

Development of a Water Infrastructure Sustainability Index (WISI): An Integrated Framework for Assessing Resilience, Circular Economy, and Long-Term Sustainability

Ayman Ramadan

ayman_as55@yahoo.com

International Business Academy of London <https://orcid.org/0009-0002-0317-7867>

Research Article

Keywords: composite indicator, wastewater treatment, sustainability assessment, circular economy, resilience, sensitivity analysis, Monte Carlo

Posted Date: October 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11530437/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Wastewater treatment plants are often assessed by compliance and cost, which leaves energy, carbon, resource recovery, asset condition, climate exposure and digital capability outside a single view. We reformulate the conceptual Water Infrastructure Sustainability Index (WISI) as a computable index for wastewater treatment plants. WISI v1.1 combines seven dimensions (technical performance, water, energy, circular economy, carbon, asset resilience, digital intelligence) through 22 normalised indicators and weighted arithmetic aggregation. The reformulation corrects inverted indicator directions, removes undefined and non-applicable indicators, and limits double counting by assigning each physical flow to one dimension. Because no complete operational dataset was available, we tested the index on 40 synthetic plants, with 50 replicate datasets, using Monte Carlo weight perturbation, varying weight distributions, and alternative model structures. Rankings were stable under moderate weight uncertainty (median Spearman $\rho = 0.98$ for Dirichlet $\kappa = 50$) and fell to 0.87 only when weights were nearly random ($\kappa = 3$); aggregation rule changed rankings most ($\rho = 0.95$ for geometric aggregation). Nominal weights did not match effective influence: technical performance, with the largest weight (0.20), accounted for about 7–8% of index variance. Class bands from the earlier framework placed a median of 70% of synthetic plants in the lowest class; data-derived bands were uncertain by about five score points. A feasibility check on two operating plants showed that routine laboratory data cover only 3 of the 22 indicators.

1. Introduction

1.1 Background

Wastewater treatment plants are often judged on compliance with effluent limits, availability and cost. These measures say little about electricity use, greenhouse gas emissions, recovery of water, energy and nutrients, ageing of assets, exposure to climate hazards or readiness for digital operation. Each of these affects whether a plant can keep serving its community under tighter standards and a changing climate (IPCC, 2023). and several of them pull in different directions: more intensive aeration may improve effluent quality while raising energy use and emissions (Drewnowski, et al., 2019). A tool that shows these trade-offs together, in a form that an operator or owner can compute from routine data, is therefore useful for screening and for setting priorities.

1.2 Existing approaches

Performance-indicator systems and standards. The IWA manual on performance indicators for wastewater services gives indicators for management and for benchmarking of wastewater utilities (Matos, et al., 2003). ISO 24511:2024 provides guidelines for managing wastewater utilities and for assessing wastewater services, with service assessment criteria and examples of performance indicators, and explicitly sets no target values or thresholds (ISO, 2024). Both provide a common language and a set of measures; neither is designed to deliver one score that weighs indicators against each other.

Composite indicators for wastewater systems. Molinos-Senante et al. (2014) built a composite indicator of economic, environmental and social issues to compare seven secondary treatment technologies for small communities. Sun et al. (2020) combined annual cost, carbon emission intensity, eutrophication and resilience, with weights from the analytic hierarchy process, to compare four wastewater management alternatives. Popović et al. (2013) examined how well sustainability indicators apply to wastewater treatment and found that environmental and economic indicators are easy to quantify whereas social indicators need a qualitative approach. These studies compare technologies or alternatives at the planning stage; they do not target the periodic assessment of an operating plant.

Frameworks for sustainability and circularity. A recent methodological roadmap for wastewater treatment plants assembled 167 sustainability indicators and 32 circularity indicators and tested them on two large plants; about 61% of its quantitative requirements could be calculated from routine plant information, but thresholds could not be defined for many indicators and guidance on data collection was lacking (Estévez, et al., 2025). Faragò et al. (2022) described a framework with 22 criteria for water resource recovery facilities. Such catalogues are comprehensive but large, and they leave open how to reduce them to a manageable, comparable score.

Infrastructure rating systems. Envision, developed by the Institute for Sustainable Infrastructure with the Zofnass Program at Harvard University, is a point-based rating system for infrastructure projects with five categories (Quality of Life, Leadership, Resource Allocation, Natural World, Climate and Resilience) and 64 credits, and it offers third-party verification (ISI, 2018). It addresses projects through planning, design and construction and not the day-to-day performance of an operating plant.

1.3 Gap and contributions

In the sources we reviewed, we found no compact, plant-level index that combines technical performance with energy, carbon, circular economy, asset resilience and digital capability and that is accompanied by an explicit test of how far its ranking depends on weights and model structure. The review covered utility indicator systems and standards, composite indicators for wastewater treatment, sustainability and circularity frameworks, and infrastructure rating systems. It was not a systematic database search, so the gap is stated relative to the sources consulted.

The conceptual WISI framework was first proposed in a book (Ramadan, 2026). That version defined the dimensions and an initial set of weights but did not specify all indicators computably or test the index. The present paper reformulates the framework for wastewater treatment plants and examines its structural robustness.

The paper makes four contributions:

1. It reformulates the conceptual WISI framework as a computable index for wastewater treatment plants, with explicit definitions, directions, normalisation and reference bounds, and with the main sources of double counting and inverted indicator direction corrected.
2. It tests the structural robustness of the index by Monte Carlo analysis of weights, emulation of expert-judgement inconsistency, alternative model structures and replicated synthetic datasets, and shows that nominal weights differ from effective influence.
3. It shows that fixed class bands are not transferable between normalisations, and proposes data-derived class edges with a bootstrap estimate of their uncertainty.
4. It checks the feasibility of applying the index to routine laboratory data from two operating plants and identifies the records needed for a full assessment.

Table 1
Positioning of WISI v1.1 relative to the approaches reviewed.

Approach	Level	Output	Thresholds / weights	Source
IWA performance indicators	Utility	Indicator set for management and benchmarking	Weighting not assessed here	Matos et al. (2003)
ISO 24511:2024	Utility	Guidelines and example indicators	No target values or thresholds set	ISO (2024)
Composite indicator of small treatment systems	Technology comparison	Composite of economic, environmental, social issues	AHP weights assigned from expert preferences	Molinos-Senante et al. (2014)
AHP composite for management alternatives	Alternative comparison	Cost, carbon intensity, eutrophication, resilience	AHP weights	Sun et al. (2020)
Sustainability and circularity roadmap	Plant	167 + 32 indicators	Thresholds undefined for many indicators	Estévez et al. (2025)
Envision	Project	Points, award levels	Levels set as shares of applicable points	ISI (2018)
WISI v1.1 (this study)	Operating plant	One score from 7 dimensions, plus profile	Baseline weights unvalidated; class edges data-derived and uncertain	This study

1.4 Structure of the paper

Section 2 describes the indicator system, the normalisation and the simulation design. Section 3 reports the results. Section 4 discusses implications and limits, and Section 5 concludes.

