Revista Brasileira de Ciências Ambientais
Brazilian Journal of Environmental Sciences

Fuzzy logic-based efficiency index for water supply systems in Northeastern Brazil

Índice de eficiência para sistemas de abastecimento de água no Nordeste Brasileiro baseado em lógica fuzzy

José Thiago Gomes da Silva[1], *Saulo de Tarso Marques Bezerra*[1]

ABSTRACT

Water supply systems (WSS) face high levels of water losses and energy consumption, leading to substantial operational costs and inefficient use of natural resources. This scenario prompted the enactment of Brazil's New Legal Framework for Sanitation, which established ambitious targets for service efficiency and universal access, reinforcing the need for robust and reliable performance assessment tools. To provide a comprehensive evaluation of system performance, this study develops a global efficiency index for assessing water supply services in municipalities across the Northeast Region of Brazil. The proposed methodology integrates official datasets, detects and removes outliers using the Isolation Forest algorithm, and combines multiple performance indicators through a fuzzy logic-based model, with membership functions defined according to the statistical distribution of the variables. The resulting index, termed the Water Supply System Assessment Index (IASA), revealed substantial disparities in efficiency among the states of the region, with Bahia achieving the highest average performance and Maranhão exhibiting the lowest efficiency levels. The analysis by population size showed that larger municipalities generally achieved higher energy efficiency but faced greater challenges in controlling water losses. Correlation analysis confirmed the internal consistency of the proposed model, demonstrating an inverse relationship between the IASA and water loss indicators. The findings provide public managers and water utilities with a practical decision-support tool, enabling more efficient resource allocation and supporting strategic planning aimed at improving the operational performance of WSS.

Keywords: Performance indicators; Water losses; Energy consumption; SNIS; SINISA.

RESUMO

Os sistemas de abastecimento de água enfrentam elevados índices de perdas e consumo de energia, gerando custos operacionais expressivos e desperdício de recursos naturais. Esse contexto impulsionou a promulgação do Novo Marco Legal do Saneamento no Brasil, estabelecendo metas de eficiência e universalização dos serviços, reforçando a necessidade de ferramentas de avaliação robustas e precisas. Com o objetivo de oferecer uma análise abrangente da eficiência desses sistemas, este estudo visa desenvolver um índice global de avaliação da eficiência dos serviços de abastecimento de água, focado nos municípios da Região Nordeste. A metodologia envolve consolidação de dados oficiais, tratamento de valores atípicos pelo algoritmo Isolation Forest e integração de indicadores de desempenho em um modelo baseado em lógica fuzzy, com funções de pertinência definidas pela distribuição estatística das variáveis. O índice resultante, denominado Índice de Avaliação de Sistemas de Abastecimento de Água (IASA), revelou desigualdades significativas entre os estados da região, com destaque para a Bahia, com a maior média, e o Maranhão, com os menores valores de eficiência. A análise por porte populacional indicou que municípios maiores tendem a apresentar maior eficiência energética, embora enfrentem desafios mais complexos no controle de perdas hidráulicas. A análise de correlação confirmou a coerência do modelo, evidenciando relação inversa entre o IASA e indicadores de perdas. Os resultados fornecem aos gestores públicos e operadores um instrumento de apoio à tomada de decisão, contribuindo para a alocação mais eficiente de recursos e para o planejamento de ações voltadas à melhoria do desempenho dos sistemas.

Palavras-chave: Indicadores de desempenho; Perdas de água; Consumo de energia; SNIS; SINISA.

[1]Universidade Federal de Pernambuco – Caruaru (PE), Brazil.
Corresponding author: José Thiago Gomes da Silva – Campus do Agreste, Universidade Federal de Pernambuco, Av. Marielle Franco, Km 59, Nova Caruaru – Caruaru (PE), Brazil, CEP 55014-900. E-mail: thiago.gsilva3@ufpe.br
Received on: 10/01/2025. Accepted on: 09/05/2026.
https://doi.org/10.67790/rbciamb.026.2820

Introduction

Universal and efficient access to water supply services is one of the primary objectives of public sanitation policies in Brazil. The enactment of the New Legal Framework for Basic Sanitation through Law No. 14,026/2020 has intensified the demand for greater operational efficiency among water utilities by establishing ambitious targets of providing 99% of the population with access to drinking water and 90% with access to wastewater collection and treatment by 2033. In addition, the legislation establishes goals related to service continuity, water loss reduction, and improvements in treatment efficiency (Brasil, 2020). These regulatory requirements compel service providers to continuously monitor and enhance their operational performance, reinforcing the need for transparent and reliable assessment tools.

Among the major challenges faced by water utilities, reducing water losses and improving energy efficiency in water supply systems (WSS) remain particularly critical. In Brazil, the monitoring and assessment of these services have traditionally relied on data from the National Sanitation Information System (SNIS) [*Sistema Nacional de Informações sobre Saneamento*], recently replaced by the National Basic Sanitation Information System (SINISA) [*Sistema Nacional de Informações em Saneamento Básico*]. Data compiled by these information systems indicate that water distribution losses in Brazil remain persistently high, reflecting both structural and operational deficiencies. In developing countries such as Brazil, this situation is further exacerbated by aging infrastructure, inefficient management practices, and limited adoption of monitoring and control technologies, resulting not only in the waste of treated water but also in unnecessary energy consumption throughout the water production and distribution processes (Liemberger and Wyatt, 2019; Sousa et al., 2025).

Regarding energy consumption, SINISA reported that expenditures on electricity for WSS and wastewater systems reached R$ 9.1 billion in 2024, with total electricity consumption of 15.7 TWh. This represents approximately 2.8% of Brazil's total electricity consumption, which amounted to 561.6 TWh in the same year, according to the Statistical Yearbook of Electric Energy (Brasil, 2025; EPE, 2025).

Given this context of high water losses and substantial energy costs, robust tools are essential for accurately assessing the efficiency of these systems. The use of performance indicators and computational models enables the integration of isolated data into synthetic indices, revealing how operational deficiencies compromise the capacity for infrastructure reinvestment. Such approaches improve planning capabilities, strengthen system management, and support strategic decision-making aimed at enhancing the sustainability of water supply services (Bezerra et al., 2019; Molinos-Senante et al., 2025).

Traditionally, the evaluation of WSS efficiency has relied on individual operational indicators or deterministic frontier models. In this context, Data Envelopment Analysis (DEA) has become one of the most widely adopted approaches because it enables the comparison of multiple inputs and outputs without requiring the specification of a functional production relationship (Banker et al., 1984; Alsharif et al., 2008; Tourinho et al., 2022). DEA-based benchmarking has been extensively applied to evaluate technical efficiency as well as to investigate the effects of external costs and environmental variables on utility performance (Villegas et al., 2019; Maziotis et al., 2020; Marques et al., 2021). More recently, frontier models and non-radial mathematical approaches have also been employed to assess energy performance in the water sector (Molinos-Senante et al., 2025). However, because deterministic models assume that input data are free from measurement errors, their application to self-reported databases – which are often affected by inconsistencies, missing information, and outliers – remains challenging. Under such conditions, sensitivity to anomalous observations may undermine the robustness of the results and lead to misleading conclusions regarding system efficiency (Tourinho et al., 2022; Sousa et al., 2025).

To overcome the limitations associated with deterministic boundaries and to better handle uncertainty in operational data, fuzzy logic, originally proposed by Zadeh (1965), offers a robust computational framework by extending classical set theory. Unlike classical binary logic, in which an element either belongs or does not belong to a set, fuzzy logic allows partial membership through a continuous membership degree ranging from zero to one. Subsequently, Mamdani (1976) established the practical application of these principles by introducing a rule-based inference system founded on linguistic if–then statements.

The ability of fuzzy logic to translate expert knowledge into interpretable mathematical models has been demonstrated in the integration of multiple parameters for developing water quality indices and groundwater sustainability indicators, where uncertainty limits the applicability of conventional methods (Sarkheil et al., 2023; Zihad et al., 2023). Fuzzy inference facilitates the handling of incomplete and uncertain data, producing classifications that more accurately represent complex environmental and operational systems (Patel and Chitnis, 2022; Abidi et al., 2024; Yazdi et al., 2024). Beyond water resources, the versatility of this methodology has enabled its application in several other fields involving multi-criteria decision-making, including predictive modeling in the financial sector (Li et al., 2022) and environmental management, such as landslide susceptibility mapping (Zhang et al., 2023), forest sustainability assessment (Lazzarotto, 2018), and urban environmental monitoring (Fernández, 2021).

