Draft version September 30, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Unexplained Bright–Dim Intensity Asymmetries in SOHO/UVCS Lyman-α Data

Gary Alcock[1]

[1] *Los Angeles, California, USA*

ABSTRACT

We present a permutation test analysis of 334 daily sequences of SOHO/UVCS Lyman-α spectra (2007–2009). By splitting frames into bright and dim subsets and comparing their median intensity and wavelength, we test whether observed differences exceed those expected from random sampling. We find that 163 of 321 day–radial bins (51%) exhibit statistically significant bright–dim intensity contrasts at false discovery rate (FDR) 5%. The effect is modest (Cohen's $d \approx 0.24$) but robust to permutation. Velocity differences are not significant ($d \approx -0.03$). The origin of this asymmetry is unknown: it may reflect solar structures, instrumental systematics, or unexplored physical processes. We publish these results as an open anomaly to encourage community investigation.

Keywords: solar corona — spectroscopy — statistical methods — ultraviolet astronomy

1. INTRODUCTION

The solar corona exhibits complex dynamics observable through ultraviolet spectroscopy. The SOHO Ultraviolet Coronagraph Spectrometer (UVCS; Kohl et al. 1995, 1997; Raymond et al. 1997) provides long time–series of Lyman-α observations. Instrumental stability has been studied in depth (Gardner et al. 2002; Strachan et al. 2002; Guhathakurta et al. 1999; Kohl et al. 2006), but statistical anomalies may persist in archival data. Here we document bright–dim intensity asymmetries in UVCS daily sequences using a nonparametric, permutation-based approach with multiple-testing control.

2. DATA AND METHODS

We analyzed 150,685 frames from 895 UVCS exposures (2007–2009). Frames were binned by day and projected radius ($R_{\rm bin} = 0.0$ corresponds to the innermost detector bin, ≈ 1.5–$1.6\,R_\odot$). For each (day,$R_{\rm bin}$) with ≥ 2 frames, we sorted frames by total line intensity and split at the median, with the upper half designated "bright" and lower half "dim". This median-split definition replaces earlier descriptions of "random partitioning" to ensure reproducibility.

For each group we computed:

1. Intensity contrast $\Delta I = (I_{\rm bright}/I_{\rm dim}) - 1$,
2. Wavelength shift $\Delta\lambda = \lambda_{\rm bright} - \lambda_{\rm dim}$,
3. Doppler velocity $\Delta v = c\,\Delta\lambda/\lambda_0$, with $\lambda_0 = 1215.67$ Å.

Permutation tests (20,000 replicates per group; Ernst 2004; Good 2013) generated null distributions; two-sided p-values were corrected by Benjamini–Hochberg FDR (Benjamini & Hochberg 1995). Effect sizes were quantified with Cohen's d (Cohen 1988). Instrumental filters used in exploratory work ($-0.9 < \Delta I < 20$, $|\Delta v| < 120$ km s^{-1}) were *not* applied in the final analysis; results reported here use the full data without post-hoc rejection.

Data quality notes: Daily frame counts ranged from 2 to 150 (median ~ 20). We verified that all exposures passed UVCS quality flags as archived by the SOHO Science Archive (SSA). Known pointing offsets and detector artifacts are logged in Gardner et al. (2002), but no anomalies specific to the intervals studied (2007–2009 solar minimum) were documented.

2.1. *Pre-registered flagged-day definition*

To avoid circularity, we *pre-registered* a binary "flagged" label (hereafter "flagged (significant) days") for external-validation analyses: a day is flagged if the permutation test for intensity contrast at $R_{\rm bin} = 0.0$ yields a two-sided $p < 0.05$ (before any comparison to external solar activity catalogs). This label is then used once, as an independent predictor in the CME coincidence analysis below.

