

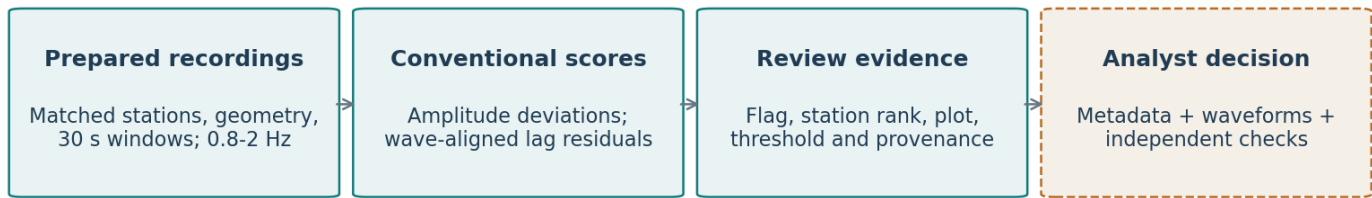
Analyst-assisted seismic quality control

Practical feasibility brief | 22 October 2026 | Research prototype

A reproducible local toolkit for reviewing amplitude and timing anomalies in compact seismic arrays. It measures how stations differ within a prepared recording window, flags unusually large residuals, and preserves the measurements and plots needed for an analyst to investigate.

The practical proposition is to make an existing technical task easier to inspect and repeat. The evidence supports a small conventional measurement component and a useful testing framework. Analyst time savings, field reliability and advantage over existing tools remain to be measured.

Measured support	What it establishes
2,496 archived synthetic records	The reduced implementation exactly reproduces its two residual vectors and maxima on all records.
68/72 weak clock injections flagged	64/72 are both flagged and assigned to the injected station; the timing flag also responds to complex wavefields.
12.2 ms median prepared-window compute	A local CPU feasibility result, with acquisition and preprocessing excluded.



Built: scripts, measurements, saved flags and illustrative reports. Proposed: an operator workflow and integrations.

Solid boxes describe implemented research components and saved artifacts. The analyst workflow and integrations require an operational pilot.

What exists today

Python measurement scripts; archived waveform/geometry fixtures; fixed review rules; per-window flag and station-index outputs; an illustrative review plot; reproducibility checks; and machine-readable evidence. There is no production interface, live station-health service or validated automatic fault diagnosis.

Recommended positioning: an inspectable array-review component for a specialist pilot, potentially integrated into an existing QC workflow. Pricing, customer demand and a distinct market advantage are open questions.

Measurements and evidence design

Established measurements, explicit calibration, retained raw inputs

Amplitude: compute station RMS over the prepared window, take its natural logarithm, subtract the array median, and rank absolute deviations. **Timing:** estimate one incident plane wave with phase-only FK beamforming, align station spectra to that estimate, and measure each station's residual lag against the normalized average of all other stations. The candidate station is the largest residual.

The reduced pipeline keeps these two measurements and omits covariance summaries, repeated station-removal FK fits and additional response features. It uses the saved, direction-checked FK implementation. Prior known-wave checks had maximum errors of 1.34 degrees internally and 0.71 degrees in four ObsPy cross-checks; these are implementation checks, not field accuracy guarantees.

Evidence design	Specification
Geometries	PDAR: 13 stations; TXAR: 9; NVAR: 11; withheld WRA: 10. Real geometries, synthetic waveforms.
Prepared observations	30-second, 10 Hz records; 0.8-2 Hz band; source SNR 2-15.
Original benchmark	480 independently seeded families; 2,496 records plus four ambiguity examples. Related conditions share a family.
Rule calibration	72 primary no-isolated-fault windows from validation families on PDAR, TXAR and NVAR.
Rule tests	72 new families on those three arrays; 120 WRA families; additional challenges on the 72 familiar-array test families.
Injected faults	Primary gain factors 2-8 or reciprocals; primary clock shifts 0.08-0.5 s. Weak gains 1.2-1.8 or reciprocals; weak clocks 0.02-0.07 s.

