



Predicting Solar Cycle 26 Using the Polar Flux as a Precursor, Spectral Analysis, and Machine Learning: Crossing a Gleissberg Minimum?

José-Víctor Rodríguez^{1,2} · Víctor Manuel Sánchez Carrasco^{3,4} ·
Ignacio Rodríguez-Rodríguez⁵ · Alejandro Jesús Pérez Aparicio^{3,4,6} ·
José Manuel Vaquero^{3,4}

Received: 22 March 2024 / Accepted: 9 August 2024 / Published online: 28 August 2024
© The Author(s) 2024

Abstract

This study introduces a novel method for predicting the sunspot number (S_N) of Solar Cycles 25 (the current cycle) and 26 using multivariate machine-learning techniques, the Sun's polar flux as a precursor parameter, and the fast Fourier transform to conduct a spectral analysis of the considered time series. Using the 13-month running average of the version 2 of the S_N provided by the World Data Center—SILSO, we are thus able to present predictive results for the S_N until January 2038, giving maximum peak values of 131.4 (in July 2024) and 121.2 (in September 2034) for Solar Cycles 25 and 26, respectively, with a root mean square error of 10.0. These predicted dates are similar to those estimated for the next two polar flux polarity reversals (April 2024 and August 2034). Furthermore, the values for the S_N maxima of Solar Cycles 25 and 26 have also been forecasted based on the known correlation between the absolute value of the difference between the polar fluxes of both hemispheres at an S_N minimum and the maximum S_N of the subsequent cycle, obtaining similar values to those achieved with the previous method: 142.3 ± 34.2 and 126.9 ± 34.2 for Cycles 25 and 26, respectively. Our results suggest that Cycle 25 will have a maximum amplitude that lies below the average and Cycle 26 will reach an even lower peak. This suggests that Solar Cycles 24 (with a peak of 116.4), 25, and 26 would belong to a minimum of the centennial Gleissberg cycle, as was the case in the final years of the 19th and the early 20th centuries (Solar Cycles 12, 13, and 14).

Keywords Sunspot number · Solar activity · Time series analysis · Machine learning · Fourier transform

1. Introduction

Analyzing solar activity is important since variations in the Sun's behavior affect both the Earth and the interplanetary environment (Pulkkinen 2007). Short-term solar activity variations, known as “space weather”, can be manifested through fluctuations in solar radiation and the magnetic field (Schwenn 2006). Furthermore, the occurrence of energetic phenomena, such as coronal mass ejections and solar flares (Kahler 1992; Howard 2014; Temmer

2021), can also trigger severe disturbances in the near-Earth space environment (Ansor, Hamidi, and Shariff 2019; Syed Zafar et al. 2021). Specifically, inclement space weather can adversely affect air travel or spaceflight and interfere with radars, satellites, the electricity grid, and high-frequency communications (Schrijver et al. 2015; Krausmann et al. 2016).

Meanwhile, the “space climate” also involves significant long-term solar activity changes spanning from decades to millennia (Solanki et al. 2004; Versteegh 2005; Usoskin 2023). It can be seen that these slower variations display 11-year solar cycles (Muñoz-Jaramillo and Vaquero 2019; Jayalekshmi, Pant, and Prince 2022; Clette et al. 2023; Usoskin 2023), which are involved, for example, in the variability of solar irradiance (Solanki and Krivova 2003; Chatzistergos et al. 2020; Kopp 2021) and magnetic flux (Usoskin et al. 2002). These fluctuations can also influence the atmosphere’s upper layers (causing warming), affect how many cosmic rays reach the Earth (Usoskin et al. 2021), and cause changes in the magnetosphere, which in turn influence the Earth’s geomagnetic activity (Mursula, Zieger, and Vilppola 2003). The current space climate context presents a scenario where the observed solar activity level has decreased from high solar activity in the mid-20th century to weak solar activity in the previous Solar Cycle 24 (Usoskin 2023). In this sense, it is worth mentioning that some authors (Feynman and Ruzmaikin 2011; Karak 2023) have proposed that we are currently at a minimum phase of the centennial solar cycle, as originally determined by Gleissberg (1939, 1945).

Predicting solar-cycle evolution is particularly relevant as it enables us to forecast how the Sun’s activity may affect the Earth. As these cycles determine the likelihood of extreme solar events like flares and coronal mass ejections, a society as dependent on technology as ours must pay attention to the Sun’s behavior in order to mitigate its impact.

With more than 400 years of systematically recorded data available, the sunspot record is the longest-lived extant dataset of direct solar observations (Clette et al. 2014; Vaquero et al. 2016; Arlt and Vaquero 2022; Clette et al. 2023). The so-called sunspot number (S_N) is thus widely used for describing future solar activity across a variety of scientific fields, including climate change, solar physics and spaceflight mission planning (Bothmer and Daglis 2007; Haigh 2007; Muñoz-Jaramillo and Vaquero 2019). This has drawn increasing scholarly and public attention to predicting solar activity via the S_N (Petrovay 2020; Overbye 2021).

Recently, significant research has been performed on this topic, with works offering various predictions for the maximum peak amplitude of S_N in the present solar cycle (Solar Cycle 25) (Nandy 2021). Various techniques have been used to make these predictions (Petrovay 2020). One approach uses physical methods to predict S_N . In this sense, Guo, Jiang, and Wang (2021) predicted a maximum S_N amplitude of 126 for Solar Cycle 25 (above that of Solar Cycle 24, which peaked at 116.4) based on a solar dynamo model. Statistical methods have also been used for sunspot prediction. This way, Aparicio, Carrasco, and Vaquero (2023), utilizing the slope of the inflection point during the cycle’s ascending part, predicted a maximum S_N of 131 ± 32 for Solar Cycle 25. Other researchers have used nonlinear modeling to make predictions. Employing the Kalman filter to calculate solar cycles under consideration of a nonlinear system of equations that modeled a dynamo system, Kitiashvili (2020) estimated a maximum S_N of 50 ± 15 in 2024–2025. Techniques based on spectral analysis have also been used to predict the S_N . Taking the annual sum of time series (from 1874 to 2013) of sunspot group areas, Javaraiah (2015) employed fast Fourier transform (FFT), the maximum entropy method, and Morlet wavelet analysis, producing a maximum S_N amplitude prediction for Solar Cycle 25 of 50 ± 10 . Several researchers have drawn on precursor methods to forecast the maximum S_N amplitude for Solar Cycle 25. For instance, considering the previous cycle’s peak and asymmetry as well as geomagnetic

indices as precursors, Lu et al. (2022) predicted an S_N peak value of 145.3 in October 2024. For their part, Asikainen and Mantere (2022) drew on solar magnetism periodicity in Hale cycles and the peak geomagnetic aa* index as a precursor to predict an S_N maximum of 171 ± 23 in May 2024.

