31. ON THE IDENTIFICATION OF 
NON-STATIONARY STOCHASTIC SYSTEMS

Tohru Ozaki
(The Institute of Statistical Mathematics, 
4-6-7 Minami-Azabu, Minato-ku, Tokyo, Japan)

1. INTRODUCTION

Successful results of application of a statistical time domain approach to the control of a stationary feedback system are reported in [16], where a newly developed statistical identification method of a multivariate autoregressive model plays an essential rôle. However, for the control of non-stationary systems, where the structure of the input or the system may change during the operation, the model of the system must be changed adaptively. Ozaki and Tong [19] gave a method of fitting a non-stationary autoregressive model to such a locally stationary scalar time series, employing Akaike's Information Criterion (AIC) which is a general decision criterion of the statistical model and is based on information theory [6]. The problem of the control of non-stationary systems can be solved by extending this procedure to the multi-dimensional case. In this paper we first survey the relation of multi-dimensional models to feedback control systems. Secondly we discuss AIC and some statistics for model identification. Finally we propose a non-stationary multi-dimensional autoregressive model fitting procedure. Some numerical results obtained by applying the procedure to artificial and real data are given.

2. AUTOREGRESSIVE MODELS FOR FEEDBACK SYSTEMS

The feedback system of Fig. 1 is usually approximated for sufficiently large $M$ by the linear model

$$y(s) = \sum_{m=1}^{M} a(m)x(s-m) + u(s)$$

(2.1)
\[ x(s) = \sum_{m=1}^{M} b(m)y(s-m) + v(s) \]  

(2.2)

where \( y(s) \) is the output of system A and at the same time is the input of system B, and where \( x(s) \) is the output of system B and is also the input of system A.

![Feedback System Diagram](image)

Fig. 1 Feedback System

\( u(s) \) is the disturbance to \( y(s) \) and \( v(s) \) is the disturbance to \( x(s) \). The least squares estimates of \( \{a(m); m=1, \ldots, M\} \) of (2.1) or \( \{b(m); m=1, \ldots, M\} \) of (2.2) are consistent only when the disturbances are white noise. When the systems are disturbed by coloured noise, which is common in real applications, we can however get consistent estimates of them by applying the whitening operator of the disturbance \( u(s) \) say

\[
(1 - \sum_{i=1}^{L} c(i)B^i)
\]

(1 - \sum_{i=1}^{L} c(i)B^i)

To the both sides of (2.1) \[1\] where \( B u(s) = u(s-1) \). Then we get

\[
y(s) - \sum_{\ell=1}^{L} c(\ell)y(s-\ell) = \sum_{m=1}^{M+L} A(m)x(s-m) + \xi(s) \]  

(2.3)

where

\[ A(1) = a(1) \]

\[ A(m) = a(m) - \sum_{\ell=1}^{m-1} c(\ell)a(m-\ell) \quad (m=2,3,\ldots,M+L) \]
and \( \xi(s) = u(s) - \sum_{\ell=1}^{L} c(\ell)u(s-\ell) \).

As the \( \xi(s) \) of (2.3) is white we can get consistent estimates \( \hat{c}(\ell) \) and \( \hat{A}(m) \) of \( c(\ell) \) and \( A(m) \) respectively. From these estimates we get consistent estimates \( \{\hat{a}(m); m=1, \ldots, M\} \) of \( \{a(m); m=1, \ldots, M\} \) as

\[
\hat{a}(1) = \hat{A}(1) \\
\hat{a}(m) = \hat{A}(m) + \sum_{\ell=1}^{m-1} \hat{c}(\ell)\hat{a}(m-\ell) \quad (m=2, 3, \ldots, M).
\]  

(2.4)

In a similar way we can get consistent estimates of impulse response functions \( \{a_{ij}(m); m=1, \ldots, M\} \) of a general feedback system of \( k \)-variables \( \{x_1(s), \ldots, x_k(s)\} \) defined by

\[
x_i(s) = \sum_{j=1}^{k} \sum_{m=1}^{M} a_{ij}(m)x_j(s-m) + u_i(s) \quad (i=1, 2, \ldots, k)
\]  

(2.5)

where \( u_i(s) \) is the proper disturbance to \( x_i(s) \). Corresponding to (2.3) we have from (2.5),

\[
x_i(s) = \sum_{\ell=1}^{L} c_{i}(\ell)x_i(s-\ell) + \sum_{j=1}^{k} \sum_{m=1}^{M} a_{ij}(m)x_j(s-m) + \varepsilon_i(s)
\]  

