THE EFFECT OF BIASED REGRESSION ON THE IDENTIFICATION OF FIR AND ARX MODELS

Barry M. Wise and N. Lawrence Ricker

Prepared for presentation at the 1990 AIChE Annual Meeting
November 12-16, 1990.
Recent Advances in Process Control

November 15, 1990

UNPUBLISHED

The Effect of Biased Regression on the Identification of FIR and ARX Models

B. M. Wise and N. L. Ricker
Center for Process Analytical Chemistry and
Department of Chemical Engineering BF-10
University of Washington
Seattle, WA 98195

ABSTRACT

This paper discusses the effect of using the biased regression technique Continuum Regression for the identification of Finite Impulse Response (FIR) and Auto-Recursive eXtensive variable (ARX) models of dynamic processes. Continuum Regression (CR) encompasses the techniques of Principal Components Regression (PCR), Partial Least Squares (PLS) and Multiple Linear Regression (MLR). The theory behind CR is discussed and its use for the identification of single input single output (SISO) systems is demonstrated. Particular attention is paid to the "artifacts" that the CR method produces in the resulting models. For the case of FIR model identification this is done by considering how PCR, (one extreme of the CR technique), decomposes the input signal into frequency "bands". The resulting models fit the actual process model in the frequency "ranges" retained in the regression. Several factors affect the relative advantage of CR over MLR for FIR model identification. These include process noise, dynamics, input excitation and incorrect specification of time delays. The effect of these factors on the location of the optimum models in the CR parameter space is also examined. For FIR model identification, CR appears to offer significant advantages over MLR. However, there appear to be only a few situations, primarily under very adverse circumstances such as high noise, where CR is advantageous for ARX model identification. The reasons for this are discussed. The use of PLS and PCR for the identification of non-linear FIR models is also demonstrated using process data. The results are compared with polynomial regression and a linearizing transformation of the data.

1.0 Introduction

Greater attention has been focused on the identification of process models in recent years, in part due to the variety of controller design methods that rely on an explicit process model. The first step in the use of many of the techniques, such as IMC, DMC, QDMC, MPC etc., is to obtain a model of the process to be controlled. Unfortunately, theoretical models of most chemical processes are often quite difficult to obtain. In these situations models must be identified from actual process data.

In this paper the identification of Finite Impulse Response (FIR) and Auto-Recursive eXtensive variable (ARX) models is considered. Very little literature exists concerning FIR model identification in particular, even though this is the model type used in several of the control schemes mentioned above. A wider variety of methods is available for the identification of ARX models. For this reason, the FIR model identification problem is emphasized here.

In the following sections, the FIR and ARX model forms will be introduced. The Continuum Regression (CR) method is outlined, and its relationship to Principal Components Regression (PCR), Partial Least Squares (PLS) and Multiple Linear Regression (MLR) is shown. Simulations are used to determine the effects that a variety of parameters have on the identification of FIR and ARX models by CR. Finally, an often

overlooked non-linear modification of PCR and PLS will be considered for identification on non-linear FIR models. Examples are given using data from a bench scale process.

2.0 FIR and ARX Models

In an FIR model the process output, y, is considered to be a function of past values of the process input, u, only:

$$y(k) = b_0u(k) + b_1u(k-1) + b_2u(k-2) + ... + b_nu(k-n) \qquad (2.1)$$

Here the b_i are constants, and y(k) and u(k) refer to the process output and input at time k, respectively. The model of equation (2.1) includes a feed through term, $b_0u(k)$, (which is equivalent to a non-zero **D** matrix in a state-space model). If the model of the system included time delays, some initial terms would be zero, up to the number of sample periods of the delay. This relationship of (2.1) is perhaps more conveniently expressed in shift operator form. (See for instance Åstrom and Wittenmark 1984, Ljung 1987 or Kwakernaak and Sivan 1972).

$$y(k) = B(q^{-1})u(k) \qquad (2.2)$$

In theory, an impulse response model would have to include an infinite number of terms in order to truly model a dynamic process. Because the effect of inputs tends to zero after long time periods, the infinite series can be truncated with little adverse affect provided that enough terms are retained to cover the process settling time. A typical FIR model might have 25-100 coefficients, i.e. the order of the polynomial B(q) might be 25 - 100. FIR models are referred to as non-parametric because they make no assumption (besides asymtotic stability of the process to be modeled) concerning the underlying dynamics of the true process. With the exception of the specification of the order of B(q), a FIR model is not a fit of parameters to a predetermined form.

The FIR models shown above are for single input/single output (SISO) systems. The FIR model form can, however, include B polynomials for multiple inputs. This allows for modeling of multiple input/single output (MISO) systems. If multiple input/multiple output (MIMO) models are required it is common to use a collection of MISO models (Ljung 1988).

In contrast to FIR models, in ARX models the current output is considered to be a function of past values of both the input and output. Thus the ARX model takes the form

$$y(k) = a_1y(k-1) + ... + a_my(k-m) + b_1u(k-1) + ... + b_nu(k-n) \qquad (2.3)$$

This is more conveniently expressed in shift operator form as

$$A(q^{-1})y(k) = B(q^{-1})u(k - d) \qquad (2.4)$$

Here the d term has been added to indicate a possible delay of d time units.

Unlike FIR models, the ARX model is considered a parametric form. Once the orders of the A and B polynomials are determined, the range of plant behaviors which it is possible to model is restricted.

3.0 Continuum Regression

The regression techniques of MLR, PCR and PLS can all be unified under one approach which will be referred to as continuum regression (CR). The basic idea behind continuum regression has been discussed among chemometricians for some time, as was pointed out in the article by Lorber et. al. (1987). The descriptive name "continuum" comes from a paper by Stone et. al. (1990).

Before going into the CR algorithm, it is useful to review the techniques it encompasses and to define the estimation problem. We assume that the output variable, y (a scalar), is a linear function of the "input" variables **x** (a row vector). Thus, we would like to find the column vector **b** such that

$$y = \mathbf{x}\,\mathbf{b} \tag{3.1}$$

Note that in the case of FIR models the input vector **x** is made up of past values of the process inputs (u values in equation 2.1). For ARX models **x** consists of past values of the process inputs and outputs (u values and y values). Of course, in practice we would generally have a series of input output pairs available, producing a column vector **y** of outputs and a matrix **X** of inputs, where the number of columns in **X** is equal to the number of input variables (number of FIR coefficients or sum of ARX model orders) and the number of rows (in **X** and **y**) is equal to the number of samples available. Thus we assume a problem of the form

$$\mathbf{y} = \mathbf{X}\,\mathbf{b} \tag{3.2}$$

3.1 Multiple Linear Regression (MLR)

The most obvious way to estimate **b** is by least-squares, i.e. MLR (see for example Strang 1980 or Sharaf 1986):

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3.3}$$

The main problem with this approach is that the matrix $\mathbf{X}'\mathbf{X}$ may be poorly conditioned. Small perturbations of **y**, possibly due to noise, may greatly affect the result. Note that the MLR method obtains the best "fit" of the model to the available data in a least squares sense. The correlation between the data and the predicted outputs will be as high as possible. This is not the same as having a model which is optimum for prediction of new data, however, as the resulting MLR model may have fit some of the noise in the data, resulting in a model quite different from the "true" underlying model.

3.2 Principal Components Regression (PCR)

To obtain a better conditioned problem for the estimation of **b** it is common to use Principal Components Regression. In this case we would use either Principal Components Analysis (PCA) (Sharaf 1986) or the Singular Value Decomposition (SVD) (Strang 1980) to decompose the **X** matrix

$$\mathbf{X} = \mathbf{T}\mathbf{P}' = \mathbf{U}\mathbf{S}\mathbf{V}' \tag{3.4}$$

where

$$\mathbf{T} = \mathbf{U}\mathbf{S} \tag{3.5}$$

and $$\mathbf{P}' = \mathbf{V}' \quad (3.6)$$

We will use the PCA decomposition, where the "scores" matrix, **T**, has orthogonal columns and the "loadings" matrix, **P'**, has orthonormal columns. The vectors in **P'** are arranged such that each lies in the direction of greatest remaining variance in **X** after the variance in the previous vector directions is removed. The PCA decomposition thus provides a new basis set **P'** for the **X** matrix where each successive basis vector describes the remaining "trend" in the **X** matrix.

Once the PCA decomposition of **X** has been obtained, estimates of **b** can be calculated. Estimates of **b** are generally obtained using the first k vectors in **P'** (principal components or PCs) and in **T**. Thus **b** is estimated as

$$\hat{\mathbf{b}} = \mathbf{P}_k(\mathbf{T}_k'\mathbf{T}_k)^{-1}\mathbf{T}_k'\mathbf{y} \quad (3.7)$$

Note that this problem is well conditioned; since the scores vectors in **T** are orthogonal the matrix $\mathbf{T}_k'\mathbf{T}_k$ is non zero on the diagonal only and its inverse is easily calculated. The PCR procedure amounts to inverting only the "major trends" in **X** when estimating **b**. The implicit assumption in this procedure is that the major trends in **X** and in **y** are causally linked.

The problem is now to determine how many principal component vectors k to retain in the estimation procedure through application of a cross-validation scheme. Typically, part of the available data is used to estimate a collection of **b** vectors using different k values. Each of these vectors is then tested against the remaining data to determine the optimum k. Generally, as PCs are added, the model error gets smaller then goes through a minimum (Geladi 1986, Hoskuldsson 1988, Lorber 1987, Ricker 1988). The value of k at the minimum model error is then used to obtain a final estimate of **b** using all the available data.

Model error is usually quantified by the Predictive Residual Error Sum of Squares (PRESS) measure. The PRESS is defined as the sum of squared residuals between actual outputs and their predicted values based on a model that is identified from a separate data set. Thus

$$\text{PRESS} = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) \quad (3.8)$$

where the $\hat{\mathbf{y}}$ are generated as in (3.2) and it is assumed that the model used is <u>not</u> derived from the test data set. The PRESS is not a measure of model fit; it is a measure of model prediction error.

It is useful to understand the relationship between PCA, SVD and eigenvector decompositions. As mentioned previously, the right singular vectors (**V'**) of **X** and the principal components (**P'**) are equal. These are also equal to the eigenvectors of **X'X**. Furthermore, the singular values of **X** are equal to the square roots of the eigenvalues of **X'X**.

If the variables (columns) in the data matrix **X** have been mean centered (scaled to have zero mean), then **X'X**/(n - 1) (where n equals the number of samples) is referred to as the covariance matrix of **X** (Sharaf 1986). If **X** is "autoscaled", (columns scaled to have zero mean and unit variance), then **X'X**/(n - 1) is the correlation matrix. In the next

section we will use the fact that the eigenvectors of the correlation matrix are equal to the principal component vectors.

3.3 Partial Least Squares (PLS)

The Partial Least Squares technique strikes a balance between the PCR and MLR approaches. PLS is very similar to PCR except that the **P'** vectors (PCs) are allowed to rotate to achieve better correlation of the resulting **T** vectors (scores) with the output variable **y**. (The rotated PCs in PLS are commonly called latent variables or LVs.) Where MLR achieves the maximum correlation between two data blocks, and the PCR factors describe the maximum variance in the input block, PLS tries to do both. Many references are available concerning the PLS algorithm (Geladi 1986, Lorber 1987, Hoskuldsson 1988) so it will not be repeated here. The main point to remember is that algorithm finds factors with maximum covariance between the input and output blocks.

PLS is generally done using an algorithm known as NIPALS. In the first step an **X** block latent variable is found that maximizes the covariance between **y** and the projection of the **X** block onto that LV. The **y** vector is then regressed on the **X** block score to determine the inner relationship. **X** and **y** are then replaced; the subspace spanned by the LV is projected out the **X** matrix and **y** is replaced by its residual. The algorithm proceeds by finding each LV in turn using the new **X** and **y**.

3.4 Relation of Continuum Regression (CR) to PCR, PLS and MLR

As mentioned previously, MLR, PLS and PCR are all special cases of Continuum Regression. CR allows the PLS "balance" between finding LVs that are descriptive of **X** block variance (like principal components) and finding LVS that are highly correlated with the output (like an MLR regression vector) to be adjusted continuously. There are, in effect, an infinite number of techniques that lie between PCR and MLR, the difference being the relative weights given to achieving correlation versus describing **X** block variance when determining the latent variables. This is shown graphically in Figure 1. The terms "canonical variance", "canonical covariance" and "canonical correlation" are from work by Stone, and are quite descriptive of the previously discussed techniques. The CR method includes these techniques and all the ones in between.

