Socio-Economic Planning Sciences 92 (2024) 101797

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: www.elsevier.com/locate/seps

Water Utility Service Quality Index: A customer-centred approach for assessing the quality of service in the water sector

Hermilio Vilarinho [a,b,*], Miguel Alves Pereira [d], Giovanna D'Inverno [e,f], Henriqueta Nóvoa [a,c], Ana S. Camanho [a]

[a] Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
[b] INESC TEC, Porto, Portugal
[c] INEGI, Porto, Portugal
[d] CEGIST, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal
[e] Department of Economics and Management, University of Pisa, Pisa, Italy
[f] Faculty of Economics and Business, KU Leuven, Leuven, Belgium

ARTICLE INFO

Keywords:
Data Envelopment Analysis
Quality of service
Benchmarking
Water regulation
Benefit-of-the-Doubt
Composite indicator

ABSTRACT

This work delves into the crucial role of service quality in the water supply and sanitation sector. Despite extensive research and implementation of quality management practices in this sector, a universally accepted definition of quality is still lacking, resulting in various service quality assessment procedures that are difficult to compare. To address this issue, the World Bank launched the 'Utility of the Future' (UoF) programme, aiming to guide water service providers in their efforts to become future-focused utilities that offer reliable, safe, inclusive, transparent, and responsive services through best-fit practices. Building upon the framework provided by the UoF programme, this study proposes the Water Utility Service Quality Index (WUSQI) - a composite indicator that reflects the quality of service provided by water supply and sanitation utilities from a customer perspective. Based on Data Envelopment Analysis, the Benefit-of-the-Doubt approach is employed to assign weights for aggregating the indicators representing the diverse performance dimensions. The study operationalises the WUSQI to assess the quality of Portuguese wholesale water and wastewater companies using data collected by the national regulator of water and waste services. A Multiple Criteria Decision Analysis technique, the Deck of Cards method, is used to specify an indicator of transparency from the information made available by the regulated utilities. The results show the effectiveness of this tool for evaluating and measuring service quality at the company level. Additionally, the findings highlight areas for improvement in the utilities' performance. By enabling companies and regulators to identify areas for improvement, the WUSQI can support the delivery of high-quality services to customers.

1. Introduction

Quality is a multifaceted concept studied extensively in management and it has become critical for the success of many industries [1,2]. However, due to its intangible nature, there is still a lack of consensus on its precise definition, which has evolved over time and in the literature. Quality encompasses dimensions such as performance, reliability, durability, aesthetics and customer satisfaction, making it complex and challenging to define universally [3]. According to van Kemenade and Hardjono [4], quality is a "fuzzy and vague concept" that cannot be measured with certainty since it depends on individual interpretation.

Defining and measuring quality becomes even more complex when it pertains to services. Unlike products, services are intangible and their quality highly depends on the perceptions of the users [5,6]. Public services, such as water supply and sanitation, involve multiple stakeholders with varying priorities and goals. As a result, the concept of quality in these contexts can be interpreted in diverse ways and requires careful consideration of the needs and expectations of all involved stakeholders [7]. In such an environment where users cannot easily switch to a different service provider, maintaining high levels of quality becomes critical to protect their interests and ensure continuity of services [8]. The lack of competition can also reduce the motivation for providers to maintain high-quality service levels, emphasising the importance of measuring quality and taking actions to ensure high-quality standards [9]. Collaboration and coordination among stakeholders are

* Corresponding author at: Faculdade de Engenharia, Universidade do Porto, Porto, Portugal.
E-mail addresses: hermilio.vilarinho@inesctec.pt (H. Vilarinho), miguelalvespereira@tecnico.ulisboa.pt (M.A. Pereira), giovanna.dinverno@unipi.it (G. D'Inverno), hnovoa@fe.up.pt (H. Nóvoa), acamanho@fe.up.pt (A.S. Camanho).

https://doi.org/10.1016/j.seps.2023.101797
Received 27 July 2023; Received in revised form 3 December 2023; Accepted 28 December 2023
Available online 2 January 2024

crucial to establishing clear standards for quality in public services, contributing to the well-being of society as a whole. The provision of safe and dependable water and sanitation services (WSS) is vital to safeguarding public health. To attain this objective, it is essential to prioritise the delivery of high-quality services [10].

Reaching a high level of service quality in the water sector requires a new management approach that ensures continuity of operations, encourages continuous improvement, develops strategic capabilities, and creates efficient and sustainable business models. To support utilities in this endeavour, the World Bank has developed the 'Utility of the Future' (UoF) programme [11], which aims to ignite, materialise and maintain transformation efforts in the water and sanitation sector. The UoF programme guides utilities, particularly in developing countries, to become future-focused and to provide high-quality services, by promoting best-fit practices that enable them to operate in an efficient, resilient, innovative and sustainable manner. It considers that the ultimate objective of water and sanitation utilities is to provide quality services that are reliable, safe, inclusive, transparent and responsive.

In the approach adopted by the World Bank, the quality of WSS is measured following a customer-centred perspective. Performance indicators are suggested for the individual dimensions of reliability, safety, inclusiveness, transparency and responsiveness. Based on the indicator values, the utilities are classified at world-class, well-performing, good, basic and elementary levels for each dimension. The programme proposes separate indicators to evaluate each dimension of the quality of the provided services but does not recommend any method to aggregate these dimensions into a single indicator reflecting the overall quality of service levels.

In fact, composite or synthetic indicators are commonly utilised to measure the quality of services as they can effectively condense complex, multi-dimensional information and support decision-makers. Composite indicators (CIs) offer advantages such as being easier to interpret than a battery of many separate indicators. CIs are also able to track progress over time and minimise the set of indicators that need to be monitored while preserving the underlying information. However, CIs may send misleading policy messages if poorly constructed or misinterpreted [12,13].

This study aims to develop a CI that reflects the quality of service provided by water supply and sanitation utilities from the customer perspective, following the UoF approach. The resulting index is named the Water Utility Service Quality Index (WUSQI). Among the numerous techniques employed to build CIs, we employ the Benefit-of-the-doubt approach (BoD) popularised by Cherchye et al. [14], based on Data Envelopment Analysis (DEA). This method was selected for its capability to assign weights that are the most favourable for the unit under consideration, in comparison to its peers in the sample for aggregating the various metrics. This approach mitigates potential objections from the entities being evaluated, making it a suitable approach for public services under regulation, such as water and sanitation services. The study also uses the Deck of Cards Method (DCM) [15,16], a Multiple Criteria Decision Analysis (MCDA) technique, to construct an indicator that reflects the transparency dimension of utilities' services. The DCM was used to specify one of the metrics that was not readily available. This approach was selected for its intuitive and user-friendly nature, facilitating effective communication with decision-makers in the process of transforming an ordinal scale into a continuous one.

The strategy developed in this study is applied to a case study of the Portuguese wholesale water and wastewater firms taking advantage of the reliable and vast data collection system provided by the Portuguese regulatory authority for this sector. To the authors' knowledge, the BoD technique has not yet been employed to construct CIs based on a customer-centred perspective of quality. The framework proposed by the World Bank in the UoF programme has not been utilised to construct CIs to express the quality of service provided by water utilities, highlighting the novelty of this research. The use of the DCM in this context represents another innovative feature of the study. The choice of Portuguese companies' data as the illustrative application of this study is motivated by the acknowledgement that Portugal presents one of the most exhaustive data sets for the water sector, extensively explored in the literature. This vast data compilation facilitated the collection of information corresponding to each dimension applied in the UoF programme, thereby enhancing the study's analytical depth.

The proposed method can support regulators in evaluating water companies' performance from a customer-centred perspective, making informed decisions that positively impact service quality levels. In some countries, such as England and Wales, service quality is a key input in setting tariffs [17,18]. This study can also provide valuable insights for water companies by identifying customer satisfaction factors, allowing them to improve service delivery. Overall, the relevance of the study relies on the potential to support improvements in the quality of service provided by water companies, benefiting customers and the broader society as a whole. While the primary focus of the UoF programme lies in improving water utilities in developing countries, the method is applied to a European context to demonstrate its practical relevance and applicability at a global level.

The structure of the remaining parts of this paper is as follows: Section 2 provides a concise literature review. Section 3 outlines the methodology proposed, while Section 4 discusses the case study, which serves as an illustrative example of the method's strength and practical application. Section 5 presents and analyses the results. Lastly, in Section 6, the conclusions of the study are presented and potential avenues for future research are explored.

2. Literature review

This section examines the literature on measuring service quality levels. Section 2.1 provides an overview of various methodologies regardless of the sector, whereas Section 2.2 focuses on the specific evaluation of service quality in the water sector.

2.1. Assessment of quality of services

Quality is a vital and intricate element of business strategy, impacting customer satisfaction, firm profitability and economic growth [19]. It drives competition among firms and shapes markets, with customers seeking high-quality products and services. However, despite its significance, there is still a lack of agreement on the precise definition of quality [2,3,8]. Reeves and Bednar [2] suggest that instead of trying to create a single definition of quality that encompasses all the aspects of existing concepts, it is more effective to weigh the trade-offs of these definitions and select the one that best suits the practitioners' requirements. Similarly, when discussing quality definitions related to tangible products, Garvin [3] recommends using the distinction between various quality perspectives for business advantage, ensuring that quality serves as a 'competitive weapon'.

The study by Reeves and Bednar [2] examined the strengths and weaknesses of the different perspectives on defining quality. These include the *excellence definition*, which views quality as a higher achievement; the *value definition*, which sees quality as an added value for the organisation; the *specification definition*, which emphasises conformance to specifications; and the *customer definition*, which focuses on meeting or exceeding expectations of the customers. These perspectives are widely discussed in management literature and applied to assess and measure the level of quality of different businesses [5,20–25].

We summarise the strengths and weaknesses of quality definitions, according to Reeves and Bednar [2], in Table 1. By examining the content of the table, we see that measuring quality as *excellence* can be challenging, while the *specification* definition primarily focuses on internal processes and may not adequately evaluate service quality. As a result, the *value* and *customer* definitions are more suitable for assessing service quality and are, indeed, the most commonly used approaches for this assessment.

Table 1
Strengths and weaknesses of quality definitions.
Source: Adapted from [2].

Definition	Strengths	Weaknesses
Excellence	– Strong marketing and human resources benefits. – Universally recognisable – High achievement.	– Little practical guidance. – Measurement difficulties. – Rapid change of excellence attribute.
Value	– Multiple attributes. – Focused on internal and external efficiency.	– Questionable inclusiveness. – Quality and value are different constructs.
Specification	– Precise measurement. – Force disaggregation of consumer needs.	– Consumers do not know or care about internal specifications. – Inappropriate for services. – Specifications may become obsolete. – Internally focused.
Customer	– Applicable across industries. – Responsive to market changes. – All-encompassing definition.	– Most complex definition. – Difficult to measure. – Customers may not know their expectations. – Confusion between customer service and customer satisfaction.

Other scholars have also been engaged in the ongoing discourse about the concept of quality, acknowledging that it is not a matter of resolution, but rather a constantly evolving idea [26–28]. This ambiguity regarding the definition of quality has also been prevalent in the context of services. In contrast to the concept of quality used for physical products, the assessment of the quality of services (QS) places a greater emphasis on customer perception and marketing [5]. As Harvey [29] points out, service quality assessment is unique to each market segment and can be classified into two main components, reflecting both the outcomes desired by the customer and the efforts that customers must undertake to achieve those outcomes. Additionally, the outcomes desired by the customer must be achieved by examining the performance of internal processes and aligning them with customer perceptions to ensure the desired outcomes are achieved. Due to the intangible nature of the service results, potential discrepancies between perception and reality are more significant in services than in goods. This fact may explain the reason why service quality has been more extensively studied in marketing than in operations, in contrast to product quality which is predominantly researched in the operations field.

Extensive literature reviews, such as Wen et al. [30] and Zhang et al. [31], indicate that there has been a growing focus on QS in quality management research. Numerous studies have highlighted its importance [31–34]. This trend is expected to continue in the future, potentially leading to further advancements in the field.

Prasad and Verma [35] presented a literature review on the main methods used to measure QS and pointed out directions for future research. According to these authors, the most popular approach to measure service quality was introduced by Parasuraman et al. [36]. This scale – entitled SERVQUAL – comprises five dimensions, namely reliability, responsiveness, empathy, assurance and tangibility. The dimension of reliability pertains to the capability to provide the assured service with consistency and precision. The responsiveness dimension focuses on the promptness and willingness to assist customers. The dimension of empathy refers to the level of care and personalised attention provided to customers by the company. The assurance dimension is linked to the expertise and politeness of employees and their capacity to encourage trust and confidence. Finally, the tangibility dimension assesses physical attributes, such as equipment, facilities and the appearance of the staff. The SERVQUAL approach sees QS as the agreement between customers' expectations and their perceptions of the provided service. SERVQUAL has been vastly used in a wide range of services and has shown the potential to be applied both to private [37–43] and public services [44,45]. For reviews of this method, see [46,47].

As noted by Asubonteng et al. [46], while SERVQUAL is a popular tool used to measure service quality, it has been criticised for its limitations, including its assumption that customers have a clear idea of what they expect from a service and the applicability of its five dimensions to all types of services. As an alternative to SERVQUAL, Cronin and Taylor [48] proposed SERVPERF method which focuses solely on service performance. In search of a better measure of service quality, researchers should continue to explore alternative approaches that are more appropriate for different types of services and better capture the nuances of customer expectations and experiences. Numerous methods tailored for specific applications have been proposed, being some derived from the SERVQUAL framework. These include SERVQHOS designed for the healthcare sector [49], HOLSERV for the hotel industry [50], HEdPERF in higher education [51], E-S-Qual for electronic services [52], and Libqual for library services [53]. The exploration of such alternatives is crucial for refining the understanding of service quality in diverse settings.

A strong measurement strategy is essential for improving QS. Metrics communicate organisational priorities and can track progress, compare performance and identify areas for improvement. A precise measurement system for the quality of services enables early detection of deviations and highlights service improvements. Ultimately, it fosters continuous learning and growth [29].

2.2. Assessment of quality of services in the water sector

The service quality in the water sector has been addressed in the literature under different approaches: (i) general performance assessments incorporating QS elements, following the quality-as-value definition, and (ii) QS measurement following the customer-centred perspective.

The first approach includes studies that aim to measure QS with a broader perspective, using the value definition. Although the results are often referred to as measures of service quality, they are actually reflections of overall performance levels. Such studies usually encompass customer-related measures and also integrate additional metrics that may not be directly tied to the customer's perspective, such as environmental sustainability, investment compliance, financial performance, asset management and human resources productivity.

Water services have been incorporating customer-centric aspects into their performance assessments since 1999 when English and Welsh water companies began implementing the overall performance assessment (OPA) methodology [18]. In the literature, according to Picazo-Tadeo et al. [54], the first paper that took a customer perspective into account when measuring water utilities' overall performance was that by Saal and Parker [55]. Since then, this strategy has been employed by several scholars [9,17,18,54,56–66]. Many of those works aim to produce composite indicators (CIs) usually referred to as indices of service quality. These studies are displayed in Table 2.

