Solar Phys (2019) 294:157
https://doi.org/10.1007/s11207-019-1536-1

Prediction of the Sunspot Number with a New Model Based on the Revised Data

Jinhuo Liu[1] · **Juan Zhao**[1] · **Haibo Lin**[1]

Received: 16 May 2018 / Accepted: 9 October 2019 / Published online: 14 November 2019

Abstract Sunspot-number prediction plays a very important role in space-weather forecasts and environmental research. In 2015, the international relative sunspot-number publisher released a new version of the data. Compared with the old version, the new data changed enough to influence sunspot-prediction methods that were based on the old data. The aim of this study is to propose a new prediction method based on the new data, and improve prediction accuracy as much as possible. A modified gaussian function, which has four parameters, was used in our method to describe each single cycle. Via four relationships among the four parameters, we found the most probable values and their uncertainties for the four parameters, and then we obtained the sunspot-number variation curve and its range for Solar Cycle 24. The results showed that the peak value should be 113.3 (the real value was 116.4), at the 57th month (between the two real peak at the 39th and the 64th month). A range of peak values was also given by this method, which range from 91 (less than the real value by 25) to 134 (greater than the real value by 18).

Keywords Solar Cycle, models · Sunspots, statistics

1. Introduction

Sunspots are important surface manifestations of solar activity. The observation of sunspots was started in the 17th century by Galileo Galilei with his telescope (Casas, Vaquero, and Vazquez, 2006), but sunspot cycles were not confirmed until 1844, when Schwabe discovered the existence of Solar Cycles 7 and 8 with his observation of sunspots from 1826 to 1843 (17 years) (Schwabe and Schwabe, 1844). In 1849, during Cycle 9, a recording method was proposed by Wolf to standardize records, which was the beginning of the international relative sunspot-number (SSN) record (Wolf, 1861). Thereafter, by collecting and scaling

✉ J. Zhao
zj@bnu.edu.cn

[1] Department of Astronomy, Beijing Normal University, Beijing 100875, China

up prior observations, the monthly averaged SSN time series extends back to 1749, and the annually averaged SSN series stretch back to 1700.

In 1957, during Cycle 19, the first spacecraft lifted off (Williamson, 2006), and the issue of the impact on spacecraft of solar activity started to attract attention. In March 1989, around the peak of Solar Cycle 22, a coronal mass ejection (CME) event caused large geomagnetic storms, which in turn led to serious impacts on multiple spacecraft (Leach, 1996). In October 1989, a flare eruption damaged solar-array electronics of GOES-5 and decreased its lifetime (Marvin and Gorney, 1992). In 2003, after the peak of Solar Cycle 23, the "Halloween storms" damaged 28 spacecraft, knocking two out of commission (Toth *et al.*, 2005). To minimize the impact on spacecraft from a hostile environment, we should enhance not only the adaptability of spacecraft to various space-weather disturbances but also our monitoring ability and early warning systems. By predicting the sunspot number, we can estimate the occurrence possibility of extreme solar events, which were the main cause of the hostile space environment.

Solar Cycle 24 started in December 2008. The maximum SSN for this cycle was predicted by several different methods. Gholipour *et al.* (2005) used a decomposition method based on singular spectrum analysis with the neural network method and found the maximum would be around 145. Wang *et al.* (2008) predicted that the peak value of Cycle 24 would be 119.5 ± 12.4 by the similarity cycle method. Aguirre, Letellier, and Maquet (2008) used a dynamic model with a deterministic part and a stochastic part and deduced the peak value of 65 ± 13 and 87 ± 13. Kakad (2011) used a new empirical model with two parameters Q_C and L, which are derived numerically solely from the information of the preceding solar cycle, and got the result 74 ± 10.

In 2016, after the peak of Cycle 24, Pesnell (2016) analyzed 105 different predictions for this cycle. These predictions were found before 2015 based on the old data, for which the peak value is around 90. The predicted results showed the predicted value distributed in a wide range from 40 to 185. Pesnell analyzed these prediction methods and classified them into six categories, which are climatology, recent climatology, spectral, dynamo model, neural network, and precursor. Among them, the climatological method predicted based on the previous statistical rules of sunspot number. Our method could be classified into this category, of which the forecast results are between 40 and 185.

In 2015, because of the growing gap between the SSN and the sunspot group number, the Royal Observatory of Belgium, the international sunspot-number publisher, revised the old data and released the new data (Clette *et al.*, 2014, 2015; Clette and Lefèvre, 2016; SILSO, 2015). Compared to the old data, the new data had significant changes, which were irregular and nonlinear. So, it is reasonable to think that this change may affect the prediction method.

Based on the new data, we propose a new method to predict sunspot numbers. A function, which was given by Hathaway, Wilson, and Reichmann (1994), was used to describe sunspot cycles. This function has been used many times in studies of solar activity. Li (1997) applied this function to predicted the maximum and other features of Solar Cycle 23. Li (1999) made efforts in revising and simplifying this function, and analyzed characteristics of the sunspot cycle shape with this function. Lasheng *et al.* (2005) analyzed the parameters of this function based on sunspot-area data. Up to now, this function has been accepted to describe the shape of the sunspots cycle. Using the data from Solar Cycle 1 to 23, we find the internal relationships and changing rules of four parameters in the function, and then we derive the prediction curve for Solar Cycle 24. Comparing the predicted curve with the real curve, we can evaluate our new method. In addition, we will use this method on Cycles 22 and 23 to check its applicability and reliability.

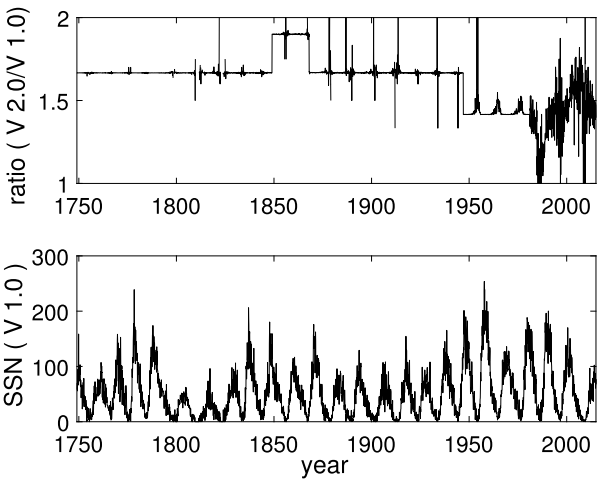

Figure 1 The *upper panel* shows the changing value of the ratio between the old sunspot-number (SSN) data (V1.0) and the new data (V2.0), while the *lower panel* shows the old data.

