Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Ocean Engineering 33 (2006) 1230–1248
OCEAN ENGINEERING
www.elsevier.com/locate/oceaneng

Technical note

Wave height forecasting by the transfer function model

Ping Chun Ho [a,*], John Z. Yim [b]

[a] *Engineering Department, Keelung Harbor Bureau, Keelung, Taiwan, R.O.C.*
[b] *Department of Harbor and River Engineering, National Taiwan Ocean University, Keelung, Taiwan, R. O. C.*

Received 1 March 2005; accepted 14 September 2005
Available online 10 November 2005

Abstract

Three completing methods were introduced in this paper. The data were completed from 1983 to 1988 at the KL and PTC stations. The completed data were called original data. The completing methods were the transfer function with fixed-parameters then built the Transfer Function (TF) model between the two stations. The TF model was used to estimate the wave data of the future. The difference of the estimated data and the original data were conferred in this paper. Two types of forecasting differences were discussed here: (1) the input data year increases from 1 to 6 years, and the difference of the estimated data and the original data were then compared. (2) Use one-year data as the input series to forecast the data of the next five years, and then compare the differences of the estimated data and the original data. After comparing the differences we can know the results. The mean value and the standard deviation of the differences decrease and go more stable as the quantity of the input data year increases. The mean value and the standard deviation of the differences increase and go more unstable as the quantity of the forecasted data year increases. This model can be used to forecast the wave data.

Keywords: Forecast; Time series analysis; ARMA model; Transfer function model

1. Introduction

The articles relevant to time series analysis are quite numerous. Among them, the most representative article will be found in the book Box–Jenkins (1976). This book introduced

* Fax: 886 2 24229044
E-mail address: pcho@klhb.gov.tw (P.C. Ho).

0029-8018/$ - see front matter
doi:10.1016/j.oceaneng.2005.09.003

ARMA/ARIMA model with three construction steps: (1) model identification, (2) model estimation, and (3) model diagnostic checking. In general, the researchers, according to these three steps can apply the ARMA/ARIMA model to the data analysis, model creation, and forecast operation. Box–Jenkins model also takes the consideration that during the same time period from a certain dynamic system it can measure the correction between two sets of time series data. The model was then constructed by inputs and outputs series. Through the analytical techniques of time series, we can make the transfer function between two sets of time series data. In the Box–Jenkins book it is known as transfer function model (TF model).

In simplicity, the TF model is meant to investigate the linear correlation about how input series affect the output series. By means of the said linear correlation, Box–Jenkins took the correlation between gas (input series) and carbon dioxide (output series). They also detailed the identifying, estimating, and diagnostic method for the forecast method. The models can be applied to forecast the carbon dioxide range of the probability, 50 or 95%. When Box and Jenkins introduced the identification stages of TF model, then we firstly make separate inputs and outputs, and prewritten the two series by the method of ARMA/ARIMA. We found out the cross-correlation function (CCF) of the residuals, and then we found the impulse response function (IRF) from the CCF. The residuals were prewritten before finding the CCF. By comparing the correlative competence between real IRF profiles and typical IRF profiles we can finally determine the numbers of parameters in the TF model. The subsequent procedures are designed for evaluation and diagnosis. We can finally reach the acceptable models applying to the future forecast. The aforesaid model creation procedures will have two defects during the actual application. *Defect 1*. This operation is required for manual determination methodology to compare the correlative graphs between IRF profiles and typical IRF profiles. For the data sets were simple and little numbers it is still reluctantly competent for manual determination availability. However, for the data sets with large numbers it is still required for manual comparison, but featured with less economic advantages. *Defect 2*. The manual determinative comparison will easily come with the subjective recognition difference, especially disadvantageous to profile determination. Any slight miss will go far wild in the final result and this said error will cause the considerable deviation of future forecast values of the model. The ways to compare profiles are also detailed with the description in the book written by Wei (1990).

Another method to determine the relation of the input series and output series is the dynamic regression model, otherwise called linear transfer function, in the identification procedure. The advantage is supposed that the input series are the polynomial of the output series. We can use the minimum squared method to find the parameter values of the input series. We do not need to find the relation of the prewritten input series and prewritten output series. The method of finding the linear relation was described in the books of Yaffee and McGee (2000) and Tsay (2002).

Pankratz (1991) introduced a method that can determine the parameters of the TF model in the identification procedure. He called the model as dynamic regression model. The advantage of that model is that it uses the rational polynomial form to represent the relation of input series and output series, the parameters of the rational polynomial form are referred to the typical IRF, then we can link them together and calculate with each other. We can know the real mean of them. The computer can use the relation to calculate the amount of the data objectively and to find the best-forecast values.

Hidalgo et al. (1995) applied the Box–Jenkins model to find the TF model of the Figueira Da Foz, Sines and La Coruña. First they used the AR model to fit the wave data of the three stations and had the residuals of the data. The graphs of the CCF of the residual at Figueira Da Foz station and Sines station are not obviously significant. But the graphs of the CCF of the Figueira Da Foz station and La Coruña station are obviously significant. Then a TF model was found to transfer the data of La Coruña station to the data of Figueira Da Foz and complete the lost data of Figueira Da Foz. When the Box–Jenkins model is applied to complete the lost wave data at different places of the same time. The CCF of the residuals, fitted the AR model of the stations, are high enough so that the effect of the TF model will be good. If the time of the lost data at both La Coruña station and Figueira Da Foz station are lost together, then the completed data will not show at the lost time section. This TF model needs improvement to complete the lost significant wave data before using.

Berthouex and Box (1996) discussed the BOD of the discharge at a sewage treatment plant. They compared the results of the univariable ARMA model and the TF model. The input effects of the TF model are the raw sewage flow rate, raw sewage, temperature, effluent suspended solids, sludge age…et al., the forecast results were compared with mean square error or square root of the mean square error. The forecast results of the TF model were better than the univariable ARMA model because the TF model takes more factors into consideration.

Young et al. (1997) used the non-linear TF model to discuss the transfer relation of the rainfall and flow data for the Canning River in Australia. The non-linear TF model is affected by many factors including temperature, moisture, seepage, sunshine, and rainfall, and the output data are discharge. They used the non-linear relation to study the rainfall and flow and used the rainfall data as the input series of the TF model. The transfer relation was called non-linear TF model. The influences of the input series affecting the output series are very complex. They removed the little affected factors, and then remained the rainfall factor as the input factor of the TF model. The measuring rainfall stations may be one or many stations. The results of the TF model were identified by coefficient of determination and Akaike information criterion (AIC) then they confirmed that TF model could be used to forecast the river flow.

The non-linear TF model used by Young et al. is the model which is basically the Box–Jenkins model. The input series and output series were replaced by non-linear functions; the calculations were more complex. The model also had to identify the orders of the parameters and had to pass the portmanteau lack-of-fit test.

