University of Pretoria- etd Cater, C R (1997)

Chapter 5

Non-linear time domain system identification: NARX

Successful implementation of linear ARX - state-space algorithms presented a further challenge: Non-linear system identification for use in response reconstruction. This could greatly improve simulation results, and eradicate the need for iterative linearization of the non-linear system. Investigation into inclusion of non-linear modelling in the QanTiM simulation system started with a survey of applicable modelling techniques. This survey is presented in Appendix A. Non-linear model requirements facilitating inclusion into response reconstruction are similar to that presented by Raath [51] for the linear case:

- Discrete.
- Multivariable.
- Time invariant.
- Black-box.
- Allow stable inversion.
- Allow use of model for simulation purposes.
- Accommodate simultaneous multiple-actuator identification.

Furthermore the model would be required to:

- Model highly non-linear systems (cubic polynomials).
- Include non-linear capabilities with minimal extra user input.
- Allow inclusion into existing QanTiM software.

This study investigates the NARX formulation, a polynomial non-linear extension of the ARX model used by QanTiM. Raath and Verwey [54], showed that the NARX model formulation satisfied all the above requirements, especially so for ease of use and possible compatibility with existing linear software. Detailed descriptions of the NARX model formulation, as well as the application thereof are given. Non-linear simulation and its limitations are discussed as well as modifications to the NARX structure for improved performance and QanTiM compatibility.

University of Pretoria- etd Cater, C R (1997)

5.1. The NARX model

NARX (Non-linear Auto Regressive with eXogenous input) is a parametric difference equation that forms a convenient linear-in-the-parameters set of equations capable of describing systems with severe non-linearity. It is a special case of the general NARMAX [18] model in which only the system dynamics are taken into account. The NARMAX model is reduced to NARX by removal of the noise model and moving average terms, presented by Billings *et al.* [5] [9] and Peyton Jones [46] [47] .

$$y(t) = F^{L}\left[y(k-1),\ldots,y(k-na),u(t-nk),\ldots,u(t-nk-nb)\right] \qquad (5\text{-}1)$$

With $F^L[\bullet]$ some non-linear polynomial function.

5.1.1. MISO-NARX formulation

Consider the MIMO non-linear dynamic system presented in Figure 5.1. As with the linear ARX model, the MIMO-NARX consists of a combination of MISO models. Combining the NARX MISO models could however not be done as elegantly as for the ARX models. A system is identified for each output channel and the MISO NARX models are then combined in a one-step-ahead simulation routine as described in Section 5.4. The most general form of the MISO-NARX model [56] as shown in equation (5-2) describes each output $y_k(t)$ as a function of inputs $u_i(t)$ and outputs $y_j(t)$. With $F^L[\bullet]$ some non-linear function.

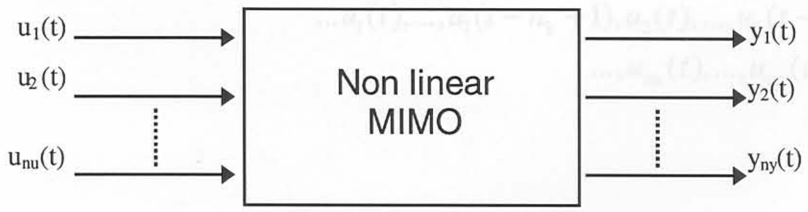


Figure 5.1. Non-linear MIMO system

University of Pretoria- etd Cater, C R (1997)

$$
\begin{aligned}
y_{\mathrm{k}}(t) = F^{\mathrm{L_k}}[&y_1(t-1),\ldots,y_1(t-na_1),y_2(t-1),\ldots,y_2(t-na_2),\ldots y_{\mathrm{ny}}(t-1),\ldots,y_{\mathrm{ny}}(t-na_{\mathrm{ny}}),\ldots \\
&\ldots u_1(t-nk_1),\ldots,u_1(t-nk_1-nb_1),u_2(t-nk_2),\ldots,u_2(t-nk_2-nb_2),\ldots \\
&\ldots,u_{\mathrm{nu}}(t-nk_{nu}),\ldots,u_{\mathrm{nu}}(t-nk_{nu}-nb_{\mathrm{nu}})]
\end{aligned}
\quad (5\text{-}2)
$$

where for channel k:
L_k = The degree of non-linearity within F
na_k = dynamic model order for output $y(t)$
nb_k = dynamic model order for input $u(t)$
nk_k = time delay

The full order approach as proposed by Raath [53] is, as with the linear ARX model, used to greatly simplify modelling. Using this approach the model format is predicted with only two parameters per output channel namely: n, the dynamic model order, and L, the degree of non-linearity. Thus for each channel $y_k(t)$ the following applies:

$$
:\quad \left.\begin{aligned}
na_i &= n_k \\
nb_j &= n_k + 1 \\
nk_j &= 0 \\
L_k &= L_k
\end{aligned}\right\}
\begin{aligned}
i &= 1,2,\cdots,ny \\
j &= 1,2,\cdots,nu
\end{aligned}
\quad (5\text{-}3)
$$

By substituting the full order model parameters (5-3) into equation (5-2) the NARX formulation is reduced as shown in equation (5-4):

$$
\begin{aligned}
y_{\mathrm{k}}(t) = F^{\mathrm{L_k}}[&y_1(t-1),\ldots,y_1(t-n_{\mathrm{k}}),y_2(t-1),\ldots,y_2(t-n_{\mathrm{k}}),\ldots y_{\mathrm{ny}}(t-1),\ldots,y_{\mathrm{ny}}(t-n_{\mathrm{k}}),\ldots \\
&\ldots u_1(t),\ldots,u_1(t-n_{\mathrm{k}}+1),u_2(t),\ldots,u_2(t-n_{\mathrm{k}}+1),\ldots \\
&\ldots,u_{\mathrm{nu}}(t),\ldots,u_{\mathrm{nu}}(t-n_{\mathrm{k}}+1)]
\end{aligned}
\quad (5\text{-}4)
$$

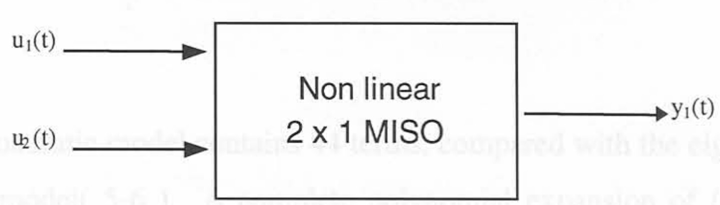
University of Pretoria- etd Cater, C R (1997)

For a value of $L_k = 1$, the NARX model is identical to the linear ARX so that $F^{L_k}[\mathrm{ARX}] \Rightarrow \mathrm{NARX}$. NARX is thus a non-linear extension of the linear ARX as shown in the next Section.

