

Revised supplementary materials

Manuscript systems-4601389 | Scientific revision: 29 September 2026 | Language edit: 6 October 2026

Digital Twin Ready Predictive Risk Management of Subsurface Injection Nodes Using Mechanics Informed Cross Fault Machine Learning

This supplement documents the existing 18 event trajectories, the recovered computational definitions, and new retrospective sensitivity analyses. This clean copy is accompanied by a comment-by-comment response containing the relevant manuscript excerpts. Frozen outcome values are retained. A separate external archive adds six slow-creep stage peaks from one independent specimen. No prospective experiments or field validation were added.

S1 Policy definitions and reproducible settings

R→S uses 12 rough source events from two specimens and six smooth target events from one specimen. S→R reverses those roles. Each source model uses scaling fitted within its training fold, balanced logistic regression ($C = 0.5$, liblinear, max_iter = 3000), and an additive source OOF intercept correction fitted by binary log loss with slope fixed at one. The intercept is a calibration-in-the-large adjustment, not evidence of complete target calibration. A score meets the threshold when probability $\geq$ threshold; an indicator is active only while its consecutive run meets the persistence rule and resets below threshold. Table 5 of the manuscript reports ungated policies.

Table S1. Original source policy settings. Full precision is retained in policy_metrics.csv and exact_thresholds.csv for computational reproduction, not statistical certainty.

Dir.	Policy	h (s)	q	Run (s)	Threshold	Logit shift
R→S	Hydraulic-only	30	0.95	3	0.38784469	-1.53621737
R→S	Acoustic-only	30	0.95	3	0.20683042	-1.66607652
R→S	Watch	60	0.90	5	0.31285211	-0.48999694
R→S	Mitigate	30	0.95	3	0.26023004	-1.56426335
R→S	Hold	10	0.99	2	0.64897965	-3.02753207
S→R	Hydraulic-only	30	0.95	3	0.41460633	-1.51227089
S→R	Acoustic-only	30	0.95	3	0.14214513	-1.50913276
S→R	Watch	60	0.90	5	0.27670086	-0.71836150
S→R	Mitigate	30	0.95	3	0.16342557	-1.34840070
S→R	Hold	10	0.99	2	0.14636248	-2.24909579

All original thresholds and target outcomes were reproduced to floating-point tolerance. The original code specifies bootstrap seed 20260914 but no logistic random_state. The revision policy sensitivity analysis fixes random_state = 20260914; the largest threshold difference is below 1×10^{-16} . There is no evidence that the original policy settings were preregistered, and the revision does not select a new optimum from target outcomes.

S2 Event definition and preprocessing audit

The revised event catalog is reconstructed directly from the raw workbooks on the original 1 Hz grid. A trailing five-sample mean pump rate above 0.1 mL/min for at least five consecutive samples identifies each physical ramp. The earliest maximum slip velocity between that ramp start and the next ramp start defines the event peak. The final search ends with the record. No manual event addition or removal is used. All 18 original peaks and 3240 model rows are reproduced; the largest state-variable error is below 1.5×10^{-11} . This is a post-review reconstruction developed with the original catalog available, not an independently registered event detector.

The original injection_stage and inj_start_index fields are numerically midpoint partitions between consecutive selected peaks; they must not be interpreted as physical injection starts. Those fields are not used by this reconstruction. Historical manual annotation decisions remain unknown. Table S2 replaces the former midpoint search bounds with measured pump-ramp bounds. Full raw-to-event code, hashes, feature checks, duplicate-maximum handling and sensitivity outputs are included in the version 3 archive.

Table S2. Reconstructed physical pump-ramp search intervals and selected peaks in elapsed grid seconds. Each model window contains lead 180,...,1; the peak is excluded.

Specimen	Stage	Start	End	Peak	Window	Ties
Rough 1	1	1714	2314	1877	1697–1876	1
Rough 1	2	2315	2916	2428	2248–2427	1
Rough 1	3	2917	3517	3033	2853–3032	1
Rough 1	4	3518	4118	3623	3443–3622	1
Rough 1	5	4119	4718	4213	4033–4212	2
Rough 1	6	4719	5314	4800	4620–4799	2
Rough 2	1	1787	2388	1923	1743–1922	1
Rough 2	2	2389	2989	2502	2322–2501	1
Rough 2	3	2990	3590	3100	2920–3099	1
Rough 2	4	3591	4190	3699	3519–3698	1
Rough 2	5	4191	4792	4291	4111–4290	1
Rough 2	6	4793	5384	4886	4706–4885	1
Smooth 1	1	1997	2598	2144	1964–2143	1
Smooth 1	2	2599	3201	2702	2522–2701	1
Smooth 1	3	3202	3805	3295	3115–3294	1
Smooth 1	4	3806	4405	3889	3709–3888	1
Smooth 1	5	4406	5022	4480	4300–4479	1
Smooth 1	6	5023	5621	5086	4906–5085	1

With a five-sample trailing mean, all nine combinations of pump thresholds 0.05/0.10/0.20 mL/min and qualifying runs of 3/5/10 s reproduce all 18 peaks on both raw-reconstructed and archived cleaned streams. Ten-sample means also retain the peaks in this grid. Unsmoothed low-threshold variants fragment some pressure ramps and select peaks only 22–91 s apart; these overlapping fragments are not new independent events. Every window selected by the primary rule has a complete six-channel history over all 209 rows needed to construct its 180 feature rows.