2. Methodology

2.1 Scope, positioning and design principles

This study reformulates the conceptual WISI framework (Ramadan, 2026) into a computable composite index for municipal wastewater treatment plants (WWTPs), designated WISI v1.1. Desalination, water reuse and distribution facilities, which appeared in the original framework, are excluded because their indicator sets, energy ranges and benchmarking logic differ materially from those of wastewater treatment. The study does not claim empirical validation. Its purpose is to (1) specify a coherent indicator system with explicit definitions, directions and normalisation rules, and (2) test the structural robustness of the resulting index through simulation on synthetic facilities.

Five design principles were applied when reformulating v1.0:

1. **Orthogonality of physical flows.** Each physical flow is assigned to one dimension only: water to WS, energy to ES, emissions to CN, and materials and nutrients to CE. This limits double counting, although it cannot eliminate dependence that arises physically (for example, electricity use drives both ES and Scope 2 emissions; this dependence is quantified in Section 2.8 rather than assumed away).
2. **Common direction and scale.** All indicators are transformed to a 0–100 scale in which higher is better.
3. **Outcome orientation.** Indicators measure performance or risk-adjusted condition, not the mere presence of tools or the importance of assets.
4. **Transparency of assumptions.** Every reference value is tagged as literature-anchored or assumed (Section 2.7).
5. **Restrained predictive claims.** Forward-looking content is limited to remaining useful life and failure risk; no forecasting model is embedded in the index.

2.2 Indicator system

WISI v1.1 aggregates seven dimensions: Technical Performance (TP), Water Sustainability (WS), Energy Sustainability (ES), Circular Economy (CE), Carbon Neutrality (CN), Asset Resilience (AR) and Digital Intelligence (DI). Table 2 lists the indicators, their definitions and their within-dimension weights.

Table 2
WISI v1.1 indicators for WWTPs.

Dim.	Indicator	Definition	Within-dim. weight
TP	TES	Treatment efficiency: $actualremoval \div targetremoval$, capped at 100	0.35
TP	PAS	Plant availability: $operatingtime \div totaltime \times 100$	0.25
TP	CUS	Capacity utilisation score (trapezoidal; 100 within 70–95% utilisation)	0.20
TP	PSS	Process stability: $100 \times (1 - CV/CV_{max})$, floored at 0	0.20
WS	WRR	Reuse ratio: reused water $\div$ treated water, scaled to a target	0.60
WS	RQC	Share of reused water meeting the required quality class	0.40
ES	SEC	Specific electricity consumption (kWh m ⁻³), inverse-linear	0.40
ES	REC	On-site renewable electricity $\div$ total consumption	0.30
ES	ERR	Recovered energy (biogas, heat) $\div$ total consumption	0.30
CE	NRR	Share of influent phosphorus recovered as a product	0.35
CE	BRR	Share of biosolids beneficially reused	0.35
CE	ROI	Resource-recovery revenue (excluding energy and water) $\div$ OPEX	0.30
CN	S1	Scope 1 intensity (kg CO ₂ e m ⁻³), inverse-linear	0.50
CN	S2	Scope 2 intensity (kg CO ₂ e m ⁻³), inverse-linear	0.50
AR	RUL	Value-weighted remaining useful life $\div$ design life	0.35
AR	RISK	Value-weighted $risk = \sum w_i \cdot FP_i \cdot AC_i$	0.35
AR	CVU	Climate resilience score (labelled climate vulnerability in v1.0)	0.30
DI	SCADA, RTM, PA, DT, AI	Maturity level 0–4 per capability, rubric-based	0.20 each

Maturity levels for DI capabilities are defined in Appendix A, and the structure of the index is summarised in Fig. 1.

CV = coefficient of variation; OPEX = operating expenditure; SCADA = supervisory control and data acquisition integration; RTM = real-time monitoring; PA = predictive analytics; DT = digital twin; AI = AI decision support; CVU = climate resilience score (labelled climate vulnerability in v1.0).

Compared with v1.0, the following changes were made. Indicators defined only by name in v1.0 (water security contribution, recovery efficiency, distribution efficiency) were removed because they are either undefined or not applicable to WWTPs. The circular-economy dimension, previously a simple sum of recovery "values", was reduced to three measurable indicators. Asset criticality, which expresses the consequence of failure rather than performance, no longer enters AR as a positive term; it is combined with failure probability into a single risk term. Energy recovered from biogas is counted in ES only.

2.3 Normalisation

All indicators x are mapped to scores s in $[0, 100]$. For indicators with a best value x_b and a worst value x_w , a linear min–max transformation with clipping is applied:

$$s = 100 \times \frac{\max(x - x_w, 0)}{\max(x_b - x_w, 0)}$$

This form applies to both increasing indicators ($x_b > x_w$, e.g. NRR) and decreasing indicators ($x_b < x_w$, e.g. SEC, S1, S2). Four indicators use specific forms:

- TES =

$$s = 100 \times \frac{\max(R - R_{min}, 0)}{\max(R_{max} - R_{min}, 0)}$$

where R is the removal efficiency of the regulated parameter (BOD, COD or TSS)

- PSS = $s = \max\left(0, 100 \times \left(1 - \frac{CV}{CV_{max}}\right)\right)$ with $CV_{max} = 50\%$.

- CUS is trapezoidal: 100 for utilisation u between 70 and 95%, falling linearly to 0 at $u = 30\%$ and $u = 130\%$.

- DI capabilities are scored $25 \times \text{level}$, where $\text{level} \in \{0, 1, 2, 3, 4\}$ follows the maturity rubric in Appendix A (Table A1).

The piecewise energy-score table of v1.0 was replaced by the linear form above because a stepwise table cannot discriminate among facilities with very low or very high energy intensity. Its effect on ranking is tested in Section 2.8.

