

A Research of Completion of Wave Height Data

Ping-Chun Ho^{1,*} and Gwo-Fong Lin²

¹Keelung Harbor Bureau, Keelung, Taiwan, R. O. C. Department of Civil Engineering, National Taiwan University, Taiwan

²Department of Civil Engineering, National Taiwan University, Taiwan, R. O. C.

Abstract: The research discusses a completing model by using the measured data on the sea surface from 1996 to 2007 at Taipei port. The properties of the measured data are monthly mean and monthly standard deviation of all. By using the two properties, we can simulate the lost properties of both monthly mean and monthly standard deviation. The deseasoned data that were used to simulate lost wave data by using ARMA model. The simulated data were inserted into the lost section. In this model, we check the some properties of the both measured wave data and simulated wave data. The properties are ACF, PACF and spectrum. The properties of the two measured wave data and simulated wave data are the same. The completed data can represent the measured data. Propositions to stimulate future research about completing data are offered.

Keywords: Wave, ARMA model, ACF, PACF and spectrum.

1. INTRODUCTION

The location of a port, the plane of the facilities or the structure design must accord to the measured wave data in nature condition that are weather, ocean and land properties. To design the criterion about other port facilities, vessel sailing and cargos up/down loading also need the measured wave data.

The most important parameters of the port while consulting are the observation of the wave data, acquirement and analysis application. However, the measurement equipments located offshore might have extraordinary data while the equipments were affected by the vessels sailing. Sometimes, the vessels anchored offshore temporally, vessels were affected by the ocean current, tide, monsoon and propeller of itself, the cable connected with land will be cut off by the anchors of the vessels, then the data can not be acquired temporally. Sometimes, fishing boats cast nets on offshore will also affect seriously the measurement equipments the data will lost the same as the anchors. Except the vessels, anchors and fishing nets, the measurement equipments may have some problems, for example, low batteries, organism sticking, organism biking and corroding, all factors affect the quantity and accuracy of the data. Therefore acquiring the wave data we should complete the lost data by reasonable methods the consequence applications will be meaningful.

Guedes Soares and Ferreira [1] researched the wave data at Frigg Platform in the North Sea, they took out the monthly mean and monthly standard deviation of the data, the non-stationary time series became stationary. They completed the

lost data by insertion method. Since they drew the periodic graphics, they understood the period of the data is 12-month. After removing the variation of the means and standard deviations, AR (5) could fit the residuals data. The AR (5) model can simulate the wave data. Hidalgo *et al.* [2], Guedes Soares and Ferreira [1], Cunha and Guedes Soares [3] used the insertion method to complete the lost data while the quantity is under five. Ferreiro [4], Hidalgo *et al.* [2], Guedes Soares and Ferreira [5] used ARMA to complete the lost data while the quantity is over five.

Jardine and Latham [6] researched the significant wave data from 1962 to 1975 at N. E. Atlantic. They acquired the monthly means and standard deviations of the data. They found the periodic variation of the data. They confirmed the results by using the harmonic analysis and represented the results as cosine function. After removing the variation of monthly means and monthly standard deviations, the residuals can be fitted by AR(1) model. The check accepted the model and it can simulate the wave data well.

O'carroll [7] practically made explains, in detail, about the procedures that were described above. The first step is transfer the non-stationary measured data to stationary data. The method used the measured data subtract the monthly mean then divide the monthly standard deviation. The procedure was also used by Guedes Soares and Ferreira [1], Guedes Soares and Ferreira [5], Cunha and Guedes Soares [3]. In 2002, Stefanakos and Athanassoulis [8]. In 2002, Stefanakos *et al.* made this standardized processes as a normal procedure to simulate the significant wave. The monthly means and monthly standard deviations of the standardized processes are consequent variations. They use the sine/cosine function to simulate the variations. The best phrase of the function is determined by the least square between measured data and simulated data. The angle of the sine/cosine function is 30 degrees in one month and 360 degrees in one year.

