Extreme value theory applied to long-term sunspot areas

RUI ZHANG[1,2,3], YAN-QING CHEN[1,2], SHU-GUANG ZENG[1,2,*] , SHENG ZHENG[1,2], YAN-SHAN XIAO[1,2], LIN-HUA DENG[4,5], XIANG-YUN ZENG[1,2] and YAO HUANG[1,2]

[1]Center of Astronomy and Space Sciences, China Three Gorges University, Yichang 443002, People's Republic of China.
[2]College of Science, China Three Gorges University, Yichang 443002, People's Republic of China.
[3]Yunnan Key Laboratory of the Solar Physics and Space Science, Kunming 650216, People's Republic of China.
[4]School of Mathematics and Computer Science, Yunnan Minzu University, Kunming 650504, People's Republic of China.
[5]Yunnan Observatories, Chinese Academy of Sciences, Kunming 650216, People's Republic of China.
*Corresponding author. E-mail: zengshuguang19@163.com

Abstract. Solar activity, such as sunspots and flares, has a great impact on humans, living beings, and technologies in the whole world. Changes in sunspots will influence high-frequency and space-navigation radio communications. Based on the full-disk, southern and northern hemispheres sunspot areas (SAs) data in 1874–2023 from the Royal Observatory, Greenwich (RGO) USAF/NOAA, extreme value theory (EVT) is applied to predict the trend of the 25th and 26th solar cycles (SCs) in this work. Two methods with EVT, the block maxima (BM) approach and the peaks-over-threshold (POT) approach, are employed to research solar extreme events. The former method focuses on each block's maximum sunspot areas value and is applied for the generalized extreme value (GEV) distribution. The latter method aims to select the extreme values exceeding a threshold value and is used to obtain the generalized Pareto (GP) distribution. It is the first time that the EVT is applied on the sunspot areas data from the Royal Observatory, Greenwich (RGO) USAF/NOAA. The analysis indicates that the estimated 8-year return levels for sunspot areas are 5701 and 6258 using the two methods, while the estimated 19-year return levels are all 7165. This suggests that the trends of the 25th and 26th solar cycles will be stronger than that of the 24th solar cycle.

Keywords. Sun: activity—methods: data analysis—sunspots.

1. Introduction

Solar activity is made of sunspots, coronal mass ejections (CMEs), filament eruptions, and flares (Chowdhury *et al.* 2021). The ebb and flow of solar activity over an 11-year cycle has important implications for high-frequency radio communications, space climate and navigation (Pala & Atici 2019). In space climate and solar physics, it is very important to predict the maximum amplitude of the following solar cycle (SC) (Kitiashvili 2021), which may help reveal the dynamical mechanism of the cycle (Babcock 1961) and plan future space missions (Kumar *et al.* 2022). Sunspots are the earliest observed solar activity and solar magnetic fields. The solar magnetic field and its interaction with solar convection drive and modulate solar activity (Bhowmik *et al.* 2023; Karak 2023).

Sunspot area (SSA) is a valuable indicator of the Sun's magnetic activity over extended periods. This metric can be effectively employed to forecast solar cycles, similar to the utilization of sunspot numbers (SSN) for this purpose (Hathaway 2015). Sunspot area stands as one of the longest observed indices of solar activity to date, rendering it almost equivalent to physical significance compared to sunspot numbers. Therefore, records of sunspot areas are important for understanding the long-term behavior of solar magnetic activity and variability (Mandal *et al.* 2020).

In general, solar activity prediction typically employs three methodologies: precursor method, model-based

method, and extrapolation method (Petrovay 2020). The precursor method, as the first approach, relies on distinct measures of solar activity or magnetism at a particular period to predict the amplitude of the upcoming solar maximum. There is a strong positive correlation between the rising rate and the peak sunspot number (Cameron & Schüssler 2008; Karak & Choudhuri 2011). Kumar *et al.* (2022) predicted the amplitude of the 25th solar cycle by using the rising rate of the sunspot cycle and found the amplitude of the 25th cycle is 138 ± 26, which will be slightly stronger than the 24th solar cycle. This comparison also affirms the accuracy of this method. The second model-based method generally entails inputting observational solar data into diverse physical dynamical models to predict solar activity. It can be further divided into two categories: predictions based on consistent dynamo models and surface flux transport models (Petrovay 2020). For instance, Schatten *et al.* (1978) employed dynamo theory to predict the sunspot number during solar cycle 21, estimating the yearly mean sunspot number at the solar maximum to be 140 ± 20. Dikpati *et al.* (2006) effectively utilized a flux transport-based dynamo tool to predict the intensity of the 24th solar cycle. Choudhuri *et al.* (2007) incorporated solar polar magnetic field data into a solar dynamo model, enabling the simulation of the most recent solar cycles. Applying a surface flux transport (SFT) model, Bhowmik & Nandy (2018) utilized a combination of an observational data-driven surface flux transport model and a dynamo model, providing ensemble predictions of solar cycle 25, and were also successful in reproducing the past eight observed solar cycles. Jiang & Cao (2018) computed correlations between key solar cycle properties, effectively forecasting upcoming cycle trends. Contrastingly, extrapolation methods, the last one mentioned, are founded on the assumption that the physical process responsible for the sunspot number record is statistically uniform. This implies that the mathematical consistencies governing its fluctuations remain consistent across all time points (Petrovay 2020). Extrapolation methods can be divided into the following categories: non-linear model-based forecast, statistical forecast, spectral methods-based forecast, machine learning and neural network-based forecast (Nandy 2021). Such as, utilizing spectral component extrapolation, Rigozo *et al.* (2011) assessed the intensity of the 25th solar cycle. Attia *et al.* (2013) introduced the Neuro–Fuzzy model for analyzing sunspot time series and predicting the sunspots for the 24th and 25th solar cycle. Applying a Bayesian approach, Noble & Wheatland (2012) projected solar cycles, while Sarp *et al.* (2018a) employed a nonlinear method to predict the 25th solar cycle.

