Orthogonal Least Squares Parameter Estimation Algorithms for Nonlinear Stochastic Systems

S.A. BILLINGS and G.N. JONES

Department of Control Engineering
University of Sheffield
Mappin Street
Sheffield S1 3JD
U.K.

Research Report No. 404

October 1990

Orthogonal Least Squares Parameter Estimation Algorithms for Nonlinear Stochastic Systems

S.A. BILLINGS and G.N. JONES

Department of Control Engineering, University of Sheffield S1 3JD, U.K.

Abstract: The derivations of orthogonal least squares algorithms based on the principle of Hsia's method and generalised least squares are presented. Extensions to the case of nonlinear stochastic systems are discussed and the performance of the algorithms is illustrated with the identification of both simulated systems and linear models of an electric arc furnace and a gas furnace.

1. Introduction

A wide class of finite-dimensional nonlinear systems can be represented by the NARMAX model (Nonlinear Autoregressive Moving Average Model with eXogenous inputs) [Leontaritis and Billings,1985]. Expanding this representation however, gives a large number of possible terms which may be required to represent the dynamic process. Estimating all of the parameters of such a model simultaneously requires excessive computation and can introduce severe numerical ill-conditioning problems. To overcome these difficulties, an auxiliary model can be introduced defined such that the candidate terms in the model are orthogonal over the data set. This has several advantages over conventional methods, the most dramatic of which is the ability to estimate the coefficient of each term independently of the other terms, and subsequently to estimate the contribution that the term makes to the overall system output, thus facilitating term selection.

This approach has been employed to derive a parameter estimation algorithm based on extended least squares [Korenberg,Billings,Liu,McIlroy, 1988] and to study various orthogonalisation methods [Chen,Billings,Luo, 1989]. Many other conventional least squares like estimation algorithms have been previously derived, mostly for linear systems, and this paper attempts to extend two such algorithms, based on Hsia's [Hsia,1976], and the generalised least squares [Clarke,1967] routines, to take advantage of using an auxiliary orthogonal model and to extend the algorithms to nonlinear systems, where this is appropriate.

Section §2 considers alternative representaions of nonlinear system dynamics. Then sections §3 and §4 consider the two algorithms debiased least squares [Hsia,1976], and generalized least squares [Clarke,1967] respectively. Section §5 illustrates the performance of the new algorithms for simulated nonlinear systems and for the identification of linear models of an electric arc furnace and a gas furnace.

2. System Representation

A dynamic system with additive output noise can be expressed in several forms within the NARMAX representaion [Leontaritis and Billings,1985; Chen and Billings,1989]. The form of this model used in the present study is

$$z(t) = F^l[z(t-1),....z(t-N_z),u(t),....u(t-N_u),$$
$$e(t-1),....e(t-N_e)] + e(t) \qquad (1)$$

which may also be written as

$$z(t) = F_a^l[z(t-1),....z(t-N_z),u(t),....u(t-N_u)] +$$
$$F_b^l[z(t-1),....z(t-N_z),u(t),....u(t-N_u),e(t-1),....e(t-N_e)] + e(t)$$
$$= F_a^l[z(t-1),....z(t-N_z),u(t),....u(t-N_u)] + \varepsilon(t) \qquad (2)$$

where all terms in $F_a^l[...]$ do not include any $e(t)$ terms $\forall t$, and

$u(t)$ is the system input

$z(t)$ is the measured system output

$e(t)$ is the residual or prediction error sequence defined as

$$e(t) = z(t) - \hat{F}_a^l[...] - \hat{F}_b^l[...]$$

In eqns (1) and (2), the lagged values of $e(t)$ are included within the model to accommodate coloured or correlated sequences of deterministic prediction errors. This is essentially the extended least squares formulation [Korenberg,Billings,Liu,McIlroy, 1988].

3. Debiased Least Squares

A system represented as in (2) can be expanded, with F_a^l defined to be a polynomial, to the form

$$z(t) = \sum_{m=0}^{m=M} \theta_m p_m(t) + \varepsilon(t) \qquad (3)$$

For example: the model

$$z(t) = \alpha z(t-1) + \beta u(t-1) + \varepsilon(t)$$

can be expressed in the form of (3) by putting

$$\theta_0 = \alpha, \quad p_0(t) = z(t-1)$$
$$\theta_1 = \beta, \quad p_1(t) = u(t-1)$$

Equation (3) can be writen in matrix form as

$$Z = P\Theta + \underline{\varepsilon} \tag{4}$$

where

$$Z^T = [z(1), z(2) \ldots\ldots z(N)]$$
$$\Theta^T = [\theta_1, \theta_2, \ldots\ldots \theta_M]$$
$$\underline{\varepsilon}^T = [\varepsilon(1), \varepsilon(2) \ldots\ldots \varepsilon(N)]$$

$$P = \begin{bmatrix} p_0(1) & p_1(1) & . \; . & p_M(1) \\ p_0(2) & p_1(2) & . \; . & p_M(2) \\ . & . & & . \\ . & . & & . \\ p_0(N) & p_1(N) & . \; . & p_M(N) \end{bmatrix} \tag{5}$$

and the $p_i(t)$ terms depend only on lagged values of z(t) and u(t).

To formulate the algorithm consider an auxiliary form of eqn (4)

$$Z = Wg + \underline{\varepsilon} \tag{6}$$

where

$$W = PT^{-1} \quad and \quad g = T\underline{\theta} \tag{7}$$

are formed such that the columns of W are mutually orthogonal, and T is a unity diagonal triangular matrix [Korenberg et al. 1988]. The formulation of this orthogonal model is described in Appendix A. Rather than following the extended least squares formulation derived previously [Korenberg et al. 1988], in the present study an alternative solution to the correlated output noise problem is proposed. This new algorithm which is derived by extending Hsia's method [Hsia 1976] is outlined below.