In 2008, a collaborative effort between the International Water Association and the Inter-American Development Bank resulted in the development of an initiative aimed at establishing a universally recognised model in the assessment of water utility performance. This initiative, known as AquaRating, serves as a performance system that enables the characterisation and evaluation of utilities through the application of key performance indicators and the implementation of best practices. AquaRating has gained recognition as a reference model by regulators, governments, and development agencies. It encompasses eight distinct areas of evaluation, including quality of service, investment planning and implementation efficiency, operating efficiency, business management efficiency, access to service, corporate governance, financial sustainability, and environmental management. This

Table 2
Studies with CIs to evaluate the quality of services in the water sector.

Reference	Quality concept	Composite indicator	Sample	Metrics	Method
Karnib [68]	Customer	Quality of service index (QSI)	4 regional water authorities in Lebanon	4 metrics for 4 years: i. Network coverage ii. Water consumption iii. Continuity of water supply iv. Water quality	Fuzzy inference model
Molinos-Senante et al. [69]	Customer	Quality of service to customers index	19 water and wastewater companies in Chile	7 metrics: i. Water supply pressure ii. Water supply quality iii. Wastewater treatment quality iv. Water supply continuity v. Wastewater collection continuity vi. Billing accuracy vii. Complaints	Ratios of Shephard's distance function to access performance changes over time
Palomero-González et al. [70]	Customer	CI to measure the quality of water supply based on users' perceptions	32 municipalities in Valencia, Spain, receiving water from the same company	6 metrics: i. Network quality ii. Water quality iii. Water price iv. Complaints v. Inconvenience caused by upgrading the network vi. Continuity of service	MCDA model with common weights based on DEA
Duarte et al. [64]	Value	Global index of service quality	15 water supply companies in Portugal	20 metrics from the regulator authority in Portugal grouped in 3 dimensions	Normalisation using fuzzy sets and aggregation by weighted average + Three different options are used to obtain weights from a panel of experts
Pinto et al. [65]	Value	QSI	99 retail water supply companies in Portugal	16 metrics from the regulator authority in Portugal	ELECTRE Tri-nC (MCDA)
Molinos-Senante et al. [66]	Value	Synthetic index of quality of service	40 rural drinking water systems in Chile	14 metrics with weights estimated by different stakeholders	Analytical Hierarchical Process (MCDA) + Monte Carlo simulation
Sala-Garrido et al. [9]	Value	CI of quality of service	24 water and wastewater companies in Chile	7 metrics: i. Investment compliance, ii. Investment to improve the QS iii. Network reposition iv. Non-revenue water v. Interruptions of water supply vi. Obstructions in the sewerage network vii. Payment accuracy	BoD using undesirable metrics + Nerlove–Luenberger super efficiency metric
Molinos-Senante et al. [18]	Value	Quality of service index	24 water and wastewater companies in Chile	10 metrics: i. Non-revenue water ii. Network reposition iii. Investment compliance iv. Water meter operability v. Interruptions in drinking water provision vi. Obstructions in sewerage network vii. Payment accuracy viii. Compliance with drinking water quality ix. Compliance with wastewater discharge x. Water supply pressure	Goal programming
D'Inverno et al. [59]	Value	Water Utility Performance CI	93 Italian water companies	8 metrics: i. Return on assets ii. Return on Equity iii. Earnings before interest, taxes, depreciation, and amortisation margin iv. Financial autonomy v. Autonomy from third parties vi. Water losses vii. Target time to do new connections viii. Target time to repair breakdowns	BoD Model with Directional Distance Function Robust and conditional approaches
Henriques et al. [61]	Value	Performance assessment CI	199 retail wastewater companies and 10 wholesale wastewater companies in Portugal	14 metrics from the regulator authority in Portugal	Directional BoD models for desirable and undesirable indicators

comprehensive approach highlights AquaRating as a system dedicated to enhancing the overall value delivered by utilities [67].

The second approach employed to address the quality of water and sanitation services (WSS) measures QS in a more focused view, using the customer definition. In this group, a prevalent method is to examine customer satisfaction, which is often achieved through surveys that employ satisfaction drivers to analyse users' perceptions of these services. This strategy was employed by Abubakar [71] and Ohwo and Agusomu [72] in Nigeria. Ammar and Saleh [73] and Murrar et al. [74] applied the SERVQUAL model in Palestine with a similar approach. SERVQUAL was also utilised by Kassa et al. [75] in Ethiopia to investigate urban water supply services. Other studies using customer satisfaction surveys are Kumasi and Agbemor [76], Tessema [77], and Rustinsyah [78]. Although surveys are commonly used to collect data, they can be expensive and subjective due to challenges in designing the survey, selecting the appropriate type and method of application, and using statistical methods for analysis. For that reason, performance indicator systems have been developed to conduct QS assessments, focusing on the customer perspective.

Both value-based and customer-based approaches utilise composite indicators as a strategy to evaluate the quality of service (QS). This approach involves gathering diverse metrics from reliable sources, such as regulators or the companies themselves, and consolidating them into a composite or synthetic indicator that effectively represents the provided service quality levels. By employing this method, a comprehensive and representative assessment of QS can be achieved.

Palomero-González et al. [70] argue that while CIs are widely used in research on services and in the water sector, they have not yet been extensively used in particular analysis of service quality under customers' perceptions. To the best of the authors' knowledge, three studies have developed CIs following the customer-centred quality concept for water utilities: Karnib [68] in Lebanon, Molinos-Senante et al. [69] in Chile and Palomero-González et al. [70] in Spain.

Table 2 summarises the main characteristics of the studies that employ CIs to measure QS. The first three rows of this table indicate the three studies that focus on customer perceptions. These studies share a common goal: to aggregate metrics that directly reflect the impact of utility performance on the user's experience into a synthetic indicator. In their respective countries, the selection of suitable metrics was based on their availability. The notable similarities across the three studies include the incorporation of metrics reflecting reliability through service continuity, as well as metrics related to water or wastewater quality, present in all three investigations. Additionally, responsiveness, reflected in metrics measuring customer complaints, is present in two studies: Palomero-González et al. [70] and Molinos-Senante et al. [69]. Furthermore, inclusiveness, either reflecting network coverage in [68] or economic affordability expressed by water prices in [70], is also addressed. These three studies that focus on the user's perspective use different methodologies for the CI computation. However, none of them employs the BoD technique. BoD models have not been used to express QS under a customer-centred perspective in water utilities, representing a novel contribution of this work.

On the contrary, the remaining studies utilising CIs listed in Table 2 do not exclusively rely on metrics related to customer perception. Instead, these studies construct composite indicators that are in many cases termed service quality indexes. However, in these contexts, quality is aligned with the value-based concept introduced by Reeves and Bednar [2]. In Table 2, the studies identified as adopting a value-concept of quality are the ones that incorporate at least one of the metrics not directly linked to the customer perspective. Those studies include measures of environmental sustainability, investment compliance, financial performance, asset management and human resources productivity. Developing an indicator that reflects the overall performance level of a utility may offer the advantage of being more comprehensive, which could explain the wider usage of this approach. Indeed, these studies typically incorporate as many available performance metrics as possible in their assessments, with the goal of capturing utility performances from a comprehensive perspective.

However, it is worth noting that customer-centred approaches to measuring quality of service (QS) can be highly valuable for managers and regulators. By considering the variables that shape customers' perceptions, decision-making processes can become more informed and aligned with the needs and expectations of the users. According to Palomero-González et al. [70], the outcomes of such assessments can significantly enhance the understanding of customers' perceived quality in an objective, quick, simple and cost-effective manner. This fact emphasises the relevance of this study.

To create a CI that accurately represents service quality, one of the crucial steps is to identify the appropriate set of performance metrics that can be combined into a single index. In the water and sanitation sector, performance indicators are frequently utilised to assess various aspects of utility performance in order to identify areas that require improvement. Alegre et al. [79] has compiled a comprehensive collection of indicators that can be used in the sector. Regulators have taken advantage of the various available metrics to better understand and support the performance of companies.

The Water and Waste Services Regulatory Authority in Portugal (ERSAR) provides one of the most widely studied sets of performance indicators in the literature. Every year, ERSAR collects a vast set of metrics from water supply, wastewater, and solid waste service providers in Portugal, and these reports can be accessed on ERSAR's website. This is part of the "sunshine regulation" policy adopted by the Portuguese regulatory authority, which involves openly publishing these metrics.

ERSAR has been reviewing and enhancing its performance indicator system over time. In 2022, the fourth generation of indicators was introduced, with the first set of results scheduled for release in 2023. The most recent data available pertains to the third generation of indicators, which covers the period spanning from 2016 to 2021. Detailed information regarding these indicators can be found in ERSAR's Technical Guide 22 [80]. ERSAR's performance indicator system for water supply and wastewater services comprises 14 primary metrics, grouped into three subsystems: (i) Adequacy of the Interaction with the User, (ii) Service Management Sustainability, and (iii) Environmental Sustainability. The first subsystem reflects the defence of user interests. The second subsystem, which reflects the sustainability of the managing entity, encompasses the economic, financial, infrastructural, operational and human resource capacity necessary to ensure regular and continuous service provision to users. The third subsystem focuses on environmental sustainability and includes aspects related to the environmental impact of the managing entity's activities, particularly with regard to the conservation of natural resources [81].

ERSAR refers to its overall performance appraisal system as a "quality of service measurement system". However, it is important to note that the evaluation method considers various factors beyond just user experience metrics. Therefore, the approach can be characterised as a value-centred quality evaluation system, as described by Reeves and Bednar [2].

Numerous publications have used ERSAR's data, including Duarte et al. [64], Pinto et al. [65], Henriques et al. [61], Mergoni et al. [82], Pereira et al. [83], and Vilarinho et al. [84,85].

After selecting the appropriate metrics to be used in constructing the CI, the next step is to decide on the aggregation technique to be employed. As indicated in Table 2, there are various methods available for this purpose. The Benefit-of-the-Doubt (BoD) approach, which is based on Data Envelopment Analysis, has been utilised in constructing CIs in numerous fields. One of its advantages is that it allows for the assignment of specific and most favourable weights for combining the various metrics. This approach is particularly suitable for regulated markets such as water and wastewater services, where there may be disagreements among operators regarding the relative importance of the different metrics. For this reason, the BoD technique is chosen to be employed in this study.

A new approach to using performance indicators for quality of service evaluation in water and sanitation utilities has been proposed by the World Bank through its 'Utility of the Future' (UoF) programme. This programme was first introduced in 2021 [11] and was updated to version 2.0 in the following year [86]. The UoF programme has set out ambitious objectives that comprise a complete management strategy to foster the development of the utilities and elevate WSS "beyond the next level". The relevance and objectives of this programme are referred to on page *ix* of Lombana Cordoba et al. [86] as follows:

> *Poor service frequently stems from a vicious cycle of dysfunctional political environments and inefficient practices. Global forces – including climate change, water scarcity, population growth, and rapid urbanisation – exacerbate these challenges to providing high-quality, sustainable WSS service delivery. Therefore, WSS utilities require a new approach to planning and sequencing reforms to provide WSS services in a sustainable manner. The UoF programme provides this new approach, building on an extensive body of knowledge on utility performance improvement.*

The UoF programme's ultimate objective is to enhance and maintain the quality of services provided by water and sanitation utilities, which is the topmost priority of the management model presented by the World Bank. This quality-based management strategy requires utilities to be reliable, secure, inclusive, transparent and responsive, which are the dimensions that form the measurement framework proposed by the programme for QS evaluation. In order to be deemed reliable, utilities must provide a continuous 24/7 supply of WSS. Adherence to water and wastewater quality standards represents safety. Inclusiveness requires that no individual or group is excluded from receiving service. To be considered transparent, WSS should provide clear and accurate information regarding their finances, operations and performance. To attain responsiveness, utilities should provide clients with timely and high-quality responses to ensure their satisfaction.

The programme examines each QS dimension using one or more performance indicators, but it does not develop a CI to reflect the overall quality of WSS services. Instead, it assigns a performance level from one to five for each metric, classified as elementary (1), basic (2), good (3), well-performing (4), or world-class (5). Those values for the metric levels are averaged for each dimension, and the utility's QS is assessed by analysing each dimension's performance level.

The UoF programme suggests metrics that can be collected by the utilities reflecting the programme dimensions. The suggested metrics can be seen in Table 3.

This study aims to propose a method to integrate the UoF programme's dimensions into a composite performance indicator, which has not yet been addressed in the literature, representing another novelty in this work.

Table 3
Suggested metrics in the World Bank's 'Utility of the Future' (UoF) Programme to assess service quality for water and sanitation systems.
Source: Adapted from [86].

Dimension	Metric
Reliability	Continuity (hours per day on average) Continuity (customers with 24/7 supply) (%) Availability (litres/capita/day) Availability of faecal sledge management emptying services (provided 24 h after service requested) (%)
Safety	Water quality (samples meeting all standards for drinking water quality) (%) Wastewater and faecal sludge treatment (%)
Inclusiveness	Drinking water coverage (%) Sanitation service coverage (%)
Transparency	Key information disclosure (%) Applications of practices to generate clear information(%) Applications of practices for ensuring accurate information (%)
Responsiveness	Customer satisfied with service (%) Grievances satisfactorily resolved within seven days (%) Sewer blockage complaints addressed within 48 h (%)

Fig. 1. DCM example.

3. Methodology

In this section, the proposed methodology is described in two stages. The first stage, in Section 3.1, describes the Deck of Card Method. In the second stage, in Section 3.2, the calculation of the Water Utility Service Quality Index (WUSQI) using the BoD technique is detailed.

3.1. The Deck of Cards method

In this subsection, the Deck of Cards method (DCM) is presented. This method will be used as a support tool to build one of the metrics used in the QS assessment.

The DCM is a Multiple Criteria Decision Analysis (MCDA) method that has gained popularity due to its simple and intuitive approach, as outlined by Corrente et al. [16]. This method is utilised to assign values to preference parameters in MCDA models, such as the relative importance of criteria in outranking methods or values representing evaluations of alternatives on considered criteria and weights of criteria. In this study, we will describe the application of the DCM to convert a scale with various levels of criteria into a continuous interval scale, while taking into account the strength of preferences between the different levels.

As explained by Corrente et al. [16], in the DCM, when using a discrete scale to evaluate a criterion, each level can be represented by a card that decision-makers can physically manipulate and arrange in their order of preference. The objective is to convert this discrete scale into a continuous scale usually ranging from 0 to 1. The conversion allows decision-makers to assign numerical values that reflect the intensity of their preferences for each level. Typically, the least preferred level is assigned a value of 0, while the most preferred level is assigned a value of 1. To determine the values of intermediate levels, decision-makers must define the strength of preference between each sequential pair of levels. The interval between two consecutive levels is filled with blank cards, and the number of cards in each position reflects the relative importance of the upper level compared to the lower level. The numerical scale can now be determined by considering the total number of cards in the deck, including both level cards and blank cards.

For better clarification, let us consider an example. Suppose that the criteria E presents four levels, l_1, l_2, l_3, l_4. The order of preference determined by the decision-makers is $l_1 \prec l_2 \prec l_3 \prec l_4$ ($\prec$ meaning "strictly less preferred than"). The decision-makers place one blank card between l_1 and l_2, two blank cards between l_2 and l_3 and four blank cards between l_3 and l_4. This means that the significance of l_3 compared to l_2 is judged to be higher than the significance of l_2 compared to l_1, and the significance of l_4 compared to l_3 is considered to be higher than the significance of l_3 compared to l_2. The resulting deck of cards is illustrated in Fig. 1.

