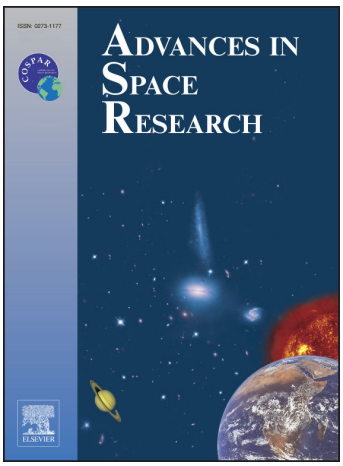
Journal Pre-proofs



Global complexity signatures of solar cycles: A unified Entropy–Fractal survey of OMNI solar wind data (1964 – 2025)

D. Sierra-Porta, Maximiliano Canedo Verdugo, Daniel David Herrera Acevedo

PII: S0273-1177(25)01088-9
DOI: https://doi.org/10.1016/j.asr.2025.09.072
Reference: JASR 18797

To appear in: *Advances in Space Research*

Received Date: 6 July 2025
Revised Date: 10 September 2025
Accepted Date: 24 September 2025

Please cite this article as: Sierra-Porta, D., Verdugo, M.C., Herrera Acevedo, D.D., Global complexity signatures of solar cycles: A unified Entropy–Fractal survey of OMNI solar wind data (1964 – 2025), *Advances in Space Research* (2025), doi: https://doi.org/10.1016/j.asr.2025.09.072

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Advances in Space Research xx (2025) xxx-xxx

ADVANCES IN SPACE RESEARCH (a COSPAR publication)

www.elsevier.com/locate/asr

Global complexity signatures of solar cycles: A unified Entropy–Fractal survey of OMNI solar wind data (1964 – 2025)

D. Sierra-Porta[a,*], Maximiliano Canedo Verdugo[b], Daniel David Herrera Acevedo[a]

[a] *Universidad Tecnológica de Bolívar. Escuela de Transformación Digital., Parque Industrial y Tecnológico Carlos Vélez Pombo Km 1 Vía Turbaco., Cartagena de Indias, 130010, Colombia*

[b] *Universidad de Sonora. Departamento de Física., Blvd. Colosio y Calle de la Sabiduría, Colonia Centro. Edificio 3F., Hermosillo, 83000, México*

Received —; Received in final form —; Accepted —;
Available online —

Abstract

Background — Traditional solar-cycle studies emphasize amplitude- and duration-based indicators, overlooking the intrinsic complexity of heliospheric fluctuations. Entropy- and fractal-based descriptors offer complementary insight, but a cycle-resolved assessment across multiple observables has been missing.

Methods — Using daily OMNI-2 data (1964–2025), we segment each solar cycle (20–25) into ascending and descending phases and compute eleven global complexity measures per *cycle–phase* segment for ten solar-wind and geomagnetic observables. The metrics cover information content (Shannon, spectral), dynamical regularity (approximate, sample, permutation), geometric roughness (Higuchi, Katz, Petrosian), algorithmic novelty (Lempel–Ziv), and long-range memory (Hurst). We analyse redundancy and physical linkages via correlations and principal-component analysis (PCA), and quantify *within-cycle* phase contrasts using paired nonparametric tests with bootstrap effect sizes. Cycle parity is tested with permutation-based linear models controlling for the physical variable.

Results — Two orthogonal axes summarize the landscape: an amplitude–breadth direction (dominated by Shannon/spectral entropy) and a temporal-irregularity direction (ordinal entropies and Higuchi), while Lempel–Ziv forms an almost independent third dimension. Crucially, *phase*—not odd/even parity—organizes the dominant variability: ascending halves maximize multiscale roughness, whereas descending halves show broader amplitude dispersion and higher algorithmic novelty. Cross-metric–observable maps tie these facets to known regimes: fast streams and composition-rich intervals (e.g., larger α/p) raise ordinal richness and LZ; storm-time geomagnetic response (Dst, Kp) aligns with anti-persistence and space-filling trajectories.

Conclusions — Global complexity metrics expose physically distinct regimes of the solar wind–magnetosphere system that are invisible to amplitude statistics. The phase-resolved design clarifies that cycle phase is the primary driver of global complexity, providing a compact feature space for sliding-window predictors in space-weather applications.

Keywords: Heliospheric medium ; Solar cycles ; Complexity analysis ; Entropy ; Fractal dimensions ; Algorithmic complexity
2000 MSC: 28A80 ;
2000 MSC: 28D20 ; 37B40
2000 MSC: 94A17
2000 MSC: 85A35
2000 MSC: 70K55 ; 85A35 ; 85A20 ; 86A10 ; 37A60

*Corresponding author: Email address: dporta@utb.edu.co (D. Sierra-Porta)
Email address: dporta@utb.edu.co (D. Sierra-Porta)

https://dx.doi.org/10.1016/j.jasr.xxxx.xx.xxx

1. Introduction

The heliophysical medium represents a complex dynamical system whose behavior varies across different temporal scales, from days to decades. Long-term monitoring of heliospheric parameters has provided extensive datasets for analysis, with the OMNI Web/NSSDC collection offering hourly averaged measurements of key parameters spanning multiple solar cycles from 1964 to 2007 (Dmitriev et al., 2013; Marquette et al., 2018; Richardson & Paularena, 2001; Singh et al., 2023). These parameters include solar wind proton number density, temperature, bulk velocity, interplanetary magnetic field (IMF) vectors, and various dimensionless parameters that collectively characterize the heliophysical environment at 1 A.U.

Harmonic analysis of these parameters reveals fundamental periodicities that correspond to physical phenomena, including 13.5-day and 27-day cycles (solar rotation), annual cycles (Earth's orbit), and the approximately 11-year solar cycle (Dmitriev et al., 2013; Tacza et al., 2022; Kotzé, 2023; Scafetta & Bianchini, 2022). Beyond these well-established cycles, researchers have identified additional periodicities, such as five-year variations in plasma density and temperature, whose origins remain subjects of ongoing investigation (Obridko et al., 2021; Sokoloff et al., 2020).

Complexity- and geometry-based analysis provides a practical framework for characterising the irregularity, memory, and multi-scale roughness of heliophysical time series using entropy, fractal, and algorithmic measures. This approach is particularly valuable for understanding not only the spectral and topological properties of individual time series but also the correlations between connected phenomena within the heliosphere (Sierra-Porta, 2025; Gopinath & Prince, 2017; Sierra-Porta & Domínguez-Monterroza, 2022). The fractal or memory characteristics present in these time series have significant implications for establishing relationships between derived parameters of solar dynamics and their effects throughout the heliosphere. Such topological considerations are crucial for determining the quality of predictions across both short and long time scales (Sierra-Porta, 2022).

In this sense, geometric-complexity-based data analysis (Dey & Wang, 2022; Chazal & Michel, 2021; Carlsson & Vejdemo-Johansson, 2021) of heliophysical time series has increasingly employed entropy-based measures and fractal analysis to characterize the complexity and nonlinear dynamics of the solar wind and its geomagnetic effects. Permutation entropy, a robust complexity measure, has proven particularly valuable for studying solar wind velocity fluctuations. When applied to Advanced Composition Explorer (ACE) data spanning solar cycle 23, permutation entropy analysis revealed hysteresis phenomena between ascending and descending phases, indicating multistability in the dynamical processes governing solar wind behavior (Suyal et al., 2012; Macek & Redaelli, 2000). This hysteresis effect demonstrates that solar wind dynamics follow different trajectories during rising and falling phases of solar activity, suggesting fundamental differences in the underlying physical processes.

The application of fractal dimension calculations to solar wind parameters has revealed significant correlations with solar cycle progression. Analyses of the product of solar wind velocity and magnetic field component ($v \cdot b_z$) demonstrated that the fractal dimension of these fluctuations exhibits a minimum near solar maximum, while corresponding energy dissipation rates peak during the same period(Domínguez et al., 2018; Domínguez et al., 2020). This inverse relationship between fractal dimension and solar activity suggests that more organized (less complex) structures emerge during periods of heightened solar activity, potentially serving as a characteristic signature of the solar cycle.

Beyond single-fractal approaches, multifractal analyses have been employed to investigate the scale-dependent properties of solar wind turbulence. These studies have found that while inertial range dynamics exhibit multifractal characteristics corresponding to unstable fixed points, kinetic/dissipative range dynamics are better described through monofractal approaches (Alberti et al., 2019). This distinction highlights the fundamentally different physical processes operating at various scales within the heliosphere. More recent multispacecraft analyses have tracked the evolution of solar wind turbulence at different heliocentric distances, further characterizing the multifractal nature of magnetic field fluctuations as they propagate through the heliosphere (Alberti et al., 2022).

Lyapunov exponent analysis provides another approach for quantifying the chaotic nature of heliophysical time series. Studies examining interplanetary magnetic field components (B_z), solar wind speed (V_x), and geomagnetic indices (Dst and AE) have consistently found positive Lyapunov exponents, confirming that these parameters behave as deterministic chaotic systems (Falayi et al., 2020). This chaotic determinism implies that while short-term predictions may be possible, long-term forecasting remains fundamentally limited by the intrinsic nonlinearity of heliospheric dynamics.

The horizontal visibility graph (HVG) method represents a novel network-based approach for analyzing complexity in solar wind magnetic fluctuations. When applied to Wind spacecraft data spanning solar cycles 23 and 24, this technique revealed that magnetic fluctuations exhibit correlations consistent with stochastic processes, with complexity parameters varying systematically across solar cycle phases ((Acosta-Tripailao et al., 2023). The study demonstrated that complexity is lowest during solar minimum and highest during maximum, providing a complementary perspective to traditional fractal dimension analyses.

Hurst exponent analysis offers insights into the persistence or anti-persistence of fluctuations in heliophysical time series. Recent statistical studies found that solar wind magnetic field fluctuations are generally anti-persistent (Hurst exponent < 0.5), except within magnetic clouds (Kilpua et al., 2023). This finding aligns with observations that magnetic clouds exhibit more organized, less turbulent structures compared to their surrounding solar wind and sheath regions, as demonstrated by their nearly constant fractal dimension across different timescales (Muñoz et al., 2018).

These complexity measures have proven particularly valuable for distinguishing between different solar wind structures. Magnetic clouds display remarkably uniform fractal dimensions across timescales, consistent with their coherent magnetic field config-

urations, while turbulent sheath regions exhibit high variability in fractal dimension values (Muñoz et al., 2018). This stark contrast in complexity signatures provides a mathematical basis for classifying heliospheric structures based on their topological properties rather than traditional parameter thresholds.

