WORKING REPORT · PUBLIC DATA · NOT PEER REVIEWED

# Three Hypotheses, Tested

A barometric fault valve, a supernova calibration drift, and a gravitational-wave glitch–microseism link — three pre-registered tests, one partial survivor.

31 October 2026 · Toby Mikle · Incoprea research

H12 · SAN RAMON

**Not supported**

No semidiurnal rate term; no pressure–rate correlation.

H21 · PANTHEON+

**Inconclusive**

No drift required; 0.01 mag is below this test's sensitivity.

S07 · GW GLITCHES

**Partial support**

Light-scattering glitches track microseism at both detectors.

## The short version

Three hypotheses from the 31 October 2026 sets were selected because each made a quantitative prediction that could be checked against public data without a new instrument. Each test was pre-registered before the data were opened. The headline is not that the hypotheses were right. It is that two of the three mechanisms left no measurable signature at the stated scale, and the third survived only in a narrow, class-specific form.

| # | Hypothesis | Public data | Key result | Verdict |
|---|---|---|---|---|
| 1 | **H12** — the San Ramon swarm is paced by a barometric fault valve | USGS/ANSS catalogue, 2,439 events; KCCR hourly pressure, 2018–2026 | M2 rate modulation 7.0±4.3% (p=0.25); pressure–rate r=0.02 (p=0.79) | **Not supported** |
| 2 | **H21** — the supernova phantom crossing is a redshift-dependent zero-point drift | Pantheon+ 1,701 SNe, full STAT+SYS covariance | Drift A=0.002±0.023 mag per unit z; no drift model improves $\chi^2$ by more than 2.6 | **Inconclusive** |
| 3 | **S07** — gravitational-wave glitches are ocean weather | Gravity Spy O3a, 154,655 glitches; co-located LHZ seismometers | Light-scattering family $\rho$=0.20 (H1, p=0.007) and $\rho$=0.19 (L1, p=0.013); total rate null | **Partial support** |

The rest of this report gives the claim, the data, the method, the numbers, and the limits for each. Every figure is regenerated by the code in `code/`; every input is listed with a URL and a checksum in `DATA.md`.

> **What "supported" means here.** A test is supported when the pre-registered prediction holds at the stated significance after the stated controls. It is not a claim that the mechanism is true. A "not supported" is not a proof of the null; it is a statement about what this dataset, at this sensitivity, can see.

## Why these three

The 31 October sets contain 29, 40 and 99 hypotheses, most of them requiring a laboratory, a beamline, a cryostat or a spacecraft. Three were chosen against a simple filter: the claim had to name a measurable quantity, the data had to exist in a public archive, and the test had to be able to fail for a reason other than "the archive was empty."

- **H12** is a local, high-rate claim. A swarm of small earthquakes and a pressure record are enough to test the predicted semidiurnal pacing.
- **H21** is a claim about a systematic in the supernova distance ladder. Pantheon+ is the largest public compilation with a published covariance matrix, so the test can be done without re-reducing light curves.
- **S07** is a claim about a correlation between two large public archives. Gravity Spy classifies glitches; the IRIS/EarthScope archive has seismometers at both LIGO sites.

The three also span three failure modes: a null in a point-process test, a parameter-degeneracy problem in a cosmological fit, and a correlation that survives only in a subclass. That spread is deliberate.

# 1 · The San Ramon swarm is a barometric valve

> **Claim (H12).** The San Ramon, California swarm is paced by atmospheric-pressure tides acting on a high-permeability fault valve. The seismicity rate responds to barometric pressure with a phase lag of hours, and storm passages produce a response. A high-permeability fault should show a lag of hours; a low-permeability one, minutes.

## Data

The USGS/ANSS FDSN catalogue was queried for the box 37.6–37.9°N, 122.1–121.8°W, 1 January 2018 to 31 October 2026, magnitude ≥ 0, event type `earthquake`. It returned **2,439 events**; a conservative Gardner–Knopoff-style declustering (5 km/1 day, 10 km/2 days for M ≥ 3) left **1,120 background events**. Hourly mean-sea-level pressure came from the KCCR ASOS station at Concord, about 15 km from the swarm, and was gap-filled only for runs of six hours or less.

