Reconstruction of Kandilli Observatory Solar Flare Index Data

Calisir, M.[1], *Gerceker, K.*[1], *Ozguc, A.*[2], *Kilcik, A.*[1]

[1]Department of Space Science and Technologies, Akdeniz University Faculty of Science, 07058, Antalya, Turkey;

[2]Kandilli Observatory and Earthquake Research Institute, Bogazici University, 34684, Istanbul, Turkey

E mail: muhammed.calisirr@gmail.com; alikilcik@akdeniz.edu.tr; ozguc@boun.edu.tr

Accepted: 27 November 2025

Abstract: A solar flare is a sudden, rapid, and intense variation in brightness in the solar atmosphere, resulting from the abrupt release of energy stored in magnetic fields. The flare index serves as a valuable metric for quantifying this short-lived solar activity. Here, we aim to reconstruct the total solar flare index data from the Kandilli Observatory back to 1937. The monthly mean flare index values were obtained from two main sources: Astronomical Institute Ondrejov Observatory of the Czech Academy of Sciences (1937–1975) and Kandilli Observatory of Boğaziçi University, Istanbul, Turkey (1976–2024). To reconstruct Kandilli Observatory total solar flare index data we performed the Multiple Regression analysis using the monthly mean Sunspot Number (SSN), Sunspot Area (SSA) and geomagnetic aa index. Then we compared our reconstruction results with the Ondrejov Observatory flare index data. Our results exhibit good agreement with the observational data and provide a promising approach for gaining a deeper understanding of solar flare activity over an extended period.

Key words: solar flare index; reconstruction; multiple regression analysis; kandilli observatory

1. Introduction

The chromosphere is one of the most dynamic layers of the solar atmosphere. In this layer, complex interactions between magnetic fields and plasma flows result in a solar flare, a transient energetic event. Solar flares are sudden events characterized by a rapid and intense brightness increase that releases large amounts of radiant and charged particle energy (Shibata and Magara, 2011; Kusano et al., 2020; Li et al., 2022). Knowing the frequency and intensity of flares is important for understanding both short-term space weather events and long-term solar activity variations, and revealing their physical relationships. However, the "exact" quantitative value of the total energy produced by flares cannot be determined (Velasco Herrera et al., 2022). Kleczek (1952) defined a quantity of the form Q = i × t to quantify daily flare activity over a 24-hour period based on observations made with the Hα (6563 Å) filter. In this relationship, "i" represents the intensity scale of importance indicating the severity of the flare, and "t" denotes the duration of the flare in minutes (see http://www.koeri.boun.edu.tr/astronomy/fi_nedir.htm) Kleczek assumed that this product approximately corresponds to the total energy released during solar flares. The Flare Index is an important metric for evaluating both the quantitative and temporal characteristics of short-term solar activity.

Different studies have partially analyzed solar flare index variations with different temporal resolutions for the entire solar disk and for each of the solar hemispheres. Švestka (1956) compiled a catalog containing the monthly and yearly mean durations of activity based on observations made using spectrohelioscopes and spectroheliographs, and analyzed chromospheric flares throughout Solar Cycles 17 and 18. Events that differed in the two cycles were found to have changed only in the maximum year, 1947. Knoška (1985) analyzed the asymmetry of the flare index for the period 1936–1976. The author noted that there is no clear relationship between the variation of flare activity over time and the phase of the 11-year cycle of solar activity, neither in terms of north-south (N-S) nor east-west (E-W) asymmetry. On the other hand, the Solar Flare Index (SFI) has been studied over a wide temporal range, from Solar Cycle 20 to 24, through various analytical approaches. Özgüç et al. (2003) analyzed the temporal variability of the SFI between 1966-2001 using Fourier and wavelet transforms. They found that, regardless of whether the fundamental periods are harmonics or not, the significance of different periodicities and their temporal occurrence patterns was clearly revealed. Velasco Herrera et al. (2018) investigated the periodic behavior of Kandilli Observatory SFI, Ground Level Enhancement (GLE), and galactic cosmic ray (GCR) data from 1966 to 2014 using the wavelet analysis method. They stated that the synchronization of GCR periods with those emerging during SFI and GLE events may be associated with the formation process of GLEs. Özgüç et al. (2021) analyzed the periodic variations of total FI data in the Northern and Southern Hemispheres during the Solar Cycle 24 using Kandilli Observatory data with MTM and Morlet wavelet analysis. They reported that while all periodic changes were