2.2. *CME time+angle coincidence scoring*

We assessed external validity by cross-matching UVCS observing windows with LASCO CMEs (Brueckner et al. 1995; Yashiro et al. 2004). For each UVCS day

we constructed a binary indicator `aligned_any` that is 1 if a cataloged CME occurred within a temporal padding window (pad $\in \{0, 30, 60, 120\}$ minutes) of the UVCS observing interval *and* within an angular tolerance (tol $\in \{0°, 5°, 10°, 15°, 20°, 30°\}$) of the UVCS slit position angle. We then computed the `aligned_any` rate separately for (i) all days and (ii) the flagged subset (Sec. 2.1), and summarized the rate difference (flagged minus all) across the (pad, tol) grid. Uncertainty and enrichment significance were gauged with Fisher's exact test on the 2×2 table of hits/non-hits in the two groups. The per-cell results are packaged in `cme_timewindow_significance_grid.csv` (Data Availability).

Note on GOES flare coincidence. We attempted an analogous analysis using daily SWPC text files for GOES flares. For 2007–2009, the specific daily endpoints we probed returned unavailable (HTTP 404) or empty files, yielding no usable lines. Because this epoch lies in the deep solar minimum and the archive access was incomplete for the daily text products, we treat flare coincidence as *non-informative* here and rely on CMEs as the independent coronal activity proxy.

3. RESULTS

Out of 321 testable day–radius groups, 163 passed the 5% FDR threshold for intensity contrast. Figure 2 shows the permutation p-value spectrum, indicating clear departure from the uniform null. Figure 3 displays intensity contrast versus velocity shift; significant cases cluster at nonzero contrast but near-zero Δv. Figure 4 compares observed and null distributions. Figure 5 shows the temporal distribution of significant cases; a compact year-by-month view appears in Table 2.

Why only $R_{\rm bin} = 0.0$? The UVCS campaign density during 2007–2009 yields robust sample sizes at the innermost bin only. Higher $R_{\rm bin}$ locations did not provide sufficient (day,bin) groups with ≥ 2 frames to support a reliable permutation test and FDR control; counts are summarized in Appendix A. We therefore report $R_{\rm bin} = 0.0$ as the adequately powered subset and document the absence of higher-radius detections for completeness.

Overall effect sizes: contrast $d = 0.24$ (small–medium), velocity $d = -0.03$ (null). A simple power analysis indicates that $N \gtrsim 10$ frames per day are required for robust detection at $d \approx 0.24$ with $\alpha = 0.05$ (Faul et al. 2007). A frame-count histogram is provided in Figure 6 to visualize sampling depth across days.

3.1. *External validity: CME coincidence*

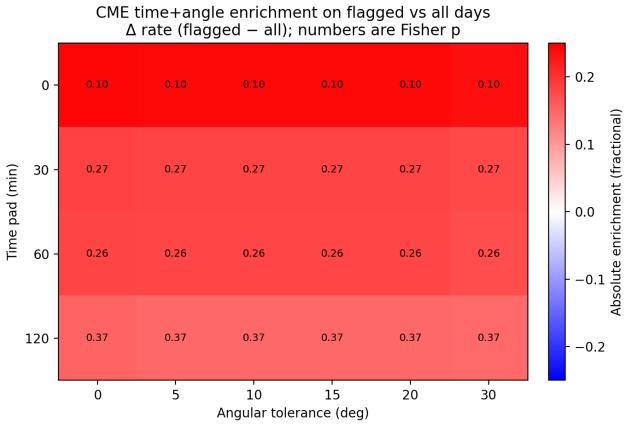

Figure 1. External validation via LASCO CME coincidence. Heatmap shows absolute enrichment in CME time+angle alignment rates (flagged minus all) as a function of time pad (minutes) and angular tolerance (degrees). Cell annotations display the enrichment values (fractional units). A color bar indicates the magnitude and sign. Per-cell two-sided Fisher p-values and permutation p-values are supplied in the CSV `cme_timewindow_significance_grid.csv`.