The label is $y_h = 1$ for $0 < \text{lead} \leq h$ and zero otherwise. Every window has exactly 180 rows, producing label fractions 1/3, 1/6, and 1/18 at 60, 30, and 10 s. A future catalog peak determines the event-window selection, so this prevalence is unsuitable as a continuous-monitoring prevalence estimate. Stages remain clustered within specimens.

Comparison of the original XLSX files with the grid data establishes linear interpolation, rather than within-second means. Maximum next-bracket delays are 0.0970 s (Rough 1), 0.0985 s (Rough 2), and 0.0996 s (Smooth). Strict zero-latency online causality is not demonstrated, and upstream slip-velocity/AE processing is not independently established as causal. The archived 1 Hz grid is the frozen input for all reported analyses.

Pressure slope is $[p(t) - p(t-29)]/29$; stress and slip changes use the same endpoints. Mean velocity and mean AE use a trailing 30-row window, with the recovered reconstruction allowing at least 24 available rows. The acoustic predictor is $\log_{10}$ of the mean raw AE rate, not the mean of $\log$ AE or an untransformed mean. The seven definitions reproduce all 3240 stored feature rows. Although the recovered mean code permits 24 rows, the raw-record audit confirms complete 30/30 history for every evaluated event row. Interstage evaluation explicitly requires complete 30/30 raw-channel history.

S3 Individual event outcomes

Warnings are seconds within the nominal horizon among intercepted events. A dash means that the tier did not intercept the event. Values describe observed events, not reliable population warning times. Support is a separate diagnostic and does not gate these warning outcomes.

Table S3. Original event warning times and empirical support at q99.

Target event	Watch (s)	Mitigate (s)	Hold (s)	Support ≤ 30 s (%)	Support ≤ 10 s (%)
Smooth_E1	22	12	—	93.3	80.0
Smooth_E2	17	8	1	93.3	80.0
Smooth_E3	16	5	—	90.0	70.0
Smooth_E4	16	2	—	93.3	80.0
Smooth_E5	18	1	—	90.0	70.0
Smooth_E6	14	1	—	90.0	70.0
Rough 1_E1	60	30	—	23.3	0.0
Rough 1_E2	60	30	10	0.0	0.0
Rough 1_E3	60	30	10	10.0	0.0
Rough 1_E4	60	30	10	30.0	0.0
Rough 1_E5	60	30	10	56.7	0.0

Target event	Watch (s)	Mitigate (s)	Hold (s)	Support ≤ 30 s (%)	Support ≤ 10 s (%)
Rough 1_E6	60	30	10	80.0	40.0
Rough 2_E1	60	30	—	100.0	100.0
Rough 2_E2	60	30	10	50.0	0.0
Rough 2_E3	60	30	10	43.3	0.0
Rough 2_E4	60	30	10	43.3	0.0
Rough 2_E5	60	30	10	0.0	0.0
Rough 2_E6	60	30	10	0.0	0.0

S4 Policy sensitivity and statistical precision

The revision evaluates $q = 0.90, 0.95, 0.99$ crossed with persistence = 1, 2, 3, 5 s at each horizon and direction: 72 configurations. Each probability threshold is calibrated from source OOF negatives; target outputs are reported without selecting a favored new policy. The nominal negative exceedance proportions before persistence are approximately 10%, 5%, and 1%, subject to ties; they are not controlled field false-alarm rates. Full settings and outcomes appear in policy_grid.csv.

Table S4. Ranges across the 12 threshold/persistence combinations per tier. These are sensitivity ranges, not confidence intervals.

Transfer	Tier	Intercept range	Early activity (s/event)	Window duty (%)
R→S	Watch	6/6–6/6	0.0–0.0	5.1–12.2
R→S	Mitigate	2/6–6/6	0.0–0.0	0.9–5.0
R→S	Hold	0/6–5/6	0.0–0.5	0.0–2.3
S→R	Watch	12/12–12/12	6.3–27.0	34.2–48.0
S→R	Mitigate	12/12–12/12	34.2–49.8	35.6–44.4
S→R	Hold	10/12–10/12	23.3–41.8	17.5–27.9

Rough to smooth Hold at q99 intercepts 2/6, 1/6, 0/6, and 0/6 as persistence rises from 1 to 2, 3, and 5 s. Watch interception is stable in this grid, but its median warning varies from 8.5 to 24.5 s. Stable observed interception does not demonstrate an optimal or reliable policy on new specimens. Display rounding in Table S4 follows the manuscript's half-up convention at one decimal place: the unchanged lower-bound values 6.25 and 23.25 s/event are shown as 6.3 and 23.3 s/event, respectively.

