Research Square

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

Cross-Regime False-Alarm Analysis and Conservative Residual Calibration for Nuclear Power Plant Transients with Limited Normal Data

Yuan Zhang
China Institute of Atomic Energy (CIAE)
Xingmin Liu
13671100478@139.com

China Institute of Atomic Energy (CIAE)
Shengguo Dong
China Institute of Atomic Energy (CIAE)

Research Article

Keywords: nuclear power plant, variable-power transients, residual-threshold mismatch, one-sided conservative calibration, false-alarm episodes, limited normal data

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11630036/v1

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

Additional Declarations: No competing interests reported.

Abstract

Nuclear power plants that adjust electrical output undergo normal thermal-hydraulic transients, yet residual-based fault monitors calibrated on a few operating histories can mistake these regime changes for anomalies. The engineering problem addressed here is the mismatch between a source-calibrated residual threshold and an unseen, still-normal operating regime under limited normal data. We evaluate a one-sided conservative commissioning residual-calibration procedure: a four-lag, multi-output Ridge model predicts the next standardized process state; an independent normal trajectory sets a 99th-percentile residual threshold; and the first 50 consecutive target samples, assumed independently verified as fault-free, estimate a median residual offset. Only a positive offset may raise the threshold. Using four normal and eight accident trajectories from the public PCTRAN nuclear power plant simulation dataset (3,808 recorded time points), we perform 12 role-disjoint whole-trajectory training, calibration and testing evaluations. On withheld normal monitoring segments, false-alarm samples decrease from 306/2,787 (10.98%) with the fixed threshold to 105/2,787 (3.77%) after conservative correction. However, distinct false-alarm episodes decrease only from 72 to 70; post-injection sample detection remains approximately 91.9%, with an 11 s median first-alarm delay. Requiring two successive exceedances reduces the corrected method's false-alarm episodes to 15 but increases median delay to 22.75 s. Moreover, one calibration/test split accounts for 193 of the 201 avoided false-alarm samples. The procedure therefore addresses sample-level false alarms arising from calibration mismatch in this dataset, but does not establish consistent event-level improvement, independent generalization, or protection-system performance.

1. Introduction

Changes in electrical load demand produce legitimate variations in nuclear plant pressure, temperature, flow and power. A residual-based fault monitor compares observed process values with predictions of normal behavior; however, a threshold calibrated on a small collection of normal trajectories may cease to represent the residual distribution in another normal operating regime. The resulting cross-regime threshold mismatch can generate false alarms even when the plant remains fault-free. This study addresses the specific engineering question of how to reduce those spurious residual exceedances with very limited normal data while checking whether fault detection or the number of distinct alarm events is adversely affected.

The publicly released Nuclear Power Plant Accident Data (NPPAD) dataset provides reproducible pressurized-water reactor transients produced with PCTRAN [1]. Work on anomaly detection under environmental variation, few-shot cross-domain adaptation, graph-conditioned density estimation and hybrid drift monitoring shows that operating-condition shifts have received considerable attention [2–5]. Zero-shot digital-twin approaches have also been explored [6]. In related thermal-power applications, conformal anomaly detection has been studied specifically for false-positive control [7], whereas formal conformal coverage for time-dependent or non-exchangeable data needs additional conditions [8]. Thus a neural predictor or a time-varying threshold alone does not establish methodological novelty. Recent

sensor-driven digital-twin monitoring and machine-learning transient prediction also provide complementary context [9, 10].

The critical difficulty is not solely inaccurate state prediction: even a useful predictor may yield a shifted residual baseline when the normal power-transition regime changes. Previous retrospective checks on these same 12 trajectories showed that a nominal 99th-percentile fixed threshold did not reliably transfer to a separate normal run. With two normal runs for model training, one different normal run for calibration and another for testing, the pooled false-alarm rate was 10.54%; a previously investigated local-quantile setting gave 11.81%. These results motivated the present research question: can a short, independently verified fault-free segment from the new operating run correct the source-to-target residual offset without requiring accident data for calibration?

To answer this question, we formulate and evaluate a one-sided conservative commissioning residual-calibration procedure. It (i) learns a one-step normal-process predictor from training-normal trajectories, (ii) calibrates a source threshold on a separate normal trajectory, (iii) estimates the target residual shift during a short externally verified normal commissioning window, and (iv) increases—but never decreases—the detection threshold when the target baseline is higher. We then assess the solution on whole-run held-out normal data and accident data using both sample-level exceedances and operationally distinct alarm episodes, with first detection delays and consecutive-hit confirmation as safety-relevant trade-offs. The individual ingredients are established techniques; the contribution is their explicitly constrained use and a reproducible evaluation of which part of the false-alarm problem is, and is not, resolved.

Prior studies cover nuclear fault diagnosis [11], power-transient recognition [12], kernel principal component analysis [13], conditional models [5], and uncertainty-aware monitoring [14]. They demonstrate that prediction errors, domain adaptation and adaptive thresholds are not new concepts by themselves. Rather than claim superiority over untested complex architectures, this paper isolates a specific failure mechanism—normal-regime residual miscalibration—and evaluates whether a simple nondecreasing threshold correction reduces its observed consequences. The distinction between fewer alarmed samples and fewer independent alarm episodes is central to that evaluation.

2. Materials and Methods

2.1. Simulation scenarios and labels

We analyze 12 trajectories from the public NPPAD variable-power extension: four normal power ramps (80→100%, 100→80%, 70→90%, and 90→70% of rated power) and eight accident trajectories (LOCA, rod withdrawal, and steam-generator A/B tube rupture under 80→100% and 100→80% ramps). Together, the trajectories comprise 3,808 recorded time points. Each record contains TIME and 96 process variables. The scenario manifest specifies the accident-injection time and is used for scoring without reference to model alarms. The simulations represent a PCTRAN pressurized-water reactor rather than field operation.

