CERTIFICATION OF APPROVAL

Time Series Modeling of Measured Wind & Wave Time Histories in Malay, Sabah & Sarawak Basin Using Autoregressive, Integrated, Moving Average (ARIMA) Method

by

Mohd. Farid bin Mohd. Shafri

A project dissertation submitted to the
Civil Engineering Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF ENGINEERING (Hons)
(CIVIL ENGINEERING)

Approved by,

(Assoc. Prof. Ir. Dr. Mohd. Shahir Liew)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

SEPTEMBER 2011

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

Certified by,

mohd farid

MOHD. FARID BIN MOHD. SHAFRI
890923-10-5095
UNIVERSITI TEKNOLOGI PETRONAS

ACKNOWLEDGEMENTS

In the name of ALLAH S.W.T, the Most Merciful and Compassionate, praise to ALLAH, He is the Almighty, eternal blessing and peace upon the Glory of the Universe, our beloved Prophet Muhammad S.A.W, his family and companions.

I wish to express my deepest appreciation to my Final Year Project supervisor, Assoc. Prof. Ir. Dr. Mohd. Shahir Liew for his confidence in me during the completion of this research of studies. His encouragement, guidance and advice help me a lot in understanding the topic and its scope. Thanks to him for leading me from the starting point which is the introduction and the background of study of this project until this dissertation is completed. All the support and guidance from him will not be forgotten.

Not to forget, deepest gratitude to postgraduate students, Zuraida Mayeetae and Wan Zalikha binti Anas for helping me directly and indirectly sharing the information which are the raw data from PETRONAS Carigali and the guideline for PASW Statistics 18 software.

I am most indebted to my parents for all encouragement and support throughout my undergraduate studies plus emphasizing to me the needs of education from the very beginning of my life.

ABSTRACT

It is very important to understand the metocean (meteorological and oceanographic) data since it is a fundamental in order to success in all marine projects as this information will provide the characteristics of the available resource for energy yield, the design requirements for survivalibility of the project and the strategy for maintenance and accessibility. So, in order to avoid bad events on the sea and to help with proceeding in situ operations, it will be very beneficial to perform the analysis of the time series modeled and obtain the forecasts of the metocean data from the time histories record which is consists of wind and wave parameters. The method that will be applied during the analysis and forecasting the measured metocean data is Autoregressive, Integrated, Moving Average (ARIMA) method.

TABLE OF CONTENT

LIST OF TABLE

LIST OF FIGURE

LIST OF GRAPH

CHAPTER 1

INTRODUCTION

1.1 Background

Metocean data is the collection of both meteorological and oceanographic data including offshore meteorological conditions (e.g. wind, sea level pressure), sea states through induced wave currents and water movements (tidal elevations and current flows). So, these data is very useful in order to access resources availability, downtime (e.g. delay to the vessel due to weather or condition can affect productivity), fatigue and extreme design values for offshore engineering design and operations (James Parker, 2010).

In order to perform the design process and operation of the offshore structures installation safely and efficiently, it is very beneficial to get important information from metocean data that provides the conditions that may affect the installation process. The metocean data mostly consists of wind, wave and current condition of a certain location. However, other parameters maybe very important for other particular locations such as visibility and ice condition.

Usually, metocean data will be represented as a summary statistic (a single statement of a long-term average) or as a time series i.e. the high frequency variation in a parameter measured over a period of time. The source data can be classified as:

1. **Past data** – archived data which may originate from either model or measurement. Hind casting is the common term for models that attempt to replicate measured data. Extreme analysis statistics depend on past data, with the length of the data archive being the prime interest for robust statistical analysis. Long-term datasets also provide the means of establishing climatic behavior in metocean conditions and inherently this requires data spanning decadal periods.
2. **Present data** – the relation between real time observations and are desirable to support operational decisions (most likely based upon in-situ measurement). If the real time data are used for forecasting short-term prediction, the resulting product is called nowcast.

3. **Future data** – provide some of prediction that based on suitable atmospheric or ocean model. Forecasting is a skill which considers all the predictions and offers an informed view on conditions in the near future. This type of data is very important in order to proceed with in-situ operations.

Metocean data can be originated from various sources. The following table will explain all the sources of the metocean data, strengths and limitations for each source.

Table 1: Strengths and limitations of metocean data sources

Data Source	Strengths	Limitations
In situ measurements	• Most accurate representation of metocean condition • Data provided as a time series should fit all analysis types • Can supply real time data for nowcast	• Site-specific, may not represent whole area of interest • Need to carefully plan deployment and servicing to avoid data gaps • Expensive to deploy and maintain, particularly with real time communications
Remote sensing	• Measured data with wide spatial coverage • Specialist providers ensure a level of quality control and provide preprocessed data • Errors expected due to high level of processing applied to raw instrument data	• Poor temporal sampling makes data inappropriate for some analysis types • Subject to errors where complex sites are represented by broader scale spatial grid points
Re-analysis/hindcast modeling	• Good coverage both in time and spatial area • Data provided as a time series should fit all analysis types	• Errors expected due to model representation of complex physical processes • Subject to errors where complex sites are represented by broader scale spatial grid points
Forecasts	• Allow planning for future events • Numerous products can be tailored to specific operations	• Risk of higher level of error with increasing forecast horizon • Spatial detail may be coarse • Need to ensure forecast delivers benefit (e.g. through validation)

The research of studies will be involving one of the metocean data sources which is modeling a time series (wind, wave and current) based on time histories record.

1.2 Problem Statement

Many say that the world's weather is driven by oceans and controlled by complex interactions between the oceans, the land and the atmosphere. It is a great benefit for us if we able to accurately monitor and predict weather, climate and oceanographic conditions to avoid bad events such as maritime accidents, pollution and disasters at sea. Plus, it will help us in order to proceed with in situ operation and also ongoing maintenance that need to be done.

In order to get the prediction of the metocean data needed, a set of time histories record need to be obtained and by using Autoregressive, Integrated, Moving Average (ARIMA) method, the analysis of the time series modeled and forecasting procedure can be done.

1.3 Objective

Basically, there are two main objectives of this study:

1. To model a time series model based on the current statistical properties from measured metocean data.
2. To determine the similarities or differences of the measured metocean properties for Malay, Sabah and Sarawak Basin.
3. To forecast with certain confidence limit of future environmental load i.e. wind and wave.
4. To ascertain the degree of stationarity of measured metocean data.

1.4 Scope of Study

This research of studies will be involving analysis of time series modeled and forecasting wind and wave based on the set of site-specific measured metocean data from PETRONAS Carigali Sdn. Bhd. that was specially generated in order to know the impact of the metocean parameters on the design, installation and operation of PETRONAS' facilities. The time series modeling and forecasting method that will be used is ARIMA method.

CHAPTER 2

LITERATURE REVIEW

2.1 Time Series

Time series is a chronological sequence of observations on a particular variable (Bowerman, O'Connell, Koehler, 2005). Time series can be composed of: trend, cycle, seasonal variations and irregular fluctuations.

Trend can be classified as ups and downs of the times series over a period of time. So, trend reflects the increment or decrement in the time series. On the other hand, cycle refers to recurring up and down movements around trend levels. These fluctuations can have a duration of anywhere from 2 to 10 years or even longer measured from peak to peak or trough to trough. Seasonal variations can be classified as periodic patterns in time series that complete themselves within a year and then repeated annually. Usually, this component can be affected by weather and customs. The last component of time series is irregular fluctuations. Irregular fluctuations are inconsistent movements in a time series that have no regular pattern. Many irregular fluctuations in a time series caused by unusual events that cannot be forecasted (e.g. earthquakes, hurricanes). Irregular fluctuations also can be caused by the error done by time series analyst.

2.1.1 Time Series Analysis

Information in time series can be obtained in time and frequency domain. But in this research of studies, only time domain analysis will be involved. In time domain analysis, the first step need to be done is generating a time series plot which is in form of observation versus time graph. This plot will visualize the statistical consistency of a time series. Next is obtaining the basic descriptive properties of the time series which are mean and variance.

$$\mu = \frac{1}{N}\sum_{t=1}^{N} x_t$$

(Equation 1)

$$\sigma^2 = \frac{1}{N-1}\sum\nolimits_{t=1}^{N-k}(x_t - \mu)^2$$

(Equation 2)

The covariance between x_t and x_{t+k} separated by k intervals in time can be estimated by:

$$C_k = \frac{1}{N}\sum\nolimits_{t=1}^{N-k}(x_t - \mu)(x_{t+k})$$

(Equation 3)

Auto-covariance function also can be obtained by using Equation 3. Autocovariance is the variance of the variable against a time-shifted version of itself. Similarly, autocorrelation can be calculated by applying the Equation 3. Basically, autocorrelation is a correlation between values of the process at different points in time, as a function of the two times or of the time difference. The following formula shows the relationship between auto-covariance and autocorrelation function:

$$r_k = \frac{C_k}{\sigma^2}$$

(Equation 4)

2.1.2 Stationarity

Time series can be considered as stationary if the statistical properties (e.g. mean and variance) of the time series are essentially constant through time which means that the properties are independent of the time origin. On the other hand, non-stationary time series do not have constant mean and variance which means that the statistical properties are dependent on the time origin. According to Nelson (1972), stationarity is a very strong condition to impose on time series since stationary time series rarely exist.

For the forecasting method that will be discussed next, stationarity is very crucial since autoregressive process will only be stable if the parameters are within a certain range, for example, if there only one autoregressive parameter then is must fall within the interval of $-1 < \phi < 1$. Otherwise, past effects would accumulate and the values of successive x_t' s would move towards infinity that will lead to unstationarity.

There are two types of stationarity: strictly stationary and weakly stationary. Strictly stationary can be occurred when the statistics of a process x_t must be constant for any time lag k, that is, the statistical properties of x_t and x_{t+k} are the same and probability distribution of the stochastic process is invariant under a shift in time. If k and m represent any integers in time lags, then

$$p(x_t, \ldots\ldots, x_{t+k}) = p(x_{t+m}, \ldots\ldots, x_{t+k+m})$$

(Equation 5)

A weaker form of stationarity requires the mean and auto-covariance of the stochastic process are finite and invariant under a shift in time.

$$\mathbf{E}\, X_t = \mu_t = \mu \qquad \mathrm{Cov}\,(X_t, X_s) = \mathbf{E}\,(X_t - \mu_t)(X_s - \mu_s) = \gamma\,(t,s) = \gamma\,(t - s)$$

(Equation 6)

Usually, most of physical time series are non-stationary or at very most is weakly stationary. Hence, in order to produce stationary time series, first difference of non-stationary time series can be done. This technique is very useful when we want to model stationary time series from the time histories record which are non-stationary. The autocorrelation of the time series produced by taking the first difference of the original time series tends to die off rapidly if this results in a stationary time series.

2.2 Time Series Modeling and Forecasting

There are two types of forecasting methods which are qualitative methods and quantitative methods. Qualitative forecasting methods generally use the opinions of experts in order to get the forecasts. Such methods are usually applied when the historical data are not available. The other type which is quantitative methods involves the historical data in order to predict the future data. One of the quantitative methods that will be involved in this research of studies is univariate forecasting model. This model predicts future values of a time series solely on the basis of the time histories record. When univariate model is used, historical data are analyzed in order to get the data pattern. Then, the data pattern will be extrapolated in order to get the forecasts with the assumption that the pattern will continue. Along this research, Box Jenkins methodology which is one of the univariate models will be used as the forecasting procedure.

2.2.1 Box-Jenkins Model

The Box-Jenkins approach was first described in the high influential book by statisticians George Box and Gwilym Jenkins in 1970. Box Jenkins modeling involves identifying an appropriate ARIMA process which is a mathematical model used for forecasting, fitting it into the data and then using the fitted model for forecasting process.

Originally, Box Jenkins modeling procedure only involves iterative-three stage process which is model selection, parameter estimation and model checking. But recently, two other stages were added to the procedure which is preliminary stage of data preparation and a final stage of model application or forecasting (Makridakis, Wheelwright and Hyndman, 1998).

2.2.2 ARIMA Processes

ARIMA processes are mathematical models used for time series modeling and forecasting. ARIMA is an acronym for Autoregressive, Integrated and Moving Average. Each of these phrases describes a different part of the mathematical model.

ARIMA processes are the most important part of time series analysis. These processes were introduced by George Box and Gwilym Jenkins in the early 1970s and sometimes known as Box Jenkins models. Box and Jenkins (1970) effectively put together in a comprehensive manner the relevant information required to understand and use ARIMA processes.

Each ARIMA process consists of three parts which are the autoregressive (AR) part or ***p***, the integrated (I) part or ***d*** and the moving average (MA) part or ***q***. The models are often written in shorthand as ARIMA ***(p,d,q)***. Next three paragraphs, will discuss more on every part of ARIMA process.

Firstly, autoregressive part describes how each observation is a function of the previous p observations. For example, if $p = 1$, then each observation is a function of only one previous observation. That is,

$$Y_t = c + \varphi_1 Y_{t-1} + e_t$$

(Equation 7)

where Y_t represents the observed value at time t, Y_{t-1} represents the previous observed value at time $t - 1$, e_t represents some random shock or random error and c and φ_1 are both constants. Other observed values of the series can be included in the right-hand side of the equation if $p > 1$:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + e_t$$

(Equation 8)

Next is integrated part that determines whether the observed values are modeled directly, or whether the differences between consecutive observations are modeled instead. For example if $d = 0$, it means that the observations are modeled directly. If $d = 1$, the differencing process is done one time before the observations are modeled. If $d = 2$, the differencing process is done two times before the observations are modeled. Usually, d is rarely more than 2. Integrated part is very useful in order to remove unstationarity since the differenced data are easier to model.

Lastly is moving average part which is a part that describes how each observation is a function based on q (error or random shock). For example, if $q = 1$, then each observation is a function of only one previous error. That is,

$$Y_t = c + \theta_1 e_{t-1} + e_t$$

(Equation 9)

Where e_t represents the random error at time t and e_{t-1} represents the previous random error at time $t - 1$. Other errors can be included in the right-hand side of the equation if $q > 1$.

