A New Decision Rule for Model Structure Identification of A Class of Nonlinear Dynamic Systems

by

Hong Gan

S. A. Billings

Dept. of Control Engineering

University of Sheffield

Sheffield

S1 3JD

Research Report 360

April 1989

A New Decision Rule for Model Structure Identification of A Class of Nonlinear Dynamic Systems

Hong Gan *S.A.Billings*
Control Engineering Department
Sheffield University
Mappin Street, Sheffield S1 3JD

Abstract

This paper is concerned with the problem of deriving a statistical decision rule for model structure identification of a class of nonlinear dynamic systems. The concept of a basis to describe the model structure is introduced and the model structure space and corresponding algebra structure are defined. Then, based on the Kullback-Leibler mean information, a new model structure decision rule is developed by maximising the average log-likelihood function. Some analytical and simulated comparisons of this decision rule with Akaike's FPE and AIC, F-test, and Bayes aposteriori decision rule are given.

I. Introduction

A key problem in applying system identification techniques is the appropriate choice of a decision rule for determining the model structure based on the information provided by a finite number of observations. The analysis of accuracy, convergence and consistency etc. of parameter estimators would be meaningless if the model structure was not tentatively known. Various techniques, ranging from statistical tests to correlation analysis and statistical decision theory, have been suggested in recent years.

Based on statistical tests and decision theory, some criteria for model structure identification have been suggested, such as FPE (Akaike 1970), AIC (Akaike 1974, Sakamoto 1986), Cp (Mallows 1973), PIC (Kashyap 1977), R^2 (Hosking 1979), FPE_{α} (Shibata 84), correlation analysis (Billings 1986) and hypothesis tests (Leontaritis and Billings 1987). After analysing the log-likelihood function via the prediction error, Söderström (1977) published results comparing Akaike's information criteria with the F-test and Cp plot etc. . This study showed that Akaike's information criteria, the F-test and the likelihood ratio test are all

$G^0 \subset \Omega$ is a measurable function set with finite elements.

$\theta^k \in \Theta$ is the corresponding parameter vector and $\Theta \subset \Omega$ is the parameter set.

$$Y^{t-1} = (y(t-1), y(t-2), \ldots, y(t-d_y))$$

$$U^{t-1} = (u(t-1), u(t-2), \ldots, u(t-d_u))$$

and $y(t) \in Y \subset \Gamma$; $u(t) \in U \subset \Gamma$; Y and U are the output and input space of s_k, respectively.

For a given system $s \in S$, an experiment record $\lambda\,(t_0, t_N)$ is given by

$$\lambda\,(t_0, t_N) = \left\{ u(t_i),\ y(t_i),\ i=0,1,\ldots,N \right\}$$

Denote

$$Y = (y(t_1), y(t_2), \ldots, y(t_N)), \qquad y(t_i) \in Y$$

$$U = (u(t_1), u(t_2), \ldots, u(t_N)), \qquad u(t_i) \in U$$

$$x_i = (g_i\,(Y^{1-1}, U^{1-1}),\ g_i\,(Y^{2-1}, U^{2-1}), \ldots, g_i\,(Y^{t_N-1}, U^{t_N-1}))$$

$$X^k = (x_{k_1}, x_{k_2}, \ldots, x_{k_m}) \quad \subset \chi^k \qquad k, i \text{ and } m \in I,$$

So X^0, $\subset \chi^0$ corresponds to the rank of all elements of G^0.

II.2 Definition of Model Structure

Definition 1: each $g_i \in G^0$ and $i \in I$ is called an element of a model structure. G^0 is referred to as the element set of a model structure of the system s. Any possible linear combination of elements in G^0 is called a model structure, denoted by c_k

$$c_k : \chi^k \times \theta^k \rightarrow Y^k \qquad k \in I \quad , Y^k \subset \chi^k.$$

All c_k, $k \in I$, comprise the model structure set for the identification of the system s, denoted by C.

Definition 2: the matrix X^0 is called a realisation of G^0 conditioned on the given experiment record $\lambda\,(t_0, t_N)$. The corresponding linear space χ^0 spanned by X^0 is called the basic structure space. Each subspace χ^k is relevant to a realisation matrix X^k and a model structure c^k.

Definition 3: G^0 is said to be an independent set if all its elements are completely exclusive.

Definition 4: a subspace $\chi^k \subset \chi^0$ is said to be of rank r if its corresponding realisation matrix X^k has rank r. A subspace $\chi \subset \chi^0$ is said to be the maximum true subspace of χ^0 if

$$Rank\ \chi = Rank\ \chi^0 = r$$

and the rank of any true subspace of χ is smaller than r.

