ANALYSIS OF PROCESS DYNAMICS WITH MONTE CARLO SINGULAR SPECTRUM ANALYSIS

C. Aldrich and M. Barkhuizen

Department of Process Engineering, University of Stellenbosch, Stellenbosch, Private Bag X1, Matieland, 7602, South Africa
Fax +27(21)8082059, E-mail: ca1@sun.ac.za

Abstract: Singular spectrum analysis of time series observations can be visualized as a sliding window of width m moving down the time series **x** of length n, to determine the orthogonal patterns that best capture the variance presented by the window. It is designed to extract information from short and noisy time series. This allows the time series to be decomposed into various components, ideally separating components containing information from those that contain mostly noise. In this paper, Monte Carlo singular spectrum analysis is used to separate the components of time series data obtained from several real world processes into signal and noise components. The methodology facilitates better understanding of the underlying dynamics of complex systems, as well as the development of automated plant systems.
Keywords: Singular Spectrum Analysis, Time Series, Electrochemical Noise.

1. INTRODUCTION

It is well known that reliable and effective process control, diagnostics of system dynamics, troubleshooting and real-time monitoring of assets is vital for the efficient and competitive operation of any process. Singular spectrum analysis (SSA) is a relatively new technique that has been developed initially in the field of climatology (Broomhead and King, 1986, Vautard and Ghil, 1989, Vautard et al., 1992), but has since been successfully expanded and applied in a variety of other areas, among which are the biosciences (Schreiber, 2000, Mineva and Popivanov, 1996), geology (Rozynski et al., 2001, Schoellhamer, 2001), economics (Kepenne, 1995, Ormerod and Campbell, 1997) and solar physics (Kepenne, 1995). In simple terms, the technique can be visualized as a sliding window of width m moving down a time series **x** of length n, to determine the orthogonal patterns that best capture the variance presented by the window. With this approach, the time series can be decomposed into various components, allowing qualitative process dynamics to be separated from noisy data (Broomhead and King, 1986; Fraedrich, 1986; Vautard and Ghil, 1989; Vautard et al., 1992).

Monte Carlo singular spectrum analysis (MC-SSA) is a methodology for discriminating between the various components of the time series, particularly between components containing meaningful information and other components containing mostly noise. Although so-called white noise (additive measurement noise) is relatively easy to detect and remove, the situation becomes more complicated when the noise also drives the system, such as is the case in autoregressive moving average processes. In this paper, application of the method in the analysis of process dynamics is considered and the paper is organized as follows. The essential methodology of SSA is first introduced in section 2, followed by several case studies in sections 3 an and 4, and conclusions and recommendations in section 5.

2. METHODOLOGY

2.1 Basic SSA.

The basic methodology of SSA consists of four main steps, as indicated in Figure 1 (Golyandina et al., 2001). The time series is first embedded in a high-dimensional lagged trajectory matrix (step 1), which is then decomposed by means of singular value decomposition into the sum of a number of bi-orthogonal matrices of rank one (step 2). These two steps represent the decomposition stage.

During the reconstruction stage, the components represented by the bi-orthogonal matrices obtained in step 2 are grouped appropriately (step 3), and finally,

in step 4, the time series components representing the various groups can be reconstructed from the matrices formed in step 3.

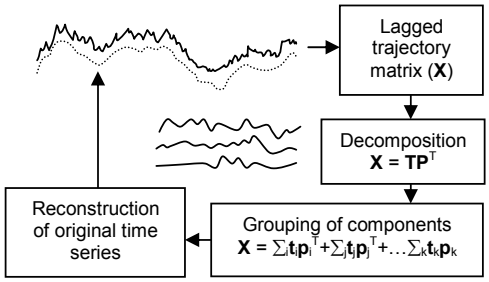

Fig. 1. Four basic steps of SSA, namely embedding of time series, decomposition by use of singular value decomposition, grouping of components and reconstruction of additive components.

More formally, SSA is based on calculating the principal directions in the lagged trajectory matrix in the phase space. An M x M covariance matrix $\mathbf{C}_X$ can be estimated directly from the data and the eigenelements λ_k and ρ_k (k = 1, 2, … M) of $\mathbf{C}_X$ are obtained from the solution of

$$\mathbf{C}_X\rho_k = \lambda_k\rho_k$$

The trace of $\mathbf{C}_X$ gives the total variance of the original time series and the eigenvalue λ_k equals the partial variance in the direction ρk. Equivalently an M x M matrix $\mathbf{E}_X$ can be formed, with columns ρ_k and the diagonal matrix $\mathbf{\Lambda}_X$ with eigenvalues λ_k as elements in decreasing order, so that

$$\mathbf{E}_X^T\mathbf{C}_X\mathbf{E}_X = \mathbf{\Lambda}_X$$

Grouping of the components of the time series is based on an analysis of the eigenspectrum $\mathbf{\Lambda}_X$.

