

Chaotic signal processing by use of second order statistical methods and surrogate data analysis

C. Aldrich

Department of Chemical Engineering, University of Stellenbosch, Stellenbosch,
Private Bag X1, Matieland, 7602, Stellenbosch, South Africa, Fax +27(21) 808 2059
CA1@maties.sun.ac.za

Abstract

The cleaning of signals contaminated by noise is a major concern in real world systems, where short noisy signals are frequently encountered. In linear analysis the problem can be dealt with by extracting sharp narrowband linear signals from broadband noise in the Fourier domain, but this cannot be used for nonlinear signals, since nonlinear structures can be difficult to distinguish from broadband noise. Under these circumstances, it is better to attempt to differentiate between the signal and the noise in the time domain, by assuming that the observed signal $s(t)$, is the sum of the desired signal $s_1(t)$ and some other signals $s_2(t), s_3(t) \dots s_m(t)$. In this paper, the use of singular spectrum analysis and related methods to this end is investigated. The signal $s(t)$ is decomposed and the constituent signals are characterized by Monte Carlo simulations in which surrogate signals are generated which can serve as a benchmark for the detection and removal of noise from the original signal.

1. Introduction

Current interest in nonlinear signal processing is driven not so much via an extension of linear analysis, but from the recognition that new concepts in chaos and information theory can play a significant role in the reinterpretation of conventional time series analysis, e.g. by providing deterministic explanations for broadband spectra. Chaotic signal processing can arise among other from harmonic distortions and inter-modulation in amplifiers, faulty behaviour in rotating machinery, electrochemical signals in corrosion sensors, avalanching in granulated systems, various contexts in speech and audio, as well as image processing and communications.

Although much progress has been made with regard to the theory and application of chaotic signal processing, major challenges still need to be resolved as far as real world systems are concerned. The reason for this is two-fold. First, most of the established methodology has been designed to deal with long, noise-free signals only and second, it is often assumed that signals are generated by purely deterministic, low-dimensional systems. At present it is not clear to what extent the theory remains valid when strict determinism is lacking.

As a consequence, the cleaning of signals contaminated by noise is a major concern in real world systems. In linear analysis the problem can be dealt with by extracting sharp narrowband linear signals from broadband noise in the Fourier domain. This cannot be used for nonlinear signals, since nonlinear structures can be very difficult to distinguish from broadband noise. Under these circumstances, it is better to attempt to differentiate between the signal and the noise in the

time domain, by assuming that the observed signal $s(t)$, is the sum of the desired signal $s_1(t)$ and some other signals $s_2(t), s_3(t) \dots s_m(t)$. Therefore, a property distinguishing $s_1(t)$ from the other signals needs to be identified, such as that $s_1(t)$ satisfies some dynamical rule in reconstructed phase space, different from the dynamical rules associated with $s_2(t), s_3(t) \dots s_m(t)$.

Unfortunately, these dynamical rules are often not known in advance, so that assumptions with regard to the signals have to be made. This can be done by identifying the deterministic part of the observed signal $s(t)$ in its proper (low) embedding dimension, so that the other (noisy) parts of the signal would appear to be nondeterministic in this embedding, owing to their higher dimensionalities.

Previous investigations in this area have considered two general strategies. The first concerns the construction of local polynomial maps using neighbourhood to neighbourhood information and using this to determine local maps with some kind of averaging of domains over state space. These averaged dynamics are then used to realign individual observations to a better deterministic map [1].

The second approach is based on the use of local or global linear filters, such as moving averages or FIR filters, to clean up the data [2-3]. In this strategy, no special knowledge of the dynamics is assumed. Rather, it is assumed that a separation in the second order statistics, such as the principal and independent components of the sample will occur between the different components of the signal. For example, it is assumed that the noise will dominate the smaller eigenvalues and associated eigenvectors extracted from the lagged sample covariance matrix, while the signal $s_1(t)$ will dominate the numerically larger eigenvalues and eigenvectors.