Some input records have discontinuous initial timestamps, for example 0 s immediately followed by 1,370 s. Prediction windows are accepted only if their four historical steps each correspond to a continuous 10 s sampling sequence. We do not synthesize missing records or bridge these gaps, and timestamps (including fractional offsets) are preserved. All data transformations and scenario mappings are distributed with the reproducibility package.

2.2. Monitoring assumption

The proposed calibration requires a clearly defined premise: an external procedure must establish that the first W = 50 consecutive valid residual-score samples of each monitored run are fault-free. During this commissioning interval, the adaptive detector issues no formal alarm. This creates a monitoring gap of approximately 500 s under continuous 10 s sampling, possibly longer when timestamps are discontinuous. In the public scenarios evaluated here, accident injection occurs after that interval. Consequently, the procedure can be examined for post-commissioning monitoring, but the data do not establish its behavior when faults arise during commissioning or contaminate the supposedly normal prefix.

2.3. Methods/Experimental: one-sided conservative residual calibration

2.3.1. Prediction residual score

The first step is to measure how far the observed process deviates from a normal prediction. Let x_t denote the 96-dimensional sensor vector. Means and standard deviations are fitted using training-normal runs only. A four-lag multi-output Ridge predictor (regularization α = 30) estimates the next standardized vector, and its root-mean-square residual defines an anomaly score; see Eqs. (1) and (2). Larger scores indicate departures from predicted normal dynamics, but a changed normal regime can also increase the score. Therefore a prediction score alone does not solve the cross-regime false-alarm problem. All predictor inputs are lagged measured signals, not a separately logged load-demand command.

$$\hat{z}_t = f_\theta\left(z_{t-4},\, z_{t-3},\, z_{t-2},\, z_{t-1}\right)$$

1

$$s_t = \sqrt{\frac{1}{96}\sum_{j=1}^{96}\left(z_{t,j} - \hat{z}_{t,j}\right)^2}$$

2

2.3.2. Whole-trajectory separation

To avoid mistaking within-run interpolation for transfer to a new regime, model fitting, source calibration and evaluation use different complete normal trajectories. For each of four normal runs, a distinct run is chosen as the source calibration set, forming 4 × 3 = 12 ordered test/calibration pairs; only the remaining

two normal runs are used to fit the scaler and Ridge model. Source residual scores determine the initial threshold; scores from the held-out test run assess normal-regime mismatch. Each of the eight accident simulations is tested only for a matching 80→100% or 100→80% transition and is reused under three calibration settings. Hence the 24 accident evaluations are repeated assessments of eight simulations, not 24 independent accident realizations.

2.3.3. Source calibration and one-sided commissioning correction

The second step corrects the particular error introduced by source-to-target residual miscalibration. The separate source calibration trajectory defines an initial threshold τ_0 at its 99th residual percentile (Eq. (3)). The first W = 50 consecutive, externally verified normal target scores form a commissioning set. The median difference Δ between target and source scores estimates a baseline residual shift (Eq. (4)). The corrected threshold $\tau_CS = \tau_0 + \max(0, \Delta)$ (Eq. (5)) is then kept fixed for the rest of the run. A positive shift raises the threshold to avoid labeling a higher, yet confirmed-normal residual baseline as a fault; a negative shift cannot lower it. This asymmetry is a conservative design restriction, not a guarantee of low false-alarm probability, and it may weaken sensitivity to low-amplitude faults. No fault labels enter threshold estimation.

$$\tau_0 = Q_{0.99}\left(S_{cal}\right) \tag{3}$$

$$\Delta = med\left(S_W\right) - med\left(S_{cal}\right) \tag{4}$$

$$\tau_{CS} = \tau_0 + max(0, \Delta) \tag{5}$$

The correction is compared with a fixed source threshold τ_0, an unrestricted two-sided median shift $\tau_MS = \max(0, \tau_0 + \Delta)$, and a warm-up maximum guard (Eq. (6)). All comparators use the same predictive model, source calibration choice, first-50-score exclusion window and monitored timestamps. Keeping the monitoring exposure identical is necessary to attribute observed alarm differences to the calibration rule rather than to unequal observation periods.

$$\tau_{MW} = max\{\tau_0, max(S_W)\}$$

6

2.3.4. Evaluation measures

We test whether the calibration solves the stated problem at three distinct levels. First, the fraction and count of alarmed samples on wholly withheld normal trajectories measure residual-threshold mismatch under changing normal regimes. Second, maximal sequences of consecutive 10 s alarm samples define separate false-alarm episodes, reflecting the number of alarm initiations rather than only their duration.

Third, accident tests report pre-injection alarms, post-injection alarmed-sample fraction, first-alarm delay relative to the manifest injection time, and detections within 300 s. Requiring r = 1, 2 or 3 consecutive exceedances provides a matched alarm-confirmation comparison. Since runs and their scores are reused across splits, the analysis is descriptive rather than an independent-sample significance test.

$$FPR = \frac{N_{false}}{N_{normal}}$$

7

3. Results

3.1. Main comparison

The primary question is whether correcting the source-to-target residual baseline decreases alarms during an unseen but normal operating transition. Under the 12 whole-trajectory evaluations, the fixed source threshold raised 306 alarms over 2,787 normal monitoring points (10.98%); the one-sided correction raised 105 (3.77%). Thus 201 alarmed time points were avoided under the same monitoring exposure. This sample-level change came with nearly identical post-injection sample detection (91.914% fixed versus 91.897% corrected), an 11 s median first-alarm delay for both methods, and 21 of 24 repeated accident evaluations detected within 300 s. The mean first-alarm delay increased slightly from 163.19 to 163.60 s. These results demonstrate a within-dataset reduction in normal residual exceedances, not proven reliability for independent future runs (Table 1; Fig. 1).

Table 1
Comparison of matched threshold calibration methods.

