Sunspot Number Prediction by an Autoregressive Model

R.Werner

Space Research & Technology Institute, Stara Zagora Department, Bulgarian Academy of Sciences

E mail: rolwer52@yahoo.co.uk

Accepted: 27 June 2012

Abstract. The prediction of the solar activity is very important and a lot of methods for the solar activity forecasting were developed because of their high relevance. Regardless of the advance in the application of physical methods for the purpose of forecasting, the results are very inconsistently spread and substantiate the application of statistical methods. In this paper using the annual sunspot number (SSN) data set for the time period of 1749 till 2010, an autoregressive model was developed, based on the Box-Jenkins methodology. A model of ninth order was obtained. Forecasts of the solar maximum and the moment of its expectation were calculated starting from 2006 up to 2010, with the data endpoints of 2005 and 2009 respectively. All obtained SSN maxima are in the range from 110 down to approximately 90. The moments when the maxima are reached increase from 2011 up to 2013 with the duration of the unexpectedly long solar minimum. Both forecasts using data up to 2009 and 2010 are very close to each other. It is expected that the SSN in the maximum in 2013 will be about 90. However the confidence band is very wide. The limits at the significance level of 0.1 are about +77/-53. The prognoses errors rapidly increase with the increasing prediction horizon. Therefore the reasonable prediction horizon is strongly limited to two to three years. For greater prediction intervals the errors are non-acceptable. This statement is in good agreement with the findings of Rozelot's study on the dynamical properties of the sunspot time series, that the accurate forecasting over a period of time longer than 2 to 4 years seems impossible.

Keywords prediction of the solar activity, statistical methods, autoregressive model

Introduction

The solar activity forecast of the next solar cycle is important for satellite drag, telecommunication outages and hazards in connection with the occurrence of strong solar wind streams producing the blackout of power plants. Also for manned space flights the prediction of the radiation risk is a requirement for a successful mission. High powerful radiation can lead to computer and computer memory upsets or failures.

As an indicator of the solar activity usually the sunspot number (SSN) is taken. It's prediction plays a crucial role also in the climate debate, where some believe that the climate change is dominated by solar variations including the modern industrial time and that the next solar cycles will be very similar to the Maunder Minimum (Schatten and Tobiska, 2003). Long solar cycles are the Gleissberg Cycle, de Vries Cycle (also called Suess Cycle) and Hallstatt Cycle (Bonev, Penev, Sello, 2004). These cycles are quasi cycles with periods of approximately 70-100, 150-230 and 2200-2400 years, respectively. Some researchers expect a deep minimum by coincidence of different long solar cycles minima during the next solar cycles. The study of long period solar activity cycles have shown that the Gleissberg cycle has a wide frequency band with a double structure consisting of 50–80 years and 90–140 year periodicities. The cycle known as the de Vries cycle is less complex showing a variation with a period of 170–260 years (Ogurtsov et al., 2002). The periods and amplitudes are very different from cycle to cycle and therefore the occurrence of the maximum/minimum is not predictable by simple multiperiod analysis. Kane pointed out that the a simple extrapolation, especially of short cycles (less than _100 years) during the present transient epoch of the 2200–2400 year cycle for solar activity prediction, is a risky procedure (Kane, 2008).

The solar cycle prediction methods can be classified in two basic categories (Hattaway, Wilson, Reichmann, 1999). The first one comprise regression techniques (including auto-regressive methods, neural networks and curve fitting) and the second one include different precursor techniques (a combination of sunspot indicators and geomagnetic field indicators (see also Baranovski, 2008). The last methods do not have a physical basis. Newer predictions are made with the help of magneto-hydrodynamic models, which describe the development of the solar activity.