DOI: 10.31401/SunGeo.2025.02.02

observed during the period when the solar cycle was maximum, no significant period was detected during the period when it was minimum. All these studies show that SFI is an important indicator of the temporal and periodic characteristics of solar activity and plays an important role in understanding the dynamic processes associated with different periods of solar cycles.

The most important problem for observational time series is the length and discontinuity of the data. Missing data, used methods, and variations in observing stations can make the index less reliable. Although datasets with the same indices may be similar in content, direct comparison may not be appropriate due to differences in observation methods and classification systems. Therefore, data reconstruction approaches are needed to ensure historical continuity and to fill in missing sections using modern analytical techniques. Karslıoğlu et al. (2019) created a new sunspot number dataset (CRz) using the total number of sunspots per day (Zurich number), based on the X-ray flare production potential. Using a linear regression equation between CRz and sunspot area, they recalculated and reconstructed the CRz dataset back to the year 1874. The reconstructed monthly data were then compared with observational data, and the results showed very good agreement (r = 0.98). Velasco Herrera et al. (2022) created a composite SFI time series for 1937–2020 by combining existing datasets from the Ondrejov Observatory of the Astronomical Institute of the Czech Academy of Sciences for the period 1937 to 1976 and the Kandilli Observatory in Istanbul for the period 1977 to 2020. In our study, Kandilli Observatory SFI data were reconstructed back to 1937 using the results of multiple regression analysis on SSN, SSA, geomagnetic aa index and Kandilli Observatory SFI data from 1976 to 2024. This provides a continuous and homogeneous SFI time series based only on Kandilli Observatory data, not two different observatories. Takalo (2023) analyzed the SFI for Solar Cycles 18 to 24 using principal component analysis based on the SFI data published by Velasco Herrera et al. (2022). They found that the first principal component of the SFI exhibits a more pronounced Gnevyshev Gap (GG) compared to other atmospheric parameters.

In this study, the mean total (northern + southern hemisphere) SFI data from the Kandilli Observatory for the period 1976–2024 were used to reconstruct the dataset back to 1937 using Multiple Regression Analysis. We compared the reconstructed dataset with the Ondrejov Observatory SFI data and found that both data sets show very good agreement (r = 0.85). The outline of the article is as follows: In Section 3, we present the data and methods used. In Section 4, we present the analysis results. In Section 5, our conclusions and discussions are presented.

2. Data and Method

2.1. Data

In this study, we used monthly mean solar activity indices (Sunspot Number, Sunspot Area) and the geomagnetic aa index to reconstruct the mean total SFI. The datasets used are described below.

a) ***Sunspot Number, (SSN):*** It is a direct indicator of the solar magnetic activity and exhibits a cycle of approximately 11 years. It has the oldest record of solar activity and is considered one of the best indicators due to its long duration (about 420 years) (Clette et al., 2014; Hathaway, 2015). Sunspot numbers are frequently used in monitoring, modeling, and forecasting solar cycles. The monthly mean SSN data were obtained from the Solar Influences Data Analysis Center (SIDC) (https://www.sidc.be/SILSO/datafiles).

b) ***Sunspot Area, (SSA):*** It refers to the total sunspot area on the surface of the Sun and is measured in surface areas expressed in millionths of the visible hemisphere of the Sun (MSH - millionths of solar hemisphere). The monthly SSA data, which show the total size of active regions in the northern and southern hemispheres, were taken from http://solarcyclescience.com/activeregions.html.