S5 Support gate diagnostics

The original Mahalanobis gate uses source 10th–90th percentile scaling, Ledoit–Wolf covariance, leave-one-event-out calibration distances, and all-source refitting for target evaluation. Original distances and cutoffs reproduce within 8×10^{-14} . A q99 cutoff is an illustrative source-tail choice, not a confidence level,

field operating limit, or 99% future-coverage guarantee. Exact distances and event percentiles are provided in CSV files.

Table S5a. Empirical distance distributions from correlated event-window seconds. Source and target fits differ.

Transfer	Distribution	Median	q95	q99	q99.5
R→S	Source OOF	2.21	3.88	5.03	5.71
R→S	target	2.30	3.47	5.33	6.14
S→R	Source OOF	1.79	3.37	8.74	9.93
S→R	target	4.38	11.62	14.62	15.12

Table S5b. Each source event's cross-fitted distance distribution and exceedance of its domain's pooled q99.

Held-out source event	Median D	q95	q99	Above pooled q99 (%)
Rough 1_E1	2.56	3.54	3.82	0.0
Rough 1_E2	2.89	4.42	4.83	0.0
Rough 1_E3	1.84	3.38	4.04	0.0
Rough 1_E4	1.83	3.53	4.06	0.0
Rough 1_E5	1.62	3.21	4.00	0.0
Rough 1_E6	2.21	3.63	4.57	0.0
Rough 2_E1	2.81	3.27	3.39	0.0
Rough 2_E2	2.15	2.78	3.97	0.0
Rough 2_E3	1.77	2.55	2.97	0.0
Rough 2_E4	2.04	3.61	4.63	0.6
Rough 2_E5	1.93	4.61	5.69	3.9
Rough 2_E6	2.42	5.68	7.22	7.8
Smooth 1_E1	2.29	8.04	10.71	4.4
Smooth 1_E2	1.69	3.11	7.89	0.6
Smooth 1_E3	1.38	3.26	5.40	0.0
Smooth 1_E4	1.30	2.32	4.55	0.0
Smooth 1_E5	1.43	2.65	6.86	0.0
Smooth 1_E6	2.48	3.12	8.47	1.1

Deleting a source event means removing the complete trajectory before recomputing the remaining cross-fitting folds, threshold, and final fit. Temporal thinning retains source rows at lead times divisible by 1, 5,

10, or 30 while all target rows remain. The diagonal alternative uses robust-scaled Euclidean distance centered on source medians. The one-class SVM uses an RBF kernel, $\gamma = \text{scale}$, $\nu = 0.01$, and negative decision function as its anomaly score. Each alternative receives source-only event-cross-fitted q99 calibration. SVM score scales can change with fit size, which adds a cross-fit-to-final-fit limitation. These analyses are diagnostic comparisons, not optimized method rankings.

Table S6a. Alternative source-calibrated support gates at q99.

Transfer	Gate	Support ≤ 30 s (%)	Support ≤ 10 s (%)
R→S	mahalanobis	91.7	75.0
R→S	diagonal robust	100.0	100.0
R→S	one class svm	97.2	91.7
S→R	mahalanobis	36.4	11.7
S→R	diagonal robust	63.1	27.5
S→R	one class svm	31.9	10.8

Table S6b. Source temporal thinning with full target evaluation. Thinning changes both temporal sampling and covariance shrinkage.

Transfer	Source step (s)	q99 cutoff	Support ≤ 30 s (%)	Support ≤ 10 s (%)
R→S	1	5.03	91.7	75.0
R→S	5	4.80	93.3	80.0
R→S	10	5.06	93.9	81.7
R→S	30	4.92	85.6	56.7
S→R	1	8.74	36.4	11.7
S→R	5	7.97	27.5	8.3
S→R	10	7.25	11.1	3.3
S→R	30	5.26	0.0	0.0

Table S7. Complete source-event deletion and recalibration. The ranges describe sensitivity; they are not uncertainty intervals.