The realisation matrix X of the maximum true subspace of χ^0 is called the basis of the structure space χ^0.

Based on these definitions, the identification of the model structure can be stated as:

For a given system $s \in S$ with experiment record $\lambda\ (t_0, t_N)$, design an independent element set G^0 and establish the basis X. Then find an optimal subset G^* from G^0, such that the output Y is represented optimally by the structure space χ^* spanned by X^*.

Remark: the structure space χ^0 is a stochastic linear space. Its basis or each coordinate $\mathbf{x}_i$, $i \in I$, is also a stochastic vector. While the parameters here are dealt with as the coordinate variables.

III. Decision Rule for Model Structure Identification

III.1 The General Form of the Decision Rule

Consider a system $s \in S$, for which the relationship of the inputs and outputs obeys the probability distribution $\gamma(\lambda)$ and the probability of the system obeying a model $c^k \in C$ is realised by the structure space $\chi^k \subset \chi$ and described by $f(\lambda|c^k, \theta^k)$. The Kullback-Leiblern (K-L) mean information (Kullback 1951) of the system can then be represented by

$$I(\gamma, f(\lambda|c^k, \theta^k)) = E_\lambda \left\{ \log \frac{\gamma(\lambda)}{f(\lambda|c^k, \theta^k)} \right\} \tag{3.1}$$

The average log-likelihood is defined by

$$S(\lambda, f(\lambda|c^k, \theta^k)) = \int \log f(\lambda|c^k, \theta^k)\ \gamma(\lambda)\ d\lambda \tag{3.2}$$

The K-L mean information can be written as

$$I(\gamma, f(\lambda|c^k, \theta^k)) = S(\lambda, \lambda) - S(\lambda, f(\lambda|c^k, \theta^k)) \tag{3.3}$$

which has the following properties:

(i) $I(\gamma, f(\lambda|c^k, \theta^k)) \geq 0$

(ii) $I(\gamma, f(\lambda|c^k, \theta^k)) = 0$ iff $\forall c^* \in C$ and $\theta^* \in \Theta$ that $\gamma(\lambda) = f(\lambda|c^*, \theta^*)$

in terms of Lebesque measure.

where, c^* and θ^* define a model which describes the " true relationship " between the input and output of the system s.

These properties imply that minimisation of the K-L mean information would be a reasonable decision rule for model structure selection. This has been widely accepted in most statistical decision problems.

For a given measurable system, the first term in the Eq (3.3) is constant. Therefore, minimising $I(\gamma, f(\lambda | c^k, \theta^k))$ in Eq (3.3) is equivalent to maximising $S(\lambda, f(\lambda | c^k, \theta^k))$ in Eq (3.2). Moreover, in any one experiment, it is assumed that the measurement record $\lambda(t_0, t_N)$ is sufficient so that the ergodic property is satisfied. Such that the ensemble average of the log-likelihood in Eq (3.2) can be approximated consistently by the average in time with probability one. Thus, the decision rule for the model structure identification problem described in section II is simplified to finding a $c^* \in \mathbf{C}$ and $\theta^* \in \mathbf{\Theta}$ such that

$$J(f(c^*, \theta^*)) = \max_{\substack{\theta^k \in \mathbf{\Theta} \\ c^k \in \mathbf{C}}} J(f(c^k, \theta^k)) \tag{3.4}$$

where,

$$J(f(c^k, \theta^k)) = \frac{1}{N} \sum_{i=1}^{i=N} \log f(\lambda(t_i, t_N) | c^k, \theta^k) \tag{3.5}$$

III.2 The Decision Rule under Gaussian Distribution

Consider the stochastic variable w^k:

$$w^k = Y - Y^k \tag{3.5}$$

$$Y^k = X^k \theta^k \tag{3.6}$$

where, $Y \in \mathbf{Y}$; $Y^k \in \chi^k \subset \chi$ and $\theta^i \in \Theta$,

which gives the difference between the measured output space and the model structure space. It is known that if the model c^k is completely controllable and observable, the average log-likelihood function of the system s obeying the model structure c^k under the condition of $\lambda(t_0, t_N)$ can be approximated as

