REVIEW.

554-FDD-91/113R0UD0
CSC/TM-91/6167R0UD0

P. 43

Future Mission Studies

Forecasting Solar Flux Directly From Its Chaotic Time Series

December 1991

554-FDD-91/113R0UD0
CSC/TM-91/6167R0UD0

Future Mission Studies

Forecasting Solar Flux Directly From Its Chaotic Time Series

December 1991

FLIGHT DYNAMICS DIVISION

CODE 550

Developed Under
the Direction of

J. Cooley
GSFC/Code 554

Through

L. Roszman
Task 404
Contract NAS 5-31500
Computer Sciences Corporation

554-FDD-91/113R0UD0
CSC/TM-91/6167R0UD0

Future Mission Studies

Forecasting Solar Flux Directly From Its Chaotic Time Series

December 1991

The primary contributor to this document is

S. Ashrafi CSC

APPROVED FOR PUBLICATION BY:

J. Cooley, ATR
Flight Dynamics Analysis Branch

M. Samii, Manager
Orbit Analysis Department

Abstract

This document presents the mathematical structure of the programs written to construct a nonlinear predictive model to forecast solar flux directly from its time series without reference to any underlying solar physics. This method and the programs are written so that one could apply the same technique to forecast other chaotic time series, such as geomagnetic data, attitude and orbit data, and even financial indexes and stock market data.

Perhaps the most important application of this technique to flight dynamics is to model Goddard Trajectory Determination System (GTDS) output of residues between observed position of spacecraft and calculated positon with no drag (Drag Flag = off). This would result in a new model of drag working directly from observed data.

TABLE OF CONTENTS

Contents (continued)

Figures

Section 1. Introduction

This document presents numerical techniques for constructing nonlinear predictive models to forecast solar flux directly from its time series. As a continuation of our previous research in understanding the dynamics of solar activity (References 1 through 5), we consider the dynamic evolution of our system (solar activity) in a re-constructed phase space that captures the (strange) attractor,* and we give a procedure for constructing parameterized maps that describe the evolution of points in the phase space into the future. The predictor would necessarily depend on past data points and different iterations of the map. Thus the map is regarded as a dynamical system and not just a fit to the data. The invariants of our dynamical system, the Lyapunov exponents and the invariant density on the attractor, are used as constraints on the choice of mapping parameters. We give a detailed analysis of methods to extract the Lyapunov exponents and show how to equate them to the values for the parametric map in the constraint optimization.

* If the data points are uniformly distributed in embedding space, our data are truly stochastic. But if the data points are distributed in a bounded region (attractor), then the data are not stochastic, and order can be extracted from data.

If the evolution of our dynamical system is on such an attractor, then the d-dimensional embedding space enclosing the attractor should be sufficiently larger than d_A that all the geometric information about the attractor will be exposed in the embedding space. In the next section we present the methods of forming this phase-space reconstruction.

2.1 Takens-Packard Phase-Space Reconstruction

Using this method, we construct from the solar flux time series x(t)'s d-dimensional vectors which, when embedded in R^d, describe the full dynamical evolution of the system. Section 3.1 is devoted to the techniques used to identify the correct value of d directly from the time series.

Suppose we have the correct value of d. We consider measuring a single scalar variable x at discrete time points x(n) for $n=1,2, \ldots ,N_D$. In our case, x is the solar flux and the observations are daily.

To illustrate the technique, we have exaggerated the sampling intervals and magnified a small portion of our time series in Figures 2-2 and 2-3. Here t_1 and t_2 should be extremely close to one another and τ should be chosen so that no information is lost. The choice of τ will be discussed later. It cannot be too big or too small. For right now, suppose we have the correct value of τ.

Figure 2-2 shows how we can construct vector kets $|y_1\rangle$, $|y_2\rangle$, ... $|y_N\rangle$.

The set of $|y(n)\rangle$, of which we have $N = N_D - d$, captures the evolution of our nonlinear dynamical system as it moves through the d-dimensional phase space. Familiar phase-space coordinates are time derivatives x(n), $\dot{x}(n)$, $\ddot{x}(n)$,..., evaluated at discrete times. The times, lagged x(n), are nonlinear combinations of the local time derivatives and are fully acceptable substitutes for the usual phase-space coordinates, as Eckmann and Ruelle have emphasized (Reference 7).

$$|y_1> = \begin{bmatrix} f_{t_1} \\ f_{t_1+\tau} \\ f_{t_1+2\tau} \\ \vdots \\ f_{t_1+(d-1)\tau} \end{bmatrix} \quad |y_2> = \begin{bmatrix} f_{t_2} \\ f_{t_2+\tau} \\ f_{t_2+2\tau} \\ \vdots \\ f_{t_2+(d-1)\tau} \end{bmatrix} \quad \ldots\ldots\ldots |y_N>$$

or

$$|y(n)> = \begin{bmatrix} f_{t_n} \\ f_{t_n+\tau} \\ f_{t_n+2\tau} \\ \vdots \\ f_{(t_n+(d-1)\tau)} \end{bmatrix} \qquad (1)$$

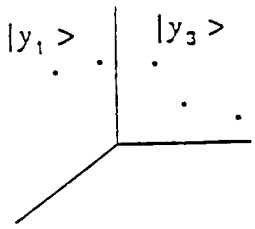


Figure 2-2. Phase-Space Taken-Packard Vectors

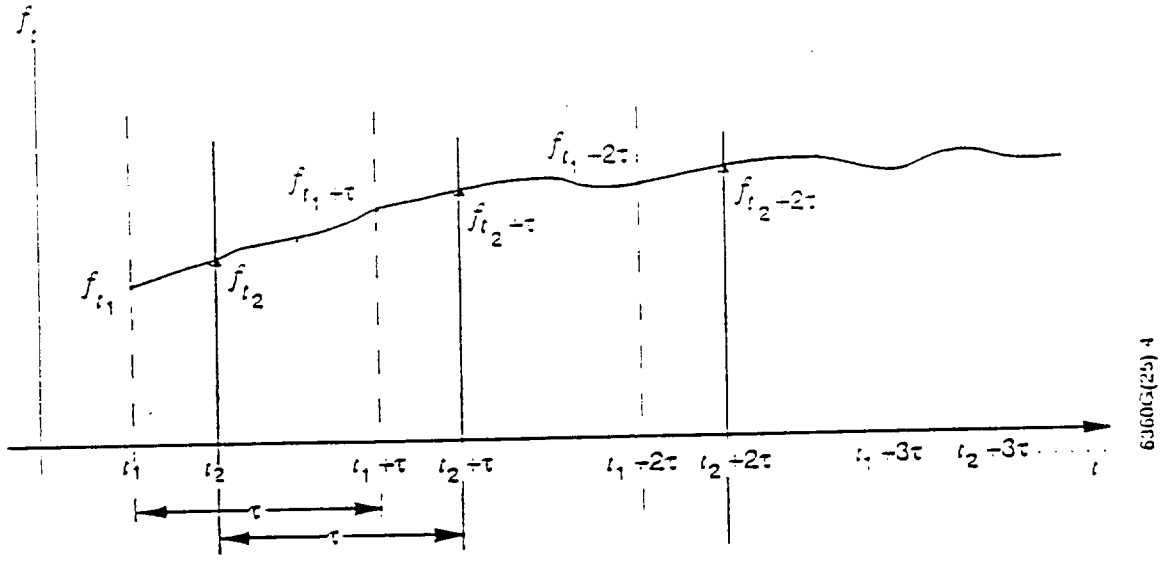


Figure 2-3. Phase-Space Reconstruction From a Time Series

Having the y(n) and the embedding space, we ask how we can use the series of y(n) to predict $|y(N+1)\rangle$, $|y(N+2)\rangle$, etc. That is, given a data set $|y(1)\rangle$, $|y(2)\rangle$, ... $|y(N)\rangle$, is there a mapping $\mathbf{F}$ from $\mathbb{R}^d$ to itself parameterized by $|a\rangle = \begin{bmatrix} a1 \\ a2 \\ . \\ . \\ ap \end{bmatrix}$ that takes us from $|y(n)\rangle$ to $|y(n+1)\rangle$?

$$|y(n+1)\rangle = \mathbf{F}(|y(n)\rangle, |a\rangle) \tag{2-1}$$

that is $|y(2)\rangle = \mathbf{F}(|y(1)\rangle, |a\rangle)$ and $|\bar{y}(3)\rangle = \mathbf{F}(|\bar{y}(2)\rangle, |a\rangle)$

For simplicity we write $\mathbf{F} = \underline{F}$ and $|y\rangle = \bar{y}$, $|a\rangle = \bar{a}$ and generate the columns in Figure 2-4.

