

Long-Term Sunspot Number Prediction based on EMD Analysis and AR Model *

Tong Xu¹, Jian Wu², Zhen-Sen Wu¹ and Qiang Li³

¹ School of Science, Xidian University, Xi'an 710071, China; xutong1104@126.com

² National Key Laboratory of Electromagnetic Environment, China Research Institute of Radio Wave Propagation, Beijing 102206, China

³ IAP, University of Rostock, Schlosstr. 6, Kuehlungsborn, 18225, Germany

Received 2007 August 19; accepted 2007 December 14

Abstract The Empirical Mode Decomposition (EMD) and Auto-Regressive model (AR) are applied to a long-term prediction of sunspot numbers. With the sample data of sunspot numbers from 1848 to 1992, the method is evaluated by examining the measured data of the solar cycle 23 with the prediction: different time scale components are obtained by the EMD method and multi-step predicted values are combined to reconstruct the sunspot number time series. The result is remarkably good in comparison to the predictions made by the solar dynamo and precursor approaches for cycle 23. Sunspot numbers of the coming solar cycle 24 are obtained with the data from 1848 to 2007, the maximum amplitude of the next solar cycle is predicted to be about 112 in 2011–2012.

Key words: Sun: sunspots — Sun: activity

1 INTRODUCTION

Most of space weather phenomena are affected by solar activity. Disturbances by the solar activity in the forms of electromagnetic radiation, solar wind and enhanced solar cosmic rays come to near-Earth space and give rise to variations of the solar-terrestrial environment in the magnetosphere, ionosphere and neutral atmosphere, which may affect radio communication, orbital lifetime of low Earth-orbit satellites and weather patterns. Thus study on solar activity prediction has been developed as a project in space environment prediction.

The sunspot number is a significant index of the solar activity because of its availability and reliability. Various numerical prediction techniques have been used for the sunspot number time series, e.g. curve fitting, artificial intelligence, neural networks and so on. These approaches, although very accurate in short-term predictions, are rather unreliable in long term (Gholipour et al. 2003). The geomagnetic precursor method proposed by Ohl (1966) has been used frequently for sunspot number prediction (Brown 1992; Thompson 1993; Lantos & Richard 1998; Schatten 2002). Because of its weak physical foundation, the reliability of the model is doubtful (Sello 2001; Gholipour et al. 2005). Schatten et al. (1987) proposed a physically more acceptable method on the basis of the solar dynamo theory and used the solar dynamo amplitude index to predict the solar activity several years in advance. This method has been used several times to predict the last two cycles (e.g. Schatten & Sofia 1993; Schatten et al. 1996; Sofia et al. 1998; Choudhuri et al. 2007), but has not been validated yet. Furthermore, some of the predictions using a non-geophysical method are successful for the recent two cycles (e.g. Thompson 1997; Kane 1999; Wang et al. 2002; Lantos 2006). Gholipour et al. (2005) proposed a method based on spectral analysis and neurofuzzy

* Supported by the National Natural Science Foundation of China.

modeling, which predicted successfully the sunspot numbers of the latest two cycles. Wang et al. (2002) obtained the maximum amplitude of solar cycle 24 using statistical characteristics of solar cycles. Echer (2004) used spectral characteristics of sunspot number and obtained the values of cycles 23 and 24. Maris & Oncica (2006) made prediction of cycle 24 based on neural network. All these methods need further verification. The long-term prediction of sunspot number, when the lead time is 2 years or more, is still an open question.

The Empirical Mode Decomposition (EMD), also known as the Hilbert-Huang Transformation (HHT) is probably one of the most important discoveries in the field of applied mathematics in the NASA's history (Huang 1998). It provides a more efficient filtering of signal from noise for non-linear and non-stationary data. For the last 200 years, Fourier analysis has been used for the analysis of time-varying data, which essentially breaks down the data into an infinite series of linear sinusoidal functions with variable coefficients. The EMD method is wholly different because the result produced by the EMD is the only one that is adaptive (with no a priori basis assigned), and time varying (the frequency is a function of time defined by differentiation rather than conventional analysis as in the Fourier type analysis, e.g. Wavelet analysis). The EMD method has been successfully applied in meteorological field, e.g. the extraction of the 11-year solar cycle in the stratosphere (Coughlin & Tung 2003), and it was also applied to solar activity (Li et al. 2007), in which the extracted periodic components of sunspot number are consistent with the well-known components. Furthermore, EMD has been extensively used in signal analysis in some other fields (Huang 1998; Chen & Li 2002).

