Least squares method of estimating the noise level in a chaotic time series

A.W. Jayawardena
International Centre for Water Hazard and Risk Management (ICHARM) under the auspices of UNESCO, Public Works Research Institute
1-6, Minamihara, Tsukuba, Ibaraki 305-8516, Japan

The International Centre for Water Hazard and Risk Management
ICHARM

Content

- Introduction
- Correlation integral
- Noise level estimation
- Applications
  - Henon map
  - Lorenz map
  - Duffing equation
  - Rossler equation
  - Chua's circuit
  - Southern Oscillation Index (SOI) data
  - Sea Surface Temperature Anomaly (SSAT) index data
  - Normalized Tahiti-Darwin mean sea level pressure difference data
- Concluding remarks

Introduction

- Almost all types of hydrological time series are contaminated with noise, some to a lesser degree than others.
- Such noise may emanate from measurement errors, human errors, transcribing errors, among others.
- For time series analysis, it is ideal, and sometimes necessary, to have data which are noise free.
- This is particularly so for non-linear time series, which have signatures of chaotic dynamics, because the techniques of analysis and prediction have been developed under the assumption that the series are noise free.

Introduction….

- The presence of noise may also limit the performance of the techniques of identification, estimation of invariant measures, model selection and prediction of deterministic dynamical systems.
- For example, the presence of noise in the time series can destroy the self-similarity of the attractor, may distort the phase-space reconstruction.
- In the past many investigators from different disciplines have attempted to deal with the problem of noise level estimation and noise reduction. Such methods, most of which employ non-linear estimation techniques, lack general applicability and are too complicated to be used in real-world situation.

Introduction….

- In this paper, a relatively simple method of estimating the noise level in a chaotic deterministic time series using a linear least squares method is presented. The method is based on the correlation integral form obtained by Diks (1999), and the special property of the Kummer's confluent hypergeometric function. It is tested with a number of theoretical time series, which are known to become chaotic under certain parameter conditions and applied to some real-world data.
- The theoretical time series used are the Hénon Map, the Lorenz equation, the Duffing equation, the Rossler equation and the Chua's circuit.
- The real-world hydrological time series used are the southern oscillation index (SOI), eastern equatorial Pacific sea surface temperature anomaly index (SSTA), and the normalized Darwin-Tahiti mean sea level pressure differences.

Introduction….

- The tests with the theoretical series show consistent satisfactory results. There is however no verification for the real-world hydrological time series because the clean signal is unknown *a priori*.
- Since the method works well with the theoretical time series, it is believed that it will be satisfactory with the real-world data as well. Another advantage of the present method is that the correlation dimension of the time series can be estimated simultaneously.

Correlation integral

- The first step in treating a time series as chaotic is to diagnose the system; i.e. to determine whether the time series is driven by a low dimensional deterministic system. It can be done by computing several invariant measures such as, the fractal dimension, the correlation dimension, the Lyapunov exponent, and the Kolmogorov entropy among others. Of these, the correlation dimension plays a significant role in identifying the system as well as for prediction of the future states of the system. For a deterministic time series generated by a dynamical system, the correlation integral $C_m(r)$, for small r and large m, is given by the scaling relationship

Correlation integral….

Eq. 1: $$C_m(r) \sim e^{-m\tau K} r^D$$

where r is the radius, m is the embedding dimension, τ is the time delay, D is the correlation dimension and K is the correlation entropy per time unit, or simply correlation entropy. The correlation dimension and the correlation entropy can respectively be interpreted as an approximate measure of the number of degrees of freedom, and a measure of the rate (T^{-1}) at which initially nearby orbits diverge.

Correlation integral….

- The standard method of calculating the correlation integral is by the correlation sum method (Grassberger and Procaccia (1983a & b)), as defined below:

Eq.2 $$C_m(r) = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=1}^{N} H(r - \|\boldsymbol{Y}_i - \boldsymbol{Y}_j\|);$$

where *H* is the Heaviside step function with *H(u)* = 1 for *u* > 0, and *H(u)* = 0 for *u* ≤ 0; *N* is the number of points in the vector time series, $\boldsymbol{Y}_i$, $\boldsymbol{Y}_j$ are points in the re-constructed phase space; *r* is the radius of sphere centered on either of the points $\boldsymbol{Y}_i$, or $\boldsymbol{Y}_j$. A point in the phase space is defined as

Eq. 3 $$\boldsymbol{Y}(t) = (x(t), x(t-\tau), \cdots, x(t-(m-1)\tau))$$

where $x(t)$, $1 \le t \le N$ is a chaotic time series embedded in the reconstructed phase space of dimension *m* and time delay τ . The norm $\|\boldsymbol{Y}_i - \boldsymbol{Y}_j\|$ may be any one of the three usual norms, the maximum norm, the diamond norm, or the Euclidean norm. Correlation integrals are calculated for a series of embedding dimensions.