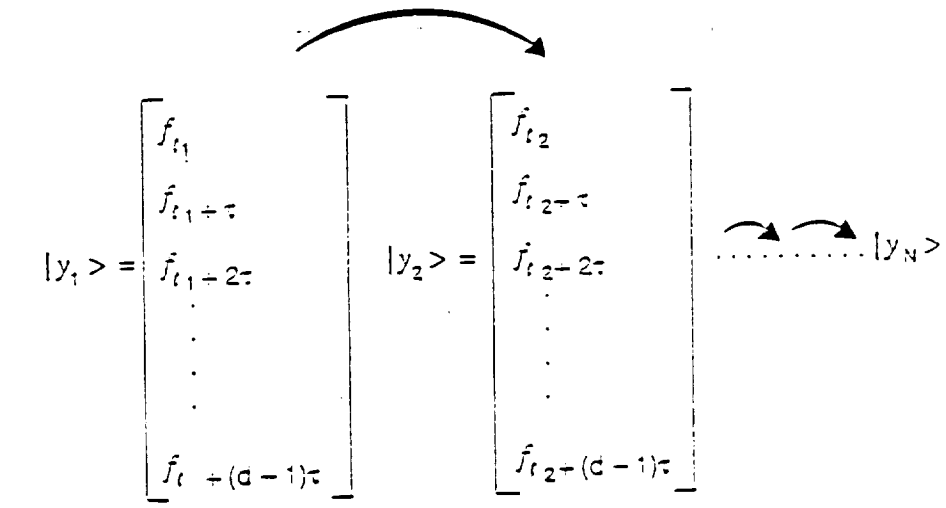
Therefore, from a time series we can form phase-space reconstruction in embedding space of dimension d, as shown in Figure 2-5.

Our function $\underline{F}(\bar{y},\bar{a})$ comes from parametrically "fitting" the right-hand column in Figure 2-4 of $\bar{y}(n+1)$ resulting from the left-hand column of $\bar{y}(n)$. Fitting the data then suggests making an estimation of $\bar{a}$ so that a cost function $\hat{C}(\bar{a})$ is minimized.

$$\hat{C}(\bar{a}) = \sum_{n=1}^{N-1} \left(\sum_{m=1}^{d} \{\bar{y}_m(n+1) - \underline{F}_m[\bar{y}(n),\bar{a}]\} \right)^2 \tag{2-2}$$

Here we are not just making a fit to data with a set of functions $\underline{F}(\bar{y},\bar{a})$. Rather, these functions evaluated along the orbit are to be related to each other in the manner of a dynamical system. That means the function $\underline{F}(\bar{y},\bar{a})$ evaluated on the data vector $\bar{y}(n)$ is required to do more than reproduce $\bar{y}(n+1)$. $\underline{F}(\bar{y},\bar{a})$ must also be a function that, when iterated, will reproduce $\bar{y}(n+2)$ after two applications to $\bar{y}(n)$, $\bar{y}(n+3)$ after three, and so on. That is, the geometrical properties of our dynamical system given by $\underline{F}(\bar{y},\bar{a})$ are used to determine the success of the fit.

The data contain invariant dynamical information that is essential for a full description of the structure of the attractor that they evolve on. We would try to impose these invariants as constraints on the fit. This emphasizes that we are creating a dynamic and not just a fit to data.



$$|y_1> = \begin{bmatrix} f_{t_1} \\ f_{t_1+\tau} \\ f_{t_1+2\tau} \\ \vdots \\ f_{t_1+(d-1)\tau} \end{bmatrix} \quad |y_2> = \begin{bmatrix} f_{t_2} \\ f_{t_2+\tau} \\ f_{t_2+2\tau} \\ \vdots \\ f_{t_2+(d-1)\tau} \end{bmatrix} \quad \cdots\cdots |y_N>$$

or

$$|y(n)> = \begin{bmatrix} f_{t_n} \\ f_{t_n+\tau} \\ f_{t_n+2\tau} \\ \vdots \\ f_{(t_n+(d-1)\tau)} \end{bmatrix}$$

Figure 2-4. Representation of a Map

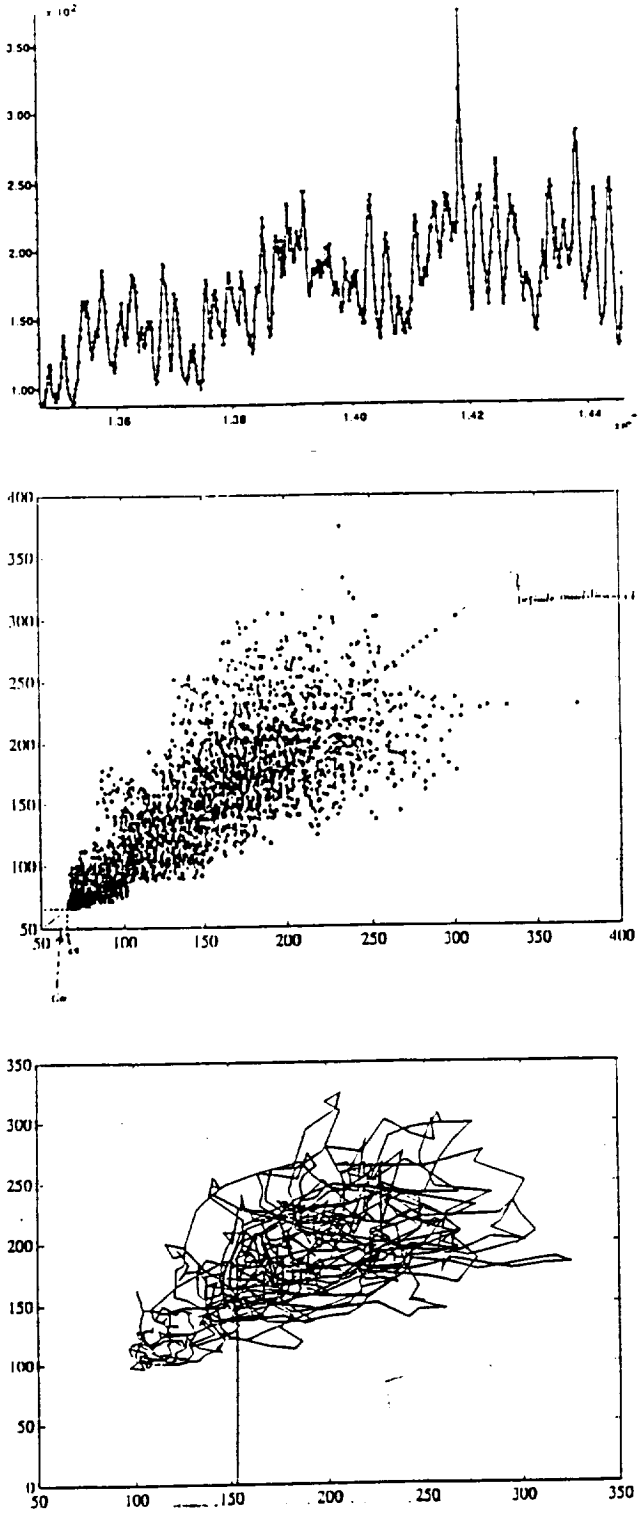


Figure 2-5. Taken-Packard Transformation

Many invariant dynamical quantities remain constant as the system evolves in time. We will concentrate on two kinds that allow us to construct a reasonable predictor. One kind of invariant, the Lyapunov exponents $\lambda_1, \lambda_2, \ldots \lambda_d$, describes the expansion or contraction of phase-space volumes under the iteration of $\underline{F}(\vec{y},\vec{a})$. The second kind of invariant is the density of points on the attractor $\rho(\vec{y})$. It captures global features of the frequency with which orbits visit various portions of the attractor. It is a different kind of invariant from Lyapunov exponents. Its integrals with smooth functions $\underline{G}(\vec{y})$ are unchanged under operation with the mapping function that underlies the dynamics $\vec{y}(n) \rightarrow \vec{y}(n+1)$.

$$\int d^d y \; \rho(\vec{y}) G(\vec{y}) = \int d^d y \rho(\vec{y}) G(F(\vec{y},\vec{a})) \tag{2-3}$$

Both the Lyapunov exponents and invariant density are independent of the initial conditions on the orbit.

Here we find the parameters $\vec{a}$ in $\underline{F}(\vec{y},\vec{a})$ minimizing a cost function subject to certain constraints. The constraints are chosen so that iterations of the mapping function $\underline{F}(\vec{y},\vec{a})$ give rise to values of dynamical invariants that are the same as those indicated by the experimentally measured data set $\vec{y}(n)$. In this way essential geometric information about the particular attractor on which the data are found will be built into the parametric mapping. In the next section we discuss the structure of the predictor.

2.2 The Structure of the Predictor

Now that we have successfully embedded the data x(n) in $\mathbb{R}^d$ by creating d-dimensional vectors $\vec{y}(n)$, n=1, ..., N, we need to choose a class of parameterized mappings, a cost function to minimize, and a means to impose the constraints on our minimization. Our maps are required to look around at the behavior of the phase-space neighbors of the point $\vec{y}(n)$ and predict forward according to how a cluster of phase-space neighbors, regardless of its temporal sequence, is moved forward in time. The map will then map forward any new point $\vec{y}$ to some weighted average of its neighbors' forward evolution.

We take our mappings to be of the form

$$\underset{\sim}{F}(\vec{y},\vec{a}) = \sum_{n=1}^{N-1} \vec{y}(n+1)\; g\;[\vec{y},\vec{y}(n);\vec{a}\;] \tag{2-4}$$

where g[$\vec{y}$,$\vec{y}$(n); $\vec{y}$] is near 1 for $\vec{y} = \vec{y}$(n) and vanishes rapidly for nonzero $|\vec{y} - \vec{y}(n)|$. This type of mapping is very similar to the form used in communications engineering under the name "kernel density estimation."