A normal least squares estimate of the parameter vector g in eqn (6) can be obtained as

$$\hat{g}_{ls} = \left[W^T W\right]^{-1} W^T Z \tag{8}$$

Substituting (6) into (8), gives

$$\hat{g}_{ls} = \left[W^T W\right]^{-1} W^T W g + \left[W^T W\right]^{-1} W^T \underline{\varepsilon}$$

$$= \underline{g} + \left[W^T W\right]^{-1} W^T \underline{\varepsilon} \tag{9}$$

Hence the least squares estimate will be biased whenever $\varepsilon(t)$ is a coloured noise sequence since

$$E[W^T \underline{\varepsilon}] \neq \underline{0} \tag{10}$$

However, rearranging (9) gives

$$\underline{g} = \hat{\underline{g}}_{ls} - \left[W^T W\right]^{-1} W^T \underline{\varepsilon} \tag{11}$$

or

$$\underline{g} = \hat{\underline{g}}_{ls} - \underline{g}_{Bias} \tag{12}$$

Hence if an estimate of $\underline{g}_{Bias}$ can be made such that

$$E[\hat{\underline{g}}_{Bias}] = \underline{g}_{Bias} \tag{13}$$

an unbiased estimate of $\underline{g}$ can be obtained.

An iterative algorithm to estimate $\underline{\varepsilon}$ and $\hat{\underline{g}}_{Bias}$ can be formulated where from eqn (6) $\underline{\varepsilon}$ is estimated as

$$\hat{\underline{\varepsilon}}^{j+1} = Z - W\hat{\underline{g}}^j \tag{14}$$

and j denotes the iteration index.

$\hat{\underline{g}}_{Bias}$ cannot be estimated directly as

$$\hat{\underline{g}}_{Bias} = \left[W^T W\right]^{-1} W^T \hat{\underline{\varepsilon}} \tag{15}$$

for substituting (14) into (15) and expanding, shows that if

$$\hat{\underline{g}}^j = \hat{\underline{g}}_{ls} - \underline{\delta g}^j \tag{16}$$

then

$$\hat{\underline{g}}^{j+1}{}_{Bias} = \underline{\delta g}^j \tag{17}$$

and hence

$$\hat{\underline{g}}^{j+1} = \hat{\underline{g}}^j \tag{18}$$

and the algorithm fails.

However,if $\varepsilon(t)$ is modeled as $F_b^l[\ldots]$ in eqn (2), expanded as a polynomial, then

$$\varepsilon(t) = \sum_{q=0}^{q=Q} c_q \omega_q + e(t) \tag{19}$$

where the c_q's are coefficients and the ω_q's are the regressor terms.

For example: the model

$$\varepsilon(t) = \alpha e(t-1) + \beta u(t-1)e(t-1) + e(t)$$

can be expressed in the form of (19) by putting

$$c_0 = \alpha, \ \omega_0(t) = e(t-1)$$
$$c_1 = \beta, \ \omega_1(t) = u(t-1)e(t-1)$$

In matrix form

$$\underline{\varepsilon}(t) = \Omega \underline{c} + \underline{e}(t) \tag{20}$$

where

$$\underline{\varepsilon}^T = [\varepsilon(1),\varepsilon(2).....\varepsilon(N)]$$
$$\underline{c}^T = [c_1,c_2,.....c_Q]$$
$$\underline{e}^T = [e(1),e(2).....e(N)]$$

$$\Omega = \begin{bmatrix} \omega_0(1) & \omega_1(1) & . \ . & \omega_Q(1) \\ \omega_0(2) & \omega_1(2) & . \ . & \omega_Q(2) \\ . & . & & . \\ . & . & & . \\ \omega_0(N) & \omega_1(N) & . \ . & \omega_Q(N) \end{bmatrix} \tag{21}$$

The bias term can therefore be iteratively computed as

$$\hat{g}_{Bias} = \left[W^T W\right]^{-1} W^T \hat{\Omega} \underline{\hat{c}} \tag{22}$$

and $\underline{\varepsilon}$ corrected accordingly. Orthogonal least squares can be used a second time to estimate $\underline{\hat{c}}$ from $\underline{\varepsilon}$ *and* $\hat{\Omega}$ which is formed from the latest residual sequence. Finaly, the true process parameters can be found recursively as

$$\hat{\Theta} = \hat{g} - (T - I)\hat{\Theta} \tag{23}$$

Note that in the current algorithm, the noise model is considered as a moving average type process. However an autoregressive type representation could also be used

$$\varepsilon(t) = F_b^l[z(t-1),....z(t-N_z),u(t),....u(t-N_u),\varepsilon(t-1),....\varepsilon(t-N_\varepsilon)] + e(t)$$

since the estimates $\hat{\underline{g}}$ and hence $\underline{\varepsilon}$ are corrected at each iteration stage.

The selection of terms to be included in both the noise and the process models in the extended least squares formulation was achieved by implementing a forward regression algorithm coupled with an error reduction ratio test (ERR) [Chen,Billings and Luo, 1989]. In Debiased Least Squares the ERR test can be applied to select the process model terms. A candidate process term is accepted into the model if the error reduction ratio is above the threshold C_d. Noise terms may also be selected by considering the contribution they make to $\varepsilon(t)$, and using a separate threshold C_{de}.

The model will be satisfactory if, and only if, the residuals are unpredictable from all linear and nonlinear combinations of past inputs and outputs. This can be demonstrated if the following correlation tests [Billings and Voon,1986] are satisfied:

$$\begin{aligned}
&\phi_{ee}(\tau) = \delta(\tau) \\
&\phi_{ue}(\tau) = 0 \quad \forall\tau \\
&\phi_{eeu}(\tau) = E[e(t)e(t-1-\tau)u(t-1-\tau)] = 0 \quad \forall\tau \\
&\phi_{u^{2\prime}e}(\tau) = 0 \quad \forall\tau \\
&\phi_{u^{2\prime}e^2}(\tau) = 0 \quad \forall\tau
\end{aligned} \tag{24}$$

The standard deviations of the process parameters can be found by using the result, that if $e(t)$ has been reduced to a white sequence

$$E\left[\underline{e}\underline{e}^T\right] = \sigma^2 I \tag{25}$$

Hence

$$\begin{aligned}
cov(\hat{g}) &= E\left[(\hat{g} - g)(\hat{g} - g)^T\right] \\
&= \sigma^2\left\{W^T W\right\}^{-1}
\end{aligned} \tag{26}$$

and

$$cov(\hat{\underline{\theta}}) = T^{-1} cov(\hat{g}) T^{-T} \tag{27}$$

[Korenberg et al.,1988].

Similarly, the standard deviations of the noise parameters can be estimated by assuming

$$\hat{F}^l_a[...] \equiv F^l_a[...] \tag{28}$$

and hence that $\underline{\varepsilon}$ is known.

The results of applying this new algorithm to two industrial processes are given in section §5.

