Forecast of Operational Sea State in Malaysian Waters

by

Abrar bin Abdul Majid
14407

Dissertation submitted in partial fulfillment of
the requirements for the
Bachelor of Engineering (Hons)
(Civil)

SEPTEMBER 2014

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Forecast of Operational Sea State in Malaysian Waters

by

Abrar bin Abdul Majid
14407

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

Approved by,

____

(Associate Professor Ir. Dr. Mohd Shahir Liew)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
September 2014

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 of persons.

____

ABRAR BIN ABDUL MAJID

ACKNOWLEDGMENT

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, Associate Professor Ir. Dr. Mohd Shahir Liew for his confidence in me during the completion of this research. His encouragement, guidance and advice help me a lot in understanding the topic and its scope. All the support and guidance from him will not be forgotten. Not to forget, deepest gratitude to postgraduate student, Suzanna Azmy for helping me directly and indirectly with the raw data from PETRONAS Carigali, installation of SPSS Statistics 21 software, and the guidelines of using the software.

Last but not least, thanks to my family and everyone who has contributed directly or indirectly in the form of their moral support and inspiration throughout my undergraduate studies.

ABSTRACT

Currently, the wave height forecast for daily operation and support activities of PETRONAS are done based on estimation method. This method is found to be inaccurate. The requirement of improving the base model for wave height are critically needed to prevent the over estimation and for cost optimization. The objective of this research is mainly to develop forecast model for Malaysian waters by studying the metocean historical data. By having this forecast model, offshore operation cost optimization can be achieved. The metocean parameter that will be focused in this study is significant wave height. The study area will be Dulang (PMO), Erb West (SBO) and Tukau (SKO) platform which located in South China Sea extended from Peninsular Malaysia, Sabah and Sarawak. In order to forecast the wave height based on the historical data, Auto-regressive Integrated Moving Average (ARIMA) method will be applied. The best ARIMA model has been selected by using Bayesian Information Criterion (BIC) value. The parameters of the estimates (p,d,q) was found to be consistent for the three regions. The operation areas can be represented by one ARIMA model. In addition, it was found that the wave time series was weakly stationary and hence no differencing is needed. ARIMA model is reliable for short-term forecast and can be used for operational by Petronas Carigali Sdn. Bhd.

Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Background

Metocean is derived from the words “Meteorology” and “Oceanography”. Meteorology is the study of atmospheric and Oceanography is the ocean physical conditions such as ocean waves, currents, tides, wind and other parameters. Metocean data is the collection of both meteorological and oceanographic data that influenced the sea state which made these data is very useful for offshore operation planning.

The ability to forecast the sea state or condition such as ocean wave can produce good information for daily offshore operational planning so that cost optimization can be done. It is a challenge to develop a forecast model for ocean wave since it acts in an unpredictable manner. In addition, waves act differently during monsoon, inter-monsoon and non-monsoon seasons.

In this research, the idea of utilizing Auto-Regressive Integrated Moving Average (ARIMA) models as the base of the developing forecast model are proposed. ARIMA are one of the most important time series models used in financial market forecasting over the past three decades [1]. Guided by these successful forecasts of market using ARIMA, the achievement of wave height forecast model using ARIMA will be our ultimate goal.

To forecast the future metocean data, the historical time series metocean data are used. A number of previous wave height data are needed for forecasting process. There are several sources of metocean data which are in-situ measurements, remote sensing, and hindcast. This study will use hindcast data.

1.2 Problem Statement

In oil and gas industry, offshore operator needs to know the sea condition especially the wave in order to have a good operation planning and decision making. Sea state is never static and very dynamic which can cause a lot of accidents and loss of life.

