows to develop *ex-post* analyses depending on defined rules (*e.g.*, to relate the QSI result with spatial features) set (modelled) within the GIS software in accordance with the focus group judgements. Thus, regarding the results obtained, it is clear that 'inland' water utilities are prone to reveal compliance shortcomings. Additionally, coastal areas seem to have, on average, an increased level of QS. In terms of management models, there is a specific focus of in-house models in lower QS categories, although also found in higher ones. The corporatized models are mostly assigned to neutral levels, despite holding the highest share of the top categories assignments. Still, one of the most straightforward results is the concentration of top QS level assignments to urban, and more densely populated, areas.

7. Concluding remarks

In this paper we propose the QSI tool. Such application, intends to provide an overall assessment obtained by suitably aggregating operational PI (partial measures of performance). The QSI tool, based on the ELECTRE TRI-nC method, also allows to define quality of service categories. A coherent set of criteria was selected through an interplay between analysts and the focus group, considering their exhaustiveness, cohesiveness, and non-redundancy [42], and modelled resorting to iso-preference logic. The conceptual framework developed, considered the stakeholders of the decision process, as well as the focus group 'cognition' (way of thinking) when choosing a particular preference elucidation mode (see, Li et al. [50], for a clear consideration of different cognitive styles). Another important issue considered was the decision problematic pursued by ERSAR [51].

The results obtained are linked to a GIS spatial structure allowing for a clearer visualization of the overall performance of water utilities and to possibly develop *ex-post* analyses. In fact, the results achieved for each utility are based on data compiled and audited by ERSAR, which highlights the quality of inputs. Additionally, it is possible to compare with the results of similar operators of different geographic areas. In the future, and acknowledging the efforts of ERSAR to improve reporting compliance, the QSI will be able to display more comprehensive results. Other future developments may rely on the possibility to include qualitative criteria, which may be relevant in a context where regulators urge to assess compliance with specific recommendations/laws, is also paramount to ERSAR. However, since we are purely developing an aggregation method of an existing PA system, the inclusion or modification of existing indicators goes beyond the scope of our paper. Further research may also rely on the adoption of a multiple criteria hierarchy process recently published [52, 53], which would allow to consider not only PI at the comprehensive level but also their specificities.

To conclude, the proposed method can be considered a suitable decision support system and useful regulatory tool. Indeed, the QSI is able to provide policy relevant outputs by combining operational features into a global performance score. The possibility to easily interpret the QSI results is a remarkable starting point for discussion, and the QSI added value rationale will surely attract public interest for policy making (as suggested by Nardo et al. [54], Pinto and Marques [55]).

Acknowledgments

The authors are grateful to the anonymous reviewers for the improvements achieved from their insightful suggestions throughout the revision of the paper. The authors would also like to thank all members of the focus group for their role in setting a stakeholder environment and valuable comments. The authors received no funding for this research and have no further conflicts of interest. Any errors and omissions are the responsibility of the authors. The usual disclaimer applies.

References

[1] Schwartz P. John Stuart Mill and laissez faire: London water. Economica 1966. 33(129):71–83.

[2] Lambert DK, Dichev D, Raffiee K. Ownership and sources of inefficiency in the provision of water services. Water Resources Research 1993. 29(6):1573–1578.

[3] OECD. Mapping Policy for Electricity, Water and Transport; vol. 2 of *Infrastructure to 2030*. Paris: OECD Publishing; 2007.

[4] Van de Walle S. What services are public? What aspects of performance are to be ranked? The case of "services of general interest". International Public Management Journal 2008. 11(3):256–274.

[5] Pointon C, Matthews K. Reprint of: Dynamic efficiency in the English and Welsh water and sewerage industry. Omega 2016. 60:98–108.

[6] Cowan S. Competition in the water industry. Oxford review of economic policy 1997. 13(1):83–92.

[7] Berg S. Water utility benchmarking measurement, methodologies, and performance incentives. Water Intelligence Online 2010. 9.

[8] Marques RC, Simões P. Does the sunshine regulatory approach work? Governance and regulation model of the urban waste services in Portugal. Resources, Conservation and Recycling 2008. 52(8):1040–1049.

[9] Alegre H, Baptista JM, Cabrera E, Cubillo F, Duarte P, Hirner W, et al. Performance indicators for water supply services. Water Intelligence Online 2013. 12.

[10] Marques R, Monteiro A. Application of performance indicators in water utilities management:a case-study in Portugal. Water science and technology 2001. 44(2-3):95–102.

[11] Van den Berg C, Danilenko A. The IBNET water supply and sanitation performance blue book: The International Benchmarking Network of Water and Sanitation Utilities Databook. World Bank Publications; 2011.

[12] Marques RC, Simões P, Pires JS. Performance benchmarking in utility regulation: the worldwide experience. Polish Journal of Environmental Studies 2011. 20(1):125–132.