c) ***aa index:*** It is a parameter used to monitor long-term variations of geomagnetic activity in the Earth's magnetic field and was defined by Mayaud (1972). The aa index measures the largest deviations of the local magnetic field for each 3-hour period and converts the magnitude of these deviations into a numerical value. These values are then combined to obtain a daily average. The aa index, which has been calculated since 1868, has served as an important tool for studying long-term variations in geomagnetic activity (Mayaud, 1972). Due to its uninterrupted continuity from 1868 to the present day, it is ideal for long-term trend analysis of geomagnetic activity. Three-hourly data are from the International Service of Geomagnetic Indices (ISGI) (https://isgi.unistra.fr/indices_aa.php). The daily and monthly average values were obtained and used for the analysis.

d) ***Solar Flare Index, (SFI):*** The monthly mean total (northern + southern hemisphere) Flare Index data were taken from two different sources: for the period 1937–1975, from the Astronomical Institute Ondřejov Observatory of the Czech Academy of Sciences, and for the period 1976–2024, from the Bogazici University, Kandilli Observatory and Earthquake Research Institute, Istanbul, Türkiye.

The data sets used in this study represent the variation of solar activity and are in good agreement with the SFI. The data, extending back to 1937, enable the analysis of the long-term solar cycle. SSN, SSA, and aa index data are correlated with the total SFI data, and all follow the sunspot cycles (see Figure 1).

While SSN and SSA have hemispheric components, the geomagnetic aa index does not have a hemispheric component. Therefore, the total SFI data set was used in the analysis.

