EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology, 2026, 27(4), 176–190
https://doi.org/10.12912/27197050/218862
ISSN 2719–7050, License CC-BY 4.0

Received: 2026.02.12
Accepted: 2026.03.16
Published: 2026.04.01

Machine-learning-assisted multi-objective environmental modelling of trace metal and mineral pollution in drinking water: A case study from Kénitra, Morocco

Latifa Ben Akka[1], Oussama Gliti[2], Zohra Bejjaji[1], Basma Naoui[1], Asmae Titafi[1], Aziz Taouraout[3], Mustapha Mouilly[4], Sakina Mehdioui[5*], Mahjoub Aouane[6]

[1] Laboratory Geosciences and Natural Ressources, Faculty of Sciences Ibn Tofail University Kenitra Morocco
[2] Laboratory of Electronic Systems, Information Processing, Mechanics and Energetics, Faculty of Sciences Ibn Tofail, Kenitra, Morocco
[3] Management and development of natural resources, Faculty of Sciences Meknès University Moulay Ismael
[4] Department of Biology, Faculty of Sciences Meknès University Moulay Ismael and Faculty of Sciences and Technologies of Errachidia, Morocco
[5] Laboratory Geosciences and Natural Ressources, Higher School of Education and Training Ibn Tofail University Kenitra Morocco
[6] Laboratory of Natural Resources and Sustainable Development, Faculty of Sciences, IbnTofail University, Kenitra, Morocco
* Corresponding author's e-mail: mehdiouisakina@gmail.com

ABSTRACT

Urban drinking-water systems increasingly face a dual challenge: trace metal contamination and disturbed mineral balance, yet many utilities still rely on descriptive monitoring rather than optimisation-driven management. Although machine learning and multi-objective evolutionary optimisation are widely applied in environmental modelling, their end-to-end integration under small-sample monitoring constraints remains under-demonstrated for actionable utility decision support. This study develops a surrogate-assisted multi-objective optimisation framework that transforms routine laboratory measurements into implementable management strategies for urban drinking-water quality. Fourteen household taps across seven distribution zones in Kénitra (Morocco) were analysed for health-relevant trace elements and macro-minerals. Gradient-boosted tree models (XGBoost) were trained under leave-one-out cross-validation to quantify predictive skill under small-sample conditions. Predictive performance was element-dependent, with $R^2 \approx 0.70$ for Ni, 0.53 for P, 0.29 for Cr, and 0.08 for Ag, consistent with stronger signal for Ni/P and attenuated learnability for near-detection and highly variable trace elements. The trained surrogates were then coupled to a four-objective NSGA-III optimisation to simultaneously reduce regulatory exceedance (sanitary risk), compress inter-zone disparities (homogeneity), improve Ca/Mg/Na/K mineral balance, and constrain intervention effort under contrasting sanitary-priority and mineral-priority profiles. The resulting Pareto fronts reveal a narrow compromise region in which sanitary risk and mineral imbalance are jointly suppressed with marginal increases in operational effort. From this region complemented by extreme non-dominated points seven operator-facing scenarios were derived, linking explicit reduction fractions and mineral adjustments to predicted system-wide outcomes (e.g., exceedance objective as low as 0.0023, inter-zone variance down to ~0.0000–0.0004, mineral deviation as low as 10.5075, and effort proxy as low as 2.4162, in the reported objective units). By demonstrating robust optimisation under small-sample conditions typical of municipal monitoring programmes, this study provides a transferable modelling architecture for data-limited urban utilities and strengthens the integration of machine learning with environmental decision-making.

Keywords: drinking-water quality, trace metals, mineral balance, XGBoost, NSGA-III, Pareto optimization, Morocco, decision support.

INTRODUCTION

Urban drinking-water systems must simultaneously manage toxicological risks associated with trace metals (e.g., Ni, Cr, Ag, Zn) and operational variability arising from distribution-network hydraulics, corrosion scales, and biofilms, all of which can compromise water quality between treatment works and the consumer's tap (WHO, 2017; IARC, 2012; Broo et al., 2001; Makris et al., 2014). Transient low-pressure events and flow reversals can mobilise deposits and resuspend accumulated materials, while aggressive corrosion-control interventions may perturb mineral balance and alter palatability. Consequently, utilities often need to arbitrate among regulatory compliance, customer acceptability, and operational effort within constrained monitoring and intervention capacities (WHO, 2017; Makris et al., 2014).

In Morocco, hydrochemical heterogeneity and coastal influences have been documented in the Rharb–Mnasra system, alongside mixed geogenic and anthropogenic pressures and seasonal variability in trace metals and bacteriological indicators (Aguedai et al., 2022; Benyoussef et al., 2021; Hicham et al., 2022). While descriptive mapping and water quality index (WQI) studies provide useful surveillance insights, they rarely convert monitoring outputs into operational control strategies that explicitly quantify trade-offs among sanitary protection, spatial homogeneity, mineral balance, and intervention cost.

Despite extensive monitoring programs and a mature regulatory discourse, the literature seldom treats inter-site homogeneity as an explicit optimisation objective, and rigorous Pareto-front assessment metrics (e.g., hypervolume, IGD, spacing, coverage, ε-additive) are infrequently reported beyond visual inspection (Zitzler et al., 2003; Emmerich and Deutz, 2018). Decision-oriented frameworks that jointly optimise sanitary exceedance, inter-site uniformity, mineral composition, and operational effort while remaining reproducible and computationally lightweight for utilities remain scarce in the Moroccan context. In parallel, recent work increasingly steers water-sector studies toward AI-integrated approaches, reflecting a broader movement toward data-guided optimisation of small- to medium-scale environmental systems (Gliti et al., 2023; Gliti et al., 2024). More broadly, prediction–optimisation workflows are now well established across engineering domains, including imaging and energy systems (Abbi et al., 2025).

The aim of this study is to develop a reproducible, standards-aware machine-learning and multi-objective optimisation framework that advances both methodological and practical decision support for urban drinking-water management in Kénitra and similar settings. We hypothesise that: (i) surrogate models trained on routine monitoring data can predict post-treatment concentrations of trace metals and minerals even under small-sample conditions; (ii) multi-objective optimisation can identify Pareto-efficient trade-offs that simultaneously reduce sanitary exceedances, improve inter-site homogeneity, and maintain mineral balance within operationally feasible limits; (iii) such a workflow can yield actionable decision portfolios that are transparent, traceable, and suitable for utilities with limited computational resources. By establishing this framework, the study moves beyond descriptive monitoring and WQI assessments toward a decision-oriented platform capable of balancing compliance, equity across sites, mineral quality, and operational cost. The following sections detail the data, model development, and optimisation methodology employed to test these hypotheses.

MATERIALS AND METHODS

Study area, water supply, and sampling design

Kenitra is embedded in the Gharb–Maâmora hydrogeological system (~6.000 km²), where groundwater generally drains northwest toward the Atlantic. Urban supply draws predominantly from the Maâmora aquifer (≈75%) and, secondarily, from the Gharb aquifer (≈25%).

Within the municipal distribution network, seven consumer zones were selected for household-tap sampling: Moulay Bousalham, Mehdia, Al Wafaa, Saknia, Ancienne Médina, Centre-Ville, and Ouled Oujih. Each zone contributed two treated tap-water samples during January–February 2022 (total n = 14). The analyte panel comprised Ag, B, Ca, Cr, Cu, K, Mg, Na, Ni, P (and Zn where reported), spanning health-relevant trace metals and taste/corrosion-relevant macro-minerals (WHO, 2017; IARC, 2012). Spatial context is shown in Figure 1.