Box–Jenkins TF model was applied to many religions about the time series analysis. The model considers many variables and increases/decreases the input factors then increases the precision of the forecast data. However, if the input and output series are lost at the same time then the TF model cannot estimate the lost data. We have to identify the order of the parameters by people checking the typical IRF profiles that lower the efficiency of the computer application.

This research uses the data of two stations: KL station and PTC station. The data are uncompleted which should be completed. After completing the lost data then the TF model can transfer the data at PTC station to the data at KL station, and then compare the statistics of the estimates. The TF model can be used to complete the lost data, and the completed data were called original data. This research uses the significant wave data of the two different stations for computer identification, suggested by Pankratz (1991). The identification can improve the Box–Jenkins model. This paper also introduces a fixed-parameters transfer function model to forecast the data of the future, and then discusses the difference of the estimates and the original data. The second section introduces the time series of the two stations. The following sections document the Box–Jenkins TF model used in this paper and the Pankiatz model. Section 4 outlines the discussion of this TF model. The final section presents the results of forecasting wave data of this research.

2. Wave data stations

This research discusses the transfer properties of two stations located at the northeastern coast of Taiwan Island. The Taiwan Island is located at the southeastern coast of China. The east coast of the island faces the Pacific Ocean directive. The monsoons blow over the island all year and the directions of monsoons are different. The northeastern monsoon blows from the northeastern coast in winter. The typhoon may affect the island from the south or the east in summer or autumn. There are measuring stations at Keelung Harbor (KL) and Pi-Tou-Chiao (PTC). Fig. 1 shows the geographical locations of these two stations. Fig. 2 shows the time series data of the two stations.

Keelung Harbor Bureau provides the data used by this study. The measuring station (25°10′20″N, 121°44′53″E) is located at the northeast of Taiwan. Wave heights were recorded for 20 min every 2 h using an ultrasonic wave meter system (Type SUTW-8600). The instrument is anchored outside the Keelung port, and is approximately 1000 m north of the eastern breakwater of the harbor. Water depth at the measuring station is 60 m. In Table 1, we have summarized the data available of the location, from 1983 to 1990. A zero in the table denotes no wave records. Hereafter, the data from Keelung Harbor will be referred to as KL for brevity.

Pi-Tou-Chiao A station (25°08′09″N, 121°55′31″E) is deployed by the Center Weather Bureau, Republic of China. The buoy wave recorder is located at PTC in Taiwan, which is approximately 2000 m offshore. The water depth is 55 m. Twenty-minute recordings of the sea state were carried out every 2 h from October 1980 to July 1988. Data availability for these months is shown in Table 2. Hereafter, we will use the abbreviation PTC for the data of PTC measuring station.

Keelung port has been in operation for one hundred years from 1886 and has loaded and unloaded for all over the world. The instrument is set outside Keelung port and is affected by the navigating boats. The unnatural values of the measuring data will come out occasionally. Sometimes, the boats waiting for the births to load/unload cargos outside the port, the boats will anchor for a long times. The boats are affected by the ocean currency, ebb and flow of the tide, monsoons and the propeller turning around. The anchor on the sea bottom will move with the boat. The anchor may hook the cable of the measuring equipment, and then pull and drag. This might break the cable and the data will be lost. Beside, there are two fishing ports, Cheng-Pin fishing port and Pa-Tou-Tzu fishing port, located in the east outside the Keelung port. There are numerous fishing boats driving

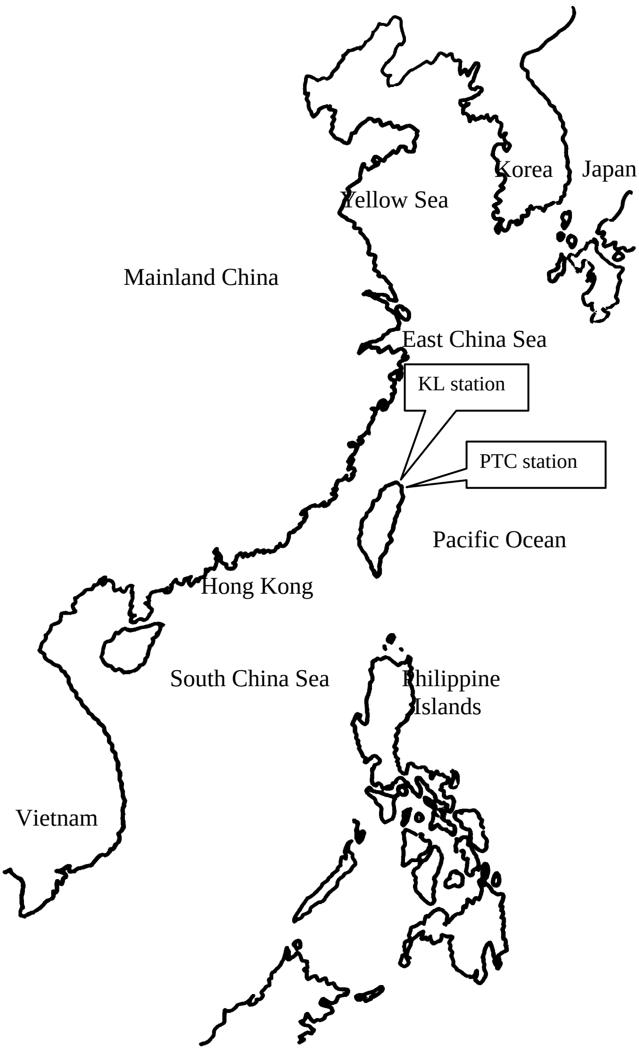

Fig. 1. The locations of the KL station and PTC station.

in/out the fishing ports and the movement of the boats may also affect the measuring equipment.

PTC station located at the sea between PTC and Yi-Lan County. The Kuroshio current affected the area. The current came from the bottom with nutrition and many planktons and attracted many kinds of fish and many shoals of fish. The fishing boats then arrive at the area to put cages or drift gill net to catch the fish. When the boats come across the buoy of the PTC station the data are affected; sometimes the data may be lost.

Apart from the effects of the boats, fishing tackles or chains of the anchors, there are other factors which may affect the measuring equipments that are put in the sea, for example, exhausting battery, equipments hooked by the trawl, stuck by seashell, covered

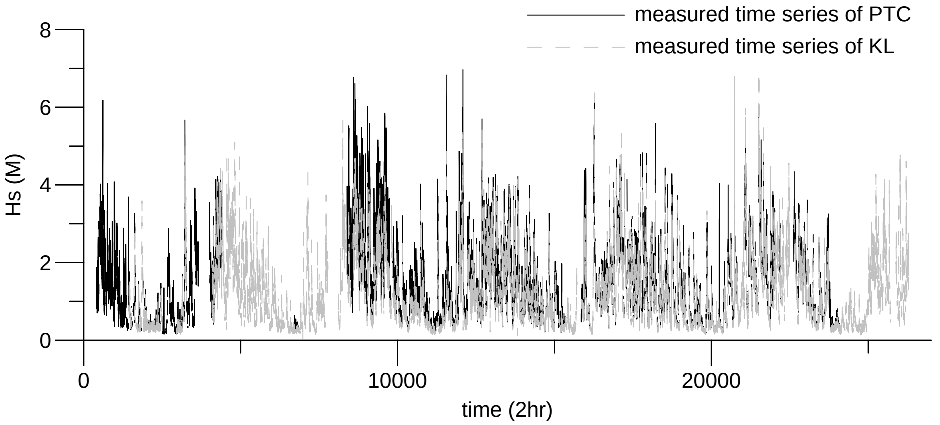

Fig. 2. The measured time series of the KL station and PTC station from 1983 to 1988.