Pesnell summarized more than 50 predictions of the solar cycle 24 (of its amplitude Rn – where R is the SSN and n the solar cycle number - and its occurrence in time for different categories (Pesnell, 2008). It is obvious that the predictions based on dynamo models, on recent climatology forecast, on neural networks and predictions using geomagnetic precursors are in the range of approximately 130-145 and significantly higher than the mean amplitude of 112 (see Tab.1 in Pesnell, 2008). In contrast, the predictions obtained with the help of the past climatology, spectral methods or solar precursors, give amplitudes or activities somewhat lower than the mean value. Some of the cited predictions have already failed because the predicted expected moments of the maximums are over (to date 15 of the predictions). The predictions of the sunspot maximum time are very questionable due to the long unexpected solar minimum. Another summary of 45 solar activity forecasts was given by Janssens (2006) with the last update from Februar 2009. He outlined a strong divergence between the statistical and more physically oriented approaches. By statistical methods the results are in average mostly for strong cycle, whereas by the physically based ones (where the prognoses of Dikpati and Hathaway where excluded), weaker cycles are obtained. The difference between the conclusions drawn in the Pesnell's summary and this of Janssens's is coming from the fact, that in the last one the statistical method neural network and spectral methods and prediction on auto-regression where included in only one category. An overview of the different solar activity prediction methods can be found e.g. in Kane (1997), Hattaway, Wilson, Reichmann (1999), Cameron and Schüssler (2006), Schüssler (2007), Kane (2008), Hathaway (2010). A method to predict not only the occurrence and amplitude of the solar cycle maximum, but also its period length was worked out by Hiremath (2006).

The NOAA and NASA panel consensus prediction of the 24 solar cycle (http://www.swpc.noaa.gov/SolarCycle/SC24/index.html) was very much changed in time. The predictions enclose the amplitude of the sunspot maximum and the date of the maximum.

In March 2007 the NOAA and NASA panel was split and predicted the solar cycle to reach a peak sunspot number of 140 in October, 2011 or a peak of 90 in August, 2012 (http://www.swpc.noaa.gov/SolarCycle/SC24/Statement_01.html). In May the prediction was revised - the Solar Cycle 24 will peak in May 2013 with 90 sunspots per day, averaged over a month (http://www.swpc.noaa.gov/SolarCycle/SC24/index.html).

The march 2011 update (http://solarscience.msfc.nasa.gov/predict.shtml) gives a smoothed sunspot number maximum of about 58 in July of 2013 and states that we are currently two years into Cycle 24 and the predicted size continues to fall.

Regardless of the advance in the application of physical methods for the purpose of forecasting, the results are very inconsistently spread and substantiate the application of statistical methods.

Some historical aspects

Yule in 1927 was the first using the method known today as auto-regression technique to describe the yearly sunspot number series, introduced in 1848 by the Swiss astronomer Johann Rudolph Wolf (Yule, 1927). Yule used the more realistic sunspot data from 1749 up to 1924. The data are reliable since 1848, questionable from 1749-1817 and are characterized as poor during 1700-1748 (Eddy, 1976). Yule solves the homogenous second difference equation, $x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2}$ where x_t are the mean removed sunspot numbers for the year t. The equation describes a damped harmonic oscillator, if the roots of the characteristic equation of (1) are conjugated complex. Yule determined the constants φ_1 and φ_2 to be means of ordinary least square estimations and he found $\varphi_1 = 1.34254$, $\varphi_2 = -0.65504$; the root of the mean squared deviation of x_t from the original sunspot values was determined and the value of 15.41 was found. The sunspot numbers (in the original Yules work called Wolfer's sunspot number, to honor of the student and later successor of Wolf at the Zürichs observatory) are not absolutly the same as the sunspot number used today from Zürich, collected since 1979 at the Sunspot Influences Data Analysis Centre (SIDC) at the Royal Observatory of Belgium.

Walker generalized Yules approach proposing autoregressive (AR) model and involved it in atmospheric data (http://en.wikipedia.org/wiki/Walker_circulation) leading him to discover the Walker circulation. After Yule many scientists have made attempts to improve the description of the sunspot time series as pure AR(p)-model or ARMA(p,q)-models of different order p and q. Some results were summarized by Mc Leod and Hipel (1977). Moran (1954) pointed out that to better describe the asymmetry development of the sunspot cycles, a model of higher order than ARMA(2,0) is need. He established a strong increase of the prediction error with rising of the prediction horizon. The coefficients of the developed autoregressive models to describe the solar activity depend on the used data period. In the following we use the yearly SIDC sunspot number data (http://sidc.oma.be/sunspot-data/) to construct an ARMA model based on the Box-Jenkins method including the 23th solar cycle data, and beginning at 1749.

