

From SCADA Pressure to Actionable Alarms: Calibration-Aware Monitoring of Kaplan Turbine Runner-Blade Actuation Friction

Xianzheng Wang^{1,2}, Shengxiang Chen^{1,2,*}, Shan Liu^{1,2}, Jianming Xia^{1,2}, Bingxue Zhang^{1,2}, Qian Wang^{1,2}, Lu Zhang^{1,2},
Wanjia Xie^{1,2}

¹ Sichuan Shengda Hydropower Development Co., Ltd., Chengdu 610041, China

² China Hydropower Construction Group Santa Hydropower Co., Ltd., Leshan 614000, China

First author: Xianzheng Wang, scslhwangxzh@powerchina.cn; Tel: +86-13981385311

Corresponding author: Shengxiang Chen, 80370693@qq.com; Tel: +86-15983385608

Received 12 July 2026; Revised 18 August 2026; Accepted 7 September 2026; Published online 20 September 2026

Abstract. Runner-blade friction and lubrication wear in a Kaplan turbine stay hard to observe directly between costly hub inspections. We introduce KFM-Cal, a unit-adaptive pressure-to-alarm algorithm that converts installed one-minute close/open servo-pressure SCADA into persistent friction and lubrication-risk alarms. The algorithm keeps three stages distinct: mechanism-mapped extraction of pressure evidence, target-normal adaptation of a frozen source ranker's operating point, and causal proportion-voting alarm formation. Target adaptation draws on a short lubricated reference segment and holds target fault labels out of training. Our tests pair two public long-duration Kaplan records with 365 days of field SCADA from Sichuan Shawan. The T1 same-unit route reaches 0.987 MDF balanced accuracy at a 0.018 false-alarm rate, and T2-to-T1 transfer reaches 0.782 balanced accuracy at a 0.005 false-alarm rate. On the field record the frozen monitor catches all three engineering events, and MDF lowers routine false alarms from 0.030872 to 0.000441 with a mean delay of 30.96 h. Route-wise ablation, grouped SHAP, block permutation, and deep-model audits single out pressure and trend as the main evidence and quantify route-specific adaptation. The results establish a calibratable SCADA route from conditional net-actuation evidence to maintenance-facing alarms.

Keywords: Kaplan turbine; SCADA monitoring; friction diagnosis; target-normal calibration; moving decision filter

1. Introduction

Kaplan turbines keep high efficiency over a wide operating range by coordinating the guide vanes and runner blades. A hydraulic servo moves the runner blades through the oil receiver, hub linkage, and blade trunnion bearings. Years of active-power regulation wear these mechanical interfaces and thin the lubrication film that protects the blade bushings. The added resistive effort then weakens regulation reliability, speeds secondary damage, and complicates maintenance planning. Direct assessment demands runner-hub disassembly, so the friction and lubrication state stays out of direct view between major inspections.

A practical route reuses signals the plant already records. Hydropower SCADA systems commonly store the average, minimum, and maximum close-side and open-side servo pressures at one-minute resolution. These extrema keep the pressure swing during blade motion: growing resistance can widen the chamber-pressure range, shift the close/open imbalance, and add persistent drift. The measurements also carry hydrodynamic, inertial, damping, dispatch, and sensor effects. Pressure thus serves as conditional evidence of net actuation effort and friction/lubrication risk. Figure 1 maps the engineering route from installed SCADA to an operator-facing alarm.

Three research gaps stand between pressure classification and deployable monitoring. First, the public glycerol indicator forms one long contiguous run per state, so random row or overlapping-window mixing spreads near-adjacent observations across training and test and inflates apparent generalization. Second, pressure varies with hydrodynamic load, inertia, damping, dispatch, and sensor history, so a useful representation must preserve the actuation mechanism while it accounts for operating context. Recent diagnosis advances underline this temporal-structure requirement: dynamic predictive modeling tracks planetary-transmission behavior under time-varying speed, and

standardized collaborative multiscale Manhattan entropy characterizes nonlinear train-bearing dynamics across scales [1, 2];

a Kaplan pressure route needs the same temporal sensitivity expressed through servo-pressure observables. Third, a ranker trained on one turbine meets a shifted score distribution on another, which moves the alarm operating point even while useful risk ordering survives; plant maintenance additionally acts on persistent alarm episodes, whereas individual window crossings supply intermediate evidence.

KFM-Cal answers these gaps. It learns an interpretable risk ordering from mechanism-bearing pressure evidence, adapts its operating point to a new unit through trusted normal operation, and forms persistent alarms with an explicit budget over false alarms, coverage, and delay. The closest runner-blade study established the pressure-friction premise, the public benchmark, established anomaly detectors, and a strict moving decision filter [3, 4]. KFM-Cal advances that route with a unit-adaptive algorithm that coordinates mechanism-mapped representation, target-normal empirical calibration, and causal proportion voting under a leakage-aware evaluation protocol.

The separation matters in practice. The source model ranks risk, target-normal calibration sets the unit-specific operating point, and MDF governs alarm persistence. Each stage then becomes measurable through the quantity it controls: ranking through PR-AUC and balanced accuracy, calibration through false alarms and risk coverage, and persistence through alarm episodes and detection delay. KFM-Cal thus forms a unit-adaptive decision operator around modular rankers, in which evidence ordering, operating-point relocation, and alarm formation keep distinct inputs, parameters, and verification targets.

The paper makes three algorithmic contributions to the pressure-monitoring route.

1. **Mechanism-grounded pressure evidence extraction.** Five coordinated groups encode chamber level, close/open difference, within-minute range, reference-relative trend, and multiscale energy and entropy. The representation ties each observable to the runner-blade torque balance and answers the

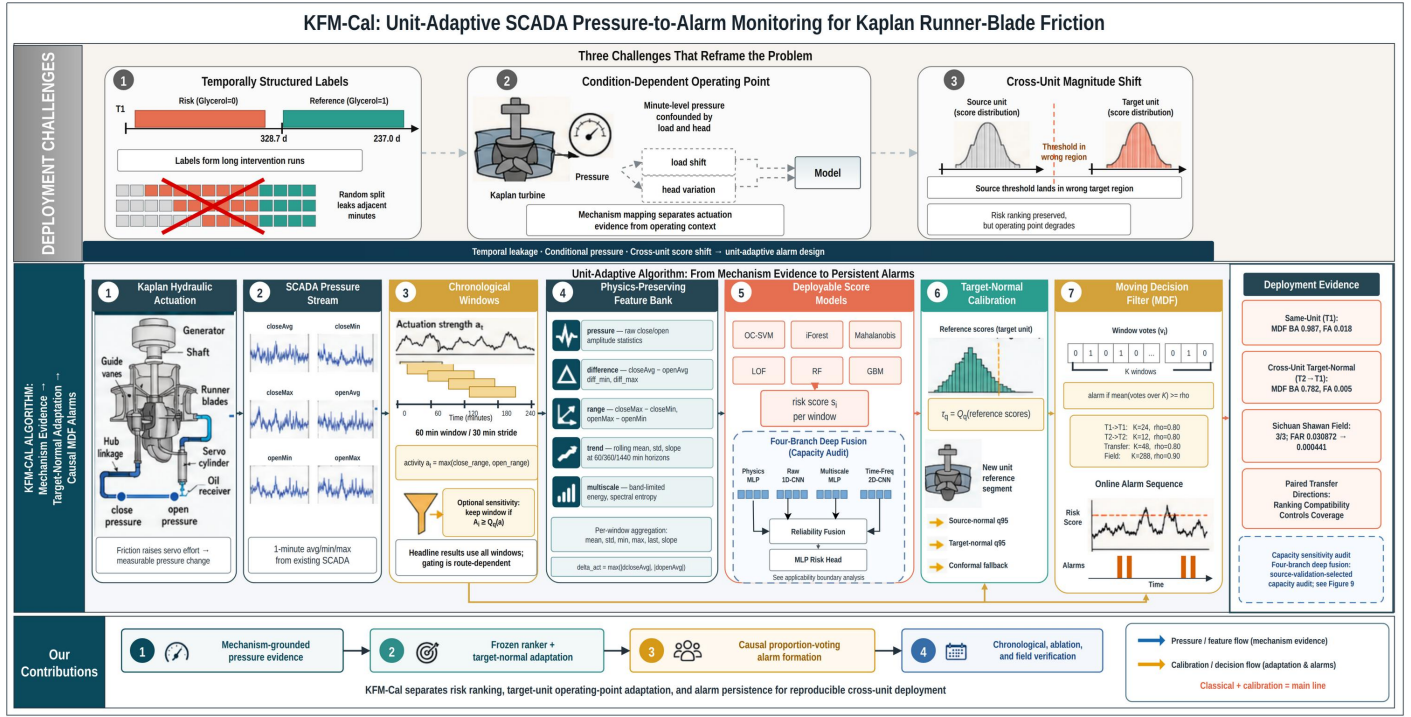


Fig. 1: KFM-Cal unit-adaptive monitoring framework. Installed close/open servo-pressure SCADA enters leakage-aware chronological windows and five mechanism-mapped evidence groups. A modular source ranker orders window risk, a short target-normal reference relocates the operating threshold, and the causal MDF converts repeated crossings into persistent maintenance alarms. The lower evidence rail summarizes the validated same-unit, cross-unit, and annual-field operating points; activity gating and deep capacity appear as route-specific sensitivity analyses.

conditional-observability gap; route-wise removal, grouped SHAP, and estimator-faithful block permutation verify its diagnostic contribution.

