ANALYSIS OF PROCESS DATA WITH SINGULAR SPECTRUM METHODS

by

Marlize Barkhuizen

Thesis presented in partial fulfilment of the requirements for the Degree

of

Masters of Science in Engineering
(Chemical Engineering)

in the Department of Process Engineering
at the University of Stellenbosch

Supervised by:
Prof C Aldrich

STELLENBOSCH
SOUTH AFRICA

DECEMBER 2003

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Marlize Barkhuizen October 2003

SYNOPSIS

The analysis of process data obtained from chemical and metallurgical engineering systems is a crucial aspect of the operating of any process, as information extracted from the data is used for control purposes, decision making and forecasting. Singular spectrum analysis (SSA) is a relatively new technique that can be used to decompose time series into their constituent components, after which a variety of further analyses can be applied to the data. The objectives of this study were to investigate the abilities of SSA regarding the filtering of data and the subsequent modelling of the filtered data, to explore the methods available to perform nonlinear SSA and finally to explore the possibilities of Monte Carlo SSA to characterize and identify process systems from observed time series data.

Although the literature indicated the widespread application of SSA in other research fields, no previous application of singular spectrum analysis to time series obtained from chemical engineering processes could be found.

SSA appeared to have a multitude of applications that could be of great benefit in the analysis of data from process systems. The first indication of this was in the filtering or noise-removal abilities of SSA. A number of case studies were filtered by various techniques related to SSA, after which a number of neural network modelling strategies were applied to the data. It was consistently found that the models built on data that have been prefiltered with SSA outperformed the other models.

The effectiveness of localized SSA and auto-associative neural networks in performing nonlinear SSA were compared. Both techniques succeeded in extracting a number of nonlinear components from the data that could not be identified from linear SSA. However, it was found that localized SSA was a more reliable approach, as the auto-associative neural networks would not train for some of the data or extracted nonsensical components for other series.

Lastly a number of time series were analysed using Monte Carlo SSA. It was found that, as is the case with all other characterization techniques, Monte Carlo SSA could not succeed in correctly classifying all the series investigated. For this reason several tests were used for the classification of the real process data.

In the light of these findings, it was concluded that singular spectrum analysis could be a valuable tool in the analysis of chemical and metallurgical process data.

OPSOMMING

Die analise van chemise en metallurgiese prosesdata wat verkry is vanaf chemiese of metallurgiese ingenieursstelsels is ‘n baie belangrike aspek in die bedryf van enige proses, aangesien die inligting wat van die data onttrek word vir prosesbeheer, besluitneming of die bou van prosesmodelle gebruik kan word. Singuliere spektrale analise is ‘n relatief nuwe tegniek wat gebruik kan word om tydreekse in hul onderliggende komponente te ontbind.

Die doelwitte van hierdie studie was om ‘n omvattende literatuuroorsig oor die ontwikkeling van die tegniek en die toepassing daarvan te doen, beide in die ingenieursindustrie en in ander navorsingsvelde, die navors van die moontlikhede van SSA aangaande die verwydering van geraas uit die data en die gevolglike modellering van die skoon data te ondersoek, ‘n ondersoek te doen na sommige van die beskikbare tegnieke vir nie-lineêre SSA en laastens ‘n studie te maak van die potensiaal van Monte Carlo SSA vir die karakterisering en identifikasie van data verkry vanaf prosesstelsels.

Ten spyte van aanduidings in die literatuur dat SSA wydverspreid toegepas word in ander navorsingsvelde, kon geen vorige toepassings gevind word van SSA op chemiese prosesse nie.

Dit wil voorkom asof die chemiese nywerhede groot baat kan vind by SSA van prosesdata. Die eerste aanduiding van hierdie voordele was in die vermoë van SSA om geraas te verwyder uit tydreekse. ‘n Aantal tipiese gevalle is ondersoek deur van verskeie benaderings tot SSA gebruik te maak. Nadat die geraas uit die tydreekse van die toetsgevalle verwyder is, is neurale netwerke gebruik om die prosesse te modelleer. Daar is herhaaldelik gevind dat die modelle wat gebou is op data wat eers deur SSA skoongemaak is, beter presteer as die wat slegs op die onverwerkte data gepas is.

Die effektiwiteit van lokale SSA en auto-assosiatiewe neurale netwerke om nie- lineêre SSA toe te pas is ook vergelyk. Albei tegnieke het daarin geslaag om nie- lineêre hoofkomponente van die data te onttrek wat nie geïdentifiseer kon word deur die lineêre benadering nie. Daar is egter gevind dat lokale SSA ‘n meer betroubare tegniek is, aangesien die auto-assosiatiewe neurale netwerke nie op sommige van die datastelle wou leer nie en vir ander tydreekse sinnelose hoofkomponente onttrek het.

Laastens is ‘n aantal tydreekse geanaliseer met behulp van Monte Carlo SSA. Soos met alle ander karakteriseringstegnieke, kon Monte Carlo SSA nie daarin slaag om al die tydreekse wat ondersoek is korrek te identifiseer nie. Om hierdie rede is ‘n kombinasie van toetse gebruik om die onbekende tydreekse te klassifiseer.

In die lig van al hierdie bevindinge, is die gevolgtrekking gemaak dat singuliere spektrale analise ‘n waardevolle hulpmiddel kan wees in die analise van chemiese en metallurgiese prosesdata.

This is not the end.
It is not even the beginning of the end.
But it is the end of the beginning.
Winston Churchill

ACKNOWLEDGEMENTS

My heavenly Father for giving me the abilities, strength and opportunities to come this far.

My study leader, Prof Chris Aldrich, for his patient guidance, continued support and invaluable input in my work.

My parents for believing in me and always having a word of support, encouragement and praise, as the need arose.

My friends for showing interest, giving advice and ‘just being there’, even though my work was largely Greek to them.

Fred for promising everything would turn out all right (it did).

Mintek for their continued financial support and interest in my work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

It is well known that reliable and effective process control, diagnostics of system dynamics, troubleshooting and real-time monitoring of assets are vital for the efficient and competitive operation of any process, with the chemical engineering industry being no exception to this. Current tendencies of companies and plants are to increasingly enlarge their capacities, both to increase their turnover in response to an ever-increasing consumer and customer demand and to enlarge their profit margins by benefiting from the cost savings associated with economy of scale. This enlargement of capacities in their turn implies a number of adjustments in operating procedures, of which the automation of control programs is quite a prominent one.

Improved control of a process by implementing automated control will not only directly impact factors such as the recovery or extraction grades under variable feed conditions, but will also compensate for operator induced disturbances in the processes thereby optimise the efficiencies of the process in general. Once the dynamic control of a production process is improved, operators can then dedicate all their attention to more value-added control of other assets.

However, the desired cost savings and improved operation would not necessarily be attained by installing any expensive control system purchased. One should rather ensure that the techniques and methods implemented would be relatively easy to integrate into the process and must be maintainable within the framework of the available engineering resources of the plant.

It is well known that the control of chemical engineering plants and processes is no simple matter, due to a number of factors. These include that

- there are many conflicting and opposing goals to be satisfied
- the processes are very often nonlinear and multi-variable in nature
- the dynamics behind the processes tend to be complex, making it harder to understand
- many processes can be described as chaotic in that variables bounce around the set-point chaotically
- fundamental or first principal models are more often than not unavailable for application to mineral processing systems

It is especially this last problem, the lack of fundamental models on which to build control systems, that introduces the need to turn to other methods to aid with the control of mineral processing systems. Due to the nature of processes and the way they are operated, the one commodity that is available in abundance in any mineral processing system is data. A single time series recorded from the output of any dynamical system, be it physical, biological, socio-economic or chemical, is the result of the combination of all the interacting variables in the process. Therefore, in principle, this single record could contain information about the dynamics of all the important variables involved in the process and the evolution of the system under consideration. The idea behind singular spectrum analysis is to exploit this inherent information in the time series and to determine some of the system's key properties by quantifying specific features of the time series.

The information extracted from the data could then serve as a platform which is used to learn more about the process and to optimize operating conditions, as well as to aid with decisions on the operations management and even the business management levels. The challenge lies therefore in finding the appropriate tools to extract this information from the data and applying this newfound insight then to model-based control systems.

Singular spectrum analysis is a relatively new technique that has been developed initially in the climatology research field, but has since been successfully expanded and applied in a variety of research fields, among which the biosciences, geology, economics and solar physics are but a few. It appears that the only significant application from which SSA has been absent, is the analysis of data obtained from process plants. However, this absence may be due to oversight, as SSA performs a number of functions that are of direct interest and advantage to the analysis of data from process plants.

The basic idea behind singular spectrum analysis is that it is a tool that embeds either a single or multivariate time series into a higher dimensional matrix, which is then decomposed into a set of base functions or constituent components.

The first major advantage that SSA holds is therefore the decomposition of the time series into the various components that constitute the basis of the time series. These components can be investigated in turn to identify major trends in the data, remove components that can be classified as pure noise and extract oscillatory components present in the data. It is especially this ability of SSA to distinguish noise components from that of trend signals that is of great interest, as that can be applied in the filtering of data, which is desirable for a great number of reasons such as data presentation, modelling and so forth.

By application of SSA to the time series, one's ability to detect change points is also greatly improved and by using refinements of the technique, it is also possible to characterize the data as being linear or nonlinear, stochastic or deterministic and so forth.

The purpose of SSA is not inherently to identify or build any particular model of the time series investigated. It is rather to provide information on the deterministic and stochastic parts of behaviour in the data, even when the time series is short and noisy (Ormerod and Campbell, 1997)

All these properties are very desirable in the analysis of time series data obtained specifically from process plants, therefore justifying the application of SSA also in the chemical engineering and metallurgical processing fields.

The objectives of this study on singular spectrum analysis can be summarized as follows:

- To perform a literature survey which investigated the methodology behind singular spectrum analysis, the available modifications to basic singular spectrum analysis and previous applications of singular spectrum analysis in both other research fields and the engineering industry.
- The application of singular spectrum analysis in the filtering of data and the evaluation of the effectiveness of this filtering by building neural network models on the data.
- To explore nonlinear singular spectrum analysis and evaluate the relevance of the different techniques for nonlinear singular spectrum analysis in application to chemical engineering process systems.
- To explore the practical applications of Monte Carlo singular spectrum analysis in the identification and characterization of time series from process systems.

The rest of this work will be structured along the following topics: Firstly, in chapter 2, a general background on the development of SSA and its application in various research fields by other researchers will be discussed. Due to the concepts behind SSA being largely unfamiliar in the chemical engineering industry, any technical discussions about the technique or applications thereof, will be left out the literature review in chapter 2, but rather discussed in the methodology in chapter 3. Except for a basic discussion about the SSA technique, the algorithms for a number of advancements, such as multivariate SSA, Monte Carlo SSA and nonlinear SSA will also be described in chapter 3.

The next four chapters will be devoted to various case studies in which different approaches to SSA will be illustrated. The most basic approach is that of chapter 4, where either univariate or multivariate SSA is applied to a time series in order to remove noise from the series after which the series is subsequently modelled. This will be done for both theoretical case studies and real process data, while comparing the results obtained from various modelling techniques.

The basic approach of chapter 4 leads on to a more complicated scenario in chapter 5, where nonlinear processes and data are being analysed by using two different approaches to nonlinear SSA. Once again the nonlinear application of SSA is applied to both a theoretical, simulated time series and real process data.

The last two chapters, chapters 6 and 7 are both concerned with the characterisation of time series by using Monte Carlo singular spectrum analysis. Chapter 6 is used as a benchmarking chapter, in that the results obtained from the Monte Carlo SSA are verified with time series with known properties. Once the reliability of the approach has been established, Monte Carlo SSA will be used in chapter 7 to characterise some data obtained from chemical engineering processes.

2 LITERATURE REVIEW

Singular spectrum analysis is a relatively new technique, developed from methods that are largely unfamiliar to the engineering and specifically the mineral processing industry. It was decided to refrain from providing a detailed literature review of the technique in this section, since the basic methodology have not yet been explained and consequently a discussion on the development might just serve to confuse the reader. Therefore, beside a brief mention of the origins of SSA and the apparent advantages of SSA compared to other spectral analysis techniques, this section is rather devoted to an investigation into the previous applications of singular spectrum analysis, both in the field of engineering and in other fields of research. A discussion on the development and refinement of singular spectrum analysis as technique will then be given in the appropriate methodology sections in chapter 3.

2.1 Origins of singular spectrum analysis

Singular spectrum analysis was developed simultaneously and independently by Broomhead and King (1986) and Fraedrich (1986). Broomhead and King (1986) applied singular spectrum analysis to the problems of dynamical systems theory and the singular spectrum approach to the method of delays was suggested to remove some of the limitations and ambiguities experienced with the method of delays. In their article they laid the mathematical basis used for singular spectrum analysis by combining PCA or SVD and embedding theorems. They also investigated some preliminary artificial time series to illustrate the advantages of using singular spectrum analysis as a statistical tool for qualitative analysis and for the removal of especially white noise from time series.

Fraedrich (1986) used observed weather and climate variables to provide information for descriptions of the properties of the attractors of these dynamical systems and to obtain an estimate of the smallest number of variables necessary to explain the system dynamics. Further groundbreaking work in the methodological development of the singular spectrum analysis toolkit and substantial research on the possibilities of the technique, was done by Robert Vautard and Michael Ghil. Vautard and Ghil (1989) extended the previous research done by Broomhead and King (1986) and refined certain aspects of the application, which will be discussed in more detail in later sections. After applying SSA to various paleoclimatic time series, they found the technique to be very flexible and incisive. They concluded that, even though SSA is related to ordinary spectral analysis, it is considerably more robust to the nonstationarities that can be found in climatic records.

Vautard et al. (1992) distinguish among three major cases encountered when performing data analysis. The first is where the evolution equations governing the data are known and these equations are relatively insensitive to the initial values of the system. The second type of data analysis occurs when the governing equations are also known, but long-term prediction of the data is impossible due to the sensitivity of the system for the initial values. The last class of data analysis is where the evolution equations for the system are completely unknown and often only noisy measurements of one of the variables in a high-dimensional system are available. It is especially with respect to this last class of data that Vautard et al. (1992) identified the potential of SSA. Even though the work focussed on single-channel SSA, they already saw the possibilities of multi-channel SSA to account for the cross-correlation between several variables that were measured simultaneously. A detailed discussion on the application of multichannel SSA (MSSA) will be given in the methodology section, but for now it will suffice to explain that a number of different variables, all relating to the same phenomena, are analysed simultaneously by using singular spectrum analysis. The purpose of this is to exploit the relationships between variables reflected in the measured time series.

Two very comprehensive works have recently been written on SSA by Elsner and Tsonis (1996) and by Golyandina et al. (2001). In both these works, the necessary mathematical background is supplied, various approaches to the theory and methodology are discussed and all the possible applications of SSA are investigated. More attention will be given to these works in the methodology section.

2.2 SSA compared to other spectral time series analysis techniques

It was found that, in principle, most processes can be characterised as a function of frequency, rather than of time. This frequency is known as either the power spectrum or the spectral density. Very irregular motions, such as noise, will have a smooth and continuous spectrum, as such a process excites all frequencies in a given band. This is contrasted by a pure periodic signal where the series can be described by one specific frequency or a limited number of frequencies (Ghil and Yiou, 1996). The challenge is therefore to determine the underlying power spectrum for real time series which lie somewhere between the two extremes just mentioned. A number of spectral analysis techniques have been developed and this section will be applied to a very brief description of some of the other techniques that can be used in the place of, or in conjunction with, SSA. Comprehensive comparisons between SSA and other spectral analysis techniques, can be found in (Ghil and Taricco, 1997) and (Ghil and Yiou, 1996).

2.2.1 Fourier analysis

Fourier analysis is similar to singular spectrum analysis, in that it also decomposes the time series into a set of base functions. However, in the case of Fourier analysis, these base functions are a linear combination of selected sine and cosine functions. These base functions are therefore fixed, making it hard to approximate localized disturbances, such as frequency pulses, in the time series, compared to the data-adaptive nature of the SSA base functions, as will be discussed later.

2.2.2 Wavelet analysis

Wavelet analysis is generally used as a basic tool for intermittent, complex and self-similar signals. The technique can be described as a mathematical microscope, in that the emphasis can be placed on a specific part of the time series and local structures and singularities can then be extracted from the small part investigated.

The basis of the technique is to reconstruct either the original time series, or a filtered version of it, by combining a family of wavelet transforms, of which the most common are sine functions convoluted with exponential functions. The wavelet basis used can be adapted to satisfy the specific requirements of the time series investigated.

2.2.3 Maximum Entropy Method (MEM)

The main benefit in using the maximum entropy method is to estimate the line frequencies for a time series that was generated by either a linear autoregressive process or an m^{th} order autoregressive process (Ghil and Yiou, 1996). The technique is performed by calculating one more autocorrelation coefficient from the time series as the order of the autoregressive model (m+1). The spectral density equivalent to the most random or least predictable process with the same autocorrelation coefficients can now be determined. If the time series being investigated is not stationary or close to autoregressive the results from MEM should preferably be verified by cross testing with other techniques. It was found by a number of researchers that the performance of MEM can be greatly improved by first applying SSA to the time series to enhance the signal to noise ratio.

2.2.4 Multi-taper method (MTM)

The estimate of the power spectrum provided by the multi-taper method is nonparametric, in that it does not require a specific, parameter dependent model of the process that had generated the time series (Ghil and Yiou, 1996). A set of tapers is used to reduce the variance of the spectral estimates. This is done by computing a set of independent estimates of the power spectrum from the pre-multiplication of the data with orthogonal tapers.
The specific advantage of MTM is its ability to detect low-amplitude oscillations in relatively short time series.

2.3 Trends analysis with SSA

Singular spectrum analysis has been successfully applied in a variety of disciplines, of which the most common is paleoclimatology and meteorology, with the biosciences, solar physics, economics and general engineering applications also showing a keen interest in the abilities of SSA. These findings from other research will now be discussed, according to the various fields in which SSA was applied.

2.3.1 Climatology and paleoclimatology

As was mentioned before, this is the area of research in which SSA has received the largest amount of attention, probably because this is also the field in which SSA was first applied to the investigation of time series. Even though these time series are not strictly related to the engineering field, there are many similarities between the natures of climatic and mineral processing time series, in that both time series tend to be relatively short with a noisy behaviour. One can therefore benefit substantially from studying the application of SSA to climatic time records.

The exploration of the effectiveness of analysing paleoclimatic time series by using SSA started simultaneously with the development of SSA in the work done by Michael Ghil and Robert Vautard in Vautard and Ghil (1989). SSA was used to describe the main physical phenomena reflected by the data (such as the periods of various oscillations observed in the data) and it was also used for adaptive spectral filters to remove the dominant oscillations of the system. When SSA was applied to the paleoclimatic time series, it also succeeded in clarifying the noise characteristics of the data. They concluded that SSA verified the need for simple nonlinear models by which the dynamic information contained in existing paleoclimatic records could be extracted and explained.

In some further work (Ghil and Vautard, 1991), SSA was applied to global temperature series with the intention of extracting global warming trends and oscillatory modes from the noise parts. The benefits of using different numbers of eigenvalues in the reconstruction of the time series were illustrated. Due to the success of their previous studies on SSA, the aim of this paper was rather to extract useful information from time series than to prove the validity of SSA. It was assumed that the benefit and relevance of SSA had already been proven.

In short succession to the article by Ghil and Vautard (1991), Elsner and Tsonis (1991) published an article also investigating the global temperature record by using SSA. This started an intriguing discussion (Allen et al., 1992b, Allen et al., 1992a, Tsonis and Elsner, 1992) on the results obtained and the conclusions derived from these results. However, the discussion focussed on technical aspects of SSA and will therefore rather be addressed in the methodology section.

Further studies focussing on oscillations in the global climate system were undertaken by Schlesinger and Ramankutty (1994). Instead of just analysing the observed global mean temperature changes, a model is used to simulate these temperature changes and the simulated values are then subtracted from the observed values. SSA was then applied to the detrended data, which revealed new oscillations not previously observed during the SSA performed on non-detrended data by Ghil and Vautard (1991), Elsner and Tsonis (1991) and Allen et al. (1992a). Their conclusion was that, when using SSA, any deterministic trends should first be removed from the data, to allow the first eigenvalues to explain dominant oscillations and not the variance explained by trends in the data.

The next step in the analysis of oscillations in weather patterns was the application of multichannel SSA (Plaut and Vautard, 1994). They used multichannel SSA to identify dynamically relevant space-time patterns and as an adaptive filtering technique.

A number of other papers were also written to explore the benefit of SSA in extracting oscillations from climatic time series, or to simply apply the technique and make conclusions on the time series from the oscillations that were extracted. In addition to those papers already mentioned, (Cortijo et al., 1995, Lall and Mann, 1995, Naidu and Malmgren, 1995, Yiou et al., 1995, Yiou et al., 1997, Melice and Rucou, 1998, Evans et al., 1999, Shun and Duffy, 1999, Dean et al., 2002, Pohjola et al., 2002) also applied SSA specifically to paleoclimatic time series obtained from ice cores, marine microfossils, corals and lakes. In all the studies, SSA was used to divide the time series into trends, oscillations and noise, from which the desired components were identified and extracted. In most of the cases, SSA was used in conjunction with other analysis techniques to verify or clarify the results obtained. In the situation where the results from SSA and other spectral techniques, specifically multitaper spectral analysis (MTM), varied slightly (Lall and Mann, 1995), these differences were attributed to the different window lengths or smoothing parameters used as well as the fact that SSA is time optimal and MTM frequency optimal. More detailed attention about the different spectral analysis techniques will be given at the end of this chapter. In the work done by Shun and Duffy (1999), attention was also given to multichannel SSA. However, the multichannel SSA was used in conjunction with single channel SSA and no comparison was therefore made about the effectiveness of the two approaches.

As it has been mentioned earlier, a substantial amount of research has been done where singular spectrum analysis was used to investigate time series relating to climatology (Allen and Smith, 1994, Corte-Real et al., 1995, Dettinger et al., 1995, Ghirardelli et al., 1995, Lall and Mann, 1995, Naidu and Malmgren, 1995, Plaut et al., 1995, Allen and Robertson, 1996, Solow and Patwardhan, 1996, Benzi et al., 1997, Zhang et al., 1997, Cook et al., 1998, Corte-Real et al., 1998, Robertson and Mechoso, 1998, Shabalova and Weber, 1998, Stahle et al., 1998, Zhang et al., 1998, Dickey et al., 1999, Elsner et al., 1999, Mo, 1999, Shun and Duffy, 1999, Vautard et al., 1999, Lee and Hang, 2000, Mo, 2000, Paegle et al., 2000, Lee, 2001, Masulli et al., 2001, Mo, 2001, Pederson et al., 2001, Prierto et al., 2001a, Prierto et al., 2001b, Ye, 2001, Ye and Cho, 2001, Yu and Mechoso, 2001, Rodo et al., 2002, Wainer and Venegas, 2002, Baratta et al., 2003, Krepper et al., 2003, Robertson and Mechoso, 2003), and these works are in addition to those works that have already been mentioned in this section. In the majority of these studies either univariate or multivariate (single channel or multichannel) SSA was used as a tool in conjunction with other spectral analysis techniques. The main aim behind the application of the technique was to extract trends or oscillations from the data, which were then related to occurrences in other climatological time series. Some exceptions to the application occurred (Allen and Robertson, 1996) where Monte Carlo SSA was used to distinguish modulated oscillations from red noise and (Masulli et al., 2001) where SSA was used for denoising of the time series to aid with forecasting, but the role that singular spectrum analysis played in most papers was relatively standard and one can only truly benefit from their varying viewpoints if one has the necessary background in climatology and meteorology. Although these works are all of great interest, they will therefore not be discussed in detail, but the interested reader is referred to them.

Benzi et al. (1997) applied SSA to an observed series of minimum and maximum temperatures and daily cumulative precipitation in the Sardinia region over a 42-year period. They tested the effectiveness of SSA as a technique to characterize the spatial and time frequency dependence of meteorological fields. Benzi et al. (1997) investigated cluster analysis on the local density maxima of principal components. In the meteorological field, the presence of a local high density of points shows that a typical climate exists or that spatial patterns recur. They described a technique developed by Molteni et al. (1990) by which to build homogeneous groups around the local density maxima in the phase space. Their approach included the characterization of seasons by using PCA and the mentioned cluster analysis technique and a spectrum analysis of minimum and maximum temperature fields by using both maximum entropy method and SSA.

Their study confirmed the suitability of SSA for the identification of significant climatological characteristics of a region and recognized that, by synthesizing the whole data set by a few representative components, the relevant characteristics of the signal from the data can be extracted effectively. They succeeded in determining the spatial patterns and time recurrences of the temperature and precipitation fields of the Sardinia region from spatially distributed data. These characteristics were not immediately recognizable from the data itself, but the results proved to be in accordance with the known general behaviour of the Sardinian climate.

Shabalova and Weber (1998) once again focused their research on temperature variability and other paleoclimatic issues. They did however experiment with a novel approach by first subjecting the time series to PCA and then using the spatial principal components as the input channels for the MSSA. The signals in the original time series were then computed by convolving the reconstructed components in PCs with the corresponding spatial modes. A number of independent tests were used to check the consistency of the reconstructed trend components and to identify the quasi-periods. Their results showed that similar results were obtained when the original time series was used directly as input channels and when the principal components were used as the input.

Three further areas of study in the group of publications on climatology that have been mentioned earlier that are worth citing specifically, is that of observing fluctuations in the hurricane frequency (Elsner et al., 1999), obtaining climatic information from tree-ring records (Cook et al., 1998, Stahle et al., 1998, Pederson et al., 2001, Gedalof et al., 2002, Pohjola et al., 2002, D'Arrigo et al., 2003) and observing a link between cholera and climatic changes (Pascual et al., 2000, Rodo et al., 2002).

Researchers used information obtained from tree ring chronologies to identify modes and oscillations in climate variabilities in various regions. (Cook et al., 1998, Stahle et al., 1998, Pederson et al., 2001, Gedalof et al., 2002, Pohjola et al., 2002, D'Arrigo et al., 2003). SSA was applied for the decomposition of the series and it was observed that oscillations that were present in the Southern Oscillation Index could successfully be extracted from the tree ring data (Stahle et al., 1998). An unconventional application of SSA was in (Cook et al., 1998), where SSA was used to examine the stability of observed oscillations. SSA was applied to both the full and truncated reconstructions of the time series. It was found that the principal components from both were in good agreement, indicating a high degree of homogeneity in the reconstruction at the specific periods.

Elsner et al. (1999) launched an investigation into the fluctuations in hurricane frequency in the North Atlantic region. The researchers combined SSA with the maximum entropy method (MEM) to obtain the leading modes of oscillation in the annual hurricane frequency.

In two papers (Pascual et al., 2000, Rodo et al., 2002) researchers investigated the relation between the El Nino-Southern Oscillation (ENSO) and the occurrence of cholera. In the first case study, they used data obtained over 18 years for the number of cholera cases reported each month in conjunction with sea surface temperatures (that provides an index for ENSO) for the same period. SSA was used to decompose both the time series and it was attempted to observe overlapping dominant frequencies between the two data series.

In the second study, two separate periods were studied, but instead of using the sea surface temperature data, the Southern Oscillation index was used to provide information about ENSO. The time series representing the cholera information was the percentage of the people that visited the clinic each month that suffered from cholera. SSA was used to isolate the main interannual variability in the data and also to compare the spectra of the two different periods in time.

A further field of research within the climatology framework that have been applying SSA to a number of time series, is that of measurements of the atmospheric temperature and pressure and hence the atmospheric circulation variability (Allen and Smith, 1994, Plaut and Vautard, 1994, Corte-Real et al., 1995, Zhang et al., 1997, Corte-Real et al., 1998, Ribera et al., 2000, Mo, 2001, Grinsted et al., 2003, Robertson and Mechoso, 2003).

An interesting work was that of Ghil and Yiou (1996) in which they gave a summary of what spectral methods can and cannot do for climatic time series. The work described the connections between time series analysis and nonlinear dynamics. They also focussed on signal-to-noise enhancement and presented some recently developed methods used for spectral analysis. The steps to follow for the various techniques, as well as the benefits and shortcomings of the techniques were illustrated by, once again, using a well known climatic time series. A further discussion of this paper will follow at the end of this chapter.

Another unusual application of SSA in the climatology field was that of Hollingsworth et al. (1997) where SSA, together with autoregressive models, was used to analyse a surface pressure time series from Mars and statistically significant spectral powers were isolated. An annual cycle simulation that corresponded to a low atmospheric dust loading, was performed by using the NASA Ames Mars general circulation model and seasonal variations of storm zones on Mars were identified. It was found that during certain seasons, localized storm zones occurred in certain areas, with the storm zones shifting into higher latitudes during other seasons. These variations in the storm zones during the seasonal cycle will have important implications for Mars' regional climate.

It can be seen that SSA has been applied by a great number of researchers to a variety of different time series relating to the meteorological and climatology fields. Even though some individual studies varied, the majority of applications of SSA to climatic time series aimed at extracting relevant trends and oscillations from the data. As it has been mentioned before, this research contains many similarities to process engineering applications, as time series obtained from both the climatology field and engineering processes tend to be relatively short and noisy, making it hard to analyse by using conventional techniques.

2.3.2 Biosciences

Singular spectrum analysis has also been applied with great success in the biological and medical research fields. One of the first studies about the advantages of SSA in the biosciences, and specifically neuroelectrical signals, was done by Mineva and Popivanov (1996). They investigated the identification of single-trial readiness by using a method based on SSA. The time series that was measured was the EEG (electroencephalogram) activity of a patient from a specific time before and until a certain period after a voluntary motor act was performed. This brain activity is known as the readiness potential of the person and indicates the preparation for the voluntary movement. The problem that faced the researchers was that this time series was also characterized as being short and noisy, making the usual techniques unsuitable. The aim of the paper was to extract certain parameters from the single-trail readiness potential and it was found that SSA separated the data records into various components, by which different dynamical stages of the movement preparatory process could be distinguished. They found that components that were hidden in the raw signal, appeared or disappeared around the onset of the readiness potential and these components were successfully revealed by SSA.

Further research on this subject was done by Popivanov et al. (1998), where they followed a combined linear and nonlinear approach. In this study it was pointed out that previous work done on EEG signal dynamics assumed linear dynamics and therefore used linear methods, such as SSA, while there existed no evidence that this type of analysis fully described the dynamics of the process. They first used two linear methods, namely SSA and time-frequency analysis, based on auto-regressive model coefficients. Four nonlinear techniques were also applied to test whether the linear techniques captured all the dynamics of the time series, and these techniques were point-wise dimension, Kolmogorov entropy, largest Lyapunov exponent and nonlinear prediction. Their results indicated that the transitions in the dynamics of the EEG activity prior to complicated voluntary activities were detected when using both linear and nonlinear characteristics. This lead to the questions of which approach is more appropriate to detect transitions in the dynamics of mental activity and how the alterations in the dynamical characteristics should be interpreted in the aspect of the mental activities that were involved. They found that due to the nature of mental processes, it was more likely that the nonlinear technique would be appropriate. They concluded that their present results did not provide any evidence that the dynamical changes that were detected reflected the mental activity involved in the voluntary movement preparation and recommended that further complex analysis were performed.

These problems were partly addressed by further research by the same authors in (Popivanov and Mineva, 1999). They pointed out once again that the majority of physiological signals, such as EEG, blood flow, human gait and ECG are characterized by complex dynamics, such as nonlinearities and nonstationarities and that it is important to be able to distinguish the characteristics of the process underlying the signal from the properties of the observed time series. Classical methods that could be used to determine possible nonlinear or chaotic dynamics are the correlation dimension, entropy analysis and Lyapunov exponents. However, these methods are not as reliable when only relatively short data series containing stochastic components and nonstationarities are available. They therefore developed several approaches that aim at determining the nonstationarities in the data and testing whether nonlinear dynamics exist.

Work in the field of applying SSA to time series of brain electrical activity was also done by Schreiber (2000). He investigated whether nonlinearity was evident in these time series, thereby investigating the first question asked by Popivanov et al. (1998). He admitted that it is unlikely that the brain functions in a linear manner, but pointed out that the nonlinear nature of the brain might not be evident in specific aspects of the brain's dynamics and one could therefore rather use the more familiar linear techniques. From his results it is seen that for many instances of brain signals one can observe a certain degree of nonlinearity. However, Schreiber also pointed out that the examples he had given, is not the same as nonlinear dynamical systems or chaos. One can therefore apply the nonlinear techniques developed from the chaos theory to great advantage, but some interpretations could also be largely misleading. In the cases where there is nonlinearity evident in the time series but the time series cannot be successfully modelled as nonlinear dynamics, the nonlinearity present is either purely static or introduced by an external event.

Another application of SSA to EEG data found was that of Celka and Colditz (2002). The aim was to develop a detection scheme by which EEG seizures could be identified. If abnormalities in the EEG are observed, there is a strong possibility that there will be a poor neurodevelopmental outcome in the newborn and infant. The potential therapeutic window for these problems is in the time span of hours, making it important to be able to automatically detect predefined patterns. However, due to reasons stated in the paper, it is not possible to fully automate the detection of these patterns, although it is very desirable to use computer-aided detection. In this paper, a new seizure detection method, based on SSA and Rissanen minimum description length model-order selection (SSA-MDL), was developed. The authors' motivation for the use of SSA were SSA's proven performance on quassi-periodic signals, as is the case for EEG, and the fact that SSA is highly robust to noise. The observed signal was first pre-processed by using a nonlinear whitening filter that spread the spectrum of the background while retaining the rhythmical features of the seizure events. The non-Gaussian shape of the probability density function is also transformed into a Gaussian shape. The signal characteristics of EEGs from newborns and infants include nonstationarity during a single recording, a non-Gaussian probability density function, various artefacts and a rhythmical background EEG of which the frequency spectrum largely overlap with the seizure one.

The suitability of SSA, as well as two other methods suggested in the literature, was investigated by using both synthetic data simulated of EEG seizures and real data from ten babies suffering from EEG seizures. The results showed that for the performance on real EEG data, the adapted SSA-MDL method constantly outperformed the other two techniques suggested, even without the pre-processing of the data series occurring. The results from both the good detection rate and the false detection rate were better than that of the other techniques, with the false detection rate being the most influenced by the pre-processing.

In the work done by Hassanpour et al. (2003), the appropriateness of SSA to detect EEG seizures in newborn infants was once again investigated by comparing the results from SSA with three other non-parametric methods. The methods applied to the data were an autocorrelation technique, a spectrum technique, a time-frequency based technique and then finally SSA. The autocorrelation method performed analysis in the time domain, using the autocorrelation function of short epochs of EEG data, the technique based on the time-frequency domain analysed the interspike intervals of EEG and the spectral analysis technique detected periodic discharges. It was found that, even though SSA gave very satisfactory results, the time-frequency method outperformed SSA in all but one case study. This is probably because a high percentage of the EEG signature occurs in the high frequency area, for which the time-frequency technique was developed.

An application removed from that of studying neurological signals, was the work done by Chiou et al. (2000) on extracting relevant components from heart beat analysis with the aid of SSA. Heartbeat interval time series have a 1/f characteristic and this characteristic can be very useful and significant in clinical situations. However, this 1/f component is often not the only component in the signal and therefore has to be separated. The 1/f signal shows many chaotic characteristics, making SSA a very suitable technique to use to extract the desired information. The authors tested the applicability of SSA by using two real-life time series. Seeing as the aim of the research was only to extract the 1/f component, they were only interested in the dominant, or first, principal component and did not have to establish any criteria by which the series should be reconstructed. They succeeded to illustrate that SSA could be used to separate the 1/f components from the sinusoidal components, allowing investigators to estimate the 1/f slope of heart disease patients.

Some further advances of SSA into other fields of medicine was made by Pereira and Maciel (2001) in the form of an effort to use SSA to estimate the mean scatter space (MSS), which is a parameter used for the quantitative characterisation of biological tissues by ultrasound. The apparent benefit of SSA in this context was its ability to decompose periodic and aperiodic structures from the time signal, even in the presence of noise. SSA was applied to simulated and real backscattered echo time series obtained from phantoms and bovine livers. A Monte-Carlo simulation was also run for both an experimental phantom and a bovine liver sample. This work was continued in (Pereira et al., 2002) where SSA was applied to MSS from ultrasonic measurements of human bone microarchitecture. The estimates obtained from SSA correlated well with estimates of the mean trabecular spacing that were obtained independently with microtomography. SSA was used to identify the periodic eigenvalue pairs, where after the time series was reconstructed with only the periodic components. The Fourier transform of the reconstructed time series was calculated and the predominant MSS was determined, providing information about the characteristics of the bone microarchitecture. The complications of the analysis lay therein that the specific interest was in periodic signals, which would not necessarily be associated with the highest eigenvalues. It has been mentioned earlier (Vautard and Ghil, 1989) that so-called ‘eigenvalue pairs’ could be associated with periodic components of the signal, but it could also be noise. In order to prevent nonoscillatory noise processes to be confused for eigenpairs belonging to the signal, a heuristic criterion was used that stated firstly that the first or the first two pairs were chosen if they represented at least 65% of the cumulative variance and secondly that the frequency associated to each of the eigenvector pair is spaced no more than 2.5% from the other. By using these criteria, SSA was found to be a powerful way to estimate MSS and to have great possibilities in the characterisation of ultrasound data.

2.3.3 Economics

The work by Kepenne (1995) takes the application of SSA to climatic time series one step further, by trying to find a relation between soybean futures prices and the ENSO signal, in an effort to illustrate the socioeconomic repercussions of the ENSO. He used multichannel SSA to isolate variability common to both the Southern Oscillation index and the normalized monthly mean time series of soybean futures prices from other variability and noise present in the data. The paper provided interesting insight into the role that ENSO could play in the soybean futures prices, both regarding to countries affected by ENSO and those that aren’t. Although it would be possible in principle to predict the monthly average soybean futures, it was pointed out by Kepenne that in practice soybean futures are bought and sold on a daily bases and not on a monthly average. The climatological implications of ENSO identified by SSA would therefore be of greater interest and advantage.

Ormerod and Campbell (1997) investigated the applicability of SSA to economic time series. They made the statement that SSA cannot be used to build models, but rather just to identify the underlying structure of the data, be it deterministic or stochastic and to give a measure of the signal-to-noise ratio of the data. Even though SSA would be able to supply information on the regularity and consistency of the factors that influence the price of a given commodity and thereby indicate whether meaningful forecasts could be carried out for the price of this commodity, SSA would not be able to specify what these factors are that influence the price. In the article, the effectiveness of SSA was tested on two time series, the Gross National Product from the USA and the Gross Domestic Product for the UK over a selected period. They stated that they had not found the autocorrelation function, suggested by Mullin (1993) useful in determining a suitable embedding dimension. It was rather decided to use the convention that economic cycles last between two to three to seven to ten years, resulting in up to forty quarterly periods. The resulting eigenspectrum was found to have little structure with no evidence of a noise floor. These results did apparently not change with variations in the embedding window. The significance of the results obtained from SSA was tested with a bootstrapping method, similar to a Monte Carlo analysis, with 500 surrogate data sets. The conclusion from this test was that the singular spectrum of the data from the UK was indistinguishable from that of an artificially generated random series, indicating that the economic time series of the Gross Domestic Product of the UK is probably a purely random series. Slightly more structure were obtained for the Gross National Product results from the US with a distinction between the eigenspectra of the surrogate series and the GNP time series. The main conclusion from their article was that SSA confirmed previous beliefs that economic time series such as the GDP and GNP exhibit too much chaos to be modelled successfully.

Ormerod extended his investigation of the application of SSA to time series relating to economic situations in (Ormerod, 2001) to the Goodwin model and the periodicity of unemployment and factor shares in the United Kingdom. The Goodwin growth cycle model is a model of the business cycle that is crucially affected by unemployment and the share of wages in national income. Ormerod saw the need to examine empirical evidence on the periodicity of these two variables by using a time series of annual data for the UK and applying two techniques of modern statistical analysis. These two techniques were firstly the application of spectral analysis after a kernel filter had removed very low frequency persistence in the data and secondly singular spectrum analysis that was based on the eigenstates of trajectory matrices obtained from the original data.

Ormerod's results indicated definite evidence that regular periodicity existed in both unemployment and the labour share. These periodicities were at the same frequencies normally associated with the business cycle. However, the cycle could only be determined weakly and a large quantity of noise is present in the data. The results from both techniques were very similar.

Thomakos et al. (2002) used singular spectrum analysis to attempt to model the realized volatility and logarithmic standard deviations of daily futures return series. Their results found that SSA could decompose the volatility series quite well and that it effectively captured both the market trend and a number of underlying market periodicities. If reliable information on periodicities that were present was available, it could play an important role in options pricing and risk management. They therefore concluded that SSA could be a valuable tool for financial practitioners. They also recommend exploring the forecasting power of SSA analysis in the context of volatility modelling.

2.3.4 Geophysics

Robertson used spectral analysis in (Robertson and Mechoso, 1998) to identify interannual and decadal cycles in the river flows of south-eastern South America. The multitaper method and SSA were used in conjunction to isolate spectral peaks, assess the statistical significance of these spectral peaks and reconstruct the underlying oscillatory components. SSA was specifically applied for the reconstruction and appeared to be a good complimentary technique to the multitaper method.

The work done by Rozynski et al. (2001) used SSA to identify and investigate temporal and spatial variations in shoreline positions in an attempt to determine characteristic patterns in the shoreline response and to see whether these patterns displayed forced or self-organized behaviour. The researchers decided that the long-term stability of the area being studied and the complicated short-term evolution of the area resulted in ideal conditions to test the effectiveness of SSA. Except for the normal decomposition into principal components and an investigation of the various reconstructed components, Rozynski et al. (2001) also calculated the correlation between the first three reconstructed components. Investigation of trends in these correlations correlated with conclusions that have been made about the behaviour of the system. The SSA allowed specific patterns to be extracted and characterized and was successfully applied to the time series.

However Reeve (2002) entered into a discussion on the article by Rozynski et al. (2001) and pointed out some constraints on the interpretation of the results from SSA that Rozynski et al. had seemingly overlooked. These were that random and systematic errors could mask system dynamics, the optimum embedding dimension could not be known in advance and could therefore only be estimated beforehand and it was not necessarily true that there would be a clear separation between the timescales of forced and natural system response or between the timescales of noise and system dynamics. Reeve cautioned the researchers not to submit to the natural desire to attempt to match the individual components to specific observable aspects of time series behaviour.

In the response from Rozynski (2002), he pointed out that by using SSA, the signal can be uniquely reconstructed in the form of reconstructed components and these reconstructed components can be investigated separately. He also mentioned that the choice of the embedding window should not be such a critical issue because it was illustrated by Vautard et al. (1992) that if the window were large enough, the dominant eigenvalues would remain close to constant.

Schoellhamer (1996) used SSA and spectral analysis to relate measurements of suspended-solids concentrations in the San Francisco Bay to the time series of several potential influences, such as the tides, freshwater run-off and wind shear. This work was continued in (Schoellhamer, 2002) where he applied SSA for time series with missing data (SSAM) to a time series of suspended-sediment concentration from San Francisco Bay. The aims of the study was to reconstruct the components, but was complicated by incomplete data series due to fouling of the sensor. The technique of SSA for time series with missing data was developed in an earlier paper of Schoellhamer (2001) and has the advantage that time series do not need to be screened, filled and subdivided prior to analysis and longer, incomplete, time series can now also be analysed reliably. Different tidal cycles and other physical processes that affected the suspended-sediment concentration were identified, along with each process' contribution to the total variance of the concentration. Schoellhamer showed that SSA, and specifically SSA with missing data, could be used successfully to extract trends from suspended-sediment concentration time series.

Further work on shores and shorelines was done by Stive et al. (2002). The evolution of shores and shorelines is variable over a wide range of different temporal and/or spatial scales, but this variability is generally still hard to understand and difficult to predict. Reliable coastal change information, however, is very important as it is used for informed decision making. The aim of the paper was to describe causes and factors for the variability of shores and shorelines dominated by waves. The work was based on a number of case studies, with the variability of the data described in terms of a range of varying time and space scales. SSA was used to detrend the data, specifically in regard to the long-term or decadal trends. After the detrending, spectral analysis was applied and cycles were successfully identified.

2.3.5 Engineering applications

One of the first applications of SSA in a field that related more directly with engineering was that of Qu et al. (1993) where a number of different nonlinear diagnostic methods were evaluated for their suitability and efficiency for application to large machinery. The methods that were investigated were SSA, pseudo-phase diagrams or limit cycle detection, the Wigner distribution and the Kullback index of the complexity based on information theory. The criteria by which the techniques were judged were their sensitivity for nonlinear phenomena, their applicability in machinery diagnostic practice, their effectiveness in computer execution and their acceptability in the enterprises. It is widely known that failures of mechanical systems are always accompanied by changes in the dynamics from being linear or weak nonlinear to strongly nonlinear. This necessitated the need for nonlinear diagnostic methods, all of which are implemented in different ways and have various advantages and disadvantages. This paper gave a short description of each of the techniques that were investigated, including their main areas of applications, their benefits and their shortcomings. It was found that, even though SSA's computer execution was relatively easy and certain fault types can be easily identified, SSA was not sensitive enough to modulated signals. It was concluded that pseudo-phase diagrams were the most effective for online diagnostics. The other methods were recommended as supplementary techniques.

A very similar study was done a few years later by Wang et al. (2001) where nonlinear diagnostic methods for rotating machinery were evaluated from the view of diagnostic practice. In this study, attention was once again paid to SSA and pseudo-phase portrait, as well as to the correlation dimension. The possibilities of all the techniques, in terms of practical applications, were illustrated with the aid of examples. It was found that each of the techniques had one or two individual elements in which their specific value lay. Pseudo-phase portraits were easily executable and were sensitive to some fault types, while the correlation dimension gave an indication of the number of state variables influencing the output from the process, which in turn represented the number of degrees of freedom of the system and allowed one to classify different faults intelligently. The benefit of the application of SSA was that the dimension of the effective subspace of the embedding space could be determined without prior knowledge of the dynamic system, providing information about the complexity of the system. Instead of trying to conclude which method is the best or the most suitable for diagnosing faults in rotating machinery, the authors concluded that all the techniques could be used successfully in conjunction with traditional methods such as Fast Fourier Transform spectra and time-frequency analysis.

The most recent work published on this topic is that by Liu and Zhao (2003) where they used multi-scale SSA to detect rotor cracks in advance, as the propagation of fatigue cracks is very undesirable for the reliability of rotating machinery. The rotor system becomes nonlinear when cracks are present and all the conventional techniques by which cracks are detected are based on Fourier analysis. However, the authors preferred SSA to traditional Fourier analysis for the analysis of nonlinear dynamics because the technique is based on the eigenelements and these eigenelements have a data-adaptive character. The research combined SSA with multi-scale approach from wavelet analysis and applied a method called multi-scale SSA (MS-SSA), which could be used for the detection of rotor cracks. Multi-scale SSA involves using the principal components of the time series as running time windows. The biggest advantage of MS-SSA is that the analysing functions are data-adaptive; meaning the shape of the analysing functions is not imposed beforehand, but is dependant on the actual time series. The MS-SSA was used to obtain the principal components of both cracked and normal rotors. The differences in the principal components obtained between the two sets of series indicated the characteristics of the cracks present in the cracked rotors and could be used for detection of future cracks.

Shaikh (1997) used local analysis techniques, such as SSA and wavelet analysis, to investigate the transition of fluids into turbulence. He found that traditional global Fourier techniques were unable to resolve the localized nature of the large-amplitude ‘events’ that precede the formation of turbulent spots in a flow field, making him unable to gain significant information about the physical processes involved in the breakdown of the laminar flow layer. His investigation involved the generation of deterministic white noise series, which was then used to excite a laminar boundary layer. These disturbance waves affected the flow patterns in a similar manner as a naturally excited situation would and also resulted in turbulent spots in the flow. The flow mechanisms responsible for the localized formation of events that lead to the breakdown of the flow to turbulence were considered to be highly nonlinear, resulting in the need for local analysis techniques, such as SSA and wavelet transforms. In his first experiment, the three-dimensional flow field associated with a localized coherent flow structure that was identified as the precursor to a turbulent spot was examined. Wavelet transforms were used in the second experiment to examine the breakdown of the identified structure and to generate high-frequency disturbances that preceded the formation of a spot. SSA was then used in the third experiment to study the natural formation and subsequent streamwise development of turbulence spots, as well as to supply an indication of the signal intermittency based on a single physically meaningful threshold criterion. The technique was adapted to detect and track emerging and fully turbulent spots in an automatic manner. Both SSA and wavelet transforms served a specific purpose in the analysis and allowed the researcher to make significant conclusions about the formation of turbulence in fluid flow.

The chaotic characteristics of underwater acoustic signals were investigated by Zhang (1998). The paper focused on noise signals radiated from ships and the techniques that were investigated were power spectrum analysis, singular spectrum analysis, Lyapunov exponents and correlation dimension. The conclusion derived from all of these techniques was that the data exhibits chaotic characteristics. It was also found that the fractal dimensions of all the different signal sources were different and this characteristic could therefore be used to classify underwater acoustic signals. By identifying that acoustic signals exhibit chaotic characteristics, a possible new means of modelling these signals was found.

Wu and Gong (2000) performed a focused investigation to test the suitability of SSA to forecast network behaviour. They used information about the network traffic going over one of the regional networks of China Education and Research Network and looked at all three of the general benefits of SSA. They identified periodic components, which could be used to optimise managing policies of the computer network, they identified and retrieved low frequency variability and trends and they applied SSA to make predictions about the network behaviour. It was concluded that SSA is extremely suitable for application on the behaviour of computer networks.

A short paper was written by Tianfang and Tianxing (2000) in which singular spectrum analysis was applied to nonlinear signal processing. It was found that weak chaotic signals could be detected, even for very undesirable signal-to-noise values. It was pointed out, however, that the technique does not allow one to determine from which dynamical system the signal had originated. Signal classification algorithms should be used for this purpose.

Another engineering application is that of SSA in plasma physics, which was done by Pasqualotto et al. (1999), Bilato et al. (2000) and Marrelli et al. (2001). The particular issues that were addressed in all three papers were SSA's suitability for denoising and detrending statistical tests. The noise series being investigated were white noise and periodic disturbances to the time series. In their discussion of a denoising algorithm, the authors proposed that the optimum number of reconstructed components is formally analogous to the problem of determining the number of independent signals present in a multichannel time series that was affected by uncorrelated white noise. They adopted the criteria known as akaike information content, that has been developed by Akaike (1974) and that has been improved by Wax and Kailath (1985) to the property known as the minimum description length. The optimal choice of the number of components is a trade-off between the ability of the model to describe the data and a penalty function that quantifies the uncertainty in the estimation of the parameters.

The paper by Tung et al. (2001) focused on using multichannel SSA to do the mixed-pixel classification of hyperspectral images. Hyperspectral images are images that were obtained from remote locations and that are used to extract information on surface cover. For a large pixel size, classification information about the land cover could be unreliable, as one pixel could include more than one type of land cover. The principal behind the mixed pixel approach is therefore to approach the image on sub-pixel level and identify proportions of the constituent material of the land cover. For the purpose of this paper, a linear mixing model is assumed, which means that the spectrum of each pixel is modelled as a combination of a set of constituent material spectra. Although the application of SSA in the paper seemed very promising, it was hard to determine from the paper how reliable the results obtained were.

A topic that is of direct process engineering interest, is that of the analysis of pressure signals by using SSA that was done by Palomo et al. (2003). It is widely known that pressure signals are vital in the control and monitoring of any process and even more so in nuclear power plants. The aim of the paper was to obtain the response time of nuclear power plant instrumented sensors signals, when non-desired oscillating components were present. These undesired components are typically caused by faults in the system or components of the plant. Once the response time was obtained, the idea was to remove the undesired oscillating contribution while retaining the information carried out by the signal. The advantage of SSA in this application was that it could extract the few components that would explain the main characteristics of the signal while simultaneously removing unwanted oscillatory components, high-frequency contributions and system noise. In conclusion they found that SSA was a powerful tool with which to improve the response time methodology used for signal-noise analysis.

2.3.6 Solar physics

Watari (1996) first applied SSA to time series obtained from solar activity in an effort to identify chaotic behaviour. A mixture of chaotic and periodic components present in the time series complicated this search. The time series was therefore first separated with the help of SSA and then examined by a nonlinear prediction method. From the analysis it was found that the so-called sunspot series that was being investigated consisted of a dominant periodic component and a highly irregular component. However, the irregular component was more representative of random noise than of chaos.

A different aspect of solar physics was addressed by Rangarajan and Iyemori (1997), Rangarajan (1998) and Rangarajan and Iyemori (1998) where SSA was used to analyse indices of geomagnetic activity, interplanetary magnetic fields, sunspot variability and solar wind parameters. It was attempted to isolate significant signal components, specifically quassi-periodic fluctuations, from background noise in the time series. Correlation and power spectral analysis was used in combination with the SSA to ascribe observed oscillations to certain models and structures.

Gallego et al. (1999) used the visual observations of the variable star Z And to estimate the background noise and detect the true signal by using an autoregressive model and SSA. From these principles of dynamical systems theory, the correlation dimension of the attractor of the time series was estimated.

The work done by Castagnoli et al. (1999b) links back to the application of SSA to climatic time series in that δ^{18}O profiles obtained from sea cores were used to identify the imprint of the solar records in a climatic time series. This work was continued in (Castagnoli et al., 1999a, Castagnoli et al., 2002a, Castagnoli et al., 2002b) and all the apparent oscillations observed from the analysis were confirmed by the performance of Monte Carlo SSA.

It had previously been established that total solar radiance changes over a range of periodicities. In the work presented by Pap and Frohlich (1999), the long-term variations of solar irradiance during specific previously identified cycles were reviewed. SSA was used to separate these long-term trends from the rest of the time series and was therefore rather used as a tool, than as the main focus of the investigation. This work was continued in (Pap et al., 2001) where SSA was used once again to smooth the data. Even though the main focus in this work was to evaluate information extracted from the data rather than to verify the applicability of SSA, it was still found that SSA was a very handy tool in the decomposition of solar activity time series.

Arzner (2002) applied SSA to identify periodic functions and smooth functions in the RHESSI light curves. The relative advantage of SSA in this application was that it did not require an independent estimate of the period of the deterministic component in the data, seeing as the RHESSI spin period is sometimes unknown. However, in the paper presented, SSA was only applied to time series for which the period of the deterministic component was indeed known, in order to verify the applicability of SSA to time series obtained from the RHESSI light curves. It was believed that using SSA had identified a number of impulsive features from the data, but these results should be substantiated by further research.

Other work done on combining SSA with the analysis of time series obtained from the solar cycles is (Rangarajan and Iyemori, 1997, Bhardwaj and Tangarajan, 1998, Rangarajan and Barreto, 2000, Juckett, 2001, Khramova et al., 2002). As with the climatology research, the majority of the applications in solar physics used SSA simply as a tool with which to extract the oscillatory components and other trends from the data, but due to the nature of the time series obtained from solar physics, these applications are still very relevant to process engineering.

2.4 Other applications of SSA

2.4.1 Process modelling and forecasting

One of the earlier applications of SSA in combination with forecasting was done by Kepenne and Ghil (1993). The data was first prefiltered by applying multichannel SSA (M-SSA), where after the time series was forecasted with the maximum entropy method (MEM). This approach allowed predictabilities of the subannual variability in atmospheric angular momentum of up to a month. The combination of these two techniques proved to be very successful, due to the nature of SSA (and therefore M-SSA) to remove variations in the data, resulting in a smooth time series and the good reputation of MEM of being able to predict smooth time series quite accurately.

Lisi et al. (1995) saw the opportunity to combine nonlinear dynamical system and artificial intelligence theory to perform forecasting of time series. They first performed adaptive noise reduction of the data by using an algorithm based on SSA and then did the forecasting by means of standard feed forward neural prediction models. They repeated the SSA of the series for a whole spectrum of retained components, calculating the normalized mean square prediction error of each forecast of the reconstruction of the original signal. The number of principal components corresponding to the smallest normalized mean square prediction error was then retained. They found that their approach was very successful for relatively short and very noisy time series, both regarding short- and long-term predictions.

The work from the above study was continued in (Lisi and Medio, 1997) where the predictability of the exchange rate was investigated, in reference to a model that implies future prices are unpredictable if the information set that are used for the predictions is the past prices. The validity of this model was taken into question and attempts were made to provide forecasts for exchange rate time series. Lisi and Medio mentioned that existing tests for nonlinearity had many shortcomings, in that they failed to generate consensus despite their relatively weak hypothesis tests. The probable reason for this was given as the lack of robustness of the tests and the differences in their power functions.

The aim of the paper from Lisi and Medio was to test a hypothesis they had labelled the Nonlinear Hypothesis. This hypothesis stated that the data does contain some structure and this structure can be exploited for short-term predictions. It was also assumed that the data have been generated by a nonlinear generating system with some additional noise added. This hypothesis was tested by combining a number of techniques from the dynamical systems theory in a novel way. This combination of techniques essentially resulted in SSA and multi-channel SSA (MSSA). The method used for the prediction of the economic time series is the nearest neighbour method, which is based on the principle that similar states over a small enough time interval will have similar successors. The successors of the past states are therefore used to predict the successors of the future similar states.

The data series that Lisi and Medio used to test their hypothesis, were the monthly spot exchange rates of seven major foreign currencies and it was attempted to do one-step ahead forecasting. These data sets were divided into training and validation data sets, with the predictions being done on both unfiltered data and data that have been filtered using MSSA, as well as a trivial prediction using the random walk hypothesis. Two-channel SSA was used for the filtering and when the prediction of the validation data was then attempted, the second currency was used as an auxiliary currency. The mean square prediction error, as well as the mean absolute prediction error was used to quantify the comparison between the random walk prediction and the two local linear or nearest neighbour prediction models. Their results indicated that the time series could successfully be modelled on the short term and that the local linear predictions on the filtered data outperformed that of the unfiltered data.

Masulli et al. (2001) presented a constructive approach to time series learning and the forecasting of individual rainfall intensities series. The specific method used was a decompositive ensemble method that was based on SSA. This method extended the constructive approach to the learning of discontinuous and/or intermittent signals.

A multi-layer perceptron neural network was used for the modelling, where the embedding dimension was determined by the Global False Nearest Neighbours method and the time lag of the input was the first minimum of the average mutual information of the signal (Abarbanel, 1996). Even though (Ormerod and Campbell, 1997) had serious misgivings about the value of SSA in a predictive environment, the application of SSA to prediction can be supported by the argument from (Masulli et al., 2001) that, since the principal components are filtered versions of the signal and are typically band-limited, they should behave more regular than the raw series and would hence be more predictable.

Computational costs were reduced during the predictions by combining the reconstructed components of similar explained variance into reconstructed waves and then predicting these waves. The prediction of the original series was then recovered as the sum of those of all the individual series components.

It was found that by using a constructive methodology, efficient predictors could be designed, even for complex signal such as those that are discontinuous or intermittent. The ensemble method combined an unsupervised and a supervised step. The unsupervised step was the decomposition where the original signal was decomposed with the aid of SSA into reconstructed waves. The supervised step was the designing and learning of the MLP predictors for each of the reconstructed waves, with the aid of suggestions from dynamical systems theory.

Singular spectrum analysis was also combined with the even more advanced technique of genetic algorithms to forecast the solar cycle (Orfila et al., 2002). The results that were obtained from this study was in good agreement with known behaviour of the solar cycle and it would therefore seem as if SSA can also be used to reconstruct data to be forecasted with genetic algorithms.

2.4.2 Change point detection

The detection of changes in time series is another area that has been the focus of a great amount of research. In (Moskvina, 2001, Moskvina and Schmidt, 2003, Moskvina and Zhigljavsky, 2003) work was done into the application of SSA to change-point detection. It was found that the sequential application of SSA could be used to detect change-points in time series. The change-point detection algorithm is based on the idea that if the mechanism generating a certain time series changes at a specific moment in time, the distance between the subspace spanned by the eigenvectors and the vectors after the change point will increase.

Moskvina identified four parameters, in addition to the normal SSA parameters (see chapter 3), that should be optimized for the change-point detection algorithm. These are the section of the time series on which SSA will be applied, two parameters concerning the test sample of which the closeness is evaluated and the threshold value by which the significance level is

determined. It was found that this technique based on singular spectrum analysis performed well when compared numerically to other change-point detection techniques.
This technique for change point detection has also been applied by Choi et al. (2002), in an effort to facilitate the timely detection of changes in traffic loads. The emphasis of their work was rather on determining suitable sampling techniques and the change point detection algorithm based on SSA was just used to validate the sampling techniques.

3 METHODOLOGY

The term ‘singular spectrum’ originates from the eigenvalue or spectral decomposition of a given matrix **A** into its spectrum or set of eigenvalues. It is these eigenvalues, λ, that make the matrix **A** - λ**I** singular. However, the use of the term singular spectrum can be slightly misleading or confusing in the context of the analysis of a single time series, as the spectral decomposition of matrices of multivariate data is also referred to as singular spectrum analysis. The application of singular spectrum analysis to time series is a relatively new approach and has its origin mainly in the study of chaos theory.

3.1 Basic Singular Spectrum Analysis

The general approach involves embedding the time series data in a high-dimensional trajectory matrix. Principal component analysis is then performed on the embedded data, introducing a new coordinate system, which moves the origin of the data to the centroid of the reconstructed system states. The dominant principal components then become the axes of the new coordinate system, each representing the maximum amount of variance in the data possible.

The procedure performed during singular spectrum analysis can be described more formally in four main steps and these are illustrated in Figure 3.1 (Golyandina et al., 2001). Step one is the embedding of the time series in a high-dimensional lagged trajectory matrix and step two involves the decomposition of the trajectory matrix into the sum of a number of bi-orthogonal matrices of rank one. These two steps are considered together to be the decomposition stage. Next comes the reconstruction stage, which can be divided into steps three and four. Step three involves the summing of the various matrices that were formed in step two into different groups, depending on the nature of the matrices. Finally, in step four, the time series representing the various groups can be reconstructed from the resulting matrices. These two stages, and the four underlying steps, will now be discussed in significantly more detail.

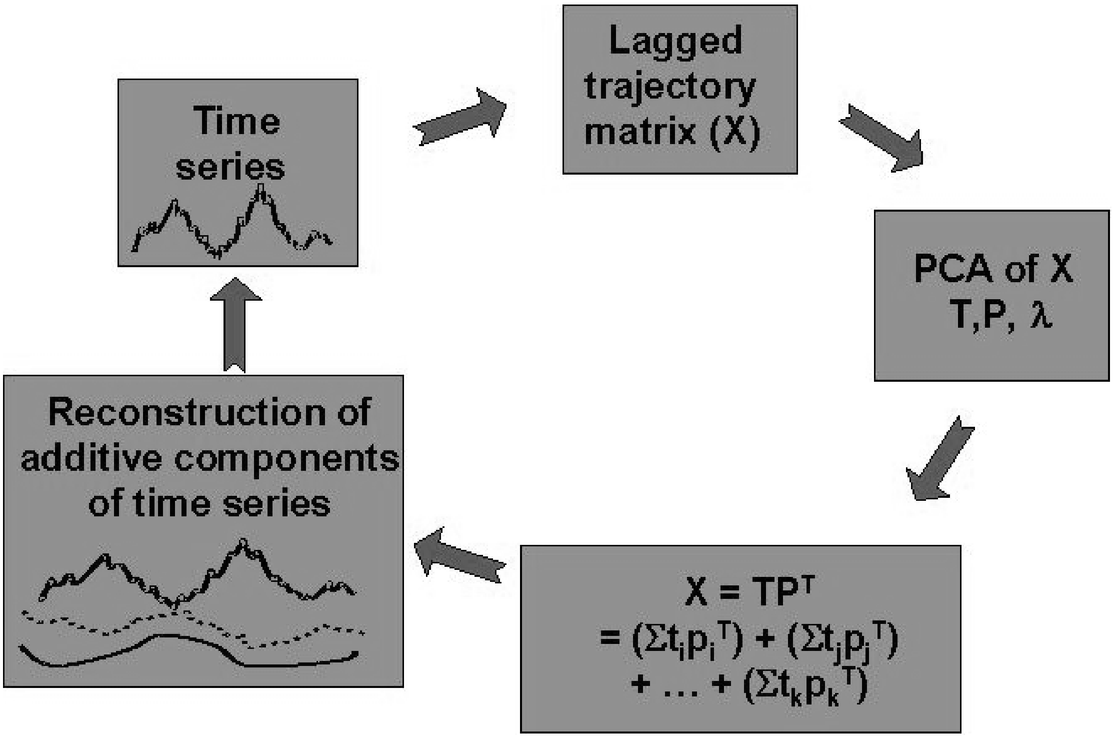

Time series
Lagged trajectory matrix (X)
PCA of X
T,P, λ
$X = TP^T$
$= (\Sigma t_i p_i^T) + (\Sigma t_j p_j^T)$
$+ \ldots + (\Sigma t_k p_k^T)$
Reconstruction of additive components of time series

Figure 3.1 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.

3.1.1 Decomposition stage

a) Embedding of time series

The purpose of the embedding stage is to expand the single time series into a multi-dimensional matrix, called a trajectory matrix, which can then in turn be decomposed into various components.

This embedding is done by giving the time series a certain lag, usually one, and then combining the resulting lagged column vectors into a matrix of a specific window length (number of columns).

Given a specific real value time series of length *n*,

$$\mathbf{Y}_t = \left[y_0, y_1, y_2, \ldots, y_{n-1}\right]$$

Let the series be embedded in a window length of *L* where *L* is an integer and 1<*L*<*n*. If the embedding is done with a unity lag, the embedding process will result in *L* lagged vectors, each of length *K* = *n* – *L* + 1 data points

$$\mathbf{X}_i = \left[y_{i-1},\ y_i,\ y_{i+1}, \ldots,\ y_{i+K-2}\right]^T \qquad 1 \leq i \leq L$$

The trajectory matrix **X** can then be constructed by combining the lagged column vectors into a single matrix

$$\mathbf{X} = \left[X_1,\ X_2,\ \ldots,\ X_L\right]^T$$

The mathematical definition of the combined trajectory matrix therefore is:

$$\mathbf{X} = \begin{bmatrix} y_0 & y_1 & y_2 & \cdots & y_{L-1} \\ y_1 & y_2 & y_3 & \cdots & y_L \\ y_2 & y_3 & y_4 & \cdots & y_{L+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{K-1} & y_K & y_{K+1} & \cdots & y_{n-1} \end{bmatrix}$$

The above principle of the construction of the trajectory matrix can be illustrated by the following example:

Let y_t be a selected time series.

y_t = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

This time series can now be given a lag of 1 and be embedded in a matrix with a window length of five, resulting in the following trajectory matrix, **X**:

$$\mathbf{X} = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 5 & 6 & 7 \\ 4 & 5 & 6 & 7 & 8 \\ 5 & 6 & 7 & 8 & 9 \\ 6 & 7 & 8 & 9 & 10 \end{bmatrix}$$

The optimal size for the embedding window depends on the nature of the time series and it is vital in the analysis of the time series to determine the most favourable window length. This window length should be wide enough to sufficiently capture the global behaviour of the system, but it should be kept in mind that the complexity of the analysis increases with the increase in the number of columns in the trajectory matrix. Because the width of the embedding window is one of the two most fundamental parameters of SSA, a substantial amount of research in the literature has also been devoted to the criteria by which the embedding window should be determined. A brief overview of the discussion the literature will be given at the end of this section.

For the purposes of this research, the optimum size of the embedding window was taken at the minimum of either the point of linear decorrelation (where the autocorrelation between the first and the last columns of the matrix is negligible) or the point where the autocorrelation between the first and the last columns reaches a first minimum, as will be illustrated later.

The autocorrelation function of a time series is an indication of the degree of correlation between the observations of a time series, *x*, that are separated by a delay of ξ. This correlation is calculated by: (Addison, 1997)

$$C = \frac{\sum_{i \in 1}^{N-\xi} (x'_i)\,(x'_{i+\xi})}{\sum_{i \in 1}^{N-\xi} (x'_i)^2} \qquad 3.1$$

and

$$x'_i = x_i - \bar{x}_i \qquad 3.2$$

Equation 3.1 is repeated for a whole spectrum of delays (ξ-values), until the delay is found at which the correlation becomes either a minimum or negligible.

b) Decomposition of time series

Once the trajectory matrix **X** has been formed, principal component analysis (PCA) or, similarly, singular value decomposition (SVD) can be performed. The analysis is performed on the lagged covariance matrix **A,** which is calculated from the trajectory matrix by using equation 3.3.

$$\mathbf{A} = \frac{\mathbf{X}^T\mathbf{X}}{(K\text{-}1)} \qquad 3.3$$

There are several different ways by which the lagged covariance matrix can be calculated (Elsner and Tsonis, 1996). The two most common structures are Hankel and Toeplitz matrices, with the Hankel structure being the structure initially suggested by Broomhead and King (1986) and the Toeplitz structure being favoured by, among others, Vautard and Ghil (1992). The format of the lagged covariance matrix used in this analysis (equation 3.3) can be classified as a Hankel matrix, characterized by the equal elements on the 'diagonals', in other words $y_{i(j-1)} = y_{(i-1)j}$ for all *i,j* > 1. The other alternative, the Toeplitz approach, calculate the lagged-covariance matrix by using:

$$\mathbf{A}_{\text{Toeplitz},\, ij} = \frac{1}{N_t - |i-j|} \sum_{t=1}^{N_t - |i-j|} x_{|i-j|+t}\, x_t \qquad 3.4$$

It was decided to use the Hankel form rather than the Toeplitz structure because the Toeplitz SSA is not aimed at nonstationary time series and has the disadvantage of producing a nonoptimal decomposition (Golyandina et al., 2001).

When singular value decomposition is applied to the lagged covariance matrix **A**, it results in the following decomposition of matrix **X**

$$\mathbf{X} = \sum_{i=1}^{d} \sqrt{\lambda_i}\, \mathbf{U}_i \mathbf{V}_i^T \qquad 3.5$$

λ_i (i = 1, . . ., L) represents the eigenvalues of matrix **A** and are arranged in decreasing order of magnitude and **U** is the corresponding orthonormal system of the eigenvectors of the matrix **A**. The value of d is determined by the rank of **X** and is the maximum such value that would include all the eigenvalues larger than zero.

In standard SVD terminology, $\sqrt{\lambda_i}$ are referred to as the singular values and it is these ordered values that are referred to as the singular spectrum of a given matrix. $\mathbf{U}_i$ and $\mathbf{V}_i$ are the left and right singular vectors of matrix **X**. The collection ($\sqrt{\lambda_i}$,$\mathbf{U}_i$, $\mathbf{V}_i$) is called the i^{th} eigentriple of the matrix **X**.