In this example, if the value 0 is assigned to l_1 and 1 is assigned to l_4, the remaining level values are given based on their position in the deck. Since there are eleven cards in the deck and ten spaces between cards, each card position is assigned a value between 0 and 1, in intervals of $(1-0)/10 = 0.1$. Therefore, the value for l_2 is 0.2 (since it is the third card in the deck, the value is calculated as $0+(3-1)\times 0.1 = 0.2$) and the value for l_3 is 0.5 (being the sixth card in the deck, the value results in $0 + (6 - 1) \times 0.1 = 0.5$).

3.2. Calculation of the WUSQI using the BoD technique

This subsection explains the strategy used to calculate the CI WUSQI, which involves using a BoD linear programming model.

BoD models are DEA-based models that can handle multiple outputs representing various metrics to be aggregated and a dummy input with a unitary value for all decision-making units (DMUs). This approach was initially proposed by Melyn and Moesen [87] to assess macroeconomic performance and popularised by Cherchye et al. [14]. The BoD model employed in this study for aggregating the chosen metrics collected from ERSAR's data set is based on a Directional Distance Function (DDF) proposed by Zanella et al. [13]. The DDF-based BoD model can handle both desirable and undesirable metrics without requiring any adjustment of measurement scales. Desirable metrics are the ones for which a better performance corresponds to higher values. Conversely, undesirable metrics are characterised by lower values indicating better performance. The set of mathematical expressions in Model (1) defines the Directional Distance Function BoD model used in this study. While the conceptual roots of BoD models lie in DEA, the evolution of DDF models has taken a distinct trajectory. A more detailed elucidation of the development of Model (1) can be found in Appendix A.

BoD models are used to perform a comparative performance assessment of a set of entities, commonly referred to as DMUs. To perform the complete assessment of the set of DMUs, Model (1) must be run and solved n times, being n the total number of DMUs. The outcomes of the model are the decision variables' values, which include v as the dummy input, and the weights u_r for the desirable metrics r, and p_k for the undesirable metrics k. The total number of desirable metrics is s, and the total number of undesirable metrics is l. The desirable and undesirable metrics are represented as y_{rj} and b_{kj}, respectively, for DMUs j (where j ranges from 1 to n). The values y_{rj_0} and b_{kj_0} correspond to the metrics of the DMU under assessment, denoted as j_0. The index r pertains to the set of desirable metrics (with r ranging from 1 to s), and the index k pertains to the set of undesirable metrics (with k ranging from 1 to l).

$$
\begin{aligned}
\text{minimise} \quad & \beta_{j_0} = -\sum_{r=1}^{s} y_{rj_0} u_r + \sum_{k=1}^{l} b_{kj_0} p_k + v \\
\text{subject to} \quad & \sum_{r=1}^{s} g_y u_r + \sum_{k=1}^{l} g_b p_k = 1 \\
& -\sum_{r=1}^{s} y_{rj} u_r + \sum_{k=1}^{l} b_{kj} p_k + v \geq 0, \qquad j = 1, \ldots, n \\
& u_r \geq 0, \qquad r = 1, \ldots, s \\
& p_k \geq 0, \qquad k = 1, \ldots, l \\
& v \in \mathbb{R}
\end{aligned}
\tag{1}
$$

The direction in which desired metrics expand and undesired ones contract towards the 'best-practice frontier' is indicated by the directional distance vector, which is specified as $(g_y, -g_b)$. The choice of the direction vector used in DEA/BoD models can impact the results obtained and has been discussed by many scholars. Depending on the objective of the study, several solutions have been proposed in the literature. Rogge et al. [88] presents alternatives for the vector to set the directions of improvement for desirable and undesirable outputs in BoD models. In this study, we adopt the values of $(g_y, -g_b)$ as $(y_{rj_0}, -b_{kj_0})$, following Zanella et al. [13] and Rogge et al. [88]. This approach enables each DMU to improve by following the path indicated by its specific metrics and the resulting CI value can be interpreted proportionally.

The performance level of DMU j_0 is represented by the factor β_{j_0} in model (1), which embodies the objective of the model. The minimum value of β_{j_0} determined by optimisation indicates the maximum possible expansion of desirable metrics and contraction of undesirable metrics while satisfying the constraints in the model. This allows DMU j_0 to choose the weights that are the most favourable to it. The associated CI, $WUSQI_{j_0}$ for j_0, is obtained as $1/(1 + \beta_{j_0})$, with a range of 0 to 1, where 1 indicates the highest level of performance. If $WUSQI_{j_0} < 1$, it means that there is a linear combination of other DMUs that performs better overall. If $WUSQI_{j_0} = 1$, DMU j_0 is on the best-practice frontier, which implies that none of the other DMUs evaluated performs better than it does.

To limit the range of the assigned weights, weight restrictions must be incorporated into the model. Zanella et al. [13] present an enhanced method for implementing weight restrictions in a directional BoD context. The weight restrictions proposed by Zanella et al. [13] are formulated using virtual weights, and enable the expression of the relative importance of the indicators in percentage terms.

Virtual weights are typically expressed as the product of each DMU's metric by its respective optimal weight. The restrictions proposed by Zanella et al. [13], instead of using the values of the outputs observed at the DMU under assessment, resort to the use of an "average DMU" whose output metrics are equal to the average of all values observed in the DMUs included in the sample, represented by $(\bar{y}_r, \bar{b}_k)$. In this strategy, percent-based constraints are formulated and included in the BoD model using the virtual weights of the average DMU. According to Zanella et al. [13], this formulation results in the specification of weight restrictions in the form of Assurance Regions type I (AR-I) weight restrictions (see [89] for further details). The use of AR-I restrictions provides the benefit of enabling the construction of a unique frontier for the assessment of all DMUs and represents the optimal choice for constructing CIs and rankings. These AR-I restrictions are the most commonly used weight restrictions in BoD models, being employed by many other researchers such as D'Inverno and De Witte [90], Henriques et al. [91], and Pereira et al. [92].

The weight restrictions, as presented in Expressions (2), are included in the BoD model following Zanella et al. [13]. Lower bounds expressed as percentages (ϕ_r and ϕ_k, respectively for desirable and undesirable indicators) are employed to ensure that no weights are equal to zero, preventing any indicator from being completely disregarded in the calculation of WUSQI, and assigning a minimum level of importance to the indicator. On the other hand, upper bounds expressed as percentages (ψ_r for desirable and ψ_k for undesirable indicators) are employed to impose maximum levels of importance on the indicators.

$$
\begin{aligned}
& \phi_r \leq \frac{u_r \bar{y}_r}{\sum_{r=1}^{s} u_r \bar{y}_r + \sum_{k=1}^{l} p_k \bar{b}_k} \leq \psi_r, \quad r = 1, \ldots, s \\
& \phi_k \leq \frac{p_k \bar{b}_k}{\sum_{r=1}^{s} u_r \bar{y}_r + \sum_{k=1}^{l} p_k \bar{b}_k} \leq \psi_k, \quad k = 1, \ldots, l
\end{aligned}
\tag{2}
$$

In order to evaluate the service quality of each DMU j_0, it is necessary to solve the BoD model separately for each DMU. The resulting $WUSQI_{j_0}$ value represents the performance of the DMU in terms of service quality.

The contribution of each metric to the determination of $WUSQI_{j_0}$ represents a piece of valuable information for decision-makers. It provides crucial insights into the performance of DMUs relative to others, highlighting areas of strength and weakness. This knowledge enables decision-makers to allocate improvement efforts more effectively. The BoD technique allows extracting this information from the model results.

In standard BoD/DEA models, which are limited to analysing situations with only desirable metrics, the strengths and weaknesses of the DMUs can be identified by the magnitude of the virtual weights, computed as the values of each metric multiplied by the associated optimal weight ($u_r y_r, r = 1, \ldots, s$). In these models, higher virtual weights indicate *strengths* of the DMUs, as the models assign higher weights to metrics with superior performance. Conversely, lower virtual weights indicate *weaknesses* of the DMUs, as they are assigned to metrics with inferior performance.

BoD models based on the Directional Distance Function (DDF), such as Model (1), include both desirable and undesirable indicators in the optimisation underlying the performance estimation. In these models, the magnitude of virtual weights based on the observed values of desirable and undesirable outputs of the DMU under assessment cannot be interpreted directly. This challenge arises because these models aim to optimise weights for desirable and undesirable metrics towards opposite directions, as indicated in the objective function of Model (1). Furthermore, the model's constraints play a pivotal role in balancing the weights assigned to the various metrics by preserving the proportional improvement paths and allowing a feasible comparison with the remaining DMU's. This transforms the use of virtual weights as a way to determine the relative contribution of each metric to the overall assessment into a complex task.

To address this limitation, we estimate the relative strengths of each DMU by the normalised values associated with the specification of the weight restrictions in the form of Assurance Regions type I (AR-I), estimated considering the "average DMU" associated to the sample under assessment as shown in (2). The AR-I weight restrictions are independent of the units of measurement and express in relative terms the contribution of each metric to the overall performance. Note that the denominator of these restrictions encompasses both undesirable and desirable metrics, ensuring that the sum of the relative virtual weights for all metrics results in 100%.

This approach provides a suitable means to break down the contribution of each output (desirable and undesirable) to the overall performance score. Accordingly, higher magnitude of the relative virtual bounds estimated in the optimisation process indicates strengths, both for desirable indicators and undesirable indicators ($\frac{u_r \bar{y}_r}{\sum_{r=1}^{s} u_r \bar{y}_r + \sum_{k=1}^{l} p_k \bar{b}_k}$, $r = 1, \ldots, s$; $\frac{p_k \bar{b}_k}{\sum_{r=1}^{s} u_r \bar{y}_r + \sum_{k=1}^{l} p_k \bar{b}_k}, k = 1$).

The formulation of AR-I weight restrictions as presented in (2) was proposed by Zanella et al. [13], and these authors reinforce the use of these restrictions to enable the representation of the relative importance of the metrics. However, it is noteworthy that, to the best of the authors' knowledge, the application of the information that can be retrieved from these restrictions to express the strengths and weaknesses of DMUs in DDF models has not been explored in the existing literature. Thus, this study makes a novel contribution by exploring the interpretation of by-products (optimal weights) of the performance assessment in directional BoD models.

4. Case study

In this section, the data set obtained from ERSAR and the metrics chosen to compute the WUSQI are introduced in Section 4.1. In Section 4.2, the computation of the metric representing the transparency dimension for water utilities is described. In Section 4.3 the final data set for WUSQI determination is presented. Finally, in Section 4.4, the bounds used in the model's weight restrictions are determined.

4.1. ERSAR's data set for the determination of the WUSQI

In this subsection, the data set used in this illustrative case study is introduced and the metrics to be aggregated composing the WUSQI are presented.

The data used to compute the WUSQI was obtained from the reports issued by ERSAR for the water sector. For this study, the data set includes the entire period covered by the "third generation" of ERSAR's metrics, spanning from 2016 to 2021. In Portugal, the water sector is characterised by a division between wholesale or bulk utilities and retail utilities, each encompassing distinct management entities. Wholesale utilities primarily focus on providing services to the retail market, while retail utilities directly cater to the final users.

The study focused on the set of water wholesale utilities in Portugal that provide both water supply and sanitation services. To evaluate the performance of these distinct business areas, two separate analyses are conducted. In the study, the DMUs are defined as a combination of the utility and the year. This means that each utility can be compared with all the other utilities in the sample, as well as with its own performance in different years.

The justification for focusing on the wholesale segment in this study stems from the highly fragmented nature of the retail market, which encompasses over 200 utilities. By choosing to concentrate on the wholesale segment, with a smaller yet representative sample, the study aims to present a more thorough demonstration of the developed method's full potential. This decision enables a deeper exploration of the method's capabilities and a more comprehensive understanding of its applicability within the water sector.

The study includes the following water supply utilities: Águas de Santo André (AdSA), Águas do Algarve (AdA), Águas do Centro Litoral (AdCL), Águas do Douro e Paiva (AdDP), Águas do Norte (AdN), Águas do Vale do Tejo (AdVT), Águas do Vouga (AdVouga), Águas Públicas do Alentejo (AgDA), Empresa Portuguesa de Águas Livres (EPAL), and Infraestruturas e Concessões da Covilhã (ICOVI). AdDP was established in 2017, so data for this utility is available only from 2017 to 2021. Therefore, the total number of DMUs for the water supply sector is 59.

The wastewater utilities considered for the study are Águas da Serra (AdSerra), AdSA, AdA, AdCL, AdN, AdVT, Águas do Tejo Atlântico (AdTA), AgDA, Associação de Municípios de Terras de Santa Maria (AMTSM), Saneamento da Península de Setúbal (SIMARSUL), Saneamento do Grande Porto (SIMDOURO), and Tratamento de Águas Residuais do Ave (TRATAVE). For AdTA, SIMARSUL, and SIMDOURO, the data available is from 2017 to 2021. Therefore, the total number of DMUs in the wastewater sector is 69.

In this study, the available metrics from ERSAR's database collected annually from Portuguese wholesale water utilities were examined. Following a thorough screening process, metrics that accurately represent the dimensions of the UoF programme were chosen. To ensure the robustness of this selection, we sought the input of experts in the water sector, with in-depth knowledge of ERSAR regulatory mechanisms.

The selected metrics are presented in Table 4 comprising the water supply and wastewater utilities. However, we found that there were no specific metrics available in ERSAR's data set to measure the dimension of transparency. To overcome this limitation, we elaborated a metric that captures the transparency dimension of the UoF programme by using the Deck of Card method (DCM), as explained in Section 3.1. The final determination of the transparency metric is presented in Section 4.2. This approach allowed us to comprehensively evaluate the Portuguese water utilities' performance across all dimensions of the UoF programme.

The Water Utility Service Quality Index (WUSQI) for the water supply business is determined by combining the metrics AA01a, AA02a, AA03a, AA04a, and AA05a, which are collected from ERSAR's data set and the new metric Transp-AA, derived from the DCM. On the other hand, the WUSQI for the wastewater business is formed by the composition of the metrics AR01a, AR02a, AR03a, AR13a, and AA04a, obtained from ERSAR's data set, and the metric Transp-AR, derived from the DCM as well. In the context of ERSAR, the acronym AA refers to water supply ("Água de Abastecimento" in Portuguese), while the acronym AR represents wastewater ("Água Residual" in Portuguese).

A careful observation of Table 4 reveals that the metrics AA02a, AR02a, AA03a, and AR03a are the only ones that are *undesirable*, meaning that their results are better when they present lower values. This characteristic is important in the aggregation process as explained in Section 3.2.

Table 4
ERSAR metrics to compose the Water Utility Service Quality Index (WUSQI).