4. Generalized Least Squares

Parameter estimates for a system of the type considered in this paper can be obtained using a Generalised Least Squares algorithm [Clarke,1967], provided that the noise can be adequately modeled as an autoregressive process such that

$$\varepsilon(t) = F_b^l[\varepsilon(t-1)...\varepsilon(t-N_\varepsilon)] + e(t) \tag{29}$$

This condition implies that, if a nonlinear process expansion is used, acceptable estimates can only be obtained in special cases.

The algorithm can be modified to use an orthogonal estimation routine at each iteration, which will aid term selection. However the data filtering used in GLS destroys the orthogonal properties of the model and a new orthogonal model must be derived at each iteration stage. In addition if an orthogonal model is used to estimate the noise parameters, this must be deorthogonalised in order to obtain the necessary filter parameters.

Such an algorithm is not described in greater detail in this paper because the inherent structure of the original algorithm means that an orthogonal version will not bring any advantages.

5. Identification Results

The Debiased Least Squares parameter estimation algorithm described above has been used to identify models from both simulated and real data sets.

5.1 EXAMPLE 1: Identification with simulated data.

Consider the system

$$\begin{aligned} y(t) &= 0.5y(t-1) + u(t-2) + 0.1y(t-2)u(t-1) \\ \varepsilon(t) &= 0.5e(t-1) + 0.2e(t-2)^2 + e(t) \\ z(t) &= y(t) + \varepsilon(t) \end{aligned} \tag{30}$$

which when rearranged takes the form of eqn (1)

$$\begin{aligned} z(t) = {} & 0.5z(t-1) + u(t-2) + 0.1z(t-2)u(t-1) - 0.25e(t-2) \\ & - 0.1u(t-1)e(t-2) + 0.2e(t-2)^2 - 0.05u(t-1)e(t-3) \end{aligned}$$

$$- 0.1e(t-3)^2 - 0.02u(t-1)e(t-4)^2 + e(t) \tag{31}$$

A sequence of 600 data pairs was generated using a gaussian input $N(0,4)$ and a noise sequence $e(t)$ with a maximum magnitude of 2. The first 400 data points were used for system identification, and the full data set for model validation.

The candidate terms for the identified model were obtained by expanding (1) as a cubic polynomial with $N_y = N_u = 2$, and $N_e = 4$ to give a total of 164 terms. The model set was reduced, by excluding all cubic terms except those linear in $u(t-1)$ and quadratic in $e(.)$, to give 54 possible terms. Identification with $C_d = 0.015$ *and* $C_{de} = 0.0075$ gave the model:

Terms	*Parameters*	*$[err]_i$'s*	*Std. dev*	
$z(t-1)$	0.4933	0.2381	0.022	
$u(t-2)$	0.9961	0.5030	0.031	
$z(t-2)u(t-1)$	0.1276	0.0319	0.012	
$e(t-4)$	0.1189	0.0107	0.051	
$u(t-1)e(t-2)$	−0.1606	0.0744	0.026	(32)
$u(t-1)e(t-3)$	−0.0642	0.0151	0.026	
$u(t-2)e(t-1)$	−0.0583	0.0103	0.027	
$e(t-2)e(t-2)$	0.1586	0.0258	0.034	
$e(t-3)e(t-3)$	−0.1061	0.0165	0.034	
$u(t-1)e(t-1)e(t-1)$	−0.0325	0.0097	0.015	

The model validity tests, defined in eqn (24), are illustrated in Figure 1 for this model. Note that $\phi_{\xi\xi}(2) \neq 0$, clearly indicating that an $e(t-2)$ term is missing from the model. Hence the model was reidentified, using the terms of (32) augmented by an $e(t-2)$ term with $C_d = 0$ *and* $C_{de} = 0.005$. Estimation yielded

Terms	*Parameters*	*$[err]_i$'s*	*Std. dev*	
$z(t-1)$	0.5143	0.2381	0.022	
$u(t-2)$	0.9944	0.5030	0.030	
$z(t-2)u(t-1)$	0.1211	0.0319	0.012	
$e(t-2)$	−0.2611	0.0522	0.051	
$u(t-1)e(t-2)$	−0.1576	0.0687	0.027	(33)
$u(t-1)e(t-3)$	−0.0665	0.0142	0.026	
$u(t-2)e(t-1)$	−0.0450	0.0065	0.025	
$e(t-2)e(t-2)$	0.1576	0.0226	0.038	
$e(t-3)e(t-3)$	−0.1052	0.0149	0.037	

Which by inspection of the model validity tests (Figure 2) can be seen to provide an acceptable representation of the system dynamics.

5.2 EXAMPLE 2: Identification of an electric arc furnace.

This data set was collected from a 135 tonne 35MVA 3-phase electric-arc steel-making furnace with an amplidyne Ward-Loenard regulator. A description of the plant and the data is given by Billings and Nicholson [Billings and Nicholson, 1975]. The input sequence is a PRBS injected into the amplidyne, and the output sequence is the electrode position. 250 data pairs were used for the identification and 350 pairs were considered for the model validity tests. Due to the presence of an integrator on the system output, the differenced output $z_d(t)$ was used throughout where

$$z_d(t) = z(t) - z(t-1) \tag{34}$$

A candidate model of 24 terms was obtained by expanding (1) as a linear polynomial with $N_y = N_u = N_e = 8$. Identification was performed with $C_d = 0.005$ and $C_{de} = 0.0025$, yielding

Terms	*Parameters*	*[err]$_i$'s*	*Std. dev*
$u(t-5)$	−0.4209	0.0113	0.081
$u(t-7)$	−0.1699	0.0073	0.090
$z_d(t-1)$	1.7927	0.6846	0.045
$z_d(t-2)$	−1.5895	0.0363	0.066
$z_d(t-3)$	1.1994	0.1058	0.065
$z_d(t-4)$	−0.4728	0.0517	0.043
$e(t-2)$	−0.7629	0.3773	0.065
$e(t-3)$	0.1128	0.0062	0.065
$e(t-4)$	0.0758	0.0033	0.065
$e(t-5)$	−0.2242	0.0297	0.066
$e(t-8)$	0.0915	0.0047	0.066

(35)

Which by inspection of the model validity tests (Figure 3) can be seen to adequately represent the system dynamics.