Despite its demonstrated versatility, studies applying fuzzy logic to evaluate the operational and energy efficiency of water utilities remain scarce. The Water Supply System Assessment Index (IASA) proposed in this study addresses this gap by integrating uncertainty treatment within a mathematical framework designed to be robust against statistical anomalies. The proposed methodology incorporates the Isolation Forest algorithm (Liu et al., 2008, 2012) to identify and remove inconsistencies commonly found in public databases (Boukerche et al., 2020). Furthermore, the model includes a temporal in-consistency penalty function based on numerical constraint-handling techniques (Deb, 2000), thereby enhancing the reliability of the efficiency assessment.

Accordingly, the objective of this study is to evaluate the efficiency of WSS in Northeastern Brazil through the development of the IASA using data obtained from the SNIS/SINISA databases. The proposed index is intended to serve as a decision-support tool for water utilities and public managers by integrating the assessment of the physical performance of the systems with the quality and consistency of sanitation reporting databases.

Methodology

This study adopts a fuzzy logic-based methodology to develop an efficiency index applicable to WSS. The methodological framework begins with the acquisition and preprocessing of publicly available data and extends to the development and application of a fuzzy inference system, followed by statistical analyses. Primary data were obtained from the historical databases of the SNIS and its successor, the SINISA, both maintained by the Brazilian Federal Government.

The entire analytical workflow was conducted using open-source software. Data consolidation and statistical analyses were performed in the R environment (version 4.5.1) using RStudio (version 2025.05), with Google Sheets initially employed for preliminary data organization. The packages FuzzyR (v2.3.2), isotree (v0.6.1.4), and corrplot (v0.95) were used to implement the fuzzy inference system, detect outliers using the Isolation Forest algorithm, and perform Spearman correlation analyses, respectively. Additional packages were employed for data management and graphical visualization.

Study area

The study area comprises the Northeast Region of Brazil, which consists of nine states and 1,794 municipalities, with more than 57.2 million inhabitants (IBGE, 2025). A substantial portion of this area lies within the so-called Brazilian semiarid region, characterized by low annual rainfall, recurrent droughts, and chronic water scarcity. These climatic conditions place continuous pressure on available water resources and require highly efficient WSS to ensure water security for the population.

Water supply services across the region exhibit considerable institutional and operational heterogeneity. Of the 1,794 municipalities located in Northeastern Brazil, SNIS/SINISA records indicate that 1,541 are served by regional or micro-regional water utilities, which constitute the focus of this study. The geographical coverage of these utilities and the companies included in the analysis are presented in Figure 1. Although the sector is predominantly operated by state-owned water utilities (Table 1), these coexist with private concessions and regional service blocks, as observed in the state of Alagoas, as well as community-based and consortium management models, such as the Integrated Rural Sanitation System (SISAR) operating in Ceará and Pernambuco and the community associations for the maintenance of services in Bahia.

Data collection and consolidation

The study analyzed data on water supply services covering the period from 2019 to 2023. Data for the years 2019–2022 were obtained from the historical SNIS database, whereas the 2023 data were extracted from the newly established SINISA. This transition reflects the modernization of Brazil's national sanitation information platform following recent regulatory reforms in the sector. Both databases are publicly available and freely accessible, allowing the long-term monitoring of service performance.

The sample comprised municipalities whose water supply services are operated by regional or micro-regional utilities, representing approximately 86% of all municipalities in Northeastern Brazil. This selection was motivated by the greater methodological standardization of these utilities and their lower susceptibility to inconsistencies in self-reported data.

Following data extraction, a statistical preprocessing procedure was conducted to ensure the quality and reliability of the analytical database. Considering the self-reported nature of the information, which may introduce inconsistencies, outlier detection was performed using the Isolation Forest algorithm. This technique was selected because of its effectiveness in identifying anomalies without requiring restrictive statistical assumptions, making it particularly suitable for the complexity of sanitation performance indicators. The algorithm isolates observations by recursively partitioning the dataset through decision trees, based on the principle that anomalous observations are numerically easier to isolate than normal data points.

The algorithm was implemented following the parameter settings recommended by Liu et al. (2008, 2012), using 100 isolation trees and a subsample size of 256 observations for each model. The threshold for classifying an observation as an outlier was statistically defined according to the three-sigma control limit. Observations with anomaly scores exceeding the mean anomaly score plus three standard deviations were classified as outliers, since values beyond this threshold are indicative of special causes rather than the natural variability inherent in the dataset (Montgomery and Runger, 2021).

Following outlier removal, a temporal consistency assessment was performed. Municipalities that did not have at least one valid observation during the five-year period for each of the five performance indicators were excluded from the analysis. For the remaining municipalities, indicator values were aggregated using population-weighted averages, thereby reflecting the demographic relevance of each municipality. Additionally, a 2.5% penalty was applied to the average value of each indicator for every year in which valid data were unavailable. This penalty, specifically defined for the proposed model, is based on the concept of penalty functions for constraint handling (Deb, 2000) and aims to discourage inconsistencies while favoring

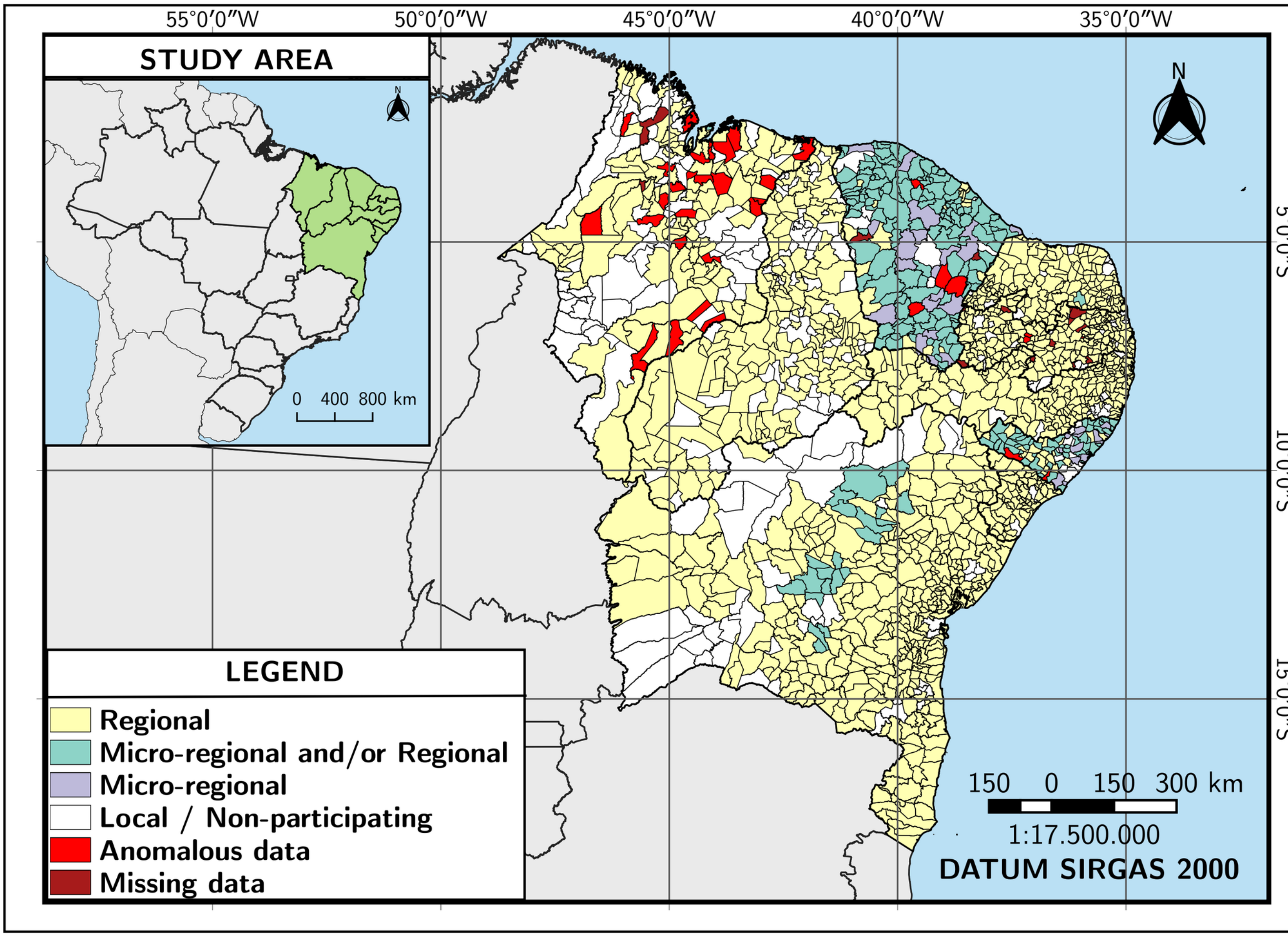

Figure 1 – Study area and geographical distribution of the analyzed water utilities

utilities that maintain regular and transparent historical records in the national public sanitation databases.