Current practice of Petronas are using weather climate which provides the information on probability of occurrence, probability of exceedance or non-exceedance, scatter plot and weather pattern to estimate the wave height. When the waves are too big, the operation will be declared as standy mode. During this mode, the operation shall be stopped. This declaration usually announced during the monsoon season. However, this declaration sometimes is inaccurate since they are not using the forecast model. The estimation can be exaggerating while the real condition is not as worse as predicted. By declaring standby mode will cause the day wasted as well as elongate the project operating time, which affects the operating cost. On certain circumstances, the operation is still on even though on the days that the waves are predicted to be big. The selection of vessel size that goes to offshore depends on the size of the forecasted wave. The inaccurate estimation may cause bad events such as maritime accidents, pollution and disasters at sea as well as inaccurate expenditure. As of now, there is still no specific forecast model that has been developed for Malaysian Waters.

The best daily offshore operational planning and decision making can be obtained by forecasting the sea state or condition such as ocean wave. It also can help us to proceed with in situ operation and also ongoing maintenance that need to be done. An efficient data management and manipulation of the ocean wave will help to optimize the operational cost. In order to forecast the future metocean data, ARIMA model will be used.

1.3 Objectives

Basically, there are three main objectives of this study :

1. To develop wave forecast model for Malaysian waters by using ARIMA (*p,d,q*).
2. To compare and contrast the ARIMA (*p,d,q*) model of wave state in Peninsular Malaysia, Sabah and Sarawak regions.
3. To measure the accuracy of the developed forecast model.

1.4 Scope of Study

The metocean parameter that will be studied in this research is significant wave height. The location of study will be in Dulang, Peninsular Malaysia Operation (PMO), Erb West, Sabah Operation (SBO) and Tukau , Sarawak Operation (SKO). The metocean data (hindcast) provided by SEAMOS-South Fine Grid Hindcast (SEAFINE) is used to forecast future data by using ARIMA model. The forecasted data will be used in offshore operation planning to optimize the daily operational cost.

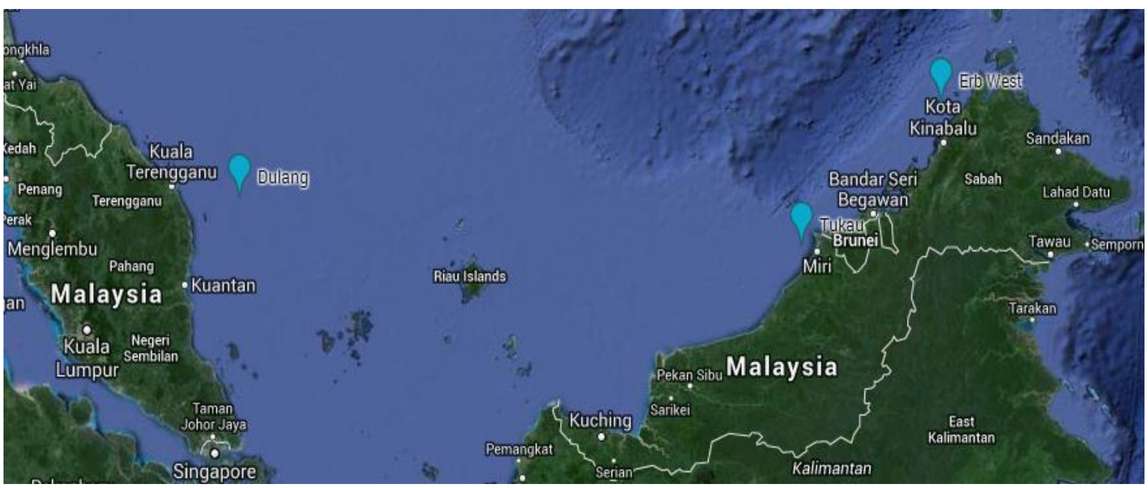


Figure 1 : *Location of platforms*

CHAPTER 2

LITERATURE REVIEW

2.1 Inter-monsoon Variation

The southern South China Sea (SSCS) is a tropical marginal sea with a complex geographical setting in Southeast Asia [2]. It is situated between the western Pacific and the eastern Indian Ocean. The SSCS connects with the Sulu Sea in the northeast through the Mindoro and Balabac Straits, with the Java Sea in the south through the Karimata Straits, and with the Andaman Sea in the west through the Malacca Strait. It is basically a shallow continental basin with the average depth of 60m.

