University of Boumerdes

Institute of Electricity and Electronics

DEPARTMENT OF RESEARCH

THESIS

Presented in partial fulfillment of the requirements of the

DEGREE OF MAGISTER

In Electronic System Engineering

By

Anissa ABBAS

Estimation of the Spectral Density of Random Signals with Missing Observations

Presented before the jury:

President: Pr. Kamel HARICHE (University of Boumerdes)

Supervisors: Dr. M. Djeddi (University of Boumerdes)

Pr. L. Refoufi (University of Boumerdes)

Members: Dr. Mohamed BENTARZI (USTHB)

Dr. BOUKHETALA (USTHB)

Dr. H. FELLAG (UTO)

ACKNOWLEDGEMENT

All my feeling of gratitude to my supervisor Dr M. Djeddi and all those who were of a great help for me.

I would like to thank all my teachers especially Dr Hariche and Dr Djeddi, and the library staff for their assistance.

Dedication

I dedicate this work to:

my parents

my sisters and brothers

and my supervisor Dr M.Djeddi

for their continuous help and engouragement

to finish this work.

Table of Content

Introduction

The spectral estimation is a preliminary data analysis tool which has been extensively used in the analysis of random signals. It consists of determining the spectral content of a random signal based on observations from that signal. Several parametric and non-parametric methods for estimating the spectral density have been suggested and investigated in the literature.

The first known methods of spectral estimation were based on the Fourier transform of either the set of data itself (periodogram) or the autocorrelation values (Blackman and Tukey estimator). In 1958, Blackman and Tukey published their well-known autocorrelative spectrum estimator. In 1965, and with the advent of the Fast Fourier Transform algorithm of Welch (1967) the periodogram estimator became more popular. However, despite the computational efficiency of the periodogram estimator, windowing was limiting its resolution.

In an attempt to alleviate the inherent limitations of the FFT approach which were particularly troublesome when analyzing short data records, new methods have been developed among them the Autroregressive spectral estimator of Parzen (1968) which is known also as the Maximum Entropy Method by Burg (1967). This estimator has been extensively investigated by many other authors among which Ulrych (1972), Wells and Chinnery (1972), Bernard (1967), and Anderson (1974). Later on, research was directed towards other time series models such as ARMA (Autoregressive Moving Average) models which are a generalization of the AR (Autoregressive) processes. It was thought that methods based on these models would provide higher performance and resolution than AR methods, Kay (1988).

Classical and parametric methods of spectral estimation consider that a full set of observations is available. In the case of realizations with missing observations, a case encountered often in practical situations, these methods can not be applied. Consider a realization X with missing values. This realization can be represented by a time series given as $\{X(j), j = 1,...N$, where M observations are missing. The estimation of the statistical characteristics of the underlying process X can not be studied using the standard methods since these latter suppose that we have a complete set of observations. Parzen (1963) proposed a method to bypass this problem by taking into consideration the problem of missing values. Dunsmir and Robinson (1981), and Ladjouze (1986) have considered the spectral estimation problem from realizations with missing values. More recently, Toloi and Morettin (1993) suggested a new method using the amplitude-modulation approach. They showed that the spectral density of the original signal $X(n)$ with missing values can be derived directly from the spectral density estimator of the modulated process $Y(n) = A(n) * X(n)$ where $A(n)$ is the modulating sequence chosen such that:

$$A(n) = \begin{cases} 1 \text{ if the process x(n) is observed at time n} \\ 0 \text{ if the process x(n) is not observed at time n} \end{cases}$$

The use of deterministic modulated signals enables us to consider the original incomplete sample as a realization with no missing values of the process $Y(t)$.

In this thesis, we propose two alternative methods for estimating the spectral density of realizations with missing values. The first method consists of using the Kalman smoother while the second one is based on the non-parametric Kernel method of curve fitting. The two proposed methods were investigated to overcome the drawback of the spectral

density estimator of Toloi and Morettin (1993) where the spectral density of the original process $X(t)$ is estimated by scaling that of the modulated process $Y(t)$. The results of the two different methods will be compared with the results of the Amplitude-modulated based approach.

In chapter 1, we introduce the concept of spectral estimation and a preview study on the methods used for spectral density estimation using complete data sets. In the second chapter, we define the concept of Amplitude modulation, and describe how it can be adapted to overcome the problem of missing observations in the spectral density estimation of the original process. Within the same chapter we will introduce the first method, proposed in this thesis, which is based essentially on the estimation of the missing values using the Kalman smoother. Because time-invariant algorithms suffer from transient effects at the beginning and the end of data records, we use a time-invariant Kalman smoothing estimator modified by a data-dependent correction term involving an initial state vector [Vaccaro and Fu Li (1990)]. This method is a parametric one i.e: its application requires the knowledge of the model of the observed time series. A second method of spectral density estimation will be discussed in the third chapter. This method is non-parametric and it consists of estimating the missing values by using the Kernel Curve Fitting approach. The Kernel smoothing method has been considered as a useful tool for identification and prediction in time series models, Kim and Cox (1996). It requires the choice of the smoothing parameter which controls the smoothness of the underlying process. In this thesis, this parameter is identified using the Cross-Validation rule as proposed by kim and Cox (1996). In the last chapter, we present the results of a simulation study which is devised for comparing the performances of the three methods. Our investigation indicates that out two stage procedures are more appropriate than the one step non-parametric method of Toloi and Morettin (1993). Furthermore, the kernel based method is more

suitable than the kalman smoothing approach since it does not require the knowledge of the parametric model of the process.

Chapter 1

Spectral density estimation of random signals

1.1- Introduction:

The general problem of spectral estimation is that of determining the spectral content of a random process based on a finite set of observations from that process. One way to estimate or describe the distribution of power with respect to frequency of a random process is to use the spectral density function. The estimation of this function known also as the spectrum of the process is usually based on procedures employing the Fast Fourier Transform function [Lacoss (1971), and Gutowski, Robinson and Treitel (1978)]. This approach to spectral analysis is computationally efficient and produces reasonable results for a large class of signal processes. In spite of these advantages, there are several inherent limitations to the FFT based methods. Its most prominent limitation is that of frequency resolution which means the non-ability to distinguish between the spectral responses of two or more signals. A second limitation is due to implicit windowing of data that occurs when using the FFT methods. These limitations have pushed forward to develop new parametric or model-based methods for spectral analysis and estimation. They are known as modern spectral analysis methods [Bingham (1967), Burg (1967,1968)]. These methods consist mainly of estimating the power spectral density using the estimated parameters of the assumed model. The existing spectral estimation methods rely on the availability of a full reliable observed data set independently of whether the signal is discrete or continuous.

1.2- Continuous and Discrete time signals:

Signals can arise in two different formats: continuous and discrete. A continuous time signal is any real or complex waveform $f(t)$ defined as a function of the continuous real temporal variable t. Similarly, a discrete time signal will be defined as a waveform $f(n)$ defined as a function of the discrete real temporal variable n.

Based on recent advances in Digital Circuit Technology, the developments of many computer-based algorithms for signal analysis have lead to the extensive use of discrete signals rather than continuous signals. In practice, continuous signals of interest are converted to discrete time signals using sampling techniques which allow us to implement solutions using the digital approach. Intensive development of the theory of sampling began in the late 1940's. The research efforts centered on the design and analysis of sampled data control systems by extending the classical techniques for designing analog control systems through the use of the Z transform.

A continuous time signal $f(t)$ can be characterized by its values $f(t_n)$ at a sequence of equidistant times $t_n = nT$ where T is the sampling period. The associated discrete signal $f(nT)$ can be regarded as the product of the continuous time signal $f(t)$ with a sampling signal $S(t)$, which is a periodic impulse train of the form:

$$S(t) = \sum_{k=-\infty}^{+\infty} \delta(t - kT)$$

where $\delta(t)$ is the Dirac function defined as:

$$\delta(t - t_0) = \begin{cases} 1 & \text{if} \quad t = t_0 \\ 0 & \text{elsewhere} \end{cases}$$

This is illustrated in Figure 1.1 given below:

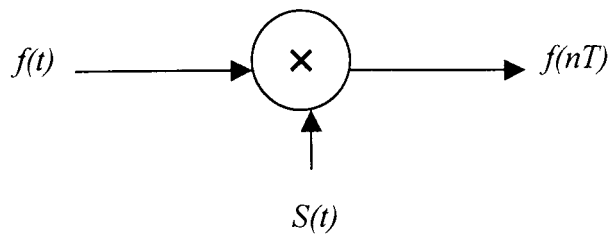

Figure 1.1: **Sampling a continuous time function**

The sampled signal can be expressed as:

$$f(nT) = \sum_{n=-\infty}^{+\infty} f(t)\delta(t - nT) \quad for\ n = 0,1,..... \tag{1.1}$$

The use of a sampled version of a continuous signal is very advantageous since we can make use of the existing computer-based algorithms as far as it is possible to reconstitute the original continuous signal from its sampled version whenever needed. In general, interpolating a set of sampled values may lead to different continuous signals i.e.: a continuous time signal can not be uniquely recovered from its samples. Lower sampling rates result in Aliasing. This causes the power occurring in the continuous signal at frequencies above the Nyquist frequency $f = 1/2T$ to appear at lower frequencies in the Discrete Fourier Transform. This effect 'Aliasing' can be avoided by applying Shannon theorm which gives a necessary and sufficient condition for recovering the original signal from its sampled signal. It states that for a bandlimited continuous signal $f(t)$ such that $F(w) = 0$ for $|w| > \sigma$, the sampling signal frequency should be at least twice σ where $F(w)$ is the Fourier Transform of the continuous signal $f(t)$.

Blackman and Tukey (1958) published classic articles describing how to estimate the power spectral density from a finite set of observations recorded at a set of uniformly spaced times $t_n = nT$ for $n = 0,...,N-1$ using the Discrete Fourier Transform. The parametric modern methods of spectral density estimation are based on Time series Modeling, to be discussed in the next section, due to the fact that a realization of a Time Series is defined as a sequential set of data measured over time.

1.3- Time series models:

A time series model aims to be a complete representation of the mechanism which may generate or account for an observed signal. It will express any observation $f(nT)$ as a function of the time, other variables, known as variables of interest, and one or more random components. In time series analysis several broad classes of models, expressed as difference equations, have been developed.

1.3.1- Autoregressive Moving Average with Exogenous Input (ARMAX) models:

The most popular linear time series model is the Autoregressive Moving Average with Exogenous Input (ARMAX). It is appropriate for signals which are generated by the following difference equation:

$$y(n) + a_1 y(n-1) + ... + a_{n_A} y(n-n_A) = b_1 u(n-1) + b_2 u(n-2) + ...b_{n_B} u(n-n_B) + ...$$
$$..... + \varepsilon(n) + c_1 \varepsilon(n-1) + ...c_{n_C} \varepsilon(n-n_C) \quad (1.2)$$

where the variables $\varepsilon(k)$ describe the noise which is usually taken to be a gaussian noise with zero mean and constant variance i.e.: $E(\varepsilon_k) = 0$

and $E\left(\varepsilon(k)^2\right)=\sigma^2$ for every k, and $u(k)$ is the values of an extra variable of interest or exogenous input.

This model is widely used in system identification, in control theory and in economic applications. It includes several interesting special cases such as:

- **Autoregressive model (AR(p)):** A discrete time signal $y(n)$ is generated by an autoregressive process of order p, if it can be described by the relation:

$$y(n)+a_1 y(n1)+...+a_p y(n-p)=\varepsilon(n) \tag{1.3}$$

This model corresponds to the ARMAX model when $b_i=0$ for $i=1,...n_B$ and $c_j=0$ for $j=1,...n_C$. The autoregressive model $AR(p)$ is found to be effective in modeling harmonics confounded with noise.

- **Moving average model (MA(q)):** A Moving Average process of order q is described by the model:

$$y(n)=\varepsilon(n)+c_1\varepsilon(n-1)+c_2\varepsilon(n-2)+.....c_q\varepsilon(n-q) \tag{1.4}$$

This model is popular in signal processing as a basis for filter design (FIR) and the identification of truncated impulse responses.

- **Autoregressive Moving Average Model (ARMA(p,q)):** The relation describing the autoregressive moving average model of order (p,q) is given by the equation:

$$y(n)+a_1 y(n-1)+...a_p y(n-p)=\varepsilon(n)+c_1\varepsilon(n-1)+....+c_q\varepsilon(n-q) \tag{1.5}$$

The autoregressive moving average model is effective in modeling the disturbance spectra with spectral peaks as well as zeros and it is therefore appropriate in model-based spectrum analysis.

Instead of dealing with difference equations given in [(1.2)- (1.5)] which are used often in the time domain analysis, an alternative method would be the use of transfer functions. Applying the Z-transform to the difference equation representing the ARMA model gives its transfer function which is defined by:

$$H(z)=\frac{C_q(z)}{A_p(z)} \quad , z \in C \tag{1.6}$$

where the p^{th} order polynomial $A_p(z)$ and q^{th} order polynomial $C_q(z)$ are given by:

$$A_p(z)=1+a_1 z+.....+a_p z^p \tag{1.7}$$

$$C_q(z)=1+c_1 z+.....+c_q z^q \tag{1.8}$$

A third method, the state space form, can be used to describe time series. We consider that the Nx1 vector of observed variables up to time t, y_t, is related to an mx1 vector, α_t, known as the state vector, via a measurement equation:

$$y_t = Z_t\alpha_t + d_t + \varepsilon_t \qquad t=1,.....T \tag{1.9}$$

where Z_t is an Nxm matrix, d_t is an Nx1 vector and ε_t is an Nx1 vector of serially uncorrelated disturbances with mean zero and covariance matrix H_t, that is:

$$E(\varepsilon_t) = 0 \quad \text{and} \quad \text{Var}(\varepsilon_t) = H \tag{1.10}$$

In general, starting from a difference equation, one can derive automatically the state space representation of the system using different methods such as Direct programming, or Nested programming. Consider the difference equation describing a system of order n:

$$a_n y(n-k) + a_{n-1} y(n-k+1) + \ldots a_1 y(k-1) + a_0 y(k) = b_0 u(k) + b_1 u(k-1) + \ldots + b_n u(k-n) \tag{1.11}$$

We define the backward shift operator E^{-i} given as:

$$E^{-i} y(k) = y(k-i)$$

Then equation (1.11) can be rewritten as:

$$\hat{A} y(k) = \hat{B} u(k) \tag{1.12}$$

where:

$$\hat{A} = \sum_{i=0}^{n} a_i E^{-i}$$

$$\hat{B} = \sum_{i=0}^{n} b_i E^{-i}$$

We define the equation:

$$\hat{A} x(k) = u(k) \tag{1.13.a}$$

$$\hat{B} x(k) = y(k) \tag{1.13b}$$

We define n state variables given by:

$$\begin{cases} x(k-n) = x_1(k) \\ x(k-n+1) = x_1(k+1) = x_2(k) \\ x(k-n+2) = x_2(k+1) = x_3(k) \\ . \\ . \\ x(k-1) = x_{n-1}(k+1) = x_n(k) \\ x(k) = x_n(k+n) \end{cases}$$

By using equation (1.13a) and (1.13b), we will get:

$$x(k) = \frac{1}{a_0}\left[u(k) - a_1 x_n(k) - a_2 x_{n-1}(k) - ..... - a_n x_1(k)\right] \qquad (1.14a)$$

$$y(k) = b_0 x_n(k+1) + b_1 x_n(k) + ... + b_n x_1(k) \qquad (1.14b)$$

$$y(k) = \frac{b_0}{a_0}u(k) + \left(b_1 - \frac{b_0 a_1}{a_0}\right)x_n(k) + \left(b_2 - \frac{b_0 a_2}{a_0}\right)x_{n-1}(k) + ... + \left(b_n - \frac{b_0 a_n}{a_0}\right)x_1(k) \qquad (1.14c)$$

In matrix form, we will have:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ . \\ . \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ . & . & . & ... & . \\ 0 & 0 & 0 & . & 1 \\ \frac{-a_n}{a_0} & \frac{-a_{n-1}}{a_0} & \frac{-a_{n-2}}{a_0} & . & \frac{-a_1}{a_0} \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ . \\ . \\ x_n(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ \frac{1}{a_0} \end{bmatrix} u(k) \qquad (1.15)$$

$$y(k) = \begin{bmatrix} b_n - \frac{a_n b_0}{a_0} & b_{n-1} - \frac{a_{n-1} b_0}{a_0} & ..... & ... & b_1 - \frac{a_1 b_0}{a_0} \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \\ \\ x_n(k) \end{bmatrix} + \frac{b_0}{a_0}u(k) \qquad (1.16)$$

As an example consider the Autoregressive process of order 2. This latter can be represented by the difference equation:

$$y(n) = \phi_1 y(n-1) + \phi_2 y(n-2) + \varepsilon_n$$

For this, we define the states:

$$\begin{cases} x(k-2) = x_1(k) \\ x(k-1) = x_1(k+1) = x_2(k) \\ x(k) = x_2(k+1) \end{cases}$$

Then, following the same development as in equations (1.14c):

$$y(k) = \varepsilon(k) + (\phi_1)x_2(k) + (\phi_2)x_1(k)$$

In matrix form, it will be represented by:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\phi_2 & -\phi_1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \varepsilon(k)$$

$$y(k) = \begin{bmatrix} \phi_2 & \phi_1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \varepsilon(k)$$

The state space representation is very useful for analyzing multivariable processes and when one is interested in identifying the sate variables $x_1(t),.....,x_p(t)$ which induce the value of the process $y(t)$.

1.4- Frequency versus time domain representation of signals:

Although a signal is a function of time, it is often possible and profitable to speak of frequency domain representation meaning that a signal consists of certain well-defined frequency components. In the definitions of the previous section, we have considered the time domain representation of signals, an alternative approach is to study the time series in the frequency domain i.e: in terms of repetitive cycles. The Fast Fourier Transform made the frequency domain approach to digital filtering competitive with the time-domain difference equation approach because it significantly reduced the computation time.

1.4.1- Power spectral density function of signals:

One way to estimate or describe the distribution of power with respect to frequency of a random process $x(n)$ is to use its power spectral density function, which is defined as the Fourier transform of the autocorellation function of the random process:

$$P_{xx}(f)=\sum_{k=-\infty}^{\infty} r_{xx}(k)\exp(-j2\pi fk) \quad -1/2 \le f \le 1/2 \tag{1.17}$$

where $r_{xx}(k)$ is the autocorrelation function of $x(n)$ defined by:

$$r_{xx}[k]=E(x(n)x(n+k)) \qquad k=0,\pm 1,..... \tag{1.18}$$

The spectral density function $P_{xx}(f)$ can be deduced automatically from the time series models discussed in the previous section. For the ARMA signals described by the model (1.5), the power spectral density can be obtained from its transfer function given by:

$$H(z)=\frac{C_q(z)}{A_p(z)} \tag{1.19}$$

The power spectrum of the output signal $y(n)$ is given as the product of the power spectrum of the input signal $\varepsilon(n)$ and the squared magnitude frequency response of the generating system that is:

$$P_{yy}(f)=P_{\varepsilon\varepsilon}(f)|H(f)|^2 \tag{1.20}$$

where $H(f)$ is the frequency response of the system defined by $H(f)=H(e^{jf})$ and $P_{\varepsilon\varepsilon}(f)$ is the input power spectrum $\varepsilon(n)$.

Assuming that the input $\varepsilon(n)$ is a white noise sequence with constant variance σ^2, we will get:

$$P_{yy}(f) = \sigma^2 |H(f)|^2 \tag{1.21}$$

Thus, the spectral density of the signal $y(n)$ generated by the $ARMA(p,q)$ model is given by:

$$\Phi(f) = \frac{|C_q(f)|^2}{|A_p(f)|^2} \sigma^2 \tag{1.22}$$

where $A_p(f)$, $C_q(f)$ are evaluated using (1.7), and (1.8) by replacing z by e^{jf}. The spectral density can easily be estimated parametrically by replacing the parameters $a_i's$, $c_i's$, and the variance σ^2 by their estimates.

1.5- Spectral estimation methods:

From equation (1.17), we see that the spectral density is a function of an infinite number of autocorrelation values, which means that the task of estimating the power density from a finite set of data is an impossible one. Alternative identification methods of spectral density function had been investigated.

1.5.1- Classical spectral estimation methods:

Classical methods use Fourier transforms on either windowed data or windowed autocorrelation estimates. The window is usually designed to give as much resolution as possible with little leakage. The two popular methods developed are:

1.5.1.1- The periodogram of a signal:

The periodiogram spectral estimator relies on the following equivalent definition of the spectral density function $P_{xx}(f)$ given by:

$$P_{xx}(f)=\lim_{M\to\infty} E\left[\frac{1}{2M+1}\left|\sum_{n=-M}^{M} x(n)\exp(-j2\pi fn)\right|^2\right] \tag{1.23}$$

If we assume that the process $x(n)$ is stationary and ergodic, the spectral density function can be estimated empirically by the periodogram which is defined by:

$$\hat{P}_{Per}(f)=\frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\exp(-j2\pi fn)\right|^2 \tag{1.24}$$

where $\{x(0),x(1),.....x(N-1)\}$ are the observed values of the signal at equidistant times $t_i=(i-1)T$ for $i=1,.....,N$.

Many of the drawbacks of the periodogram spectral estimator (1.24) are due to the assumptions made about the data outside the measurement interval $[0,(N-1)T]$. The finite data sequence $\{x(0),...,x(N-1)\}$ may be viewed as being obtained by windowing an infinite length sample sequence with a rectangular window. The use of only this finite set of data reduces to assume implicitly that the unmeasured data are zero. In other words, the available data set $\{x(0),x(1),.....x(N-1)\}$ can be regarded as a truncated discrete time signal $x^N(k)$ defined by:

$$x^N(k)=\begin{cases} x(k) & k=0,1,.....N\text{-}1 \\ 0 & otherwise \end{cases} \tag{1.25}$$

Using (1.17), the spectral density function of the truncated signal $x^N(k)$ is defined by:

$$P_{x^N x^N}(f) = \sum_{k=-\infty}^{\infty} r[k] \exp(-j2\pi fk) \tag{1.26}$$

where

$$r(k) = \frac{1}{N} \sum_{n=-\infty}^{+\infty} x^N[n] x^N[n+k] \tag{1.27}$$

Substituting $r(k)$ by its expression in (1.26), we obtain the spectral density of the truncated signal $x^N(k)$:

$$P_{x^N x^N}(f) = \sum_{k=-\infty}^{\infty} \left[\frac{1}{N} \sum_{n=-\infty}^{+\infty} x^N(n) x^N(n+k) \right] e^{-j2\pi fk} \tag{1.28}$$

Multiplying the right hand side of equation (1.28) by the term $e^{j2\pi n} e^{-j2\pi n}$ and interchanging the order of the two sums will result in:

$$P_{x^N x^N}(f) = \frac{1}{N} \sum_{n=-\infty}^{+\infty} x^N(n) e^{j2\pi fn} \sum_{k=-\infty}^{+\infty} x^N(n+k) e^{-j2\pi f(n+k)} \tag{1.29}$$

$$P_{x^N x^N}(f) = \frac{1}{N} X_N(2\pi f) X_N^*(2\pi f) = \frac{1}{N} |X_N(2\pi f)|^2 \tag{1.30}$$

where $X_N^*(2\pi f)$ is the complex conjugate of the Fourier Transform $X_N(2\pi f)$ of $x(n)$.

Therefore, the periodogram estimator (1.24) of the spectral density of the original signal $x(n)$ is equal to the spectral density function of the truncated signal $x^N(n)$.

We will investigate the mean and variance of the periodogram (1.24) using the mean and variance of the autocorrelation function as follows:

Using the fact that:

$$E\left(\hat{P}_{Per}(f)\right)=E\left(\frac{1}{N}\left|\sum_{n=0}^{N-1}x(n)\exp(-j2\pi fn)\right|^2\right) \tag{1.31}$$

we can show that:

$$E\left[\hat{P}_{Per}(f)\right]=\sum_{k=-(N-1)}^{N-1}E\left\{\hat{r}_{xx}(k)\right\}e^{-j2\pi fk} \tag{1.32}$$

where $\hat{r}_{xx}(k)$ is the sample autocorellation function defined by:

$$\hat{r}_{xx}(k)=\frac{1}{N}\sum_{n=0}^{N-1-|k|}x(n)x(n+k) \qquad k=0,\pm1,.....\pm N\text{-}1 \tag{1.33}$$

Taking into account that:

$$E\left\{\hat{r}_{xx}(k)\right\}=\frac{1}{N}\sum_{n=0}^{N-1-|k|}E\{x(n)x(n+k)\}=1-\frac{|k|}{N}r_{xx}(k) \tag{1.34}$$

$$=1-\frac{|k|}{N}r_{xx}(k)$$

we deduce that the expectation of the periodogram estimator $P_{Per}(f)$ is given by:

$$E\left[\hat{P}_{Per}(f)\right] = \sum_{k=-(N-1)}^{N-1} \frac{N-|k|}{N} r_{xx}(k) e^{-j2\pi fk} \tag{1.35}$$

As the number of observations increases, we obtain:

$$\lim_{N\to\infty} E\left\{\hat{P}_{Per}(f)\right\} = \sum_{k=-\infty}^{\infty} r_{xx}(k)\exp(-j2\pi fk) \tag{1.36}$$

That is $\lim_{N\to\infty} E\left\{\hat{P}_{Per}(f)\right\} = P_{xx}(f)$ which shows that the periodogram estimator is an asymptotically unbiased estimator of the unknown spectral density function of the signal $x(n)$.

Similarly, we can show that the variance of the periodogram can be written as [Vasegin (1996)]:

$$Var\left[\hat{P}_{Per}(f)\right] = P_{Per}^2(f)\left(1+\left(\frac{\sin 2\pi fN}{N\sin 2\pi f}\right)^2\right) \tag{1.37}$$

Hence, the variance of $\hat{P}_{Per}(f)$ converges to $P_{Per}^2(f)$. In other words, the periodogram estimator is not a consistent estimator of $P_{xx}(f)$ since its variance does not converge to zero. This estimator has been extensively used and many modifications have been proposed to improve its performance. The two most popular modifications are the Averaged periodogram, and the one which is based on data windowing.

Since the variance of the periodogram given in (1.37) is non-zero independently of the data set length N, one way to improve the statistical properties of the periodogram is to average a set of K different periodogram estimators of the same process. If K independent data records are available, the averaged periodogram estimator is defined as:

$$\hat{P}_{avr}(f)=\frac{1}{K}\sum_{m=0}^{K-1}\hat{P}_{Per}^{(m)}(f) \tag{1.38}$$

where

$\hat{P}_{Per}^{(m)}(f)$ is the periodogram of the m^{th} data set.

The mean of $\hat{P}_{avr}(f)$ is the same as that of $\hat{P}_{Per}(f)$ but its variance will be decreased by a factor of K and it is given by:

$$Var\left[\hat{P}_{Per}(f)\right]=\frac{1}{K}P_{Per}^{2}(f)\left(1+\left(\frac{\sin 2\pi fN}{N\sin 2\pi f}\right)^{2}\right) \tag{1.39}$$

This estimator is useful for the situations where the number of observations N is large enough so that one can split the series in K subseries of reasonable sizes.

When N is not large enough or when the signal of interest is a sinusoidal or narrowband signal corrupted by white noise, it is advantageous to filter the data $x(n)$ prior to computing its Fourier Transform. Data windowing (filtering) will prevent from masking a low-level signal by sidelobes of a higher level signal close in frequency. The periodogram is given by:

$$\hat{P}_{Per}(f)=\frac{1}{U}\frac{1}{N}\left|\sum_{n=0}^{N-1}x[n]w(n)\exp(-j2\pi fn)\right|^{2} \tag{1.40}$$

where U is the window energy defined by:

$$U=\frac{1}{N}\sum_{n=0}^{N-1}w^{2}(n)$$

and $w(n)$ is a real sequence termed a lag window [Johansson (1993) and Kay (1988)] satisfying the classical conditions:

i) $0 \leq w[n] \leq w[0] \leq 1$

ii) $w[-n] = w[n]$

iii) $w[n] = 0$ for $|n| > M$ where M is to be specified such that $M \leq N\text{-}1$.

1.5.1.2- Blackman-Tukey estimator:

The Blackman-Tukey power spectral density estimator is the Fourier transform of the windowed autocorrelation estimate.

$$\hat{P}_{BT}(f) = \sum_{k=-(N-1)}^{N-1} w(k) r_{xx}(k) \exp(-j2\pi fk) \tag{1.41}$$

where $w(k)$ is a lag window satisfying the conditions (i), (ii), and (iii) given above.

The estimator (1.41) is equivalent to the periodogram if $w(k) = 1$ for every k. The use of the window $w(k)$ reduces the mean square error of the power spectral estimate thus increasing its statistical stability. This method is also known as the moving average method. Various window functions have been proposed which are usually unrelated to the data or random process to be analyzed.

The expectation $\hat{P}_{BT}(f)$ is given by:

$$E\left[\hat{P}_{BT}(f)\right] = \sum_{k=-(N-1)}^{N-1} E[\hat{r}_{xx}(k)] w(k) e^{-j2\pi fk} \tag{1.42}$$

Using the fact that:

$$E\left\{\hat{r}_{xx}[k]\right\} = \frac{1}{N}\sum_{n=0}^{N-1-|k|} E\{x(n)x(n+k)\}$$
$$= 1 - \frac{|k|}{N} r_{xx}(k) \qquad (1.43)$$
$$= r_{xx}(k) w_B(k)$$

where $w_B(k)$ is the Barlett or triangular window.

Then, equation (1.42) can be written as:

$$E\left[\hat{P}_{BT}(f)\right] = \sum_{k=-(N-1)}^{N-1} r_{xx}(k) w_c(k) e^{-j2\pi fk} \qquad (1.44)$$

where:

$$w_c(k) = w_B(k) w(k) \qquad (1.45)$$

Using the duality property of the Fourier Transform, it can easily be seen that [Vasegin (1996)]:

$$\lim_{N\to\infty} E\left[\hat{P}_{BT}(f)\right] = \int P_{xx}(v) W_c(f-v) dv \qquad (1.46)$$

On the other hand, using similar arguments, we can show that:

$$Var\left[\hat{P}_{BT}(f)\right] = \frac{U}{N} P_{xx}^2(f) \qquad (1.47)$$

where $W_c(f)$, U are the Fourier transform and the energy of $w_c(k)$ respectively.

Therefore, we conclude that the use of the window $w(k)$ reduces the variance of the spectral estimator at the expense of increasing the bias

unless the process is white noise, for which the bias is zero for any lag window.

1.5.2- Modern spectral estimation methods:

The popular classical methods of the last sections rely mainly on windowing. The window used does not depend upon the true spectral density to be estimated. Because of this, misleading conclusions may sometimes be drawn when the estimated spectral density is used, especially if the available data set is limited so that no fixed window function can be found with sufficient resolution to resolve components of interest in the set of data itself. For this, two nonlinear procedures for spectral estimation with increased resolution were proposed: The Maximum Entropy Method and the Maximum Likelihood Method.

1.5.2.1- The Autoregressive or Maximum Entropy Method:

The nonparametric spectral density estimators $\hat{P}_{Per}(f)$, $\hat{P}_{Avr}(f)$, $\hat{P}_{BT}(f)$ are based on replacing the unknown autocorrelation function $r_{xx}(k)$, $k \in Z$ of the generating process $x(n)$ by that of the truncated signal $x^N(n)$ given in (1.25). In other words, the values of the autocorrelation function are considered to be zero for lags k beyond which no empiprical estimates can be computed. This assumption results in spectral leakage and loss of frequency resolution. This serious drawback motivated Burg (1967), Parzen (1968), and Kay (1988) to develop a parametric approach, known as the autoregressive or Maximum Entropy Method, for estimating the unknown spectral density of the signal. This approach is based on, instead of appending zeros to increase the length of the estimated autocorrelation function, estimating the autocorrelation function beyond the data limited range from a fitted parametric model to the process of

interest. They suggested to fit, using least squares method, the autoregressive model $AR(p)$ given in (1.3), and to estimate the spectral density by that of the estimated parametric model $AR(p)$ i.e.:

$$P(f) = \frac{2\sigma^2}{\left|1 - \sum_{k=1}^{p} \hat{a}_k e^{-2\pi jf}\right|^2} \tag{1.48}$$

where $\hat{a}_k$'s are the estimated parameters.

It has been shown that this approach has proven to be useful particularly for short realizations. Despite its superiority with respect to other spectral estimation methods, the usefulness of this method is marred to a considerable extent by a lack of a criterion for choosing the order p of the autoregressive process. It is known from experience that if p is chosen too small, it will result in a highly smoothed estimate obviating the resolution advantages of the Maximum Entropy Method whereas the use of a large value for p introduces spurious details in the estimated spectral density. Furthermore, it may be applied but with difficulty to the case where the sampling is either non-uniform or where various autocorrelation samples may be missing

1.5.2.2- The Maximum likelihood method:

The Maximum Likelihood Method may be derived by solving a classical optimal filtering problem. The filter is designed to pass the power in a narrowband about the signal frequency of interest and minimize the power due to interfering spectral components such as noise. The Maximum Likelihood Method is also known as a minimum variance spectral estimator Kay (1988), and it is defined by:

$$\hat{P}_{MV}(f) = \frac{1}{e^H \hat{R}_{xx}^{-1} e} \tag{1.49}$$

where e^H is the conjugate transpose of the vector e defined by:

$$e = [1 \quad \exp(j2\pi f) \quad \exp(j4\pi f) ..... \exp(j2\pi f(p-1))]$$

and $\hat{R}_{xx}$ is the pxp sample autocorellatoion matrix.

It has been concluded by using test data that the spectral density estimator (1.49) exhibits less variability and less resolution than the autoregressive spectral estimator given in (1.48) for large data sets. On the other hand, it exhibits more resolution than the periodogram and Blackman-Tukey estimators.

1.6- Illustrative example:

In order to illustrate the above spectral density methods of estimation, we simulated 128 uniformly spaced observations from the autoregressive process of order 2 represented by the mathematical model:

$$y(n) = 0.75y(n-1) - 0.3y(n-2) + \varepsilon_n$$

where ε_n is a gaussian white noise series with zero mean and variance $\sigma^2 = 1$.

The simulated series is shown in Figure 1.2 whereas Figure 1.3 shows the periodogram estimator which is not smooth at all with several peaks. In order to reduce the spectral leakage due to the truncation of measurements, we used the Hamming window with $M = 14$, and $M = 20$ where M represents the width of the window. The windowed periodogram, the Blackman-Tukey estimators along with the theoretical power spectral density, and the Maximum Entropy Method estimator,

assuming that we know the order of the model, are shown in Figure 1.4, and Figure 1.5. They indicate clearly that windowing improves considerably the smoothness of the periodogram estimator. Furthermore the Blackman-Tukey estimator follows exactly the shape of the true spectral density, and it is more sensitive to the choice of M than the periodogram.

The Maximum Entropy Method estimator is close in performance to that of the Blakman-Tukey estimator. In order to illustrate the effect of the cruciality of the choice of the order of parametric model $AR(p)$ which is required for the Maximum Entropy Method we applied the method by assuming that $p=1$, 3, and 4. The estimated spectral density are shown in Figure 1.6 which indicates that the underestimating of p will produce an oversmoothed estimate whereas overestimating the parameter p will produce an undersmooth spectral density.

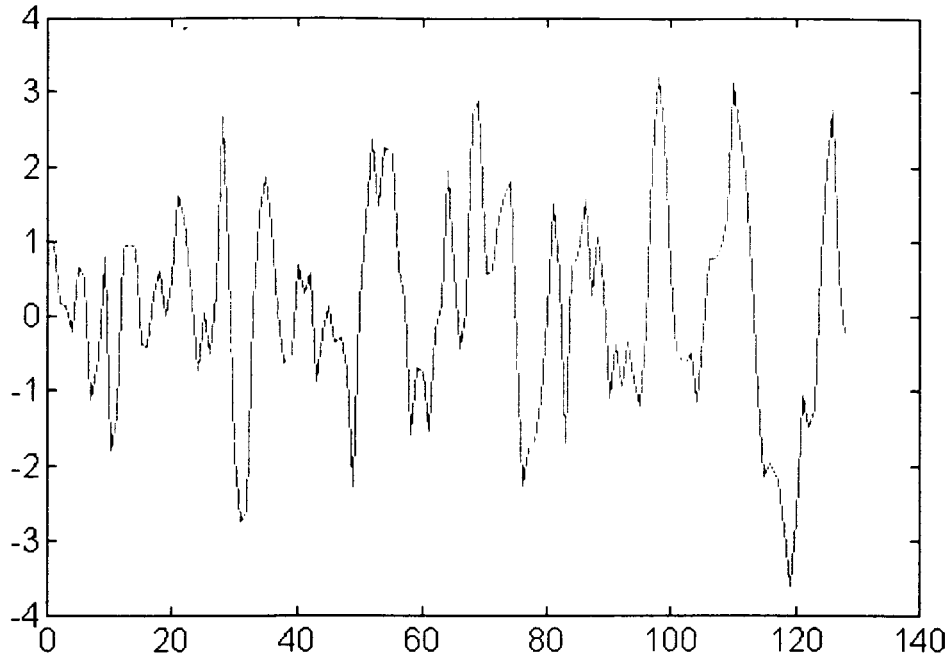

Figure 1.2: A realization of the second order AR process $y(n) = 0.75y(n-1) - 0.3y(n-2) + \varepsilon_n$

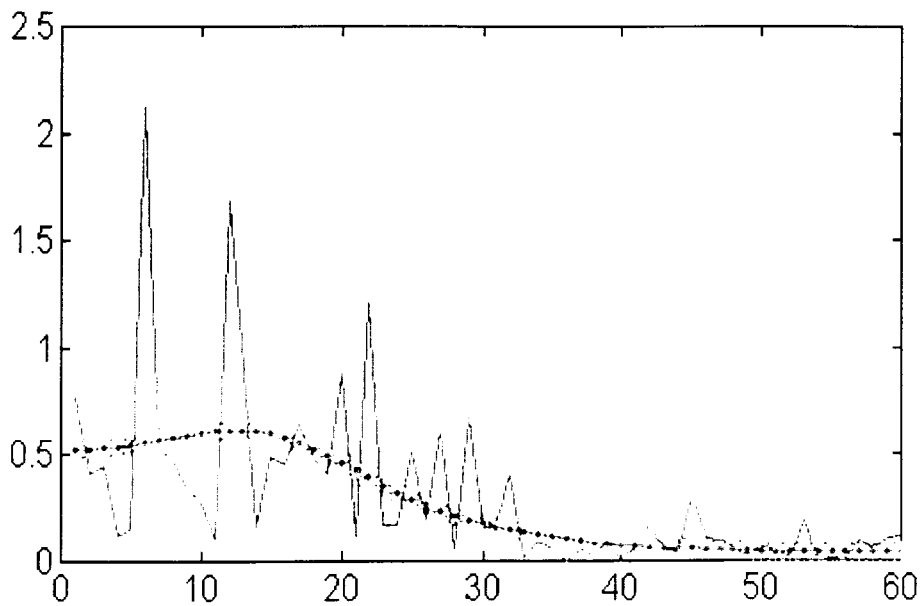

Figure 1.3: Periodogram spectral density estimator (-) without windowing, Theoretical spectral estimator (*)

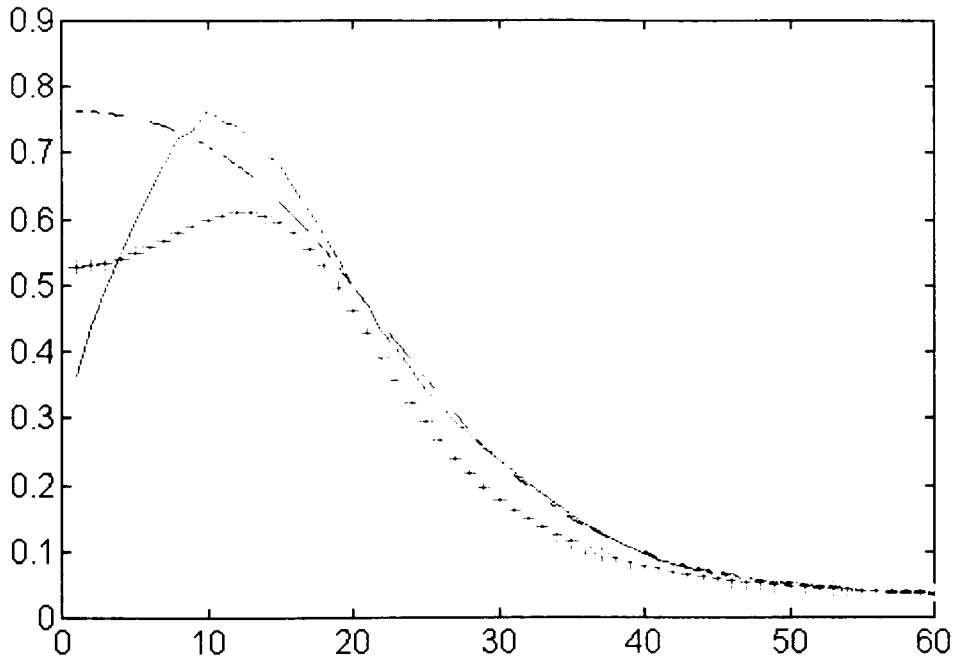

Figure 1.4: Periodogram spectral density estimator (-), M=14, Hamming, Blackman _Tuckey estimator (-.), M=14 Hamming,, Theoretical spectral estimator (x), MEM spectral density estimator (o)

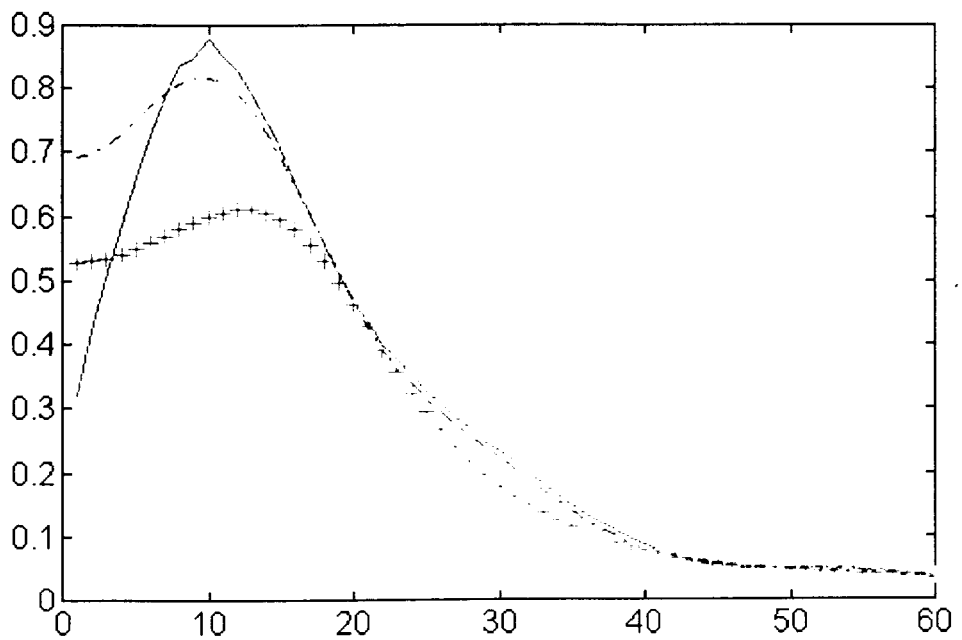

Figure 1.5: Periodogram spectral density estimator (-), M=20, Hamming, Blackman _Tuckey estimator (-.), M=20 Hamming, Theoretical spectral estimator (x), MEM density spectral estimator (o)

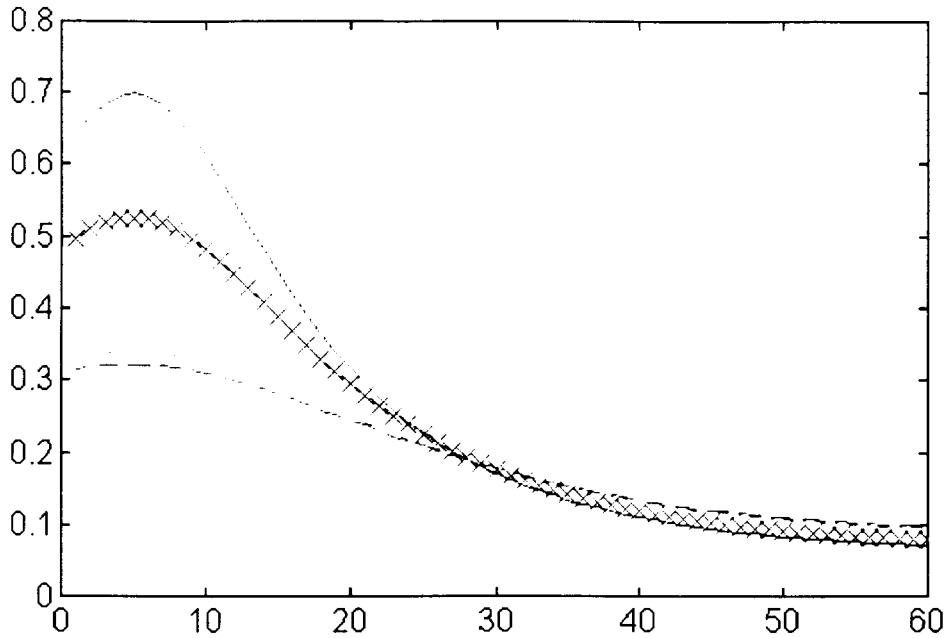

Figure 1.6: MEM estimator *p=1* (-), MEM estimator *p=2* (x), MEM estimator *p=3* (-.), MEM spectral density estimator *p=4* *(- -)*.

CHAPTER 2

Spectral density estimation of signals with missing observations using the Kalman smoother approach

2.1 Introduction:

In the last chapter, we reviewed some of the proposed methods for estimating the spectral density of random processes from a completely known realization $\{x(1),x(2),....x(n)\}$. However, in some practical situations, the realizations may not be fully observed on the time interval $[1,nT]$ on which they were recorded. In other words, the observations $x_i's$ are not available for some $i's$ of the set $\{1,.....,n$.

The estimation of the spectral density of the generating process $x(t)$ from incomplete realizations has been considered by Parzen (1981), Dunsmir and Robinson (1981), and Toloi and Moretin (1993). In this chapter, we discuss how the proposed method of Toloi and Morettin (1993) can be adapted for overcoming the difficulties caused by the missing observations. They considered the problem of estimating the spectral density of $x(t)$ from that of an observed amplitude modulated process $y(t)$ which is defined as:

$$y(t) = A(t)x(t)$$

where $A(t)$ is a modulation process.

It is obvious that the available observations of the incomplete original realization $\{x(1),x(2),....x(n)\}$ can be regarded as a modulated realization $\{y(n)\}$ where $y(n) = A(n)x(n)$ and $A(n)$ is given by:

$$A(n)=\begin{cases}1 & \text{if the process } x(n) \text{ is observed at time } n \\ 0 & \text{if the process } x(n) \text{ is not observed at time } n\end{cases} \quad (2.1)$$

Therefore, the incomplete realization augmented with a sequence of zero i.e.:

$$y(n)=\begin{cases}x(n) & \text{if the process } x(n) \text{ is observed at time } n \\ 0 & \text{if the process } x(n) \text{ is not observed at time } n\end{cases} \quad \text{for } n=1,...N \quad (2.2)$$

is a complete realization of the process $y(n)$ which can be used to estimate its spectral density using one of the estimation methods of chapter 1. The approach of Toloi and Morettin will be discussed in detail in the next section. In section 3, we suggest another method based on kalman filtering approach. This alternative approach consists mainly of estimating the missing observations using the well known kalman smoothing procedure and then estimating directly the spectral density of the process $x(n)$ using the augmented realization.

2.2 Spectral density estimation based on amplitude Modulated signals:

Toloi and Morettin (1993) have investigated the problem of spectral estimation of amplitude modulated signals and derived three different estimators of the spectral density of the unobserved process $x(n)$ using the modulated process $y(n)$ and the modulating sequence $A(n)$. They considered the cases where $A(n)$ is a deterministic sequence, random and uncorrelated process or random and correlated process. They derived directly the spectral density estimator of the signal $x(n)$ from the spectral estimator of the modulated signal $y(n)$ as follows:

Consider the realization $\{y(n)\}$ defined by (2.2). The periodogram of this realization is given by:

$$\hat{P}_{Per}^{y}(f)=\frac{1}{N}\left|\sum_{n=0}^{N-1}y(n)\exp(-j2\pi fn)\right|^{2} \tag{2.3}$$

The aim is to derive the periodogram $\hat{P}_{Per}^{x}(f)$ of the process $x(n)$ from that of $y(n)$. Using equations (2.3), and the fact that $y(n)=A(n)x(n)$ we obtain:

$$\hat{P}_{Per}^{y}(f)=\frac{1}{N}Y(n)Y^{T}(n)=\frac{1}{N}A(n)X(n)X^{T}(n)A^{T}(n) \tag{2.4}$$

where $Y^{T}(n)$ is the conjugate transpose of the vector $Y(n)=\left[y(1)e^{-j2\pi f},.....y(n)e^{-j2\pi fn}\right]$ and $A(n)=\left[a(1),.....a(n)\right]$.

Since the modulating sequence takes values 1 or 0, then (2.4) will be equivalent to:

$$\hat{P}_{Per}^{y}(f)=A(n)_{X}\hat{P}_{Per}^{x}(f)A^{T}(n)=A(n)A^{T}(n)\hat{P}_{Per}^{x}(f) \tag{2.5}$$

Hence, the periodogram of the process $x(n)$ can be evaluated as:

$$\hat{P}_{Per}^{x}(f)=\left\{2\pi\sum_{n=0}^{N-1}a^{2}(n)\right\}^{-1}\hat{P}_{Per}^{y}(f) \tag{2.6}$$

The factor 2π is added for normalization purposes. By regarding $\hat{P}_{Per}^{x}$ as a satisfactory approximation of the periodogram, the spectral density of $x(n)$ can be identified by windowing $\hat{P}_{Per}^{x}(f)$:

$$P_{AM}(f)=\frac{2\pi}{N}\sum_{s=0}^{N-1}W^{(N)}\left(f-\frac{2\pi s}{N}\right)\hat{P}_{Per}^{x}\left(\frac{2\pi s}{N}\right) \tag{2.7}$$

where $W^{(N)}(f)$ is a deterministic real even function of bounded variations in the frequency domain satisfying:

$$\int_{-\infty}^{+\infty} W(\beta)d\beta = 1$$

$$\int_{-\infty}^{+\infty} |W(\beta)|d\beta = 1$$

The statistical properties of the estimator $P_{AM}(f)$ are inherited from those of the periodogram $\hat{P}^x_{Per}(f)$. The expectation $E[P_{AM}(f)]$ can be computed as:

$$E[P_{AM}(f)] = E\left[\frac{2\pi}{N}\sum_{s=0}^{N-1} W^{(N)}\left(f - \frac{2\pi s}{N}\right)\hat{P}^x_{Per}\left(\frac{2\pi s}{N}\right)\right]$$

$$= \frac{2\pi}{N}\sum_{s=0}^{N-1} W^{(N)}\left(f - \frac{2\pi s}{N}\right)E\left[\hat{P}^x_{Per}\left(\frac{2\pi s}{N}\right)\right] \qquad (2.8)$$

Using the results obtained in equation (1.35) of the previous chapter and which gives the mean of the periodogram as:

$$E[\hat{P}_{Per}(f)] = \sum_{m=-(N-1)}^{N-1} \frac{N-|m|}{N} r_{xx}[m]e^{-j2\pi fm} \qquad (2.9)$$

we will deduce that:

$$E[P_{AM}(f)] = \frac{2\pi}{N}\sum_{s=0}^{N-1} W^{(N)}\left(f - \frac{2\pi s}{N}\right)\sum_{m=-(N-1)}^{N-1} \frac{N-|m|}{N} r_{xx}[m]e^{-j2\pi mf} \qquad (2.10)$$

As N tends to infinity, $E[P_{AM}(f)]$ tends to the unknown spectral density of the process $x(n)$. Toloi and Morettin have shown that $\text{var}[P_{AM}(f)] \to 0$ as $N \to 0$.

2.2.1- Illustrative example:

We simulated 128 observations from the autoregressive process of order 2 used in chapter 1, i.e. $y(n)=0.75y(n-1)-0.3y(n-2)+\varepsilon_n$

where ε_n is a gaussian white noise series with zero mean and constant variance $\sigma^2=1$. The theoretical spectral density is given by:

$$P^{AR}(f)=\frac{1}{\left|1-0.75e^{-j2\pi f}+0.3e^{-j4\pi f}\right|^2}$$

We estimated the spectral density of the complete realization using equation (1.41) of chapter 1. On the other hand, we computed the spectral density estimator (2.7) using 10, 20 and 40 missing values respectively for the same realization to get an insight on the effect of the number of missing values. The obtained results are shown in Figures 2.1-2.3 from which we deduce that the estimator (2.7) gives good results when the number of missing values is relatively small. However, its performance decreases when the number of missing values is increased. We notice that the peak is shifted away from that of the original realization.

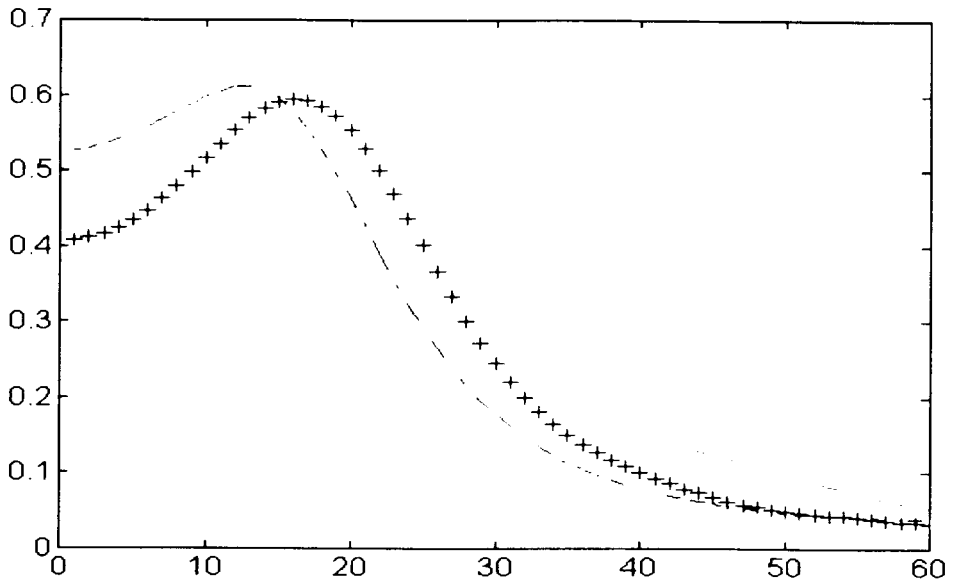

Figure 2.1: Theoretical spectral estimator of the original signal (x), spectral density using the original time series (-.), Amplitude-Modulated estimator with 10 missing values (-)

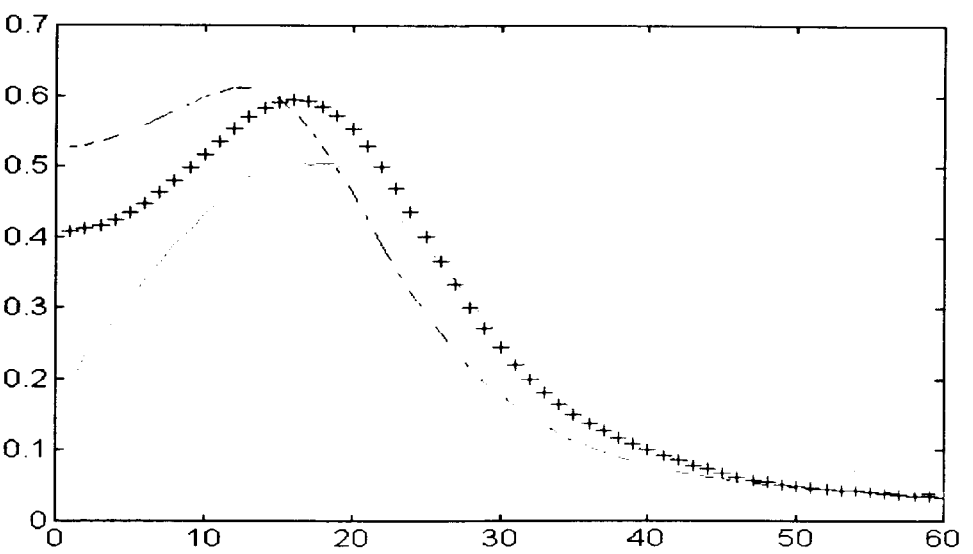

Figure 2.2: Theoretical spectral estimator of the original signal (x), Spectral density estimator of the original time series (-.), Amplitude modulated estimator with 20 missing values (-)

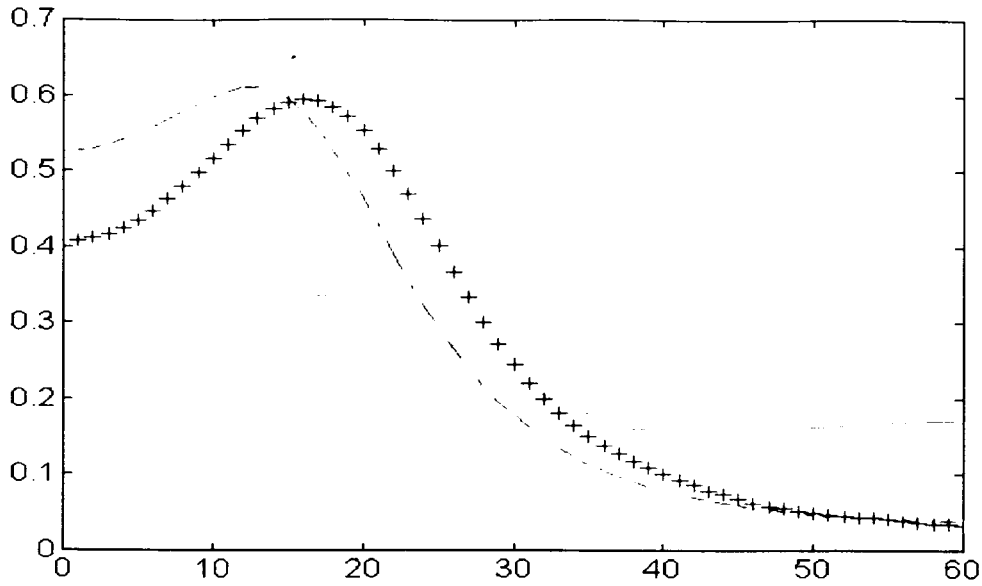

Figure 2.3: Theoretical spectral estimator of the original signal (x), spectral density estimator of the original time series (-.), Amplitude modulated estimator with 40 missing values (-)

2.3- Kalman filter approach:

The Kalman filter approach is used to estimate the state of a linear dynamic system which is perturbed by gaussian white noise using measurements corrupted by additive gaussian white noise. It is based on a state-space representation of systems. The first equation which is known as the state equation, models the dynamics of the process whereas the second equation, known as the observation equation, models the noisy observation signal. The Kalman filter [Grewal and Andrews (1993)] was derived initially as the solution to the Weiner filtering problem, using the state space model for discrete dynamic and random processes. It has been extended later on to the continuous time problem by Kalman and Bucy [Vaccaro and Fu Li (1990)]. It is derived as a recursive minimum mean square error predictor of the state $x(n)$ given the observation signal $y(n)$.

When we have a set of measurements $\{y(1), y(2),....y(n)\}$ that represents a stationary zero mean process in a white noise environment, the generating process can be described by the following state space model:

$$x(k) = F(k,k-1)x(k-1) + g(k)u(k) \tag{2.11}$$

$$y(k) = H(k)x(k) + \eta(k) \tag{2.12}$$

where:

$x(k)$ is the value of a one or p-dimensional state vector at time k.

$F(k,k-1)$ is a scalar or PxP state transition matrix that relates the states of the process at times $k-1$ and k.

$H(k)$ is a scalar or a $1xP$ vector.

$y(k)$ is the k^{th} observation.

$u(k)$, $\eta(k)$ are uncorrelated input and observation noises with covariance matrices $Q(k)$, and $R(k)$ respectively.

2.3.1- Derivation of the Kalman filter:

The kalman filter is an estimator $\hat{x}(k+1)$ of the state which is computed using the available data up to time k. If we denote by $\hat{y}(k/k-1)$ the least square error predictor of $y(k)$ based on the previous observations $\{y(1), y(2),....y(k-1)\}$, the associated innovation signal is defined as:

$$v(k) = y(k) - \hat{y}(k/k-1) \tag{2.13}$$

The signal $v(k)$ contains all that is unpredictable from the past observations including the noises $u(.)$, and $\eta(.)$. The concept of innovations is central to the derivation of the Kalman filter. For an optimal linear mean squared error estimate, the innovation signal must be orthogonal to the past observation vectors i.e.:

$$E\left[v(m)y^T(m-k)\right] = 0 \qquad k > 0 \tag{2.14}$$

$$E\left[v(m)v^T(k)\right] = 0 \qquad m \neq k \tag{2.15}$$

Using (2.13), and the fact that:

$$\hat{y}(k/k-1) = H(k)\hat{x}(k/k-1) \tag{2.16}$$

we obtain:

$$v(k) = H(k)\tilde{x}(k) + \eta(k) \tag{2.17}$$

where:

$$\tilde{x}(k) = x(k) - \hat{x}(k/k-1)$$

If we denote the covariance matrix of the innovation signal as Qv i.e.:

$$Qv = E\left[v(k)v^T(k)\right] \tag{2.18}$$

we obtain:

$$Qv = E\left[H(k)\tilde{x}(k)\tilde{x}^T(k)H^T(k)\right] + E\left[\eta(k)\eta^T(k)\right] \tag{2.19}$$

$$Qv = H(k)P(k)H^T(k) + R(k) \tag{2.20}$$

where $R(k)$, and $P(k)$ are the covariance matrices of the observation noise $\eta(k)$, and that of the signal prediction error $\tilde{x}(k)$ respectively.

If we denote by $\hat{x}(k+1/k)$ the Minimum Mean Square Error one step predictor of $x(k+1)$, the prediction of $x(k+1)$ based on the samples available up to time k can be expressed recursively as a linear combination of the prediction based on the samples available up to time $k-1$ and the innovation signal at time k as:

$$\hat{x}(k+1/k) = \hat{x}(k+1/k-1) + K(k)v(k) \tag{2.21}$$

where $K(k)$ is a $Px1$ matrix which is known as the Kalman gain matrix.

Equation (2.21) can be rewritten as:

$$\hat{x}(k+1/k) = F(k+1/k)\hat{x}(k/k-1) + K(k)v(k) \tag{2.22}$$

Multiplying both sides of equation (2.22) by $v^T(k)$ and computing the expectation of the two sides of the equation, we will get:

$$E[\hat{x}(k+1/k)v^T(k)] = E[F(k+1/k)\hat{x}(k/k-1)v^T(k)] + K(k)E[v(k)v^T(k)] \tag{2.23}$$

Since $\hat{x}(k+1/k)$ is orthogonal to the past values $v^T(k)$, the Kalman gain matrix will be given by:

$$K(k) = E[\hat{x}(k+1/k)v^T(k)]Qv^{-1} \tag{2.24}$$

Using the fact that:

$$E[\hat{x}(k+1/k)v^T(k)] = E[(x(k+1) - \tilde{x}(k+1/k))v^T(k)] \tag{2.25}$$

and $E[\tilde{x}(k+1/k)v^T(k)]=0$, the equation (2.25) becomes:

$$E[\hat{x}(k+1/k)v^T(k)]=F(k+1,k)E\left[\tilde{x}(k/k-1)\tilde{x}^T(k/k-1)\right]H^T(k) \qquad (2.26)$$

We can easily show that the Kalman gain matrix $K(k)$ may be recursively computed using:

$$K(k)=F(k+1,k)P(k)H^T(k)[H(k)P(k)H^T(k)+R(k)]^{-1} \qquad (2.27)$$

where:

$$P(k)=L(k)P(k-1)L^T(k)+g(k)Q(k)g^T(k)+K(k-1)R(k)K^T(k-1) \qquad (2.28)$$

and $L(k)$ is a PxP matrix defined by:

$$L(k)=[F(k,k-1)-K(k-1)H(k-1)] \qquad (2.29)$$

As a consequence, the Kalman filter can be summarized as:

Step 0: We assume that we have an output observation sample $\{y(0),y(1),....y(N-1)\}$, and an output state $\hat{x}(k)$ with initial conditions set to:

$\hat{x}(0/0)=0$, and $P(1)=\text{cov}\{x(1),x(1)\}$ for $k=1,2,.....$

Step1: Compute the Innovation signal $v(k)=y(k)-H(k)\hat{x}(k/k-1)$

Step 2: Compute the Kalman gain matrix $K(k)$:

$$K(k)=F(k+1,k)P(k)H^T(k)[H(k)P(k)H^T(k)+R(k)]^{-1}$$

Step 3: Compute the estimate $\hat{x}(k+1/k)$:

$$\hat{x}(k+1/k) = F(k+1/k)\hat{x}(k/k-1) + K(k)v(k)$$

Step 4: Update the prediction error correlation matrix $P(k+1)$ using the formula:

$$P(k+1) = L(k+1)P(k)L^T(k+1) + g(k+1)Q(k+1)g^T(k+1) + K(k)R(k+1)K^T(k)$$

where:

$$L(k+1) = [F(k+1,k) - K(k)H(k)]$$

Step 5: Increment k and start again from step 1.

As a consequence, if $y(l)$ is a missing observation, then it can be estimated by $\hat{y}(l/l-1)$ using the Kalman filtering algorithm. In the next section, we present the Kalman smoother algorithm which is more appropriate for estimating the missing values especially if l is small.

2.3.2-Kalman smoother:

The Kalman smoother algorithm is devised to estimate the state $x(t)$ of a system at time t using measurements made before and after time t whereas the Kaman filter uses only the past values. The accuracy of the Kalman smoother is generally superior to that of Kalman filter because it uses more measurements for computing the estimate. It can be derived from the Kalman filter by introducing a backward term to compute an estimate $\hat{x}(t)$ using future values as follows:

Using the state space model (2.11), the optimal estimate of $x(n+1)$ obtained from the data observed up to time n is given by:

$$\hat{x}(n+1/n) = \sum_{k=1}^{n} E\{x(n+1)v^T(k)\}Q_v^{-1}(k)v(k) \qquad (2.30)$$

where v is the innovation vector defined in (2.16) and Q_v is its covariance matrix given in (2.20).

Similarly, $x(i)$ can be estimated from the available data set $\{y(1), y(2), ...y(n)\}$ for $1 \le i \le n$ by:

$$\hat{x}(i/n) = \sum_{k=1}^{i} E\{x(i)v^T(k)\}Q_v^{-1}(k)v(k) + \sum_{k=i+1}^{n} E\{x(i)v^T(k)\}Q_v^{-1}(k)v(k) \qquad (2.31)$$

Thus, we deduce that the smoothed estimate $\hat{x}(i/n)$ for $i < n$ can be derived by modifying the filtered estimate $\hat{x}(i/i)$. To illustrate this, we will derive a computational formula to compute the smoothed estimate $\hat{x}(i/i)$ backward starting from time n.

Using equations (2.17), and (2.31), we can prove that:

$$v(k) = H(k)F(k,k-1)\tilde{x}(k-1/k-1) + g(k-1)u(k-1) + n(k) \qquad (2.32)$$

and

$$E\{x(i)v^T(k)\} = E\{\tilde{x}(i/i)\tilde{x}^T(k-1/k-1)\}F^T(k,k-1)H^T(k) \qquad (2.33)$$

Therefore, equation (2.31) becomes:

$$\hat{x}(i/n) = \hat{x}(i/i) + \sum_{k=i+1}^{n} E\{\tilde{x}(i/i)\tilde{x}^T(k-1/k-1)\}F^T(k,k-1)H^T(k)Q_v^{-1}(k)v(k) \qquad (2.34)$$

Using equation (2.22), and (2.17), we obtain:

$$\hat{x}(k+1/k) = F(k+1,k)\hat{x}(k/k-1) + K(k)[y(k) - H(k)\hat{x}(k/k-1)] \qquad (2.35)$$

where:

$$\hat{x}(k+1/k) = F(k+1,k)\hat{x}(k/k) \tag{2.36}$$

If we assume that $F(k+1,k)$ is non-singular, after some mathematical simple substitutions, we deduce that:

$$\hat{x}(i/n) = \hat{x}(i/i) + \sum_{k=i+1}^{n} E\{\tilde{x}(i/i)\tilde{x}^T(k-1/k-1)\}F^T(k,k-1)P^{-1}(k)[\hat{x}(k/k) - \hat{x}(k/k-1)] \tag{2.37}$$

this can be expressed in the closed form as:

$$\hat{x}(i/n) = G(i)\hat{x}(i/i) + A(i)\hat{x}(i+1,n) \qquad for\ \ i = 1,....n-1 \tag{2.38}$$

where:

$$A(i) = \Gamma(i)F^T(i+1,i)P^{-1}(i+1)$$

$$G(i) = I - A(i)F(i+1,i)$$

and $$\Gamma(i) = E\{\tilde{x}(i/i)\tilde{x}^T(i/i)\ .$$

This equation provides a way for computing recursively the smoothed estimate $\hat{x}(i/n)$ backward from time n once the solutions $\hat{x}(i/i)$ for $i = 1,.....n$ have been computed.

As a consequence, the predicted value $\hat{y}(k/N)$ can be estimated recursively by replacing the state values by their estimates in the model (2.11), i.e.:

$$\hat{y}(k/N) = H\hat{x}(k/N) \tag{2.39}$$

where:

$$\hat{x}(k/N) = A(k)\hat{x}(k+1/N) + G(k)\hat{x}(k/k) \tag{2.40}$$

$$\hat{x}(k/k) = \phi(k)\hat{x}(k-1/k-1) + K(k)y(k) \tag{2.41}$$

$$K(k) = P(k)H^T\left[HP(k)H^T + R\right]^{-1} \tag{2.42}$$

$$\phi(k) = [I - K(k)H]F \tag{2.43}$$

$$A(k) = \Gamma(k)F^T P^{-1}(k+1) \tag{2.44}$$

$$G(k) = I - A(k)F \tag{2.45}$$

$$\Gamma(k) = P(k) - P(k)H^T\left[HP(k)H^T + R\right]^{-1} HP(k) \tag{2.46}$$

and the unknown state-error covariance matrix $P(k)$ is computed recursively by solving the following algebraic RICCATI equation:

$$P(k+1) = F\left\{P(k) - P(k)H^T\left[HP(k)H^T + R\right]^{-1} HP(k)\right\}F^T + gQg^T \tag{2.47}$$

where $P(0) = E\{x(0)x(0)^T$ and the initial state $\hat{x}(0/0)$ is taken to be zero.

Since this estimation procedure is computationally expensive, we suggest to use time-invariant filtering method, which was proposed by Vaccarro and Fu Li (1990) to reduce the effect of transients.

They showed that, using time-invariant filtering, the states $\hat{x}(k/k)$ could be expressed by:

$$\hat{x}(k/k) = \phi\,\hat{x}(k-1/k-1) + Ky(k) \tag{2.48}$$

$$\hat{x}(k/k) = \phi^k \hat{x}(0/0) + \sum_{i=1}^{k} \phi^{k-i} K y(i) \tag{2.49}$$

where $\phi = [I - KH]F$

Bearing in mind that the initial state vector $x(0)$ summarizes the past history of the signal for $n \leq 0$, it is more appropriate to estimate $\hat{x}(0/0)$ from the observed data $\{y(1), y(2), ...y(n)\}$ rather than setting it to be equal to zero.

They have showed that the observed value $y(k)$ can be expressed as:

$$y(k) = H\phi^k \hat{x}(0/0) + H\sum_{j=1}^{k} \phi^{k-j} K y(j) + e(k) + n(k) \tag{2.50}$$

where $e(k)$ is the error of the estimated signal when forward filtering is used with $\hat{x}(0/0) = 0$.

If we define the residual data $z(k)$ as:

$$z(k) = y(k) - H\sum_{j=1}^{k} \phi^{k-j} K y(j) \tag{2.51}$$

which can be written as:

$$z(k) = H\phi^k \hat{x}(0/0) + w(k) \tag{2.52}$$

where $w(k)$ is the sum of the observation and estimation errors $\eta(k)$ and $e(k)$.

As a consequence, we notice that the initial estimate $\hat{x}(0/0)$ can be computed using the Least Square Error method. In other words, we

estimate $\hat{x}(0/0)$ by the value which minimizes the sum of the squares of the total error $\mathbf{w}$:

$$J = \sum_{n=1}^{N} w^2(n) = \mathbf{w}^T \mathbf{w} \tag{2.53}$$

It can easily be verified that the least square estimate of the initial state vector is given by:

$$\hat{x}(0/0) = (\theta_f^T \theta_f)^{-1} \theta_f^T \mathbf{Z} \tag{2.54}$$

where:

$$\mathbf{Z} = [z(1), ..., z(N)]^T$$

and

$$\theta_f = [H\phi^1, ..., H\phi^N]^T$$

The modified Kalman smoother algorithm using a data-dependent correction term for estimating the initial state vector is summarized in the following four steps:

i) -Compute the estimate $\hat{x}(k/k)$ by using equation (2.41) and $\hat{x}(0/0) = 0$.

ii) -Evaluate $\hat{x}(0/0)$ from equation (2.54)

iii) -Update the states $\hat{x}(k/k)$ using the equation:

$$\hat{x}(k/k) = \hat{x}(k/k) + \phi^k \hat{x}(0/0) \tag{2.55}$$

iv) -Compute the states $\hat{x}(k/N)$, and $y(k)$ by using (2.40), and (2.39) respectively.

2.3.3- Spectral density estimation of signals with missing values using the Kalman approach:

Assume that we have a sample of N observations $\{y(1),.....y(N)\}$ with some missing observations $y(i)$ where $1<i<N$. In order to estimate the spectral density of this process using the different available spectral estimation methods, we construct a new augmented realization, where the missing values are replaced by their estimated values using the Kalman smoother algorithm described in the previous section. In other words, we first estimate the missing values $\hat{y}(i)$ by using the available data set $\{y(i-M),..,y(i-1),y(i+1),...,y(i+M)\}$ in the neighborhood of the missing value $y(i)$, and then we compute the spectral density using the windowed autocorrelation function given in the first chapter by equation (1.41):

$$\hat{P}_y(f)=\sum_{k=-(N-1)}^{N-1} w[k]\hat{r}_{yy}[k]\exp(-j2\pi fk) \tag{2.56}$$

where $\hat{r}_{yy}(k)$ is the sample autocorrelation function of augmented realization $y(n)$ defined by:

$$\hat{r}_{yy}(k)=\frac{1}{N}\sum_{n=0}^{N-1-|k|} y[n]y[n+k] \qquad k=0,\pm 1,.....\pm N\text{-}1$$

and $w(k)$ is the hamming window defined by:

$$w(k)=\begin{cases}\left(0.54+0.46\cos\left(\pi\dfrac{k}{M}\right)\right) & ,|k|\le M\\ 0 & ,|k|>M\end{cases} \tag{2.57}$$

The statistical performances of this proposed updated estimator are inherited from the classical estimator. For instance, its expectation is given by:

$$E[P_y(f)] = \sum_{k=-(N-1)}^{N-1} w(k) E\left[\hat{r}_{yy}(k)\right] \exp(-j2\pi fk) \qquad (2.58)$$

Since the Kalman smoother algorithm is asymptotically consistent, the augmented series can be regarded as a realization of the process of interest. Then the sample autocorrelation function $\hat{r}_{yy}(k)$, which is computed from the augmented sample, is also an asymptotically consistent estimate of the unknown autocorrelation function $r_{yy}(k)$. Thus,

as $N \to \infty$, $E[P_y(f)] = \sum_{k=-\infty}^{\infty} r_{yy}(k) \exp(-j2\pi fk)$

That is $P_y(f)$ is asymptotically an unbiased estimate of the unknown spectral density. On the other hand, using the same argument, we can show that its asymptotic variance tends to that of the classical estimate which is computed from the complete realization.

2.3.4- Illustrative example:

We generated 128 observations from the same autoregressive model of order 2 used in the previous examples. The following figures show the spectral density of the original realization without the missing values, the amplitude-modulation based approach, and the estimator (2.56) which is based on the Kalman smoother approach. In order to compute a smoothed estimate of the missing value $y(i)$, we use the data points $\{y(i-M),...,y(i-1),y(i+1),...,y(i+M)\}$ where M was taken to be equal to 6 in this example. The estimators are compared when we have 10, 20, and 40 uniformly placed missing values respectively. In the case of 10 missing values, the missing observations were:

$\{y(5),\ y(12),\ y(19),\ y(20),\ y(33),\ y(45),\ y(50),\ y(64),\ y(79),\ y(93)\}$

The Kalman filter based spectral estimator performs better than the amplitude modulated based spectral density estimator. However, this approach requires the knowledge of the parameters of the model of the system. Unfortunately, in practical situations, these parameters are not known and estimating them is not an easy task especially using incomplete realizations. In order to bypass this difficulty, it is more suitable to use a non-parametric method to estimate the missing values. This approach will be discussed in the next chapter.

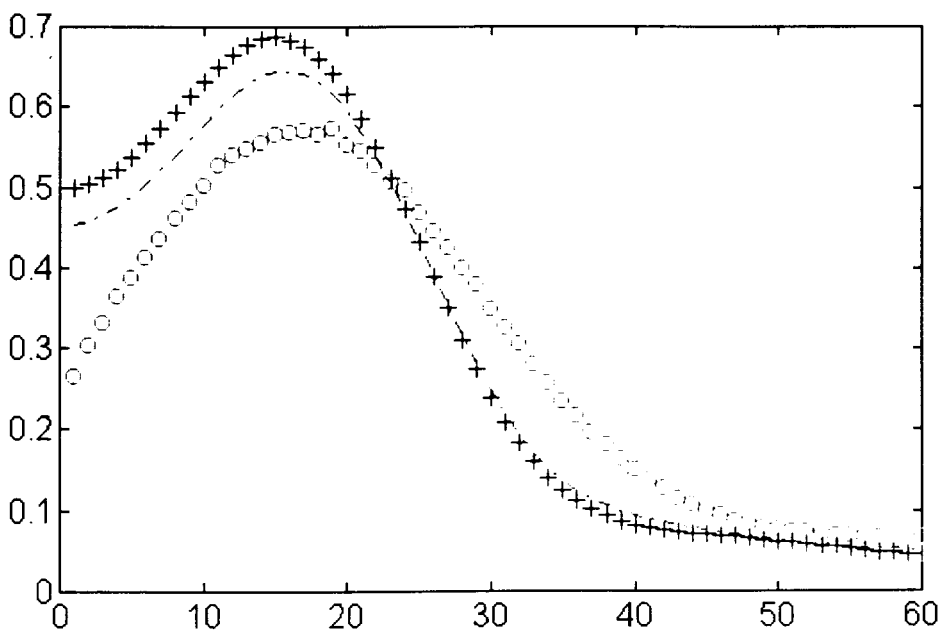

Figure 2.4: Spectral density using the original time series (x), Kalman approach spectral estimator with 10 missing values (-.), Amplitude-Modulated based spectral estimator (o)

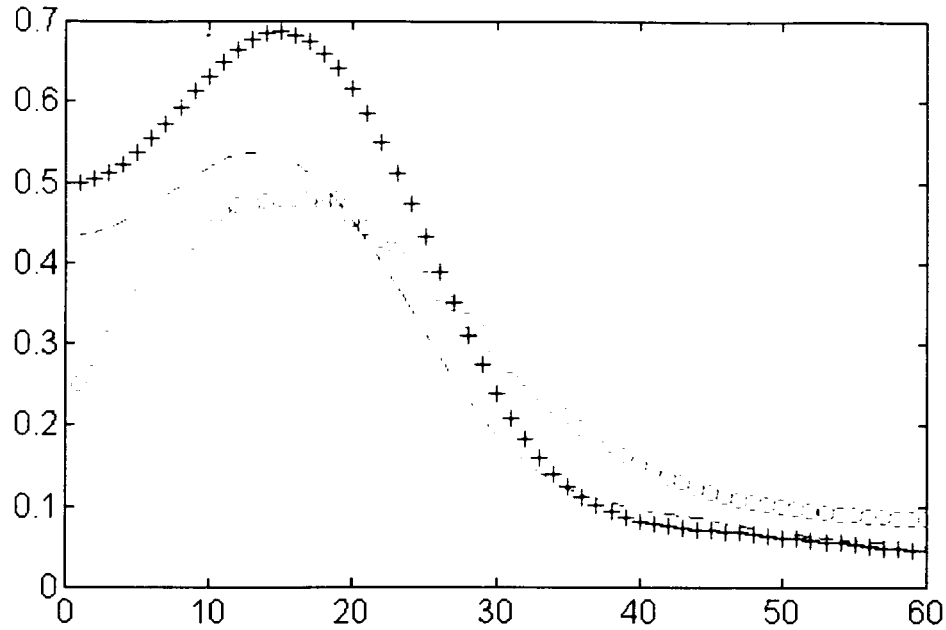

Figure 2.5: Spectral density using the original time series (x), Kalman approach spectral estimator with 20 missing values (-.), amplitude-modulated based spectral estimator (-)

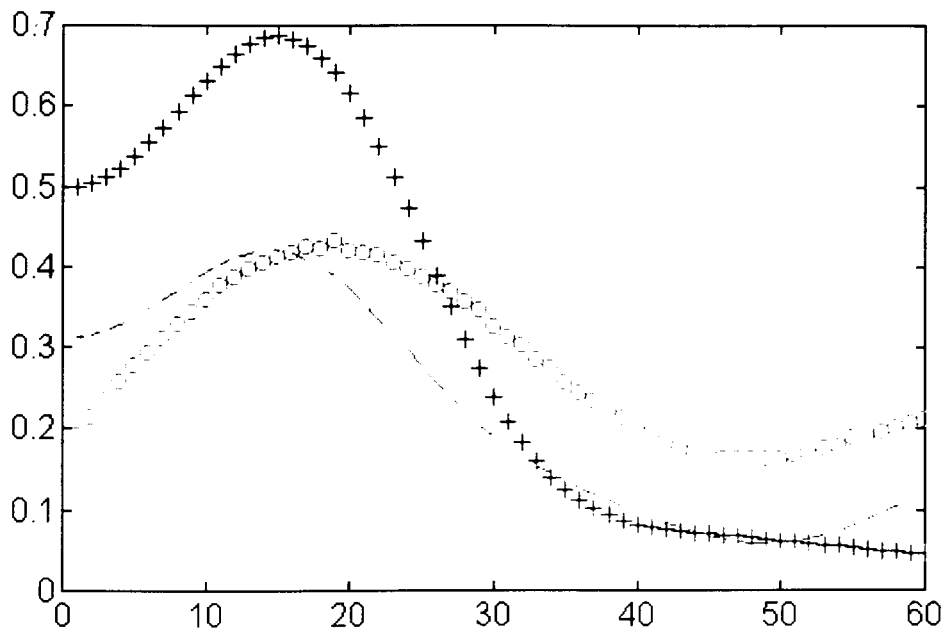

Figure 2.6: Spectral density using the original time series (x), Kalman approach spectral estimator with 40 missing values (-.), amplitude-modulated based spectral estimator (-)

Chapter 3

Spectral density estimation of signals with missing values using the Non-parametric kernel smoothing approach

3.1- Introduction:

The Kalman filter approach discussed in the previous chapter is a parametric method which requires the knowledge of the model of the process whose spectral density is to be estimated. Investigations have been done by several authors to estimate the model of a process where the main difficulty is how to determine the family of the model (AR, ARMA, or MA) which is more suitable for the process of interest. In the previous chapter we followed the principle given by Parzen (1963) which consists in taking into consideration the problem of missing observations before estimating the spectral density. We estimated first the missing values using the Kalman smoother, and then evaluated the spectral density of the process based on the new constructed realization. In this chapter we will follow the same principle but instead of using the parametric kalman method to estimate the missing values, we use a non-parametric method to bypass the problem of estimating the model of the realization. This non-parametric method consists in using the kernel method of curve fitting to be discussed in the following sections.

3.2- Non-parametric kernel method of curve fitting:

Assume that we have N observations $(x_1, y_1), (x_2 y_2), \ldots (x_N, y_N)$ which are generated from the following model:

$$y_j = \mu(x_j) + e_j \qquad j=1.....N \tag{3.1}$$

where e_j are zero mean, uncorrelated random variables with common variance σ^2 and the x_j are non-stochastic design points assumed to satisfy $a \le x_1 < x_2 ..... < x_N \le b$, for fixed finite constants a, and b. The problem is to estimate the unknown function $\mu(x)$ non-parametrically. The first nonparametric estimator $\hat{\mu}(.)$ is the Fourier estimator defined by:

$$\hat{\mu}(x) = \sum_{j=1}^{N} y_j K_h(x - x_j) \tag{3.2}$$

where $K_h(.)$ is the Dirichlet kernel defined by:

$$K_h(u) = \sin(2\pi u / h) \Big/ \sin(\pi u) \tag{3.3}$$

and h is a parameter to be specified. $K_h(u)$ is symmetric about 0 and have a maximum at zero.

When the design points x_j are not equally spaced, the estimator $\hat{\mu}(.)$ of (3.2) will no longer be a Fourier series estimator of $\mu(.)$, but it can be considered as an estimator of $\mu(.)$ due to the properties of K_h (symmetric and centred at zero) since $\hat{\mu}(.)$ is a type of weighted average of observations y_j's which are in the vicinity of x_j's.

This motivated the development of the kernel estimator of $\mu(.)$ which is defined by:

$$\hat{\mu}(x) = \sum_{j=1}^{N} y_j K_h(x - x_j) \tag{3.4}$$

when the design points x_j's are non-equidistant and $0 < x_j < 1$. The kernel function $K_h(.)$ should have finite support e.g. [-1,1] in order to ensure that the observation y_j will contribute most to its own estimate. Furthermore, to improve the asymptotic properties of $\hat{\mu}(.)$, $K_h(.)$ should satisfy the following moment conditions:

$$\int_{-1}^{1} K(u)du = 1 \tag{3.5.a}$$

$$\int_{-1}^{1} uK(u)du = 0 \tag{3.5.b}$$

$$\int_{-1}^{1} u^2 K(u)du = \alpha \neq 0 \tag{3.5.c}$$

$$\int_{-1}^{1} K(u)^2 du < \infty \tag{3.5.d}$$

Kernels of higher orders require higher order moments conditions. They are used when derivatives of the estimator $\hat{\mu}(.)$ are required. The most commonly used kernels are the uniform, triangular and quadratic kernel which are defined below respectively:

$$K(u) = \begin{cases} \frac{1}{2} & ,|u| \leq 1 \\ 0 & ,|u| > 1 \end{cases} \tag{3.6.a}$$

$$K(u) = \begin{cases} 1-|u| & ,|u| \leq 1 \\ 0 & ,|u| > 1 \end{cases} \tag{3.6.b}$$

$$K(u)=\begin{cases}\frac{3}{4}\left(1-u^2\right) & ,|u|\le 1\\ 0 & ,|u|>1\end{cases} \tag{3.6.c}$$

$$K(u)=\begin{cases}\frac{3}{8}\left(3-5u^2\right) & ,|u|\le 1\\ 0 & ,|u|>1\end{cases} \tag{3.6.d}$$

The quadratic kernel of (3.6.d) is known as the minimum variance kernel. There exist different variants of the kernel estimator given in (3.2). These variants are given below:

$$\hat{\mu}_1(x)=(Nh)^{-1}\sum_{i=1}^{N}K\left(\frac{x-x_i}{h}\right)y_i \tag{3.7.a}$$

$$\hat{\mu}_2(x)=\sum_{i=1}^{N}(x-x_i)K\left(\frac{x-x_i}{h}\right)y_i \tag{3.7.b}$$

The estimator (3.7.b) was proposed by Priestley and Chao (1972). It is a modification of the one given by (3.7.a) where the factor Nh is replaced by $(x\text{-}x_i)$. It is more appropriate for unequally spaced data. In order to ensure that the weights sum to one, Nadaray (1964) and Watson (1964) introduced a local weighted average of the y_i's given by:

$$\hat{\mu}_3(x)=\left.(Nh)^{-1}\sum_{i=1}^{N}K\left(\frac{x-x_i}{h}\right)y_i\right/\hat{f}(x) \tag{3.7c}$$

where $\hat{f}(.)$ is the kernel density estimator of the marginal density $f(x)$ of X given by:

$$\hat{f}(x)=(Nh)^{-1}\sum_{i=1}^{N}K\left(\frac{x-x_i}{h}\right)$$

The estimator (3.7.c) is more appropriate than the other estimators when the x_j are random. Clark (1977), Gasser and Muller (1979) have studied and investigated other kernel estimators. These estimators are defined by (3.7.d) and (3.7.e).

$$\hat{\mu}_4(x) = \sum_{i=1}^{N}\left(\frac{1}{h}\int_{s_{j-1}}^{s_j} K\left(h^{-1}(x-s)\right)ds\right)y_j \qquad (3.7.d)$$

where

$$s_0 = a, s_{j-1} \le x_j \le s_j, \quad j = 1,\ldots\ldots, N-1, \; s_N = b$$

and:

$$\hat{\mu}_5(x) = \int_b^a \frac{1}{h} K\left(h^{-1}(x-s)\right)h_y(s)ds \qquad (3.7.e)$$

where

$$h_y(s) = \begin{cases} y_1 & \text{if} \quad s \le x_1 \\ y_j\left(\dfrac{x_{j+1}-s}{x_{j+1}-x_j}\right) + y_{j+1}\left(\dfrac{s-x_j}{x_{j+1}-x_j}\right) & \text{if} \quad x_j \le s \le x_{j+1} \\ y_N & \text{if} \quad s \ge x_N \end{cases}$$

Before discussing their asymptotic properties, it is worth noting that this approach is applicable for modeling time series when $y_j's$ are the observed records and the design points x_j will be just the instants $t_j's$ at which the $y_j's$ had been recorded since the model of Time series $\{y_i's\}$ can always be expressed as:

$$y_j = \mu(t_j) + \varepsilon_j$$

where $\mu(t_j)$ may depend on the previous records y_{j-k}, and the errors ε_{j-k}'s for $k = 1,.....,j-1$.

It is also important to note that the missing records $y(j)$ are not considered as a drawback specially if the kernel estimator used is appropriate for non-equidistant t_j's since, in this case, they can simply be ignored.

3.3- Asymptotic theory:

In the asymptotic study of kernel estimators, we will focus on the kernel estimator given by (3.7.d) and refer to the work of Gasser and Muller (1979). For simplicity, we assume that $[a,b]=[0,1]$ and $x_j = j/N$.

If μ is continuously differentiable, then it can be shown that the expected mean of the estimator is given by:

$$E\{\hat{\mu}_h(x)\} = \frac{1}{h}\int_0^1 K\left(h^{-1}(x-s)\right)\mu(s)ds + O\left(n^{-1}\right) \qquad (3.8)$$

and the Bias of the estimator can be expressed as:

$$E\{\mu_h(x)\} - \mu(x) = \frac{h^2}{2}\mu''(x)\int_0^1 u^2 K(u)du + o(h^2)O\left(n^{-1}\right) \qquad (3.9)$$

From equation (3.9), we can conclude that the bias of the estimator given by (3.7d) will be larger in areas where the term $\mu''(x)$ changes rapidly. If the bias term is positive, then the estimator $\hat{\mu}(x)$ is overestimating the estimate whereas when it is negative, it simply underestimates it. The inspection of data will give an idea about the areas where the estimator will not behave as good as other estimators.

The variance of $\hat{\mu}_h(x)$ can be expressed as [Eubank (1988)]:

$$Var\ \hat{\mu}_h(j/N) = \frac{\sigma^2}{Nh} \int_{(j-N)/Nh}^{j/Nh} K(u)^2\, du + O\left((Nh)^{-2}\right) \tag{3.10}$$

where σ^2 is the variance of the error signal ε_j.

The mean square error $R(j/N,h)$ at a point $x \in [0,1]$ and the average mean square error $R_N(h)$ are defined by:

$$R(j/N,h) = E\left(\mu(j/N) - \hat{\mu}_h(j/N)\right)^2 \tag{3.11}$$

$$R_N(h) = \frac{1}{N}\sum_{j=1}^{N} E\left(\mu(j/N) - \hat{\mu}_h(j/N)\right)^2 \tag{3.12}$$

Equation (3.11) can be rewritten as:

$$R(j/N,h) = \left(\mu(j/N) - E\hat{\mu}(j/N)\right)^2 + Var\left[\hat{\mu}(j/N)\right] \tag{3.13}$$

Using (3.13), (3.8), and (3.9), we deduce that (3.12) can be expressed as:

$$R_N(h) \approx \frac{\sigma^2}{Nh} \int_{(j-N)/Nh}^{j/hN} K(u)^2\, du + \frac{h^4\left(\mu''(j/N)\right)^2}{4}\left[\int_{-1}^{1} u^2 K(u)^2\, du\right]^2 \tag{3.14}$$

we notice that $R_N(h)$ decays to zero if $h \to 0$ and $Nh \to 0$ simultaneously as $N \to \infty$.

As a consequence, the kernel nonparametric estimator is asymptotically unbiased and consistent estimator of the unknown function $\mu(x)$. However, its computation requires the choice of the smoothing parameter h.

The optimal value of the parameter h, which minimizes (3.14) is given by:

$$h_{opt} = \left\{ \frac{\sigma^2 N^{-1} \int_{-1}^{1} K(u)^2 du}{\left[\int_0^1 \mu''(x)^2 dx \right] \left[\int u^2 K(u) du \right]^3} \right\}^{1/5} \tag{3.15}$$

The associated asymptotically optimal risk is given by:

$$R_N(x, h_{opt}) \approx \frac{1.25}{N^{4/5}} \left\{ \int_0^1 \mu''(x)^2 dx \right\}^{1/5} \left\{ \sigma^4 \left[\int_{-1}^{1} K(u)^2 du \right]^2 \left[\int_{-1}^{1} u^2 K(u) du \right] \right\}^{2/5} \tag{3.16}$$

Unfortunately, h_{opt} can not be computed easily since it depends on the second derivative of the unknown function $\mu(x)$. This problem can be overcome by using one of the existing data-based methods such as the generalised cross validation rule to be discussed in the next section.

3.4- Selection of the bandwidth parameter *h*:

The smoothing parameter h determines the number, in a certain series, of the observations near y_j which will be considered to compute $\hat{\mu}(.,h)$. The choice of the smoothing parameter is very important since it governs the smoothness of the estimator. The smaller is the bandwidth parameter h the wigglier is the estimator $\hat{\mu}(.,h)$. In this case $\hat{\mu}(.,h)$ relies heavily on the data near x. On the other hand, larger bandwidth parameters will allow more averaging of data. This results in producing very smooth

estimators $\hat{\mu}(.,h)$ but it increases the bias. One can conclude that one of the crucial points in applying the kernel estimator is the choice of the bandwidth h. Several methods known as data driven techniques have been suggested to overcome this problem. One of these methods is the method of Cross-Validation rule (CV).

Suppose that h is restricted to some set $H_N \subset R^+$. We define the CV rule as the minimizer of the function $CV(h)$ over the set H_N given by:

$$CV(h)=\frac{1}{N}\sum_{j=1}^{N}\{\, y_j-\hat{\mu}_j(x_j,h)\,\}^2 w(x_j) \tag{3.17}$$

where $w(.)$ is a non-negative weighing function and $\hat{\mu}_j(.,h)$ is the leave-one-out estimator i.e: the estimator computed from the data by omitting the j^{th} observation. Kim and Cox (1996) applied this criterion using the kernel estimator (3.7.c) where:

$$\hat{\mu}_j(x)=\left.(N-1)^{-1}\sum_{i\neq j}h^{-1}K\left(\frac{x-x_i}{h}\right)y_i \right/ \hat{f}_j(x) \tag{3.18}$$

and

$$\hat{f}_j(x)=(N-1)^{-1}\sum_{i\neq j}^{N}h^{-1}K\left(\frac{x-x_i}{h}\right) \tag{3.19}$$

By using the CV method given by (3.17) and the kernel estimator (3.7.c), then the loss function defined by:

$$N^{-1}\sum_{j=1}^{n}\left(\hat{\mu}_h(x_j,h)-\mu(x_j)\right)^2 \qquad (3.20)$$

converges in probability to zero. Wong (1983) has shown that if the x_j are uniformly spaced, the selection of the smoothing parameter h using the Cross validation method will produce consistent estimators.

3.5- Spectral density estimation of signals with missing values using the Kernel smoothing approach:

In case where we have a realization of a process Y with missing values, then the sequence $\{x_1, x_2 ....., x_N\}$ will not be necessarily equidistant. For this, we will use the estimator given by (3.7.c) to estimate the missing observations of our realization. Knowing that the most crucial point in using the kernel estimator, is the selection of the smoothing parameter h, we use the Cross Validation method described in the previous section to improve the properties of the spectral density estimate. Assume that we have a realization $\{y(n)\}$, $n = t_1, t_2, ...t_N$ with missing observations at some t_j's. We will use the kernel estimator to estimate the missing observations at t_j by considering the realization with missing observations as a realization with non-equidistant design points x_j simply equal to the available t_j's (The missing t_j's are simply ignored). Once the non-parametric estimator $\hat{\mu}(.)$ is obtained, we estimate the missing observation y_j at t_j by:

$$\hat{\mu}(t_j) = \sum_{i=1}^{N} K\left(\frac{t_j - t_i}{h}\right) y_i \Big/ \hat{f}(t_j) \qquad (3.21)$$

$$\hat{f}(t_j) = (Nh)^{-1} \sum_{i=1}^{N} K\left(\frac{t_j - t_i}{h}\right) \tag{3.22}$$

Once the missing values are estimated, one can estimate the spectral density of the new realization without missing values by using the spectral density estimator given by the windowed autocorrelation function defined in the first chapter by equation (1.41):

$$\hat{P}_y(f) = \sum_{k=-(N-1)}^{N-1} w[k]\hat{r}_{yy}[k]\exp(-j2\pi fk) \tag{3.23}$$

where $\hat{r}_{yy}(k)$ is the sample autocorrelation function of the augmented realization $\{y(n)\}$ defined by:

$$\hat{r}_{yy}(k) = \frac{1}{N} \sum_{n=0}^{N-1-|k|} y[n]y[n+k] \qquad k = 0, \pm 1, \ldots . \pm N\text{-}1 \tag{3.24}$$

and $w(k)$ is the hamming window defined by:

$$w(k) = \begin{cases} \left(0.54 + 0.46\cos\left(\pi \frac{k}{M}\right)\right) & ,|k| \le M \\ 0 & ,|k| > M \end{cases} \tag{3.25}$$

As for the Kalman based estimator, the statistical properties of the estimator $\hat{P}_y(f)$ are inherited from the nice asymptotic properties of the Kernel method of Curve Fitting.

3.6- Illustrative example:

In order to illustrate the performance of the kernel approach in spectral density estimation, we simulated a series of 128 observations from the model:

$$Y(n) = 0.75Y(n-1) - 0.3Y(n-2) + \varepsilon_n$$

where ε_n is a gaussian white noise series with zero mean and constant variance $\sigma^2 = 1$.

For comparative purposes, the estimated spectral density using the complete series and the amplitude-modulated, the kalman, and the kernel smoothing approaches are shown in Figures 3.1-3.3 for 10, 20 and 40 missing values respectively.

These figures indicate that the kernel approach performs better than the amplitude-modulated approach does. It follows more precisely the shape of the spectral density of the original signal than what does the amplitude-modulated signal. It is almost equivalent in performance to the kalman-based approach when the true parameters are used.

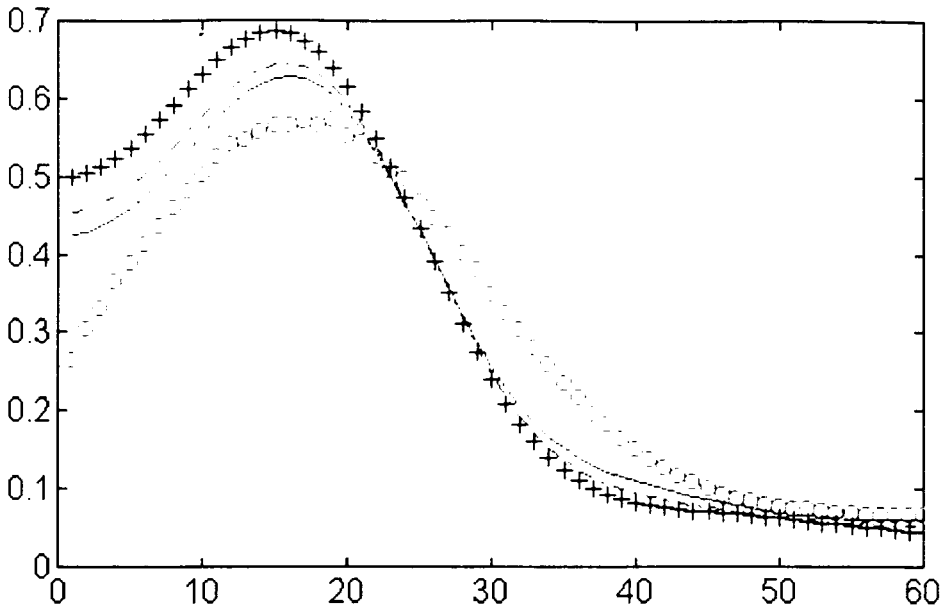

Figure 3.1: Spectral density using the original time series (x), Kalman approach spectral estimator with 10 missing values (-.), amplitude-modulated based spectral estimator with 10 missing values (o), Kernel approach spectral estimator with 10 missing values (-)

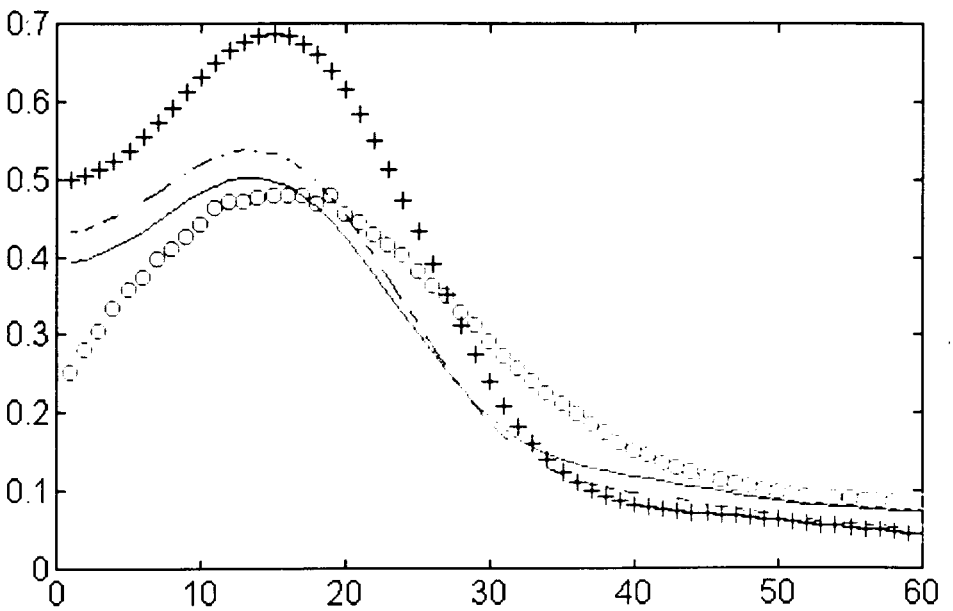

Figure 3.2: Spectral density using the original time series (x), Kalman approach spectral estimator with 20 missing values (-.), amplitude-modulated based spectral estimator with 20 missing values (o), Kernel approach spectral estimator with 20 missing values (-).

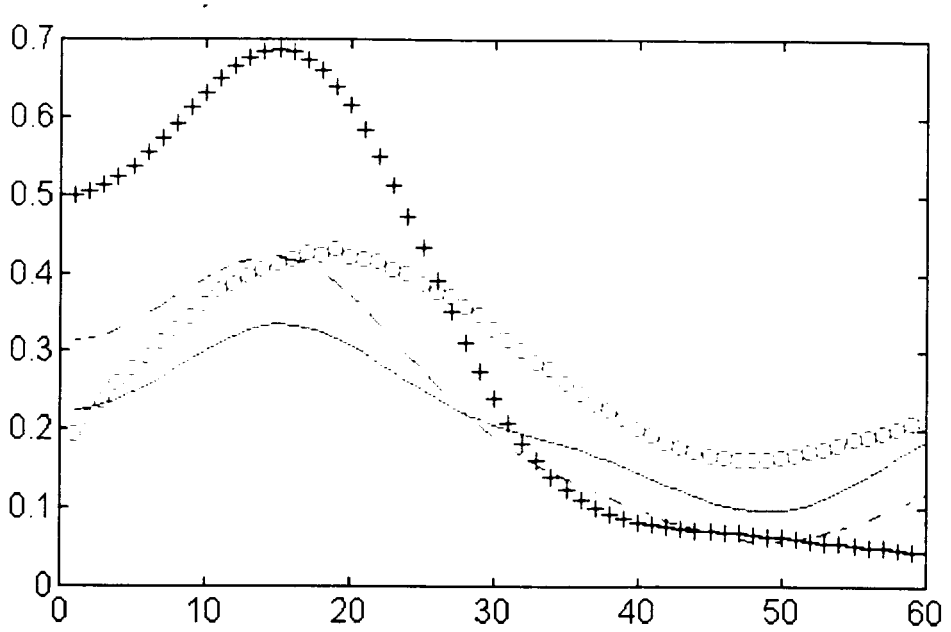

Figure 3.3: Spectral density using the original time series (x), Kalman approach spectral estimator with 40 missing values (-.), amplitude-modulated based spectral estimator with 40 missing values (o), Kernel approach spectral estimator with 40 missing values (-).

Chapter 4

Numerical experiments

4.1- Introduction:

In this chapter, we present the results of intensive simulation studies which had been devised for comparing the performances of the three spectral estimators using incomplete simulated series. To investigate the effect of the model of the generating process, and the number of missing values, we simulated 100 realizations of 128 observations from each of the following models:

$$y_n = 0.75y_{n-1} - 0.3y_{n-2} + \varepsilon_n \tag{4.1}$$

$$y_n = 0.75y_{n-1} - 0.5y_{n-2} + \varepsilon_n \tag{4.2}$$

$$y_n = 0.75y_{n-1} + 0.6y_{n-2} + 0.2y_{n-3} + 0.3y_{n-4} + \varepsilon_n \tag{4.3}$$

$$y_n = 0.5y_{n-1} - 0.25y_{n-2} + \varepsilon_n - 1.1\varepsilon_{n-1} + 0.24\varepsilon_{n-2} \tag{4.4}$$

where ε_n is a zero-mean gaussian noise with variance σ^2.

We considered the cases of 10, 20, and 40 missing observations and σ^2=0.25, 1, and 4. For each simulated series, we computed the spectral density estimators $f_{ORG}(.)$, $f_{AM}(.)$, $f_{KAL}(.)$, $f_{KER}(.)$ where:

$f_{ORG}(.)$ is the Blakman-Tukey estimator (1.41) computed using the complete realization, $f_{AM}(.)$ is the estimator of Toloi and Morettin given in (2.7), and $f_{KAL}(.)$, $f_{KER}(.)$ are the proposed estimators given in (2.56), (3.27) where the missing observations are estimated using the Kalman

smoother, and the nonparametric Kernel method of curve fitting respectively.

The comparison of these methods was based on the average mean square error defined by:

$$MSE_{AM}(i) = \frac{1}{N}\sum_{n=1}^{N}\left(f_{AM}(f_n) - f_{ORG}(f_n)\right)^2$$

$$MSE_{KAL}(i) = \frac{1}{N}\sum_{n=1}^{N}\left(f_{KAL}(f_n) - f_{ORG}(f_n)\right)^2 \quad \text{for} \quad i = 1,...,100$$

$$MSE_{KER}(i) = \frac{1}{N}\sum_{n=1}^{N}\left(f_{KER}(f_n) - f_{ORG}(f_n)\right)^2$$

In addition to that, we computed the confidence interval $[a,b]$ for each method where a, and b are given by:

$$a_{AM} = \overline{MSE}_{AM} - 2\sqrt{Var_{AM}}$$
$$b_{AM} = \overline{MSE}_{AM} + 2\sqrt{Var_{AM}}$$

$$a_{KAL} = \overline{MSE}_{KAL} - 2\sqrt{Var_{KAL}}$$
$$b_{KAL} = \overline{MSE}_{KAL} + 2\sqrt{Var_{KAL}}$$

$$a_{KER} = \overline{MSE}_{KER} - 2\sqrt{Var_{KER}}$$
$$b_{KER} = \overline{MSE}_{KER} + 2\sqrt{Var_{KER}}$$

where $\overline{MSE}$ denotes the mean of $\{MSE(i)\}$ $i = 1,...100$ and

$$Var_{AM} = \frac{1}{100}\sum_{n=1}^{100}\left(MSE_{AM}(n) - \overline{MSE}_{AM}\right)^2$$

$$Var_{KAL} = \frac{1}{100}\sum_{n=1}^{100}\left(MSE_{KAL}(n) - \overline{MSE}_{KAL}\right)^2$$

$$Var_{KER} = \frac{1}{100}\sum_{n=1}^{100}\left(MSE_{KER}(n) - \overline{MSE}_{KER}\right)^2$$

4.2- Presentation of the results of simulation study 1:

We simulated a series of 128 observations from the second order autoregressive models given in (4.1) and (4.2). The computed estimators are shown in Figures 4.1-4.3 and 4.4-4.6. They indicate that both the kalman approach estimator $f_{KAL}(.)$, and the kernel approach estimator $f_{KER}(.)$perform better than the amplitude-modulation approach estimator $f_{AM}(.)$. They follow exactly the shape of the spectral density of the original time series with no missing values. This is confirmed by the results of table 4.1 which were obtained from the computation of the mean square error and the confidence interval computed from 100 simulated realizations from model (4.1). We notice that the Kernel Curve Fitting approach is almost equivalent to the kalman smoother approach using the true parameters, when the optimal smoothing parameter h is computed using the cross-validation method criterion. This table shows that the performance of the three methods decreases as the number of missing values increases. It also indicates that the variance of the noise has an influence on the precision of the estimators as expected from the asymptotic results.

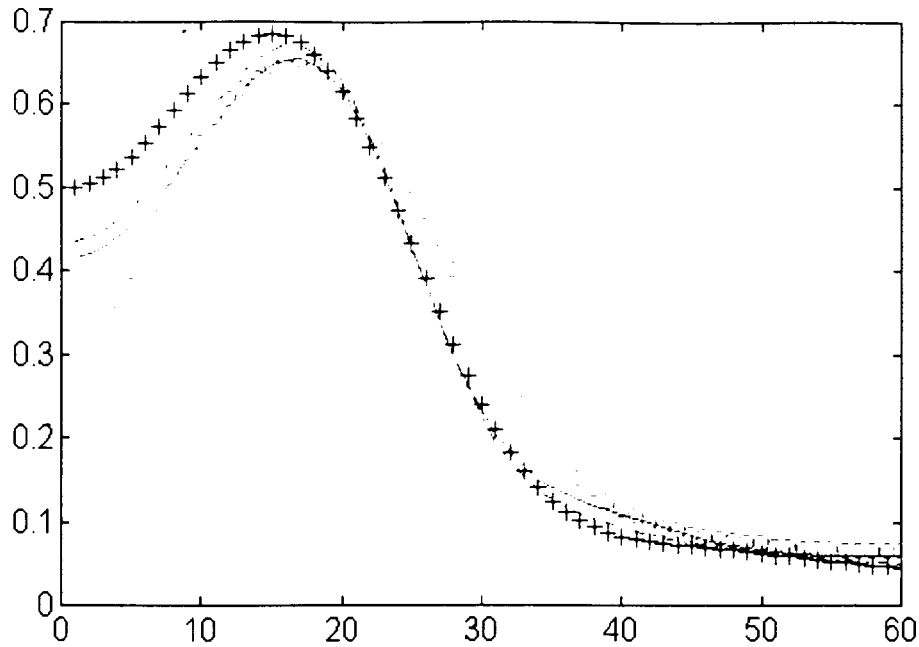

Figure 4.1: Spectral density using the original time series (x), Kalman approach spectral estimator with 10 missing values (-.), amplitude-modulated based spectral estimator with 10 missing values (o), Kernel approach spectral estimator with 10 missing values (-)

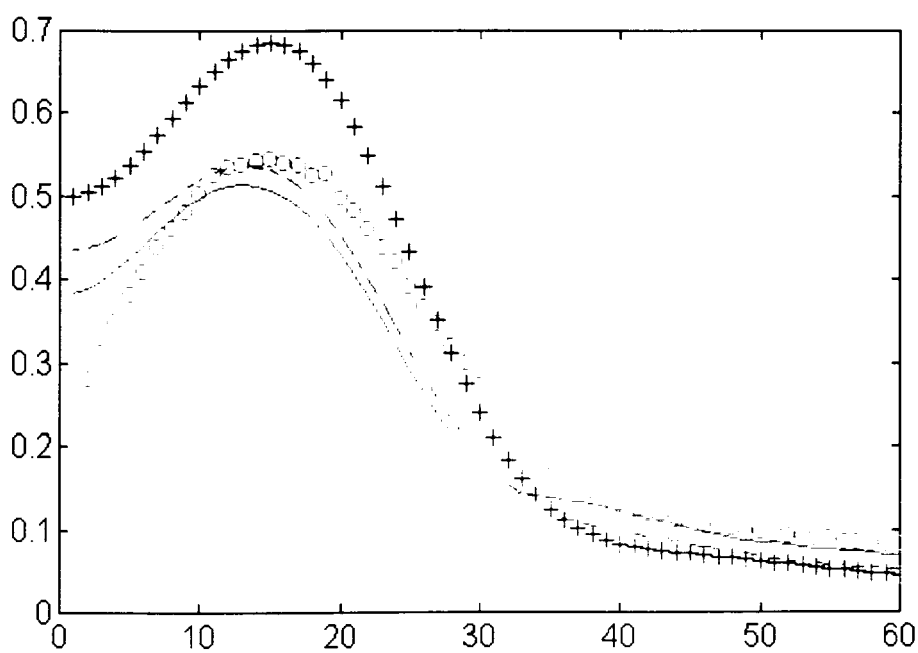

Figure 4.2: Spectral density using the original time series (x), Kalman approach spectral estimator with 20 missing values (-.), amplitude-modulated based spectral estimator with 20 missing values (o), Kernel approach spectral estimator with 20 missing values (-).

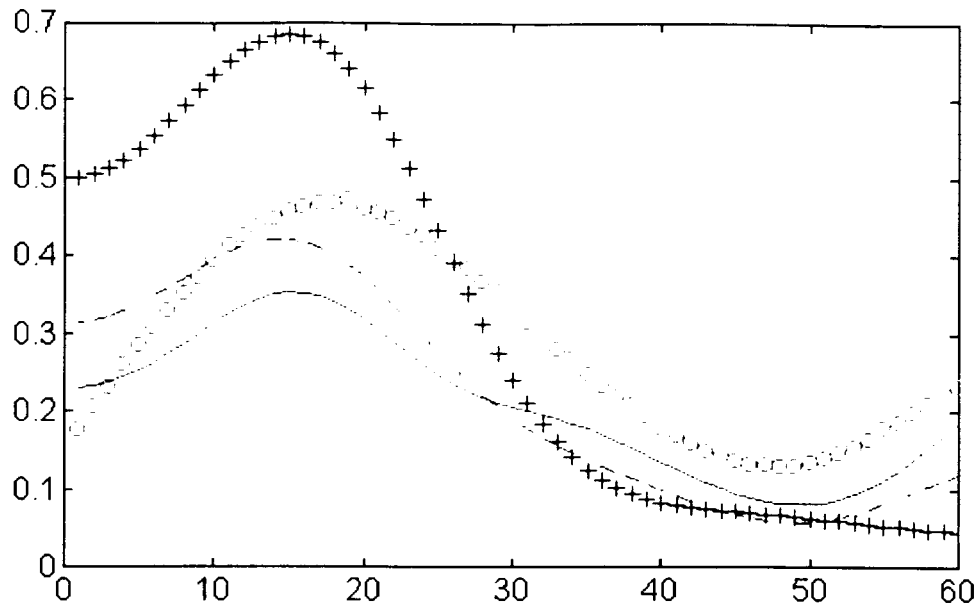

Figure4.3: Spectral density using the original time series (x), Kalman approach spectral estimator with 40 missing values (-.), amplitude-modulated based spectral estimator with 40 missing values (o), Kernel approach spectral estimator with 40 missing values (-).

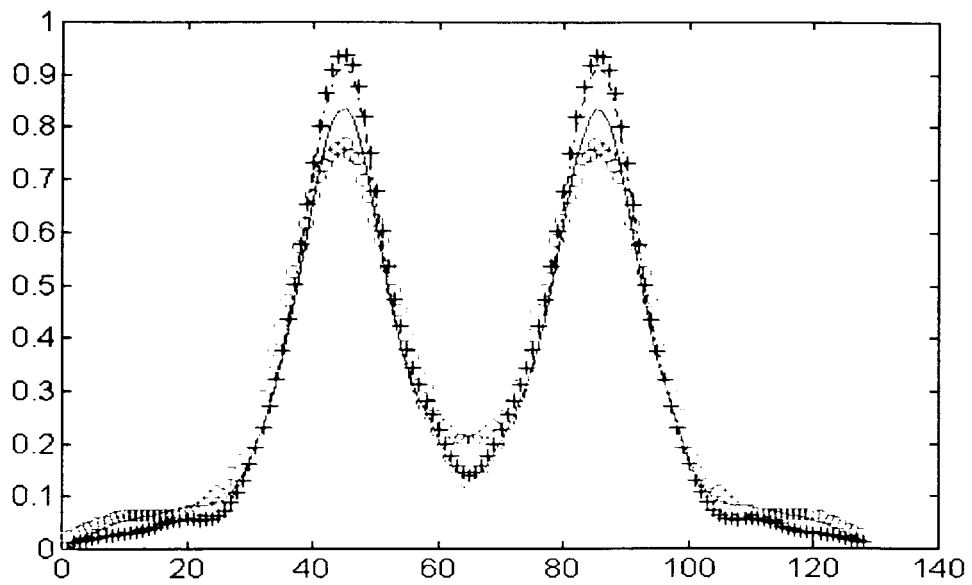

Figure 4.4: Spectral density using the original time series (x), Kalman approach spectral estimator with 10 missing values (-.), amplitude-modulated based spectral estimator with 10 missing values (*), Kernel approach spectral estimator with 10 missing values (-).

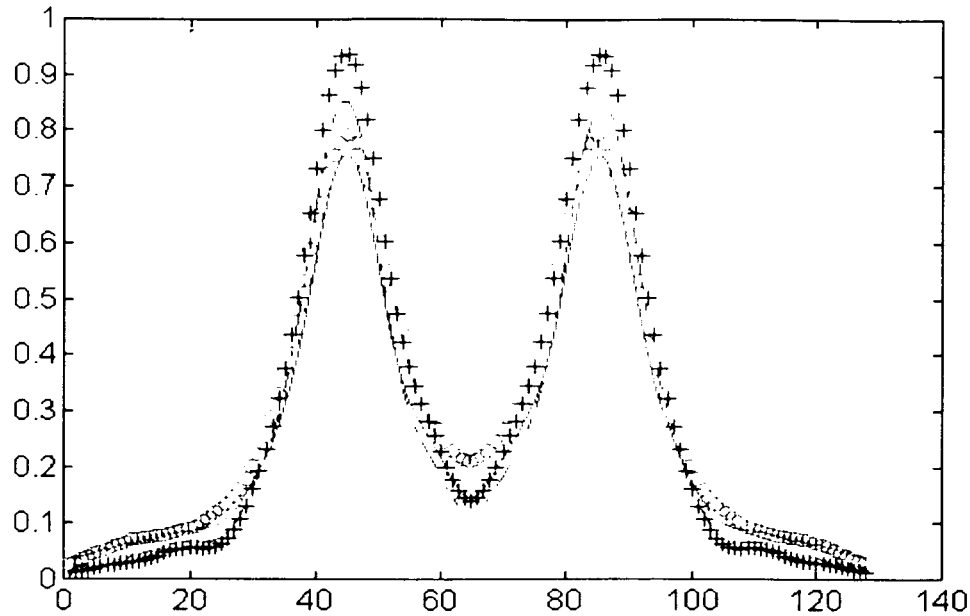

Figure 4.5: Spectral density using the original time series (x), Kalman approach spectral estimator with 20 missing values (-.), amplitude-modulated based spectral estimator with 20 missing values (o), Kernel approach spectral estimator with 20 missing values (-).

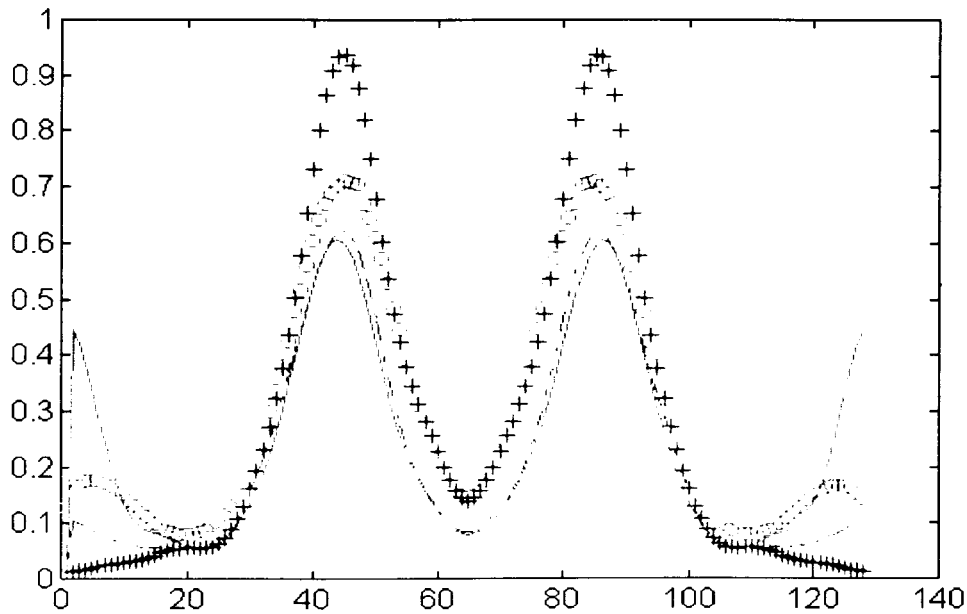

Figure 4.6: Spectral density using the original time series (x), Kalman approach spectral estimator with 40 missing values (-.), amplitude-modulated based spectral estimator with 40 missing values (*), Kernel approach spectral estimator with 40 missing values (-).

Missing Values	Variance	Average Mean square error			Confidence Interval					
		MSE_{KAL}	MSE_{KER}	MSE_{AM}	A_{KAL}	b_{KAL}	a_{KER}	b_{KER}	a_{AM}	b_{AM}
10 Missing Values	$(0.5)^2$	0.0001	0.0002	0.0007	-0.00011	0.00029	-0.0002.1	0.00048	-0.00025	0.0015
	1	$7.8665e^{-4}$	0.0013	0.0078	-0.0018	0.0048	-0.0035	0.0077	-0.0040	0.024-6
	$(2)^2$	0.0232	0.0355	0.1734	-0.0286	0.0756	-0.0556	0.1234	-0.0659	0.3934
20 Missing Values	$(0.5)^2$	0.000162	0.00027	0.00042	0.00046	0.00059	-0.000014	0.00059	-0.000015	0.00088
	1	0.0054	0.0046	0.0117	-0.0011	0.0071	-0.0017	0.0107	-0.0037	0.0244
	$(2)^2$	0.0658	0.1106	0.2234	-0.0208	0.1246	-0.0318	0.1804	-0.0496	0.3742
40 missing values	$(0.5)^2$	0.0010	0.0017	0.0015	-0.00012	0.0022	-0.00031	0.0031	-0.00053	0.0030
	1	0.0192	0.0266	0.0216	-0.0017	0.0375	-0.0031	0.0504	-0.0044	0.0591
	$(2)^2$	0.2061	0.3729	0.3572	0.0768	0.4932	0.0602	0.7648	-0.0177	0.8353

Table 4.1: Average Mean Square Error and Confidence Interval for the AR(2) process $y_n = 0.75y_{n-1} - 0.3y_{n-2} + \varepsilon_n$

4.3- Presentation of the results of simulation study 2:

In this study, we simulated a series of 128 observations from the models (4.3) and (4.4). The estimated spectral densities for 10, 20 and 40 missing values are shown in Figure 4.7-4.9, and 4.10-4.12 respectively. As in the previous simulation study, we notice that the proposed methods perform better than the one based on amplitude-modulated signals. Our study suggests that the kernel approach gives better results than that of the kalman approach when the smoothing parameter is correctly estimated.

We also computed the mean square error, and the associated confidence interval using 100 realizations from model (4.3), and (4.4) by considering 10, 20 and 40 missing values and different values for the variance σ^2 of the noise. The results are obtained in tables 4.2 , and 4.3 respectively which confirm the results of the previous section.

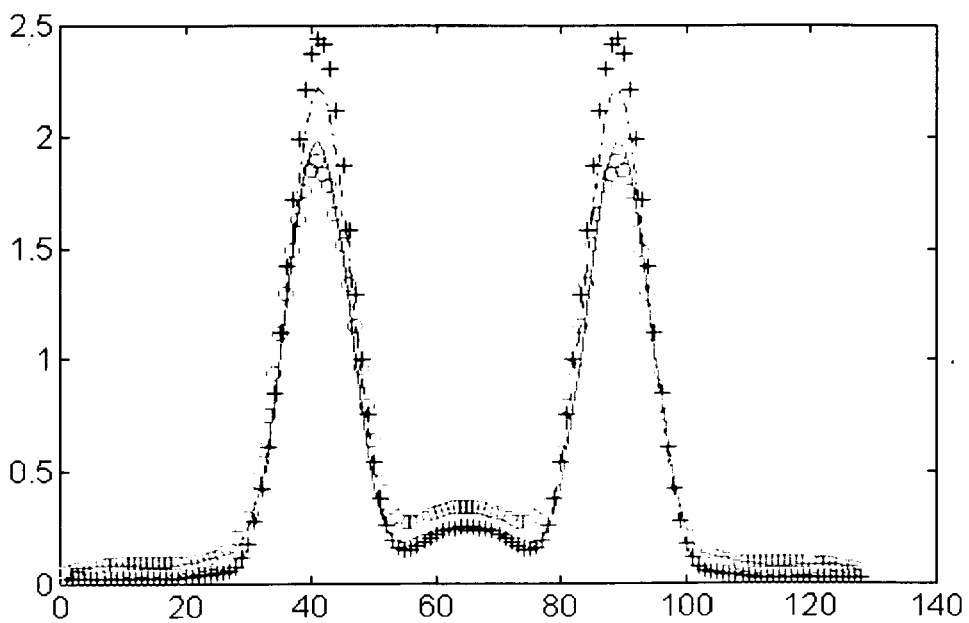

Figure 4.7: Spectral density using the original time series (x), Kalman approach spectral estimator with 10 missing values (-.), amplitude-modulated based spectral estimator with 10 missing values (*), Kernel approach spectral estimator with 10 missing values (-).

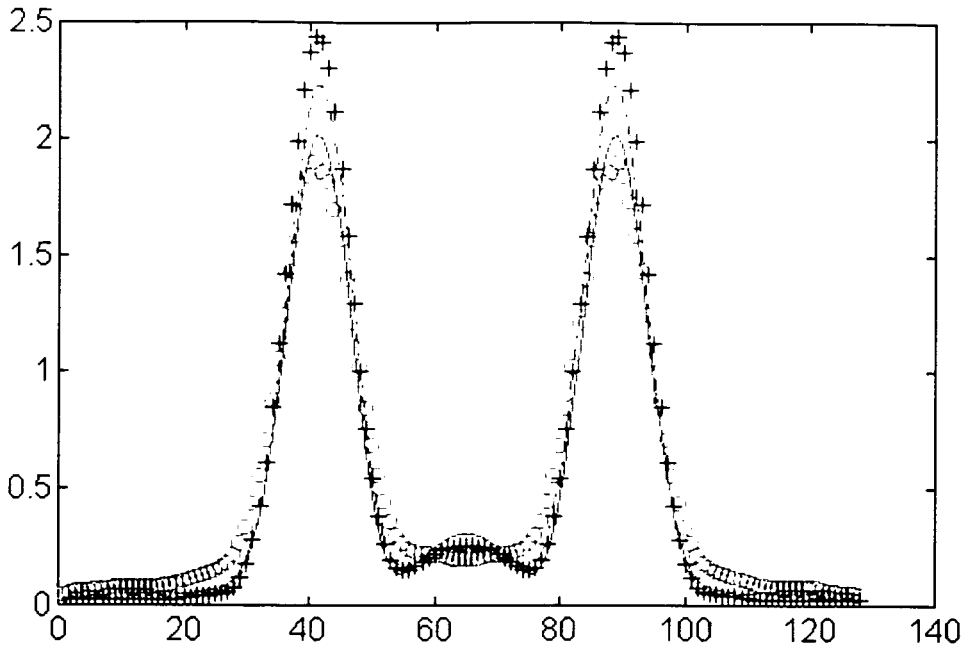

Figure 4.8: Spectral density using the original time series (x), Kalman approach spectral estimator with 20 missing values (-.), amplitude-modulated based spectral estimator with 20 missing values (o), Kernel approach spectral estimator with 20 missing values (-).

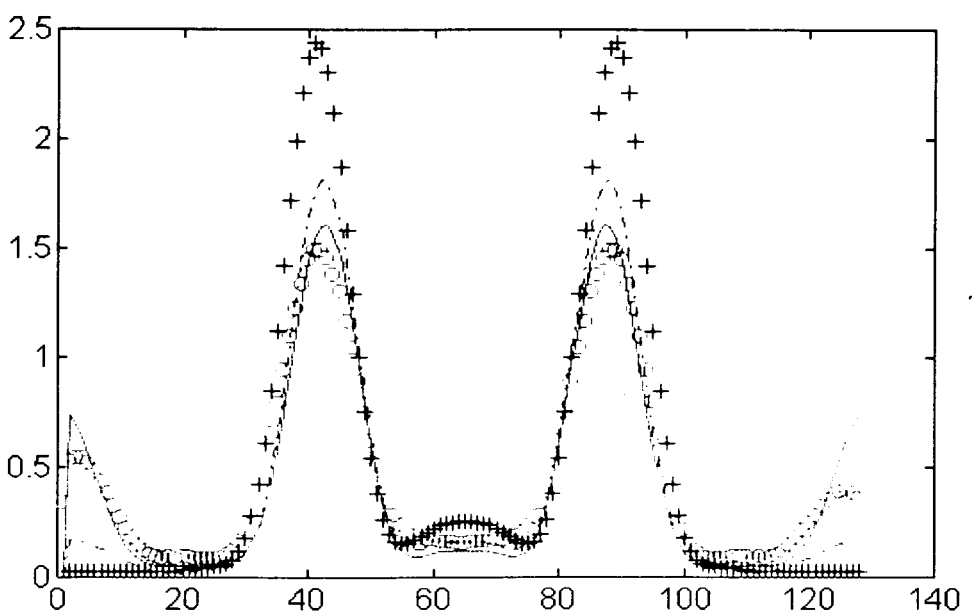

Figure 4.9: Spectral density using the original time series (x), Kalman approach spectral estimator with 40 missing values (-.), amplitude-modulated based spectral estimator with 40 missing values (*), Kernel approach spectral estimator with 40 missing values (-).

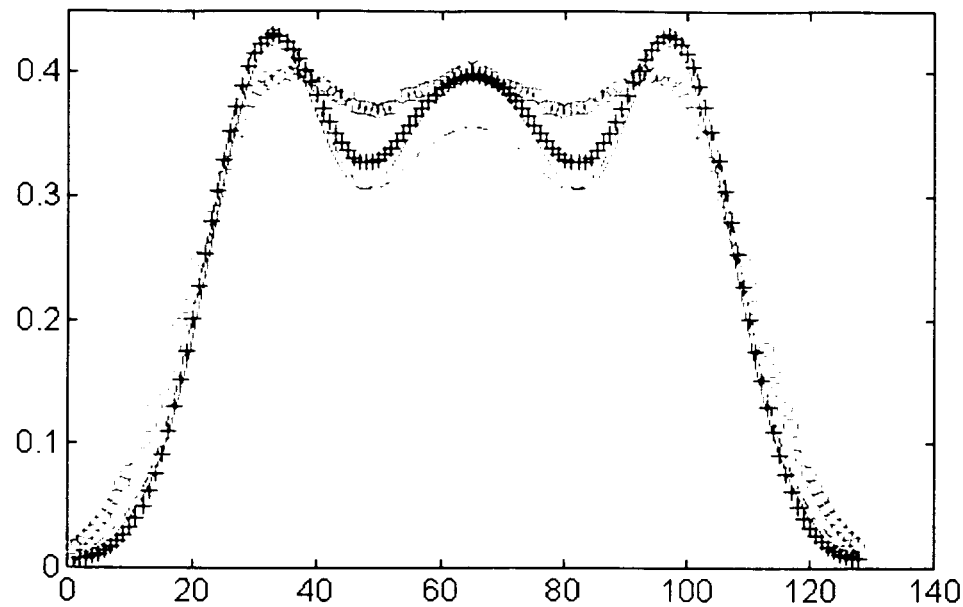

Figure 4.10: Spectral density using the original time series (x), Kalman approach spectral estimator with 10 missing values (-.), amplitude-modulated based spectral estimator with 10 missing values (o), Kernel approach spectral estimator with 10 missing values (-).

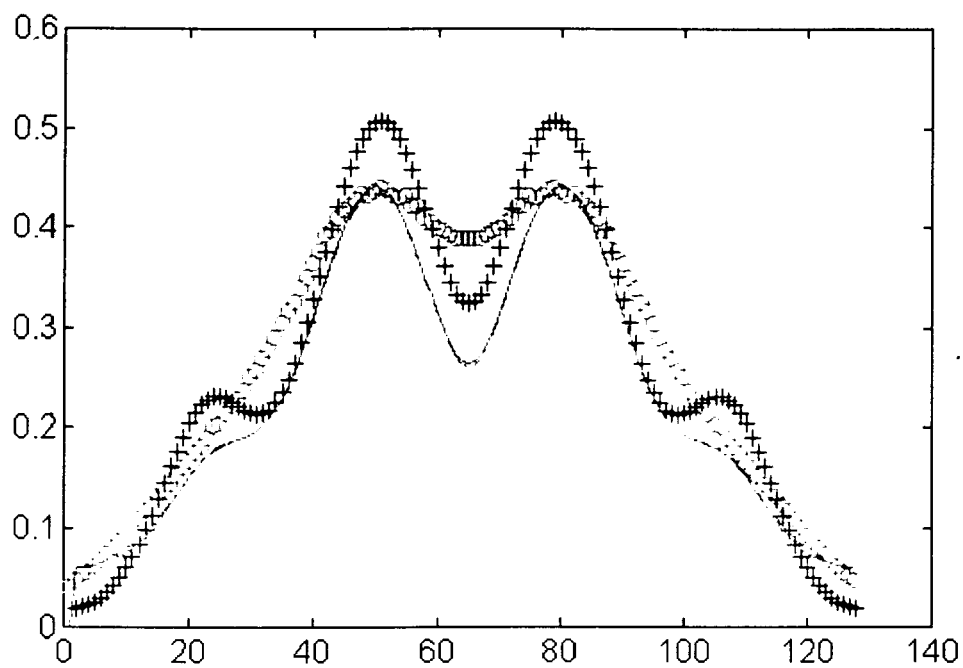

Figure 4.11: Spectral density using the original time series (x), Kalman approach spectral density estimator with 20 missing values (-.), amplitude-modulated based spectral density estimator with 20 missing values (o), Kernel approach spectral density estimator with 20 missing values (-).

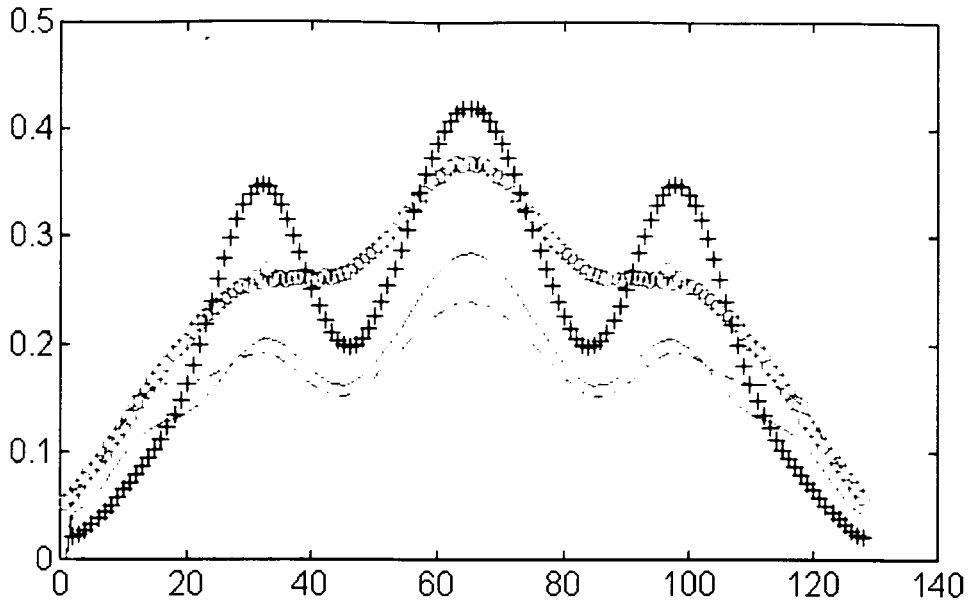

Figure 4.12: Spectral density using the original time series (x), Kalman approach spectral estimator with 40 missing values (-.), amplitude-modulated based spectral estimator with 40 missing values (o), Kernel approach spectral estimator with 40 missing values (-).

Missing Values	Variance	Average Mean Square Error			Confidence Interval					
		MSE_{KAL}	MSE_{KER}	MSE_{AM}	a_{KAL}	b_{KAL}	a_{KER}	b_{KER}	a_{AM}	b_{AM}
10 Missing Values	$(0.5)^2$	0.0005	0.0013	0.0015	-0.0006	0.0016	-0.0010	0.0035	-0.0058	0.0228
	1	0.0143	0.0297	0.0612	-0.0058	0.0230	-0.0014	0.0424	-0.0195	0.0816
	$(2)^2$	0.1474	0.3105	0.3516	-0.0325	0.1751	-0.1288	0.5099	-0.4812	4.9956
20 Missing Values	$(0.5)^2$	0.0014	0.0032	0.0031	-0.0017	0.0054	-0.0072	0.0206	-0.0185	0.0468
	1	0.0413	0.0979	0.0924	-0.0355	0.1361	-0.1002	0.3994	-0.0459	0.2378
	$(2)^2$	0.4671	1.4185	1.0167	-0.4856	1.4435	-2.2032	5.0187	-2.9507	9.7431
40 missing values	$(0.5)^2$	0.0180	0.0378	0.0150	-0.0219	0.0486	-0.0795	0.1624	-0.0136	0.0524
	1	0.2239	0.3058	0.2312	-0.1105	0.4255	-0.3163	1.0785	-0.0406	0.1667
	$(2)^2$	3.5421	8.5456	2.4369	-2.8071	7.9740	-4.1179	15.8998	-2.2316	8.9463

Table 4.2: Average Mean Square Error and confidence interval for the AR(4) process

$$y_n = 0.75y_{n-1} + 0.6y_{n-2} + 0.2y_{n-3} + 0.3y_{n-4} + \varepsilon_n$$

Missing Values	Variance	Average Mean square error			Confidence interval					
		MSE_{KAL}	MSE_{KER}	MSE_{AM}	a_{KAL}	b_{KAL}	a_{KER}	b_{KER}	a_{AM}	b_{AM}
10 Missing Values	$(0.5)^2$	0.000064	0.000058	0.000076	0.000016	0.00012	0.000008	0.00011	0.000012	0.00017
	1	0.000921	0.000816	0.0012	-0.000421	0.0026	-0.00059	0.0028	-0.0014	0.0050
	$(2)^2$	0.0145	0.0160	0.0260	-0.0080	0.0401	-0.0101	0.0440	-0.0351	0.1072
20 Missing Values	$(0.5)^2$	0.00016	0.00017	0.00011	-0.000091	0.00042	-0.0000013	0.00027	-0.000012	0.00017
	1	0.0021	0.0022	0.0020	0.000039	0.0034	-0.00038	0.0041	-0.0012	0.0057
	$(2)^2$	0.0247	0.0263	0.0233	-0.0039	0.1182	0.00077	0.0888	0.0013	0.0648
40 missing values	$(0.5)^2$	0.00068	0.00050	0.00024	0.00016	0.0013	-0.000084	0.0014	-0.000055	0.00066
	1	0.0093	0.0080	0.0035	-0.0031	0.0283	-0.0015	0.0229	-0.0022	0.0102
	$(2)^2$	0.1727	0.1582	0.0614	0.0504	0.2992	-0.0348	0.4128	-0.0574	0.2970

Table 4.3: Average Mean Square Error and confidence interval for the *ARMA(2,2)* process

$$y_n = 0.5y_{n-1} - 0.25y_{n-2} + \varepsilon_n - 1.1\varepsilon_{n-1} + 0.24\varepsilon_{n-2}$$

4.4- Estimation of the parameters of AR(2) model from incomplete realizations:

In this case study, we investigated the performance of our nonparametric approach in the estimation of the parameters of the parametric model from incomplete realizations. As for the spectral density estimation, one can estimate the parameters using any identification method once the missing values are estimated using the kernel method of Curve Fitting. We simulated 100 realizations of 128 observations from model (4.1) with $\sigma^2 = 1$ and by considering 10, and 20 missing values. For each realization, we estimated the missing values from the Fitted nonparametric regression curve $\hat{\mu}(.,h)$, and then we estimated the parameters of the *AR(2)* model using the built-in Matlab identification procedures.

The mean and the mean square error of the parameters a_1 and a_2 of the model are $\overline{a}_{1KER} = -0.6696$ and $\overline{a}_{2KER} = 0.2828$; and $\overline{MSEa}_{1KER} = 0.00061$ and $\overline{MSEa}_{2KER} = 0.0017$ respectively when there are 10 missing values. They indicate that the confidence intervals include the true values of the parameters. When the number of missing values is increased to 20 observations, the mean and the mean square error of the parameters a_1 and a_2 of the model are $\overline{a}_{1KER} = -0.5505$ and $\overline{a}_{2KER} = 0.1122$; and $\overline{MSEa}_{1KER} = 0.0082$ and $\overline{MSEa}_{2KER} = 0.0150$ respectively. Furthermore, we noticed that the confidence intervals are bigger when the number of missing values is increased.

Conclusion

In this thesis, we investigated the problem of spectral density estimation of random signals from observed realizations with missing observations. When some observations are missing (nonavailable), we can not use the popular approach which is based on Fourier Transform since it requires equidistant sampled observations. The main objective of our research work is to develop methods so that the classical Fourier approach can be used even when the realization is incomplete. In the firs part of our investigation, we considered the method of spectral density estimation using amplitude-modulated signals which was proposed by Toloi and Morettin (1993). Despite its theoretcal properties, this estimator is derived by scaling the spectral density estimator which is computed by replacing the missing values by zeros. This drawback motivated us to consider the possibility of estimating the missing values from the available observations. For instance, we investigated the use of Kalman smoother algorithm to estimate the missing values. This suggestion is motivated by Vaccaro and Fu Li (1990) which states that the kalman smoother is powerfull even when the size of data is relatively short. Our simulation results show that this approach is encouraging even though it requires the knowledge of some of the parameters of the generating process. This drawback motivated us to estimate the missing values using the kernel nonparametric method of Curve Fitting which was applied by Kim and Cox (1996) in Time series modeling. This approach which is a data-based approach does not require prior knowledge on the generating process. The performance of this approach is investigated using simulated series. Our simulation studies indicate that it performs almost as good as the Kalman's approach when the true parameters of the model are used.

The last simulated case study of cahpter 4 suggests that this approach is not limited to spectral density estimation but it can also be used in system identification since the estimated parameters of the process which were computed from the original sample are very close to those which were computed once the missing values were replaced by their nonparametrically predicted values.

As a further research work, it is interesting to investigate the possibility of using Bootstrap techniques to improve the performance of the nonparametric approach for the cases where the number of missing values is relatively large. In this approach the missing values will not be estimated from the kernel regression curve $\hat{\mu}(.,h)$ but from the average of B simulated regression curves.

References:

- Ables J.G., Maximum entropy spectral analysis, (Astronomy and Astrophisics supplemnet series, June 1974.

- Balckman R.B. and Tukey J.W., "Measurement of power spectra from the point of view of communications engineering", New York, Dover, 1959.

- Bingham C., Frey M.D. God, and Tukey J.W, Modern Techniques of Spectral Estimation, IEEE Transactions on Audio and Electroacoustics, June 1967.

- Burg, J.P., "Maximum entropy spectral analysis", paper presented at the 37th annual international meeting, Soc. of Explor. Geophysics, Oklahoma city. Okla Oct 31, 1967.

- Burg, J.P., "New analysis technique for time series data", paper presented at Advanced Study Institute on Signal Processing, NATO. Enschede, Netherlands 1968.

- Caines Peter E., "Linear stochastic systems", 1988.

- Craven, P. and Wahba G., "Smoothing noisy data with spline functions:estimating the correct degree of smoothing by method of cross-validation", Numer. Math. 31, 377-403, 1979.

- Cooley, J.W, and Tukey, J.W, "An algorithm for the machine calculations of complex Fourier Series", Math. of Comp., v.19. p.297-301, 1965.

- Dunsmir W., Robinson P.M., "Asymptotic theory of Time series containing missing and amplitude modulated observations", Sankhya, Ser. A. pp260-281, 1981.

- Dunsmir W., Robinson P.M., "Parametric estimators for stationary time series with missing values", Advances in Applied Probability, 13, pp 129-146, 1981.

- Gasser Th. and Muller H.G., "Kernel estimation of regression functions in smoothing techniques for curve estimation (Th. Gasser and M. Rosenbalt, eds), 23-68, 1979.

- Grewal Mohinda, S., Andrews Angus P., "Kalman filtering, theory and practice", Prentice hall, 1993.

- Gutowski Paul R., Robinson Enders A., and Treitel Sven, "Spectral estimation: Fact or fiction", IEEE transactions Geosci. Electron. , vol. GE-16, pp80-84. April 1978.

- Haddad Richard A., "Digital Signal Processing, theory applications and hardware", 1991.

- Johansson Rolf, "System modeling and identification", Prentice Hall, 1993.

- Kay Steven M., and Marple Stanley Lawrence, JR., "Spectrum Analysis, A modern Perspective", Proceeding of the IEEE, Vol. 69, NO. 11, November 1981.

- Kay Steven M., "Modern Spectral estimation, Theory and application", Prentice hall, 1988.

- Lacoss R.T, "Data Adaptive Spectral Analysis methods", Geophysics, vol. 36, pp.661-675, Aug. 1971.

- Ladjouze Salim, "Problemes d'estimation dans les series temporelles stationnaires avec donnees manquantes", Thesis presented at 'L'universite Scientifique et medicale de Grenoble', 1986.

- Lewis Frank L., "Optimal estimation with an introduction to stochastic control theory", 1986.

- Nadaray, E. A., "On estimating regression", Theor. Probab. Appl. 9, 141-142, 1964.

- Parzen, E., "On spectral analysis with missing observations and amplitude modulation", Sankhya Ser. A, pp. 383-392, 1981.

- Parzen E., "Statistical spectral analysis methods", Stanford Univ., Stanford, Calif., Tech Rep. 11 on Contract Nonr-225(80), June 10, 1968.

- Priestly, M. B. and Chao M. T. "Nonparametric function fitting", J. Roy. Statist. Soc. B 34, (1972)

- Srinath M. D., Rajasekaran P. K., "An introduction to statistical signal processing with applications", 1979.

- Kim T. Y. and Cox Dennis D., "Bandwidth selection in Kernel smoothing of time series", Journal of TIME SERIES Vol. 17, No. 1, (1996).

- Titterinton, D. M. "Common structure of smoothing thecniques in statistics", Intern. Statist. Rev. 53, 141-170. (1985).

- Toloi Clelia M. C. and Morettin Pedro A., "Spectral analysis of amplitude modulated time series", Journal of TIME SERIES Vol. 14, No. 4, (1993).

- Ulrych Tad J., Bishop Thomas N., "Maximum Entropy Spectral Analysis and Autoregressive decomposition", Rev. Geophysics and Space Phys., vol. 13 pp. 183-200, Feb 1975.

- Ulrych Tad J., "Maximum entropy power spectrum of truncated sinusoids", J. Geophysics. Res., 77, 1396-1400, 1972.

- Vasegin Saeed., "Advanced signal processing and digital noise reduction", Wiley Teubner communications , 1996.

- Vaccaro Richard J. and Fu Li., "Signal estimation over short data records: A data-dependent time-invariant algorithm:, Journal of the Franklin Institute Vol. 327, No. 3, pp. 439-455, 1990.

- Watson, G. S., "Smooth regression Analysis", Sankhaya A26, 359-372, (1964)

- Welch P.D., "The use of fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short modified periodograms", IEEE transactions on Audio and Electroacoustics, June 1967.