ON-LINE MONITORING AND CONTROL OF INDUSTRIAL MINERAL PROCESSING PLANTS BY USE OF SINGULAR SPECTRUM ANALYSIS

C. Aldrich
Department of Chemical Engineering, University of Stellenbosch, Stellenbosch, Private Bag X1, Matieland, 7602, Stellenbosch, South Africa, Fax +27(21) 808 2059
E-mail:

ABSTRACT

System identification from on-line measurements is a key issue in the development of automatic control systems for flotation plants. Unfortunately, noise in the data poses a significant hurdle in the development of reliable nonlinear dynamic models. In this investigation, some preliminary results on the use of singular spectrum analysis are considered and it is shown that the approach provides a feasible means of neutralizing the deleterious effects of noise on process models and by implication also on the quality of subsequent process control structures.

INTRODUCTION

In the last five years, considerable progress has been made with control systems based on direct monitoring of the froth (Moolman et al., 1995). 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 froths. In addition, the presence of reagents or mineral species can also be related to the appearance of the froth.

The identification of the underlying dynamics of froth flotation systems from experimental data is complicated by the fact that erratic fluctuations in the observed behaviour are typically derived from a mixture of various plant disturbances that can be notoriously difficult to disentangle. In this example, the use of singular spectrum analysis (Vautard et al., 1992) and the method of delay coordinates to identify the dynamics of flotation systems are considered.

Since flotation processes are complex and in practice the control of industrial flotation plants is often based on the visual appearance of the froth phase, the quality of control depends largely on the experience and ability of a human operator. A standard control cycle comprises the fixing of initial set points, a settling period for transient dynamics to subside, a period of measurement and evaluation, and a final estimate of appropriate set points. Some studies have indicated that operators often tend to make the periods for settling and measurement too small. Apart from these aspects, the inexperience or inability of the operator can have a further significant impact on the control of the plant. Consequently, optimal control is not usually maintained, especially where incipient erratic behaviour in the plant is difficult to detect.

This information has changed the way in which flotation plants can be monitored and controlled. Although initial control strategies centred around multivariate statistical process control schemes, it is realized that full automation may also be a feasible option, given the quality of present instrumentation.

High-quality on-line data means reliable on-line identification of the flotation process, from noisy nonlinear measured data. The high level of noise poses a particular problem, as it has a severely deleterious effect on traditional system identification algorithms. In this investigation, the merits of singular spectrum analysis to identify flotation systems is considered and it is shown that this approach provides a feasible means for the development of automatic control systems for flotation plants.

STATE SPACE APPROACH TO SYSTEM IDENTIFICATION

The identification of the underlying dynamics of froth flotation systems from experimental data is complicated by the fact that erratic fluctuations in the observed behaviour are typically derived from a mixture of various plant disturbances that can be notoriously difficult to disentangle. Singular spectrum analysis is based on the singular value decomposition of data matrices representing the froth images (or any other
sensor data) and is comparatively simple to use, as will be shown by way of a case study in the paper. Owing to the complexity of floritation, identification based on phenomenological models is not feasible and as a consequence dynamic process models have to be developed direct from input-output data.

A single record (time series) from a dynamic system is the result of all interacting variables associated with the process. In principle, it should therefore contain information about the dynamics of all the key variables involved in the evolution of the system. Mathematically, this can be justified by assuming that the variables of a dynamic process system satisfy a single $p$th order differential equations, i.e.

$$\frac{dx_1}{dt} = f_1(x_1, x_2, \ldots, x_p)$$
$$\frac{dx_2}{dt} = f_2(x_1, x_2, \ldots, x_p)$$
$$\frac{dx_p}{dt} = f_p(x_1, x_2, \ldots, x_p) \quad (1)$$

This system of differential equations can be reduced to a single $p$th order differential equation via successive differentiation of any one of the $1^{st}$ order differential equations, that is

$$d^n x_i/dt^n = f(x_1, dx_1/dt, d^2x_1/dt^2, \ldots, d^{n-1}x_1/dt^{n-1}) \quad (2)$$

This equation represents the entire set of $1^{st}$ order equations without any loss of information regarding the dynamics of the system. In other words, the single variable system (set of $1^{st}$ order differential equations) is equivalent to the multivariable system ($p$th order differential equation).

Moreover, instead of a continuous variable and its derivatives, a discrete time series and its successive shifts by a lag parameter $\tau$ should be sufficient to capture the dynamics of the system (Ruelle, 1980). The lagging is equivalent to first order differencing of the time series, which is analogous to differentiation. In an observed time series, the lagged copies of the series can thus be seen as additional system variables.

One of the difficulties in dealing with empirical systems is that one does not know a priori whether deterministic dynamics underlie the data, that is whether dynamic attractors exist. Clearly, reliable classification of the data as a first step in system identification is important, otherwise the resulting model will not generalize beyond the training data set.

