



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/77050/>

Monograph:

Billings, S.A. (1987) Parametric Methods of Analysing Linear and Nonlinear Systems. Research Report. Acse Report 318 . Dept of Automatic Control and System Engineering. University of Sheffield

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



PARAMETRIC METHODS OF ANALYSING
LINEAR AND NONLINEAR SYSTEMS

S A Billings, BEng, PhD, CEng, MIEE,
AFIMA, MInstMC

Research Report No 318

August 1987

Invited Lecture on the short course
"Vibration Analysis and Identification
of Nonlinear Structures", Heriot-Watt
University, 16-18 Sept 1987

PARAMETRIC METHODS OF ANALYSING
LINEAR AND NONLINEAR SYSTEMS

S A Billings

Department of Control Engineering
University of Sheffield
Mappin Street, Sheffield S1 3JD

200064221



1. Introduction

Parameter estimation methods for linear systems are now a well established branch of control theory and the present study is an attempt to provide a brief introduction to these methods, to show how they can be extended to nonlinear systems and to demonstrate the application of the algorithms to real systems.

2. Linear Systems

Parameter estimation provides a means of estimating the parameters of a model of a process directly from the process input and output sequences. The basic approach is illustrated in Fig.1 where the measured inputs to the process are also inputs to the mathematical model. The outputs \hat{y} of this model depend on the structure assumed and on the values of the parameters β .

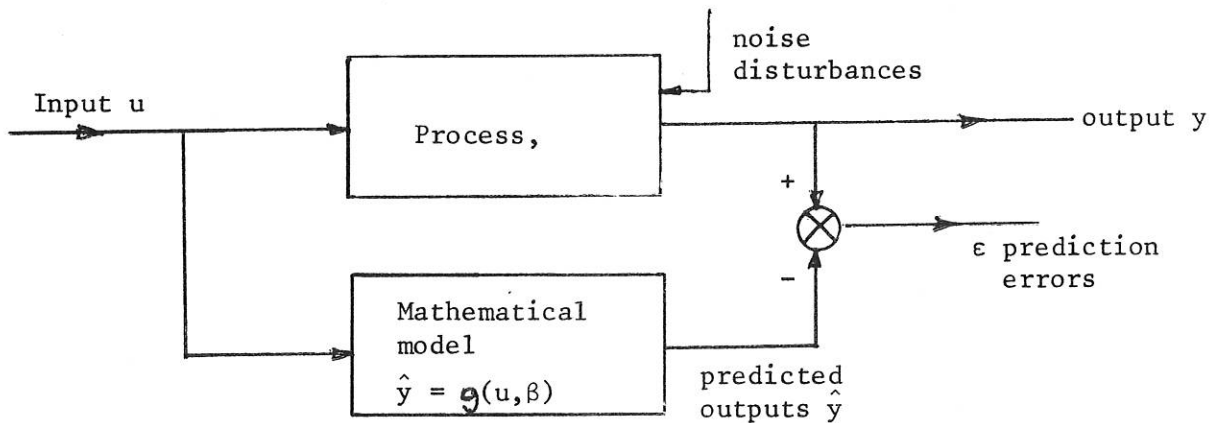


Fig.1

By comparing \hat{y} with the measured plant outputs y prediction errors $\epsilon = y - \hat{y}$ can be defined which depend on the unknown parameters β . Some suitable scalar function of ϵ can then be proposed as a cost function and the best values of β can then be estimated.

2.1. System Models

2.1.1. Difference Equations

A general single-input single-output linear system can be represented by the differential equation

$$p_n \frac{d^n y}{dt^n} + \dots + p_1 \frac{dy}{dt} + p_0 y(t) = q_m \frac{d^m u}{dt^m} + \dots$$

$$\dots + q_1 \frac{du}{dt} + q_0 u(t) \tag{1}$$

Assuming zero initial conditions the operator $S = \frac{d}{dt}$ can be introduced to yield the system transfer function

$$y(s) = \frac{q_m s^m + \dots + q_1 s + q_0}{p_n s^n + \dots + p_1 s + p_0} u(s) \tag{2}$$

or $y(s) = e^{-S\Delta} \frac{B(s)}{A(s)} u(s)$

where a time delay Δ has been introduced to make the description slightly more general. Almost all parameter estimation techniques are implemented digitally so the model of the process which is identified is the Z-transform of the system transfer function after a ZOH has been inserted.

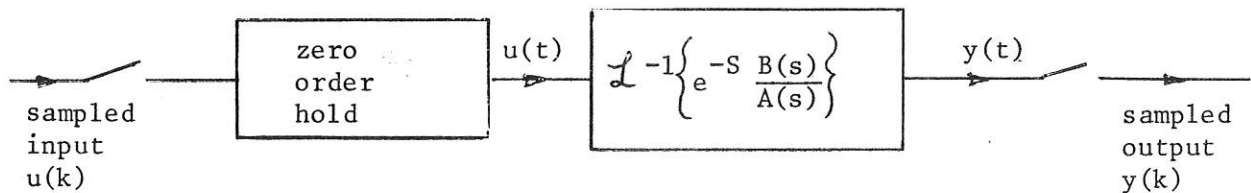


Fig.2

$$y(k) = \frac{z^{-\ell} (b_1 z^{-1} + b_2 z^{-2} + \dots + b_n z^{-n})}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} u(k)$$

$$= z^{-\ell} \frac{B(z^{-1})}{A(z^{-1})} u(k) \quad (3)$$

where n is the process model order, and ℓ is the system time delay. The system time delay is accommodated by shifting the output sequence back ℓ steps to give the model

$$y(k) = \frac{B(z^{-1})}{A(z^{-1})} u(k) \quad (4)$$

Multiplying out and interpreting z^{-j} as the delay operator $z^{-j}u(t) = u(t-j)$ gives

$$y(k) = b_1 u(k-1) + \dots + b_n u(k-n) - a_1 y(k-1) - \dots - a_n y(k-n) \quad (5)$$

This is called a difference equation. Notice that the impulse response can be obtained directly from eqn (5) by setting $u(0) = 1$, $u(k) = 0$ for $k > 0$.

If a set of sampled inputs and outputs of length N are available $u(1), u(2), \dots, u(N)$

$y(1), y(2), \dots, y(N)$

then eqn (5) can be expressed as

$$\begin{pmatrix} y(n+1) \\ y(n+2) \\ \vdots \\ y(N) \end{pmatrix} = \begin{pmatrix} -y(n) & \dots & -y(1) & | & u(n) & \dots & u(1) \\ & & & | & & & \\ & & & | & & & \\ & & & | & & & \\ -y(N-1) & & -y(N-n) & | & u(N-1) & & u(N-n) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

or $Y = \phi\beta$

(6)

2.1.2. Noise Models

The data sequences will normally contain additive noise and this is represented by a rational transfer function $N(z^{-1})$ driven by an uncorrelated white noise sequence $\xi(k)$.

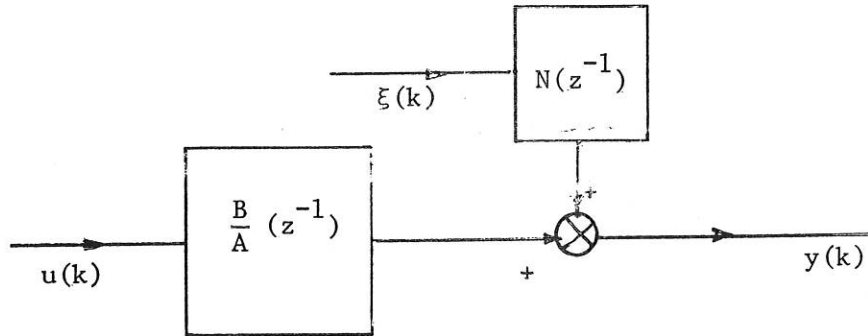


Fig.3

The most general noise model is an AutoRegressive Moving Average (ARMA) model

$$N(z^{-1}) \xi(k) = \frac{C(z^{-1})}{D(z^{-1})} \xi(k) = \frac{1+c_1 z^{-1} + \dots + c_m z^{-m}}{1+d_1 z^{-1} + \dots + d_m z^{-m}} \xi(k) \quad (7)$$

Special cases of this model are:-

- (a) $e(k) = C(z^{-1})\xi(k)$ - Moving Average (MA) model ($d_i = 0 \forall i$)
- (b) $e(k) = \frac{\xi(k)}{D(z^{-1})}$ - AutoRegressive (AR) model ($c_i = 0 \forall i$)

Expanding the AR model

$$e(k+m) = \xi(k+m) - d_1 e(k+m-1) \dots - d_m e(k)$$

$$\begin{pmatrix} e(k+1) \\ \vdots \\ e(N) \end{pmatrix} = \begin{pmatrix} -e(m) & \dots & -e(1) \\ \vdots & & \vdots \\ -e(N-1) & \dots & -e(N-m) \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{pmatrix} + \begin{pmatrix} \xi(m+1) \\ \vdots \\ \xi(N) \end{pmatrix} \quad (8)$$

or $\underline{e} = -\underline{E}\underline{f} + \underline{\xi}$

which is very similar to the process model.

