Memoirs of the Faculty of Engineering, Kyushu University, Vol. 55, No. 1, March 1995

Reconstruction of Chaotic System Equations and Prediction of Monthly Sunspot Time Series

by

Kenji JINNO*, Shiguo XU**, Ronny BERNDTSSON***,

Akira KAWAMURA**** and Minoru MATSUMOTO*****

(Received January 6, 1995)

Abstract

Modeling of sunspots is important since they indicate the relative activity of the sun which in turn may influence different terrestrial properties. To predict the nonlinear and chaotic behavior of sunspot time series the problem of reconstructing underlying system equations is studied. The proposed procedure for this is; 1) based on the behavior of observed time series and strange attractor, find reference system equations that show similar basic features as the time series, 2) assume a general structure of the governing system equations by Taylor series expansion, 3) use the reference system equations as initial state in an updating procedure (extended Kalman filter) to estimate the structure of the governing system equations. Results show that predictions up to eight months ahead can be made with good agreement after identifying the governing system equations. The extended Kalman filter was shown to be an efficient tool to identify parameter values and to make updated predictions of the chaotic system.

Keywords: Sunspots, Chaos, Time series, System equation, Prediction, Reconstruction, Extended Kalman filter, Linear transformation

1. Introduction

Analyses and prediction of sunspots have been made by numerous researchers during the latest decades[1,2,3,4]. The sunspot number is a quantitative coefficient of sun activity and therefore important for, e.g., global weather, satellite trajectories, geomagnetic variations,

* Professor, Department of Civil Engineering
** Visiting Professor, Department of Civil Engineering, on leave from Department of Civil Engineering, Dalian University of Technology, Dalian, China
*** Associate Professor, Department of Water Resources Engineering, Lund University, Lund, Sweden
**** Associate Professor, Department of Civil Engineering
***** Graduate Student, Department of Civil Engineering Hydraulics and Soil Mechanics

etc. Recently, studies that consider the chaotic behavior of the sun's behavior have shown that better predictions can be made by using developments within dynamical systems theory[4,5,6].

Many studies during recent years have indicated chaotic characteristics for sunspot time series[4,5,6,7]. The nonlinear evolution in time for sunspots may be one reason why models based on periodic behavior fail to predict the time series accurately[4]. Current sunspot forecasting methods can be grouped in three broad categories[4,23]; (1) statistical methods that assume fundamental periods in the solar cycle, (2) statistical methods which assume a certain behavior of the sun in a current cycle and together with the behavior in past cycles will give the future of the current cycle, and (3) precursor techniques which assume that the behavior of the solar magnetic field in the previous cycle determines the behavior of the present cycle. In this paper, we attempt to exploit a somewhat different possibility, namely, a method that embraces the entire and long-term behavior of the observed time series. We aim at trying to reconstruct the unknown governing equations for the observed time series. With the knowledge of these underlying equations it will be possible to make predictions of the future behavior of sunspots.

Although chaos limits predictability, repeated short-term forecasts may prove feasible[8]. Models based on chaotic premises do not contain any explicit random components. These models are based on purely deterministic but nonlinear equations, and pseudo-periodical behavior is observed if the system contains a so called strange attractor. If similar characteristics can be found in the original data, it is reasonable to believe that a description of the time series can be made using chaotic principles.

In Chapter 2 we outline a methodology that can be used to reconstruct unknown system equations that display chaotic behavior based on observed time series. The only information that we have about the unknown governing system equations is observed time series. Consequently, analysis of basic properties for these observations become crucial for the modeling. Several studies have shown that nonlinear analyses methods are extremely noise sensitive[9,10]. Therefore, observed time series must firstly be cleaned by a noise reduction scheme. The procedure for this and how to find initial conditions for the system equations are described in Chapter 3. In Chapter 4 we describe the application of the extended Kalman filter algorithm for updating parameter values and to determine the structure of the system equations. Chapter 5 is describing the prediction results and the identified governing system equations for sunspot time series. We close with a summary and discussion of how results may be interpreted.

2. Methodology and Basic Model Structure

One interesting property of low-dimensional dissipated systems is that their dimension may provide the number of equations needed to describe the system. These systems are completely deterministic, however, extremely sensitive to initial conditions and thus they are called chaotic. Although the chaotic system is unpredictable in the long term, the system's settling on a fractal trajectory (strange attractor) may be used for short-term predictions.

The dimension d of the strange attractor indicates how many variables that are necessary to describe the evolution in time. For example, $d=2.5$, indicates that the time series can be described by a system equation containing three independent variables. The structure of the equation system, however, is unknown. Moreover, we do not know whether the observed component is a single independent variable of the system or an element composed by several variables. Generally, it is very difficult and indeed impossible to find the exact original system equations. Therefore, an equivalent system equation (reconstructed system equation) which can generate a time series similar to the observed one is what at best can be expected[11,12,13,14].

The state variable with dimension $d = n$ for an unknown system may be expressed as a vector system $X = [x_1, x_2, x_3, \cdots, x_n]^T$. The general form of the system equations may be expressed as

$$\dot{X} = F(X) \tag{1}$$

or

$$\begin{cases} \dot{x}_1 = f_1(x_1, x_2, x_3, \cdots, x_n) \\ \dot{x}_2 = f_2(x_1, x_2, x_3, \cdots, x_n) \\ \cdots \cdots \cdots \cdots \cdots \cdots \\ \dot{x}_n = f_n(x_1, x_2, x_3, \cdots, x_n) \end{cases} \tag{2}$$

The necessary information needed to determine the system equations includes three parts, namely: a) the number of independent variables (the dimension), b) the structure of the equations, and c) the parameter values. To determine the number of necessary independent variables, we can use the dimension d of the strange attractor for the observed time series. A relevant structure of system equations means proper functions $f_1, f_2, f_3, \cdots, f_n$, so that the number of parameters for the system equations is minimized. The work of Gouesbet[12,13] employs the ratio of polynomials as a general form of a three-dimensional equation system. In this paper, we choose a Taylor series expansion to define a general form for the system equations and then we employ extended Kalman filter to identify parameter values.

