Forecast of the Decadal Average Sunspot Number

D.M. Volobuev, N.G. Makarenko

Pulkovo Observatory, St.-Petersburg, Russia

dmitry.volobuev@mail.ru

Abstract. The forecast of the decadal average sunspot number (SN) becomes possible with an extension of telescopic observations based on proxy reconstructions using the tree ring radiocarbon data during the Holocene. These decadal numbers (SN_{RC}) provide a powerful statistic to verify the forecasting methods. Complicated dynamics of long-term solar activity and noise of proxy-based reconstruction make the one-step-ahead forecast to be challenging for any forecasting method. Here we construct a continuous dataset of SN_{RC} which extends the group sunspot number and the international sunspot number. The known technique of non-linear forecast, the local linear approximation, is adapted to estimate the coming SN. Both the method and the continuous dataset were tested and tuned to obtain the minimum of a normalized average prediction error (E) during the last millennium using several past millennia as a training dataset. $E = 0.58\sigma_D$ is achieved to forecast the SN successive differences whose standard deviation is σ_D =7.39 for the period of training. This corresponds to the correlation (r=0.97) between true and forecasted SN. This error is significantly smaller than the prediction error when the surrogate data were used for the training dataset, and proves the non-linearity in the decadal SN. The estimated coming SN is smaller than the previous one.

Keywords: Solar activity, non-linear forecast, radiocarbon

1. Introduction

Solar activity forecasting is important for various scientific areas, mostly related to operations of low-Earth orbiting satellites, long-term forcing in climate models and geophysical applications. Many attempts were made to predict the future behavior of the coming solar cycle (see *e.g.* Kane (2007) for a brief review). Numerous prediction techniques have been developed to accurately predict the phase and amplitude of future solar cycles, but with limited success. The solar cycle is very difficult to predict due to noise contamination, high dispersion level and high variability both in phase and amplitude at different time-scales. The number of the observed 11-year solar cycles is too small to provide a statistical confidence for any predictive rule or theoretical model for the cycle-to-cycle variations. Another question is to forecast the envelope of 11-year cycles on the basis of the proxy data. A variety of solar activity reconstructions which are based

on different proxies such as historical observations of sunspots and aurora, the ^{10}Be content in polar ice, and the ^{14}C content in tree rings (see *e.g.* Usoskin and Kovaltsov (2004) for a brief review) postulate independently the persistence of long-term variations, mostly of the Gleissberg and Suess cycles. These proxies are contaminated with noise but still provide some perspective to forecast the long-term solar activity variations. Among these proxies, the ^{14}C content in tree rings is the only global continuous dataset which is statistically useful, not only to build the forecast model but also to verify its result. Several attempts were made to extract the isotope production rates and respective long-term variations in solar activity from the ^{14}C content in tree rings. Relatively short (period<400 years) variations agree well among different reconstructions when these reconstructions are scaled to the sunspot number via an appropriate semi-empirical (Stuiver and Quay,1980; Volobuev *et al.*, 2004; Marchal, 2005; Ogurtsov, 2005) or theoretical (Solanki *et al.*, 2004) models. Hereafter we call this sunspot number originating from the tree ring radiocarbon "SN_{RC}". The SN_{RC} proxies provide a powerful statistic of decadal changes of the envelope of solar activity during the Holocene (more than 10000 years before present), but their long-term trend is uncertain mostly due to high uncertainties in the estimation of the dipole magnetic moment of the Earth (Muscheler *et al.*, 2005). An attempt to forecast several forthcoming decadal averages of the group sunspot number on the basis of different SN_{RC} proxies was made by Ogurtsov (2005), but the accuracy of the method proposed for the one-step-ahead forecast was below the level of inertial forecast ($SN_{RC}(t+1)= SN_{RC}(t)$). Another attempt by Clilverd *et al.* (2006) is based on the extrapolation of the Fourier components of SN_{RC}. This approach cannot provide an excellent prediction of the non-linear time series like the sunspot number (Ostryakov and Usoskin, 1990; Sello, 2001) because the Fourier components represent a stochastic rather than a non-linear process. However, the non-linearity of the decadal sunspot number is not obvious and is one of the subjects of this investigation. The scope of the present investigation is the one-step-ahead forecast of the decadal numbers. The result of such forecast may be useful to forecast the size of the coming 11-year cycles with an appropriate model. Within this scope we address the following questions:

- Is it possible to make a continuous composite "sunspot number" during the Holocene up to the present time?

- Can the noisy SN_{RC} series predict the sunspot number based on telescopic observations?
- Is it possible to produce the decadal sunspot number from a model of stochastic or non-linear process?

These questions we try to answer with a computer experiment. We make statistical predictions using a non-linear forecast method and consider the normalized average prediction error (E) (defined in Section 3) for the step-by-step forecast as a criterion to incorporate the data into a continuous dataset, to adjust the method and to verify the hypotheses about the nature of the long-term variations. The paper is organized as follows: In Section 2 we build a continuous sunspot number from 9455 BC to the present. In Section 3 we describe the forecast method and in Section 4 we discuss the errors in data, errors of the forecast and provide the surrogate data test of non-linearity. Section 5 discusses the constraints to physical hypotheses about the nature of the long-term solar cycles.