$$J(f(c^k, \theta^k)) = J(f(w^k)) \tag{3.7}$$

where,

$$J(f(w^k)) = \frac{1}{N} \sum_{i=1}^{i=N} \log f(w(i)^k) \tag{3.8}$$

assumption 1 Just the first and the second moments of the distribution of $f(w^k)$ are considered and $f(w^k)$ can be approached by the Gaussian distribution

$$f(w^k) \approx N(\overline{w}^k, R)$$

where,

$$\overline{w}^k \triangleq E\left\{ w^k \right\}$$

$$R^k \triangleq E\left\{ (w^k - \overline{w}^k)(w^k - \overline{w}^k)^T \right\}$$

assumption 2 the statistical properties of Y and X^k satisfy

$$\overline{w}^k = 0$$

$$R_y \triangleq E\left\{ (Y - EY)(Y - EY)^T \right\} = \rho_y \Xi$$

$$R_x^k \triangleq E\left\{ (X^k - EX^k)(X^k - EX^k)^T \right\} = diag(R_{x_i^k})$$

$$R_{x_i^k} = \rho_x \Xi$$

$$R_{yx} \triangleq E\left\{ (Y - EY)(X^k - EX^k)^T \right\} = 0$$

Therefore,

$$\begin{aligned} R^k &= R_y^k + R_{yx}^k \otimes \theta^k + R_x^k \otimes \theta^k \theta^{k^T} \\ &= (\rho_y + \theta^{k^T} R_x^k \theta^k)\Xi \end{aligned} \qquad (3.9)$$

where, $i,k \in I$ and Ξ is the unit matrix.

The decision rule defined in Eq(3.4) can be obtained by

$$\begin{aligned} D(\lambda) &= c^* \\ &= argument \left\{ \max_{\substack{\theta^k \in \Theta \\ c^k \in C}} J(f(c^k, \theta^k)) \right\} \\ &= argument \left\{ \min_{c^k \in C} J(f(c^k)) \right\} \end{aligned} \qquad (3.10)$$

$$J(f(c^k)) = \frac{1}{2} \log \det \hat{R}^k + \frac{1}{2N} \| Y - X^k \theta^* \|^2_{\hat{R}^{k^{-1}}} \qquad (3.11)$$

where,

$$\hat{R}^k = (\hat{\rho}_y + \theta^{k^T} \hat{R}_x^k \theta^k)\ \Xi \qquad (3.12)$$

$$\hat{\rho}_y = \frac{1}{N} \|Y - EY\|^2 \qquad (3.12)$$

$$\hat{R}_x^k = E\left\{(X^k - EX^k)^T (X^k - EX^k)\right\} = diag(\hat{\rho}_{x_1^k}, \hat{\rho}_{x_2^k}, \ldots, \hat{\rho}_{x_i^k}) \qquad (3.13)$$

$$\hat{\rho}_{x_i^k} = \frac{1}{N} \|X_i^k - EX_i^k\|^2 \qquad (3.14)$$

$$\theta^* = (X^{k^T} X^k)^{-1} X^{k^T} Y \qquad i,k \in I \qquad (3.15)$$

Comment 1 The assumption of a Gaussian distribution refers just to w^k. Nothing is assumed about the probability density of Y and Y^k.

Comment 2 For each functional structure in G^0, there is a corresponding realisation in the structure space.

Comment 3 The variances $\hat{\rho}_y$ and $\hat{R}_x^k$ can be computed independently from the parameter θ^*. Consequently, many effective filtering

and estimation algorithms can be used for this purpose.

Comment 4 Eq (3.15) is a basic formulation in regression analysis. Therefore, most techniques of regression analysis can be applied to the decision procedure.

IV. Discussion and Comparison

IV.1 The Principle of Parsimony

The principle of parsimony suggests a compromise choice between the residual variance and the number of parameters, which is stated as follows "given two models fitted to the same data with residual variance ρ_1 and ρ_2 which are close to each other, choose the model which involves the smaller number of parameters". In most current suggested criteria, like FPE, AIC, FPE_α and the Bayesian criterion given by Kashyap, this problem is approached by adding some function of the number of parameters to the criterion. This has proved to be useful in the determination of the order of time series models and transfer functions. But for most multivariable systems and nonlinear systems, the candidates of model structure for the identification decision are unnested. It is evident that just the variance of the residuals and the number of parameters is not sufficient for the decision concerning the model structure .

The decision rule proposed in this paper includes not only the dimension of the model structure and the residuals, but also the the variance of each coordinate of the structure basis.

