Available online at www.sciencedirect.com

ScienceDirect

Advances in Space Research xx (2024) xxx-xxx

ADVANCES IN SPACE RESEARCH (a COSPAR publication)

www.elsevier.com/locate/asr

Revised Cross-Correlation and Time-Lag between Cosmic Ray Intensity and Solar Activity Using Chatterjee's Correlation Coefficient

D. Sierra-Porta[a,*]

[a] *Universidad Tecnológica de Bolívar, UTB. Facultad de Ciencias Básicas, Parque Industrial y Tecnológico Carlos Vélez Pombo Km 1 Vía Turbaco, Cartagena de Indias, 130010, Bolívar, Colombia*

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

Abstract

This study revisits the cross-correlation between cosmic ray intensity (CRI) and solar activity (SA) by comparing traditional Pearson correlation with Chatterjee's correlation coefficient. Traditional analyses using Pearson correlation are useful for identifying linear relationships and time lags. However, they may not fully capture more complex interactions in the data. Chatterjee's correlation coefficient, while sensitive to different types of relationships, including nonlinear ones, provides a complementary perspective on the temporal relationships between CRI and SA. This approach broadens our understanding of potential dependencies, offering additional insights that may not be captured through Pearson correlation alone.

The findings reveal that Chatterjee's correlation complements Pearson's insights by providing an alternative view of the relationship between cosmic ray intensity (CRI) and solar activity (SA). The results show that Chatterjee's correlation coefficients are, on average, approximately 45-50% smaller than Pearson's, which could reflect different sensitivities to the underlying data structure rather than solely indicating a nonlinear component. Additionally, the time lags identified using Chatterjee's correlation are generally shorter and more consistent across different solar cycles compared to those obtained with Pearson's correlation, suggesting that CCC may capture temporal patterns in a distinct manner.

Further analysis using Dynamic Time Warping (DTW) and Mean Absolute Percentage Error (MAPE) metrics demonstrated that, in more than half of the scenarios considered, alignment based on Chatterjee's time lags resulted in lower errors and better alignment of the series compared to Pearson's lags. This indicates that Chatterjee's method is particularly effective for capturing the immediate and nuanced responses of CRI to SA changes, especially in recent solar cycles.

This comprehensive approach provides broader insights into the dynamic interactions between cosmic ray intensity (CRI) and solar activity (SA), highlighting the importance of considering multiple correlation measures, including both linear and nonlinear approaches, in space weather research. The results suggest that Chatterjee's correlation offers a complementary perspective on these interactions, providing additional details about how SA influences CRI over time, which may not be fully captured by Pearson's correlation alone.

Cosmic Rays ; Solar Activity ; Cross-Correlation ; Chatterjee's correlation ; Pearson correlation ; Space Weather

*Corresponding author: orcid=0000-0003-3461-1347;
Email address: dporta@utb.edu.co (D. Sierra-Porta)

https://dx.doi.org/10.1016/j.jasr.xxxx.xx.xxx
0273-1177/

1. Introduction

The study of the relationship between cosmic ray intensity (CRI) and Solar Activity (SA), particularly using sunspot numbers (SN), has been extensively researched. Several studies have established a significant correlation between CRI and SA, frequently employing cross-correlation analysis to identify the time lags between these time series. These investigations consistently demonstrate that SA influences CRI, with well-documented time lags reflecting the delayed response of cosmic rays to changes in SA. Traditional analyses, primarily using Pearson correlation, have provided valuable insights but often fall short in capturing the full complexity of these interactions due to their linear nature and sensitivity to outliers.

Understanding the correlation between CRI and SA is crucial in space science and space weather forecasting. This relationship affects our comprehension of the Earth's magnetosphere and ionosphere, influencing satellite operations, communication systems, and power grids systems in the Earth (Dorman (2021); Höeffgen et al. (2020); Zheng et al. (2024); Dobynde et al. (2023)). Accurate identification of these correlations helps in predicting space weather events, mitigating their impact, and protecting technological infrastructure.

In their study, Sierra-Porta (2018) examined the covariability between daily CRI and three parameters: SN, flare index, and the Ap index, over three solar cycles (1976–2008). While the Ap index is a geomagnetic activity index and not directly of solar origin, it still reflects the impact of solar wind and interplanetary magnetic field variations on Earth's magnetic field. Using cross-correlation analysis, they found significant negative correlations with lags of 181 days for sunspots, 156 days for the flare index, and 2 days for the Ap index. These results shows the strong influence of solar activity and related geomagnetic disturbances on CRI, with daily resolution data adding robustness to the findings.

In another contribution, Iskra et al. (2019) analyzed long-term changes in galactic cosmic ray (GCR) intensity across different solar magnetic cycles from 1959 to 2014. They focused on the delay time (DT) between GCR intensity changes and various heliospheric parameters. The study reveals varying DTs between CRI and SA parameters like sunspot number and tilt angle, depending on the polarity of the solar magnetic field. These DTs are crucial for understanding GCR transport in the heliosphere. They observed that SA, characterized by SN, significantly impacts heliospheric conditions, including the interplanetary magnetic field (IMF), its turbulence, and other solar phenomena. They found that DTs vary with the 22-year solar cycle and are influenced by the structure of IMF turbulence and the drift of GCRs. The study shows the importance of modeling and experimentally analyzing GCR propagation to understand DTs across different solar cycles.

Finally, Koldobskiy et al. (2022) analyzed the time lag between cosmic-ray intensity and solar variability, focusing on SN and open solar flux (OSF) from 1959 to 2014. They employed cross-correlation and wavelet coherence methods to identify time lags. The study found that GCR intensity lags behind SSN by about eight 27-day Bartels rotations (BRs) on average, with the delay varying over the 22-year solar cycle. The OSF lags behind SSN by about one year, showing a clear 22-year cycle in the delay. GCRs and OSF also exhibit a strong 22-year cycle, confirming that OSF is a good proxy for heliospheric modulation of GCRs (Usoskin et al. (2004); Wu et al. (2018); Usoskin et al. (2021)). The study found high coherence between the indices over a few years, with GCR variability lagging SSN by about eight 27-day Bartels rotations on average, varying with the 22-year cycle. The OSF lags behind SSN by about one year, influenced by the lifetimes of active regions and the transport of the surface magnetic field. These findings underscore the importance of understanding these delays for accurately modeling the transport of galactic cosmic rays (GCR) in the heliosphere. The results are consistent with theoretical models of heliospheric modulation and offer significant observational constraints for cosmic ray studies.

However, Pearson correlation coefficient (PCC), while useful and employed in multiple research studies (Maghrabi et al. (2021); Mishra & Mishra (2018) and references therein), has notable limitations. Pearson correlation primarily measures linear relationships and can be significantly affected by outliers (Sullivan et al. (2021)). Additionally, it assumes a linear dependency structure (Schober et al. (2018)), which might not capture the complexity of the relationships between variables like CRI and SA, known to exhibit interactions and temporal lags influenced by various solar and heliospheric processes (Koldobskiy et al. (2022)).

The observed time lag between solar activity and cosmic-ray intensity can be explained by the underlying physical processes that govern their interaction (Potgieter (2013)). While sunspots themselves do not directly influence cosmic-ray intensity, the magnetic disturbances produced by large, long-lived active regions on the Sun can significantly impact the properties of the solar wind (Stansby et al. (2021)). These changes in the solar wind affect how galactic cosmic rays (GCRs) travel through the heliosphere. The delay in cosmic-ray variations near Earth, relative to solar activity, is mainly due to the outward movement of solar wind and heliospheric magnetic field (HMF) fluctuations, along with the inward diffusion of GCR particles, which is influenced by large-scale drift processes (Jokipii & Levy (1977); Gervasi et al. (1999); Boschini et al. (2018)).