Deleted source event	Recalibrated q99	Target support ≤ 30 s (%)	Target support ≤ 10 s (%)
Rough 1_E1	4.99	92.2	76.7
Rough 1_E2	5.03	92.2	76.7
Rough 1_E3	5.22	93.9	81.7
Rough 1_E4	5.13	93.9	81.7

Deleted source event	Recalibrated q99	Target support ≤ 30 s (%)	Target support ≤ 10 s (%)
Rough 1_E5	5.01	92.8	78.3
Rough 1_E6	5.15	91.7	75.0
Rough 2_E1	5.04	91.1	73.3
Rough 2_E2	5.06	92.2	76.7
Rough 2_E3	5.08	92.8	78.3
Rough 2_E4	5.41	93.9	81.7
Rough 2_E5	5.60	93.9	81.7
Rough 2_E6	5.00	90.6	71.7
Smooth 1_E1	7.98	19.7	0.8
Smooth 1_E2	13.36	88.6	74.2
Smooth 1_E3	8.77	41.7	11.7
Smooth 1_E4	8.56	40.0	11.7
Smooth 1_E5	8.51	38.6	11.7
Smooth 1_E6	8.29	36.1	10.8

Deleting Smooth 1_E2 increases rough target final-30-s coverage to 88.6%; across all smooth-source deletions coverage spans 19.7–88.6%. Rough-source deletions produce 90.6–93.9%. Whole-rough-specimen cross-fitting changes q99 to 9.47 and retains 100% of smooth states. Only two rough specimens are included in that diagnostic, and no analogous smooth-specimen calibration can be performed with one specimen. Median within-event lag-one distance correlation is 0.947 for rough events and 0.963 for smooth events. The correlation and thinning analyses are descriptive because these trajectories are nonstationary.

Table S8. Mean signed coordinate contributions to D^2 during the final 30 s. The AE coordinate is log-transformed as defined in S2.

Coordinate	R→S contribution	S→R contribution
hydraulic loading ratio	1.71	-0.46
pressure slope 30s	1.42	-3.68
shear delta 30s	7.44	-93.13
normal delta 30s	2.21	-8.08
slip delta 30s	-3.39	53.50
velocity mean 30s	0.79	92.15
ae mean 30s	0.82	67.01
Total mean D^2	11.00	107.30

Let z be the target state transformed by source-fitted scaling. For $\delta = z - \mu_s$ and $P = \Sigma_s^{-1}$, $c_j = \delta_j(P\delta)_j$ and $\Sigma_j c_j = D^2$. Negative terms are covariance cross-term allocations, not protective effects. These values are not causal importance or probabilities. Exact per-second decomposition and its algebraic verification are in the revision archive.

S6 Interstage exposure and washout sensitivity

Models and persistence are evaluated over all valid continuous observations. Metrics count only eligible intervals, and excluded gaps separate counted episodes. Eligibility requires complete 30/30 histories for six raw channels and excludes [peak-180, catalog post_end+washout], inclusive. The 30 s primary washout follows feature memory, not a proven physical reset. All retained intervals were archived as interstage_locked, with a pump rate ≤ 0.1 mL/min. They are not independently certified free of smaller slip or AE events. Original event predictions reproduce within 2×10^{-15} before continuous application; no target exposure tunes the source parameters.

Table S9. Actual short laboratory interstage exposure. Values are alert seconds (percent of eligible exposure), not field false-alarm rates.

Transfer	Washout (s)	Exposure (s)	Watch s (%)	Mitigate s (%)	Hold s
R→S	0	2277	124 (5.45)	70 (3.07)	0
R→S	30	2097	6 (0.29)	0 (0.00)	0
R→S	60	1917	0 (0.00)	0 (0.00)	0
S→R	0	4166	87 (2.09)	83 (1.99)	0
S→R	30	3806	61 (1.60)	34 (0.89)	0
S→R	60	3446	43 (1.25)	23 (0.67)	0

At 30 s washout, smooth Watch contains one episode; rough Watch five and rough Mitigate three. All active seconds pass q99 support. Rough 1 contributes 1783 s with no alerts; Rough 2 contributes 2023 s and all reverse-direction alerts. The substantial washout sensitivity and only 1.64 h combined exposure preclude claims about long-term field false alarms. Zero Hold activity here does not establish a zero underlying rate.

S7 Discrimination and conditional resampling

Table S10. AP with descriptive event-resampling percentile ranges. AP is average_precision_score, not trapezoidal PR area.

Transfer	h (s)	AP	Resampling 95% range	Positive/total rows	Specimens/event s
R→S	60	0.98	0.97–1.00	360/1080	1/6
R→S	30	0.81	0.72–0.95	180/1080	1/6
R→S	10	0.57	0.40–0.82	60/1080	1/6

Transfer	h (s)	AP	Resampling 95% range	Positive/total rows	Specimens/event s
S→R	60	0.97	0.93–0.99	720/2160	2/12
S→R	30	0.89	0.80–0.97	360/2160	2/12
S→R	10	0.53	0.42–0.73	120/2160	2/12

AP ranges use 5000 whole-event bootstrap resamples with seed 20260914 and fixed fitted predictions. They omit model-refit and between-specimen uncertainty. Original warning/burden/support summaries use 10,000 event resamples; Wilson ranges are nominal binomial event-count intervals, not valid population confidence bounds under unquantified stage dependence. Fixed-window prevalence is equal between directions at a given horizon despite different counts; it differs across horizons. All score comparisons remain descriptive.