By now setting $\mathbf{P}_i = \sqrt{\lambda_i}\mathbf{V}_i$ and $\mathbf{T}_i = \mathbf{U}_i$, equation 3.5 can be converted into the result from principal component analysis

$$\mathbf{X} = \sum_{i=1}^{d} \mathbf{T}_i \mathbf{P}_i^T \qquad 3.6$$

This illustrates the similarity between PCA and SVD and shows that by applying principal component analysis, the trajectory matrix ($\mathbf{X} \in \Re^{kxl}$) is decomposed into the product of a score ($\mathbf{T} \in \Re^{kxl}$) and a transposed loading ($\mathbf{P} \in \Re^{lxl}$) matrix.

The trajectory matrix can now be expressed as the sum of the outer products of the individual pairs of vectors $\mathbf{t}_i$ and $\mathbf{p}_i$, from which matrices **t** and **p** are composed, as shown in equation 3.7.

$$\mathbf{X} = \mathbf{T}_i\mathbf{P}_i^T = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \ldots + \mathbf{t}_d\mathbf{p}_d^T \qquad 3.7$$

By now the time series has been decomposed into its basic constituents and the time series can be analysed and reconstructed according to the relevance and importance of the various principal components.

One of the benefits of using SVD is related to the properties of the directions that are determined by the eigenvectors $\mathbf{u}_1$, ..., $\mathbf{u}_d$. The eigenvectors are created in such a manner that the variation of the projections of the lagged vectors in the direction of the first eigenvector is a maximum and thereafter the direction of every subsequent eigenvector is orthogonal to all previous directions. The variation of the projections of the lagged vectors onto all other directions is also maximal. This means that the direction of the ith eigenvector, $\mathbf{u}_i$ can be called the ith principal direction.

This characteristic also allows one to determine the amount of the total variance accounted for by each of the principal components or eigenvectors, as the eigenvalues associated with each eigenvector specifies the relative magnitude of the variance explained by the relevant eigenvector.

The most significant properties of the PCA decomposition are summarized by Benzi et al. (1997) as:

a) The normalized eigenvalue spectrum is proportional to the fraction of the total variance that each principal component explains.

b) Due to the fact the eigenvalues are in decreasing order, the tail of the spectrum shows a plateau that corresponds to the noise components that result from small-scale fluctuations.

c) The highest eigenvalues represent the large-scale fluctuations of the deterministic fields and by using only these components, one can filter the original signal from the small-scale fluctuations.

d) The principal components themselves are time-dependent coefficients that describe the parameter evolution and can also be used in other types of analysis, such as spectral analysis and cluster analysis.

In the rest of the discussion, the parameters obtained from PCA (**T** and **P**) will be used, rather than those from SVD (**U** and **V**). The so-called principal components of PCA are represented by the ‘score’-values or **T**-vectors.

3.1.2 Reconstruction stage

c) Grouping of components

Once the decomposition of matrix **X,** described in equation 3.7, has been obtained, the *d* different matrices ($\mathbf{t}_i\mathbf{p}_i^T$ combinations) can be summed into a smaller number of groups representing different underlying aspects of the time series. Any time series, **Y**, can be seen as the sum of a number of basic time series, $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$, …, $\mathbf{Y}^{(i)}$, each with their own fundamental behaviour, influences and effects. The aim of this step is to combine the various principal components into the appropriate underlying time series.

These basic time series or additive components are typically the 'trend' components, which are smooth parts of the series that show long-term variance, the various 'oscillatory' components that indicate periodicities in the time series and finally the largely undesired 'noise' components. The process of combining the matrices into different groups or subsets is known as the eigentriple grouping.

It was seen from chapter two that by far the most common application of SSA in the literature was to extract trends. However, for engineering purposes the emphasis is more on dividing the given time series into valuable signal and underlying noise. Some work was also done specifically on noise reduction by Broomhead and King (1986), Vautard and Ghil (1989) and Vautard et al. (1992) and the more technical details of their findings will be discussed at the end of this section, along with some research on the interpretations of eigenelements themselves.

It should be mentioned at this point that this pure division of a measured time series into distinct underlying time series consisting of noise and signal is only true for a system that is linearly separable. This requires that the time series was contaminated by a purely white noise process. If the nature of the noise was red noise, the effect of the noise would be embedded throughout the distribution of the eigenelements. However, for the purposes of this discussion, the assumption will be made that the time series has been generated by a linear process and is therefore separable. The more complicated situation where the noise contaminated the time series is characterized as red noise, will be discussed in one of the case studies in a later section (section 4.7).

If it is assumed that a time series **Y** is composed of two underlying time series, they can be seen as being separable by the decomposition in equation 3.7 if there exists a collection of indices $\mathbf{I} \subset \{1, . . ., d\}$ such that $\mathbf{X}^{(1)} = \Sigma_{i\in I}\mathbf{X}_i$ and $\mathbf{X}^{(2)} = \Sigma_{i\notin I}\mathbf{X}_i$. The contribution of the first underlying time series $\mathbf{X}^{(1)}$ can then be quantified by using the eigenvalues associated with the eigenvectors that were combined to construct time series. The share of the relevant eigenvalues, ϕ, and therefore of the individual series is measured by ϕ

$$\phi = \frac{\sum\limits_{i\in I}\sqrt{\lambda_i}}{\sum\limits_{i=1}^{L}\sqrt{\lambda_i}} \tag{3.8}$$

which is also indicative of the combined amount of variance explained by the grouped eigenvectors, as was explained in a previous section.

The above example is typical of the approach followed in the analysis being considered, where the emphasis was placed mainly on separating the underlying signal from noise present in the measurement. In this case it has been assumed that the measured signal consisted only of two underlying time series, viz. signal and noise, and not so much importance was attached to the identification of oscillatory components.

The exact position of the grouping of the various components is a very subjective decision and depends strongly on the aims of the analysis. Components that might be considered as 'noise' for one application could be seen as valuable signal information for another analysis. If the analysis is done for visualization purposes, it might be sufficient to retain components that explain as little as 60% of the variance. However, if the time series is analysed and refined for modelling purposes, it is often necessary to explain in excess of 95% of the variance, to ensure that none of the essential information from the data is lost. This requires that a larger number of principal components or eigenvectors be retained.

In the general application of this work, any significant breaks in the eigenspectrum of the time series or the obvious presence of a noise floor served as a first estimate of the group of principal components to retain. These breaks in the eigenspectra were determined by visual inspection of the relevant figures and the main feature that was searched for during the analysis was the appearance of clusters in the spectra. These clusters were indicative of eigenvalues with similar relevance. The so-called noise floor could therefore usually be identified as the large group of eigenvalues towards the tail of the eigenspectrum that all explain roughly the same (small) amount of variance in the data. Although this approach can be criticized as being very subjective, it was found that the visual judgements were relatively easy and reliable.

d) Reconstruction of original time series

Once the individual principal components of the time series have been separated into the relevant groups, the original time series can be reconstructed with a smaller number of principal components. This is done by performing diagonal averaging on the matrices resulting from the summation of the relevant $\mathbf{t}_i\mathbf{p}_i^T$ products in each of the groups.
For a $K \times L$ matrix, $\mathbf{X_{rec}}$, the elements of the matrix are defined as x_{ij} with $1 \leq i \leq K$, $1 \leq j \leq L$. Let $L^* = \max(L, K)$, $K^* = \min(L, K)$ and $N = L + K - 1$. Let $x^*_{ij} = x_{ij}$ if $L<K$ and $x^*_{ij} = x_{ji}$ otherwise. The use of diagonal averaging will transfer the matrix $\mathbf{X_{rec}}$ to a single time series $z_0, . z_k. . , z_{N-1}$, according to the following formula (Vautard et al, 1992):

$$z_k = \begin{cases} \dfrac{1}{k+1}\sum_{m=1}^{k+1} x^*_{m,k-m+2} & \text{for } 0 \leq k \leq K^* - 1, \\ \dfrac{1}{K^*}\sum_{m=1}^{K^*} x^*_{m,k-m+2} & \text{for } K^* \leq k \leq L^* \\ \dfrac{1}{N-k}\sum_{m=k-L^*+2}^{N-L^*+1} x^*_{m,k-m+2} & \text{for } L^* + 1 \leq k \leq N \end{cases} \qquad 3.9$$

Equation 3.9 can be explained in a more practical manner by the following example. If matrix $\mathbf{X_{rec}}$ were obtained by summation of the main principal components to be retained, the original signal would be extracted by calculating the averages of the respective diagonals (following the dotted lines), as illustrated.

$$\mathbf{X_{rec}} = \begin{bmatrix} 4 & 3 & 4 & 5 & 1 \\ 3 & 5 & 6 & 5 & 6 \\ 3 & 9 & 4 & 5 & 8 \\ 9 & 7 & 8 & 9 & 8 \\ 2 & 7 & 5 & 4 & 1 \\ 3 & 4 & 9 & 6 & 4 \\ 9 & 5 & 6 & 8 & 7 \end{bmatrix}$$

The resulting time series will then be:
$\mathbf{z_{rec}} = [\,4 \;\; 3 \;\; 4 \;\; 7.25 \;\; 3.8 \;\; 4.8 \;\; 7 \;\; 6.5 \;\; 4.3 \;\; 6 \;\; 7\,]$
If the series has been properly embedded, skilfully grouped and carefully reconstructed, the time series should now represent the underlying process dynamics more reliably and this should expedite and improve the process modelling, optimisation and control.

3.1.3 Literature review on embedding and window length

The concepts of embedding space and embedding dimension were first introduced by Broomhead and King (1986) in relation to a discussion on the method of delays. They highlight the advantage of using a higher embedding delay than 1, which would cancel the effect of the highly correlated samples resulting from the very short sampling times employed in practice. However, they also identified the need to establish a basis to use to determine the optimum embedding dimension, as they mentioned some of the problems experienced with the current application of the method of delays. By studying the Lorenz model for various step sizes and sampling sizes while using a constant window length, it was found that the shape of the singular spectrum is insensitive to the range of sampling times used. However, this claim was contradicted by Vautard and Ghil (1989) when they found that the number of significant eigenvalues also depended on the sampling time used. A reduction in the sampling time increased the information about both the noise and the true signal, leading to more eigenvalues in both the noise floor and the significant part of the spectrum.
Vautard and Ghil (1989) made the statement that no optimal window length exists, as the choice of the window length is a compromise between including more significant information

about the time series in a large window size and achieving a high degree of statistical confidence in a small window size. They suggest rather varying the window length over a reasonable range and evaluating the stable features of the eigenset from this range of window lengths.

One of the discoveries of Vautard et al. (1992) from their investigation of the optimum window length was that SSA does not resolve periods longer than the window length, leading to the conclusion that the larger *m* is, the better it would be for the construction of strange attractors whose spectrum includes periods of arbitrary length. However, to prevent statistical errors, they recommend that the window length does not exceed one third of the length of the time series.

3.1.4 Noise reduction

Broomhead and King (1986) identified the possibility that the noise floor in the eigenvalue spectrum of a series contaminated by white noise can be used to extract the deterministic component from the data. Vautard and Ghil (1989) thereafter illustrated that if the noise is not purely white, a number of plateaus occur in each eigenspectrum, resulting in a more complicated spectrum than that which Broomhead and King (1986) suggested.

Vautard and Ghil (1989) defined the statistical dimension of the data set as the number of significant mutually-orthogonal directions of a reconstructed attractor. This statistical dimension thus provided an upper bound for the minimum number of degrees of freedom (d-o-f) of the measured system. In a practical sense relating back to SSA and the initial work done by Broomhead and King (1986), the statistical dimension is the number of singular values above the noise floor.

In order to show that SSA is a powerful tool for signal reconstruction from noisy data, Vautard et al. (1992) developed a systematic method to determine the break in the eigenvalue spectrum, by which the ‘noisy’ eigenvalues in the tail of the spectrum can be identified.

$$n(p) = \frac{\sum_{i=\mathrm{l}}^{N}(y_i - \sum_{k=\mathrm{l}}^{p} x_i^k)^2}{\sum_{\mathrm{i}=1}^{N} \{ y_i - x_i \}^2} \qquad 3.\ 10$$

They suggested using equation 3.10 to determine the noise reduction ratio (*n*(*p*)) when *p* reconstructed components are considered and where y_i is the noisy time series and x_i is the underlying signal. The average of this ratio was determined for 100 realizations obtained from each of four different processes with simple and known properties. These averages of the ratios for each process allowed the optimum number of components necessary to retain for the removal of the noise from the data to be calculated. However, they also pointed out that for real-life processes one does not have 100 realizations available and nor does one not know what the underlying clean signal should be, making this approach hard to follow.

The more realistic technique that they suggested to test whether the section of eigenvalues labelled as noise is in fact noise, was to compare the results of the SSA analysis of the noise component with that from a Monte Carlo simulation of a pure Gaussian noise process. If all the relevant components have been retained, the SSA from the pure Gaussian white noise process and that from the rejected eigenvalues should provide statistically indistinguishable results. The term statistical dimension was defined as the smallest number of reconstructed components *p* for which the statistical results of the pure Gaussian white noise process and the noise component obtained from the time series will be identical. When the original signal is reconstructed by a reduced number of principal components, the difference between this reconstruction and the original signal is due to two factors. The first is the part of the signal that was removed by the filter and the second is the part of the noise that was retained with the signal. The skill of the analysis lies in minimizing both these quantities.

However, as Vautard and Ghil (1989) pointed out, it should be remembered that the number of variables necessary to fit the data is not an invariant of the underlying dynamical system, but it is also a function of the characteristics of the data.

Due to the nature of the paleoclimatic time series that Vautard and Ghil (1989) studied, they also addressed the topic of data of which the measurements were not performed regularly in time and interpolation was necessary to provide sampling uniformity. The problem with these time series is that the noisy component in the data was also interpolated and thereby the data cannot be viewed as purely white noise anymore.

After applying SSA to these series, they found that the technique is very flexible and incisive. It allowed an assessment of the reliability of the estimates for the dynamical dimension by providing two criteria. The first is that the estimate for the dynamical dimension must be smaller than the statistical dimension and, secondly, when the data are projected onto the eigenvectors, the correlation histogram of the raw data must provide the same dynamical dimension estimate as that from the data with variance normalized components. A further benefit that was identified was that SSA allowed a diagnosis of the reasons why the two tests mentioned above failed, such as insufficient sample length or severe nonstationarity in the data. The other conclusion from the research was that SSA provided a data-dependent but reliable estimate of the statistical dimension, either by investigating the singular spectrum of simple cases or by combining a study of the eigenspectrum with that of the principal components for more complex time series. They concluded that, even though SSA is related to ordinary spectral analysis, it is considerably more robust to the nonstationarities that can be found in climatic records.

3.1.5 Interpretation of eigenelements

a) Trends and nonstationarities

It was pointed out by Vautard et al. (1992) that if only one realization of a process was available, it is impossible to determine whether apparent behaviour of the data (such as an increase in the mean value) is due to a trend or nonstationarity or due to the presence of ultra-low frequencies. However, if the purpose of the analysis is to study higher frequencies that are clearly manifested, the presence of any one of the above phenomena is very undesirable. Vautard et al. (1992) derived a systematic data-adaptive algorithm, based on the same principles as noise reduction, to remove trends or ultra-low frequencies from data.

b) Pairs of eigenelements

According to Elsner and Tsonis (1996) a record containing a significant oscillation will produce a dominant eigenvalue pair (that explains in excess of 65% of the total variance) that has nearly identical frequencies. This is a simplification of the two natural criteria based on the spectral properties of the eigenvectors that was proposed by Vautard et al. (1992) to determine the statistical degeneracy of paired eigenvectors. The first criterion is based on the observation that oscillating pairs of eigenelements must be spectrally localized around the same frequency. The other criterion is that the amplitude of the peaks of the spectra must be high. If the criterion that the pair has the same frequency is not met, it is likely that the dominant pair was rather the result of first order autoregressive noise present in the data.

3.1.6 Prediction

It was pointed out by Vautard et al. (1992) that the because the PC's are filtered versions of the signal, their behaviour is generally more regular and only a selected subset of these eigenvalues could be predicted. The forecasts of the individual PCs can then be combined with the reconstruction algorithm to obtain a forecast of the whole series

3.2 Multichannel Singular Spectrum Analysis

Generally, in applications in process engineering, one is presented with a series of measurements on a set of variables, rather than just a single variable. Multichannel or multivariate SSA, as it is sometimes referred to, is a natural extension of the approach discussed previously for a single time series. When considering *n* observations at equal time intervals on a set of Q variables or a Q-channel time series ($X_q(t)$:, q = 1, 2, … Q), the generalization of SSA to a multivariate time series can be approached by first computing a trajectory matrix for each variable X, i.e. for the k^{th} variable

$$\mathbf{X}_k = \begin{bmatrix} X_{k,1} & X_{k,2} & \cdots & \cdots & X_{k,m} \\ X_{k,2} & X_{k,3} & \cdots & \cdots & X_{k,m+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{k,n'-1} & \cdots & \cdots & \cdots & X_{k,n-1} \\ X_{k,n'} & X_{k,n'+1} & \cdots & \cdots & X_{k,n} \end{bmatrix} \qquad 3.11$$

with 1 ≤ k ≤ Q. These lagged trajectory matrices are subsequently used to form an augmented trajectory matrix, D = ($\mathbf{X}_1$, $\mathbf{X}_2$, …, $\mathbf{X}_k$, …, $\mathbf{X}_Q$). The embedding dimensions of the various variables do not have to be identical, as the resulting combined trajectory matrix will just have the number of rows of the shortest individual trajectory matrix (determined by the largest embedding window). The excess values from the other individual trajectory matrices will simply be removed.

A grand lag covariance matrix, C_X, can then be constructed from the augmented trajectory matrix, analogous to that obtained with a single time series. Each block $C_{i,j}$ is a matrix containing estimates of the lag covariance between channels i and j.

$$\mathbf{C}_x = \begin{bmatrix} C_{1,1} & C_{1,2} & \cdots & \cdots & C_{1,Q} \\ C_{2,1} & C_{2,2} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots & C_{Q-1,Q} \\ C_{Q,1} & \cdots & \cdots & C_{Q-1,Q} & C_{Q,Q} \end{bmatrix} \qquad 3.12$$

The rest of the analysis is performed similar to that of standard SSA in that PCA is used to extract the principal components and the time series is reconstructed with a reduced number of principal components. It is worth noting that if the different series are contaminated with white noise to varying extents, the cleaner series will automatically benefit more from the analysis (Lisi and Medio, 1997). It should just be kept in mind during the ‘unembedding’ or diagonal averaging of the time series that the matrix of summed components must be separated into different blocks representing each of the original variables. This is accomplished by simply extracting the number of columns of the matrix that correspond with each of the original individual trajectory matrices. The diagonal averaging can then be done independently on each block of the reconstructed matrix to ensure that the reconstruction of each variable is obtained separately.

3.3 Monte Carlo SSA

Monte Carlo singular spectrum analysis (MC-SSA) is a methodology for discriminating between various components of the time series, particularly between components containing meaningful information and other components containing mostly noise. This problem is especially important in process engineering applications, such as modelling, control, data validation and filtering. 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. These stochastic processes have frequency spectra that decrease monotonically with frequency and are often referred to as warm-coloured.

Generally speaking, MC-SSA involves a null hypothesis against which the data are tested, as well as a discriminating statistic, such as the autocorrelation, the correlation dimension or some other discriminating property of the data. The data are first assumed to belong to a specific class of dynamic processes, e.g. 1st order autoregressive processes or more broadly stationary linear Gaussian processes in general, perhaps distorted by some nonlinear measurement system (sensor). Surrogate data are subsequently generated from this process and various statistics are calculated from both the surrogate and the original data (Theiler et al., 1992). The calculated statistics of the surrogate and the original data can then be compared according to the null hypothesis, which has been postulated to say that the process that has generated the original data is of the same class as the system that has generated the surrogate data. The null hypothesis will be rejected if the discriminating statistic of the surrogate data differs by more than a predetermined margin from that of the original time series, as will be discussed in more detail later.

These concepts can be better illustrated by a simple example.

Time series **x** has been generated by a first order autoregressive model

$$x_t = 0.92x_{t-1} + \varepsilon_t \qquad 3.\ 13$$

and three surrogate series were generated by just randomly shuffling the observations of time series **x**. These surrogate sets are therefore of the data class of random white noise. The resulting time series, **x**, and surrogate series are supplied in Table 3. 1. The hypothesis postulated that the time series has been generated by a purely random white noise process and the autocorrelation coefficient was used as a test statistic. For a purely random white noise process, the autocorrelation of the data should be very low, with the correlation reaching zero very quickly.

Table 3. 1 Values for artificial first order autoregressive time series, x and the three random surrogates generated from the time series.

$\mathbf{x}_t$	Surrogate 1	Surrogate 2	Surrogate 3
0.500	0.371	0.485	0.213
0.485	0.500	0.394	0.371
0.450	0.359	0.213	0.384
0.425	0.194	0.228	0.500
0.394	0.390	0.500	0.188
0.390	0.450	0.282	0.425
0.371	0.220	0.207	0.320
0.384	0.213	0.210	0.210
0.359	0.384	0.384	0.183
0.320	0.183	0.183	0.220
0.282	0.188	0.194	0.359
0.235	0.210	0.450	0.207
0.228	0.394	0.320	0.282
0.220	0.207	0.425	0.228
0.207	0.282	0.220	0.485
0.194	0.425	0.359	0.235
0.188	0.228	0.235	0.394
0.183	0.320	0.188	0.194
0.210	0.220	0.371	0.390
0.213	0.485	0.390	0.220
0.220	0.235	0.220	0.450

The autocorrelation functions of the time series and the three surrogates are illustrated in Figure 3.2. It can be seen that the autocorrelation of the first order autoregressive series is significantly higher than that of the random, white noise series, leading to the conclusion that the null hypothesis should be rejected.

The algorithm and simple example described above can be summarized in Figure 3.3 with a simplified flow diagram of the performance of Monte Carlo SSA, before the technique will be discussed in more detail in the rest of the section.

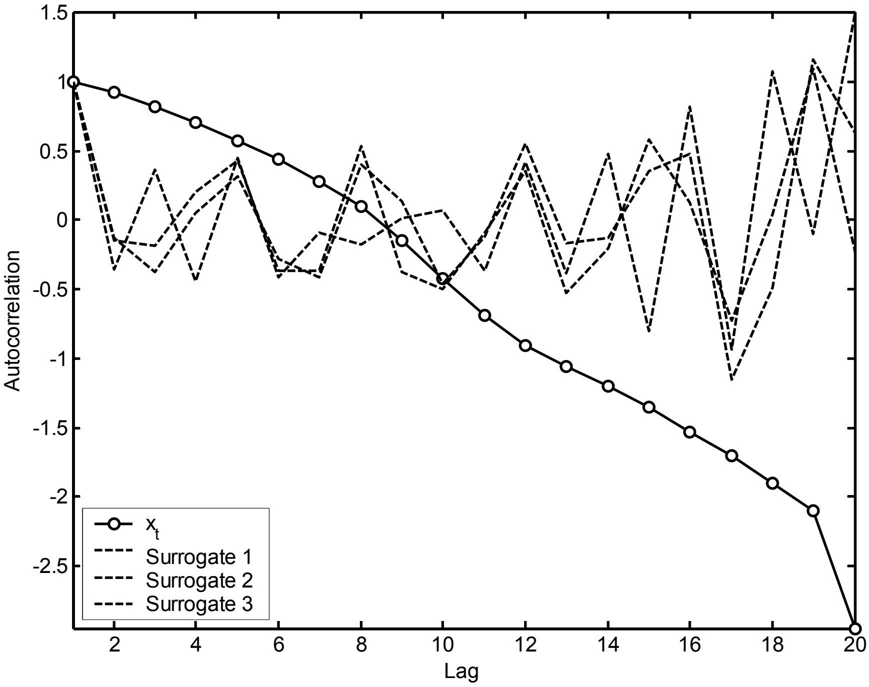


Figure 3.2 Autocorrelation function of artificial series, x, and three random, white noise surrogates generated from this series.

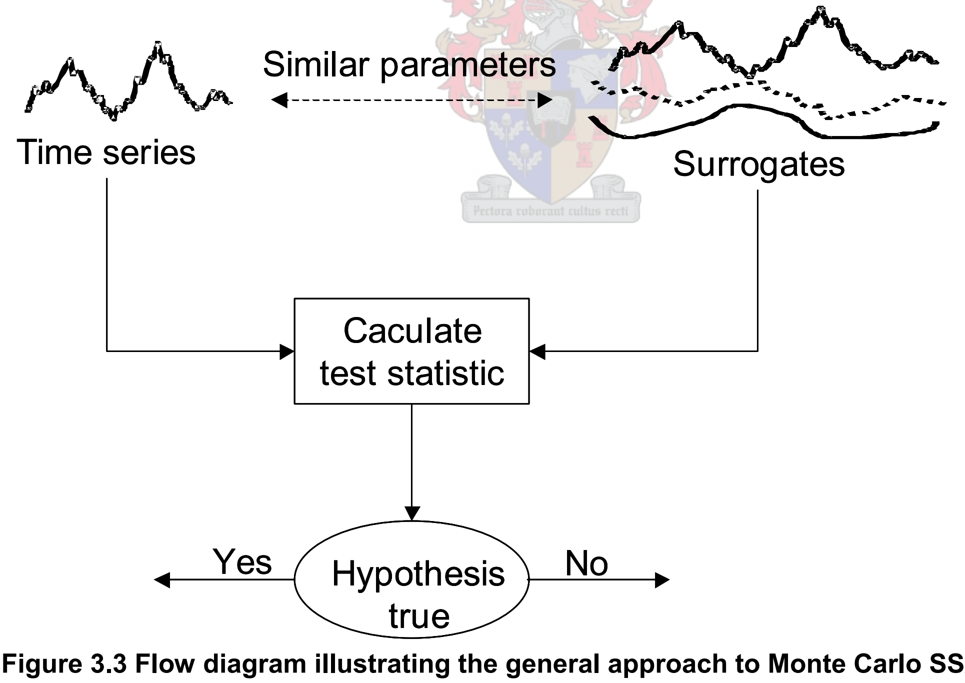


Figure 3.3 Flow diagram illustrating the general approach to Monte Carlo SSA in that surrogate data sets with similar parameters than that of the original time series are generated and by using a test statistic, both are tested against the postulated null hypothesis.

3.3.1 Monte Carlo in the literature

As was mentioned earlier in chapter 2, the Monte Carlo technique has been applied quite often in relation with singular spectrum analysis. This section will be used to provide a more detailed discussion of the various implementations of MC-SSA in the work done by other researchers.

The first work that mentioned using Monte-Carlo simulations to generate confidence limits for the discriminating statistic was Barnard (1963), with further developments made by Hope (1968), Besag and Diggle (1977), Hall and Titterington (1989), Noreen (1989), Fisher and Hall (1990) and Tsay (1992).

Theiler et al. (1992) did not initially use Monte Carlo analysis as such, but provided a very useful overview on the method of surrogate data to use as a test for nonlinearity. He presented a comprehensive discussion on the various levels of hypothesis testing, different test statistics that can be used and the available methods by which to generate the surrogate data sets.

This work was then extended in (Theiler and Prichard, 1996) where the use of the Monte-Carlo method for hypothesis testing was investigated. They pointed out that the questions typically asked about data sets in relations to hypothesis testing are:

- Is the distribution of the data non-Gaussian?
- Is the mean of the data significantly nonzero?
- Are there any temporal correlations?
- Is there any nonlinear structure in the temporal correlations?
- It the time series chaotic?

The null hypothesis is formulated to accept an answer of ‘no’. This is also the default answer by lack of any contrary proof. A discriminating test statistic is used and this test statistic is then evaluated to determine if it falls within the bounds that would be expected if the null hypothesis were true. As was explained earlier, Monte-Carlo analysis is based on the principle of calculating values for the discriminating test statistic for a great many realizations of the null hypothesis. This collection of estimates is then used to determine the boundaries for the test statistic.

When addressing the questions of whether a time series is non-Gaussian and whether there are any temporal correlations, Theiler and Prichard (1996) suggest two techniques by which the Monte Carlo realizations can be generated. They named these approaches typical and constrained realizations. They also suggested typical hypotheses to use when testing for different properties in the data. For instance, when trying to determine whether the time series is nonlinear, the hypothesis should be that the data had arisen from a linear stochastic process. For this hypothesis, two different approaches can be used for the surrogate data.

The first method to generate the surrogate data is to fit a linear model to the original data series and to then use different realizations of Gaussian white noise for the residual terms and thereby reconstruct the surrogate data from the linear model. The linear model approach could use either an autoregressive moving average (ARMA), a purely autoregressive (AR) or purely moving average model to simulate the linear stochastic processes.

The second approach to obtain the surrogate data would be to take a Fourier Transform of the data, randomise the phases and then invert the transform again. Both techniques have advantages and disadvantages and one should consider the practical trade-offs when deciding which technique to use. In the terms of the two techniques mentioned earlier, the ARMA method would be a typical realization-approach and the Fourier Transform would generate constrained realizations. It should be noted that the sample Fourier spectrum obtained from Fourier Transform of the original data is a poor estimator of the underlying frequency spectrum. However, as long as the spectrum is not the main focus of the calculation, this would not necessarily present a flaw in the Fourier Transform based method of calculating the surrogate data. The biggest advantage of ARMA is in fitting the model, where as Fourier Transform is more useful for fitting the data. Theiler and Prichard (1992) mentioned that if one wanted to calculate error bars or confidence limits rather than test the null hypothesis for a certain test statistic, the Fourier Transform method would be very undesirable to determine the surrogate sets and the ARMA method should rather be used.

It is important to distinguish between the different problems of the estimation of confidence intervals and testing the null hypothesis (Theiler and Prichard, 1996). For the estimation of confidence intervals, a statistic of some intrinsic value, such as the mean or the fractal dimension, is calculated for the data and certain ‘error bars’ for the calculated value is specified. These confidence limits enclose, within certain probability limits, the actual mean of the true underlying distribution. However, when the null hypothesis is tested, it is done for a carefully specified hypothesis and the aim is to determine whether the data are actually consistent with this hypothesis.

Allen and Smith (1996) used Monte-Carlo SSA to detect irregular oscillations in the presence of coloured noise. They identified the need for a statistical tool by which discrimination could be made between possible oscillation signals and other signals present in the time series. The null hypothesis used was that of the data being coloured noise and the basic formalism of SSA provided a natural test for modulated oscillations against this hypothesis. Even though the presence of coloured noise will be discussed in significantly more detail in a later section (section 4.7), it would perhaps be appropriate at this stage to just explain the term. The name 'coloured' or 'red' noise is a popular term with no mathematical significance at all, that has been assigned to noise series with certain characteristics. This term is related to noise series or phenomena of which the power spectral densities are proportional to $1/f^{\beta}$ (Addison, 1997, Aldrich, 2002).

Monte Carlo SSA was tested for three different types of artificial data. The first situation was where the power spectral characteristics of the noise were known prior to the analysis, in the second situation it was tested whether the data consisted of only white or coloured noise and lastly a composite hypothetical noise model was tested by which it was assumed that some deterministic components were found in the data and the aim was to determine if the remainder of the components were noise.

According to Allen and Smith (1996), there can be two different approaches when Monte Carlo hypothesis testing is applied to the analysis of nonlinear systems. One approach is that the null hypothesis should be well understood and the other is that the null hypothesis should be physically interesting. For example, if it is known beforehand, due to the physical situation, that a system could not appear to be white noise, one does not gain any new information from rejecting the white noise null hypothesis. However, this situation is not so simple for first order autoregressive processes or so-called red or coloured noise. The output from many systems, both in the engineering industry and in many other research areas, is often indistinguishable from purely red-noise systems. The complexity of the test procedure therefore depends on how much prior knowledge about the properties of the noise is available.

The application of Monte Carlo SSA was further extended by Palus and Novotna (1998) in that it was also used to evaluate and test the regularity of dynamics, in addition to the normal test performed by inspecting the eigenvalues. This was done by evaluating the SSA modes against the coloured noise null hypothesis. This approach resulted in enhanced test sensitivity and reliability in detecting the relevant modes.

3.3.2 General approach to Monte Carlo SSA

The general approach in Monte Carlo SSA or surrogate data analysis has been illustrated in Figure 3.3 and can be summarized as follows.

i. Generation of sets of surrogate data, each similar to the original time series, i.e. of the same length and statistically indistinguishable from the original time series with regard to certain specified characteristics.

ii. Calculation of a discriminating statistic (test statistic) for the measured time series and the sets of surrogate data. Any statistic quantifying some aspect of the time series can be used, such as forecasting error, largest Lyapunov exponent, correlation dimension etc, although some test statistics are more suitable for certain analyses, as will be discussed shortly.

iii. Setting up of a hypothesis that there is no difference between the discriminating statistics of the original time series and the surrogates (null hypothesis).

iv. Testing of the null hypothesis based on the values of the discriminating statistics and acceptance or rejection of the null hypothesis.

Implementation of the approach therefore requires specification of the measures by which the surrogate data and the original time series should be identical (i), the specification of a discriminating statistic (ii) and specification of the test parameters (iii-iv).

The noise data found embedded in a time series signal can be considered as a stochastic process consisting of random variables. These values evolve in time according to certain probabilistic rules, which means that a noise process is essentially a collection of random variables that have been ordered in time. The statistical moments of a noise process are defined with respect to this distribution of the random variables u_1, u_2, . . . , u_{NT}.

More formally, let $\mathbf{x} \in \Re^n$ be a time series consisting of n observations, ψ a specific hypothesis, $\Im_\psi$ the set of process systems consistent with the hypothesis, and $\tau: \Re^N \rightarrow \mathbf{U}$ a statistic that will be used to evaluate the hypothesis ψ that $\mathbf{x}$ was generated by some process $\Im \in \Im_\psi$.

It is then possible to use this statistic to discriminate between the original data **x** and the surrogate data $\mathbf{x}_s$ consistent with the hypothesis given by the probability density of τ, given $\Im$, i.e. $\mathbf{p}_{T,\Im(t)}$.
For the time series $\mathbf{x} \in \Re^n$ mentioned above, the first two steps of basic SSA can be performed normally, just as discussed previously in paragraph 3.1. The algorithm described in section 3.3.3 can now be used to generate a collection of *p* surrogate data sets $\mathbf{x}_i^{surr}$ (i = 1, 2, . . . , *p*), all of the same dimension as time series **x,** which is then analysed using SSA with exactly the same parameters as the original time series.

3.3.3 Generation of surrogate data

As mentioned before, surrogate data are nondeterministic artificially generated data that mimic certain features of an observed time series. For instance, surrogate data may have the same mean, variance, Fourier power spectrum or autocorrelation function as the measured time series. The idea is to see whether the measured time series, which are similar to the surrogate data, have the same value of the selected measure.
A number of techniques are available to generate surrogate data. The power of these methods to reject the null hypothesis is dependent on a number of key issues such as the computational complexity of the algorithm and the accuracy with which the statistical properties of the data are being analysed (Broomhead and King, 1986).
There are three general algorithms by which surrogate data are generated (Popivanov and Mineva, 1999).

i. The simplest algorithm for generating surrogate data is by random shuffling (shuffled surrogates). The surrogate data thus consist of a random permutation of the original data. This method preserves the distribution of the original time series and is consistent with the null hypothesis of Type 0 (identically, independently distributed noise). However, it is seldom used, as other tests for white noise (normally distributed data) are commonly available.

ii. The second algorithm was first introduced by Theiler et al. (1992). This algorithm generates Fourier-transformed (FT) surrogates and it is based on the null hypothesis that the data are linearly filtered noise (Type 1). This algorithm can only be applied to data that are normally distributed (Broomhead and King, 1986). Fourier transformed surrogates have the same power spectra as the original data, but by randomising the phases, the nonlinear structures of the data are removed.

iii. The third algorithm was also first presented by Theiler et al. (1992) and in this case the linear correlation between the data and the possible static nonlinear transformation is retained.

During the performance of Monte Carlo SSA in this work, both the amplitude adjusted Fourier transform (AAFT) (algorithm (ii)) and the iterative amplitude adjusted Fourier transform (IAAFT) algorithms were used. The AAFT algorithm was developed with the purpose of testing the null hypothesis that a monotonic non-linear transformation of a linear Gaussian process has generated the observed time series. The algorithm consists of the following four steps (Popivanov and Mineva, 1999):

i. Generating a Gaussian time series with elements developed independently from the Gaussian pseudorandom number generator.

ii. Reorder the time series so that the new time series has the same spectrum as the real data, only with a Gaussian distribution. This is achieved by reordering the generated time series so that the ranks of the samples of the Gaussian and real time series coincide.

iii. Produce the first surrogate by applying the Fourier transform algorithm to the reordered time series.

iv. Reorder the real data with respect to the first surrogate to obtain the final surrogate. This surrogate will preserve both the power spectrum and the amplitude distribution of the real data.

An iterative method (IAAFT) is described in which the power spectra of the AAFT surrogate is adjusted back to that of the original data, in a series of iterations, before the distribution is rescaled back to that of the original data. The algorithm was first introduced by Vautard et al. (1992). It is important to note that the IAAFT algorithm makes no assumptions with regard to the form of the measurement function (Vautard et al., 1992). It has been argued (Rozynski et al., 2001) that the test with IAAFT surrogates is “almost too strong”.

Since a linear stochastic time series is completely characterized by its Fourier spectrum (or autocorrelation function), it should be a minimum requirement that the surrogate data and the original time series have the same power spectra (or autocorrelation functions).

3.3.4 Choosing a test statistic

Different processes might give rise to power spectra and distributions that would not necessarily be unique (Vautard et al., 1992). When a test statistic is thus considered, it is essential that it should be able to discriminate between the variations in the power spectra and distributions and deviations from the null hypothesis.
The test statistic, T, can be considered as a single number, estimating the characteristics of the data and its variations in such a way as to be able to decide whether the time series is consistent with the null hypothesis (Vautard and Ghil, 1989, Vautard et al., 1992).
The correlation dimension and Lyapunov exponent are considered to be pivotal test statistics, since the probability distribution of these quantities would be the same for all processes, regardless of the source of the noise of the estimated model. The Lyapunov exponent has been shown to be misleading in the presence of noise and therefore the correlation dimension has gained favour as the pivotal statistic of choice (Theiler et al., 1992, Takens, 1993, Theiler, 1995, Theiler and Prichard, 1996).
With specific application to singular spectrum analysis, the discriminating statistic can either be the eigenspectrum of the time series or the shape of the eigenvectors (Elsner and Tsonis, 1996). For the purpose of this investigation, it was decided to use the spectrum of eigenvalues rather than the eigenvector shape for a number of reasons. Firstly, the eigenvector shape criterion can only be considered reliable for longer time series and secondly the theoretical separation of frequencies for eigenvector pairs becomes invalid if the deterministic component is contaminated by red noise.
This particular test allows one to compare the overall shape of the singular spectrum of a given time series with the singular spectrum of a number of artificially generated surrogate data sets, but does not take into account the frequencies of the corresponding eigenvectors. By investigating whether the eigenvalues of the original time series fall within the confidence limits generated by the eigenvalues of the surrogate data sets, one can reject or fail to reject the null hypothesis.

a) Correlation dimension

The correlation dimension is defined by

$$d_{\text{corr}} = \lim_{\varepsilon \to \infty} \frac{\log C(\varepsilon)}{\log \varepsilon} \qquad 3.\ 14$$

where $C(\varepsilon)$ is

$$C(\varepsilon) = \lim_{N \to \infty} \frac{1}{N^2} C_N(\varepsilon) \qquad 3.\ 15$$

where N is the number of available points and $C_N(\varepsilon)$ is the number of pairs of points on the attractor whose distance from one another is less than ε (Golia and Sandri, 2001).

b) Confidence limits

The confidence limits of the eigenspectrum provide an area within which a certain percentage of the eigenvalues of the original time series must fall, if the surrogate series that specified the confidence limits have been generated by the same type of process that generated the original data. These confidence limits (φ) for each eigenvalue can be calculated by:

$$\varphi(\lambda_i) = \bar{\lambda}_i \pm \sigma_\lambda t_{\alpha/2} \qquad 3.\ 16$$

where $\bar{\lambda}_i$ is the average of all the eigenvalues at the same position (i) in the eigenspectra that were obtained from the surrogate series (i.e. the average value of the first eigenvalue, the average value of the second eigenvalue from the surrogates, etc.), σ_λ is the standard deviation of the eigenvalues that were used to calculate the average eigenvalue and t is the Student t-score, depending on the confidence level, α, and the number of surrogate series that was used to generate the eigenvalues. If the two terms in equation 3.16 is added, the upper confidence limit is obtained and when they are subtracted, the lower confidence limit is calculated.

3.3.5 Classes of hypothesis

The data set is assumed to belong to a class of null hypothesis, which may or may not adequately explain the data. The surrogate data generated should mimic the specific class of null hypothesis.

The first step in analysing a time series, using hypothesis testing, is to determine whether there exist any dynamics, i.e. whether the data are simply white noise or whether they are correlated. The simplest null hypothesis in this case is that the time series x_i, i = 1,…,N, is uncorrelated white noise with an unspecified distribution. When it is assumed that the data has a Gaussian distribution, a surrogate data set can be generated, consisting out of the original data, but that would be random (any temporal correlation destroyed) in every other sense. The surrogate will by construction have the same amplitude distribution as the original data.

If serial correlations have been found in the time series (rejection of dependence, Type 0) it would be important to know what the nature of the correlations are (Elsner and Tsonis, 1996). Possibly the simplest way is to explain the observed structure by linear two-point autocorrelations.

A corresponding null hypothesis (Type 1) would be to assume that the observed time series $\{x_i\}$ was generated by a normal, linear stochastic process. The residuals of a linear fit could be tested for correlation, but it has been shown to be more suitable to test the time series directly for possible non-linear correlation (Rozynski et al., 2001).

The most general null hypothesis (Type 3) would be one in which the possibility could be included that the data were measured by a static (instantaneous) invertible measurement function *h*, which would be independent of time (Vautard et al., 1992).

A time series x_N is said to be consistent with this null hypothesis if there exists any underlying Gaussian linear stochastic signal y_N such that $y_N = h(x_N)$. Such signals as y_N would share the same power spectrum and amplitude distribution as the original signal x_N (Vautard et al., 1992).

An important prerequisite of the amplitude adjusted Fourier-transformed (AAFT) surrogates is that *h* needs to be monotonic (invertible) i.e. h^{-1} exists. The iterated and corrected AAFT as well as the DFS algorithms however are not dependent on this assumption on the invertibility of *h*.

The final surrogates y_s have by construction the same amplitude distribution, for finite N, but do not necessarily have the same sample power spectra (Vautard et al., 1992).

3.4 Nonlinear singular spectrum analysis

Principal component analysis, and therefore singular spectrum analysis, are both linear techniques, which often causes nonlinear components of time series to be overlooked. However, due to recent advances in time series analysis, many traditional multivariate and time series analysis techniques have been expanded to allow the study of nonlinear components (Schreiber, , Schreiber, 1998, Schreiber, 2000). These advances have been done mainly in the field of neural networks on nonlinear principal component analysis (NLPCA) (Kramer, 1991, Kirby and Miranda, 1996, Hsieh, 2001, Hsieh and Wu, 2001a, Hsieh and Wu, 2001b, Hsieh and Wu, 2002, Hsieh and Hamilton, 2003, Newbigging et al., 2003) and localised principal component analysis (LPCA) (Aldrich, 2002).

3.4.1 Nonlinear Principal Component Analysis

In paragraph 3.1.1, traditional principal component analysis was described as a technique in which a hyperplane is fitted to the data in an attempt to approximate the data being investigated with a hyperplane that explains the largest possible amount of variance. However, where the subspace in which the data are embedded shows significant curvature, the linear hyperplane may not be a good (compact) approximation of the subspace. Recently developed neural networks will rather approximate the data using a continuous curve, thereby allowing the classification of nonlinear components too. This idea can be illustrated by means of the sample graph in Figure 3.4. Both subspaces are one-dimensional, but case 2, shown in Figure 3.4(b) shows significant curvature and can only be approximated by two components or one nonlinear component.

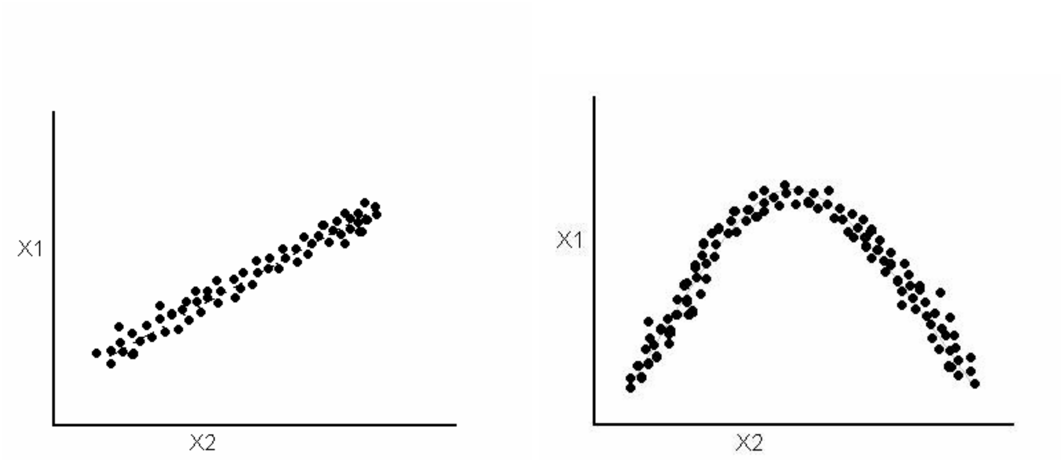


Figure 3.4 Illustration of the differences between (a) linear and (b) nonlinear time series where (a) can be fitted by a linear hyperplane, but (b) requires a curved approach.

Although many different methods can be used to extract nonlinear principal components, only two will be considered here, viz. by use of auto-associative neural networks and by means of localized PCA. Two different approaches can be followed to perform nonlinear principal component analysis with auto-associative neural networks. The first approach would be simply to perform nonlinear principal component analysis on the original time series, as one would apply normal principal component analysis. The second approach would be that favoured by Hsieh (2001), Hsieh and Wu (2001a), Hsieh and Wu (2001b), Hsieh and Wu (2002), Hsieh and Hamilton (2003) and Newbigging et al. (2003), in that normal principal component analysis is performed first and the principal components that have been extracted are then used as input for the nonlinear principal component analysis. The purpose behind this is to extract the nonlinear components of the data from the linear principal components that have been identified.

The technique involves a bottleneck process and is illustrated by Figure 3.5 (Hsieh and Hamilton, 2003).

The data being used as input for the network are in the form $\mathbf{x}(t) = [x_1, \dots x_l]$, where each variable, x_i ($i = 1, \dots l$) is a separate time series of length n. If the first approach is followed, the separate time series are the individual columns from the lagged trajectory matrix. For the second approach, the matrix of principal components is the input matrix. The information from the input nodes is then mapped by the neural network through the bottleneck, onto the output **x'**. This output is then considered to be the nonlinear principal component that has been extracted from the data that were used as input.

The nonlinear SSA (NLSSA) method that has been developed by Hsieh and Wu (2001b) is based on NLPCA and the second approach to the input matrix. The data are pre-filtered by performing the first three steps of singular spectrum analysis, described in paragraph 3.1 and only retaining a reduced number of principal components. The principal components of the first few leading SSA modes are then used as input ($x_1, \dots x_l$) for the NLPCA network.

The neural network used to perform NLPCA consists of three hidden layers of variables or nodes between the input and output layers (Figure 3.5). The number of nodes in each of the layers depends strongly on the time series being analysed and the nature of the network being built.

The number of nodes in the input layer is equal to the number of principal components (l) that are used as input. A transfer function, f_1, maps from the **x** to the first hidden layer, known as the encoding layer. The encoding layer is represented by a column vector of length m, $\mathbf{h}^{(x)}$, with elements

$$h_k^{(x)} = f_1((\mathbf{W}^{(x)}\mathbf{x} + \mathbf{b}^{(x)})_k) \qquad 3.17$$

where $\mathbf{W}^{(x)}$ is defined as a m x l weight matrix, $\mathbf{b}^{(x)}$ is a column vector of length m containing the bias parameters and $k = 1, \dots, m$. In a similar fashion, the second transfer function f_2 will map from the encoding layer to the bottleneck layer. The bottleneck layer is specified to contain a single node.

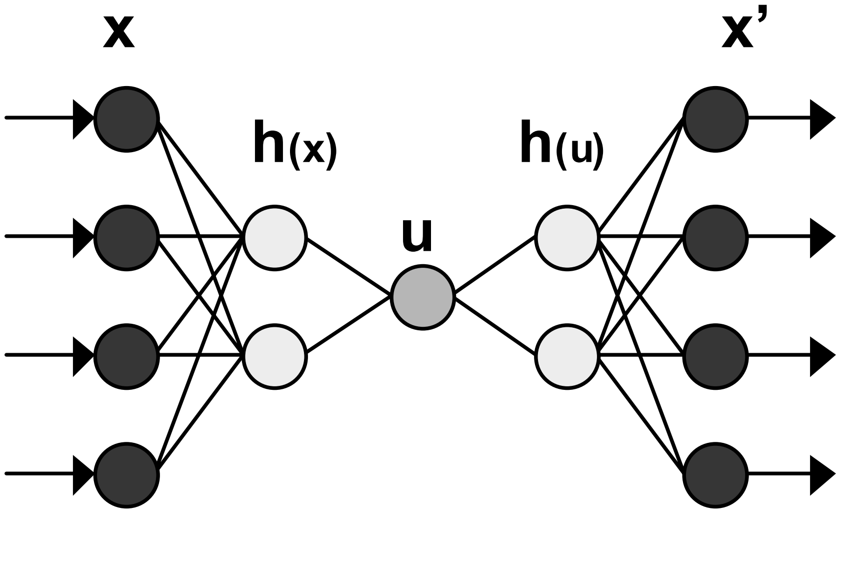


Figure 3.5 Layout of neural network for nonlinear principal component analysis.

This bottleneck node represents the nonlinear principal component, *u*

$$u = f_2(\mathbf{w}^{(x)} \cdot \mathbf{h}^{(x)} + \bar{b}^{(x)}) \qquad 3.18$$

The transfer function f_1 that is used to encode the data is usually nonlinear, such as the hyperbolic tangent or the sigmoidal function, while f_2 is generally taken to be the identity function.

A third transfer function now maps from *u* to the decoding layer $\mathbf{h}^{(u)}$, which is the final hidden layer

$$h_k^{(u)} = f_3((\mathbf{w}^{(u)}u + \mathbf{b}^{(u)})_k) \qquad 3.19$$

This is then followed with the final mapping by f_4 to the output layer, **x'**.

$$x_i' = f_4((\mathbf{W}^{(u)}\mathbf{h}^{(u)} + \bar{\mathbf{b}}^{(x)})_i) \qquad 3.20$$

The length of the output column vector is equivalent to that of the input column vector, i.e. there are *l* nodes in both the input and output layers.

The specification of the mapping parameters is done by minimizing the cost function $J = \langle \| \mathbf{x} - \mathbf{x}' \|^2 \rangle$. To achieve this, the optimal values of $\mathbf{W}^{(x)}$, $\mathbf{b}^{(x)}$, $\mathbf{w}^{(x)}$, $\mathbf{w}^{(u)}$, $\mathbf{b}^{(u)}$, $\mathbf{W}^{(u)}$, $\mathbf{b}^{(u)}$ need to be determined. By minimizing the cost function, the mean squared error (MSE) between the original data **x** and the NN output **x'** is minimized.

For this analysis, the hyperbolic tangent function was used for f_1 and f_3 and the identity function for f_2 and f_4. If the constraint $\langle u \rangle = 0$ is imposed, the total number of free weight and bias parameters become $2lm + 4m + l$.

It can therefore be seen that the choice of *m*, which is the number of nodes in both the encoding and decoding layer is quite a significant decision. The principle by which this choice is made, is parsimony. For a large *m*, the network will have a high nonlinear modelling capability, but simultaneously, the risk of over fitting the model (to also fit the noise in the model) is increased. If $m = 1$ and an identity mapping function is used for f_4, it is implied that all x_i' are linearly related to a single hidden neuron, which is equivalent to only a linear relation between the x_i' variables. This means that in order to obtain a nonlinear solution, $m \geq 2$.

A fixed network configuration has been used for all the analysis in this thesis. The numbers of input and output nodes were variable, depending on the nature of the time series being analysed, but for the hidden layers a [2 1 2] configuration was used.

Once the network has trained on the data, the output of the network could be unembedded in a similar fashion to how the unembedding was done for normal SSA after the time series was reconstructed. However, the time series obtained from the unembedding is only a single nonlinear principal component and not a whole reconstructed series. By subtracting the output matrix of the network from the input, one obtains a residual matrix, which would contain the remainder of the nonlinear components that have not yet been extracted. If this residual matrix is then sent through the network again, the next nonlinear component could be extracted. This procedure can be repeated until all the desired components have been extracted. The nonlinear reconstructed series can then be obtained by adding all the desired components. This is illustrated diagrammatically by Figure 3.6.

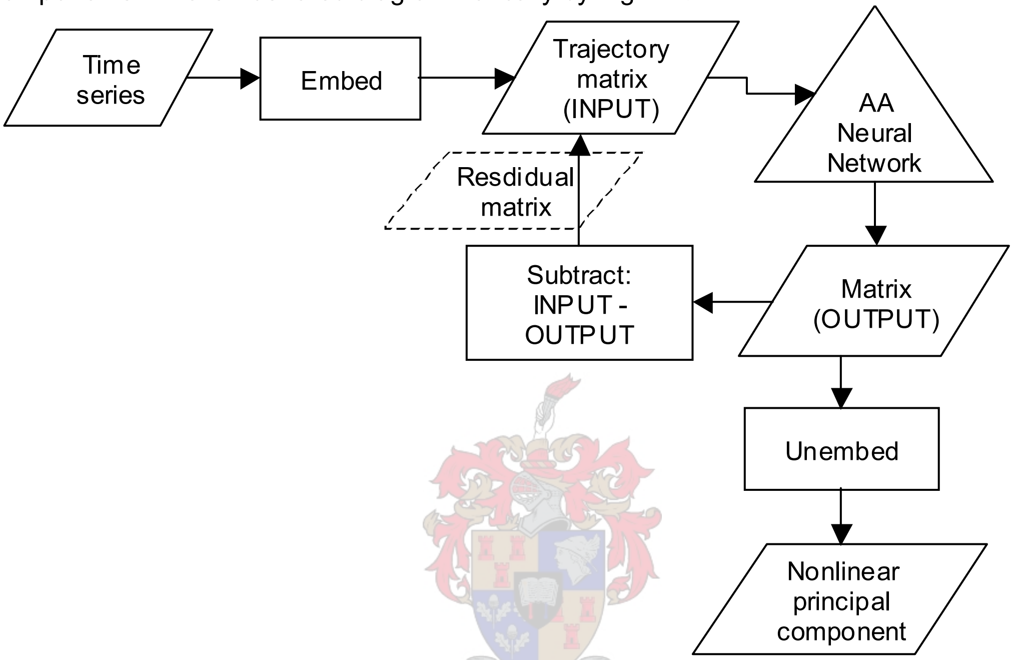


Figure 3.6 Flow diagram of algorithm of auto-associative neural network analysis.

3.4.2 Localized Principal Component Analysis

The second technique with which nonlinear time series can be analysed is localized principal component analysis. The main principle behind this approach is to divide the nonlinear time series $\mathbf{y} = [y_1, \ldots, y_N]$ into a number of localized sections, $\mathbf{y}_i$ ($i = 1, \ldots, k$), each of which can then be treated as an independent linear time series.

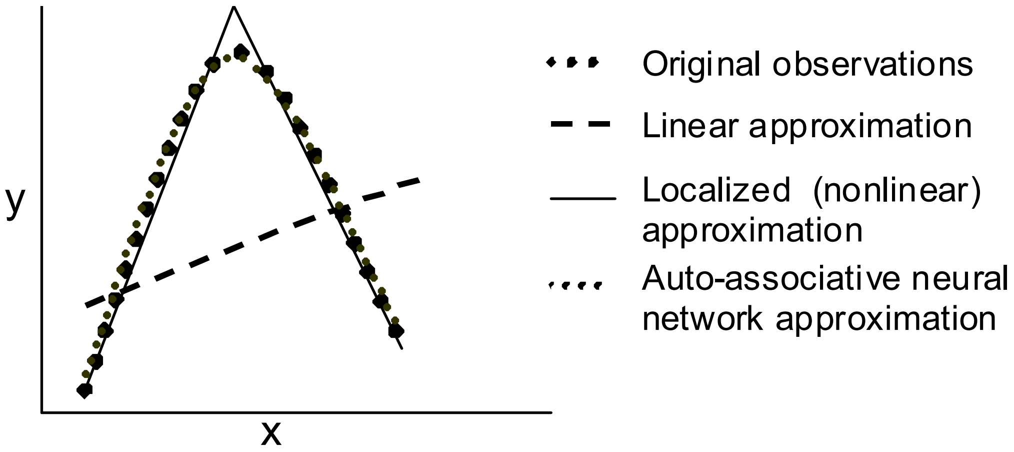


Figure 3.7 Illustration of localized nonlinear approach to singular spectrum analysis compared to the linear approximation.

The benefit of following this approach can be seen from the elementary example in Figure 3.7. If the parabola were to be approximated by a linear method, a significant part of the structure of the data will be lost. However, by dividing the time series into two separate smaller series, a much closer approximation of the true nature of the series can be obtained, even with linear techniques, with the best approximation in Figure 3.7 obtained by using auto-associative neural networks.

The division of the time series can be done either by splitting all the observations into a number of parts all of equal length or, more appropriately, the time series can be sectioned at the points of major changes in the trend of the data by doing visual inspection.

Regardless of which approach is followed to segment the data, the rest of the analysis can be performed in exactly the same manner as that described in section 3.1. Once all the parts of the time series have then been reconstructed separately with a reduced number of principal components, the original time series can be obtained again by simply combining the reconstructed parts in order. This approach is illustrated diagrammatically in Figure 3.8.

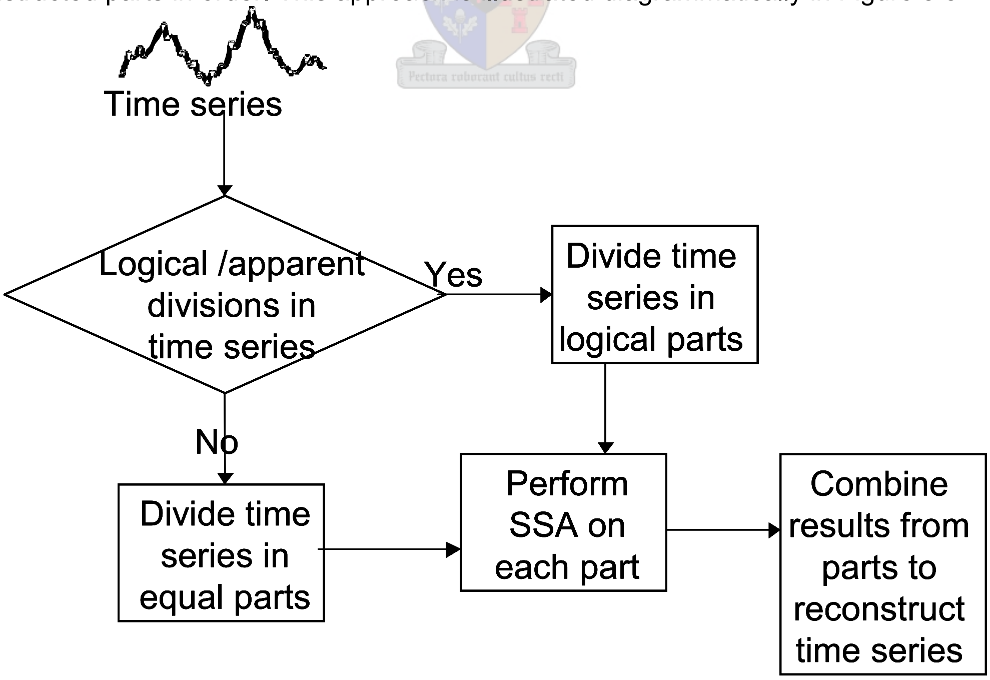

Time series
Logical /apparent divisions in time series
Yes
No
Divide time series in logical parts
Divide time series in equal parts
Perform SSA on each part
Combine results from parts to reconstruct time series

Figure 3.8 Diagrammatical presentation of steps involved in localized SSA approach to time series.

3.5 Other approaches to SSA

A number of other approaches to SSA have also been found in the literature. However, these approaches fall outside the scope of the current research and will only be mentioned in the interest of completeness.

3.5.1 Multi-scale singular spectrum analysis

This technique has been applied by a number of researchers, with (Yiou et al., 2000) providing a detailed discussion on the development of the technique and how it is implemented.

The main focus of multi-scale SSA is in the study of nonstationary time series, which is the reason for the implementation of the multi-scale ideas from wavelet analysis. Multi-scale SSA is very similar to localized SSA (section 3.4.2) in that only a certain window of the time series is evaluated at a time. However, in multi-scale SSA the window is moving along the time series, whereas for localized SSA a number of separate, fixed windows are analysed simultaneously.

3.5.2 Random lag singular spectrum analysis

Most of the work on random lag SSA has been done by Varadi et al. (1999) and Varadi et al. (2000) on solar oscillations. The only variation from standard SSA is to rather use random lags when dealing with the autocorrelations, instead of using a fixed lag (usually one) for the whole time series.

The advantage of using random lags rather than a fixed lag, is that it allows one to carry out SSA for wide-frequency bands.

3.5.3 Approximate projectors

Moskvina and Schmidt (2003) proposed a completely different approach to SSA in that an approximate projector, *p*, is determined. This projector is the orthogonal projector onto the subspace spanned by the eigenvectors and if only a reduced number of vectors are to be retained (as is usually the case with SSA), this projector only has to span a reduced subspace. When the reduced projector is multiplied with the original trajectory matrix, the reconstructed time series can be obtained, without having to explicitly perform PCA. Moskvina and Schmidt introduced an algorithm by which this projector can be estimated from a polynomial approximation of the characteristic function of the data.

4 FILTERING OF DATA WITH SSA

One of the most fundamental applications of singular spectrum analysis is the filtering of time series for a number of purposes, such as visual presentation, modelling of the data or system identification. In this chapter, strategies towards system identification with singular spectrum analysis are proposed. With these strategies the data are smoothed prior to fitting models, which can lead to a significant improvement over models built on the original data.

These filtering techniques are applied to a number of different data sets for illustration purposes. The first data set is the simulated second order response of the flow between two noninteracting tanks in series, the second case study is the observations from a simulated carbon-in-leach gold leaching process. The next two case studies are from a copper flotation plant, with the third being the froth measurements and the fourth example the measurements of the precious metal recoveries. The last case study was concerned with features measured from a froth monitoring system on a lead flotation plant. The chapter is ended with a discussion on the difference in the behaviour of SSA when handling white vs. red noise, using a simulated sine wave as illustration.

4.1 Identification of process system

A predictive model is used to identify the underlying process dynamics represented by the time series being investigated. Different classes of models can be used to fit the data, whether local or global. In this chapter, neural networks will be used to identify the process systems. As was described earlier in section 3.4.1, neural network models typically consists of an input layer, which is presented with variables derived from the embedded time series, a chosen number of hidden layers, with a specified number of nodes in each layer, and an output layer, which generates predicted values for the time series. The neural network models built for the identification of the process system consisted of only one hidden layer, using sigmoidal activation functions and with the number of nodes in the layer being variable.

Care should be taken in determining the number of hidden nodes, as the more nodes there are, the more accurately the data can be modelled, but the less general the model will be. For application in this chapter, the modified Schwartz information criterion (Schwartz, 1978, Barnard et al., 2001) was used to determine the optimal number of hidden nodes. It balances the trade-off between model accuracy and complexity by encoding the model parameters (weights of the neural network) and prediction errors as bit streams of information. The more complex the model and the larger the prediction error, the more bits would be required in the encoding, so that the model structure corresponding with the lowest MSIC was considered optimal. The objective function can be summarized as follows

$$\text{MSIC} = n\sum[\log(\text{MSE})] + [u \text{ x } \log(n)] \qquad 4.1$$

$$u = (\text{m}_{in} + 2\text{m}_{\text{out}} + 2)S \qquad 4.2$$

where u is the model order (number of model parameters), n is the number of samples in the training data set, MSE is the mean square error of prediction, S is the number of hidden nodes in the neural network, and m_{in} and m_{out} are the number of nodes in the input and output layers respectively, not counting bias nodes. The quality of the models was assessed by means of validation data sets not used during the development of the models.

The different strategies for the identification of process systems were all based on singular spectrum analysis and the different approaches can be summarized as follows:

Univariate embedding

i. Given a set of time series **Y** = [$\mathbf{y}_1$, $\mathbf{y}_2$, … $\mathbf{y}_p$], construct the trajectory matrix of each time series. This is done by embedding the data of each time series with a unity lag and embedding dimension equal to the lag index where the autocorrelation function of the time series is small (< 0.2). This yields the set of trajectory matrices $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_p]$, each with different dimensions in general, i.e. $\mathbf{X}_j \in \Re^{nj \times mj}$.

ii. Compute the lagged covariance matrices of **X** and decompose these matrices (**C**) into their principal component vectors (**P**) and scores (**T**), so that $\mathbf{X}_j = \mathbf{T}_j\mathbf{P}_j^T$, for j = 1, 2, … p.

iii. Reduce the dimensionality of the embedding if possible by retaining k ≤ p principal components to retain and reconstruct the original trajectory matrices from the reduced number of principal components, yielding $\mathbf{X}_{rec}$.

iv. Reconstruct the original time series from $\mathbf{X}_{rec}$ by unembedding, as explained previously. This gives the original set of time series **Y**, the reconstructed set of time series, $\mathbf{Y}_{rec}$, as well as the individual components of the time series, $Y^{\#} = [y^{\#}_1, y^{\#}_{2, \ldots} y^{\#}_k] = \{\mathbf{t}_j\mathbf{p}_j^T\}$, for j = 1, 2, … k.

Multivariate embedding

i. See step i above.

ii. Assemble the trajectory matrices of the individual time series, $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_p]$, so that $\mathbf{A} \in \Re^{nA}$ x ^{mA}, i.e. the matrices are joined row-wise, so that nA = min(nj), for j = 1, 2, … p and mA = m1 + m2 + …+ mp.

iii. Compute the lagged covariance matrix of **A** and use this matrix ($\mathbf{C}_A$) as a basis for the decomposition of **A** into its principal component vectors ($\mathbf{P}_A$) and scores ($\mathbf{T}_A$), so that $\mathbf{A} = \mathbf{T}_A\mathbf{P}_A^T$.

iv. Remove noise from the data by approximating **A** by the first k principal components and score vectors, i.e. $\mathbf{A}^* = \mathbf{T}_{A,k}\mathbf{P}_{A,k}^T$. (similar to step iii above). The number of principal components to be retained is determined via inspection of the eigenspectrum of **A**.

v. Separate **A*** into p different blocks of which the number of columns in the respective blocks equals that of the trajectory matrices of the p individual time series, i.e. $\mathbf{A}^* = [\mathbf{A}_1, \mathbf{A}_2, \ldots \mathbf{A}_p,]$.

vi. Reconstruct the original time series from the appropriate parts of **A*** by unembedding, as explained previously. This gives the original set of time series **Y**, the reconstructed set of time series, $\mathbf{Y}_{rec}$, as well as the individual components of each time series in the original set, $Y_i^{\#} = [y^{\#}_{i,1}, y^{\#}_{i,2, \ldots} y^{\#}_{i,kA^*}] = \{\mathbf{t}_j\mathbf{p}_j^T\}$, for i = 1, 2, … p and j = 1, 2, … kA*.

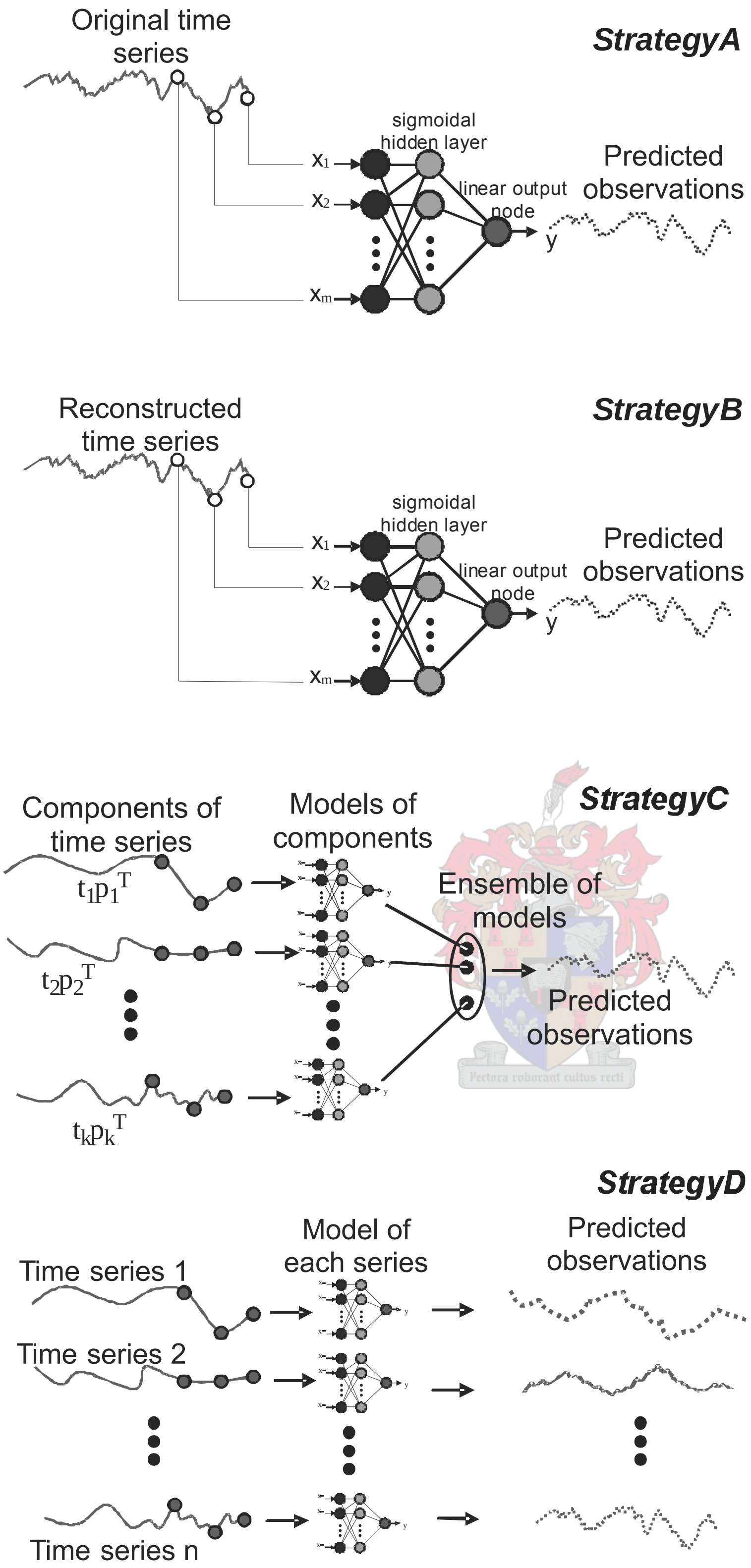


Figure 4.1 Modelling strategies (A) - (D), where the original time series is predicted in strategy A, the time series reconstructed by SSA predicted in strategy B, the individual components extracted by SSA predicted in strategy C and the reconstructions obtained from multichannel SSA predicted in strategy D.