Dimension	Water Supply (AA)		Wastewater (AR)	
	Metric ERSAR	Definition	Metric ERSAR	Definition
Inclusiveness	AA01a - Service physical accessibility (%)	Percentage of the total number of households located in the area of intervention of the utility for which there are wholesale infrastructures connected or with the possibility of connection to the retail system.	AR01a - Service physical accessibility (%)	Percentage of the total number of households located in the area of intervention of the utility for which there are wholesale infrastructures connected or with the possibility of connection to the retail system.
	AA02a - Service economic accessibility (%)	Average proportion of income spent on the water supply service based on a consumption of 120 m^3/year and an average income per household in the system's area of intervention.	AR02a - Service economical accessibility (%)	Average proportion of income spent on the sanitation service based on a consumption of 120 m^3/year and an average income per household in the system's area of intervention.
Reliability	AA03a - Occurrence of supply failures (no./delivery point/year)	Weighted average number of supply failures per delivery point per year. The weighting factor is the number of households with effective service depending on each delivery point.	AR03a - Flood occurrence (no./100 km sewer year)	Frequency of flooding incidents originating from the public sewer network, calculated as the number of incidents per 100 kilometres of sewer on public roads and/or properties per year.
Safety	AA04a - Safe water (%)	Percentage of water that is controlled and of good quality, determined by multiplying the compliance rate of required sampling with the percentage of compliance with the specification values set forth in the legislation.	AR13a - Effectiveness in accomplishing legal parameters of wastewater discharge (%)	Percentage of the equivalent population served by treatment facilities that ensure compliance with the discharge requirements, both in terms of periodicity of monitoring and compliance with discharge legal limits.
Responsiveness	AA05a - Reply to suggestions and complaints (%)	Percentage of written complaints and suggestions that received a written response within the legal deadline	AR04a - Reply to suggestions and complaints (%)	Percentage of written complaints and suggestions that received a written response within the legal deadline
Transparency	Indicators are not available		Indicators are not available	

4.2. Determination of the transparency metric

This subsection details how the metric to reflect transparency in the quality of services of water utilities is developed through the application of the Deck of Cards method (DCM).

The current data set of metrics gathered by ERSAR from service providers lacks a metric that represents the dimension of transparency. According to the definition of transparency provided by the UoF programme, it refers to the availability, reliability and accuracy of the information that a utility provides about its operations. To address this gap, a new metric for transparency was developed using the information contained in ERSAR's data set.

In the data set reported from the utilities to ERSAR, a classification of the estimated reliability level for each reported metric is included, as outlined in Table 5. If a metric is not reported, it is indicated as "NR". When determining a utility's transparency level, the amount of missing data and the reliability of the reported information are both taken into account. A service provider is considered more transparent if it reports a higher proportion of data and the reported information is more reliable.

To create a transparency scale, the DCM was utilised, as detailed in Section 3.1. The first step in this process is to define the levels of transparency in order of preference. For this study, the preferred order is straightforward: $NR \prec * \prec ** \prec ***$. The second step involves determining the strength of preference between each sequential pair of levels by placing blank cards between each pair. To ensure the credibility of this step, ERSAR's staff was consulted. Their involvement ensures the robustness of the decision-making process undertaken by water sector experts.

In this study, according to the opinion collected from the experts, five blank cards were inserted between the NR and * levels, three between the * and ** levels, and one between the ** and *** levels, as illustrated in Fig. 2. Using this approach, a continuous transparency scale ranging from zero to one was obtained, with values of 0.000 for NR, 0.500 for *, 0.833 for **, and 1.000 for ***.

Table 5
Metric reliability in ERSAR data set.
Source: Adapted from [80].

Reliability band of the Information source	Associated concept
***	Data based on extensive measurements, reliable records, procedures, investigations or analyses adequately documented and recognised as the best method of calculation.
**	Generally the same as above, but with some non-significant flaws in the data, such as some documentation being missing, old calculations, reliance on unconfirmed records, or the inclusion of some data by extrapolation.
*	Data based on estimates or extrapolations from a limited sample.

A new metric for transparency is then created for each reported metric, and the average of all the transparency values provided by a utility in a given year is used to determine the annual transparency metric for that utility.

4.3. Final data set for the determination of the WUSQI

This subsection presents the final data set comprising the metrics that will be aggregated to construct the Water Utility Service Quality Index (WUSQI).

The data sets for both groups of utilities are not always complete, as some metrics were not reported by the utilities. To address this issue, the study first attempted to use the metric reported by the same utility in the previous year, recognising that it may provide the best available representation of its performance. If the metric was also not reported in the previous year, the approach recommended by Kuosmanen et al. [93], Morais and Camanho [94], and Henriques et al. [91] was employed to handle the missing data instances, which consists of using a small value equal to the minimum value of each

Fig. 2. Transparency metric construction via the Deck of Cards method.

Table 6
Metrics for constructing Water Utility Service Quality Index (WUSQI).

Utilities' group	Dimension	Metric code	Metric description	N	Average	Standard Deviation	Minimum	Maximum
Water Supply (AA)	Inclusiveness	AA01a	Service physical accessibility (%)	59	93.47	9.03	79.00	100.00
		AA02a	Service economical accessibility (%)	59	0.18	0.05	0.12	0.28
	Reliability	AA03a	Occurrence of supply failures (no./delivery point/year)	59	0.01	0.01	0.00	0.02
	Safety	AA04a	Safe water (%)	59	99.79	0.20	99.36	100.00
	Responsiveness	AA05a	Reply to suggestions and complaints (%)	59	77.80	27.47	40.00	100.00
	Transparency	Transp-AA	Transparency metric from DCM	59	0.89	0.04	0.80	0.97
Wastewater (AR)	Inclusiveness	AR01a	Service physical accessibility (%)	69	96.17	5.32	82.50	100.00
		AR02a	Service economical accessibility (%)	69	0.19	0.09	0.02	0.37
	Reliability	AR03a	Flood occurrence (no./100 km sewer year)	69	7.96	8.36	0.00	25.55
	Safety	AR13a	Effectiveness in accomplishing legal	69	96.21	4.32	86.50	100.00
	Responsiveness	AR04a	Reply to suggestions and complaints (%)	69	77.45	38.44	0.00	100.00
	Transparency	Transp-AR	Transparency metric from DCM	69	0.89	0.06	0.71	0.99

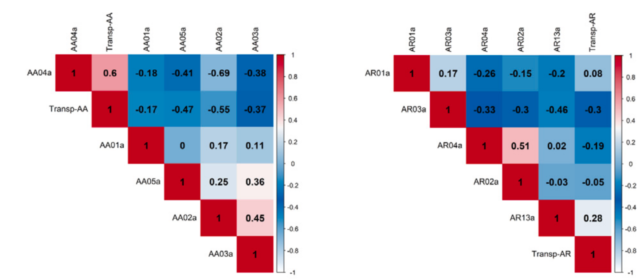
desirable metric as a replacement. In the case of undesirable metrics, the missing instances were replaced with a large number equivalent to the maximum value of each metric. This process ensures that the DMU's performance evaluation is not unfairly affected by the lack of data.

Due to the sensitivity of the DEA method to extreme values in the data set, outliers were replaced following the method proposed by Zanella et al. [95]. An outlier is identified, according to Montgomery [96], as an observation that lies beyond the limits of 1.5 times the distance between the third quartile and the first quartile of the data, known as the interquartile range (IQR). Therefore, values higher than each metric's median plus 1.5 times IQR were replaced by the median plus 1.5 times IQR, and values lower than each metric's median minus 1.5 times IQR were replaced by the median minus 1.5 times IQR. This ensures that atypical observations are replaced with values closer to the centre of the distribution.

DEA formulations typically require positive inputs and outputs, although this requirement can be relaxed, as discussed by Charnes et al. [97]. In the study, the zero values were replaced with a small positive number of 0.0001, following Bowlin [98] and Sarkis [99].

The descriptive statistics for the metrics that compose the WUSQI are displayed in Table 6.

The study examined the correlation between the metrics utilised in building the WUSQI. The calculated Pearson correlation coefficients indicate that there is no strong relationship between the metric pairs used to construct each CI. The absolute values of the coefficients in Fig. 3 are considerably far from one, with only four coefficients marginally exceeding 0.5. Consequently, the lack of a strong correlation provides evidence for incorporating all the variables into the models.



(a) Correlation between pairs of metrics - water supply (AA)

(b) Correlation between pairs of metrics - wastewater (AR)

Fig. 3. Pearson correlation coefficients between pairs of metrics.

4.4. Determination of bounds used in weight restrictions

This subsection details the determination of the lower and upper bounds used in the AR-I weight restriction of the BoD models.

In order to solve the BoD models, it is necessary to define the lower bounds ϕ_r and ϕ_k for desirable and undesirable metrics, respectively, as well as the upper bounds ψ_r and ψ_k for desirable and undesirable metrics, respectively. These bounds are utilised in the weight restrictions, as outlined in Expressions (2).

The lower bounds in weight restrictions guarantee that no weights are assigned null values, thereby ensuring that all metrics are taken into account in the computation of WUSQI. This is crucial for enabling suitable discrimination of the WUSQI value because, if many weights are set to zero, a considerable number of DMUs could be considered best-performing and the comparative performance evaluation would not discriminate differences among performance levels. In this study, weight restrictions are utilised to determine the relative significance of the metrics. It is important to highlight that the lower bounds ϕ_r for desirable metrics and ϕ_k for undesirable metrics set a minimum threshold for the relative contribution of a metric. In fact, the AR-I weight restrictions in the BoD model impose that the model cannot assign weights that result in lower relative significance than those specified thresholds. Consequently, lower values of ϕ_r and ϕ_k contribute to better discrimination of lower performances of the DMUs, effectively revealing their weaknesses. However, if ϕ_r and ϕ_k are set too low, the WUSQI values may lack discrimination. Therefore, it is necessary to search for a balance that allows for both the differentiation of WUSQI values and more precise identification of DMUs' weaknesses. The lower bounds ϕ_r and ϕ_k were set to 0.05 in the study for both water supply and wastewater analysis. Different values ranging from 0.01 and 0.10 were tested. After running this sensitivity analysis, the intermediate value of 0.05 was chosen to strike a balance between the discrimination of WUSQI values and the identification of weaknesses.

The upper bounds ψ_r and ψ_k are utilised to restrict the maximum level of relative importance assigned to different metrics in the performance assessment. In consultation with experts with extensive knowledge of regulation in the Portuguese water sector, it was determined that the specific characteristics of the wholesale market recommend that the responsiveness dimension should be given less relative importance than the other dimensions reflecting inclusiveness, reliability, safety and transparency. This is due to the limited real-time

Table 7
Descriptive statistics for Water Utility Service Quality Index (WUSQI).

Year	Water Supply (AA)				Wastewater (AR)			
	Average	Std. Dev.	Min.	Max.	Average	Std. Dev.	Min.	Max.
2016	0.917	0.046	0.851	0.972	0.951	0.043	0.891	1.000
2017	0.925	0.059	0.840	0.993	0.931	0.050	0.845	1.000
2018	0.917	0.067	0.840	0.998	0.926	0.049	0.836	1.000
2019	0.941	0.045	0.856	0.995	0.928	0.060	0.828	1.000
2020	0.943	0.039	0.875	0.996	0.938	0.064	0.800	1.000
2021	0.929	0.060	0.830	1.000	0.938	0.049	0.878	1.000
Overall	0.929	0.052	0.830	1.000	0.935	0.052	0.800	1.000

contact with the final user that occurs in wholesale water services. This decision is based on the recognition that complaints and suggestions from customers in the wholesale segment are infrequent and typically not considered critical factors for service quality assessments. Treating this dimension with equal significance as the others could potentially yield outcomes that do not accurately reflect the actual requirements for quality service in this specific market segment. Consequently, the upper bound for the responsiveness dimension, represented by metrics AA05a and AR04a for water supply and wastewater utilities respectively, was set at 0.15. This implies that the relative importance of responsiveness in the assessment ranges from 5% to 15%. In contrast, the upper bounds for the remaining metrics were not specified, indicating that the relative importance of these metrics can be greater or equal to 5%. This ensures that the relative importance assigned to these metrics remains flexible while adhering to reasonable limits in terms of the lower bounds.

By incorporating expert input and setting these upper bounds, the assessment framework achieves a balanced consideration of the metrics, taking into account the unique characteristics of the wholesale market in Portugal. This approach ensures that the evaluation remains aligned with the actual requirements and expectations for quality service in this segment.

5. Results and discussion

The BoD models were computed for the two different businesses, water supply and wastewater. Since the DMUs comprise a combination of utility and year, each model computation included the data for all utilities over the six years of analysis. Descriptive statistics, presented in Table 7, offer a summary of the obtained results. The statistics of WUSQI reveal that there is potential for improvement in both sectors. Furthermore, when examining the yearly averages of WUSQI for each year, a relative stability in the performance in both sectors is observed, indicating that the businesses have maintained a certain level of service quality over time.

Results of the WUSQI's computation, including the relative importance or contributions for each dimension, are displayed in Table B.1 (Appendix B) and Table C.1 (Appendix C) for water supply and wastewater utilities, respectively. The relative importance of the dimensions is determined by the contributions of the metrics related to each dimension. Note that the relative importance of the dimension of Inclusiveness is determined by summing the contributions of the two metrics that form this dimension.

In the following subsections, a more detailed analysis of the results is provided for each group of utilities. Specifically, Section 5.1 examines the quality of service evaluations for water supply utilities, while Section 5.2 focuses on the same evaluations for wastewater utilities. Finally, Section 5.3 offers insights derived from the geographical distribution analysis of both groups of utilities for 2021, the most recent year under examination. These insights aim to provide support for future improvement initiatives.

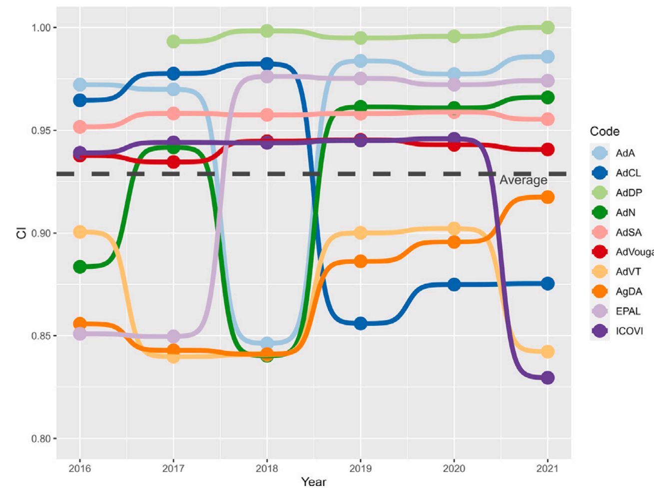


Fig. 4. Evolution of the Water Utility Service Quality Index (WUSQI) from 2016 to 2021 in water supply utilities.

5.1. Results for water supply utilities

In this subsection, the results of the assessment for the group of utilities that provide water supply services are presented and discussed.

The evolution of the WUSQI in water supply utilities from 2016 to 2021 is depicted in Fig. 4. It is evident from the graph that AdDP consistently outperformed all other utilities during this period. Moreover, AdDP's performance showed an upward trend over time, reaching the maximum WUSQI score of 1 in 2021, making it the only water supply utility to achieve this feat. Therefore, AdDP is an ideal candidate for identifying best practices in service quality within the sector.

Throughout the analysis period, three utilities consistently outperformed the sector average, displaying the lowest variation in performance over the years: AdDP, AdSA, and AdVouga. AdA also performed above average throughout the period, with the exception of 2018.