Correlation integral….

- A more simple form for the correlation integral has been obtained by Schouten *et al.*, (1994) also based on the maximum norm. Using an upper bound of radius r as r_o , and δ as the maximum noise amplitude, they obtained the correlation integral as

Eq. 4 $$C_m(r) = [\frac{r-\delta}{r_0-\delta}]^D$$

for $\delta \le r \le r_0$

- Subsequently, Diks (1999) derived the following expression for the correlation integral in the presence of noise when $C_m(r)$ is based on Euclidean norm:

Eq. 5 $$C_m(r) = \frac{\phi e^{-m\tau K} m^{-D/2} 2^{-m} \sigma^{D-m} r^m}{\Gamma(m/2+1)} M(\frac{m-D}{2}, \frac{m}{2}+1, -\frac{r^2}{4\sigma^2})$$

Correlation integral….

- Eq. 2, as well as other methods, is generally applicable to noise free time series. The presence of noise (dynamical and observational) strongly affects the correlation integral and the results may become distorted or even completely wrong. Several authors (for example, Ott and Hanson, 1981; Ott *et al.*, 1985; Smith, 1992; Schreiber, 1993 & 1997; Oltmans and Verheijen, 1997; and, Diks, 1999) have addressed the problem resulting from the presence of noise but it still basically remains a topic of current research interest.

- A notable contribution on this topic is that of Schreiber (1993a) who obtained the following approximate formula based on the maximum norm for the correlation integral for a time series contaminated with Gaussian noise:

Eq. 6
$$\frac{d[\ln(C_{m+1}(r))]}{d[\ln(r)]} = \frac{d[\ln(C_m(r))]}{d[\ln(r)]} + \frac{r\exp(-r^2/4\sigma^2)}{\sigma\sqrt{\pi}\,erf(r/2\sigma)}$$

where σ is the standard deviation of the Gaussian distribution, i.e., the noise level of the time series and *erf* refers to the error function.

Correlation integral….

- In Eq. 6, ϕ is a constant and M is Kummer's confluent hypergeometric function which has the following integral representation:

Eq. 7:

$$M(a,b,z)=\frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)}\int_0^1 e^{zt}t^{a-1}(1-t)^{b-a-1}\,dt$$

Correlation integral....

- Using Eq. (5) and (7), the correlation dimension and noise level for the time series can be estimated by a non-linear least squares method, at least in theory. However, because of the strong non-linearity in the equations, it is difficult in practice.

Noise level estimation

A relationship linking the correlation dimension D, the correlation sum $C_m(r)$, and the noise level σ can be shown to be

$$D + 2\left[m\frac{d[\ln(C_m(r))]}{d[\ln(r)]} - \frac{d^2[\ln(C_m(r))]}{d[\ln(r)]^2} - \left(\frac{d[\ln(C_m(r))]}{d[\ln(r)]}\right)^2\right]\frac{\sigma^2}{r^2} = \frac{d[\ln(C_m(r))]}{d[\ln(r)]}$$

In this equation (Eq.8), the correlation dimension is linear with respect to σ^2 . By substituting $\sigma = 0$, Eq. 8 simplifies to

$$D = \frac{d[\ln C_m(r)]}{d[\ln r]}$$

and, for noisy data, as $r \to 0$, as

$$\lim_{r \to 0} \frac{d[\ln C_m(r)]}{d[\ln r]} = m$$

Noise level estimation....

- Eq. 8 can be written in the form:

$$y = D + 2\sigma^2 x$$

where

$$y = \frac{d[\ln C_m(r)]}{d[\ln r]}$$

$$x = \frac{1}{r^2}\left[m\frac{d[\ln C_m(r)]}{d[\ln r]} - \frac{d^2[\ln C_m(r)]}{(d[\ln r])^2} - \left(\frac{d[\ln C_m(r)]}{d[\ln r]}\right)^2\right]$$

Noise level estimation….

$$y_n \approx r_n \frac{c_n - c_{n-1}}{\Delta r}$$

$$x_n \approx \frac{(m-1)\Delta r(c_n - c_{n-1}) - r_n(c_{n-1} - 2c_n + c_{n+1}) - r_n(c_n - c_{n-1})^2}{r_n(\Delta r)^2}$$

$$r_{n+1} - r_n = \Delta r$$

$$c_n = \ln C_m(r_n)$$

r_n, $(1 \le n \le L)$ given radius, and

L, the number of test values of the radius r.

