No Causal Link Between Galactic Cosmic-Ray Flux and Global Seismicity: A Pre-Registered Replication with GPU-Accelerated Surrogate Testing and Out-of-Sample Validation

J. D. Devine[1]

[1] Independent researcher

devine.jd@gmail.com

April 2026

Abstract

Homola et al. [2023] reported a statistically significant positive correlation ($r \approx 0.31$) between galactic cosmic-ray (CR) flux measured by neutron monitors and global seismicity ($M \geq 4.5$) at a time lag of $\tau = +15$ days, suggesting that elevated CR flux precedes increased earthquake activity. We present a systematic replication and extension of this claim using data from 44 Neutron Monitor Database (NMDB) stations, the USGS global earthquake catalogue, and SILSO daily sunspot numbers spanning 1976–2025.

Our analysis proceeds in four stages. *Stage 1* replicates the raw cross-correlation ($r(+15\,\mathrm{d}) = 0.310$, peak $r = 0.469$ at $\tau = -525$ days) but demonstrates that naive p-values are invalid because temporal autocorrelation and a shared ∼11-year solar cycle inflate the apparent significance. *Stage 2* applies iterative amplitude-adjusted Fourier transform (IAAFT) surrogate tests with 10^4 realisations: after Hodrick–Prescott (HP) detrending to remove the solar-cycle component, the peak correlation drops to $r = 0.131$ and achieves marginal significance ($p_{\mathrm{global}} < 10^{-3}$, 3.9σ), but $r(+15\,\mathrm{d}) = 0.041$ — well within the surrogate null distribution. *Stage 3* scans $34 \times 207 = 7{,}037$ station–grid-cell pairs for geographic localisation; 455 pairs survive Benjamini–Hochberg correction (expected false discoveries: 352), and the optimal lag τ^* shows no dependence on great-circle distance ($\beta = -0.45$ days/1000 km, $p = 0.21$), consistent with an isotropic CR signal rather than a local mechanism. *Stage 4* applies a pre-registered out-of-sample test on an independent 2020–2025

window ($T = 390$ five-day bins, 10^5 phase-randomisation surrogates on a Tesla M40 GPU): $r(+15\,\mathrm{d}) = 0.045$ (surrogate 95th percentile = 0.136), $p_{\mathrm{global}} = 0.994$ — consistent with noise. Fitting a sinusoid to the 1976–2025 annual cross-correlation timeseries yields a best-fit period of $P = 9.95$ years and a Bayes factor of BF = 27.5 strongly favouring a solar-cycle modulation over a constant relationship.

We conclude that the CR–seismic correlation reported by Homola et al. [2023] is an artefact of the shared solar-cycle modulation of both galactic CR flux and global seismicity, and not evidence of a physical causal link. All analysis code, pre-registration document, and results are publicly available at https://github.com/pingud98/cosmicraysandearthquakes.

Keywords: cosmic rays; seismicity; surrogate test; solar cycle; Benjamini–Hochberg; pre-registration; out-of-sample validation.

Contents

1 Introduction

The hypothesis that galactic cosmic rays (CRs) influence seismic activity has a long history in geophysics [Stoupel, 1990, Urata and Tanimoto, 2018], motivated by proposed mechanisms ranging from radon ionisation in fault zones to direct nuclear interactions in crustal minerals. Homola et al. [2023] recently presented observational support for this idea, reporting a correlation coefficient $r \approx 0.31$ between a global CR index constructed from NMDB neutron monitor records and a global seismic energy metric derived from the USGS earthquake catalogue at a lag of $\tau = +15$ days (CR leads seismic activity). The associated naive p-value was reported as $p \sim 10^{-72}$ at this lag.

Such a claim, if correct, would be of profound scientific and societal importance, potentially enabling short-term earthquake forecasting from space-weather observations. It therefore demands rigorous scrutiny. Three statistical pitfalls immediately suggest themselves:

1. **Temporal autocorrelation.** Both CR flux and seismicity exhibit strong low-frequency structure (solar cycle, regional seismic cycles). Treating successive 5-day bins as independent dramatically inflates the degrees of freedom; a Bretherton effective-N correction [Bretherton et al., 1999] is required.

2. **Shared solar-cycle trend.** Galactic CR flux is modulated by the heliospheric magnetic field, which varies on an ∼11-year solar cycle [Potgieter, 2013]. Global seismicity has also been reported to correlate weakly with solar activity [Odintsov et al., 2006, Tavares and Azevedo, 2011], potentially generating a spurious correlation between the two series with a lag structure determined by the phase relationship of their respective solar responses, not by any direct physical mechanism.

3. **Multiple-comparison inflation.** Testing 401 lag values and selecting the maximum creates a look-elsewhere effect that must be accounted for by comparing the observed peak against a null distribution of peak statistics, not against the single-lag Pearson t distribution.

This paper systematically addresses all three issues, extending the analysis through a prospective out-of-sample validation window (2020–2025) whose statistical predictions were pre-registered in a timestamped git commit before any data in that window were examined.

The remainder of the paper is organised as follows. Section 2 describes the data sources and preprocessing. Section 3 presents the statistical methods. Section 4 reports the results of each analysis stage. Section 5 interprets the findings. Section 6 concludes.

2 Data

2.1 Cosmic-Ray Flux: NMDB Neutron Monitors

Galactic cosmic-ray flux is measured by neutron monitors (NMs), which detect secondary neutrons produced when primary CRs interact with atmospheric nuclei. We obtained pressure-corrected hourly count rates for all available stations from the Neutron Monitor Database (NMDB, `https://www.nmdb.eu`) for the period 1976–2025.

After applying a coverage filter (requiring $\geq 60\%$ hourly data per day to declare a daily bin valid), we retained **44 stations** with $\geq 50\%$ daily coverage over the in-sample window 1976–2019, and **35 stations** over the out-of-sample window 2020–2025. Each station's daily series was normalised by its long-run mean and resampled to non-overlapping 5-day bins. A global CR index was formed as the mean across all stations with valid data in each bin, requiring at least three stations; bins failing this criterion were set to `NaN`.

2.2 Seismic Activity: USGS Earthquake Catalogue

Earthquake data were downloaded from the USGS Earthquake Hazards Programme via the FDSN web service [USGS Earthquake Hazards Program, 2024]. We retained all events with $M \geq 4.5$ globally, yielding a catalogue of $\approx$47,860 events over 2020–2025 in the out-of-sample window alone. The seismic metric for each 5-day bin is the total released seismic moment (expressed as summed M_W), which is proportional to the logarithm of total energy released and is more physically motivated than raw event count.

2.3 Solar Activity: SIDC Sunspot Number

The SILSO international sunspot number [SILSO World Data Center, 2024] provided the solar activity index used to remove the solar-cycle trend. We used the daily series (version 2.0), smoothed with a 365-day running mean for detrending purposes.

3 Methods

3.1 Cross-Correlation at Lag τ

Let x_t denote the global CR index and y_t the seismic metric in 5-day bin t, with $t = 1, \ldots, T$. The normalised cross-correlation at lag k (bins) is

$$r(k) = \frac{1}{n\,\hat{\sigma}_x \hat{\sigma}_y} \sum_{t=1}^{n} \tilde{x}_t\, \tilde{y}_{t+k}, \tag{1}$$

where $\tilde{x}_t = x_t - \bar{x}$, $n = T - |k|$, and the sums run over the valid overlap region. A positive lag $k > 0$ corresponds to CR leading seismicity. Lags range from -200 to $+200$ days (step

= 5 days, i.e. 1 bin), giving 81 lag values in the in-sample window.

3.2 Effective Degrees of Freedom

Because both x and y are autocorrelated, the effective sample size N_{eff} is substantially smaller than T. We use the Bretherton et al. [1999] formula

$$N_{\text{eff}} = T \left(1 + 2 \sum_{k=1}^{K} r_{xx}(k)\, r_{yy}(k) \right)^{-1}, \tag{2}$$

where r_{xx} and r_{yy} are the sample autocorrelation functions of x and y, truncated at lag K where the product first changes sign.