*Address correspondence to this author at the Keelung Harbor Bureau, Keelung, Taiwan, R. O. C. Department of Civil Engineering, National Taiwan University, Taiwan; Tel: 886-2-24229044; Fax: 886-2-24229044; E-mail: pcho@klhb.gov.tw

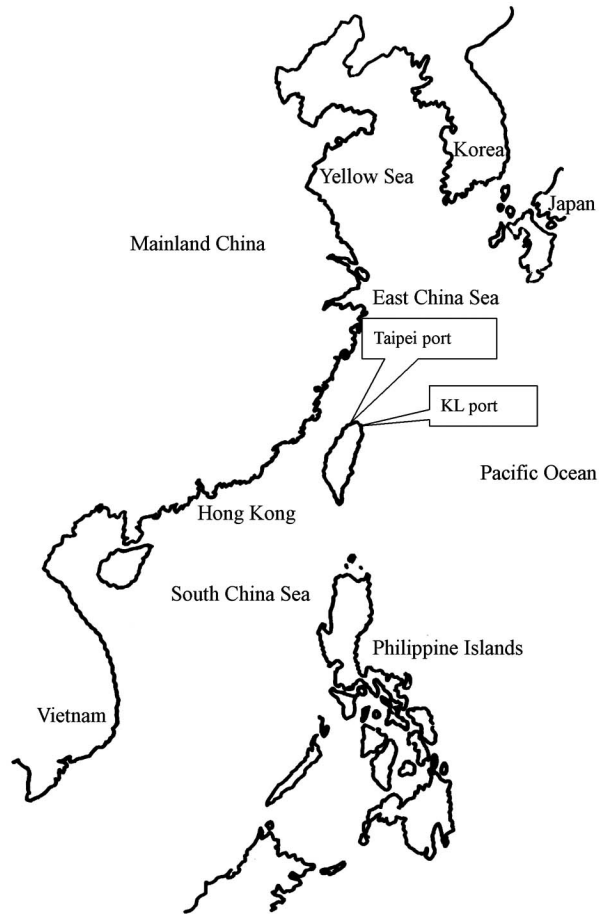


Fig. (1). The location of the Taipei port.

The function can simulate the variation in one year as a periodic change. Then the lost mean and standard deviation of the special month can be available.

The discharge of the river has the similarity as the means of the wave. For example, Lungu and Sefe [9] researched the record of a river located on the southeast Botswana. They found the first step harmonic analysis could describe the 97% ~ 99% variation of the discharge. They removed the first step periodic series and then found the stochastic model of the discharge. The model can forecast the discharge of the river. Lin and Chen [10] had some researches in this area of time series.

In this paper the model uses measured wave data to find the properties those have been suggested. An ARMA model then used the properties to simulate the data. The model can simulate the lost data. We compared the properties of ACF and PACF between measured data and simulated data. The properties of the data are the same. Then the model can be a reference in the future.

The next section, in this paper, will introduce the location of the measuring site and measurement of the wave data. The third section will describe the completing method. The fourth section will make a description of the results and checks of the model. The last section is the conclusions of the paper.



Fig. (2). The figure of the measuring data pile.

2. THE MEASUREMENT AND LOCATION OF THE MEASURING SITE

Taiwan island locates at south-east ocean of mainland China. The east of the island faces the Pacific Ocean directly. The directions of monsoons through the island are different. The northeast monsoon flows from the northeastern sea of the island and then into the coast in winter. On average, three to four typhoons attack the island from south-eastern or eastern sea each year. Keelung Harbor Bureau installed the measurement on the northern Taiwan Sea. Fig. (1) has shown the location of the site.

The wave data of this research came from the northern Taiwan island, the location has a port called Taipei port. The object of constructing the port was transporting sand from the eastern island to the northern island. The consulting method is pumping the sand in the port and refilling them to the banks of wharfs. The breakwaters outside the wharfs have to extend to the sea. Then the most important parameter that decides the height of the breakwater is significant wave height. Therefore, it is very important to study the properties of the wave height.

