Self-Tuning Control of Nonlinear Armax Models

by

K R Sales[+]

S A Billings[‡]

[+] Unilever Research, Port Sunlight Laboratory
Quarry Road East, Bebington, Wirral
Merseyside, UK

[‡] Department of Control Engineering,
University of Sheffield,
Sheffield, UK

Research Report 362

May 1989

SELF-TUNING CONTROL OF NONLINEAR ARMAX MODELS

K. R. SALES† and S. A. BILLINGS‡

A control weighted self-tuning minimum-variance controller with a nonlinear difference equation structure is described. An extended recursive least-squares estimation algorithm is employed to provide the adaptiveness. Performance analysis of the controller is discussed in terms of a cumulative loss function and high-order correlation functions of the system input, output and residual sequences. Simulation results from an experiment using a model identified from real a system are also provided.

1. Introduction

The need for self-tuning control essentially arises from the desire to control processes whose parameters are either unknown or slowly time-varying. The self-tuner has three main elements (*Clarke 1980*). There is a standard feedback law, given as a difference equation, which acts upon a set of measured values, (system output, set-point, control input, noise estimates), and which generates a new control input signal. An on-line recursive estimator monitors these values and calculates a set of parameter estimates for a prescribed structural model of the plant. The parameter estimates are then fed into a control design algorithm which computes the parameters of the feedback control law. Self-tuning control, which was originally developed by Åström and Wittenmark (*Åström and Wittenmark 1973*), and later extended to the Generalised Minimum-Variance Self-Tuning Controller of Clarke and Gawthrop (*Clarke, Cope and Gawthrop 1975; Clarke and Gawthrop 1975*), has now become very popular and widely applied in the process industries (*Harris and Billings 1985*).

In recent years several nonlinear self-tuners have been described. These have though been based on the Hammerstein (*Anbumani, Patnaik and Sarma 1981; Anbumani, Sarma and Patnaik 1981; Grimble 1982, 1983; Keviczky and Haber 1974; Keviczky and Vajk 1978; Keviczky, Vajk and Hethéssy 1979; Korpela and Koivo*

† Unilever Research, Port Sunlight Laboratory, Quarry Road East, Bebington, Wirral, Merseyside, UK.

‡ Department of Control Engineering, University of Sheffield, Sheffield, UK.

1982) and bilinear (*Dochain and Bastin 1984; Goodwin, McInnes and Long 1982; Goodwin and Sin 1984; Svoronos, Stephanopoulos and Aris 1981*) models. The failings of these lie in the specific nature of the Hammerstein and bilinear models. The Hammerstein model restricts the process to being modelled by a static input nonlinearity followed by a dynamic linear element. Although some of this work could be extended, quite trivially, for use with the general cascade model (*Billings and Fakhouri 1982*) it would still only be applicable to systems that could be characterized as seperate blocks of the form described. The Hammerstein model may occur naturally where a nonlinear actuator is used to control a linear process, though this has successfully been overcome in the past by putting a linearizing controller around the actuator. Anbumani, Patnaik and Sarma (*Anbumani, Patnaik and Sarma 1981; Anbumani, Sarma and Patnaik 1981*) successfully applied a minimum-variance self-tuner to a variety of Hammerstein models and used simulations to demonstrate the practicality of their method. However, they erroneously concluded that control weighting could not be applied to nonlinear minimum-variance self-tuning controllers, by attempting to weight a linear function of the control input. This was overcome by Grimble (*Grimble 1982, 1983*), who applied control weighting to a nonlinear function of the control input. Korpela and Koivo (*Korpela and Koivo 1982*) put a linear controller around a Hammerstein model by assuming that the internal state variable between the static input nonlinearity and the dynamic linear element could be calculated from the known or estimated nonlinear term. Keviczky and co-workers (*Keviczky and Haber 1974; Keviczky and Vajk 1978; Keviczky, Vajk and Hethéssy 1979*) considered minimum-variance extremum control of the Generalized Hammerstein model. The aim, in this instance, is to drive some physical variable to its extremum (minimum or maximum). However, well tuned stochastic minimum-variance controllers suffer from bias from the extremum and increased process output variance compared with the linear case (*Haber, Keviczky and Unbehauen 1987*).

Minimum-variance self-tuning regulation of a bilinear model was first reported by Svoronos, Stephanopoulos and Aris (*Svoronos, Stephanopoulos and Aris 1981*), who used simulation experiments to demonstrate the effectiveness of their work. Goodwin, McInnes and Long (*Goodwin, McInnes and Long 1982; Goodwin and Sin 1984*) described bilinear minimum-variance controllers for the control of waste water treatment and pH neutralization and showed the superiority of their methods (over linear minimum-variance and PID controllers) with practical applications.

This paper introduces a minimum-variance self-tuning algorithm based on the recently reported **NARMAX** model (*Billings and Leontaritis 1981; Leontaritis and Billings 1985*) and has a more general appeal than the work reviewed above. The relationship between the various model structures is discussed and their advantages and disadvantages outlined. Several practical problems are addressed and the superiority of the methodolgy over linear self-tuners is clearly demonstrated using a simulation experiment on a model identified from a practical process. The problems of model validation and performance analysis are reviewed using cumulative loss functions and correlation functions of the system input, output and residuals specifically developed for detection of bias in estimation algorithms structured to find the parameters of nonlinear models.

2. Nonlinear models - An overview

There have now been several mathematical model types proposed which have been applied with varying degrees of success to the identification and control of systems. One initial division which is traditionally made is into linear and nonlinear models, though linear models are merely a subset of the nonlinear ones with the degree of nonlinearity equal to unity. Another possible divsion is:

(i) Functional Series models (eg. Volterra series, Wiener series),

(ii) Difference Equation models (eg. Bilinear, NARMAX, Generalized Hammerstein),

(iii) Specialized Structure models (eg. block orientated models (Hammerstein, Wiener, General Cascade model)).

It is quite often possible to represent particular nonlinear structures in more than one of these divisions.

The functional series models have the advantage that explicit knowledge of the model order and dead time is not necessary but this is far outweighed by the fact that several hundred parameters would have to be estimated for a Volterra series model and possibly many times more for even low degrees of nonlinearity in the case of the Wiener series model. The resultant computational burden is heavy, especially as the probability of being able to define realistic parametric models from the results is low and prior knowledge of the system is not easily incorporated into that model. This is further exacerbated by the fact that some simple and common nonlinearites (deadzone, saturation, etc.) lead to complex kernals.

A common occurance with nonlinear systems is to have linear dynamic and nonlinear static elements connected in cascade. The general cascade block-orientated model (*Billings 1980; Billings and Fakhouri 1982*) allows for the two block types to be cascaded in any sequence as long as the model is representative of the system. Two commonly considered forms are the Hammerstein (consisting of a static nonlinear element followed by a linear dynamic element and Wiener (consisting of a dynamic linear element followed by a static nonlinearity) models.

Difference equation models are by far the most important class of model, for both linear and nonlinear systems, since they are the most widely implemented. This has arisen for several reasons:

(i) they arise naturally from the physical laws of the system,

(ii) they have relatively few parameters and are numerically easy to handle,

(iii) the model can be used to represent a broad class of nonlinearities using a manageable set of terms,

(iv) the coefficients of the model can be estimated using established parameter estimation algorithms, which are not dependent on specialized input signals,

(v) the model is in a form that may be readily used for implementation on a digital computer, especially relevant for control purposes.