If a low-dimensional equation system can be assumed (e.g., $d < 3$), it would be sufficient to expand the function $F(X)$ in Eq. (1) into a Taylor series up to second-order terms according to:

$$\begin{cases} \dot{x} = a_{11} + a_{12}x + a_{13}y + a_{14}z + a_{15}xy + a_{16}xz + a_{17}yz + a_{18}x^2 + a_{19}y^2 + a_{110}z^2 \\ \dot{y} = a_{21} + a_{22}x + a_{23}y + a_{24}z + a_{25}xy + a_{26}xz + a_{27}yz + a_{28}x^2 + a_{29}y^2 + a_{210}z^2 \\ \dot{z} = a_{31} + a_{32}x + a_{33}y + a_{34}z + a_{35}xy + a_{36}xz + a_{37}yz + a_{38}x^2 + a_{39}y^2 + a_{310}z^2 \end{cases} \tag{3}$$

Equation (3) contains 30 parameters. It is obviously difficult to identify these many parameters at the same time by use of only one observed time series. Consequently, some simplifications are necessary. In general, efficient estimation of the parameter values may not only decrease the number of unknown parameters, but may also simplify the structure of the system. For this, we apply an assumed structure of the system as initial conditions. All 30 parameters in Eq. (3) are then updated by the extended Kalman filter. As initial conditions for the system equations of sunspot time series we have chosen a modified form of the Rössler equations. The reasons and advantages for this are further elaborated on below.

3. Properties of the Attractor

Figure 1 shows monthly mean Wolf sunspot numbers from Chernosky and Hagan[16] and consecutive volumes of J. Geophys. Res. The figure shows raw and noise-reduced sunspot numbers using the noise reduction algorithm of Schreiber[7,17]. The noise reduction algorithm was especially developed for dimension estimations. Known methods for correlation dimension estimations are extremely noise sensitive, and it is therefore necessary to work with cleaned data.

The general idea of the noise reduction method is to replace each coordinate in the time series by an average value over a suitable neighborhood in the phase space. The main effect

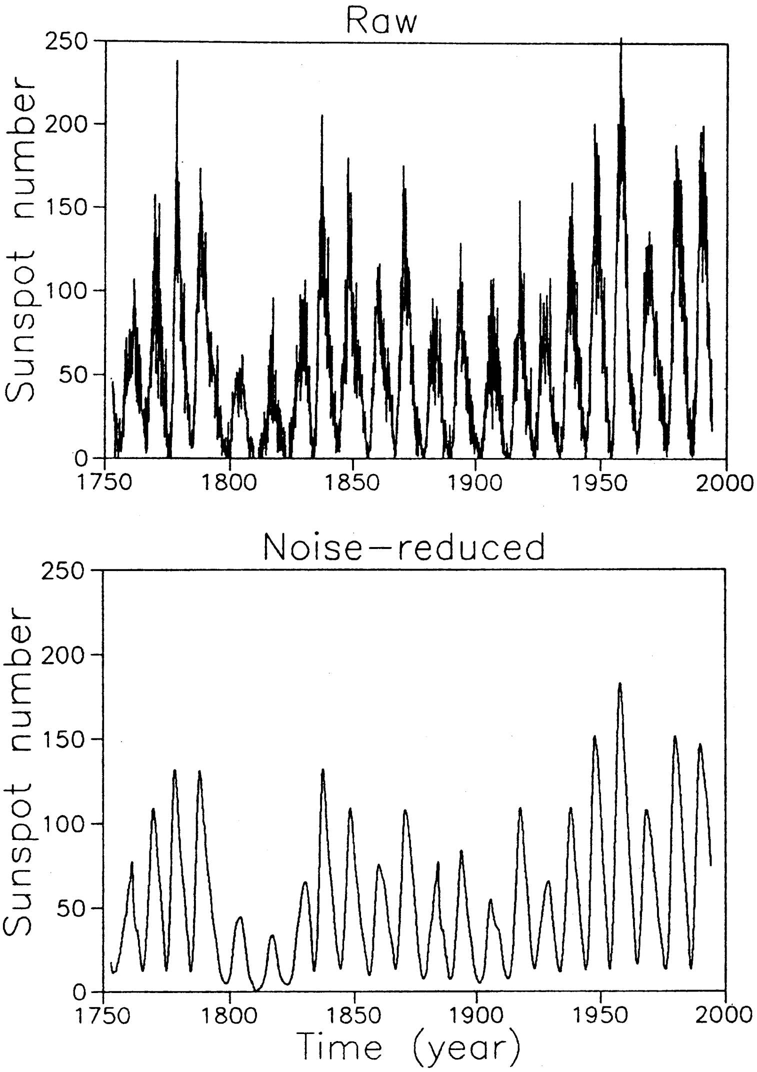

Fig. 1 Raw and noise-reduced monthly sunspot time series from Jan. 1753 to May 1994.

of the noise reduction is to remove high frequencies from the time series and leave low frequencies more or less unaffected. Consequently, general properties of the time series are kept while smail-scale variations are evened out.