Combining these three parts gives the diverse range of ARIMA models,

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1 - L)^d X^t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) e_t$$

(Equation 10)

where L is the lag operator, an element of a time series to produce the previous element.

CHAPTER 3

METHODOLOGY

3.1 Box-Jenkins Methodology

Since this research involving time series modeling and forecasting data from time histories, Box-Jenkins methodology will be applied. There are 5 stages in this methodology. Before applying the stages, the data received will be sorted first in order to ensure that the analysis will become smooth. From the sorted data, a time series will be generated from selected parameter with respect to particular time range. After plotting the time series, 5 stages of Box-Jenkins methodology can be applied. The stages are:

1. Data preparation. It involves transformations and differencing. Transformations of the data (such as square roots or logarithms) can help stabilize the variance in a series where the variation changes with the level. Then the data are differenced until there are no obvious patterns such as trend or seasonality left in the data. "Differencing" means taking the difference between consecutive observations, or between observations a year apart. Usually, the differenced data are easier to model rather than original data.
2. Model selection in the Box-Jenkins framework uses various graphs based on the transformed and differenced data to try to identify potential ARIMA processes which might provide a good fit to the data.
3. Parameter estimation which means that the values of the model coefficients which provide the best fit to the data will be determined. To perform this estimation, sophisticated computational algorithms can be used.
4. Model checking. It involves testing the assumptions of the model to identify any inadequacy. If the model is found to be inadequate, it is necessary to go back to Step 2 and try to identify a better model.
5. Forecasting is what the whole procedure is designed to accomplish. Once the model has been selected, estimated and checked, it is usually a straight forward task to compute forecasts by using computer.

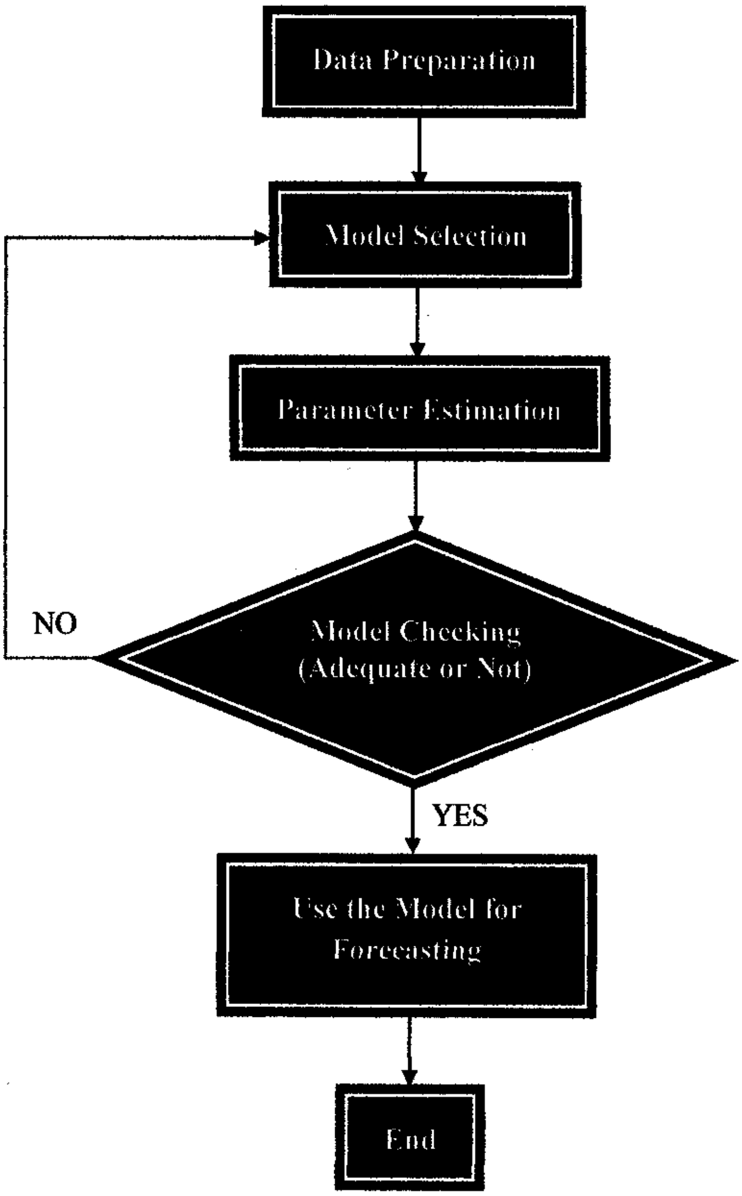


Figure 1: Illustration of Box-Jenkins Methodology

3.2 Project Methodology

First of all, raw measured metocean data got from PETRONAS Carigali Sdn. Bhd. will be filtered to remove all the irrelevant values due to error that come from measurement instrument in order to ensure that the time series analysis and forecasting work can be done smoothly. The table below shows the availability of the data according to month and types of data which are wind speed and significant wave height. The red marked box represents the fault data and will be eliminated from the analysis since the data contain too many irrelevant values.

Table 2: Data availability for each month for all regions

Field	Year	Data	Month											
Dulang (PMO)	2002	Wind Speed	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
		Wave Height	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Samarang (SBO)	2002	Wind Speed	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
		Wave Height	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Tukau (SKO)	2002	Wind Speed	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
		Wave Height	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec

Second, linear interpolation will be done to the less fault data in order to repair the data. Linear interpolation can be done by summing the upper and lower limit of the error value and then take the average of the summation as the replacement for the error value. The interpolation can be performed by using PASW Statistics 18 software.

After modifying the raw metocean data, time series modeling can be done by using ARIMA method. From the output produced by the PASW Statistics 18 software, we can identify each parameter for ARIMA which are autoregressive (***p***), integrated (degree of differencing, ***d***) and moving average (***q***) parameters for wind speed and wave height. Then, these parameters will be summarized into a table in order to determine the similarities and differences between Malay, Sabah and Sarawak Basin and also between monsoon and non-monsoon season.

One example of the time series modeled can be seen as follows:

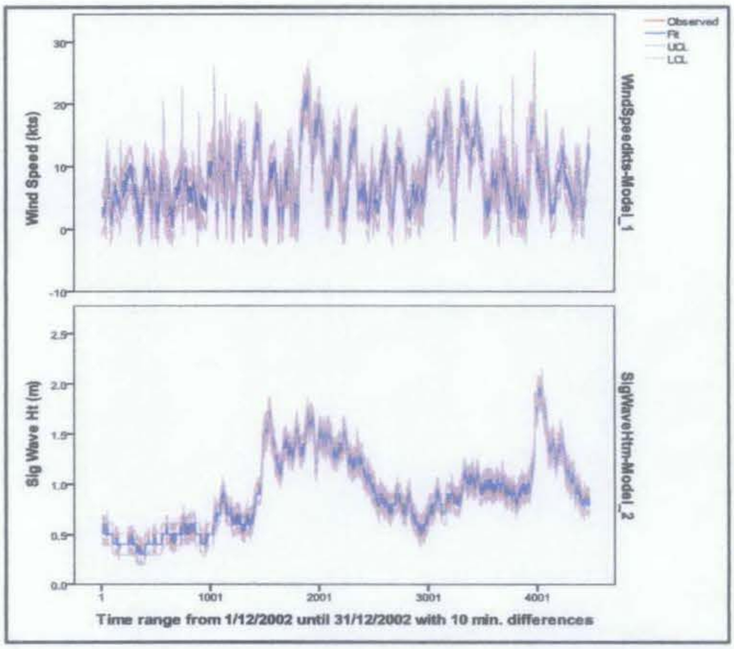


Figure 2: Time series of wind and wave in December 2002 at TKQ-A platform

The final stage of this project is forecasting of the metocean data that will be done for 2 weeks period based on one month data. Then, the forecasted data will be compared with the actual data and the 95% confidence limit (CL) that can be obtained by the summation of the forecasted data and the standard deviation of the actual data.

95% CL = forecasted data + $(2 \times \sigma_{\text{real data}})$ (Equation 11)

Time	Date	WindSpeedkts	SigWaveHtm	Predicted_WindSpeedkts_Model_1	Predicted_SigWaveHtm_Model_2
0:00:00.00	01-Sep-2002	8.50	.70	.	.
0:10:00.00	01-Sep-2002	7.80	.70	8.50	.70
0:20:00.00	01-Sep-2002	8.90	.70	7.90	.70
0:30:00.00	01-Sep-2002	8.50	.70	8.79	.70
0:40:00.00	01-Sep-2002	8.70	.70	8.52	.70
0:50:00.00	01-Sep-2002	9.10	.70	8.66	.70
1:00:00.00	01-Sep-2002	9.30	.70	9.01	.70
1:10:00.00	01-Sep-2002	8.30	.70	9.20	.70
1:20:00.00	01-Sep-2002	9.50	.70	8.39	.70
1:30:00.00	01-Sep-2002	8.70	.70	9.33	.70
1:40:00.00	01-Sep-2002	7.00	.70	8.74	.70

Figure 3: Screenshot of the output from PASW Statistics 18 software shows the forecast values for wind and wave

3.3 Program and Software

3.3.1 PASW Statistics 18 Computer Software

PASW Statistics 18 is the main software that will be used. It can help in order to model the time series and proceed with forecasting the data. It is a comprehensive system used for analyzing system.

3.3.2 Microsoft Office Excel

This software can helps a lot in order to provide precise value in modifying data and plotting graph.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Data Analyzed and Validation

In this research, there are three different locations represent three regions in Malaysia are used. First is Dulang-B platform which is situated in the Dulang field represents Malay Basin. Second is TKQ-A platform which is represents Sarawak Basin situated in Tukau field. Lastly is SMQ-A platform which is located in the Samarang field represents Sabah Basin. These three platforms have its own function and all measurements are done in these platforms to get the metocean data that shows every region's metocean condition and situation.

The following table gives the details of each platform mentioned.

Table 3: Details of Dulang-B, TKQ-A and SMQ-A platform

DETAILS	DULANG (PMO)	TUKAU (SKO)	SAMARANG (SBO)
Platform	Dulang-B	TKQ-A	SMQ-A
Field	Dulang	Tukau	Samarang
Platform Type	Fixed Steel Jacket	Fixed Steel Jacket	Fixed Steel Jacket
Platform Function	Production, Drilling, Accommodation	Accommodation	Accommodation
Installed	1/1/1990	1/1/1982	1/1/1984
Oil Prod. (avg)	12428 BOPD	0.0 BOPD	0.0 BOPD
Gas Prod. (avg)	0.0 MCFD	0.0 MCFD	0.0 MCFD
Latitude	5° 49' 44.947" N	4° 24' 47.181" N	5° 37' 07.686" N
Longitude	104° 09' 25.900" E	113° 43' 40.767" E	114° 53' 18.959" E
Water Depth	79.2 m	46.3 m	10.1 m

According to Malaysian Meteorological Department, the northeast monsoon usually commences in the month of November and ends in March. This monsoon actually gives a significant impact to the wind and wave condition of the three regions mentioned before. Plus, during non-monsoon season, typhoons that frequently developed over west Pacific and head to the Philippines region will affect the wind and wave condition at the northwest coast of Sabah and Sarawak. So, the characteristics of these situations hopefully can be seen during the analysis of the time series modeled for Malay, Sabah and Sarawak Basin.

4.2 Time Series Modeling

4.2.1 Autocorrelation of the Time Series

The autocorrelation function plots of a time series can give clear visual about how to understand the nature of the time series. If the original time series is already stationary, the autocorrelation function plot shape of the original time series and first difference time series will be identically same. In the other words, both original and first differenced series will cut off and die off rapidly as the time lags increasing. On the other hand, for non-stationary time series, the autocorrelation function plot tends to dies off slowly along with the increasing time lags. Only after first differencing, the autocorrelation of the time series tends to dies off rapidly; hence, this results in a stationary time series. In an autocorrelation function plot, two horizontal lines will represent to standard errors. These two lines are located slightly above and below zero correlation horizontal line.

The following are the example of the autocorrelation plot of the original and first differenced time series for wind speed and wave height in January 2002 at Dulang-B platform (Malay Basin).

