

Evaluating a post-incident drinking water quality monitoring network using consumer complaints and verified leak events

Dongwoo Jang

jdw@inu.ac.kr

Incheon National University <https://orcid.org/0000-0003-2487-8451>

Research Article

Keywords: water distribution system, district metered area, water-quality monitoring, customer complaints, discoloration, event detection

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11221808/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

After a major discoloration (“red water”) incident in 2019, Incheon, Republic of Korea, deployed high-frequency multi-parameter water-quality sensors at the district metered area (DMA) scale across the affected Gongchon service area. This study evaluates whether that monitoring investment is spatially aligned with the water-quality problems consumers actually experience, using customer complaints as an independent, consumer-referenced benchmark. Two validation windows were assembled: 23 stations over 151 observation days in 2022 (83,352 hourly records), intersected with a georeferenced ledger of 649 complaints including field-inspection outcomes; and five continuously operated blocks in 2023, intersected with 284 verified leak repairs. Three findings emerge. First, during a September 2022 event, 22 of 23 stations registered a sequential, network-wide turbidity anomaly beginning approximately two days before a single-day surge of 140 complaints, showing that the network can detect a network-scale disturbance before the complaint peak. Second, this capability did not extend to the block scale: 84% of perceptible, block-attributable complaints arose in unmonitored blocks, and block-day tests of sensor anomalies against complaints were null under every event definition—a spatial mismatch between sensor placement and complaint burden, not a sensor failure. Third, verified leak events showed a modest but statistically detectable excess of turbidity anomalies within 48 h (26% vs. 12% placebo; 10,000 draws, $p = 0.024$). An in-sample placement audit shows the current layout covers 16.0% of the complaint burden—the 49th percentile of random placement—whereas a complaint-informed 23-block selection would have covered 89.0% retrospectively. Placement, as much as sensing, determines monitoring value.

1. Introduction

Drinking water distribution networks are increasingly instrumented with continuous, multi-parameter water-quality sensors, yet the question of what these networks actually detect—and, equally, what they systematically miss—is rarely answered against independent evidence. The most consequential deployments are often reactive: after a highly publicized water-quality incident, utilities install monitoring hardware across the affected area under strong public and regulatory pressure. The Republic of Korea offers a prominent example. In 2019, a valve operation during an emergency supply conversion triggered a discoloration (“red water”) incident in Seo-gu, Incheon, that persisted for approximately twenty days, affected large parts of the Gongchon water treatment plant service area, and prompted nationwide regulatory change, including statutory designation of intensively managed vulnerable areas (Lee et al. 2025b; Park et al. 2023; Lee et al. 2024). The economic damage of such incidents is substantial, with estimated daily losses during the Incheon event exceeding 900 million KRW (Lee et al. 2025a). In its aftermath, block-scale (district metered area, DMA) multi-parameter water-quality stations were installed throughout the affected service area. This paper asks whether such a post-incident network delivers the early-warning function it was installed for, using the utility’s own customer complaint records as the reference.

Discoloration remains the dominant water-quality reason customers contact their water supplier (Vreeburg and Boxall 2007). Its physical basis—gradual accumulation of particulate material on pipe walls and its mobilization by hydraulic disturbance—is well established (Vreeburg and Boxall 2007; Husband and Boxall 2011; Boxall et al. 2023), including for trunk mains whose disturbances propagate across entire downstream networks (Cook et al. 2016), and for transient events that mobilize material in the first dynamic surge (Weston et al. 2021). Because mobilized material manifests as turbidity, continuous turbidity monitoring is the natural instrument for detecting discoloration risk, although the turbidity–solids relationship is variable (Pourcel et al. 2020). Critically, the literature has long treated customer contacts as the de facto performance indicator of distribution water quality: utilities “use customers as water quality monitors” where sensor coverage is sparse, and recent work has begun to formalize customer feedback as an input to monitoring design—for sampling-programme refinement (Dakša and Kokina 2025) and for machine-learning prediction of complaint occurrence from water-quality parameters (Lamolla et al. 2026). Multi-parameter, multi-sensor fusion studies likewise report that water-quality events are “commonly first detected by consumers” and seek to reverse that order (Gleeson et al. 2025).

A parallel literature addresses event detection algorithms for online water-quality data, from the CANARY family of tools (Hart et al. 2007; Nafsin and Li 2021; Leow et al. 2017; Burkhardt et al. 2022; Cronin et al. 2025) to deep-learning and hybrid detectors (El-Shafeiy et al. 2023; Li et al. 2023; Liao et al. 2024; Chen et al. 2026; Hu et al. 2022b) and national-scale water-quality prediction systems in Korea (Im et al. 2022). A third strand optimizes sensor placement for contamination detection, typically minimizing detection time or maximizing coverage over simulated intrusion ensembles (Hu et al. 2018; Hu et al. 2022a; Diao et al. 2023; Gautam et al. 2022; Zanelli et al. 2025; Palleti 2025; Yao et al. 2026; Acharya et al. 2026). What is missing at the intersection of these strands is an empirical, system-level evaluation: a deployed block-scale network, an extended multi-station record, and an independent consumer-referenced benchmark of experienced water quality. Simulation-based placement studies rarely confront where complaints actually arise; detection-algorithm studies are usually validated on synthetic anomalies or single-site records; and complaint studies rarely have co-located continuous sensor data. Physical disturbances that should produce detectable signatures—pipe breaks, repairs, and low-pressure events—are documented causes of water-quality degradation (Qi et al. 2018; Besner et al. 2008; Mattioli et al. 2025), and burst detection from hydraulic data is well studied (Xu et al. 2020; Wéber et al. 2020; Wang et al. 2022), but the water-quality footprint of verified leak events in operational sensor records remains largely unquantified at the DMA scale under placebo-controlled designs. Complaints, moreover, are an impact-based observation process rather than a direct measurement of network water quality, and therefore require perceptibility screening, origin classification, and careful spatial attribution before they can serve as an evaluation benchmark or placement objective—steps that this study makes explicit.

This study evaluates the post-incident monitoring network of the Gongchon service area (Seo-gu, Incheon) along three axes, each matched to a distinct failure or success mode of a monitoring programme:

1. **Network-scale event detection.** Can the network detect a genuine, large-scale disturbance before consumers react? A September 2022 event provides a revealing test case: a network-wide turbidity cascade traversed 22 of 23 stations approximately two days before a one-day surge of 140 complaints.
2. **Spatial coverage.** Does sensor placement coincide with where consumers experience problems? Block-attributed complaint burden is compared with the sensor layout, and block-day panel tests quantify the (absent) statistical association between sensor anomalies and complaints.
3. **Physical-event signatures.** Do verified network-origin physical events (leaks and their repairs) leave detectable water-quality signatures? A superposed epoch analysis with a placebo (random-date) control quantifies the turbidity response within 48 h of 31 verified leak events inside monitored blocks.

The contribution is threefold. First, empirical evaluations combining operational sensor records, spatially resolved consumer complaints, and verified physical network events remain limited; this study provides such an evaluation for an operational, post-incident DMA network, based on 83,352 hourly multi-parameter records and a complaint ledger with field-inspection outcomes. Second, it reports an honest null result—block-day sensor anomalies do not associate with complaints under any tested event definition—and diagnoses its cause as a spatial placement mismatch (84% of perceptible complaints arise in unmonitored blocks) rather than sensor insensitivity, with direct design implications. Third, it demonstrates a placebo-controlled epoch method for attributing water-quality disturbances to verified physical network events, yielding a modest but statistically detectable turbidity signature ($p = 0.024$, 10,000 draws), robust in direction across window, threshold, and baseline choices. Together, the three axes convert a routinely collected complaint stream into a rigorous evaluation instrument for monitoring-network design, and motivate complaint-burden-weighted sensor placement as a complement to simulation-based optimization.

2. Materials and methods

2.1 Study area and monitoring network

The study area is the mainland Seo-gu portion of the Gongchon water treatment plant service area in Incheon, Republic of Korea—the area affected by the 2019 discoloration incident. The distribution network is organized into 85 planning blocks (DMAs) delineated in the utility’s block-system GIS (Fig. 1); island subareas (Ganghwa, Yeongjong) served by the same plant were excluded using the GIS supply-zone attributes so that all data layers (complaints, sensors, network) share one consistent boundary. Following the 2019 incident, multi-parameter water-quality stations were installed at block inlets or representative points within blocks. Each station records turbidity, residual chlorine, pH, electrical conductivity, and water temperature at hourly (or finer, aggregated to hourly) resolution. Station measurements are point observations at the installation location and are not assumed to represent the full spatial distribution of quality within a block; this point-versus-block distinction is retained throughout the analysis and revisited in the limitations (Section 4.4).

2.2 Data: two validation windows and the complaint ledger

Two independent validation windows were assembled, each pairing sensor data with an independent record of consumer- or field-verified events (Table 1).

2022 window (network-wide, complaint-referenced). Daily operational reports from the utility's SCADA archive were parsed into an hourly database covering 23 stations within the 1 August–31 December 2022 window. Two archive dates (13–14 December) are missing for all stations, leaving 151 complete observation days and 83,352 hourly records (23 stations × 151 days × 24 h). All 23 stations reported on every archived day. This window overlaps the utility's 2022 water-quality complaint ledger, which records 2,428 complaints city-wide, of which 649 originate in Seo-gu. Ledger fields include reception date-time, administrative neighbourhood (dong), free-text complaint description, complaint type, a free-text field-response record (including on-site measurements of residual chlorine and pH where taken), and—in 46% of Seo-gu records—an explicit block identifier. City-wide, the dominant complaint types are discolored water (n = 650), foreign particles (n = 330), and filter discoloration (n = 103).