Correlation dimensions were calculated according to the algorithm given by Grassberger[18]. Berndtsson *et al.*[7] showed that noise-reduced sunspots display saturation at a correlation dimension $d<2$. Mundt *et al.*[4] found $d \approx 2.3$ for the same sunspot data. However, they used a different noise reduction scheme. Consequently, in this study we will assume that $d<3$.

Due to the complex nonlinear behavior of the sunspot time series, it is difficult to assess the structure of the governing system equations. However, by studying the properties of the attractor, the general behavior of the system's time evolution can be evaluated. **Fig. 2** shows the phase space portrait of noise-reduced sunspot time series and the strange attractor for a time lag $\tau = 10$ months. Modeling and prediction of the system depends on the ability to describe this attractor.

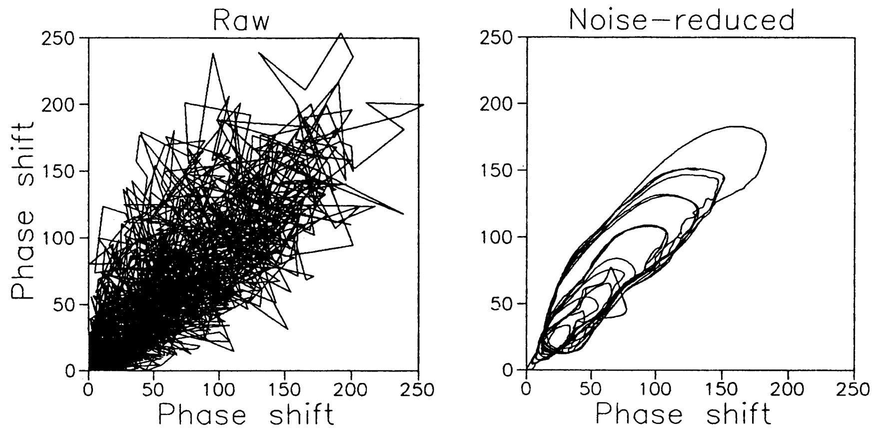

Fig. 2 Strange attractor of raw and noise-reduced sunspot time series (lag time $\tau=10$ months; monthly data during 241 years).

A remarkable impression from **Fig. 2** is that the different sunspot cycles appear to follow about five preferential development or attraction lines. Once such an attraction line, which the present cycle is following, is identified, it is possible to make accurate short-term predictions of future states in the cycle.

As mentioned before, it is possible to describe a chaotic behavior by purely deterministic and nonlinear differential equations. Since, however, it is extremely difficult to assess the structure of an unknown nonlinear equation system[11,12,13,14], further simplifications are necessary. Consequently, we suggest a method that significantly reduces the computational difficulties in estimating the structure of the unknown nonlinear system. The method involves the choice of an equation system with similar time evolution as the observed time series as initial condition for the system parameter identification scheme. By drawing the phase space portrait for time series $z(t)$ generated by the Rössler equations it is found that the attractor pattern of the noise-reduced sunspot time series and the one for $z(t)$ display obvious similarities. Therefore, we assume that a modified form of the Rössler equations can be used as an initial state for the system identification procedure according to below.

The Rössler equations[11] are a set of differential equations according to:

$$\begin{cases} \dot{x} = -y-z \\ \dot{y} = x+ay \\ \dot{z} = b+z(x-c) \end{cases} \tag{4}$$

where a, b, and c are constants. For different values of the constants, the system has different behavior and the pattern of the attractor changes as well. To use the Rössler equations as an initial state for the system to be identified, a linear transformation is performed. The transformation $z^* = \gamma z$ is applied to adjust the amplitude of $z(t)$ of the Rössler equations to the noise-reduced sunspot time series, and $t^* = Tt$ to temporally synchronize the two series. As a result, the modified Rössler equations can be expressed as:

$$\begin{cases} \dfrac{dx}{dt^*} = \dfrac{1}{T}\left(-y-\dfrac{1}{\gamma}z^*\right) \\ \dfrac{dy}{dt^*} = \dfrac{1}{T}(x+ay) \\ \dfrac{dz^*}{dt^*} = \dfrac{1}{T}[\gamma b+z^*(x-c)] \end{cases} \tag{5}$$

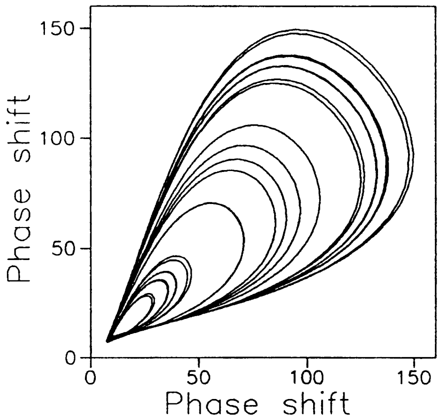

Fig. 3 Generated strange attractor from the linear transformed Rössler equations (parameters of the Rössler equations are $a=0.398$, $b=2.0$, $c=4.0$, $\gamma=26.0$, $T=21.12$; $\Delta t=1$; initial values are $(x_0, y_0, z_0)=(1.0, 1.0, 26.0)$; lag time $\tau=10$ time steps. The number of the data is same as that of sunspot time series).

where the linear transformation parameters γ and T are determined from the sunspot time series. In this case, $\gamma = 26.0$ and $T = 21.12$ months. **Fig. 3** shows the phase space portrait for the time series $z^*(t)$ generated by the Eq. (5).

