MODEL UNCERTAINTY FOR IDENTIFICATION WITH FINITE DATA

Graham GOODWIN*, Michel GEVERS†, Brett NINNESS‡

Abstract

We assume that the unmodelled dynamics of a system are a realisation of a random process with parameterised second order properties. We show that these parameters may be estimated from the data. We show how these second order properties may then be used for model order evaluation of uncertainty bounds.

1 Introduction

The need of robust control designers for model descriptions including uncertainty bounds is a major theoretical challenge in identification. The bounds required account for two error terms. The first one, often called the bias error, is caused by the fact that the parametrized model structure is, at best, a simplified (low order) version of the true system. The second component, often called variance error, is caused by the noise in the data that make up the particular realization that is used for identification.

The key tool used for the computation of variance errors is the Cramér-Rao lower bound on the estimated parameters. In the case of exact model structure, this tool has produced reasonable variance error expressions for the estimated transfer fuctions : see e.g.[1], [2]. This variance error typically decreases like $\frac{1}{N}$, when N is the number of data.

The characterization of the bias error on the estimated transfer function is much more difficult. A first result, due to Wahlberg and Ljung [3] produces an implicit characterization of the asymptotic bias. Whereas it yields useful insights into input design problems, it does not produce explicit bias error bounds. Recent work has focused on producing such explicit error bounds : [4]-[5]. However, the results so far are based on a complete prior specification of the characteristics of the unmodelled dynamics. These characteristics typically involve assumptions on the magnitude and smoothness of the variation of the transfer function, $G_\Delta(e^{j\omega})$, of the unmodelled dynamics in the frequency domain.

The contributions of this paper are twofold. First, following the tradition of [2],[6] , we adopt a stochastic embedding approach. However, a novel feature is that we show how the parameters of these prior distribution functions can be estimated from the data by maximum likelihood. We also compute the Cramer-Rao bound for these estimates.

Our second contribution is to show that the above procedure leads naturally to a method for optimal model order selection with finite data. The optimal order is obtained by minimizing some suitable criterion of the total mean square error between the true transfer function $G_T(e^{j\omega})$ and the estimated model $G(e^{j\omega}, \hat{\theta}_N)$ based on N data.

We show that, in the presence of undermodelling and with finite data, this new criterion performs better than the classical Final Prediction Error (FPE) and AIC criteria.

We should like to make it very clear that, contrary to popular beliefs, with finite noisy data the optimal model order is typically smaller than the "exact" model order if such an exact order exists, and that the traditional quest for a true model order on the basis of finite data is a misguided pursuit.

2 Model Assumptions

We shall consider discrete-time single-input single-output systems for simplicity, and we shall assume throughout that the true system is described by

$$y_t = G_T(q)u_t + v_t \qquad (1)$$

where q is the forward shift operator, $G_T(q)$ is the "true" transfer function and v_t is additive

*Department of Electrical Engineering, University of Newcastle, Shortland AUSTRALIA

†Laboratoire d'Automatique, Dynamique et Analyse des Systemes, UCL, 3 place du Levant, B1348 Louvain-la-Neuve, BELGIUM

‡Thanks to the Australian Telecommunications and Electronics Research Board for scholarship assistance

$$G_T(q) = G(q,\theta) + G_\Delta(q). \quad (2)$$

We shall make the following assumptions concerning $G(q,\theta)$, $G_\Delta(q)$, and v_t.

Assumptions concerning the nominal model

The nominal model is a finite linear combination of known rational basis functions,

$$G(q,\theta) = B(q)\theta \quad (3)$$

with

$$B(q) = [B_1,(q),\ldots,B_p(q)] \quad (4)$$

$$\theta = (\theta_1,\ldots,\theta_p)^T \quad (5)$$

The vector $\theta \in \mathbb{R}^l$ is unknown but fixed.

Assumptions on the unmodelled dynamics. The transfer function $G_\Delta(q)$ of the unmodelled dynamics is considered as a particular realization of a zero mean random variable having a probability density function $f_\Delta(\beta)$ with known structure, but parametrized by an unknown vector β. Without loss of generality, $G_\Delta(q)$ can be written as

$$G_\Delta(q) = \sum_1^\infty \eta_k q^{-k} \quad (6)$$

The important point is that we shall not attempt to estimate G_Δ (i.e. we shall not estimate a large number of parameters η_k) : this would be modelling the unmodelled dynamics. Rather we shall estimate the low-dimensional parameter vector β that describes the properties of G_Δ (eg. magnitude and smoothness).