2.2 Monte Carlo SSA.

Monte Carlo SSA (MC-SSA) is an extension of basic SSA and is used to gain a better understanding of the time series or some of its components by testing various hypotheses about the data. This is done by using a discriminating statistic, such as the autocorrelation, the correlation dimension or some other statistic to compare the time series under investigation with that of a number simulated realizations of stochastic processes derived from the original time series. In the nonlinear dynamics literature, such realizations are often referred to as surrogate data.

As a trivial example, consider a time series $\mathbf{x}(t)$, t = 1, 2, 3 … N, suspected to be white noise. To test this hypothesis by means of Monte Carlo singular spectrum analysis, the discriminating statistic could be the autocorrelation of the time series at a certain lag, say $\tau = 1$, or the predictability of the time series by some model, etc. Surrogate data are subsequently generated from the data, by randomizing the time series, with each surrogate data set representing particular instance of the randomized time series, i.e. $\mathbf{x}_s^j(t)$, with t = 1, 2, ... N and j = 1, 2, .. M, the number of surrogate time series.

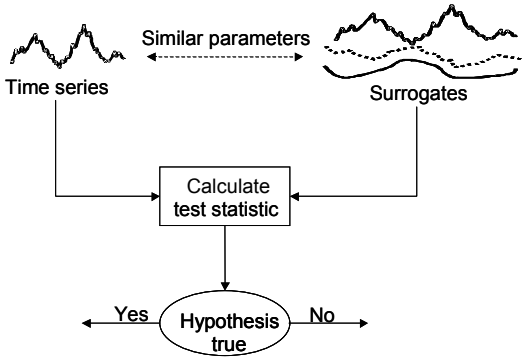

Fig. 2. Four basic steps of SSA, namely embedding of time series, decomposition by use of PCA or SVD, grouping of components and reconstruction of additive components.

The discriminating statistic is then calculated for the original time series (e.g. the one-step ahead predictability of the time series), as well as for each surrogate time series. If the predictabilities of the original time series and its surrogates do not differ, it has to be concluded that the original time series is indistinguishable from its surrogates and may well be white noise. The idea is illustrated in Figure 2.

More formally, the covariance matrix $\mathbf{C}_R$ of the surrogate data are projected onto the eigenvector basis $\mathbf{E}_X$ of the original time series, so that

$$\mathbf{\Lambda}_R = \mathbf{E}_X^T\mathbf{C}_R\mathbf{E}_X$$

The matrix $\mathbf{\Lambda}_R$ is generally not diagonal, since ΛR is not the result of singular value decomposition of the surrogate data set, but is a measure of the similarity of the given surrogate data set with the original time series. The degree of similarity can be quantified by computing the statistics of the diagonal elements of $\mathbf{\Lambda}_R$ via Monte Carlo simulation.

3. SULPHURIC ACID CONCENTRATION IN A LEACH CIRCUIT

3.1 Background

The leaching of a valuable metal on an industrial plant is controlled by addition of acid to a series of leaching tanks. Manual dosage of the acid by an operator is complicated by the large residence time of the ore in the vessels, so that the effects of over- or underdosage are only discovered after the fact. A better understanding of the dynamics of the metal and the acid concentration could therefore lead to large improvement in the control of the leaching process by means of a model-based control system.

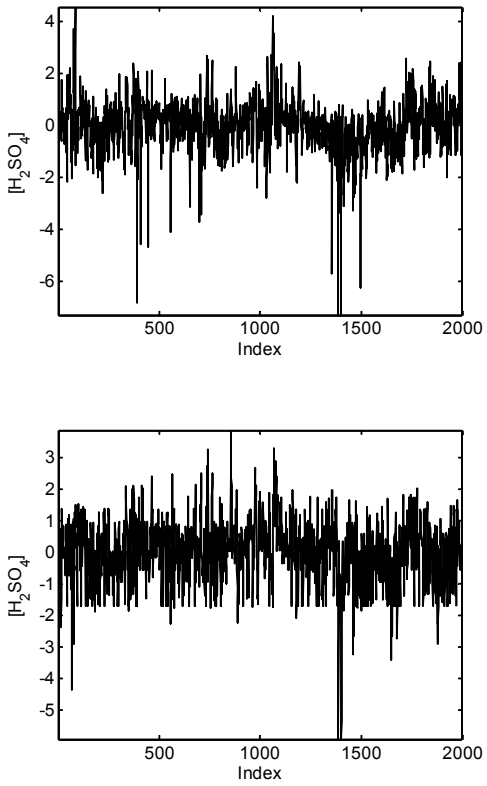

Fig. 3. Sample of twice-daily observations of scaled sulphuric acid concentration in the anolyte (top) and feed (bottom) of an industrial leach plant.