The one step ahead predicted output $\hat{z}(t)$, defined as

$$\hat{z}_d(t) = \hat{F}^l_a[z_d(t-1),....z_d(t-N_z),u(t),....u(t-N_u)] + \hat{F}^l_b[z_d(t-1),....z_d(t-N_z),u(t),....u(t-N_u),e(t-1),....e(t-N_e)] \tag{36}$$

where $e(t) = z_d(t) - \hat{z}_d(t)$, is compared with the measured output in Figure 4.

The model predicted output is defined as

$$\hat{z}_d^{po}(t) = \hat{F}^l_a[\hat{z}_d(t-1),....\hat{z}_d(t-N_z),u(t),....u(t-N_u)] \tag{37}$$

and is shown in Figure 5.

5.3 EXAMPLE 3: Identification of a gas furnace.

This data set was collected from a gas furnace producing carbon dioxide by Jenkins and Watts [Jenkins and Watts, 1968]. The input sequence is the input gas rate (a second input, air rate was kept constant) and the output sequence is the percentage of carbon dioxide in the outlet gas. The first 200 data pairs were used for identification and 295 for model validation. Due to a trend in the data the identification was performed on differenced outputs and inputs $z_d(t)$ and $u_d(t)$ where

$$z_d(t) \;=\; z(t) - z(t-1) \;\; and \;\; u_d(t) \;=\; u(t) - u(t-1) \tag{38}$$

A candidate model of 24 terms formed by expanding (1) as a linear polynomial with $N_y = N_u = N_e = 8$ was used with $C_d = C_{de} = 0.01$. The estimated model is given below

Terms	*Parameters*	*[err]$_i$'s*	*Std. dev*
$u_d(t-3)$	−0.0733	0.0823	0.003
$u_d(t-5)$	−0.0711	0.7366	0.004
$z_d(t-1)$	0.5381	0.0642	0.017
$e(t-3)$	−0.1825	0.0295	0.031
$e(t-4)$	−0.2334	0.0496	0.031
$e(t-5)$	−0.1892	0.0314	0.031

(39)

Inspection of the model validity tests (Figure 6) shows this model to be an acceptable representation of the system dynamics. The one step ahead predicted output is compared with the measured output in Figure 7, and the model predicted output is shown in Figure 8.

6. Conclusion

An established system identification algorithm has been extended to the nonlinear case, and reformulated to take advantage of an orthogonal auxiliary model.

The advantages of the new algorithm include computational simplicity and the ability to implement a simple term selection routine, based on an error reduction ratio test.

The results of applying the new Orthogonal Debiased Least Squares algorithm to identify models of simulated systems and to both an electric arc furnace and a gas furnace have been presented.

Acknowledgments

The authors gratefully acknowledge the support of the UK Science and Engineering Research Council under grant GK/F24177.

APPENDIX A: Formulation of the Auxhillary Orthogonal Model

The matrix decomposition theorem [Fox, 1964] states that a positive definite square matrix A can be decomposed as

$$A = LDU \tag{A1}$$

in which L *and* U are unity lower and upper triangular matrices, D is diagonal with all positive elements and the decomposition is unique. If A is constrained to be symmetric then it can be shown that

$$L = U^T \tag{A2}$$

and hence

$$A = U^T D U \tag{A3}$$

The correlation matrix $P^T P$ associated with the model

$$Z = P\Theta + \underline{\varepsilon} \tag{A4}$$

is symmetric and positive definite and can therefore be expressed as

$$P^T P = T^T D T \tag{A5}$$

where T is unity upper triangular and D is diagonal with all positive elements.

Now eqn A4 can be expressed as

$$Z = P(T^{-1}T)\Theta + \underline{\varepsilon} \tag{A6}$$

or

$$Z = W\underline{g} + \underline{\varepsilon} \tag{A7}$$

where

$$W = PT^{-1}, \; \underline{g} = T\Theta \tag{A8}$$

and it can be shown that

$$W^T W = D \tag{A9}$$

Defining W as

$$W = \begin{bmatrix} w_0(1) & w_1(1) & . & . & w_M(1) \\ w_0(2) & w_1(2) & . & . & w_M(2) \\ . & . & & & . \\ . & . & & & . \\ w_0(N) & w_1(N) & . & . & w_M(N) \end{bmatrix} \qquad \text{A10}$$

where

$$\sum_{t=1}^{N} w_i(t) w_j(t) = 0 \quad \forall\, i \neq j \qquad \text{A11}$$

specifies D in eqn A9 as

$$D = \begin{bmatrix} \sum_{t=1}^{N} w_0^2(t) & & & 0 \\ & \sum_{t=1}^{N} w_1^2(t) & & \\ & & . & \\ 0 & & . & \sum_{t=1}^{N} w_M^2(t) \end{bmatrix} \qquad \text{A12}$$

Premultipling the definition of W in eqn A8 by W^T and postmultipling by T gives

$$W^T W T = W^T P T^{-1} T = W^T P \qquad \text{A13}$$

or

$$T = \left(W^T W\right)^{-1} W^T P = D^{-1} W^T P \qquad \text{A14}$$

Multipling this expression out gives

$$T = \begin{bmatrix} \dfrac{\sum_{t=1}^{N} w_0(t) p_0(t)}{\sum_{t=1}^{N} w_0^2(t)} & \dfrac{\sum_{t=1}^{N} w_0(t) p_1(t)}{\sum_{t=1}^{N} w_0^2(t)} & . \; . & \dfrac{\sum_{t=1}^{N} w_0(t) p_M(t)}{\sum_{t=1}^{N} w_0^2(t)} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ \dfrac{\sum_{t=1}^{N} w_M(t) p_0(t)}{\sum_{t=1}^{N} w_M^2(t)} & \dfrac{\sum_{t=1}^{N} w_M(t) p_1(t)}{\sum_{t=1}^{N} w_M^2(t)} & . \; . & \dfrac{\sum_{t=1}^{N} w_M(t) p_M(t)}{\sum_{t=1}^{N} w_M^2(t)} \end{bmatrix}$$

$$
= \begin{bmatrix} \alpha_{00} & \alpha_{01} & . & . & \alpha_{0M} \\ \alpha_{10} & \alpha_{11} & . & . & \alpha_{1M} \\ . & . & & & . \\ . & . & & & . \\ \alpha_{M0} & \alpha_{M1} & . & . & \alpha_{MM} \end{bmatrix} \qquad \text{A15}
$$

