Prediction of solar activity for the next 500 years

Friedhelm Steinhilber and Jürg Beer

Received 18 May 2012; revised 18 February 2013; accepted 2 March 2013; published 23 May 2013.

Recently, a new low-noise record of solar activity has been reconstructed for the past 9400 years by combining two $^{10}$Be records from Greenland and Antarctica with $^{14}$C from tree rings [Steinhilber et al., 2012]. This record confirms earlier results, namely, that the Sun has varied with distinct periodicities in the past. We present a prediction of mean solar magnetic activity averaged over 22 years for the next 500 years mainly based on the spectral information derived from the solar activity record of the past. Assuming that the Sun will continue to vary with the same periodicities for the next centuries, we extract the spectral information from the past and apply it to two different methods to predict the future of solar magnetic activity. First, the two methods are tested by predicting past changes. Our methods are able to predict periods of high and low solar activities for a few centuries in the past. However, they are less successful in predicting the correct amplitude. Then, the methods were used to predict the period 2000–2500. Both methods predict a period of low activity around 2100 A.D. Between 2100 and 2350 A.D., the results are inconsistent regarding the duration of the low-activity state in 2100 A.D. and the level of activity until 2250 A.D. Around 2250 A.D., both methods predict a period of moderate activity. After 2350 A.D., both methods point to a period of high activity. The period of high activity will end around 2400 A.D. and will be followed by a period of moderate activity.


1. Introduction

[2] The Sun plays a fundamental role in life on Earth. The electromagnetic radiation emitted by the Sun is our primary source of energy. The uneven spatial distribution of the heating by that radiation over the Earth’s surface leads to temperature gradients, which drive the climate system. The Sun also emits continuously and sometimes explosively plasma from its surface. This plasma, called solar wind, carries solar magnetic fields. It forms the heliosphere, which acts as a magnetic shield against the galactic cosmic rays, reducing the cosmic ray intensity reaching near-Earth space. The intensity decreases a second time due to the shielding effect of the geomagnetic field. When the solar wind approaches the Earth, it interacts with the geomagnetic field and the atmosphere, inducing electrical currents and producing auroral lights at high latitudes.

[3] It has been known since ancient times that the Sun is a variable star. The number of sunspots varies with the so-called Schwabe cycle ranging in duration from 8 to 14 years [Eddy, 1976]. This raises the important question of to what extent solar variability affects the Earth and its immediate vicinity. A quantitative answer is difficult and requires not only a detailed knowledge of the past variability but also a better understanding of the complex nonlinear interactions and the involved feedback mechanisms. So far, there is growing evidence that solar variability has been an important factor among others in past climate change [Gray et al., 2010; Lockwood, 2012; Hegerl et al., 2007].

[4] Solar flares, which are sudden releases of huge amounts of energy on the Sun, can lead to Earth-directed coronal mass ejections which can cause dramatic magnetic storms on Earth with potentially devastating effects on power grids, communication satellites, and GPS systems. This topic is a new and not yet fully acknowledged thread for the human society that grows rapidly with the progress in modern technology vulnerable to such space events.

[5] It is therefore not surprising that the fields of solar-terrestrial relationship, space weather, and space climate got increasing attention in recent years [Schwenn, 2006; Lockwood et al., 2009; Solanki et al., 2004]. An important task with large implications for space research and risk management is the development of tools to predict the future of solar activity. So far, most attempts have concentrated on predicting amplitude and duration of the next 11 year Schwabe cycle [Dikpati et al., 2006; Pesnell, 2008; Petrovay, 2010]. These attempts basically failed probably due to the fact that the solar dynamo has chaotic properties, at least as quantified by sunspot numbers [Charbonneau, 2010]. A similar argumentation against long-term prediction is given by Solanki and Krivova [2011], who argue that
predicting the solar activity is unreliable because of the nonlinearity of the solar dynamo responsible for the solar activity. However, this study does neglect that the long-term records of solar activity derived from cosmogenic radionuclides reveal cycles with well-defined decadal to centennial periodicities but variable amplitudes (e.g., 88 year Gleissberg and 208 year de Vries cycles) [Damon and Sonett, 1991; Feynman and Ruzmaikin, 2001; Knudsen et al., 2009]. These periodicities point to the existence of a deterministic component in the long-term solar activity and therefore provide the means for prediction.