Noise level estimation....

The least squares estimates of the noise level and the correlation dimension can then be shown to be

$$\overline{\sigma}^2 = \frac{\sum_{n=2}^{L-2}(y_{n+1} - y_n)(x_{n+1} - x_n)}{2\sum_{n=2}^{L-2}(x_{n+1} - x_n)^2}$$

$$\overline{D} = \frac{1}{L-2}\sum_{n=2}^{L-1}(y_n - 2\overline{\sigma}^2 x_n)$$

Applications (data known to be chaotic)

- Henon map: $$x_{n+1} = 1 - ax_n^2 + bx_{n-1} + n_n$$
- Lorenz map:

$$\begin{cases} \dfrac{dx}{dt} = \sigma(y - x) \\ \dfrac{dy}{dt} = -xz + rx - y \\ \dfrac{dz}{dt} = xy - bz \end{cases}$$

- Duffing equation:

$$\begin{cases} \dfrac{dx}{dt} = y \\ \dfrac{dy}{dt} = -ax - x^3 + b\cos(t) \end{cases}$$

Applications (data known to be chaotic)....

- Rossler equation:

$$\begin{cases} \dfrac{dx}{dt} = -y - z \\ \dfrac{dy}{dt} = x + ay \\ \dfrac{dz}{dt} = b + z(x - c) \end{cases}$$

- Chua's circuit:

$$\begin{cases} \dfrac{dx}{dt} = (1 - K - \varepsilon_r K)y - (1 + \varepsilon_r)x + \varepsilon_r z \\ \dfrac{dy}{dt} = x + (K - 2)y \\ \varepsilon_c \dfrac{dz}{dt} = \varepsilon_r (x + Ky - z) - \alpha_1 z - \alpha_2 (|z + 1| - |z - 1|) \end{cases}$$

Added and actual noise

- Although the noise levels (σ) added to the artificial data sets generated by the above systems are known, the actual noise levels for the noisy data would be somewhat different. In this paper, the actual noise level is calculated as follows:

$$\sigma_{Actual} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(s_i - \bar{s}_i)^2} \qquad (19)$$

where N is the sample number, and s_i and $\bar{s}_i$ are the noisy and clean data respectively.

Applications (data known to be chaotic)….