2.4 Asset resilience and the climate resilience sub-index

The climate resilience component CVU is the climate resilience index of v1.0 with its digital-preparedness term removed, since digital capability is already represented by DI:

$$\text{CVU} = 0.333 \cdot \text{EX} + 0.278 \cdot \text{AD} + 0.222 \cdot \text{RD} + 0.167 \cdot \text{RC}$$

where EX is exposure management, AD adaptation capacity, RD redundancy and RC recovery capability, each scored 0–100 with higher = better. These components follow common concepts of climate adaptation and resilience (OECD, 2013; Hollnagel, 2011; ISO, 2019). EX is defined as the degree to which exposure to flooding, heat, drought and coastal hazards is controlled, not as the exposure itself (IPCC, 2023). The weights are the original CRI weights (0.30, 0.25, 0.20, 0.15) renormalised to sum to one.

The risk term of AR follows asset-management risk concepts (ISO, 2014) and is computed as:

$$\text{AR} = 100 \times (1 - \sum_i w_i \cdot \text{FP}_i \cdot \text{AC}_i \cdot R_{\max})$$

where w_i is the share of asset value in group i ($\sum w_i = 1$), FP_i is the failure probability of group i , AC_i its normalised criticality (Campbell, et al., 2024), and $R_{\max} = 0.60$ is the risk level treated as the worst case (assumption; see Table 3).

2.5 Aggregation

Dimension scores are weighted arithmetic means of their indicators (Table 2). The overall index is:

$$\text{WISI} = 0.20 \cdot \text{ES} + 0.15 \cdot \text{CN} + 0.15 \cdot \text{CVU} + 0.15 \cdot \text{AR}$$

The dimension weights are inherited from v1.0 and are treated as a working hypothesis, not as validated values. Weighted arithmetic aggregation is fully compensatory; a geometric alternative is therefore evaluated as a sensitivity case (Nardo, et al., 2008).

2.6 Classification

v1.0 used five performance bands with fixed edges at 90, 80, 70 and 60. Because those edges had no empirical basis, two data-derived alternatives were compared:

- quintile edges, set at the 20th, 40th, 60th and 80th percentiles of the WISI scores;
- k-means edges, set at the midpoints between the centres of a five-cluster k-means partition of the WISI scores.

Classes are labelled A (highest) to E (lowest); the label “Autonomous” used for the top class in v1.0 is dropped because the class is defined by score, not by operational autonomy. Stability of the k-means edges was assessed by bootstrap resampling of facilities (500 replicates).

2.7 Synthetic facility data and reference values

Because no operational dataset covering all indicators was available, 40 synthetic WWTPs were generated (seed 2026), with 50 additional replicate datasets (seeds 100–149) used to check dependence on a single random draw.

All results derived from these data describe the behaviour of the index structure; they do not describe real facilities.

Each facility has a technology (diffused-air activated sludge, mechanical-aeration activated sludge, membrane bioreactor, trickling filter), a capacity (log-normal, median 60,000 m³ d⁻¹), an age (uniform, 3–45 years), and a grid emission factor (uniform, 0.10–0.70 kg CO₂e kWh⁻¹). A latent management-quality factor, partly dependent on age, shifts all indicators in a plausible direction, so that better-managed plants tend to perform better across dimensions. This is a simplifying assumption: it induces positive inter-dimension correlation, which may be stronger or weaker in real populations. Scope 2 intensity is calculated from the electricity bought from the grid ($\text{SEC} \times (1 - \text{REC} - \text{ERR})$) multiplied by the grid factor, so the dependence between ES and CN is built into the generator, as it is physically.

The generator was adjusted once after a first test run: the design life used for RUL was raised from 30 to 50 years and the worst-case risk $R_{\max}$ from 0.40 to 0.60, because the initial values produced many zero scores. No further tuning was done to shape the score distribution.

Table 3
Reference values for normalisation.

Indicator	Best	Worst	Basis
SEC (kWh m ⁻³)	0.25	1.00	Best: excellent-performance range for diffused-air activated sludge 0.23–0.39 kWh m ⁻³ (P25–P75), mechanical aerators 0.33–0.80 (Silva & Rosa, 2022). Worst bound of 1.00 is an assumption
NRR (fraction of influent P)	0.70	0	Literature: recovery 10–60% from liquid phase, 35–70% from sludge, 70–98% from ash (Chrispim, et al., 2019).
S1 (kg CO ₂ e m ⁻³)	0.02	0.30	Assumption; published values vary widely by method
S2 (kg CO ₂ e m ⁻³)	0.00	0.35	Assumption
REC	0.30	0	Assumption (cap near the > 20% target of v1.0)
ERR	0.60	0	Assumption
ROI	0.10	0	Assumption (v1.0 target of 10% OPEX offset)
WRR	0.70	0	Assumption (v1.0 target)
RISK (FP × AC)	0	0.60	Assumption
CV_max (%)	–	50	Assumption

A common best bound derived from diffused-air plants disadvantages plants with mechanical aeration, whose reference range is higher; technology-specific bounds are a refinement for v2.0.

Assumed bounds require calibration with local data before any practical use. Carbon bounds in particular depend on the electricity grid and the accounting boundary.

2.8 Uncertainty and sensitivity analysis

Five analyses were performed on the synthetic data.

(1) Dimension dependence. Spearman correlations among the seven dimension scores were computed, with particular attention to $ES - CN$. The effective influence of each dimension was expressed as its share of $WISI$ variance, $w_k \cdot Cov(D_k, WISI) / Var(WISI)$, which sums to one and can be compared with the nominal weight.

(2) Weight uncertainty by Monte Carlo. Dimension weights were drawn from Dirichlet distributions centred on the baseline weights, $\alpha = \kappa \cdot w$, with $\kappa = 50$ (moderate), 20 (high) and 3 (near-random uncertainty), 10,000 draws each. For each draw the ranking was compared with the baseline ranking (Spearman's ρ), together with overlap of the five best facilities, 5–95% rank intervals and class retention. Retention is the share of draws in which a facility stays in its baseline class.

(3) Sensitivity to weight distributions. To assess the impact of expert disagreement on the index, latent weights were sampled from a Dirichlet distribution ($\kappa = 30$) around the baseline weights. These weights were then used to generate pairwise comparison matrices on the Saaty 1–9 scale, with log-normal noise added to simulate variation in expert opinions. Matrices with a consistency ratio (CR) below 0.10 were retained (random index 1.32 for $n = 7$; Saaty, 2008).

(4) Structural sensitivity. Rankings from the baseline were compared with those from alternative model choices: the stepwise energy table of v1.0; geometric aggregation; removal of DI; removal of S2 from CN; removal of CN; z-score instead of min–max normalisation; equal dimension weights; and looser worst-case bounds for SEC and S2. A one-at-a-time scan changed each dimension weight by $\pm 50\%$ with renormalisation.