The assumptions underlying Pearson's correlation have been widely discussed and debated (Binder (1959); Ventura-León et al. (2023); Puth et al. (2014); Bishara & Hittner (2012); Knief & Forstmeier (2021)). It's not surprising, though it can be confusing, that various statistical sources offer differing interpretations of these assumptions. In practice, the coefficient can be used to quantify the strength of a linear relationship between two variables without requiring specific assumptions.

For the calculation of Pearson's correlation, the data do not need to be normally distributed (Schober et al. (2018); Kowalski (1972); Binder (1959); Bland & Altman (1994)); however, normality is an important assumption when making statistical inferences, such as constructing confidence intervals or generalizing results to the population. In many real-world datasets, this assumption is often violated, which can affect the reliability of inferences based on the correlation coefficient.

This study seeks to enhance our understanding of the temporal dynamics between cosmic ray intensity (CRI) and solar activity (SA) by incorporating Chatterjee's correlation coefficient (CCC) (Chatterjee (2021)), a statistical method that complements traditional approaches. Chatterjee's correlation does not assume linearity and is less sensitive to outliers, offering insights into a broader range of dependencies. By comparing it with Pearson's correlation, we aim to explore how CCC can provide additional perspectives and contribute to our overall interpretation of the interactions between CRI and SA. Previous studies have shows the utility of Chatterjee's correlation in various contexts, making it a valuable tool for this investigation (Dong et al. (2023); Shi et al. (2022); Sadeghi (2022); Lin & Han (2023)).

By revisiting the cross-correlation analysis with this advanced statistical tool, we aim to uncover deeper insights into the interaction between these variables. This approach not only validates previous findings but also enhances our understanding of the complex relationships that may exist. The application of Chatterjee's correlation coefficient is expected to reveal additional dependencies that traditional methods might overlook, offering a more detailed and accurate depiction of the cosmic ray and SA interplay.

The primary objectives of this study are to: (a) Evaluate the performance of CCC in capturing different types of relationships between cosmic ray intensity (CRI) and solar activity (SA); (b) Compare the findings with those obtained using Pearson's correlation (PCC) to assess the benefits and limitations of each method; and (c) Provide new insights into the temporal dynamics and interactions between CRI and SA.

2. Sources of datasets in this study

2.1. Sources

In this study, we utilize neutron monitor count data from the Neutron Monitor Data Database (NMDB) across various global stations. The stations included are JUNG, NEWK, KERG and OULU. The dataset spans from October 1, 1964, to May 31, 2024. The data is initially provided at an hourly resolution, which is validated and subsequently averaged to daily values. The data type used is corrected for efficiency, and timestamps correspond to the beginning of each time interval.

NMDB (`https://www.nmdb.eu/`) serves as a comprehensive source for cosmic ray data, providing access to real-time and historical Neutron Monitor measurements from stations worldwide through an easy-to-use interface. The neutron monitor stations used in this study are located across various geographic regions, providing a broad range of data. The details of the neutron monitor stations used in this study are summarized in Table 1. This table includes the location, latitude (Lat, N), longitude (Lon, E), altitude (Alt, in m.a.s.l), and cutoff rigidity (R_c in GV) for each station, along with their managing institutions.

Although there are more than 40 neutron monitor stations around the world at different locations, latitudes, longitudes and with different magnetic rigidity cutoff conditions, the latter have been selected due to the criterion that they are the stations (according to the data that can be downloaded) that have the minimum number of missing data throughout the dataset. These stations represent a balance between good data quality and less missing data. Some stations around the world have undergone some updating or upgrading processes that usually leave significant jumps in measured intensities and counts, affecting the temporal continuity of the data. The above criteria finally include having reliable data for which a minimum of data mining and processing has to be done without structurally changing the data.

Additionally, we incorporate SN data from the World Data Center SILSO (`https://www.sidc.be/SILSO/datafiles`), Royal Observatory of Belgium, Brussels SILSO World Data Center (1964-2024). This data, revised and presented in version 2.0 since July 1, 2015, is available at an hourly resolution, aligning with the neutron monitor data's minimum resolution. Proper credit for using this data is required, as specified by WDC-SILSO.

2.2. Data processing

Our analysis covers solar cycles 20 through 25, with cycle 25 still ongoing (See Table 2). To mitigate short-term transient effects, all datasets were averaged. For the initial cross-correlation analysis, data gaps were left unfilled (marked as NaN) to prevent introducing any biases. However, for subsequent analyses requiring complete datasets, we employed random forest imputation (Breiman (2001)), a machine learning strategy known for its robustness in handling complex data structures. This method allows for more accurate and reliable completion of missing values. The imputation results demonstrated that the filled data maintained the overall trends and patterns of the original dataset, ensuring the integrity of further analyses. Days with solar energetic-particle events registered by neutron monitors (ground level enhancements, GLEs) Usoskin et al. (2020) were removed from both the neutron monitor and sunspot data.

Additionally, we standardized the neutron monitor counts and SN using min-max scaling to transform all values to a [0,1] scale. This was done using the formula:

$$N_{scaled} = \frac{N_i - N_{min}}{N_{max} - N_{min}}, \tag{1}$$

where $N = [N_i]$ represents each time series with i the temporal stamp. This preprocessing step ensures consistency and comparability across the different datasets used in the analysis.

After selecting the four neutron monitor stations mentioned and performing the scaling, we constructed a new CRI variable by taking the average value of all the monitors. This approach aims to create a composite CRI variable that incorporates information from all stations, providing a more robust estimate of cosmic ray intensity.

The newly created CRI variable maintains the same distribution characteristics as the individual station variables, ensuring that the statistical properties are preserved. However, by averaging the values from multiple monitors, this estimator integrates the information from all stations, potentially reducing

Table 1. Details of Neutron Monitor Stations used in this study.

Monitor	Location	Lat(N)	Lon(E)	Alt(masl)	R_c	Managing by	Detector
JUNG	Switzerland	46.55	7.98	3570	4.5	Physikalisches Institut University of Bern	Standard 18-IGY
NEWK	USA	39.68	-75.75	50	2.45	Bartol Research Institute University of Delaware	6 BF_3 detectors and 3 ^{3}He detectors
KERG	Kerguelen Indian Ocean	-49.35	70.25	33	1.14	French Polar Institute (IPEV, Brest) Paris Observatory, FRANCE	18NM64
OULU	Finland	65.05	25.47	15	0.81	Sodankyla Geophysical Observatory University of Oulu	9-NM-64 (Chalk River tubes)

Table 2. Periods of Solar Cycles used in this study.

SC	Start (Y-M)	Min SSN	Max (Y-M)	Max SSN	Length (Y-M)
20	1964-10	14.3	1968-11	157	11-5
21	1976-03	17.8	1979-12	233	10-6
22	1986-09	13.5	1989-11	213	9-11
23	1996-08	11.2	2001-11	180	12-4
24	2008-12	2.2	2014-04	116	11-0
25	2019-12	1.8	2023-12	Progr: 129	

measurement noise and increasing the reliability of the CRI estimates.