2023 window (five blocks, leak-referenced). Year-round hourly water-quality data for 2023 were available for the five blocks whose stations operated continuously through the year (blocks 866, 871, 874, 876, 884; 8,760 h each; pH unavailable). This window intersects a georeferenced 2023 leak-complaint register for Seo-gu (641 reports with coordinates). False reports (n = 252) and duplicate reports (n = 20) were excluded, as—conservatively—were records transferred to other authorities or still open at year-end; the 284 reports confirmed by completed repair works constitute the verified leak set. Coordinate-based spatial joining against the block GIS located 31 verified leak events inside the five monitored blocks; these form the treatment set for the epoch analysis (Section 2.6).

Using two windows converts an apparent weakness—the complaint ledger (2022) and the continuous year-round records (2023) do not coincide for the full network—into two independent tests: the 2022 window tests complaint correspondence network-wide, and the 2023 window tests physical-event signatures where continuous data exist.

Table 1
Data assets and validation windows.

Item	2022 window	2023 window
Period	Aug–Dec 2022 (151 observation days)	Jan–Dec 2023 (8,760 h)
Water-quality stations	23 (all operational)	5 blocks (866, 871, 874, 876, 884)
Parameters	Turbidity, residual Cl, pH, EC, temperature	Turbidity, residual Cl, EC, temperature (pH unavailable)
Records	83,352 hourly records	5 × 8,760 hourly records
Reference data	Complaint ledger: 649 Seo-gu complaints (2,428 city-wide); block field in 46%; field-response text incl. on-site Cl/pH	Leak register: 641 georeferenced reports; 284 verified by completed repairs; 31 within monitored blocks
Role	Network-wide, complaint-referenced evaluation	Leak-referenced physical-event signature test

2.3 Robust anomaly scoring and magnitude-aware event tiers

Raw parameter values are non-stationary across seasons and heterogeneous across stations, so anomalies were scored with a robust, station-specific z-score. For each station and parameter, a rolling 14-day baseline window (shifted by one hour so the current observation never contributes to its own baseline) provides the median and the median absolute deviation (MAD); the robust z-score is

$$z = (x - \text{median}) / (1.4826 \cdot \text{MAD}),$$

where 1.4826 scales MAD to the standard deviation under normality. Median/MAD scoring is insensitive to the outliers it is designed to detect (Leys et al. 2013). Before scoring, records were quality-controlled with fixed validity ranges: zero readings, which the telemetry system uses as a missing-data code, were treated as missing, as were physically implausible values (turbidity outside 0–50 NTU, residual chlorine outside 0–2 mg/L, water temperature outside 0–40°C). Hours in which the rolling MAD equals zero leave the z-score undefined; such hours were excluded from event detection rather than assigned an arbitrary floor. An *event* at a station is defined as $z \geq 3$ sustained for at least 2 consecutive hours, suppressing single-sample spikes typical of sensor noise.

Because a purely statistical threshold can flag fluctuations that are far below human perception (Section 3.3), two magnitude-aware tiers were additionally defined for the panel tests: **Tier A** (absolute-magnitude criterion: turbidity ≥ 0.5 NTU or residual chlorine < 0.1 mg/L) and **Tier B** (hybrid criterion: $z \geq 5$ with turbidity ≥ 0.2 NTU). The Tier A anchors correspond to Korean statutory and operational reference points —0.1 mg/L is the minimum free residual chlorine required in distribution under the Water Supply and Waterworks Installation Act, and 0.5 NTU is the operational turbidity standard applied to treated water—

used here as magnitude thresholds for perceptibility-relevant excursions, not as compliance determinations at the tap; Tier B requires a statistically extreme excursion to also have non-trivial absolute magnitude.

2.4 Complaint processing: origin stratification and three-tier spatial matching

Complaints were processed in three steps, with the following flow: of the 649 Seo-gu complaints, 326 were retained as perceptible (315 routine water-quality test requests and 8 non-sensory records excluded); all 326 could be attributed to a block—301 (92%) directly from the ledger’s block field (T1) and 25 via the neighbourhood-to-block mapping (T2)—and origin stratification then classified 90 as network-origin, 19 as premise-origin, and 217 as undetermined. First, *perceptibility screening*: complaint types reflecting direct sensory experience of water at the tap (discolored water, foreign particles, filter discoloration, odour/taste) were retained as “perceptible” complaints; administrative and non-sensory records were excluded. Second, *origin stratification*: the free-text field-response records were classified by rule-based text analysis into network-origin (mains-side cause identified or remediated by network action, e.g., flushing), premise-origin (in-building plumbing cause identified), and undetermined. Premise-origin complaints act as a negative control—no distribution-side sensor should be expected to anticipate them—directly addressing the concern that consumer complaints conflate distribution and building-plumbing problems. The classification rules were fixed before the association analyses and are summarized in Table 2; classification was performed by a single coder applying the written rules, without independent double-coding—a reproducibility limitation acknowledged in Section 4.4. Among 326 perceptible, block-attributable Seo-gu complaints, 90 were classified network-origin, 19 premise-origin, and 217 undetermined (the field-response text is often terse, which is itself reported as a data limitation).

Table 2

Rule-based classification of complaint origin from the ledger’s free-text field-response records.

Category	Criterion (field-response content)	Typical indicators
Network-origin	A distribution-side cause was identified, or the complaint was resolved by network action	Mains flushing/drain-off performed; hydrant or mains work noted; discoloration linked to network operation
Premise-origin	An in-building plumbing cause was identified	Internal/service piping named as cause; building tank or private-side valve issue; resolved by the premise owner
Undetermined	Neither side established from the record	Water-quality test only; advice given; no cause stated; absent or terse response text

2.5 Block-day panel analysis

To test whether sensor-detected anomalies associate with complaints at the operationally relevant block scale, a block-day panel was constructed over monitored blocks × observation days in the 2022 window.

Each block-day is labelled (i) event/no-event by each detection definition (z-only; Tier A; Tier B) and (ii) complaint/no-complaint if at least one perceptible complaint was matched to that block within ± 2 days. Association was tested with Fisher's exact test and odds ratios. Analysing block-days rather than individual complaints avoids pseudo-replication from complaint clusters and gives every monitored block-day, including quiet ones, its correct weight. Covariate-adjusted regression (e.g., mixed-effects logistic models with block and calendar terms) was considered but is uninformative here: the panel contains 74 complaint block-days, and no more than one of them coincides with an anomaly under any detection definition, so such models separate rather than estimate; exact tests with full contingency cells and a matching-window sensitivity analysis (± 1 to ± 5 days) are reported instead. To complement the binary panel, block-level complaint counts were additionally modelled cross-sectionally with a negative binomial regression (robust errors, log pipe length as exposure offset) on sensor presence, length-weighted pipe age, and metallic-mains share, testing whether complaint burden is explained by monitoring status or asset covariates.

2.6 Superposed epoch analysis of verified leak events with placebo control

For the 2023 window, each of the 31 verified leak events was aligned at its report time ($t = 0$), and station z-scores for the leak's host block were superposed from $t = -72$ h to $t = +72$ h. Two statistics were computed: (i) the epoch-mean turbidity z-score before (-72 to 0 h) and after (0 to $+48$ h) the event; and (ii) a *hit rate*, the fraction of events followed within 48 h by a sustained turbidity anomaly ($z \geq 3$, ≥ 2 h) in the host block. Because turbidity anomalies also occur in the ordinary course of operations, the hit rate is meaningless without a null reference. A placebo distribution was therefore generated by drawing 10,000 sets of 31 random dates (matched to the same blocks and avoiding ± 48 h around the true event windows) and recomputing the hit rate for each set; the empirical p-value is the fraction of placebo sets whose hit rate equals or exceeds the observed rate. Because leak occurrence and background turbidity both vary seasonally, a second, stratified placebo was constructed by additionally matching each placebo date to its event's calendar month and day of week (10,000 draws). Two same-site duplicate reports (same block, within 72 h and 200 m) were also identified, and the analysis was repeated on the deduplicated event set. This design attributes signatures to leak events only in excess of the network's background anomaly rate. To guard against the arbitrariness of any single detection definition, the entire analysis—including the placebo—was repeated over a sensitivity grid of detection windows (24, 48, 72 h) and z-thresholds (2, 3, 4), with the 48 h/ $z \geq 3$ combination pre-designated as the primary definition before hit rates were computed: 48 h spans the typical report-to-repair interval in the register, and $z \geq 3$ is the same anomaly threshold used throughout Section 2.3. Event alignment uses the report time, not the (unobserved) physical onset of the leak, so lags between onset, reporting, and repair blur the epoch alignment; this can only dilute a true post-event signal, making the estimated excess conservative.

2.7 Complaint-weighted placement experiment

To quantify what the identified coverage gap implies for sensor deployment, a retrospective placement experiment was performed. Monitored blocks were selected greedily by descending perceptible

complaint burden, and, for each budget k , the fraction of the total block-attributable perceptible complaint burden falling inside the selected blocks was computed and compared with the coverage of the actual 23-station layout. Because the selection and the evaluation use the same 2022 complaint data, the resulting curve is an in-sample upper bound rather than a predictive guarantee; it is presented as a diagnostic of how far the current layout sits from a complaint-informed one, not as an operational design, which would additionally weigh hydraulic representativeness, detection objectives, and installation constraints. Three robustness checks accompany the in-sample curve: (i) a temporal split, selecting blocks on January–June complaints and evaluating coverage on July–December complaints (and the reverse); (ii) a bootstrap of the 326 complaints (1,000 resamples) to assess the stability of the burden ranking; and (iii) a random-placement baseline, the coverage distribution of 23 blocks drawn uniformly from the block set (10,000 draws), against which the actual layout’s coverage is located as a percentile; bootstrap 95% confidence intervals for the coverage curve were obtained from the complaint resamples.

