

Extreme value theory and the solar cycle

A. Asensio Ramos

Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain
e-mail: aasensio@iac.es

Received 30 March 2007 / Accepted 14 June 2007

ABSTRACT

Aims. We investigate the statistical properties of the extreme events of the solar cycle as measured by the sunspot number.

Methods. The recent advances in the methodology of the theory of extreme values are applied to the maximal extremes of the time series of sunspots. We focus on the extreme events that exceed a carefully chosen threshold and a generalized Pareto distribution is fitted to the tail of the empirical cumulative distribution. A maximum likelihood method is used to estimate the parameters of the generalized Pareto distribution and confidence levels are also given to the parameters. Due to the lack of an automatic procedure for selecting the threshold, we analyze the sensitivity of the fitted generalized Pareto distribution to the exact value of the threshold.

Results. According to the available data, which only span the previous ~250 years, the cumulative distribution of the time series is bounded, yielding an upper limit of 324 for the sunspot number. We also estimate that the return value for each solar cycle is ~188, while the return value for a century increases to ~228. Finally, the results also indicate that the most probable return time for a large event such as the maximum at solar cycle 19, happens once every ~700 years and that the probability of finding such a large event with a frequency smaller than ~50 years is very small. In spite of the essentially extrapolative character of these results, their statistical significance is very large.

Key words. methods: data analysis – methods: statistical – Sun: activity – Sun: magnetic fields

1. Introduction

When analyzing a given physical process, repetitive measurement opens up the possibility of applying the power of statistical techniques. The well-developed field of statistics has devised a panoply of methods which allow us to infer properties from the observed phenomena. Sometimes, these statistical methods are so powerful that one can extract statistically significant information from noisy or a reduced set of observations.

One of the most striking examples of this is the case of extreme events. In spite of their inherent rarity, extreme events sometimes play important roles and turn out to be of fundamental importance. In certain fields (analysis of precipitation and floods, maximum temperatures, global climate, etc.), extreme events are those that produce radical and serious changes. For this reason, the extreme value theory is well-developed in these fields and has been applied in recent decades with great success.

We find few applications of the theory of extreme values in the field of Astrophysics (e.g., Bhavsar & Barrow 1985; Bhavsar 1990; Bernstein & Bhavsar 2001). This situation is somewhat surprising because usually the most interesting events that astrophysicists study are the most extreme ones. In this paper, we apply the theory of extreme values and its most recent advances to the investigation of the solar activity cycle. One of the best-known indicators of solar activity is the sunspot relative number. This indicator closely follows the 11-year solar activity cycle and has been continuously tabulated since ~1750. During the recent decades, there has been increasing interest in the prediction of the upcoming solar cycles and different techniques have been applied (e.g., Li et al. 2001; Orfila et al. 2002; Dikpati & Gilman 2006; Du 2006). The main cause for this interest is to be found, not only on the pure scientific curiosity of knowing in advance the amplitude of the following solar maximum, but in the

influence of a strong (or weak) solar maximum on the interplanetary medium. A strong solar maximum can induce solar storms which can damage the enormous number of satellites around the Earth. Another reason is the feasibility of a tripulated mission to Mars. The required long journey has to be carried out away from strong solar maxima to minimize the exposure to dangerous doses of radiation. Our work analyzes the statistical properties of these extreme activity events to estimate their frequency and amplitude. This could serve to gain some insight into the efficient amount of shielding needed to protect satellites and/or tripulated missions.

For this reason, the ability to predict when such extreme solar cycle events happen is of interest. A few studies have been oriented towards the investigation of the statistics of such extreme events. Siscoe (1976a,b) used the theory of extreme value to analyze the largest sunspot number per solar cycle. Later, Willis & Tulunay (1979) extended these studies by analyzing the data from sunspots umbrae, complete sunspots (umbrae+penumbrae) and faculae during nine solar cycles. However, this type of analysis has not been repeated during the last 25 years, where more than two solar cycles have occurred. Additionally, the previous studies were based on older approaches to the statistics of extreme values.