When the variance of the basis for two model structures and the residuals are close, the model structure with the smaller parameter dimension will be chosen. If the dimension of the two model structures are equal and their residuals are close to each other, the model structure which involves the smaller structure variance will be selected. This property makes the criterion more effective to the structure decision of MIMO and nonlinear systems.

IV. 2 Robustness

The Robustness of a criterion reflects its ability to tolerate deviations which occur in practical applications from the theoretical assumptions and uncertainties. In AIC, the derivation was based on the assumption that the parameters are " very close to the true ones " and the aposteriori distribution is Gaussian. Similarly, in Kashyap's Bayesion criterion, a priori probability of the parameters was assumed. This limited the design of the structure space, especially for nonlinear systems. Whereas, no requirement on the probability of the parameters in the decision rule is suggested here. Almost all previous criteria used just one kind of information measure for goodness of fit, the residuals or the prediction errors. But in practice, especially for nonlinear systems, the residuals and prediction errors do not just include information about the model. They are often disturbed by measurement noise both in inputs and in outputs and the estimation error of the initial states. Consequently, the single measure for goodness of fit, either by residual or prediction error, may fail to give a reasonable assessment.

The decision rule suggested here includes and combines three kinds of information. The first is the residual, which measures the distance between the output space and the structure space. The second is the prediction errors, which measure the distance of the predicted model output to the output space. The third is the equation error, which gives the deviation of the predicted basis from the measured basis. These three pieces of information describe the behaviour of the system obeying the model in different directions and provide a coinciding measure of goodness of fit. This property makes this criterion more robust when the observed data are disturbed.

V. Numerical Examples

The simulation computation and practical utility of the methods are demonstrated in this section. The numerical comparative results of seven criteria are presented. These seven criteria are J_1 the variance of residual; J_2 the Cp statistic; J_3 the FPE test; J_4 the AIC test; J_5 the Bayesian aposteriori criterion; J_6 the variance of prediction errors; J_7 the new decision rule. The forward orthogonal regression algorithm and the one step backward orthogonal regression algorithm are used.

Example 1 : A linear dynamical system with coloured noise.

The simulated system model is

$$y(t) = 1.61y(t-1) - 1.61y(t-2) + 0.78y(t-3) + 1.2u(t-1) - 0.95u(t-2) + 0.2u(t-3)$$
$$+ e(t) + 0.1e(t-1) + 0.25e(t-2) + 0.87e(t-3)$$

Where, the noise e(t) is a Gaussian sequence with mean zero and variance one, (e(t) = N(0,1)). An ARMA model set with $d_y = 5$ and $d_u = 5$ is used. The results are shown in Fig. (1) and Table (1).

Example 2 A nonlinear dynamical system .

The simulated system model is

$$y(t) = 0.75y(t-1) + 0.22y(t-2) + 0.2u(t-1) - 0.8u(t-2)$$
$$- 0.024y(t-2)y(t-2)y(t-2) - 0.043y(t-1)y(t-1) - 0.19u(t-1)y(t-1)$$
$$- 0.016u(t-2)u(t-2)y(t-2) - 0.016u(t-1)u(t-1)y(t-1)$$
$$- 0.15u(t-2)y(t-2)y(t-2) - 0.16u(t-2)u(t-2)u(t-2) + \eta(t)$$
$$\eta(t) = e(t) + 0.1e(t-1)$$

Where, the noise e(t) is a Gaussian sequence with mean zero and variance one, (e(t) = N(0, 0.2)). A polynomial NARMAX (Nonlinear AutoRegression and Moving Average eXogenous) model set with $l = 3$ (l is the maximum order of the polynomial set), $d_y = 2$ and $d_u = 2$ was used. The results are shown in Fig. (2) and Tables (2) and (3) respectively.

Example 3 : The empirical data used here was collected from an experiment on a pilot scale liquid level system. The model set is the same as that in the example 2. The first 700 recorded data were used for model structure identification and the first 500 records were used for parameter estimation. The results are shown in Table (4) - Table (7) and Fig. (3) - Fig. (10) respectively.

VI. Conclusions

A new decision rule for model structure identification has been developed based on the Kullback-Leibler mean information. Some analytical and simulated comparisons of this decision rule with Akaike's FPE and AIC, Mallows's statistical Cp plot and Bayes aposteriori decision rule have been given. The results suggest that this new criterion may be superior in the case of disturbed inputs and outputs, especially for nonlinear systems. Since the identification procedure based on this new proposed decision rule can be implemented without the aid of any subjective judgement, its application to problems in nonlinear system identification should simplify the often complex procedure of determining the model structure.