In the beginning of the consulting Taipei port, there was not any wave data at the site. The Engineer of Taipei port used other measurements near the site to estimate the wave height of the port. One of the measurements is located on the offshore of the Tao-Yuan county, the other is the CBK of China Petroleum Company outside Tsin-Chu county. However, the reliability of the estimated data is not good enough. For really understanding the properties of the wave data, the consulting bureau then installed a measurement pile outside Taipei port, the depth of the pile under water is 15 meters. The coordination of the pile is $25^{\circ}10'44''$ N and $121^{\circ}22'41''$ E. Fig. (2) shows the graph of the pile. A wind meter is on top of the pile and which is 15.4 meters above the water level. There are Tide and Current Meter under the water level. The meter can also acquire the data of wind, wave, tide and current at the same time. The meter records the wave data for 17 minutes in every one hour. The frequency of the meter is 2 Hz.

The data of the research is from 1996 to 2007. Because the variation of the offshore and fishing boats cast nets that caused the malfunction of the meters, the available rate of the data is not good enough. Table 1 shows the data quantity

Table 1. The Available Rate of the Wave Data at Taipei Port

Month	Year											
	2002		2003		2004		2005		2006		2007	
January	0	0.00%	743	99.90%	744	100.00%	513	69.00%	699	94.00%	744	100.00%
February	0	0.00%	672	100.00%	646	96.10%	254	37.80%	630	93.80%	540	80.40%
March	0	0.00%	134	18.00%	719	96.60%	542	72.80%	702	94.40%	740	99.50%
April	609	84.60%	719	99.90%	697	96.80%	578	80.30%	271	37.60%	717	99.60%
May	696	93.50%	719	96.60%	720	96.80%	671	90.20%	0	0.00%	737	99.10%
June	694	96.40%	695	96.50%	588	81.70%	422	58.60%	230	31.90%	711	98.80%
July	719	96.60%	718	96.50%	720	96.80%	719	96.60%	744	100.00%	625	84.00%
August	0	0.00%	718	96.50%	730	98.10%	744	100.00%	717	96.40%	576	77.40%
September	0	0.00%	695	96.50%	410	56.90%	717	99.60%	692	96.10%	639	88.70%
October	0	0.00%	684	91.90%	0	0.00%	701	94.20%	718	96.50%	0	0.00%
November	0	0.00%	131	18.20%	0	0.00%	492	68.30%	718	99.70%	0	0.00%
December	0	0.00%	598	80.40%	0	0.00%	381	51.20%	408	54.80%	0	0.00%

and available rate of all the data. The least data quantity is the year of 1996. The unstable meter was the major problem at the beginning. Because Keelung harbor bureau installed the pile in the sea that affected the pile for 6 years, then the pile inclined a lot. In 2002, Keelung harbor bureau installed another pile that caused less data. The quantity of the data at the Taipei Port measurement is almost 97, 000, the available rate is 58%. That is a good help to support the basic parameters for consulting Taipei Port.

3. COMPLETING MODEL

In this research, we survey all the data in Taipei port at first. If the lost quantity is under five then we use the linear interpolation to complete the lost data. If the lost quantity is more than five then we use the ARMA model to complete the lost data. If all data in one month are lost then we use the data in the month before lost data month to complete the lost data. If the lost records were more than two successive months then we use the data of the same month in the year before the lost year to complete the blank.

The ARMA model simulates the data that should be stationary. That meant the statistics properties of time series would not change with time. In general, the data of all nature records are not stationary. In this research, we transfer the non-stationary wave data to stationary data. The transfer method is Bruce [11] method.

We observe the figures of the monthly mean and monthly standard deviation, we know that the values of the mean and standard deviation are smaller in summer compared with that in winter. In general, the period that used in time series analyzes depend on different data properties or subjects. For example, Hirsch [12] use monthly data to analyze the reliability between water supply and hydrology. Lee and Liu [13] analyzed the discharge data that the monthly period were removed, and then used the time series model to analysis the discharge of the river. Guedes Soares and Ferreira [5] discussed the time series analyze and simulation of long-term significant wave height data at North Sea, in the model

they used the period is one month. We used the period in this model is one month. We used the Bruce model [11] to transfer the original data. We showed the first step as:

$$Y_{t,p}^* = \frac{(Y_{t,p} - \bar{Y}_p)}{S_p} \quad t = 1, 2, 3, \dots; p = 1, 2, 3, \dots, 12 \quad (1)$$

where $Y_{t,p}$ is the wave high value at time t in month p . \bar{Y}_p and S_p is the mean and standard deviation at month p , respectively. $Y_{t,p}^*$ is the value that removed the influence of the mean and standard deviation.