Please send a good quality drawings by e-mail

Figure 1. Spatial context and household-tap sampling footprints across Kenitra's distribution zones

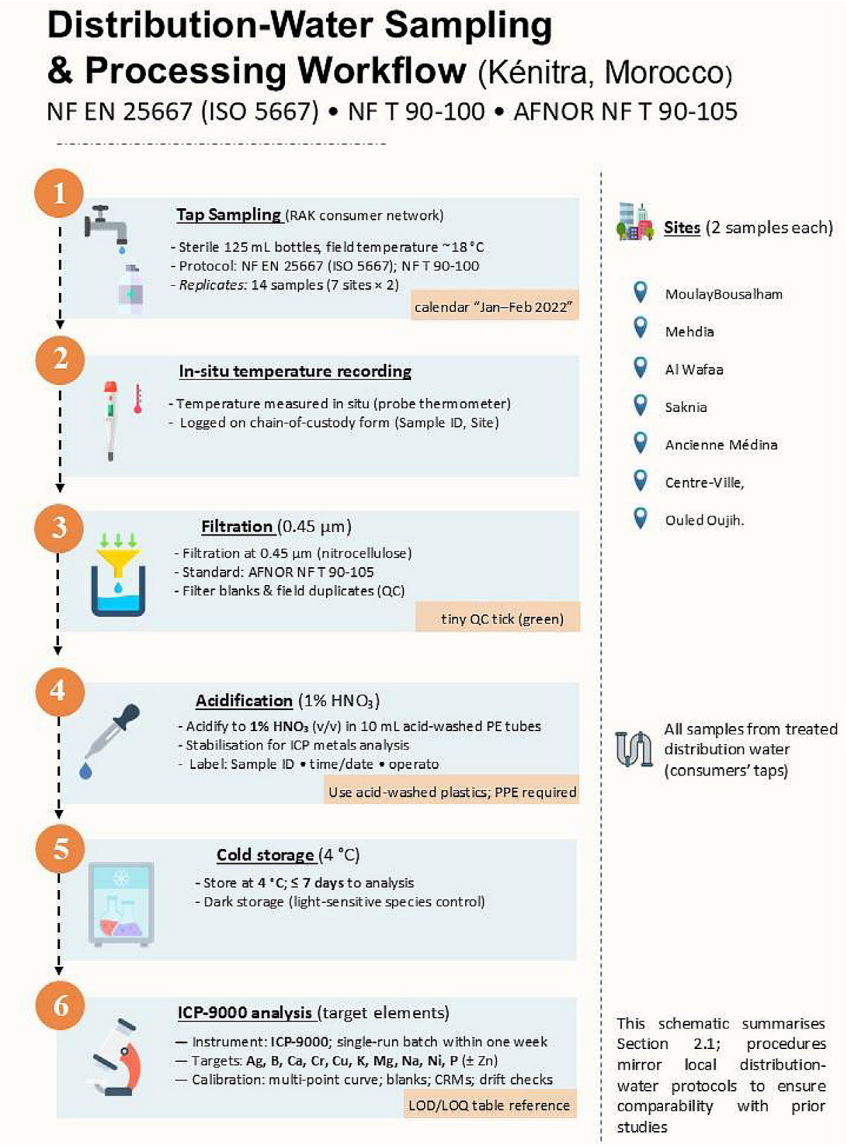

Figure 2. *Distribution-water sampling and processing workflow (Kenitra).* Tap samples (125 mL, ~18 °C) were collected to NF EN 25667 (ISO 5667) and NF T 90-100, filtered at 0.45 µm (AFNOR NF T 90-105), acidified to 1% HNO_3, stored at 4 °C (≤ 7 days), and analyzed on ICP-9000 for Ag, B, Ca, Cr, Cu, K, Mg, Na, Ni, P (± Zn). All samples originated from treated distribution networks; two replicates per site across seven sites (Jan–Feb 2022)

Sample collection and laboratory analysis

Tap samples were collected by the local water utility (RAK) in 125 mL sterile bottles in accordance with NF EN 25667 (ISO 5667) and NF T 90-100. Temperature was recorded in situ. Waters were filtered at 0.45 µm (AFNOR NF T 90-105), acidified to 1% HNO_3, stored at 4 °C, and analysed within one week using ICP-9000 for the target elements (Figure 2).

Supplementary Table S1 show experimental tap-water concentrations (mg/L) measured across seven distribution zones in Kénitra (n = 14; Jan–Feb 2022). Values reported as "<DL" indicate concentrations below the analytical detection/quantification threshold. This table is the primary source for baseline visualisations (Figures 3–5) and for constructing the modelling dataset used in the ML–NSGA-III workflow.

Data restructuring and preprocessing

The raw laboratory export was organised as repeating non-uniform blocks ("Element name → Unit → Average") indexed by a numeric Sample ID, which is not directly suitable for modelling. The worksheet was converted into a strict samples × variables table using the following protocol: (i) each "Unit" row was mapped to its preceding "Element name" header, and the

subsequent "Average" row was read as numeric values associated with the nearest Sample ID; (ii) unit strings (mg/L, µg/L) were harmonised and concentrations were expressed in mg L^{-1} (µg L^{-1} converted accordingly); (iii) numeric coercion was enforced and non-parsable cells were retained as NA; (iv) duplicated element labels were consolidated; (v) given the balanced design and high completeness, no imputation was performed; and (vi) implausible entries were flagged using robust median/MAD checks and unit verification (Helsel, 2020), without removing samples solely based on statistical flags. Each sample ID was mapped to one of the seven zones (two samples per zone). The full experimental tap-water concentrations (per-sample results) are provided in Table S1, and Figures 3–5.

Feature set and site encoding

To represent spatial/infrastructure heterogeneity without imposing artificial ordinality, the categorical zone variable ("Site") was encoded using one-hot vectors (seven dummy variables), allowing the models to learn zone-specific offsets and interactions (James et al., 2021). The predictor set consisted of all available measured analytes plus the one-hot Site factor; targets were selected at-risk elements $e \in \{Ni,Cr,Ag,Zn,P\}$ where available.

Surrogate modelling (purpose, specification, validation)

The machine-learning component is used as a surrogate to approximate the mapping between measured water-quality descriptors (including Site) and concentrations of selected at-risk elements, enabling rapid evaluation within the optimisation loop. For each target element, an independent XGBoost regressor (XGBRegressor) was trained because gradient-boosted trees capture non-linearities and feature interactions, tolerate heterogeneous predictors, and remain robust under modest sample sizes typical of monitoring campaigns (Chen and Guestrin, 2016). Hyperparameters were restricted to compact values to reduce over-parameterisation under small-sample conditions: n_estimators=300–400,max_depth=3–4, learning_rate=0.07–0.08, subsample=0.9, colsample_bytree=0.8, with a fixed random seed. Given n=14, leave-one-out cross-validation (LOOCV) was used to estimate out-of-sample performance. Predictive skill was reported using on held-out predictions; RMSE was also reported to provide error scale in concentration units. After validation, each target model was refit on the full dataset using fixed hyperparameters for deterministic surrogate use during multi-objective optimization (Table 1).

Domain of applicability (no extrapolation)

The workflow is constrained to the empirical domain represented by the monitoring campaign. Optimisation is evaluated on the monitored zones/samples and bounded intervention variables; no extrapolation is claimed to unobserved zones, substantially different networks, or concentration regimes outside the observed range without new data.

Decision variables and intervention parameterisation

We encode interventions by two continuous classes:

1. For each $e \in \{Ni, Cr, Ag, Zn, P\}$, a fractional reduction $r_e = [0,0.8]$ scales the predicted concentration,

$$C'_e = (1 - r_e)\,\hat{C}_e \quad (1)$$

where: $r_e = 0.8$ representing an 80 % decrease relative to the modelled baseline.

Table 1. Hyperparameter ranges for XGBRegressor models

Hyperparameter	Value range	Description
n_estimators	300–400	Number of boosting trees
max_depth	3–4	Maximum tree depth
learning_rate	0.07–0.08	Shrinkage per tree
subsample	0.90	Row subsampling
colsample_bytree	0.80	Feature subsampling
random_state	42	Reproducibility seed

2. For each essential mineral $m \in$ {Ca, Mg, Na,K}, a relative adjustment $\Delta_m \in [-0.2,0.2]$ modifies the measured level

$$C'_m = (1 + \Delta_m)\, C_m \quad (2)$$

The decision vector

$$\mathbf{X} = (r_{Ni}, r_{Cr}, r_{Ag}, r_{Zn}, r_P, \Delta_{Ca}, \Delta_{Mg}, \Delta_{Na}, \Delta_K)$$

Defines the search space explored by the optimizer.