Reference

H. Akaike, "A new look at the statistical model identification "; IEEE, AC-19, pp 716-723, 1974.

S. A. Billings and W. S. F. Voon, " Correlation based model validity tests for non-linear models "; INT. J. Control, 44, pp 235-244, 1986.

J. R. M. Hosking, " The asymptotic distribution of R^2 for autoregressive-moving average time series model when parameter are estimated "; Biometrika, 66, pp 166-167, 1979.

R. L. Kashyap, "A bayesian comparison of different classes of dynamic models using empirical data "; IEEE AC-22, pp 715-727, 1977.

A S. Kullback and R. A. Leibler "On information and sufficiency", Ann. Math. 22, 1951.

I. J. Leontaritis and S. A. Billings, " Model selection and validation methods for nonlinear systems "; INT. J. Control, 45, pp 311-341, 1987.

C. L. Mallows, "Some comments on Cp "; Technometrics, 15, pp 661-675, 1973.

Y. Sakamoto and M. Ishigouro and G. Ktagawa, " Akaike information criterion statistics"; KTK Scientific Publisher/Tokyo, 1986

R. Shibata, " Approximate efficiency of a selection procedure for the number of regression variables "; Biometrika, 71, pp 43-49, 1984.

T. Söderström, " On model structure testing in system identification "; INT. J. Control, 26, pp 1-18 1977.

table 1. Decision Results for Example 1

J_1	J_2	J_3	J_4	J_5	J_6	J_7
y(t-1)	y(t-1)	y(t-1)	y(t-1)	y(t-1)	y(t-1)	y(t-1)
y(t-2)	y(t-2)	y(t-2)	y(t-2)	y(t-2)	y(t-2)	y(t-2)
y(t-3)	y(t-3)	y(t-3)	y(t-3)	y(t-3)	y(t-3)	y(t-3)
u(t-1)	u(t-1)	u(t-1)	u(t-1)	u(t-1)	u(t-1)	u(t-1)
u(t-2)	u(t-2)	u(t-2)	u(t-2)	u(t-2)	u(t-2)	u(t-2)
u(t-3)	u(t-3)	u(t-3)	u(t-3)	u(t-3)	u(t-3)	u(t-3)
u(t-5)	u(t-5)	u(t-5)	u(t-5)	u(t-5)		u(t-5)
y(t-4)	y(t-4)	y(t-4)	y(t-4)	y(t-4)		
y(t-5)	y(t-5)	y(t-5)	y(t-5)	y(t-5)		
u(t-4)	-					

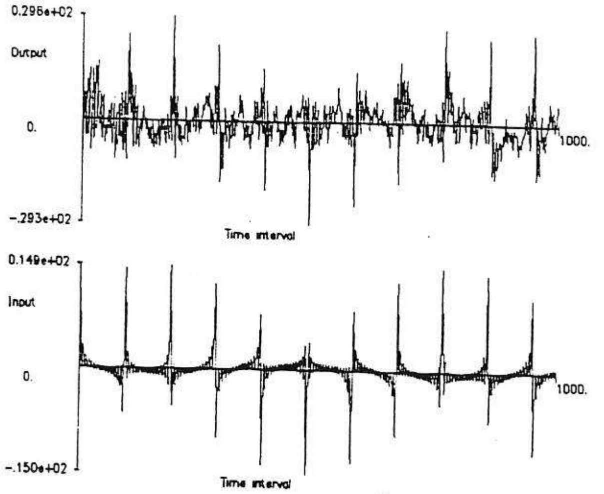

fig. 1. input and output of example 1.

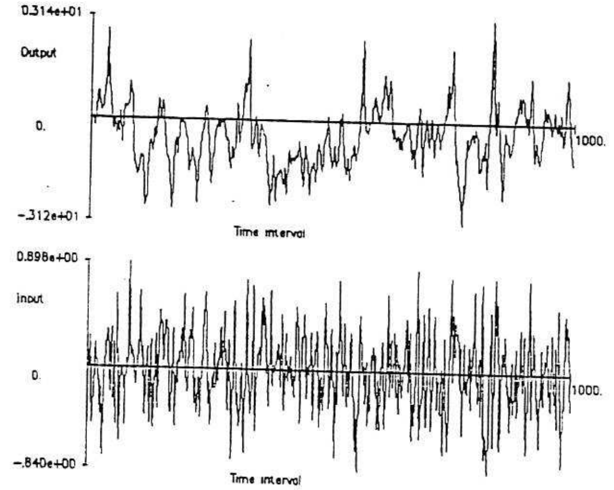

fig. 2. input and output of example 2.