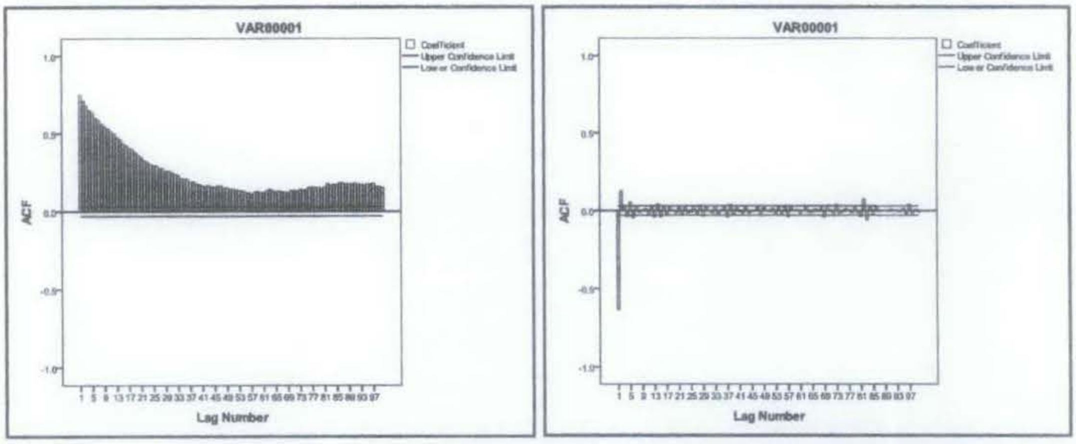


Figure 4: Autocorrelation plot of the original (left side) and first differenced (right side) time series for wind in January 2002 at Dulang-B platform

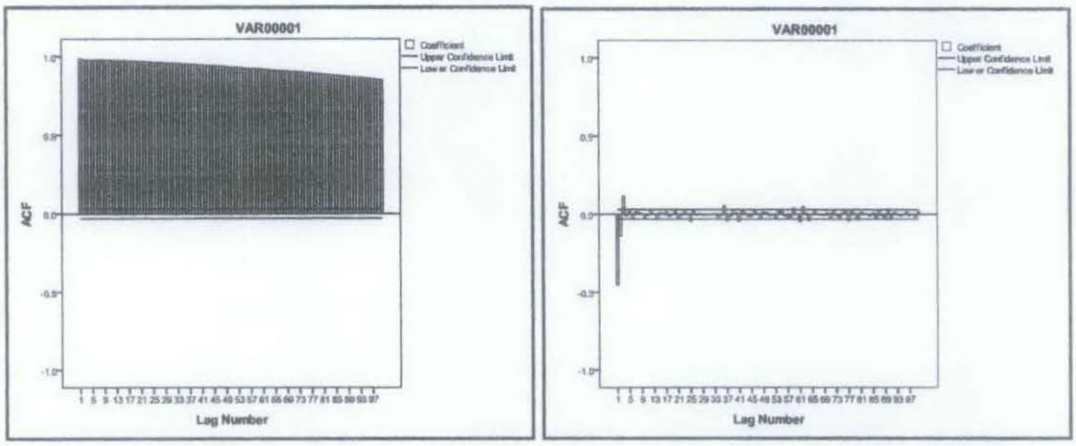


Figure 5: Autocorrelation plot of the original (left side) and first differenced (right side) time series for wave in January 2002 at Dulang-B platform

From the autocorrelation plots above, we can classify that time series for wind speed is weakly stationary (more stationary than time series for wave height) as we can see that the autocorrelation plot for original wind speed time series easily dies off if compared to wave height time series. So, the time series for wave height can be defined as non-stationary.

4.2.2 Time Series for Wind and Wave

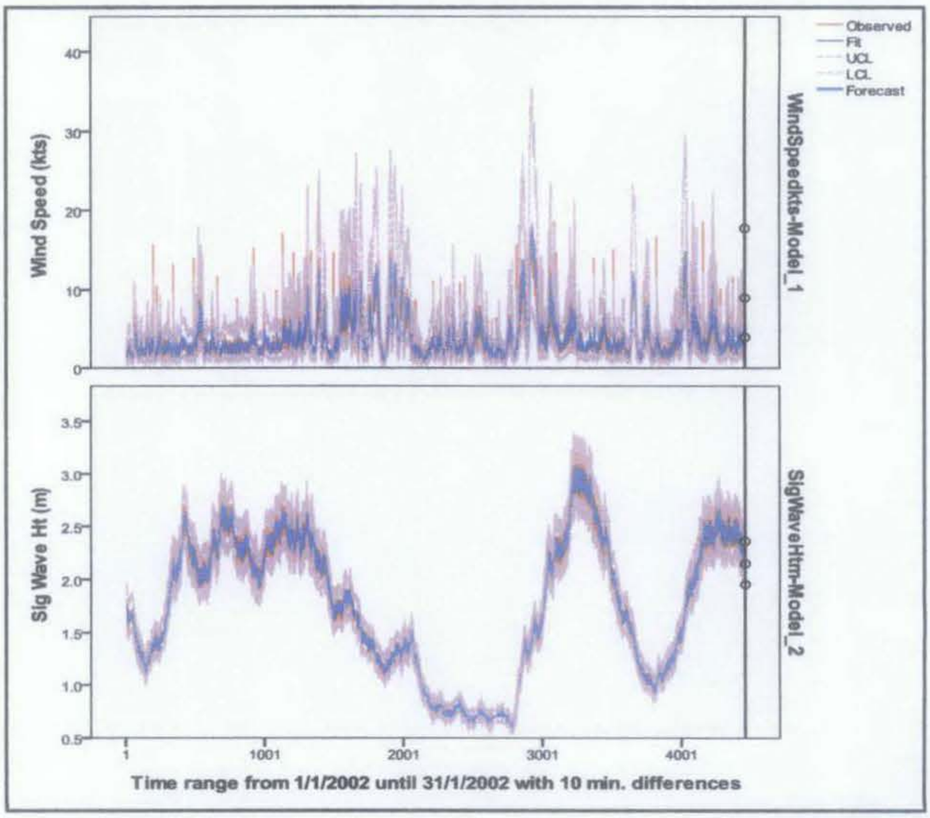


Figure 6: Time series of wind and wave in January 2002 at Dulang-B platform

Above is the example of time series for wind speed and wave height in January 2002 at Dulang-B platform modeled using ARIMA method. For other months and platforms, the time series are attached in the appendices section at the end of this report. From the time series modeled, we can see that the pattern for wind time series is more stationary than wave. For overall observation, we know that time series analysis for wind and wave can be regarded as stochastic processes since the most obvious characteristic of the time series is random in nature and cannot be duplicated or resemble to one another. In fact, time series usually provides important statistical information regarding the physical phenomena represented by the time series.

4.3 Parametric Comparison

After obtaining the time series for wind speed and wave height for every month in the year 2002, the ARIMA parameters will be summarized into a table to make the parametric comparison between Malay, Sabah and Sarawak Basin, plus, between monsoon and non-monsoon season.

These are the summary of the ARIMA parameters for both wind and wave time series:

Table 4: Summary of the ARIMA parameters for wind time series for each region

Month	Dulang (PMO)			Samarang (SBO)			Tukau (SKO)		
	ARIMA Parameters								
	AR (p)	I (d)	MA (q)	AR (p)	I (d)	MA (q)	AR (p)	I (d)	MA (q)
January	0	1	4	0	1	3	4	1	4
February	-	-	-	0	1	12	0	1	4
March	1	1	1	0	1	2	0	1	1
April	0	1	7	0	1	10	-	-	-
May	-	-	-	0	1	4	0	1	14
June	1	1	9	2	1	2	1	1	6
July	0	1	10	-	-	-	1	1	15
August	1	1	16	-	-	-	1	1	2
September	0	1	11	2	1	1	1	1	14
October	0	1	7	0	1	6	2	1	2
November	0	1	8	-	-	-	1	1	1
December	0	1	2	-	-	-	1	1	1

Notes: 1. Green-highlighted rows represent monsoon season

2. "-" represents the data is not available for the modeling

Table 5: Summary of the ARIMA parameters for wave time series for each region

Month	Dulang (PMO)			Samarang (SKO)			Tukau (SKO)		
	ARIMA Parameters								
	AR (p)	I (d)	MA (q)	AR (p)	I (d)	MA (q)	AR (p)	I (d)	MA (q)
January	0	1	17	0	1	7	0	1	11
February	-	-	-	0	1	5	-	-	-
March	1	1	9	0	1	4	1	1	9
April	1	1	3	0	1	10	-	-	-
May	-	-	-	1	1	9	0	1	2
June	0	1	3	0	1	2	0	1	8
July	0	1	2	-	-	-	2	1	14
August	1	1	3	-	-	-	0	1	3
September	0	1	6	1	1	12	0	1	2
October	0	1	2	0	1	2	1	1	3
November	1	1	3	-	-	-	1	1	3
December	0	1	10	-	-	-	0	1	10

Notes: 1. Green-highlighted rows represent monsoon season

2. "-" represents the data is not available for the modeling

4.3.1 Comparison between Malay, Sabah and Sarawak Basin

From the summary above, we can define that wave exhibit higher irregularity and variation during monsoon season at Dulang field (Malay Basin) based on the moving average (MA) parameters since the value range is the highest among the other region. It means that the time series for wave at Malay Basin highly depends on the previous noise that leads to the irregularity and variety of the data.

Tukau field (Sarawak Basin) is expected to experience higher sustain wind speed during monsoon and non-monsoon period as compared with Dulang field (Malay Basin) and Samarang field (Sabah Basin) based on the autoregressive (AR) parameter that sustain at one value for almost every month in the year 2002.

Because of the typhoon that developed over west Pacific and head towards Philippines region during non-monsoon season, the northwest coast of Sabah and Sarawak will be affected too since the location of Sabah and Sarawak near the Philippines region. From the ARIMA parameters summary for wave time series above, we can see that the value of the moving average (MA) parameters for Sabah and Sarawak Basin is in the larger range if compared to the Malay Basin. Hence, we can justify that typhoon near the Philippines region affect the wave condition at Sabah and Sarawak Basin that leads to the irregularity and variation of the wave.

Generally, the wind speeds are uniform throughout the three regions (Malay, Sabah and Sarawak Basin) based on the high value of the autoregressive (AR) and moving average (MA) parameters.

4.3.2 Comparison between Monsoon and Non-Monsoon Season

We can identify from the summary above that wind during non-monsoon is proven to be non-stationary and during monsoon to be weakly stationary since the moving average (MA) parameters for wind during non-monsoon is generally very high if compared to monsoon season. Thus, it is clearly seen that past noise really affected the wind speed time series that lead to the unstationarity. Logically, during non-monsoon season, the wind is not continuously come to the area unlike during monsoon season which is the wind constantly come to the area.

On the other hand, wave is non-stationary during both monsoon and non-monsoon season. This is because the moving average (MA) parameters for wave height time series is constantly in large value for both seasons if compared to the wind speed time series that only has the large value for moving average (MA) parameters during non-monsoon season.

4.4 Forecasting Results

By implementing ARIMA method, forecasting future environmental loads (i.e. wind and wave) can be done according to the desired period of time e.g. one day, one week or one month. Plus, this method provides 95% confidence limit from the forecast so that it will give better description for the upcoming condition of the environmental loads. However, to perform long-term forecasting, the data need to be updated regularly as the characteristics and statistical properties of the upcoming environmental loads will vary along with time.

The following figures show the forecasts and the confidence limits for two weeks for month October 2002 at Samarang field (Sabah Basin).

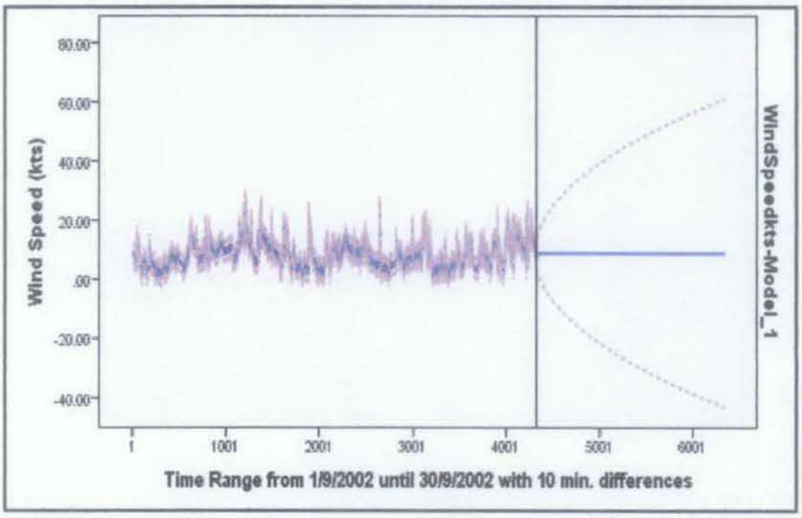


Figure 7: Time series of the wind speed with 2 weeks forecasts and confidence limits for October 2002 at SMQ-A platform