5.1.2. Non-linearity in the NARX model

The NARX-model of (5-4) is described in terms of some non-linear function $F^L[\bullet]$. This function finds all unique combinations of L_k-degree multiples of the linear model terms. The non-linear structure is introduced in Example 5.1 for a specific MISO system to precede the more general formulation [12].

Example 5.1: NARX formulation for a simple MISO system



The second order linear ARX difference equation for a two-input single output system is:

$$\begin{aligned} y_1(t) = {} & y_1(t-1)\cdot a_{1_1} + y_1(t-2)\cdot a_{1_2} + \\ & u_1(t)\cdot b_{1_1} + u_1(t-1)\cdot b_{1_2} + u_1(t-2)\cdot b_{1_3} + \\ & u_2(t)\cdot b_{1_4} + u_2(t-1)\cdot b_{1_5} + u_2(t-2)\cdot b_{1_6} \end{aligned} \tag{5-5}$$

Using vector notation, we may further reduce it to:

$$\begin{aligned} & X_{L1_1}(t) = \begin{bmatrix} y_1(t-1) & y_1(t-2) & u_1(t) & \cdots & u_2(t-2) \end{bmatrix} \text{ and} \\ & \Theta_1 = \begin{bmatrix} a_{1_1} & a_{1_2} & b_{1_1} & \cdots & b_{1_6} \end{bmatrix}^T \\ & \quad = \begin{bmatrix} \theta_{1_1} & \theta_{1_2} & \theta_{1_3} & \cdots & \theta_{1_8} \end{bmatrix}^T \\ & \Rightarrow \\ & y_1(t) = \sum_{i=1}^{8} X_{L1_1}(t,i)\cdot \theta_{1_i} \end{aligned} \tag{5-6}$$

University of Pretoria- etd Cater, C R (1997)

Now extend this model to the non-linear second order quadratic NARX-formulation by adding all quadratic combinations of the linear regression vector $X_{L1_1}(t)$.

$$
\begin{aligned}
y_1(t) &= \sum_{i=1}^{8} X_{L1_1}(t,i)\cdot\theta_{1_i} + \overbrace{\sum_{R_1=1}^{8}\sum_{R_2=1}^{R_1} X_{L1_1}(t,R_1)\cdot X_{L1_1}(t,R_2)\cdot\theta_{1_j}}^{j=(9,10,\cdots,44)} \\
&= \sum_{i=1}^{44}\mathbf{X}_1(t,i)\cdot\theta_{1_i} \\
&= \mathbf{X}_1(t)\bullet\Theta_1
\end{aligned}
\qquad (5\text{-}7)
$$

where

$$
\mathbf{X}_1(t) = \left[\underbrace{X_{L1_1}(t)}_{\text{Linear}}\quad\underbrace{X_{L2_1}(t)}_{\text{Quadratic}}\right],\ \text{and}
$$

$$
X_{L2_1}(t) = \left[y_1(t-1)^2\quad y_1(t-1)\times y_1(t-2)\quad\cdots\quad u_2(t-2)^2\right]
$$

This quadratic model contains 44 terms, compared with the eight of the linear model(5-6). A complete polynomial expansion of (5-7) is given in Appendix E.

The general form of the non-linear model is shown below for cubic ($L=3$) non-linearity [14][56]:

$$
\begin{aligned}
y_k(t) &= \underbrace{\sum_{i=1}^{R} X_{L1_k}(t,i)\cdot\theta_{k_i}}_{\substack{\text{Linear part}\\ L=1\\ R=ny\cdot n+nu\cdot(n+1)}} + \underbrace{\sum_{R_1=1}^{R}\sum_{R_2=1}^{R_1} X_{L1_k}(t,R_1)\cdot X_{L1_k}(t,R_2)\cdot\theta_{k_j}}_{\substack{\text{Quadratic part}\\ L=2\\ j=(M_{L1}+1,M_{L1}+2,\cdots,M_{L2})}} + \underbrace{\sum_{R_1=1}^{R}\sum_{R_2=1}^{R_1}\sum_{R_3=1}^{R_2} X_{L1_k}(t,R_1)\cdot X_{L1_k}(t,R_2)\cdot X_{L1_k}(t,R_3)\cdot\theta_{k_j}}_{\substack{\text{Cubic part}\\ L=3\\ j=(M_{L2}+1,M_{L2}+2,\cdots,M_{L3})}} + \ldots \\
&= \sum_{i=1}^{M}\mathbf{X}_k(t,i)\cdot\theta_{k_i}
\end{aligned}
\qquad (5\text{-}8)
$$

University of Pretoria- etd Cater, C R (1997)

The variable M represents the maximum number of coefficients in the NARX-equation. In Equation(5-8) the terms M_{L2} and M_{L3} respectively represent the number of terms for the quadratic and cubic parts of the NARX-model. The non-linear model is still linear-in-the-parameters even though it can describe systems with severe non-linearity.