Threshold method	Withheld normal FPR	Pre-injection FPR	Post-fault recall	Median delay
Fixed source percentile	306/2787 (10.98%)	0.87%	91.91%	11 s
Two-sided median shift	234/2787 (8.40%)	1.09%	92.04%	11 s
One-sided conservative shift	105/2787 (3.77%)	0.87%	91.90%	11 s
Warm-up maximum guard	290/2787 (10.41%)	0.00%	91.84%	11 s

3.2. Per-regime disaggregation

The mechanism is most visible when the target and source normal regimes are strongly mismatched. For withheld NORM_70_to_90, alarms across three source calibrations decreased from 240/609 to 39/609; in the most severe NORM_90_to_70-to-NORM_70_to_90 calibration pair, they decreased from

203/203 to 10/203. Conversely, the correction did not change the observed alarm counts for NORM_80_to_100 (35), NORM_90_to_70 (31), or NORM_100_to_80 (0). Hence the method does not systematically improve every normal regime; its measured benefit is tied to a particular residual mismatch. Figure 2 displays only test/calibration pairs with at least one alarm; all-zero pairs remain included in the statistics.

Note

Zero-alarm test–calibration pairs are omitted from this plot only; they remain included in aggregate analyses.

3.3. Accident detection and delay

Reducing false alarms matters only if a change in threshold does not conceal important accident responses. We therefore compare the two thresholds using injection times read directly from the scenario metadata. Median first detection delays for LOCA, rod withdrawal, SGATR and SGBTR are, respectively, 3.25, 3.25, 75 and 70 s under both methods. Of 24 repeated accident evaluations, 23 had unchanged first-alarm times after correction and one was delayed by 10 s. This limited test suggests little change in first detection for the observed cases, but the longer steam-generator tube-rupture delays demonstrate why similar pooled detection rates cannot substitute for accident-specific assessments (Fig. 3).

3.4. Ablation and failure modes

The ablation tests explain why the direction constraint matters. An unrestricted downward median correction increased false alarms in one NORM_90_to_70 evaluation from approximately 13.4% to 67.7% by lowering the threshold, whereas the one-sided rule forbids such a decrease. Conversely, protecting the threshold only by the largest warm-up score failed to compensate for the severe NORM_70_to_90 mismatch. The data therefore support a narrower interpretation: adjusting for a positive baseline-location shift can alleviate some false alarms, but neither the chosen guard nor one-sided calibration provides reliable coverage across all regimes.

3.5. Alarm episodes and consecutive confirmation

The second question is whether fewer alarmed samples actually translate into fewer alarm events. A false-alarm episode is defined as a maximal run of alarms at consecutive 10 s timestamps; an alarm-free point or a timestamp discontinuity terminates the episode. We additionally test confirmation after r = 1, 2 or 3 successive threshold exceedances on valid consecutive observations. All variants use the same initial commissioning exclusion and identical test streams.

The answer differs sharply from the sample-level result. Although one-sided calibration reduced normal alarmed samples by 65.7%, it reduced distinct false-alarm episodes only from 72 to 70 under single-hit confirmation. This indicates that the correction often shortened sustained alarms rather than preventing

an alarm from starting. Requiring two consecutive exceedances reduced the corrected method's episode count from 70 to 15, but increased the median first detection delay from 11 to 22.75 s; three-hit confirmation reduces alarms further with additional delay (Table 2; Fig. 4). Thus calibration correction and alarm confirmation address different parts of the false-alarm burden, and an engineering assessment must report both.

Table 2
Alarm episodes and detection under matched confirmation rules.

Threshold	Hits	Normal alarm samples	False episodes	Postfault recall	Delay (s)	≤ 300 s
Fixed	1	306/2787	72	91.91%	11.00	21/24
Fixed	2	234/2787	14	91.41%	22.75	20/24
Fixed	3	220/2787	5	90.99%	32.75	20/24
Conservative	1	105/2787	70	91.90%	11.00	21/24
Conservative	2	35/2787	15	91.39%	22.75	20/24
Conservative	3	20/2787	5	90.98%	32.75	20/24

3.6. Domain-wise effect concentration and effective sample size

The apparent benefit must also be examined at the level of independent normal regimes. Across the four distinct withheld normal trajectories, fixed and corrected alarm totals are identical for 80→100% (35/519), 100→80% (0/963), and 90→70% (31/696), and differ only for 70→90% (240/609 versus 39/609). Furthermore, a single calibration/test pair accounts for 193 of the 201 avoided alarmed samples (96.0%). Accordingly, the pooled 10.98%-to-3.77% reduction is driven by one difficult normal regime and one especially mismatched calibration pair; it cannot be interpreted as a repeatable improvement across independent plants or runs (Fig. 5).

The equal-weight average of the four regime-specific false-alarm fractions is 12.65% before correction and 4.40% after correction. This descriptive macro average avoids allowing longer records to dominate the percentage but does not create new independent observations. Together with the nearly unchanged episode count, it reinforces the distinction between reducing the duration of false-alarm bursts and reliably preventing their initiation.

Note

The 100%→80% normal ramp is omitted from the chart because both methods produced zero false alarms.

3.7. Window-length sensitivity and the untested warm-up failure mode

To test sensitivity to the length of the externally verified normal prefix, we repeat the retrospective calculation for commissioning windows of 10, 20, …, 80 valid samples. Corrected withheld-normal false-alarm fractions range from approximately 3.00% to 3.97%, compared with 10.59% to 11.21% for the matched fixed-threshold evaluations. Since the number of monitored samples changes with window length, those percentages must be read with their corresponding exposure denominators. More importantly, all eight available accident runs inject faults only after the 50-sample commissioning interval. The present experiments therefore do not answer whether the calibration remains safe if the prefix contains a fault or whether it can detect an accident during that interval.

3.8. Operational interpretation

The proposed procedure should be interpreted as an auxiliary monitoring algorithm for addressing normal-regime threshold mismatch, not as a replacement for reactor protection or plant interlocks. Its target-window adaptation requires independent confirmation of a fault-free prefix; the predictor and residual score cannot certify that assumption. A separately validated monitoring or protection layer would still be required during the commissioning gap. No early-injection dataset is available here to test that operational arrangement (Fig. 6).

