JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 100, NO. A8, PAGES 14,773–14,781, AUGUST 1, 1995

Prediction of sunspots using reconstructed chaotic system equations

Kenji Jinno,[1] Shiguo Xu,[2,3] Ronny Berndtsson,[4] Akira Kawamura,[1] and Minoru Matsumoto[1]

Abstract. Modeling of sunspots is important since they indicate the relative activity of the Sun which in turn 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 dimension of strange attractor, find reference system equations (in our case the modified Rössler equations) that show similar basic features as the time series (e.g., appearance of attractor, amplitude and pseudoperiod), (2) assume a general structure of the governing system equations by Taylor series expansion, (3) use the reference equation systems as initial state in an updating procedure (extended Kalman filtering) to estimate the structure of the governing system equations. Using this procedure, results show that predictions on an average up to eight months ahead can be made with good agreement for sunspot time series 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.

1. Introduction

Analyses and prediction of sunspots have been made by numerous researchers during the latest decades [*Bray and Loughhead*, 1964; *Brown*, 1988; *Butcher*, 1990; *Mundt et al.*, 1991]. The sunspot number is a quantitative coefficient of sun activity and therefore important for, for example, global weather, satellite trajectories, geomagnetic variations, etc. Recently, studies that consider the chaotic behavior of the Sun's behavior have indicated that better predictions can be made using developments within dynamical systems theory [*Kurths and Herzel*, 1987; *Weiss*, 1988; *Mundt et al.*, 1991].

Many studies during recent years have indicated chaotic characteristics for sunspot time series and solar activity in general [*Ruzmaikin*, 1981; *Zeldovich and Ruzmaikin*, 1983; *Gilman*, 1986; *Kurths and Herzel*, 1987; *Weiss*, 1988; *Feynman and Gabriel*, 1990; *Mundt et al.*, 1991; *Berndtsson et al.*, 1994]. None of these studies may be said to constitute an absolute evidence of a chaotic Sun since no universal method exists at present to discriminate between colored noise with power law spectra and underlying dynamical processes in data [*Theiler et al.*, 1992]. However, observed time series may at least tentatively be viewed as deterministically chaotic if prediction methods based on underlying deterministic properties are significantly better as compared to autoregressive linear models based on stochastic theory [*Farmer and Sidorowich*, 1987, 1988; *Sugihara and May*, 1990]. As noted by *Mundt et al.* [1991], one reason why models based on periodic behavior fail to predict sunspot time series accurately may be the nonlinear behavior.

Current sunspot forecasting methods can be grouped in three broad categories [*Withbroe*, 1989; *Mundt et al.*, 1991]: (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 [*Farmer and Sidorowich*, 1987]. Models based on chaotic premises do not contain any explicit random components. These models are based on purely deterministic but nonlinear equations, and pseudoperiodical 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 section 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

[1]Department of Civil Engineering, Kyushu University, Fukuoka, Japan.
[2]Department of Civil Engineering, Dalian University of Technology, Dalian, China.
[3]Now at Department of Civil Engineering, Kyushu University, Fukuoka, Japan.
[4]Department of Water Resources Engineering, Lund University, Lund, Sweden.

Paper number 95JA01167.
0148-0227/95/95JA-01167$05.00

basic properties for these observations becomes crucial for the modeling. Several studies have shown that nonlinear analyses methods are extremely noise sensitive [e.g., *Grassberger et al.*, 1991, 1993]. 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 section 3. In section 4 we describe the application of the extended Kalman filter algorithm for updating parameter values and to determine the structure of the system equations. Sections 5 and 6 describe the prediction results and the identified governing system equations for monthly as well as annual 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 dissipative 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 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 [e.g., *Rössler*, 1976; *Gouesbet*, 1991a, b; *Xu et al.*, 1993a].

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: (1) the number of independent variables (the dimension), (2) the structure of the equations, and (3) 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* [1991a, b] 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 we then employ an extended Kalman filter to identify the parameter values.

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

$$\dot{x} = a_{11} + a_{12}x + a_{13}y + a_{14}z + a_{15}xy + a_{16}xz + a_{17}yz + a_{18}x^2 + a_{19}y^2 + a_{110}z^2 \tag{3a}$$

$$\dot{y} = a_{21} + a_{22}x + a_{23}y + a_{24}z + a_{25}xy + a_{26}xz + a_{27}yz + a_{28}x^2 + a_{29}y^2 + a_{210}z^2 \tag{3b}$$

$$\dot{z} = a_{31} + a_{32}x + a_{33}y + a_{34}z + a_{35}xy + a_{36}xz + a_{37}yz + a_{38}x^2 + a_{39}y^2 + a_{310}z^2 \tag{3c}$$

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 (3) are then updated by the extended Kalman filter. As initial conditions for the system, 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

The sunspot number data used in the paper are monthly mean Wolf sunspot numbers from *Chernosky and Hagan* [1958]. Figure 1 shows raw and noise-reduced sunspot numbers using the noise reduction algorithm of *Schreiber* [1993]. This algorithm was especially developed for dimension estimations. Known methods for dimension estimations are extremely noise sensitive, and it is therefore necessary to work with cleaned data [*Grassberger et al.*, 1991].

The general idea of the noise reduction method is to replace each coordinate in the time series $\mathbf{x}_i$ by an average value over a suitable neighborhood in the phase space. A radius η is chosen and for each coordinate in $\mathbf{x}_i$ a set Ω_i^η for all neighbors $\mathbf{x}_j$ is selected so that

$$sup\{|x_{j-k} - x_{i-k}|, \cdots, |x_{j+l} - x_{i+l}|\} \equiv \|\mathbf{x}_j - \mathbf{x}_i\|_{sup} < \eta \tag{4}$$

where k and l denote past and future coordinates. Consequently, coordinates in $\mathbf{x}_i$ are replaced by mean values in Ω_i^η:

$$x_i^{corr} = \frac{1}{|\Omega_i^\eta|} \sum_{\Omega_i^\eta} x_j \tag{5}$$

The noise reduction results in removal of high frequencies from the time series and leave low frequencies more or less unaffected as seen in the power spectrum diagram (Figure 2).