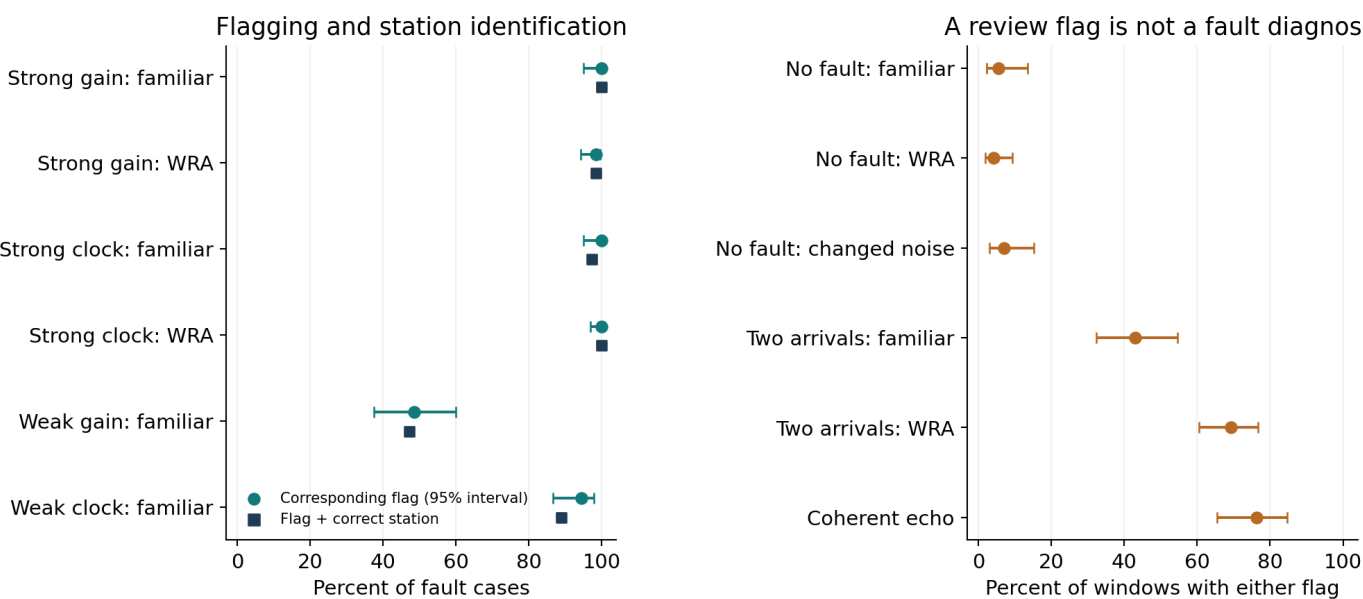
Each score's cutoff is its calibration 95th percentile, using the higher observed order statistic. Strict exceedance triggers a review flag: amplitude > 0.443735 natural-log units; absolute lag > 0.01 s. The lag-search step is 0.01 s. This step and cutoff do not establish 10 ms field accuracy.

The illustrative 5% per-score calibration target is not an industry standard or a combined false-alarm guarantee. Calibration flags were 3/72 for each score and 6/72 for either score. Thresholds were fixed before examining this extension's test results. The dataset and underlying residuals had been examined previously, so this is exploratory reuse, not a fresh confirmation.

What the screening test supports

Detection of an anomaly is evaluated separately from locating its station

Injected condition	Corresponding flag	Flag + correct station
Strong gain, familiar arrays	72/72 (100.0%)	72/72 (100.0%)
Strong gain, withheld WRA	118/120 (98.3%)	118/120 (98.3%)
Strong clock, familiar arrays	72/72 (100.0%)	70/72 (97.2%)
Strong clock, withheld WRA	120/120 (100.0%)	120/120 (100.0%)
Weak gain, familiar arrays	35/72 (48.6%)	34/72 (47.2%)
Weak clock, familiar arrays	68/72 (94.4%)	64/72 (88.9%)



Left: corresponding amplitude or timing flag, with Wilson 95% intervals; squares require both flagging and the correct station. Right: either flag in windows without an injected isolated instrument fault. Complex arrivals legitimately create residuals but prevent a fault-specific interpretation.