There are a few general requirements that g[$\vec{a}$, $\vec{y}$(n); $\vec{a}$] has to satisfy. These requirements are common in estimation theory (e.g., as $\sigma \to 0$, g[$\vec{y}$,$\vec{y}$(n);$\vec{a}$]→ delta function). The value of g[$\vec{y}$,$\vec{y}$(n);$\vec{y}$] should be numerically stable and computationally efficient. These requirements are all satisfied by the following choice.

$$g(\vec{y},\vec{y}(n);\vec{a}) =$$

$$\frac{e^{\frac{-|\vec{y}-\vec{y}(n)|^2}{\sigma}}\left[a_1 + a_2\vec{y}(n)(\vec{y}-\vec{y}(n)) + \sum_{k=3}^{P} a_k\left(\frac{|\vec{y}-\vec{y}(n)|^2}{\sigma}\right)^{m_k}\right]}{\sum e^{\frac{-|\vec{y}-\vec{y}(n)|^2}{\sigma}}\left[a_1 + \sum_{k=3}^{P} a_k\left(\frac{|\vec{y}-\vec{y}(n)|^2}{\sigma}\right)^{m_k}\right]} \tag{2-5}$$

The parameter $\vec{a}$ is P-dimensional, $\vec{a} = [a_1,a_2,a_3,\ldots a_p]^\dagger$, and σ is a fixed parameter for scaling. The variable m_k is also fixed at various values. We can treat σ and m_k as parameters to be optimized in the same sense as the $\vec{a}$.

The choice of the cost function is also up to us. Since we treat $\underset{\sim}{F}(\vec{y},\vec{a})$ as a dynamical system evolving points $\vec{y}$(n) into $\vec{y}$(n+1), our map should reproduce from $\vec{y}$(n) not only $\vec{y}$(n+1) but, via application of the map, a sequence of $\vec{y}$(n+1), $\vec{y}$(n+2), . . . $\vec{y}$(n+L) up to some L beyond which we simply do not trust the accuracy of our algorithm $\underset{\sim}{F}$ or of the computer we use to compute the future $\vec{y}$s.

In electrical engineering it is a common practice to form predictors of linear systems as

$$\vec{y}(m + 1) = \sum_{k=1}^{L} X_k \, \vec{y}(m - k + 1) \tag{2-6}$$

This suggests that the predictor for future points of a nonlinear system is a linear combination of iterated powers of the map $\underset{\sim}{F}(\vec{y},\vec{a})$ as a generalization to the linear predictor. The nonlinear generalization is

$$\vec{y}(m + 1) = \sum_{k=1}^{L} \vec{X}_K \, \underset{\sim}{F}^k [\, \vec{y}(m - k + 1),\vec{a}\,] \tag{2-7}$$

where $\underset{\sim}{F}^k$ is the kth application of $\underset{\sim}{F}$. If $\underset{\sim}{F}(\vec{y},\vec{a})$ were the exact mapping, we would require $\sum_{k=1}^{L} X_k = 1$. The X_k weight the various iterations of $\underset{\sim}{F}$ and are used to determine which iterations of $\underset{\sim}{F}$ we believe are the most accurate. We should require $x_j \geq x_{j+1}$ to indicate that the lower iterations of $\underset{\sim}{F}$ are believed to be more accurate than higher iterations.

Our predictor combines both past information from times m-k+1 and k = 1,2,...L and information from all the phase-space neighbors of the orbit points $\vec{y}$(m-k+1) because of the structure of $\underset{\sim}{F}(\vec{y},\vec{a})$.

By extracting the phase-space information in $\underset{\sim}{F}(\vec{y},\vec{a})$, we efficiently tap properties of the full data set. In the next section we study the form of cost function for our predictor.

2.3 Cost Function for Nonlinear Predictor

The cost function is nothing but a normalized rms deviation from our prediction to the actual data points. Therefore, the cost function associated with our nonlinear predictor is

$$\underset{\sim}{C}(\vec{X},\vec{a}) = \frac{\sum_{n=l}^{N-1} \left[\left| \vec{y}(n+1) - \sum_{k=1}^{L} \vec{X}_k \, \underset{\sim}{F}^k [\vec{y}(n - k + 1),\vec{a}] \right|^2 \right]}{\sum_{n=1}^{N} |\vec{y}(n) \cdot \vec{y}(n)|^2} \tag{2-8}$$

This kind of cost function will automatically contain information on the Lyapunov exponents, which themselves are expressions of the dynamics as iterations of the map. Some information on the invariant density function on the attractor is also contained in this improved cost function. Different choices for the function $g[\bar{y},\bar{y}(n);\ \bar{a}]$ can be taken. The Gaussian we work with could be replaced by a Lorentzian or other choices that weight neighbors more.

Section 3 Invariants of Dynamical Systems

Having a map and a cost function, we are ready for the constraints. We discuss how to determine Lyapunov exponents and invariant density from the $\underline{F}(\bar{y},\bar{a})$. Equating the numerical values for the Lyapunov exponents λ_i extracted from data to their expression found from $\underline{F}(\bar{y},\bar{a})$ will give us our first set of constraints on the minimization of $\underline{C}(\bar{x},\bar{a})$. These constraints produce the optimum values of parameters $\bar{a}$.

Since we have a finite amount of information, we choose to express this in terms of the projection of $\rho(\bar{y})$ on a set of dual basis functions that is a complete set in $\mathbb{R}^d$. By projecting the $\rho(\bar{y})$ determined from the data onto these basis functions, we can determine the coefficients of the expansion of $\rho(\bar{y})$ in this basis. Similarly, we can project the $\rho(\bar{y})$ determined from the map $\underline{F}(\bar{y},\bar{a})$ onto these basis functions and determine the expansion coefficient of the map. Equating the coefficients from the data to the ones determined from the map gives our final constraints on the minimization of $\underline{C}(\bar{x},\bar{a})$.

One can test these techniques by numerically generating a data set of x(n) from the known attractors like Henon or others and treating these data as having come from an unknown source. As the dimension of the phase space increases, the amount of data necessary for accurate prediction increases dramatically. Once $\underline{F}(\bar{y},\bar{a})$ is found, the details of $\underline{F}(\bar{y},\bar{a})$ should reflect the known features of the phenomena giving the signal.

3.1 Choice of the Embedding Dimension d

Here we would like to determine the correct value of the embedding dimension d from the scalar time series x(n), $n=1,2,\ldots,N_D$. We assume that there are enough data that we need not be concerned with statistical issues about numerical accuracy. We also assume that extrinsic noise is absent from the x(n) when we receive them. We further assume that by following Taken's phase-space reconstruction technique we have successfully captured the dynamics and embedded our time series. This requires a correct choice of τ, which will be discussed in the next section.

For now, let's further assume we have a correct τ to construct the attractor in the phase space. To establish dimension d, we need some characteristic of the attractor that becomes unchanging as d becomes large enough, thus indicating that the attractor can be embedded in $\mathbb{R}^d$. This invariant characteristic of the attractor is the attractor dimension d_A. One increases d until d_A remains constant and identifies the minimum d where d_A "saturates" as the embedding dimension. But computation of d_A is difficult, so we use the correlation function D(r) proposed by Takens (Reference 8).

$$D(r,N,d) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} U(r-\|\vec{y}(j) - \vec{y}(i)\|) \quad i \neq j \tag{3-1}$$

where $U(z)$ is just the unit step function $U(z) = \begin{cases} 1 & z>0 \\ 0 & z<0 \end{cases}$.

For N large enough, the behavior of D(r,N,d) for r becomes independent of N and D(r,N,d) takes the form

$$D(r,N,d) = \Phi(r,d) r^{\nu(d)} \tag{3-2}$$

If we plot D(r,N,d) versus r we can single out the correct value of dimension d as in Figure 3-1.

From Figure 3-1 it is concluded that the minimum value of d=3 is the right choice beyond which attractor dimension d_A does not change or the slope of our graph becomes constant.

In the next section we study the correct choice of τ to reconstruct the phase-space attractor.

