Non-linear system identification of an autocatalytic reactor using least squares support vector machines

by G.T. Jemwa* and C. Aldrich* Paper written on project work carried out in partial fulfilment of B.Sc. Eng. (Extractive Metallurgy) degree

Synopsis

The concepts behind support vector machines are very interesting both in theory and in practice, as they are based on a universal constructive learning procedure derived from the statistical learning theory developed by Vapnik. In this paper, their application to time series modelling is considered by means of simulated data from an autocatalytic reactor. In particular, reconstruction methods from non-linear dynamics are used to define a state space model for the process. Multivariate embedding techniques are compared to scalar embedding with respect to modelling.

Keywords: System identification, Support vector machines, Multilayer perceptrons, Non-linear dynamics

Introduction

Many chemical process systems exhibit complex non-linear dynamical behaviour. For these systems adequate mathematical models are difficult to derive from first principles. Until recently, process control strategies for such processes employed simple PID design rules or linear state space models. With the business environment becoming increasingly competitive, process control is commensurately relied on to reach and maintain new levels of plant operation. With this, the shift to model-based control systems has gained momentum. In turn, these model-predictive control systems are crucially dependent on high-quality process models, especially where complicated process systems are concerned.

A new paradigm of learning, viz. support vector machines, was established in the early 1990s. Although used at first for classification problems, they were rapidly generalized to regression problems and other predictive modelling applications. The strength of these systems lies in their ability to simultaneously mimimize the error of estimation in the training data (so-called empirical risk) and the complexity of the model (so-called structural risk). When time series models are developed, embedding of the data usually plays a crucial role in the ultimate performance of the model. However, since support vector machines can be designed to deal with sparse data (many variables and few data), it is not clear what role embedding of the data plays in the performance of these types of models. The main objective of this paper is therefore to investigate the role of various embedding strategies on the performance of support vector machines (a least squares variant is used). A well-known non-linear system is simulated as a case study and the models are benchmarked against multilayer perceptron neural networks, to get a better idea of their performance.

Reconstruction of system dynamics from observed time series

Dynamical systems can be represented by geometrical objects in coordinate spaces defined by the degrees of freedom of the system, also called state variables. In experimental systems it may be difficult or impossible to access the state space of the system. Delay co-ordinate embedding[1] allows one to relate the experimental time series and the underlying geometrical objects that describes the dynamics. In particular, given an observed time series $\{x(t)\}_{t=1}^{N}$ of length N, the reconstructed state space vector at time t is given by

$$x_t = \left[x_t, x_{t-t_d}, x_{t-2t_d}, \ldots, x_{t-(m-1)t_d} \right] \qquad [1]$$

where m and t_d are respectively referred to as the embedding dimension and the time delay. Embedding theorems require that $m > 2D + 1$ where D is the actual state space in which the original system is defined. Because the

* *Department of Chemical Engineering, University of Stellenbosch, Matieland, South Africa.*

 SA ISSN 0038–223X/3.00 + 0.00. Paper received Jan. 2003; revised paper received Feb. 2003.

underlying state space is inaccessible (and hence D is unknown) and the observed data are noisy and finite, heuristic algorithms have been developed that are useful in determining the embedding parameters m, and t_d. These approaches include the method of average mutual information and the decay of the autocorrelation function for determining t_d, and the method of false nearest neighbours and its variants for selecting an optimal m[2]. Essentially, whatever approach is used, it must ensure that successive components in the reconstructed vector space are as independent as possible and that the attractor is completely unfolded in the embedded space. (The attractor is the geometrical object in state space that the dynamics of the system asymptotically collapse onto.) Since the same t_d is used between successive vector components we refer to the approach as uniform (delay-coordinate) embedding.

The delay-coordinate embedding theorems assume that an infinite amount of noise-free data are available, and therefore, guarantee the capture of all the dynamics using only a single or scalar time series. In practice, these assumptions are never met. Measured time series invariably are contaminated by both measurement and dynamical noise. This artefact results in uncertainties in the reconstructed attractor. Also, a scalar time series may not be able to explain all the dynamics if it is not related to some of the underlying state variables[3,4]. Since in most processes more than one measured response variable is available, it is logical to attempt and exploit redundant information available in all measured time series. By concatenating previously embedded individual time series, a high-dimensional reconstructed state space is obtained. By exploiting correlation that may exist among multiple observed time series, it is possible to obtain lower-dimensional embeddings using dimensionality reduction algorithms, such as principal component analysis.