In this paper, we propose an approach based on the EMD method and the Auto-Regressive method (AR), and significant results have been achieved.

2 EMD ANALYSIS OF SUNSPOT NUMBER

Sunspot number data are obtained from the Space Environment Center of NOAA. We selected 12-month running means of monthly mean sunspot number from January 1848 to December 1992 as our sample and the predicted values are tested against data from January 1990 to May 2007. Lastly, we used data from January 1848 to May 2007 to predict the sunspot numbers of the cycle 24.

Using the EMD method to decompose the data, we selected a cubic spline line as the upper envelope following Huang (1998). The sunspot numbers are decomposed into six modes and one trend, see Figure 1. Here IMF1 to IMF2 are the high frequency components obtained through Fast Fourier Transformation (FFT), with periods 1.3- to 1.4-year, IMF3 is a Quasi-biennial Oscillation (QBO) component. IMF4 is the primary component of the original data with an amplitude much bigger than the others: it has an obvious 11-year variation, with characteristically shorter ascending times than descending times. IMF4 is very similar to the original sunspot numbers and their correlation coefficient is 0.8963. The rest are the Hale periodic component, Double-Hale periodic component and the trend, respectively. The instantaneous frequency of each IMF can be read off the Hilbert spectrum shown in Figure 2. It is obvious that the top line varies around the average frequency ~ 0.1 (11-year), and the others vary around ~ 0.05 and ~ 0.025 (22-year and about 40-year, respectively). IMF1 to IMF3 are not shown adequately because of their high frequency. The physical mechanism of each periodic component has been outlined in Li et al. (2007).

As shown in Figure 3, time series is reconstructed by IMF4–IMF7 and smoother than the observed series, which is easier to predict consequently. The correlation coefficient of the reconstructed and observed series reaches 0.9958, thus IMF4–IMF7 almost represent the variation of sunspot number series.

3 AR PREDICTION

AR is commonly used in nonlinear prediction, and is also often used in long-term prediction of sunspot number (e.g. Yule 1927; Ghaddar & Tong 1981; Jian et al. 1996; Muslim 2001). For a stationary and zero-mean time series, it can be written as

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_n x_{t-n} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_n \varepsilon_{t-n} + \varepsilon_t, \quad (1)$$

where φ_i is the auto-regressive coefficient, θ_i the smoothed mean parameter, and ε_t the random noise $\sim N(0, \sigma^2)$. When $\theta_i = 0$, the model is called the n order auto-regressive model and is given by (Kay &

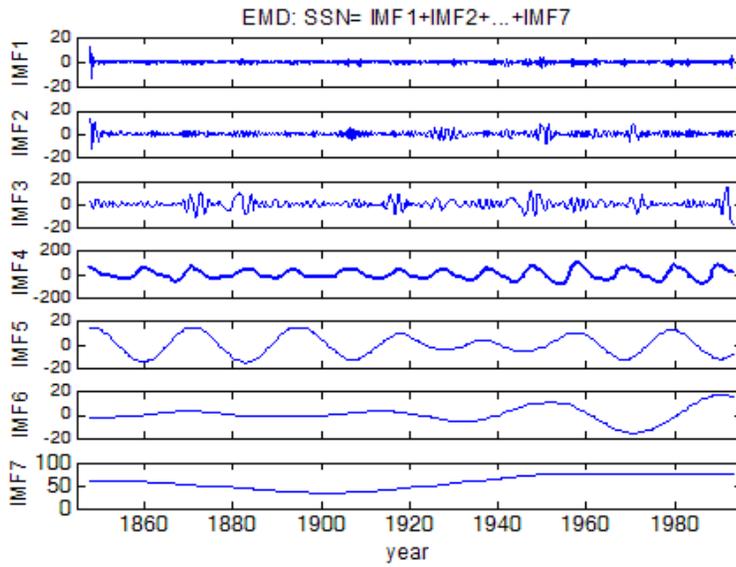


Fig. 1 Sunspot number data from 1848 to 1992 decomposed into six modes and one trend by the EMD method.