The data in Figure 3 shows 2000 twice-daily observations of the (scaled) concentrations of the H_2SO_4 in the anolyte (solid line), while the broken line shows the acid concentration in the feed. Figure 4 shows the confidence limits for an SSA eigenspectrum, given the time series, where the null hypothesis tested was that the data have been generated by a deterministic signal composed of the 1st eigenvector, contaminated by a 1st order autoregressive process, AR(1), with unknown mean. The parameters for the AR(1) process were estimated from the data, after removal of the signal, and then used to generate realizations of AR(1) noise. The covariance matrix was computed after adding the signal to these surrogate realizations, while the eigenspectra were computed by projection onto the basis **E**. These surrogate eigenspectra were subsequently used to construct the 90% and 95% confidence interval shown in Figure 4.

As can be seen from Figure 4, the null hypothesis appears to be valid in the case of the acid concentration in the feed (bottom), but may not be acceptable in the case of the acid in the anolyte. This information can be used to considerable advantage in the modelling and control of the process, especially in the case of the acid in the anolyte.

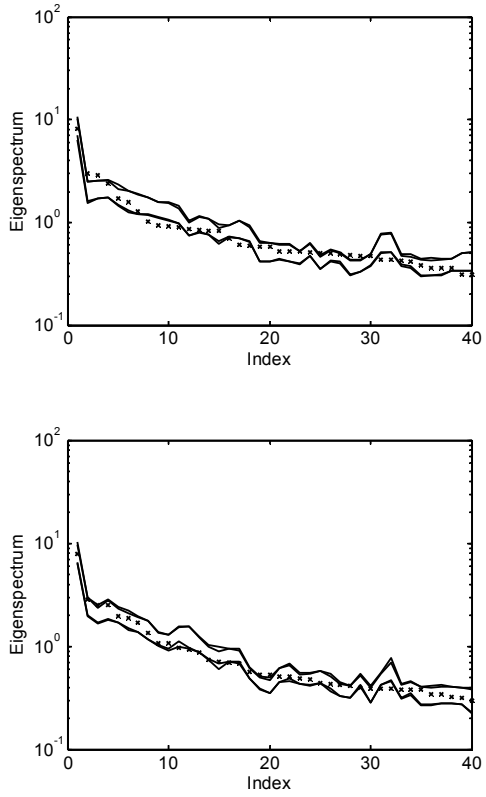

Fig. 4. Sample of twice-daily observations of scaled sulphuric acid concentration in the anolyte (top) and feed (bottom) of an industrial leach plant.

4. MODELLING OF MILLING CIRCUIT ON A GOLD PLANT

The grinding circuit on the Leeudoorn gold mine in South Africa (capacity approximately 8 ton per annum) consists of two Polysius 5 m x 11 m single-stage semi-autogenous grinding mills, each equipped with Siemens variable speed gearless ring motor drives. Classification is accomplished by two-stage hydrocyclones, with a final product size of 80% passing 75 μm. The circuit has a Polysius stacker and reclaimer system with a total live stock pile capacity of 100 000 tons. The grinding circuit is followed by conventional thickening, cyanide leaching and carbon-in-pulp (CIP) circuits. The instrumentation of the circuit includes a belt weigher to measure the mill feed rate, sump level measurement, flow rates, density and pressure measurement on the cyclone feed streams, mill power, mass (via load cells) and speed measurements. The final product size is measured by means of an Outokumpu particle size indicator.

Figure 5 shows the scaled power draw of the semi-autogenous (SAG) mill, which consisted of 1140 hourly measurements. These data were embedded in a lagged trajectory matrix with 40 columns, which was decomposed to give a singular spectrum as indicated in Figure 6. Surrogate data representing red noise (1st order autoregressive process) were projected onto the eigenbasis of the original data to yield the 90% and 95% confidence limits as shown on Figure 6.