4. Application of Extended Kalman Filter

The extended Kalman filter[19,20] is used to identify parameters and update predictions based on observed time series. Even though the time series possesses chaotic characteristics the extended Kalman filter is effective to identify parameter values of the system[14,15]. In the extended Kalman filter, Eq. (3) is used as system equation for the sunspot prediction. The system vector X includes three state variables and 30 parameters according to:

$$X = [x_1, x_2, x_3, \cdots, x_{33}]^T = [x, y, z, a_{11}, a_{12}, a_{13}, \cdots, a_{310}]^T \tag{6}$$

As a result, the system equation becomes:

$$\begin{cases} \dot{x}_1 = f_1(X) = x_4 + x_5x_1 + x_6x_2 + x_7x_3 + x_8x_1x_2 + x_9x_1x_3 + x_{10}x_2x_3 + x_{11}x_1^2 + x_{12}x_2^2 + x_{13}x_3^2 \\ \dot{x}_2 = f_2(X) = x_{14} + x_{15}x_1 + x_{16}x_2 + x_{17}x_3 + x_{18}x_1x_2 + x_{19}x_1x_3 + x_{20}x_2x_3 + x_{21}x_1^2 + x_{22}x_2^2 + x_{23}x_3^2 \\ \dot{x}_3 = f_3(X) = x_{24} + x_{25}x_1 + x_{26}x_2 + x_{27}x_3 + x_{28}x_1x_2 + x_{29}x_1x_3 + x_{30}x_2x_3 + x_{31}x_1^2 + x_{32}x_2^2 + x_{33}x_3^2 \\ \dot{x}_i = f_i(X) = 0, \qquad 4 \le i \le 33. \end{cases} \tag{7}$$

In the extended Kalman filter the nonlinear function $f(X)$ is expanded into a Taylor series at a point X^*. Usually X^* takes the value $\hat{X}(k|k-1)$. The $\hat{X}(k|k-1)$ is the estimation of the state vector X at time step k based on the observation at time step $k-1$. Therefore, at the first calculation step the initial value $X(0)$ should be given. According to above, the initial values of parameters $a_{11}, a_{12}, \cdots, a_{310}$ are the coefficients of Eq. (5), i.e., $x_6 = -1/21.12$, $x_7 = -1/26.0/21.12$, $x_{15} = 1/21.12$, $x_{16} = 0.398/21.12$, $x_{24} = 2.0 \times 26.0/21.12$, $x_{27} = -4.0/21.12$,

$x_{29} = 1.0/21.12$, and others are equal to zero.

The initial values of the variables $x_1(= x)$ and $x_2(= y)$ were set to 1.0 and 1.0 (same as that for generating the Rössler time series). The initial value of z was set to the same value as the initial sunspot value, $x_3(= z) = 31$.

The observation equation of the extended Kalman filter is a vector function expressed as:

$$Y = g(X) \tag{8}$$

In this case, the observation equation then becomes:

$$y_1 = g_1(X) = x_3 \tag{9}$$

Consequently, the variable x_3 is noise-reduced monthly sunspot time series.

5. Prediction Results

Given a preliminary structure of system equations and initial values for parameters, the

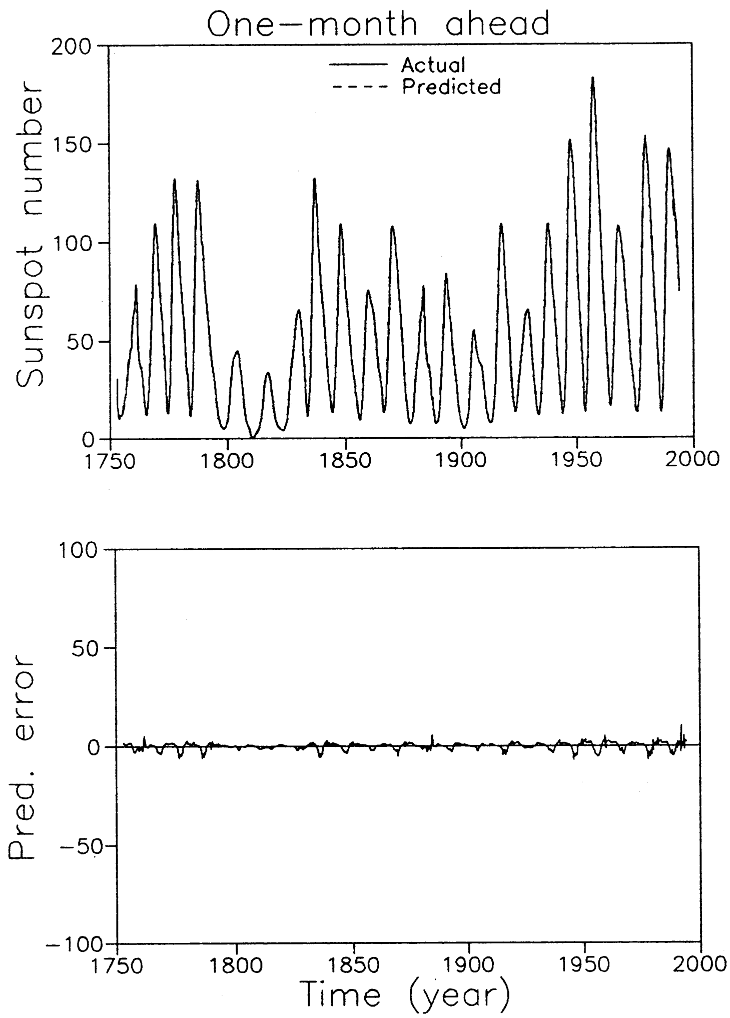

Fig. 4a One-month ahead prediction of noise-reduced sunspot time series using Equations (5) to (9).

extended Kalman filter is updating the system equations at every time step with observations and identifying parameters based on the observed time series. Correspondingly, any step ahead prediction can be made by the updated system equations. One-month, three-month, and six-month ahead predictions are shown in **Figs. 4a–4c**. From the figures it is clear that prediction results are acceptable. The average relative error of the one-month ahead predictions is 3%. Corresponding values for three-month and six-month ahead predictions are 7% and 15%, respectively.