(5) Replication and thresholds. Analyses (2) and (4) were repeated across 50 independent synthetic datasets to check dependence on the random draw, and class edges were examined by bootstrap (Section 2.6).

All code was written in Python (NumPy, SciPy, scikit-learn) and is available with the supplementary material; every random process used a fixed seed.

2.9 Limitations of the design

- The data are synthetic; ranking stability shows that the index is not fragile with respect to the tested perturbations, not that it measures sustainability correctly.
- Positive correlation between dimensions is partly produced by the generator's latent factor, and this tends to favour ranking stability.
- The inherited dimension weights, the assumed reference bounds and the class edges derived from 40 synthetic plants have not been validated with experts or operational data.
- Scope 3 emissions, social indicators and cost are outside the index.
- DI is an enabling dimension, scored on capability rather than outcomes; its inclusion is a modelling choice that is examined, not defended, in the sensitivity analysis.

3. Results

3.1 Score distribution and dependence between dimensions

Baseline WISI scores of the 40 synthetic plants ranged from 29.1 to 76.0 (mean 53.5, SD 10.6). Mean dimension scores were highest for TP (88.3) and WS (63.2) and lowest for CE (28.8) and ES (29.9) (Table 4). The narrow spread of TP (SD 6.2) compared with the other dimensions (SD 14.6–20.4) is a consequence of the generator and of the compliance-oriented indicators it contains, which are bounded near their maximum.

Table 4
Dimension scores and influence on WISI (n = 40, baseline weights).

Dimension	Nominal weight	Mean	SD	Pearson r with WISI	Effective share of WISI variance
TP	0.20	88.3	6.2	0.72	0.083
WS	0.15	63.2	20.4	0.66	0.188
ES	0.15	29.9	16.5	0.73	0.168
CE	0.15	28.8	17.9	0.62	0.155
CN	0.10	55.3	20.1	0.65	0.122
AR	0.15	56.4	14.6	0.66	0.135
DI	0.10	35.4	20.4	0.79	0.150

r = Pearson correlation between the dimension score and WISI.

Nominal weights did not match effective influence (Fig. 3). TP carries the largest nominal weight (20%) but accounted for only 8.3% of the variance of WISI, whereas DI, with a nominal weight of 10%, accounted for 15.0%. Across 50 replicate datasets the effective share of TP was 4.6–9.6% (5th–95th percentile), so this gap is a stable property of the specification and not an artefact of one sample.

The ES–CN correlation was 0.47 in the main dataset and 0.58 (5th–95th percentile 0.40–0.71) across replicates. This partly reflects the physical link built into the generator, since purchased electricity enters Scope 2. The mean off-diagonal correlation among all dimensions was 0.38 (0.24–0.46) across replicates, a level that comes partly from the shared latent factor of the generator (Section 2.9).

3.2 Robustness to weight uncertainty

Under Dirichlet perturbation of the dimension weights, ranking remained stable for moderate and high uncertainty and degraded only when the weights were close to random (Table 5, Fig. 4–5).

Table 5
Ranking stability under weight uncertainty (10,000 draws; main dataset).

Scenario	Median Spearman ρ	5th pct ρ	Draws with $\rho \geq 0.9$	Top-5 overlap	Mean 5–95% rank-interval width
Dirichlet $\kappa = 50$	0.981	0.955	100%	0.95	5.9
Dirichlet $\kappa = 20$	0.963	0.909	96.9%	0.90	8.8
Dirichlet $\kappa = 3$	0.867	0.713	28.7%	0.76	17.5

With $\kappa = 50$ the typical facility shifted by about six rank positions between the 5th and 95th percentile of the simulated weight sets (Fig. 4). Rank intervals were widest for plants in the middle of the distribution and narrowest at both ends, which is expected when scores are densely packed in the centre. Across 50 replicate datasets the median ρ under $\kappa = 50$ was 0.985 (0.977–0.990) and the 5th-percentile ρ was 0.964 (0.946–0.974) (Fig. 6).

When expert disagreement was simulated through pairwise-comparison matrices (Section 2.8, item (3)), the consistency filter became selective only at higher judgement noise: acceptance was 100% at $\sigma = 0.25$, 81% at $\sigma = 0.5$, 10% at $\sigma = 0.8$ and 0.3% at $\sigma = 1.2$ (Table 6). Rankings derived from the accepted matrices remained close to the baseline at every level, with the 5th-percentile ρ between 0.905 and 0.934.

Table 6
Ranking stability under simulated AHP judgement noise (accepted matrices, CR < 0.10).

Judgement noise σ	Acceptance rate	Median ρ	5th pct ρ
0.25	100%	0.973	0.934
0.5	81.2%	0.970	0.925
0.8	9.9%	0.965	0.914
1.2	0.3%	0.950	0.905

A one-at-a-time scan, changing each dimension weight by – 50% or + 50% with renormalisation, gave a minimum ρ of 0.979.

3.3 Structural sensitivity

Aggregation form and z-score normalisation affected rankings more than moderate weight perturbation, whereas the other structural variants changed rankings very little, and all comparisons remained above $p = 0.94$ (Table 7, Fig. 7). Geometric aggregation changed rankings most ($p = 0.949$; maximum rank shift 12 positions; mean absolute score shift 7.1 points), which indicates that low-scoring dimensions, mainly ES, CE and DI in this dataset, are penalised more strongly than under compensatory aggregation. The stepwise energy table of v1.0 and looser worst-case bounds for SEC and S2 changed rankings very little ($p = 0.995$ and 0.999), so the choice of the linear form and the exact bounds had little effect on order, although they do change absolute scores.

Table 7
Structural sensitivity relative to the baseline specification.

Variant	Spearman ρ	Kendall τ	Mean abs. score shift	Max. rank shift
Looser worst bounds (SEC 1.2, S2 0.45)	0.999	0.987	1.12	1
SEC step table of v1.0	0.995	0.959	1.58	5
Equal dimension weights	0.992	0.949	2.43	4
Without DI	0.981	0.910	2.13	7
Without CN	0.979	0.910	1.44	6
CN = S1 only	0.974	0.885	1.40	7
z-score normalisation	0.962	0.859	5.05	8
Geometric aggregation	0.949	0.844	7.09	12

Removing S2 from CN or removing CN altogether produced rank correlations of 0.974 and 0.979, which shows that the electricity-driven overlap between ES and CN does not by itself dominate the ordering. Removing DI produced $\rho = 0.981$.