The linear difference equation form appears as:

$$A(z^{-1})z(t) = B(z^{-1})u(t-k) + C(z^{-1})e(t) \tag{1}$$

where $z(t)$ is the system output, $u(t)$ a controllable input signal, $e(t)$ a zero-mean Gaussian noise sequence, $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ polynomials in the in the backward shift operator, z^{-1} and k the system delay.

In the last twenty years or so the bilinear model has received a great deal of attention (*Brockett 1972; Bruni, Di Pillo and Koch 1974; Mohler 1973*) for the reasons given above. In difference equation notation it takes the form:

$$A(z^{-1})z(t) = \{B(z^{-1}) + \sum_{i=0}^{\infty}[S_i(z^{-1})z(t-k)]z^{-i}\}u(t-k) + C(z^{-1})e(t) \tag{2}$$

where $S_i(z^{-1})$ is a parameter polynomial and otherwise the notation is as above.

A non-seperable form of the Hammerstein model, the Generalized Hammerstein model has been studied by Báyánz, Haber and Keviczky (*Báyánz, Haber and Keviczky 1973*) and is given by the equation:

$$A(z^{-1})z(t) = B_1(z^{-1})u(t-k) + B_2(z^{-1})u^2(t-k) + ..$$

$$..+ B_n u^n(t-k) + C(z^{-1})e(t) \qquad (3)$$

Sontag (*Sontag 1976, 1979*) has dealt in considerable detail with the realization of polynomial response functions. Finite realization implies that the function is a rational difference equation of the form:

$$A_f[y(t-1) ,..., y(t-p);u(t) ,..., u(t-p)]y(t) = B_f[y(t-1) ,..$$

$$.., y(t-p);u(t) ,..., u(t-p)] \qquad (4)$$

where A_f and B_f are polynomials of finite degree. If the response function is bounded and finitely realizable then a special form of it occurs, the output-affine difference equation which is given by:

$$\sum_{i=0}^{p} A_i[u(t) ,..., u(t-p)]y(t-i) = A_{p+1}[u(t) ,..., u(t-p)] \qquad (5)$$

where the A_i are polynomials.

This may also be represented as a finite Volterra series (*Leontaritis and Billings 1985*).

The output-affine model has a more general appeal than the above models and is globally valid, but its main drawback is that it is not linear-in-the-parameters, and hence does not lend itself to simple recursive estimation procedures (*Chen and Billings 1987*). The state space representation of the output-affine model, has successfully been used in the identification of Electrical Power Plants (*Cyrot-Normand and Dang van Mien 1980; Dang van Mien and Cyrot-Normand 1984*), by patching together a series of linear signal dependent state space models. Both a prediction error estimation algorithm and a recursive maximum liklihood estimator have beeen derived for the output affine

model (*Chen and Billings 1988a, 1988b*).

This paper is particularly concerned with control using a new nonlinear structure, the **NARMAX** model (*Billings and Leontaritis 1981, 1982; Billings and Voon 1983; Chen and Billings 1989*). Leontaritis and Billings (*Leontaritis and Billings 1985*) have shown that a suitable representation for a broad class of nonlinear systems is given by:

$$y(t) = F[y(t-1) \ ,..., \ y(t-n_y);u(t-k) \ ,..., \ u(t-k-n_u+1)] \tag{6}$$

where y(t) is the system noise-free output, n_u and n_y the orders of u(t) and y(t) respectively and F[•] some nonlinear function of u(i) and y(i). The Hammerstein, Wiener, bilinear and several other well-known nonlinear structures can be shown to be special cases of the NARMAX model (*Billings and Leontaritis 1981, 1982*).

Expanding F[•] as a polynomial function and defining

$$V_1 = y(t-1) \ ,..., \ V_{n_y} = y(t-n_y),$$

$$V_{n_y+1} = u(t-k) \ ,..., \ V_{n_y+n_u} = u(t-k-n_u+1)$$

then

$$y(t) = F[V_1,V_2 \ ,..., \ V_{n_y+n_u}] \tag{7}$$

and the polynomial expansion is:

$$y(t) = \sum_{i=1}^{n_s}\theta_i V_i + \sum_{i=1}^{n_s}\sum_{j=1}^{n_s}\theta_{ij}V_iV_j + \ ... \ + \sum_{i=1}^{n_s}\sum_{j=1}^{n_s} \ ,..$$
$$.., \sum_{m=1}^{n_s}\sum_{n=1}^{n_s}\theta_{ij \ ,..., \ mn}V_iV_j \ ,..., \ V_mV_n \tag{8}$$

where $n_s = n_y + n_u$

In the identification of linear systems it is assumed that any internal noise can be translated to an output noise source. This does not hold for nonlinear systems and even output additive noise produces multiplicative terms. In order to introduce noise to the model assume that the output is corrupted by

a zero-mean sequence, e(t), such that:

$$z(t) = y(t) + e(t) \tag{9}$$

Substituting equation (9) into equation (8) and rewriting in the form of equation (6) gives:

$$z(t) = F^l[z(t-1) \ ,..., \ z(t-n_y);u(t-k) \ ,..., \ u(t-k-n_u+1);e(t-1) \ ,..$$

$$.., \ e(t-n_y)] + e(t) \tag{10}$$

where l is the degree of nonlinearity.

As the superposition principle does not apply to nonlinear systems the internal noise cannot be translated to be additive at the output. Hence, a more realistic formulation is:

$$z(t) = F^l[z(t-1) \ ,..., \ z(t-n_y);u(t-k) \ ,..., \ u(t-k-n_u+1);e(t),e(t-1) \ ,..$$

$$.., \ e(t-n_y)] \tag{11}$$

In practice, however, equation (10) can be used to represent a wide class of nonlinear systems.

The applicability of the block-orientated and bilinear models is more limited than that of the output-affine and NARMAX models, because of the greater generality of the latter two. One additional failing of the bilinear model is that although bilinearity often occurs naturally in continuous time (*Mohler 1973*), upon discretization the model becomes output-affine or NARMAX (*Fleiss and Cyrot-Normand 1982; Fraiech and Ljung 1987; Chen and Billings 1989*). A first order approximation will yield a discrete bilinear model (*Goodwin, McInnes and Long 1982*), but this may introduce serious approximations. Because the NARMAX model is linear-in-the-parameters, it lends itself easily to the well established and easily implementable recursive estimation methods and hence provides a natural basis upon which to develop nonlinear self-tuning controllers.

3. A control algorithm for the NARMAX model.

From equation (10), the NARMAX model is defined as:

$$z(t) = F^l[z(t-1)\ ,...,\ z(t-n_z);u(t-k)\ ,...,\ u(t-k-n_u+1);e(t-1)\ ,..$$

$$..,\ e(t-n_e)] + e(t) \qquad (12)$$

$$= F^l[\bullet,(t)]$$

where n_z and n_e are the orders of z(t) and e(t) respectively, and are introduced for convenience in practical applications.

The basic minimum-variance controllers are designed to satisfy the criterion (*Keviczky, Vajk and Hethéssy 1979*):

$$J_2(t-k) = E[(z(t+k) - w(t))^2] \qquad (13)$$

where z(t) is the system output, w(t) the demand input and $E[\bullet]$ the expectation operator.