After completing the data consolidation process, 1,491 municipalities were retained for the final analysis, as summarized in Table 1.

Fuzzy model variables: performance indicators

Considering that the objective of this study is to evaluate the hydraulic and energy efficiency of the systems, the main performance indicators associated with these dimensions were selected from the SNIS and SINISA databases. Each indicator was treated as a *crisp* input variable (i.e., a deterministic numerical value) in the proposed fuzzy inference system. The performance indicators adopted in this study for the development of the fuzzy model are presented in Table 2.

The water loss indicators (IN013, IN049, and IN051) capture different dimensions of operational inefficiency, encompassing both economic losses and physical water losses. Conversely, the energy-related indicators (IN058 and IN060) assess both the amount of electricity consumed and its associated cost. Collectively, these five indicators provide a comprehensive assessment of the technical and financial performance of WSS.

The relevance of these indicators has been demonstrated in previous studies. The report published by Trata Brasil and GO Associados (2025) employed indicators IAG2012, IAG2013, and IAG2015 (formerly IN013, IN049, and IN051) to evaluate hydraulic efficiency, highlighting their effectiveness in characterizing water losses in distribution systems. Similarly, Bezerra et al. (2019) applied the same indicators to assess the performance of the systems in Northeastern Brazil, demonstrating their suitability for diagnosing water losses. In addition, the National Secretariat for Basic Sanitation annually reports indicators IN049, IN051, IN058, and IN060 in the *Water and Wastewater Services Diagnostic Report* as key references for evaluating the operational and energy efficiency of services. Equations 1–5 defining the performance indicators adopted in this study are presented below.

Table 1 – Water utilities and municipalities included in the study

State	Water utility	No. of municipalities
Alagoas	CASAL, BRK, VAA and Águas do Sertão	89
Bahia	EMBASA and Central	370
Ceará	CAGECE and SISAR	171
Maranhão	CAEMA	115
Paraíba	CAGEPA	187
Pernambuco	COMPESA	173
Piauí	AGESPISA	157
Rio Grande do Norte	CAERN	157
Sergipe	DESO	72

Table 2 – Characteristics of the performance indicators adopted in this study

SNIS Performance Indicator	SINISA Performance Indicator	Abbreviation	Unit	Description
Non-Revenue Water Index (IN013)	Non-Revenue Water Losses (IAG2012)	IPF	%	Percentage of non-revenue water relative to the total volume of water produced (excluding the volume used for utility operations).
Water Distribution Loss Index (IN049)	Total Water Losses in Distribution (IAG2013)	IPD	%	Percentage of water not delivered to consumers relative to the total volume of water produced.
Water Loss per Service Connection Index (IN051)	Total Water Losses per Service Connection (IAG2015)	IPL	L/connection/day	Annual volume of water losses in the distribution system per service connection.
Electric Energy Consumption Index for Water Supply Systems (IN058)	Average Electricity Consumption in Water Supply Services (IAG2023)	ICE	kWh/m³	Amount of electrical energy required to produce one cubic meter of treated water.
Electricity Expenditure Index for Water Supply and Wastewater Systems (IN060)	Average Electricity Expenditure in Water Supply Services (IFA2007)	IDE	R$/kWh	Annual average cost per kilowatt-hour of electricity consumed by the analyzed WSS.

$$IPF = abs[(A_{disp} - A_{fat}) \times 100 / A_{disp}] \quad (1)$$

$$IPD = (A_{disp} - A_{cons}) \times 100 / A_{disp} \quad (2)$$

$$IPL = (A_{disp} - A_{cons}) / (N_{con} / 0.365) \quad (3)$$

$$ICE = E_{water} / (A_{prod} + A_{imp}) \quad (4)$$

$$IDE = (D_{energy} \times 0.001) / (E_{water} + E_{wastewater}) \quad (5)$$

where A_{cons} is the annual volume of water consumed (m³); A_{disp} is the annual volume of water available for distribution (m³), corresponding to the volume of water produced plus the volume of imported treated water, excluding the volume used for utility operations; A_{fat} is the annual billed water volume (m³); A_{imp} is the annual volume of imported treated water (m³); A_{prod} is the annual volume of water produced (m³); D_{energy} is the annual expenditure on electricity (R$); E_{water} is the total electricity consumption of the WSS (kWh); $E_{wastewater}$ is the total electricity consumption of the wastewater system (kWh); and N_{con} is the number of active water service connections linked to the public distribution network.

The Non-Revenue Water Index (IN013) required specific preprocessing before being incorporated into the proposed model because it may assume negative values when the billed water volume exceeds the volume available for distribution. Trata Brasil and GO Associados (2025) reported that this situation occurs in municipalities where a large proportion of service connections are billed based on the minimum tariff volume regardless of actual consumption, thereby distorting the indicator. Bezerra et al. (2019) further observed that this distortion is commonly associated with small municipalities characterized by intermittent water supply and low per capita water consumption, often below the minimum billable volume. Therefore, the absolute value of the indicator was used in the proposed model so that large positive or negative values consistently represent operational inefficiency in the services.

Statistical analysis and definition of membership functions

In fuzzy logic, each element of a set is assigned a membership degree within the interval [0, 1], representing its relative degree of belonging to that set. This approach allows a continuous transition between 0 (false) and 1 (true), enabling partial membership in multiple sets with different membership degrees.

In this study, two classical types of membership functions were adopted:

Triangular membership functions, defined by the triplet (*a, b, c*), where *a* and *c* represent the lower and upper bounds of the interval over which the function assumes nonzero values, and *b* corresponds to the point at which the membership degree reaches its maximum value (1);

Trapezoidal membership functions, defined by the quadruplet (*a, b, c, d*), where *a* and *d* delimit the interval over which the function is nonzero, while the interval between *b* and *c* defines the region of full membership (membership degree equal to 1).

Five linguistic terms were defined for each performance indicator: *Very Good*, *Good*, *Fair*, Poor, and *Very Poor*. These fuzzy sets, which do not possess rigid boundaries, represent the continuous behavior of the variables and enable a linguistic interpretation of the performance of the analyzed WSS. To define the boundaries of each fuzzy set and their corresponding membership functions, the mean and standard deviation of each performance indicator were calculated from the consolidated dataset. The resulting statistical measures are presented in Table 3.

Accordingly, two approaches were adopted to define the membership function intervals. The first was applied when the lower bound was smaller than the mean minus two standard deviations, as shown in Equation 6:

Table 3 – Minimum and maximum values and descriptive statistics of the performance indicators adopted in this study

Indicator	Minimum Value (Inf)	Maximum Value (Sup)	Mean ($\bar{x}$)	Standard Deviation (σ)
IPF [%]	0.01	98.93	29.85	20.28
IPD [%]	5.50	88.38	40.53	15.13
IPL [L/con/day]	5.19	1,837.23	286.63	268.34
ICE [kWh/m³]	0.00	5.59	1.11	0.92
IDE [R$/kWh]	0.06	6.24	0.73	0.45

$$inf < (\bar{x} - 2 \cdot \sigma) \quad (6)$$

whereas the second was applied when the lower bound exceeded the mean minus two standard deviations, according to Equation 7:

$$inf > (\bar{x} - 2 \cdot \sigma) \quad (7)$$

In the first case, when the lower bound (inf) is smaller than the mean minus two standard deviations, the maximum membership degree of the Fair membership function is centered at the sample mean, whereas the peaks of the remaining membership functions are defined by values obtained from the mean minus one or two standard deviations. Under this configuration, the Very Good and Very Poor membership functions are trapezoidal, whereas the remaining functions are triangular, as summarized in Table 4. Among the selected indicators, only IPD exhibited this behavior (Figure 2b).

In the second case, when the lower bound exceeds $(\bar{x} - 2 \cdot \sigma)$, the interval between inf and $(\bar{x} + 2 \cdot \sigma)$ is divided into four equal segments, and the resulting values define the points of maximum membership for the corresponding fuzzy sets. Under this configuration, only the Very Poor membership function assumes a trapezoidal shape (Table 5). This procedure was applied to the IPF, IPL, ICE, and IDE indicators (Figures 2a, 2c, 2d, and 2e). Figure 2f presents the output membership functions of both the intermediate and the final fuzzy models.