All processing used Python (pandas, geopandas, SciPy); coordinates were handled in the Korean TM projection of the utility GIS.

3. Results

3.1 Seasonal baseline dynamics of the monitored blocks

The 2023 five-block record establishes the seasonal envelope against which anomalies must be judged. Residual chlorine shows a pronounced seasonal cycle: the summer median across the five blocks is 0.62 mg/L versus 0.74 mg/L in winter (Fig. 2), and the hourly temperature–chlorine correlation is moderate ($r = -0.49$), consistent with temperature-accelerated residual decay along the network with substantial operational variation superimposed on the seasonal drift. Turbidity baselines at all five stations sit far below the visibility threshold (medians of approximately 0.05 NTU), which is the precondition for the detection paradox reported in Section 3.3: statistically extreme excursions can remain physically trivial. These baseline properties motivated both the rolling robust- z design (Section 2.3) and the magnitude-aware tiers.

3.2 A network-scale event: the September 2022 cascade

The 2022 window contains one revealing large-scale episode (Fig. 3). Beginning at 04:00 on 18 September 2022, sustained turbidity anomalies ($z \geq 3$, ≥ 2 h) appeared sequentially at 22 of the 23 stations over a 15-hour span—first at stations 874 and 835 (04:00), then 876 and 837 (05:00), 802 and 840 (06:00), and progressively across the network until station 884 (17:00)—with a maximum z of 42.5. The ordered onset times are consistent with a network-scale disturbance propagating through the transmission–distribution system rather than with independent local events, although the specific operational cause could not be confirmed from the available records (Section 4.4).

Two days later, on 20 September, the complaint stream surged to 140 complaints in a single day—the largest daily count in the ledger—concentrated in the Geomam neighbourhood, which the empirical mapping (146 ledger cases) attributes to block 847. Block 847 has no water-quality station. The nearest monitored evidence is indirect: adjacent station 848 recorded $z = 11.9$ during the cascade, and station 869 recorded an extreme excursion to 17.7 NTU (roughly 380 times its baseline) on 23 September, temporally consistent with, but not independently verified as, a post-event flushing or drain-off response, although the ledger’s terse field-response entries for these days (“inquiries on cause and stabilization”) do not document the specific operational action, and utility work orders were not available to confirm it.

The September event thus demonstrates both faces of the network in a single episode. The sensors delivered a coherent early signal—a network-wide anomaly pattern approximately two days before the consumer response peaked—yet the block where consumers were most affected was outside the sensor layout, so the signal could support only network-level, not block-level, situational awareness.

3.3 The coverage gap: complaints arise where sensors are not

Across the full 2022 window, block-day panel tests found no association between sensor-detected anomalies and complaints under any definition (Table 3). With the purely statistical definition ($z \geq 5$; 186 event block-days), the odds ratio was 0.74 (Fisher’s exact $p = 0.80$).

Table 3
Block-day panel tests of sensor-detected events against perceptible complaints (± 2 days), 2022 window (3,473 block-days), Fisher’s exact test with full contingency cells (a = event & complaint, b = event only, c = complaint only, d = neither).

Event definition	a	b	c	d	Odds ratio	p
z-only (turbidity $z \geq 5$, ≥ 2 h)	3	183	71	3,216	0.74	0.80
Tier A (turbidity ≥ 0.5 NTU or $Cl < 0.1$ mg/L)	0	25	74	3,374	0	1.0
Tier B ($z \geq 5$ and turbidity ≥ 0.2 NTU)	0	37	74	3,362	0	1.0

The null association is unchanged when the matching window is varied from ± 1 to ± 5 days (z-only OR 0.70–0.82, $p \geq 0.66$; Tier A and B: zero co-occurrences at every window). Inspection showed that most z-defined anomalies correspond to absolute turbidity of only 0.06–0.11 NTU—statistically extreme against ultra-low baselines but two orders of magnitude below the ~ 5 NTU visibility threshold; a z-score alone cannot distinguish “unusual for this sensor” from “noticeable to a person”. However, the magnitude-aware tiers did not rescue the association: Tier A produced 25 event block-days and Tier B 37, and neither coincided with matched complaints (both OR = 0). The magnitude-aware tiers are nonetheless operationally tractable: over the 3,473 monitored block-days of the window, Tier A fired on 25 block-days and Tier B on 37, equivalent to approximately 2.6 and 3.9 alarm-days per station-year respectively—an alarm burden a control room can realistically triage.

The explanation is spatial, not algorithmic (Fig. 4). Of the 326 perceptible, block-attributable complaints, 84% originate in blocks with no water-quality station; only 16% arise in monitored blocks. The seven highest complaint-burden blocks (847, 880, 813, 872, 843, 867, 881) all lack sensors—including block 847, the epicentre of the September event. Under this placement, block-level complaint verification is nearly impossible in principle: the panel's null result is a structural consequence of the coverage gap and should be read as a diagnosis of the sensor layout rather than a failure of the sensors or the detector. This is the central finding for monitoring-programme design, and it was reachable only because the evaluation used an independent, spatially resolved consumer benchmark rather than the sensor data alone.

The concentration of complaints in unmonitored blocks is not an artifact of population size. For the 17 blocks of the Geomdan subregion, the utility's block ledger provides official resident-registration counts, allowing per-capita rates to be computed (Table 4). The ranking is preserved—and sharpened—under normalization: unmonitored blocks 881 and 872 carry 36.1 and 17.8 perceptible complaints per 1,000 residents, whereas the monitored blocks in the same subregion lie at 0.35 (block 866) and 0.18 (block 871) per 1,000, with three further monitored blocks (874, 876, 884) recording no attributable perceptible complaints at all. Conversely, the most populous block of the subregion (873, 22,692 residents, unmonitored) shows a low rate of 0.22 per 1,000—population does not drive burden. Rates for the smallest blocks (42–675 residents) are statistically unstable and are reported for completeness rather than ranking; the qualitative conclusion—that the high-burden blocks are not simply the populous ones—does not depend on them.

The 84% figure is also insensitive to how the large undetermined-origin class is resolved. Restricting to the 90 network-origin complaints raises the unmonitored share to 88.9%; assigning all 217 undetermined complaints to network origin gives 86.3%; and the 19 premise-origin complaints—the negative control—show no such concentration (47.4% in unmonitored blocks, indistinguishable from a spatially unstructured pattern at this sample size). Whatever the true composition of the undetermined class, the spatial mismatch stands.

A cross-sectional negative binomial regression of block-level complaint counts (83 blocks; log pipe length as exposure offset; robust errors) confirms the descriptive pattern statistically: complaint burden rises steeply with metallic-mains share (incidence rate ratio 1.44 per 10-percentage-point increase, 95% CI 1.19–1.72, $p < 0.001$), is unrelated to length-weighted pipe age once material is conditioned on (IRR 0.97 per year, $p = 0.48$), and—critically—is unrelated to sensor presence (IRR 0.82, 95% CI 0.41–1.65, $p = 0.58$). Monitored and unmonitored blocks are therefore statistically indistinguishable in complaint burden after asset covariates, which is precisely the problem: the sensor layout carries no information about where consumers experience harm.

Nor is the gap an artifact of measuring coverage strictly within block boundaries. The block layout is spatially dense: 53 of the 62 unmonitored blocks share a boundary with at least one monitored block, and extending coverage to monitored-plus-adjacent blocks would nominally span 96.6% of the

perceptible complaint burden—including block 847, which adjoins monitored block 848. The September event shows that adjacent stations do register network-scale disturbances (848 reached $z = 11.9$ during the cascade). Yet the block-day association remains null: a point sensor samples the water of its own block, and the localized mobilization events that generate most discoloration complaints (Vreeburg and Boxall 2007; Husband and Boxall 2011; Boxall et al. 2023) need not propagate across a block boundary at detectable magnitude. Physical adjacency is therefore no substitute for in-block sensing where localized events dominate, while network-scale disturbances are the (rarer) case that neighbouring stations can cover.

The high-burden blocks are also structurally distinct in ways that support a mains-side origin of their complaint load. From the utility’s pipe GIS (installation year and material per segment, length-weighted), the seven highest-burden blocks average 17.2 years of pipe age with metallic mains—predominantly ductile or cast iron and coated steel—making up 89.4% of pipe length, against 12.0 years and 52.3% for the service area as a whole (and 14.0 years, 53.9% for the monitored blocks). Older, ferrous-dominated networks are exactly where wall-bound material accumulates and mobilizes (Vreeburg and Boxall 2007; Husband and Boxall 2011; Boxall et al. 2023), so the complaint geography is consistent with the asset-condition geography—reinforcing that these blocks are rational monitoring targets rather than statistical noise.

Table 4
Perceptible complaint burden normalized by registered population, Geomdan subregion
(the 17 blocks with official per-block resident-registration counts).

Block	Monitored	Population	Perceptible complaints	Per 1,000 residents
885	No	42	2	47.6*
881	No	194	7	36.1*
872	No	675	12	17.8*
896	No	173	1	5.8*
880	No	4,473	22	4.9
873	No	22,692	5	0.22
866	Yes	8,577	3	0.35
871	Yes	16,776	3	0.18
874	Yes	11,300	0	0
876	Yes	5,873	0	0
884	Yes	2,282	0	0

*Small-denominator rates (< 1,000 residents); unstable, shown for completeness.