Consequently, general properties of the time series are kept while small-scale variations are evened out. This is also seen in Figure 3, which shows the phase space portrait of raw and noise-reduced sunspot time series and the strange attractor for a time lag $\tau = 10$ months. Modeling and prediction of the system depend on the ability to describe this attractor. From the figure, it is in a striking way clear that raw data cannot be used to gain information on the attractor. No pattern is discernible for raw data. Noise-reduced data, however, display a clear structure that can be used to investigate the attractor dimension.

A remarkable impression from Figure 3 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 may be possible to make accurate short-term predictions of future states in the cycle.

As a result of the noise-reduction technique, k (in this case, $k = 30$) values at the beginning of the time series and k values at the end, are not filtered. In order to still keep these data for the below analyses, a simple filter was applied for these 60 values.

Attractor dimensions were calculated according to the algorithm given by *Grassberger* [1990]. This is displayed in Figure 4 which shows the slope of log $C(r)$ for the slope of log r according to

$$\log C(r) = d|\log r| \tag{6}$$

where $C(r)$ is the correlation integral defining the density of points around a specific coordinate x_i within a distance r related as (for small r):

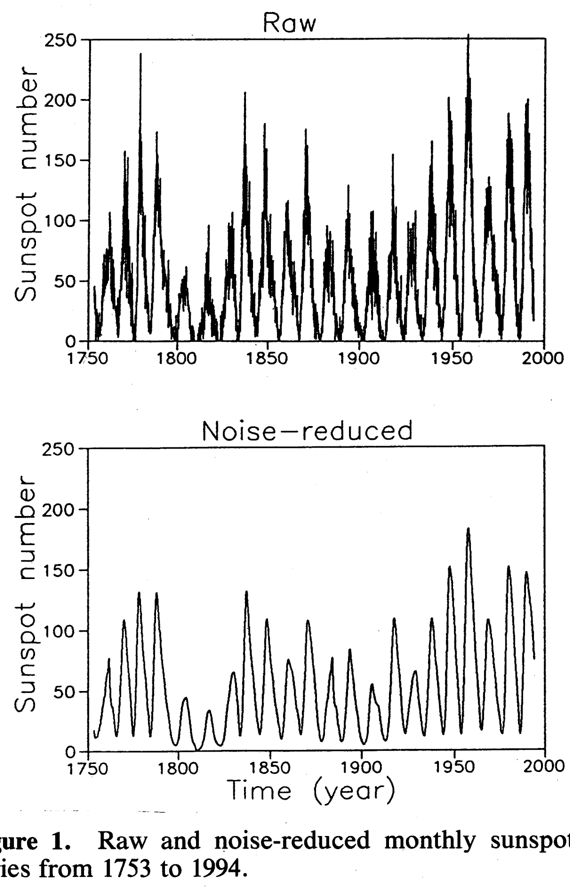

Figure 1. Raw and noise-reduced monthly sunspot time series from 1753 to 1994.

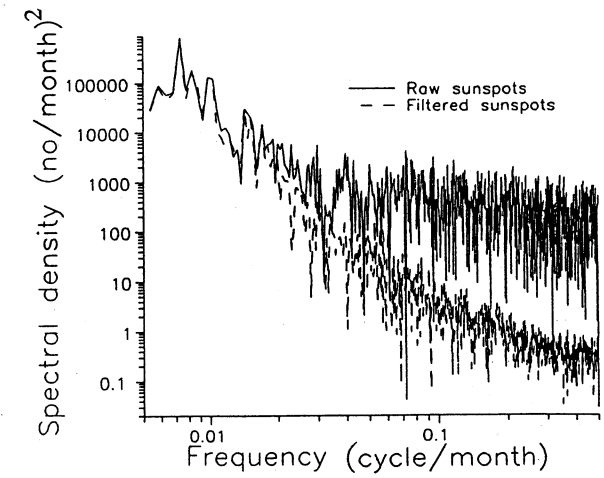

Figure 2. Power spectrum of raw and noise-reduced monthly sunspot time series.

$$C(r) \sim r^d \tag{7}$$

The correlation integral $C(r)$ is calculated as

$$C(r) = \frac{1}{N^2} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \theta\,(r - |x_i - x_j|) \tag{8}$$

where θ is the Heaviside function defined by $\theta(x) = 0$ if $x < 0$ and $\theta(x) = 1$ if $x > 0$. As seen from Figure 4, there is a clear scaling region for $-2.0 < \log r < 0$($d \log C(r)/d \log r$ does not change for changing values of log r). According to Figure 4, noise-reduced sunspots display saturation at a correlation dimension $d < 2$ [*Berndtsson et al.*, 1994].

Mundt et al. [1991] found $d \approx 2.3$ for the same sunspot data. However, they used a second-order, digital Butterworth filter with a cutoff frequency of 1/6 yr^{-1}. When comparing their power spectrum for the filtered data with the power spectrum in Figure 2, it appears that the algorithm by *Schreiber* [1993] preserves more of the high-frequency variation compared to the Butterworth filter. A reason for this is that the method of *Schreiber* [1993] is a nonlinear technique, while the Butterworth filter works as a linear procedure. Nevertheless, for the modeling work below, we will assume that $d < 3$.

As a result of 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.