The climate over SSCS is controlled by South China Sea (SCS) monsoon. Weaker southwesterly summer monsoon winds occur from April to August which drives a northward coastal jet off Vietnam. Stronger northeasterly winter monsoon winds usually occur from November to March which causes a southward coastal jet in the SCS. Current moves southward along Peninsular Malaysia during northeast monsoon and moves in the opposite way during southwest monsoon. Thus, the pattern of sea surface circulation is different according to monsoon seasons. During southwest monsoon, the main surface current moves from the Indonesian Seas (Karimata and Malacca Straits) to the north (South China Sea). On the opposite, during the northeast monsoon, the sea surface circulation comes from the north to the south through the small islands.

2.2 Time Series

Chronological sequence of observation on a particular variable is called time series [3]. There are several types of time series which are trend, cycle, seasonal variations and unpredictable variations.

Trend can be classified as ups and downs of the time series over a period of time. Besides that, trend reflects the increment or decrement in the time series. For cycle, it 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.

Furthermore, seasonal variations can be classified as periodic patterns in time series that complete themselves within a year and then repeat annually. This component usually affected by weather. Lastly, the unpredictable fluctuation is inconsistent movements in a time series that have no regular pattern. This type of time series cannot be forecasted since it is caused by unusual events like earthquakes, hurricanes and other natural phenomena.

2.2.1 Times Series Analysis

Historical data form a time series. Example of a time series is a set of observations on a variable measured at successive points in time or over successive periods of time. There are basically two main goals of time series analysis:

1. Identifying the nature of the phenomenon, represented by the sequence of observation
2. Forecasting (predicting future values of the time series variable)

For this study, only time domain analysis that will be involved. The first step in time domain analysis is generating a time series plot which is in form of observation against time graph. Next step is obtaining the basic descriptive properties of the time series which are mean (2.1) and variance (2.2).

$$\mu = \frac{1}{N}\sum_{t=1}^{N} x_t \tag{2.1}$$

$$\sigma^2 = \frac{1}{N-1}\sum_{t=1}^{N-k} (x_t - \mu)^2 \tag{2.2}$$

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

$$C_k = \frac{1}{N-1}\sum_{t=1}^{N-k} (x_t - \mu)(x_{t+k}) \tag{2.3}$$

Autocovariance function also can be obtained by using equation 2.3. Autocovariance is the variance of the variable against a time-shifted version of itself. In similar, autocorrelation also can be calculated by using equation 2.3. Basically, autocorrelation is a correlation between values of the process at different points in time, as a function of two times or the time difference. The following formula shows the relationship between autocovariance and autocorrelation function.

$$r_k = \frac{C_k}{\sigma^2} \tag{2.4}$$

2.3 Time Series Forecasting

Forecasting methods can be classified as quantitative or qualitative. Quantitative forecasting method can be used when:

1. Past information about the variable being forecast is available
2. The information can be quantified
3. It can be assumed that the pattern of the past will continue into the future

For quantitative forecasting method, a forecast model can be developed by using a time series method or causal method. Time series method is the historical data are restricted to past values of the variable. The objective is to discover a pattern in the historical data and then extrapolate the pattern in the future. One of the examples of this method is ARIMA model.

For qualitative method, generally it involves the use of expert judgment to develop forecasts. In this study, ARIMA model will be used to forecast future data.

2.3.1 ARIMA (p,d,q) Models

ARIMA is a mathematical models used for time series forecasting. ARIMA stands for Autoregressive, Integrated and Moving Average. Each of these phrases describes a different portion of the mathematical model

According to [4], ARIMA model of a time series is defined by three terms (p,d,q). Identification of a time series is the process of finding integer, usually very small (e.g., 0, 1, or 2) values of p, d, and q that model the patterns in the data. When the value is 0, the element is not needed in the model. The middle element, d is identified before p and q. The goal is to determine whether the process is stationary. Stationary process has a constant mean and variance over the time period of the study.