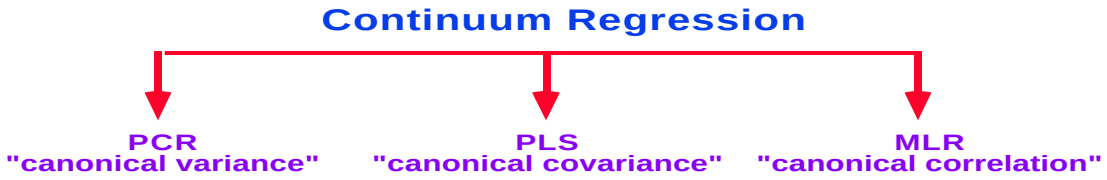

Figure 1. Relationship Between Continuum Regression and the Techniques of PCR, PLS and MLR.

There are several ways to implement the CR method. The implementation chosen here is different than Stone's or Lorber's but is functionally equivalent. The first step is to use the SVD to decompose the input data block **X**

$$\mathbf{X} = \mathbf{USV'} \qquad (3.9)$$

A new **X** data block, defined here as $\mathbf{X}^m$, is then formed by taking the singular values to the power m

$$\mathbf{X}^m = \mathbf{U}\mathbf{S}^m\mathbf{V}' \tag{3.10}$$

$\mathbf{X}^m$ and the original $\mathbf{y}$ are then used as inputs to a conventional PLS routine, and a set of regression vectors $\mathbf{b}$ is calculated using different numbers of latent variables. These regression vectors must be scaled to account for the modification of $\mathbf{X}$ given in (3.10)

$$\mathbf{b}_{scl} = \mathbf{V}\mathbf{S}^m\mathbf{S}^{-1}\mathbf{V}'\mathbf{b} \tag{3.11}$$

where it is assumed that the all zero rows of $\mathbf{S}$ below the main diagonal have been eliminated. Adjusting the power of m adjusts the PLS balance. As m goes to 0, the MLR solution is obtained; as m goes to infinity, the PCR solution results.

At this point, a few comments concerning the chosen CR algorithm are in order. This algorithm is particularly easy to implement, as it can be written as a shell around an existing PLS program. Furthermore, it is instructive because it is easy to understand. As m goes to zero, the resulting singular values all approach 1 and their differences become small. Under these circumstances, there is no "preferred direction" in the $\mathbf{X}$ block and covariance is maximized when correlation is maximized. In this case the MLR "correlation" solution results. When m becomes very large, the $\mathbf{X}$ block becomes highly directional and maximizing covariance becomes synonymous with finding the directions of $\mathbf{X}$ block variance. In this case the latent variables become equal to the principal components and the PCR "variance" solution results.

There are other algorithms for CR that are certainly more computationally efficient; this one requires that a complete SVD be performed. However, it is necessary to do only one SVD, and it can be used many times by taking the singular values to different powers. Typically, the PLS part of the algorithm takes much less computer time than performing the SVD.

In PCR and PLS a search is performed which uses cross validation to determine the number of latent variables that minimizes the model error (PRESS). In CR, however, we must now search over two variables: the number of latent variables and the continuum parameter. It is possible to view the PRESS as a function of the number of LVs and the continuum parameter as shown in Figure 2. In the figure the height of the surface represents the error for the model corresponding to the given number of latent variables and power of the singular values. The location of the PCR, PLS and MLR models is shown. In the figure the power of the singular values has been varied from 8 (next to PCR) to 1/8 (next to MLR) in logarithmically spaced intervals. In reality there is only one MLR model, however, this model error has been repeated down the back right side of the figure for continuity.

Figure 2 illustrates some of the features that are common to most CR PRESS surfaces. The level surface to the right in the figure, the "MLR plain", represents models identified with so many latent variables that they have converged to the MLR solution. All PLS and PCR techniques converge to the MLR solution as latent variables are added. The more correlation is factored in, the faster the convergence. On the left of the figure are models with large error, the "PCR mountain", identified with too few latent variables. In between is a "valley of best models" that have minimum PRESS. What we want to find is the model that represents the "bottom" of the "valley", i.e. the model with the minimum prediction error.

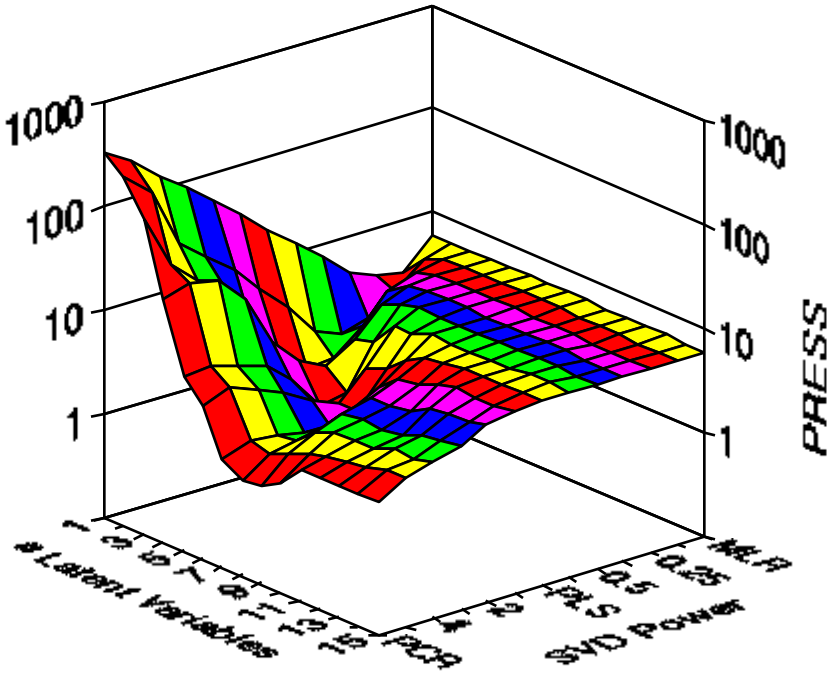

Figure 2. Continuum Regression PRESS Surface

4.0 Continuum Regression for FIR Model Identification

We would like to develop an understanding of how the CR method performs when used for the identification of FIR models and identify and any potential problems concerning its application. One concern in particular involves the creation of "artifacts" in the identified models. The approach we will take to understanding CR will be to first study the behavior of PCR, one extreme of the CR method. Specifically, we will look at how the decomposition of the input block proceeds, and relate this decomposition to the frequency domain behavior of the resulting PCR models. Once the effects of PCR are understood, the effects of the CR method will be much clearer. We will then look at how CR models are affected by variables such as process noise, input excitation and dynamic characteristics.

4.1 Frequency Domain Effects of PCR

When the data are arranged for identification of an FIR model, the resulting matrices resemble those shown in Figure 3 below. Here the first 3 samples are shown for identification a 6 coefficient FIR model. Note how the values of the input in the **X** matrix are repeated along diagonals. As pointed out in Box and Jenkins (p. 53), the correlation matrix of this **X**, [**X'X**/(n - 1)], approaches the Autocorrelation (or Autocovariance depending on scaling) Matrix (ACM) as more data are collected. Each entry in the Autocorrelation (Autocovariance) Matrix, $\mathbf{a}_{ij}$, is equal to the correlation coefficient (covariance) between u(t+j) and u(t+i). Thus, every value in the ACM depends upon only upon the difference of the indices.

---------- **X** ----------					**y**
u(5)	u(4)	u(3)	u(2)	u(1)	y(6)
u(6)	u(5)	u(4)	u(3)	u(2)	y(7)
u(7)	u(6)	u(5)	u(4)	u(3)	y(8)
:	:	:	:	:	:

Figure 3. Arrangement of Data for FIR Model Identification.

The ACM is special type of matrix known as a Toeplitz matrix, in which values are repeated along diagonals. The ACM is also symmetric. Furthermore, if the characteristics of the input signal u are known, the expected value of the ACM can be calculated. For instance, if u(k) is generated by passing a white noise signal c(k) through a first order filter

$$u(k) = \alpha\, u(k-1) + (1 - \alpha)\, c(k) \qquad (4.1)$$

then the expected values of the entries in the ACM are

$$\mathbf{a}_{ij} = (\alpha)^{|i-j|} \qquad (4.2)$$

As mentioned above, the PCR method relies on an eigenvector decomposition of the covariance or correlation matrix $[\mathbf{X}'\mathbf{X}/(n - 1)]$ which for FIR models is equal to the Autocovariance or Autocorrelation Matrix. Typically, this leads to eigenvectors (PCs) that have coefficients that look like sine and cosine curves. Such a case is plotted below in Figure 4, which shows the coefficients of the first five eigenvectors of an ideal ACM (where the entries in the ACM are equal to the expected values for large data sets). Because of the equality of principal component vectors and eigenvectors of the corresponding covariance matrix, the plots are identical to plots of the entries in the **p** vectors from PCA. This particular ACM was calculated assuming that the input signal was white noise filtered with a first order filter as in (4.1) where $\alpha = 0.8$. There would be 101 coefficients in the FIR model corresponding to the given ACM.

It can be shown that the eigenvector coefficients of an ACM resulting from white noise through a first order filter are identically sine and cosine functions for the continuous case (as the sample rate goes to zero). (This proof is given in the forthcoming dissertation by Wise.) For typical finite cases, where the number of coefficients is >10 and there are several hundred samples, the agreement between calculated principal components and pure sines and cosines is very good. Other ACM forms arrived at through higher order filters produce similar, though not identical, results. Certain cases, such as white noise through a second order under-damped filter, produce eigenvectors that still have periodic behavior, but are essentially combinations of frequencies and can be rather complex.

We can see that in some sense PCR breaks the input signal up into components of differing frequencies. This has a direct effect on the models obtained from the technique. Two identification experiments are used here to illustrate this effect. In the first case the true system is first order, while in the second case the true system is second order under-damped. Both systems have unit gain at steady state. In both identification experiments a Pseudo Random Binary Sequence (PRBS) input signal was generated by filtering white noise though a second order Butterworth filter and taking the sign of the result. The input signal was considered to be 1 when this result was positive and -1 when it was negative. The process models and filter parameters are given in Table 1. The calculated PRBS was used to generate a calibration set of 500 samples from both processes. PCR was then used

to identify models of the (noise free) processes. The frequency behavior of these models using different numbers of PCs was then tested.

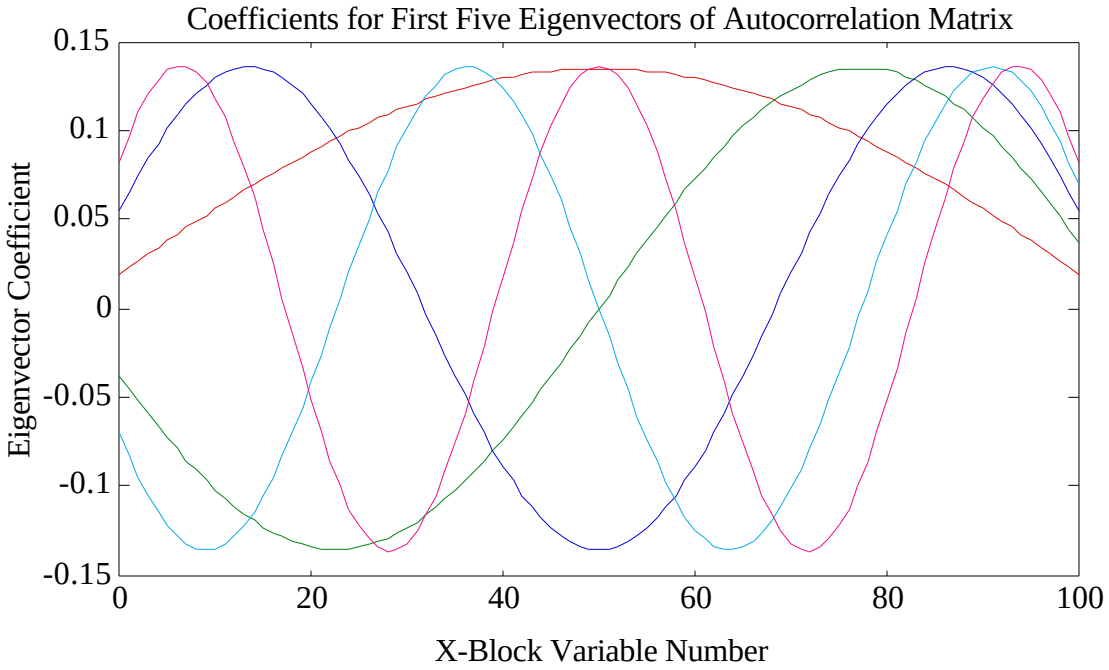

Figure 4. Coefficients in First Five Eigenvectors of Autocorrelation Matrix White Noise Process Through First Order Filter.