5.1.3. Coefficients in the NARX model

For the SISO system the maximum number of coefficients is defined by Billings & Voon [7] as M:

$$M = \sum_{i=1}^{L} n_i$$
$$n_i = \frac{n_{i-1} \cdot (na + nb + i - 1)}{i}\ , \quad \text{where } n_0 = 1 \qquad (5\text{-}9)$$

The author [14] showed that this equation does not hold for non-linear MIMO systems. Such a system's maximum number of coefficients for a specific output channel can be expressed by the following equations:

$$M_k = R + \sum_{R_1=1}^{R}\left[R_1 + \sum_{R_2=1}^{R_1}\left[R_2 + \sum_{R_3=1}^{R_2}\left[R_3 + \ldots + \sum_{R_L=1}^{R_L-1} R_L\right]\right]\right]$$
$$R = ny \cdot n_k + nu \cdot (n_k + 1) \qquad (5\text{-}10)$$

The number of coefficients explodes with increasing degrees of non-linearity, prompting investigation into reduced parameter modelling [8] [38][39]. Reduction of the number of model parameters can be done by structure selection prior to modelling, as discussed in Section 5.5, or by making use of a reduced parameter estimation technique, as discussed in Section 5.3

University of Pretoria- etd Cater, C R (1997)

5.2. NARX regression

Transformation of input-output data into the NARX formulation of Equation (5-4) and (5-8) is done in such a manner as to maintain generality at all times. Based on the work by Leontaritis [38] for NARMAX systems, Equation (5-8) is written in matrix formulation, with the non-linear extensions appended to the linear ARX regression matrix so as to maintain a linear-in-the-parameters set of equations.

$$\mathbf{Y}_k = \mathbf{X}_k \bullet \Theta_k \qquad (5\text{-}11)$$

where: $\mathbf{Y}_k^{\mathrm{T}} = [y_k(1)\ y_k(2) \ldots y_k(t) \ldots y_k(N)]$

$\Theta_k^{\mathrm{T}} = [\theta_{k_1}\ \theta_{k_2} \ldots \theta_{k_M}]$

$$\mathbf{X}_k = [X_{L1} \quad X_{L2} \quad X_{L3} \quad \cdots]$$

$$= \begin{bmatrix} X_k(1,1) & X_k(1,2) & \cdots & X_k(1,M) \\ X_k(2,1) & X_k(2,2) & \cdots & X_k(1,M) \\ \vdots & \vdots & & \vdots \\ X_k(t,1) & X_k(t,2) & \cdots & X_k(t,M) \\ \vdots & \vdots & & \vdots \\ X_k(N,1) & X_k(N,2) & \cdots & X_k(N,M) \end{bmatrix}$$

The procedure of finding a MISO NARX regression matrix is to be repeated for each output channel of the MIMO system. A number of such regression methods were developed by the author, a summary of these algorithms is presented in Appendix B. The next step in the system identification process is to find the NARX coefficient matrix Θ_k. Methods in which these coefficients can be estimated are the subject of the next Section.

5.3. NARX parameter estimation

Parameter estimation is the process of finding the unknown coefficients θ_i for the NARX equation and thus identifying the dynamic characteristics of the system. A number of well-studied methods are available for parameter estimation, most of which are well suited to the NARX model.

The majority of the parameter estimation techniques can be placed into two categories: Prediction error methods and Correlation methods. [7][8][18] [36] The first being concerned with minimising some error function for the identification model. Correlation techniques are however concerned with finding the solution for some function. Various authors [1][4][7][51] showed prediction error methods to be best suited to practical system identification. In this thesis the focus will thus remain on the prediction error methods for finding the parameter vector of the NARX-model.

A loss function $J[\bullet]$ can be defined with the following form:

$$J = \frac{1}{N} \cdot \sum_{\tau=1}^{N} f\left(\varepsilon(\tau)\right) \qquad (5\text{-}12)$$

With $f(\bullet)$ some positive function of the identification error $\varepsilon(t) = y(t) - X(t)\cdot\Theta$.

The basis of the prediction error methods now lies with finding a set of parameters Θ associated with minimising this loss function. Various techniques are based on this principle, including Least Squares [52] , Extended Least Squares [8], and Maximum Likelihood [7] parameter estimation. The Least Squares technique is inherently suited to the NARX-model and computationally simple to apply. According to Strejc [58] "It may be stressed that in the field of parameter estimation the Least Squares technique has reached a significant level of popularity and perfection." It is thus the only technique that will be covered in this study. Various solutions to the Least Squares problem are however given.

University of Pretoria- etd Cater, C R (1997)

5.3.1. The Least Squares problem

Gauss defined: "the most probable value of the unknown quantities will be that one for which the sum of the squares of the differences between the actually observed and computed values multiplied by numbers that measure the degree of precision is a minimum". In a Least Squares form the loss function of Equation (5-12) is thus described with $f(\bullet)$ a quadratic function of $\varepsilon(t)$ which can be minimised with respect to Θ: Equation [29]

$$J = \frac{1}{N} \cdot \sum_{\tau=1}^{N} \left(\varepsilon(\tau)\right)^2 \qquad (5\text{-}13)$$

This defines the Least Squares equation (5-13) for a specific output channel.

$$y_k(t) = X_k(t) \cdot \Theta_k + \varepsilon_k(t) \ :$$

$$\begin{aligned} J_k &= \frac{1}{N} \sum_{\tau=1}^{N} \left(\varepsilon_k(\tau)\right)^2 \\ &= \frac{\varepsilon_k^{\mathrm{T}} \cdot \varepsilon_k}{2} \qquad \{\text{if: } \mathbf{Y}_k = \mathbf{X}_k \cdot \Theta_k + \varepsilon_k\} \end{aligned} \qquad (5\text{-}14)$$

Substitute into $\varepsilon_k = \mathbf{Y}_k - \mathbf{X}_k \Theta_k$ into (5-14) and expand the expressions to obtain the following equation for the loss function.

$$\begin{aligned} J_k &= \frac{\varepsilon_k^{\mathrm{T}} \cdot \varepsilon_k}{2} \\ &= \frac{(\mathbf{Y}_k - \mathbf{X}_k \cdot \Theta_k)^{\mathrm{T}} \cdot (\mathbf{Y}_k - \mathbf{X}_k \cdot \Theta_k)}{2} \\ &= \frac{\mathbf{Y}_k^{\mathrm{T}} \cdot \mathbf{Y}_k - 2 \cdot \left(\mathbf{Y}_k^{\mathrm{T}} \cdot \mathbf{X}_k \cdot \Theta_k\right) + \left(\mathbf{X}_k \cdot \Theta_k\right)^{\mathrm{T}} \mathbf{X}_k \cdot \Theta_k}{2} \end{aligned} \qquad (5\text{-}15)$$

University of Pretoria- etd Cater, C R (1997)