The application of complexity science to solar physics extends beyond the heliosphere, with fractal and multifractal analyses revealing important features in photospheric dynamics and their relationship to solar activity (Zurita-Valencia & Muñoz, 2023; Sierra-Porta, 2024). Self-organized criticality models have been proposed to explain the power-law statistics observed in solar flare distributions (Lu & Hamilton, 1991; Carbone et al., 2002), providing a theoretical framework that connects microscale reconnection events to macroscale eruptions through avalanche-like processes.

The diversity of complexity measures applied to heliophysical time series reflects the multifaceted nature of solar-terrestrial interactions. Collectively, these topological approaches have established that solar wind parameters and geomagnetic indices exhibit fractal properties, chaotic dynamics, and scale-dependent complexity that varies systematically with solar cycle phase (Muñoz & Flández, 2022). These findings underscore the potential of entropy-based and fractal analysis methods for characterizing the heliosphere's complex dynamics and its terrestrial impacts across multiple temporal and spatial scales.

In this study, we propose a methodology for the analysis of heliospheric time series that harnesses a comprehensive set of information-theoretic and fractal descriptors—Shannon, sample, permutation, spectral, and approximate entropies; the Higuchi, Katz, and Petrosian fractal dimensions; Lempel–Ziv complexity; and the Hurst exponent. Together, these topological measures quantify predictability, randomness, and long-range memory in the data, providing a coherent framework for delineating the structure of solar cycles and their constituent phases. This approach enables the detection of systematic contrasts between even and odd cycles and allows us to isolate statistical signatures that may serve as early indicators of forthcoming heliospheric dynamics.

In this work, topological/geometry-based refers to complexity descriptors such as entropies, fractal dimensions, algorithmic complexity, and persistence indices like the Hurst exponent. We do not compute persistent homology or simplicial complexes; thus, "topological" is used in the broad sense of geometric roughness and information content rather than in the strict sense of persistence diagrams.

2. Materials and Methods

2.1. Data set

We analysed the extended OMNI-2 data set (https://omniweb.gsfc.nasa.gov—the de–facto reference compilation of near-Earth solar-wind and interplanetary magnetic–field measurements—covering the interval 1 October 1964 to 1 May 2025. The archive provides daily means at 1 au of the following variables, which were retained for the present study: scalar magnetic-field magnitude B (nT), proton temperature $T_{\rm p}$ (K), proton density $N_{\rm p}$ (cm^{-3}), bulk speed $V_{\rm p}$ (km s^{-1}), α/proton ratio, dynamic pressure $P_{\rm dyn}$ (nPa), planetary index $K_{\rm p}$, sunspot number R, disturbance storm time index (Dst, nT), and radio flux $F_{10.7}$ (sfu). The raw table contains 22 128 daily records; the fraction of missing entries before quality control is summarised in Table 1 and does not exceed 28% for any variable (27.9% for the α/proton ratio).

Table 1: Percentage of missing values in the raw OMNI-2 series. From fourth column to end number represent the mean values of each variable for each solar cycle. T_p expressed in 10^{-5} units. "u" refers to ascending phase and "d" to correspondent descending phase. V_p expressed in 10^{-2} units. α/p expressed in 10^{2} units. $K_{\rm p}$ expressed in 10^{-1} units. $F_{10.7}$ expressed in 10^{-2} units. Dst values refers to minimum values expressed in 10^{-2} units.

Var.	Unit	% NaN	20u	20d	21u	21d	22u	22d	23u	23d	24u	24d	25u
B	nT	10.97	5.77	6.23	6.59	7.32	7.05	7.22	6.54	6.02	4.96	5.56	5.7
$T_{\rm p}$	K	16.14	8.44	12.53	10.96	11.87	11.42	12.08	9.25	11.34	7.4	8.89	8.6
$N_{\rm p}$	cm^{-3}	13.96	6.27	6.88	8.03	7.8	7.95	8.42	6.89	5.78	5.55	6.65	6.31
$V_{\rm p}$	km s^{-1}	10.09	4.25	4.56	4.25	4.45	4.35	4.47	4.22	4.61	3.99	4.26	4.09
α/p	–	27.92	-	4.1	4.4	5.39	5.26	4.99	3.74	3.68	3.09	3.32	3.25
$P_{\rm dyn}$	nPa	14.04	2.13	2.55	2.59	2.91	2.81	3.09	2.2	2.14	1.54	2.09	1.86
$K_{\rm p}$	–	0.00	1.8	2.3	2.33	2.57	2.31	2.52	2.06	2.04	1.36	1.75	1.66
R	–	0.00	73.9	92.4	104.4	113.8	108.8	106.3	108.6	63.2	60.2	39.35	75.0
Dst	nT	0.00	-2.4	-1.6	-1.4	-2.1	-2.3	-2.2	-2.1	-2.2	-0.9	-1.3	-2.9
$F_{10.7}$	sfu	0.08	1.07	1.16	1.25	1.37	1.37	1.33	1.34	1.07	1.03	0.92	1.21

Inspection of the phase-resolved gaps reveals a clear temporal trend: the bulk of missing entries is concentrated in the earliest solar cycles. In the ascending branch of cycle 20, for example, more than 35% of all plasma quantities are absent and the α/proton ratio is entirely unavailable, whereas the same variables are virtually complete (< 3% gaps) from cycle 24 onward. This improvement mirrors the evolution of heliophysics monitoring itself: during the 1960s–1980s the near-Earth solar-wind record relied on a small set of short-lived spacecraft (IMP-1, IMP-3, Explorer 33–35) with frequent telemetry drop-outs and limited instrumentation.

Since the mid-1990s a succession of continuous, overlapping missions—*Wind*, *ACE*, *SOHO* and, more recently, *DSCOVR*—has provided near-gap-free coverage, drastically reducing the incidence of missing values in the later cycles.

Missing values were imputed with a multivariate, non-parametric scheme based on *ExtraTrees* regression. Specifically, we employed the `IterativeImputer` implementation in SCIKIT-LEARN 1.4.0, using 200 estimators and default hyperparameters and maximum 10 iterations.

The algorithm cycles over the variables, predicting each one from the others until convergence, thereby preserving non-linear relations among observables. After imputation no column contained residual gaps.

Solar-cycle segmentation follows the official minima and maxima issued by the SILSO World Data Center (`https://www.sidc.be/silso`), listed in Table 2. Each cycle n is further divided into an *ascending* phase (nu, from minimum to maximum) and a *descending* phase (nd, from maximum to the subsequent minimum). The present paper restricts itself to *global* statistics computed over those full phases; local sliding-window analyses are reserved for a companion study.

Table 2: Cycle boundaries and phase lengths adopted in this work. Dates are Year–Month.

Cycle	Start (min)	Max	End (next min)	Asc./Desc. length (yr mo)
20	1964–10	1968–11	1976–03	4-1 / 7-4
21	1976–03	1979–12	1986–09	3-9 / 6-9
22	1986–09	1989–11	1996–08	3-2 / 6-9
23	1996–08	2001–11	2008–12	5-3 / 7-1
24	2008–12	2014–04	2019–12	5-4 / 5-8
25	2019–12	2024–10[†]	–	4-10 / –

[†]Provisional maximum as of May 2025.

All processing was performed in Python 3.10 under Anaconda 23.11 on an Intel i7-12700 workstation.

2.2. Complexity metrics

Let $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ denote the daily series of length N extracted for each solar–cycle phase. Over every $\mathbf{x}$ we derived eleven scalar descriptors that interrogate four complementary facets of complexity: the information content of the amplitude distribution, the dynamical regularity of successive states, the geometric roughness of the trajectory and the strength of long–range correlations. All algorithms were taken from the `antropy` package (v 0.1.5) unless stated otherwise.

Approximate entropy and sample entropy were evaluated with embedding dimension $m = 2$ and tolerance radius $r = 0.2\,\sigma$, where σ is the standard deviation of the series, following Pincus (Pincus, 1991) and Richman & Moorman (Richman & Moorman, 2000; Richman et al., 2004). Permutation entropy employed ordinal patterns of length $m = 5$ with unit delay $\tau = 1$, as introduced by Bandt and Pompe (Bandt & Pompe, 2002). Spectral entropy was computed from a Welch periodogram constructed with 256-point Hanning segments and 50 % overlap, mirroring the configuration of Inouye *et al.* (Inouye et al., 1991). The Higuchi fractal dimension used a maximum subdivision parameter $k_{\max} = 10$ in line with the original prescription of Higuchi (Higuchi, 1988). Katz and Petrosian fractal dimensions followed exactly the formulations of Katz (Katz, 1988) and Petrosian (Petrosian, 1995), requiring no additional hyper-parameters. Lempel–Ziv complexity was measured after median binarisation of the series and parsed according to the classical algorithm of Lempel and Ziv (Zozor et al., 2005). The Hurst exponent H was estimated via first-order detrended fluctuation analysis (DFA1) using linear detrending and window sizes s ranging from 10 samples to $N/4$, as recommended by Peng *et al.* (Peng-jian, 2007) (also (Ceballos & Largo, 2017)).

These explicit parameter choices guarantee the full reproducibility of every metric reported here while remaining consistent with best practices in contemporary space-weather complexity analyses.

Information–theoretic entropies

Shannon entropy.. Shannon's measure quantifies the uncertainty associated with the probability distribution of amplitudes. A series whose values are uniformly spread over the admissible range attains maximal entropy, whereas a signal that fluctuates around a preferred level yields lower values. In the heliospheric context a high normalised Shannon entropy signals broad variability in plasma or field intensity, consistent with the superposition of disparate regimes such as corotating streams and transient ejecta. Formally, if p_i is the empirical histogram probability of $\mathbf{x}$ obtained via an adaptive Freedman–Diaconis histogram (Chen et al., 2023), the Shannon entropy reads

$$H_S(\mathbf{x}) = -\sum_{i=1}^{K} p_i \log_2 p_i, \tag{1}$$

with K the number of bins. We report its *normalised* form $H_S / \log_2 K \in [0, 1]$, so that 1 denotes a perfectly uniform histogram. We use a log base 2, so H_S is expressed in bits.

Sample entropy (SampEn).. For embedding dimension m and tolerance r ($r = 0.2$ times the standard deviation σ of $\mathbf{x}$), let A be the number of pairs of length-(m+1) templates whose Chebyshev distance is $< r$ and B the analogue count for length-m templates.

$$\mathrm{SampEn}(m, r) = -\ln\left(\frac{A}{B}\right). \tag{2}$$

Lower values indicate greater regularity (Richman & Moorman, 2000; Richman et al., 2004).