(2.6)

where

\[
u_i(s) = \sum_{\ell=1}^{L} c_{i}(\ell)u_i(s-\ell) + \varepsilon_i(s)
\]

and \( \varepsilon_i(s) \) is white noise, and corresponding to (2.6) we have

\[
\hat{a}_{ij} = \hat{A}_{ij}(1)
\]

\[
\hat{a}_{ij}(m) = \hat{A}_{ij}(m) + \sum_{\ell=1}^{m-1} \hat{c}_{i}(\ell)\hat{a}_{ij}(m-\ell) \quad (m=2, 3, \ldots).
\]

(2.7)
The model (2.5) is a special form of the well-known general \( k \)-dimensional autoregressive model

\[
X(s) = \sum_{m=1}^{M} A(m) X(s-m) + E(s) \tag{2.8}
\]

where

\[
X(s) = \begin{bmatrix}
x_1(s) \\
\vdots \\
x_k(s)
\end{bmatrix} , \tag{2.9}
\]

\[
A(m) = \begin{bmatrix}
A_{11}(m) & \cdots & A_{kk}(m) \\
\vdots & \ddots & \vdots \\
A_{kl}(m) & \cdots & A_{kk}(m)
\end{bmatrix} , \tag{2.10}
\]

\[
E(s) = \begin{bmatrix}
\epsilon_1(s) \\
\vdots \\
\epsilon_k(s)
\end{bmatrix} , \tag{2.11}
\]

and where \( M \) corresponds to \( L+M \) in (2.5). The spectral density matrix \( P(f) \) of the model (2.8) is given by

\[
P(f) = A(f)^{-1} \Sigma (A(f)^*)^{-1} \tag{2.12}
\]

where \((i,h)\)th element of \( A(f) \) is given by

\[
A_{ih}(f) = \sum_{m=0}^{M} A_{ih}(m) \exp(-i2\pi fm) \tag{2.13}
\]

where \( A_{ih}(0) \) is supposed to be the \((i,h)\)th element of the negative of the identity matrix, and \( \Sigma \) is the covariance matrix of the noise \( E(s) \). As is well-known \( p_{ij}(f) \), the \((i,j)\)th element of \( P(f) \), gives the cross spectral density between \( x_i(s) \) and \( x_j(s) \) and \( p_{ii}(f) \) gives the power spectral density of \( x_i(s) \). The noise
\( \Sigma = \begin{pmatrix}
\sigma_i^2 & 0 \\
\cdot & \cdot \\
0 & \sigma_k^2
\end{pmatrix}, \)

where \( \sigma_i^2 = E[\epsilon_i^2] \)

so that

\[
P_{ii}(f) = \sum_{j=1}^{k} |(A(f))^{-1}_{ij}|^2 \sigma_j^2. \quad (2.14)
\]

\( |(A(f))^{-1}_{ij}|^2 \) of (2.14) indicates the degree of contribution of noise \( u_j(s) \) to the power spectrum of \( x_i(s) \). This gives us useful information in selecting effective variables for feedback systems when we design control systems [1].

For the feedback control system, the \( k \)-dimensional vector

\[
X(s) = \begin{pmatrix}
x_1(s) \\
\vdots \\
x_k(s)
\end{pmatrix}
\]

of (2.8) consists of \( r \)-dimensional vector \( x(s) \) of control variables and \( l \)-dimensional vector \( y(s) \) of manipulated variables as

\[
X(s) = \begin{pmatrix}
x(s) \uparrow \\
\downarrow \vdots \\
x(s) \uparrow \\
\downarrow \end{pmatrix}
\]

\[
\begin{pmatrix}
x_1(s) \\
\vdots \\
x_k(s)
\end{pmatrix} = \begin{pmatrix}
r \uparrow \\
\downarrow \vdots \\
\downarrow \end{pmatrix}
\]

\( y(s) \)

where \( r+l = k \). Thus (2.8) allows the following representation of the controlled variables, which is directly used for the controller design [16];
\( x(s) = \sum_{m=1}^{M} a(m)x(s-m) + \sum_{m=1}^{M} b(m)y(s-m) + \varepsilon(s) \)  \hspace{1cm} (2.16)

where \( a(m) \), \( b(m) \) and \( \varepsilon(s) \) are given by

\[
A(m) = \begin{pmatrix}
\uparrow & \leftarrow r & \leftarrow l \\
\downarrow & a(m) & b(m) \\
\downarrow & l & * & *
\end{pmatrix}
\hspace{1cm} (2.17)
\]

\[
E(s) = \begin{pmatrix}
\uparrow & \leftarrow 1 \\
\downarrow & \varepsilon(s) \\
\downarrow & l & *
\end{pmatrix}
\hspace{1cm} (2.18)
\]