2.1.3. The ARMAX model

The ARMAX model is a combination of the difference equation model eqn (5) and a MA noise model

$$\underbrace{A(z^{-1})}_{\text{AR}} y(k) = \underbrace{B(z^{-1})}_{\text{control or exogenous variable}} u(k) + \underbrace{C(z^{-1})}_{\text{MA}} \xi(k) \quad (9)$$

2.3. Properties of Estimates

If $\hat{\beta}$ is an estimate the bias of $\hat{\beta}$ is defined as

$$b[\hat{\beta}] = E[\hat{\beta}] - \beta \quad (10)$$

where $E[\cdot]$ represents the expectation operator. Ideally the estimates should be unbiased $b[\hat{\beta}] = 0$ so that $E[\hat{\beta}] = \beta$ or at least asymptotically unbiased

$$\lim_{N \rightarrow \infty} b[\hat{\beta}] = 0$$

If $\hat{\beta}$ is scalar the variance of $\hat{\beta}$ is defined as

$$\text{Var}[\hat{\beta}] = E[(\hat{\beta} - E[\hat{\beta}])^2]$$

When the estimate is unbiased $E[\hat{\beta}] = \beta$ and

$$\text{Var}[\hat{\beta}] = E[(\hat{\beta} - \beta)^2] \quad (11)$$

When $\hat{\beta}$ is a vector the covariance is defined as

$$\text{Cov}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \quad (12)$$

Notice that the mean squared error can be expressed as

$$E[(\hat{\beta} - \beta)^2] = \text{Var}(\hat{\beta}) + \text{bias}^2(\hat{\beta})$$

2.4. Parameter Estimation Algorithms

2.4.1. Simple Least Squares

Consider initially the idealised process shown below

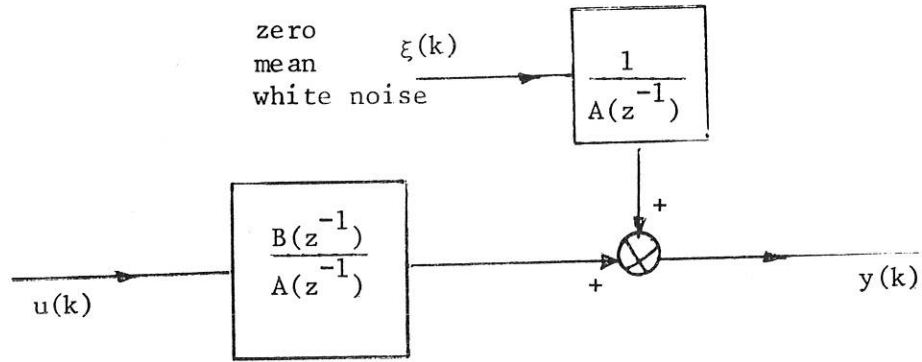


Fig.4

$$y(k) = \frac{B}{A} (z^{-1})u(k) + \frac{1}{A(z^{-1})} \xi(k)$$

$$\therefore y(k)(1+a_1 z^{-1} + \dots + a_n z^{-n}) = (b_1 z^{-1} + \dots + b_n z^{-n})u(k) + \xi(k) \quad (13)$$

giving

$$y(k) = x_k^T \beta + \xi(k) = [-y(k-1) \dots -y(k-n), u(k-1) \dots] \begin{pmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \end{pmatrix} + \xi(k)$$

or for N measurements

$$Y = \phi\beta + \xi$$

The least squares estimate is obtained by minimising the sum of squares of the model errors

$$\begin{aligned} J &= \frac{1}{N} \sum_{m=1}^N (y(m) - x_m^T \beta)^2 \\ &= \frac{1}{N} (Y - \hat{\phi}\beta)^T (Y - \hat{\phi}\beta) \end{aligned} \quad (14)$$

Computing $\frac{dJ}{d\hat{\beta}^T}$ and setting this to zero yields the simple least squares estimate

$$\hat{\beta} = (\phi^T \phi)^{-1} \phi^T Y \quad (15)$$

To check if the estimate is unbiased substitute $Y = \phi\beta + \xi$ into the estimate eqn (15)

$$\hat{\beta} = (\phi^T \phi)^{-1} \phi^T (\phi\beta + \xi)$$

or $\phi^T \phi (\hat{\beta} - \beta) = \phi^T \xi$ (16)

taking expected value

$$E[\phi^T \phi (\hat{\beta} - \beta)] = E[\phi^T \xi] = 0$$

hence $E[\hat{\beta}] = \beta$ and the estimate is unbiased. Since the additive noise on the model is zero mean and white

$$E[\xi \xi^T] = \sigma^2 I$$

and therefore

$$\begin{aligned} \text{Cov}(\hat{\beta}) &= E[(\phi^T \phi)^{-1} \phi^T \xi \xi^T \phi (\phi^T \phi)^{-1}] \\ &= \sigma^2 (\phi^T \phi)^{-1} \end{aligned} \quad (17)$$

An estimate of σ^2 the variance of the sequence $\xi(k)$ can be calculated as

$$\begin{aligned} \hat{\xi} &= Y - \phi \hat{\beta} \\ \therefore \sigma^2 &= \left(\sum_{k=1}^N \hat{\xi}^2(k) \right) / (N-n) \end{aligned} \quad (18)$$

2.4.2. Least Squares with Correlated Noise

Consider the more realistic situation where the plant output is corrupted by additive noise that is non-white

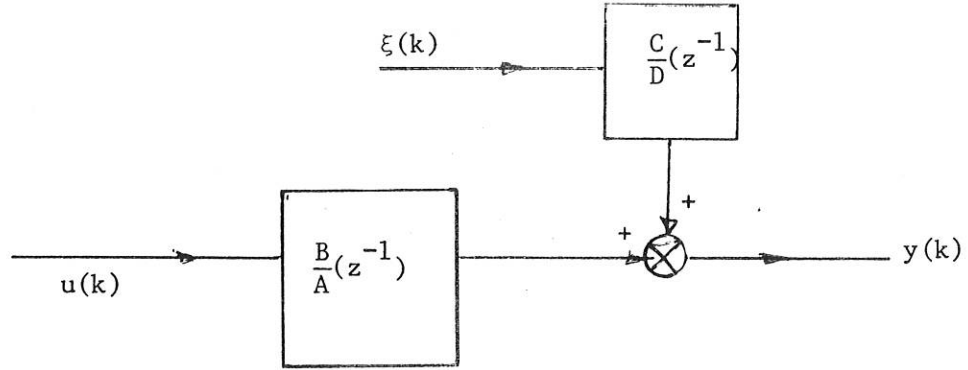


Fig.5

$$y(k) = \frac{B}{A} u(k) + \frac{C}{D} \xi(k) \quad (19)$$

Multiplying out to make the model linear-in-the-parameters

$$\begin{aligned} Ay(k) &= Bu(k) + \frac{AC}{D} \xi(k) \\ &= Bu(k) + e(k) \end{aligned} \quad (20)$$

or in matrix form

$$Y = \phi\beta + e$$

Analysing the bias of the estimate yields

$$E[(\phi^T \phi)(\hat{\beta} - \beta)] = E[\phi^T e] \neq 0$$

because $\phi = [-y; u]$ where $y(k)$ is a function of $e(k)$ which is non-white. Thus $E[\hat{\beta}] \neq \beta$ and the estimate will be biased.

There are numerous least squares based algorithms which have been developed to overcome this problem but only the extended least squares (ELS) method will be described in the present analysis.

Consider the ARMAX model of eqn (20) and set $C' = AC/D$ to give

$$A(z^{-1})y(k) = B(z^{-1})u(k) + C'(z^{-1})\xi(k) \quad (21)$$

Assuming initially that $\xi(k)$ is measurable, eqn (21) can be expressed as

$$y(k) = x_k^T \beta + \xi(k)$$

where

$$\begin{aligned} x_k^T &= [-y(k-1), \dots, -y(k-n), u(k-1), \dots, u(k-n), \xi(k-1) \dots \xi(k-n')] \\ \beta^T &= [a_1, \dots, a_n, b_1, \dots, b_n, c_1', \dots, c_n'] \end{aligned} \quad (22)$$

However, $\xi(k)$ is unknown so define x_k^T as

$$\hat{x}_k^T = [-y(k-1), \dots, -y(k-n), u(k-1), \dots, u(k-n), \hat{\xi}(k-1), \dots, \hat{\xi}(k-n')] \quad (23)$$

or for N measurements

$$Y = \psi \beta + \xi$$

where ψ is defined as

$$\psi = \begin{pmatrix} \hat{x}_{n+1}^T \\ \vdots \\ \hat{x}_N^T \end{pmatrix} \quad (24)$$

and $\hat{\xi}(k)$ is the residual sequence or prediction error defined by

$$\hat{\xi}(k) = y(k) - \hat{x}_k^T \hat{\beta}_{k-1} \quad (25)$$

This leads to the following algorithm which is known as extended least squares.