Reconstruction of the state space of the system facilitates successful identification of the system and can be applied to all deterministic systems. A state space model ensures the complete description of the dynamics of a system and allows prediction of the system outputs, even without complete knowledge of the fundamental system dynamics.

### State Space Reconstruction

Broadly speaking, three methods can be used for state space reconstruction from a scalar time series, viz. derivative coordinates, delay coordinates and what is known as singular spectrum analysis. Derivative coordinates (Packard et al., 1980) are based on higher order derivatives of the time series, but are not particularly useful for experimental data, owing to their susceptibility to noise. In contrast, delay coordinate methods and singular spectrum analysis are used extensively on experimental systems. Of the two, singular spectrum analysis appears to be the more robust in the presence of noise, as explained below.

### Singular Spectrum Analysis

The term singular spectrum comes from the spectral (eigenvalue) decomposition of a matrix $A$ into its spectrum (set) of eigenvalues. The eigenvalues $\lambda$ are the values that make the matrix $A - \lambda I$ singular. Actually, the term singular spectrum is somewhat unfortunate in the context of the analysis of time series, since the traditional eigenvalue decomposition (principal component analysis) of matrices representing multivariate data is also an analysis of the singular spectrum. The spectral decomposition of matrices has only recently been applied to time series analysis and has had its roots mostly in the application of chaos theory (Broomhead and King, 1986; Landa and Rosenblum, 1991).

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

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

$$X = \begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4
\end{bmatrix} = \begin{bmatrix}
x_1 & x_2 \\
x_2 & x_3 \\
x_3 & x_4 \\
x_4 & x_5
\end{bmatrix}$$

It is not necessary to use successive values of the observations. For example, with every second
observation only (i.e. $k = 2$) three new variables would have been created, $z_1 = [x_1, x_2]^T$, $z_2 = [x_2, x_3]^T$, and $z_3 = [x_3, x_4]^T$. The construction of the matrix $X$ is based on an $(m,k)$ window, i.e. $m$ indicates the number of time series observations in each new snapshot and $k$ indicates the sample times between observed elements. In the first case, a $(3,1)$-window was used, while a $(3,2)$-window was used in the second case. If $k = 1$ is used, then reference is made to an $m$-window only. Matrix $X$ is referred to as the augmented matrix or trajectory matrix and contains a complete record of patterns that have occurred in a window of size $m$. That is $X_i \in \mathbb{R}^{(n+m-1) \times m}$.

Since $X_{i+j} = X_{i+j}$ for all $i > 1$ and $j > 1$, the columns of the matrix are highly correlated. The embedding space $N \in \mathbb{R}^n$ is the space of all $m$-element patterns.

Rather than investigating the trajectory matrix $X$ for repetitive patterns in the original time series, it is more effective to consider the lagged covariance matrix computed from the trajectory matrix $X$ and its transpose, i.e.

$$S = X^TX/(m-1).$$

Reconstruction of the attractor of the system is accomplished by a simple decomposition of $X$ as

$$X = TP^T$$

by solving for $Sp = \lambda_p p$, subject to the orthonormality constraints $pp^T = 1$ $(i = j)$ and $pp^T = 1$ $(i \neq j)$.

The loading matrix $P = [p_1, p_2, \ldots, p_p]^T$ constitute the phase space in which the time series is embedded, via the score matrix $T = [t_1, t_2, \ldots, t_i]^T$. Selection of the number of load vectors to obtain the dimensionality of the embedding is based on the variance explained by the retained loading vectors (principal components), as represented by the diagonal eigenvalue matrix of the system, $\Lambda$. Although the attractor can be visualized quite reliably with principal components explaining 60-70% or more of the total variance of the lagged covariance matrix, it is better to retain a larger number of principal components for building predictive models. Approaches to ensure that only the components associated with noise are omitted from the model are discussed elsewhere.

Recursive Prediction

Once embedding of the observed data has been completed, predictive models can be built, based on the multiple input-single output (MISO) system $E: \mathbb{R}^p \rightarrow \mathbb{R}^l$, defined by

$$y'(t+1) = f[y(t)]$$

Equation (5) is the crux of the dynamic model. Once it is established, the evolution $y(t) \rightarrow y(t+1)$ is known and this in turn determines the unknown evolution of $x(t) \rightarrow x(t+1)$. Unlike linear models, where minimization of the mean square of the prediction error leads to accurate models, minimization of the mean square of the prediction error is a necessary, but not a sufficient criterion for the reliability of the models of nonlinear systems, since trajectories in the same attractor can differ vastly from one sample to another.

The model is of the form $\mathbb{R}^n: x_t \rightarrow y_{t+1}$, where $x$ is the embedding vector at time $t$. A state space parameterization of $Y$ is formed by using time series embedding, as follows:

$$x_{t+1} = f(x_t)$$

$$y_t = g(x_t)$$

In the above nonlinear state space formulation, $x = \Omega(y, m, k), \vec{y} = \Omega'$, where $\Omega$ is the embedding operator.