On the other hand, AgDA and AdVT consistently performed below the sector average, with some variation observed in both companies. Notably, AgDA displayed an upward trend in WUSQI over the past four years.

In terms of performance variation, Fig. 4 highlights that AdCL, AdN, EPAL, and ICOVI exhibited higher fluctuations over the years. AdCL experienced a significant decline in performance in the last three years, while ICOVI's performance dropped only in the most recent year. However, AdN and EPAL managed to improve their performance and achieve relative stability in recent years.

The analysis presented in Fig. 4 provides a comparative view of the performance variations and trends for each utility throughout the analysed period. By employing the BoD technique, the assessment highlights the strengths of each utility which minimises objections or complaints that may arise regarding the importance of the various metrics used in the evaluation. By examining the trends depicted in Fig. 4, decision-makers can discern the utilities' performance trajectory and identify notable patterns. The graph reveals the utilities' ability to maintain or improve their standings, as well as areas where they may require additional attention.

The relative importance of each metric to utilities' performance was obtained based on AR-I weight restrictions, using the resulting weights of each metric from the computation of the BoD model. Fig. 5 presents the importance of each dimension in the case of water supply utilities. In Fig. 5, each utility is presented through a bar chart that showcases its performance across different years of analysis. The height of each bar corresponds to 100% total performance. Within each bar, coloured regions indicate the relative importance of different dimensions in determining the quality of service (QS) performance.

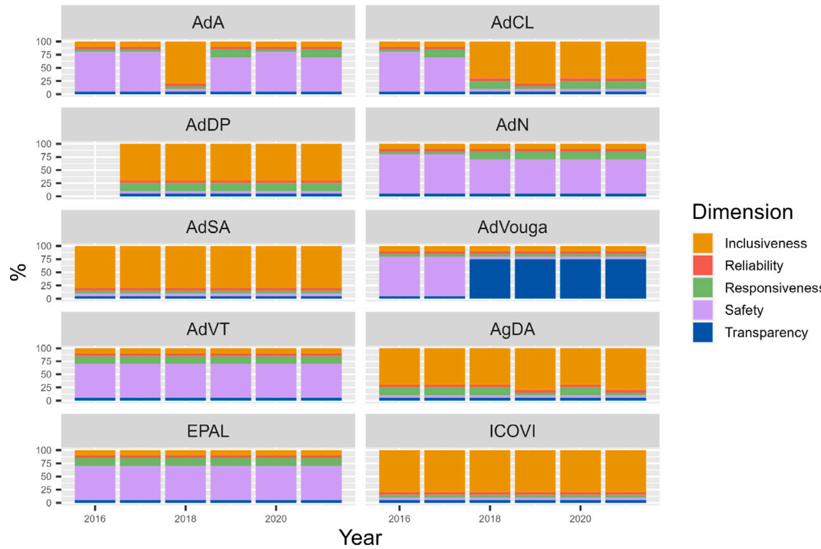


Fig. 5. Relative importance of each dimension in the Water Utility Service Quality Index (WUSQI) - water supply utilities.

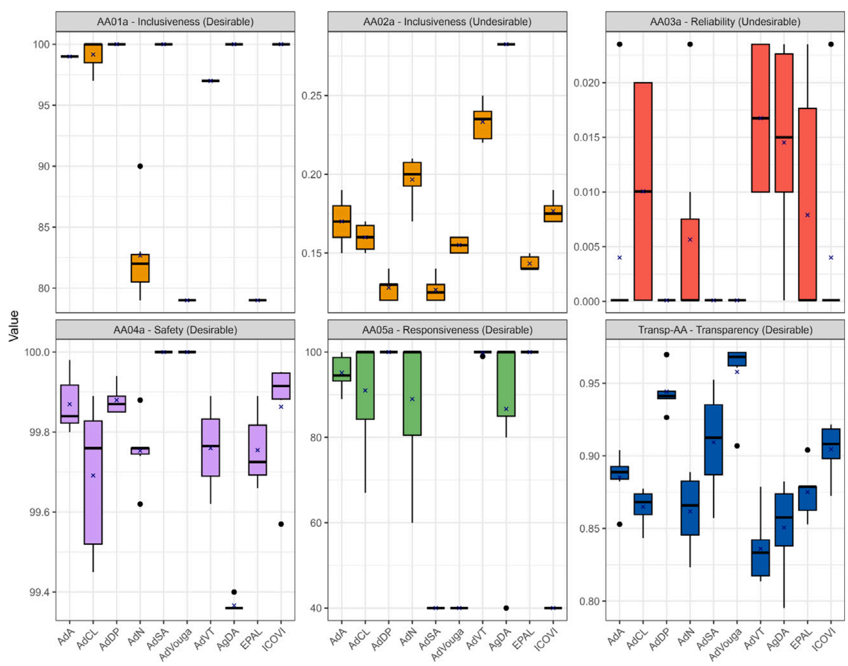


Fig. 6. Metrics of water supply utilities from 2016 to 2021.

By examining the size of these coloured regions, we can easily grasp the respective contributions and relative significance of each dimension to the utility's performance. Larger coloured regions within the bars indicate superior performance in the corresponding dimensions, highlighting the strengths of the utility in a given year. Conversely, smaller coloured regions in the bars signify weaknesses of the utilities in the dimensions they represent. This visual representation provides a concise and intuitive means of comprehending the impact of each dimension on the QS delivered by the utilities.

The contribution of each dimension to the QS performance of the utilities effectively highlights the strengths and weaknesses of the utilities, as evidenced by several examples. For instance, the bar chart of the top-performing AdDP utility in Fig. 5 emphasised its inclusiveness, which remained consistently strong over the years. Inclusiveness was also the primary strength for AdSA, AgDA and ICOVI across the whole period. In contrast, for AdN, AdVT and EPAL, the dimension of safety emerges as the most significant throughout the analysed period. This finding highlights the importance placed on safety in their QS performance.

The importance of dimensions varies for the remaining utilities over the years, indicating fluctuations in areas of improvement. This variation underscores the dynamic nature of the utilities' performance, with dimensions exhibiting different levels of importance at different points in time. It is worth noting that in this analysis, only one dimension emerged as a significant strength for each utility in a given year. The weaknesses of the utilities can be identified among the remaining dimensions.

Analysing this information is not always straightforward, as the contribution values in a DMU assessment are not easily comparable between different DMUs. These values represent relative importance, holding significance only for the performance of each specific DMU. We can examine the values of the safety dimension as an example of this complexity. The results of the performance metrics of the water supply utilities are displayed in Fig. 6. A look at the graph reveals that

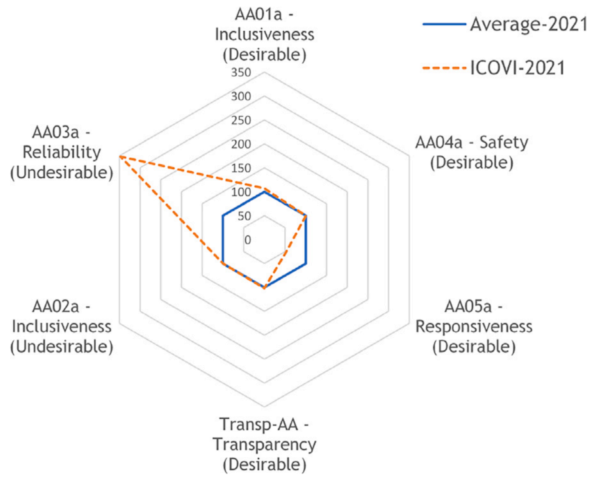


Fig. 7. Comparison between the performance of ICOVI and the average performance of the sector in all metrics for 2021.

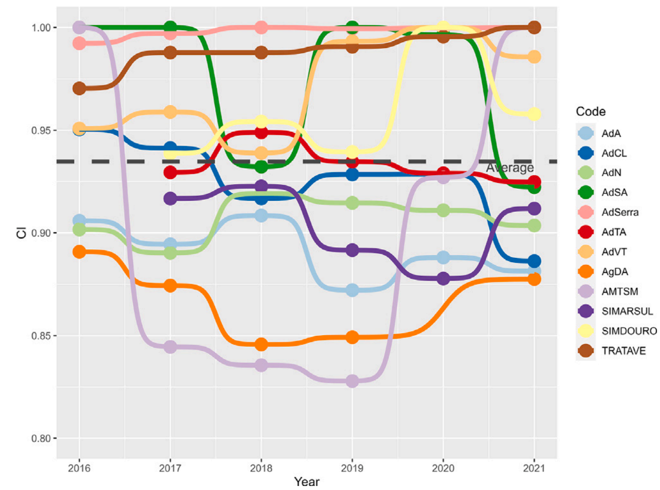


Fig. 8. Evolution of the Water Utility Service Quality Index (WUSQI) from 2016 to 2021 in wastewater utilities.

the safety levels, represented by the indicator AA04a (% safe water), are consistently high in Portugal. All the wholesale utilities maintained safety levels above 99% throughout the entire period. In particular, as shown in Fig. 6, AdSA achieved the maximum value of 100% in safety for all years.

However, in the comparative analysis with other DMUs, AdSA's main strength was identified as inclusiveness in the performance assessment, rather than safety, as displayed in Fig. 5. Furthermore, it is important to note that while AdSA performed better in terms of safety compared to other utilities such as AdVT, EPAL, and AdN, which identified safety as their main strength, we cannot definitively claim that those utilities are safer than AdSA. In fact, it can be stated that inclusiveness emerged as the strongest dimension for AdSA, while safety remained the dominant dimension for AdVT, EPAL, and AdN. Upon analysing the plots for the inclusiveness metrics, AA01a and AA02a, in Fig. 6, it is clear that AdSA stands out as a top performer for both of them. It is important to note that while AA04a and AA01a are desirable metrics, with higher values indicating better results, AA02a is an undesirable metric where lower values indicate better performance.

Lavigne et al. [100] highlight the complexity of such comparisons, pointing out that a poorly performing DMU may exhibit a relative strength in a specific metric, which could be a weakness for a highly-performing DMU, despite the latter performing better overall in that metric.

By considering both the importance of dimensions and the results of the WUSQI, utilities can pinpoint areas that require improvement and make informed decisions to enhance their overall service quality. A notable example is ICOVI's decline in performance in 2021, where decision-makers should concentrate on addressing identified weaknesses. In this specific case, a closer examination of metric values reveals a significant deterioration in ICOVI's reliability. Specifically, the number of supply failures per delivery point (AA03a) reached the highest level within the entire sample, indicating the worst performance in this regard. This clear indication highlights a critical area for ICOVI to prioritise and improve upon. The radar chart shown in Fig. 7 provides a visual representation of ICOVI's metrics in 2021, as well as the sector's average for the same year. These metrics have been normalised on a scale that considers the sector's average as 100 to ensure comparability. Notably, the undesirable metric AA03a demonstrates a significantly poorer performance compared to the sector's average in the same year.

One important aspect to consider when analysing the presented results is the role of the regulator in suggesting continuous improvement actions for the utilities. Based on the performance of each utility in different dimensions, the regulator can provide guidance to companies on how to improve their service quality. For instance, for utilities that consistently underperform in specific metrics, the regulator can provide targeted support to address the issue. This could include setting specific targets for improvement, technical assistance or even imposing fines for non-compliance. On the other hand, for utilities that excel in certain dimensions, such as AdDP in inclusiveness, the regulator can recognise their success and encourage them to share their best practices with other utilities.

5.2. Results for wastewater utilities

This subsection displays and discusses the results of the assessment for the group of utilities that provide wastewater services.

Fig. 8 illustrates the progression of the WUSQI in wastewater utilities between 2016 and 2021. In this analysis, ten DMUs emerged as top performers, achieving the highest WUSQI score of 1. Notably, each year included in the analysis featured its own set of top performers. In 2016, AMSTM and AdSA stood out, followed by AdSA in 2017, AdSerra in 2018, AdSA in 2019, and both AdVT and SIMDOURO in 2020. Lastly, in 2021 AMSTM, AdSerra, and TRATAVE claimed the top spot in 2021. These utilities' metrics for quality of service in those years can serve as benchmarks for the wastewater sector. Additionally, AdSerra, AdVT, SIMDOURO and TRATAVE consistently performed above the sector's average throughout the entire analysis period. On the other hand, AdA, AdN, AgDA, and SIMARSUL, consistently fell below the sector's average. The other utilities exhibited more variability in their performance across the period.

The visualisation presented in Fig. 8 proves to be a powerful tool for detecting significant changes in performance. It enables the identification of utilities that have achieved relative stability over time, exemplified by AdSerra, which consistently maintained a WUSQI value between 0.992 and 1.000. On the other hand, it also highlights utilities that have experienced remarkable variations, such as AMTSM. By closely examining and analysing these variations and trends, decision-makers can gain valuable insights into the underlying factors driving performance fluctuations within the utilities. These insights can inform strategic decision-making processes, allowing for targeted interventions and improvement initiatives where they are most needed.

In the context of wastewater utilities, Fig. 9 displays the relative importance assigned to each dimension based on their respective metrics, which can be interpreted in the same way as in the water supply sector.

By examining the bar charts, we can gain insights into the relative importance of various dimensions in the quality of service (QS) performance of different utilities across different areas. AgDA, AMTSM and

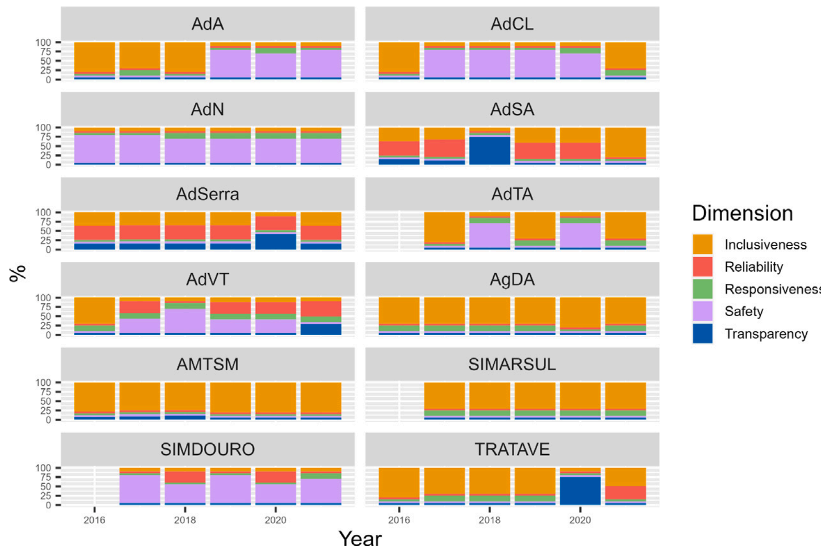


Fig. 9. Relative importance of each dimension in the Water Utility Service Quality Index (WUSQI) - wastewater utilities.

SIMARSUL, for example, exhibited consistent strength in inclusiveness throughout the entire analysis period. On the other hand, AdN and SIMDOURO consistently showcased strong safety as their most significant dimension across all the years under evaluation. The other utilities, however, demonstrated more diversity in their primary strengths.