2. Change of International Relative Sunspot Number

The SSN data is from the Royal Observatory of Belgium (SILSO, 2015). The monthly averaged SSN data start January 1749 and end March 2017, covering 23 completed cycles (Solar Cycles 1 – 23) and the current cycle (Cycle 24). In 2015, Clette and co-workers revised the data and released the version 2.0 data on 01 July 2015. This version contains several corrections of past inhomogeneities in the time series (Clette and Lefèvre, 2016). We can see the data changes in Figure 1. From 1749 to 1947, the new time series is 1.67 times higher than the old time series for the most part. However, the ratio rises to 1.9 from 1849 to 1868, and falls to 1.41 from 1947 to 1981. After 1981, the ratio varies with time. In addition, the ratio fluctuates greatly for low values of the original. These changes may have a certain impact on the prediction methods based on the old data.

3. Method

3.1. The Basic Function of the Model

A modified gaussian function was given by Hathaway, Wilson, and Reichmann (1994), which was used to describe the curve of the SSN of each single solar cycle:

$$f(t) = \frac{a(t-d)^3}{\mathrm{e}^{\frac{(t-d)^2}{b^2}} - c} \tag{1}$$

This function has four parameters: a, b, c, and d. The parameter a stretches the curve along the y-axis; b stretches the curve along both the y- and the x-directions; c makes the curve asymmetrical, and d moves the curve along the x-axis. The first three parameters a, b, and c influence the height of the peak, while the last three parameters b, c, and d determine the position of the peak.

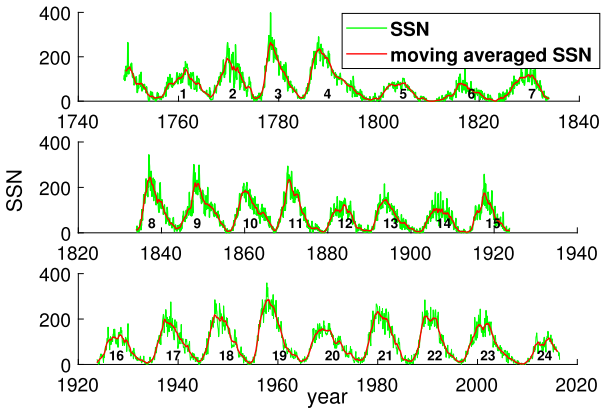

Figure 2 The monthly averaged SSN data and the data after removal of a 13-point moving average. The *red curve* is obviously smoother than the *green curve*.

3.2. Moving Average and Fitting Curve

Compared with the daily SSN data, the monthly averaged data have less fluctuation. However, as can be seen in Figure 2, especially around the peak, the curve still has strong fluctuation, which may complicate the prediction and cause unexpected errors during fitting. In order to avoid these situations, the data need to be further smoothed before prediction. Because what we predict is monthly average SSN data, we use a 13-point moving average.

The smoothed sequence is divided into 23 complete cycles and the ongoing Cycle 24. We cut the cycle at the median position where there are several zero points around the valley between two adjacent cycles. The Levenberg–Marquardt nonlinear least-square algorithm (LMA) is applied to each of the 23 complete cycles with Equation 1, respectively. The LMA is more robust than the Gauss–Newton algorithm, and it could obtain the minimum value even if initial coefficient values are far from the final result (Seber and Wild, 1989). However, as with most nonlinear fitting algorithms, this method could only calculate the local minimum. Therefore, the initial coefficients should be carefully considered. After testing, we set the initial coefficients to

$$a_0 = 0.005, \qquad b_0 = 50, \qquad c_0 = 0, \qquad d_0 = 0.$$

The initial values of coefficients a and b (a_0 and b_0) are around its mean value in the trial fitting with the values of 0.005 and 50. The parameter c affects the asymmetry of the curve, and 0 is the watershed for c where the peak shifts neither toward earlier nor later times. So, we set the initial value of c [c_0] to zero. The parameter d affects the location of the curve directly. For the same reason as c, we also set its initial value to zero.

The errors of the original data are used to estimate the errors in the derived fitting parameters with Monte Carlo method. For each cycle, we generated 10,000 curves and fitted them to get 10,000 groups of parameters. The errors were obtained to be

$$\sigma_{X_i}^2 = \frac{\sum_{j=1}^{10000}(X_{ij} - X_i)^2}{10000 - 1}$$

where i is the cycle number, and j is the number of the simulated curves in this cycle [i]. X represents one of the parameters a, b, c, and d. X_i is the fitting result of the curve without error of this cycle [i]. Fitting results and errors are presented in Table 1.

Table 1 The fitting results and the errors of four parameters in the basic function from Solar Cycles 1 – 23.

Cycle	$a \times 10^{-2}$	$\sigma_a \times 10^{-2}$	b	σ_b	c	σ_c	d	σ_d
1	0.18	0.07	69.6	2.4	−7.67	7.05	−35	6
2	0.28	0.04	51.7	1.6	0.40	0.38	−14	4
3	0.71	0.09	39.4	1.2	0.86	0.03	1.5	1.2
4	0.18	0.02	59.3	1.8	0.90	0.02	−10.6	1.3
5	0.18	0.12	59.4	2.1	−5.68	13.63	−24	10
6	0.13	0.03	54.1	2.4	−1.02	1.11	8	5
7	0.80	0.28	58.1	1.0	−51.67	28.59	−35	5
8	0.42	0.02	45.6	0.6	0.88	0.01	−1.7	0.6
9	0.15	0.01	65.2	0.8	0.26	0.11	−12.4	1.4
10	0.17	0.01	56.4	0.7	0.88	0.01	−1.1	0.6
11	0.42	0.02	47.2	0.5	0.55	0.08	−6.3	1.0
12	0.16	0.06	55.7	1.2	0.04	2.63	−17	9
13	0.20	0.01	50.3	0.6	0.88	0.02	−6.9	0.7
14	0.12	0.01	54.8	1.4	0.90	0.19	−2	4
15	0.35	0.03	49.3	0.5	−0.66	0.36	−11.9	1.3
16	0.19	0.01	52.8	0.7	0.45	0.25	−12	3
17	0.26	0.01	50.7	0.4	0.84	0.02	0.9	0.7
18	0.35	0.01	49.4	0.4	0.74	0.04	−4.0	0.8
19	0.59	0.01	44.3	0.2	0.87	0.02	0.9	0.4
20	0.13	0.01	60.0	0.4	0.87	0.01	−7.2	0.5
21	0.39	0.01	49.1	0.4	0.63	0.07	−3.0	1.0
22	0.44	0.01	45.6	0.3	0.79	0.04	−4.0	0.7
23	0.23	0.01	55.9	0.4	0.27	0.16	−10.0	1.4

3.3. Parameters of the Special Cycle 7

Before seeking the relationships between parameters, a special cycle will be discussed first. Cycle 7 started in November 1823 and ended in May 1834, after the valley of the Gleissberg cycle. This cycle seems very different from the other cycles in terms of variation trend. The peak of Cycle 7 appeared obviously later than other cycles (Figure 2). The ratio between the length of the rising and falling part can be found in Table 2. The ratio for Cycle 7 is much larger than those other cycles (even more than 1.5 times the secondary large value in the table). The delay of the peak has a large effect on the fitting result. Figures 3 show the variation of the parameters a, b, and c. Cycle 7 deviates clearly from these parameters of other cycles.