From a statistical perspective, this composite variable offers several advantages. Averaging across multiple stations can mitigate the effects of local anomalies or errors specific to individual monitors, resulting in a more accurate and stable representation of CRI. This method leverages the central limit theorem, where the mean of several independent and identically distributed random variables tends to follow a normal distribution, even if the original variables themselves are not normally distributed.

Moreover, this approach enhances the signal-to-noise ratio by averaging out random fluctuations and emphasizing the underlying trend common to all stations (Sierra-Porta & Domínguez-Monterroza (2022)). This composite CRI variable can therefore provide a more consistent and reliable basis for subsequent analyses, such as examining correlations with solar activity or geophysical phenomena.

3. Materials and Methods

To explore the relationships between neutron monitor counts and SN, we employed Chatterjee's correlation coefficient, a robust method for measuring dependence between two variables. Chatterjee's correlation coefficient (Chatterjee (2021)), denoted usually as ξ, is defined as following. Consider a pair of variables R and S:

$$\mathrm{CCC}_1 := \xi_n(R,S) = 1 - \frac{3\sum_{i=1}^{n-1}|k_{i+1}-k_i|}{n^2-1}, \tag{2}$$

where k_i is the ranks of the i-th observation in the respective S variable, when assuming that the R_i's and the S_i's have no ties, the data is rearranged as $(R_{(1)}, S_{(1)}), \cdots, (R_{(n)}, S_{(n)})$ such that $R_{(1)} \leq \cdots \leq R_{(n)}$ and $S_{(1)} \leq \cdots \leq S_{(n)}$. This unique ranking is possible due to the absence of ties in R_i and S_i.

In rank-based statistical methods, the term 'ties' refers to instances where two or more observations share the same value, resulting in the same rank. The presence of ties can complicate the calculation of rank-based metrics, such as correlation coefficients, as special adjustments are often needed to handle these cases. In contrast, when a dataset has no ties, meaning all values are unique and no two observations share the same rank, the calculations are simplified, and the interpretation of the results becomes more straightforward. The absence of ties eliminates the need for correction factors that account for tied ranks, making the analysis cleaner and more direct.

When dealing with ties, ξ_n is defined in the following manner. If the X_i's have ties, rearrange them in ascending order, breaking the ties randomly and uniformly. Let k_i remain as previously defined, and introduce l_i as the count of j where $Y(j) \geq Y(i)$. Then, define it as follows:

$$\mathrm{CCC}_2 := \xi_n(X,Y) = 1 - \frac{n\sum_{i=1}^{n-1}|k_{i+1}-k_i|}{2\sum_{i=1}^{n} l_i(n-l_i)}. \tag{3}$$

When there are no ties among the S_i's, $l_1, \ldots, l_n$ becomes a permutation of $1, \ldots, n$, thus the denominator in the above expression equals $n(n^2-1)/3$, reducing this definition to the earlier expression in eq. (2).

There are significant differences when compared to earlier correlation measures. CCC ranges from 0 to 1. It becomes zero only if both variables are independent and reaches one if one variable is a function of the other. Unlike the another coefficients, it is not symmetric in R and S, indicating that $CCC(R,S) \neq CCC(S,R)$. This asymmetry is logical since the focus is on whether R is a function of S, without implying the reverse relationship. Additionally, the author introduces an asymptotic theory for computing p-values, though some researchers have questioned the power of the coefficient. This coefficient has several advantages over traditional correlation methods such as Pearson's or Spearman's. In this study we use the version of equation (2) for the calculations hereafter.

To apply CCC to our time series data, we used the xicorrelation package (`https://pypi.org/project/xicorrelation/`), which integrates the original `XICOR` library written by the method's author (Chatterjee (2021)). This package is a `Python` adaptation of the `R` implementation detailed in the original code (`https://statweb.stanford.edu/~souravc/xi.R`) and the corresponding `R` package (`https://`

github.com/cran/XICOR). Additionally, our approach employed vectorization and parallel computing techniques to enhance efficiency, particularly for large-scale computations. We customized a slice window routine within the algorithm to calculate the lag time, sliding the time series relative to each other and obtaining CCC values at various lags.

A notable modification in the experimental design of this study involves the inclusion of a procedure to evaluate the correlation and time lag between cosmic ray series and sunspot numbers, with an additional step of smoothing the time series. The smoothing was performed using the exponentially weighted moving average (EWMA) method (Lucas & Saccucci (1990); Hunter (1986)). EWMA is a technique that applies exponentially decreasing weights to past observations, giving more importance to recent data points. The smoothing parameter, alpha (α), determines the rate at which these weights decrease. Mathematically, the EWMA at time t, $S(t) = S_t$, is given by:

$$S_t = \alpha Y_t + (1 - \alpha) S_{t-1}, \tag{4}$$

where Y_t is the value of the time series at time t and S_{t-1} is the EWMA at the previous time point. The parameter α ranges between 0 and 1, where a smaller α gives more weight to older observations, resulting in smoother time series, while a larger α emphasizes recent observations. For this study, various alpha factors were applied: α=[0.005, 0.01, 0.05, 0.1, 0.5, 0.75, 0.9, 1]. When $\alpha = 1$, the exponentially weighted moving average (EWMA) does not perform any smoothing, and the smoothed series S_t simply equals the original series Y_t. This means that each point in the time series retains its original value without any influence from previous points.

The figure 1 illustrates the cosmic ray intensity time series smoothed using different Exponentially Weighted Moving Average α coefficients to demonstrate how the results vary depending on the smoothing parameter. In the first plot, covering solar cycles 20 to 25, the series without smoothing is compared with series smoothed using $\alpha = 0.01$ and $\alpha = 0.25$. It is evident that higher smoothing levels (larger α values) attenuate high-frequency fluctuations. The second plot focuses on solar cycle 24 and shows a similar trend: the data smoothed with $\alpha = 0.01$ exhibit greater reduction in short-term fluctuations compared to $\alpha = 0.25$. These visualizations underscore the importance of selecting an appropriate smoothing parameter to capture both long-term trends and short-term variations in CRI time series.

The exponentially weighted moving average (EWM) technique introduces a form of smoothing that, while highly responsive to recent data points, does not introduce a true time lag in the traditional sense. The EWM algorithm applies a decay factor to prior observations based on a smoothing parameter α, allowing recent values to influence the smoothed result more heavily than earlier values. However, this selective weighting can create the impression of a lag (see Fig. 1), particularly when the series experiences rapid fluctuations. In such cases, the EWM may appear to "trail" the original series, although it remains aligned temporally and simply reflects a dampened response to abrupt changes. This feature of the EWM is valuable for filtering high-frequency noise without the forward displacement associated with fixed-window moving averages.

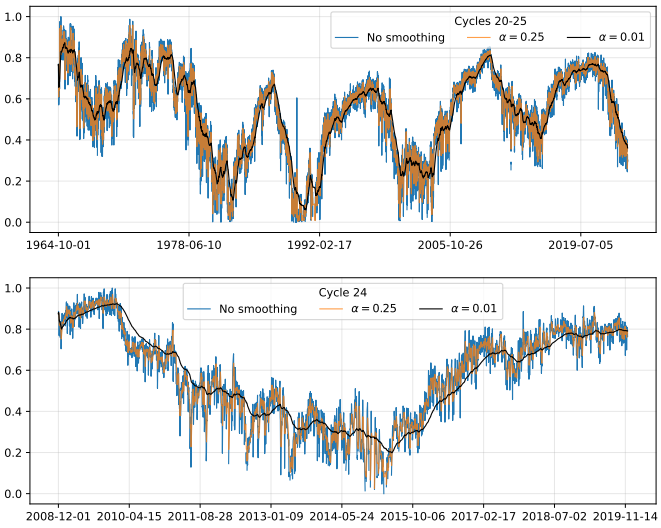

Fig. 1. Comparison of cosmic ray intensity series smoothed using different EWMA alpha coefficients. The top plot shows the CRI time series for solar cycles 20-25, comparing no smoothing (blue line), $\alpha = 0.25$ (orange line), and $\alpha = 0.01$ (black line). The bottom plot focuses on cycle 24 with the same comparison between smoothing levels. As the smoothing factor increases, high-frequency variations are attenuated, providing a clearer view of long-term trends.