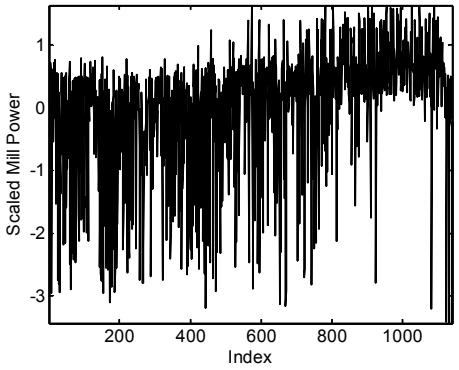

Fig. 5. Scaled power consumption of the SAG mill.

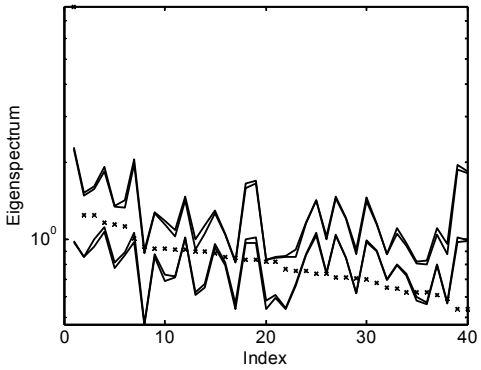

Fig. 6. Eigenspectra of the data shown in Fig. 5, as well as surrogate data generated from these data.

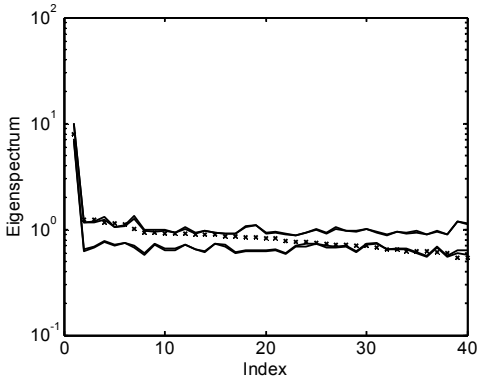

Fig. 7. Eigenspectra of the data shown in Fig. 5, as well as surrogate data generated from these data, after removal of the signal component of the data, associated with the 1st eigenvalue shown in Fig. 6.

The most striking feature of Figure 6, is the location of the first eigenvalue, which lies far above the confidence limits of the AR(1) process. This suggests that the mill power draw has characteristic dynamics, which may be contaminated with correlated noise.

Figure 7 shows that this is indeed the case. In this figure, the process dynamics (signal) associated with the first eigenvalue has been separated from the remainder of the components and as before the parameters for the AR(1) process were estimated from the data, after removal of the signal. The residual data were used to generate realizations of AR(1) noise, the signal was added back to the surrogate data before the covariance matrix was computed. The eigenspectra of the surrogate data were computed by projection onto the basis **E**. These surrogate eigenspectra were subsequently used to construct the 90% and 95% confidence interval shown in Figure 6. As can be seen from Figure 7, the eigenspectrum of the time series without the signal is confined within the confidence limits of the AR(1) process. Therefore, the mill power draw can be decomposed into a signal component and noise, which follows a 1st order autoregressive process. The signal component is shown in Figure 8, while the AR(1) noise component is shown in Figure 9.

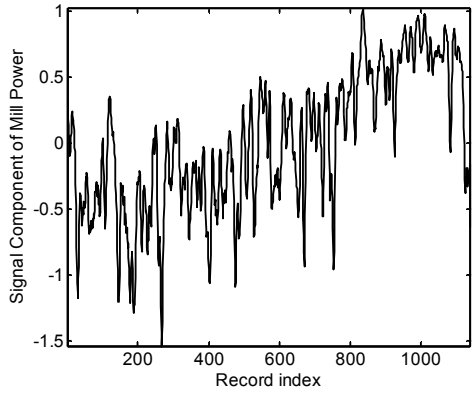

Fig. 8. Signal component of mill power consumption.

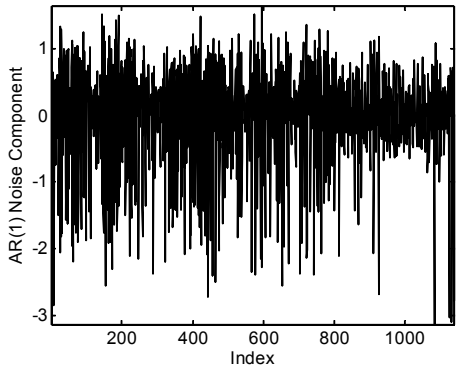

Fig. 9. AR(1) noise component of mill power consumption.