2. **Target-normal operating-point adaptation.** A short chronological lubricated-reference segment relocates the threshold of a frozen source ranker while target fault labels stay held out for evaluation. This adaptation answers the cross-unit score-shift gap; identical-score counterfactuals isolate the gain from threshold relocation and give the empirical quantile a direct commissioning and post-maintenance reading.
3. **Causal proportion-voting alarm formation.** A $\lceil pK \rceil$ -of- K rule generalizes strict consecutive voting and separates score threshold, persistence horizon, and vote ratio. It closes the distance between window crossings and maintenance-grade persistent episodes; its causal construction turns window scores into event-level alarms with a measurable budget over false alarms, coverage, and delay.

The evaluation builds evidence in deployment order: protocol validity, pressure observability, same-unit and cross-unit operating points, the calibration and persistence mechanisms behind them, annual field behavior, component contributions, and deep-model capacity. Two long public turbine records support the chronology, same-unit, cross-unit, ablation, and attribution tests. A read-only 365-day record from the Sichuan Dadu River Shawan Hydropower Station provides held-out field evidence. A guarded deep-model audit reports tuning, regularization, and training-size behavior and places model capacity beside the calibrated classical route.

2. Related Work and Contribution Scope

2.1 Pressure evidence for hydraulic actuation faults

Sandstrom et al. tied runner-blade lubrication state to close/open servo pressure, compared established anomaly detectors, and used a moving decision filter to form alarms [3]. Their companion dataset holds two long Kaplan SCADA records with glycerol presence as a lubrication proxy [4]. Guide-vane work likewise combines differential pressure and vane position to monitor distributor-mechanism friction [5]. These studies establish the physical premise that actuation pressure carries mechanical-condition information. KFM-Cal adds an explicit torque-balance reading, chronological leakage control with a disclosed boundary overlap, target-unit calibration from trusted normal operation, and route-wise component evidence.

Hydropower condition monitoring has drawn on KICA-PCA [6], moving-window PCA with Bayesian isolation [7], composite health indices [8], minimum-spanning-tree anomaly distance [9], and autoencoders [10]. Reviews connect useful plant alarms to measurement semantics and maintenance decisions [11, 12]. Cavitation studies show the value of vibration, acoustic emission, pressure pulsation, and SCADA under different sensing constraints [13–15], while hydraulic-governor models show that installed pressure and control variables reveal actuation anomalies [16]. For runner-blade friction, close/open pressure carries the mechanism-bearing evidence and operating variables define the context.

2.2 Dynamic representation, transfer, and alarm formation

Section 1 cited dynamic predictive modeling under time-varying speed and standardized collaborative multiscale Manhattan entropy as recent exemplars of temporal-structure-aware

diagnosis [1, 2]. KFM-Cal addresses a complementary deployment problem. It maps minute-level Kaplan servo pressure to friction/lubrication-risk evidence, adapts the operating point through target-normal data, and evaluates persistent operator alarms. Its trend and multiscale groups carry the shared insight that temporal structure matters, while the torque-balance mapping, chronological protocol, and route-wise attribution tie that structure to pressure observability.

KFM-Cal accepts one-class SVM [17], isolation forest [18], local outlier factor [19], Mahalanobis distance, random forests, and gradient boosting as modular score generators. Domain-adversarial and multi-source methods learn aligned representations through target-domain access during training [20–22], while transfer-learning reviews organize domain adaptation, domain generalization, and knowledge-guided transfer under sparse fault-event data [23]. KFM-Cal exploits a practical operational asymmetry: commissioning and maintenance readily provide a lubricated target reference, while target fault evidence stays scarce. The algorithm joins frozen source ranking, target-normal quantile relocation, and causal proportion voting into a unit-adaptive pressure-to-alarm map. Identical-score counterfactuals verify threshold relocation, and event-level counterfactuals verify persistence, giving the methodological contribution a direct training, commissioning, and online-decision interface.

3. Method

Figure 2 summarizes the physical-to-computational chain. The balance model maps pressure observables to combined actuation effort. Chronological windows yield mechanism-grounded features, modular rankers yield frozen scores, target-normal calibration sets the unit-specific threshold, and MDF determines the persistent alarm state.

3.1 Physical motivation and monitoring target

Let θ denote the runner-blade/linkage angle and ℓ the lubrication state. We use the simplified balance in Eq. (1):

$$J_{eq}\ddot{\theta} + c_{eq}\dot{\theta} + \tau_{fr}(\theta, \dot{\theta}, \ell) + \tau_{hyd}(\theta, Q, H) = K_{\tau}\Delta p, \quad (1)$$

$$\Delta p = P_{close} - P_{open}$$

In this balance, J_{eq} and c_{eq} collect equivalent inertia and damping, τ_{fr} is the friction term affected by lubrication, τ_{hyd} is the load-, flow-, head-, and blade-angle-dependent hydrodynamic torque, and K_{τ} is the effective pressure-to-torque gain. Minute extrema aggregate the combined actuation effort over each sampling interval. Signed and absolute differential pressure describe net hydraulic effort; pressure range and short slopes describe motion-related transients; long residuals describe persistent drift after reference and operating context are considered. Equation (1) therefore provides an observability model for conditional friction/lubrication-risk evidence at the available sensing rate.

The expected signatures overlap because each pressure sample reflects the combined balance. Hydrodynamic torque primarily shifts signed chamber imbalance with the operating point. Inertia appears as short impulses or slopes during commands, and damping links pressure range to motion rate. Friction and lubrication changes produce persistent resistance-related shifts once measured operating context is conditioned, while sensor bias and dispatch changes create alternative level shifts that held-

out normal references can assess. The feature bank preserves this structure and evaluates stability across reference, event, and unit conditions.

The public records include a glycerol indicator g_t , where $g_t = 1$ denotes the lubricated reference. The positive monitoring label in Eq. (2) is aligned with the alarm condition:

$$r_t = 1 - g_t. \quad (2)$$

Thus, $r_t = 1$ denotes the higher friction/lubrication-risk state. Glycerol status acts as the benchmark state proxy, while Eq. (1) connects the pressure response to the actuation balance.

3.2 Leakage-aware chronological protocols

Each public label run is divided chronologically into 60% training, 20% validation, and 20% test blocks. Windows contain 60 min and advance with a 30-min stride. Chronological assignment keeps each long state run ordered and removes direct random mixing across blocks. The implemented boundary retains 29 shared raw minutes between the last window of one block and the first window of the next, which defines the disclosed raw-overlap scope of the protocol. Model, feature, threshold, and MDF choices use training and validation evidence, while target-test labels remain reserved for final evaluation.

Three deployment routes follow. Same-unit calibration trains and validates on one turbine before its held-out chronological test. Cross-unit calibration fits the ranker on a source turbine and evaluates it on another. Target-normal calibration uses the first chronological 20% of target-normal windows as an upstream reference, with the remaining target chronology yielding test metrics. The headline results use the all-window population, and activity gating forms a separately labeled route sensitivity.