Figure 5 shows the correlation coefficient between observed and predicted sunspot time series versus the forecast time. It is seen that the correlation coefficients for predictions up to eight months ahead remain above 0.9. After that the prediction accuracy reduces remarkably. This kind of decrease in the prediction accuracy is known as a typical feature of chaotic systems as pointed out by Sugihara and May[21].

The identification process for the two system variables x and y is shown in **Fig. 6** (updated for the same observation period). Though their physical interpretation is concealed, they keep an appearance common for chaotic time series. Equation (10) is the finally

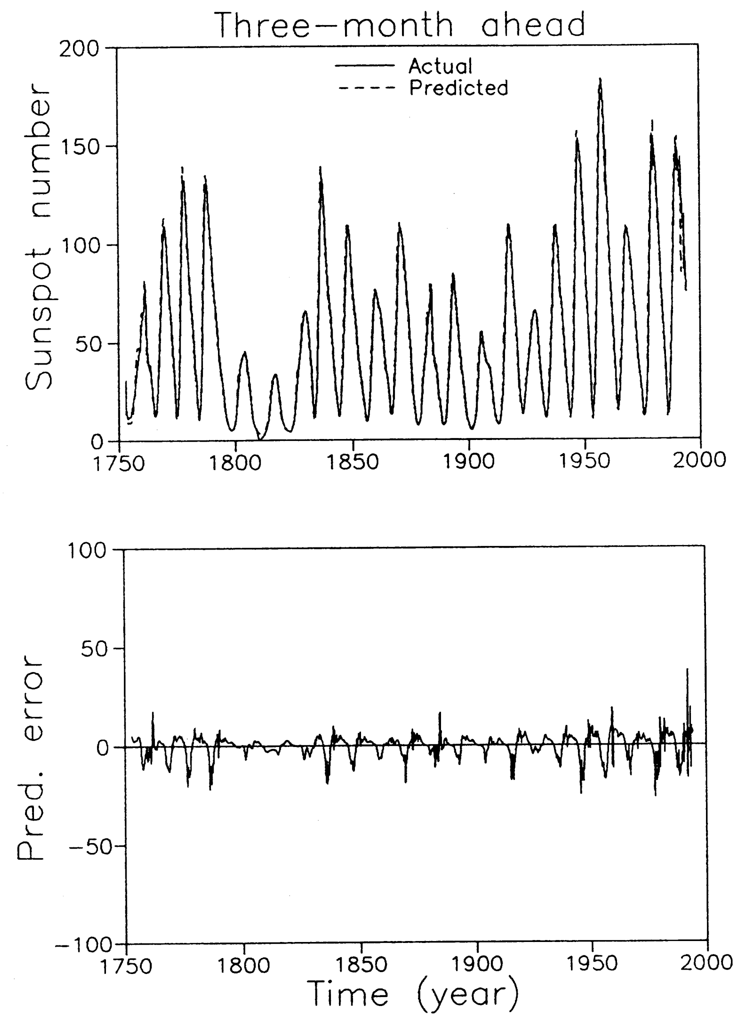

Fig. 4b Same as **Fig. 4a** but for three-month ahead prediction.

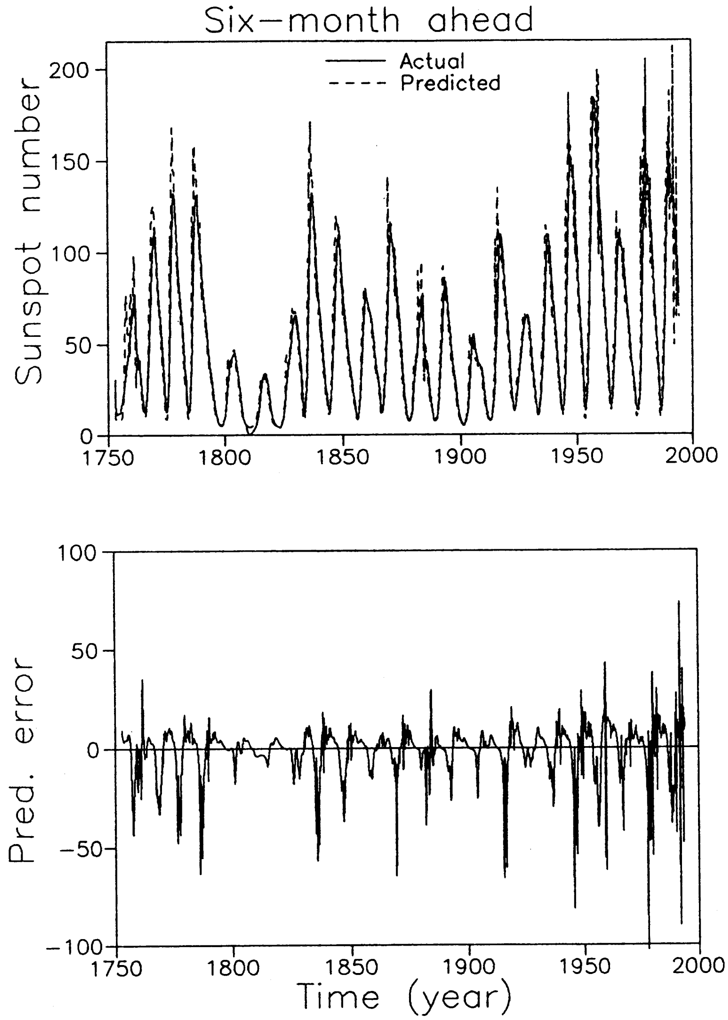

Fig. 4c Same as **Fig. 4a** but for sixth-month ahead prediction.