3.4 Verified leak events leave a detectable turbidity signature

The 2023 leak-referenced window provides positive evidence that the sensors do register genuine network-origin physical events when those events occur inside monitored blocks (Fig. 5). Superposed over the 31 verified leak events, the epoch-mean turbidity z-score rises from + 0.9 in the 72 h before the report to + 1.7 in the 48 h after it, and 26% of events (8/31) were followed within 48 h by a sustained turbidity anomaly ($z \geq 3$, ≥ 2 h) in the host block. Against 10,000 random-date placebo draws (mean hit rate 12%, 95% band 3–23%), the empirical p-value is 0.024; the risk difference is approximately 14 percentage points, corresponding to a roughly twofold elevation over the background rate. The excess is robust in direction, though modest in size, across every specification tested. Under the stringent stratified placebo—each placebo date matched to its event’s block, calendar month, and day of week—the background rate rises to 15% and the excess becomes borderline ($p = 0.052$). Deduplicating the two same-site repeat reports (Section 2.6) strengthens the result (8/29, 28% vs. 12%, $p = 0.014$). Across the window–threshold grid of Table 5, the observed hit rate exceeds the placebo mean in all nine combinations, reaching significance at the primary definition and near-significance ($p = 0.06$ – 0.09) at most adjacent settings, with only the narrowest-window, highest-threshold cell (2 hits) clearly underpowered. Varying the baseline window preserves the direction of the excess (observed-to-placebo ratios of 1.2 and 1.7 under 7- and 28-day baselines) but not its significance ($p = 0.41$ and 0.21), so the signature is best characterized as a modest excess that is statistically detectable at the 14-day baseline scale used throughout this study—pre-specified before the leak analysis—and suggestive rather than definitive at other scales. The fine structure of the epoch curve (Fig. 5a) also addresses the alignment caveat of Section 2.6. The 24 h immediately before the report is the quietest sub-window of the epoch (mean $z = 0.49$, versus 1.16 in -72 to -24 h, the latter inflated by a single extreme excursion at one event), and only 3 of 31 events show a sustained pre-report anomaly against 8 of 31 afterwards: the turbidity signature follows the report—consistent with mobilization during the leak’s later stages or the repair works themselves (Besner et al. 2008)—rather than building up before it. Repair-completion timestamps are not recorded in the register for the verified events, so the report-to-repair lag cannot be quantified; the 48 h primary window is intended to span it. Residual chlorine showed no analogous epoch response. Roughly one in four verified leak-and-repair sequences therefore leaves a turbidity fingerprint detectable at the block station over and above the background anomaly rate—physical confirmation that network-origin disturbances of exactly the kind implicated in discoloration (mobilization by hydraulic disturbance during the leak or its repair (Vreeburg and Boxall 2007; Weston et al. 2021; Qi et al. 2018; Besner et al. 2008)) are within the network’s detection reach, and a counterpoint to the premise-origin complaints that no distribution-side sensor could anticipate.

Table 5

Sensitivity of the leak-signature hit rate to the detection window and z-threshold (31 verified leak events; placebo: 300 random-date draws per combination for the grid, 10,000 draws for the primary definition; primary definition in bold).

Window	z-threshold	Hits	Observed	Placebo mean [95% band]	Empirical p
24 h	2	6/31	19%	15% [3–27%]	0.30
24 h	3	5/31	16%	8% [0–16%]	0.09
24 h	4	2/31	6%	5% [0–13%]	0.42
48 h	2	10/31	32%	22% [10–35%]	0.09
48 h	3	8/31	26%	12% [3–23%]	0.024
48 h	4	5/31	16%	7% [0–19%]	0.08
72 h	2	12/31	39%	26% [13–42%]	0.09
72 h	3	8/31	26%	15% [3–26%]	0.07
72 h	4	6/31	19%	10% [2–23%]	0.06

3.5 How far is the current layout from a complaint-informed one?

The placement experiment of Section 2.7 puts a number on the coverage gap (Fig. 6). Selecting monitored blocks by descending complaint burden, a single block—847, the September epicentre—would cover 47.9% of the block-attributable perceptible complaint burden, three times the 16.0% covered by the entire current 23-station network. Seven blocks reach 70.6%, and an equal budget of 23 blocks reaches 89.0%. The steep initial slope reflects the heavy concentration of consumer-perceived problems in a handful of blocks; the flat tail shows that most of the sensor budget in a complaint-informed layout buys marginal coverage, which is precisely the budget that hydraulic and security objectives (Hu et al. 2018; Hu et al. 2022a; Diao et al. 2023; Gautam et al. 2022; Zanelli et al. 2025; Palleti 2025; Yao et al. 2026; Acharya et al. 2026) should govern. These figures are in-sample (Section 2.7); the robustness checks bound how far they generalize. The full-year burden ranking itself is stable: in 1,000 bootstrap resamples of the complaints, block 847 ranks first in 100% of draws and the top-7 set retains on average 5.9 of its 7 members. The temporal split is more nuanced, and honestly reported: blocks selected on January–June complaints ($n = 69$) covered 68.1% of July–December complaints at $k = 7$ —close to the 70.6% in-sample figure, and far above the current layout’s 8.6% coverage of the same period—but the single top block chosen from the pre-event half-year (848, not 847) covered 0% of the event-dominated second half, and selection in the reverse direction transferred weakly (24.6% at $k = 7$ against the current layout’s 43.5% on January–June). Complaint geography is thus event-driven and time-varying at the single-block scale, while the multi-block burden set transfers usefully; a complaint-informed layout should rest on multi-year burden, not on one year’s epicentre. Finally, the current layout is no better than uninformed placement:

over 10,000 random draws of 23 blocks, mean coverage is 27% and the actual network's 16.0% sits at the 49th percentile of the random distribution—the layout performs like a random one with respect to consumer-impact coverage. Bootstrap 95% confidence intervals on the greedy curve (1,000 complaint resamples; medians 47.9% [42.6–53.7] at $k = 1$, 71.8% [67.2–76.4] at $k = 7$, and 92.0% [89.3–94.8] at $k = 23$) are tight relative to these contrasts, so none of the headline comparisons depends on sampling noise in the complaint set.

4. Discussion

4.1 What the network sees

The three axes yield a coherent picture. At the network scale, the monitoring system performs the function it was installed for: the September 2022 cascade was registered by 22 of 23 stations with an ordered spatial progression, and the anomaly pattern led the consumer response by roughly two days—operationally meaningful lead time for source tracing, valve strategy, and targeted flushing before complaints peak (Oh et al. 2026). At the block scale, and only where stations exist, the network also registers physical events: the placebo-controlled epoch analysis shows that verified leaks are followed by excess turbidity anomalies within 48 h, significant at the primary detection definition and directionally consistent across all definitions tested. These results place empirical, field-scale numbers on two capabilities that are usually asserted from simulation: network-scale disturbance detection and physical-event signature capture.

4.2 What it misses, and why

The same evaluation shows what the network misses. The null block-day panel result is not a marginal effect that better algorithms might recover; under the current layout, 84% of the perceptible complaint burden—including the September epicentre and all seven of the highest-burden blocks—falls where there is no station to test. The random-placement benchmark deserves emphasis of its own: with respect to complaint-burden coverage, the existing network did not outperform uninformed placement (49th percentile of random 23-block draws). This does not mean the deployment was careless—it was governed by hydraulic and administrative criteria that random placement does not capture—but it does mean the deployment was poorly aligned with the spatial distribution of observed consumer impacts, and that no amount of sensor or algorithm improvement can close a gap that is located in the layout. The z-only detection experiment adds an algorithmic caution with general relevance: over ultra-low turbidity baselines, purely statistical anomaly definitions generate events with no physical significance, and event detectors intended to relate to consumer experience must carry absolute-magnitude criteria alongside statistical ones. Reporting this null result matters: monitoring evaluations that report only successes systematically overstate what post-incident instrumentation achieves, and the failure mode identified here—placement mismatch—is invisible to any evaluation that uses sensor data alone.

Why did the placement miss the complaint burden? The layout was chosen for hydraulic and administrative convenience (block inlets, existing chambers, telemetry availability) in the immediate

aftermath of the 2019 incident, not against a spatial model of consumer risk. Simulation-based placement optimization (Hu et al. 2018; Hu et al. 2022a; Diao et al. 2023; Gautam et al. 2022; Zanelli et al. 2025; Palleti 2025; Yao et al. 2026; Acharya et al. 2026) would not necessarily have prevented the mismatch either, because its objective functions are built on intrusion ensembles over network topology, not on the empirical geography of consumer-perceived problems, which reflects pipe age and material, local hydraulics, and demographics that standard ensembles do not encode. The complaint ledger itself is the missing objective: it is spatially resolved, continuously collected at no additional cost, and—as shown here—stratifiable by origin so that premise-side noise can be filtered out.

4.3 Design implications

Three implications follow for utilities operating or planning post-incident monitoring networks. First, **complaint-burden-weighted placement**: allocate at least part of the sensor budget to blocks ranked by perceptible, network-origin complaint burden, rather than exclusively by hydraulic convenience or simulated intrusion coverage; in this network, monitoring the seven highest-burden blocks alone would have covered 70.6% of the observed perceptible complaint stream—including the September epicentre—against 16.0% for the current 23-station layout (Section 3.5), and a complaint-informed 23-block budget would have covered 89.0%. Second, **magnitude-aware multi-tier detection**: operational alarming should combine robust statistical scores (which adapt to station-specific baselines) with absolute thresholds tied to perception and regulation, as in the Tier A/B definitions; the z-only tier is useful for network-scale pattern detection (Section 3.2) but should not drive block-level consumer-impact alarms. Third, **institutionalize the complaint ledger as an evaluation instrument**: with modest improvements—systematic block attribution at intake (raising the 46% fill rate), structured field-response coding instead of free text, and routine georeferencing—the ledger would support continuous, quantitative auditing of monitoring performance of the kind performed once in this study, closing the loop between consumer experience and network operation (Dakša and Kokina 2025; Lamolla et al. 2026; Gleeson et al. 2025). Under-reporting biases this loop in a knowable direction: complaints under-represent problems in areas with lower reporting propensity, so a purely complaint-weighted objective would under-serve exactly those areas; this argues for complaint burden as one term in a hybrid placement objective—alongside hydraulic representativeness, asset condition, and population served (Hu et al. 2018; Hu et al. 2022a; Diao et al. 2023; Gautam et al. 2022; Zanelli et al. 2025; Palleti 2025; Yao et al. 2026; Acharya et al. 2026)—rather than a replacement for them.