This criterion minimises the variance of the error in the output. Unfortunately, in practical applications, satisfaction of such a criteria often requires a large control action that may result in damage to plant equipment, excessive expense or operation of the model outside its validity bound, particularly in the early stages of parameter tuning. The Generalized Minimum-Variance controller (*Clarke, Cope and Gawthrop 1975; Clarke and Gawthrop 1975, 1979*) overcomes this by including a weighting on the control input within the cost function. When the GMV cost function is related to the NARMAX model the following cost function is considered:

$$J_3(t-k) = E[P(z(t+k) - Rw(t))^2 + (Q'f(u(t)))^2] \qquad (14)$$

where P, R and Q´ are polynomials in the backward shift operator z^{-1}, and f(u(t)) is a function that includes all the terms in u(t) that are in the model, ie:

$$f(u(t)) = F^l[z(t+k-1)\ ,...,\ z(t+k-n_z);u(t)\ ,...,\ u(t-n_u+1);e(t+k-1)\ ,..$$

$$..,\ e(t+k-n_e)] - F^l[z(t+k-1)\ ,...,\ z(t+k-n_z);u(t-1)\ ,..$$

$$..,\ u(t-n_u+1);e(t+k-1)\ ,...,\ e(t+k-n_e)] \qquad (15)$$

The need for the weighting to be on a function such as f(u(t)) arises from the fact that if the weighting is solely on a linear function of u(t) then the estimation does not produce parameter estimates expected for minimum-variance control (*Anbumani, Patnaik and Sarma 1981; Grimble 1982*), (proof of this is given in Appendix A1).

Expanding equation (14):

$$J_3(t-k) = E[(PF^l[z(t+k-1)\ ,...,\ z(t+k-n_z);u(t)\ ,...,\ u(t-n_u+1);e(t+k-1)\ ,..$$

$$..,\ e(t+k-n_e)] - Rw(t) + Pe(t+k))^2 + (Q'f(u(t)))^2] \qquad (16)$$

e(t) is unknown and unpredictable, however, the only element of Pe(t+k) that is uncorrelated with other terms in $F^l[\bullet,(t+k)]$ is $p_0e(t+k)$. Hence,

$$J_3(t-k) = E[(PF^l[z(t+k-1)\ ,...,\ z(t+k-n_z);u(t)\ ,...,\ u(t-n_u+1);e(t+k-1)\ ,..$$

$$..,\ e(t+k-n_e)] - Rw(t) + (P-p_0)e(t+k))^2 + (Q'f(u(t)))^2]$$

$$+ E[p_0(e(t+k))^2] \qquad (17)$$

Differentiating equation (17) with respect to u(t) to find the new control action gives:

$$\frac{dJ_3(t-k)}{du(t)} = E[(PF^l[\bullet,(t+k)] + (P-p_0)e(t+k) - Rw(t))\ 2\ \frac{dF^l[\bullet,(t+k)]}{du(t)}$$

$$+ Q'f(u(t))\ 2q_0'\ \frac{df(u(t))}{du(t)}] = 0 \qquad (18)$$

Since

$$\frac{dF^l[\bullet,(t+k)]}{du(t)} = \frac{df(u(t))}{du(t)}$$

the control law becomes:

$$E[PF'[\bullet,(t+k)] + (P-p_0)e(t+k) - Rw(t) + Q'q_0'f(u(t))] = 0 \qquad (19)$$

which may be written in terms of a new polynomial, $F''[\bullet,(t+k)]$, as:

$$E[F''[z(t+k-1) ,..., z(t+k-n_1);u(t) ,..., u(t-n_2);e(t+k-1) ,..$$

$$.., e(t+k-n_3)] - Rw(t)] = 0 \qquad (20)$$

where $n_1 = n_z + n_p$

$n_2 = n_u + n_p - 1$ or $n_u + n_q - 1$ whichever is the greater

$n_3 = n_e + n_p$

Now, if $k > 1$, then there will be future, unknown terms $z(t+k-d)$ and $e(t+k-d)$, where $d = 1, 2, ..., k-1$. The future terms in $z(t)$ may be predicted but $e(t)$ is an unknown random sequence and hence is unpredictable. These values are set to zero; their conditional mean. Thus, equation (20) reduces to:

$$E[F''[z(t+k-1) ,..., z(t+k-n_1);u(t) ,..., u(t-n_2);e(t) ,..$$

$$.., e(t+k-n_3)] - Rw(t)] \qquad (21)$$

Clarke and Gawthrop (*Clarke and Gawthrop 1975, 1979*) showed that minimization of the GMV cost function is equivalent to minimising the variance of an auxiliary output function, $\Phi(t)$. The same can be shown in the NARMAX case:

$$\Phi(t+k) = Pz(t+k) - Rw(t) + Qf(u(t)) \qquad (22)$$

where $Q = Q'q_0'$.

Consider

$$J_4(t+k) = E[(\Phi(t+k))^2] \qquad (23)$$

$$= E[(Pz(t+k) - Rw(t) + Qf(u(t)))^2] \qquad (24)$$

$$= E[(PF^l[z(t+k-1)\ ,...,\ z(t+k-n_z);u(t)\ ,...,\ u(t-n_u+1);e(t+k-1)\ ,..$$

$$..,\ e(t+k-n_e)] + Pe(t+k) - Rw(t) + Qf(u(t)))^2] \qquad (25)$$

The only element in Pe(t+k) that is not correlated with terms in $F^l[\bullet,(t+k)]$ is $p_0e(t+k)$. Thus equation (25) becomes:

$$J_4(t+k) = E[(PF^l[z(t+k-1)\ ,...,\ z(t+k-n_z);u(t)\ ,...,\ u(t-n_u+1);$$

$$e(t+k-1)\ ,...,\ e(t+k-n_e)] + (P-p_0)e(t+k) - Rw(t)$$

$$+ Qf(u(t)))^2] + E[(p_0e(t+k))^2] \qquad (26)$$

Differentiating equation (26) with respect to u(t) gives:

$$\frac{dJ_4(t+k)}{du(t)} = E[PF^l[z(t+k-1)\ ,...,\ z(t+k-n_z);u(t)\ ,...,\ u(t-n_u+1);$$

$$e(t+k-1)\ ,...,\ e(t+k-n_e)] + (P-p_0)e(t+k) - Rw(t)$$

$$+ Qf(u(t))]\ 2(1+q_0)\ \frac{dF^l[\bullet,(t+k)]}{du(t)} = 0 \qquad (27)$$

Adapting the notation of equation (27) to that of equation (21), the two can be seen to be equivalent. The auxiliary output, $\Phi(t)$, in the linear case is physically realizable, however, in the NARMAX case f(u(t)) is not extractable as a unique function and hence, although it is feasible to minimise the auxiliary output it cannot be measured directly from the system.