The attractors of the signal and noise components of the mill power draw are shown in Figures 10 and 11 respectively. Note the smoother attractor of the signal, as opposed to the more spiky appearance of the attractor of the noise component, as would be expected of a signal more predictable than the noise.

A more in-depth analysis of the two components is possible by further decomposition, but this analysis is beyond the scope of the paper.

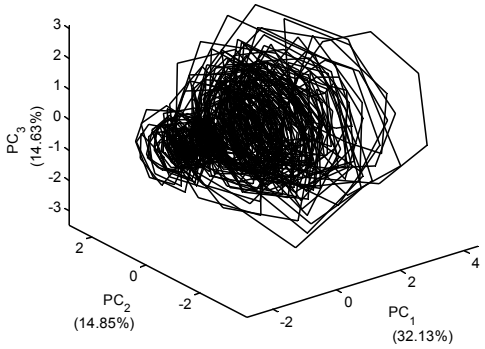

Fig. 10. Attractor of signal component of mill power consumption.

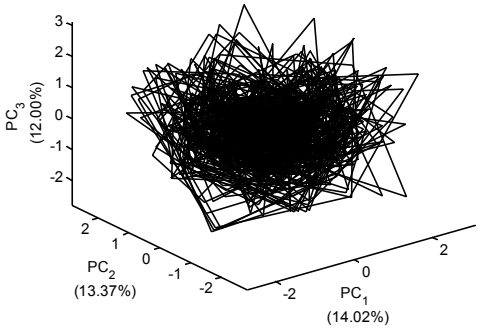

Fig. 11. Attractor of AR(1) noise component of mill power consumption.

5. ELECTROCHEMICAL NOISE DATA

5.1 Background

In the final example, data generated by a nonlinear deterministic process are considered. The data were obtained from a laboratory experiment, and represent electrochemical noise resulting from the corrosion of 304 stainless steel in hydrochloric acid (De Wet, 2001). The reasoning behind the measurement of the electrochemical noise properties was that the corrosion of metals is an electrochemical phenomenon and therefore these parameters can be used to provide an estimate of the corrosive process.

The material used was stainless steel 304 and both the electrochemical current noise and the electrochemical potential noise were measured simultaneously, by using a zero resistance ammeter and a high impedance voltmeter respectively. In order to measure the potential and current noise simultaneously and accurately, a three-electrode sensor was used. The current was then measured between two of the sensor elements and while the potential was measured between the third element, used as a reference, and the two coupled elements. The sampling rate for the measured time series was 0.432s, which led to the collection of 3156 observations. A sample of the observed time series from the current noise is displayed in Figure 12.

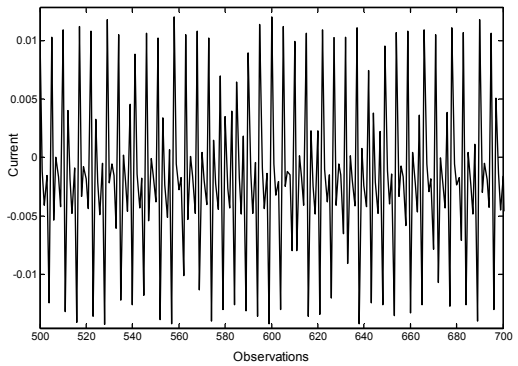

Fig. 12. Typical sample of electrochemical current noise measurements.

5.2 Classification of data

Monte Carlo SSA was used to classify the data as discussed above. The confidence limits for the eigenspectra of 1st order autoregressive processes, along with the eigenspectrum of the electrochemical current noise series, are shown in Figure 13.

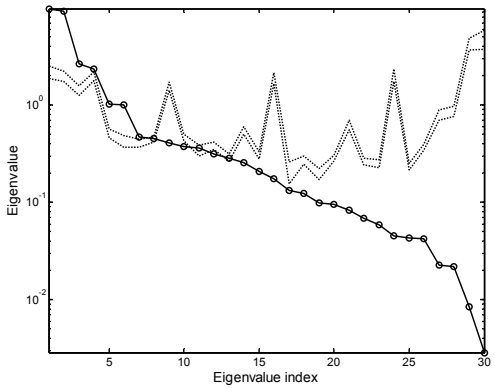

Fig. 13. Eigenspectrum of electrochemical current noise series along with 95% confidence limits generated from 15 first order autoregressive surrogate series.

