Classifying What Changed Between Two Revisions of a Spectral Dataset

Satyajeeth Suresh Kannan
Innovation Academy · Mary's Lab
github.com/JeetuSK0808/maryslab

Preprint · August 2026 · Mary's Lab technical report TR-12

ABSTRACT

Version control reports that bytes changed. For a spectral dataset that is not the useful question: a re-measurement, a recalibration, and a silent smoothing of a claimed measurement all look identical to a text diff, and they have entirely different implications for anyone citing the data. We classify each changed record by its numerical signature — a least-squares gain-and-offset fit identifies recalibration; a collapse in Savitzky–Golay residual energy identifies smoothing; preserved noise energy on a matching shape identifies re-measurement; anything else is a replacement — and report perceptual impact separately from numerical change. On five constructed revisions with known ground truth the classifier recovers the intended category in 5 of 5 cases. On 200 measured spectra from the USGS Spectral Library it recovers 0.66: recalibration and smoothing hold, while re-measurement is read as replacement and replacement as recalibration. The second failure is a property of the domain, not a threshold — mineral spectra differ mostly in brightness, so swapping the sample genuinely looks like a gain change — and it means a release gate must treat a recalibration verdict as needing confirmation rather than as a pass. We also show why the perceptual and numerical measures must both be reported: they dissociate in both directions.

Keywords: data provenance; spectral datasets; reproducibility; Savitzky–Golay; version control

1. Introduction

Published results rest on datasets, and datasets are revised. A citation names a collection, rarely a version, and the collection changes underneath the citation. When a result fails to reproduce, establishing whether the data moved is often impossible after the fact.

The obstacle is that a curator comparing two revisions of a spectral record sees two arrays of thirty-eight numbers. Whether the difference represents a fresh measurement of the same sample, an instrumental recalibration, a filter applied in place, or a different sample entirely is not visible by inspection — and those four cases warrant four different responses.

2. Method

Records are matched by identifier and each changed pair is classified by signature.

Recalibration is a two-parameter fit: least-squares gain and offset mapping old to new. If that model absorbs nearly all the difference, the change is instrumental rather than physical.

Smoothing is measured in the noise band. We take the residual against a Savitzky–Golay quadratic fit — chosen because a local polynomial fit passes real curvature such as absorption features while retaining per-sample noise, whereas a moving average would smear real structure into the residual and make measured spectra look processed. A collapse in residual energy indicates a filter was applied.

Re-measurement is the complement: comparable residual energy on a near-identical shape. **Replacement** is what remains.

Every matched record additionally carries its perceptual impact as ΔE_{OK} under D65, reported separately from the RMS change.

3. Evaluation

Inputs are synthetic with known ground truth, which is the only way to evaluate a classifier of this kind: we construct five revisions whose category we chose, then ask what the instrument reports.

Table 1. Classification against constructed ground truth. RMS is the numerical change; ΔE_OK is the perceptual change under D65. The two dissociate, which is why both are reported.

record	ground truth	classified as	RMS	ΔE_OK
r0	identical	identical	0	0
r1	renoised	renoised	0.0195	0.0078
r2	smoothed	smoothed	0.036	0.0225
r3	rescaled	rescaled	0.0727	0.0532
r4	replaced	replaced	0.2848	0.3461

The classifier recovers 5 of the five constructed categories. We want to be exact about what a perfect score on five cases does and does not show. The five revisions were built to have clearly separated signatures — a 0.82 gain for the recalibration, two smoothing passes for the filter, a different base curve for the replacement — so this establishes that the signatures separate when they are distinct, and nothing about where the decision boundaries actually lie. A recalibration with a gain of 1.005, or a single light smoothing pass, would sit near a threshold, and this experiment does not report how such a case is handled.

The dissociation between the two reported measures is visible in the table and is the practical argument for reporting both. A change can be numerically large and perceptually invisible, or numerically modest and perceptually obvious. A single "amount changed" figure would collapse these and mislead in whichever direction the reader assumed.

The intended deployment is a release gate rather than a report: failing a release when records were smoothed without declaration, or when appearance moved beyond budget without being declared a replacement, converts a data-handling norm into a check that runs.

Reproducing these numbers

Clone `github.com/JeetuSK0808/maryslab` and run `node papers/build/experiments.mjs`. Every figure in this report is written by that script into `papers/build/results.json`; the instrument itself is exercised by `diff_spectral_datasets`. The engines carry a 401-vector conformance suite shared by a TypeScript reference implementation, a C++20 core, and a WebAssembly build, so a reported number is a property of the specification rather than of one implementation.