The Least Squares estimation of Θ_k is found by minimising the loss function J_k with respect to Θ_k

$$\frac{d}{d\,\Theta_k}(J_k) = \frac{\mathbf{X}_k{}^{\mathrm{T}} \cdot \mathbf{X}_k \cdot \Theta_k - \mathbf{X}_k{}^{\mathrm{T}} \cdot \mathbf{Y}_k}{2} \qquad (5\text{-}16)$$

and $J_k(\Theta_k)$ is thus a minimum for :

$$\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X} \cdot \Theta - \mathbf{X}^{\mathrm{T}} \cdot \mathbf{Y} = 0$$

$$\Downarrow$$

$$\boxed{\Theta = \left[\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right]^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{Y}} \qquad (5\text{-}17)$$

This is the familiar Least Squares equation [8][52][58] which is valid only if $[\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}]^{-1}$ exists which proved not trivial for various reasons, including:

Table 5.1: Obstacles in finding the inverse $[X^T \cdot X]$-1

Size	The large number of terms in the NARX regression matrix presented a problem for existing inversion processes.
Ill-conditioning	More often than not, the NARX regression matrix showed signs of ill conditioning
Linear dependency	Incorrect sampling rates captured data with similar samples, resulting in a regression matrix with linearly dependant rows.

A survey of finding the solutions to the Least Squares equation (5-17) is presented in Appendix C and D for full, as well as condensed parameter sets. This survey of parameter estimation techniques showed that methods based on orthogonal decomposition to be the most effective for finding the inverse $[\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}]^{-1}$. The concepts of a condensed model structure and the associated parameter estimation techniques are discussed in Section 5.3.2.

University of Pretoria- etd Cater, C R (1997)

5.3.1.1. Orthogonal decomposition

Parameter estimation techniques based on orthogonal decomposition [36], [58] of the NARMAX and, similarly, the NARX regression matrix proved the most effective. These methods all have in common that a matrix **X** can be transformed into an orthogonal matrix **Q** and an upper triangular matrix **R** so that.

$$\mathbf{X} = \mathbf{Q} \cdot \mathbf{R} \tag{5-18}$$

Transform the Least Squares equation, in order to find the parameter vector Θ.

$$\Theta = \left[\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right]^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{Y}$$
$$\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X} \cdot \Theta = \mathbf{X}^{\mathrm{T}} \cdot \mathbf{Y} \tag{5-19}$$

Substitute $\mathbf{X} = \mathbf{Q} \cdot \mathbf{R}$

$$[\mathbf{Q} \cdot \mathbf{R}]^{\mathrm{T}} \cdot \mathbf{Q} \cdot \mathbf{R} \cdot \Theta = [\mathbf{Q} \cdot \mathbf{R}]^{\mathrm{T}} \cdot \mathbf{Y} \tag{5-20}$$
$$\mathbf{Q}^{\mathrm{T}} \cdot \mathbf{R}^{\mathrm{T}} \cdot \mathbf{Q} \cdot \mathbf{R} \cdot \Theta = \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{R}^{\mathrm{T}} \cdot \mathbf{Y}$$

but $\mathbf{Q}^{\mathrm{T}} \cdot \mathbf{Q} = \mathbf{I}$

$$\mathbf{R}^{\mathrm{T}} \cdot \mathbf{R} \cdot \Theta = \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{R}^{\mathrm{T}} \cdot \mathbf{Y} \tag{5-21}$$
$$\mathbf{R} \cdot \Theta = \left[\mathbf{R}^{\mathrm{T}}\right]^{-1} \cdot \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{R}^{\mathrm{T}} \cdot \mathbf{Y}$$

thus

$$\Theta = \left[\mathbf{R}^{\mathrm{T}}\right]^{-1} \cdot \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{Y} \tag{5-22}$$

University of Pretoria- etd Cater, C R (1997)

Methods for solving the Least Squares equation using orthogonal decomposition include:

- Classical Gram Schmidt methods [11]
- Modified Gram Schmidt methods [10], [32] and
- Householder transformations [32]

These methods are discussed in Appendix C and D

5.3.2. Full vs. reduced parameter modelling

Korenberg [36] indicated that "provided the significant terms in the model can be detected, models with fewer than ten terms are usually sufficient to capture the dynamics of highly non-linear processes." This presents the problem of structure detection for the NARX equation. Appendix D details various methods for finding a reduced set of parameters i.e. discarding terms in the NARX equation which do not contribute to the dynamic behaviour of the system. These methods all require the full set of NARX coefficients to be available for evaluation, thus an initial full set regression and parameter estimation is required prior to structure detection. Further the structure detection processes proved computationally cumbersome. More importantly simulation of the NARX models proved insensitive to the number of terms involved. The author found that the amount of effort concerned with model reduction does not warrant the implementation thereof. Full parameter set modelling proved more practical to investigate the implementation of NARX in structural response reconstruction.

University of Pretoria- etd Cater, C R (1997)

5.4. Simulation of NARX systems

In the linear case a state space representation of the ARX model was calculated prior to simulation. This MIMO state space system proved convenient, especially so for linear application within Matlab [53]. A non-linear equivalent of this model format conversion, that is NARX to state space, can be done in the same way for a specific system. Formulation of a general NARX to state space procedure is however not trivial and the simulation of non-linear differential state space equations is computationally taxing. The non-linear state space simulation algorithms created and implemented by the author proved too complex and mathematically expensive to warrant further investigation, application or discussion. The purpose of this study does not warrant an extensive investigation into general NARX to state space conversion algorithms. A simulation algorithm that made use of a step-ahead-predictor nested in a sample point loop proved an effective method for simulation non-linear systems. This non-linear simulation algorithm makes full use of the linear-in-the-parameters structure of the NARX model, in principle done according to Algorithm 5.1. It is a direct implementation of Equation (5-4).