As a branch of statistics, extreme value theory (EVT) has played an important role in various science fields in recent times, such as climate change (Nogaj *et al.* 2006; Coelho *et al.* 2008; Acero *et al.* 2011, 2014), economics (Gilli & Keellezi 2006), engineering (Castillo *et al.* 2004), and especially astrophysics (Bhavsar & Barrow 1985; Bernstein & Bhavsar 2001; Deng *et al.* 2020). In these fields, extreme events are rare occurrences that can produce dramatic and severe changes. Due to its importance and specificity, EVT offers techniques for studying and estimating the probability of predicting rare events (Coles 2001). Solar activity, as a part of astronomy, can also experience rare extreme events. For instance, although the typical duration of the solar activity cycle is approximately 11 years, the most probable return time for a large event such as the maximum at solar cycle 19, happens once every about 700 years and the probability of finding (Asensio Ramos 2007). Additionally, considerations must be made for the Maunder minimum period, while hardly any sunspots were observed. Consequently, the applicability of the EVT has been extended to the study of solar sunspot data.

In 2017, EVT was applied to observe the new sunspot number index at three different timescales, and the results showed that sunspot numbers are hardly >550 in the coming years (Acero *et al.* 2017). Elvidge & Angling (2018) used EVT to determine the probability of Carrington-like solar flares and they found that the return level (RL) of the expected 150 years was approximately an $\times 60$ flare. EVT was employed to focus on extreme solar flare events to calculate the return levels of Carrington or Halloween storm-like events (Tsiftsi & De la Luz 2018). Acero *et al.* (2018a) used 10Be and 14C time series to study solar activity with EVT. The extreme events law was estimated with the solar radio of flux at the daily scale, and the results suggested that there was a bound for the index (Acero *et al.* 2018b) in 2019. Besides, four projects to predict and mitigate the harmful effects of extreme space weather storms in 2020 employed EVT statistical model to examine the intensities of the magnetic superstorms recorded in the disturbance storm time (Dst) index time series, and the result showed that the choice of data and statistical model could significantly affect the extrapolation probabilities of extreme events (Love 2020). In 2021, the sunspot number series from the Purple Mountain Observatory was studied with EVT to conclude that the trend of the 25th solar cycle will be stronger than the 24th solar cycle (Chen *et al.* 2021).

Extreme value theory has been used to study the data from sunspots, sunspots umbrae and faculae (Willis & Tulunay 1979). However, no one has employed EVT

on sunspot areas to study solar activity. To freshen up the research, EVT is first applied to study the sunspot areas in this work, examining data from three scopes: the entire disk, the southern hemisphere, and the northern hemisphere. The daily sunspot areas dataset used in this paper are from the Royal Observatory, Greenwich (RGO) USAF/NOAA during May 1874 to July 2023. In Section 2, the data and methods are introduced. We present the results from our experimental analyses in Section 3. Finally, our concluding discussion is given in Section 4.

2. Data and methods

2.1 *Data*

In this work, the daily sunspot areas dataset for the period of May 1874 to July 2023 are from the Royal Observatory, Greenwich (RGO) USAF/NOAA. The website, http://solarcyclescience.com/activeregions. html, covering 149 years of observations. The statistical analysis of sunspot areas data is important for understanding solar activities and their impact on the Earth. The long-term evolution of the sunspot areas during this period is shown in Figure 1.
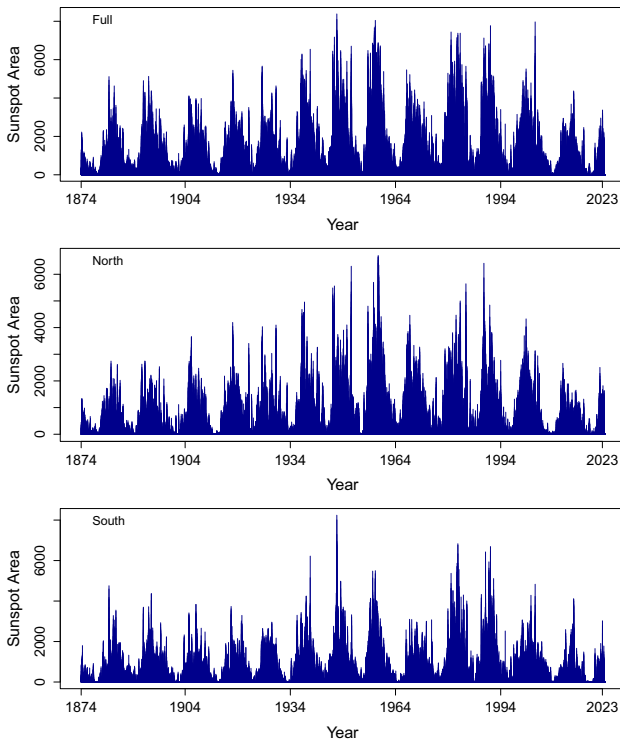

Figure 1. Distribution of the daily sunspot areas dataset during the period of May 1874 to July 2023. These three figures correspond to datasets of the full disk, the northern and southern hemispheres.

2.2 *Methods*

Extreme value theory is a branch of statistics that provides a tool for estimating the probability of events outside the observed range and characterizes the probability of random processes (Coles 2001).