Objectives and formal problem statement (all minimised)

Let L_e denote statutory limits (WHO, 2022): $L_{Ni} = 0.02$ L, $L_{Cr} = 0.05$ L, $L_{Ag} = 0.10$ L, $L_{Zn} = 3.0$ L, $L_p = 5.0$ mg L^{-1}. Let the mineral targets be $T_{Ca} = 45$, $T_{Mg} = 8$, $T_{Na} = 20$, $T_K = 0.8$ mg L^{-1} (Cotruvo, 2017).

1) Sanitary compliance, f_1. Total exceedance across all samples and elements:

$$f_1(x) = \sum_{e \in E} \sum_i \max(0, C'_{e,i} - L_e), \quad E = \{Ni, Cr, Ag, Zn, P\} \quad (3)$$

2) Spatial homogeneity, f_1. Variance across sites of site-mean concentrations (operational equity):

$$f_2(x) = \sum_{e \in E} Var_s(\mu'_{s,e}), \mu'_{s,e} = \frac{1}{n_s} \sum_{i \in s} C'_{e,i} \quad (4)$$

3) Mineral balance, f_3. Sum of RMSE terms relative to palatability/benefit targets (Cotruvo, 2017):

$$f_3(x) = \sum_{m \in M} \frac{1}{n} \sum_i \left(C'_{m,i} - T_m\right)^2, \quad M = \{Ca, Mg, Na, K\} \quad (5)$$

4) Treatment-effort proxy, f_4. A convex surrogate for operational/economic burden:

$$f_4(x) = \sum_{e \in E} \omega_e |r_e| + \sum_{m \in M} \omega_m |\Delta_m| \quad (6)$$

with element-specific weights $\{\omega_e, \omega_m\}$ taken from the approved methodological note.

To avoid scale dominance, objective vectors were min/max normalized within each generation prior to survival selection. We solved two policy-relevant scenarios: Run-A (Sanitary Priority) softly emphasizes f_1 and Run-B (Mineral Priority) softly emphasizes f_3. This ML-NSGA-III pipeline is thus directly aligned with the management problem – selecting implementable combinations xthat balance compliance, spatial equity, taste/corrosion considerations, and effort.

NSGA-III configuration and workflow

We used NSGA-III with Das–Dennis reference directions (Deb and Jain, 2014; Das and Dennis, 1998), as implemented in pymoo (Blank and Deb, 2020). Configuration – population ≈ 165, 250 generations, simulated binary crossover ($\eta_c = 15$), polynomial mutation ($\eta_m = 20$), duplicate elimination enabled, and seed = 42. Two policy profiles were analysed – Run-A (sanitary priority) and Run-B (mineral priority). Outputs included Pareto-optimal decision vectors and objective values in raw and normalised units, along with diagnostic plots; scripts are provided as Supplementary code.

RESULTS

Baseline patterns across distribution zones

Baseline measurements (n = 14 household taps across seven distribution zones; two samples per zone) indicate measurable inter-zone heterogeneity despite sampling treated distribution water throughout the same municipal system. Figure 3 summarizes standardized contrasts (z-scores) across analytes and zones, showing structured spatial patterns for several macro-minerals (notably Ca, Mg, and Na), while some trace metals appear at low or near-detection levels in multiple zones.

Mineral composition relative to the palatability/corrosion-oriented targets used in the optimisation objective f_3 is shown in Figure 4. Across zones, deviations from the Ca/Mg/Na/K targets suggest that mineral balance is not uniformly achieved under the observed operating state. Figure 5 positions trace-metal concentrations (Ni, Cr, Ag, P; Zn where reported) against guideline values, providing a baseline view of compliance headroom and any localized risk signal prior to optimisation. Together, Figures 3–5 establish that the management problem extends beyond pointwise compliance to include spatial uniformity and mineral-balance control at the distribution level.

Surrogate model performance under small-sample validation

Predictive performance was assessed using leave-one-out cross-validation (LOOCV) to quantify generalisation under a small-sample monitoring regime. Performance varied across target elements (Table 2). Nickel exhibited the strongest explained variance (LOOCV $R^2 \approx 0.70$),

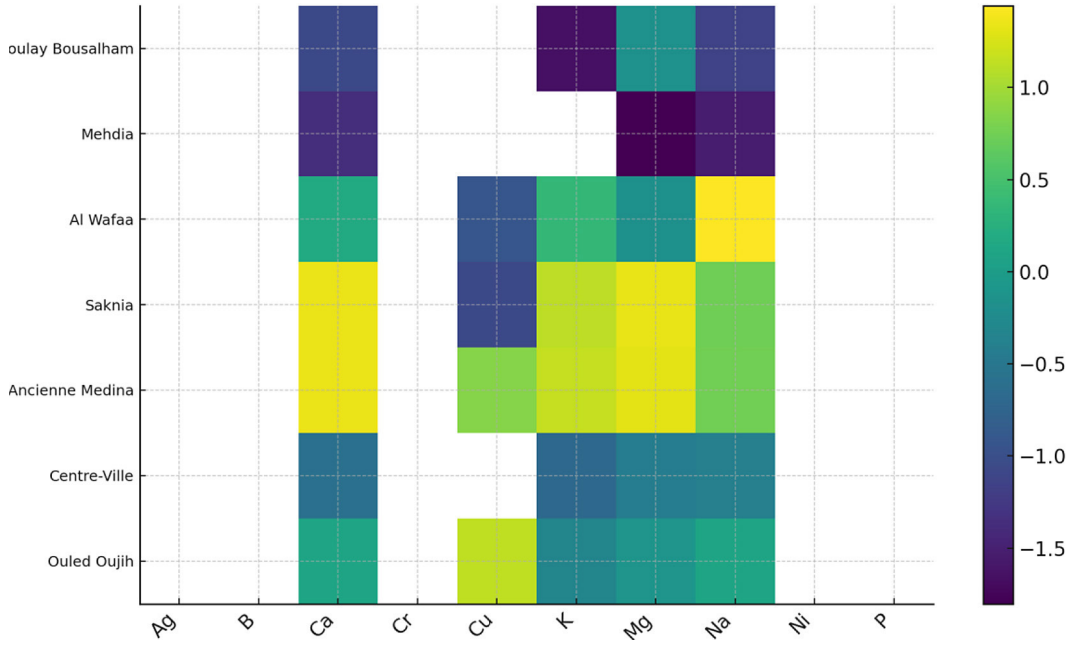

Figure 3. Exploratory heat-map: per-site z-scores by analyte

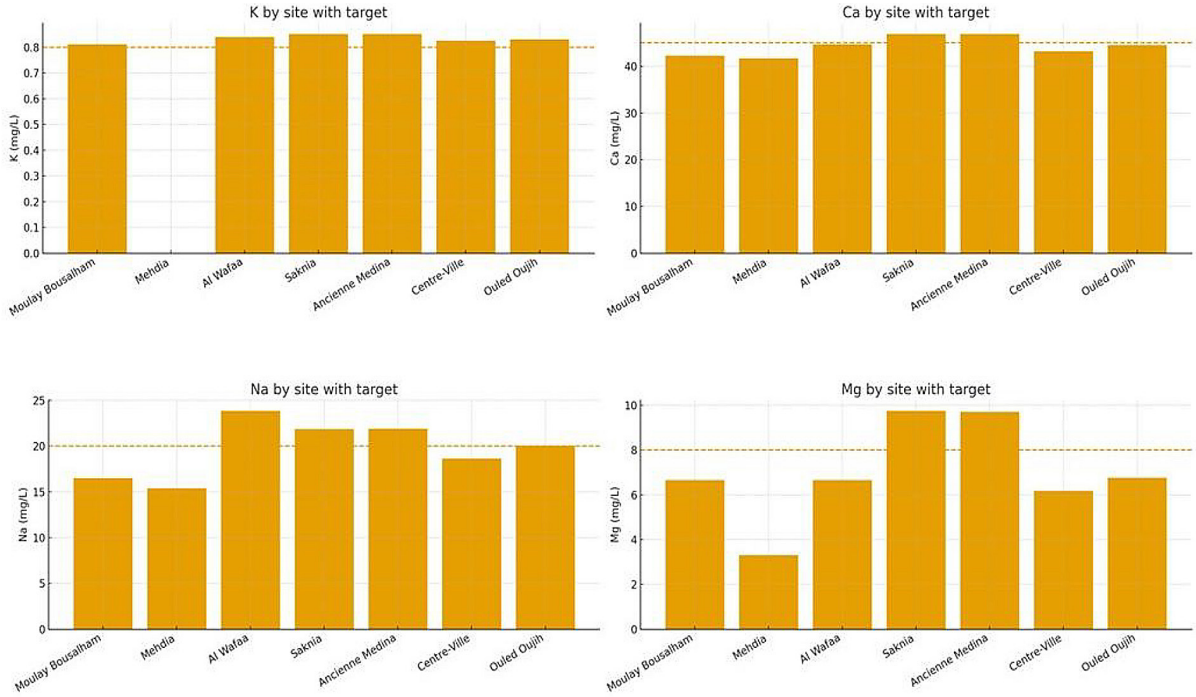

Figure 4. Minerals vs palatability/corrosion targets by site (Ca, Mg, Na, K)