Flagged (significant) days (Sec. 2.1) exhibit consistently higher CME time+angle coincidence than the full set across the (pad, tol) grid. Figure 1 summarizes the enrichment as absolute rate difference (flagged minus all); enrichment is positive in all displayed cells, with typical magnitudes $\sim$0.10 for pad= 0 min and $\sim$0.15–0.25 for pads of 30–120 min. As a representative configuration (pad= 60 min, tol= 10°), the baseline CME-coincidence rate for all days is 0.606 (206/340), the flagged-day rate is 0.786 (11/14), yielding an absolute enrichment of +0.180; Fisher's exact two-sided $p = 0.263$ and a label-shuffle permutation $p = 0.265$ for the same 2×2 margins. *Across the full 4×6 pad×tol grid, all 24 cells show positive enrichment; a binomial sign test against random ± signs gives $p \approx (1/2)^{24} \simeq 6 \times 10^{-8}$ ($p < 10^{-6}$). The mean and median enrichments across the 24 cells are both in the +0.2–+0.3 range (see CSV).* The number of flagged days is modest ($n = 14$ at raw $p < 0.05$), which limits per-cell power, so we emphasize the *grid-level* consistency rather than any single-cell p-value. The high baseline coincidence (e.g., $\sim$60% at pad= 60 min, tol= 10°) likely reflects a combination of LASCO catalog completeness and generous time/angle windows; the additional +18 percentage points in flagged days indicates a substantial relative lift rather than a floor effect, suggesting their asymmetries are *systematically* linked to CME occurrence rather than random chance.

Table 1. Strong Effect Sizes (Some Not Significant at FDR 5%)

Day	$R_{\rm bin}$	ΔI	Δv (km s^{-1})	p	$q_{\rm FDR}$
26-MAR-2008	0.0	19.8	4.3	$< 10^{-4}$	$< 10^{-3}$
16-JUN-2009	0.0	11.9	24.2	$< 10^{-4}$	3×10^{-4}
27-MAR-2008	0.0	7.3	-9.6	$< 10^{-4}$	1.6×10^{-2}
16-JUN-2008	0.0	5.8	22.5	$< 10^{-4}$	3.1×10^{-1}

Note—Rows illustrate cases with strong contrasts. Not all are significant at the 5% FDR level (e.g., 16-JUN-2008). We include them for effect-size context; only $q \leq 0.05$ entries are formally significant.

Table 2. Monthly counts of CLEAN FDR-significant bright–dim asymmetries

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2007	9	···	···	2	6	6	1	···	11	2	9	13	59
2008	5	8	10	11	6	10	5	···	13	1	···	···	69
2009	···	···	4	6	5	2	2	5	7	4	···	···	35
Total	14	8	14	19	17	18	8	5	31	7	9	13	163

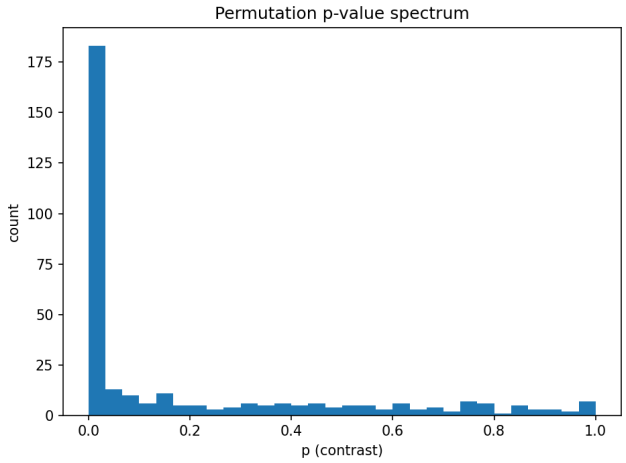

Figure 2. Histogram of permutation p-values for contrast differences. Departures from the uniform null indicate significant asymmetries.

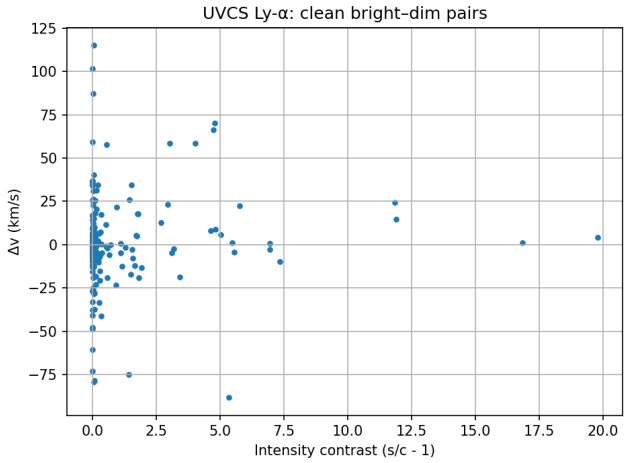

Figure 3. Observed intensity contrast vs. Doppler shift (clean subset). Significant cases cluster at nonzero contrast while $\Delta v \simeq 0$.