S8 Reproduction and remaining provenance limits

The revision archive is version 3.0.0 dated 2026-09-29. README.md specifies the complete run order. The original computational environment record lists Python 3.13.5, NumPy 2.3.5, pandas 2.2.3, SciPy 1.17.0, scikit-learn 1.8.0, and matplotlib 3.10.8. Revision calculations ran in Python 3.12.14 with the recorded package versions; the archive contains exact environment JSON files. SciPy, required by the original script but omitted from its requirements list, is explicitly included. Original files are retained in analysis_package; new analyses occupy separate directories. SHA256SUMS records the complete file inventory; repository/verify_checksums.py verifies it. Refined figure exports, their plotting script, and unchanged plotting inputs are supplied separately in Systems_4601389_Refined_Figure_Sources.zip; no analysis file in the original version 3 archive is overwritten.

A public or anonymous repository URL has not been created. The complete archive is provided for journal upload, with no fabricated link or DOI. The original annotation intent was not recovered; the revision supplies a complete raw-pump-to-catalog reconstruction that reproduces the original peaks without the retrospective midpoint boundaries. The companion submitted manuscript is disclosed as unpublished, and its current editorial status should be confirmed by the authors. One additional independent specimen has been evaluated; further replication and compatible field validation remain needed.

S9 Event end annotation sensitivity

This check evaluates the existing frozen continuous predictions after excluding [peak–180, peak+L] with L = 180, 210 or 240 s, complete feature history and pump rate ≤ 0.1 mL/min. It uses no historical post_end annotations. It is a sensitivity analysis of selected locked exposure, not a field false-alarm rate. Fixed durations are not verified recovery times.

Table S11. Alert seconds (eligible-exposure duty percent) under fixed post-peak exclusions.

Transfer	L (s)	Exposure (s)	Watch s (%)	Mitigate s (%)	Hold s
R→S	180	1610	0 (0.00)	0 (0.00)	0
R→S	210	1430	0 (0.00)	0 (0.00)	0
R→S	240	1250	0 (0.00)	0 (0.00)	0

Transfer	L (s)	Exposure (s)	Watch s (%)	Mitigate s (%)	Hold s
S→R	180	3388	33 (0.97)	16 (0.47)	0
S→R	210	3028	33 (1.09)	16 (0.53)	0
S→R	240	2668	33 (1.24)	16 (0.60)	0

The rough target has four Watch episodes and one Mitigate episode under all three settings. At $L = 210$ s, 375 rough-target seconds are unsupported, but all 33 Watch seconds and 16 Mitigate seconds are supported; zero smooth-target exposure seconds are unsupported. These results identify residual alert activity without relying on the archived event-end annotations.