Table 2. Cross-validated (LOO) R^2 by target element

Target element	LOOCV (R^2)	Note
Ni	0.70	Highest predictive skill
P	0.53	Moderate predictive skill
Cr	0.29	Limited predictive skill
Ag	0.08	Weak predictive skill
Zn	—	Modelled only when reported; otherwise excluded

followed by phosphorus ($R^2 \approx 0.53$). Chromium showed limited predictive strength ($R^2 \approx 0.29$), and silver performance was weak ($R^2 \approx 0.08$), consistent with lower learnability under trace-level variability and near-detection behavior in routine monitoring. Zinc was modelled only when reported; otherwise it was not included as a surrogate target. After validation, each target-specific model was refit on the full dataset with fixed hyperparameters to provide deterministic surrogate predictions for the optimisation stage within the observed operating domain.

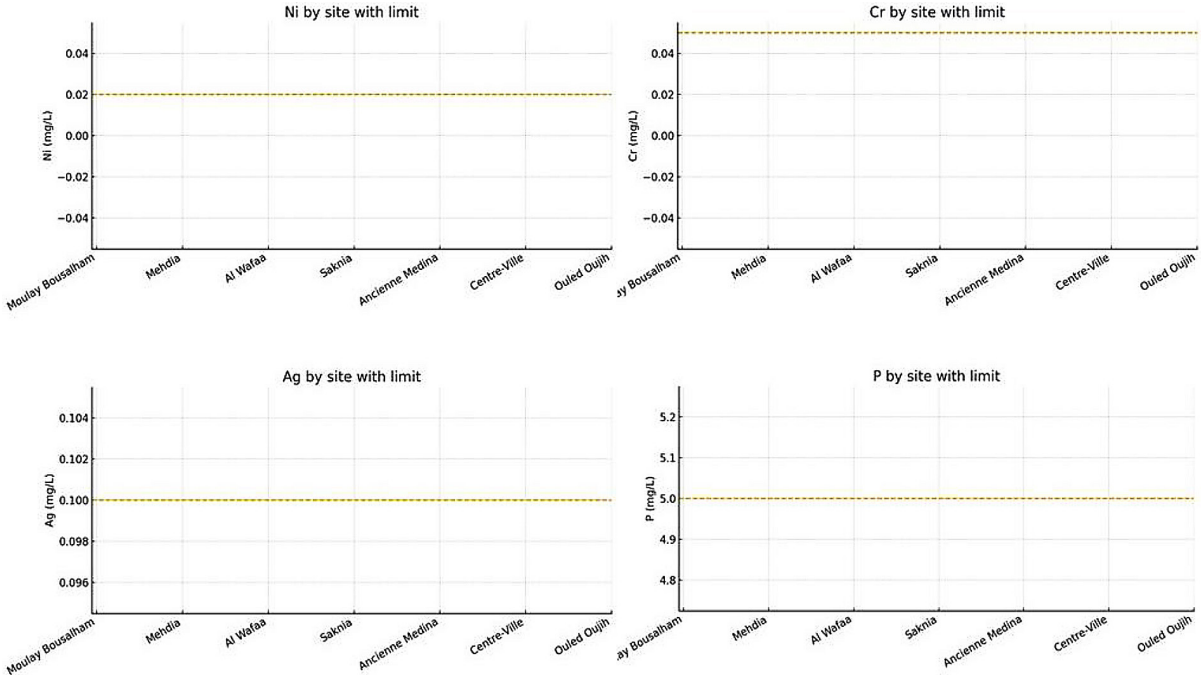

Figure 5. Trace metals vs guideline values by site (Ni, Cr, Ag, P; Zn where reported)

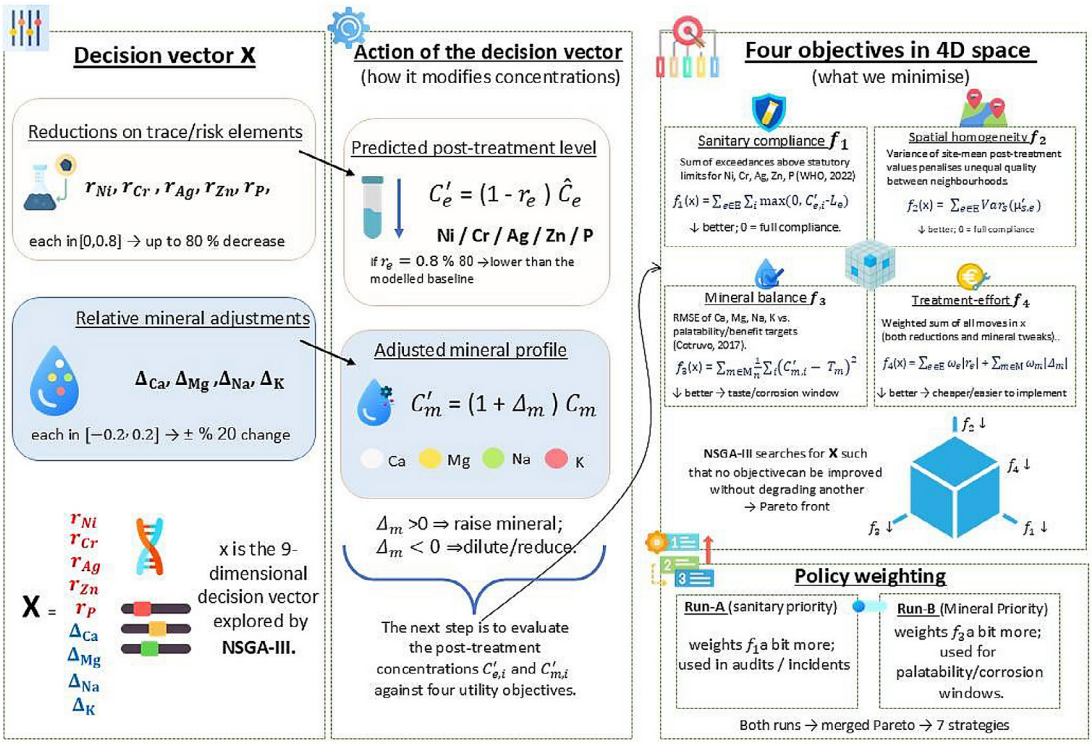

Figure 6. Linking the decision vector to the four-objective optimisation space (ML + NSGA-III workflow for Kenitra)