Two key methods for studying extreme events are the block maxima (BM) approach and the peaks-over-threshold (POT) approach. The former achieves the generalized extreme value (GEV) distribution and obtains the generalized Pareto (GP) distribution. In the BM approach, the observed data are divided into some blocks with equal size n, and the maximum value of each block is obtained. Next, extreme sample points are fitted to GEV distribution, a type of law of large numbers with a probability distribution. When the sample size is sufficiently large, GEV is used to model the distribution of the maximum (or minimum) values from a sample of random variables. In the realm of extreme value theory, it is commonly used to represent the ultimate distribution of the maxima in a sequence of independent and identically distributed random variables (Coles 2001; Acero *et al.* 2017). The probability distribution function of the GEV distribution is expressed as follows:

$$\text{GEV}(x; \mu, \delta, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]_+^{-\frac{1}{\xi}}\right\}. \tag{1}$$

It should satisfy $-\infty < \mu, \xi < +\infty, \sigma > 0$, while μ is the location parameter, σ is the scale parameter and ξ is the shape parameter.

As for the POT approach, the value of u for the upper threshold is set and sample values over it are extracted. Then, extreme sample points are fitted to the GP distribution. More details about GP distribution can be found in Coles (2001), Acero *et al.* (2017, 2018a). A brief description follows. In the asymptotic limit for sufficiently large thresholds, the distribution function of $(X - u)$, conditional on $X > u$, is approximately as follows:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{\frac{-1}{\xi}}, \tag{2}$$

defined on $\{y: y > 0\}$ and

$$\left\{y: \left(1 + \frac{\xi y}{\tilde{\sigma}}\right) > 0\right\},$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$ with σ is the scale parameter and ξ is the shape parameter ($\xi \neq 0$).

The choices of block length in the BM approach and threshold value in the POT approach should

be objective, ensuring a balance between bias and variance (Tsiftsi & De la Luz 2018). The dataset are daily sunspot areas, so a block length of one year is set in BM. The stability assessment of the parameter estimates and the mean residual life plot are two ways to choose the threshold in the POT approach.

Some differences exist between the two methods. One is the defined way of the extreme events (Coles 2001; Acero *et al.* 2017; Tsiftsi & De la Luz 2018; Acero *et al.* 2018a). Another difference is the utilization of data. For the BM approach, the choice of the block length could result in a large amount of wasted data when the block length is large. However, the POT approach can allow for effective use of data, which will not waste valuable extreme data. Meanwhile, strong links exist in two approaches. As the block maximum is fitted to the GEV distribution, exceedances over the threshold will be fitted to the GP distribution. The shape parameters of these two distributions are expected to be asymptotically identical (Coles 2001; Tsiftsi & De la Luz 2018).

3. Results

3.1 *Block maxima approach*

Here, we take the full-disk data for the prediction of the 25th solar cycle as an example to illustrate the methods and workflow of data processing in the prediction process. Following the below steps, first of all, the dataset are cut into some blocks, all of which have equal length. Given the use of daily sunspot areas, the value for the block size is approximately one year. A series of values are tried for the block length: 360–370 are tried, and the value of 365 is the best choice, considering higher estimation efficiency. According to the Figure 2, it can be found that the sunspot areas data are divided into 150 blocks with vertical red lines, indicating that there are 150 maxima values.

Then 150 maxima values are fitted to the GEV distribution. In the fitting process, the maximum likelihood estimation (MLE) method is used to estimate three parameters for the GEV distribution: μ (location parameter), σ (scale parameter), and ξ (shape parameter). In Figure 3, three standard diagnostic plots are utilized for estimating the accuracy of GEV model. The quantile–quantile (QQ) plot compares the empirical quantiles of 150 sample points with GEV model quantiles and confirms the validity of the GEV model in the top panel (a), as the plotted point set is close to linear. The QQ-plot compares randomly generated data
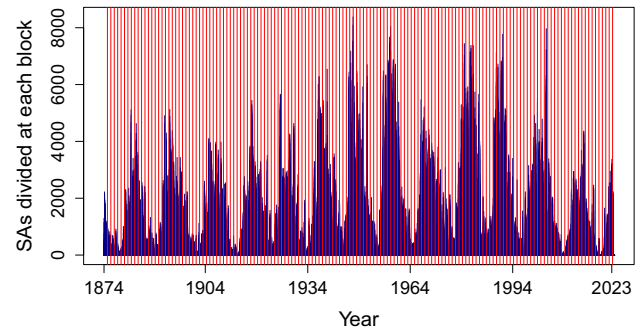

Figure 2. The full-disk sunspot areas (SAs) cut into 150 blocks with the BM approach. The vertical axis represents the daily sunspot areas. Each pair of red lines is separated by approximately 1 year.

from GEV distribution against the empirical quantiles. In middle panel (b), we can observe that all 150 sample points are uniformly distributed within the 95% confidence bands, displaying a linear relationship, which confirms the validity of the model. The shape of the empirical density line is similar to the model density line in the bottom panel (c). Therefore, three diagnostic plots provide support for the GEV model.

The next step is to use the bootstrap technique to obtain the fitting results. Return level is a crucial property predicting extreme events in the future, meaning that the N-year return level is the level expected to be surpassed, on average, once every N years.