In this paper we extend the previous analysis of the solar cycle based on the monthly variation of the sunspot number. The interest of this study resides in the application of the more recent techniques to infer the statistical properties of the extreme values of the activity cycle. Our purpose is to identify clues that help forecast the extremes of the solar cycle and their occurrence frequency for extended periods of time (not only in the future, but also in the past). In more detail, once the probability distribution function of extreme events (largest number of sunspots) is characterized, we investigate whether this distribution is limited

and which are the typical events that we can expect for a given amount of time.

2. Extreme value theory

It is well-known that the Central Limit theorem (e.g., Feller 1971) states the asymptotic distribution that a sum of identically distributed random variables with finite variances will follow when the number of these variables is sufficiently large. What is less known is that a similar theory exists for the distribution of the maximum values taken by a random variable. This apparent lack of awareness has to be found in the fact that, although the theoretical formalism has been known for several decades, it has taken a long time to devise practical methods to apply the formalism to real data. In addition, while the Central Limit theorem deals with the functional form of the part of the distribution with largest probability (where the majority of the events occur) when a sufficiently large number of random variables with finite variance are summed, the extreme value theory deals with the tails of such distributions. As a consequence, since by definition very few events occur in the tails of the distribution, this lack of observables makes the estimation of the probability tail a very difficult task. Fortunately, statisticians have developed efficient methods which can make use of the few events in the tails to estimate the statistical properties of such extreme events.

Two different approaches are used mainly for the analysis of extreme events. The essential difference is in the way extreme events are defined. Let $\{X_i, i = 1 \dots n\}$ be a sequence of independent random variables which have a common distribution function F . Each X_i can be considered as a measure of a random process taken with a certain timestep Δt . The first approach, termed the *block maxima* approach (Fisher & Tippett 1928; Coles 2001), takes as extreme events the maximum (or minimum) value of the random variable in fixed intervals of time. For instance, if the timestep is considered to be one hour, the maximum among 24 consecutive measurements is the daily maximum. The second approach, termed the *peaks over threshold* (POT) approach (e.g., Coles 2001), takes as extreme events all the values of the time series that exceed a given threshold. The first approach was that taken by Siscoe (1976a,b), and Willis & Tulunay (1979) for the analysis of extreme values in the temporal evolution of sunspot numbers. Traditionally it has been selected for the analysis of time series in which a clear seasonal (periodic) behavior is detected. However, in spite of the theoretical simplicity of the block maxima approach, it suffers from important drawbacks. One of the most important limitations is that it tends to make inefficient use of the data. The importance of the necessity to overcome this lack of efficiency lies in the fact that the extreme value theory deals with extreme events, which are, by definition, scarce. For this reason, POT is becoming the fundamental method of choice in recent applications of the theory of extreme events thanks to the efficient use of the reduced amount of data available.

We now briefly present the theoretical results which we use in this paper. Assume that we measure a random variable at constant time intervals and that we obtain a sequence $\{X_i, i = 1 \dots n\}$. The measurements have to be independent and they have a common distribution function. The sequence can then be described with the aid of the cumulative distribution function F . Since we are focusing on extreme events, we are interested in the tail of such distribution. The POT approach is based on analyzing what is known as the *conditional excess distribution function*, $F_u(y)$, defined as:

$$F_u(y) = P(y \geq X - u | X > u), \quad 0 \geq y < \infty, \quad (1)$$

where X is the random variable, u is the threshold used to distinguish the maximum values and we have assumed (for simplicity) that X can take infinitely large values. This $F_u(y)$ function therefore describes the cumulative probability that, given a value of the random variable larger than the threshold, it exceeds the threshold by a quantity y . In the case that the complete cumulative distribution function F is perfectly known, $F_u(y)$ would also be known. In realistic applications, the cumulative distribution has to be empirically estimated and the tail of the distribution is often poorly sampled. For this reason, the estimation of $F_u(y)$ is usually not possible or uncertain. Theoretically, there is a relation between F and $F_u(y)$:

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad y > 0. \quad (2)$$

The feasibility of the POT approach finds its roots in the powerful theoretical result by Pickands (1975), who derived that, for a very large class of underlying F distributions¹, the conditional excess distribution function is well approximated by the generalized Pareto distribution (GPD):

$$F_u(y) \approx \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma}y\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-y/\sigma} & \text{if } \xi = 0, \end{cases} \quad (3)$$

with $y \in [0, \infty]$ if $\xi \geq 0$ and $y \in [0, -\sigma/\xi]$ if $\xi < 0$. The GPD is a general cumulative distribution function which is able to model the behavior of different tails depending on the exact value of the parameters. The quantity ξ gives information about the shape, in particular, the “strength” of the tail. An exponential-type distribution (normal, exponential, log-normal) is found for $\xi = 0$, a bounded beta-type distribution (beta or uniform) is found for $\xi < 0$ (zero probability is assigned for events above a certain limit) while a heavy-tailed Pareto-type distribution (power law, Pareto, Cauchy) is found for $\xi > 0$ (the tail falls more slowly than an exponential). The quantity σ gives information about the scale of the distribution. Although not applied in this study, it is important to remember that there is a duality between these distributions and the extreme value Gumbel, Fréchet and Weibull distributions of the block maxima approach.

With the aid of the theorem developed by Pickands (1975), the functional form of the cumulative distribution function for events above u can be written. Making $x = u + y$ and solving for $F(x)$ in Eq. (2), the cumulative distribution function for events above the threshold u can be written as:

$$F(x) = 1 - \frac{N_u}{n} \left[1 + \frac{\xi}{\sigma}(x - u)\right]^{-1/\xi}. \quad (4)$$

The previous expression assumes that the value of the cumulative distribution at u is given by the estimation $(n - N_u)/n$ (the so-called historical simulation), with n the total number of points in the time series we are analyzing and N_u the number of points above the threshold. This estimation is expected to be accurate if the threshold is high enough. Obviously, the functional form described by Eq. (4) is only valid for $x \geq u$. Once the function F is known, all the statistical properties of the extreme events can be calculated. One of the most interesting statistical properties is the so-called “return time”. This is defined as the typical time that one has to wait until an event of amplitude x_{ret} happens again. It can be estimated from Eq. (4) by setting $x = x_{\text{ret}}$ and making the

¹ According to Pickands (1975), the class of distributions F that fulfill the theorem are those whose (block) maxima follow one of the three families of extreme value distributions (Fisher & Tippett 1928). These are the Gumbel, Fréchet or Weibull distributions.

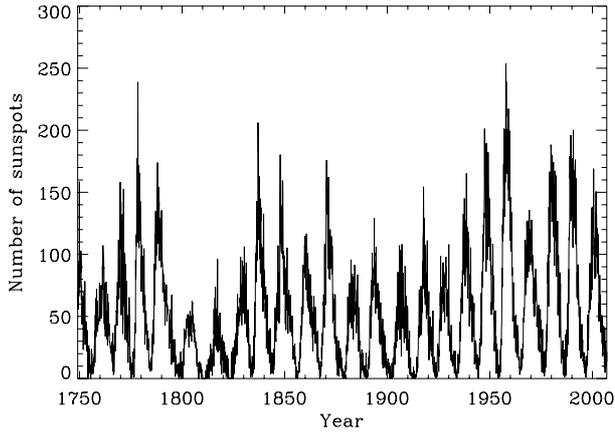


Fig. 1. Solar cycle variation of the International Sunspot Number.

identification $t_{\text{ret}}^{-1} = 1 - F(x_{\text{ret}})$. The units of this time variable depend on the units of the timestep Δt .

