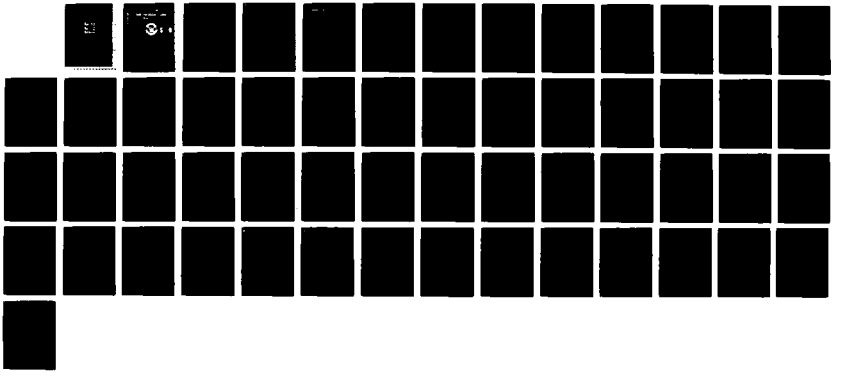
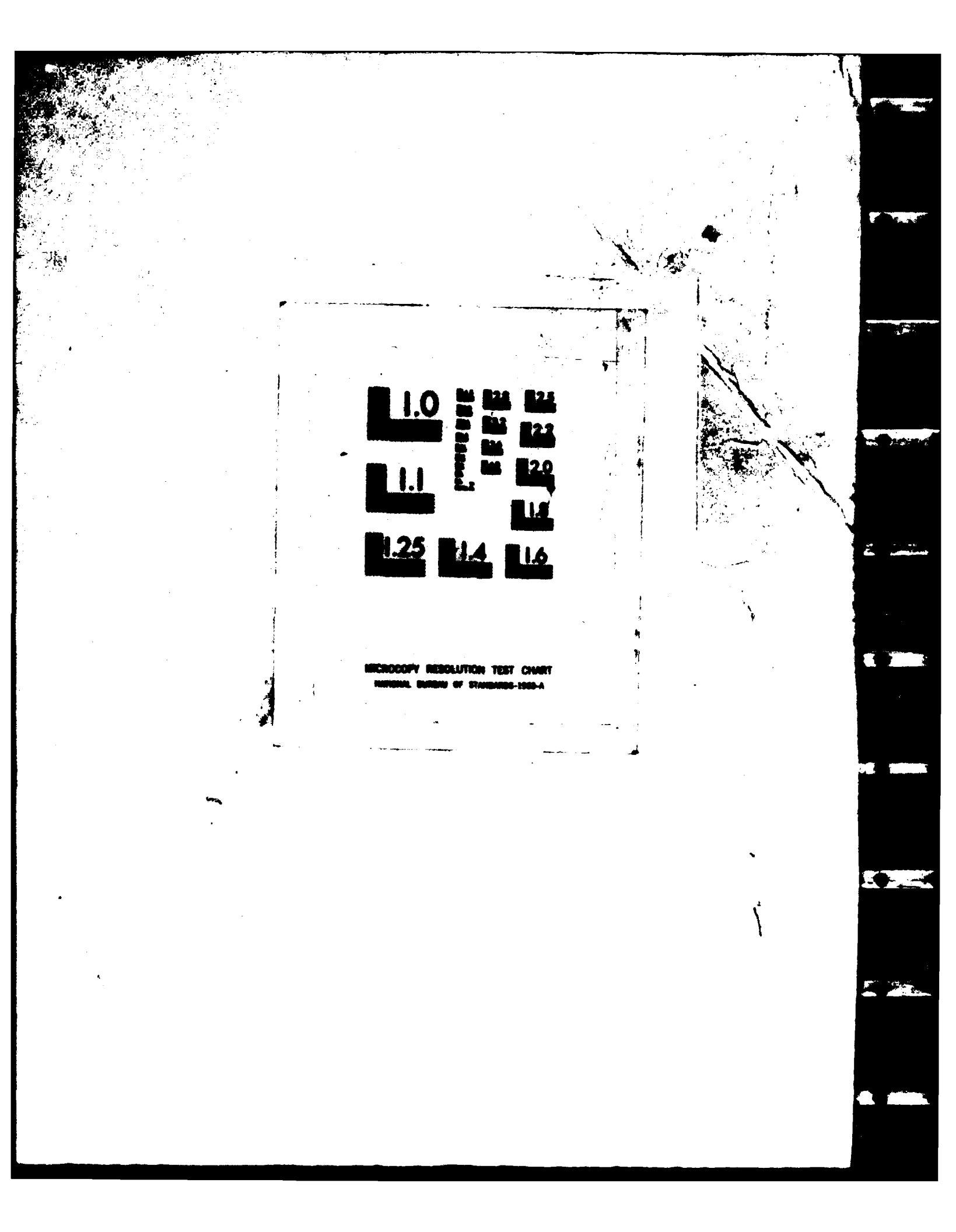
AD-A181 607 SMOOTHNESS PRIORS TRANSFER FUNCTION ESTIMATION(U) NAVAL 1/1
POSTGRADUATE SCHOOL MONTEREY CA W GERSCH ET AL. JUN 87
NPS55-87-005

UNCLASSIFIED F/G 12/1 NL




1.0
1.1
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

NPS55-87-005

NAVAL POSTGRADUATE SCHOOL

Monterey, California

SMOOTHNESS PRIORS TRANSFER FUNCTION ESTIMATION

WILL GERSCH
GENSHIRO KITAGAWA

JUNE 1987

Prepared for:
Naval Postgraduate School
Monterey, CA 93943-5000

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral R. C. Austin
Superintendent

D. A. Schrady
Provost

Reproduction of all or part of this report is authorized.

This report was prepared by:

WILL GERSCH
National Research Council
Senior Associate
Naval Postgraduate School

Reviewed by:

PETER PURDUE
Professor and Chairman
Department of Operations Research

Released by:

KNEALE T. MARSHALL
Dean of Information and
Policy Sciences

SECURITY CLASSIFICATION OF THIS PAGE

AD-A181 607

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NPS55-87-005	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b OFFICE SYMBOL (If applicable) Code 55	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING SPONSORING ORGANIZATION 8b OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
SMOOTHNESS PRIORS TRANSFER FUNCTION ESTIMATION

12. PERSONAL AUTHOR(S)
Gersch, Will and Kitagawa, Genshiro (Instit of Statistical Mathematics - Tokyo, Japan)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical Report	FROM _____ TO _____	1987 June	50

16 SUPPLEMENTARY NOTATION

17. COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Bayesian model, smoothness priors, time series analysis, transfer function estimation,

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A smoothness priors approach to transfer function estimation from stationary time series is shown. An infinite order impulse response model plus an infinite order additive AR noise model is assumed. This is algebraically equivalent to an infinite order ARMAX plus white noise model. A finite order ARMAX model approximation to this model is actually fitted to data. Frequency domain smoothness priors are assumed on the ARMAX polynomials and smoothness hyperparameters balance the tradeoff between the infidelity of the model to the data and the infidelity of the model to the smoothness constraints. The likelihood of the hyperparameters is maximized by a least squares gradient search computational procedure. The method is illustrated by the analysis of the Box-Jenkins series J data. Some of the statistical properties of the method are explored in Monte-Carlo simulation studies. keywords-

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Will Gersch	22b TELEPHONE (Include Area Code) (408)646-2801 22c OFFICE SYMBOL Code 55

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

☆ U.S. Government Printing Office: 1986—606-243

SMOOTHNESS PRIORS TRANSFER FUNCTION ESTIMATION*

Will Gersch
Department of Information and Computer Sciences
University of Hawaii
Honolulu. HI 96822

and

Genshiro Kitagawa
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-Ku
Tokyo 106, Japan

ABSTRACT: A smoothness priors approach to transfer function estimation from stationary time series is shown. An infinite order impulse response model plus an infinite order additive AR noise model is assumed. This is algebraically equivalent to an infinite order ARMAX plus white noise model. A finite order ARMAX model approximation to this model is actually fitted to data. Frequency domain smoothness priors are assumed on the ARMAX polynomials and smoothness hyperparameters balance the tradeoff between the infidelity of the model to the data and the infidelity of the model to the smoothness constraints. The likelihood of the hyperparameters is maximized by a least squares gradient search computational procedure. The method is illustrated by the analysis of the Box-Jenkins series J data. Some of the statistical properties of the method are explored in Monte-Carlo simulation studies.

KEYWORDS: Bayesian model, smoothness priors. time series analysis, transfer function estimation.

* This work was completed while the first author was a National Research Council Senior Associate at the Naval Postgraduate School, Monterey California.

1. INTRODUCTION

A Bayesian, smoothness priors approach is introduced in this paper for transfer function estimation. Jointly stationary input and output data is assumed to be observed in the presence of additive colored noise. The method is particularly relevant when only short spans of data are available, when the impulse response is relatively long-tailed and when the low order polynomial ARMAX type model can not capture the true model structure. The method is illustrated by the analysis of the Box-Jenkins Series J data. The statistical performance of the method is explored in Monte-Carlo simulation studies.

The models in Astrom and Bohlin (1965) and Box and Jenkins (1970), are the classical parametric time domain transfer function models. In that method, ARMAX type models characterised by polynomial operators on the input, the output and the observation noise are fitted to the observed input and output data. (The observation noise in the Astrom -Bohlin model is MA noise. It is AR noise in the Box-Jenkins model.) That method requires the specification of three polynomial operator orders, one each for the input, output and noise polynomials and the estimation of the unknown polynomials coefficients via the minimisation of a performance functional. Typically that computation is achieved by a computationally costly nonlinear optimisation procedure. In such procedures it is only feasible to search for solutions over low polynomial orders.

Despite the fact that conventional transfer estimation methods have been extensively used, the influence of the sampling variability in the polynomial model order selection on the transfer function estimation performance remains to be explored. Another objection to the use of low order polynomial ARMAX models is that the "parsimonious" parametric model may not be a good characterisation of the system that generated the data. An elaboration of this objection to the conventional parametric modeling method, from a Bayesian point of view, is that the conventional parametric modeling methods can not yield "correct" models. That is, if there is information in the data to select, by some best model order selection procedure. an ARMAX p,q,r model, then there is also information in the data to select alternative ARMAX p',q',r' models. There the best

Bayesian model requires that the transfer functions computed with different model orders be averaged, with respect to the likelihood of each fitted model and the prior probabilities of the model orders. Akaike (1979a) is perhaps the first example of a Bayesian time series modeling with averaging of the computational results over models with different model orders. The smoothness priors method of transfer function estimation introduced here obviates the 3 stringent parameter model order search problem in conventional transfer estimation procedures. Our procedure also uses 3 parameters. One is a 1 dimensional model order parameter, the other two are smoothness differential order parameters. We demonstrate that the values of those parameters are not critical in determining the estimated transfer function properties.

The Mayne-Firsoon (1978) three stage least squares (3SLS) procedure was developed to avoid the costliness of the maximisation of the nonlinear likelihood for the transfer function modeled wirth additive MA noise.. Almost contemporaneous alternative linear computational transfer function estimation procedures were generalised least squares (Clarke, 1967) and extended least squares (Panuska, 1968). Durbin (1961), a 2SLS procedure, was the conceptual predecessor of the Mayne-Firsoon procedure. Astrom and Mayne (1982) and Hannan et al. (1986) are recursive procedure realisations of the Mayne-Firsoon 3SLS transfer function estimation procedure. Other recent noteworthy publications on or related to transfer function estimation include Hannan and Rissanen (1982), a recursive method for finding model orders, and Ljung (1985), a study of the statistical properties of time and frequency domain transfer function estimation procedures.

Jordinson et al. (1970), Newbold (1973), Wegman (1980), Jakeman and Young (1982), and Kruc et al. (1982) are examples of the literature on statistical regularisation and Bayesian smoothed deconvolution procedures for the estimation of transfer functions. Applications of that literature include the estimation of the transfer function of the vascular system, applications in radiology, dispersive relations in streams etc. This activity is not summarised here . Our own smoothness priors method is a variation on that Bayesian theme.

In our method, an Mth order impulse response between input and output plus an Mth order autoregressive (AR) model for the additive noise is assumed. with M "quite large". This model is equivalent to an ARMAX plus white noise model. We assume integrated square seroth and kth order derivative frequency domain smoothness constraints on the polynomial operators. In the least squares framework, the resultant model strikes a balance between the infidelity of the solution to the data and the infidelity of the solution to the smoothness constraints. That balance or tradeoff is characterised by one parameter for each of four smoothness constraints. In Bayesian terminology, those are referred to as hyperparameters, (Lindley and Smith, 1972). The likelihood of the hyperparameters that characterise the class of smoothness priors is maximised to yield the best transfer function model with the best data dependent priors.

The approach taken in this paper is an application of our frequency domain smoothness priors AR model spectral estimation method, Kitagawa and Gersch (1985a). Some of our other time-domain smoothness priors papers on the modeling of nonstationary time series are Brotherton and Gersch (1981), Kitagawa (1981) and Kitagawa and Gersch (1984, 1985b). Additional Bayesian methods of time series analysis are in TIMSAC-84, Akaike et al. (1985). Gersch and Kitagawa (1987), is a review of our smoothness priors modeling of time series. The papers by Shiller (1973) and Akaike (1979b,1980) are predecessors to our own work. In particular, Akaike (1980), motivated our interest in this subject. Additional related work, that is better known in the statistics literature, are the methods of regularization, Tikhonov (1965), and the maximised penalized likelihood method (I.J. Good 1970, and Good and Gaskins 1980), Wahba (1977),(1982),(1983) and O'Sullivan (1987).

The smoothness priors method of transfer function estimation is treated in Section 2. An example of the smoothness priors analysis of the Box-Jenkins Series J data is shown in Section 3. Studies of the statistical performance of the smoothness priors method and comparisons of the SP and 3SLS methods of transfer estimation are also in Section 3. An interpretation of the results, summary and discussion in Section 4 conclude the paper.

2 ANALYSIS

In Section 2.1, several features of a Bayesian model for linear regression are shown. A variety of other analyses of the Bayesian model are shown for example in Zellner (1971) and Broemeling (1985). The presentation here uses assumptions on the priors of the parameter vector similar to those used earlier in Lindley and Smith (1972) and Akaike (1980). Illustrations of smoothness priors, a particularization of the Bayesian linear model analysis, are shown in applications to the time series analysis problems considered by Whittaker (1923) and Shiller (1973) in Section 2.2. Our transfer function model is described in Section 2.3. In contrast with the time domain priors on the model coefficients in the Whittaker and Shiller problems, in Section 2.4 we show an application of smoothness priors in the frequency domain, in transfer function estimation.