[6] A similar situation is known from weather forecasting. Predicting the daily weather is very complex and only possible for a few days. However, the average weather over a long period, which is the climate, can be predicted more reliably based on, for example, the annual orbit of the Earth around the Sun (seasons) and the planetary effects on the orbital parameters (eccentricity, obliquity, and precession) according to the Milankovitch theory (glacial/interglacial periods).

[7] A few predictions of long-term solar activity exist [e.g., Abreu et al., 2010; Barnard et al., 2011; Bonev et al., 2004; Lockwood, 2010; Lockwood et al., 2011]. All these studies are based either on the 400 year long sunspot record or on millennia long solar activity records derived from cosmogenic radionuclides ($^{10}$Be and $^{14}$C). From these data, information about solar variability in the past is extracted. Assuming that the Sun will not change its behavior, the future activity can be predicted. However, the studies differ in the way they extract and use the information.

[8] The study of Bonev et al. [2004] analyses sunspots as well as data of the cosmogenic radionuclide $^{14}$C. By superimposing the prominent 88 year cycle (Gleissberg) and the 208 year cycle (de Vries), this study predicts a period of low activity at the end of the 21st century. The studies [Lockwood, 2010; Lockwood et al., 2011; Barnard et al., 2011] are based on a solar activity reconstruction mainly derived from the cosmogenic radionuclide $^{10}$Be. By applying superposed epoch analyses to the solar activity after a grand solar maximum, these studies come to the conclusion that a Maunder Minimum–type grand solar minimum will occur in roughly 40 years from now with a probability of 8% (rising to 50% in about 100 years). The study [Barnard et al., 2011] applies the prediction of solar activity to determine future occurrence rates of solar energetic particle events. Abreu et al. [2008, 2010] use two independent approaches to predict the future by analyzing the same $^{10}$Be record as Lockwood [2010], Lockwood et al. [2011], and Barnard et al., [2011]. The first approach predicts the future activity using the waiting-time distributions of the occurrence of grand maxima and minima. They find that the current state of high solar activity will end soon and calculate that there is a high probability that the next grand minimum will occur around the year 2100 A.D. The second approach superposes the most important periodicities in the data to predict the future. This approach predicts a period of low activity around the year 2100 A.D., which confirms the first approach.

[9] The different studies generally agree in predicting a period of low solar activity during the 21st century in spite of the fact that all these studies are based on single individual records of cosmogenic radionuclides. As has been shown recently, about 70% of the variability of a radionuclide record reflects solar activity, while the remaining 30% can be attributed to climate effects [Steinhilber et al., 2012]. By combining several radionuclide records, this noise level has been reduced [Steinhilber et al., 2012].

[10] Here we follow the approaches by Abreu et al. [2010] and Bonev et al. [2004] and make use of the spectral information contained in the past solar activity to predict the future 500 years. The low-noise record [Steinhilber et al., 2012] based on several radionuclide data sets ($^{10}$Be from Greenland and Antarctica and $^{14}$C from tree rings) is used as the record of solar activity. This record has less noise than the individual records used in previous predictions and therefore can be considered as an improvement of these former studies.

[11] Finally, we would like to note that the term “solar activity” is not well defined. While in the past it was often considered as a synonym for sunspots, we use it in this work as a measure of the general solar magnetic activity, which includes both the closed magnetic field associated with sunspots and the open magnetic field responsible for the modulation of the galactic cosmic rays and, therewith, the production rate of cosmogenic radionuclides.