System identification

The system identification problem is concerned with the search for a mathematical model for dynamical systems using only observed data[5]. The best model with respect to a specified criterion is chosen from a hypothesis space that contains possible model structures. The hypothesis space is selected according to *prior* knowledge of the system generating the data.

Traditionally, in system identification the data are assumed to be generated by an unknown *linear* system. However, most processes are typically non-linear, which makes the linear assumptions restrictive in real applications. By assuming non-linearity, the space in which the approximation to the underlying model can be found is enlarged. Several model structures for non-linear hypothesis spaces have been proposed in the literature. In this study we investigate the predictive modelling capabilities of a relatively new kernel-based learning machine called the least squares support vector machine (LSSVM). A class of neural network algorithms called multilayer perceptrons are used as a benchmark.

Feed-forward multilayer perceptron neural networks

Multilayer perceptrons (MLPs) are a subset of powerful universal non-linear function estimators called neural networks[6]. Neural networks learn patterns in data using the brain metaphor and seek to learn patterns in data using parallel distributed structure. When presented with examples, or the *training data*, MLPs learn the functional map between input and output pairs using the empirical risk minimization principle. That is, given a hypothesis space, the ERM principle selects the best possible model that minimizes the error between predicted and target output.

The MLP architecture is composed of three sub-structures, viz. an input layer, one or more hidden layers, and output layer. It is common to include a bias node for each hidden layer. Interneuron connection strengths or synaptic weights are used to store knowledge obtained from a *learning* process. The non-linearity is introduced by defining appropriate basis functions for the hidden layer nodes, usually the sigmoid functions. The weights and bias terms constitute the function parameters that are estimated iteratively from the training data.

Although MLPs can approximate any non-linear function, the training algorithms are prone to trapping in local minima of the error surface of the cost function used. Also they have poor generalization capabilities—a direct result of the use of the ERM principle. Figure 1 shows a typical MLP architecture.

Least squares support vector machines

Support vector machines are kernel-based learning machines proposed within the area of statistical learning theory and structural risk minimization (SRM)[7]. The SRM principle seeks an upper bound on the expected risk or error associated with a given hypothesis space. Hence, unlike multilayer perceptrons, SVMs have better generalization characteristics. The basic idea in support vector machines is to map a low-dimensional input space into a high-dimensional space in which linear regression is possible, as illustrated in Figure 2.

The least squares support vector machines is a variant of Vapnik's standard SVM formulation, which uses an ε-insensitive loss function[8]. In the LSSVM, the ε-insensitive loss function is replaced by a squared loss function and equality constraints are used instead of inequality constraints in the formulation of the problem. Thus, given the training set $\{x_k y_k\} \in \Re^n$, k = 1, 2, … N, a non-linear mapping function $\varphi : \Re^n \rightarrow \Gamma$, that facilitates linear regression in the high-dimensional space Γ, the learning machine is defined as

$$f(x) = w \cdot \varphi(x) + b \qquad [2]$$

The separating plane is defined by w and b is some threshold. From the SRM principle, the convex optimization problem is formulated as

$$\min_{w,b,e} \Im(w,e) = \tfrac{1}{2}\left(w^T w\right) + \lambda \sum_{k=1}^{N} e_k^2$$
$$subject\ to\ y_k = w^T \varphi\left(x_k\right) + b + e_k \qquad [3]$$

The cost function with a squared error term and a regularization term λ corresponds to a form of ridge regression. Details on formulating the dual form of the problem can be found in[8–9].