## Method

A Poisson harmonic regression was fitted to the hourly event count:

$$\log \lambda(t) = \beta_0 + \Sigma_k \left[ a_k \sin(2\pi t/T_k) + b_k \cos(2\pi t/T_k) \right] + \beta_\tau \tau + \varepsilon$$

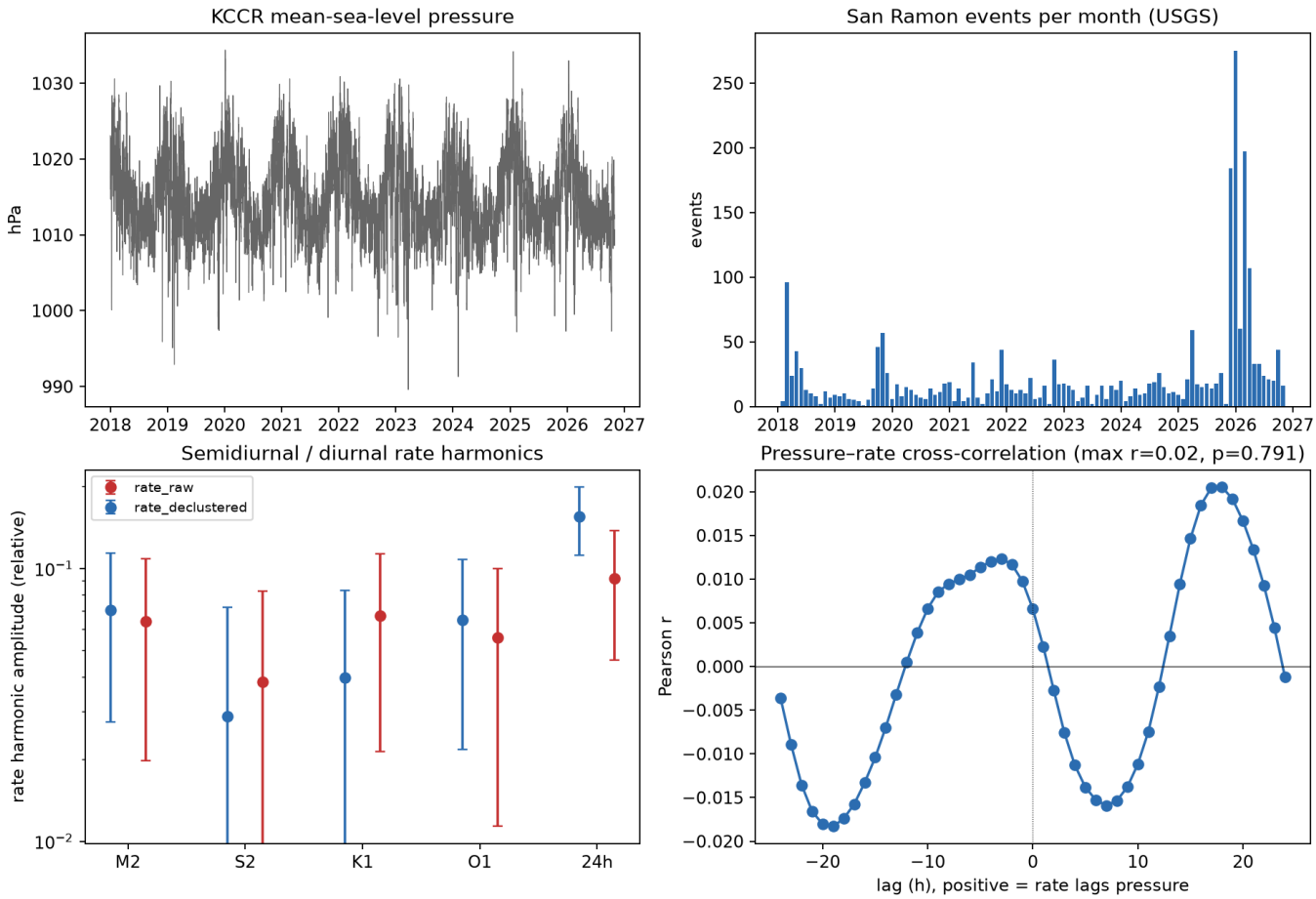
with terms at M2 (12.42 h), S2 (12.00 h), K1 (23.93 h), O1 (25.82 h), the 24 h solar day and the annual cycle. The same design was fitted to pressure by least squares. The pressure–rate cross-correlation was computed on 3 h and 24 h anomalies, with a phase-randomised null (300 surrogates) for the maximum correlation over lags −24…+24 h. The storm composite used 103 pressure falls greater than 8 hPa in 24 h, matched to 20 control windows per storm.