3.4 Classification

With the v1.0 edges (90, 80, 70, 60), 31 of the 40 plants (78%) fell in the lowest class, 5 in D, 4 in C and none in A or B; across replicates a median of 70% of plants (5th–95th percentile 62.5–78.9%) were in the lowest class. These edges therefore did not discriminate among plants once indicators were normalised against literature-anchored best and worst values. Data-derived edges were more discriminating (Table 8, Fig. 8): quintile edges are 62.7, 54.9, 50.3 and 46.3; k-means edges are 69.3, 60.0, 51.5 and 42.0 and yield class sizes of 4, 5, 13, 13 and 5.

Table 8
Class edges and class retention under weight uncertainty (Dirichlet $\kappa = 50$).

Edge set	Edges (A/B, B/C, C/D, D/E)	Mean class retention	Replicates: retention (5th–95th pct)
v1.0	90, 80, 70, 60	0.927	0.867–0.914
Quintile	62.7, 54.9, 50.3, 46.3	0.646	0.616–0.712
k-means	69.3, 60.0, 51.5, 42.0	0.761	0.694–0.801

The high retention of the v1.0 edges is not evidence of a robust classification, because most

plants are already in the lowest class and cannot move down. The main dataset lies slightly above the replicate range because its share of plants in the lowest class (78%) is higher than the replicate median (70%). For the k-means edges only 35% of plants kept their class in at least 80% of the draws. Bootstrap resampling of the facilities showed that the k-means edges are uncertain by roughly ± 5 points: 67.4 (95% interval 60.5–70.5), 57.6 (51.6–60.9), 49.5 (44.8–53.0) and 40.0 (35.1–46.0). Class assignment near an edge should therefore be reported together with its score and an uncertainty band.

3.5 Feasibility check on two operating plants

To check what the index requires in practice, the daily laboratory reports of two operating wastewater treatment plants (Plant A and Plant B; names withheld) for 1 January to 30 September 2026 (273 days) were examined. The data were cleaned without imputation: unambiguous entry errors were corrected, doubtful values were set to blank and logged, and days without a report were excluded. The reports contain influent and effluent quality only. Flow records were missing for January–March, and energy, chemicals, sludge, asset and digital records were not available. Of the 22 indicators, three (TES, PAS and PSS, all in TP) can be derived from such data: TES only if a removal target is defined, and PAS only as a proxy based on reporting days. CUS additionally needs flow and design capacity. No removal target was available; with observed removals of 96–99%, TES would be at or near its cap for any target at or below the observed removal. No WISI score was computed, because six of the seven dimensions have no data and a score built on TP alone would not be meaningful. Table 9 gives the quantities that could be calculated.

Table 9
Quantities computable from routine laboratory data (273 days).

Quantity	Plant A	Plant B
Reporting days (proxy for PAS)	264 (96.7%)	273 (100%)
Removal of mean concentration, BOD / COD / TSS / NH ₃ (%)	97.2 / 96.0 / 98.7 / 99.8	97.8 / 97.1 / 98.1 / 98.1
CV of effluent, BOD / COD / TSS / NH ₃ (%)	10 / 23 / 45 / 63	17 / 11 / 33 / 41
PSS by parameter, BOD / COD / TSS / NH ₃	81 / 53 / 11 / 0	66 / 78 / 35 / 17
PSS, mean of the four parameters	36	49

Three points follow. Removals were close to their ceiling in both plants, so TES would discriminate little, as the synthetic results for TP suggested. PSS depends strongly on the effluent parameter used: for Plant A it ranges from 81 (BOD) to 0 (NH₃), which is why the specification averages it over the monitored parameters. Finally, routine laboratory data cover three of the 22 indicators, and the remaining dimensions need flow, electricity, sludge, emission and asset records that most plants hold but do not report together. The check shows feasibility and data needs, not validity.

3.6 Summary of findings

1. Rankings were insensitive to moderate weight perturbation (median $\rho \geq 0.96$ for $\kappa \geq 20$) and degraded only when weights were nearly random (median $\rho = 0.87$).
2. Aggregation form and z-score normalisation influenced ranking more than moderate weight changes, whereas the other structural variants changed it very little; no tested variant fell below $\rho = 0.94$.
3. Nominal weights differ from effective influence; the largest nominal weight (TP) had the smallest effective share.
4. Fixed v1.0 class edges are not appropriate for the new normalisation, and data-derived edges are themselves uncertain with samples of this size.
5. The ES–CN overlap is moderate ($\rho \approx 0.4$ – 0.7) and does not dominate rankings.

These results apply to synthetic data generated with a shared latent factor. A real facility sample may show weaker inter-dimension correlation and therefore lower ranking stability.

4. Discussion

4.1 What the simulation does and does not show

The simulation shows that the reformulated index is structurally stable: rankings changed little under moderate weight uncertainty, under simulated expert disagreement and under the tested changes of model structure. It does not show that WISI measures sustainability correctly, because the data were generated for this study and share a latent quality factor that raises the correlation between dimensions. When dimensions move together, different weight sets tend to produce similar rankings, so part of the observed stability is a property of the data and not of the index. The stability figures should therefore be read as an upper bound for populations with strongly correlated performance, and a test on real plants is the necessary next step.

4.2 Nominal weights are not effective influence

The largest nominal weight, 0.20 for Technical Performance, corresponded to only about 7–8% of the variance of the index. Technical indicators of compliant plants cluster near their maximum, so they separate facilities poorly, however heavily they are weighted. Conversely, Digital Intelligence and Water Sustainability, which vary widely among plants, had a larger effective share than their nominal weight. This is a known feature of weighted composite indices, in which weights are interpreted as trade-offs rather than as importance coefficients (Nardo, et al., 2008). It has a practical consequence: experts who assign weights in a Delphi–AHP exercise should be told that a weight expresses how much one score point in a dimension is worth in the aggregate, and should be shown how effective influence changes when weights are altered. Otherwise, a high weight on a dimension where all plants score alike will not produce the intended emphasis.

4.3 Compensation and aggregation

Among the tested structural variants, geometric aggregation changed the ranking most. A geometric mean limits compensation, so a plant with a very weak dimension is penalised more than under the arithmetic mean. This matters for the intended use of WISI: if the aim is to flag plants that are strong on average but fail in one area (for example, no resource recovery), a non-compensatory or partly compensatory rule is better aligned with that aim. The choice should be made explicitly with the end users of the index. We retained arithmetic aggregation as baseline for continuity with v1.0 and recommend reporting the dimension profile alongside the single score.