AR(p)-model

The time series value, observed at equidistant moments *t* in an autoregressive model is the weighted sum of the previous ones.

$$x_t = \mu + \sum_{i=1}^{p} \varphi_i x_{t-i} + u_t , \qquad t=1,\ldots\ldots,n \qquad (1)$$

where the moments *t* was numbered from *1* to *n*, μ is a constant and the residual u_t presents a white noise process and *p* determines the order of the autoregressive process. The white noise term holds the assumptions of the cross-sectional regression, the mean of the error are zero, the variance is constant and correlations between the residual and itself and also between variables x_{t-i} and the residual u_t do not exist. In contrast x_t is defined by a weak stationary process, in other words the mean expected value of *x* is μ, the variance σ^2 of the series x_t is constant and does not depend on time (that is to say the series is homoskedastic) and the autocorrelation depends only of the lag (e.g. Box and Jenkis, p.26,). These assumptions implicate that time series shows not have a trend. In some cases it is also necessary to de-seasonalize the time series before it's description with an autoregressive model. The structure of the regressive equation, where the value at the time *t* are determined by the past observations, is predestined for forecasts, because the future value at the moment *t*+1, is defined by the preceding *p* time-series values.

Using the original data x_i and the back transformed values $\hat{x}_i$ the residuals ε_i and the standard deviation σ can be calculated easily. Not measured values at the time point *n+1* can be estimated from the last p values of the time series:

$$\hat{x}_{n+1} = \hat{x}(h) = \varphi_1 x_n + \varphi_2 x_{n-1} + \ldots + \varphi_p x_{n-p} \,. \qquad (2)$$

The prediction values are determined, besides from the model (in the case of an AR -model), also by the last *p* data, called for this reason pivotal values. The forecasted value $\hat{x}_{n+1}$ is often denote by *x(h)* and *h* is referred as prediction horizon. By means of the estimated future value $\hat{x}(h)$, the next future value $\hat{x}_{n+2}$ can be determined:

$$\hat{x}(2h) = \varphi_1 \hat{x}(h) + \varphi_2 x_n + \ldots + \varphi_p x_{n-(p-1)} \,. \qquad (3)$$

In this way it is possible to estimate successively m forecast values $\hat{x}(mh)$. The forecast series $\hat{x}(mh)$ is named prediction profile. It can be shown that the prediction values asymptotically converge to *μ* with increasing prediction horizon (Thome, 2005).

To obtain a measure of the quality of the forecast, forecast values can be calculated with the first *p* values

$$\hat{x}(h)_{p+1} = \varphi_1 x_p + \varphi_2 x_{p-1} + \ldots + \varphi_p x_1 \,. \qquad (4)$$

In an analogous manner $\hat{x}_{p+2}$ can be estimated using the values from x_{p+1} till x_2 and so on. Of course, values which are not observed, can be calculated by forecasting of the reverse time series or by setting to zero. By means of the forecast estimation $\hat{x}(h)_{p+1}$ and the *p-1* last measurement values the future data $\hat{x}(2h)_{p+2}$ can be obtained. Then the residual for a determined prediction horizon *kh* is determined from the difference of the observed value x_i and its estimation $\hat{x}_i$:

$$\varepsilon(kh)_i = x_i - \hat{x}(kh)_i \,. \qquad (5)$$

The forecast quality can be estimated then by the sum of the square residuals ε_i^2 and by the mean standard deviation:

$$\sigma(kh) = \sqrt{\frac{1}{n-p-p'} \sum_{i=p+1+h}^{n} (x_i - \hat{x}(kh)_i)^2} \,, \qquad (6)$$

where *n* is the number of the observations, *p* represents the order of the AR model and ***p'*** is the number of used coefficients to estimate and *kh* is the prediction horizon.