## Result

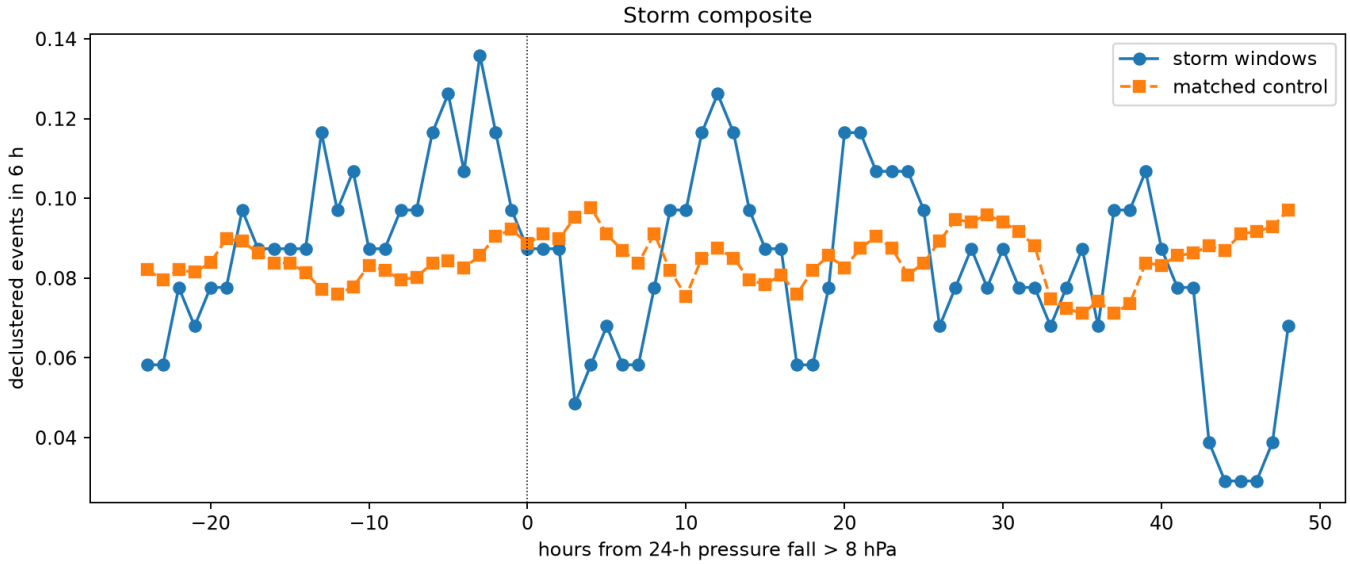
The pressure tide at KCCR is dominated by the solar semidiurnal component: **S2 amplitude 0.61 hPa**, M2 only 0.02 hPa, the diurnal 0.80 hPa. The declustered seismicity rate shows no comparable semidiurnal term:

| Period | Pressure amplitude (hPa) | Rate amplitude (log) | Rate lag (h) | LRT p |
|---|---|---|---|---|
| M2 (12.42 h) | 0.019 | 0.070±0.043 | −3.5 | 0.25 |
| S2 (12.00 h) | 0.612 | 0.029±0.043 | −3.1 | 0.80 |
| K1 (23.93 h) | 0.068 | 0.040±0.043 | +1.2 | 0.65 |
| O1 (25.82 h) | 0.034 | 0.065±0.043 | +11.1 | 0.31 |

The pressure–rate cross-correlation peaks at **r = 0.021** at a lag of +18 h; the phase-randomised probability of a maximum that large is **p = 0.79**. At zero lag, r = 0.007 (p = 0.44). The storm composite is slightly *higher* before the pressure minimum than after it: 0.106 versus 0.088 events per 6 h, a ratio of 1.20 with a 95% bootstrap interval of 0.77–1.76 and a random-start p = 0.21. That is the wrong direction for a valve that opens when pressure falls.