Fuzzy inference system

To develop the fuzzy inference system, the performance indicators were divided into two subgroups according to the type of efficiency being evaluated. The first subgroup comprises the water loss indicators IPF, IPD, and IPL, which were used to assess hydraulic efficiency, whereas the second subgroup includes the energy consumption and expenditure indicators ICE and IDE, which were used to evaluate energy efficiency.

The input values were processed independently by each subgroup in the first inference stage, producing two intermediate outputs: one representing hydraulic efficiency and the other representing energy efficiency. These intermediate outputs were subsequently used as inputs to a second inference stage responsible for computing the IASA. The rule base consists of 125 (5^3) rules for the hydraulic subsystem and 25 (5^2) rules for the energy subsystem.

In the second inference stage, responsible for integrating the intermediate outputs, an additional 25 rules (5^2) were implemented. This hierarchical architecture was adopted to reduce computational complexity, since a single fuzzy model including all five input variables would require 3,125 rules (5^5), making the rule-definition process considerably more complex.

The inference process is based on linguistic if–then rules that associate antecedent conditions with their corresponding consequents.

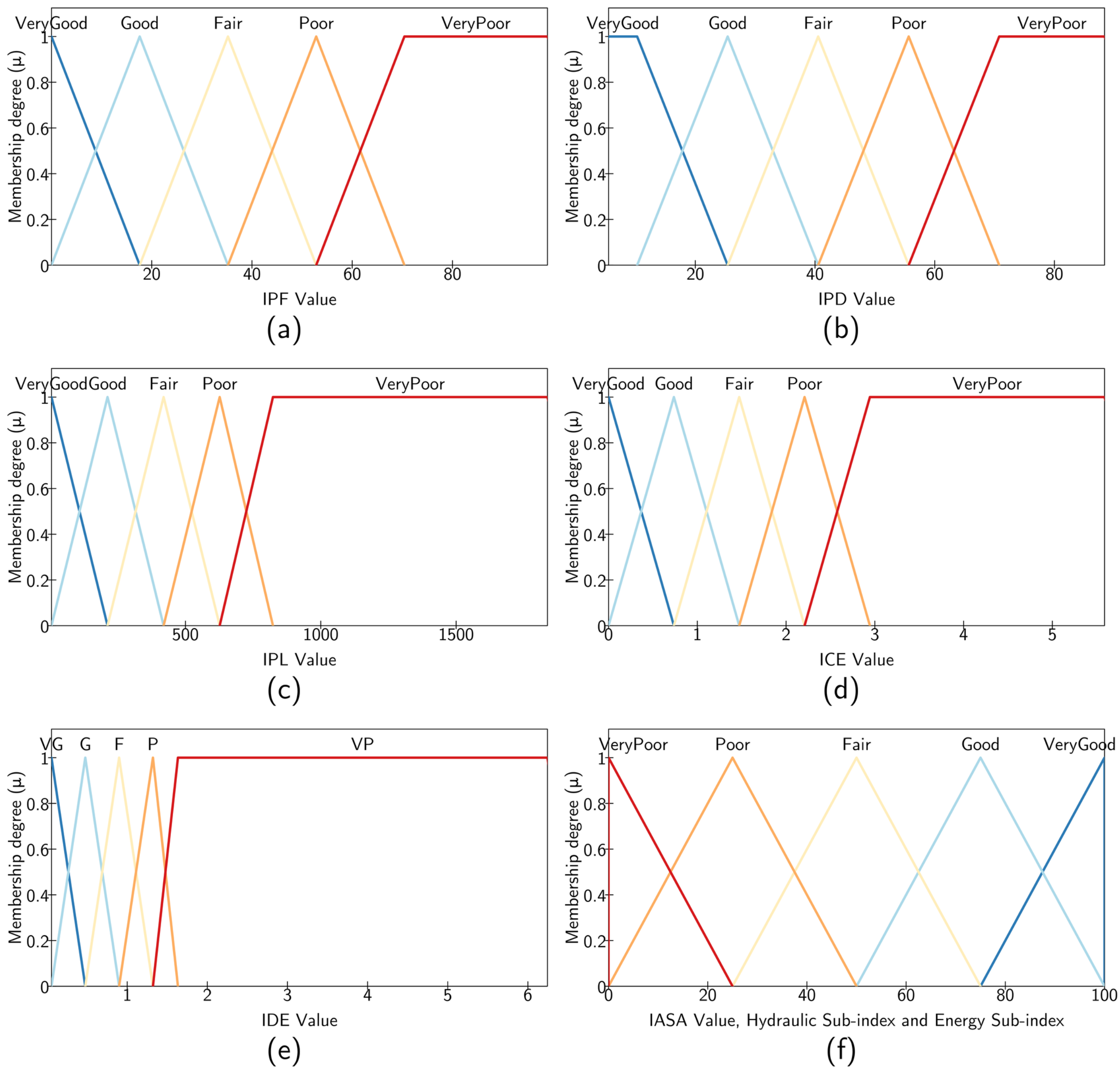

Figure 2 – Membership functions and their corresponding boundaries for (a) IPF, (b) IPD, (c) IPL, (d) ICE, (e) IDE, and (f) IASA

Table 4 – Membership function intervals for Case 1

Membership Function	Shape	Intervals
Very Good	Trapezoidal	$[\inf - \sigma, \inf, \bar{x} - 2\cdot\sigma, \bar{x} - \sigma]$
Good	Triangular	$[\bar{x} - 2\cdot\sigma, \bar{x} - \sigma, \bar{x}]$
Fair	Triangular	$[\bar{x} - \sigma, \bar{x}, \bar{x} + \sigma]$
Poor	Triangular	$[\bar{x}, \bar{x} + \sigma, \bar{x} + 2\cdot\sigma]$
Very Poor	Trapezoidal	$[\bar{x} + \sigma, \bar{x} + 2\cdot\sigma, \sup, \sup + \sigma]$

Table 5 – Membership function intervals for Case 2

Membership Function	Shape	Intervals
Very Good	Triangular	[inf – Q, inf, inf + Q]
Good	Triangular	[inf, inf + Q, inf + 2·Q]
Fair	Triangular	[inf + Q, inf + 2·Q, inf + 3·Q]
Poor	Triangular	[inf + 2·Q, inf + 3·Q, $\bar{x}$ + 2·σ]
Very Poor	Trapezoidal	[inf + 3·Q, $\bar{x}$ + 2·σ, sup, sup + Q]

These rules employ logical operators to combine input conditions and determine the resulting output. An example of a rule is:

If (ICE is Very Good) and (IDE is Very Good), then (Energy Efficiency is Very Good).

All remaining rules follow the same logical structure. Figure 3 presents the rule surfaces, which describe the overall behavior of the fuzzy inference system. In these surfaces, the output variable on the Z-axis (efficiency index) responds continuously and nonlinearly to all possible combinations of the input variables represented on the X- and Y-axes. Figure 3a illustrates the hydraulic efficiency surface obtained by fixing IPF while varying IPD and IPL. Figure 3b shows the energy efficiency surface as a function of ICE and IDE, whereas Figure 3c presents the final IASA output generated in the second inference stage.

The Mamdani inference method was adopted, in which the membership degrees of the input variables are combined using the minimum (min) operator, while the outputs of the individual rules are aggregated using the maximum (max) operator (Mamdani, 1976). Figure 4 illustrates the flowchart of the proposed fuzzy inference system.

Following the inference process, defuzzification is performed to convert the fuzzy output into a crisp value that can be readily interpreted by users. For this purpose, the centroid (center-of-area) method was employed, as it is one of the most widely used defuzzification techniques: it considers the entire support of the fuzzy set and computes the weighted centroid of the area under the membership function. This method produces a single representative value that can be interpreted as the expected value of the output variable. By accounting for the complete shape of the aggregated fuzzy set, the centroid method provides a balanced and consistent solution, making it particularly suitable for applications requiring robustness and representativeness.

As a complementary step for validating the proposed fuzzy model, a correlation analysis was performed involving the input variables, the intermediate sub-indices, and the final efficiency index (IASA). Spearman's rank correlation coefficient was selected because of its suitability for empirical datasets commonly encountered in the sanitation sector, which frequently exhibit nonlinear relationships, asymmetric distributions, and the presence of outliers. This coefficient provides a robust measure of the direction and strength of monotonic relationships between variables, even when the underlying data do not follow linear patterns. Furthermore, its lower sensitivity to extreme values enhances the robustness and reliability of the statistical interpretation of the results (Bishara and Hittner, 2012; Medupe and Letshwenyo, 2024).