Table 1. Numerator [$A(q^{-1})$] and Denominator [$B(q^{-1})$] Polynomial Coefficients

	Numerator	Denominator
Case 1	.1426	1 -0.8574
Case 2	0.1129 0.1038	1 -1.5622 0.7788
Filter	0.0015 0.0029 0.0015	1 -1.8890 0.8949

The results of the identification are shown in Figures 5 and 6 which give the Bode gain magnitudes as a function of input frequency for the true system and the FIR models. The true response of the system is shown as the hatched line. The frequency response of the PCR identified FIR models, using from 1 to 6 PCs are also shown. Note how the 1 PC models accurately describe the process behavior at low frequency only. As PCs are added to the regression, the model matches the actual system response to progressively higher frequencies. The "dips" in the gain for the PCR models are a consequence of the sinusoidal nature of the PCs used to construct the FIR models. For instance, the FIR model identified using just 1 PC has coefficients that are a cosine function of a particular frequency. Other frequencies, therefore, can be orthogonal to this frequency and are not passed by the model. The gain "dips" occur, in fact, at frequency intervals of 2π, which would be expected based on the behavior of orthogonal cosine functions.

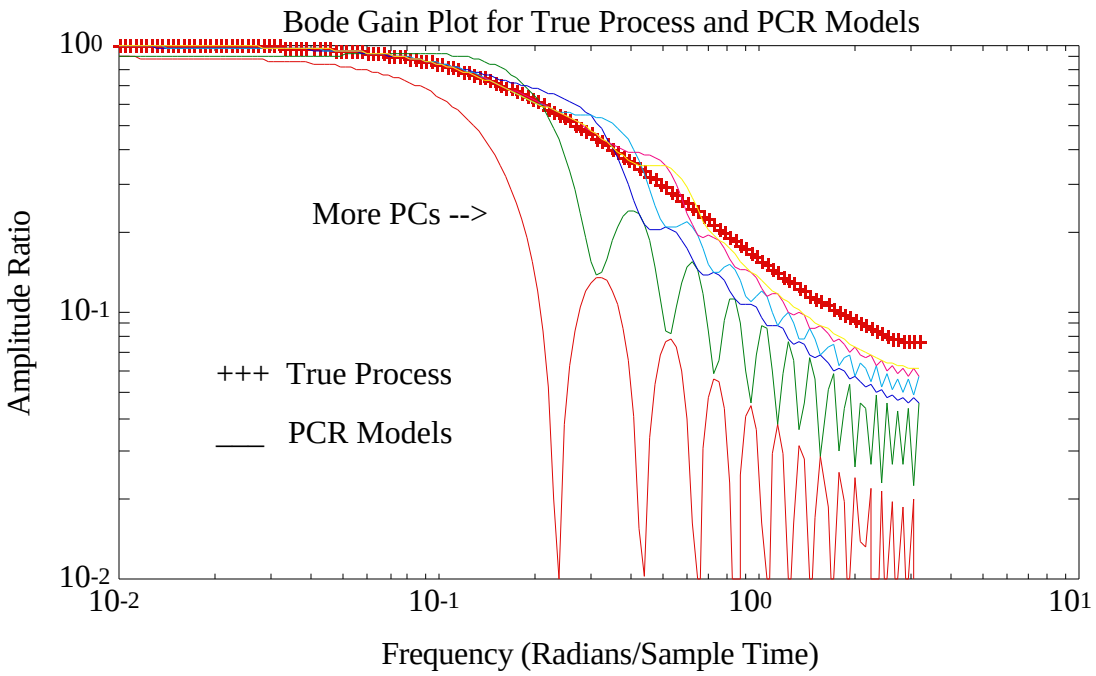

Figure 5. Bode Gain Magnitude Plot for First Order Process and PCR Models.

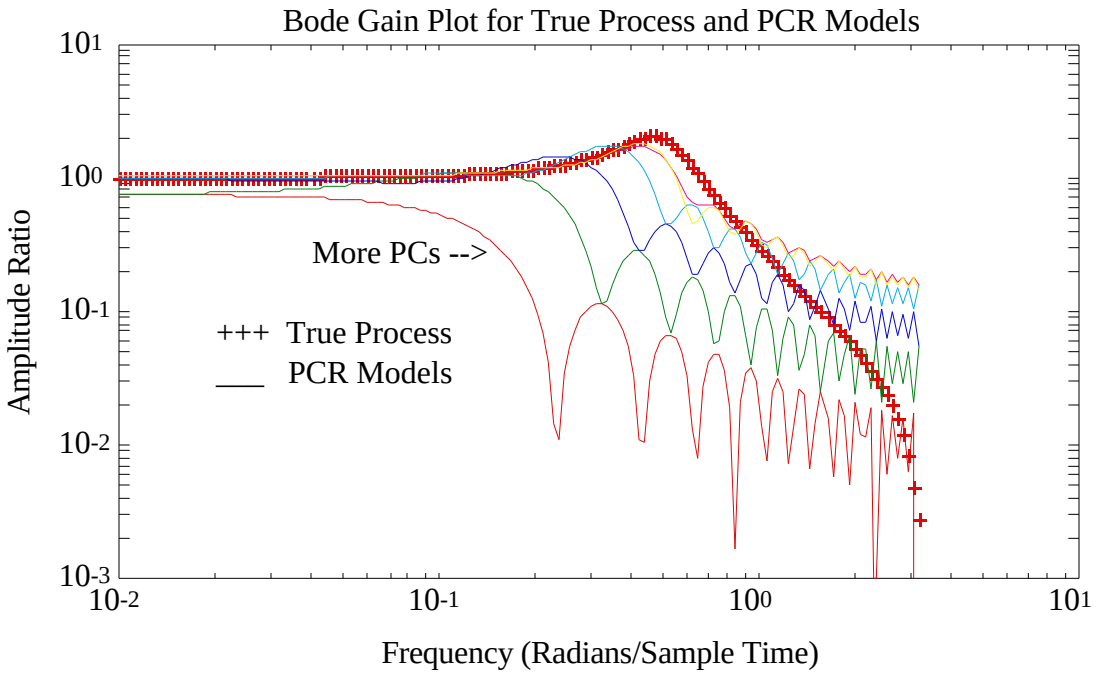

Figure 6. Bode Gain Magnitude for Second Order Process and PCR Models

The point concerning the consequences of the sinusoidal nature of the PCs is an important one and deserves some further investigation and clarification. Suppose for a moment that the first PC from the decomposition of the ACM can be represented as a

continuous cosine function of period 4π (frequency = 0.5) over the interval from $-\pi$ to π. Here the interval and period are chosen so that the function will go through one half cycle over the interval. For a sinusoidal input the process output y(t) of the 1 PC (continuous) FIR model must then be

$$y_1(t) = c_1 \int_{-\pi}^{\pi} \cos(0.5\,x)\,\cos(m(x+t))\,dx \qquad (4.3)$$

where m is the frequency of the input signal and c_1 is the constant determined from regressing $\mathbf{y}$ onto the scores vector $\mathbf{t}_1$. It can be shown that for values of m that are equal to an integer + 0.5 this integral is zero for all t. Thus we expect that frequencies of 1.5, 2.5, 3.5 etc. will not pass through the 1 PC model.

Let us further assume that the second PC can be represented as a sine function of period 2π (frequency 1) over the same interval. The contribution of the second PC to the model will then be

$$y_2(t) = c_2 \int_{-\pi}^{\pi} \cos(x)\,\cos(m(x+t))\,dx \qquad (4.4)$$

It can be shown that when m is an integer greater than 1 then the value of the integral in (4.4) is zero for all t. Thus we expect that the second PC will make no contribution to the 2 PC model at frequencies of 2, 4, 6 etc. and the response at these points will be equal to the 1 PC model response. It is also interesting to note that there are no values of m less than 1 for which the integral vanishes.

Considering once again Figures 4 and 5 we see that the behavior shown is expected based on the mathematical argument given above. The 1 PC model does not pass certain frequencies that occur at even intervals. The 2 PC model passes these frequencies but adds nothing to the centers of the intervals of the 1 PC model. Similar behavior continues as more PCs are added to the model. Because the "period" of the first and subsequent PCs can change given changes in the input signal, the exact location of the "dips" will also change. The behavior of the models will be similar in any case, however.

It was mentioned above that ACMs that are generated by processes other than white noise through first order filters have more complex decompositions. As Figures 4 and 5 demonstrate, even in the case where the data record is finite and the input signal was generated though a higher order filter, results are similar to the ideal ACM case. Numerical results indicate that FIR models identified from PCR will tend to fit first in the frequency ranges where there was the most power in the input signal. For instance, if a band-pass filter is used to generate the input signal, the models identified will be fit first in the frequency range where the input signal had the most power. As PCs are added to the regression, the models will fit in regions further away from the "band".

4.2 The Effect of Process Parameters on FIR Identification using CR

Now that we have an understanding of how the PCR method works for FIR model identification, we can proceed with investigating the effect of process parameters on models identified by the CR method. We would like to understand the effect that process

measurement noise level, process dynamics, amount of input excitation and mis-estimated time delays have on the location of the "best models" in the CR parameter space. We would also like to know how these factors affect the relative advantage of the CR method over MLR.

A collection of 7 representative process models will be used to test the CR procedure. The numerator and denominator polynomials of the models is given in Table 2. These models are intended to span a variety of dynamic behaviors typical of chemical processes. The processes are as follows: first order, second order over-damped, first order over second order over-damped, second order over-damped with right half plane zero, fifth order over-damped and fifth order under-damped. The Bode Gain plots of the test processes are shown in Figure 7. The frequency axis is scaled such that the Nyquist frequency (twice the sample frequency) corresponds to 2π on the plot.

Table 2. Numerator and Denominator Polynomial Coefficients for Test Models.

	Numerator	Denominator
Model 1	.1426	1 -0.8574
Model 2	0.0256 0.0215	1 -1.5503 0.5974
Model 3	-0.0742 0.1256	1 -1.5353 0.5866
Model 4	0.1707 -0.1330	1 -1.5834 0.6211
Model 5	0.1129 0.1038	1 -1.5622 0.7788
Model 6 numerator	0.0354 -0.0475 0.0222 -0.0041 0.0002	
Model 6 denominator	1 -2.9004 3.2427 -1.7335 0.4395 -0.0421	
Model 7 numerator	0.1976 -0.2723 0.1345 -0.0285 0.0023	
Model 7 denominator	1 -2.7819 3.0946 -1.6267 0.3753 -0.0278	

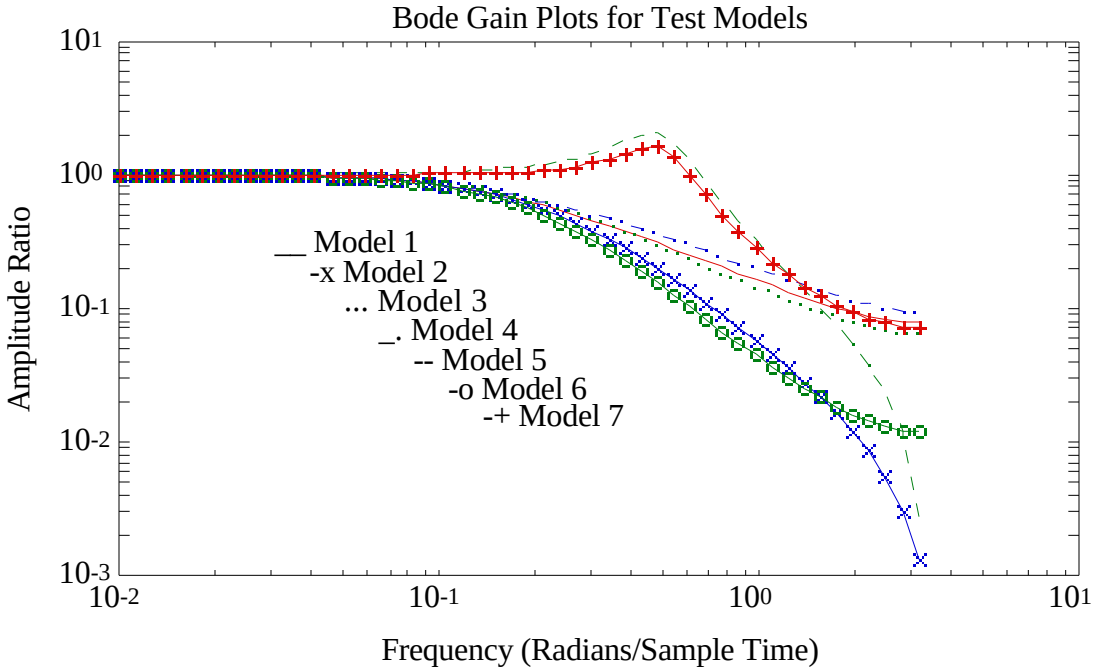

Figure 7. Bode Gain Magnitude Plots for the Seven Test Processes.