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 [e.g., *Rössler*, 1976; *Gouesbet*, 1991a, b; *Xu et al.*, 1993a], 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

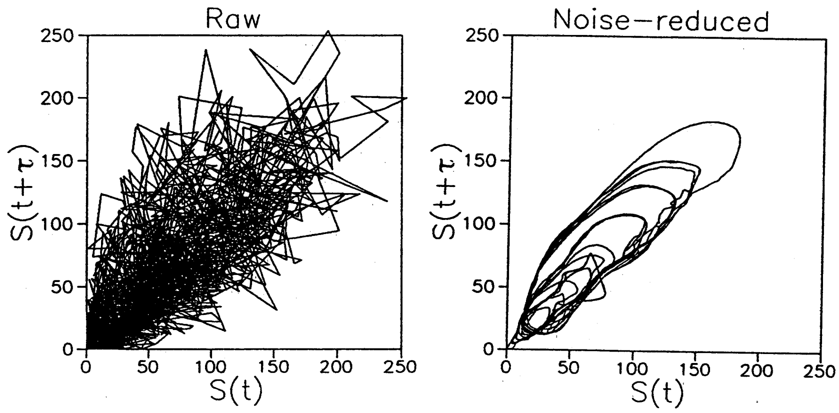

Figure 3. Strange attractor of raw and noise-reduced sunspot time series (lag time τ = 10 months; monthly data during 241 years).

series $z(t)$ generated by the Rössler equations it is found that the attractor pattern of the smoothed 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 [1976] equations are a set of differential equations according to:

$$\begin{cases} \dot{x} = -y - z \\ \dot{y} = x + ay \\ \dot{z} = b + z(x - c) \end{cases} \tag{9}$$

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

$$\frac{dx}{dt^*} = \frac{1}{T}\left(-y - \frac{1}{\gamma} z^*\right) \tag{10a}$$

$$\frac{dy}{dt^*} = \frac{1}{T}(x + ay) \tag{10b}$$

$$\frac{dz}{dt^*} = \frac{1}{T}[\gamma b + z^*(x - c)] \tag{10c}$$

where the linear transformation parameters γ and T are determined from the sunspot time series. In this case, γ = 26.0 and T = 21.12 months. Figure 5 shows the time series and the phase space portrait for the time series $z^*(t)$ generated by (10). As seen from Figure 3, there are obvious similarities between the two attractors. Especially appealing is the appearance of several attraction lines which the system follows, that is, the same temporal behavior as the noise-reduced sunspot data (Figure 3). Below, we exploit this similarity to make real-time predictions of the sunspot time series.

4. Application of the Extended Kalman Filter

The system and observation equations of the Kalman filter are expressed in discrete form as

$$X(k+1) = \Phi(k)X(k) + \alpha(k) + v(k) \tag{11}$$

$$Y(k+i) = H(k+i)X(k+i) + \beta(k+i) + w(k+i) \tag{12}$$

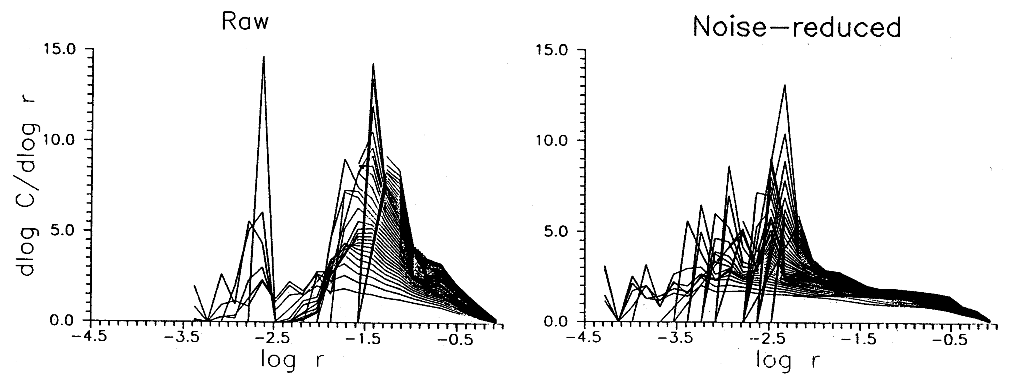

Figure 4. Slopes d log $C(r)/d$ log r versus log r for raw and noise-reduced sunspots (embedding dimensions are m = 2, 40).

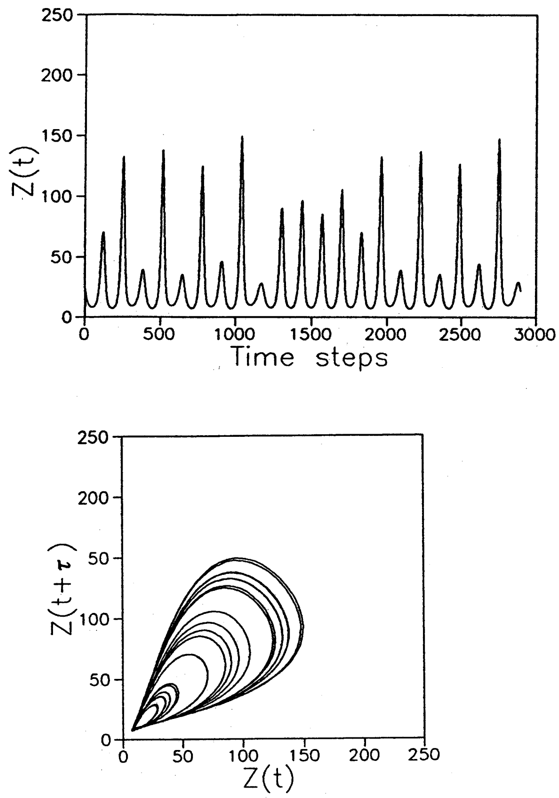

Figure 5. Generated time series and strange attractor from the linear transformed Rössler equations (parameters of the Rössler equations are $a = 0.398$, $b = 2.0$, $c = 4.0$, $\gamma = 26.0$, $T = 21.12$; $\Delta t = 1$; initial values are $(x_0, y_0, z_0) = (1.0, 1.0, 26.0)$; lag time $\tau = 10$ time steps).

where

$X(k+1)$ system vector to be estimated;
$\Phi(k)$ known state transition matrix;
$\alpha(k)$ known constant vector;
$v(k)$ white Gaussian system noise vector;
$Y(k+i)$ observation vector;
$H(k+i)$ known observation matrix;
$\beta(k+i)$ known constant vector;
$w(k+i)$ white Gaussian observation noise vector;
k calculation time step;
i time step of observations.

