ON THE SELECTION OF VARIABLES FOR QUALITATIVE MODELLING OF DYNAMICAL SYSTEMS

JOSEP MARIA MIRATS I TUR[a,*], FRANÇOIS E. CELLIER[b,†], RAFAEL M. HUBER[a,‡] and S. JOE QIN[c,¶]

[a]*IRII, Institut de Robòtica i Informàtica Industrial, Universitat Politècnica de Catalunya—Consejo Superior de Investigaciones Científicas, 08034 Barcelona, Spain;* [b]*Electrical and Computer Engineering Department, The University of Arizona, Tucson, AZ 85721, USA;* [c]*Department of Chemical Engineering, University of Texas at Austin, Austin, TX 78712, USA*

Behavioural modelling of physical systems from observations of their input/output behaviour is an important task in engineering. Such models are needed for fault monitoring as well as intelligent control of these systems. The paper addresses one subtask of behavioural modelling, namely the selection of input variables to be used in predicting the behaviour of an output variable. A technique that is well suited for qualitative behavioural modelling and simulation of physical systems is Fuzzy Inductive Reasoning (FIR), a methodology based on General System Theory. Yet, the FIR modelling methodology is of exponential computational complexity, and therefore, it may be useful to consider other approaches as booster techniques for FIR. Different variable selection algorithms: the method of the unreconstructed variance for the best reconstruction, methods based on regression coefficients (OLS, PCR and PLS) and other methods as Multiple Correlation Coefficients (MCC), Principal Components Analysis (PCA) and Cluster analysis are discussed and compared to each other for use in predicting the behaviour of a steam generator. The different variable selection algorithms previously named are then used as booster techniques for FIR. Some of the used linear techniques have been found to be non-effective in the task of selecting variables in order to compute a posterior FIR model. Methods based on clustering seem particularly well suited for pre-selecting subsets of variables to be used in a FIR modelling and simulation effort.

Keywords: Fuzzy inductive reasoning; Variable selection; Behavioural modelling; Inductive modelling; Qualitative modelling; Input/output modelling

1. INTRODUCTION

Intelligent controllers frequently operate with look-ahead data in order to compensate for system delays and/or improve their performance. For example, the controllers that regulate the water distribution system of a city may, on the one hand, work with predicted values of water flows, because the water incurs a delay from the time it is released at the reservoir until it arrives at the city where it is to be used; and on the other hand, they may work with predictions of water needs at the time when the water that is currently being released will

*Corresponding author. Tel.: +34-93-401-58-05. E-mail: jmirats@iri.upc.es
†E-mail: cellier@ece.arizona.edu
‡Tel.: +34-93-401-57-57. E-mail: rhuber@iri.upc.es
¶E-mail: qin@che.utexas.edu

arrive at the city. Hence, tools and techniques for predicting future values of observed trajectory behaviour constitute important elements of intelligent control architectures.

A difficult problem when trying to model the output or outputs of a system from its inputs, is knowing which inputs to use in order to make a good prediction of the output or outputs. All potential inputs are not always necessary, because some may be redundant, whereas others may not provide information that is useful for predicting the behaviour of the output or outputs of the system being studied. The problem becomes worse when dealing with large-scale systems, such as nuclear power plants, airplanes, or water distribution systems, since the list of potential inputs may be formidable.

Many approaches for the selection of variables have been presented in the literature using classical and Bayesian statistical techniques as well as other mathematical modelling tools such as neural networks. Principal Components Analysis (PCA) has been studied with this aim by Jolliffe (1972; 1973), applied to artificial as well as real data sets. In Allen (1971), variable selection is performed using the mean square error of prediction of different possible regression models. In Mansfield *et al.* (1977), a regression model based on principal components is suggested, in which one variable is temporarily eliminated at a time computing the least square error for each of these regression models. The model offering the smallest least square error is then selected, and the corresponding variable is permanently discarded. The procedure is repeated until the smallest least square error becomes too big. The admissible procedures to perform variable selection when a regression model is used are analysed and discussed by Kabaila (1997). Different regression methods are compared in the task of selecting variables by Hoeting *et al.* (1996), McShane *et al.* (1997) and Adams and Allen (1998). Chipman *et al.* (1997) and Hoeting and Ibrahim (1998) used other approaches such as Bayesian and heuristic techniques with the purpose of selecting variables of a system. The canonical correlation analysis is explained and used to select variables in a study of Al-Kandari and Jolliffe (1997). Also, work has been reported in this area using neural networks (Lisboa and Mehri-Dehnavi, 1996; Muñoz and Czernichow, 1998).

In this paper, a qualitative modelling methodology, called Fuzzy Inductive Reasoning (FIR) (Cellier, 1991), is investigated with the aim of providing forecasts of trajectory behaviour of measured variables for control purposes. In particular, the paper deals with the problem of pre-selecting a set of candidate input variables in order to reduce the model search space of FIR. This is important since FIR employs an algorithm of exponential computational complexity in the identification of the best qualitative input/output model. Section 2 provides a brief review of this modelling technique and applies it to compute both dynamic and static models of an industrial steam generator process.

Then, different procedures are applied to first, model the steam generator process, and, afterwards, perform variable selection as a booster technique for FIR. The method of the unreconstructed variance for the best reconstruction, described by Dunia and Qin (1998) in a different context, is one of the methods presented for the purpose of selecting which input variables to use to model a given system output. This technique is reviewed in Section 3 of the paper, while in Appendix A, a brief review of its mathematical underpinnings is given. Section 4 discusses the use of statistical methods, based on regression coefficients, for the purpose of selecting input variables. Section 5 provides a brief overview of other variable selection methods advocated in the open literature. All statistical techniques presented in this paper employ only *static* models so the predictions made by those techniques, as presented in Sections 3–5, are obviously worst when compared against the prediction made by the dynamic FIR model in Section 2.

Yet, the purpose of this paper is not to compare the different predictions of the steam generator data the advocated methods are capable to give. The purpose is to find a set of bootstrapping techniques for FIR that, with little computational effort, can encounter subsets

of variables to be considered in a FIR optimal mask search. In Section 6 the different variable selection approaches presented in earlier sections, used to find static models, are applied to the problem of qualitatively modelling the behaviour of a steam generator (boiler) of a chemical process using FIR. To this end, the sets of proposed inputs obtained by the different variable selection algorithms are offered to the modelling engine inside FIR, and the predictions obtained by the FIR dynamic qualitative models when using the so obtained static sets are compared to each other.

2. FUZZY INDUCTIVE REASONING

2.1. Methodology Review

A very brief verbal review of the FIR methodology is provided in this section. For a deeper insight into the methodology, the reader is encouraged to review the referenced publications.

Inductive reasoning (IR) is an inductive modelling technique designed by Klir as part of his General System Problem Solving (GSPS) framework (Klir, 1985). A first implementation of IR was made available by Uyttenhove as part of his Ph.D. dissertation (Uyttenhove, 1979). This implementation was called Systems Approach Problem Solver (SAPS). An improved version of the original SAPS program was developed by Cellier and Yandell (1987), and later extended in Li and Cellier (1990) to offer fuzzy reasoning capabilities. Accordingly, the enhanced methodology is now called FIR (Cellier *et al.*, 1992). In the sequel, a number of different authors used FIR to qualitatively model and simulate different kinds of systems and time series, while constantly improving the methodology (Mugica, 1995; de Albornoz, 1996; López 1999; Mirats Tur and Huber 1999). The most recent comprehensive description of the FIR methodology can be found in Nebot *et al.* (1998).

FIR operates on observations of input/output behaviour of a system, or on observations of time series. In order to qualitatively reason about these observed behaviours, real-valued trajectory behaviour needs to be fuzzified, i.e. mapped into a set of fuzzy classes. In FIR, the process of fuzzification is called *recoding*. In this process, real-valued data are mapped into qualitative triples, consisting of a class value, a fuzzy membership value, and a side value. The side value is a speciality of the FIR methodology. It describes whether the original real-valued data point lies to the left, at the centre, or to the right of the maximum of the Gaussian membership function governing the chosen class. The side value makes it possible to defuzzify qualitative triples into real-valued quantitative data without information loss.

In FIR, quantitative data are usually recoded into either three or five classes using equal frequency partitioning to determine the landmarks (borderlines) between neighbouring classes. Once the landmarks have been found, the Gaussian fuzzy membership functions associated with each class are automatically determined by letting them assume a maximum value of 1.0 in the centre between the two landmarks that limit the class, and by letting them decay to a value of 0.5 at the two landmarks themselves.

FIR operates initially on a real-valued raw data matrix. Each column of this matrix contains one equidistantly sampled trajectory of an observed variable, whereas each row contains one time-stamped record of all the observed variables. In the process of recoding, the raw data matrix is converted to three separate matrices: a multi-valued qualitative class matrix, a real-valued fuzzy membership function matrix with values ranging between 0.5 and 1.0, and a ternary side matrix.

Once the observed data have been recoded, FIR attempts to find behavioural patterns among the observations, using the information stored in the class values. To this end, it tries to find relationships among these class values that are as deterministic as possible. For example,

a set of observations may contain five variables named x_1, x_2, x_3, x_4, and y. A qualitative model is to be determined that is capable of predicting future values of y as a function of past values of y as well as past and future values of x_1, x_2, x_3, and x_4. FIR may determine that the most successful predictions of y at the current time can be made by making use of x_3 two time steps back, x_1 and x_4 one time step back, as well as y one time step back. Such a relationship can be written as

$$y(t) = f\{x_3(t-2), x_1(t-1), x_4(t-1), y(t-1)\}$$

where f is a qualitative tabular function specified by the observations made, i.e. by means of the training data.

In FIR, such a qualitative relationship is encoded in the form of a so-called *mask*. A mask is a matrix that contains as many columns as there are observed variables, and as many rows as the qualitative relationship covers time instants. Inputs of the qualitative relationship (so-called *m-inputs*) are encoded as negative integers, whereas the mask output (the so-called *m-output*) is encoded as $+1$. The mask corresponding to the previously introduced qualitative relationship is shown below.

$$\begin{array}{c} \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & y \end{array} \\ \begin{pmatrix} 0 & 0 & -1 & 0 & 0 \\ -2 & 0 & 0 & -3 & -4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Possible mark

FIR searches for the best qualitative relationship (qualitative model) by either an exhaustive search or one of several heuristics applied to a set of mask candidates. The set of mask candidates is encoded in the form of a so-called *candidate mask* that contains -1 elements at the locations of potential m-inputs, and a $+1$ at the location of the m-output. A possible candidate mask for the five-variable system is shown below.

$$\begin{array}{c} \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & y \end{array} \\ \begin{pmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 \end{pmatrix} \end{array}$$

Candidate mask

The index used to compare the masks is an entropy-based measure called the *quality of the mask* in the FIR context. The Shannon entropy measure is used to determine the uncertainty associated with predicting a particular output given any legal input state. The optimality of the mask is evaluated with respect to the maximisation of its predictive power. Details of how the quality of each mask is determined can be found in Cellier (1991) and Nebot *et al.* (1998). The optimal mask represents a compromise between predictiveness and specificity of the model. A simpler mask, i.e. a mask with a smaller number of m-inputs, makes it easier to make predictions, but the predictions obtained in this way are not very specific. On the other hand, a more complex model, if applicable, can make highly specific predictions, but often, there may not be enough evidence gathered from the training data to justify such a prediction.

Once the optimal mask has been found, it can be used to flatten dynamic relations into static ones. To this end, the mask is shifted along the data matrix during k time intervals in order to construct the so-called behaviour matrix embracing the behaviour of the system

(the behaviour of the system is learnt by means of the optimal mask). In fact, there are three separate matrices making up the behaviour: a class, a membership and a side matrix. In the behaviour matrix, columns represent the m-inputs and the m-output of the mask, and rows represent fuzzy rules that can be linguistically sorted. Hence, the behaviour matrix constitutes a fuzzy rule based that FIR automatically synthesizes from the training data and the optimal mask.

Once the fuzzy rule base has been synthesized, it is possible to make predictions. To this end, a new record of m-inputs (a testing data record) is compared with the m-input records contained in the fuzzy rule base. The five nearest neighbours are found. The predicted m-output is then computed as a weighted average of the m-outputs of the five nearest neighbours among the training data, whereby the weights are determined based on the relative relevance (proximity or similarity) of each of the five nearest training data to the testing data record in the m-input space.