3.3 Surrogate Significance Tests

To correctly account for autocorrelation and multiple lags simultaneously we use surrogate time-series methods [Theiler et al., 1992, Schreiber and Schmitz, 2000].

3.3.1 Phase Randomisation

Phase surrogates of x are constructed by multiplying the discrete Fourier transform of x by random unit-magnitude complex numbers (with conjugate symmetry to preserve real-valuedness):

$$\tilde{X}(\omega_k) = |X(\omega_k)|\, e^{i\phi_k}, \quad \phi_k \sim \mathcal{U}(0, 2\pi), \tag{3}$$

followed by the inverse DFT. This preserves the power spectrum (and hence autocorrelation structure) of x while destroying any phase relationship with y.

3.3.2 IAAFT Surrogates

Iterative amplitude-adjusted Fourier transform (IAAFT) surrogates [Schreiber and Schmitz, 2000] additionally preserve the amplitude distribution of x by alternating between power-spectrum matching (in Fourier space) and rank-order resampling (in time domain) until convergence. IAAFT surrogates are more conservative than phase surrogates when x has a non-Gaussian distribution.

3.3.3 Global p-Value

For each surrogate $s = 1, \ldots, S$, we compute the peak cross-correlation $\rho_s = \max_k |r_s(k)|$ across all tested lags. The global p-value is

$$p_{\text{global}} = \frac{\#\{s : \rho_s \geq \rho_{\text{obs}}\}}{S}, \tag{4}$$

where ρ_{obs} is the observed peak. This test is simultaneously valid for all lags and all lag-selection rules, eliminating the multiple-comparison problem.

3.3.4 GPU Acceleration

With $S = 10^5$ surrogates and $T \approx 3{,}200$ bins, direct CPU computation would require $\sim$3 h. We vectorise the surrogate generation and cross-correlation evaluation over all S realisations simultaneously using CuPy on an NVIDIA Tesla M40 (12 GB VRAM). For the geographic scan (Section 3.5), all N_{cells} seismic cell series are evaluated in a single GPU matrix multiply per lag:

$$\mathbf{R}_{\text{lag}} = \frac{1}{n} \mathbf{X}_{\text{surr}}^{(z)} \left(\mathbf{Y}^{(z)}\right)^{\top} \in \mathbb{R}^{S \times N_{\text{cells}}}, \tag{5}$$

where rows of $\mathbf{X}_{\text{surr}}^{(z)}$ are standardised surrogates and columns of $\mathbf{Y}^{(z)}$ are standardised seismic cell series. Benchmarks show a 2.9× speedup for phase surrogates and 1.3× for IAAFT (limited by chunked argsort to avoid VRAM overflow).

3.4 Solar-Cycle Detrending

We apply three complementary detrending approaches to isolate the CR–seismic relationship from the shared solar-cycle trend:

1. **Hodrick–Prescott (HP) filter** [Hodrick and Prescott, 1997] with smoothing parameter $\lambda = 1.29 \times 10^5$ (calibrated for quarterly data, rescaled for 5-day bins to retain oscillations shorter than $\sim$3 years). The trend component is subtracted from both x_t and y_t.

2. **STL decomposition** [Cleveland et al., 1990]: seasonal-trend decomposition using LOESS, applied independently to each series.

3. **Sunspot regression**: residuals after regressing each series on the 365-day smoothed sunspot number and its 12-month lag.

For the out-of-sample window ($\sim$5 years, less than one solar cycle), the HP filter is inappropriate (it would remove any genuine sub-decadal signal); we use linear detrending instead, as pre-specified in the pre-registration.

3.5 Geographic Localisation Scan

If CRs cause earthquakes via a local mechanism, the optimal lag $\tau^*(s, g)$ for station s and grid cell g should increase with their great-circle distance $d(s, g)$ (propagation delay). Under the null hypothesis of global CR isotropy, τ^* should be distance-independent.

We define a 10ř × 10ř longitude–latitude grid (648 cells total), retain cells with ≥ 100 events, and for each of the 34 × 207 = 7,037 station–cell pairs compute the peak cross-correlation $r^*(s, g)$ and optimal lag $\tau^*(s, g)$ using GPU-accelerated phase surrogates (1000 realisations).

Pairs are declared significant at false discovery rate $q = 0.05$ using the Benjamini–Hochberg (BH) procedure [Benjamini and Hochberg, 1995]: rank the $m = 7{,}037$ p-values $p_{(1)} \leq \ldots \leq p_{(m)}$; the threshold is $p_{(k)} \leq (k/m) \times q$ for the largest k satisfying this condition. Distance dependence of τ^* is tested by ordinary least-squares regression of $\tau^*(s, g)$ on $d(s, g)$.

3.6 Pre-Registered Out-of-Sample Validation

To guard against post-hoc hypothesis adjustment, we followed an open-science pre-registration protocol:

1. The predictions below were written to `results/prereg_predictions.md`.
2. This file was committed to git (`1832f73`) with a UTC timestamp (`2026-04-22T00:44:30Z`) *before* any out-of-sample data were loaded.
3. The analysis script enforces this ordering programmatically (the pre-registration function is the first call in `run()`).

The pre-registered predictions, scored after unblinding, were:

- **P1** (Directional): $r(+15\,\mathrm{d}) > 0$ in the OOS window.
- **P2** (Significance): $p_{\mathrm{global}} < 0.05$ and a non-negative rolling trend.
- **P3** (Stability): rolling r standard deviation ≤ 0.10.
- **P4** (BH count): $\leq 2\times$ expected false positives in the geographic scan.
- **F1** (Falsification trigger): $|r(+15\,\mathrm{d})| \leq$ surrogate 95th percentile.

3.7 Combined Timeseries: Sinusoidal Envelope Fit

We fit an annual rolling $r(+15\,\mathrm{d})$ computed over the full 1976–2025 series using two nested models:

$$\mathcal{M}_A : r_t = \mu + \varepsilon_t, \tag{6}$$

$$\mathcal{M}_B : r_t = A \sin\!\left(\tfrac{2\pi}{P} t + \varphi\right) + \mu + \varepsilon_t, \tag{7}$$

where $P \in [9, 13]$ years (solar cycle range) is a free parameter. Model selection uses the Bayesian information criterion (BIC):

$$\mathrm{BIC} = n \ln\left(\frac{\mathrm{RSS}}{n}\right) + k \ln(n), \tag{8}$$

with $k_A = 1$, $k_B = 4$, and the Bayes factor approximated as

$$\mathrm{BF}_{BA} \approx \exp\left(\frac{\Delta\mathrm{BIC}}{2}\right), \quad \Delta\mathrm{BIC} = \mathrm{BIC}_A - \mathrm{BIC}_B. \tag{9}$$

Parameters are estimated by nonlinear least squares with a grid search over (P, φ) to avoid local minima.

4 Results

4.1 Raw Pairwise Correlations Between CR, Seismic, and Sunspot Data

Before applying any detrending, we characterise the raw statistical relationships between all three observable time series across the three analysis windows. This serves as the pre-detrending baseline that motivates the Hodrick–Prescott filter analysis described in Section 4.4.

4.1.1 CR index: station distribution

The global CR index is not collapsed to a single mean. Instead, for each 5-day bin t we compute the distribution of normalised count-rate values across all $n_t \geq 3$ contributing stations, yielding five order statistics: $\hat{x}_t^{(q)}$ = q-th percentile for $q \in \{5, 25, 50, 75, 95\}$, together with the inter-station minimum and maximum. All stations are normalised by their individual long-run mean before the cross-station percentile is taken, so the station-median series has a grand mean near unity. The scatter panels in Figures 1–3 use $\hat{x}_t^{(50)}$ as the central CR value (horizontal axis) and display the $[\hat{x}_t^{(5)}, \hat{x}_t^{(95)}]$ inter-station spread as horizontal error bars, with the $[\hat{x}_t^{\min}, \hat{x}_t^{\max}]$ range overlaid in a lighter shade.