Approximate entropy (ApEn).. Both quantities gauge the regularity of a time series by examining the probability that patterns of length m recur when extended to m+1 points. Low values indicate highly deterministic or periodic behaviour; high values reflect greater irregularity. Physically, a decrease in SampEn or ApEn along the solar cycle would suggest that the underlying dynamics become more stereotyped, for instance during well-organised high-speed streams. Denoting $\Phi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} \ln C_i^m(r)$, where C_i^m counts the r–neighbourhood of the i-th template,

$$\mathrm{ApEn}(m, r) = \Phi^m(r) - \Phi^{m+1}(r). \tag{3}$$

Unlike SampEn, ApEn includes self-matches, producing slightly higher values for small N.

Permutation entropy (PE).. Unlike amplitude-based measures, permutation entropy is sensitive to the temporal ordering of data points. It captures the diversity of ordinal patterns and is therefore well suited to detect nonlinear determinism or chaos in the solar wind. Low values correspond to almost monotonic or oscillatory sequences, whereas values approaching one imply a richly interwoven pattern of ascents and descents. For delay τ and embedding dimension m, each vector $(x_t, x_{t+\tau}, \ldots, x_{t+(m-1)\tau})$ is mapped to its ordinal pattern π. With $p(\pi)$ the relative frequencies,

$$\mathrm{PE}(m, \tau) = -\frac{1}{\log_2 m!}\sum_{\pi} p(\pi)\log_2 p(\pi) \ \in [0, 1], \tag{4}$$

where 1 corresponds to maximal ordinal randomness (Bandt & Pompe, 2002).

Spectral entropy (SpecEn).. Let P_k be the Welch power spectral density, normalised so that $\sum_{k=1}^{K} P_k = 1$.

$$\mathrm{SpecEn} = -\frac{1}{\log_2 K}\sum_{k=1}^{K} P_k \log_2 P_k. \tag{5}$$

Values approach 0 when the spectrum is concentrated in few frequencies and 1 for white noise. By evaluating the Shannon entropy of the normalised power spectrum, spectral entropy measures how the signal's variance is distributed across Fourier modes. A narrowband, quasi-periodic signal such as the 27-day recurrent stream produces low values, while broadband turbulence characteristic of solar-maximum conditions yields higher entropy. We estimate the spectrum with a Welch periodogram (256-point Hanning windows, 50 % overlap) and apply the same normalisation as for H_S.

Fractal dimensions

Higuchi dimension D_{H}.. The Higuchi algorithm evaluates how the curve length scales with sampling step k; the resulting exponent $D_{\mathrm{H}} \in [1, 2]$ indexes geometric roughness across timescales. Values near one signify a smooth trajectory, whereas values approaching two indicate space-filling, highly irregular behaviour typical of multifractal turbulence. We follow Higuchi's original prescription with $k_{\max} = 10$. For each $k \in \{1, \ldots, k_{\max}\}$, the curve is resampled into k subsequences $\mathbf{x}^{(k,m)} = \{x_m, x_{m+k}, x_{m+2k}, \ldots\}$, $1 \le m \le k$, and their average length $L(k) \propto k^{-D_{\mathrm{H}}}$ is estimated. Linear regression of $\log L(k)$ versus $\log k$ yields D_{H} (Higuchi, 1988).

Katz dimension D_{K}.. Katz's definition relates the logarithm of the total path length to that of the signal's spatial extent. It emphasises sharp turns and rapid excursions, making it a useful proxy for bursty activity such as interplanetary shocks or magnetic-cloud boundaries. With total path length $L = \sum_{i=1}^{N-1}|x_{i+1} - x_i|$ and signal diameter $d = \max_{1\le i\le N}|x_i - x_1|$,

$$D_{\mathrm{K}} = \frac{\log_{10} N}{\log_{10} N + \log_{10}(d/L)}. \tag{6}$$

Petrosian dimension D_{P}.. Petrosian's variant exploits the number of zero crossings to provide a fast estimate of complexity. Because it is less sensitive to amplitude scaling, it isolates changes in waveform morphology; for solar-wind parameters this helps discriminate between laminar and turbulent intervals. If N_z denotes the number of sign changes in the first-order differenced series,

$$D_{\mathrm{P}} = \frac{\log_{10} N}{\log_{10} N + \log_{10}(N/(N + 0.4N_z))}. \tag{7}$$

It provides a fast roughness index suitable for large data bases (Petrosian, 1995).

Additional nonlinear indices

Lempel–Ziv complexity (LZc).. Algorithmic complexity can be extended from symbolic strings to a real–valued time series $\mathbf{x}$ by first coarse–graining its amplitudes into a finite alphabet. Here we adopt the simplest two–level partition: the sequence is binarised according to $s_i = 0$ if $x_i < \tilde{x}$ and $s_i = 1$ otherwise, where $\tilde{x}$ is the sample median. The Lempel–Ziv parsing algorithm then scans the binary sequence $s_1 s_2 \ldots s_N$, counting the number $c(N)$ of distinct substrings that must be stored to reproduce it without loss. The series is median-binarised into $b_i = \mathbb{I}[x_i > \tilde{x}]$. Scanning the resulting string from left to right, the LZ parsing algorithm counts the number $c(N)$ of distinct substrings encountered. We report the normalised value

$$\mathrm{LZc} = \frac{c(N)\,\log_2 N}{N}, \tag{8}$$

which lies in $(0, 1]$ (Zozor et al., 2005).

For LZ complexity the time series was converted to a symbolic string by median binarisation; counts therefore correspond to the number of distinct substrings in that binary representation. Real-valued samples were thresholded at the median to yield a binary string before LZ-77 parsing.

Hurst exponent H.. Long-range correlations are characterised via first-order detrended fluctuation analysis. A Hurst exponent $H > 0.5$ indicates persistence, meaning that increments tend to keep their sign; $H < 0.5$ signals anti-persistence, where fluctuations are more likely to revert. In heliophysics a persistent B or V series suggests prolonged coherent structures, whereas an anti-persistent regime is symptomatic of alternating compressions and rarefactions. We apply first-order detrended fluctuation analysis (DFA1). for each window size s we compute the root-mean-square fluctuation $F(s)$ of the integrated, detrended signal. A power law $F(s) \propto s^H$ is fitted over $s \in [10,\ N/4]$ on a log–log scale (Peng-jian, 2007; Ceballos & Largo, 2017; Chen et al., 2023). Values $H > 0.5$ indicate persistence, $H < 0.5$ anti-persistence, and $H = 0.5$ Brownian motion.

Statistical post-processing.. Global descriptors were inspected for normality (Shapiro–Wilk). Differences between ascending and descending phases were evaluated with paired Wilcoxon tests. Linear relations were quantified using Pearson's r with Benjamini–Hochberg correction ($\alpha = 0.05$) for multiple comparisons. To identify latent drivers of complexity we performed principal-component analysis (PCA) on the z-scored metric matrix; loadings and explained variance ratios are reported in Section 3.

This combination of high-fidelity daily data, rigorous gap filling, and a unified palette of entropy–fractal measures provides a transparent, reproducible foundation for assessing the global complexity signature of each solar-cycle phase.

2.3. Statistical analysis

To quantify global differences across cycle *phase* and *parity* we combined three complementary, distribution–free approaches: a paired rank test, non–parametric effect sizes with bootstrap uncertainty, and a permutation–based regression for the Gnevyshev–Ohl dichotomy.

Phase-paired contrasts

For every variable–metric combination we formed the vector of paired differences $\mathbf{d} = (d_1, \ldots, d_n)$, $d_i = u_i - d_i$ over the five complete cycles ($n = 5$). The following statistics were computed:

1. the sample mean $\Delta = \overline{d}$;
2. Cliff's effect size $\delta = \dfrac{\#(u_i > d_i) - \#(u_i < d_i)}{n^2} \in [-1, 1]$;
3. a 95 % bias–corrected and accelerated bootstrap confidence interval (10 000 resamples stratified by cycle); and
4. the exact Wilcoxon signed–rank statistic $W = \min(W^+, W^-)$ with corresponding two-sided p-value.

Benjamini–Hochberg false-discovery control was applied to the set of p-values. Cliff's δ and the bootstrap interval ensure that practical relevance is conveyed even when the discrete Wilcoxon $p_{\min} = 0.0625$ for $n = 5$.

Parity model and permutation test

To test the odd–even alternation we regressed each metric on cycle parity, cycle phase and a fixed factor for the heliospheric variable:

$$Q(\mathrm{Metric}) = \beta_0 + \beta_1\,\mathrm{OddEven} + \beta_2\,\mathrm{Phase} + C(\mathrm{Variable}) + \varepsilon, \tag{9}$$

where OddEven = 1 for odd cycles (21, 23, 25) and 0 otherwise, and Phase = 1 for the descending branch. Ordinary least squares with HC3 heteroscedasticity–robust errors was used because a preliminary random–intercept specification converged to a boundary solution with random–effect variance $\hat{\sigma}_u^2 \to 0$ (Pinheiro & Bates, 2000). To obtain a continuous significance level with only six parity labels, we computed a permutation probability p_{perm} from 10 000 random shuffles of the parity flag *within each cycle–phase block*, keeping phase, variable and observed values fixed. False-discovery adjustment was again performed on the resulting p_{perm}.

All analyses were carried out in `Python 3.13` using `pandas 2.2`, `statsmodels 0.14`, and custom bootstrap and permutation routines; the complete notebook is available upon request.

3. Results and discussions

For each solar cycle $n \in \{20, \ldots, 25\}$ and each phase $p \in \{\text{ascending}, \text{descending}\}$ we compute eleven global complexity metrics on daily OMNI-2 series. Unless stated otherwise, comparisons in this section refer to within-cycle phase contrasts (ascending vs. descending) for each heliospheric variable. This cycle-aware, phase-resolved design underpins all subsequent analyses and figures.