The high quality of predictions obtainable using the FIR methodology has been demonstrated in many publications. In addition, FIR has been compared both qualitatively and quantitatively to other competing methodologies in López (1999). The results obtained by López show that FIR is indeed among the very best qualitative forecasting tools available today.

The problem with the methodology lies in the computational complexity of its algorithms. The exhaustive model search algorithm is of exponential complexity, and even the currently implemented heuristics are of polynomial complexity, and therefore quite slow. Whereas FIR works exceedingly well when applied to a five-variable system, it is doomed to failure when dealing with a 50-variable system.

The purpose of the ongoing investigation is to find variable pre-selection algorithms of lower computational complexity than FIR that would permit to automatically extract a subset of variables from the total set of available sensory information that would contain good candidates of m-inputs that, in a subsequent FIR analysis, would lead to qualitative models with high predictiveness and specificity.

2.2. FIR Model of a Steam Generator

Data from a steam generator process will be used throughout the report as an example, in order to compare the results obtained using different methods of variable selection. Using a sampling interval of 5 min, 632 data points were collected during a period of significant change in the boiler throughput so as to cover a wide range of the process behaviour. Figure 1 shows a schematic of the boiler process. Although no methodology can be fully tested if it is only applied to just one data set, the given process can be seen as a fairly generic process in the industry. When working with industrial processes we normally find big plants with lots of variables to measure, somehow related between them usually under the rules of determinate physical laws. This is the case of the studied process.

The variable to be predicted is the *NOx* content sampled from the boiler stack. Eight input variables are considered to have influence on the *NOx* emission level. Table I shows the considered variables.

A FIR dynamic model was constructed using 85% of the gathered data. The remaining 15% of the measurement data were used to validate the model. Since the system under investigation contains only nine variables, the FIR approach can be used directly to model the system under study, and to select a group of input variables to predict the output of the boiler.

It was decided to recode each of the nine variables separately into three classes using equal frequency partitioning to determine the landmarks. A mask candidate matrix of depth 5 was

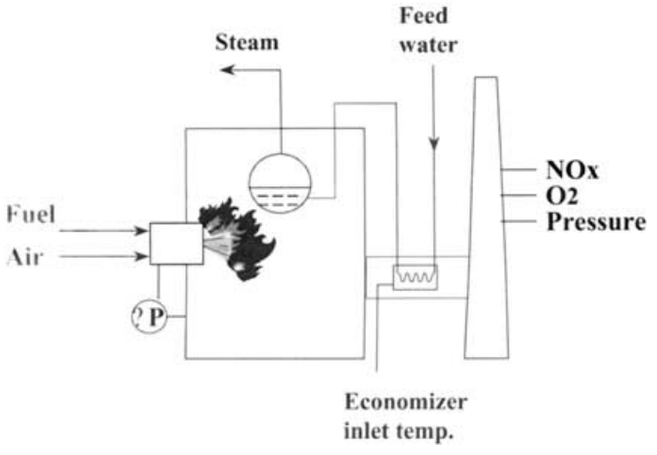

FIGURE 1 Schematic for the boiler process.

proposed, postulating that significant m-inputs may not lag more than 20 min behind the m-output. All elements of the mask candidate matrix were preset to -1, except for the element (5,9), which was set to $+1$, as it represents the location of the m-output. The optimisation problem was solved using exhaustive search, except that the maximum allowed complexity of the mask (the maximum number of mask elements different from 0) was limited to five. The set of best masks of each complexity was retained as promising masks to be investigated further.

The retained masks are shown below. The mask of complexity 4 exhibits the highest quality, followed by the mask of complexity 5, followed by that of complexity 3. The mask of complexity 3 treats the *NOx* level as a univariate time series, since it proposes that the future behaviour of the *NOx* level can best be predicted taking into consideration its own past behaviour only.

$$m4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

TABLE I Variables of the boiler system

Variable	*Physical meaning*
1 (input)	Airflow (KPPH)
2 (input)	Fuel flow (Pct)
3 (input)	Stack oxygen (%)
4 (input)	Steam flow (KPPH)
5 (input)	Economiser inlet temp. (F)
6 (input)	Stack pressure (in H_2O)
7 (input)	Windbox pressure (in H_2O)
8 (input)	Feedwater flow (KPPH)
9 (input)	NOx, PPM

$$m5 = \begin{pmatrix} 0 & 0 & -1 & 0 & 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -4 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$m3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

None of the masks made use of any variables except variables 2, 3, 8, and 9. Hence, any successful variable pre-selection algorithm should retain these same variables, filtering out the ones that FIR would not consider in an optimal mask analysis.

FIR can itself be used as a variable pre-selection algorithm, as long as the overall number of variables is not too large. In the example presented here, it is not evident that a mask depth of five suffices to capture the best possible masks. Yet, a candidate mask of greater depth would already in the case of a 9 variable system lead to unacceptably large optimisation cost. Hence, the previously obtained results were used to postulate a new candidate mask, this time of depth 16 spanning a time period of 75 min, in which the variables 1, 4, 5, 6, and 7 were disabled by presetting all elements of the mask candidate matrix located in those columns to 0.

The same optimisation approach was used as before to determine the set of best masks. Evidently, the masks found earlier are still within the search space, i.e. if other masks are retained, they must be better than those found earlier. The three retained models are as follows:

$$\text{complexity 5:} \quad y = f\{x_3(t), x_3(t-15), x_8(t-9), y(t-1)\}$$

$$\text{complexity 4:} \quad y = f\{x_2(\mathrm{t}-1), x_3(t-15), y(t-1)\}$$

$$\text{complexity 3:} \quad y = f\{y(t-1), y(t-5)\}$$

None of the retained masks is identical to any of the ones found earlier, i.e. the larger mask depth indeed paid off. As before, FIR chose an autoregressive model in the case of the mask of complexity 3.

Figure 2 shows the output data validation set, i.e. the last 15% of the available data in continuous line, and the prediction of the output variable, depicted in dashed line, using the three retained masks. To this end, another facet of the FIR methodology is being used. In a prediction, i.e. a qualitative simulation, FIR not only forecasts future values of the m-output; in addition, it generates estimates of the quality of these predictions in the form of a confidence value. How FIR estimates the confidence in its own predictions is explained in detail in López (1999). The predicted value at each time instant is determined as follows. In each simulation step, the three masks are used to compute three separate predictions

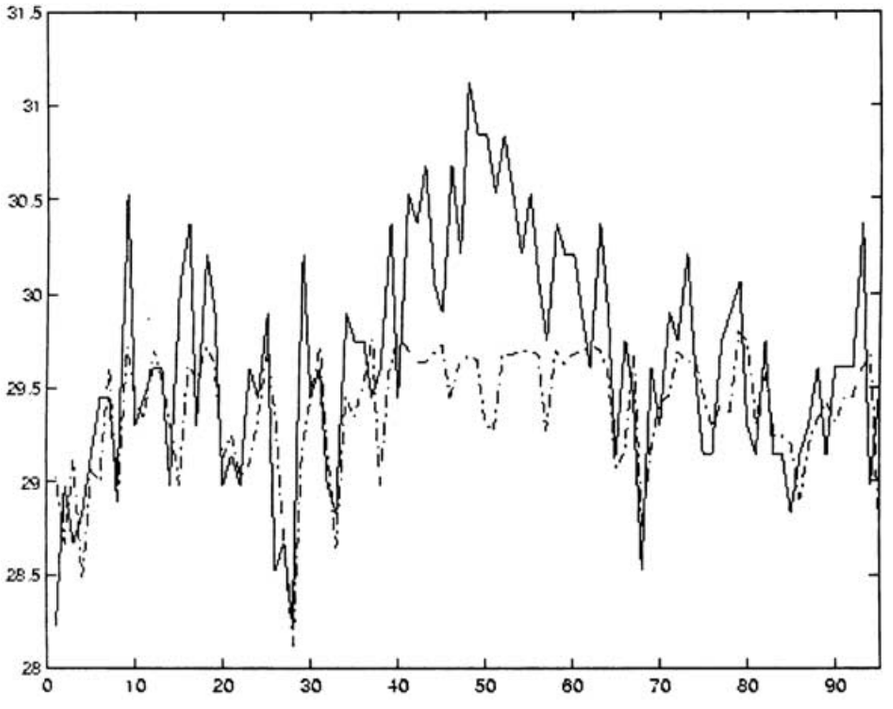

FIGURE 2 Original (continuous line) and predicted (dashed line) validation set for the output using FIR.

of the m-output value. Each prediction is accompanied by a confidence value. Here, the prediction with the highest confidence value has been retained as the true prediction for that step. Of course, the true prediction could have been established using other approach as for example weighting all the obtained predictions by their confidence, but how the prediction is to be obtained lies beyond the scope of this paper.

The reader may notice that no *NOx* value beyond 29.7 PPM was ever predicted, i.e. whereas the lower *NOx* levels are predicted (dashed line) fairly accurately, the higher values are not. The reason is that the training data do not contain any *NOx* levels beyond 29.7 PPM. FIR can only predict patterns that it has observed before. Since it has never seen such high *NOx* levels, it cannot predict their existence. The MSE error of this forecast is 0.5522.

It may be important to note that, for all other methods presented in this report, the original data have to be normalised to zero mean and unit variance. With the FIR methodology, this is not necessary. This explains why the ordinate axis in Fig. 2 has a different scale than for all other methods. In order for the prediction errors to be comparable, it is necessary to divide the FIR MSE value by the variance of the output. The normalised FIR MSE error assumes a value of 0.2817.

2.3. FIR Models Excluding Temporal Relations

In order to be able to compare the quality and the computing reduction achieved when using different variable selection techniques, static FIR models of the boiler are needed, that is, models excluding temporal relationships. The candidate mask of depth 1 proposed for this

TABLE II FIR models without temporal relations

Complexity	*Model*									*Quality*	*MSE*
2	(0	0	0	0	0	0	0	−1	1)	0.1532	1.1510
3	(0	0	0	0	0	0	−1	−2	1)	0.1523	1.2087
4	(0	0	−1	−2	0	0	0	−3	1)	0.1418	1.0974
5	(0	−1	−2	0	−3	0	0	−4	1)	0.0900	0.8234

purpose is:

$$\text{mcan} = (-1 \quad -1 \quad -1 \quad -1 \quad -1 \quad -1 \quad -1 \quad -1 \quad 1)$$

Table II shows the retained masks for each of the allowed complexities, as well as the normalised MSE value of the prediction using each one of them.

Since the search space is now much smaller, the optimal masks can be found much more rapidly. However, the information obtained is also less valuable. FIR concludes that variables 1 and 6 can be safely discarded, whereas all other variables need to be retained for the time being. All static models found are of fairly low quality, and the MSE values resulting from their use are consequently rather high.

Notice that, in the FIR methodology, the quality of masks can only be truly compared to each other as long as their complexities are the same. Moreover, for a given mask, the MSE value depends on the validation data set, and both considerations must be taken into account when attempting to extract conclusions from Table II.

3. METHOD OF THE UNRECONSTRUCTED VARIANCE FOR THE BEST RECONSTRUCTION

The method described in this section was developed at the University of Texas in Austin (Dunia, 1997; Dunia and Qin, 1998; Qin and Dunia, 1998). It was previously used to select the number of principal components to keep in a PCA model, based on the best reconstruction of the variables. The purpose of the PCA model was to identify faulty sensors in a system, and to reconstruct sensor data values from measurements of sensors attached to other signals, exploiting the redundancy inherent in multiple sensor data streams. The methodology had not been designed as a tool for finding an input/output model of a system, though the two tasks are evidently related to each other.

When a PCA model is used to reconstruct missing or faulty values, the reconstruction error is a function of the number of intervening principal components. In order to determine the number of principal components (PCs) to be used, the methodology proposes making use of the variance that the model cannot reconstruct; that is, it uses the variance of the reconstruction error.

Prior to determining the number of PCs to be retained in the model, the available measurement data are analysed to determine what variables are well reconstructed from which others. For the PCA analysis to be applicable, the data must first be normalised to zero mean and unit variance. Given a system with k variables, every variable is reconstructed from the other $k - 1$ variables, and its unreconstructed variance is computed as a function of the number of retained principal components. The results are tabulated as shown in Table III for the case of the boiler data. In order to obtain a fair comparison, only the training data were used in the analysis.

