Clock misalignment or real hydrogen loss?

A bounded simulation study for a short research stay (8–12 weeks) · Teodor Dimov (Trakia University) · **Proposed host (under discussion): I²CNER, Kyushu University — not confirmed.** The Japanese proposal is the full document; this English page is a supplement.

Illustrative model. All numbers below come from a public, reproducible benchmark (Python = JavaScript = expected_output.json). They do not describe any real station, alarm setting or product.

1. Question and core contribution

Question. When can an outlet clock offset be told apart from a real hydrogen loss (leak) in a dispenser-line mass balance — and when it cannot, which additional information makes a reliable decision possible?

Core contribution. To establish *when the available measurements do not allow a reliable decision*, and *what extra information changes that* (an independent time reference, a faster thermometer, a better flowmeter). Showing that a particular detector "wins" is not the goal; a negative result is an acceptable conclusion.

Measurable success criterion. At an equal budget of unnecessary shutdowns, identify the conditions under which a leak at or above a *minimum significant leak* is detected *within a pre-chosen deadline*. The deadline and the minimum significant leak are agreed with the proposed host (if the stay goes ahead) and frozen before evaluation.

2. Method (one system, one report)

- Line-inventory mass balance R_W over a 5 s window between inlet and outlet meters, Abel–Noble density checked against NIST (Leachman EOS), first-order meter lags; mass-balance leak monitoring as in API RP 1130 and the NRC Canada HRS mass-balance report.
- Timing faults kept separate: clock offset and drift, jitter (±0.1 s per sample), late packets handled as *hold-last-value*, and physical meter lag. Ordinary faults: frozen meter, frozen pressure gauge, meter-bias step.
- Calibration protocol: every threshold is set in advance on clean (no-leak) calibration fills with another seed and other stations, for an **equal nominal target** of 1% per fill in the worst of 7 declared clock-offset strata (±0.3 s). False-alarm rates are then measured on separate test fills. Confidence intervals: station-cluster bootstrap (400 fills per condition, 50 stations).
- Detectors: $u \cdot |\Delta q|$ is a first-order timing allowance, not a bound. CUSUM (Page 1954), sequential testing (Wald 1945) and GLR with a nuisance offset (Basseville & Nikiforov 1993). EKF / profile-likelihood identifiability (Särkkä 2013; Raue et al. 2009) are extensions only.

3. Trade-off (illustrative benchmark v1.5.0)

False alarms on no-leak test fills (unnecessary shutdowns) and correct leak diagnosis (sign included); TTD = median time to detection, detected cases only

Detector	FA, declared ±0.3 s (max)	FA +0.5 s	FA drift	FA meter lag 1 s	FA meter-bias step	2% leak	5% leak	TTD 2%	Mean mass lost 2%
A — inventory-corrected balance	1%	81.5%	57%	100%	0%	**0**/400	**2**/400	—	34.2 g
B2 — fixed timing allowance u = 0.4 s	0.5%	99.75%	35.75%	100%	13%	**36**/400	**307**/400	56.5 s	33.0 g
B3 — state-dependent allowance u_t	0.75%	99%	31.25%	100%	3.5%	**1**/400	**252**/400	153.0 s	34.2 g
B4 — two-stage: B2 (low threshold) → CUSUM	0.25%	40%	26.75%	1.5%	59.5%	**223**/400	**387**/400	5.3 s	19.1 g
B5 — GLR: leak + admissible offset	0%	8.25%	13.25%	100%	24.25%	**64**/400	**320**/400	41.7 s	31.4 g
C — one-sided CUSUM pair (1 s blocks)	0.25%	37%	28%	0.75%	82%	**322**/400	**386**/400	4.4 s	10.6 g
A, calibrated at δ = 0 only (as in v1.4)	100%	(not re-evaluated in this row)				**280**/400	**397**/400	—	—

Leak onset uniform during the fill; mass lost for missed leaks is the whole leak. 2% / 5% of the hold flow, δ = 0. Full tables, ROC with three outcomes (any alarm / first alarm after onset / correct diagnosis), sensitivity and cost views: uniquesupernova.com/ja-i2cner/#/benchmark