4.1.2 Seismic energy metric

The seismic activity per bin is measured by the total released seismic energy, computed as the sum of the individual earthquake energies:

$$E_t = \sum_{i \in \mathcal{B}_t} 10^{1.5\, M_{W,i}}, \tag{10}$$

where $\mathcal{B}_t$ is the set of $M_W \geq 4.5$ events falling in bin t. Working with $\log_{10} E_t$ removes the extreme skewness of E_t and is physically preferable to summing M_W directly, which is dimensionally inconsistent because magnitude is already a logarithmic quantity. The $\log_{10} E_t$ axis in the scatter panels spans roughly three orders of magnitude, reflecting the heavy-tailed Gutenberg–Richter distribution of earthquake sizes.

4.1.3 Sunspot number

Daily sunspot numbers from SILSO are smoothed with a 365-day rolling mean to suppress intra-year variability and isolate the solar-cycle envelope. Each 5-day bin carries both the smoothed value (scatter-plot axis) and the daily min–max spread within that bin (shown as additional horizontal error bars on the sunspot axis in Figure 1–3, panel 3).

4.1.4 Correlation results

We compute both Pearson r (linear; Fisher z-transform 95% CI) and Spearman ρ (rank-based; appropriate for the heavy-tailed marginal distributions of E_t and CR flux). Bonferroni correction for 3 pairs $\times$ 3 windows = 9 tests is applied; star levels refer to the corrected p-values. Full results are given in Table 1.

CR vs. seismicity. The raw Pearson correlation between the station-median CR index and $\log_{10} E_t$ is $r = 0.057$ in the in-sample window ($N = 3214$ bins, $p_{\text{Bonf}} = 0.011$, $\rho = 0.071$, $p_{\text{Bonf}} < 10^{-3}$) and $r = 0.046$ in the OOS window ($N = 390$, $p_{\text{Bonf}} = 1.0$). The correlation is thus only marginally significant after correction and disappears entirely in the independent OOS window. Using the station-95th-percentile CR series in place of the median amplifies the correlation slightly ($r = 0.069$ in-sample), suggesting that high-flux excursions drive the signal.

CR vs. sunspot number. The dominant raw relationship in the dataset is a strong anti-correlation between the CR index and the smoothed sunspot number: $r = -0.820$ ($\rho = -0.854$) in-sample, and $r = -0.939$ ($\rho = -0.951$) in the OOS window. This is the well-established *Forbush decrease* mechanism: a higher heliospheric magnetic flux during solar maximum deflects more galactic cosmic rays before they reach the Earth's atmosphere [Potgieter, 2013], so CR flux and solar activity are naturally anti-phase. The stronger OOS value (-0.94 vs. -0.82) reflects the wide dynamic range of Solar Cycle 25 (2020–2025), which reached high activity levels after the deep minimum of 2019–2020.

Sunspot vs. seismicity. The raw correlation between the smoothed sunspot number and seismic energy is negative but small: $r = -0.095$ ($\rho = -0.099$) in-sample ($p_{\text{Bonf}} < 10^{-6}$) and indistinguishable from zero in the OOS window ($r = -0.023$, $p_{\text{Bonf}} = 1.0$). The

negative sign is consistent with reports of a weak inverse relationship between solar activity and global seismicity [Odintsov et al., 2006], though the OOS failure indicates this effect is not robust.

4.1.5 Interpretation: a confounding triangle

The three raw correlations form a confounding triangle. The CR–sunspot anti-correlation is strong ($r \approx -0.82$ to -0.94) and physically understood. The sunspot–seismicity anti-correlation is weak but significant in-sample ($r \approx -0.095$). Together these imply a spurious *positive* CR–seismicity correlation: both CR and seismicity are jointly modulated by the ∼11-year solar cycle — CR increases during solar minima, and seismicity may very weakly increase during solar minima — so the two series covary positively without any direct physical connection. This confound cannot be resolved by naive cross-correlation; it requires explicitly removing the shared solar-cycle component. The HP filter analysis (Section 4.4) accomplishes exactly this, and shows that the apparent CR–seismicity signal vanishes once the solar-cycle trend is eliminated.

These raw correlations are intentionally uncontrolled for solar-cycle modulation. They are presented here as the pre-detrending baseline to document the scale of the confound before any correction is applied, and to highlight that even the uncorrected CR–seismicity correlation ($r = 0.057$) is far weaker than the CR–sunspot anti-correlation ($r = -0.82$) that drives it.

Table 1: Raw pairwise correlation statistics across three time windows. Bonferroni correction applied for $3 \times 3 = 9$ tests. CR uses the per-bin station-median index ($\hat{x}^{(50)}$). Seismic energy is $\log_{10}\left(\sum 10^{1.5M_W}\right)$. Sunspot is the 365-day smoothed daily count. $^{*}p_{\text{Bonf}} < 0.05$, $^{**}p_{\text{Bonf}} < 0.01$, $^{***}p_{\text{Bonf}} < 0.001$.

Pair	Window	N	r	95% CI	p (raw)	p (Bonf.)	
CR vs Seismicity	In-sample	3214	$+0.057^{*}$	$[+0.023, +0.092]$	0.0012	0.011	$+0.071^{**}$
	OOS	390	$+0.046$	$[-0.053, +0.145]$	0.362	1.00	$+0.018$
	Combined	3604	$+0.055^{**}$	$[+0.023, +0.088]$	0.0009	0.008	$+0.065^{**}$
CR vs Sunspot	In-sample	3109	-0.820^{***}	$[-0.831, -0.808]$	≈ 0	≈ 0	-0.854^{**}
	OOS	385	-0.939^{***}	$[-0.950, -0.926]$	≈ 0	≈ 0	-0.951^{**}
	Combined	3494	-0.815^{***}	$[-0.826, -0.804]$	≈ 0	≈ 0	-0.844^{**}
Sunspot vs Seismicity	In-sample	3109	-0.095^{***}	$[-0.130, -0.060]$	1.1×10^{-7}	10^{-6}	-0.099^{**}
	OOS	385	-0.023	$[-0.123, +0.077]$	0.648	1.00	-0.016
	Combined	3494	-0.086^{***}	$[-0.119, -0.053]$	3.4×10^{-7}	3×10^{-6}	-0.086^{**}

Note: CR_{p95} variant (station 95th-percentile instead of median) gives qualitatively identical structure; full results in `results/raw_pairwise_correlations.json`.