The first factor to be investigated is the effect of process measurement noise. In this series of tests the measurement noise level was varied from 5% to 100% of the process

gain. The FIR window width used in all of these experiments was 30 time steps. All processes were adjusted to reach 99% of steady state gain after 30 steps. A PRBS input was used to generate 500 input/output samples from all seven processes and with noise levels of 5, 10, 20, 40 and 100%. Models were tested with a similar PRBS.

Some typical model error (PRESS) surfaces for varying noise levels are shown in Figures 8 and 9. Figure 8 shows the PRESS surface for identification of a first order process with 5% noise, while Figure 9 shows the PRESS surface when 40% noise is added. Note how, at the 5% noise level, the optimum models are not much better than the MLR models. At higher noise levels the "valley" is much deeper relative to the MLR "plain" which has risen. Note also that the location of the "valley" shifts to fewer latent variables at the higher noise levels. This is consistent with PCR and PLS experience; when more noise is added to the data the best models generally are those with fewer latent variables. This same information is presented in more compact form in Figure 10, which shows the location of the valley bottom for all the noise levels tested. In Figure 10 each line gives the number of latent variables in the optimum model for each continuum parameter. For example, the figure shows that at a noise levels of 5, 10 and 20 % there are 8 latent variables in the optimum PCR model. When the noise level is changed to 40% there are 6 LVs in the optimum model and only 4 LVs in the optimum PCR model for a noise level equal to 100% of the process gain.

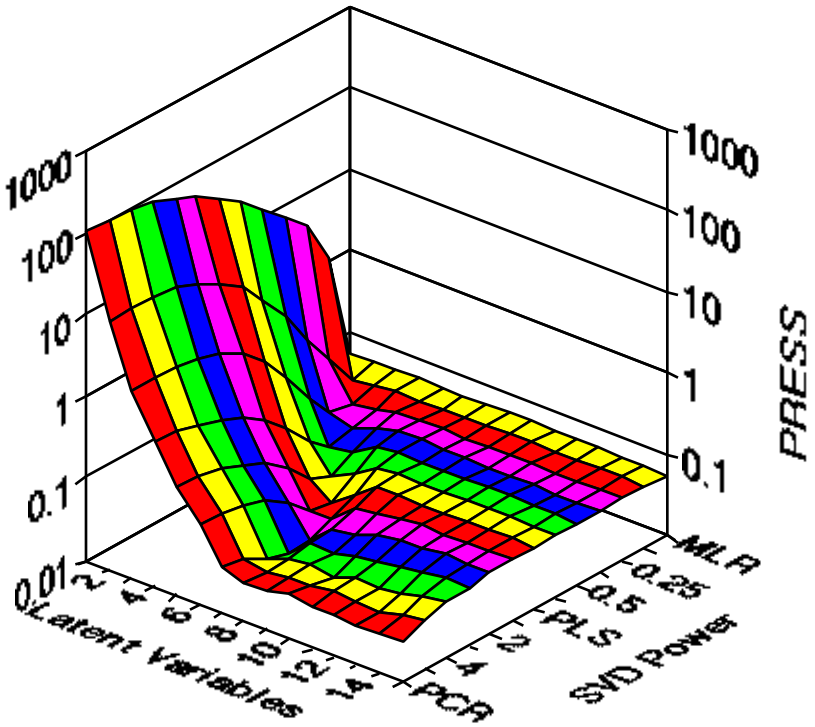

Figure 8. Continuum Regression PRESS Surface From Models of First Order Process with 5% Process Noise.

The effect of process dynamics on the location of the best models was investigated by identifying models using an identical input sequence, test sequence and measurement noise. Once again, 500 samples were generated for each process and and the same noise sequence equal to 20% of the process gain was added to each output. Figure 11 shows the location of the valley bottom for each process. Note how some processes require more latent variables to fit than others. The reason for the differences becomes more apparent after

consideration of the Bode Gain plots of the processes shown in Figure 7 and the Bode Gain plots of typical PCR models shown in Figures 5 and 6. The PCR models tend to fit the gain behavior up to a certain frequency (determined by the frequency of the last PC used) then drop off at the rate of about a decade per decade. Processes like the second order under-damped system which go through a maximum gain then drop off rapidly require a large number of PCs to fit this behavior accurately. On the other hand, processes such as the 4th over 5th order system with gain that starts to drop off at relatively low frequency and decline with slope = 1 are fit very well using only a few PCs.

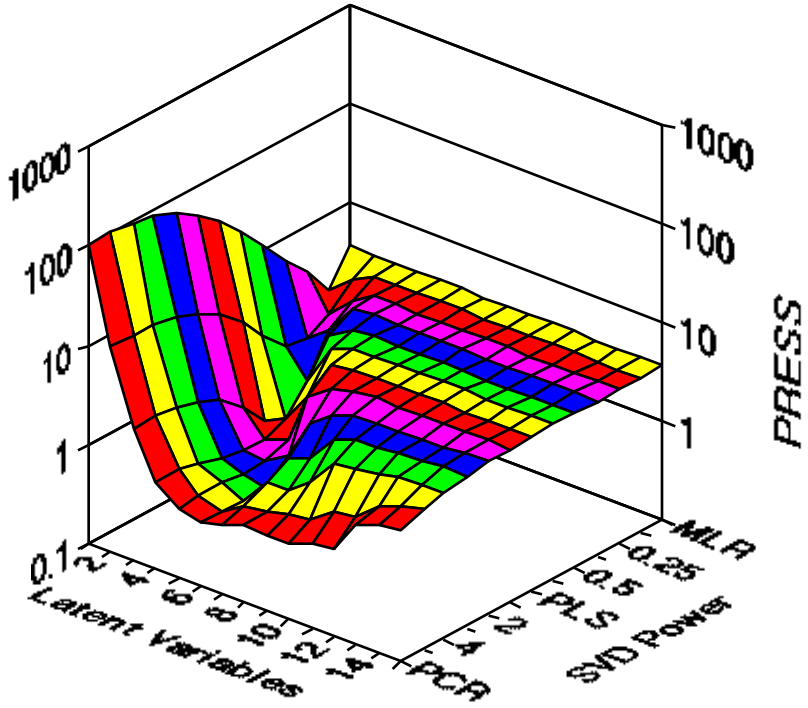

Figure 9. Continuum Regression PRESS Surface From Models of First Order Process with 40% Process Noise.

The effect of differing levels of input excitation on the location of the best CR models is shown in Figure 12. The lines plotted correspond to different filter cutoff points of the white noise input signal. The filter cutoff of 0.025 corresponds to an input signal with very little excitation, where a cutoff of 0.4 constitutes a large amount of excitation. Note how the lines in the figure cross. It appears that as input excitation is increased, more latent variables are required in the PCR side models while fewer latent variables are used in the MLR side models. This trend is understandable when one considers how the techniques work. For MLR models, a greater amount of input excitation keeps the autocorrelation matrix better conditioned and results in a good solution. For PCR models, when there is nearly equal amounts of power at all frequencies, the PCs for a finite data record may be descriptive of any frequency or combination of frequencies. To some extent these PCs appear quite arbitrary, and it may take a large number of them to adequately model the system response.

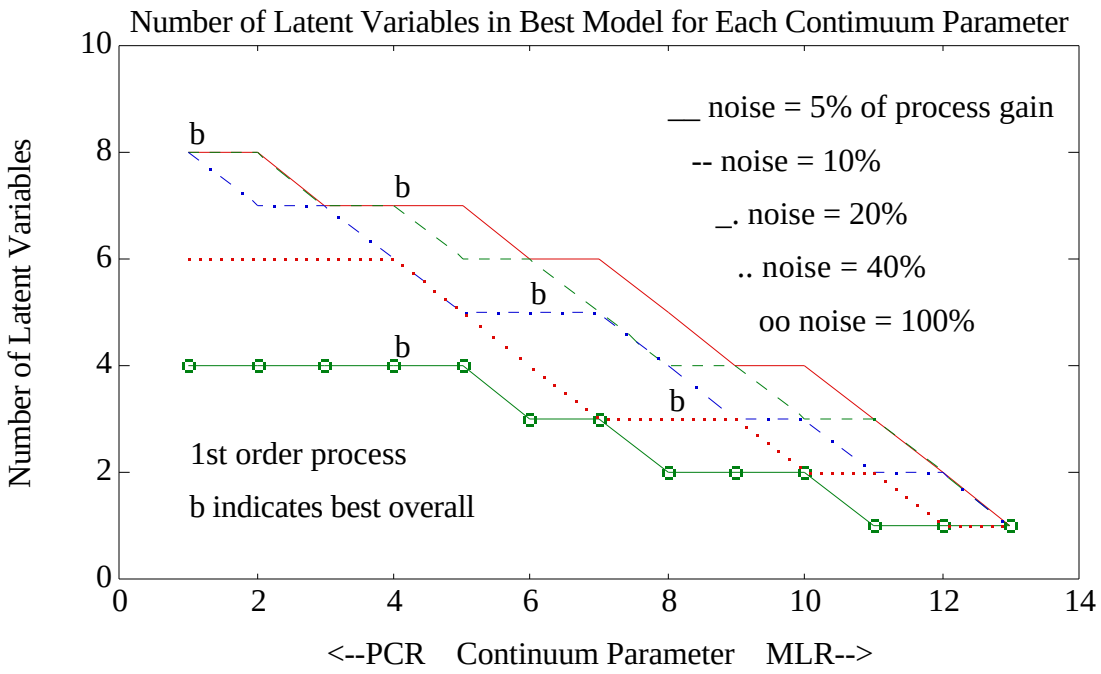

Figure 10. Effect of Process Noise Level on Number of Latent Variables in Best Model for Each Continuum Parameter

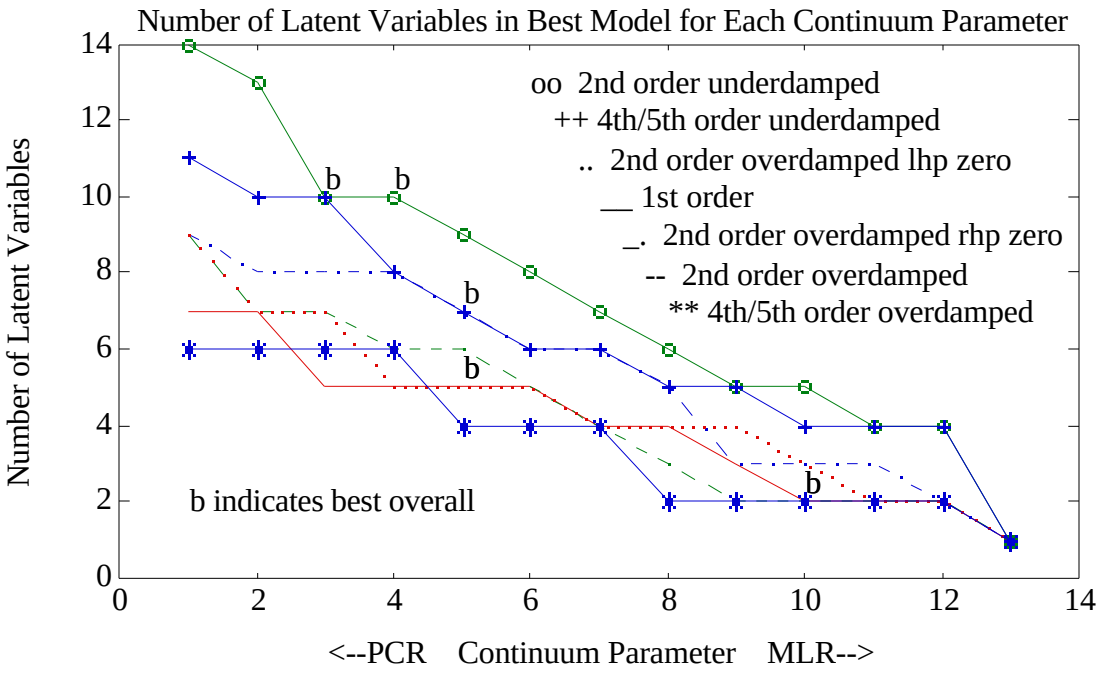

Figure 11. Effect of Process Dynamics on Number of Latent Variables in Best Model for Each Continuum Parameter

Processes often have pure time delays between input changes and output responses. If the length of these delays is known this is taken into account during the model identification. Only inputs that can effect the output are included in the regression matrix, which may mean that one or several of the left hand columns of **X** in section 4.1 are

eliminated The objective is to avoid estimating any coefficients that are known to be zero since any estimate is necessarily less accurate than the true value of zero.