4. Self-tuning aspects.

The algorithm may be made self-tuning by employing an extended recursive least squares estimator (*Goodwin and Payne 1977; Billings and Voon 1984*) to directly estimate the controller parameters. The data and parameter estimate vectors may be constructed thus:

$$\Phi(t/t-k) = \Phi^*(t/t-k) + e(t) \qquad (28)$$

where $\Phi^*(t/t-k)$ is the optimal prediction of $\Phi(t)$ at time (t-k). Hence,

$$\Phi^*(t/t\text{-}k) = F^{-l}[z(t\text{-}1)\ ,...,\ z(t\text{-}n_1);u(t\text{-}k)\ ,...,\ u(t\text{-}k\text{-}n_2);e(t\text{-}1)\ ,..$$

$$..,\ e(t\text{-}n_3)] - Rw(t) \tag{29}$$

$$= \chi^T(t)\ \hat{\theta}(t) \tag{30}$$

where $\chi^T(t)$ is the data vector and $\hat{\theta}(t)$ the parameter estimate vector at time t.

$$\chi^T(t) = [z(t\text{-}1)\ ,...,\ z(t\text{-}n_1);u(t\text{-}k)\ ,...,\ u(t\text{-}k\text{-}n_2);e(t\text{-}1)\ ,...,\ e(t\text{-}n_3);w(t)\ ,...,\ w(t\text{-}n_r)$$

$$z^2(t\text{-}1)\ ,..\qquad z(t\text{-}1)u(t\text{-}k)\ ,..\qquad z(t\text{-}1)e(t\text{-}1)\ ,..$$

$$....., \text{higher terms}]$$

$$\hat{\theta} = [\hat{\theta}_1\ ,..\quad \hat{\theta}_{n_1},\quad \hat{\theta}_{n_1+1}\ ,..\quad \hat{\theta}_{n_1+n_2},\quad \hat{\theta}_{n_1+n_2+1}\ ,..\ \hat{\theta}_{n_1+n_2+n_3},\ r_0\ ,....,\ r_{n_r}$$

$$\hat{\theta}_{11}\ ,..\qquad \hat{\theta}_{1n_1+1}\ ,..\qquad \hat{\theta}_{1n_1+n_2+1}\ ,..$$

$$....., \text{higher terms}]$$

The optimal control law is given by equation (20), however, it is not practically implementable since the future noise terms are unknown and unpredictable. Hence, the control law is reduced to equation (21). Note though that the estimator finds parameter estimates for all the terms in equation (20). Solely estimating the parameters of equation (21) would result in bias in the estimator, since it would attempt to force the information contained in the missing terms into the structure it is estimating.

5. Calculation of the noise estimates.

With linear systems the noise is very often disregarded if the signal to noise ratio is high. However, with nonlinear systems where the noise is internally additive it may contribute a more significant part of the output signal and should in general be estimated and included in the control law where possible. The noise terms are unknown, but may estimated at each iteration. By considering equations (28) and (30) an estimate may be found:

$$\hat{e}(t) = \Phi(t) - \chi(t-k)\ \hat{\theta}(t-k)$$

Note that this is actually the system residual.

Substituting the above result into the control law, equation (21), gives a new control law:

$$F'^{l}[z(t+k-1)\ ,...,\ z(t+k-n_1);u(t)\ ,...,\ u(t-n_2);\hat{e}(t+k-1)\ ,..$$

$$..,\ \hat{e}(t+k-n_3)] - Rw(t) = 0 \qquad (31)$$

6. Prediction of the future output terms.

The control law of equation (31) contains future unknown output terms in $z(t+k-1)\ ,...,\ z(t+1)$ that occur if $k > 1$. These terms may be proxied by their predictions. The predictions can be found by employing a sequential algorithm based on setting the predictions of the auxiliary output to zero (the ultimate goal of the controller and its conditional mean upon convergence of the parameter estimates) and knowledge of the current parameter estimates.

From equation (22):

$$\Phi(t) = Pz(t) - Rw(t-k) + Qf(u(t-k))$$

$$= F'^{l}[z(t-1)\ ,...,\ z(t-n_1);u(t-k)\ ,...,\ u(t-k-n_2);\hat{e}(t-1)\ ,..$$

$$..,\ \hat{e}(t-n_3)] - Rw(t-k) + \hat{e}(t)$$

The prediction of $\Phi(t+1)$, $\Phi^*(t+1)$, may be derived and equated to zero.

$$\Phi^*(t+1) = F'^{l}[z(t)\ ,...,\ z(t-n_1+1);u(t-k+1)\ ,...,\ u(t-k-n_2+1);\hat{e}(t)\ ,..$$

$$..,\ \hat{e}(t-n_3+1)] - Rw(t-k+1) = 0 \qquad (32)$$

Extending this one step further ahead $\Phi_p(t+2)$ may be derived:

$$\Phi^*(t+2) = F^{-1}[z_p(t+1) \ ,..., \ z(t-n_1+2);u(t-k+2) \ ,..., \ u(t-k-n_2+2);\hat{e}(t) \ ,..$$

$$.., \ \hat{e}(t-n_3+2)] - Rw(t-k+2) = 0 \qquad (33)$$

where $z_p(t+1)$ is the prediction of $z(t+1)$ and may be found by solution of equation (33).

The algorithm may be stepped ahead as many times as are necessary in order to predict the required values of $z_p(t+d)$, where $d = 1, 2, ..., k-1$. Clearly the accurracy of the predictions will decrease the further ahead they are predicted since future noise terms, or their estimates, have to be omitted from the calculations.

Using the above algorithm may generate complex solutions if the degree of nonlinearity of $z_p(t+d)$ is even. A solution to this is to use an interpretation of equation (22). $z_p(t+d)$ may also be found from:

$$\Phi^*(t+d) = Pz_p(t+d) - Rw(t-k+d) + Qf(u(t-k+d)) = 0 \qquad (34)$$

Finding $z_p(t+d)$ from equation (34) implies that the function $\Phi(t+d)$ will be minimised at the next iteration and that the parameter estimates will have converged, whereas equations (33) and (34) take account of the current knowledge that the estimator has of the parameters.

7. Performance analysis.

Performance analysis is concerned with assessing the "well-being" of the controller and the estimator. If the controller is performing satisfactorily and the parameter estimates have converged to their true values then the system output should be close to the desired value, (there will always be a slight discrepancy due to the unknown noise terms). One common and simple test for continually monitoring this difference is a cumulative loss function. This is generally recorded as the sum of the squares of the errors. Large steps in the function would suggest that something abnormal has occurred. Although simple, this test gives little insight as to why a failure in the system has occurred or any assurance that the controller will continue to perform satisfactorily. Model validation has the ability to point the analyst at the cause of a failure. Model validity tests, as their name implies, check the validity of a model for a

system, but more than that, they can indicate the structure of the terms, (ie. are they input/output or noise model terms, are they even or odd ordered), that are missing or whose parameter estimates have not converged. If the model is in error bias in the residuals occurs; it is this bias that model validity tests detect. Bias means that future "good control" cannot be guaranteed - the control loop may suddenly "blow-up", even though it has been running for some time giving seemingly good control.