To study the trend of the solar activity of the 25th solar cycle, which is started from 2020 to 2031, $N = 8$ years is suitable to be set for the last year of this solar cycle, as our data covers the period from 1874 to 2023 ($2023 + 8 = 2031$). That is to say, in the future, 8 years after 2023, the return level value of sunspot areas can be obtained by bootstrap. Similarly, for the 26th solar activity cycle from 2031 to 2042, $N = 19$ could be set, representing the future 19 years after 2023. In Table 1, there are three parameters of GEV distribution and their 95% confidence intervals (CIs). It is seen that the shape parameter is negative, presenting that an upper bound exists in the distribution and that there is a maximum extreme value. Table 2 presents the estimates of the return level of 8 years and 19 years. Combining the Figure 4 with Table 2, the return level for $N = 8$ and $N = 19$ is about 5701 and 7165, showing that the maximum value of sunspot areas is about 5701 and 7165 for the 25th and 26th solar cycle. Comparing with the amplitude of cycle 24, the trend of both return levels is upward. The estimates of the return level for the southern and northern hemispheres also confirm this point, as seen in the estimated values in Table 2.
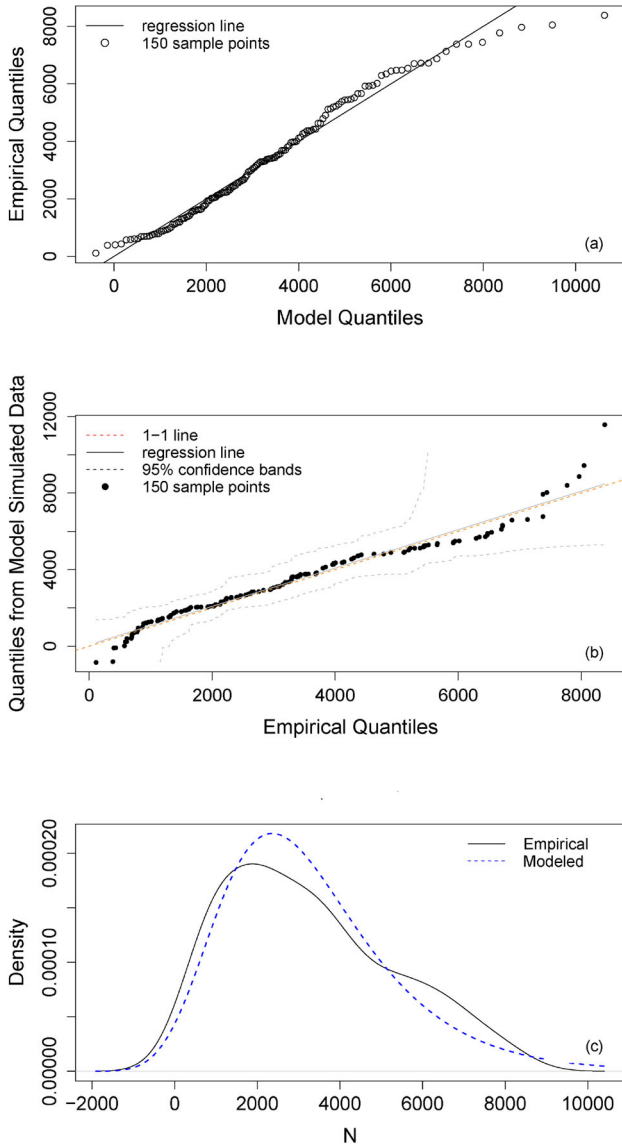
Figure 3. The GEV distribution has three diagnostic plots of full-disk sunspot area. The sample points represent the maximum value of each block. (a) The QQ-plot compares the empirical quantiles of 150 sample points with the GEV model quantiles. (b) The QQ-plot compares the randomly generated data against the empirical data quantiles with the 95% confidence bands (black dashed line). (c) The plot shows that the empirical density line of the observed maximum (black solid line) and GEV model density line (blue dashed line) are close to coinciding.

3.2 *Peaks-over-threshold approach*

Although the BM method is theoretically simple and easy to implement and is better suited for situations where data is not completely independent and identically distributed, it still has a drawback of inefficient data utilization (Ferreira & de Haan 2013).

Therefore, the POT method is considered the fundamental approach for extreme value analysis because of efficient utilization (Chen *et al.* 2021). The POT method picks up all 'relevant' high observations. Choosing u, the value of a threshold is vital in this process. The choice of threshold is necessary but difficult; u cannot be too high or too low. As the analysis methods for these three datasets relying on POT are identical, we have chosen the full-disk data as an example to reveal the detailed analysis process. Figure 5 shows the mean residual life of full-disk data. To determine the most suitable threshold, it is necessary to identify a stable interval from Figure 5 for the experiment. It can be observed that the image between [1000, 5000] is stable. As the confidence interval becomes larger, there seems to be a linear relationship in the interval [3000, 4500]. There is the stability of the parameter estimation, which is fitted by the GP distribution over a range of thresholds. In Figure 6, the lines of the shape parameter and reparameterized scale parameter are stable when u is set to 3300. Besides, other values for u are tried, and the best fit is achieved while $u = 3300$. Employing a similar analysis, we have determined the threshold values for the southern and northern hemispheres data, resulting in $u = 2135$ and $u = 2412$, respectively.

Due to the clustering of extreme values in the sunspot area, there are often consecutive days with maximum values exceeding the threshold, making it difficult to obtain approximately independent clusters of extreme observations. Therefore, these potential exceedance clusters must be processed with a declustering procedure to avoid short-term dependencies in the time series. The method called 'runs declustering'. This method involves marking exceedances as belonging to the same cluster when the gap between them is less than a fixed number of observations, referred to as the run length (Acero *et al.* 2017). Figure 7 shows the data above a threshold u. Then, we can get the independent threshold exceedances after the declustering procedure. The run length value is set to 13 because the solar global rotation is 27, and half is 13.5 (Heristchi & Mouradian 2009). During the declustering process, exceedances are marked as one cluster if the exceedances are separated by less than a number of run lengths. Here, we take the full-disk data as an example. The number of new independent threshold exceedance is 221. Then 221 exceedance is applied to the GP distribution. The method of estimating the parameters is also the MLE and the estimations of the two parameters σ, ξ of the GP distribution are estimated as 1311.47 and -0.07. The same analysis method is applied to the northern and southern hemispheres, and the resulting parameters are

Table 1. Estimates and their 95% confidence intervals (CIs) of three GEV parameters obtained by bootstrapping with the BM approach. The data includes full-disk, southern hemisphere, and northern hemisphere results.