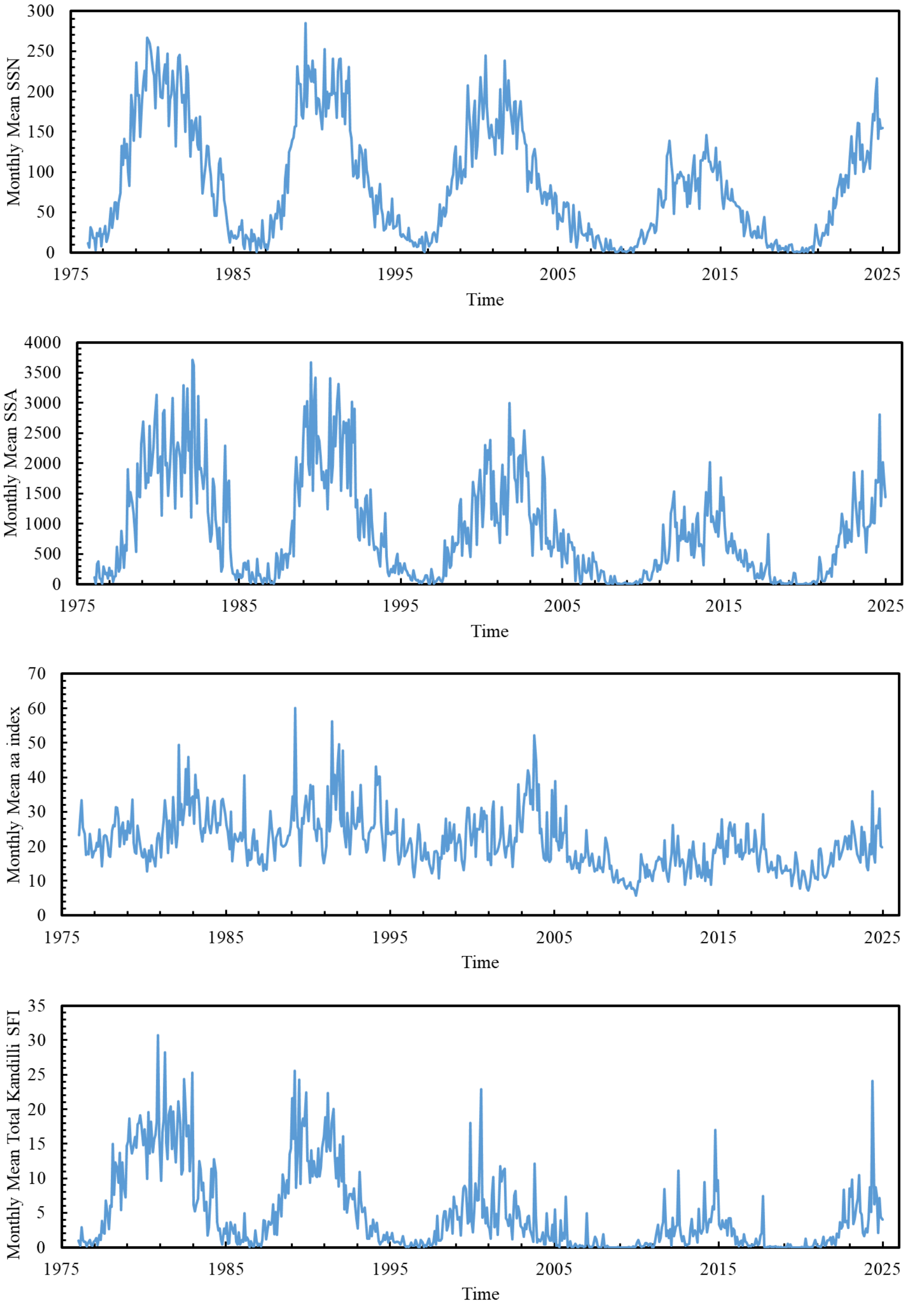

Figure 1. Temporal variations of SSN, SSA, aa index and Kandilli Observatory total SFI for 1976-2024

2.2. Multiple Regression Analysis

Multiple regression analysis (MRA) is used to examine the relationship between one dependent variable and multiple independent variables (Köksal, 1985; Tabachnick, 1996). This analysis helps to explain, predict, or model the dependent variable. It reconstructs the relationships between a dependent variable and a set of independent variables. The dependent variable is a phenomenon that we are trying to explain. The independent variable, on the other hand, is a phenomenon that is considered a causal factor and is assumed to influence or cause the dependent variable (Nayebi, 2020). Insufficient results can be obtained with only a single independent variable. This indicates that one independent variable does not provide enough information to predict the corresponding value of the dependent variable. A linear regression model can be used when the dependent variable is influenced by two or more controlled variables (Coker, 1995). In this study, we employed a MRA approach by incorporating additional independent variables in order to establish a meaningful relationship. The multiple regression equation is expressed as:

$$y = c_1 x_1 + c_2 x_2 + \ldots + c_k x_k + c_0 \qquad (1)$$

Where y is the dependent variable, x1,x2...xk are the independent variables, k is the number of independent variables, and c0,c1,c2...ck are the regression coefficients. In this study, the dependent variable is the SFI, while the independent variables are SSN, SSA, and aa index.

The MRA was applied between the monthly mean SSN, SSA, aa index, and Kandilli Observatory total SFI for the period 1976–2024. We obtained a regression equation using MRA (see Equation 2 and Table 1). Then, we reconstructed the Kandilli Observatory flare index data for the years 1937–1975 using the derived multiple regression equation. Finally, the reconstructed data were compared with the Ondřejov Observatory flare index data.

$$y = 0.005\, x_1 + 0.013\, x_2 + 0.083\, x_3 - 2.122 \qquad (2)$$

Table 1. Multiple Regression Analysis Equation Coefficients

Dependent Variable	İndependent Variables	Regression Coefficients
SFI = y	*SSA* = x_1	$c_1 = 0.005$
	SSN = x_2	$c_2 = 0.013$
	Aa = x_3	$c_3 = 0.083$
		$c_0 = -2.122$ *(Intercept)*

3. Results

We used Multiple Regression Analysis to obtain the Kandilli Observatory's monthly mean total SFI data. The monthly mean SFI values of the Astronomical Institute Ondrejov Observatory (1937–1975) were compared with the reconstructed Kandilli data. Within the scope of Multiple Regression Analysis, monthly mean SSN and SSA were used as variables representing solar activity, while the monthly mean aa index was used as an indicator of geomagnetic activity. These three variables were included in the regression analysis as independent variables in order to model the total SFI data from the Kandilli Observatory. A prediction model was created using the regression coefficients obtained from the analysis, and this model was used to reconstruct the missing Kandilli SFI data for the period 1937–1975 (Figure 2). The reconstructed values obtained to test the accuracy of the model were compared with data obtained from the Ondrejov Observatory for the same period. As a result of this comparison, a strong correlation (r = 0.85) was found between the two data sets (Figure 3). This finding shows that the regression model successfully represents the SFI in past periods and provides an effective method for reliably completing missing data.