3.3 SCADA semantics and activity sensitivity

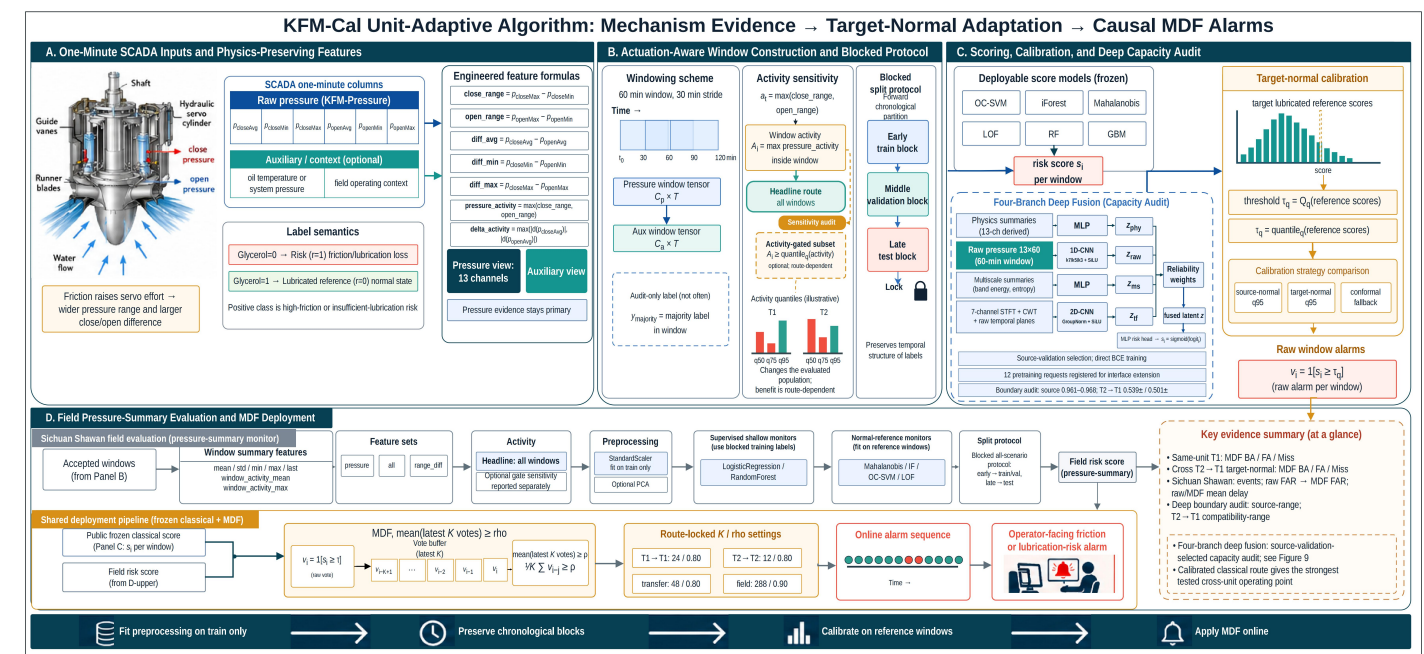
Each minute contains the average, minimum, and maximum close-side and open-side pressures. The extrema preserve the excursion within a minute, while the close/open difference preserves the signed chamber imbalance. For activity sensitivity, Eq. (3) defines the larger chamber range:

$$a_t = \max(p_{closeMax,t} - p_{closeMin,t}, p_{openMax,t} - p_{openMin,t}). \quad (3)$$

A window is retained when its peak a_t exceeds a threshold estimated from source-training raw minutes. We test q0.75 and q0.90 thresholds. This gate concentrates actuation-rich windows and simultaneously changes the evaluated population. Retained counts and matched-window controls therefore accompany every gate result and separate evidence concentration from population composition.

3.4 Mechanism-mapped feature bank

Each window yields five feature groups. **Pressure** retains close/open levels, extrema, and robust summaries. **Difference** contains signed and absolute close/open imbalance, including reference-relative residuals. **Range** contains chamber excursions, peak activity, and short-term regulation variability. **Trend** contains rolling levels, variability, slopes, and exponentially weighted residuals over 60-, 360-, and 1440-min horizons. **Multiscale** contains high-, medium-, and low-frequency energy



records 12 pretraining requests as compatibility entries. Stage A3 freezes the top two configurations per source before source-test or target-test scoring. Selection uses source-validation MDF balanced accuracy under a 0.05 false-alarm budget, followed by event, miss, and PR-AUC tie-breaks.

The audit runs seeds 0, 42, and 123; chronological training fractions of 10%, 25%, 50%, 75%, and 100%; weight decay of 0, 10^{-5} , 10^{-4} , 5×10^{-4} , and 10^{-3} ; and dropout of 0.15, 0.25, 0.35, and 0.50. Across the four frozen configurations, learning rate spans 3×10^{-4} to 10^{-3} , batch size 64 to 128, training length 12 to 48 epochs, patience 5 to 7, latent width 32 to 64, branch dropout 0 to 0.30, and label smoothing 0 to 0.06. Together, these settings specify the architecture, selection rule, regularization ranges, and compatibility conditions for the 122 guarded jobs; Section 4.8 and Figure 9 report the corresponding capacity audit. The main paper uses the audit to position deep capacity relative to the frozen calibrated route.

4. Experiments and Results

The experiments move from evaluation validity to deployment evidence, and each contribution from Section 1 meets its test here. Section 4.2 audits the chronological protocol that every later result inherits. Section 4.3 establishes pressure observability, Sections 4.4-4.5 measure the deployment operating points together with the calibration and persistence mechanisms behind them, and Section 4.6 follows the frozen monitor through a full field year. Section 4.7 attributes performance to individual components and feature groups, and Section 4.8 bounds the deep-fusion alternative. The mechanism-mapped representation is verified in Sections 4.3 and 4.7, target-normal adaptation in Sections 4.4-4.5, and proportion-voting alarm formation in Sections 4.5-4.6.

4.1 Datasets, protocols, and metrics

Table 1 summarizes the three evidence sources. Public Turbine 1 (T1) contains 814,512 one-minute rows over 565.7 days: the $g = 0$ run contributes 473,280 rows and the $g = 1$ run contributes 341,232 rows. Public Turbine 2 (T2) contains 1,052,493 rows over 731.0 days, with 817,740 $g = 0$ rows and 234,753 $g = 1$ rows. Each state forms one contiguous run. Within every run, the first 60% is assigned to training, the next 20% to validation, and the final 20% to held-out chronological testing. The field source contains 525,600 one-minute rows over 365 days from Sichuan Shawan. Its 14-day fit/calibration reference is separated from two held-out 7-day references, three engineering event intervals, and routine operation. Field source files are treated as read-only, and event annotations are evaluation evidence only.

Balanced accuracy (BA) averages sensitivity and specificity. False-alarm rate (FAR) is the fraction of reference windows alarmed; miss rate is the fraction of risk windows remaining below the alarm decision. PR-AUC evaluates ranking under class imbalance. We also report false-alarm episodes, event coverage, within-event alarm recall, and delay from the annotated event-interval start. Standard deviations use independent seeds. Overlapping public windows and the three field events receive descriptive metrics matched to their evidence structure.

Selection and evaluation occupy separate chronological roles. Public model and threshold choices use source training and validation or the designated target-normal reference; the final

chronological test blocks yield held-out metrics. Field threshold and MDF choices use the fit/calibration reference, while event outcomes remain reserved for evaluation. The frozen monitor remains the field headline, and the feature-group comparison forms a post-hoc normal-reference sensitivity. This separation preserves the three engineering events as independent field evidence.

4.2 Protocol audit: chronological blocking as the evaluation basis

Figure 3 establishes the evaluation protocol. The label audit shows one long intervention run per state, so random row or window mixing distributes close temporal neighbors across partitions and inflates apparent generalization. Chronological blocking assigns early history to model development and later history to evaluation, making source and target roles explicit.

The protocol result concerns evaluation validity. Random-split accuracy serves as a leakage diagnostic, while chronological results represent forward deployment more closely. The disclosed 29-min adjacent-boundary overlap defines the raw-sample scope, and independent-seed or descriptive uncertainty reporting follows that structure. Every later experiment inherits the same frozen chronology.

4.3 Actuation-aware pressure observability

With the protocol fixed, the next step is to establish that installed pressure carries condition evidence at all. Figure 4 develops the feature-led physical argument. Pressure difference separates the two public states, range describes regulation activity, trend variables expose persistent changes, and multiscale descriptors separate fast response from slow drift. The unit offset visible in feature and PCA space explains the cross-unit score shift that target-normal calibration addresses.

The feature landscape establishes observability and route-dependent gate value. Under $q0.90$ gating, the T1 same-unit test retains 14 windows and reaches BA 0.50. T2 same-unit and T2-to-T1 reach higher BA on their gated populations, while matched-window controls attribute much of the apparent change to retained-window composition. The all-window population therefore forms the formal headline routes, and Section 4.7 quantifies gating as a sensitivity.

4.4 Same-unit and cross-unit deployment performance

Observability established, the deployment question follows: what operating points does the frozen algorithm reach on its own unit and on a new one? Figure 5 and Table 2 report the four frozen all-window seed-0 routes. Same-unit T1 monitoring uses trend features, a GBM ranker, source-normal $q95$, and MDF (k, p) = (24, 0.80); it reaches BA 0.987252, FAR 0.017575, and miss rate 0.007921 (Fig. 5a). T2 same-unit uses the same feature/model/calibration family with (k, p) = (12, 0.80) and reaches BA 0.822110, FAR 0.068966, and miss rate 0.286815. The spread between the two same-unit points motivates unit-specific operating-point assessment across a fleet.

Cross-unit transfer demonstrates directional operating-point adaptation. T2-to-T1 uses trend features, an OCSVM ranker, target-normal $q95$, and (k, p) = (48, 0.80). Useful ranking combines with threshold relocation to reach FAR 0.005056, BA 0.782113, and risk-window coverage 0.569282. T1-to-T2 uses difference features, iForest, target-normal $q95$, and the same MDF setting; its conservative zero-FAR point reaches BA 0.509098, miss rate 0.981804, and risk-window coverage

Table 1: Evidence sources and evaluation boundaries. Public labels form long contiguous runs; field calibration, reference, routine, and event evidence keep separate roles.