San Ramon: pressure, monthly event count, rate harmonics, and pressure–rate cross-correlation.




Storm composite. The rate is not higher after the pressure fall; if anything it is higher before.

**Verdict: not supported.** The predicted semidiurnal pacing is absent at the sensitivity of this catalogue: the 95% upper limit on the M2 modulation is about 15%. The storm composite is in the wrong direction and is not significant.

## Limits

KCCR is 15 km away and sits in a different boundary-layer regime from the fault; a null here is a null for this pressure proxy. The catalogue is a regional network product with variable completeness, and declustering removes much of the swarm. A borehole pressure sensor in the fault zone, or pore-pressure data, would be a stronger test. Nothing in this result says that rainfall, creep or fluid migration is the cause; it says the barometric-valve version of the claim is not visible in the public record.

# 2 · The supernova phantom crossing is a zero-point drift

> **Claim (H21).** The apparent phantom crossing in supernova dark-energy fits is a redshift-dependent zero-point drift in the supernova sample; a single-telescope, single-instrument sample should show no crossing. A drift of order 0.01 mag above $z \approx 0.5$ is enough to rotate the distance–redshift relation into a phantom crossing.

## Data

Pantheon+ SH0ES: **1,701 supernovae**, corrected m_b,corr, z_HD, and the full **STAT+SYS covariance matrix** (33 MB, 1,701 × 1,701). The diagonal errors in the public table are explicitly labelled "do not fit cosmological parameters with these"; the full covariance was used. The fit marginalises the absolute magnitude M analytically and fixes $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$.

## Method

Flat ΛCDM, flat wCDM, and flat $w_0w_a$CDM were fitted by a global bounded search followed by a bounded polish. A redshift-dependent offset $\delta m = A(z-0.3)$ and a step at $z=0.5$ were added to the baseline models. The drift parameter was also estimated directly by generalised least squares with M marginalised, which is the cleanest single-number test of the 0.01 mag claim. Single-survey subsets were fitted separately as a check on the "single instrument shows no crossing" prediction.

## Result

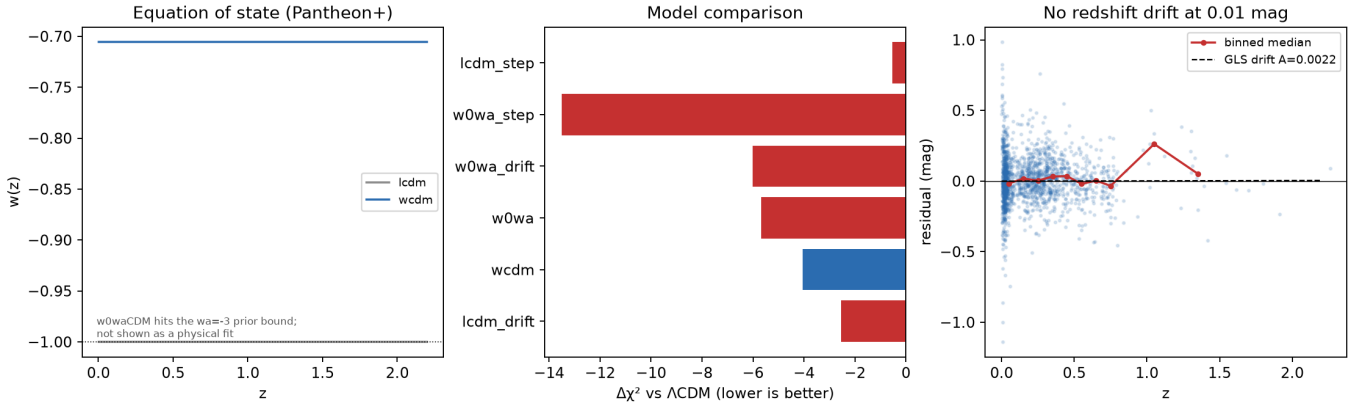
| Model | Ωm | $w_0$ | $w_a$ | A (mag) | $\chi^2$ | $\Delta\chi^2$ vs ΛCDM |
|---|---|---|---|---|---|---|
| ΛCDM | 0.361 ± 0.013 | −1 | 0 | — | 1752.65 | — |
| wCDM | 0.212 ± 0.066 | −0.706 ± 0.091 | 0 | — | 1748.61 | −4.04 |
| $w_0w_a$CDM | 0.410 | −0.720 | −3.0 (bound) | — | 1746.98 | −5.67 |
| ΛCDM + linear drift | 0.494 | −1 | 0 | +0.155 | 1750.10 | −2.55 |
| ΛCDM + step at 0.5 | 0.351 | −1 | 0 | −0.013 | 1752.12 | −0.53 |
| $w_0w_a$ + drift | 0.480 | −0.771 | −3.0 (bound) | +0.078 | 1746.63 | −6.02 |