An example of a suitable model for the characteristics of the unmodelled dynamics would be to assume that $G_\Delta(e^{j\omega})$ is a zero mean Gaussian process with covariance function:

$$E\{G_\Delta(e^{j\omega_1})G_\Delta(e^{-j\omega_2})\} = \frac{\alpha e^{j\omega}}{e^{j\omega}-\lambda} \quad ;\omega \triangleq \omega_1 - \omega_2. \quad (7)$$

This corresponds to the time-domain assumption that the impulse response coefficients η_k are independent but non identically distributed random variables drawn from a zero mean Gaussian distribution having variance:

$$E\{\eta_k^2\} = \alpha\lambda^k \quad \alpha \in \mathbb{R}^+, \lambda \in (0,1) \quad (8)$$

We stress that the stochastic embedding of the unmodelled $G_\Delta(q)$ is just one of several ways of imposing some smoothness assumptions on the variation of $G_\Delta(e^{j\omega})$ with ω. So the model (8) abilistic model is characterized by only two parameters $\beta^T = (\alpha,\lambda)$ whereas the impulse response itself, $G_\Delta(q)$, is possibly infinite dimensional.

Assumption on the noise

The additive noise $\{v_t\}$ is a zero mean stochastic process whose probability density function $f_{v_t}(\gamma,t)$ is also assumed to be known except for some unknown finite parameter vector γ. The conceptual similarity of this assumption to that used for the unmodelled dynamics should be noted. A special case considered in the examples later is summarized as:

Assumption A

$G_\Delta(e^{j\omega})$ is a zero mean Gaussian process, stationary in the frequency domain, nonstationary in the time-domain, described by (7). $\{v_t\}$ is independent of $G_\Delta(e^{j\omega})$ and is an independent identically distributed zero mean Gaussian process with variance σ^2.

3 Estimation of the nominal model and the characteristics of the residual

In this section we show how to estimate the parameter vector θ of the nominal model and the parameter vectors β and γ that parametrize the probability density functions of the unmodelled dynamics and of the noise, respectively, on the basis of a finite set of input and output data. These parameters can in principle be estimated jointly by maximizing the likelihood function of the data, which is parametrized by θ,β and γ. However, we shall instead estimate θ by Least Squares; this is closer to people's familiar way of estimating a nominal model. We shall then use the residuals as our new data for the maximum likelihood estimation of β and γ. We shall consider that N output data $y_1,\ldots,y_N$ are available and we shall henceforth assume that this data length N is larger than the settling time of the unmodelled transfer function $G_\Delta(q)$, i.e.

$$G_\Delta(q) = \sum_1^L \eta_k q^{-k},\ L \le N. \quad (9)$$

We shall further assume that the input signal and is available from some sufficient time instant in the past (compatible with the settling time of G_Δ and of the basis functions) up to time N. Taking into account (1), (2), (3) and (9), we can then write the input-output relations in vector

$$
\begin{aligned}
Y &= (y_1, \ldots, y_N)^T && (11)\\
V &= (v_1, \ldots, v_N)^T && (12)\\
\Phi &= [\phi_{ij}] \text{ with } \phi_{ij} \triangleq B_j(q)u_i && (13)\\
\Psi &= [\psi_{ij}] \text{ with } \psi_{ij} \triangleq q^{-j}u_i = u_{i-j} && (14)\\
\eta &= [\eta_1, \ldots, \eta_L]^T && (15)
\end{aligned}
$$

and $\theta \in \mathbb{R}^{\text{l}}$ is as in (5). Here $B_j(q)u_i$ is to be interpreted as the operator $B_j(q)$ acting on u_i. The Least Squares (LS) estimate of θ is

$$\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T Y \qquad (16)$$

with N-vector of residuals:

$$
\begin{aligned}
\varepsilon &\triangleq Y - \Phi\hat{\theta} && (17)\\
&= [I - \Phi(\Phi^T\Phi)^{-1}\Phi^T]Y \triangleq PY. && (18)
\end{aligned}
$$

The matrix in (18) has rank N-p. Therefore ε has a singular distribution of rank N-p. To obtain a new full rank data vector, we respresent ε in a new coordinate system that forms a basis for the space orthogonal to the columns of Φ.