The estimate of the state vector at time step $k + i$, calculated using the observation obtained at time step k, is denoted $\hat{X}(k+i|k)$ and at time step $k + i$, $\hat{X}(k+i|k+i)$. If $\hat{X}(k|k)$ is known after the observation is obtained at time step k, $\hat{X}(k+i|k)$ and $\hat{X}(k+i|k+i)$ are calculated as

$$\hat{X}(k+i|k) = \Phi(k+i-1)\hat{X}(k+i-1|k) + \alpha(k+i-1) \tag{13}$$

$$\hat{X}(k+i|k+i) = \hat{X}(k+i|k) + K(k+i)\tilde{Y}(k+i|k) \tag{14}$$

where

$$K(k+i) = P(k+i|k)H^T(k+i) \cdot [H(k+i)P(k+i|k)H^T(k+i) + W(k+i)]^{-1} \tag{15}$$

$$\tilde{Y}(k+i|k) = Y(k+i) - \hat{Y}(k+i|k) \tag{16}$$

$$\hat{Y}(k+i) = H(k+i)\hat{X}(k+i|k) + \beta(k+i) \tag{17}$$

in which $W(k+i)$ is the covariance matrix of the observation noise $w(k+i)$ and T denotes transposition.

Similarly, the state estimate error covariance matrix at time step $k + i$, calculated using the observation obtained at time step k, is denoted $P(k+i|k)$ and at time step $k + i$, $P(k+i|k+i)$. They are calculated as

$$P(k+i|k) = \Phi(k+i-1)P(k+i-1|k)\Phi^T(k+i-1) + V(k+i-1) \tag{18}$$

$$P(k+i|k+i) = [I_U - K(k+i)H(k+i)]P(k+i|k) \tag{19}$$

where $V(k+i-1)$ is the covariance matrix of the system noise $v(k)$ and I_U is a unit matrix. Further details on the above procedure can be found in the works by *Athans et al.* [1968], *Bras and Rodriguez-Iturbe* [1985], or *Kawamura et al.* [1984, 1989].

The extended Kalman filter [*Athans et al.*, 1968] 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 [*Xu et al.*, 1993a, b]. In the extended Kalman filter, (3) is used as system equation for the sunspot prediction. The system vector X includes three state variables and 30 parameters according to

$$X = [x_1, x_2, x_3, \cdots, x_{33}]^T = [x, y, z, a_{11}, a_{12}, a_{13}, \cdots, a_{310}]^T \tag{20}$$

As a result, the system equation becomes

$$\dot{x}_1 = f_1(X) = x_4 + x_5x_1 + x_6x_2 + x_7x_3 + x_8x_1x_2 + x_9x_1x_3 + x_{10}x_2x_3 + x_{11}x_1^2 + x_{12}x_2^2 + x_{13}x_3^2 \tag{21a}$$

$$\dot{x}_2 = f_2(X) = x_{14} + x_{15}x_1 + x_{16}x_2 + x_{17}x_3 + x_{18}x_1x_2 + x_{19}x_1x_3 + x_{20}x_2x_3 + x_{21}x_1^2 + x_{22}x_2^2 + x_{23}x_3^2 \tag{21b}$$

$$\dot{x}_3 = f_3(X) = x_{24} + x_{25}x_1 + x_{26}x_2 + x_{27}x_3 + x_{28}x_1x_2 + x_{29}x_1x_3 + x_{30}x_2x_3 + x_{31}x_1^2 + x_{32}x_2^2 + x_{33}x_3^2 \tag{21c}$$

$$\dot{x}_i = f_i(X) = 0, \qquad 4 \le i \le 33 \tag{21d}$$

Equations (21) can be rewritten according to

$$\dot{X} = F(X) \tag{22}$$

where $F(X)$ is a nonlinear function of the vector X. If we expand $F(X)$ at X^* (close to X) into Taylor series and take first-order terms, then

$$F(X) = J(X^*)X + B(X^*) \tag{23}$$

where $J(X^*)$ is the Jacobian matrix and

$$B(X^*) = F(X^*) - J(X^*)X^* \tag{24}$$

Equation (22) then becomes

$$\dot{X} = J(X^*)X + B(X^*) \tag{25}$$

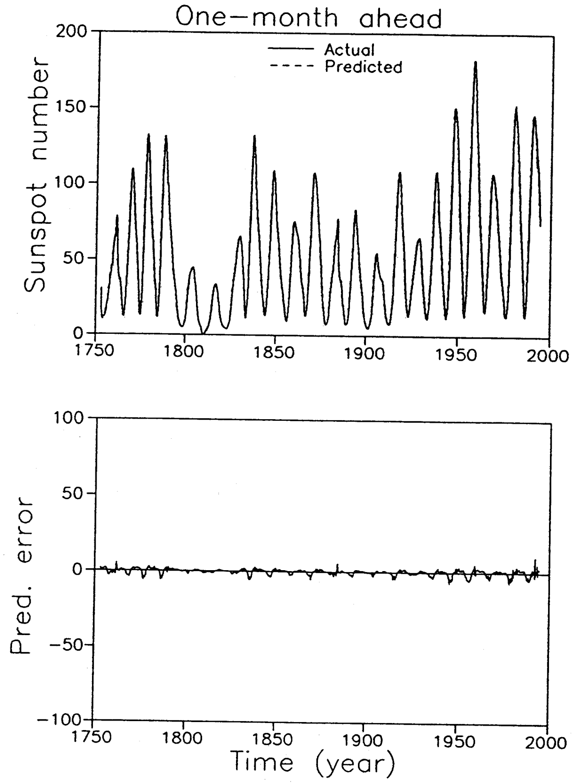

Figure 6a. One-month ahead prediction of noise-reduced sunspot time series using (10) to (14).