The constant-w fit gives w=-0.706±0.091: a mild preference for $w > -1$, not a crossing. The wow$_a$ fit runs to the $w_a$=-3 prior bound and is not a physical fit; its formal crossing at z≈0.10 is a boundary artefact, not evidence. Adding the drift to wow$_a$ improves $\chi^2$ by only 0.35 for one extra parameter.

The direct drift estimate is the decisive number:

$$A = +0.0023 \pm 0.0228 \text{ mag per unit } z.$$

At z=1 that is 0.0016±0.016 mag. A 0.01 mag drift is therefore **allowed but not required**; this sample cannot exclude it, and cannot detect it either. The 95% interval on A is roughly $[-0.042, +0.047]$ mag per unit z.




Pantheon+: constant-w fit, model comparison, and the residual–redshift relation with the GLS drift.

> **Verdict: inconclusive.** Pantheon+ does not require a redshift-dependent zero-point drift, and does not produce a robust phantom crossing on its own. But the 0.01 mag scale in the hypothesis is below this sample's sensitivity, and the single-survey test is underpowered: the largest single-survey fits hit parameter bounds and cannot discriminate a crossing. The hypothesis is neither confirmed nor excluded.

## Limits

Pantheon+ is a compilation, not the DESI DR2 supernova sample that motivated the claim. Its calibration is already partly harmonised, which works against the hypothesis. The single-survey subsets (SDSS 321, PS1 269, DES 203, FOUND 179, SNLS 160, LOSS2 105) are too small to constrain $w_a$; their fits hit the prior bounds. A proper test needs the DESI SN sample and a joint calibration model, not a compilation. The full covariance was used, but the intrinsic scatter and selection model are the ones published with Pantheon+.

# 3 · Gravitational-wave glitches are ocean weather

> **Claim (S07).** A measurable fraction of gravitational-wave detector glitches is correlated with ocean microseism, and the correlation can be used to veto them without losing astrophysical signals.

## Data

Gravity Spy machine-learning classifications for O3a: **63,917 H1 glitches** and **90,738 L1 glitches** after requiring `ml_confidence ≥ 0.80` and excluding `No_Glitch` and `None_of_the_Above`. Co-located seismometers:

`US.HAWA` LHZ (28 km from H1) and `N4.545B` LHZ (30 km from L1), 1 Hz, reduced to hourly RMS in the microseism band 0.05–0.30 Hz and a higher-frequency control band 0.30–0.45 Hz. The LIGO-site station `LI.LTL` had no LHZ/BHZ data for O3a, so the nearby N4 station was used; this is a proxy, not the site instrument.