2. Continuous Composite Dataset

The forecast requires a continuous dataset which we need to construct from three different time series (Table 1).

Table 1 Components of the composite continuous dataset

Dataset	Period [years]	Transform	Web source
SN_{RC} (Solanki *et al.*, 2004)	9455BC—1895AD	SN_{RC}+7.6	http://gcmd.nasa.gov/records/GCMD_NOAA_NCDC_PALEO_2005-015.html
GSN (Hoyt and Schatten, 1996)	1885—1995 AD	$<GSN>_{10}$	http://www.ngdc.noaa.gov/stp/SOLAR/ftpsunspotnumber.html
ISN	1996—2006 AD	$< ISN>_{10}$	http://solarscience.msfc.nasa.gov/greenwch/spot_num.txt

The historical period 9455BC—1895AD is covered with SN_{RC} dataset. The stop point of SN_{RC} is 1895 because the Suess effect (a sudden growth of atmospheric CO_2 content which has continued up to the present) started there which is difficult to take into account. The period of intersection (1610—1895) was studied in detail by Solanki *et al.* (2004). During this period our composite continuous dataset is based on SN_{RC} because this period is the most recent and most accurate period in SN_{RC}, whereas GSN may not be so accurate because regular Greenwich

observations started in 1848 only. SN_{RC} is shifted up in amplitude by 7.6 to make its ending points at 1885 and 1895 equal to the corresponding points in GSN (Figure 1). This shifting up helps us to remove the negative values which were in SN_{RC} during the deep solar activity minima (*e.g.* Spörer and Maunder minima, Figure 3) because these values are false by the definition of the sunspot number. It also provides a continuity condition for the composite dataset and its derivative which is strongly required by our forecast method. This shifting up is consistent with the finding by Muscheler *et al.* (2007) that the 20-th century maximum of solar activity is not so exceptional as it was stated by Solanki *et al.* (2004) and the average level of solar activity during the Holocene might have been somewhat higher. The method itself also provides a statistical result which verifies the structure of the composite dataset. Namely the prediction error of GSN since 1895 increases considerably if the continuity condition is not satisfied or the average value is used instead of the continuation during the period of intersection. The recent part of the composite dataset is the GSN which is smoothed by an appropriate smoothing spline to maximize the correlation with SN_{RC} during the period (1610 - 1895) of SN_{RC} and GSN intersection. The GSN dataset was connected to the international sunspot number (ISN) during the period (1995-2007). GSN is very close to ISN for the cycles of moderate amplitude as cycle 23, but not for other cycles (Nagovitsyn, 2005). Hereafter we introduce the notation of *composite* data set for the decadal averaged continuous data set {SN_{RC}, GSN, ISN}. The smoothing spline was used to adjust the GSN smoothness to SN_{RC}.It is a commonly used algorithm of data approximation (*e.g.* Fortran subroutine CUBSPL or MatLab Spline Toolbox function CSAPS). It minimizes the weighted sum $p\sum(y_i - S_i)^2 + (1-p)\int (D^2 S(t))^2 dt$ of errors between the data (y) and the spline fitting (S) and the norm of the second derivative of the spline ($D^2 S(t)$) Here p is a smoothing parameter in the form of $p= 1/(1+h^3/a)$, and the sampling rate h=10 years. We found that parameter a =0.6 maximizes the correlation during the period of intersection. Although the result of the forecast is not very sensitive to the degree of smoothing, some consideration is due here. Over-smoothed data can be predicted more easily but are not so interesting. Namely a=0.01 leads to the loss of the gap in decadal GSN at 1975 but decreases the average prediction error E from $0.58\sigma_D$ to $0.53\sigma_D$; here σ_D is the standard deviation of the successive differences in decadal numbers for the training period (see Sections 3, 4). In the

case of less smoothing with a=6, the error increases to E= 0.81σ_D. Nevertheless, this case closely reproduces the decadal average sunspot number (Figure 1). The SN_{RC} data look more smooth, partly because some smoothing has been introduced by INTCAL98 which was constructed as an average of regional datasets with their errors (Reimer, 2004), and partly due to physical ocean smoothing (reservoir effect) which may not have been completely removed by the model (Solanki *et al.*, 2004). Even with this selection of a=6, the coming decadal number is also predicted smaller as in the case of a=0.6. The last three spline-averaged points deviate significantly from the decadal average (Figure 1), mostly due to non-Gaussian distribution of the annual numbers. The deviations are much smaller if we compare the values of yearly differences, so that the differentiation operation provides the stability of the forecast results (Section 3).