Despite these advances, much work still needs to be done. For example, in the latter approach, one has to make drastic assumptions with regard to the nature of the noise. When these assumptions are valid, the methods work rather well, but otherwise they may break down completely, such as when the noise is assumed to be white and Gaussian and it is red and Poisson distributed [4].

In this paper, an approach similar to the latter strategy above is explored, by using the effect of singular spectrum analysis and variants thereof on the correlation dimension of the signals as a guide towards noise reduction in the signals. The strategy differs from those proposed previously, in that the distinction between the deterministic and stochastic components of the signal $s(t)$ is based on the dynamic structure of the data on various scales. The paper is organized as follows. The relevant theory on singular spectrum analysis is summarized in Section 2, while the characterization of time series in reconstructed state space is discussed in Section 3. In

Sections 4 and 5 two cases studies are considered and the results and conclusions are summarized in Section 6.

2. Singular spectrum analysis and other 2nd order statistical methods

The term singular spectrum comes from the spectral (eigenvalue) decomposition of a matrix \mathbf{A} into its spectrum (set) of eigenvalues. The eigenvalues λ are the values that make the matrix $\mathbf{A} - \lambda \mathbf{I}$ singular. Actually, the term singular spectrum is somewhat unfortunate in the context of the analysis of time series, since the traditional eigenvalue decomposition of matrices representing multivariate data is also an analysis of the singular spectrum. The spectral decomposition of matrices has only recently been applied to time series analysis and has had its roots mostly in the application of chaos theory [5-7].

In essence, the data are embedded in a very high-dimensional reconstruction, followed by the introduction of a new coordinate system, where the origin is moved to the centroid of the reconstructed system states and the axes are represented by the dominant principal components of the states (points).

The idea is explained by way of a simple example. Consider a uniformly sampled univariate time series with $n = 5$ observations $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5]^T$. With an embedding dimension of $m = 2$, there will be $n - (m - 1) = 4$ snapshots of the time series, i.e. $y_1 = [x_1, x_2]^T$, $y_2 = [x_2, x_3]^T$, $y_3 = [x_3, x_4]^T$ and $y_4 = [x_4, x_5]^T$. These snapshots can be arranged as row vectors, as shown below.

$$\mathbf{X} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} x_1 & x_1 \\ x_2 & x_2 \\ x_3 & x_3 \\ x_4 & x_4 \end{bmatrix} \quad (1)$$

It is not necessary to use successive values of the observations. For example, with every second observation only two new variables would have been created, $z_1 = [x_1, x_3]^T$, $z_2 = [x_3, x_5]^T$. The construction of the matrix \mathbf{X} is based on an (m, J) window, i.e. m indicates the number of time series observations in each new snapshot and J indicates the sample times between observed elements. In the first case, a $(3, 1)$ -window was used, while a $(3, 2)$ -window was used in the second case. If $J = 1$ is used, then reference is made to an m -window only. Matrix \mathbf{X} is referred to as the augmented or trajectory matrix and contains a complete record of patterns that have occurred in a window of size m , i.e. $X_{ij} \in \mathcal{R}^{n-m+1 \times m}$. Since $X_{(i-1)j} = X_{ij(-1)}$ for all $i > 1$ and $j > 1$, the columns of the matrix are highly correlated. The embedding space $\mathcal{K} \in \mathcal{R}^m$ is the space of all m -element patterns. Rather than investigating the trajectory matrix \mathbf{X} for repetitive patterns in the original time series, it is more effective to consider the *lagged covariance matrix* computed from the trajectory matrix \mathbf{X} and its transpose, i.e. $\mathbf{S} = \mathbf{X}^T \mathbf{X}$.

For time series with very peculiar patterns, such as series consisting of short sharp spikes, delay embedding may lead to very inhomogeneous sets of points in the reconstructed state space, that may be difficult to analyze. Under these circumstances, embedding can be based on the pattern features of the time series, e.g. the time intervals between spikes.