2.4 Application of ARIMA

In the journal 'Selection of Best ARIMA Model for Forecasting Average Daily Share Price Index of Pharmaceutical Companies in Bangladesh,' [5] examine empirically the best ARIMA model for forecasting. The data used to meet the purpose is average daily share price indices of the data series of Square Pharmaceuticals Limited (SPL). Firstly, stationary condition of the data series are observed by ACF and PACF plots. To check for stationarity, statistics method such as Ljung-Box-Pierce Q-statistics and Dickey-Fuller are used. They found that the average daily share price indices of SPL data series are non-stationary even after log-transformation. The best ARIMA model has been selected by using AIC, SIC, AME, RMSE and MAPE criteria. The best model to forecast the SPL data series is ARIMA (2,1,2).

Besides that, ARIMA method also been used in stock market by Acturial Consultants. The other application of ARIMA is in chemical processing and control. In conclusion, Science and Arts used this method.

CHAPTER 3

METHODOLOGY

3.1 Box-Jenkins Method

Box-Jenkins methodology will be used in this research since it involving time series forecasting data based on the historical data. There are 5 stages in this methodology. Before that, the historical data received will be sorted first in order to ensure the analysis can be done smoothly. From the sorted data, a wave height time series will be generated with respect to a particular time range. After plotting the time series, 5 stages of Box-Jenkins Methodology can be applied.

3.1.1 Data Preparation

Data preparation involves transformation and differencing. Transformations of the data such as square roots or logarithms can help to stabilize the variance in a series where the variation changes with the level. Then, the data will be differenced until there are no obvious patterns such as trend or seasonality left in the data. Differencing of a time series $x(t)$in discrete time t is the transformation of the series $x(t)$to a new time series $d(t)$where the values $d(t)$are the differences between consecutive values of $x(t)$. This procedure may be applied consecutively more than once, giving rise to the "first differences", "second differences" and etc.

The first differences $d^{(1)}(t)$of a time series $x(t)$ are described by the following expression:

$$d^{(1)}(t) = x(t) - x(t-1) \tag{3.1}$$

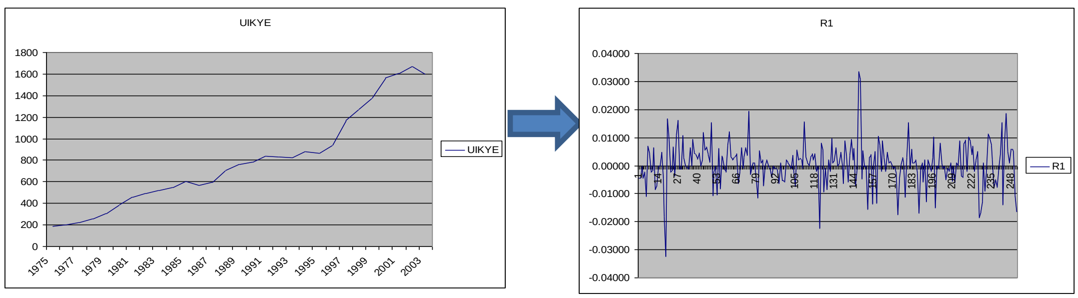
The second differences $d^{(2)}(t)$ may be computed from the first differences $d^{(1)}(t)$ according to the expression:

$$d^{(2)}(t) = d^{(1)}(t) - d^{(1)}(t-1) \tag{3.2}$$

The general expression for the differences of order m is given by the recursive formula

$$d^{(m)}(t) = d^{(m-1)}(t) - d^{(m-1)}(t-1) \tag{3.3}$$

where the top index means the order of the difference. The purpose of doing differencing process is to stationarize the non-stationary data.



Non- Stationary series　　　　Stationary series

Figure 2 : *Conversion of non-stationary series to stationary series*

3.1.1.1 Checking the Series Stationarity Using Correlogram

The stationarity of the series can be checked by using Correlogram. Correlogram is known as an autocorrelation plot, is a plot of the sample autocorrelations, ACF versus time lag, k.