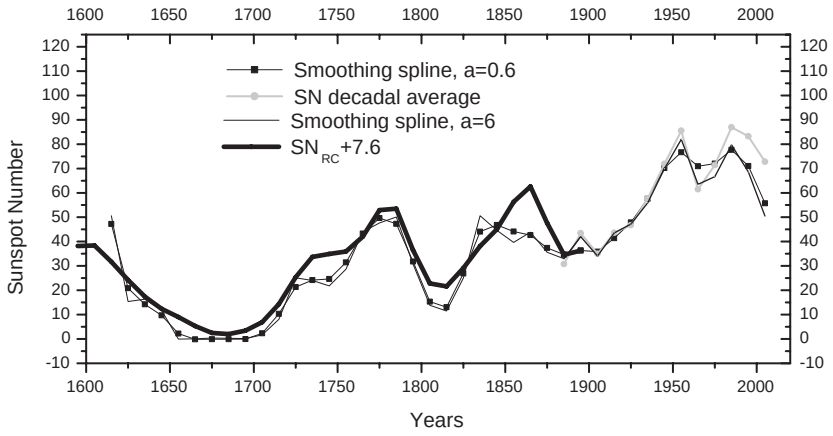

Figure 1. The composite data set. The curve 'Smoothing spline, a=0.6' best fits the SN_{RC} data. The curve 'Smoothing spline, a=6' best reproduces the SN decadal average which is the annual group sunspot number averaged during the decade. The last point is estimated from the incomplete decade.

We have three time series to describe the dynamics of solar activity during 9455BC—2005AD in the over-decadal time scales. These data have been produced from various sources, so that they have different levels of the noise contamination and physical structure of their errors. Strictly speaking, each time series should be considered as a separate index of solar activity. We will, however, construct a *composite* (synthetic) series by introducing some formal criteria for the continuation of data. How can one be sure if any physics is preserved under this heterogeneous time series? We suppose that the answer could

be provided with the investigation of the predictability of this time series. Namely if we can obtain an adequate prediction error, we may conclude that this synthetic series preserves the level of determinism, *i.e.* it may be regarded as an *observable* of some hypothetical continuous dynamical process.

3. Method and Results

The method of the local linear (LL) approximation in the phase-space first proposed by Farmer and Sidorowich (1987) was used to forecast our composite time series. The hypothesis in the background is that the low-dimensionality chaos (non-linearity) is present in the long-term solar activity variations or that they can be described with some model system of several non-liner ordinary differential equations (ODEs). This model system can be derived as a truncation of the mean field dynamo equations (Weiss, 1985) or a disc dynamo (Volobuev, 2006) or dipole-quadrupole interaction dynamo (Knobloch and Landsberg, 1996). The hypothesis of the non-linearity model can be verified with the success of the LL-method to predict a time series (Farmer and Sidorowich, 1987) because the construction of the method assumes its ability to predict also a stochastic time series. The LL-method has originated from the famous method of "analogs" by Lorenz (1969). Edward Lorenz is mostly known for his research on the chaotic system of non-linear ODEs which describes the appearance of chaotic regimes in the atmosphere mixing. "Lorenz analogs" are recurrent durations in the evolution of a time series. A sudden appearance of an analog in the chaotic time series can be explained by the preserved shape of a chaotic attractor in the phase space. In the phase space the analog is a nearest point (a neighbor) from the closest phase trajectory, and the dimension of the phase space corresponds to the length of the analog. The closer is the analog the better is the forecast.

An interesting paradox is that the Lyapunov time is often about the length of one quasi-cycle (a pseudo-period) of the chaotic system, whereas the nearest analog may appear after 10^4 quasi-cycles or more. Therefore, in order to find a good analog for short-term weather forecast, the span of available data (the so-called ' libraries') which is10-100 years is considered "rather short" (Nicolis, 1998). So, despite quick memory loss and the fundamental constraint for the predictability horizon, noise-free chaotic systems have no restriction for the recurrence time of the analog, and therefore a longer library is preferable for a better prediction of a

time series. Farmer and Sidorowich (1987) improved this approach with local (within analogs) linear regression and generalized it with the Takens-Packard procedure of the phase-space reconstruction (Packard *et al.*, 1980).

We will make several improvements of the LL-method when applying it to predict the composite time series. First of all, successive differences $D_k=C_k-C_{k-1}$ instead of amplitudes C_k themselves of the composite dataset will be forecasted. The procedure of the differentiation significantly reduces the amplitude of the long-term variation (Figures 3, 4) which is questionable in the SN_{RC} proxy due to high uncertainty in the geomagnetic dipole moment (Muscheler *et al.*, 2005) and other significant systematic errors (Section 4). This procedure also helps us to reduce the inertial component (*i.e.* a trivial forecast when $C_k = C_{k-1}$) which is problematic because of the high one-step autocorrelation $r(C_{k-1},C_k)$=0.9. As a result the autocorrelation $r(D_{k-1}, D_k)$=0.5 is much smaller. The second improvement is that we will adopt one-step delay (τ=1) that maximizes the library of the historical analogs, the minimum dimension of the reconstructed phase space (m=3), and minimum number of analogs to be found ($N_a=m+1$). Apart from that we will avoid the multiple usage of the same data points, which is often the feature of the delay embeddings. The reconstructed phase space is:

$$D = \begin{matrix} D_1,D_4,D_7\ldots,D_{[N/m]-m+1} \\ D_2,D_5,D_8\ldots,D_{[N/m]-m+2} \\ D_3,D_6,D_9\ldots,D_{[N/m]} \end{matrix} \qquad (1)$$