- (a) Compute the ordinary least squares estimate eqn (15) to provide start values
- (b) Estimate the prediction errors eqn (25)
- (c) Compute the extended least squares estimate

$$\beta_{\text{els}} = (\psi^T \psi)^{-1} \psi^T Y \quad (26)$$
- (d) Go to (b) and repeat until convergence.

Providing the algorithm converges the estimate $\hat{\beta}_{els}$ should be unbiased because ξ in eqn (23) will be a zero mean white noise sequence.

2.4.4. Maximum Likelihood and Prediction Error Methods

Suppose that $P(X:\theta_1, \theta_2 \dots)$ is a probability density function of known form for the random variable X , a function which contains several unknown parameters $\theta_1, \theta_2 \dots$. If an experiment is performed in which X takes particular values $(x_1, x_2 \dots x_n)$ the probability density function becomes a function of the unknown parameters only and a likelihood function can be defined

$$L(\theta_1, \theta_2 \dots | x_1, x_2 \dots x_n) = P(x_1, x_2 \dots x_n; \theta_1, \theta_2 \dots) \quad (27)$$

The likelihood function thus describes the probability that the parameters caused the measurements to occur. Maximising $L(\cdot)$ gives an estimate of the parameters which most likely caused the measurements to occur - Maximum Likelihood Estimates.

To apply maximum likelihood to the ARMAX model eqn (9) the central limit theorem is invoked so that the probability density function of the residuals $\xi(k)$ is Gaussian and hence of a known form.

Maximum likelihood estimates have minimum variance, are asymptotically unbiased, asymptotically efficient, consistent and asymptotically normal.

The idea of estimating the coefficients in a model to minimise the error between the predicted and measured outputs leads to a unification of parameter estimation algorithms called prediction error methods. Consider the ARMAX model eqn (9)

$$\frac{A}{C} y(k) = \frac{B}{C} u(k) + \xi(k)$$

$$\text{or } y(k) = \left(1 - \frac{A}{C}\right) y(k) + \left(\frac{B}{C}\right) u(k) + \xi(k) \quad (28)$$

It can be shown that the optimal predictor is

$$\hat{y}(k|k-1;\beta) = \left(1 - \frac{A}{C}\right) y(k) + \left(\frac{B}{C}\right) u(k) \quad (29)$$

and the prediction errors or innovations are defined as

$$\varepsilon(k) = y(k) - \hat{y}(k|k-1;\beta) \quad (30)$$

The principle of prediction error identification methods is to choose the parameter vector so that the prediction errors become small.

Two cost functions are normally used

$$J_1(\beta) = \text{trace}(WR_N(\beta)); \quad W = \text{positive definite} \quad (31)$$

$$J_2(\beta) = \log \det(R_N(\beta)) \quad (32)$$

where

$$R_N(\beta) = \frac{1}{N} \sum_{k=1}^N \varepsilon(k;\beta) \varepsilon^T(k;\beta) \quad (33)$$

Minimisation of J_1 or J_2 leads to a prediction error estimate of β .

It can be shown that under weak assumptions, the PE estimates are consistent and asymptotically Gaussian distributed.

If the noise on the model is white Gaussian then the simple least squares, maximum likelihood and prediction error estimates will all be equal. If the noise is coloured but Gaussian then the maximum likelihood and prediction error estimates will be equal. Many of the asymptotic properties which hold for maximum likelihood estimates also apply to prediction error estimates even though no probabilistic knowledge is assumed.

Notice that all the algorithms discussed above can be implemented in recursive form so that the estimates are updated each time a new measurement becomes available.

2.4.5. Model Validation

Once a model has been estimated it must be validated to ensure that it is a representative description of the system. Numerous model validity tests are available and at the very least the experimenter should check that:-

- (a) the predicted output of the model indicates that the model is a reasonable fit
- (b) the autocorrelation of the residuals $\xi(k)$ or prediction errors is an impulse at the origin
- (c) the cross correlation between the input and the residuals is zero.

2.4.6. Frequency Domain Estimation

Classical spectral analysis uses the FFT in conjunction with the formula

$$S_{uy}(j\omega) = H(j\omega)S_{uu}(\omega) + S_{ue}(j\omega) \quad (34)$$

to estimate the system frequency response $H(j\omega)$. Whilst this approach is widely used the performance is poor for short data lengths and leakage distorts the spectrum and masks weak signals. Autoregressive and maximum entropy based methods were introduced to overcome this problem when estimating the power spectrum of a signal but are not appropriate for the cross spectrum estimate $S_{uy}(\omega)$ required in eqn (34). However, once the coefficients in the model

$$y(k) = z^{-\ell} \frac{B(z^{-1})}{A(z^{-1})} u(k) + \frac{C(z^{-1})}{D(z^{-1})} \xi(k) \quad (35)$$

have been estimated using any one of the parameter estimation methods described above the system frequency response $H(j\omega)$ can be readily computed. To achieve this objective discard the estimated noise model in eqn (35) to leave

$$y(k) = z^{-\ell} \frac{B(z^{-1})}{A(z^{-1})} u(k) \quad (36)$$

and substitute $z = e^{j\omega T}$, $T =$ sampling interval, to yield the frequency response estimate directly

$$\hat{H}(j\omega) = \left\{ z^{-\ell} \frac{B(z^{-1})}{A(z^{-1})} \right\}_{z=e^{j\omega T}} \quad (37)$$

Notice that the estimates based on eqn (34) involve averaging, windowing and smoothing with the assumption that $S_{ue}(j\omega)$ will tend to zero for a large enough sample. In contrast to this the parametric approach using eqn (37) is based on estimation in the time domain and allows the estimation of the system frequency response without any noise. The noise, the right hand term in eqn (35) is discarded, not averaged out, before estimating the frequency response using eqn (37) and it might be anticipated that this will result in smoother frequency response estimates particularly for short record lengths.

To illustrate these ideas the system S_1

$$y(k) = \frac{z^{-1} + 0.5z^{-2}}{1 - 1.5z^{-1} + 0.7z^{-2}} u(k) + \frac{1 + 0.3z^{-1}}{1 + 0.6z^{-1}} \xi(k) \quad (38)$$

was simulated to generate 500 data pairs whenever $u(k)$ was a 6th order prbs and $\xi(k)$ was a random signal uniformly distributed between -1.0 and 1.0. A least squares based estimator using a tenth order AR noise model produced the estimated process model

$$x(k) = \frac{1.012z^{-1} + 0.4884z^{-2}}{1 - 1.498z^{-1} + 0.698z^{-2}} u(k) \quad (39)$$

The choice of model order and time delay were identified using loss function analysis, pole-zero cancellation, autocorrelation of the residuals and cross-correlation between the input and residual tests. The system frequency response was computed by substituting $z = e^{j\omega T}$ into eqn (39). Estimates of the spectra using eqn (34) were obtained by padding the 500 data pairs with twelve zeros, applying a Hamming window and smoothing the estimates. A comparison of the estimates is shown below.

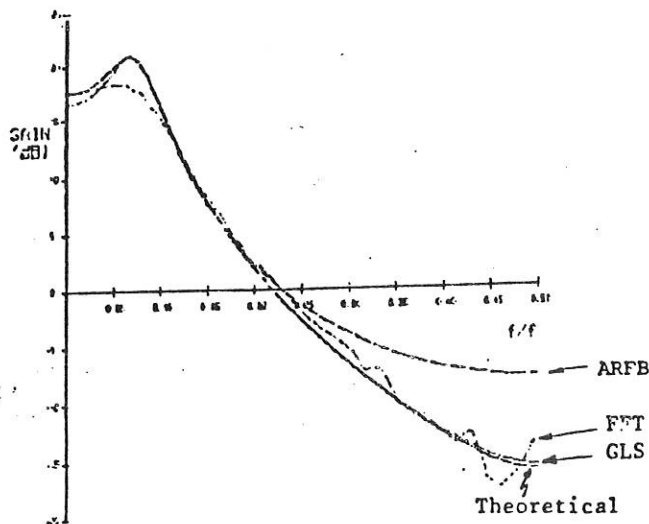


Fig.6

Notice that the parametric estimate tends to be much more accurate for higher frequencies and this becomes more evident for lower S/N ratios and shorter record lengths. Numerous simulated examples and applications to real data have produced similar results.

3. Nonlinear Systems

Parameter estimation methods for nonlinear systems where the structural form of the describing differential equations are known are now well established. When little a priori information is available and the process is treated as a black-box, the usual approach is to expand the input/output using a suitable model representation. Whilst the Volterra or Wiener series approach is widely documented this almost always leads to a model with an enormous number of parameters even for simple nonlinear systems. The NARMAX (Nonlinear ARMAX) model was introduced to overcome this problem and results relating to this model will be described below.