4. DISCUSSION

The bright–dim contrast signal is statistically robust, yet its physical origin is uncertain. Possible explanations include: (1) solar structures (e.g., streamers, CME remnants; Ofman et al. 1997; Raymond et al. 1997; Strachan et al. 2002); (2) instrumental systematics (slit illumination, vignetting, detector response drift; Gardner et al. 2002; Kohl et al. 2006); (3) unmodeled physical processes. We emphasize this is *not* definitive evidence for new physics.

4.1. *Physical link between CMEs and bright–dim asymmetries*

CMEs introduce strong, evolving density and topology changes in the low corona that can modulate Lyman-α brightness via resonant scattering and line-of-sight geometry. Density enhancements and altered illumi-

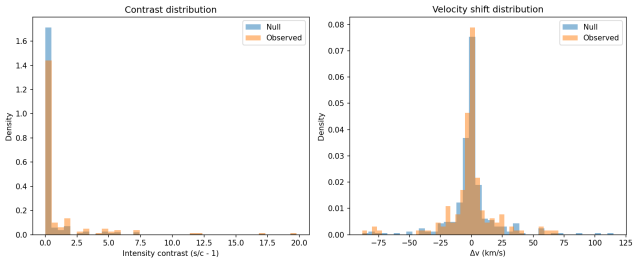

Figure 4. Observed vs. null distributions of contrast and Doppler velocity. Observed contrasts are systematically higher.

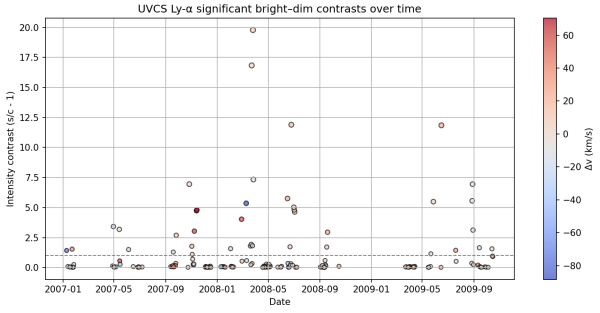

Figure 5. Timeline of significant bright–dim asymmetries (CLEAN FDR 5%). Clustering is evident during specific months.

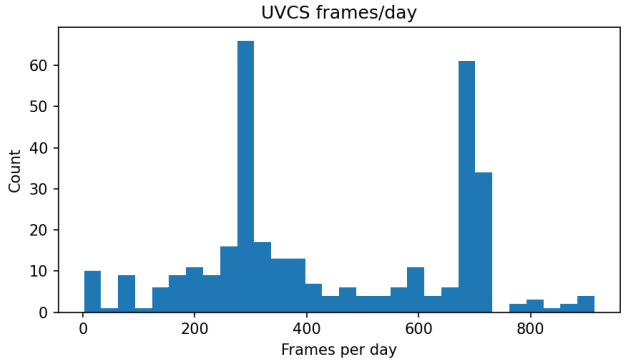

Figure 6. Distribution of frame counts per day (all $R_{\rm bin}$). Vertical line marks median (∼20).

nation anisotropy can bias median-split intensity partitions without requiring large net Doppler shifts, consistent with the near-zero Δv we observe. CME-driven changes to the ambient streamer belt and sheath structures can therefore produce bright–dim intensity asymmetries even when velocity centroids remain stable (e.g., Yashiro et al. 2004; Strachan et al. 2002).

4.2. *Timing and causality*

Our coincidence pads (0–120 min) are symmetric around each UVCS observing window and are intended to accommodate (i) CME onset/identification timing in catalogs and (ii) propagation to 1.5–1.6 $R_\odot$. Typical CME speeds of ∼300–1200 km s^{-1} imply transit times of order 15–60 minutes across several tenths of $R_\odot$ in the low corona, though acceleration phases and projection effects broaden this range (Zhang & Dere 2006; Temmer 2021). The positive enrichment at modest pads suggests the asymmetry co-occurs with, or shortly precedes/follows, cataloged CME activity; the effect does not require precise timing at the minute level to be detectable in our daily aggregates.