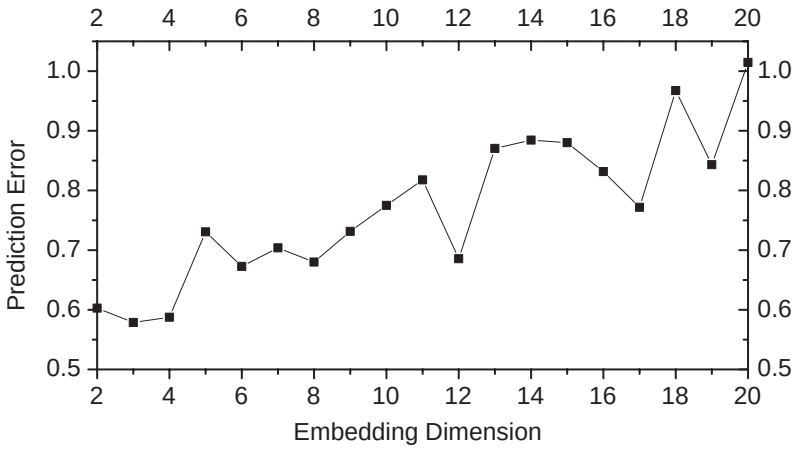

Figure 2. Variation of the prediction error with the embedding dimension.

There indices D_i are numbered from the present to the past, square brackets mean the round part, columns are vectors in the phase space (the library of analogs), N is the total length of the training set. Dimension m = 3 (Figure 2) was chosen here to minimize the prediction error. We tested dimensions m=2—20 and found that m=3 provides the best accuracy of the forecast. This is inconsistent with a theoretical consideration, namely, if the length of the time series is infinite and data error is close to zero, we should expect a reduction in the prediction error for higher embedding dimensions and stabilization of the error at the constant level when the true dimension is achieved. On the contrary, Figure 2 shows only one point (dimension 3) of this tendency but significant increase in the error at higher dimensions which is caused by the limited length of the time series and noise contamination which leads to the smaller number of perfect analogs at higher dimensions.

Nearest neighbors (Lorenz analogs) A are the columns in Equation (1) which are closest to the first one in the sense of the squared Euclidean distance. They form a matrix similar to the matrix (1) with an additional upper row X for the prediction. A constrained linear least square problem should be solved to find the vector of regression coefficients g:

$$A^{\mathrm{t}}g = X^{\mathrm{t}} \quad (2)$$

$$-\sigma < g_i < \sigma$$

Here σ is the standard deviation of D, t is the transpose operator. Now we can find the quantity $D_0 = \sum_{i=1}^{10} g_i D_i$ and $C_0 = C_1 + D_0$ which is the forecast of the derivative and the composite data themselves, respectively[1]. The one-step-ahead forecast of the composite data is evaluated during the last millennium (Figure 3). The correlation between actual C_a and forecasted C_0 is extremely high for a step-by-step forecast (see the inset in Figure 3), but this is partially due to slow changes of C_a (high contribution of the inertial forecast). More informative is the correlation between the actual D_a and forecasted D_0 of successive differences (r=0.74) which is significant for 100 independent points.

[1] MatLab scripts for the method can be found at http://www.mathworks.com/matlabcentral/fileexchange/ , File Id 17276.

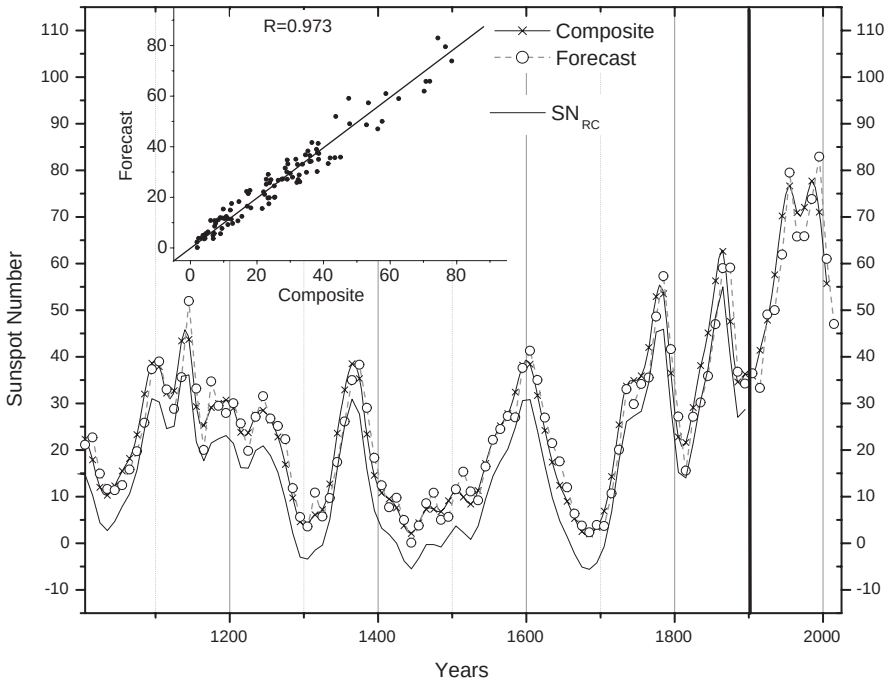

Figure 3. Continued envelope of sunspot number and its 10-year-ahead forecast for the last millennium. Here SN_{RC} means the sunspot number derived from tree ring radiocarbon content by Solanki *et al*. (2004). The composite data after the vertical thick line shows the group sunspot numbers (GSN), smoothed and 10-year sampled to provide the best correlation with SN_{RC}. The composite data before the vertical thick line shows SN_{RC} shifted up to provide the continuity of the composite at the point of the vertical line. 'Forecast' means the 10-year-ahead forecast of the composite data; each predicted point is allowed to use all points before it back to 9455 BC. The inset shows a scatter plot between actual and forecasted decadal means, and the diagonal line.