The method then proceeds by summing up the unreconstructed variances of each column. The number of PCs for which the sum of the unreconstructed variances is a minimum is

TABLE III Unreconstructed variance

Var \ *#PC*	1 *PC*	2 *PCs*	3 *PCs*	4 *PCs*	5 *PCs*	6 *PCs*	7 *PCs*	8 *PCs*
X_1	0.0128	0.0128	0.0047	0.0036	0.0033	0.0025	0.0024	0.0062
X_2	0.0190	0.0118	0.0078	0.0074	0.0063	0.0047	0.0036	0.0107
X_3	0.9957	18.2113	1.4427	1.0621	1.0522	1.0407	0.8821	6.8374
X_4	0.0109	0.0079	0.0050	0.0038	0.0036	0.0032	0.0019	0.0019
X_5	0.0305	0.0304	0.0286	0.0283	0.0551	0.0538	0.0357	0.0508
X_6	0.0338	0.0335	0.0229	0.0189	0.0190	0.2857	0.2253	0.3810
X_7	0.0307	0.0306	0.0121	0.0120	0.0105	0.0107	0.0266	0.0278
X_8	0.0688	0.0676	0.0630	0.2825	0.2822	1.7516	4.7219	14.4137
Y	0.5763	0.5768	2.2764	2.2718	2.2282	2.2844	12.3508	147.682

determined. It turns out that, in this system, only one PC needs to be retained. The method then throws out those variables that exhibit, for the optimal number of retained PCs, unreconstructed variances that are bigger than the value that would be obtained if the measurement data for those variables were replaced by their mean values. Due to normalisation, the theoretical threshold value is 1. In practice, it may be better to throw out variables with an unreconstructed variance above 0.8 or 0.9.

In the boiler example and looking at the 1 PC column of Table III, it can be seen that variable 3 needs to be thrown out. This means that any variable of the system, except for variable 3, can be reconstructed using a PCA model with a single PC that is made up from the information available through all variables except for variable 3, which should be ignored.

The mathematical underpinnings of the methodology outlined in this section are summarized in Appendix A.

3.1. Modelling *NOx* Output Using the Unreconstructed Variance Method

It can be seen from Table III that reconstructing the *NOx* level (variable 9) from the other variables is considerably more difficult than reconstructing any of the other signals with the exception of variable 3. Yet, the approach suggests that it is meaningful to construct a PCA model for variable 9, using a single PC made up from variables 1, 2, 4, 5, 6, 7, and 8.

The results of the analysis are shown in Fig. 3. The continuous line shows the validation data of the *NOx* measured variable. The dashed line shows the predictions made by the PCA model made up from the training data using a single PC composed from the variables 1, 2, 4, 5, 6, 7, and 8. The MSE value obtained from this model for the training data period is 0.7073, and for the validation period, the one depicted in the figure, it is 0.9979. The reader may notice that the prediction is indeed better for the training data than for the validation data.

The MSE value is considerably higher than in the case of the FIR model. The reason is that the PCA model makes no attempt at replicating the high frequency oscillations exhibited by the real data. The model has clearly low-pass characteristics. Hence, even if the prediction looks good to the naked eye, the distances between the prediction and the real data are large most of the time, and consequently, the MSE value cannot be made small.

Did the reconstructed variance analysis do a good job at deciding which variables to retain in the PCA analysis? To answer this question, a second PCA model was made, also consisting of a single PC, but made using the variables 2, 3, and 8, as proposed by FIR. The results of this prediction are shown in Fig. 4. This time, the MSE values are 0.5703 for the training data set, and 0.8306 for the validation data set.

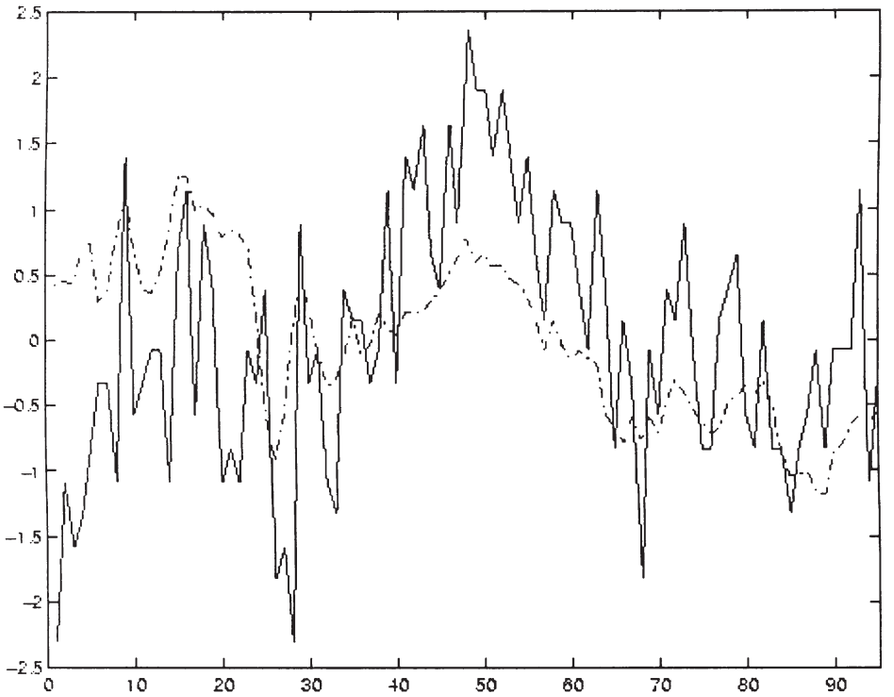

FIGURE 3 Prediction given by a PCA model using input variables 1, 2, 4, 5, 6, 7, and 8.

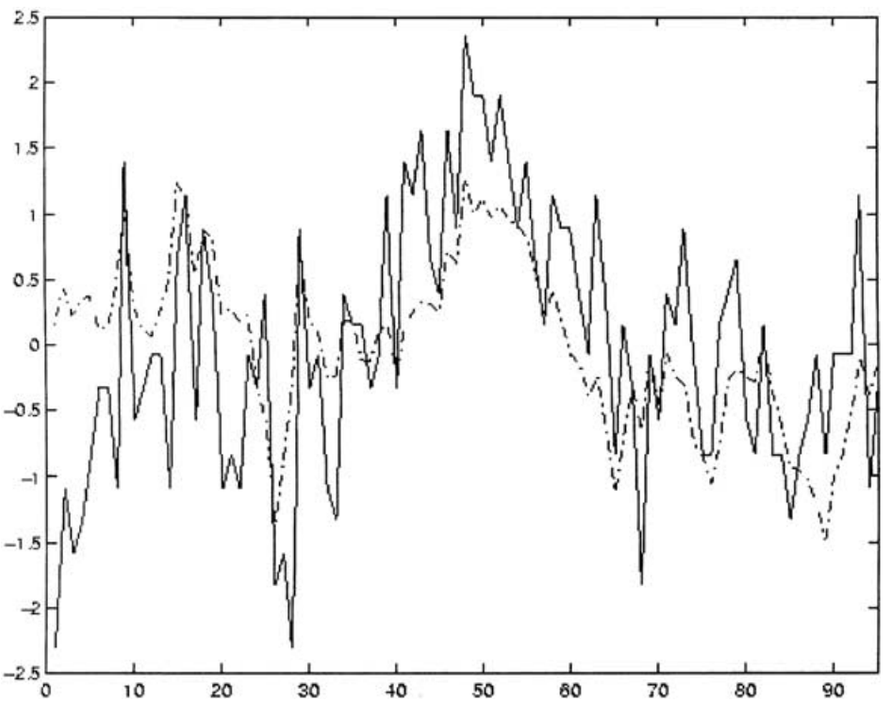

FIGURE 4 *NOx* output predicted using a PCA model built from variables 2, 3, and 8.

Comparing Figs. 3 and 4 with the naked eye, the predictions look kind of similar. Yet, the MSE values of the prediction of Fig. 4 are considerably lower, i.e. the reconstructed variance analysis did not work as well as we thought at deciding which variables need to be retained in order to obtain a decent prediction, even if the prediction is to be made by a PCA model.

4. METHODS BASED ON REGRESSION COEFFICIENTS

Three different methods based on regression coefficients have been used to analyse how to select a subset of variables to be kept within a model. These methods are ordinary least squares (OLS), principal components regression (PCR), and partial least squares (PLS). A general review of these methods can be found in Geladi and Kowalski (1986) and Jackson (1991).

4.1. Ordinary Least Squares Method

Given n observations of an input/output system with k input or predictor variables and one output or response variable, the traditional regression model can be written as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

where $\mathbf{X}$ denotes an $n \times k$ matrix of observations of the input variables normalised to zero mean and unit variance, $\mathbf{y}$ denotes an $n \times 1$ vector of the output also normalised, $\mathbf{b}$ a $k \times 1$ vector of regression coefficients, and $\mathbf{e}$ an $n \times 1$ vector of residuals (also called perturbations in the literature, that is, $\mathbf{e}$ is the effect of all the variables that affect variable y and that are not included in the model). The least squares solution for $\mathbf{b}$ is:

$$\mathbf{b} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y}$$

This expression is obtained under the following considerations about the residuals:

Their expectation is zero, $E[e_i] = 0$, and their variance is constant, σ^2.
Perturbations are independent, $E[e_i e_j] = 0$, $i \neq j$, and their distributions are normal.

Applying this method to our data, the vector of regression coefficients $\mathbf{b}$, is computed using the training data set. The 95% confidence interval for each regression coefficient, computed in accordance with Eq. (1), and the percentage that each coefficient contributes to variable y are compute as well in order to perform a variable selection.

$$\hat{\beta}_i \pm t_{n-k-1}(\alpha/2)\hat{S}_R\sqrt{q_{ii}} \tag{1}$$

In Eq. (1), $\hat{\beta}_i$ accounts for the regression coefficients, $t(\alpha/2)$ is the t distribution with $n - k - 1$ degrees of freedom, $\hat{S}_R$ is the estimation of the residual standard deviation, and q_{ii} are the diagonal elements of the matrix $(\mathbf{X}'\mathbf{X})^{-1}$.

The OLS technique is a very simple technique, and consequently, more refined techniques should be rejected if they cannot outperform OLS. Using an OLS model, it might be expected that the best prediction can be obtained when all variables (all regression coefficients) are being used. Figure 5 shows an OLS prediction of the validation data of the boiler system using all the input variables. The MSE value for the training data set is 0.6466, and for the validation data set, it is 1.1973.

Contrary to the PCA model, the OLS model also represents the oscillatory components of the behaviour. Yet, the MSE values are still larger than in the case of the PCA model.

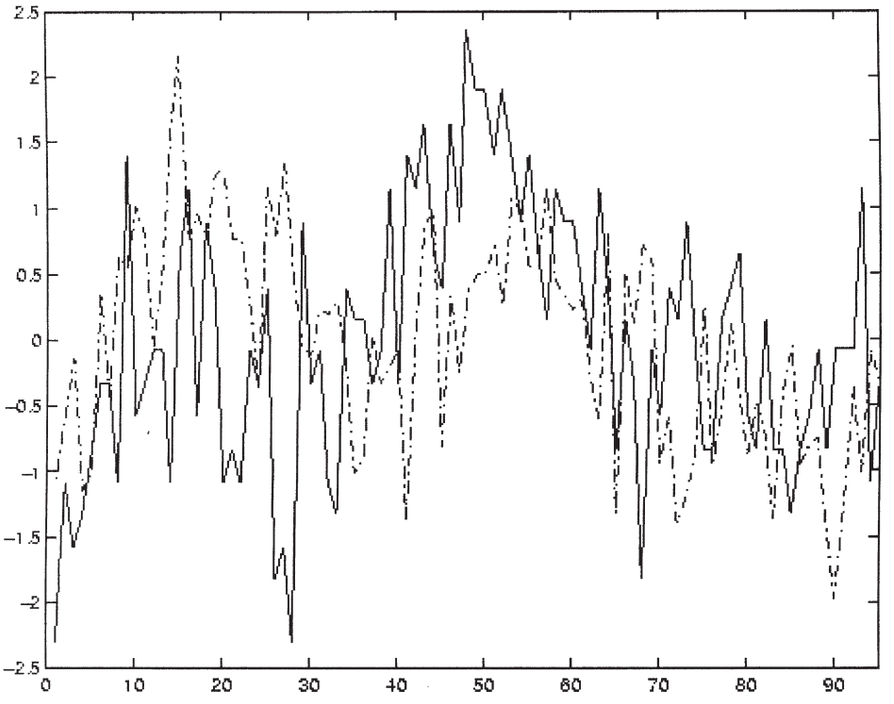

FIGURE 5 OLS Prediction with all input variables.