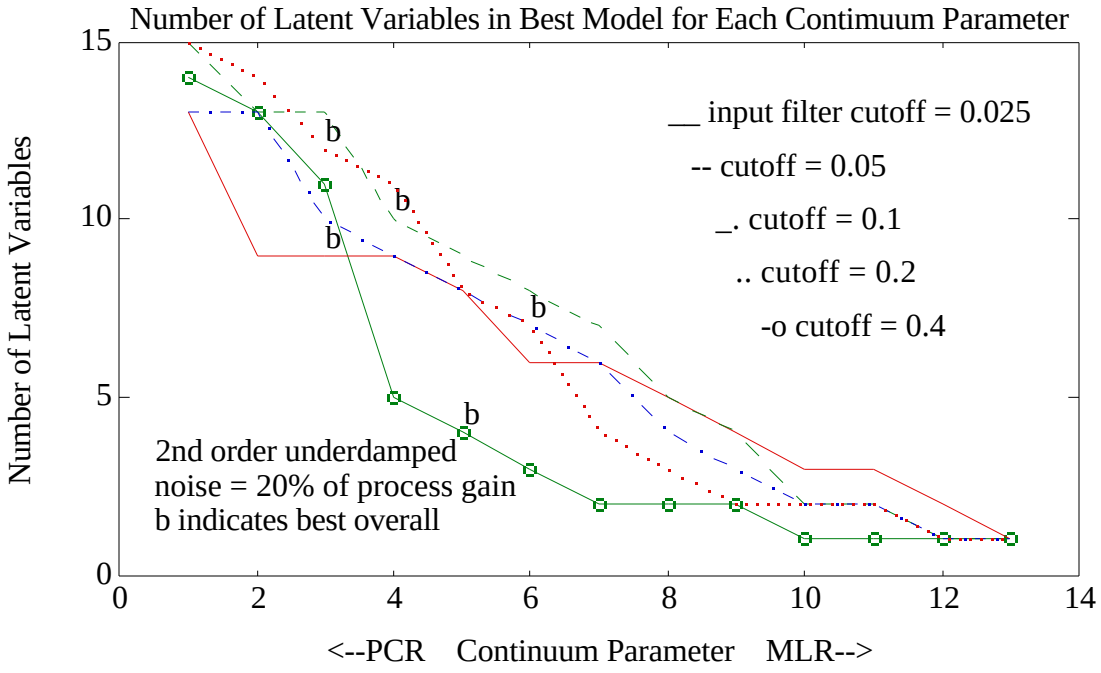

Figure 12. Effect of Input Excitation on Number of Latent Variables in Best Model for Each Continuum Parameter

The effect of incorrect estimation of time delays is shown in Figure 13, which shows the location of the best models for time delay errors of 0 (time delay equal to time delay assumed) to 5 units (time delay 5 units greater than that assumed). As can be seen from the figure, the best models for the mis-estimated cases typically have more latent variables than for the case where the time delay estimate was correct. This might be expected due to the way that the CR method imposes correlation on the FIR coefficients. The method can be thought of as approximating the true response with smooth functions that look like sines and cosines. The jump from a coefficient of 0 to some larger value is not smooth, however, and requires PCs that look like higher frequency terms in order to fit it. Incorrect estimation of the time delay by progressively larger amounts does not further increase the number of latent variables in the optimum models, in fact, there is some evidence that the number of latent variables actually decreases.

The actual model error for the best models is shown in Figure 14. It can be seen that the models identified by CR suffered a relatively larger adverse effect than the MLR models, but they are still better in an absolute sense. Note how the model error does not increase as the assumed time delay is incorrect by progressively larger amounts. It is important to point out that in the event of an incorrectly assumed time delay the leading FIR coefficients may not appear to be very near zero, and will also tend to look smooth as if they were modelling a true response. Thus, a cursory look at the leading coefficients may not lead one to check for a time delay. This is an artifact of the method, and the user should be aware of this potential problem. Furthermore, as Figure 14 shows, the model error improves dramatically when the system time delay is correctly estimated.

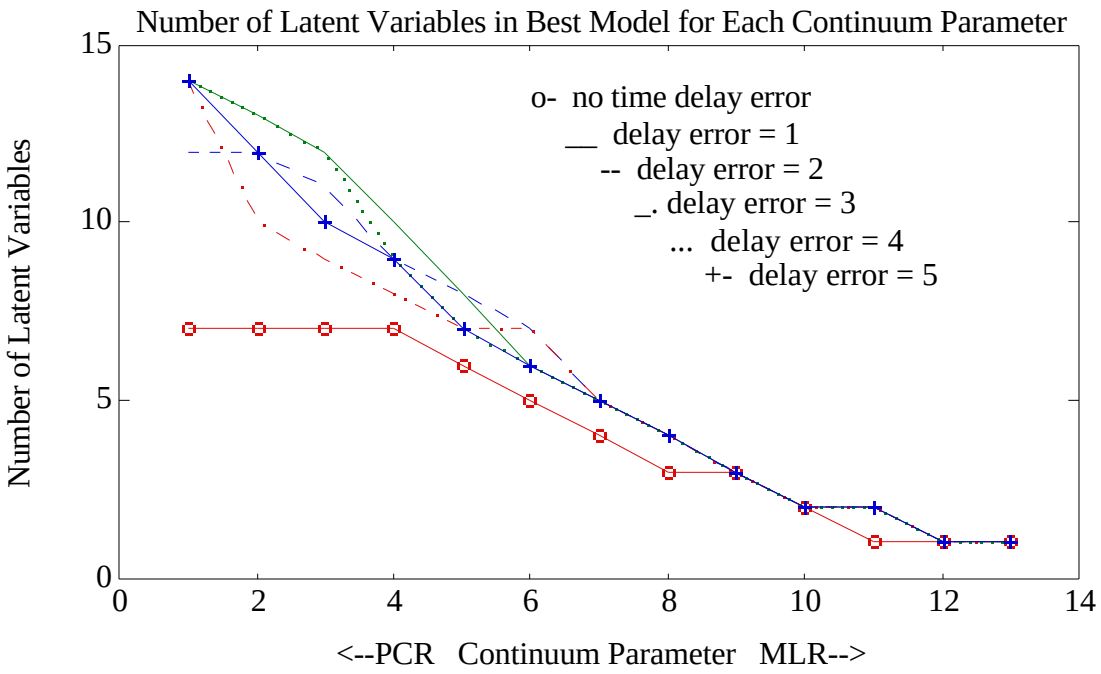

Figure 13. Effect of Mis-Estimated Time Delay on Number of Latent Variables in Best Model for Each Continuum Parameter

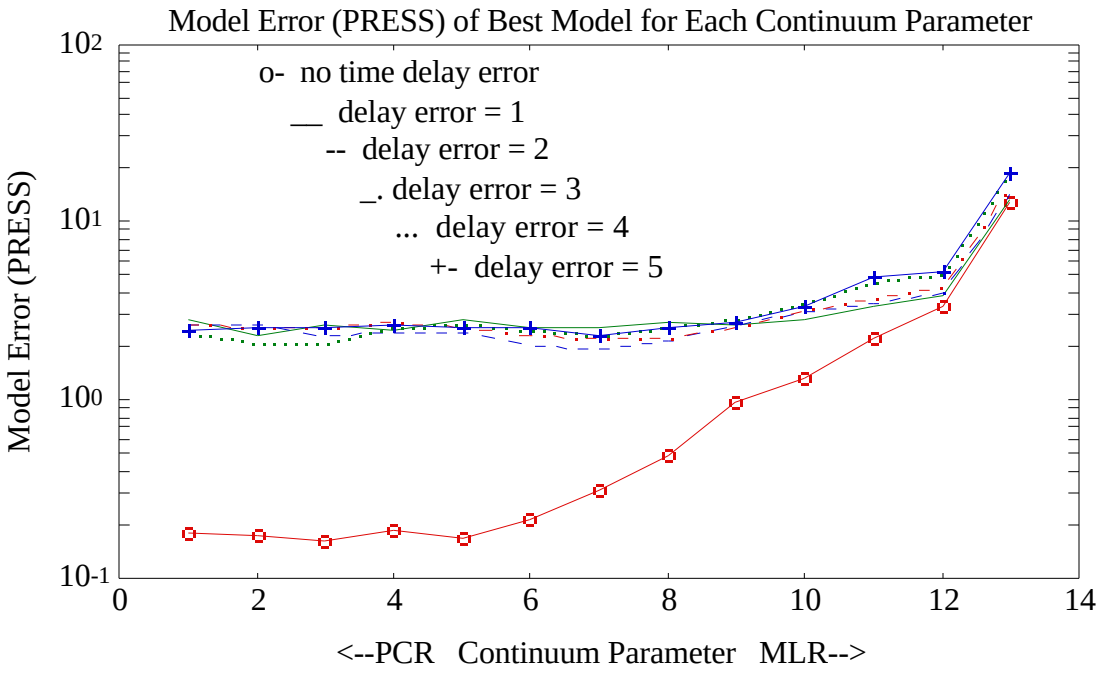

Figure 14. Model Errors for Correct and Incorrectly Estimated Time Delays.

Before leaving the subject of FIR model identification, it is interesting to look at the effect of the CR identification method on the FIR coefficients for a typical case. A simulation was performed where a first order model was used to generate an input/output data set consisting of 500 samples. The input was a PRBS and the process noise standard deviation was equal to 20% of the process gain. CR was used to identify FIR models of

the process with 30 coefficients. The models were then tested against the true model using three different inputs: white noise, a PRBS similar to the identification data and a step.

Figure 15 shows the true FIR coefficients of the true first order process along with some of the models identified from input/output data. The models shown are each "best" in some way, illustrating the point that the "best" model depends upon the test one chooses. The second model shown is most accurate relative to the true model when tested with a white noise input, while the third is best when tested against a PRBS similar to the one used for its identification. The fourth model has minimum error for step tests. Finally, the MLR model is the one that "fits" the calibration data best. Note the jaggedness of the MLR model relative to the others. This is a result of the "ill conditioning" of the problem due to the near rank deficiency of the input autocorrelation matrix and the correlation in the FIR parameters themselves.

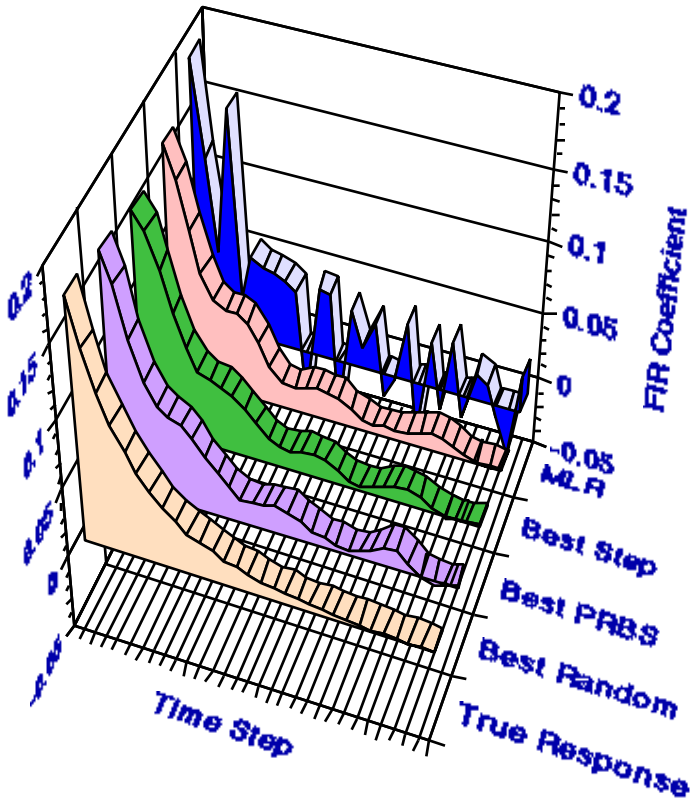

Figure 15. FIR Coefficients Identified from Second Order Over-Damped System with Noise = 20% of Process Gain

5.0 Continuum Regression for ARX Model Identification

In this section the effect of using Continuum Regression for the identification of ARX models is considered. It might be expected that there is less to be gained from using CR for ARX identification. There are typically fewer parameters to estimate and the correlation between them is less significant. Furthermore, from a theoretical standpoint, a properly posed ARX regression problem is not ill conditioned. If the proposed orders are either correct or less than the true orders of the process, the **X** matrix should be of full rank (provided that adequate input excitation has been used to generate the data). In fact, a rank

deficient **X** matrix is an indicator that the problem has been over-parameterized, as pointed out in Ljung (1987).