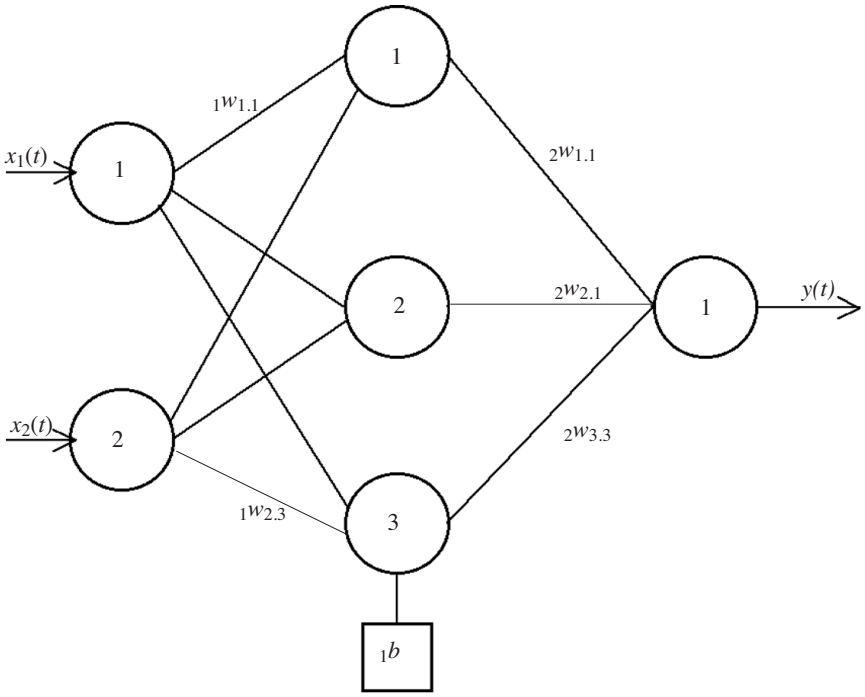

Figure 1—A MLP with a single hidden layer. Here, a 2-dimensional input space is mapped into a 1-dimensional output space via a hidden layer with 3 nodes (with non-linear basis functions) and a bias node

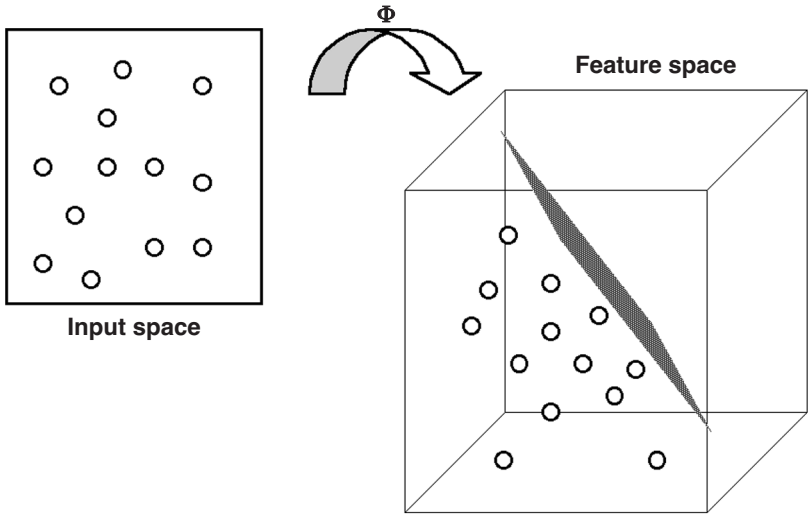

Figure 2— The basic idea in support vector learning. An input low-dimensional space is implicitly mapped by a function Φ into a high-dimensional space in which linear regression is possible

Identification of the autocatalytic reaction system

Process description

The cubic autocatalytic reactor is widely used to illustrate the occurrence of complex behaviour in chemical engineering. This includes chaotic dynamics[10]. The reactions occurring in autocatalytic continuously stirred tank reactor are described by

$$A + 2B \rightarrow 3B - r_A = k_1[A][B]^2 \quad [4]$$

$$D + 2B \rightarrow 3B - r_D = k_2[D][B]^2 \quad [5]$$

$$B \rightarrow C \quad r_C = k_3[B] \quad [6]$$

where A, B, C, and D are reacting species; $[i]$ the concentration of species i; r_i is the rate of reaction with respect to species i and k is a reaction constant. Defining dimensionless concentrations X, Y, and Z for species A, D, and B respectively, it can be shown that the model equations describing the system are given by[10]:

$$\frac{dX}{d\tau} = 1 - X - Da_X XZ^2 \quad [7]$$

$$\frac{dY}{d\tau} = \beta - Y - Da_Y YZ^2 \quad [8]$$

$$\frac{dZ}{d\tau} = 1 - \left(1 + Da_Z\right)Z + \alpha\left(Da_X XZ^2 + Da_Y YZ^2\right) \quad [9]$$

where, $\alpha = [A]_f/[B]_f$, and $\beta = [C]_f/[A]_f$ are ratios of species in the feed; τ is the dimensionless time; and Da_i the Damköhler numbers for corresponding species. The system can be shown to exhibit chaotic behaviour for $\alpha = 1.5$, $\beta = 2.81$, $Da_X = 18000$, $Da_Y = 400$ and $Da_Z = 80$[10].

Using initial random conditions and an integration time step of 0.01, time signals of length 15000 points were generated by numerically solving the system (Equations [7]–[8]) using a fifth-order Runge-Kutta ODE solver. The data were sampled at a time interval of 5 time steps. Furthermore, signals were obtained by adding normally distributed noise for different percentages of the standard deviations of individual time series (variables X, Y and Z). Figure 3 shows the original attractor of the autocatalytic process with additive Gaussian noise with a standard deviation of 20% of the standard deviation of the time series.