The following four modelling strategies (depicted diagrammatically in Figure 4.1) were considered.

A. r-step ahead prediction of the original time series based on embedding of the original time series $\mathbf{y}(t+r) = f(\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots \mathbf{y}_{t-mA})$.

B. r-step ahead prediction of the original time series based on a model of the reconstructed time series $\mathbf{y}(t+r) = f(\mathbf{y^*}_{t-1}, \mathbf{y^*}_{t-2}, \ldots \mathbf{y^*}_{t-mA})$.

C. r-step ahead prediction of the original time series based on an ensemble of models of the individual $\mathbf{tp}^T$ components of each time series $\mathbf{y}(t+r) = f(\mathbf{y}^{\#}_{t-1}, \mathbf{y}^{\#}_{t-2}, \ldots \mathbf{y}^{\#}_{t-mA})$.

D. Similar to strategy B, except that it is based on the multivariate embedding ($\mathbf{A}^*$) of a set of time series.

4.2 Flow between two noninteracting tanks in series

4.2.1 Generation of series and performing SSA

To illustrate the use of singular spectrum analysis it was decided to start with a time series of which the characteristics are known beforehand. Consider the 2nd order response of the flow of two noninteracting tanks in series, the overall transfer functions of which can be described by

$$G(s) = \frac{1}{(0.38s + 1)(2.62s + 1)} = \frac{1}{(s^2 + 3s + 1)} \qquad 4.3$$

The actual response to a pulsed input (broken line) in Figure 4.2 is indicated by the solid line, while measurements are simulated ('+' markers) with zero mean Gaussian noise with a standard deviation of 0.1.

The autocorrelation function of the time series reached the point of weak to negligible correlation (< 0.2) at an index of 23, as can be seen in Figure 4.3. Based on this criterion, the trajectory matrix consisted of 23 columns of the time series, each copy delayed by a time step of one. This matrix formed the basis from which 23 principal components were extracted. The eigenvalues associated with each of the 23 principal components (eigenspectrum) are shown in Figure 4.4.

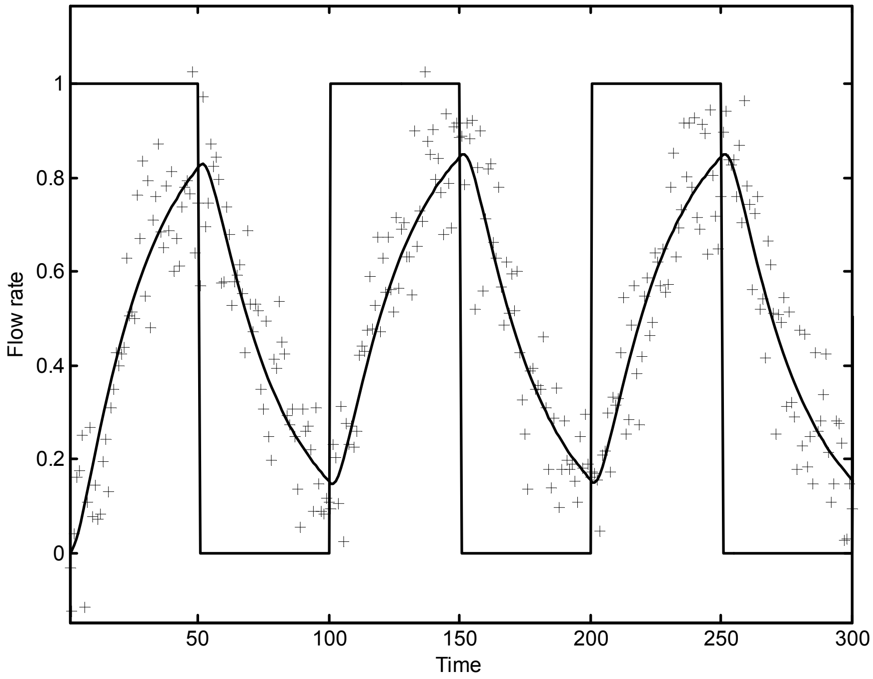


Figure 4.2 Actual response (solid line) of a 2nd order system to a pulsed input (dashed line) and simulated observations ('+' markers).

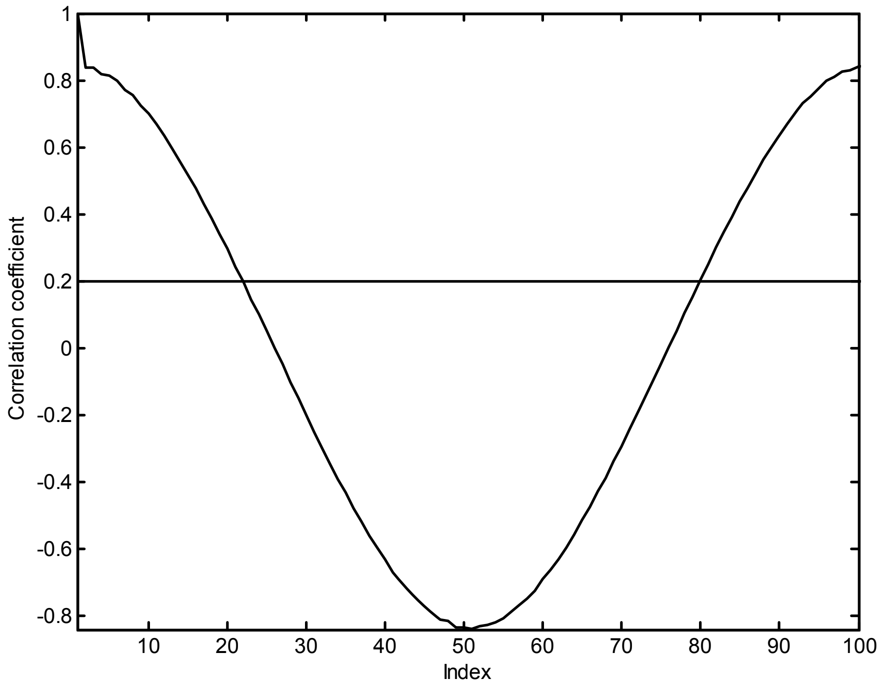


Figure 4.3 Autocorrelation function for two-tank series.

The first four principal components explained 94.1%, 2.87%, 0.27% and 0.16% of the variance in the time series respectively and only these four components were retained. In fact, only the first two components could probably have been retained, since they capture more than 97% of the variance in the data.

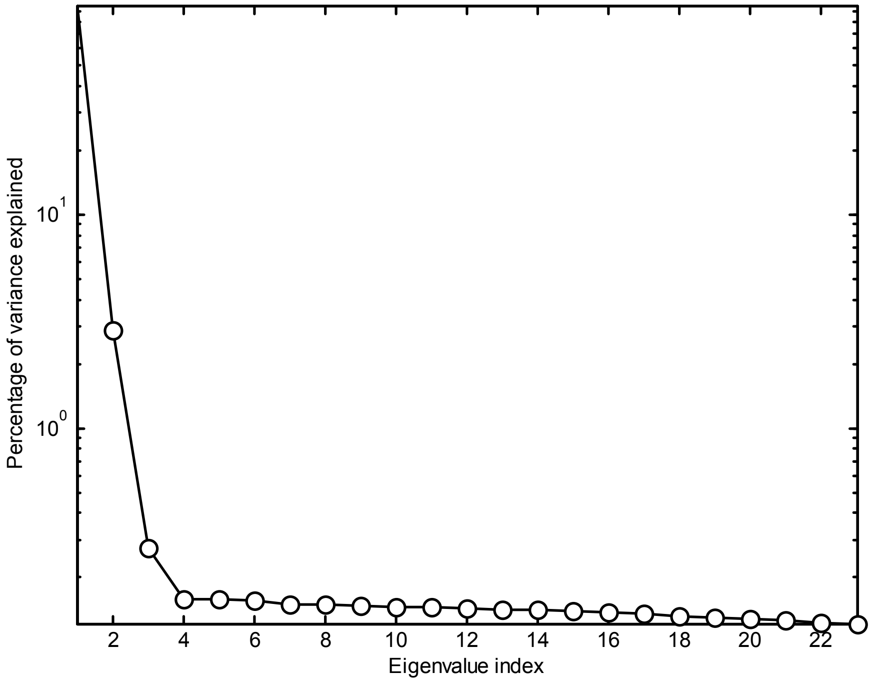


Figure 4.4 Eigenspectrum for the response (solid line) shown in Figure 4.2.

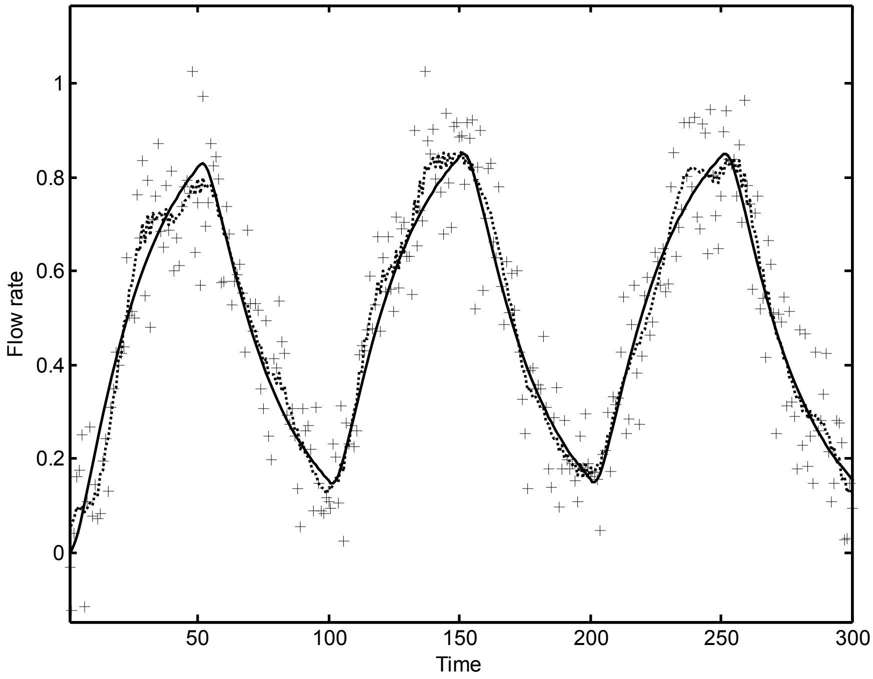


Figure 4.5 Reconstruction of the observed time series (dotted line), true process dynamics (solid line) and observations ('+' markers).

The reconstructed time series is shown as the broken line in Figure 4.5. The true response (solid line) and measured response ('+' markers) are also indicated in this figure. It can be seen from Figure 4.5 that the reconstruction of the data by using SSA gives a very close approximation of the true process dynamics.

4.2.2 Modelling of data series

Thereafter neural network models were fitted to the data, according to strategies A-C discussed above, and the results that were obtained by the modelling are shown in Figure 4.6. The numbers in parentheses in the legend represent the squared Pearson correlation coefficients of the models (r^2), i.e. the fraction of the variance explained by the model, calculated by using equation 4.4.

$$r^2 = \frac{\sum_{i=1}^{n} (y_{pred} - y_i)^2}{\sum_{i=1}^{n} (y_{avg} - y_i)^2} \qquad 4.4$$

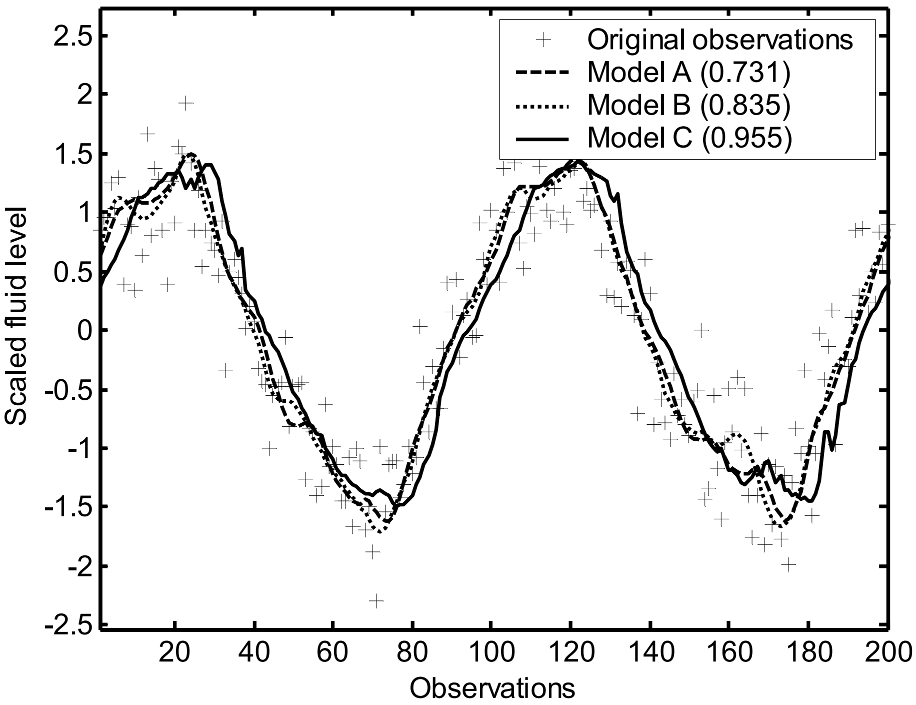


Figure 4.6 Prediction of the output flow rate from the two-tank system in section 4.2. Values in parentheses in the legend are the r^2-values associated with the models.

The characteristics and results obtained for the two-tank series are summarized in Table 4.1. In this table, the different network configurations for the various strategies can be seen, where the configuration refers to the number of nodes in the input, hidden and output layers respectively. The squared Pearson's correlation coefficient for each strategy and the autocorrelation of the first two values in the series are also supplied. This autocorrelation value can be used as an estimate of the degree of correlation between the observations in the time series, and therefore the level of noise affecting the time series. The lower the correlation, the more noise the time series. For a purely linear time series, one would expect the neural network model to perform at least as well as the autocorrelation value, as the autocorrelation is an indication of the amount of variance that can be explained with a simple linear regression model, using $y_t = \alpha y_{t-1}$.

Table 4.1 Summary of results obtained from modelling with different strategies (A-D), as well as the autocorrelation function, AC(1), at a lag of one for the two-tank in series data set. Network configurations refer to the number of nodes in the input, hidden and output layers respectively

Time series	Network configuration	Validation data R^2	AC(1)
Tanks in series (A)	4:3:1	0.731	0.84
Tanks in series (B)	4:5:1	0.835	
Tanks in series (C)	Multiple	0.855	

Since the system is linear and the sampling rate is high, strategy B did not perform better than a simple linear model and strategy C performed only slightly better than one would expect from a linear model. However, both strategies B and C gave significantly better results than strategy A, where the model was compromised by the noise in the unfiltered data.

4.2.3 Influence of window length

One final set of tests that was performed on this two tank in series data set was to embed the time series into trajectory matrices of differing window lengths in an effort to determine the sensitivity of SSA for the window length parameter. The chosen window lengths were 5, 10, 20, 30, 40 and 80 columns, while it has been determined in the previous section that the optimum window length for this series is 23 columns, according to the criterion that has been specified.

The resulting eigenspectra for all six different window lengths are supplied in Figure 4.7, with an enlargement of the first few eigenvalues in Figure 4.8. Even though the time series were embedded up to a window length of 80, only the first 60 eigenvalues are shown. This was seen as sufficient, as the last twenty eigenvalues behaved similar to the eigenvalues normally found in the tail-end of the eigenspectra in that it flattened into a noise plateau. It can be seen that for the first two window lengths, namely 5 and 10, it is hard to distinguish between the relevant eigenvalues and those representing noise. This is probably due to the fact that the embedding dimension of the data is not high enough to capture all the variance in the data, as the autocorrelation function in Figure 4.3 is still in excess of a correlation of 0.75 at an index of 10. For a window length of 20, which is relatively close to the optimum embedding window of 23, it gets easier to distinguish the first two eigenvalues from the rest of the eigenspectrum and, as it was mentioned earlier, these two eigenvalues would probably be sufficient to characterize the data. For all three of the larger window lengths the first two eigenvalues can also be identified as being significant. However, for the window length of 40, the second two eigenvalues also seem relevant, whereas this is not necessarily true and for the window length of 80 columns, the eigenspectrum exhibits very strange behaviour in the region of the 52nd eigenvalue. The eigenspectrum makes an unexpected break, almost indicating two different groupings of eigenvalues. This ‘break’ appears to be associated with the minimum in the autocorrelation function (see Figure 4.3) at an index or lag of 52 and it would appear as if the eigenspectrum is repeating itself. However, a more detailed explanation for this behaviour would require further investigation that is beyond the scope of the present study.

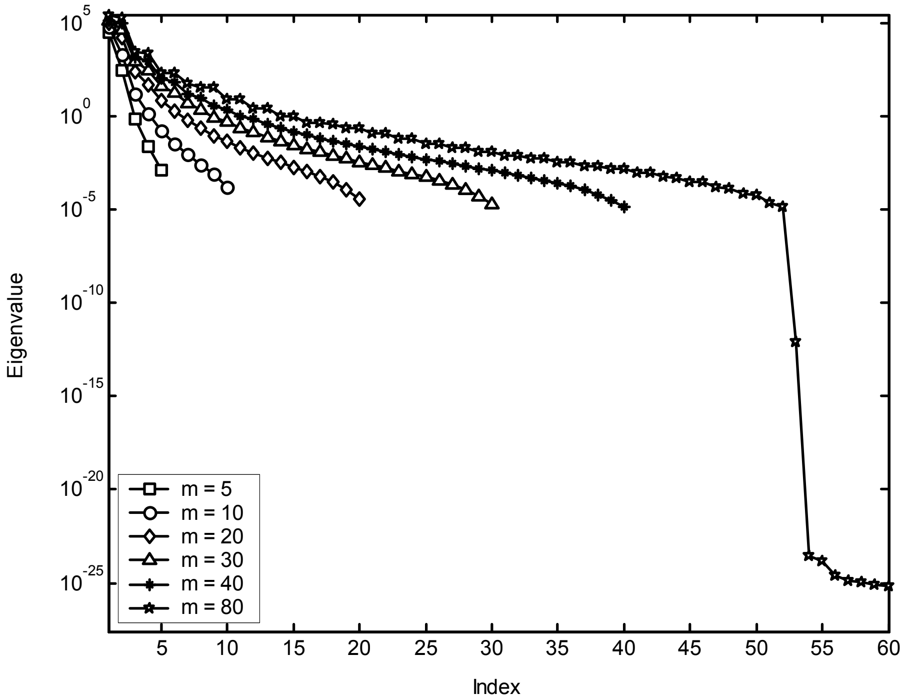


Figure 4.7 Eigenspectra obtained from singular spectrum analysis of two tanks in series time series by using different window lengths.

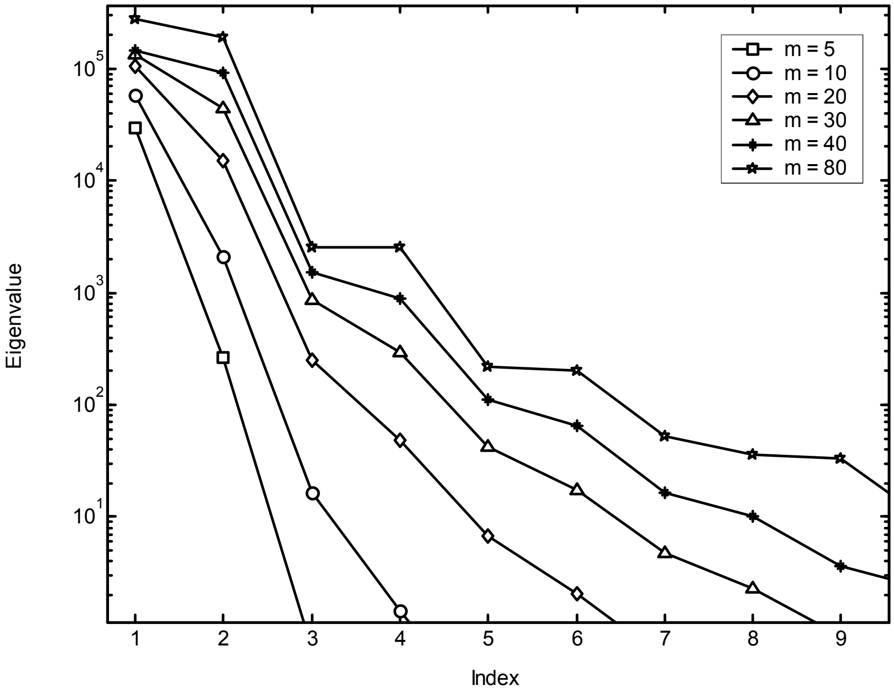


Figure 4.8 Enlargement of sections of eigenspectra obtained from singular spectrum analysis of two tanks in series time series by using different window lengths.

4.2.4 Reconstructed attractor

The differences between the different embedding dimensions can also be seen from the various reconstructed attractors, displayed in Figure 4.9.
The attractor is a visual representation of the behaviour of the system in geometric form and is obtained by plotting the first three principal components as functions of each other.

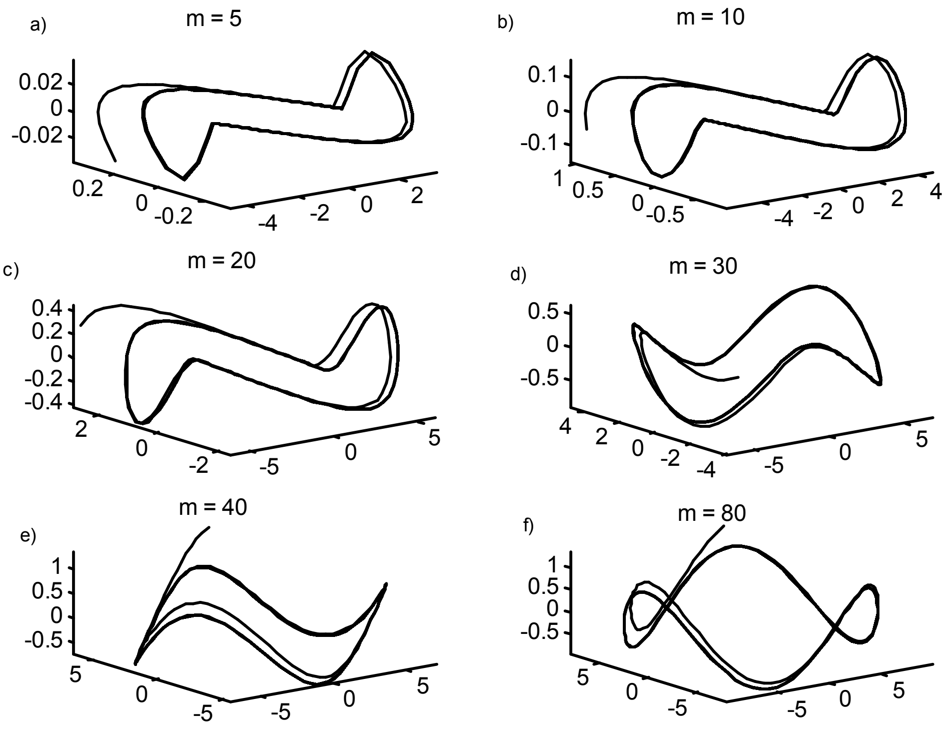


Figure 4.9 Reconstructed attractor of two tanks in series time series for different window lengths of the embedding window.

Figure 4.9(a)-(c) indicates that for the three window lengths smaller than the optimum window length of 23, the reconstructed attractors have basically the same shape. However, as the embedding window size increases, in Figure 4.9(d)-(f) one can see some change in the topology of the reconstructed attractors. The most conspicuous change is the appearance of loops in the attractor associated with a window length of 80. These changes in the appearance of the attractor associated with larger window lengths are not the results of more information being retrieved in the reconstruction, as it has been found that the time series can be reconstructed almost perfectly within an embedding window larger than ten components. The change in topology could therefore be attributed to spin present in the reconstruction of the time series.

4.3 Carbon-in-leach process

4.3.1 Background

In this second case study, a simple linear signal is considered in order to further demonstrate the methodology, but certain complications are presented in this time series. The data were obtained from a simulated carbon-in-leach cascaded continuous stirred tank reactors system (Van der Walt, 1992), as shown by the broken line in Figure 4.10. The data were specifically concerned with the extraction of gold from leached or leaching slurries. Note the discontinuity in the observations, which complicates filtering of the data. The true dynamics of the system were corrupted by measurement noise with a normal distribution (zero mean and standard deviation of 0.25), as indicated by the dots in Figure 4.10.

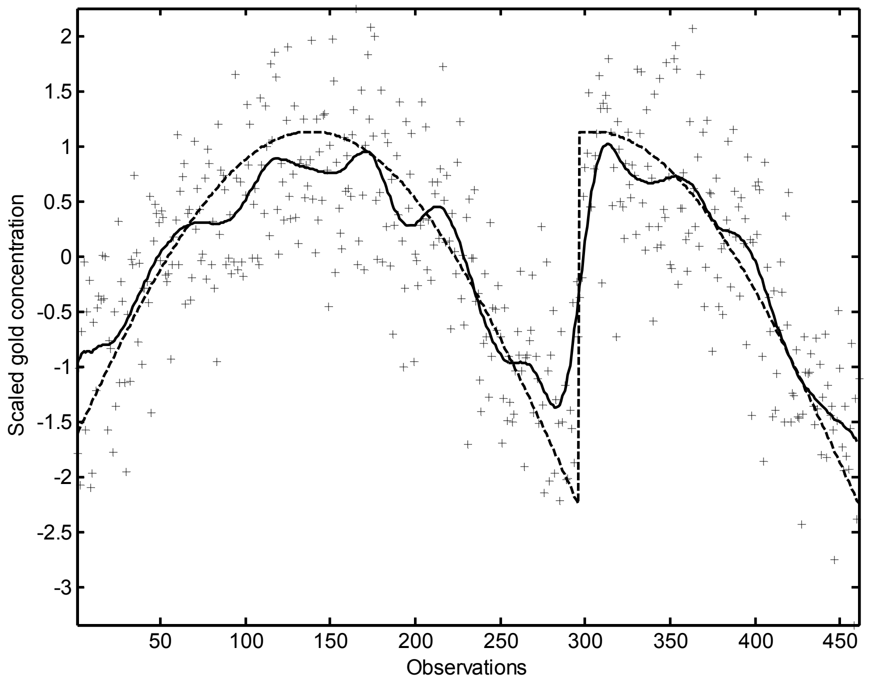


Figure 4.10 Simulated carbon-in-leach process with noisy observations ('+'-markers) and true dynamics (broken line).

From this corrupted time series, a trajectory matrix with 50 columns and a 50 x 50 lagged covariance matrix were constructed. Even though the autocorrelation function in Figure 4.11 would suggest rather using a window size in the order of 35 (where the autocorrelation function reaches 0.2), it was decided for this case study to instead use the window size where the autocorrelation function reaches the real point of decorrelation, namely at an autocorrelation of 0. This then explains the window length of 50 in this case study.
It should be kept in mind that the window length determined from the autocorrelation function is only an indication of the minimum length that would be sufficient to capture all the variance from the data. However, this by no means implies that this minimum value would always be the optimum window length and one should still use discretion in the choice of the embedding dimension.

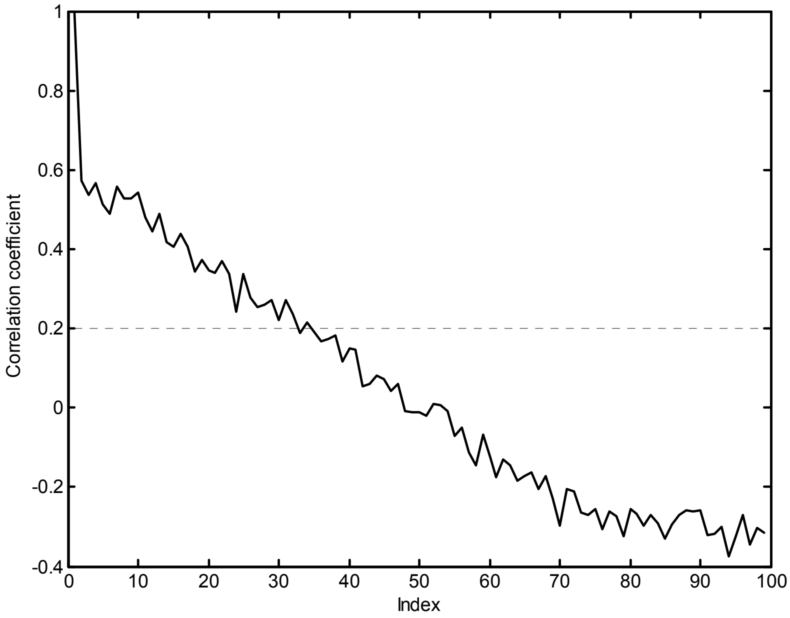


Figure 4.11 Autocorrelation function of carbon-in-leach simulated CSTR process.

4.3.2 Grouping of eigenvalues

The eigenvalue spectrum of the observations is shown in Figure 4.12. The eigenspectrum of the data shows distinct groupings of the eigenvalues, with the first three eigenvalues located on a line with a steep slope, followed by the next two groups of three eigenvalues lying on different line segments, with the remainder of the eigenvalues being relatively small and therefore ignored for the purposes of filtering. This suggests the decomposition of the time series into three large components, of which the first one is the most important in terms of filtering the data.

The groupings of the eigenvalues in Figure 4.12 shows the first 9 cumulatively reconstructed components of the decomposed time series (RC$_1$, RC$_2$, … RC$_9$. This figure confirms the observation made from the inspection of the eigenspectrum itself. The first three cumulatively reconstructed components can be seen to approximate the original noise-free dynamics of the time series quite well with the additional components added after the third component tending to be more representative of small variations, which in this case is purely noise. The grouping of the eigenvalues can also be seen when the reconstructions with four to six components are compared with those from seven to nine components. For the smaller eigenvalues (reconstructed with four to six components), the variations added to the signal seemed a lot ‘smoother’ than those added by the seventh to the ninth component.

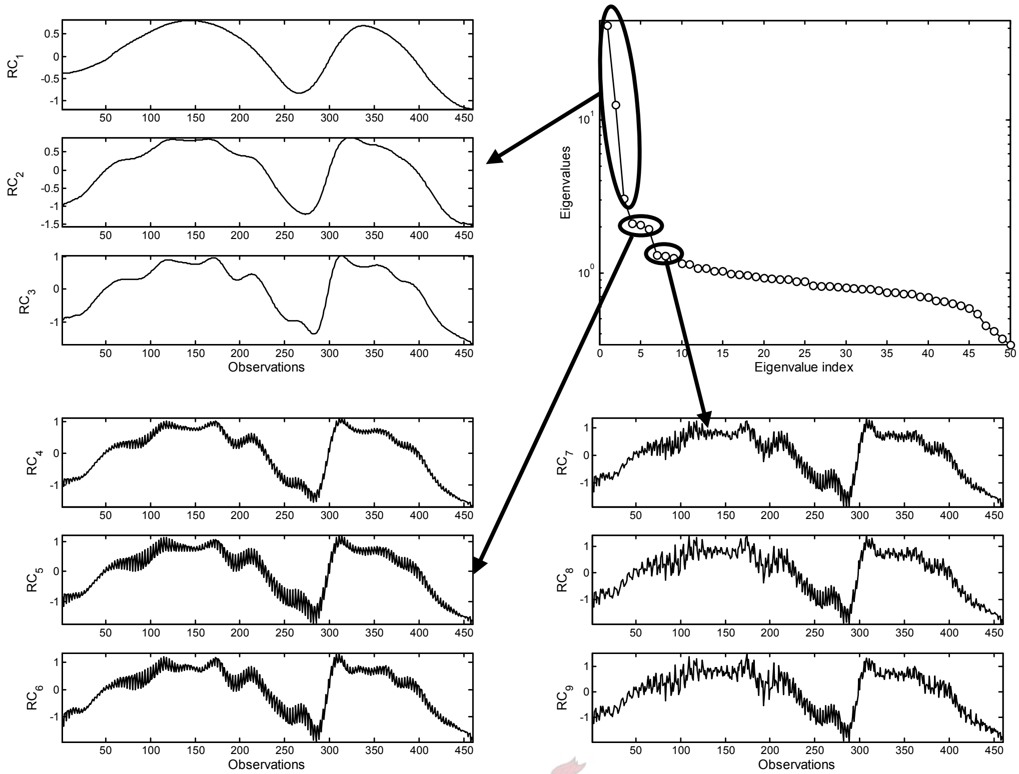


Figure 4.12 Eigenspectrum of the data in Figure 4.10 and the cumulative reconstructed time series (RC_1–RC_9) associated with the first 9 eigenvalues.

This difference between the principal components in different positions in the eigenspectrum can also be seen from Figure 4.13, where the individual reconstructed components are illustrated. However, it is interesting to note that even though eigenvalues four, five and six and eigenvalues seven, eight and nine appear to be divided into two groups in which it would be expected that all the components would behave similarly, in both cases the third eigenvalue in the group produced a completely different reconstructed component from the rest. This confirms the need that was identified by Allen and Robertson (1996) to confirm or discredit apparent oscillatory components by first applying Monte Carlo SSA. More attention will be given to this approach in a later chapter.

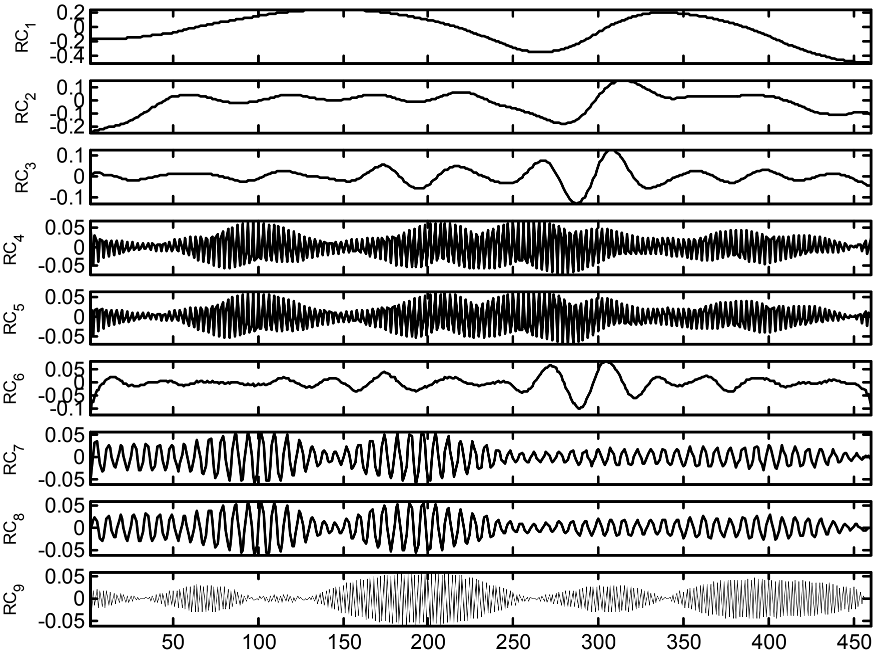


Figure 4.13 Individually reconstructed components for the first nine eigenvectors extracted from the carbon-in-leach cascaded CSTR process.

4.3.3 Filtering of data

The filtered version of the data obtained from SSA was compared with the results obtained from other filtering techniques, in order to determine the comparative filtering efficiency of SSA.

Because the process simulated was a simple linear signal, a mean filtering technique was used. This involved taking the average over a selected number of values from the data and then moving the 'window' in which the average was taken along the time series. For this application, the time was taken to determine the theoretically optimal filter window size by comparing the results from a number of different filtering windows. The results for three of the filter windows, namely 5, 11 and 15 are illustrated in Figure 4.14, where the filtered time series are compared with the original clean signal. It can be seen from the correlation coefficients, supplied in brackets in the axes labels, that a moving average of 11 gave the best correlation with the original, clean process dynamics. It would appear from the figure that the smaller moving averages still contained too much of the variation induced by the added noise, and the larger moving averages, represented by the window of 15 in the figure, were not adaptable enough to define the discontinuity and lost information around the start and the end of the time series.

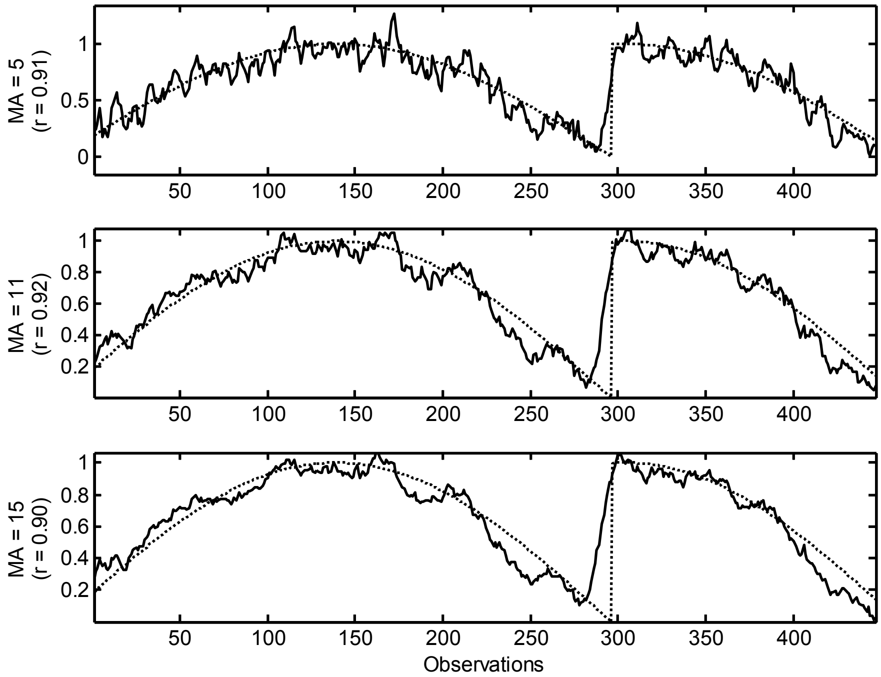


Figure 4.14 Moving average filters of filter window sizes of 5, 11 and 15 which were built on the noise-corrupted time series (solid lines) are compared with the original clean signal (dotted line). Correlation of the filtered series with the original time series is supplied in brackets.

Both the output from the moving average filter and the singular spectrum analysis were then compared with the original clean process dynamics. This allowed one to judge how effectively the two techniques could identify the true process dynamics from a noisy signal.
The results obtained from SSA for only three reconstructed components (r = 0.96) are displayed in Figure 4.15 and are somewhat better than those obtained with a theoretically optimal mean filter of 11, which showed a correlation of r = 0.92 with the original signal. In practice, the results obtained with the mean filter would likely be significantly worse, as one does not have the luxury to find the optimum window length by comparing results with the known process dynamics.

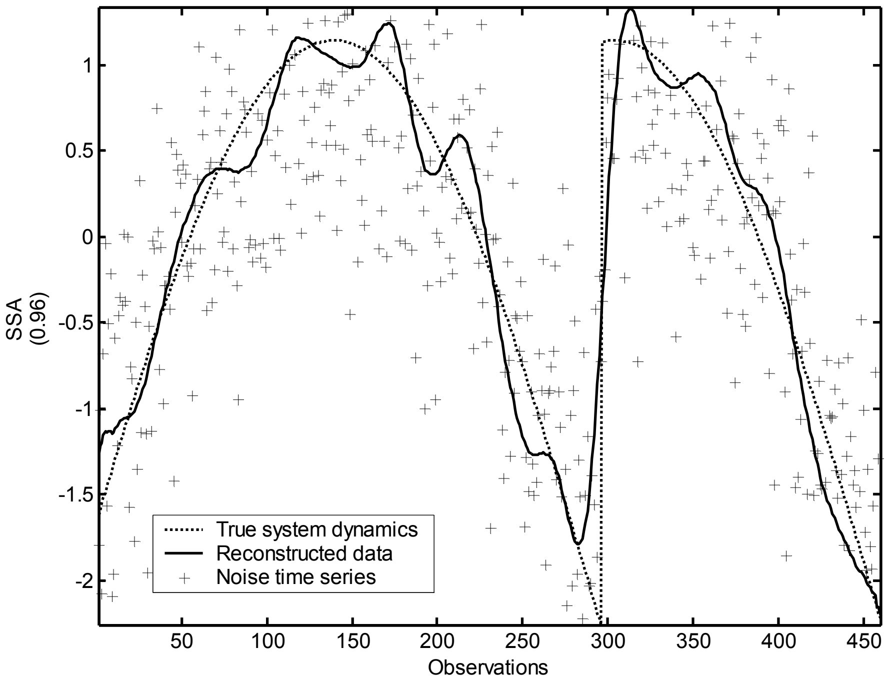


Figure 4.15 Filtering of the carbon-in-leach cascaded CSTR process obtained by using SSA and a reconstruction with only three components (solid line), in comparison with the true process dynamics (dotted line) and the noisy time series on which the filtering was applied ('+'-markers).

4.4 Base metal flotation plant – rougher, cleaner and scavenger circuits

4.4.1 Background

Frothing is a common phenomenon in mineral engineering operations and especially in flotation it is of fundamental importance to the efficiency of grades and recoveries. In the last five years, considerable progress has been made concerning the use of control systems based on direct monitoring of the froth. At present, state-of-the-art digital image processing systems are based on sophisticated algorithms for the measurement of bubble size distributions in the froth, the analysis of flow patterns in flotation cells, as well as measurement of the stability of the froth surface near the concentrate overflow. In addition, the presence of reagents or mineral species can also be related to the appearance of the froth.

The data in this example were obtained from a South African copper flotation plant. The plant consists of a crushing section and milling circuit, followed by a magnetic separation section. The purpose of the magnetic separation is to remove the high percentage of magnetic material in the ore and thereby reduce the load on the flotation circuit. The flotation circuit itself is designed to operate with feed grades of 0.6% Cu, 9.0% Pb, 2.4% Zn and 130g/t silver. The circuit configuration consists of two conditioners, in which sulphuric acid, two copper collectors and a frother is added. From the two rougher banks, the concentrate is circulated to the three cleaner banks, where zinc is depressed by acid in the first cleaner and lead depressed in the second and the third cleaners by adding lime. The cleaner tails and the scavenger concentrate are returned to the copper feed of the flotation circuit, and these two streams make up the bulk of the feed.

4.4.2 Embedding of plant data

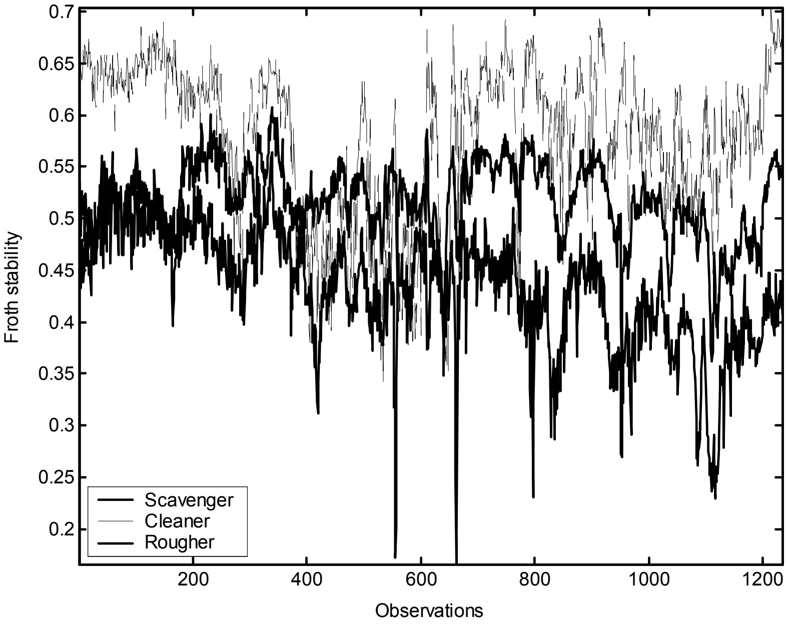


Figure 4.16 Time series observations of the copper froth stability in the rougher, cleaner and scavenger circuits.

The time series for the rougher, cleaner and scavenger units consisted of 1234 observations each, obtained at 12-minute intervals, as indicated in Figure 4.16. These observations were stability measurements of the froth in the 2nd cell in each circuit, which could be related to the recovery and grade of the circuit.

The optimal sizes of the embedding windows for each of the time series were determined by means of autocorrelation analysis as described earlier and were specified as 140, 29 and 64 columns respectively. The cleaner, rougher and scavenger time series were then scaled to zero mean and unit variance and embedded in trajectory matrices (X_C, X_R, X_S). From these trajectory matrices, the lagged covariance matrices ($C_C = X_C^TX_C$, $C_R = X_R^TX_R$ and $C_S = X_S^TX_S$) were formed and decomposed by means of principal component analysis. The eigenspectra of these trajectory matrices are shown in Figure 4.17.

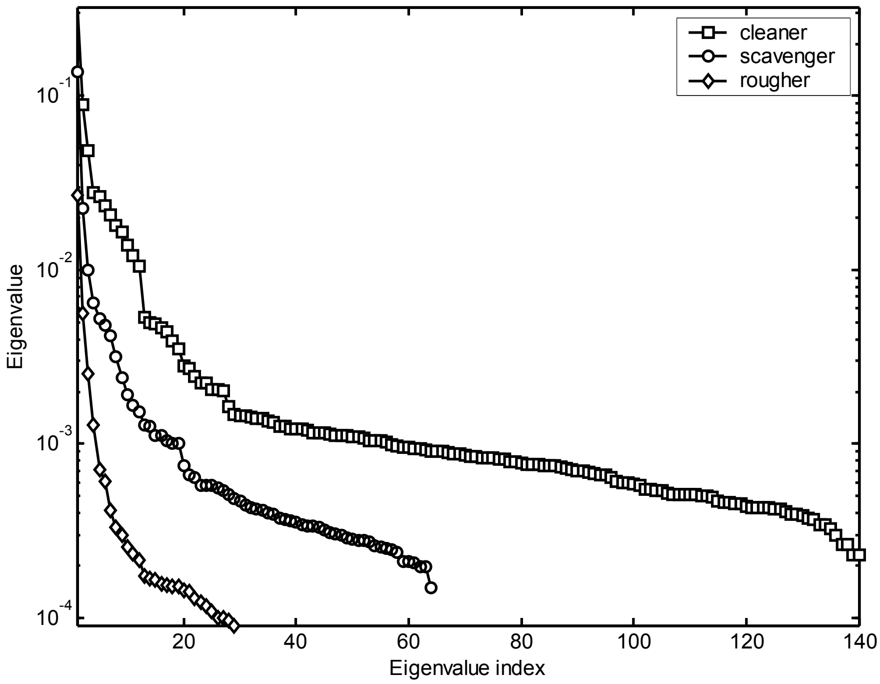


Figure 4.17 Eigenspectra of the cleaner, scavenger and rougher trajectory matrices.

By using the eigenspectra of the three time series, as presented in Figure 4.17, as the main source of information, one can attempt to extract the significant principal components of the time series and discard those components representative of noise. The split between the components of useful signal and those of noise is usually done where the spectrum either reaches a plateau (flattens out) or where a significant change in the shape of the spectrum occurs. This can be better described as the point on the graph where the contribution of each consecutive eigenvalue towards explaining the variance in the data became relatively small and did not differ significantly from that of the previous eigenvalue.

Figure 4.18 - Figure 4.20 show the components of the rougher, cleaner and scavenger data, as well as the cumulative reconstruction of each time series with these components. This cumulative reconstruction is achieved by adding the appropriate number of corresponding $t_ip_i^T$-components and then performing the unembedding operation to obtain the reconstruction with the desired number of reconstructed components. The most dominant component, representing the first principal component, gives the basic shape or trend of the time series. The rest of the components either represent other factors influencing the time series or indicate noise. The progressive reconstruction of each time series is based on the use of increasing numbers of principal components. The number of principal components used to reconstruct each time series is indicated next to each time series.

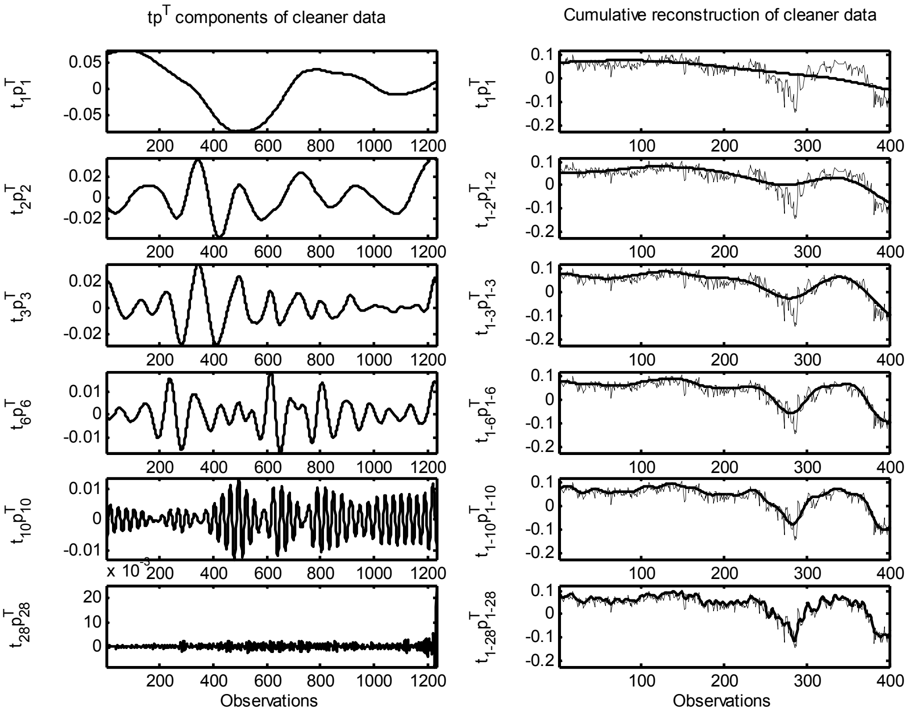


Figure 4.18 tp^T components of the scaled cleaner data (left column) and cumulative reconstruction.

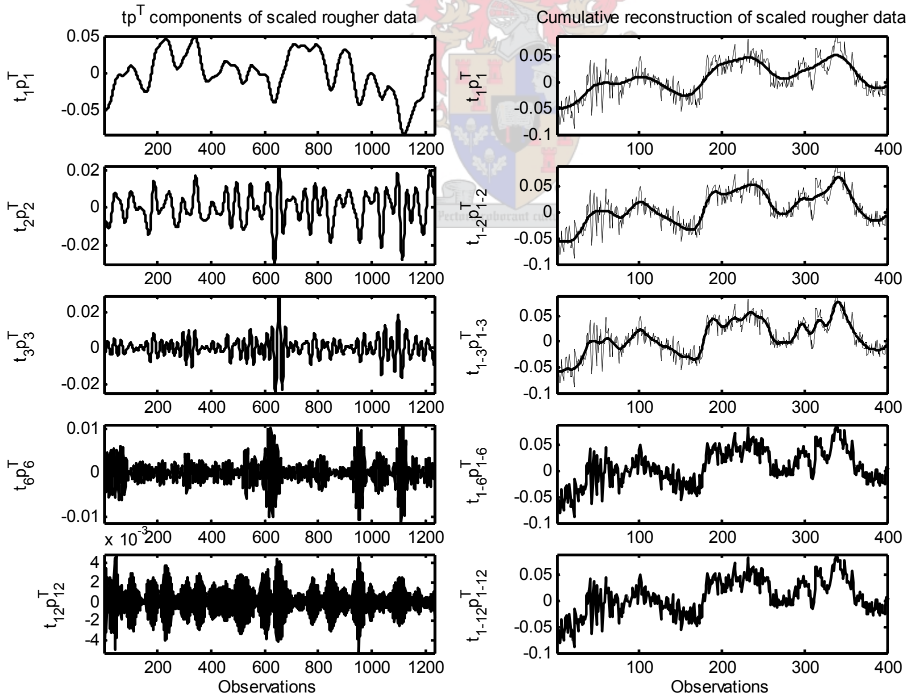


Figure 4.19 tp^T components of the scaled rougher data (left column) and cumulative reconstruction.

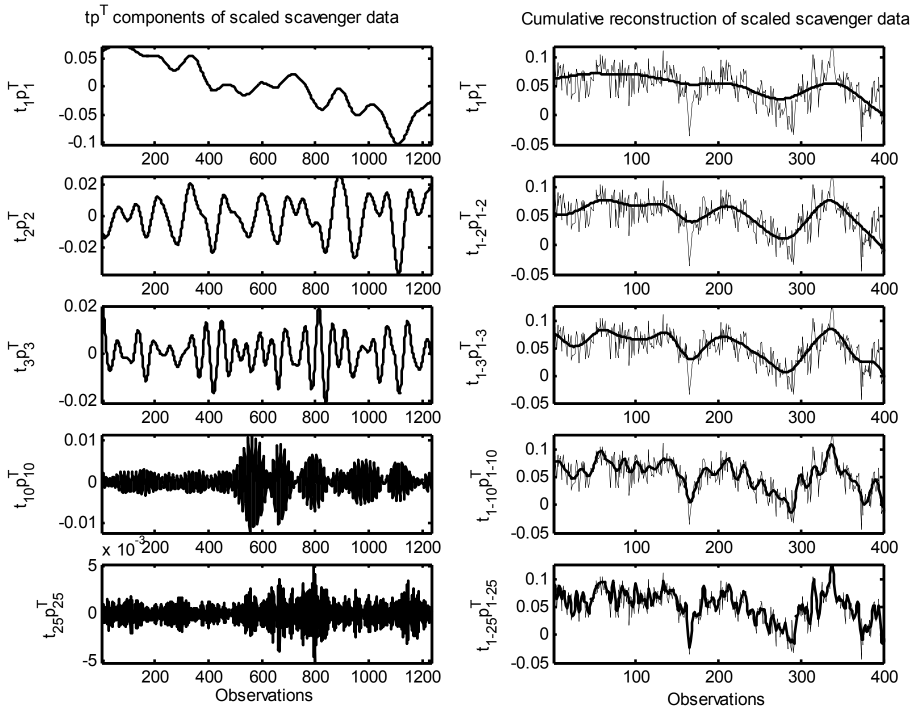


Figure 4.20 tp^T components of the scaled scavenger data (left column) and cumulative reconstruction.

4.4.3 Modelling of the cleaner, rougher and scavenger

Multilayer perceptron neural network models were built for each of the three reconstructed time series, as well as the three original time series according to strategies A and B previously discussed. In addition, models were also fitted to the components of each time series (strategy C), as well as to the time series reconstructed from the multivariate trajectory matrix of the system (strategy D). The single hidden layer neural networks with sigmoidal activation functions were automatically constructed based on a modified Schwarz information criterion and the Levenberg-Marquardt optimization algorithm. The quality of the models (A-D) was assessed by means of validation data sets not used during the development of the models.

The output of each model was validated against data from the *original* time series. This data against which the model was validated included a section of the time series (usually 20% of the observations) that has not been used for training purposes (validation data set). As can be seen from Figure 4.21- Figure 4.23, the correlations obtained from modelling strategies B and C were appreciably better than that from modelling strategy A. With strategy A, the noise in the data tended to confound the model. Also note that for one-step ahead predictions, only the cleaner data could be predicted better with all three of the SSA-based neural network approaches than one would expect from the autocorrelation between two consecutive observations. This implies that, definitely in the case of the scavenger and the rougher series, and most probably for the cleaner time series, one could simply have used a linear model, such as regression, to predict the filtered time series with the same level of accuracy.

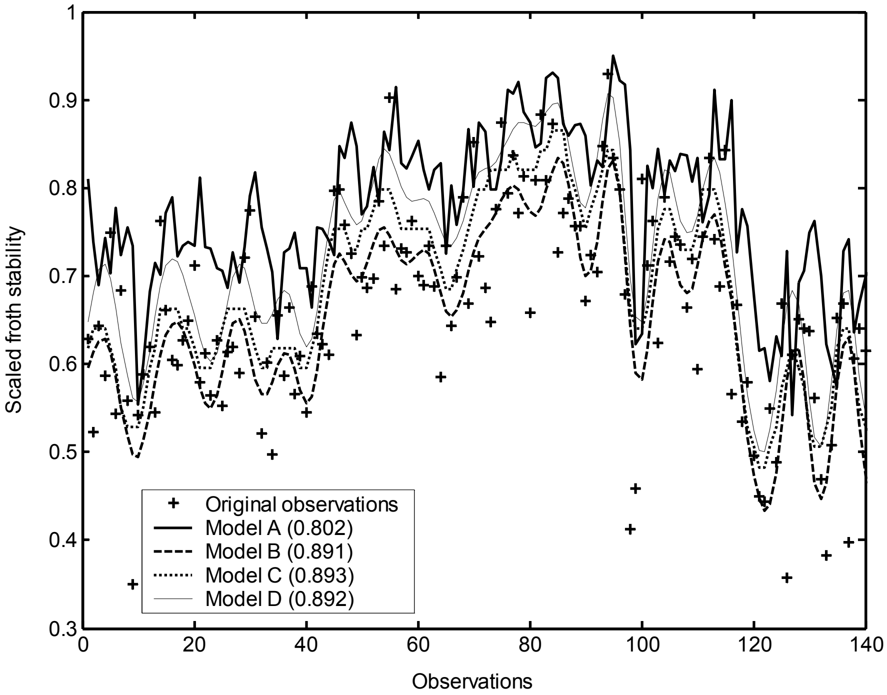


Figure 4.21 One-step ahead prediction of the froth stability in the cleaner for all four modelling strategies. The number supplied in brackets next to each model indicate the correlation of the modelling results with the original time series.

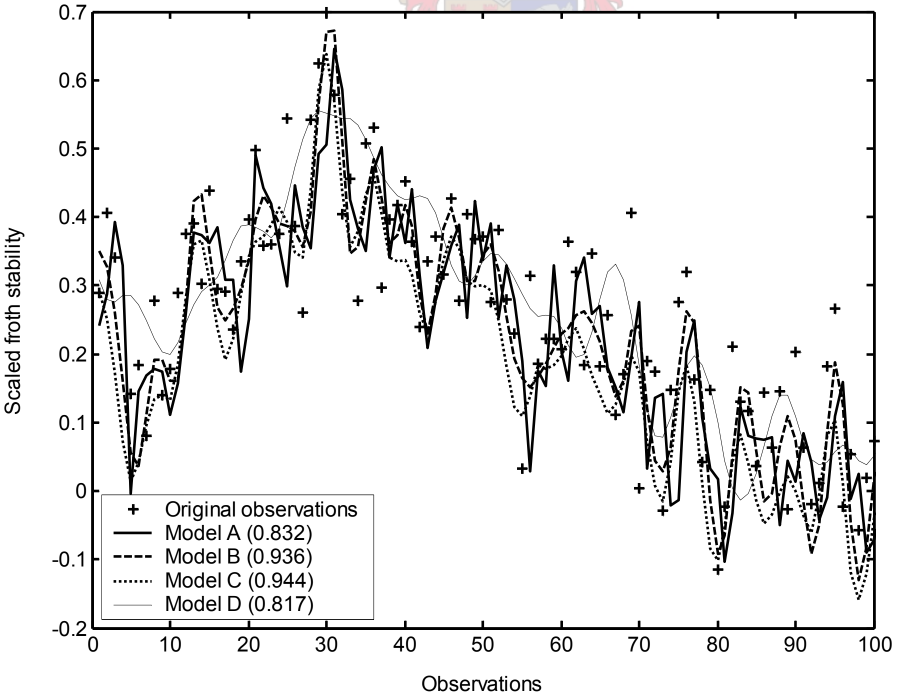


Figure 4.22 One-step ahead prediction of the froth stability in the rougher for all four modelling strategies. The number supplied in brackets next to each model indicate the correlation of the modelling results with the original time series.

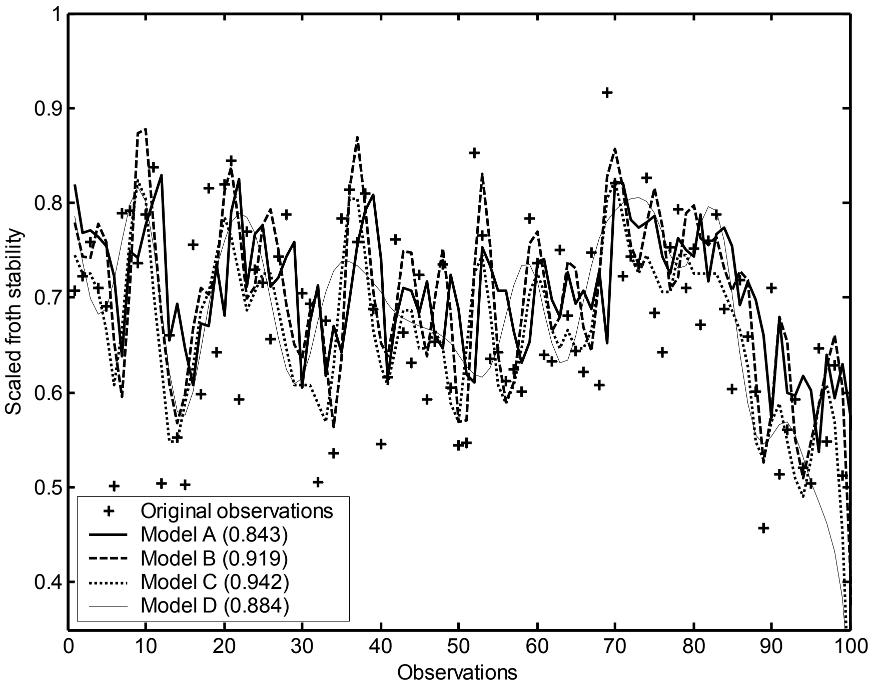


Figure 4.23 One-step ahead prediction of the froth stability in the scavenger for all four modelling strategies. The number supplied in brackets next to each model indicate the correlation of the modelling results with the original time series.

The summary of the modelling results and specifications are supplied in Table 4.2, where the significance of all the columns is similar to that in paragraph 0 and Table 4.1.
In this case study, strategies B and C proved to be superior to strategy A on all the time series considered and slightly or significantly better than the benchmark model, AC(1). In this case strategy D, based on a multivariate embedding of the data, did not perform as well as strategies B and C. In theory it is supposed to exploit the redundancy in the data, but in practice the reduced presentation of the trajectory matrices led to a net loss in information, which impacted on the models.

Table 4.2 Summary of results obtained from modelling with different strategies (A-D), as well as the autocorrelation function, AC(1), at a lag of one for each data set from the copper flotation plant. Network configurations refer to the number of nodes in the input, hidden and output layers respectively

Time series	Network configuration	Validation data R^2	AC(1)
Scavenger (A)	14:3:1	0.843	0.889
Scavenger (B)	11:4:1	0.819	
Scavenger (C)	Multiple	0.942	
Scavenger (D)	16:3:1	0.884	
Cleaner (A)	22:2:1	0.802	0.866
Cleaner (B)	8:4:1	0.891	
Cleaner (C)	Multiple	0.893	
Cleaner (D)	8:3:1	0.892	
Rougher (A)	34:2:1	0.832	0.909
Rougher (B)	26:2:1	0.936	
Rougher (C)	Multiple	0.944	
Rougher (D)	19:5:1	0.817	

4.5 Base metal flotation plant – recovery in scavenger circuit

4.5.1 Background and performance of SSA

The data used for this case study was obtained from the same base metal flotation plant described in section 4.4.1 but for this time series the recovery grade of the precious metals, Cu, Pb and Zn, in the scavenger circuit was measured. Figure 4.24 shows the 12-minute interval measurements of the Zn, Pb and Cu concentrations. Each time series shown in Figure 4.24 consisted of 1234 measurements and was scaled to zero mean and unit variance before doing the analysis.

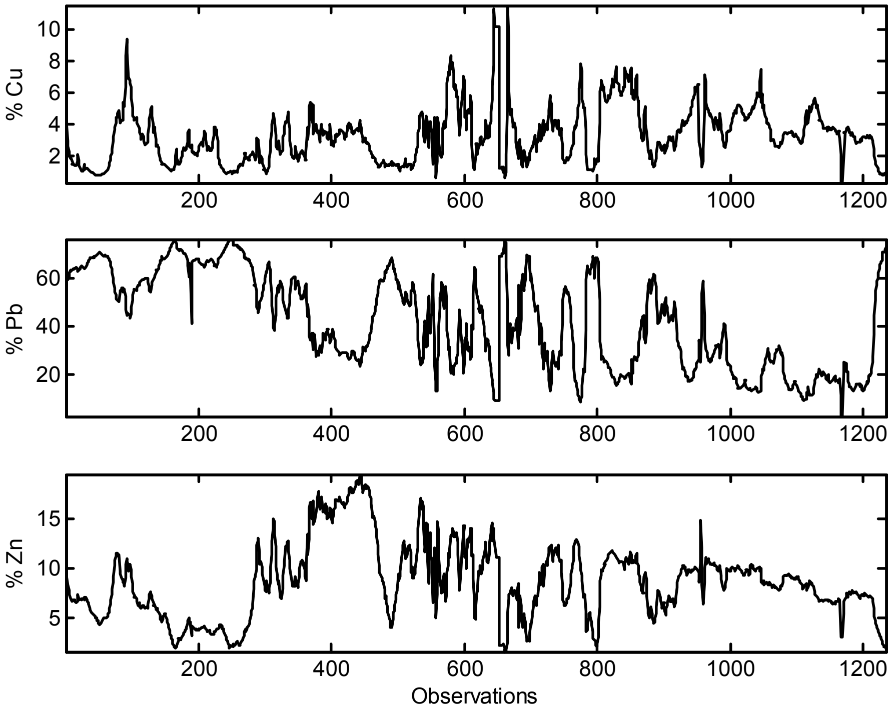


Figure 4.24. Measurements of Zn, Pb and Cu in the scavenger circuit collected at 12-minute intervals.

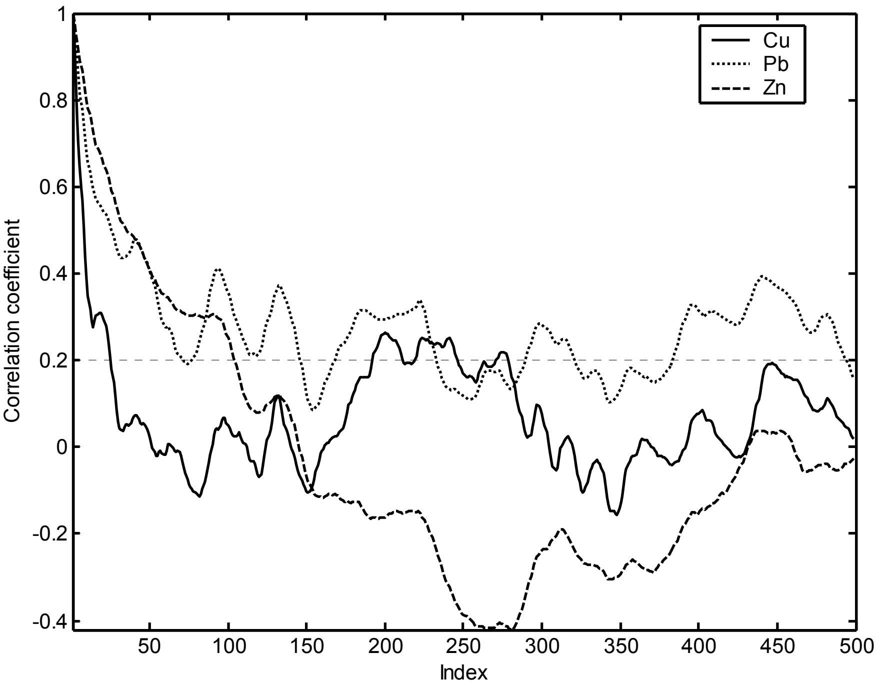


Figure 4.25 Autocorrelation functions of data from scavenger circuit of base metal flotation plant.

The dimensions of the covariance matrices were determined by inspecting Figure 4.25. For the Zn, Pb and Cu series, these dimensions were 105 x 105, 31 x 31 and 26 x 26 respectively. Even though the Pb series only reaches the point of linear decorrelation (correlation coefficient < 0.2) at a window length of approximately 75, a significant minimum can be observed at embedding dimension of 31, and therefore this smaller dimension was specified for the trajectory matrix.

For the eigenvalue distributions of the time series, displayed in Figure 4.26, one can see that none of the time series has such a clearly defined noise plateau as that of the 'artificial' time series investigated in sections 4.2 and 0. This necessitated greater care in the selection of the number of principal components to retain, often requiring a number of trial and error attempts to find the optimum number of eigenvalues.

These trial and error attempts were performed by choosing an apparent ideal number of components to retain, then building a neural network model on the filtered data. The degree to which the model could predict a set of data not used during the training of the neural network served as an indication of the accuracy of the model. This procedure was repeated for a number of retained components for each time series and the number of components giving the best correlation of the predicted validation data with the original time series was seen as the optimum reconstruction.

The number of components retained for the Cu, Pb and Zn series, were finally specified as 11, 14 and 12 components respectively. The time series reconstructed with a smaller number of principal components, still explained 98.1% of the variance in the copper data, 99.5% of the variance in the lead data and 95.7% of the variance in the zinc data. It is these reconstructed series, as well as the original time series, that were used to build models on the data, as will be discussed in the following section.