Schwabe and Schwabe (1844) showed that the peak of the sunspot group number of Cycle 7 was in 1828, which is obviously earlier than that of our SSN data at May 1830. Normally, there are differences between the sunspot group number and the sunspot relative number, but, compared with the 11-year total cycle length, the two years difference of peak position seems too long. For SSN data, the more reliable data start at 1850, after Wolf put forward the standard formula for the sunspot relative number. Cycle 7 is before 1850, so we consider that some of the particularity of Cycle 7 might to be caused by the reliability of data. This cycle will not be used in the prediction.

Table 2 The ratio between the length of the rising part and the falling part. The value of Cycle 7 is obviously larger than other cycles.

Cycle	Ratio	Cycle	Ratio
1	1.25	13	0.48
2	0.59	14	0.66
3	0.51	15	0.69
4	0.37	16	0.85
5	1.21	17	0.56
6	0.83	18	0.48
7	2.18	19	0.54
8	0.55	20	0.57
9	0.79	21	0.57
10	0.59	22	0.47
11	0.40	23	0.79
12	0.80		

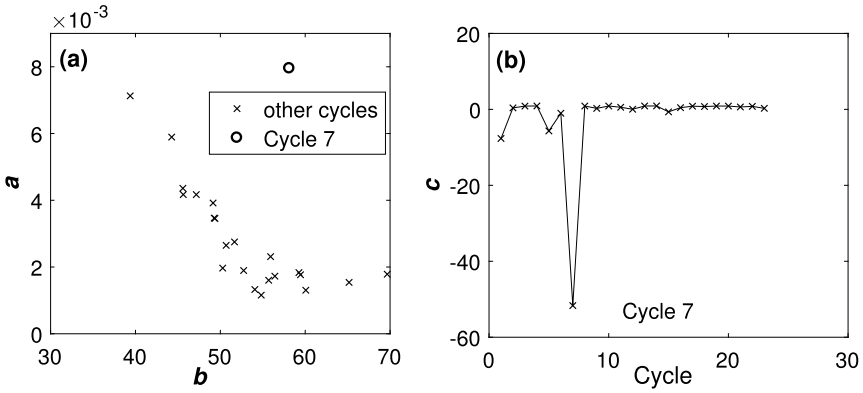

Figure 3 (**a**) The parameters *a vs*. *b*. The *circle* is Cycle 7, which deviates from the crosses of other cycles. (**b**) Value of parameter *c* as a function of cycle number. The point for Cycle 7 is obviously smaller than the other points.

3.4. Prediction of Parameters *a* and *b* for Cycle 24

The prediction model starts with parameters a and b for Cycle 24. To predict these two parameters we need at least two relationships about them. The first relationship established between parameters a and b will be introduced in Section 3.4.1. The second relationship, which establish the connection between a of different cycles, will be introduced in Section 3.4.2. In Section 3.4.3, we will discuss two special cycles that were found in Section 3.4.2. In Section 3.4.4, the prediction about parameters a and b for Cycle 24 will be given.

3.4.1. Relationship Between a and b

The first relationship is between a and b, but it is not a direct relationship. In Equation 1, parameter a multiplies t^3, while parameter b is in the denominator of t. It would be useful

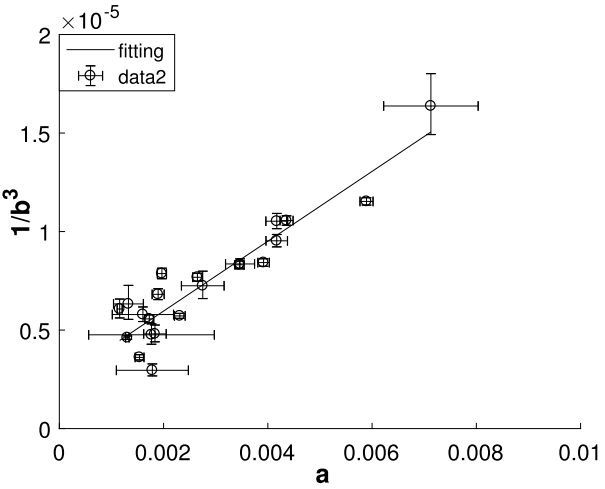

Figure 4 Scatter plot of the parameter $b' \equiv 1/b^3$ *vs.* parameter a. The errors have been added for each point, and the *solid line* is the linear relationship between parameters a and b'.

to define a new parameter b':

$$b' = \frac{1}{b^3} \tag{2}$$

There is a linear relationship between a and b', and we can see it in Figure 4. The relationship between a and b' is

$$b' = 1.8(\pm 0.3) \times 10^{-3} a + 2.4(\pm 1.1) \times 10^{-6} \tag{3}$$

with root-mean-square error (RMSE) $\sigma = 1.16 \times 10^{-6}$. The Pearson correlation coefficient r, 0.93, is higher than the 99.9% confidence level.

3.4.2. Relationship Between a and δa

The second relationship should connect the parameters of different cycles so that the parameters for Cycle 24 can be calculated from the known parameters. That is between a and δa, which is defined as

$$\delta a_i = a_i - a_{i+1} \tag{4}$$

where i is the cycle number.

The a and δa of 23 cycles are shown in Table 3, and the relationship can be seen in Figure 5. In the figure, the dot–dashed line is the linear-fit result of all of the points (Cycle 7 was eliminated). It is:

$$\delta a = 1.2(\pm 0.5) a - 3.3(\pm 1.6) \times 10^{-3} \tag{5}$$

with RMSE $\sigma = 1.6 \times 10^{-3}$. The value of the correlation coefficient r is 0.762.

In Figure 5, there are two points that deviate from the other points (in the lower-right part of the solid line). They are parameters a for Cycles 2 and 18. They deviate from the linear relationship by 2.75σ and 2.03σ, while the third-largest deviation is just 1.16σ. In addition, they are on the same side of the fitting curve. So, their impact on fitting is considerable. To reduce the effect caused by these two points, we used bisquare-weight robust-fitting to

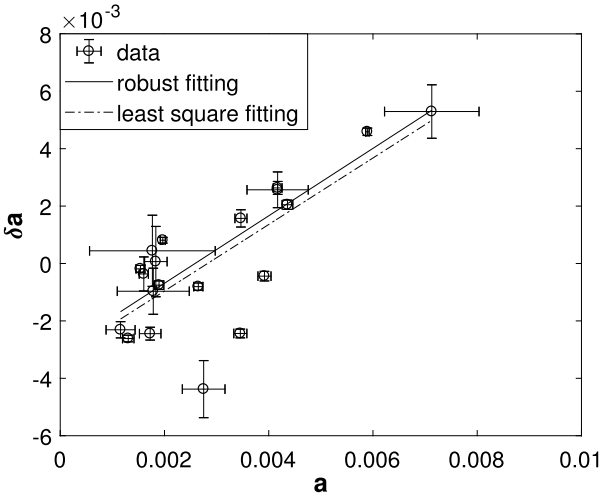

Figure 5 Scatter plot of δa *vs*. a. The *circles* are the original data in Table 3, and the errors have been added for each point. They have an obviously linear relationship. The *solid line* is the robust-fitting result, while the *dot–dashed* line is the least-squares fitting result.