Following Farmer and Sidorowich (1987) we compute the root-mean-square error $\sigma_\Delta = < [C_a(t) - C_0(t)]^2 >_{\text{test}}^{1/2}$ during the test period, normalize it by the standard deviation of the difference series $\sigma_D = < [D(t) - < D(t) >]^2 >_{\text{training}}^{1/2}$ during the period 9455BC—1005AD and calculate the normalized prediction error $E = \sigma_\Delta / \sigma_D$. Here < > means the time average, σ_D=7.39 is a constant in our calculations to make the results of different predictions equally scaled. The normalized prediction error *E* is the main criterion which we use in this work to estimate the accuracy of various predictions; this is the most common criterion which is widely used in the forecast practice. If *E*=0 the prediction is perfect, *E*=1 indicates that the performance is no better than a constant predictor

$D_0(t_{\text{test}}) = < D(t) >_{\text{training}} \equiv \text{const}$. Constant predictor obviously is very close to zero for the difference of large enough but constrained training dataset $< D(t) >_{\text{training}} = 0$ which corresponds to the inertial forecast $C_0(t) \equiv C_a(t\text{-}1)$ for the amplitude of the decadal sunspot number.

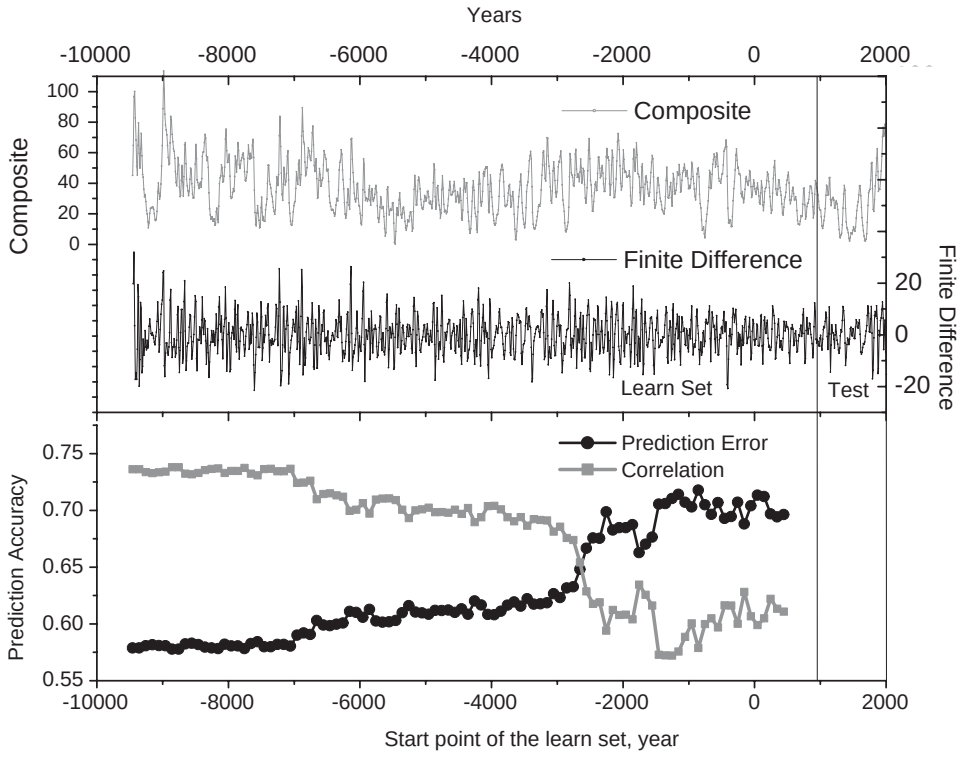

Figure 4. Variation of prediction accuracy with progressive removal of points from the training set. The curve labeled 'Composite' is the same as in Figure 2; 'Finite Difference' shows the successive differences of the composite data series; 'Prediction Error' is the normalized prediction error averaged during the test period (after vertical line); 'Correlation' is the correlation between actual and predicted successive differences.

From the theoretical consideration by Farmer and Sidorowich (1987) we should expect the power-law variation of the prediction error *E* as a function of the number of data points *N*; the prediction error decreases with the increase of the length of the training set. On the contrary, this decrease in *E* is limited by the growth in the noise amplitude in the past. The confrontation between these two tendencies leads to the stabilization of the error at 7000-8000 BC, Figure 4, bottom panel. This leads also to the conclusion that the accuracy of the prediction now is limited by the errors in data, not by the length of the SN_{RC} time series. So

an increase in the accuracy of the radiocarbon calibration curve in future may lead to an increase in the accuracy of this solar activity prediction method.