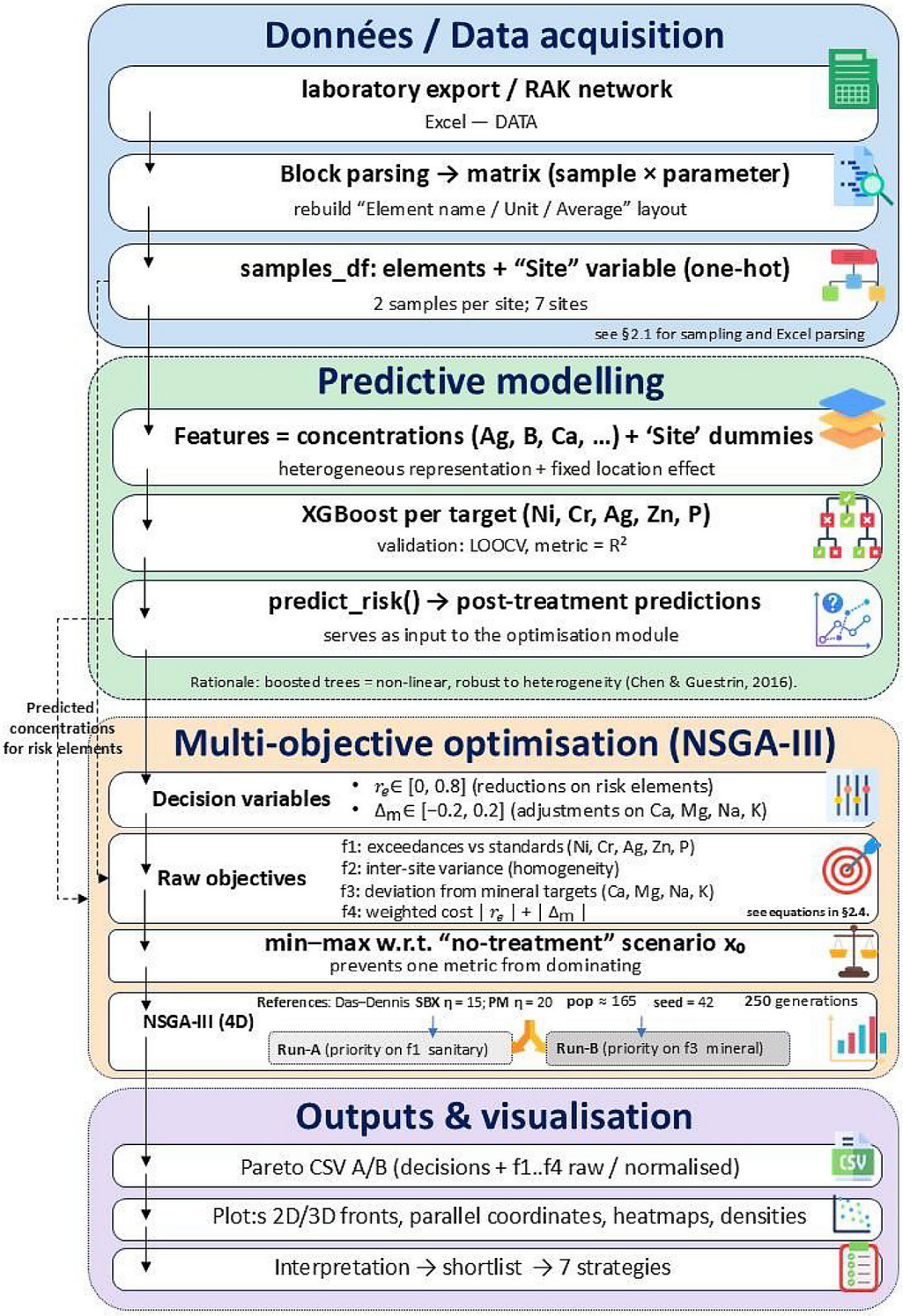

Figure 7. ML + NSGA-III worklow for portabme-water optimisation (Kénitra)

Pareto fronts and decision-relevant trade-offs

Front geometry under contrasting priority profiles

Both optimisation profiles produced dense non-dominated solution sets spanning the admissible region in the four-objective space (f1,f2,f3,f4) (f_1, f_2, f_3, f_4)(f1,f2,f3,f4). Figure 8 provides a two-panel visualisation that enables direct comparison between the sanitary-priority profile (Run-A) and the mineral-priority profile (Run-B). In Run-A, solutions populate a low-exceedance region with consistently low inter-site variance, while mineral deviation and effort vary along the front. In Run-B, solutions shift toward lower mineral

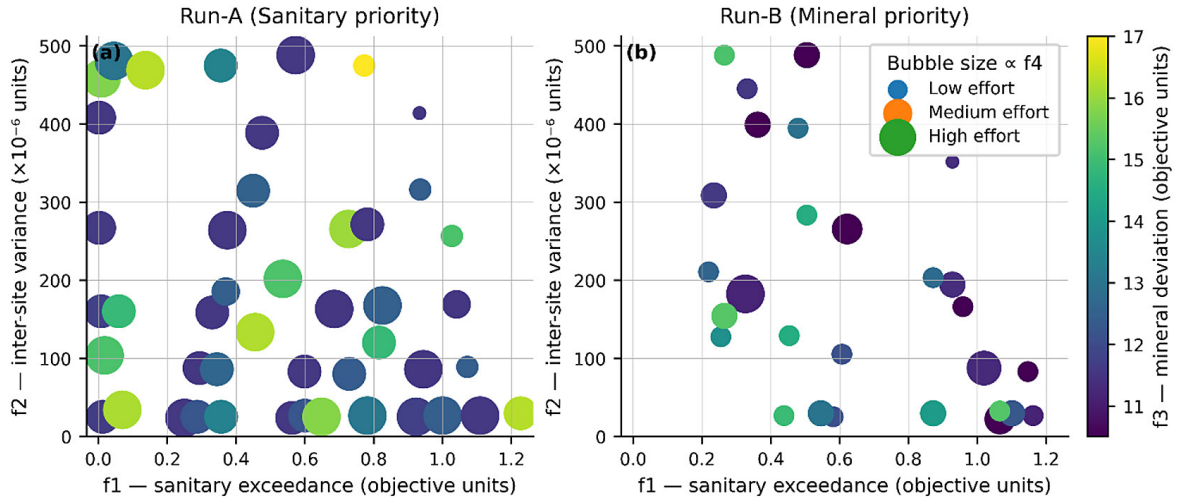

Figure 8. (a–b). 2-D bubble fronts (Run-A: x = f_1, y = f_2, colour = f_3, size = f_4; Run-B: x = f_3, y = f_2, colour = f_1, size = f_4)

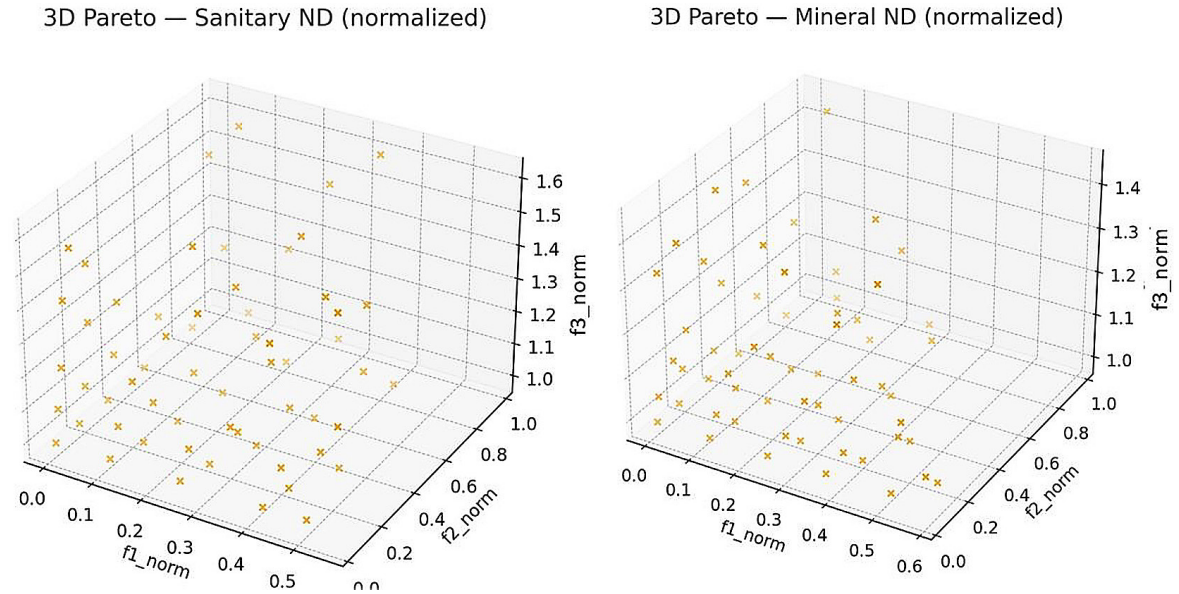

Figure 9. 3D Normalized 3D Pareto fronts (Run-A sanitary-priority vs Run-B mineral-priority)