One of the most successful and robust precursor methods for predicting the maximum S_N amplitude is probably the so-called polar precursor method, which considers the polar magnetic field as a precursor parameter for solar activity (Nandy 2021). The Sun's polar magnetic field plays a crucial role in solar activity dynamics and is fundamental to our understanding of our star's magnetic behavior (Muñoz-Jaramillo et al. 2012). This field undergoes a cyclical reversal approximately every 11 years, coinciding with the solar cycle. During this cycle, the fields weaken, reverse polarity, and then strengthen again, a phenomenon that is intricately linked to the generation and evolution of sunspots as well as the overall solar activity (Tobias 2023). On the other hand, a measure of the strength of the polar magnetic field over a specific area is the so-called polar flux. Thus, the polar flux quantifies the total magnetic field passing through an area near the Sun's poles. In this sense, it is essentially the product of the polar magnetic field and the area over which this field extends. Therefore, the polar flux is directly dependent on the strength of the polar magnetic field. Likewise, the polar flux can also be obtained from the so-called polar faculae, as shown in Muñoz-Jaramillo et al. (2012), since their formation is directly related to the Sun's magnetic field. In particular, the polar precursor technique exploits the notion that the amplitude of a solar cycle is likely proportional to the poloidal field's amplitude (or that of the polar flux) near the cycle's start. However, Kumar et al. (2021) used the polar field (and the proxy) data four years after the polar field reversal to predict an S_N peak for Cycle 25 of 120 ± 25 (in addition to forecast an S_N maximum of 126 ± 3 using the usual prediction of the polar precursor at cycle minimum). Subsequently, Kumar, Biswas, and Karak (2022) predicted a peak in Cycle 25 of 137 ± 23 , exploiting the fact that the polar solar field's growth rate correlates highly with the next cycle's amplitude and growth rate. Similarly, Biswas, Karak, and Kumar (2023) explored the reliability and robustness of using the polar field rise rate as a precursor for an early prediction of solar cycle. After that, Javaraiah (2023) predicted 125 ± 7 as the maximum S_N amplitude for Solar Cycle 25, using the polar magnetic field as a precursor and considering the fact that the main and secondary peaks of a solar cycle maximum (Gnevyshev peaks) correlate. Recently, Upton and Hathaway (2023) predicted an S_N peak for Solar Cycle 25 of 134 ± 8 , occurring in the fall of 2024. These authors used geomagnetic activity levels and the Sun's polar magnetic field configuration—particularly the polar fields and axial dipole moment at cycle minimum—as predictors.

Lastly, methods using machine learning (ML) have also been applied to sunspot prediction in past works. For example, Moustafa and Khodairy (2023) employed a hybrid model comprising Long Short-Term Memory (LSTM) and AutoRegressive Integrated Moving Average (ARIMA) algorithms to predict Solar Cycle 25's S_N peak, arriving at 137.04 in September 2024. Meanwhile, through a deep learning (DL) approach, Su et al. (2023) estimated 133.9 in February 2024, while Prasad et al. (2023) predicted 136.9 in April 2023 based on an LSTM model. Finally, Peguero and Carrasco (2023) also employed an LSTM model to forecast the maximum amplitude of Solar Cycle 25, concluding that it would be a new minimum of the Gleissberg cycle.

As the previous research outlined above shows, and as highlighted in Nandy's (2021) review, substantial disparity exists in the prediction results concerning Solar Cycle's peak S_N value and its timing, even when using the same method. This inconsistency reflects the inherent challenges in grasping the finer details of our star's behavior. To be specific, the mentioned 11-year period emerges clearly in the S_N series, remaining almost constant throughout. However, there are changes in the solar dynamo's behavior within the different cycles,

producing amplitude variations. Consequently, reliably predicting how the S_N will evolve—not just in the current cycle, but also generally in future cycles—is still a challenge. Hereby, we should note that the prediction techniques outlined above have certain limitations. For instance, not all statistical models account for the underpinning physical processes that produce sunspot cycles, often leading to inaccurate predictions (Nandy 2021). Meanwhile, the absence of detailed and accurate data on the solar system's initial conditions and parameters can diminish the effectiveness of physical or precursor models (Upton and Hathaway 2018). Finally, S_N prediction models that rely solely on ML approaches are unable to consider the fundamental solar physics driving sunspot generation (Camporeale 2019), limiting their applicability. Based on the discrepancies and limitations described above, there is an urgent need to deepen the research in this field by developing techniques that enable highly accurate solar activity forecasting (Asensio Ramos et al. 2023).

Seeking to overcome the limitations of single estimation models, we here combine several methods based on currently available data (up to September 2023) to develop an innovative technique for S_N prediction. Specifically, we consider a precursor parameter of solar activity, namely the polar fluxes of the northern and southern hemispheres (a proxy of the polar magnetic field), conduct a spectral analysis of the time series under consideration (to identify periodicities that can be used to lag the series, thereby producing new attributes that can serve as predictors in our prediction model), and lastly apply multivariate ML algorithms (where the dependent variable is S_N and the independent variables are the two series of the northern and southern polar fluxes, respectively, subjected to spectral analysis). The purpose of this approach is to create a method that incorporates all the strengths of the abovementioned techniques, thereby gathering the benefits from three distinct approaches under a single umbrella. We expect this method to offer substantially better predictions for solar cycles, thereby improving on the findings of prior work that solely used a univariate ML approach without precursor parameters (Rodríguez, Rodríguez-Rodríguez, and Woo 2022), ultimately offering a deeper understanding of the solar dynamo's long-term behavior.

Our focus lies, in particular, on forecasting the remainder of Solar Cycle 25 and Solar Cycle 26 using that innovative method. The outline of this article is as follows: the data and methodology used in this work are explained in Section 2. We present the results obtained along with our discussion in Section 3. Section 4 is dedicated to showing the main conclusions derived from this work.

2. Methodology

2.1. Data Description

The sunspot number data for this work were taken from the values of the S_N (version 2) index available on the website of the World Data Center SILSO (Clette and Lefèvre 2016) belonging to the Royal Observatory of Belgium (ROB) in Brussels: www.sidc.be/silso/datafiles.

To predict the S_N via the multivariate ML algorithms of the final proposed model, we consider two time series corresponding to the northern and southern polar fluxes. These hemispheric solar flux series, recognized as precursor parameters of solar activity, are used as the predictors (independent variables) of our dependent variable (S_N). The polar flux series corresponding to both hemispheres were extracted from Muñoz-Jaramillo et al. (2012).¹

¹dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/KF96B2.

2.2. Data Processing

Firstly, it should be noted that data from the above-mentioned site, corresponding to the polar fluxes, are annual. Therefore, a linear interpolation of the data of these series is carried out to obtain monthly estimated values (this work attempts a long-term forecast of solar cycle evolution, in terms of the 13-month smoothed S_N values). Next, we estimate a 13-month centered running average for the time series considered, where all elements of the averages have the same weight (1), except for the first and last elements, whose weight is 0.5 (as we have considered a 13-month centered running average, for each series, neither the first six months nor the last six months have smoothed averages). By adhering to this convention, we ensure that the results can be easily compared and contrasted across different investigations, promoting collaboration and knowledge advancement in this area. Moreover, by processing the data using such an average, along with the avoidance of a model too complex, we avert potential overfitting in the ML algorithms. Overfitting is possible when a model fits the training phase data too well and cannot correctly generalized to new data that were not incorporated during the training (this can be due, among other things, to the presence of noise in the training data).