More formally, basic singular spectrum analysis can be seen as a four-step process, as indicated in Figure 1 [8,9]. In the first step (embedding), the one-dimensional time series is recast as an L -dimensional time series (trajectory matrix). In the second step (singular value decomposition), the trajectory matrix is decomposed into a sum of bi-orthogonal matrices of rank one. These two steps constitute the decomposition stage of singular spectrum analysis.

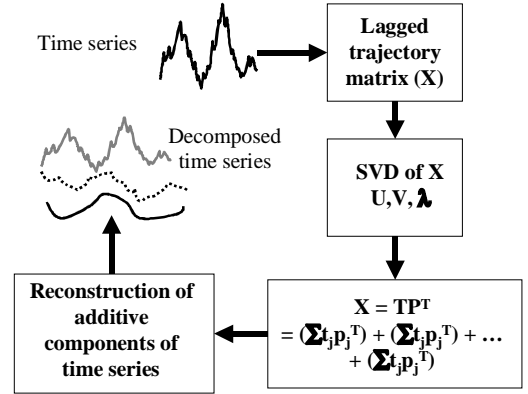


Figure 1: Decomposition and reconstruction of a time series by use of singular spectrum analysis.

The third step concerns splitting of the matrices computed in step 2 into several groups. The trajectory matrix is thus represented as the sum of several resultant matrices. In the final step, a time series is reconstructed for each resultant matrix by means of diagonal averaging [8]. The original time series is thus decomposed into a series of additive components, which can be characterized by use of surrogate data analysis, the theory of which is summarized below.

3. Surrogate data analysis

The method of surrogate data [10-13] involves a null hypothesis against which the data are tested, as well as a discriminating statistic. The data are first assumed to belong to a specific class of dynamic processes. Surrogate data are subsequently generated from this process and various statistics are calculated from both the surrogate and the original data [10]. If the calculated statistics of the surrogate and the original data are different, then the null hypothesis that the process that has generated the original data is of the same class as the system that has generated the surrogate data is rejected.

More specifically, let $\mathbf{x} \in \mathcal{R}^n$ be a time series consisting of n observations, ψ a specific hypothesis, \mathcal{S}_ψ the set of process systems consistent with the hypothesis, and $T: \mathcal{R}^N \rightarrow \mathcal{U}$ be a statistic that will be used to evaluate the hypothesis ψ that \mathbf{x} was generated by some process $\mathcal{S} \in \mathcal{S}_\psi$. Generally the statistic $U \in \mathcal{U}$, and it will be possible to discriminate between the original data \mathbf{x} and the surrogate data \mathbf{x}_s consistent with the hypothesis given by the probability density of T , given \mathcal{S} , i. e. $p_{T, \mathcal{S}}(t)$.

3.1. Pivotal test statistics

Theiler [14] has suggested that a distinction can be made between so-called pivotal and non-pivotal statistics. A test

statistic T is considered to be pivotal, if the probability distribution $p_{T,\mathfrak{S}}$ is the same for all processes \mathfrak{S} consistent with the hypothesis ψ . Moreover, a distinction can be made between simple and composite hypotheses. If the set of all processes consistent with the hypothesis (\mathfrak{S}_ψ) is a singleton, then the hypothesis is simple. Otherwise, the hypothesis is composite and can be used not only to generate surrogate data consistent with a particular process \mathfrak{S} , but also to estimate $\mathfrak{S} \in \mathfrak{S}_\psi$. In fact, \mathfrak{S} has to be specified when the hypothesis is composite, unless T is a pivotal statistic [14], meaning that $p_{T,\mathfrak{S}}$ is the same for all $\mathfrak{S} \in \mathfrak{S}_\psi$.

