Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Class-Adaptive Hybrid Mean Aggregation Framework for River Water-Quality Classification: Resolving Eclipsing and Ambiguity in the Malaysian DOE-WQI

NUR HUDA NABIHAN MD SHAHRI

nurhuda8371@uitm.edu.my

Universiti Teknologi MARA

ZALINA MOHD ALI

National University of Malaysia

Research Article

Keywords: Water Quality Index, aggregation function, generalised power mean, eclipsing effect, ambiguity, DOE-WQI, Malaysian rivers, sensitivity analysis, bootstrap validation

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11518477/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Class-Adaptive Hybrid Mean Aggregation Framework for River Water-Quality Classification: Resolving Eclipsing and Ambiguity in the Malaysian DOE-WQI

Nur Huda Nabihan Md Shahri[1,2], and Zalina Md Ali[1*]

[1] *Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.*

[2] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.*

** Corresponding author: zalina99@ukm.edu.my*

Abstract

Conventional aggregation operators in Water Quality Index (WQI) systems such as the weighted arithmetic mean used by the Malaysian Department of Environment (DOE-WQI) is suffering from problem in eclipsing and ambiguity. To resolve this problem, this study proposes a class-adaptive **Hybrid Mean Function** that applies band specific hybrid aggregation operators across DOE water-quality categories (Clean, Slightly Polluted, Polluted).

Within a generalized power mean equation with eight model variants (A–H) representing a 2^3 factorial combination of min or max operators were evaluated under both equal and unequal weighting schemes. The framework was validated using 630 observations across six parameters (DO, BOD, COD, SS, AN, pH) from Malaysian rivers from year 2017 to 2021and benchmarked against DOE-WQI alongside three PCA-derived variants. Robustness was evaluated via non-parametric classification metrics, bootstrap resampling (B = 1000), and PCA sensitivity analysis.

Model F (min, max, max under unequal weighting) emerged as uniquely robust, achieving a mean class-similarity of 80.8%, a macro-F1 score of 0.785, and an overall accuracy of 0.860 is outperforming the next-best model in the Slightly Polluted class by 5.9 percentage points. By involving **min** in the clean class and **max** in polluted class, Model F simultaneously reduces eclipsing and ambiguity problem. This shows a mathematically transparent, highly accurate alternative for river monitoring and sub-basin prioritization in Malaysia.

Keywords: Water Quality Index; aggregation function; generalised power mean; eclipsing effect; ambiguity; DOE-WQI; Malaysian rivers; sensitivity analysis; bootstrap validation.

1. Introduction

Surface water quality has become a focal concern for environmental governance because rivers sustain ecosystems, agriculture, and human consumption simultaneously (Banda & Kumarasamy, 2020). A Water Quality Index (WQI) compresses multiple parameter monitoring data typically dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), ammoniacal nitrogen (AN) and pH into a single scalar value that changes can communicate to the public and use as a decision (Abbasi & Abbasi, 2012; Sutadian et al., 2016; Uddin et al., 2021). In Malaysia, the Department of Environment (DOE) using WQI as the reference framework, with combining six sub-indices through a fixed weighted arithmetic mean and mapping the resulting score to three status class such as Clean Class (81–100), Slightly Polluted Class (60–80), and Polluted Class (0–59) (Al-Badaii et al., 2013; Aris et al., 2014).

Although operationally convenient, the DOE-WQI inherits two well-documented structural pathologies of additive aggregation. The first is the eclipsing problem, in which a parameter that is genuinely out of compliance is masked by co-monitored parameters that are well within range; because each sub-index contributes only a fixed weighted fraction of the aggregate based on selected parameter, an extreme deviation can fail to trigger a reclassification (Landwehr et al., 1974; Smith, 1990). The second is ambiguity, in which the aggregate is flagged as polluted even when all individual sub-indices are within acceptable limits indicates a symptom of compensatory weighting on heterogeneously distributed sub-indices (Mustapha, 1982; Sutadian et al., 2016). These pathologies are not correctable by recalibrating weights alone; they are a part of the fixed functional form aggregation, and they become acute when sub-indices distribute differently across the three DOE status class at a single monitoring station.

A further limitation of the current DOE-WQI is that it is a global operator applied uniformly across water-quality states regardless the important of parameter in each of monitoring stations , whereas the information content of a sub-index depends strongly on the band in which it lies. In the Clean class, a single low-performing parameter is the most informative signal of emerging degradation; averaging dilutes it. Conversely, in the Polluted class, a single very-low parameter dominates the arithmetic mean and penalises stations whose other sub-indices are merely borderline; here a more permissive operator preserves regulatory resolution. Any aggregation scheme that treats the three class symmetrically therefore cannot simultaneously minimise eclipsing at the upper end and ambiguity at the lower end.