Let R be any matrix whose columns span the subspace orthogonal to the columns of Φ. One way of constructing such R is to take any N-p independent linear combinations of the columns of P. Now define $Z \in \mathbb{R}^{\mathbb{N}-\text{l}}$ as follows :

$$Z \triangleq R^T\varepsilon. \qquad (19)$$

Now Z has a nonsingular distribution and, by the construction of R,

$$Z = R^T Y = R^T\Psi\eta + R^T V. \qquad (20)$$

Since R^T and Ψ depend on the input signal only, we observe that Z is the sum of two independent random vectors whose probability density functions are computable functions of the unknown parameter vectors β and γ. We can therefore compute the probability density function of Z, conditioned on the input data vector U, and on $\xi^T \triangleq (\beta^T, \gamma^T)$. We denote the corresponding likelihood function by $L(\xi \mid Z, U)$. Maximimizing this likelihood function yields the desired estimate for the unknown parameters :

$$\max_{\xi} L(\xi \mid Z, U) \Rightarrow \hat{\xi} \qquad (21)$$

4 Computation of uncertainty bounds

In this section we first show that the error in the transfer function estimate, $G_T(e^{j\omega}) - G(e^{j\omega}, \hat{\theta})$, replacing β and γ by their maximum likelihood estimates obtained in the previous section, we can therefore compute an estimate of the p.d.f. of $G_T(e^{j\omega}) - G(e^{j\omega}, \hat{\theta})$, and in particular of its mean square error over the ensemble of possible realizations. From (2), (3) and (9) we have

$$G_T(e^{j\omega}) - G(e^{j\omega}, \hat{\theta}) = B(e^{j\omega})(\theta - \hat{\theta}) + \Omega(e^{j\omega})\eta, \qquad (22)$$

where

$$\Omega(e^{j\omega}) \triangleq [e^{-j\omega}, \ldots, e^{-jL\omega}] \qquad (23)$$

It also follows from (16) and (10) that

$$\theta - \hat{\theta} = M\Psi\eta + MV \qquad (24)$$

where

$$M \triangleq -(\Phi^T\Phi)^{-1}\Phi^T. \qquad (25)$$

Combining (22) and (24) yields

$$
\begin{aligned}
\tilde{G}(e^{j\omega}) &\triangleq G_T(e^{j\omega}) - G(e^{j\omega}, \hat{\theta}) && (26)\\
&= [\Omega(e^{j\omega}) + B(e^{j\omega})M\Psi]\eta + B(e^{j\omega})MV. && (27)
\end{aligned}
$$

We note that Ω and B are known functions of ω , while M and Ψ are known functions of the input signal. Therefore, since η and V are two independent random vectors whose p.d.f., $f_\Delta(\beta)$ and $f_v(\gamma)$, are known up to the parameter vectors β and γ, one can compute the p.d.f. of $\tilde{G}(e^{j\omega})$ and compute such quantities as confidence regions, etc... In particular, when the Gaussian conditions of assumption A hold we have: have the following result.

$$
\begin{aligned}
V(\omega) &\triangleq E\{\mid G_T(e^{j\omega}) - G(e^{j\omega}, \hat{\theta}) \mid^2\} = && (28)\\
&[\Omega(e^{j\omega}) + B(e^{j\omega})M\Psi] \times C_\eta(\beta)[\Omega(e^{j\omega}) + && (29)\\
&B(e^{j\omega})M\Psi]^* + B(e^{j\omega})MC_v(\gamma)M^T B^*(e^{j\omega}) && (30)
\end{aligned}
$$

M is defined by (25), Φ by (13), Ψ by (14), $*$ denotes conjugate transpose and

$$C_\eta(\beta) = diag\{\alpha\lambda, \ldots, \alpha\lambda^L\} \qquad (31)$$

An estimate of the Mean Square Error of $G(e^{j\omega}, \hat{\theta})$ is then obtained by replacing β and γ by their maximum likelihood estimates obtained as in section 3.

The expression (30) produces bounds on the amplitude of the error in the transfer function estimate. Alternatively, from (27) one can compute the 2×2 covariance matrix of:

$$\tilde{g}(e^{j\omega}) \triangleq \begin{bmatrix} Re\ \tilde{G}(e^{j\omega}) \\ Im\ \tilde{G}(e^{j\omega}) \end{bmatrix}. \qquad (32)$$

For Gaussian distributions, say, one can then compute a confidence ellipse at each ω around the Nyquist diagram of $G(e^{j\omega}, \hat{\theta})$. This will be illustrated in the simulations of Section 7.