Reconstruction of the state space

Table I summarizes the results for the embedding parameters used in the reconstruction of the state space. The same values were assumed in all cases investigated, as it was observed that the addition of noise had little effect on the embedding parameters.

Fitting non-linear models to observed data

In fitting non-linear models, a non-linear function that maps a state vector to a future point occurring one time-step ahead was constructed using both MLPs and LSSVMs. The total data set of length 3000 points was split into two: (i) the first 2000 points for training and validation, and (ii) the last 1000 points as an independent test set for model verification.

Model parameterization is non-trivial. A good model must be able to predict future outputs of 'unseen' inputs to some degree of accuracy. At the same time, it has to explain variations in data induced by the dynamics of the system only. In other words, an optimal model has both good accuracy and generalization properties. This is achieved in different ways.

In the case of MLPs, a simple Schwarz Information Criterion (SIC) was used, which suggested an optimal MLP structure with approximately 18 nodes in the hidden layer. An optimal weight matrix was determined by 10-fold cross validation: the training data were randomly permuted and divided into 10 folds of similar lengths. For the kth (k = 1, 2, … 10) iteration the kth fold was used to validate the model trained on the other folds. The model structure with the least mean square error was chosen.

The generalization of SVMs is implicitly defined in the model formulation. However, a sparse model description is not obtained for LSSVMs. This can be done by selectively pruning those support vectors that do not contribute significantly to the model. In this case pruning did not lead to any

Table I

Determining embedding parameters

Variable	Parameter	
	T	*m*
X	1	3
Y	1	4
Z	1	8

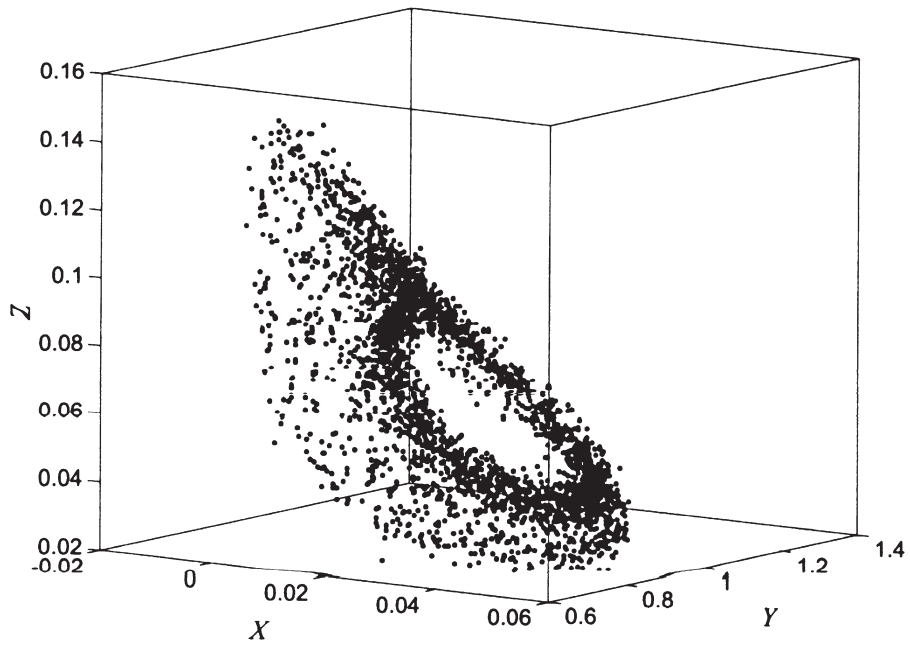

Figure 3—Original attractor of the autocatalytic process for the case of additive noise σ = 20%