S10 Supplementary diagnostic figures

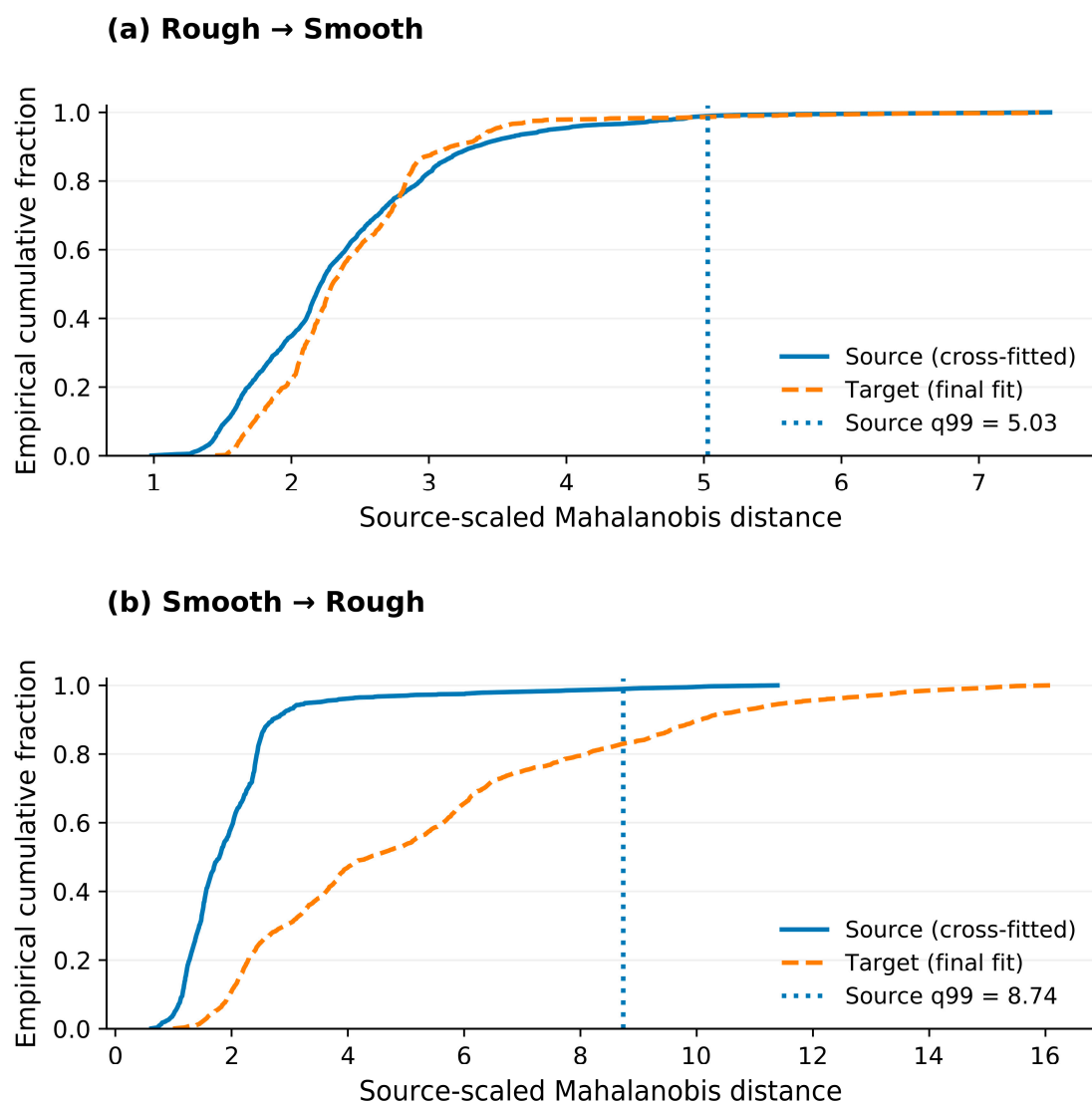


Figure S1. Empirical cumulative distributions of source cross-fitted and final-fit target distances for (a) rough-to-smooth and (b) smooth-to-rough transfer. Dotted lines identify the respective source q99 cutoffs. Each curve uses the unchanged correlated one-second event-window states, not independent operating samples. The two panels have separately fitted source geometries.