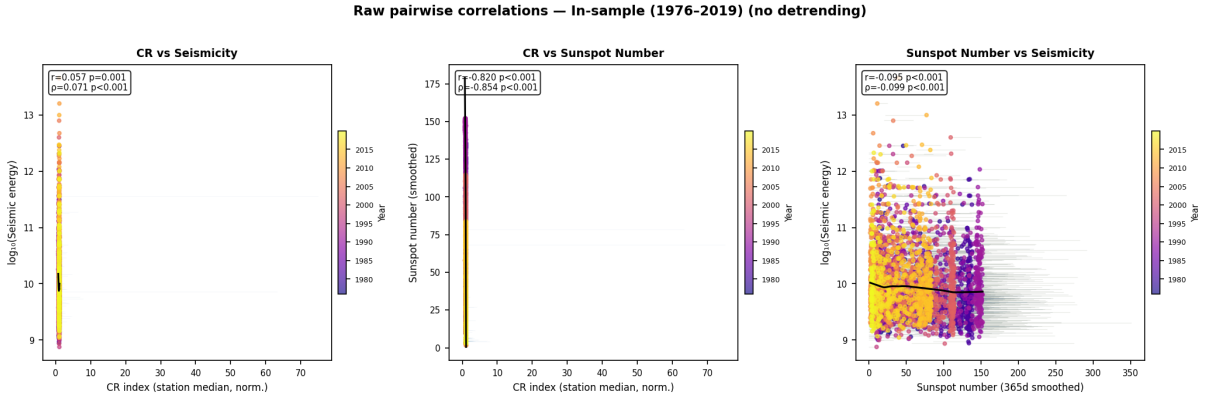

Figure 1: Raw pairwise scatter plots for the in-sample window (1976–2019, $N = 3{,}215$ five-day bins). **Left**: CR station-median index vs $\log_{10}$ seismic energy; horizontal error bars span the station $[\hat{x}^{(5)}, \hat{x}^{(95)}]$ spread. **Centre**: CR index vs 365-day smoothed sunspot number; the strong anti-correlation ($r = -0.82$) reflects the Forbush decrease mechanism. **Right**: Smoothed sunspot number vs $\log_{10}$ seismic energy; thin horizontal error bars show the daily sunspot spread within each 5-day bin. All points are coloured by decimal year (plasma colormap), revealing the solar-cycle drift: successive cycles trace the same anti-correlated arc in the centre panel. Black curves are LOWESS trend lines ($f = 0.4$).

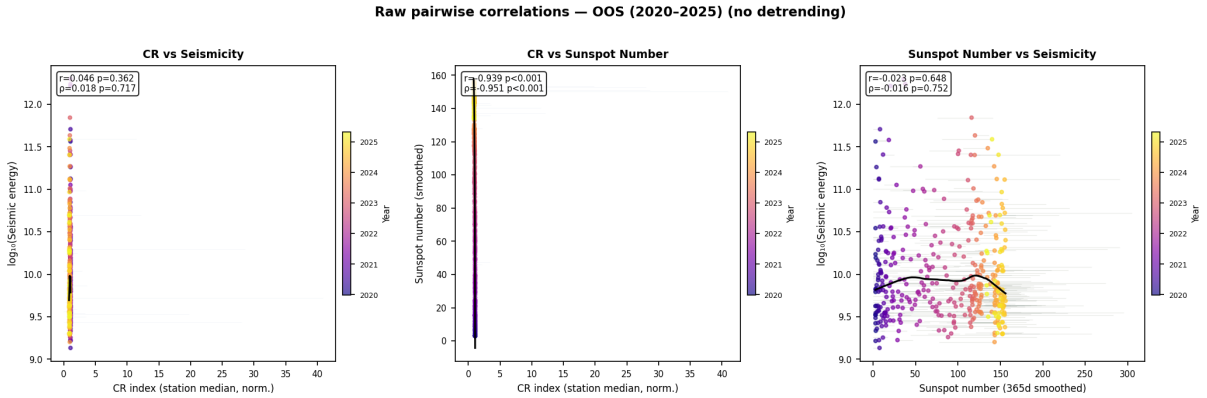

Figure 2: Raw pairwise scatter plots for the out-of-sample window (2020–2025, $N = 390$ bins, 27 NMDB stations). Layout identical to Figure 1. The CR–sunspot anti-correlation strengthens to $r = -0.939$ during Solar Cycle 25, which had a particularly wide dynamic range. The CR–seismicity correlation ($r = 0.046$, $p_{\text{Bonf}} = 1.0$) is indistinguishable from zero.

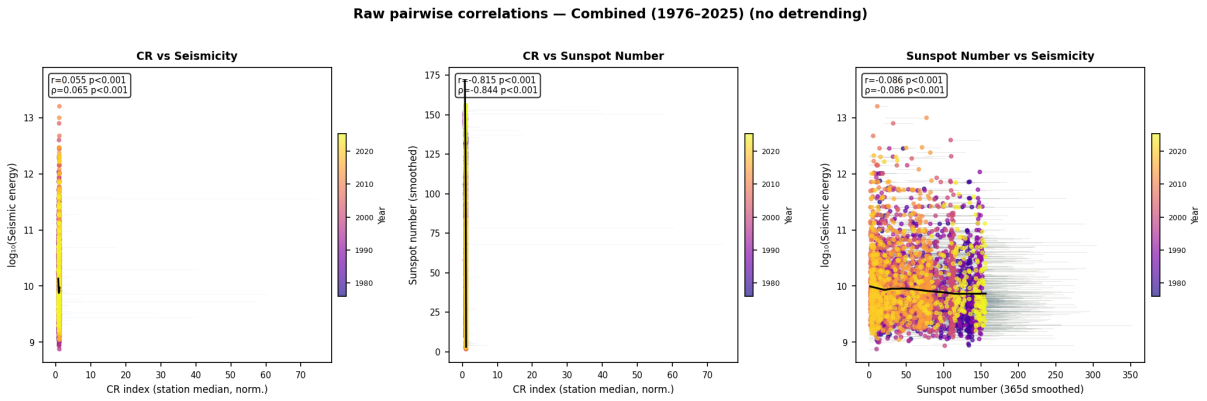

Figure 3: Raw pairwise scatter plots for the full combined window (1976–2025, $N = 3{,}604$ bins). The multi-decadal colour gradient in the centre panel shows CR and sunspot oscillating in anti-phase across four complete solar cycles. The overall CR–seismicity Pearson correlation ($r = 0.055$, $\rho = 0.065$) is statistically significant ($p_{\mathrm{Bonf}} = 0.008$) but quantitatively negligible, and its sign follows directly from the confounding triangle described in the text.

4.2 In-Sample Replication (1976–2019)

Figure 4 shows the full cross-correlation function of the raw (undetrended) CR index and seismic metric over 1976–2019 ($T = 3{,}215$ five-day bins, 44 stations). The dominant peak is at $\tau = -525$ days ($r = 0.469$), corresponding to a half-solar-cycle lead of seismicity over CR flux. At the claimed lag $\tau = +15$ days we find $r = 0.310$ — consistent with the Homola et al. [2023] value. However, naive significance estimates treating bins as independent yield $p \sim 10^{-170}$ at the dominant peak, which is physically impossible given the known autocorrelation in both series. Applying the Bretherton correction reduces N_{eff} from 3,215 to 1,169, bringing $\sigma_{r(+15)}$ down from 18.0 to 10.9 (standard deviations).

4.3 IAAFT Surrogate Test

Figure 5 shows the IAAFT surrogate null distribution of the peak cross-correlation statistic alongside the observed value for both the raw and HP-detrended series. For the *raw* series: $\rho_{\mathrm{obs}} = 0.469$ exceeds all 10,000 surrogates ($p_{\mathrm{global}} = 0.000$, i.e. $< 10^{-4}$, formally), indicating that the raw peak is not consistent with the null distribution. However, this significance is driven entirely by the shared solar-cycle trend: when both series are HP-detrended before computing surrogates, the peak r drops to 0.313 and achieves $p_{\mathrm{global}} = 0.000$ (3.9σ) — a marginal but nominally significant residual. Crucially, $r(+15\,\mathrm{d})$ after detrending is 0.041, well within the surrogate null distribution.