Table 1
The available data are from May 1983 to October 1990 at the KL station

Month/year	1983	1984	1985	1986	1987	1988	1989	1990
1	0	357	314	369	372	354	369	365
2	0	346	334	336	329	170	264	332
3	0	372	361	372	349	329	371	339
4	0	360	337	360	360	359	351	292
5	268	371	347	364	353	371	363	371
6	349	297	360	273	357	294	350	357
7	305	371	354	363	346	346	369	372
8	0	384	347	211	370	370	372	341
9	328	348	360	326	292	357	353	358
10	0	123	369	371	367	360	366	220
11	0	227	360	343	358	180	358	0
12	293	255	372	371	368	368	359	0

by cake, animals biting cable, corroded by the salt, etc. and these will affect the data availability and precision. The data may be lost. Therefore, after having the significant wave data, we must complete the lost data, so that the future research will be meaningful.

3. The TF model of the Box–Jenkins and Pankratz model

Box–Jenkins introduced the ARMA/ARIMA model in 1976. The model was also extended as the TF model at that time. The ARMA model was written as:

$$\phi(B)y_t = \theta(B)a_t \quad (1)$$

where y_t is the significant wave height at time t.

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \quad (2)$$

Table 2
The available data are from October 1980 to June 1988 at PTC station

Month/ year	1980	1981	1982	1983	1984	1985	1986	1987	1988
1	0	360	362	0	1	370	316	259	112
2	0	290	288	291	0	334	302	255	0
3	0	335	312	323	0	300	314	244	230
4	0	317	343	273	0	344	331	259	193
5	0	334	349	183	0	335	300	252	196
6	0	315	290	145	0	335	291	259	208
7	0	305	1	94	127	323	41	213	1
8	0	318	231	228	0	325	167	136	0
9	0	302	316	291	0	353	198	221	0
10	226	335	1	335	0	358	249	65	0
11	350	343	0	1	0	354	239	197	0
12	368	352	0	339	356	335	223	263	0

$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q \tag{3}$$

where B is the lag operator and can be rewritten as follows

$$By_t = y_{t-1} \qquad B^i y_t = y_{t-i} \tag{4}$$

The parameter ϕ_p and parameter θ_q are the parameters of the AR and MA model, respectively. The values p and q are the order of the AR and MA model, respectively. The parameter a is the independent variable in the model and the Gaussian white noise process, $N(0, \sigma_a^2)$. The orders p and q are referred to the graphs of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of y_t series. The parameter values were calculated by minimum mean square error method. The model has to be examined by the AIC test and portmanteau test and the residuals have to be checked by the autocorrelation check. The books of Wei (1990), Brockwell and Davis (1991), Shumway and Stoffer (2001) described the details.

A TF model where an influence series (called input series) was put inflects the ARMA model. It basically includes one or more input series and one or more output series. The direction of the influence is unidirectional, the input series can infect the output series, otherwise does not. Hereafter we consider a TF model with a single input series and a single output series, the model can be shown as:

$$y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} a_t \tag{5}$$

where

$$\omega(B) = 1 - \omega_1 B - \cdots - \omega_s B^s \tag{6}$$

$$\delta(B) = 1 - \delta_1 B - \cdots - \delta_r B^r \tag{7}$$

The means of $\theta(B)$, $\phi(B)$ and a_t are the same as function (1). The input series $x_{t-\mathrm{b}}$ are the significant wave height at time $t-b$, b is the lag time. The output series y_t

are the significant wave height at time t. The function (5) is sometimes called the dynamic regression model. The values r, s, and b are the orders of the model, respectively. The orders are referred to the typical IRFs of the input series and output series. The details can be seen in Box and Jenkins (1976), Wei (1990) and Pankratz (1991).

First of all, we have to know the meaning of the time series at the two stations in the TF model. The input time series are at the location of PTC station. The output time series are at the location of KL station. The two time series have to been prewritten before using the TF model, then we call the procedure is normalization. After finding the normal time series, we have to calculate the CCF of the two residuals. Then calculate the IRF by the CCF. We compare the profiles of the calculated IRF with the profiles of typical IRF, the orders r, s, and b can be known. The meaning of IRF is that the input series will affect the output series after a few times (call lag time), the change will show out. The change can be shown in the profiles of IRF. The orders of the $\theta(B)$ and $\phi(B)$ in the TF model are the same as the meaning of the ARMA model, and the procedure of estimation is the same as the ARMA model.

This paper used two yearly sets of time series to find the profiles of the IRF, wish to find the typical IRF to match, however, we cannot find the specific profiles to fit. Thereafter, construct the yearly times series cannot be used to the TF model in the area. We modified the method, not to use the yearly data, but used the monthly data and referred to the techniques of Pankratz (1991, p184) written as:

judge b: Find the bar charge of the IRF profile, the height value higher than double standard deviation at the first lag time, and then determine b. The value b was called the dead time.

judge r: There is not any downfall pattern in the IRF profile, then mostly r is 0, 1 or 2. In experience $r \geq 2$ is rarely appeared. The IRF profile does not descend slowly but rather cutoff to zero obliviously then $r=0$; when the IRF profiles show a simple exponential downfall, either instantly or in the end, then $r=1$; when the IRF profile show a compound exponential decay, or damped sine wave decay, either instantly or in the end, then $r=2$.

judge s: set $s-u+r-1$; and when $r>0$ then u is the location of non-zero highest value in the IRF profile; when $r=0$ and we cannot determine the pattern in the IRF profile, then $u=0$.

The paper changed the determination from yearly IRF profiles to monthly IRF profiles. The results of comparing the monthly IRF profiles are better than the yearly IRF profiles. The paper used the monthly IRF method and the techniques of the determination, suggested by Pankratz; a computer can then be used for the amount of the data calculation, and then the parameters of the model can be determined exactly and quickly. The techniques can also adopt the monthly data and improve the precision of the yearly data.

This research used six years, total 72 monthly transfer functions to calculate the mean values of the same month parameters in different years. Then we have one set (12-month) parameters to present the TF model of the two stations, called fixed-parameter TF model. The input data year increases from 1 to 6 years in the model. Then the differences of the estimated data and the original data of two-, three-, or six-year data can be calculated.

The method to calculate the differences is shown below:

$$d = |y_t - \hat{y}_t| \tag{8}$$

where d, the differences of the estimated data and original data; y_t, the original wave height data at time t; $\hat{y}_t$, the estimated wave height data at time t.