This approach allows for a more nuanced analysis by reducing noise and capturing underlying trends in the data, thereby potentially enhancing the accuracy and interpretability of the correlation and time lag assessments between the two variables. By examining the effect of different smoothing levels, the study aims to identify the most effective smoothing parameter for accurately depicting the relationship between cosmic ray intensity and sunspot activity.

4. Results and discussions

4.1. Correlation coefficients

To begin our analysis, we first examine the distributions of cosmic ray counts and sunspot numbers. Figure 2 presents two histograms: the upper histogram shows the probability density distribution of CRI counts and SN for solar cycle 24 (black curve) and for all cycles considered (stepfilled curve), while the lower histogram illustrates the cumulative distribution for both variables. The distributions reveal that neither cosmic ray counts nor SN follow a normal distribution; instead, this seems closer to a rather asymmetric distribution, in the case of cosmic rays negative asymmetric ('Burr distribution' or 'Skewnorm distribution') and viceversa (positive asymmetric) for sunspots ('Asymmetric Laplace distribution' or 'Exponnorm distribution'). This observation is crucial because it suggests that traditional correlation and cross-correlation techniques, such as Pearson, Spearman, or Kendall coefficients, may not be appropriate for capturing the nonlinear characteristics of the relationships between these variables.

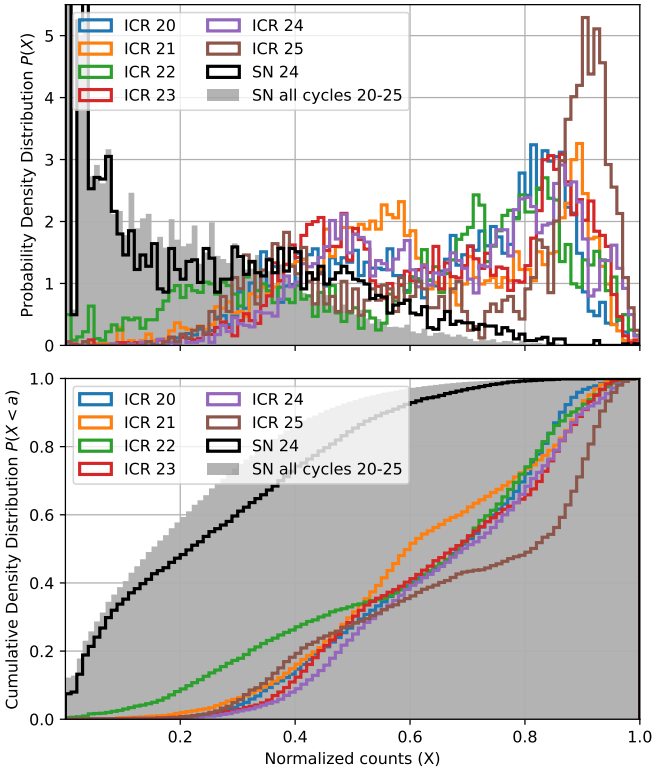

Fig. 2. Histograms showing the distribution of cosmic ray counts (for Oulu station) and SN. The upper panel displays the probability density distribution, while the lower panel shows the cumulative distribution. The cosmic ray data spans solar cycles 20 to 24, with the SN corresponding to cycle 24 and all years considered.

Figure 2 clearly shows the counts of cosmic rays over several solar cycles (20 to 24) and the SN for cycle 24 as a reference. The distributions of cosmic ray counts and SN are not normally distributed but rather to the irregular nature of space weather phenomena (Laken & Čalogović (2013)). CRI is influenced by various solar activities, particularly by coronal mass ejections (CMEs), which cause disturbances in the heliospheric magnetic field and modulate the propagation of galactic cosmic rays (Idosa et al. (2023); Mishra et al. (2006)). While solar flares can release energetic particles, they do not directly affect galactic cosmic ray intensity. Sunspots (SN), on the other hand, are indicators of the magnetic activity that can lead to solar flares and CMEs, rather than being influenced by these events.. This randomness results in distributions that are less uniform, reflecting the fluctuating and unpredictable characteristics of space weather events. Additionally, the 11-year solar cycle, characterized by alternating periods of high and low SA, plays a significant role in modulating CRI and SN. The fluctuations in solar activity, including the frequency and intensity of solar events, influence the variability of CRI and SN throughout the solar cycle. These variations are particularly notable during transitions between solar maximum and minimum, where changes in the heliospheric environment can affect cosmic ray propagation and sunspot dynamics.

The use of alternative methods, such as CCC, can provide valuable insights into different types of dependencies, as various correlation coefficients may yield distinct results. The analysis provided by these histograms lays the groundwork for further exploring the interactions between cosmic ray intensity (CRI) and solar activity (SA), highlighting how the choice of correlation method can influence the interpretation of relationships in the data. This underscores the importance of considering complementary approaches to standard correlation techniques when analyzing complex, real-world datasets.

Inspired by these observations, we present the findings from the application of cross-correlation analysis to measure the time lag between cosmic ray counts and SN. The cross-correlation function was calculated for various time lags, revealing significant temporal relationships between these two variables. The analysis aims how changes in SA, as indicated by SN, correlate with variations in CRI detected by neutron monitors.

In this section, we present the findings from the application of cross-correlation analysis to measure the time lag between cosmic ray counts and SN, using both Pearson and Chatterjee correlation coefficients. Table 3 below summarizes the correlation coefficients and time lags for each solar cycle, along with the differences between the Pearson and Chatterjee methods. The above table presents the results only for strong (α=0.05), medium (α=0.5) and no smoothing factor (α=1).

Table 3. Cross-Correlation Results Using Pearson and Chatterjee Coefficients. The table shows the correlation coefficients and time lags for each solar cycle, along with the differences between the two methods and for the different smoothing factor α. In last column ΔLag is expressed in days.