How can the regression coefficients be used to determine a subset of variables to be retained in a simplified model? Different variable selection methods can be found in the literature. In Peña (1989), a statistical analysis based on the t distribution is performed to check whether the ith regression coefficient can assume a value of 0 or not with a given probability. In Daling and Tamura (1970) and Lindgren *et al.* (1995), those variables with smaller coefficients in the regression equation are discarded. These two methods are similar to each other. When computing the confidence interval for a small coefficient, the probability that 0 is within this interval is high. In other words, when performing a t-test on this coefficient, the observed t value, computed in accordance with Eq. (2), will not be of significant magnitude, and therefore, the hypothesis $\hat{\beta}_i = 0$ cannot be refused. Consequently, this variable should not be taken into consideration within the model.

$$t_{\text{obs}} = \frac{\hat{\beta}_i}{\hat{S}_{\text{R}}\sqrt{q_{ii}}} \tag{2}$$

The criterion adopted here has been to select those variables with significant contribution to the regression equation, and to drop those with smaller contributions. The cut-off between selected and discarded variables is set to be 5% of contribution to the total regression line.

Applying this criterion to the previously obtained coefficients, variables 1, 2, 4, 5, and 7 are to be retained. Their regression coefficients are recomputed after eliminating from the **X** matrix those columns corresponding to discarded input variables. The top portion of Fig. 6 shows the real and predicted validation output values when using a regression model with variables 1, 2, 4, 5, and 7. The MSE value is 0.7483 when predicting the training data

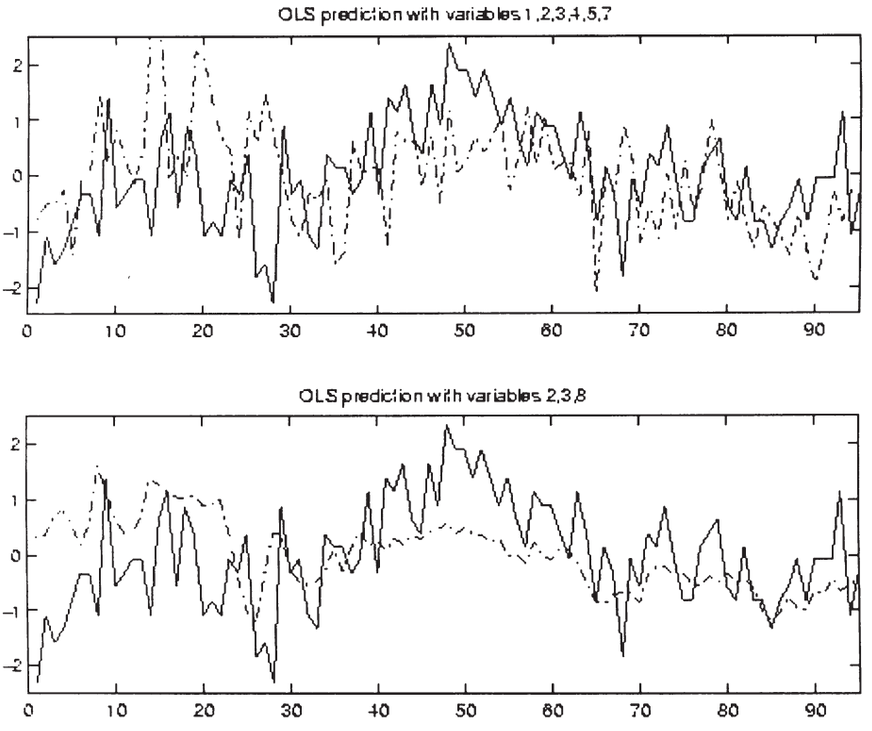

FIGURE 6 OLS predictions using variables 1, 2, 4, 5, and 7 (top), and variables 2, 3, and 8 (bottom).

set, and 1.3250 when predicting the validation data set. The prediction turns out to be a little poorer than that obtained when keeping all variables in the regression analysis.

Remembering the results obtained with the PCA model, it may be of interest to check what happens when the variables kept in the regression model are those that FIR suggested to retain, i.e. variables 2, 3, and 8. Again, the coefficients need to be recomputed. The bottom portion of Fig. 6 shows the corresponding prediction. In this case, the MSE value is 0.6482 for the training data set and 1.0703 for the validation data set. The results are clearly better than using the variables that the OLS modelling approach suggested to retain, and in the case of the validation data set, the results are even better than keeping all variables in the regression model.

As in the case of the PCA analysis, the OLS modelling approach resulted to be non-effective in terms of deciding which variables should be kept in the model and which should be discarded from it (Tables IV–VII).

4.2. Principal Components Regression Model

In this methodology, the input variables are transformed to principal components before calculating the regression coefficients. The PCR model thus pre-processes the input data,

TABLE IV Total unreconstructed variance for each Principal Component

1.7786	18.9828	3.8633	3.6903	3.6603	5.4373	18.2502	169.412

TABLE V Regression coefficients using OLS

Coefficient	95% *Confidence interval*		*% Contribution*
−1.8925	−3.0524	−0.7326	20.1015
2.3633	1.3890	3.3377	25.1023
0.3493	−0.1639	0.4346	3.7098
1.8467	0.5612	3.1322	19.6146
0.5028	0.0598	0.9459	5.3408
−0.1847	−0.6177	0.2483	1.9620
−2.1369	−2.7344	−1.5394	22.6976
0.1385	−0.0976	0.3746	1.4714

converting them to a set of equivalent PCs. It then uses those PCs to estimate the output by means of an OLS approach.

A PCR analysis was performed on the training data set of the boiler system, and a cross-validation method (Wold, 1978; Osten, 1988) was used to determine the number of Latent Variables (LVs) to be kept in the regression model. To this end, the training data were split into 10 blocks of equal size, and the Predictive Residual Error Sum of Squares (PRESS) value was calculated for each of them.

Analysing the results obtained, it was decided to retain four of the possible eight LVs in the regression model, and the regression coefficients were subsequently calculated as well as the percentage that each coefficient contributes to the regression equation. Since the PCA analysis results in a linear transformation on the input space, it is possible to transform the resulting regression coefficients for the four PCs back to eight equivalent regression coefficients for the original input variables.

Figure 7 shows the predictions obtained for the validation data set using a PCR model made of all input variables whereby the four most important PCs (LVs) were retained. The resulting MSE values are 0.7189 for the training data set and 1.2798 for the validation data set.

It may make sense to again throw some of the input variables out from the beginning. Using the same criterion that was applied in the case of the OLS model, we found that variables 1 and 3 could be discarded from the model, i.e. the model retains variables 2, 4, 5, 6, 7, and 8. A PCR model for this reduced set of input variables was subsequently computed. Cross-validation revealed that, in this case, five of the six possible LVs ought to be retained in the regression model. The top portion of Fig. 8 shows the prediction using this PCR model. The resulting MSE values are 0.6866 for the training data set and 1.1587 for the validation data set. The results are slightly better than for the PCR model involving all input variables.

Just like in the previous two sections, a third PCR model was then calculated using input variables 2, 3, and 8, as proposed by FIR. In this case, two of the possible three LVs are to be retained. The bottom portion of Fig. 8 shows the predictions obtained. In this case, the resulting MSE values are 0.7532 for the training data and 1.0645 for the validation data.

The results here are less good for the training data, but consistent with the results obtained for the previous two methods, they are better in the case of the validation data set.

TABLE VI Regression coefficients for var. 1, 2, 4, 5, and 7 (left col.) and for var. 2, 3, and 8 (right col.)

Coef. 1,2,4,5,7	*Coef.* 2,3,8
0.7389	0.6693
0.7051	0.0844
0.7575	−0.013
0.6024	
−2.152	

TABLE VII Regression coefficients for all variables when using PCR

Coefficients	*% Contribution*
−0.0404	1.5098
−0.1418	5.2951
0.0618	2.3079
0.2670	9.9665
1.2325	46.0144
−0.3242	12.1031
−0.4718	17.6139
0.1390	5.1893

These results could be interpreted in such a way as to suggest that statistical techniques offer decent *interpolation* capabilities, but FIR exhibits better *generalisation* power.

4.3. Partial Least Squares Regression Method

The PLS method operates in similar ways as the PCR method. The PCR method transforms the input space into a set of PCs. However, it does not do anything to the outputs. The PLS method transforms both the inputs and the outputs to sets of PCs using a PCA analysis, and in addition, takes into account the relationship between the input and the output spaces. A brief description of the PLS technique is given in Appendix B.

The PLS technique was applied to the training data of the boiler system. Just like in the PCR method, it is necessary to decide how many LVs are to be retained in the regression model. Hence, cross-validation was used, splitting the available data into 10 blocks of equal

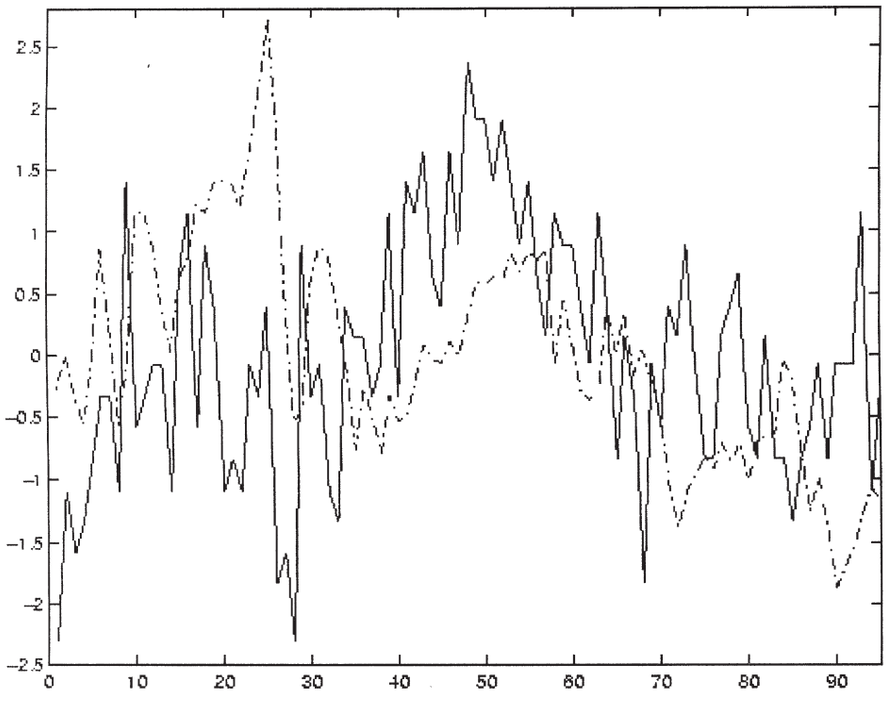

FIGURE 7 PCR model of all input variables retaining four LVs.

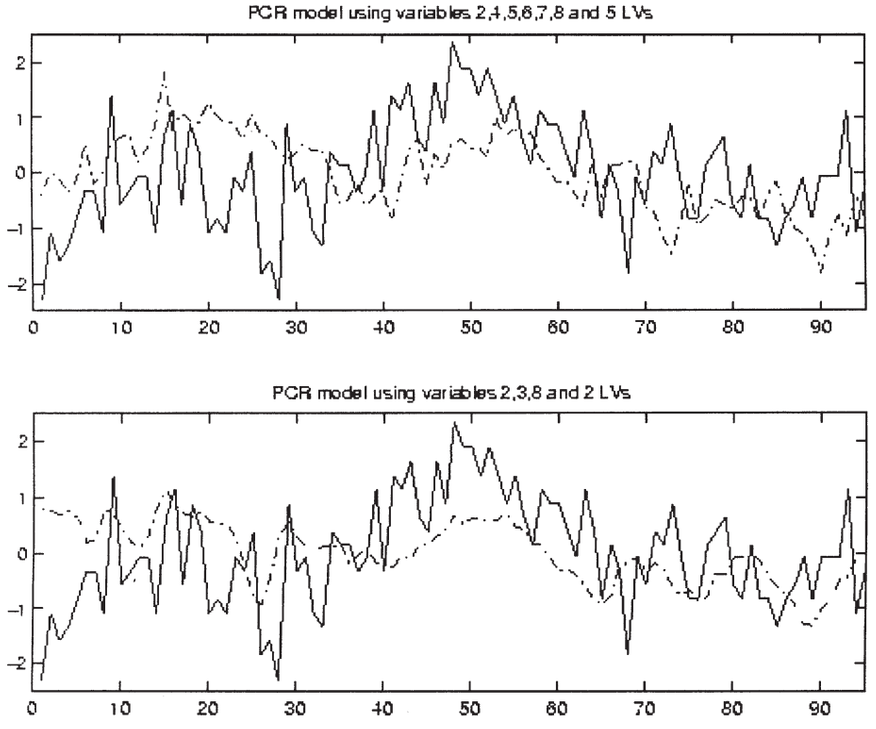

FIGURE 8 Prediction of PCR model using variables 2, 4, 5, 6, 7, and 8 (top); and 2, 3, and 8 (bottom).

sizes. The result of this study was that four of the possible eight LVs are to be retained in the regression model. As in the previous method, the regression coefficients found for the four retained PCs were then translated back to equivalent regression coefficients for the eight original input variables.