table 3. Decision Results for Example 2

decision rule	model structure
J_1	1 4 11 6 28 19 3 2 18 5 29 22 30 13 12 21 7 14 32 23 25 20 31 27 15 16 17 24 26 8 10 33 9 34
J_2	1 4 11 6
J_3	1 4 11 6 28 19 3 2 18 5 29 22 30
J_4	1 4 11 6
J_5	1 4
J_6	1 4 11 6 28 19 3 2 5 18 29
J_7	1 4 11 28 19 3 2 5 18 29

table 2. Element Set of Model Structure of example 2

1 y(t-1)	8 y(t-1)y(t-1)y(t-1)	15 u(t-1)u(t-1)u(t-1)	22 y(t-2)u(t-2)	29 u(t-1)u(t-1)y(t-1)
2 y(t-2)	9 y(t-1)y(t-1)y(t-2)	16 u(t-1)u(t-1)u(t-2)	23 y(t-1)y(t-1)u(t-1)	30 u(t-1)u(t-1)y(t-2)
3 u(t-1)	10 y(t-1)y(t-2)y(t-2)	17 u(t-1)u(t-2)u(t-2)	24 y(t-1)y(t-1)u(t-2)	31 u(t-1)u(t-2)y(t-1)
4 u(t-2)	11 y(t-2)y(t-2)y(t-2)	18 u(t-2)u(t-2)u(t-2)	25 y(t-1)y(t-2)u(t-1)	32 u(t-1)u(t-2)y(t-2)
5 y(t-1)y(t-1)	12 u(t-1)u(t-1)	19 y(t-1)u(t-1)	26 y(t-1)y(t-2)u(t-2)	33 u(t-2)u(t-2)y(t-1)
6 y(t-1)y(t-2)	13 u(t-1)u(t-2)	20 y(t-1)u(t-2)	27 y(t-2)y(t-2)u(t-1)	34 u(t-2)u(t-2)y(t-2)
7 y(t-2)y(t-2)	14 u(t-2)u(t-2)	21 y(t-2)u(t-1)	28 y(t-2)y(t-2)u(t-2)	

table 4. Decision Results for Example 3

decision rule	model structure
J_1	1 3 4 11 19 28 2 22 6 18 30 21 20 31 5 14 13 24 17 33 16 15 12 27 25 26 10 23 9 8 29 34 7 32
J_2	1 3 4 11 19 28 2 22 6
J_3	1 4 11 6 28 19 3 2 5 18 29
J_4	1 3 4 11 19 28 2 22 6 18 30 21 20 31 5 14 13 24 17 33 16 15 12 27 25 26 10 23 9 8
J_5	1 3 4 11 19 28
J_6	1 3 4 11 19 28 2 22 6 18 30 21 20 31 5 14 13 24 17 33 16 15 12 27 25 26 10 23 9 8
J_7	1 3 4 11 19 28 2 6 18 30 21

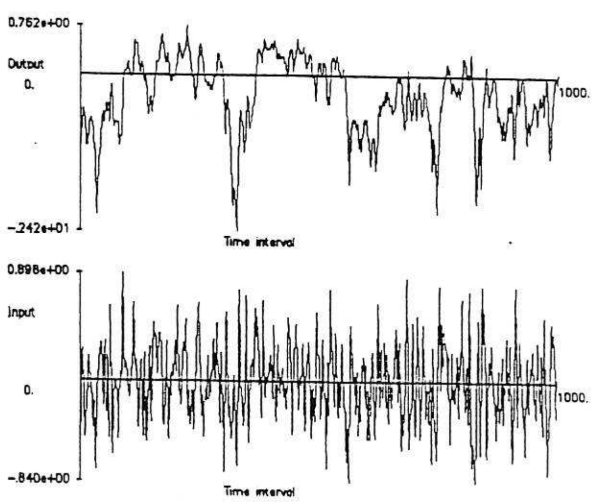

fig. 3. input and output of example 3.

