Classification of Process Dynamics with Monte Carlo Singular Spectrum Analysis

Marlize Barkhuizen and Chris Aldrich

Department of Process Engineering, University of Stellenbosch, Private Bag X1, Matieland, 7602, Stellenbosch, South Africa, Fax: +27(21)808 2059

Abstract

Singular spectrum analysis is a linear multivariate method for the analysis of time series data, based on principal component analysis of an augmented data set consisting of the original time series data and lagged copies of the data. It can be used to decompose the time series into a set of component time series, each of which could be investigated individually to gain a better understanding of the process dynamics, or to allow for the removal of noise from the data.

Keywords: Singular spectrum analysis, time series analysis, predictive modelling

1. Introduction

It is well known that reliable, effective process control is vital to the efficient and competitive operation of chemical process systems. With the increased emphasis on more advanced (model-based) control systems in particular, accurate control depends on a solid grasp of the dynamic behaviour of the system, or failing that, at least some timely, reliable diagnostics of the process dynamics of the system. In particular, the search for evidence of predictability in the data provides a starting point for system identification and the construction of advanced control systems. Although such predictability may be limited in nonlinear systems, even limited predictability may be informative and hence the value of techniques designed for the detection of periodicities or intermittent trends in the data.

In this context, singular spectrum analysis is a relatively new technique that has been developed initially in the field of climatology (Broomhead and King, 1986, Vautard and Ghil, 1989, Vautard et al., 1992), but have since been applied in a variety of other research fields, among which are the biosciences (Mineva and Popivanov, 1996), geology (Rozyński et al., 2001, Schoellhamer, 2001), economics (Ormerod and Campbell, 1997) and solar physics (Kepenne, 1995). It is based on the idea of sliding a window down a time series in order to identify patterns which account for a large proportion of the variance in the views of the time series.

Monte Carlo singular spectrum analysis (MC-SSA) is a methodology for discriminating between various components of time series data, 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. The paper is organized as follows. First, the basics of SSA are introduced in section 2, followed by the methodology of Monte Carlo SSA. These techniques are subsequently considered in a couple of case studies on industrial processes.

2. Methodology

2.1 The basics of singular spectrum analysis

The general approach behind SSA can be described in four main steps (Golyandina et al., 2001), as indicated in figure 1. 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, during which the principal components and eigenvalues of the series are extracted. These two steps represent the decomposition stage. The reconstruction stage 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 matrices formed in step three.

\[ X = TP^T \]

**Fig. 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.

2.2 Monte Carlo singular spectrum analysis

Generally speaking, MC-SSA is an extension of basic SSA that involves a null hypothesis against which the data are tested. This is done by using a discriminating statistic, such as the autocorrelation, the correlation dimension or some other discriminating property of the data and comparing this property of the time series under investigation with that of a number of surrogate series. 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 similar to the original time series are subsequently generated from this process and various statistics are calculated from both the surrogate and the original data. If these statistics do not differ, the original time series is assumed to have the same character as the surrogate data. As a trivial example, to test whether a time series contains any structure, surrogate data would be generated by randomizing the time series and comparing the autocorrelations of the original time series and its surrogates. If these statistics do not differ, it could be concluded that the original time series is indeed the same as the surrogates, which contain no structure.

3. Recovery of Base Metals from Scavenger Cells of Flotation Plant.

3.1 Background
The data in this case study 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 time series investigated were the measurements of the recovery grades of the precious metals, Cu and Pb, in the scavenger circuit. Figure 2 shows the 12-minute interval measurements of the Pb and Cu concentrations. Each time series consisted of 1234 measurements and was scaled to zero mean and unit variance before doing the analysis.

3.2 Classification of time series
The time series were being tested against the hypothesis that the data had been generated by a first order autoregressive process and the test statistic being used was the eigenspectrum of the series, obtained from the PCA decomposition of the lagged trajectory matrix of the time series. The 15 first order autoregressive surrogate series were used to generate 95% confidence limits for the eigenspectra of the series, which
are displayed in figure 3. From this figure it can be seen that the eigenspectrum of the lead time series is almost completely within the confidence limits, with those of the copper time series falling partly inside and partly outside the confidence limits.

One can conclude from these results that the copper time series had not been generated by a first order autoregressive process, but that the lead series can probably be characterized as a first order autoregressive process.

More information about the nature of the time series can be obtained from the appearance of the reconstructed attractors (figure 4), which are the first three principal components plotted as a function of each other. From the attractor of the zinc time series, it would seem as if the time series is not stationary. This property is not obvious from the appearance of the time series itself (figure 2) and more refined tests are usually required to determine the stationarity of a time series than visual inspection.

Fig. 3. Eigenvalue spectra of copper (a) and lead (b) time series along with the 95% confidence limits generated from first order autoregressive surrogate series.

Fig. 4. Attractors of copper (a) and lead (b) recoveries in the scavenger cell of a base metal flotation plant. The percentage of the total variance explained by each principal component is shown in parenthesis in the appropriate axis label.
4. Acid Concentration in a Leach Circuit.

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

Fig. 5. Sample of twice-daily observations of scaled sulphuric acid concentration in the anolyte (solid line) and feed (broken line) of an industrial leach plant.

The data in figure 5 shows the (scaled) concentrations of the $H_2SO_4$ in the anolyte (solid line), while the broken line shows the acid concentration in the feed. 2282 twice-daily samples were considered. Figure 6 shows the results of the Monte Carlo simulations to test the null hypothesis that the time series are 1st order autoregressive processes. These results suggest that the observed acid concentration in the feed is of a lower order autoregressive process than that of the acid in the anolyte. The results are confirmed by linear autoregressive models fitted to the data, the results of which are shown in figure 7. In this figure, the broken line shows that the model order for the acid in the anolyte is approximately 3 or 4, while that of the acid in the feed is approximately 2 or 3.

5. Conclusions

The basic formalism of SSA provides a natural test for periodic components in the time series against arbitrary stochastic models (null hypotheses). In this paper, testing by means of Monte Carlo SSA was considered with simple noise models. Although univariate time series only were considered, the method is readily extendable to multivariate time series and can serve as a useful guide the modelling of dynamic processes.
Fig. 6. Eigenspectra (cross markers) of sulphuric acid in the anolyte (top) and feed (bottom) shown in Figure 5. The bold lines indicate the estimated 95% confidence limits of the eigenvalues. Note that the first eigenvalue in each case lies outside the confidence limits by a large margin, suggesting higher than 1st order processes.

Fig. 7. Predictive errors and model order of autoregressive models fitted to the sulphuric acid in the anolyte (broken line) and the feed (solid line).

References