For linear control loops the criteria to be satisfied are given by (*Clarke and Gawthrop 1975*):

$$\left.\begin{aligned} E[\hat{e}(t),\ \hat{e}(t+\tau)] &= \phi_{\hat{e}\hat{e}}(\tau) = \delta(\tau) \\ E[u(t),\hat{e}(t+\tau)] &= \phi_{u\hat{e}}(\tau) = 0 \quad \forall\ \tau \geq k \\ E[z(t),\hat{e}(t+\tau)] &= \phi_{z\hat{e}}(\tau) = 0 \quad \forall\ \tau \geq k \\ E[w(t),\hat{e}(t+\tau)] &= \phi_{w\hat{e}}(\tau) = 0 \quad \forall\ \tau \end{aligned}\right\} \tag{35}$$

In model validation Billings and Voon (*Billings and Voon 1983*) have shown that the conventional linear correlation functions do not necessarily detect all possible nonlinear terms and hence have defined a set of higher order correlation functions suitable for use with the NARMAX model. Clearly their arguments will apply to the residuals in this case; the same features are being sought. However, the system under analysis is in a closed-loop and hence the range for τ is as above, equations (35). In summary, for a system with internally additive noise, the controller can be said to performing satisfactorily if and only if:

$$\left.\begin{aligned} &\phi_{\hat{e}\hat{e}}(\tau) = \delta(t) \\ &\phi_{u\hat{e}}(\tau),\ \phi_{\hat{e}(\hat{e}u)}(\tau),\ \phi_{u^{2}\hat{e}}(\tau),\ \phi_{u^{2}\hat{e}^{2}}(\tau),\ = 0 \quad \forall\ \tau \geq k \\ &\phi_{z\hat{e}}(\tau),\ \phi_{\hat{e}(\hat{e}z)}(\tau),\ \phi_{z^{2}\hat{e}}(\tau),\ \phi_{z^{2}\hat{e}^{2}}(\tau),\ = 0 \quad \forall\ \tau \geq \kappa \\ &\phi_{w\hat{e}}(\tau) = 0 \quad \forall\ \tau \end{aligned}\right\} \tag{36}$$

Higher order correlation functions in w(t) are not necessary as w(t) enters the system linearly.

8. Simulation results.

The results of the simulation experiment discussed below have been performed on a model identified from a practical system. The initial parameter estimates were preset to zero, with the exception of the estimate associated with w(t) in the control law; this is set to 1.0 as is consistent with the work on linear systems (*Clarke and Gawthrop 1975*), and the diagonal of the covariance matrix to 10^4, which implies that there is little knowledge of the parameter values. The experiment was run over 1000 samples and the model validation correlation functions were computed from the final 500 samples, allowing time for parameter convergence. This implies that the 95% confidence bound for the correlation plots will be 8.76×10^{-2}.

Graphical plots of the demand input, w(t), control input, u(t), system output, z(t), residuals, $\hat{e}(t)$, the actual and theoretical loss functions, the parameter estimates and the diaganal of the covariance matrix are presented and discussed.

The model used in this experiment was identified from a laboratory scale liquid level system. The system consists of a DC water pump feeding a conical flask which in turn feeds a square tank, giving the system second order dynamics. The controllable input is the voltage to the pump motor and the system output is the height of water in the conical flask. The aim, under simulation conditions, is for the water height to follow some demand signal.

The plant model was identified as:

$$\begin{aligned} z(t) = {} & 0.9722z(t-1) + 0.3578u(t-1) - 0.1295u(t-2) \\ & - 0.3103z(t-1)u(t-1) - 0.04228z^2(t-2) + 0.1663z(t-2)u(t-2) \\ & + 0.2573z(t-2)e(t-1) - 0.03259z^2(t-1)z(t-2) - 0.3513z^2(t-1)u(t-2) \\ & + 0.3084z(t-1)z(t-2)u(t-2) + 0.2939z^2(t-2)e(t-1) \\ & + 0.1087z(t-2)u(t-1)u(t-2) + 0.4770z(t-2)u(t-1)e(t-1) \\ & + 0.6389u^2(t-2)e(t-1) + e(t) \end{aligned} \qquad \text{(S1.1)}$$

Setting $P(z^{-1})$ and $R(z^{-1})$ to 1.0, for simplicity, $Q(z^{-1})$ to 0.2 and using

the derivation of §3. the following controller results:

$$0.9722z(t) + 0.4293u(t) - 0.1295u(t-1) - 0.3724z(t)u(t)$$
$$- 0.04228z^2(t-1) + 0.1663z(t-1)u(t-1) + 0.2573z(t-2)\hat{e}(t-1)$$
$$- 0.03259z^2(t-1)z(t-2) - 0.3513z^2(t-1)u(t-2) + 0.3084z(t)z(t-1)u(t-1)$$
$$+ 0.2939z^2(t-1)\hat{e}(t) + 0.1304z(t-1)u(t)u(t-1) + 0.5724z(t-1)u(t)\hat{e}(t)$$
$$- 0.6389u^2(t-1)\hat{e}(t) - w(t) = 0 \qquad \text{(S1.2)}$$

The noise sequence, e(t), was Gaussian and given as $N(0.0, 0.05)$.

Figures 1.1-4 give the input-output data for a typical run. The output, z(t), quickly tracks the demand, w(t), with some slight attentuation due to the effect of the control weighting. The actual cumulative loss function, figure 1.5, and the theoretical loss function, figure 1.6, (the cumulative sum of the noise sequence), become near parallel within 100 samples, implying that the residual sequence is approximately equal to the noise sequence. The process model parameters, figures 2.1-6, 2.8-10 and 2.12 have tuned well, and are close to their expected values given in equation S1.2. Figures 2.7, 2.11 and 2.13-14 are for the noise model and as with linear systems these take longer to converge, though none of them is radically in error. The model validation correlation functions, figures 3.1-10, are extremely encouraging. There are two points which are outside the 95% confidence bound - $\phi_{\hat{e}(\hat{e}u)}(\tau)$ at $\tau = 1$ and $\phi_{u^2\hat{e}^2}(\tau)$ at $\tau = 6$ - though these errors are slight and indicate that the noise model parameters have not fully tuned.

A "best fit" linear model of the liquid level system has also been identified. This model was used to design a linear GMV controller. Using the linear controller to control the nonlinear model resulted in the estimation blowing up within 20-30 samples, and no control being achieved. Presetting the parameters estimates of the linear controller to their desired values still failed to prevent loss of control.

9. Conclusion.

A control weighted self-tuning minimum-variance controller with a nonlinear structure has been derived. Discussion has shown that the NARMAX based controller is more generally applicable than those nonlinear controllers that have so far been reported, and that using the NARMAX structure is a more

practicable approach than using functional series or block structured models. The superiority of the nonlinear structure is shown by the failure of the linear GMV to provide any control of the nonlinear models in the simulation experiments. The use of the high order correlation functions allows for a quantative assessment of the performance of the nonlinear controller by detecting bias in the residuals. Though used statically in these simulations the correlation functions may easily be updated to provide continuous on-line assessment.

Appendix A1.

Anbumani, Patnaik and Sarma (*Anbumani, Patnaik and Sarma 1981*) showed that control weighting of a linear function of u(t) was not possible. Here their result is reworked for the NARMAX model based controller.

If a linear control function were to be weighted, equation (14) would become:

$$J_3(t+k) = E[(Pz(t+k) - Rw(t))^2 + (Q'u(t))^2] \qquad A(1.1)$$

Expanding equation (A1.1):

$$J_3(t+k) = E[(PF^l[z(t+k-1) \ ,..., \ z(t+k-n_z);u(t) \ ,..., \ u(t-n_u-1);e(t+k-1) \ ,..$$

$$.., \ e(t+k-n_e)] - Rw(t) + Pe(t+k))^2 + (Q'u(t))^2] \qquad (A1.2)$$

Again $p_0e(t+k)$ is the only term in (A1.2) that is uncorrelated with other terms. Now, differentiating with respect to u(t) to find the new control action gives:

$$\frac{dJ_3(t+k)}{du(t)} = E[(PF^l[\bullet,(t+k)] + (P-p_0)e(t+k) - Rw(t)) \ 2 \ \frac{dF^l[\bullet,(t+k)]}{du(t)}$$

$$+ 2Q'q'_0u(t)] = 0 \qquad (A1.3)$$

The recursive least-squares estimator is of the form (*Soderstrom, Ljung and Gustavsson 1978; Billings and Voon 1984*):

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\varepsilon(t) \qquad \text{(A1.4)}$$

where

$$\varepsilon(t) = \Phi(t) - \chi^T(t)\hat{\theta}(t-1) \qquad \text{(A1.5)}$$

Thus, for equations (A1.3) and (A1.5) to have the same form $\frac{F^l[\bullet,(t+k)]}{du(t)}$ must equal unity, which cannot be guaranteed.