Unlike the FIR case, it is difficult to characterize the result of a PCA decomposition of the input correlation matrix for ARX regression, i.e. there does not appear to be a simple frequency domain interpretation of this decomposition. This is due to the more complex nature of the correlation matrix for ARX models. The **X** block includes both input and output values; this leads to a correlation matrix that includes correlations between lagged inputs and lagged outputs.

Numerical results indicate that ARX models identified with PCR show behavior that is similar to that seen in FIR identification. An example of this is Figure 16, which shows the Bode Gain plots of ARX models identified with PCR. The true system is model 7 from Table 2, a 4th over 5th order under-damped system. The process orders were correctly specified in the model identification procedure, and the system was noise free. A PRBS input was generated using the filter in Table 1. The ARX models identified using 1-5 PCs are shown. As in the FIR case (compare with Figures 5 and 6), the 1 PC model accurately describes only the low frequency behavior of the system. Additional PCs improve the high frequency accuracy. There is also some evidence of the pattern of gain "dips" as seen in the FIR case, but this artifact is not nearly as pronounced here.

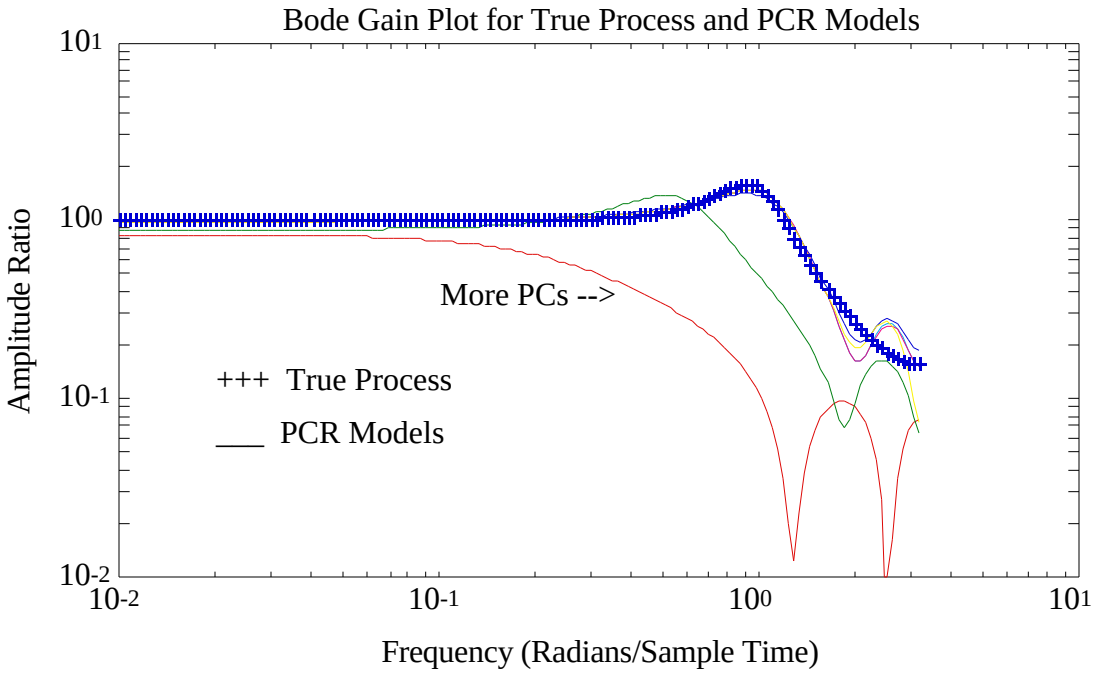

Figure 16. Bode Gain Plots of Process Model 7 and ARX Models Identified with PCR.

Tests were performed to investigate the effect of different factors on the quality of ARX models identified by CR. The same factors used in the FIR model identification experiments were considered, along with the additional possibility of an incorrect choice of model order. This was not a factor for FIR model identification because FIR models make no assumption about process order.

In almost all the cases considered, there was little difference between the best CR models and those obtained with MLR. In particular, when test processes with low orders were used, differences between the best CR models and the MLR models were typically

very small. In these situations the optimum CR models were often those that used all the latent variables, i.e. the MLR solution. The largest differences were for the cases of very high noise, over-parameterized models and very little input excitation. Generally, the trends observed in FIR identification with CR were similar to those observed in ARX model identification, but were usually the less clear than in the FIR case. Some highlights of the simulations performed are given in the paragraphs that follow.

PRESS surface plots for a very high noise test are shown in Figures 17 and 18. This test was performed using model 7 in Table 2 as the true system. The model orders were correctly specified and the noise standard deviation was 40% of the process gain. The input was a PRBS. In Figure 16, a PRBS with frequency content similar to that of the calibration data input was used to test the models. In Figure 17 a white noise signal was used for the test. Note how most of the best CR models for each continuum parameter are only slightly better than the MLR model. The best overall model (with an SVD power of 2 and 5 latent variables) has about a 35% smaller error in the PRBS test. Also, while the PRESS surface is similar to that seen in the FIR case, the "valley" is not nearly so well defined and is somewhat fragmented. There are also multiple local minima.

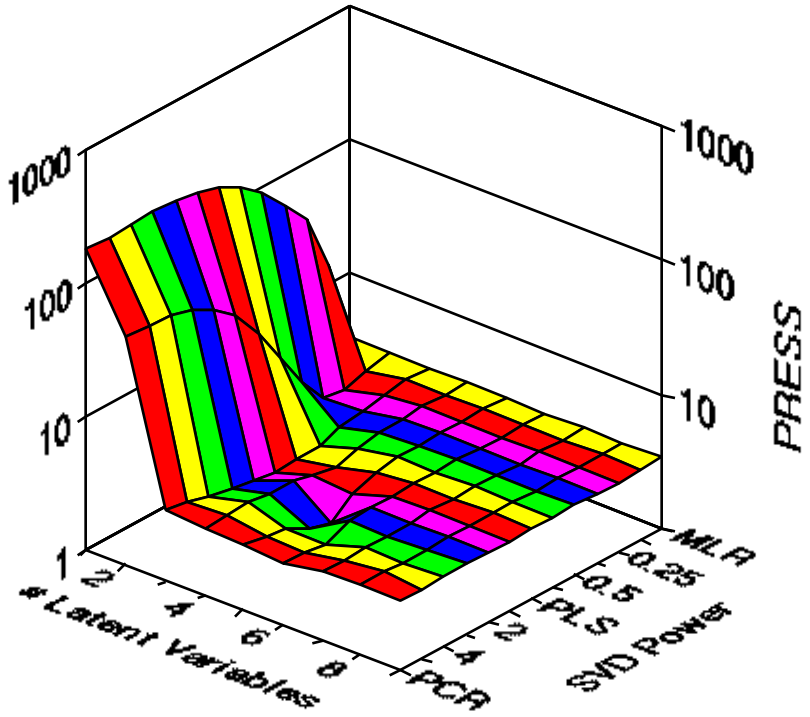

Figure 17. PRESS Surface for ARX Models Tested Against PRBS.

It is also apparent that the valley is more pronounced in Figure 18 where a random input was used to test the model. This is consistent with the FIR identification experience in the previous sections. A random input signal emphasizes higher frequencies more than a PRBS, and in the tests performed here CR tends to get models with better behavior in the high frequencies. Thus the PRESS valley is more pronounced when a random signal was used as a test.

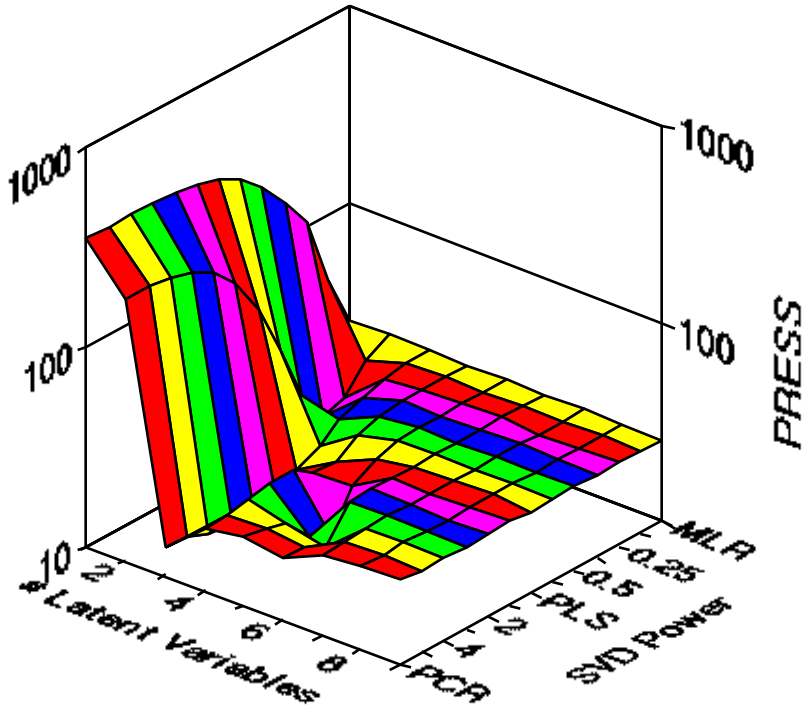

Figure 18. PRESS Surface for ARX Models Tested Against Random Input.

The effect of process noise on the location of the best models in the CR parameter space is shown in Figure 19. The figure shows the number of latent variables in the best model for each value of the continuum parameter. The figure is based on a test using a PRBS. Note that while the general trend is as we expect, i.e., there are fewer latent variables used at higher noise values, there is some irregular behavior in the best model location. In particular, there are several instances where there are jumps to more latent variables as the regression technique moves towards MLR. This was seen in FIR identification only very rarely. In the ARX identification experiments performed here it was very common. Tests with over-parameterized models showed similar behavior with added noise.

The effect of over-parameterization was also investigated. As mentioned previously, CR had a larger relative advantage over MLR with over-parameterized models. Even in highly over-parameterized models such as specifying sixth over seventh order when the true process is fourth over fifth order the difference is not great. The location of the best models in the CR parameter space tended towards more latent variables as higher process orders were specified, though not as fast as the order increased. For instance, if both the numerator and denominator orders were specified to be higher than the actual process order by 2, the number of latent variables at optimum might increase by 2, but not by 4.

The effect of input excitation was similar to that found in FIR model identification. ARX models identified by MLR degraded more rapidly than those identified by CR as input excitation was decreased. The optimum models also tended towards fewer latent variables, though this effect was not as pronounced as with FIR. As an example, the PRESS surface for a very high noise case with very low input excitation is shown in Figure 20. The true model is number 7 from Table 2. The input for the calibration set was 7 random steps (4 up and 3 down) over a 500 sample time period. Process noise was 40% of process gain and the model orders were correctly specified. Only in these very adverse

situations is the difference between the best CR models and the MLR model this pronounced.