In Figure 1 (bottom), the considered times series corresponding to the polar fluxes from both hemispheres can be observed, after applying the 13-month running average. The 13-

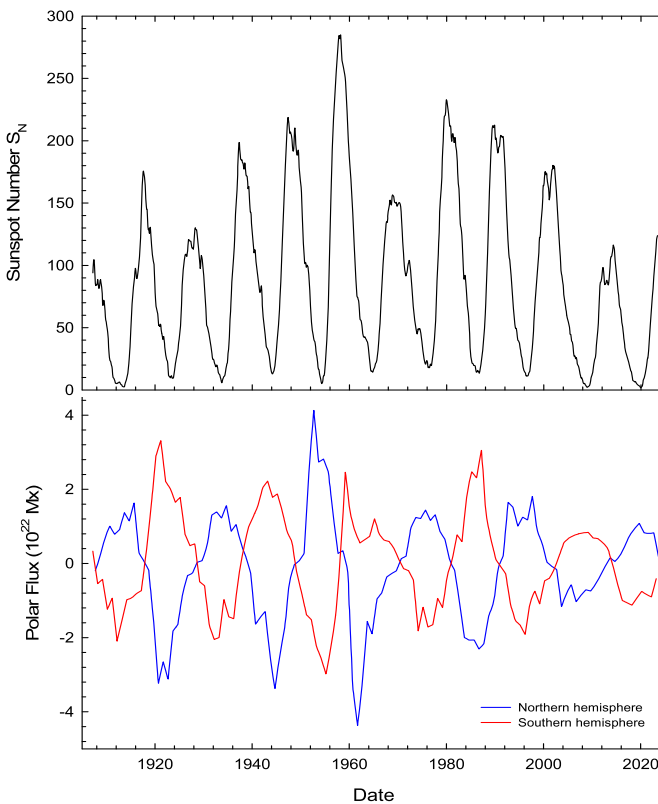


Figure 1 S_N times series (top) and considered time series corresponding to the northern and southern polar fluxes (bottom) after applying the 13-month running average for the same range of dates.

Table 1 Period and number of records for each considered series (after the 13-month running average has been applied).

Series	Period	Number of records
S_N	July 1749 – May 2023	3287
Northern Polar Flux	September 1907 – September 2023	1393
Southern Polar Flux	March 1907 – March 2023	1393

month smoothed S_N is also shown at the top of Figure 1. Table 1 displays date ranges for each series, in addition to the number of records of each, after the 13-month running average has been applied. It is worth mentioning that the original annual data for both polar flux series are published each year in different months for each hemisphere (September for the Northern Hemisphere and March for the Southern Hemisphere). This fact implies that although the same number of records is available for both series (1393), there will be a different number of predicted values for the polar flux in each hemisphere up to the stipulated common time horizon (January 2038), as shown below.

2.3. Time Series Fourier Transform Analysis

As mentioned above, Fourier-transform-based spectral analysis permits the extraction of certain characteristics of a given time series, which enhances predictive ML algorithms' performance (Koç and Koç 2022). Hence, we apply FFT to all 13-month smoothed data series (including the dependent variable S_N), seeking to identify suitable periodicities with which these series can be lagged (by a number of positions equal to these periods or multiples thereof) as frequently as possible for as long as the series have data. This produces lagged series, which in turn offer new predictors or attributes that can be incorporated into the input dataset for the predictive ML algorithms. This approach is based on the idea that for time series containing certain identifiable periods, their behavior at instants lagged by a time equivalent to those periods—or multiples thereof—can offer useful information for estimating these series' behavior in the future. We conduct this analysis for all series, because when forecasting S_N (target), the multivariate ML algorithm uses as predictors (inputs) not only the S_N series itself, lagged in line with its periodicities, but also the polar flux series already forecasted until the time horizon for which the S_N prediction is to be made. Thus, in order to previously carry out such polar flux series predictions, a univariate ML forecast will be performed for each polar flux series, wherein the predictors are, in turn, the series lagged according to their periodicities. Figure 2 displays the periodograms for the series considered here, indicating the different peaks that represent the various identified periodicities. Table 2 presents the main periodicities' values alongside the number of positions they represent regarding these series' displacement.

2.4. ML Algorithm Implementation, Selection, and Application

The different ML models employed in our S_N prediction approach are implemented in Python (scikit-learn library). We select linear regression (LR), random forest (RF), support vector machines (SVM) and Gaussian processes (GP) as these provide a set of diverse strategies that permit us to perform multivariate analysis on time series data that have cyclical as well as nonlinear patterns, i.e., the S_N and the two polar flux series. Moreover, these algorithms are well known for their capability for modeling sequential data. A more detailed

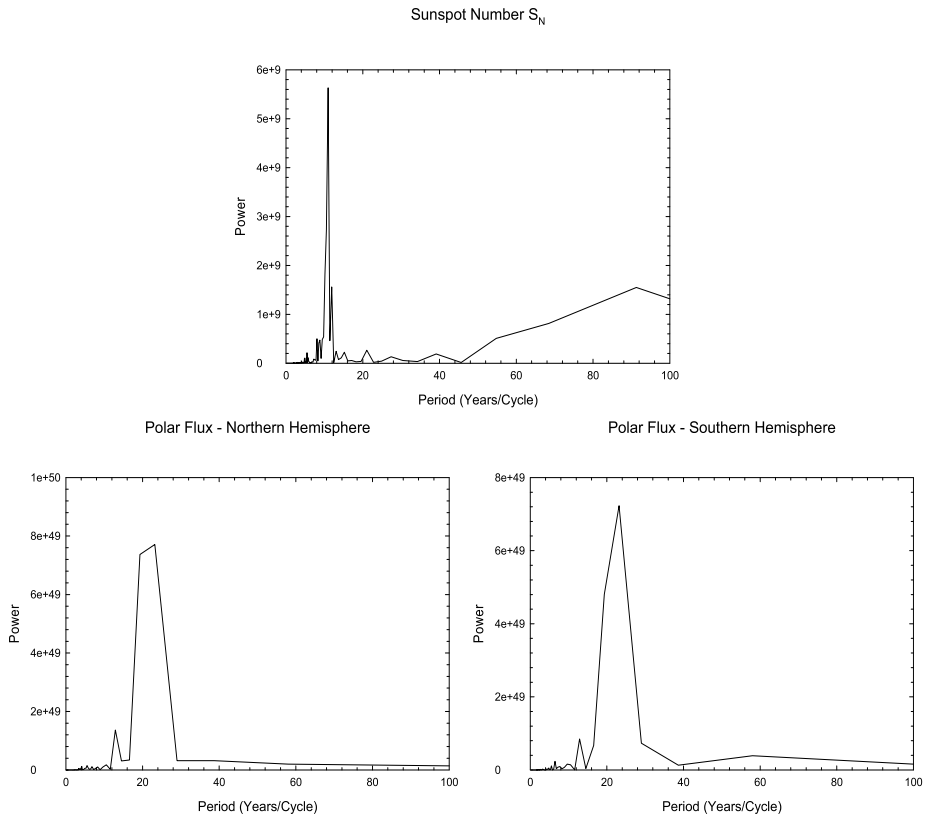


Figure 2 Periodograms of the sunspot number (top panel), and the northern (bottom-left panel) and southern (bottom-right panel) polar fluxes.

Table 2 Main periodicities of the considered time series (obtained from their respective periodograms) and number of positions to be lagged that these periods represent.

Series	Periods (years/cycle)	Number of corresponding lag positions
S_N	5.70, 8.83, 10.95, 15.21, 54.76, 91.27	68, 106, 131, 182, 657, 1095
Northern Polar Flux	12.88, 19.33, 23.2	154, 232, 278
Southern Polar Flux	12.88, 19.33, 23.2, 58	154, 232, 278, 696

description of these algorithms can be found in Shmueli and Lichtendahl (2016) for LR, Vapnik (2013) for SVM, Oshiro, Perez, and Baranauskas (2012) for RF, and Seeger (2004) for GP. The root mean square error (RMSE) estimated in the test phase is used to assess the goodness of fit of each algorithm. Then, we select the algorithm with the best performance to predict the polar flux series and S_N .

Notably, the series' predictive time horizon is January 2038. In this sense, the prediction period for the S_N will be from June 2023 to January 2038 (176 records), that of the northern polar flux will be from October 2023 to January 2038 (172 records), and that of the southern polar flux will be from April 2023 to January 2038 (178 records). This way, the number of

Table 3 Initialization parameters for the different ML algorithms employed.