Notably, in the case of wastewater utilities, multiple strengths were identified in each utility for a particular year, which distinguishes them from the water supply sector. For instance, in 2021, AdSerra showcased a relative importance of 38.97% in reliability, making it its major strength for that year. This utility was also strong in inclusiveness that accounted for 35.10% of the relative importance in this year. Transparency accounted for 15.83%, and responsiveness and safety for 5.09% and 5.00%, respectively. These figures reveal the main weaknesses for AdSerra in 2021, as safety and responsiveness scored relatively lower compared to the other dimensions. Therefore, AdSerra should prioritise improvement actions aimed at addressing these weaknesses.

To gain a deeper understanding of the wastewater utilities' results, it would be worth examining the factors that contributed to the top-performing utilities' success. Specifically, investigating the specific policies or practices that the top performers implemented to reinforce their strengths and achieve the highest WUSQI scores could provide valuable insights. If commonalities among these utilities are identified, they could be replicated by other wastewater utilities to enhance their service quality.

Furthermore, it would be advantageous to investigate the variations in the performance of other utilities and determine the factors that contributed to their inconsistent service quality, such as the weak reliability levels of many of the low-performing wastewater utilities. It is possible that external factors, such as severe weather events like heavy rainfall, may have influenced the reliability metric, which is linked with flood occurrence in wastewater systems. Moreover, internal factors like management practices and resource allocation could have also impacted their performance.

Regulators can identify best practices and areas for improvement by analysing the top-performing and bottom-performing utilities, as well as those with more variability in their performance. This analysis can inform the development of policies and guidelines that promote continuous improvement in service quality across the wastewater sector.

5.3. Geographical distribution insights comparing water supply and wastewater utilities

This subsection provides additional insights into the assessment results by highlighting the geographical distribution of the water utilities. Specifically focusing on the year 2021, the analysis combines both groups of utilities, aiming to support future improvement initiatives.

In Fig. 10, the maps depict the locations of each utility's headquarters along with their evaluation results for the latest year of analysis, 2021. The symbols used in the maps represent different performance levels: a star indicates top-performing utilities, a top-pointing triangle represents utilities performing above average except for the top performers, and a bottom-pointing triangle represents utilities performing below average. Additionally, the colour of each symbol signifies the dimension in which the utilities excel, reflecting their main strength.

In Fig. 10, it is evident that wastewater utilities in the southern regions of Portugal tend to perform below the sector's average. This observation is in line with the findings of Mergoni et al. [82], who reported on the seasonal imbalances and droughts affecting this part of the country and their impact on the waste collection and treatment processes.

The results of the most recent year analysed reveal that inclusiveness and safety are the predominant strengths among utilities in Portugal. Additionally, six companies, namely AdA, AdCl, AdN, AdSA, AdVT, and AgdA, operate in both the water supply and wastewater sectors. Remarkably, five of these companies consistently excel in the same strength for both water supply and wastewater services. Specifically, AdCl, AdSA, and AgdA demonstrate a strong emphasis on inclusiveness, while AdA and AdN prioritise safety as their main strength. As these companies operate in both water supply and wastewater segments, it can be inferred that there is a consistent approach to management practices across both areas. This alignment in priorities may indicate a deliberate and strategic focus on inclusiveness and safety throughout their operations. Such consistency in managerial practices across sectors reflects a shared commitment to excellence and suggests the presence of effective strategies in place to address these key dimensions.

6. Conclusion

This study proposes the Water Utility Service Quality Index (WUSQI) as a composite indicator that reflects the quality of service provided by water supply and sanitation utilities from a customer perspective. The WUSQI rests on the framework introduced by the World

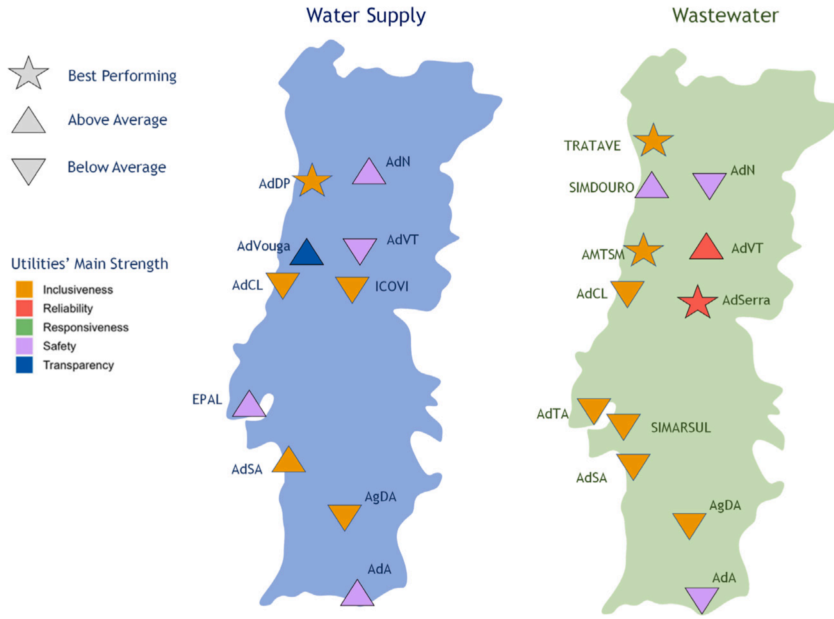


Fig. 10. Results for Water Utility Service Quality Index (WUSQI) assessment in 2021.

Bank's 'Utility of the Future' (UoF) programme, which aims to reflect the reliability, safety, inclusiveness, transparency and responsiveness of the services offered by water sector companies. The Benefit-of-the-Doubt (BoD) approach, based on Data Envelopment Analysis (DEA), is employed to assign weights for aggregating various metrics. The Deck of Cards method (DCM), a Multiple Criteria Decision Analysis (MCDA) technique, is used to define the transparency metric, which was not available in the examined data set.

We apply the WUSQI to assess the quality of Portuguese wholesale water and wastewater firms and find that it is an effective tool for performing a quality service level benchmarking exercise and uncovering performance trends over time. While the UoF programme was originally designed to address water sector needs in developing countries, this study showcases its applicability in a European context, specifically within the water sector of Portugal. By selecting Portugal as a case study, the study highlights the method's capabilities and its relevance in assessing service quality in a developed country setting.

The study's findings have significant implications for the water supply and sanitation sector, where measuring and ensuring high levels of service quality is crucial for public health and well-being. The WUSQI can help utilities and regulators identify strengths and weaknesses, set targets and track performance over time. By adopting a customer-centred perspective and measuring quality along multiple dimensions, the WUSQI encourages utilities to improve their services continuously and to foster trust and satisfaction among customers. Compared to value-centred broader approaches to determining service quality, the WUSQI's focus on the customer perspective represents a significant advantage in terms of enhancing the knowledge of customers' perceived quality in an objective and effective way.

This study makes several innovative contributions to the literature. First, it applies the BoD technique to measure the service quality of water utilities from a customer perspective. Moreover, the AR-I weight restrictions in the BoD model are utilised to reveal the relative importance of the various dimensions. The study also uses the UoF framework as a basis for developing the methodology, which is a novel application of this framework. Finally, the study develops a transparency metric using DCM using ERSAR's data set.

One notable strength of this study is the integration of expert opinion in key stages of the methodology, such as the definition of metrics and the construction of the transparency metric using the DCM. This inclusion of expert input enhances the applicability of the study to the context of water wholesale utilities in Portugal. By incorporating expert knowledge and insights, the study benefits from a comprehensive understanding of the industry and can provide more accurate and meaningful results.

One limitation of the study is that it does not consider the diverse environments in which utilities operate. The regulator may need to take this into account when analysing the results and developing improvement strategies for the utilities. Additionally, changes in the regulatory framework or market conditions may also affect the performance of the utilities. Since these aspects were not the primary focus of our study, we opted not to pursue them. Further works can explore contextual analysis, providing more insights into the utilities' performance and addressing this limitation.

Subsequent studies could also focus on the group of utilities that operate retail systems that are closer to end-users. Analysing the performance of these utilities would provide valuable insights into the specific challenges and opportunities they face in delivering water services directly to consumers, thereby complementing the findings of this study focused on wholesale utilities.

In conclusion, the WUSQI represents a valuable tool for assessing and measuring the quality of water supply and sanitation services, thereby contributing to the water sector's improvement and the achievement of the Sustainable Development Goals. The study's relevance lies in its emphasis on the importance of adopting a customer-centred approach to service quality measurement, which encourages further research on the subject, especially in the context of other public services. It is expected that the WUSQI fosters collaboration and coordination among stakeholders, leading to the provision of high-quality and reliable water and sanitation services.

CRediT authorship contribution statement

Hermilio Vilarinho: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. **Miguel Alves Pereira:** Formal analysis, Methodology, Validation, Writing – review & editing. **Giovanna D'Inverno:** Methodology, Validation, Writing – review & editing, Formal analysis. **Henriqueta Nóvoa:** Funding acquisition, Project administration, Resources, Supervision, Validation, Formal analysis, Methodology, Writing – review & editing. **Ana S. Camanho:** Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work is financed by National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within the project eduBEST, with reference 2022.08686.PTDC. Hermilio Vilarinho acknowledges the financial support provided by FCT through PhD research grant 2021.05244.BD and INESC TEC – Institute for Systems and Computer Engineering. Miguel Alves Pereira also thanks the FCT for supporting this research via project UIDB/00097/2020. Giovanna D'Inverno acknowledges financial support from the University of Pisa, Italy (PRA 2023 Project).

Appendix A. Clarification on the directional distance function BoD model

The Directional Distance Function (DDF) models are built with an objective of simultaneously expanding outputs and contracting inputs in accordance to a directional vector. Chung et al. [101] proposed Model (A.1) including also undesirable outputs. In this model, i represents the number of inputs, ranging from 1 to m, k denotes the number of undesirable outputs ranging from 1 to l, r represents the desirable outputs ranging from 1 to s, and j stands for the number of DMUs ranging from 1 to n. The variables are represented as follows: x_{ij} for inputs, b_{kj} for undesirable outputs, and y_{rj} for desirable ones. Vector g, presenting the components g_x, g_b and g_y, indicate the direction of change towards the model's objective. j_0 represents the DMU under assessment. The objective function parameter θ_{j_0} expresses the simultaneous contraction of inputs and undesirable outputs and the expansion of desirable outputs. The vector λ_j signifies the intensity variables, indicating the extent to which all DMUs contribute to the assessment of j_0.

$$
\begin{aligned}
\text{maximise} \quad & \theta_{j_0} \\
\text{subject to} \quad & \sum_{j=1}^{n} x_{ij}\lambda_j \le x_{rj_0} - \theta_{j_0} g_x, && i = 1,\dots,m \\
& \sum_{j=1}^{n} b_{kj}\lambda_j = b_{kj_0} - \theta_{j_0} g_b, && k = 1,\dots,l \\
& \sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{rj_0} - \theta_{j_0} g_y, && r = 1,\dots,s \\
& \lambda_j \ge 0, && j = 1,\dots,n \\
& \theta_{j_0} \in \mathbb{R}
\end{aligned}
\tag{A.1}
$$

In this model, the equality constraint indicates that the undesirable outputs are weakly disposable. This implies that an improvement in undesirable outputs is only possible if a desirable output is reduced or an input is increased. Inputs and desirable outputs are assumed to be strongly disposable.

Zanella et al. [13] presented an adapted version of Model (A.1), transforming it into a BoD model as shown in (A.2). This adaptation involves considering a unitary level for input. Now, the vector g has only two dimensions, namely g_b and g_y.

$$
\begin{aligned}
\text{maximise} \quad & \theta_{j_0} \\
\text{subject to} \quad & \sum_{j=1}^{n} \lambda_j \le 1, \\
& \sum_{j=1}^{n} b_{kj}\lambda_j = b_{kj_0} - \theta_{j_0} g_b, && k = 1,\dots,l \\
& \sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{rj_0} - \theta_{j_0} g_y, && r = 1,\dots,s \\
& \lambda_j \ge 0, && j = 1,\dots,n \\
& \theta_{j_0} \in \mathbb{R}
\end{aligned}
\tag{A.2}
$$

Furthermore, Zanella et al. [13] introduced an enhancement to Model (A.2), presented as (A.3). This modification involves transforming the equality constraint into an inequality, relaxing the weak disposability property. This adaptation enables the classification of a DMU as best-performing when no further improvements in both desirable and undesirable metrics are achievable. Additionally, the constraint $\sum_{j=1}^{n} \lambda_j \le 1$ is converted to an equality to allow the aggregation of performance indicators expressed as ratios, as required in composite indicators. A detailed explanation of the assumptions behind this enhancement can be found in [13].

$$
\begin{aligned}
\text{maximise} \quad & \theta_{j_0} \\
\text{subject to} \quad & \sum_{j=1}^{n} \lambda_j = 1, \\
& \sum_{j=1}^{n} b_{kj}\lambda_j \le b_{kj_0} - \theta_{j_0} g_b, && k = 1,\dots,l \\
& \sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{rj_0} - \theta_{j_0} g_y, && r = 1,\dots,s \\
& \lambda_j \ge 0, && j = 1,\dots,n \\
& \theta_{j_0} \in \mathbb{R}
\end{aligned}
\tag{A.3}
$$

By the duality theory, the same model can be written in its dual form as Model (A.4). The three constraints in (A.3) are associated to the decision variables v, p_k and u_r in Model (A.4).

$$
\begin{aligned}
\text{minimise} \quad & -\sum_{r=1}^{s} y_{rj_0} u_r + \sum_{k=1}^{l} b_{kj_0} p_k + v \\
\text{subject to} \quad & \sum_{r=1}^{s} g_y u_r + \sum_{k=1}^{l} g_b p_k = 1 \\
& -\sum_{r=1}^{s} y_{rj} u_r + \sum_{k=1}^{l} b_{kj} p_k + v \ge 0, && j = 1,\dots,n \\
& u_r \ge 0, && r = 1,\dots,s \\
& p_k \ge 0, && k = 1,\dots,l \\
& v \in \mathbb{R}
\end{aligned}
\tag{A.4}
$$

Model (A.4) is equivalent to Model (1) employed in the calculation of the composite indicator in this study. The variables u_r and p_k in the dual model correspond to the weights associated to the desirable outputs (y_{rj}) and undesirable outputs (b_{kj}). The variable v corresponds to the weight associated with the dummy input, which can be interpreted as a "helmsman", endeavouring to guide the DMU under assessment towards an optimised condition.

Appendix B. Results for water supply utilities - WUSQI and contributions of each dimension.

See Table B.1.

Appendix C. Results for wastewater utilities - water utility service quality index (WUSQI) and contributions of each dimension.

See Table C.1.

Table B.1
Water supply utilities (WUSQI) and contributions of each dimension.