Data	Location (μ) [95% CI]	Scale (σ) [95% CI]	Shape (ξ) [95% CI]
Full	2344.47 [2026.30, 2662.69]	1685.97 [1421.32, 1950.63]	-0.008 [-0.20, 0.18]
South	2016.79 [1671.34, 2362.23]	1356.07 [1088.63, 1623.52]	-0.01 [-0.20, 0.18]
North	2089.58 [1731.98, 2447.18]	1328.22 [1062.57, 1593.87]	-0.05 [-0.27, 0.17]

Table 2. Estimate and its 95% confidence interval of the 8 year and 19 year return levels (RL) by bootstrapping with BM approach for the three datasets.

Data	8 year-RL	19 year-RL
Full	5700.65 (5192.53, 6178.53)	7164.88 (6373.87, 8023.58)
South	4707.05 (4251.71, 5208.14)	5839.86 (4949.56, 6796.80)
North	4600.94 (4104.91, 5046.86)	5671.08 (4899.87, 6484.26)

shown in Table 3. In Figure 8, the precision of the GP model evaluated by standard diagnostic plots for full-disk data are shown. These three plots also support the accuracy of the GP model.

In Figure 9, the confidence interval of the return level for $N = 8$ and $N = 19$ years is obtained by bootstrapping with a confidence of 95%. Table 3 shows the data by bootstrapping with the two different parameters of GP distribution. The shape parameter and the 95% confidence interval are negative. Combining the results of three standard diagnostic plots in Figure 8 and these two parameters of the full-disk data in Table 3, it can be concluded that there is a good GP distribution. Table 4 lists the estimates of the return level for the daily time series with $N = 8$ and $N = 19$ years, indicating that the maximum of sunspot area value is about 6258 and 7165 for 25th and 26th solar cycle. Compared to the strength of the 24th solar cycle, the solar activity is stronger during the 25th and 26th solar cycle.

4. Discussions and conclusions

Extreme value theory is commonly used in astrophysics as an excellent statistical tool for estimating rare events. To predict the solar activity during the 25th and 26th solar cycles, EVT (with two different methods: BM and POT) is used to analyze the full-disk, southern and northern hemispheres daily sunspot areas from the Royal Observatory, Greenwich (RGO) USAF/NOAA during May 1874 to July 2023. It is the first time EVT is employed to study the solar activity with sunspot areas observations.

In the BM approach, the return levels were estimated by obtaining the maximum block data values and fitting them to the GEV distribution. According to the results in Table 1, it can be seen that shape parameters are negative, which means that the existence of the extreme upper bound of time series is revealed. The return levels for $N = 8$ years were estimated as 5701, 4707 and 4601, respectively (see Table 2) applying to data from the full-disk, southern hemisphere and northern hemisphere. This implies that the maximum values for the 25th solar cycle are approximately 5701, 4707 and 4601. The return levels for $N = 19$ years, which is approximately the maximum values for the 26th solar cycle, were estimated as 7165, 5840 and 5671, respectively, applying to the full, south and north data. These results we can conclude that the 25th and 26th solar cycles will be stronger than cycle 24. In addition, the maxima in the northern and southern hemispheres add up to a much larger maximum than the full-disk data, suggesting that the peak activities in the northern and southern hemispheres are not within the same time frame. It is possible, although less likely, that this phenomenon could result in a double peak in the 25th and 26th solar cycles (Karak *et al.* 2018).

According to the application of the POT approach the mean residual life plot and the parameter estimates were used to obtain the best threshold u. Hence, we selected the thresholds $u = 3300$, 2135 and 2412 for the data from the full-disk, southern and northern hemispheres, respectively. This approach results in a negative shape parameter, indicating an upper bound in the distribution (see Table 3). The 8-year return levels were estimated as
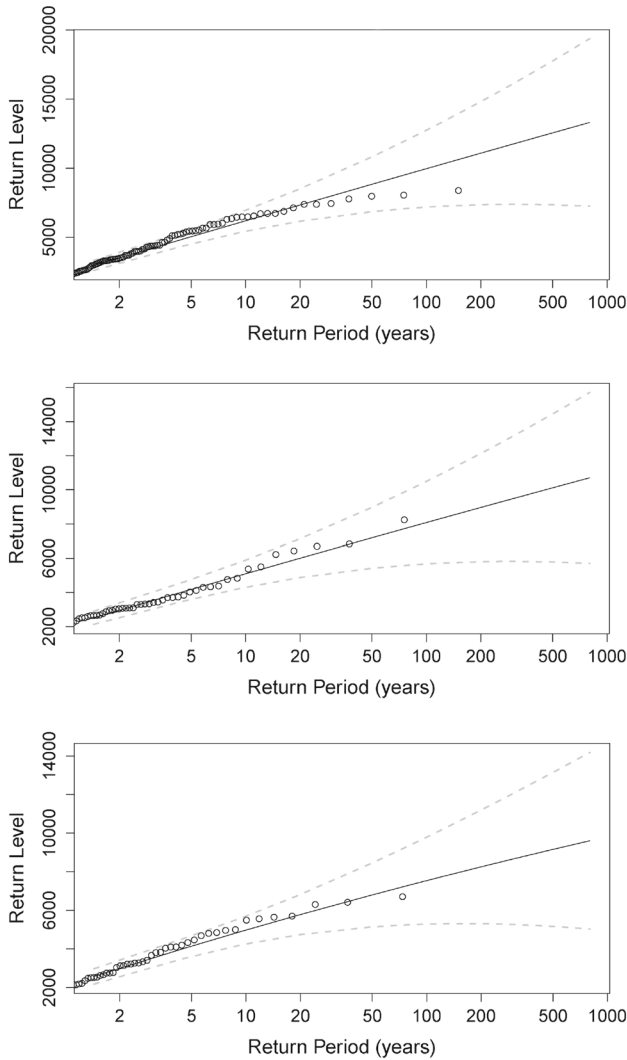
Figure 4. Return level plot (log scale) of the maxima values for full-disk (top), southern hemisphere (middle) and northern hemisphere (bottom) daily sunspot areas with 95% confidence intervals (dashed lines).
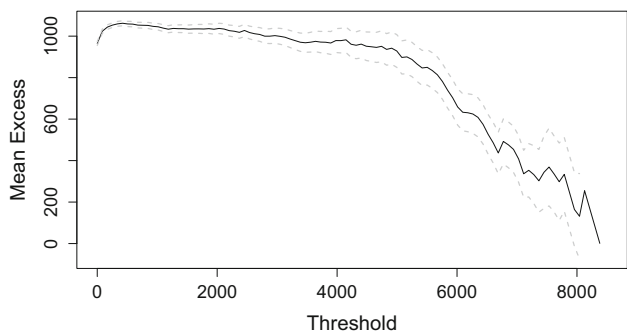