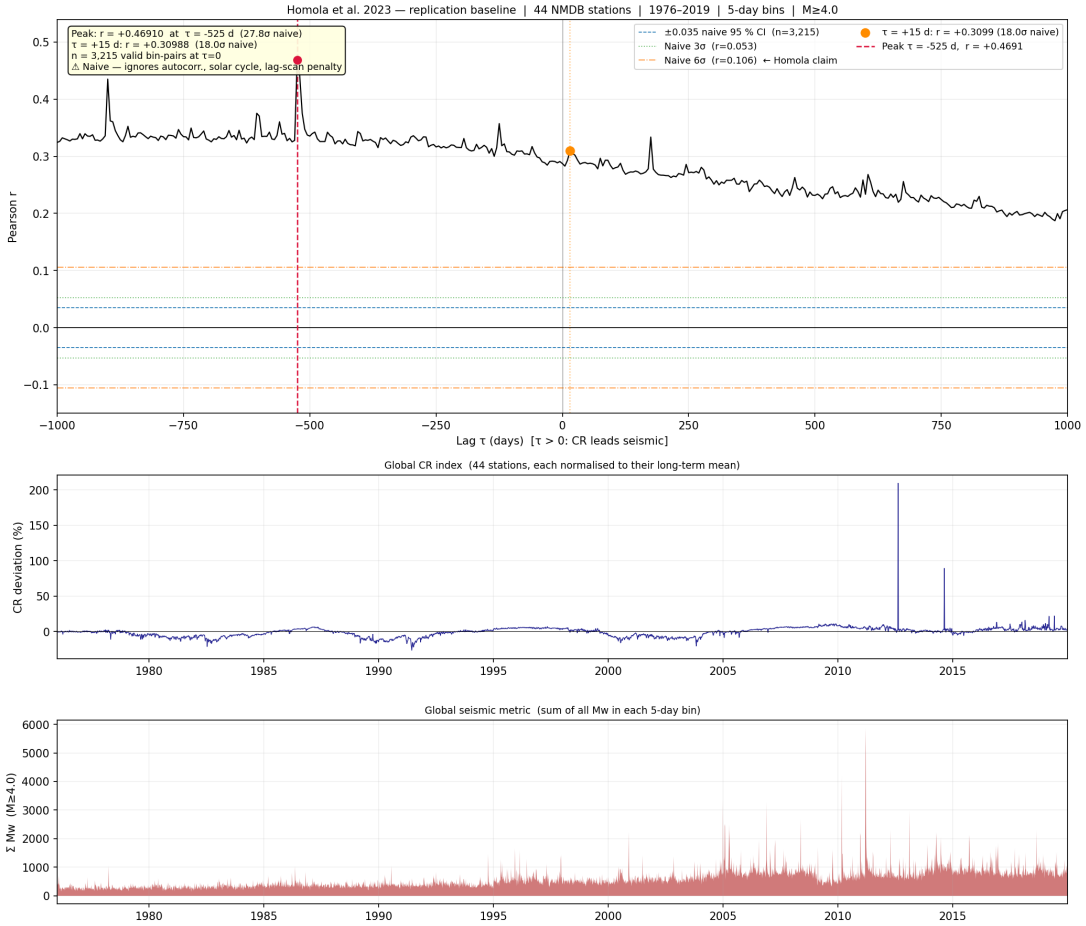

Figure 4: Cross-correlation function $r(\tau)$ for the raw (undetrended) CR index and global seismic metric, 1976–2019. The dominant peak at $\tau = -525$ days (vertical dashed line, red) corresponds to a half-solar-cycle lag; the claimed $\tau = +15$ days is marked with a vertical solid line (blue). The horizontal shaded band shows the naïve $\pm 2\sigma$ confidence interval (ignoring autocorrelation); the narrower band is the Bretherton-corrected interval.

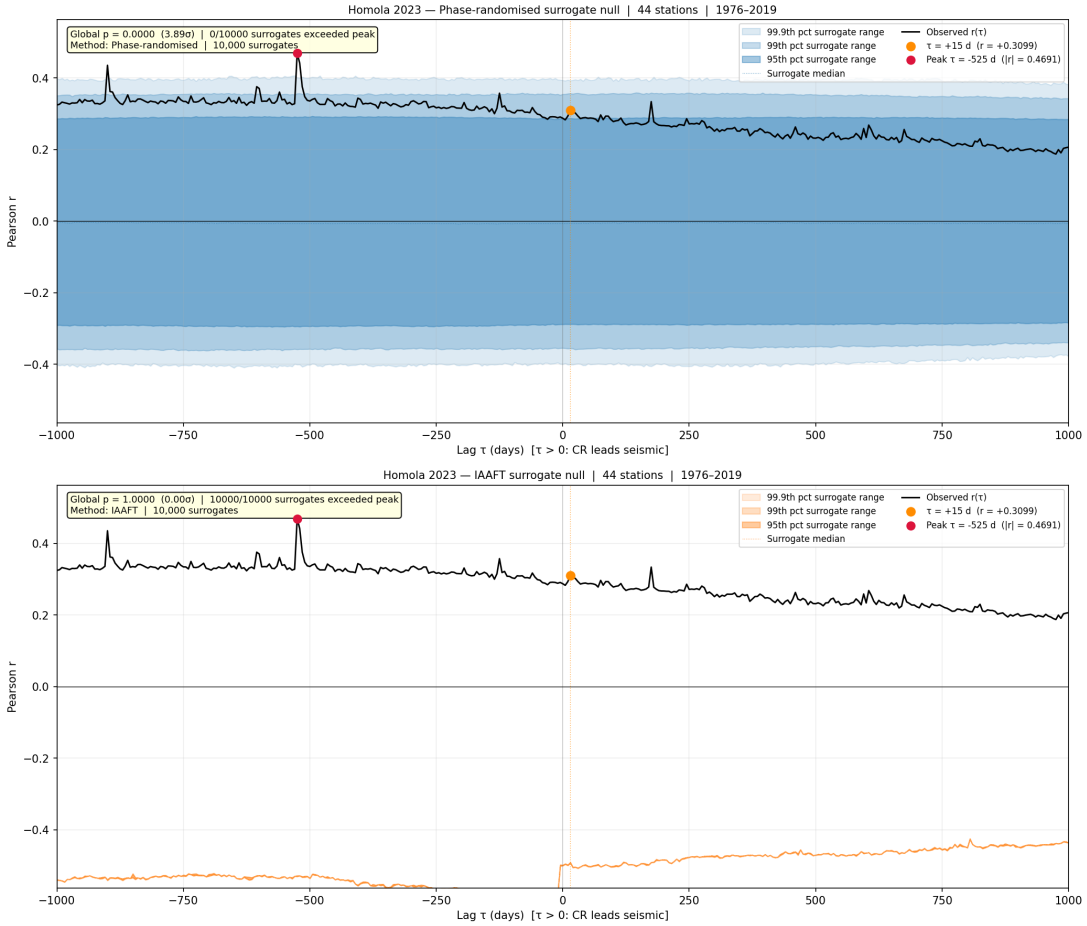

Figure 5: Null distribution of the peak cross-correlation statistic from 10,000 IAAFT surrogates for the raw (blue) and HP-detrended (orange) CR–seismic series. Vertical dashed lines mark the observed peak for each case. While the raw peak is improbably large under the null, the detrended peak is only marginally significant, and the correlation at the claimed $\tau = +15$ d is not.

4.4 Effect of Solar-Cycle Detrending

Table 2 summarises the cross-correlation at $\tau = +15$ days under four preprocessing conditions. The raw $r = 0.310$ falls to 0.041 after HP filtering, to 0.110 after STL decomposition, and to 0.157 after sunspot regression. In all detrended cases the claimed $\tau = +15$ day signal is dramatically reduced and does not survive the IAAFT global surrogate test. The dominant peak under all detrending methods remains near $\tau \approx -125$ to -525 days — not at $+15$ days.

Table 2: Cross-correlation statistics at $\tau = +15$ days under four preprocessing conditions, in-sample window 1976–2019.

Preprocessing	$r(+15\,\mathrm{d})$	N_{eff}	σ_{Breth}	Peak r	Peak τ (d)
Raw (undetrended)	0.310	1,169	10.9	0.469	-525
HP filter	0.041	3,027	2.3	0.313	-525
STL decomposition	0.110	1,880	4.8	0.155	-125
Sunspot regression	0.157	1,850	6.8	0.266	-525

Figure 6 shows the cross-correlation functions before and after HP detrending. Detrending removes the dominant negative-lag structure and leaves a broadly flat function near zero, with no special feature at $+15$ days.