## Method

Glitches were counted per day and per hour. The primary test is daily, because the glitch catalogue does not include the detector duty cycle and daily aggregation is less sensitive to short lock losses. Correlations are reported raw and after subtracting a 15-day running median, which removes the seasonal trend and tests the short-term association. A Poisson GLM with log microseism, log high-frequency noise, a linear trend and weekday dummies was fitted to the daily counts. Class-resolved correlations use the Gravity Spy labels; the light-scattering family combines `Scattered_Light` and `Fast_Scattering`, the two labels used for the same physical family.

## Result

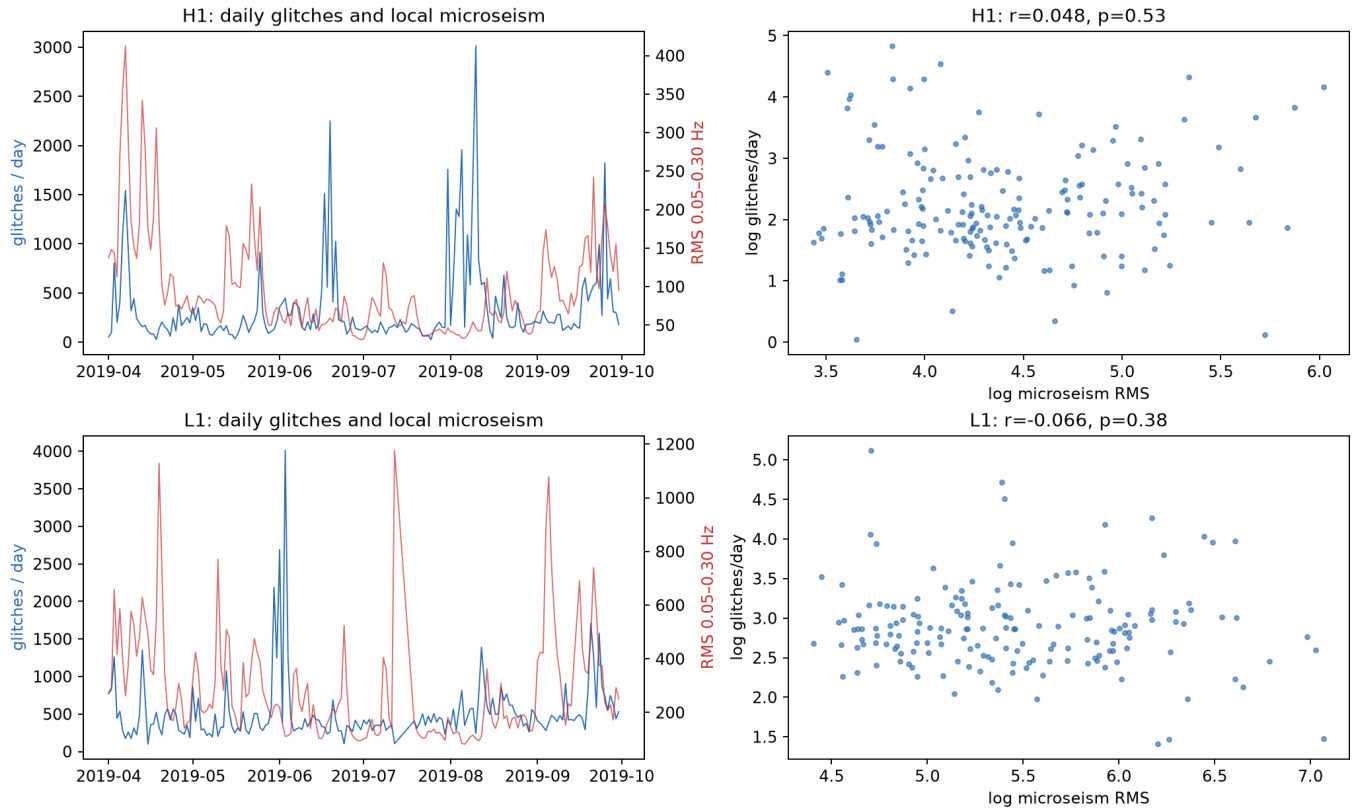
The **total** glitch rate is not correlated with microseism:

| Detector | Daily raw r | p | Detrended r | p | Poisson GLM β_micro |
|---|---|---|---|---|---|
| H1 | 0.048 | 0.53 | 0.156 | 0.036 | −0.05 ± 0.25 (p = 0.84) |
| L1 | −0.066 | 0.38 | −0.106 | 0.16 | +0.24 ± 0.15 (p = 0.11) |