From expression (2.16) a state space representation preferable for the on-line control of the system is introduced [8] as

\[
Z(s) = \Phi z(s-1) + \Gamma y(s-1) + W(s)
\hspace{1cm} (2.19)
\]

where

\[
Z(s) = \begin{pmatrix}
\uparrow & \leftarrow 1 \\
\downarrow & Z_1(s) \\
\downarrow & \vdots \\
\downarrow & Z_M(s)
\end{pmatrix}, \quad \Gamma = \begin{pmatrix}
\uparrow & \leftarrow l \\
\downarrow & b_1 \\
\downarrow & \vdots \\
\downarrow & b_M
\end{pmatrix}, \quad W(s) = \begin{pmatrix}
\uparrow & \leftarrow \varepsilon(s) \\
\downarrow & 0 \\
\downarrow & \vdots \\
\downarrow & 0
\end{pmatrix},
\]

\[
Y(s-1) = \lambda y(s-1)
\]
\[ Z_{i,s} \text{ 's } (i=2,3,\ldots,M) \text{ are computed from the past history of } x(s) \text{ and } y(s) \text{ by (2.19) and } Z_{1,s} \text{ is given as the present observation of } x(s). \text{ If } Z_{1,s} \text{ is computed as defined by the right hand side of (2.19) assuming } \epsilon(s)=0, \text{ this gives the one-step-ahead prediction of } x(s). \text{ The controller is designed so as to give the minimum of the quadratic criterion with appropriate positive semi-definite matrices } Q \text{ and } R \]

\[ J_h = E\left\{ \sum_{n=1}^{H} (z_n^T Q z_n + y_{n-1}^T R y_{n-1}) \right\} \quad (2.20) \]

where \( H \) is predetermined positive integer and \( E \) denotes expectation. The optimal controller is given in a form

\[ Y(s) = GZ(s) \quad (2.21) \]

where, as is well-known, \( G \) is obtained by employing the dynamic programming technique [24],[16].

A computer program for controller design through AR models is given in [7].

3. ORDER DETERMINATION OF MODELS

Order determination of an autoregressive model is the crucial point for the success of the identification of the feedback systems. To solve this Akaike [2] introduced a static, Final Prediction Error (FPE)

\[ \text{FPE}(M) = \hat{\sigma}^2 (M) \left( 1 + \frac{M+1}{N} \right) \left( 1 - \frac{M+1}{N} \right)^{-1} \quad (3.1) \]
which is an estimate of one-step-ahead prediction error variance of a scalar autoregressive model of order \( M \) fitted to the observations \( x_1, \ldots, x_N \). \( \hat{\sigma}^2(M) \) is the least square estimate of the residual variance of the model and is given by

\[
\hat{\sigma}^2(M) = \hat{c}(0) - \hat{a}_M \hat{c}_1 - \ldots - \hat{a}_M \hat{c}_M
\]

(3.2)

where \( \hat{c}_\ell = \frac{1}{N} \sum_{i=1}^{N-\ell} x_i x_{i+\ell} \) and the \( \hat{a}_M \) 's are obtained by solving the Yule-Walker equation defined by the \( \hat{c}_\ell \)'s. If the order of the model is increased too much the estimated residual variance \( \hat{\sigma}^2(M) \) of (3.1) usually decreases, while

\[
(1 + \frac{M+1}{N})(1 - \frac{M+1}{N})^{-1}
\]

of (3.1), which is the indication of the unreliability of the model, increases. However, if the order is decreased too much the model's inadequacy usually increases. He proposed the minimum FPE procedure which selects the model which minimizes FPE. The procedure is extended to the identification of the multi-dimensional autoregressive model (2.8) and also to the identification of the autoregressive model (2.16) for control [3]. These procedures have been shown to work remarkably well with practical data [2],[3],[12],[13],[14],[16].

Akaike [6] proposed a general decision criterion for statistical model building based on the success of the minimum FPE procedure for the autoregressive model fitting in the time series. He suggested taking the Kullback-Leibler information quantity

\[
I(f(\cdot|\theta), g(\cdot)) = \int \log \left( \frac{f(x|\theta)}{g(x)} \right) g(x) \, dx
\]

(3.3)

as a natural measure of the deviation of the model \( f(x|\theta) \) from the true structure \( g(x) \) and suggested adopting the model \( f(x|\theta) \) which minimizes this quantity, or equivalently maximizes the mean log-likelihood

\[
\int g(x) \log f(x|\theta) \, dx
\]

(3.4)