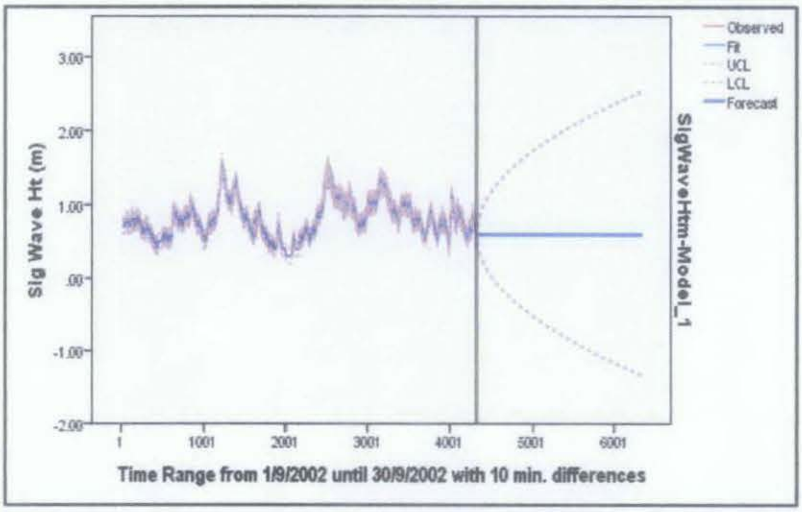
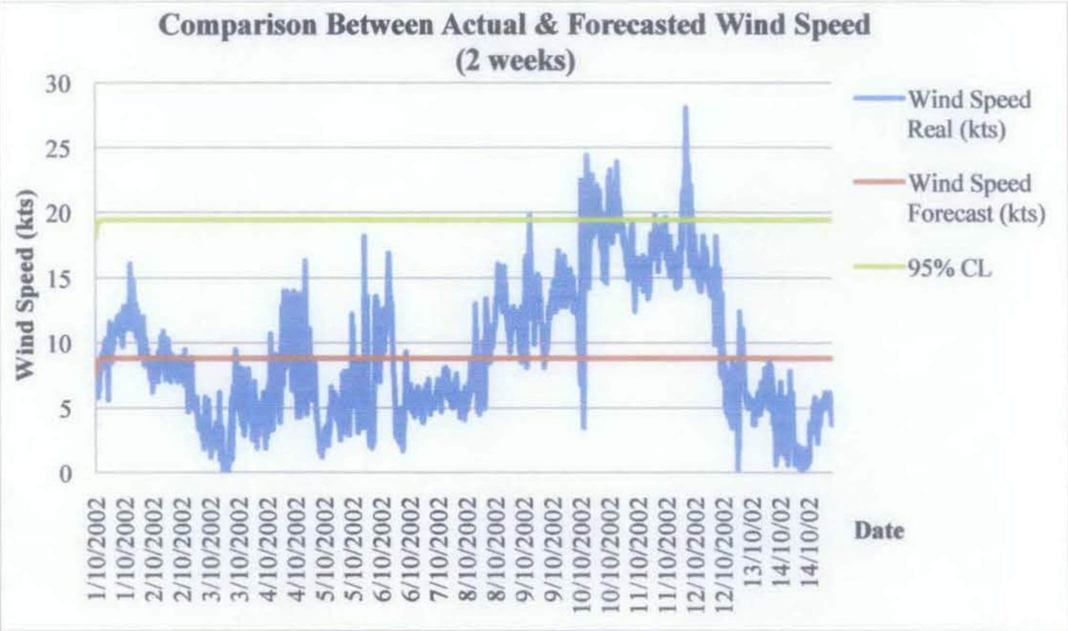
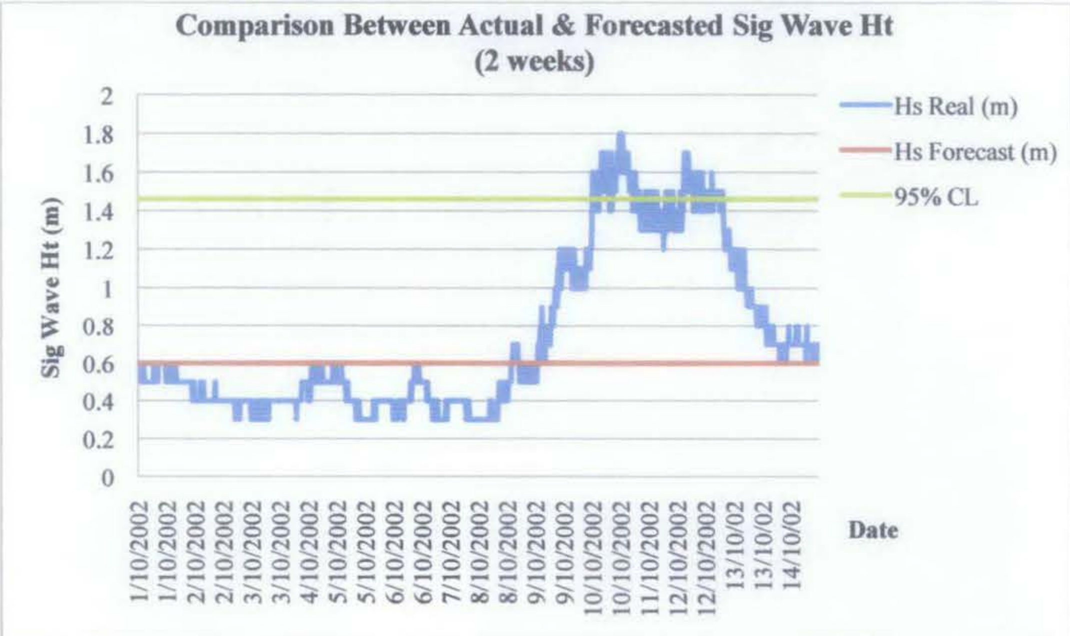


Figure 8: Time series of the wave height with 2 weeks forecasts and confidence limits for October 2002 at SMQ-A platform

For the next 2 figures, we can see the comparison between the actual measured metocean data with the forecasts data. Plus, we can identify where the 95% confidence limit is for each environmental load.



Graph 1: Comparison between actual and forecasted wind speed for 2 weeks in month of October 2002 at SMQ-A platform



Graph 2: Comparison between actual and forecasted wave height for 2 weeks in month of October 2002 at SMQ-A platform

From the graphs above, we can identify that almost 95% of overall actual data falls under the 95% confidence limit from the forecast. So, in order to consider worst case scenario for obtaining extreme design values for offshore engineering design and operations, we can take the 95% confidence limit from the forecast by using ARIMA method.

4.5 Comparison between Current Practices and ARIMA Method

With the development of new oil fields, the design of the offshore structures will be emphasized on its design efficiency and economics due to escalating material cost. In order to obtain maximum quality and extreme design value of the offshore structures, it is very important to analyze and study the interaction and current condition of the environmental loads (i.e. wind and wave). In Malaysia, the usage of PTS and API code is very wide and this practice is not updated and conservative since it is based on many past years of criteria and not regularly updated.

By applying ARIMA method in order to predict and get the current statistical properties of the environmental loads involved, it will be easier for us in order to get the required design needed for a certain condition or event based on current situation without forgetting the worst case scenario (extreme event) of the certain location.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

There are several points can be concluded from the overall analysis of this project. First, wind and wave are clearly stochastic (non-stationary) processes as we can see from the autocorrelation plots and the summary for ARIMA parameters. So, in order to improvise and make the analysis to become easier, first differencing needs to be performed before proceeding with the analysis of the time series.

Second, wind is proven to be non-stationary during non-monsoon season and weakly stationary during monsoon season as we can see from the moving average parameters for wind are higher during non-monsoon rather than monsoon season. Plus, if we think logically, the wind will be constantly come to the area during monsoon season if compared to non-monsoon season.

Third, wave is non-stationary during monsoon and non-monsoon season because of the moving average (MA) parameters are constantly high in the monsoon and non-monsoon. Thus, wave is practically highly depends on the previous noise if compared to wind that highly depends on previous noise during non-monsoon season.

Fourth, ARIMA (p,d,q) can be used to forecast future environmental loads (i.e. wind and wave) by providing expected and sustainable parameters. Furthermore, it also provides 95% confidence limit from the forecast to preclude any possibility of exceedence.

5.2 Recommendations

There are some recommendations can be taken into account in order to improvise the findings of this research. First of all, other environmental load (i.e. current) should be considered during the analysis to ensure that we can study the interaction between current and the offshore structures, plus, we can further analyze the metocean properties of a certain location very well.

Second, additional platform and data with minimum error are needed so that more precise analysis for every region can be done. In addition, the characteristics and statistical properties of every region can be seen clearly since the data provided come from many locations in a particular region. So, stronger conclusion and justification can be produced from the analysis of the various locations.

Third, longer period of metocean data is needed to detect the seasonal trend happened in the ocean which is El Nino and La Nina events that occurred once in every eight to nine years.

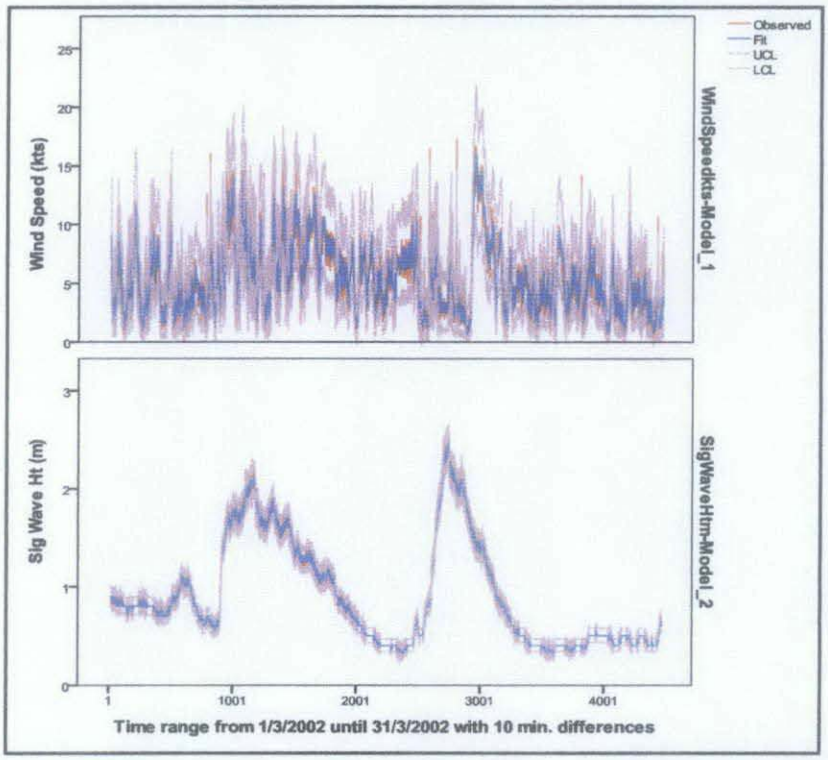
Finally, in the future, this method can be used to benchmark against current practices by oil and gas industry in their offshore operations such as launching, operation, maintenance and in establishing design criteria within certain tolerance.

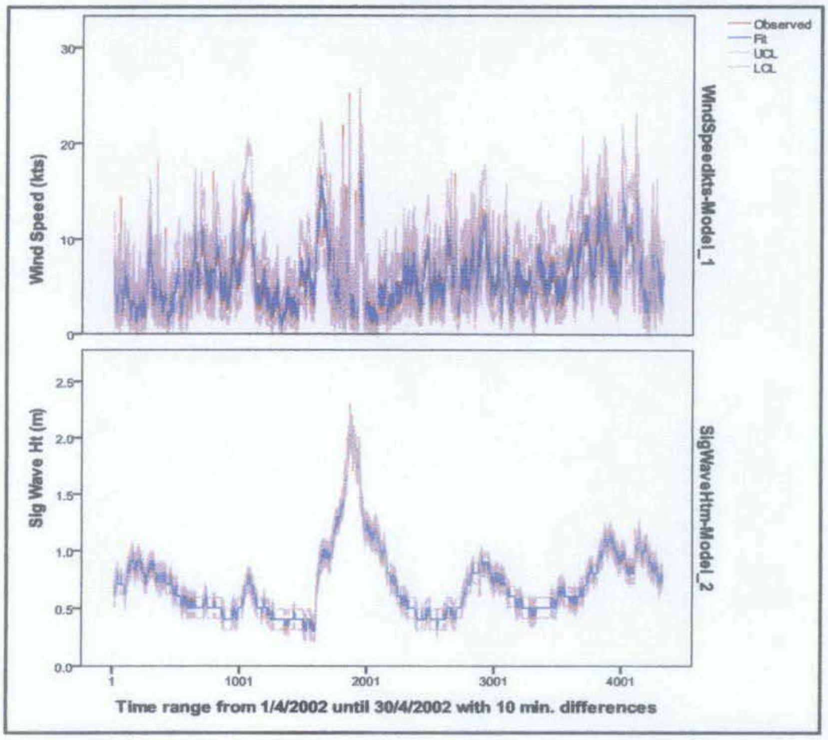
REFERENCES

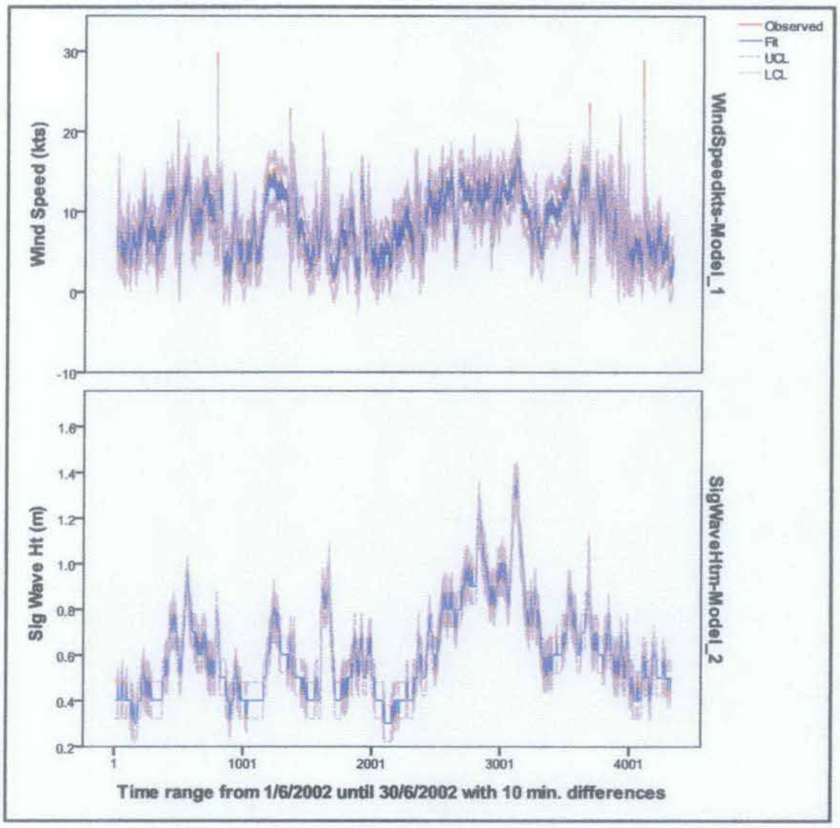
1. James Parker, 2010, *The Importance of Metocean Data at Marine Renewable Energy Sites*, Gardline Environmental Ltd.
2. Trevor Gilbert, July 1998, *Maritime Response Operations - Requirements for Metocean Data and Services,* Conference and Workshop on Meteorological and Oceanographic Services for Marine Pollution Emergency Response Operations.
3. W. Cooper, A. Saulter, P. Hodgetts, 2008, *Guideline for the Use of Metocean Data Through the Life Cycle of a Marine Renewable Energy Development*, Ciria.
4. S.H. Liew, May 1988, *Statistical Analysis of Wind Loadings and Responses of a Transmission Tower Structure*, Texas Tech University, USA.
5. George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, 2008, *Time Series Analysis: Forecasting and Control,* John Wiley & Sons, Inc., Canada.
6. Rob J. Hyndman, May 2001, *Box Jenkins Modelling*.
7. Rob J. Hyndman, May 2001, *ARIMA Processes*.
8. Bowerman, O'Connell, Koehler, 2005, *Forecasting, Time Series and Regression: An Applied Approch, 4th Edition,* Duxbury Applied Series, Thomson Brooks/Cole.
9. J. Scott Armstrong, 1984, *Forecasting by Extrapolation: Conclusion from 25 Years of Research,* University of Pennsylvania.
10. www.statsoft.com/textbook/time-series-analysis/
11. www-stat.wharton.upenn.edu/~stine/stat910/.../02_stationarity.pdf
12. www.met.gov.my/index.php?option=com_content&task=view&id=75&Itemid=1089&limit=1&limitstart=0