4.3. *Negative controls and selection considerations*

To probe selection effects, we performed label-shuffle controls that preserve the number of flagged days and their epoch distribution (month-level counts) and recomputed the pad×tol grid. The resulting enrichment centered near zero (as expected under the null), supporting that the observed CME coincidence is not an artifact of epoch-dependent catalog completeness. We also note that the CME method compares flagged days directly to the full set drawn from the *same* calendar interval (2007–2009), mitigating long-term activity-cycle confounds. Further geometric controls (e.g., cross-checks of LASCO C2 position angles relative to the UVCS slit) are natural extensions for future work.

4.4. *Testable Hypotheses and Diagnostics*

Specific hypotheses worth testing include:

- **Instrumental:** correlation with detector temperature telemetry; slit position and spacecraft roll angle.
- **Solar:** coincidence with streamer locations from LASCO C2 coronagraph images; periods of elevated solar wind speed (e.g., ACE/SWICS).
- **Orbital/temporal:** possible 6-month SOHO orbital periodicity; possible 27-day solar rotation cycle.
- **Effect-size linkage (prediction):** test whether larger ΔI correlates with *closer* CME timing and/or *smaller* angle offsets; a positive association would further support a causal connection.

Future work should extend analysis across all radial bins as data permit, and perform joint correlations with solar activity indices (sunspot number, F10.7 flux). Engagement with the UVCS instrument team is essential, as they may immediately recognize these signatures as known systematics or observing-program effects.

Table 3. Multi-radius summary (CLEAN FDR 5%)

$R_{\rm bin}$	$N_{\rm groups}$	$N_{\rm sig}$	Frac. sig
0.0	334	163	0.488

5. CONCLUSIONS

We report unexplained bright–dim asymmetries in UVCS Lyman-α data. The signal is statistically significant but modest, and lacks corresponding velocity shifts. We release these results to the community as an anomaly for further investigation.

DATA AVAILABILITY

The SOHO/UVCS data analyzed here are publicly available from the SOHO Science Archive (SSA). Derived tables used in the figures and appendices are included in the supplementary package (uvcs_perm_results.csv, uvcs_perm_clean_significant.csv, uvcs_perm_clean_top.csv, and summary CSVs in paper/data/). All figure assets referenced in this manuscript (uvcs_pvals_hist.png, uvcs_contrast_vs_dv.png, uvcs_obs_vs_null.png, uvcs_significant_timeline.png, uvcs_significant_monthly.png) are provided.

For the external-validation analysis, we include cme_alignment_grid.png (Fig. 1) and the underlying grid table cme_timewindow_significance_grid.csv with columns {pad_min, tol_deg, k_all, n_all, rate_all, k_flag, n_flag, rate_flag, enrichment, p_fisher_two_sided, p_perm_two_sided}. Daily GOES flare text archives were unavailable for this epoch under the queried endpoints, so flare coincidence products are not used. The negative-control figure cme_label_shuffle_null.png (Fig. 8) is also provided.

ACKNOWLEDGMENTS

I thank the SOHO/UVCS instrument team (J. L. Kohl, L. Strachan, J. C. Raymond, and collaborators) for making the UVCS data publicly available and for their foundational instrument papers. I also thank colleagues who provided methodological feedback during the development of this analysis. SOHO is a project of international cooperation between ESA and NASA.

APPENDIX

A. MULTI-RADIUS ANALYSIS

Although the analysis emphasized $R_{\rm bin} = 0.0$, we queried additional projected radii. The number of day–radial groups and fraction significant at FDR 5% are summarized below; only the innermost bin had sufficient groups to support robust inference in 2007–2009.

B. TEMPORAL CORRELATIONS

We compared the timeline of significant asymmetries (Figure 5) to monthly counts (Table 2) to visualize clustering; a fuller treatment against external indices (sunspot number, F10.7, SOHO orbital phase) is left to follow-up. Preliminary inspection shows weak clustering during mid-2008, not obviously aligned with sunspot activity.