The ACF lies between -1 and +1, the correlogram also lies between these values. If the ACF falls immediately from 1 to 0, then equals about 0 thereafter, the series is stationary. If the ACF declines gradually from 1 to 0 over a prolonged period of time, then it is not stationary.

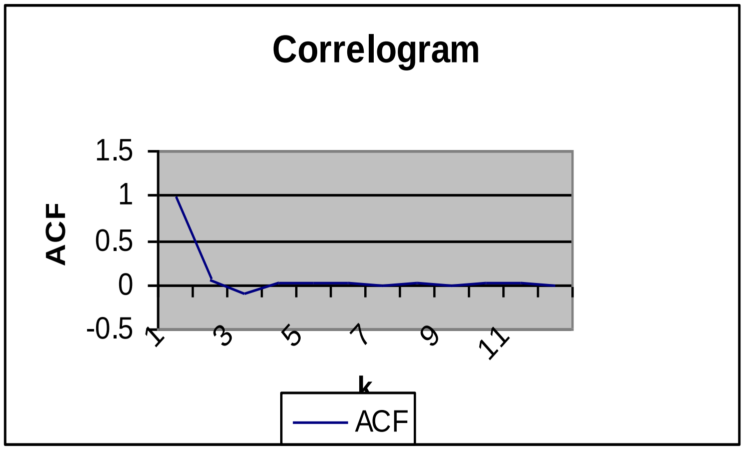


Figure 3 : *Correlogram of Stationary Series*

3.1.2 Model Selection

Model selection in the Box-Jenkins framework uses various graphs based on the transformed or differenced data in order to identify potential ARIMA processes which might provide a good fit to the data. The selection of model is based on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plot. There are several types of model which are model for non-stationary series, Autoregressive (AR), Moving Average (MA) and also Autoregressive Integrated Moving Average (ARIMA). The best model will be selected based on Bayesian Information Criterion (BIC) value. Smaller value of BIC indicates better fit of model.

3.1.2.1 Autoregressive Integrated Moving Average, ARIMA (*p,d,q*)

Models for non-stationary series are called Autoregressive Integrated Moving Average models, or ARIMA (*p*,*d*,*q*), where *d* indicates the amount of differencing.

3.1.2.2 Auto-regressive, AR (*p*)

It is a series of current values which depend on its own previous values. The value of *p* is the number of auto-regressive components in an ARIMA (*p*,*d*,*q*) model.The value of *p* is 0 if there is no relationship between adjacent observations. When the value of *p* is 1, there is a relationship between observation at lag 1 and the correlation coefficient, $\emptyset_1$ is the magnitude of the relationship. When the value of *p* is 2, there is a relationship between observation at lag 2 and the correlation coefficient, $\emptyset_2$ is the magnitude of the relationship. Thus p is the number of correlations that needed to model the relationship. Equation below is the example of a model with p = 2 , ARIMA (2,0,0).

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \varepsilon_t \tag{3.4}$$

where $Y_t =$ Response variable at time *t*

$\emptyset_i =$ Regression coefficient to be estimated

$\varepsilon_t =$ Error term at time *t*

3.1.2.3 Moving Average, MA (q)

Moving average is a regression model with the dependent variable. Y_t depending on previous values of the error rather the variable itself. The value of q is the number of moving average components in an ARIMA *(p, d, q)*. When q is zero, there is no moving average component. When q is 1, there is a relationship between the current data and the random shock at lag 1 and the correlation coefficient ω_1 represents the magnitude of the relationship. When q is 2, there is a relationship between the current data and the random shock at lag 2 and the correlation coefficient ω_2 represents the magnitude of the relationship. For example, an ARIMA (0,0,2) model is

$$Y_t = \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} \tag{3.5}$$

where $Y_t =$ Response variable at time

$\omega_i =$ Regression coefficient to be estimated

$\varepsilon_t =$ Error term at time *t*

3.1.2.4 Mixed Model, ARMA (*p,q*)

Usually a series has both auto-regressive and moving average components, thus both types of correlations are required to model the patterns. If both elements are present only at lag 1, the equation is expressed as :

$$Y_t = \emptyset_1 Y_{t-1} - \omega_1 \varepsilon_{t-1} + \varepsilon_t \tag{3.6}$$

The above equation is the combination of both AR and MA components.