3.1. The NARMAX Model

If a system is linear then it is finitely realizable and can be represented by the linear difference equation model

$$x(k) = \sum_{i=1}^{n_y} (a_i x(k-i)) + \sum_{i=1}^{n_u} (b_i u(k-i)) \quad (40)$$

if the Hankel matrix of the system has finite rank. When the system is nonlinear a similar representation can be derived by utilizing results from differential geometry to yield the nonlinear difference equation model

$$y(k) = F^* [x(k-1), \dots, x(k-n_x), u(k-1), \dots, u(k-n_u)] \quad (41)$$

where $F^*[\cdot]$ is some nonlinear function of $u(\cdot)$ and $x(\cdot)$. The model of eqn (41) can be shown to exist whenever

- (a) the state-space of the Nerode realization does not have infinite dimensions (i.e. we exclude distributed parameter systems), and
- (b) the linearized system around the origin has a Hankel matrix of maximum rank (i.e. a linearized model would exist if the system were operated close to an equilibrium point).

Equation (41) represents the single-input single-output case but the results have been extended to include multivariable systems. The Hammerstein, Wiener, bilinear, Volterra and other well known nonlinear models can be shown to be special cases of eqn (41).

An equivalent representation for nonlinear stochastic systems can be derived by considering input-output maps based on conditional probability density functions to yield the model

$$y(k) = F[y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u), \epsilon(k-1), \dots, \epsilon(k-n_\epsilon)] + \epsilon(k) \quad (42)$$

where $\epsilon(k)$ is the prediction error. This model is referred to as the Nonlinear AutoRegressive Moving Average model with eXogenous inputs or NARMAX model.

A NARMAX model with first order dynamics expanded as a second order polynomial nonlinearity would for example be represented as

$$\begin{aligned} x(k) &= F_2[x(k-1), u(k-1)] \\ &= C_1 x(k-1) + C_2 u(k-1) + C_{11} x^2(k-1) + C_{12} x(k-1)u(k-1) + C_{22} u^2(k-1) \end{aligned} \quad (43)$$

Assuming that the output measurements are corrupted by additive noise

$$y(k) = x(k) + e(k)$$

gives the input-output model

$$\begin{aligned}
 y(k) = & C_1 y(k-1) + C_2 u(k-1) + C_{11} y^2(k-1) + C_{12} y(k-1)u(k-1) \\
 & + C_{22} u^2(k-1) + e(k) - C_1 e(k-1) - 2C_{11} y(k-1)e(k-1) \\
 & + C_{11} e^2(k-1) - C_{12} e(k-1)u(k-1)
 \end{aligned} \tag{44}$$

Because the NARMAX model maps the past input and output into the present output multiplicative noise terms are induced in the model even though the noise was additive at the output. In general the noise may enter the system internally and because the system is nonlinear it will not always be possible to translate it to be additive at the output as in the example shown below.

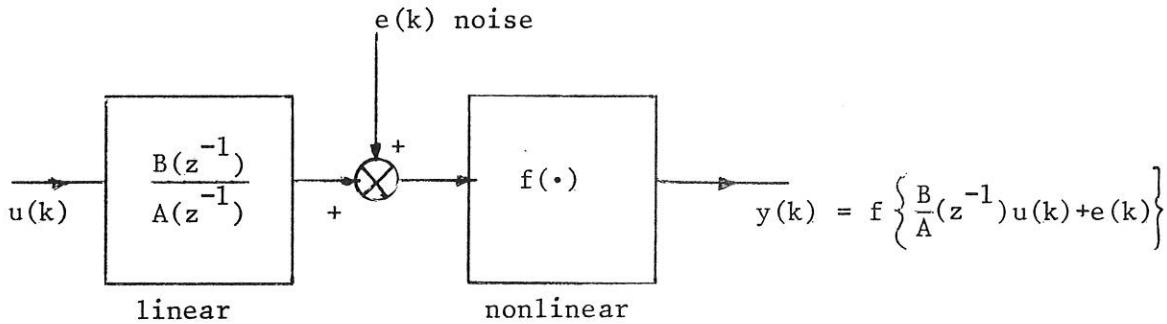


Fig.7

This situation will again result in multiplicative noise terms in the NARMAX model with the added complication that the noise source and the prediction error will not in general be equal. Since most of the parameter estimation techniques derived for linear systems assume that the noise is independent of the input, biased estimates results when they are applied to nonlinear systems eqn (42).

The recursive extended least squares (RELS) algorithm can however be readily adapted to the NARMAX model, by defining the following vectors

$$\begin{aligned} \mathbf{x}_k^T &= [y(k-1), u(k-1), y^2(k-1), y(k-1)u(k-1), u^2(k-1), \varepsilon(k-1), \\ &\quad \varepsilon(k-1)y(k-1), u(k-1)\varepsilon(k-1), \varepsilon^2(k-1)] \\ \hat{\beta}^T &= [\hat{C}_1, \hat{C}_2 \dots \hat{C}_s] \\ \varepsilon(k+1) &= y(k+1) - \mathbf{x}_{k+1}^T \hat{\beta} \end{aligned} \quad (45)$$

for the model of eqn (44) for example. With these definitions the standard RELS algorithm can be applied to yield unbiased parameter estimates. The development of recursive maximum likelihood and instrumental variable algorithms for the NARMAX model is not quite so straightforward.

The direct application of an offline maximum likelihood algorithm is not possible because in general the prediction errors will not have a Gaussian distribution. However, by considering the loss function

$$J_2(\beta) = \frac{1}{N} \log_e \det \sum_{k=1}^N \varepsilon(k; \beta) \varepsilon^T(k; \beta) \quad (46)$$

it can be shown that the prediction error estimates obtained by minimising eqn (46) have very similar asymptotic properties to the maximum likelihood estimates even when $\varepsilon(k)$ is non-gaussian. A prediction error algorithm has been developed for the NARMAX model based on this result.

Notice that the determination of the model structure or which variables to include in the NARMAX model is vital if a parsimonious representation of the system is to be identified. Simply increasing the order of the dynamic terms (n_y, n_u, n_ε) in eqn (42) and the order

of the polynomial expansion to achieve the desired prediction accuracy will in general result in an excessively complex model and numerical ill-conditioned computations. Consequently all the parameter estimation algorithms derived for the NARMAX have been augmented with a stepwise regression algorithm, a likelihood ratio test and other tests to detect the model structure or significant terms in the model prior to final estimation.

These results have recently been extended to include a new orthogonal estimation algorithm for the NARMAX model. This algorithm allows each coefficient in the model to be estimated recursively and quite independently of the other terms in the model because of the orthogonal property which holds for any input and automatically shows the contribution that each term makes to the output variance.

Alternative model expansions are being investigated, and algorithms based on a rational model expansion of $F(\cdot)$ in eqn (42) rather than a polynomial expansion and other globally valid nonlinear difference equations have been derived. The approximation of nonlinear systems by fitting a series of locally linear models and patching these together to form a nonlinear description have also been studied. Unfortunately whilst this appears initially to be an attractive approach it often produces input dependent models which are not representative of the system for inputs other than those used in the identification experiment.

3.1.1. Nonlinear model validation

Whichever model formulation or identification algorithm is implemented it is important to test that the identified model does adequately describe the data set. When the system is nonlinear the

residuals $\xi(k)$ should be unpredictable from all linear and nonlinear combinations of past inputs and outputs and this condition will hold iff

$$\begin{aligned}
 \phi_{\xi\xi}(\tau) &= \delta(\tau) & \phi_{u\xi}(\tau) &= 0 \quad \forall \tau \\
 \phi_{u\xi}(\tau) &= 0 \quad \forall \tau & \phi_{\xi\xi u}(\tau) &= 0 \quad \forall \tau \\
 \phi_{\xi\xi u}(\tau) &= E[\xi(k)\xi(k-1-\tau)u(k-1-\tau)] = 0 & & \tau \geq 0
 \end{aligned} \tag{47}$$

Notice that for nonlinear systems the traditional linear tests $\phi_{\xi\xi}(\tau)$ and $\phi_{u\xi}(\tau)$ are not sufficient.

Experience has shown that when using a prediction error algorithm the tests in both eqn (47) often give the experimenter a great deal of information regarding the deficiencies in the fitted model and can indicate which terms should be included in the model to improve the fit.

3.1.2. Experiment design for nonlinear systems

The design of input sequences appropriate for the identification of nonlinear systems has been studied in some detail. The analysis has shown that the common choice of a binary input sequence if used in nonlinear system identification can lead to disastrous results. Binary sequences not only do not maximise the determinant of the average information matrix but they also almost always make it singular so that even identifiability is lost. The input excitation for nonlinear systems must excite all the dynamic modes over the complete amplitude range of interest. The first requirement specifies the spectral density of the input, the second the probability density function.

The design of inputs for nonlinear systems when the only constraint is on the input is complex but the following rules can

be derived from information theoretic arguments

- (i) for a power or amplitude constraint on the input, the input should be an independent sequence
- (ii) for a power constraint on the input the input should be Gaussian
- (iii) for an amplitude constraint on the input the input should be uniformly distributed
- (iv) for a multivariable system the inputs should be mutually independent.

The design of inputs when there are constraints on the output has been investigated but the analytical solution to such problems seems to be intractable. Some practical guidelines that can be followed can however be formulated.