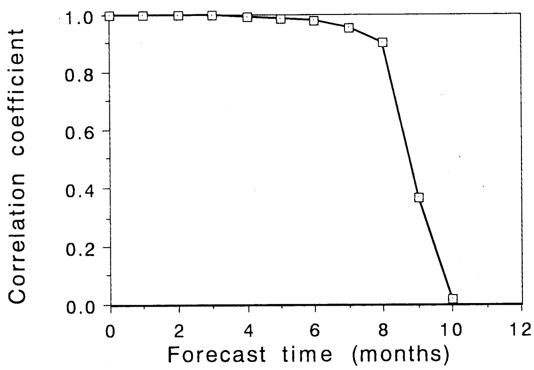

Fig. 5 Correlation coefficients between observed and predicted sunspot time series versus the forecast time.

identified system equation for the sunspot prediction at the end of the observation period:

$$\begin{cases} \frac{dx}{dt^*} = \frac{1}{T}(-0.00013-0.00041x-\underline{1.3y}-\underline{0.047z^*}+0.00052xy-0.038xz^* \\ \quad -0.0064yz^*+0.0010x^2+0.0029y^2-0.0019(z^*)^2) \\ \frac{dy}{dt^*} = \frac{1}{T}(-0.0029+\underline{0.97x}+\underline{0.33y}-0.021z^*-0.00055xy+0.020xz^* \\ \quad -0.044yz^*-0.0014x^2+0.0029y^2-0.0013(z^*)^2) \\ \frac{dz^*}{dt^*} = \frac{1}{T}(\underline{11.0}+0.00059x-0.00048y-\underline{1.5z^*}-0.00042xy+\underline{1.6xz^*} \\ \quad -0.016yz^*-0.000089x^2+0.0012y^2+0.059(z^*)^2) \end{cases} \tag{10}$$

When comparing Eqs. (10) and (5), it is seen that underlined parameters in Eq. (10) have similar magnitude as corresponding parameters of Eq. (5). This suggests that these terms may be of dominant importance to keep similar overall characteristics for the strange attractor of the sunspots as those for the Rössler equations. However, even so, the structure of Eq. (10) is not equal that of Eq. (5). Also, when comparing prediction results by using Equs.

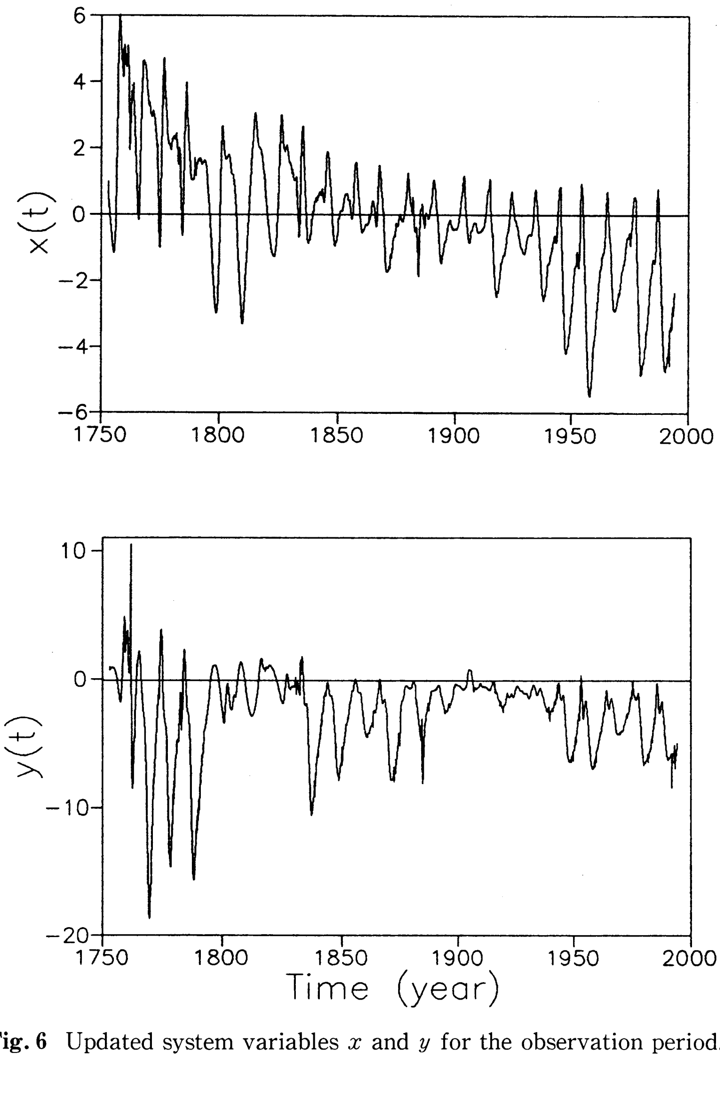

Fig. 6 Updated system variables x and y for the observation period.