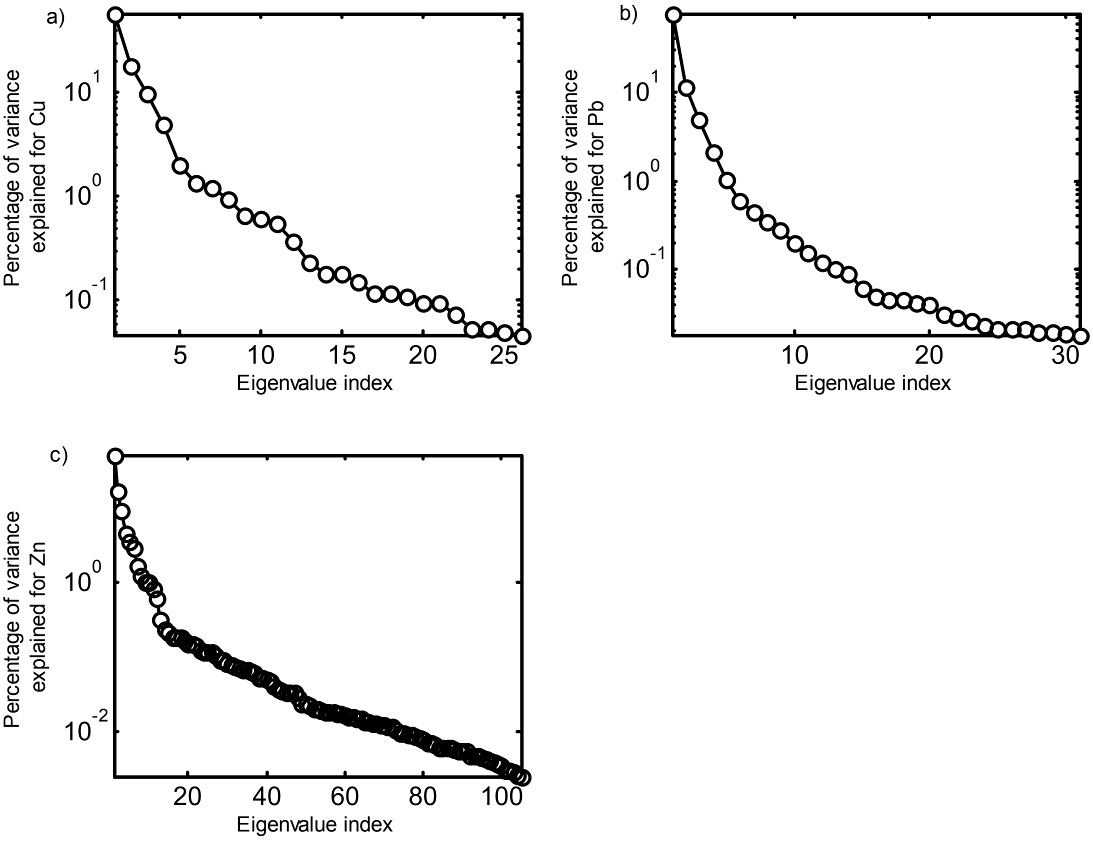


Figure 4.26 Percentage of variance explained by each eigenvalue for a) Cu, b) Pb and c) Zn time series.

4.5.2 Modelling of the Cu, Pb and Zn in the scavenger

Only two of the modelling strategies described in section 4.1 were considered, namely strategies A and B. The results from this modelling are summarized in Table 4.3 and sections of the modelled time series are shown in Figure 4.27, with the variance accounted for by each model indicated in parentheses in the legend of the figure.

The multilayer perceptron neural network (5 hidden nodes) fitted to the original data from the scavenger, could account for 87.5% of the variance of the observed copper concentrations. In contrast, the neural network model (8 hidden nodes) fitted to the smoothed copper concentration data, accounted for 92.2% of the variance. Likewise, the neural network models fitted to the original data could explain 97.3% (4 hidden nodes) and 80.5% (4 hidden nodes) of the variance of the lead and zinc concentrations respectively, while the models fitted to the smoothed data could explain 97.9% (5 hidden nodes) and 99.0% (20 hidden nodes) of the variance of the measured lead and zinc concentrations.

Table 4.3 Summary of results obtained from modelling with different strategies (A and B), as well as the autocorrelation function, AC(1), at a lag of one for each data set from scavenger circuit of the copper flotation plant. Network configurations refer to the number of nodes in the input, hidden and output layers respectively

Time series	Network configuration	Validation data R^2	AC(1)
Copper (A)	15:5:1	0.875	0.945
Copper (B)	14:8:1	0.922	
Lead (A)	16:4:1	0.973	0.982
Lead (B)	15:4:1	0.979	
Zinc (A)	23:5:1	0.805	0.979
Zinc (B)	10:20:1	0.990	

It can be seen that, especially for the zinc time series, the model built on the reconstructed time series significantly outperformed that built on the original time series. The trends with regard to these two modelling strategies that are noticed here, are similar to that observed in the other case studies presented earlier in this thesis. These trends are also analogous to the ideas on which principal component regression is based outside the field of dynamic modelling or system identification, except that in this case the idea is not to decorrelate the inputs, but to reduce the deleterious effect of noise.

However, from Table 4.3 it can be seen that the only model that performed better than the AC(1) value for that time series, was the neural network model built on the reconstructed zinc time series, with all the other models built on reconstructed time series performing just slightly worse than their respective AC(1) values. The models built on both the original copper and the original zinc time series performed significantly worse than their AC(1) values, with the model built on the original lead series coming closest to its AC(1) value. The reason why the zinc gave the best results is probably because the reduced number of principal components retained for the copper and the lead were not sufficient to explain all of the variance in the data. Some of the important process information was probably lost in the principal components that were classified as noise and therefore discarded. This is an interesting situation, seeing as the percentage of the eigenvalues themselves that were retained was the smallest in the case of the zinc. Both the lead and the copper were embedded in relatively small trajectory matrices and therefore a much larger percentage of their eigenvalues was retained, even though the number is more or less the same as for zinc.

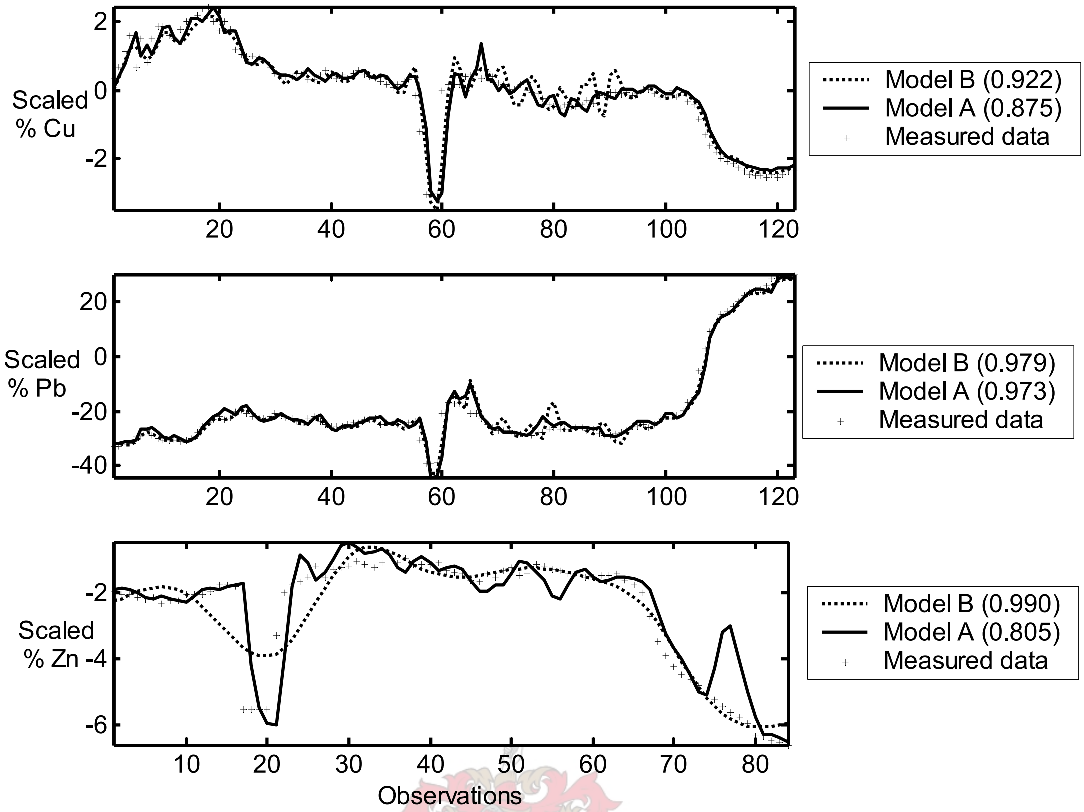


Figure 4.27 Model predictions of the Cu, Pb and Zn concentrations in the scavenger circuit. Numbers in brackets indicate the fraction of the variance in the original data explained by the model.

The success of the model built on the original zinc time series can also be seen from the free-run prediction of this series. In Figure 4.28, the free-run predictions of Model A for the Zn in the scavenger circuit are shown. Free-run predictions are obtained by using the predictions of the model at time *t* as its input for prediction at time *t*+l, where l is the chosen embedding lag of the time series. It is a very stringent test, because small prediction errors accumulate rapidly, eventually leading to catastrophic failure of a less-than-perfect model. Nonetheless, as shown in the figure, the model can predict the Zn concentration for more than 30 time steps quite accurately. In a time series such as this, where the interval between observations is 12 minutes, 30 time steps culminates into quite a lengthy period of time that can be predicted in advance. This can be interpreted as a conservative estimate of the control horizon of a model-based control system. The free-run models for the Cu and Pb failed almost immediately and those results are shown in Figure 4.29.

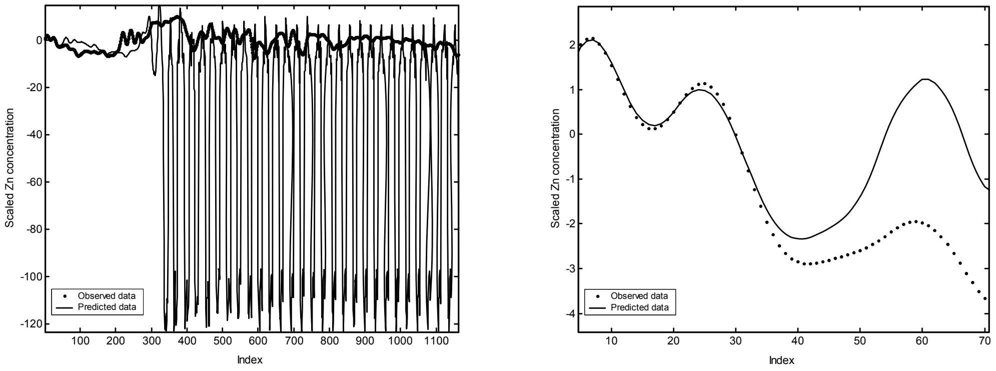


Figure 4.28 (a) Free-run prediction of Zn in the scavenger circuit by use of Model A, and (b) a close-up of the data shown in (a).

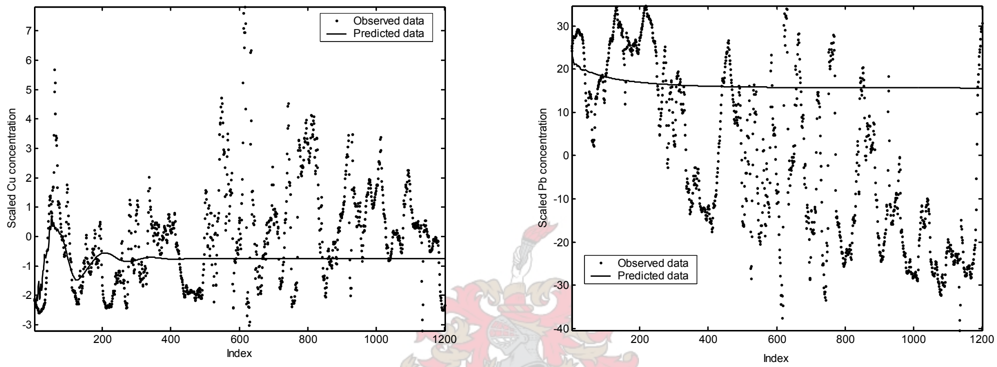


Figure 4.29 Free-run predictions of (a) Cu and (b) Pb concentrations in the scavenger circuit of the base-metal flotation plant.

4.6 Lead flotation plant

4.6.1 Background and singular spectrum analysis

In the final example, a froth monitoring system on a lead flotation plant is considered. Images from the froths in a zinc rougher cell were captured and digitised, after which image features were extracted from the images. Five such features were extracted, viz. x_1, x_2, x_3, x_4 and x_5. These features were identical to the AGL, SNE, ENT, SM and INSTAB features respectively, as described by Moolman (1995) and Moolman et al. (1995). AGL was the average grey level of the froth, the SNE, ENT and SM features could be related to the bubble size of the froth, and as before INSTAB gave an indication of the instability of the froth.

The autocorrelation functions of the various time series (x_1 to x_5) are presented in Figure 4.30. It is clear from this figure that the degrees of correlation between the consecutive observations of the time series vary significantly among the various series. The three variables relating to the bubble size of the froth, represented by x_2, x_3 and x_4 illustrate a very low degree of correlation, which is probably indicative of a high percentage of measured noise present in the measurement of this data. The other two variables, x_1 and x_5, show a much higher level of correlation, with x_5 only reaching the point of linear decorrelation at a window length of 340 (not shown in Figure 4.30). For all the time series the embedding dimensions were specified as the points of linear decorrelation, except for x_1 and x_5, which were taken at the minima occurring at window lengths of 76 and 151 respectively. The embedding dimension for x_2 was taken as 36, for x_3 as 42 and for x_4 as 29.

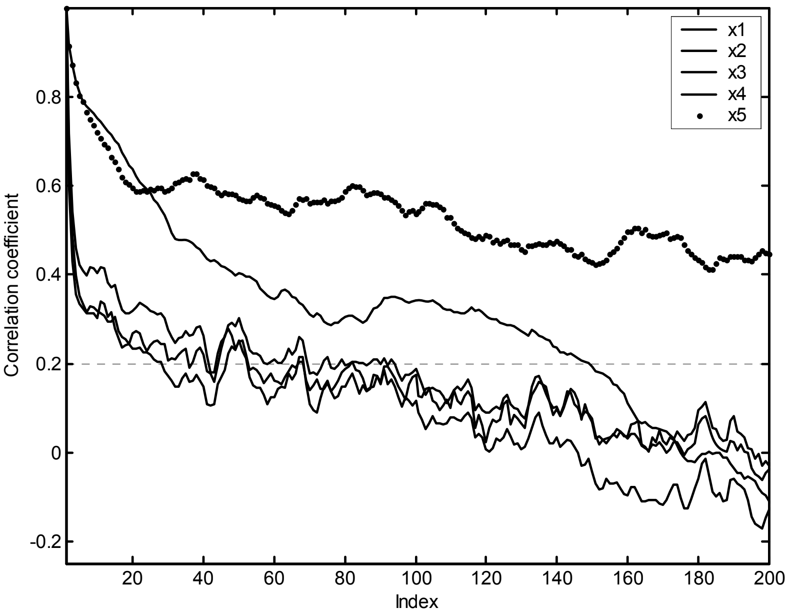


Figure 4.30 Autocorrelation function of variables x1 to x5 from lead flotation plant.

The resulting eigenvalue spectra from series x_1, x_2, x_3, x_4 and x_5 are shown in Figure 4.31. For this set of data, a relatively large number of principal components of each series were retained in an effort to preserve as much of the variance observed in the data as possible. Therefore, for x_1, x_2, x_3, x_4 and x_5 there were 35, 22, 32, 18 and 68 components retained respectively. It can be seen from Figure 4.31 that for each of the time series, this is more than half of the principal components that were calculated – significantly more than that in the previous case studies in this thesis. However, even with such a large fraction of the components being retained, the amount of variance explained by the reconstructed time series was relatively small, varying between 87.8% for time series x_2 and 95.7% for x_1.

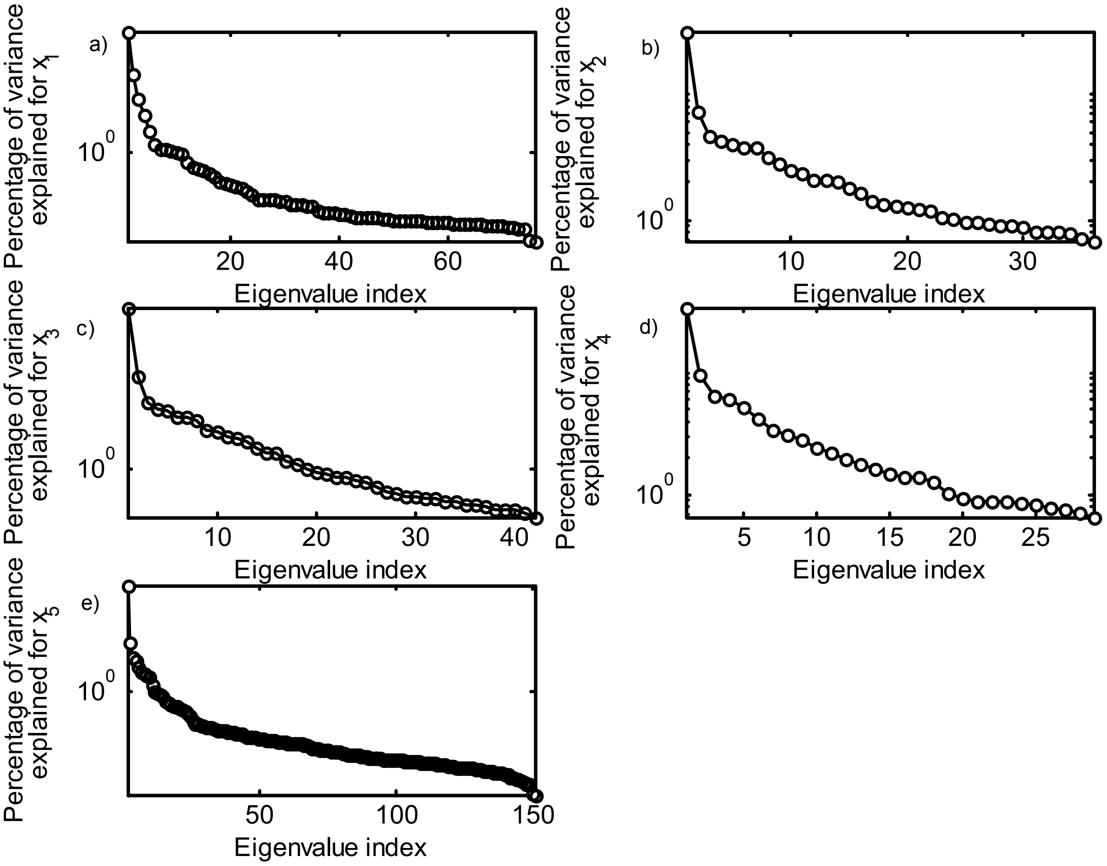


Figure 4.31 Percentage of variance explained by each eigenvalue and principal component for a) x_1, b) x_2, c) x_3, d) x_4 and e) x_5 time series.

4.6.2 Modelling

Modelling strategies A, B and D described in section 4.1 were implemented by following the same steps as before. In this case study, the time series were also considered as a multivariate system, which was previously described as the basis for modelling strategy D. The results are summarized once again in Table 4.4, with sections of the resulting model fits shown in Figure 4.32, and the r^2 values (equation 4.4) shown in the legends of the subplots.

Table 4.4 Summary of results obtained from modelling with different strategies (A-D), as well as the autocorrelation function, AC(1), at a lag of one for each data set from the lead flotation plant. Network configurations refer to the number of nodes in the input, hidden and output layers respectively

Time series	Network configuration	Validation data R^2	AC(1)
x_1 (A)	25:3:1	0.950	0.903
x_1 (B)	44:3:1	0.954	
x_1 (D)	16:3:1	0.902	
x_2 (A)	17:3:1	0.459	0.589
x_2 (B)	18:3:1	0.243	
x_2 (D)	10:4:1	0.335	
x_3 (A)	21:3:1	0.506	0.697
x_3 (B)	6:3:1	0.390	
x_3 (D)	25:3:1	0.487	
x_4 (A)	14:3:1	0.533	0.653
x_4 (B)	15:3:1	0.779	
x_4 (D)	15:3:1	0.460	
x_5 (A)	20:3:1	0.804	0.887
x_5 (B)	16:3:1	0.910	
x_5 (D)	13:3:1	0.906	

In the case of feature x_1, both strategies A and B performed similarly and significantly better than the AC(1) criterion. This was not the case with feature x_2, where strategies B and D led to models that performed markedly worse than the model built with strategy A. This can be attributed to the excessive smoothing that was done prior to reconstruction of the data. The principal component model could explain only 87.8% of the variance of the trajectory matrix, which meant that significant process information was lost, even though a large number of principal components were retained. Retaining even more principal components improved the results considerably. The same goes for the models used to predict x_3. Like x_2, the time series exhibited a high level of noise, as was seen in Figure 4.30 and is reflected by the relatively low AC(1) values for x_2 and x_3. Although x_4 showed a similarly high level of noise when compared to that of x_2 and x_3 and the smallest number of principal components were retained, strategy B performed considerably better than strategies A and D. With feature x_5, modelling strategies B and D again showed results superior to that of modelling strategy A.

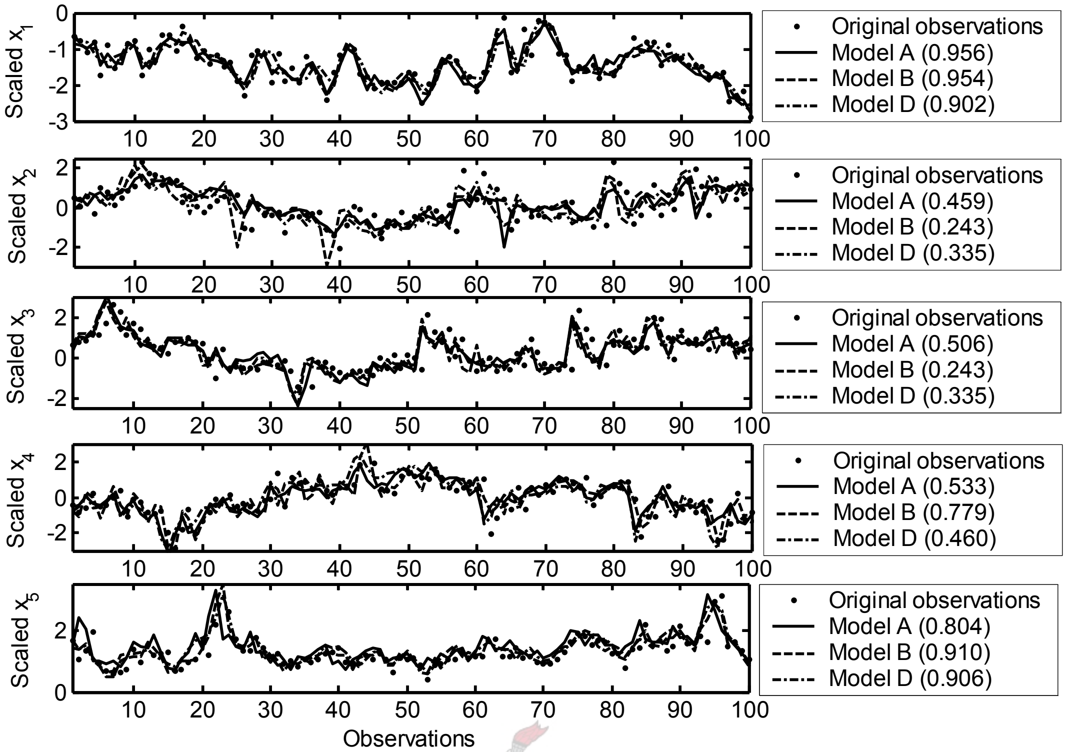


Figure 4.32 Model fits to froth features (x1-x5) with modelling strategies A, B and D. The number supplied in brackets next to each model indicate the correlation of the respective modelling results with the original time series.

4.7 Comparison between behaviour of SSA when analysing red vs. white noise

It can be seen from the above examples that SSA is extremely effective in removing noise from signals. However, all the series that were investigated were affected by noise known as 'white noise' or measurement noise. This is an identically distributed noise series that is independent from both the behaviour of the observed signal and the past behaviour of the noise signal. In theory, these two series, i.e. the noise and the signal, can be completely separated, as illustrated in equation

$$x_{t,\text{ observed}} = x_{t,\text{signal}} + \in_{t,\text{ noise}} \qquad 4.5$$

A more troublesome occurrence is the presence of 'coloured' or 'red' noise where the noise process at a particular observation is correlated to both the noise measured in nearby observations and the signal observations at nearby time steps.

The simplest form of red noise can be approximated by a first order linear autoregressive process, described in equation 4.6.

$$x_t = \alpha x_{t-1} + \in_{t,\text{ noise}} \qquad 4.6$$

where x_t is the observed variable at time t and x_{t-1} is the same variable one observation earlier. The constant, α, represents the lag correlation between the successive measurements of the time series and $\in_t$ is a random error term with zero mean and variance of σ^2. If $|\alpha| < 1$, the generated time series can be assumed to be stationary and the higher the value of α the further the noise fluctuate from the mean and the lower the fluctuations of the frequency will be. This autoregressive noise series can then be added to the original signal to simulate a time series contaminated by red noise.

Another approach to simulate a series contaminated by red noise would be to embed the noise in the measurement of the signal itself. An example of this type of noise would be where a sine wave was generated as a function of time, but the time measurement was contaminated by a random noise value, as illustrated in equation 4.7. This type of noise is referred to as dynamic noise.

$x_t = \sin(t + \in_t)$ 4.7

One of the problems experienced when analysing time series contaminated by an autocorrelated noise process was that the generated eigenvalues representing the noise process would not all be equal. This means that the noise floor would be sloping rather than being nearly-flat as can be seen with white noise processes, making it harder to distinguish between those eigenvalues representing signal and those associated with noise (Elsner and Tsonis, 1996).

4.7.1 Sine wave contaminated by various noise series

The differences between the three noise processes and the results obtained when performing SSA can be illustrated by the following example. A sinusoidal time series was generated and then contaminated separately by noise generated by all three techniques mentioned above. The relevant parameters, such as the value of α and the variance of the white noise process were also varied to illustrate their effect for each of the noise types.

For the white noise process, described by equation 4.5, the signal series was generated by the sine function and the random error functions $\in_{t,i}$ were generated with zero mean and variances of 0.15, 0.3 and 0.7 respectively. The resulting 'measured' time series x_t were calculated by equation 4.8 and are illustrated in Figure 4.33.

$x_{t,i} = \sin(t) + \in_t$ 4.8

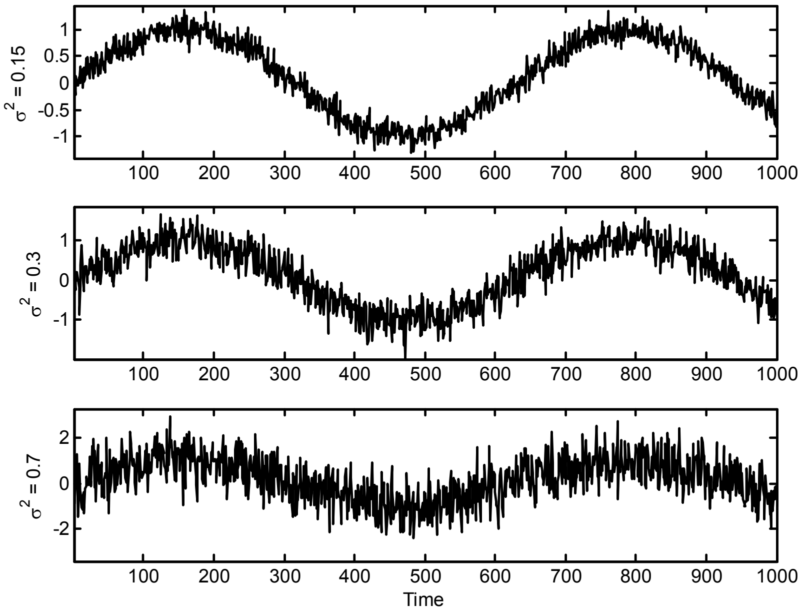


Figure 4.33 Sine wave contaminated by white noise of different variance.

For the red noise process, equation 4.6 was used to generate the noise. α was taken as the variable parameter and was specified at 0.2, 0.7 and 0.9 respectively. The signal series x_t and the random error function $\in_t$ were generated in a similar fashion as for the white noise process, with the variance of $\in_t$ taken as 0.3. The two series were added in a linear fashion, resulting in the series illustrated in Figure 4.34.

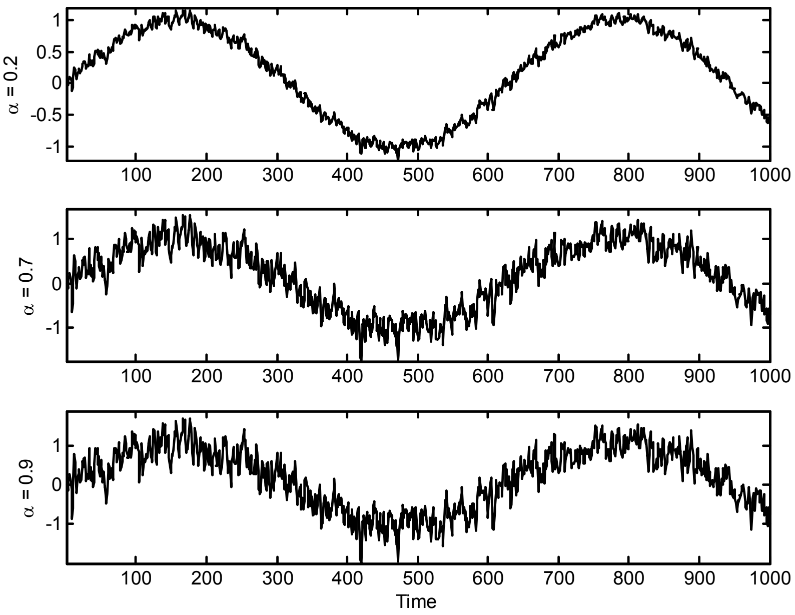


Figure 4.34 Sine wave contaminated by autocorrelated noise using different values of α.

Lastly, for the dynamic noise process, equation 4.7 was used, with the error function $\in_{t,i}$ once again generated with a number of variances, specifically 0.15, 0.3 and 0.7. The resulting time series are shown in Figure 4.35.

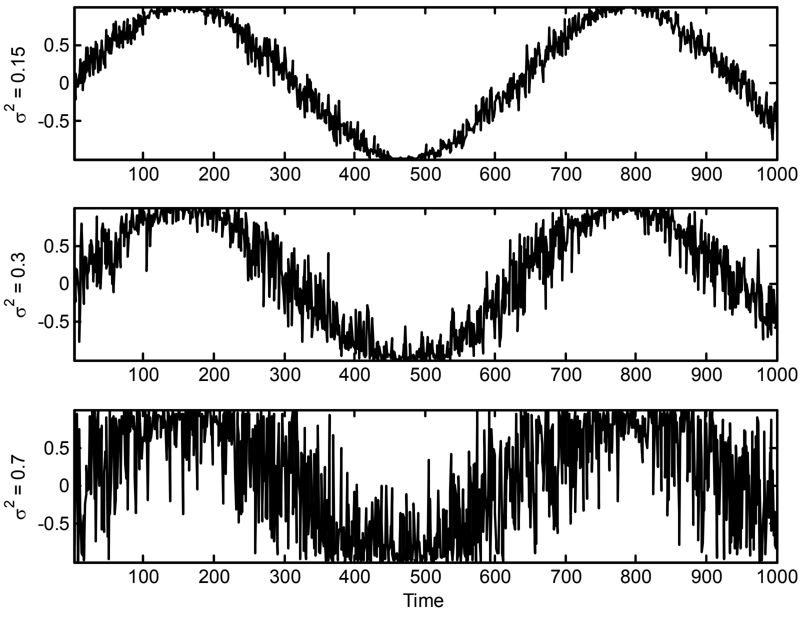


Figure 4.35 Sine wave contaminated by dynamic noise using different variances for the randomly generated error function.

It can be seen from Figure 4.33 - Figure 4.35 that, although the same basic time series was used, all the resulting 'measured' time series have been affected by the various forms of noise to differing extents.

4.7.2 SSA of sine wave

Standard SSA was applied to all nine time series shown in the previous section. The time series were all embedded in their optimum embedding window, which varied between a dimension of 118 and 140, depending on the nature of the noise that has been added to the sine wave. However, for all nine time series, only two principal components were retained, as one should be able to fully describe the dynamics of a sine wave by using two components. Table 4.5 supplies the amount of the variance in the contaminated series explained by the two retained components for each of the time series, as well as the correlation coefficient of the reconstructed series with the original, clean sine wave.

Table 4.5 Summary of results from SSA analysis on simulated sine wave time series supplying the percentage of the variance in the contaminated signal explained by two retained components, as well as the correlation of the reconstructed series with the clean sine wave.

Series	Percentage of variance explained	Correlation coefficient
White noise (σ^2 = 0.15)	95.9%	0.9998
White noise (σ^2 = 0.3)	91.6%	0.9989
White noise (σ^2 = 0.7)	52.3%	0.9939
Red noise (α = 0.2)	99.1%	0.9998
Red noise (α = 0.7)	90.0%	0.9979
Red noise (α = 0.9)	84.6%	0.9965
Dynamic noise (σ^2 = 0.15)	97.9%	0.9999
Dynamic noise (σ^2 = 0.3)	91.6%	0.9995
Dynamic noise (σ^2 = 0.7)	62.4%	0.9963

As would be expected, the amount of the variance explained by the two retained components differed for each of the series and the variance explained for the series with the higher noise variances was significantly less than for the ‘cleaner’ series. It is interesting to note that, even though the amount of the variance of the contaminated signals that was explained by the reconstructed series varied over a whole range of percentages, the correlation of all the reconstructed series with the original signal were within less than one percentage point of one another and were all extremely high. Unfortunately though, the signal to noise ratios of all the time series, especially those containing red and dynamic noise, were too high for the true effect (or lack thereof) of SSA on red noise to become apparent. In other words, the effect of the sine wave was too dominant and that of the red noise too weak and therefore SSA could analyse the signals contaminated by red noise just as successfully as the signals contaminated by white noise.

4.7.3 Alternative techniques to deal with coloured noise

The problem that has been attempted to be addressed in the section above, namely the inadequacy of SSA to discriminate clearly between red noise and true signal combined in a time series, has been studied and discussed in detail by Allen and Smith (1997). The rest of this chapter will be used to provide a brief summary of their suggested technique to deal with coloured noise.

It has been mentioned earlier (section 3.1.2) and in the discussion in section 4.7.1 that the eigenvectors of the covariance matrix used for principal component analysis ($\mathbf{C}_D$) are only the eigenvectors of the covariance matrix of the desired signal ($\mathbf{C}_S$) if the noise contaminating the signal is white noise. This complication necessitates the so-called ‘pre-whitening’ of the time series, which is similar to the pre-whitening performed in generalized regression or canonical analysis. The performance of this pre-whitening can shortly be described as follows.

If vector **e** is chosen to define a state space direction, the expected data variance to noise variance ratio in that direction can be defined as

$$\rho \equiv \frac{\mathbf{e}^{T}\mathbf{C}_{D}\mathbf{e}}{\mathbf{e}^{T}\mathbf{C}_{N}\mathbf{e}} \tag{4.9}$$

where $\mathbf{C}_N$ is the noise covariance and $\mathbf{C}_D$ the covariance matrix of the whole time series. Given that the noise covariance $\mathbf{C}_N$ is positive-definite, with eigenvalues forming the diagonal elements of Λ_N and eigenvectors $\mathbf{E}_N$, a coordinate transformation is defined

$$\mathbf{e}' \equiv \Lambda_N^{1/2} \mathbf{E}_N^T \mathbf{e}, \quad \mathbf{e} \equiv \mathbf{E}_N \Lambda_N^{-1/2} \mathbf{e}' \qquad 4.10$$

In these transformed coordinates, the noise has equal variance in all directions, so

$$\rho \equiv \frac{\mathbf{e}'^T \mathbf{C}'_D \mathbf{e}'}{\mathbf{e}'^T \mathbf{e}'} \qquad 4.11$$

where $\mathbf{C}'_D$ and $\mathbf{C}'_S$ are the transformed covariance matrices and are defined by

$$\mathbf{C}_D' \equiv \Lambda_N^{1/2} \mathbf{E}_N^T \mathbf{C}_D \mathbf{E}_N \Lambda_N^{-1/2} \qquad 4.12$$

$$\mathbf{C}_S' \equiv \Lambda_N^{1/2} \mathbf{E}_N^T \mathbf{C}_S \mathbf{E}_N \Lambda_N^{-1/2} \qquad 4.13$$

The vector $\mathbf{e}'$ which will maximise ρ in equation 4.11 is simply the eigenvector of $\mathbf{C}'_D$ that has the largest eigenvalue. As a result of the coordinate transformation, the process covariance of $\mathbf{C}'_D$ will be equal to the process covariance of $\mathbf{C}'_S$ plus the identity matrix. The eigenvectors of $\mathbf{C}'_D$ will now be consistent estimators of the eigenvectors of $\mathbf{C}'_S$. Thus, calculating the $\mathbf{e}'$ with the eigenvectors of $\mathbf{C}'_D$ (the columns of $\mathbf{E}'_D$ arranged in order of decreasing eigenvalue) will provide an optimal and consistent set of signal-to-noise maximising vectors, where the signal to noise ratios are given by the eigenvectors

$$\Lambda'_D = \mathbf{E}'^T_D \mathbf{C}'_D \mathbf{E}'_D \qquad 4.14$$

Once the dominant eigenvectors have been identified, these vectors can be transformed back to the original coordinates to ease the interpretation of the vectors. This is done by

$$\bar{\mathbf{E}}_D \equiv \mathbf{E}_N \Lambda_N^{-1/2} \mathbf{E}'_D \qquad 4.15$$

The most useful property of these signal-to-noise maximising eigenvectors is that their expected orientation is independent of the noise variance, if $\mathbf{C}_N$ correctly reflects the noise autocorrelation. One could thus obtain a consistent estimate of the patterns that would be observed when analysing the time series in the absence of any noise. This makes this technique an optimal linear filter for the reconstruction of signal in the presence of correlated noise.

4.8 Summary

Singular spectrum analysis has proved a very useful tool to perform the filtering of data before the data are modelled by using neural networks. In all the case studies considered, the models built on the data after SSA was applied, outperformed the models that were built on the time series alone.

In the case of the carbon-in-leach gold leaching process, it was shown that a reconstruction of the time series with only three components outperformed an optimised moving average filter in terms of their correlation with the original clean signal.

It was found that the success of the various modelling strategies in which SSA was involved, varied according to the nature of the time series. For some of the series, the models built on the plain reconstructed series performed best, while for others the extra information that was extracted during multivariate embedding proved to make a difference and for still other series the models built on the individual reconstructed components were the most successful.

A secondary investigation also proved the importance of choosing a long enough window length for the time series. This will prevent that some of the information from the time series get lost during the embedding and subsequent decomposition.

In conclusion, the difference in the behaviour of SSA, when faced with data contaminated with red noise vs. white noise, was investigated. Even though the results obtained from the case study did not illustrate the problem as clearly, a comprehensive discussion from the literature was supplied to serve as a solution to the inadequacy of SSA to handle red noise satisfactorily, if the signal to noise ratio is low enough.

5 NONLINEAR SSA

It has been shown in many fields of society that the viewpoint of small causes leading to small effects is merely wishful thinking from a school of thought that would like to explain the world in linear terms (Schreiber, 1998). The reality is that nonlinear processes can be found in abundance, with the engineering industry being no exception. It would therefore make sense to extend the successful analysis of linear processes from the previous chapter to also describe or analyse nonlinear processes or systems that display nonlinear dynamics. This will be done by implementing two nonlinear techniques, namely the localized principal component analysis and auto-associative neural networks. Two case studies that will be investigated, of which one is a simulated flow in a series of tanks and the other real electrochemical noise data obtained from a corrosion measurement system.

5.1 Flow in a series of tanks

5.1.1 Background

To illustrate the use of nonlinear singular spectrum analysis, consider the response of the flow of four tanks in series.

Figure 5.1 shows the actual (solid line) and simulated measured (+) response to a pulsed input (broken line). The measured response was simulated by adding an artificially generated nonlinear error to the actual response. The error function was generated by

$$\varepsilon'_i = \varepsilon_i + (\varepsilon')_{i-1}^{1/2} \qquad 5.1$$

where ε was generated randomly with a normal distribution, zero mean and a standard deviation of 0.1.

The simulated time series consisted of 6000 observations (only 3000 of which are shown in Figure 5.1 for clarity purposes), with a period of 300 observations between each of the step inputs.

For the performance of localized SSA, the time series was divided into twenty equal sized parts, each consisting of 300 observations and the break between the different parts coinciding with a change in the step input signal. Each of these parts was then analysed independently from the other parts.

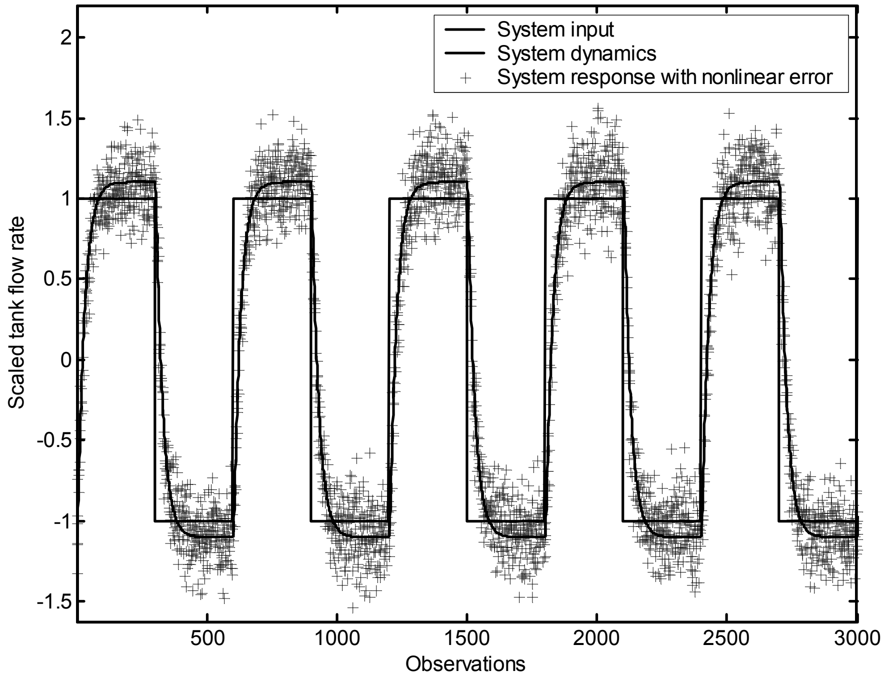


Figure 5.1 Actual system response (solid line), simulated measured response with nonlinear error (+) and pulsed input signal (broken line) obtained from flow in four tanks in series.

5.1.2 Linear and localized SSA of four tanks in series process

a) Autocorrelation

A profound difference between the behaviour of the series as a whole and the individual parts could be observed during the analysis. The first evidence of this could be seen from the autocorrelation functions, which were used to determine the optimum embedding window for the time series.

The autocorrelation function of the complete time series, displayed in Figure 5.2, appears significantly smoother than those of the individual parts of which four representative autocorrelation functions are illustrated in Figure 5.3. The correlation between the observations of the whole series is also higher than that between the observations of the individual parts.

Due to the periodic nature of the whole series, the autocorrelation function would start increasing again at a lag of 300. However, as the series reaches a point of decorrelation long before that, the trouble was not taken to embed the complete time series into a dimension that high.

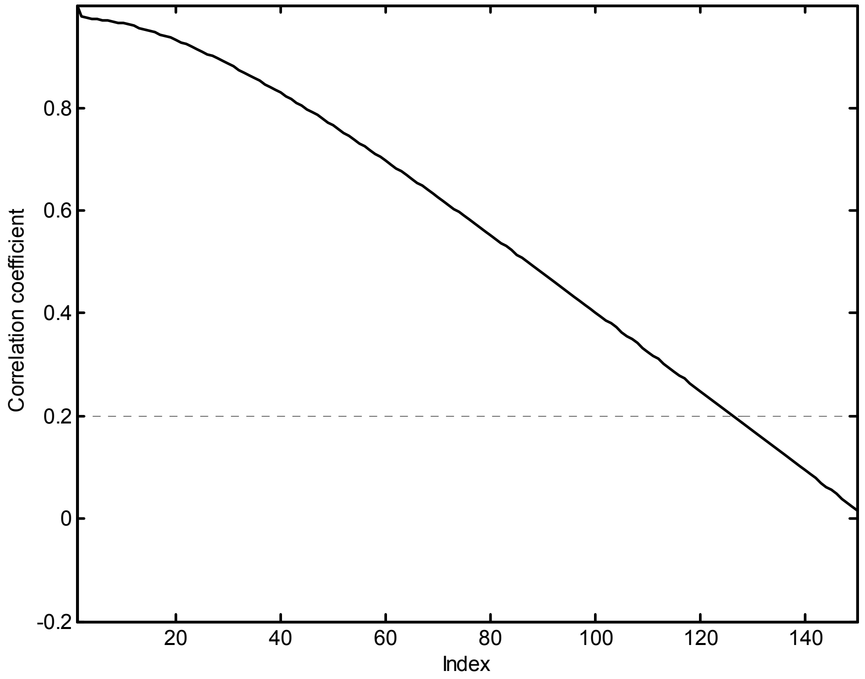


Figure 5.2 Autocorrelation function of four tanks in series time series obtained from linear SSA.

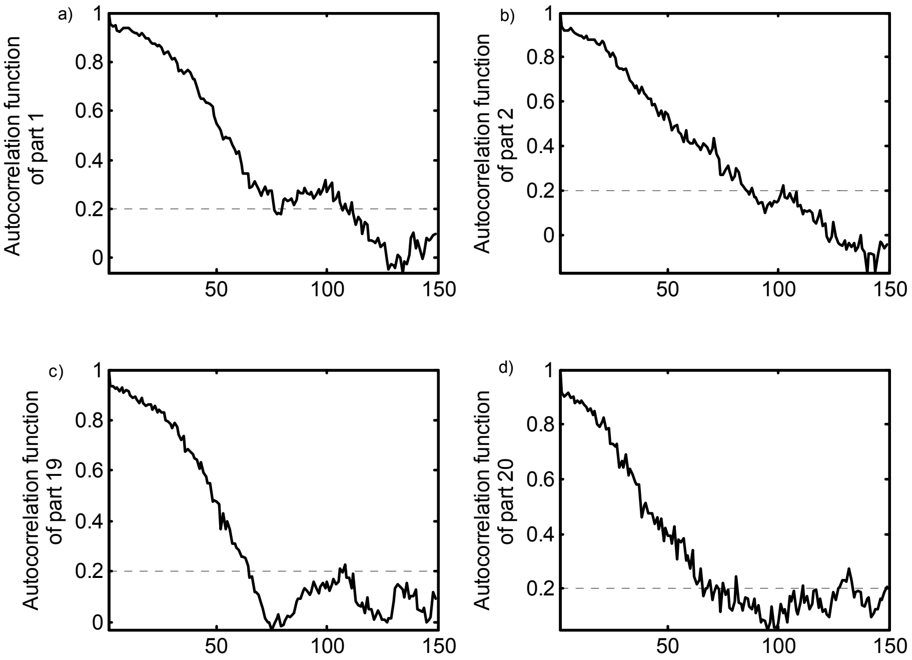


Figure 5.3 Autocorrelation function of a) part 1, b) part 2, c) part 19 and d) part 20 of four tanks in series time series obtained from localized SSA.

Based on the observed autocorrelation functions, both the original time series and the separate parts were embedded in trajectory matrices. The embedding dimension of the original time series was 127 and that of the individual parts varied between 63 and 96.

b) Eigenvalue distribution

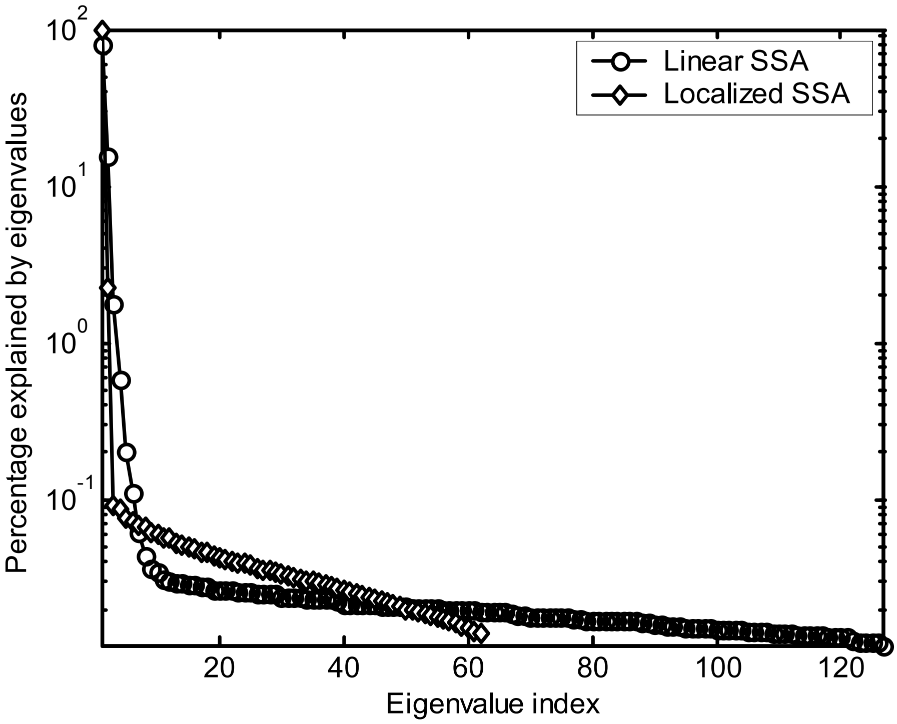


Figure 5.4 Percentage of variance explained by eigenvalues of four tank time series analysed by linear SSA.

Another marked difference between the two techniques could be seen from the distribution of the eigenvalues. In Figure 5.4 the eigenvalues gradually decrease to a relatively long noise floor with the first nine eigenvalues more or less evenly spaced along this decrease. In comparison to this, the eigenvalue spectra taken from the first, second, nineteenth and twentieth parts of the whole time series, presented in Figure 5.5, show a much sharper drop before a more gradual noise floor is reached. It would also seem as if only the first two eigenvalues, compared to the first nine, is relevant. Although Figure 5.5 only displays the eigenvalues of four of the parts of the time series, this behaviour was observed in all of the parts of the four tank time series analysed by localized SSA. Two components were retained for each of the parts of the four tanks time series and these two components explained between 97% and 98% of the variance that was observed in the part of the noisy time series, whereas nine components from the linear SSA were necessary to explain 97.7% of the variance.

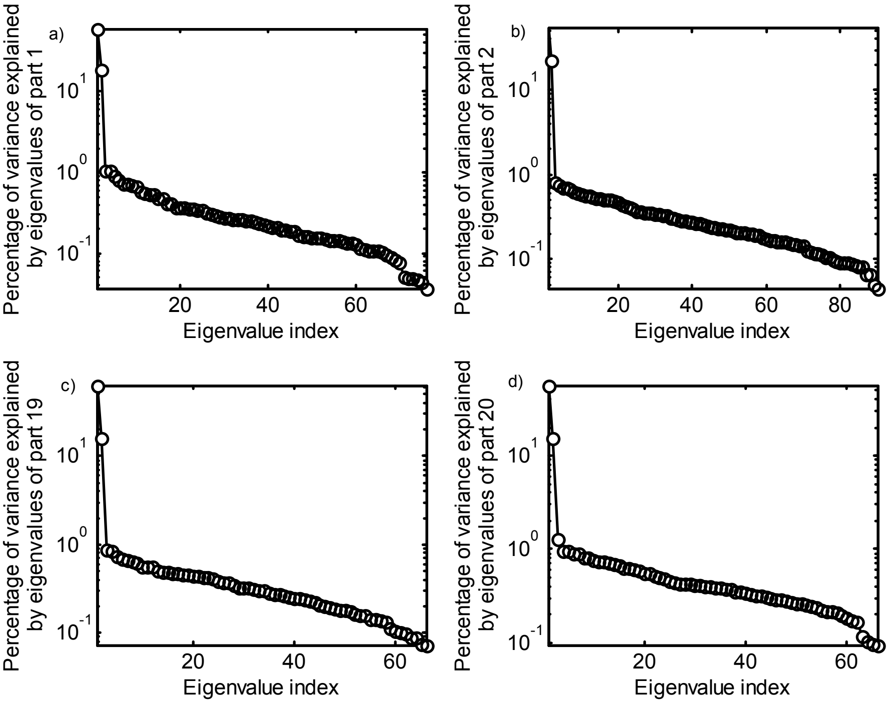


Figure 5.5 Percentage of variance explained by eigenvalues of a) part 1, b) part 2, c) part 19 and d) part 20 of the four tank time series analysed by localized SSA.

c) Reconstruction

The time series were then reconstructed from the reduced number of principal components and the reconstructed parts were combined in order to obtain the localized SSA version of the reconstructed time series. For comparison purposes, only two of the components from the linear (complete) analysis, similar to the individual parts of the localized analysis, were retained. Sections of these reconstructions are displayed in Figure 5.6, along with the original system dynamics, uncorrupted by the nonlinear noise.

It was decided to rather compare the resulting reconstructions with the uncorrupted series than the simulated noisy response because this would be a more accurate way to determine if the nonlinear noise has been successfully removed. It can be seen in Figure 5.6 that, except for spiky behaviour in the region of the changes in the step input, the results from the localized SSA follows the path of the system dynamics significantly closer than the reconstruction obtained from linear SSA, which struggles to follow the system dynamics accurately. This observation is supported by the correlation coefficient where the reconstructed series from localized SSA has a correlation of 0.999 with the original ‘clean’ data series and the reconstructed time series from the linear SSA has a correlation of 0.992. The difference in the correlation with the original system dynamics of the two time series is almost insignificant in magnitude; however localized SSA has the added advantage that a significantly smaller number of principal components have to be retained to obtain the same representation of the data.

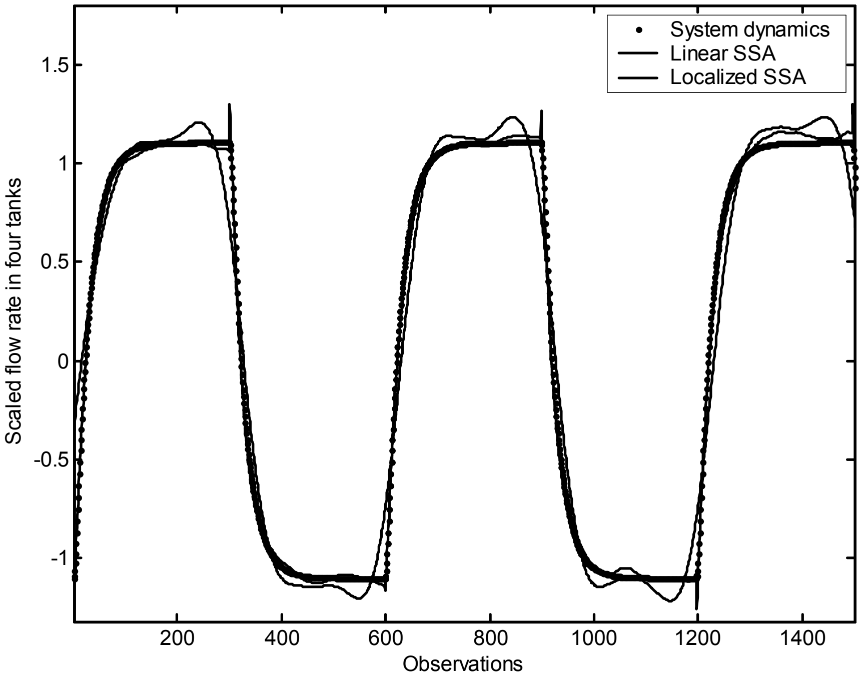


Figure 5.6 Section of the reconstructed time series obtained from linear SSA (solid line) and localized SSA (dotted line) compared to the true system dynamics (dotted markers).

The final indication of the difference in the results obtained when analysing the four-tank time series with linear SSA compared to localized SSA is the individual tp^T or reconstructed components extracted. These components have been presented in Figure 5.7 and Figure 5.8, where the localized SSA has been expanded to include more principal components in each part for comparative purposes. When the first reconstructed component in the two figures are compared, it can be seen that where linear SSA first approaches the time series from the basis of a sine wave, localized SSA could already with the first component identify the underlying saw-tooth structure of the time series.

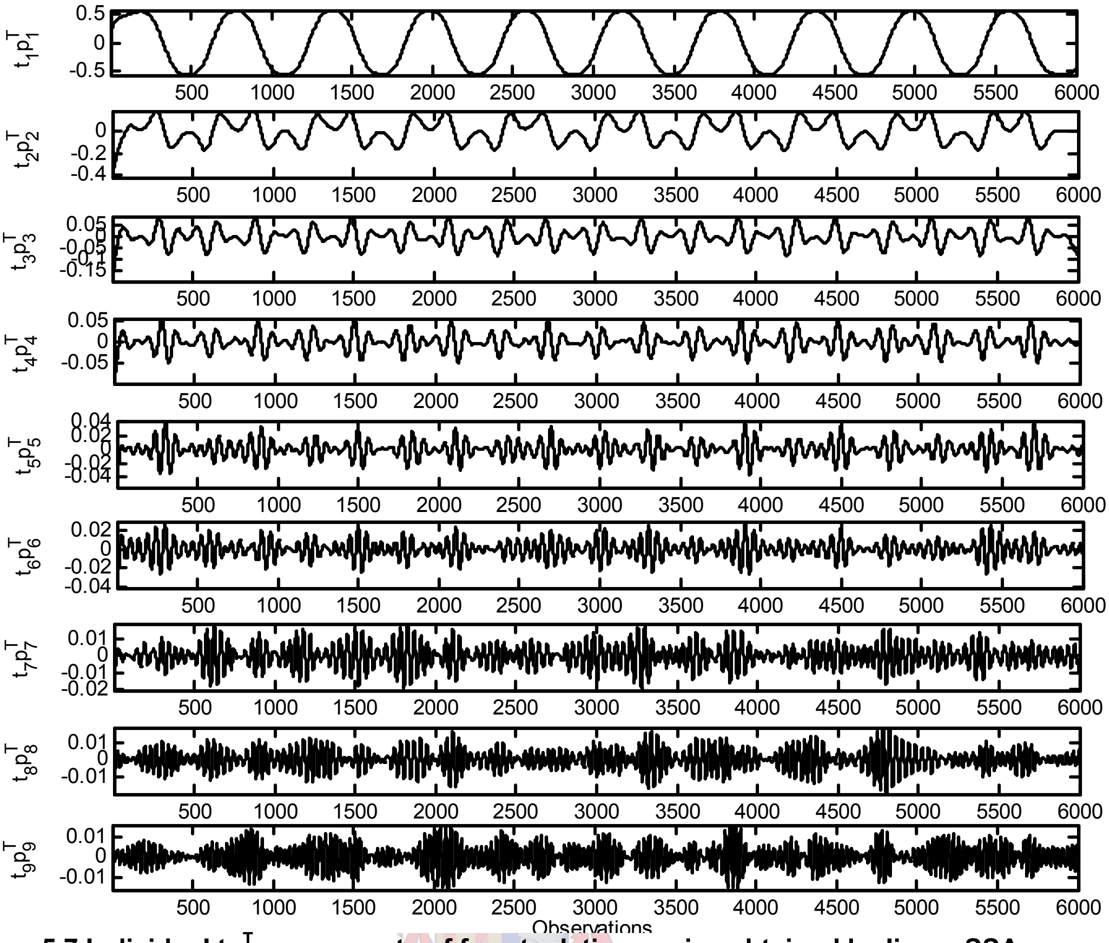


Figure 5.7 Individual tp^T components of four-tank time series obtained by linear SSA.

From Figure 5.7 it can be seen that the reconstructed time series from linear SSA is steadily 'built up' into a more 'square' series by adding more and more components, but is also simultaneously being contaminated more by the embedded noise. On the other hand, from Figure 5.8, it seems as if it could have been sufficient to retain only one component from the localized SSA, as the unwanted 'spikes' are introduced by the second component.

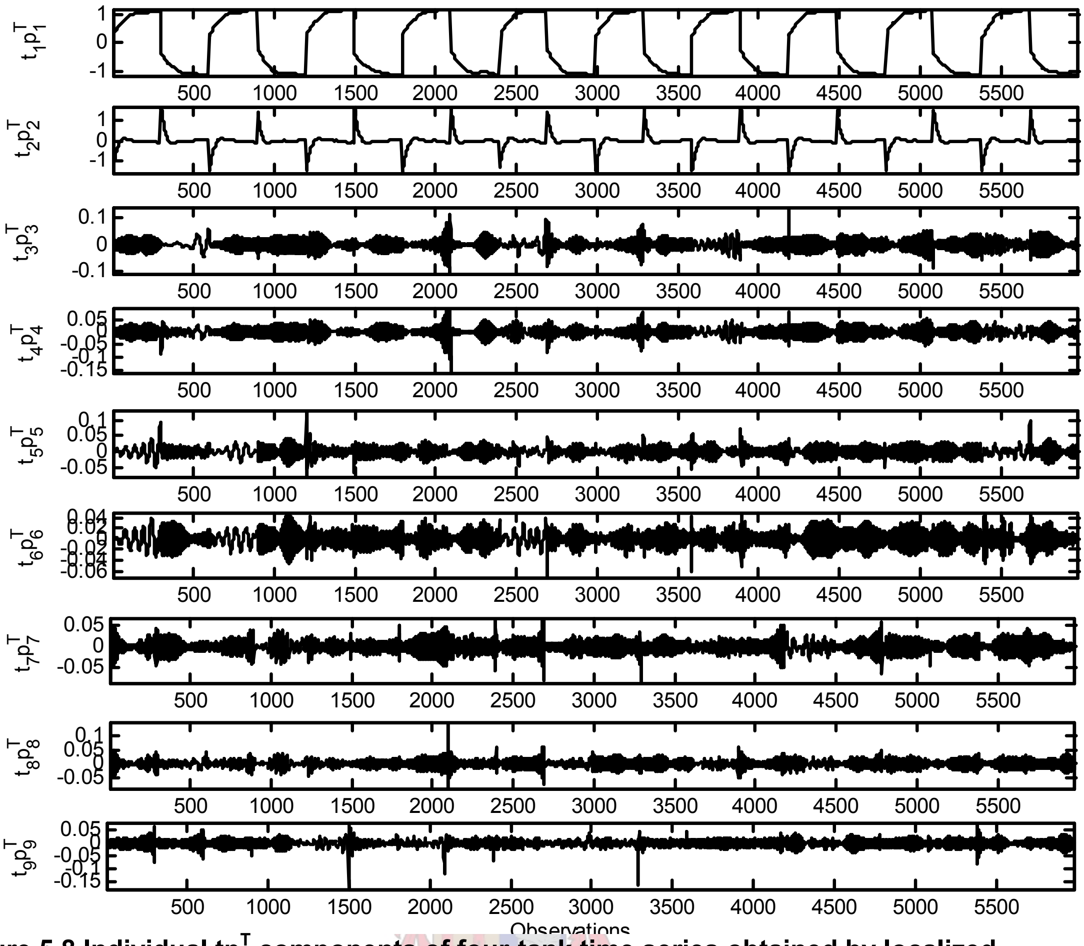


Figure 5.8 Individual tp^T components of four-tank time series obtained by localized SSA.

The possibility that the weaker correlation between the linear SSA results and the true process dynamics of the four-tank system was a result of too few principal components being retained, was also considered. The more eigenvalues one retain, the larger amount of the original variance is included in the reconstructed time series, gradually moving from including more and more relevant system information to including more and more irrelevant noise. The reconstruction was also done for all nine of the components obtained from linear SSA that seem to lie on the steep slope of the eigenspectrum, but these results just showed the influence of more of the nonlinear noise embedded in the simulated data. For a 'real' time series obtained from an industrial plant, one does not have the advantage of knowing the true process dynamics and one is therefore not able to choose the number of eigenvalues that are the most convenient to best describe the dynamics. One needs to analyse the time series based on available information, which in this situation would have been the eigenvalue spectra and in which case the localized SSA provided a significantly sounder base for judgement than the linear SSA.

5.1.3 Auto-associative neural network

It was attempted to also extract the nonlinear principal components from the time series by using an auto-associative neural network. This network was configured in a way similar to the approach used by Hsieh (2001), Hsieh and Wu (2001a), Hsieh and Wu (2001b), Hsieh and Wu (2002), Hsieh and Hamilton (2003) and Newbigging et al. (2003), in that the nine principal components that were extracted by linear SSA were used as the input series for the network. However, despite trying a number of different approaches and network configurations, it was found that this network could not extract nonlinear principal components from the four tanks in series data set. Although this is very disappointing, it is not wholly unexpected, as it is well known (Tan, 1996, Chang, 2001) that auto-associative neural networks may have trouble during the training process, especially if more than one hidden layer is used.

5.2 Electrochemical Noise Process

5.2.1 Background

These data were obtained from an experimental set-up where the electrochemical noise occurring during the corrosion of material was measured and recorded (shown in Figure 5.9 and Figure 5.10 (De Wet, 2001)). For the purposes of this case study, corrosion was considered as the destruction or deterioration of a material as a result of interaction with the environment. 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.

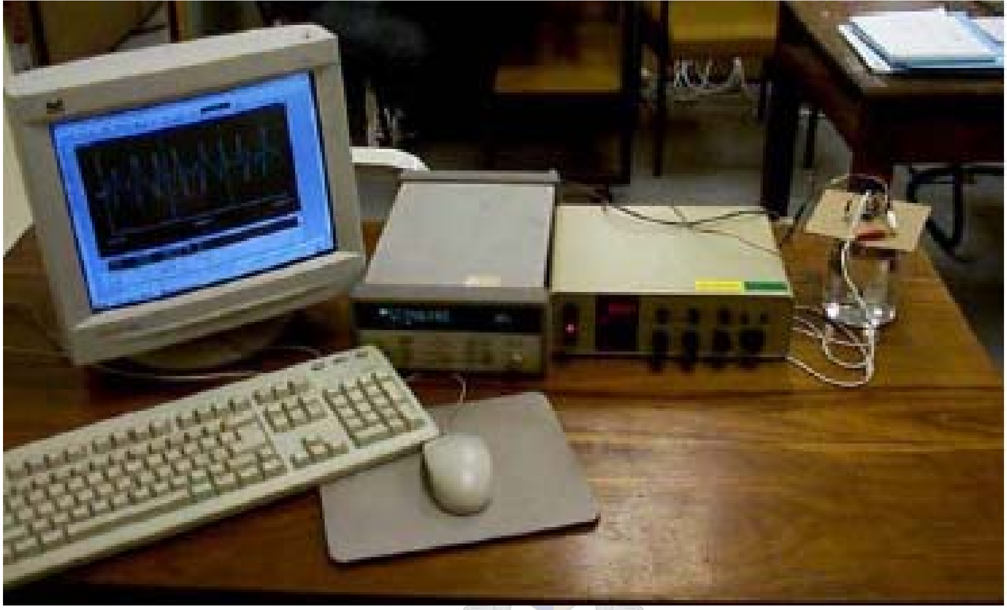


Figure 5.9 Experimental set-up used to measure electrochemical noise data.

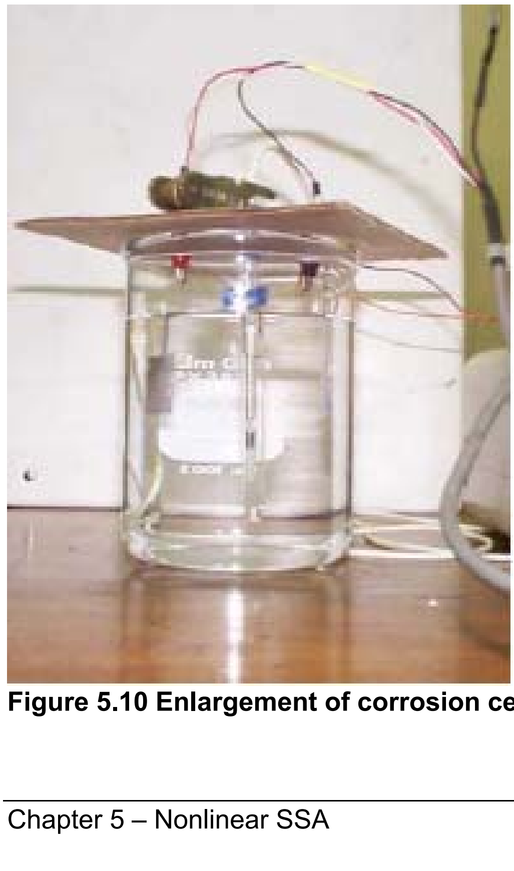


Figure 5.10 Enlargement of corrosion cell

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 accurately measure the potential and current noise simultaneously, a three-electrode sensor was required. 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 and the resulting length of the time series was 3156 observations.

The observed time series from the current noise is displayed in Figure 5.11. However, it can be seen that the data appear very ‘spiky’ and it is hard to observe any trends just from the inspection of the time series. Figure 5.12 displays a representative section of this time series to provide a close-up view of the behaviour of the observations, once again illustrating the spiking nature of the current noise time series.

The time series measured from the voltage noise of the process is displayed in Figure 5.13. It is clear that this is a non-stationary process and that the voltage steadily decreased over time. However, from the close-up of a section of the voltage time series, displayed in Figure 5.14, one can see that this time series also display a certain level of ‘spiky’ behaviour, but not as severe as that observed for the current data.

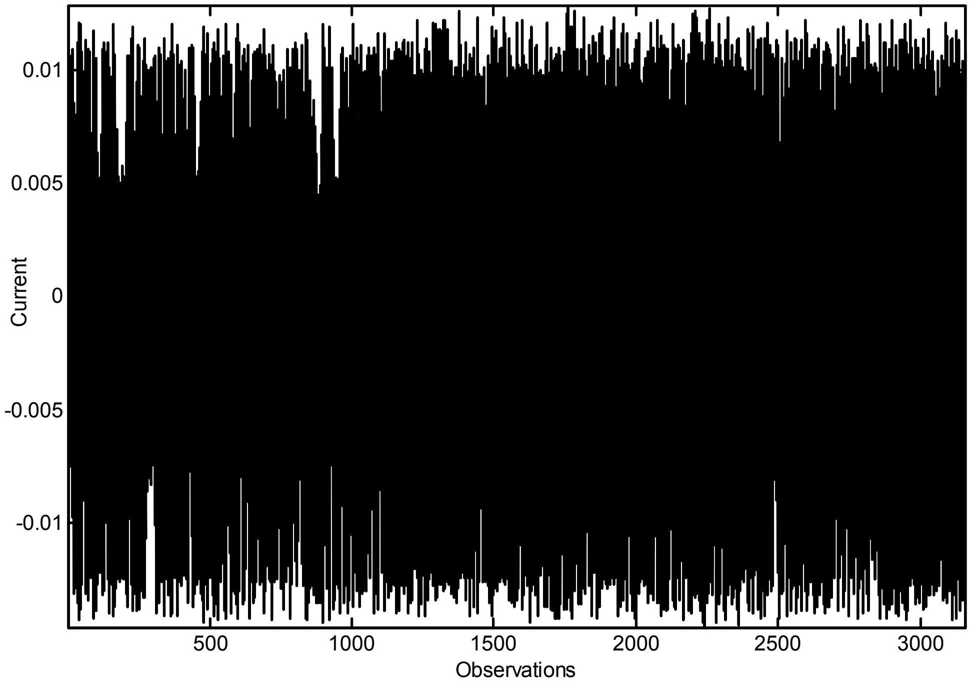


Figure 5.11 Original current observations from the electrochemical noise process.