[12] This paper is structured as follows. In section 2, we shortly present the solar activity reconstruction and its properties. The two prediction methods that are the basis of our approach are described in section 3. In section 4, first, the prediction methods are tested on data from the past and then are applied to predict solar activity for the next 500 years. Section 5 gives a discussion, and in section 6, some conclusions are drawn.

2. Data

[13] To make a long-term prediction of solar activity, long time series of past solar activity are required. The longest continuous direct record of solar activity, the sunspot record, is not suitable because it covers only the past ~400 years. In addition to sunspots, cosmogenic radionuclides are also commonly used to reconstruct solar activity [Beer et al., 2011]. Cosmogenic radionuclides are produced by the interaction of cosmic ray particles with the gases of the Earth’s atmosphere [Masarik and Beer, 2009]. After production, the radionuclides are transported to ground where they are stored. The two most prominent radionuclides suitable for reconstructing solar activity are $^{14}$C stored in tree rings and $^{10}$Be extracted from polar ice cores [Beer et al., 2011]. Solar activity has been reconstructed for the past ~10,000 years [Beer et al., 2011; Solanki et al., 2004; Steinhilber et al., 2008, 2012; Usoskin et al., 2007; Vonmoos et al., 2006] from these radionuclides. The radionuclide signal is composed of two production components (solar and geomagnetic activities) and a system component (climatic noise). On a decadal time scale, up to 30% of the radionuclide signal may be due to the system effects. The system effects are mostly unknown and cannot be removed from a single radionuclide record. However, system effects, for example, induced by regional climate fluctuations can be reduced by combining several radionuclide records and applying principal component analysis [Abreu et al., 2012a; Steinhilber et al., 2012].

[14] Here we use the solar activity reconstruction obtained by Steinhilber et al. [2012]. This reconstruction consists of 22 year averages and covers the period 9400 to ~38 B.P. (1988 A.D.). The reconstructed parameter is the solar
modulation potential \( \Phi \), which describes the cosmic ray intensity modulation in the heliosphere due to the variable interplanetary magnetic field.

[15] Figure 1 shows the data, including results from Fourier and wavelet analyses. Fourier analysis shows that the long-term solar activity has varied with distinct periodicities. Fourier analysis also shows that the main spectral lines are very sharp, representing well-defined periodicities, and that the phase is preserved during periods of low activity. This stability of periodicities and phases over the past 9400 years leads us to assume that the Sun will continue to vary with the same periodicities for the next centuries. However, the amplitudes of these distinct periodicities are modulated, as shown by the wavelet analysis. This makes it difficult to reliably predict the amplitude of future solar activity.

3. Prediction Methods

[16] We used two different prediction methods. Both methods are based on the spectral information obtained from the 9400 year long solar activity reconstruction, assuming that the periodicities found in the past will also exist in the near future (500 years). The two methods differ in the way they deal with the observed amplitude modulation. The first method neglects the amplitude modulation. It calculates the power spectrum from data of the past and calculates the prediction by superposing the important periodicities that are the periodicities with the largest average amplitudes, presuming that the phase of each periodicity is continuous across the meeting of the past and the future. The second method makes a wavelet decomposition of the original time series and predicts the single wavelet scales individually using an autoregressive (AR) model. This method considers the amplitude modulation and therefore, in principle, should be better than the first method. However, it has several free unknown parameters that have to be determined numerically which introduce some uncertainty.

3.1. Method 1: Constant Amplitudes—FFT Method

[17] We used Fourier spectrum analysis [fast Fourier transformation (FFT)] to find the most prominent lines (periodicity, amplitude, and phase) in the data. We applied FFT in a stepwise manner, searching for the periodicity with the largest amplitude and its phase angle. Then, this line is removed from the data. This procedure is repeated until a predefined number of lines have been reached. The number of lines is a free parameter in this method, which varies between 10 and 100. We found that the prediction depends only weakly on the number of lines if more than 10 lines are used. A second free parameter is the length of the time window, which serves as the calibration period. In the following, this method will be called the FFT method.