APPENDICES

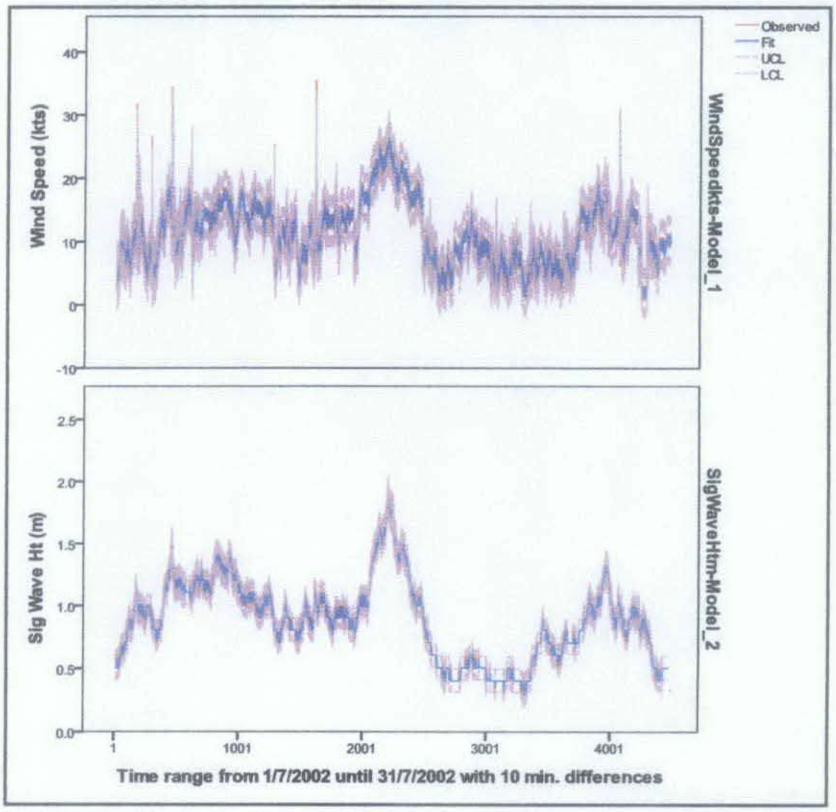
Time Series for Dulang-B platform from March 2002 until December 2002



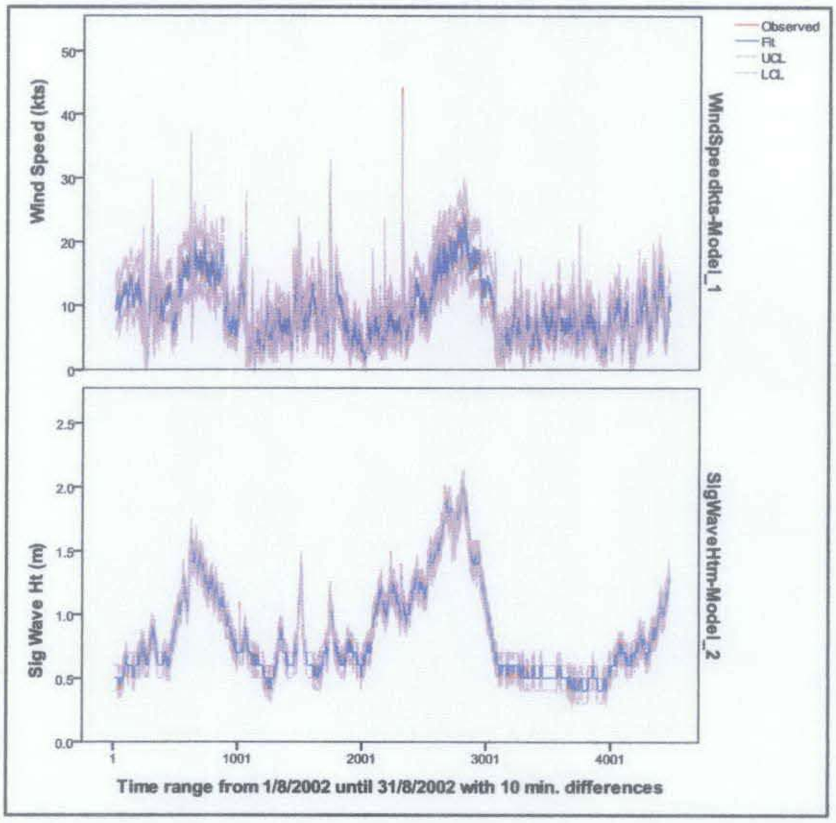


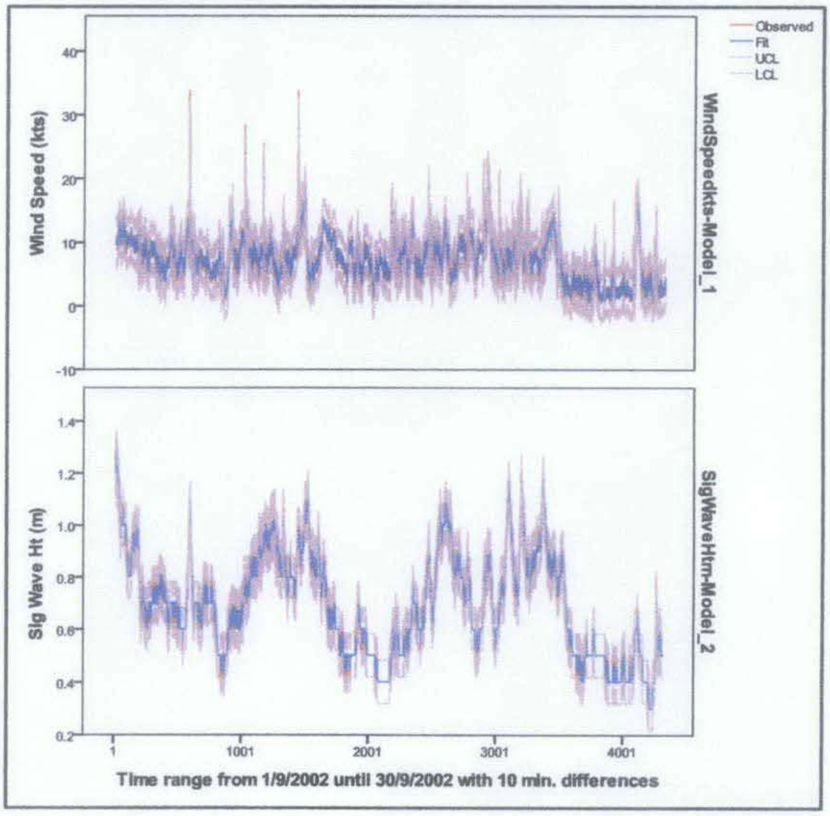


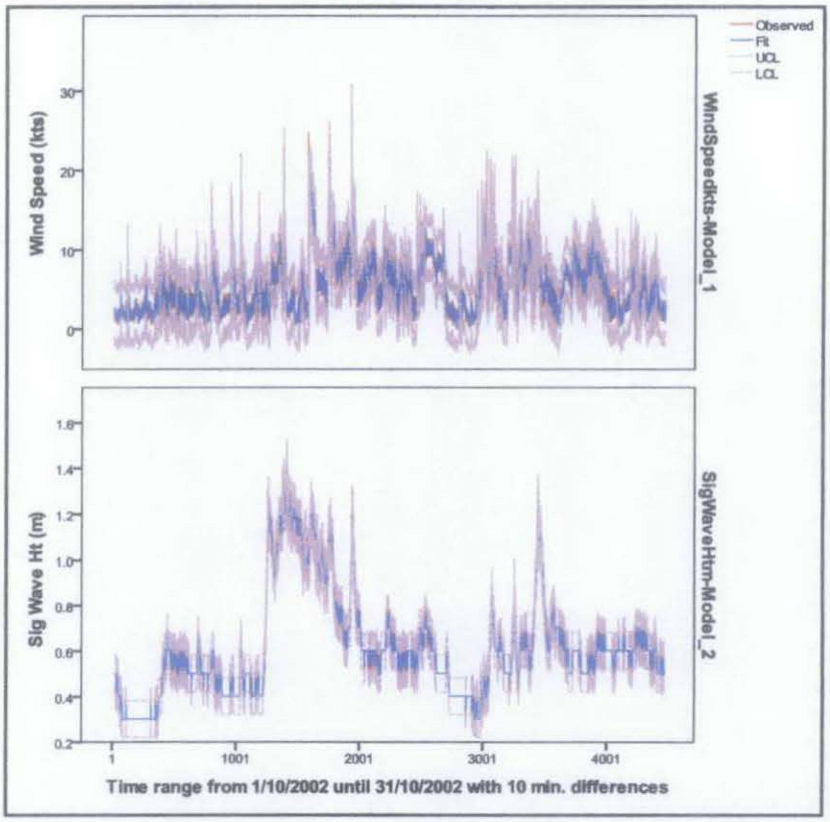
Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/6/2002 until 30/6/2002 with 10 min. differences

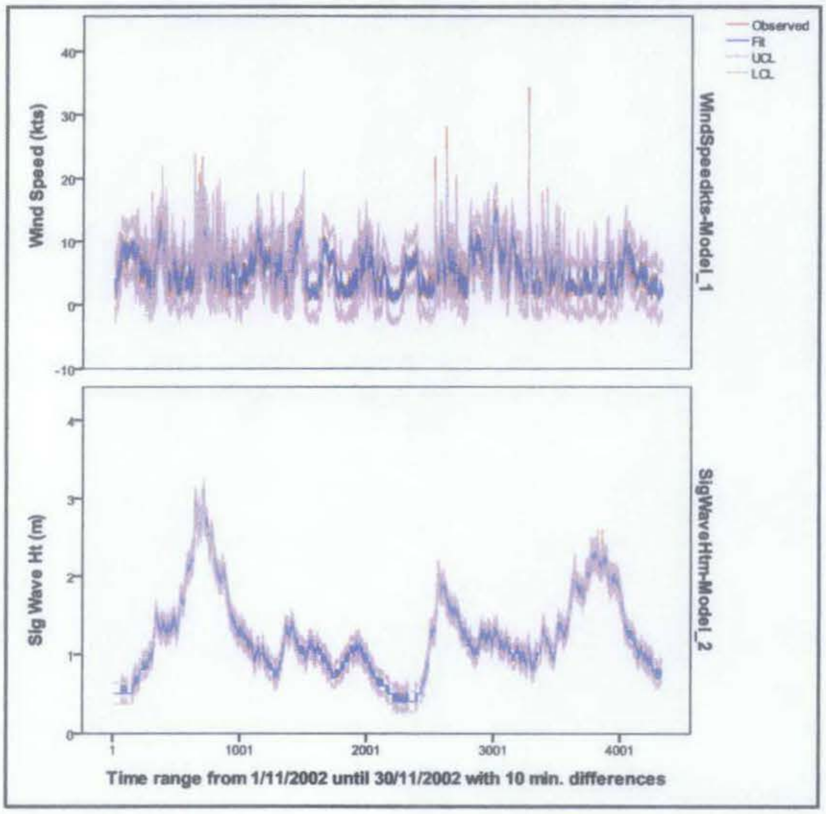


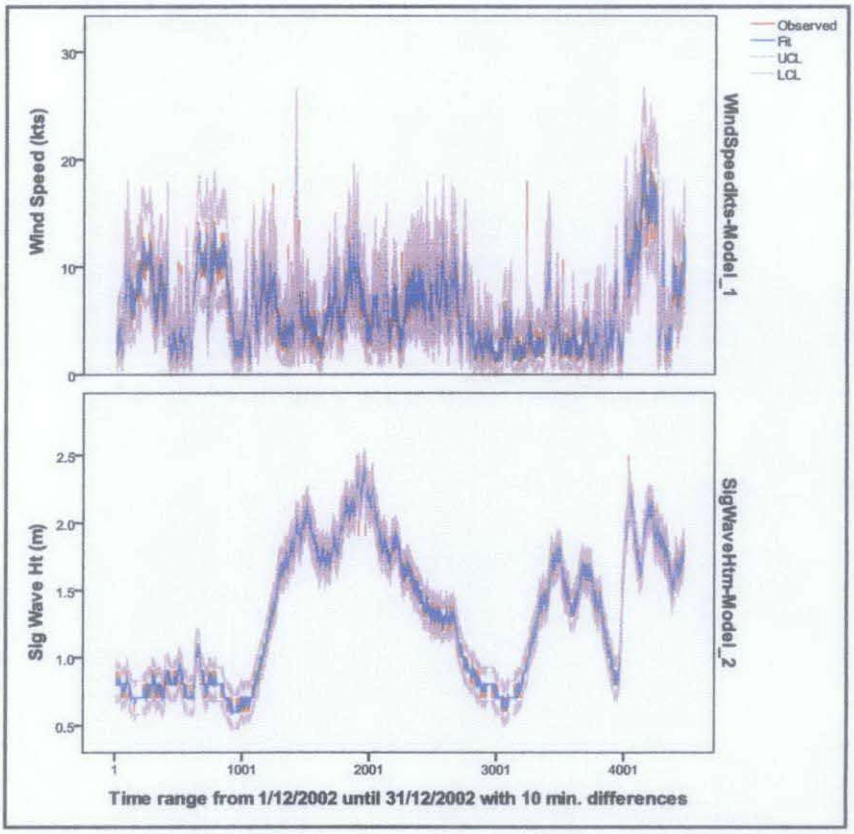
Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/7/2002 until 31/7/2002 with 10 min. differences


Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/8/2002 until 31/8/2002 with 10 min. differences

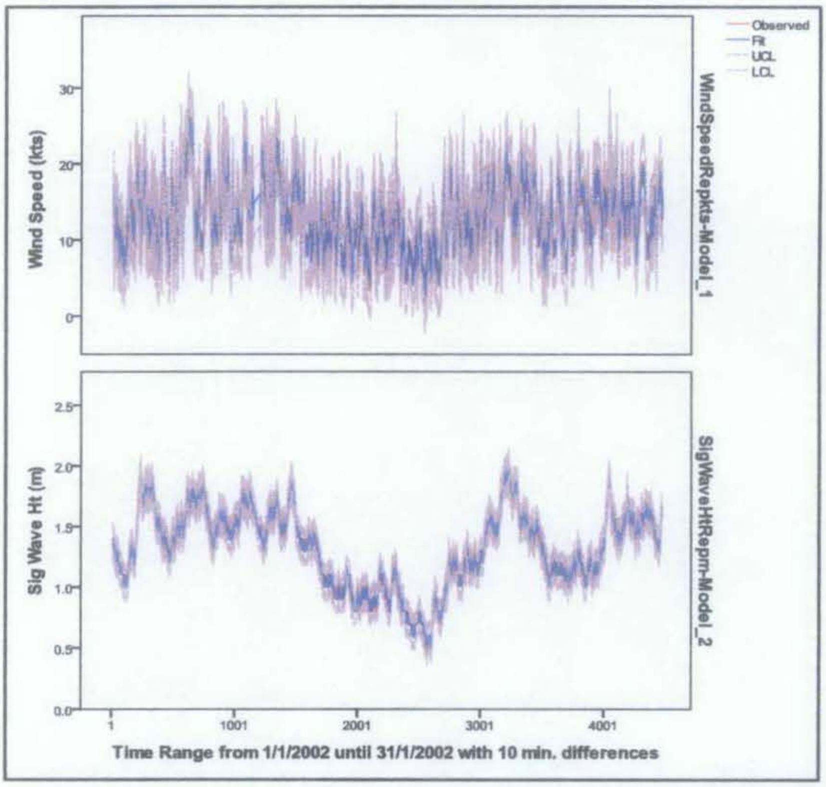

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/9/2002 until 30/9/2002 with 10 min. differences

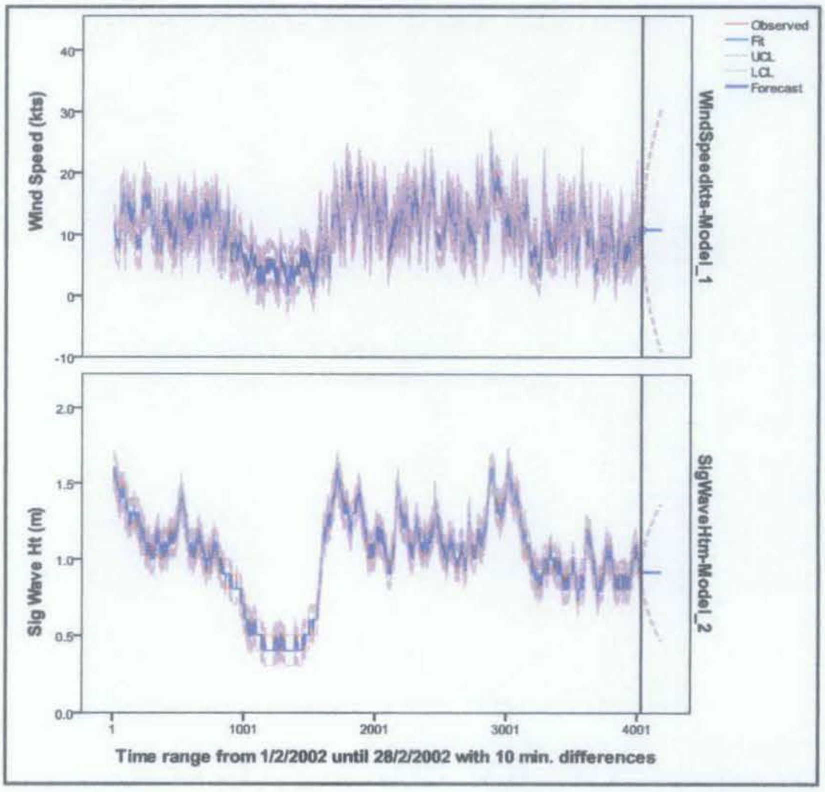

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/10/2002 until 31/10/2002 with 10 min. differences

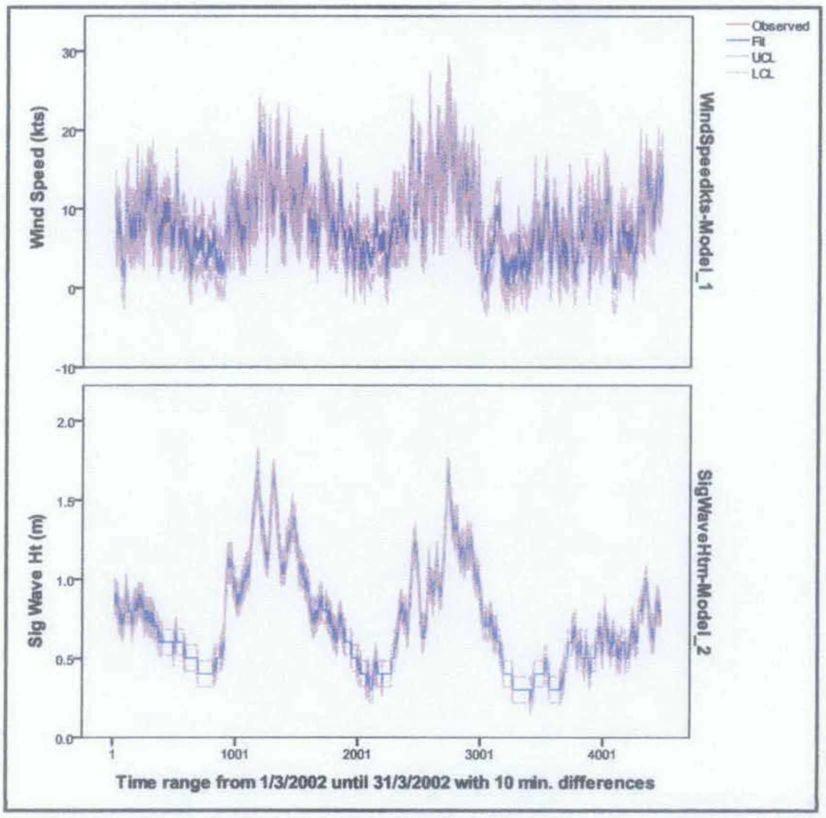

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/11/2002 until 30/11/2002 with 10 min. differences

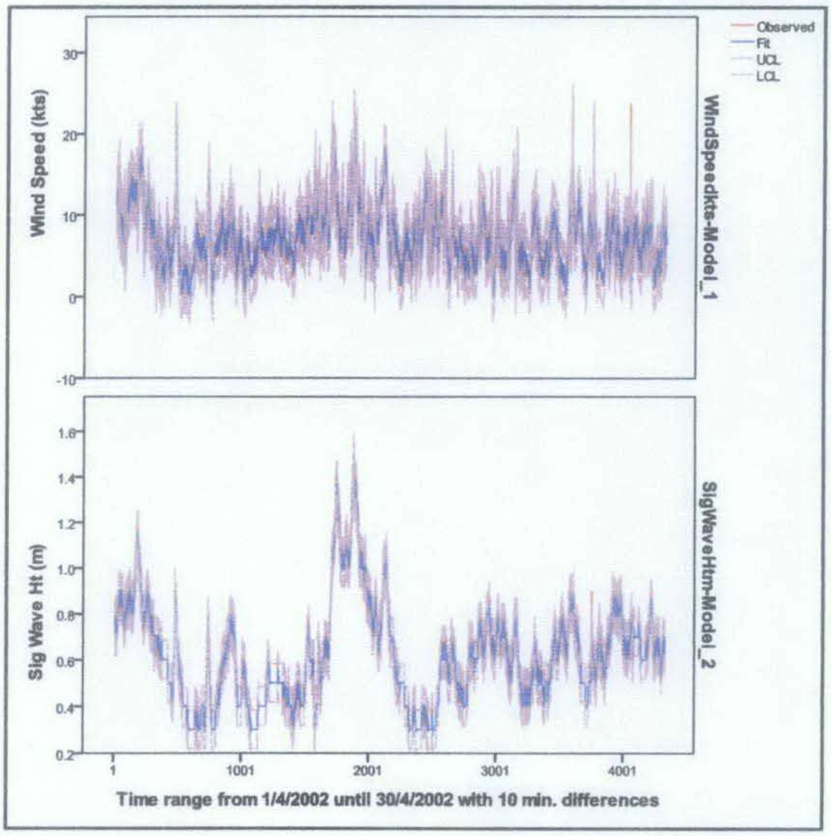

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/12/2002 until 31/12/2002 with 10 min. differences

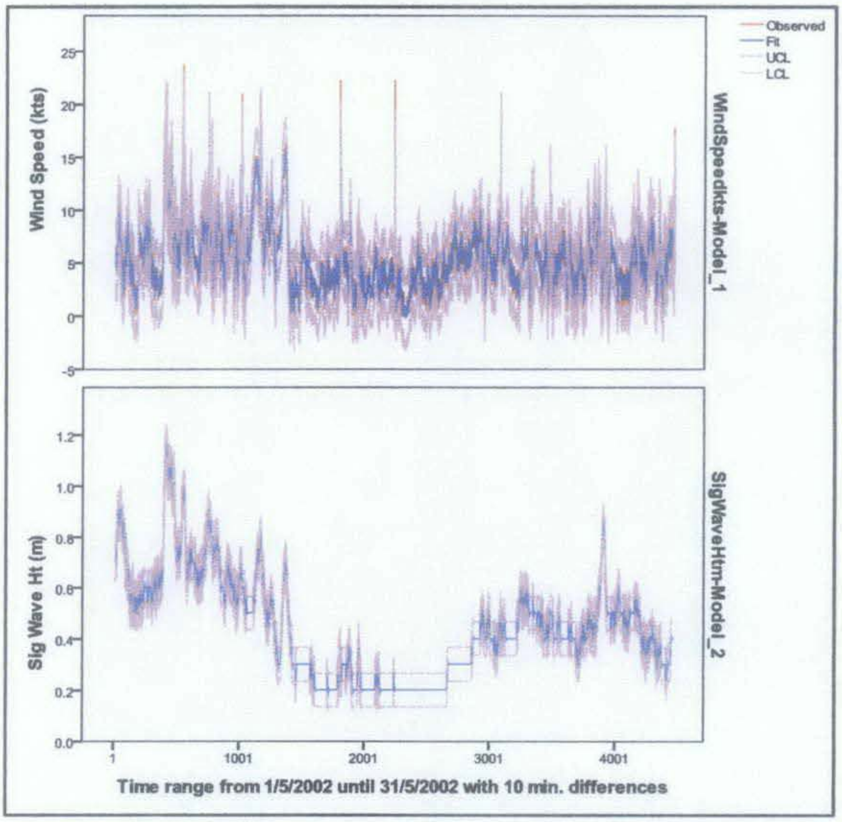
Time Series for SMQ-A platform from January 2002 until October 2002