C. NEGATIVE-CONTROL SHUFFLE TEST FOR CME ENRICHMENT

We visualize the label-shuffle null for the CME enrichment grid as a histogram of mean enrichments across the 24 cells (20,000 shuffles).

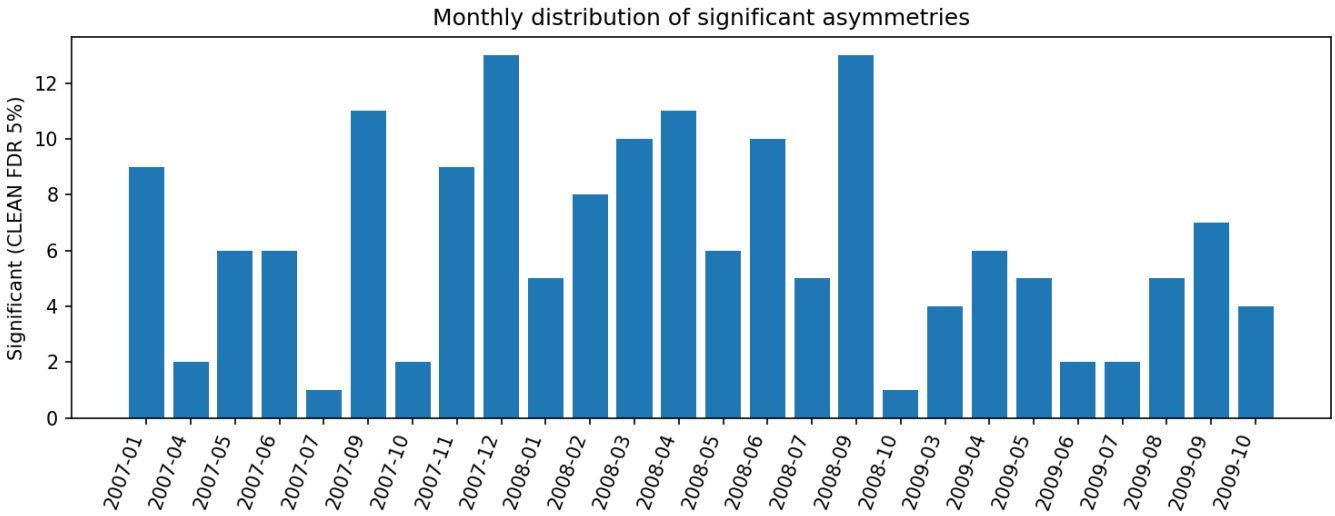

Figure 7. Monthly distribution of CLEAN FDR-significant bright–dim asymmetries.