4. Real-data evaluation: the same operations on measured spectra

Five constructed cases show the signatures separate when they are made distinct. They say nothing about where the decision boundaries sit on real material spectra, which is what a curator actually faces. We repeated the experiment on 200 records drawn from snapshot `usgs-splib07a-1`, applying the same five operations — and, for the replacement class, substituting a *different real mineral* rather than a synthetic curve, which is what a replacement in a real library looks like.

Table 2. Per-class recall on measured spectra, 40 records per class. The two classes that collapse do so for reasons specific to real material data.

ground truth	n	correct	recall	mistaken for
identical	40	40	1	—
rescaled (recalibration)	40	40	1	—
smoothed	40	37	0.925	renoised (3)
renoised (re-measurement)	40	4	0.1	replaced (36)
replaced	40	11	0.275	rescaled (26)

Overall accuracy falls from 5 of 5 on constructed cases to 0.66 on measured ones (132 of 200). Three classes hold up — *identical* and *rescaled* are perfect, *smoothed* is 0.925 — and two collapse. Both failures are informative rather than random, and both are specific to real data.

Re-measurement is read as replacement in 36 of 40 cases. The test for re-measurement asks for a matching shape carrying a comparable amount of noise. But as TR-13 establishes on this same corpus, published USGS spectra carry almost no per-sample noise at all — so adding a realistic re-measurement jitter multiplies the high-frequency energy rather than matching it, and the record fails the "same noise level" test it was supposed to pass. The classifier was calibrated against data that has a noise floor; this library does not have one.

Replacement is read as recalibration in 26 of 40 cases, and this one is a genuine property of the domain rather than a calibration error. Mineral reflectance spectra differ from one another substantially in overall brightness and comparatively less in shape. Substituting one mineral for another therefore often *is* well approximated by a gain and an offset, which is exactly the signature the recalibration test looks for. On this corpus, "we swapped the sample" and "we recalibrated the instrument" are frequently not distinguishable by a two-parameter fit, no matter how the threshold is set.

That second finding is the one worth carrying forward. It is not fixable by tuning: it says the feature is genuinely ambiguous on material libraries, and a release gate built on it will wave through swapped samples as recalibrations. A gate should therefore treat *rescaled* as requiring human confirmation rather than as an acceptable automatic verdict — the opposite of the reading a clean synthetic result would have encouraged.

Reproducing these numbers

Clone `github.com/JeetuSK0808/maryslab` and run `node papers/build/experiments.mjs`. Every figure in this report is written by that script into `papers/build/results.json`; the instrument itself is exercised by `diff_spectral_datasets`. The engines carry a 401-vector conformance suite shared by a TypeScript reference implementation, a C++20 core, and a WebAssembly build, so a reported number is a property of the specification rather than of one implementation.

5. Limitations

This section states what the result does not establish. It is written to be read before the abstract is quoted.

- **Signature matching is not chain of custody.** Verdicts come from thresholded numerical fingerprints. A smoothed curve with realistic synthetic noise added back will read as a re-measurement. The instrument detects careless changes reliably and deliberate disguise not at all.
- **Five cases is a demonstration, not a validation.** The evaluation establishes that the signatures separate on constructed examples with known truth. It does not establish accuracy on a real corpus of revisions, which would require a corpus with labelled history that we do not have.
- **Perceptual impact is under D65 for the standard observer only.** A change invisible there may matter under another light.
- **Matching is by identifier.** A renamed record appears as one removal and one addition; renames are not tracked.
- **Both revisions must sit on the canonical grid.** Comparing differently-sampled data would require resampling, which is itself a change, so the tool refuses rather than silently introducing one.
- **Prior art was surveyed informally.** Data versioning tools are numerous; a search found none classifying spectral revisions by physical signature.

6. Acknowledgements and disclosure

The instruments described here were implemented in Mary's Lab, an open-source computational color science project. Software design, implementation, and the preparation of this report were carried out with substantial assistance from a large language model (Anthropic Claude), under the author's direction and review. All numerical results were produced by executing the released code; none were generated by a language model. The author is responsible for the claims made here.

References

1. Savitzky, A. and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639.
2. Ottosson, B. (2020). *A perceptual color space for image processing (Oklab)*. bottosson.github.io.
3. CIE (2004). *Colorimetry, 3rd edition*. CIE 15:2004. Commission Internationale de l'Éclairage, Vienna.
4. Wyszecki, G. and Stiles, W. S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd edition. Wiley.
5. Schanda, J. (ed.) (2007). *Colorimetry: Understanding the CIE System*. Wiley-Interscience.
6. Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.
7. Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226–1227.

Source, engines, and the script that produced every number in this report: `github.com/JeetuSK0808/maryslab`. Results regenerate with `node papers/build/experiments.mjs`. Inputs: synthetic where §3 says so, and measured USGS splib07a spectra (snapshot `usgs-splib07a-1`) where a real-data section says so. Prepared with AI assistance (see Acknowledgements).