This paper also discussed the differences of the original data and forecasted data that the input one-year data to forecast the next one-year data, next two-year data, and next five-year data. The determination of the difference is the same as the function (8).

4. The results of the fix parameter TF model and discussion

The wave data are lost at KL station and PTC station that are the same as others. This research uses three methods to complete the lost data under three different conditions: (1) insert method: the lost data are few; (2) ARMA model: the lost data are longer but less than the month should have; (3) O'Carroll (1984) method: the lost data are longer than one month.

The method to complete the lost data by O'Carroll model is basically suggested; the mean and standard deviation of the month in one year can be fitted by the trigonometric function. The earth goes around the sun and causes the change of the airstreams; the change of the wave will be changed at last. The earth goes around the sun with period and the mean and the standard deviation of the wave data may have the period too. We use the trigonometric function to simulate the period of the mean and standard deviation. The mean and standard deviation of the months in one year cannot change much quickly. The amplitude of vibration does not change so much too. The lost mean and standard deviation of the month can be completed by the simulated mean and standard deviation in the same year. That is the mean and standard deviation can be estimated by the fitted trigonometric function. O'Carroll (1984) suggested the method. This paper used the O'Carroll method to complete the lost month and compare the other years with the same month that did not lose the data, we can find the difference is little, and there is no extraordinary value in the completed data. The competed data can be used for the future study.

The measured data at the KL and PTC stations were completed by three methods. The completed data were called original data. The ACF, PACF and moments of the measured data and original data were compared, and then we can know the statistics of the original data are the same as the measured data. We use the ACF and PACF of the KL station in 1985 to compare the differences (Fig. 3). The ACF lines of the original data and measuring data were decay and neared to each other. The PACF of the original data and measured data were cut off at lag time 1. The results of the moments to compare the differences of the original data and measured data were shown in Table 3. The data at the KL station were widely distributed from May to September 1985. The wave height was smaller in that period, and the mean value was 0.78 m, the simulated mean was 0.92 m. There were only data in January, July, and December 1984 at the PTC station, 73.55% data were concentrated in December, and in the particular period the wave height was more.

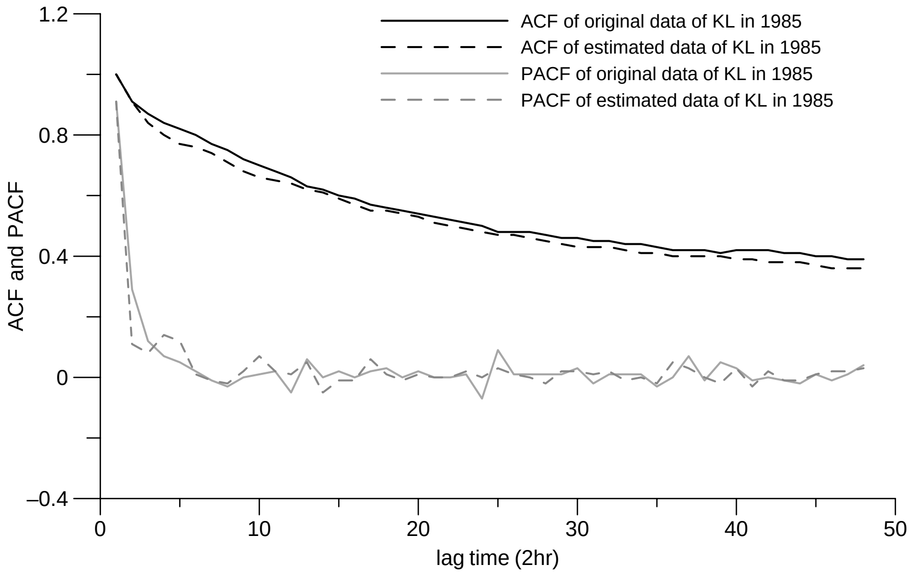

Fig. 3. The ACF and PACF of the original and estimated times series at the KL station in 1985.

The mean value in the table was 2.12 m. The completed data compared with other data, and we know the differences were not much. We can see from the ACF, PACF and moments to compare the measured data with the original data, we can know the statistics of the two data are the same. We completed the data at the KL station from 1983 to 1988 the time series of the figure was shown in Fig. 4. The time series of the completed data at PTC station from 1983 to 1988 is shown in Fig. 5.

Then we used the method suggested by Pankratz (1991) to transfer the data from PTC station to KL station every month. The data at the PTC station overlap the data at KL station from 1983 to 1988. There are six overlapped years and we then have six sets of parameters as shown in Table 4. The wave time series data estimated by the six sets of

Table 3
The moments of the original and completing wave data

	Mean	Standard deviation	Skewness coefficient	Kurtosis coefficient
Original data at KL station in 1983	0.78	0.88	1.86	3.23
Completed data at KL station in 1983	0.92	0.96	1.97	2.74
Original data at PTC station in 1983	1.21	0.91	0.89	3.49
Completed data at PTC station in 1983	1.26	0.88	1.11	4.69
Original data at PTC station in 1984	2.12	1.55	0.34	2.35
Completed data at PTC station in 1984	1.15	1.54	0.60	2.10
Original data at PTC station in 1988	1.12	0.85	0.73	2.70
Completed data at PTC station in 1988	1.33	0.92	0.38	4.43

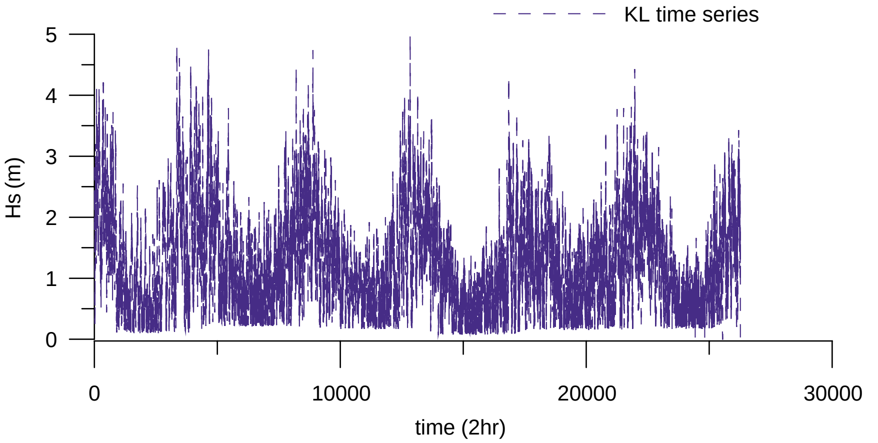

Fig. 4. The original times series from 1983 to 1988 at the KL station.