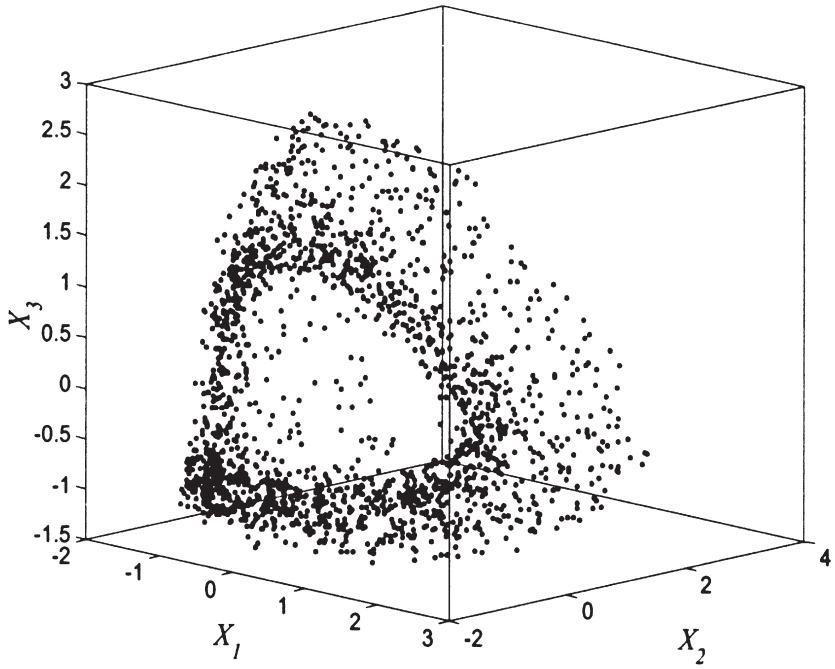

Figure 4—Reconstructed attractor of the autocatalytic process using uniform embedding of the *X* observed time series

improvement in the model. A similar weight optimization scheme as described for the MLP was used in determining the optimal hyper-parameters α,γ using grid-search numerical optimization. The following Tables II–V and Figures 5–7 summarize the results obtained in the study.

Discussion

From the results presented in Tables II–V it is evident that in all cases the LSSVM model structures performed better than the MLPs under the assumption that the training data were independent and identically distributed (hence random permutation of signals). MLPs have been reported to give significantly better predictive results in modelling this particular system[11].The only difference between the procedures used in[11] to those used here is the additional step of random permutation of the input-output pairs. Therefore, it may be argued that the MLP structure could be learning the patterns in data using more of the temporal correlations information than the underlying dynamic relationship. Otherwise, the model structure used was not within the target space of the underlying function.

Table II

One-step-ahead prediction for scaler embedding using variable *X*

Noise Level (% σ)	MLP		LSSVM	
	MSE	R^2	MSE	R^2
0	5.85E-04	0.0191	5.52E-10	1.0000
1	5.96E-04	0.1241	1.19E-07	0.9994
5	1.03E-03	0.2564	2.45E-06	0.9870
10	6.98E-04	0.0001	1.01E-05	0.9467
20	1.29E-03	0.0540	3.02E-05	0.8456

Table III

One-step-ahead prediction for multivariate embedding with *X* and *Y*

Noise Level (% σ)	MLP		LSSVM	
	MSE	R^2	MSE	R^2
0	4.10E-04	0.0255	7.42E-11	1.0000
1	4.52E-04	0.0949	3.38E-08	0.9998
5	3.38E-04	0.0342	8.31E-07	0.9956
10	1.11E-03	0.2774	3.29E-06	0.9826
20	1.45E-03	0.0767	1.37E-05	0.9299

Table IV

One-step-ahead prediction for multivariate embedding with *X* and *Z*

Noise Level (% σ)	MLP		LSSVM	
	MSE	R^2	MSE	R^2
0	3.20E-04	0.1118	7.26E-11	1.0000
1	3.47E-04	0.0917	2.77E-08	0.9999
5	6.07E-04	0.3101	6.73E-07	0.9964

Table V

Free-run prediction for scalar embedding

Noise Level (% σ)	MLP		LSSVM	
	MSE	R^2	MSE	R^2
0	5.11E-04	0.015	1.86E-05	0.9054
1	9.63E-04	0.099	2.02E-04	0.2150
5	7.06E-04	0.260	1.41E-03	0.3831
10	2.71E-03	0.000	3.88E-04	0.0135
20	2.10E-03	0.053	3.84E-05	0.0105