The second step of the Bruce model is shown as:

$$Z_{t,p} = \Phi^{-1} \left[F_{Y_{t,p}^*} (Y_{t,p}^*) \right] \quad (2)$$

Where $F_{Y_{t,p}^*} (Y_{t,p}^*)$ is the cumulative distribution of the original data, Φ^{-1} is the inverse transform of the normal cumulative distribution, $Z_{t,p}$ is the wave height data of the Bruce model at time t in month p . We connect the values at consequent time t in consequent months, and then we can have consequent wave values. At last, we can use the ARMA model to simulate the wave data.

4. RESULTS AND ANALYSIS

4.1. The Fitted Results of the Monthly Mean Value

In this research, we can calculate the mean value of every month from 1996 to 2007. From the figure, we know the values are between 0.3 m and 1.23 m in summer. The values of the mean are between 0.62 m and 2.12 m in winter.

The time series of the wave data is non-stationary in the data of all years. We used the method to calculate the mean values of the measured data. The quantity of data in this model is 11 years. If one month of the values is lost, the value of the others mean can replace the lost value. Therefore, we can simulate the data of the lost month by the mean value of the same month.

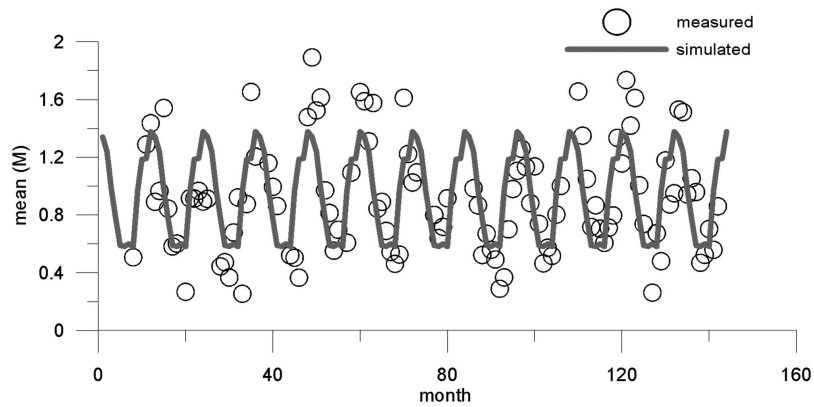


Fig. (3). The monthly measured and simulated mean data.

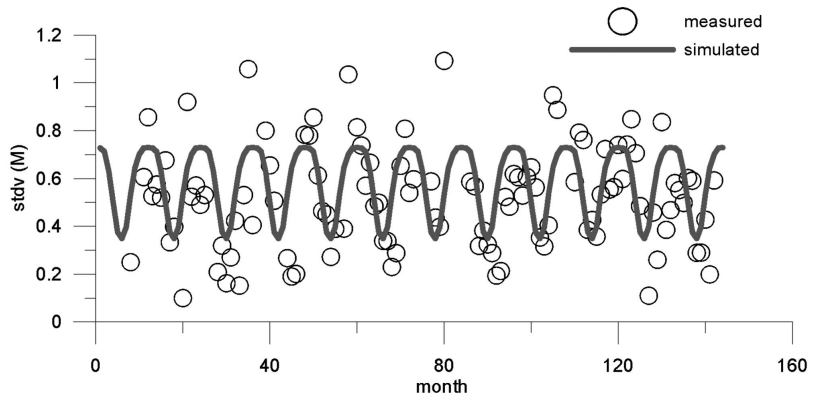


Fig. (4). The monthly measured and simulated standard deviation data.