To satisfy the requirements of A5, T should be upper triangular, and this can be achieved by defining

$$
\alpha_{ij} = \begin{cases} 0 & \forall\, i > j \\ 1 & \forall\, i = j \\ \dfrac{\sum_{t=1}^{N} w_i(t) p_j(t)}{\sum_{t=1}^{N} w_i^2(t)} & \forall\, i < j \end{cases} \qquad \text{A16}
$$

which implies

$$
T = \begin{bmatrix} 1 & \alpha_{01} & \alpha_{02} & . & . & \alpha_{0M} \\ & 1 & \alpha_{12} & . & . & \alpha_{1M} \\ & & 1 & & & . \\ & 0 & & & & . \\ & & & & & 1 \end{bmatrix} \qquad \text{A17}
$$

The elements of W in eqn A10 can now be determined by writing

$$
WT = P
$$

as

$$
W = W - W(T - I) \qquad \text{A18}
$$

Eqn A16 implies that the k^{th} column of T can be found from P and the first $k-1$ columns of W. Similarly eqn A18 implies that the k^{th} column of W can be found from P, the first $k-1$ columns of W, and the first k columns of T. Hence the columns of T and W can be found sequentially to form the two matrices.

Once an estimate $\hat{g}$ has been found, $\hat{\Theta}$ can be obtained by rearranging

$$
\hat{g} = T\hat{\Theta} \qquad \text{A19}
$$

as

$$\hat{\Theta} = \hat{g} - (T - I)\hat{\Theta} \quad \text{A20}$$

which enables the estimates θ_i to be found sequentially from $\hat{g}$, T and $\theta_j \; \forall j > i$ without any need to invert T.

References

BILLINGS, S.A. AND NICHOLSON, H., (1975). "Identification of an Electric-arc Furnace Electrode-control System," *Proc. IEE*, vol. 122, no. 8, pp. 849-856.

BILLINGS, S. A. AND VOON, W. S. F., (1986). "Correlation Based Model Validity Tests for Nonlinear Systems," *Int. J. Control*, vol. 44, no. 1, pp. 235-244.

CHEN, S., BILLINGS, S.A., AND LUO, W., (1989). "Orthogonal Least Squares Methods and their Application to Nonlinear System Identification.," *Int. J. Control*, vol. 50, pp. 1873-1896.

CHEN, S. AND BILLINGS, S.A., (1989). "Representations of Nonlinear Systems: The NARMAX Model.," *Int. J. Control*, vol. 49, no. 3, pp. 1013-1032.

CLARKE, D.W., (1967). "Generalized Least Squares Estimation of the Parameters of a Dynamic Model," *Proc. of the IFAC symp. Identification in Automatic Control Systems.*.

FOX, L., (1964). *An Introduction to Numerical Linear Algebra*, (Oxford: Clarendon Press).

HSIA, T.C., (Feb. 1976). "On Least Squares algorithms for System Parameter Estimation," *IEEE Transactions on Automatic Control*, pp. 104-108.

JENKINS, G.M. AND WATTS, D.G., (1968). *Spectral Analysis and its Applications*, (Holden Day).

KORENBERG, M., BILLINGS, S.A., LIU, Y.P., AND MCILROY, P.J., (1988). "Orthogonal Parameter estimation algorithm for Non-linear Stochastic Systems," *Int. J. Control*, vol. 48, no. 1, pp. 193-210.

LEONTARITIS, I.J. AND BILLINGS, S.A., (1985). "Input-Output Parametric Models for Nonlinear Systems.Parts I and II.," *Int. J. Control*, vol. 41, no. 2, pp. 303-344.