Source	Rows / duration	State or event structure	Calibration/reference evidence	Held-out evaluation evidence
Public T1	814,512 / 565.7 d	g = 0: 473,280 rows; g = 1: 341,232 rows; 2 runs	Chronological 60% train and 20% validation within each run	Final chronological 20% within each run
Public T2	1,052,493 / 731.0 d	g = 0: 817,740 rows; g = 1: 234,753 rows; 2 runs	Chronological 60% train and 20% validation within each run	Final chronological 20% within each run
Sichuan Shawan field	525,600 / 365.0 d	3 engineering event intervals and routine operation	14-d fit/calibration reference	Two held-out 7-d references, 3 events, and routine periods

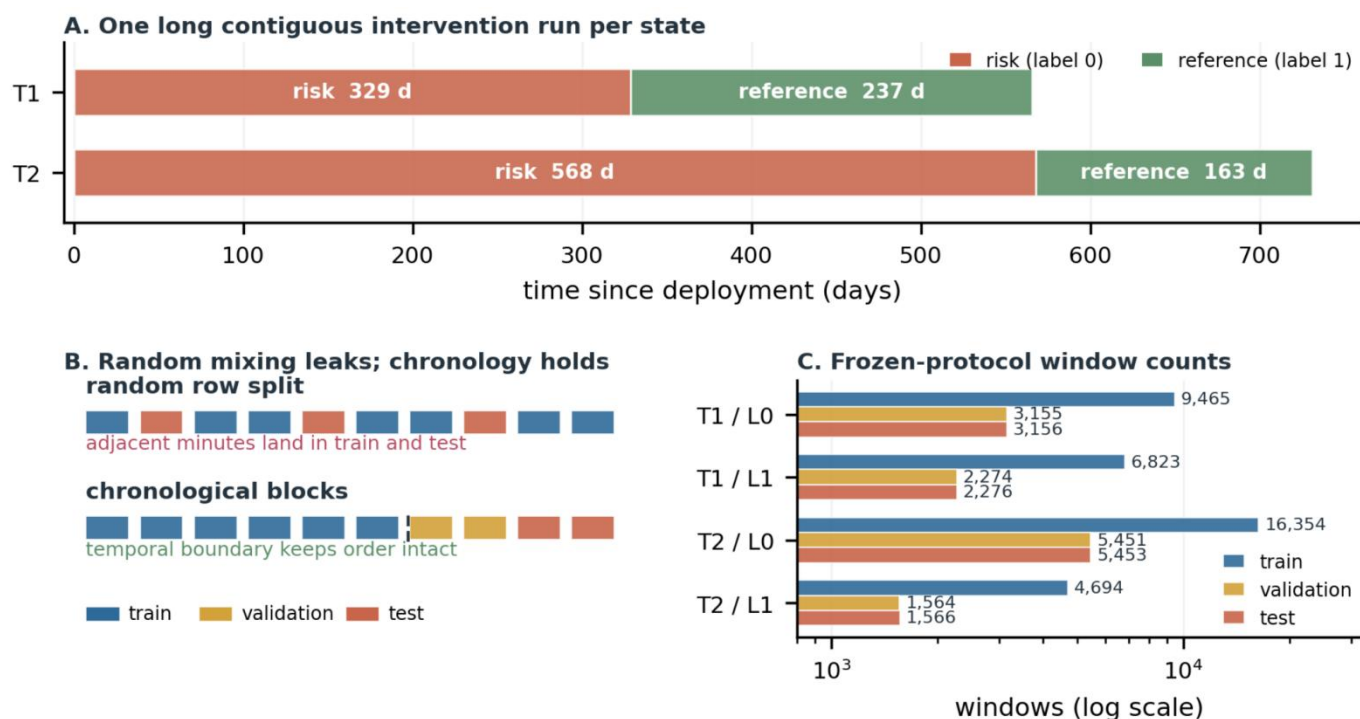


Fig. 3: Public-data protocol audit. Panel A shows the long contiguous glycerol-state runs; Panel B contrasts random mixing with forward chronological assignment; Panel C reports the resulting training, validation, and test window counts. All reported routes use 60-min windows with a 30-min stride and retain a disclosed 29-min overlap at adjacent block boundaries. The figure establishes block-level chronology as the common evaluation basis for every later route.

Table 2: All-window seed-0 public operating points. Threshold and MDF settings are frozen by route; the conservative T1-to-T2 point is retained as the transfer-compatibility boundary.

Route	Feature / model	Threshold reference	MDF (K, p)	BA	FAR	Miss
T1->T1	Trend / GBM	Source-normal q95	(24, 0.80)	0.987252	0.017575	0.007921
T2->T2	Trend / GBM	Source-normal q95	(12, 0.80)	0.822110	0.068966	0.286815
T2->T1	Trend / OCSVM	Target-normal q95	(48, 0.80)	0.782113	0.005056	0.430718
T1->T2	Difference / iForest	Target-normal q95	(48, 0.80)	0.509098	0.000000	0.981804

0.018196. Figure 5b makes the direction explicit: threshold relocation alone moves T2-to-T1 from FAR 0.117 at three-seed BA 0.869 to FAR 0.005 at BA 0.782, roughly a 23-fold reduction in alarm load for a modest ranking cost, while T1-to-T2 moves from FAR 0.063 to the conservative zero-FAR point. Fig-

ures 5c-5d place both main-line picks on the Pareto frontier near the low-FAR corner. The paired routes establish two complementary requirements for new-unit deployment: target-normal calibration aligns the score operating point, and source-target ranking compatibility determines usable coverage.

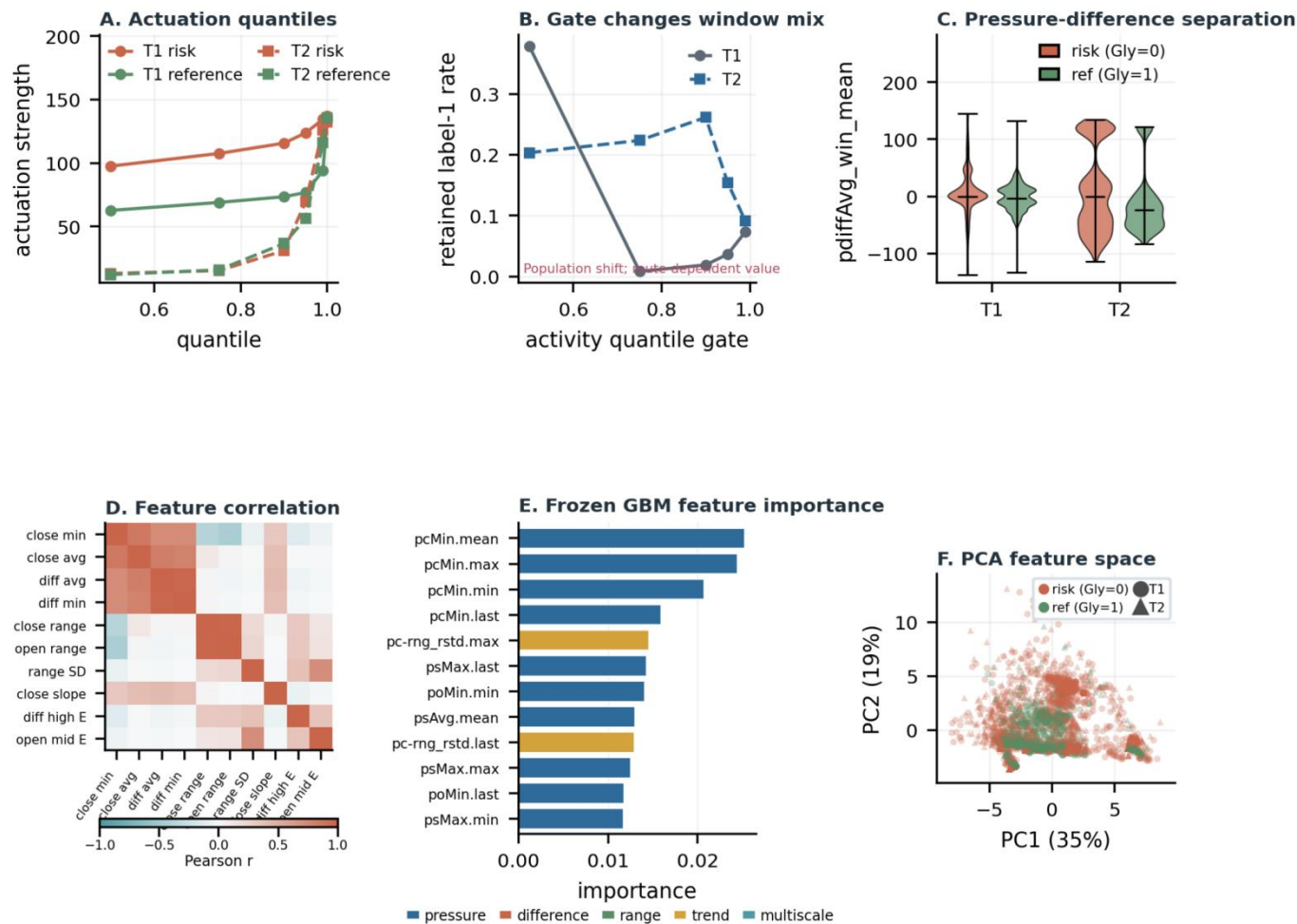


Fig. 4: Actuation and feature landscape. Panels A-B quantify activity strength and the population composition retained by activity gates. Panel C shows state separation in pressure difference; Panel D maps feature-family correlation; Panel E reports frozen GBM importance; Panel F visualizes state and turbine geometry. Together, the panels support the five mechanism-grounded evidence groups and reveal the unit-specific offset that motivates target-normal operating-point adaptation.