Constrained realization [15] schemes can be employed when non-pivotal statistics are applied to composite hypotheses. That is, apart from generating surrogate data that represent typical realizations of a model of the system, the surrogate data should also be representative of a process yielding identical estimates of the parameters of the process compared to the estimates of the process parameters obtained from the original data. Put in a different way, if $\mathfrak{S}_{est} \in \mathfrak{S}_\psi$ is the process estimated from the original data \mathbf{x} , and \mathbf{x}_s is a surrogate data set generated by $\mathfrak{S}' \in \mathfrak{S}_\psi$, then \mathbf{x}_s is a constrained realization of $\mathfrak{S}_{est} \in \mathfrak{S}'$.

As an example, if ψ is the hypothesis that \mathbf{x} is generated by linearly filtered independent identically distributed noise, then *non-constrained* surrogate data \mathbf{x}_s' can be generated from a Monte Carlo simulation based on the best linear model estimated from \mathbf{x} . The data \mathbf{x}_s' can be constrained by shuffling the phases of the Fourier transform of the data, producing a set of random data \mathbf{x}_s'' with the same power spectra (and autocorrelation), rank order statistics, nonlinear prediction error, etc., would all be non-pivotal test statistics characterizing dynamic manifold structures, since the distributions of these statistics would all depend on the form of the noise source and the type of linear filter. In contrast, the Lyapunov exponents and the correlation dimension (fractal dimension) would be pivotal test statistics, since the probability distributions of these quantities would be the same for all processes, regardless of the source of the noise or the estimated model. Since recent investigations have shown that Lyapunov exponents can be misleading in the presence of noise, the correlation dimension has gained favour as the pivotal statistic of choice.

3.2. Classes of hypotheses

Three classes of hypotheses are used widely, viz. those equivalent to the assumption that the data are identically, independently distributed noise (0), linearly filtered noise (1) and a monotonic nonlinear transformation of linearly filtered noise (2), generated by randomizing the data (0), randomizing or shuffling the phases of the Fourier transform of the data (1) and randomizing the phases of amplitude adjusted Gaussian noise (2).

In the latter case (hypotheses of class 2), the procedure for generating surrogate data consists of the following steps

- i. Generation of an identical independently distributed Gaussian data set \mathbf{y} , and reordering of \mathbf{y} to have the same rank distribution as \mathbf{x} .
- ii. Generation of a surrogate data set \mathbf{y}_s of \mathbf{y} (by shuffling of the phases of the Fourier transform of the data set \mathbf{y})

- iii. Finally, reordering of the original data set \mathbf{x} , to generate a surrogate data set \mathbf{x}_s , which has the same rank distribution as \mathbf{y}_s .

Surrogates generated by these procedures are referred to as algorithm 0, algorithm 1 and algorithm 2 surrogates, while algorithm 2 surrogates are also known as amplitude adjusted Fourier transform (AAFT) surrogates [16].

In order to assess the nature of the data, the surrogates and the original time series are displayed in bivariate plots of the correlation dimension of the data sets against the scale. No rigorous statistical tests are as yet available to discriminate between the original and the surrogate data sets and evaluation takes place by means of inspection. These calculations are outlined below by means of a simple example.

4. Noise reduction in a linear Gaussian process

In the 1st case study, a simple linear signal is considered in order to demonstrate the methodology. The signal is generated by the function $y = \sin(\theta) + \sin(\theta/2) + \epsilon$, with $\theta \in [0, 20\pi]$ and $\epsilon \sim N(0, 0.25)$ representing Gaussian noise with zero mean and standard deviation of 0.25 added to the curve.

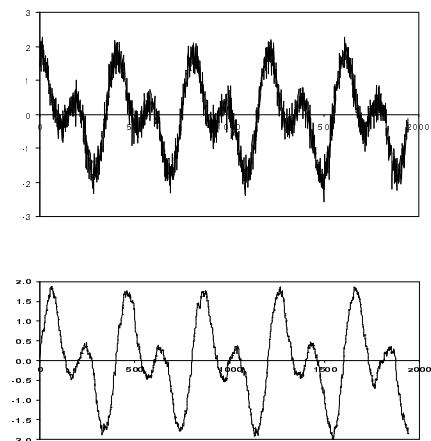


Figure 2: Noisy signal (top) and reconstructed signal (bottom) with reduced noise.