4. What the preliminary results say (and do not say)

- negative **State-dependent allowance (B3) did not help:** 2% leak -8.75 pp (95% CI -16 to -2.25) versus B2.
- **Two-stage detection (B4) recovers small-leak detection** while keeping false alarms ≤ 0.25% inside the declared range: 2% leak +46.75 pp (95% CI 36.25 to 57.25) versus B2. **Cost:** a meter-bias step raises alarms in 59.5% of fills — B4 more readily reads a meter fault as a leak. CUSUM detects a change during the fill; with a leak present from the start of the fill, B4 detects only 16% of 2% leaks.
- **Calibration matters more than the detector:** calibrated at δ = 0 only, A detects 280/400 of 2% leaks but shuts down 100% of no-leak fills at a +0.2 s offset; calibrated over the declared range, A detects 0/400.
- **Failures are detector failures, not proven indistinguishability:** under a 1 s meter lag B2 alarms on 100% of no-leak fills, but B4 and C on 1.5% and 0.75%. (Correction of v1.4: under drift, B2 gives 35.75%, not "≈100%".)
- **Value of added measurements** (200 fills): an independent time reference (±0.02 s) lets A detect 122/200 of 2% leaks (C 182/200); a ±0.5% flowmeter (hypothetical) lets B2 detect 175/200; a faster thermometer helps less (B2 37/200 vs 21/200).
- **Meter error:** OIML R 139-1 gives 1.5% for the *meter* (class 2) and 2% / 3% for the *whole measuring system*. Applying 2% / 3% per meter is shown only as a stress case. A v1.4 bug (populations not regenerated) is fixed; with δ = 0 calibration B2 then detects 5% leaks in 109/200 and 69/200 fills (v1.4 reported the same 78.5% for every MPE). The v1.4 claim that conclusions do not change is withdrawn.
- **Not evaluated:** independent pressure/temperature confirmation (in this model a leak does not affect pressure). Weighting unnecessary shutdowns against misses (energy economics) is a question for the proposed host, not a social study.

5. Plan

8 weeks (report in week 8) or 12 weeks: confirm assumptions with experts at the proposed host (if agreed) → adapt the model → freeze code and thresholds (hash recorded) → a **blind test whose conditions would be set by the proposed host (if agreed)** → performance map, value of added measurements, failure cases → one technical report (code, data, uncertainty, failures). No connection to real equipment; no operator-confidential data.

6. Our ask

- Introduction to a researcher who could advise on the measurement configuration and on instrument response characteristics.
- Agreement, before evaluation, on the detection deadline and the minimum significant leak (the success criterion).
- A person (a student or researcher is fine) who could set blind-test conditions without telling the applicant.
- Discussion of hosting for a short stay (status, supervision, 8–12 weeks) and a review of the report's wording before publication.

Sources (verified 20–21 Oct 2026)

API RP 1130, Computational Pipeline Monitoring for Liquids, 2nd ed. (2022) · L. Zhang et al., Hydrogen Mass Balance of Hydrogen Refuelling Stations in British Columbia, NRC-EME-56557 (2023), doi:10.4224/40003216 · E. S. Page, Biometrika 41 (1954) 100–115 · A. Wald, Ann. Math. Stat. 16 (1945) 117–186 · M. Basseville, I. V. Nikiforov, Detection of Abrupt Changes, Prentice Hall (1993) · S. Särkkä, Bayesian Filtering and Smoothing, CUP (2013) · A. Raue et al., Bioinformatics 25 (2009) 1923–1929 · OIML R 139-1:2018 · J. W. Leachman et al., J. Phys. Chem. Ref. Data 38 (2009) 721–748 · NIST Chemistry WebBook SRD 69 · SAE J2601. Full list (91 sources, not 91 independent pieces of evidence) in the Japanese / Bulgarian documents.

Code (MIT): uniquesupernova.com/ja-i2cner/ · benchmark/ (mass_balance_clock.py, .mjs, expected_output.json, SHA256SUMS.txt).