4.4 Dependence between dimensions and the role of digital capability

Electricity links Energy Sustainability and Scope 2 emissions, and the ES–CN correlation in the simulation (about 0.4–0.7) reflects this. Removing S2 or CN altogether changed rankings only slightly, which suggests that the overlap does not dominate the index. It still means that a plant that reduces electricity use is rewarded twice, once in ES and once in CN, and this should be stated when scores are communicated. Users who prefer strict independence can drop S2 from CN, as tested in Section 3.3, so that CN reflects direct emissions only, or report ES and CN separately.

Digital Intelligence differs in kind from the other dimensions, because it describes capability to act and not an outcome. Its removal altered rankings little ($p = 0.981$), but its effective share of variance was larger than its nominal weight, so it can have visible influence on scores of plants with few digital tools. Whether such a dimension belongs inside the index or beside it, as an enabling indicator reported separately, deserves discussion with practitioners before v2.0.

4.5 Classification

Bands taken over from v1.0 did not fit the new normalisation: most synthetic plants fell into the lowest class. Data-derived edges spread plants better but were uncertain by about five score points with 40 plants, and with k-means edges only about a third of facilities kept their class in at least 80% of the weight draws. Quintile edges were less stable (retention 0.62–0.71 across replicates).

Two conclusions follow. First, class edges should be set from a sufficiently large reference sample of real plants of comparable type and size, and updated as the sample grows. Second, reported classes should be accompanied by the score and a band of uncertainty, because a class label hides how close a plant is to an edge.

4.6 Positioning relative to existing frameworks

WISI is intended to complement, not replace, established utility assessment systems. Performance-indicator systems for wastewater service (Matos, et al., 2003) and ISO 24511:2024 provide indicators and assessment criteria for utility management; ISO 24511 sets no target values or thresholds (ISO, 2024). Envision rates infrastructure projects through planning, design and construction (ISI, 2018). What WISI proposes is a single score that combines resilience, resource recovery, carbon and digital capability with technical performance, and that records, indicator by indicator, whether the direction, bounds and weight are literature-based or assumed. Table 1 compares these approaches by level, output and basis for thresholds and weights.

4.7 Implications for practice and roadmap

For utilities, the reformulated index provides a transparent calculation procedure that can be run with routine plant data. Indicators that are difficult to obtain, such as phosphorus recovery or criticality-weighted failure risk, can be reported as not available and the dimension rescaled, provided the missing share is disclosed. Regulators and owners should use the score for screening and prioritisation and not for ranking plants of different type or size without comparable reference bounds.

A feasibility check on two operating plants (Section 3.5) showed that routine laboratory data cover only three of the 22 indicators, so a full application requires flow, electricity, sludge and asset records.

The path from this study to a validated index has four steps:

1. **Expert weighting (v2.0).** Delphi–AHP with a panel of practising engineers and utility managers, whose size and selection criteria are reported, with results compared with the baseline weights and with objective weights such as CRITIC or entropy.
2. **Calibration with real data.** Reference bounds and class edges derived from operational data of a sample of plants, with local grid emission factors.
3. **Structural validation.** Principal component analysis, internal consistency and correlation with independent outcomes such as operating cost per unit of treated water, using real plant data.
4. **Extension.** Adaptation of indicator sets and bounds to desalination, reuse and distribution facilities, each with its own calibration.

4.8 Limitations

Beyond the design limitations in Section 2.9, the study did not test whether other indicator choices within a dimension would change conclusions, did not include uncertainty in the indicator values themselves (measurement error, missing data), and did not examine temporal stability of scores. These are suitable subjects for work with real operational data.

5. Conclusions

This study reformulated the conceptual WISI framework into a computable index for wastewater treatment plants and examined its structural robustness through simulation on synthetic facilities. Rankings were stable under moderate weight uncertainty (median Spearman $\rho = 0.98$ for Dirichlet $\kappa = 50$) and under simulated expert disagreement, and varied most with aggregation form and normalisation method among the tested variants. Nominal weights did not reflect effective influence: technical performance, with the largest weight, accounted for about 7–8% of index variance. Class edges inherited from the earlier framework placed a median of 70% of synthetic plants in the lowest class, and data-derived edges were uncertain by about five score points. These findings concern the structure of the index and not its validity for real plants, which requires expert weighting and calibration with operational data. A feasibility check on two operating plants showed that routine laboratory data cover only three of the 22 indicators, so applying the index requires flow, energy, sludge and asset records.

The indicator definitions, the reference-value table and the accompanying code provide a basis for the staged validation described in Section 4.7 and for comparison with existing assessment systems.

Declarations

Conflict of Interest

The author declares no conflict of interest.

AI Usage Statement

Artificial intelligence tools were used to support language refinement, editing assistance, and manuscript development. All technical concepts, methodologies, interpretations, and conclusions remain the responsibility of the author.

Author Affiliation

AYMAN RAMADAN

International Business Academy of London

Kingdom of Saudi Arabia

Author Contributions

Conceptualization, Methodology, Formal Analysis, Framework Development, Writing – Original Draft, Writing – Review and Editing.

Data Availability Statement

No proprietary datasets were used in the development of this manuscript. All examples are illustrative unless otherwise stated.