He introduced a statistic, which we call Akaike's Information Criterion (AIC)
AIC = \(-2\)log(maximum likelihood) + 2(number of parameters) \ (3.5)

as a sample estimate of the information quantity \(I(f(\cdot | \theta); g(\cdot))\)

omitting terms common to every model. The procedure which
adopts the model whose AIC is minimum is called the Minimum
AIC Estimation (MAICE) procedure. When we wish to test the
adequacy of fit of the model \(f(x|\hat{\theta})\) against a more general
model \(f(x|\theta)\) where

\[
\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_{k_1}, O, \ldots, O) \quad \text{and} \quad \hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_{k_1}, \hat{\theta}_{k_1+1}, \ldots, \hat{\theta}_{k_2}),
\]

the procedure tells us to adopt the model \(f(x|\hat{\theta})\) if

\[
(-2) \log\left(\frac{f(x|\hat{\theta})}{f(x|\hat{\theta}_2)}\right) \leq 2(k_2-k_1) \quad (3.6)
\]

and otherwise the model \(f(x|\hat{\theta}_2)\). When \(k_2-k_1=1\) the above
criterion gives 15.73\% as a level of significance of the like-
lihood ratio test. Thus the MAICE procedure can be seen as an
extension of the likelihood ratio testing procedure in the
sense that it chooses the level of significance depending on
the difference in numbers of parameters of the two models.

It is easy to check that

\[
AIC(M) = N \log(FPE(M)) + \text{Constant} + O\left(\frac{1}{N}\right) \quad (3.7)
\]

for the autoregressive model fitting of a Gaussian scalar time
series of observations \(x_1, \ldots, x_N\), so that the MAICE procedure
is asymptotically equivalent to the minimum FPE procedure.

AIC provides a versatile procedure for statistical model
identification and has been successfully applied to the order
determination of polynomial regressions \([4],[22]\), order
determination of Markov chain models \([23]\), order determination
of factor analysis \([4],[5]\), order determination of ARIMA models
of time series \([18]\), modelling of non-stationary autoregressive
models \([19]\) and stochastic control of a ship's motion \([17]\).

The approximate maximum likelihood of the Gaussian multi-
dimensional autoregressive model (2.8) is given by Whittle
\([25]\) as

\[
\hat{\Theta}(x) = (2\pi e)^{-\frac{Nk}{2}} \det(\hat{S}_M)^{-\frac{N}{2}}
\]

\( (3.8) \)
where \( \det(S_M) \) is the determinant of \( S_M \) which is the maximum likelihood estimate of the covariance matrix of the residual vectors of the model of order \( M \). So the AIC of the model is given by

\[
AIC(M) = N \log \det(S_M) + 2(Mk^2 + k^2 + k) + Nk \log 2\pi + Nk
\]

(3.9)

where the parameters are the \( M \) coefficient matrices, the residual variance matrix and the mean vector of the series of dimension \( k \). The maximum likelihood estimate of the coefficient matrix and the residual variance matrix \( S_M \) for each model of order \( M \) is given by the efficient recursive procedure due to Whittle [26], which is a multi-dimensional generalization of the procedure of the scalar autoregressive model fitting schemes of successively increasing order proposed by Levinson [15] and Durbin [9].

4. NON-STATIONARY AUTOREGRESSIVE MODEL FITTING USING AIC

Our idea of modelling the non-stationary process is to consider the process to be locally stationary and to fit stationary autoregressive models locally by means of the MAICE procedure (or equivalently by the minimum FPE procedure). The global goodness of fit of this joined model is measured again by AIC. So the partition which gives the minimum AIC is adopted as the best model. The approximate maximum log-likelihood of the joined model composed of \( r \) stationary scalar autoregressive models \( A_i \) \( (i=1, \ldots, r) \) is given by

\[
-\frac{1}{2} \sum_{i=1}^{r} \{n_i \log \sigma_i^2 + n_i \log 2\pi + n_i\},
\]

where \( n_i \) and \( \sigma_i^2 \) denote, respectively, number of observations and the least squares estimate of the residual \( A_i \). So the AIC of this joined model \( A_1 \theta \cdots \theta A_r \) is given by

\[
AIC(A_1 \theta A_2 \cdots \theta A_r) = \sum_{i=1}^{r} \{n_i \log \theta_i^2 + 2(M_i+2)\} + n \log 2\pi + n
\]

where \( n = \sum_{i=1}^{r} n_i \), and parameters involved in this joined model are autoregressive coefficients, the residual variance of each local model and the mean of each local series. This idea of
modelling is an extension of the segmented regression model identification [10],[11],[20],[21], where both joining points and coefficients of each local regression are estimated by the maximum likelihood estimation method. However it needs enormous computation to seek the minimum AIC model for such a general joined series composed of \(i\) (\(i=1,...,r\)) segments of stationary series, because the number of partitions is very large. A practical on-line identification procedure is presented in Ozaki et al. [19], in which the length of each of the mutually independent local series is supposed to be an integral multiple of sufficiently large integer \(N\). Hereafter we will call this \(N\) the basic span of the series. In such a case it is required, each time a new set of observations of length \(N\) is obtained, to decide whether we should fit the new observations by a new model separately or fit them together with the preceding observations. A typical situation is illustrated in Fig. 2. It is now supposed that we have a model

\[ mN \quad N \]

\[ LN \quad (l+m)N \quad (l+m+1)N \]

\[ M(LN+1,(l+m)N) \quad M((l+m)N+1,(l+m+1)N) \]

\[ M(LN+1,(l+m+1)N) \]