References

Anbumani, K., Patnaik, L.M. and Sarma, I.K., 1981, Self-tuning minimum-variance control of nonlinear systems of the Hammerstein model. IEEE Trans. on Automatic Control, Vol AC-26, No. 4, pp959-961.

Anbumani, K., Sarma, I.K. and Patnaik, L.M. 1981, Self-tuning control of nonlinear systems characterized by the Hammerstein model. 8th IFAC World Triennial Congress, Kyoto, Japan, pp79-84.

Åström, K.J and Wittenmark, B. 1973, On self-tuning regulators, Automatica, Vol. 9, pp185-199.

Báyánz, Cs. Haber, R. and Keviczky, L. 1973, Some estimation methods for nonlinear discrete-time identification. 3rd IFAC Symp. on Sys. Ident., Deflt, Holland.

Billings, S.A. 1980, Identification of nonlinear systems - A survey. IEE Proc., Vol. 127, Pt. D, No.6, pp272-284.

Billings, S.A. and Fakhouri, S.Y. 1982, Identification of systems composed of linear dynamic and static nonlinear elements. Automatica, Vol. 18, pp15-26.

Billings, S.A. and Leontaritis, I.J. 1981, Identification of nonlinear systems using parameter estimation techniques. Proc. IEE Conf. 194: Control and Its Applications, University of Warwick, pp183-187.

Billings, S.A. and Leontaritis, I.J. 1982, Parameter estimation techniques for nonlinear systems. Proc. of the 6th IFAC Symp. on Ident. and Sys. Par. Est., Washington DC, USA. pp505-510.

Billings, S.A. and Voon, W.S.F. 1983, Structure detection and model validity tests in the identification of nonlinear systems. IEE Proc., Vol. 130, Pt. D, No. 4, pp193-199.

Billings, S.A. and Voon, W.S.F. 1984, Least squares parameter estimation algorithms for non-linear systems. Int. J. Sys. Sci., Vol. 15, No. 6, pp601-615

Brockett, S.A. 1972, On the algebraic structure of bilinear systems in "Theory and Application of Variable Structure Systems" by Mohler, R.R. and Ruberti, A. (Eds), Acedemic Press.

Bruni, C., Di Pillo, G. and Koch, G. 1974, Bilinear systems: An appealing class of nearly linear systems in theory and application. IEEE Trans, on Automatic Control, Vol. AC-19, pp334-348.

Chen, S. and Billings, S.A. 1988a, Prediction-error estimation algorithm for non-linear output-affine systems. Int. J. Control, Vol. 47, No. 1, pp309-322.

Chen, S. and Billings, S.A. 1988b, Recursive maximum liklihood identification of a nonlinear output-affine model. Int. J. Control, Vol. 48, No. 4, pp1605-1629.

Chen, S. and Billings, S.A. 1989, Representations of nonlinear systems: the NARMAX model. Int. J. Control, Vol. 49, No 3, pp1013-1032.

Clarke, D.W. 1980, Some implementation considerations for self-tuning controllers in 'Numerical Techniques for Stochastic Control' by Archetti, F. and Cugiani, M. (Eds).

Clarke, D.W., Cope, S.N. and Gawthrop, P.J. 1975, Feasibilty study of the application of microprocessors to self-tuning controllers. Tech. Rep. 1137/75, Dept of Engineering Science, University of Oxford.

Clarke, D.W. and Gawthrop, P.J. 1975, Self-tuning controller. IEE Proceedings, Vol 122, No 9, pp929-934.

Cyrot-Normand, D. and Dang van Mien, H. 1980, Nonlinear state-affine identifcation methods; application to electical power plants. Proc. of IFAC Symp. on Automatic Control in Power Generation, Distibution and Protection, Pretoria, South Africa. pp449-462.

Dang van Mien, H. and Cyrot-Normand, D. 1984, Nonlinear state-affine identification methods; applications to electrical power plants. Automatica, Vol. 20, No. 2, pp175-188.

Dochain, D. and Bastin, G. 1984, Adaptive identification and control algorithms for nonlinear bacterial growth systems. Automatica, Vol. 20, No. 5, pp175-188.

Fliess, M. and Cyrot-Normand, D. 1982, On the approximation of nonlinear systems by some simple state-space models. Proc. of the 6th IFAC Symp. on Ident. and Sys. Par. Est., Washington DC, USA. pp511-514.

Fraiech, F. and Ljung, L. 1987, Recursive identification of bilinear systems. Int. J. Control, Vol. 45, No. 2, pp453-470.

Goodwin, G.C., McInnes, B. and Long, R.S. 1982, Adaptive control algorithms for wastewater treatment and pH neutralization. Optimal Control Applications and Methods, Vol. 3, pp443-459.

Goodwin, G.C. and Payne, R.L. 1977, Dynamic System Identification - Experiment Design and Data Analysis. Acedemic Press.

Goodwin, G.C. and Sin, K.S. 1984, Adaptive Filtering, Prediction and Control. Prentice Hall.

Grimble, M.J. 1982, Weighted minimum-variance self-tuning control. Int. J. Control, Vol. 36, No. 4, pp597-606.

Grimble, M.J. 1983, Weighted minimum-variance control of nonlinear systems. Research Report ICU/13/November 1982, Dept of Electronic and Electrical Engineering, University of Strathclyde.

Haber, R., Keviczky, L. and Unbehauen, H. 1987, Applications of adaptive control on nonlinear dynamic processes - a survey of input-output approaches. Proc. of the 7th IFAC Symp. on Ident. and Sys. Par. Est., York, UK.

Harris, C.J. and Billings, S.A. 1985, Self-Tuning and Adaptive Control: Theory and Applications. 2nd Edition, Peter Perigrinus.

Keviczky, L. and Haber, R. 1974, Adaptive dual extremum control by the Hammerstein model. IFAC Symp. on Stochastic Control, Budapest, Hungary. pp333-341.

Keviczky, L. and Vajk, I. 1978, A self-tuning extremum regulator. Conf. Record of the 12th Asilomar Conf. on Circuits, Systems and Computers, Pacific Grove, CA, USA. pp69-72.

Keviczky, L., Vajk, I. and Hethéssy, J. 1979, A self-tuning extremal controller for the generalized Hammerstein model. Proc. of the 5th IFAC Symp. on Ident. and Sys. Par. Est., Darmstadt, West Germany. pp1147-52.

Korpela, K. and Koivo, H.N. 1982, On self-tuning control of a class of nonlinear scalar processes. 6th IFAC Symp. on Ident. and Sys. Par. Est., Washington DC, USA. pp1401-1405.

Leontaritis, I.J. and Billings, S.A. 1985, Input-ouput parametric models for nonlinear systems: Part I: Deterministic nonlinear models, - Part II: Stochastic nonlinear systems. Int J. Control, Vol. 41, No. 2, pp303-344.