Company	Year	CI	Relative importance of WUSQI dimensions				
			Inclusiveness	Reliability	Safety	Responsiveness	Transparency
AdSA	2016	0.952	80.0%	5.0%	5.0%	5.0%	5.0%
	2017	0.958	80.0%	5.0%	5.0%	5.0%	5.0%
	2018	0.957	80.0%	5.0%	5.0%	5.0%	5.0%
	2019	0.958	80.0%	5.0%	5.0%	5.0%	5.0%
	2020	0.959	80.0%	5.0%	5.0%	5.0%	5.0%
	2021	0.955	80.0%	5.0%	5.0%	5.0%	5.0%
AdA	2016	0.972	10.0%	5.0%	75.0%	5.0%	5.0%
	2017	0.970	10.0%	5.0%	75.0%	5.0%	5.0%
	2018	0.846	80.0%	5.0%	5.0%	5.0%	5.0%
	2019	0.984	10.0%	5.0%	65.0%	15.0%	5.0%
	2020	0.977	10.0%	5.0%	75.0%	5.0%	5.0%
	2021	0.986	10.0%	5.0%	65.0%	15.0%	5.0%
AdCL	2016	0.965	10.0%	5.0%	75.0%	5.0%	5.0%
	2017	0.978	10.0%	5.0%	65.0%	15.0%	5.0%
	2018	0.982	70.0%	5.0%	5.0%	15.0%	5.0%
	2019	0.856	80.0%	5.0%	5.0%	5.0%	5.0%
	2020	0.875	70.0%	5.0%	5.0%	15.0%	5.0%
	2021	0.875	70.0%	5.0%	5.0%	15.0%	5.0%
AdDP	2017	0.993	70.0%	5.0%	5.0%	15.0%	5.0%
	2018	0.998	70.0%	5.0%	5.0%	15.0%	5.0%
	2019	0.995	70.0%	5.0%	5.0%	15.0%	5.0%
	2020	0.996	70.0%	5.0%	5.0%	15.0%	5.0%
	2021	1.000	70.0%	5.0%	5.0%	15.0%	5.0%
AdN	2016	0.884	10.0%	5.0%	75.0%	5.0%	5.0%
	2017	0.942	10.0%	5.0%	75.0%	5.0%	5.0%
	2018	0.840	10.0%	5.0%	65.0%	15.0%	5.0%
	2019	0.961	10.0%	5.0%	65.0%	15.0%	5.0%
	2020	0.961	10.0%	5.0%	65.0%	15.0%	5.0%
	2021	0.966	10.0%	5.0%	65.0%	15.0%	5.0%
AdVT	2016	0.901	10.0%	5.0%	65.0%	15.0%	5.0%
	2017	0.840	10.0%	5.0%	65.0%	15.0%	5.0%
	2018	0.841	10.0%	5.0%	65.0%	15.0%	5.0%
	2019	0.900	10.0%	5.0%	65.0%	15.0%	5.0%
	2020	0.902	10.0%	5.0%	65.0%	15.0%	5.0%
	2021	0.842	10.0%	5.0%	65.0%	15.0%	5.0%
AdVouga	2016	0.938	10.0%	5.0%	75.0%	5.0%	5.0%
	2017	0.935	10.0%	5.0%	75.0%	5.0%	5.0%
	2018	0.945	10.0%	5.0%	5.0%	5.0%	75.0%
	2019	0.945	10.0%	5.0%	5.0%	5.0%	75.0%
	2020	0.943	10.0%	5.0%	5.0%	5.0%	75.0%
	2021	0.941	10.0%	5.0%	5.0%	5.0%	75.0%
AgDA	2016	0.856	70.0%	5.0%	5.0%	15.0%	5.0%
	2017	0.843	70.0%	5.0%	5.0%	15.0%	5.0%
	2018	0.841	70.0%	5.0%	5.0%	15.0%	5.0%
	2019	0.886	80.0%	5.0%	5.0%	5.0%	5.0%
	2020	0.896	70.0%	5.0%	5.0%	15.0%	5.0%
	2021	0.917	80.0%	5.0%	5.0%	5.0%	5.0%
EPAL	2016	0.851	10.0%	5.0%	65.0%	15.0%	5.0%
	2017	0.850	10.0%	5.0%	65.0%	15.0%	5.0%
	2018	0.976	10.0%	5.0%	65.0%	15.0%	5.0%
	2019	0.975	10.0%	5.0%	65.0%	15.0%	5.0%
	2020	0.972	10.0%	5.0%	65.0%	15.0%	5.0%
	2021	0.974	10.0%	5.0%	65.0%	15.0%	5.0%
ICOVI	2016	0.939	80.0%	5.0%	5.0%	5.0%	5.0%
	2017	0.944	80.0%	5.0%	5.0%	5.0%	5.0%
	2018	0.944	80.0%	5.0%	5.0%	5.0%	5.0%
	2019	0.945	80.0%	5.0%	5.0%	5.0%	5.0%
	2020	0.946	80.0%	5.0%	5.0%	5.0%	5.0%
	2021	0.830	80.0%	5.0%	5.0%	5.0%	5.0%

Table C.1
Wastewater utilities - WUSQI and contributions of each dimension.

Company	Year	CI	Relative importance of WUSQI dimensions				
			Inclusiveness	Reliability	Safety	Responsiveness	Transparency
AdSerra	2016	0.992	35.10%	38.97%	5.00%	5.09%	15.83%
	2017	0.997	34.08%	39.03%	5.65%	5.10%	16.14%
	2018	1.000	34.08%	39.03%	5.65%	5.10%	16.14%
	2019	0.999	34.08%	39.03%	5.65%	5.10%	16.14%
	2020	1.000	10.00%	38.30%	5.00%	5.50%	41.20%
	2021	1.000	35.10%	38.97%	5.00%	5.09%	15.83%
AdSA	2016	1.000	37.12%	38.57%	5.00%	5.00%	14.31%
	2017	1.000	32.57%	46.37%	5.00%	5.00%	11.06%
	2018	0.932	10.00%	5.00%	5.00%	5.00%	75.00%
	2019	1.000	40.83%	44.17%	5.00%	5.00%	5.00%
	2020	0.996	40.83%	44.17%	5.00%	5.00%	5.00%
	2021	0.922	80.00%	5.00%	5.00%	5.00%	5.00%
AdA	2016	0.906	80.00%	5.00%	5.00%	5.00%	5.00%
	2017	0.894	70.00%	5.00%	5.00%	15.00%	5.00%
	2018	0.908	80.00%	5.00%	5.00%	5.00%	5.00%
	2019	0.872	10.00%	5.00%	75.00%	5.00%	5.00%
	2020	0.888	10.00%	5.00%	65.00%	15.00%	5.00%
	2021	0.881	10.00%	5.00%	75.00%	5.00%	5.00%
AdCL	2016	0.950	80.00%	5.00%	5.00%	5.00%	5.00%
	2017	0.941	10.00%	5.00%	75.00%	5.00%	5.00%
	2018	0.917	10.00%	5.00%	75.00%	5.00%	5.00%
	2019	0.928	10.00%	5.00%	75.00%	5.00%	5.00%
	2020	0.929	10.00%	5.00%	65.00%	15.00%	5.00%
	2021	0.886	70.00%	5.00%	5.00%	15.00%	5.00%
AdN	2016	0.902	10.00%	5.00%	75.00%	5.00%	5.00%
	2017	0.890	10.00%	5.00%	75.00%	5.00%	5.00%
	2018	0.919	10.00%	5.00%	65.00%	15.00%	5.00%
	2019	0.915	10.00%	5.00%	65.00%	15.00%	5.00%
	2020	0.911	10.00%	5.00%	65.00%	15.00%	5.00%
	2021	0.904	10.00%	5.00%	65.00%	15.00%	5.00%
AdTA	2017	0.929	80.00%	5.00%	5.00%	5.00%	5.00%
	2018	0.949	10.00%	5.00%	65.00%	15.00%	5.00%
	2019	0.935	70.00%	5.00%	5.00%	15.00%	5.00%
	2020	0.929	10.00%	5.00%	65.00%	15.00%	5.00%
	2021	0.925	70.00%	5.00%	5.00%	15.00%	5.00%
AdVT	2016	0.951	70.00%	5.00%	5.00%	15.00%	5.00%
	2017	0.959	10.00%	31.46%	38.54%	15.00%	5.00%
	2018	0.939	10.00%	5.00%	65.00%	15.00%	5.00%
	2019	0.993	11.68%	32.12%	36.20%	15.00%	5.00%
	2021	1.000	11.68%	32.12%	36.20%	15.00%	5.00%
	2020	0.986	10.25%	40.55%	5.00%	15.00%	29.21%
AgDA	2016	0.891	70.00%	5.00%	5.00%	15.00%	5.00%
	2017	0.874	70.00%	5.00%	5.00%	15.00%	5.00%
	2018	0.846	70.00%	5.00%	5.00%	15.00%	5.00%
	2019	0.849	70.00%	5.00%	5.00%	15.00%	5.00%
	2020	0.800	80.00%	5.00%	5.00%	5.00%	5.00%
	2021	0.878	70.00%	5.00%	5.00%	15.00%	5.00%
AMTSM	2016	1.000	77.87%	5.00%	5.00%	5.00%	7.13%
	2017	0.845	74.76%	5.00%	6.28%	5.00%	8.96%
	2018	0.836	74.30%	5.00%	5.00%	5.00%	10.70%
	2019	0.828	80.00%	5.00%	5.00%	5.00%	5.00%
	2020	0.927	80.00%	5.00%	5.00%	5.00%	5.00%
	2021	1.000	80.00%	5.00%	5.00%	5.00%	5.00%
SIMARSUL	2017	0.917	70.00%	5.00%	5.00%	15.00%	5.00%
	2018	0.923	70.00%	5.00%	5.00%	15.00%	5.00%
	2019	0.892	70.00%	5.00%	5.00%	15.00%	5.00%
	2020	0.878	70.00%	5.00%	5.00%	15.00%	5.00%
	2021	0.912	70.00%	5.00%	5.00%	15.00%	5.00%
SIMDOURO	2017	0.939	10.00%	5.00%	75.00%	5.00%	5.00%
	2018	0.954	10.00%	29.51%	50.49%	5.00%	5.00%
	2019	0.939	10.00%	5.00%	75.00%	5.00%	5.00%
	2020	1.000	10.00%	29.51%	50.49%	5.00%	5.00%
	2021	0.958	10.00%	5.00%	65.00%	15.00%	5.00%
TRATAVE	2016	0.970	80.00%	5.00%	5.00%	5.00%	5.00%
	2017	0.988	70.00%	5.00%	5.00%	15.00%	5.00%
	2018	0.988	70.00%	5.00%	5.00%	15.00%	5.00%
	2019	0.991	70.00%	5.00%	5.00%	15.00%	5.00%
	2020	0.996	10.00%	5.00%	5.00%	5.00%	75.00%
	2021	1.000	49.03%	35.97%	5.00%	5.00%	5.00%