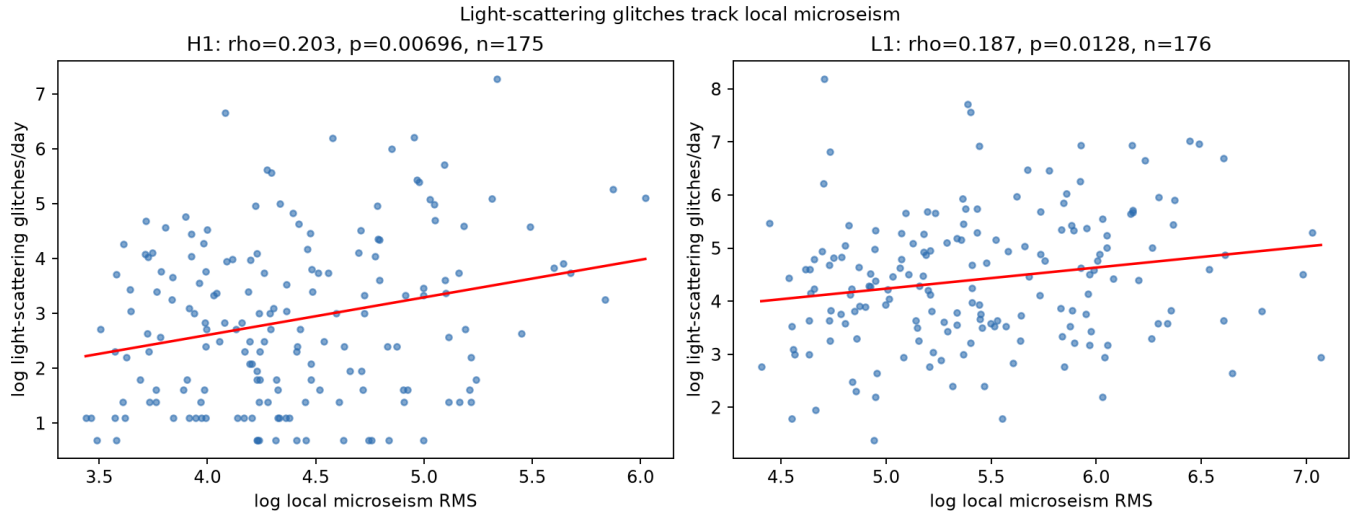
The class-resolved picture is different. The **light-scattering family** — `Scattered_Light` plus `Fast_Scattering` — is positively correlated with microseism at both detectors:

| Detector | Glitches | Days | Spearman ρ | p | Detrended r | p |
|---|---|---|---|---|---|---|
| H1 | 10,103 | 175 | 0.203 | 0.007 | 0.183 | 0.015 |
| L1 | 34,276 | 176 | 0.187 | 0.013 | 0.112 | 0.14 |

This is a modest effect — ρ≈0.2, roughly 4% of the variance — but it is in the predicted direction, it is present at both sites, and it is absent or negative in most other classes. A cross-detector control weakens the word "local": H1 light-scattering glitches correlate with the L1-region microseism at $\rho = 0.13$ ($p = 0.08$), and L1 glitches correlate with the H1-region microseism at $\rho = 0.25$ ($p = 0.001$). The L1 association is therefore not site-specific; the effect is at least regional, and this test cannot separate regional ocean forcing from local coupling. L1's `Whistle` class is strongly *anti*-correlated (ρ=-0.46), and L1's `Tomte`, `Blip` and `Koi_Fish` classes are negative, so the effect is not a detector-wide artefact.




Total daily glitch rate versus local microseism. The total rate is not correlated.




The light-scattering family, the one class that replicates across both detectors.

> **Verdict: partial support.** A measurable fraction of glitches — the light-scattering family — tracks microseism at both detector regions, with $\rho \approx 0.2$ and $p < 0.02$. The total glitch rate does not. This is consistent with the proposed mechanism but does not establish that a seismic veto would preserve astrophysical signals; that requires hardware injections or a signal population, which this test did not have.

