Journal of Atmospheric and Solar-Terrestrial Physics 67 (2005) 595–603

Journal of ATMOSPHERIC AND SOLAR-TERRESTRIAL PHYSICS

www.elsevier.com/locate/jastp

Solar activity forecast: Spectral analysis and neurofuzzy prediction

Ali Gholipour[a,b,*], Caro Lucas[a,b], Babak N. Araabi[a,b], Masoud Shafiee[c]

[a]Electrical and Computer Engineering Department, Control and Intelligent Processing Center of Excellence, University of Tehran, Tehran, Iran

[b]School of Cognitive Sciences, Institute for studies in theoretical Physics and Mathematics (IPM), Tehran, Iran

[c]Faculty of Electrical Engineering, Amir Kabir University of Technology, Tehran, Iran

Received 22 October 2003; received in revised form 24 November 2004; accepted 7 December 2004

Abstract

Active research in the last two decades indicates that the physical precursor and solar dynamo techniques are preferred as practical tools for long-term prediction of solar activity. But why should we omit more than 23 cycles of solar activity history, and just use empirical methods or simple autoregressive methods on the basis of observations for the latest eight cycles? In this article, a method based on spectral analysis and neurofuzzy modeling is proposed that is capable of issuing very accurate long-term prediction of sunspot number time series. A locally linear neurofuzzy model is optimized for each of the principal components obtained from singular spectrum analysis, and the multi-step predicted values are recombined to make the sunspot number time series. The proposed method is used for solar cycles 22 and 23 and the results are remarkably good in comparison to the predictions made by solar dynamo and precursor methods. An early prediction of the maximum smoothed international sunspot number for cycle 24 is 145 in 2011–2012.

Keywords: Solar activity; Sunspot numbers; Forecasting; Singular spectrum analysis; Neurofuzzy modeling; Long-term prediction

1. Introduction

Most of the space weather phenomena are influenced by variations in solar activity. During the years of solar maximum there are more solar flares causing significant increase in solar cosmic ray intensity. The high-energy particles disturb communication systems and affect the lifetime of satellites. Coronal mass ejections and solar flares are the origin of shocks in solar wind and cause geomagnetic disturbances in the earth's magnetosphere. The high rate of geomagnetic storms and sub-storms results in atmosphere heating and drag of Low Earth Orbit (LEO) satellites. Long-term solar activity forecasting is especially useful to space mission centres because the orbital trajectory parameters of satellites are greatly affected by the changing solar activity. There are also many arguments in recent years on correlations between solar activity and terrestrial climate (Laut, 2003; Oh et al., 2003; Tsiropoula, 2003). The level of solar activity is usually expressed by Zurich or International sunspot number (R_Z) or the 2800 MHz (10.7 cm) radio flux ($F10.7$) index. The sunspot numbers are commonly used in numerical analysis because they are

*Corresponding author. Electrical and Computer Engineering Department, Control and Intelligent Processing Center of Excellence, University of Tehran, Tehran, Iran. Fax: (+9821) 872 5029.

E-mail addresses: gholipoor@ut.ac.ir, ali_gholipour@yahoo.com (A. Gholipour), lucas@ipm.ir (C. Lucas).

1364-6826/$ - see front matter
doi:10.1016/j.jastp.2004.12.001

available for more than 23 cycles, while the $F10.7$ index is preferred in space weather studies because of its higher signal-to-noise ratio and direct relationship to the solar EUV-UV emission. The correlation coefficient between the two indices is 0.987 and a simple least-squares estimation yields the following formula (Sofia et al., 1998):

$$F10.7 = 58.7 + 0.918 R_Z.$$

Much effort has been dedicated to find physical evidence for the solar activity, to discover relationships between the level of activity and the effects of solar events on earth and environment, which is useful in reconstructing the sunspot number time series; and, interestingly, observations like the records of the ^{11}Be concentration in polar ice show that the period of high solar activity during the last 60 years is unique throughout the past 1150 years (Usokin et al., 2003). So, although the solar activity presents some clear periodicities, its prediction is quiet difficult; and despite the many statistical and experimental arguments of finding meaningful structures in the observations, it more and more shows long-term and large-scale changes. However, with the current accurate records of up to 300 years, seeking for long-term predictions of about 10 years is reasonable and seems feasible.