Results and discussion

The data obtained from the SNIS/SINISA databases proved to be consistent for 1,491 of the 1,541 municipalities analyzed, representing more than 80% of the municipalities in Northeastern Brazil. These

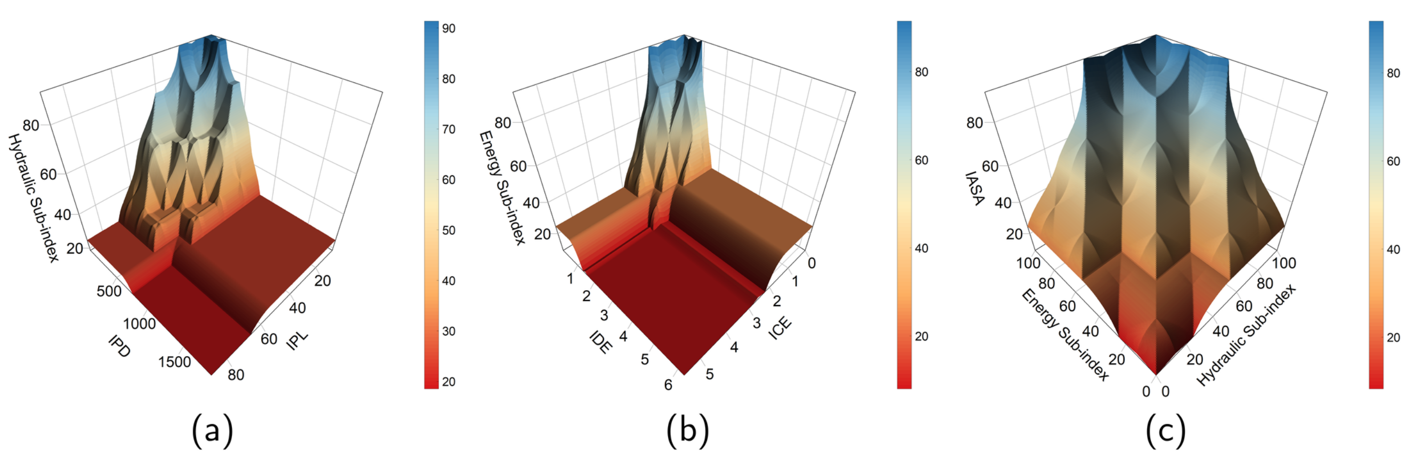

Figure 3 – Rule surfaces of the fuzzy inference system

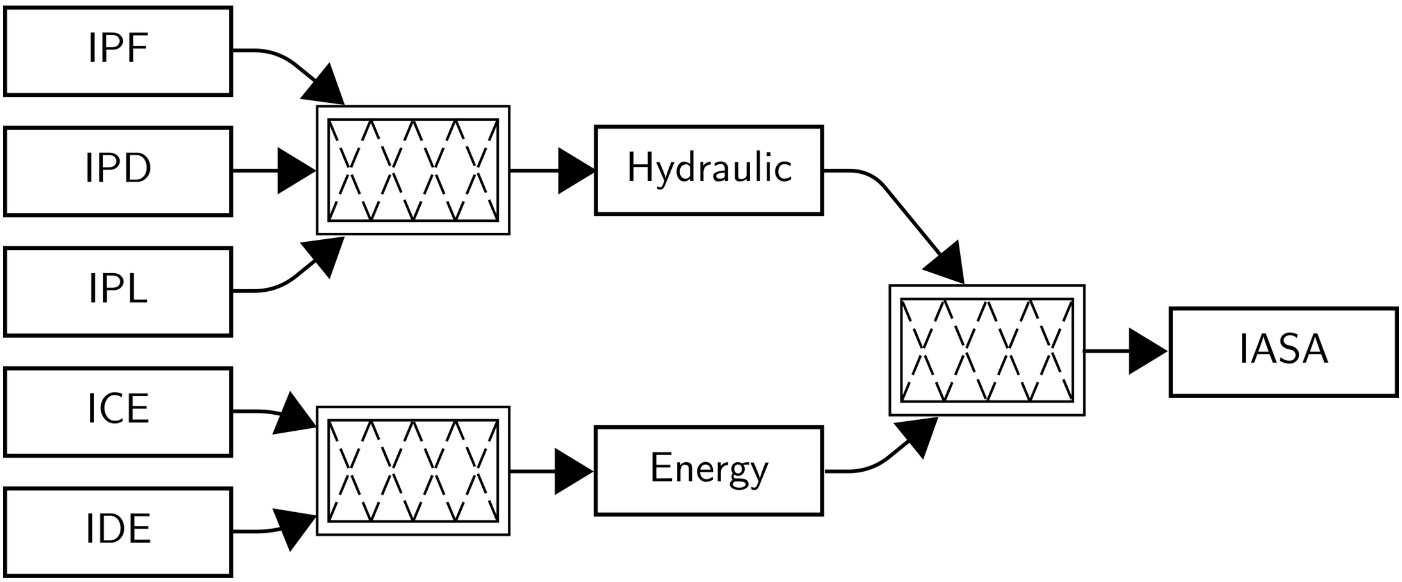

Figure 4 – Flowchart of the proposed fuzzy inference system

municipalities are distributed across all nine states of the region, ensuring broad spatial representativeness of the study sample. The IASA was applied to this set of municipalities, and the resulting values are presented in Figure 5 as a thematic map, facilitating the spatial visualization of system performance.

The application of the proposed index revealed marked regional disparities, with the state of Bahia achieving the highest average performance (61.70) and Maranhão exhibiting the lowest score (38.58). The superior performance of Bahia is consistent with the findings of Bezerra et al. (2019), who identified the state's water utility as having the best operational performance among the utilities evaluated in Northeastern Brazil. Likewise, the spatial analysis conducted by Araújo et al. (2016) had already indicated that Maranhão and other northeastern states exhibited persistently high levels of distribution water losses, reflecting long-standing structural and operational deficiencies throughout the region. The poor performance observed for Maranhão in the proposed model reflects institutional vulnerabilities that directly constrain the investment capacity of water utilities, thereby limiting the universal provision of safe drinking water, particularly in socially vulnerable areas (Cetrulo et al., 2019). These findings suggest that operational inefficiency undermines the long-term sustainability of water utilities and highlights the need for targeted action plans focused on strengthening corporate governance and improving operational management (Trata Brasil and GO Associados, 2025).

Table 6 presents the average values of the performance indicators and the IASA for each state, whereas Table 7 lists the ten municipalities with the highest and lowest scores according to the proposed index.

Effect of population size on the performance indicators

To evaluate the influence of population size on system efficiency, the municipalities were grouped into five population classes according to the classification proposed by Bezerra et al. (2019). Figure 6 presents the boxplots of the performance indicators for the respective population groups, whereas Table 8 summarizes the mean values for each class. Overall, larger municipalities exhibited lower data dispersion, indicating more homogeneous performance across the evaluated indicators. The hydraulic indicators showed greater symmetry among the population classes, whereas the energy indicators displayed positive skewness in the smaller municipalities, suggesting greater operational heterogeneity in these systems.

Distinct trends were observed between the two groups of indicators. Larger municipalities generally exhibited greater energy efficiency, while remaining vulnerable to substantial physical water losses. This behavior can be explained by the fact that systems serving a large number of service connections are able to optimize the operation of pumping stations and distribute operational costs over a larger customer base (Tourinho et al., 2022; Molinos-Senante et al., 2025). However, the complexity of operating extensive distribution networks and maintaining adequate pressure control makes it more difficult to reduce real water losses, demonstrating that system scale alone does not guarantee lower water losses (Marques et al., 2021).

Regarding the final index (IASA), both the mean and median values decreased with increasing population size, suggesting that larger systems do not necessarily achieve higher overall efficiency. This trend is largely driven by the higher levels of physical water losses observed in these municipalities.