α	Cycle	PCC	CCC	Lag (PCC)	Lag (CCC)	ΔLag (Days)
0.005	20	-0.936	0.783	58	67	-9
0.005	21	-0.966	0.827	500	444	56
0.005	22	-0.986	0.89	41	46	-5
0.005	23	-0.97	0.828	344	340	4
0.005	24	-0.979	0.868	130	89	41
0.005	25	-0.997	0.983	207	518	-311
0.005	2021	-0.874	0.63	334	368	-34
0.005	2223	-0.952	0.775	131	83	48
0.005	2425	-0.984	0.87	153	166	-13
0.5	20	-0.732	0.389	5	7	-2
0.5	21	-0.757	0.453	339	485	-146
0.5	22	-0.849	0.584	3	9	-6
0.5	23	-0.796	0.499	413	168	245
0.5	24	-0.793	0.521	124	8	116
0.5	25	-0.868	0.573	247	199	48
0.5	2021	-0.697	0.335	308	313	-5
0.5	2223	-0.808	0.501	112	110	2
0.5	2425	-0.817	0.52	125	65	60
1	20	-0.71	0.361	5	5	0
1	21	-0.738	0.43	339	362	-23
1	22	-0.834	0.556	4	16	-12
1	23	-0.775	0.477	413	169	244
1	24	-0.769	0.505	125	31	94
1	25	-0.848	0.549	247	221	26
1	2021	-0.678	0.31	308	252	56
1	2223	-0.791	0.485	112	121	-9
1	2425	-0.794	0.507	125	91	34

Notably, the PCC (in absolute values) are consistently higher (more negative) than the CCC, which is consistent with literature, suggesting a stronger linear relationship between the variables. However, the Chatterjee coefficient, also indicates substantial correlations, though systematically and generally weaker than Pearson's.

One of the key observations in our study is the consistent decrease in both Pearson and Chatterjee correlation coefficients as

the smoothing factor increases. This result can be expected, as applying heavier smoothing (using lower values of the EWMA smoothing coefficient alpha) can introduce artificial correlations into the modified data. The smoother the time series becomes, the more the variability is reduced, leading to higher correlations between the variables.

Smoothing is a technique used to reduce short-term fluctuations and highlight longer-term trends in time series data. By applying an EWMA with various α values, we control the degree of smoothing. A lower alpha value results in greater smoothing, which reduces the impact of transient noise and random fluctuations in the data. While smoothing helps in reducing noise, it also causes a loss of detailed variability and finer structures in the time series. This can lead to a decrease in the measured correlation coefficients. In the context of CRI and SN, this means that the nuanced, short-term interactions between these variables may become less pronounced, resulting in lower correlation values.

When the time series are smoothed, the high-frequency variations that contribute to the overall correlation are diminished. As a result, the linear correlation, measured by Pearson's correlation coefficient (PCC), tends to decrease. This reduction in PCC with increasing α is expected, as smoothing operations attenuate sharp changes. Chatterjee's correlation coefficient (CCC), while sensitive to different types of relationships, including non-linear dependencies, is also affected by the smoothing process. As the smoothing factor increases, both linear and non-linear components are smoothed out, leading to a decrease in CCC, although the behavior may differ from that of PCC.

From the data, we observed that with a lower smoothing factor (e.g., $\alpha = 0.005$), both Pearson and Chatterjee coefficients tend to be higher. This is because the raw time series data retain more of their inherent variability and noise, which can contribute to higher correlation values. As the smoothing factor increases (e.g., $\alpha = 0.5$ or $\alpha = 1$), the correlation coefficients decrease. The time series become smoother, highlighting long-term trends while suppressing short-term variations, leading to lower correlation values. This behaviour can be seen graphically and more fully in the figure 3.

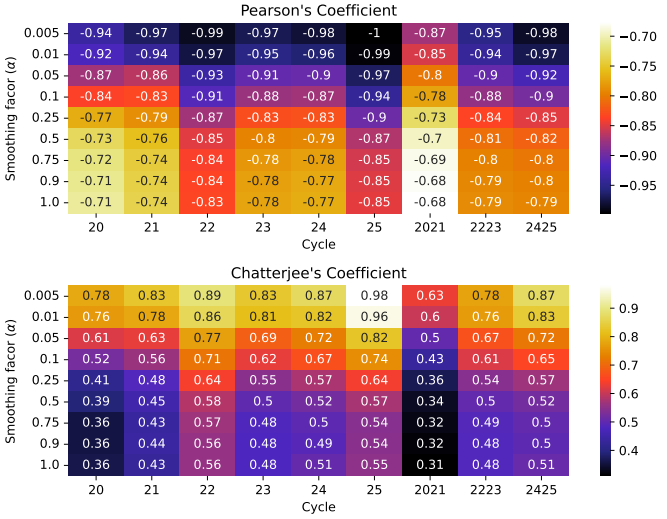

Fig. 3. Comparison of Pearson's and Chatterjee's correlation coefficients across different solar cycles and smoothing factors. The upper panel shows the PCC for CRI and SN across various solar cycles, smoothed using different α values. The lower panel presents the corresponding CCC. Note the variability and differences in correlation strengths and patterns between the two methods, highlighting Chatterjee's sensitivity to the impact of different smoothing factors. The color scales for Pearson and Chatterjee correlation coefficients appear inverted because Pearson correlation can take negative values, while Chatterjee correlation is always positive. This leads to different ranges of values in the color map for each coefficient.

4.2. Time Lag Analysis

The time lag analysis between CRI and SN using both PCC and CCC provides insights into the temporal dynamics of these relationships. We examined the time lags across different solar cycles and smoothing factors, comparing the results from both correlation methods.

As the smoothing factor changes, the time lag estimates from both PCC and CCC also vary. Generally, the time lags calculated using Pearson's method decrease systematically across different cycles and smoothing factors. This systematic decrease in Pearson's time lags suggests a robust capture of the underlying temporal dynamics, even as noise and fluctuations in the data are reduced through smoothing.

Conversely, the time lags calculated using Chatterjee's method do not exhibit the same systematic decrease as those calculated using Pearson's correlation. This variability in Chatterjee's time lags reflects its sensitivity to different patterns in the data and the impact of varying smoothing factors. While this can offer a complementary perspective on the data, it also introduces greater variability in the lag estimates.

Across different solar cycles, additional patterns emerge. Even solar cycles generally exhibit shorter time lags compared to their corresponding odd-numbered cycles. For instance, in cycle 22, the time lags are consistently shorter than those in cycle 21. This pattern holds true across various smoothing factors, indicating a possible underlying temporal structure that differs between even and odd solar cycles.

When considering pairs of solar cycles together (e.g., cycles 20-21, 22-23, and 24-25), the results show consistent patterns. The time lags for Pearson's coefficient tend to be greater and more stable (for all α) when two cycles are combined, compared to analyzing individual cycles. For the use of Chatterjee's coefficient this is not exactly the same. For example, in the combined analysis of cycles 20-21, the time lag for Pearson's coefficient shows a systematic decrease, while Chatterjee's coefficient exhibits a shorter and less stable lag. This pattern is repeated in the combined analyses of cycles 22-23 and 24-25, where Chatterjee's method consistently provides shorter time lags compared to Pearson's method. Figure 4 shows this behaviour visually.

For instance, in cycle 24, the differences persisted. For $\alpha = 0.005$, Lag PCC was 130 days while Lag CCC was 89 days, with a difference of 41 days. As smoothing increases, these differences become more evident, but Pearson's method

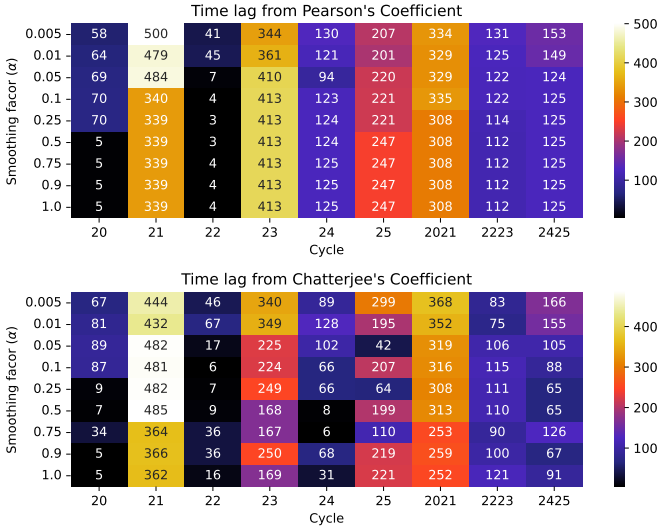

Fig. 4. Comparison of time lags calculated using Pearson's and Chatterjee's correlation coefficients across different solar cycles and smoothing factors.

still showed greater stability. Cycle 25 is still in its ascending phase, making it an atypical case. For alpha = 0.005, Lag PCC was 207 days while Lag CCC was 518 days, showing a significant difference of -311 days (almost a year's difference). This large discrepancy could be attributed to the ongoing increase in solar activity, which affects the lag estimates differently for each method. However, substantial differences are achieved over all smoothing cycles and smoothing factors.