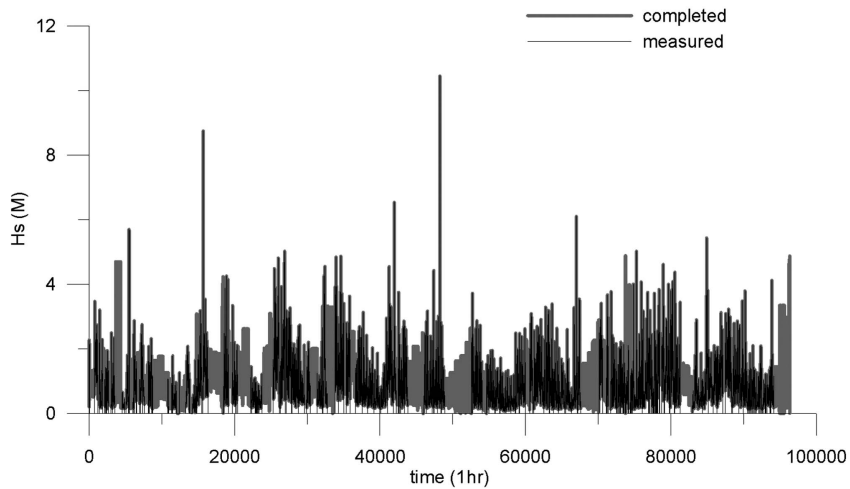


Fig. (5). The measured and simulated data.

Use the method that described above we can have simulated mean, and then combined the measured monthly mean, the mean values figure has shown in Fig. (3). In the Fig. (3), the change of the simulated value is the same as the change of the measured value, the simulated value line overlaps the measured value line in many places.

4.2. The Fitted Results of the Monthly Standard Deviation Value

The values of the measured monthly standard deviation in summer from 1996 to 2007 are between 0.35 m to 0.54 m. They are between 0.68 m to 0.74 m in winter. The method to simulate the standard deviation is the same as the method to

simulate the mean. Fig. (4) showed the results. From the Fig. (4), we can know the verity of the simulated standard deviation values were change with the measured standard deviation values. The simulated value line overlaps the measured value line in many places.

4.3. The Results of the Simulated Lost Data by ARMA Model

We use the ARMA model to simulate to lost data. Fig. (5) showed the results. In Fig. (5), the black line is measured significant wave height. There are many blanks between the two black lines. The blanks are lost data.

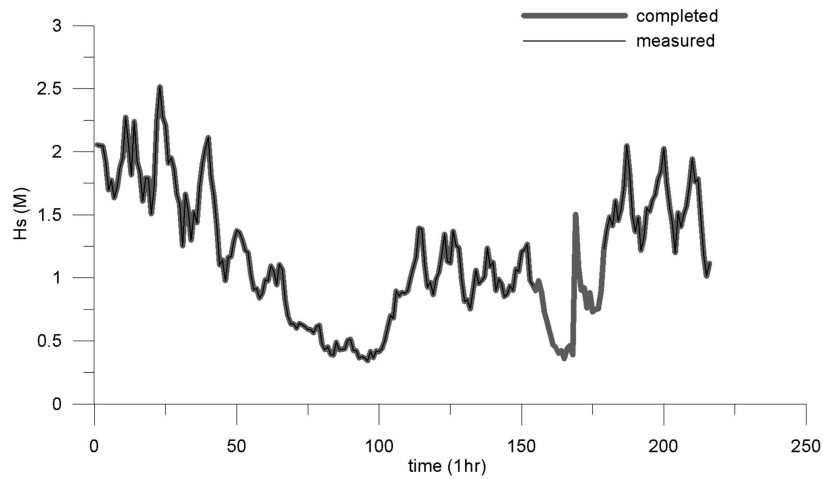


Fig. (6). The one month of measured and simulated data.