3.1.3. An example

To illustrate some of the ideas associated with parameter estimation based on the NARMAX model consider the identification of a model relating the input volume flow rate $u(t)$ and the level of liquid $y(t)$ in the interconnected tanks illustrated below.

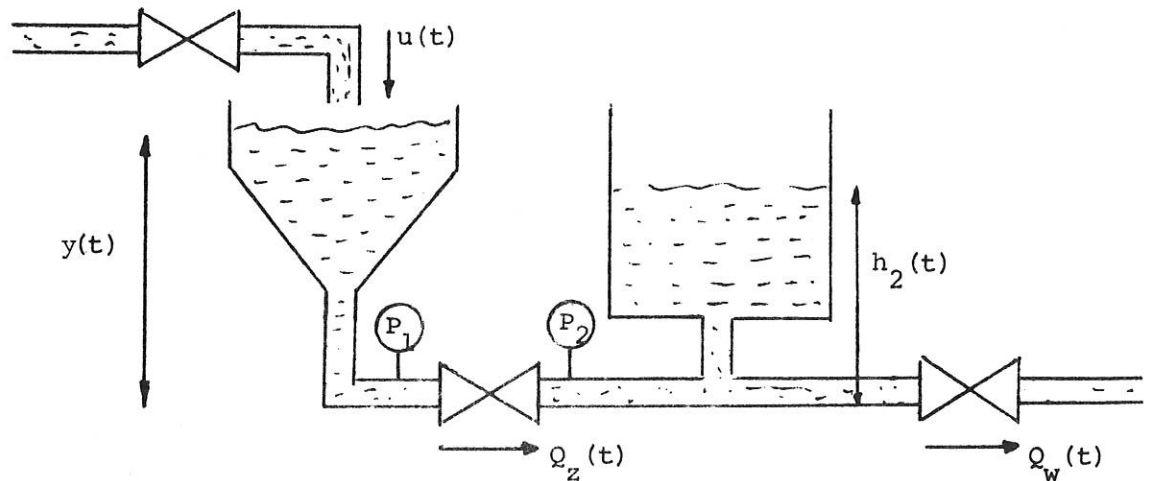


Fig.8. Interconnected Tanks

A zero mean Gaussian signal was used to perturb the input $u(t)$ and 1000 data pairs were recorded by sampling the input and output at 9.6 secs.

In the early stages of any identification procedure it is important to establish if the process under test exhibits nonlinear characteristics which will warrant a nonlinear model. This can readily be achieved using a simple correlation test. If the third order moments of the input are zero and all even order moments exist (a sine wave, gaussian or ternary sequence would for example satisfy the properties) then the process is linear iff

$$\phi_{y'y'^2}(\tau) = E[(y(k)-\bar{y})(y(k+\tau)-\bar{y})^2] = 0 \quad \forall \tau \quad (48)$$

$\phi_{y'y'^2}(\tau)$ for the liquid level system is illustrated below and clearly shows that, as expected, the liquid level system is highly nonlinear.

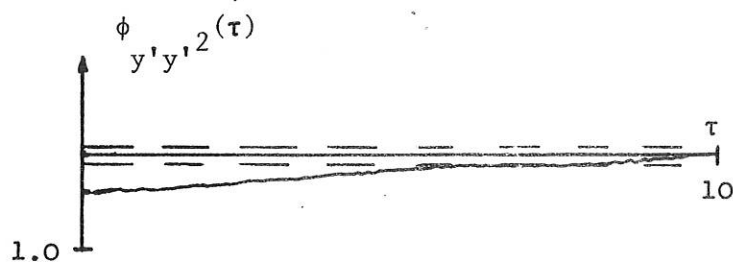


Fig.9. Nonlinear detection test

Initially a linear model was fitted to the data using a maximum likelihood algorithm to give the representation

$$\begin{aligned} y(k) = & 0.746y(k-1) + 0.340y(k-2) - 0.122y(k-3) \\ & + 0.471u(k-1) - 0.174u(k-2) - 0.040u(k-3) \\ & + \epsilon(k) + 0.423\epsilon(k-1) + 0.038\epsilon(k-2) \end{aligned} \quad (49)$$

A comparison of the process and linear model predicted output is illustrated in Fig.10. The model validity tests eqn (47), for

this model are illustrated in Fig.11. Notice that although $\phi_{\xi\xi}(\tau)$ and $\phi_{u\xi}(\tau)$ indicate linear adequacy for the model eqn (49), $\phi_{u^2\xi}(\tau)$ and $\phi_{u^2\xi^2}(\tau)$ are well outside the 95% confidence bands indicating that nonlinear terms should be included in the model description.

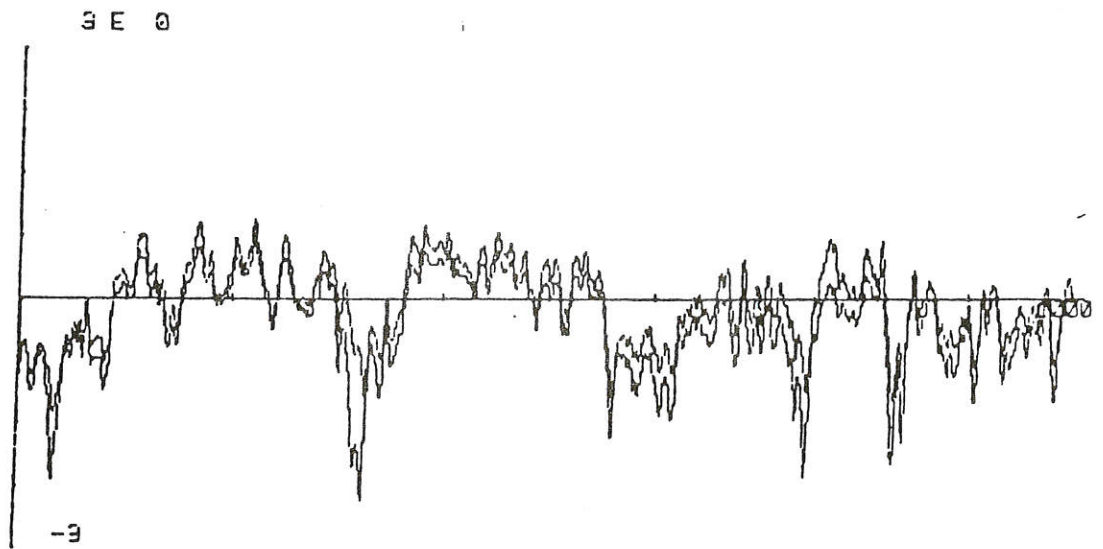


Fig.10. Process and Predicted output-best linear model

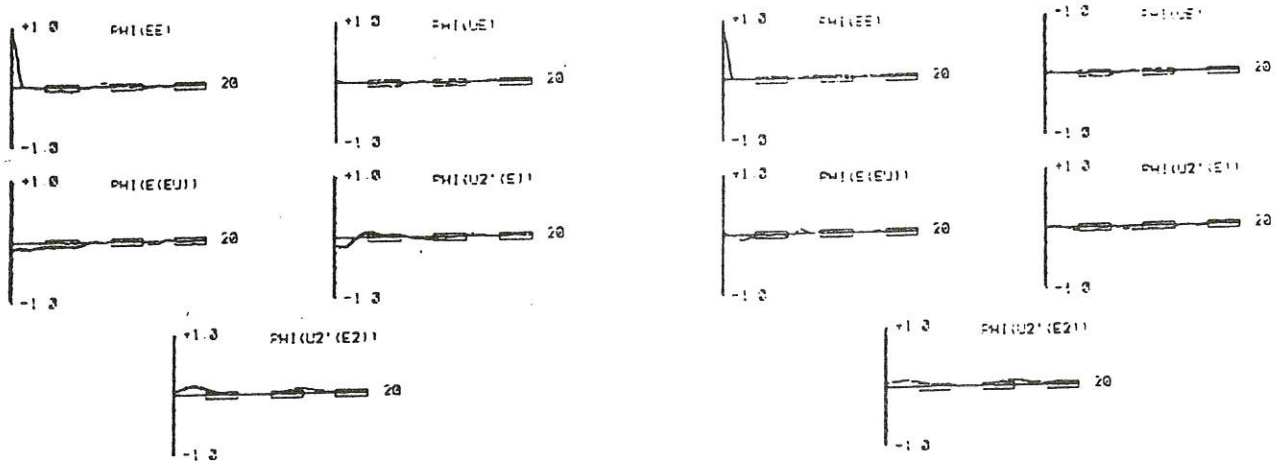


Fig.11. Model Validation - best linear model

Fig.13. Model validation - best nonlinear model

The effect of introducing nonlinear terms into the model was therefore investigated and a prediction error algorithm yielded the NARMAX model representation

$$\begin{aligned} y(k) = & 0.436y(k-1)+0.681y(k-2)-0.149y(k-3) \\ & +0.396u(k-1)+0.014u(k-2)-0.071u(k-3) \\ & -0.351y(k-1)u(k-1)-0.034y^2(k-2) \\ & -0.135y(k-2)u(k-2)-0.027y^3(k-2)-0.108y^2(k-2)u(k-2) \\ & -0.099u^3(k-2)+\varepsilon(k)+0.344\varepsilon(k-1)-0.201\varepsilon(k-2) \end{aligned} \quad (50)$$

The model validity tests for the model of eqn (50) are illustrated in Fig.13 and these together with the comparison of the NARMAX model predicted and process output Fig.12 show the considerable improvement in the prediction capabilities of the estimated NARMAX model eqn (50) compared with the best linear model eqn (49).