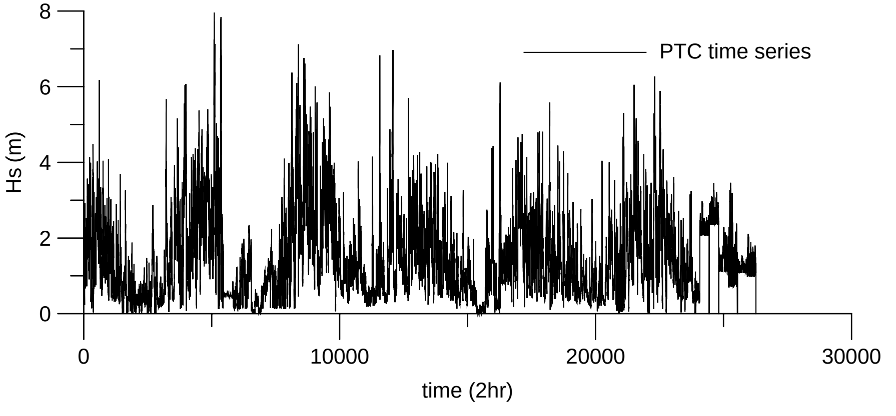

Fig. 5. The original times series from 1983 to 1988 at the PTC station.

parameters (changed-parameters model) at KL station from 1983 to 1988 is shown in Fig. 6. The mean values of the six sets of parameters called fixed-parameters are also shown in Table 4. The wave time series data estimated by the fixed-parameters model are shown in Fig. 7. We compared the data estimated by the changed-parameters model and fixed-parameters model, and we know the statistics of the estimated data is the same. The statistic results are compared the ACF, PACF and moments of the two sets. We show the ACF and PACF of the original data (KL), estimated data by changed-parameters model (Sim KL) and estimated data by fixed-parameters model (Fix model Sim KL) from 1983 to 1988 in Figs. 8–13, respectively. There are some common properties in the figures. (1) All of the ACF lines are decayed exponentially and without any periodic change. (2) All locations of the lag time at 8 where the ACF lines are under 0.5 which means the relation is little. (3) The change of the ACF lines of the original data and estimated data by changed-

Table 4
The values of parameters *b, r, s, p, q* at KL the station

Year/*b, r, s, p, q* / Month/	1983	1984	1985	1986	1987	1988	Fixed parameters
1	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0
2	1,2,1,2,0	1,2,1,2,0	1,2,1,6,0	1,2,1,1,0	1,2,1,1,0	1,2,1,2,0	1,2,1,2,0
3	1,2,1,2,0	1,2,1,2,0	1,2,1,5,0	1,2,1,3,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0
4	1,2,1,2,0	1,2,1,2,0	1,2,1,6,0	1,2,1,2,0	1,2,1,8,0	2,2,1,2,0	1,2,1,4,0
5	1,2,1,2,0	1,2,1,2,0	1,2,1,4,0	1,2,1,3,0	1,2,1,2,0	2,2,1,6,0	1,2,1,3,0
6	1,2,1,2,0	1,2,1,2,0	1,2,1,6,0	3,2,1,1,0	1,2,1,4,0	2,2,1,1,0	1,2,1,3,0
7	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0	1,2,1,3,0	1,2,1,2,0	1,2,1,1,0
8	1,2,1,2,0	1,2,1,2,0	2,2,1,5,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0	1,2,1,1,0
9	1,2,1,6,0	1,2,1,2,0	1,2,1,3,0	2,2,1,4,0	2,2,1,3,0	1,2,1,2,0	1,2,1,2,0
10	1,2,1,1,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0
11	1,2,1,1,0	1,2,1,2,0	1,2,1,3,0	1,2,1,4,0	1,2,1,2,0	1,2,1,2,0	1,2,1,2,0
12	1,2,1,2,0	1,2,1,2,0	1,2,1,6,0	1,2,1,2,0	1,2,1,4,0	1,2,1,2,0	1,2,1,3,0

parameters model and fixed-parameters model are the same. There is no rise or drop all of a sudden and that means the ACF is steady and the same. (4) The locations of the PACF at lag time 2 where the PACF are dropped quickly to 0.1 means the statistics of the three data are the same. The first and second moments are shown in Figs. 15 and 16. The lines are the means and standard deviations of the original data at PTC station (PTC), original data at KL station (KL), estimated data by the changed-parameters model at KL station (Sim KL) and estimated data by the fixed-parameters model at KL station (Fix model Sim KL). The changes of the mean values in Fig. 14 are similar to each other. The result, in Fig. 15, the changes of the standard deviation is the same as the change as the mean values.

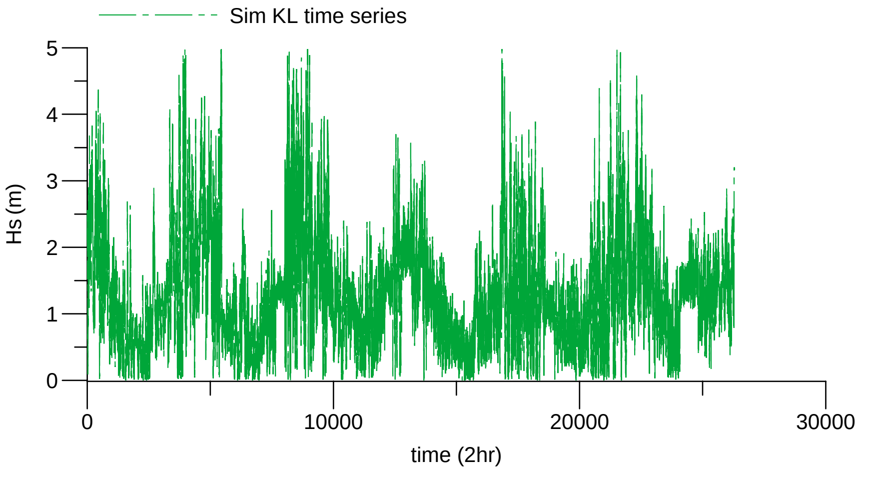

Fig. 6. The estimated times series by the changed-parameters model at the KL station from 1983 to 1988.

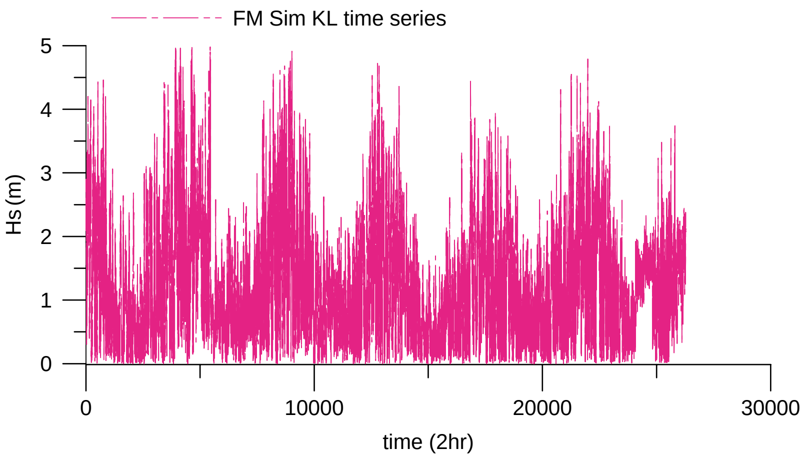

Fig. 7. The estimated times series by the fixed-parameters model at the KL station from 1983 to 1988.