References

1. Campbell JD, Jardine AKS, McGlynn J, Barry DM (2024) Asset Management Excellence: Optimizing Equipment Life-Cycle Decisions, 3rd edn. CRC, Boca Raton, FL
2. Chrispim MC, Scholz M, Nolasco MA (2019) Phosphorus Recovery from Municipal Wastewater Treatment: Critical Review of Challenges and Opportunities for Developing Countries. *J Environ Manage* 248:109268
3. Drewnowski J, Remiszewska-Skwarek A, Duda S, Łagód G (2019) Aeration Process in Bioreactors as the Main Energy Consumer in a Wastewater Treatment Plant: Review of Solutions and Methods of Process Optimization. *Processes* 7(5):311
4. Estévez S, Arias A, Feijoo G, Moreira MT (2025) Methodological Guide and Roadmap to Assess the Compliance of Wastewater Treatment Plants with Sustainability and Circularity Criteria. *Water Res* 274:123125
5. Fuller A, Fan Z, Day C, Barlow C (2020) Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access* Volume 8:108952–108971
6. Hollnagel E (2011) *Resilience Engineering in Practice: A Guidebook*. Farnham. Ashgate Publishing, Surrey, England
7. IPCC (2023) Climate Change 2023: Synthesis Report. IPCC, Geneva
8. ISI (2018) Envision Sustainable Infrastructure Framework: Guidance Manual. Washington, DC, Institute for Sustainable Infrastructure
9. ISO (2014) *ISO 55001:2014. Asset Management – Management Systems – Requirements*. International Organization for Standardization (ISO), Geneva, Switzerland
10. ISO (2019) *ISO 14090: Adaptation to Climate Change – Principles, Requirements and Guidelines*. International Organization for Standardization, Geneva
11. ISO (2024) *ISO 24511:2024. Activities Relating to Drinking Water and Wastewater Services – Guidelines for the Management of Wastewater Utilities and for the Assessment of Wastewater Services*. International Organization for Standardization (ISO), Geneva
12. Matos R et al (2003) Performance Indicators for Wastewater Services. IWA Publishing, London
13. Molinos-Senante M et al (2014) Assessing the Sustainability of Small Wastewater Treatment Systems: A Composite Indicator Approach. *Sci Total Environ* 497–498. 10.1016/j.scitotenv.2014.08.026
14. Nardo M et al (2008) Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing, Paris, France
15. OECD (2013) Water and Climate Change Adaptation: Policies to Navigate Uncharted Waters. OECD Publishing, Paris
16. Ramadan ASA (2026) *Water Infrastructure Sustainability Index (WISI): A Predictive Engineering Framework for Resilient, Circular, and Sustainable Water Systems*. <https://doi.org/10.5281/zenodo.23405289> ed. s.l.:Zenodo, 174 pp
17. Silva C, Rosa MJ (2022) A Comprehensive Derivation and Application of Reference Values for Benchmarking the Energy Performance of Activated Sludge Wastewater Treatment. *Water* 14(10):1620
18. Faragò M, Rebsdorf M, Damgaard A, Rygaard M (2022) An Operational Framework to Quantify the Sustainability of Water Resource Recovery Facilities. Singapore International Water Week 2022, Singapore
19. Molinos-Senante M et al (2014) Assessing the Sustainability of Small Wastewater Treatment Systems: A Composite Indicator Approach. *Sci Total Environ* 497–498. <https://doi.org/10.1016/j.scitotenv.2014.08.026>
20. Popović T, Kraslawski A, Avramenko Y (2013) Applicability of Sustainability Indicators to Wastewater Treatment Processes. In: 23rd European Symposium on Computer Aided Process Engineering. Elsevier, pp. 931–936

21. Saaty TL (2008) Decision Making with the Analytic Hierarchy Process. Int J Serv Sci 1(1):83–98

22. Sun Y et al (2020) A Composite Indicator Approach to Assess the Sustainability and Resilience of Wastewater Management Alternatives. Sci Total Environ 725:138286

Figures

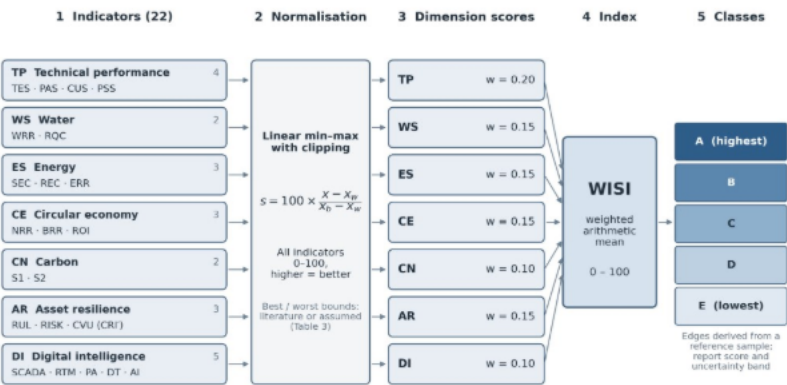


Figure 1

Structure of WISI v1.1 for wastewater treatment plants: 22 indicators in seven dimensions are normalised to 0–100, aggregated by weighted arithmetic means into dimension scores and the index, and assigned to classes A–E. Dimension weights are working hypotheses inherited from v1.0.

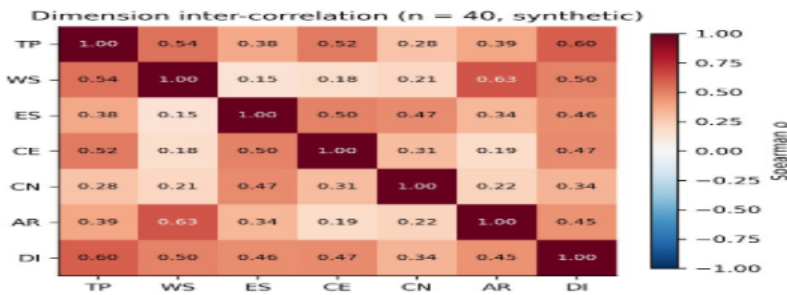


Figure 2

Spearman rank correlations among the seven WISI dimension scores (n = 40 synthetic plants).

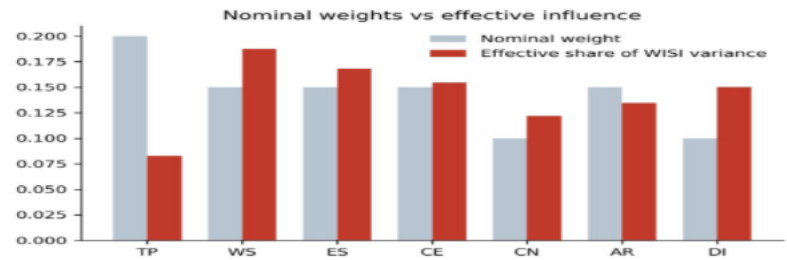


Figure 3

Nominal dimension weights compared with each dimension's effective share of WISI variance.

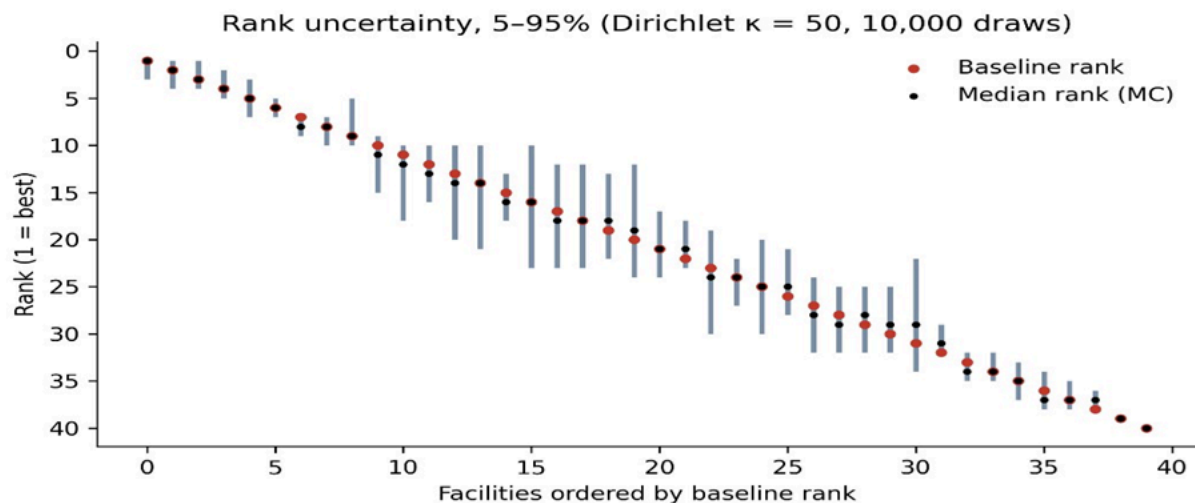


Figure 4

Rank uncertainty of each plant under weight perturbation (Dirichlet $\kappa = 50$, 10,000 draws); bars show the 5th–95th percentile of rank.