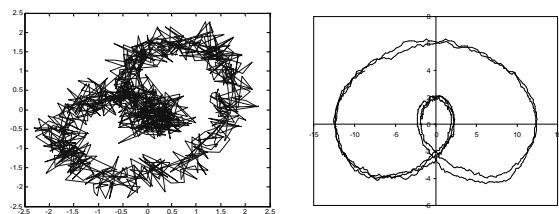


Figure 3: Reconstructed attractors of the respective systems shown in Figure 1, i.e. original signal at the top and reconstructed signal at the bottom.

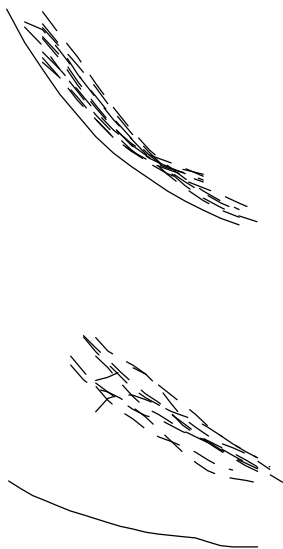
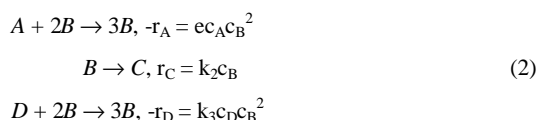


Figure 4: Surrogate data of the respective systems shown in Figure 1, i.e. the original signal at the top and the same signal after smoothing at the bottom.

The original signal and the smoothed signal reconstructed from the first two principal components extracted from the lagged covariance matrix of the signal are shown in Figure 2. The reconstructed attractors of the respective signals are shown in Figure 3, while the corresponding correlation dimension curves vs scale are shown in Figure 4. The curves in Figure 4 show the stochastic nature of the original signal, i.e. the correlation dimension curve of the original signal (solid line) lies very close to the correlation dimension curves of the surrogate signals (broken lines). Smoothing via decomposition and reconstruction of the signal yield a lower correlation dimension, which differ markedly from the associated surrogate signals, especially at smaller scale values.

5. Autocatalysis in a continuous stirred tank reactor

The second case study concerns an autocatalytic process in a continuous stirred tank reactor originally considered by Gray and Scott [17] and subsequently investigated by Lynch [18]. The system is capable of producing self-sustained oscillations based on cubic autocatalysis with catalyst decay and proceeds mechanistically as follows.



where A , B , C and D are the participating chemical species and k_1 , k_2 and k_3 the rate constants for the chemical reactions. This process is represented by the following set of ordinary differential equations.

$$dX/dt = 1 - X - aXZ^2$$

$$dY/dt = 1 - Y - bYZ^2 \tag{3}$$

$$dZ/dt = 1 - (1+c)Z + daXZ^2 + fbYZ^2$$

where X , Y , and Z denote the dimensionless concentrations of species A , B and D , while a , b and c denote the Damköhler numbers for A , B and D respectively. The ratio of feed concentration of A to that of B is denoted by d and the same ratio of D to B by f . The process is chaotic, with a well-defined attractor for specific ranges of the two parameters, d and e . For the settings $a = 18000$; $b = 400$; $c = 80$; $d = 1.5$; $f = 4.2$, and initial conditions $[0, 0, 0]^T$, the set of equations was solved by using a 5th order Runge Kutta numerical method over 100 simulated seconds. This gave approximately 10 000 observations, which were resampled with a constant sampling period of 0.01 s. The Y state was taken as the output variable. Figure 5 shows the attractor of the process reconstructed from the process states X , Y and Z . Gaussian noise $N(0, \sigma)$ was added to these observations, with the standard deviation equal to 10% of the average of each signal.