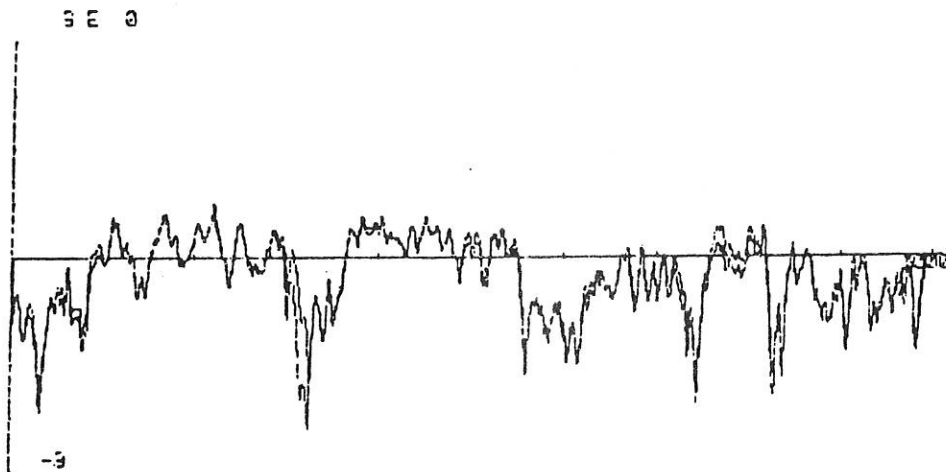


Fig.12. Process and Predicted output for the estimated NARMAX model

Numerous other simulated and industrial processes have been identified using these techniques including a heat exchanger, a 6996 bhp diesel generator and a turbo-charged lorry engine. In all cases the fitted model contained typically six to ten process model terms, a considerable reduction compared to the thousands of terms required to describe the same systems by Volterra or Wiener series.

3.2. Spectral Analysis for Nonlinear Systems

The application of linear spectral estimation procedures to data generated from nonlinear systems can introduce significant errors. For example, estimation of the frequency response function for a system represented by the model

$$x(k) = ax(k-1) + bu(k-1) + cu^2(k-1) \quad (51)$$

where $u(k)$ is a signal whose third order moments are zero (eg zero mean Gaussian or sine wave inputs) yields

$$\hat{H}_1(j\omega) = \frac{S_{ux}(j\omega)}{S_{uu}(\omega)} = \frac{b}{e^{j\omega} - a} \quad (52)$$

The estimate of the frequency response function is therefore completely independent of c the nonlinear term in eqn (51) showing the limitations of linear methods when applied to nonlinear systems.

Efforts to resolve these problems have in the past concentrated on functional series methods and higher order spectra. Introducing the Volterra series

$$x(t) = \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} \dots \int h_n(\tau_1, \tau_2, \dots, \tau_n) \prod_{i=1}^n u(t-\tau_i) d\tau_i \quad (53)$$

and taking the Fourier transform of the kernels defines the generalised

transfer functions

$$H_n(\omega_1, \omega_2, \dots, \omega_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \tau_2, \dots, \tau_n) \cdot \exp[-j(\omega_1 \tau_1 + \dots + \omega_n \tau_n)] d\tau_1 \dots d\tau_n \quad (54)$$

If the system were linear only the first term, the standard convolution integral, would exist in eqn (53) and eqn (54) would yield $H_1(j\omega)$ which is the linear frequency response function. When the system is nonlinear there is no single function which characterizes the frequency response behaviour and generalised frequency response functions of the form of eqn (54) must be evaluated up to order n the degree of nonlinearity. Traditionally multidimensional FFT based algorithms have been used to estimate the generalized transfer functions but unfortunately all the algorithms require excessive data lengths, make unrealistic assumptions about the system, require special inputs (eg Gaussian white), and produce results which are input dependent.

Virtually all of these restrictions can be avoided by fitting a NARMAX model and using the probing or harmonic input method to compute the $H_n(\omega_1, \omega_2, \dots, \omega_n)$ functions. This approach represents an extension of the parametric linear frequency response estimation scheme of section 2.4.6.

To illustrate the ideas involved consider the nonlinear Hammerstein model S_2 illustrated in Fig.14 where the noise $e(t)$ was $N(0,0.08)$ the input was $N(0,1.0)$ and 1000 data pairs were considered.

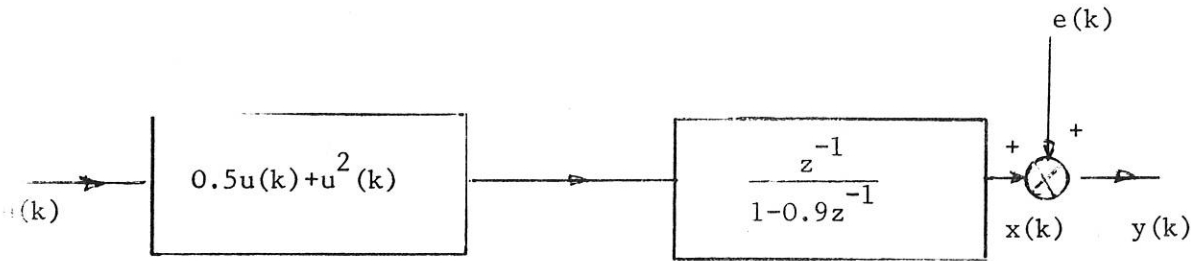


Fig.14. System S_2

From Fig.14 the true system model can be expressed as a NARMAX model

$$\begin{aligned} x(k) &= 0.9x(k-1) + 0.5u(k-1) + u^2(k-1) \\ y(k) &= x(k) + e(k) \end{aligned} \quad (55)$$

The application of the nonlinear detection test $\phi_{y'y'^2}(\tau)$ eqn (48) clearly showed the system was nonlinear and the orthogonal estimation algorithm identified the following model structure and parameters

$$\begin{aligned} y(k) &= 0.00158z + 0.8997y(k-1) + 0.5007u(k-1) \\ &\quad + 1.001u^2(k-1) + e(k) - 0.874e(k-1) \end{aligned} \quad (56)$$

The algorithm estimated that the terms in eqn (56) contributed 99.954% to the variation in $y(k)$ thus indicating why all other higher order dynamic and nonlinear terms were deleted from the model.

Note that the Volterra series representation for this system would require the estimation of about 500 parameters!

The generalized transfer functions defined by eqn (54) can now be computed. Probing eqn (56) with a single exponential $u(k) = e^{j\omega kT}$ ignoring the almost zero constant term are the noise terms and setting $T = 1$ yields

$$H_1(j\omega)e^{j\omega k} = 0.8997H_1(j\omega)e^{j\omega(k-1)} + 0.5007e^{j\omega(k-1)} + 1.00[e^{j\omega(k-1)}]^2 \quad (57)$$

Equating coefficients of $e^{j\omega k}$ yields the first order frequency response function

$$H_1(j\omega) = \frac{0.5007}{e^{j\omega} - 0.8997} \quad (58)$$

Probing with the input $u(k) = e^{j\omega_1 k} + e^{j\omega_2 k}$ and equating coefficients of $e^{j(\omega_1 + \omega_2)T}$ yields in a similar manner the second order generalized transfer function

$$H_2(j\omega_1, j\omega_2) = \frac{1.001}{e^{j(\omega_1 + \omega_2)} - 0.8997} \quad (59)$$

Both $H_1(j\omega)$ and $H_2(j\omega_1, j\omega_2)$ are plotted in Fig.15 below.