Figure 5. Mean residual life plot for the full-disk daily sunspot areas. The graph allows a relatively stable threshold interval to be found, making it easier to determine the final threshold. The interval seems to have a linear relationship [3000, 4500].
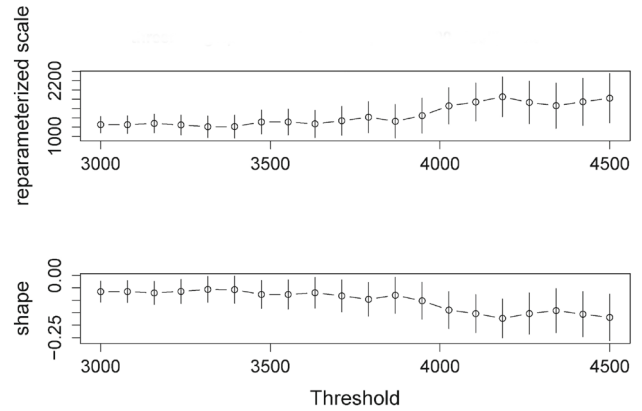

Figure 6. Estimates of two parameters (shape and scale) with different thresholds using full-disk data. There is a linear relationship of about 3300. The selection of the threshold range is based on observations from Figure 5 and multiple experiments.
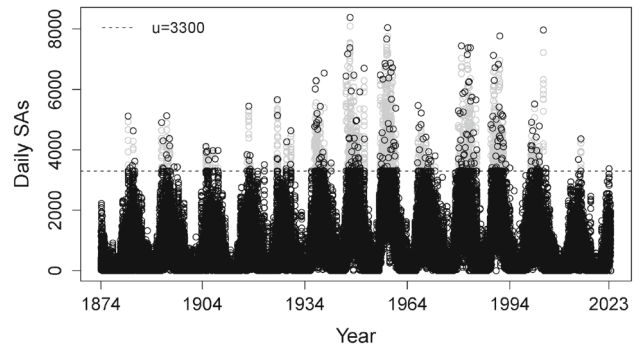
Figure 7. Decluster data above a threshold: $u = 3300$, and newly independent exceedances are produced. The clear points above the dashed line are the values after the declustering procedure.

6258, 4501 and 4344 for the full-disk, south and north data, while the 19-year return levels were estimated as 7165, 5363 and 5086 for the three datasets, respectively. The results of the POT technique are generally higher than the BM, indicating that the 25th and 26th solar cycles are stronger than the 24th solar cycles.

As discussed in Bhowmik *et al.* (2023), the solar cycle could be understood as a weakly nonlinear limit cycle influenced by stochastic noise. These models have effectively replicated the characteristics of the solar cycle, including the occurrence of grand minima, when observed over specific time scales. The predictability of the solar cycle is jointly influenced by stochasticity and nonlinearity, making long-term predictions notably challenging. The solar activity cycle is inherently predictable; however, due to its intrinsically dynamical complexity, it can only be predicted with high accuracy

Table 3. Estimates and their 95% confidence intervals (CIs) of two GP parameters obtained by bootstrapping with the POT approach for the three datasets.