Figure 9 shows the prediction of the validation data set for the PLS model using all input variables and retaining four LVs. The resulting MSE values are 0.6536 for the training data set, and 1.1192 for the validation data set.

Next, variable selection is performed. The same criterion is used as in the two previous sections. This time, variables 2, 4, 5, 6, and 7 are selected. The PLS model for these variables was then calculated. After performing the cross-validation test, it was decided to retain two of the possible five LVs. The top portion of Fig. 10 shows the prediction of the validation data set using a PLS model in the five selected variables with two LVs retained. The resulting MSE values are 0.6967 for the training data set, and 1.2429 for the validation data set. Hence, the results are a little poorer than in the previous case, where all variables had been used (Tables VIII and IX).

As in the previous two sections, a comparison was made with a PLS model in variables 2, 3, and 8, as proposed by FIR. Cross-validation revealed that two of the possible three LVs ought to be retained. The bottom portion of Fig. 10 shows the prediction obtained using this model. The resulting MSE values are 0.7527 for the training data set, and 1.0484 for the validation data set. As in the case of the PCR analysis, the results are poorer for the training data (reduced interpolation capability) but better for the validation data (improved generalisation power).

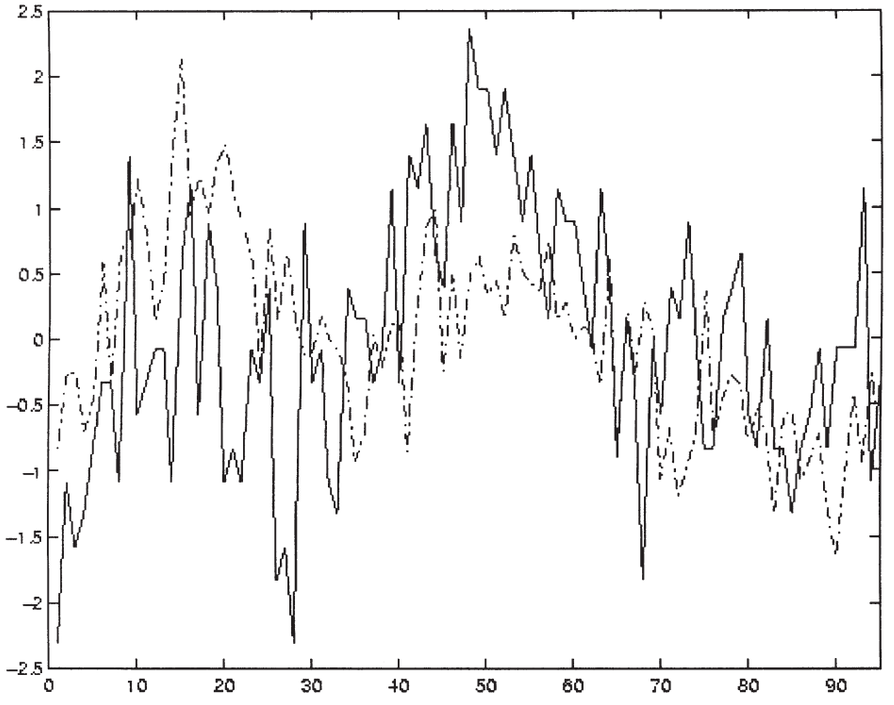

FIGURE 9 PLS model using four LVs and all physical variables.

5. OTHER METHODS

The results obtained using different methods of selecting variables advocated in Jolliffe (1972; 1973) are compared in this section using the same data set. One of those methods uses multiple correlation analysis, two methods are based on principal component analysis, and two further methods are based on cluster analysis. As with all the other methods presented, only the training data set was used to decide, which variables are to be retained to construct the model.

5.1. Multiple Correlation Coefficients

The first method used is based on Multiple Correlation Coefficients (MCC). The algorithm is named **A2** in the above referenced papers. The method works as follows: Suppose there are

TABLE VIII Regression coefficients for variables 2, 4, 5, 6, 7, and 8 (left); and 2, 3, and 8 (right)

Coef. 2,4,5,6,7,8	*Coef.* 2,3,8
0.7260	0.3271
0.9852	0.0697
0.6313	0.3297
0.0787	
− 1.9695	
0.2074	

TABLE IX Regression coefficients for all variables when using PLS

Coefficients	*% Contribution*
−0.1660	2.4436
1.6099	22.6983
0.2671	3.9315
1.5364	22.6154
0.4340	6.3880
−0.3468	5.1052
−2.4332	35.8170
0.0001	0.0009

p variables. The first discarded variable is the one that has the maximum multiple correlation with the remaining $p - 1$ variables. The multiple correlation of a variable is calculated as the sum of individual correlations between that variable and any other variable. The algorithm is repeated with the remaining variables. At each stage with q variables remaining, the variable having the largest multiple correlation with the other $q - 1$ variables is the next to be discarded. The algorithm terminates when the maximum multiple correlation between variables has decreased to a value below R_0. According to the authors of the papers, a good value for R_0 is 0.15.

When applying this method to the boiler data, it is found that variables 2 and 3 are to be retained. The discarded variables are, in the order of discarding them, variables 1, 7, 6, 5, 4, and 8. Notice that the last variable discarded was variable 8. Hence, with a value of

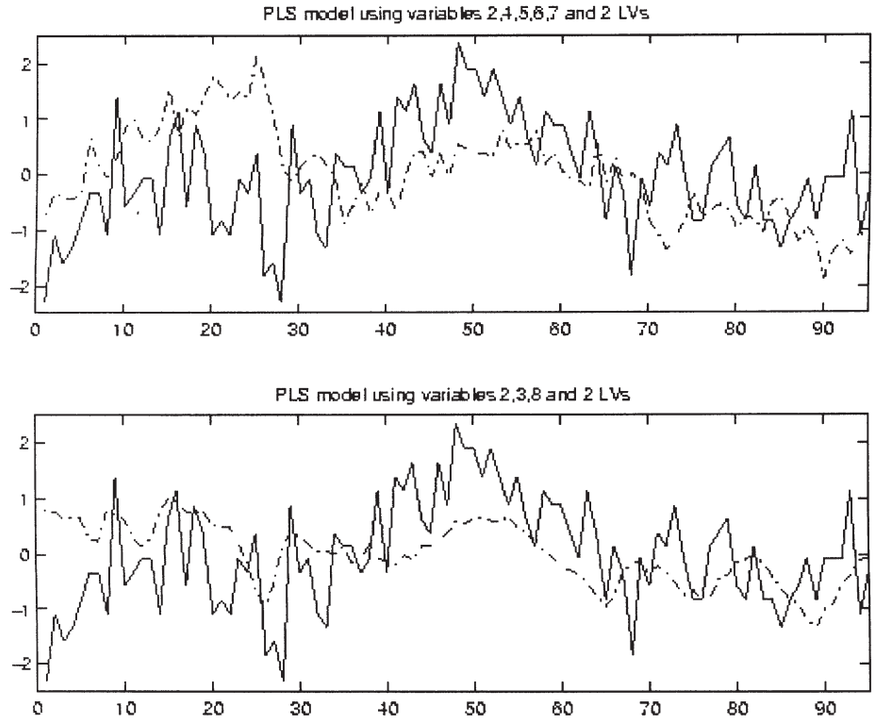

FIGURE 10 Predictions using PLS with physical variables 2, 4, 5, 6, and 7 (top); and 2, 3, and 8 (bottom).

R_0 slightly larger, this method would have found the same set of variables to be retained as FIR.

5.2. Methods Based on Principal Component Analysis

The next two methods in Jolliffe (1972; 1973) are based on principal component analysis.

5.2.1. B2 Method

A principal component analysis is performed on the data. Those PCs with eigenvalues smaller than L_0 are taken into account. Then, beginning with the PC corresponding to the smallest eigenvalue, the variable with the largest coefficient in this PC that has not been eliminated before is discarded. The algorithm proceeds in this way with all the PCs chosen. According to the references, a good value for L_0 is 0.7.

Two versions of this method have been implemented. The first of them, named **B2a**, uses the components of the PCs as they are, i.e. preserving their sign. When applying the method to the boiler data, it is found that variables 3 and 8 must be retained, whereas variables 4, 1, 7, 2, 5, and 6 (in this order) must be discarded.

The second version, referred to as **B2b**, makes use of the absolute value of the coefficients. When applying this method to the boiler data, it is found that variables 2 and 3 must be retained, whereas variables 4, 1, 7, 6, 5, 8 (in this order) must be discarded. Notice that also this method would have found the optimal set of variables to be retained if only a slightly smaller value of L_0 would have been chosen.

5.2.2. B4 Method

The second method based on principal component analysis is the one named **B4**. The basic idea is the same as in the previous case, but operates in backward mode. A principal component analysis is performed on the data. Those PCs with eigenvalues larger than L_0 are taken into account. Then, beginning with the vector corresponding to the largest eigenvalue, the variable with the largest coefficient in this PC that has not been selected already before is preserved. The method proceeds with all the PCs chosen in this way. A good value for L_0 is also 0.7.

Once more, the results of two versions of this algorithm are presented. The first one, names **B4a**, uses the coefficients of the PCs as they are, preserving their sign. When applying the method to the boiler data, it is found that variables 1 and 2 (in order of acceptance) are to be retained, while the remaining variables are to be discarded.

The second version, named **B4b**, uses the absolute value of the coefficients. When applying this method to the boiler data, it is found that variables 1 and 3 must be retained with all other variables being discarded.

5.3. Methods Based on Cluster Analysis

The two last methods to be discussed in this report are based on cluster analysis. Although many clustering methods have been reported in the literature, only two of those methods have been included in this paper. Before discussing the results obtained, a sketch of the four steps of the clustering methods applied in this study is presented.

- First, it is necessary to define a measure of similarity, say R_{XY}, between two groups of variables X and Y. Two such metrics will be used here in order to compare the two different clustering methods.

The first metric is for a complete-linkage method and is given by

$$R_{XY} = \max_{\substack{i \in X \\ j \in Y}} r_{ij} \tag{3}$$

The second measure is for an average-linkage method and is given by

$$R_{XY} = \frac{\left(\sum_{i \in X} \sum_{j \in Y} r_{ij} \right)}{(n_1 n_2)} \tag{4}$$

where r_{ij} are the correlation coefficients between variables i and j; and n_1 and n_2 are the number of variables contained in groups X and Y, respectively.

- Second, once the measure of similarity has been chosen, R_{XY} is computed for each of the $p(p-1)/2$ pairs of single variable groups of a p variable system.
- Third step: if A and B are the two groups for which R_{XY} is a maximum, replace A and B by the single group $C = A \cup B$.
- Fourth and last step: for each group not involved in the previous step, calculate R_{XC} and return to the third step.

The clustering process cycles between steps three and four until the number of groups has decreased to a sufficiently small value. Two more decisions need to be made: (1) when to stop with the clustering algorithm, and (2) which variable to choose from each of the obtained clusters.

A criterion for terminating the clustering algorithm is to continue with steps three and four until all R_{XY} between the remaining clusters fall below some level R_0. In accordance with Jolliffe (1972; 1973), a good value for R_0 is 0.55, when the complete-linkage method is used, and 0.45, when the average-linkage method is applied. Two methods of choosing variables from each cluster have been implemented:

Inner-clustering selects one of the first two variables forming each cluster.
Outer-clustering selects the last variable that joined each cluster.

Table X summarises the results obtained when applying this method to the boiler system. Using either the single-linkage method or the average-linkage method leads to the same results, thus, Table X is valid for both metrics. The variables in each cluster are given in the order in which they joined the cluster.