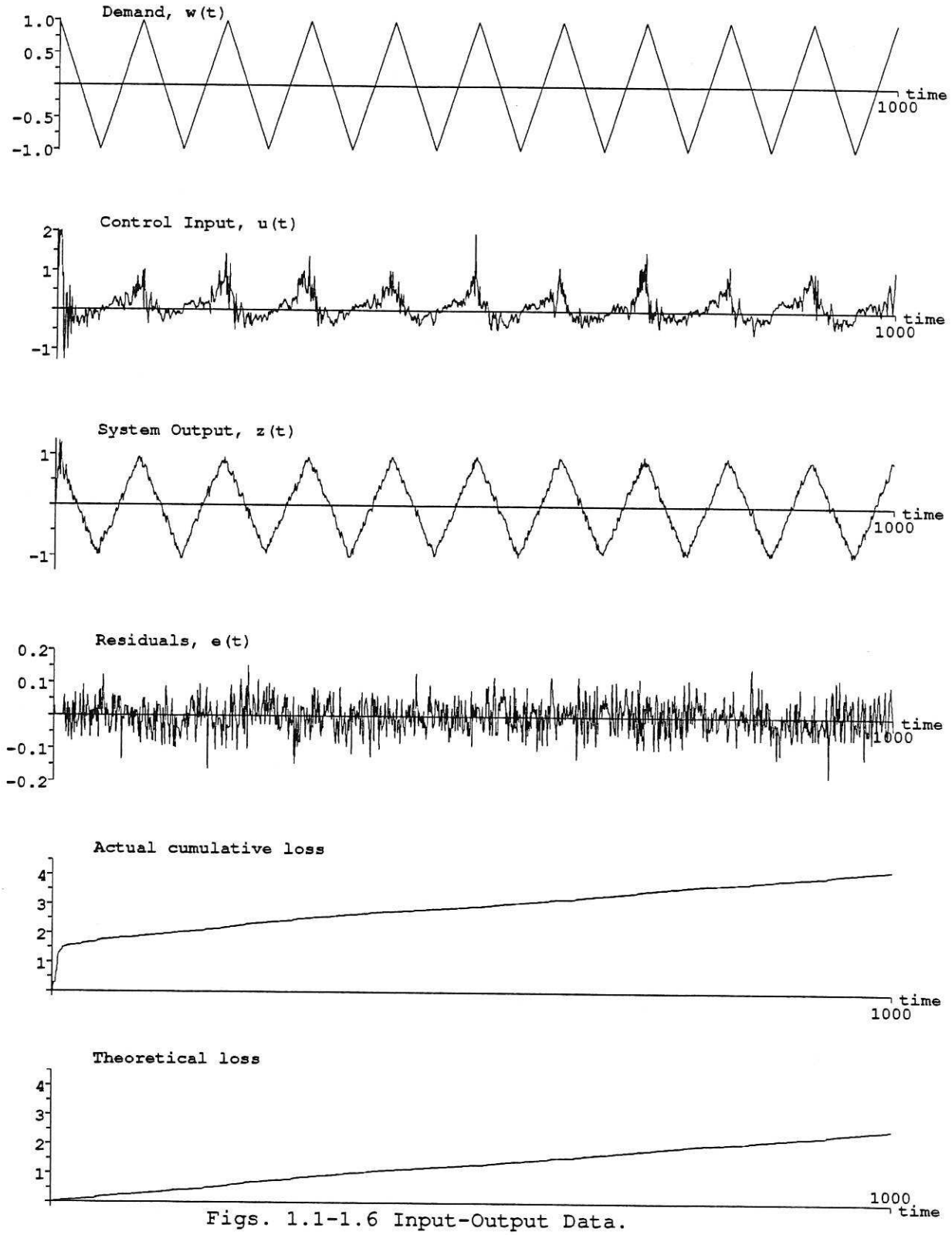
Mohler, R.R., 1973, Bilinear Control Processes. Acedemic Press.

Soderstrom, T., Ljung, L. and Gustavsson, I. 1978, A theoretical analysis of recursive estimation methods. Automatica, Vol. 14, No. 3, pp231-244.

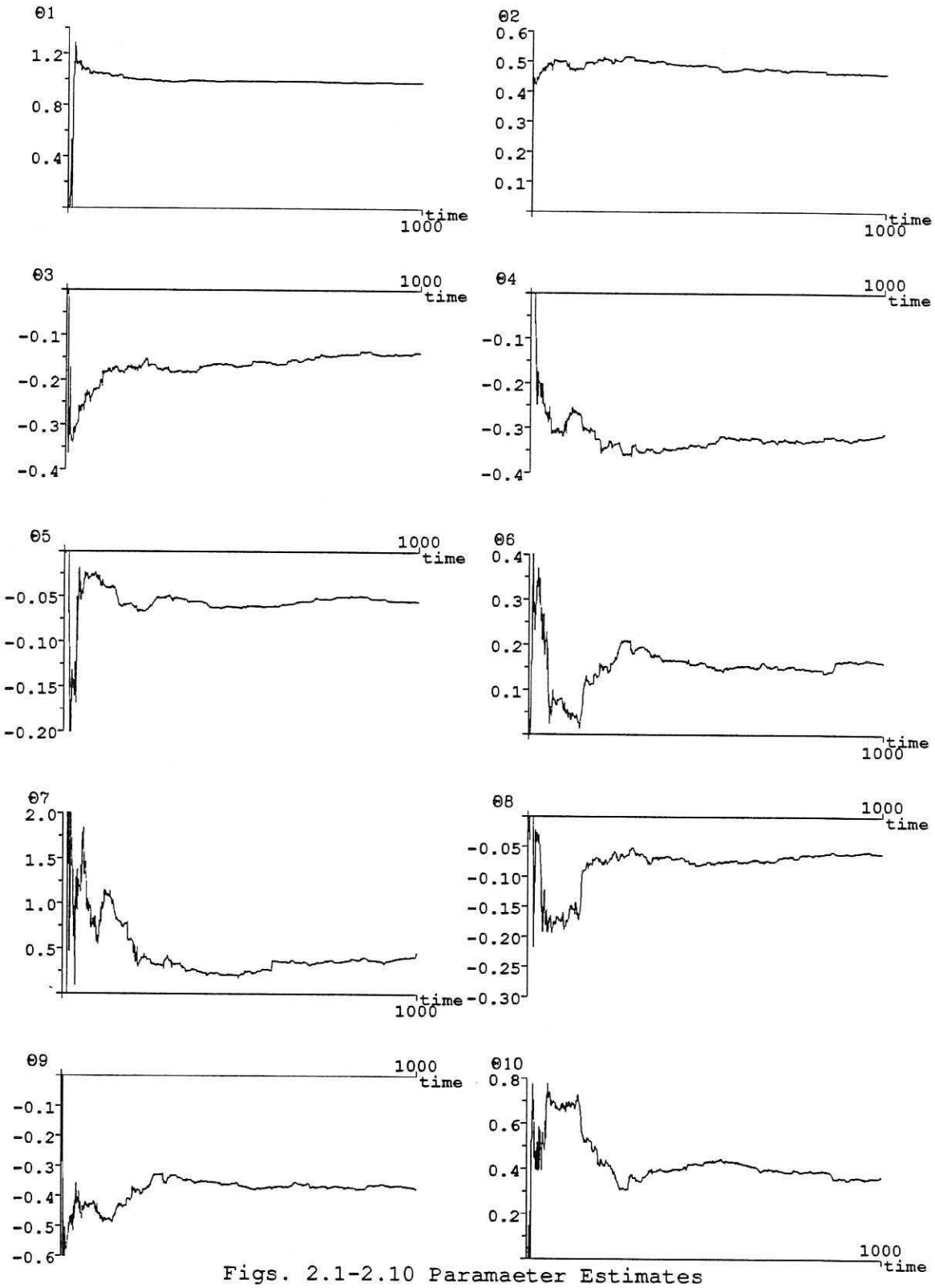
Sontag, E.D. 1976, On realisations of discrete-time nonlinear systems. "Notices", A.M.S. Vol. 23, pp19-23.