None of these prescriptions is specific to Incheon. The evaluation framework—perceptibility screening and origin stratification of a routinely kept complaint ledger, block-day association tests, placebo-controlled epoch analysis of verified physical events, and a complaint-weighted placement audit—uses only data that most utilities already collect, and the failure mode it exposes (post-incident sensor deployments placed by hydraulic and administrative convenience, misaligned with consumer-impact geography) is a generic risk of reactive monitoring investment, not a local peculiarity. The transferable finding is that monitoring-network value must be assessed jointly on sensor sensitivity and spatial

representativeness relative to observed impacts; the Incheon network is simply the first operational system on which both dimensions have been measured against the same consumer benchmark.

4.4 Limitations

Four limitations bound the conclusions. (i) Station measurements are point observations; a block-level “anomaly” here means an anomaly at the station, and within-block heterogeneity is unobserved. (ii) The complaint ledger under-reports true consumer experience (not all affected consumers complain), its block attribution is partly empirical (T2 mapping), and 217 of 326 perceptible complaints could not be assigned an origin because field-response text is terse; the origin-stratified analyses therefore rest on the classifiable subset. (iii) The 2023 leak-signature analysis covers only the five continuously operated blocks, and 2022 and 2023 windows differ in station sets, so the three axes are internally consistent but not a single joint model. (iv) The September 2022 cascade’s root cause could not be confirmed from the available records; the propagation pattern and timing support a network-scale hydraulic disturbance, and the near-simultaneous, ordered onset across the whole service area is most consistent with a source-side or trunk-main disturbance (a treatment-plant process upset, a trunk-main operation, or a large valve manoeuvre) rather than coincident local causes (Cook et al. 2016); attribution to a specific operation remains open, and targeted retrieval of plant and valve operation logs for 17–18 September 2022 is the direct test. (v) The placement experiment of Section 3.5 is evaluated primarily in-sample; the temporal split shows that the multi-block burden set transfers forward usefully while single-block selection is event-sensitive, so complaint-informed placement should be built on multi-year records. Per-block registered-population counts were available only for the Geomdan subregion, so the per-capita robustness check covers 17 of the 85 blocks. (vi) Complaint-origin classification followed written rules (Table 2) but was applied by a single coder without independent double-coding, and leak epochs are aligned on report time rather than physical onset (completion timestamps are unrecorded for the verified events), which dilutes rather than inflates the estimated signature. None of these limitations affects the central spatial finding, which depends only on where complaints arise relative to where stations exist.

5. Conclusions

This study used customer complaints—perceptibility-screened, origin-stratified, and spatially matched—as an independent benchmark to evaluate a DMA-scale water-quality monitoring network installed after the 2019 Incheon discoloration incident. The network demonstrably works at the network scale: a September 2022 disturbance was captured as a sequential turbidity cascade across 22 of 23 stations, with roughly two days of lead time over the consumer response, and verified leak events inside monitored blocks left a modest excess of turbidity signatures within 48 h (26% vs. 12% placebo, $p = 0.024$; borderline under a season-matched null), directionally consistent across all windows, thresholds, and baseline choices tested. But the evaluation equally shows that the network cannot verify—or protect—where its sensors are not: 84% of perceptible, block-attributable complaints, the seven highest-burden blocks, and the September epicentre itself all lie in unmonitored blocks, and block-day tests of sensor anomalies against complaints are null at every event definition, including magnitude-aware ones. A

retrospective placement experiment makes the cost of the mismatch concrete: one complaint-informed block would have covered 47.9% of the perceptible complaint burden, and seven would have covered 70.6%, against 16.0% for the entire current network. The practical conclusions are that post-incident sensor deployments should weight placement by empirically observed, network-origin complaint burden; that event definitions must combine robust statistical scoring with absolute-magnitude criteria; and that routinely collected complaint ledgers, modestly upgraded, can serve as standing evaluation instruments for monitoring networks. More broadly, the study illustrates that the value of a monitoring network depends not only on sensor performance but also on where the sensors are deployed, and that honest evaluation against a consumer-referenced benchmark—including its null results—is a prerequisite for the next generation of water-quality early-warning systems.

Declarations

Competing Interests

The author declares no competing financial or non-financial interests that are relevant to the content of this article.

Funding

This work was supported by Incheon National University Research Grant in 2023.

Author Contributions

Dong-Woo Jang is the sole author and is responsible for the conceptualization, methodology, data curation, formal analysis, visualization, and writing of this article.

Acknowledgements

This work was supported by Incheon National University Research Grant in 2023.

Data Availability

The water-quality, complaint, and GIS data were provided by the Incheon Metropolitan City Waterworks Headquarters under a research agreement and cannot be shared publicly. Aggregated data supporting the findings are available from the author on reasonable request.

References

1. Acharya A, Ghimire AB, Ale Magar B, Shin S (2026) Sensor placement for contamination detection in urban water distribution system based on multidimensional resilience. *Systems* 14(4):422. <https://doi.org/10.3390/systems14040422>
2. Besner MC, Lavoie J, Morissette C, Payment P, Prévost M (2008) Effect of water main repairs on water quality. *J Am Water Works Assoc* 100(7):95–109. <https://doi.org/10.1002/j.1551-8833.2008.tb09679.x>
3. Boxall J, Blokker M, Schaap P, Speight V, Husband S (2023) Managing discolouration in drinking water distribution systems by integrating understanding of material behaviour. *Water Res* 243:120416. <https://doi.org/10.1016/j.watres.2023.120416>
4. Burkhardt JB, Sahoo D, Hammond B, Long M, Haxton T, Murray R (2022) Near real-time event detection for watershed monitoring with CANARY. *Environ Sci Adv* 1(2):170–181. <https://doi.org/10.1039/D2VA00014H>
5. Chen X, Guo J, Zhao Y, Feng J (2026) Research on water quality monitoring and abnormal early warning from the dual-driven perspective of knowledge and data. *PLoS ONE* 21(4):e0345405. <https://doi.org/10.1371/journal.pone.0345405>
6. Cook DM, Husband PS, Boxall JB (2016) Operational management of trunk main discolouration risk. *Urban Water J* 13(4):382–395. <https://doi.org/10.1080/1573062X.2014.993994>
7. Cronin L, Taylor CM, Briciu-Burghina C, Lucy FE, Regan F (2025) An open-source workflow for identifying hydrodynamic water quality events in rivers by continuous water quality monitoring and time-series data processing using R and US EPA CANARY. *MethodsX* 15:103538. <https://doi.org/10.1016/j.mex.2025.103538>
8. Dakša G, Kokina K (2025) Supplementing tap water quality monitoring through customer feedback: a GIS-centered approach. *Water* 17(21):3103. <https://doi.org/10.3390/w17213103>
9. Diao K, Emmerich M, Lan J, Yevseyeva I, Sitzenfrei R (2023) Sensor placement in water distribution networks using centrality-guided multi-objective optimisation. *J Hydroinform* 25(6):2291–2303. <https://doi.org/10.2166/hydro.2023.057>
10. El-Shafeiy E, Alsabaan M, Ibrahim MI, Elwahsh H (2023) Real-time anomaly detection for water quality sensor monitoring based on multivariate deep learning technique. *Sensors* 23(20):8613. <https://doi.org/10.3390/s23208613>
11. Gautam DK, Kotecha P, Subbiah S (2022) Efficient k-means clustering and greedy selection-based reduction of nodal search space for optimization of sensor placement in the water distribution networks. *Water Res* 220:118666. <https://doi.org/10.1016/j.watres.2022.118666>
12. Gleeson K, Husband S, Boxall J (2025) Multi-parameter multi-sensor data fusion for drinking water distribution system water quality management. *AQUA Water Infrastruct Ecosyst Soc* 74(7):451–465. <https://doi.org/10.2166/aqua.2025.002>
13. Hart D, McKenna SA, Klise K, Cruz V, Wilson M (2007) CANARY: a water quality event detection algorithm development tool. *World Environmental and Water Resources Congress 2007*. ASCE, pp 1–9. [https://doi.org/10.1061/40927\(243\)517](https://doi.org/10.1061/40927(243)517)