Various numerical prediction techniques have been used for the sunspot number time series, e.g. Fourier analyses, curve fitting, artificial intelligence, and neural networks (Gholipour et al., 2003). These methods, although very accurate in short-term predictions, are not reliable in long term. The geomagnetic precursor techniques are based on statistical relationships between geomagnetic activity indices in the declining phase of one cycle (at the end of the cycle) and the activity level of the next solar maximum (Brown, 1992; Thompson, 1993; Joselyn, 1997). The first studies have used the geomagnetic *aa* index for prediction, while it is argued that the predictions on the basis of *Ap* index are more accurate (Sofia et al., 1998). All the precursor methods are investigated via the statistical tests for just eight solar cycles including eight numbers for solar maximums, and with such few data samples the best fitting is linear autoregressive estimation, and the reliability of the model is doubtful. Schatten et al. proposed a more acceptable method on the physical basis of solar dynamo theory and introduced the solar dynamo amplitude (SODA) index to predict the solar activity several years in advance (Schatten et al., 1978; Schatten and Sofia, 1987). This method has been used several times to forecast the last two cycles (Schatten and Pesnell, 1993; Schatten et al., 1996; Schatten, 2002), and seems to be more reliable than the common precursor methods, but has not been validated yet. The current forecasting approaches are developed in the direction of ignoring the long history of sunspot number time series and trying to find simple empirical models on the basis of a few observations of the peak of the latest cycles and the geomagnetic indices. The timescale of these indices is much smaller than the variations in solar cycle, and their long-term trends cannot reveal the nonlinearities or complex periodicities of solar activity.

In this article, a decomposition method based on singular spectrum analysis is used to make an intuitive nonlinear black box modeling technique applicable to long-term prediction of sunspot number time series. The singular spectrum analysis (SSA) (Vautard and Ghil, 1989; Vautard et al., 1992) performs a data adaptive filtering in the lag coordinate space of data and yields the principal components of the time series, which have narrow band frequency spectra and obvious temporal patterns. The principal components can be classified into linear or nonlinear trends, periodic or quasi-periodic patterns, and colored noise components, most of which are long-term predictable. Thus, after adaptive noise cancellation, one can design appropriate models for the components, and recombine the predicted components to obtain the original time series. The proposed method is successfully tested for the 22nd and 23rd solar cycles and the results are compared to the predictions by precursor and solar dynamo techniques. The mathematical description of the SSA algorithm is presented in Section 2; the locally linear neurofuzzy modeling technique along with the appropriate learning algorithm is considered in Section 3, and the results and their comparison are presented in Section 4. The last section contains some concluding remarks.

2. Spectral analysis

SSA is defined as a new tool to extract information from short and noisy chaotic time series (Vautard et al., 1992). It relies on the Karhunen–Loeve decomposition of an estimate of covariance matrix based on M lagged copies of the time series. Thus, as the first step, the embedding procedure is applied to construct a sequence $\{\tilde{X}(t)\}$ of M-dimensiona vectors from time series $\{X(t) : t = 1, \ldots, N\}$:

$$\tilde{X}(t) = (X(t), X(t+1), \ldots, X(t+M-1)),$$
$$t = 1, \ldots, N', \quad N' = N - M + 1. \tag{1}$$

The $N' \times M$ trajectory matrix (D) of the time series has the M-dimensional vectors as its columns, and is obviously a Hankel matrix (the elements on the diagonals $i+j = const$ are equal). In the second step, the $M \times M$ covariance matrix C_X is calculated and its eigenelements $\{(\lambda_k, \rho_k) : k = 1, \ldots, M\}$ are determined by singular value decomposition (SVD). Each eigenvalue, λ_k, estimates the partial variance in the direction of ρ_k, and the sum of all eigenvalues equals the total