WHAT A QUALITY FLAG DOES NOT ENCODE: A PRE-REGISTERED CERTIFICATE AUDIT OF THE MAST MAGNETICS CHANGEPOINT AT SHOT 16493

JUAN CARLOS PAREDES

Abstract. Archives of fusion diagnostics carry per-shot, per-signal quality flags, and it is tempting to read a flagged era as a calibrated-differently era. We audit that reading at the sharpest documented boundary we could find in the FAIR-MAST archive: the poloidal-field probe array `xma`/`xmb` shows exactly one Bad-to-good quality transition in 394,649 catalogued source rows, at shot 16492 → 16493, coinciding with a three-month between-campaign maintenance gap. We run, unchanged and with its selection rules pre-registered before any cross-era number was seen, a calibration- consistency certificate whose detection class is proven — it fires, with distribution-free semantics, on polarity/sign-structure faults, and a positive control on this very data shows it does (separation AUC 0.95 on a synthetic single-probe inversion). Against the real boundary the certificate is silent: cross-era session graphs are statistically indistinguishable from within-era matched nulls (separation AUC 0.485), no correlation sign flips exist across the boundary (0 of 12 audited probes; 3.1% of all 780 probe pairs, all at near-zero correlation), and the flagged era's data is structurally healthy — same variance, same correlation structure, same dropout behavior as the good era. The conclusion is a characterization, not a detection: **the archive's Bad flag at this boundary encodes something orthogonal to calibration sign-structure** — plausibly absolute gain/offset accuracy or curation judgment — and the flagged data is, for relational analyses, more usable than the flag suggests. We argue that pre-registered certificate audits of documented quality boundaries are a cheap, portable practice: either they find the first un-planted structural fault, or they sharpen — with proof of sensitivity — what a community's quality metadata does and does not mean.

1. The planted-fault gap, and what would close it

The companion method paper validates a cycle-frustration certificate for inter-calibration inconsistency on MAST magnetics and LHD bolometry: planted polarity faults are detected at AUC 0.976–1.000 and localized at 0.93–1.00 while remaining statistically invisible to pairwise monitoring. Its stated limitation (§8 there) is that every fault is planted. The natural closing move is a *documented* incident: a fault the archive itself records, detected by the same untuned pipeline. This note reports the first such attempt, run under pre-registration so that every outcome — fires-and-localizes, fires-without-localizing, does-not-fire — was reportable before the data spoke. The outcome was the third, and the design converts it into content.

2. The incident

A programmatic scan of the FAIR-MAST source catalogue (394,649 rows; per-shot, per-diagnostic `quality`) finds exactly one Bad→good transition for the `xma` poloidal-field array: shots $\leq$ 16492 are flagged Bad, shots $\geq$ 16493 good, with the boundary falling on the M6 campaign's three-month maintenance gap. The mirror array `xmb` shows the same boundary. [Evidence]

Date: July 5, 2026.

3. Method: the certificate, unchanged

The pipeline is Paper 1's, with scoring, thresholds, and gates untouched — "same untuned pipeline" is a load-bearing property of the program and of this audit. Sessions are discharges; nodes are probes; edges carry the sign and Fisher-z-bounded confidence of per-session correlations, N_eff-corrected for autocorrelation ($\rho_1 \approx 0.95$ here; N_eff $\approx$ 165 of 7,500 nominal samples — reported, and not the cause of anything below: the retained edges at $|r| \approx 0.9$ are significant under N_eff). The certificate condition and the leave-one-out localization are as in the companion paper. [Theorem]

Pre-registered choices (frozen in `RUN_LOG.md` before any cross-era result): 12 of 40 CCBV probes selected by the companion paper's availability rule (finite and non-constant in all 36 sampled shots — the rule, not the experimenter, picks); 18 evenly spaced usable shots per era near the boundary ([15000, 16492] Bad; [16493, 17800] good); within-era session pools as matched nulls; cross-era pools as the test; 60 trials per condition, matched RNG.

4. Result: silence, with proof the instrument was listening

- **No separation.** Cross-era frustration radius 0.046 ± 0.022 against within-era nulls 0.031 (pre) and 0.063 (post): separation AUC **0.485** — chance; the within-post null actually sits higher than the cross condition.
- **No sign structure change.** Correlation sign flips across the boundary: **0 of 12** audited probes; extending to all 40 probes, 24 of 780 pairs (3.1%) flip, all at near-zero correlation (median $|\Delta R| = 0.0075$) — bookkeeping noise, not polarity.
- **The flagged era is structurally healthy.** Median per-channel std 0.330 (Bad era) vs 0.314 (good); mean $|r|$ 0.786 vs 0.787; per-channel dropout occurs in both eras at similar rates. The Bad era is not dead, not noisy, not decorrelated.
- **The instrument works.** A synthetic sign inversion planted on one post-era probe — the positive control — lifts the radius to 0.110 and the separation AUC to **0.95** on this exact data. The silence on the real boundary is a property of the boundary, not of the certificate. [Evidence]

(Certificate firing rates are high in every condition — 0.68–1.00 — because CCBV probes co-view the plasma at mean $|r| \approx 0.79$ and odd cycles occur within single eras; as in the companion paper's envelope study, the *contrast* against matched nulls is the signal, and here the contrast is flat.)