2.1 A BAYESIAN LINEAR REGRESSION MODEL

Consider first, the linear regression model

$$y = X\theta + \epsilon \,. \tag{2.1.1}$$

In (2.1.1), $y = (y_1,\ldots,y_N)$, is an nx1 vector of observations, θ is a px1 parameter vector, X is an nxp known design matrix and ϵ is an i.i.d. nx1 random vector with $\epsilon \sim N(0,\sigma^2 I)$. Then, the conditional data distribution is

$$p(y|\, X,\theta,\sigma) = (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}(y - X\theta)^T(y - X\theta)\right\} . \tag{2.1.2}$$

In the stochastic regression problem, θ is considered to be a random vector with distribution $\pi(\theta)$. From Bayes theorem, the posterior distribution of the parameter vector θ is proportional to the product of the conditional data distribution (the likelihood), $p(y|\, X,\theta,\sigma^2)$, and the prior distribution, $\pi(\theta)$,

$$\pi(\theta|\, y,\sigma) \propto p(y|\, X,\theta,\sigma)\pi(\theta) \,. \tag{2.1.3}$$

Let the prior distribution of θ be $D\theta \sim N(D\theta_0,\lambda^2 I)$. In the application of Bayesian regression

methods to smoothness priors problems, we shall consider the special case of $\theta_0 = 0$ Also, it is convenient to introduce the parameterization, $\lambda = \tau/\sigma$. In that case, the prior distribution on θ is

$$\pi(\theta|\,\tau,\sigma) = (2\pi\sigma^2)^{-p/2}\left|\tau^2 D^T D\right|^{1/2}\exp\left\{-\frac{\tau^2}{2\sigma^2}\theta^T D^T D\theta\right\}. \tag{2.1.4}$$

τ is referred to as the hyperparameter of the prior distribution, (Lindley and Smith, 1972). In this conjugate family Bayesian situation, in which both the priors and conditional data distribution are normally distributed, the posterior distribution is also normally distributed, (Zellner 1971). The mean of that distribution is easily computed as the minimizer of

$$\left\|\begin{pmatrix} y \\ 0 \end{pmatrix} - \begin{pmatrix} X \\ \tau D \end{pmatrix}\theta\right\|^2. \tag{2.1.5}$$

If τ were known, the computational problem in (2.1.5) could be solved by an ordinary least squares computation. The solution for $\hat{\theta}$ is

$$\hat{\theta} = \left[X^T X + \tau^2 D^T D\right]^{-1} X^T y \tag{2.1.6}$$

with the residual sum of squares,

$$SSE(\tau) = y^T y - \hat{\theta}^T\left[X^T X + \tau^2 D^T D\right]\hat{\theta}. \tag{2.1.7}$$

The posterior distribution of θ, $\pi(\theta|\,y,\tau,\sigma)$ is a proper distribution, therefore the likelihood for the unknown parameter τ can be determined by

$$L(\tau,\sigma) = \int_{-\infty}^{\infty}\pi(\theta|\,y,\tau,\sigma)\,d\theta\,. \tag{2.1.8}$$

I.J. Good (1965) referred to the maximisation of (2.1.8) as a Type II maximum likelihood method. Since $\pi(\theta|\,y,\tau,\sigma)$ is normally distributed, (2.1.8) can be expressed in the closed form, (Akaike 1980),

$$L(\tau,\sigma) = (2\pi\sigma^2)^{-N/2}|\,\tau^2 D^T D|^{1/2}|\,X^T X + \tau^2 D^T D|^{-1/2}\exp\left\{\frac{-1}{2\sigma^2}SSE(\tau)\right\} \tag{2.1.9}$$

The maximum likelihood estimator of σ^2 is

$$\hat{\sigma}^2 = SSE(r)/N \,. \tag{2.1.10}$$

It is convenient to work with -2 log likelihood. Using (2.1.10) in (2.1.9) yields

(2.1.11)

$$-2logL(r,\theta) = Nlog2\pi + Nlog(SSE(r)/N) + \log|X^TX + r^2D^TD| - \log|r^2D^TD| + N.$$

This is the basic relation that we use in our smoothness priors least squares analysis. A practical way to determine the value of r^2 for which the -2log-likelihood is minimised, is to compute the likelihood for discrete values of r^2 and search the discrete -2log likelihood-hyperparameter space for the minimum. If there are more than say 2 hyperparameters, it might be more expeditious to use a gradient search algorithm to determine the hyperparameters that maximise the likelihood. Akaike (1980) demonstrated the first practical use of the likelihood of the Bayesian model and the use of the likelihood of the hyperparameters, as a measure of the goodness of fit of a model to data.

Several other facets of stochastic regression may be of interest. The solution of the ordinary least squares regression problem in (2.1.1) is

$$\hat{\theta}_{LS} = (X^TX)^{-1}X^Ty \,. \tag{2.1.12}$$

Matrix algebra yields

$$\hat{\theta} = (X^TX + r^2D^TD)^{-1}(X^TX\hat{\theta}_{LS} + r^2D^TD\theta_0). \tag{2.1.13}$$

That is, the posterior parameter estimate is a weighted sum of the least squares solution and the prior mean, θ_0. Let θ_t be the true value of the parameter vector θ. Then, direct evaluation of the mean square parameter vector error, $MSE(\hat{\theta}) = Var(\hat{\theta}) + E(\hat{\theta}-\theta_t)^TE((\hat{\theta}-\theta_t)$, yields the result

$$MSE(\hat{\theta}) \leq MSE(\theta_{LS}) \tag{2.1.14}$$

iff

$$tr(X^TX)^{-1} \geq tr(X^TX + r^2D^TD)^{-1} + r^2(\theta_t-\theta_0)^TD^T(X^TX + r^2D^TD)^{-2}D(\theta_t-\theta_0) \,.$$

The first term on the RHS of (2.1.14) is not larger than the LHS of (2.1.13). Depending upon how close θ_0 is to θ_t, the $MSE(\hat{\theta})$ may or may not be less than $MSE(\hat{\theta}_{LS})$. The Bayesian method

minimizes expected loss. Therefore the expected value of $MSE(\hat{\theta})$ will be less than or equal to the expected value of $MSE(\hat{\theta}_{LS})$.

2.2 SOME EXAMPLES OF SMOOTHNESS PRIORS MODELING

Two of the earliest smoothness priors problems are illustrated here. We refer to those as the Whittaker problem, (Whittaker 1923), and the Shiller problem (Shiller 1973),

The Whittaker Problem: In the problem treated by Whittaker, the observations $y_n, n=1,...,N$ are given. They are assumed to consist of the sum of a "smooth" function and observation noise,

$$y_n = f_n + \epsilon_n. \tag{2.2.1}$$

The problem is to estimate the unknown $f_n, n=1,...,N$. In a time series interpretation of this problem, $f_n, n=1,...,N$ is the trend of a nonstationary mean time series. A typical approach to this problem is to use a class of parametric models. The quality of the analysis is completely dependent upon the appropriateness of the assumed model class. A flexible model is desirable. In this context, Whittaker suggested that the solution balance a tradeoff of goodness of fit to the data and goodness of fit to a smoothness criterion. This idea was realized by determining the $f_n, n=1,...,N$ to minimize

$$[\sum_{n=1}^{N} (y_n - f_n)^2 + \mu^2 \sum_{n=1}^{N} (\nabla^k f_n)^2] \tag{2.2.2}$$

for some appropriately chosen smoothness tradeoff parameter μ^2. In (2.2.2) $\nabla^k f_n$ expresses a kth-order difference constraint on the solution f, with $\nabla f_n = f_n - f_{n-1}$. $\nabla f_n^2 = \nabla(\nabla f_n)$, etc. (Whittaker's original solution was not expressed in a Bayesian context. Whittaker and Robinson, 1924 does invoke a Bayesian interpretation of this problem.)

The properties of the solution to the problem in (2.2.1)-(2.2.2) are clear. If $\mu^2=0$, $f_n = y_n$ and the solution is a replica of the observations. As μ^2 becomes increasingly large, the smoothness constraint dominates the solution and the solution satisfies a kth order constraint. For large μ^2

and $k=1$, the solution is a constant, for $k=2$, it is a straight line etc.. Whittaker left the choice of μ^2 to the investigator.

From the Bayesian point of view, the difference equation constraints on the parameter vector problem are stochastic. That is, $\nabla^k f_n = w_n$, with w_n assumed to be an i.i.d. normally distributed zero-mean sequence with unknown variance τ^2 . For example for $k=1$ and $k=2$ those constraints are:

$$f_n = f_{n-1} + w_n; \tag{2.2.3}$$

$$f_n = 2f_{n-1} - f_{n-2} + w_n.$$

Corresponding to the matrix D in (2.1.6), for difference orders $k=1$ and $k=2$ respectively, the smoothness constraints can be expressed in terms of the NxN constraint matrices D_1 and D_2,

$$D_1 = \begin{bmatrix} \alpha & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & & \\ & 0 & & \ddots & \\ & & & \ddots & \ddots \\ & & & & -1 \quad 1 \end{bmatrix}, \qquad D_2 = \begin{bmatrix} \alpha & & & & & & \\ -\beta & \beta & & & & & \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & 0 & & \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ & & & & \ddots & \ddots & \\ & & & & & 1 \quad -2 \quad 1 \end{bmatrix}. \tag{2.2.4}$$

In (2.2.4) α and β are small numbers that are chosen to satisfy initial conditions. (An alternative to the ad hockery in specifying α and β in (2.2.4) is to estimate f_0 and f_1 by a maximum likelihood method.)

We use the parameterization $\mu=\sigma/\tau$. Therefore, μ^2 has a noise-to-signal interpretation. Larger τ corresponds to smoother trends. For fixed k and fixed τ^2 the least squares solution can be expressed in the form of (2.1.5). The matrix X and the parameter vector θ in (2.1.5) are replaced by the identity matrix I and the parameter vector $f=(f_1,\ldots f_N)$. Then for example, with $k = 2$ and $D = D_2$, the solution $\{f_n, n = 1,\ldots,N\}$ satisfies

$$\left\| \begin{pmatrix} y \\ 0 \end{pmatrix} - \begin{pmatrix} I \\ \tau D_2 \end{pmatrix} f \right\|^2 . \tag{2.2.5}$$

From (2.1.6),the solution to (2.2.5), with $D=D_2$, is

$$\hat{f} = [I + \tau^2 D_2^T D_2]^{-1} y \ , \tag{2.2.6}$$

and the value of $SSE(\tau)$ is given by (2.1.8) with $\hat{\theta}=\hat{f}, X=I, D=D_2$. The minimized value of -2log likelihood for this problem is:

(2.2.7)

$$-2logL(\hat{\tau},\hat{\sigma}) = Nlog2\pi + Nlog(\frac{1}{N}SSE(\hat{\tau})) + log|\hat{\tau}^2 D_2^T D_2 + I| - log|\hat{\tau}^2 D_2^T D_2| + N.$$

The Shiller Problem: The problem treated by Shiller is the estimation of the distributed lag (impulse response), given the jointly stationary time series observations, $\{y_n, x_n;\ n=1,...,N\}$. The distributed lag model is,

$$y_n = \sum_{m=0}^{M} h_m x_{n-m} + \epsilon_n \tag{2.2.8}$$

Frequently and also in the case of the data analyzed by Shiller, econometric data is short duration. As a result, econometricians have been motivated to Bayesian analyses. They assume a prior distribution on the parameters of the model and thereby increase the effective data length. The smoothness priors assumed on the distributed lag coefficients by Shiller were of the form, $\nabla^k h_m = w_m$, with $\{w_m, m=0,...,M\}$, a zero mean, normally distributed zero mean i.i.d sequence. Those are the same priors assumed for the Whitaker example. We take h_0 to be 0 and for simplicity consider the initial conditions, $x_0,...x_{1-M}$, as known. Then. the computational problem for the smoothness priors distributed lag parameters, is as in (2.1.6). In the application of (2.1.6) to the Shiller problem, the parameter vector $\theta=h$, and the Nxm matrix X as given below in (2.2.9) and the matrix D either D_1 or D_2 as in (2.2.4).

$$X = \begin{bmatrix} x_0 & \cdots & x_{1-M} \\ x_1 & \cdots & x_{2-M} \\ \vdots & & \vdots \\ x_{N-1} & \cdots & x_{N-M} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, h = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_M \end{bmatrix}. \tag{2.2.9}$$

2.3 A TRANSFER FUNCTION MODEL

Assume that input/output jointly stationary time series data x_n, y_n, $n=1,\ldots,N$ is observed. Assume that the output y_n is observed in the presence of additive colored noise, w_n. Consider a representation of the input/output plus noise in the impulse response plus colored noise form,

$$y_n = \sum_{m=1}^{\infty} b_m x_{n-m} + w_n \tag{2.3.1a}$$

$$w_n = \sum_{m=1}^{\infty} a_m w_{n-m} + u_n \,. \tag{2.3.1b}$$

For convenience in (2.3.1b), u_n is assumed to be a Gaussian zero-mean uncorrelated sequence with unknown variance σ_u^2. In (2.3.1a) b_m is an impulse response sequence and w_n is assumed to be in AR model form.

Using the assumed stationarity, (2.3.1a) yields

$$w_{n-j} = y_{n-j} - \sum_{i=1}^{\infty} b_i x_{n-i-j} \,. \tag{2.3.2}$$

Substituting the expression for w_{n-j} into (2.3.1b) yields the model

$$y_n = \sum_{m=1}^{\infty} c_m y_{n-m} + \sum_{m=1}^{\infty} d_m x_{n-m} + u_n \tag{2.3.3}$$

with

$$c_m = a_m, \; m=1,\ldots \tag{2.3.4}$$

$$d_m = b_m - \sum_{l=1}^{m-1} b_l a_{m-l}, \; m=1,\ldots$$

Equation (2.3.3) is an ARMAX model with additive white noise u_n. The models in (2.3.1a),(2.3.1b) are estimated using (2.3.3) and (2.3.4). The infinite order transfer function model in (2.3.3) is approximated by finite transfer function model

$$y_n = \sum_{m=1}^{M} c_m y_{n-m} + \sum_{m=1}^{M} d_m x_{n-m} + u_n \,, \tag{2.3.5}$$

with M assumed to be "large". (The choice of M may be determined by the maximization of a likelihood and Akaike's AIC.) The coefficients c_m, d_m $m=1,...,M$ are directly estimated by the Bayesian procedure described in the following section. The estimates of the coefficients of the model (2.3.1a) are then obtained by the formulas

$$a_m = c_m, \; m=1,...,M \tag{2.3.6}$$

$$a_m = 0, \; m=M+1,...$$

$$b_m = d_m + \sum_{i=1}^{m-1} a_i b_{m-i}, \; m=1,...,M$$

$$b_m = \sum_{i=1}^{M} a_i b_{m-i}, \;\; m=M+1,M+2,... \,.$$

From (2.3.1a), the frequency response function from the input x_n to the output y_n can be obtained from

$$h(f) = \sum_{m=1}^{\infty} b_m \exp[-2\pi i f m] \tag{2.3.7}$$

where $i^2=-1$. The power spectrum of the noise w_n is given by

$$S(f) = \frac{\sigma^2}{|1 - \sum_{m=1}^{M} a_m \exp[-2\pi i f m]|^{-2}} \tag{2.3.8}$$

where σ^2 is the innovations variance of the estimated model (2.3.5).

Identify the quantities $C(f)$ and $D(f)$

$$C(f) = 1 - \sum_{m=1}^{M} c_m exp[-2\pi imf], \tag{2.3.9}$$

$$D(f) = \sum_{m=1}^{M} d_m exp[-2\pi ifm] .$$

Then, a more convenient form for the frequency response function is

$$h(f) = \frac{D(f)}{C(f)} \tag{2.3.10}$$

$$= \frac{\sum_{m=1}^{M} d_m exp[-2\pi ifm]}{\left[1 - \sum_{m=1}^{M} c_m \exp[-2\pi ifm]\right]} .$$

2.4 A SMOOTHNESS PRIORS TRANSFER FUNCTION MODEL

The Whittaker and Shiller problems, Section 2.2, are examples of smoothness priors model parameter constraints in the time domain. Akaike (1979b) is very likely the first example of frequency domain smoothness priors constraints. In this section, we employ frequency domain parameter constraints that are similar to those that were successfully used in our smoothness priors modeling of AR models for spectral estimation, Kitagawa and Gersch (1985a). (Gersch and Kitagawa 1984 was an earlier frequency domain priors version of the SP transfer function method.)