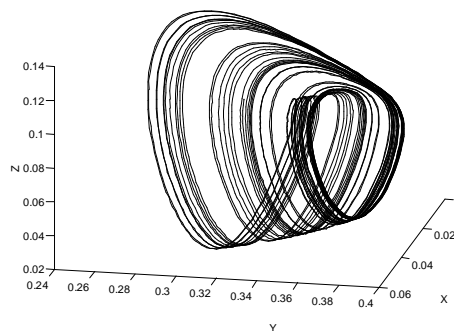


Figure 5: The attractor of the autocatalytic system without noise.

The noisy data could be reconstructed from the principal components of the 10-column lagged covariance matrix of the observations. The variance explained by each of the first six of these ten principal components is shown in Table 1. The cumulative variance explained by the first six principal components was 99.60%.

Table 1: Principal component analysis of the based on the lagged covariance matrix of the noisy autocatalytic data.

PC	1	2	3	4	5	6
Var(%)	66.3	31.0	2.2	0.46	0.15	0.11

The correlation dimension curves and associated amplitude adjusted surrogate data are shown in Figures 6 and 7. Note that the correlation dimension curves for the reconstructed signal components based on the first and second principal components are more or less identical. Successive signal components reconstructed from higher index principal components show increasingly higher correlation dimensions.

The 6th signal component appears to be entirely stochastic, which means that the signal can be reconstructed from the first 5 principal components, while the noise resides in all subsequent components. Note that classic criteria would have underestimated the number of principal components to be retained.

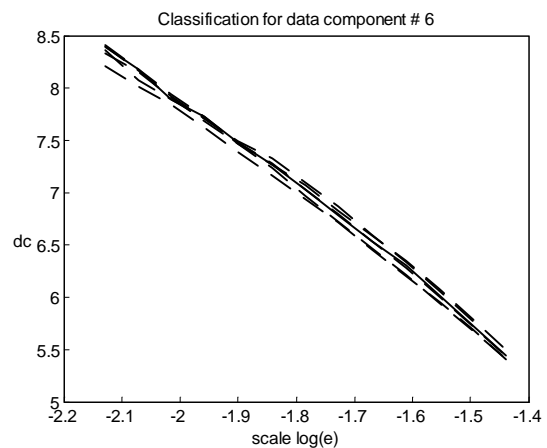
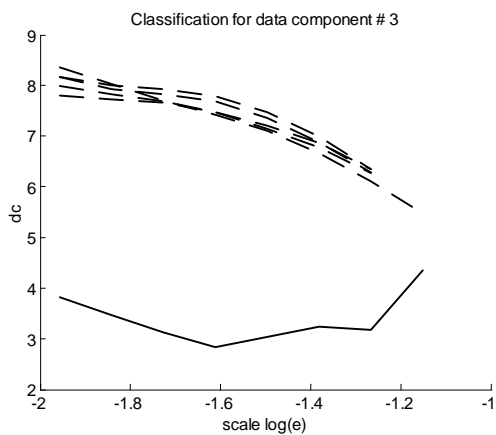
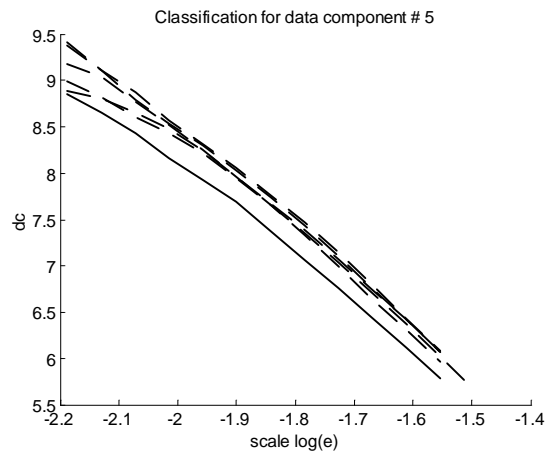
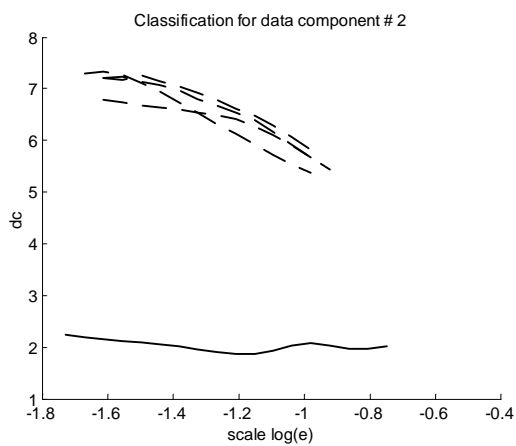
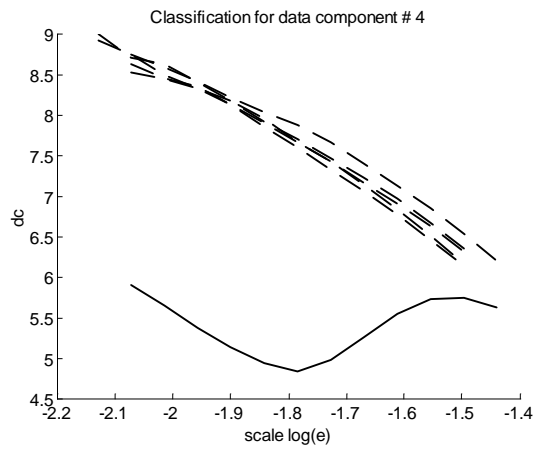
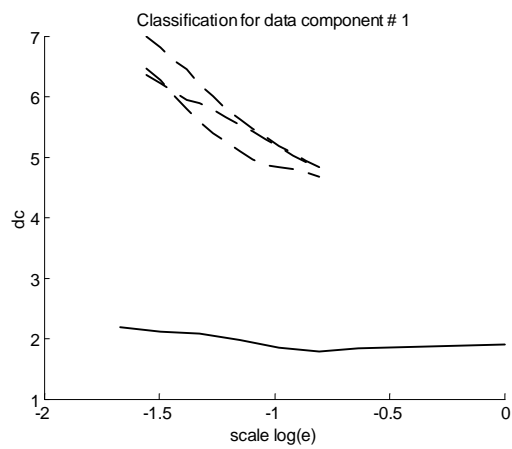