The comparative analysis of time lags using Pearson and Chatterjee correlation coefficients across different solar cycles and smoothing factors reveals distinct strengths for each method. Pearson's method demonstrates superior consistency and systematic behavior, making it a valuable tool for understanding the temporal interactions between cosmic rays and solar activity. In contrast, Chatterjee's method provides shorter time lags when longer periods are analyzed, though it comes with greater variability in estimates. These findings underscore the importance of selecting the appropriate correlation method based on the specific research goals and characteristics of the data.

4.3. Measuring the Effectiveness of Alternative Correlation Estimators

To evaluate the effectiveness of using an alternative estimator for the correlation coefficient, we propose measuring the similarity of the time series both without applying any displacement and with displacements based on the time lags calculated using Pearson's and Chatterjee's methods. This approach will help us understand the impact of the different lag estimations on the alignment and similarity of the CRI and SN series.

To measure the similarity between the CRI and SN time series, we will employ two techniques: Dynamic Time Warping (DTW) (Senin (2008); Müller (2007); Sakoe & Chiba (1978)) and Mean Absolute Percentage Error (MAPE). These methods provide robust metrics for comparing the alignment and differences between the time series under various conditions.

DTW is a technique used to find the optimal alignment between two time series, allowing for warping in the time dimension. It calculates the minimum distance required to align the two series, considering both amplitude and temporal shifts. The formula for DTW is as follows:

$$\mathrm{DTW}(X, Y) = \sqrt{\sum_{(i,j)\in\pi} \|X_i - Y_j\|^2} \tag{5}$$

where X and Y are the two time series being compared, π represents the optimal alignment path between the points of X_i and Y_i, are the corresponding points in the series, and $\|\cdot\|$ denotes the Euclidean norm. By minimizing this distance, DTW provides a measure of similarity that is particularly useful for time series with varying temporal dynamics.

A higher DTW value indicates a greater dissimilarity between the two time series, meaning that more warping (i.e., adjustments in time and amplitude) is required to align them. Conversely, a lower DTW value signifies higher similarity, with less warping needed to achieve alignment. Therefore, lower DTW values are desirable as they reflect a closer alignment between the series.

Mean Absolute Percentage Error (MAPE) is a straightforward metric that computes the average absolute percentage differences between corresponding points in two time series. It provides a measure of the overall percentage discrepancy between the series, with lower values indicating higher similarity. MAPE is sensitive to both amplitude differences and phase shifts, making it a complementary metric to DTW.

The experimental procedure involves three main steps. First, we will measure the similarity between the CRI and SN time series without applying any displacement. This baseline measurement will provide a reference for comparing the effects of the displacement methods. Next, we will apply the time lags calculated using Pearson's correlation method to shift the CRI series relative to the SN series. We will then measure the similarity of the displaced series using DTW and MAPE. This step will help us assess how well the Pearson-estimated lags align the series. Finally, we will apply the time lags calculated using Chatterjee's correlation method to shift the CRI series relative to the SN series. Similar to the previous step, we will measure the similarity of the displaced series using DTW and MAE. This will allow us to compare the effectiveness of Chatterjee's lag estimates in aligning the series.

By comparing the DTW and MAPE values across these three conditions, we aim to determine whether Chatterjee's correlation coefficient provides a more effective estimation of time lags for aligning the CRI and SN series. To understand the potential benefits and limitations of using Chatterjee's method as an alternative to Pearson's method in this context, the results are graphically presented in Figure 5, which shows the ratio of the DTW and MAPE metrics for all the datasets and smoothing factors considered in this study.

The left panel of Figure 5 presents the ratio of DTW values obtained by aligning the CRI and SN series using the time lags calculated from Pearson's and Chatterjee's correlation coefficients across various smoothing factors. A ratio close to 1

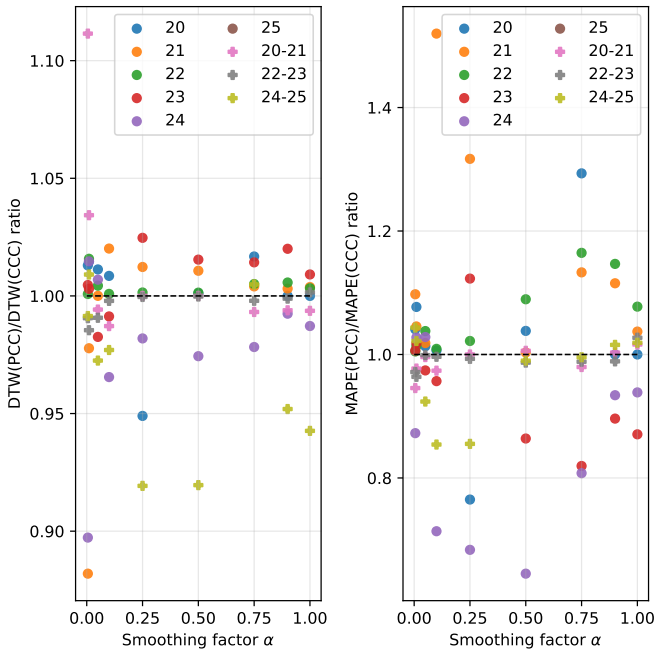

Fig. 5. Comparison of DTW and MAPE ratios for time series aligned using Pearson and Chatterjee correlation coefficients. The left panel shows the ratio of DTW values when the series are aligned using Pearson's time lag (numerator) versus Chatterjee's time lag (denominator) across different smoothing factors. The right panel displays the corresponding ratio of MAPE values. Both panels illustrate the ratios for individual solar cycles (20, 21, 22, 23, 24, 25) and combined cycles (20-21, 22-23, 24-25), and for all smoothing factor considered.

indicates similar performance between the two methods, while a ratio greater than 1 suggests that the DTW value using Pearson's time lag is higher (indicating less similarity) than using Chatterjee's time lag. Conversely, a ratio less than 1 implies that the DTW value using Pearson's time lag is lower (indicating more similarity) than using Chatterjee's time lag.

Observing the left panel, it is evident that for most cycles and smoothing factors, the ratios hover around 1, indicating comparable performance between the two methods in aligning the series. However, there are some deviations where the ratio significantly exceeds or falls below 1, suggesting differences in the alignment effectiveness of the two methods. Specifically, for cycles 20-21 and 24-25, the DTW ratios indicate a generally consistent performance across different smoothing factors, while cycles 22-23 show more variability, reflecting the differences in the temporal dynamics captured by the two correlation coefficients.