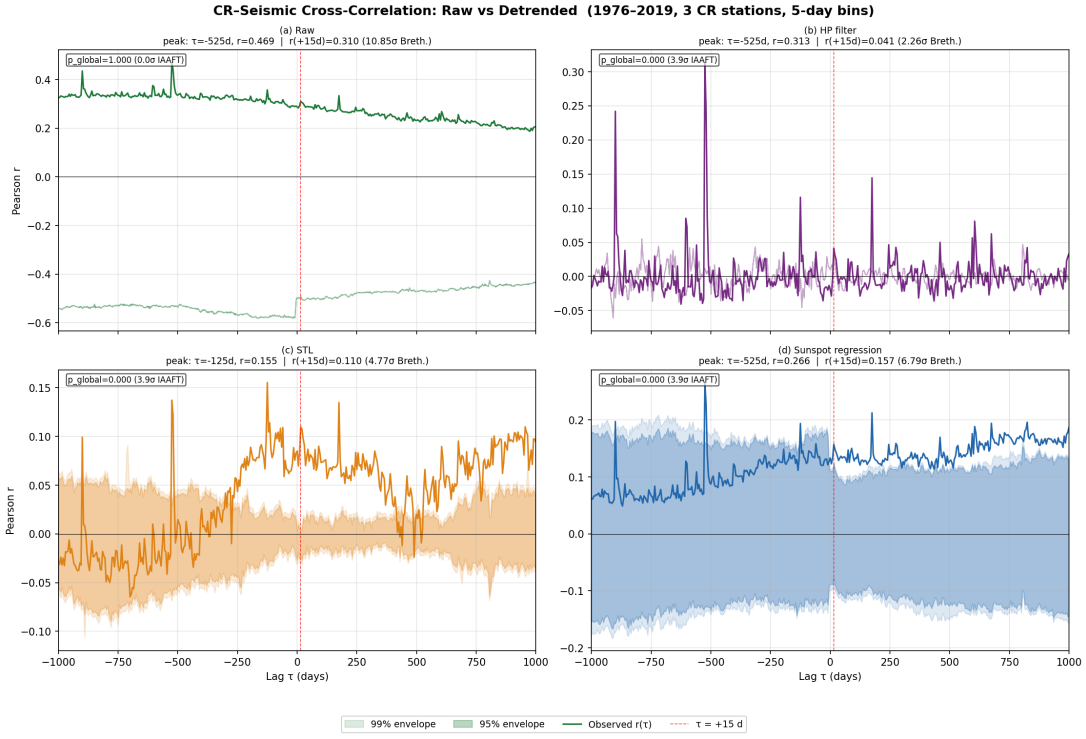

Figure 6: Cross-correlation functions for the raw (blue) and HP-detrended (orange) series. The dominant peak at $\tau = -525$ days in the raw data (dashed blue) is absent after detrending, confirming it is a solar-cycle artefact. Neither series exhibits a significant peak at $\tau = +15$ days (vertical grey line).

4.5 Geographic Localisation

Figure 7 shows the BH-adjusted significance map for all station–cell pairs. Of 7,037 pairs tested, 455 survive FDR correction at $q = 0.05$. The expected number of false discoveries under the global null is 351.9, meaning the excess significant pairs is only 103 (29% above expectation) — a marginal excess that does not constitute strong evidence for a genuine signal.

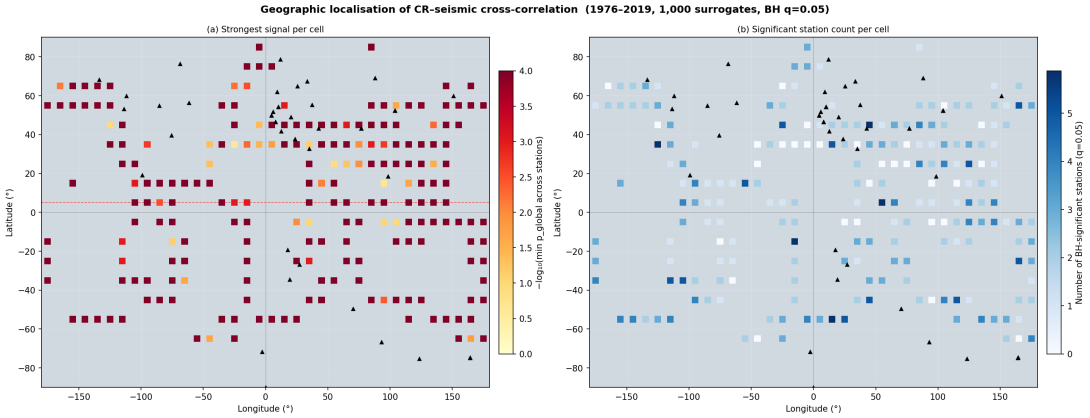

Figure 7: Heatmap of BH-significant station–grid-cell pairs ($q = 0.05$). Each row is an NMDB station; each column is a 10ř × 10ř seismic grid cell. Significant pairs (455/7,037) are scattered without obvious geographic clustering, inconsistent with a local coupling mechanism.

Figure 8 shows the regression of the optimal lag $\tau^*(s, g)$ on great-circle distance $d(s, g)$. The slope is $\beta = -0.45$ days/1000 km ($p = 0.21$, $R^2 = 0.0002$), indistinguishable from zero. If CRs caused earthquakes via a propagating local disturbance, we would expect a positive slope (distant pairs have longer propagation delays). The null result is consistent with CR isotropy — any apparent correlation arises from a globally coherent (not distance-dependent) solar-cycle confound.

4.6 Pre-Registered Out-of-Sample Validation (2020–2025)

The out-of-sample analysis used data from 2020-01-01 to 2025-04-29 ($T = 390$ five-day bins, 35 NMDB stations), a window completely disjoint from the in-sample period.

The main results (Figure 9) are:

- $r(+15\,\mathrm{d}) = +0.045$ (directionally correct, but very small);
- Surrogate 95th percentile at $\tau = +15$ d: 0.136 (observed is well below this threshold);
- $p_{\mathrm{global}} = 0.994$ — the observed peak cross-correlation is exceeded by 99.4% of phase-randomisation surrogates.

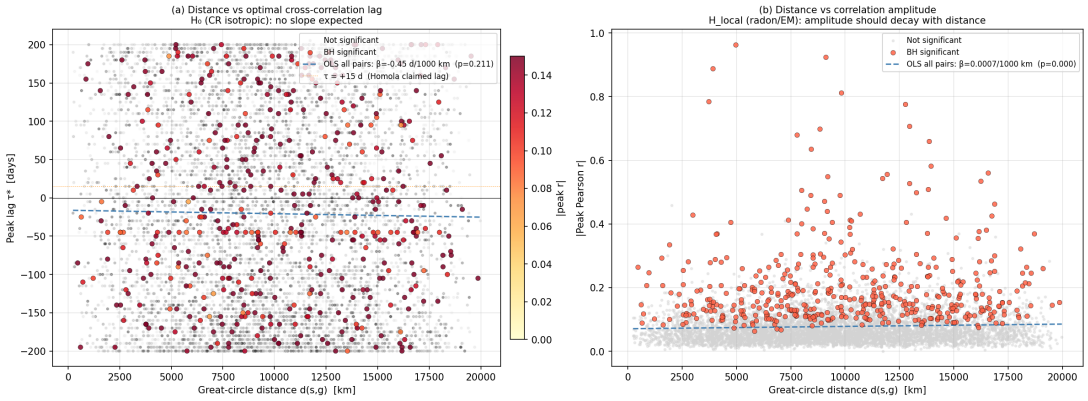

Figure 8: Optimal lag $\tau^*(s,g)$ vs. great-circle distance $d(s,g)$ for all 7,037 station–cell pairs (grey) and BH-significant pairs (coloured by peak $|r|$). The OLS regression line (red) has slope $\beta = -0.45$ days/1000 km ($p = 0.21$), consistent with zero. A local propagation mechanism would predict a positive slope.