Data	Scale (σ) [95% CI]	Shape (ξ) [95% CI]
Full	1311.47 [1027.67, 1595.26]	-0.07 [-0.24, 0.11]
South	1023.29 [819.24, 1227.33]	-0.001 [-0.14, 0.14]
North	902.03 [680.18, 1123.88]	-0.001 [-0.19, 0.18]
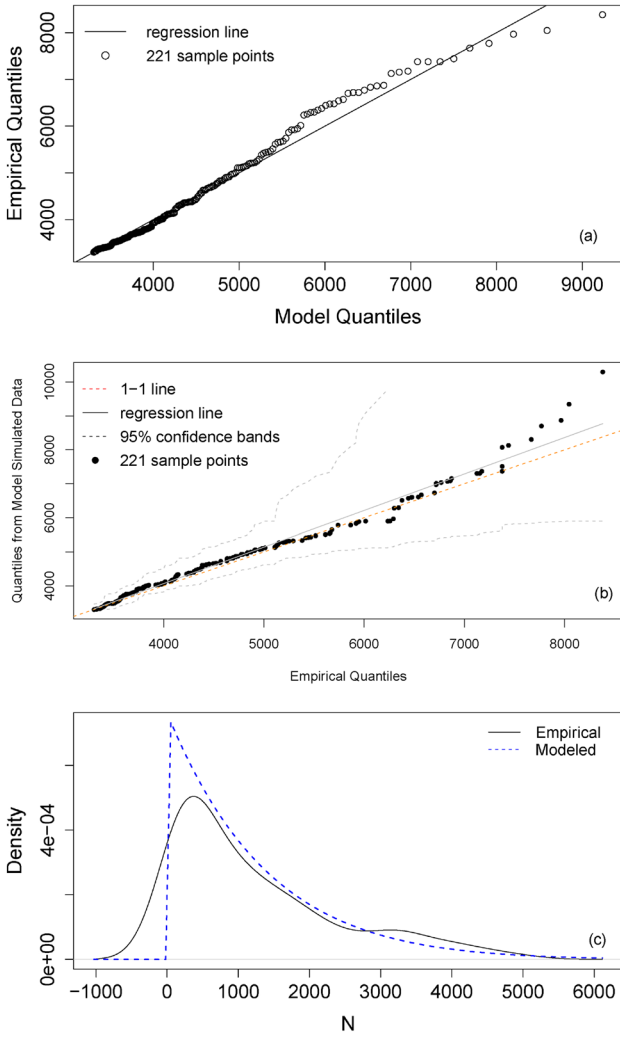
Figure 8. There are three diagnostic plots in the GP distribution for the full-disk data. (a) The QQ-plot compares the empirical quantiles of 150 sample points with the GP model quantiles. (b) The QQ-plot shows the randomly generated data against the empirical data quantiles with the 95% confidence bands (black dashed line). (c) The plot shows that the empirical density line of the observed maximum (black solid line) and GP model density line (blue dashed line) are close to coinciding.
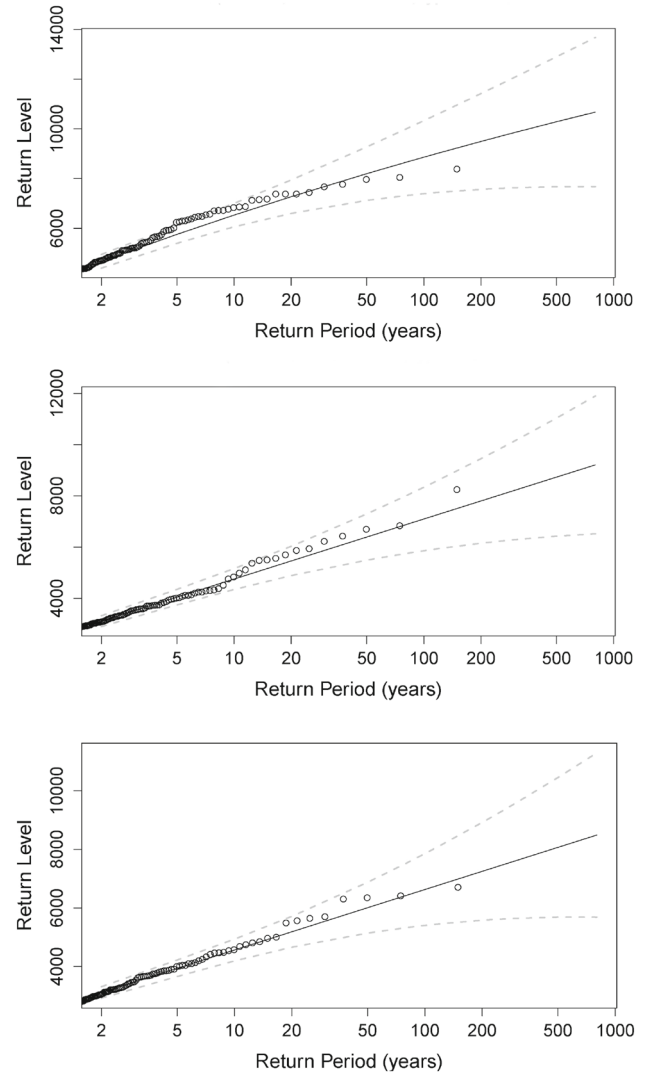

Figure 9. Return level plot (log scale) of the maxima values for full-disk (top), southern hemisphere (middle) and northern hemisphere (bottom) daily sunspot areas with 95% confidence intervals (dashed lines).

for the short- to mid-term (Deng *et al.* 2016). There exist differences between the predictions of the 25th solar cycle made by various data and methods. On the one hand, spectral analysis was used on sunspot numbers to predict the amplitude of the 25th solar cycle, and the result showed that the amplitude would be lower than the amplitude of the 24th solar cycle (Javaraiah

Table 4. Estimate and its 95% confidence interval of the 8 year return level (RL) by bootstrapping with POT approach for the three datasets.

Data	8 year-RL	19 year-RL
Full	6258.12 (5968.08, 6554.83)	7164.92 (6639.36, 7680.65)
South	4500.89 (4246.05, 4761.29)	5363.01 (4849.28, 5959.76)
North	4343.53 (4124.74, 4564.32)	5086.11 (4669.88, 5595.11)

2015). Gopalswamy *et al.* (2018) used microwave imaging on 17 GHz data and predicted that the two cycles are similar. On the other hand, the results of many other works are consistent with our results. For example, Sarp *et al.* (2018b) predicted that the solar maximum of the 25th solar cycle was greater than that of the 24th solar cycle. Prasad *et al.* (2022) used the LSTM model to predict the strength of the 25th solar cycle by using the monthly smoothed sunspot number from the Royal Observatory of Belgium, Brussels, suggesting that the upcoming 25th solar cycle is stronger. The works mentioned above pointed towards our conclusion: solar cycle 25 will be stronger than cycle 24.