The parameters of the GPD are usually obtained from the empirical data by means of a maximum likelihood estimation. Assuming that y_1, y_2, \dots, y_N are the N values of the original time series that exceed the threshold u , the log-likelihood is (Coles 2001):

$$\ell(\sigma, \xi) = -N \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^N \log \left(1 + \frac{\xi y_i}{\sigma}\right). \quad (5)$$

Since the value of the σ and ξ parameters that maximize the log-likelihood cannot be found analytically, the $\ell(\sigma, \xi)$ function is maximized using standard numerical optimization methods (e.g., Press et al. 1986). Once the maximum likelihood value of the parameters are found, it is possible to calculate confidence intervals. A standard method relies on the assumption that the likelihood function is given approximately by a normal distribution (Coles 2001). Under this assumption, the confidence interval can be estimated with the aid of the estimated curvature of the likelihood function, which is proportional to the Hessian matrix of Eq. (5). A more refined method, and the one that we use in this paper, is to use the information from the likelihood function itself. This allows us to give asymmetric confidence intervals in the parameters due to the skew of the likelihood function.

3. Application to the solar cycle

The time series representing the solar activity cycle during the last 257 years is shown in Fig. 1. The data represent the sunspot number, an estimation of the number of individual sunspots (counting individual sunspots and group of sunspots). In order to decrease the dispersion, we focus here on the monthly averaged value. The data have been tabulated since 1750 and it is now known as the International Sunspot Number². The time series presents a clear regularity with time as a consequence of the influence of the solar cycle on the surface magnetism. Under these circumstances, the premises on which the POT formalism lies are not fulfilled. In particular the random variables are not independent because there is a certain degree of correlation between consecutive events: a large sunspot number is typically followed by another large sunspot number.

Several techniques have been devised to overcome this difficulty. One of the most applied methods is the “de-clustering” of

the time series (e.g., Coles 2001). It consists of locating clusters in the exceedance over the threshold and representing them by the maximum value inside each cluster. This has two undesired consequences: (i) the number of events available for the GPD analysis is reduced and (ii) a somewhat arbitrary criterion for the cluster definition has to be included. Recently, Fawcett & Walshaw (2007) have shown that this de-clustering technique introduces biases in the maximum likelihood estimations of σ and ξ . They also show that the direct application of the POT method using the whole time series neglecting any temporal periodicity leads to negligible biases. The price to pay is that the confidence intervals for the GPD parameters are larger than those obtained using standard techniques (Coles 2001). Following Fawcett & Walshaw (2007), we apply the POT method to the sunspot number time series without any de-clustering technique. The POT formalism also requires the underlying distribution of the random variables to be stationary (see, e.g., Coles 2001). The large number of studies that are successful in reproducing the time evolution of the International Sunspot Number using deterministic methods (e.g., Verdes et al. 2000; Dikpati & Gilman 2006; Choudhuri et al. 2007; Cameron & Schüssler 2007) suggest that this is the case. Therefore, we can safely consider that the physics (probability distribution function of the random variables) driving the solar cycle does not vary appreciably with time.

In this paper we only focus on the statistics of the upper tail of the distribution, i.e., maximal values of the sunspot number. Although the application of the POT formalism to the analysis of the distribution of minimal values is also possible, larger time series are needed. We briefly discuss this issue in Sect. 4. In order to apply the POT formalism, a threshold u has to be fixed. The threshold has to be sufficiently large so that the generalized Pareto distribution is a suitable functional form for describing the tail of the cumulative distribution and it has to be sufficiently small so that enough values are available to give an accurate estimation of the parameters of the GPD. There is not any known automatic procedure for the selection of the threshold. In this paper we choose a value of the threshold based on reasonable ideas and we verify the behavior of the parameters of the GPD for different values of the threshold. In our case, u has been chosen as the value that leaves 96% of the points of the time series below and only 4% of the points are considered as extreme values. For the dataset shown in Fig. 1, we find $u = 149.4$. From the original set of 3096 data points, we leave 122 points above the threshold which are used to fit the GPD neglecting any time dependence. The empirical cumulative distribution function for points above the threshold is built and the values of σ and ξ that give the best fit are obtained.