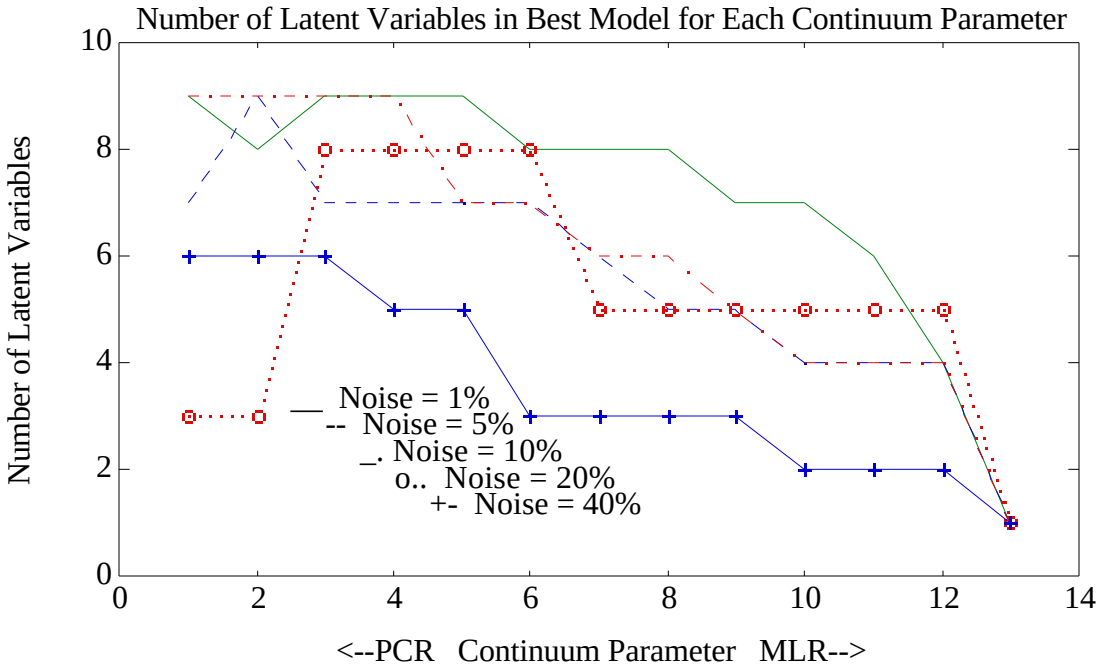

Figure 19. Number of Latent Variables in Best Models for Different Noise Levels, PRBS Test.

The Bode Gain plots of the best models from the high noise low input excitation case are shown in Figure 21. This figure demonstrates that the CR models are closer to the true response at most points, but not at all. Once again, the best model depends upon the test specified.

When using CR for ARX model identification in practice, finding the optimum model may be quite difficult. As we have seen, the minima can be quite shallow. These would be even more difficult to find when testing the models against noisy data sets. We have had the luxury here of testing the models against the true system. Even so, the location of the best model is often unclear.

In summary, it appears that CR offers advantages for ARX model identification only in very adverse cases. Under most circumstances, the relatively small potential gains in model accuracy combined with the difficulty in identifying the best models make MLR a more predictable choice of methods.

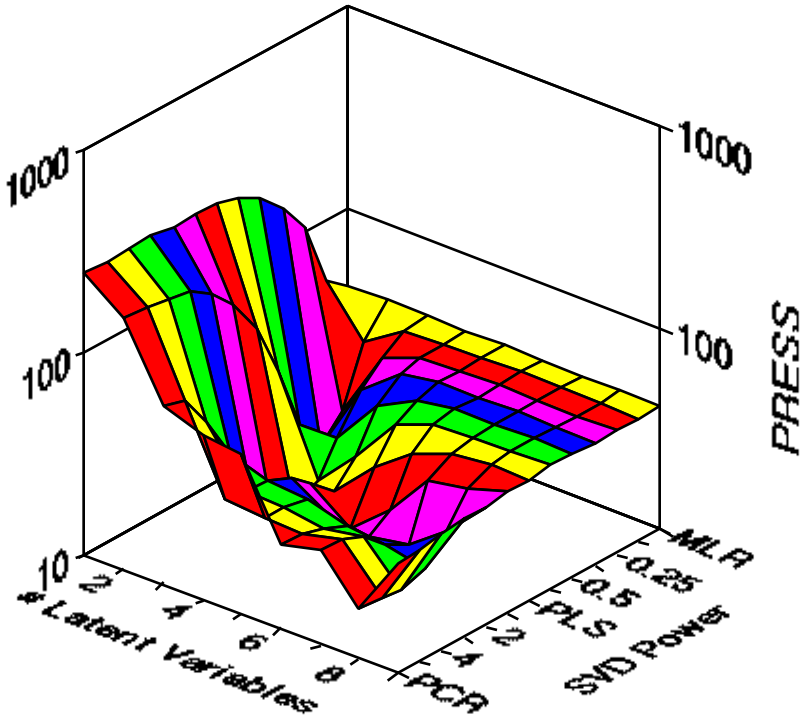

Figure 20. PRESS Surface for High Noise Low Input Excitation Case.

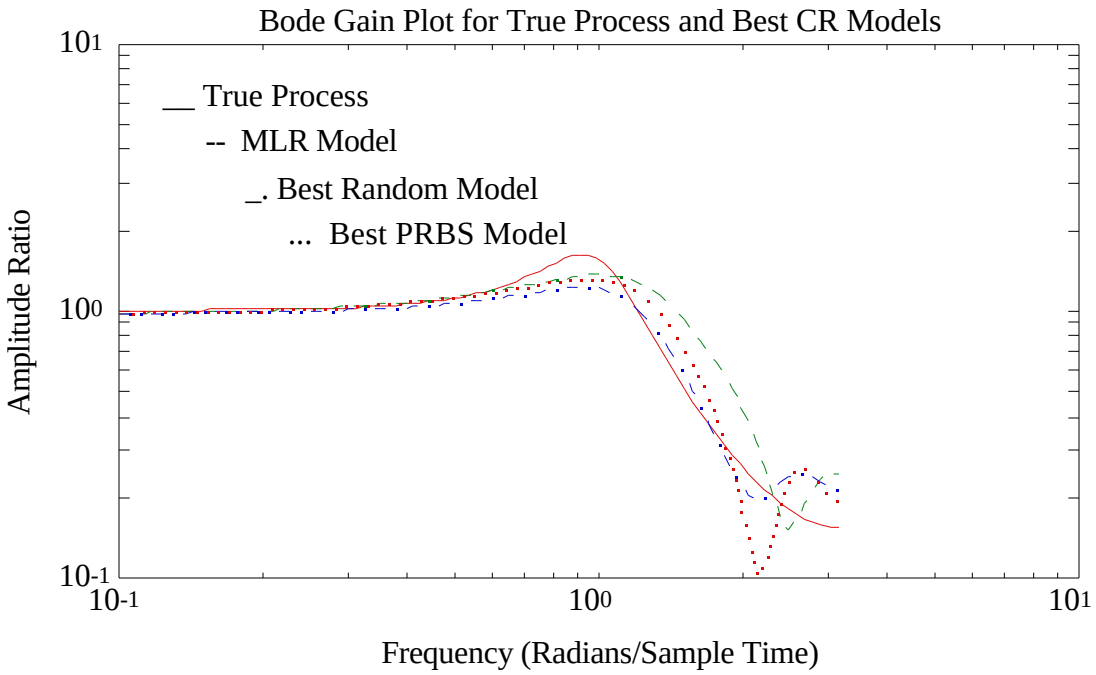

Figure 21. Bode Gain Plots for True Process and ARX Models Determined by CR and MLR.

6.0 An Example of PCR and PLS for Non-Linear FIR Model Identification

While PCR and PLS are linear techniques, it is easy to modify them for non-linear model identification. In this section the basic ideas behind non-linear PCR and PLS are discussed and an example of identification of a non-linear process model is given. It is not intended that this section provide a comprehensive treatment of the subject of non-linear biased regression. Instead, this example serves as an indicator of the potential of biased non-linear techniques.

A non-linear version of PCR can be implemented by proposing a non-linear relationship between the **X** block scores **T** and the output **y**. The form of the non-linear relationship can be arbitrary (such as a polynomial), or it may be arrived at through theoretical consideration of the process. Often, plots of the **X** block scores versus **y** will suggest a particular non-linear relationship.

As an example of identifying a non-linear PCR model, imagine that we have proposed that the output y should be fit to the first k **X** block scores $\mathbf{T}_k$ using a second order polynomial. Thus we are proposing that

$$y = [t_1^2 \cdots t_k^2 \; t_1 \; \ldots \; t_k \; 1] \, \mathbf{b} \tag{6.1}$$

where **b** is a vector of regression coefficients to be determined. Taking some liberties with the notation, this can be rewritten for a collection of input output pairs as

$$\mathbf{y} = [\mathbf{T}_k^2 \; \mathbf{T}_k \; \mathbf{1}] \, \mathbf{b} \tag{6.2}$$

where it is understood that $\mathbf{T}_k^2$ indicates squaring the elements of $\mathbf{T}_k$, and that **1** indicates a vector of ones of appropriate length. Once the PCA decomposition of the **X** block has been obtained, the vector **b** can be estimated with the normal equations

$$\hat{\mathbf{b}} = ([\mathbf{T}_k^2 \; \mathbf{T}_k \; \mathbf{1}]' \, [\mathbf{T}_k^2 \; \mathbf{T}_k \; \mathbf{1}])^{-1} [\mathbf{T}_k^2 \; \mathbf{T}_k \; \mathbf{1}]' \, \mathbf{y} \tag{6.3}$$

This procedure can be contrasted with polynomial regression. If a second order polynomial fit is proposed, then the resulting relationship is

$$\mathbf{y} = [\mathbf{X}^2 \; \mathbf{X} \; \mathbf{1}] \, \mathbf{b} \tag{6.4}$$

where the notation is as in equation (6.2). The matrix in brackets in (6.4), however, is no better conditioned than **X** alone, and is possibly worse. Thus it can be seen that this problem is at least as unstable as the typical linear model problem.

Non-linear versions of PLS are accomplished in an analogous manner. A non-linear form is proposed for the PLS inner relationship. At each step the algorithm determines vector in the X block whose scores have the highest covariance with the y block residual. The scores are fit using the desired non-linear relationship, the estimates of y are calculated and then subtracted off the residual.

As an example, non-linear PLS and PCR were applied to data from the tank apparatus described in Haesloop and Holt (1990). This system consists of a tank with an outlet designed expressly so the tank outflow rate would be a highly non-linear function of the tank level. The output from this system is the voltage signal from the level measurement

device. The input is the voltage signal to a pump which supplies water to the tank. The system gives a highly non-linear input/output relationship. Haesloop and Holt used this process to test a neural-net identification method. We have used it here to test non-linear PLS and PCR for identifying a non-linear FIR model.

Haesloop performed a series of identification experiments where the input was varied in a pseudo-random fashion and output data were collected. An example of such an experiment, which is used here as calibration data, is shown in Figure 22. The values of the input and output correspond to voltages.

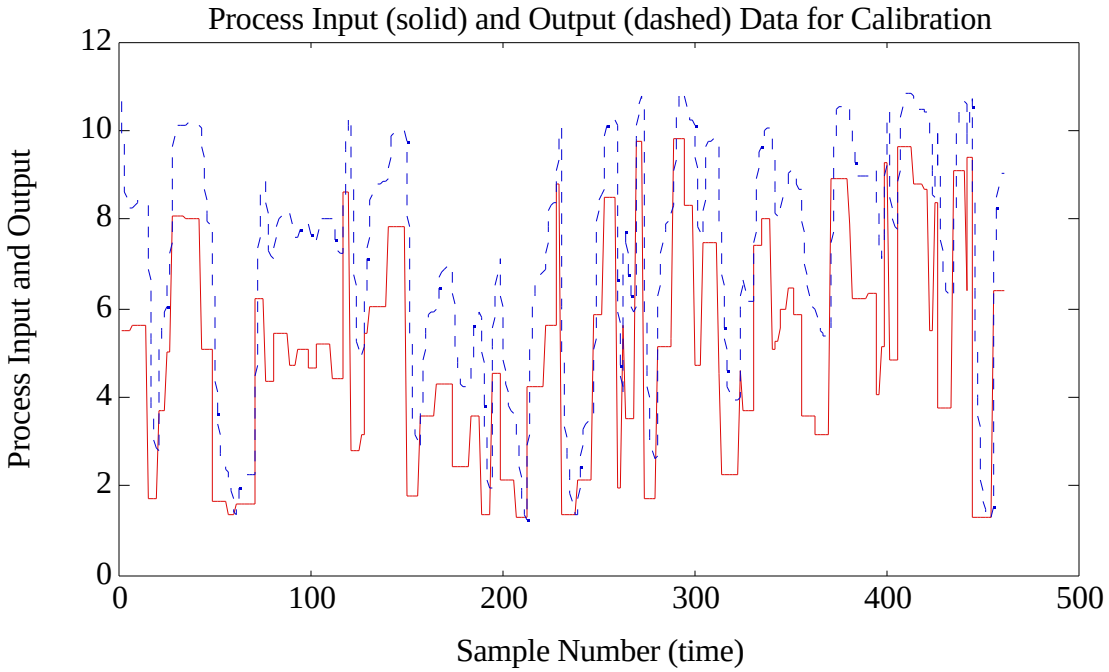

Figure 22. Non-Linear Tank Process Input/Output Data Used for Calibration

After scaling the input/output data to zero mean and unit variance several linear FIR models were identified using different numbers of past input values. The best results were when the last 6 values of the process input were used. At this point PLS and PCR models with polynomial inner relationships were tried with the data. Several modeling attempts showed that fixing the polynomial to be second order provided better predictive models than higher order. A series of cross validation tests using the calibration data showed that the PLS model with 3 latent variables provided the minimum prediction error while the PCR model with 5 latent variables did best. For comparison, a model was formed using second order polynomial regression as in equation (6.4). The estimate of b was obtained using the normal equations. Several attempts were also made to linearize the input/output data and fit a linear model. The best choice observed was to simply square the process output.