FIGURE 1: Model validity plots for the initial model estimation of example1

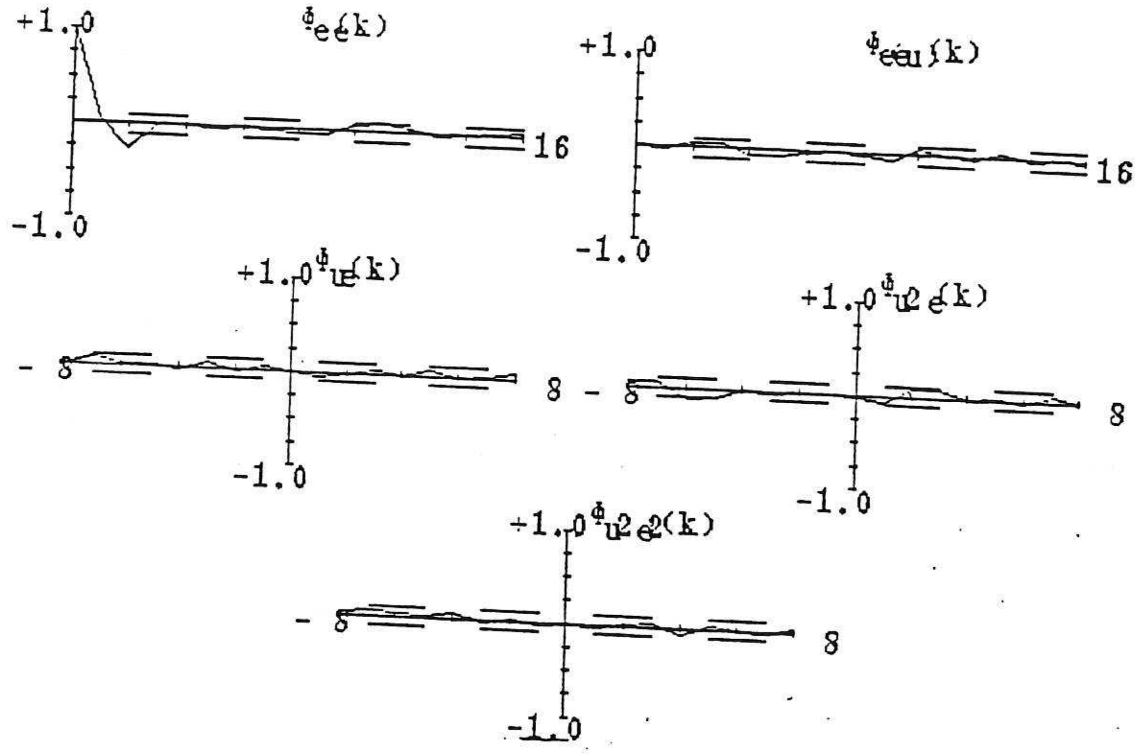

FIGURE 2: Model validity plots for the final model estimation of example 1

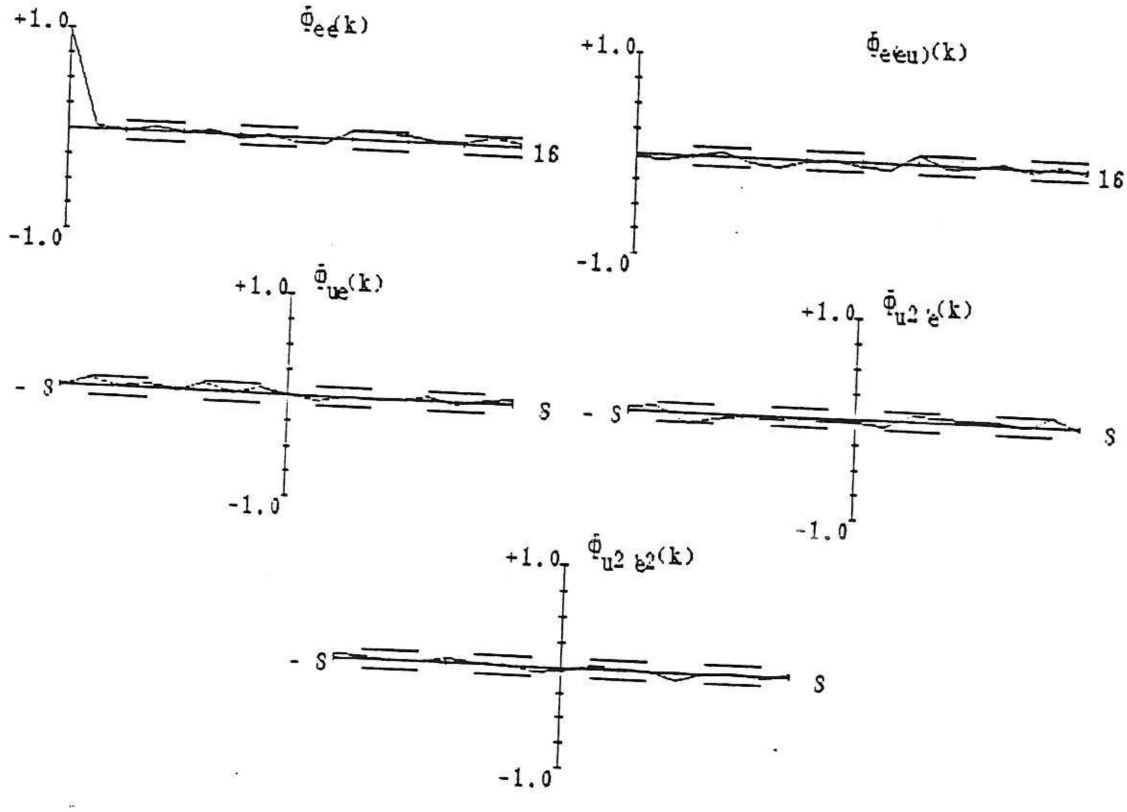

FIGURE 3: Model validity plots for the model estimation of example 2

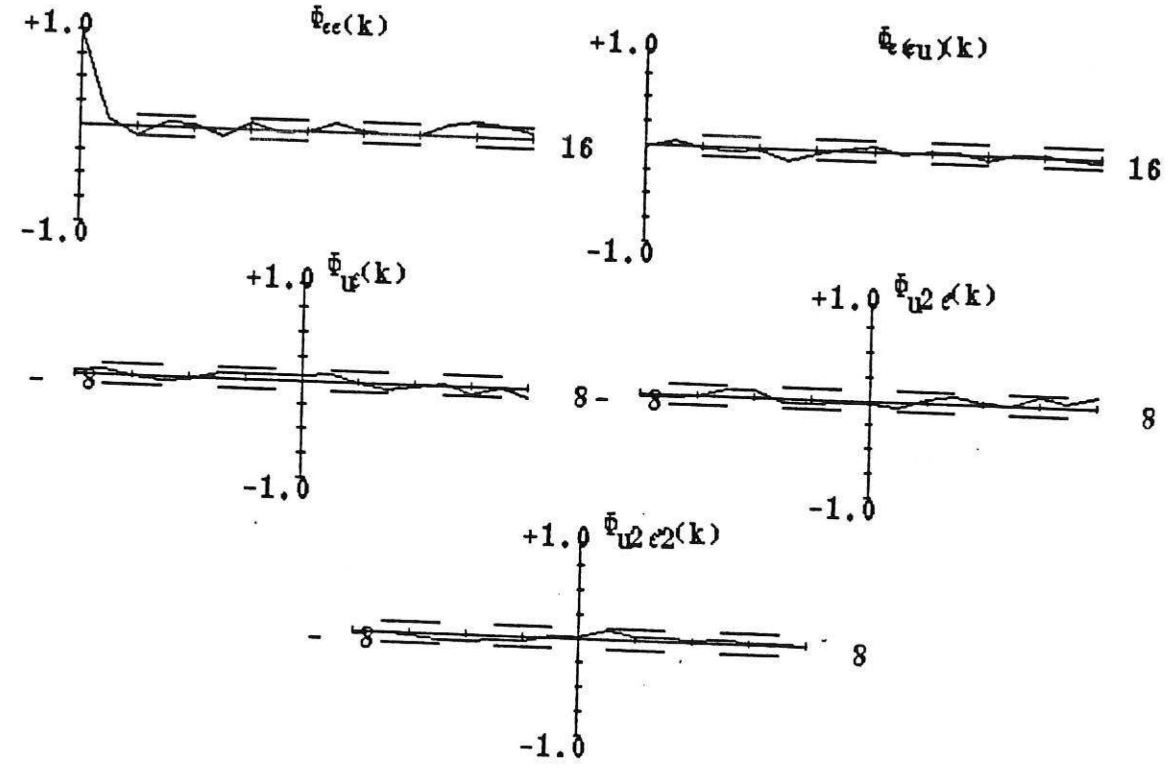

FIGURE 4: One step ahead and measured outputs of an electric arc furnace (example 2)