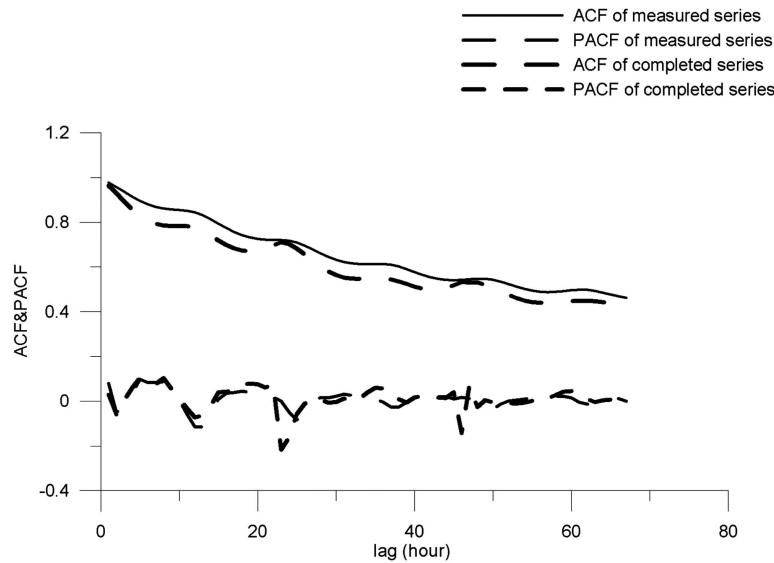


Fig. (7). The ACF and PACF of the measured and simulated data.

The gray line represents the simulated value that we used in the simulated method. Observe the black line and the gray line we can know the verity of the line in summer are smaller than winter. From the verity of the simulated data, they are the same as the measured data.

We use one-month wave height data shown in Fig. (6). In Fig. (6), the wave height are decrease in the first 100 hours, the verity are also decrease. The lost data are between 150-hour and 180-hour. The grey line shows the simulated data. The line also showed the verity of the data. That means the simulated method is feasible.

4.4. The Properties of the ACF and PACF

Fig. (7) showed the lines of the ACF and PACF of measured data and simulated data. From ACF-lines of the measured value and simulated value in the Fig. (7), both of the lines decrease as time lag increase. There is no abruptly rising up or falling down of the two lines. In other words, the properties of the two ACF are the same. From the PACF of the measured and simulated lines, we know they are similar with each other. The verities of the two lines are decrease with times.

4.5. The Properties of the Spectrum

We want to know the properties of spectrum of the measured and simulated data. Fig. (8) showed the two lines of the spectrums. The spectrums at low frequencies are increase in the figure, the highest is at 0.8, and then gradually decrease. The verities of the two lines are the same. Obviously, they showed abruptly a peak at frequency 8. The reason is the cold front coming from the northeastern sea is at an interval of 8 days. When the cold front passed the verities of the sea wave are increase, the peak then shown in the figure. The ARMA model can simulate the properties.

5. CONCLUSIONS

This paper introduces the competition of the lost data by ARMA model. First of all, we have to find the seasonal properties of the measured data. The properties are mean and standard deviation. Then remove the seasonal properties of the measured data. The ARMA model used the removed data to simulate the lost data. We replace the lost blank in the measured data to the simulated data. At last, we check the properties of the two sets of the data. The properties are