4. Discussion

4.1. Originality relative to the literature

The problem-solution chain is deliberately narrow: sparse normal histories produce a source threshold that can be too low for a different but fault-free regime; a verified normal prefix provides a target residual baseline; a one-sided median-offset rule prevents downward threshold movement and can mitigate some resulting exceedances. Prior research has already studied condition-aware monitoring, transfer learning and drift adjustment [2–5]; therefore the simple Ridge predictor and median operation are not claimed as new mathematical primitives. The engineering contribution of this case study is the explicitly bounded calibration procedure and its joint, auditable evaluation in terms of alarm samples, alarm episodes and accident delay. The observed event-level limitation remains a substantive result, not a claim to be concealed.

4.2. Causal and operational assumptions

This solution has a non-negotiable assumption: the initial target scores must be verified as fault-free independently of the detector. Otherwise an early fault could raise the estimated baseline and the detection threshold, potentially suppressing later warnings. The 50-score blind period is therefore unsuitable as a stand-alone safety-protection scheme. In addition, the available model uses lagged measured variables, not a separately logged electrical-load command; it does not solve the distinct

problem of fault-contaminated regime identification. These conditions define what the proposed calibration does and does not address.

4.2.1. False-alarm samples are not false-alarm episodes

The experiment resolves an important evaluation ambiguity. A reduced false-positive sample fraction does not necessarily imply fewer notifications to an operator, because the same prolonged alarm episode can contain many time points. The correction mainly affects the number of alarmed samples, whereas consecutive-hit confirmation is more effective at suppressing short events but incurs detection delay. Future monitoring studies should therefore report sample false-alarm fraction, episode initiation count, episode duration and fault-specific latency under matching monitoring exposure.

4.3. External validity and future validation

Only four independent normal trajectories and eight independent accident trajectories are available. The 12 whole-run role assignments and 24 accident evaluations repeatedly reuse those histories, and the procedure was developed through retrospective inspection of this dataset. Consequently, these results support transparent within-dataset comparisons but not independent generalization or a statistical false-alarm guarantee. Independent normal regimes, varied initial conditions and fault injections during commissioning would be required to test whether the observed correction remains effective. No such new simulator experiments are claimed here.

5. Conclusions

This study addresses cross-regime false alarms caused by residual-threshold mismatch when only a few normal nuclear power-plant transients are available. It evaluates a one-sided conservative commissioning residual-calibration procedure that combines a normal-data predictor, an independent source quantile, and an externally verified target-normal median offset. On twelve public PCTRAN trajectories, normal alarmed samples declined from 306/2,787 (10.98%) to 105/2,787 (3.77%), while distinct false-alarm episodes declined only from 72 to 70; post-injection detection remained approximately 91.9%. Two-hit confirmation reduced the corrected episode count to 15 at the cost of increasing median first detection delay from 11 to 22.75 s. Because 193 of 201 avoided alarm samples came from a single calibration/test pair, the observed benefit is regime-dependent and does not demonstrate independent reliability. In principle, the procedure could be adapted to other reactor types, including fast reactors, because it is based on measurable time-series residuals and a verified normal calibration window rather than pressurized-water-reactor-specific equations. Such transfer would require reactor-specific variable selection, model retraining, threshold calibration, safety-constraint review and independent validation; no cross-reactor performance is established by this study.

Abbreviations

NPP

nuclear power plant
NPPAD
Nuclear Power Plant Accident Data
PCTRAN
PC—based Transient Analyzer
PWR
pressurized—water reactor
CIAE
China Institute of Atomic Energy
LOCA
loss—of—coolant accident
RW
rod withdrawal
SGATR
steam—generator A tube rupture
SGBTR
steam—generator B tube rupture
FPR
false—positive rate
CSV
comma—separated values
KPCA
kernel principal component analysis
CS
conservative shift
MS
median shift
MW
warm—up maximum.

Declarations

Competing Interests

No competing interests.

Ethical Declaration

Not applicable.

Consent to Participate

Not applicable.

Consent to Publish

Not applicable.

Clinical Trial Number

Not applicable.

Funding

This research received no external funding.

Author Contribution

Yuan Zhang was responsible for study conceptualization, methodological design, data curation and interpretation, preparation of figures and tables, and drafting and revision of the manuscript. Yuan Zhang and Xingmin Liu are responsible for reviewing and approving the submitted version.

Acknowledgements

Not applicable.

Availability of Data and Materials

The NPPAD simulator dataset and accident metadata are publicly accessible from the original dataset repository: https://github.com/thu-inet/NuclearPowerPlantAccidentData (Qi et al., 2022; reference [1]). Derived score tables, split definitions, and analysis scripts may be obtained from the corresponding author on reasonable request. No additional independent simulator runs were generated for this study.

References

1. Qi B, Xiao X, Liang J, Po L-C, Zhang L, Tong J (2022) An open time-series simulated dataset covering various accidents for nuclear power plants. Sci Data 9:766. https://doi.org/10.1038/s41597-022-01879-1