Equation (25) is of linear differential type. By considering system noise $u(k)$, (25) is transformed into a difference equation for numerical solving according to

$$X(k+1) = \Phi(X^*)X(k) + \Gamma(X^*)B(X^*) + u(k) \quad (26)$$

where

$$\Phi(X^*) = e^{J(X^*)\Delta t} = I + J(X^*)\Delta t + \cdots + \{J(X^*)\Delta t\}^m/m! + \cdots \quad (27)$$

$$\Gamma(X^*) = (e^{J(X^*)\Delta t} - I)J^{-1}(X^*) = \Delta t[I + J(X^*)\Delta t/2! + \cdots + \{J(X^*)\Delta t\}^{m-1}/m! + \cdots] \quad (28)$$

By comparing, in a successive way, with the state equation (11) of the extended Kalman filter, the terms $\Phi(k)$ and $\alpha(k)$ in (11) will correspond to the terms $\Phi(X^*)$ and $\Gamma(X^*)B(X^*)$ in (26). Here up to fourth-order terms are used in the calculations.

The term X^* usually 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 (10), that is, $x_6 = -1/21.12$, $x_7 = -1/26.0/21.12$, $x_{15} = 1/21.12$, $x_{16} = 0.398/21.12$, $x_{24} = 2 \times 26.0/21.12$, $x_{27} = -4/21.12$, $x_{29} = 1/21.12$, 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 in the extended Kalman filter is shown in its general form in (12). When the observed time series of sunspot numbers is regarded as observed data of the system variable x_3, then the observation matrix becomes $H = [0, 0, 1, 0, \cdots, 0]$ (33-element vector), and the constant vector $\beta = 0$. The variable x_3 is, consequently, the monthly sunspot time series.

5. Prediction Results for Monthly Data

Given the preliminary structure of system equations and initial values for parameters, the 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, 3-month, and 6-month ahead predictions are shown in Figures 6a–6c. From the figures it is clear that prediction results are acceptable for all time steps ahead. The average relative error of the one-month ahead predictions is 3%. Corresponding values for 3-month and 6-month ahead predictions are 7 and 15%, respectively.

Figure 7 shows the correlation coefficient between observed and predicted sunspot time series versus the prediction lead time. It is seen that, on an average for the entire observation period, the correlation coefficients for predictions up to eight months ahead remain above 0.9. After this lead time, 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* [1990]. Also shown in Figure 7, is correlation for two periods of 30 years each, when the prediction was very good (best case) and very bad (worst case), respectively.

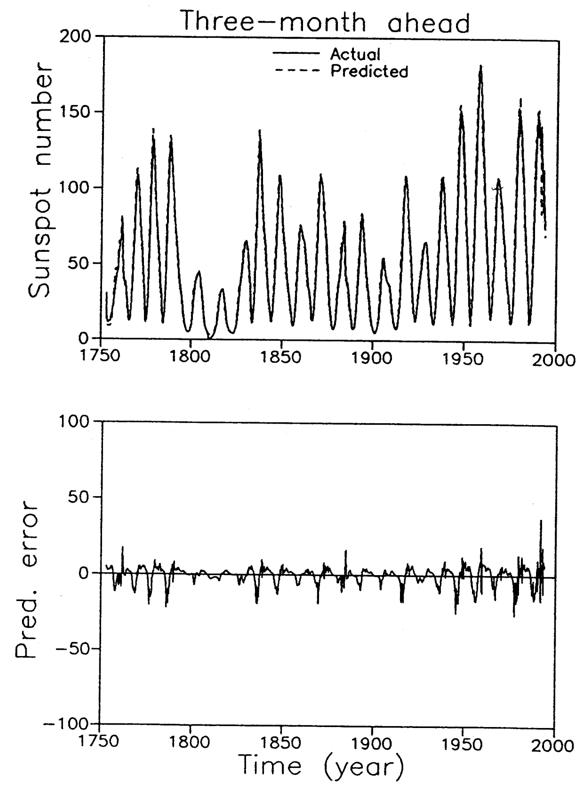

Figure 6b. Same as Figure 6a but for 3-month ahead prediction.

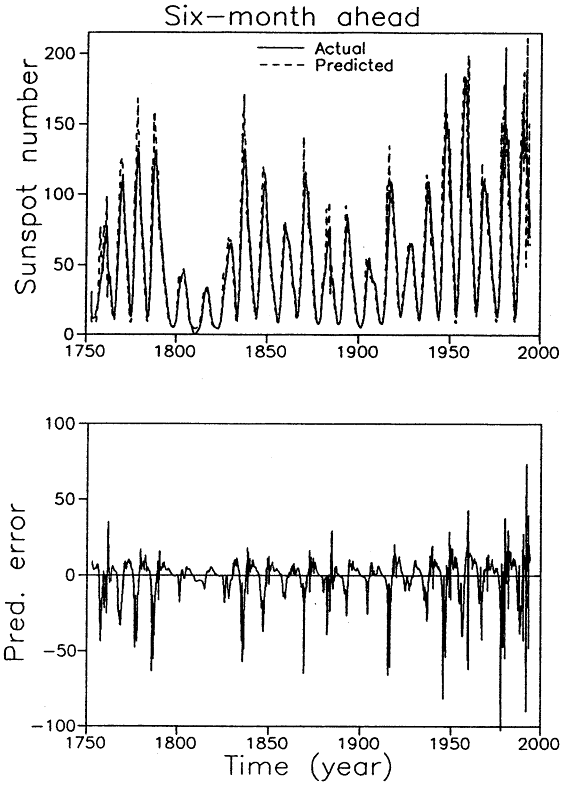

Figure 6c. Same as Figure 6a but for 6-month ahead prediction.