The GPD parameters have been estimated maximizing the log-likelihood given by Eq. (5). As noted above, such an approach permits the most probable values and their confidence intervals to be obtained. In our case, we obtain $\sigma = 27.157^{+4.587}_{-3.632}$ and $\xi = -0.155^{+0.19}_{-0.09}$, as shown in Table 1. With these values, the ensuing cumulative distribution function is shown in the left panel of Fig. 2, where we show the value over the threshold on the horizontal axis and the value of the GPD on the vertical axis. The right panel of Fig. 2 shows the empirical cumulative distribution versus the fit $F_u(y)$ using the GPD. This is the so-called probability plot, and it clearly indicates that the GPD produces a good approximation to the empirical cumulative distribution.

Several interesting points deserve comment. Firstly, the confidence intervals presented in Table 1 are probably underestimating the true confidence intervals because we have neglected the temporal dependence. Secondly, a negative value for the shape

² <http://sidc.oma.be/>

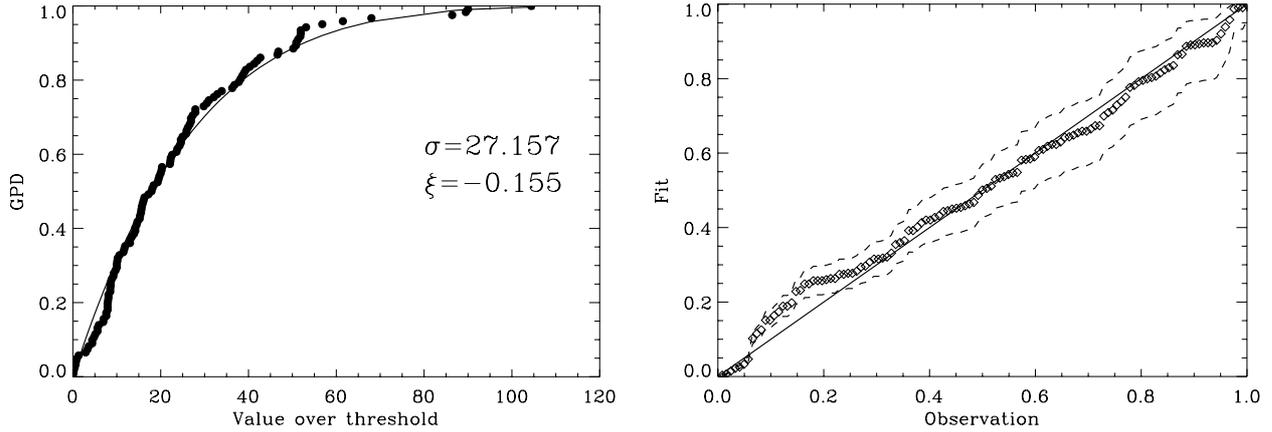


Fig. 2. Summary of the quality of the fit of the empirical cumulative distribution to a generalized Pareto distribution. The *left panel* shows the value of the cumulative distribution for different values of the sunspot number above the selected threshold of 149.4. The *right panel* shows the variation of the quality of the fit when the uncertainties in the σ and ξ parameters of the GPD are taken into account.

Table 1. Parameters and return level estimates. The values are obtained for a threshold of $u = 149.4$.

Parameter	Estimate	95% confidence interval
σ	27.157	[23.525, 31.745]
ξ	-0.155	[-0.251, 0.035]
$-\sigma/\xi + u$	324.0	[270.0, 990.0]
11-year return level	187.8	[180.2, 203.0]
100-year return level	228.4	[208.1, 253.8]