**Fig. 2 Models for observations** \(x(LN+1),...,x((l+m+1)N)\).

fitted to the observations \(x(LN+1), x(LN+2),..., x((l+m)N)\). The notation \(M(LN+1,(l+m)N)\) is used to represent the model. We have two possible choices for the modelling of the newly obtained observations \(x((l+m)N+1),...,x((l+m+1)N)\): one is to fit the data with the model \(M((l+m)N+1,(l+m+1)N)\) and the other is to fit the data with the model \(M(LN+1,(l+m+1)N)\), both models being identified by the use of the MAICE procedure. The choice of a model is made according to the values of AIC of the respective models computed from the observations \(x(LN+1),..., x((l+m)N), x((l+m)N+1),..., x((l+m+1)N)\). The AIC of the model \(M(LN+1,(l+m+1)N)\) is given by

\[
(m+1)N \log \hat{\sigma}_1^2 + 2(M_1 + 2)
\]

where \(\hat{\sigma}_1^2\) is the least squares estimate of residual variance

\[(4.1)\]
of the autoregressive model $M(\lambda N+1, (\lambda+m+1)N)$ and $M_1$ is the order of the model. The AIC of the joined model of $M(\lambda N+1, (\lambda+m)N)$ and $M((\lambda+m)N+1, (\lambda+m+1)N)$ which we denote by

$$M(\lambda N+1, (\lambda+m)N) \otimes M((\lambda+m)N+1, (\lambda+m+1)N)$$

is given by

$$mN \log \hat{\sigma}^2 - N \log \hat{\sigma}^2 + 2(M_2 + M_3 + 4)$$

(4.2)

where $\hat{\sigma}^2$ and $\hat{\sigma}^2$ are the least squares estimates of residual variance of the autoregressive models $M(\lambda N+1, (\lambda+m)N)$ and $M((\lambda+m)N+1, (\lambda+m+1)N)$ respectively; and $M_2$ and $M_3$ are the orders of these models. Therefore, if the AIC of the model $M(\lambda N+1, (\lambda+m+1)N)$ is less than the AIC of the joined model $M(\lambda N+1, (\lambda+m)N) \otimes M((\lambda+m)N+1, (\lambda+m+1)N)$, then the model $M(\lambda N+1, (\lambda+m+1)N)$ is chosen for the whole series of observations $x(\lambda N+1), \ldots, x((\lambda+m+1)N)$, otherwise the model is switched from $M(\lambda N+1, (\lambda+m)N)$ to $M((\lambda+m)N+1, (\lambda+m+1)N)$ for the observations $x((\lambda+m)N+1), \ldots, x((\lambda+m+1)N)$. Using this method, estimates of the power spectra of non-stationary scalar time series were obtained [19].

The procedure is extended to the multi-dimensional case as follows. The maximum log-likelihood of a k-dimensional autoregressive model is given by (3.9), so that the maximum log-likelihood of the joined model of k-dimensional autoregressive model $M(N_1+1, N_2) \otimes M(N_2+1, N_3)$ is given by

$$-\frac{(N_2-N_1)}{2} \log \det \hat{S}_1 - \frac{(N_3-N_2)}{2} \log \det \hat{S}_2 - \frac{(N_3-N_1)k}{2} \log 2\pi$$

(4.3)

where $\det \hat{S}_1$ and $\det \hat{S}_2$ are the determinants of the maximum likelihood estimates of variance matrix of the models $M(N_1+1, N_2)$ and $M(N_2+1, N_3)$, respectively. The AIC of the joined model is, then, given by

$$-\frac{(N_2-N_1)}{2} \log \det \hat{S}_1 + (N_3-N_2) \log \det \hat{S}_2 + (N_3-N_1)k$$

$$+ (N_3-N_1)k \log 2\pi + 2(M_1+1)k^2 + (M_2+1)k^2 + 2k$$

(4.4)
Fig. 3  Power spectral density $P_{x_1 x_1}(f)$ of the model (4.5)