table 5. model decided by criterion J_7

Terms	Estimates	ERR	Stdev
y(t- 1)=	0.7835e+00	(0.974e+02)	(0.495e-01)
u(t- 1)=	0.4013e+00	(0.144e+01)	(0.157e-01)
u(t- 2)=	-.7240e-01	(0.271e+00)	(0.192e-01)
y(t- 2)*y(t- 2)*y(t- 2)=	-.2281e-01	(0.127e+00)	(0.610e-02)
y(t- 1)*u(t- 1)=	-.2466e+00	(0.593e-01)	(0.544e-01)
y(t- 2)*y(t- 2)*u(t- 2)=	-.1362e+00	(0.133e+00)	(0.147e-01)
y(t- 2)=	0.1941e+00	(0.996e-02)	(0.484e-01)
y(t- 1)*y(t- 2)=	-.3755e-01	(0.209e-01)	(0.105e-01)
u(t- 2)*u(t- 2)*u(t- 2)=	-.1599e+00	(0.203e-01)	(0.405e-01)
y(t- 2)*u(t- 1)=	0.4674e-01	(0.821e-03)	(0.546e-01)

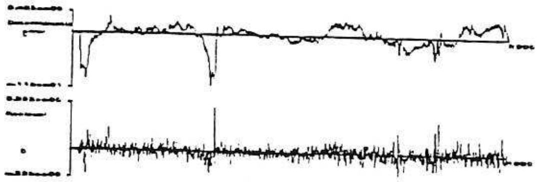

fig. 5. prediction errors and residuals by model J_7

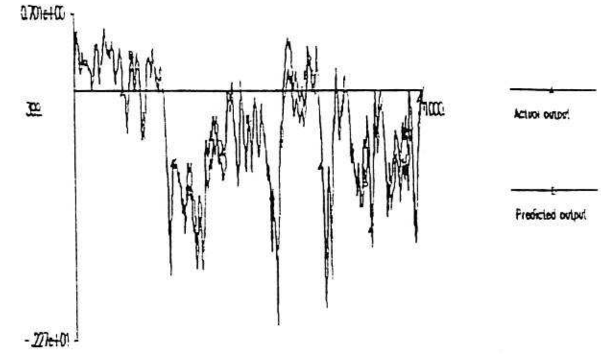

fig. 6. predicted output superimposed on actual output using model from J_7

table 6. model decided by criterion J_4

Terms	Estimates	ERR	Stdev
y(t- 1)=	0.7661e+00	(0.974e+02)	(0.501e-01)
u(t- 1)=	0.3574e+00	(0.144e+01)	(0.160e-01)
u(t- 2)=	-.1037e+00	(0.271e+00)	(0.165e-01)
y(t- 2)*y(t- 2)*y(t- 2)	-.2787e-01	(0.127e+00)	(0.600e-02)
y(t- 1)*u(t- 1)=	-.2252e+00	(0.593e-01)	(0.224e-01)
y(t- 2)*y(t- 2)*u(t- 2)	-.1220e+00	(0.133e+00)	(0.172e-01)
y(t- 2)=	0.2104e+00	(0.996e-02)	(0.486e-01)
y(t- 2)*u(t- 2)=	0.6306e-01	(0.136e-02)	(0.260e-01)
y(t- 1)*y(t- 2)=	-.4906e-01	(0.264e-01)	(0.103e-01)