Table 3 The parameter δa and its error used to compute the second relation. Cycle 7 has been eliminated. Because $\delta a_6 = a_6 - a_7$ and a_7 has been eliminated, Cycle 6 is eliminated too.

Cycle	$\delta a \times 10^{-4}$	$\sigma_{\delta a \times 10^{-4}}$	Cycle	$\delta a \times 10^{-4}$	$\sigma_{\delta a \times 10^{-4}}$
1	−10	8	13	8.1	1.0
2	−44	9	14	−23.1	2.8
3	53	9	15	15.7	3.0
4	1	12	16	−7.5	1.5
5	4	12	17	−8.1	1.3
8	26.3	2.3	18	−24.4	1.5
9	−1.9	1.2	19	45.9	1.3
10	−24.4	2.2	20	−26.1	1.2
11	25	6	21	−4.4	1.7
12	−3	5	22	20.5	1.6

replace the least-square fitting. The robust-fitting result is shown as solid line in Figure 5. It is:

$$\delta a = 1.2(\pm 0.4)a - 3.038(\pm 1.5) \times 10^{-3} \tag{6}$$

with a RMSE $\sigma = 1.482 \times 10^{-3}$. The correlation coefficient r, 0.811, is obviously higher than the least-square fitting result and higher than 99.9% confidence level.

3.4.3. Discussion About Cycles 2 and 18

In the last section, we found Cycle 2 and 18 deviate from the linear relationship. In addition, they have other special characteristics.

The curves for Cycle 2 and 18 can be seen in Figure 6. They seem very similar on variation tendency. Because there are only two cycles, we cannot confirm whether it is just a coincidence. We will discuss this problem further in the future.

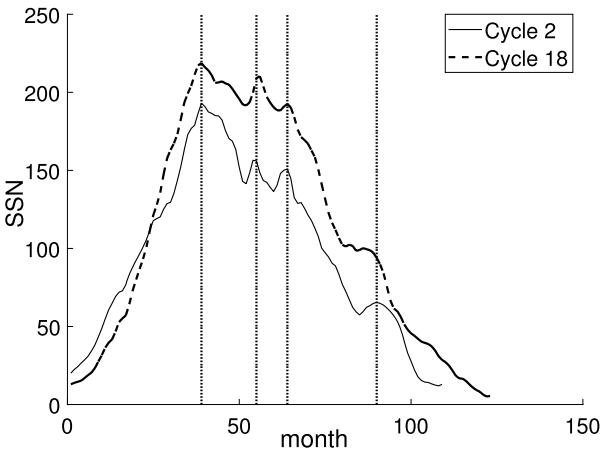

Figure 6 The curves of Cycle 2 and 18 are similar in many ways. There are *four dotted lines*, which correspond to the four peaks of the curve. They all have the primary peak around the 40th month, two secondary peak around the 55th month and the 65th month, and have a little secondary peak at around the 90th month in the falling part. Beside the peaks, the changing trend of the two peak are very similar.

3.4.4. Prediction of a_{24} and b_{24}

Using Equation 6, the δa_{23} could be calculated as

$$\delta a_{23} = 1.175\, a_{23} - 3.038 \times 10^{-3} = -3.227 \times 10^{-4}$$

where a_{23} could be found in Table 3. Then we calculate a_{24} with Equation 4:

$$a_{24} = a_{23} - \delta a_{23} = 2.634 \times 10^{-3}$$

Afterward, we calculate the parameter b'_{24} with Equation 3:

$$b'_{24} = 1.772 \times 10^{-3}\, a_{24} + 2.420 \times 10^{-6} = 7.088 \times 10^{-6}$$

The most probable values of a_{24} and b_{24} for Cycle 24 are

$$a_{24} = 2.634 \times 10^{-3}$$

$$b_{24} = {b'_{24}}^{-\frac{1}{3}} = 52.0584$$

The errors of a_{24} and b_{24} will be discussed in Section 3.6 using the Monte Carlo (MC) method.

3.5. Prediction of c_{24} and d_{24}

In Section 3.4, the parameters a_{24} and b_{24} have been found by two relations (Equation 3 and Equation 6). The other two parameters in the model [c_{24} and d_{24}] should be calculated using another two relations, but it is quite a challenge to find a direct relationship between them. As an alternative, two new parameters were introduced to establish two new relationships for prediction and to find c_{24} and d_{24} indirectly. These two new parameters are $f_{\max}$ and the f_{ratio}, and the two relations are:

i) The linear relationship between the U_{slope} and the $f_{\max}$
ii) The linear relationship between the U_{slope} and the f_{ratio}

We defined them as the third and fourth relationship, where the third relationship is the Waldmeier effect (Waldmeier, 1955).

In order to use these two relations, we first need to define two new parameters. There are several maxima for each cycle, such as the maximum of the moving-averaged data, the average of the primary and secondary peaks, and the maximum of the fitting curve. We have chosen the peak value of the fitting curve as $f_{\max}$, and the f_{ratio} is defined as

$$f_{\text{ratio}} = \frac{f_{\max}}{\textit{Position of the max}} \tag{7}$$

where the *Position of the max* means the month number of $f_{\max}$ in the current cycle.

The definition of the U_{slope} will be introduced in Section 3.5.1. These two relations will be introduced in Section 3.5.2 and Section 3.5.3. In Section 3.5.4, we will give our prediction of parameters c and d, and the most probable curve for Cycle 24.

3.5.1. Definition of the U_{slope}

U_{slope} is the slope of linear fitting of the rising part. Different ranges of the rising part will have different fits. Therefore, a unified fitting range is necessary. For this calculation, the range should satisfy the following three conditions:

i) The end of the range should not be later than the end of second year of current cycle.
ii) The fitting slope of this range should have good relationship with the $f_{\max}$ and f_{ratio}.
iii) The range should be long enough for fitting.

Based on these conditions, establish the following restrictions:

i) The range should end before the 25th month in the cycle.
ii) The range should not be shorter than five months.
iii) the correlation coefficient between the slope of this range and the maximum should be greater than the 99.9% confidence level.
iv) On the basis of the above three restrictions, the correlation coefficient between the rising slope and the maximum should be as large as possible.