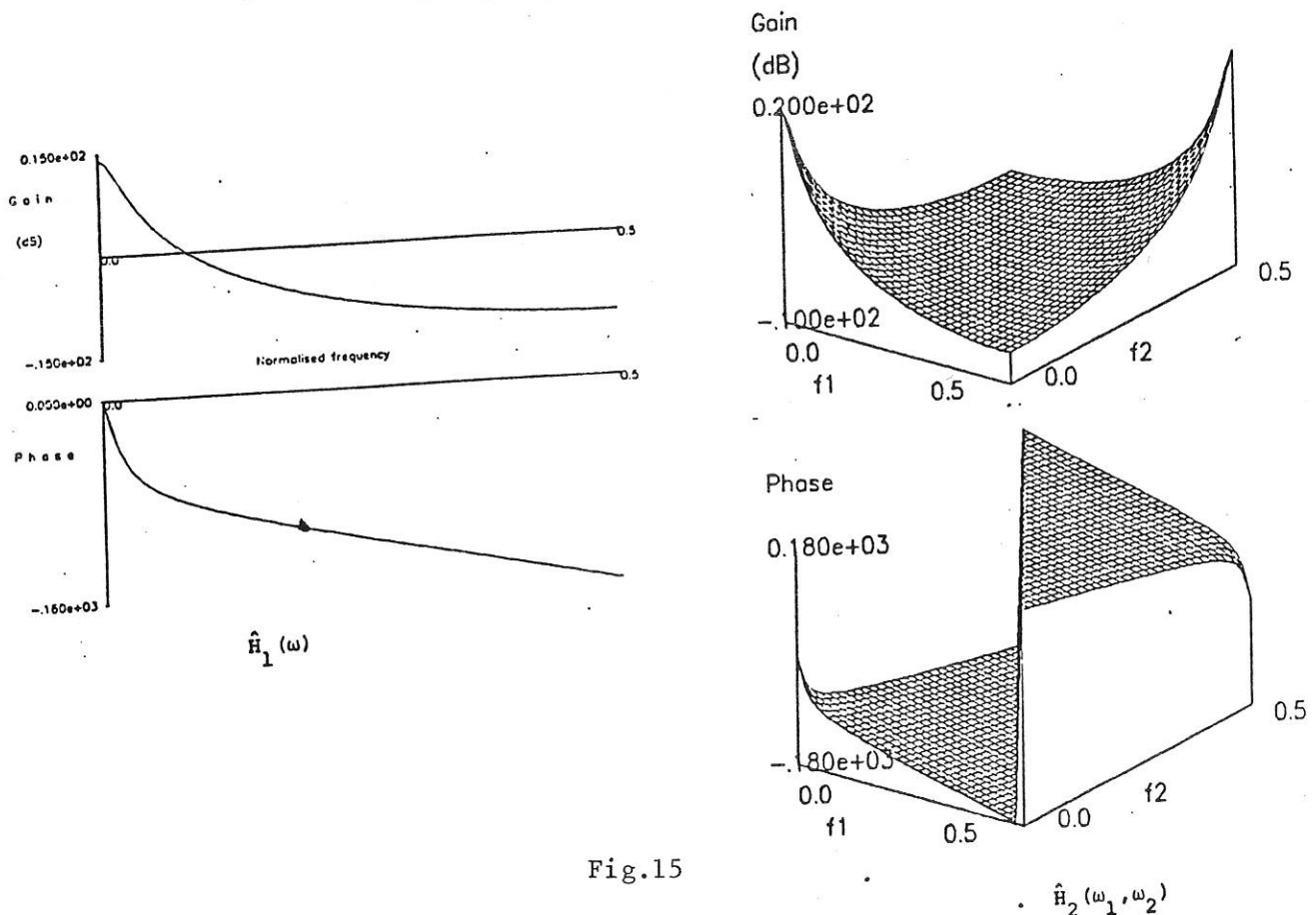


Fig.15

Linearizing the estimated model eqn (56) about different operating points yields the evolution of the linearized spectra illustrated in Fig.16.

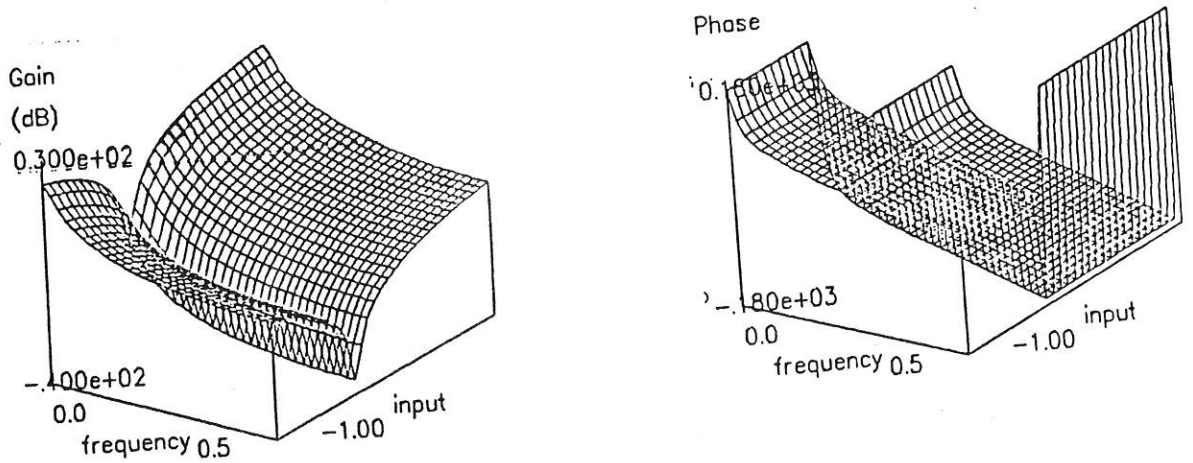


Fig.16

As a second example consider the system S_3

$$\ddot{y} + 20\dot{y} + 10^4 y + 3 \times 10^3 u y = u \quad (60)$$

This can be simulated and a NARMAX representation estimated to yield the estimates of $H_1(j\omega)$, $H_2(j\omega_1, j\omega_2)$, $H_3(j\omega_1, j\omega_2, j\omega_3)$ illustrated below.