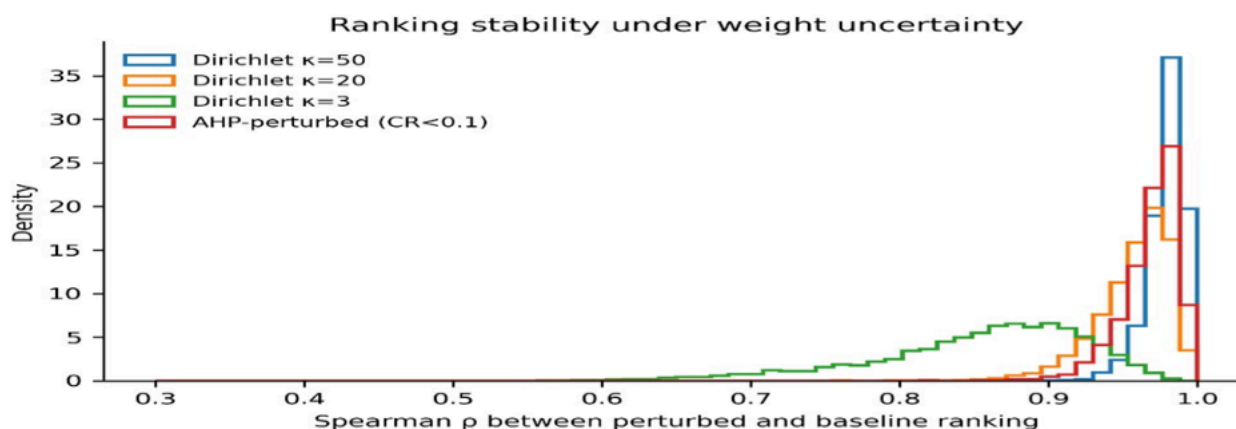


Figure 5

Distribution of Spearman ρ between perturbed and baseline rankings for Dirichlet weight perturbation ($\kappa = 50, 20, 3$) and for Simulated AHP weights ($CR < 0.10$).

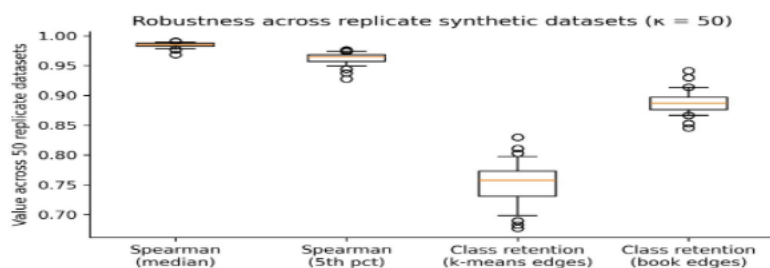


Figure 6

Robustness measures across 50 replicate synthetic datasets (Dirichlet $\kappa = 50$).

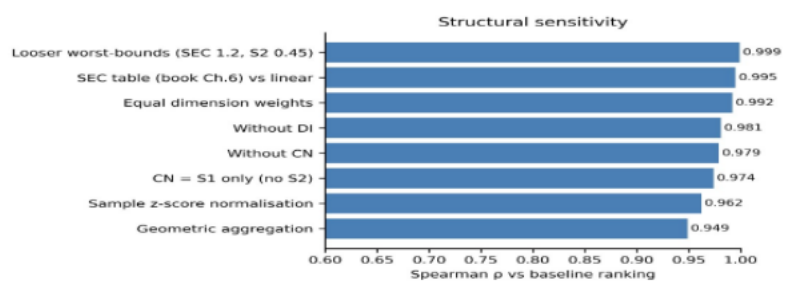


Figure 7

Rank correlation (Spearman ρ) between the baseline ranking and rankings under alternative model structures.

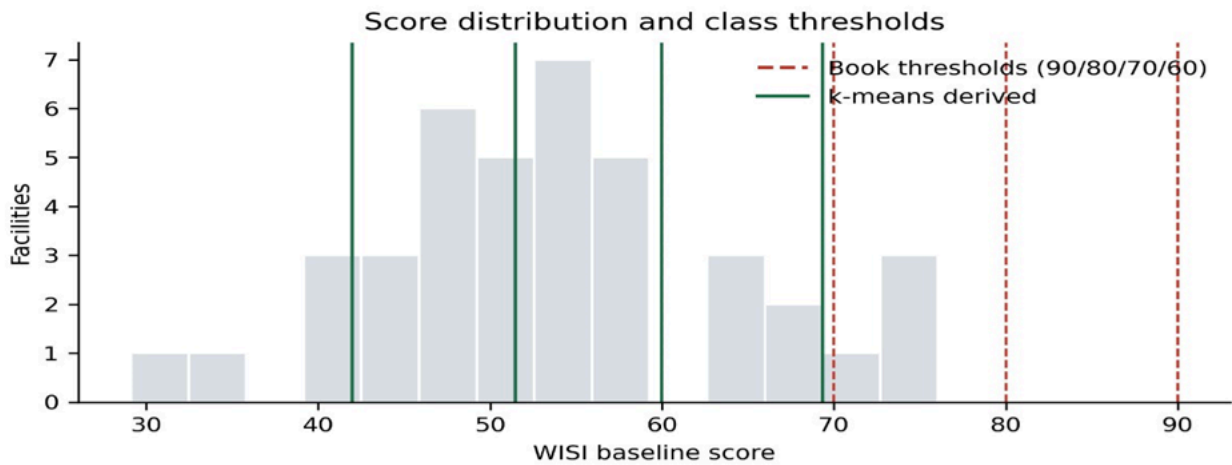


Figure 8

Distribution of baseline WISI scores with the v1.0 class edges and the k-means-derived edges.