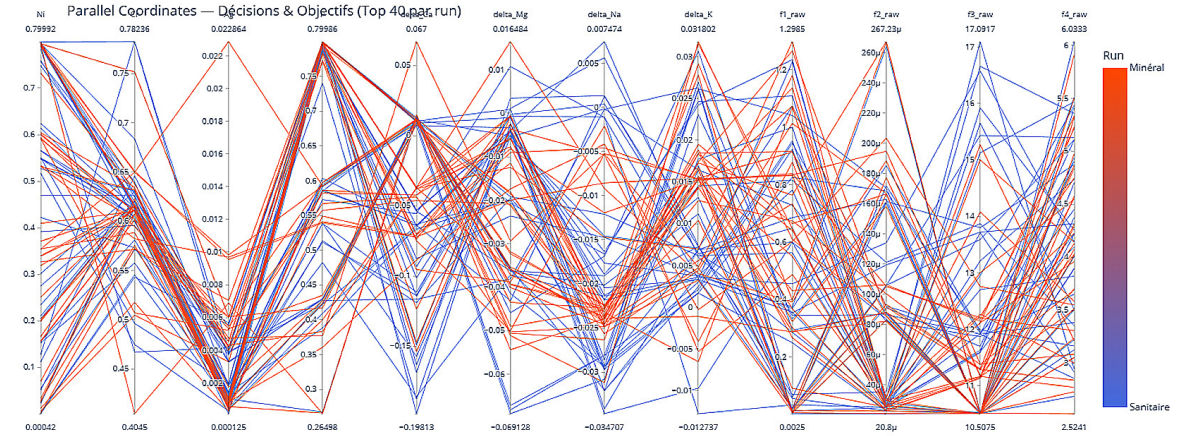

Figure 10. Parallel-coordinates view of decision vectors(r_e, Δ_m) and raw objectives($f_1 \ldots f_4$) (top non-dominated 40 per run)

deviation while maintaining similarly low inter-site variance; in this region, tighter mineral control may coincide with modest relaxation in sanitary exceedance unless additional effort is accepted.

The normalized 3D front visualisation (Figure 9) confirms that the compromise structure is stable in multi-dimensional objective space rather than a projection artifact. Across both runs, inter-site variance (f_2) remains uniformly small, while the sanitary–mineral trade-off (between f_1 1 and f_3) defines the dominant compromise corridor.

Decision-objective relationships and diagnostic views

Parallel-coordinates analysis (Figure 10) links intervention levers (r_e, Δm) to objective outcomes. Across the non-dominated sets, sanitary improvements (lower f1) are primarily associated with stronger reductions in Ni and Cr (r_{Ni}, r_{Cr}), while lower mineral deviation (f_3) is mainly driven by moderate mineral adjustments, particularly Δ_{Mg} and Δ_{Na}. The effort proxy f_4 increases monotonically with the overall magnitude of decision changes, consistent with its definition. The summarized parallel view (Figure 11) highlights systematic differences between profiles, with Run-A relying more strongly on risk-element reductions and Run-B relying more strongly on mineral tuning, while both preserve low inter-site variance.

Correlation analysis (Figure 12) corroborates these patterns at the ensemble level, showing directionally consistent associations between

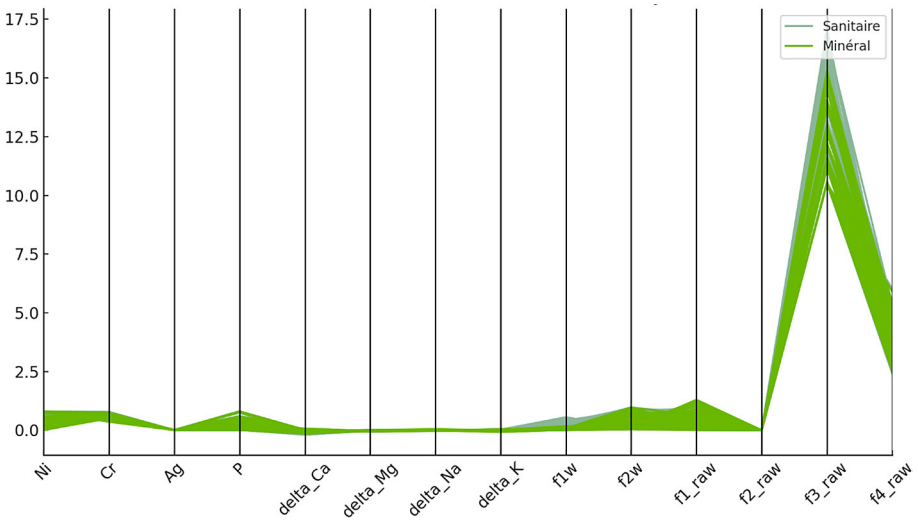

Figure 11. Parallel-coordinates summary of decision variables and raw objectives (Run-A vs Run-B)

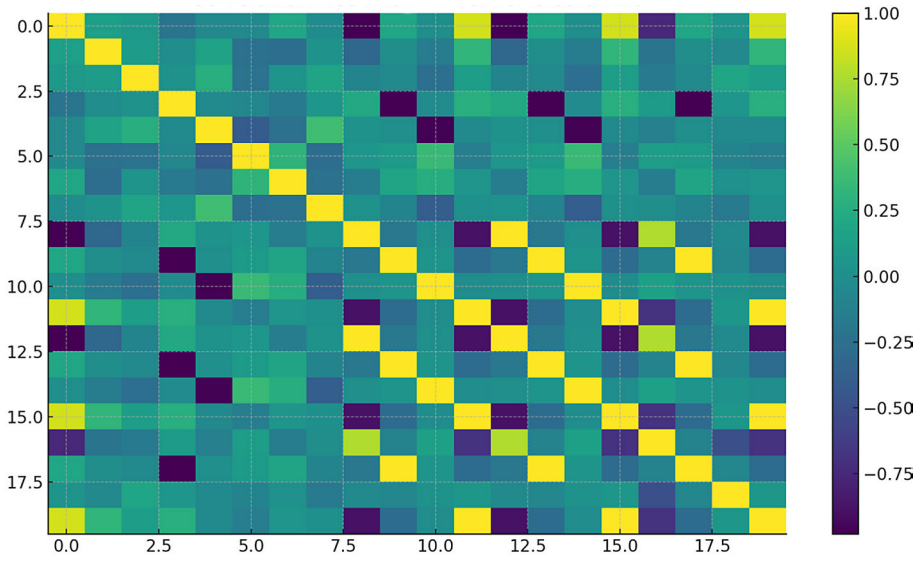

Figure 12. Correlation matrix (decisions r_e, Δ_m and objectives f_1–f_4)

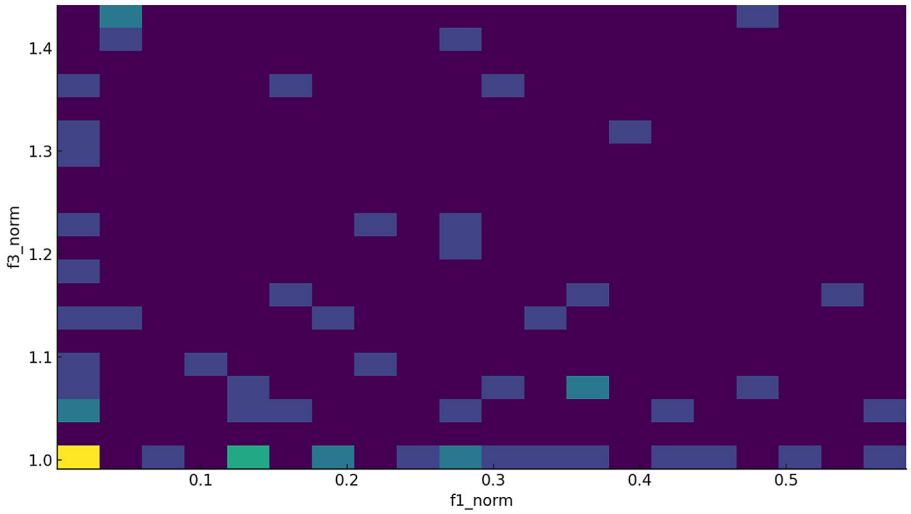

Figure 13. 2-D density in (f_1, f_3) showing the compromise knee

key levers and their corresponding objectives. Finally, the 2D density in the sanitary–mineral plane (Figure 13) identifies a high-density "knee" region in which substantial improvements are achieved before penalties in competing objectives rise sharply. This region provides a principled anchor for selecting balanced operator presets.