Figure 6: From top to bottom: Correlation dimension curves, (a) for reconstructed X-state of the noisy autocatalytic system (solid line) based on 1st, 2nd and 3rd principal components and their associated Type 2 surrogates (broken lines).

Figure 7: From top to bottom: Correlation dimension curves, (a) for reconstructed X-state of the noisy autocatalytic system (solid line) based on 4th, 5th and 6th principal components and their associated Type 2 surrogates (broken lines).

The Kaizer criterion (eigenvalues larger than unity for scaled data) would have suggested two only, while interpretation of a scree diagram would have suggested the retention of three principal components.

6. Summary and conclusions

This paper has considered the use of second order statistics and surrogate data analysis to reduce noise in observed signals, which may or may not be chaotic. By means of Monte Carlo simulations (surrogate data), the reconstructed components of the signal can be assessed systematically, which allows for efficient removal of noise in nonlinear systems. The approach is general, but its reliability depends on the reliability of the surrogate data that are used to test hypotheses regarding the different signal components. Although principal component (singular spectrum) analysis was used to decompose the signals, other methods such as independent component analysis can also be used, when the observations have a non-Gaussian distribution.

Although other experiments not discussed here have indicated that the approach outlined above is better suited to dealing with nonlinear chaotic data than other methods, it has certain drawbacks. The first is the high computational cost and the second concerns the surrogate data analyses.

Surrogate data methods are not infallible and can yield more false rejections of the null hypothesis when the data are strongly correlated [17]. Moreover, while it is true that Gaussian linear processes are fully specified by either the autocorrelation function or the power spectrum, only an *estimate* of the spectra is used with the analysis. This can give rise to additional high frequency elements in the spectrum. Another problem is that nonstationarity of the time series can also cause false positives. For example, spikes or changes in the process over time means the null hypothesis will be rejected. This can lead to a large class of formally correct rejections, which are interpreted in the wrong way [17].