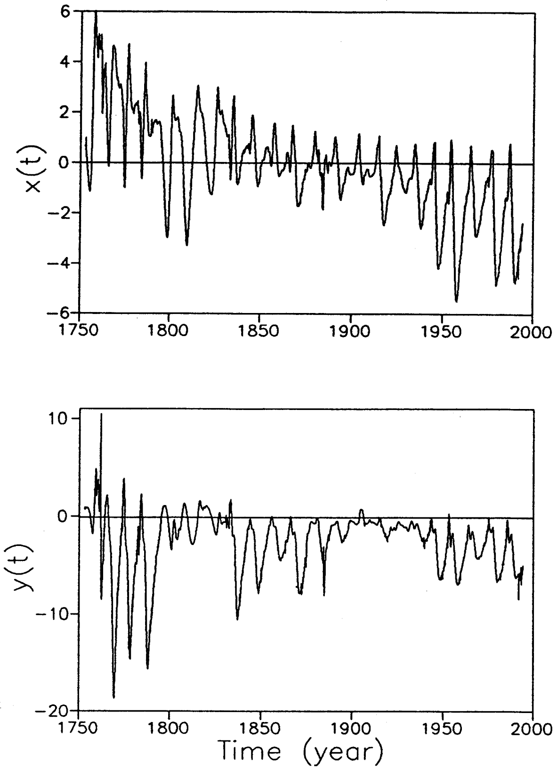

Figure 8. Updated system variables x and y for the observation period.

The best case was for a period 1790–1820 when sunspots displayed a very regular behavior and the worst case for a period toward the end of the observation period when there was an extreme variation (Figure 1). From Figure 7 it is seen that even under a worst case scenario, the model is able to predict sunspots with a lead time of 6 months. For the best case the correlation gradually drops below 0.6 after a lead time of 13 months.

The updated two system variables x and y are shown in Figure 8. Though, it is not possible to interpret them physically, they keep an appearance common for chaotic time series. Equation (29) is the finally identified equation for the sunspot prediction at the end of the observation period:

$$\frac{dx}{dt^*} = \frac{1}{T}(-0.00013 - 0.00041x - \underline{1.3y} - \underline{0.047z^*}$$

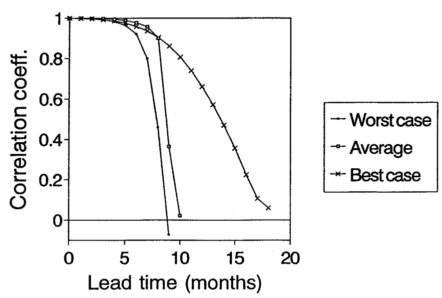

Figure 7. Correlation coefficients between observed and predicted sunspot time series versus lead time for monthly data (''average'' indicates the entire observation period, while ''worst case'' and ''best case'' correspond to the worst and the best 30-year period, respectively).

$$+ 0.00052xy - 0.038xz^* - 0.0064yz^* + 0.0010x^2$$

$$+ 0.0029y^2 - 0.0019(z^*)^2) \qquad (29a)$$

$$\frac{dy}{dt^*} = \frac{1}{T}(-0.0029 + \underline{0.97x} + \underline{0.33y} - 0.021z^*$$

$$- 0.00055xy + 0.020xz^* - 0.044yz^* - 0.0014x^2$$

$$+ 0.0029y^2 - 0.0013(z^*)^2) \qquad (29b)$$

$$\frac{dz^*}{dt^*} = \frac{1}{T}(\underline{11.0} + 0.00059x - 0.00048y - \underline{1.5z^*}$$

$$- 0.00042xy + \underline{1.6xz^*} - 0.016yz^* - 0.000089x^2$$

$$+ 0.0012y^2 + 0.059(z^*)^2) \qquad (29c)$$

When comparing (29) and (10), it is seen that underlined parameters in (29) have similar magnitude as corresponding parameters of (10). This suggests that these terms are 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 (29) is not equal that of (10). Also, when comparing prediction results by using (10) and (29), respectively, prediction results improve substantially when using (29). Figure 9 shows a comparison between parameters of (10) (denoted ''initial'' in the figure) and of underlined parameter values in (29). It is seen that several of these change significantly during the identification period. This gives support to the relevance of (29).