parameter ξ is found. The evidence for this is strong because the 95% confidence interval is almost exclusively in the negative domain. According to the results presented above, this suggests that the cumulative distribution is bounded and that zero probability is assigned to excesses above a certain limit, given by the ratio $x_{lim} - u = -\sigma/\xi$. In our case, we find that the fitted GPD assigns zero probability to extreme values larger than 175.6 above the threshold. Taking into account the chosen threshold, this gives a limit of $x_{lim} = 324.0$. This value is consistent with the data presented in Fig. 1. Note also that the 95% confidence interval is [270.0, 990.0], which has been obtained from the likelihood function. It gives a very stringent value of the lower limit (as is obvious because of the presence of a large amount of data below the threshold) but a very large upper limit. A 68% confidence interval is estimated to be [288, 420]. This is a clear indicative of the asymmetry of the likelihood function with a very long tail towards larger x_{lim} , giving the idea that the information for a strong upper limit is hardly present in the data. It is however important to take into account that these results can fluctuate depending on the exact value of the chosen threshold. Furthermore, it can also fluctuate in the future if the same calculation is carried out with more data from subsequent solar cycles.

In order to analyze the strength of this conclusion, we carried out the fit of the cumulative distribution for different values of the threshold. The results are shown in Fig. 3 for different values of the threshold u , the upper panel showing the values obtained for σ and the lower panel the values for ξ . If the GPD is a reasonable model for the exceedances above a certain threshold u_0 , it should remain reasonable for larger values of the threshold (e.g., Coles 2001). As a consequence, the estimates of σ and ξ should remain constant above u_0 provided u_0 is a valid threshold. We see that this happens in our case for $u > 150$, the value we have chosen above. It is interesting to note that for $u \sim 120$, approximately 9% of the points lie above the threshold, while

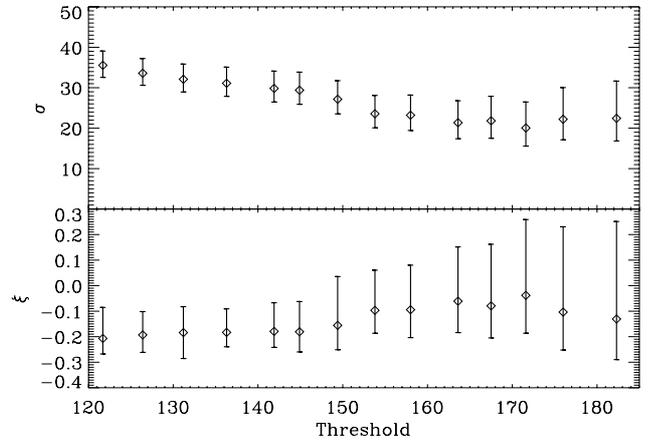


Fig. 3. Values of the σ and ξ parameters obtained for different values of the threshold.

for $u \sim 180$, less than 1% of the points lie above. There is a clear increase in the uncertainty of the retrieved parameters when the threshold increases due to the lack of points. What seems clear is that, for $u < 150$, the ξ parameter has only negative values inside the error bars. For larger values of the threshold, both negative and positive values for ξ are obtained, although negative values tend to be more probable.

Once the cumulative distribution given by Eq. (4) is obtained, several statistical properties of the extreme values can be inferred. One of the most interesting in terms of prediction of future events are the so-called return times and return level. The return time is the typical time one has to wait until an event reaches and surpasses a threshold. Similarly, the return level is the typical extreme event one would find after waiting for a given amount of time. These quantities are obtained easily from Eq. (4) and they are shown in Fig. 4. For consistency, we only show the results for values above the chosen threshold. The dashed lines present the confidence interval which is induced by the uncertainty in the inferred parameters of the GPD. The vertical dotted line approximately indicates a solar cycle, of the order of 11 years. For such an amount of time, the return level equals ~ 187 . If we take into account the confidence interval, we find that the return level lies in the interval [180, 203] with 95% confidence. These values appear to be consistent with the empirical results from the past, according to Fig. 1. This is produced by the similarity between