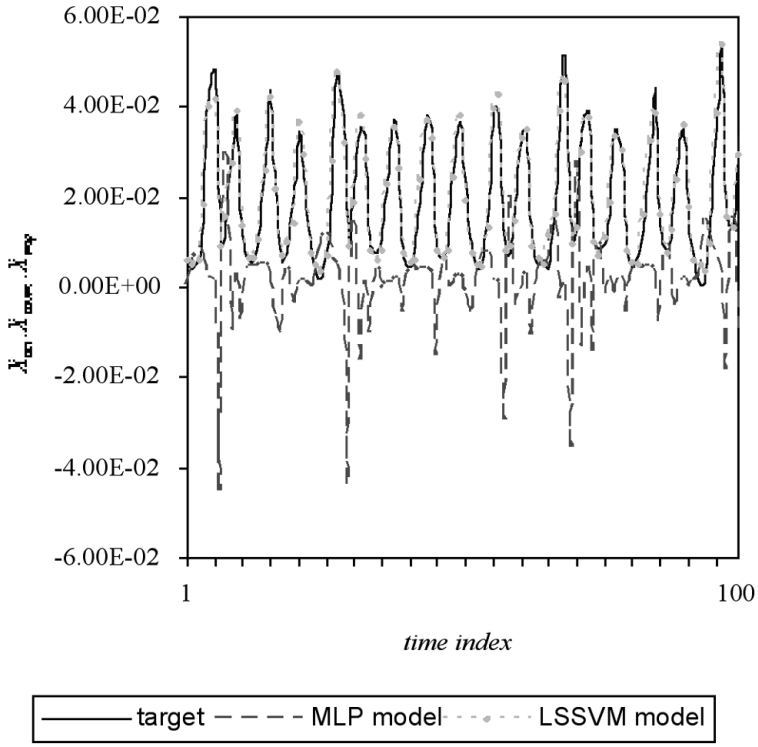

Figure 5—One-step-ahead prediction for scalar embedding with σ = 10%

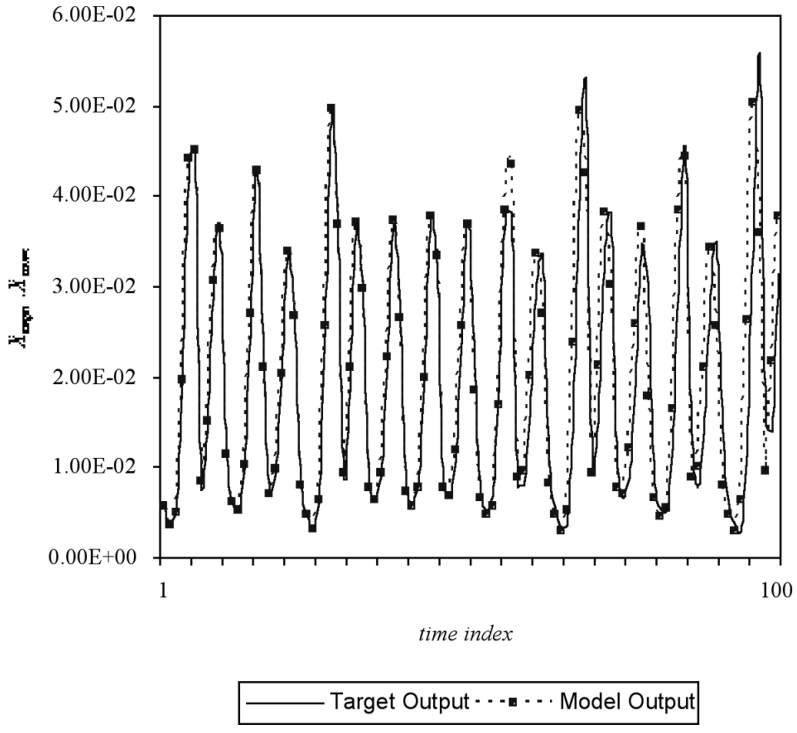

Figure 6—Free-run prediction results for an LSSVM model using a scalar embedding with σ = 0%

An interesting observation was that LSSVM model built using reconstructed state space vectors from a univariate embedding had very good long-term prediction properties, as shown in Table V*5* and Figures. 6, 7. Iterative or free-run predictions are a more powerful test of how successful the model building process has been[11–12]. A poor model tends to a fixed point on iterative prediction. As shown in Figure 6, the LSSVM model for the case shown traced the trajectory of the actual output for about 50 time steps. Furthermore, although the deviations between predicted and actual