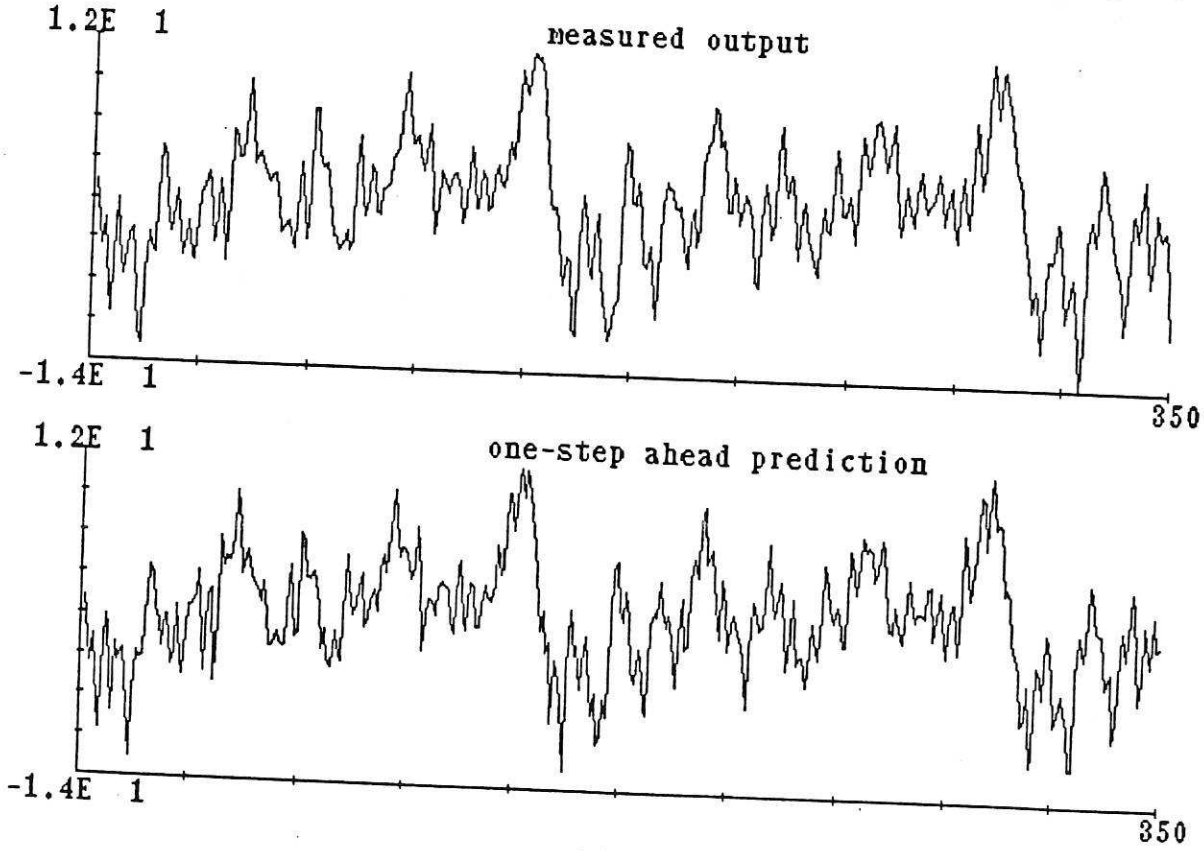

FIGURE 5: Predicted output of an electric arc furnace (example 2)