Algorithm 5.1: NARX simulation

INPUT:	Dynamic system input data:	$u_1(t),\ u_2(t) \dots,\ u_{nu}(t)$
	NARX model coefficients for each o/p channel	Θ_k
	Dynamic model order for each o/p channel:	n_k
	Degree of non-linearity for each o/p channel:	L_k
	Number of sample points to use in regression	N
	Number if input and output channels	$nu,\ ny$
OUTPUT:	Dynamic system output data:	$y_1(t),\ y_2(t) \dots,\ y_{ny}(t)$

FOR $t = 1, 2, \dots, N$	*Sample point loop*
FOR $k = 1, 2, \dots, ny$	*Channel loop*
$X_k(t) = f(u,y,na,L)$	*Calculate non-linear regression vector for channel k at sample point t*
$y_k(t) = X_k(t) \cdot \Theta_k$	*Calculate the value for output channel* y_k *at sample point t. Note that* $\cdot$ *denotes a vector product operation so that* $y_k(t) = X_k(t) \cdot \Theta_k = \sum_{i=1}^{M} \mathbf{X}_k(t,i) \cdot \theta_{k_i}$ *with M the number of coefficients in the NARX equation.*

University of Pretoria- etd Cater, C R (1997)

This basic simulation loop (Algorithm 5.1) however proved extremely slow due to the regression operation at each sample point. This regression includes the process of finding the appropriate combinations of linear model terms to form the non-linear extensions to the ARX format. Another limitation is the use of a sample point main loop, which implies the simulation may become slow for large data sets. Due to the one-step-ahead nature of the algorithm a more elegant method than this sample point loop could not be found. The problem of regression and NARX structure formulation at each sample point is solved by careful use of Matlab's matrix capabilities. This was done by defining a condensed NARX model structure which, once identified, could easily be passed to the simulation algorithm and thus eradicate the need to recalculate the NARX structure.

5.4.1. Condensed NARX model structure

The NARX coefficient vector Θ_k completely characterises the k^{th} channel of the system. Thus if all *ny* coefficient vectors are known the system can be considered thoroughly identified. This format is however not convenient for simulation purposes since for each channel a separate set of coefficients must be stored, and passed to various functions. Furthermore, for the purpose of simulation, the NARX model structure, i.e. the non-linear combinations of linear ARX terms to be multiplied must be recalculated.

The author defined a model structure that contains the coefficient vectors for all system channels, as well as model information such as model order and degree of non-linearity for each channel. This condensed NARX structure further contained the combinations of all non-linear terms within the model. The elements contained in the condensed model structure for each channel are described in Table 5.2.

University of Pretoria- etd Cater, C R (1997)

Table 5.2. Model descriptors for output channel k

NARX coefficient vector	$\Theta_k^T = [\theta_{k_1}\ \theta_{k_2}\ \ldots\ \theta_{k_M}]$		$(1 \times M_k)$
Model parameters	*Model order:*	n_k	(1×1)
	Degree of non-linearity:	L_k	(1×1)
	Number of model terms:	M_k	(1×1)
	Number of quadratic terms:	$ML2_k$	(1×1)
	Number of cubic terms:	$ML3_k$	(1×1)
Non-linear combination vectors	*Quadratic combinations:*	$\Pi2_k$	$(2 \times ML2_k)$
	Cubic combinations:	$\Pi3_k$	$(3 \times ML3_k)$

The model parameters of Table 5.2 are written into a model parameter matrix Mpar_k for each channel k. These MISO models are then appended into a single model parameter matrix for all system channels, as shown in Equation (5-23) The condensed NARX model structure was implemented to accommodate, at most, cubic ($L=3$) non-linear models.

$$\mathbf{Mpar}_k = \begin{bmatrix} \begin{bmatrix} \cdots & \Theta_{k_i}^T & \cdots \end{bmatrix} \\ \begin{bmatrix} n_k & L_k & M_k & M_{L2_k} & M_{L3_k} \end{bmatrix} \\ \left.\begin{bmatrix} 1 & 1 & \cdots & 2 & \cdots & M_k \\ 1 & 2 & \cdots & 2 & \cdots & M_k \end{bmatrix}\right\} \Pi 2_k \\ \left.\begin{bmatrix} 1 & 1 & \cdots & 1 & \cdots & 2 & \cdots & M_k \\ 1 & 1 & \cdots & 2 & \cdots & 2 & \cdots & M_k \\ 1 & 2 & \cdots & 2 & \cdots & 2 & \cdots & M_k \end{bmatrix}\right\} \Pi 3_k \end{bmatrix} (7 \times M_k)$$

$$\mathbf{Mpar} = \begin{bmatrix} \mathbf{Mpar}_1 \\ \mathbf{Mpar}_2 \\ \vdots \\ \mathbf{Mpar}_{ny} \end{bmatrix}$$

(5-23)

University of Pretoria- etd Cater, C R (1997)

The condensed NARX model structure has the capability to model so called non-square systems. That is a dynamic system of which the number of inputs differs from the number of outputs. The author developed simulation algorithms using this condensed NARX model structure that are sufficiently fast to allow implementation in general non-linear system identification and response reconstruction. The computation time of the algorithm is directly proportional to the length of the data. A linear regression operation must still be performed at each sample point, but the non-linear combinations thereof need not be done. A revised simulation procedure is presented in Algorithm 5.2

Algorithm 5.2: NARX condensed model structure simulation

INPUT:	Dynamic system input data:	$u_1(t),\ u_2(t) \ldots,\ u_{nu}(t)$
	NARX condensed model parameter matrix	$Mpar$
OUTPUT:	Dynamic system output data:	$y_1(t),\ y_2(t) \ldots,\ y_{ny}(t)$

FOR k = 1, 2, ..., ny	*Channel loop*
$Mpar_k = f(Mpar)$	*Extract model parameter matrix for channel k*
$\Theta_k = f(Mpar_k)$	*Extract NARX coefficient vector for channel k*
$na_k = f(Mpar_k)$	*Extract model parameters for channel k*
$L_k = f(Mpar_k)$	
$M_k = f(Mpar_k)$	
$ML2_k = f(Mpar_k)$	
$ML3_k = f(Mpar_k)$	
$\Pi2_k = f(Mpar_k)$	*Extract non-linear combination matrices for channel k*
$\Pi3_k = f(Mpar_k)$	
FOR t = 1, 2, ..., N	*Sample point loop*
FOR k = 1, 2, ..., ny	*Channel loop*
$XL1_k(t) = f(u,y,na)$	*Calculate linear regression vector for channel k at sample point t*
$XL2_k(t) = f(XL1_k(t), \Pi2_k)$	*Calculate the non-linear terms of the regression matrix using the linear ARX terms and the non-linear combination vectors.*
$XL3_k(t) = f(XL1_k(t), \Pi3_k)$	
$X_k(t) = [XL1_k(t)\ XL2_k(t)\ XL3_k(t)]$	
$y_k(t) = X_k(t) \cdot \Theta_k$	*Calculate the value for output channel y_k at sample point t.*