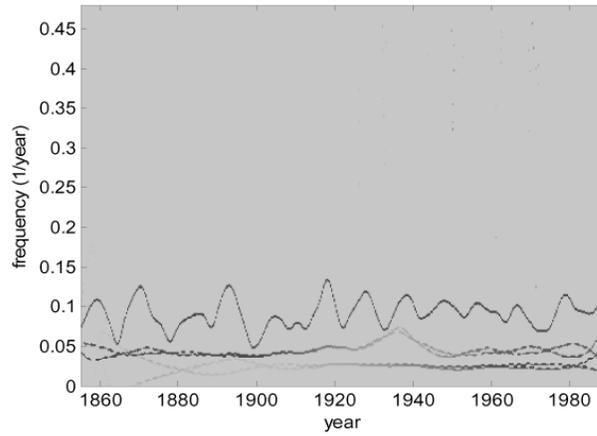


Fig. 2 Hilbert spectrum for the sunspot number. The skeleton lines represent each of the IMFs.

Mamplé 1981)

$$x_t = \sum_{i=1}^n \varphi_i x_{t-i} + \varepsilon_t, \tag{2}$$

in which $n = 132$.

In order to evaluate the validity of AR prediction, we examined the predicted values of the 11-year component (IMF4) with the “real” 11-year component values of the cycle 23 based on the data from January 1848 to May 2007. Because IMF4 almost represents the sunspot number variation, its prediction is critical

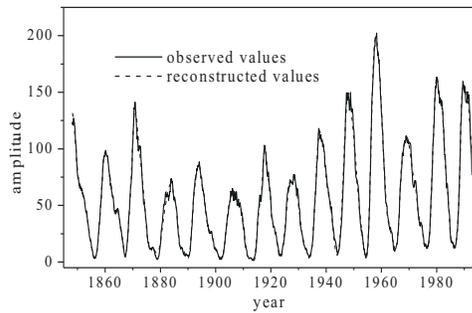


Fig. 3 Original time series and reconstructed series using EMD.

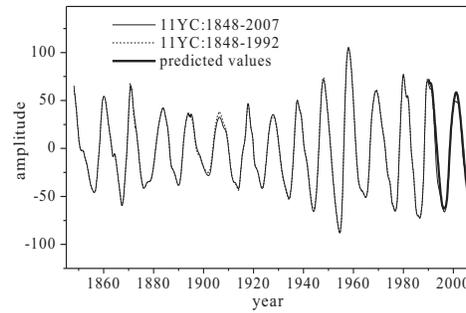


Fig. 4 Predicted values from 1990 and 11-year cycle components from 1848 to 1992 and 1848 to 2007.

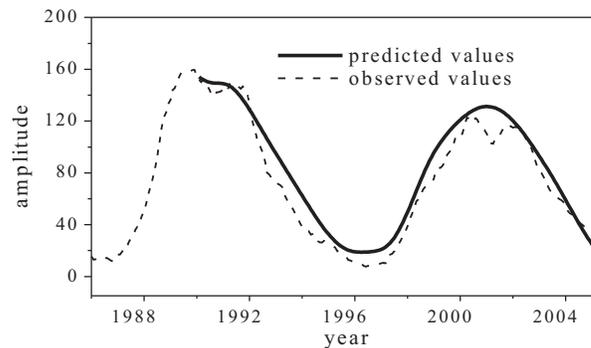


Fig. 5 Predicted values from 1990 and the observed values.

for the whole prediction. The predicted values are shown in Figure 4 as well as the 11-year component from January 1848 to May 2007 decomposed by the EMD method. The dashed line is the 11-year component from January 1848 to December 1992 and the thin solid line is the 11-year component from January 1848 to May 2007, decomposed by EMD. From Figure 4, it is evident that the two are almost coincident for the overlapped interval, indicating high reliability of the EMD result. The predicted values from 1990 are shown by the bold solid line. The predicted maximum is 58 and the “real” maximum is 49, so the relative error is 18.4% and the root-mean-square error is 11.6.