Let R_k and Q_k respectively, measures of the roughness of the $C(f)$ and $D(f)$ polynomials, be characterised by the integrated square *kth* derivative of those operators,

$$R_k = \int_{-1/2}^{1/2} \left| \frac{d^k C(f)}{df^k} \right|^2 df, \quad Q_k = \int_{-1/2}^{1/2} \left| \frac{d^k D(f)}{df^k} \right|^2 df . \tag{2.4.1}$$

Then using the definition of $C(f)$,$D(f)$, equation (2.3.9), direct evaluation of (2.4.1) yields

$$R_k = (2\pi)^{2k} \sum_{m=1}^{M} m^{2k} c_m^2 , \quad Q_k = (2\pi)^{2k} \sum_{m=1}^{M} m^{2k} d_m^2 , \tag{2.4.2}$$

From the definitions in (2.4.1), large values of R_k and Q_k, respectively mean an unsmooth, in the sense of *kth* differential, frequency domain measure of the $c(.)$ and $d(.)$ polynomials. We also

introduce the *zero−th* derivative smoothness constraints

$$R_0 = \int_{-1/2}^{1/2} |C(f)|^2 df = 1 + \sum_{m=1}^{M} c_m^2 \tag{2.4.3}$$

$$Q_0 = \int_{-1/2}^{1/2} |D(f)|^2 df = \sum_{m=1}^{M} d_m^2$$

Let the differential orders for the numerator and denominator polynomials to be k_1, and k_2 respectively. With these "frequency domain" priors we then have the constrained least squares problem which for fixed values of k_1, k_2 and τ_j^2, $j=1,...,4$ determines the $\{c_m, d_m,\ m=1,...,M\}$ that minimizes

(2.4.4)

$$\sum_{n=1}^{N} [y_n - \sum_{m=1}^{M} c_m y_{n-m} - \sum_{m=1}^{M} d_m x_{n-m}]^2 + \sum_{m=1}^{M} [\tau_1^2 c_m^2 + \tau_2^2 m^{2k_1} c_m^2] + \sum_{m=1}^{M} [\tau_3^2 d_m^2 + \tau_4^2 m^{2k_2} d_m^2].$$

In (2.4.4) τ_j^2 $j=1,....4$ are the tradeoff parameters. By a proper choice of the tradeoff parameters, our estimate of the model parameters, $\{c_m, d_m,\ m=1,...,M\}$, balance a tradeoff between infidelity of the transfer function solution to the data and infidelity to the smoothness constraints.

2.5 THE SMOOTHNESS PRIORS LEAST SQUARES PROBLEM, DETERMINING THE TRADEOFF PARAMETERS

As indicated in Section 2.1, the constrained least squares problem has a Bayesian interpretation which facilitates the determination of the tradeoff parameters in the criterion. We apply that approach to the particular smoothness prior transfer function problem at hand.

In detail, the minimization of (2.4.4) is equivalent to the maximization of

(2.5.1)

$$\exp\left\{\frac{-1}{2\sigma^2}\sum_{n=1}^{N}[y_n - \sum_{m=1}^{M} c_m y_{n-m} - \sum_{m=1}^{M} d_m x_{n-m}]^2\right\}\exp\left\{\frac{-1}{2\sigma^2}\sum_{m=1}^{M}(\tau_1^2 + \tau_2^2 m^{2k_1})c_m^2\right\}\exp\left\{\frac{-1}{2\sigma^2}\sum_{m=1}^{M}(\tau_3^2 + \tau_4^2 m^{2k_2})d_m^2\right\}.$$

In that form, the constrained least squares solution has a Bayesian interpretation as the maximum a posteriori estimate of the model with the data distribution

$$p(y|\, X,\theta,\sigma) = (2\pi\sigma^2)^{-N/2}\exp\left\{\frac{-1}{2\sigma^2}\sum_{n=1}^{N}|y_n - \sum_{m=1}^{M} c_m y_{n-m} - \sum_{m=1}^{M} d_m z_{n-m}|^2\right\} \tag{2.5.2}$$

and the prior distribution

$$\pi(\theta|\, \tau,\sigma) = (2\pi\sigma^2)^{-M}|\, D^T D|^{\,1/2} exp[\frac{-1}{2\sigma^2}\theta^T D^T D\theta] \tag{2.5.3}$$

where τ denotes the vector of hyperparameters $\tau_1,\ldots\tau_4$, and θ denotes the model parameters, $c_m, d_m, m=1,\ldots,M$. In (2.5.3)

$$D = \begin{bmatrix} (\tau_1^2+\tau_3^2)^{1/2} & & & & \\ & (\tau_2^2+\tau_4^2)^{1/2} & & & \\ & & (\tau_1^2+2^{2k_1}\tau_3^2)^{1/2} & & \\ & & & \ddots & \\ & & & & (\tau_2^2+M^{2k_2}\tau_4^2)^{1/2} \end{bmatrix}, \theta = \begin{bmatrix} c_1 \\ d_1 \\ \vdots \\ c_M \\ d_M \end{bmatrix}, \tag{2.5.4}$$

From (2.5.2) and (2.5.3) it follows that

$$p(y|\, X,\theta,\sigma)\pi(\theta|\, \tau,\sigma) = \tag{2.5.5}$$

$$(2\pi\sigma^2)^{-(N+2M)/2}|\, D^T D|^{\,1/2}\exp[\frac{-1}{2\sigma^2}(\hat{\theta}-\theta)^T(X^T X + D^T D)^{-1}(\hat{\theta} - \theta)]\exp[\frac{-1}{2\sigma^2}SS(\tau)],$$

where

$$X = \begin{bmatrix} y_0 & z_0 & \cdots & y_{1-M} & z_{1M} \\ y_1 & z_1 & \cdots & y_{2-M} & z_{2-M} \\ \vdots & \vdots & & \vdots & \vdots \\ y_{N-1} & z_{N-1} & \cdots & y_{N-M} & z_{N-M} \end{bmatrix}, \quad z = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \tag{2.5.6}$$

and

$$\hat{\theta} = (X^T X + D^T D)^{-1} X^T z\,, \tag{2.5.7}$$

$$SS(\tau) = z^T z - \hat{\theta}^T(X^T X + D^T D)\hat{\theta}\,.$$

In (2.5.6), the initial condition data $\{z_i, y_i,\ i = -M,-M+1,\ldots,0\}$ are assumed known. Integrating

(2.5.5) with respect to the parameter vector yields the likelihood of the hyperparameters,

$$L(\tau,\sigma) = (2\pi\sigma^2)^{-N/2}|\,D^TD|^{1/2}|\,X^TX + D^TD\,|^{-1/2}\exp[\frac{-SS(\tau)}{2\sigma^2}]. \qquad (2.5.8)$$

Then

$$\log L(\tau,\sigma) = \frac{-N}{2}\log 2\pi\sigma^2 + \frac{1}{2}\log|\,D^TD| \;\; \frac{-1}{2}\log|\,X^TX + D^TD| \; - \frac{1}{2\sigma^2}SS(\tau), \qquad (2.5.9)$$

with the maximum likelihood estimate of σ^2 given by

$$\hat{\sigma}^2 = \frac{1}{N}SS(\tau)\,. \qquad (2.5.10)$$

Substituting the estimated value of σ^2 from (2.5.10) into (2.5.9) yields

$$\log L(\tau,\sigma) = \frac{-N}{2}\log 2\pi\sigma^2 + \frac{1}{2}\log|\,D^TD| \; - \frac{1}{2}\log|\,X^TX + D^TD| \; - \frac{N}{2}\,, \qquad (2.5.11)$$

which is to be maximized to obtain the maximum likelihood estimates of τ_j^2 $j=1,\ldots,4$. The likelihood for the hyperparameters is maximised via a Davidon-Fletcher-Powell gradient search algorithm. That algorithm is exterior to a Householder transformation least squares solution of the constrained least squares problem.

We use the AIC statistic, (Akaike 1973),

$$AIC = -2\log L(\tau,\sigma) + 2(\textit{number of parameters estimated}) \qquad (2.5.12)$$

$$= N(\log 2\pi\sigma^2 + 1) - \log|\,D^TD| \; + \log|\,X^TX + D^TD| \; + 2(\textit{number of parameters estimated}),$$

to determine the order M for the transfer function. Akaike (1980) referred to (2.5.12) as the likelihood of the Bayesian model. The analysis indicated here is referred to as a "quasi-Bayesian " analaysis. A more thorough Bayesian solution of the transfer function estimation problem, would require that priors be specified on the model order M. A completely orthodox Bayesian analysis would require priors on the model order and parameters of the stochastic input to the system.

3. ANALYSIS OF BOX-JENKINS SERIES J DATA AND MONTE CARLO RESULTS

In this section the results of the transfer function analysis of the Box-Jenkins Series J gas furnace data by the smoothness priors and Box-Jenkins methods are described and compared. Some properties of the smoothness priors method of analysis are shown. Also, we show the results of Monte Carlo studies of the statistical performance of the smoothness priors, (SP), method and the 3SLS asymptotically maximum likelihood method of Mayne and Firzoon ,(1978), based on models derived from the BJ series J data.

The input output BJ series J data are shown in Figure 1. The variances of the input and output data are 1.14727 and 10.25357 respectively. For illustrative purposes, the data shown in Figure 1 was normalised to have zero mean and the same variance. (Inverting the output data and superimposing it over the input data reveal the output to be a delayed-low pass filtered version of the input.) The generic Box-Jenkins transfer function model is

$$y_n = a_1 y_{n-1} + \cdots - a_p y_{n-p} + b_1 x_{n-1-d} + \cdots + b_q x_{n-q-d} + w_n \tag{3.1}$$

$$w_n = c_1 w_{n-1} + \cdots + c_r w_{n-r} + u_n.$$

In (3.1) $\{x_n, y_n, w_n, , n=1,\ldots,N\}$ are respectively the observed input and output and the unobserved added noise. Also in (3.1), $\{u_n, n=1,\ldots,N\}$ is a normal zero mean i.i.d. random variable with variance σ_u^2. The dimensional parameters of the BJ model are $d=2$, $p=2$, $q=3$, $r=2$. The published vectors of BJ model coefficients are: $a=(0.57,0.01); b=(0.53,-0.37,-0.51); c=(1.53,-0.63)$, $\sigma_u^2 = 0.05058$, (Box-Jenkins, Section 11.4). For the AIC optimal, $k_1=4, k_2=2$, SP model, the dimensional parameters are $p=q=r=4$. The $d=0$ model is the AIC best shift parameter model. (In this data example, the higher order SP model automatically accounts for the delay between input and output data, without requiring an additional non-zero d parameter.) The SP a,b polynomial coefficients are $a=(1.58824, -0.70509, -0.13198, 0.14861)$, $b=(0.17090, -0.43813, -0.17497, 0.08608)$. The vector of smoothness priors tradeoff parameters was $r^2=(0.00886, 0.10305, 0.71914, 0.10686)$.

Table 1 shows the values of the AIC for the differential orders $k_1,k_2=1,...,4$ for the order $M=4$ model.

INSERT FIGURE 1 HERE: BOX-JENKINS SERIES J DATA

Table 1. AIC's d=0, M=4, SP Model, Parametric in k1 and k2.

k2	k1 = 1	k1 = 2	k1 = 3	k1 = 4	k1=5
1	46.886	45.697	45.663	46.345	47.375
2	46.436	46.186	45.316	45.995	46.630
3	46.273	45.031	45.120	45.794	46.436
4	46.672	45.013	45.096	45.769	46.414
5	46.410	45.072	45.072	45.721	46.388

In Figure 2, the original Box-Jenkins Series J output data, and the SP model tracked output data are shown superimposed. A vertical scale displaced version of the difference between the original and tracked data is also shown in Figure 2. The tracked output data is computed by passing the input through the estimated model. The appearance of the $N=294$ tracked data version of the BJ modeled data, incorporating the $d=2$ parameter, appears very similar to that of the SP modeled data and is not shown here.

In the sense of minimum mean square tracking error, the performance of the AIC optimum SP and BJ models were similar. The variances of the tracking error for the BJ and SP models (the sums of squares of the residuals, SSE), were 0.70187 and 0.68662 respectively. The ratio of the relative variance of the tracking error to the variance of the true output was 0.06798 and 0.06696 for the BJ and SP model respectively.

INSERT FIGURE 2 HERE: TRACKED DATA

The impulse response, transfer function amplitude response, phase response and power spectral densities (psds) of the residuals associated with the BJ and SP models are shown in Figure 3. (Compare the transfer function and residual spectrum of a windowed periodogram analysis of the B-J data, Jenkins and Watts, p446.) The residual psds were computed from Householder transformation-Akaike AIC criterion AR models. The coefficients of the corresponding AIC best AR_4 models of the BJ and SP residuals were {1.53458,−0.55879,−0.21378,0.16280} and {1.58329,−0.65524,−0.17195,0.17461} respectively with corresponding innovations variances 0.05995 and 0.05755. In Figure 3, the SP and BJ modeled psds of the residuals are almost identical. The $AR-AIC$ model of the residual of the SP modeled BJ Series J data shown in Figure 3 is quite similar in appearance to that obtained automatically by the SP modeling procedure and computed directly by equation (2.3.8).

Also in Figure 3, after the first 3 time points, when the impulse response for the BJ model is zero, the impulse responses of the BJ model and SP model appear similar. The SP model impulse response is smoother than the BJ impulse response. The initial non zero going part of the SP impulse response is a consequence of the fact that the optimal SP model delay parameter d is zero.

The BJ modeled transfer function and phase function versus frequency each have some relatively abrupt kinks in their responses as compared to those for the SP modeled results.

INSERT FIGURE 3 HERE: IMPULSE RESPONSE, TRANSFER FCN & PHASE & NOISE PSD's SP & BJ MODELS

Some comments on the stationarity of the Box-Jenkins Series J data are in order here. In Figure 2, the true BJ output data and the SP modeled data appear more discrepant in the latter part than the earlier part of the time series. Also in Figure 2, there are relatively large excursions in the latter part of the residual time series. That evidence suggests that the Series J data is nonstationary. To examine that conjecture, we examined the residuals of the AR modeled tracked data, fitted SP models to the first 200 data points and to the first 75 data points of the Series J

data and examined the tracking behavior of those models. The properties of the BJ series J data do appear to change slightly after $n=225$.

The impulse response, noise spectrum, transfer function and phase function for the $k_1=1,k_2=1$ SP models for $n=200$ and $n=75$ data are shown superimposed in Figure 4A and 4B respectively. As expected from Figure 2, the properties of the SP modeled $n=296$ and $n=200$ series J data are quite similar. The properties of the SP modeled $n=296$ and $n=75$ data are also quite similar. The conjecture that the SP modeling method might be reasonable for relatively short length data spans is supported by the evidence shown in Figure 4. The apparent property of the SP procedure to yield reliable parameter estimates with relatively short data length time series is a consequence of the assumption of priors on the model parameters. The priors are equivalent to the observation of additional data.

For completeness, the a,b polynomial coefficients corresponding to SP $M=4,k_1=1,k_2=1$ modeled data are respectively: {1.13620,−0.15871,−0.29041,0.13128}, {1.08946,−0.24032,−0.55796,0.11972} for the $n=200$ data and {1.25647,−0.20495,−0.35709,0.16760}, {1.18340,−0.54000,−0.36671,0.28644} for the $n=75$ data. The respective residual variances of the SP modeled $M=4$, $n=200$ and $n=75$ data point models were 0.09900 and 0.04629 respectively. The corresponding relative tracking variance ratios were 0.01078 and 0.00620.

INSERT FIGURE 4 HERE: IMPULSE RESPONSE etc n=200 and n=75 Models

The effects of the choice of model order, M, and the differential orders k_1,k_2 on the impulse response, the noise spectrum and the transfer function amplitude and phase of the SP model are also of interest.

First, to illustrate the effect of model order M on transfer function model properties, graphs of impulse response, amplitude and phase response are shown superimposed in Figure 5A,B,C for the optimal SP $M=4$ model and the likelihood best SP models of orders $M=10,20$ and $M=30$. The

graphical results for the higher order SP models wiggle only slightly around those for the order M=4 model. Those results indicate that, on the provision that the model order is sufficiently large, the specification of the order of the SP model does not very critically influence the transfer function characteristics. For completeness, some of the computed results for those models are: M=10: k_1=1,k_2=4,AIC=47.789; M=20: k_1=2,k_2=4,AIC=50.224; M=30: k_1=2,k_2=4,AIC=56.886. The similarity in the appearance of the model properties for the different model orders is compatible with the similarity in the values of the AIC for the different models.