The model is an approximation of $y_{t+1} = g(f(x_t))$ as $y_{t+1,es} = g_{es}(x_t)$. The model structure that is estimated can be any appropriate non-linear regressor, e.g. a multilayer perceptron or a radial-basis function neural network. The phase variables that constitute the embedding vector are not directly related to the original process states, but are an equivalent parameterization of the system.

Owing to its flexibility with regard to stochastic or deterministic systems (be they linear or nonlinear), singular spectrum analysis provides a very useful approach towards the modelling of flotation systems, as will be illustrated below.

IDENTIFICATION OF AN INDUSTRIAL BASE METAL FLOTATION PLANT.

A machine vision system consisting of a videocamera in a protective casing mounted on top of a flotation cell on a South African base metal plant (Van Olst, 1998) was connected to a personal computer equipped with a frame grabber, which monitored one of the cells in a bank of primary roughers. The identification of the system was based on the stability measurements of the froth, a sample of which is shown in Figure I below.
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

0 500 1000 1500 2000 2500 3000
Observation

Figure 1. Bubble size measurements of froth on a base metal flotation plant. Samples were collected with intervals of one minute over a period of 20 hours.

Takens’ Embedding of the Observations

As a first step to identification, the data have to be embedded. As mentioned above, this means the determination of a suitable lag and embedding dimension. Selection of an appropriate lag is usually guided by the principle that embedding vector components should be independent, i.e. the correlation or dependence between lag variables should be low. The average mutual information function of this time series is shown in Figure 2. This figure indicates a lag of $k = 2$ (at the first minimum value of the function). The embedding dimension of $m = 4$ was determined by the false nearest neighbour algorithm, which unfolds (embeds) the data in successively larger dimensions, until the observations become stationary. With a lag of $k = 2$ and an embedding dimension of $m = 4$, the attractor of the time series can be reconstructed, as shown in Figure 3.

As can be seen from Figure 3, the attractor is extremely spiky, with little discernible structure, except for the two clustered regions. Although allowance should be made for distortion caused by the projection of the attractor from four to three dimensions, the appearance of the attractor indicates a high level of stochasticity, suggesting that system identification can be expected to be poor.

Figure 2. Average mutual information (AMI) of the time series depicted in Figure 1, indicating a lag of 2.

Figure 3. 3-D reconstruction of the attractor of the time series shown in Figure 1.

Reconstruction of the same data with singular spectrum analysis does not require the explicit specification of a lag, although some indication of this is required in order to select a window of observations of sufficient length. The autocorrelation function shown in Figure 4 suggests a lag of approximately 80, where the autocorrelation is weak (less than 0.3).

Principal components were consequently extracted from a lagged covariance matrix with 80 lag variables. The first three principal components explained 70.6% of the variance of the data, while the remaining principal components were all of more or less equal size, indicating noise in the data.

Figure 4. Autocorrelation function of the time series shown in Figure 1.
Reconstruction of the attractor by projection of the lagged variables onto the first three principal components is shown in Figure 5. This attractor has visibly more structure than the comparable one shown in Figure 3.

The reconstructed time series based on the first three principal components is shown in Figure 6. The two time series have been shifted for better visualization.

A predictive model (multilayer perceptron neural network) was subsequently fitted to the data, according to equations 5 and 6. The single hidden layer of the neural network was automatically optimized by use of the Schwarz information criterion and the trained neural network was used to predict the last few hundred observations in the time series. Note that these last observations were uncharacteristic of the preceding observations (as indicated in Figures 1 and 6).

---

**Figure 5.** Reconstructed attractor of the time series in Figure 1, by projection of the first 80 lag variables (k=1) onto the first three principal components.

**Figure 6.** Reconstructed observations (solid line at bottom) based on the first three principal components extracted from the lagged covariance matrix of the data (broken line at top). The data have been shifted for better visualization.

The predictions generated by the models are shown in Figure 7. The model can be improved considerably by the addition of input variables in the form of a multiple-input-single-output (MISO) system, which can also form the basis for a model predictive control system.

---

**CONTROLLER DESIGN**

Since the plant is bound to drift, the model will have to be updated periodically, while regular controller tuning will also be necessary. This can be done by conventional means, i.e. repeating the modeling procedure at regular intervals and retuning the controller.

---

**CONCLUSIONS**

- The classical state space approach to system identification is not robust with regard to the short noisy time series data typically generated by mineral processing operations, such as flotation plants.
- By making use of singular spectrum analysis (principal components extracted from the lagged covariance matrix of the measured data), the effect of noise on the characterization
of the dynamics of these processes can be alleviated substantially.

- Current work on the design of model-based predictive control systems based on the above-mentioned dynamic models of mineral processing plants looks promising and will be pursued in future studies.

Figure 8. Closed loop architecture of a neurocontroller designed by reinforcement learning, as proposed by Aldrich and Conradie (2000).

REFERENCES