According to these restrictions, the possible range of the starting point is from the first to the 20th month, and the possible range of the ending point is from the fifth to the 25th month. We investigate all of the possible combinations of starting and ending points. The total event space was shown in Figure 7. Each small square in the figure represents a possible combination, and the color represents the correlation coefficients between the slopes of this range and the $f_{\max}$. The darkest red square, containing the 13th to the 25th month, is the best fitting range.

3.5.2. Relationship Between U_{Slope} and $f_{\max}$

There is a linear relationship between the rising slope and the $f_{\max}$, which can be seen in Figure 8. This relationship can be described by

$$f_{\max} = 15(\pm 4) \times U_{\text{Slope}} + 85(\pm 24) \tag{8}$$

with RMSE $\sigma = 26$. The correlation coefficient r, 0.877, is higher than the 99.9% confidence level.

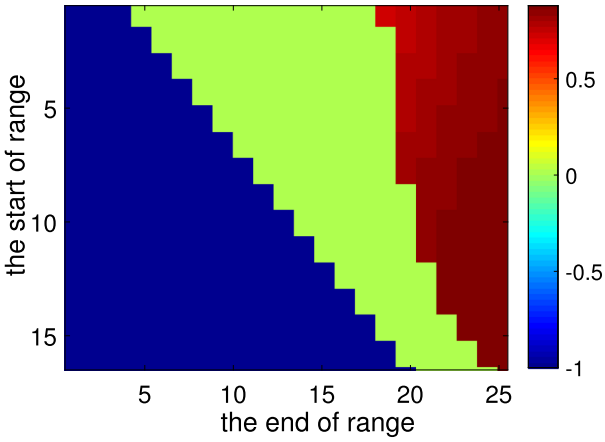

Figure 7 The x-axis is the position of the ending month of the range, while the y-axis is the position of the starting month of the range. A starting time could be combined with each ending and become possible selected ranges (each grid in the figure). The *color reflects the correlation coefficient between the maximum and the slope of this selected range* from Cycle 1 to Cycle 23. Because the starting month should be later than ending month, the *lower-left part of the figure* is impossible and we assign it the value -1 (indicated by *dark blue*). We can calculate the 99.9% confidence line of the possible ranges according to its length. For the range in which the correlation coefficient is less than its 99.9% confidence level, we assign it the value 0 indicated by *green*. The remaining fitting ranges (*red squares*) could be chosen as the range. The *darker area* indicates the higher correlation coefficient.

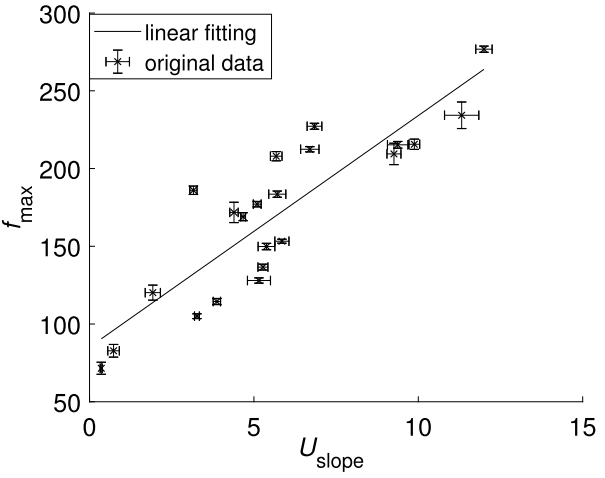

Figure 8 The linear relationship between the U_{slope} and the f_{max}. The x-axis is the *slope* of select section (the 13th to the 25th month) for each of the 23 cycles. The y-axis is the f_{max} for each of the 23 cycles. We have defined the f_{max} in 3.5.

3.5.3. Relationship Between U_{Slope} and f_{ratio}

The relationship between the selected U_{Slope} and the f_{ratio} has been defined in Section 3.5, and can be seen in Figure 9, described by

$$f_{\text{ratio}} = 4.1(\pm 0.7) \times 10^{-1} U_{\text{Slope}} + 1.0(\pm 0.4) \tag{9}$$

with RMSE $\sigma = 0.5$. The correlation coefficient r, 0.938, is higher than the 99.9% confidence level.

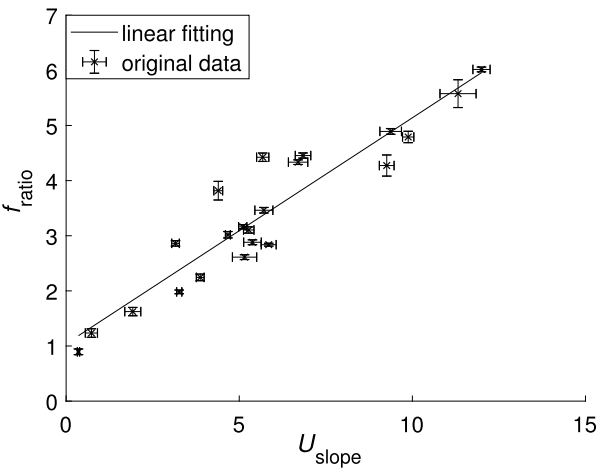

Figure 9 The linear relationship between the U_{slope} and the f_{ratio}. The x-axis is the U_{slope} of each selected section. The y-axis is the f_{ratio}, which has been defined in Section 3.5.

3.5.4. Calculating c_{24} and d_{24}, and Most Probable Curve

We have already found the four parameters for Cycle 24. The next step is to calculate c and d. To simplify the calculation, we can transform the basic function into

$$f(r) = \frac{ab^3r^3}{e^{r^2} - c} \tag{10}$$

where r is

$$r = \frac{t-d}{b}$$

assuming $f'(r) = 0$ and yields

$$c = \left(1 - \frac{2}{3}r^2\right)e^{r^2} \tag{11}$$

replacing c in Equation 10, which yields

$$f_{max} \times e^{r^2} = 1.5ab^3r$$

where r is the variable.

We calculate the numerical solution r of the equation. c can be calculated by Equation 11 and d can be expressed as

$$d = -r \times b + Loc$$

where the Loc is $\frac{f_{max}}{f_{ratio}}$. The most probable c_{24} and d_{24} are

$$c_{24} = -1.176,$$
$$d_{24} = -11.02.$$

Then we obtain the most-probable curve (calculated) for Cycle 24 using a_{24} and b_{24} in Section 3.4.4 and c_{24} and d_{24} in Section 3.5.4. The curve can be seen in Figure 10.