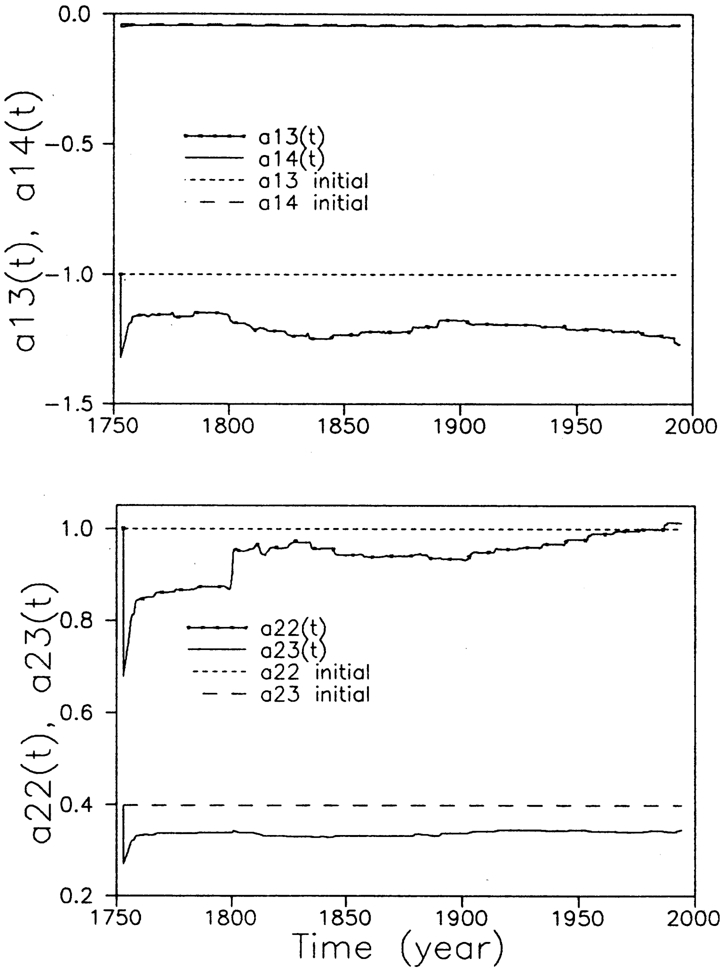

Fig. 7a Comparison between updated parameters of Equation (5) and underlined parameter values in Equation (10) (a_{13}, a_{14}, a_{22} and a_{23}).

(5) and (10), respectively, it is seen that prediction results improve substantially when using Eq. (10). **Figure 7** shows a comparison between updated parameters of Eq. (5) and underlined parameter values in Eq. (10). It is seen that several of these change significantly during the identification period. This gives support to the relevance of Eq. (10).

6. Conclusion

The sunspot number is a quantitative coefficient of the activity of the sun. Therefore, it is important make predictions of its future behavior which may affect terrestrial conditions. We have shown that improved prediction can be made using chaotic system equations for the sunspot time series behavior. For this, we introduced a procedure to reconstruct unknown nonlinear system equations from an observed chaotic time series. The procedure for this was ; 1) based on the behavior of the observed time series and strange attractor, find reference system equations that show similar basic features as the time series (e.g., the appearance of strange attractor, the amplitude, the pseudo period, and so on), 2) assume a general system equation by applying Taylor series expansion, 3) use the reference system

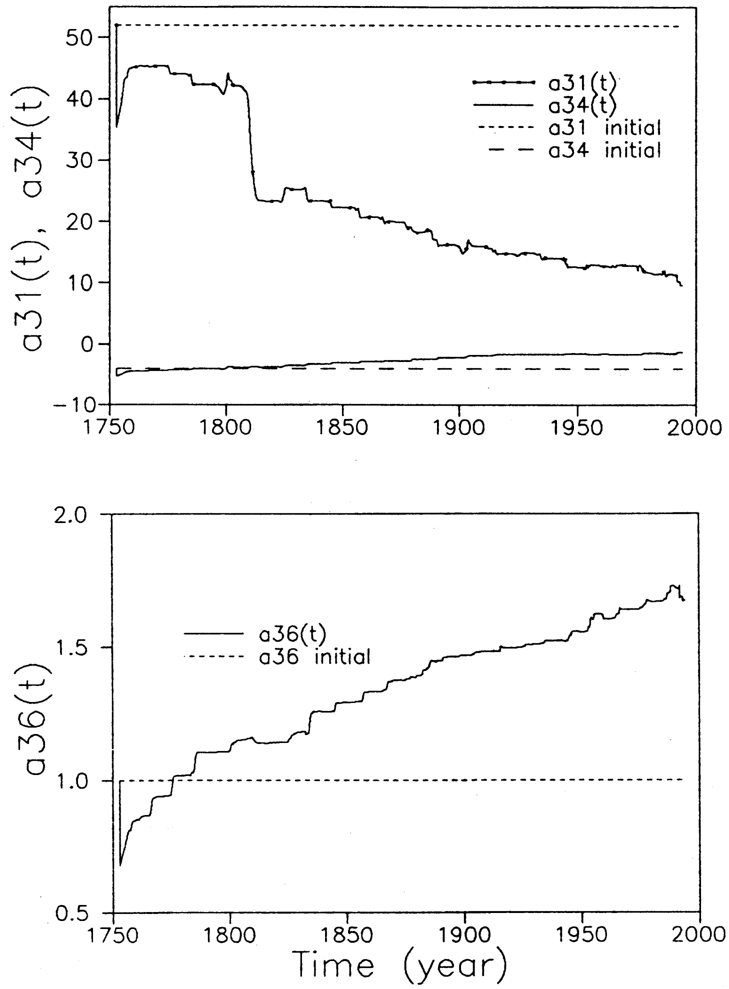

Fig. 7b Same as **Fig. 7a** but for the parameters a_{31}, a_{34}, and a_{36}.

equation as initial state for the general system equation and use it in an updating procedure. Predictions of future states of the observed variable can then be made by means of the updated and identified system equation. The outcome of the procedure shows promising results and prediction of noise-reduced sunspot time series indicates an effective approach.