The right panel of Figure 5 shows the ratio of MAPE values for series aligned using the time lags derived from Pearson's and Chatterjee's correlation coefficients. Similar to the DTW ratios, a MAPE ratio close to 1 indicates comparable performance, while ratios significantly different from 1 suggest discrepancies in the alignment effectiveness. The MAPE ratios exhibit greater variability than the DTW ratios, with some cycles showing a pronounced difference in the error metrics between the two methods. For instance, cycles 22 and 23 display higher MAPE ratios, indicating that the alignment based on Chatterjee's time lags may result in lower prediction errors compared to Pearson's lags. In contrast, cycles 20-21 and 24-25 show ratios closer to 1, suggesting similar performance.

For all scenarios, more than half of the experiments show that alignment according to the time lag calculated by Chatterjee results in less error or more aligned series compared to using Pearson's coefficient. This indicates that Chatterjee's method often provides a more accurate alignment of the CRI and SN series.

These visualizations shows that while both correlation methods provide valuable insights into time lag estimation for aligning CRI and SN series, Chatterjee's method can offer improved alignment in specific cases. The observed variability across different cycles and smoothing factors underscores the importance of considering the specific characteristics of the time series when selecting the appropriate correlation method for time lag estimation.

The figure 5 shows that, in general, for the recent solar cycles 23, 24, and 25, as well as for data containing two combined cycles (23-24 and 24-25), Chatterjee's coefficient produces better results when considering the error metrics analyzed. This is reflected in the comparisons of DTW and MAPE, where alignments based on the lags calculated with Chatterjee's method show lower errors and greater similarity between the time series of cosmic ray intensity and sunspot numbers, compared to alignments based on the lags calculated with Pearson's coefficient. These results suggest that Chatterjee's correlation coefficient is particularly effective in improving the accuracy of predictions and alignments during these periods.

5. Conclusions

In this study, we analyzed the temporal relationship between cosmic ray counts and sunspot numbers using both Pearson and Chatterjee correlation coefficients. Our findings highlight the complexity and dynamic nature of the interactions between solar activity and cosmic ray intensity.

The results indicate that while Pearson's correlation provides a strong relationship with notable time lags, Chatterjee's correlation reveals shorter and potentially more accurate time lags. This dual approach allows for a more comprehensive understanding of the relationship between solar activity (SA) and cosmic ray modulation. The Pearson correlation coefficients showed consistently strong inverse relationships between cosmic ray counts and sunspot numbers across different solar cycles, with time lags varying significantly, sometimes reaching several hundred days, in line with similar findings reported in the literature. These large time lags suggest extended periods over which changes in SA influence CRI, aligning with the gradual propagation of solar effects through the heliosphere.

The Pearson correlation coefficients demonstrated consistently strong inverse relationships between cosmic ray counts and sunspot numbers across different solar cycles, with significant time lags that sometimes extended to several hundred days. These extended lags align with the gradual propagation of solar effects through the heliosphere. In contrast, Chatterjee's

correlation coefficients, while also indicating inverse relationships, were generally weaker but exhibited shorter and more consistent time lags. This difference shows the ability of Chatterjee's method to capture more immediate responses of CRI to SA changes. The differences in time lags and correlation strengths between the two methods emphasize the complex nature of these interactions and the necessity for multiple analytical approaches to fully understand them.

Additionally, our analysis using Dynamic Time Warping and Mean Absolute Percentage Error metrics showed that in more than half of the experiments, alignment based on time lags calculated with Chatterjee's correlation coefficient resulted in lower error and better-aligned series compared to using Pearson's coefficient. The practical utility of Chatterjee's method in improving the alignment of CRI and SN time series provide more accurate representations of their temporal relationships.

This study contributes to the field by demonstrating the utility of combining traditional correlation methods with more robust non-linear techniques like Chatterjee's correlation. This combined approach provides a deeper and more nuanced understanding of the temporal dynamics between SA and CRI. The insights gained from this study can help improve predictive models of space weather and enhance our understanding of the heliospheric processes affecting cosmic ray modulation.

Future studies could incorporate additional variables that influence cosmic ray intensity, such as geomagnetic indices, solar wind parameters, and interplanetary magnetic field data (See ref Tomassetti et al. (2022)), addition to extrapolating the benefit of other more robust correlation coefficients. This multivariate approach would enable a more holistic analysis of the factors affecting cosmic ray modulation. An interesting avenue for future work could be to establish how our results change when considering cosmic ray time series and try to look at comparative results trying to extract information about whether cutoff rigidity introduces significant changes (Sierra-Porta (2018, 2022); Sierra-Porta & Domínguez-Monterroza (2022); Tomassetti et al. (2022)).

Acknowledgments

The authors would like to thank the UTB Research Department for its unconditional support in this research. While the research has not been funded by any institution or entity, the author is grateful to UTB for the computer equipment provided for the development of the research activity and the generation of scientific computing results.

We gratefully acknowledge the Neutron Monitor DataBase (NMDB) and the World Data Center SILSO, Royal Observatory of Belgium, Brussels, for providing the invaluable datasets that made this study possible. The extensive and meticulously curated data from NMDB and SILSO have been crucial for our analysis of cosmic ray intensity and solar activity.

References

Binder, A. (1959). Considerations of the place of assumptions in correlational analysis. *American Psychologist*, *14*(8), 504. doi:https://psycnet.apa.org/doi/10.1037/h0048094.

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: comparison of pearson, spearman, transformation, and resampling approaches. *Psychological methods*, *17*(3), 399. doi:https://doi.org/10.1037/a0028087.

Bland, J. M., & Altman, D. G. (1994). Correlation, regression, and repeated data. *BMJ: British Medical Journal*, *308*(6933), 896.

Boschini, M., Della Torre, S., Gervasi, M. et al. (2018). Propagation of cosmic rays in heliosphere: The helmod model. *Advances in Space Research*, *62*(10), 2859–2879. doi:https://doi.org/10.1016/j.asr.2017.04.017.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32. doi:https://doi.org/10.1023/A:1010933404324.

Chatterjee, S. (2021). A new coefficient of correlation. *Journal of the American Statistical Association*, *116*(536), 2009–2022. doi:https://doi.org/10.1080/01621459.2020.1758115.

Dobynde, M., Harikumaran, J., Guo, J. et al. (2023). Cosmic radiation reliability analysis for aircraft power electronics. *IEEE Transactions on Transportation Electrification*, *10*(1), 344–352. doi:https://doi.org/10.1109/TTE.2023.3278319.

Dong, M., Wang, B., Wei, J. et al. (2023). Causal identification of single-cell experimental perturbation effects with cinema-ot. *Nature Methods*, *20*(11), 1769–1779. doi:https://doi.org/10.1038/s41592-023-02040-5.

Dorman, L. I. (2021). Space weather and cosmic ray effects. In *Climate change* (pp. 711–768). Elsevier. doi:https://doi.org/10.1016/B978-0-12-821575-3.00033-5.

Gervasi, M., Rancoita, P., Usoskin, I. et al. (1999). Monte-carlo approach to galactic cosmic ray propagation in the heliosphere. *Nuclear Physics B-Proceedings Supplements*, *78*(1-3), 26–31. doi:https://doi.org/10.1016/S0920-5632(99)00518-6.

Höeffgen, S. K., Metzger, S., & Steffens, M. (2020). Investigating the effects of cosmic rays on space electronics. *Frontiers in Physics*, *8*, 318. doi:https://doi.org/10.3389/fphy.2020.00318.