University of Pretoria- etd Cater, C R (1997)

5.5. Modified NARX systems

The number of terms in the NARX formulation tends to explode for systems with high degrees of non-linearity and large numbers of input and output channels. Furthermore, numerical techniques tend to become unstable for models with large numbers of non-linear terms. It would thus be ideal to limit the number of terms within the NARX model. The concept of reduced parameter modelling was introduced in Section 5.3.2, which concluded sub-set selection techniques to be computationally too taxing to warrant implementation. An alternative route is to select a non-linear model structure prior to identification. A number of special cases of the NARX model were defined and are presented in Sections 5.5.1. through 5.5.4.

5.5.1. Purely Quadratic NARX

An approach similar to the bi-linear model (Appendix A.3.2) was used to limit the number of model terms, yet maintain a high degree of non-linear modelling capability. The purely quadratic NARX, as shown in Equation (5-24) for a SISO system, includes all quadratic combinations of the linear ARX model, without any non-linear cross-coupling terms.

$$y_k(t) = a_0 + \sum_{i=1}^{n_k} a_i \cdot y(t-i) + \sum_{i=1}^{n_k+1} b_i \cdot u(t-i) + \sum_{i=1}^{2n_k+1} c_i \cdot \left(u(t-i)\right)^2 \sum_{i=1}^{2n_k+1} d_i \cdot \left(y(t-i)\right)^2 \qquad (5\text{-}24)$$

Parameter estimation for the purely quadratic NARX is done exactly the same as for the full parameter set NARX model.

University of Pretoria- etd Cater, C R (1997)

5.5.2. Quasi-Static NARX

Block orientated models (Appendix A.2) include the concept of modelling system dynamics linearly and only the static system behaviour with a non-linear model. An implementation of such a modelling approach was presented by Billings and Fakhouri [4]. Their approach utilised the first two kernels in the Volterra series expansion, which proved too complex for practical application. The author implemented a similar approach for NARX models, where only the static part of the system behaviour is modelled non-linearly. A simple method to implement this quasi-static NARX, is to include the entire linear ARX model but append only the static combinations (dynamic order = 0) in the non-linear extension. Consider again the two input single output system presented in Example 5.1 Equation (5-25) shows only the static relation between the two input channels and the output squared.

$$\begin{aligned} y_1(t) = {} & y_1(t-1)\cdot a_{1_1} + y_1(t-2)\cdot a_{1_2} + \\ & u_1(t)\cdot b_{1_1} + u_1(t-1)\cdot b_{1_2} + u_1(t-2)\cdot b_{1_3} + \\ & u_2(t)\cdot b_{1_4} + u_2(t-1)\cdot b_{1_5} + u_2(t-2)\cdot b_{1_6} + \\ & (u_1(t))^2\cdot c_{1_1} + (u_2(t))^2\cdot c_{1_2} \end{aligned} \qquad (\text{5-25})$$

A more direct approach is to apply the split-spectra modelling concept as presented in Section 3.2 to non-linear systems.

University of Pretoria- etd Cater, C R (1997)

5.5.3. Split spectra linear-non-linear modelling

The linear-non-linear split spectra modelling procedure presented in Figure 5.2 is similar to the linear system split spectra approach, only here the low frequency part of the data is modelled with non-linear algorithms, and the high frequency dynamics with the conventional linear ARX formulation.