Fig. 4  Power spectral density $P_{x_2 x_2}(f)$ of the model (4.5)

Input: $X_1(S)$  
Output: $X_2(S)$  
$N = 300$  
$H = 100$

Fig. 5  Frequency response function of the model (4.5)
Fig. 6 Power spectral density $p_{x_1x_1}(f)$ of the model (4.6)

Fig. 7 Power spectral density $p_{x_2x_2}(f)$ of the model (4.6)

Fig. 8 Frequency response function of the model (4.6)
where $M_1$ is the order of the model $M(N+1,N_1)$ and $M_2$ is the order of the model $M(N+1,N_2)$, and where parameters involved in the joined model are $(M+M)$ coefficient matrices, 2 variance matrices of residual vectors and 2 mean vectors.

To test the practical effectiveness of the above extended procedure for the analysis of feedback systems, we apply it to an artificial two-dimensional autoregressive series of length 2000 whose former half is generated from the model

\[
\begin{pmatrix}
  x_1(s) \\
  x_2(s)
\end{pmatrix}
= \begin{pmatrix}
  0.202 & 0.395 \\
  -0.039 & 0.592
\end{pmatrix}
\begin{pmatrix}
  x_1(s-1) \\
  x_2(s-1)
\end{pmatrix}
+ \begin{pmatrix}
  -0.122 & -0.186 \\
  -0.009 & 0.035
\end{pmatrix}
\begin{pmatrix}
  x_1(s-2) \\
  x_2(s-2)
\end{pmatrix}
+ \begin{pmatrix}
  -0.017 & 0.025 \\
  -0.198 & -0.132
\end{pmatrix}
\begin{pmatrix}
  x_1(s-3) \\
  x_2(s-3)
\end{pmatrix}
+ \begin{pmatrix}
  0.077 & -0.145 \\
  -0.217 & -0.182
\end{pmatrix}
\begin{pmatrix}
  x_1(s-4) \\
  x_2(s-4)
\end{pmatrix}
+ \begin{pmatrix}
  0.031 & 0.224 \\
  -0.029 & 0.284
\end{pmatrix}
\begin{pmatrix}
  x_1(s-5) \\
  x_2(s-5)
\end{pmatrix}
+ \begin{pmatrix}
  \epsilon_1(s) \\
  \epsilon_2(s)
\end{pmatrix}
\tag{4.5}
\]

where $(\epsilon_1(s), \epsilon_2(s))^\prime$ is normally distributed with mean vector $\mathbf{0}=(0,0)^\prime$ and covariance matrix

\[
\Sigma = \begin{pmatrix}
  0.466 \times 10^5 & -0.214 \times 10^3 \\
  -0.214 \times 10^3 & 0.876 \times 10^4
\end{pmatrix}
\]

and the latter half generated from the model

\[
\begin{pmatrix}
  x_1(s) \\
  x_2(s)
\end{pmatrix}
= \begin{pmatrix}
  0.120 & 0.069 \\
  -0.023 & 0.745
\end{pmatrix}
\begin{pmatrix}
  x_1(s-1) \\
  x_2(s-1)
\end{pmatrix}
+ \begin{pmatrix}
  -0.139 & -0.038 \\
  -0.021 & -0.372
\end{pmatrix}
\begin{pmatrix}
  x_1(s-2) \\
  x_2(s-2)
\end{pmatrix}
+ \begin{pmatrix}
  -0.064 & 0.134 \\
  -0.160 & -0.012
\end{pmatrix}
\begin{pmatrix}
  x_1(s-3) \\
  x_2(s-3)
\end{pmatrix}
+ \begin{pmatrix}
  -0.029 & -0.075 \\
  -0.228 & -0.188
\end{pmatrix}
\begin{pmatrix}
  x_1(s-4) \\
  x_2(s-4)
\end{pmatrix}
+ \begin{pmatrix}
  \epsilon_1(s) \\
  \epsilon_2(s)
\end{pmatrix}
\tag{4.6}
\]
where \((\varepsilon_1(s), \varepsilon_2(s))^t\) is normally distributed with zero mean vector and covariance matrix
\[
\Sigma = \begin{pmatrix}
0.401 \times 10^5 & 0.732 \times 10^3 \\
0.732 \times 10^3 & 0.123 \times 10^5
\end{pmatrix}
\]

The basic span \(N\) which must be specified beforehand was set to be 500. The possible order of each local model was assumed to be less than 26. As a result, the first two segments of each length 500 were joined and fitted by a single model separately from the last two segments of each length 500, which were also joined and fitted by a single model. The AIC of the models at each stage are illustrated in Table I.