2. Kim S, Zhang F (2025) A robust anomaly detection system for nuclear power plants under varying environmental conditions and malfunction levels. Nucl Technol 211(8):1625–1644. https://doi.org/10.1080/00295450.2024.2422234
3. He J, Zheng S, Yi S, Yang S, Huan Z (2025) Cross-domain few-shot anomaly detection for equipment in nuclear power plants. Nucl Eng Des 436:113956. https://doi.org/10.1016/j.nucengdes.2025.113956
4. Li J, Guo J, Guo C, Zhang T, Huang X (2025) Online anomaly detection for nuclear power plants via hybrid concept drift. Energies 18(17):4491. https://doi.org/10.3390/en18174491
5. Zhang L, Cheng W, Zhang S, Xing J, Chen X, Gao L et al (2024) Spatial-temporal graph conditionalized normalizing flows for nuclear power plant multivariate anomaly detection. IEEE Trans Industr Inf 20(11):12945–12957. https://doi.org/10.1109/TII.2024.3431027
6. Qi B, Sun J, Sui Z et al (2025) A transient detection framework in nuclear power plants using zero-shot learning based on digital twins. Prog Nucl Energy 189:105848. https://doi.org/10.1016/j.pnucene.2025.105848
7. Kundacina O, Vincan V, Gojic G, Ninkovic V, Miskovic D (2025) Conformal anomaly detection for predictive maintenance in thermal power plants. IEEE Access 13:39738–39752. https://doi.org/10.1109/ACCESS.2025.3546451
8. Oliveira RI, Orenstein P, Ramos T, Romano JV (2024) Split conformal prediction and non-exchangeable data. J Mach Learn Res 25(225):1–38. https://jmlr.org/papers/v25/23-1553.html
9. Liu S, Ye G, Zhao X (2025) Edge temporal digital twin network for sensor-driven fault detection in nuclear power systems. Sensors 25(24):7510. https://doi.org/10.3390/s25247510
10. Tang Y, Liu T, Chang Y, Liu Z, Peng G (2025) Transient parameter prediction and fault diagnosis for nuclear power plants based on machine learning. J Nucl Sci Technol 62(10):1006–1022. https://doi.org/10.1080/00223131.2025.2522347
11. Wang Z, Wei H, Tian R, Tan S (2025) A review of data-driven fault diagnosis method for nuclear power plant. Prog Nucl Energy 186:105785. https://doi.org/10.1016/j.pnucene.2025.105785
12. Mendoza M, Tsvetkov PV (2023) An intelligent fault detection and diagnosis monitoring system for reactor operational resilience: Power transient identification. Prog Nucl Energy 156:104529. https://doi.org/10.1016/j.pnucene.2022.104529
13. Wang H, Peng M-J, Yu Y, Saeed H, Hao C-M, Liu Y-K (2021) Fault identification and diagnosis based on KPCA and similarity clustering for nuclear power plants. Ann Nucl Energy 150:107786. https://doi.org/10.1016/j.anucene.2020.107786
14. Li J, Zhang T, Zheng S, Xu J, Lin M, Tan S, Tian R (2026) Reliable fault diagnosis framework for nuclear power plants based on Bayesian neural network and uncertainty analysis. Nucl Eng Des 448:114707. https://doi.org/10.1016/j.nucengdes.2025.114707

Figures

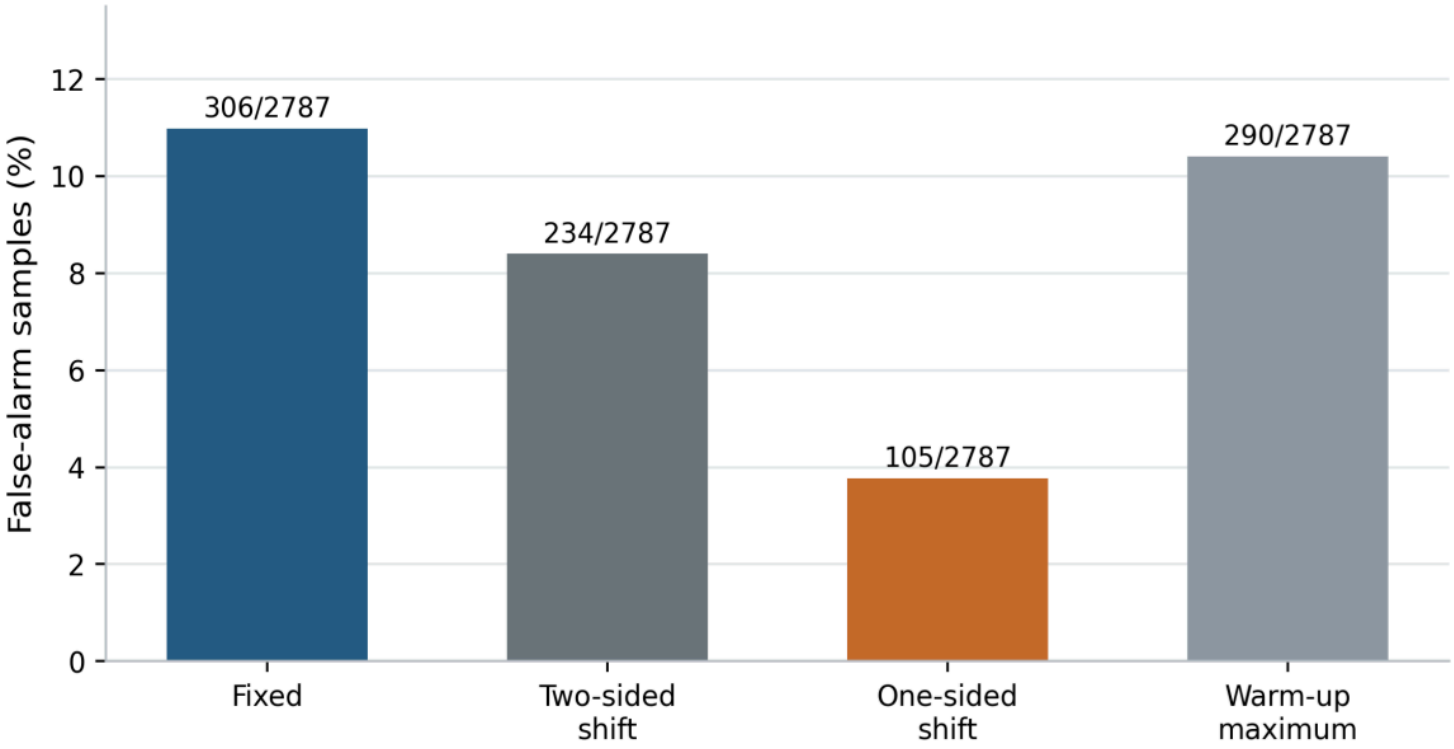

Figure 1

Comparison of false-alarm sample rates under four threshold rules.