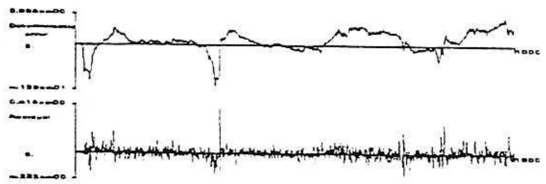

fig. 7. prediction errors and residuals by model J_4

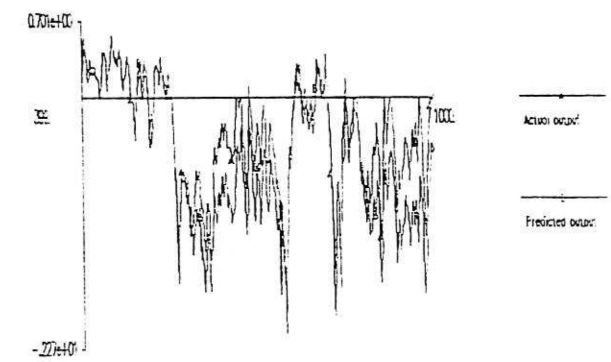

fig. 8. predicted output superimposed on actual output using model from J_4

table 7. model decided by criterion J_3

Terms	Estimates	ERR	Stdev
y(t- 1)=	0.7705e+00	(0.974e+02)	(0.698e-01)
u(t- 1)=	0.3956e+00	(0.144e+01)	(0.275e-01)
u(t- 2)=	-.9727e-01	(0.271e+00)	(0.267e-01)
y(t- 2)*y(t- 2)*y(t- 2)=	-.6209e+00	(0.127e+00)	(0.591e+00)
y(t- 1)*u(t- 1)=	-.5284e+00	(0.593e-01)	(0.386e+00)
y(t- 2)*y(t- 2)*u(t- 2)=	-.1402e+01	(0.133e+00)	(0.102e+01)
y(t- 2)=	0.2027e+00	(0.996e-02)	(0.677e-01)
y(t- 2)*u(t- 2)=	-.4432e+00	(0.136e-02)	(0.371e+00)
y(t- 1)*y(t- 2)=	0.2751e+00	(0.264e-01)	(0.167e+00)
u(t- 2)*u(t- 2)*u(t- 2)=	0.2607e+00	(0.165e-01)	(0.237e+00)
u(t-.1)*u(t- 1)*y(t- 2)=	-.2526e+00	(0.804e-02)	(0.123e+00)
y(t- 2)*u(t- 1)=	0.2654e+00	(0.405e-02)	(0.382e+00)
y(t- 1)*u(t- 2)=	0.6200e+00	(0.316e-02)	(0.387e+00)
u(t- 1)*u(t- 2)*y(t- 1)=	0.2762e+00	(0.545e-03)	(0.262e+00)
y(t- 1)*y(t- 1)=	-.3226e+00	(0.126e-01)	(0.172e+00)
u(t- 2)*u(t- 2)=	-.1016e+00	(0.161e-02)	(0.841e-01)
u(t- 1)*u(t- 2)=	0.2581e-03	(0.470e-02)	(0.132e+00)
y(t- 1)*y(t- 1)*u(t- 2)=	-.1655e+01	(0.366e-05)	(0.112e+01)
u(t- 1)*u(t- 2)*u(t- 2)=	-.6785e+00	(0.282e-02)	(0.560e+00)
u(t- 2)*u(t- 2)*y(t- 1)=	-.9449e-01	(0.200e-03)	(0.152e+00)
u(t- 1)*u(t- 1)*u(t- 2)=	0.4365e+00	(0.286e-02)	(0.543e+00)
u(t- 1)*u(t- 1)*u(t- 1)=	0.6074e-01	(0.604e-03)	(0.222e+00)
u(t- 1)*u(t- 1)=	0.2019e-01	(0.326e-02)	(0.799e-01)
y(t- 2)*y(t- 2)*u(t- 1)=	-.1328e+01	(0.788e-02)	(0.943e+00)
y(t- 1)*y(t- 2)*u(t- 1)=	0.2417e+01	(0.127e-02)	(0.201e+01)
y(t- 1)*y(t- 2)*u(t- 2)=	0.3013e+01	(0.143e-02)	(0.209e+01)
y(t- 1)*y(t- 2)*y(t- 2)=	0.2561e+01	(0.148e-02)	(0.194e+01)
y(t- 1)*y(t- 1)*u(t- 1)=	-.1183e+01	(0.118e-03)	(0.111e+01)
y(t- 1)*y(t- 1)*y(t- 2)=	-.3354e+01	(0.313e-03)	(0.223e+01)
y(t- 1)*y(t- 1)*y(t- 1)=	0.1392e+01	(0.254e-02)	(0.877e+00)

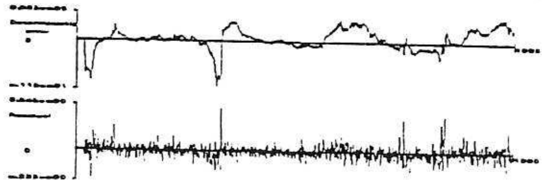

fig. 9. prediction errors and residuals by model J_3

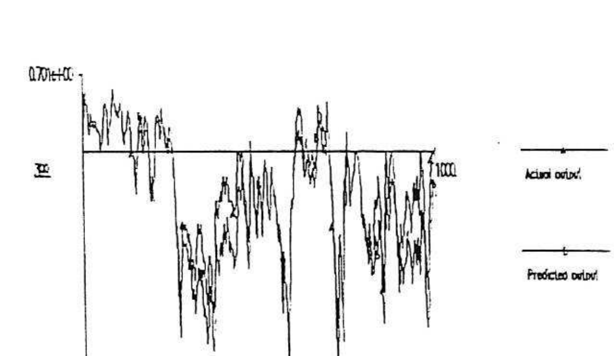

fig. 10. predicted output superimposed on actual output using model from J_3