Shortlisting implementable operator-facing scenarios

To translate dense Pareto sets into operationally implementable options, a shortlist was constructed by combining (i) extreme non-dominated points corresponding to minimum sanitary exceedance (min f_1), minimum inter-site variance (min f_2), minimum mineral deviation (min f_3), and minimum effort (min f_4), and (ii) additional balanced compromise points selected from the knee region. Table 3 reports the shortlisted strategies with their decision vectors and objective values, providing an audit-ready basis for operator selection. For communication and training, the shortlist is summarized as a set of operator-facing strategy archetypes in objective space (Figure 14), while Table 3 remains the numerical reference.

DISCUSSION

Practical implications for operators in Kénitra

The optimisation outputs translate into two operational regimes that can be implemented as priority-dependent presets. First, when regulatory compliance is the overriding constraint (e.g., audit periods or after transient hydraulic disturbances), operators should select solutions from the

Table 3. Shortlisted optimization strategies for Kénitra: trade-offs in sanitary, homogeneity, mineral, and cost space

ID	Profile	Selection	r_{Ni}	r_{Cr}	r_{Ag}	r_{P}	Δ_{Ca}	Δ_{Mg}	Δ_{Na}	Δ_{K}	f_1 (exceed.)	f_2 (site var.)	f_3 (mineral)	f_4(cost)	Composite score
14	Sanitary-priority	Extreme sanitary (min f_1)	0.7997	0.6751	0.0129	0.0845	0.0090	-0.0024	-0.0184	0.0259	0.0023	0.0004	10.5306	4.6003	0.5457
62	Mineral-priority	Extreme mineral (min *f3*)	0.7969	0.6496	0.0098	0.5922	0.0114	-0.0544	-0.0239	0.0278	0.0043	0.0001	10.5075	5.3168	0.5217
1	Sanitary-priority	Max homogeneity (min f_2)	0.7994	0.7824	0.0040	0.7995	0.0098	-0.0672	-0.0281	0.0261	0.0025	0.0000	10.5102	6.0333	0.5243
52	Sanitary-priority	Minimal cost (min f_4)	0.1619	0.5864	0.0018	0.0820	0.0138	-0.0211	-0.0223	0.0323	0.9333	0.0004	10.5185	2.4162	0.7064
61	Mineral-priority	Balanced compromise	0.7951	0.7515	0.0061	0.7998	0.0119	-0.0393	-0.0226	0.0165	0.0057	0.0000	10.5096	5.9054	0.5233
2	Sanitary-priority	Balanced compromise	0.7957	0.6835	0.0018	0.5913	0.0070	-0.0691	-0.0308	0.0261	0.0052	0.0001	10.5156	5.4102	0.5234
5	Sanitary-priority	Balanced compromise	0.7955	0.6284	0.0033	0.4292	0.0138	-0.0364	-0.0201	0.0112	0.0078	0.0002	10.5144	4.9718	0.5261

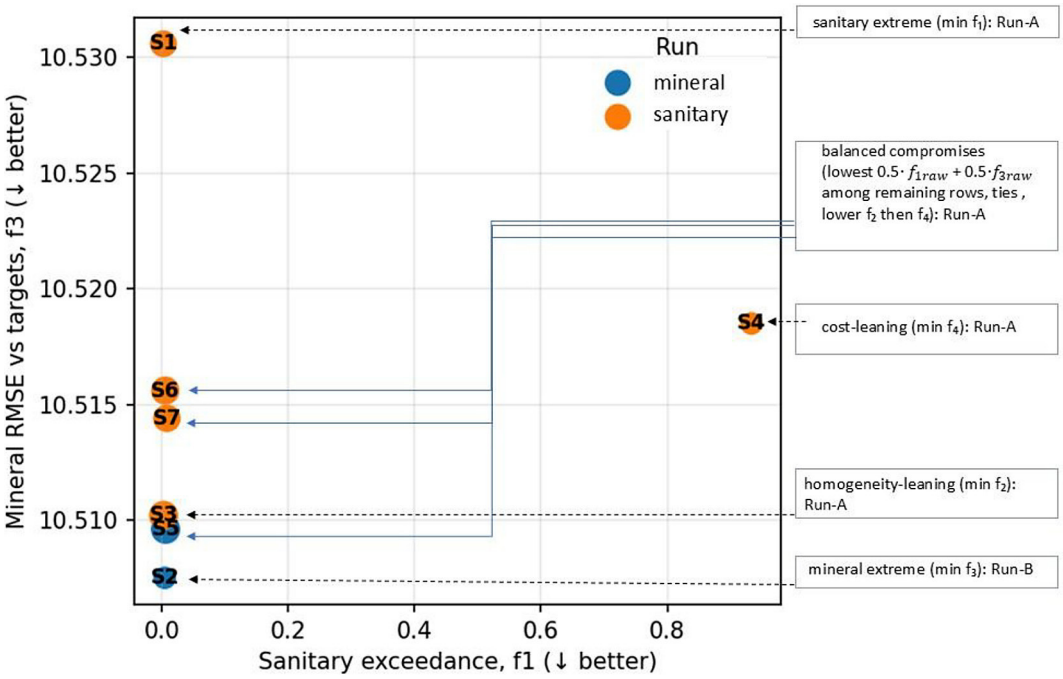

Figure 14. Seven operator-facing strategies in objective space

sanitary-priority profile that minimise the exceedance objective f_1 (Table 3; Figure 8a). These settings typically rely on stronger reductions in Ni and Cr (r_{Ni}, r_{Cr}) while maintaining very low inter-zone variance f_2. The expected trade-offs are a moderate increase in the effort proxy f_4 and, in some cases, a small deterioration in mineral-balance deviation f_3, which is consistent with evidence that aggressive control actions can perturb buffering conditions and corrosion equilibria within distribution systems (Liu et al., 2013; Makris et al., 2014).

Second, when palatability, scaling and corrosion control are the dominant drivers under stable sanitary conditions, the mineral-priority profile provides solutions with markedly lower mineral deviation f_3 at still-low f_2, primarily via targeted adjustments in Mg and Na (Δ_{Mg}, Δ_{Na}) (Table 3; Figure 8b). In this regime, modest relaxation in f_1 may occur unless additional risk-element reductions are accepted (and f_4 allowed to rise), reflecting the operational reality that mineral conditioning and sanitary barriers must be jointly managed rather than optimised in isolation (WHO, 2017; WHO, 2022). Baseline summaries used to motivate these regimes (Figures 3–5) are computed directly from the cleaned samples × analytes matrix provided in Supplementary Data S1

Between these two poles, both profiles consistently identify a compromise corridor in which sanitary exceedance and mineral deviation are jointly suppressed at moderate effort while preserving low inter-zone variance (Figures 9 and 13). This corridor provides defensible default operating points for routine conditions, and the seven-row shortlist (Table 3) offers a ready-to-pilot portfolio spanning the extremes and the most efficient compromises (Figure 14).

Why inter-site variance f_2 remains persistently low

Two complementary mechanisms explain the systematically low values of f_2 across both priority profiles. First, the intervention levers are system-wide (uniform r_e and Δ_m across zones), so multiplicative reductions and additive mineral shifts act coherently on all samples, compressing between-site dispersion by construction. Second, many distribution-system responses (e.g., conditioning effects mediated by biofilms, disinfectant residual behaviour, and pipe-scale interactions) can exhibit network-level coherence when interventions are applied uniformly, which further suppresses inter-zone disparities even when absolute concentrations shift (Liu et al., 2013; Makris et al., 2014). This interpretation also clarifies a practical limitation: if operators require targeted equity control under strongly heterogeneous hydraulics or localised sources, future extensions should introduce site-specific levers rather than only global adjustments.