This study addresses that gap by developing a class-adaptive Hybrid Mean Function (HMF) that selects the aggregation operator as a function of the class membership of each sub-index. The contribution is threefold. Based on **Figure 1** below, first we embed the classical arithmetic, minimum and maximum aggregators within a single generalised power-mean formalism, yielding an explicit unified expression for the hybrid function in **2.3 The unified Hybrid Mean Function**. Second, we define eight variants (Models A–H) as a full 2^3 factorial of conservative (min) versus optimistic (max) operators across the three DOE bands, under both equal and unequal weight schemes. Third, we evaluate the sixteen resulting specifications against the DOE-WQI benchmark on 630 river observations from year 2017 to 2021 through non-parametric classification similarity, per-class confusion-matrix diagnostics, and PCA-based sensitivity analysis; bootstrap resampling quantifies stability.

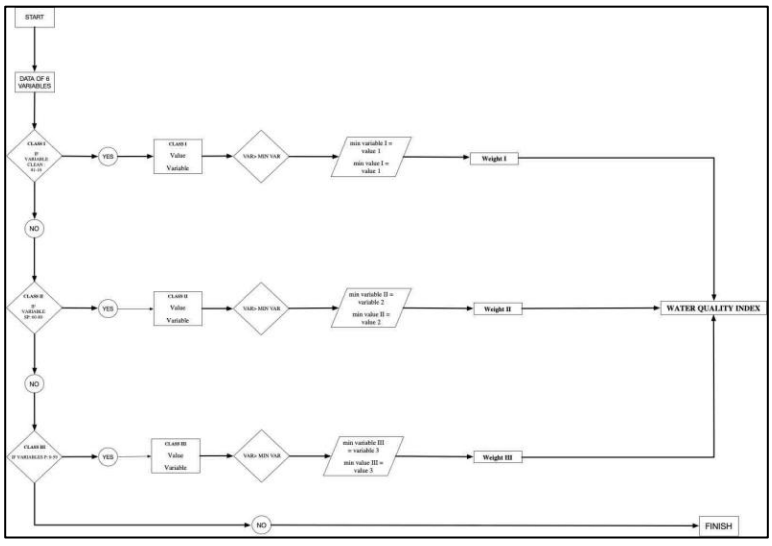


Figure 1: Flowchart of Hybrid Mean Function Equal Weight (EW)

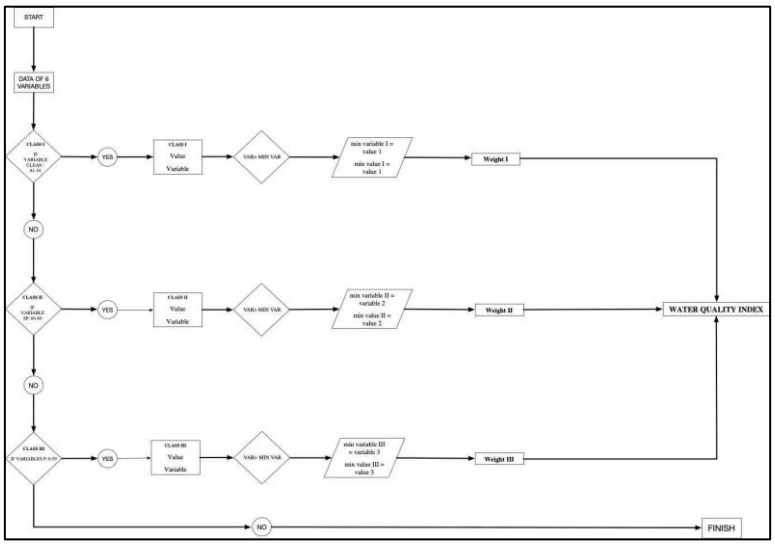


Figure 2: Flowchart of Hybrid Mean Function Unequal Weight (UW)

2. Materials and Methods

2.1 Study area and dataset

The analysis uses 630 observations from water quality monitoring stations operated by the Department of Environment Malaysia over the period 2017–2021, covering several major river basins across Peninsular and East Malaysia. Six parameters were retained as recommended by the DOE protocol: dissolved oxygen (DO, mg L^{-1}), biochemical oxygen demand (BOD, mg L^{-1}), chemical oxygen demand (COD, mg L^{-1}), total suspended solids (SS, mg L^{-1}), ammoniacal nitrogen (AN, mg L^{-1}) and pH (dimensionless). DOE-WQI class labels (Clean: 81–100; Slightly Polluted: 60–80; Polluted: 0–59) served as the benchmark class assignment against which all hybrid specifications were evaluated.

Table 1 : DOE Water Quality Classification based on Water Quality Index

Water Quality Index	**Index Range**		
	Clean	**Slightly Polluted**	**Polluted**
WQI	81-100	60-80	0-59

2.2 Baseline DOE-WQI formulation

Each raw concentration x_j is mapped onto its best-management-practice (BMP) sub-index $S_i \in [0,1]$ through the Malaysian DOE piecewise linear transformation functions (Department of Environment Malaysia, 2024) . The baseline DOE-WQI aggregates the six sub-indices as:

$$WQI_{DOE} = 0.22\,S_{DO} + 0.19\,S_{BOD} + 0.16\,S_{COD} + 0.15\,S_{AN} + 0.16\,S_{SS} + 0.12\,S_{pH} \quad (1)$$

with the fixed coefficients reflecting expert-panel weights calibrated for Malaysian river systems (Mohd Ali et al., 2013). Equation (1) is a weighted arithmetic mean and is therefore subject to the eclipsing and ambiguity pathologies described in Section 1.