Sontag, E.D. 1979, Polynomial Response Maps. Spinger Verlag.

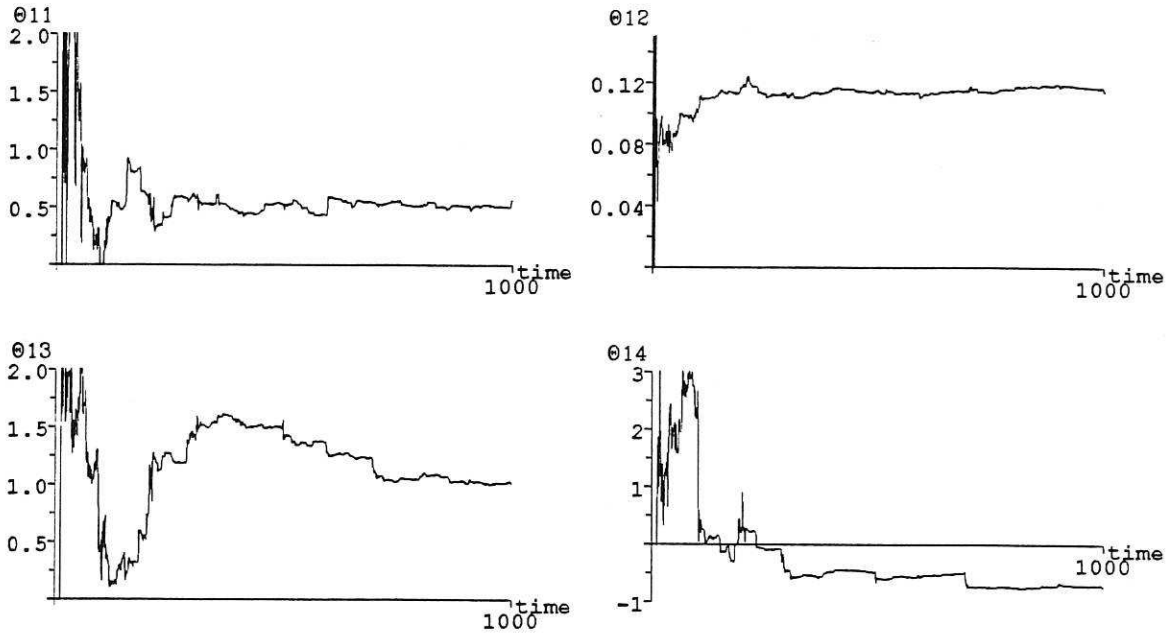
Svoronos, S., Stephanopoulos, G. and Aris, R. 1981, On bilinear estimation and control. Int. J. Control, Vol. 34, No. 4, pp651-684.



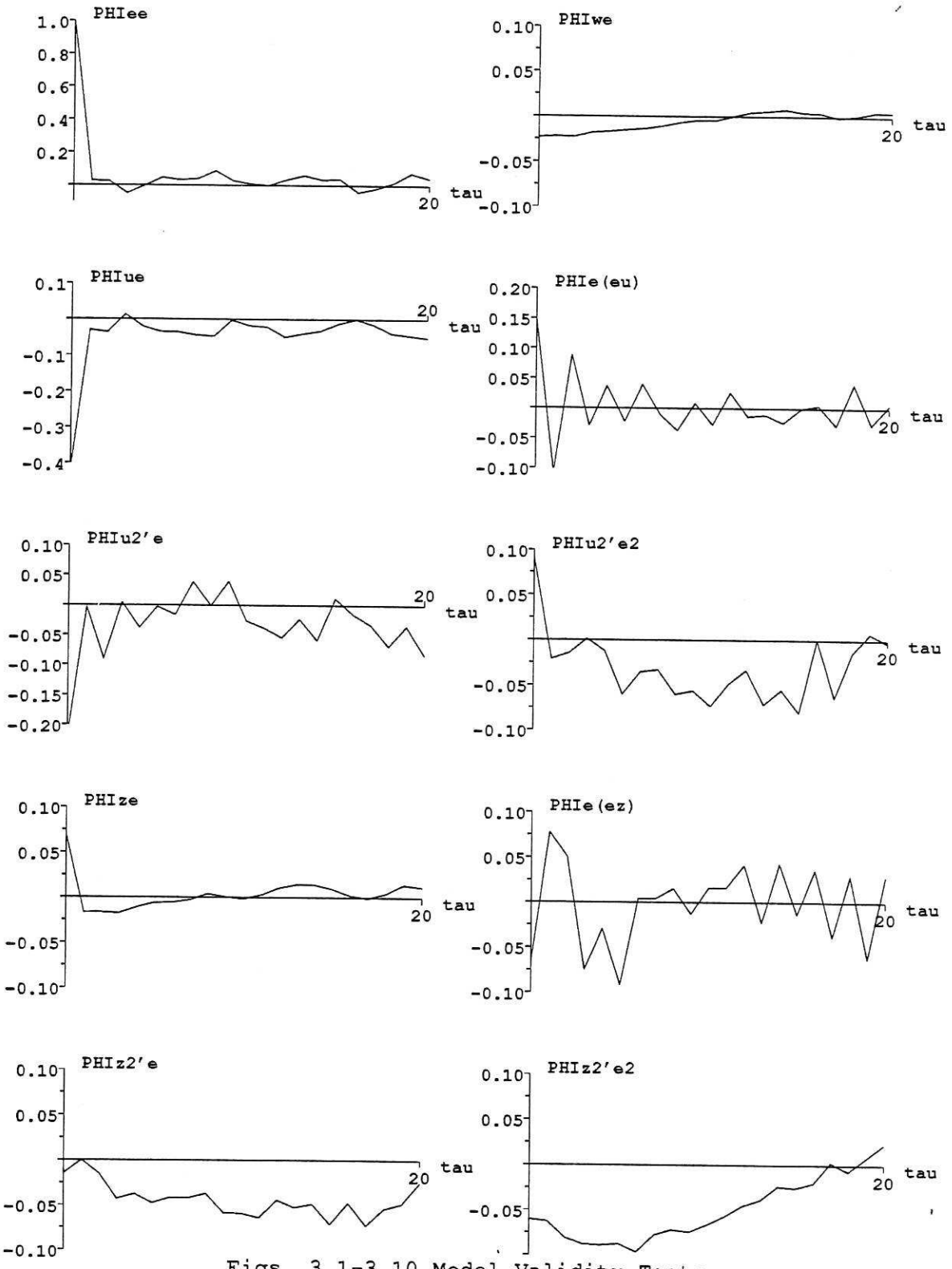
Figs. 1.1-1.6 Input-Output Data.



Figs. 2.1-2.10 Paramaeter Estimates



Figs. 2.11-2.14 Parameter Estimates



Figs. 3.1-3.10 Model Validity Tests