3.2. Method 2: Modulating Amplitudes—WTAR Method

[18] The second prediction method makes use of the wavelet decomposition method combined with an AR model. We refer to references [Murtagh et al., 2004; Renaud et al., 2002] for a detailed explanation of the method. In brief, the time series is decomposed in \( J \) wavelet scales. Then, an AR model of order \( N \) is applied to the individual scales. This implies that \( J \times N \) model coefficients of the AR model have to be determined in the calibration window. To predict the signal at time \( t+1 \), the scales are individually predicted using their individual AR model coefficients and finally summed up. To calculate the signal at \( t+2 \), the AR coefficients are recalculated considering the new data at time \( t+1 \). Free parameters of this method are the number of scale \( J \), the order of the AR model \( N \), and the time window of the calibration. We used different combinations of the parameters \( N \) and \( J=(N=2,J=8),(3,7),(3,6),(2,9),(4,6) \)—and different calibration windows. As will be shown below, the calibration window length has only little influence on the results. In the following, this method will be called the WTAR method.

Figure 1. Solar activity over the past 9400 years (solar modulation potential \( \Phi \)). Time is given in years B.P. (with B.P. = before 1950). (a) Time series, (b) wavelet spectrum, (c) Fourier power spectrum, and (d) zoom-in of Figure 1c.
4. Results

4.1. Tests of the Methods’ Reliability

[19] An important advantage of our time series of solar activity (Figure 1a) is that it is long enough to test our methods. It offers the opportunity to use different time windows in the past to calibrate our method and to predict the remaining part of the 9400 years. Before applying the methods, first, we subtracted the linear long-term trend from the data to make them more stationary and then interpolated the data to a resolution of 1 year in the case of the FFT method and approximately 4.56 years in the case of the WTAR method. The value of 4.56 years is due to the fact that the WTAR method uses $2^{11} = 2048$ data points. Our main goal is to predict solar activity for the next several hundred years ($\leq 500$ years). To be consistent with this goal, we tested the ability of our methods to predict a variety of periods of up to 500 years in the past. The predictions were compared with the observations by calculating the $r$. The correlation $r$ was calculated for different prediction periods: 50, 75, 100, 200, and 500 years. The temporal evolution of $r$ (exemplarily shown for the 500 year prediction in Figure 2) was obtained for both methods [FFT method (Figure 2a) and WTAR (Figure 2b)] by moving a calibration window of 6000 years stepwise through the 9400 years of data and predicting for each step the following 500 years for the different method parameters.

[20] We repeated the test of the method’s reliability in the past with normally distributed random numbers with the same mean value and variance as the observational data using the same time windows for the calibration and method parameters as before for the predictions with real data. The results for the 500 year predictions are also shown in Figure 2. As expected, the average correlation for the random data is zero for both methods. In the case of the FFT method, the correlation coefficients are in the range of $-0.5$ to $0.5$ for the 50 year prediction and get smaller for longer predictions (not shown). The coefficients vary in the range of $-0.1$ to $0.1$ for the 500 year prediction. The distribution of $r$ is wider for the WTAR method, where the correlation coefficients vary almost in the entire range, i.e., from $-1$ to 1. We applied the Kolmogorov-Smirnov test to determine whether the distributions of the predictions based on observational data are equal to the distributions of the predictions based on random data. The alternative hypothesis is that the distributions of the predictions based on observational data are shifted to larger values (i.e., positive correlation) compared to the predictions based on random data. For all different parameter sets and for the two methods, the predictions based on observational data are significantly different from the random data. This finding implies that both methods are suitable to predict solar activity. We note that the correlation coefficients are on average ~0.3 for the 500 year predictions, which is significantly less than 1 (perfect correlation). This result indicates that our methods are able to predict the trend in solar activity reasonably well (positive $r$) but less so the amplitude. Although most of the time the correlation is positive, there are a few periods with negative correlation coefficients. First, we compared the time series of correlation coefficients $r$ from the two methods with each other, but