The values of the M=10 optimal SP model, a and b polynomial coefficients are:

a = {1.56980,−0.70201,−0.02424,0.06248.0.01102,−0.00000,−0.00152,−0.00046,0.00004,0.0010},

b = {0.15110,−0.38420,−0.17713,−0.00672.0.03029,0.08646,0.04176.0.00577,−0.01716,−0.00717}.

The pattern of a,b polynomial coefficients is similar for the larger order SP models. The tapering toward zero values effect of the smoothness priors constraints on the model parameters, particularly on the higher order a polynomial parameters and the relatively long tail b model parameters helps explain the similarity of the M=4,10,20 and M=30 model properties. The b model parameters in the numerator of the rational polynomial description of the model do not have as dramatic an effect as do the a polynomial denominator polynomial parameters on the model properties. The long tail b polynomial parameters and the short tail a polynomial parameters are well approximated by the SP M = 4 model.

For the purposes of comparison, we also fitted ordinary least squares, (OLS), models of orders M=5,10 and M=20 to the BJ series J data. Graphical results of the impulse response, transfer function and phase function for those models are shown in Figures 5D,E,F. The M=5 LS model properties are very similar to the optimal SP model properties. (The relative variance of the tracking error was 0.06706. The OLS M=5 model is actually a superior model of the BJ series J data than the original BJ model.) The computed properties of the OLS M=10 and M=20 models wiggle a lot more around the SP M=4 model than the SP M=30 model. This is very clear evidence that the SP model properties are relatively insensitive to model order in comparison with other,

conventional transfer function modeling methods.

INSERT FIGURE 5 HERE: EFFECT OF MODEL: ORDER M

In order to illustrate the effects of increasing differential orders k_1, k_2 on transfer function properties, graphs of those properties are shown for SP $k_1=1, k_2=1$ and $k_1=9, k_2=9$ models of order $M = 4$ in Figure 6. The amplitude response, noise spectrum and phase response of the higher order smoothness constraint differential models are smoother than those for the lower differential order model. This behavior could be anticipated because the priors are frequency domain roughness constraints.

INSERT FIGURE 6 HERE EFFECT OF DIFFERENTIAL ORDERS k_1, k_2

We note that the formula for the AIC in (2.5.12) can not be applied to determine the best of alternative models with different initial orders M. In the notation of (2.5.6), the "initial conditions" values of $(z_0, y_0, t = 1-M, 2-M \ldots 0)$ were assumed known. In modeling data with an SP model of order M, we customarily take $(z_0, y_0, t = 1, \ldots M)$ as initial conditions and model the data on the remaining $N-M$ data points. The likelihood is actually computed for the last $N-M$ observations, $(y_{M+1}, \ldots y_N)$. In that case, models of different M orders are modeled on different data and it is not appropriate to use the AIC to distinguish between models. A formula that permits the AIC's of models of different orders to be compared is.

$$AIC = \frac{N}{N-M} 2\log\text{-likelihood} + 2(\text{number of parameters estimated}). \qquad (3.2)$$

The formula in (3.2) is reasonable under the assumption that the data is stationary, i.e. the properties of the first M vales of the data do not change very much with different values of M.

Monte Carlo Results

Monte Carlo simulation studies were performed to explore the statistical performance of the SP method of transfer function estimation. Some comparisons of the SP and Mayne and Firzoon (1978) 3SLS procedure were also done. We chose to compare the SP with the 3SLS procedure because the latter procedure has asymptotic MLE properties and is easy to program. The principal topics of interest are the bias and MSE parameter estimation properties and transfer function confidence interval properties of the SP and the 3SLS procedures. We show computational results of simulation studies of the SP and 3SLS models with additive AR and additive MA observation noises.

Consider the transfer function model

$$y_n = a_1 y_{n-1} \quad \cdots + a_p y_{n-p} + b_1 x_{n-1} + \cdots + b_q x_{n-q} + v_n \tag{3.3}$$

with v_n an added noise process. The MA_r noise model is: $v_n = \sum_{m=0}^{r} c_m u_{n-m}$ where $\{u_n\}$ is a zero mean i.i.d. process. The AR_r noise model is $v_n = \sum_{m=1}^{r} c_m v_{n-m} + u_n$, where again $\{u_n\}$ is a zero mean i.i.d. process. The AR observation noise model is the Box-Jenkins model. The MA observation noise model was used in Astrom and Bohlin (1965). Since then it has been used extensively in engineering applications.

The 3SLS procedure was developed as an alternative to the comptationally extensive maximum likelihood method for the MA observation noise model. For convenience, the 3SLS procedure is as follows:

Let a,b.c denote the AR, MA and added MA noise polynomials respectively in equation (3.3).

i)Using least squares (LS), fit a "long" a.b polynomial model, to the $\{x_n, y_n, n=1,\ldots,N\}$ input-output data and compute the residual time series.

ii) Fix the orders of the a.b and c polynomials to their final model orders and use the original input-output data and the residuals from stage i) to estimate the a.b and c polynomial coefficients

by LS.

iii) Prewhiten the input and output data using the inverse of the c polynomial determined in stage ii) and estimate the fixed order a,b polynomials coefficients by LS.

We fit the 3SLS model to the original Box-Jenkins Series J data in order to verify the relevance of that procedure for a comparison of results AR observation noise Monte Carlo study. An first stage 3SLS a,b polynomial model order $p,q=8$ was determined by trial and error. A similar procedure for determining the stage one model order was used in Hannan et al. (1986). The 3SLS $M=4$ model parameters were: $a=(1.58607,-0.67426,-0.17672,0.16647)$, $b=(0.19765,-0.44775,-0.24883,0.17300)$. The appearance of the superimposed SP $M=4$, $k_1=1,k_2=1$ and 3SLS modeled impulse response, transfer function and phase response were visually indistinguishable. On the basis of this evidence, it was thought reasonable to examine the performance of the 3SLS transfer function model with AR observation noise that was similar to the observation noise in the BJ series J data.

The model that we used to synthesize data for the Monte Carlo simulations is a variation of the model of the BJ series J data. The input data for the simulation was the Box-Jenkins Series J input data. For the first set of trials, the added noise was a stochastic version of an AR_4 model of the residual noise from the SP fit to the Box-Jenkins series J data. The a,b coefficients of the noiseless simulation model were $a=(1.66283,-0.64256,-0.30648,0.22377)$ $b=(-0.83218,-0.47872,-0.24869,0.12831)$. The AR_4 model coefficients were $c=(1.69069,-0.69023,-0.28565,0.022507)$, $\sigma^2=0.284$. The (biased) SP model parameters fitted to a noiseless version of that data were $a=(1.73481,-0.21383,-0.92282,0.40184)$, $b=(-0.03550,-0.05907,0.04443,0.05015)$. The vector of hyperparameters was $r=(0.000062,0.000004,0.000027,0.000003)$. Such small values should not be surprising because to within roundoff errors, the noiseless data is exactly an $AR_4 MA_4$ model.

Results of the statistical properties of 25 replications from the SP and 3SLS models for $n=296$, AR_4, data points are shown in Table 2. The output data is regressed partially upon itself

so the 3SLS procedure as well as the SP procedure, will yield biased coefficient estimates. (The magnitude of the biases is model dependent.) The bias of the SP modeled parameters is defined as the difference between the mean SP parameters and the zero added noise 3SLS model parameters. The standard deviation and bias errors are comparable for both the the SP and 3SLS procedures. (The standard deviation of the SP modeled estimated b polynomial parameters are actually somewhat smaller than those for the 3SLS procedure.)

Table 2 AIC's, SP & 3SLS M=4 Models

	SP M = 4			3SLS M = 4		
param	mean	std. dev.	bias	mean	std. dev.	bias
ar1	-1.6436	0.0541	-0.0820	-1.6641	0.0554	-0.0615
ar2	-0.7746	0.1093	-0.5833	-0.8395	0.1052	-0.6483
ar3	-0.1577	0.1082	-0.7836	-0.0911	0.1054	0.8502
ar4	0.2040	0.0524	-0.2029	0.1775	0.0538	-0.2294
ma1	-0.0404	0.0217	-0.0050	-0.0596	0.0484	0.0243
ma2	-0.0284	0.0146	0.0310	-0.0181	0.0893	0.0414
ma3	-0.0161	0.0126	0.0281	-0.0078	0.0861	0.0519
ma4	-0.0192	0.0106	-0.0200	-0.0511	0.0494	-0.1019

Figure 7 is an illustration of the mean and plus and minus on sigma of results estimated from the Monte Carlo trials. The illustrations correspond to the simulation results reported in Table 2. From Figure 7 it appears that the overall mean square error in transfer function estimate is slightly smaller for the SP than for the 3SLS method. The similarity of the SP and 3SLS simulation results is compatible with the similarity of performance of those models on the BJ series J data.

INSERT FIGURE 7 HERE SP AND 3SLS TRANSFER FCN MEAN AND STD. DEVS.

The order $4a,b$ polynomial 3SLS model tends to be overparameterized. In order to verify that the 3SLS statistical results shown in Table 2 were representative, an ARMAX 2,2,2 model was also simulated and modeled by the 3SLS and SP M=4 model procedures. As before, the input was the Box-Jenkins series J input data. The additive stochastically modeled noise was an AR_2

version of the residual of the SP modeled series J data. The simulation model parameters were $a=(1.46778,-0.55192), b=(-0.03586,-0.06788), c=(1.58778,-0.72461)$, $\sigma^2=0.078$. The statistical results of 25 replications of fitting the SP and 3SLS models to the trial data are shown in Table 3. Here, the bias of the SP modeled parameters is defined as the difference between the mean and zero added noise SP model parameters. The standard deviation and bias of the SP $M=4$ model parameters are comparable to those in Table 2. The standard deviation and bias of the $M=2$ 3SLS model parameters are similar to those observed in Table 2 for the $M=4$ model.

Table 3. AIC SP M=4, & 3SLS M=2 Models

	SP M = 4			3SLS M = 2		
param	mean	std. dev.	bias	mean	std. dev.	bias
ar1	1.5846	0.0566	-0.1502	-1.1562	0.0386	-0.1633
ar2	-0.7048	0.1072	-0.4910	-0.0095	0.0356	-0.1171
ar3	-0.0323	0.0984	0.8905	-	-	-
ar4	0.0486	0.0415	-0.3532	-	-	-
ma1	-0.0443	0.0222	-0.0088	-0.0095	0.0406	0.0264
ma2	-0.0362	0.0383	0.0229	-0.0918	0.0436	-0.0231
ma3	-0.0001	0.0243	-0.0451	-	-	-
ma4	-0.0299	0.0236	0.0600	-	-	-

We recall that the added noise for the Table 2 data was AR_4 and that for Table 3 was AR_2. The consistency of the tabulated results for the SP model in Tables 2 and 3 suggest that the SP model is reasonably robust with respect to noise color and noise model order.

Finally, we show results of simulation studies with added MA observation noise. The primary objects of interests in these computational experiments were a comparison of the SP and 3SLS modeling performance and the sensitivity of the SP transfer function modeling method to choices of model order M. the differential orders k_1, k_2, the observation noise level and the sensitivity to data length.

The model for these simulations was a slight variation of the model for the AR observation noise simulations. The superimposed impulse response, transfer function and phase function of the 3SLS and SP $M=4, M=5$ and $M=10$ on the noise free data were visually indistinguishable. In the first stage of the 3SLS method a model order of 20 was used. The stochastic input signal was an

AR_4 version of the input of the series J data. The 3SLS M=4 noise free data a,b polynomials were: a = {1.72566,−0.19127,−0.94126,0.40697}, b = {−0.03550,−0.05940,0.04412,0.05070}.

The observation noise was an MA_3 model, (inadvertently MA_3, instead of MA_4). The noise model parameters were c = {1.0,0.6,0.3,0.1}. This model has a fairly flat psd spectrum. A sample run of the MA_3 noise yielded an AR_{10} $AIC-AR$ model of that noise.

Monte Carlo trials with n=296 and n=78 were done. Superimposed output and output plus noise and a displaced version of the noise are shown in Figure 8 for typical sample trials of data lengths n=296 and n=78. Figure 8 illustrates the noise level of the experiments and how dramatically short the n=78 data span is.

Differential orders k_1 and k_2 best SP M=5,M=10 and M=20 transfer function models were fitted to a sample trial of the n=296 data with the following results: M=5, k_1=1, k_2=1, AIC=−1567.545, M=10, k_1=4, k_2=1, AIC=−1609.348, M=20, k_1=4, k_2=3, AIC=−1557.410. The similarity of the AIC values suggest that these models will have similar properties.

The computed results of the MA_3 noise simulation runs for the $3SLS$ M=4, SP M=10, k_1=1,k_2=4 and SP M=20, k_1=2,k_2=4, SP M=20, k_1=1,k_2=3, SP M=20, k_1=3,k_2=5 and an SP M=20, k_1=2,k_2=4 model of data with observation noise variance 9 times that of the previous three simulation trials are shown in Figure 9A,B,C,D,E,F. Tabulated values of the mean and standard deviation of the parameter estimates for the 3SLS and SP M=10, M=20 k_1=2,k_2=4 simulation trials are shown in Table 4. The 3SLS and SP M=10 results transfer function and phase function results are reasonable similar. The relative jaggedness of the impulse response for the 3SLS model indicates that a lower order 3SLS model might be more suitable for this data. The SP M=20, k_1,k_2=4 model results appear very similar to the results for the M=10 models. The results of the Monte Carlo runs, with SP models M=20,k_1=1,k_2=3 and M=20,k_1=3,k_2=5, (Figures 9D,E), are very similar to those shown for the M=20,k_1=2,k_2=4, (Figure 9C), results. Those illustrations show the relative insenstivity of SP modeling to variations in the differential orders k_1,k_2.

As expected, the results in Figure 9F for the SP M=20 model the 9 times variance trials show more dispersion than the results for the Figure 9C trials. The graphical results in Figure9 are compatible with the performance of the 3SLS and SP models of the Box-Jenkins series J data.

Table 4. AIC, 3SLS M=4 & SP M=10,20,20 Models

	3SLS M=4		SP M=10		SP M=20		SP M=20, VAR x 9	
param	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
ar1	0.7217	0.0469	0.6420	0.0356	0.6100	0.0472	0.5836	0.0615
ar2	0.0001	0.0588	-0.0109	0.0329	-0.0154	0.0321	-0.0459	0.0532
ar3	0.0490	0.0630	-0.0013	0.0111	-0.0000	0.0019	-0.0036	0.0068
ar4	-0.0041	0.0397	0.0031	0.0135	0.0002	0.0001	-0.0001	0.0005
ar5	-	-	0.0016	0.0075	0.0001	0.0001	0.0001	0.0002
ma1	-0.0387	0.0235	-0.0363	0.0210	-0.0437	0.0248	-0.0547	0.0463
ma2	-0.1115	0.0498	-0.0958	0.0366	-0.0906	0.0366	-0.0907	0.0621
ma3	0.0060	0.0557	-0.0798	0.0318	-0.0884	0.0308	-0.0675	0.1091
ma4	-0.1526	0.0333	-0.0719	0.0277	-0.0674	0.0252	-0.0879	0.0608
ma5	-	-	-0.0467	0.0243	-0.0628	0.0253	-0.0704	0.0539

In Table 4 only the results for the first 5 a,b polynomial coefficients are shown. That is sufficient to understand the results of the Monte Carlo trials. The standard deviations for the 3SLS and M=10 SP trials are comparable. Those results are also comparable to those shown in Table 2 for the AR noise trials. The a polynomial coefficients for the SP trials tend to taper quickly and correspondingly, the standard deviations for those coefficients tend to be smaller than the standard deviations for the larger valued coefficients. The values of the b polynomial coefficients do not taper and correspondingly neither do the standard deviations of those coefficient estimates.