Algorithm	Parameters
RF	<ul style="list-style-type: none"> - Bag size (percentage of training set size): 100. - Thread number: 1. - Tree number: 100. - Maximum tree depth: unlimited. - Number of random features: 0. - Seed: 1.
LR	<ul style="list-style-type: none"> - Ridge parameter: 10^{-8}.
GP	<ul style="list-style-type: none"> - Gaussian noise level: 1.0. - Seed: 1. - Polynomial kernel: size of the cache = 250,007; exponent = 1.0.
SVM	<ul style="list-style-type: none"> - C parameter: 1.0. - Polynomial kernel: size of cache = 250,007; exponent = 1.0. - Sequential minimal optimization (SMO) with epsilon for round-off error = 10^{-12}; epsilon parameter for epsilon insensitive loss function = 0.001; tolerance = 0.001; seed = 1.

records reserved for the test phase in each case is assumed to be identical to the number of records to be predicted from known data up until January 2038: 176 for the S_N , 172 for the northern polar flux and 178 for the southern polar flux. We dedicate the remaining data in each series (not used for the test phase) to training and validation, using a n -split expanding window cross-validation in the training/validation phase to adjust the hyperparameters. The number of splits is selected in order to provide a balance between the amount of data used for training and the robustness of the validation, ensuring that the model is evaluated on different data segments to improve its generalization. Therefore, n will be different for the S_N series ($n = 5$) and for the northern and southern polar fluxes ($n = 3$). The validation window is assumed to consist of a number of records (v) equal to that in the testing phase and the training dataset of each split/iteration consists of a number of records of $\text{round}[i/n \cdot (T - v)]$ (where i is the number of split/iteration—from 1 to n —and T is the total number of records reserved for training and validation). Table 3 shows the initialization parameters for the different ML algorithms. It should be noted that the choice of these parameters responds to the call for a balance between performance, accuracy, and computational complexity.

3. Results and Discussion

3.1. Multivariate Prediction Model Construction

To begin with, we perform a univariate prediction of the two polar flux time series until January 2038. We hereby assume the series themselves as the predictors of each, lagged by the number of positions equal to their periodicities, as outlined previously. Thus, each prediction is an ultimately autoregressive/multivariate problem (although we only consider the time series data). To this end, we draw on the four abovementioned ML algorithms, i.e., RF, LR, GP and SVM. Table 4 presents the RMSE values for each algorithm in the test phase of the respective prediction models constructed for the two polar flux series (for each series,

Table 4 RMSE values (in units of 10^{21} Mx) for four algorithms considered for the test phases of each prediction model for both polar flux series.

Series	Number of records for the Test Phase	RF	LR	GP	SVM
Northern Polar Flux	172	4.52	5.33	6.35	5.64
Southern Polar Flux	178	3.95	3.55	4.24	2.98

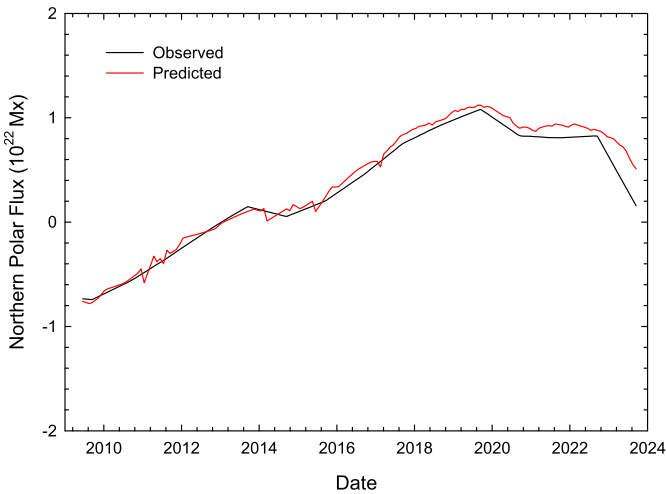


Figure 3 Observed data compared with predicted data for the northern polar flux univariate prediction model’s testing phase (172 records) using the RF algorithm.

the lowest RMSE value from among the four algorithms is highlighted in bold). Furthermore, the above-mentioned number of records considered for the test phase in each case is also provided (which is different for each hemisphere, since the number of records for the test phase has been assumed to be the same as the number of values to be predicted in each series until January 2038, and as discussed above at the end of section ‘Data processing’, the latter number differs for each hemisphere). It clearly emerges that for the northern polar flux series, the RF technique makes the best prediction based on RMSE, while the southern polar flux series is most reliably predicted by SVM.

It should be noted that although the data type of the two series is the same, factors related to the specific behavior of each series, as well as to the particular characteristics of the ML model considered, may cause the lower RMSE to be obtained with a different algorithm in each case. In this sense, RF models perform better when trying to capture particularly nonlinear and complex interactions in the data, as well as more abrupt fluctuations in the series (Breiman 2001), as is the case with the northern polar flux series (see cycles between 1950 and 1970). Nevertheless, SVM models are more effective in series where the behavior of the data is more stable and less subject to strong fluctuations (Cristianini and Shawe-Taylor 2000), as is the case for the southern polar flux series.

Figures 3 and 4 show a comparison between predicted and observed values for the test phases of both the northern polar flux (172 records, from June 2009 to September 2023,

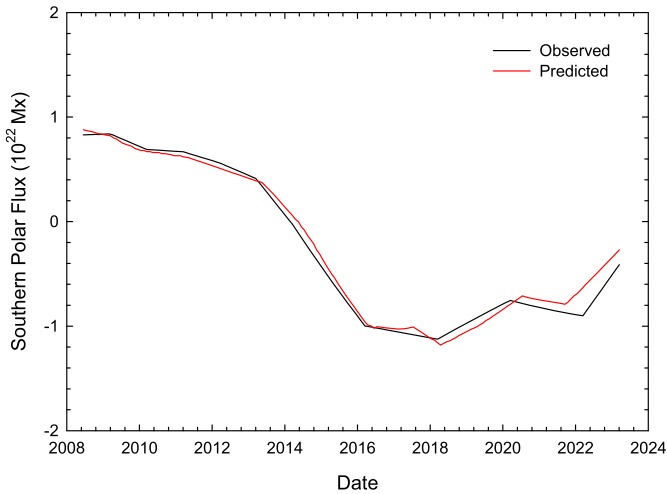


Figure 4 Observed data compared with predicted data for the southern polar flux univariate prediction model's testing phase (178 records) using the SVM algorithm.

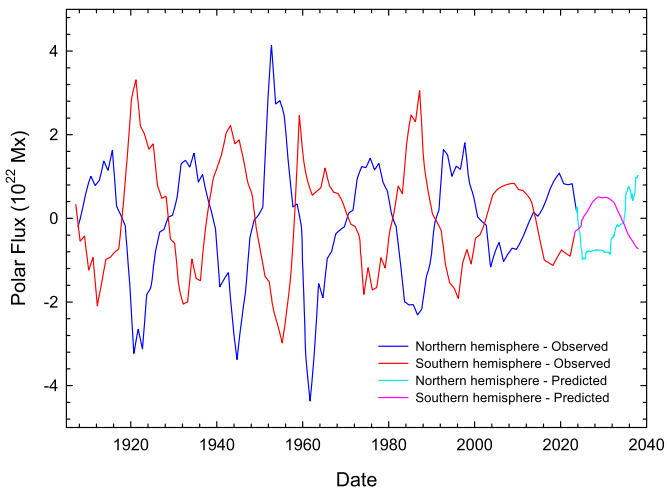


Figure 5 Prediction curves for the two considered polar flux series until January 2038.

using the RF algorithm) and the southern polar flux (178 records, from June 2008 to March 2023, using the SVM algorithm), respectively.