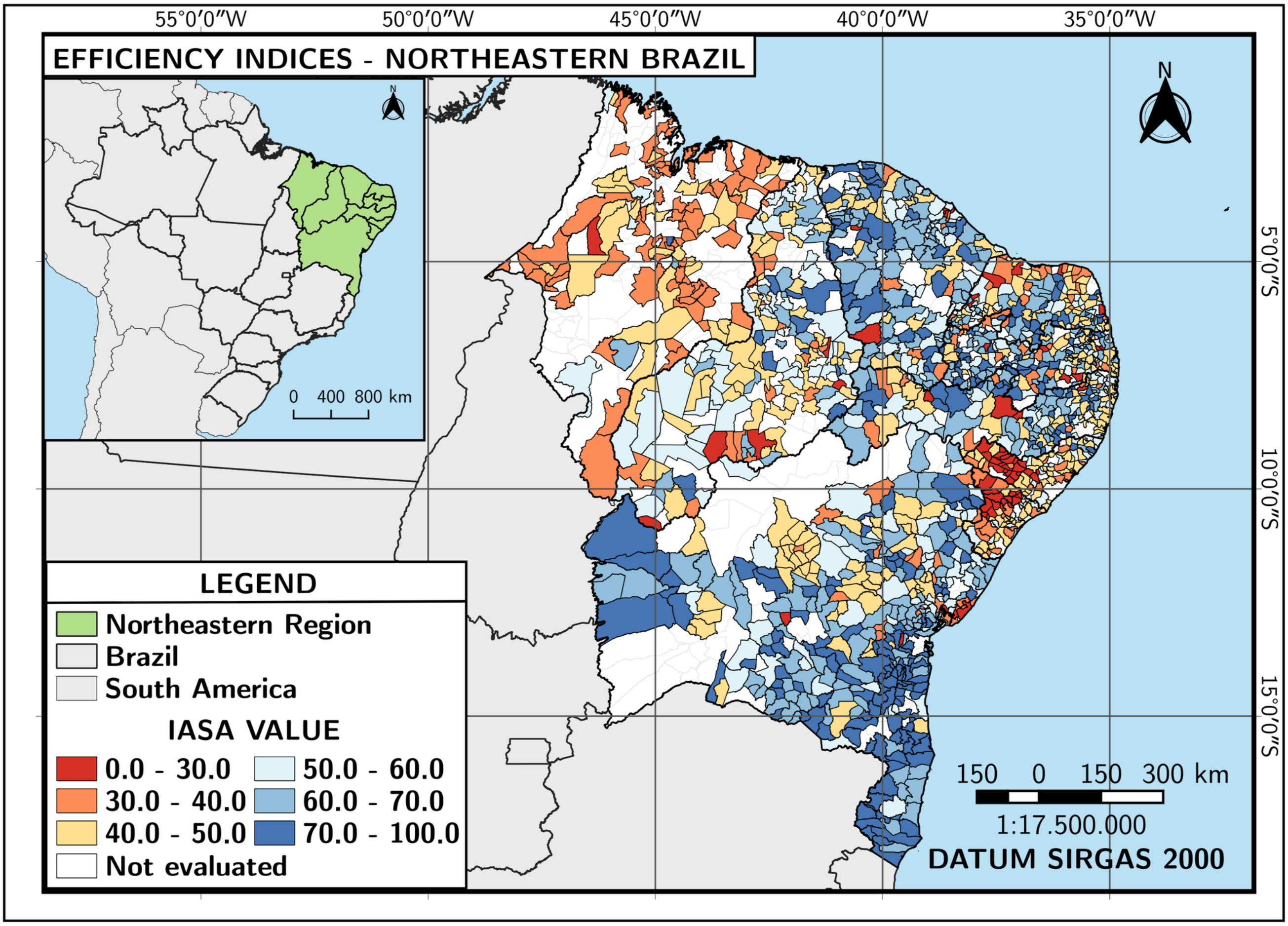

Figure 5 – Spatial distribution of the IASA across municipalities in Northeastern Brazil

Table 6 – Mean values of the performance indicators and the IASA by state

State	IPF (%)	IPD (%)	IPL (L/con/day)	ICE (kWh/m³)	IDE (R$/kWh)	IASA
Alagoas	45.68	52.65	416.44	1.39	0.76	42.24
Bahia	23.05	28.22	123.47	1.21	0.66	61.70
Ceará	19.68	35.90	184.70	0.92	0.94	59.67
Maranhão	66.88	65.02	924.31	0.60	0.65	38.58
Paraíba	18.85	37.75	200.78	1.03	0.87	57.86
Pernambuco	26.84	38.95	233.90	1.38	0.51	57.29
Piauí	36.78	47.84	354.61	0.72	0.87	50.89
Rio Grande do Norte	24.32	41.54	291.06	1.34	0.74	53.34
Sergipe	42.90	53.54	380.07	1.43	0.56	41.02

Validation of the fuzzy model through correlation analysis

The correlation matrix (Figure 7) further demonstrates the coherence and internal consistency of the proposed fuzzy model. As expected, the IASA exhibited a strong negative correlation with all water loss indicators, while also showing strong internal consistency among the variables composing the hydraulic sub-index. These findings confirm that increasing water losses are directly associated with

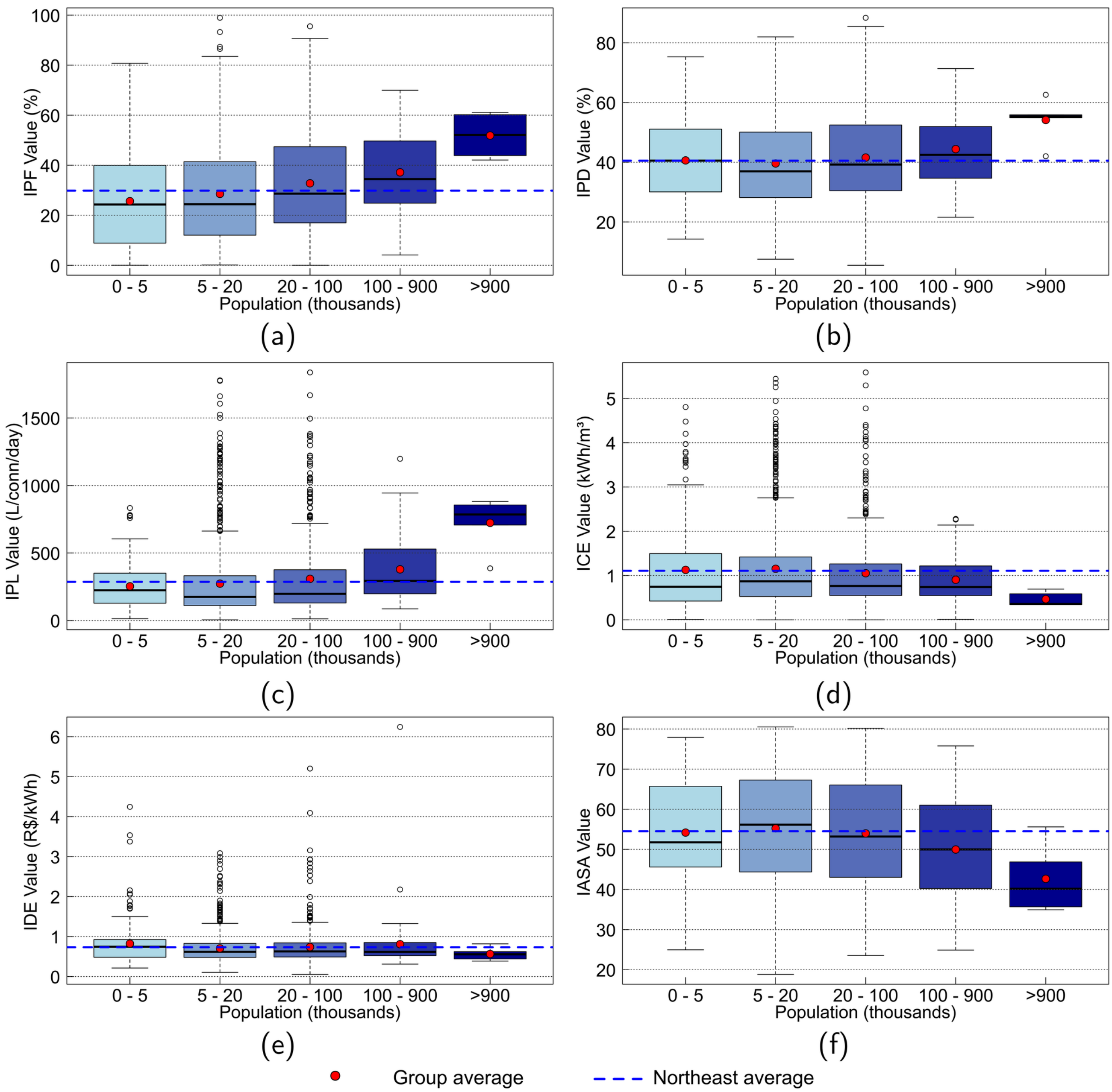

Figure 6 – Comparison of the performance indicators across population classes: (a) IPF, (b) IPD, (c) IPL, (d) ICE, (e) IDE, and (f) IASA