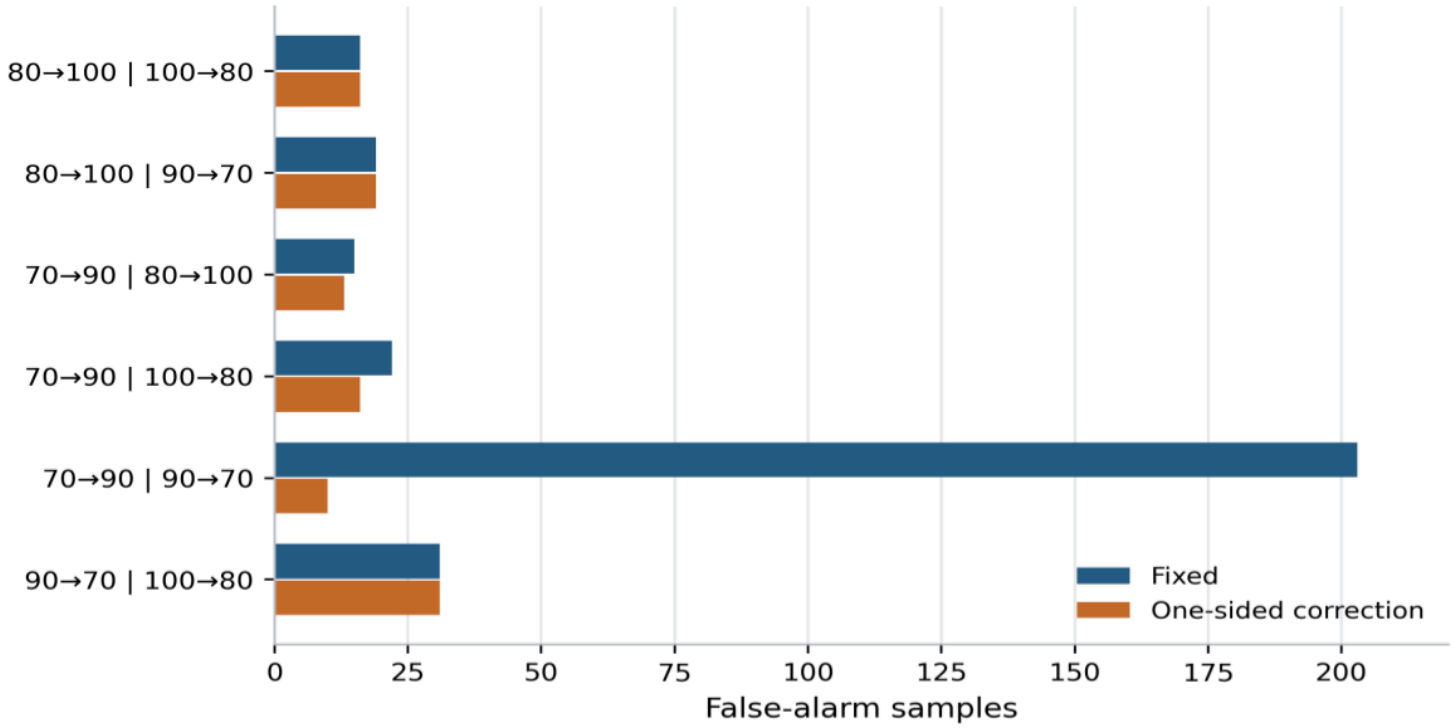

Figure 2

False-alarm samples for nonzero normal test–calibration pairs.

Note: Zero-alarm test–calibration pairs are omitted from this plot only; they remain included in aggregate analyses.