As can be seen in Figures 3 and 4, the prediction for the test phases of both the northern and southern polar flux series is consistent with the observed data, as indicated by the RMSE values obtained for both cases: $9.7 \cdot 10^{20}$ Mx and $8.3 \cdot 10^{20}$ Mx, respectively.

Therefore, Figure 5 presents the prediction curves (until January 2038) for the two polar flux series, each obtained with the algorithm that produced the lowest RMSE value. Observed data are blue for the Northern Hemisphere and red for the Southern Hemisphere. Predictions are light blue for the Northern Hemisphere and pink for the Southern Hemisphere.

Table 5 RSME values obtained in the test phase by the four algorithms considered to build the multivariate model to predict the S_N series.

Algorithm	RMSE
RF	10.04
LR	16.48
GP	15.44
SVM	11.13

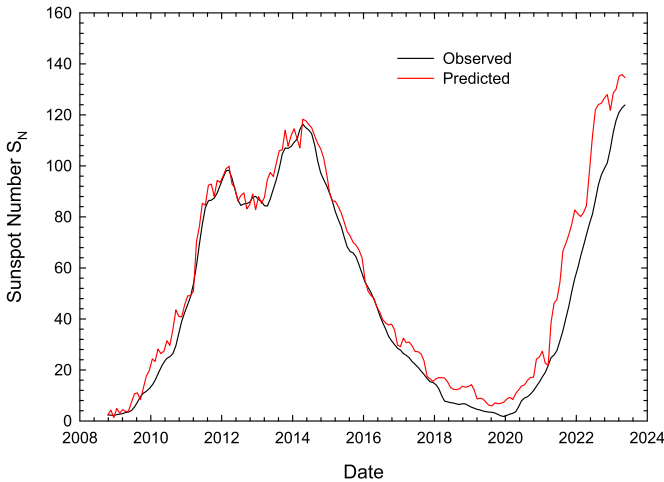


Figure 6 Observed data compared with predicted data for the S_N multivariate prediction model’s testing phase (176 records) using the RF algorithm.

Lastly, we construct the multivariate model for predicting the S_N (until January 2038) using the four considered algorithms (RF, LR, GP and SVM). As predictors, we assume both the S_N series itself, lagged by various numbers of positions according to periodicity (Table 2), and the two polar flux precursor series, as predicted until January 2038. Table 5 presents the RMSE values obtained during the testing phase (176 records) for the four algorithms. As shown in Table 5, the lowest RMSE for the test phase (10.04) is produced by the RF algorithm; hence, we consider this algorithm for the multivariate prediction of S_N .

Figure 6 presents a comparison between the observed values and our predictions for the testing phase (red line) (from October 2008 to May 2023) based on RF.

The curves are in good agreement in terms of the cyclic behavior of the series and, in particular, throughout the entire Solar Cycle 24. The model hereby offers a good fit regarding the S_N values and the maxima’s occurrence dates. For Solar Cycle 24, the observed and predicted maxima are 116.4 and 118.3 ± 10.0 , respectively, both of which occurred in April 2014.

Figure 7 compares the modeled and observed data in a training phase time slot based on RF (only since 1970—as it is merely a sample to illustrate the goodness of fit of the model in the training phase—and up to September 2008, where the data reserved for the test phase begins). There is excellent agreement, as shown by the RMSE for this phase of 2.02.

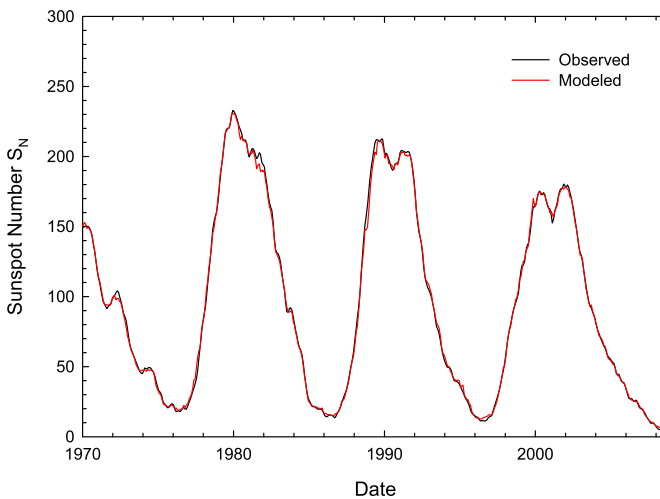


Figure 7 Example comparison between observed data and modeled data for a training phase time slot of the S_N multivariate prediction model (since 1970 and up to September 2008) using RF.

3.2. Predicting Activity for Solar Cycles 25 and 26

Figure 8 provides the final multivariate prediction (grey line, top panel) of the S_N series until January 2038 (176 records), alongside the previously made predictions for the two selected polar flux series (light blue and pink lines, bottom panel).

The model predicts a maximum S_N of 131.4 ± 10.0 (in July 2024) for Solar Cycle 25 and 121.2 ± 10.0 (in September 2034) for Solar Cycle 26. The S_N values obtained in both solar cycles are significantly below the average of the maximum amplitudes considering Solar Cycles 1–24, which is ~ 180 . In this sense, we note that the observations made during Solar Cycle 25 so far hints that this cycle will be below average within the lowest 25th percentile (Carrasco and Vaquero 2023; Ridley 2023). Furthermore, although there have been some predictions regarding the maximum amplitude of Cycle 25 that forecasted that this cycle will be a strong cycle (Han and Yin 2019; McIntosh et al. 2020; McIntosh, Leamon, and Egeland 2023), our prediction for Cycle 25 is in agreement with most of the predictions covered in Nandy (2021). Meanwhile, the predicted maximum of Solar Cycle 26 is slightly lower than that of the preceding Solar Cycle 25 and is similar to the maximum amplitude observed in Solar Cycle 24 (116.4). This would imply that Solar Cycles 24, 25, and 26 belong to a minimum in the centennial Gleissberg cycle. As an example, this was the case for Cycles 12, 13, and 14 in the final years of the 19th century and near the beginning of the 20th century. Notably, the average maximum amplitude for Solar Cycles 12–14 was 126.0, whereas based on our results it would be 123.0 for Solar Cycles 24–26. Moreover, our findings point to the same direction as those obtained by Feynman and Ruzmaikin (2012, 2014) and Karak (2023), who contended that Solar Cycles 23 and 24 could belong to a minimum of the Gleissberg cycle. In addition, the predicted dates for the two S_N maxima for Solar Cycle 25 and 26 (July 2024 and September 2034, respectively) are similar to those obtained for the two polar flux polarity reversals for Solar Cycles 25 and 26 (April 2024 and August 2034, respectively). In this regard, it is worth mentioning that our prediction for the next polar flux reversal for Cycle 25 (April 2024) would be in agreement with the latest data