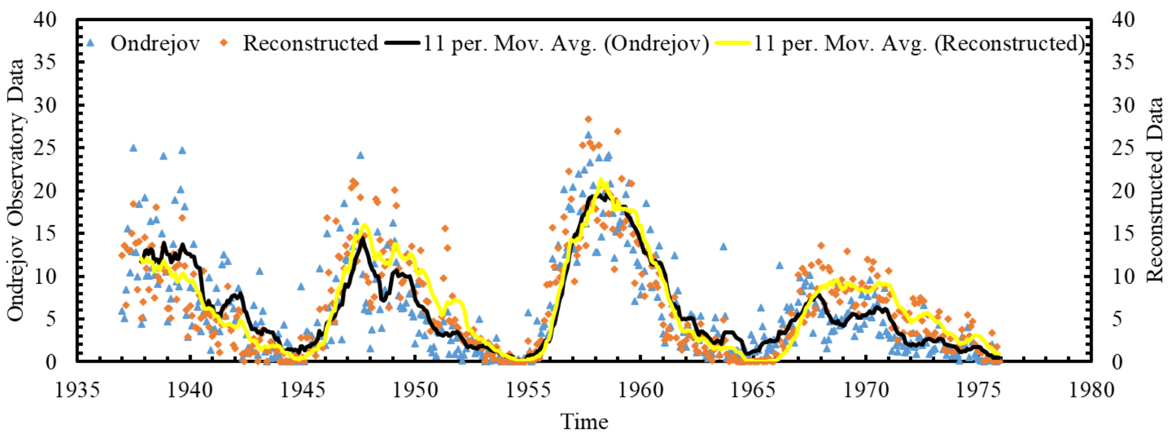

Figure 2. Comparison between the Kandilli Observatory monthly mean total SFI reconstructed for the years 1937–1975 using MRA and the Ondřejov Observatory data

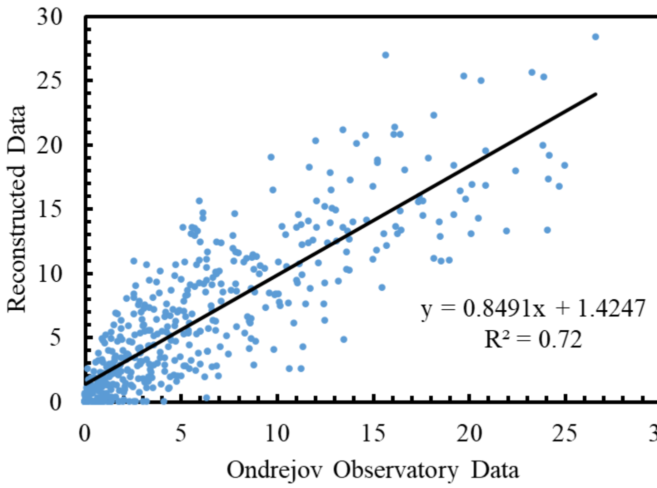

Figure 3. Relationship between the Ondřejov Observatory and the reconstructed Kandilli Observatory data

The performance of reconstruction was examined for periods of minimum and maximum activity. The minimum and maximum intervals between 1937 and 1976 were shown in Figure 4, as black (maximum) and green (minimum) vertical lines. To determine these intervals, first, the 6 steps moving average of the monthly mean SSN was calculated. Then, to determine the maximum intervals: (i) the approximate trough on the moving-average curve before reaching the maximum was located, and the corresponding point in the monthly mean data was taken as the start of the interval; (ii) the end points of the intervals were approximately determined by adding the maximum durations given by Kilcik et al. (2014) to the starting points. Low-activity intervals were defined in a similar way. For each interval, a correlation analysis was performed between the Ondřejov observations and the reconstructed data, and the correlation coefficients were obtained (Table 2). These results show that the MRA model performs better during the minimum phase compared to maximum.