In Figure 5, the predicted values of IMF1–IMF7 by use of AR multi-step prediction are reconstructed and the predicted values are largely consistent with the observed values. The predicted maximum of solar 23 is 132 and the observed is 122, so the relative error is 9.8% and the root-mean-square error 13.9. Because the maximum amplitude of cycle 23 and cycle 22 both exceeded 150, it means much difficult in predicting accurately the sunspot number of cycle 23, and many predicted results are higher than the observed value. Schatten et al. (1993) obtained a maximum amplitude of 170 ± 25 by use of solar dynamo prediction. Joselyn (1997) predicted the maximum amplitude to be 160 ± 30 using geomagnetic precursor. With non-geophysical methods, Tian (1997) claimed that the maximum amplitude was to be about 190 based on back-propagation (BP) neural network. Using singular spectrum analysis (SSA) and neurofuzzy analysis, Gholipour et al. (2005) predicted a maximum of 141.

The data from January 1848 to May 2007 give the 12-running mean values of sunspot numbers of cycle 24 using EMD and AR prediction, see Figure 6. Because the short time scale components (IMF1–IMF3) were ignored, the predicted values are smooth. The maximum amplitude of cycle 24 is predicted to be about 112 and the peak time is 2011–2012. Wang et al. (2002) predicted a maximum of 83.2–119.4 and Echer et

al. (2004) obtained a maximum amplitude of 115 ± 13.2 in 2012. These results are comparable with ours. However, Gholipour et al. (2005) argued that the maximum amplitude was to be 145, while Badalyan et al. (2001) predicted a low cycle with a maximum amplitude not exceeding 50. Time will tell which one is the more accurate.

The sum of the ignored components IMF1–IMF3 is shown in Figure 7, in which the vibration amplitudes are dramatic near the peak time. We segment the 14-cycle data into 14 sections starting from 1856, the beginning of solar cycle 10, and use the calculated root-mean-square as the prediction error for solar cycle 24,

$$\sigma(j) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N [\text{IMF}(j)]^2}, \quad (3)$$

where $j = 1, 2, \dots, 132$ and $N = 14$. The RMS error is shown in Figure 6. Considering the prediction error, the upper limit of the maximum amplitude is about 120.

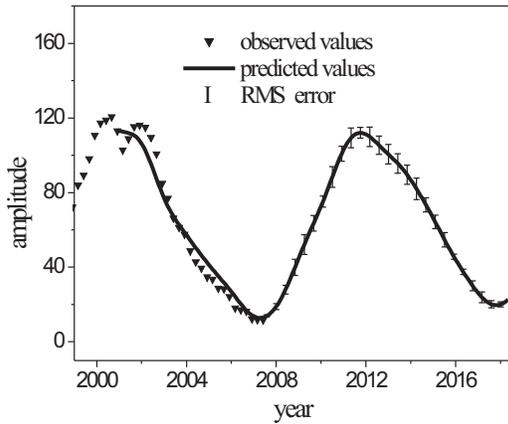


Fig. 6 Predicted values from 2001 and the observed values.

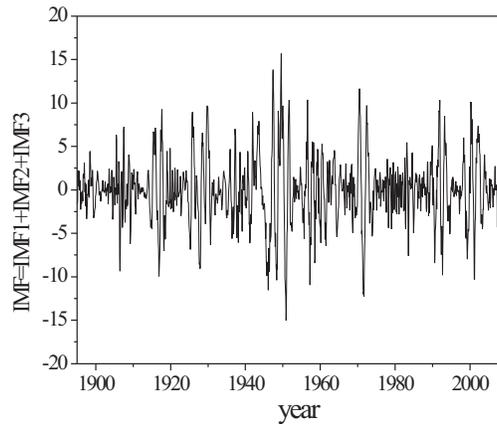


Fig. 7 Summation of IMF1 to IMF3.