References

[1] Harvey L, Green D. Defining quality. Assess Eval High Educ 1993;18(1):9–34.
[2] Reeves CA, Bednar DA. Defining quality: Alternatives and implications. Acad Manag Rev 1994;19(3):419–45.
[3] Garvin DA. What does "product quality" really mean? Sloan Manage Rev 1984;25:25–43.
[4] van Kemenade E, Hardjono TW. Twenty-first century total quality management: The emergence paradigm. TQM J 2019;31(2):150–66.
[5] Sousa R, Voss CA. Quality management re-visited: A reflective review and agenda for future research. J Oper Manage 2002;20(1):91–109.
[6] Gupta A, McDaniel JC, Herath SK. Quality management in service firms: Sustaining structures of total quality service. Manag Serv Qual 2005;15(4):389–402.
[7] Rowley J. Quality measurement in the public sector: Some perspectives from the service quality literature. Total Qual Manag 1998;9(2–3):321–33.
[8] Salleh A, Othman N, et al. An importance-performance analysis of sustainable service quality in water and sewerage companies. Ind Eng Manag Syst 2019;18(1):89–103.
[9] Sala-Garrido R, Mocholí-Arce M, Molinos-Senante M. Assessing the quality of service of water companies: a 'benefit of the doubt' composite indicator. Soc Indic Res 2021;155(1):371–87.
[10] Margerum RD, Robinson CJ. Collaborative partnerships and the challenges for sustainable water management. Curr Opin Environ Sustain 2015;12:53–8.
[11] Lombana Cordoba C, Saltiel G, Sadik N, Perez Penalosa F. Utility of the future, taking water and sanitation utilities beyond the next level. A methodology to ignite transformation in water and sanitation utilities. Washington, DC: World Bank; 2021.
[12] Nardo M, Saltelli A, Tarantola S, Hoffman A, Giovannini E. Handbook on constructing composite indicators: methodology and user guide. OECD and European Union and Joint Research Centre-European Commission; 2008, p. 162.
[13] Zanella A, Camanho AS, Dias TG. Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. European J Oper Res 2015;245(2):517–30.
[14] Cherchye L, Moesen W, Rogge N, Puyenbroeck TV. An introduction to 'benefit of the doubt' composite indicators. Soc Indic Res 2007;82(1):111–45.
[15] Figueira J, Roy B. Determining the weights of criteria in the ELECTRE type methods with a revised simos' procedure. European J Oper Res 2002;139(2):317–26.
[16] Corrente S, Figueira J, Greco S. Pairwise comparison tables within the deck of cards method in multiple criteria decision aiding. European J Oper Res 2021;291(2):738–56.
[17] Maziotis A, Molinos-Senante M, Sala-Garrido R. Assesing the impact of quality of service on the productivity of water industry: A Malmquist–Luenberger approach for England and Wales. Water Resour Manag 2017;31:2407–27.
[18] Molinos-Senante M, Delgado-Antequera L, Gómez T. Measuring the quality of service of water companies: A two-stage goal programming synthetic index proposal. Soc-Econ Plan Sci 2022;79.
[19] Forker LB, Vickery SK, Droge CL. The contribution of quality to business performance. Int J Oper Prod Manage 1996;16(8):44–62.
[20] Contreras CEM. La calidad del servicio y la satisfacción del consumidor. Remark-Rev Brasil Mark 2011;10(2):146–62.
[21] Gorla N, Somers TM, Wong B. Organizational impact of system quality, information quality, and service quality. J Strateg Inf Syst 2010;19(3):207–28.
[22] Janahi MA, Al Mubarak MMS. The impact of customer service quality on customer satisfaction in islamic banking. J Islam Mark 2017;8(4):595–604.
[23] Hjelt M, Syynimaa N. Quality management in service desk: How does service desk managers define and measure quality. In: International conference on enterprise information systems. SCITEPRESS Science And Technology Publications; 2018.
[24] Adam EE, Corbett LM, Flores BE, Harrison NJ, Lee T, Rho B-H, Ribera J, Samson D, Westbrook R. An international study of quality improvement approach and firm performance. Int J Oper Prod Manage 1997;17(9):842–73.
[25] Sahney S, Banwet DK, Karunes S. Conceptualizing total quality management in higher education. TQM Mag 2004;16(2):145–59.
[26] Golder PN, Mitra D, Moorman C. What is quality? An integrative framework of processes and states. J Mark 2012;76(4):1–23.
[27] Elshaer I. What is the meaning of quality? Technical report, University Library of Munich, Germany; 2012.
[28] Kenyon GN, Sen KC. Perception of quality. Springer; 2016.
[29] Harvey J. Service quality: A tutorial. J Oper Manage 1998;16(5):583–97.
[30] Wen D, Sun X, Yan D. The quality movement: Where are we going? Past, present and future. Total Qual Manage Bus Excell 2022;33(1–2):92–112.
[31] Zhang C, Moreira MR, Sousa PS. A bibliometric view on the use of total quality management in services. Total Qual Manage Bus Excell 2021;32(13–14):1466–93.
[32] Carnerud D. Exploring research on quality and reliability management through text mining methodology. Int J Qual Reliab Manag 2017.
[33] Lo Q-Q, Chai K-H. Quantitative analysis of quality management literature published in total quality management and business excellence (1996–2010). Total Qual Manage Bus Excell 2012;23(5–6):629–51.
[34] Sánchez-Franco MJ, Calvo-Mora A, Periáñez-Cristobal R. Clustering abstracts from the literature on quality management (1980–2020). Total Qual Manage Bus Excell 2022.
[35] Prasad RK, Verma S. Service quality measures: Systematic literature review and future research directions. Int J Manag Pract 2022;15(1):9–23.
[36] Parasuraman A, Zeithaml VA, Berry LL. SERVQUAL: a multi-item scale for measuring consumer perceptions of service. J Retail 1988;64(1):12–40.
[37] Saleh F. Analysing service quality in the hospitality industry using the SERVQUAL model. Serv Ind J 1991;11(3):324–45.
[38] Reidenbach R, Sandifer-Smallwood B. Exploring perceptions of hospital operations by a modified SERVQUAL approach. J Health Care Mark 1990;10(4):47–55.
[39] Newman K. Interrogating SERVQUAL: A critical assessment of service quality measurement in a high street retail bank. Int J Bank Mark 2001;19(3):126–39.
[40] Awasthi A, Chauhan S, Omrani H, Panahi A. A hybrid approach based on SERVQUAL and fuzzy TOPSIS for evaluating transportation service quality. Comput Ind Eng 2011;61(3):637–46.
[41] Pakdil F, Aydin O. Expectations and perceptions in airline services: An analysis using weighted SERVQUAL scores. J Air Transp Manag 2007;13(4):229–37.
[42] Tan K, Kek S. Service quality in higher education using an enhanced SERVQUAL approach. Qual High Educ 2004;10(1):17–24.
[43] Bojanic D, Rosen L. Measuring service quality in restaurants: An application of the SERVQUAL instrument. J Hosp Tour Res 1994;18(1):3–14.
[44] Wisniewski M. Measuring service quality in the public sector: The potential for SERVQUAL. Total Qual Manag 1996;7(4):357–66.
[45] Wisniewski M. Using SERVQUAL to assess customer satisfaction with public sector services. Manag Serv Qual: Int J 2001;11(6):380–8.
[46] Asubonteng P, Mccleary K, Swan J. SERVQUAL revisited: A critical review of service quality. J Serv Mark 1996;10(6):62–81.
[47] Ladhari R. A review of twenty years of SERVQUAL research. Int J Qual Serv Sci 2009;1(2):172–98.
[48] Cronin Jr JJ, Taylor SA. Measuring service quality: a reexamination and extension. J Mark 1992;56(3):55–68.
[49] Solves JM, Aina JB, Rodríguez-Marín J, Andrés JA. Calidad percibida del cuidado hospitalario. Gac Sanit 1997;11(4):176–89.
[50] Mei AWO, Dean AM, White CJ. Analysing service quality in the hospitality industry. Manag Serv Qual: Int J 1999;9(2):136–43.
[51] Abdullah F. The development of HEdPERF: a new measuring instrument of service quality for the higher education sector. Int J Consum Stud 2006;30(6):569–81.
[52] Parasuraman A, Zeithaml VA, Malhotra A. ES-QUAL: A multiple-item scale for assessing electronic service quality. J Serv Res 2005;7(3):213–33.
[53] Dryden J. Do we care what users want? Evaluating user satisfaction and the LibQUAL+™ experience. J Arch Organ 2005;2(4):83–8.
[54] Picazo-Tadeo AJ, Sáez-Fernández FJ, González-Gómez F. Does service quality matter in measuring the performance of water utilities? Util Policy 2008;16(1):30–8.
[55] Saal DS, Parker D. Productivity and price performance in the privatized water and sewerage companies of England and Wales. J Regul Econ 2001;20(1):61–90.
[56] Woodbury K, Dollery B. Efficiency measurement in Australian local government: The case of new south Wales municipal water services. Rev Policy Res 2004;21(5):615–36.
[57] Kumar S, Managi S. Service quality and performance measurement: Evidence from the Indian water sector. Int J Water Resour Dev 2010;26(2):173–91.
[58] Molinos-Senante M, Maziotis A. Assessing the influence of exogenous and quality of service variables on water companies' performance using a true-fixed stochastic frontier approach. Urban Water J 2018;15(7):682–91.
[59] D'Inverno G, Carosi L, Romano G. Environmental sustainability and service quality beyond economic and financial indicators: A performance evaluation of Italian water utilities. Soc-Econ Plan Sci 2021;75:100852.
[60] Mocholi-Arce M, Sala-Garrido R, Molinos-Senante M, Maziotis A. Performance assessment of water companies: A metafrontier approach accounting for quality of service and group heterogeneities. Soc-Econ Plan Sci 2021;74:100948.
[61] Henriques AA, Fontes M, Camanho AS, D'Inverno G, Amorim P, Silva JG. Performance evaluation of problematic samples: A robust nonparametric approach for wastewater treatment plants. Ann Oper Res 2022.
[62] Molinos-Senante M, Maziotis A, Villegas A. Performance analysis of Chilean water companies after the privatization of the industry: The influence of ownership. Water Int 2022;47(1):114–31.
[63] Chang J, Li W, Zhou Y, Zhang P, Zhang H. Impact of public service quality on the efficiency of the water industry: evidence from 147 cities in China. Sustainability 2022;14(22):15160.
[64] Duarte AA, Rodrigues GMdC, Ramos RA. A global service quality index to evaluate the performance and sustainability in water supply utilities. WSEAS Trans Environ Dev 2009;5(12):759–69.
[65] Pinto FS, Costa AS, Figueira JR, Marques RC. The quality of service: An overall performance assessment for water utilities. Omega (United Kingdom) 2017;69:115–25.
[66] Molinos-Senante M, Muñoz S, Chamorro A. Assessing the quality of service for drinking water supplies in rural settings: A synthetic index approach. J Environ Manag 2019;247:613–23.

[67] Krause MK, Cabrera Jr E, Cubillo F, Diaz C, Ducci J. Aquarating: an international standard for assessing water and wastewater services. IWA Publishing; 2018.

[68] Karnib A. Evaluation of the quality of service of the water supply delivery in Lebanon. J Water Sanitation Hyg Dev 2015;5(1):17–27.

[69] Molinos-Senante M, Gómez T, Caballero R, Sala-Garrido R. Assessing the quality of service to customers provided by water utilities: A synthetic index approach. Ecol Indic 2017;78:214–20.

[70] Palomero-González J, Almenar-Llongo V, Fuentes-Pascual R. A composite indicator index as a proxy for measuring the quality of water supply as perceived by users for urban water services. Technol Forecast Soc Change 2022;174:121300.

[71] Abubakar IR. Quality dimensions of public water services in Abuja, Nigeria. Util Policy 2016;38:43–51.

[72] Ohwo O, Agusomu TD. Residential customers satisfaction with public water provision in ojota, Nigeria. Eur Sci J ESJ 2018;14(23):119–20.

[73] Ammar SM, Saleh YA. Grey relational evaluation of bulk water provision service quality using the SERVQUAL framework. Int J Public Adm 2021.

[74] Murrar A, Batra M, Rodger J. Service quality and customer satisfaction as antecedents of financial sustainability of the water service providers. TQM J 2020;33(8):1867–85.

[75] Kassa K, Chernet M, Kelemework G, Zewde B, Woldemedhin A. Customer satisfaction survey: The case of urban water supply services in southern ethiopia. Water Pract Technol 2017;12(4):1009–17.

[76] Kumasi T, Agbemor BD. Tracking user satisfaction of rural water services in northern ghana. J Water Sanitation Hyg Dev 2018;8(2):307–18.

[77] Tessema MM. Assessment of customer's satisfaction on water supply service delivery in bahir dar city, Ethiopia. Int J Eng Res Adv Technol 2020;6(10):20–9.

[78] Rustinsyah R. Determining the satisfaction level of water user association service quality for supporting sustainable rural development. Dev Stud Res 2019;6(1):118–28.

[79] Alegre H, Baptista JM, Cabrera Jr H, Cubillo F, Duarte P, Hirner W, Merkel. Performance indicators for water supply services. 3rd ed.. London: IWA Publishing; 2017, p. 1–404.

[80] Ersar, Lnec. Guia de avaliação da qualidade dos serviços de águas e resíduos prestados aos utilizadores – 3. a geração do sistema de avaliação. - guia técnico 22. 2021, Issued by ERSAR - Entidade Reguladora dos Serviços de Águas e Resíduos and LNEC - Laboratório Nacional de Engenharia Civil. Available on http://www.ersar.pt/pt/publicacoes/publicacoes-tecnicas/guias.

[81] ERSAR. Relatório anual dos serviços de águas e resíduos em Portugal (2021): Volume 1 - caracterização do setor de águas e resíduos. 2021.

[82] Mergoni A, D'Inverno G, Carosi L. A composite indicator for measuring the environmental performance of water, wastewater, and solid waste utilities. Util Policy 2022;74:101285.

[83] Pereira MA, Vilarinho H, D'Inverno G, Camanho AS. A regulatory robust conditional approach to measuring the efficiency of wholesale water supply and wastewater treatment services. Util Policy 2023;83:101594.

[84] Vilarinho H, D'Inverno G, Nóvoa H, Camanho AS. The measurement of asset management performance of water companies. Soc-Econ Plan Sci 2023;87:101545.

[85] Vilarinho H, D'Inverno G, Nóvoa H, Camanho AS. Performance analytics for regulation in retail water utilities: Guiding asset management by identifying peers and targets. Util Policy 2023;82:101559.

[86] Lombana Cordoba C, Saltiel G, Penalosa FP. Utility of the future 2.0, taking water and sanitation utilities beyond the next level. A methodology to ignite transformation in water and sanitation utilities. Washington DC: World Bank; 2022.

[87] Melyn W, Moesen W. Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available, public economics research paper 17. KU Leuven: Center for Economic Studies; 1991.

[88] Rogge N, Jaeger SD, Lavigne C. Waste performance of NUTS 2-regions in the EU: A conditional directional distance benefit-of-the-doubt model. Ecol Econom 2017;139:19–32.

[89] Allen R, Athanassopoulos A, Dyson RG, Thanassoulis E. Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions. Ann Oper Res 1997;73:13–34.

[90] D'Inverno G, De Witte K. Service level provision in municipalities: A flexible directional distance composite indicator. European J Oper Res 2020;286(3):1129–41.

[91] Henriques AA, Camanho AS, Amorim P, Silva JG. Performance benchmarking using composite indicators to support regulation of the portuguese wastewater sector. Util Policy 2020;66:101082.

[92] Pereira MA, Camanho AS, Figueira JR, Marques RC. Incorporating preference information in a range directional composite indicator: The case of portuguese public hospitals. European J Oper Res 2021;294(2):633–50.

[93] Kuosmanen T, et al. Modeling blank data entries in data envelopment analysis. Econ-WPA working paper at WUSTL, 2002, 210001.

[94] Morais P, Camanho AS. Evaluation of performance of European cities with the aim to promote quality of life improvements. Omega 2011;39(4):398–409.

[95] Zanella A, Camanho AS, Dias TG. Benchmarking countries' environmental performance. J Oper Res Soc 2013;64(3):426–38.

[96] Montgomery DC. Introduction to statistical quality control. John Wiley & Sons; 2012.

[97] Charnes A, Cooper W, Thrall RM. A structure for classifying and characterizing efficiency and inefficiency in data envelopment analysis. J Prod Anal 1991;2(3):197–237.

[98] Bowlin WF. Measuring performance: An introduction to data envelopment analysis (DEA). J Cost Anal 1998;15(2):3–27.

[99] Sarkis J. Preparing your data for DEA. In: Modeling data irregularities and structural complexities in data envelopment analysis. Springer; 2007, p. 305–20.

[100] Lavigne C, Jaeger SD, Rogge N. Identifying the most relevant peers for benchmarking waste management performance: A conditional directional distance benefit-of-the-doubt approach. Waste Manag 2019;89:418–29.

[101] Chung YH, Färe R, Grosskopf S. Productivity and undesirable outputs: A directional distance function approach. J Environ Manag 1997;51(3):229–40.

Hermilio Vilarinho is a Post-doctoral Researcher at the Institute for Systems and Computer Engineering, Technology and Science and an Invited Assistant Professor at the Faculty of Engineering (FEUP) of the University of Porto (Portugal). He holds a Ph.D. in Industrial Engineering and Management from FEUP (2023). He has a degree in Mechanical Engineering from the Federal University of Bahia (Brazil) and in Business Administration from the Catholic University of Salvador (Brazil). He has worked as an Engineering professional in leading positions in Brazil, USA, Angola and Japan, with emphasis on Quality Management and Project Management.

Miguel Alves Pereira is a tenure-track Assistant Professor in Systems Engineering and Management at the Department of Engineering and Management of the Instituto Superior Técnico - Universidade de Lisboa, as well as an Integrated Member of the Centre for Management Studies at the same institution. His main interests concern analytics and operational research / management science, emphasising the development of efficiency, productivity, and decision analysis models, and their applications to health, water and sanitation, sustainable development, energy, and climate change. He is the author of 19 scientific publications, as well as an ad hoc referee in 53 international peer-reviewed journals.

Giovanna D'Inverno is a Tenure-track Assistant Professor in Mathematical Methods for Economics, Actuarial Science and Finance at the Department of Economics and Management, University of Pisa (Italy). She is also Research Associate in the Leuven Economics of Education Research center at KU Leuven (Belgium). She has a joint PhD degree (2018) from IMT School for Advanced Studies Lucca (Italy) and KU Leuven. She is author of several publications in top-ranked journals. Her research focuses on the development of quantitative methods and policy evaluation techniques to address policy relevant issues. Recurrent empirical applications mostly cover public economic issues and sustainability challenges.

Henriqueta Nóvoa is an Assistant Professor at the Faculty of Engineering (FEUP) of the University of Porto (Portugal). She holds a Ph.D. in Engineering Sciences from FEUP (2000) and a first degree in Electrical Engineering from FEUP. She has worked as a production test engineer at Texas Instruments for 6 years. She has been teaching courses in Quality Management, Statistics and Strategic Planning of Information Systems since 1990. She has been involved in research projects in diverse areas, such as wine industry, construction industry and packaging. She published more than 20 refereed papers in international conferences and scientific journals.

Ana S. Camanho is Vice-Rector of the University of Porto and Associate Professor with Habilitation at the Faculty of Engineering of the University of Porto (FEUP). She has a degree in Industrial Engineering and Management from FEUP and a Ph.D. in Industrial and Business Studies from Warwick Business School, UK. She is the author of over 100 scientific publications in Operations Research and Data Science. Her research focuses on the development of quantitative methods for evaluating efficiency and productivity using Data Envelopment Analysis. She was vice-president of the Portuguese Operational Research Society (2011–2012), Pro-Director of FEUP (2018-2022), and Director of the Bachelor and Master Degrees in Industrial Engineering and Management of FEUP. She is Associate Editor of the Journal of Productivity Analysis, RAIRO-Operations Research, and Sustainability Analytics and Modeling (Elsevier).