Table 7 – Municipalities with the best and worst performance according to the IASA

Highest Scores				Lowest Scores			
Rank	Municipality	State	IASA	Rank	Municipality	State	IASA
1	Marcionílio Souza	BA	80.52	1491	Bacabeira	MA	18.85
2	Santa Cruz	RN	80.19	1490	Serra do Mel	RN	19.08
3	Marizópolis	PB	79.91	1489	Jacaré dos Homens	AL	19.18
4	Jucuruçu	BA	79.72	1488	Buriticupu	MA	23.53
5	Marco	CE	79.65	1487	Belo Monte	AL	23.82

Highest Scores				Lowest Scores			
Rank	Municipality	State	IASA	Rank	Municipality	State	IASA
6	Frei Miguelinho	PE	79.44	1486	Severiano Melo	RN	24.89
7	Caiçara	PB	79.27	1485	Simões Filho	BA	24.89
8	Ibirapuã	BA	79.06	1484	Batalha	AL	24.90
9	Passira	PE	78.90	1483	Monte Alegre de Sergipe	SE	24.90
10	Nova Ibiá	BA	78.63	1482	Major Isidoro	AL	24.91

Table 8 – Mean values of the performance indicators and the IASA by population class

Population Class (inhabitants)	Number of Municipalities	IPF (%)	IPD (%)	IPL (L/con/day)	ICE (kWh/m³)	IDE (R$/kWh)	IASA
< 5 thousand	180	24.62	40.62	253.33	1.13	0.82	54.18
5-20 thousand	801	27.64	39.59	273.13	1.15	0.71	55.25
20-100 thousand	454	32.48	41.58	308.50	1.05	0.74	53.98
100-900 thousand	51	37.15	44.39	378.81	0.90	0.81	49.96
> 900 thousand	5	51.88	54.16	723.15	0.47	0.56	42.65

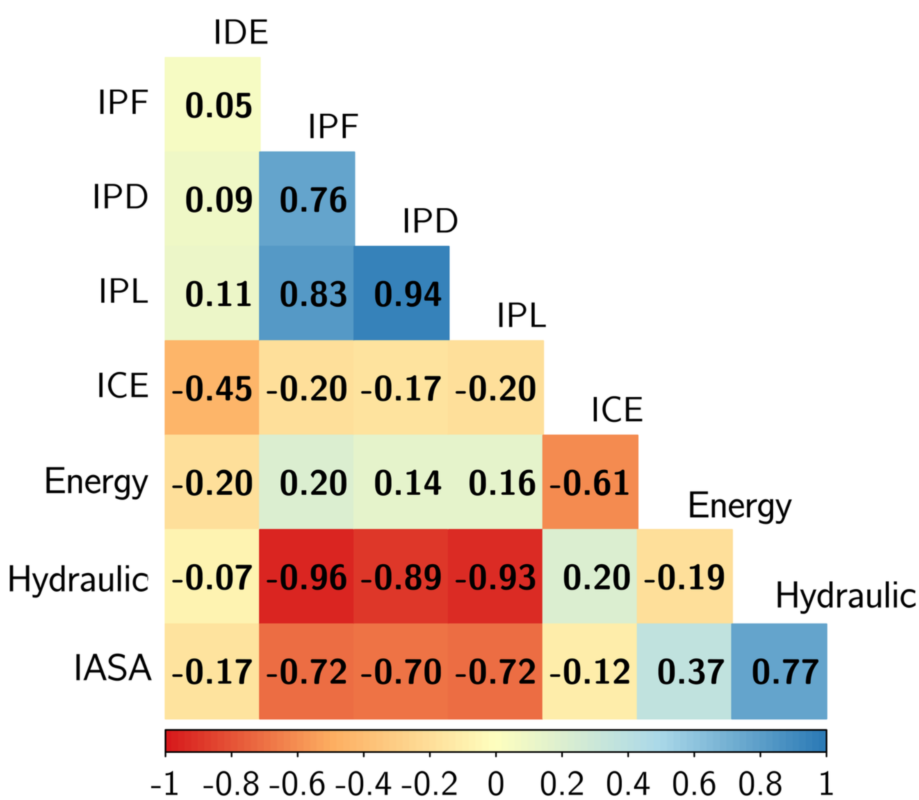

Figure 7 – Correlation matrix of the performance indicators, intermediate sub-indices, and the IASA

lower overall system efficiency, consistent with the behavior expected from the proposed model.

Furthermore, the IASA showed a strong positive correlation with the hydraulic sub-index (r = 0.77) and a moderate positive correlation with the energy sub-index (r = 0.37). This difference suggests that, for the municipalities analyzed and under the adopted inference rules, hydraulic efficiency variables exert a greater influence on the final index than energy efficiency variables. This finding highlights that preserving the physical integrity of distribution networks constitutes the foundation for the long-term sustainability of WSS. Reducing non-revenue water provides a dual benefit by preventing water losses while simultaneously decreasing the energy required to pump water that ultimately does not reach consumers (Ferreira et al., 2019; Liemberger and Wyatt, 2019). The integrated assessment provided by the proposed index also indicates to regulators that strategies focused exclusively on improving equipment efficiency are likely to have limited effectiveness unless accompanied by measures aimed at reducing real water losses (Marques et al., 2021; Sousa et al., 2025).

Conclusions

This study successfully achieved its primary objective by developing and validating the IASA for WSS in Northeastern Brazil. The application of the proposed model revealed clear differences in performance among the states, with Bahia presenting the highest mean IASA value (61.70) and Maranhão the lowest (38.58). The results further demonstrate that system efficiency depends less on water availability than on the management capacity of water utilities. The analysis also showed that larger municipalities generally exhibit lower electricity consumption per unit of water produced; however, these same systems experience higher levels of water losses due to the complexity and aging of their distribution networks. Correlation analysis confirmed both the coherence and sensitivity of the proposed model, validating the expected relationships among the variables and demonstrating the logical and operational consistency of the index.

Beyond its practical findings, this study contributes to advancing methodologies for evaluating water utility performance. The proposed framework integrates fuzzy logic with anomaly detection algorithms, enabling the treatment of uncertainty inherent in self-reported public databases. Unlike traditional deterministic approaches, which generally assume that utility-reported information is accurate, the proposed methodology explicitly recognizes the possibility of errors, inconsistencies, and missing information, incorporating these aspects into the assessment of management transparency. Consequently, the study demonstrates that missing or inconsistent information should not be interpreted merely as statistical noise but rather as an important indicator of operational inefficiency and structural weaknesses in the governance of services.

The proposed tool represents a valuable decision-support instrument for the regulation of water supply services based on publicly available databases. The methodology provides managers and regulators with a performance metric capable of supporting investment prioritization and guiding universal service policies in a manner that more accurately reflects operational reality. An inherent limitation of this type of model is the need to periodically recalibrate the membership function boundaries as the underlying database evolves over time. Finally, the proposed framework provides opportunities for future research applying the same uncertainty-handling approach to other efficiency assessment contexts, either by adapting the methodology to different geographical regions or by extending it to other dimensions of urban infrastructure.

Authors' Contributions: Silva, J.T.G.: Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft; Bezerra, S.T.M.: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Conflicts of interest: The authors declare no conflicts of interest.

Funding: This work was supported by the National Council for Scientific and Technological Development (CNPq) [Conselho Nacional de Desenvolvimento Científico e Tecnológico], grant numbers 420012/2021-4 and 308202/2022-8.

Declaration on the use of generative artificial intelligence: During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to review and improve the overall writing quality. After using this tool, the authors reviewed and edited the content as necessary and take full responsibility for the content of the published article.

References

Abidi, J.H.; Elzain, H.E.; Sabarathinam, C.; Selmane, T.; Selvam, S. et al., 2024. Evaluation of groundwater quality indices using multi-criteria decision-making techniques and a Fuzzy logic model in an irrigated area. Groundwater for Sustainable Development, v. 25, 101122. https://doi.org/10.1016/j.gsd.2024.101122

Alsharif, K.; Feroz, E.H.; Klemer, A.; Raab, R., 2008. Governance of water supply systems in the Palestinian Territories: A data envelopment analysis approach to the management of water resources. Journal of Environmental Management, v. 87 (1), 80-94. https://doi.org/10.1016/j.jenvman.2007.01.008.