**Table I.**

AIC and identified orders of models for observations generated from (4.5) and (4.6)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>identified order</th>
<th>number of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M(1,500))</td>
<td>0.19638 \times 10^5</td>
<td>5</td>
<td>500</td>
</tr>
<tr>
<td>(M(501,1000))</td>
<td>(0.19616 \times 10^5)</td>
<td>5</td>
<td>500</td>
</tr>
<tr>
<td>(M(1,1000))</td>
<td>0.19671 \times 10^5</td>
<td>5</td>
<td>1000</td>
</tr>
<tr>
<td>(M(1,1000))</td>
<td>(0.20070 \times 10^5)</td>
<td>4</td>
<td>500</td>
</tr>
<tr>
<td>(M(1,1500))</td>
<td>(0.20108 \times 10^5)</td>
<td>4</td>
<td>500</td>
</tr>
<tr>
<td>(M(1001,1500))</td>
<td>(0.20070 \times 10^5)</td>
<td>4</td>
<td>1000</td>
</tr>
<tr>
<td>(M(1001,2000))</td>
<td>(0.20070 \times 10^5)</td>
<td>4</td>
<td>1000</td>
</tr>
</tbody>
</table>

Estimated parameters of the identified models at each stage are given in Tables II to V. Figures of the estimated power spectra \(p_{x_1 x_1}^2(f)\) and \(p_{x_2 x_2}^2(f)\) and the frequency response function corresponding to these local models are illustrated in Figs. 9 to 20, where the frequency response function \(A_{21}(f)\)
### Table II
Estimated parameters of the model M(1,500)

<table>
<thead>
<tr>
<th>Order M</th>
<th>covariance matrix of residuals $\Sigma$</th>
<th>coefficient matrix $A(1)$</th>
<th>coefficient matrix $A(2)$</th>
<th>coefficient matrix $A(3)$</th>
<th>coefficient matrix $A(4)$</th>
<th>coefficient matrix $A(5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.384x10^5 0.527x10^3 0.529x10^3 0.740x10^4</td>
<td>0.122 0.486</td>
<td>-0.052 -0.167</td>
<td>-0.030 -0.142</td>
<td>0.177 -0.133</td>
<td>0.063 0.270</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.040 0.707</td>
<td>-0.019 0.004</td>
<td>-0.191 -0.083</td>
<td>-0.200 -0.247</td>
<td>0.004 0.313</td>
</tr>
</tbody>
</table>

### Table III
Estimated parameters of the model M(1,1000)

<table>
<thead>
<tr>
<th>Order M</th>
<th>covariance matrix of residuals $\Sigma$</th>
<th>coefficient matrix $A(1)$</th>
<th>coefficient matrix $A(2)$</th>
<th>coefficient matrix $A(3)$</th>
<th>coefficient matrix $A(4)$</th>
<th>coefficient matrix $A(5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.411x10^5 -0.475x10^5 0.475x10^2 0.759x10^4</td>
<td>0.175 0.370</td>
<td>-0.090 -0.152</td>
<td>-0.043 0.030</td>
<td>0.104 -0.218</td>
<td>0.032 0.270</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.045 0.621</td>
<td>0.001 0.067</td>
<td>-0.207 -0.161</td>
<td>-0.224 -0.177</td>
<td>-0.010 0.305</td>
</tr>
</tbody>
</table>
### Table IV
Estimated parameters of the model $M(1001,1500)$

<table>
<thead>
<tr>
<th>Order M</th>
<th>covariance matrix of residuals $\Sigma$</th>
<th>coefficient matrix $A(1)$</th>
<th>coefficient matrix $A(2)$</th>
<th>coefficient matrix $A(3)$</th>
<th>coefficient matrix $A(4)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$0.385 \times 10^5$ $-0.282 \times 10^3$</td>
<td>0.151 $-0.049$</td>
<td>$-0.096$ $0.146$</td>
<td>$-0.070$ $0.079$</td>
<td>$-0.021$ $-0.042$</td>
</tr>
<tr>
<td></td>
<td>$-0.282 \times 10^3$ $0.130 \times 10^5$</td>
<td>$-0.035$ $0.765$</td>
<td>$-0.015$ $-0.379$</td>
<td>$-0.182$ $-0.015$</td>
<td>$-0.198$ $-0.186$</td>
</tr>
</tbody>
</table>