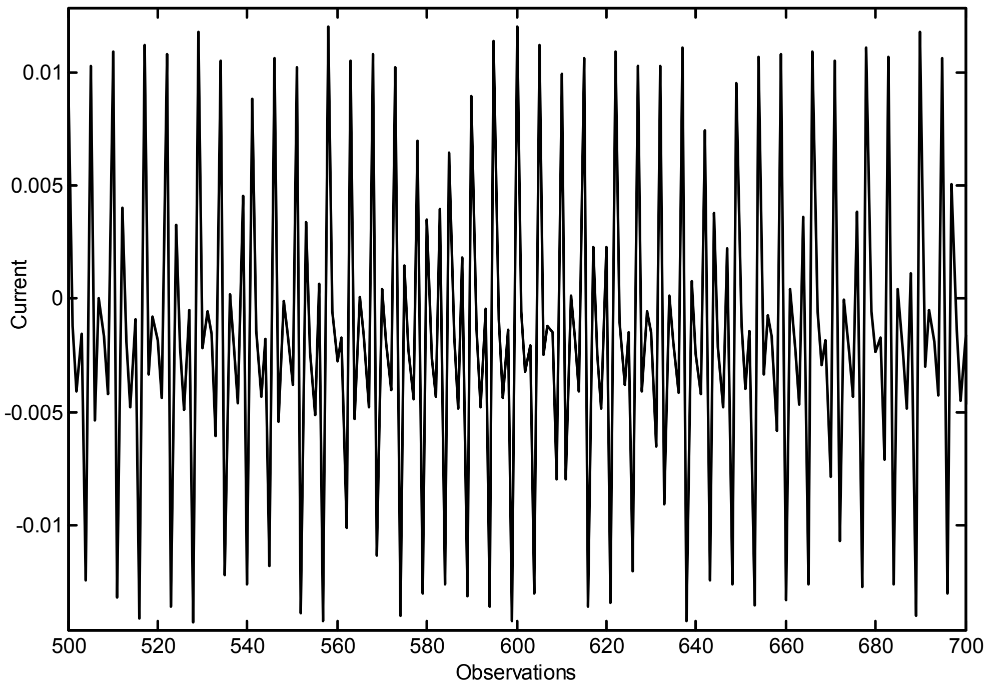


Figure 5.12 Section of the original current observations from the electrochemical noise process.

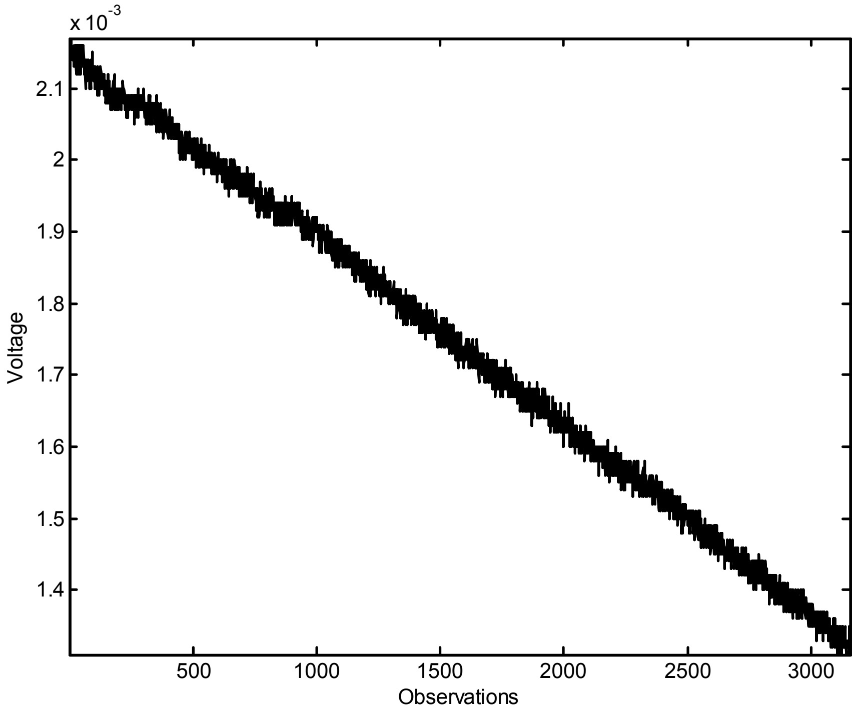


Figure 5.13 Original voltage observations from the electrochemical noise process.

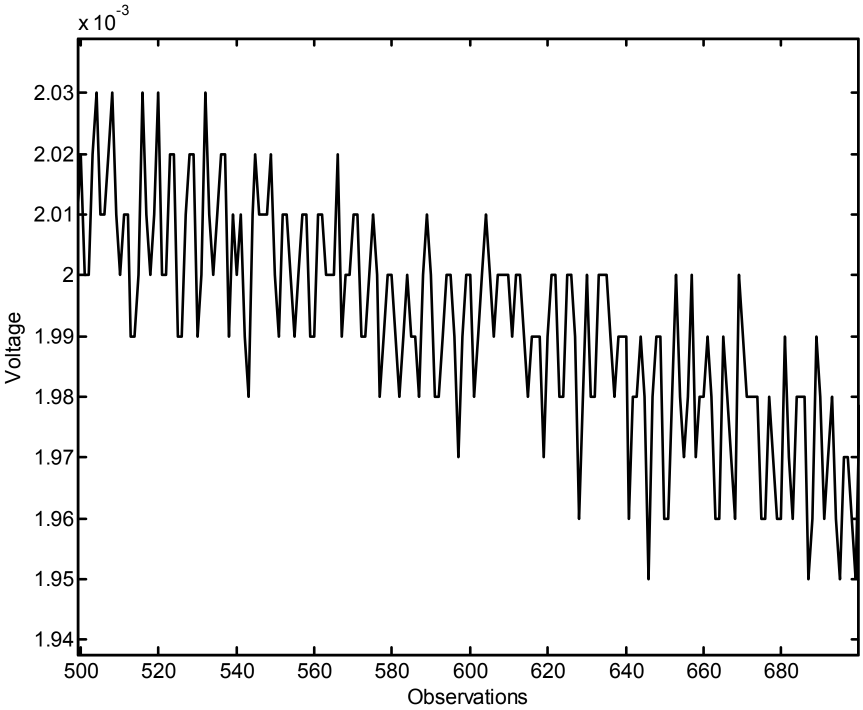


Figure 5.14 Section of the original voltage observations from the electrochemical noise process.

In order to perform the localized SSA, the time series were also divided into a number of parts, each of which could be analysed independently. The usual approach used to determine the positions at which the time series should be split into different parts, is to visually inspect the original time series and try to identify obvious changes in the behaviour of the time series, as was done for the previous case study. However, due to the large variance of the current time series seen in Figure 5.11, this approach was inappropriate. It was therefore decided rather to visually inspect the first and second reconstructed components obtained from linear singular spectrum analysis. This simplified representation of the original time series can be seen in Figure 5.15 and was used to identify the four so-called split points at which the current time series was divided into five parts. These split points occurred at 220, 790, 1000 and 2500 observations. The five different sections of the time series were therefore not all of equal length but rather divided to represent different sections of behaviour in the current time series.

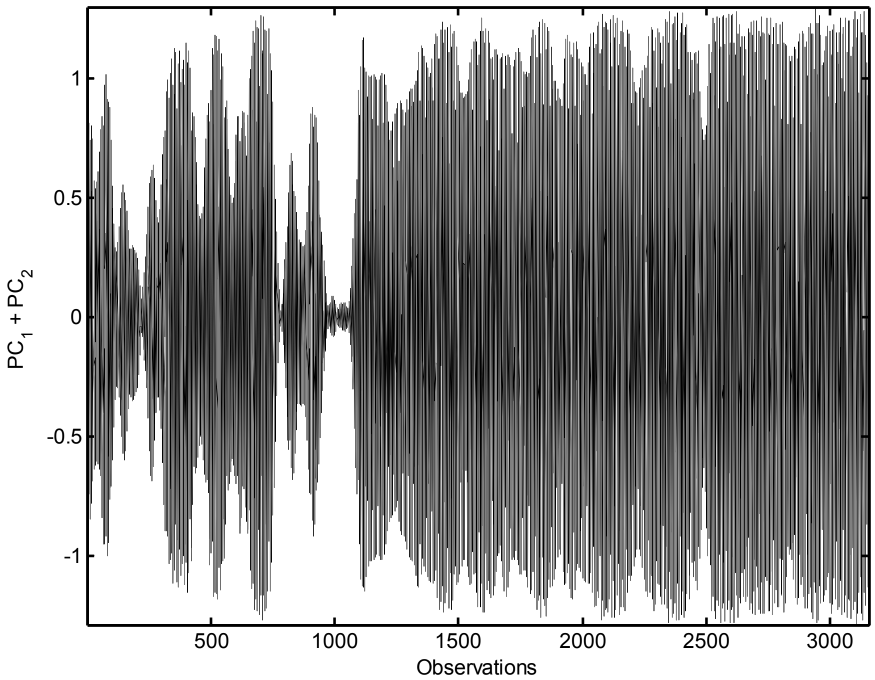


Figure 5.15 Combined reconstruction of first and second principal components of the current time series from the electrochemical noise process, indicating the approximate positions of the break points for the different parts of the time series for localized SSA purposes.

5.2.2 Linear and localized SSA of electrochemical noise process

In order to provide a benchmark with which to compare the results from the various approaches, 'normal' or linear singular spectrum analysis was first performed on the time series, where-after localized singular spectrum analysis, as was discussed in section 3.4.2, was performed. The results of the current time series will be presented simultaneously to better illustrate comparisons and differences between the two sets of results. The results from the voltage analysis were very similar to that of the current time series, and because no new insights or discussions could be derived from the voltage results, it was decided not to include these figures.

a) Embedding

Figure 5.16 illustrates the autocorrelation function of the current time series. From this correlation function, it would appear that the data has a very low degree of correlation, which is probably due to the distinct spikes that can be observed in the time series. These 'spikes' are most likely the result of a too low sampling rate during the experimental measurements, which introduces a certain degree of unreliability into the data.

However, it would also seem as if the 'band' of correlation coefficients is decreasing with an increase in window length, and it was decided rather to use this 'band' than the actual autocorrelation function as a guideline for the window length. Based on this criterion, the optimum embedding window for the current electrochemical noise process was estimated to be 30.

Each of the parts of the time series was embedded separately after the optimum embedding window for that part of the time series has been determined. The behaviour of the autocorrelation function of the individual parts were very similar to that of the original time series illustrated in Figure 5.16 and the embedding dimension specified for sections 1, 2, 3, 4 and 5 of the current time series were 30, 30, 40, 40 and 50, respectively.

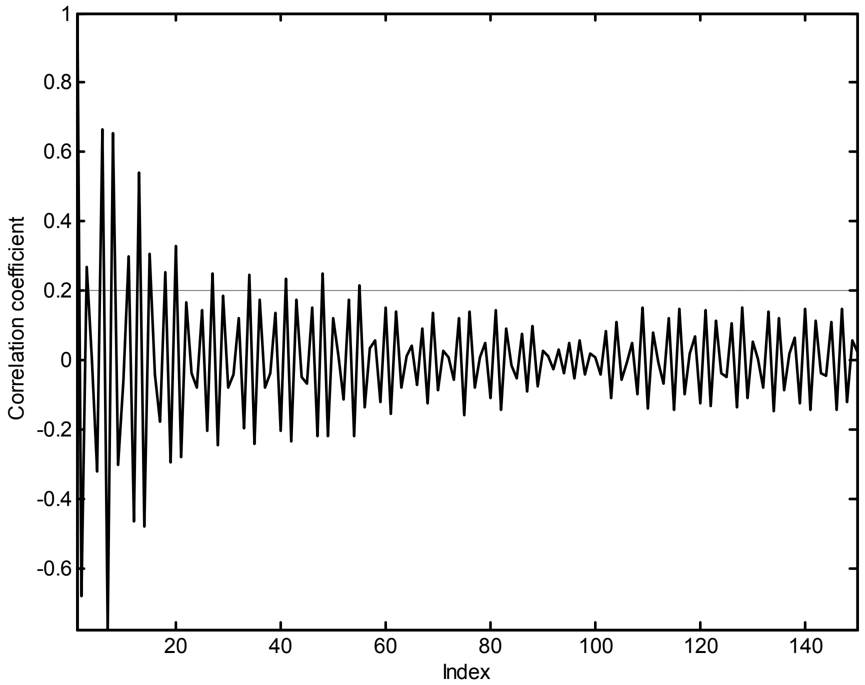


Figure 5.16 Autocorrelation function of current time series.

b) Eigenvalues

Once principal component analysis has been performed on the lagged covariance matrix of either the whole original series or of the relevant part of the original series, the eigenvalues could be extracted. In order to obtain the eigenvalue distribution of the localized principal component analysis, the minimum number of eigenvalues among the different parts was determined. In this case study, this constituted 30 eigenvalues. The first 30 eigenvalues extracted from each of the time series were then summed, after it were weighted to represent the fraction that each respective part forms of the whole time series. It was necessary to assume that the amount of variance explained by the last few eigenvalues are insignificant compared to that of the first 30 and after inspection of the relevant eigenvalue series, this assumption was confirmed.

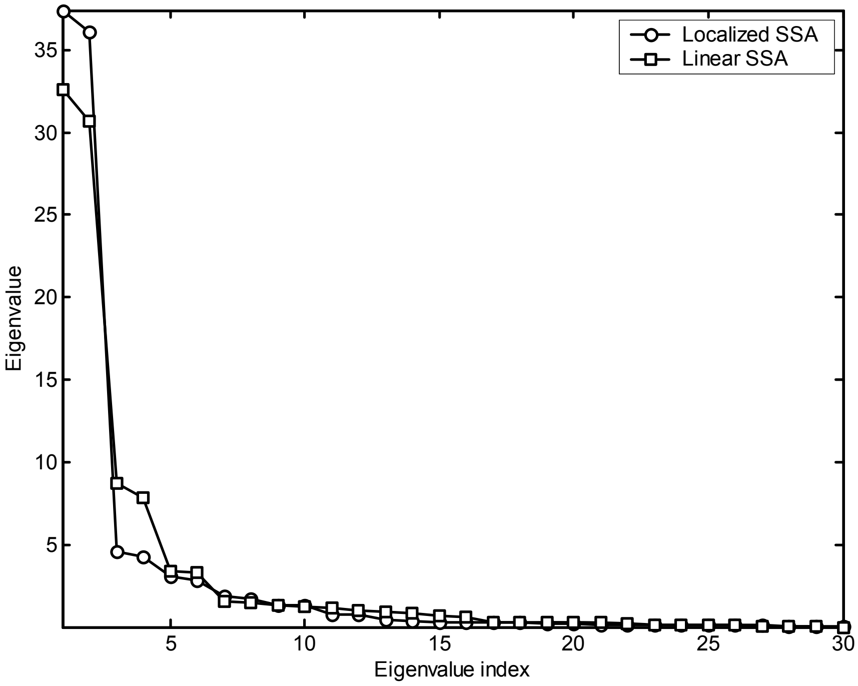


Figure 5.17 Eigenvalues of current time series from electrochemical noise observations, extracted by using localized (circles) and linear (square) SSA.

After this process was completed, one could compare the eigenvalue distributions obtained from the linear singular spectrum analysis and that from localized singular spectrum analysis, which was more representative of underlying nonlinear structures in the data. The resulting eigenvalue spectra are presented in Figure 5.17. From this figure one can clearly see the difference in the results obtained from the two techniques, especially in the values of the first four eigenvalues. The percentage of the variance explained by the first two principal components obtained from localized SSA is significantly higher than that from linear SSA. The noise floor appears also more defined for the localized SSA than for linear SSA, in that the third and fourth components obtained from the linear technique seem as if they could also be relevant.

Even though there was a distinct separation between the first few eigenvectors and the rest of the components of the time series, it was decided that, in order to ensure that all the information in the time series is retained, a relatively large number of the principal components should be retained for the reconstructed time series. By investigating the percentage of the variance explained by each of the eigenvalues, rather than the eigenvalues themselves, displayed on a logarithmic scale, as in Figure 5.18 one can obtain a more comprehensive impression of the distribution of the eigenvalues. These values are supplied for both the linear SSA results and the percentages explained by each of the eigenvalues, combined from the different parts of the localized SSA. The percentages were weighted according to the weight of each part of the time series, similar to the construction of Figure 5.17. The final number of principal components retained in the linear singular spectrum analysis was 16 and these components explained 97.6% of the variance observed in the original time series.

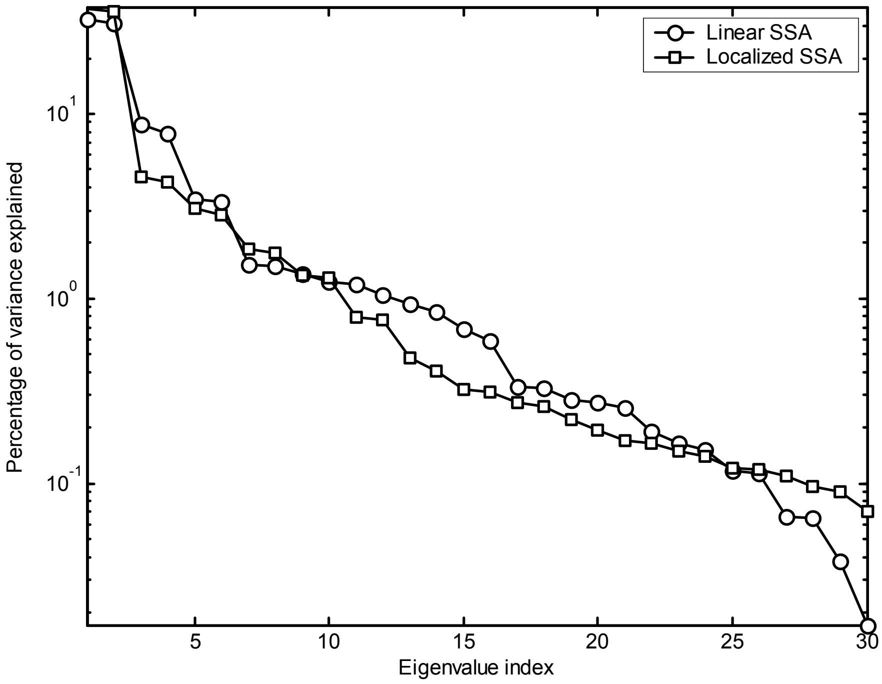


Figure 5.18 Percentage of variance explained by each eigenvalue obtained from linear SSA of current time series from electrochemical noise process.

The combined percentages explained by each eigenvalue of the eigenvalue spectrum in Figure 5.18 have been constructed by combining the percentages explained by the eigenvalues of each part of the series. These percentages for each part are presented on a semi-logarithmic scale (which allows a better distinction of the eigenvalues lying on the so-called floor of the spectrum) in Figure 5.19.

During the localized singular spectrum analysis, each of the sections that were analysed separately retained a different number of principal components, in order to better represent the behaviour of that section of the time series. The number of components retained for sections 1, 2, 3, 4 and 5 were 10, 10, 14, 12 and 12 respectively and explained 98.3, 96.3, 98.2, 94.7 and 96.0% of the variance in each respective part. It can be seen that, even though a smaller number of components were used for each part of the time series, the average amount of variance explained by localized SSA is similar to that explained by linear SSA.

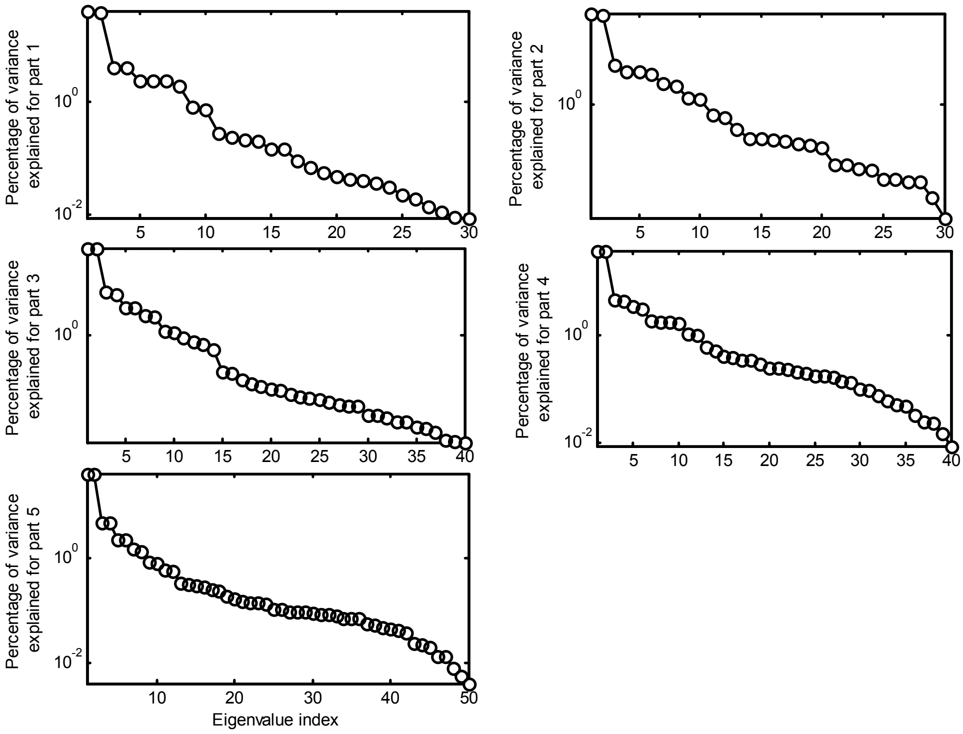


Figure 5.19 Percentage of variance explained by eigenvalue for each part of current time series from electrochemical noise process obtained from localized SSA.

The resulting reconstructed time series from both the localized and linear SSA are presented in Figure 5.20. From this figure it can be seen that, regarding the reconstruction of the original time series, the linear SSA technique fares slightly better than the localized SSA technique, with a correlation coefficient of 0.982 between the reconstructed time series from linear SSA and the original time series and a coefficient of 0.981 between that from localized SSA and the original time series. However, even though the difference between the two techniques is barely noticeable, it can be seen that localized (nonlinear) analysis of the time series enables the identification of components explaining larger amounts of variance than the linear analysis. This could lead to a more compact reconstruction of the data using fewer principal components.

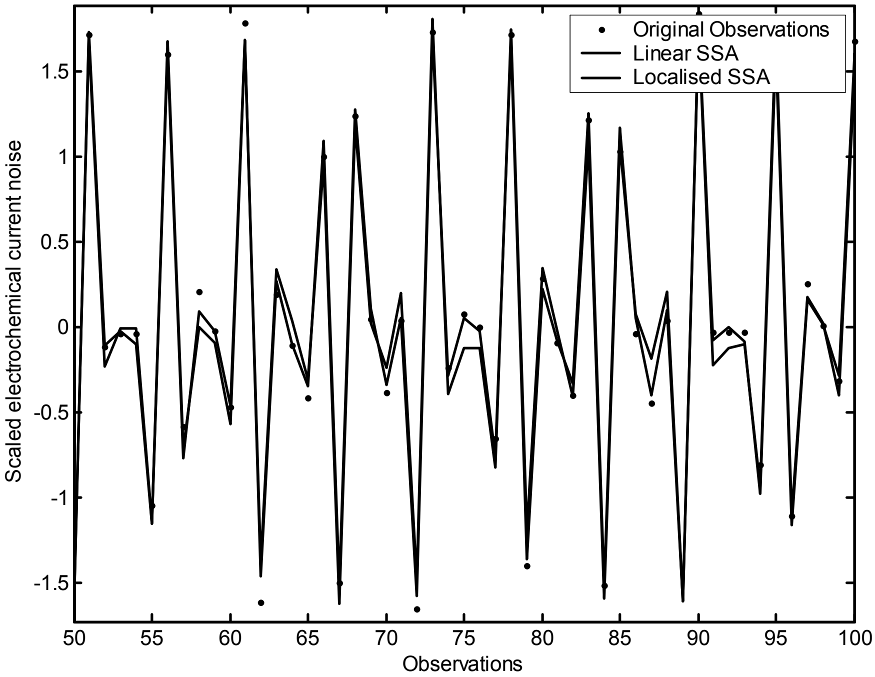


Figure 5.20 Reconstructed time series obtained by linear SSA (solid line) and localized SSA (dotted line) from original observations (dotted markers) of electrochemical current noise process.

If the two time series were to be reconstructed with only two principal components retained for each of the decompositions, it was found that the localized SSA reconstruction correlated with the original time series with a factor of 0.868, while the correlation of the linear SSA reconstruction with the original time series was only 0.853. This observation is in good agreement with the results from the eigenvalue spectra, which indicated that the first two components extracted by localized SSA were more important than the first two extracted by linear SSA.

Another part of the investigation studied the difference in the principal components extracted by using the two techniques. This was achieved by also embedding the parts of the localized SSA into their optimum embedding window, depending on the behaviour of the parts, but then to retain an equal number of principal components for each part of the time series. This was necessary to allow the $\mathbf{TP}^T$ matrices of the different parts to be joined again and to extract the principal components. In order to err on the side of caution, the number of components retained in each part was 16. The reconstructed components from both localized and linear SSA were then grouped, based on their behaviour observed in the respective eigenspectra (Figure 5.18 and Figure 5.19) and the resulting series are presented in Figure 5.21 and Figure 5.22. Due to the number of different eigenspectra obtained for the localized SSA, it was harder to determine the component groups, but a general trend that was observed from Figure 5.19 was that the eigenvalues tended to be divided into pairs.

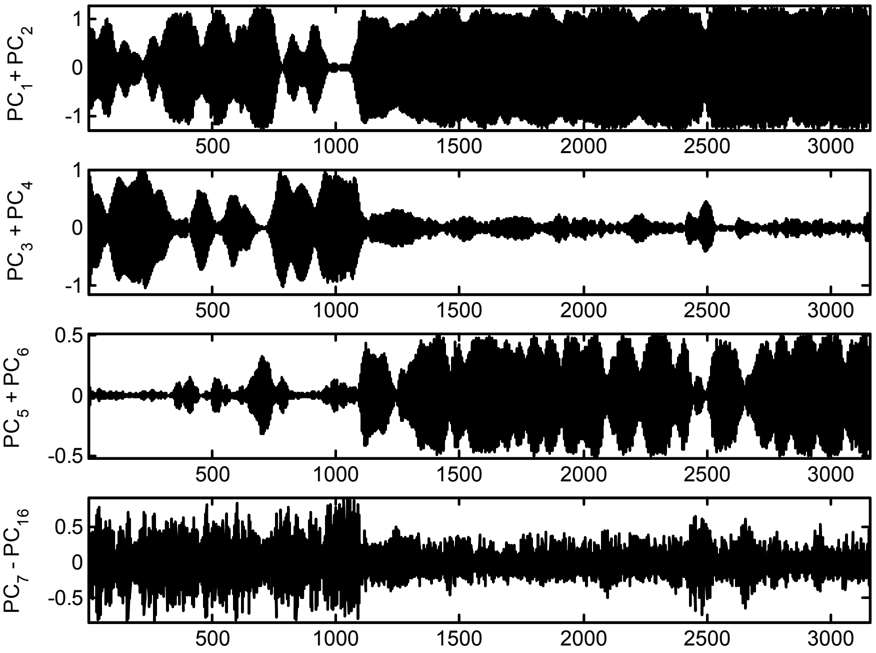


Figure 5.21 Individual principal component pairs obtained from linear singular spectrum analysis of current data from electrochemical noise process.

It can be seen that, as was expected from the difference in eigenvalue distributions, the first two reconstructed components obtained by localized SSA appeared closer to the original time series than the first two components from linear SSA. The periodicities of the different components are also clear from these two figures. It is interesting to note that the different components seem 'complimentary' in that during time intervals when the variance of one time series is relatively small, another would have a sudden interval of high variance.

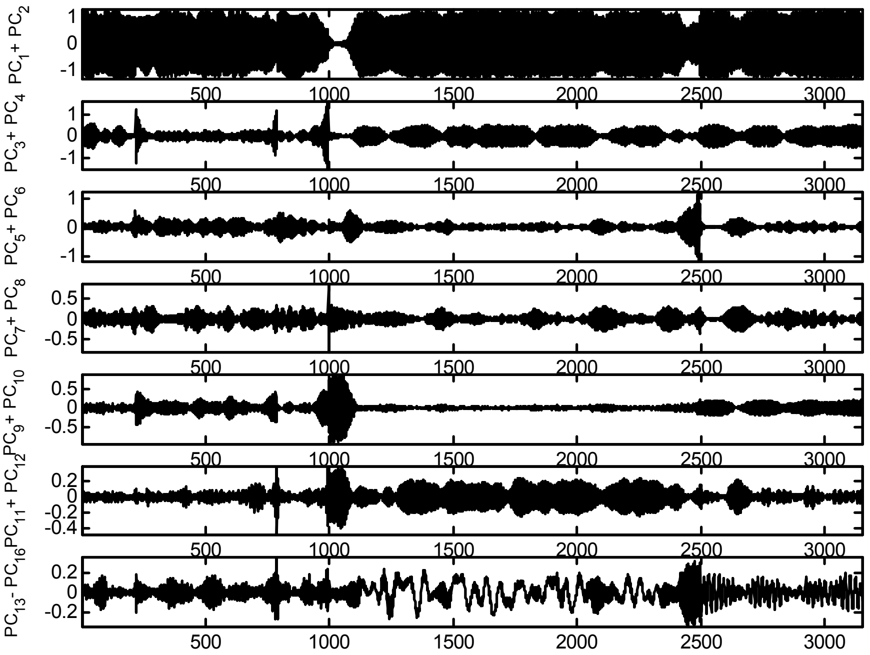


Figure 5.22 Individual principal component pairs obtained from localized singular spectrum analysis of current data from electrochemical noise process.

5.2.3 Auto-associative neural network analysis

a) Background

The second possible technique by which nonlinear time series can be analysed is to use auto-associative neural networks to perform the analysis, as has been described in section 3.4.2. This technique was also applied to the current series from the measured electrochemical noise process, extracting 16 tp^T-combinations.

As with the linear SSA, the time series was embedded in a 30-dimensional trajectory matrix. This matrix was then given as input to the auto-associative neural network in an effort to extract the first nonlinear component. After the first component has been extracted, the output from the network was subtracted from the input to obtain the residual values. These residual values were then supplied to the same network configuration to extract the second nonlinear component, and so forth until all sixteen nonlinear components were extracted. The hidden layers of the network had a [2 1 2] configuration, with the number of input and output nodes equal to the number of columns in the trajectory matrix, in other words, 30 each.

b) Results

The resulting nonlinear principal component combinations, similar to those for the linear SSA in Figure 5.21 and for the localized SSA in Figure 5.22, are supplied in Figure 5.23.

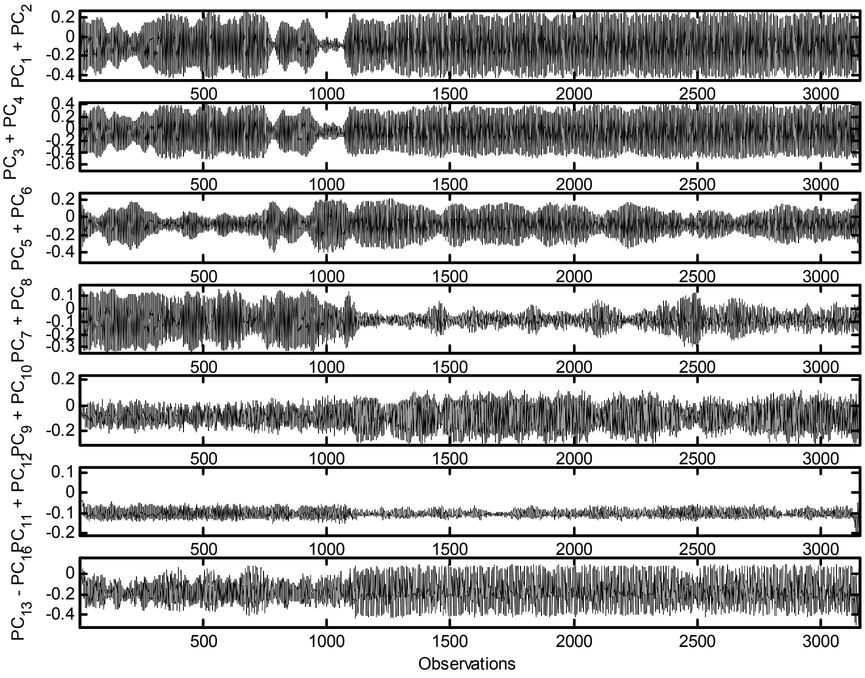


Figure 5.23 Individual nonlinear principal component pairs obtained from auto-associative neural network analysis of current data from electrochemical noise process.

It can be seen from the figure that the nonlinear components extracted from the time series differ slightly from the linear and localized ones, in that some of the later nonlinear components still explain relatively large amounts of variance. The variance in the nonlinear components themselves is also spread more evenly along the components, unlike the linear analysis where it seemed much more as if certain components were only active over certain parts of the time series (Figure 5.7).

Another point worth noting was the behaviour of the neural network in the training and extraction of the first nonlinear component, displayed in Figure 5.24. It can be seen that no useful information at all could be extracted from this output. The reason for this seemed to be instabilities in the network and the composition of the time series.

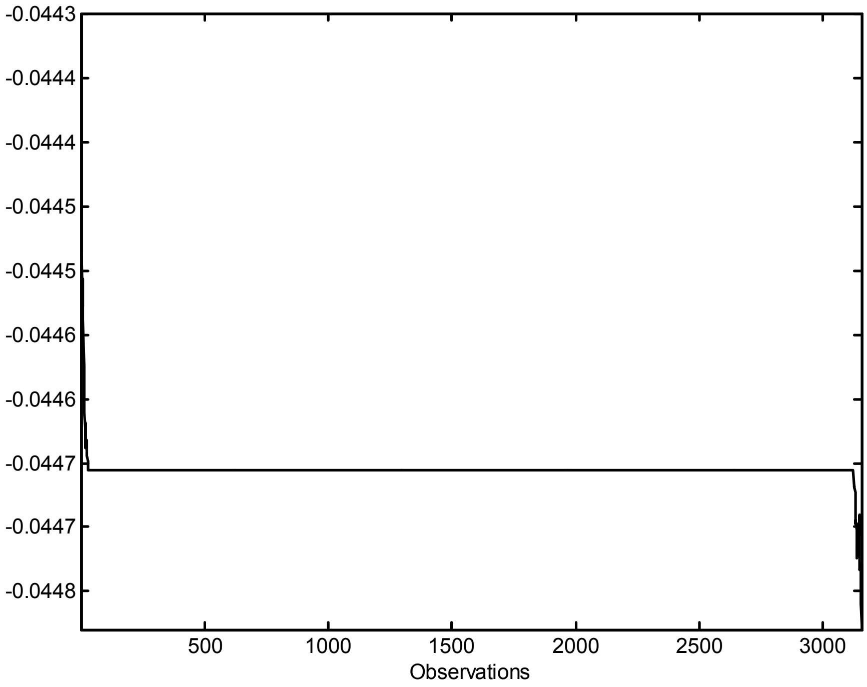


Figure 5.24 First nonlinear principal component extracted from the current series of the electrochemical noise data set by using an auto-associative neural network.

However, once the information that the network considered as the first nonlinear component was removed from the data, the residual values did not present any trouble at all in extracting the second nonlinear principal component (shown in Figure 5.25) and the subsequent nonlinear components (not shown individually).

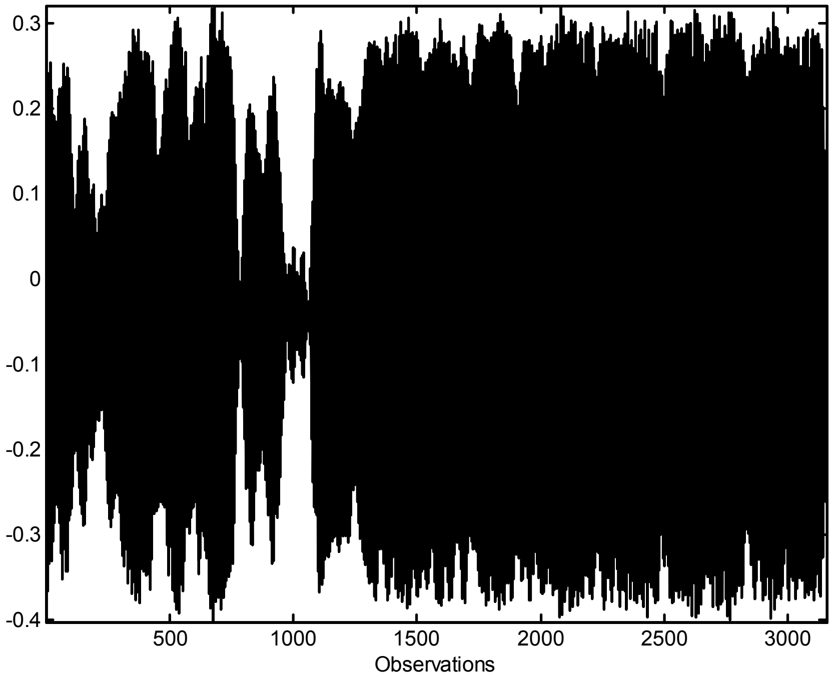


Figure 5.25 Second nonlinear principal component extracted from the current series of the electrochemical noise data set by using an auto-associative neural network.

5.3 Summary

In this chapter two different approaches to extract nonlinear information from the time series were investigated. The first method used was that of localized SSA, where the time series was divided into a number of different parts, each of which were analysed separately by SSA. The second approach was to use auto-associative neural networks.

It was found from both case studies, that a distinct difference between the eigenvaluespectra of the linear and localized SSA could be observed. In both cases the localized SSA resulted in a more defined distinction between valuable and noise components. The correlation of the reconstructed series from localized SSA with the original series also tended to be better than that of the reconstruction from linear SSA.

Unfortunately the auto-associative neural networks did not perform as well as one would have hoped, as the network had trouble training some of the time series. However, reliable results could be obtained for the electrochemical noise data and once again a marked difference was observed in the nature of the principal components that were extracted.

6 MONTE CARLO SSA

Monte Carlo singular spectrum analysis (MC-SSA) is a methodology for discriminating between various components of the time series, particularly components containing meaningful information and other components containing mostly noise. This problem is especially important in process engineering applications, such as modelling, control, data validation and filtering. In practice, time series data are often assumed to be linear and stochastic (ARMA models) or, conversely, assumed to be nonlinear, because the process is known to be nonlinear from first principles. In the latter case, this does not mean that the observed data would also be nonlinear by default. This incorrect assumption could lead to less than optimal modelling, systematic process errors during filtering of the data, etc 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, as is the case in autoregressive moving average processes. These stochastic processes have frequency spectra that decrease monotonically with frequency and are often referred to as warm-coloured.

There are a great number of tests in the literature by which to characterize the dynamics of processes by means of observed data. However, none of these tests are infallible. Monte Carlo SSA is another such test, and is no exception to the imperfect nature of other tests. Although this technique has been used extensively in the literature (Allen and Smith, 1996, Theiler and Prichard, 1996, Palus and Novotna, 1998), little has been done to assess the reliability of these tests and the associated characteristics.

By first applying Monte Carlo SSA in this chapter to a number of time series of which the characteristics are known beforehand, one can obtain an idea of the reliability of Monte Carlo SSA, where-after Monte Carlo SSA can be used in the next chapter to characterize real data series.

6.1 Artificial data sets with known properties

6.1.1 Properties of artificial time series

The applicability and the validity of the Monte Carlo approach to singular spectrum analysis are first investigated by using a series of artificial data sets. These data sets were generated in such a manner that their properties were all known beforehand, which means they could serve as benchmark series by which to evaluate the performance of Monte Carlo SSA. These time series were generated in such a fashion to represent all the types of data that are presented Figure 6.1 and can also be used for comparison purposes among the different classes of stationary data. Therefore, the resulting seven time series are LGX, BGX, TGX, LUX, BUX, TUX and NonLin. These time series were then compared with an first order autoregressive (AR(1)) series, to see if Monte Carlo singular spectrum analysis could differentiate between the different classes.

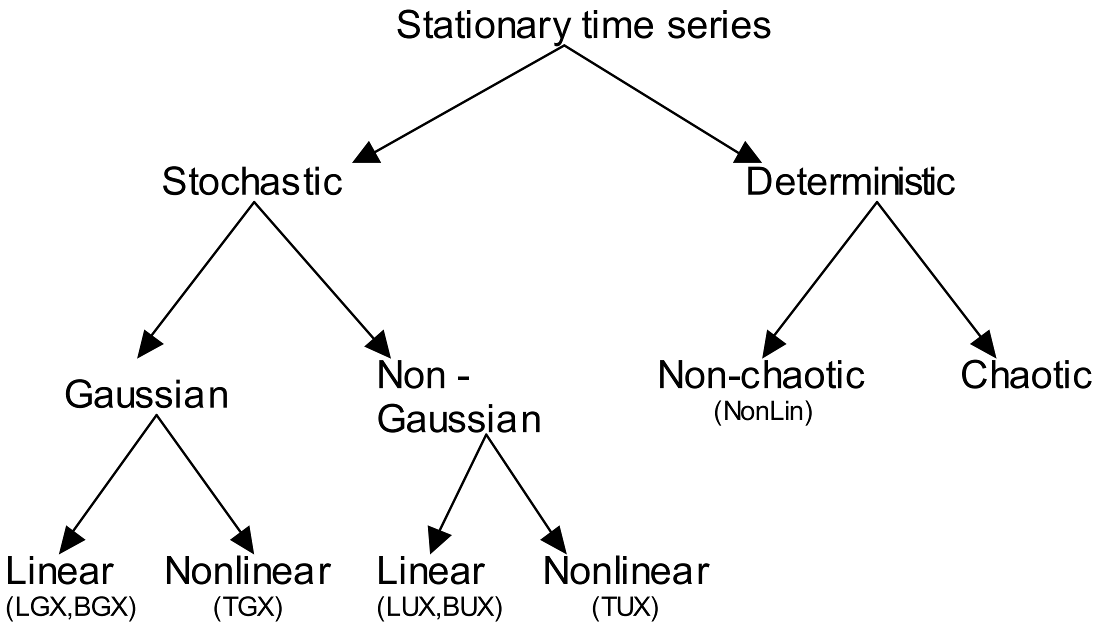


Figure 6.1 Classification of stationary time series into different classes.

Table 6.1 provides a summary of the characteristics of each of the artificial time series and the appropriate classification of each time series has also been indicated on Figure 6.1.

Table 6.1 Description of data characteristics of artificial time series used for benchmarking of Monte Carlo SSA

Name of Series	Characteristics
AR(1)	First order autoregressive time series
LGX	Linear Gaussian time series
BGX	Bilinear Gaussian time series
TGX	Nonlinear Gaussian time series of the threshold autoregressive type (TAR)
LUX	Linear time series with uniform noise
BUX	Bilinear time series with uniform noise
TUX	Nonlinear time series with uniform noise of the threshold autoregressive type (TAR)
NonLin	Nonchaotic nonlinear system

The time series investigated (see Figure 6.2) were all generated artificially by using the following equations:

a) AR(1)

$$x_t = 0.92x_{t-1} + \varepsilon_t \tag{6.1}$$

ε_t is a Gaussian randomly generated noise series with zero mean and a standard deviation of 0.15.

b) LGX

$$x_t = 0.12x_{t-1} + 0.08x_{t-2} - 0.2x_{t-3} + \varepsilon_t + 0.15\varepsilon_{t-1} + 0.36\varepsilon_{t-2} \tag{6.2}$$

ε_t is a Gaussian randomly generated noise series with zero mean and a standard deviation of 0.15.

c) BGX

$$x_t = 0.12x_{t-1} + 0.08x_{t-2} - 0.2x_{t-3} + \varepsilon_t + 0.15\varepsilon_{t-1} + 0.36\varepsilon_{t-2} + 0.04x_{t-1} \times \varepsilon_{t-1}$$

$$+ 0.16x_{t-1} \times \varepsilon_{t-2} - 0.35x_{t-2} \times \varepsilon_{t-2} \qquad 6.3$$

ε_t is a Gaussian randomly generated noise series with zero mean and a standard deviation of 0.15. This bilinear time series is a mildly nonlinear version of the linear time series, LGX. The nonlinearity is introduced by the bilinear terms at the end of equation 6.3, where the time series observations and the error terms are multiplied with each other. However, for this time series, the coefficients of the bilinear terms are relatively small, compared to that of the linear terms. This leads to the expectation that the LGX and BGX time series will behave similarly.

d) TGX

$$x_t = 0.1x_{t-1} + \varepsilon_t \text{ if } x_{t-1} < 0.5 \qquad 6.4$$
$$x_t = 0.9x_{t-1} + \varepsilon_t \text{ if } x_{t-1} \geq 0.5$$

ε_t is a Gaussian randomly generated noise series with zero mean and a standard deviation of 0.15. Where the bilinear time series is mildly nonlinear, this time series has a strongly nonlinear character.

e) LUX

$$x_t = 0.12x_{t-1} + 0.08x_{t-2} - 0.2x_{t-3} + \mathrm{u}_t + 0.15\mathrm{u}_{t-1} + 0.36\mathrm{u}_{t-2} \qquad 6.5$$

u_t is a randomly generated uniform noise series with $–0.5 \leq u_t \leq 0.5$.

f) BUX

$$x_t = 0.12x_{t-1} + 0.08x_{t-2} - 0.2x_{t-3} + \mathrm{u}_t + 0.15\mathrm{u}_{t-1} + 0.36\mathrm{u}_{t-2} + 0.04x_{t-1} \times \mathrm{u}_{t-1}$$
$$+ 0.16x_{t-1} \times \mathrm{u}_{t-2} - 0.35x_{t-2} \times \mathrm{u}_{t-2} \qquad 6.6$$

u_t is a randomly generated uniform noise series with $–0.5 \leq u_t \leq 0.5$. The comparison between the linear and bilinear series with a Gaussian distribution also holds for the linear and bilinear series with uniform distributions.

g) TUX

$$x_t = 0.1x_{t-1} + \mathrm{u}_t \text{ if } x_{t-1} < 0.5 \qquad 6.7$$
$$x_t = 0.9x_{t-1} + \mathrm{u}_t \text{ if } x_{t-1} \geq 0.5$$

u_t is a randomly generated uniform noise series with $–0.5 \leq u_t \leq 0.5$.

h) NonLin

$$x_t = \sin(t) + \cos(\tfrac{t}{2}) - \sin(\tfrac{t}{4}) + \sin(\tfrac{t}{8}) \qquad 6.8$$

with *t* measured in radians.
It will be noted that none of the time series in this section can be classified as being purely deterministic and chaotic. This class of data will be discussed separately in a later section in this chapter (section 6.3).

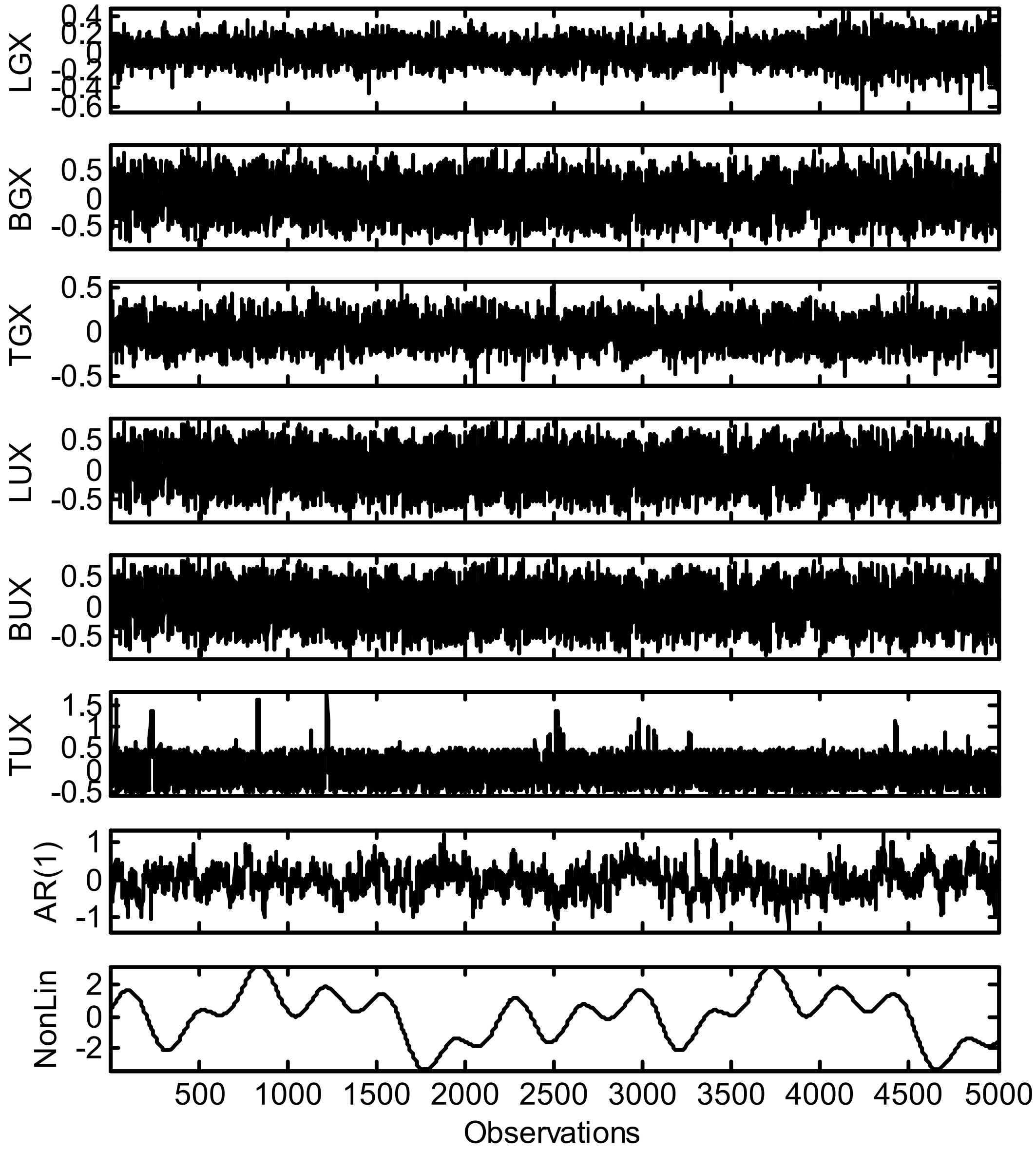


Figure 6.2 Illustration of the artificial time series used for benchmarking of the Monte Carlo process.

15 surrogate data sets were generated for each time series, by using the amplitude adjusted Fourier transform algorithm. These data sets, along with the original time series, were tested against the hypothesis of being linear, Gaussian, stochastic time series, both by using the eigenspectra and the correlation dimensions as test statistics.

6.1.2 Characterization of time series

Figure 6.3 illustrates the histograms of the frequency distributions of the eight time series under consideration in this section. It can be seen that the distribution of the three Gaussian time series (LGX, BGX and TGX), fits the normal distribution curve of the data much closer than the other time series, except for the AR(1) process, which also seems to have a normal distribution. The largest deviations from the normal curve can be seen, as would be expected, from the nonlinear time series. This series seem not to follow the normal distribution curve at all.

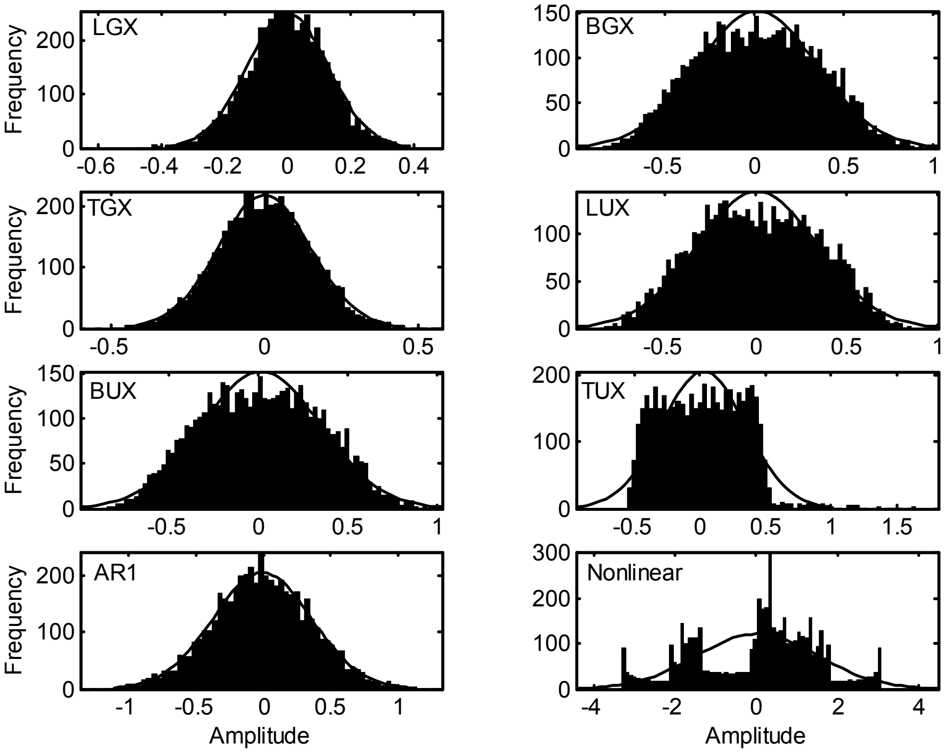


Figure 6.3 Frequency distribution of observations in original time series with various characteristics.

A similar figure (Figure 6.4) has been constructed for three of the surrogate data sets that have been generated for each of the time series. One can see from this figure that, even though the individual surrogates vary slightly, the general frequency distribution of the surrogates is very similar to that of the various original time series. This would be expected, as one of the characteristics of the AAFT algorithm is that the amplitude distributions of the surrogates will be the same as that of the original series. Although it is not shown here, the frequency distribution from surrogate data generated by the IAAFT algorithm did not differ at all from those generated by the AAFT algorithm (shown in Figure 6.4), illustrating the validity of using either one of the techniques to generate the surrogate data.

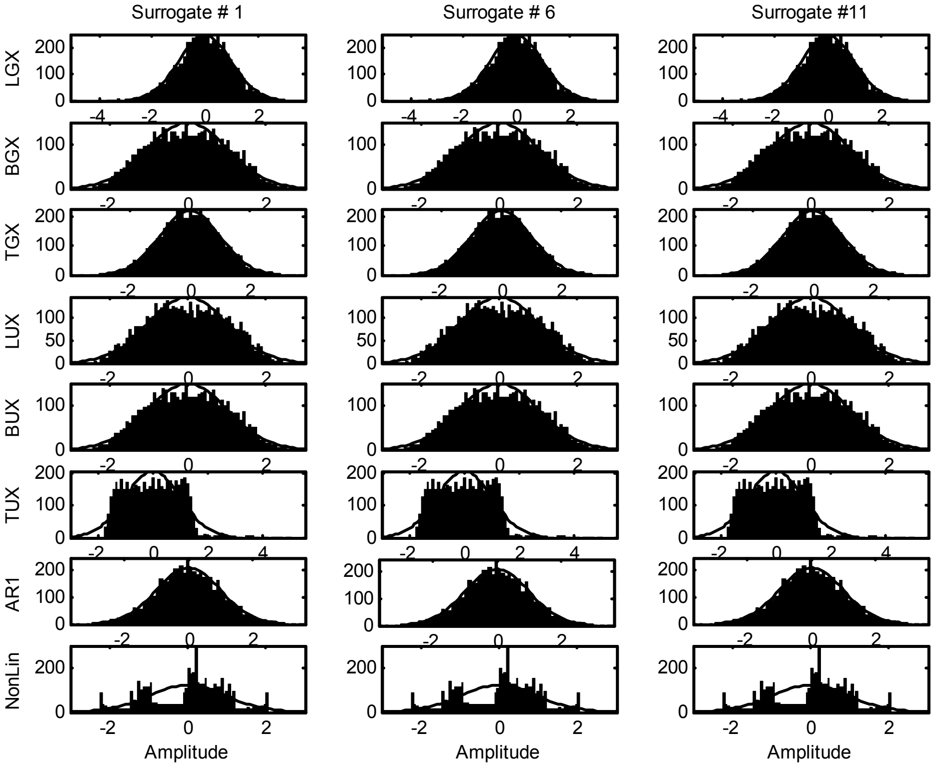


Figure 6.4 Frequency distributions of observations from three separate surrogate data sets for each artificial time series by using the AAFT algorithm to generate the surrogate data sets.

The eigenvalue spectra, along with the respective confidence limits, for each of the time series are shown in Figure 6.5 to Figure 6.27. Two sets of confidence limits (shown in two separate figures for each time series) have been generated. The one tests for a linear, stochastic, Gaussian series and the other for a first order autoregressive series. The surrogates adhering to the linear, stochastic, Gaussian time series characteristics were generated by using the AAFT algorithm. The surrogates used to test for first order autoregressive noise were obtained by using an algorithm developed by Allen and Smith (1997), which has been discussed in an earlier section (section 4.7.3).

The correlation dimension curves of the original time series, together with that of fifteen surrogate data sets that were generated by the AAFT algorithm, were also used for characterisation purposes. The relevance and calculation of the correlation dimension have been discussed in section 3.3.4, but in short the correlation dimension can be seen as an indication of the number of points of an embedded object (the attractor) within a certain radius (e) of a point.

When the results from the LGX and the BGX time series that are displayed in Figure 6.5 to Figure 6.10 are compared, it can be seen that, especially in the behaviour of the eigenspectra towards their respective confidence limits, there is not much difference between the two time series. This would be expected, as the bilinear Gaussian time series actually exhibit characteristics very similar to these of the linear Gaussian time series. However, there is still a marked difference between the correlation dimension behaviour of the two time series. The correlation dimension from the bilinear time series (Figure 6.10) behaved significantly more like that of the surrogate series than the correlation dimension of the linear time series (Figure 6.7).

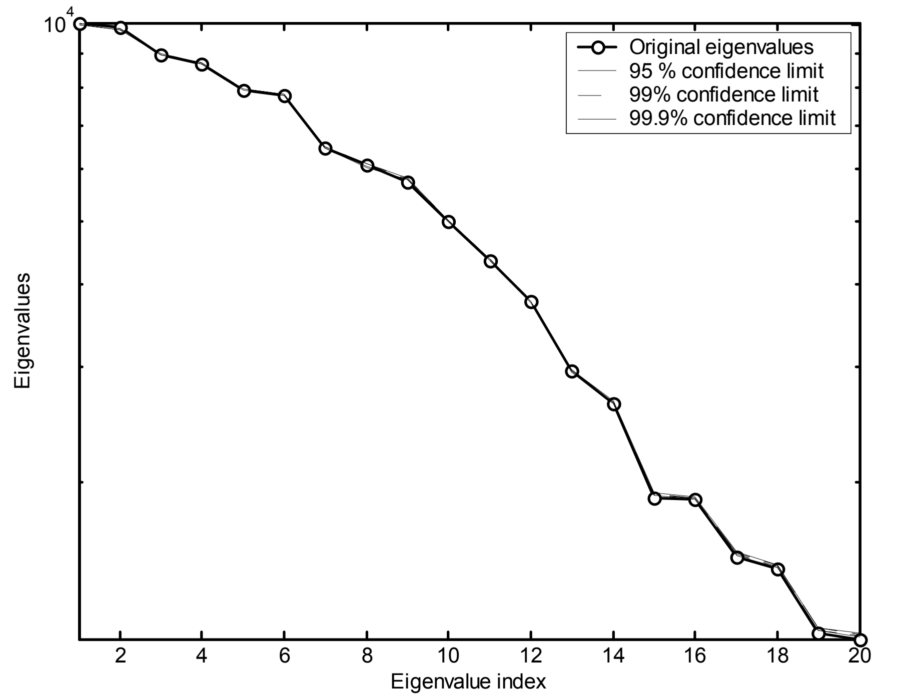


Figure 6.5 Eigenspectra of LGX time series along with confidence bands calculated from surrogates generated by the AAFT algorithm.

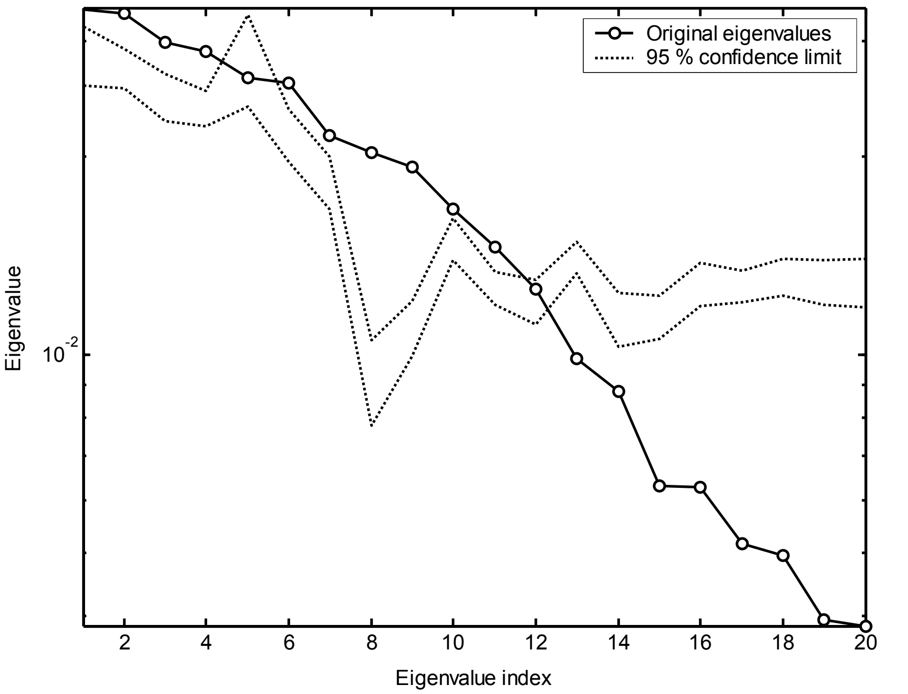


Figure 6.6 Eigenspectra and confidence bands testing if the LGX time series has properties similar to an AR(1) process.

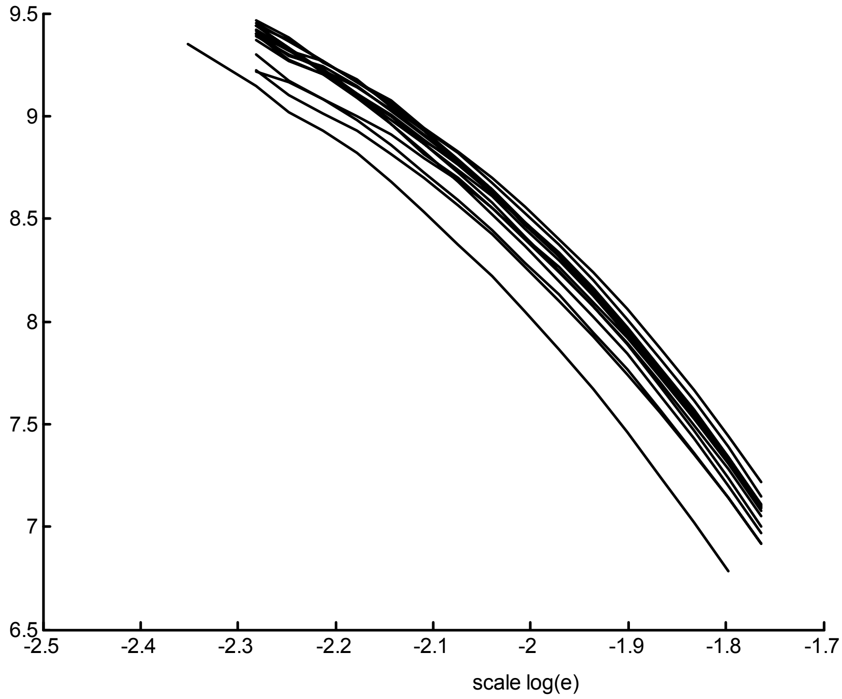


Figure 6.7 Correlation dimension of the LGX time series (solid line) along with the correlation dimensions of 15 surrogate data sets (broken lines) generated by using the AAFT algorithm.

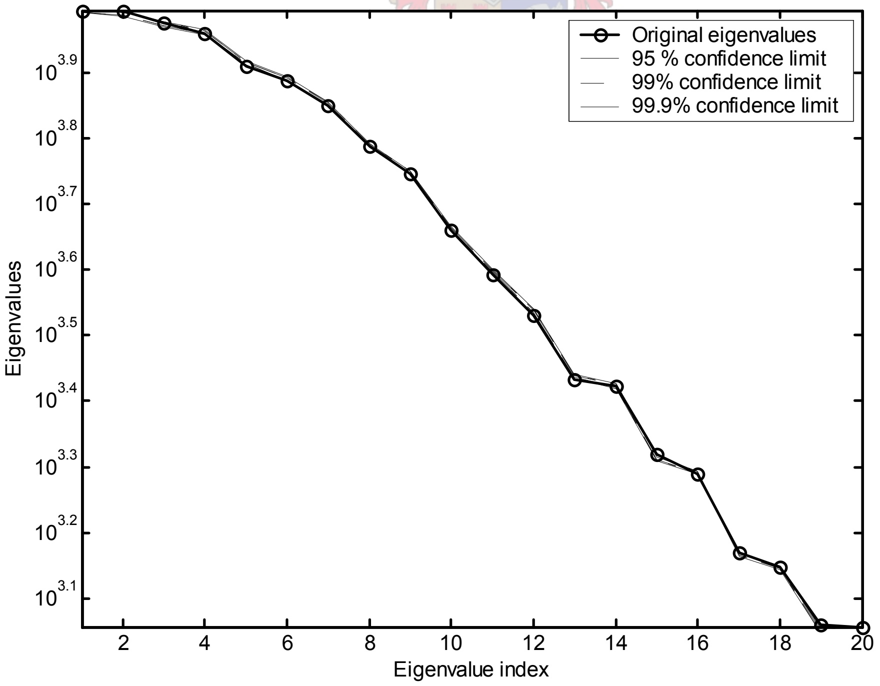


Figure 6.8 Eigenspectra of BGX time series along with confidence bands calculated from surrogates generated by the AAFT algorithm.

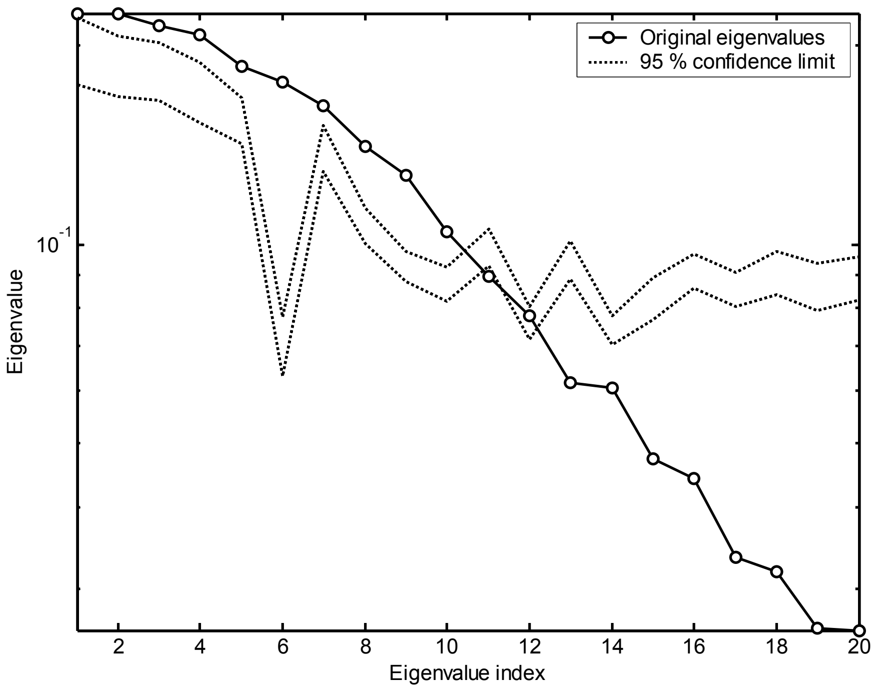


Figure 6.9 Eigenspectra and confidence bands testing if the BGX time series has similar properties than an AR(1) process.

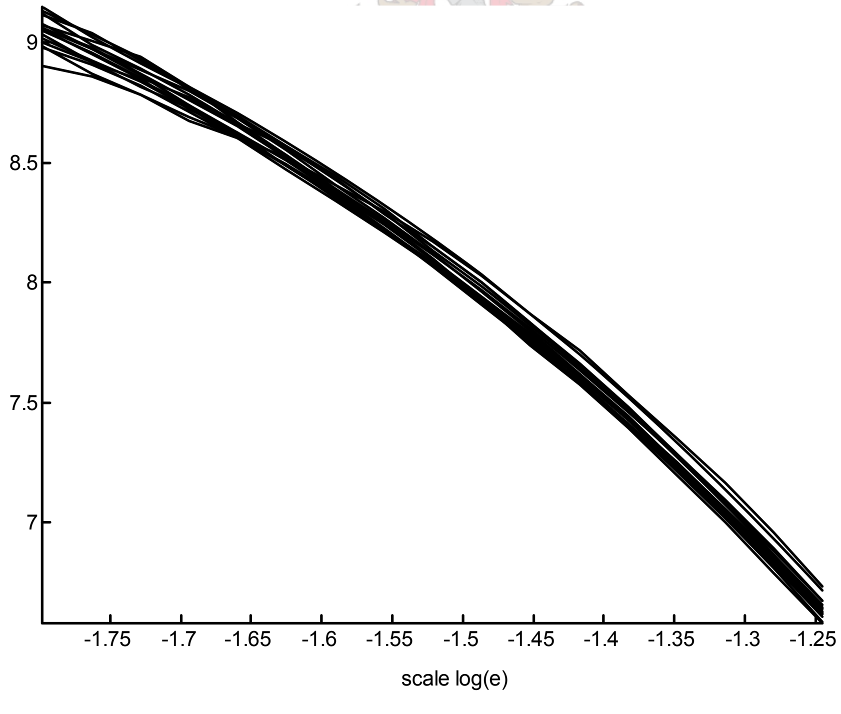


Figure 6.10 Correlation dimension of the BGX time series (solid line) along with the correlation dimensions of 15 surrogate data sets (broken lines) generated by using the AAFT algorithm.