3.2 Choice of the Time Shift τ

The choice of time shifts τ is not well agreed upon. If the underlying system were a differential equation and a scalar variable x(t) were measured at discrete times $x(n) = x(t_o + n\,\Delta t)$, then we would be, by the choice of logged variables, trying to find a discrete replacement for the usual phase-space coordinates:

$$x(t),\ \dot{x}(t),\ddot{x}(t),\ \ldots\ \frac{d^{d-1}x(t)}{dt^{d-1}}$$

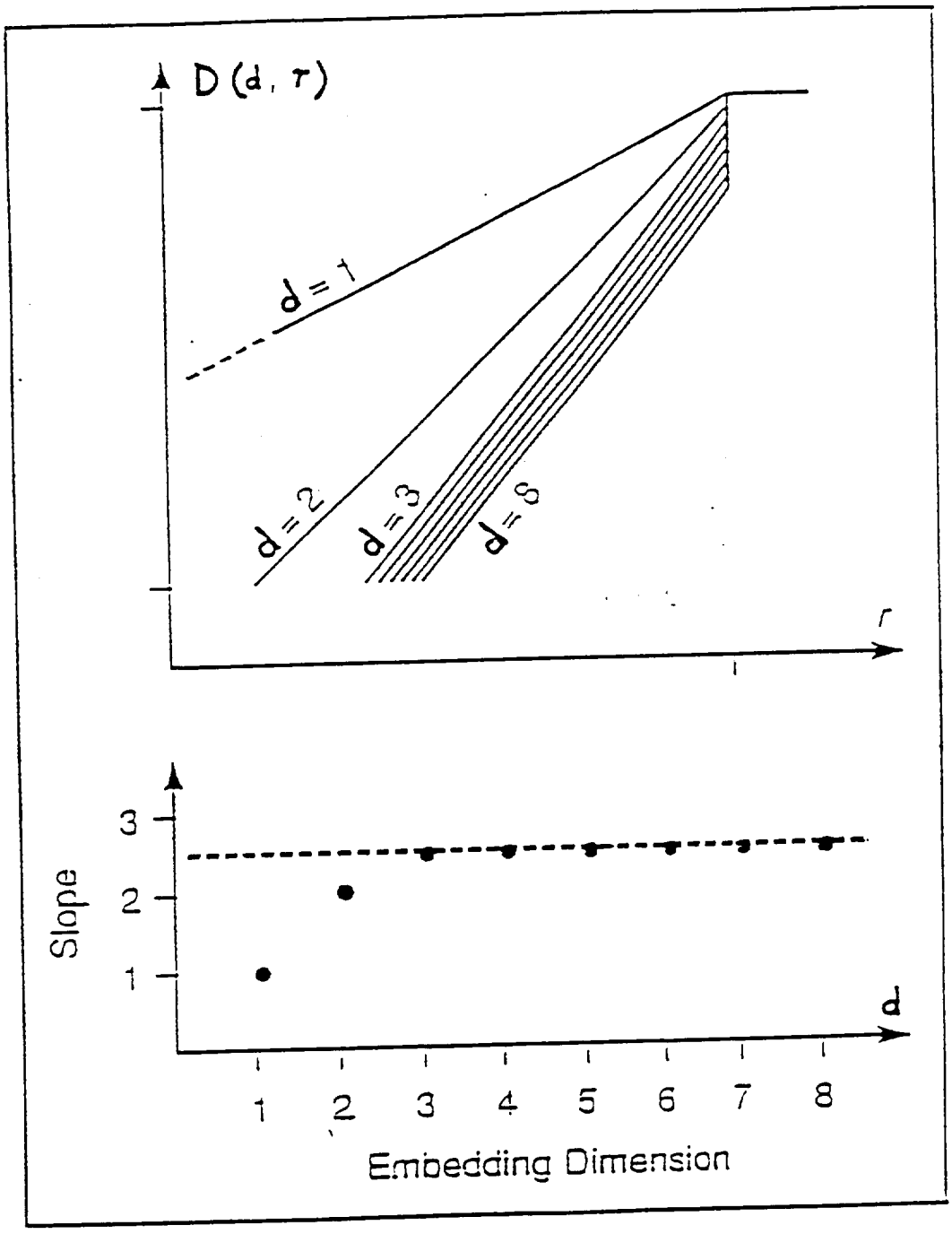
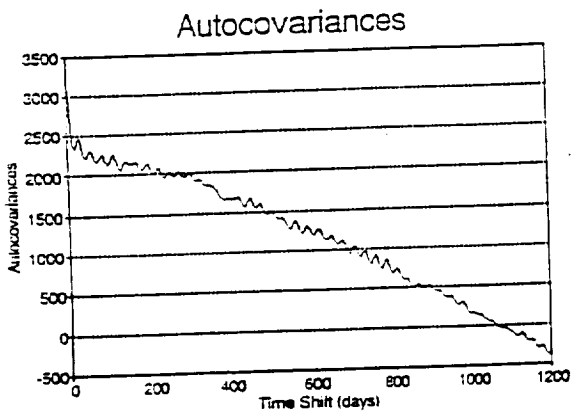


Figure 3-1. Correlation and Attractor Dimensions (Reference 9)

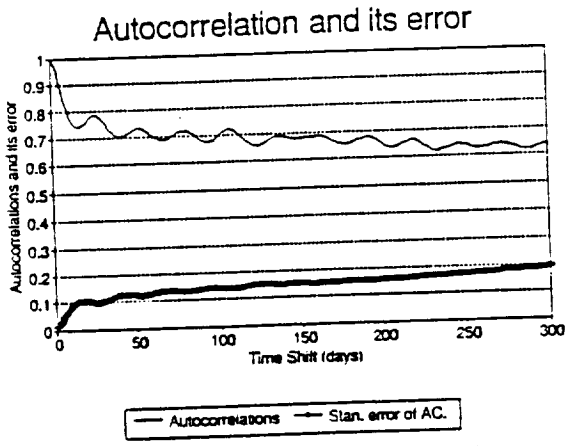
The best choice for time shift τ is a fraction of the time associated with the first local minimum of the autocorrelation function

$$\frac{1}{T}\int_0^T x(t+\tau)x(t)dt \tag{3-3}$$

We find that this choice, although somewhat arbitrary, works well in practice and gives a simple systematic way of determining τ. The autocorrelation of more than 4000 data points of solar flux time series is shown in Figure 3-2. The first local minimum occurs at t=13 days, so τ=3 days is a good choice.



Plot of Autocovariances for Solar Flux Time Series of 1200 Shifts



Plot of Autocorrelation and Standard Error of Autocorrelation for Solar Flux Time Series of 300 Shifts

Figure 3-2. Autocovariance and Autocorrelation of Solar Flux Series

Section 4. The Most Important Invariant of a Time Series

4.1 Extracting the Largest Lyapunov Exponent from a Time Series

4.1.1 Description Of Extraction

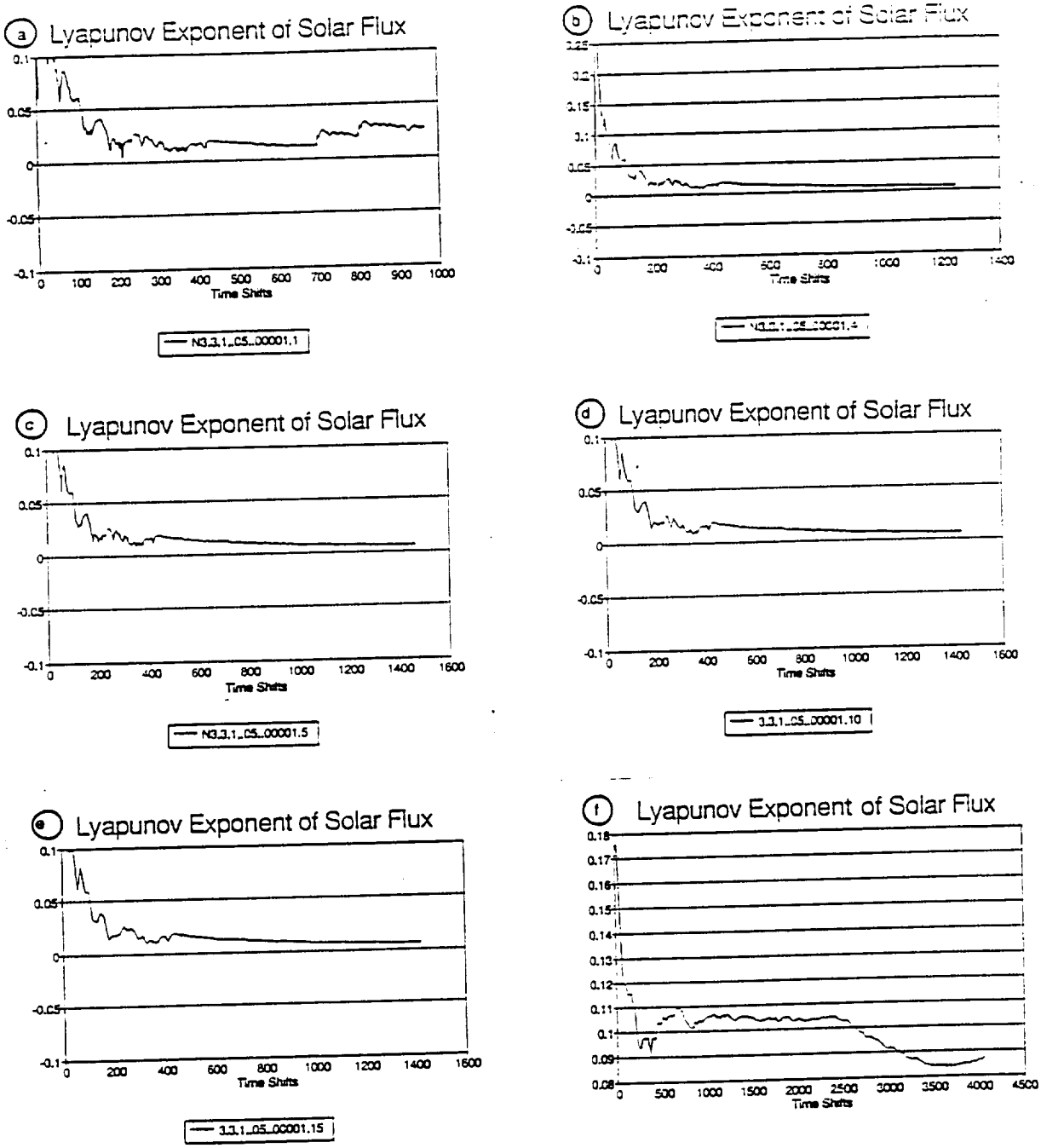
The sum of the Lyapunov exponents is the time-averaged divergence of the phase space trajectory; hence any dissipative dynamic system will have at least one negative exponent. Any dynamic system without a fixed point will have at least one zero Lyapunov exponent.