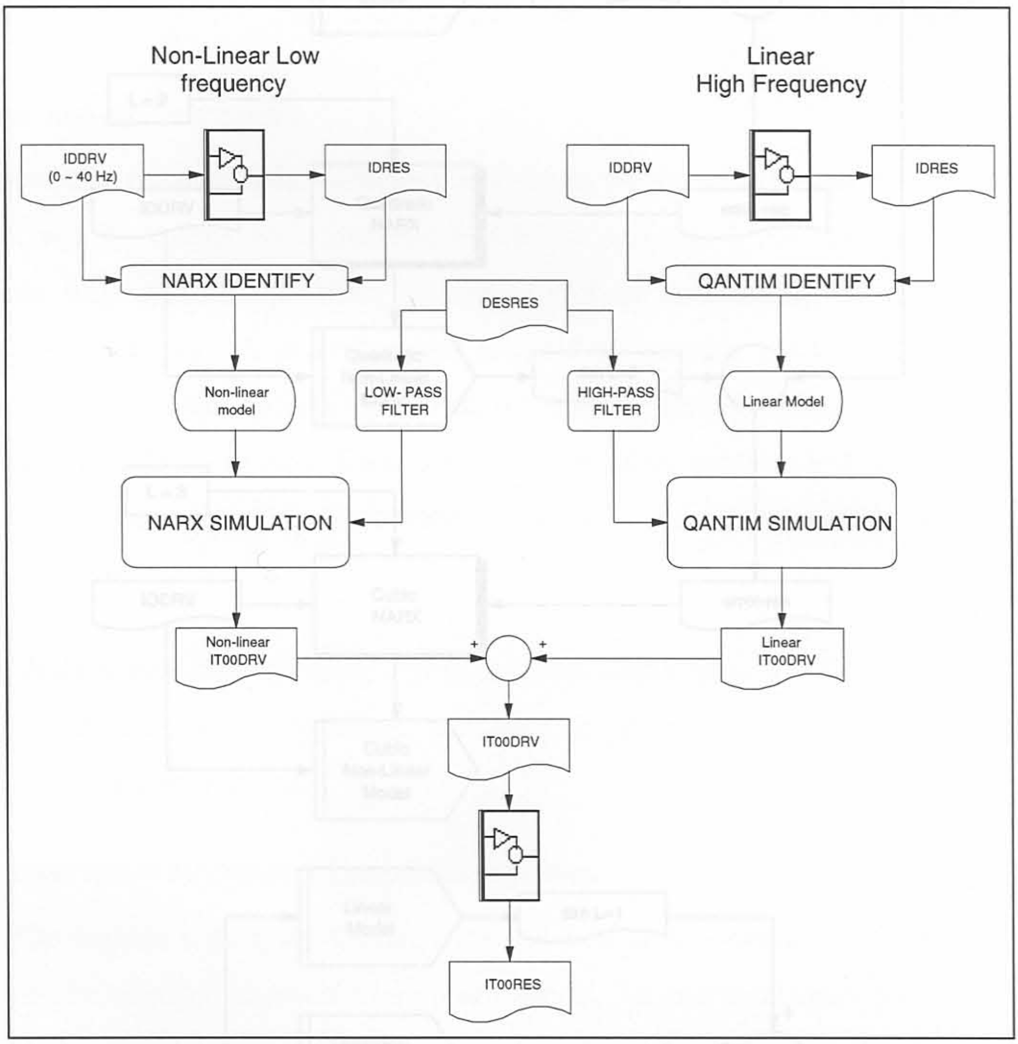


Figure 5.2: Linear-non-linear split spectra modelling

The linear-non-linear split spectra modelling technique proved ideal to model the low frequency high amplitudes associated with non-linear response, as well as the high frequency linear dynamics of a system.

University of Pretoria- etd Cater, C R (1997)

5.5.4. Non-linear error signal modelling

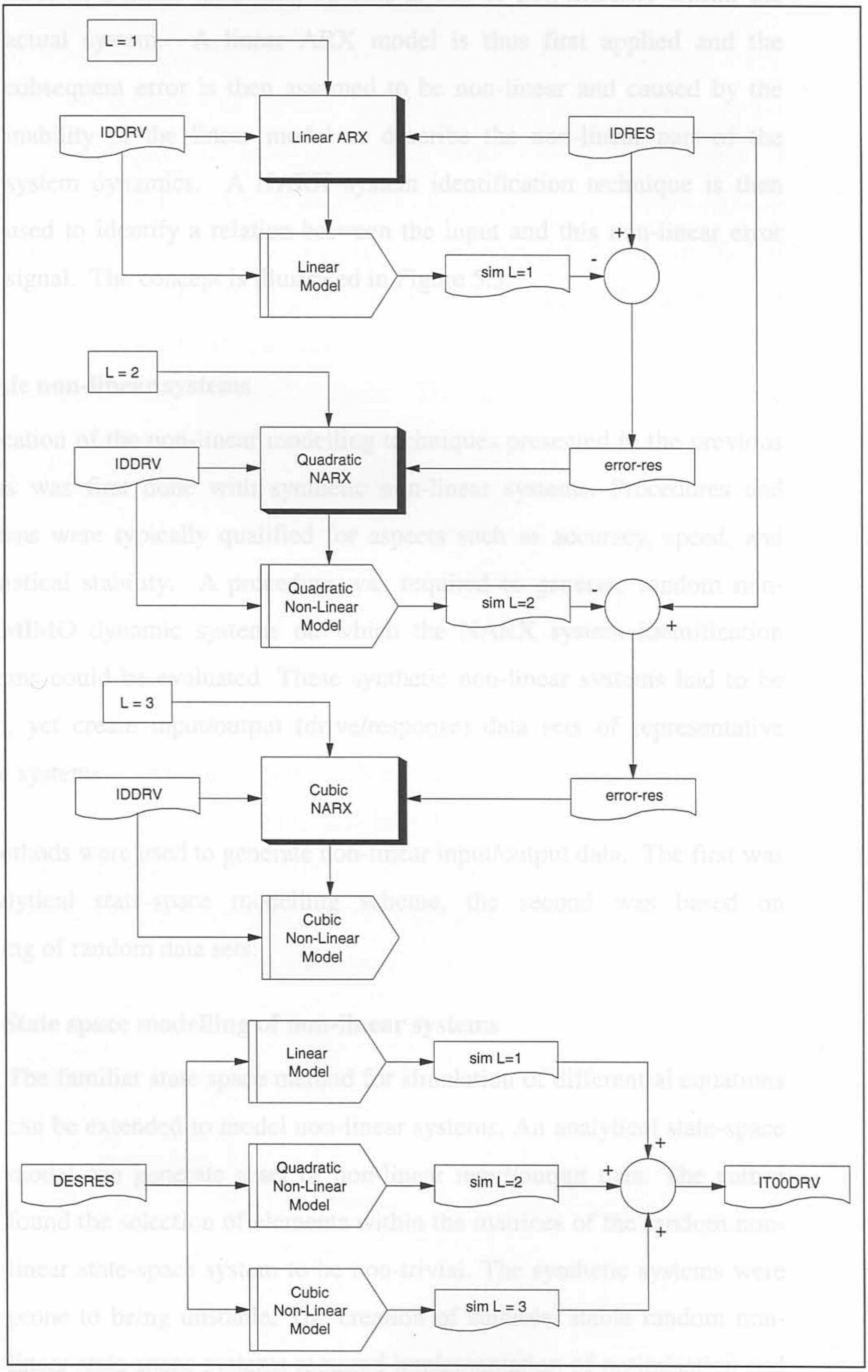


Figure 5.3: Non-linear error signal modelling

University of Pretoria- etd Cater, C R (1997)

Non-linear error signal modelling is based on the assumption that the error in a linear modelling scheme is due to non-linearity within the actual system. A linear ARX model is thus first applied and the subsequent error is then assumed to be non-linear and caused by the inability of the linear model to describe the non-linear part of the system dynamics. A NARX system identification technique is then used to identify a relation between the input and this non-linear error signal. The concept is illustrated in Figure 5.3.

5.6. Synthetic non-linear systems

Qualification of the non-linear modelling techniques presented in the previous Sections was first done with synthetic non-linear systems. Procedures and algorithms were typically qualified for aspects such as accuracy, speed, and mathematical stability. A procedure was required to generate random non-linear MIMO dynamic systems on which the NARX system identification algorithms could be evaluated. These synthetic non-linear systems had to be random, yet create input/output (drive/response) data sets of representative real-life systems.