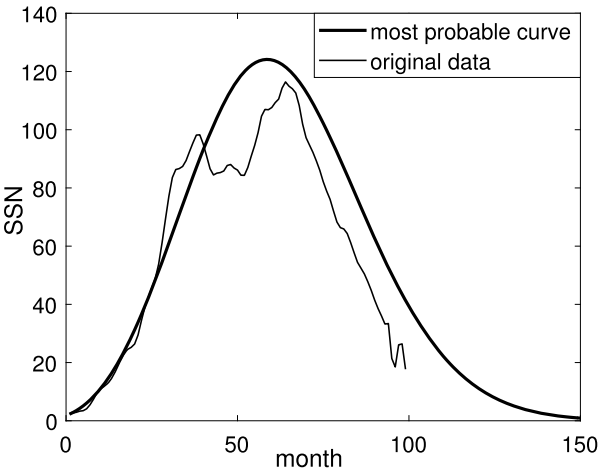

Figure 10 The x-axis is the month of Cycle 24. The y-axis is SSN. The *thin line* is the smoothed monthly averaged SSN data. The *thick line* is the most probable curve we predicted. The real peak appears at the 64th month, with the value of 116.4. The predicted peak appears at the 59th month, with the value of 124.2.

3.6. Predicting the Cycle 24 Curve with the Monte Carlo Method

The most probable curve (calculated) for Cycle 24 has been given in Section 3.5, but the error ranges of the four relationships (Equations 3, 6, 8, and 9) have not been considered. In this section, these error ranges will be used to simulate the possible curves by the Monte Carlo method (MC). Then, from the simulated result, we will obtain the most probable curve (statistics from MC) and the range of prediction curve. The four relationships and their RMSE, which will be used in MC, have been summarized in Equation 12.

$$\begin{aligned} b' &= 1.8 \times 10^{-3} a + 2.4 \times 10^{-6}, \quad \sigma = 1.2 \times 10^{-6} \\ \delta a &= 1.18a - 3.0 \times 10^{-3}, \quad \sigma = 1.5 \times 10^{-3} \\ f_{\max} &= 15 \times U_{\text{Slope}} + 85, \quad \sigma = 26 \\ f_{\text{ratio}} &= 4.1 \times 10^{-1} U_{\text{Slope}} + 1.0, \quad \sigma = 0.5 \end{aligned} \tag{12}$$

3.6.1. The Structure of MC

There are four layers of loops in the MC, and each loop generates one of the four parameters. From Equation 6, we can generate the a_{24} as

$$\begin{aligned} a_{24} &\sim N\left(\mu_{a24}, {\sigma_{a24}}^2\right) \\ \mu_{a24} &= 2.6 \times 10^{-3} \\ \sigma_{a24}^2 &= \left(1.5 \times 10^{-3}\right)^2 = 2.2 \times 10^{-6} \end{aligned}$$

This is the first loop of the MC.

For each randomly generated a_{24}, we calculated the μ_{b24} via Equation 3, and then generate b_{24} as

$$b_{24} \sim N\left(\mu_{b24}, {\sigma_{b24}}^2\right)$$

We set this as the second loop of the MC calculation.

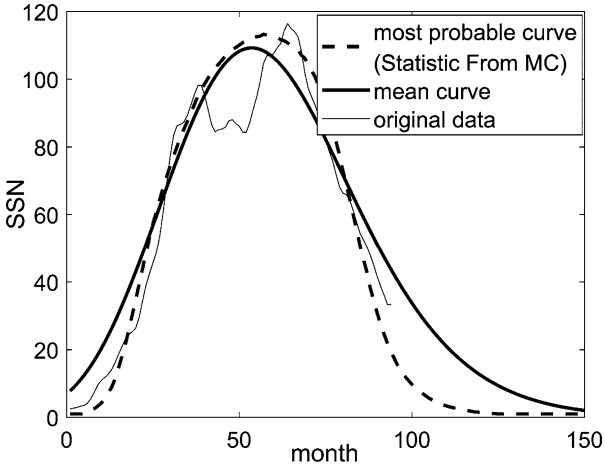

Figure 11 SSN *vs*. month for Cycle 24, showing the original data, the most-probable curve and mean curve from the Monte Carlo calculation.

Then, with Equation 8 and 9, we derive the expectations $[\mu]$ of $f_{\max}$ and f_{slope}, generate them to obey the Gaussian distribution, and set these two ranges as the third and fourth loops of the MC calculation.

If each loop runs 70 times (each parameter generates 70 random numbers), there will be 70^4 (24,010,000) prediction curves.

3.6.2. Eliminating the Unrealistic Curves

The MC method gives us the possible curves for Cycle 24, but some of them are obviously in conflict with the characteristics of the SSN curve. The unrealistic curves have the following characteristics:

i) At least one point on the curve lower than zero.
ii) The first month (the valley value) of the curve has a large value.
iii) The peak is much higher than the real curve (some values even reach 10^3).

Some restriction is needed to eliminate these kinds of curves. The mean value of the peaks μ_{p} of the 23 cycles is 168.2, and the variance σ_{p} is 52.8. The mean value of the valley $[\mu_{\text{v}}]$ of the 23 cycles is 9.7, and the variance σ_{v} is 5.8. It is reasonable to restrict the value of peak and valley with the range of $\mu \pm 3\sigma$. So, we restrict the curve as follows:

i) There are no points less than zeros on the curve.
ii) The peak value of the curve should be within the range of 10.0 to 326.4 ($\mu_{\text{p}} \pm 3\sigma_{\text{p}}$).
iii) The first point and the 151st point (the end position of the longest cycle) of the curve should be within the range of 0 ($\mu_{\text{v}} - 3\sigma_{\text{v}} = -7.8 < 0$) to 27.3 ($\mu_{\text{v}} + 3\sigma_{\text{v}}$).

Finally, 20,685,127 curves were selected from all 24,010,000 curves.

3.6.3. The Most Probable Curve and Mean Curve from Monte Carlo Method

The results will be introduced in two parts. In this section, we will discuss the most probable curve and the mean curve of the MC method.

In the previous section, we selected the possible curves. Taking the average of each month, we obtain the mean curve. Moreover, we extract the SSN values of all possible

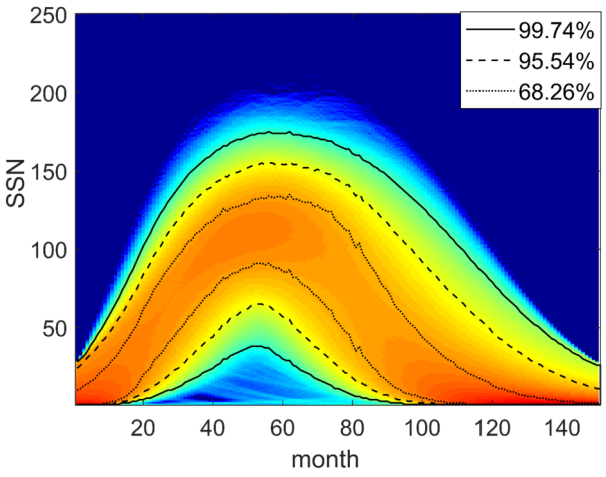

Figure 12 The range of the predicted curve for Cycle 24. The *color* reflects the frequency of the curve passing through the position.

curves monthly, and count their distribution with the bin $\Delta SSN = 1$. Taking the peaks of distribution of every month, we can get the most probable curve (statistics from MC). These two curves can be seen in Figure 11.