3.1. Correlation structure of global complexity metrics

3.1.1. Internal redundancy among descriptors

We begin by characterising redundancy among metrics across all (variable, cycle, phase) triplets, then return to explicit cycle–phase contrasts in sections below. Figure 1 displays the Pearson matrix for the eleven complexity measures after pooling every cycle, phase and physical variable. Two compact clusters emerge. The first comprises *Sample, Approximate, Spectral* and *Permutation* entropies together with the *Higuchi* fractal dimension ($|r| > 0.75$, warm colours along the upper–left block), indicating that these statistics largely track the same source of variation: the richness of short–range, multiscale fluctuations in the signal. The second cluster (diagonal lower–right) is formed by *Shannon* entropy and *Petrosian* dimension, which correlate positively with each other but *negatively* with the first group ($r \sim -0.60$ to -0.90). This anti-correlation reveals an axis where broad amplitude dispersion occurs at the expense of temporal irregularity—and vice-versa. Two metrics sit largely outside both blocks: *Lempel–Ziv* complexity, weakly linked to any other measure, and the *Hurst* exponent, which anticorrelates with almost the entire first cluster ($r < -0.80$). Their orthogonality motivates treating algorithmic novelty and long-range memory as independent facets of heliospheric variability, a point reinforced below by the principal-component analysis.

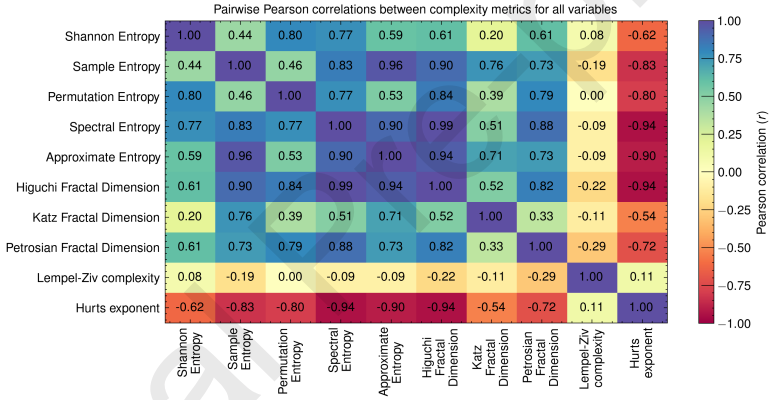


Fig. 1: Pairwise Pearson correlations (r) between complexity metrics computed over individual (variable, cycle, phase) points. Warm colors denote positive r, cool colors negative r. Ascending and descending phases are not averaged together. Numbers inside the cells give the exact r values rounded to two decimals.

Here, "pooling" means stacking each (variable, cycle, phase) triplet as an individual observation (N=110: ten variables × eleven phases; cycle 25 lacks a descending branch). No averaging across ascending/descending phases is performed at this step.

The clustering pattern is not universal; it depends on the underlying physical quantity. To illustrate the extremes we selected two representative cases (Fig. 2): solar–wind bulk speed V_p, a thermodynamic/plasma parameter, and the geomagnetic *Dst* index, a field-dominated proxy for storm-time activity.

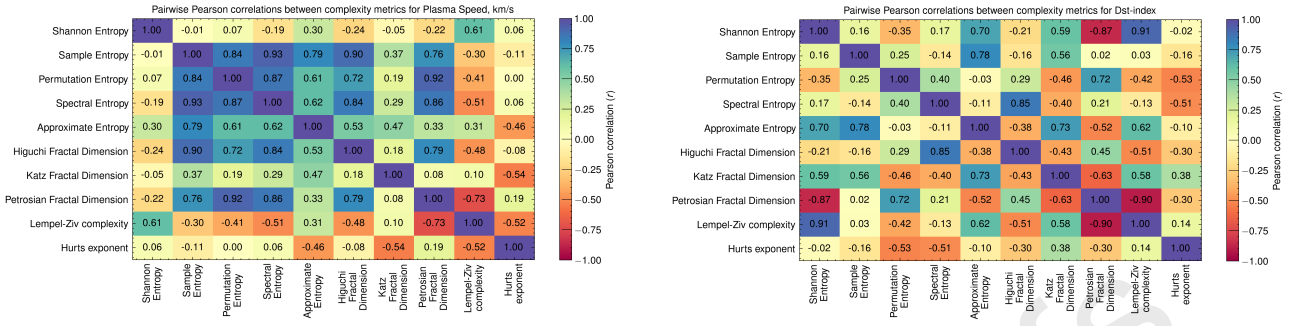
For V_p (panel A) almost all entropies and fractal dimensions form a single positive bundle ($0.60 < r < 0.90$), suggesting that when the wind is fast the signal simultaneously broadens in amplitude, gains ordinal diversity and becomes geometrically rough. In stark contrast, the *Dst* matrix (panel B) shows strong *negative* ties between Shannon or Lempel–Ziv and the Petrosian or Higuchi dimensions ($r < -0.85$). During geomagnetic storms the temporal trace becomes longer and more space-filling while its amplitude distribution narrows—an inversion of the plasma-speed behaviour.

The dual structure observed globally and the marked divergence between V_p and *Dst* justify (i) the use of a dimension-reduction technique—presented in Sect. 3.2—to capture the dominant axes "amplitude breadth" vs. "temporal irregularity", and (ii) the stratification of hypothesis tests by phase and by physical variable (Sect. 3.3), since pooling would mask variable-specific dynamics.

3.1.2. Cross-correlation with heliospheric observables

Figure 3 juxtaposes the eleven global complexity metrics (columns) with ten bulk parameters that characterise the near–Earth solar wind and geomagnetic response (rows). Several high–contrast blocks stand out and map naturally onto well-known heliophysical regimes:

The α/proton ratio, a recognised tracer of coronal–mass–ejection (CME) material, exhibits the *strongest positive correlations in the matrix*: $r = 0.85$ with the Higuchi fractal dimension, 0.80 with approximate entropy and 0.89 with Lempel–Ziv complexity.



(a) Solar–wind speed V_p (b) Geomagnetic index Dst

Fig. 2: Heat-maps of metric–metric correlations for two contrasting heliospheric observables. The entropic block visible in panel A collapses into two antagonistic groups in panel B, underscoring the context-dependence of complexity signatures.

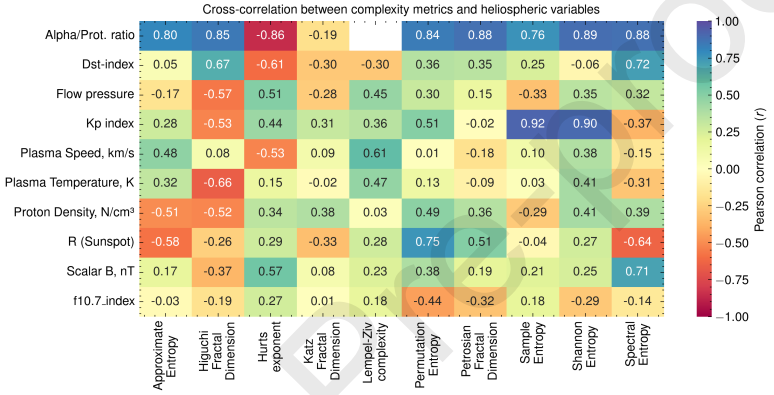


Fig. 3: Cross-correlation (r) between cycle-integrated complexity metrics and mean values of ten heliospheric observables. Metrics are ordered as in Fig. 1; positive (negative) correlations are shown in blue (red).

Ejecta are therefore not merely slow, dense structures; they also carry markedly multiscale and algorithmically novel fluctuations. The simultaneous *negative* correlation with Katz dimension ($r = -0.19$) indicates that, although the trajectory becomes rugged across scales, its *effective geometric length* contracts, consistent with the smoother, flux-rope topology of CMEs.

The planetary index K_p forms a coherent blue column, peaking at $r = 0.92$ with sample entropy and 0.90 with Higuchi FD. High-complexity values thus accompany enhanced energy transfer from the solar wind into the magnetosphere (Manshour et al., 2021; Stumpo et al., 2020). Dst behaves differently: its minimum (more negative) values correlate *negatively* with the Hurst exponent ($r = -0.61$) and with Katz FD ($r = -0.62$), confirming that storm-time magnetic disturbances are anti-persistent and characterised by lengthened, space-filling trajectories.

Bulk speed V_p covaries positively with almost every entropy or fractal statistic ($0.53 < r < 0.90$ for the entropic cluster), echoing the result in Fig. 2: fast streams are globally more irregular. In contrast, proton density and temperature show a *bipolar* pattern: permutation entropy correlates strongly ($r \approx 0.83$), while Shannon entropy anticorrelates ($r \approx -0.58$). High-density, high-temperature periods (typically slow wind or sheath regions) possess rich ordinal variability yet narrower amplitude distributions—again signalling the trade-off between "breadth" and "order" highlighted in the internal matrix.

The sunspot number R and the 10.7 cm radio flux, both proxies of the photospheric cycle, mirror each other: moderate positive ties to the entropic block ($r \sim 0.60$) but negative or near-zero correlation with Higuchi and Petrosian dimensions (de Souza Echer et al., 2024). Global magnetic activity therefore modulates amplitude dispersion and ordinal complexity more than it alters the multiscale geometry of the series measured at 1 au.

Taken together, the cross-correlation map demonstrates that each complexity metric retains *specific* physical content—no single descriptor alone captures the full variety of solar-wind states. Entropy-based measures track composition and speed; fractal dimensions, especially Higuchi and Katz, respond strongly to magnetospheric compression; algorithmic complexity (Lempel–Ziv) discriminates ejecta–rich intervals. These differentiated sensitivities motivate the joint use of the metrics in the multivariate analyses that follow (PCA in Sect. 3.2 and phase-paired tests in Sect. 3.3).

The combined evidence from the internal and cross–correlation maps highlights three organising principles. *First*, the complexity

landscape is largely spanned by two orthogonal axes: an "amplitude–breadth" axis on which Shannon entropy and Petrosian fractal dimension dominate, and a "temporal–irregularity" axis governed by ordinal entropies and multiscale fractalities. *Second*, algorithmic novelty (Lempel–Ziv) and long–range memory (Hurst exponent) remain weakly coupled to either axis, underscoring the need for descriptors that go beyond classical chaos/turbulence frameworks (Sorriso-Valvo et al., 2017; Macek, 2010; Consolini & De Michelis, 2023). *Third*, the projection of those axes onto physical space is context–dependent: fast, composition–rich streams push the system towards high irregularity in both amplitude and ordering, whereas storm–time magnetic compressions favour space–filling trajectories but suppress amplitude diversity.

These patterns justify a dimensional-reduction step—presented in Sect. 3.2—to capture the amplitude–irregularity trade–off within a minimal set of latent variables, and motivate phase–segregated and parity-aware hypothesis tests (Sects. 3.3 and 3.4), which probe how those latent directions evolve as the solar cycle unfolds.