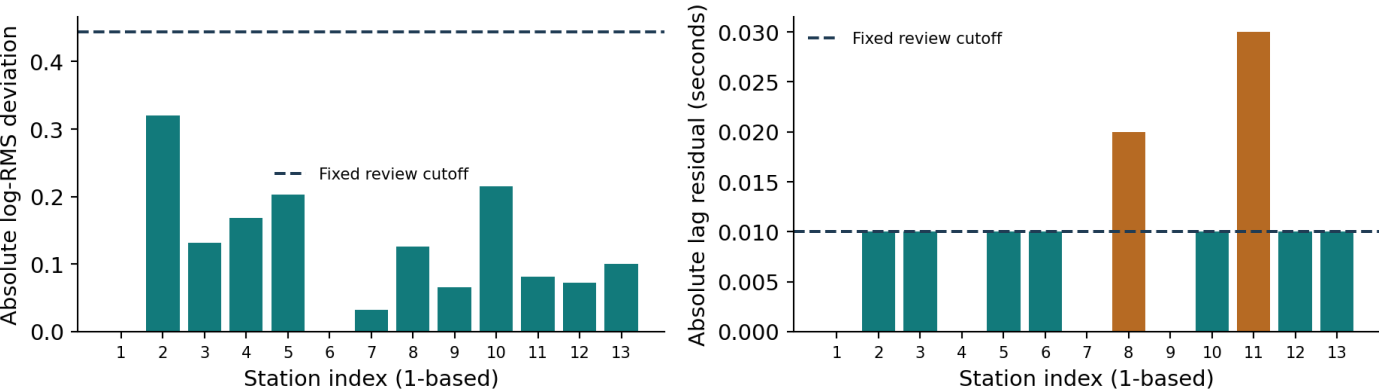
For weak clocks the timing-flag rate is 68/72 (94.4%) (95% interval 86.6-97.8%). On familiar primary no-fault windows, timing flags occur in 2/72 (2.8%) (95% interval 0.8-9.6%). These rates are conditional on this generator and fixed cutoffs; they do not estimate real-world positive predictive value.

Each plotted condition contains one window per independent seed family. Intervals exclude threshold-fit uncertainty and do not establish generalization to a population of arrays. Related rows must not be pooled as independent field observations. All regimes, including changed pulse/noise and combined faults, are retained in the linked full results table.

An analyst review item, with limits visible

Illustrative output from a saved synthetic window

Illustrative saved window: PDAR, first test family, weak clock injection



Example selected as PDAR's first test family, weak-clock condition. It illustrates the evidence an analyst could inspect; it was selected after evaluation and is not a representative-performance claim.

Review field	Saved value or interpretation
Window	PDAR_096_challenge_weak_clock; 13 stations, 30 seconds.
Timing finding	Largest absolute residual: 0.03 s at station index 11; fixed cutoff 0.01 s; review flag raised.
Amplitude finding	Maximum log-RMS deviation: 0.320; below cutoff 0.444.
Known synthetic truth	Injected station index 11; imposed delay -0.0471 s. This truth is available in the experiment only.

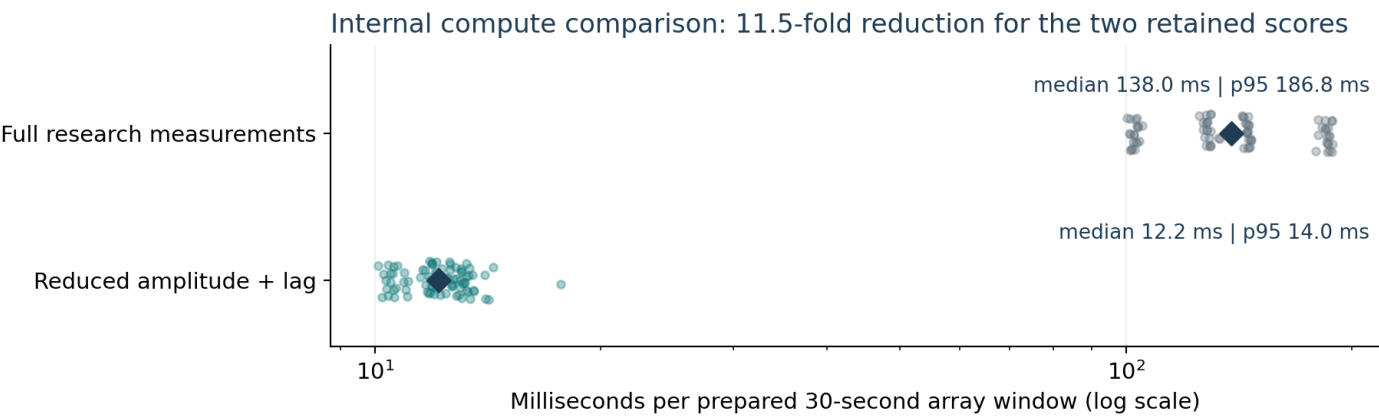
The output should say: **timing inconsistency; inspect this station and window.** An analyst would then examine phase identity, sensor and response metadata, timing status, site behavior, neighboring windows and alternative arrivals. A flag provides evidence for review; it does not establish the physical cause.