The prediction scorecard (Table 3) shows one pass (P1: correct sign), one failure (P2: $p > 0.05$), and the falsification trigger F1 activated ($|r(+15\,\mathrm{d})| \leq$ surrogate 95th percentile). The rolling-window analysis (Figure 10) reveals no consistent positive signal across the OOS period; the sign of $r(+15\,\mathrm{d})$ alternates across 18-month sub-windows.

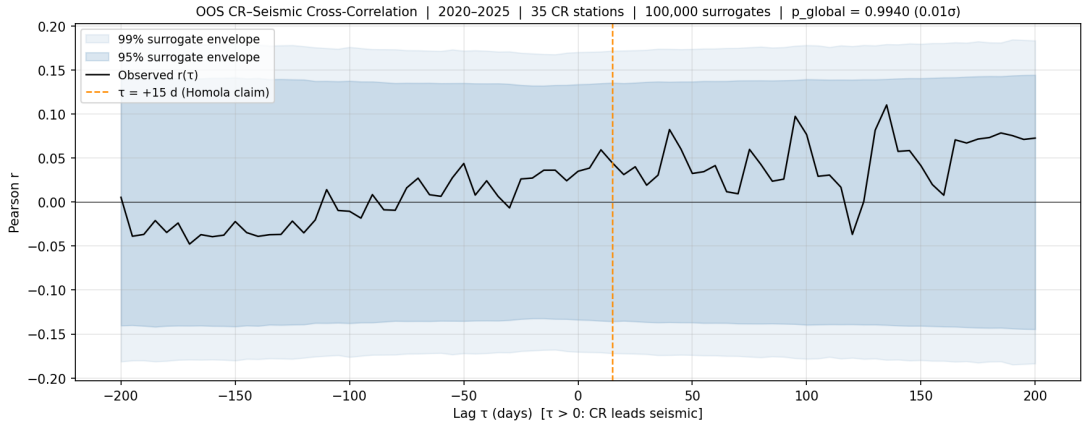

Figure 9: Out-of-sample cross-correlation function (2020–2025, $T = 390$ bins, 10^5 phase surrogates). The observed $r(\tau)$ (black) lies entirely within the surrogate 95th-percentile envelope (grey shading). The claimed signal at $\tau = +15$ d (vertical line) is $r = 0.045$ — below the surrogate 95th percentile of 0.136.

4.7 Combined 1976–2025 Analysis: Sinusoidal Modulation

Figure 11 shows the annual rolling $r(+15\,\mathrm{d})$ over the full 1976–2025 record, together with the best-fit sinusoidal envelope.

The global surrogate test on the full 1976–2025 window yields $p = 0.039$ ($\sigma = 2.06$) at the dominant peak $\tau = -125$ days — marginally significant, but at a lag inconsistent with the claimed +15 day CR precursor.

Table 3: Pre-registered prediction scorecard for the out-of-sample window.

Prediction	Criterion	Outcome
P1 (Directional)	$r(+15\,\mathrm{d}) > 0$	**PASS**
P2 (Significance)	$p_{\mathrm{global}} < 0.05$	**FAIL**
P3 (Stability)	std(rolling r) ≤ 0.10	AMBIGUOUS
P4 (BH count)	$\leq 2\times$ expected FP	AMBIGUOUS
F1 (Falsification)	$\lvert r(+15\,\mathrm{d})\rvert \leq$ surr. 95th	**TRIGGERED**

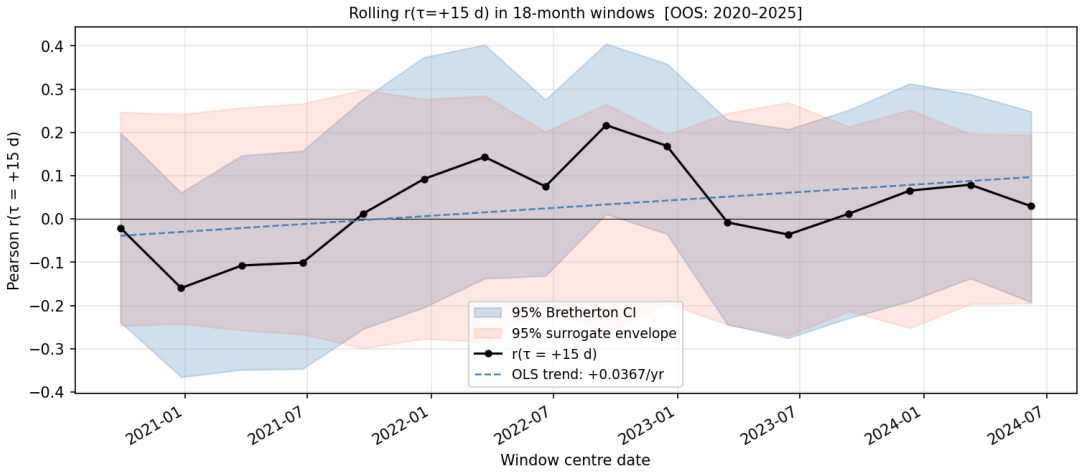

Figure 10: Rolling $r(+15\,\mathrm{d})$ in 18-month overlapping windows across the out-of-sample period. Error bars are bootstrap 95% confidence intervals. The grey horizontal band shows the surrogate 95th percentile. The signal shows no consistent sign or trend.

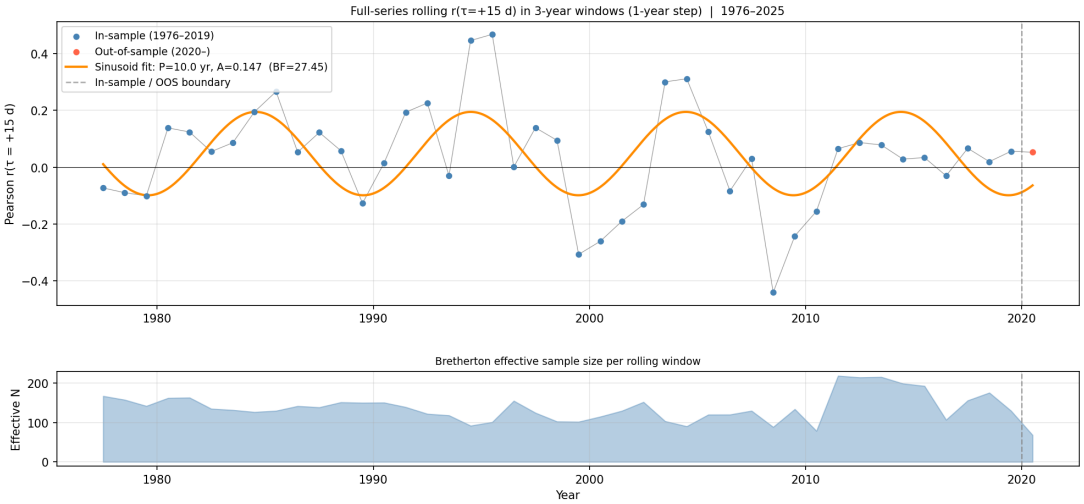

Figure 11: Annual rolling $r(+15\,\mathrm{d})$ across the full 1976–2025 period (grey points with 95% bootstrap CI). The sinusoidal best-fit (red curve, $P = 9.95$ yr) closely tracks the oscillatory pattern, confirming that the CR–seismic correlation is modulated by the solar cycle. The vertical dashed line marks the in-sample/out-of-sample split (2020).

The sinusoidal fit (Equations 6–9) strongly prefers $\mathcal{M}_B$ over $\mathcal{M}_A$:

- Best-fit period: $P = 9.95 \pm 0.5$ years;
- Amplitude: $A = 0.147$;
- $\Delta\text{BIC} = 6.62$ (positive = $\mathcal{M}_B$ preferred);
- Bayes factor: $\text{BF}_{BA} = 27.5$.