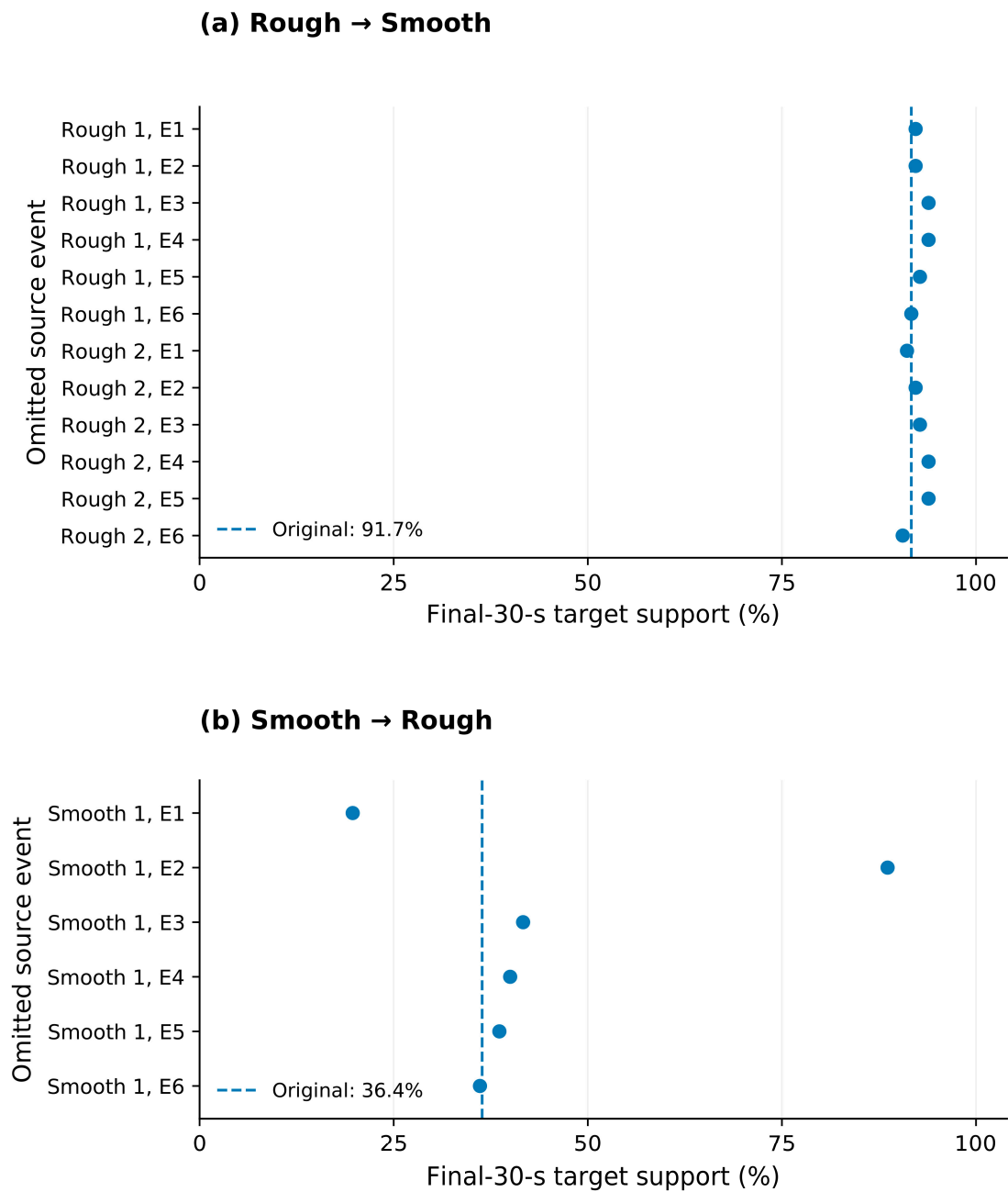


Figure S2. Final-30-s target support after deleting one source event and completely recalibrating: (a) rough-source deletions; (b) smooth-source deletions. Each point labels the omitted event; dashed lines show the original full-source result. Both panels use the same support scale. These sensitivity cases reuse the same target observations and do not constitute additional validation datasets.