The reader may notice that both techniques selected a subset of the variables that FIR proposed to retain. Combining the two methods, the optimal set of variables would have been found in the case of the example at hand.

TABLE X Regression coefficients of PLS model for variables 2, 4, 5, 6, and 7 (left); and 2, 3, and 8 (right)

Coef. 2,4,5,6,7	*Coef.* 2,3,8
0.4469	0.3461
0.7674	0.0790
1.0934	0.3100
−0.4233	
−1.2280	

6. USING SUBSETS OF VARIABLES FOR STATIC FIR PREDICTIONS

To this end, the predictions made by statistical techniques, as presented in Sections 3–5, were compared against the much better prediction made by FIR in "Fuzzy Inductive Reasoning" section (Fig. 2). Such a comparison is unfair, because all statistical techniques presented in this paper employed *static* models only, whereas FIR made use of a *dynamic* (time dependent) model. All techniques discussed in Sections 3–5 could also have been used to generate dynamic models by simply duplicating and triplicating the sets of variables, i.e. the columns of the data matrix, shifting them down each time by one row. The computational work for these methods would have been enlarged, because they would have had to deal with more variables in this way, but the predictions would certainly have been improved.

Yet, this was not the purpose of this paper. The purpose was to find a set of bootstrapping techniques for FIR that, with little computational effort, could encounter subsets of variables to be considered in a FIR optimal mask search. To this end, static models, which can be obtained easily at low computational cost, have been generated, in the hope that the variables not used by these models would be less likely candidates also in a dynamic model search. That is, if the performed static analysis is used in a conservative way, only throwing out the worst variables, FIR will be able to perform a good estimate of important variables even dynamically. Such an assumption makes sense due to the autocorrelation inherent in any physical signal.

Table XI lists the results of employing the static methods obtained in previous sections as bootstrapping techniques for a dynamic model search using FIR.

The first row of Table XI shows the results of performing an optimal mask search using all nine variables and a candidate mask of depth 16. Hence, the corresponding candidate mask contains $16 \cdot 9 - 1 = 143$ potential inputs (" -1" elements). An exhaustive search was performed analysing the quality of masks consisting of up to four of these inputs plus the output. To this end, $1 \cdot 581 \cdot 580$ masks had to be evaluated. The search consumed 160 min of computation time on a Sun Ultra Sparc II Workstation. The optimal masks of complexities 4 and 5 (columns C4 and C5) were tabulated with respect to their resulting mask qualities (columns 3 and 4). The number of different masks visited in the process of searching for the optimal mask of complexity 4 is listed in column 5, and the number of masks visited in search of the optimal mask of complexity 5 is presented in column 6. Column 7 shows the MSE error obtained in a prediction that combines the predictions made by the optimal masks of complexities 4 and 5. As discussed in López (1999), FIR not only makes a prediction of an output variable; it simultaneously provides a measure of confidence in its own prediction. In the simulation leading to the MSE value reported in column 7, predictions were made in parallel with the optimal masks of complexities 4 and 5, and in each step, the prediction accompanied by the larger confidence value was kept.

Each of the subsequent rows tabulates one suboptimal search algorithm, making use of the results of the static model searches performed in Sections 2–5. For example, the first of these

TABLE XI Selection of variables achieved with cluster analysis

Method	*Inner clustering*	*Outer clustering*
Variables kept	2,3	3,8
Clusters found	2,4,1,7,6,5,8,3	2,4,1,7,6,5,8,3

rows tabulates a suboptimal search, whereby the variables to be considered were obtained using the static FIR models found in "Fuzzy Inductive Reasoning" section of the paper. Those models suggested that variables 1 and 6 are less likely candidates for input variables. Consequently, variables 1 and 6 were discarded from the set of potential inputs. This is shown in column 2. The mask candidate now contains 0 elements (forbidden connections) in columns 1 and 6, and consequently, it only contains $16{\cdot}7 - 1 = 111$ potential inputs. The subsequent exhaustive search through all masks compatible with the mask candidate matrix and the constraint of not having more than four inputs resulted in a search through 557,845 masks. Hence, the computational effort was about 1/3 of the one needed for the experiment described in the previous paragraph. It consumed 53 min of execution time on a Sun Ultra Sparc II Workstation. It turned out that FIR did a good job at discarding variables. The resulting masks of complexities 4 and 5 are exactly the same as found using the three times more expensive search through all possible masks.

The subsequent rows revisit every one of the techniques discussed in Sections 3–5 of the paper. The techniques of "Method of the Unreconstructed Variance for the Best Reconstruction" and "Methods Based on Regression Coefficients" sections were not well suited for the task at hand. All of these techniques threw out variable 3, which turned out to be essential in making good FIR predictions. This variable was discarded, because it exhibits a relatively poor cross-correlation with the other variables. Consequently, these statistical techniques considered the variable of lesser relevance. This decision let to optimal masks of reduced quality, and as was to be expected, the use of these masks in a FIR prediction led to substantially larger prediction errors.

Obviously, cross-correlation only evaluates the strengths of *linear* relationships, whereas the FIR forecasting engine exploits also *non-linear* relationships among variables. Yet, this does not fully explain the comparatively poor performance of these techniques, since even the statistical modelling techniques of "Method of the Unreconstructed Variance for the Best Reconstruction" and "Methods Based on Regression Coefficients" sections, using perfectly linear regression models, exhibit better prediction results when they are applied to the set of variables selected by FIR (which includes variable 3) than when they are based on their own variable selection. Evidently, and in spite of its relatively poor cross-correlation with the other variables, variable 3 still contained valuable information that could be exploited in predictions.

The final set of rows summarizes the performance of the techniques presented in "Other Methods" section. These techniques performed considerably better than those presented in "Method of the Unreconstructed Variance for the Best Reconstruction" and "Methods Based on Regression Coefficients" sections. Except for method **B4A**, all methods resulted in optimal masks that were either the truly optimal ones, or at least of almost equal qualities. In accordance with this finding, also the resulting MSE values were close to optimal. Why did these techniques work better? The reason is that they did not attempt to eliminate variables with poor cross-correlation to the output. Instead, they eliminate variables with strong cross-correlation to other inputs. This makes sense, because if two inputs are strongly correlated with each other, they contain almost identical information, and therefore, either one of them will suffice to explain the output. This strategy works even in the case of non-linear systems and for use by non-linear prediction algorithms.

The techniques presented in "Other Methods" section are considerably more aggressive in throwing out variables than the algorithms presented earlier. Since the set of remaining variables is small, the optimal mask search, for the given example, can be performed quickly. These searches are completed on a Sun Sparc II Workstation within less than 2 min, i.e. they execute about 100 times faster. All of these techniques exhibit another important advantage. They sort the variables in order of increasing or decreasing importance. Hence, it would be

easy to add one more variable and repeat the optimal mask search to check whether or not the mask quality improves. This could still be done rather inexpensively.

The loss of prediction quality incurred when computing FIR dynamic models from the different subsets of variables is summarized in Table XII. The columns labelled *P* stand for the percentage of mask quality lost by the different variable subsets relative to that of the exhaustive FIR model of equal complexity. *P* is computed as:

$$P = \frac{Q_{\text{FIR dynamic}} - Q_{\text{MODELi}}}{Q_{\text{FIR dynamic}}} \times 100$$

The columns labelled *N* represent the percentage of MSE error increase due to the selection of different subsets of variables relative to that of the exhaustive FIR model. *N* is computed as:

$$N = \frac{\text{MSE}_{\text{MODELi}} - \text{MSE}_{\text{FIR dynamic}}}{\text{MSE}_{\text{FIR dynamic}}} \times 100$$

The values in the rows labelled C* were computed combining the predictions made by the optimal masks of complexities 4 and 5 as explained earlier.

From Table XII, it is evident that there exists a strong positive correlation between the percentage-wise reduction in mask quality and the corresponding increase in prediction error, at least for the example at hand.

Table XIII shows the reduction in computing effort attained when using FIR with each one of the proposed subsets of variables. Each column stands for the reduction in the number of masks to compute and it has been calculated as:

$$R = \frac{\#_{\text{FIR dynamic}} - \#_{\text{MODELi}}}{\text{FIR dynamic}} \times 100$$

where # is the number of masks to be evaluated for a given complexity. The last column in Table XIII relates to methods *A2*, *B2a*, *B2b*, *B4a*, *B4b*, inner and outer clustering, because they all achieve the same reduction of the FIR model search space (Table XIV).

TABLE XII Dynamical models obtained from reduced sets of variables

Method	*Selected variables*	*Mask qualities*		*Number of computed masks*		*MSE*
		C4	*C5*	*C4*	*C5*	
None	All	0.6080	0.6196	82160	1581580	0.5522
FIR (Static)	2,3,4,5,7,8	0.6080	0.6196	37820	557845	0.5522
Unreconstr. var.	1,2,4,5,6,7,8	0.5943	0.4290	59640	1028790	0.7324
OLS	1,2,4,5,7	0.5943	0.4258	23426	292825	0.7516
PCR	2,4,5,6,7,8	0.5943	0.4290	37820	557845	0.7324
PLS	2,4,5,6,7	0.5943	0.4258	23426	292825	0.7516
A2	2,3	0.6080	0.6167	2600	14950	0.5529
B2a	3,8	0.6057	0.6177	2600	14950	0.5845
B2b	2,3	0.6080	0.6167	2600	14950	0.5529
B4a	1,2	0.5943	0.4258	2600	14950	0.7516
B4b	1,3	0.6049	0.6123	2600	14950	0.5640
Inner clust.	2,3	0.6080	0.6167	2600	14950	0.5529
Outer clust.	3,8	0.6057	0.6177	2600	14950	0.5845

TABLE XIII Loss of prediction quality due to selection of variable subsets

	FIR (dynamic)		*FIR (static)*		*Unrecons. variance*		*PCR*		*OLS, PLS*		*A2, B2b, Inner clustering*		*B2a, Outer clustering*		*B4a*		*B4b*	
	Q	*P*	*Q*	*P*	*Q*	*P*	*Q*	*P*	*Q*	*P*	*Q*	*P*	*Q*	*P*	*Q*	*P*	*Q*	*P*
C4	0.6080	0	0.6080	0	0.5943	2.2	0.5943	2.2	0.5943	2.2	0.6080	0	0.6057	0.4	0.5943	2.2	0.6049	0.51
C5	0.6196	0	0.6196	0	0.4290	30.7	0.4290	30.7	0.4258	31.3	0.6167	0.5	0.6177	0.3	0.4258	31.3	0.6123	1.2
	MSE	N	MSE	N	MSE	N	MSE	N	MSE	N	MSE	N	MSE	N	MSE	N	MSE	N
C*	0.5522	0	0.5522	0	0.7324	32.6	0.7324	32.6	0.7516	36.1	0.5529	0.1	0.5845	5.9	0.7516	36.1	0.5640	2.14

TABLE XIV Model search space reduction attained with each of the methods

	FIR (*dynamic*)	*FIR* (*static*) *PCR*	*Unreconstructed variance*	*OLS*, *PLS*	*A2*, *B2a*...
C4	0	53.97	27.41	71.49	96.84
C5	0	64.73	34.95	81.49	99.05

7. CONCLUSIONS

The behaviour of systems can be predicted using either *a priori* knowledge (deductive techniques) or observations (inductive techniques). Only inductive prediction techniques were analysed in this article. All but the simplest of those techniques make predictions in two steps. In the first step, an input/output model is created based on the observations made; in the second step, a simulation of the previously made model is then performed with the purpose of making predictions. All of the techniques surveyed in this paper first make a model that is then being used in a simulation.

When creating a model, the observations can either be used directly (quantitative modelling techniques), or they can first be discretised or at least fuzzified (qualitative modelling techniques). All of the techniques studied in this article make use of quantitative modelling techniques, except for FIR, which embraces a qualitative modelling approach.

The modelling process, be it quantitative or qualitative in nature, usually occurs in two steps. In a first step, the model *structure* is being identified. In a second step, the model *parameters* are being estimated. Most of the techniques discussed in this paper operate in such a fashion. The model structure determines the set of variables to be used by the model. In the case of the techniques discussed in Sections 3–5 of the paper, these variables are then being used, either directly or indirectly, in a linear regression model. The parameter estimation step determines the regression coefficients. FIR also starts out by determining the model structure, i.e. by selecting the set of variables to be used in the simulation. However, no parameter estimation takes place, since FIR is a non-parametric technique. During its qualitative simulation, FIR refers directly back to the training data, rather than capturing the knowledge contained in the training data in a set of parameter values.