We can know from Figs. 8 to 15 that the fixed-parameters model can make good use of transferring the data from PTC station to KL station.

We use the fixed-parameters model to input one-year data into the model and forecast the data of the next year. The input data year increases from 1 to 6 years then we can know the differences of the original data and the forecasted data. The differences are shown in Table 5. We can know the following results from the table. (1) The mean of the difference decreases as the input data year increases. (2) The parenthesis is the standard deviation of the difference. The difference values have much change when the input data year is one.

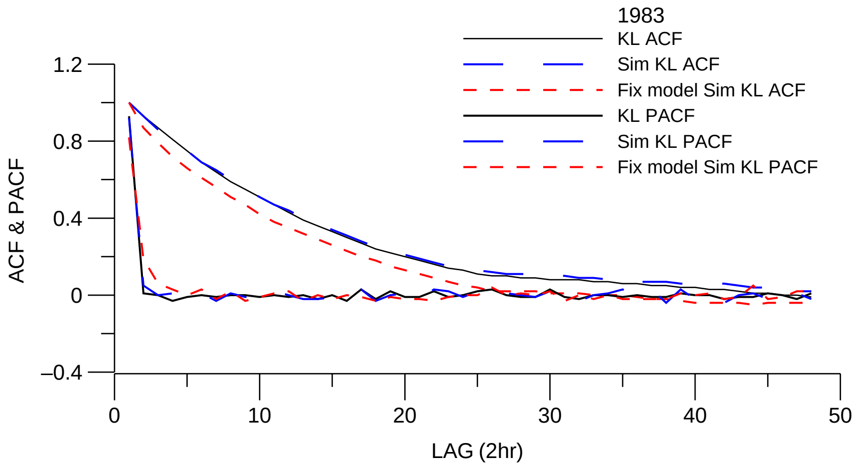

Fig. 8. The ACF and PACF of the original data (KL) and the estimated times series by the changed-parameters (Sim KL) and fixed-parameters model (Fix model Sim KL) at the KL station in 1983.

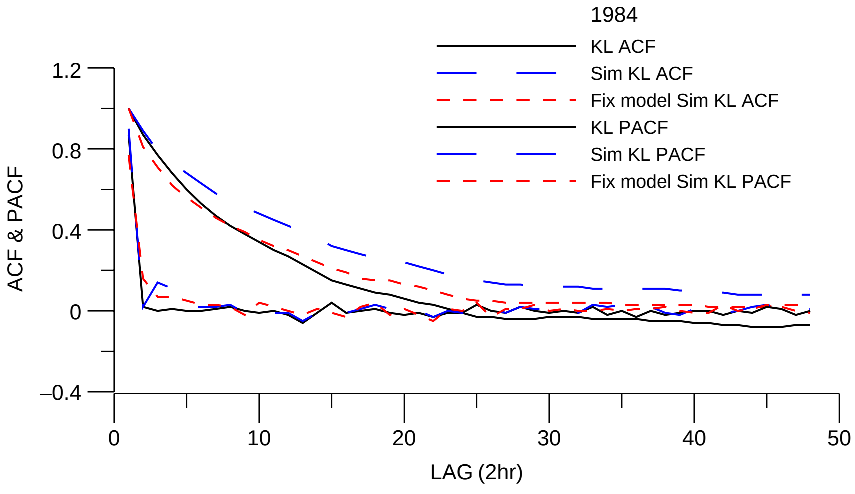

Fig. 9. The ACF and PACF of the original data (KL) and the estimated times series by the changed-parameters (Sim KL) and fixed-parameters model (Fix model Sim KL) at the KL station in 1984.

When the input data year increases to six years the standard deviation of the difference decrease and seems more stable. (3) When the input data year is little then the mean and standard deviation is smaller in winter, and the summer is larger. (4) The mean of the differences are between 1.38 and 0.84 m when the input data year is one. When the input data year increases to six years the difference values are between 0.64 and 0.62 m. (5) The standard deviations of the differences are between 1.36 and 0.96 m when the input

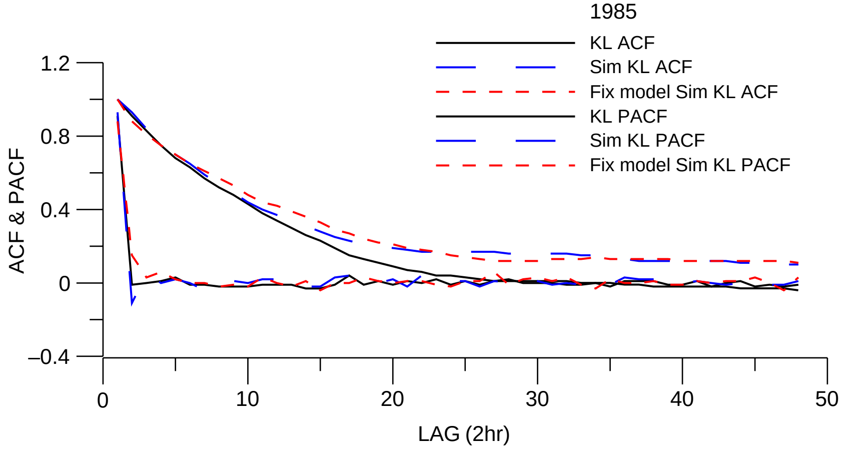

Fig. 10. The ACF and PACF of the original data (KL) and the estimated times series by the changed-parameters (Sim KL) and fixed-parameters model (Fix model Sim KL) at the KL station in 1985.

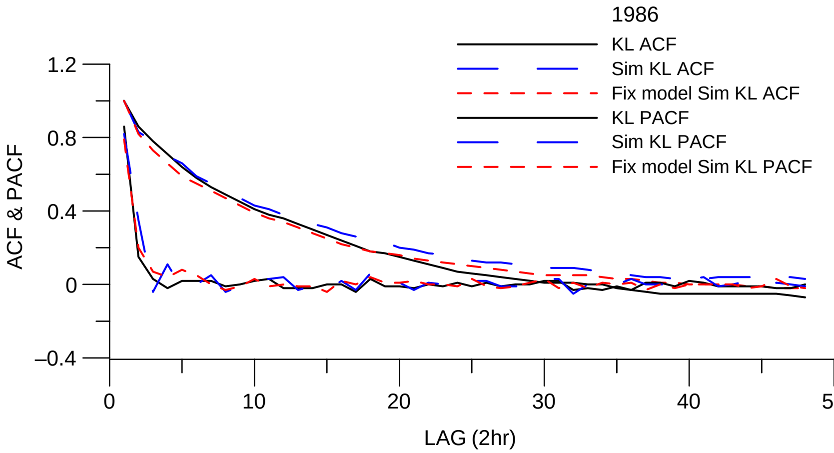

Fig. 11. The ACF and PACF of the original data (KL) and the estimated times series by the changed-parameters (Sim KL) and fixed-parameters model (Fix model Sim KL) at the KL station in 1986.

data year is one. When the input data year increases to six years the values are between 0.69 and 0.67 m.