The extended Kalman filter was shown to be an efficient tool to identify parameter values and to make updated predictions of the chaotic system. Owing to the sensitivity of the chaotic system to initial values of variables and parameters, the updating capability of the extended Kalman filter should be stressed.

Prediction results for noise-reduced sunspot time series were encouraging. Continued studies will be made to increase the lead time and to apply the present method to climatological and hydrological time series.

Acknowledgments : The International Exchange Foundation at the Engineering Faculty of Kyushu University supported this research by giving S. Xu the possibility to work at Kyushu University. The Japanese-German Center Berlin Special Exchange Program, the Swedish Natural Science Research Council, and the OK Environmental Foundation gave support to R. Berndtsson. This support is gratefully acknowledged.

Reference

1) Bray, R. J., and R. E. Loughhead, 1964. *Sunspots*, chapman and Hall, London.
2) Brown, G. M., 1988. Solar Cycle 22 to be one of the largest on record? *Nature*, 333, 121-122.
3) Butcher, E. C., 1990. The prediction of the magnitude of sunspot maxima for Cycle 22 using abnormal quiet days in Sq (H), *Geophys. Res. Lett.*, 17, 117-118.
4) Mundt, M. D., W. B. Maguire II, and R. R. P. Chase, 1991. Chaos in the sunspot cycle: analysis and prediction, *J. Geophys. Res.*, 96, 1705-1716.
5) Kurths, J., and H. Herzel, 1987. An attractor in a solar time series, *Physica*, 25D, 165-172.
6) Weiss, N. O., 1988. Is the solar cycle an example of deterministic chaos?, in *Secular Solar and Geomagnetic Variations in the Last 10,000 Years*, edited by F. R. Stephenson and A. W. Wolfendale, pp. 69-78, Kluwer, Boston.
7) Berndtsson, R., K. Jinno, A. Kawamura, S. Xu, 1994. Dynamical systems theory applied to long-term temperature and precipitation time series, In: J. Menon (Ed.), *Trends in Hydrology*, Counc. Sci. Res. Integr., Trivandrum, India, (in press).
8) Farmer, J. D., and J. J. Sidorowich, 1987. Predicting chaotic time series, *Phys. Rev. Lett.*, 59, 845-848.
9) Grassberger, P., T. Schreiber, and C. Schaffrath, 1991. Nonlinear time sequence analysis, *Int. J. Bifurc. Chaos*, 1, 521-547.
10) Grassberger, P., R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber, 1993. On noise reduction methods for chaotic data, *Chaos*, 3, 127-143.
11) Rössler, O. E., 1976. Different types of chaos in two simple differential equations, *Z. Naturforsch.*, 31, 1664-1670.
12) Gouesbet, G., 1991. Reconstruction of the vector fields of continuous dynamical system from numerical scalar time series, *Phys. Rev. A*, 43, 5321-5331.
13) Gouesbet, G., 1991. Automatic reconstruction of dynamical system equations from numerical scalar time series, *Eighth Symp. on Turbulent Shear Flows, Techn. Univ. of Munich*, Sep. 9-11.
14) Xu, S., K. Jinno, A. Kawamura, R. Berndtsson, and J. Olsson, 1993. Reconstructing systems from chaotic numerical time series, *Proc. of Hydraul. Eng., Japan Soc. of Civ. Eng.*, 37, 853-856.
15) Xu, S., K. Jinno, A. Kawamura, M. Matsumoto, 1993. Reconstructivity of system equations from a chaotic numerical time series, *Proc. of Annual Conf. of Japan Soc. of Civ. Eng.*, 276-277.
16) Chernosky, E. J., and M. P. Hagan, 1958. The Zurich sunspot number and its variations for 1700-1957, *J. Geophys. Res.*, 63, 755-788.
17) Schreiber, T., 1993. An extremely simple nonlinear noise reduction method, *Phys. Rev. E. Stat. Physics*, 47, 2401-2404.
18) Grassberger, P., 1990. An optimized box-assisted algorithm for fractal dimensions, *Phys. Lett. A*, 148, 63-68.
19) Athans, M., R. P. Wishner, A. Bertolini, 1968. Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy-measurements, *IEEE Transactions on automatic control, AC-13*, No. 5, 504-514.
20) Kawamura, A., K. Jinno, T. Ueda, and H. Yoshinaga, 1989. On the on-line prediction of the concentration distribution of one-dimensional constant coefficient stochastic

convective-dispersion equation based on Kalman filter, *Techn. Rep. of Kyushu Univ.*, 62, No. 1, (in Japanese with English abstract).
21) Sugihara, G., and R. M. May, 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature*, 344, 734-741.
22) Lorenz, E. N., 1963. Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 130-141.
23) Withbroe, G. L., 1989. Solar activity cycle : history and predictions, *J. Spacecr. Rockets*, 26, 394-402.