Model identification

It is well known that the sunspot number time series is not a strong periodic process. The period length varies from cycle to cycle. The 23th cycle was one of the longest cycles with a very wide minimum. By inspection of the plot of the yearly sunspot number as a function of the time is observed that the time series is not stationary. The series can easily be split in parts, where the solar activity is lower or higher during some subsequent solar cycles (see Figure 1). As the beginning of an interval was use a sunspot minimum and the part is ending just before a minimum. For every part the arithmetic mean and the standard deviation of the sunspot numbers were calculated. As it is clearly shown at Figure 2, the standard deviation found increases linearly with the means of higher sunspot numbers indicating a multiplicative time series model (Schlittgen and Streitberg, 2001, p 11).

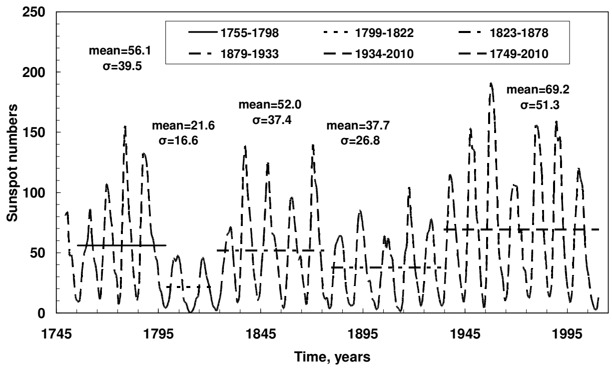

Fig.1. The development of the solar activity is displayed indicated by the sunspot number from 1749 up to 2010. The series was separated in some parts and the mean value and the standard deviation for these intervals are specified.

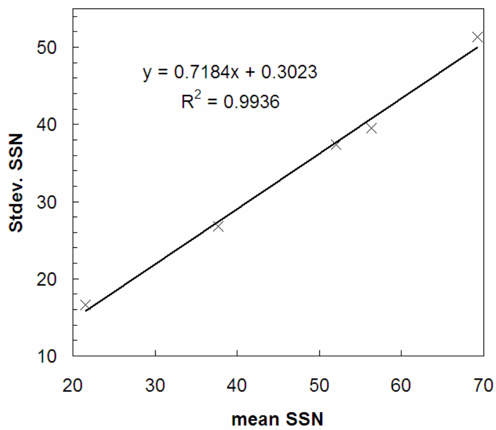

Fig. 2. Dependences of the standard deviations of the mean sunspot numbers for the different time intervals (see Figure 1.)

A logarithmic transformation transfers the model in an additive model again. During the Dalton Minimum in 1810 a SSN value of zero was observed. To overcome this problem a constant can be added to x_t and after this the log-transformation can be carried out. However the inverse transformation is more complex because the dispersion function of the original and the transformed series are not the same and prediction values and the prognoses errors have to be corrected also (Granger and Newbold, 1976; Pankratz 1991). McLeod (1977) compared different models including a multiplicative AR model and showed that a AR(9) model is to be preferred, but no forecasting was performed. Tests to made forecast using the log-transformation have shown that the prediction errors after the inverse transformation are non-acceptably high. Therefore a simple square root transformation was used. It is obvious in figure 3 that the square root transformed time series is not fully homoskedastic but it is also well known that the Box-Cox square root transformation remove the heteroskedasticity in satisfactory manner and save the normality of the residuals.

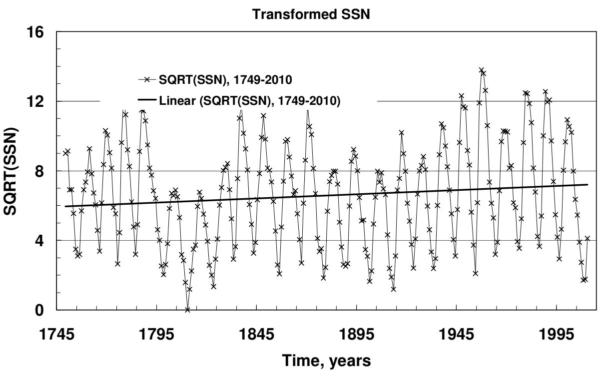

Fig. 3. The square root transformed Sunspot number time series and its linear trend.