5. What the flag therefore encodes — and the two-sided value

The certificate's detection class is sign/polarity structure: lead swaps, inverted conventions, re-mounted probes. The audit shows the Bad flag at this boundary encodes none of that. The remaining candidates are absolute calibration accuracy (gain or offset against a reference — invisible to a sign-and-$|r|$ certificate by construction), or curation judgment predating the modern pipeline. Either way, two communities gain: archive users learn that Bad-era `xma` data is relationally coherent — usable for correlation- and structure-based analyses even where absolute accuracy is unwarranted; and method users learn that **a documented quality changepoint need not be a structure changepoint** — quality metadata and calibration structure are different observables, and only the second is the certificate's target. A cheap sign-structure invariance check should precede any future claim that a flag boundary is a calibration incident; we have adopted this as a grading rule in our own dossier.

6. The baseline and the cross-domain twin

Two companion records frame the negative. First, the fault-free gluing baseline: the program's MAST record coheres across every attempted obstruction (25 notebook entries; the companion paper's null radii of 0.001–0.002), so certificate silence on honest data is the norm, not an anomaly of this run. Second, the space-domain twin: on ESA Swarm the same pipeline, under the same discipline, detected a documented hard instrument failure un-injected (the 2014 Swarm-C ASM death), returned a proven mechanism for why the tokamak-style money shot cannot exist on ρ-guarded spacecraft (the model-term wedge), and constructed the pairwise-invisible case exactly where the constellation's scalar guard is structurally absent. The pair of studies — one silence with proof of sensitivity, one detection with a documented cause — is the honest empirical boundary of what "real-fault validation" currently means for this method. [Cross-cite; the Swarm study is reported separately.]

7. Limitations and next steps

The mechanism behind the Bad flag remains uncharacterized; a one-line inquiry to the FAIR-MAST maintainers could settle it and is the single cheapest upgrade to this note (outward move, gated). One boundary, one machine, one diagnostic class was audited; the dossier's remaining candidates (MAST bolometers' end-of-life degradation; LHD inter-campaign bolometer recalibration) are graded B and would exercise different fault classes — the LHD candidate needs changepoint metadata we do not hold. The certificate's blindness to gain/offset faults is by design and is the companion program's transport-valued frontier, not a defect of the audit. And the audit's pre-registration is its load-bearing honesty device: the selection rules, session pools, and all three outcomes were committed before the first cross-era number existed; nothing here was chosen after the fact.

Figures

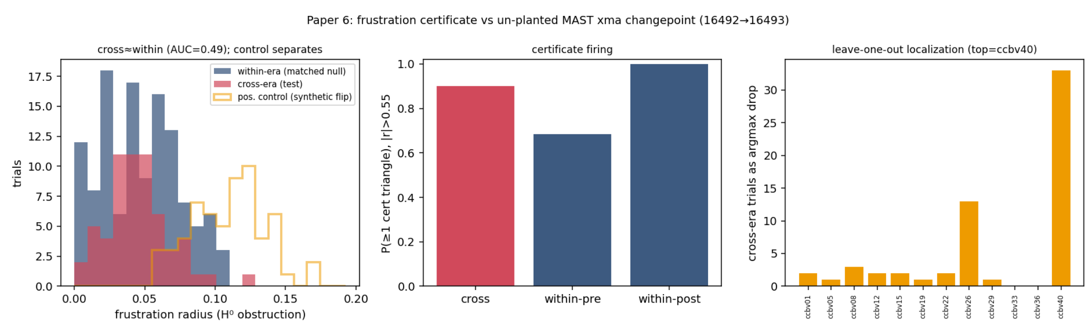

Figure 1. Cross-era vs within-era frustration radius, certificate firing rates, and the positive control (synthetic single-probe inversion, AUC 0.95): the real boundary at shot 16493 produces no separation (AUC 0.485).

Authorship and AI-assistance disclosure

This research program — its conception, its original ideas, its questions, and its direction — is the work of the author, Juan Carlos Paredes. The strategic pivots, the falsification discipline, the decisions about what to pursue, what to test, what to claim, and what not to claim are his throughout. In the course of the program he directed large language models (Anthropic Claude — Opus and Fable model families) as research assistants: implementing and running numerical experiments he commissioned, formalizing proofs in Lean 4, drafting and revising prose, and retrieving literature, all under his direction and review. Formal results are machine-checked (Lean 4) where stated; numerical claims are reproducible from the released scripts. Juan Carlos Paredes is the author of this work; the assistance is disclosed in the spirit of the full transparency this venue requires.

Independent researcher
Email address: cpt66778811@gmail.com