In the figure, the peak value of the mean curve appears at the 54th month, with a value of 109.2, and the peak value of the most probable curve (statistics from MC) is at the 57th month with the value of 113.3. The real peak appears at the 64th month, with the value of 116.4. The two predicted peak values are very close to the real value. Meanwhile, we should note that there are two peaks in Cycle 24, and the two predicted peaks position is between the two peaks of original data (at the 39th and the 64th months). Our prediction results are consistent with the changing characteristics of the real curve.

3.6.4. The Range of Prediction from the Monte Carlo Method

The range of the predicted curve can be seen in Figure 12. The color reflects the frequency of the curve passing through the position. We count the distribution of each month. According to the distribution, we can get the 99.74%, 95.54%, and 68.26% confidence interval ($\alpha =$ 0.0026, 0.0446, and 0.3474). The range of peak for the 99.74% line is from 38 to 175. For the 95.54% line, the range is from 65 to 155, and for the 68.26% line, it is from 91 to 134.

3.6.5. Discussion of the Two Kinds of Predicted Curves

In this section, we will discuss the bias, consistency, and correlation of the mean curve and most probable curve (statistics from MC).

By subtracting the real curve, the residuals of the two curves could be obtained, respectively. For the mean curve, the mean value and the variance of the residual are 5.52 and 95, and, for the most probable curve, they are 3.61 and 99. To evaluate them, the residuals between the fitting curve and the real curve for the past 23 cycles are calculated. The mean values and variances of the variance of the residuals for these 23 cycles can be found in Table 4.

The mean values of the variance of the residuals of the for Cycle 23 range from -3.73 to 0.60. The mean values of two predicted curves are out of this range, and even out of the 3 σ range, which means that these two predicted curves are biased. However, obviously, it is

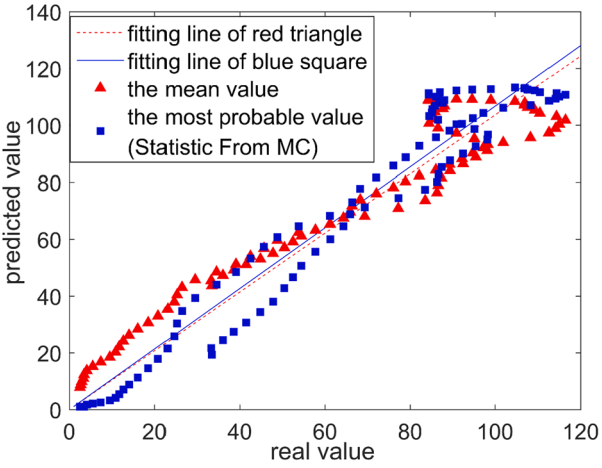

Figure 13 Predicted (mean and most probable) *vs*. real data. The *solid line* and the *dashed line* show the results of linear fits to the most probable value and the mean value, respectively.

too strict to use the statistical results of the fitting residual to judge a prediction. We should notice that the bias is only 3.61 and 5.52, which is acceptable in the prediction. The most probable curve (statistics from MC) has less bias than mean curve.

The variances range from 33 to 378. The median and the mean value are 76 and 121. The variances of mean curves and most probable curve (statistics from MC) are obviously in this range and between the median and the mean value. It can be considered that the two predicted curves agree well with the real curve, and the mean curve is a little bit better than the most probable curve (statistics from MC) at this point.

The linear relationship between the original and predicted data can be seen in Figure 13. The red-triangle-shaped points are the real values *vs*. the mean values, while the blue-square-shaped points are the real values *vs*. the most probable values. The red-dashed line and the blue-solid line are their fitting lines (set the intercept to 0) of the points in the same color. The R-square of the blue line is 0.945, while that of the red line is only 0.880. However, both of them are higher than 99.9% confidence level. So, we consider that the most probable curve (statistics from MC) has better correlation with the real data.

To summarize, comparing with the mean curve, the most probable curve has less bias and correlates better with the real curve. The variation of the residuals between the mean curve and the real curve is less than that of the most probable curve (statistics from MC). We considered that the shape of mean curve is more similar to the fitting curve (least square) of the real curve. Also, the most probable curve reproduces the variation of the real curve better and has less bias.

3.6.6. Verifying the Method with Cycle 22 and 23

In this section, we will verify the applicability and reliability of our method by applying to Cycles 22 and 23. As shown in Figure 14 and Figure 15, the MC predicted results and original data have been given.

For Cycle 23 in Figure 14, the predicted curve is slightly different from the original data. The differences include the slightly lower peak and the slightly earlier peak position. However, the difference in height of the peak is no more than 15, which is acceptable in the prediction for sunspot number. Moreover, including the rising part and the peak part, the MC 68.26% confidence interval, which means there is 68.26% probability that the real curve will fall in this interval, nearly covers the total observed data.

Table 4 The mean values and variances of residuals between the fitting curves and real curves of 23 cycles.

Cycle	Mean	Variance of residuals	Cycle	Mean	Variance of residuals
1	0.04	69	13	−1.87	57
2	−0.13	102	14	0.60	45
3	−3.70	345	15	−0.47	111
4	−3.73	378	16	0.17	46
5	0.34	35	17	−1.04	76
6	−0.54	33	18	−0.44	82
7	0.59	35	19	−2.91	93
8	−2.23	278	20	−0.75	74
9	−0.81	269	21	−0.54	69
10	−1.24	213	22	−0.88	114
11	−2.03	141	23	−0.50	54
12	0.26	75			

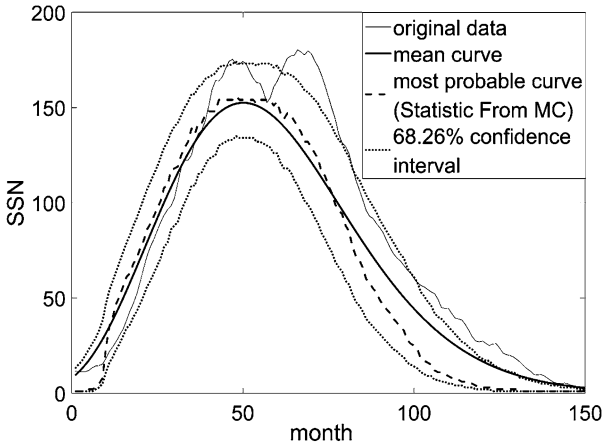

Figure 14 Comparison of the real and predicted SSN curves for Cycle 23.

For Cycle 22 in Figure 15, the result is even better than for Cycle 23. The predicted curve perfectly reproduces the real curve, and nearly all of the parts of the curve are in the interval of 68.26% confidence. The little part out of this interval is between the main peak and secondary peak. However, what was given by our method is the predicted interval of the fitting curve, and this part of the fitting curve still falls in the interval.