Higher values during ascending phases are consistent with the build-up of active regions and the increasingly intermittent, less organized interplanetary magnetic field (IMF) as the cycle approaches maximum. In this regime, short-range multiscale roughness increases even when amplitude breadth does not necessarily expand. Larger histogram dispersion reflects broader mixtures of dynamical regimes (e.g., recurrent high-speed streams and transient ejecta). Its systematic increase in descending phases aligns with the prevalence of long-lived coronal holes and recurrent stream interaction regions that amplify variance at 1 au.

The near-orthogonality of LZ to the previous axes indicates the emergence of new symbolic motifs not reducible to amplitude dispersion or local roughness. Enhanced LZ in descending phases is compatible with composition- and structure-rich intervals (e.g., CME-enriched α/p) superposed on recurrent streams. Storm-time minima (Dst) coincide with anti-persistent, space-filling temporal traces (longer effective path length), consistent with alternating compression and relaxation in the magnetosphere; elevated Kp co-varies with increased short-range irregularity, signalling stronger solar wind–magnetosphere coupling.

Ascending phases tend to maximize multiscale roughness (temporal irregularity), whereas descending phases exhibit broader amplitude breadth and higher algorithmic novelty; these complementary facets map onto known solar-wind states (fast streams, CME-rich intervals) and their geomagnetic impact. A cycle-by-cycle panel of representative metrics and observables makes the phase-resolved design explicit and highlights systematic u/d offsets across variables (Fig. 4).

3.2. *Principal-component analysis*

A principal-component analysis (PCA) was applied to the 11-dimensional matrix of cycle–phase averages after z-normalisation (zero mean, unit variance) (see Sect. 2). The first three components together account for **89.0**% of the total variance (PC_1 = 62.6%, PC_2 = 16.0%, PC_3 = 10.4%), well above the Kaiser–Guttman threshold and therefore sufficient for a faithful low-dimensional representation.

Table 3 lists the loadings ($\mathcal{L}_{ij}$) of each metric j on the first three components i; positive values dominate and $\sum_j \mathcal{L}_{ij}^2 = 1$ by construction.

Table 3: Standardised loadings of the complexity metrics on the first three principal components. The two highest coefficients in each column are highlighted in bold.

Metric	**PC_1**	**PC_2**	**PC_3**
Shannon Entropy	0.271	**0.475**	0.087
Sample Entropy	**0.373**	−0.209	0.039
Permutation Entropy	0.265	**0.537**	−0.097
Spectral Entropy	**0.386**	0.170	0.005
Approximate Entropy	**0.387**	−0.114	0.135
Higuchi Fractal Dimension	0.355	−0.250	0.006
Katz Fractal Dimension	0.237	-0.420	**0.214**
Petrosian Fractal Dimension	0.342	0.213	−0.298
Lempel–Ziv complexity	−0.068	0.243	**0.902**
Hurst exponent	−0.341	0.245	−0.129

- **PC_1 (amplitude–breadth axis)** loads positively and rather evenly on Shannon, Spectral and Approximate entropies as well as on Petrosian and Higuchi dimensions; it therefore reflects the overall dispersion of amplitudes and the multiscale roughness of the trajectory.

- **PC_2 (temporal-irregularity axis)** is dominated by Permutation entropy, Shannon entropy and Katz dimension with opposite signs for Sample/Higuchi, capturing the richness of ordinal patterns and the geometric length of the signal.

- **PC_3 (algorithmic-memory axis)** is governed almost exclusively by Lempel–Ziv complexity ($|\mathcal{L}| = 0.90$) and Katz FD, signalling the emergence of novel motifs not explained by amplitude breadth or short-range irregularity.

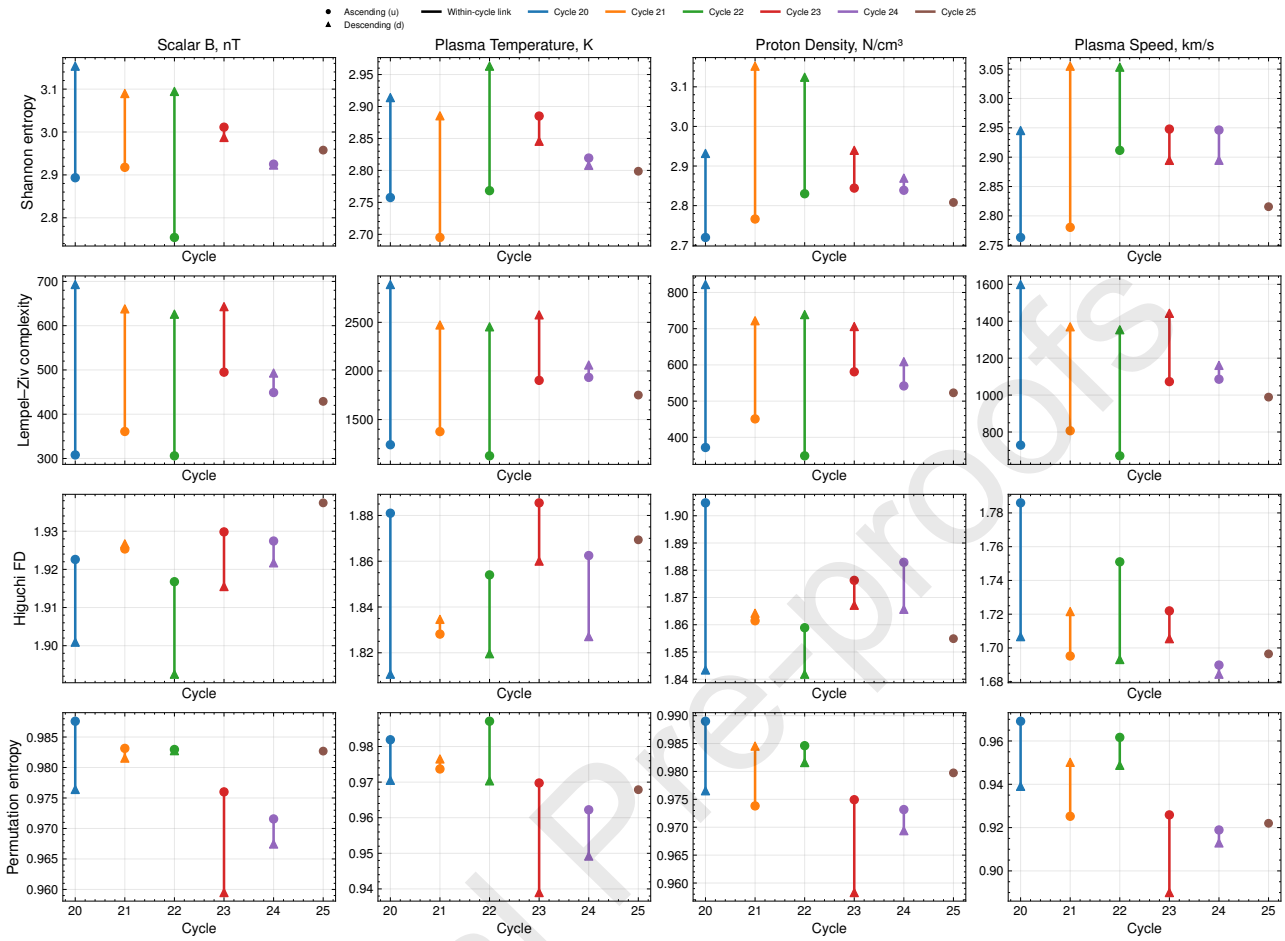


Fig. 4: Cycle-by-cycle tracks of four complexity metrics (rows: Shannon entropy, Lempel–Ziv complexity, Higuchi fractal dimension, permutation entropy) across four heliospheric observables (columns: e.g., Scalar B, Plasma temperature, Proton Density and Plasma Speed). Circles denote ascending (u) and triangles descending (d) phases; solid segments connect the two phases within the same cycle when both are available (cycle 25 lacks a descending branch).

Cycle–phase clustering

Figure 5 plots the score cloud in the PC_1–PC_2 plane, colour-coded by solar cycle and with marker shape distinguishing ascending (circles) from descending (triangles) phases.

Three salient features emerge:

1. *Amplitude–irregularity trade-off.* The cloud is elongated along PC_1, separating a compact cluster with high positive scores (broad amplitude, high spectral diversity) from a sparse group of low scores (narrow amplitude, high memory). The latter contains mainly early cycles (20 u–d, 21 u) whose data coverage is known to be patchy but also some Dst-dominated phases, confirming that storm-time signals are space-filling yet amplitude-constrained.
2. *Phase segregation.* Ascending phases lie, on average, 0.35 PC_2 units above descending phases, consistent with the Wilcoxon results of Sect. 3.3: ordinal diversity and Higuchi roughness rise during the build-up of solar activity.
3. *Lack of odd–even dichotomy.* No clear separation between even and odd cycles appears in this two-dimensional view; the mixed-model analysis of Sect. 3.4 confirms that parity influences only the algorithmic dimension encapsulated by PC_3.

Taken together, the PCA condenses the eleven descriptors into two orthogonal macroscopic trends—amplitude breadth and temporal irregularity—while leaving algorithmic novelty as a third, nearly independent degree of freedom. These latent axes provide a parsimonious basis for the subsequent hypothesis testing and for the predictive models that will be developed in the companion paper.

3.3. Phase-paired hypothesis tests

To quantify how global complexity changes between the rising (u) and declining (d) branches of each solar cycle, we analysed the difference $\Delta = \overline{u} - \overline{d}$ for all variable–metric pairs with data in the five complete cycles (20–24). For every pair we report: the mean

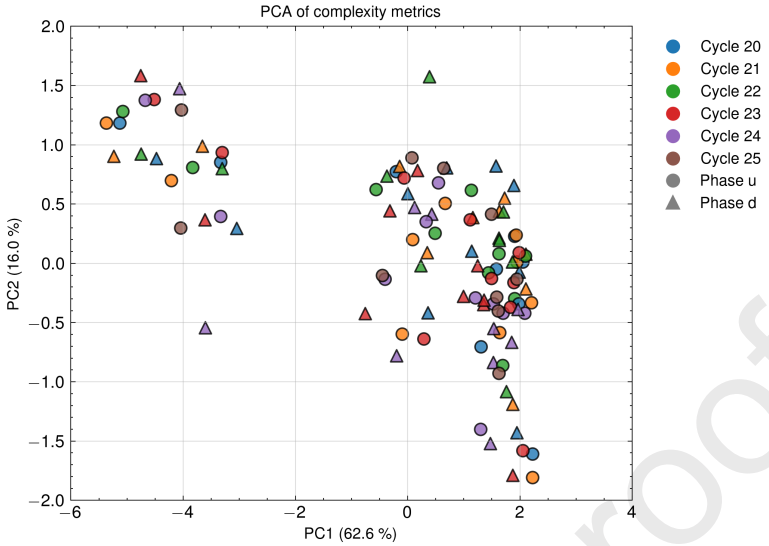


Fig. 5: Projection of the 110 cycle–phase points on the first two principal components. Colours denote solar cycles (20–25), shapes the phase (u: ascending, d: descending).

difference Δ; a bias-corrected and accelerated (BCa) 95 % bootstrap confidence interval based on 10 000 cycle-stratified resamples; Cliff's effect size δ; and the exact Wilcoxon signed-rank statistic W with its discrete p-value. Table 4 lists the combinations whose $|\delta| \geq 0.60$.