A small positive Lyapunov exponent is an indication of chaos and a very large positive Lyapunov exponent is an indication of a totally stochastic or random system. Therefore, the sign of the exponent provides a qualitative picture of a system's dynamics. A positive exponent represents chaos, a zero exponent represents a marginally stable system, and a negative exponent, a periodic system.

Figure 4-1 shows the actual solar flux data and their largest Lyapunov exponent for more than 4000 points. Here we have used the technique of phase-space reconstruction with delay coordinates shown in Figure 2-2 for a small portion of a time series.

To check whether the program that generates Figure 4-1 is really functioning well, we have plotted in Figure 4-2 the Lyapunov exponent of a time series that was generated from a sinusoidal function. Because a sinusoidal function is well behaved and totally deterministic, its Lyapunov exponent should approach zero. (Remember that we can generate only a truncated sinusoid and not an infinitely extended sinusoid.)

This technique of extracting the largest Lyapunov exponent was originally developed by Wolf, Swift, Swinney, and Vastano (Reference 5). The disadvantage of this method is that it gives only the positive largest Lyapunov exponent. To improve the technique that can enable us to determine all the Lyapunov exponents, Eckmann, Kamphorst, Ruelle, and Ciliberto developed a QR matrix decomposition method of calculating the exponents. Their method is more complicated than the Wolf-Swift-Swinney method and is computationally more expensive. For



Largest Lyapunov exponent for N=4090 data points, d=3 three dimensional embedding space, T=3 days time shift, DT=1 one day time difference between the data samples, SCalmx=0.05 length scale, SCalmn=0.00001 smallest length scale, and different evolution time

(a) Evolv=1 (b) Evolv=4 (c) Evolv=5 (d) Evolv=10 (e) Evolv=15

(f) a plot of Lyapunov exponent for much larger T, SCalmx, SCalmn, Evolv.

Figure 4-1. Lyapunov Exponents of Solar Flux

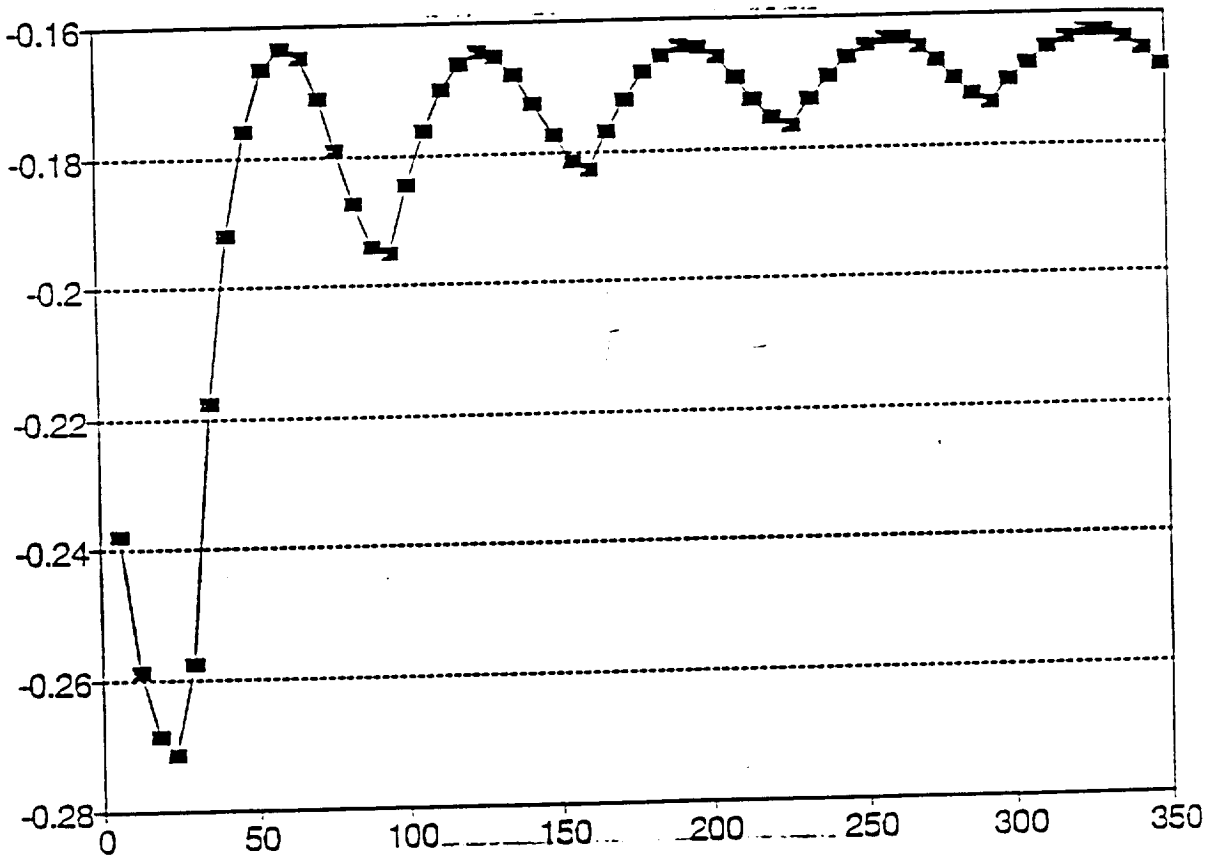


Figure 4-2. The Lyapunov Exponent of a Truncated Sinusoid

this reason we refer the reader to Reference 11 for further investigation of their method. However, because the Wolf-Swift-Swinney method is very simple, we will use it in our construction of a predictor and will discuss this method in ten steps.

4.1.2 Procedure for Extraction: The Wolf-Swift-Swinney Method

NPT = length of our solar flux time series

DIM = dimension of the phase-space reconstruction

TAU = time delay reconstruction

DT = time between the data samples → to normalize the exponent

SCALMX = length scale we consider to be larger → no more information is probed $\ell >$ scalmx

SCALMN = length scale we consider to be small → noise dominates $\ell <$ scalmnx

EVOLV = constant propagation time ($\Delta = t_{k+1} - t_k$)

ANGLMX = maximum angular error (not a free parameter): does not really change the exponent estimate and is usually between 0.2 and 0.3 radians

1) Search data for nearest neighbor to the first essential point. (This is not easy.)

2) Disregard points closer than SCALMN.

3) Propagate current pair of points EVOLV steps through the attractor and compute the final separation.

4) Find the log of the ratio of final to initial separation of this pair. This updates a running average rate of orbital divergence.

5) Attempt a replacement step.

6) Calculate the distance of each delay coordinate point to the evolved **essential point.**

7) Check whether or not the angular orientation is less than ANGLMX radians for points closer than SCALMX but farther away than SCALMN.

8) If more than one point is found, use the point defining the smallest angular change for replacement.

9) If no points are found, take replacement points as far as 2 x SCALMX away.

10) Repeat this process until the essential trajectory gets to the end of the data file, by which time we hope to see stationary behavior in the Lyapunov exponent.

$$\lambda = \frac{1}{t_M - t_o} \sum_{k=1}^{M} \log_2 \frac{L'(t_k)}{L(t_{K-1})} \tag{4-1}$$

where M is the total number of replacement steps.

Insert it from next page

As Figure 4-1 shows, we have a good convergence of the largest Lyapunov exponent for parameter values of N=4090 data points, d=3 three-dimensional embedding space, τ=3 three day time shift, DT=1 one day time difference between the samples, SCALMX = 0.05 maximum length scale, SCALMN=0.00001 minimum length scale, EVOLV=4,5,10,15. As we increase the EVOLV, we reach a point where the exponent does not change its convergence behavior. The first value of EVOLV that makes a satisfactory convergence is the optimum EVOLV. In our case, this value is EVOLV=4 and the Lyapunov exponent is about λ_{max} = 0.01. Now that we have the Lyapunov exponent directly from our time series, we are ready to calculate it from the map. Setting equal λ calculated from the map and λ calculated from the time series would give us a set of constraint equations to calculate parameters $\vec{a}$ to minimize the cost function $\underline{C}(\vec{x},\vec{a})$.