4. Errors in Data and Prediction Error

The prediction error consists of errors in the data and in the prediction method or model. The total average error of SN_{RC} reported by Solanki *et al.* (2004) is somewhat greater than the standard deviation σ_D of the SN_{RC} difference series during the training period. The coming decadal value seems to be unpredictable because it is inside the error of the previous value. However, the reported error is mostly the systematic (long-term) error and can be reduced significantly after the differentiation procedure (Section 3). The total error is defined by Solanki *et al.* (2004) as follows:

$$\sigma_0 = \Delta_m + \Delta_\tau + \sqrt{\sigma_{14C}^2 + \sigma_{GEO}^2 + \sigma_{CR}^2 + \sigma_{OF}^2 + \sigma_{NL}^2} \qquad (3)$$

We summarize various sources of the total error in Table 2.

Table 2 Sources of total SN_{RC} error.

Error	Average value, SN	Description	Systematic /short term
Δ_m	2-10	Uncertainties in the pre-Holocene ^{14}C production rate	Systematic
Δ_τ	1.2	Conversion of open magnetic flux into sunspot number	Systematic
σ_{14C}	4.3	$\Delta^{14}C$ measurement errors	Short term
σ_{GEO}	5.6	Uncertainties in the geomagnetic dipole moment	Systematic
σ_{CR}	4.6	Conversion of 14C production rate into cosmic ray flux	Systematic
σ_{OF}	0.5	Conversion of cosmic-ray flux into open magnetic flux	Systematic
σ_{NL}	2	Non-linear model for computing the open flux from the SN	Systematic

If we use the procedure of differentiation (Section 3) to remove the systematic errors, the principal source of error in the difference series turned out to be due to $\Delta^{14}C$ measurement error, which is approximately 30% of total error. $\Delta^{14}C$ measurement errors are from INTCAL98 (Stuiver *et al.*, 1998) errors. These errors are mostly non-systematic, but they may be overestimated during their conversion to SN_{RC} errors. INTCAL98 was constructed by taking an average of all data binned inside of each 10- year window. Sub-decadal measurements, *e.g.*

measurements of annual one-ring or semi-decadal five-ring samples were treated as if they were decadal (Reimer *et al.*, 2004). This procedure may lead to over-estimation of 5 - 10 times of the error in SN_{RC} due to significant contribution from the regular 11-year solar activity variation to this error estimation. We summarize that the error in the SN_{RC} difference series may be only a few percent of the total error reported for the decadal data because of two reasons:

- Differentiation removes the systematic errors
- SN_{RC} errors may be overestimated because sub-decadal measurements were treated as if they were decadal

Unfortunately it is impossible to quantify the remaining percent of error (the true error in the difference series) which would influence our prediction method from the consideration of the data error only. This approach would require too many details about the error evolution of all sources of data (Table 2) which are involved in the calculations of SN_{RC}. A brief summary is that it is much smaller than the reported total error.

The question arises, how to estimate the true error of the SN_{RC} difference series? We calculate the prediction error during the test period as a function of the variable data error in the training set using the Monte-Carlo approach. We construct a number of synthetic series $Y_i(t)$ during the training period, so that:

$$Y_i(t) = SN_{RC}(t) + \alpha N_i \sigma_0(t) \qquad (4)$$

where N_i are vectors of random numbers from a normal distribution with unit variance, α is a regulator of noise amplitude, *e.g.* α=0.5 simulates the series $Y_i(t)$ where noise amplitude is equivalent to 50% of reported total error. From each series we calculate the corresponding one-step-ahead forecast of 100 points of the test series and its prediction error E_i with the method of Section 3. Then E_i are averaged after 20 random runs with the fixed parameter α.

These average values as a function of α are shown in Figure 5. We should expect that, at some small α the decrease in the prediction error will stop and this α will show the true error of the data. However, as a matter of fact we can see only some smoother decrease in α<0.02 (2% of total error). The asymptote of unpredictability E=1 with 50-100% noise added is consistent with σ_D=7.39 which

is 67% of the reported error $<\sigma_0>=11$.

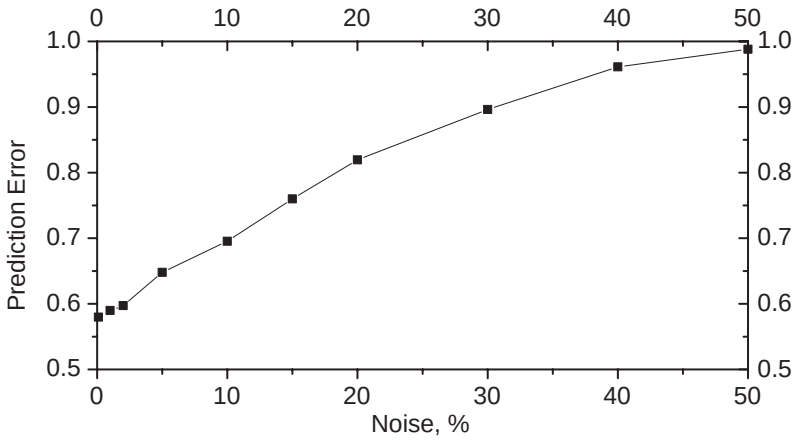

Figure 5. Variation of the prediction error with noise added to the training set. The abscissa is the Gaussian noise added to the training set relative to the error bars calculated by Solanki *et al.* (2004).