The fit of all of the models to the data is shown for a 140 point segment in Figure 23. The fit errors (sum of squared residuals) are given in Table 3. Note that the fit error of all the non-linear models and transformed linear model are about equal. The linear model fit is considerably worse. The polynomial regression model has the best fit to the data.

The models were then tested using data from another input/output experiment. Each of the models was used to predict the process output and the PRESS was calculated for

each model. The PRESS numbers are given in Table 3. In this test the FIR model identified using non-linear PLS model was found to have the smallest prediction error, followed by the non-linear PCR model and the polynomial regression. The linear model applied after the data transformation was somewhat worse than the non-linear models, and the strictly linear model had the largest error.

Table 3. Model Fit Errors and Prediction Errors for Non-Linear Process

Model	Fit Error	Prediction Error (PRESS)
Linear FIR	211.0	105.5
Non-linear PLS	26.1	3.9
Non-linear PCR	27.8	7.2
Polynomial Regression	20.1	8.4
Transformed Linear	27.5	11.8

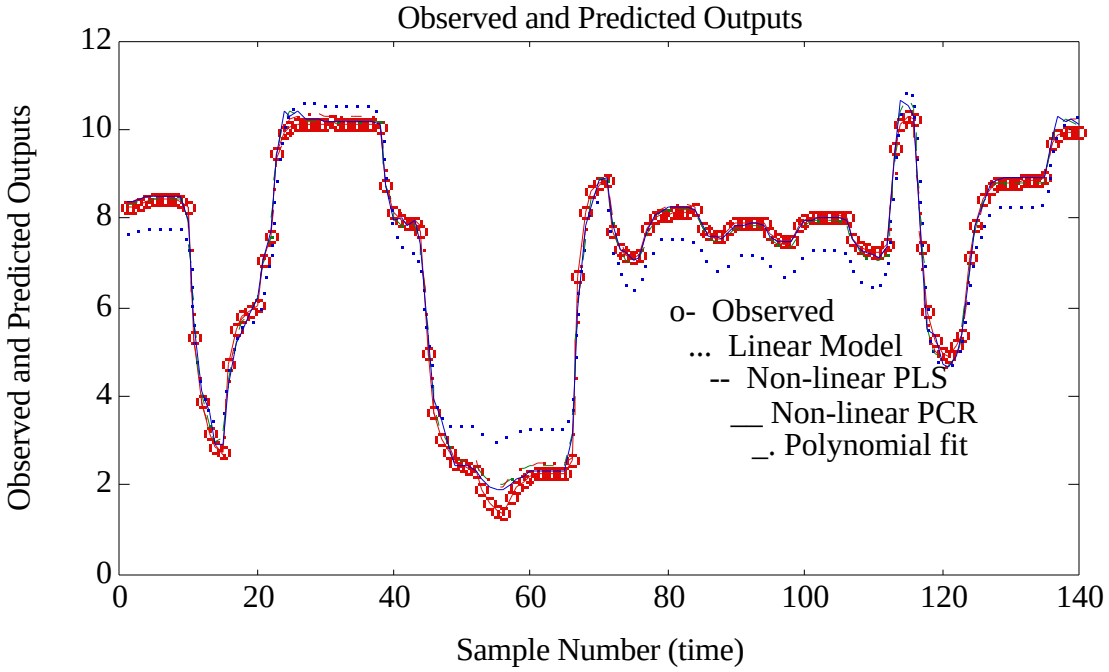

Figure 23. Segment of Calibration Data Set Showing Model Fits.

A segment of the actual and predicted data is shown in Figure 24. While it is apparent that all models were offset in places, the model identified by non-linear PLS almost always lies closest to the actual process output. All of the models are apparently free of any fit artifacts, such as drastic overshoots after step changes. This type of behavior was observed in the non-linear PLS models using 4 or more latent variables. When using this technique, it is prudent to look for fit residuals with odd behavior. This can be a sign of over-fitting.

This data set, in fact, provides a classic example of the perils of over-fitting a data set. Note that while the polynomial model fit the calibration data the best, it did not predict the test data as well as either the PLS or PCR models.

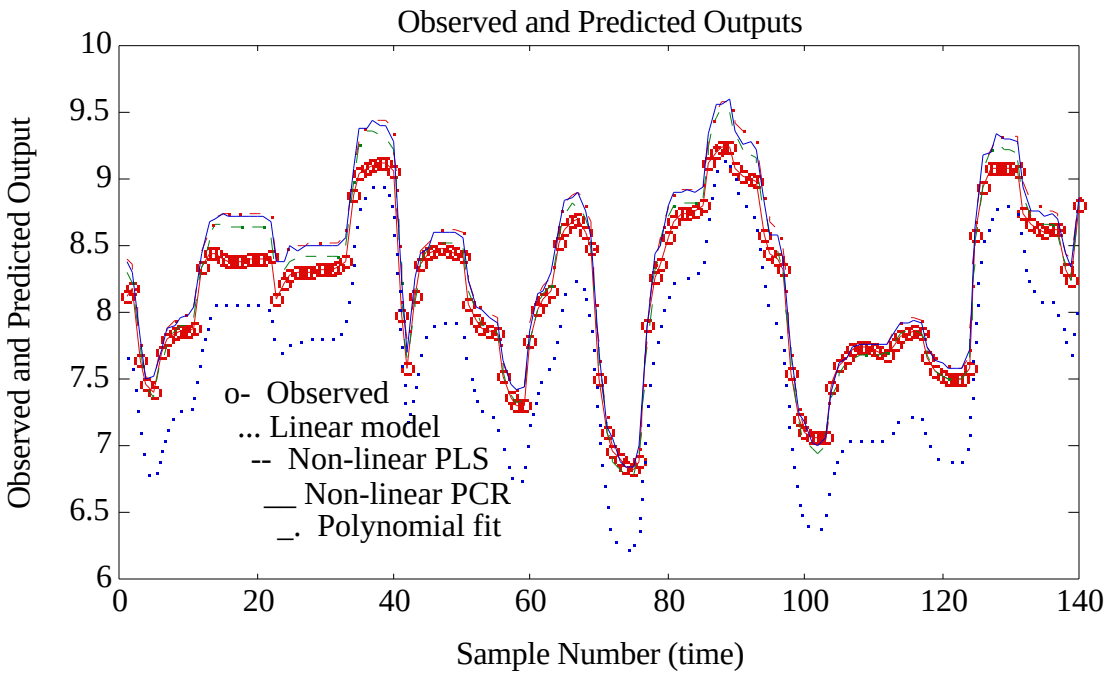

Figure 24. Test Data Set Showing Actual and Predicted Process Output

7.0 Conclusions

Several major conclusions may be reached based on the analysis and simulations performed here. It is apparent the CR is significantly better than MLR for identifying FIR models. While the CR method does produce some artifacts, these are understood and do not impact significantly on the resulting model performance. One possible pitfall concerns incorrect specification of system time delays, and care should be taken to account for this.

CR appears to offer no real advantages for the identification of ARX models in typical identification problems. Properly parameterized models identified by MLR are generally as good or nearly as good as models identified by CR. The exception to this is when conditions are very adverse, e.g. there is very little input excitation in the calibration data and the noise level is very high. Even then, finding the best CR model may be difficult.

There appears to be great potential for the application of PCR and PLS with non-linear inner relationships to the identification of non-linear process models. PCR and PLS help avoid over-fitting the data, a problem of even greater significance for non-linear data. While this text does not provide a complete analysis of the methods, it is clear that further research is warranted

Acknowledgements

The authors would like to thank Dan Haesloop and Brad Holt for the use of the non-linear tank data. We also acknowledge West Valley Nuclear Services Co., Inc. for their continued support.

References

Åstrom, K. J. and Wittenmark, B., *Computer Controlled Systems: Theory and Design,* Prentice-Hall, New Jersey, 1984.

Box, G. E. P. and Jenkins, G. M. *Time Series Analysis: forcasting and control,* Holden Day, Oakland CA, 1976.

Diaz, H. and Desrochers, A. A., "Modeling of Nonlinear Discrete-time Systems from Input-Output Data", *Automatical,* Vol. 24, No. 5, pp. 629-641, 1988.

Frank, I. E., Feikema, J., Constantine, N. and Kowalski, B. R., "Prediction of Product Quality from Spectral Data Using the Partial Least-Squares Method," *J. Chem. Info. and Comp. Sci*, **24**(20), 1984.

Geladi, P. and Kowalski, B. R. "PLS tutorial" *Anal. Chim. Acta.,* **185**(1) (1986).

Geladi, P., "Notes on the History and Nature of Partial Least Squates (PLS) Modelling", *Journal of Chemometrics*, Vol. 2, 231-246 (1988)

Haesloop, D. G. and Holt, B. R., "A Neual Network Structure for System Identification", *Proceeding of the Amercan Control Conference*, San Diego, CA pg. 2460, 1990.

Hoskuldsson, A., "PLS Regression Methods", Journal of Chemometrics, Vol. 2, 211-228 (1988)

Kwakernaak, H. and Sivan, R., *Linear Optimal Control Systems,* Wiler-Interscience, New York 1972.

Ljung, L., *System Identification Toolbox for use with MATLAB,* The MathWorks, Inc., 1988.

Ljung, L., *System Identification: Theory for the User,* Prentice-Hall, Inc., New Jersey, 1987.

Lorber, A. and Kowalski, B. R., "A Note on the Use of the Partial Least-Squares Method for Multivariate Calibration", *Applied Spectroscopy,* Vol. 42, No. 8, 1988.

Lorber, A., Wangen, L. E. and Kowalski, B. R., "A Theoretical Foundation for the PLS Algorithm," *J. Chemometrics* **1**, 19 (1987).

Mejdell T. and Skogestad, S., "Estimate of process outputs from multiple secondary measurements," 1989 American Control Conference Proceedings"

Prett, D. M., Skrovanek, T. A. and Pollard, J. F., "Process Identification - Past, Present, Future", *Shell Process Control Workshop,* ????

Ricker, N. L., "The Use of Biased Least-Squares Estimators for Parameters in Discrete-Time Pulse-Response Models," *Ind. Eng. Chem. Res.*, **27**(2), p.343, 1988.

Sage, A. P. and White, C. C., III, *Optimum Systems Control, Second Edition,* Prentice-Hall, New Jersey, 1977.

Sharaf, M. A., Illman, D. L., and Kowalski, B. R., *Chemometrics*, John Wiley and Sons, New York (1986).

Soderstrom, T. and Stoica, P. G., *Instrumental Variable Methods for System Identification*, Springer-Verlag, New York (1983)

Stone, M. and Brooks, R. J., "Continuum Regression: Cross-validated Sequentially-constructed Prediction embracing Ordinary Least Squares, Partial Least Squares, and Principal COmponents Regression", Unpublished.

Strang, G., *Linear Algebra and Its Applications*, Academic Press New York, 1980.

Wangen, L. E. and Kowalski, B. R., "A Multiblock Partial Least Squates Algorithm for Investigating Complex Chemical Systems," *J. Chemometrics*, Vol 3, 1988.

* The eigenvector decomposition of the input autocorrelation matrix effectively breaks the signal into frequency bands. The effect of this decomposition appears in the frequency response behavior of the resulting PCR FIR models.

* Mis-estimation of time delays degrades FIR models identified by CR to a greater extent than models identified by MLR. Typically, however, the CR models are still better than the MLR models.

* The addition of noise to the process output increases the relative advantage of the CR identifed models over MLR identified models.

* The true process dynamics greatly affect the location of the best models in CR space.

* As process input excitation is increased, the optimum models

* The CR model deemed best depends upon the test chosen. Step tests tend to favor models near the MLR side, while white noise tests favor PCR models.

* The frequency domain artifacts associated with PCR identification tend to be smoothed out in the CR models nearer PLS. Some evidence of these artifacts typically remains, however.

* When identifying ARX models using CR, the blha lbha

* PCR and PLS with non-linear inner relationships offer an advantage of polynomial regression for non-linear model identification.