4.2 Determining the Lyapunov Exponents from the Map $\underline{F}(\vec{y},\vec{a})$

Whatever method we use (the Wolf-Swift-Swinney method or the Eckmann-Kamphorst-Ruelle method) to determine the Lyapunov exponents from the time series, we must now establish a way to express these same quantities in terms of our map $\underline{F}(\vec{y},\vec{a})$. This was originally done by Shimada and Nagashima (Reference 12). But to easily use the results in our optimizations, we

INSERT A

A point on the attractor is given by

Figure 4-3. Evolution and Replacement Procedures Used to Estimate Lyapunov Exponent Directly From a Time Series

$$|y_1\rangle = \begin{bmatrix} f_{t_1} \\ f_{t_1+\tau} \\ f_{t_1+2\tau} \\ \vdots \\ f_{t_1+(c-1)\tau} \end{bmatrix} \quad |y_2\rangle = \begin{bmatrix} f_{t_2} \\ f_{t_2+\tau} \\ f_{t_2+2\tau} \\ \vdots \\ f_{t_2+(c-1)\tau} \end{bmatrix} \quad \ldots\ldots\ldots |y_N\rangle$$

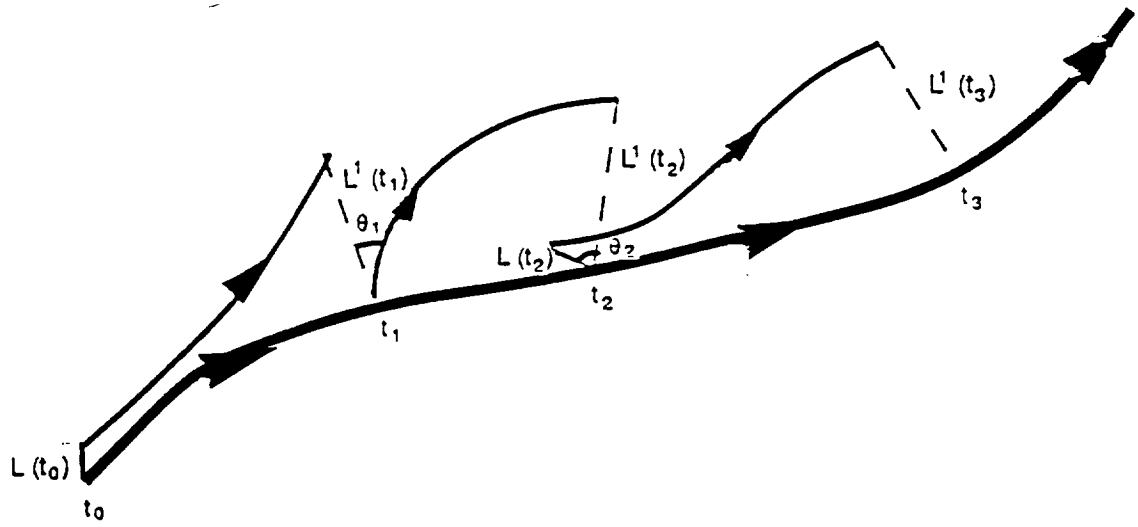


Figure 4-3

can use the slightly different approach presented by Abarbanel et al. (Reference 13).

Lyapunov exponents characterize the way neighboring points, small areas, or small volumes near the orbit of interest evolve under the mapping. To find them one linearizes the mapping $\vec{y}(n+1) = \underline{F}[\vec{y}(n),\vec{a}]$ around a given orbit $\vec{y}(1)$, $\vec{y}(2)$, $\vec{y}$, . . . $\vec{y}(n)$. Small deviations from this orbit, called $\delta\vec{y}(n)$, evolve as

$$\delta\vec{y}(n+1) = \underline{M}\,[\vec{y}(n)]\,\delta\vec{y}(n)$$

where

$$[\underline{M}(\vec{y})]_{jl} = \frac{\partial}{\partial y_l}\,\underline{F}_j(\vec{y},\vec{a}) \tag{4-3}$$

is evaluated along the orbit of interest. The Lyapunov exponents are found from the eigenvalues of matrix $\underline{M}^k[y(1)]$

$$\underline{M}^k\left[\vec{y}(1)\right] = \underline{M}\left[\vec{y}(k)\right]\underline{M}\left[\vec{y}(k-1)\right]\underline{M}\left[\vec{y}(k-2)\right]\ldots\underline{M}\left[\vec{y}(1)\right] \tag{4-4}$$

Now, to find the largest Lyapunov exponent, we apply the Matrix $\underline{M}^k$ to an arbitrary vector $\vec{w}^k$. Then, forming

$$\lambda_1 = \lambda_{max} = \frac{1}{\tau k}\,Ln\,\frac{\|\underline{M}^k(\vec{W})\|}{\|\vec{w}\|} \tag{4-5}$$

from the trace of the Matrix $\underline{M}^k$ we can write

$$\lambda_1 = \frac{1}{\tau k}\,Ln\,[tr(\underline{M}^k)] \qquad \text{(for large k)} \tag{4-6}$$

Remember that $\underline{M}$ is a function of $\bar{a}$. To find the next largest exponent λ_2 we can use

$$\lambda_1 + \lambda_2 = \frac{1}{\tau_k} Ln \{[tr(\underline{M}^k)]^2 - tr(\underline{M}^{2k})\} \qquad (4\text{-}7)$$

Section 5. Invariant Measure On The Attractor

The frequency with which orbits y(n) visit regions of the phase space $\mathbb{R}^d$ defines an invariant distribution function $\rho(\bar{y})$, which is formally defined for the mapping $\bar{y}(n+1) = \mathbf{F}[\bar{y}(n)]$ as

$$\rho(\bar{y}) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \delta^d \{y - F^k[y(1)]\} = \lim_{N\to\infty} \rho_N(y) \qquad \text{(for maps)} \qquad (5\text{-}1)$$

In similar fashion, the invariant distribution for a numerical data set y(n), n=1,2,...,N is given by

$$\rho(\bar{y}) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \delta^d [\bar{y} - \bar{y}(k)] \qquad \text{(for time series)} \qquad (5\text{-}2)$$

Any finite sequence of N points has a finite resolution on the attractor. That resolution is approximately N^{-1/d_A}, which is the order of the mean distance of N points on a d_A-dimensional set. To handle this matter of finite resolution, we introduce a complete orthonormal set of functions $\psi_\mu(\bar{y})$ defined on $\mathbb{R}^d$, which can serve as a basis set. Truncating this expansion at some finite order $\mu=G$ provides a finite-resolution representation corresponding to whatever information we have on $\rho(\bar{y})$. We then expand

$$\rho(\bar{y}) = \sum_{\mu=1}^{G} B_\mu \, \psi_\mu(\bar{y}) \qquad (5\text{-}3)$$

The coefficients B_μ will be the invariants of the dynamic process that characterizes $\rho(\bar{y})$ within a given basis $\psi_\mu(\bar{y})$.

Now, the challenge is to extract B_μ from the data vectors $\bar{y}(n)$ and from the parameterized map $\underline{F}(\bar{y},\bar{a})$. Equating the B_μ from the data to those from the map will produce our final constraints on the cost function $\underline{C}(\bar{x},\bar{a})$.

5.1 Coefficients B_μ Of The Invariant Measure From The Time Series

Any choice of basis $\psi_\mu(\vec{y})$ would do the job, such as the complete orthogonal (or orthonormal) set used in the Fourier Series $\psi_{\vec{m}}(\vec{y}) = e^{i\vec{m}.\vec{y}}$, where $\vec{m} = [m_1, m_2, m_3, \ldots m_\alpha]^\dagger$. But since our attractor is bounded, most of the work performed by the Fourier representation of $\rho(\vec{y})$ will be expended in making $\rho(\vec{y})$ vanish from the attractor. Thus, what we need is orthonormal functions concentrated on the attractor. Now we construct an optimal choice for $\psi_\mu(\vec{y})$. Take the total data set $\vec{y}(n)$, $n = 1,2, \ldots$. and divide it into two portions as in Figure 5-1.

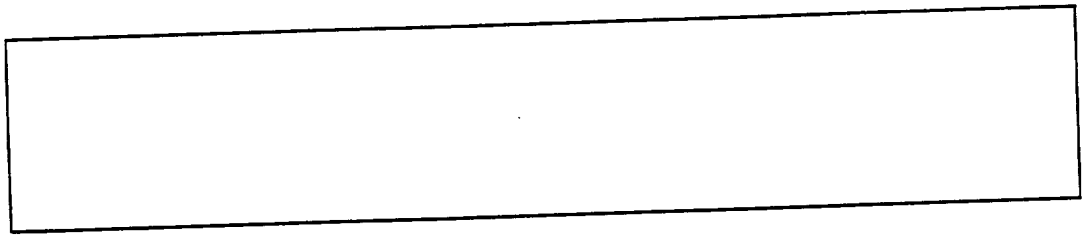


Figure 5-1. Breaking the Time Series Into Small Sections for Proper Treatment

Now take the second portion of length Q and further divide it into G subsections, each L long (Q=GL). Each group is a sample of the invariant attractor. Treat each of the G data sets as an independent sample of $\rho(\vec{y})$ and form the invariant distribution for the αth sample (see Figure 5-2).