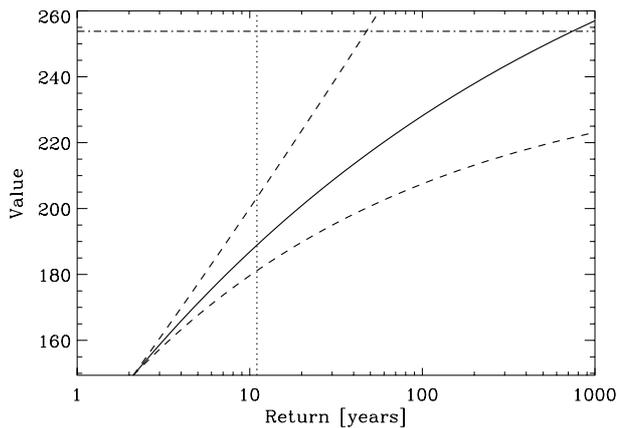


Fig. 4. Return value for a given amount of time (or return time for a given level) for the fit obtained with a threshold of $u = 149.4$. The dashed lines indicate the confidence levels of the return value taking into account the uncertainty in the GPD parametrization. The vertical dotted line indicates 11 years, approximately one solar cycle. The horizontal line indicates the value of the maximum found on solar cycle 19.

the empirical tail of the extreme distribution and the GPD, which makes the statistical properties inferred from the GPD a good estimation of the statistical properties of the empirical distribution. If we assume that the previous 23 solar cycles are representative of the behavior of the Sun during a longer time period and unless a long-period modification of the solar cycle exists, the previous estimations of the return levels are statistically significant. For the case of the return level for 100 years, we find a value of ~ 228 , with a confidence interval of $[208, 254]$. Again, this value is consistent with the data.

Concerning the return times, it is interesting to estimate them for the most extreme cases in the observed dataset. It is important to have in mind that relying on the GPD for very long-term extrapolation can be dangerous. The dataset on which we have applied the extreme value theory spans only ~ 250 years, so one should not blindly rely on the return values for large events if they only happen once in the original data. For instance, according to the GPD, an event such as the extremely strong peak of cycle 19 (the maximum of 253.8 took place during 1957) occurs once every ~ 700 years. Taking into account the uncertainty in the GPD fit, we find that it is possible to give only a lower limit to this return time because the upper limit is unbounded. The horizontal line indicates the maximum value of 253.8 obtained during cycle 19. The intersection of this horizontal line with the solid line given by the GPD extrapolation occurs at ~ 700 years, showing the most probable value of the return time. The intersection with the dashed lines indicate the intervals of 95% confidence. In this case, we find that the GPD extrapolation implies that such an extreme event happens, with 95% probability, with a frequency above once every ~ 50 years, with an apparently unbounded upper limit.

We want to leave a word of caution on the values obtained above because of their extrapolative character. However, we also want to stress the fact that this extrapolative character is based on strong theoretical roots. A small variation on the results obtained in this work might be expected as soon as more data showing more extreme events is available. This variation with respect to the values presented above should be very small if the underlying probability distribution function (essentially, the physics that drives the solar cycle) that we have calculated does not change appreciably with time.

4. Concluding remarks

We have presented an extreme value theory analysis of the solar cycle as measured by the sunspot number. The peak-over-threshold method has been applied. The analyzed time series presents a certain degree of correlation because a large number of sunspots is usually followed by another large number of sunspots and these variables cannot be considered to be uncorrelated. Following Fawcett & Walshaw (2007), we have applied the POT method to the time series without any de-clustering technique. As a consequence, the confidence intervals presented in this study could be an underestimation of the real confidence intervals because the likelihood function is affected by the presence of time correlation in the time series. Our results indicate that the distribution of extreme solar cycle events is bounded so that the value of 324 cannot be exceeded. The analysis of the confidence intervals give that, with 95% confidence, this maximum value is larger than 288. Of more relevance are the 11-year and 100-year return values. The results indicate that there is a very high probability of finding values in the range $[180, 203]$ every solar cycle, and values in the range $[208, 254]$ every ~ 10 solar cycles. Additionally, we have shown that an extreme event such as that on solar cycle 19 (during 1957) occurs with a frequency above once every ~ 50 years.