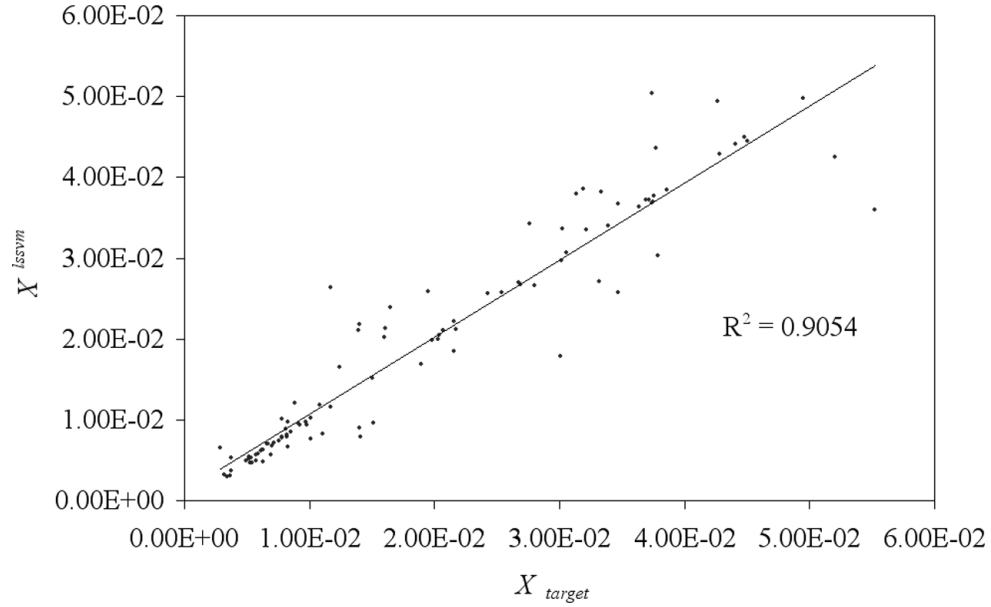

Figure 7—Correlation plots for the free-run prediction results shown in Figure 6

increased with time it is observed that the dynamical behaviour is essentially the same. The deviation could be explained by the fact that chaotic and non-linear systems are very sensitive to infinitesimal perturbations.

As expected, increasing noise levels had a negative effect on the model. Table V shows that noise levels greater than 5% of the standard deviation of the measured data results in non-robust models. It is important, therefore, to minimize the influence of noise in signals when non-linear modelling approaches such as the ones used herewith are used.

With respect to the inclusion of information from simultaneously observed response variables, Tables III and IV indicate that a multivariate embedding is potentially superior to the univariate case. This confirms results reported elsewhere[13]. Observe that when information from more than one signal is included, the effect of noise is minimized.

Conclusions

We investigated system identification of an autocatalytic reactor using non-linear state space models. It was found that under the assumption of independent and identically distributed data, the least squares support vector machine outperformed an alternative parameterized set of models, namely multilayer perceptrons. Excessive noise levels (> 5%) in observed signals had a negative effect on the building of accurate and reliable mathematical models obtained. Inclusion of information from other response variables improved model performance and, in particular, reduced the effect of noise on the obtained models.

References

1. Sauer, T., Yorke, J.A., and Casdagli, M. 'Embedology', *J. Stat. Phys,* vol. 65, nos. 3/4, 1991. pp. 579–616.
2. Kantz, H. and Schreiber, T. *Nonlinear Time Series Analysis,* Cambridge University Press, Cambridge, 1997.
3. Cao, L., Mees, A., Judd, K. 'Dynamics from multivariate time series,' *Physica D.,* vol 121, 1998. pp. 75–88.
4. Barnard, J.P., Aldrich, C., and Gerber, M. 'Embedding of multidimensional time-dependent observations , *Phys. Rev.*, E., vol. 64, no. 4, 2001. pp. 1–4.
5. Ljung, L. *System Identification*, Prentice Hall, New Jersey, 1987.
6. Haykin, S. *Neural Networks—A Comprehensive Foundation*, Macmillan College Publishing, New Jersey, 1994.
7. Cristianni, N. and Shawe-Taylor, J. *An Introduction to Support Vector Machines,* Cambridge University Press, 2000.
8. Suykens, J.A.K. 'Nonlinear modelling and support vector machines,' *Proceedings of the IEEE Instrumentation and Measurement Technology Conference,* Budapest, Hungary, May 21–23, 2001.
9. Suykens, J.A.K., De Brabanter, J., Lucas, L., and Vandewalle, J. 'Weighted least squares support vector machines: robustness and sparse approximation,' Internal Report 00-37, ESAT-SISTA, K.U. Lueven, 2000.
10. Lynch, D.T. 'Chaotic behaviour of reaction systems: mixed cubic and quadratic autocatalysis,' *Chem. Eng. Sci.*, vol. 47, no. 17/18, 1992. pp. 4435–4444.
11. Barnard, J.P. Empirical State Space Modelling with Applications in Online diagnosis of Multivariate Dynamic Systems, Ph.D. thesis, University of Stellenbosch, 1999.
12. Judd, K., Small, M., and Mees, A.I. 'Achieving good nonlinear models: Keep it simple, vary the embedding and get the dynamics right,' in *Nonlinear Dynamics and Statistics,* A.I. Mees, Ed. Boston: Birkhauser, 2001, pp. 65–80.
13. Jemwa, G.T. Multivariate Nonlinear Time Series Analysis of Dynamic Process Systems, M.Sc. Eng thesis, University of Stellenbosch, 2003. ◆

III International Congress of Prospectors and explorers ProEXPLO 2003—Peru*

Over 100 milllion dollars will be invested in mining exploration in Peru this year

Between 100 and 120 million dollars will be invested this year in mining exploration in Peru, seeking to generate a local market and convert exploration into a business in its own right, accelerating the rate of discovery of new deposits and strengthening the development of national mining.