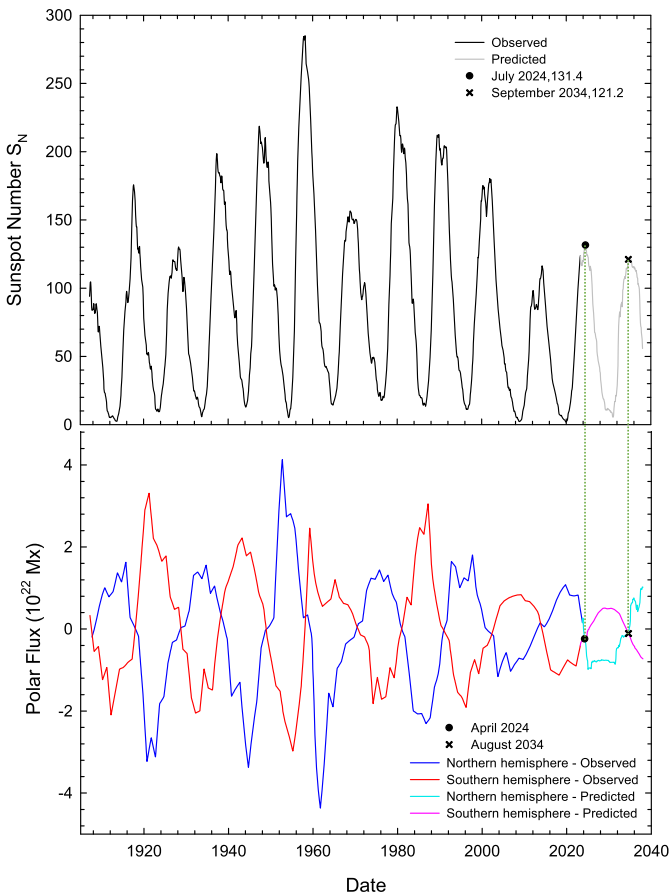


Figure 8 Final multivariate prediction of the S_N series (grey line, top panel) until January 2038, as well as the previously made predictions of the selected two polar flux series (light blue and pink lines, bottom panel).

provided by the Wilcox Solar Observatory (WSO), since signs of such a polar flux reversal have already been observed.

On the other hand, there is a well-known relationship between the absolute value of the difference between the polar fluxes of both hemispheres at a minimum of the S_N and the maximum of the S_N of the subsequent cycle (Upton and Hathaway 2023). In this sense, it is worth mentioning that the polar field at the S_N cycle minima exclusively impacts the magnitude of the subsequent S_N cycle (Karak and Nandy 2012; Nandy 2021). Figure 9 shows the S_N peak values of Solar Cycles 15 to 24 as a function of the absolute value of the difference between the polar fluxes of both hemispheres at the minimum S_N preceding each cycle.

The values in Figure 9 can be fitted by a linear regression line with the equation $y = 3 \cdot 10^{-21}x + 87.933$ (with a correlation coefficient $r = 0.7$ and a p -value = 0.025), being x the absolute value of the difference between polar fluxes of both hemispheres at the minimum S_N preceding each cycle and y the maximum amplitudes of the S_N cycles. Therefore, if we apply such an equation to predict the S_N maxima for Solar Cycles 25 and 26, we obtain maximum amplitudes of 142.3 for Solar Cycle 25 (with an absolute value of the difference

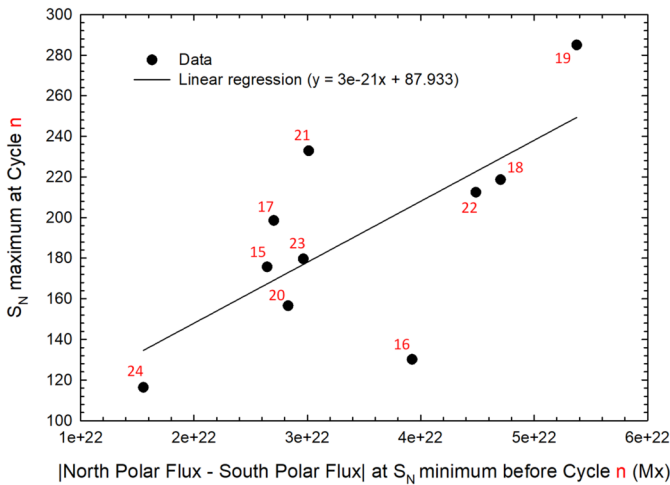


Figure 9 S_N peak values of Cycles 15 to 24 as a function of the absolute value of the difference between polar fluxes of both hemispheres at the minimum S_N preceding each cycle.

between the polar fluxes at the preceding S_N minimum of $1.81 \cdot 10^{22}$ Mx, which is already known from observed data) and 126.9 for Solar Cycle 26 (with an absolute value of the difference between polar fluxes at the preceding S_N minimum of $1.29 \cdot 10^{22}$ Mx, which, in this case, is obtained from the univariate predictions previously carried out for both polar fluxes). Taking into account that the RMSE obtained when calculating the regression line is 34.22, the S_N peak values estimated using this method for Solar Cycles 25 and 26 closely align (within the margin of error) with those predicted with the multivariate ML prediction (131.4 ± 10.0 and 121.2 ± 10.0 , respectively). Therefore, it is noteworthy to point out that we obtain similar predictions for the S_N maximum amplitudes of Solar Cycles 25 and 26 using two different approaches.

4. Conclusions

This work has proposed a novel and innovative technique for solar cycle prediction, specifically the S_N , employing multivariate ML techniques, considering both hemispheres' polar fluxes as precursor parameters, and subjecting the considered time series to spectral analysis via FFT.

The RF algorithm yielded the lowest RMSE in the testing phase (10.0) and was therefore used in the final multivariate ML prediction of S_N . We thus present our predictive results for the S_N of Solar Cycles 25 and 26, namely maxima of 131.4 (in July 2024) and 121.2 (in September 2034), respectively. Our results suggest that Cycle 25 will have a below-average maximum, while Cycle 26 will have a lower maximum than the preceding cycle. This could indicate that we are at present in the minimum of a centennial Gleissberg cycle.

The predicted dates for S_N maxima in Solar Cycles 25 and 26 (July 2024 and September 2034, respectively) are similar to those obtained for the next two polar flux polarity reversals for Solar Cycles 25 and 26 (April 2024 and August 2034, respectively).

Moreover, the S_N peak values for Cycles 25 and 26 have also been estimated using the equation of the linear regression resulting from the known relationship between the absolute

value of the difference between the polar fluxes of both hemispheres at an S_N minimum and the S_N maximum of the subsequent cycle, resulting in values of 142.3 ± 34.2 and 126.9 ± 34.2 , respectively. Hence, it is noteworthy that, using two different approaches, we produce similar predictions for the S_N maximum amplitudes of Solar Cycles 25 and 26.

Finally, we would like to draw an analogy between the evolution of weather prediction and solar cycle prediction. Over the last century, theoretical meteorologists have focused their efforts on weather prediction using the physical equations of meteorological models, achieving truly astonishing results (Fleming 2016). However, recently, exceptional advances have been made in weather prediction using machine learning techniques, driven by observed or quasi-observed data (Lam et al. 2023). Similarly, solar physicists have made notable strides in recent decades in the theoretical modeling of the solar dynamo based on magnetohydrodynamic equations for predicting the solar cycle (Choudhuri 2015). Nevertheless, we firmly believe that the time has definitively come for machine learning techniques to predict the solar cycle, driven by observed or quasi-observed data. This approach proves to be a crucial milestone, and this work is a compelling example of it.

Author contributions J.-V.R., V.-M.S.-C., and I.R.-R. had the idea of the manuscript. J.-V.R. made the calculations included in the manuscript and wrote the manuscript. All the authors discussed the results and revised the document, proposing improvements and providing enriching opinions.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research was supported by the Economy and Infrastructure Counselling of the Junta of Extremadura through project IB20080 (co-financed by the European Regional Development Fund). A.J.P. Aparicio thanks Universidad de Extremadura and Ministerio de Universidades of the Spanish Government for the award of a postdoctoral fellowship “Margarita Salas para la formación de jóvenes doctores (MS-11)”.