4 SUMMARY

The EMD method has advantages in the extraction of different time scale components. We attempted to apply the method to a long-term prediction of sunspots, and found the proposed method to be successfully tested for solar cycle 23. With the EMD method and AR multi-prediction, we obtained for solar cycle 24 a maximum amplitude of sunspot number of about 112 in 2011–2012. We also specified the prediction error for each month of the coming solar cycle using root-mean-square of the ignored high-frequency components IMF1–IMF3.

Acknowledgements The authors thank the unknown referees for helpful comments. This work is supported by the NSFC (National Natural Science Foundation of China) under Grant 40310223, and by the National Key Laboratory of Electromagnetic Environment (LEME) under Grant 9140C0801060803.

References

- Badalyan O. J., Obridko V., Sykora N. J., 2001, *Solar Phys.*, 199, 421
- Brown G. M., 1992, *Ann. Geophys.*, 10, 453
- Coughlin K. T., Tung K. K., 2003, *Adv. Space. Res.*, 34, 323
- Chen C. H., Li C. P., *Terr. Atmos. Ocean. Sci.*, 2002, 13, 171
- Choudhuri A. R., Piyali C., Jie J., 2007, *Phys. Rev. Lett.*, 98, 131103
- Echer E., Rigozo N. R., Nordemann D. R. et al., 2004, *Ann. Geophys.*, 22, 2239
- Ghaddar D. K., Tong H., 1981, *Appli. Statistics*, 30, 238
- Gholipour A., Abbaspour A., Araabi B. N. et al., 2003, *Proceedings of MIC2003*, p. 158
- Gholipour A., Caro Lucas, Babak N. A. et al., 2005, *J. Atmos. Terr. Phys.*, 67, 595
- Huang N. E., 1998, *Prog. Roy. Soc. Lond.*, 454, 903
- Jan L., Hanna K., Esa R., 1996, *Oikos*, 3, 565
- Joselyn J. A., 1997, *Eos Transactions American Geophysical Union*78, 211
- Lantos P., Richard O., 1998, *Solar Phys.*, 182, 231
- Kane R. P., 1999, *Solar Phys.*, 189, 217
- Kay S. M., Mample S. L., 1981, *Proc. IEEE*, 69, 1380
- Lantos P., 2006, *Solar Phys.*, 229, 373
- Li Q., Wu J., Xu Z. W. et al., 2007, *Chin. J. Space Sci.*, 27, 1 (in Chinese)
- Sperl M., 1998, *Comm. in Asteroseismology*, 111, 1 (Univ. of Vienna)
- Maris G., Oncica A., 2006, *Sun & Geospace.*, 1
- Schatten K. H., Pesnell W. D., 1993, *Geophys. Res. Lett.*, 20, 2257
- Schatten K. H., Sofia S., 1987, *Geophys. Res. Lett.*, 14, 623
- Schatten K. H., Myers D. J., Sofia S., 1996, *Geophys. Res. Lett.*, 23, 605
- Schatten K. H., 2002, *J. Geophys. Res.*, 107, 1377
- Sello S., 2001, *A&A*, 377, 312
- Sofia S., Fox P., Schatten K. H., 1998, *Geophys. Res. Lett.*, 25, 4149
- Tian J. H., 1997, *Chin. J. Space Sci.*, 17, 255 (in Chinese)
- Thompson R. J., 1993, *Solar Phys.*, 148, 383
- Thompson R. J., 1997, In: G. Hechman, K. Marubashi, M. A. Shea, et al. eds., *Solar-Terrestrial Predictions-V*, Tokyo: Hiraiso Center, p.97
- Wang J. L., Gong J. C., Liu S. Q. et al., 2002, *Chin. J. Astron. Astrophys. (ChJAA)*, 2, 396
- Wang J. L., Gong J. C., Liu S. Q. et al., 2002, *Chin. J. Astron. Astrophys. (ChJAA)*, 2, 557
- Yule G. U., 1927, *Phil. Trans. Roy. Soc. London*, 226, 267