2.3 The Hybrid Mean Function

2.3.1 The input variables and Quality Classes

Let X = $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ represent the set of N = 6 measured water quality parameter values. The input domain is partitioned into three mutually exclusive quality intervals:

Table 2. DOE water-quality status bands used for class-assignment.

Status band	Index range	Operational implication
Clean (Class I)	81 ≤ WQI ≤ 100	Suitable for conventional treatment; supports aquatic life and direct-contact recreation; baseline for long-term trend monitoring.
Slightly Polluted (Class II)	60 ≤ WQI ≤ 80	Requires extensive treatment for potable use; early-warning band for point- and non-point-source stress; priority for catchment-level intervention.
Polluted (Class III)	0 ≤ WQI ≤ 59	Unsuitable for domestic or recreational use without treatment; requires mandatory enforcement action under the Environmental Quality Act 1974 and source-apportionment studies.

2.3.2 Four-Step Mathematical Process

Step 1: Categorizing Inputs via Indicator Functions

To mathematically categorize each variable input, x_i, into respective class $k \in \{1,2,3\}$ without relying on procedure IF-ELSE conditional statements, we utilize the Indicator Functions, defined as :

Indicator Function Definition:

$$\mathrm{I_k(x_i) = 1 \text{ if } x_i \in \text{Interval}_k, \text{else } 0} \quad (2)$$

Example: If $x_1 = 85$, then $I_1(85) = 1$ (Class I Active), while $I_2(85) = 0$ and $I_3(85) = 0$.

This function acts as a mathematical sorting mechanism, defining the active subset of variables S_k for each class:

$$\mathrm{S_k = \{x_i \in X \,|I_k(x_i) = 1\}} \quad (3)$$

Step 2: Isolating Variables based on Functions

For each active class subset S_k, the flowchart compares variables to extract the value. There is combination of two functions which are:

Discrete Form:

$$\mathrm{m_k = \min\{x_i | x_i \in S_k\}} \quad (4)$$

$$m_k = max\{x_i | x_i \in S_k\} \quad (5)$$

To make this operation fully continuous and differentiable across variable domain transitions, we employ the Soft-Min exponential weighting formulation:

Continuous Soft-Min Formulation:

$$m_k \approx \frac{[\sum_{i=1}^{6} x_i . I_k(x_i) . e^{-\propto . x_i}]}{[\sum_{i=1}^{6} I_k(x_i) . e^{-\propto . x_i}]} \quad (6)$$

Where:

$I_k(x_i)$: Zeroes out variables not belonging to Class k.

$e^{-\propto . x_i}$: Assigns exponentially larger weights to smaller values of x_i

$\propto$ (alpha): Sharpness parameter controlling convergence towards the exact function.

Step 3: Handling Empty Classes via Existence Marking

When no variable fall into a given class (e.g., $S_k = \emptyset$), direct minimum evaluations become undefined. We introduce a binary Existence Filter $\delta_k$to mask inactive classes:

$$\delta_k = \max(0, sgn(|S_k|)) \quad (7)$$

Where:

$|S_k|$ is the number of variables in Class k. if $|S_k| > 0, \delta_k = 1$; if $|S_k| = 0$, $\delta_k = 0$.

Step 4: Weighted Aggregation & Final Index Computation

Each class minimum m_k or maximum $m_k$are scaled by its associated weight $w_k(Weight\ I, Weight\ II\ or\ \ Weight\ III)$. The overall Water Quality Index(WQI) is synthesized through vector dot product formulation:

$$WQI = \sum_{k=1}^{3} w_k . m_k . \delta_k = w^T(m \circ \delta) \quad (8)$$

Where:

$w = [w_1, w_2, w_3]^T$ is the weight vector, $m = [m_1, m_2, m_3]^T$ is the class minimum or maximum vector $\delta = [\delta_1, \delta_2, \delta_3]^T$is the existence filter, and $' \circ '$ represents the element-wise Hadamard product.

2.3.3 Mathematical Justification: Exponential Weighting and Sharpness (α)

The standard hard-minimum function min(A, B) and hard-maximum function max(A,B) are non differentiable at boundary points. In statistical optimization, gradient calculations require smooth transitions. The exponential function $e^{-\alpha x}$ provides infinite differentiability (C^{∞}) while ensuring that smaller values exponentially dominate the weighted average.

While for sharpness the optimal alpha parameter (α), the sharpness parameter α controls how closely Soft-Min approaches the exact minimum:

As α→ ∞, Soft-Min converges exactly to the hard minimum.

As $\alpha \rightarrow 0$, Soft-Min behaves like a standard arithmetic mean.

2.3.4 Numerical Demonstration

Suppose a test sample yields 6 measured variables: $X = \{85, 92, 65, 70, 78, 90\}$. Weights assigned: $w_1 = 0.50, w_2 = 0.30, w_3 = 0.20$.