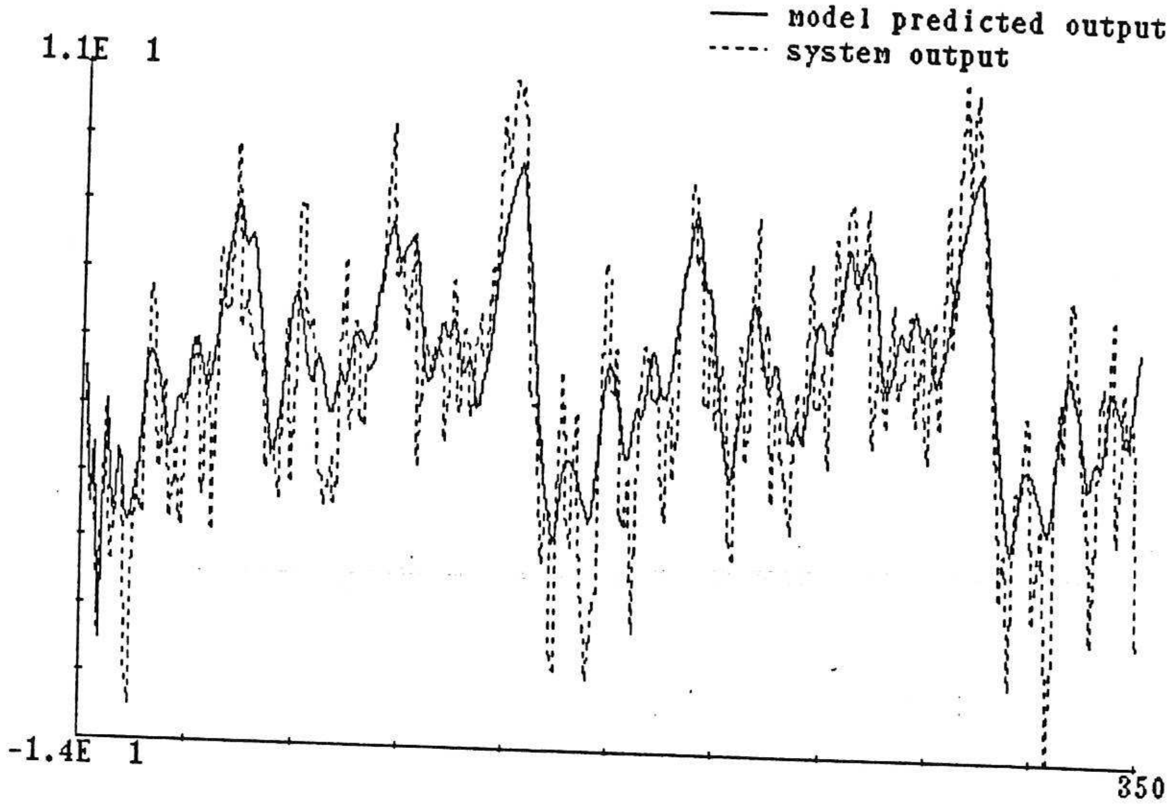

FIGURE 6: Model validity plots for the model estimation of example 3

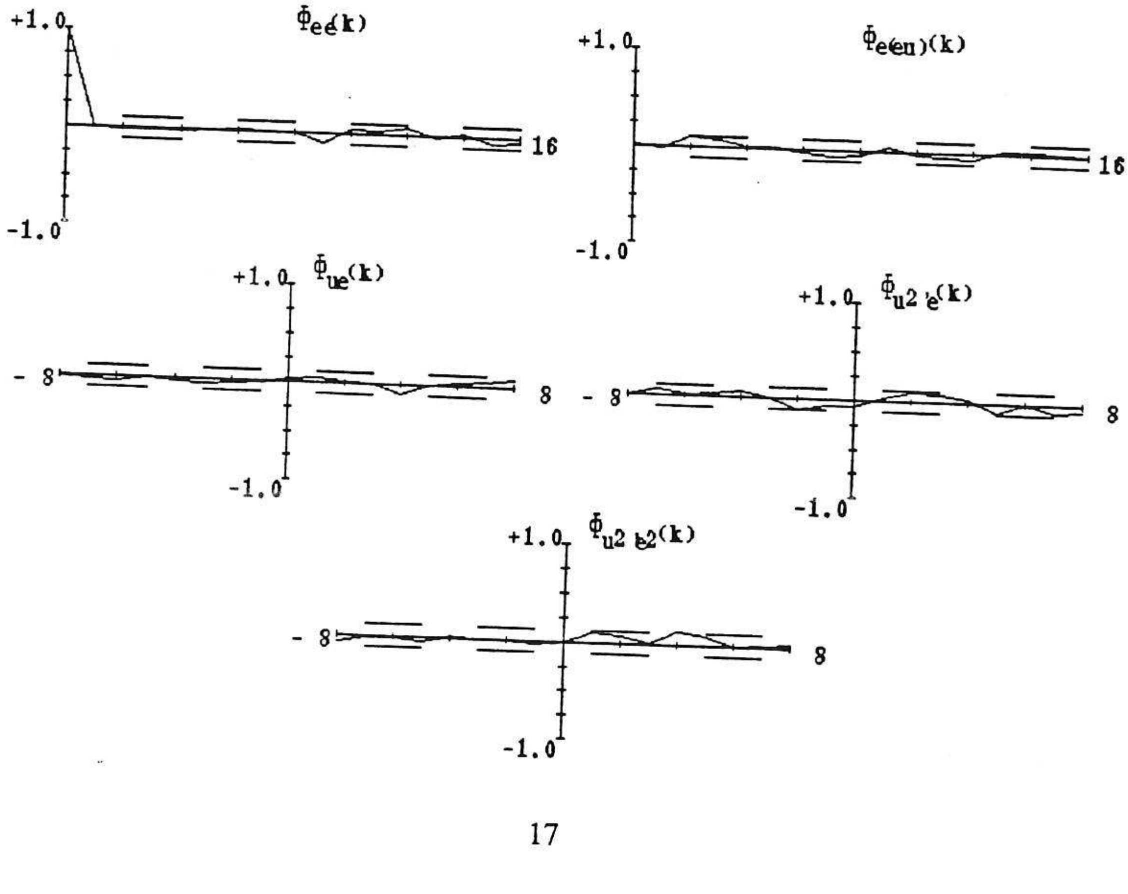

FIGURE 7: One step ahead and measured outputs of a gas furnace (example 3)

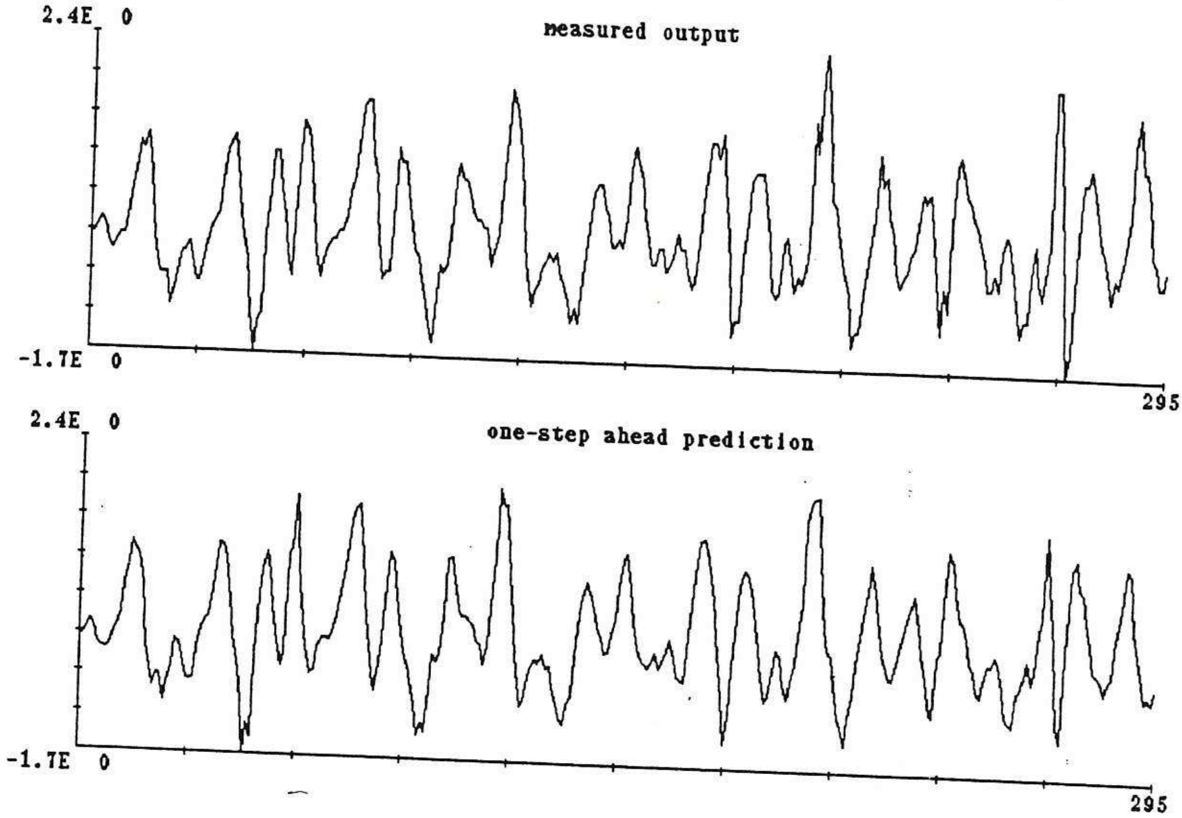

FIGURE 8: Predicted output of a gas furnace (example 3)