If we input the only one-year data to forecast the next few year data, what the differences are? We then input one-year data into the fixed-parameters model and forecast the next one-year, next two-year…and next five-year data. The differences are shown in Table 6. From the table we can have the results as follows. (1) The mean of the differences increase as the input data years increase. (2) The standard deviation values of the differences have less change when the forecasted data year is one. When the forecasted

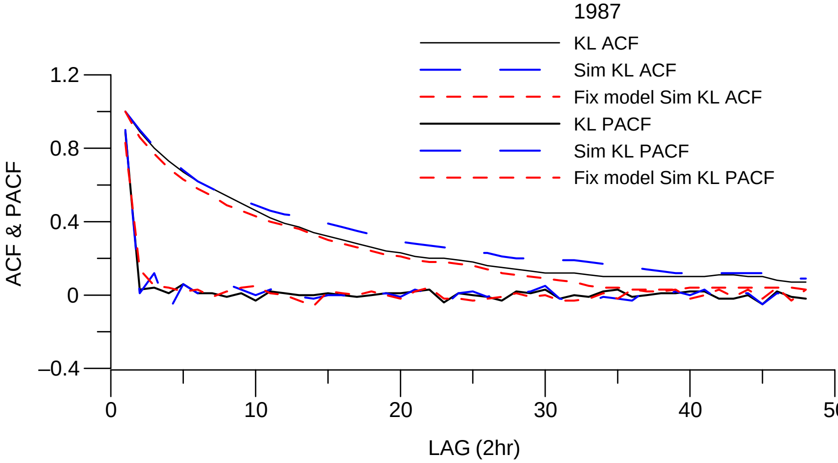

Fig. 12. The ACF and PACF of the original data (KL) and the estimated times series by the changed-parameters (Sim KL) and fixed-parameters model (Fix model Sim KL) at the KL station in 1987.

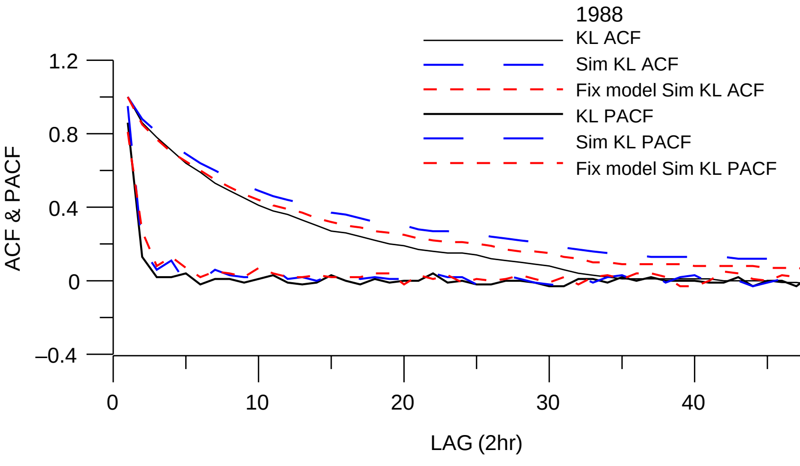

Fig. 13. The ACF and PACF of the original data (KL) and the estimated times series by the changed-parameters (Sim KL) and fixed-parameters model (Fix model Sim KL) at the KL station in 1988.

data years increase to five years the standard deviation value of the difference decreases and seems unstable. (3) When the forecasted data year is little the mean and the standard deviation of the difference are small in summer, but the values are large in winter. (4) The means of the differences are between 1.14 and 0.52 m when the forecasted data year is one. The forecasted data year increases to five, the mean values are between 1.10 and 0.45 m. (5) The standard deviations of the differences are between 1.09 and 0.35 m as the forecasted data year is one. When the forecasted data years increase to five the values increase and between 1.12 and 0.48 m.

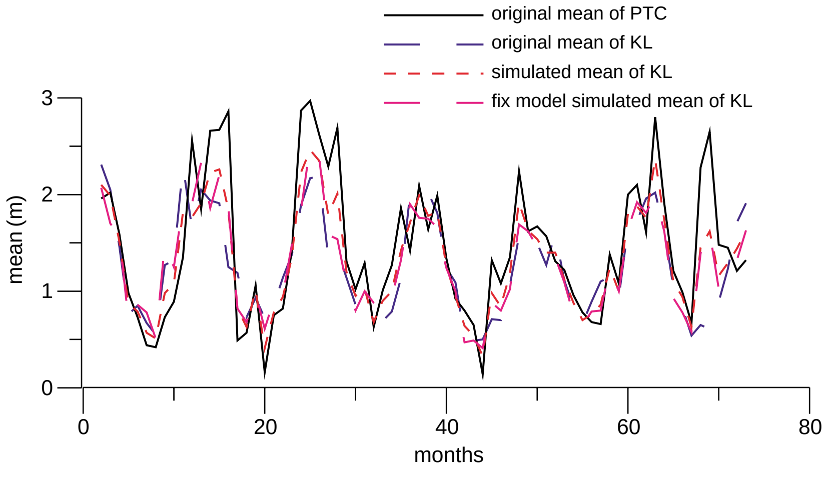

Fig. 14. The mean of the original times series (KL & PTC) and the estimated data by the changed-parameters (Sim KL) and fixed-parameters model (Fix model Sim KL) at the KL station.

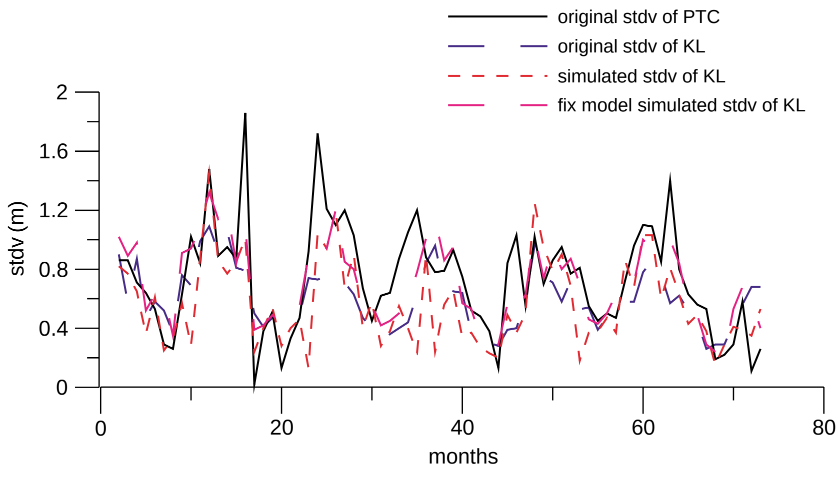

Fig. 15. The standard deviation of the original times series (KL & PTC) and the estimated data by the changed-parameters (Sim KL) and fixed-parameters model (Fix model Sim KL) at the KL station.