Table 2. Maximum and minimum activity intervals and obtained correlation coefficients (r)

	Maximum Activity Intervals					
	Start	**End**	**Start**	**End**	**Start**	**End**
	1946.92	1949.75	1956.58	1959.50	1967.33	1971.25
r:	0.66		0.51		0.47	
	Minimum Activity Intervals					
	Start	**End**	**Start**	**End**	**Start**	**End**
r:	1942.17	1945.25	1952.75	1955.67	1963.17	1966.08
	0.76		0.76		0.79	

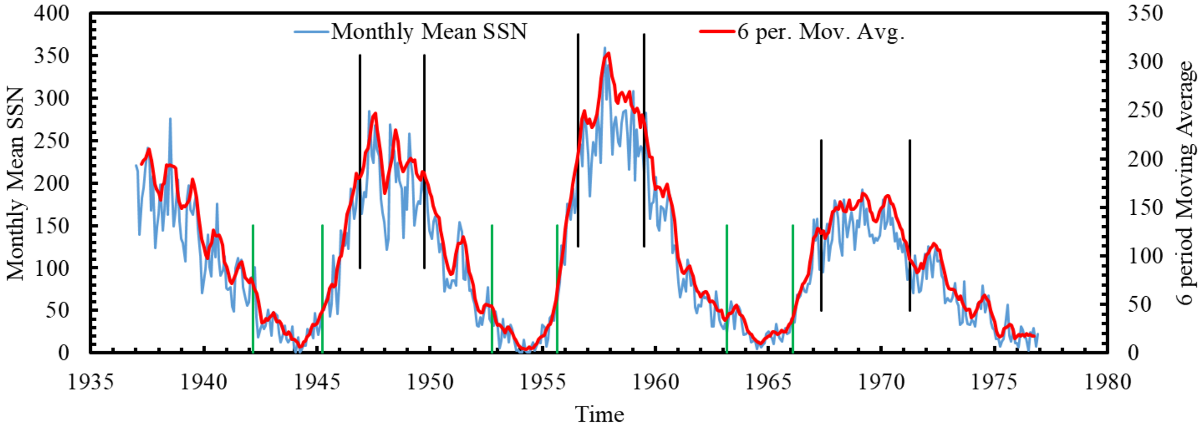

Figure 4. Monthly mean SSN between 1937 and 1976 (blue line) and its 6 steps moving average (red line). In the figure green vertical lines describe the intervals of minimum, and black vertical lines describe the intervals of maximum

4. Conclusions and Discussions

In this study, the monthly mean total (North + South Hemisphere) SFI data from the Kandilli Observatory for the period of 1937–1975 were reconstructed. MRA approach was applied to the Kandilli SFI data for the period of 1976–2024. The reconstructed Kandilli data were compared with the observational monthly mean SFI data from the Astronomical Institute Ondřejov Observatory (1937–1975). Our main results are as follows:

- Kandilli Observatory data have been restructured for the years 1937-1975 using MRA and a reliable correlation found between reconstructed data and Ondrejov Observatory SFI data (r=0.85).
- The reconstructed Kandilli Observatory total SFI data, ensuring data continuity, enhancing statistical reliability, and supporting in-depth analyses of long-term solar activity and space weather modeling.

SFI plays an important role in understanding the temporal and periodic characteristics of solar activity and has been widely studied in the literature (Švestka, 1956; Knoška, 1985; Ozguç et al., 2003; Ozguc et al., 2022; Kilcik, Rozelot and Ozguc 2024). The creation of long-term data sets and the completion of missing parts in the analysis of cyclical behavior provide a critical resource for both retrospective analysis and forward-looking forecasting models. Velasco Herrera et al. (2022) and Takalo (2023) studied on SFI datasets created between the years 1937 and 2020. These studies demonstrate that how important it is to reconstruct missing data using reliable methods for long-term analyses of solar activity. In this study, Kandilli SFI data was reconstructed to cover the period between 1937 and 1975. Using a multiple regression analysis approach that combines three different indices (SSA, SSN, aa), we have developed a comprehensive model based not only on solar-derived variables but also on parameters related to geomagnetic activity.