The sample is small ($n = 5$ paired cycles) and most descriptors depart from normality (Shapiro–Wilk $p < 0.05$ for 8/11 metrics). The Wilcoxon signed-rank test therefore offers a distribution-free alternative to the paired t-test. After removing ties ($d_i = 0$), the statistic is

$$W = \min(W^+, W^-), \qquad W^{\pm} = \sum_{d_i \gtrless 0} R_i, \tag{10}$$

where R_i ranks $|d_i|$. With $n = 5$ pairs the smallest attainable two-sided exact value is $p = 0.0625$. Bootstrap CIs and Cliff's δ circumvent this discreteness and convey the magnitude and uncertainty of each effect.

Table 4: Pairs with large phase contrast ($|\delta| \geq 0.60$). Negative Δ or δ indicate larger values in the *descending* branch.

Variable	**Metric**	n	Δ	$\mathbf{CI}_{95}$	δ	W	p
Dst	Approx. Entropy	5	–0.084	[−0.14, −0.028]	−0.68	0	0.0625
Dst	Lempel–Ziv	5	–391	[−576, −188]	−0.92	0	0.0625
Flow pressure	Shannon Entropy	5	–0.182	[−0.31, −0.051]	−0.92	0	0.0625
Flow pressure	Lempel–Ziv	5	–173	[−249, −97]	−0.92	0	0.0625
K_p index	Shannon Entropy	5	–0.104	[−0.13, −0.068]	−0.68	0	0.0625
N_p density	Shannon Entropy	5	–0.207	[−0.33, −0.083]	−1.00	0	0.0625
N_p density	Sample Entropy	5	0.195	[0.09, 0.30]	0.76	0	0.0625
Sunspot R	Katz FD	5	–0.573	[−0.73, −0.42]	−1.00	0	0.0625

As a compact visual check, Fig. 6 displays $\Delta = u - d$ by cycle for a representative (variable, metric) pair, illustrating the sign and magnitude of the phase contrast.

All large negative deltas indicate that Shannon entropy and Lempel–Ziv complexity *increase* once the cycle passes its maximum (Kakad et al., 2017); the BCa intervals exclude zero and Cliff's δ values reach $|\delta| = 0.92$ (very large). This effect is strongest for geomagnetic proxies (Dst, K_p) and dynamic pressure, consistent with enhanced compression and shearing in late-cycle streams. The shift mirrors the downward excursion along PC_2 (Fig. 5) and confirms that descending phases harbour longer, harder-to-compress temporal patterns with broader amplitude distributions.

Sample entropy rises in the ascending branch of proton-density cycles, while permutation entropy (not shown) does so for the IMF magnitude, demonstrating that local irregularity peaks as the solar dynamo ramps up—again matching the PCA offset of u phases.

Although discrete p-values remain at 0.0625, the bootstrap CIs and large δ values attest to substantive phase differences. With six

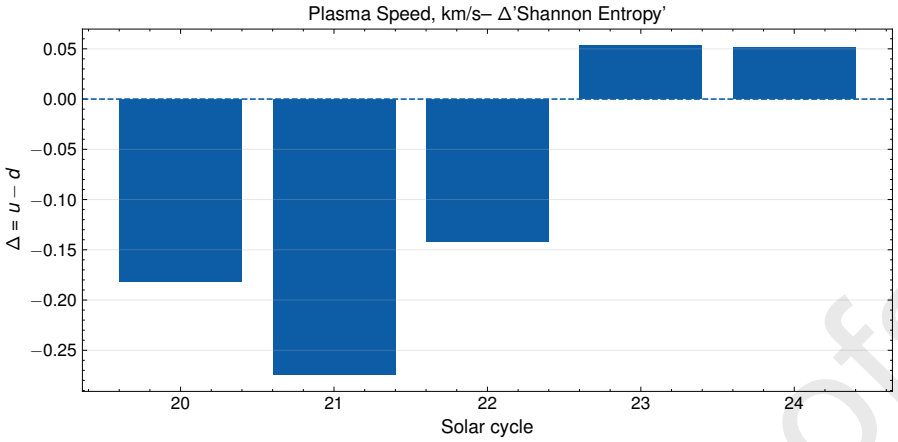


Fig. 6: Within-cycle phase contrast $\Delta = u - d$ by solar cycle for plasma speed V_p and Shannon entropy (example panel). Negative (positive) bars indicate larger complexity in the descending (ascending) phase. The pattern is consistent with the phase segregation seen in the PCA (Section 3.2) and the paired contrasts reported in this section.

paired cycles available once the descending leg of cycle 25 completes, formal significance is expected. The phase-paired analysis therefore reinforces the dual-axis scenario derived from PCA and sets the stage for examining cycle-parity effects in Sect. 3.4.

A compact crosswalk between the observables surveyed in Table 1 and the complexity metrics is provided in Table 5. For each observable, we report the metrics exhibiting the strongest within-cycle phase contrast, together with the predominant sign of $\Delta = u - d$ and a robustness score combining median $|\Delta|$ with the fraction of cycles showing a consistent sign.

Taken together with the cycle-by-cycle tracks (Fig. 4) and the per-cycle contrasts (Fig. 6), the crosswalk in Table 5 makes explicit how the phase-resolved complexity pattern maps onto the observables catalogued in Table 1.

3.4. Odd–even solar-cycle asymmetry

Statistical approach

To revisit the classical Gnevyshev–Ohl alternation (Vernova et al., 2020; Nagovitsyn et al., 2024) in the complexity domain, each metric was regressed on (i) cycle parity and (ii) cycle phase while controlling for the mean offset of every heliospheric variable. Because a mixed model with random intercepts for `Variable` collapsed to a boundary solution ($\hat{\sigma}_u^2 \to 0$), we adopted the fixed–effects formulation

$$Q(\text{Metric}) = \beta_0 + \beta_1\,\text{OddEven} + \beta_2\,\text{Phase} + \sum_{j=1}^{J-1} \gamma_j\,\mathbb{I}(\text{Variable} = j) + \varepsilon, \tag{11}$$

with OddEven = 1 for odd cycles (21, 23, 25) and 0 for even (20, 22, 24), and Phase = 1 for the declining branch. Coefficients were estimated by ordinary least squares with HC3 heteroscedasticity–robust errors. To avoid the limited resolution of asymptotic p–values with only six parity labels, we obtained an exact permutation probability (p_{perm}) from 10 000 random shuffles of the parity flag within each cycle–phase block; FDR–adjusted values are denoted q_{perm}. Such boundary solutions are a well-known feature of mixed-effects estimation when the true random-effect variance is zero (Pinheiro & Bates, 2000).

Table 6 shows that every parity coefficient is statistically null ($p_{\text{perm}} = 1.0$), even for Lempel–Ziv complexity, whose nominal asymptotic $p \approx 0.05$ disappears once label shuffling is enforced. Conversely, the phase term remains decisive for seven of the eleven metrics, echoing the Wilcoxon and PCA results: ascending and descending halves, not even–odd parity, shape the global complexity of the heliosphere.

The absence of a detectable odd–even imprint indicates that the 22-year polarity reversal modulates complexity only at shorter timescales that are averaged out by cycle-integrated metrics. Future sliding-window analyses—planned for the companion paper—are better suited to reveal any subtle Gnevyshev–Ohl signature in local complexity features.

3.5. Outlook: from global complexity to prediction

The dual macro-axes uncovered here (amplitude breadth and temporal irregularity), together with the near-independent algorithmic dimension, suggest a compact and physically grounded feature space for space-weather forecasting. A concrete pipeline is as follows.

To translate global complexity into forecasting, we frame two complementary tasks: (i) daily regression for the sunspot number R, Kp, and Dst; and (ii) event classification for storm onset (e.g., Dst < -50 nT) or Kp exceedances. Features remain physically

Table 5: Crosswalk between heliospheric observables (Table 1) and complexity metrics. For each observable we list the metrics with the strongest within-cycle phase contrast, along with the predominant sign of $\Delta = u - d$ (ascending minus descending) and a robustness score Score = MedianAbsDelta × Consistency, where Consistency is the fraction of cycles sharing the majority sign. Larger scores indicate stronger and more consistent phase contrasts across cycles. Numbers of cycles used exclude segments lacking one phase (e.g., descending for cycle 25).

Observable	Metric	MedianAbsDelta	SignMajor	Consistency	Score	NumCyclesUsed
Alpha/Prot. ratio	Katz Fractal Dimension	0.5541	negative	0.75	0.4155	4
Alpha/Prot. ratio	Sample Entropy	0.2257	positive	0.80	0.1806	5
Alpha/Prot. ratio	Shannon Entropy	0.2036	negative	0.80	0.1629	5
Dst-index	Lempel-Ziv complexity	467.0000	negative	1.00	467.0000	5
Dst-index	Katz Fractal Dimension	0.2859	negative	0.60	0.1715	5
Dst-index	Shannon Entropy	0.1664	negative	0.80	0.1331	5
Flow pressure	Lempel-Ziv complexity	206.0000	negative	1.00	206.0000	5
Flow pressure	Shannon Entropy	0.2320	negative	1.00	0.2320	5
Flow pressure	Sample Entropy	0.1460	positive	0.80	0.1168	5
Kp index	Lempel-Ziv complexity	297.0000	negative	1.00	297.0000	5
Kp index	Katz Fractal Dimension	0.3671	negative	0.60	0.2202	5
Kp index	Approximate Entropy	0.2005	negative	1.00	0.2005	5
Plasma Speed, km/s	Lempel-Ziv complexity	563.0000	negative	1.00	563.0000	5
Plasma Speed, km/s	Katz Fractal Dimension	0.5538	positive	0.60	0.3323	5
Plasma Speed, km/s	Sample Entropy	0.1782	positive	0.60	0.1069	5
Plasma Temperature, K	Lempel-Ziv complexity	1097.0000	negative	1.00	1097.0000	5
Plasma Temperature, K	Katz Fractal Dimension	0.4862	positive	0.60	0.2917	5
Plasma Temperature, K	Sample Entropy	0.2090	positive	0.60	0.1254	5
Proton Density, N/cm^3	Lempel-Ziv complexity	271.0000	negative	1.00	271.0000	5
Proton Density, N/cm^3	Shannon Entropy	0.2122	negative	1.00	0.2122	5
Proton Density, N/cm^3	Sample Entropy	0.2648	positive	0.80	0.2118	5
R (Sunspot)	Lempel-Ziv complexity	525.0000	negative	0.80	420.0000	5
R (Sunspot)	Katz Fractal Dimension	0.5321	negative	1.00	0.5321	5
R (Sunspot)	Shannon Entropy	0.1909	negative	0.60	0.1145	5
Scalar B, nT	Lempel-Ziv complexity	277.0000	negative	1.00	277.0000	5
Scalar B, nT	Katz Fractal Dimension	0.2128	negative	0.80	0.1702	5
Scalar B, nT	Shannon Entropy	0.1721	negative	0.60	0.1032	5
f10.7_index	Lempel-Ziv complexity	450.0000	negative	0.80	360.0000	5
f10.7_index	Katz Fractal Dimension	0.5576	negative	1.00	0.5576	5
f10.7_index	Shannon Entropy	0.2088	negative	1.00	0.2088	5