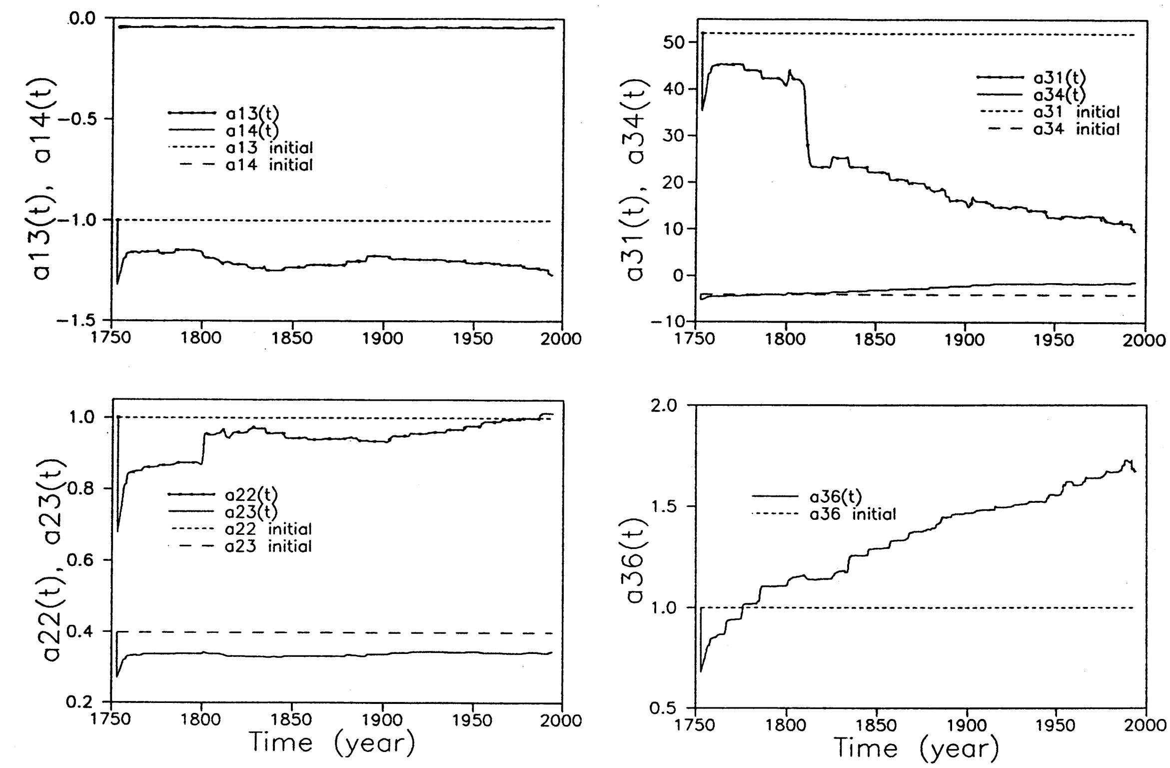

Figure 9. Comparison between parameters of (10) (denoted "initial") and of underlined parameter values in (29) for monthly data (a_{13}, a_{14}, a_{22}, a_{23}, a_{31}, a_{34}, and a_{36}).

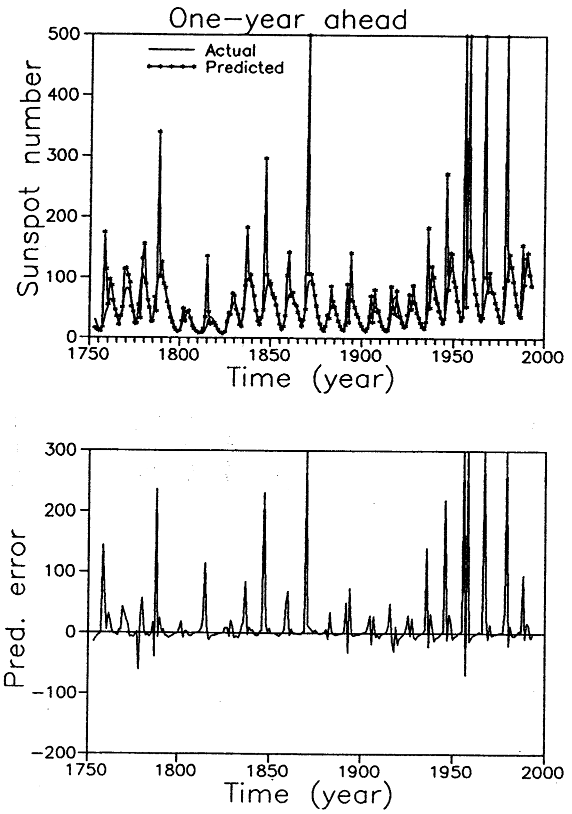

Figure 10. One-year ahead prediction of annual sunspot time series.

6. Prediction Results for Annual Data

Prediction of annual sunspot data was also made using noise-reduced annually accumulated monthly sunspots (monthly data as in Figure 1). For the annual data, the transformation parameters take the values $\gamma = 26.0$ and $T = 1.76$ years. Figure 10 shows the outcome of the calculations. As seen from Figure 10, prediction results are not satisfactory for one year lead time. Large discrepancies are found especially around the peaks of the sunspots time series.

By analyzing the correlation coefficients between observed and predicted sunspot time series vs. lead time for annual data, it is found that only for 1-year ahead time steps the correlation remains high (correlation coefficient is 0.904). This is a small improvement compared to the monthly predictions for the same lead time. As a result of the similar appearance of the monthly and annual time series, it is not possible to improve the prediction results substantially, however. The annual data contain less information compared to the monthly data, and thus less prediction steps are possible with good agreement.

7. Conclusion and Discussion

The sunspot number is a quantitative coefficient of the activity of the Sun. Therefore it is important to 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 dimension estimation of the strange attractor, find reference system equations that show similar basic features as the time series (e.g., general appearance of the attractor, the amplitude, the pseudoperiod, etc), (2) assume a general system equation by applying Taylor series expansion, (3) use the reference system 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.

Most of previously published sunspot prediction methods use what may be called local predictions. That is, based on the recent past, reasonable predictions for the future are sought. To make repeated short-term forecasts are consistent with the assumption of nonlinearity and chaotic properties. However, the method presented here differs from previous techniques in the sense that it tries to encompass also the long-term information for the prediction. This is done by keeping the structure of the governing system equations through the entire prediction period.

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 Scandinavia-Japan Sasakawa Foundation, the Swedish Natural Science Research Council, and the OK Environmental Foundation gave support to R. Berndtsson. This support is gratefully acknowledged.

The Editor thanks E. C. Butcher and S. W. Kahler for their assistance in evaluating this paper.

References

Athans, M., R. P. Wishner, and A. Bertolini, Suboptimal state estimation for continuous-time nonlinear system from discrete noisy measurement, *IEEE Trans. Automatic Control*, *AG-13*, 504–514, 1968.