Figure 2. (a) Comparison of solar activity with the 500 year predictions of the two methods. Time is given in years B.P. (with B.P. = before 1950). Figures 2a and 2c show the correlation coefficients between the data and the predictions of the two methods as a function of time. Grey areas represent the range of the results using different model parameters. The black curves are obtained using random data with the same mean and variance as the solar activity record, as shown in Figure 2a, for the calibration of the methods. On average, the correlations between observation and prediction are positive and significantly different from the random-data-based predictions. (b) Using the FFT method with a calibration window length of 6000 years and a line number varying between 20 and 100. (c) Using the WTAR method with a calibration window length of 6000 years and a varying number of scales and orders of the AR model.
we could not identify any clear relationship between the results of the two methods. Second, we investigated whether changes in the correlation coefficients coincide with changes in solar activity, i.e., grand solar minima and maxima. No clear relationship could be found.

[21] Examples of how the predictions compare with the observational data are shown in Figure 3. Figure 3a shows the 500 year predictions calibrated in a window with the most recent point in 495 B.P. The prediction covers the period of 473 to -27 B.P. (the last point in the solar activity record). This prediction shows agreement with the observational data. The three grand solar minima, namely, Maunder, Dalton, and Gleissberg, are predicted. The timing of these minima is correct, and also, the amplitudes agree reasonably well, especially in the case of the FFT method. The present period of a grand solar maximum observed since the 1950s is also predicted. The WTAR method predicts generally smaller amplitude variations than the FFT method. In Figure 3a, the correlation coefficients of the 500 year predictions with the solar activity record are on average $r=0.42$ (FFT) and $r=0.51$ (WTAR). These results are encouraging, regarding that the predictions were calibrated in a window until 495 B.P. The second example (Figure 3b) shows the prediction starting in 754 B.P. (most recent calibration point is 1084 B.P.). As can be also seen in this case, both methods agree with the timing of changes in solar activity. The two grand solar minima, namely, Wolf and Spörer, show up clearly. While the predicted amplitudes of the FFT method are of the same order as the observation, the amplitudes predicted by the WTAR method are smaller. These results are consistent with the correlation coefficients for the 500 year predictions and the solar activity record that are much lower for the WTAR method compared to the FFT method ($r=0.33$ for FFT and $r=0.11$ for WTAR). The third example (Figure 3c) shows the prediction starting in 1062 B.P. (most recent calibration point is 1084 B.P.). As before, the predictions of both methods are positively correlated with the solar activity record. The correlation coefficients for the 500 year prediction are $r=0.43$ (FFT) and $r=0.29$ (WTAR). The timing of activity changes is similar in the predictions and the solar activity record. The amplitudes of solar activity are better predicted by the FFT method than by the WTAR method. Overall, the three examples show that our methods are more successful in predicting the shape of the solar activity than its amplitude.

4.2. Prediction of Solar Activity, Solar Forcing, and Cosmic Ray Intensity for the Next 500 Years

[22] For the prediction of solar activity for the next 500 years, we have chosen the same parameters used in the previous section. The most recent 22 year average time point in the solar activity reconstruction is centered at 1977 A.D. Thus, future means beyond this point. We used again two calibration windows of length 4000 and 6000 years, resulting in several predictions for both methods. By combining all predictions, we constructed bands covering the full range of values at every point in time.