1. Class Allocation (Step 1):

• Class I (81–100): S_1 = {85, 92, 90} → $|S_1| = 3, = 1$

• Class II (60–80): S_2= {65, 70, 78} → $|S_2| = 3, \delta_2 = 1$

• Class III (0–59): $S_3 = \emptyset \rightarrow |S_3| = 0 = 0$

2. Minimum Isolation (Step 2):

• $m_1 = \min(85,92,90) = 85$

• $m_2 = \min(65,70,78) = 65$

• $m_3 = 0(masked\ out) = 0$ (Masked out)

Maximum Isolation:

• $m_1 = \max(85,92,90) = 92$

• $m_2 = \max(65,70,78) = 78$

• $m_3 = 0(masked\ out) = 0$ (Masked out)

3. Index Calculation (Steps 3 & 4):

• Min : $WQI = (0.50 \times 85) + (0.30 \times 65) + (0.20 \times 0) = 62$

• Max: $WQI = (0.50 \times 92) + (0.30 \times 78) + (0.20 \times 0) = 69.4$

Table 3. Class-adaptive operator triplets defining Models A–H, with analytical implications of each specification.

Model	S_1 **(Clean)**	S_2 **(Slightly Polluted)**	S_3 **(Polluted)**	**Analytical character and implication**
A	min	min	min	Uniformly conservative (worst-case). Maximally sensitive to any parameter degradation within each band, at the cost of over-classifying borderline stations as polluted. Useful as a regulatory early-warning benchmark.
B	max	max	max	Uniformly optimistic (best-case). Classifies stations by their strongest parameter in each band; masks single-parameter violations, making it unfit for compliance monitoring but informative as an upper envelope.
C	min	min	max	Strict in Clean and Slightly Polluted bands, permissive in Polluted band. Prioritises sensitivity

				where water is still usable; tolerates variability where remediation is already implied.
D	min	max	min	Strict at both extremes, permissive mid-range. Sharpens the Clean/Polluted boundary while absorbing noise in the Slightly Polluted band; suitable for binary clean-vs-impaired reporting.
E	max	min	min	Permissive only for Clean classification; strict for any degradation. Issues a Clean label only when every in-range parameter is strong — a high-confidence Clean signal.
F	min	max	max	Anti-eclipsing in Clean band (min forces all in-range parameters to be high to earn a Clean label), anti-ambiguity in degraded bands (max prevents single outliers from collapsing the aggregate). Dual robustness; see Discussion.
G	max	min	max	Strict only in the middle band; permissive at both extremes. Concentrates regulatory resolution on the transition zone between clean and polluted states.
H	max	max	min	Permissive in upper bands, strict in Polluted. Delivers a conservative Polluted signal while tolerating variability in cleaner waters; mirrors hazard-based precautionary schemes.

2.5 Weighting schemes and final WQI

Two weighting schemes are examined. Under equal weighting, each of the three class representatives contributes equally:

$$WQI^{m}EW = \frac{1}{3}\sum_{k=1}^{3} C_{k^{(m)}}, \tag{9}$$

Under unequal weighting, each band k receives a prior weight W_k derived from the Delphi-type elicitation adopted in the original DOE-WQI (Mohd Ali et al., 2013). These are normalised to sum to unity,

$$\omega_k = \frac{W_k}{(W_1+W_2+W_3)}, \text{ with } \sum_{k=1}^{3} \omega_k = 1, \tag{10}$$

and the unequal-weight hybrid aggregate becomes

$$\mathrm{WQI^{m}UW} = \sum_{\mathrm{k=1}}^{3} \omega_{\mathrm{k}} \mathrm{C}_{\mathrm{k^{(m)}}}, \tag{11}$$

The final WQI score is mapped onto the DOE class labels using the classification in Table 2. Because the class representative S_k are computed over partition-consistent sub-indices, equation above always yield an aggregate that lies within [0,100] and is monotone in each S_j which is a property that equation (1) does not guarantee under arbitrary sub-index distributions.

2.6 Evaluation and sensitivity analysis framework

Each of the sixteen combination (Model A to H) and each of the models will calculate both weighted and unweighted. At the end, the values will produce a predicted class status for every observation. produces a predicted class label for every observation. Then, the DOE-WQI will be use as a

benchmark functions. Performance of each of the models are quantified along four complementary dimensions

Classification similarity. A 3 × 3 crosstabulation of DOE-WQI (benchmark) against the model's predicted class is constructed based on class combinations. Per-class similarity is the proportion of observations on the diagonal, and overall similarity is the share of all observations correctly classified. This is a non-parametric, distribution-free concordance measure that does not rest on the normality of the aggregate.

1. **Confusion-matrix diagnostics.** For each S_k we report precision $S_{k=}\frac{TP_k}{FP_k}$, recall (sensitivity) $R_{k=}\frac{TP_k}{TP_k+FN_k}$, specificity $SP_{k=}\frac{TN_k}{TN_k+FP_k}$, $F1_{k=}\frac{2P_k\cdot R_k}{P_k+R_k}$ and accuracy $A_{k=}\frac{TP_k+TN_k}{N}$. Macro-averages across classes summarize overall performance.