14. Hu C, Li M, Zeng D, Guo S (2018) A survey on sensor placement for contamination detection in water distribution systems. *Wirel Netw* 24(2):647–661. <https://doi.org/10.1007/s11276-016-1358-0>
15. Hu Z, Chen W, Tan D, Ye S, Shen D (2022a) Multi-objective model for optimal sensor placement in water distribution systems considering contamination probability variation-based contaminant impact. *J Clean Prod* 372:133445. <https://doi.org/10.1016/j.jclepro.2022.133445>
16. Hu Z, Chen W, Wang H, Tian P, Shen D (2022b) Integrated data-driven framework for anomaly detection and early warning in water distribution system. *J Clean Prod* 373:133977. <https://doi.org/10.1016/j.jclepro.2022.133977>
17. Husband PS, Boxall JB (2011) Asset deterioration and discolouration in water distribution systems. *Water Res* 45(1):113–124. <https://doi.org/10.1016/j.watres.2010.08.021>
18. Im Y, Song G, Lee J, Cho M (2022) Deep learning methods for predicting tap-water quality time series in South Korea. *Water* 14(22):3766. <https://doi.org/10.3390/w14223766>
19. Lamolla PV, Páez-Izquierdo A, Carrasco F, Paraira M, Ganzer M, Montes S, Botta F, Pacheco D, Farmani R, Tobella J, Martínez-Gomariz E, Berberana N, Poch M, Molinos-Senante M (2026) Linking drinking water quality to customer complaints: a data-driven study in the Barcelona metropolitan area. *Appl Water Sci* 16(8):234. <https://doi.org/10.1007/s13201-026-02882-2>
20. Lee H, Lim D (2025a) A study on the regional economic impact of water supply damage: focusing on damage cost estimation. *J Korean Regional Economics*
21. Lee J, Nam SH, Koo JW, Shin Y, Kim E, Song J, Hwang TM (2024) Detecting clusters and tracing management of water distribution system using space–time scan statistics and utility network modeling. *Process Saf Environ Prot* 187:249–258. <https://doi.org/10.1016/j.psep.2024.04.125>
22. Lee SI, Kim J, Kim A, Jun H (2025b) Risk pipe assessment and flushing block prioritization in water distribution networks based on DEA. *J Korean Soc Hazard Mitig* 25(3):85–95. <https://doi.org/10.9798/kosham.2025.25.3.85>
23. Leow A, Burkhardt J, Platten WE III, Zimmerman B, Brinkman NE, Turner A, Murray R, Sorial G, Garland J (2017) Application of the CANARY event detection software for real-time performance monitoring of decentralized water reuse systems. *Environ Sci Water Res Technol* 3(2):224–234. <https://doi.org/10.1039/C6EW00226A>
24. Leys C, Ley C, Klein O, Bernard P, Licata L (2013) Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J Exp Soc Psychol* 49(4):764–766. <https://doi.org/10.1016/j.jesp.2013.03.013>
25. Li Z, Liu H, Zhang C, Fu G (2023) Generative adversarial networks for detecting contamination events in water distribution systems using multi-parameter, multi-site water quality monitoring. *Environ Sci Ecotechnol* 14:100231. <https://doi.org/10.1016/j.ese.2022.100231>
26. Liao Z, Zhang M, Chen Y, Zhang Z, Wang H (2024) A Prediction – Detection – Judgment framework for sudden water contamination event detection with online monitoring. *J Environ Manage* 355:120496. <https://doi.org/10.1016/j.jenvman.2024.120496>

27. Mattioli MC, Benedict KM, Miko S, Barrett CE, Roundtree A, Kim S, Collier SA, Adam E, Gargano JW, Yoder JS, Vacs Renwick DA, Rotert K, Sullivan M, Sweeney S, Beach M, Hill VR (2025) Health effects and water quality following low pressure events in drinking water distribution systems in the United States. *Environ Health Perspect* 133(6):067002. <https://doi.org/10.1289/EHP15564>
28. Nafsin N, Li J (2021) Using CANARY event detection software for water quality analysis in the Milwaukee River. *J Hydro-environ Res* 38:117–128. <https://doi.org/10.1016/j.jher.2021.06.003>
29. Oh Y, Park H, Kim T, Koo J, Yoo DG (2026) Rapid response and impact assessment of water quality incidents in urban water distribution systems using forensic techniques. *Sci Rep* 16:15839. <https://doi.org/10.1038/s41598-026-44014-5>
30. Palleti VR (2025) Optimal sensor placement for contamination detection and identification in water distribution networks under demand uncertainty. *Comput Chem Eng* 194:109003. <https://doi.org/10.1016/j.compchemeng.2025.109003>
31. Park H, Kim K, Hyung J, Kim T, Koo J (2023) Decision-making for the hazard ranking of water distribution networks using the TOPSIS method. *Water Supply* 23(2):715–726. <https://doi.org/10.2166/ws.2023.013>
32. Pourcel F, Duchesne S, Ouellet M (2020) Evolution of the relationship between total suspended solids concentration and turbidity during flushing sequences of water pipes. *J Water Supply Res Technol AQUA* 69(4):376–386. <https://doi.org/10.2166/aqua.2020.127>
33. Qi Z, Zheng F, Guo D, Zhang T, Shao Y, Yu T, Zhang K, Maier HR (2018) A comprehensive framework to evaluate hydraulic and water quality impacts of pipe breaks on water distribution systems. *Water Resour Res* 54(10):8174–8195. <https://doi.org/10.1029/2018WR022736>
34. Vreeburg JHG, Boxall JB (2007) Discolouration in potable water distribution systems: a review. *Water Res* 41(3):519–529. <https://doi.org/10.1016/j.watres.2006.09.028>
35. Wang C, Xu Q, Qiang Z, Zhou Y (2022) Research on pipe burst in water distribution systems: knowledge structure and emerging trends. *J Water Supply Res Technol AQUA* 71(12):1408–1424. <https://doi.org/10.2166/aqua.2022.150>
36. Wéber R, Huzsvár T, Hős C (2020) Vulnerability analysis of water distribution networks to accidental pipe burst. *Water Res* 184:116178. <https://doi.org/10.1016/j.watres.2020.116178>
37. Weston SL, Collins RP, Boxall JB (2021) An experimental study of how hydraulic transients cause mobilisation of material within drinking water distribution systems. *Water Res* 194:116890. <https://doi.org/10.1016/j.watres.2021.116890>
38. Xu W, Zhou X, Xin K, Boxall J, Yan H, Tao T (2020) Disturbance extraction for burst detection in water distribution networks using pressure measurements. *Water Resour Res* 56(5). <https://doi.org/10.1029/2019WR025526>. e2019WR025526
39. Yao Y, Wallner M, Klawonn F (2026) Optimal sensor and sampling placement for contaminant detection: a comprehensive review across water distribution and wastewater collection systems. *Eng* 7(3):121. <https://doi.org/10.3390/eng7030121>

40. Zanelli E, Nicolini M, Goi D (2025) Optimal water quality sensor placement in water distribution systems: a computationally cost-effective genetic algorithm framework. Water 17(18):2786. <https://doi.org/10.3390/w17182786>

Figures

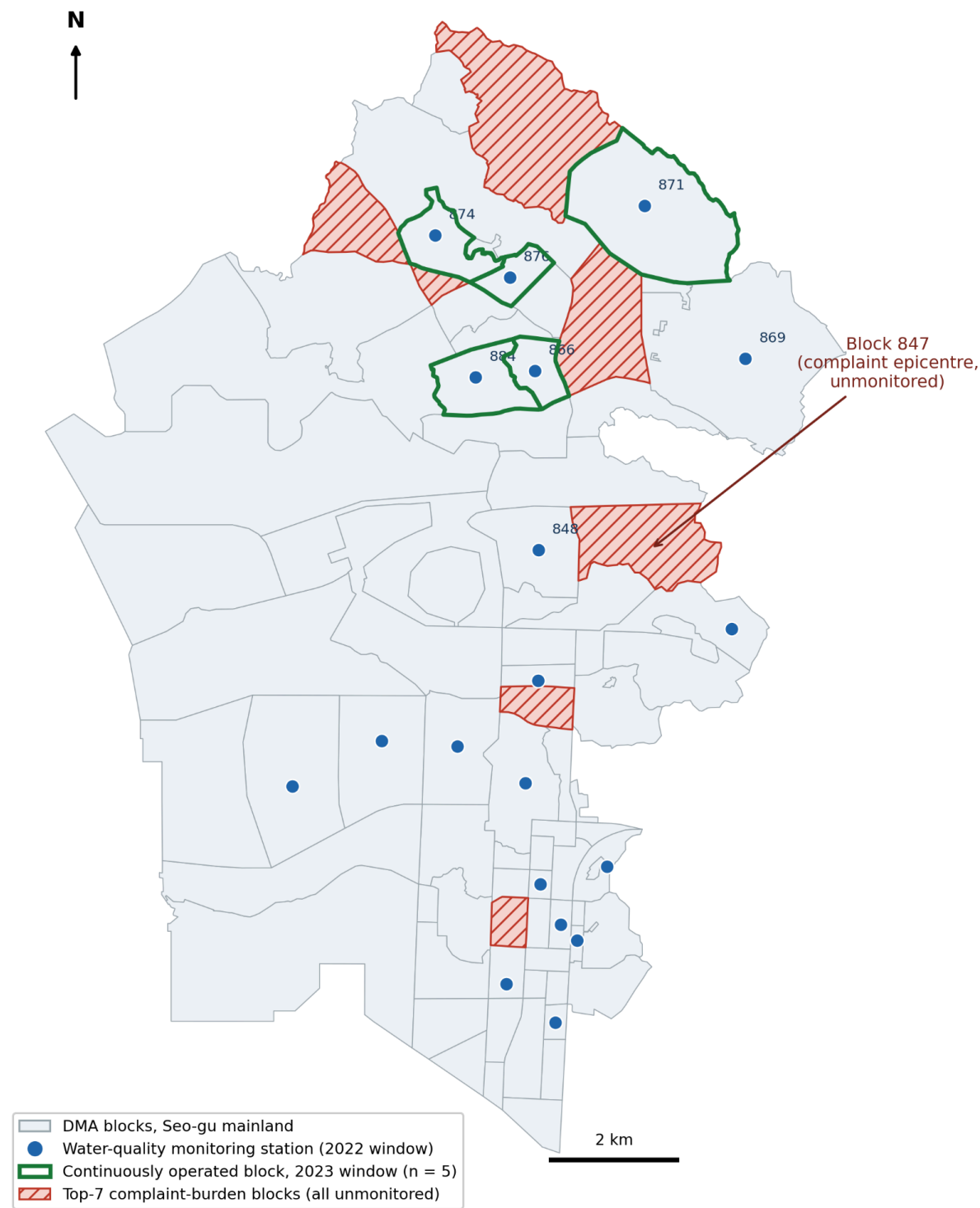


Figure 1

Study area: DMA blocks of the Seo-gu mainland portion of the Gongchon water treatment plant service area, Incheon, Republic of Korea. Blue dots mark water-quality monitoring stations of the 2022 window (23 stations; stations whose sub-blocks are not delineated in the planning GIS are shown at their host-block location). Green outlines mark the five blocks operated continuously through 2023. Hatched polygons are the top-7 complaint-burden blocks, none of which contains a station.