Table 5
The means and standard deviations of the differences are between original data and estimated data in different input years

Input series / years/month	One year	Two years	Three years	Four years	Five years	Six years
1	1.37 (1.08)	1.00 (1.16)	0.92 (0.85)	0.71 (0.72)	0.61 (0.68)	0.62 (0.67)
2	1.36 (1.00)	0.88 (1.03)	0.89 (0.82)	0.69 (0.71)	0.62 (0.68)	0.62 (0.67)
3	1.38 (1.36)	0.97 (1.08)	0.88 (0.81)	0.69 (0.71)	0.63 (0.68)	0.62 (0.68)
4	1.35 (1.33)	0.89 (1.03)	0.83 (0.79)	0.68 (0.70)	0.63 (0.68)	0.63 (0.68)
5	1.12 (1.24)	0.81 (0.99)	0.79 (0.76)	0.66 (0.69)	0.62 (0.68)	0.63 (0.69)
6	1.03 (1.14)	0.77 (0.96)	0.76 (0.75)	0.64 (0.69)	0.61 (0.68)	0.63 (0.69)
7	0.94 (1.08)	0.70 (0.93)	0.74 (0.74)	0.62 (0.68)	0.62 (0.67)	0.64 (0.69)
8	0.87 (1.02)	0.66 (0.91)	0.72 (0.74)	0.61 (0.67)	0.62 (0.67)	0.64 (0.69)
9	0.84 (0.98)	0.97 (0.89)	0.71 (0.74)	0.60 (0.67)	0.61 (0.67)	0.63 (0.68)
10	0.87 (0.96)	0.96 (0.89)	0.72 (0.74)	0.59 (0.67)	0.61 (0.67)	0.63 (0.68)
11	0.97 (1.04)	0.65 (0.92)	0.72 (0.74)	0.61 (0.67)	0.62 (0.67)	0.63 (0.68)
12	1.19 (1.35)	0.93 (0.86)	0.72 (0.73)	0.61 (0.68)	0.62 (0.67)	0.63 (0.68)

The values in parentheses are the standard deviations.

5. Conclusion

This research used three completions to complete the lost data before applying the transfer function. The completed data were called original data and the statistics of the original data are the same as that of the measured data. We can complete the lost data by the three completions even if the input data are missing. The completed data then can be a set of input series and transfer to the output series.

The estimation of the TF model was suggested by Box–Jenkins. However, the estimation must use people to check the typical IRF and determine the parameters.

Table 6
The differences of the original data and estimated data of forecasting next year's data by using the fixed-parameters model

Month/forecast years	One year	Two years	Three years	Four years	Five years
1	0.97 (0.79)	0.81 (0.81)	0.89 (0.86)	0.93 (0.88)	0.84 (0.79)
2	0.86 (0.64)	0.69 (0.62)	0.70 (0.57)	0.71 (0.56)	0.75 (0.51)
3	1.08 (1.09)	1.08 (1.01)	1.03 (0.99)	1.02 (0.96)	1.01 (0.88)
4	1.14 (0.60)	0.99 (0.70)	1.02 (0.69)	1.01 (0.67)	1.01 (0.63)
5	0.52 (0.45)	0.45 (0.47)	0.47 (0.47)	0.51 (0.50)	0.52 (0.49)
6	0.50 (0.35)	0.43 (0.39)	0.46 (0.39)	0.41 (0.48)	0.45 (0.48)
7	0.69 (0.61)	0.64 (0.67)	0.72 (0.71)	0.77 (0.73)	1.01 (0.76)
8	0.98 (0.84)	0.89 (0.91)	1.00 (0.94)	1.08 (0.99)	1.09 (1.01)
9	0.63 (0.54)	0.56 (0.59)	0.59 (0.54)	0.60 (0.57)	0.61 (0.52)
10	0.81 (0.61)	0.70 (0.66)	0.72 (0.68)	0.75 (0.70)	0.83 (0.74)
11	1.03 (0.79)	0.91 (0.79)	0.99 (0.83)	1.03 (0.83)	1.05 (0.78)
12	1.04 (0.97)	1.00 (0.97)	1.06 (0.94)	1.11 (0.98)	1.10 (1.12)

The values in the parenthesis are the standard deviations.

This paper applied the techniques of Pankratz to determine the parameters and used the monthly data to transfer the amount of data. The results are perfect. The advantages of the computer are applied to the calculation.

When we used the changed-parameters transfer function model to transfer the wave data the parameters are changeable depending on different months and years. We used the mean values of the parameters of the same month in different years to find a set of fixed-parameters to transfer the wave data. The statistics of the two estimated data are the same after checking the ACF, PACF, and moments of the two data.

This paper used different data years as input series to input the fixed-parameters model and to forecast the data of the next year. The results of comparing the forecasted data with the original data are written as below. The mean of the difference decreases as the input data year increases. The input data year increases then the standard deviation of the difference decrease and seems more stable. The input data year is more less than the mean and standard deviation is smaller in summer, and larger in winter. The means of the differences are between 1.38 and 0.62 m. The standard deviations of the differences are between 1.36 and 0.67 m.

We use one-year data as input series to forecast the next year data and next few year data; the results of the comparisons are shown as below. The mean of the differences increases as the input data years increase. The forecasted data years increase and the standard deviation value of the difference decreases and seems unstable. The forecasted data years are little, the mean and the standard deviations of the differences are small in summer, but the values are large in winter. The mean of the differences are between 1.14 and 0.45 m. The standard deviations of the differences are between 1.12 and 0.35 m.

The results include the competed data model, monthly changed-parameters TF model, monthly fixed-parameters TF model and the differences of the forecasted data. We can know the fixed-parameters model can make good use of forecasting the data in the future years.

References

Berthouex, P.M., Box, G.E., 1996. Time series models for forecasting wastewater treatment plant performance. Water Resource 30 (8), 1865–1875.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, CA p. 575.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, second ed. Springer, New York p. 577.

Hidalgo, O.S., Borge, J.C.N., Cunha, C., Guedes Soares, C., 1995. Filling missing observations in time series of significant wave height. Proceedings of the 14th Conference on Offshore Mechanics and Artic Engineering, vol. 2, pp. 9–17.

Pankratz, A., 1991. Forecasting with Dynamic Regression Models. Wiley, New York p. 386.

Shumway, R.H., Stoffer, D.S., 2001. Time Series Analysis and its Applications. Springer, New York p. 549.

Tsay, R.S., 2002. Analysis of Financial Time Series. Wiley, New York p. 448.

Wei, W.W.S., 1990. Time Series Analysis: Univariate and Multivariate Methods. Addison-Wesley, Redwood City, CA p. 478.

Yaffee, R.A., McGee, M., 2000. Time Series Analysis and Forecasting. Academic Press, New York p. 528.

Young, P.C., Jakeman, A.J., Post, D.A., 1997. Recent advances in the data-based modelling and analysis of hydrological systems. Water Science Technology 36 (5), 99–116.