- Standard errors between estimated and added (or actual)noise for the data sets

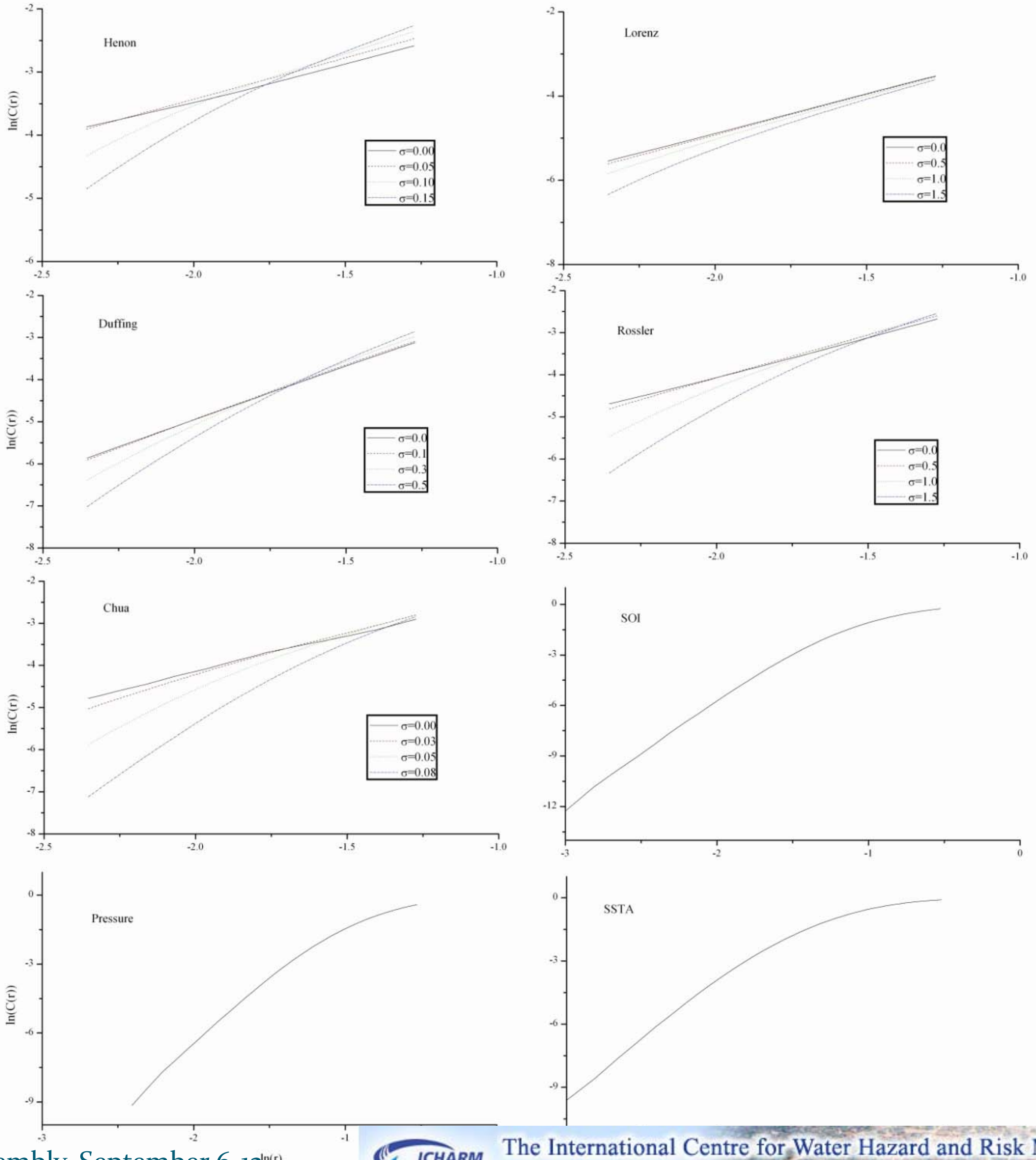
Data set	Added noise		Actual noise	
	Proposed method	Schreiber's method	Proposed method	Schreiber's method
Hénon	0.01042	0.05941	0.01059	0.05986
Lorenz	0.2248	1.6044	0.2207	1.6020
Duffing	0.07908	0.1724	0.07797	0.1732
Rossler	0.02566	0.4533	0.02550	0.4577
Chua	0.0006442	0.03162	0.0007906	0.03179