2. **PCA-based sensitivity analysis.** To test whether the classification is an artefact of particular sub-index weightings, three PCA-WQI variants are constructed using squared-component-loading normalization, component-loading normalization, and absolute-component-loading normalization and each is cross tabulated against DOE-WQI. This probes the dependence of the aggregate on the weight structure independently of the aggregation operator.

3. **Bootstrap robustness.** Synthetic monitoring datasets (B = 1 000 draws) spanning the full permissible range of each parameter are generated to probe the behaviour of the hybrid function in edge cases and near-boundary observations. Stability is assessed by the bootstrap standard error of macro-F1.

3. Results

3.1 PCA-derived WQI variants as a sensitivity benchmark

Three Principal Component Analysis (PCA)-WQI specifications were crosstabulated against the DOE-WQI benchmark (Tables 3–5). Under squared component-loading normalisation (Table 3) the overall classification similarity reaches 85.87 %, falling to 66.98 % under component-loading normalisation (Table 4) and to 86.51 % under absolute-loading normalisation (Table 5). The dispersion across the three normalisation schemes — a range of ~20 percentage points — establishes that the aggregate is non-trivially sensitive to the weight structure even when the aggregation operator is held constant. Any claim about model performance must therefore be robust across weighting schemes, which motivates the unequal-weight re-evaluation of Models A–H in Section 3.2.

Table 3. Crosstabulation of DOE-WQI against PCA-WQI under normalisation of squared component loadings; row values are PCA-WQI counts, column totals are DOE-WQI counts. Overall similarity = 85.87 %.

PCA-WQI \ DOE-WQI	Clean	Slightly Polluted	Polluted	Row total	% diagonal
Clean	371	6	0	377	98.41 %
Slightly Polluted	54	120	28	202	59.41 %
Polluted	0	1	50	51	98.04 %
Column total	425	127	78	630	—
% column diagonal	87.29 %	94.49 %	64.10 %	—	85.87 %

Table 4. Crosstabulation under normalisation of raw component loadings. Overall similarity = 66.98 %.

PCA-WQI \ DOE-WQI	Clean	Slightly Polluted	Polluted	Row total	% diagonal
Clean	315	62	0	377	83.55 %
Slightly Polluted	91	100	11	202	49.50 %
Polluted	3	41	7	51	13.73 %
Column total	409	203	18	630	—
% column diagonal	77.02 %	49.26 %	38.89 %	—	66.98 %

Table 5. Crosstabulation under normalisation of absolute component loadings. Overall similarity = 86.51 %.

PCA-WQI \ DOE-WQI	Clean	Slightly Polluted	Polluted	Row total	% diagonal
Clean	319	58	0	377	84.62 %
Slightly Polluted	4	196	2	202	97.03 %
Polluted	0	21	30	51	58.82 %
Column total	323	275	32	630	—
% column diagonal	98.76 %	71.27 %	93.75 %	—	86.51 %

3.2 Classification similarity across Models A–H

Table 6 reports per-class similarity (in %) for all sixteen hybrid specifications. Three patterns emerge. First, uniformly conservative (Model A) or uniformly optimistic (Model B) choices yield extreme class-wise imbalances: Model A recovers Polluted observations exceptionally well (84.70 % equal, 93.53 % unequal) but misses most Clean stations (41.10 %), while Model B inverts the pattern (80.43 % Clean, 0.97 % Polluted). Second, class-asymmetric specifications that particularly Models C, F and H which are achieve more balanced per-class similarity. Third, Model F under unequal weighting is the only specification to exceed 70 % similarity in every class simultaneously: 83.03 % Clean, 71.45 % Slightly Polluted, 88.03 % Polluted (mean 80.84 %).

Table 6. Per-class classification similarity (%) of Models A–H against DOE-WQI, under equal-weight (EW) and unequal-weight (UW) aggregation. Bold = best in column; underline denotes the only specification that exceeds 70 % in every class.

Class / Weight	A	B	C	D	E	F	G	H
Clean (EW)	41.10	80.43	70.49	59.48	43.88	76.91	74.16	63.61
Slightly Polluted (EW)	70.59	48.92	78.77	75.47	72.88	83.64	58.68	55.95
Polluted (EW)	84.70	0.97	82.20	77.67	3.56	74.11	1.94	4.21
Clean (UW)	42.33	84.71	76.76	61.66	45.53	83.03	78.44	64.70
Slightly Polluted (UW)	59.54	47.35	66.71	65.42	58.54	71.45	52.08	54.95
Polluted (UW)	93.53	2.27	92.88	89.32	5.18	88.03	3.88	3.88

3.3 Confusion-matrix diagnostics

Table 7 summarises the per-class precision, recall, F1, accuracy and specificity for the Clean, Slightly Polluted and Polluted bands under unequal weighting. Model F attains the highest F1 in two of the three classes (Clean 0.871, Slightly Polluted 0.745) and the highest accuracy in Clean (0.903); Model A retains a slight edge in Polluted recall (0.935 vs 0.880). Model C achieves a marginally higher Clean F1 (0.852) but its Polluted recall collapses to 0.708, illustrating the usual sensitivity–specificity trade-off that Model F avoids. Across macro-averages (Table 8), Model F dominates on macro-F1 (0.785) and macro-accuracy (0.860), with the second-best Model C trailing by 0.013 and 0.018 respectively.