Two methods were used to generate non-linear input/output data. The first was an analytical state-space modelling scheme, the second was based on modelling of random data sets.

5.6.1. State space modelling of non-linear systems

The familiar state space method for simulation of differential equations can be extended to model non-linear systems. An analytical state-space model can generate a set of non-linear input/output data. The author found the selection of elements within the matrices of the random non-linear state-space system to be non-trivial. The synthetic systems were prone to being unstable. The creation of suitable, stable random non-linear state space systems required implementation of optimisation and non-linear control system techniques [48]. Subsequently, a more simplistic alternative approach was devised.

University of Pretoria- etd Cater, C R (1997)

5.6.2. Random non-linear input/output data

A more general and easy to use system for creating synthetic non-linear dynamic systems was needed. The desired input parameters to such a random model generation technique would include the number of channels, model order for each channel, as well as the degree of non-linearity for each channel. Randomly selecting a set of NARX coefficients within these given parameters will generally result in an unstable system. The author devised a technique to create random stable non-linear dynamic systems. The input/output data from these systems could then be used to test the various non-linear algorithms. The random non-linear model generator as shown in Figure 5.4 makes use of a random signal generator to create two sets of random data. The first set will be used as a pseudo drive and the second as a pseudo response for the random model. These random data sets comply with the desired number of system channels. A NARX identification routine is then used to find a model parameter matrix which best describes the relation between the pseudo drive and response data, resulting in a random non-linear model. Valid non-linear input/output data can now be obtained by simulating drive data through the model and recording the subsequent responses. These random models proved ideal for evaluation of the various parameter estimation techniques as described in Appendix C.

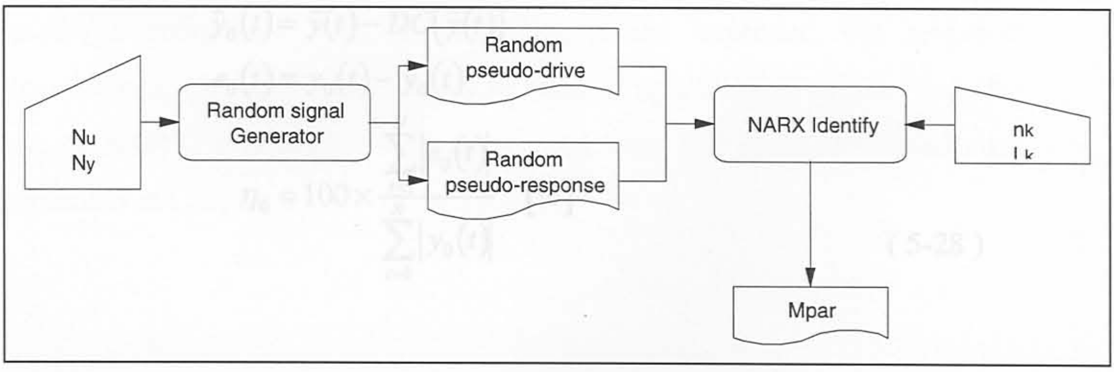


Figure 5.4: Random non-linear model generator

University of Pretoria- etd Cater, C R (1997)

5.7. Error functions

It is important to be able to quantify the success of a simulation exercise. The QanTiM fit value [34] as used in the linear simulation package also proved ideally suited to judge non-linear simulation accuracy. An error function e_y is defined by subtracting the achieved simulation response from the desired response.

$$e_y(t) = y(t) - \hat{y}(t) \qquad (5\text{-}26)$$

The QanTiM fit value is then calculated for each channel for N data points.

$$\eta_y = 100 \times \frac{\sum_{t=1}^{N} \left| e_y(t) \right|}{\sum_{t=1}^{N} \left| y(t) \right|} \qquad (5\text{-}27)$$

The error function as presented in (5-27) is inappropriate for responses with high DC offsets. The error value, η_y for a response signal with a DC offset will be factored by the DC value and thus be lower than actually is the case. Removing all DC offsets prior error quantification can however rectify this problem. In Equation (5-28) the author revised the QanTiM fit value. This DC-corrected function is used for evaluation of all simulation results.

$$y_0(t) = y(t) - DC\left(y(t)\right)$$

$$\hat{y}_0(t) = \hat{y}(t) - DC\left(\hat{y}(t)\right)$$

$$e_0(t) = y_0(t) - \hat{y}_0(t)$$

$$\eta_0 = 100 \times \frac{\sum_{t=1}^{N} \left| e_0(t) \right|}{\sum_{t=1}^{N} \left| y_0(t) \right|} \quad [\%] \qquad (5\text{-}28)$$

University of Pretoria- etd Cater, C R (1997)

5.8. Detecting non-linearity

Ideally a system should be classified as linear, or non-linear prior to modelling and thus warrant the use of a non-linear model. The general problem with non-linear modelling is determining which type of non-linearity, if any, is applicable. This could be polynomial, exponential, dead bands, signum functions or any other type of non-linearity. This difficulty is not encountered in linear systems modelling, which has largely attributed to the popularity of linear modelling.

Billings and Fadzil [6] suggested to plot the system gain against amplitude for a series of step inputs of varying amplitudes. This method is however not suited to practical mechanical systems, especially so for servo-hydraulic testing applications. Another method as suggested by Billings and Voon [8] showed that whenever the input: $u(t) + b$, $\overline{u}(t)=0$, $b \neq 0$ is applied to a system, the system cannot be linear if $\overline{z}_b(t) \neq \overline{z}(t)$ where $\overline{z}_b(t)$ and $\overline{z}(t)$ are the mean levels of the system output for the inputs b (i.e. $u(t) = 0$) and $u(t)+b$ respectively. In theory, this method based on evaluation of system mean responses is applicable to most servo-hydraulic test systems. In practice, however, comparison between linear and non-linear modelling of a system proved the best practical method of detecting non-linearity. Typically a system which proved difficult to model linearly was then identified using the NARX formulation. Generally the optimal model order as found in the linear case was then used along with a quadratic polynomial. If the modelling results improved, but remained unsatisfactory, a cubic model could be used. Application examples were the use of NARX techniques did improve modelling and simulation results are presented in Chapter 6.