Condition without an isolated fault	Either flag	95% interval
Familiar-array primary single wave	4/72 (5.6%)	2.2-13.4%
Withheld WRA primary single wave	5/120 (4.2%)	1.8-9.4%
Changed pulse/noise, single wave	5/72 (6.9%)	3.0-15.2%
Two independent arrivals, WRA	83/120 (69.2%)	60.4-76.7%
Coherent echo, familiar arrays	55/72 (76.4%)	65.4-84.7%

A quiet result does not certify correct data. Real site amplification can resemble a gain change; propagation structure can resemble timing residuals. Four saved ambiguity constructions show that a spatial timing gradient and a changed incident wave can produce identical observations. Independent constraints are needed to distinguish them. No real fault diagnosis or earthquake hazard/early-warning capability has been validated.

A smaller computational component

Internal profiling supports simplification; external cost advantage is untested



Sixteen fixed primary records (family 96; four arrays and four conditions), each repeated five times per implementation. Order alternates. Points are warm compute timings, not independent performance trials or operational latency measurements.

Prepared-window computation	Median	95th percentile
Reduced amplitude + lag	12.2 ms	14.0 ms
Full research measurement function	138.0 ms	186.8 ms

The full/reduced ratio of summed measured compute times is 11.50. The reduced function reproduces both residual vectors and their maxima with maximum absolute difference 0.0 across all 2,496 archived recordings. It also retains all four beam-boundary records and four lag-boundary station readings. The reduced function does not retain the full pipeline's other measurements or capabilities.

Measured on the current macOS arm64 host, with one numerical thread requested per library; the CPU model string was unavailable. Archives were decoded and geometry engines warmed before timing. Download, decoding, filtering, engine setup, report writing and analyst review are excluded. The input is a complete 30-second recording; milliseconds of computation do not imply milliseconds of observation-to-decision latency.

Where economic value would have to appear

The most relevant pilot metric is total effort per correctly resolved review item: analyst time + setup/maintenance + compute/storage + investigation of false flags. Potential savings could come from evidence preparation, traceability or fewer manual steps. No staff-time, hosting-cost, price or return-on-investment measurement has been made.

This timing comparison is against our own broader research function. Established tools may already run comparably fast, and implement different tasks. It establishes that a lighter implementation is possible here; it does not establish a cheaper market offering.

Existing tools and the comparison we owe

Primary documentation checked 22 October 2026; no competitor benchmark executed

Existing option	Documented capability	Implication for this prototype
ISPAQ / MUSTANG [S1]	Seismic quality measurements using local files or FDSN data; local metrics can feed QuARG.	Use as an established metric source/comparator; avoid recreating routine QC.
QuARG [S2]	Find issues, inspect them, track tickets and generate QA reports; configurable thresholds and local ISPAQ input.	Closest workflow comparator. A report generator alone is not a distinct proposition.
SeisComP scqc [S3]	Stream-quality metrics and alarms, including RMS, gaps, latency and digitizer-reported timing quality.	Compare or integrate within an operator's existing network workflow.
ObsPy array_processing [S4]	Public FK beamforming implementation and example workflow.	Independent computational reference; the original diagnostic includes limited known-wave cross-checks.
Cardinal [S5]	Seismic/geoacoustic array-processing code with an adaptive-array stage and example clients.	Relevant comparator for array interpretation and complex arrivals; not yet tested head-to-head.

A possible contribution is a compact, event-window array review packet with measured inter-station residuals, explicit ambiguity and replayable evidence. These ingredients are conventional. Their usefulness as a package, and whether existing tools already supply the needed functionality, must be established with an operator. The documentation review is not evidence that competitors lack an equivalent feature.

Digitizer-reported timing quality and waveform-derived residual lag are different measurements. That distinction motivates a comparison, not a superiority claim. Which one helps depends on the operator's actual failure cases and access to independent timing information.

Integration first: evaluate this as an optional array-specific component alongside an existing QC system. A standalone product should be considered only after a measured workflow gap is established. User access, provenance checks, station mappings and reviewed data preparation are still needed for a supported operational implementation.

Sources: [S1 ISPAQ](#), [S2 QuARG](#), [S3 scqc](#), [S4 ObsPy](#), [S5 Cardinal](#).