Models can be either *dynamic* or *static*. In a dynamic model, the current value of the output may depend on its own past, as well as on current and past values of the inputs. In a static model, the current value of the output only depends on the current values of the inputs. All of the techniques advocated in this article may be used to create static or dynamic models, though only FIR was actually used in the paper for creating dynamic models.

Since all of the techniques discussed in this article first select a set of variables to be used, they can be arbitrarily combined with each other, i.e. any one of the techniques can be used to select the set of variables, which can then be used by either the same or any other technique to make predictions.

In "Fuzzy Inductive Reasoning" section of the paper, FIR was used to create both static and dynamic models. The static FIR models suggested elimination of variables 1 and 6 of the 9-variable system used throughout the paper as an example. The dynamic FIR model, due to its better resolution, suggested elimination of variables 4, 5, and 7 in addition to variables 1 and 6, preserving only variables 2, 3, and 8 as the most relevant inputs.

"Method of the Unreconstructed Variance for the Best Reconstruction" and "Methods Based on Regression Coefficients" sections of the paper analysed a set of statistical modelling and simulation techniques. All of these techniques were used exclusively for the creation of static models. In each subsection, a technique was used to select a subset

of variables to be subsequently used in a linear regression analysis for making predictions. The simulation results obtained in this way were compared against simulations making use of the variables proposed by FIR, i.e. the first step of each technique was replaced by a FIR model selection, whereas the subsequent parameter identification and regression techniques were preserved from the methods discussed. It turned out that none of these *linear* modelling techniques did a very good job at choosing a pertinent subset of variables of the *non-linear* plant used as an example. The variables proposed by FIR worked better, even for the purpose of being used in *linear* regression models.

"Other Methods" section of the paper discussed a set of clustering techniques for the purpose of variable selection. These are pure modelling techniques that can be combined with any of the previously discussed simulation approaches. No simulations were performed in "Other Methods" section. All of the techniques discussed in this section were used for static modelling only. It turned out that the techniques advocated in "Other Methods" section were excellently suited for the purpose of variable selection.

"Using Subsets of Variables for Static FIR Predictions" section of the paper made use of the subsets of variables proposed by the different techniques presented in the earlier sections for the purpose of creating dynamic FIR models to be used in subsequent FIR simulations. The techniques proposed in "Method of the Unreconstructed Variance for the Best Reconstruction" and "Methods Based on Regression Coefficients" sections of the paper were least suitable for the task at hand. They eliminated important variables early on, while keeping a fairly large set of less important variables in the model. The techniques presented in "Other Methods" section were excellently suited for the purpose of variable pre-selection. They are fairly fast, work well also in the case of non-linear applications, and order the variables in terms of either increasing or decreasing importance.

Of all the techniques discussed in this article, FIR is by far the best both in terms of its modelling capabilities as well as the power of its simulation engine. Hence, FIR can be used as a gauge against which the other techniques can be measured. Yet, FIR is deplorably *slow* both during modelling and during simulation. FIR's modelling engine is of exponential computational complexity, at least if an exhaustive mask search is being used, and consequently, FIR is unsuited for dealing with large-scale models. Only neural networks are yet slower in terms of creating models from observations. Hence, FIR needs a booster technique. Some of the approaches discussed in "Other Methods" section revealed themselves as excellently suited for such purpose.

Like all non-parametric approaches, FIR is also slow during simulation, but this is unfortunately inevitable. No booster technique can help with this problem.

Only a single application was used throughout the paper to demonstrate the advantages and shortcomings of the various methodologies discussed. Yet, the chemical process discussed in this paper is fairly generic, and the results obtained can indeed be generalised beyond this single application. Other applications have been studied, and the results obtained are consistent with those reported in this paper.

Acknowledgements

The research reported in this article was made possible, thanks to a Ph.D. fellowship of the Ministry for Education and Culture from the Spanish Government funded within the frame of the TAP96-0882 project that enabled the first author of this paper to spend 3 months at the University of Texas in Austin with the research group of Dr Joe Qin during the fall of 1998.

References

Adams, M.J. and Allen, J.R. (1998) "Variable selection and multivariate calibration models for X-ray fluorescence spectrometry", *Journal of Analytical Atomic Spectrometry* **13**(2), 119–124, ISSN: 0267-9477.

de Albornoz, A. (1996) *Inductive Reasoning and Reconstruction Analysis: Two Complementary Tools for Qualitative Fault Monitoring of Large-Scale Systems*, Ph.D. Dissertation, Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya (Barcelona, Spain).

Al-Kandari, N. and Jolliffe, I.T. (1997) "Variable selection and interpretation in canonical correlation analysis", *Communications in Statistics, Part B: Simulation and Computation* **26**(3), 873–900.

Allen, D.M. (1971) "Mean square error of prediction as a criterion for selecting variables", *Technometrics* **13**, 469–475.

Cellier, F.E. (1991) *Continuous System Modeling* (Springer, New York).

Cellier, F.E. and Yandell, D.W. (1987) "SAPS II: a new implementation of the systems approach problem solver", *International Journal of General Systems* **13**(4), 307–322.

Cellier, F.E., Nebot, A., Mugica, F. and de Albornoz, A. (1992) "Combined qualitative/quantitative simulation models of continuous-time processes using FIR techniques," Proceedings of the SICICA'92, IFAC Symposium on Intelligent Components and Instruments for Control Applications, Malaga, Spain, May 22–24, pp. 589–593.

Chipman, H., Hamada, M. and Wu, C.F.J. (1997) "Bayesian variable-selection approach for analyzing designed experiments with complex aliasing", *Technometrics* **39**(4), 372–381.

Daling, J.R. and Tamura, H. (1970) "Use of orthogonal factors for selection of variables in a regression equation—an illustration", *Applied Statistics* **19**(3), 260–268.

Dunia, R. (1997) *A Unified Geometric Approach for Process Monitoring and Control*, Ph.D. Dissertation, Department of Chemical Engineering, The University of Texas at Austin (Austin, TX).

Dunia, R. and Qin, S.J. (1998) "A unified geometric approach to process and sensor fault identification and reconstruction: the unidimensional fault case", *Computers in Chemical Engineering* **22**(7–8), 927–943.

Dunia, R., Qin, S.J., Edgar, T.F. and McAvoy, T.J. (1996) "Identification of faulty sensors using principal component analysis", *AIChE Journal* **42**, 2797–2812.

Geladi, P. and Kowalski, B.R. (1986) "Partial least squares regression: a tutorial", *Analytica Chimica Acta* **185**, 1–17.

Heikka, R., Minkkinen, P. and Taavitsainen, V.-M. (1994) "Comparison of variable selection and regression methods in multivariate calibration of a process analyzer", *Process Control and Quality* **6**(1), 47–54.

Hoeting, J. and Ibrahim, J.G. (1998) "Bayesian predictive simultaneous variable and transformation selection in the linear model", *Computational Statistics and Data Analysis* **28**(1), 87–103.

Hoeting, J., Raftery, A.E. and Madigan, D. (1996) "Method for simultaneous variable selection and outlier identification in linear regression", *Computational Statistics and Data Analysis* **22**(3), 251–270.

Jackson, J.E. (1991) *A User's Guide to Principal Components* (John Wiley Interscience, New York).

Jolliffe, I.T. (1972) "Discarding variables in a principal component analysis. I: Artificial Data", *Applied Statistics* **21**, 160–173.

Jolliffe, I.T. (1973) "Discarding variables in a principal component analysis. II: Real Data", *Applied Statistics* **22**, 21–31.

Kabaila, P. (1997) "Admissible variable-selection procedures when fitting misspecified regression models by least squares", *Communications in Statistics Theory and Methods* **26**(10), 2303–2306.

Klir, G.J. (1985) *Architecture of System Problem Solving* (Plenum Press, New York).

Li, D. and Cellier, F.E. (1990) "Fuzzy measures in inductive reasoning", Proceedings of the Winter Simulation Conference, (New Orleans, LA), pp 527–538.

Lindgren, F., Geladi, P., Berglund, A., Sjöström, M. and Wold, S. (1995) "Interactive variable selection (IVS) for PLS. Part II: Chemical applications", *Journal of Chemometrics* **9**(5), 331–342.

Lisboa, P. and Mehri-Dehnavi, A.R. (1996) "Sensitivity methods for variable selection using the MLP," Proceedings of International Workshop on Neural Networks for Identification, Control, Robotics, and Signal/Image Processing, NICROSP. IEEE, Los Alamitos, CA, USA. pp. 330–338.

López, J. (1999) *Time Series Prediction Using Inductive Reasoning Techniques*, Ph.D. Dissertation, Organització i Control de Sistemes Industrials, Universitat Politècnica de Catalunya (Barcelona, Spain).

Mansfield, E.R., Webster, J.T. and Gunst, R.F. (1977) "An analytic variable selection technique for principal component regression", *Applied Statistics* **26**(1), 34–40.

McShane, M.J., Cote, G.L. and Spiegelman, C. (1997) "Variable selection in multivariate calibration of a spectroscopic glucose sensor", *Applied Spectroscopy* **51**(10), 1559–1564, ISSN: 0003-7028.

Mirats Tur, J.M. and Huber, R.M. (2000) "Fuzzy inductive reasoning model based fault detection applied to a commercial aircraft", *Simulation* **75**(4), 188–198.

Mugica, F. (1995) *Diseño Sistemático de Controladores Difusos Usando Razonamiento Inductivo*, Ph.D. Dissertation, Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya (Barcelona, Spain).

Muñoz, A. and Czernichow, T. (1998) "Variable selection using feedforward and recurrent neural networks", *International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications* **6**(2), 91–102.

Nebot, A., Cellier, F.E. and Vallverdú, M. (1998) "Mixed quantitative/qualitative modeling and simulation of the cardiovascular system", *Computer Methods and Programs in Medicine* **55**, 127–155.

Osten, D.V. (1988) "Selection of optimal regression models via cross-validation", *Journal of Chemometrics* **2**, 39–48.

Peña, D. (1989) *Estadística modelos y métodos. Modelos lineales y series temporales*, Alianza editorial, 2a Ed.
Qin, S.J. and Dunia, R. (1998) "Determination of the number of principal components for best reconstruction," Proceedings of the 5th IFAC Symposium on Dynamics and Control of Process Systems, Corfu, Greece, June 8–10, pp. 359–364.
Uyttenhove, H.J. (1979) *SAPS—System Approach Problem Solver*, Ph.D. Dissertation, SUNY (Binghamton, NY).
Wold, S. (1978) "Cross-validatory estimation of the number of components in factor and principal components models", *Technometrics* **20**(4), 397–405.

APPENDIX A

The mathematical foundations underpinning the methodology for fault identification by variance reconstruction, advocated in section 3 of the paper, are reviewed here. A full description of the method can be found in Dunia (1997), Dunia and Qin (1998) and Qin and Dunia (1998). The approach discussed in those papers makes use of a normal process model to decompose the sample vector into two parts:

$$x = \hat{x} + \tilde{x} \tag{1}$$

where $x \varepsilon \Re^m$ represents a normalised sample vector of zero mean and unit variance. The vectors $\hat{x}$ and $\tilde{x}$ are the modelled and residual portions of x, respectively. PCA is used to calculate $\hat{x}$,

$$\hat{x} = Pt = PP^{\mathrm{T}}x = Cx \tag{2}$$

where $P \varepsilon \Re^{m \times l}$ is the loading matrix, and $t \varepsilon \Re^l$ is the score vector. The number of PCs retained are $l \geq 1$. The matrix $C = PP^T$ represents the projection on the l-dimensional principal component subspace. The residual $\tilde{x}$ lies in the residual subspace of m-l dimensions

$$\tilde{x} = (\mathbf{I}^{(\mathrm{m})} - \mathrm{C})x \tag{3}$$

The PCA model partitions the measurement space ($\Re^m$) into two orthogonal subspaces: the principal component subspace, and the residual subspace.