Denote the model structures $M_1(\theta_1), \ldots, M_r(\theta_r)$. Note that $\theta_1, \theta_2, \ldots, \theta_r$ may or may not have different dimensions. For each model structure $M_i(\theta_i)$ one can estimate $\hat{\theta}_i$ by Least Squares. With the assumed prior distribution for $G_\Delta(q)$ and for v_t , we can then compute, for each estimated nominal model, the corresponding maximum likelihood estimates of β and γ. We shall denote by $\hat{\beta}_i$ and $\hat{\gamma}_i$ the estimates coresponding to $M_i(\hat{\theta}_i)$, and by $\hat{V}_i(\omega)$ the estimate of the Mean Square Error (30) in which $C_\eta(\beta)$ and $C_v(\gamma)$ are replaced by $C_\eta(\hat{\beta}_i)$ and $C_v(\hat{\gamma}_i)$. Note that $\hat{V}_i(\omega)$ depends on the particular model $M_i(\hat{\theta}_i)$ in two ways : through the particular choice of basis vector $B(e^{j\omega})$ that has been selected in model M_i, and through the parameter vectors $\hat{\beta}_i$ and $\hat{\gamma}_i$.

To select among the r candidate models, we shall now consider any one of the following three criteria.

$$J_i^1 = \sup_i \hat{V}_i(\omega),\ i \in [1, r] \tag{33}$$

$$J_i^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{V}_i(\omega)d\omega,\ i \in [1, r] \tag{34}$$

$$J_i^3 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{V}_i(\omega)S_u(\omega)d\omega,\ i \in [1, r] \tag{35}$$

Here $S_u(\omega)$ denotes the power spectral density of a possibly new input sequence to which the model will be applied. The three criteria obviously cover three different applications in which the model may be used. Other criteria can easily be formulated. Depending on the application (or on one's favourite criterion) the optimal structure will be obtained as M_{i*} where

$$i^* = \arg \min_{i=1,\ldots,r} J_i^k \qquad k = 1, 2 \text{ or } 3. \tag{36}$$

We now consider the situation where the family of nominal models is a sequence of models generated from an increasing sequence of basis functions $B_1(q), B_2(q), \ldots$. The nominal models are then of the form $\sum_1^p B_i(q)\theta_i$ with $p = 1, 2, \ldots$, and the problem of structure selection becomes one of model order selection.

We shall consider the special case where the additive noise v_t is white with variance σ^2, and where the model is intended to be used on the same input data as was used for identification. For this special case we consider the optimal model order selection problem for a criterion which is a slight modification of J_3, and which we shall call the Generalized Information Criterion nominal model. For this special case $\hat{V}(\omega)$ reduces to (see (30)):

$$\hat{V}_p(\omega) = (\Omega + BM\Psi)C_\eta(\hat{\beta}_p)(\Omega + BM\Psi)^* + \tag{38}$$

$$\hat{\sigma}^2 tr[(\Phi^T\Phi)^{-1}B^*B] \tag{39}$$

By Parseval's theorem,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}(B^*B)S_u(\omega)d\omega = \frac{1}{N}(\Phi^T\Phi). \tag{40}$$

Therefore

$$GIC(p) = \hat{\sigma}^2 + \frac{p}{N}\hat{\sigma}^2 + \tag{41}$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}(\Omega + BM\Psi)C_\eta(\hat{\beta}_p)(\Omega + BM\Psi)^*d\omega \tag{42}$$

Akaike considers a similar criterion in his Final Prediction Error (FPE) test for model order selection [9]. His FPE criterion differs from our $GIC(p)$ criterion in two ways. First, FPE does not contain a term due to undermodelling. Second, in our GIC criterion, the value used for $\hat{\sigma}^2$ is obtained independently of the particular nominal model under consideration. Akaike's criterion explicitly depends on obtaining a new estimate of σ^2, denoted $\hat{\sigma}_p^2$, for each model dimension using

$$\hat{\sigma}_p^2 = \frac{1}{N-p}(Y - \Phi\hat{\theta}_p)^T(Y - \Phi\hat{\theta}_p). \tag{43}$$

This leads to the FPE criterion :

$$FPE(p) = \frac{N+p}{N-p} \times \frac{1}{N}(Y - \Phi\hat{\theta}_p)^T(Y - \Phi\hat{\theta}_p). \tag{44}$$