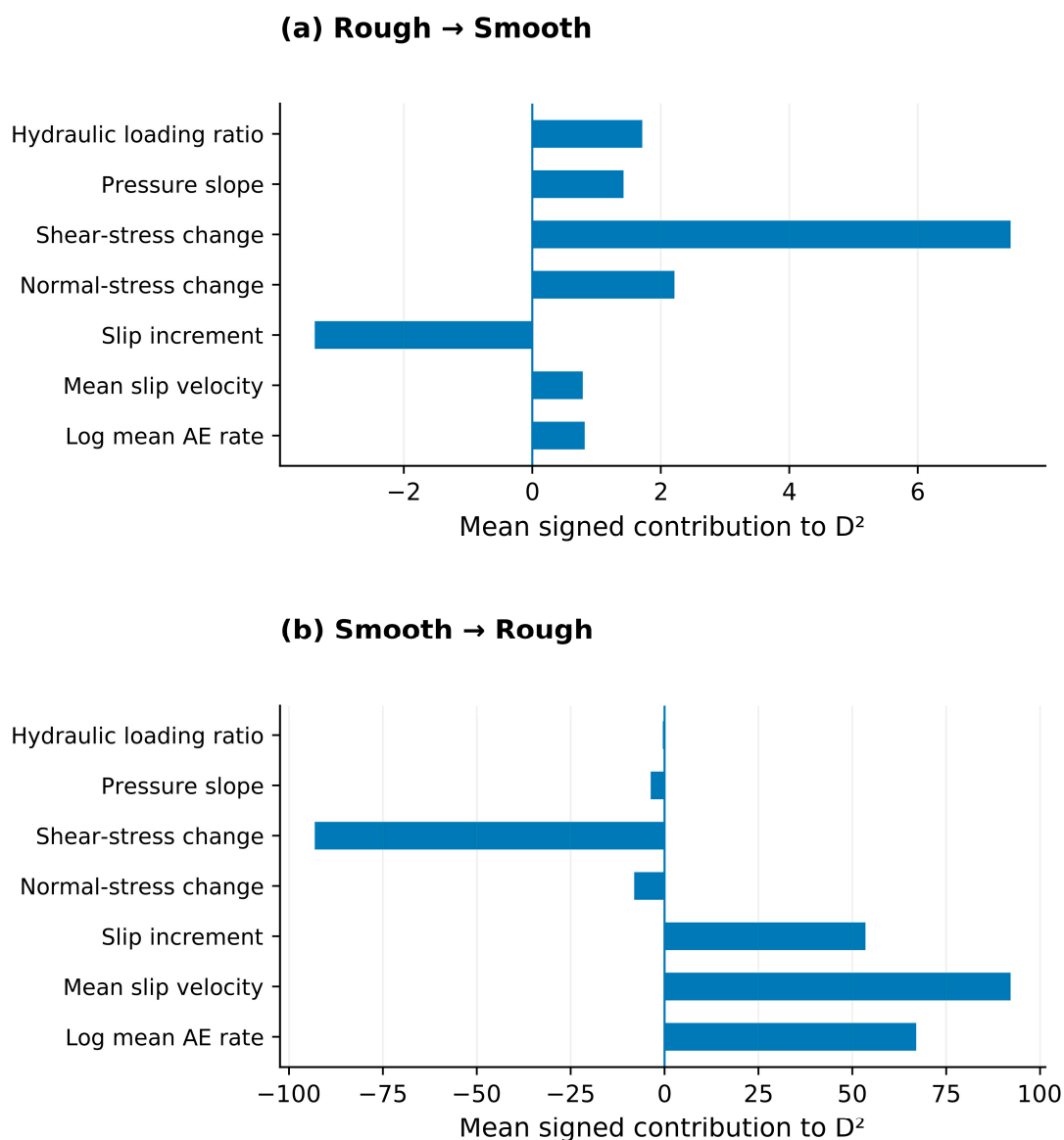


Figure S3. Signed coordinate contributions to mean D^2 in the final 30 s for (a) rough-to-smooth and (b) smooth-to-rough transfer. The acoustic coordinate is the log-transformed mean AE rate. Negative terms allocate covariance cross-terms; they are not protective or causal effects. The panels use different horizontal scales. Values are unchanged from Table S8.

S11 Independent external laboratory evaluation

Wang et al. (2020), Figshare version 3, DOI 10.6084/m9.figshare.11948739.v3, supplies the independent SC2 slow-pressurization specimen. The associated methods article is DOI 10.1029/2019GL086627. SC1 is excluded because its AE and hydraulic histories match the original Smooth 1 experiment; 39,500 shared AE timestamps agree within 5.2×10^{-13} counts/s. This exclusion prevents repeated publication of the same observations from increasing the sample count.

SC2 contains six nonoverlapping 180-row windows preceding the earliest archived velocity maximum in each published pressure-stage. Stage markers are the first 5.10/9.10/13.10/17.10/21.10/25.10 MPa crossings. Each window has 29 additional preceding rows for complete feature history and no earlier cataloged stage maximum in its pre-180-s interval. Separate mechanical, hydraulic, velocity and AE records are interpolated to common integer experimental seconds 3200–7200 without extrapolation. Pressure is

converted from bar to MPa; archived effective normal stress is used directly; slip, velocity and AE rate are mm, mm/s and counts/s. Processed AE/velocity conventions are not certified as identical or online causal.

Both source models and all existing source-only calibrations, thresholds, persistence settings and support cutoffs were frozen. Original 18-event outputs were reproduced before external prediction. All six SC2 peaks were retained before outcomes were computed, and no external data are used to fit or select the model. These are six repeated slow-creep stage peaks on one specimen, not six independent earthquakes. The original 18-event cohort and external six-peak cohort remain separate.

Table S12. Independent SC2 outcomes for unchanged source policies. Warning times are calculated for intercepted peaks only. Early exposure and duty refer to event windows, not field false-alarm burden.

Source	Tier	Hits	Median warning (s)	Early (s/event)	Duty (%)	AP
Rough	Watch	5/6	45	3.2	20.6	0.85
Rough	Mitigate	0/6	—	0.0	0.0	0.76
Rough	Hold	0/6	—	0.0	0.0	0.34
Smooth	Watch	4/6	60	39.2	43.0	0.68
Smooth	Mitigate	6/6	30	72.8	56.7	0.60
Smooth	Hold	3/6	10	26.8	17.7	0.23

Rough-source support is 1077/1080 rows overall (99.72%), 357/360 within 60 s (99.17%), 177/180 within 30 s (98.33%), and 59/60 within 10 s (98.33%). Smooth-source support is 100% in each window. Supported-hit counts equal the displayed hit counts. Rough-source Mitigate never activates despite AP 0.76 and high support; smooth-source Mitigate intercepts every peak but is active early for 72.8 s/event. These outcomes show that domain support and ranking performance do not establish portable source-calibrated thresholds.

Additional published archives were screened but not counted as validation. Cao (2023), DOI 10.17632/nz8hm5z4y5.3, contains four tests on one shale specimen and lacks archived velocity and AE-rate channels. Geng (2026), DOI 10.17632/kvwnbnrpg.2, provides eight specimens, but short cycles create ambiguous or overlapping 180 s targets and its archived AE rate counts all sensor-channel hits. Of 82 cycle maxima, eight pass basic history/spacing screening, but none was accepted as a fully equivalent labeled input; no model outcomes were computed. Mont Terri field data, DOI 10.5281/zenodo.6811595, provide hydraulic, strain and seismic records but need calibrated slip/stress reconstruction, clock reconciliation and compatible event definitions. These specific obstacles prevent treating available records as validation of the unchanged seven-variable classifier. The archive includes the full screening audit and primary-source links.