### Table V
Estimated parameters of the model $M(1001,2000)$

<table>
<thead>
<tr>
<th>Order M</th>
<th>covariance matrix of residuals $\Sigma$</th>
<th>coefficient matrix $A(1)$</th>
<th>coefficient matrix $A(2)$</th>
<th>coefficient matrix $A(3)$</th>
<th>coefficient matrix $A(4)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$0.387 \times 10^5$ $-0.492 \times 10^3$</td>
<td>0.108 $0.020$</td>
<td>$-0.130$ $-0.003$</td>
<td>$-0.087$ $0.196$</td>
<td>$-0.032$ $-0.068$</td>
</tr>
<tr>
<td></td>
<td>$-0.492 \times 10^3$ $0.129 \times 10^5$</td>
<td>$-0.035$ $0.720$</td>
<td>$-0.041$ $-0.357$</td>
<td>$-0.173$ $-0.029$</td>
<td>$-0.213$ $-0.173$</td>
</tr>
</tbody>
</table>
Fig. 9 Power spectral density \( p_{x_1x_1}(f) \) estimated through the model \( M(1,500) \)

Fig. 10 Power spectral density \( p_{x_2x_2}(f) \) estimated through the model \( M(1,500) \)

Input: \( X_1(S) \)
Output: \( X_2(S) \)
\( N = 500 \)
\( H = 100 \)

Fig. 11 Frequency response function estimated through the model \( M(1,500) \)
Fig. 12 Power spectral density $p_{X_1X_1}(f)$ estimated through the model $M(1,1000)$

Fig. 13 Power spectral density $p_{X_2X_2}(f)$ estimated through the model $M(1,1000)$

Input: $X_1(S)$
Output: $X_2(S)$
$N = 1000$
$H = 100$

Fig. 14 Frequency response function estimated through the model $M(1,1000)$
Fig. 15 Power spectral density $p_{X_1X_1}(f)$ estimated through the model $M(1001, 1500)$

Fig. 16 Power spectral density $p_{X_2X_2}(f)$ estimated through the model $M(1001, 1500)$

Input: $X_1(S)$
Output: $X_2(S)S$
N = 500
H = 100

Fig. 17 Frequency response function estimated through the model $M(1001, 1500)$
Fig. 18. Power spectral density $p_{x_1x_1}(f)$ estimated through the model $M(1001,1500)$.

Fig. 19. Power spectral density $p_{x_2x_2}(f)$ estimated through the model $M(1001,2000)$.

Input: $X_1(S)$
Output: $X_2(S)$
$N = 1000$
$H = 100$

Fig. 20. Frequency response function estimated through the model $M(1001,2000)$. 
Fig. 21 2-dimensional car vibration data of length 1600
Fig. 22 Estimated power spectral density $p_{x_1 x_1}(f)$ of the car vibration data $(1-1200)'_1$.

Fig. 23 Estimated power spectral density $p_{x_2 x_2}(f)$ of the car vibration data $(1-1200)'_2$. 
Fig. 24  Estimated power spectral density $p_{X_1X_1}(f)$ of the car vibration data (1201-1600)

Fig. 25  Estimated power spectral density $p_{X_2X_2}(f)$ of the car vibration data (1201-1600)
of each of these figures is given by the cross spectrum
\( p_{x_2x_1}(f) \) and the power spectrum \( p_{x_1x_1}(f) \) as

\[
A_{21}(f) = \frac{p_{x_2x_1}(f)}{p_{x_1x_1}(f)}.
\]

We note the similarity of these figures to the corresponding figures Figs. 3 to 8 of the theoretical models. Although we took 250 or 200 as the basic span, we arrived at the same two models, \( M(1,1000) \) and \( M(1001,2000) \).

The procedure was also applied to two-dimensional car vibration data of length 1600 (Fig. 21). The basic span of this series was chosen to be 400. As a result, observations after the 1200th were fitted separately from the preceding observations. The data were again divided into two parts when we chose the basic span to be 300. The separation looks reasonable because the amplitude of the later part of the data is rather large (see Fig. 21). This part of the car vibration data may be inflated by large disturbances from the road. The estimated power spectra \( p_{x_1x_1}(f) \) and \( p_{x_2x_2}(f) \) of the model \( M(1,1200) \) for the preceding part and those of the model \( M(1201,1600) \) for the succeeding part also suggest the difference of the level of noises of both models (see Figs. 22 to 25).

5. CONCLUSIONS

The multi-dimensional autoregressive models are useful, in real applications, for the analysis of feedback systems and the controller design. Some newly introduced statistics for the order determination of the model make this approach powerful. This approach can be extended to the non-stationary situations, in which the autoregressive model for control has to be modified adaptively as the time passes. The adaptive modification of the model is performed using AIC with the assumption that the length of each of the mutually independent local series is supposed to be an integral multiple of the basic span \( N \). However the objective choice of the basic span \( N \) remains unsolved. Of course, instead of assuming that the possible joining points are predetermined, we can use the MAICE principle to estimate their locations, but at the cost of parting with the "on-line" property of the procedure.
REFERENCES