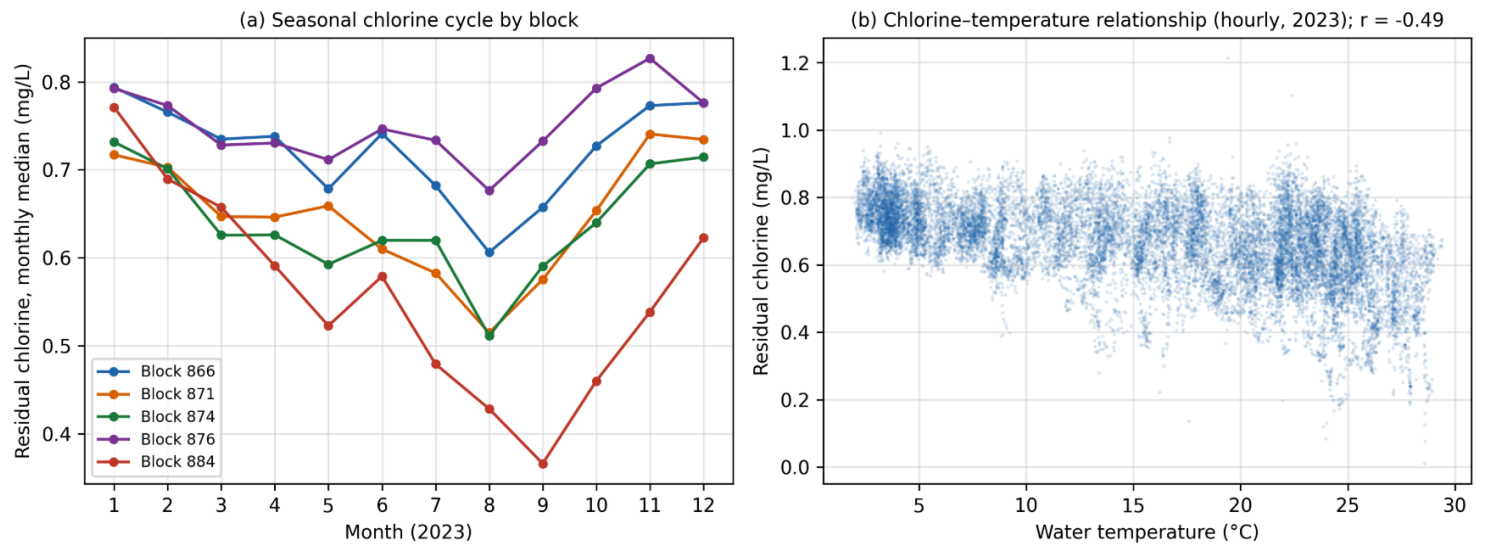


Figure 2

Seasonal water-quality baseline of the five continuously operated blocks (2023). (a) Monthly median residual chlorine by block, showing the summer depression of residuals (summer median 0.62 mg/L vs. winter 0.74 mg/L). (b) Hourly chlorine-temperature relationship ($r = -0.49$). Zero-coded missing values and physically implausible readings were excluded. The seasonal envelope motivates rolling, station-specific baselines.

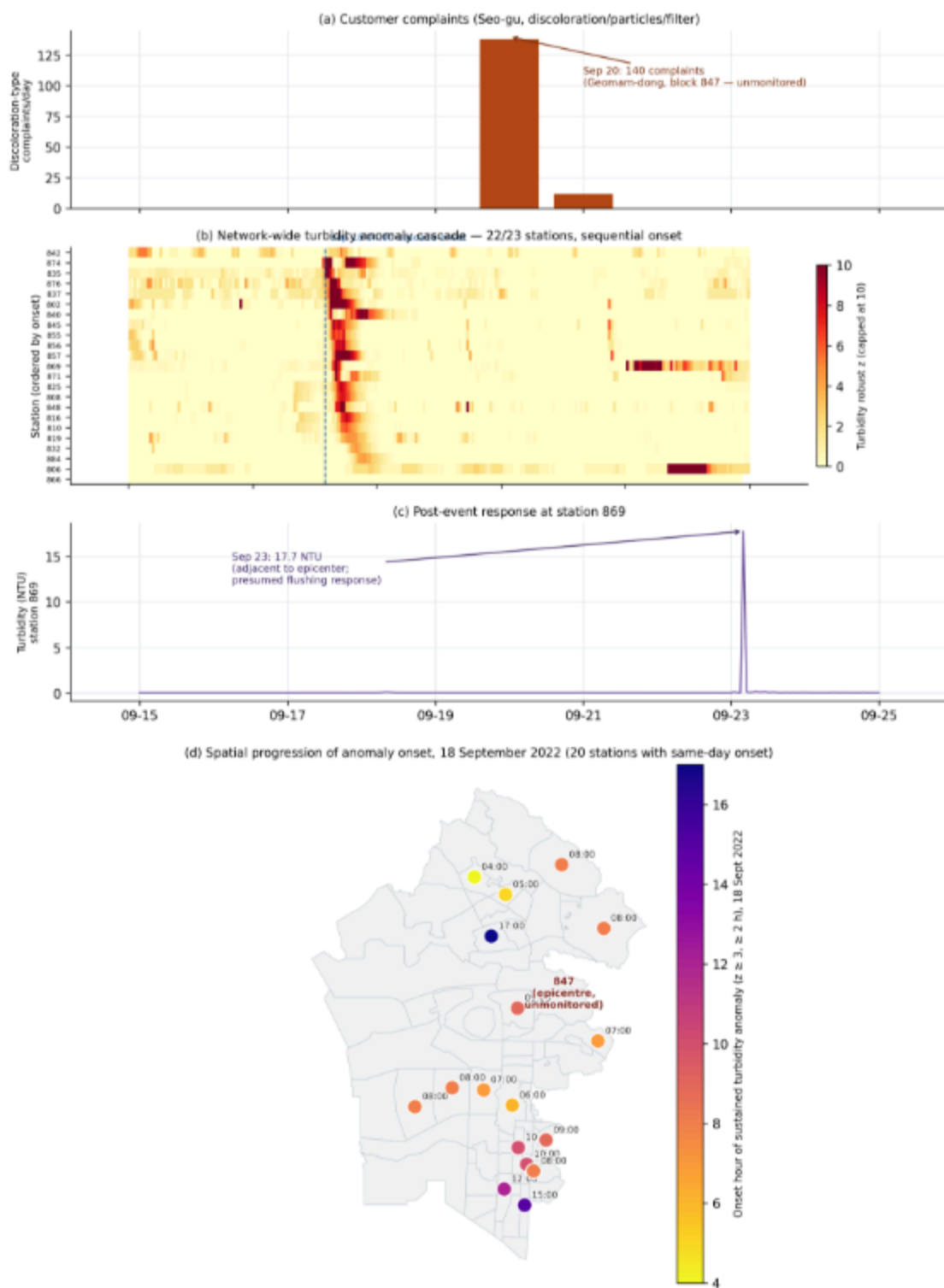


Figure 3

The September 2022 network-scale event. (a) Daily Seo-gu complaint counts, peaking at 140 on 20 September. (b) Robust turbidity z-scores across the 22 responding stations, showing the sequential onset from 04:00 on 18 September. (c) The extreme post-event excursion at station 869 (17.7 NTU on 23 September), temporally consistent with, but not independently verified as, a post-event flushing or drain-off response. (d) Spatial progression of anomaly onset on 18 September across the 20 stations with

same-day onset, coloured by onset hour; the ordered spatial pattern is consistent with a single propagating disturbance.