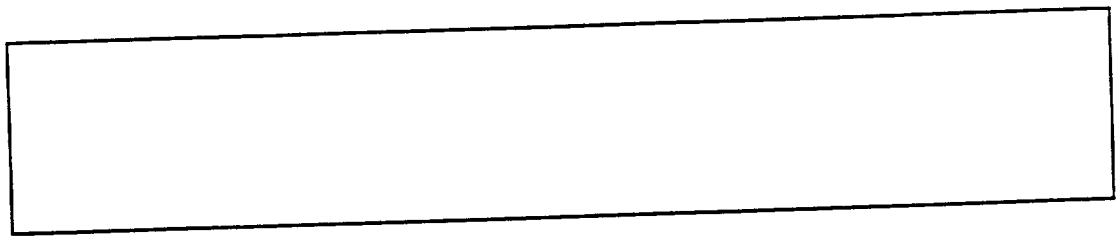


Figure 5-2. One Portion of the Time Series

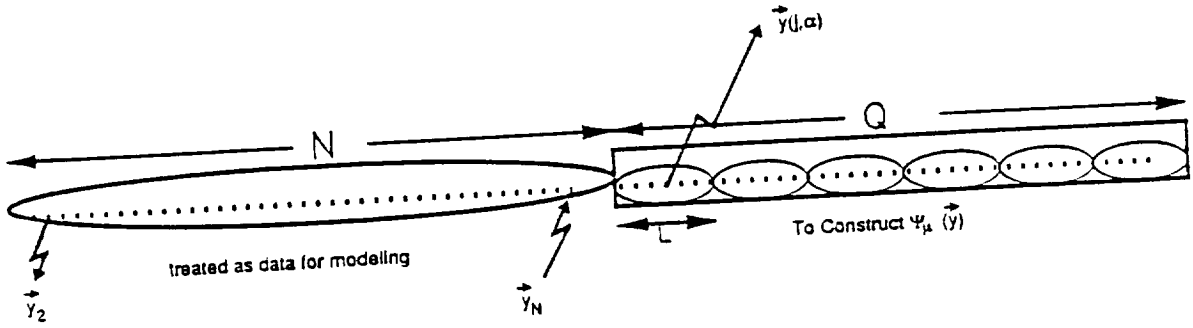
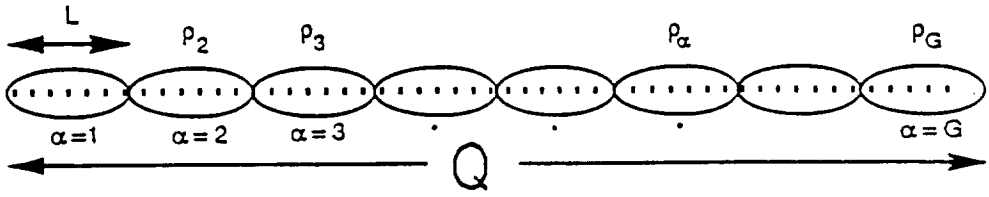

$\vec{y}(l,\alpha)$
N
Q
L
To Construct $\Psi_\mu(\vec{y})$
treated as data for modeling
$\vec{y}_2$
$\vec{y}_N$

Figure 5-1



To calculate $\Psi_\mu(\vec{y})$

Figure 5-2

$$\rho_\alpha(\vec{y}) = \frac{1}{L}\sum_{k=1}^{L} \delta^d\left[\vec{y} - \vec{y}(k,\alpha)\right] \tag{5-4}$$

with $\alpha = 1, 2, \ldots G$. The data point $\vec{y}(k,\alpha)$ is the kth member of the αth sample. Of course, the mean density of the G samples is just the total invariant density of the data set of length Q.

$$\rho(\vec{y}) = \frac{1}{G}\sum_{\alpha=1}^{L} \rho_\alpha(\vec{y}) \tag{5-5}$$

Now, from G samples $\rho\alpha(y)$ we form the following phase-space correlation function:

$$R(\vec{z},\vec{w}) = \frac{1}{G}\sum_{\alpha=1}^{L} \rho_\alpha(\vec{z})\ \rho_\alpha(\vec{w}) \qquad \text{(Every iteration is within a given sample.} \tag{5-6}$$

It can be shown that the normalized eigenfunctions of this correlation function are the optimal $\psi_\mu(\vec{y})$ for expansion of functions localized on the attractor.

The requirement that $\psi_\mu(\vec{y})$ be an eigenfunction of $\mathbb{R}(\vec{z},\vec{w})$ leads to

$$\int d^{d_z} R(\vec{w},\vec{z})\ \psi_\mu(z) = \mu\,\psi_\mu(\vec{w}) \tag{5-7}$$

The $\psi_\mu(\vec{y})$'s are normalized as follows

$$\int d^{d_w} \psi_\mu(\vec{w})\,\psi_{\mu'}(\vec{w}) = \delta_{\mu\mu'} \tag{5-8}$$

For an infinite G, the set of eigenfunctions becomes complete. For finite G, $R(\vec{w},\vec{z})$ becomes a finite sum of separable kernels. Thus

$$\psi_\mu(\vec{y}) = \sum_{\alpha=1}^{G} C_\alpha^\mu\,\rho_\alpha(\vec{y}) \tag{5-9}$$

Now they are localized near the attractor, because $\rho_\alpha(\vec{y})$'s vanish beyond the attractor. Simple substitutions reduce the problem to a finite matrix problem.

$$\int d^d z \left[\frac{1}{G}\sum_{\alpha=1}^{G}\rho_\alpha(\vec{z})\rho_\alpha(\vec{w})\right]\left[\sum_{\alpha=1}^{G}C_\alpha^\mu\rho_\alpha(\vec{z})\right] = \mu\sum_{\alpha=1}^{G}C_\alpha^\mu\rho_\alpha(\vec{w}) \tag{5-10}$$

$$\sum_{\beta=1}^{G}A_{\alpha_\beta}C_\beta^\mu = \mu C_\alpha^\mu \quad where \quad A_{\alpha_\beta} = \frac{1}{G}\int d^d z\,\rho_\alpha(z)\rho_\beta(z) \tag{5-11}$$

using the normalization condition we have

$$\sum_{\alpha=1}^{G}C_\alpha^\mu C_\alpha^{\mu'} = \frac{1}{\mu G}\delta\mu\mu' \tag{5-12}$$

This shows that all the eigenvalues μ are positive.

For all the calculations and computations of $\rho_\alpha(\vec{y})$ we use the Gaussian representation of Δfunction

$$\delta^d(\vec{x}) = \frac{1}{\sqrt{(\pi\bar{\omega})^d}}^{\rho-|x|^2\bar{\omega}} \equiv f_{\bar{\omega}}(|x|) \tag{5-13}$$

therefore

$$\rho_\alpha(\vec{y}) = \frac{1}{L}\sum_{\kappa=1}^{L}f_{\bar{\omega}}\left(|\vec{y}-\vec{y}(k,\alpha))|\right) = \frac{1}{L}\sum_{k=1}^{L}\frac{1}{\sqrt{(\pi\bar{\omega})^d}}e^{-|\vec{y}-\vec{y}(k,\alpha)|^2\bar{\omega}} \tag{5-14}$$

thus

$$A_{\alpha\beta} = \frac{1}{\sqrt{(\pi\bar{\omega})^d}}\frac{1}{GL^2}\sum_{kj=1}^{L}e^{-|\vec{y}(k,\alpha)-\vec{y}(j,\beta)|^2/\bar{\omega}} \tag{5-15}$$

when $\bar{\omega}$ is small, we have only a small loss of resolution in calculating $\rho_\alpha(\vec{y})$. Now we are ready to find our optimal $\psi_\mu(\vec{y})$ from the G data sets. First we calculate the G $\times$ G matrix $A_{\alpha\beta}$. Next we calculate the eigenvalues μ and eigenvectors C_α^μ of this matrix, being sure to normalize them according to Equation (5-12). Now we can calculate $\psi_\mu(\vec{y})$ by using the normalized C_α^μ and the

$\rho(\vec{y})$ from Equations (5-9 and 5-14). A typical value for L is 750 $\vec{y}$ points and for G is 5 samples. These are good values for N=4500 $\vec{x}$points of our time series. We now have G orthonormal functions $\psi_\mu(\vec{y})$ extracted from G samples $\rho(\vec{y})$ of the invariant distribution. We can use Equation (5-8) to project a particular B_μ out of Equation (5-3),

$$B_\mu = \int d^d\vec{y}\ \rho(\vec{y})\ \psi_\mu(\vec{y}) \tag{5-16}$$

This shows that B_μ are invariants of the dynamics since they are integrals of ψ_μ with the density $\rho(\vec{y})$.