This fact was stressed today by Dr Miguel Cardozo Goytizolo, Chair of the organizing committee of the III International Congress of Prospectors and Explorers ProEXPLO 2003, which will be held in Lima from April 22nd through 25th next, under the motto **Exploration: The Future of Mining?**

He stated that this International Congress will have an innovative technical programme and that the papers will be delivered by specialists of international prestige in the various branches of geology applied to mining exploration.

ProEXPLO 2003 will favour development becoming a fruitful local and regional business. He also stated that there will be professionals experienced in the financing of mining exploration elsewhere who will be able to apply their experience in Peru and other countries in the region.

Cardozo maintained that investments in mining exploration are carried out despite the notable crisis in the sector due to the fall in the price of minerals and the reduction in the number of companies involved in the mining business.

He explained that according to figures given by the Metals Economics Group (MEG), world-wide expenditure in exploration during 2001 was only 2,200 million, the lowest in the last 10 years, amounting to a reduction of 15% compared to the previous year and 58% in comparison with 1997.

However, Peru continues to be one of the most attractive countries for the exploration market in Latin America. In 2001 it regained the 4th place which it first attained in 1999, and displaced Chile as a destination for exploration investment.

During 2001, Latin America received approximately 634 million dollars in exploration, 130 of which came to Peru; and the figure for this year is estimated at between $100 million and 120 million, taking into account that Peru is one of the main destinations for this activity.

According to the MEG, five large companies disappeared from the exploration market during the year 2000 due to mergers and acquisitions; and during 2001 the budgets of merged companies have been considerably smaller. It is foreseen that exploration budgets will continue to decrease.

Furthermore, he stated that Peru, following the example of leading countries such as the United States, Canada and Australia, should direct its efforts towards making exploration a 'mass product', turning it into a sustainable business, which would accelerate the discovery of new deposits.

He said that if exploration effort is limited to the capacity of international companies which operate in the country, the rate of discovery will continue to be very slow. However, if explorers go out to seek new deposits in the field without necessarily working for the large companies, the process can become more rapid.

He revealed that this will be one of the important topics dealt with at ProEXPLO 2003. Talks are being held between representatives of the Toronto Stock Exchange, which will send speakers to the event, and the Lima Stock Exchange, to generate debate and take actions such that exploration in Peru can become a great business.

ProEXPLO 2003 will be a great forum to learn what is going on in Peru and the region as a whole. It will provide information on the state of projects throughout the region, those which have progressed, those which have better prospects, their geological characteristics, and where the greatest exploration effort is headed.

There will be information on technical progress in explorations, geophysics, geochemistry and structural geology, specialist technical topics which are making rapid progress worldwide and in which it is vital to share experience and knowledge.

ProEXPLO 2003 papers will be given by geologists of international prestige and will focus on subjects such as: The latest discoveries and exploration projects in the region, Geological and exploration concepts and models at local and regional level, and the role of geological services in encouraging exploration.

Other subjects to be dealt with are: The role of structural geology in mining exploration and its possibilities of predicting the location of new ore deposits; the renovation of traditional exploration tools and cutting-edge technology which makes remarkable new discoveries possible; the exploration business, risk, financing sources, project value, global, regional and local trends in investment, etc. environmental protection and the sustainable development aim.

Complementary to the papers, there will be valuable pre- and post-congress courses as well as guided field visits to ore deposits in the region.

In addition, the Technical Exhibition which was highly successful in the previous congress, will be one of the main attractions in ProEXPLO 2003, providing a meeting point between suppliers of equipment and services, and mining and exploration companies.

The main exploration projects and deposits in operation will be present at the Core Shack, which will exhibit a wide range of drill cores for observation and discussion

The event will draw to Lima a large number of geoscientists, prospectors, executives of companies that are world leaders in exploration, suppliers, government entities and investors from the world over. ◆

---

* *Contact: Miguel Cardozo, President of the Organizing Committee, ProEXPLO 2003*
www.iimp.org.pe