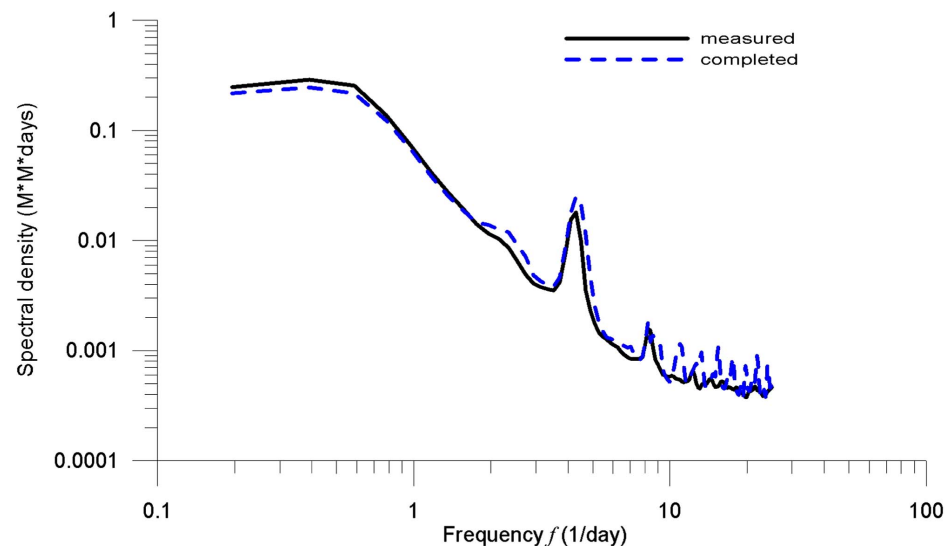


Fig. (8). The spectrum of the measured and simulated data.

ACF, PACF and spectrum. From the results, we know that the properties of the measured data and simulated data are the same. The simulated data can represent the measured data.

The measured data are from 1996 to 2007 at Taipei port. The length of the records is 11 years. By using such data to analysis we have excellent results. The method suggested in the paper is reasonable and feasible, and it can be a reference in the future.

ACKNOWLEDGEMENT

The authors are thankful to Harbor & Marine Technology Center, I. O. T., M. O. T. C., for kindly providing the valuable field data.

REFERENCES

- [1] C. Guedes Soares, and A. M. Ferreira, "Analysis of the seasonality in non-stationary time series of significant wave height", *Comput. Stochastic Mech.*, Spanos (ed.), Balkema: Rotterdam, 1995, pp. 559-68.
- [2] O. S. Hidalgo, J. C. Neito Borge, C. C. Cunha, and C. Guedes Soares, "Filling missing observations in time series of significant wave height", *Proc. 14th Int Conf Offshore Mech. Artic Eng.*, vol. 2, pp. 9-17, 1995.
- [3] C. Cunha, and C. Guedes Soares, "On the choice of data transformation for modelling time series of significant wave height", *Ocean Eng.*, vol. 26, pp. 489-506, 1999.
- [4] O. Ferreiro, "Methodologies for the estimation of missing observations in time series", *Stat. Probab. Lett.*, vol. 5, pp. 65-69, 1987.
- [5] C. Guedes Soares, and A. M. Ferreira, "Representation of nonstationary time series of significant wave height with autoregressive models", *Probab. Eng. Mech.*, vol. 11, pp. 139-148, 1996.
- [6] T. P. Jardine, and F. R. Latham, "An analysis of wave height records for the N. E. Atlantic", *Quart. J. R. Met. Soc.*, vol. 107, pp. 427-435, 1981.
- [7] F. M. O'Carroll, "Weather modelling for offshore operations", *Statistician*, vol. 33, pp. 161-169, 1984.
- [8] C. N. Stefanakos, and G. A. Athanassoulis, "A unified methodology for the analysis, completion and simulation of nonstationary time series with missing values, with application to wave data", *Appl. Ocean Res.*, vol. 23, pp. 207-220, 2002.
- [9] E. M. Lungu, and F. T. K. Sefe, "Stochastic analysis of monthly streamflows", *J. Hydrol.*, vol. 126, pp. 171-182, 1991.
- [10] G. F. Lin, and L. H. Chen, "Time series forecasting by combining the radial basis function network and self-organizing map", *Hydrol. Process.* vol. 19, no. 10, pp. 1025-1937, 2005.
- [11] R. L. Bruce, *On the Simulation of Long-Term Ocean Wave Data*, MSc thesis, University of Liverpool, 1982.
- [12] R. M. Hirsch, "Synthetic hydrology and water supply reliability", *Water Resour. Res.*, vol. 15, no. 6, pp. 1603-1615, 1979.
- [13] T. Y. Lee, and C. L. Liu, "Deseasonalized ARMA Model for Monthly Streamflows", *J. Civil Hydraulic Eng.*, vol. 16, no. 1, pp. 23-38, 1989.

Received: February 8, 2011

Revised: April 16, 2011

Accepted: June 15, 2011

© Ho and Lin; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.