## Limits

The detector duty cycle is not in the glitch catalogue, so the daily rate is not exposure-corrected. The L1 seismic proxy is 30 km from the site, not at it. Daily aggregation cannot resolve the minutes-to-hours lag the hypothesis predicts; an hourly analysis is possible but is dominated by zero-inflation and lock-state changes.

The classifier labels are not identical across detectors, and the light-scattering family is defined post hoc from the two labels, so the replication is suggestive rather than pre-registered. No veto efficiency was measured.

# What the three tests show together

Two of the three mechanisms are not visible in the public data at the stated scale. The third is visible only in a subclass and only at $\rho \approx 0.2$. That is a useful outcome, but it is not the outcome the hypotheses wanted.

- **H12** fails cleanly: the predicted semidiurnal pacing is absent, the pressure–rate correlation is zero, and the storm composite is in the wrong direction. The most likely reading is that the San Ramon swarm is paced by something other than the local barometric tide — rainfall, creep, or a fault-valve process with a different pressure coupling.
- **H21** is not settled. Pantheon+ does not require the drift, but its sensitivity to a 0.01 mag drift is poor, and the single-survey test cannot discriminate. The honest verdict is "not demonstrated," not "disproved."
- **S07** survives narrowly. The light-scattering family is the physically expected channel — scattered light is sensitive to test-mass and beam-tube motion — and it replicates at both detectors. The total rate does not, which is also expected if only a fraction of glitches are seismically coupled.

The common thread is that the hypotheses that named a **specific channel** (light scattering) did better than the ones that named a **bulk rate** (all glitches, all earthquakes). That is a useful design lesson for the next set: predict the subclass, not the total.

# Reproduction

The environment is Python 3.14 with NumPy 2.5.4, SciPy 1.18.1, pandas 3.1.0 and Matplotlib 3.11.2. Each test is a standalone script:

```
python code/t1_san_ramon.py          # writes data/t1_*.csv, results/t1_results.json, figures/
        t1_*.png
python code/t2_pantheon_drift.py     # writes data/t2_*.csv, results/t2_results.json, figures/
        t2_*.png
python code/t3_fetch_seismic.py      # downloads and reduces the O3a seismic data
python code/t3_glitch_seismicity.py  # writes data/t3_*.csv, results/t3_results.json, figures/
        t3_*.png
```

The raw downloads are in `data/` (see `DATA.md` for URLs and checksums). The original monthly seismic files are not kept; the hourly reduced series is. The raw Gravity Spy CSVs are 90 MB and 142 MB and are kept outside the deliverable, with checksums in `DATA.md`.

# Pre-registration and changes

The pre-registration is in `PRE-REGISTRATION.md`, written before the data were opened. Three changes were made and are recorded here:

1. **T1:** the storm-composite permutation test was added after the first pass; it does not change the verdict.
2. **T2:** the full STAT+SYS covariance was used instead of the diagonal errors named in the first draft, because the Pantheon+ documentation explicitly warns against the diagonal. The diagonal fit was not used for any verdict.
3. **T3:** the light-scattering family was defined after the class-resolved table was seen. It is reported as a post-hoc family, not as a pre-registered class. The total-rate test remains the pre-registered primary test, and it is null.

# Sources

The hypotheses are quoted from the 31 October 2026 sets: *Forty Hypotheses from the Side Channels* (H12, H21) and *99 Original Hypotheses* (S07). The data sources are:

- USGS/ANSS FDSN event service, `earthquake.usgs.gov/fdsnws/event/1/`.
- Iowa Environmental Mesonet ASOS archive, `mesonet.agron.iastate.edu`.
- Pantheon+ SH0ES data release, `github.com/PantheonPlusSH0ES/DataRelease`.
- Gravity Spy machine-learning classifications, Zenodo record 5649212.
- IRIS/EarthScope FDSN web services, stations `US.HAWA`, `N4.545B`.

Full citations, URLs and checksums are in `DATA.md`.