The final Monte Carlo trials are for the shorter length n=78 data. An exploratory computation on a single trial of an N=78 data SP M=10 model, (that yields an N=68 data points for actual modeling), indicated that the k_1=2,k_2=1 model was optimum for that data. No experiment was performed for the SP M=5 model. Graphical results for 25 SP M=5,k_1=1,k_2=1 and SP M=10,k_1=1,k_2=1 models *MA* noise Monte Carlo trials are shown in Figure 10. Numerical results corresponding to those shown in Figure 10 are in Table 5. (Only results for the first 6 polynomial

coefficients are shown.) The standard deviations for the $n=78$ data coefficients are in general larger than those for the $n=296$ data coefficients. The SP $M=10$ $n=78$ data parameter mean values and standard deviations indicate the same kind of tapering effects as was observed for the $n=296$ data. The graphical results for the SP transfer function modeled $n=296$ and $n=78$ data look very similar. This is additional evidence to support the conjecture that Bayesian smoothness priors transfer function modeling can yield reliable transfer function estimation with relatively short duration data.

Table 5. AIC, 3SLS M=4 & SP M=10,20,20 Models

	SP M = 4		SP M = 10	
param	mean	std. dev.	mean	std. dev.
ar1	0.7217	0.0848	0.6623	0.0816
ar2	0.0196	0.0850	0.0537	0.0671
ar3	0.0382	0.0678	0.0186	0.0605
ar4	-0.0030	0.0331	0.0164	0.0318
ar5	-0.0151	0.0297	0.0153	0.0222
ar6	-	-	0.0062	0.0220
ma1	-0.0448	0.0248	-0.0416	0.0551
ma2	-0.0772	0.0344	-0.0971	0.0753
ma3	-0.0735	0.0335	-0.0663	0.0426
ma4	-0.0521	0.0322	-0.0616	0.0352
ma5	-0.0512	0.0342	-0.0330	0.0382
ma6	-	-	-0.0125	0.0162

4. INTERPRETATION OF RESULTS, SUMMARY AND DISCUSSION

A smoothness priors approach to the problem of transfer function estimation was demonstrated. The smoothness priors are stochastic constraints on the parameters of the linear model. The Bayesian smoothness priors procedure is one possible way of utilising the information in the likelihood function. The critical computation in the Bayesian smoothness priors approach is the likelihood of the Bayesian model. The likelihood is used as an objective measure of the goodness of a model and the hyperparameters which maximise the likelihood are determined by a gradient search method.

Some of the consequences of the smoothness priors approach to transfer function estimation are:

1) The complex multiple model orders selection procedures of other transfer function estimation methods are obviated by the smoothness priors method. Instead, the specification of the values of model order, M, and the differential orders, k_1,k_2, appear not to be very critical in determining the impulse response, transfer function and phase properties of the model.

2) Least squares and conventional maximum likelihood are abandoned as a criterion for modeling data. Instead, constrained least squares criterion or equivalently penalized likelihood methods are used in the smoothness prior method.

3) The smoothness priors model will in general not be as parsimonious, in the number of model parameters, as the conventional least squares or maximum likelihood transfer function estimation methods.

4) Smoothness priors modeling will tend to be more robust with respect to the specification of the observation noise color than the conventional methods.

5) Smoothness priors modeling will tend to yield reliable parameter estimates with shorter length data than conventional methods.

The relatively large model order non-parsimoniousness property of the Bayesian smoothness priors model can be justified. In any alternative Bayesian transfer function modeling, the Bayesian

model would be an average of the transfer function computed by models of the different a,b,c polynomial model orders weighted by the likelihood and the priors on model orders. The Bayesian transfer function model would have contributions from the largest a,b,c polynomial model orders considered. As a result, the overall Bayesian model orders would be the identical to the orders of that largest orders model. Thus, the Bayesian procedure will have a,b,c polynomial model orders larger than the model selected say by the minimum AIC procedure. By the specification of a large model order and smoothness priors constraints on the model parameters, the smoothness priors method achieves the effect of the alternative Bayesian methods of averaging the computational results of different order models.

The use of frequency domain priors is a relatively new idea. The class of frequency domain priors that we used lack a definite physical interpretation. They seem reasonable because they penalise the higher order polynomial coefficients with increasing weights. As expected, the SP model properties are increasingly smooth with increasing kth derivative constraints. Also, the overall optimum solution is not necessarily the smoothest solution. An important property of our frequency domain priors is that they permit the problem of transfer function estimation to be cast within the framework of the linear model.

The Monte Carlo results suggest that the smoothness priors method of transfer function estimation achieves comparable statistical performance with the asymptotically efficient procedures.

In summary. the smoothness priors method of transfer function estimation appears to yield statistical performance results that are comparable and perhaps superior to other well known methods while enjoying a more forgiving parameter computational search procedure than the search procedures of other methods. In general, the results shown attest to the flexibility of the Bayesian modeling approach.

REFERENCES

Akaike, H. (1973), Information Theory and an Extension of the Maximum Likelihood Principle, in Second International Symposium in Information Theory, B.N. Petroc and F. Caski eds. Budapest, Akademiai Kiado, 267-281.

Akaike, H. (1979),A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Building, Biometrika, 66, 237-242.

Akaike, H.(1979), Smoothness Priors and the Distributed Lag Estimator, Dept. of Statistics, Stanford Univ., Stanford, CA., T.A. Anderson, Project Director, Tech. Report 40.

Akaike, H. (1980), Likelihood and the Bayes procedure, In, Bayesian Statistics, J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith eds., University Press, Valencia , Spain, 143-166.

Akaike H., Ozaki T.,Ishiguro M.,Ogata Y.,Kitagawa G.,Tamura Y-H.,Arahata E., Katsura K. and Tamura Y.,(1985), TIMSAC-84,Part 1 and Part 2, Vols. 22 and 23, Computer Science Monographs, The Institute of Statistical Mathematics, Tokyo, Japan.

Astrom K.J.A. and Bohlin T. (1965), Numerical Identification of a Linear Dynamical System from Normal Operating Records, IFAC Symposium on the Theory of Self-Adaptive Control, Teddington, 96-111.

Astrom K.J.A. and Mayne D.Q. (1982), A New Algorithm for Recursive Estimation of Controlled ARMA Processes, Proceedings 6th IFAC Symposium on Identification and System Parameter Estimation, ed. G.A. Bekey and G.W. Saridis, Oxford, Pergamon, 175-179.

Box G.E.P and Jenkins G.M.(1970),Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco.

Broemeling , L.D. (1985), Bayesian Analysis of Linear Models, Marcel Dekker, New York.

Brotherton T. and Gersch W. (1981), A Data Analytic Approach to the Smoothing Problem and Some of its Variations. Proceedings of the 20th IEEE Conference on Decision and Control, 1061-1069.

Clarke, D.W. (1967), Generalized Least-squares Elimination of the Parameters of a Dynamic Model, Proceedings IFAC Symposium in Identifcation in Automatic Control Systems, Prague, Paper 3.17.

Durbin, J. The Fitting of Time Series Models, (1960), International Statistical Review, 28, 233-244.

Gersch, W. and Kitagawa, G. (1983), The Prediction of Time Series with Trends and Seasonalities, Journal of Business and Economic Statistics, 1,253-264.

Gersch, W. and Kitagawa, G. (1984), A Smoothness Priors Method for Transfer Function Estimation, Proceedings of the 23rd IEEE Conference On Decision and Control, xxx-xxx.

Good I.J. (1965), The Estimation of Probabilities, Cambridge, Mass., MIT Press.

Good I.J. and Gaskins J.R. (1980), Density Estimation and Bump Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data, 75,42-73 Journal of the American Statistical Association, 75,42-73.

Hannan, E.J. and Rissanen J. (1982), Recursive Estimation of Mixed Autoregressive-Moving Average Order, Biometrika, 69,81-94.

Hannan, E.J., Kavalieris L. and Mackisack M. (1986), Recursive Estimation of Linear Systems, Biometrika, 73,119-13. (1986)

Jordinson, R., Arnold, S., Hinde, B. Miller, G.F., Rodgers, .G. and Kitchin, A.M. (1976), Steady site transport function analysis of portions of the central circulation in man. Computers and

Biomedical Research. 9.192-305.

Kitagawa, G. (1981), A Nonstationary Time Series Model and its Fitting by a Recursive Filter, Journal of Time Series Analysis, 2,103-116.

Kitagawa G. and Gersch W. (1984), A Smoothness Priors-State Space Modeling of Time Series with Trend and Seasonality, Journal of the American Statistical Association, 79, 378-389.

Kitagawa G.and Gersch W. (1985a), A Smoothness Priors Long AR Model Method for Spectral Estimation, IEEE Transactions on Automatic Control, AC-30,57-65.

Kitagawa G.and Gersch W. (1985b), A Smoothness Priors Time Varying AR Coefficient Modeling of Nonstationary Time Series, IEEE Transactions on Automatic Control, AC-30,48-56.

Kuruc, A., Caldicott, W.J.H and Treves, S. (1982), An improved Deconvolution Technique for the Calculation of Renal Functions, Computers and Biomedical Research, 15,46-56.

Lindley D.V. and Smith A.F.M. (1972), Bayes Estimate for the Linear Model, Journal of the Royal Statistical Society, B,34,1-41.

Ljung L. (1985), On the Estimation of Transfer Functions, Automatica, 21, 677-696.

Mayne D.Q. and Firzoon F. (1978), Linear Estimation of ARMA Systems, Proc. IFAC, 7th World Congress, Helsinki.

Newbold, P. (1973), Bayesian Extension of Box-Jenkins Transfer Function Noise Models, Journal of the Royal Statistical Society, B, 35,323-336.

O'Sullivan F.,(1987), A Statistical Perspective on Ill-Posed Problems, Statistical Science, 1,502-527.

Panuska, V. (1968), An Adaptive Recursive Least Squares Identification Algorithm. Proceedings

Joint Automatic Control Conference.

Shiller R. (1973), A Distributed Lag Estimator Derived from Smoothness Priors, Econometrica, 41,775-778.

Tikhonov A. N. (1965), Incorrect Problems of Linear Algebra and a Stable Method for their Solution. Soviet Math. Dokl. 6, 988-991.

Wahba G. (1977) A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for Solving Them, in: Applications of Statistics, ed. P.R. Krishnaih, North Holland, 507-524.

Wahba G. (1982),Constrained Regularization for Ill Posed Linear Operator Equations with Applications in Meteorology and Medicine, Statistical Theory and Related Topics III,2,eds S.S. Gupta and J.O. Berger, New York, Academic Press, 383-418.

Wahba G. (1983) Bayesian confidence intervals for the cross-validated smoothing spline, Journal of the Royal Statistical Society B,45,133-150.

Wegman, E.J. (1980), Optimal Estimation of Time Series Functions, IEEE Transactions Acoust., Speech, Signal Process., ASSP 28,763-767.

Whittaker E.T. (1923), On a New Method of Graduation, Proceedings of the Edinborough Mathematical Association, 78,81-89.

Whittaker E.T. and Robinson G. (1924), Calculus of Observations, A Treasure on Numerical Calculations, Blackie and Son, Lmtd., London. pp 303-306

Zellner, A. (1971), An Introduction to Bayesian Influence in Econometrics, Wiley, New York.

LEGENDS

FIGURE 1. Box-Jenkins Series J Gas Furnace Data

Figure 2. Original and Modeled Data Superimposed and Difference Data Smoothness Priors Model, $n=296$

FIGURE 3. Impulse Response, Amplitude and Phase Response and Residual PSD versus Frequency, Superimposed Box-Jenkins and Smoothness Priors Model Results, (Dotted Lines).

FIGURE 4. Superimposed Smoothness Priors Impulse Response, Amplitude Response, Phase Response and Noise Spectrum for

A: $n=296$ and $n=200$ data points, (dotted lines). B: $n=296$ and $n=75$ data points, (dotted lines).

FIGURE 5. Superimposed Impulse Response, Amplitude Response, Phase Response and Noise Spectrum:

A: SP $M=4$ and $M=10$ models. B: SP $M=4$ and $M=20$ models.

C: SP $M=4$ and $M=30$ models. D: SP $M=4$ and LS $M=5$ models.

E: SP $M=4$ and LS $M=10$ models. F: SP $M=4$ and LS $M=20$ models.

FIGURE 6. Superimposed Smoothness Priors Impulse Response, Amplitude Response, Phase Response and Noise Spectrum for SP $M=4$ $k_1=1, k_2=1$ and $k_1=9, k_2=9$ (dotted lines) models.

FIGURE 7. Mean and Plus and Minus One Standard Deviation Impulse Response, Amplitude Response, Phase Response. AR Observation Noise:

A: Smoothness Priors Model. B: 3SLS Model.

FIGURE 8. Input, Output and Ouput Plus Noise:

A: $N=296$, Input, Noise, Superimposed output and output plus noise, Superimposed output and output plus larger variance noise, (in descending order).

B: $N=78$, Superimposed output and output plus noise, Noise.

FIGURE 9. Mean and Plus and Minus One Standard Deviation Impulse Response, Amplitude Response and Phase Response, MA Observation Noise, N = 296:

A: 3SLS M=4 Model. B: SP M=10, k_1=2,k_2=4 Model.

C: SP M=20, k_1=1,k_2=3 Model. D: SP M=20, k_1=3,k_2=5 Model.

E: SP M=20, k_1=2,k_2=4 Model. F: SP M=20, k_1=2,k_2=4 Model, Variance x 9.

FIGURE 10. Mean and Plus and Minus One Standard Deviation Impulse Response, Amplitude Response and Phase Response, MA Observation Noise, N = 78:

A: SP M=5 k_1=2,k_2=1 Model. B: SP M=10, k_1=2,k_2=1 Model.

FIG 1

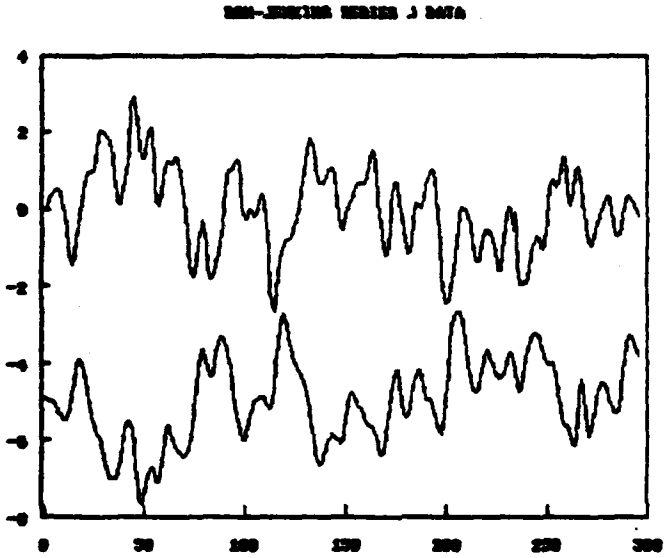


FIG 2

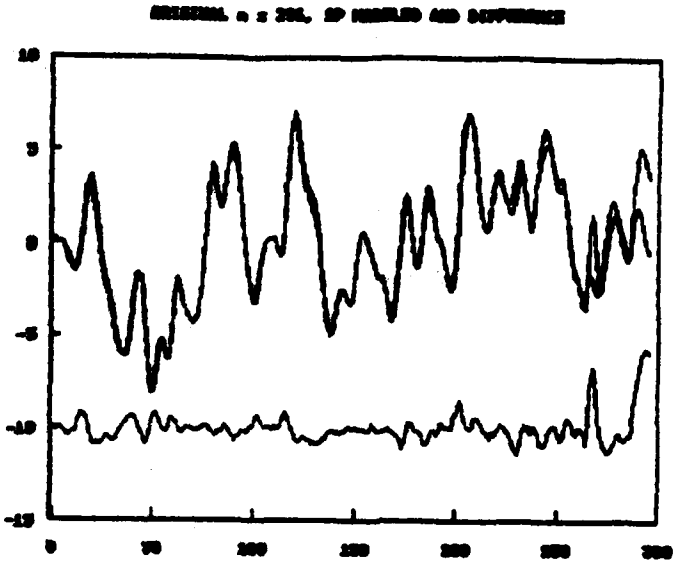
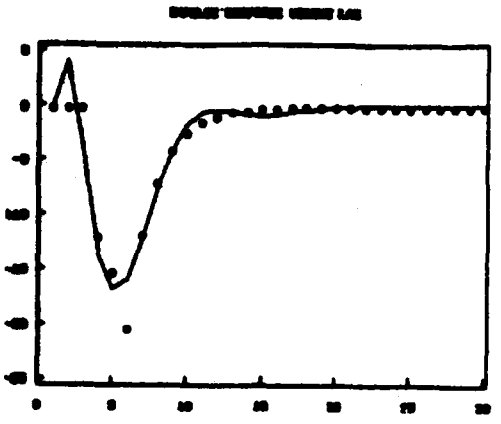
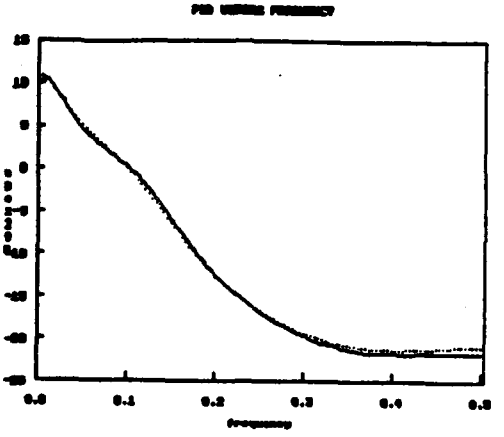
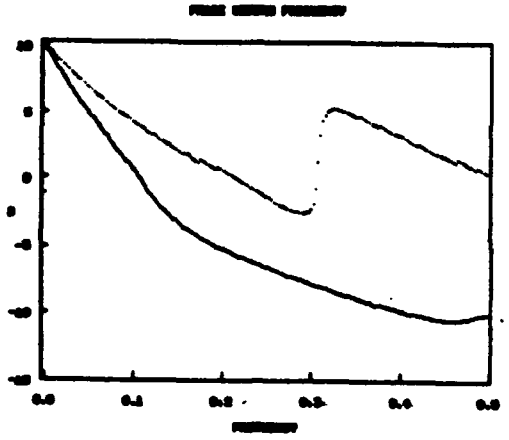


FIG 3







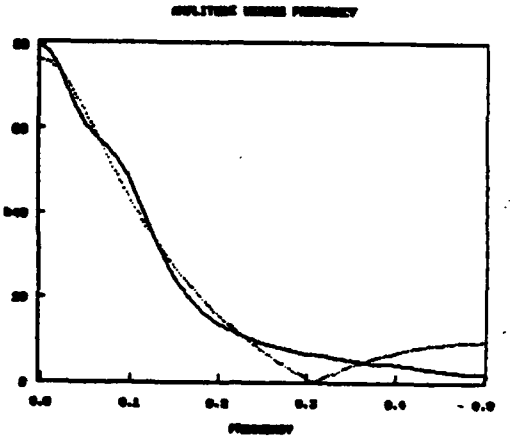
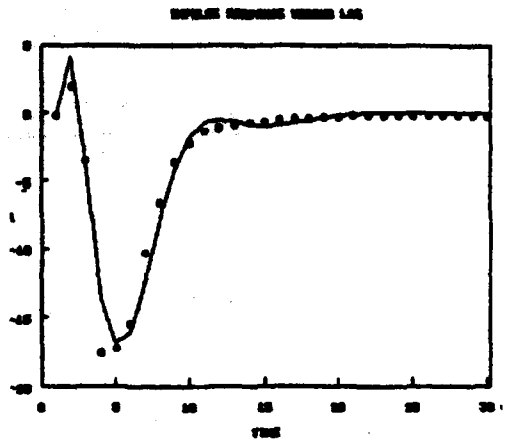
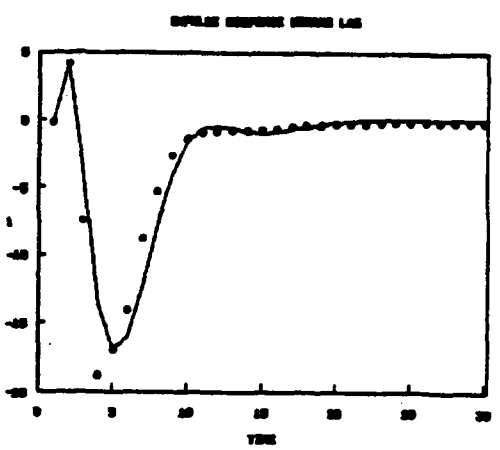
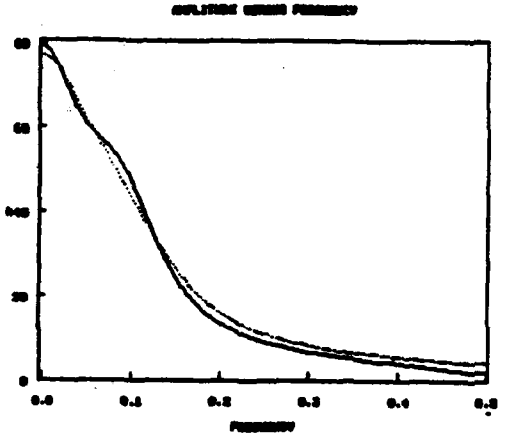
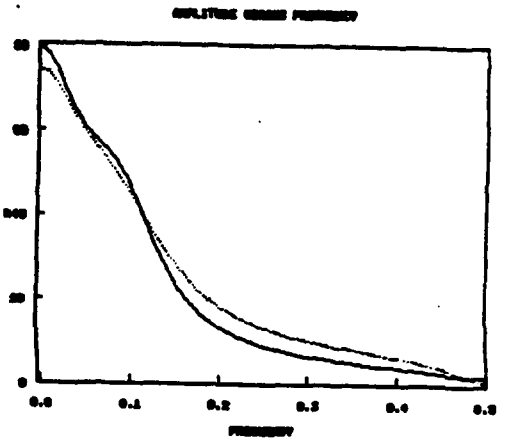


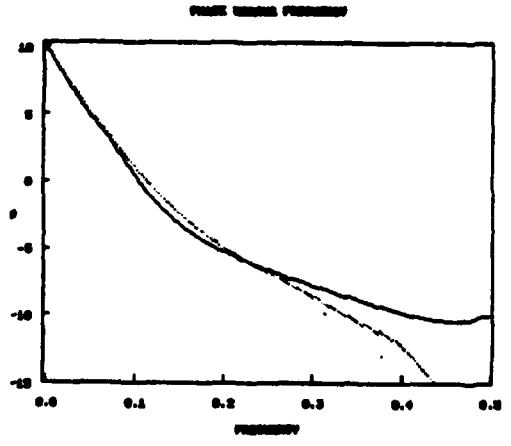
FIG 4A,B











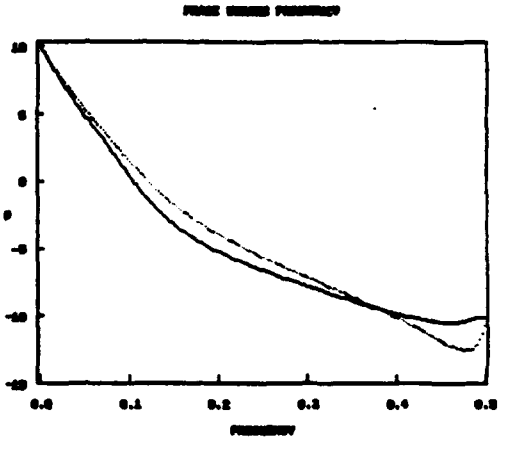
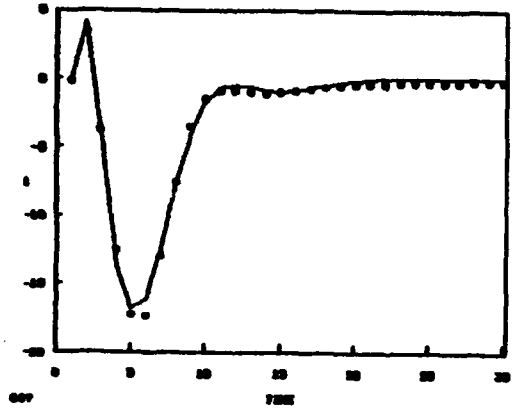
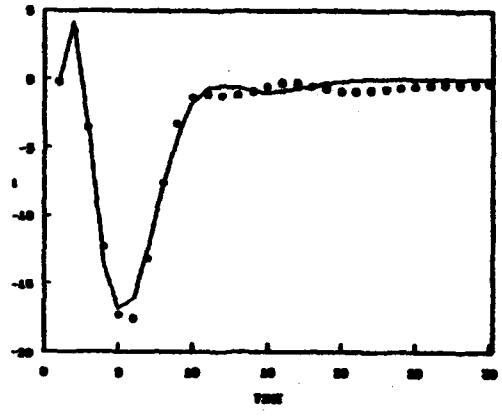
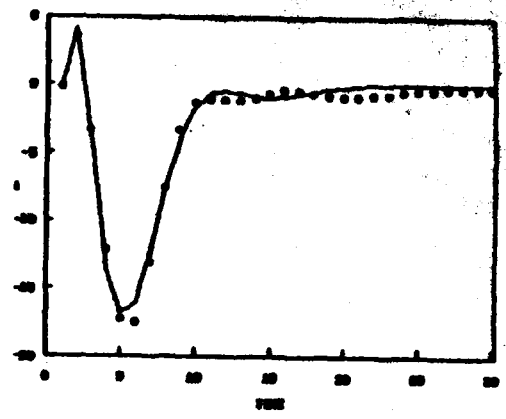
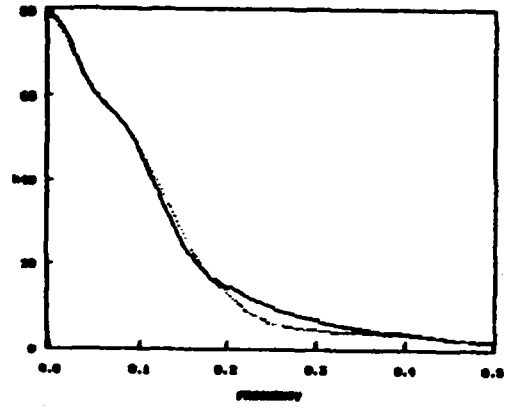


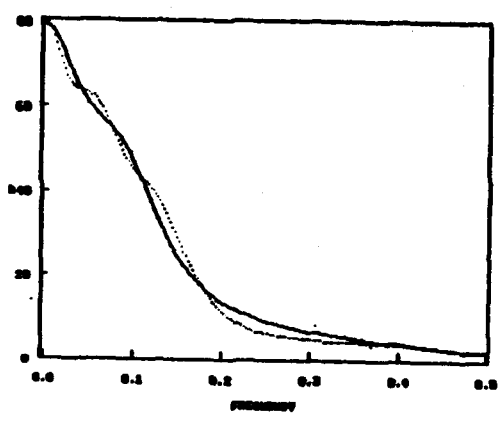
FIG 5A B C

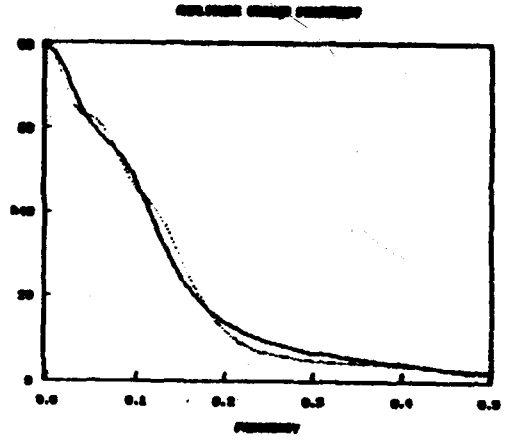


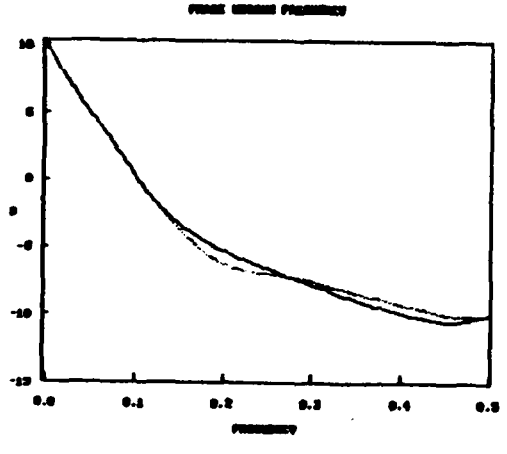


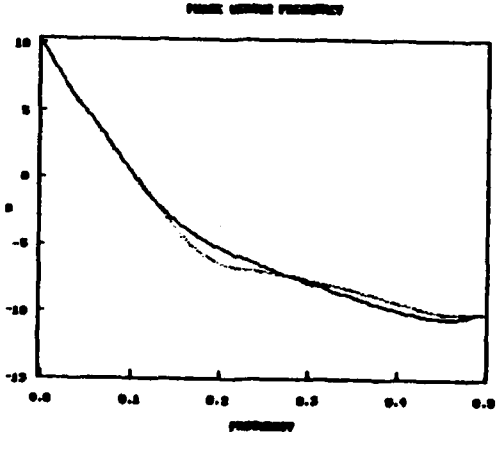












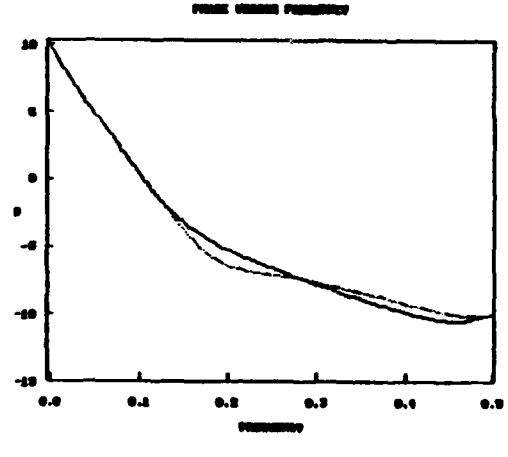
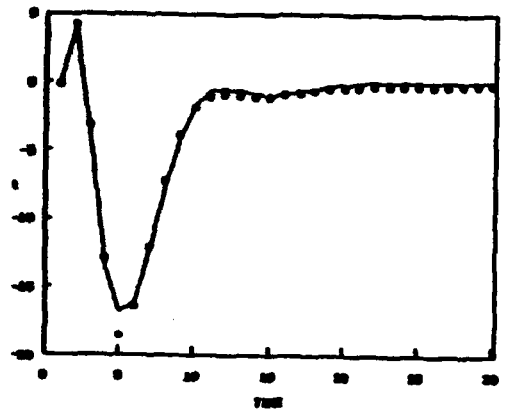
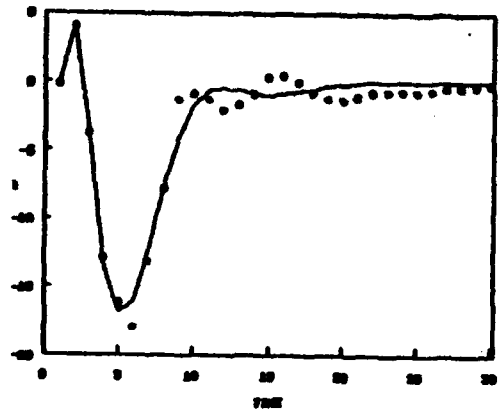
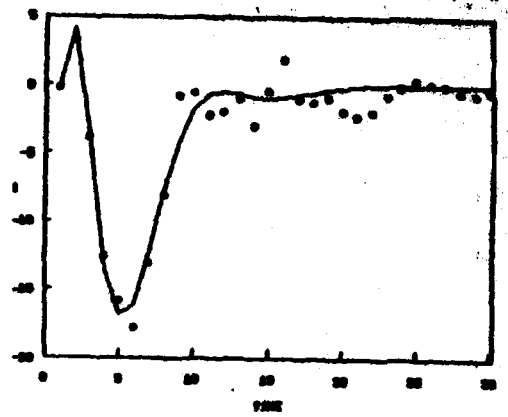
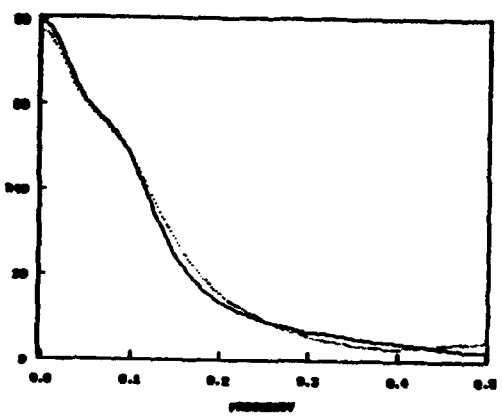


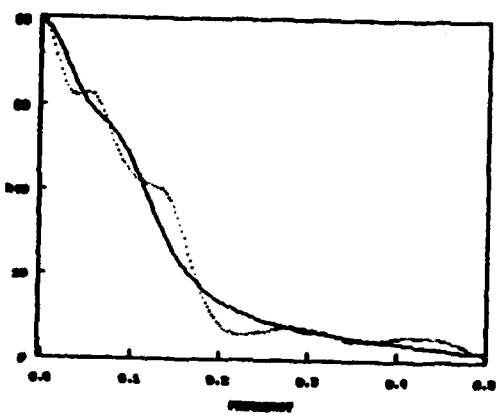
FIG 5(E,F

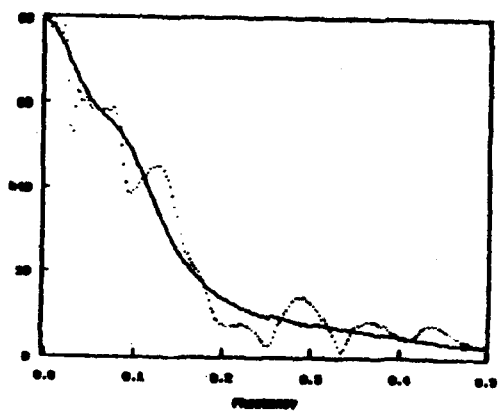


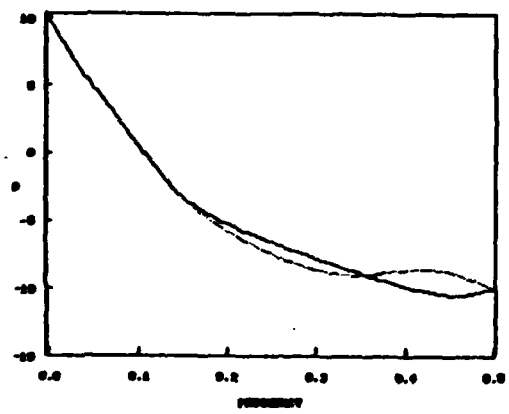


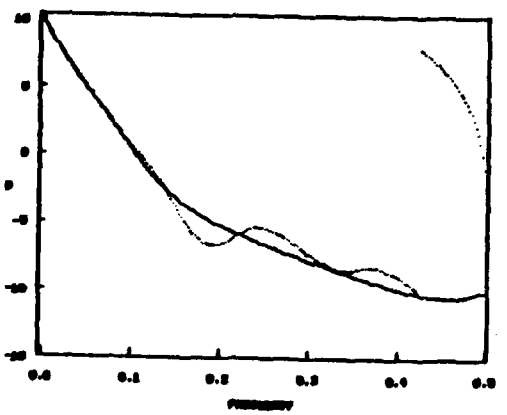












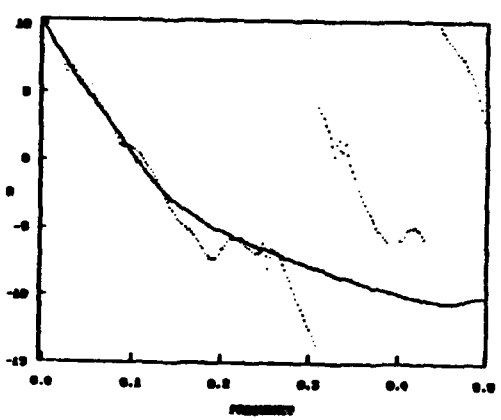
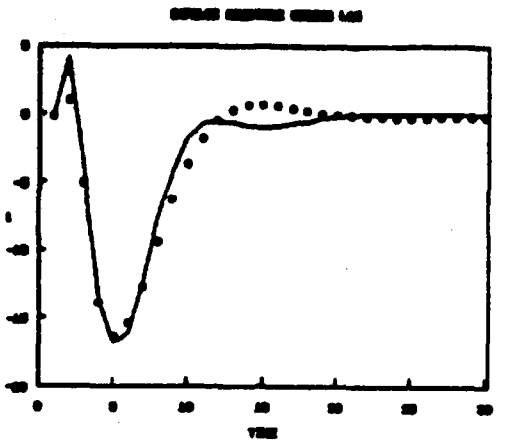
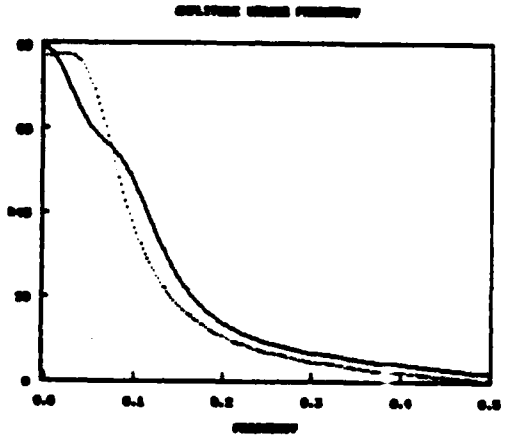


FIG 6





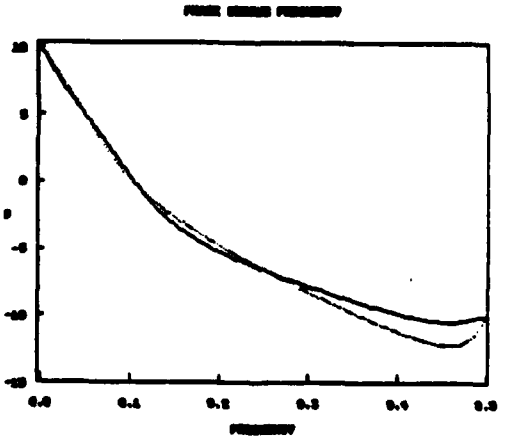
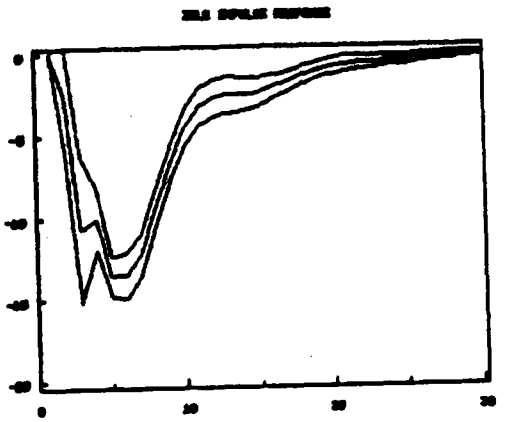
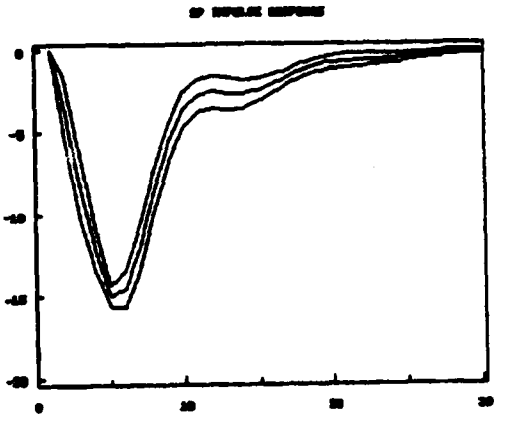
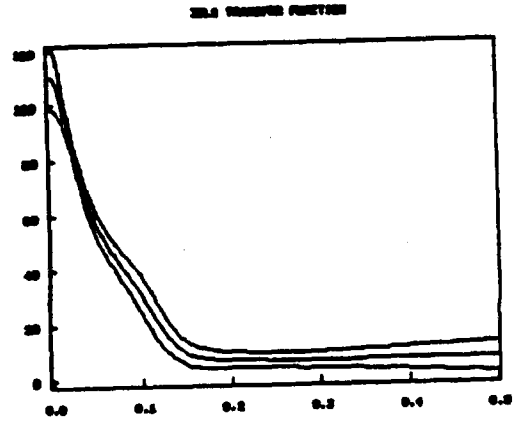
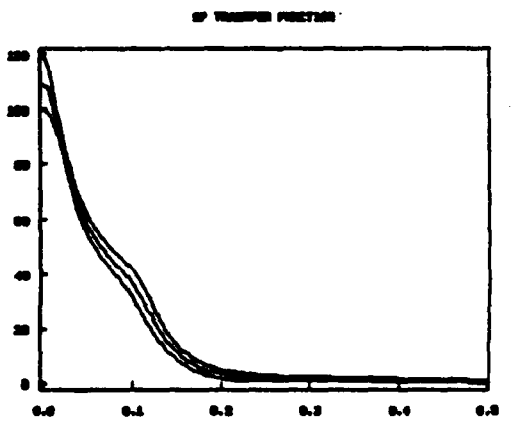


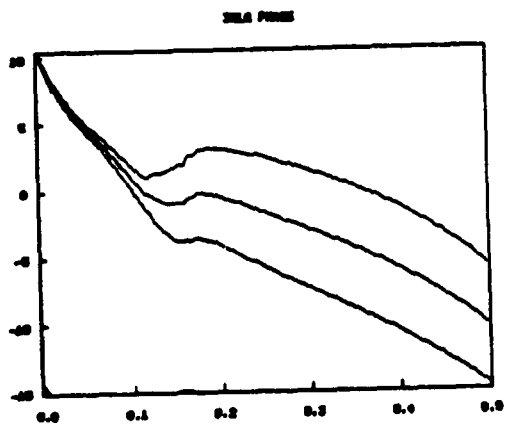
FIG 7











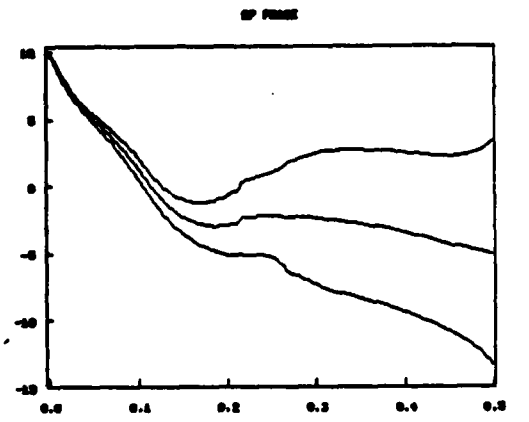
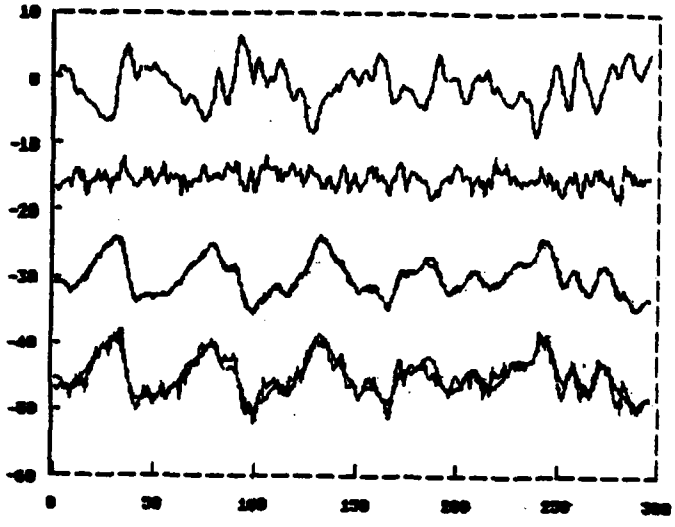


Fig 8A,B



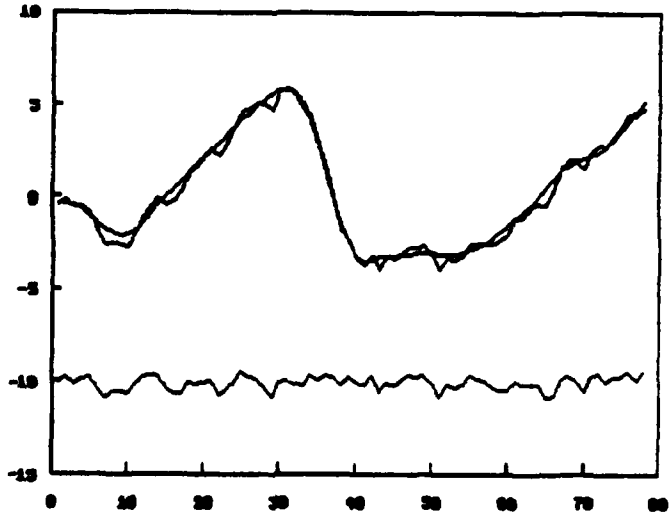
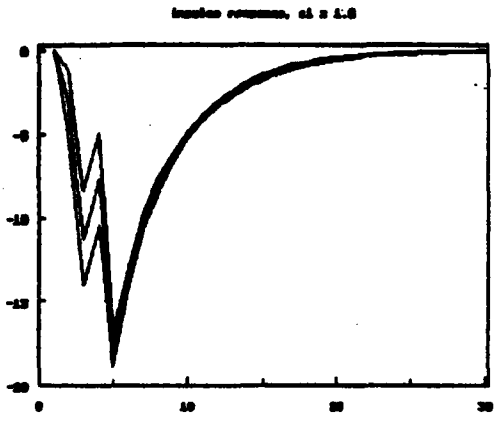
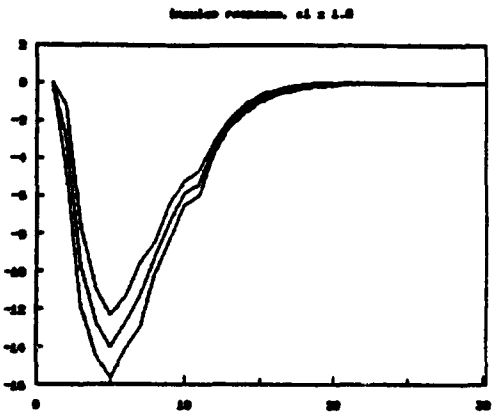
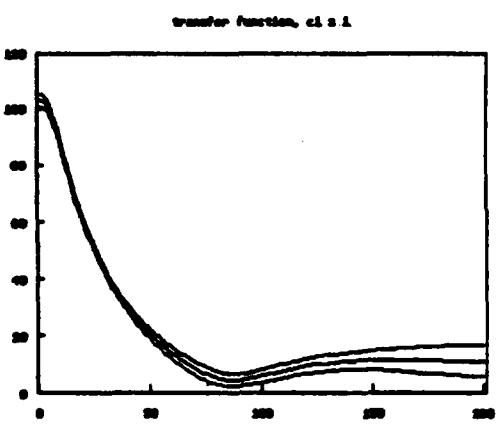
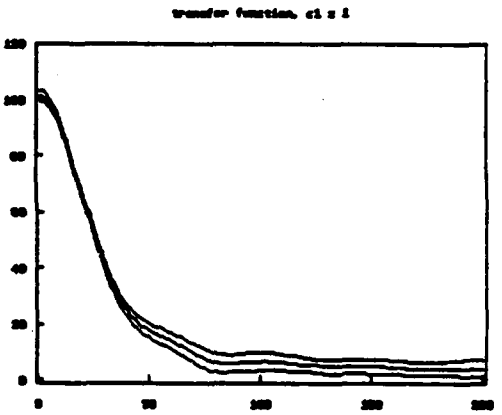


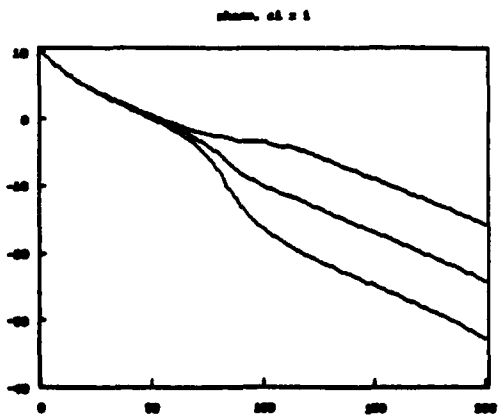
FIG 9A,B











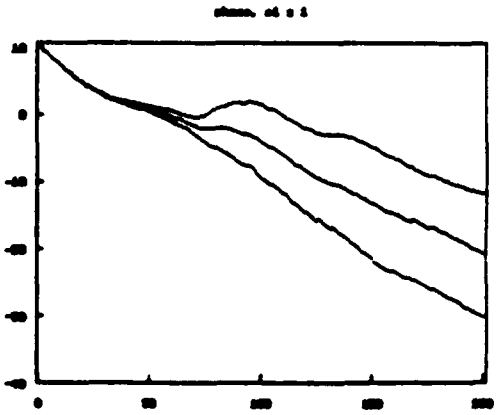
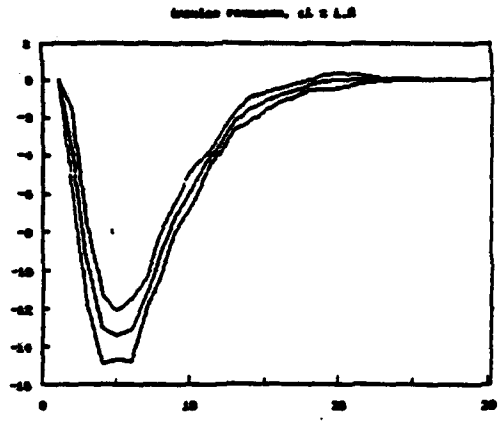
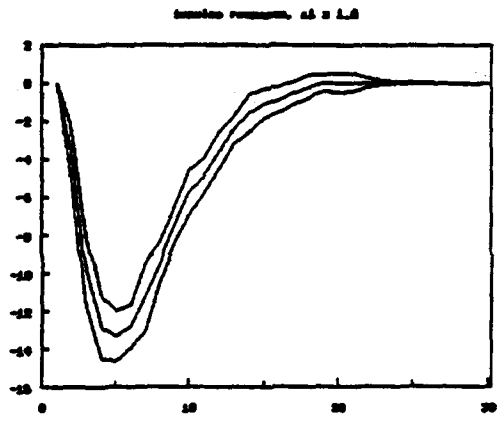
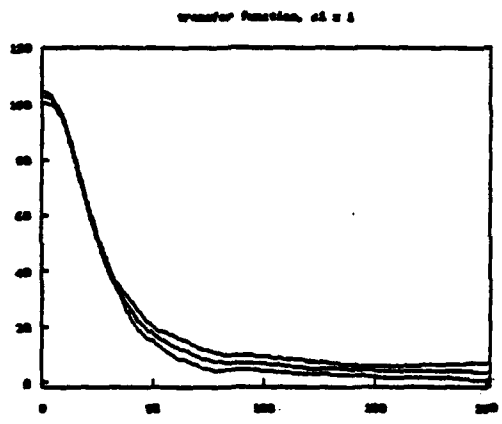
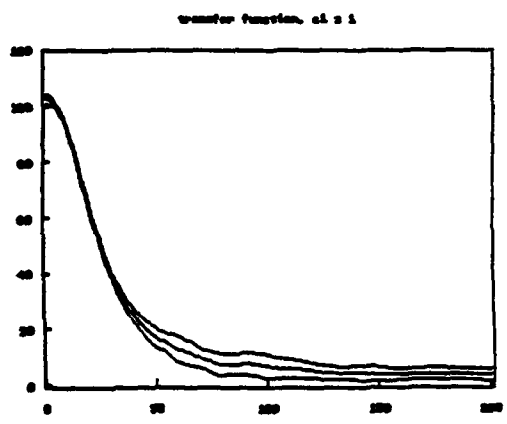


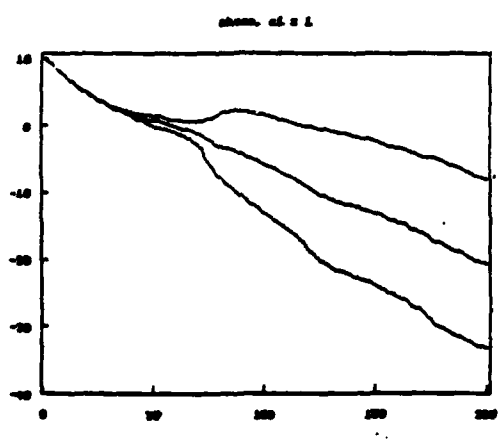
FIG 9C,D











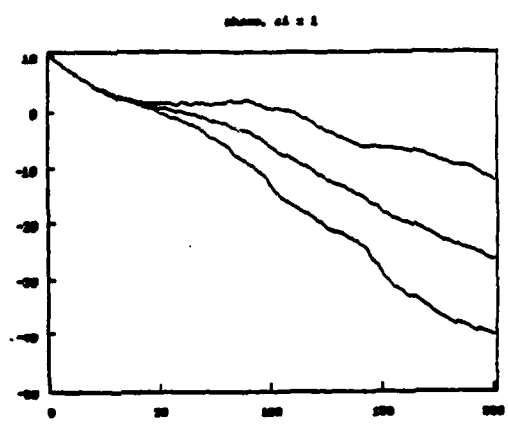
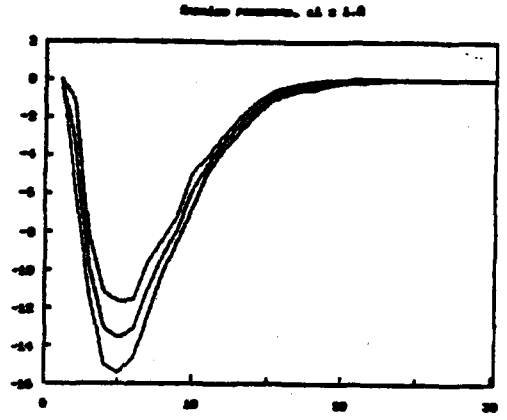
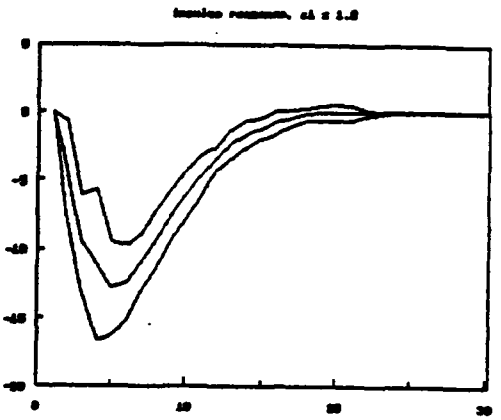
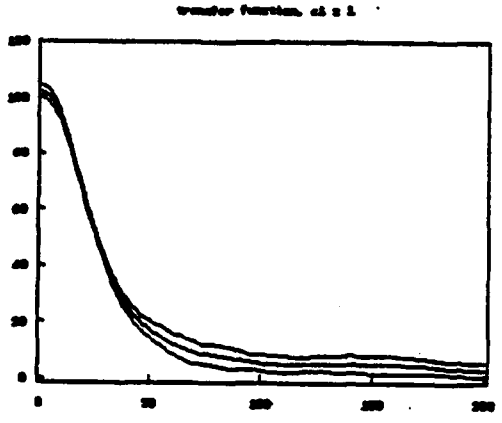
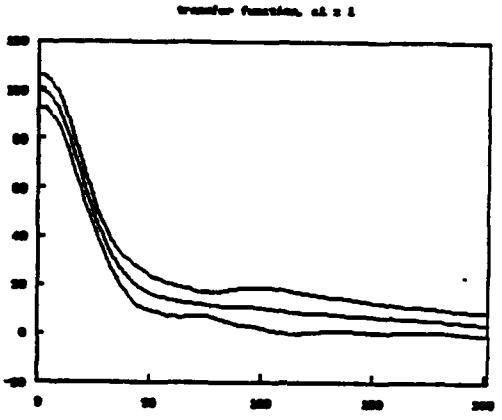


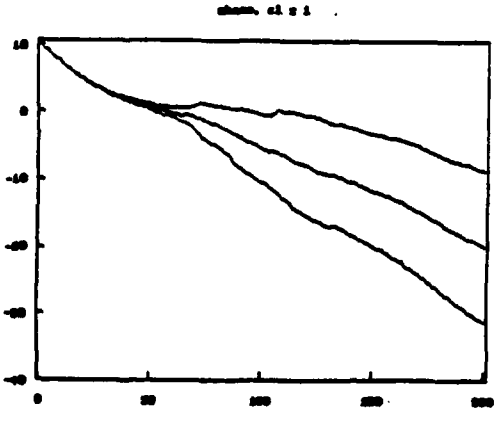
FIG 9E,F











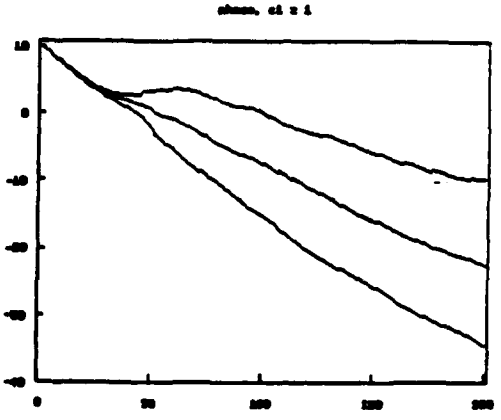
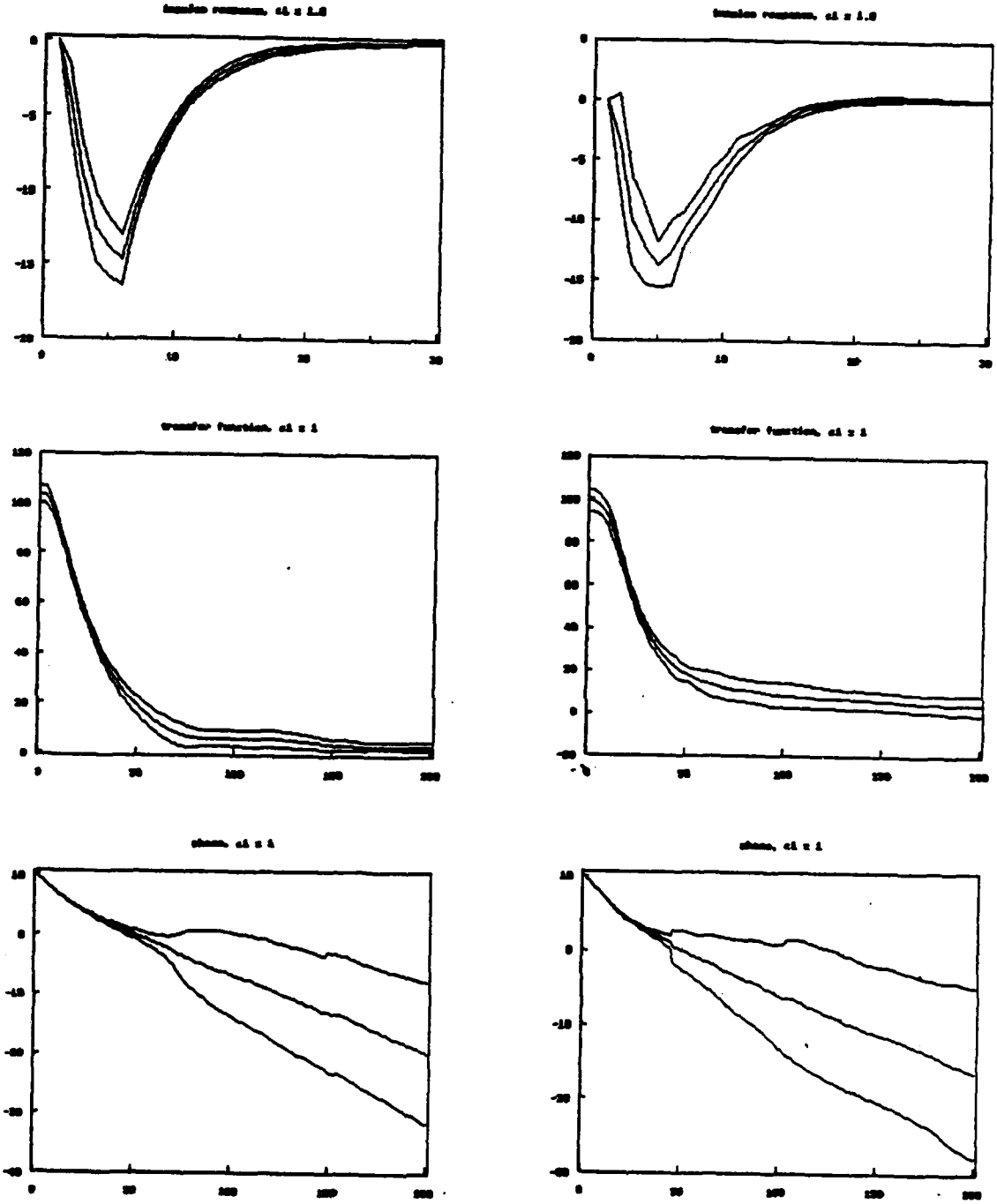


FIG 10A,B



DISTRIBUTION LIST

	NO. OF COPIES
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Library (Code 0142) Naval Postgraduate School Monterey, CA 93943-5000	2
Director of Research Administration (Code 012) Naval Postgraduate School Monterey, CA 93943-5000	1
Library (Code 55) Naval Postgraduate School Monterey, CA 93943-5000	1
Center for Naval Analyses 2000 Beauregard Street Alexandria, VA 22311	1
Operations Research Center, Room E40-164 Massachusetts Institute of Technology Attn: R.C. Larson and J.F. Shapiro Cambridge, MA 02139	1
Koh Peng Hong Ministry of Defense Blk A, Stockport Road Singapore 0511	1
Arthur P. Hurter, Jr. Professor and Chairman Northwestern University Dept of Industrial Engineering and Management Sciences Evanston, IL 60201-9990	1

END

7-87

DTIC