Data availability The sources of the data used in this paper are provided within the manuscript.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ansor, N.M., Hamidi, Z.S., Shariff, N.N.M.: 2019, The impact on climate change due to the effect of global electromagnetic waves of solar flare and coronal mass ejections (CMEs) phenomena. *J. Phys. Conf. Ser.* **1298**(1), 012019. DOI.
- Aparicio, A.J.P., Carrasco, V.M.S., Vaquero, J.M.: 2023, Prediction of the maximum amplitude of solar cycle 25 using the ascending inflection point. *Solar Phys.* **298**, 100. DOI.
- Arlt, R., Vaquero, J.M.: 2022, Historical sunspot records. *Living Rev. Solar Phys.* **17**, 1. DOI.
- Asensio Ramos, A., Cheung, M.C., Chifu, I., Gafeira, R.: 2023, Machine learning in solar physics. *Living Rev. Solar Phys.* **20**(1), 4. DOI.
- Asikainen, T., Mantere, J.: 2022, Prediction of even and odd sunspot cycles. *Solar Phys.* DOI.
- Biswas, A., Karak, B.B., Kumar, P.: 2023, Exploring the reliability of polar field rise rate as a precursor for an early prediction of solar cycle. *Mon. Not. Roy. Astron. Soc.* **526**(3), 3994.

- Bothmer, V., Daglis, I.A.: 2007, *Space Weather. Physics and Effects*, Springer, Heidelberg.
- Breiman, L.: 2001, Random forests. *Mach. Learn.* **45**, 5.
- Camporeale, E.: 2019, The challenge of machine learning in space weather: nowcasting and forecasting. *Space Weather* **17**(8), 1166. DOI.
- Carrasco, V.M.S., Vaquero, J.M.: 2023, Solar cycle 25 will be a weak-moderate cycle: an update. *Res. Notes AAS* **7**, 162. DOI.
- Chatzistergos, T., Ermolli, I., Giorgi, F., Krivova, N.A., Puiu, C.C.: 2020, Modelling solar irradiance from ground-based photometric observations. *J. Space Weather Space Clim.* **10**, 45. DOI.
- Choudhuri, A.R.: 2015, *Nature's Third Cycle: A Story of Sunspots*. Oxford. DOI.
- Clette, F., Lefèvre, L.: 2016, The new sunspot number: assembling all corrections. *Solar Phys.* **291**, 2629. DOI.
- Clette, F., Svalgaard, L., Vaquero, J.M., Cliver, E.W.: 2014, Revisiting the sunspot number. A 400-year perspective on the solar cycle. *Space Sci. Rev.* **186**, 35. DOI.
- Clette, F., Lefèvre, L., Chatzistergos, T., Hayakawa, H., Carrasco, V.M.S., et al.: 2023, Re-calibration of the sunspot number: status report. *Solar Phys.* **298**, 44. DOI.
- Cristianini, N., Shawe-Taylor, J.: 2000, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge.
- Feynman, J., Ruzmaikin, A.: 2011, The Sun's strange behavior: Maunder minimum or Gleissberg cycle? *Solar Phys.* **272**, 351. DOI.
- Feynman, J., Ruzmaikin, A.: 2012, The centennial gleissberg cycle in space weather. *AIP Conf. Proc.* **1500**(1), 44. DOI.
- Feynman, J., Ruzmaikin, A.: 2014, The centennial Gleissberg cycle and its association with extended minima. *J. Geophys. Res. Space Phys.* **119**(8), 6027. DOI.
- Fleming, J.R.: 2016, *Investing Atmospheric Science. Bjerknes, Rossby, Wexler, and the Foundations of Modern Meteorology*, MIT Press, Cambridge.
- Gleissberg, W.: 1939, A long-periodic fluctuation of the sun-spot numbers. *Observatory* **62**, 158.
- Gleissberg, W.: 1945, Evidence for a long solar cycle. *Observatory* **66**, 123.
- Guo, W., Jiang, J., Wang, J.X.: 2021, A dynamo-based prediction of solar cycle 25. *Solar Phys.* **296**, 1. DOI.
- Haigh, J.D.: 2007, The Sun and the Earth's climate. *Living Rev. Solar Phys.* **4**, 2. DOI.
- Han, Y.B., Yin, Z.Q.: 2019, A decline phase modeling for the prediction of solar cycle 25. *Solar Phys.* **294**, 107. DOI.
- Howard, T.: 2014, *Space Weather and Coronal Mass Ejections*, Springer, New York.
- Javaraiah, J.: 2015, Long-term variations in the north-south asymmetry of solar activity and solar cycle prediction, III: prediction for the amplitude of solar cycle 25. *New Astron.* **34**, 54. DOI.
- Javaraiah, J.: 2023, Prediction for the amplitude and second maximum of solar cycle 25 and a comparison of the predictions based on strength of polar magnetic field and low-latitude sunspot area. *Mon. Not. Roy. Astron. Soc.* **520**(4), 5586. DOI.
- Jayalekshmi, G.L., Pant, T.K., Prince, P.R.: 2022, Sunspot-cycle evolution of major periodicities of solar activity. *Solar Phys.* **297**(7), 85. DOI.
- Kahler, S.W.: 1992, Solar flares and coronal mass ejections. *Annu. Rev. Astron.* **30**(1), 113.
- Karak, B.B.: 2023, Models for the long-term variations of solar activity. *Living Rev. Solar Phys.* **20**(1), 3.
- Karak, B.B., Nandy, D.: 2012, Turbulent pumping of magnetic flux reduces solar cycle memory and thus impacts predictability of the Sun's activity. *Astrophys. J. Lett.* **761**(1), L13.
- Kitiashvili, I.N.: 2020, Application of synoptic magnetograms to global solar activity forecast. *Astrophys. J.* **890**(1), 36. DOI.
- Koç, E., Koç, A.: 2022, Fractional Fourier transform in time series prediction. *IEEE Signal Process. Lett.* **9**, 2542. DOI.
- Kopp, G.: 2021, Science highlights and final updates from 17 years of total solar irradiance measurements from the Solar Radiation and Climate Experiment/Total Irradiance Monitor (SORCE/TIM). *Solar Phys.* **296**, 133. DOI.
- Krausmann, E., Andersson, E., Gibbs, M., Murtagh, W.: 2016, Space weather y critical infrastructures: findings and outlook. EUR 28237 EN. *Publications Office of the European Union* DOI.
- Kumar, P., Biswas, A., Karak, B.B.: 2022, Physical link of the polar field buildup with the Waldmeier effect broadens the scope of early solar cycle prediction: cycle 25 is likely to be slightly stronger than cycle 24. *Mon. Not. Roy. Astron. Soc. Lett.* **513**(1), 112. DOI.
- Kumar, P., Nagy, M., Lemerle, A., Karak, B., Petrovay, K.: 2021, The polar precursor method for solar cycle prediction: comparison of predictors and their temporal range. *Astrophys. J.* **909**(1), 87. DOI.
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirthsberger, P., Fortunato, M., Alet, F., et al.: 2023, Learning skillful medium-range global weather forecasting. *Science* **382**(6677), 1416.
- Lu, J.Y., Xiong, Y.T., Zhao, K., Wang, M., Li, J.Y., Peng, G.S., Sun, M.: 2022, A novel bimodal forecasting model for solar cycle 25. *Astrophys. J.* **924**(2), 59. DOI.