The confidence limits generated by the surrogates from the AAFT algorithm are relatively narrow, making it hard to determine at first glance if the eigenspectrum falls inside or outside the confidence limits (Figure 6.5 and Figure 6.8). If the figures are enlarged (not shown here), it can be seen that for both the LGX and the BGX time series the distribution of the eigenvalues is about 50% inside and 50% outside the confidence limits for the linear, stochastic Gaussian series. This is not quite as would be expected, as the LGX series was generated to adhere to all these characteristics and the small coefficients of the bilinear terms in equation 6.3 would suggest that the BGX series would also behave very similar to the LGX series. This illustrates the known fact that no characterisation technique produces reliable results all the time.

A much more conclusive result could be obtained from the confidence limits that were generated for the eigenspectra of first order autoregressive time series (Figure 6.6 and Figure 6.9). It can be seen that the eigenspectra of both the LGX and the BGX series are largely outside the confidence bands, concurring with the known fact that the series are not first order autoregressive series.

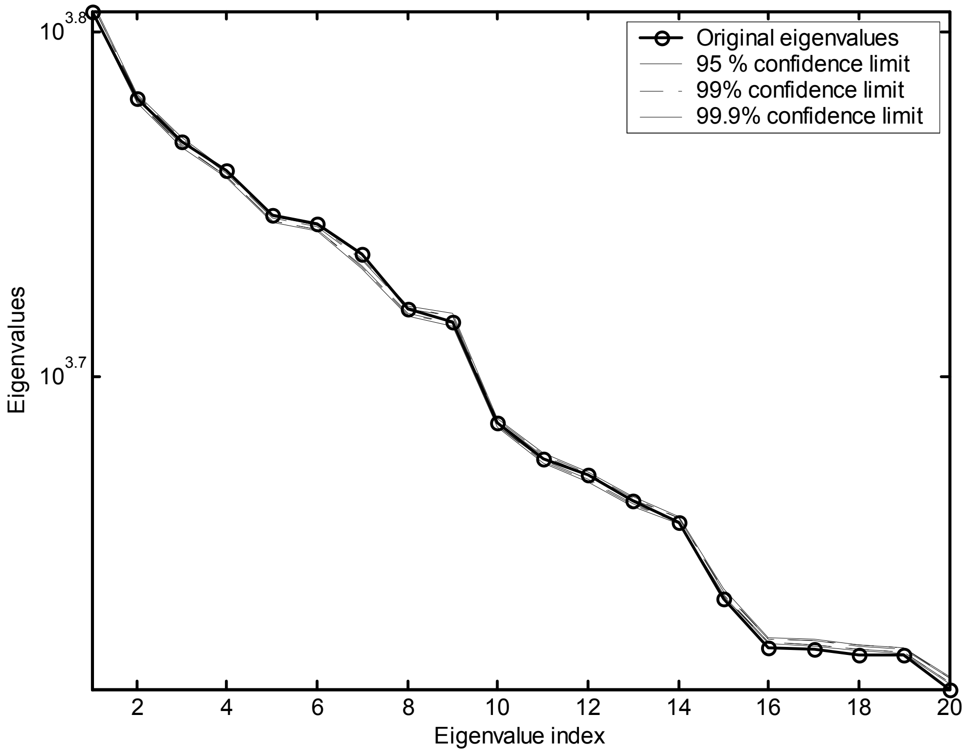


Figure 6.11 Eigenspectra of TGX time series along with confidence bands calculated from surrogates generated by the AAFT algorithm.

Once again in Figure 6.11 it is found that an almost equal amount of eigenvalues can be found inside and outside the confidence bands. However, this is in agreement with the expected results, as the TGX time series has nonlinear properties and therefore does not correspond with the null hypothesis of a linear, stochastic, Gaussian time series. This agreement, however, is contradicted by the correlation dimension, as the correlation dimension curve of the TGX series is indistinguishable from that of the surrogates (Figure 6.13).

Another interesting observation is that a large number of the eigenvalues was inside the confidence bands for a first order autoregressive model (Figure 6.12). This could also be attributed to the inability of any one technique to reliably characterize all time series.

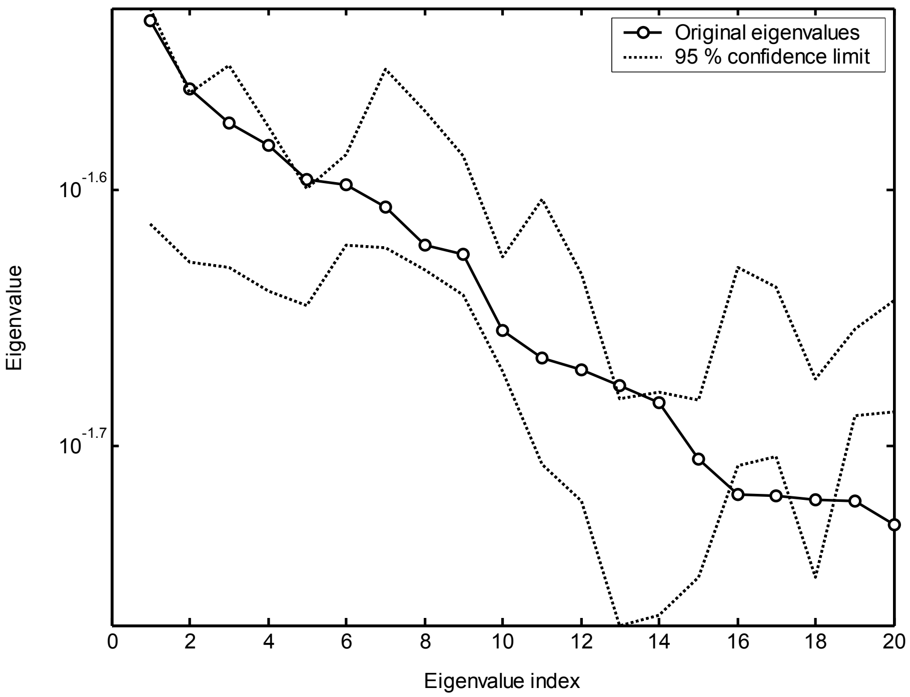


Figure 6.12 Eigenspectra and confidence bands testing if the TGX time series has similar properties than an AR(1) process.

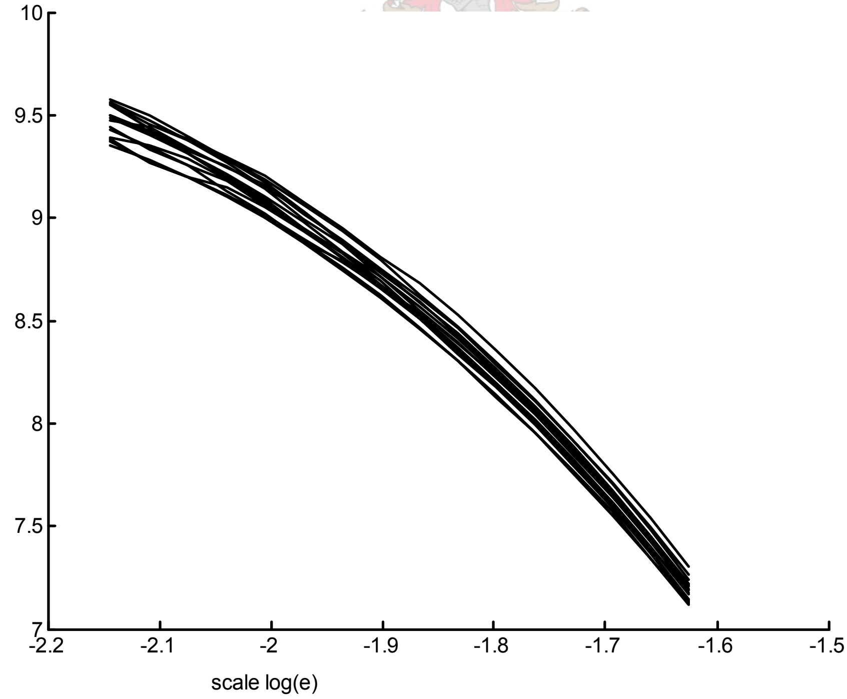


Figure 6.13 Correlation dimension of the TGX time series (solid line) along with the correlation dimensions of 15 surrogate data sets (broken lines) generated by using the AAFT algorithm.

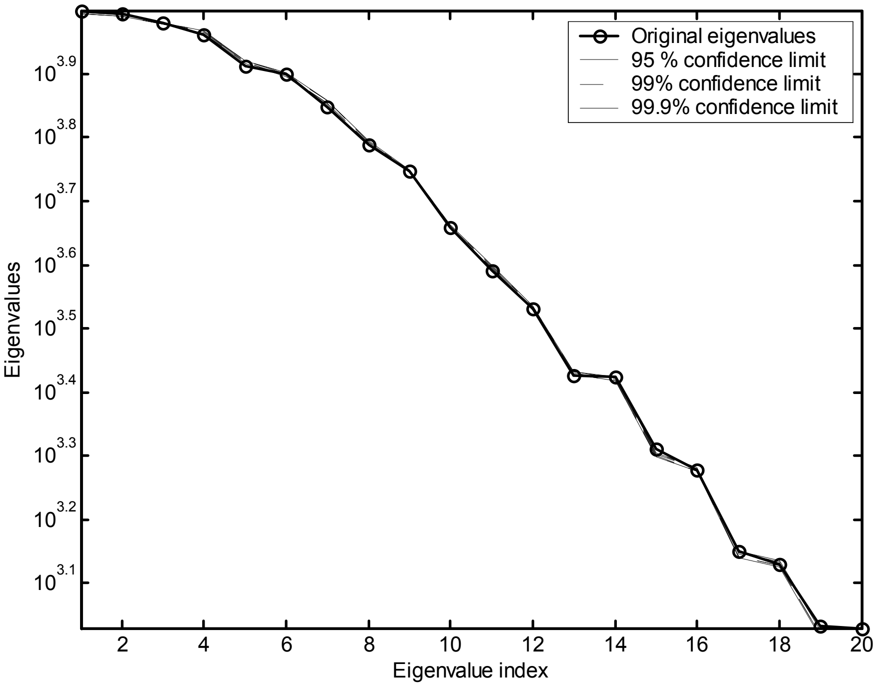


Figure 6.14 Eigenspectra of LUX time series along with confidence bands calculated from surrogates generated by the AAFT algorithm.

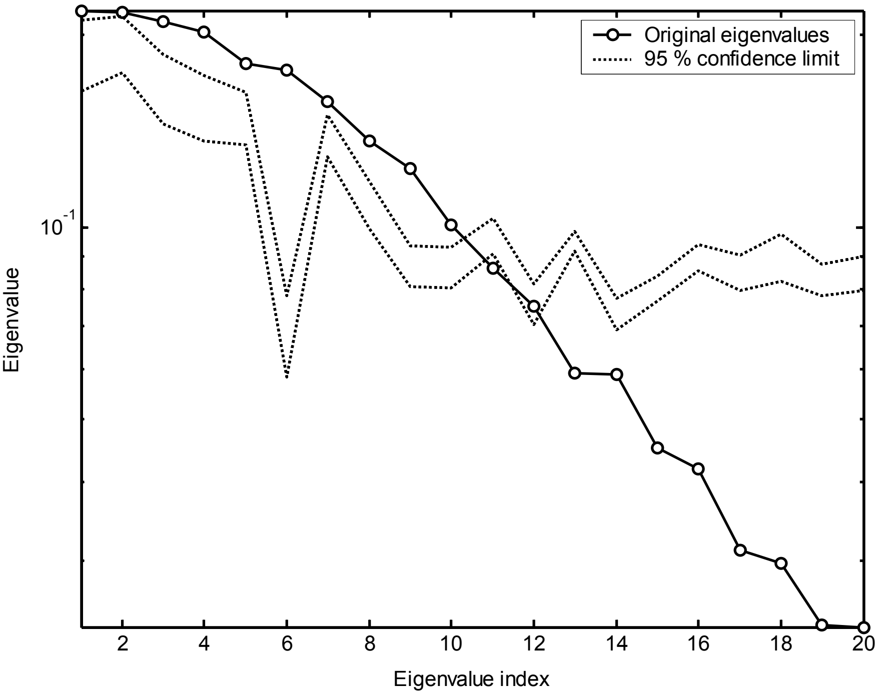


Figure 6.15 Eigenspectra and confidence bands testing if the LUX time series has similar properties than an AR(1) process.

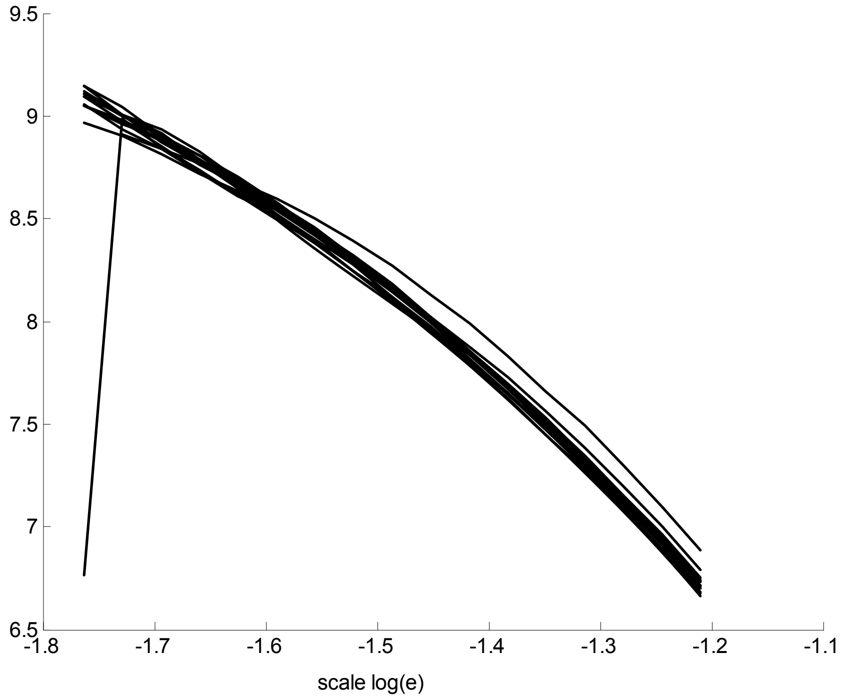


Figure 6.16 Correlation dimension of the LUX time series (solid line) along with the correlation dimensions of 15 surrogate data sets (broken lines) generated by using the AAFT algorithm.

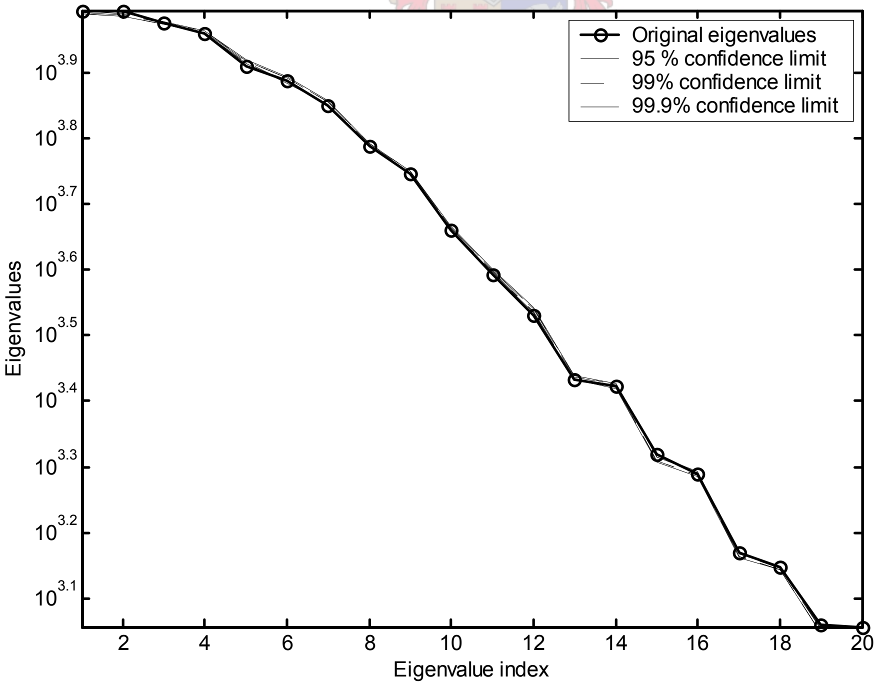


Figure 6.17 Eigenspectra of BUX time series along with confidence bands calculated from surrogates generated by the AAFT algorithm.

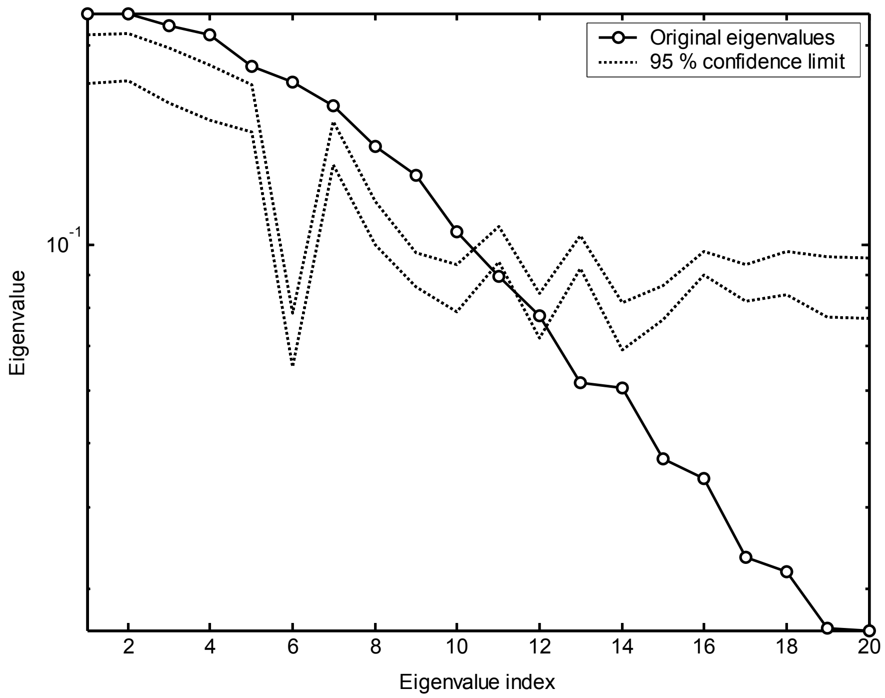


Figure 6.18 Eigenspectra and confidence bands testing if the BUX time series has similar properties than an AR(1) process.

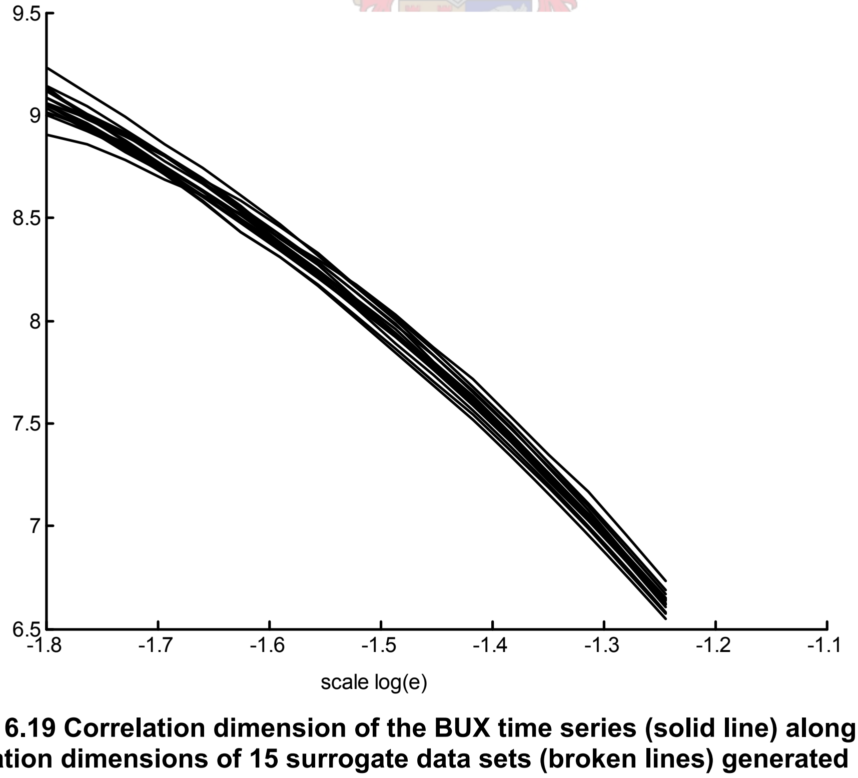


Figure 6.19 Correlation dimension of the BUX time series (solid line) along with the correlation dimensions of 15 surrogate data sets (broken lines) generated by using the AAFT algorithm.

As was the case for the two linear Gaussian series, LGX and BGX, there is not a large difference between the results for the two linear series with uniform noise distribution, LUX and BUX. Even though the eigenvalues still fall both inside and outside the confidence bands

(Figure 6.14 and Figure 6.17), a larger number of the eigenvalues could be found outside the confidence limits for these two Non-Gaussian time series. This can also be seen from the correlation dimension curves, in that, even though they are still relatively close to the curves of the surrogate series, the curves from the LUX and BUX series could be distinguished from those of their respective surrogates (Figure 6.16 and Figure 6.19).

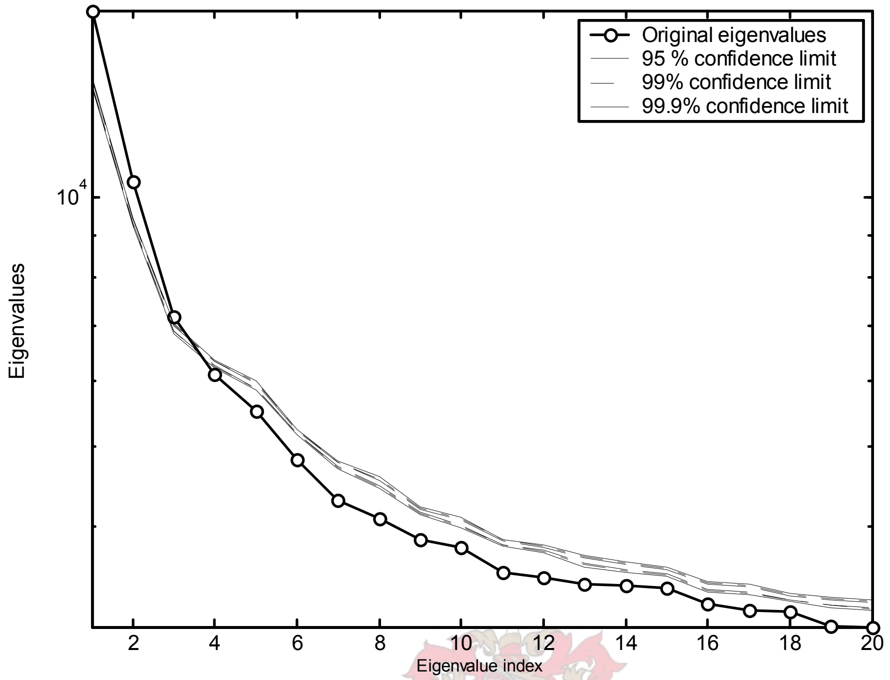


Figure 6.20 Eigenspectra of TUX time series along with confidence bands calculated from surrogates generated by the AAFT algorithm.

One of the sharper deviations from the confidence limits of the eigenspectra can be seen for the TUX time series in Figure 6.20. In light of the knowledge that this time series is neither linear, nor normally distributed, one can expect the null hypothesis of a stochastic, linear, Gaussian time series to be strongly rejected. This is then also the case for the eigenspectrum where none of the eigenvalues is bounded by the confidence bands. Unfortunately, this is not reflected in the surrogate analysis of the correlation dimension of the series (Figure 6.22), as that of the original time series lies very clearly among that of the surrogate series.

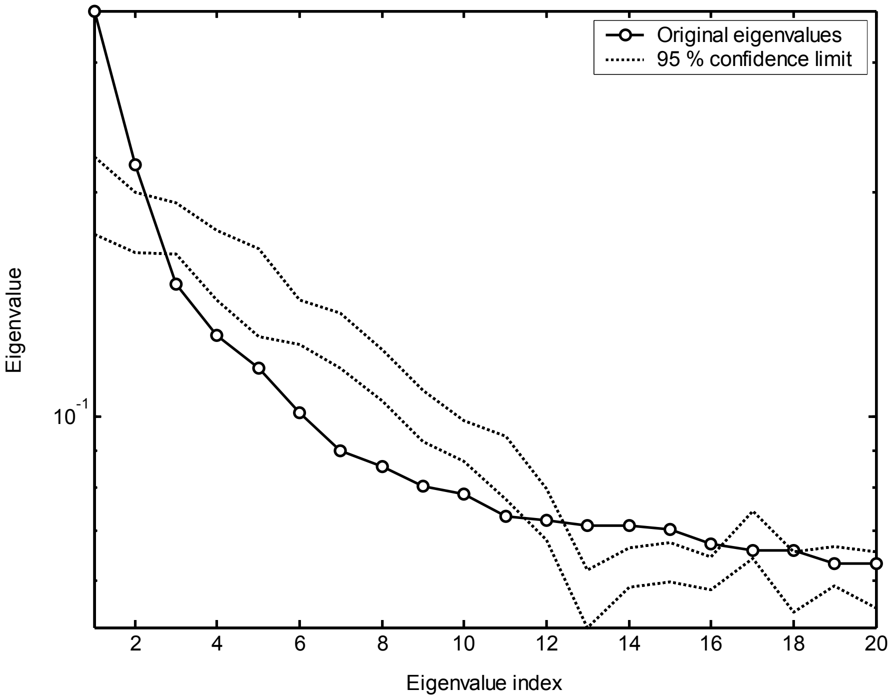


Figure 6.21 Eigenspectra and confidence bands testing if the TUX time series has similar properties than an AR(1) process.

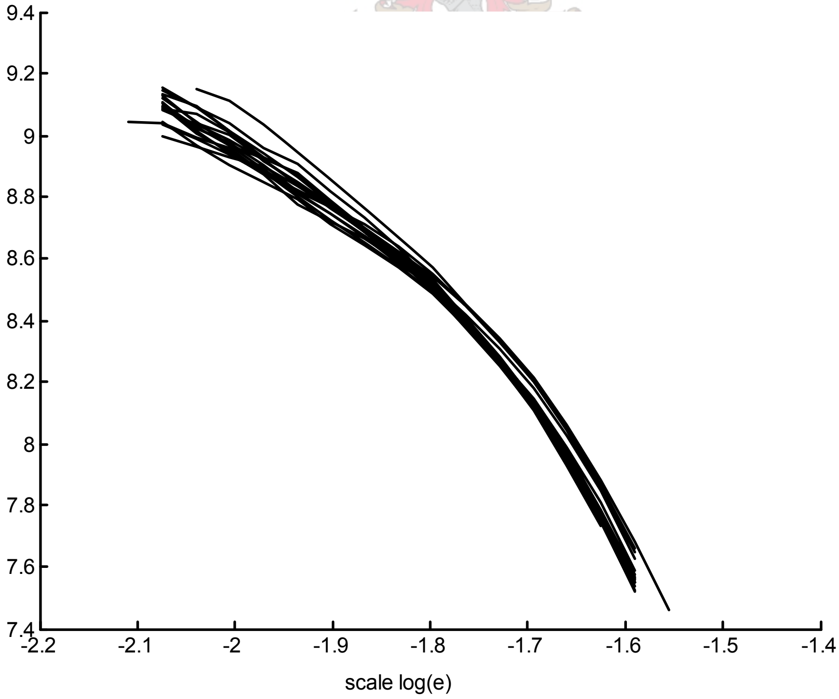


Figure 6.22 Correlation dimension of the TUX time series (solid line) along with the correlation dimensions of 15 surrogate data sets (broken lines) generated by using the AAFT algorithm.

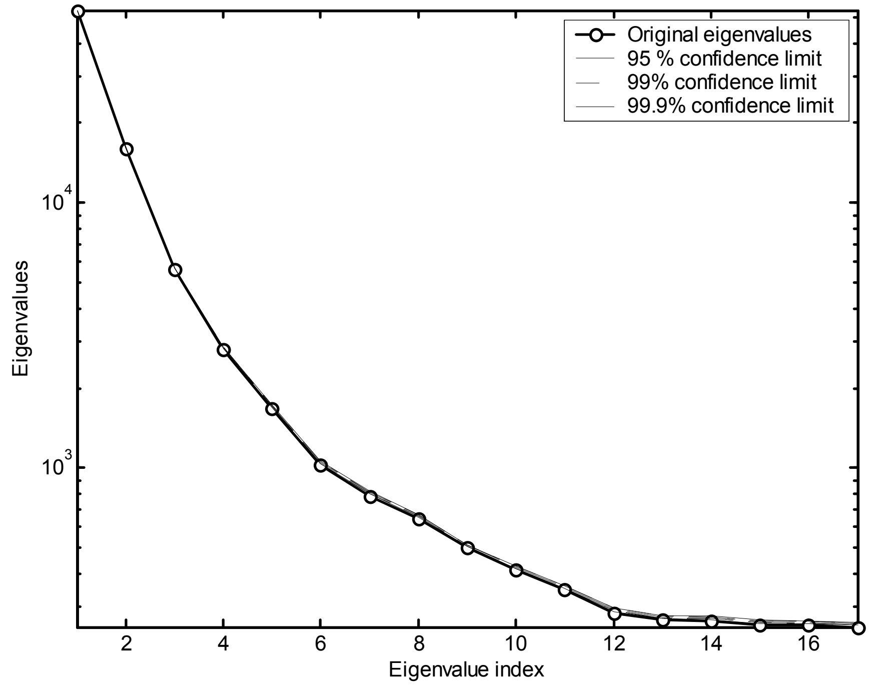


Figure 6.23 Eigenspectra of AR(1) time series along with confidence bands calculated from surrogates generated by the AAFT algorithm.

When the position of the eigenspectrum of the simulated first order autoregressive series relative to the two sets of confidence limits is inspected, the difference in the two test statistics become clear. Upon closer investigation, it was found that only one of the seventeen eigenvalues was bounded by the confidence limits for a linear, stochastic, Gaussian process (Figure 6.23). However, all of the eigenvalues could be found within the confidence bands for a first order autoregressive process (Figure 6.24), as would rightly be expected.

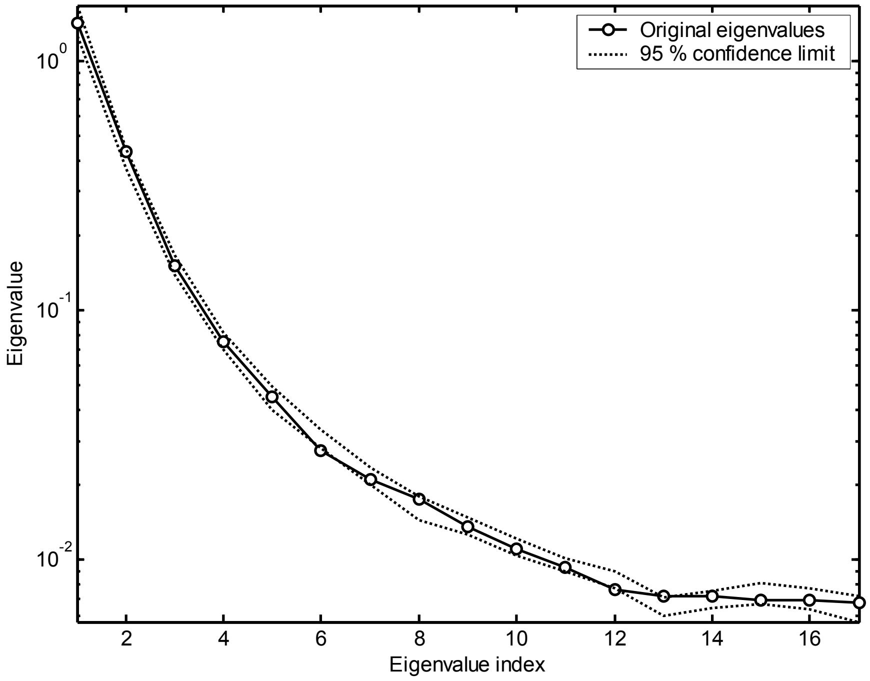


Figure 6.24 Eigenspectra and confidence bands confirming that the generated AR(1) time series has similar properties than the surrogate AR(1) series.

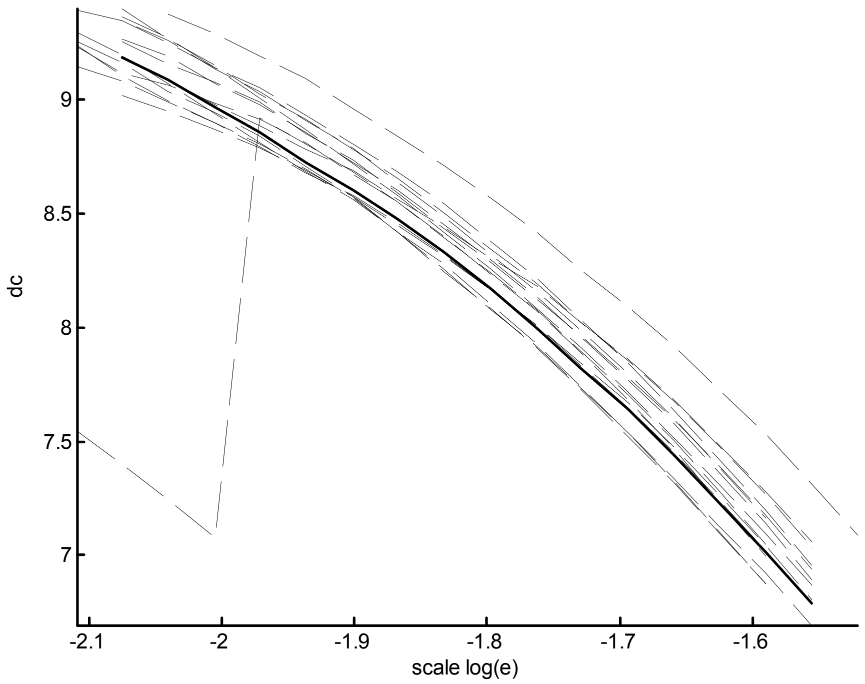


Figure 6.25 Correlation dimension of the AR(1) time series (solid line) along with the correlation dimensions of 15 surrogate data sets (broken lines) generated by using the AAFT algorithm.

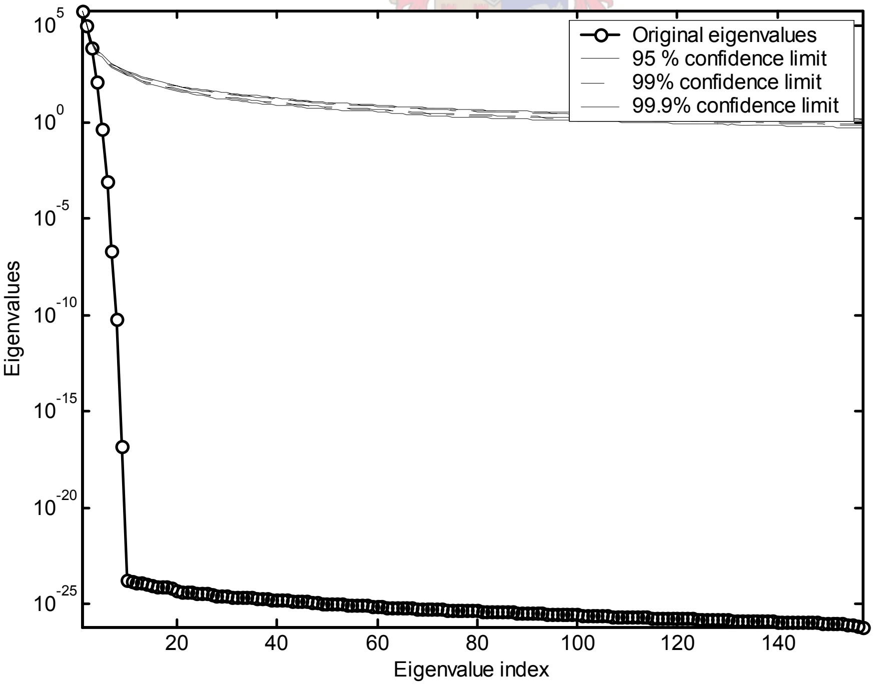


Figure 6.26 Eigenspectra of NonLin time series along with confidence bands calculated from surrogates generated by the AAFT algorithm.

The eigenspectrum of the NonLin time series, shown in Figure 6.26 displays an even more marked deviation from the confidence limits than that of the nonlinear, non-Gaussian series (TUX). This can be attributed to the truly nonlinear and deterministic nature of this time series. For this extreme case in terms of data characteristics, it can also be seen that the results from the surrogate analysis of the correlation dimension (Figure 6.27) correspond with that from the eigenspectrum, in that it indicates a clear difference between the correlation dimension of the linear, stochastic, Gaussian surrogate sets and that of the time series. The confidence limits of the first order autoregressive model could not be calculated for the NonLin time series, as a problem in the software produced negative (and therefore nonsensical) eigenvalues.

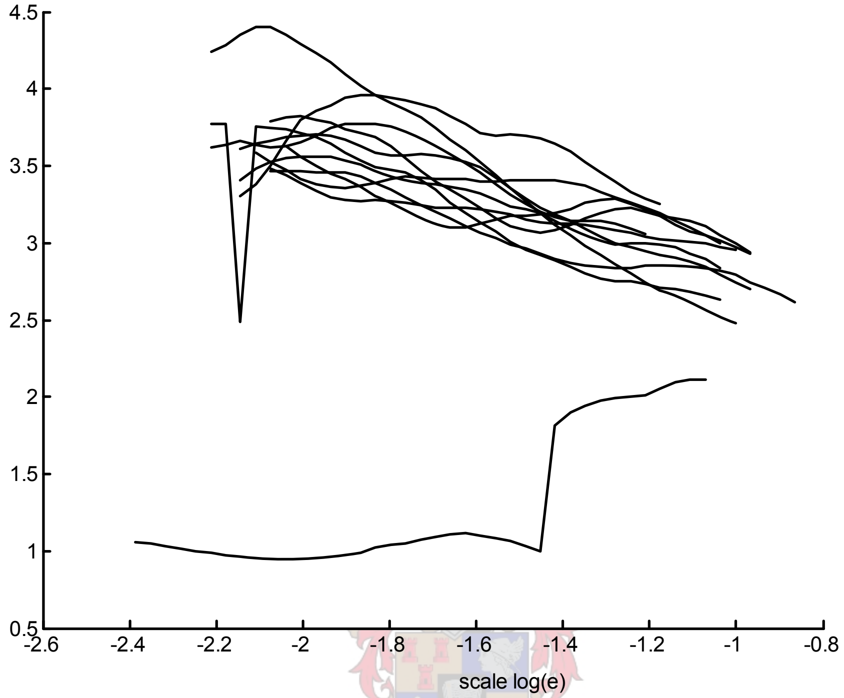


Figure 6.27 Correlation dimension of the NonLin time series (solid line) along with the correlation dimensions of 15 surrogate data sets (broken lines) generated by using the AAFT algorithm.

The observations and results from a number of the above case studies once again illustrated that no one test can provide a reliable characterisation for all time series all the time. This was apparent from the correct identification of some series by one test but a failure to obtain a conclusion from another test for the same series. One should therefore rather use a number of tests in combination, as was done in this section and is proposed by Barnett et al. (1997).

6.2 Flow in a series of tanks

6.2.1 Background

In order to further investigate the accuracy of Monte Carlo SSA in characterising time series, the response of the flow of four tanks in series, considered in section 4.2 on page 75, is once again investigated. The overall transfer function is once again given as

$$G(s) \equiv \frac{1}{(0.38s+1)(2s+1)(2.62s+1)(0.1s+1)} \qquad 6.9$$

Figure 6.28 shows the actual (solid line) and simulated measured (+) response to a pulsed input (broken line). The measured response was simulated by adding zero mean Gaussian noise with a standard deviation of 0.1 to the actual response. The trajectory matrix derived from the time series data consisted of 22 columns, each copy delayed by a time step of one. This matrix formed the basis from which 22 principal components were extracted.

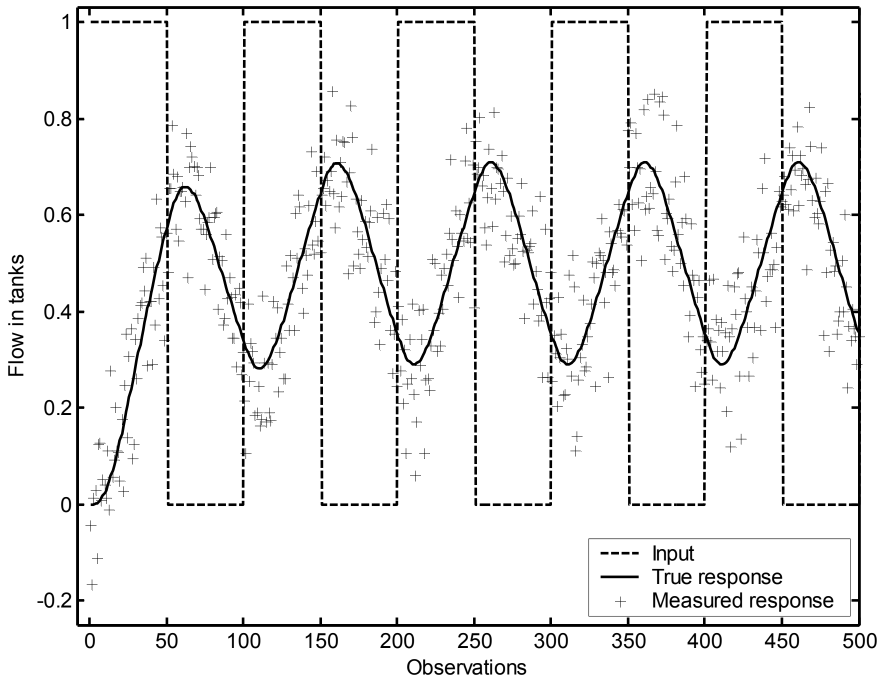


Figure 6.28 Actual system response (solid line), simulated measured response (+) and pulsed input signal (broken line) obtained from flow in four tanks in series.

6.2.2 Monte Carlo simulations

The Monte Carlo SSA was performed on the time series obtained from the flow in four tanks in series. The eigenvalues associated with each of the 22 principal components (eigenspectrum) of the time series are shown in Figure 6.29. Surrogate data were subsequently generated from the measured response time series by using the amplitude adjusted Fourier transform (AAFT) and confidence limits for the eigenspectrum were estimated by means of Monte Carlo simulations. These are also indicated in Figure 6.29.

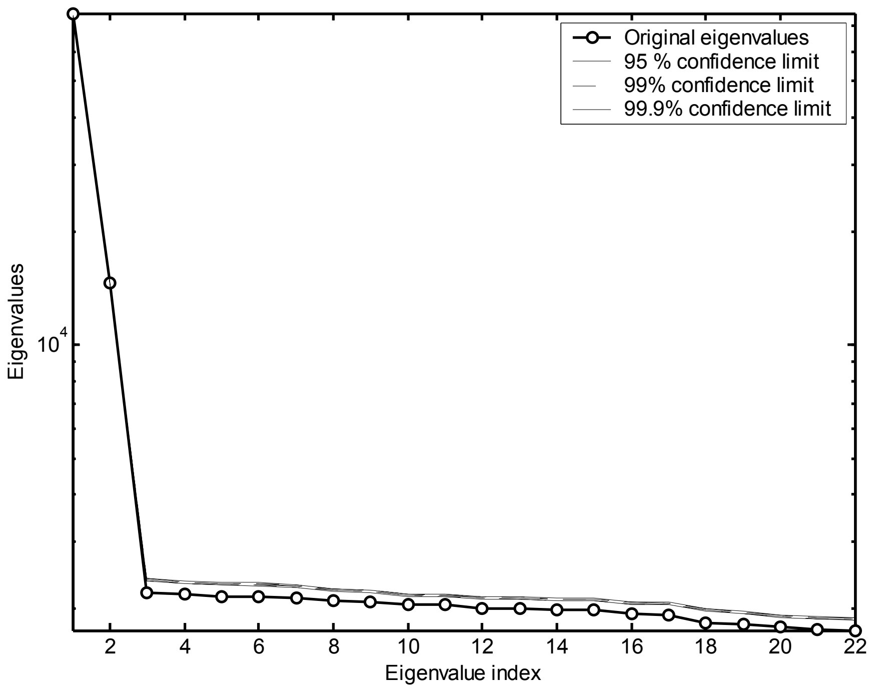


Figure 6.29 Eigenspectrum generated from complete series of four tanks in series data set, along with the confidence limits of the eigenspectrum obtained from Monte Carlo SSA on surrogate data sets generated by using the AAFT algorithm.

It can be seen from Figure 6.29 that the eigenspectrum of the nonstationary system falls outside its estimated confidence limits. If the figure is enlarged, this could be seen to be true even for the first two eigenvalues, even though it would not appear that way in Figure 6.29. These results therefore suggest that the null hypothesis of a stationary, linear Gaussian system has to be rejected. Since the system is Gaussian and linear, this rejection can only be attributed to the nonstationarity of the system, seen in the first few observations in Figure 6.28.

The effect of the initial transient part of the time series can be seen in the reconstructed attractor of the time series, displayed in Figure 6.30. The toroidal shape of the attractor described by the scores of the first three principal components reflects the roughly periodic behaviour of the system. The loose end portrayed at the bottom left of Figure 6.30 indicates the initial transient behaviour of the system.

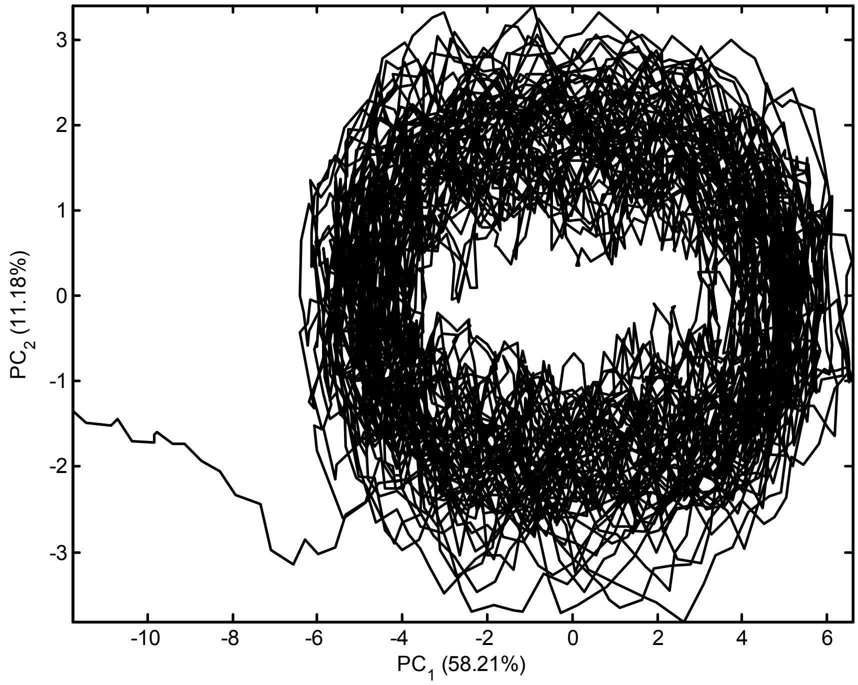


Figure 6.30 Attractor of the measured response shown in Figure 6.28, including the transient part of the time series from approximately 0-100 observations.

6.3 Autocatalysis in a continuous stirred tank reactor

6.3.1 Simulation of time series

The third case study concerns an autocatalytic process in a continuous stirred tank reactor originally considered by Gray and Scott (1983) and Gray and Scott (1984) and subsequently investigated by Lynch (1992). The system is capable of producing self-sustained oscillations based on cubic autocatalysis with catalyst decay and proceeds mechanistically as follows:

$$\mathrm{A} + 2\mathrm{B} \rightarrow 3\mathrm{B}, \quad -r_A = k_1 c_A c_B^2 \qquad 6.10$$

$$\mathrm{B} \rightarrow \mathrm{C}, \quad -r_C = k_2 c_B \qquad 6.11$$

$$\mathrm{D} + 2\mathrm{B} \rightarrow 3\mathrm{B}, \quad -r_D = k_3 c_D c_B^2 \qquad 6.12$$

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:

$$\frac{dX}{dt} = 1 - X - aXZ^2 \qquad 6.13$$

$$\frac{dY}{dt} = 1 - Y - bYZ^2 \qquad 6.14$$

$$\frac{dZ}{dt} = 1 - (1+c)Z + daXZ^2 + ebYZ^2 \qquad 6.15$$

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 e.
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; e = 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 X state was taken as the output variable.

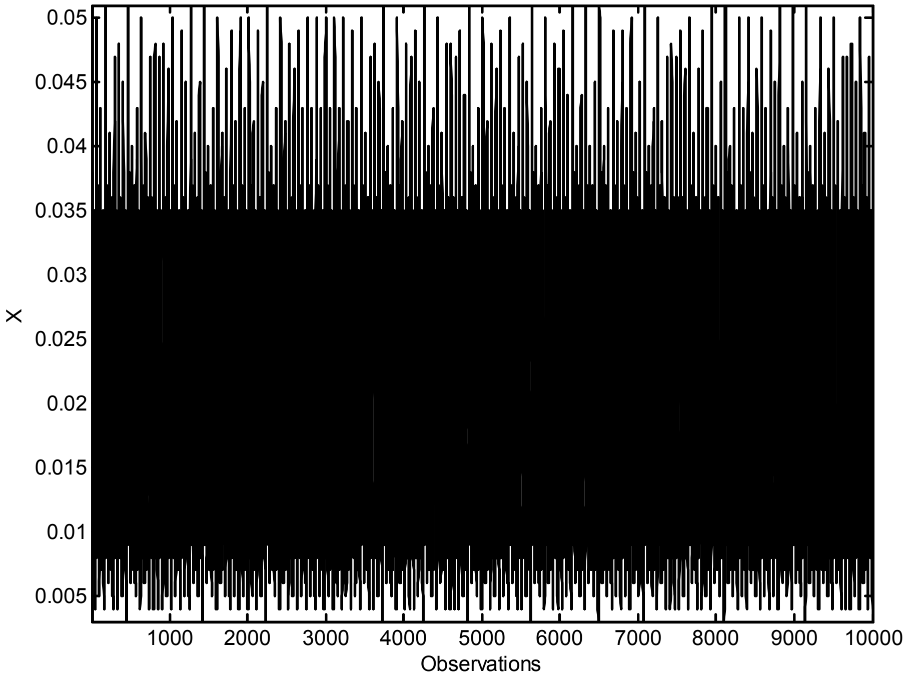


Figure 6.31 Dimensionless concentration time series of species A (represented by X) used for the analysis.

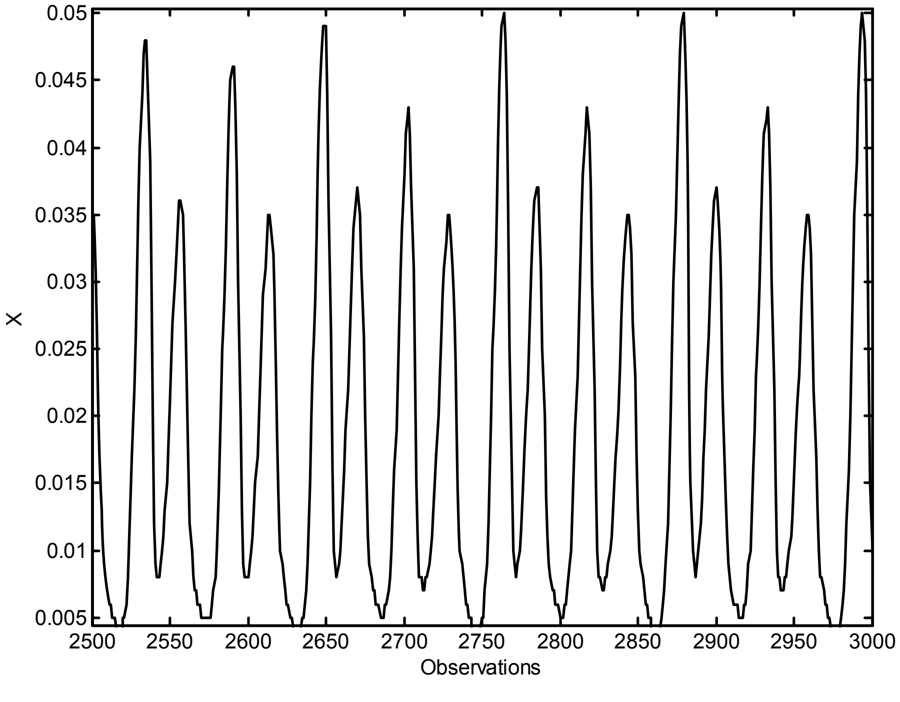


Figure 6.32 Close-up of a section of the dimensionless concentration time series of species A (represented by X) to illustrate the behaviour of the time series.

Figure 6.31 and Figure 6.32 respectively provide an illustration of the whole time series obtained from the autocatalysis in a CSTR and a close-up of a representative section of the time series.

6.3.2 Characterization of time series by use of Monte Carlo SSA

Due to the relatively low level of correlation between the observations in the time series, which is displayed in Figure 6.31 and Figure 6.32, it was found that it would be sufficient to embed the time series in eight dimensions.

Figure 6.33 shows the attractor of the process reconstructed from the first three principal components with the amount of variance represented by each of the components supplied in brackets next to the appropriate axis. As it has been mentioned earlier, a closed attractor is representative of an underlying periodic nature in the time series. However, this attractor never returns exactly to where it started from and can therefore not be classified as being closed. The appearance could rather be described to be a broad band, which is also indicative of vaguely periodic behaviour. The clearly defined shape indicates the deterministic nature of the time series in contrast to a completely stochastic time series, which would display very little regularity in the attractor. The large amount of variance represented by the first few eigenvalues can also be seen from the percentages supplied in the figure.

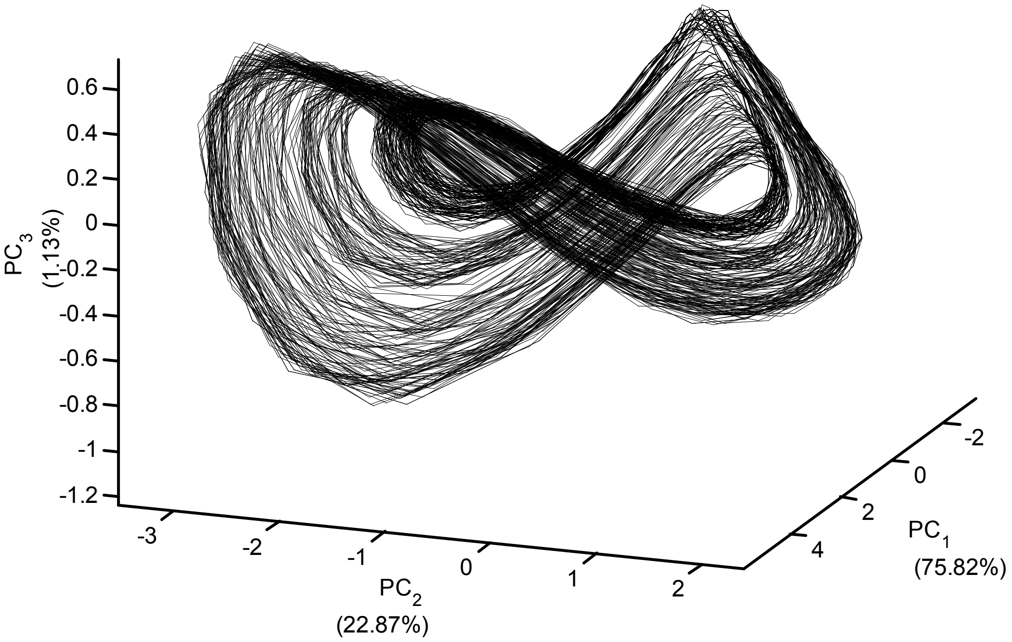


Figure 6.33 Reconstructed attractor from the first three principal components of the X process state. The percentage of the variance represented by each principal component is supplied in parenthesis next to the appropriate axis.

In Figure 6.34 and Figure 6.35, the eigenspectrum and estimated confidence limits of the simulated measurements from the autocatalytic system are shown. The confidence limits in Figure 6.34 were calculated from surrogate data sets generated by the AAFT algorithm and those in Figure 6.35 from surrogate data sets obtained from the IAAFT algorithm. As it has been mentioned earlier, the IAAFT algorithm place more restrictions on the generation of the surrogate data sets and is therefore a significantly more stringent test. This can be seen from the narrower confidence limits generated by the IAAFT surrogates than by the AAFT surrogates.

However, the eigenspectrum of the original observations falls outside the confidence limits generated by both algorithms by a wide margin, owing to the nonlinearity of the data (which are otherwise known to be stationary and non-Gaussian (not a stochastic time series)).

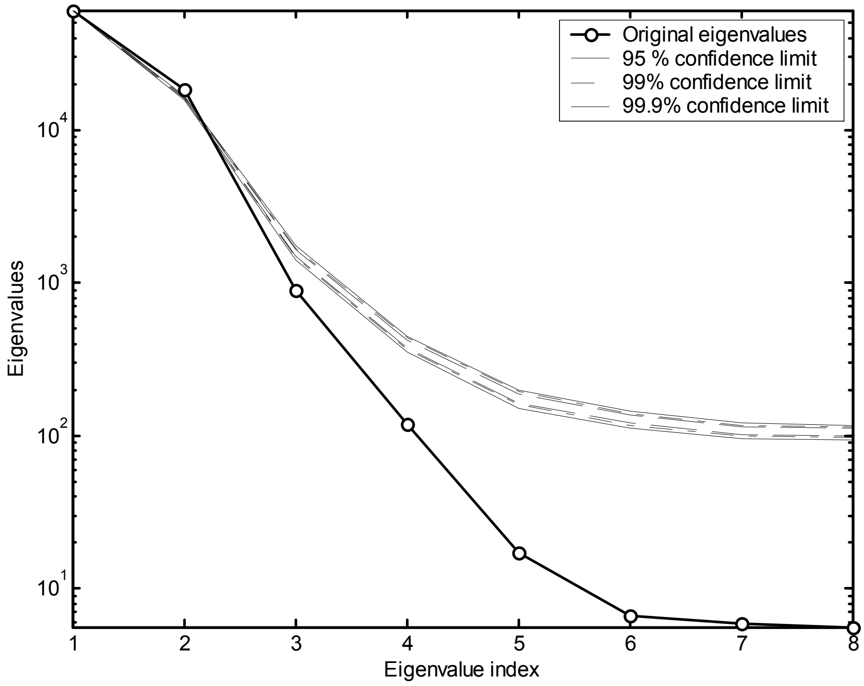


Figure 6.34 Eigenvalue distribution and confidence intervals for the eigenspectrum of the X process state of the autocatalytic CSTR reactor. The confidence intervals were calculated by means of surrogate data sets calculated from the AAFT algorithm.

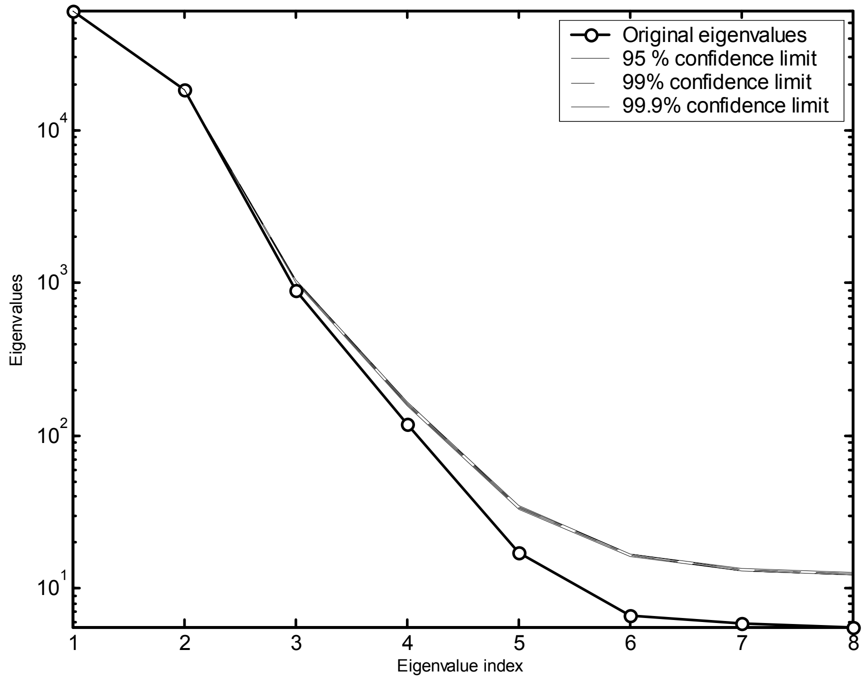


Figure 6.35 Eigenvalue distribution and confidence intervals for the eigenspectrum of the X process state of the autocatalytic CSTR reactor. The confidence intervals were calculated by means of surrogate data sets calculated from the IAAFT algorithm.

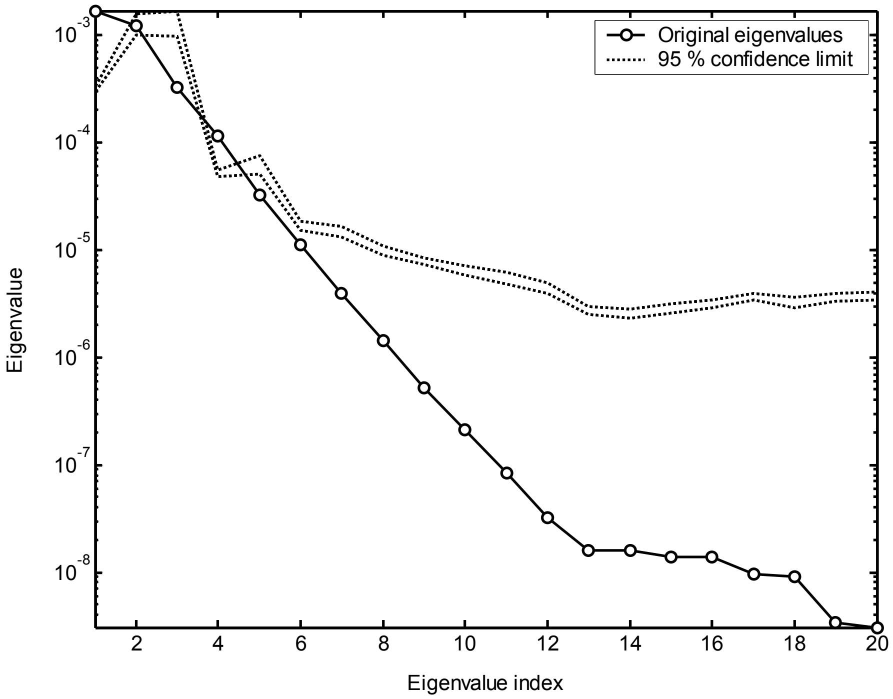


Figure 6.36 Eigenspectra and confidence bands testing if the autocatalytic CSTR reactor time series has similar properties than an AR(1) process.

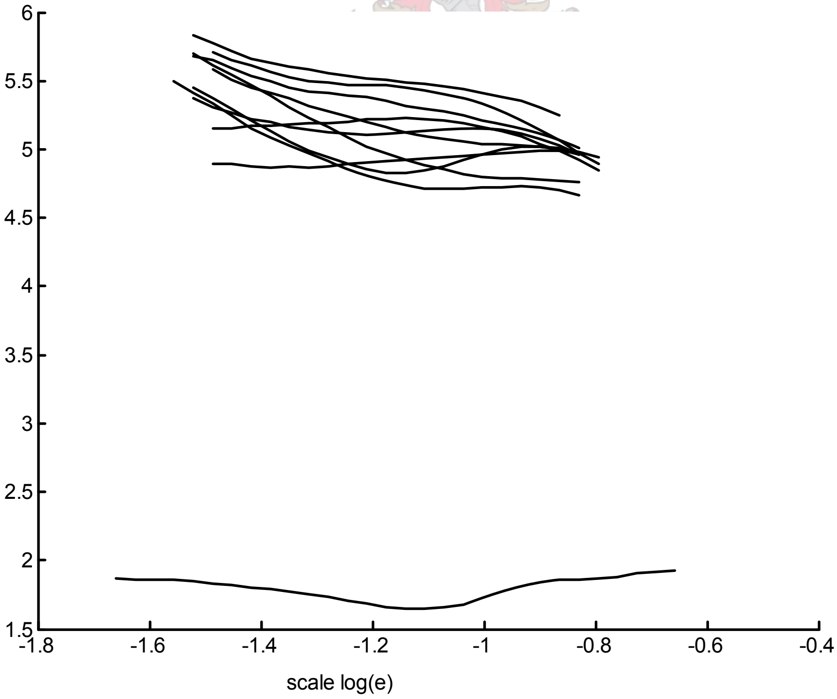


Figure 6.37 Correlation dimension of the X process state of the autocatalytic CSTR reactor (solid line) along with the correlation dimensions of 15 surrogate data sets (broken lines) generated by using the AAFT algorithm.

The test to see whether the data exhibit first order autoregressive properties also supplied negative results (Figure 6.36) with the eigenvalues falling outside the confidence limits, as would be expected for this series.

The results from the eigenspectra are confirmed by Figure 6.37, from which it can be seen quite clearly that the correlation dimension of the autocatalytic data is significantly different from that of the linear, stochastic, Gaussian surrogate data sets.

6.4 Summary

Monte Carlo SSA has been applied to three case studies, consisting of a variety of different types of time series. The underlying system characteristics of these series were known beforehand and they were used to determine the effectiveness and reliability of Monte Carlo SSA to characterise the nature of time series. The time series were tested against the null hypothesis of the series being from a linear, stochastic, Gaussian process. The test statistics used were both the eigenspectra of the series and the correlation dimensions.

It was found from this chapter that, especially in ‘extreme’ cases, such as highly nonlinear or chaotic time series, Monte Carlo SSA could be used very successfully to characterise time series. For the time series of which the characteristics were not as profound or marked, Monte Carlo did not succeed in all the applications. It should be remembered, however, that no single test to classify data sets has proved to be successful all the time for all time series. It is therefore recommended rather to use a combination of tests.

It was also found that, as was expected, IAAFT gave a much stronger test for the given null hypothesis than AAFT.

In practice, systems are usually contaminated with measurement noise, and not as readily classified as the previous two simulated systems. Such real-life systems will therefore be considered in the next chapter.

7 APPLICATIONS OF MONTE CARLO SSA

It has been established in the previous chapter that Monte Carlo SSA can be an extremely hand tool in characterizing time series by using observed data. This chapter will therefore be dedicated to the application of Monte Carlo SSA to two sets of time series obtained from real plants and processes, in an effort to obtain more information about the characteristics of these systems. Monte Carlo SSA will be combined with other tools that are regularly used for system characterisation in order to obtain more detailed or specific information about the various time series and underlying process dynamics, as it was found in the previous chapter that no single characterization technique can characterize all the systems correctly all the time.

7.1 Composition of scavenger circuit from base metal flotation plant

7.1.1 Background and SSA

The first real-life case study used data obtained from the recovery of precious metal in the scavenger circuit of a South African copper flotation plant. This particular case study has already been investigated in a previous section (section 4.5, on page 61) where SSA as a filtering technique was applied to the data.
The Cu, Pb and Zn time series were embedded in trajectory matrices of the same dimension as in section 4.5.1, viz. 26, 31 and 105 respectively. These matrices were subsequently decomposed into principal components, each with its own eigenspectrum, with upper and lower confidence limits computed from 30 surrogate data sets generated by both the amplitude adjusted Fourier transform algorithm (AAFT) and the iterative amplitude adjusted Fourier transform algorithm (IAAFT). The surrogate data sets were constructed to have the same power spectra as the original data set, which is consistent with the hypothesis of a stationary linear stochastic (Gaussian) process. If the eigenspectra of the time series are mostly outside the outer (99.9%) confidence limits, the hypothesis of a stationary linear stochastic process (possibly distorted by nonlinear measurement) has to be rejected, i.e. fitting of an autoregressive moving average model would not yield optimal results.
The confidence limits of both techniques are shown in Figure 7.1 - Figure 7.3, along with an enlargement of the first eigenvalue and its confidence limits, as calculated by the IAAFT algorithm. It can be seen that the confidence bands calculated by the two techniques differ significantly. The AAFT confidence limits seem broader than those of the IAAFT algorithm and for all three the time series the eigenvalues quite clearly fall outside the AAFT confidence limits, but more detailed inspection is required to determine the position with respect to the IAAFT confidence limits. The main focus will therefore rather be on the confidence limits calculated from IAAFT, as this seem to be the more stringent test.

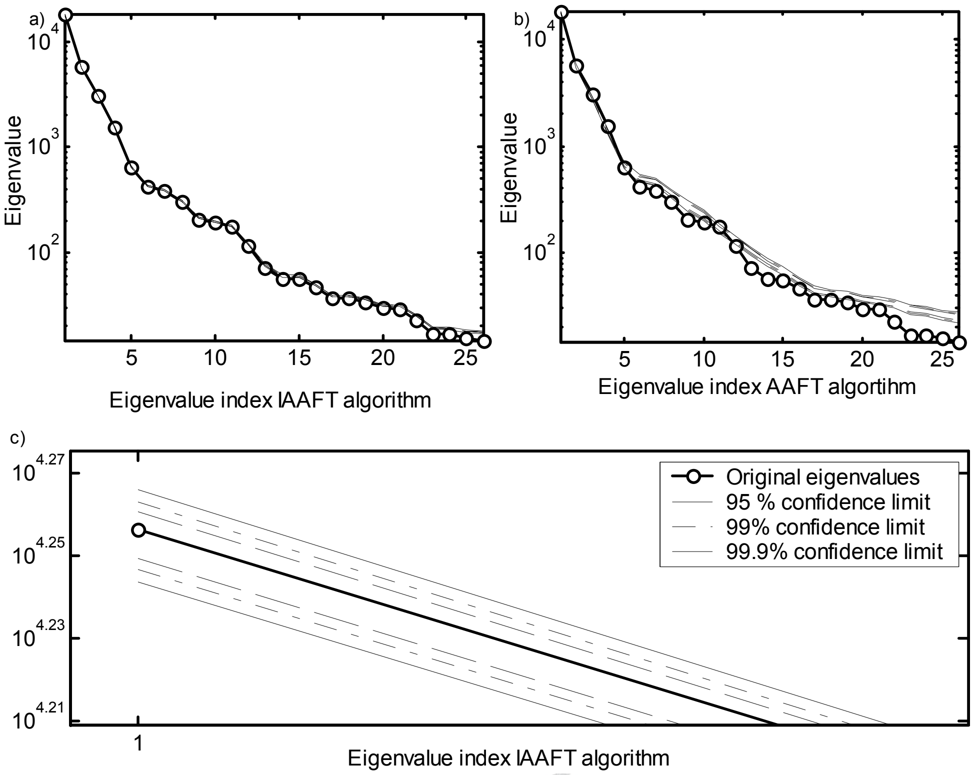


Figure 7.1 Eigenspectrum of copper time series along with confidence limits of the eigenspectrum, generated with a) IAAFT and b) AAFT algorithms and c) enlargement of the first principal component with the IAAFT confidence limits.