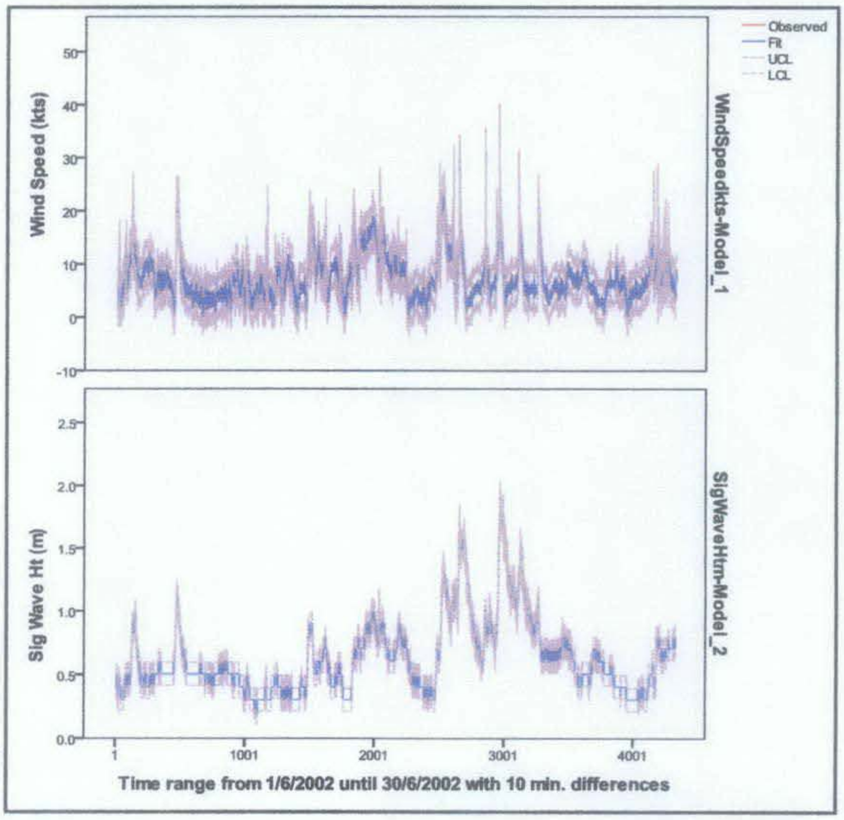


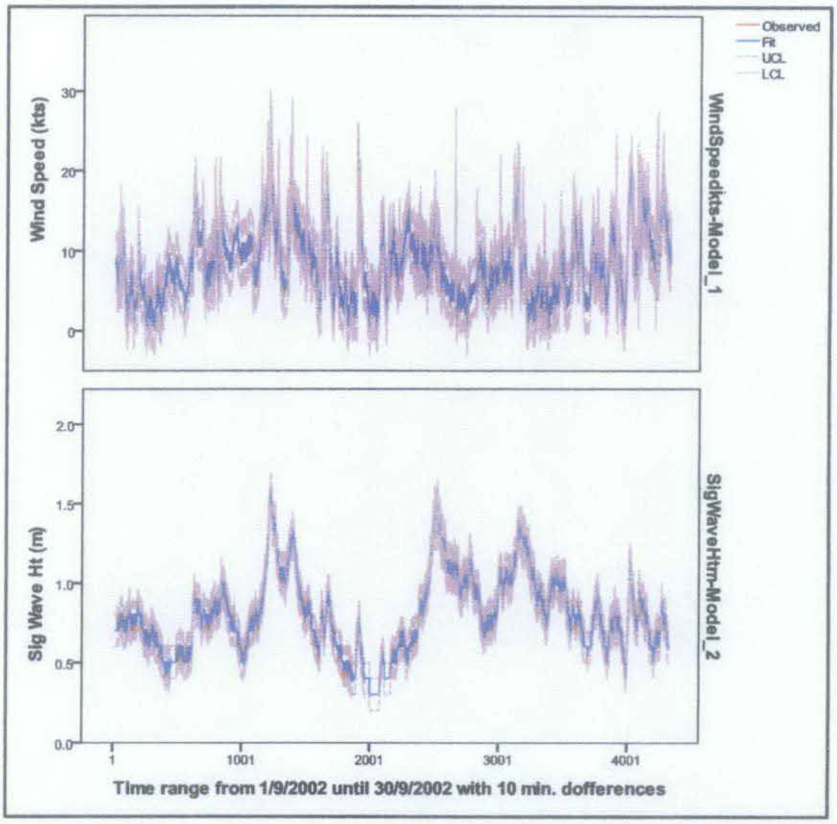


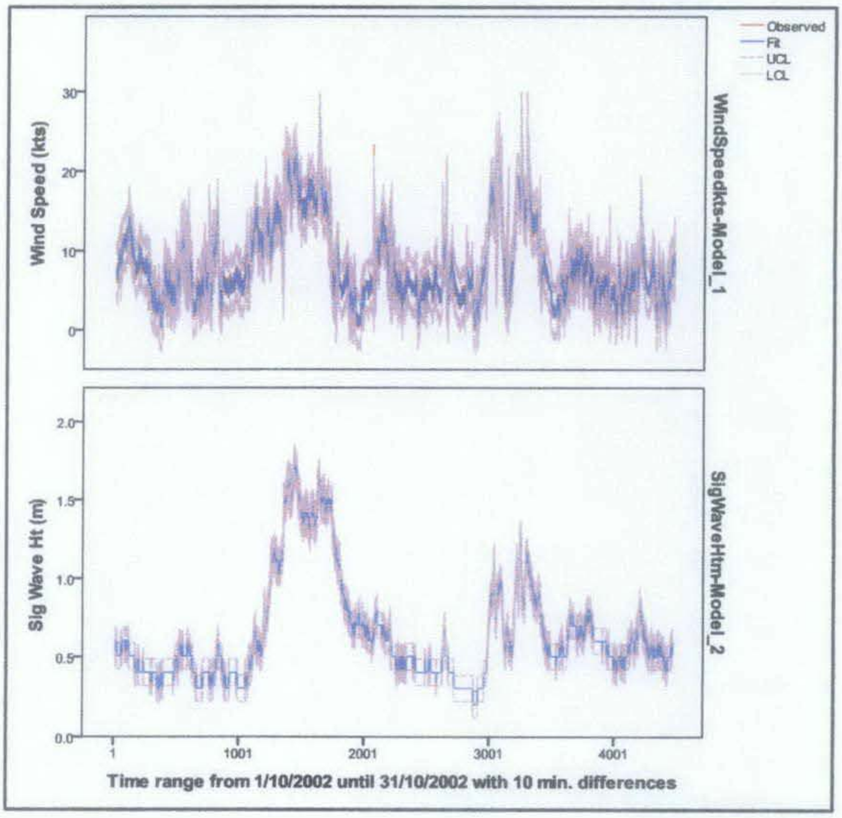

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/3/2002 until 31/3/2002 with 10 min. differences


Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/4/2002 until 30/4/2002 with 10 min. differences

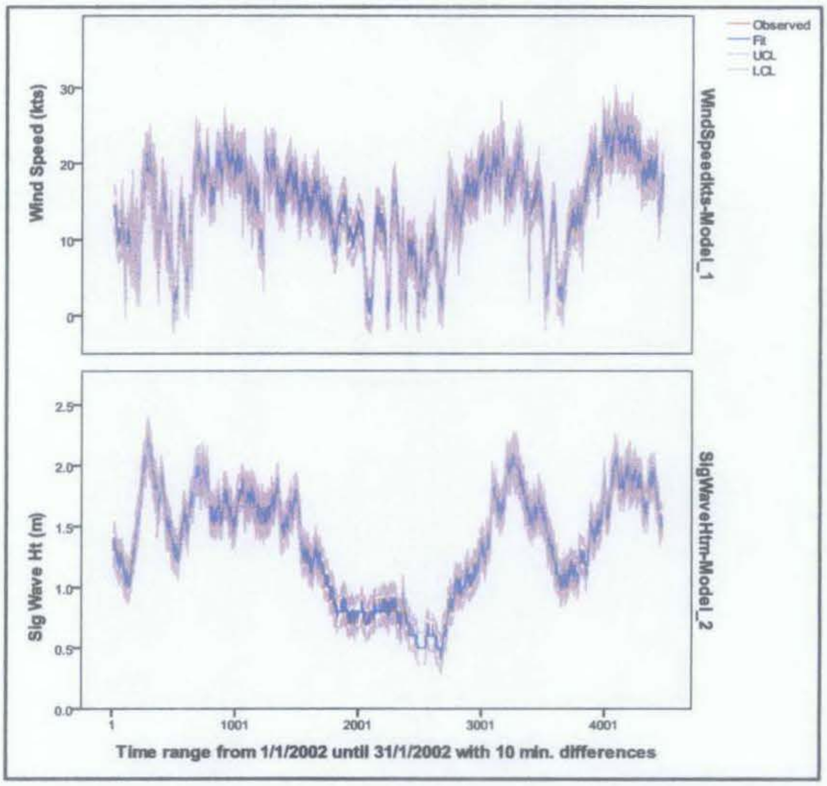

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/5/2002 until 31/5/2002 with 10 min. differences

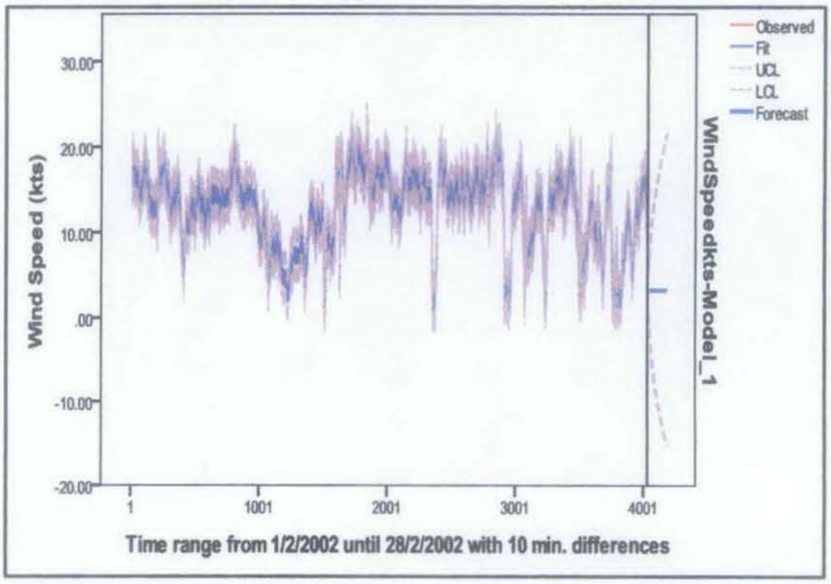

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/6/2002 until 30/6/2002 with 10 min. differences

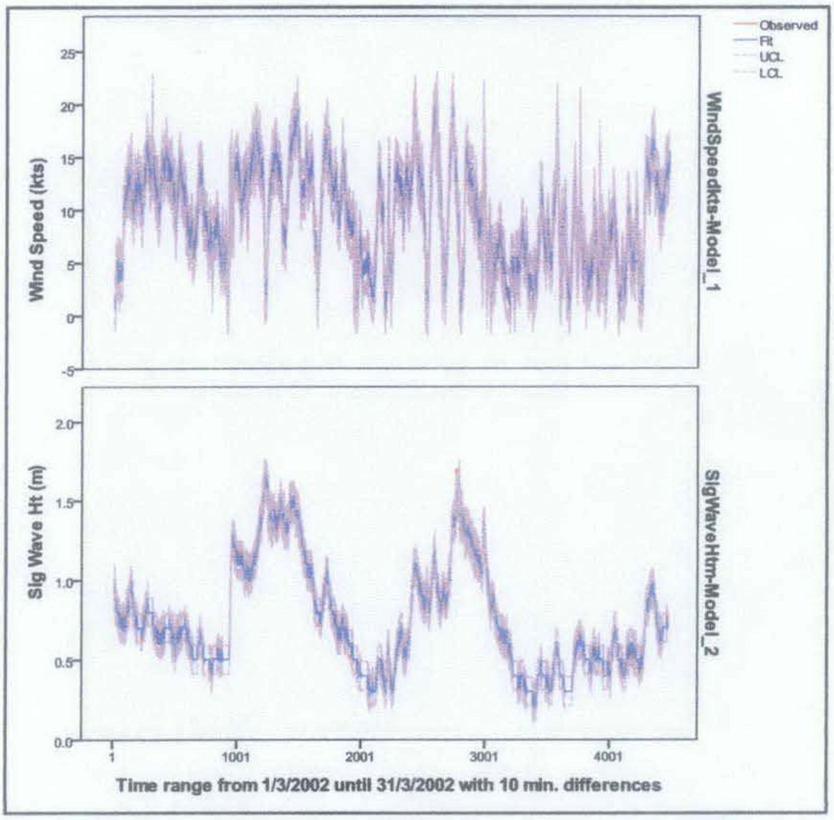

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/9/2002 until 30/9/2002 with 10 min. dofferences

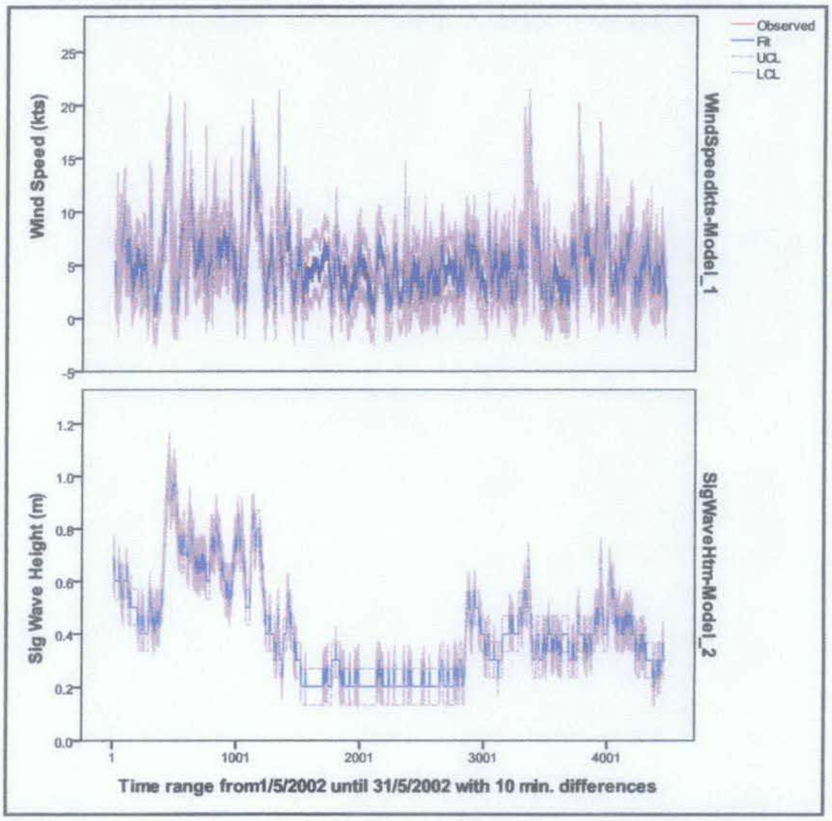

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/10/2002 until 31/10/2002 with 10 min. differences