The sample vector for normal operating conditions is denoted by x^* (unknown when a fault has occurred). In the presence of a process fault $\Im_i$, the sample vector can be represented as:

$$x = x^* + f\xi_i \tag{4}$$

where ξ_i is a normalised fault direction vector, and the scalar f represents the magnitude of the fault. The fault direction vector can be projected on the two subspaces:

$$\xi_i = \hat{\xi}_i + \tilde{\xi}_i \tag{5}$$

where $\Im_j$ has been assumed. Along all possible fault directions x^* is reconstructed from x, the vector x_j is obtained moving x in the ξ_j direction,

$$x_j = x - f_j\xi_j \tag{6}$$

where f_j is an estimate for f. The reconstructed vector is expected to be close to x^*, the distance between x_j and the principal component subspace is given by the magnitude of the SPE for the reconstructed vector. The fault magnitude f_j is obtained by minimising SPE_j along the direction ξ_j

$$\mathrm{SPE}_j \equiv \|\tilde{x}\|^2 = \|\tilde{x} - \tilde{f}_j\tilde{\xi}_j^0\|^2 \tag{7}$$

$$\frac{\mathrm{dSPE}_j}{\widetilde{\mathrm{d}f_j}} = 0 \text{ leading to } \tilde{f}_j = \tilde{\xi}_j^{0^T}\tilde{x} \tag{8}$$

Now the method of unreconstructed variance can be presented. If the assumed fault is the actual fault in Eq. (8) $j = i$,

$$\tilde{f}_i = \tilde{\xi}_i^{0^T}(\tilde{x}^* + \tilde{f}\tilde{\xi}_i^0) \tag{9}$$

the addition of Eqs. (4) and (6) illustrates the effect of $f_i - f$ when comparing x_i with x^*

$$\|x^* - x_i\|^2 = (f - f_i)^2 = \left(\frac{\tilde{\xi}_i^T\tilde{x}^*}{\tilde{\xi}_i^T\tilde{\xi}_i}\right)^2 \tag{10}$$

The unreconstructed variance, u_i, in the direction ξ_i represents the variance of the projection $x^* - x_i$ on the fault direction ξ_i

$$u_i \equiv \mathrm{var}\{\xi_i^T(x^* - x_i)\} = \varepsilon\{\|x^* - x_i\|^2\} = \frac{\tilde{\xi}_i^T\varepsilon\{x^*x^{*T}\}\tilde{\xi}_i}{(\tilde{\xi}_i^T\tilde{\xi}_i)^2} = \left(\frac{\tilde{\xi}_i^T\tilde{\mathbf{R}}\tilde{\xi}_i}{(\tilde{\xi}_i^T\tilde{\xi}_i)^2}\right) \tag{11}$$

where $\tilde{\mathbf{R}}$ denotes the covariance matrix of the normal residual. Minimising u_i with respect to l

$$\min_l u_i \tag{12}$$

can be used to determine the number of principal components and the set of sensors to keep for process monitoring. The unreconstructed variance can be projected on the two subspaces

$$u_i = \hat{u}_i + \tilde{u}_i \tag{13}$$

In Dunia and Qin (1998), it is shown that $\tilde{u}_i$ is monotonically decreasing with respect to l, and $\hat{u}_i$ tends to infinity as l tends to m. Figure A1 illustrates this effect. Equation (12) only provides the optimal l for $\mathscr{J}_i$, considering the set of all possible faults $\{\mathscr{J}_j\}$,

$$\min_l q^T u = \min_l(q^T\tilde{u} + q^T\hat{u}) \tag{14}$$

where u represents the vector of unreconstructed variances for all $\mathfrak{J}_i\varepsilon\{\mathfrak{J}_j\}$, and q is a weighting vector with positive entries.

APPENDIX B

A brief description of the PLS technique is given in this appendix. For a full description, the reader is encouraged to review the extensive literature written on this methodology, for example Geladi and Kowalski (1986) and Jackson (1991).

The PLS (PLS based regression) technique operates in a similar form as PCR in the sense that a set of vectors is obtained from the predictor (input) variables. The main difference is that as each vector is obtained, it is related to the responses and the reduction of variability of the inputs. The estimation of the next vector takes into account this relationship, and

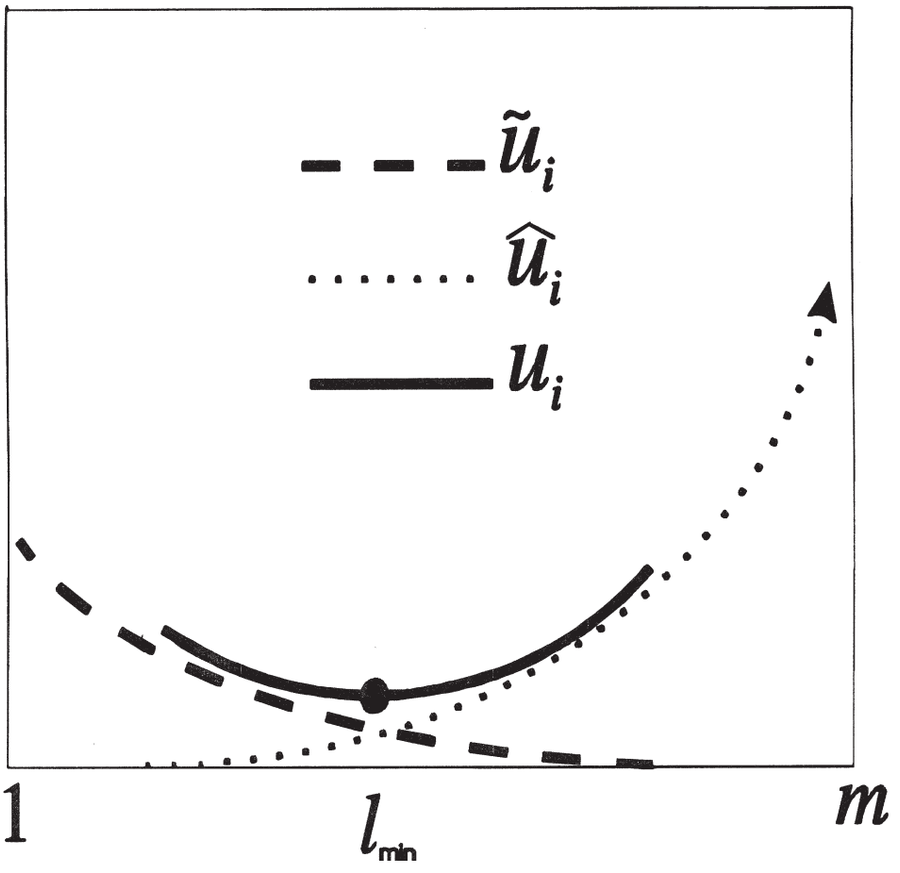

FIGURE A1 Unreconstructed variance as the summation of $\hat{u}_i$ and $\tilde{u}_i$.

simultaneously, a set of vectors for the outputs it is also obtained that takes into account such a relationship.

PLS has often been presented as an algorithm rather than a linear model, it is based on the NIPALS algorithm (a least squares algorithm for obtaining principal components). In this brief review of the method, the notation offered in Geladi and Kowalski (1986) has been used.

Consider **X** and **Y** real data matrices of sizes $n \times p$ and $n \times q$, respectively, representing n observations on p input and q output variables. The first step is to normalise both **X** and **Y** to zero mean and unit variance, then two operations are carried out together:

$$\mathbf{X} = \mathbf{TP} + \mathbf{E} (\mathbf{T} \text{ has size } n \times k, \ \mathbf{P} \text{ has size } k \times p, \text{ and } \mathbf{E} \text{ has size } n \times p)$$

$$\mathbf{Y} = \mathbf{UQ} + \mathbf{F}^* (\mathbf{U} \text{ has size } n \times k, \ \mathbf{Q} \text{ has size } k \times q, \text{ and } \mathbf{F}^* \text{ has size } n \times q)$$

$k \leq q$ is the number of vectors associated with **X**. **E** is the matrix of residuals of **X** at the kth stage (when $k = p$, $\mathbf{E} = 0$). $\mathbf{F}^*$ is an intermediate step in obtaining the residuals for **Y** at the kth stage.

In the singular value decomposition associated with PCA, matrices **Q** and **P** would be the characteristic vectors, and matrices **T** and **U** the principal component scores. These matrices

do not have the same properties in PLS, but may still be thought of in the same vein; **T** and **U** are referred to as *X-scores* and *Y-scores*, respectively.

It is possible to use regression to predict the output block of variables from the input one. This is done decomposing the **X** block and building up the **Y** block. In PLS, a prediction equation is formed by:

$$\mathbf{Y} = \mathbf{TBQ} + \mathbf{F}$$

where **F** is the actual matrix of residuals for **Y** at the kth stage, and **B** is a transformation matrix of size $k \times k$.

It is possible to calculate as many PLS components as the rank of the **X** matrix, but not all of them are normally used. In order to decide how many components (also referred to as latent variables) to use there are several methods advocated in the literature. One of them is using the number of components that minimises a measure of PRESS (predictive residual error sum of squares).

Josep M. Mirats i Tur received his title of Enginyer de Telecomunicacions (Electrical Engineering, specialised in electronics) in 1995, from the Universitat Politècnica de Catalunya (UPC). He finished his Ph.D. on qualitative modelling in the Institute of Robotics depending of both the UPC and CSIC, Centro Superior de investigaciones científicas (Scientist research Spanish council) in November 2001. Before joining, the Institute he was working for the private industry in the research department of the Seat-Volkswagen Company. He has been involved as research support engineer within the Institute for different European and CICYT (Comisión Interministerial de Ciencia y Tecnología) projects. His main scientific interests concerns simplifying the computation cost inherent to the existent qualitative modelling and simulation methodologies, concretely with the FIR methodology, and use it to model and simulate large-scale systems.

François E. Cellier received his B.S. degree in Electrical Engineering from the Swiss Federal Institute of Technology (ETH) Zürich in 1972, his M.S. degree in Automatic Control in 1973, and his Ph.D. degree in Technical Sciences in 1979, all from the same university. Dr Cellier joined the University of Arizona in 1984 as Associate Professor. His main scientific interests concern modelling and simulation methodologies, and the design of advanced software systems for simulation, computer-aided modelling, and computer-aided design. Dr Cellier has authored or co-authored more than 80 technical publications, and he has edited four books. He recently published his first textbook on Continuous System Modeling (Springer-Verlag, New York, 1991). He served as General Chairman or Program Chairman of many international conferences, most recently ICBGM'93 (SCS International Conference on Bond Graph Modeling, San Diego, January 1993), CACSD'94 (IEEE/IFAC Symposium on Computer-Aided Control System Design, Tucson, March 1994), ICQFN'94 (SCS International Conference on Qualitative Information, Fuzzy Techniques, and Neural Networks in Simulation, Barcelona, June 1994), ICBGM'95 (Las Vegas, January 1995), WMC'96

(SCS Western Simulation MultiConference, San Diego, January 1996), WMC'97 (Tucson, January 1997). He is Associate Editor of several simulation related journals, and he served as vice-chairman on two committees for standardization of simulation and modeling software. Dr Cellier was promoted to the rank of Full Professor in 1997.

Rafael M. Huber received his Ingeniero Industral (Electrical Engineering branch) and his Ph.D. in Ingeniería Industrial in 1976, both from the Universitat Politècnica de Catalunya (UPC). His present position is Catedrático de Universidad (Professor) at the Automatic Control Department of the UPC and nowadays he is serving as director of the Instituto de Robótica e Informática Industrial (IRI) depending of the UPC and the Spanish Council of Scientific Research (CSIC). His main scientific interests concern modelling and simulation methodology and the design of advanced simulation environments. Its present research focus qualitative modelling and simulation and its application to dynamic systems fault detection and diagnosis. He has been involved as research engineer or research head in projects with Spanish industry, the Comisión Interministerial de Ciencia y Tecnología (CICYT), the CSIC, the European Space Agency and the U.S. National Science Foundation. Prof. Huber has authored or co-authored more than 40 technical publications and edited two books related to continuous system modelling.

Dr S. Joe Qin is currently an Associate Professor in Chemical Engineering and Quantum Teaching Fellow in Chemical Engineering at University of Texas at Austin. He obtained his BS and MS degrees in Automatic Control from Tsinghua University in Beijing, China, in 1984 and 1987, respectively. He received his Ph.D. degree in Chemical Engineering from University of Maryland in 1992. His current research interests include process monitoring and fault identification, model predictive control, run-to-run control, system identification, microelectronics process control and diagnosis, chemical process monitoring and control, and control performance monitoring. He is a recipient of the NSF CAREER Award, DuPont Young Professor Award, and is currently an Editor for *Control Engineering Practice* and a Member of the Editorial Board of *Journal of Chemometrics*.