The next comparison: operator effort and correctness

A bounded decision about practical usefulness

- 1. Select one real job.** Start with retrospective review of suspicious compact-array windows, if an operator identifies this as costly. Agree on the decision: flag for review, identify an affected station, or confirm a cause. These require different evidence. The present tool supports the first two as hypotheses for testing.
- 2. Build an independently labeled sample.** Use resolved maintenance/timing records and specialist review, along with naturally occurring normal periods, real multiple arrivals and site effects. Keep synthetic injections as separate stress tests. Split complete events, fault episodes, arrays and time periods; hold an untouched final set. Include all windows from a representative monitoring period to estimate operational false-alert burden.
- 3. Compare the actual workflows.** Reference: the operator's current tools, including QuARG/ISPAQ or SeisComP where used. Candidate: the same workflow plus our review packet. A secondary comparison can use simple standalone residual plots. Provide identical recordings and metadata. Counterbalance case order and analyst assignment; avoid teaching an analyst the answer by showing the same case twice.

Measure	Why it decides value
Correctness by fault severity	Recall of independently confirmed actionable problems; correct station/cause distinguished.
False-review burden	Nonactionable flags per station-day or array-day, with episode deduplication and defined exposure.
Decision time and disagreement	Active analyst minutes to a correct disposition, unresolved cases and inter-analyst variation.
Full cost and latency	Acquisition/preprocessing, setup, maintenance, compute and report/review time.
Coverage and failure handling	Missing/dead channels, bad metadata, unfamiliar arrays, mixed arrivals and unresolved physical ambiguity.

- 4. Agree on acceptance before testing.** The operator sets acceptable missed-problem risk, review workload and material effort savings. Size the evaluation around the needed uncertainty; rare failures require sufficient exposure and fault episodes. Use grouped uncertainty estimates and a frozen final test. There is no universal accuracy percentage that substitutes for this agreement.
- 5. Make a stop-or-integrate decision.** Advance only if the added component preserves acceptable decisions and measurably reduces total work, or reveals useful independently confirmed problems at an acceptable review cost. If existing tools already do the job equally well, retain the benchmark and avoid a redundant product. A negative commercial comparison does not erase the research asset's reproducibility value.

Evidence, reproduction and scope

A traceable technical asset, with commercial claims left open

Claim / artifact	Producing script	Saved evidence
Frozen extension and environment	qc_00_prepare.py	QC_FEASIBILITY_PROTOCOL.md; results/qc_feasibility/baseline.json
Threshold flags and station outcomes	qc_01_screen.py	screen.json; screen_predictions.json; screen_table.csv
2,496 raw-score checks and timings	qc_02_profile.py	profile.json; archived diagnostic waveform NPZ files
Four figures and illustrative case	qc_03_assets.py	figures/qc_feasibility/; illustrative_case.json
This brief in three formats	qc_04_publish.py	SEISMIC_QC_BRIEF.md, .html and .pdf
Preservation and replay checks	qc_05_verify.py	results/qc_feasibility/final_checks.json
Original synthetic design/checks	diag_common.py; diag_00_prepare.py; diag_01_measure.py	DIAGNOSTIC_TEST_PROTOCOL.md; results/diagnostic/
Software comparison sources	Primary documentation review	results/qc_feasibility/sources.json

All abbreviated script paths above are under outputs/scripts/; evidence paths are relative to outputs/ unless noted. The [extension lab book](#) records run order, reproduction commands, mistakes and the complete evidence index. The [screen table](#) includes every evaluated regime. The [original diagnostic lab book](#) retains the broader classifier results and limitations.

Scientific runtime: Python 3.14.3; NumPy 2.4.2; SciPy 1.17.0; ObsPy 1.5.0; scikit-learn 1.8.0. Source scripts, versions, per-record outputs, raw archive hashes and protocol hashes are retained. Numerical replay is distinct from independent scientific replication.

Public internet access was used for primary software documentation. No new seismic recordings were downloaded for this extension. Existing inputs are treated as read-only; the frozen Phase 1 audit is preserved. The before/after checksum verification is saved separately with the final artifacts.

Interpretation of the evidence

Observed: the saved recordings, geometry, measurements and existing implementations. **Computed:** fixed-threshold screening, conditional station identification, exact residual replay and internal compute timings. **Inferred:** a lightweight array-review component is technically feasible. **Unestablished:** field fault diagnosis, reliable confidence under changed conditions, specialist time savings, competitive advantage and willingness to pay.

Next commitment: a small operator-led comparison of a real review task. Further feature invention is not a prerequisite. The useful asset is a transparent measurement-and-evidence package whose practical value can be tested directly.