4.5 Mechanism analysis: threshold relocation and causal persistence

Table 2 reports what the operating points are; this section isolates why they move. Figure 6 follows identical frozen scores through source calibration, target-normal calibration, and MDF. For T2-to-T1, the source q95 threshold gives raw FAR 0.2718 and miss rate 0.0797. Target-normal q95 relocates the threshold and gives raw FAR 0.0898 with miss rate 0.3209. The frozen (48, 0.80) MDF then gives FAR 0.0051, one false-alarm episode, and miss rate 0.4307. Since scores and test population remain fixed, this sequence isolates operating-point relocation followed by causal persistence as the mechanism.

The reciprocal route maps the same calibration-persistence sequence under a different score ordering. Target-normal q95 lowers raw FAR from 0.2885 to 0.0492, and MDF reaches the conservative zero-FAR setting. The accompanying coverage trajectory quantifies the ranking-compatibility budget. Reference-based standardization and CORAL produce direction-specific operating points under the same frozen-score protocol. These comparisons position ranking compatibility and calibration as distinct design variables.

4.6 Annual field validation at Sichuan Shawan

The public routes establish the algorithm; the field record tests its yearly operating behavior. The field route is evaluated on real hydropower-station SCADA from Sichuan Shawan. The source files are read-only. A 14-day normal reference supports fitting and calibration; two 7-day normal references, routine periods, and three engineering event intervals are held out. The frozen score threshold is 2.481642 and the MDF uses $(k, p) = (288, 0.90)$. Figure 7 and Table 3 preserve the original annual-dashboard role: the full-year score, event zooms, first-alarm timing, reference stability, and alarm-load trade-off are read as one operational story.

Field annotations specify event-interval starts, so delay is measured from each annotated start; Table 3 lists the event-by-event outcomes. For the early friction increase on days 120-125, raw and MDF delays are 5.70 and 32.88 h, with event recalls 0.776 and 0.414. For lubrication loss with intervention on days 225-233, the corresponding delays are 1.62 and 10.55 h and recalls are 0.946 and 0.819. For sensor drift plus light friction on days 325-328, delays are 0.67 and 49.45 h and recalls are 0.501 and 0.119. The raw monitor covers 3/3 events with mean delay 2.6611 h, routine FAR 0.030872, and

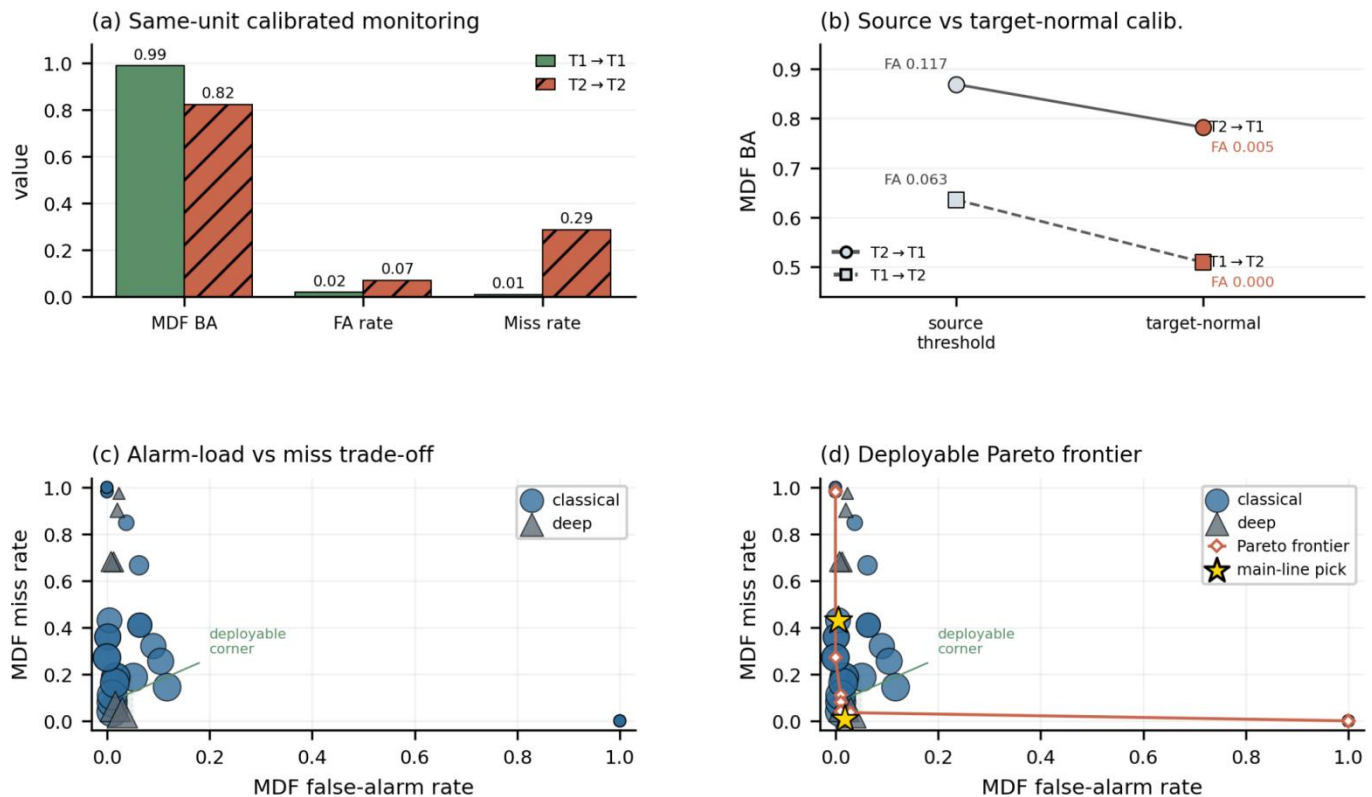


Fig. 5: Main deployment results across protocols. (a) Same-unit calibrated MDF operating points for T1 and T2. (b) Source-threshold counterfactual versus target-normal calibration for both transfer directions; the three-seed source points move to the frozen target-normal points. (c) Alarm-load versus miss trade-off across the frozen classical configurations and deep-audit runs, with the low-FAR deployable corner shaded. (d) The same plane with the Pareto frontier and the two main-line operating points.

held-out-reference FAR 0.008680. MDF also covers 3/3 events and reduces routine FAR by about 70-fold to 0.000441, reaches zero held-out-reference FAR, and gives mean delay 30.9611 h. These delays remain within the day-scale development horizon of the annotated degradation intervals and express the intended persistence budget.

A normal-reference sensitivity compares field selectivity across feature groups. The difference-group F2 detector covers 3/3 events with mean MDF delay 24.98 h, routine FAR 0.000455, and zero held-out-reference FAR. The all-five-group F9 detector also covers 3/3 events with routine FAR 0.0961 and held-out-reference FAR 0.2104. The frozen monitor remains the preregistered headline; F2 identifies a compact representation for prospective validation.

Physical-proxy checks support the conditional interpretation. Absolute differential pressure has event Cliff's delta 0.342 and maximum held-out-normal drift 0.014; multiscale energy has event delta 0.824 and maximum normal drift 0.141. Peak activity and the context-conditioned residual show normal-reference drift approaching their event effects. Differential pressure is the most stable field proxy in this comparison, while the other variables add diagnostic context across the combined actuation balance.

4.7 Component ablation and feature contribution

The deployment and field results rest on several coordinated components, and this section separates their individual effects. Figure 8 combines component ablation and feature contribution. The counterfactuals cover activity gating, trend, multiscale

descriptors, target calibration, and MDF. On T2-to-T1, the source-threshold counterfactual changes three-seed mean BA from about 0.782 to 0.869 and FAR from about 0.005 to 0.117. Raw decisions give BA 0.795, FAR 0.090, and miss rate 0.320. Joint reading of BA and alarm load shows how target calibration and MDF move the route toward the low-FAR maintenance objective.

Feature contribution is equally route-dependent. Under q0.90 T2-to-T1, trend-removed and multiscale-removed routes reach BA 0.50, while the corresponding q0.90 T2 same-unit routes reach BA 0.940 and 0.954, above the gated all-group route. The q0.90 T1 same-unit population contains 14 test windows and therefore defines a narrowly scoped sensitivity. Matched-window comparisons separate population selection from the gate operation.