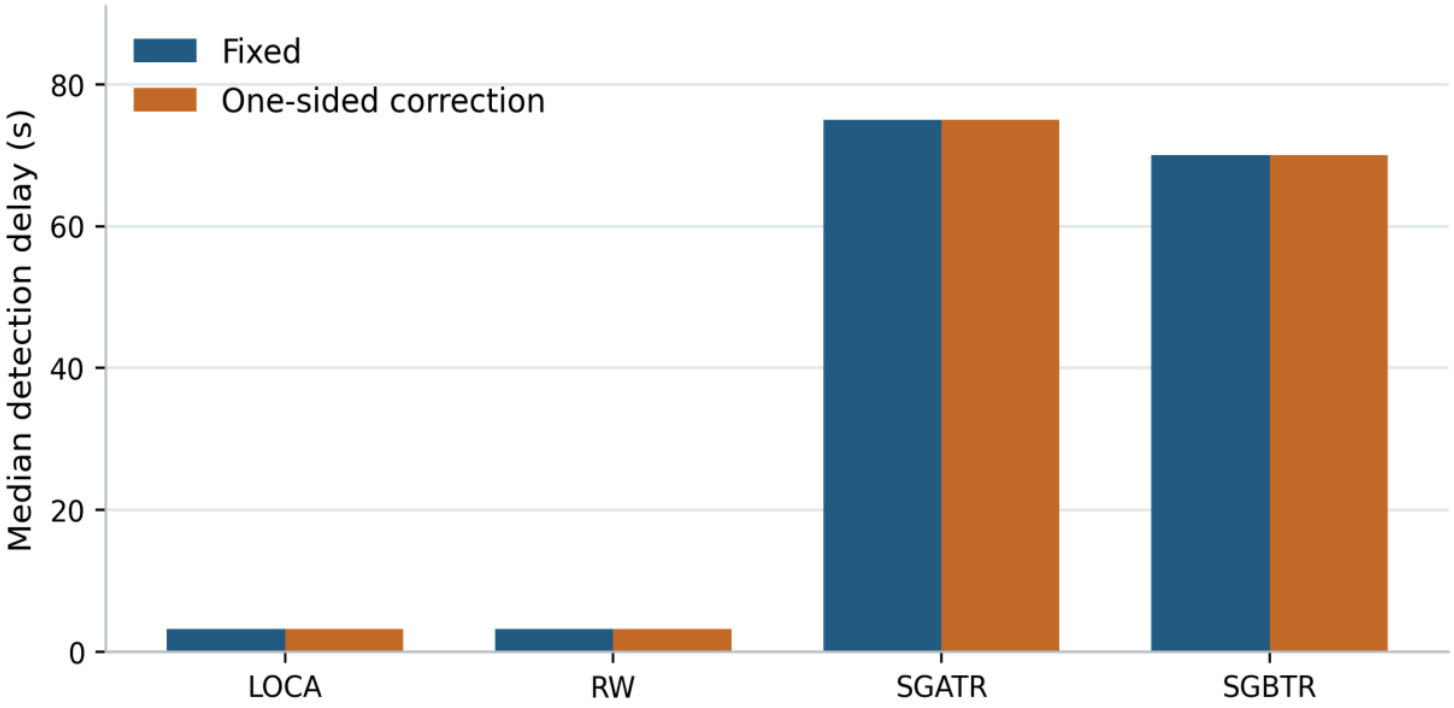

Figure 3

Median accident detection delay by fault category.

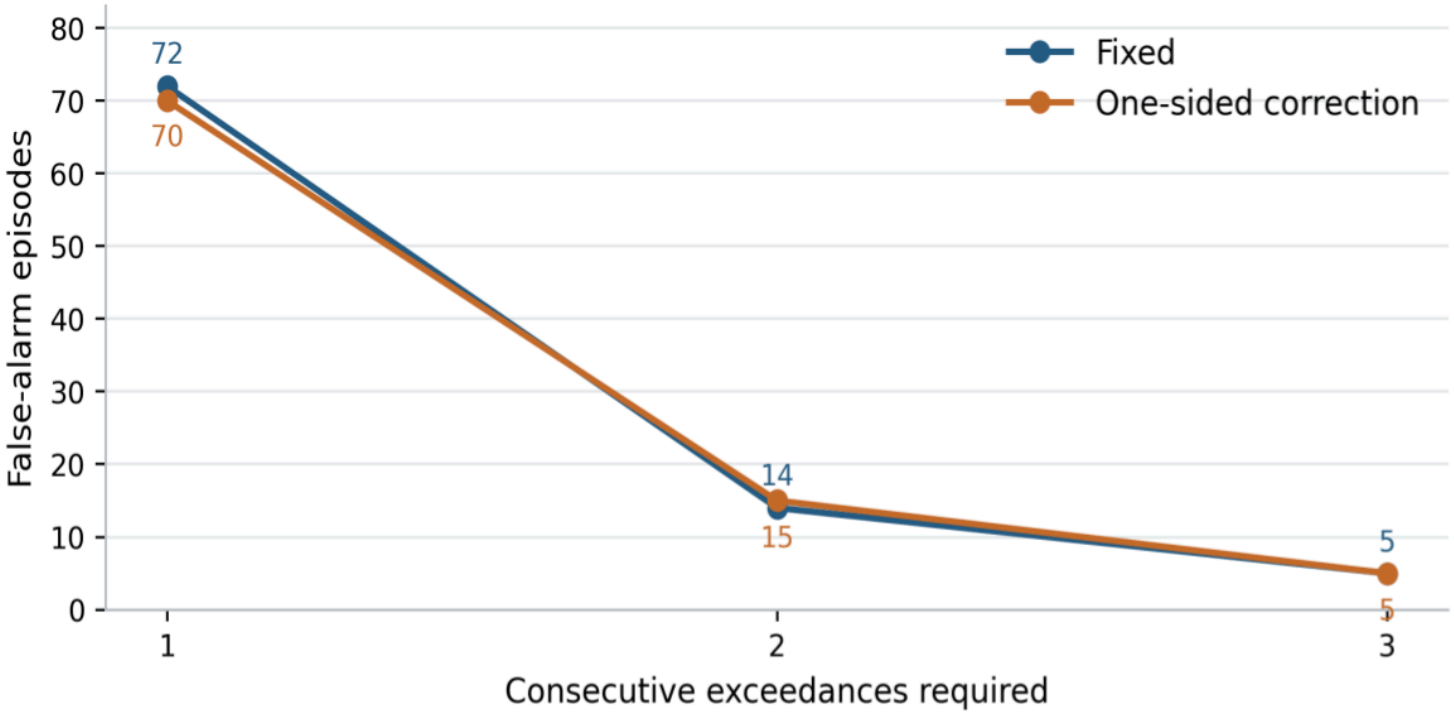

Figure 4

False-alarm episodes under consecutive-hit confirmation.