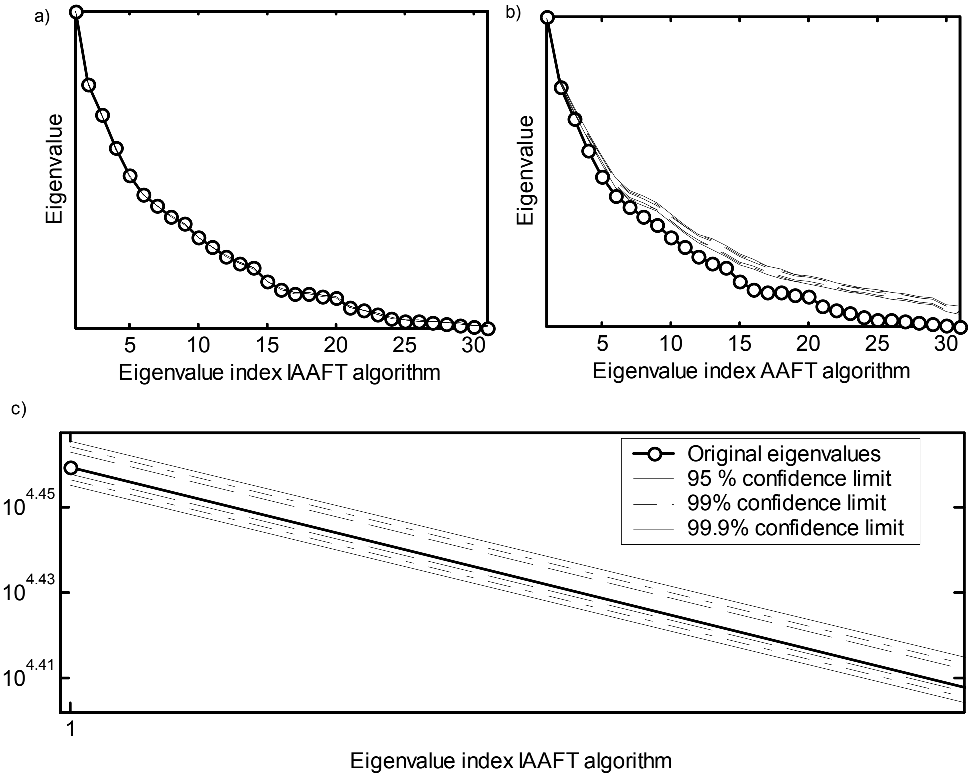


Figure 7.2 Eigenspectrum of lead time series along with confidence limits of the eigenspectrum, generated with a) IAAFT and b) AAFT algorithms and c) enlargement of the first principal component with the IAAFT confidence limits.

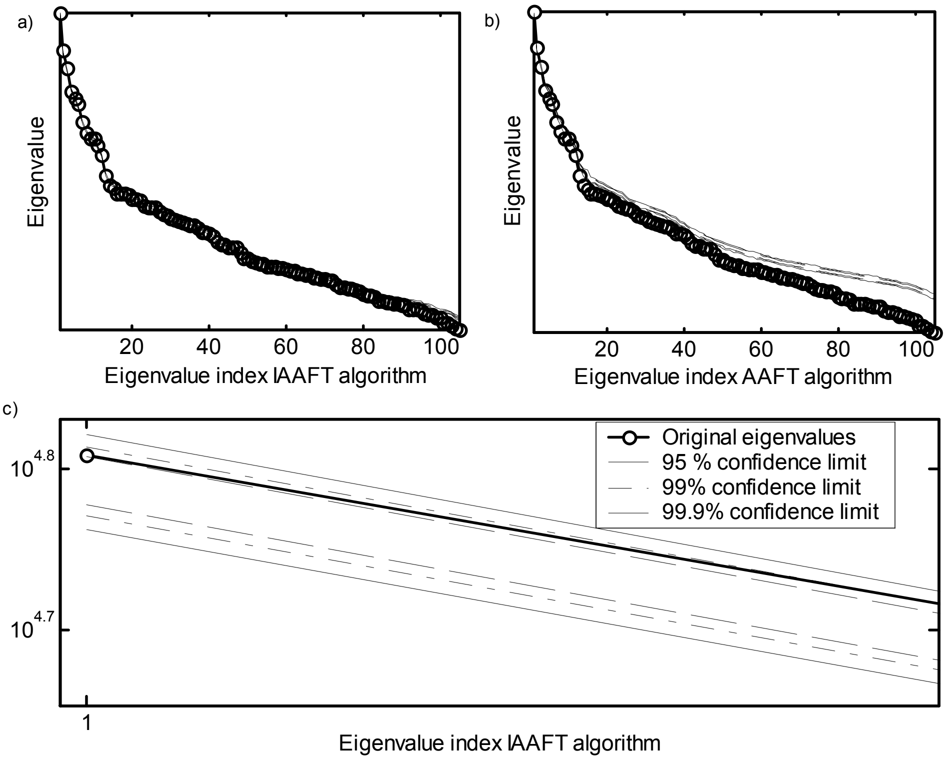


Figure 7.3 Eigenspectrum of zinc time series along with confidence limits of the eigenspectrum, generated with a) IAAFT and b) AAFT algorithms and c) enlargement of the first principal component with the IAAFT confidence limits.

7.1.2 Characterization of time series based on Monte Carlo SSA

As can be seen from Figure 7.1, the eigenspectrum of the copper data fell very close to the confidence limits for the eigenspectrum, especially to those of the IAAFT algorithm. This meant close inspection was necessary to determine whether the null hypothesis of a stationary linear Gaussian system could be rejected. Figure 7.1 (c) shows that the first eigenvalue falls inside the 95% confidence band and this was also found to be true for the third and fourth eigenvalues (not shown in detail), whereas all the other eigenvalues could be found to be outside the 99.9% confidence band. The eigenvalues became more removed from the confidence bands, the lower down the eigenspectrum these values were placed.
The eigenspectrum of the percentage lead in Figure 7.2 falls about 50% inside and 50% outside the confidence limits derived from the surrogate data. Although not shown in detail, only the second, sixth, sixteenth to nineteenth and twenty-first to thirtieth eigenvalues are outside the limits. The rest of the values are all at least 99.9% within the limits, with the majority falling within the 95% confidence band. On this basis, these data are also not strictly stationary and Gaussian (as suggested by further analysis of the data). It should be borne in mind that each eigenvalue represents a component of the time series and that each of these can be considered individually in the Monte Carlo tests. In this thesis, the entire eigenspectrum is considered, without formally differentiating between the different eigenvalues (time series components). Nonetheless, where a large number of eigenvalues are considered, more emphasis should be placed on the first few eigenvalues than on the last few. This has been discussed on an ad hoc basis where applicable in the thesis (Chapter 6, AR(1) process).
In the eigenspectrum of the percentage zinc in Figure 7.3, once again the eigenvalues are distributed both inside and outside the confidence limits. The last half of the eigenvalues (i.e. from eigenvalue 49 to 105) all fall outside the confidence limits derived from the surrogate data, while the first number of eigenvalues are mostly inside the confidence bands, with values 2, 3, 6, 12, 15-17, 19-20, 24, 33-34 and 37-39 found outside the 99.9% confidence limits. Again, strictly speaking, the data are therefore not entirely stochastic, or stationary and further tests are necessary.

The lack of stationarity in the data is confirmed by Figure 7.4. This figure was generated by performing standard principal component analysis on the system of three time series (Cu, Pb and Zn) that were shown initially in Figure 4.24. The first two components (t-values) were plotted as a function of each other, with a distinction being made between the first (o-markers) and the second (+-markers) halves of the time series. 95% and 99% confidence bands were also generated for the score plot. The nonstationarity is clearly evident, the extent of which is highlighted by Figure 7.4, where most of the deviations between the first and second half of the time series can be attributed to the lead and zinc contents in the feed. If the data were stationary, a much larger percentage of the data from the second half of the time series would have been within the confidence bands.

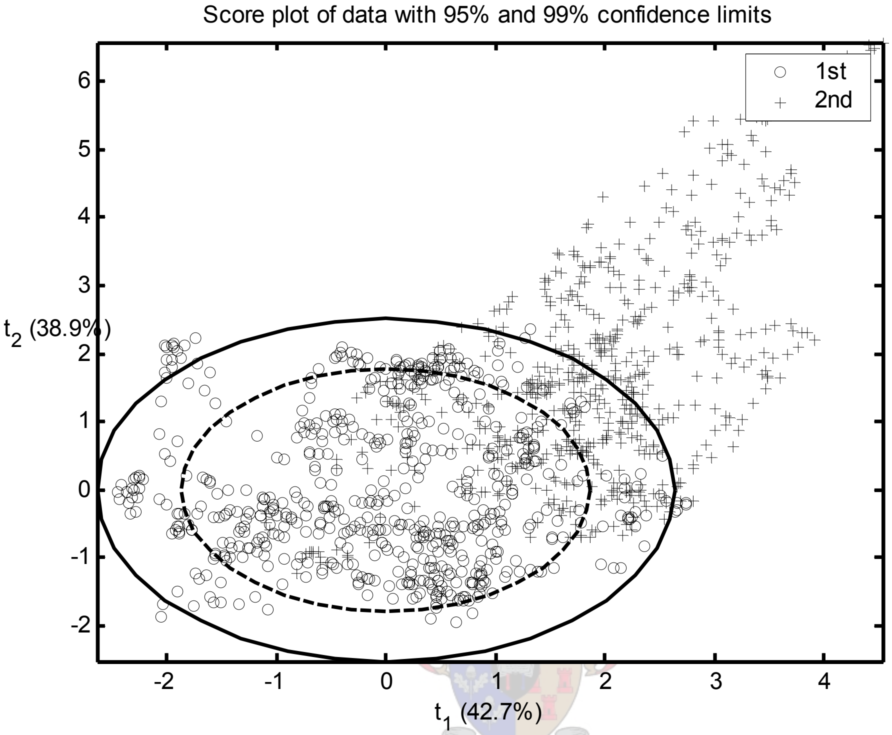


Figure 7.4 Nonstationarity of the feed to the flotation plant, showing the principal component scores of the first two weeks (o) and the last two weeks of the feed (+). The percentage of the total variance explained by each principal component is shown in parenthesis in the appropriate axis label.

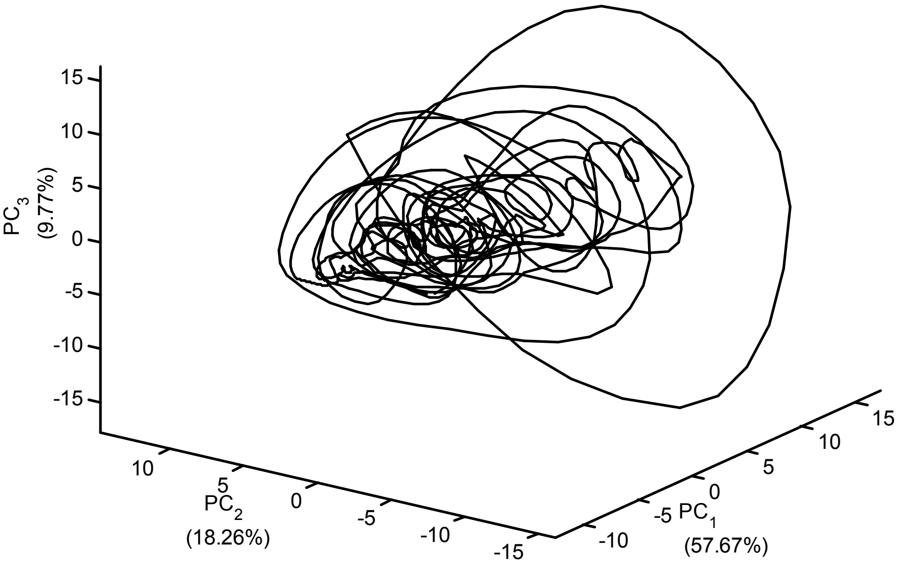


Figure 7.5 Attractor of copper time series. The percentage of the total variance explained by each principal component is shown in parenthesis in the appropriate axis label.

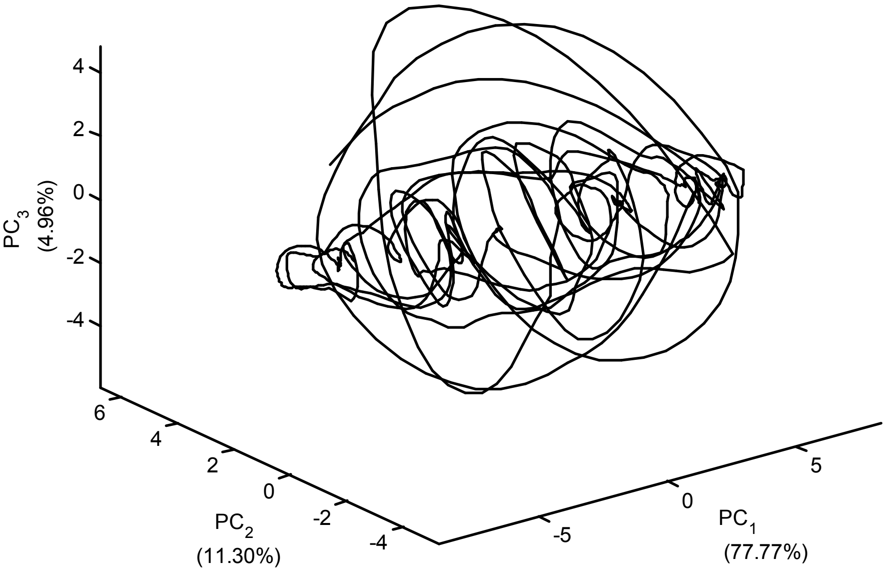


Figure 7.6 Attractor of lead time series. The percentage of the total variance explained by each principal component is shown in parenthesis in the appropriate axis label.

The reconstructed attractor of each data set is shown in principal component space in Figure 7.5 - Figure 7.7. Inspection of the attractors can give important clues with regard to the stationarity of the data (closed recurring orbits), the predictability of the system (smoothness and boundedness of the attractor), etc. When the three attractors are compared, the zinc time series appears to be less stationary than the other two series, seeing as no evidence can be found of recurring orbits in the attractor, while the attractor of the copper time series indicates very pronounced recurring orbits.

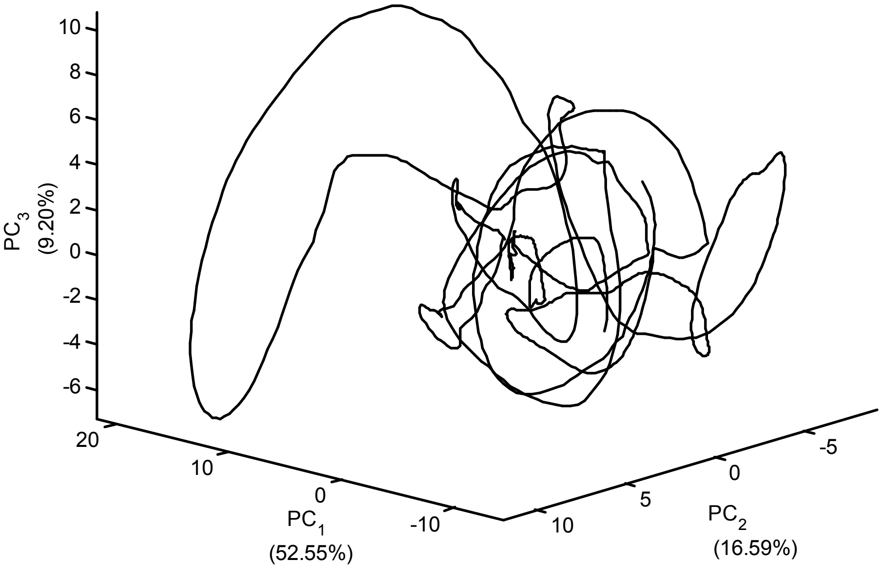


Figure 7.7 Attractor of zinc time series. The percentage of the total variance explained by each principal component is shown in parenthesis in the appropriate axis label.

Quantile-quantile plots (Figure 7.8) were generated for the data in order to provide more information on whether the data are normal or not. These plots suggest that the data are roughly normal. Quantile-quantile plots, such as these, are often used in practice to assess models of the distribution of time series. In Figure 7.8, the data were plotted against standard normal quantiles. Deviations from the dotted lines in these plots would then indicate deviations from a normal distribution, especially if the deviations had occurred towards the middle part of the plots.

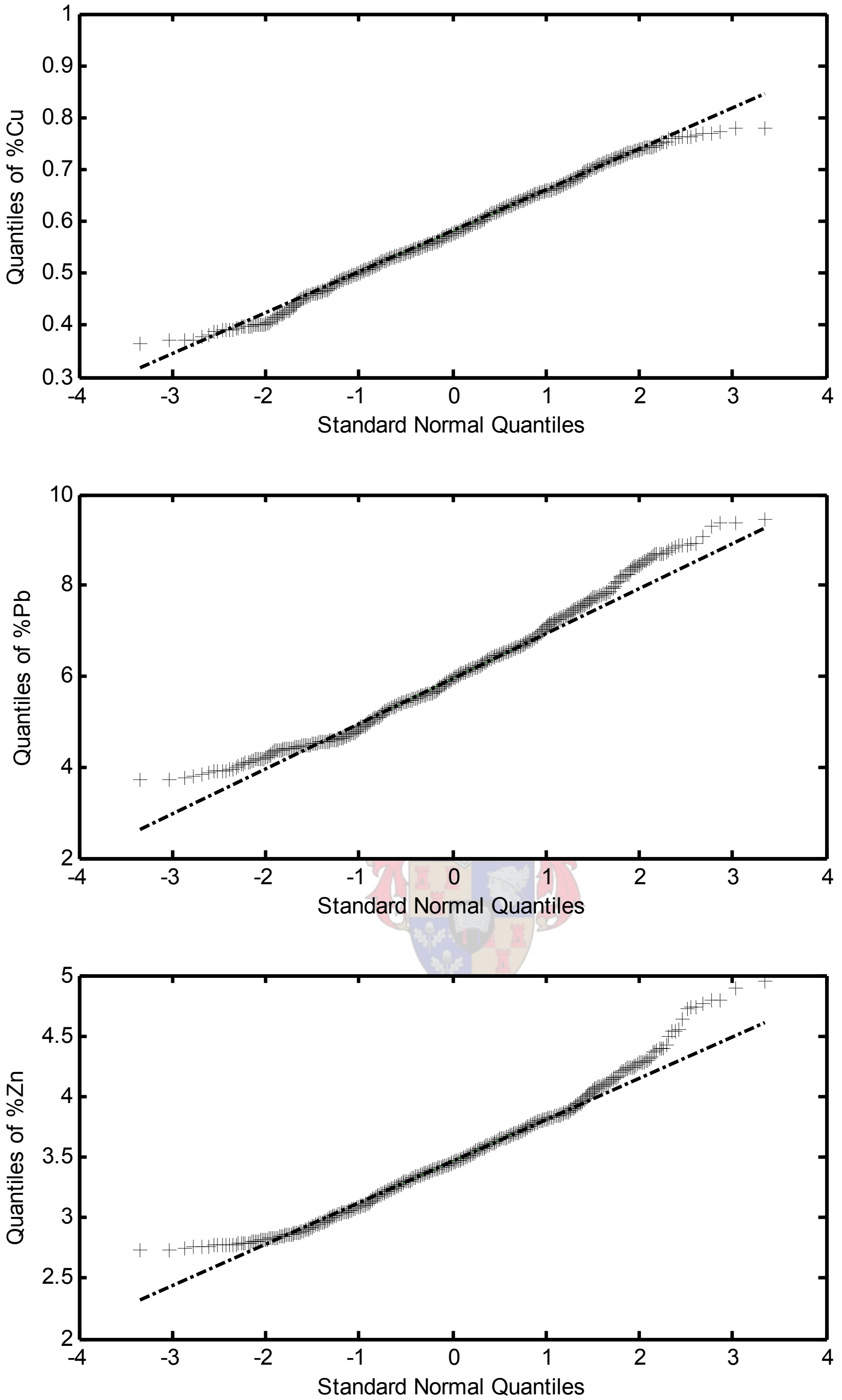


Figure 7.8 Q-Q plots of Cu, Pb and Zn feed data versus standard normal.

The stationarity of the system can be assessed by comparing the mean of the first half of the observations with the mean of the second half in an ANOVA test. The results (with those for the other two metals) are summarized in Table 7.1. In this table the total sums of squares (SS) is partitioned into the sums of squares for the variables (Period 1 and Period 2), as well as the error terms. As can be seen from the results, the probability of the Pb and Zn being stationary is negligible, while that of the copper is approximately 12%. It is a crude test, but nonetheless an indication of the stationarity of the data. On this basis alone, the eigenspectra of all three metals should fall outside the confidence limits generated by the Monte Carlo analyses. Although none of the eigenspectra displayed in Figure 7.1 to Figure 7.3 fell completely outside the confidence limits, but were rather partly inside and partly outside, the fact that they were not completely inside correlates with the expected characteristics.

Table 7.1 Analysis of variance for Cu, Pb and Zn

Cu	Source	SS	df	MS	F	Prob>F
	Periods	0.0403	1	0.0403	6.41	0.012
	Error	7.76	1232	0.0063		
	Total	7.80	1233			
Pb	Source	SS	df	MS	F	Prob>F
	Periods	700.47	1	700.47	1268.6	0
	Error	680.23	1232	0.5521		
	Total	1380.7	1233			
Zn	Source	SS	df	MS	F	Prob>F
	Periods	35.45	1	35.45	341.08	0
	Error	128.06	1232	0.104		
	Total	163.52	1233			

7.2 Avalanching behaviour of particles

7.2.1 Background

The behaviour of granular substances is a very complex field of study and has therefore been the subject of a great amount of research. The specific phenomenon that was under investigation for this case study was the avalanching behaviour of powders and particles. Avalanching is an occurrence that could, on the one hand, lead to disastrous results if it takes place on a large enough scale. However, the studying of avalanches on a much smaller scale is just as important, as this affects the successful operation of a number of processes, such as the discharging of grains from silos and the feeding of particulate raw materials to a process. Erratic discharge can lead to problems in processing or ultimately in the product quality.

Previous work on the subject of avalanches has been done by a number of researchers. Held et al. (1990) devised an experimental set-up that allowed sand to fall one grain at a time onto the pan of a high-precision balance. This set-up was used to study the avalanches of sand cascading down a sand pile, and as sand fell of the edges of the plate, the fluctuations in the mass of sand on the balance were measured.

Avalanches in a wide range of sizes were observed over the time the experiment was run. From the results it was concluded that the sand pile had organized itself to a critical state. However, when it was tried to repeat the same experiments with a larger base plate, only large avalanches were observed. This lead to the conclusion that only small piles will naturally evolve to a critical state. No explanation was given for this conclusion.

Rastogi and Klinzing (1994) aimed to extend the work of Held et al. (1990) to a larger scale. This was done by studying a large pile, which better represented actual avalanches than when one grain was dropped at a time. The set-up consisted of a hopper, feeding solids onto a conveyor belt, from which the solids fell down at a trickle. An aluminium plate served as a base on which the solids pile formed. This plate was suspended by three chains and below the plate a hollow cone directed the fallen particles towards the balance below. The balance was connected to a computer that took readings every five hundred milliseconds. Glass beads of different sizes and shapes, as well as different sized aluminium particles were used for the various runs.

The weight of the avalanches and the frequency of occurrences were noted. The results were illustrated as the weight of the avalanche versus the percentage of avalanche with weight greater than that specified. A low percentage of avalanches with small weights was noticed. It was assumed that small avalanches became distributed on the slope and were absorbed moving towards some inequilibrium. This continued until a threshold was reached, which resulted in a large avalanche.

The angle of repose during the experiments was determined by videotaping the experiment and then using an image analyser. Their results concluded that the system’s fractal dimension (which is indicated by the number of different slopes of the cumulative weight data plots) indicated the measure of the tendency of the solid to flow freely. It was seen once again that larger and smaller particles showed different flow tendencies.

Two dimensional strange attractor plots of the times and weights of consecutive avalanches were constructed. These plots confirmed the process as chaotic and showed the existence of attractor sets. It was concluded that an avalanche can only be correlated with a succeeding or preceding one, but not necessarily with one that occurs several steps away.

Some of the most recent work was done by Smith and Tüzün (2002) specifically on the stress, voidage and velocity coupling in an avalanching granular heap. This paper followed a more theoretical approach, in that the advantages of the application of wavelet transforms were investigated, rather than just making conclusions on the behaviour of particles during avalanches. Their experimental set-up consisted of only two dimensions, as they stated that phenomena visible in two dimensions would generally be attenuated into three dimensions. Particles were fed as a trickle onto a heap and the output from the simulation was a time-referenced set of variables at particle level. These variables were analysed using discrete wavelet transforms and it was found that the correlation between time-lagged wavelet transform coefficients can be much more informative than the correlation functions derived from the original time series themselves. Thereby they have identified a novel way to determine time constants in the context of discrete events, by correlating wavelet coefficients.

In this section, another alternative for the characterisation of the flow behaviour of granular materials is investigated. When particles are completely free flowing, the flow behaviour can be seen as a stochastic process. However, once conditions of self-organized criticality develop, the particles cannot be considered as free flowing anymore and alternative analysis methods should be investigated. The correlation dimension of a time series can be seen as a good indication of the degree of determinism in the time series and can therefore be used to characterize the time series. In this section, an attempt was made to characterize the flow and avalanche behaviour of a number of different materials under various conditions, with the use of the correlation dimension and Monte Carlo singular spectrum analysis.

7.2.2 Experimental set-up and resulting time series

The main experimental set-up that was used is illustrated in Figure 7.9. The equipment consisted of a cylinder that contained the granular substance and from which the particles flowed through a nozzle, onto a mounted disc. The particles would then accumulate on this disc, until an avalanche occurred and the fallen particles cascaded onto an electronic scale. The scale was connected to a computer, which recorded the reading on the scale every 200 ms, resulting in a time series that represented the accumulation of weight on the scale. The scale could only weigh a maximum of 2 kg accurately, and thereby limited the amount of particulate substance used in each experiment.

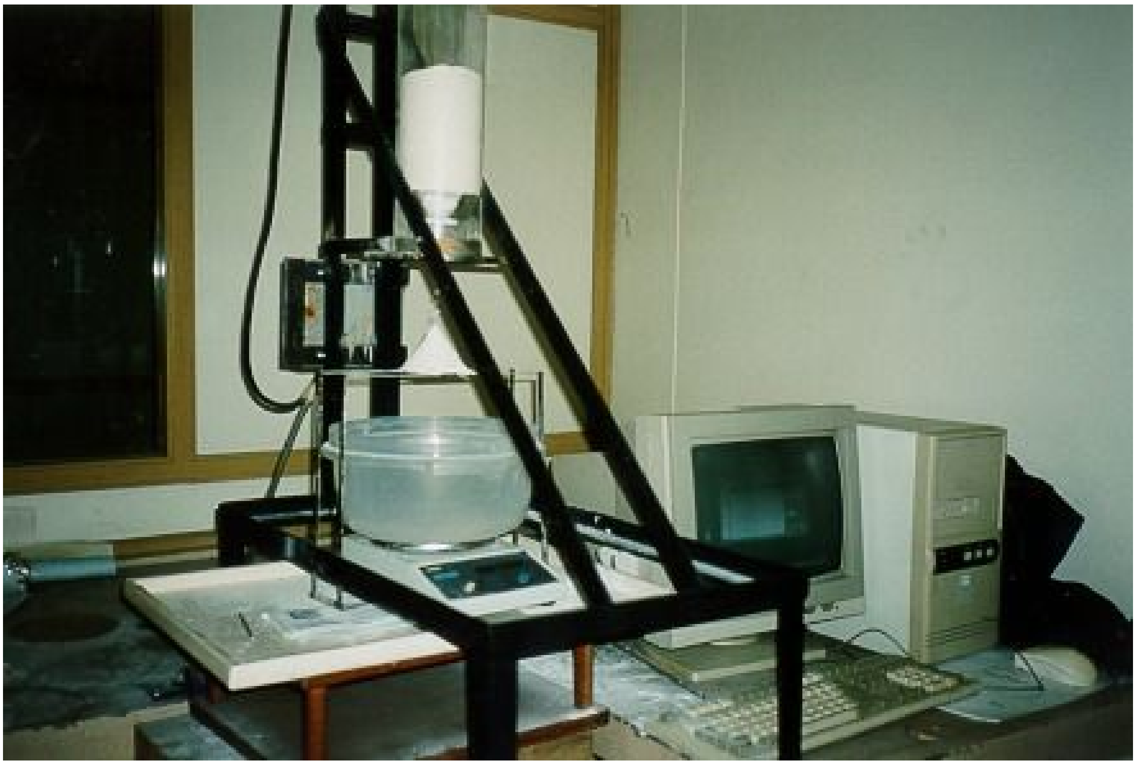


Figure 7.9 Experimental set-up used to investigate avalanching behaviour of various types of particles.

The cylinder containing the particles was equipped with a motorized vibrator to aid the flow of the more stable particles. Various sized nozzles were used and the disc underneath the nozzle could also have one of two different diameters. The available nozzle sizes were 5mm, 8mm, 10mm, 15mm, and 20mm and two different sized discs that were used had diameters of 80mm and 160mm respectively. The substances that were investigated were tiling cement, salt, sand, maize-meal and cake flour.

The experiments were also adapted to investigate the effect of temperature and moisture content on the flow of particles. The first was achieved by slightly increasing the moisture content of the sand and then comparing the flow behaviour to that of very dry sand. The moisture content of the ‘wet’ sand was 0.33% higher than that of the dry sand.

By respectively heating and freezing a batch of the maize-meal, the effect of temperature could be investigated. The temperature differences achieved in this part of the experiment varied from -17˚C to 5˚C for the cold maize-meal and while the warm maize-meal cooled down from 114˚C to 60˚C during the course of the experiment. The experiments were all run for a substantial period of time, and therefore the temperatures of the particles before and after the experiments were measured. This allowed one to obtain an average temperature difference.

The time series obtained directly from the measurements during the experiments had a non-stochastic nature, due to the accumulation of weight on the scale during the avalanching process, as can be seen in Figure 7.10. This figure represents sections of the original measured time series obtained during the experiments. The difference in the flow behaviour of the various particles can already be seen from the variations in the appearance of the measurements.

The non-stochastic nature of all the series necessitated the performance of a basic transformation of the measured time series to ensure that the time series used for further analysis was stationary. A linear regression curve was modelled to the measured data and then the residuals of the fitting were calculated. These residuals were used as the stochastic time series. The transformed time series for the various granular substances are illustrated in Figure 7.11. Due to the variations in flow rate and flow behaviour of the different types of particles, the number of observations for each of the series are not the same.

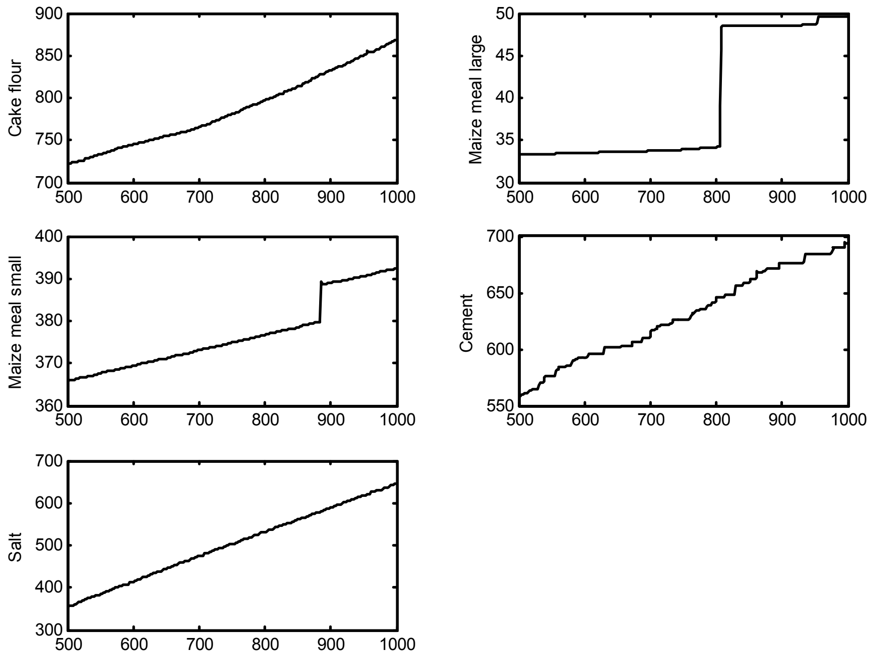


Figure 7.10 Sections of original time series obtained from avalanching experiments to illustrate difference in flow behaviour.

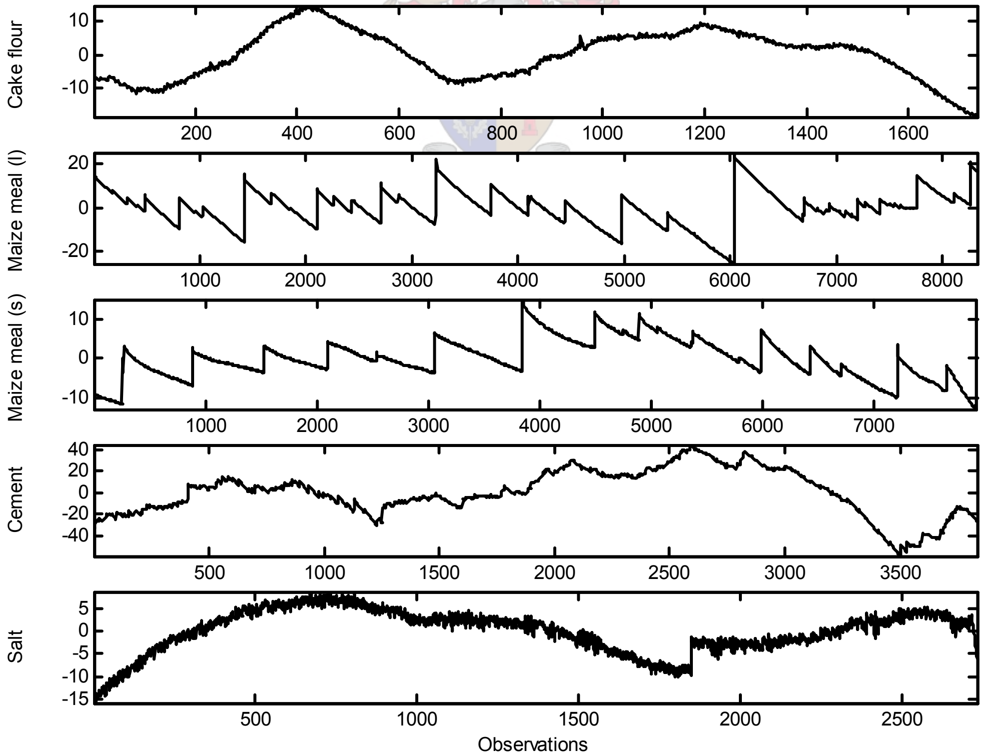


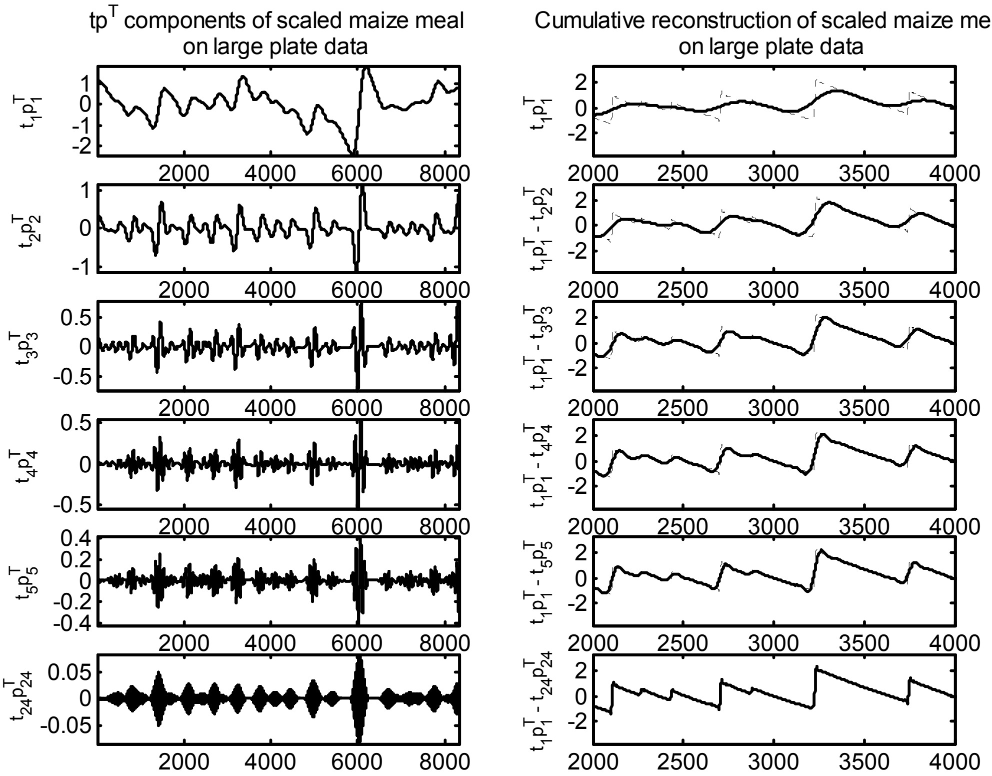
Figure 7.11 Transformed time series obtained from the experiments performed on the various granular substances and used for subsequent analysis.

7.2.3 Singular spectrum analysis results

Each of the five particle systems presented above were analysed using normal linear singular spectrum analysis. During the analysis it was found that these series were some of the more ‘problematic’ case studies investigated thus far. The time series either had very high autocorrelations, in that the embedding dimensions of the trajectory matrices were very high, or there was no obvious noise floor in the eigenvalue spectrum, making it very complicated to find a criterion for the number of eigenvalues to retain in the reconstruction of each series. It was found, however, that the series all behaved quite differently in accordance with the variety of particle flow behaviours observed in the experimental set-up.

All five the time series were embedded in relatively large dimensions, with the window length of the cake flour, maize-meal on a large plate, maize-meal on a small plate, cement and salt time series being 160, 188, 266, 401 and 336, respectively. As could be expected from the significantly different visual appearance of the time series, the eigenvalue spectra displayed a variety of behaviours. Each time series was then reconstructed with the optimum number of principal components for that series.

The cumulative reconstructions for each of the series, along with an illustration of a number of the separate reconstructed components are supplied in Figure 7.12 to Figure 7.16. It can be seen that all the time series can be reconstructed quite well with only a limited number of retained components, even though most series were expanded into quite a significant trajectory matrix. It is worth noting the similarities in the individual components of the two maize meal series, as both series exhibit the same development of the components in terms of the amount of variance explained by each progressive component and the frequency of the variance.


tp^T components of scaled maize meal on large plate data

Cumulative reconstruction of scaled maize meal on large plate data

Figure 7.12 Individual and cumulative reconstructed components of maize-meal on large plate data series.

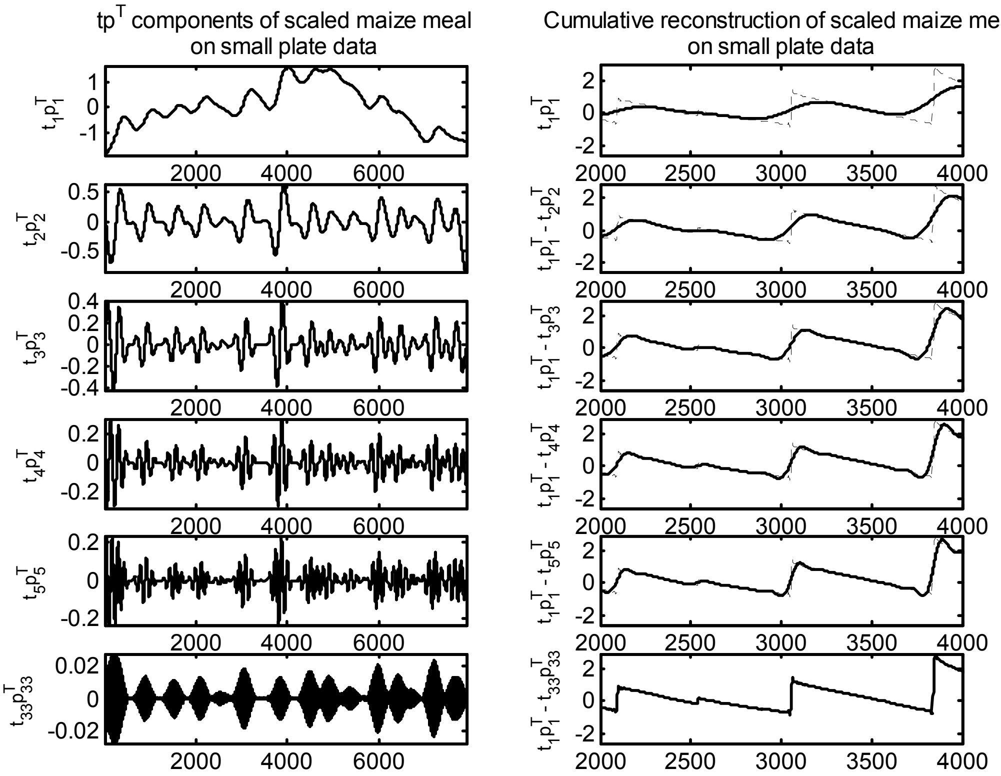


Figure 7.13 Individual and cumulative reconstructed components of maize-meal on large plate data series.

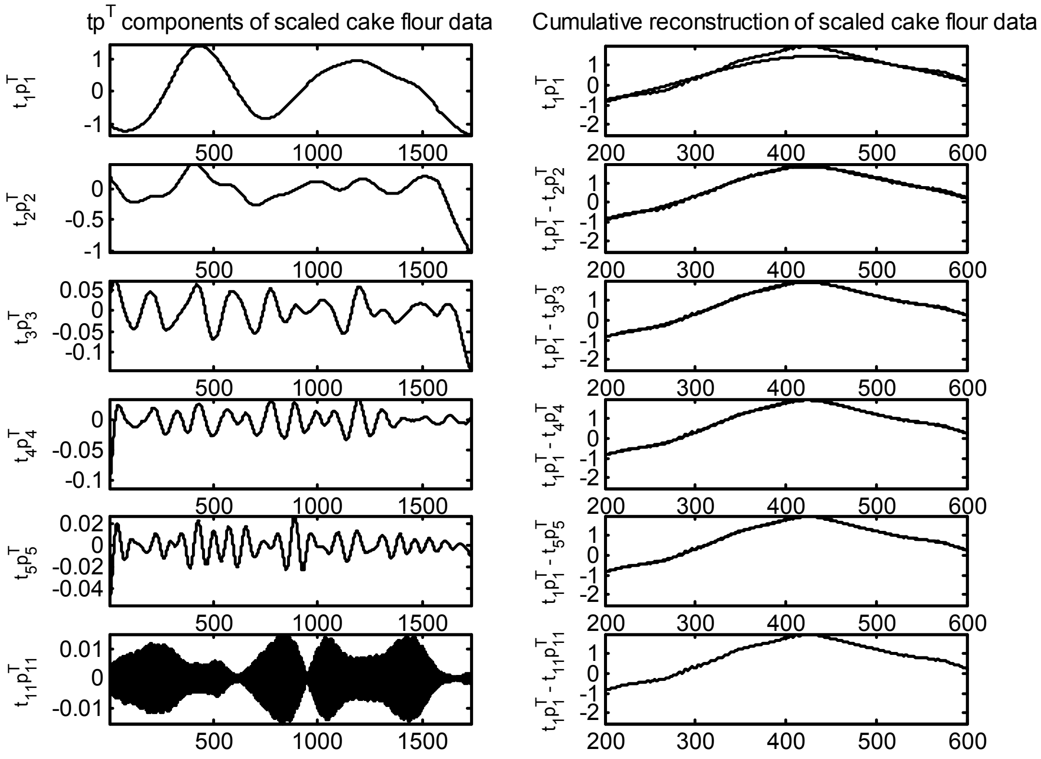


Figure 7.14 Individual and cumulative reconstructed components of cake flour data series.

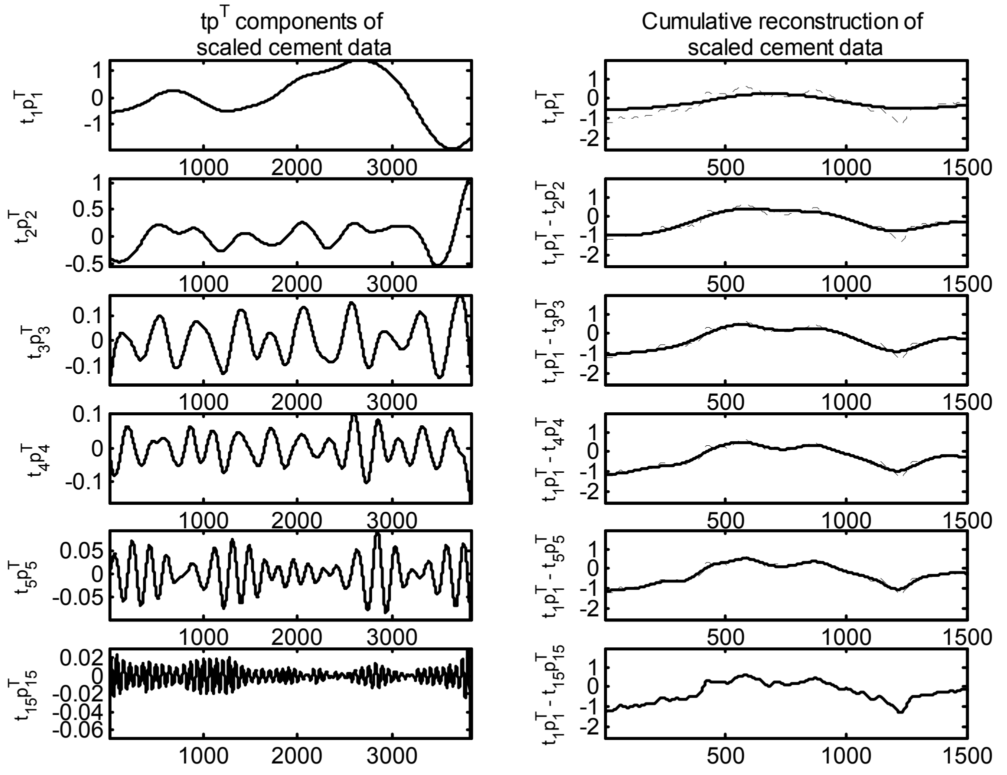


Figure 7.15 Individual and cumulative reconstructed components of maize-meal on large plate data series.

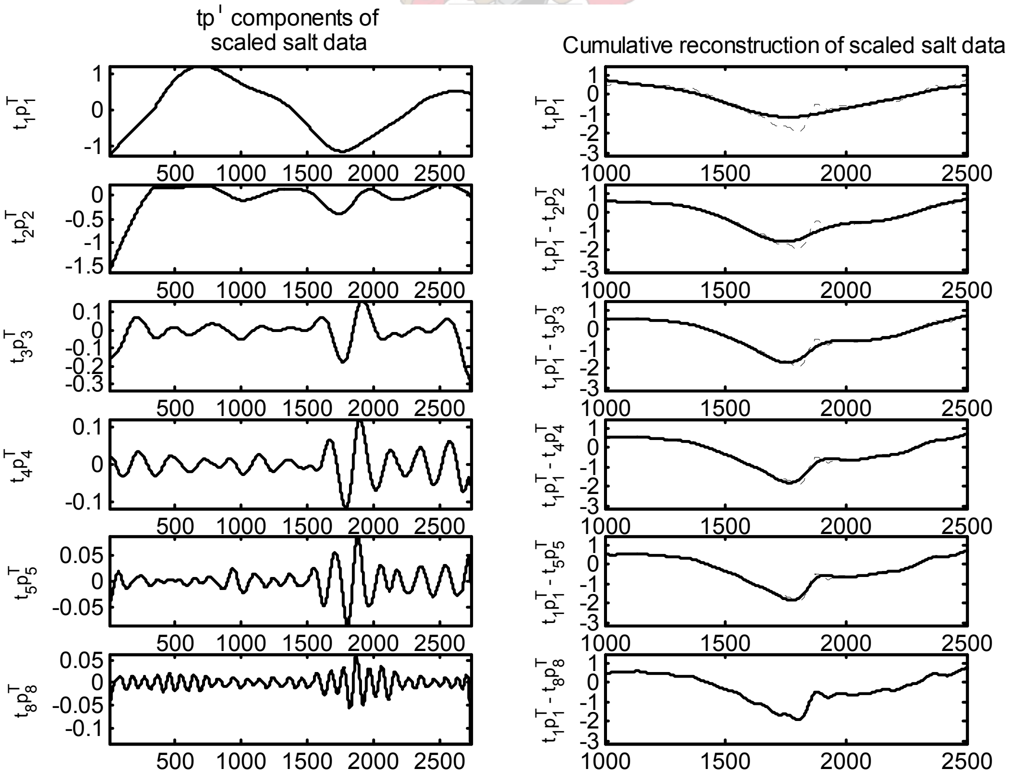


Figure 7.16 Individual and cumulative reconstructed components of maize-meal on large plate data series.

7.2.4 Classification of flow behaviour by means of Monte Carlo SSA

The eigenspectrum for each embedded time series, as well as the confidence limits for the eigenspectrum, calculated by using 15 surrogate data sets generated by the IAAFT algorithm, are presented in Figure 7.17 - Figure 7.26. Each eigenspectrum is also followed by an enlargement of the first few eigenvalues to provide a better illustration of the position of the eigenspectrum relative to the confidence limits calculated for that spectrum.

When these figures are studied and compared, it is observed that, not only does the shape of the eigenspectrum differ among the series, but the position of the eigenspectra relative to the confidence limits of the respective eigenspectra is also quite different. For a series such as the cake flour time series, the eigenspectrum of the time series is positioned relatively far outside the 99.9% confidence limits, while for both the maize-meal time series a substantial part of the eigenspectra are within the 95% confidence bands with the rest of the maize-meal eigenspectra still following the confidence limits closely. The eigenspectra of the other two time series displays behaviour somewhat in between these two extremes.

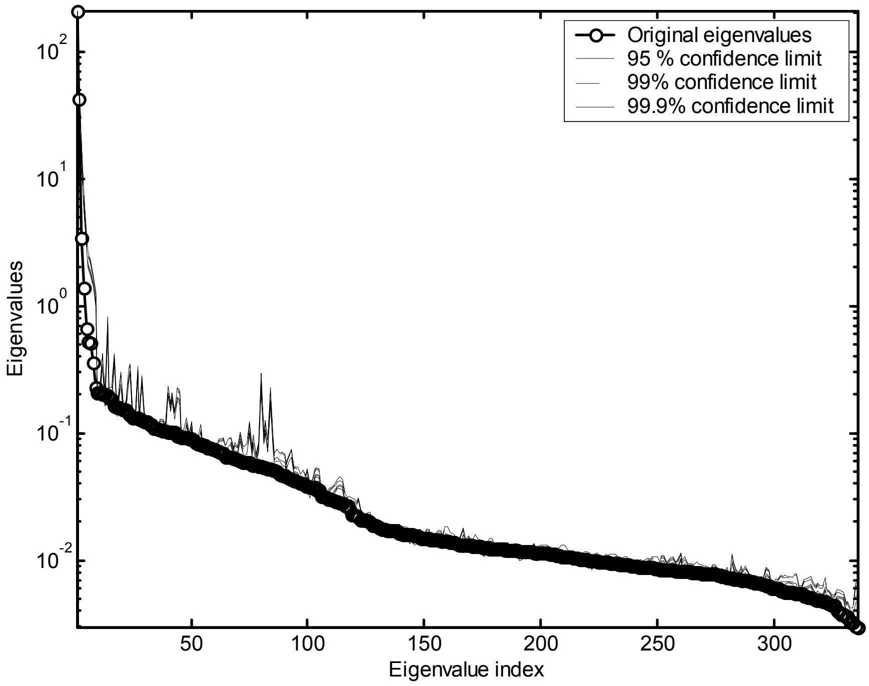


Figure 7.17 Eigenvalue spectrum of salt time series along with confidence limits for the eigenvalues calculated from surrogate data generated by the IAAFT algorithm.

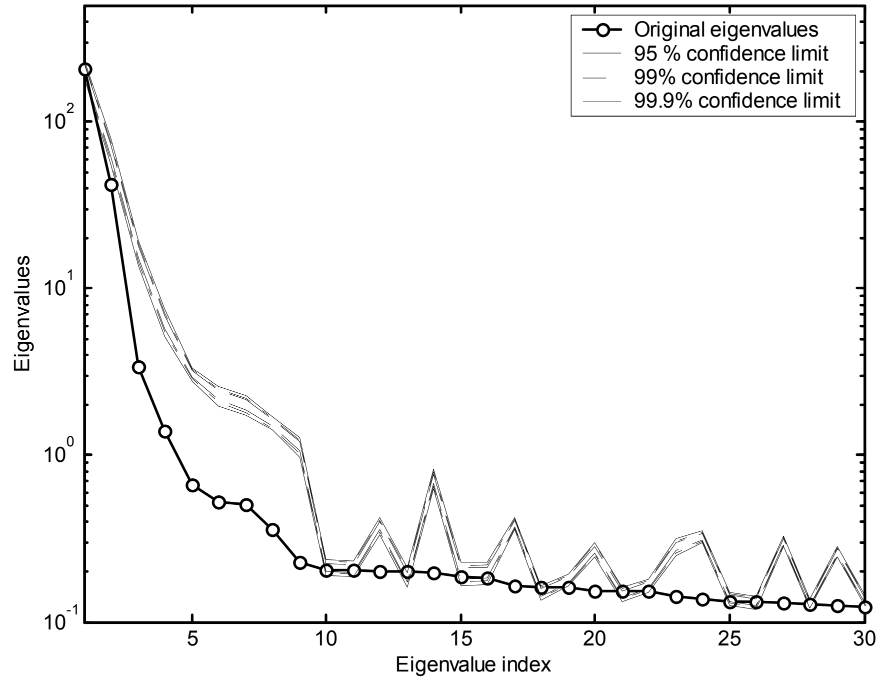


Figure 7.18 Enlarged section of eigenvalue spectrum of salt time series along with confidence limits for the eigenvalues calculated from surrogate data generated by the IAAFT algorithm.

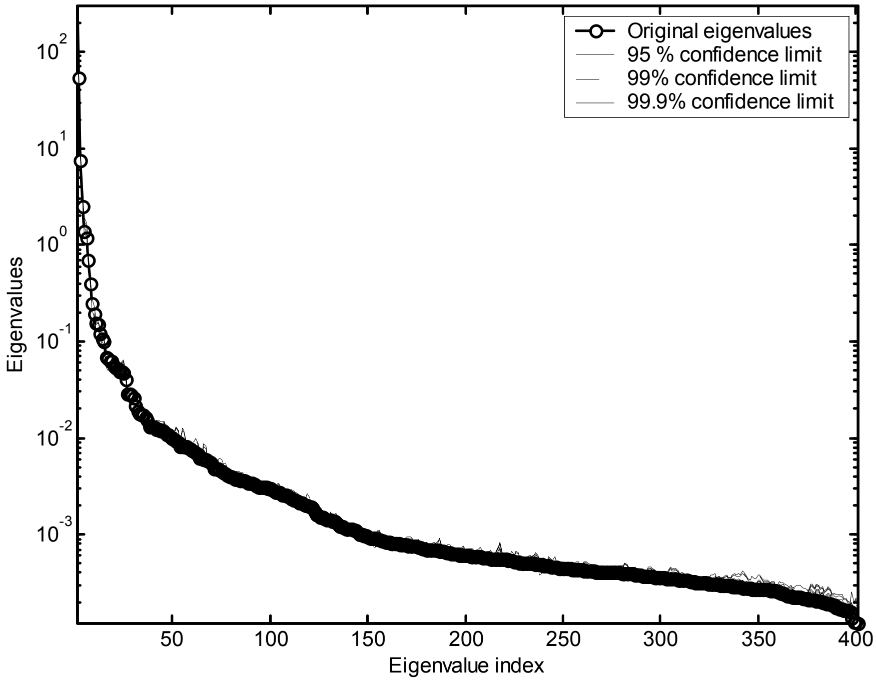


Figure 7.19 Eigenvalue spectrum of cement time series along with confidence limits for the eigenvalues calculated from surrogate data generated by the IAAFT algorithm.

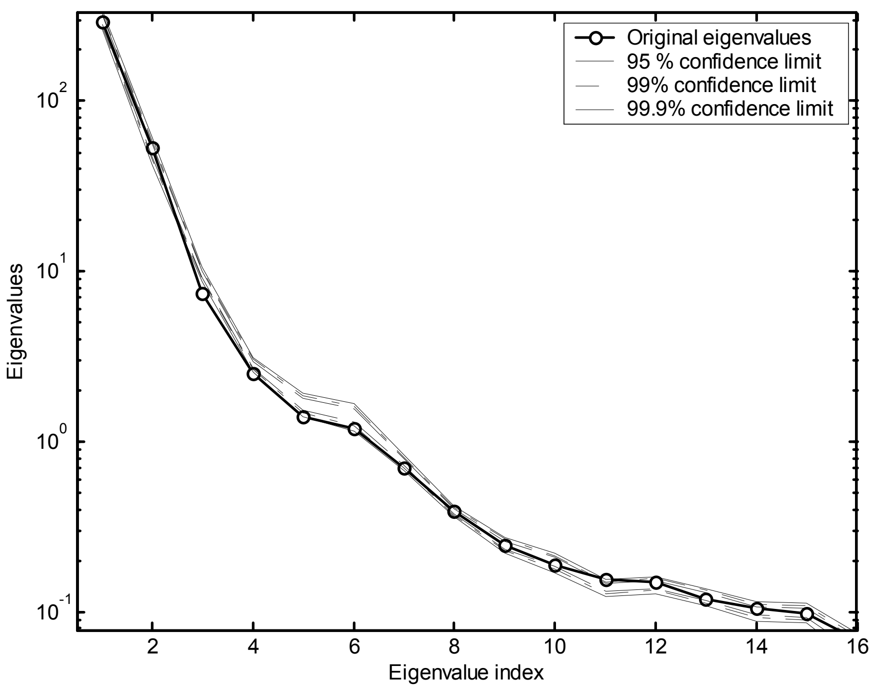


Figure 7.20 Enlarged section of eigenvalue spectrum of cement time series along with confidence limits for the eigenvalues calculated from surrogate data generated by the IAAFT algorithm.

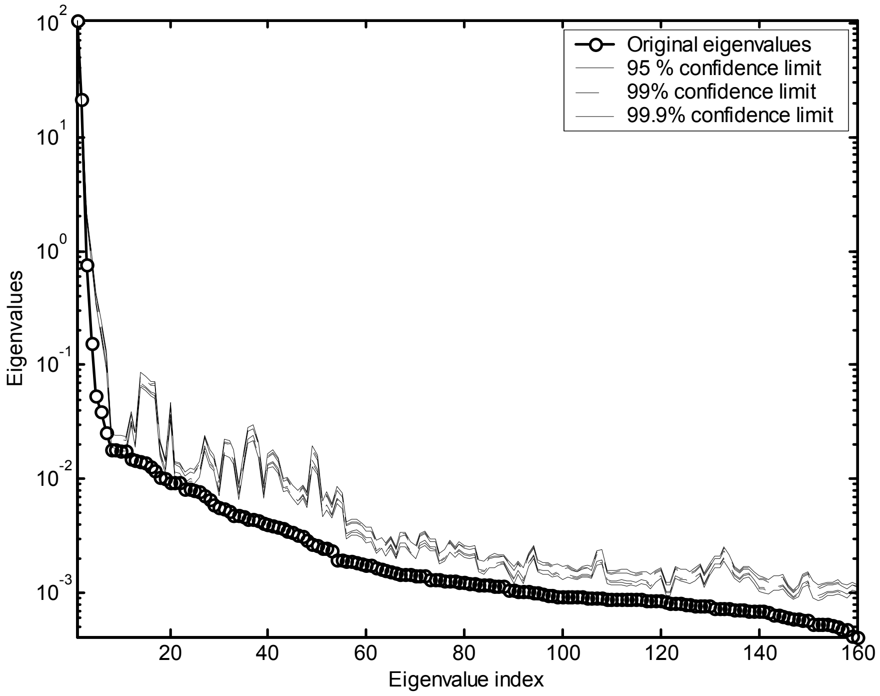


Figure 7.21 Eigenvalue spectrum of cake flour time series along with confidence limits for the eigenvalues calculated from surrogate data generated by the IAAFT algorithm.

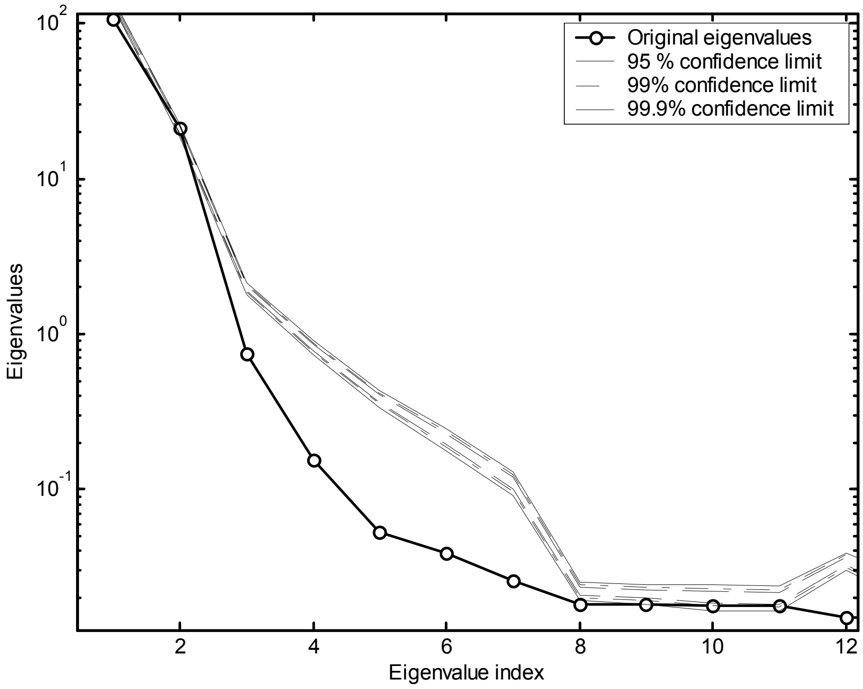


Figure 7.22 Enlarged section of eigenvalue spectrum of cake flour time series along with confidence limits for the eigenvalues calculated from surrogate data generated by the IAAFT algorithm.

When comparing Figure 7.23 and Figure 7.25, it can be seen that the eigenspectra of the two maize-meal time series are very similar, regardless of the size of the plate on which the particles built up before avalanching. Both time series were embedded in relatively large window lengths and the resulting eigenspectra shows a very smooth curve with the percentage of variance explained by each eigenvalue very gradually decreasing. As can be expected, this made it really difficult to determine the cut-off point for the eigenvalues to be retained. However, it was found that only a very small number of these components had to be retained to still explain a significant amount of the variance in the data. For the maize-meal using a large base plate, 24 of the 188 components were retained, explaining 99.1% of the variance, while for the maize-meal with the small base plate 33 of the 267 components was retained and 99.6% of the variance was explained.

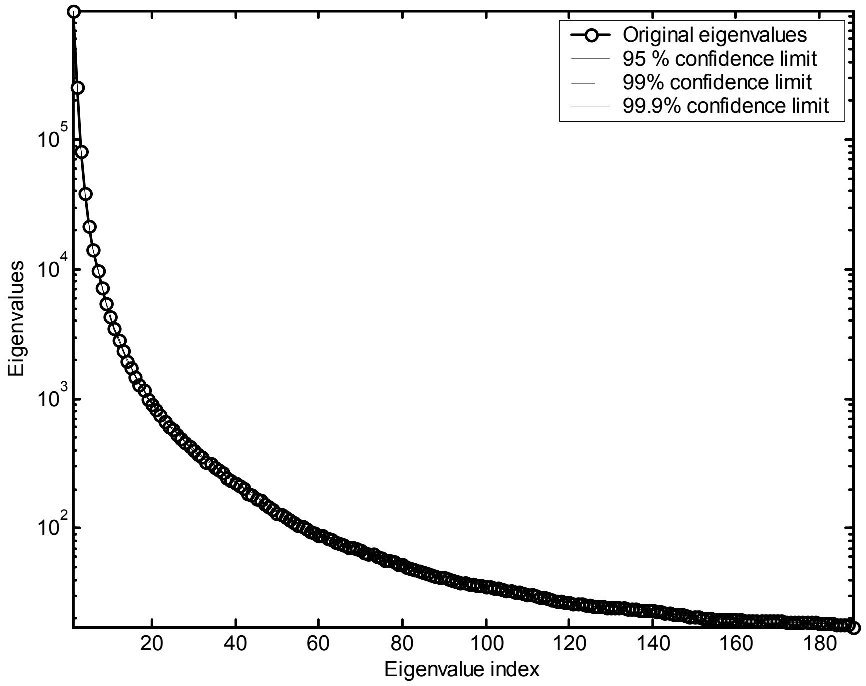


Figure 7.23 Eigenvalue spectrum of maize-meal on large plate time series along with confidence limits for the eigenvalues calculated from surrogate data generated by the IAAFT algorithm.

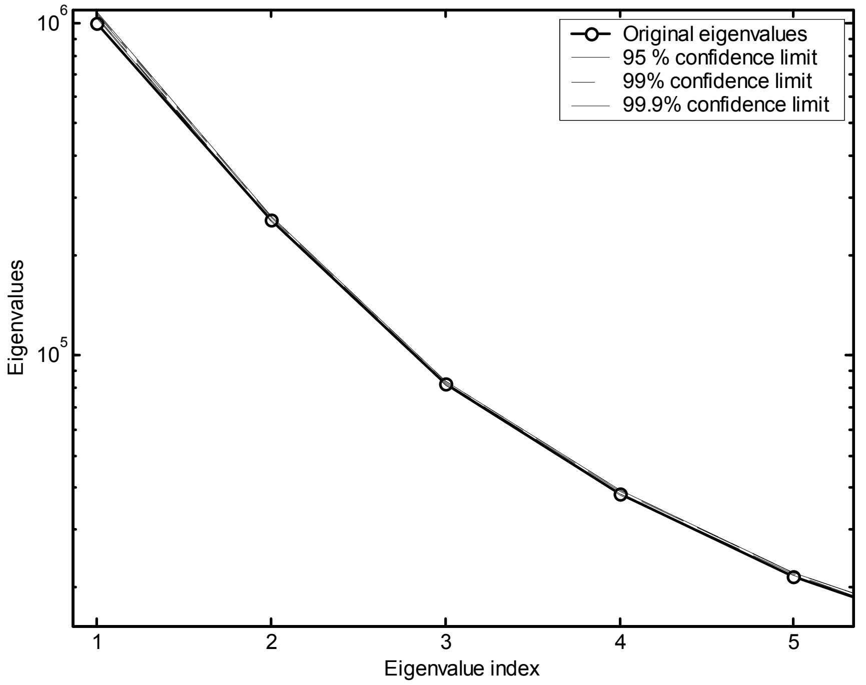


Figure 7.24 Enlarged section of eigenvalue spectrum of maize-meal on large plate time series along with confidence limits for the eigenvalues calculated from surrogate data generated by the IAAFT algorithm.

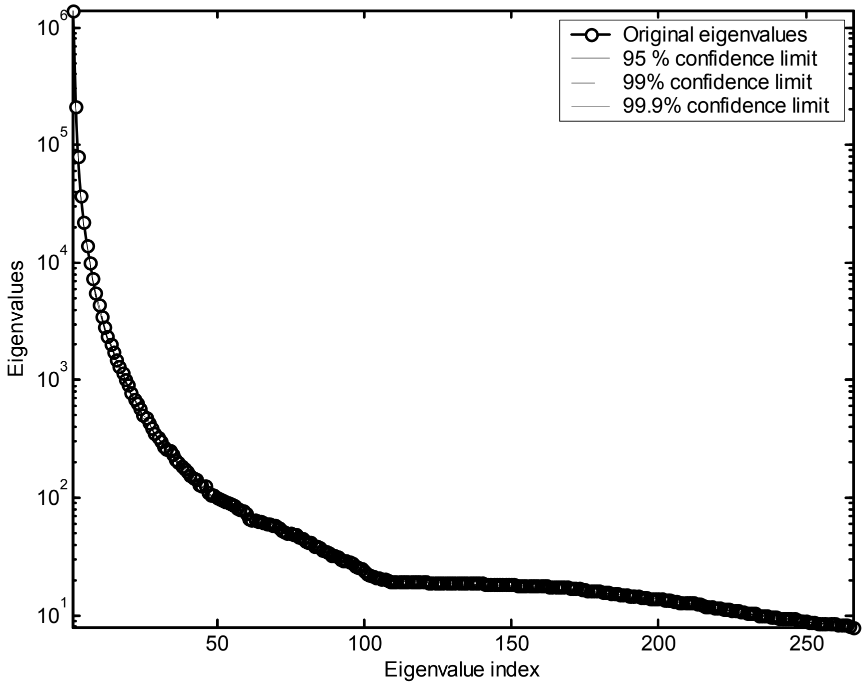


Figure 7.25 Eigenvalue spectrum of maize-meal on small plate time series along with confidence limits for the eigenvalues calculated from surrogate data generated by the IAAFT algorithm.

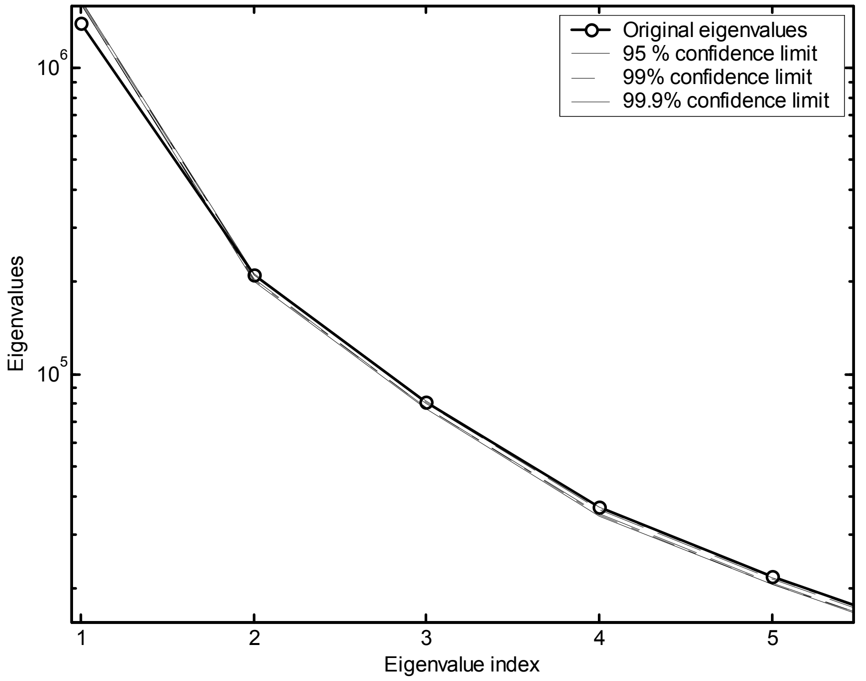


Figure 7.26 Enlarged section of eigenvalue spectrum of maize-meal on small plate time series along with confidence limits for the eigenvalues calculated from surrogate data generated by the IAAFT algorithm.