From this figure, it is clear that the eigenspectrum of electrochemical noise process falls outside the confidence bands for the eigenspectrum of a first order autoregressive process, and that the data cannot be treated as an autoregressive process. Although not shown here in detail, separation of the components into signal and noise were not successful and it appears as if the signal contains little noise, as suggested by the accuracy with which the process could be modelled, as discussed in more detail below.

5.3 Modelling of time series

The accuracy of the classification by Monte Carlo SSA was investigated by fitting nonlinear one-step ahead models to the data. Figure 14 illustrates the accuracy of a multilayer perceptron neural network, as well as a regression tree. An AR(1) model is included for comparative purposes. The neural network has ten inputs with a lag of unity, while the

regrssion tree had one such input only, so the performance of the models is not comparable.

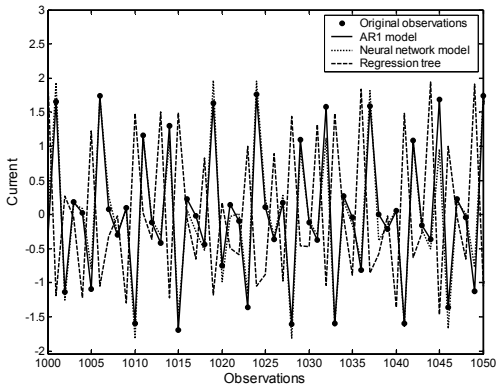

Fig. 14. Modelling predictions of the electrochemical current noise series (solid markers) by a 1st order autoregressive model (solid line), multilayer perceptron neural network (dotted line) and regression tree (dashed line).

Figure 15 shows the free-run prediction by the neural network model. With this approach, predicted values are used as input for future predictions. Small errors in prediction accumulate and can lead to rapid deterioration in the performance of the model. This is not the case in this instance and it appears as if the neural network has managed to capture the underlying dynamics of the process with a very high degree of accuracy, suggesting a highly deterministic process with little noise.

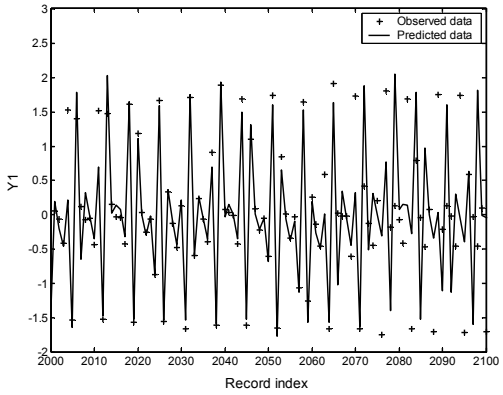

Fig. 15. Free-run prediction of electrochemical current noise data using a multi-layer perceptron neural network.

6. CONCLUSIONS

The analysis of dynamic process data provides crucial information in the automation of plants and process systems. By decomposing time series data into various components, a better understanding of the process is possible. Separation of the time series observation into signal and noise components is particularly important and in this paper the approach was demonstrated via several case studies on real world data. Although univariate time series were considered only, the approach is readily extendable to multivariate systems, as well as cross-spectral analysis of the data and further work in this regard is currently being pursued by the authors.

ACKNOWLEDGEMENTS

The authors are grateful for the financial support provided by Mintek, in particular the bursary afforded to Ms Barkhuizen.

REFERENCES

Broomhead, D. S. and King, G. P., (1986), *Physica D,* **20,** 217-236.

Golyandini et al. 2001

Elsner and Tsonis, 2000.

Kepenne, C. L., (1995), *Journal of Climate,* **8,** 1685-1689.

Mineva, A. and Popivanov, D., (1996), *Journal of Neuroscience Methods,* **68,** 91-99.

Ormerod, P. and Campbell, M., (1997), In *System dynamics in economic and financial models* (Eds, Heij, C., Schuacher, J. m., Hanzon, B. and Praagman, C.) John Wiley.

Rozynski, G., Larson, M. and Pruszak, Z., (2001), *Coastal Engineering,* **43,** 41-58.

Schoellhamer, D. H., (2001), *Geophysical Research Letters,* **28,** 3187-3190.

Schreiber, T., (2000), In *Chaos in Brain?* (Eds, Lehnertz, K., Elger, C. E., Arnhold, J. and Grassberger, P.) World Scientific, Singapore.

Vautard, R. and Ghil, M., (1989), *Physica D: Nonlinear Phenomena,* **35,** 395-424.

Vautard, R., Yiou, P. and Ghil, M., (1992), *Physica D: Nonlinear Phenomena,* **58,** 95-126.