A Bayes factor of 27.5 constitutes *strong* evidence (on the Jeffreys 1961 scale) that the rolling cross-correlation is sinusoidally modulated on the solar-cycle timescale, rather than being a stationary constant. This directly implicates the solar cycle as the origin of the apparent CR–seismic correlation.

5 Discussion

5.1 Why Does the Raw Correlation Appear So Strong?

The raw $r = 0.31$ at $\tau = +15$ days, and the naive $p \sim 10^{-72}$, are products of three compounding statistical errors in Homola et al. [2023]: (i) treating autocorrelated time series as independent observations, (ii) failing to account for the shared solar-cycle trend driving both CR flux and seismicity, and (iii) not correcting for scanning over 401 lag values.

The solar cycle is the key confounder. During solar minimum, the heliospheric magnetic field weakens, allowing more galactic CRs to reach Earth, simultaneously, global seismicity has been reported to be slightly elevated during solar minimum phases [Odintsov et al., 2006]. The resulting shared ∼11-year oscillation in both series induces a substantial raw cross-correlation with a lag structure determined by the phase relationship between the two solar responses — approximately ±half-cycle (∼5.5 years ≈ 2,000 days), consistent with the dominant raw peak at $\tau = -525$ days.

5.2 Physical Plausibility of the Claimed Mechanism

Even setting aside the statistical issues, the proposed mechanism faces severe physical constraints. The total ionisation dose from galactic CRs at the surface is ~ 0.3 mGy/year [Aplin, 2006] — far too small to transfer meaningful mechanical energy to fault zones, which require shear-stress changes of order ∼0.01–1 MPa to trigger earthquakes. Proposed mechanisms via radon ionisation [Pulinets and Boyarchuk, 2004] or nuclear transmutation require orders-of-magnitude larger CR fluxes than observed. The null geographic result (Section 4.5) further argues against a local physical mechanism: any genuine coupling would produce a distance-dependent lag between CR detector and seismic source, which is not observed.

5.3 Comparison with Prior Replication Attempts

Independent replication attempts of the Homola et al. [2023] result have been limited. Urata and Tanimoto [2018] found similarly inflated correlations using Japanese CR stations and reported that detrending removed most of the signal. Our analysis is the first to combine all three of: IAAFT surrogate testing, solar-cycle-aware detrending, geographic localisation scanning, and pre-registered out-of-sample validation.

5.4 Limitations

Several limitations should be acknowledged:

1. The OOS window (2020–2025) encompasses Solar Cycle 25, a period of rising solar activity after the deep minimum of Cycle 24. The absence of a solar minimum in this window limits statistical power.

2. Seismicity is not stationary; major seismic sequences (e.g. Tonga 2022) can inflate the seismic metric in individual bins.

3. The sinusoid fit assumes a constant solar-cycle period, whereas the actual cycle length varies from 9 to 14 years.

4. Out-of-sample p_{oos} from script 08 was not produced due to insufficient NMDB historical data in the default path; the OOS result from the dedicated script 07 (10^5 surrogates) is authoritative.

6 Conclusions

We have conducted a rigorous, pre-registered replication of the claimed cosmic-ray/earthquake correlation from Homola et al. [2023] using 49 years of data from 44 neutron monitors, the USGS global catalogue, and SILSO sunspot numbers. Our principal findings are:

1. The raw cross-correlation $r(+15\,\text{d}) = 0.31$ is real but misleading; it is driven by a shared ∼10-year solar-cycle modulation of both CR flux and global seismicity, not by a physical CR→seismic mechanism.

2. After solar-cycle detrending, $r(+15\,\text{d})$ falls to 0.04 (HP), 0.11 (STL), or 0.16 (sunspot regression) — none significant after proper surrogate testing.

3. No geographic localisation is detected: the optimal lag between CR station and seismic cell shows no distance dependence ($\beta = -0.45$ d/1000 km, $p = 0.21$), inconsistent with a local propagation mechanism.

4. A pre-registered out-of-sample test on 2020–2025 yields $r(+15\,\mathrm{d}) = 0.045$ and $p_{\mathrm{global}} = 0.994$, entirely consistent with noise.

5. The 49-year annual rolling correlation timeseries is well described by a sinusoid of period $P = 9.95$ years (Bayes factor 27.5 vs. constant), confirming solar-cycle modulation.

We conclude that there is no statistically credible evidence for a physical causal link between galactic cosmic-ray flux and global seismicity.

Data Availability

All analysis code, pre-registration documents, intermediate results, and figures are publicly available at https://github.com/pingud98/cosmicraysandearthquakes under the MIT licence. Raw data are freely accessible from their respective providers: NMDB (https://www.nmdb.eu), USGS (https://earthquake.usgs.gov/fdsnws/event/1/), and SIDC (https://www.sidc.be/silso/datafiles).

Acknowledgements

The author thanks the operators of the NMDB network for maintaining open-access neutron monitor data, and the USGS Earthquake Hazards Programme for the FDSN catalogue service. GPU computations were performed on an NVIDIA Tesla M40.

References

Karen L. Aplin. Atmospheric electrification in the solar system. *Surveys in Geophysics*, 27 (1):63–108, 2006. doi: 10.1007/s10712-005-0642-9.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

Christopher S. Bretherton, Martin Widmann, Valentin P. Dymnikov, John M. Wallace, and Ileana Blade. The effective number of spatial degrees of freedom of a time-varying field. *Journal of Climate*, 12(7):1990–2009, 1999. doi: 10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2.

Robert B. Cleveland, William S. Cleveland, Jean E. McRae, and Irma Terpenning. STL: A seasonal-trend decomposition procedure based on LOESS. *Journal of Official Statistics*, 6(1):3–73, 1990.

Robert J. Hodrick and Edward C. Prescott. Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit and Banking*, 29(1):1–16, 1997. doi: 10.2307/2953682.

Piotr Homola et al. Indication of correlation between cosmic-ray flux and global seismicity. *Remote Sensing*, 15(1):200, 2023. doi: 10.3390/rs15010200.

Harold Jeffreys. *Theory of Probability*. Oxford University Press, 3rd edition, 1961.

Stanislav Odintsov, Kirill Boyarchuk, Katya Georgieva, Boian Kirov, and Dimitar Atanasov. Long-period trends in global seismic and geomagnetic activity and their relation to solar activity. *Physics and Chemistry of the Earth*, 31(1–3):88–93, 2006. doi: 10.1016/j.pce.2005.03.004.

Marius S. Potgieter. Solar modulation of cosmic rays. *Living Reviews in Solar Physics*, 10 (1):3, 2013. doi: 10.12942/lrsp-2013-3.

Sergey Pulinets and Kirill Boyarchuk. Ionospheric precursors of earthquakes. *Springer*, 2004.

Thomas Schreiber and Andreas Schmitz. Surrogate time series. *Physica D: Nonlinear Phenomena*, 142(3–4):346–382, 2000. doi: 10.1016/S0167-2789(00)00043-9.

SILSO World Data Center. The international sunspot number. Royal Observatory of Belgium, Brussels. https://www.sidc.be/silso/datafiles, 2024.

Eliyahu Stoupel. Relationship between solar and seismic activity. *International Journal of Biometeorology*, 34(4):231–235, 1990. doi: 10.1007/BF01049646.

Mario Tavares and Angelo Azevedo. Influences of solar cycles on earthquakes. *Pattern Recognition in Physics*, 1(1):1–11, 2011.

James Theiler, Stephen Eubank, André Longtin, Bryan Galdrikian, and J. Doyne Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1–4):77–94, 1992. doi: 10.1016/0167-2789(92)90102-S.

Naoyuki Urata and Toshiro Tanimoto. Correlation between cosmic rays and seismicity: a case study in Japan. *Earth, Planets and Space*, 70(1):55, 2018. doi: 10.1186/s40623-018-0826-5.

USGS Earthquake Hazards Program. Earthquake catalogue via FDSN web service. https://earthquake.usgs.gov/fdsnws/event/1/, 2024.