The performance of the MRA model was evaluated separately for periods of minimum and maximum solar activity. Three minimum and three maximum activity intervals were identified between 1937 and 1976, and the correlation coefficients between the observed and reconstructed SFI were calculated for each interval: for minimum periods, r = 0.76, 0.76, 0.79; for maximum periods, r = 0.66, 0.51, 0.47. These results indicate that the model performs more consistently during minimum periods, whereas the higher scatter during maximum periods leads to lower correlations. This difference can be considered a fundamental limitation of the MRA approach. MRA is a powerful method for analyzing the effects of multiple independent variables; however, its reliability and accuracy are limited by the linear nature of data, the size of the dataset, and the period it covers. In this study, the Ondřejov observational data (1937–1975) covers a relatively short time span compared to long-term solar activity, which imposes constraints on the MRA analysis. Furthermore, the relationships between the independent and dependent variables, as well as cyclical variability, affect the MRA results. The nonlinear, autocorrelated, and non-stationary nature of solar activity indices, particularly during periods of high solar activity, limits the model's performance. Despite these limitations, the MRA approach can reliably reconstruct SFI over multiple solar cycles. However, the varying performance across different activity levels highlights the method's sensitivity to high-variability conditions and should be considered when interpreting the reconstructed data. As a result, minimum periods are reproduced more reliably, whereas performance during maximum periods decreases due to increased variability and the complexity of solar flare activity.

Our study demonstrates that the Kandilli SFI data reconstructed using the MRA approach show very good agreement with the Ondrejov observations. This contributes significantly to ensuring data continuity, improving statistical consistency, and monitoring the long-term evolution of solar activity. At the same time, evaluating multiple independent variables together offers a more comprehensive modeling approach. The combined use of SSA, SSN, and aa indices provides a multidimensional analysis perspective that considers not only solar-derived variables but also geomagnetic effects. In particular, filling in missing or irregular data intervals for historical periods allows for a more robust examination of the dynamics of solar cycles. It also enables statistical testing of relationships between different indicators. This methodological approach goes beyond studies in the literature that generally focus on single-variable models, contributing to the development of prediction models that analyze the combined effect of multiple variables. The results obtained enable more detailed statistical analyses of past solar cycles and provide a solid data foundation for future space weather modeling. The restructuring methods applied have not only confirmed the accuracy of the data but also demonstrated an effective and applicable approach to extending the historical records of the flare index. Consequently, this provides a stronger foundation for conducting more reliable and comprehensive analyses of long-term solar activity trends.

5. Acknowledgements

The sunspot number data are taken from Solar Influences Data analysis Center (SIDC) (https://www.sidc.be/SILSO/datafiles). The aa index data are taken from (https://isgi.unistra.fr/indices_aa.php). The sunspot area index data are taken from (http://solarcyclescience.com/activeregions.html). The SFI data are taken from Kandilli (https://astronomi.bogazici.edu.tr/flare-index-arsivi) and Ondrejov Observatories (https://space.asu.cas.cz/~sos/flare_archive.html).

6. References

Ataç, T., Özgüç, A. Flare Index of Solar Cycle 22. Solar Physics 180, 397–407 (1998). DOI. 10.1023/A:1005047805338

Clette, F., Svalgaard, L., Vaquero, J.M. ve Cliver, E.W. 2014. Revisiting the sunspot number: A 400-year perspective on the solar cycle. Space Science Reviews, 186: s.35–103.

Coker, A. K. (1995). Fortran Programs for Chemical Process Design, Analysis, and Simulation. Gulf Professional Publishing, CHAPTER 1 - Numerical Computation, Pages 1-102, DOI. 10.1016/B978-088415280-4/50002-9.

Hathaway, D.H. 2015. The solar cycle. Living Reviews in Solar Physics, 12(1): s. 4.