3.1.2.5 ACF and PACF

Models are identified through patterns in their autocorrelation functions (ACF) and partial autocorrelation functions (PACF). Both autocorrelations and partial autocorrelations are computed for sequential lags in the series. The first lag has an autocorrelation between Y_{t-1} and Y_t , the second lag has both an autocorrelation and partial autocorrelation between Y_{t-2} and Y_t , and so on. ACFs and PACFS are the functions across the lags. The general equation for autocorrelation is expressed as :

$$r_k = \frac{\frac{1}{N-K}\sum_{t=1}^{N-k}(Y_t - \bar{Y})\,(Y_{t-k} - \bar{Y})}{\frac{1}{N-1}\sum_{t=1}^{N}(Y_t - \bar{Y})^2} \tag{3.7}$$

where N is the number of observation in the whole series, k is the lag. $\bar{Y}$ is the mean of the whole series and the denominator is the variance of the whole series.

The equations for computing partial autocorrelation are much more complex, and involve a recursive technique. These are the expressions showing the relationship between ACF and PACF for the first three lag.

$$PACF(1) = ACF(1) \tag{3.8}$$

$$PACF(2) = \frac{ACF(2) - \left(ACF(1)\right)^2}{(1 - [ACF(1)]^2} \tag{3.9}$$

$$PACF(3) = \frac{-2(ACF(1))ACF(2) - [ACF(1)]^2 ACF(3)}{1 + 2[ACF(1)]^2 ACF(2) - [ACF(2)]^2 - 2[ACF(1)]^2} \quad (3.10)$$

3.1.2.6 Model Selection

Based on the graph of ACF and PACF, the model of AR, MA or ARMA will be selected. Below is the table of summary for ACF and PACF properties.

Table 1 : *ACF and PACF properties*

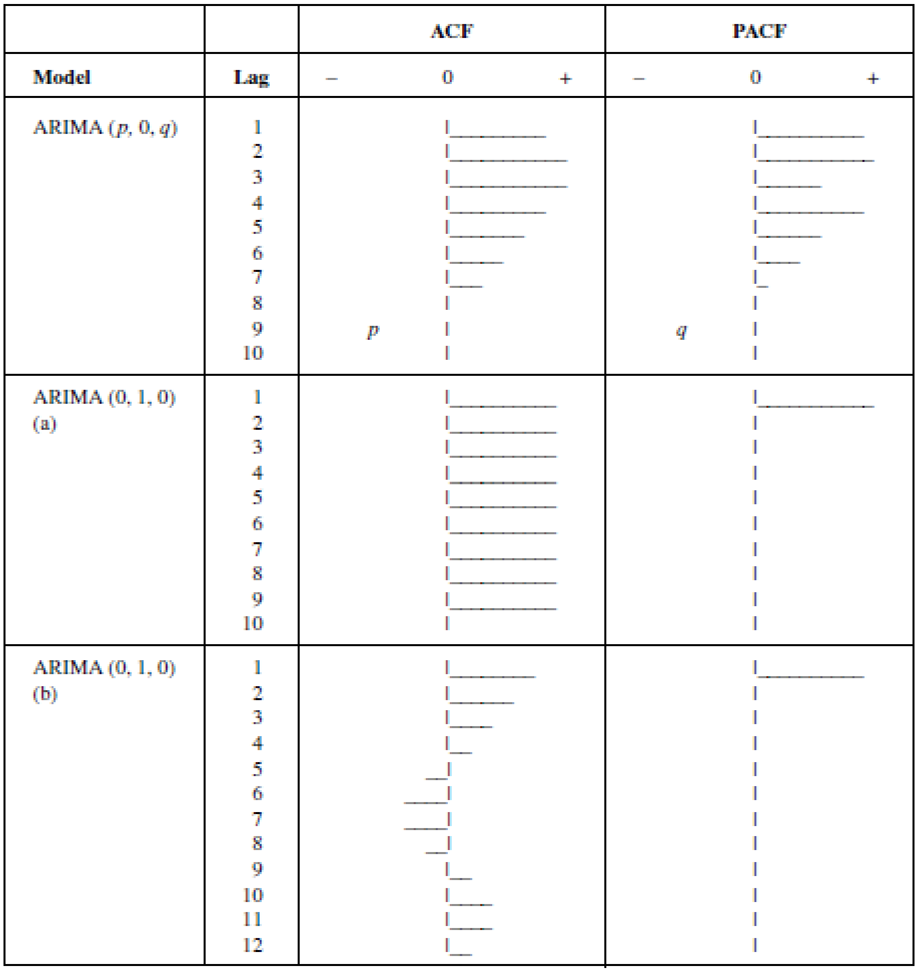
	ACF	PACF
AR(p)	Die out	Cut off after the order p of the process
MA(q)	Cut off after the order q of the process	Die out
ARMA(p,q)	Die out	Die out