Hunter, J. S. (1986). The exponentially weighted moving average. *Journal of quality technology*, *18*(4), 203–210. doi:https://doi.org/10.1080/00224065.1986.11979014.

Idosa, C., Giri, A., Adhikari, B. et al. (2023). Variations of cosmic ray intensity with the solar flare index, coronal index, and geomagnetic indices: Wavelet and cross correlation approaches. *Physics of Plasmas*, *30*(8). doi:https://doi.org/10.1063/5.0157553.

Iskra, K., Siluszyk, M., Alania, M. et al. (2019). Experimental investigation of the delay time in galactic cosmic ray flux in different epochs of solar magnetic cycles: 1959–2014. *Solar Physics*, *294*(9), 115. doi:https://doi.org/10.1007/s11207-019-1509-4.

Jokipii, J., & Levy, E. (1977). Effects of particle drifts on the solar modulation of galactic cosmic rays. *Astrophysical Journal, Part 2-Letters to the Editor, vol. 213, Apr. 15, 1977, p. L85-L88.*, *213*, L85–L88. doi:https://doi.org/10.1086/182415.

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, *53*(6), 2576–2590. doi:https://doi.org/10.3758/s13428-021-01587-5.

Koldobskiy, S. A., Kähkönen, R., Hofer, B. et al. (2022). Time lag between cosmic-ray and solar variability: Sunspot numbers and open solar magnetic flux. *Solar Physics*, *297*(3), 38. doi:https://doi.org/10.1007/s11207-022-01970-1.

Kowalski, C. J. (1972). On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *21*(1), 1–12. doi:https://doi.org/10.2307/2346598.

Laken, B. A., & Čalogović, J. (2013). Composite analysis with monte carlo methods: an example with cosmic rays and clouds. *Journal of Space Weather and Space Climate*, *3*, A29. doi:https://doi.org/10.1051/swsc/2013051.

Lin, Z., & Han, F. (2023). On boosting the power of chatterjee's rank correlation. *Biometrika*, *110*(2), 283–299. doi:https://doi.org/10.1093/biomet/asac048.

Lucas, J. M., & Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, *32*(1), 1–12. doi:https://doi.org/10.1080/00401706.1990.10484583.

Maghrabi, A., Aldosari, A., & Almutairi, M. (2021). Correlation analyses between solar activity parameters and cosmic ray muons between 2002 and 2012 at high cutoff rigidity. *Advances in Space Research*, *68*(7), 2941–2952.

doi:https://doi.org/10.1016/j.asr.2021.05.016.

Mishra, A., Gupta, M., & Mishra, V. (2006). Cosmic ray intensity variations in relation with solar flare index and sunspot numbers. *Solar Physics*, *239*, 475–491. doi:https://doi.org/10.1007/s11207-006-0138-x.

Mishra, V., & Mishra, A. (2018). Long-term modulation of cosmic-ray intensity and correlation analysis using solar and heliospheric parameters. *Solar Physics*, *293*, 1–22. doi:https://doi.org/10.1007/s11207-018-1357-7.

Müller, M. (2007). Dynamic time warping. *Information retrieval for music and motion*, (pp. 69–84). doi:https://doi.org/10.1007/978-3-540-74048-3_4.

Potgieter, M. S. (2013). Solar modulation of cosmic rays. *Living Reviews in Solar Physics*, *10*, 1–66. doi:https://doi.org/10.12942/lrsp-2013-3.

Puth, M.-T., Neuhäuser, M., & Ruxton, G. D. (2014). Effective use of pearson's product–moment correlation coefficient. *Animal behaviour*, *93*, 183–189. doi:https://doi.org/10.1016/j.anbehav.2014.05.003.

Sadeghi, B. (2022). Chatterjee correlation coefficient: a robust alternative for classic correlation methods in geochemical studies-(including "triplecpy" python package). *Ore Geology Reviews*, *146*, 104954. doi:https://doi.org/10.1016/j.oregeorev.2022.104954.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, *26*(1), 43–49. doi:https://doi.org/10.1109/TASSP.1978.1163055.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia*, *126*(5), 1763–1768. doi:https://doi.org/10.1213/ANE.0000000000002864.

Senin, P. (2008). Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, *855*(1-23), 40. URL: https://www.researchgate.net/profile/Pavel-Senin/publication/228785661_Dynamic_Time_Warping_Algorithm_Review/links/02bfe5100f11a7929f000000/Dynamic-Time-Warping-Algorithm-Review.pdf.

Shi, H., Drton, M., & Han, F. (2022). On the power of chatterjee's rank correlation. *Biometrika*, *109*(2), 317–333. doi:https://doi.org/10.1093/biomet/asab028.

Sierra-Porta, D. (2018). Cross correlation and time-lag between cosmic ray intensity and solar activity during solar cycles 21, 22 and 23. *Astrophysics and Space Science*, *363*, 1–5. doi:https://doi.org/10.1007/s10509-018-3360-8.

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., & 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.

SILSO World Data Center (1964-2024). The international sunspot number. *International Sunspot Number Monthly Bulletin and online catalogue*, .

Stansby, D., Green, L. M., van Driel-Gesztelyi, L. et al. (2021). Active region contributions to the solar wind over multiple solar cycles. *Solar Physics*, *296*(8), 116. doi:https://doi.org/10.1007/s11207-021-01861-x.

Sullivan, J. H., Warkentin, M., & Wallace, L. (2021). So many ways for assessing outliers: What really works and does it matter? *Journal of Business Research*, *132*, 530–543. doi:https://doi.org/10.1016/j.jbusres.2021.03.066.

Tomassetti, N., Bertucci, B., & Fiandrini, E. (2022). Temporal evolution and rigidity dependence of the solar modulation lag of galactic cosmic rays. *Physical Review D*, *106*(10), 103022. doi:https://doi.org/10.1103/PhysRevD.106.103022.

Usoskin, I., Koldobskiy, S., Kovaltsov, G. et al. (2020). Revised gle database: Fluences of solar energetic particles as measured by the neutron-monitor network since 1956. *Astronomy & Astrophysics*, *640*, A17. doi:https://doi.org/10.1051/0004-6361/202038272.

Usoskin, I. G., Mursula, K., Solanki, S. et al. (2004). Reconstruction of solar activity for the last millennium using be data. *Astronomy & Astrophysics*, *413*(2), 745–751. doi:https://doi.org/10.1051/0004-6361:20031533.

Usoskin, I. G., Solanki, S. K., Krivova, N. A. et al. (2021). Solar cyclic activity over the last millennium reconstructed from annual 14c data. *Astronomy & Astrophysics*, *649*, A141. doi:https://doi.org/10.1051/0004-6361/202140711.

Ventura-León, J., Peña-Calero, B. N., & Burga-León, A. (2023). The effect of normality and outliers on bivariate correlation coefficients in psychology: A monte carlo simulation. *The Journal of General Psychology*, *150*(4), 405–422. doi:https://doi.org/10.1080/00221309.2022.2094310.

Wu, C. J., Usoskin, I. G., Krivova, N. et al. (2018). Solar activity over nine millennia: A consistent multi-proxy reconstruction. *Astronomy & Astrophysics*, *615*, A93. doi:https://doi.org/10.1051/0004-6361/201731892.

Zheng, Y., Jun, I., Tu, W. et al. (2024). Overview, progress and next steps for our understanding of the near-earth space radiation and plasma environment: Science and applications. *Advances in Space Research*, . doi:https://doi.org/10.1016/j.asr.2024.05.017.