The LL forecast method can also help to test the non-linearity of time series (Kantz and Schreiber, 2004) using the surrogate data approach. This non-linearity test serves to distinguish between the stochastic (autoregressive moving average - ARMA) process and low-dimensionality chaos via the calculation of the prediction error. The time series is termed non-linear if the test set is predicted significantly better when the training is made on the actual time series compared with the training on its surrogate data which simulate multiple realizations of a stochastic process. First, we assume the null hypothesis of a monotonically rescaled Gaussian linear stochastic process. Next we need surrogates with a given distribution and given linear correlations which are similar to that in our composite data. This is approximately achieved by the amplitude-adjusted Fourier transform (AAFT) algorithm. The Fourier transform of the data is multiplied by random phases and then transformed back, conserving the sample periodogram, see *e.g.* Kugiumtzis (2000). Here we use the realization of AAFT provided by Barnett and Wolff (2005) to produce the surrogate data from the training set[2]. The prediction error is averaged during the test period for each of 100 randomized training sets.

[2] The algorithm is available at http://www.mathworks.com/matlabcentral/fileexchange/ , File Id 16062.

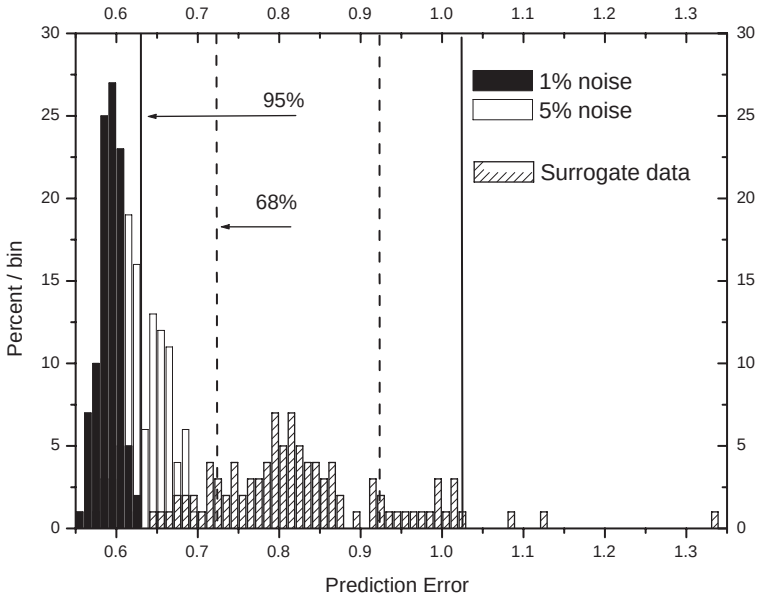

Figure 6. The non-linearity test. The '% Noise'-histogram shows the distribution of prediction error over 100 random runs with noise of corresponding amplitude is added to the training set. The 'Surrogate data'-histogram shows the distribution of prediction error over 100 random runs of AAFT algorithm to randomize the learn set. The vertical solid and dashed lines show 95% and 68% intervals, respectively, for the distribution of the prediction error with the surrogate data.

The histogram (Figure 6) shows a broad distribution with average 0.83 which is significantly smaller than expected for pure random (Gaussian) processes which are unpredictable (average error not smaller than 1). The histogram deviates significantly from the histogram of the prediction error for the composite difference series if the noise amplitude does not exceed 1% of total error, *i.e.* the hypothesis of non-linearity is strongly supported for this level of error in the data. The non-linearity hypothesis is still significant for 5% error in the data (Figure 6). For a higher level of the noise the non-linearity cannot be proven with this method.

5. Discussion

In this investigation we developed the "state of the art" prediction technique on the basis of the local linear approximation method. The main result is the statistical evidence that step-by-step changes of decadal SN_{RC} contain enough information to predict step-by-step changes of corresponding decadal numbers derived from the telescopic observations, despite these changes are within the

error reported by Solanki *et al.* (2004). Assuming that the true error in SN_{RC} difference series is small enough we can also prove the non-linearity of the decadal changes of sunspot numbers. These results are mathematically defined and statistically quantified but they have several interesting consequences for understanding of the physics under the long-term variations of solar activity and need an additional discussion.

Currently there are several hypotheses of how the secular variation of solar activity (Gleissberg cycle) is originated. Random fluctuation or "eruptive" hypothesis, with "memory" not exceeding the length of the 11-year cycle was started by Waldmeier. Assuming this hypothesis we should have the normalized prediction error $E=1$ for any statistical method, so that our calculations do not support this hypothesis. The random number generator which was included in the Babcock-Leighton scheme to modulate the cycles is not supported either, because we should have found no "analogs" in the historical record and would have obtained $E=1$ with the LL-method. Also a sum of the hypothetical periodical processes with noise contamination should be rejected as a candidate to describe the long-term variation, as it was shown by the non-linearity test. Our calculations are consistent with chaotic modulation that appeared as an envelope of the 11-year cycle produced by some deterministic chaotic system (Weiss, 1985), or with a hypothesis of torsion oscillations which produces the long-term cycles (Kitchatinov *et al.*, 1999) by non-linearity (Volobuev, 2006).