The results obtained in this paper are based on the statistical analysis of the tail of the sunspot number distribution. This analysis is driven by the extreme value theory that is based on strong theoretical roots developed during the last 50 years. Such application relies on the assumption that the limiting behavior of the stochastic process behind the observed time series can be obtained from the application of certain mathematical limits. A functional form for the tails of distributions is available and we only have to fit this tail distribution to our dataset. However, this approach has limitations. One of the strongest is that it is not yet clear whether the limiting mathematical models that we have used can be directly applied to finite time series. It is expected that, in the limit of an infinitely large time series, the models correctly reproduce the tails of the underlying distribution. However, when the time series is of limited size, fluctuations can be of importance and lead to inaccuracies. In our case, the results that we present appear to support the fact that the statistics of extreme events are correctly reproduced under the framework of the extreme value theory. The fact that the theory explains the already observed extreme events is also favorable. The validity of the theory is also supported by the large number of practical applications of the theory which we found in the literature.

The extreme value theory can also be applied to minima. One of the most interesting future applications of this approach is to estimate the return time for long low-activity periods, with the aim of estimating the frequency with which extreme events such as the Maunder minimum may take place. A much longer time series of solar activity would be needed for such an estimation. Furthermore, an appropriate re-definition of the time series has to be carried out in order to transform these low-activity periods into extremes cases. A possibility that could be of interest is to use the solar sunspot number reconstructed from ^{14}C activity during the last 11 000 years (Solanki et al. 2004). This reconstruction shows periods of low activity similar to the Maunder minimum whose probability (and ensuing return time) could be calculated in the framework of the extreme value theory.

Acknowledgements. I thank A. López Ariste, R. Manso Sainz and V. Martínez Pillet for helpful discussions and the anonymous referee for constructive remarks. This research has been funded by the Spanish Ministerio de Educación y Ciencia through project AYA2004-05792.

References

- Bernstein, J. P., & Bhavsar, S. P. 2001, MNRAS, 322, 625
- Bhavsar, S. P. 1990, in *Evolution of the Universe of Galaxies*, ed. R. G. Kron, ASP Conf. Ser., 10, 209
- Bhavsar, S. P., & Barrow, J. D. 1985, MNRAS, 213, 857
- Cameron, R., & Schüssler, M. 2007, ApJ, 659, 801
- Choudhuri, A. R., Chatterjee, P., & Jiang, J. 2007, Phys. Rev. Lett., 98, 131103
- Coles, S. 2001, *An Introduction to Statistical Modeling of Extreme Values* (London, UK: Springer-Verlag)
- Dikpati, M., & Gilman, P. A. 2006, ApJ, 649, 498
- Du, Z. L. 2006, A&A, 457, 309
- Fawcett, L., & Walshaw, D. 2007, *Environmetrics*, in press
- Feller, W. 1971, *An introduction to probability theory and its applications*, Vol. 2, 3rd Ed. (New York: Wiley)
- Fisher, R. A., & Tippett, L. H. C. 1928, Proc. Cambridge Phil. Soc., 24, 180
- Li, K. J., Yun, H. S., & Gu, X. M. 2001, A&A, 368, 285
- Orfila, A., Ballester, J. L., Oliver, R., Alvarez, A., & Tintoré, J. 2002, A&A, 386, 313
- Pickands, J. 1975, Ann. Stat., 3, 119
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1986, *Numerical Recipes* (Cambridge: Cambridge University Press)
- Siscoe, G. L. 1976a, J. Geophys. Res., 81, 4782
- Siscoe, G. L. 1976b, J. Geophys. Res., 81, 6224
- Solanki, S. K., Usoskin, I. G., Kromer, B., Schüssler, M., & Beer, J. 2004, Nature, 431, 1084
- Verdes, P. F., Parodi, M. A., Granitto, P. M., et al. 2000, Sol. Phys., 191, 419
- Willis, D. M., & Tulunay, Y. K. 1979, Sol. Phys., 64, 237