The estimate (43) would be an unbiased estimate of σ^2 if there were no undermodelling. Our rationale for using a high-dimensional model for the estimation of σ^2 is to ensure that undermodelling does not affect our estimate. Our criterion explicitly and, we believe, correctly accounts for undermodelling through the third term in (42). Indeed,

$$Y - \Phi\hat{\theta}_p = P(\Psi\eta + V) \tag{45}$$

Substituting in (44) and taking the expected value w.r.t. the noise v_t, assuming the undermodelling to be a deterministic quantity, yields

$$E_v\{FPE(p)\} = \tag{46}$$

$$= \sigma^2 + \frac{p}{N}\sigma^2 + \frac{N+p}{N-p} \times \frac{1}{N} \times tr\{P\Psi\eta\eta^T\Psi^T P\} \tag{47}$$

Comparing with (42) shows that the FPE criterion *on average* captures the variance effects correctly. However, the bias term is incorrectly scaled by a factor $\frac{N+p}{N-p}$.

In this section, we specialize our results to the case where the unmodelled dynamics and the noise satisfy the conditions of Assumption A, where the nominal models are Finite Impulse Response models, $FIR(p)$, and where the input sequence is deterministic and has the following orthogonality property :

$$\frac{1}{N}(\Psi^T\Psi) = \sigma_u^2 I \tag{48}$$

where σ_u^2 is the input power. Without loss of generality, we take $L = N$ (see (9)). In line with our developments of Section 3 we let R be the last $N - p$ columns of Ψ Note also that, by (48), $R^T\Phi = 0$.

The prediction errors $\{\varepsilon_k\}$ in this case thus form an iid sequence due to the orthogonality of the input with

$$\sigma_\varepsilon^2 = N\sigma_u^2\alpha\sum_{k=1}^{L}\lambda^k + \sigma_0^2 \tag{49}$$

The asymptotic properties of the estimates follow from the well known work of Wald:

$$\hat{\alpha}_N \xrightarrow{a.s.} \alpha \quad \hat{\lambda}_N \xrightarrow{a.s.} \lambda \quad \hat{\sigma^2}_N \xrightarrow{a.s.} \sigma^2 \tag{50}$$

But more interestingly, if we define

$$\beta^T = \left[\sqrt{\frac{-\ln N}{\ln\lambda_0}}\frac{\alpha}{\alpha_0}, \left(\frac{-\ln N}{\ln\lambda_0}\right)^{3/2}\frac{\lambda}{\lambda_0}, \sqrt{N}\frac{\sigma^2}{\sigma_0^2}\right] \tag{51}$$

then

$$\sqrt{N}(\hat{\beta}_N - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P) \tag{52}$$

with

$$P = 2\begin{bmatrix} 4 & -6 & 0 \\ -6 & 12 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{53}$$

So that

$$\mathrm{Var}\left(\frac{\hat{\alpha}_N}{\alpha_0}\right) = \mathcal{O}\left(\frac{1}{N\ln N}\right) \tag{54}$$

$$\mathrm{Var}\left(\frac{\hat{\lambda}_N}{\lambda_0}\right) = \mathcal{O}\left(\frac{1}{(N\ln N)^3}\right) \tag{55}$$

A proof of these results may be found in [10].

fitted with white noise input for $\{u_k\}$

$$\frac{e^{-\Delta s}}{\tau s + 1} \qquad \tau = 0.1, \quad \Delta = 5$$

The output sequence $\{y_k\}$ was disturbed by $N(0, 0.005)$ distributed white noise. The following calculations were made:

1. The parameters $\alpha, \lambda, \sigma^2$ were estimated.
2. These estimates were used to evaluate the integral of the squared frequency response estimation error both with and without weighting by the input power spectrum. This was done for a range of nominal FIR model orders.
3. The true mean square errors for various model orders were calculated for comparison with the estimates in 2, using the known $G_T(e^{j\omega})$. This would, of course, be impossible in practice.
4. Akaike's AIC and FPE model order determination criteria [8]-[9] were calculated from the data.

Typical results for these calculations are shown in Figure 1.