The square root transformed SSN time series was detrended and the autocorrelation function was determined (see Figure 4). This way the estimated autocorrelation function (EACF) was found. The EACF displays a behavior like an aperiodic oscillation. The values after lag 9 of the estimated partial autocorrelation function (EPACF) are not significant. The EACF and PEACF (both calculated with the Statistica 6 program) indicate characteristics typical for an AR process of 9th order.

Determing the model parameters and evaluation of the model

The *p* constants of the AR-equation (1) φ can be determined in different ways. First, is it possible to do this by the recursive solution of the Yule-Walker (e.g. Thome, p. 90) equation, related to the first p autocorrelation coefficients. The second way for the determination of the constants φ in principle consists in the interpretation of equation (1) as a linear regression problem, which can be solved by a lot of standard methods, minimizing the mean squared errors. Using the original data set publicised in the Yule's work without performaing the Box-Cox transformation on the base of the Yule Walker equation for the AR(2) model, the constants were found as $\varphi_1 = 1.3398$ $\varphi_2 = -0.6505$. These values are in very good agreement with the values noticed by Yule: $\varphi_1 = 1.3425$ $\varphi_2 = -0.6550$. Applying the detrended and square root transformated sunspot number SIDC data set from 1749 up to 1924, we found $\varphi_1 = 1.3571$ and $\varphi_2 = -0.6601$, very close values to these from Yule as well as.

To determine the AR model parameter, Statistica 6 programme was used. The estimation method in this software maximized the likelihood of the data. The McLeod and Sales approximation was used. The resulting parameters of the AR(9) model of the detrended square root transformed SSN data from 1749 up to 2010 are summarized in the table 1.

(Parameters of other models were calculated including moving average as well, however no significant differences were found). The constant μ was not estimated, because the series was detrended before the AR model parameters were determined. Four of the autoregression parameters are significant, highlighted in the table 1. The constant φ_4 is very close/near at the significance level of 0.05.

The residuals for kh=1 were determined using equation 5 and the fit to a normal distribution is shown in figure 4 together with the results for the tests of normality. By the three tests, the Kolmogorov-Smirnov, the Liliefors and the Chi-squared test, the hypothesis of normality cannot be rejected and as mentioned above the Cox-Box transformation really saves the normality of the residuals. Moreover the residuals are not auto-correlative, as displayed by the autocorrelation function of the residuals shown at figure 5.

Tab. 1. Parameter of the AR(9) model

order	φi	Asymp. std. err.	Asymp t(253)	Sign. level
1	**1.22**	0.06	20.1	0.00
2	**-0.50**	0.10	-5.26	0.00
3	-0.10	0.10	-1.01	0.32
4	0.20	0.10	1.95	0.05
5	**-0.22**	0.10	-2.13	0.03
6	0.06	0.10	0.63	0.53
7	0.08	0.10	0.83	0.40
8	0.16	0.10	-1.67	0.10
9	**0.27**	0.06	4.47	0.00

For *kh*=1, 2, 3, 4 the standard deviations values are found in line to the equation (6) 1.10, 1.74, 2.04 and 2.11 respectively. The standard deviations increase asymptotically very fast with increasing prediction horizon. To illustrate the standard deviation in the measure of the original sunspot series, every prediction value was transformed back and the standard deviation was estimated. For the upper prediction horizons the following sigma values were obtained: 14.4, 23.3, 29.1, and 32.3. It is clear that with a value greater than 30, the meaning of a prediction is questionable. The mean of the sunspot numbers is 51.9 and the two sigma confidence interval is 60, also greater than the mean sunspot number. Of course the confidence limit is not symmetric about the prediction value as a result of the nonlinear transformation as can be seen in figure 6, where the prediction for the sunspots for the 24 sunspot cycle is shown together with the upper and lower confidence limit for the significance level of p= 0.9. For comparison, the prediction for the last four cycles is drawn. This forecast was calculated on the basis of the data up to the previous solar sunspot minimum. While the sunspot number maximum is obtained only three to four years after the solar maximum, the prognoses are very unascertained. For example the forecasts for the 21 and 22 cycle demonstrate the uncertainty.