Araújo, S.C.; Silva Filho, J.A.; Silva, G.M.S.; Cabral Filha, M.C.S.; Nogueira, V.F.B., 2016. Distribuição espacial de indicadores operacionais de serviço de abastecimento de água no Nordeste Brasileiro. Revista Verde de Agroecologia e Desenvolvimento Sustentável, v. 11 (1), 20-28. https://doi.org/10.18378/rvads.v11i1.4470.

Banker, R.D.; Charnes, A.; Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, v. 30 (9), 1078-1092. https://doi.org/10.1287/mnsc.30.9.1078.

Bezerra, S.T.M.; Pertel, M.; Macêdo, J.E.S., 2019. Avaliação de desempenho dos sistemas de abastecimento de água do Agreste brasileiro. Ambiente Construído, v. 19 (3), 249-258. https://doi.org/10.1590/s1678-86212019000300336.

Bishara, A.J.; Hittner, J.B., 2012. Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. Psychological Methods, v. 17 (3), 399-417. https://doi.org/10.1037/a0028087.

Boukerche, A.; Zheng, L.; Alfandi, O., 2020. Outlier detection: Methods, models, and classification. ACM Computing Surveys, v. 53 (3), 1-37. https://doi.org/10.1145/3381028

Brasil, 2020. Presidência da República. Lei nº 14.026, de 15 de julho de 2020.

Diário Oficial da União, Brasília.

Brasil, 2025. Ministério das Cidades. Resultados SINISA 2025. https://www.gov.br/cidades/pt-br/acesso-a-informacao/acoes-e-programas/saneamento/sinisa/resultados-sinisa/resultados-sinisa-2025.

Cetrulo, T.B.; Marques, R.C.; Malheiros, T.F., 2019. An analytical review of the efficiency of water and sanitation utilities in developing countries. Water Research, v. 161, 372-380. https://doi.org/10.1016/j.watres.2019.05.044.

Deb, K., 2000. An efficient constraint handling method for genetic algorithms. Computer Methods in Applied Mechanics and Engineering, v. 186 (2-4), 311-338. https://doi.org/10.1016/S0045-7825(99)00389-8.

Empresa de Pesquisa Energética (EPE), 2025. Anuário estatístico de energia elétrica 2025: Ano base 2024. EPE, Rio de Janeiro.

Fernández, L.P.S., 2021. Environmental noise indicators and acoustic indexes based on Fuzzy modelling for urban spaces. Ecological Indicators, v. 126, 107631. https://doi.org/10.1016/j.ecolind.2021.107631.

Ferreira, R.C.; Silveira, A.G.; Cabral, C.; Gama, J.R.; Silva, B., 2019. Caderno Temático 1: Perdas de água e eficiência energética. Ministério das Cidades, Brasília, 35 p.

Instituto Brasileiro de Geografia e Estatística (IBGE), 2025. Diretoria de Pesquisas (DPE). Coordenação de População e Indicadores Sociais (COPIS). Estimativas da população para 2025. IBGE, Brasília.

Lazzarotto, D.R., 2018. Avaliação da Sustentabilidade da Floresta Nacional de Irati por Meio de Lógica Fuzzy. Floresta e Ambiente, v. 25 (1), e00014712. https://doi.org/10.1590/2179-8087.014712.

Li, P.; Gu, H.; Yin, L.; Li, B., 2022. Research on trend prediction of component stock in Fuzzy time series based on deep forest. CAAI Transactions on Intelligence Technology, v. 7 (4), 617-626. https://doi.org/10.1049/cit2.12139.

Liemberger, R.; Wyatt, A., 2019. Quantifying the global non-revenue water problem. Water Supply, v. 19 (3), 831-837. https://doi.org/10.2166/ws.2018.129.

Liu, F.T.; Ting, K.M.; Zhou, Z.-H., 2008. Isolation Forest. In: 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, pp. 413-422. https://doi.org/10.1109/ICDM.2008.17.

Liu, F.T.; Ting, K.M.; Zhou, Z.-H., 2012. Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data (TKDD), v. 6 (1), 1-39. https://doi.org/10.1145/2133360.2133363

Mamdani, E.H., 1976. Advances in the linguistic synthesis of Fuzzy controllers. International Journal of Man-Machine Studies, v. 8 (6), 669-678. https://doi.org/10.1016/S0020-7373(76)80028-4.

Marques, L.O.A.; Carvalho, R.S.; Sa, M.O.M.; Malheiros, T.F., 2021. Benchmarking enquanto ferramenta de diminuição das perdas físicas em sistemas de abastecimento de água. Ambiente & Sociedade, v. 24, 1-18. https://doi.org/10.1590/1809-4422asoc20200025vu2021L4DE.

Maziotis, A.; Villegas, A.; Molinos-Senante, M.; Sala-Garrido, R., 2020. Impact of external costs of unplanned supply interruptions on water company efficiency: Evidence from Chile. Utilities Policy, v. 66, 101087. https://doi.org/10.1016/j.jup.2020.101087.

Medupe, M.; Letshwenyo, M.W., 2024. Investigation of self-purification and water quality index during dry and rainy seasons in the Khurumela Stream (Botswana). Journal of Ecohydraulics, v. 10 (1), 1-18. https://doi.org/10.1080/24705357.2024.2363755.

Molinos-Senante, M.; Maziotis, A.; Sala-Garrido, R.; Mocholi-Arce, M., 2025. Benchmarking energy efficiency in water utilities: Evidence from England and Wales. Computers & Industrial Engineering, v. 209, 111457. https://doi.org/10.1016/j.cie.2025.111457.

Montgomery, D.C.; Runger, G.C., 2021. Estatística aplicada e probabilidade para engenheiros. 7. ed. LTC, Rio de Janeiro, 416 p.

Patel, A.; Chitnis, K., 2022. Application of Fuzzy logic in river water quality modelling for analysis of industrialization and climate change impact on Sabarmati river. Water Supply, v. 22 (1), 238-250. https://doi.org/10.2166/ws.2021.275.

Sarkheil, H.; Noughabi, K.S.; Azimi, Y.; Rahbari, S., 2023. Fuzzy soil quality index using resistivity and induced polarization for contamination assessment in a lead and zinc drainage irrigation field study. Ecological Indicators, v. 152, 110362. https://doi.org/10.1016/j.ecolind.2023.110362.

Sousa, C.O.M.; Souza, R.F.; Fouto, N.M.M.D., 2025. Key drivers of non-revenue water in developing countries: insights from a multilevel study in Brazil. Cleaner Water, v. 3, 1-11. https://doi.org/10.1016/j.clwat.2025.100078.

Tourinho, M.; Santos, P.R.; Pinto, F.T.; Camanho, A.S., 2022. Performance assessment of water services in Brazilian municipalities: An integrated view of efficiency and access. Socio-Economic Planning Sciences, v. 79, 101139. https://doi.org/10.1016/j.seps.2021.101139.

Trata Brasil; GO Associados, 2025. Estudo de perdas de água 2025 (SINISA, 2023): desafios na eficiência do saneamento básico no Brasil. Instituto Trata Brasil, São Paulo.

Villegas, A.; Molinos-Senante, M.; Maziotis, A., 2019. Impact of environmental variables on the efficiency of water companies in England and Wales: a double-bootstrap approach. Environmental Science and Pollution Research, v. 26, 35557-35573. https://doi.org/10.1007/s11356-019-06238-z.

Yazdi, S.H.; Robati, M.; Samani, S.; Hargalani, F.Z., 2024. Assessing the sustainability of groundwater quality for irrigation purposes using a Fuzzy logic approach. Environmental and Sustainability Indicators, v. 22, 100342. https://doi.org/10.1016/j.indic.2024.100342.

Zadeh, L.A., 1965. Fuzzy sets. Information and Control, v. 8 (3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X.

Zhang, Y.; Zhang, J.; Dong, L., 2023. Fuzzy logic regional landslide susceptibility multi-field information map representation analysis method constrained by spatial characteristics of mining factors in mining areas. Processes, v. 11 (4), 985. https://doi.org/10.3390/pr11040985.

Zihad, S.M.R.A.; Islam, A.R.M.T.; Siddique, M.A.B.; Mia, M.Y.; Islam, M.S. et al., 2023. Fuzzy logic, geostatistics, and multiple linear models to evaluate irrigation metrics and their influencing factors in a drought-prone agricultural region. Environmental Research, v. 234, 116509. https://doi.org/10.1016/j.envres.2023.116509.