The possibility was investigated that the eigenspectra of the time series could provide some information as to the flow behaviour of the various types of particles, by relating the nature of the time series to the observed avalanching characteristics. For a time series such as the cake flour that can be observed in Figure 7.21 and Figure 7.22, it can be safely stated that the eigenspectrum falls outside the 99.9% confidence limits and therefore the null hypothesis that the data have been generated by a stationary linear Gaussian system has to be rejected. This would then necessitate further investigation of the time series to determine on which of these grounds (stationary, linear or Gaussian) the null hypothesis was rejected. For all of the case studies being investigated in this section, it can be assumed that the time series is stationary. The basis for this assumption is that the original observations obtained from the experimental set-up have been processed, as described earlier, with the specific purpose of making the data stationary. The two remaining possibilities are therefore that the flow behaviour of cake flour is either non-Gaussian or nonlinear, or both.

The other extreme is represented by the two maize-meal time series, where the eigenspectra fall largely within the 95% confidence band and therefore the null hypothesis fails to be rejected. This means that the avalanching process of maize-meal from both a small and a large plate is possibly a stationary, linear Gaussian process.

The behaviour of the eigenspectrum of the cement time series was very similar to that of the two maize-meal time series, while the eigenspectrum of salt was relatively far outside the confidence limits, which was closer to the behaviour of the cake flour time series, although not as extreme. The shape of the eigenspectrum of the cement also closely corresponds with that of the two maize-meal time series whereas the cake flour and salt time series both seem to have more of a noise floor. These similarities in the behaviour of different time series could unfortunately not be related directly to the visual observations of the flow behaviour of the various systems. Visually it was found that the salt and cement behaved in a similar manner with the cake flour acting more like the maize-meal, although not as likely to ‘build up’ before avalanching.

7.2.5 Investigation of correlation dimension

For this case study, the correlation dimension curves of the time series were also investigated in an effort to obtain more information on the behaviour of the time series and to obtain another criterion by which to compare the behaviour of the various particle systems.

Figure 7.27 illustrated the correlation dimension of the five time series presented in Figure 7.11, along with the correlation dimension of 15 surrogate data sets generated by the IAAFT algorithm. These figures can be used in a similar manner as those in Figure 7.17 to Figure 7.26, where a selected output from the time series under inspection is tested against a null hypothesis that the time series were generated by a stationary linear Gaussian process.

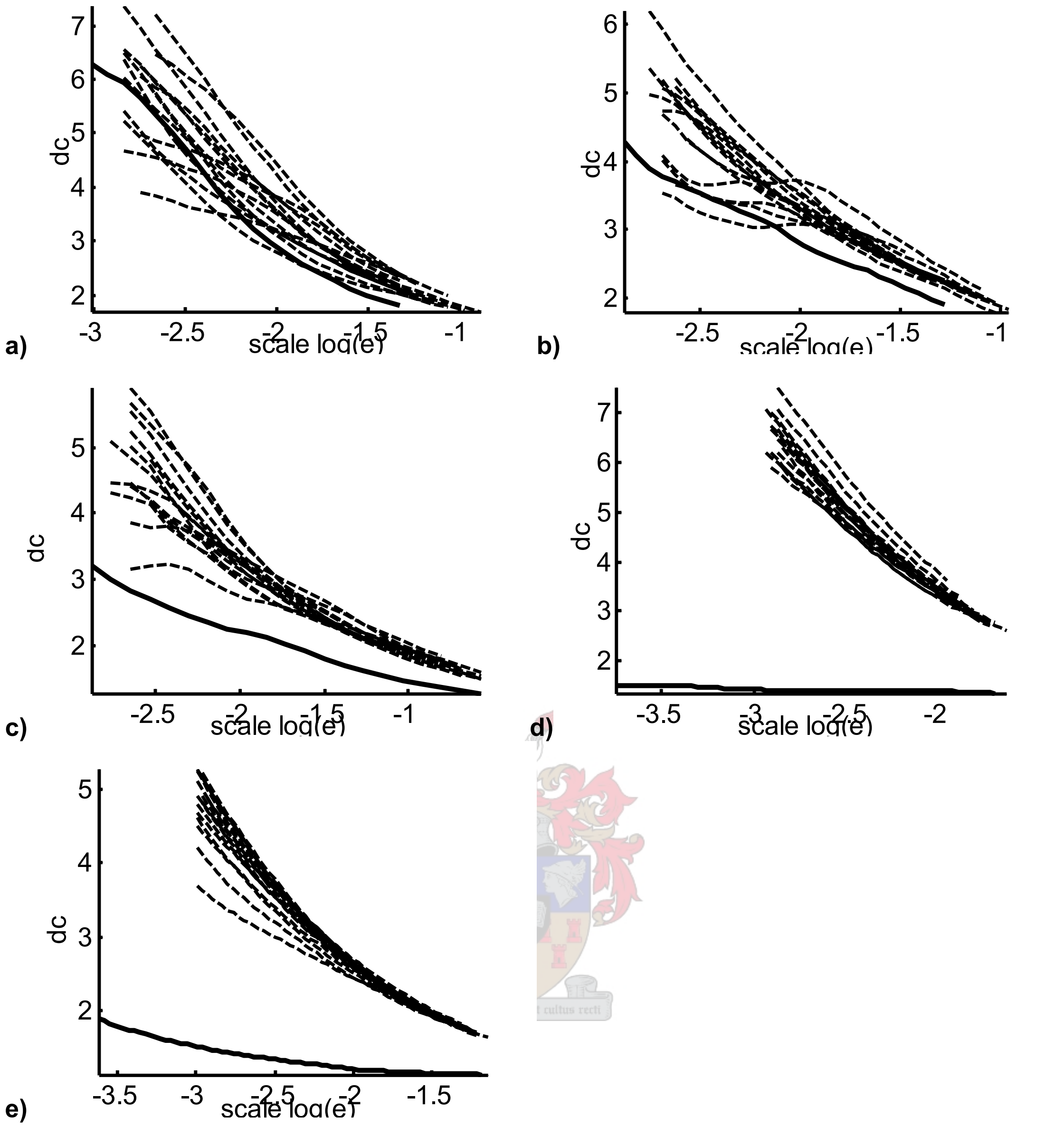


Figure 7.27 Correlation dimension of a) Salt, b) Cement, c) Cake flour, d) Maize-meal on large disc and e) maize-meal on small disc time series, as well as the correlation dimension of 15 surrogate data sets generated by IAAFT algorithm.

Unlike the results obtained for the eigenspectra in section 7.2.3, the appearance of the correlation dimensions in Figure 7.27 corresponds more accurately to the visual observations during the experiment. The correlation dimension curves of the salt and cement with their more free-flowing natures are practically indistinguishable from that of the linear, stochastic, Gaussian surrogates, while the correlation dimension curve of the maize meal, which flowed significantly less regularly, is quite far removed from that of the surrogates.

Included in Figure 7.28 and Figure 7.29 respectively, are also the correlation dimensions from the time series obtained from the temperature and humidity tests, as have been described in section 7.2.2, presented in a similar fashion as the other time series. The correlation dimension curves for these series will be discussed, even though the eigenvalue spectra have not been shown. The eigenvalue spectra have been omitted because the difference in the behaviour of the same particle system under different operating conditions can be seen quite clearly from these correlation dimension curves.

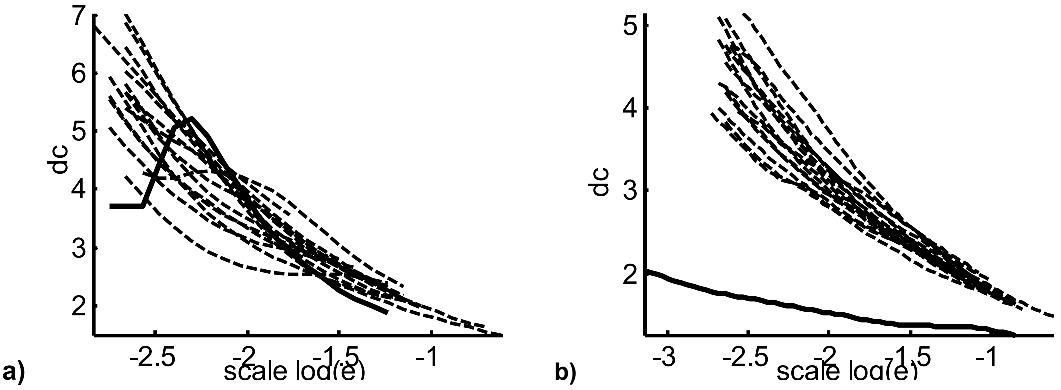


Figure 7.28 Correlation dimension of a) cold maize-meal and b) warm maize-meal, as well as the correlation dimension of 15 surrogate data sets generated by IAAFT algorithm.

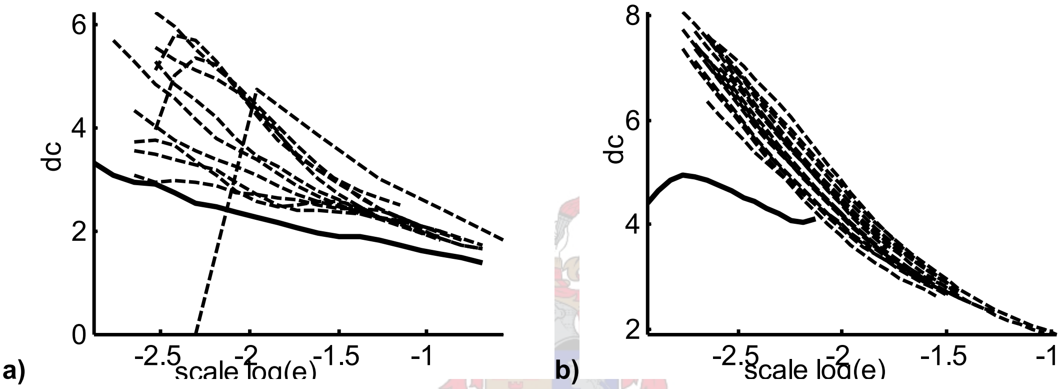


Figure 7.29 Correlation dimension of a) wet sand and b) dry sand, as well as the correlation dimension of 15 surrogate data sets generated by IAAFT algorithm.

It is seen from Figure 7.28 that the correlation dimension of cold maize meal behaves much more like that of a more free-flowing particle, such as cement or salt, than it does to that of normal maize meal. The relationship of the correlation dimension of warm maize meal, on the other hand, corresponds more with that of the normal maize meal correlation curve, both in respect of the shape and values of the curve and the position of the curve relative to that of the surrogate data sets. The difference between the flow behaviour of wet and dry sand is not as pronounced as that between warm and cold maize meal (Figure 7.29).

A further test was done where the slope of the correlation dimension curves of all the time series in Figure 7.27 to Figure 7.29 were evaluated. This evaluation of the slope was done by taking the value of the correlation dimension at a low value for *e* (small distance between the points) and a high value for *e* (large distance between the points) and plotting these values for each time series as a function of each other (Figure 7.30). In order to be able to make a comparison, the high and low values for *e* for all the time series had to be the same values. Due to the differing natures of the time series, this resulted in a relatively short span of the correlation dimension curves over which all the curves overlapped, making the results in Figure 7.30 slightly less reliable than one would desire.

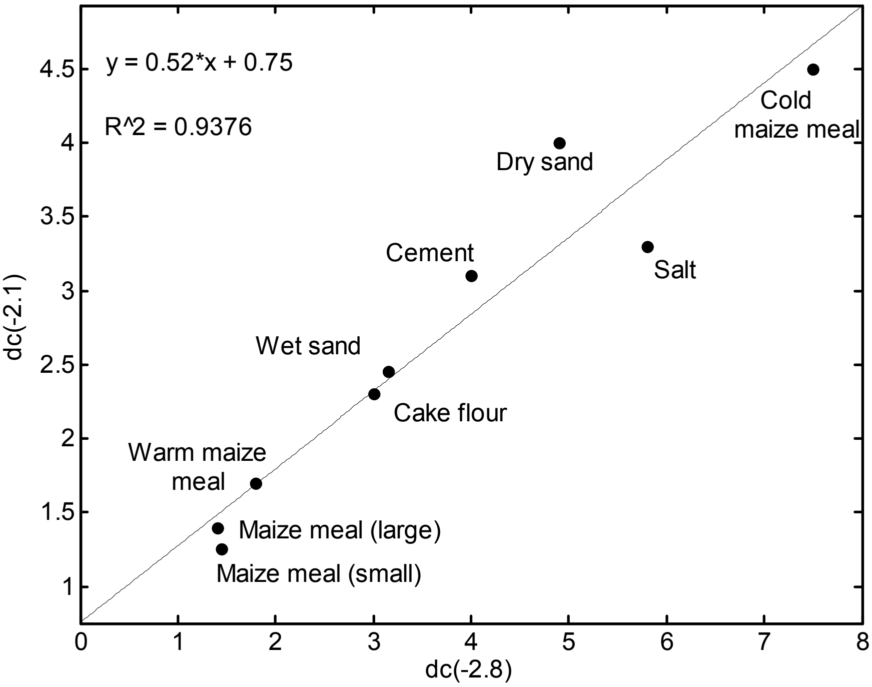


Figure 7.30 Comparison of slope of correlation dimension curves for different particles by plotting high and low measurements.

However, even with the short sections of the curves for which the slopes were calculated, the distribution of the different types of particles in Figure 7.30 still correlates well with the visual observations made regarding the flow behaviour of the particles during the experiments. It is interesting to note the large difference between the flow of warm maize meal and cold maize meal, both in the slope of their correlation dimensions and the visual observations.
Figure 7.30 is probably the best representation of the actual flow behaviour of the particles that were observed.

7.2.6 Reconstructed attractors

The final set of information about the different particle systems can be obtained from a visual inspection of their reconstructed attractors in Figure 7.31 to Figure 7.35. As it has been mentioned in a previous section (section 7.1.2), the appearance of the attractor can provide valuable information about the nature of the time series from which the components have been extracted, such as the periodicity in the time series, how deterministic the series is and the level of noise in the series.

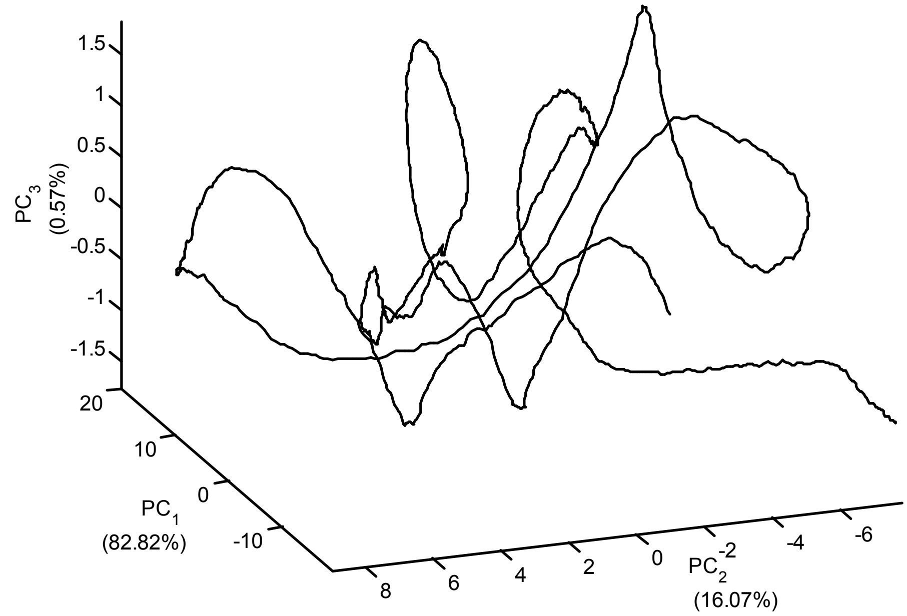


Figure 7.31 Reconstructed attractor for cake flour time series with the amount of variance explained by each principal component supplied in brackets next to the appropriate axis.

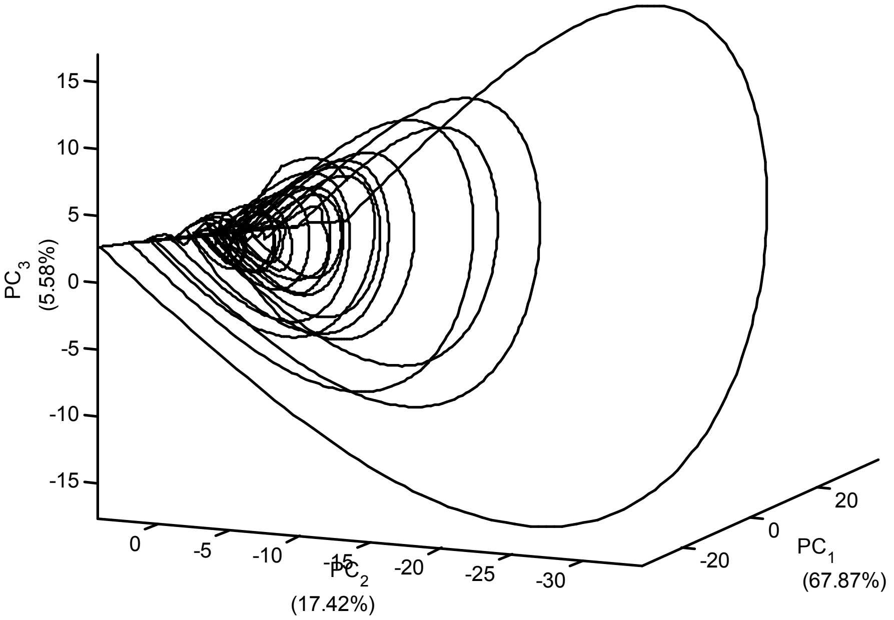


Figure 7.32 Reconstructed attractor for maize meal on a large plate series with the amount of variance explained by each principal component supplied in brackets next to the appropriate axis.

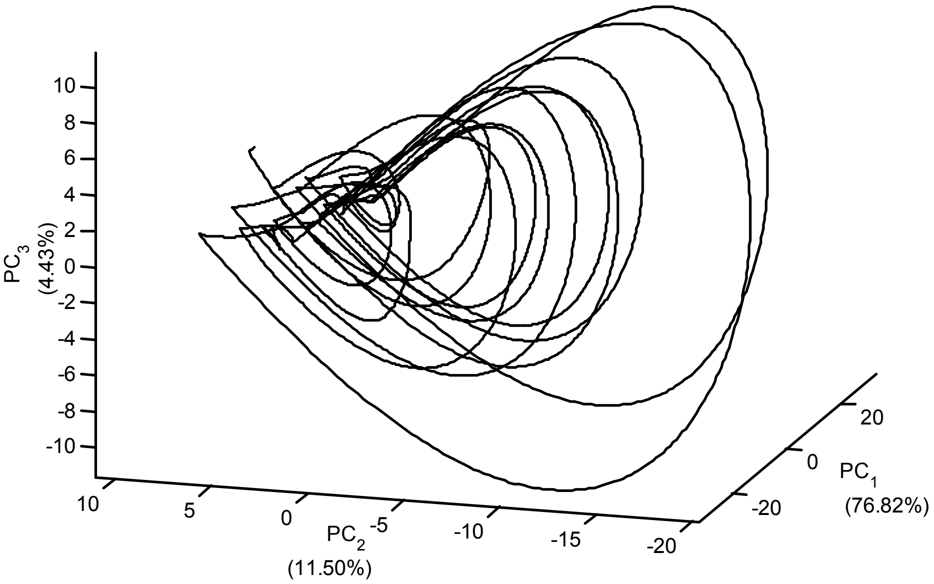


Figure 7.33 Reconstructed attractor for maize meal on a small plate series with the amount of variance explained by each principal component supplied in brackets next to the appropriate axis.

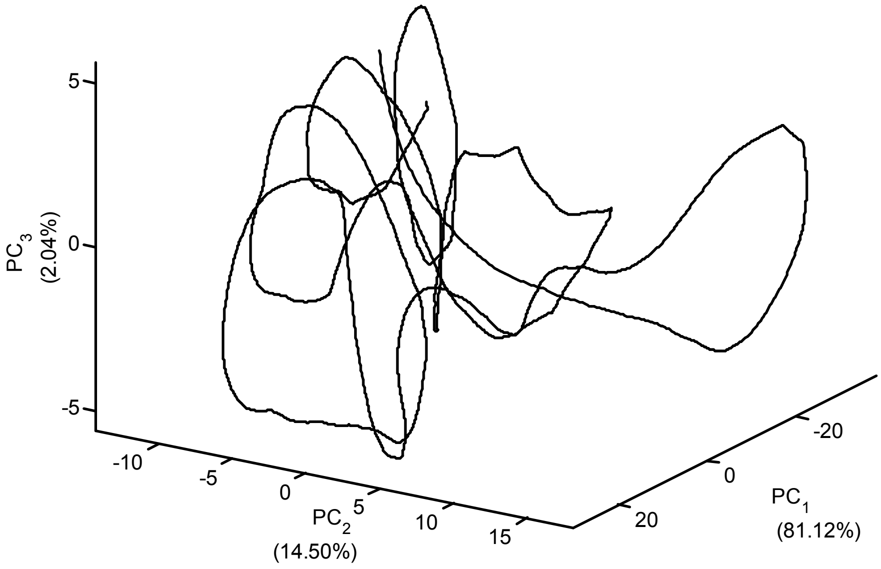


Figure 7.34 Reconstructed attractor for tiling cement time series with the amount of variance explained by each principal component supplied in brackets next to the appropriate axis.

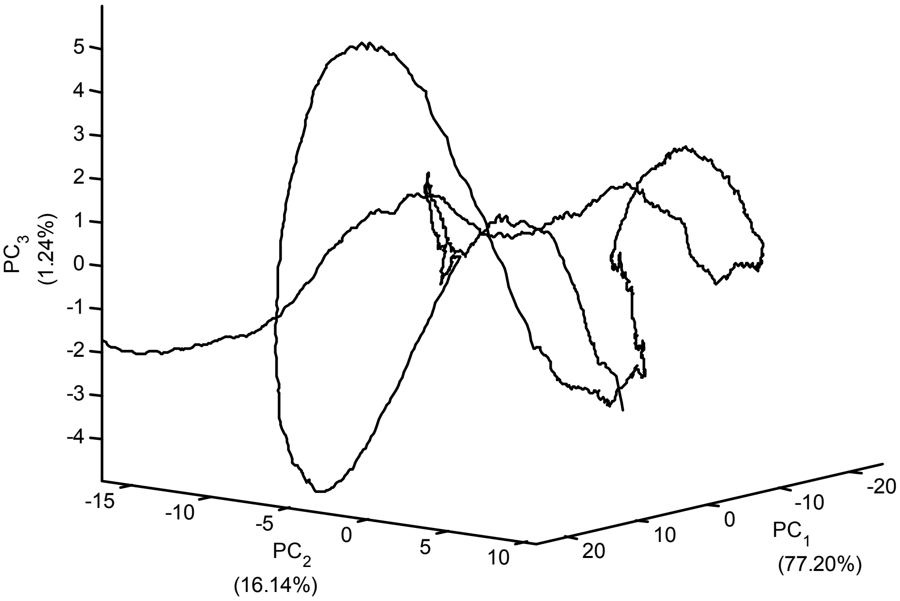


Figure 7.35 Reconstructed attractor for salt time series with the amount of variance explained by each principal component supplied in brackets next to the appropriate axis.

Once again the difference in the flow behaviours can be seen very clearly from the various attractors. The more deterministic series with a more restricted flow pattern, such as the two maize meal series (Figure 7.32and Figure 7.33), exhibit significantly more regularity in their attractors than those of the free-flowing, stochastic series, such as salt (Figure 7.35). It is also interesting to note the similarity in the shapes of the maize meal attractors from the small and the large base plates. This once again indicates the lack of sensitivity of the particles for the size of the surface they avalanche from.

7.3 Summary

Monte Carlo SSA has been applied to two case studies of real process data in an effort to classify the processes generating the data. The Monte Carlo SSA was combined with a number of other criteria, such as the reconstructed attractors, score plots and the quantile-quantile plots of the data.

For a discrete time series of finite length, singular spectrum analysis makes use of the principal component decomposition of an estimate of the correlation matrix that is based on m lagged copies of the time series, which forms the trajectory matrix of the time series. The resultant eigenvectors form an optimal basis that is orthonormal at zero lag and permit the signal to be decomposed into its possibly oscillatory and aperiodic components.

The eigenspectrum associated with these eigenvectors is a generalized statistic that characterizes the nature of the time series and can be used to discriminate between time series or time series components on the basis of stochasticity/determinism, linearity/nonlinearity, etc. In this and the previous chapter, the eigenspectrum as a whole was considered in the characterization of the time series. This is a very stringent approach, as the eigenspectrum consists of a ranked series of eigenvalues and more sophisticated analysis is possible by testing the individual eigenvalues. For example, the first few eigenvalues of the time series is usually associated with the trend or major variation in the time series, while eigenvalues with higher indices are associated with the fine structure (noise) of the time series. By requiring the null hypothesis to be valid for all components of the time series, it may mean that the time series is classified on the basis of the nature of its least significant components as well.

Note however, that distinguishing between significant and insignificant components is not a trivial matter. Moreover, the interpretation of the statistical tests should be done very carefully, to account for possible biases in the tests and flawed surrogate data. Despite these caveats, the general approach outlines in these chapters constitutes a promising route towards the classification and analysis of time series, and ultimately better system identification, process control and optimisation of plant operations through better data analysis.

8 CONCLUSIONS

Singular spectrum analysis appears to be a useful tool to analyse data generated by dynamic process systems, since it can be used to decompose time series data into different components.

- It was found in the literature survey that singular spectrum analysis as a technique has already attracted a vast amount of attention from a multitude of research disciplines. These include applications in the physical natural sciences, such as climatology, the biosciences, solar physics and geology, as well as some work in the economics and general engineering fields. However, it was found that the application of SSA to chemical and metallurgical engineering processes, has been a great oversight, in that no references could be found.
- The advantage of using SSA to perform the filtering of data before the data are modelled by using neural networks has been investigated in chapter 4. In all the case studies considered, the models built on the data after SSA was applied, outperformed the models that were built on the time series alone, indicating that SSA has great value in filtering data prior to modelling. It was also found for the carbon-in-leach process that a reconstruction of the time series with only three components outperformed an optimised moving average filter in terms of the series' correlation with the original clean signal.

  A secondary investigation also proved the importance of choosing a long enough time series window when performing the filtering. This will prevent the loss of some of the information from the time series during the embedding and subsequent decomposition. If the window length is too short, valuable information could still be contained in the tail eigenvalues, which are usually regarded as irrelevant.

  When the aspect of the filtering of data contaminated with red noise vs. white noise, was investigated with SSA, the results obtained from the case study did not illustrate the problem as clearly, due to problems with the signal to noise ratio. However, a comprehensive discussion from the literature was supplied to serve as a solution to the inadequacy of SSA to handle red noise satisfactorily.
- During the investigation of the different methods by which to perform nonlinear SSA, attention was given to localized SSA and auto-associative neural networks. Although both techniques succeeded in identifying nonlinear components from the data that basic SSA could not extract, it would be recommended rather to use localized SSA than auto-associative neural networks. Localized SSA could easily be applied to the case studies and gave satisfactory results, while many problems were experienced with the implementation of auto-associative neural networks, despite the apparent advantages from the literature.
- Monte Carlo SSA was first applied to a number of time series with known characteristics, in order to serve as an indication of the reliability of Monte Carlo SSA to correctly classify and identify data series. The time series were being tested against two different hypotheses, the one was that the data were from a first-order autoregressive process and the other that a linear, stochastic, Gaussian process generated the data. During these benchmarking tests, it was once again found that no single test could give reliable results for all the time series that is characterized by it. However, Monte Carlo SSA did succeed in correctly classifying a number of time series that were incorrectly classified by techniques such as the surrogate analysis of the correlation dimension curves.

  When the Monte Carlo SSA technique was applied to real data series, a number of other criteria, such as the reconstructed attractors, score plots and the quantile-quantile plots of the data, were used in conjunction with Monte Carlo SSA to compensate for the shortages found in the previous chapter. When the eigenspectrum as a whole is considered in the characterization of the time series, it presents a very stringent approach, as the eigenspectrum consists of a ranked series of eigenvalues and more sophisticated analysis would possible by testing the individual eigenvalues. By requiring the null hypothesis to be valid for all components of the time series, it may mean that the time series is classified on the basis of the nature of its least significant components as well.

Note however, that distinguishing between significant and insignificant components is not a trivial matter. Moreover, the interpretation of the statistical tests should be done very carefully, to account for possible biases in the tests and flawed surrogate data. Despite these caveats, the general approach outlines in these chapters constitutes a promising route towards the classification and analysis of time series, and ultimately better system identification, process control and optimisation of plant operations through better data analysis.

REFERENCES

Abarbanel, H. D. I. (1996) *Analysis of Observed Chaotic Data,* Springer, New York.

Addison, P. S. (1997) *Fractals and chaos: An illustrated course,* Institute of Physics Publishing, Bristol, Philadelphia.

Akaike, H. (1974) A new look at statistical model identification *IEEE Transactions on Automatic Control,* **19,** (6) 716-723.

Aldrich, C. (2002) *Exploratory analysis of metallurgical process data with neural networks and related methods,* Elsevier, Amsterdam, Boston, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo.

Allen, M. R., Read, P. L. and Smith, L. A. (1992a) Temperature oscillations *Nature,* **359,** 679.

Allen, M. R., Read, P. L. and Smith, L. A. (1992b) Temperature time-series? *Nature,* **355,** 686.

Allen, M. R. and Robertson, A. W. (1996) Distinguishing modulated oscillations from coloured noise in multivariate datasets *Climate dynamics,* **12,** 775-784.

Allen, M. R. and Smith, L. A. (1994) Investigating the origins and significance of low-frequency modes of climate variability *1994,* **21,** (10) 883-886.

Allen, M. R. and Smith, L. A. (1996) Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise *Journal of Climate,* **9,** 3373 - 3404.

Allen, M. R. and Smith, L. A. (1997) Optimal filtering in singular spectrum analysis *Physics Letters A,* **234,** (6) 419-428.

Arzner, K. (2002) Time-domain demodulation of RHESSI light curves *Solar Physics,* **210,** 213-227.

Baratta, D., Cicioni, G., Masulli, F. and Studer, L. (2003) Application of an ensemble technique based on singular spectrum analysis to daily rainfall forecasting *Neural Networks,* **16,** 375 - 387.

Barnard, G. A. (1963) Discussion on: The spectral analysis of point processes (by M S Bartlett) *Journal of Royal Statistical Society,* **25,** 294.

Barnard, J. P., Aldrich, C. and Gerber, M. (2001) Identification of dynamic process models with surrogate data *AIChE Journal,* **47,** (9) 2064 - 2075.

Barnett, W.A., Gallant, A.R., Hinich, M.J., Jungeilges, J.A., Kaplan, D.T., Jensen, M.J. (1997) A single-blind controlled competition among tests for nonlinearity and chaos. *Journal of Econometrics*, **82**, 157-192.

Benzi, R., Deidda, R. and Marrocu, M. (1997) Characterization of temperature and precipitation fields over Sardinia with principal component analysis and singular spectrum analysis *International Journal of Climatology,* **17,** (11) 1231-1262.

Besag, J. and Diggle, P. J. (1977) Simple Monte-Carlo tests for spatial pattern *Applied Statistics,* **26,** 327-333.

Bhardwaj, S. K. and Tangarajan, G. K. (1998) A model for solar quiet day variation at low latitude from past observations using singular spectrum analysis *Proceedings of the Indian Academy of Sciences - Earth and Planetary Sciences,* **107,** (3) 217-224.

Bilato, R., Marrelli, L., Martin, P., Franz, P., Spizzo, G., Murari, A. and Zabeo, L. (2000) Time series statistical analysis for electron temperature fluctuations measurements in plasmas In *Conference on Control Fusion and Plasma Physics*, Vol. 24B Budapest, pp. 1376-1379.

Broomhead, D. S. and King, G. P. (1986) Extracting qualitative dynamics from experimental data *Physica D,* **20,** 217-236.

Castagnoli, G. C., Bernasconi, S. M., Bonino, G., Monica, P. D. and Taricco, C. (1999a) 700 year record of the 11 year solar cycle by planktonic foraminifera of a shallow water Mediterranean core *Advanced Space Research,* **24,** (2) 233-236.

Castagnoli, G. C., Bonino, G., Monica, P. D., Taricco, C. and Bernasconi, S. M. (1999b) Solar activity in the last millennium recorded in the δ^{18}O profile of planktonic foraminifera of a shallow water Ionian sea core *Solar Physics,* **188,** 191-202.

Castagnoli, G. C., Bonino, G. and Taricco, C. (2002a) Long term solar-terrestrial records from sediments: carbon isotopes in planktonic foraminifera during the last minimum *Advanced Space Research,* **29,** (10) 1537 - 1549.

Castagnoli, G. C., Bonino, G., Taricco, C. and Bernasconi, S. M. (2002b) Solar radiation variability in the last 1400 years recorded in the carbon isotope ratio of a Mediterranean sea core *Advanced Space Research,* **29,** (12) 1989-1994.

Celka, P. and Colditz, P. (2002) A computer-aided detection of EEG seizures in infants: A singular-spectrum approach and performance comparison *IEEE Transactions on Biomedical Engineering,* **49,** (5) 455-462.

Chang, K.-y. and Ghosh, J. (2001) A unified model for probabilistic principal surfaces *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **23,** (1) 22-41.

Chiou, D.-C., Huang, H.-H., Chan, H.-L. and Wu, C.-P. (2000) Extraction of 1/f component from heartbeat interval signal by singular spectrum analysis *IEICE Transactions on Information and Systems,* **E38D,** (2) 302-304.

Choi, B.-Y., Park, J. and Zhang, Z.-L. (2002) Adaptive random sampling for load change detection In *SIGMETRICS 2002, International Conference on Measurements and Modelling of Computer Systems* California.

Cook, E. R., D'Arrigo, R. D. and Briffa, K. R. (1998) A reconstruction of the North Atlantic Oscillation using tree-ring chronologies from North-America and Europe *The Holocene,* **8,** (1) 9-17.

Corte-Real, J., Qian, B. and Xu, H. (1998) Regional climate change in Portugal: precipitation variability associated with large-scale atmospheric circulation *International journal of climatology,* **18,** 619-635.

Corte-Real, J., Wang, X. and Zhang, X. (1995) Modes of variability in the Northern Hemisphere's mid-tropospheric large-scale circulation *Theoretical and applied climatology,* **50,** 133-146.

Cortijo, E., Yiou, P., Labeyrie, L. and Cremer, M. (1995) Sedimentary record of rapid climatic variability in the North-Atlantic ocean during the last glacial cycle *Paleoceanography,* **10,** (5) 911-926.

D'Arrigo, R., Buckley, B., Kaplan, S. and Woollett, J. (2003) Interannual to multidecadal modes of Labrador climate variability inferred from tree rings *Climate Dynamics,* **20,** 219-228.

Dean, W., Anderson, R., Bradbury, J. P. and Anderson, D. (2002) A 1500-year record of climatic and environmental change in Elk Lake, Minnesota - I: Varve thickness and grey-scale density *Journal of Paleolimnology,* **27,** (3) 287-299.

Dettinger, M. D., Ghil, M. and Keppenne, C. L. (1995) Interannual and interdecadal variability in United States surface-air temperatures, 1910-87 *Climatic Change,* **31,** 35-66.

De Wet, M. (2001) Monitoring of localized corrosion phenomena by use of electrochemical noise measurements: Design of experimental set-up. *Final Year Laboratory Project Report*, Stellenbosch University, Stellenbosch

Dickey, J. O., Gegout, P. and Marcus, S. L. (1999) Earth-atmosphere angular momentum exchange and ENSO: The rotational signature of the 1997-98 event *Geophysical Research Letters,* **26,** (16) 2477-2480.

Elsner, J. B., Kara, A. B. and Owens, M. A. (1999) Fluctuations in the North Atlantic hurricane frequency *Journal of Climate,* **12,** (2) 427-437.

Elsner, J. B. and Tsonis, A. A. (1991) Do bidecadal oscillations exist in the global temperature record? *Nature,* **353,** (6344) 551-553.

Elsner, J. B. and Tsonis, A. A. (1996) *Singular Spectrum Analysis. A New Tool in Time Series Analysis,* Plenum Press, New York, London.

Evans, M. N., Fairbanks, R. G. and Rubenstone, J. L. (1999) The thermal oceanographic signal of El Nino reconstructed from a Kiritimati Island coral *Journal of Geophysical Research - Oceans,* **104,** (C6) 13409-13421.

Fisher, N. I. and Hall, P. (1990) On bootstrap hypotheses testing *Australian Journal of Statistics,* **32,** 177-190.

Fraedrich, K. (1986) Estimating the dimension of weather and climate attractors *Journal of Atmospheric Science,* **43,** (5) 419-432.

Gallego, M. C., Garcia, J. A., Vaquero, J. M. and Sanchez Bajo, F. (1999) A strange attractor on a daily timescale in the visual observations of Z and ? *Acta Astronomica,* **49,** 171-180.

Gedalof, Z., Mantua, N. J. and Peterson, D. L. (2002) A multi-century perspective of variability in the Pacific Decadal Oscillation: new insights from tree rings and coral *Geophysical Research Letters,* **29,** (24) 57-1 - 57-4.

Ghil, M. and Taricco, C. (1997) Advanced spectral analysis methods In *Past and Present Variability of the Solar-Terrestrial System: Measurement, Data Analysis and Theoretical Models*(Eds, Castagnoli, G. C. and Provenzale, A.) Societa Italiana di Fisica, IOS Press, Bologna, Amsterdam.

Ghil, M. and Vautard, R. (1991) Interdecadal oscillations and the warming trend in global temperature time series *Nature,* **350,** 324-327.

Ghil, M. and Yiou, P. (1996) Spectral methods: What they can and cannot do for climatic time series In *Decadal Climate Variability: Dynamics and Predictability*(Eds, Anderson, D. and Willebrand, J.) Elsevier, Amsterdam, pp. 445-482.

Ghirardelli, J. E., Rienecker, M. M. and Adamec, D. (1995) Meridional Ekman heat transport: Estimates from satellite data *Journal of Physical Oceanography,* **25,** 2741-2755.

Golia, S. and Sandri, M. (2001) A resampling algorithm for chaotic time series *Statistics and Computing,* **11,** 241-255.

Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001) *Analysis of Time Series Structure - SSA and related techniques,* Chapman & Hall, CRC, Boca Raton, London, New York, Washington DC.

Gray, P. and Scott, S. K. (1983) Autocatalytic reactions in the isothermal, continuous stirred tank reactor. Isolas and other forms of multi-stability *Chemical Engineering Science,* **38,** 29-43.

Gray, P. and Scott, S. K. (1984) Autocatalytic reactions in the isothermal, continuous stirred tank reactor. Oscillations and instabilities in the system A + 2B -> 3B, B-> C *Chemical Engineering Science,* **39,** 1087-1097.

Grinsted, A., Flather, R., Jevrejeva, S., Moore, J. C., Wakelin, S., Williams, J. and Woodworth, P. (2003) Influence of large-scale atmospheric circulation on the European sea level: Results from singular spectrum analysis and wavelet transform *Geophysical Research Abstracts,* **5**.

Hall, P. and Titterington, D. M. (1989) The effect of simulation order on level accuracy and power of Monte-Carlo tests *Journal of Royal Statistical Society B,* **51,** 459-467.

Hassanpour, H., Mesbah, M. and Boashash, B. (2003) Comparative performance of time-frequency based newborn EEG seizure detection using spike signatures In *ICASSP*IEEE, pp. 389-392.

Held, G. A., Solina, D. H., Keane, D. T., Haag, W. J., Horn, P. M. and Grinstein, G. (1990) Experimental study of critical mass fluctuations in an evolving sandpile. *Physical Review Letters,* **65,** 1120-1123.

Hollingsworth, J. L., Haberle, R. M. and Schaeffer, J. (1997) Seasonal variations of storm zones on Mars *Advanced Space Research,* **19,** (8) 1237-1240.

Hope, A. C. A. (1968) A simplified Monte-Carlo significance test procedure. *Journal of the Royal Statistical Society B,* **30,** 582-598.

Hsieh, W. W. (2001) Nonlinear principal component analysis *Tellus,* **53A,** 599-615.

Hsieh, W. W. and Hamilton, K. (2003) Nonlinear singular spectrum analysis of the tropical stratospheric wind *Quarterly Journal Review of the Meteorological Society,* **129**.

Hsieh, W. W. and Wu, A. (2001a) Nonlinear multichannel singular spectrum analysis of the tropical Pacific climate variability using a neural network approach *Journal of Geophysical Research (Oceans),* **107,** (C7).

Hsieh, W. W. and Wu, A. (2001b) Nonlinear singular spectrum analysis by neural networks *Neural Networks*.

Hsieh, W. W. and Wu, A. (2002) Nonlinear singular spectrum analysis In *International Joint Conference on Neural Networks '02*, Vol. 3, pp. 2819-2824.

Juckett, D. A. (2001) Period and phase comparisons of near-decadal oscillations in solar, geomagnetic, and cosmic ray time series *Journal of Geophysical Research - Space Physics,* **2001,** (106) A9.

Kepenne, C. L. (1995) An ENSO signal in soybean futures prices *Journal of Climate,* **8,** 1685-1689.

Kepenne, C. L. and Ghil, M. (1993) Adaptive filtering and prediction of noisy multivariate signals: an application to subannual variability in atmospheric angular momentum *International Journal of Bifurcation and Chaos,* **3,** (3) 625-634.

Khramova, M. N., Kononovich, E. V. and Krasotkin, S. A. (2002) Quasi-biennial oscillations of global solar-activity indices *Solar System Research,* **36,** (6) 507-512.

Kirby, M. J. and Miranda, R. (1996) Circular nodes in neural networks *Neural Computations,* **8,** 390-402.

Kramer, M. A. (1991) Nonlinear principal component analysis using auto-associative neural networks *AIChE Journal,* **37,** 223-243.

Krepper, C. M., Garcia, N. O. and Jones, P. D. (2003) Interannual variability in Uruguay river basin *International Journal of Climatology,* **23,** (1) 103-115.

Lall, U. and Mann, M. (1995) The Great Salt Lake: A barometer of low-frequency climatic variability *Water Resources Research,* **31,** (10) 2503-2515.

Lee, Y.-A. (2001) A T-EOF based prediction method *Journal of Climate,* **15,** (2) 226 - 234.

Lee, Y.-A. and Hang, L.-J. (2000) A possible volcanic signal in NCEP/NCAR reanalysis data *TAO,* **11,** (4) 895-908.

Lisi, F. and Medio, A. (1997) Is random walk the best exchange rate predictor? *International Journal of Forecasting,* **13,** 255-267.

Lisi, F., Nicolis, O. and Sandri, M. (1995) Combining singular-spectrum analysis and neural networks for time series forecasting *Neural Processing Letters,* **2,** (4) 6-10.

Liu, Y.-f. and Zhao, M. (2003) Detection of rotor cracks based on multi-scale singular spectrum analysis In *Damage Assessment of Structures, Proceedings. Key Engineering Materials*, Vol. 245-246, pp. 273-278.

Lynch, D. T. (1992) Chaotic behaviour of reaction systems: parallel cubic autocatalators. *Chemical Engineering Science,* **47,** (2) 347-355.

Marrelli, L., Bilato, R., Franz, P., Martin, P., Murari, A. and O'Gorman, M. (2001) Singular spectrum analysis as a tool for plasma fluctuations analysis *Review of scientific instruments,* **72,** (1) 499-502.

Masulli, F., Baratta, D., Cicioni, G. and Studer, L. (2001) Daily rainfall forecasting using an ensemble technique based on singular spectrum analysis In *International Joint Conference on Neural Networks 01*, Vol. 1 Piscataway, NJ, USA, pp. 263-268.

Melice, J. L. and Rucou, P. (1998) Decadal time scale variability recorded in the Quelccaya summit ice core delta O-18 isotopic ration series and its relation with the sea surface temperature *Climate Dynamics,* **14,** (2) 117-132.

Mineva, A. and Popivanov, D. (1996) Method for single-trial readiness potential identification, based on singular spectrum analysis *Journal of Neuroscience Methods,* **68,** (1) 91-99.

Mo, K. (2001) Adaptive filtering and prediction of intraseasonal oscillations *Monthly Weather Review,* **129,** (4) 802-817.

Mo, K. C. (1999) Alternating wet and dry episodes over California and intraseasonal oscillations *Monthly Weather Review,* **127,** 2759-2776.

Mo, K. C. (2000) Intraseasonal modulation of summer precipitation over North America *Monthly Weather Review,* **128,** (5) 1490-1505.

Molteni, F., Tibaldi, S. and Palmer, T. N. (1990) Regimes in the wintertime circulation over northern extratropics. I: observational evidence *Quarterly Journal of the Royal Meteorological Society,* **116,** 31-67.

Moolman, D. W. (1995) In *Department of Chemical Engineering* University of Stellenbosch, Stellenbosch, South Africa.

Moolman, D. W., Aldrich, C. and Deventer, J. S. J. v. (1995) The interpretation of flotation froth surfaces by using digital image analysis and neural networks *Chemical Engineering Science,* **55,** (22) 3501 - 3513.

Moskvina, V. (2001) In *School of Mathematics* Cardiff University, Cardiff.

Moskvina, V. and Schmidt, K. M. (2003) Approximate projectors in singular spectrum analysis *SIAM Journal on Matrix Analysis & Applications,* **24,** (4) 932-942.

Moskvina, V. and Zhigljavsky, A. (2003) An algorithm based on singular spectrum analysis for change-point detection *Communications in Statistics,* **32,** (2) 319-352.

Mullin, T. (1993) A dynamical systems approach to time series analysis In *The Nature of Chaos* Oxford Scientific Publications.

Murotani, K. and Sugihara, K. (2003) Watermarking 3D polygonal meshes using the singular spectrum analysis In *ISM Symposium on Statistics, Combinatorics and Geometry* Tokyo, pp. 7-21.

Naidu, P. D. and Malmgren, B. A. (1995) A 2 200 years periodicity in the Asian Monsoon system *Geophysical research letters,* **22,** (17) 2361-2364.

Newbigging, S., Mysak, L. A. and Hsieh, W. (2003) Improvements to the nonlinear principal component analysis method, with applications to ENSO and QBO *Atmosphere-Ocean,* **41**.

Noreen, E. W. (1989) *Computer intensive methods for testing hypotheses,* Wiley, New York.

Orfila, A., Ballester, J. L., Oliver, R., Alvarez, A. and Tintore, J. (2002) Forecasting the solar cycle with genetic algorithms *Astronomy & Astrophysics,* **386,** 313-318.

Ormerod, P. (2001) The Goodwin model and the periodicity of unemployment and factor shares in the UK.

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

Paegle, J. N., Byerle, L. A. and Mo, K. C. (2000) Intraseasonal modulation of South American summer precipitation *Monthly Weather Review,* **128,** (3) 837-850.

Palomo, M. J., Sanchis, R., Verdu, G. and Ginestar, D. (2003) Analysis of pressure signals using a singular system analysis (SSA) methodology *Progress in Nuclear Energy,* **43,** (1-4) 329-336.

Palus, M. and Novotna, D. (1998) Detecting modes with nontrivial dynamics embedded in coloured noise: enhanced Monte Carlo SSA and the case of climate oscillations *Physics Letters A,* **248,** (2-4) 191-202.

Pap, J., Rozelot, J. P., Godier, S. and Varadi, F. (2001) On the relation between total irradiance and radius variations *Astronomy & Astrophysics,* **372,** 1005-1018.

Pap, J. M. and Frohlich, C. (1999) Total solar irradiance variations *Journal of Atmospheric and Solar-Terrestrial Physics,* **61,** 15-24.

Pascual, M., Rodo, X., Ellner, S. P., Colwell, R. and Bouma, M. J. (2000) Cholera Dynamics and El Nino-Southern Oscillation *Science,* **289,** 1766-1769.

Pasqualotto, R., Marelli, L., Bilato, R., Franz, P., Giudicotti, L., Intravaia, A., Martin, P., Murari, A., Nielsen, P., Spizzo, G., Zabeo, L., Canton, A. and Terranova, D. (1999) Energy transport in RFP enhanced confined regimes In *Conference on Control Fusion and Plasma Physics*, Vol. 23J Maastricht, pp. 1153-1156.

Pederson, N., Jaoby, G. C., D'Arrigo, R. D., Cook, E. R., Buckley, B. M., Dugarjav, C. and Mijiddorj, R. (2001) Hydrometeorological reconstructions for North-eastern Mongolia derived from tree rings: 1651-1995 *Journal of Climate,* **14,** (5) 872-881.

Pereira, W. C. A., Bridal, S. L., Coron, A. and Laugier, P. (2002) Mean scatterer spacing of backscattered ultrasound signals from *in vitro* human cancellous bone specimens In *IEEE Ultrasonics Symposium*, Vol. 2.

Pereira, W. C. d. A. and Maciel, C. D. (2001) Performance of ultrasound echo decomposition using singular spectrum analysis *Ultrasound in Medicine & Biology,* **27,** (9) 1231-1238.

Plaut, G., Ghil, M. and Vautard, R. (1995) Interannual and interdecadal variability in 335 years of Central England temperatures *Science,* **268,** 710-713.

Plaut, G. and Vautard, R. (1994) Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere *Journal of the Atmospheric Sciences,* **51,** (2) 210-236.

Pohjola, V. A., Martma, T. A., Meijer, H. A. J., Moore, J. C., Isaksson, E., Vaikmae, R. and Wal, R. S. W. v. d. (2002) Reconstruction of three centuries of annual accumulation rates based on the record of stable isotopes of water from Lomonosovfonna, Svalbard In *Annals of Glaciology*, Vol. 35 International Glaciology Society, Cambridge, pp. 57-62.

Popivanov, D. and Mineva, A. (1999) Testing procedures for non-stationarity and non-linearity in physiological signals *Mathematical Biosciences,* **157,** (1-2) 303-320.

Popivanov, D., Mineva, A. and Dushanova, J. (1998) Tracking EEG signal dynamics during mental tasks *IEEE Engineering in Medicine and Biology,* **17,** (2) 89-95.

Prierto, R., Herrera, R., Doussel, P., Gimeno, L., Ribera, P., Garcia, R. and Hernandez, E. (2001a) Interannual oscillations and trend of snow occurrences in the Andes region since 1885 *Australian Meteorological Magazine,* **50,** (2) 164-168.

Prierto, R., Herrera, R., Doussel, P., Gimeno, L., Ribera, P., Garcia, R. and Hernandez, E. (2001b) Looking for periodicities in the hail intensity in the Andes region *Atmosfera,* **14,** (2) 87-93.

Qu, L., Xie, A. and Li, X. (1993) Study and performance evaluation of some nonlinear diagnostic methods for large rotating machinery *Mechanical Machine Theory,* **28,** (5) 699-713.

Rangarajan, G. K. (1998) Sunspot variability and an attempt to predict solar cycle 23 by adaptive filtering *Earth Planets Space,* **50,** 91-100.

Rangarajan, G. K. and Araki, T. (1997) Multiple timescales in the fluctuations of the equatorial dst index through singular spectrum analysis *Journal of Geomagnetism and Geoelectricity,* **49,** (1) 3-20.

Rangarajan, G. K. and Barreto, L. M. (2000) Long term variability in solar wind velocity and IMF intensity and the relationship between solar wind parameters & geomagnetic activity *Earth Planets and Space,* **52,** (2) 121-132.

Rangarajan, G. K. and Iyemori, T. (1997) Time variations of geomagnetic activity indices Kp and Ap: an update *Annales Geophysicae - Atmospheres Hydrospheres and Space Sciences,* **15,** (10) 1271-1290.

Rangarajan, G. K. and Iyemori, T. (1998) Hemispherical and latitudinal differences in the response of geomagnetic activity to recurrent solar wind streams *Journal of Atmospheric and Solar-Terrestrial Physics,* **60,** 1035-1046.

Rastogi, S. and Klinzing, G. E. (1994) Characterizing the rheology of powders by studying dynamic avalanching of the powder. *Particle and Particle Systems Characterization,* **11,** 453-456.

Reeve, D. E. (2002) Comments on "Forced and self-organized shoreline response for a beach in the Southern Baltic Sea determined through singular spectrum analysis" *Coastal Engineering,* **44,** 267-269.

Ribera, P., Garcia, R., Diaz, H. F., Gimeno, L. and Hernandez, E. (2000) Trends and interannual oscillations in the main sea-level surface pressure patterns over the Mediterranean *Geophysical research Letters,* **27,** (8) 1143-1146.

Robertson, A. W. and Mechoso, C. R. (1998) Interannual and decadal cycles of river flows of South-eastern South America *Journal of Climate,* **11,** (10) 2570-2581.

Robertson, A. W. and Mechoso, C. R. (2003) Circulation regimes and low-frequency oscillations in the South Pacific sector *Monthly Weather Review,* **131,** 1566-1576.

Rodo, X., Pascual, M., Fuchs, G. and Faruque, A. S. G. (2002) ENSO and cholera: A nonstationary link related to climate change? *PNAS,* **99,** (20) 12901-12906.

Rozynski, G. (2002) Reply to comments on "Forced and self-organized shoreline response for a beach in the Southern Baltic Sea determined through singular spectrum analysis" *Coastal Engineering,* **44,** 271-272.

Rozynski, G., Larson, M. and Pruszak, Z. (2001) Forced and self-organized shoreline response for a beach in the southern Baltic Sea determined through singular spectrum analysis *Coastal Engineering,* **43,** (1) 41-58.

Schlesinger, M. E. and Ramankutty, N. (1994) An oscillation in the global climate system of period 65-70 years *Nature,* **367,** 723-727.

Schoellhamer, D. H. (1996) Factors affecting suspended-solids concentrations in South San Francisco Bay, California *Journal of Geophysical Research,* **101,** (C5) 12087-12095.

Schoellhamer, D. H. (2001) Singular spectrum analysis for time series with missing data *Geophysical Research Letters,* **28,** (16) 3187-3190.

Schoellhamer, D. H. (2002) Variability of suspended-sediment concentration at tidal to annual time scales in San Francisco Bay, California *Continental Shelf Research,* **22,** 1857-1866.

Schreiber, T. Measuring nonlinear dynamics.

Schreiber, T. (1998) Measuring nonlinear dynamics In *Non-linear Dynamics in Mechanical Processing* Dortmund.

Schreiber, T. (2000) Is nonlinearity evident in time series of brain electrical activity? In *Chaos in Brain?*(Eds, Lehnertz, K., Elger, C. E., Arnhold, J. and Grassberger, P.) World Scientific, Singapore.

Schwartz, G. (1978) Estimating the dimension of a model *The Annals of Statistics 6*, 461 - 464.

Shabalova, M. V. and Weber, S. L. (1998) Seasonality of low-frequency variability in early-instrumental European temperatures *Geophysical Research Letters,* **25,** (20) 3859 - 3862.

Shaikh, F. N. (1997) Investigation of transition to turbulence using white-noise excitation and local analysis techniques *Journal of Fluid Mechanics,* **348,** 29-83.

Shun, T. and Duffy, C. J. (1999) Low-frequency oscillations in precipitation, temperature and run-off on a west-facing mountain front: A hydrogeologic interpretation *Water Resources Research,* **35,** (1) 191-201.

Smith, L. and Tüzün, U. (2002) Stress, voidage and velocity coupling in an avalanching granular heap *Chemical Engineering Science,* **57,** 3795-3807.

Solow, A. R. and Patwardhan, A. (1996) Extracting a smooth trend from a time series: A modification of singular spectrum analysis *Journal of Climate,* **9,** 2163-2166.

Stahle, D. W., D'Arrigo, R. D., Krusic, P. J., Cleaveland, M. K., Cook, E. R., Allan, R. J., Cole, J. E., Dunbar, R. B., Therrel, M. D., Gay, D. A., Moore, M. D., Stokes, M. A., Burns, B. T., Villaneuva-Diaz, J. and Thompson, L. G. (1998) Experimental dendroclimatic reconstruction of the Southern Oscillation *Bulletin of the American Meteorological Society,* **79,** (10) 2137-2150.

Stive, M. J. F., Aarninkhof, S. G. J., Hamm, L., Hanson, H., Larson, M., Wijnberg, K. M., Nicholls, R. J. and Capobianco, M. (2002) Variability of shore and shoreline evolution *Coastal Engineering,* **47,** 211-235.

Takens, F. (1993) Detecting nonlinearities in stationary time series *International Journal of Bifurcation and Chaos,* **3,** 241-256.

Tan, S. and Mavrovouniotis, M. L. (1995) Reducing data dimensionality through optimizing neural network inputs *AIChE Journal,* **41,** (6) 1471-1480.

Theiler, J. (1995) On the evidence for low-dimensional chaos in an epileptic encephalogram *Physics Letters A,* **196,** 335-341.

Theiler, J., Eubank, E., Longtin, A. and Galdrikian, B. (1992) Testing for nonlinearity in time series: The method of surrogate data *Physica D,* **58,** 77-94.

Theiler, J. and Prichard, D. (1996) Constrained-realization Monte Carlo method for hypothesis testing *Physica D,* **94,** 221-235.

Thomakos, D. D., Wang, T. and Wille, L. T. (2002) Modelling daily realized futures volatility with singular spectrum analysis *Physica A: Statistical Mechanics and its Applications,* **312,** (3-4) 505-519.

Tianfang, C. and Tianxing, C. (2000) Detecting dynamical signals using singular spectrum analysis In *International Conference on Signal Processing*, Vol. 1, pp. 290-293.

Tsay, R. S. (1992) Model checking via parametric bootstraps in time series analysis *Applied Statistics,* **41,** 1-15.

Tsonis, A. A. and Elsner, J. B. (1992) Oscillating global temperature *Nature,* **356,** 751.

Tung, C.-T., Tseng, D.-C. and Tsai, Y.-L. (2001) Mixed-pixel classification for hyperspectral images based on multichannel singular spectrum analysis *International Geoscience and Remote Sensing Symposium,* **5,** 2370-2372.

Van der Walt, T. J. (1992) In *Department of Chemical Engineering* Stellenbosch University, Stellenbosch.

Varadi, F., Pap, J. M., Ulrich, R. K., Bertello, L. and Henney, C. J. (1999) Searching for signal in noise by random-lag singular spectrum analysis *The Astrophysical Journal,* **526,** 1052 - 1061.

Varadi, F., Ulrich, R. K., Bertello, L. and Henney, C. J. (2000) Random-lag singular cross-spectrum analysis *The astrophysical journal,* **528,** 53-56.

Vautard, R. and Ghil, M. (1989) Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series *Physica D: Nonlinear Phenomena,* **35,** (3) 395-424.

Vautard, R., Plaut, G., Wang, R. and Brunet, G. (1999) Seasonal prediction of North American surface air temperatures using space-time principal components *Journal of Climate,* **12,** (2) 380-394.

Vautard, R., Yiou, P. and Ghil, M. (1992) Singular-spectrum analysis: A toolkit for short, noisy chaotic signals *Physica D: Nonlinear Phenomena,* **58,** (1-4) 95-126.

Wainer, I. and Venegas, S. A. (2002) South Atlantic multidecadal variability in the climatic system model *Journal of Climate,* **15,** (12) 1408-1420.

Wang, W. J., Chen, J., Wu, X. K. and Wu, Z. T. (2001) The application of some nonlinear methods in rotating machinery fault diagnosis *Mechanical Systems and Signal Processing,* **15,** (4) 697-705.

Watari, S. (1996) Separation of periodic, chaotic and random components in solar activity *Solar Physics,* **168,** 413-422.

Wax, M. and Kailath, T. (1985) Detection of signals by information theoretic criteria *IEEE Transactions on Acoustics, Speech and Signal Processing,* **33,** (2) 387-392.

Wu, H. and Gong, J. (2000) Forecast of network behaviour based on singular-spectrum analysis In *International Conference on Communication Technology*, Vol. 1, pp. 702-706.

Ye, H. (2001) Characteristics of winter precipitation variation over Northern Central Eurasia and their connections to sea surface temperatures over the Atlantic and Pacific Oceans *Journal of Climate,* **14,** 3140-3155.

Ye, H. and Cho, H. R. (2001) Spatial and temporal characteristics of intraseasonal oscillations of precipitation over the United States *Theoretical and Applied Climatology,* **68,** (1-2) 51-66.

Yiou, P., Fuhrer, K., Meeker, L. D., Jouzel, J., Johnsen, S. and Mayewski, P. A. (1997) Paleoclimatic variability inferred from the spectral analysis of Greenland and Antarctic ice-core data *Journal of Geophysical Research - Oceans,* **102,** (C12) 26441-26454.

Yiou, P., Jouzel, J., Johnsen, S. and Rognvaldsson, O. E. (1995) Rapid oscillations in Vostok and GRIP ice cores *Geophysical Research Letters,* **22,** (16) 2179-2182.

Yiou, P., Sornette, D. and Ghil, M. (2000) Data-adaptive wavelets and multi-scale singular-spectrum analysis *Physica D: Nonlinear Phenomena,* **142,** (3-4) 254-290.

Yu, J.-Y. and Mechoso, C. R. (2001) A coupled atmosphere-ocean GCM study of the ENSO cycle *Journal of Climate,* **14,** (10) 2329-2350.

Zhang, X. (1998) Chaotic characteristics analyses of underwater acoustic signals *International Conference on Signal Processing Proceedings,* **2,** 1451-1454.

Zhang, X., Corte-Real, J. and Wang, X. L. (1997) Low-frequency oscillations in the Northern Hemisphere *Theoretical and applied climatology,* **57,** 125-133.

Zhang, X., Sheng, J. and Shabbar, A. (1998) Modes of interannual and interdecadal variability of Pacific SST *Journal of Climate,* **11,** (10) 2556-2569.

INDEX

APPENDIX A SOFTWARE DEVELOPED

A.1 MATLAB software

A MATLAB toolbox, called *ssa_calc*, was developed to perform normal SSA, multichannel SSA, nonlinear SSA and Monte Carlo SSA, along with a number of other options, such as the generation of reconstructed attractors and selected other figures.

The programming code for this toolbox would be too voluminous too include in its entirety in the appendix. It was therefore decided to rather provide the toolbox on an enclosed compact disc and just provide the programming code of the main menu of the toolbox. This menu illustrates all the options that the user of the toolbox has, along with the required input parameters and the format of the output parameters obtained. The program itself will not be discussed, but the relevance of the variance options should be clear from the commenting in the programming code itself.

A.2 Programming code of *ssa_calc* toolbox

```
%[Series, Model, Data] = ssa_calc(TS, SeriesInfo)
%
% Perform SSA calculations on a chosen time series and generate relevant
% figures to illustrate the results
%
% Input:
%   TS - Timeseries. Currently only single timeseries can be accommodated
%   but future provision will be made for multivariate timeseries
%
%   SeriesInfo - Optional input argument containing information on series,
%   obtained as output argument when the function was previously used on
%   the same timeseries (see Output)
%
% Output:
%   Series - Structured array containing all relevant information on the
%   timeseries, consisting of the following fields:
%       OriginalTS = Original timeseries
%       TSEmbedded = Trajectory matrix
%       Window = Window size of trajectory matrix
%       P = Loading values obtained from PCA
%       T = Scores obtained from PCA
%       L = Eigenvalues from PCA;
%       Tsq = Hoteling T-Square statistic;
%       PersVar = Vector containing cumulative percentage variance explained
%       NrComponents = Nr of principal components retained;
%       pers = Percentage variance explained by retained principal
%           components
%       TSrec = Matrix obtained by multiplying T and P' for fewer components
%       TSUnembed = Reconstructed TS with fewer principal components;
%       CumRecon = Cumulative reconstruction of T*P' values (RC's)(i.e. 1 RC in
%           column 1, 2 RC's in column 2, 3 RC's in column 3, etc;)
%       TPRecon = Individual reconstruction of T*P' values (RC's)(i.e. RC 1 in
%           column 1, RC 2 in column 2, RC 3 in column 3, etc)
%       Conlim = Structured array containing 95%, 99% and 99.9% confidence
%           limits for eigenspectra of surrogate data (optional)
%       Surrogate  = Surrogate data sets generated (optional)
%   Model - Model definition obtained from quick_ident
%   Data - Data definition for Model, obtained from quick_ident
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% c. Marlize Barkhuizen, Oct 2003



function[Series, Model, Data] = ssa_calc(TS,varargin)

NrSeries = size(TS,2);

clc
non_lin = 'n';
mchoice = ' ';
Model = 0;
Data = 0;


tf = isempty(varargin);
if tf == 0
   Series = varargin{1}(1);
   disp (' ')
   disp (' ')
   non_lin = input('Was the supplied analysis for the time series done using a non-linear
technique? (y/n)  ','s')
   disp (' ')
   if isempty(non_lin)
      non_lin = 'n';
   end

   if non_lin ~= 'y'
      Window = Series.Window;
   end
else
   if NrSeries ==1
      SeriesName = input('What is the name of the time series?     ','s');
   else
      GenName = input ('What is the name of the combined time series?  ','s');
      for k = 1:NrSeries
         disp (' ')
         disp (' ')
         boodsk = sprintf('What is the name of individual time series nr %d   ',k);
         naam = input(boodsk,'s');
         SeriesName(1,1) = {GenName};
         SeriesName(2,k) = {naam};
      end
   end
   Window = 0;
   Series.Name = SeriesName;
   Series.OriginalTS = auto(TS);
```

```
    Series.TSEmbedded = 0;

end


while mchoice ~= '0'

    clc

    disp ('Please select one of the following options: ')
    disp (' ')
    disp ('1. Determine embedding window size')
    disp ('2. Extract principal components')
    disp ('3. Monte Carlo simulation')
    disp ('4. Generate selected figures')
    disp ('5. Perform modelling of the time series using Quick Ident')
    disp ('6. Save results')
    disp (' ')
    disp ('0. Quit')
    disp (' ')

    mchoice = input ('Choice: ','s');

    switch (mchoice)
        case '1'
            clc
            if non_lin == 'y'
                disp ('This option is not valid for the non-linear analysis of the time series')
                disp ('Press any key to return to the main menu')
                pause
            else
                if NrSeries ==1
                    [Window] = window_length(Series, 500);
                else
                    [Window] = window_length_mult(Series, 500);
                end

                disp (' ')
                disp ('Press any key to return to the main menu ')
                pause
            end

        case '2'
            clc
            disp ('Please select one of the following options: ')
            disp (' ')
            disp ('1. Linear Principal Component Analysis')
            disp ('2. Localized Principal Component Analysis')
            disp ('3. Auto-associative MLP components')
            disp (' ')

            schoice2 = input ('Choice: ','s');

            if isempty(schoice2)
                schoice2 = '1';
            end

            if schoice2 == '1'

                if NrSeries ==1
```

```
          [Series] = pca_calc(Series, auto(TS), Window);
        else
          [Series] = pca_calc_mult(Series, auto(TS), Window);
        end
        non_lin = 'n';
      end

      if schoice2 == '2'

        [Series] = pca_calc_non_lin_random(Series, auto(TS));
        non_lin = 'y';

      elseif schoice2 == '3'
        clc
        if non_lin == 'y'
          disp ('The results from the localized principal components will now be deleted')
          toestem = input ('Do you want to continue? y/n');
          if toestem =='y'
            Series = rmfield(Series,'Ind_Part_Info');
            Series = rmfield (Series, 'NrParts');
            Window == 0;
          else
            continue
          end

        end
        if Window == 0
          [Window] = window_length(Series, 500);
          if Series.TSEmbedded == 0
            [x,y] = embed(TS,Window,-1);          % Embed data with window length
calculated above
            Series.TSEmbedded = y';
          end
        end

        [Series] = nn_mlp_mod_ssa(Series);

      end

      disp (' ')
      disp ('Press any key to return to the main menu ')
      pause


    case '3'
      clc
      if non_lin == 'y'
        [Series] = eigen_confidence_lim_non_lin (Series)
      else
        [Series] = eigen_confidence_lim (TS, Series);
      end
      disp (' ')
      disp ('Press any key to return to the main menu ')
      pause


    case '4'
      clc
      [Series] = figure_menu(Series, NrSeries, auto(TS), non_lin);
    case '5'
      clc
      [Model, Data] = modelling(Series, auto(TS));
      disp (' ')
      disp ('Press any key to return to the main menu ')
      pause


    case '6'
      clc
      save_results(Model, Data, Series)


    case '0'
      return

    otherwise
      disp('Invalid input! Please choose a number between 0 and 9')
      disp(' ')

  end %end case

end
clc

return
```

APPENDIX B PUBLICATIONS

Aldrich, C. and **Barkhuizen, M.** (2003). Process system identification strategies based on the use of singular spectrum analysis. *Minerals Engineering*, **16**(9), 815-826

Aldrich, C. and **Barkhuizen, M.** (2003). Identification of process systems with singular spectrum analysis and neural networks. *Proceedings of the 22nd International Mineral Processing Congress*, (XXII IMPC), **3**, 1627-1637 [Cape Town, South Africa, 28 September – 3 October 2003]

Aldrich, C and **Barkhuizen, M.** (2003). Identification of nonlinear process dynamics with Monte Carlo singular spectrum analysis. *Journal of the South African Institute of Mining and Metallurgy*, **103**(2), 127-137

Barkhuizen, M. and Aldrich, C. (2003). Classification of process dynamics with Monte Carlo singular spectrum analysis. *Proceedings of the 14th ESCAPE Conference,* (ESCAPE 14), [Lisbon, Portugal, 16-19 May 2004] (In press)

Aldrich, C. (2003) and **Barkhuizen, M**. Analysis of process dynamics with Monte Carlo singular spectrum analysis. *Proceedings of the 1st African Control Conference,* [Cape Town, South Africa, 3-5 December 2003] (In press)

Barkhuizen, M. and Aldrich, C. (2003). Analysis and process modelling with singular spectrum analysis and neural networks. *Chemical Engineering R&D Conference '03*, Western Cape Branch of SAIChE, Cape Town, South Africa, 4 April 2003