[13] Haider H, Sadiq R, Tesfamariam S. Performance indicators for small-and medium-sized water supply systems: a review. Environmental reviews 2013. 22(1):1–40.

[14] Wiecek MM, Ehrgott M, Fadel G, Figueira JR. Editorial: Multiple criteria decision making for engineering. Omega 2008. 36(3):337–339.

[15] Roy B, Slowinski R. Questions guiding the choice of a multicriteria decision aiding method. EURO Journal on Decision Processes 2013. 1(1-2):69–97.

[16] Haider H, Sadiq R, Tesfamariam S. Selecting performance indicators for small and medium sized water utilities: Multi-criteria analysis using ELECTRE method. Urban water Journal 2015. 12(4):305–327.

[17] Marques RC, da Cruz NF, Pires J. Measuring the sustainability of urban water services. Environmental Science & Policy 2015. 54:142–151.

[18] Nafi A, Tcheng J, Beau P. Comprehensive methodology for overall performance assessment of water utilities. Water Resources Management 2015. 29(15):5429–5450.

[19] Molinos-Senante M, Gómez T, Garrido-Baserba M, Caballero R, Sala-Garrido R. Assessing the sustainability of small wastewater treatment systems: A composite indicator approach. Science of the Total Environment 2014. 497:607–617.

[20] Molinos-Senante M, Marques RC, Pérez F, Gómez T, Sala-Garrido R, Caballero R. Assessing the sustainability of water companies: A synthetic indicator approach. Ecological Indicators 2016. 61:577–587.

[21] Duarte AALS, Rodrigues GMC, Ramos RAR. A global service quality index to evaluate the performance and sustainability in water supply utilities. WSEAS Transactions on Environment and Development 2009. 5(12).

[22] Almeida-Dias J, Figueira JR, Roy B. A multiple criteria sorting method where each category is characterized by several reference actions: The ELECTRE TRI-NC method. European Journal of Operational Research 2012. 217(3):567–579.

[23] Kadziński M, Tervonen T, Figueira JR. Robust multi-criteria sorting with the outranking preference model and characteristic profiles. Omega 2015. 55:126–140.

[24] Corrente S, Greco S, Ishizaka A. Combining analytical hierarchy process and Choquet integral within non-additive robust ordinal regression. Omega 2016. 61:2–18.

[25] Mulliner E, Malys N, Maliene V. Comparative analysis of MCDM methods for the assessment of sustainable housing affordability. Omega 2016. 59:146–156.

[26] Phillips LD. A theory of requisite decision models. Acta psychologica 1984. 56(1):29–48.

[27] Figueira JR, Greco S, Roy B, Słowiński R. An overview of ELECTRE methods and their recent extensions. Journal of Multi-Criteria Decision Analysis 2013. 20(1-2):61–85.

[28] Figueira J, Mousseau V, Roy B. ELECTRE methods. In: Figueira J, Greco S, Ehrgott M, editors. Multiple Criteria Decision Analysis: State

of the Art Surveys.; vol. 1. New York, USA: Springer Science+Business Media Inc; 2016; p. 155–185.

[29] Roy B. Multicriteria methodology for decision aiding. Dordrecht, The Netherlands: Kluwer Academic Publieshers; 1996.

[30] Almeida-Dias J, Figueira JR, Roy B. ELECTRE TRI-C: A multiple criteria sorting method based on characteristic reference actions. European Journal of Operational Research 2010. 204(3):565–580.

[31] Roy B, Figueira J, Almeida-Dias J. Discriminating thresholds as a tool to cope with imperfect knowledge in multiple criteria decision aiding: Theoretical results and practical issues. Omega 2014. 43:9–20.

[32] Roy B. The outranking approach and the foundations of ELECTRE methods. Theory and Decision 1991. 31(1):49–73.

[33] ERSAR. Annual Report on Water and Waste Services in Portugal (2012). Entidade Reguladora dos Serviços de Águas e Resíduos; 2013.

[34] Pinto FS, Marques RC. Tariff recommendations: A panacea for the portuguese water sector? Utilities Policy 2015. 34:36–44.

[35] ERSAR. Caraterização do setor de águas e resíduos; vol. 1 of *Relatório Anual dos Serviços de Águas e Resíduos em Portugal (Annual Report on Water and Waste Services in Portugal, 2015)*. Entidade Reguladora dos Serviços de Águas e Resíduos; 2015.

[36] Pinto FS, da Cruz N, Marques R. Contracting water services with public and private partners: A case study approach. Journal of Water Supply: Research and Technology-AQUA 2015. 64(2):194–210.