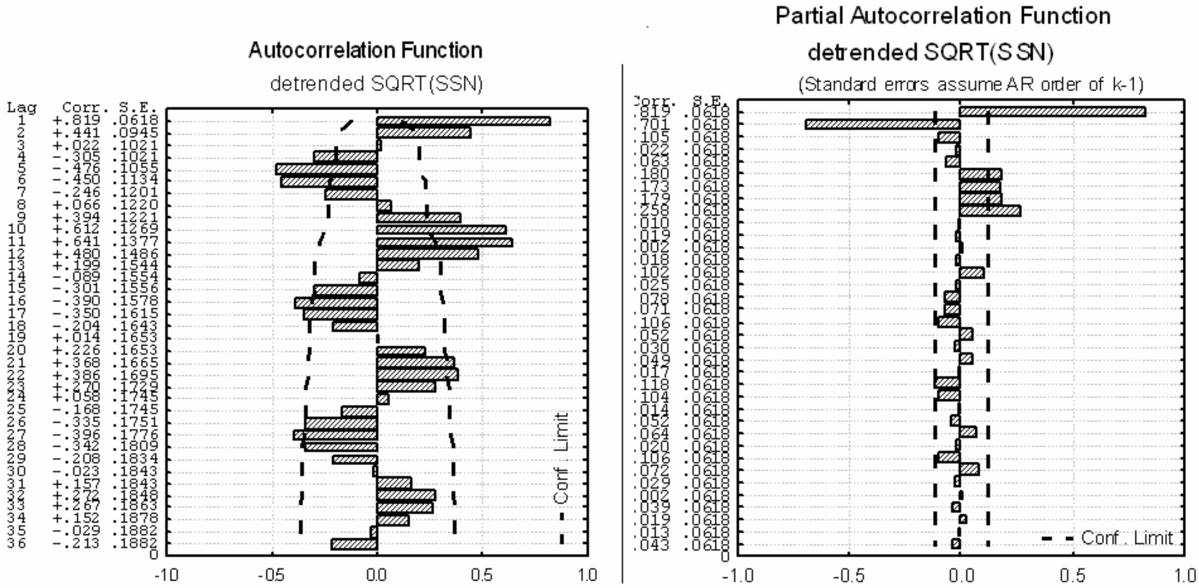

Fig.4. Estimations of the autocorrelation and the partial autocorrelation function of the Sunspot numbers from 1749 till 2010. (For discussion see text.)