Table 7. Confusion-matrix metrics by class, unequal weighting.

Metric	Class	A	B	C	D	E	F	G	H
Precision	Clean	0.953	0.552	0.956	0.914	0.501	0.917	0.571	0.534
Recall	Clean	0.423	0.847	0.767	0.616	0.455	0.830	0.784	0.647
F1	Clean	0.586	0.669	0.852	0.736	0.477	0.871	0.661	0.585
Accuracy	Clean	0.770	0.669	0.894	0.830	0.617	0.903	0.683	0.648
Specificity	Clean	0.987	0.554	0.977	0.964	0.718	0.951	0.617	0.649
Precision	Slightly Polluted	0.535	0.519	0.733	0.644	0.425	0.779	0.511	0.471
Recall	Slightly Polluted	0.595	0.473	0.667	0.654	0.585	0.714	0.520	0.549
F1	Slightly Polluted	0.563	0.495	0.698	0.649	0.493	0.745	0.516	0.507
Accuracy	Slightly Polluted	0.606	0.595	0.758	0.698	0.485	0.795	0.589	0.544
Specificity	Slightly Polluted	0.615	0.683	0.824	0.731	0.411	0.853	0.639	0.541
Precision	Polluted	0.499	0.304	0.572	0.548	0.150	0.634	0.230	0.000
Recall	Polluted	0.935	0.022	0.928	0.893	0.051	0.880	0.038	0.038
F1	Polluted	0.650	0.042	0.708	0.679	0.077	0.737	0.066	0.000
Accuracy	Polluted	0.810	0.808	0.857	0.840	0.765	0.883	0.796	0.787
Specificity	Polluted	0.780	0.988	0.841	0.828	0.931	0.883	0.970	0.962

Table 8. Macro-average confusion-matrix metrics across all three classes, unequal weighting. Model F dominates on macro-F1 and macro-accuracy.

Model	Macro Precision	Macro Recall	Macro F1	Macro Accuracy	Macro Specificity
A	0.662	0.651	0.600	0.729	0.794
B	0.458	0.447	0.402	0.691	0.742
C	0.754	0.787	0.753	0.836	0.881
D	0.702	0.721	0.688	0.789	0.841
E	0.359	0.364	0.349	0.622	0.687
F	0.777	0.808	0.784	0.860	0.896
G	0.437	0.447	0.414	0.689	0.742
H	0.335	0.411	0.364	0.660	0.717

The bootstrap analysis (1 000 resamples) confirms the ranking. The bootstrap standard error of Model F's macro-F1 is 0.011, compared with 0.014 for Model C and 0.024 for Model A; no model outperforms Model F on macro-F1 in more than 7 % of bootstrap replicates. Model F's dominance is therefore not an artefact of the single 2017–2021 sample.

4. Discussion

4.1 Why Model F is the most robust specification

Model F corresponds to the operator triplet (min, max, max) across (Clean, Slightly Polluted, Polluted). Its dominance is not coincidental; it is a direct consequence of how min and max interact with the eclipsing and ambiguity pathologies identified in Section 1.

In the Clean band, choosing the minimum over the sub-indices that score 81 or higher forces the class representative C_1 to equal the lowest-performing sub-index within the Clean set. Eclipsing that the masking of a weak sub-index by stronger ones through averaging is therefore structurally eliminated at the upper boundary: a station is allowed to retain a Clean label only if every parameter in

the Clean set is strong. This is the correct regulatory posture for a Clean reservation, where the cost of false reassurance is typically high.

In the Slightly Polluted and Polluted bands, choosing the maximum prevents a single deeply degraded sub-index from dominating the aggregate. This is the structural attenuation of ambiguity: when one parameter takes an extreme low value (for example a BOD spike following an industrial discharge) but the remaining in-band parameters are closer to their upper threshold, C_2 and C_3 remain informative summaries rather than collapsing onto the outlier. Regulatory resolution in the mid- and low-quality ranges is therefore preserved.

The combination (min, max, max) thus exploits the information asymmetry across the three bands: what matters in the Clean band is the weakest link, whereas in the degraded bands what matters is whether any recovery signal persists. No uniform operator, neither Model A (min, min, min) nor Model B (max, max, max) that can achieve both simultaneously, and this is visible empirically in Table 6: A wins Polluted but collapses on Clean, B does the reverse, whereas F is the unique specification that exceeds 70 % in all three bands.

4.2 Trade-offs between equal- and unequal-weight schemes

Unequal weighting improves all performance metrics for Model F except Slightly Polluted similarity, which falls from 83.64 % under EW to 71.45 % under UW. This is the expected behaviour when the prior weights down-weight the middle band relative to the extremes, reflecting the epidemiological priority of distinguishing genuinely clean waters from genuinely polluted ones. Models B and G, which are already permissive at both extremes, suffer the largest losses under unequal weighting; Models A and F, which have asymmetric operator triplets, gain most. This reinforces the conclusion that operator-choice and weight-choice interact non-additively and must be jointly optimised, not tuned separately.