- McIntosh, S.W., Leamon, R.J., Egeland, R.: 2023, Deciphering solar magnetic activity: the (solar) Hale cycle terminator of 2021. *Front. Astron. Space Sci.* **10**, 1050523. DOI.
- McIntosh, S.W., Chapman, S., Leamon, R.J., Egeland, R., Watkins, N.W.: 2020, Overlapping magnetic activity cycles and the sunspot number: forecasting sunspot cycle 25 amplitude. *Solar Phys.* **295**, 163. DOI.
- Moustafa, S.S., Khodairy, S.S.: 2023, Comparison of different predictive models and their effectiveness in sunspot number prediction. *Phys. Scr.* **98**(4), 045022. DOI.
- Muñoz-Jaramillo, A., Vaquero, J.M.: 2019, Visualization of the challenges and limitations of the long-term sunspot number record. *Nat. Astron.* **3**(3), 205. DOI.
- Muñoz-Jaramillo, A., Sheeley, N.R., Zhang, J., DeLuca, E.E.: 2012, Calibrating 100 years of polar faculae measurements: implications for the evolution of the heliospheric magnetic field. *Astrophys. J.* **753**, 146.
- Mursula, K., Zieger, B., Vilppola, J.H.: 2003, Mid-term quasi-periodicities in geomagnetic activity during the last 15 solar cycles: connection to solar dynamo strength—to the memory of Karolen I. Paularena (1957–2001). *Solar Phys.* **212**, 201. DOI.
- Nandy, D.: 2021, Progress in solar cycle predictions: sunspot cycles 24–25 in perspective. *Solar Phys.* **296**, 54. DOI.
- Oshiro, T.M., Perez, P.S., Baranauskas, J.A.: 2012, *How Many Trees in a Random Forest? International Workshop on Machine Learning and Data Mining in Pattern Recognition*, Springer, Berlin, 154.
- Overbye, D.: 2021, Will the Next Space-Weather Season Be Stormy or Fair? *The New York Times*, New York.
- Peguero, J.C., Carrasco, V.M.S.: 2023, A critical comment on “Can solar cycle 25 be a new Dalton minimum?”. *Solar Phys.* **298**, 48. DOI.
- Petrovay, K.: 2020, Solar cycle prediction. *Living Rev. Solar Phys.* **17**, 2. DOI.
- Prasad, A., Roy, S., Sarkar, A., Panja, S.C., Patra, S.N.: 2023, An improved prediction of solar cycle 25 using deep learning based neural network. *Solar Phys.* **298**(3), 50. DOI.
- Pulkkinen, T.: 2007, Space weather: terrestrial perspective. *Living Rev. Solar Phys.* **4**, 1. DOI.
- Ridley, P.: 2023, On the strength and duration of solar cycle 25: a novel quantile-based superposed-epoch analysis. *Solar Phys.* **298**, 66. DOI.
- Rodríguez, J.V., Rodríguez-Rodríguez, I., Woo, W.L.: 2022, Machine learning-based prediction of sunspots using Fourier transform analysis of the time series. *Publ. Astron. Soc. Pac.* **134**(1042), 124201. DOI.
- Schrijver, C.J., Kauristie, K., Aylward, A.D., Denardini, C.M., Gibson, S.E., Glover, A., et al.: 2015, Understanding space weather to shield society: a global road map for 2015–2025 commissioned by COSPAR and ILWS. *Adv. Space Res.* **55**(12), 2745. DOI.
- Schwenn, R.: 2006, Space weather: the solar perspective. *Living Rev. Solar Phys.* **3**(1), 1. DOI.
- Seeger, M.: 2004, Gaussian processes for machine learning. *Int. J. Neural Syst.* **14**(02), 69.
- Shmueli, G., Lichtendahl, K.C. Jr.: 2016, *Practical Time Series Forecasting with R: A Hands-on Guide*, Axelrod Schnall Publishers, Green Cove Springs.
- Solanki, S.K., Krivova, N.A.: 2003, Can solar variability explain global warming since 1970? *J. Geophys. Res. Space Phys.* **108**(A5), 1200. DOI.
- Solanki, S.K., Usoskin, I.G., Kromer, B., Schüssler, M., Beer, J.: 2004, Unusual activity of the Sun during recent decades compared to the previous 11,000 years. *Nature* **431**(7012), 1084. DOI.
- Su, X., Liang, B., Feng, S., Dai, W., Yang, Y.: 2023, Solar cycle 25 prediction using N-BEATS. *Astrophys. J.* **947**(2), 50. DOI.
- Syed Zafar, S.N.A., Umar, R., Sabri, N.H., Jusoh, M.H., Dagang, A.N., Yoshikawa, A.: 2021, Effects of solar flares and coronal mass ejections on Earth’s horizontal magnetic field and solar wind parameters during the minimum solar cycle 24. *Mon. Not. Roy. Astron. Soc.* **504**(3), 3812. DOI.
- Temmer, M.: 2021, Space weather: the solar perspective: an update to Schwenn (2006). *Living Rev. Solar Phys.* **18**(1), 4. DOI.
- Tobias, S.: 2023, Turbulent times for the Sun’s magnetic field. *Nat. Astron.* **7**, 644. DOI.
- Upton, L.A., Hathaway, D.H.: 2018, An updated solar cycle 25 prediction with AFT: the modern minimum. *Geophys. Res. Lett.* **45**(16), 8091. DOI.
- Upton, L.A., Hathaway, D.H.: 2023, Solar cycle precursors and the outlook for cycle 25. *J. Geophys. Res. Space Phys.* **128**, e2023JA031681. DOI.
- Usoskin, I.G.: 2023, A history of solar activity over millennia. *Living Rev. Solar Phys.* **20**, 2. DOI.
- Usoskin, I.G., Mursula, K., Solanki, S.K., Schüssler, M., Kovaltsov, G.A.: 2002, A physical reconstruction of cosmic ray intensity since 1610. *J. Geophys. Res. Space Phys.* **107**(A11), SSH-13. DOI.
- Usoskin, I.G., Solanki, S.K., Krivova, N.A., Hofer, B., Kovaltsov, G.A., Wacker, L., et al.: 2021, Solar cyclic activity over the last millennium reconstructed from annual 14C data. *Astron. Astrophys.* **649**, A141.
- Vapnik, V.: 2013, *The Nature of Statistical Learning Theory*, Springer, Berlin.
- Vaquero, J.M., Svalgaard, L., Carrasco, V.M.S., Clette, F., Lefèvre, L., Gallego, M.C., Arlt, R., Aparicio, A.J.P., Richard, J.-G., Howe, R.: 2016, A revised collection of sunspot group numbers. *Solar Phys.* **291**, 3061. DOI.
- Versteegh, G.J.: 2005, Solar forcing of climate. 2: evidence from the past. *Space Sci. Rev.* **120**, 243. DOI.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

José-Víctor Rodríguez^{1,2}  · Víctor Manuel Sánchez Carrasco^{3,4}  ·
Ignacio Rodríguez-Rodríguez⁵  · Alejandro Jesús Pérez Aparicio^{3,4,6}  ·
José Manuel Vaquero^{3,4} 

✉ J.-V. Rodríguez
jvictor.rodriguez@upct.es; josevictor.rodriguez@edu.uah.es

- ¹ Departamento de Tecnologías de la Información y las Comunicaciones, Universidad Politécnica de Cartagena, 30202, Cartagena, Spain
- ² Department of Computer Engineering, Universidad de Alcalá, 28801, Alcalá de Henares, Madrid, Spain
- ³ Departamento de Física, Universidad de Extremadura, 06006 Badajoz, Spain
- ⁴ Instituto Universitario de Investigación del Agua, Cambio Climático y Sostenibilidad (IACYS), Universidad de Extremadura, 06006 Badajoz, Spain
- ⁵ Departamento de Ingeniería de Comunicaciones, Universidad de Málaga, 29071, Málaga, Spain
- ⁶ Earth Remote Sensing Laboratory (EaRSLab) and Center for Sci-Tech Research in EArth sysTem and Energy-CREATE, Instituto de Investigação e Formação Avançada (IIFA), Universidade de Évora, 7000-671 Évora, Portugal