Despite these disadvantages, the use of surrogate data analysis provides a more sophisticated basis for dealing with noisy signals, as well as the assessment of other constituent structures in the signals and future work will be conducted towards a broader classification of the signal components.

7. References

- [1] E. Kostelich and J.A. York, "Noise reduction - finding the simplest dynamical system consistent with the data", *Physica D*, vol. 41, pp. 183-196, 1990.
- [2] R. Cawley and G. Hsu, "Local-geometric projection method for noise reduction in chaotic maps and flows", *Physical Review A*, vol. 46, pp. 3057-3082, 1992.
- [3] J.A. Sauer, "Noise reduction - finding the simplest dynamical system consistent with the data", *Physica D*, vol. 41, pp. 183-196, 1989.
- [4] H.D.I. Abarbanel, "Analysis of Observed Chaotic Data", Springer Verlag, New York, NY, USA, 1996.
- [5] D.S. Broomhead, and G.P. King, "Extracting qualitative dynamics from experimental data", *Physica D*, vol. 20, pp. 217-236, 1986.
- [6] P.S. Landa, and M.G. Rosenblum, "Time series analysis for system identification and diagnostics", *Physica D*, vol. 48, pp. 232-254, 1991.
- [7] R. Vautard, P. Yiou, and M. Ghil, "Singular spectrum analysis: A toolkit for short, noisy and chaotic time series", *Physica D*, vol. 58, pp. 95-126, 1992.
- [8] N. Golyandina, V. Nekrutkin, and A. Zhiglavsky, *Analysis of Time Series Structure - SSA and Related Techniques*, Chapman and Hall/CRC Press, Boca Raton, FL, USA, 2001.
- [9] J.B. Elsner, and A.A. Tsonis, *Singular Spectrum Analysis: A New Tool in Time Series Analysis*, Plenum Press, New York, NY, USA, 1996.
- [10] J. Theiler, S. Eubank, A. Longtin, and B. Galdrikian, "Testing for non-linearity in time series: The method of surrogate data", *Physica D*, vol. 58, pp. 77-94, 1992.
- [11] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank and J.D. Farmer, "Using surrogate data to detect nonlinearity in time series" In: *Nonlinear Modeling and Forecasting*, eds, M. Casdagli and S. Eubank, SFI Studies in the Sciences of Complexity, vol. XII, pp. 163-188, Addison-Wesley, Reading, MA, USA, 1992.
- [12] F. Takens, "Detecting non-linearities in stationary time series", *International Journal of Bifurcations and Chaos*, vol. 3, pp. 241-256, 1993.
- [13] J. Theiler and D. Pritchard, "Constrained realization Monte Carlo method for hypothesis testing", *Physica A*, vol. 94, pp. 221-235, 1996.
- [14] J. Theiler, "On the evidence for low-dimensional chaos in an epileptic encephalogram", *Physics Letters A*, vol. 196, pp. 335-341, 1995.
- [15] T. Schreiber and A. Schmitz, "Improved surrogate data for non-linearity tests", *Physical Review Letters*, vol. 77, no. 4, pp. 635-638, 1996.
- [16] M. Small and K. Judd, "Correlation dimension: A pivotal statistic for non-constrained realizations of composite hypotheses in surrogate data analysis", *Physica D*, vol. 129, pp. 386-400, 1998.
- [17] P. Gray and S.K. Scott, "Autocatalytic reactions in the isothermal, continuous stirred tank reactor. Isolates and other forms of multi-stability", *Chemical Engineering Science*, vol. 38, pp. 29-43, 1983.
- [18] D.T. Lynch, "Chaotic behavior of reaction systems: parallel cubic autocatalators", *Chemical Engineering Science*, vol. 47, pp. 347-355, 1992.