Berndtsson, R., K. Jinno, A. Kawamura, J. Olsson, and S. Xu, Dynamical systems theory applied to long-term temperature and precipitation time series, in *Trends in Hydrology*, edited by J. Menon, pp. 291–297, Counc. Sci. Res. Integr., Trivandrum, India, 1994.

Bras, R. L., and I. Rodriguez-Iturbe, *Random Functions and Hydrology*, pp. 1–559, Addison-Wesley, Reading, Mass., 1985.

Bray, R. J., and R. E. Loughhead, *Sunspots*, Chapman and Hall, London, 1964.

Brown, G. M., Solar cycle 22 to be one of the largest on record?, *Nature*, *333*, 121–122, 1988.

Butcher, E. C., The prediction of the magnitude of sunspot maxima for Cycle 22 using abnormal quiet days in *Sq*(H), *Geophys. Res. Lett.*, *17*, 117–118, 1990.

Chernosky, E. J., and M. P. Hagan, The Zurich sunspot number and its variations for 1700–1957, *J. Geophys. Res.*, *63*, 755–788, 1958.

Farmer, J. D., and J. J. Sidorowich, Predicting chaotic time series, *Phys. Rev. Lett.*, *59*, 845–848, 1987.

Farmer, J. D., and J. J. Sidorowich, Exploiting chaos to predict the future and reduce noise, *Rep. LA-UR-88-901*, Los Alamos Natl. Lab., Los Alamos, N. M., 1988.

Feynman, J., and S. B. Gabriel, Period and phase of the 88-year solar cycle and the Maunder minimum: Evidence for a chaotic sun, *Sol. Phys.*, *127*, 393–403, 1990.

Gilman, P. A., The solar dynamo: Observations and theories of solar convection, global circulation, and magnetic fields, in *Physics of the Sun*, edited by P. A. Sturrock, pp. 95–160, D. Reidel, Norwell, Mass., 1986.

Gouesbet, G., Reconstruction of the vector fields of continuous dynamical system from numerical scalar time series, *Phys. Rev. A*, *43*, 5321–5331, 1991a.

Gouesbet, G., Automatic reconstruction of dynamical system equations from numerical scalar time series, paper presented at Eighth Symposium on Turbulent Shear Flows, Tech. Univ. of Munich, Munich, Sept. 9–11, 1991b.

Grassberger, P., An optimized box-assisted algorithm for fractal dimensions, *Phys. Lett. A*, *148*, 63–68, 1990.

Grassberger, P., T. Schreiber, and C. Schaffrath, Nonlinear time sequence analysis, *Int. J. Bifurc. Chaos*, *1*, 521–547, 1991.

Grassberger, P., R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber, On noise reduction methods for chaotic data, *Chaos*, *3*, 127–143, 1993.

Kawamura, A., K. Jinno, and T. Ueda, Detection of abnormality by the adaptive Kalman filter, (in Japanese with English abstract), *Proc. Jpn. Soc. Civ. Eng.*, *345/II-1*, 111–121, 1984.

Kawamura, A., K. Jinno, T. Ueda, and H. Yoshinaga, On the on-line prediction of the concentration distribution of one-dimensional constant coefficient stochastic convective-dispersion equation based on Kalman filter, (in Japanese with English abstract), *Techn. Rep. Kyushu Univ.*, *62*(1), 1989.

Kurths, J., and H. Herzel, An attractor in a solar time series, *Physica D Amsterdam*, *25*, 165–172, 1987.

Mundt, M. D., W. B. Maguire II, and R. R. P. Chase, Chaos in the sunspot cycle: Analysis and prediction, *J. Geophys. Res.*, *96*, 1705–1716, 1991.

Rössler, O. E., Different types of chaos in two simple differential equations, *Z. Naturforsch.*, *31*, 1664–1670, 1976.

Ruzmaikin, A. A., The solar cycle as a strange attractor, *Comment Astrophys.*, *9*, 85–93, 1981.

Schreiber, T., An extremely simple nonlinear noise reduction method, *Phys. Rev. E. Stat. Phys.*, *47*, 2401–2404, 1993.

Sugihara, G., and R. M. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature*, *344*, 734–741, 1990.

Theiler, J., S. Eubank, A. Longtin, B. Galdrakin, and J. D. Farmer, Testing for nonlinearity in time series: The method of surrogate data, *Physica D, Amsterdam*, *58*, 77–94, 1992.

Weiss, N. O., 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 Academic, Norwell, Mass., 1988.

Withbroe, G. L., Solar activity cycle: History and predictions, *J. Spacecr. Rockets*, *26*, 394–402, 1989.

Xu, S., K. Jinno, A. Kawamura, R. Berndtsson, and J. Olsson, Reconstructing systems from chaotic numerical time series, *Proc. Hydraul. Eng., Jpn. Soc. Civ. Eng.*, *37*, 853–856, 1993a.

Xu, S., K. Jinno, A. Kawamura, and M. Matsumoto, Reconstructivity of systems equations from a chaotic numerical time series, *Proc. Annu. Conf. Jpn. Soc. Civ. Eng.*, 276–277, 1993b.

Zeldovich, Y. B., and A. A. Ruzmaikin, Dynamo problems in astrophysics, in *Soviet Sci. Rev., Sect. E*, edited by R. A. Syunyaev, pp. 333–383, Harwood, New York, 1983.

R. Berndtsson, Department of Water Resources Engineering, Lund University, Box 118, S-221 00, Lund, Sweden.

K. Jinno, A. Kawamura, M. Matsumoto, and S. Xu, Department of Civil Engineering, Kyushu University, Hakozaki, Higashi-ku, Fukuoka 812, Japan.

(Received July 20, 1994; revised February 4, 1995; accepted March 10, 1995.)