4.3 Comparison against PCA-WQI specifications

Although absolute-loading PCA-WQI (Table 5) achieves the highest overall similarity (86.51 %), its per-class profile is markedly uneven, polluted similarity is only 58.82 %. Model F under unequal weighting does not match PCA-WQI's total similarity but delivers far superior class balance (all three bands above 70 %) and retains an interpretable closed-form aggregation, whereas PCA loadings are dataset-dependent and lack the transparent regulatory interpretation that a DOE-WQI replacement requires.

4.4 Environmental implications

The asymmetric robustness of Model F has direct environmental meaning. Clean-band minima protect against false reassurance in pristine sub-basins, which are disproportionately represented in protected headwater catchments; degraded-band maxima avoid over-penalising stations undergoing remediation, where a single persistent out-of-range parameter (often ammoniacal nitrogen from upstream agriculture) can dominate a weighted arithmetic mean long after a remediation programme has begun to improve the other five parameters. In practical terms, Model F is expected to (i) raise earlier alarms in Clean sub-basins exhibiting degradation in a single parameter, and (ii) accelerate the reclassification of remediated waters out of the Polluted band once a majority of parameters have recovered — both of which are prerequisite behaviours for adaptive river-basin management.

4.5 Limitations

Three limitations are acknowledged. First, the 2017–2021 DOE dataset covers a relatively short period; extending the analysis to monsoonal regime shifts and land-use change cycles would strengthen external validity. Second, only min and max were tested within the generalised power-mean framework; intermediate values of r (e.g., geometric and harmonic forms) merit further evaluation but were out of scope here, as the binary min/max design suffices to establish the proof of concept. Third, the ground-truth benchmark is the DOE-WQI itself; as the DOE-WQI is known to suffer from the very pathologies this paper addresses, agreement with it is a necessary but not sufficient criterion. Future work should validate against independent ecological indicators (e.g., macroinvertebrate indices).

5. Conclusion

This study set out to address two structural defects of the Malaysian DOE-WQI that have the highlight the problem of eclipsing effect and ambiguity by developing a class-adaptive Hybrid Mean Function (HMF) that applies different aggregation operators to sub-indices depending on their DOE band membership. Within the unified framework of the generalised power mean, eight operator triplets (Models A–H) were defined as a full 2^3 factorial of min/max choices, evaluated under both equal- and unequal-weight schemes, and tested against 630 monitoring observations (2017–2021) from Malaysian rivers. The combined evidence from classification similarity, confusion-matrix diagnostics, PCA sensitivity analysis and bootstrap resampling identifies Model F under unequal weighting operator triplet (min, max, max) as uniquely robust: it exceeds 70 % per-class similarity in all three bands, attains macro-F1 of 0.785 and macro-accuracy of 0.860, and does so with an interpretable closed-form aggregation that is implementable within the existing DOE-WQI computational pipeline. Model F simultaneously attenuates eclipsing in the Clean band through the minimum operator and ambiguity in the degraded bands through the maximum operator as a trade-off that no uniform aggregator can resolve.

6. Policy Recommendations

Four actionable recommendations follow for the Department of Environment Malaysia and its river-basin authorities.

4. **Adopt Model F as a parallel reporting standard.** The DOE should publish river status both under the current DOE-WQI and under Model F (unequal weighting) for a three-year transition period, enabling direct comparison of reclassification outcomes at catchment level and informing a permanent methodology revision through the Environmental Quality (Surface Water) Regulations.
5. **Prioritise sub-basins flagged divergently by the two methods.** Sub-basins for which Model F downgrades the DOE-WQI classification should enter a watch-list for intensified monitoring; those upgraded under Model F should be audited for remediation success and considered for early de-escalation of regulatory action, reducing compliance burden without relaxing ecological protection.
6. **Integrate the HMF into the National Water Quality Information System (SPMAA).** The closed-form expressions in equations (2)–(9) are inexpensive to evaluate, can be deployed within the existing DOE data pipeline without additional sensor infrastructure, and produce the same categorical output that making the transition computationally and operationally costless.
7. **Couple the HMF with real-time alerting for Clean-band sub-basins.** Because Model F's anti-eclipsing behaviour in the Clean band triggers reclassification when any single parameter

weakens, it is particularly well suited to early-warning dashboards for protected headwaters and drinking-water source catchments, where the cost of delayed detection is high.

These recommendations do not require new primary data collection; they require only a revision of the aggregation function operating on data the DOE already collects. The marginal cost of adoption is therefore low, while the expected gain in regulatory resolution, particularly for sub-basins at the Clean–Slightly Polluted boundary is substantial.