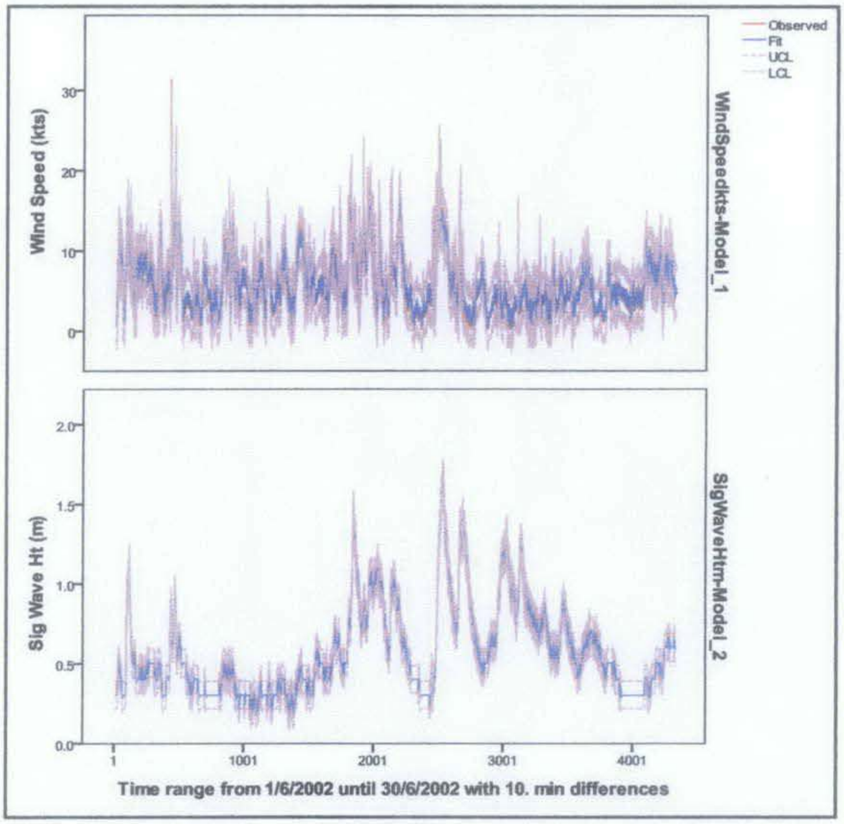
Time Series for TKQ-A platform from January 2002 until November 2002

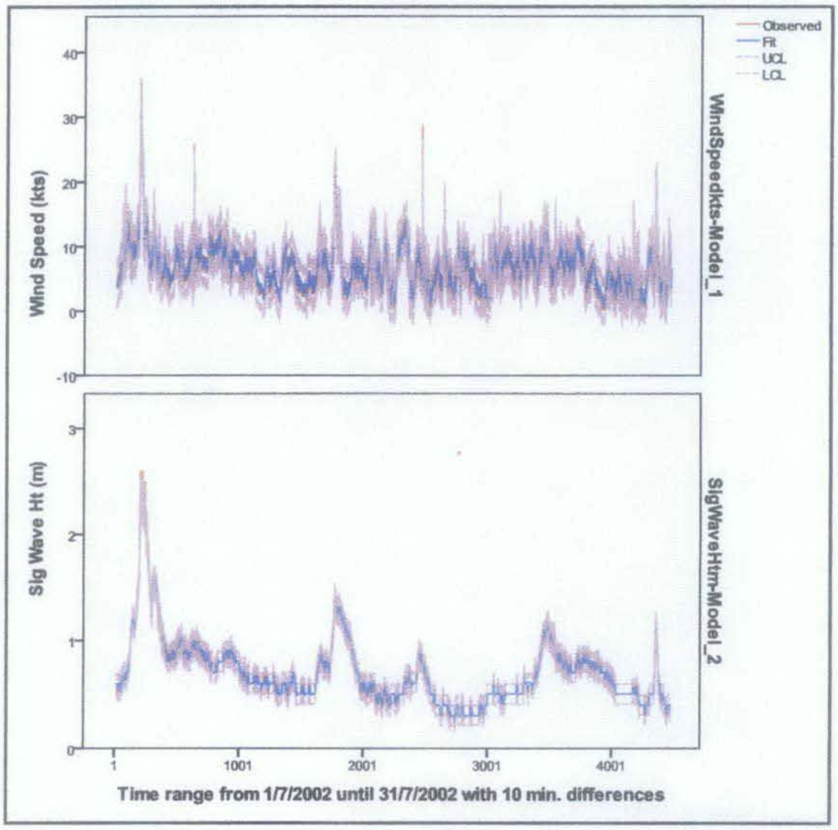


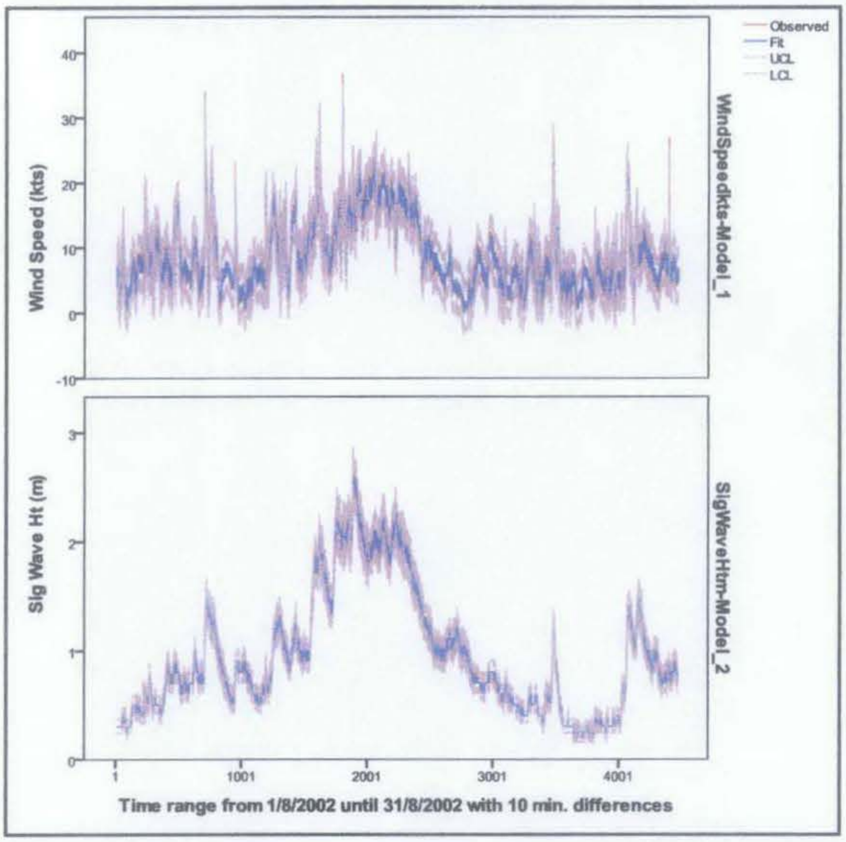


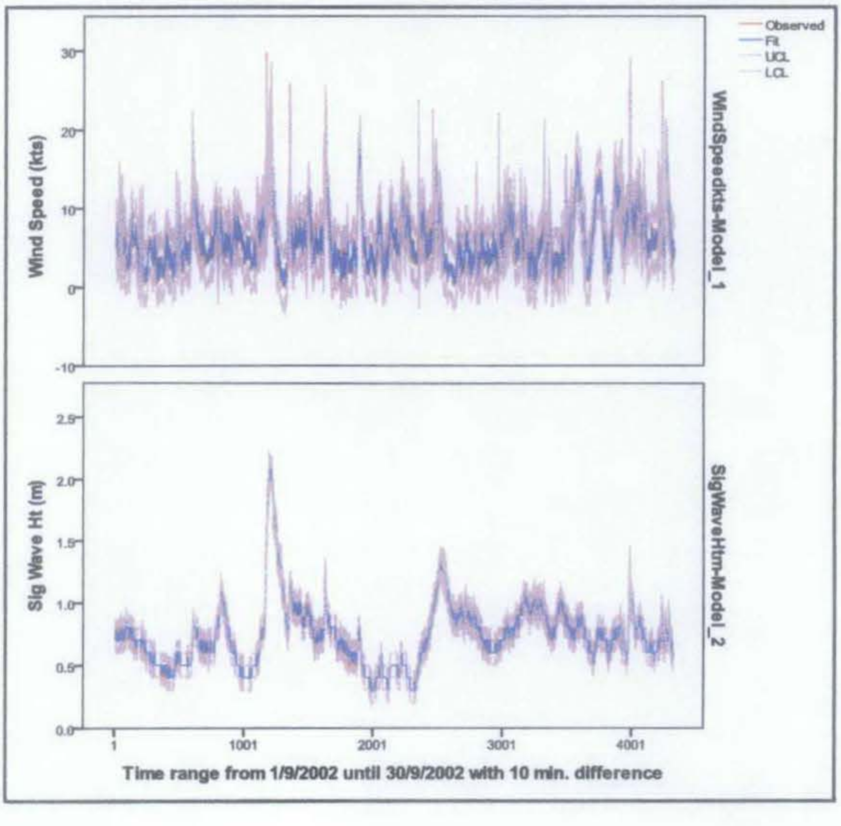

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
1
1001
2001
3001
4001
Time range from 1/3/2002 until 31/3/2002 with 10 min. differences

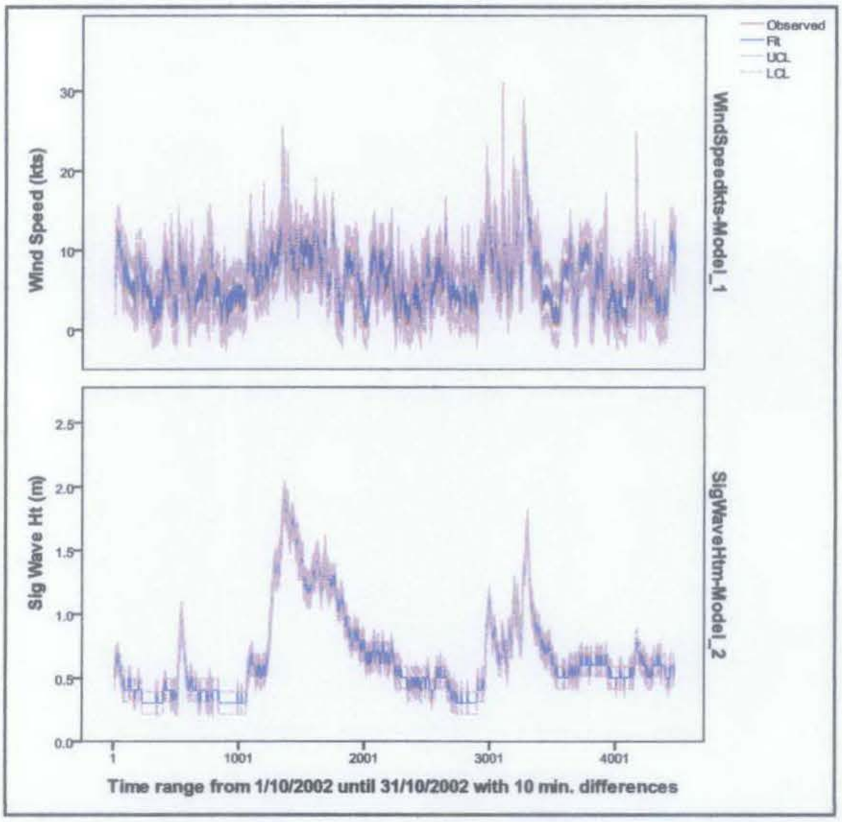

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Height (m)
SigWaveHtm-Model_2
1
1001
2001
3001
4001
Time range from1/5/2002 until 31/5/2002 with 10 min. differences

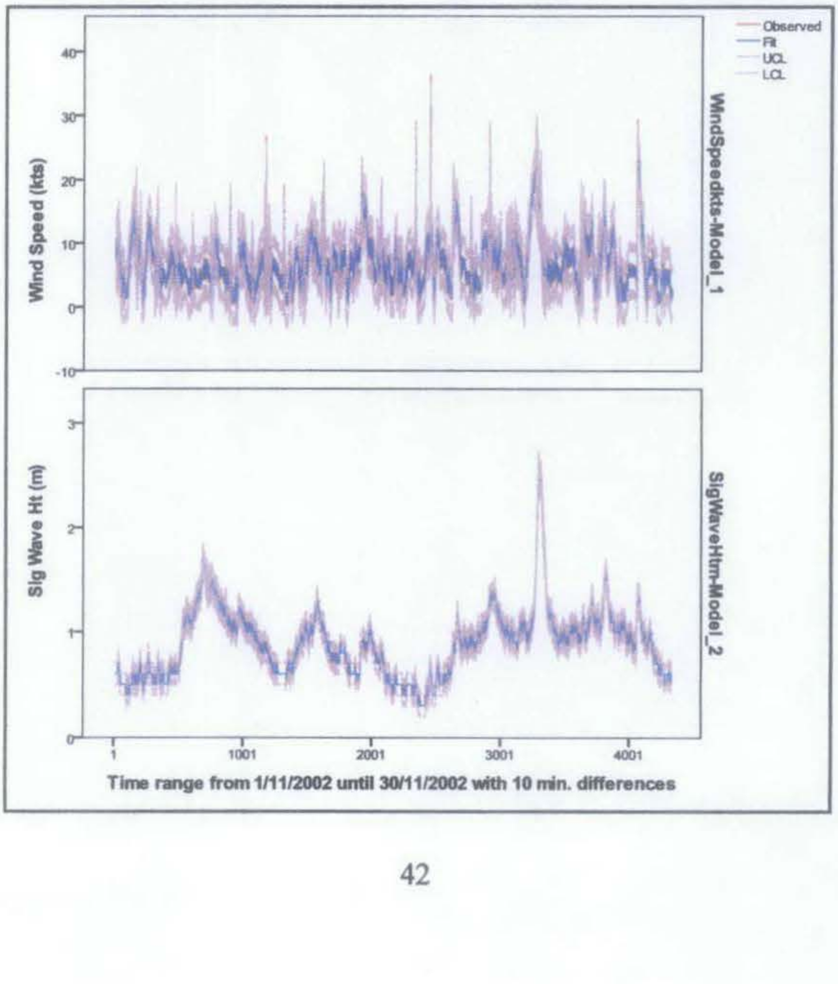

Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/6/2002 until 30/6/2002 with 10. min differences


Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/7/2002 until 31/7/2002 with 10 min. differences


Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/8/2002 until 31/8/2002 with 10 min. differences


Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
Time range from 1/9/2002 until 30/9/2002 with 10 min. difference


Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
1
1001
2001
3001
4001
Time range from 1/10/2002 until 31/10/2002 with 10 min. differences


Observed
Fit
UCL
LCL
Wind Speed (kts)
WindSpeedkts-Model_1
Sig Wave Ht (m)
SigWaveHtm-Model_2
1
1001
2001
3001
4001
Time range from 1/11/2002 until 30/11/2002 with 10 min. differences