Figure 8 therefore retains route and population labels throughout the analysis. Calibration and MDF counterfactuals use all windows and hold scores fixed, while gate and feature-removal rows use the explicitly marked q0.90 population. This design supports a precise conclusion: trend and multiscale structure provide essential evidence for one transferred route, compact alternatives for another route, and a sensitivity estimate when aggressive gating yields a small test population.

Across 12 route-seed units, grouped SHAP assigns median importance fractions of 0.5668 to pressure, 0.3762 to trend, and 0.0420 to multiscale features. Estimator-faithful 24-h block permutation ranks trend first with median PR-AUC drop 0.0621, followed by pressure at 0.0226 and multiscale at 0.0133. The

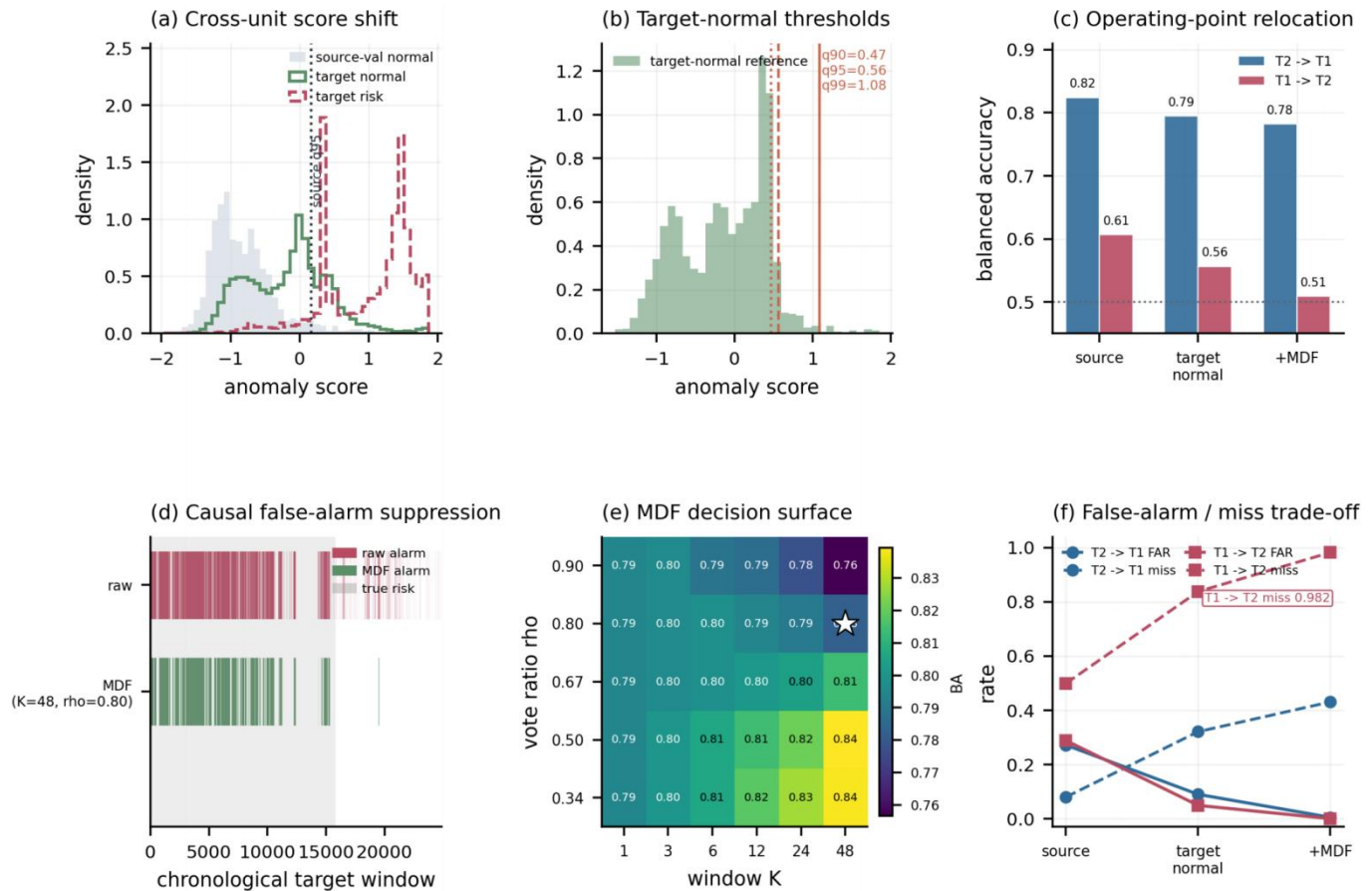


Fig. 6: Target-normal calibration and MDF mechanism. Panels A-B show the source-target score shift and the target-normal quantiles. Panel C traces BA as the operating point moves; Panel D shows how causal voting consolidates repeated threshold crossings; Panel E maps the (k, p) decision surface; Panel F aligns FAR and miss-rate trade-offs for both transfer directions. T2-to-T1 demonstrates low-FAR adaptation with useful coverage, and the paired direction defines ranking compatibility as a source-selection criterion.

Table 3: Event-by-event results for the frozen field monitor. Delay is measured from each annotated event-interval start, the reference point the field annotations provide.

Event interval	Days	Raw delay (h)	MDF delay (h)	Raw event recall	MDF event recall
Early friction increase	120-125	5.70	32.88	0.776	0.414
Lubrication loss with intervention	225-233	1.62	10.55	0.946	0.819
Sensor drift plus light friction	325-328	0.67	49.45	0.501	0.119

rankings differ because SHAP uses a route tree model or fixed tree companion, whereas permutation acts directly on the frozen main estimator. All five OCSVM drops are zero on the T2-to-T1 explanation slice, indicating saturation of that explanation slice and motivating the paired use of SHAP and estimator-faithful permutation.

4.8 Deep-fusion capacity audit

A remaining design question is whether added model capacity changes the transfer picture. Figure 9 reports the four-branch deep-fusion audit after source-only selection. The two selected T2 configurations reach source-test MDF BA 0.961 ± 0.005 and 0.968 ± 0.014 , while T2-to-T1 transfer reaches 0.539 ± 0.056 and 0.501 ± 0.006 . The source-to-target contrast identifies representation adaptation and multi-unit diversity as the pri-

mary requirements for deep cross-unit deployment at the available sensing rate.

Training-size sensitivity follows a peaked pattern. T2-to-T1 transfer rises from 0.526 ± 0.042 at 10% history to 0.636 ± 0.059 at 50%, then reaches 0.540 ± 0.058 at 75% and 0.540 ± 0.055 at 100%. Weight-decay and dropout sweeps preserve the same unit sensitivity across the recorded configurations. For this registry, minute-level sensing, two public units, and the recorded search budget, the physically structured calibrated route achieves the stronger transfer operating point; higher-rate multimodal data and broader multi-unit training provide the natural expansion path for deep fusion.