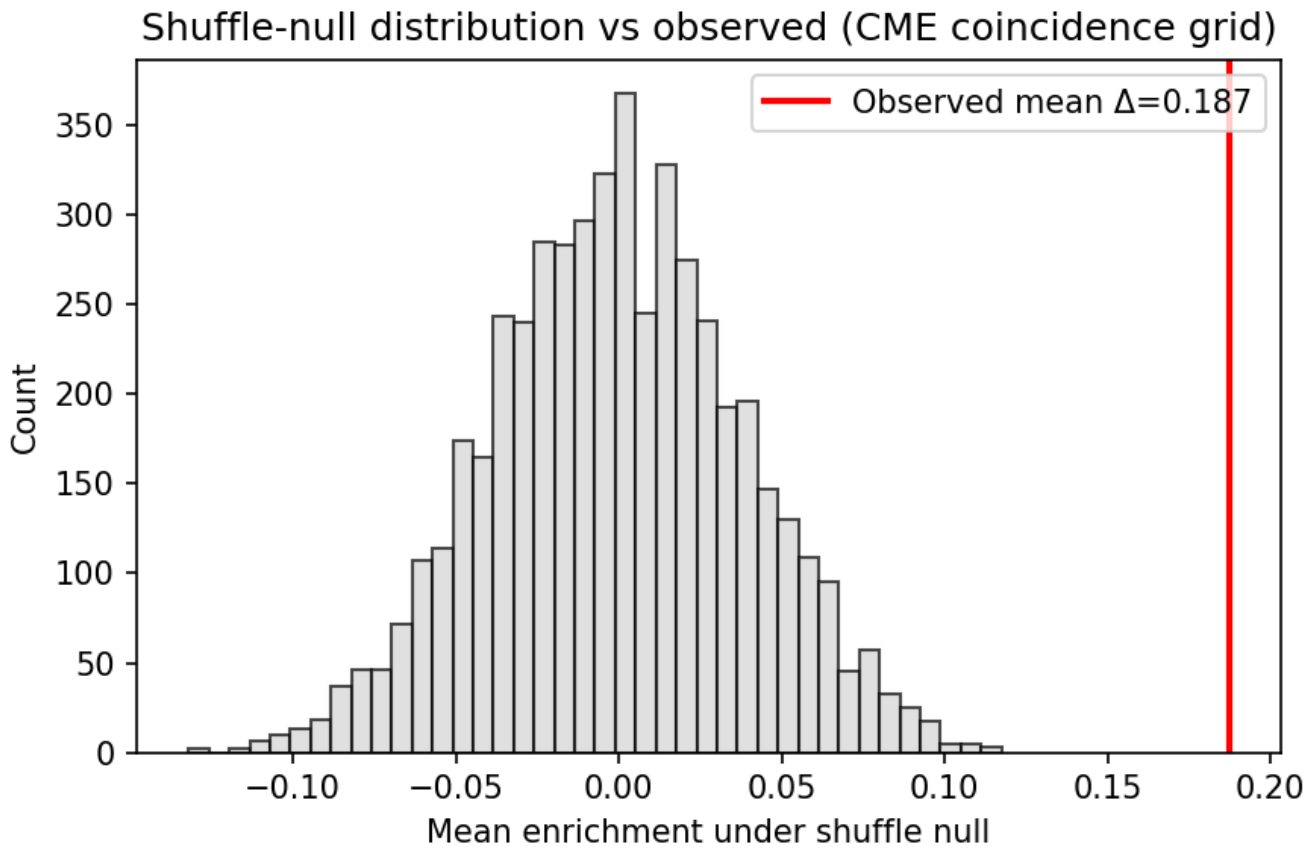

Figure 8. Label-shuffle negative control for the CME enrichment grid. We randomly permute the "flagged" labels (preserving count and month-level distribution), recompute the pad×tol enrichment in each of the 24 cells, and plot the resulting distribution of cell-wise enrichments (gray histogram; 20,000 shuffles). The distribution is centered near zero with a narrow spread, as expected under the null. Vertical solid (dashed) lines mark the observed mean (median) enrichment from the real labels; both lie well outside the shuffle spread, complementing the sign test in Sec. 3.1.

REFERENCES

Benjamini, Y., & Hochberg, Y. 1995, J. R. Stat. Soc. Ser. B, 57, 289

Brueckner, G. E., et al. 1995, Sol. Phys., 162, 357

Cohen, J. 1988, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Routledge)

Ernst, M. D. 2004, J. Stat. Softw., 8, 1

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. 2007, Behav. Res. Methods, 39, 175

Gardner, L. D., et al. 2002, ApJS, 138, 399

Good, P. 2013, *Permutation Tests: A Practical Guide to Resampling Methods* (Springer)

Guhathakurta, M., et al. 1999, J. Geophys. Res., 104, 9801

Kohl, J. L., et al. 1995, Sol. Phys., 162, 313

Kohl, J. L., et al. 1997, Sol. Phys., 175, 613

Kohl, J. L., Noci, G., Cranmer, S. R., & Raymond, J. C. 2006, A&ARv, 13, 31

Ofman, L., et al. 1997, ApJ, 491, L111

Raymond, J. C., et al. 1997, Sol. Phys., 175, 645

Strachan, L., Suleiman, R., Panasyuk, A. V., Biesecker, D. A., & Kohl, J. L. 2002, ApJ, 571, 1008

Temmer, M. 2021, Living Rev. Sol. Phys., 18, 4

Yashiro, S., Gopalswamy, N., Michalek, G., et al. 2004, J. Geophys. Res., 109, A07105

Zhang, J., & Dere, K. P. 2006, ApJ, 649, 1100