As mentioned above, predictions for solar activity are important. In addition to the distributions used in the aforementioned work, extreme value theory encompasses various other distributions, such as the Fréchet distribution, Gumbel distribution, Weibull distribution, and more. These distributions have been widely applied in diverse fields, including flood prediction, rainfall estimation and stock market forecasting. It can be attempted to use these distributions to study sunspot areas for predicting the strength of solar cycle. We hope that this study would attract more authors to use newly methods to predict the strength of the solar cycle to make a great impact on our life in the future.

Acknowledgements

References

Acero F. J., Carrasco V. M. S., Gallego M. C., García J. A., Vaquero J. M. 2017, Astrophys. J., 839, 98

Acero F. J., Gallego M. C., García J. A., Usoskin I. G., Vaquero J. M. 2018a, The Astrophysics Journal, 853, 80

Acero F. J., García J. A., Gallego M. C. 2011, J. Clim., 24, 1089

Acero F. J., García J. A., Gallego M. C., Parey S., Dacunha-Castelle D. 2014, J. Geophys. Res. (Atmos.), 119, 39

Acero F. J., Vaquero J. M., Gallego M. C., García J. A. 2018b, Geophys. Res. Lett., 45, 9435

Asensio Ramos A. 2007, Astron. Astrophys., 472, 293

Attia A.-F., Ismail H. A., Basurah H. M. 2013, Astrophys. Space Sci., 344, 5

Babcock H. W. 1961, Astrophys. J., 133, 572

Bernstein J. P., Bhavsar S. P. 2001, Mon. Not. R. Astron. Soc., 322, 625

Bhavsar S. P., Barrow J. D. 1985, Mon. Not. R. Astron. Soc., 213, 857

Bhowmik P., Jiang J., Upton L., Lemerle A., Nandy D. 2023, Space Sci. Rev., 219, 40

Bhowmik P., Nandy D. 2018, Natu. Commun. 9, 5209

Cameron R., Schüssler M. 2008, Astrophys. J., 685, 1291

Castillo E., Hadi A. S., Balakrishnan N., Sarabia J. M. 2004, Wiley

Chen Y.-Q., Zheng S., Xiao Y.-S. *et al.* 2021, Atmosphere, 12, 1176

Choudhuri A. R., Chatterjee P., Jiang J. 2007, Phys. Rev. Lett., 98, 131103

Chowdhury P., Jain R., Ray P. C., Burud D., Chakrabarti A. 2021, Sol. Phys., 296, 1

Coelho C. A. S., Ferro C. A. T., Stephenson D. B., Steinskog D. J. 2008, J. Clim., 21, 2072

Coles S. G. 2001, An Introduction to Statistical Modeling of Extreme Values (Berlin: Springer)

Deng L. H., Li B., Xiang Y. Y., Dun G. T. 2016, Astron. J., 151, 2

Deng L. H., Zhang X. J., Deng H., Mei Y., Wang F. 2020, Mon. Not. R. Astron. Soc., 491, 848

Dikpati M., de Toma G., Gilman P. A. 2006, Geophys. Res. Lett., 33, L05102

Elvidge S., Angling M. J. 2018, Space Weather Int. J. Res. Appl.

Ferreira A., de Haan L. 2013, arXiv e-prints, arXiv:1310.3222

Gilli M., Keellezi E. 2006, Comput. Econ., 27, 207

Gopalswamy N., Makela P., Akiyama S., Yashiro S., Thakur N. 2018, J. Atmos. Sol. Terr. Phys., 179, 225

Hathaway D. H. 2015, Living Rev. Sol. Phys., 12, 4

Heristchi D., Mouradian Z. 2009, Astron. Astrophys., 497, 835

Javaraiah J. 2015, New Astron., 34, 54

Jiang J., Cao J. 2018, J. Atmos. Sol. Terr. Phys., 176, 34

Karak B. B. 2023, Living Rev. Sol. Phys., 20, 3

Karak B. B., Choudhuri A. R. 2011, Mon. Not. R. Astron. Soc., 410, 1503

Karak B. B., Mandal S., Banerjee D. 2018, Astrophys. J., 866, 17

Kitiashvili I. N. 2021, Mon. Not. R. Astron. Soc., 505, 6085

Kumar P., Biswas A., Karak B. B. 2022, Mon. Not. R. Astron. Soc., 513, L112

Love J. J. 2020, Space Weather, 18, e02255

Mandal S., Krivova N. A., Solanki S. K., Sinha N., Banerjee D. 2020, Astron. Astrophys., 640, A78

Nandy D. 2021, Sol. Phys., 296, 54

Noble P. L., Wheatland M. S. 2012, Sol. Phys., 276, 363

Nogaj M., Yiou P., Parey S., Malek F., Naveau P. 2006, Geophys. Res. Lett., 33, L10801

Pala Z., Atici R. 2019, Sol. Phys., 294, 50

Petrovay K. 2020, Living Rev. Sol. Phys., 17, 2

Prasad A., Roy S., Sarkar A., Chandra Panja S., Narayan Patra S. 2022, Adv. Space Res., 69, 798

Rigozo N. R., Souza Echer M. P., Evangelista H., Nordemann D. J. R., Echer E. 2011, J. Atmos. Sol. Terr. Phys., 73, 1294

Sarp V., Kilcik A., Yurchyshyn V., Rozelot J. P., Ozguc A. 2018a, Mon. Not. R. Astron. Soc., 481, 2981

Sarp V., Kilcik A., Yurchyshyn V., Rozelot J. P., Ozguc A. 2018b, Mon. Not. R. Astron. Soc., 481, 2981

Schatten K. H., Scherrer P. H., Svalgaard L., Wilcox J. M. 1978, GRL, 5, 411

Tsiftsi T., De la Luz V. 2018, Space Weather, 16, 1984

Willis D. M., Tulunay Y. K. 1979, Sol. Phys., 64, 237