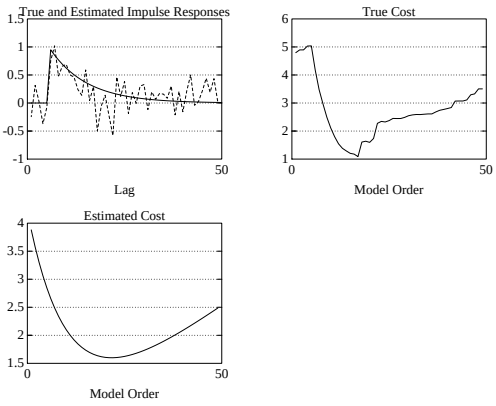

Figure 1: *Full order estimation, Akaike criterion, and true and estimated weighted and non-weighted mean square error.*

The top left quadrant shows the true impulse response together with the full 30th order model found via least squares. The true impulse response is the solid line, and the estimated impulse response is the dashed line. The Akaike information criterion (dotted line) and final prediction error tests (solid line) are shown in the

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|G(e^{j\omega})-\hat{G}(e^{j\omega})|^2|U(e^{j\omega})|^2 d\omega \qquad (56)$$

versus nominal model order as a solid line, and the estimate of (56) as a dashed line. The estimate was obtained from the estimates $\hat{\theta}$, $\hat{\alpha}$, $\hat{\lambda}$ and $\hat{\sigma^2}$ via equation (35). Notice that the estimate of the value of the integral is a very good indication of the true value of the integral. Furthermore, notice that the integrals (true and estimated) give a very clear criterion for the best model order to fit to the data using least squares depending on the intended use for the model. Other results are reported in [13].

8 Conclusion

In most identification applications, the nominal model is at best an approximation to the true system, whose structure is more complex than that of the parametrized model. This induces an error between the true transfer function and the estimated nominal model, which is called the unmodelled dynamics. One way of treating this error is to estimate it by further parametrizing it, but this amounts to replacing the nominal model by a more complex one, i.e. it amounts to modelling the unmodelled dynamics. Instead, we have shown in this paper that, by assuming that the unmodelled dynamics is a realization of a stochastic process described by a parametrized probability density function, one can estimate these parameters.

Our simulations have shown that very simple probability density models coupled with straightforward estimation procedures produce very reliable error bounds.

Our procedure produces an estimate of the mean square error between the true and estimated nominal transfer functions. This estimate is the sum of two clearly distinguishable terms, one due to the undermodelling (which decreases with model complexity) and one due to noise in the data (which increases with model complexity and decreases with the number of data). This expression has allowed us not only to produce error bounds, but also to develop a new optimal model order estimation criterion, GIC. This criterion, explicitly and, we believe, correctly incorporates the effect of undermodelling. It compares favorably with Akaike's FPE, as is demonstrated by both our theoretical analysis and simulations.

[2] G.C. Goodwin, M. Gevers and D.Q. Mayne, *"Bias and variance distribution in transfer function estimates"*, submitted to 9th IFAC Symp. on Identification and System Parameter Estimation, Budapest, July 1991.

[3] B. Wahlberg and L. Ljung, *"Design variables for bias distribution in transfer function estimation"*, IEEE Trans. Autom. Control, vol.AC-31, February 1986, pp. 134-144.

[4] R.O. La Maire, L. Valavani, M. Athans and G. Stein,*"A frequency domain estimator for use in adaptive control systems"*, Proceedings of ACC, pp. 238-244, 1987.

[5] R.L. Kosut,*"Adaptive robust control via transfer function uncertainty estimation*, Proceedings ACC, Atlanta, 1988.

[6] G.C. Goodwin, D.Q. Mayne and M. Salgado,*"Uncertainty, information and estimation"*, IFAC Symposium on Adaptive Control and Signal Processing, Glasgow, 1989.

[7] H. Akaike, *"Fitting autoregressive models for prediction"*, Ann. Inst. Stat. Math., vol. 21, pp. 243-347, 1969.

[8] H. Akaike, *"Information theory and an extension of the maximum likelihood principle"*, Proc. 2nd Int. Symp. Information Theory, pp. 267-281, 1972.

[9] H. Akaike,*"A new look at the statistical model identification"*, IEEE Trans. Autom. Control, vol. AC-19, pp. 716-723.

[10] B.M.Ninness,*'Robust Estimation'* , Ph.D. Thesis, Department of Electrical Engineering, University of Newcastle, AUSTRALIA

[11] R.H. Middleton and G.C. Goodwin, *"Digital Estimation and Control : A Unified Approach"*, Prentice Hall, New Jersey, 1990.

[12] B. Wahlberg, *"System identification using high-order models, revisited"*, Technical Report LITH-ISY-I-0999, Dept. of Electrical Engineering, Linköping University, Sweden, 1989.

[13] G.C. Goodwin M.Gevers, B.M.Ninness, *"Optimal Model Order Selection and Estimation of Model Uncertainty for Identification with Finite Data"*, IEEE Trans. Autom. Control, To Appear.