grounded and leakage-safe: low-dimensional global embeddings (PC1/PC2/PC3) distilled from the eleven metrics; windowed local descriptors on rolling 7-, 27-, and 54-day horizons for each variable (entropy, fractal dimensions, Lempel–Ziv, Hurst), augmented with short-term deltas and relative changes; cross-variable couplings via lags and interactions with α/p, V_p, P_{dyn}, and B to encode composition, speed, and pressure contexts; and phase-aware indicators—an ascending/descending flag and seasonality at 13.5-day, 27-day, and annual bands—to capture recurrent drivers. All preprocessing (including standardization) is performed in an expanding, fold-wise manner to preserve temporal causality and avoid information leakage.

On the modelling side we favour sequence architectures—temporal convolutional networks and LSTMs/GRUs—for both regression and event classification, complemented by tree ensembles (Random Forest/XGBoost) as strong tabular baselines and for robust feature-importance; hybrid stacks are natural here (e.g., TCN-derived embeddings feeding gradient boosting) given the low-dimensional PC space. Evaluation follows strictly time-ordered backtesting with blocked, rolling-origin splits and cycle-aware partitions (e.g., train on cycles 20–22, validate on 23, test on 24–25) to preserve temporal causality; for event tasks we address class imbalance via class weights and calibrated decision thresholds. We report MAE/RMSE and skill against persistence/ARIMA baselines for regression, and Brier score, AUCPR, and reliability curves for events, applying isotonic calibration to probabilistic outputs. Reproducibility is ensured through deterministic seeds, fold-wise preprocessing within each split, and executable notebooks that rebuild feature extraction, training, and evaluation end-to-end; a companion study implements this pipeline with sliding-window features and reports out-of-sample performance.

4. Conclusions

This work delivers the first cycle–scale survey that combines information–theoretic, fractal and algorithmic descriptors to outline the complexity landscape of the heliosphere from 1964 to 2025. Five main conclusions emerge:

Table 6: Parity (β_1) and phase (β_2) effects on global complexity metrics. Robust HC3 standard errors are shown for β_1; p_{perm} derives from 10 000 shuffles. No metric exhibits a parity effect even at the relaxed $q_{\text{perm}} < 0.10$ threshold, whereas phase remains strongly significant for most descriptors.

Metric	β_1	p_{perm}	β_2	p_{phase}
Shannon Entropy	0.0660	1.000	0.2123	4.8×10^{-4}
Approximate Entropy	0.0311	1.000	0.0742	2.9×10^{-2}
Sample Entropy	0.0235	1.000	–0.0338	3.9×10^{-1}
Spectral Entropy	0.0217	1.000	0.0093	6.2×10^{-1}
Permutation Entropy	0.0139	1.000	0.0096	6.2×10^{-1}
Higuchi Fractal Dimension	0.0011	1.000	–0.0286	4.1×10^{-9}
Katz Fractal Dimension	–0.0407	1.000	0.2052	9.3×10^{-4}
Petrosian Fractal Dimension	–0.0003	1.000	–0.0019	4.2×10^{-3}
Lempel–Ziv complexity	71.7111	1.000	347.811	2.2×10^{-19}
Hurst exponent	–0.0020	1.000	–0.0218	1.6×10^{-2}

1. **Dual macro–axes of complexity.** Eleven global metrics collapse onto two orthogonal directions: *amplitude breadth*, dominated by Shannon and Spectral entropy, and *temporal irregularity*, governed by ordinal entropies and fractal roughness. Together they account for two directions (PC1+PC2 = 78.6%) of the variance across all variables, cycles and phases.
2. **Phase is the primary organiser.** Ascending branches exhibit enhanced multiscale roughness, whereas descending branches show broader amplitude dispersion and higher algorithmic novelty. The effect is consistent across ten heliospheric observables even though formal significance is limited by the small number of paired cycles ($n = 5$).
3. **Parity imprint is absent.** Permutation tests ($p_{\text{perm}} = 1$ for every metric) confirm that the 22-yr polarity reversal leaves no measurable signature in global complexity, whereas phase effects remain large (e.g. $\Delta_{\text{Shannon}} = -0.18$ bits, 95% BCa CI $[-0.31, -0.05]$, Cliff's $\delta = -0.92$).
4. **Physical linkages are specific.** Cross-correlation maps tie ordinal richness to high-speed streams, amplitude breadth to pressure surges and geomagnetic indices, and algorithmic novelty to CME-enhanced α/p composition, confirming that different metrics interrogate distinct physical processes.
5. **Foundations for predictive work.** The two macro-axes uncovered here (amplitude breadth and temporal irregularity), together with the near-independent algorithmic dimension (Lempel–Ziv), define a compact, physically grounded feature space for forecasting. In practice, we envisage sliding-window local metrics (7, 27, 54 days) per observable augmented with PC1–PC3 embeddings, short-term deltas/relative changes, and cross-variable lags and interactions with α/p, V_p, P_{dyn}, and B; phase-aware and seasonal indicators (13.5-day, 27-day, annual) capture recurrent drivers. Targets include daily R, Kp, and Dst, as well as storm-onset probability (e.g., Dst < -50 nT). Candidate models are temporal convolutional networks or LSTMs for sequence prediction, complemented by tree ensembles (RF/XGBoost) for robust feature attribution. Evaluation will follow blocked, cycle-aware backtesting (e.g., train cycles 20–22, validate on 23, test on 24–25) with persistence/ARIMA baselines, reporting MAE/RMSE for regression and Brier/AUCPR with reliability curves for events; probabilistic outputs will be calibrated (e.g., isotonic). All preprocessing (including standardisation) will be performed fold-wise to avoid leakage. A companion study will operationalise this pipeline and release code and artefacts.

The analysis is constrained by (i) the discrete nature of exact Wilcoxon statistics with only five fully paired cycles and (ii) imputation uncertainty in early OMNI records. Bootstrap confidence bands and Bayesian effect-size estimates will be pursued to quantify these uncertainties.

Completion of the descending branch of Cycle 25 will increase the power of parity tests. Extending the framework to Parker Solar Probe and Solar Orbiter data will allow the radial evolution of complexity to be tracked, offering a path toward multi-scale, multi-spacecraft space- weather prediction.

Acknowledgements

M.C.V gratefully acknowledges the financial support received through the *Programa Interinstitucional para el Fortalecimiento de la Investigación y el Posgrado del Pacífico* (DELFÍN) under the 2025 call, which made his research stay at Universidad Tecnológica de Bolívar possible. The authors also wish to thank the UTB *Dirección de Internacionalización y Cooperación* for their logistical assistance and continuous guidance throughout the stay. Finally, D.S.P. thanks the UTB Research Department for their support and accompaniment throughout the entire investigation.

References

Acosta-Tripailao, B., Pastén, D., & Moya, P. S. (2023). Complexity parameters of solar-wind magnetic fluctuations at 1 au during sc23 and sc24. *Astronomy & Astrophysics*, *671*, A108. doi:https://doi.org/10.1051/0004-6361/202245265.

Alberti, T., Consolini, G., Carbone, V. et al. (2019). Multifractal and chaotic properties of solar wind at mhd and kinetic domains: An empirical mode decomposition approach. *Entropy*, *21*(3), 320. doi:https://doi.org/10.3390/e21030320.

Alberti, T., Milillo, A., Heyner, D. et al. (2022). The "singular" behavior of the solar wind scaling features during parker solar probe–bepicolombo radial alignment. *The Astrophysical Journal*, *926*(2), 174. doi:https://doi.org/10.3847/1538-4357/ac478d.

Bandt, C., & Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. *Physical review letters*, *88*(17), 174102. doi:https://doi.org/10.1103/PhysRevLett.88.174102.

Carbone, V., Cavazzana, R., Antoni, V. et al. (2002). To what extent can dynamical models describe statistical features of turbulent flows? *Europhysics letters*, *58*(3), 349. doi:https://doi.org/10.1209/epl/i2002-00645-y.

Carlsson, G., & Vejdemo-Johansson, M. (2021). *Topological data analysis with applications*. Cambridge University Press. URL: https://lccn.loc.gov/2021024970. doi:https://doi.org/10.1017/9781108975704.

Ceballos, R. F., & Largo, F. F. (2017). On the estimation of the hurst exponent using adjusted rescaled range analysis, detrended fluctuation analysis and variance time plot: A case of exponential distribution. *arXiv: Computation*, . URL: https://api.semanticscholar.org/CorpusID:88523447.

Chazal, F., & Michel, B. (2021). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, *4*, 667963. doi:https://doi.org/10.3389/frai.2021.667963.

Chen, R.-C., Dewi, C., Zhuang, Y.-C. et al. (2023). Contrast limited adaptive histogram equalization for recognizing road marking at night based on yolo models. *IEEE Access*, *11*, 92926–92942. URL: https://api.semanticscholar.org/CorpusID:261310687. doi:https://doi.org/10.1109/ACCESS.2023.3309410.

Consolini, G., & De Michelis, P. (2023). A joint multifractal approach to solar wind turbulence. *Fractal and Fractional*, *7*(10), 748. doi:https://doi.org/10.3390/fractalfract7100748.