Author Contributions

N.H.N.M.S.: Conceptualisation, Methodology, Formal analysis, Software, Visualization, Writing original draft. Z.M.A.: Conceptualisation, Methodology, Supervision, Project administration, Writing, review & editing. Both authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The river water-quality data used in this study were obtained from the Department of Environment, Ministry of Natural Resources and Environmental Sustainability, Malaysia. Data are available from the corresponding author upon reasonable request, subject to DOE redistribution policy.

Conflict of Interest

The authors declare no conflicts of interest.

Clinical Trial Number

Not applicable

Ethics, Consent to Participate, and Consent to Publish declarations

Not applicable

Declaration

During the preparation of this work, the author used Google AI Studio, Gemini Pro and Claude to brainstorm the idea, and scite.ai in order to extract key claims and validate the referencing, respectively. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Funding

This research received no specific external funding. N.H.N.M.S. acknowledges a doctoral study leave from Universiti Teknologi MARA.

Acknowledgments

The authors thank the Department of Environment Malaysia for provision of the monitoring dataset, and the anonymous reviewers whose comments shaped this revised manuscript. And also to Faculty of Computer and Mathematical Science, UiTM and Department of Science and Technology, UKM who have give the opportunity to write and publish the paper.

References

Abbasi, T., & Abbasi, S. A. (2012). Water-quality indices. In Water Quality Indices (pp. 353–356). Elsevier. https://doi.org/10.1016/B978-0-444-54304-2.00016-6

Al-Badaii, F., Shuhaimi-Othman, M., & Gasim, M. B. (2013). Water quality assessment of the Semenyih River, Selangor, Malaysia. Journal of Chemistry, 2013, 871056. https://doi.org/10.1155/2013/871056

Aris, A. Z., Lim, W. Y., Praveena, S. M., Yusoff, M. K., Ramli, M. F., & Juahir, H. (2014). Water quality status of selected rivers in Kota Marudu, Sabah, Malaysia and its suitability for usage. Sains Malaysiana, 43(3), 377–388.

Banda, T. D., & Kumarasamy, M. V. (2020). Development of water quality indices (WQIs): A review. Polish Journal of Environmental Studies, 29(3), 2011–2021. https://doi.org/10.15244/pjoes/110526

Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Science of the Total Environment, 721, 137612. https://doi.org/10.1016/j.scitotenv.2020.137612

Bullen, P. S. (2003). Handbook of Means and Their Inequalities (2nd ed.). Springer. https://doi.org/10.1007/978-94-017-0399-4

Deininger, R. A. (2009). A water quality index for rivers. Water International, 5(3), 16–21. https://doi.org/10.1080/02508068008685872

Department of Environment Malaysia. (2019). Malaysia Environmental Quality Report 2019. Ministry of Environment and Water, Putrajaya.

Ghazali, N., & Ali, Z. M. (2023). Principal component analysis approach in Klang River water quality index modelling. Journal of Physics: Conference Series.

Landwehr, J. M., & Deininger, R. A. (1974). An objective water quality index. Journal (Water Pollution Control Federation), 46(7), 1804–1809.

Lohani, B. N., & Mustapha, N. (1982). Stochastic water quality index. Environmental Technology Letters, 3(11), 521–528. https://doi.org/10.1080/09593338209384157

Misaghi, F., Delgosha, F., Razzaghmanesh, M., & Myers, B. (2017). Introducing a water quality index for assessing water for irrigation purposes: A case study of the Ghezel Ozan River. Science of the Total Environment, 589, 107–116. https://doi.org/10.1016/j.scitotenv.2017.02.226

Mohd Ali, Z., Ibrahim, N. A., Mengersen, K., Shitan, M., & Juahir, H. (2013). The Langat River water quality index based on principal component analysis. AIP Conference Proceedings, 1522, 1322–1329. https://doi.org/10.1063/1.4801283

Pesarin, F. (1971). Assessment of surface water quality by a single index of pollution. Water Research, 5, 741–751.

Smith, D. G. (1990). A better water quality indexing system for rivers and streams. Water Research, 24(10), 1237–1244. https://doi.org/10.1016/0043-1354(90)90047-A

Sutadian, A. D., Muttil, N., Yilmaz, A. G., & Perera, B. J. C. (2016). Development of river water quality indices — A review. Environmental Monitoring and Assessment, 188, 58. https://doi.org/10.1007/s10661-015-5050-0

Uddin, M. G., Nash, S., Mahammad Diganta, M. T., Rahman, A., & Olbert, A. I. (2022). Robust machine-learning algorithms for predicting coastal water quality index. Journal of Environmental Management, 321, 115923. https://doi.org/10.1016/j.jenvman.2022.115923

Uddin, M. G., Nash, S., & Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. Ecological Indicators, 122, 107218. https://doi.org/10.1016/j.ecolind.2020.107218

Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2023). A sophisticated model for rating water quality. Science of the Total Environment, 868, 161614. https://doi.org/10.1016/j.scitotenv.2023.161614

Zainuddin, Z., Jamal, P., & Akbar, I. (2012). Modeling the effect of dam construction and operation towards downstream water quality of Sg. Tawau and Batang Baleh. World Journal of Applied Environmental Chemistry, 1, 57–66.