Therefore,

$$B_\mu = \frac{1}{N}\sum_{\alpha=1}^{G}\sum_{\kappa=1}^{N} C_\alpha^\mu\, \rho_\alpha(\vec{y}(\kappa)) = \frac{1}{NL}\sum_{\kappa=1}^{N}\left[\sum_{\alpha=1}^{G}\sum_{j=1}^{L}\frac{C_\alpha^\mu}{\sqrt{(\pi\bar{\omega})^d}}\, e^{-|\vec{y}(j,\alpha)-\vec{y}(\kappa)|^2/\bar{\omega}}\right] \tag{5-17}$$

(from data)

because from Equation (5-2, 5-9, and 5-16) we have

$$\beta_\mu = \int d^d\vec{y}\rho(\vec{y})\psi_\mu(\vec{y}) = \int d^d\vec{y}\left[\lim_{N\to\infty}\frac{1}{N}\sum_{\kappa=1}^{N}\delta^d(\vec{y}-\vec{y}(k))\right]\left[\sum_{\alpha=1}^{G}C_\alpha^\mu\rho_\alpha(\vec{y})\right]$$

$$= \lim_{N\to\infty}\sum_{\alpha=1}^{G}\sum_{\kappa=1}^{N}C_\alpha^\mu\int d^d\vec{y}\,\delta^d(\vec{y}-\vec{y}(k))\rho_\alpha(\vec{y}) = \frac{1}{N}\sum_{\alpha=1}^{G}\sum_{\kappa=1}^{N}C_\alpha^\mu\rho_\alpha(\vec{y}(\kappa)) \tag{5-18}$$

$$\rho_\alpha(\vec{y}(\kappa))$$

Thus, Equation (5-17) is used to calculate B_μ from the data, where B_μ are the G numbers characterizing the invariant density $\rho(\vec{y})$ by its projection on the optimum basis vectors $\psi_\mu(\vec{y})$. In the next section we discuss how one can find B_μ from the map. The G equalities between these two evaluations of B_μ form our final constraints on the minimization of the cost function $\underline{C}(\vec{x},\vec{a})$.

5.2 Coefficients B_μ Of The Invariant Measure From The Map

To determine $B\mu$ from the map $\underline{F}(\bar{y})$, we look at the definition of invariant density as expressed by Equation (5-1) and Equation (5-2). Define A_K the projection on $\psi_\mu(\bar{y})$

$$A_\kappa(\mu) = \int d^d\bar{y}\psi_\mu(\bar{y})\delta^d\left(\bar{y}-F^\kappa(\bar{y}(1))\right) = \psi_\mu(F^\kappa(\bar{y}(1))) \tag{5-19}$$

This is the projection of $\delta^d(y-\underline{F}^\kappa(\bar{y}(1)))$ onto the orthonormal $\psi_\mu(\bar{y})$. That is, we can expand the δ function in terms of $\psi_\mu(\bar{y})$ to get

$$\delta^d\left(\bar{y}-F^\kappa(\bar{y}(1))\right) = \sum_{\mu=1}^{G} A_\kappa(\mu)\psi_\mu(\bar{y}) \tag{5-20}$$

Therefore Equation (5-1) can be written as:

$$\rho(\bar{y}) = \lim_{N\to\infty}\frac{1}{N}\sum_{\kappa=1}^{N}\sum_{\mu=1}^{G} A_\kappa(\mu)\psi_\mu(\bar{y}) = \sum_{\mu=1}^{G}\left[\frac{1}{N}\sum_{\kappa=1}^{N} A_\kappa(\mu)\right]\psi_\mu(\bar{y}) \qquad \begin{matrix}(5\text{-}21)\\(\text{for large N})\end{matrix}$$

Comparing this with Equation (5-3), we have

$$B_\mu = \frac{1}{N}\sum_{\kappa=1}^{N} A_\kappa(\mu) = \frac{1}{N}\sum_{\kappa=1}^{N} \psi_\mu(F^\kappa(y(1))) \tag{5-22}$$

From Equation (5-9) and the definition of $\rho(\bar{y})$ we can write this as

$$B_\mu = \frac{1}{NL}\sum_{\kappa=1}^{N}\left[\sum_{j=1}^{L}\sum_{\alpha=1}^{G} C_\alpha^\mu\ \delta^d\left(\bar{y}(j,\alpha) - F^\kappa(y(1))\right)\right] \tag{5-23}$$

We can replace $F^{\kappa}(y(1))$ by $F(y(\kappa))$ and therefore

$$B_{\mu} = \frac{1}{NL}\sum_{\kappa=1}^{N}\left[\sum_{\alpha=1}^{G}\sum_{j=1}^{L} C_{\alpha}^{\mu}\,\delta^{d}\left(\vec{y}(j,\alpha) - \underset{\sim}{F}(\vec{y}(\kappa))\right)\right] \tag{5-24}$$

$$= \frac{1}{NL}\sum_{\kappa=1}^{N}\left[\sum_{\alpha=1}^{G}\sum_{j=1}^{L} \frac{C_{\alpha}^{\mu}}{\sqrt{(\pi\bar{\omega})^{d}}}\, e^{-|\vec{y}(j,\alpha) - \underset{\sim}{F}[\vec{y}(\kappa)]|^{2}/\bar{\omega}}\right] \quad \text{(from a map)}$$

We also observe that

$$\delta^{d}\left(\vec{y} - F^{\kappa+1}(\vec{y}(1))\right) = \int d^{d}w\, \delta^{d}(y - \underset{\sim}{F}(\vec{w}))\, \delta^{d}\left(\vec{w} - F^{\kappa}(y(1))\right) \tag{5-25}$$

thus

$$A_{\kappa+1}(\mu) = \sum_{\mu'} T_{\mu\mu'}\, A_{\kappa}(\mu') \tag{5-26}$$

where

$$T_{\mu\mu'} = \int d^{d}\vec{y}\, \psi_{\mu}\left(\underset{\sim}{F}(\vec{y})\right)\psi_{\mu}(\vec{y}) \tag{5-27}$$

and

$$\int d^{d}y\, \rho_{L}(\vec{y})\, \psi_{\mu}(\vec{y}) = \rho_{L}(\mu) = \frac{1}{L}\sum_{\kappa=1}^{L} A_{\kappa}(\mu) \tag{5-28}$$

which is

$$(1-T)\rho_{L}(\mu) = \frac{1}{L}(1-T^{L})A_{1} \tag{5-29}$$

since

$$\lim_{L\to\infty} \rho_L(\mu) = \int d^d y \rho(\bar{y}) \psi_\mu(\bar{y}) = B_\mu \tag{5-30}$$

This shows that the B_μ are the components of the eigenvector of T with eigenvalue unity, if Equation (5-1) for $\rho(\bar{y})$ converges.

Section 6. Conclusion

In this document we developed a set of procedures for forecasting solar flux without reference to any underlying solar physics. These procedures are applicable to any time series with a broadband power spectrum. We introduced techniques of extracting dynamical invariant directly from solar flux time series. The existence of positive, small Lyapunov exponents supports our hypothesis that solar dynamics is not a stochastic process; it is indeed chaotic.

A FORTRAN computer program for procedures outlined in this document is under development, and forecasting results will appear in future documents or technical papers.

When we have perfected these programs, we will be able to apply these techniques to any time-series predictions (orbit, attitude, atmospheric density, geomagnetic index, ... stock market). Perhaps the most important application will be to a time series generated by subtracting the predicted positon of a spacecraft using GTDS (with density flag = off) from the observed position of a spacecraft from telemetry or ground track data. This allows us to model all perturbations on the trajectory of the spacecraft directly from data.

PAPERS IN CHAOS THEORY AND APPLICATIONS

1. Ashrafi, S., *Future Missions Studies on Preliminary Comparisons of Solar Flux Models*, Goddard Space Flight Center, Flight Dynamics Division, FDD/544-91-004, December 1990

2. Ashrafi, S., *Future Missions Studies on Solar Flux Analysis Using Chaos*, Goddard Space Flight Center, Flight Dynamics Division, FDD/544-912-006, December 1990

3. Ashrafi, S., and L. Roszman, "Evidence of Chaotic Pattern in Solar Flux Through a Reproducible Sequence of Period-Doubling-Type Bifurcations," Proceeding of Flight Mechanics/Estimation Theory Symposium, May 1991

4. Ashrafi, S., *Future Missions Studies on Chaotic Solar Flux (Structural Stability, Attractor Dimension, Application of Catastrophe)*, Goddard Space Flight Center, Flight Dynamics Division, FDD/544-91-112, June 1991

5. Ashrafi, S, and L. Roszman, "Limits on the Predictability of Solar Flux time Series," Publication Release for *Journal of Geophysical Research*, June 1991

6. Packard, N., et al., "Geometry From a Time Series," *Physical Review Letters*, 1980, vol. 45, no. 712

7. Eckmann, J., and D. Ruelle, "Ergodic Theory Chaos and Strange Attractors," *Review of Modern Physics*, 1985, vol. 57, no. 617

8. Takens, F., "Dynamical Systems and Turbulence," *Lecture Notes in Mathematics , Volume 898*, Berlin: Springer, 1981, pp. 366 - 81

9. Theiler, J., "Quantifying Chaos: Practical Estimation of the Correlation Dimension," Ph.D. Dissertation, California Institute of Technology, 1987

10. Wolf, A., et al., "Determining Lyapunov Exponents From Time Series," Physical, 1985, 16D, no. 285

11. Eckmann, J., et al., "Lyapunov Exponents From a Time Series," *Physical Review*, 1986, vol. 34A, no. 4971

12. Shimada, I., and T. Nagashima, "Lyapunov Exponents," *Program in Theoretical Physics*, 1979, vol. 61, no. 1605

13. Abarbanel, H., R. Brown, and P. Bryant, "Computing the Lyapunov Spectrum of a Dynamical System From an Observed Time Series," *Physical Review*, 1991, vol. 43A, no. 6

p/s delete "el" and italicize Physica