[23] Figure 4 shows the prediction bands for the next 500 years. Both methods predict a solar activity minimum around 2100 A.D. (-150 B.P.). The level will be comparable to the Dalton minimum. However, as indicated in the figure, the timing of the future solar activity is better constrained than its level. The minimum around 2100 A.D. is followed by a 200 year long period for which the

Figure 3. Comparison of observation and 500 year predictions in the last millennium. The FFT method predictions were obtained using different numbers of lines and calibration window lengths, and the WTAR method predictions were obtained using different combinations of scales and orders of the AR model and calibration window lengths. Time is given in years B.P. (with B.P. = before 1950). The range of the results for the FFT method is indicated by bright grey and that for the WTAR method is indicated by dark grey. (a) From 473 B.P. to -27 B.P. (1977 A.D.), (b) from 754 to 254 B.P., and (c) from 1062 to 562 B.P. Grand solar minima are marked in the top panel: O: Oort, W: Wolf, S: Spörer, M: Maunder, D: Dalton, G: Gleissberg.
predictions differ. The WTAR method predicts that the minimum around 2100 A.D. will last about 50 years with a subsequent period of higher activity. In the case of FFT, the solar minimum lasts until about 2200 A.D. and is then followed by a period of moderate activity. Around 2400 A.D., the two predictions agree again, pointing to the occurrence of a period characterized by higher activity.

5. Discussion

Recently, Abreu et al. have put forward the hypothesis that the planets exert a torque on the tachocline of the Sun which modulates the decadal to centennial cycles observed in the cosmogenic radionuclides records [Abreu et al., 2012b; Charbonneau, 2013]. If correct, the well-defined periodicities of these cycles justify the methods applied here to make predictions. However, the varying amplitudes remain a serious source of uncertainty.

As has been shown by [Frohlich, 2009; Lockwood, 2002; Lockwood and Stamper, 1999], the solar modulation potential is related to total solar irradiance (TSI) via the strength of the interplanetary magnetic field. The relationship has been applied to derive TSI for the past [Steinhilber et al., 2009, 2012]. Now the same relationship can be used to derive TSI for the future. The prediction of TSI is given in Figure 4 on the right y axis. Note that TSI is given as the difference to the value of the PMOD composite during the solar cycle minimum of the year 1986 A.D. (1365.57 W m⁻²) [Frohlich, 2009].

If one is interested in the effect of charged particles either from the galaxy (galactic cosmic rays) or from the Sun (solar energetic particles) at Earth, the shielding of the geomagnetic field has to be taken into account. The geomagnetic field strength has also changed in the past, generally showing a decreasing trend during the past century [Gubbins et al., 2006]. However, it seems impossible to accurately determine its future change due to the large uncertainty in geodynamo models and the input data [Hulot et al., 2010; Lhuillier et al., 2011]. Generally, a decrease in the geomagnetic field strength amplifies the increase in cosmic ray intensity due to low solar activity and attenuates the decrease in cosmic ray intensity due to high solar activity. Finally, we note that the prediction of total solar irradiance is not influenced by future changes in the geomagnetic field strength.

6. Conclusions

Based on the past millennia of solar magnetic activity derived from cosmogenic radionuclides, our two methods predict a clear decrease in solar activity, reaching a minimum comparable to the Dalton minimum around 2100 A.D., in good agreement with previous predictions. This minimum will be followed by a slow more or less steady increase until 2400 A.D. As a consequence, the increase of global warming will be slightly attenuated until 2100 A.D. However, the subsequent increase in solar activity will further enhance the global warming.

This prediction of the solar modulation potential Φ can be used as the basis to estimate further parameters of solar activity such as spectral solar irradiance [Shapiro et al., 2011; Thuillier et al., 2012; Vieira and Vieira, 2011], the cosmic ray intensity, the occurrence rates of solar energetic particle events [Barnard et al., 2011], the sunspot number [22, 24, 38], and the interplanetary magnetic field [Barnard et al., 2011; Steinhilber et al., 2010; Usoskin et al., 2002; Vieira and Solanki, 2010].

The 500 year predictions of solar modulation potential and of total solar irradiance will be available online at the NOAA paleo server (http://www.ncdc.noaa.gov/paleo/forcing.html).
Acknowledgments. We thank two anonymous reviewers for their useful comments. We acknowledge financial support by NCCR Climate the Swiss climate research and by the Swiss National Science Foundation under grant CRSI122-130642 (FUPSol).

References