Karslıoğlu, A. B, Dönmez, B, & Kilcik, A (2019). Total sunspot number calculation based on flare production potential of sunspot groups: reconstruction of SSN and its relation to some geomagnetic indices. Turkish Journal of Physics 43 (2): 187-195. DOI. 10.3906/fiz-1811-7

Kilcik, A., Ozguc, A. 2014. One Possible Reason for Double-Peaked Maxima in Solar Cycles: Is a Second Maximum of Solar Cycle 24 Expected? Sol Phys 289, 1379–1386. DOI: 10.1007/s11207-013-0407-4

Kilcik, A., Rozelot, J., and Ozguc, A., 2024, Periodic Behavior of Selected Solar, Geomagnetic and Cosmic Activity Indices during Solar Cycle 24. Universe, vol.10, no.3, id.107

Kusano, K., Iju, T., Bamba, Y., Inoue, S.: 2020, A physics-based method that can predict imminent large solar flares. Science 369, 587. DOI. 10.1126/science.aaz2511

Kleczek, J.: 1952, Catalogue de L'Activité des Éruptions Chomosphériques (Premiré Parties) 22, Publ. Inst. Centr Astron, Prague.

Knoška, Š.: 1985, Distribution of flare activity on the solar disk in the years 1937 – 1976. Contrib. Astron. Obs. Skaln. Pleso 13, 217.

Köksal, B. A. (1985). İstatistik analiz metodları. Türkiye,İstanbul: Çağlayan Kitabevi.

Li, T., Sun, X., Hou, Y., Chen, A., Yang, S., Zhang, J.: 2022, A new magnetic parameter of active regions distinguishing large eruptive and confined solar flares. Astrophys. J. 926, L14. DOI. 10.3847/2041-8213/ac5251

Mayaud, P.N. 1972. The aa indices: A 100-year series characterizing the magnetic activity. Journal of Geophysical Research, 77(34): s. 6870–6874.

Nayebi, H. (2020). Advanced statistics for testing assumed casual relationships. Springer International Publishing. ISSN:2367-1092, DOI. 10.1007/978-3-030-54754-7

Ozguç, A., Ataç, T. & Rybák, J. 2003, Temporal variability of the flare index (1966–2001). Solar Physics 214, 375–396,. DOI. 10.1023/A:1024225802080

Ozguç, A., Kilcik, A., Sarp, V., Yesilyaprak, H., Pekta¸ s, R. 2021, Periodic variation of solar flare index for
the last solar cycle (cycle 24). Adv. Astron. 8, 5391091. DOI. 10.1155/2021/5391091

Ozguc, A., Kilcik, A. and Yurchyshyn, V. 2022, Temporal and Periodic Variations of the Solar Flare Index During the Last Four Solar Cycles and Their Association with Selected Geomagnetic-Activity Parameters. Sol Phys 297, 112. doi:10.1007/s11207-022-02049-7

Shibata, K., Magara, T.: 2011, Solar flares: magnetohydrodynamics processes. Living Rev. Solar Phys. 8, 6. DOI. 10.12942/lrsp-2011-6

Švestka, Z.: 1956, Several notes of the statistics of chromospheric flares. Bull. Astron. Inst. Czechoslov. 7, 9.

Takalo, J. Analysis of the Solar Flare Index for Solar Cycles 18 – 24: Extremely Deep Gnevyshev Gap in the Chromosphere. Sol Phys 298, 86 (2023). DOI. 10.1007/s11207-023-02177-8

Tabachnick, B. G., & Fidell, L. S. (1996). Using Multivariate Statistics (3rd ed.). New York: Harper Collins.

Velasco Herrera, V.M., Perez-Peraza, J., Soon, W., Marquez-Adame, J.C.: 2018, The quasi-biennial oscillation of 1.7 years in ground level enhancement events. New Astron. 60, 7. DOI. 10.1016/j.newast.2017.09.007

Velasco Herrera, V.M., Soon, W., Knoška, Š. et al. The New Composite Solar Flare Index from Solar Cycle 17 to Cycle 24 (1937 – 2020). Sol Phys 297, 108 (2022). DOI. 10.1007/s11207-022-02035-z