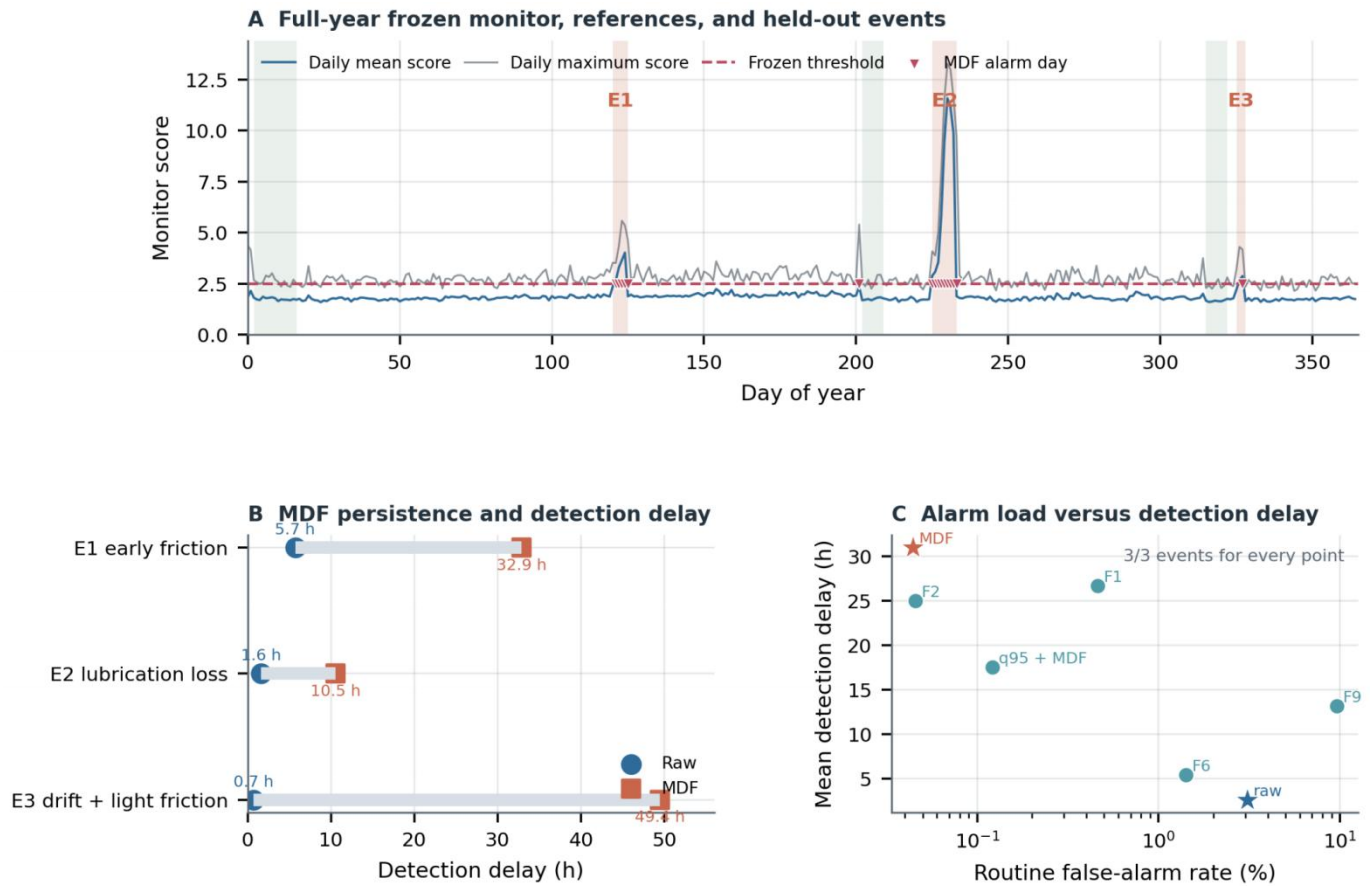


Fig. 7: Annual Sichuan Shawan field evidence. Panel A places the frozen score, threshold, two held-out references, and three engineering events on the full-year timeline. Panel B reports first-alarm delay for each event, and Panel C relates routine alarm load to mean delay across decision settings. Both raw and MDF routes cover all three events; MDF changes routine FAR from 0.030872 to 0.000441 and held-out-reference FAR from 0.008680 to 0, with mean delay changing from 2.66 to 30.96 h. Event annotations remain reserved for held-out engineering evaluation.

5. Discussion

KFM-Cal contributes a unit-adaptive monitoring algorithm built on modular risk rankers. Its novelty lies in the coordinated adaptation layer: mechanism-grounded evidence defines what the ranker sees, target-normal empirical calibration defines where a new unit operates on the score axis, and causal proportion voting defines when repeated risk becomes a maintenance alarm. This separation turns established classifiers into interchangeable ranking backbones and gives the full method a reproducible training, commissioning, and online-decision interface.

Pressure is useful because it sits in the runner-blade actuation path, and Eq. (1) explains its conditional meaning. Public attribution emphasizes pressure and trend; the field sensitivity identifies a selective difference-group detector; peak activity and context residuals also vary during held-out normal periods. These results form a coherent physical picture. Each route observes a different mixture of net actuation effort, operating context, and sensor history, while mechanism mapping anchors the alarm semantics to friction/lubrication risk.

Target-normal calibration and MDF serve distinct deployment functions. Calibration sharply reduces cross-unit false alarms when source ranking transfers, and MDF converts isolated crossings into persistent alarm episodes. Their operating parameters determine risk coverage and delay. The paired transfer direc-

tions show why a new unit requires joint inspection of ranking quality, threshold location, persistence, reference quality, and maintenance horizon. T2-to-T1 demonstrates the intended low-FAR adaptation, while the reciprocal route turns source-target ranking compatibility into an explicit source-selection criterion.

This separation also defines the theoretical contribution. Target-normal calibration uses the empirical target-normal score distribution to relocate a frozen ranker's operating point and remains invariant to monotone score reparameterization. The causal MDF generalizes strict consecutive voting through independent threshold, horizon, and vote-ratio controls. Together they provide a documented adaptation interface, frozen selection rule, and measurable alarm budget that can be reproduced on another unit and compared directly with statistical thresholding or representation-alignment alternatives.

The field record tests the route over long operation. Three events are covered with a low routine alarm load after MDF, and the two held-out references remain quiet. This evidence supports a maintenance-facing risk monitor whose alarms trigger engineering review. The event intervals establish operational feasibility, while broader maintenance histories will estimate population event rates and refine causal specificity.

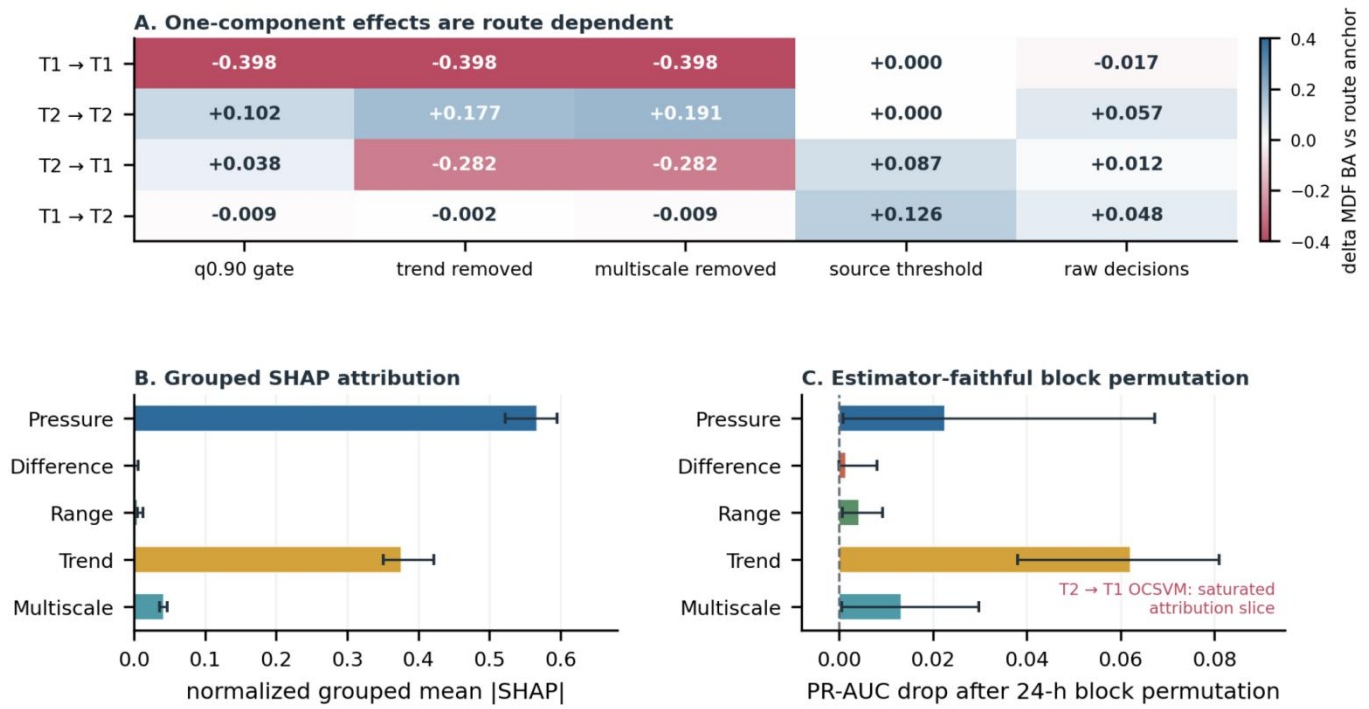


Fig. 8: Route-wise component ablation and grouped feature contribution. Panel A reports component effects relative to each frozen route while retaining all-window, q0.90-gated, and matched-window population labels. Panel B gives grouped SHAP attribution, and Panel C gives estimator-faithful 24-h block-permutation importance. Attribution is conducted after route selection; together, the panels identify pressure and trend as the principal evidence and show how calibration and MDF trade balanced accuracy for a lower maintenance alarm load.