Robustness and recommended sensitivity analyses

Indicator-based diagnostics support the interpretation that differences between Run-A and Run-B reflect genuine priority trade-offs rather than optimisation artefacts. Comparable hypervolume and IGD values, alongside moderate spacing behaviour, indicate similar convergence and coverage of the non-dominated sets in normalised objective space (Zitzler et al., 2003; Emmerich and Deutz, 2018). For operational adoption and reviewer-level robustness, the following sensitivity analyses are recommended:

a) multi-seed repeats of NSGA-III to quantify front stability, reporting the distribution of HV/IGD/spacing across seeds;
b) bootstrap envelopes of indicators computed on resampled non-dominated sets to communicate uncertainty in coverage and spread;
c) preference robustness tests (perturbing weights and reference directions) to verify that shortlisted strategies remain competitive under realistic shifts in value judgments; and
d) decision-space diversity checks to ensure the shortlist spans implementation-diverse lever combinations so that practical constraints (e.g., reagent availability) do not eliminate all top candidates.

Positioning within the Moroccan hydro-environmental literature

The identified compromise corridor and the dual-regime interpretation are consistent with the hydrochemical complexity reported in Moroccan systems. Coastal salinisation and marine influence have been documented in the Rharb–Mnasra system, complicating mineral management and spatial uniformity (Aguedai et al., 2022). Other Moroccan case studies report seasonal co-occurrence of metallic and bacteriological indicators (El Hammioui et al., 2022; Hicham et al., 2022), and mixed geogenic–anthropogenic contributions to metals in surface waters (Ferdaous et al., 2015). Against this backdrop, the present study advances beyond descriptive WQI/GIS-style surveillance by delivering a prescriptive optimisation workflow that returns implementable decision vectors with explicit, quantifiable trade-offs among sanitary exceedance (f_1), homogeneity (f_2), mineral balance (f_3), and effort (f_4). It also complements distribution-system literature on episodic quality deterioration driven by hydraulics, scales, and biofilms (Liu et al., 2013; Makris et al., 2014) by mapping the frontier of achievable states under controllable levers, thereby providing a transferable decision-support structure for data-limited utilities.

CONCLUSIONS

This study demonstrates that routine municipal monitoring can be converted from descriptive reporting into an optimisation-driven decision workflow for urban drinking-water management under data-limited conditions. The proposed ML + many-objective framework successfully generated policy-relevant portfolios that make trade-offs explicit between sanitary exceedance, inter-site uniformity, mineral balance, and intervention effort for the Kénitra distribution network. In this sense, the study objective was achieved: the workflow produces implementable operator-facing scenarios rather than a single nominal "best" setting.

From a scientific standpoint, the main contribution is the explicit integration of spatial homogeneity and mineral-balance control as optimisation objectives alongside regulatory compliance, within a lightweight surrogate-assisted architecture suitable for small monitoring datasets. The results reveal a consistent compromise corridor where both sanitary risk and mineral imbalance can be jointly reduced at moderate effort, and they translate this corridor into a concise set of actionable intervention profiles with predictable system-wide outcomes. The surrogate stage showed heterogeneous predictive reliability across elements: performance was strongest for the most structured targets in this dataset (notably Ni and P), whereas near-detection variability and small-sample constraints attenuated accuracy for some metals (e.g., Cr and Ag). This limitation does not invalidate the optimisation within the observed operating domain, but it motivates targeted data enrichment for the weakest targets. Future work should (i) extend monitoring across seasons and operational states to strengthen surrogate stability, (ii) introduce site-specific levers to better reflect network heterogeneity, (iii) replace the current effort proxy with a monetised cost model aligned with utility budgets, and (iv) report uncertainty envelopes (multi-seed optimisation and resampling-based indicator stability) to support risk-aware adoption. Overall, the framework fills the gap between monitoring and decision support

by providing a reproducible pathway from measurements to implementable, priority-dependent operating choices.

Data availability. The dataset used to generate Figures 3–5 (raw laboratory export and the cleaned samples × analytes matrix; n = 14 household taps across seven distribution zones) is provided with this revision as Supplementary Data S1 (Excel). A formatted snapshot of the cleaned dataset is also included in the manuscript (Table S1) to facilitate reading and reuse.

REFERENCES

1. Abbi, B., Touazit, A., Gliti, O., Igouzal, M., Pontie, M., Lemenand, T., Charki, A., (2025). modelling salt rejection in nanofiltration and reverse osmosis membranes using the Spiegler-Kedem model enhanced by a bio-inspired metaheuristic algorithms: particle swarm optimization and grey wolf optimization. *J. sustain. dev. energy water environ. syst. 13*, 1–22. https://doi.org/10.13044/j.sdewes.d13.0565
2. Aguedai, H., Jelbi, M., Lahlou, F., Abdelaziz, M., (2022). Hydrochemical and geophysical characterization of the Mnasra coastal aquifers (Rharb basin NW Morocco). *Arab J Geosci 15*, 1480. https://doi.org/10.1007/s12517-022-10738-7
3. Benyoussef, S., El Ouarghi, H., Arabi, M., El Yousfi, Y., Azirar, M., Ait Boughrous, A., (2021). Assessment of the impact of Imzouren city's WWTP on the quality of the surrounding groundwater, at the Rif Central (Northern Morocco). *E3S Web Conf. 240*, 01008. https://doi.org/10.1051/e3sconf/202124001008
4. Blank, J., Deb, K., (2020). Pymoo: Multi-objective optimization in python. *IEEE Access 8*, 89497–89509. https://doi.org/10.1109/ACCESS.2020.2990567
5. Deb, K., Jain, H., (2014). An Evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints. *IEEE Trans. Evol. Computat. 18*, 577–601. https://doi.org/10.1109/TEVC.2013.2281535
6. El Hammioui, Y., Anarghou, H., Belghiti, M., Hachi, T., Elhamdouni, D., Laiboud, C., Essabiri, H., Boumalkha, O., Khaffou, M., Abba, E.H., (2022). Evaluation of the metallic and bacteriological quality of well water in Khenifra province (Morocco). *IOP Conf. Ser.: Earth Environ. Sci. 1090*, 012031. https://doi.org/10.1088/1755-1315/1090/1/012031
7. Ferdaous, L., Mohammed, B., Nouzha, B., Hamid, L., Abrerrahim, L., (2015). *Etude de la qualite physicochimique et de la contamination metallique des eaux de surface du bassin versant de Beht (Maroc).*
8. Gliti, O., Chafik El Idrissi, M., Igouzal, M., (2024). Intelligent design of an ultra-thin near-ideal multilayer solar selective absorber using grey wolf optimization linked to deep learning. *Ecol. Eng. Environ. Technol. 25*, 70–87. https://doi.org/10.12912/27197050/175785
9. Gliti, O., Igouzal, M., El Idrissi, M.C., (2023). Development of a sustainable solar water desalinator using a novel hollow hemispherical grid shell solar selective absorber designed via phasor particle swarm algorithm. *Ecol. Eng. Environ. Technol. 24*, 130–149. https://doi.org/10.12912/27197050/173510
10. Hicham, G., Mustapha, A., Mourad, B., Abdelmajid, M., Ali, S., Yassine, E.Y., Mohamed, C., Ghizlane, A., Zahid, M., (2022). Assessment of the physico-chemical and bacteriological quality of groundwater in the Kert Plain, northeastern Morocco. *International Journal of Energy and Water Resources 6*, 133–147. https://doi.org/10.1007/s42108-021-00157-x
11. Liu, G., Verberk, J.Q.J.C., Van Dijk, J.C., (2013). Bacteriology of drinking water distribution systems: an integral and multidimensional review. *Appl Microbiol Biotechnol 97*, 9265–9276. https://doi.org/10.1007/s00253-013-5217-y
12. Makris, K.C., Andra, S.S., Botsaris, G., (2014). Pipe scales and biofilms in drinking-water distribution systems: Undermining finished water quality. *Critical Reviews in Environmental Science and Technology 44*, 1477–1523. https://doi.org/10.1080/10643389.2013.790746
13. World Health Organization (Ed.), (2022). Guidelines for drinking-water quality, Fourth edition incorporating the first and second addenda. World Health Organization, Geneva.
14. World Health Organization (Ed.), (2017). *Guidelines for drinking-water quality, Fourth edition incorporating the first addendum.* ed. World Health Organization, Geneva.
15. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Da Fonseca, V.G., (2003). Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Trans. Evol. Computat. 7*, 117–132. https://doi.org/10.1109/TEVC.2003.810758