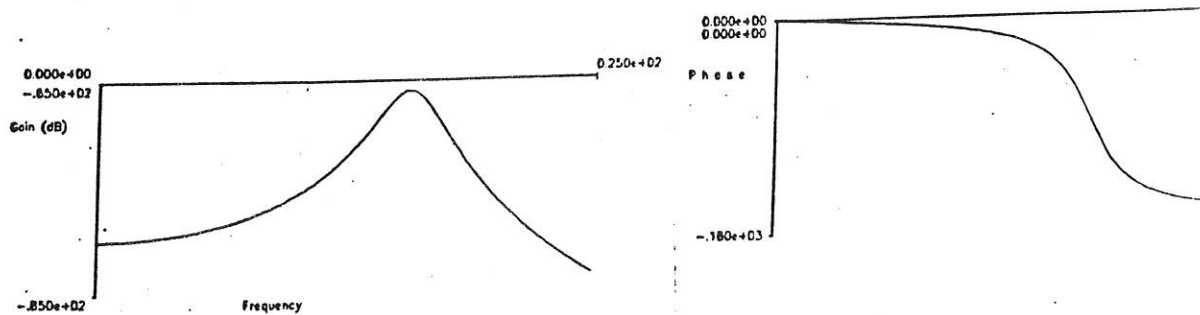


Fig.17. $H_1(\cdot)$

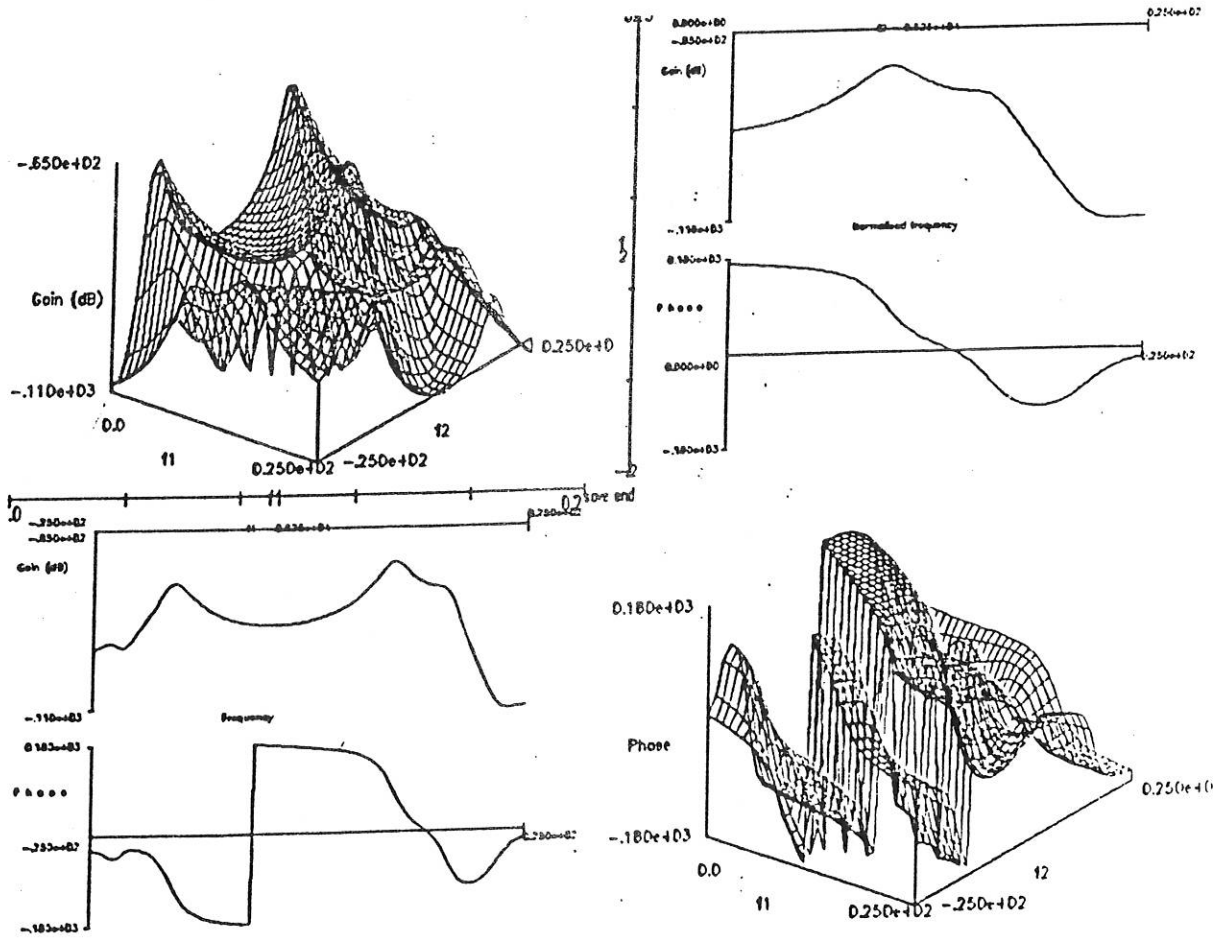


Fig.18. $H_2(\cdot, \cdot)$

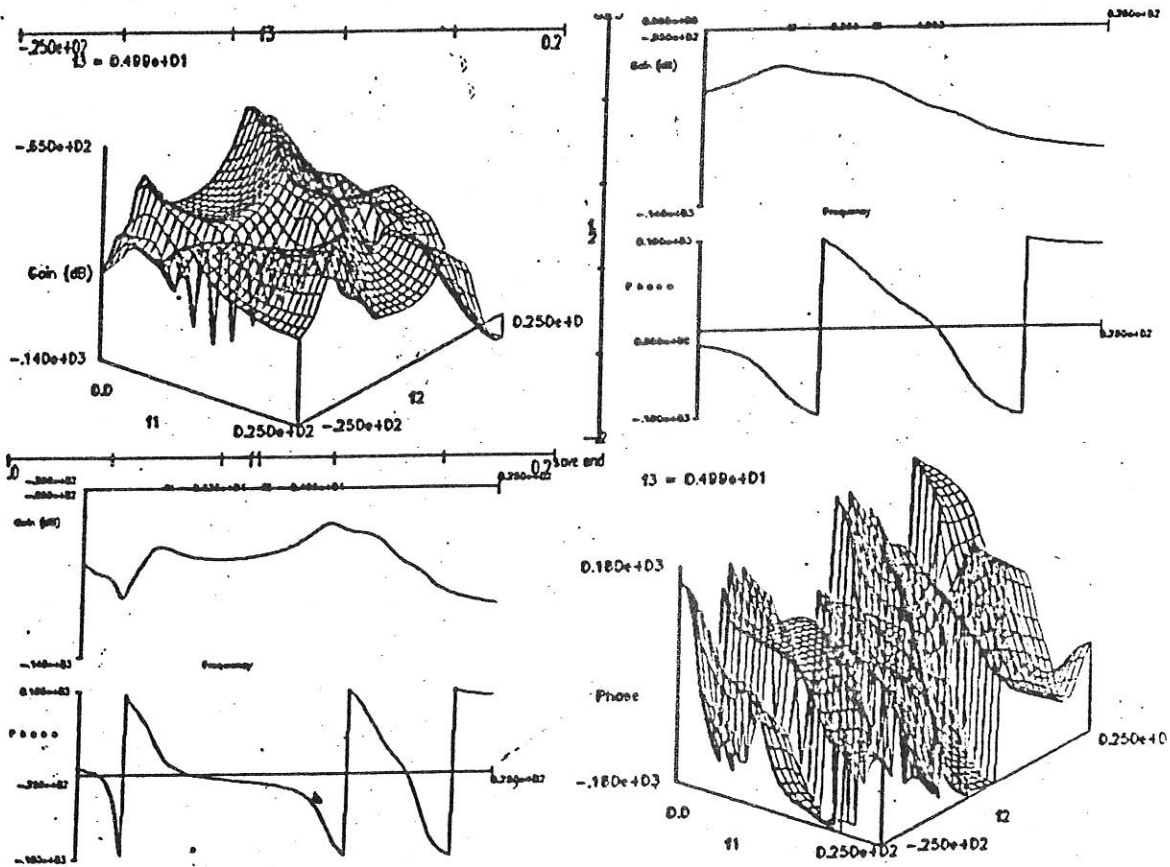


Fig.19. $H_3(\cdot, \cdot)$ $f_3 = 5\text{Hz}$

The system harmonics $H_1(j\omega)$, $H_2(j\omega, j\omega)$ (1st harmonic) and $H_3(j\omega, j\omega, j\omega)$ (2nd harmonic) can be plotted on one graph as illustrated in Fig.20.

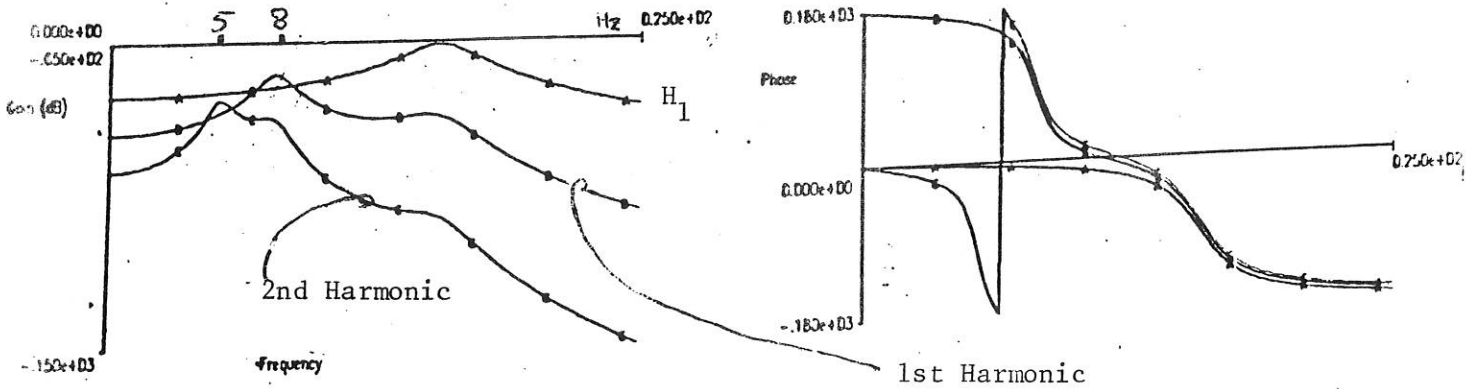


Fig.20. Harmonics

Inspection of Fig.20 shows that there is a resonant peak around 5Hz for the second harmonic and around 8Hz for the first harmonic. This can be verified by injecting an 8Hz sine wave into the estimated NARMAX model to illustrate for example how the system would respond to this input, Fig.21.

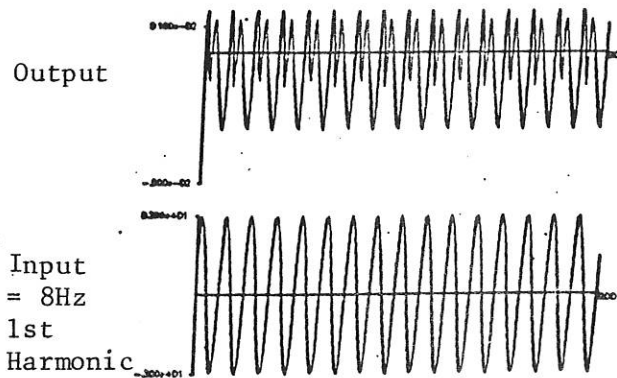


Fig.21

The advantages of the parametric approach based on the NARMAX model are therefore:-

- (a) $H_i(j\omega_1, j\omega_2, \dots)$ can be easily computed for all i
- (b) Because the method is based on estimating a NARMAX model it

is not input dependent, works for sensible record lengths and often results in a small parameter set

- (c) The evolution of the linearized spectra can be computed
- (d) The frequency response functions can be decomposed to show the contribution of $H_1(\cdot)$, $H_2(\cdot\cdot)$, harmonics etc
- (e) Once identified the system can be simulated in the time or frequency domain to show the response to any input.

Bibliography

- Norton, J.P.: An introduction to identification; Academic Press 1986.
- Billings, S.A., Voon, W.S.F.: A prediction error and stepwise regression estimation algorithm for nonlinear systems; Int.J.Control, 44, 803-822, 1986.
- Billings, S.A., Voon, W.S.F.: Correlation based model validity tests for nonlinear models; Int.J.Control, 44, 1986, 235-244.
- Leontaritis, I.J., Billings, S.A.: Experimental design and identifiability for nonlinear systems; Int.J.Systems Science, 1987, 18, 189-202.
- Billings, S.A., Voon, W.S.F.: Piecewise linear identification of nonlinear systems; Int.J.Control, 46, 1987, 215-235.
- Billings, S.A., Tsang, K.M.: Estimating higher order spectra; 5th Modal Anal. Conf., 1987.