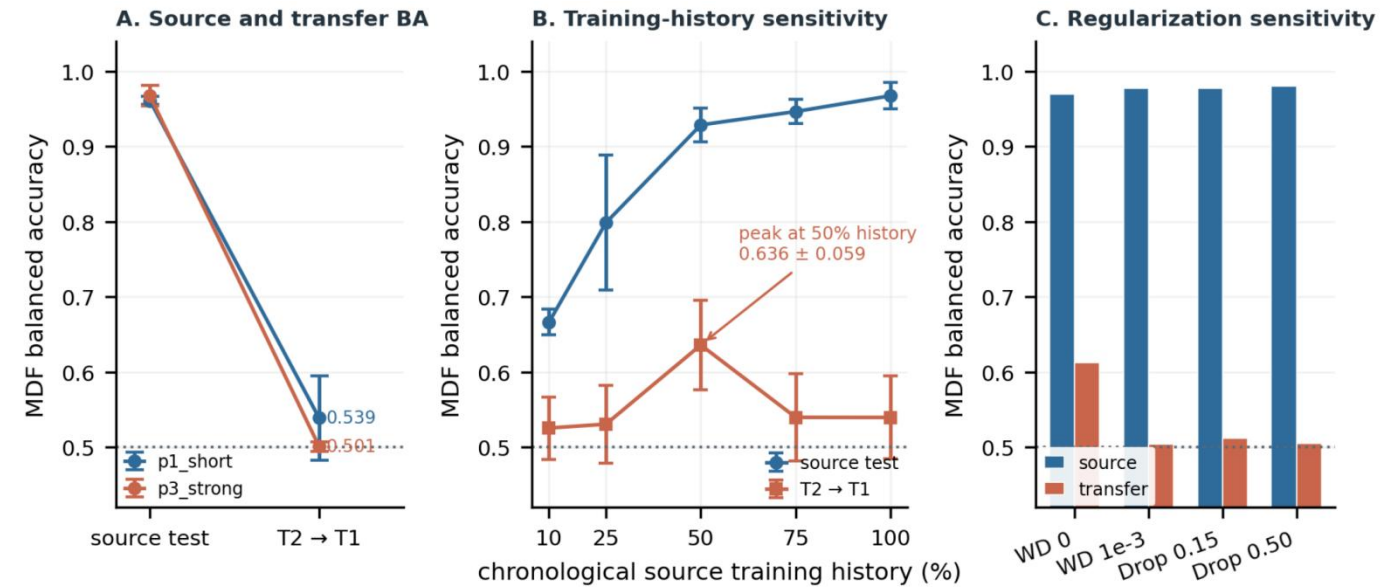


Fig. 9: Four-branch deep-fusion capacity audit. Panel A compares source-test and T2-to-T1 performance for validation-selected configurations. Panel B shows that 50% chronological source history gives the highest recorded transfer BA, 0.636 ± 0.059 . Panel C maps weight-decay and dropout sensitivity. The figure positions the calibrated classical route as the stronger deployment choice for the tested architecture, one-minute sensing rate, two public units, and recorded search budget.

6. Limitations

The evidence scope contains two public turbines and one long run per state. Its 60-min windows use a 30-min stride, and adjacent chronological blocks retain the disclosed 29-min raw overlap. Glycerol presence serves as a lubrication-state proxy. Equa-

tion (1) is therefore interpreted at the level of combined actuation effort; synchronized high-rate blade acceleration, hydrodynamic torque, and pressure-to-torque calibration would support future decomposition of the individual torque terms.

Target-normal calibration requires a trustworthy normal reference on every new unit. The field study supplies three

engineering event intervals with annotated interval starts. Additional units, years, maintenance outcomes, and independent mechanical measurements will estimate fleet-level transfer and causal specificity. The deep audit covers the existing registry, one-minute sensing rate, and recorded search budget; higher-rate multimodal sensing and substantially larger multi-unit training data define the next evaluation regime.

7. Conclusion

KFM-Cal converts installed Kaplan servo-pressure SCADA into persistent friction/lubrication-risk alarms through a unit-adaptive algorithm with three coordinated stages: mechanism-grounded pressure evidence, target-normal operating-point adaptation of a frozen source ranker, and causal proportion-voting MDF. The all-window seed-0 T1 same-unit route reaches BA 0.987 at FAR 0.018, and T2-to-T1 transfer reaches BA 0.782 at FAR 0.005. On 365 days of Sichuan Shawan SCADA, the frozen monitor covers 3/3 engineering events and lowers routine FAR from 0.030872 to 0.000441 with mean delay 30.96 h. Ablation and attribution identify pressure and trend as the principal public-data evidence, while the transfer and deep audits quantify source-target compatibility. The resulting method provides a leakage-aware, physically interpretable, and calibratable route from existing pressure records to maintenance-facing alarms.

Project Information

This work is associated with Contract No. SW2026/M1-39.

Declaration of AI-Assisted Language Editing

During preparation of this manuscript, the authors used ChatGPT only for English grammar checking and language polishing. No generative AI or AI-assisted tool was used to produce the scientific content, data, analyses, figures, tables, results, interpretations, conclusions, or references. The authors reviewed and edited all language changes and take full responsibility for the content of the submitted manuscript.

References

- [1] Ran Gong, Jiamin He, and Cheng Wang. Dynamic predictive modeling driven fault diagnosis method for planetary transmissions under time varying rotational speeds. *Measurement*, 289:122659, 2026.
- [2] Hongchuang Tan, Yiheng Su, Suchao Xie, Enci Yan, and Deqiang He. Standardized collaborative multiscale manhattan entropy: A nonlinear time series metric for train bearing fault diagnosis. *Digital Signal Processing*, 182:106322, 2026.
- [3] Lars-Johan Sandstrom, Kim Berglund, Par Marklund, and Gregory F. Simmons. Friction monitoring in kaplan turbines. *Machines*, 13(4):313, 2025.
- [4] Lars-Johan Sandstrom, Kim Berglund, Par Marklund, and Gregory F. Simmons. Hydropower turbine friction monitoring dataset, 2025.
- [5] A. Åsnes, A. Willersrud, F. Kretz, and L. Imsland. Predictive maintenance and life cycle estimation for hydro power plants with real-time analytics. In *Proceedings of the 2018 HYDRO Conference*, Gdansk, Poland, 2018. 15–17 October 2018.
- [6] Wenlong Zhu, Jianzhong Zhou, Xin Xia, Chaoshun Li, Jian Xiao, Han Xiao, and Xinxin Zhang. A novel KICA–PCA fault detection model for condition process of hydro-electric generating unit. *Measurement*, 58:197–206, 2014.
- [7] Arthur Henrique de Andrade Melani, Miguel Angelo de Carvalho Michalski, Renan Favarão da Silva, and Gilberto Francisco Martha de Souza. A framework to automate fault detection and diagnosis based on moving window principal component analysis and Bayesian network. *Reliability Engineering & System Safety*, 215:107837, 2021.
- [8] Wei Jiang, Yanhe Xu, Zhong Chen, Nan Zhang, Xiaoming Xue, Jie Liu, and Jianzhong Zhou. A feature-level degradation measurement method for composite health index construction and trend prediction modeling. *Measurement*, 206:112324, 2023.
- [9] Imtiaz Ahmed, Aldo Dagnino, and Yu Ding. Unsupervised anomaly detection based on minimum spanning tree approximated distance measures and its application to hydropower turbines. *IEEE Transactions on Automation Science and Engineering*, 16(2):654–667, 2019.
- [10] Hongteng Wang, Xuewei Liu, Liyong Ma, and Yong Zhang. Anomaly detection for hydropower turbine unit based on variational modal decomposition and deep autoencoder. *Energy Reports*, 7:938–946, 2021.
- [11] Mallikarjun Kande, AlfIsaksson, Rajeev Thottappillil, and Nathaniel Taylor. Rotating electrical machine condition monitoring automation—a review. *Machines*, 5(4):24, 2017.
- [12] Alessandro Betti, Emanuele Crisostomi, Gianluca Paoletti, Antonio Piazzi, Fabrizio Ruffini, and Mauro Tucci. Condition monitoring and predictive maintenance methodologies for hydropower plants equipment. *Renewable Energy*, 171:246–253, 2021.
- [13] Xavier Escaler, Eduard Egusquiza, Mohamed Farhat, Francois Avellan, and Miguel Coussirat. Detection of cavitation in hydraulic turbines. *Mechanical Systems and Signal Processing*, 20(4):983–1007, 2006.
- [14] L. Gaisser, O. Kirschner, and S. Riedelbauch. Cavitation detection in hydraulic machinery by analyzing acoustic emissions under strong domain shifts using neural networks. *Physics of Fluids*, 35(2), 2023.
- [15] Abderrauof Benabdesselam, Martin Gagnon, and Antoine Tahan. Diagnostic of erosive cavitation in hydraulic turbines from indirect measurements using scada data with uncertainty quantification. *Results in Engineering*, 30:110456, 2026.
- [16] Mehmet Akif Bütüner, İlhan Koşalay, and Doğan Gezer. Machine-learning-based modeling of a hydraulic speed governor for anomaly detection in hydropower plants. *Energies*, 15(21):7974, 2022.

- [17] Bernhard Scholkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.
- [18] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE, 2008.
- [19] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jorg Sander. Lof: identifying density-based local outliers. *ACM SIGMOD Record*, 29(2):93–104, 2000.
- [20] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation, 2014.
- [21] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, Francois Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *arXiv preprint arXiv:1505.07818*, 2015.
- [22] Chengjian Yuan, Xiangqu Xiao, and Chaoshun Li. A two-stage multi-source subdomain adaptation network for fault diagnosis of rotating machinery under incomplete health state. *Expert Systems with Applications*, 320:132153, 2026.
- [23] Ruqiang Yan, Weihua Li, Siliang Lu, Min Xia, Zhuyun Chen, Zheng Zhou, Yasong Li, and Jingfeng Lu. Transfer learning for prognostics and health management: Advances, challenges, and opportunities. *Journal of Dynamics, Monitoring and Diagnostics*, 3(2):60–82, 2024.