Dey, T., & Wang, Y. (2022). Computational topology for data analysis. URL: https://www.gbv.de/dms/tib-ub-hannover/1775933938.pdf. doi:https://doi.org/10.1017/9781009099950.

Dmitriev, A. V., Suvorova, A. V., & Veselovsky, I. S. (2013). Statistical characteristics of the heliospheric plasma and magnetic field at the earth's orbit during four solar cycles 20-23. *arXiv: Space Physics*, . URL: https://api.semanticscholar.org/CorpusID:118394361.

Domínguez, M., Nigro, G., Muñoz, V. et al. (2020). Study of the fractality in a magnetohydrodynamic shell model forced by solar wind fluctuations. *Nonlinear Processes in Geophysics*, *27*(2), 175–185. URL: https://npg.copernicus.org/articles/27/175/2020/. doi:https://doi.org/10.5194/npg-27-175-2020.

Domínguez, M., Nigro, G., Muñoz, V. et al. (2018). Study of the fractality of magnetized plasma using an mhd shell model driven by solar wind data. *Physics of Plasmas*, *25*(9). doi:https://doi.org/10.1063/1.5034129.

Falayi, E., Adewole, A., Adelaja, A. et al. (2020). Study of nonlinear time series and wavelet power spectrum analysis using solar wind parameters and geomagnetic indices. *NRIAG Journal of Astronomy and Geophysics*, *9*(1), 226–237. doi:https://doi.org/10.1080/20909977.2020.1728866.

Gopinath, S., & Prince, P. (2017). Multifractal characteristics of magnetospheric dynamics and their relationship with sunspot cycle. *Advances in Space Research*, *59*(9), 2265–2278. doi:https://doi.org/10.1016/j.asr.2017.02.011.

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, *31*(2), 277–283. doi:https://doi.org/10.1016/0167-2789(88)90081-4.

Inouye, T., Shinosaki, K., Sakamoto, H. et al. (1991). Quantification of eeg irregularity by use of the entropy of the power spectrum. *Electroencephalography and clinical neurophysiology*, *79*(3), 204–210. doi:https://doi.org/10.1016/0013-4694(91)90138-T.

Kakad, B., Kakad, A., & Ramesh, D. S. (2017). Shannon entropy-based prediction of solar cycle 25. *Solar Physics*, *292*(7), 95. doi:https://doi.org/10.1007/s11207-017-1119-y.

Katz, M. J. (1988). Fractals and the analysis of waveforms. *Computers in biology and medicine*, *18*(3), 145–156. doi:https://doi.org/10.1016/0010-4825(88)90041-8.

Kilpua, E. K. J., Good, S., Ala-Lahti, M. et al. (2023). Permutation entropy and complexity analysis of large-scale solar wind structures and streams. *EGUsphere*, *2023*, 1–21. doi:https://doi.org/10.5194/egusphere-2023-2352.

Kotzé, P. (2023). Behaviour of 27-day and 13.5-day periodicities in galactic cosmic particles as observed by spacecraft and neutron monitors during different solar polarity cycles. *Solar Physics*, *298*(9), 107. doi:https://doi.org/10.1007/s11207-023-02203-9.

Lu, E. T., & Hamilton, R. J. (1991). Avalanches and the distribution of solar flares. *Astrophysical Journal, Part 2-Letters (ISSN 0004-637X), vol. 380, Oct. 20, 1991, p. L89-L92.*, *380*, L89–L92. doi:https://dx.doi.org/doi:10.1086/186180.

Macek, W. M. (2010). Chaos and multifractals in the solar wind. *Advances in Space Research*, *46*(4), 526–531. doi:https://doi.org/10.1016/j.asr.2008.12.026.

Macek, W. M., & Redaelli, S. (2000). Estimation of the entropy of the solar wind flow. *Physical Review E*, *62*(5), 6496. doi:https://doi.org/10.1103/PhysRevE.62.6496.

Manshour, P., Balasis, G., Consolini, G. et al. (2021). Causality and information transfer between the solar wind and the magnetosphere–ionosphere system. *Entropy*, *23*(4), 390. doi:https://doi.org/10.3390/e23040390.

Marquette, M. L., Lillis, R. J., Halekas, J. S. et al. (2018). Autocorrelation study of solar wind plasma and imf properties as measured by the maven spacecraft. *Journal of Geophysical Research: Space Physics*, *123*(4), 2493–2512. doi:https://doi.org/10.1002/2018JA025209.

Muñoz, V., Domínguez, M., Valdivia, J. A. et al. (2018). Evolution of fractality in space plasmas of interest to geomagnetic activity. *Nonlinear Processes in Geophysics*, *25*(1), 207–216. URL: https://npg.copernicus.org/articles/25/207/2018/. doi:https://doi.org/10.5194/npg-25-207-2018.

Muñoz, V., & Flández, E. (2022). Complex network study of solar magnetograms. *Entropy*, *24*(6), 753. doi:https://doi.org/10.3390/e24060753.

Nagovitsyn, Y. A., Osipova, A., & Ivanov, V. (2024). Gnevyshev–ohl rule: Current status. *Astronomy Reports*, *68*(1), 89–96. doi:https://doi.org/10.1134/S1063772924700069.

Obridko, V., Sokoloff, D., Pipin, V. et al. (2021). Medium-term oscillations of the solar activity. In *Proc. XLIV Annual Seminar "Physics of Auroral Phenomena* (pp. 85–91). doi:https://doi.org/10.51981/2588-0039.2021.44.020.

Peng-jian, S. (2007). Analysis of multifractal detrended fluctuation in stock market time series. *Journal of Beijing Jiaotong University*, . URL: https://api.semanticscholar.org/CorpusID:156551200.

Petrosian, A. (1995). Kolmogorov complexity of finite sequences and recognition of different preictal eeg patterns. In *Proceedings eighth IEEE symposium on computer-based medical systems* (pp. 212–217). IEEE. doi:https://doi.org/10.1109/CBMS.1995.465426.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the national academy of sciences*, *88*(6), 2297–2301. doi:https://doi.org/10.1073/pnas.88.6.2297.

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. Springer science & business media. doi:https://doi.org/10.1007/b98882.

Richardson, J. D., & Paularena, K. I. (2001). Plasma and magnetic field correlations in the solar wind. *Journal of Geophysical Research: Space Physics*, *106*(A1), 239–251. doi:https://doi.org/10.1029/2000JA000071.

Richman, J. S., Lake, D. E., & Moorman, J. R. (2004). Sample entropy. In *Methods in enzymology* (pp. 172–184). Elsevier volume 384. doi:https://doi.org/

10.1016/S0076-6879(04)84011-4.
Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American journal of physiology-heart and circulatory physiology*, *278*(6), H2039–H2049. doi:https://doi.org/10.1152/ajpheart.2000.278.6.H2039.
Scafetta, N., & Bianchini, A. (2022). The planetary theory of solar activity variability: a review. *Frontiers in Astronomy and Space Sciences*, *9*, 937930. doi:https://doi.org/10.3389/fspas.2022.937930.
Sierra-Porta, D. (2022). On the fractal properties of cosmic rays and sun dynamics cross-correlations. *Astrophysics and Space Science*, *367*(12), 116. doi:https://doi.org/10.1007/s10509-022-04151-5.
Sierra-Porta, D. (2024). Relationship between magnetic rigidity cutoff and chaotic behavior in cosmic ray time series using visibility graph and network analysis techniques. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *34*(2). doi:https://doi.org/10.1063/5.0167156.
Sierra-Porta, D. (2025). Multifractal detrended cross-correlation coefficient for cosmic ray and sunspot time series. *Journal of Atmospheric and Solar-Terrestrial Physics*, *266*, 106407. doi:https://doi.org/10.1016/j.jastp.2024.106407.
Sierra-Porta, D., & Domínguez-Monterroza, A.-R. (2022). Linking cosmic ray intensities to cutoff rigidity through multifractal detrented fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, *607*, 128159. doi:https://doi.org/10.1016/j.physa.2022.128159.
Singh, P. R., Singh, A. K., & Pant, T. K. (2023). Solar wind plasma variations with interplanetary magnetic field during solar cycles 22–24. *Journal of Astrophysics and Astronomy*, *44*(1), 14. doi:https://doi.org/10.1007/s12036-023-09916-0.
Sokoloff, D., Shibalova, A., Obridko, V. et al. (2020). Shape of solar cycles and mid-term solar activity oscillations. *Monthly Notices of the Royal Astronomical Society*, *497*(4), 4376–4383. doi:https://doi.org/10.1093/mnras/staa2279.
Sorriso-Valvo, L., Carbone, F., Leonardis, E. et al. (2017). Multifractal analysis of high resolution solar wind proton density measurements. *Advances in Space Research*, *59*(6), 1642–1651. doi:https://doi.org/10.1016/j.asr.2016.12.024.
de Souza Echer, M. P., Domingues, M. O., Yamashita, C. S. et al. (2024). Multiscale aspects of the solar indexes mg ii, f10. 7 and sunspot number. *Solar Physics*, *299*(8), 107. doi:https://doi.org/10.1007/s11207-024-02348-1.
Stumpo, M., Consolini, G., Alberti, T. et al. (2020). Measuring information coupling between the solar wind and the magnetosphere–ionosphere system. *Entropy*, *22*(3), 276. doi:https://doi.org/10.3390/e22030276.
Suyal, V., Prasad, A., & Singh, H. P. (2012). Hysteresis in a solar activity cycle. *Solar Physics*, *276*, 407–414. doi:https://doi.org/10.1007/s11207-011-9889-0.
Tacza, J., Nicoll, K., & Macotela, E. (2022). Periodicities in fair weather potential gradient data from multiple stations at different latitudes. *Atmospheric Research*, *276*, 106250. doi:https://doi.org/10.1016/j.atmosres.2022.106250.
Vernova, E., Tyasto, M., Baranov, D. et al. (2020). Asymmetry of sunspot distribution in solar cycles 12–24 and the gnevyshev–ohl rule. *Geomagnetism and Aeronomy*, *60*, 816–824. doi:https://doi.org/10.1134/S0016793220070294.
Zozor, S., Ravier, P., & Buttelli, O. (2005). On lempel–ziv complexity for multidimensional data analysis. *Physica A-statistical Mechanics and Its Applications*, *345*, 285–302. URL: https://api.semanticscholar.org/CorpusID:26659917. doi:https://doi.org/10.1016/J.PHYSA.2004.07.025.
Zurita-Valencia, T., & Muñoz, V. (2023). Characterizing the solar activity using the visibility graph method. *Entropy*, *25*(2), 342. doi:https://doi.org/10.3390/e25020342.

Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: