



OPTIMAL PARAMETRIC MODELLING OF MEASURED SHORT WAVES

Sukomal Mandal

Assistant Director, Ocean Engineering Division, National Institute of Oceanography, Dona Paula, Goa-403 004, India.

SYNOPSIS : The spectral analysis of measured short waves can efficiently be carried out by the fast Fourier transform technique. Even though many present techniques can be used for the simulation of time series waves, these may not provide accurate and computationally efficient results for specific applications. Therefore, an alternate method is described here to provide spectral relationships with minimum loss of information. The measured short waves are expressed by the parametric models namely, autoregressive or auto-regressive moving average algorithms which are the polynomial functions. Further reduced order autoregressive moving average spectral estimations are successfully applied to time series waves with different sampling rates. This paper emphasises the importance of selecting a suitable sampling interval for better estimates of parametric modelling and also for better statistical representation. Implementation of the above algorithms in a structural monitoring system has the potential advantage of storing more representations of wave environment than is currently possible because of the reduced size of the data sets.

INTRODUCTION

It is often necessary that in the absence of measured time series short waves, theoretical power spectra (Pierson-Moskowitz, JONSWAP, etc.) are considered for various offshore applications. Even though many present techniques are used for the simulation of time series waves [Borgman (1969), Tucker, Challenor and Carter (1984)], these may not provide either accurate or computationally efficient results for specific offshore applications. Presently available methods (digital time series or spectral values) can not store short wave information for long periods in remote places. All these facts have led to the development of parametric modelling.

Various parametric system identification techniques are applied to ocean wave dynamic systems. Spanos and Hansen (1981) used maximum entropy method which was referred by the authors as Linear Prediction Theory (LPT). This is an autoregressive (AR) algorithm for digital simulation of sea waves and can be used as an efficient alternate method. They used Pierson-Moskowitz spectrum as the target spectrum which can be obtained as the output of the AR model. Time series waves were determined as the output of the recursive digital filter to a white noise input. Although AR spectral estimates fluctuated, in an average sense the AR model approximated reliably the Pierson-Moskowitz spectrum. It was emphasized that the numerical studies were considered as preliminary, and the important aim of the study was to make potential usefulness of AR algorithm for ocean engineering applications. Even though Fines *et al* (1981) described autoregressive moving average (ARMA) algorithm, they have made some statistical studies applying AR algorithm to the time series waves measured off Norwegian coast. Medina and Sanchez-Carratala (1991) established a robust AR model of JONSWAP wave spectrum, where the addition of white noise with a variance of $0.0025m_0$ (m_0 is the spectral moment) makes an acceptable fitting of the AR spectrum to JONSWAP spectrum. Witz and Mandal (1991)

modified the LPT method and then applied to time series for offshore dynamic processes. Mandal *et al* (1992) had showed that the initial accurate AR estimates would lead to estimate the better and reliable reduced order ARMA model.

In this paper, AR algorithm is formulated based on the least square approach and applied to North Sea time series waves. The time series data set is divided into three data sets with different sampling rates. Then optimal model order selection of the AR model is carried out using Akaike Information Criterion and Minimum Description Length methods, which are derived based on the least squares error estimation. Sampling interval plays an important role for optimal parametric modelling of ocean waves. The spectral comparisons of the FFT and AR models are examined for selecting suitable sampling interval of the time series waves. Then the reduced order ARMA spectral estimation technique is applied to describe the time series with minimum loss of information.

AUTOREGRESSIVE MODEL

The time series $\{y_i\}$ is said to be an autoregressive process, AR(p) of the order p and is generated from the relationship

$$y_r = - \sum_{k=1}^p a_k y_{r-k} + \sqrt{\rho_w} w_r \quad (1)$$

Here y_r is the r th sample of the discrete stochastic process and w_r is the white noise with variance ρ_w .

The transfer function of the AR(p) process can be expressed as

$$H_{AR}(Z) = \frac{\sqrt{\rho_w}}{1 + \sum_{k=1}^p a_k Z^{-k}} \quad (2)$$

Therefore, the estimated power spectrum, $S_{yy}(\omega)$ of the time series y_t can be written as

$$S_{yy}(\omega) = |H_{AR}(e^{i\omega T})|^2 \quad (3)$$

$$= \frac{\rho_w}{|1 + \sum_{k=1}^p a_k e^{-ik\omega T}|^2}$$

Where sampling period, T is generally defined by the cut-off frequency, ω_b , through the Nyquist relationship

$$T = \frac{\pi}{\omega_b}$$

The autocorrelation function, R_λ of the finite set of the time series data, i.e., $\{0 < Y_t < M\}$, can be calculated from the relationships

$$R_\lambda = f(M, \lambda) \sum_{t=\lambda+1}^M Y_t Y_{t-\lambda} \quad (4)$$

where, $f(M, \lambda) = 1/(M - \lambda)$ for the unbiased sample correlations, and $f(M, \lambda) = 1/M$ for the biased sample correlations.

Let ϵ be the error between the target spectrum, $S(\omega)$, and the estimate power spectrum of the AR output, $S_{yy}(\omega)$, is expressed as

$$\epsilon = \frac{b_0^2}{2\omega_b} \int_{-\omega_b}^{\omega_b} \frac{S(\omega)}{S_{yy}(\omega)} d\omega \quad (5)$$

By minimizing ϵ and substituting $S_{yy}(\omega)$ [equation (3)] in above equation, the parameters a_k are determined. Therefore, minimisation of error, ϵ , brings the following important relationship

$$\frac{d\epsilon}{da_i} = 0$$

$$\sum_{k=1}^p a_k R_{|i-k|} = -R_i \quad (6)$$

Equation (6) can be written in Toeplitz matrix form as

$$\begin{bmatrix} R_0 & R_1 & R_2 & \dots & R_{p-1} \\ R_1 & R_0 & R_1 & \dots & R_{p-2} \\ R_2 & R_1 & R_0 & \dots & R_{p-3} \\ \dots & \dots & \dots & \dots & \dots \\ R_{p-1} & R_{p-2} & R_{p-3} & \dots & R_0 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \dots \\ R_p \end{bmatrix} \quad (7)$$

These are also known as Yule-Walker equations (Mandal *et al.*, 1992). Using equations (7), i.e., when the parameters a_1, a_2, a_3, \dots

a_k and noise variance ρ_w are determined, then the time series y_t can be generated digitally by using recursive equation (1).

AUTOREGRESSIVE MOVING AVERAGE MODEL

Let the time series $\{y_t\}$ be an autoregressive moving average (ARMA) process of the orders (p, q) with $p > q$. It is defined by the difference equations

$$y_t = - \sum_{k=1}^p a_k y_{t-k} + \sum_{i=0}^q b_i w_{t-i} \quad (8)$$

Here y_t is the t th sample of the discrete stochastic process and w_t is the Gaussian white noise process.

The ARMA(p, q) process can also be expressed as the output of a digital filter whose transfer function in terms of z -transform is defined as

$$H_{ARMA}(z) = \frac{\sum_{i=0}^q b_i z^{-i}}{1 + \sum_{k=1}^p a_k z^{-k}} \quad (9)$$

The estimated power spectrum, $S_{yy}(\omega)$ is written as

$$S_{yy}(\omega) = |H_{ARMA}(e^{i\omega T})|^2 \quad (10)$$

There are various procedures available to obtain the unknown coefficients a_k and b_i [see Graupe *et al.*, 1975, Spanos, 1983 and Mandal *et al.*, 1992]. As the reduced order ARMA method described by Mandal *et al.* (1992) shows potential usefulness in offshore application, this method of obtaining coefficients of the reduced order ARMA model is applied for the present study. The quality of the reduced order ARMA spectral estimates will depend on how accurately the initial higher order AR parameters are selected.

OPTIMAL MODEL ORDER SELECTION

The Yule-Walker method requires a minimum of p (where p is the AR model order) correlation coefficients for the solution of the well-determined set of equations of an AR process. Witz and Mandal (1991) have demonstrated the improvement in spectral estimation accuracy with a larger number of equations and more correlation coefficients. However, there are higher computational involvement associated with a higher order model. The question is how high a model order should one select so that the solution remains computationally efficient from the application point of view and the model can represent an acceptable target.

The Least Squares Error (LSE) method is commonly used owing to its simplicity, but can not be applied to any random process. If the process is purely convergent, the least squares error will decrease with increasing model order. Hence, using the LSE method, it is

sometimes difficult to select the model order which is the optimal representation of the stochastic process. Other method based on Akaike Information Criterion (AIC) (Akaike, 1974) or Minimum Description Length (MDL) (Rissanen, 1983) overcome this problem.

The AIC and MDL methods are expressed as

$$AIC = M [\ln(ER)] + 2p \quad (11a)$$

$$MDL = M [\ln(ER)] + p \ln(M) \quad (11b)$$

where M is the number of time series data points and ER is the average value of the sum of the squared error of the data or target spectrum with the estimated values from 1 to M .

APPLICATIONS

The wave measurements (Heteren et al, 1988) were carried out from the Noordwijk platform (MPN) at 17.3m water depth, and about 10km away from the Dutch coast. For the present study a time series wave data (M12) measured from North East location of the platform is used. This study considers 4096 discrete wave data points collected at a sampling frequency of 4Hz.

The time series wave data set of same duration are categorised into three time series of different sampling intervals: (a) time series of 4096 data points (denoted as M4096) with sampling rate of 4Hz, (b) data set of same duration as in case of M4096 with selecting alternate points forms the time series of 2048 data points (denoted as M2048) with sampling rate of 2Hz and (c) data set of same duration as in case of M2048 with selecting alternate points forms the time series of 1024 data points (denoted as M1024) with sampling rate of 1Hz.

Fig.1a shows the comparison of the two time series M4096 and M2048. Similarly Fig.1b shows the comparison between the two time series M4096 and M1024. This emphasizes that the time series (M2048) with sampling rate of 2Hz gives better representation than that of M1024. From these figures one can easily identify the effect

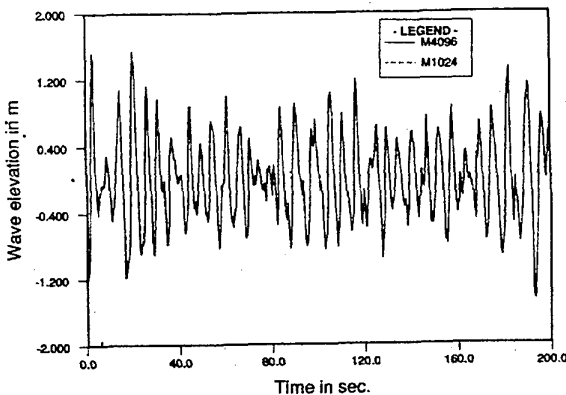


Fig.1a. North Sea time series waves: M4096 and M1024.

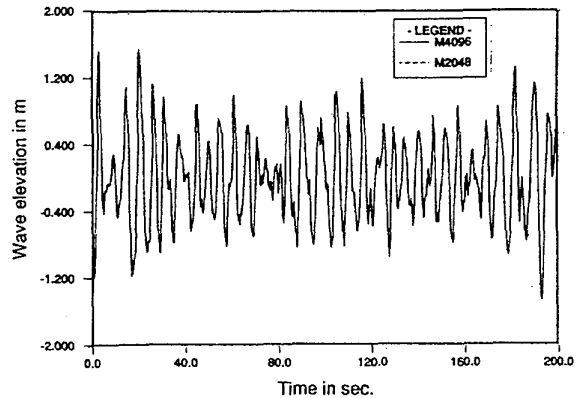


Fig.1b. North Sea time series waves: M4096 and M2048.

on sampling rate on accuracy of the digitized time series. The auto-correlation variation of the three time series is shown in Fig.2. Fig.3 shows variation of AIC [equation (11a)] and MDL [equation (11b)] of the time series (M4096). From Figures 2 and 3 the optimal model order is selected as 44. Therefore, the AR(44) model optimally represents the time series (M4096). The spectral estimate comparison of the AR(44) model and FFT is shown in Fig.4.

Similarly, variations of AIC and MDL for the time series M2048 and M1024 are estimated and the spectral comparisons between FFT and AR methods are shown in Figures 5 and 6. While examining all the

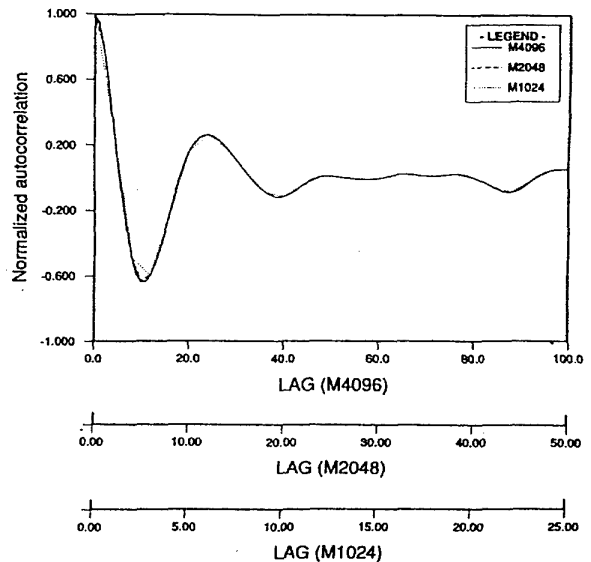


Fig.2. Normalized Autocorrelations of M4096, M2048 and M1024.

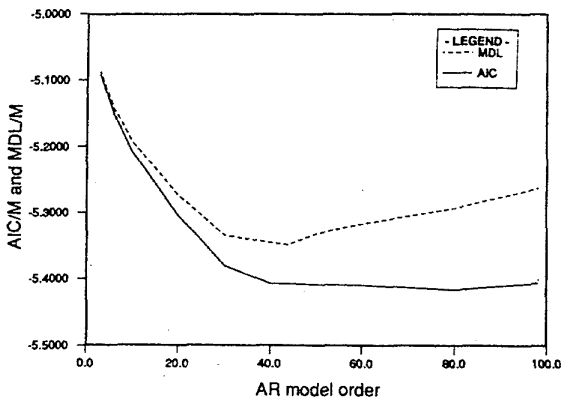


Fig.3. Variation of AIC and MDL for M4096.

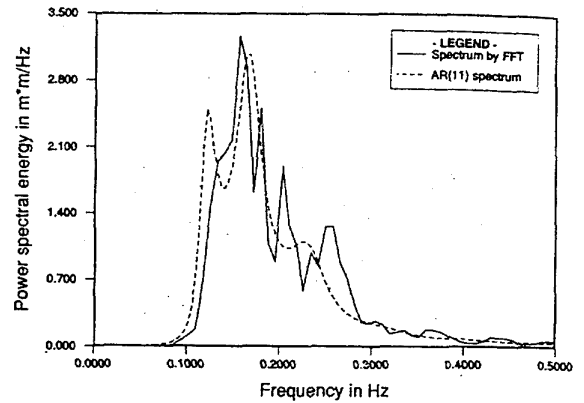


Fig.6. Spectral estimates of M1024 by FFT and AR(11).

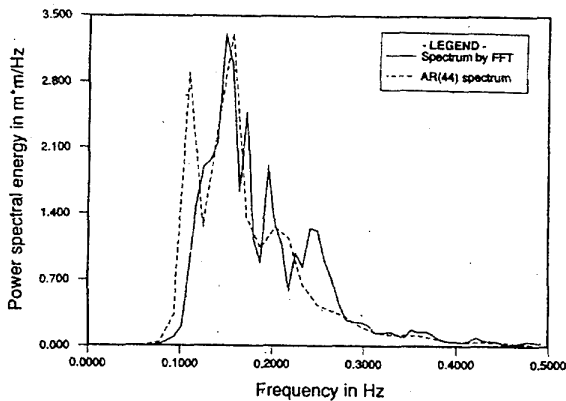


Fig.4. Spectral estimates of M4096 by FFT and AR(44).

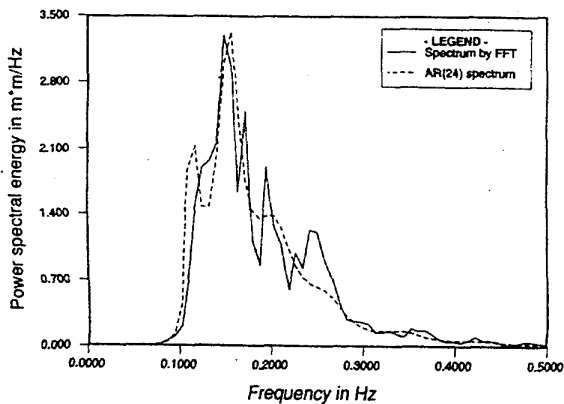


Fig.5. Spectral estimates of M2048 by FFT and AR(24).

three cases, the AR(11) spectrum of the time series (M1024) is fairly represented. The AR(24) spectrum of the time series (M2048) yields better representation of the dynamic system because the energy at second highest peak (0.11Hz) is much less as compared to other spectra of AR(11) and AR(44), and the energy at peak frequency region is very closed to the one determined by FFT. Therefore, it is observed that the very high sampling rate may give a better representation of the system in time domain. However, it is not been observed any improved statistical estimates as compared to other time series with lower sampling rates. The FFT spectra of the above three data sets are examined. It shows hardly any significant differences. While superimposing autocorrelation functions and using x-axis scales for 1, 1/2 and 1/4 times of M4096, M2048 and M1024 respectively (Fig.2), autocorrelation variation of M1024 is distinguishable as compared to other two cases owing to its roughness. This also shows that variation of autocorrelation lag is directly related to the sampling interval of the time series waves. Hence from the statistical point of view one should select a sampling rate which will optimize the collection of long term time series data.

The reduced order ARMA spectral estimation techniques based on Mandal *et al* (1992) are applied to M4096 and M2048 data sets. Based on these techniques the coefficients of the initial optimal AR polynomials are estimated from equation (6). After calculating roots of the AR polynomial forming real modes, energy in each mode is determined as shown in Tables 1 and 2. Selected modes which are contributing higher energies form the reduced order AR part of the ARMA model. The modes 9, 10 and 11 (Table 1) are multiplied to form the reduced order AR part for M4096 data set. Similarly modes 5 and 6 (Table 2) are multiplied to form the reduced order AR part for M2048 data set. The moving average part is then calculated based on the modal decomposition and partial fraction methods as described by Mandal *et al* (1992). The coefficients, a_k and b_l of the reduced order ARMA models are estimated, and the reduced order ARMA(6,6) and ARMA(4,4) models represent M4096 and M2048 data sets respectively. The ARMA(6,6) and ARMA(4,4) spectra give reasonable agreement with the FFT spectra as shown in Figures 7 and 8 respectively. It is observed that the time series with higher sampling rate yields better reduced order ARMA spectral estimates.

Table 1 Modes and energy levels for M4096.

Mode	Function	Normalized Energy
1	$1 + 2.0146z^{-1} + 1.0470z^{-2}$	< 1E-06
2	$1 + 1.7136z^{-1} + 0.9728z^{-2}$	< 1E-06
3	$1 + 1.3518z^{-1} + 0.9540z^{-2}$	< 1E-06
4	$1 + 0.9331z^{-1} + 1.0519z^{-2}$	< 1E-06
5	$1 - 0.1283z^{-1} + 1.2933z^{-2}$	< 1E-06
6	$1 - 1.4063z^{-1} + 1.5560z^{-2}$	< 1E-06
7	$1 - 0.4225z^{-1} + 0.6942z^{-2}$	< 1E-06
8	$1 - 2.5952z^{-1} + 1.7261z^{-2}$	< 1E-06
9	$1 - 1.8786z^{-1} + 0.9311z^{-2}$	1.0000
10	$1 - 1.7675z^{-1} + 0.8763z^{-2}$	0.1482
11	$1 - 1.0679z^{-1} + 0.5032z^{-2}$	0.0002

Table 2 Modes and energy levels for M2048.

Mode	Function	Normalized Energy
1	$1 + 1.8658z^{-1} + 0.9206z^{-2}$	< 1E-06
2	$1 + 0.3829z^{-1} + 1.3881z^{-2}$	< 1E-06
3	$1 + 0.4805z^{-1} + 0.4986z^{-2}$	0.00001
4	$1 - 1.4486z^{-1} + 1.2172z^{-2}$	< 1E-06
5	$1 - 1.6376z^{-1} + 0.7902z^{-2}$	0.9977
6	$1 - 1.2788z^{-1} + 0.5705z^{-2}$	1.0000

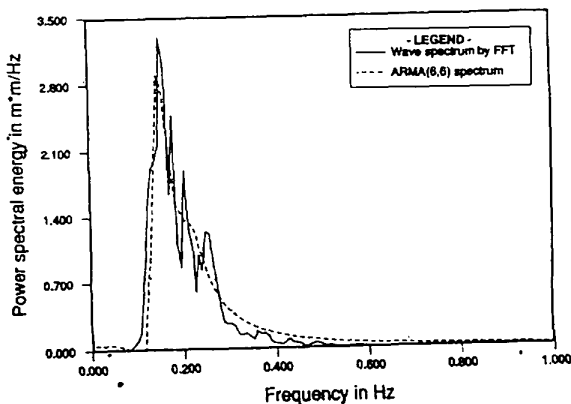


Fig.7. Spectral estimates of M4096 by FFT and ARMA(6,6).

However, this comparison shows that the reduced order ARMA(4,4) model of M2048 can be used to represent the measured sea waves. Using model order reduction techniques, here 4096 or 2048 data points are reduced to a few values which are the coefficients of the ARMA(6,6) or ARMA(4,4) model respectively.

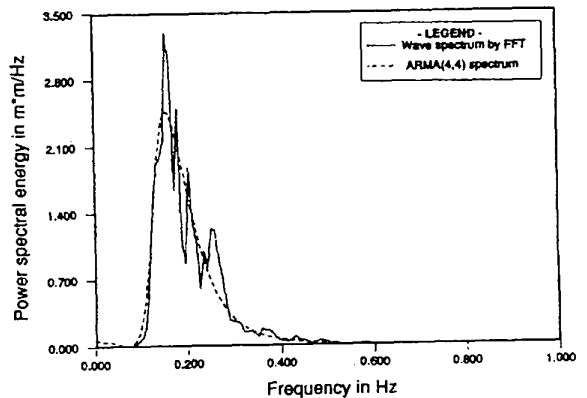


Fig.8. Spectral estimates of M2048 by FFT and ARMA(4,4).

Using optimal model order and appropriate sampling rate, the number of AR or ARMA coefficients is much less as compared to the number of FFT values. For on-line data collection programme implementation of the AR or ARMA algorithms has the potential advantage of storing more representations of wave environment than is currently possible because of the reduced size of the data sets.

CONCLUSIONS

An AR algorithm is formulated based on least squares error method. Optimal AR model orders are selected based on the AIC and MDL methods.

This paper establishes a better understanding of time series data analysis over sampling rate. One can generally examine such study using only FFT technique. But, here a study using AR algorithm strongly emphasises the need of suitable selection of sampling rate. The comparative study between the FFT and AR methods confirms the selection of an appropriate sampling rate to obtain a better statistical representation.

The reduced order ARMA spectral estimation techniques are successfully applied for M4096 and M2048 data sets. Time series of 4096 or 2048 wave data points can be represented by few values which are the reduced order ARMA coefficients.

The present study demonstrates that implementation of the AR or ARMA algorithms has the potential advantage of storing more representations of wave environment than is currently possible because of the reduced size of the data sets.

ACKNOWLEDGEMENTS

The author would like to gratefully acknowledge the support of Dr G J Lyons and Dr J A Witz, Santa Fe Laboratory for offshore engineering, University College London.

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19(6), December, pp 716-723.
- Borgman, L.E. (1969). Ocean wave simulation for engineering design. *Journal of Waterways and Harbor Division*, ASCE, 94(4), November, pp 557-593.
- Fines, S, Houb, O.G., Mo, K. and Overtick, T. (1981). On the estimation of wind wave spectra by parametric methods, *Coastal Engineering*, 5, pp 97-109.
- Graupe, D; Krause, D J and Moore, J.B. (1975). Identification of autoregressive moving average parameters of time series. *IEEE Transactions on Automatic Control*, Vol-AC-20, pp 104-106.
- Heteren, J.V., Keijser, H. and Schaap, B. (1988). Comparison wave directional measuring systems. *Applied Ocean Res.*, 10(3), pp 129-143.
- Mandal, S., Witz, J. A. and Lyons, G. J. (1992). Reduced order ARMA spectral estimation of ocean waves. *Applied Ocean Research*, 14(5), pp 303-312.
- Medina, J R and Sanchez-Carratala, C R (1991). Robust AR representation of ocean spectra. *Journal of Engineering Mechanics*, ASCE, Vol-117, No-12, pp 2926-2930.
- Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11(2), pp 416-431.

Spanos, P.D. and Hansen, J.E. (1981). Linear prediction theory for digital simulation of sea waves. *Journal of Energy Resources Technology*, ASME, 103, pp 243-249.

Spanos, P.D. (1983). ARMA algorithms for ocean wave modelling. *Journal of Energy Resources Technology*, ASME, 105, pp 300-309.

Tucker, M.J., Challenor, P.G. and Carter, D.J.T. (1984). Numerical simulation of a random sea: a common error and its effect on wave group statistics. *Applied Ocean Research*, 6(2), pp 118-122.

Witz, J.A. and Mandal, S. (1991). Application of system identification to the reduction of structural monitoring data. Final Integration of Design Techniques Project Report MON5, Santa Fe laboratory for Offshore Engineering, Dept. of Mechanical Eng., University College London, October, 1991.

NOMENCLATURE

a_k	Coefficients of AR process
b_i	Coefficients of ARMA process (numerator)
$a(z^{-1})$	Polynomial of AR process
$H_{AR}(z)$	Transfer function of AR process
$H_{ARMA}(z)$	Transfer function of ARMA process
M4096	Time series with 4096 points and 4Hz sampling rate
M2048	Time series with 2048 points and 2Hz sampling rate
M1024	Time series with 1024 points and 1Hz sampling rate
p	AR model order (denominator)
q	Model order of ARMA process (numerator)
R_i	Correlation coefficients of the random process
$S(\omega)$	Target power spectrum
$S_{yy}(\omega)$	Estimated power spectrum of AR model