[37] LNEC, IRAR. Guia de avaliação da qualidade dos serviços de águas e resíduos prestados aos utilizadores: 1.ª geração do sistema de avaliação. No. 12 in Série GUIAS TÉCNICOS; 1 ed.; IRAR; 2013.

[38] LNEC, ERSAR. Guia de avaliação da qualidade dos serviços de águas e resíduos prestados aos utilizadores: 2.ª geração do sistema de avaliação. No. 19 in Série GUIAS TÉCNICOS; 2 ed.; ERSAR; 2013.

[39] Sun B, Ma W. An approach to consensus measurement of linguistic preference relations in multi-attribute group decision making and application. Omega 2015. 51:83–92.

[40] Saaty TL, Ergu D. When is a decision-making method trustworthy? Criteria for evaluating multi-criteria decision-making methods. International Journal of Information Technology & Decision Making 2015. 14:1–17.

[41] Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value trade-offs. Cambridge university press; 1993.

[42] Keeney RL. Value-focused thinking: A path to creative decisionmaking. Harvard University Press; 1992.

[43] De Feo G, De Gisi S. Using MCDA and GIS for hazardous waste landfill siting considering land scarcity for waste disposal. Waste Management 2014. 34(11):2225–2238.

[44] Figueira J, Roy B. Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. European Journal of Operational Research 2002. 139(2):317–326.

[45] Bottero M, Ferretti V, Figueira JR, Greco S, Roy B. Dealing with a multiple criteria environmental problem with interaction effects between criteria through an extension of the ELECTRE III method. European Journal of Operational Research 2015. 245(3):837–850.

[46] Macary F, Dias JA, Figueira J, Roy B. A multiple criteria decision analysis model based on ELECTRE TRI-C for erosion risk assessment in agricultural areas. Environmental Modeling & Assessment 2014. 19(3):221–242.

[47] Dodgson J, Spackman M, Pearman A, Phillips L. Multi-criteria analysis: a manual. London: Department for Communities and Local Government; 2009.

[48] Porteous AH, Rammohan SV, Lee HL. Carrots or sticks? improving social and environmental compliance at suppliers through incentives and penalties. Production and Operations Management 2015. 24(9):1402–1413.

[49] Konisky DM, Teodoro MP. When governments regulate governments. American Journal of Political Science 2015.

[50] Li Z, Xu J, Lev B, Gang J. Multi-criteria group individual research output evaluation based on context-free grammar judgments with assessing attitude. Omega 2015. 57:282–293.

[51] Guitouni A, Martel JM. Tentative guidelines to help choosing an appropriate MCDA method. European Journal of Operational Research 1998. 109(2):501–521.

[52] Corrente S, Greco S, Słowiński R. Multiple Criteria Hierarchy Process with ELECTRE and PROMETHEE. Omega 2013. 41(5):820–846.

[53] Corrente S, Greco S, Słowiński R. Multiple criteria hierarchy process for ELECTRE TRI methods. European Journal of Operational Research 2016. 252:191–203.

[54] Nardo M, Saisana M, Saltelli A, Tarantola S, Hoffman A, Giovannini E. Handbook on constructing composite indicators. Paris: OECD publishing; 2005.

[55] Pinto FS, Marques RC. Tariff suitability framework for water supply services. Water Resources Management 2016:1–17.