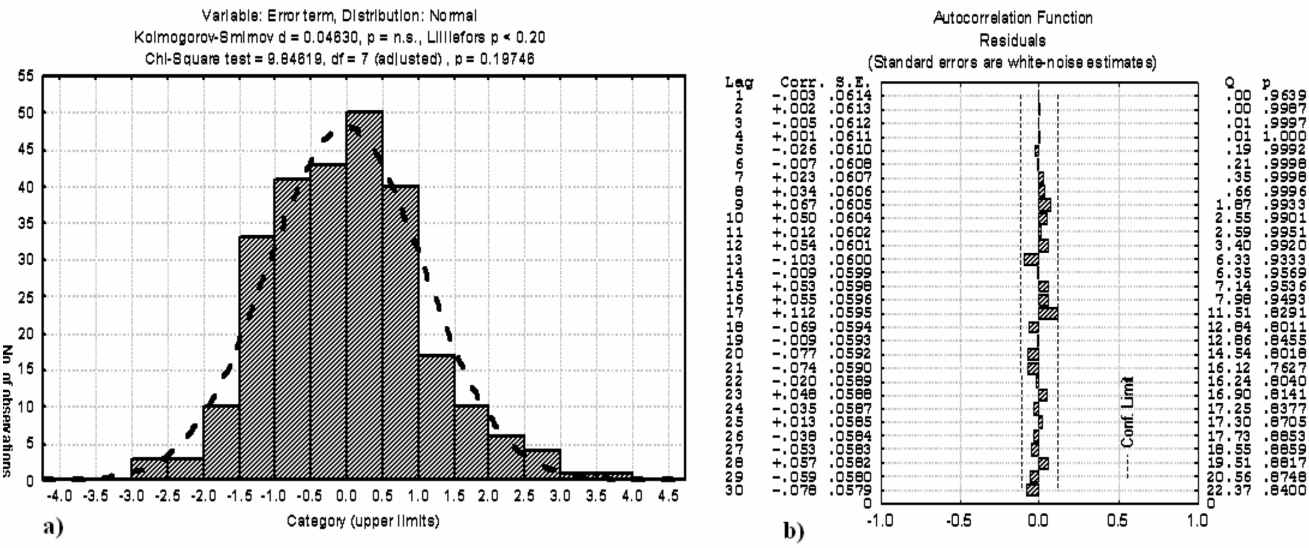

Fig.5. The distribution of the residuals and its approximation as normal distribution, where the test results are displayed in the graphic header a) and the estimated autocorrelation function of the residuals b).

The maximum of the 24th solar cycle was forecasted to be reached in 2012 with the maximum of approximately 90 sunspots yearly mean with the confidence interval of +77/-53.

In praxis the moment when the solar activity reaches its minimum is not know. It can be determined only after the appearance of new sunspots with reversed polarity of the magnetic field of/between the following and preceding sunspots with respect to the old ones. Therefore forecasts since 2006 (with the data from1749 including 2005) were calculated for every year. The forecasts for the following 11 years are shown in figure 6. The obtained sunspot maxima are all in the interval from 110 down to 87. And in general with the progress of the sunspot minimum the prediction moment is shifted to the following year. Only for the last prediction (the forecast from 2010), which was calculated after reaching the sunspot minimum in 2009, the prediction values from the previous forecast were used. The reason is that the first prediction sunspot number (of 16.2) for 2010 of the forecast from 2009 is very close to the really observed value (of 16.9). The last predictions show that the sunspot maximum the figure 7 the observed monthly sunspot number up to June 2011 is also drawn for comparison. The prediction values based on the prognoses started in

2009 and in 2010 are in very good agreement with the observed monthly SSN to date.

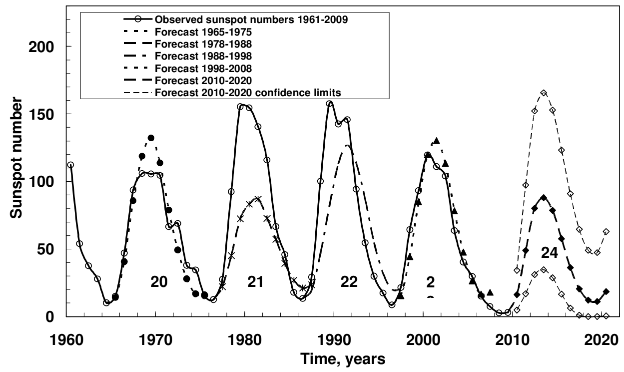

Fig.6. The observed and prognosed sunspot numbers. The prognoses of the next sunspot cycle were started at the sunspot minimum before.

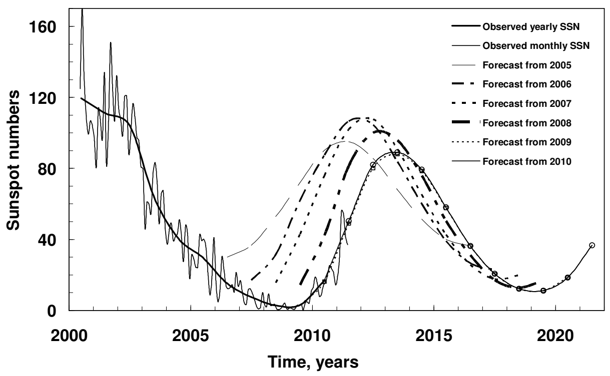

Fig. 7. Forecasting of the sunspot numbers for the 24th solar cycle. The forecast starts in different years. The squares mark the values for prediction with start after the observations of the SSN 2009, and the circles- the prediction calculated on the basis of the SSN data up to 2010. For comparison the observed monthly sunspot numbers up to June 2011are drawn.