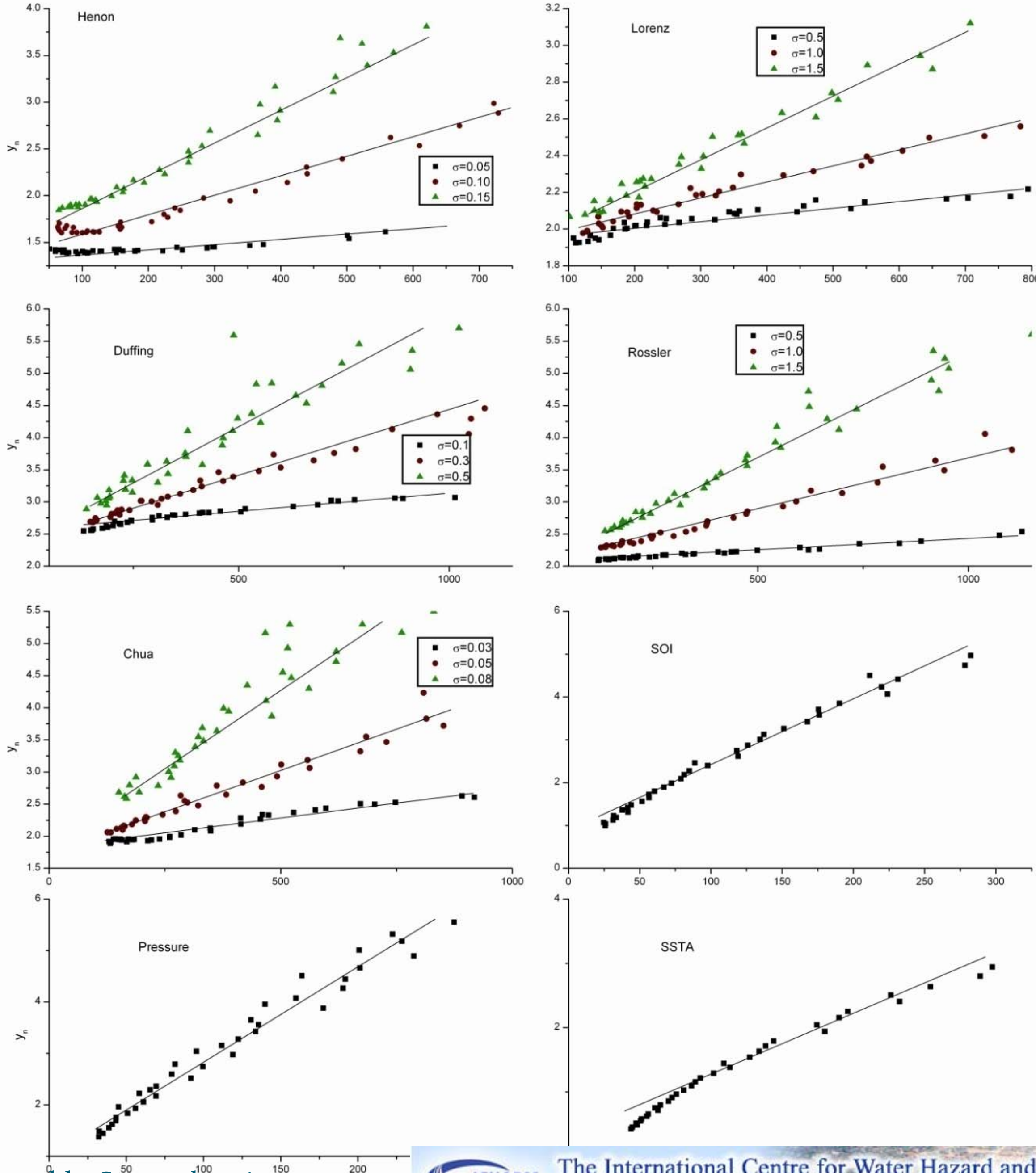
Application to real data

- Southern oscillation index (SOI) which is a normalized Darwin-Tahiti sea level pressure difference (January 1876 to December 2006)
- Eastern equatorial Pacific sea surface temperature anomaly (SSTA) index (January 1893 to December 1998)
- Normalized Darwin-Tahiti sea level pressure difference (January 1876 to December 1998)

Application to real data….

- The noise levels for the SOI, sea surface temperature anomaly index and the sea level pressure differences are respectively 1.21, 3.13 and 1.26, and the corresponding correlation dimensions are 0.8015, 0.4575 and 0.9926.
- Although these results cannot be verified, it is reasonable to expect them to be acceptable since the method has been extensively verified using several examples.


Henon
σ=0.00
σ=0.05
σ=0.10
σ=0.15
Lorenz
σ=0.0
σ=0.5
σ=1.0
σ=1.5
Duffing
σ=0.0
σ=0.1
σ=0.3
σ=0.5
Rossler
σ=0.0
σ=0.5
σ=1.0
σ=1.5
Chua
σ=0.00
σ=0.03
σ=0.05
σ=0.08
SOI
Pressure
SSTA
ln(C(r))
ln(r)



Henon
σ=0.05
σ=0.10
σ=0.15
Lorenz
σ=0.5
σ=1.0
σ=1.5
Duffing
σ=0.1
σ=0.3
σ=0.5
Rossler
σ=0.5
σ=1.0
σ=1.5
Chua
σ=0.03
σ=0.05
σ=0.08
SOI
Pressure
SSTA

Concluding remarks

- In this study, a method of estimating the noise level present in a chaotic time series is proposed by employing the linear least squares method. This is an improvement over previous methods of estimating the noise level all of which use the nonlinear least squares method.
- In the present method, a linear form connecting the correlation sum, the noise level and the correlation dimension is obtained. It is easier to apply and is expected to lead to less computing error compared to a nonlinear method.
- The method is verified using some artificial chaotic time series generated by Hénon map, Lorenz equation, Duffing's equation, Rossler equation and Chua's equation with added Gaussian noise. The numerical results consistently show that the proposed method give better estimates of the noise level for these chaotic time series than those obtained by the nonlinear method introduced by Schreiber (1993).
- The application part includes noise level estimations of monthly SOI, monthly eastern equatorial Pacific sea surface temperature anomaly index, and normalized monthly Darwin-Tahiti sea level pressure differences.

Thank you for your attention!

Reference

- Jayawardena, A. W., Xu, P.C., and Li, W. K. (2009): *A method of estimating the noise level in a chaotic time series,* Chaos, American Institute of Physics, DOI:10.1063/1.2903757 (published online on May 13, 2009)