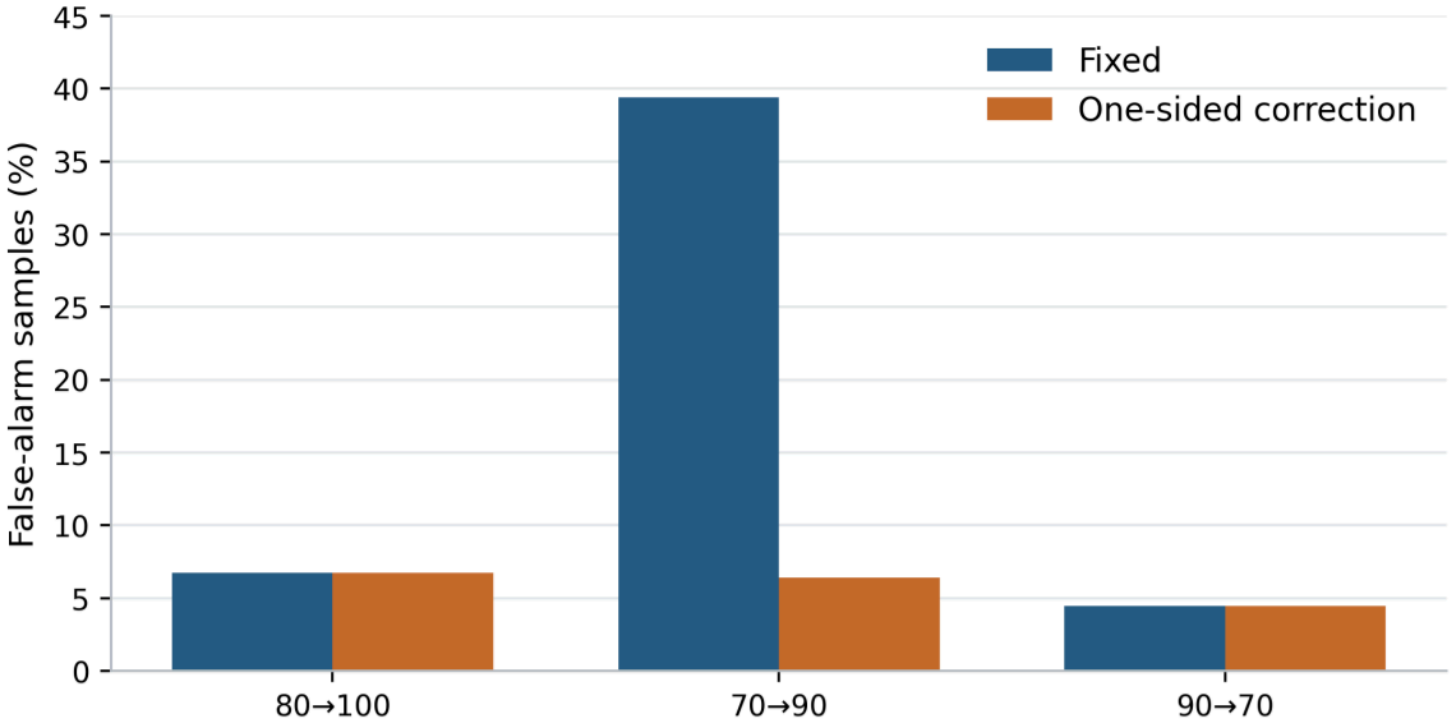

Figure 5

False-alarm rates by withheld normal power-transition regime.

Note: The 100%→80% normal ramp is omitted from the chart because both methods produced zero false alarms.

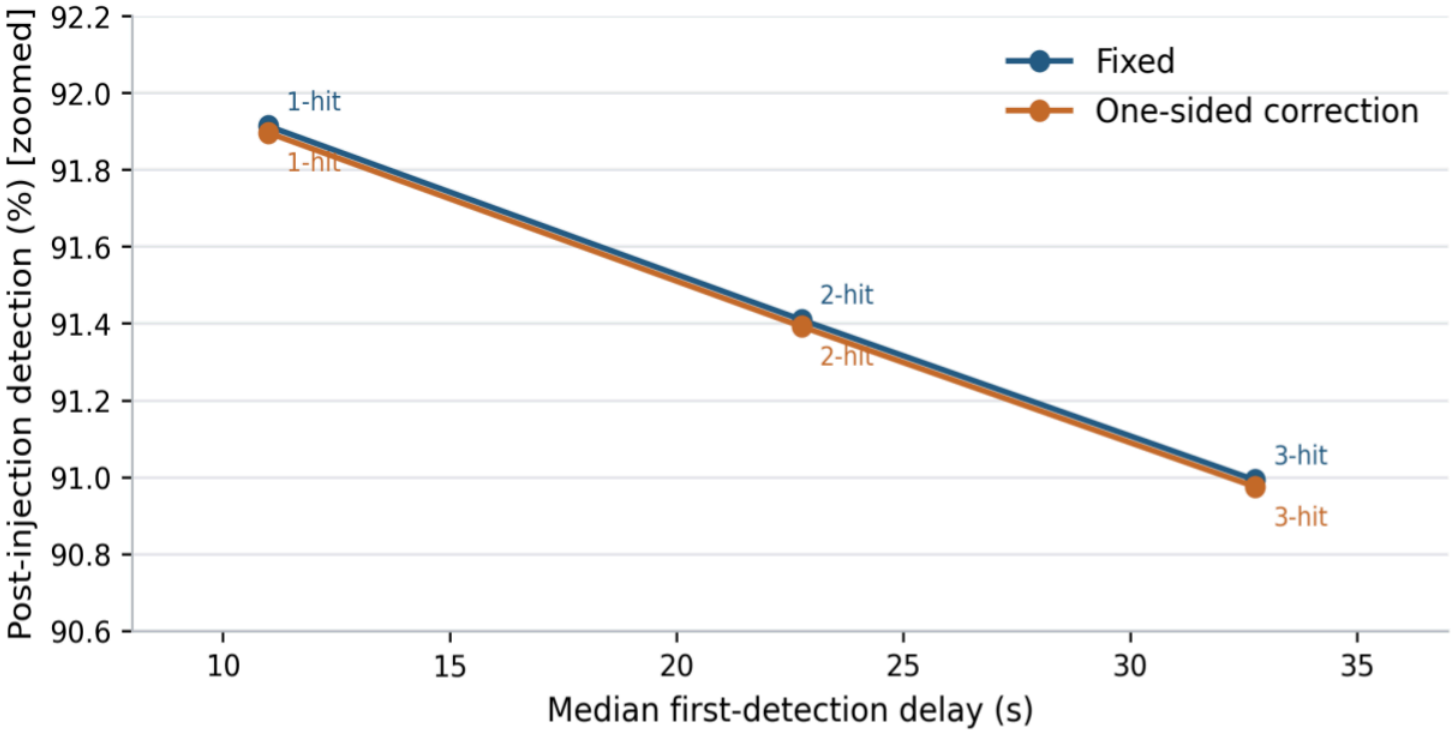

Figure 6

Detection sensitivity and delay under confirmation rules.