After the back transformation of the forecast values of the transformed observed data, applying the equation 4, the residua and the mean standard deviation (equations 5 and 6) was calculated. 14.4 is the obtained standard deviation with the full set of parameters. The significant coefficients can be dropped from the model, consequently the standard deviation is increasing slightly to 15.8. It is significantly smaller than the standard deviation (σ = 41.5) from the mean (μ= 51.9) of the original time series.

Conclusions and summary

While the sunspot number data from 1749 up to the modern time are to be described by an AR(9) model, the forecast of the solar activity is limited to two - three years. For longer prediction horizons the prediction errors rise to above 30, less meaningful for solar activity forecasts. This statement is in good agreement with the findings of Rozelot that accurate forecasting over a period of time longer than 2 to 4 years seems impossible due to a Lyapounov exponent of 0.5 if the 11-year activity cycle is excepted (Rozelot, 1995; see also Li, 2007). Rozelot pointed out that every solar cycle seems to be dominated by its own physical logic. Despite this the forecasting from 2009 and 2010 is in good agreement with the observed monthly SSN to date. The activity maximum of the 24th solar cycle is expected for the year 2013 with a SSN of approximately 90.

References:

Baranovski, A.L., Clette, F., Nollau V.: 2008, Ann. Geophys. 26 (2), 231.
Bonev, B., Penev, K., Sello, S.: 2004, Astrophys. J. 605 (1), 81.
Box, G.E.P., Jenkins, G. M.: 1976, Time series analysis, forecasting and control, Holden-Day Inc., p. 26.
Cameron, R., Schüssler, M.: 2006: arXiv:astro-ph/0612693v1
Eddy, J. A.: 1976, Science 192, 1189.
Granger, C.W.J., J. Newbold: 1976, J. Roy. Statist. Soc. Ser. B. 28 (2), 189.
Hathaway, D.H.: 2010, Living Rev. Solar Phys. 7 (1), 1.
Hathaway, D.H., Wilson, R. M., Reichmann E. J.: 1999, J. Geophys. Res. 104 (A10), 22,375.
Hiremath, K.M.: 2008, Astrophys. Space Sci. 314 (1-3), 45.
Janssens, J.: 2006, Solar Cycle 24: Overview of predictions on the start and amplitude of a new solar cycle, accessed as http://members.chello.be/j.janssens/SC24.html.
Kane, R.P.: 1997, Geophys. Res. Lett. 24 (15), 1899.
Kane, R.P.: 2008, Ann. Geophys. 26 (11), 3329.
McLeon, A.I., Hipel, K.W., Lennox, W.C.: 1977, Water Resources Researches 13 (3), 577.
Moran, P.A.P.: 1954, Statist. Soc. Ser. B 16 (1), 112.
Li Q.-X., Li, K.-J.: 2007, Chin. J. Astron. Astrophys. 7 (3), 435.
Ogurtsov, M. G., Nagovitsyn, Yu. A., Kocharov, G. E., Jungner, H.: 2002, Solar Phys. 211 (1), 371.
Pankratz, A.: 1991, Forecasting with dynamic regression models, New York, John Wiley, p. 338.
Pesnell, W.D.: 2008, Solar Phys. 252 (1), 209.
Rozelot, J.P.: 1995, Astron. Astrophys. 297, L45.
Schatten, K. H., Tobiska, W. K.: 2003, B. Am. Astron. Soc. 35, 817.
Schlittgen, R., Streitberg, B.H.J., Zeitreihenanalyse: 2001, Aufl., R.Oldenburg Verlag München Wien, p. 11.
Schüssler, M.: 2007, Astron. Nachr 328, (10), 1087.
Thome, H., Zeitreihenanalyse, R.: 2005, Oldenburg Verlag München Wien, p. 287.
Yule, G. U.: 1927, Phil. Trans. Royal Soc. Ser. A 226, 267.