Another question to be asked is: why the LL-method was used? The LL-method continues to be one of the most powerful methods for the statistical forecast. A similar method was used by Sello (2001) to forecast the smoothed monthly sunspot number. The accuracy of the method is much higher if it is compared with the approach by Ogurtsov (2005) and probably higher than the accuracy that can be achieved in the approach by Clilverd *et al*. (2006) which is simulated with our surrogate data test. It cannot be compared with a traditional method, *e.g.* with the "precursor" method, because a very limited length (1-5 cycles) was available for these methods to be tested. Obviously, any other non-linear method can be used to forecast the difference series of synthetic composite data. The prediction error may be better if some techniques based on the artificial intelligence are applied. The LL-method, however, is often included in these techniques and our specifications may help to adjust them. On the other hand, the non-linear LL-

method helps us to understand the physics of the long-term variations of solar activity.

Conclusions

The method of local linear approximation has been adapted to forecast the coming decadal value of sunspot number. We conclude that:

- The coming decadal sunspot number is predictable with the use of the radiocarbon-based sunspot numbers
- Most likely, the decadal sunspot number reflects a non-linear process, not a stochastic process
- The next decadal sunspot number will not be higher than the previous one.

Acknowledgements

We thank Yury Nagovitsyn for helpful discussions. We also are indebted to an anonymous referee for several interesting questions which had led to considerable extension of the initial content of this work. The work is supported by grants: Program of the Presidium of Russian Academy of Science (RAS): “Solar activity and physical processes in the Sun-Earth system”, Russian Fund of Basic Research N 07-02-00379-a, 05-07-90107, Russian Science Support Foundation, Saint-Petersburg Scientific Center of RAS.

References

Barnett, A.G., Wolff, R.: 2005, *IEEE Trans. Signal Processing* **53**, 26.

Clilverd, M.A., Clarke, E., Ulich, T., Rishbeth, H., Jarvis, M.J.: 2006, *Space Weather* **4**, S09005.

Farmer, J.D., Sidorowich, J.J.: 1987, *Phys. Rev. Lett.* **59**, 845.

Hoyt, D. V., Schatten, K. H.: 1996, *Solar Phys.* **165**, 181.

Kane, R.P.: 2007, *Solar Phys.* **243**, 205.

Kantz, H., Schreiber, T.: 2004, *Nonlinear Time Series Analysis*, 2nd Ed., Cambridge University Press, Cambridge, 369.

Kitchatinov, L.L., Pipin, V.V., Makarov, V.I., Tlatov, A.G.: 1999, *Solar Phys.* **189**, 227.

Knobloch, E., Landsberg, A. S.: 1996, *Monthly Notices Roy. Astron. Soc.* **278**, 294.

Kugiumtzis, D.: 2000, *Phys. Rev. E* **62**, 1.

Lorenz, E.N.: 1969, *J. Atmos. Sci.* **26**, 636.

Marchal, O.: 2005, *Climate Dyn.* **24**, 71.

Muscheler, R., Joos, F., Müller, S.A., Snowball, I.: 2005, *Nature* **436**, E3.

Muscheler, R., Joos, F., Beer, J., Müller, S., Vonmoos, M., Snowball, I.: 2007, *Quart. Sci. Rev.* **26**, 82

Nagovitsyn, Yu.A.: 2005, *Astron. Lett.* **31**, 557.

Nicolis, C.: 1998, *J. Atmos. Sci.* **55**,465.

Ogurtsov, M.G.: 2005, *Solar Phys.* **231**, 167.

Ostryakov, V. M., Usoskin, I.G.: 1990, *Solar Phys.* **127**, 405.

Packard, N. H., Grutchfield, J. P., Farmer, J. D., Shaw, R. S.: 1980, *Phys. Rev. Lett*. **45**, 712.

Reimer, P. J., Baillie, M. G. L., Bard, E., Bayliss, A., Beck J. W., Bertrand, C. J. H., *et al.* : 2004, *Radiocarbon* **46**, 1029.

Sello, S.: 2001, *Astron. Astrophys.* **377**, 312

Solanki, S. K., Usoskin, I. G., Kromer, B., Schüssler, M., Beer, J.: 2004, *Nature* **431**, 1084.

Stuiver, M., Quay, P. D.: 1980, *Science* **207**, 11.

Stuiver, M., Reimer, P.J., Bard, E., Beck, J.W., Burr, G.S,, Hughen, K.A., Kromer, B., McCormac, G., van der Plicht, J., Spurk, M.:1998, *Radiocarbon* **40**, 1041.

Usoskin, I.G., Kovaltsov, G.A.: 2004, *Solar Phys.* **224**, 37.

Volobuev, D.: 2006, *Solar Phys.* **238**, 421

Volobuev, D.M., Nagovitsyn Yu.A., Jungner, H., Ogurtsov, M.G., Miletsky, E.V.: 2004, In: Stepanov, A.V., Benevolenskaya, E.E., Kosivichev, A.G.(eds.) *Multiwavelength Investigations of Solar Activity, IAU Symp.* **223**, *565.*

Weiss, N.O.: 1985, *J. Stat. Phys.* **39**, 477.