Complaint burden vs. sensor coverage — Seo-gu blocks, 2022

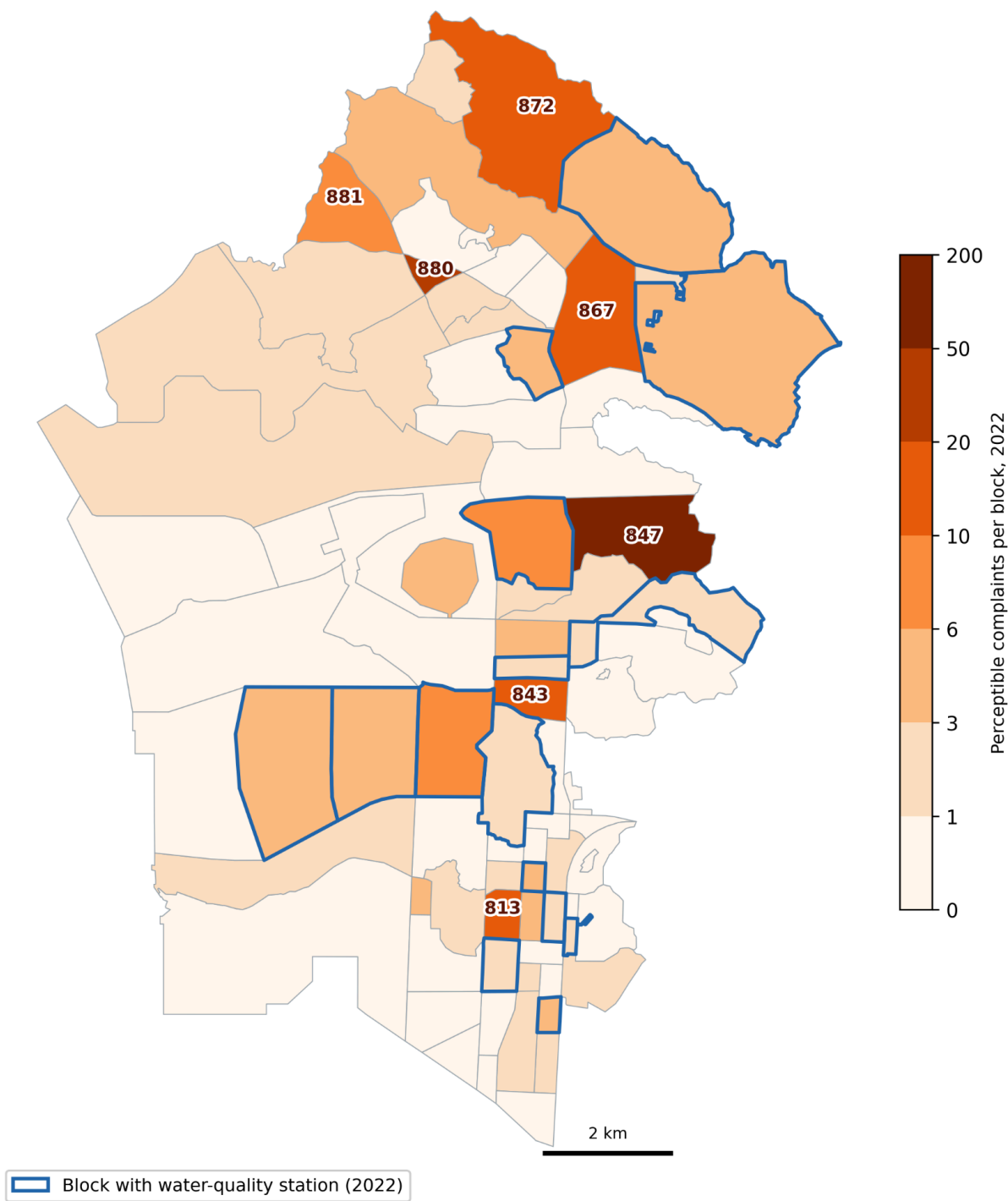


Figure 4

The coverage gap: block-attributed perceptible complaint burden (choropleth) versus sensor placement (blue outlines). 84% of perceptible, block-attributable complaints originate in unmonitored blocks; the

seven highest-burden blocks (labelled) all lack stations.

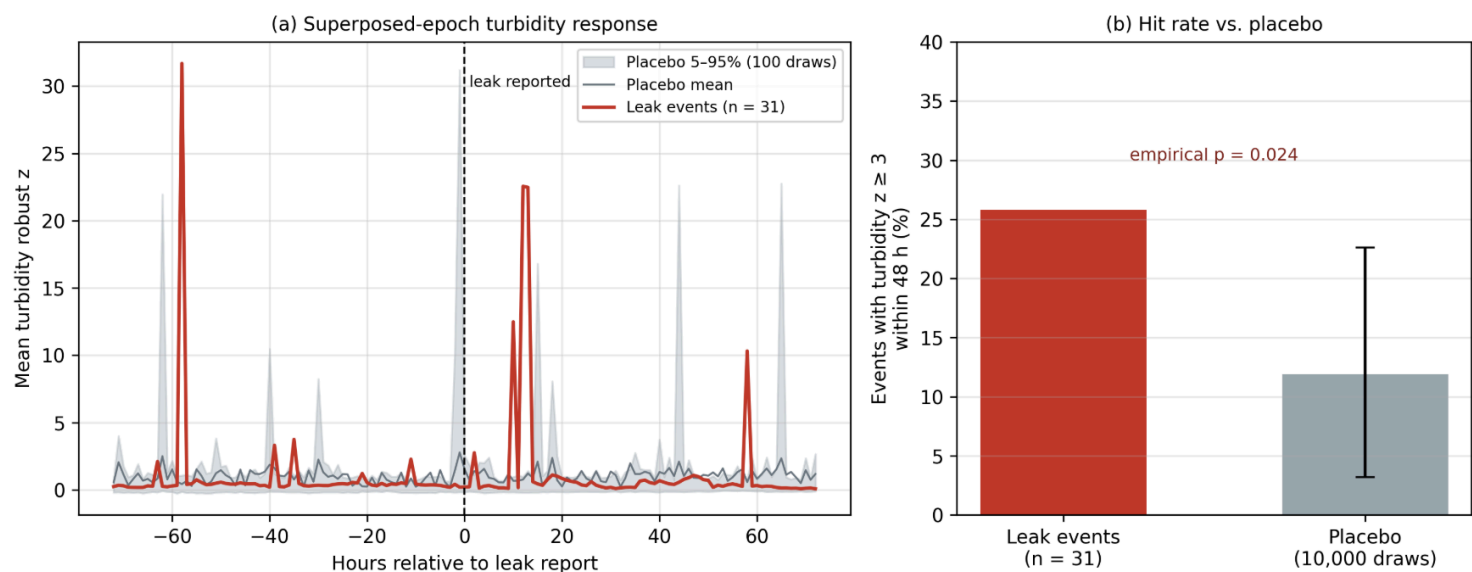


Figure 5

Superposed epoch analysis of the 31 verified leak events in monitored blocks (2023). (a) Epoch-mean turbidity z-score relative to leak report time (red), against the placebo mean and 5–95% band (100 random-date draws); the epoch mean rises from +0.9 (–72–0 h) to +1.7 (0–48 h). (b) Post-48 h hit rate (sustained turbidity $z \geq 3$, ≥ 2 h): observed 26% (8/31) versus the placebo distribution (mean 12%; 95% band 3–23%, 10,000 draws); empirical $p = 0.024$. See Table 5 for the window/threshold sensitivity grid.

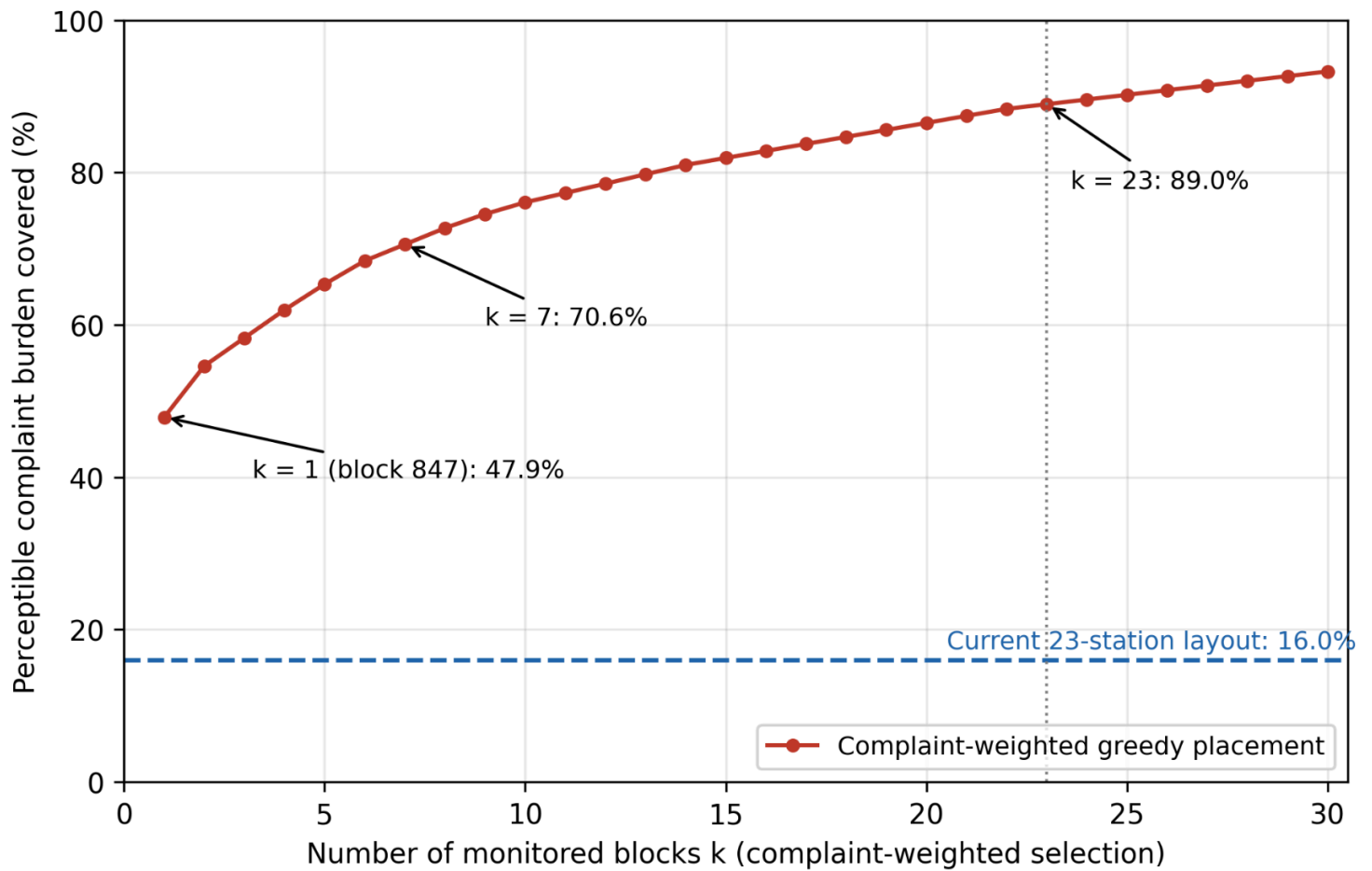


Figure 6

Retrospective complaint-weighted placement: share of block-attributable perceptible complaint burden covered as a function of the number of monitored blocks k (greedy selection by burden), compared with the actual 23-station layout (dashed line, 16.0%). In-sample evaluation on the 2022 ledger; see Section 2.7 for caveats.