Appendix A. Pictorial description of the remaining QSI criteria modelled

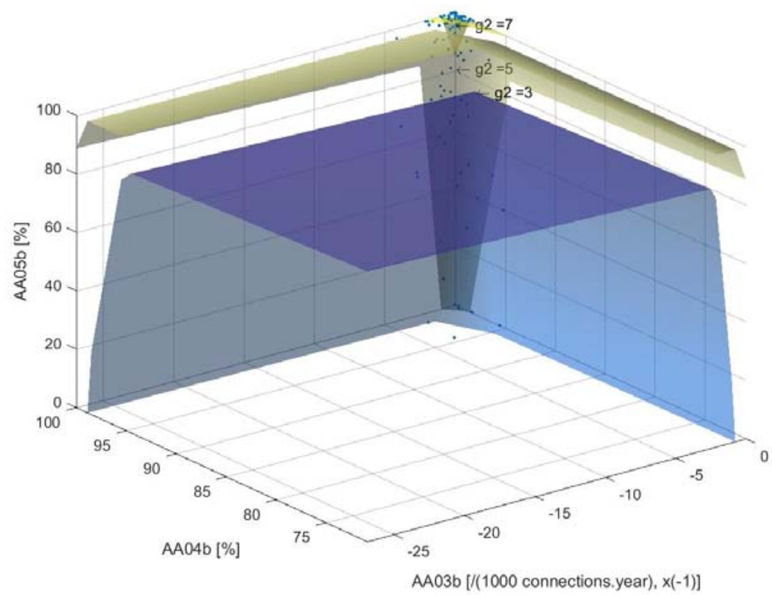


Figure A.4: Criteria g_2, set of three attributes

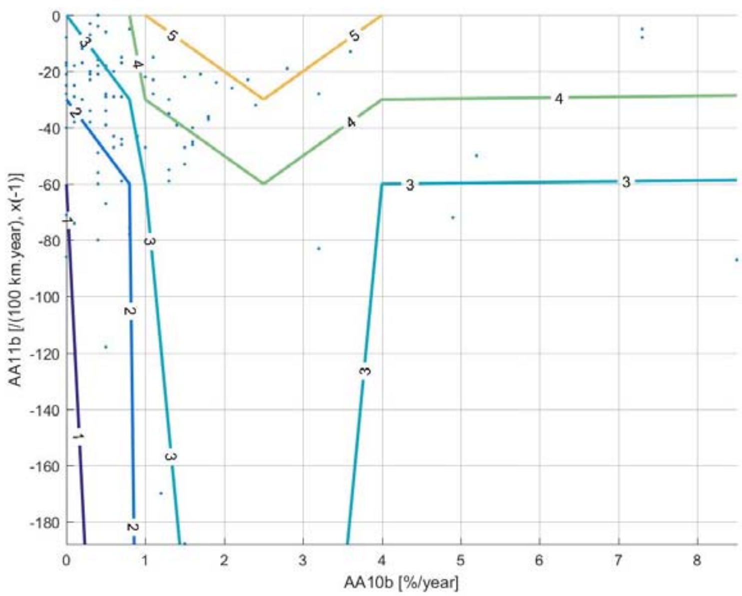


Figure A.5: Criteria g_4, set of two attributes

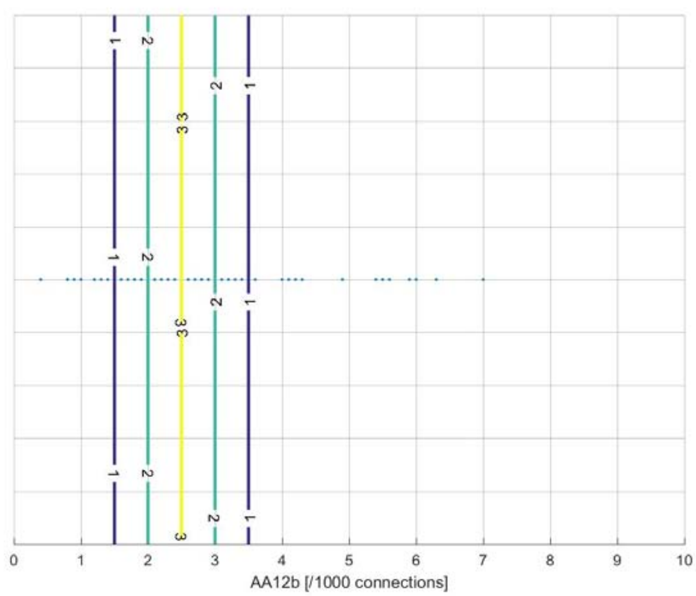


Figure A.6: Criteria g_5, single attribute

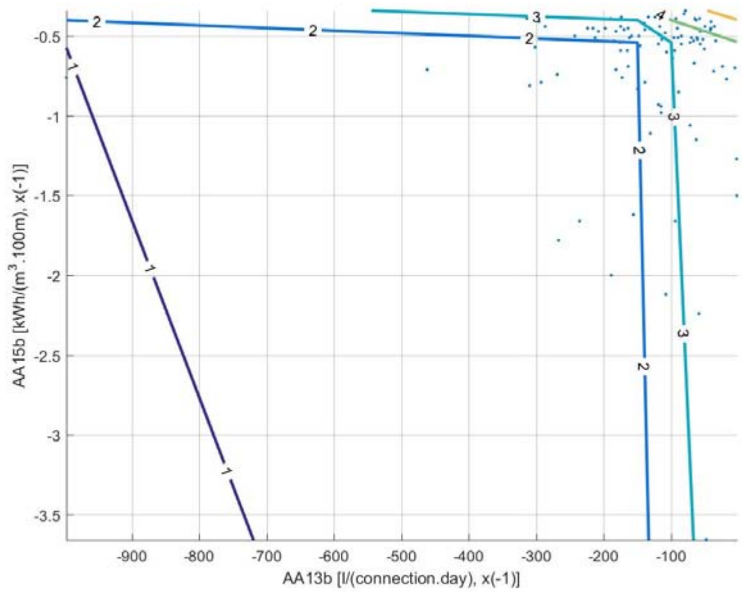


Figure A.7: Criteria g_6, set of two attributes

Appendix B. Variation of the results achieved for different credibility tests and sets of weights

Figure B.8: QSI result variations for the defined credibility levels and sets of weights