4. Discussion and Conclusion

i) A function with four parameters was used to describe each cycle, so, to predict the curve of Cycle 24, we found four relationships, which are Equations 3, 6, 8, and 9, which have been summarized in Equation 12

ii) From the above equations, we found four parameters for Cycle 24, and then we obtain the most probable predicted curve for Cycle 24. The peak of this curve is at the 59th month, with the value of 124.2. The real peak appears at the 64th month, with the value of 116.4.

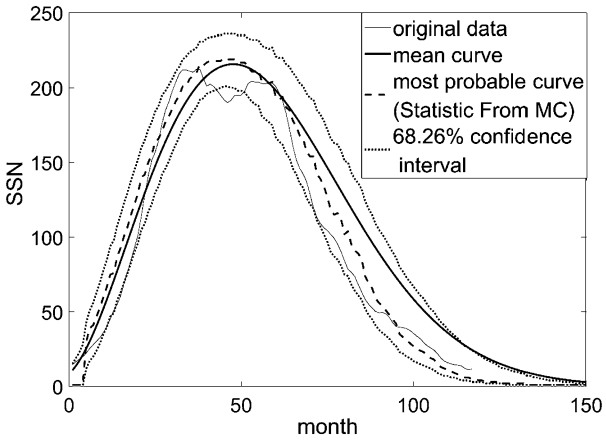

Figure 15 Comparison of the real and predicted SSN curve for Cycle 22.

iii) Using the Monte Carlo method and the error ranges of the four relationships, we predicted the mean curve and the most probable curve for Cycle 24. The peak of the mean curve is at the 54th month, and the peak of the most probable curve is at the 57th month, which are both between the primary and the secondary peaks of the real curve. The peak value of the most probable curve is 113.3, a little bit less than the value 116 of the real peak. The peak value of the mean curve is 109.2, which is less than that of the most probable curve but higher than the secondary peak of the real curve.

iv) The range of peak values for the 99.74% line is from 38 to 175. For the 95.54% line, the range is from 65 to 155, and, for the 68.26% line is from 91 to 134.

v) We found two similar cycles (Cycle 2 and Cycle 18), but we lack a reasonable explanation for the behavior of those two cycles (Section 3.4.3). They will be studied further.

vi) In the last two relationships, we used the 13th to the 25th months of the current predicted cycle, so the model was unable to give predicted results before the end of the second year.

vii) The prediction model that we use established by the four factors is effective. We considered that the shape of the mean curve is more similar to the fitting curve (least square) of the real curve. Also, the most probable curve (statistics from MC) follows the changing detail of the real curve better and has less bias.

viii) We have tested the model, which contains four parameters, through the data of Cycle 24. For the next step, we will use this method on Cycle 25.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant No. 11203004) and BNU Interdisciplinary Research Foundation for the First-Year Doctoral Candidates (Grant No. BNUXKJC1812). The data that we used are published by SILSO.

Disclosure of Potential Conflicts of Interest The authors declare that they have no conflicts of interest.

References

Aguirre, L.A., Letellier, C., Maquet, J.: 2008, Forecasting the time series of sunspot numbers. *Solar Phys.* **249**, 103. DOI. ADS.

Casas, R., Vaquero, J.M., Vazquez, M.: 2006, Solar rotation in the 17th century. *Solar Phys.* **234**, 379. DOI. ADS.

Clette, F., Lefèvre, L.: 2016, The new sunspot number: Assembling all corrections. *Solar Phys.* **291**(9), 2629. DOI. ADS.

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. ADS.

Clette, F., Cliver, E.W., Lefèvre, L., Svalgaard, L., Vaquero, J.M.: 2015, Revision of the sunspot number(s). *Space Weather* **13**, 529. DOI. ADS.

Gholipour, A., Lucas, C., Araabi, B.N., Shafiee, M.: 2005, Solar activity forecast: Spectral analysis and neurofuzzy prediction. *J. Atmos. Solar-Terr. Phys.* **67**, 595. DOI. ADS.

Hathaway, D.H., Wilson, R.M., Reichmann, E.J.: 1994, The shape of the sunspot cycle. *Solar Phys.* **151**, 177. DOI. ADS.

Kakad, B.: 2011, A new method for prediction of peak sunspot number and ascent time of the solar cycle. *Solar Phys.* **270**, 393. DOI. ADS.

Lasheng, Z., Li, G., Haijuan, Z., Liuming, H.: 2005, The shape of sunspot cycles described by monthly sunspot areas. *Solar Phys.* **232**, 143. DOI. ADS.

Leach, R.D.: 1996, Spacecraft system failures and anomalies attributed to the natural space environment. *Space Programs and Technologies Conf.* DOI.

Li, Y.: 1997, Predictions of the features for sunspot cycle 23. *Solar Phys.* **170**(2), 437. DOI. ADS.

Li, K.: 1999, The shape of the sunspot cycle described by sunspot areas. *Astron. Astrophys.* **345**(345), 1006.

Marvin, D.C., Gorney, D.J.: 1992, Solar proton events of 1989: Effects on spacecraft solar arrays. *NASA STI/Recon Technical Report* **93**. ADS.

Pesnell, W.: 2016, Predictions of solar cycle 24: How are we doing? *Space Weather* **14**(1), 10. DOI.

Schwabe, H., Schwabe, H.: 1844, Sonnen-beobachtungen im jahre 1843. *Astron. Nachr.* **21**(15), 234. DOI.

Seber, G.A.F., Wild, C.J.: 1989, *Nonlinear Regression*, *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*, Wiley, New York. 0-471-61760-1. DOI.

SILSO: 2015, *The international relative sunspot number*, www.sidc.be/silso/datafiles. Accessed 20 November 2017.

Toth, G., de Zeeuw, D.L., Gombosi, T.I., Manchester, W.B., Ridley, A.J., Roussev, I.I., Sokolov, I.V.: 2005, Sun-to-Thermosphere Simulation of the October 28, 2003 Event With the Space Weather Modeling Framework. *AGU Fall Meet. Abs.* DOI. ADS.

Waldmeier, M.: 1955, *Ergebnisse und Probleme der Sonnenforschung*, Geest and Portig, 1955. 2. erweiterte Aufl., Leipzig. ADS.

Wang, J., Miao, J., Liu, S., Gong, J., Zhu, C.: 2008, Prediction of the smoothed monthly mean sunspot numbers for solar cycle 24. *Sci. China Ser. G, Phys. Mech. Astron.* **51**, 1938. DOI. ADS.

Williamson, M.: 2006, *Spacecraft Technology: The Early Years*, *History of Technology*, Institution of Engineering and Technology, London. DOI.

Wolf, R.: 1861, Abstract of his latest results. *Mon. Not. Roy. Astron. Soc.* **21**, 77. DOI. ADS.