In Table 1 context, ‘die out’ means tend to be zero gradually while ‘cut off’ means disappear or is zero. Besides that, ACF and PACF plot for common ARIMA models are summarized in Table 2 which is adopted from [4].

Table 2 : *Common ACF and PACF plots for ARIMA model*

Model	Lag	ACF − 0 +	PACF − 0 +
ARIMA (1, 0, 0)	1	I___________	I_________
	2	I_________	I
	3	I______	I
	4	I____	I
	5	I_	I
	6	I	I
	7	I	I
	8	I	I
	9	I	I
	10	I	I
ARIMA (0, 0, 1)	1	I_________	I___________
	2	I	I___________
	3	I	I________
	4	I	I____
	5	I	I_
	6	I	I
	7	I	I
	8	I	I
	9	I	I
	10	I	I
ARIMA (2, 0, 0)	1	I___________	I__________
	2	I____________	I___________
	3	I________	I
	4	I_____	I
	5	I_	I
	6	I	I
	7	I	I
	8	I	I
	9	I	I
	10	I	I
ARIMA (0, 0, 2)	1	I___________	I____________
	2	I____________	I____________
	3	I________	I________
	4	I_____	I______
	5	I_	I____
	6	I	I_
	7	I	I
	8	I	I
	9	I	I
	10	I	I

Table 3 : *Common ACF and PACF plots for ARIMA model (cont.)*



3.1.3 Parameter Estimation

In parameter estimation stage, the values of the model coefficients which provide the best fit to the data will be determined. To perform this estimation, sophisticated computational algorithms such as Yule Walker, Method of Moments and Partial Inverse Matrices will be used.

The parameter of each mathematical model depends on the set of time series data which will show the characteristics of either autoregressive (AR), moving average (MA) or the combination of both. The determination of the model adapted can be obviously shown by the ACF PACF graph and value. However, there are some hierarchies in model selection.

The model selection will reflect on the number of coefficient that will be used in each model. AR and MA model will only have one type of coefficient plus one error coefficient, while ARIMA model will have 2 types of coefficient plus one error coefficient.

On the other hand, model degree will determined the repetition of each coefficient in one model. For example, AR 2 model will have 2 values of Ø coefficient.

3.1.4 Model Checking

Before we start using the model to forecast the data, the model needs to be checked first. It involves testing the assumptions of the model in order to identify any inadequacy. If the model is found to be inadequate, it is a must to go back to model selection stage and try to identify a better model. The model is adequate if the differences between actual and predicted values are too large.

3.1.5 Forecasting

The model will be used for forecasting once the the best ARIMA (p,d,q) model has been selected. It is a straight forward task computes by using IBM SPSS Statistics 21 software. Forecasting is the ultimate purpose of the whole procedure is designed for.

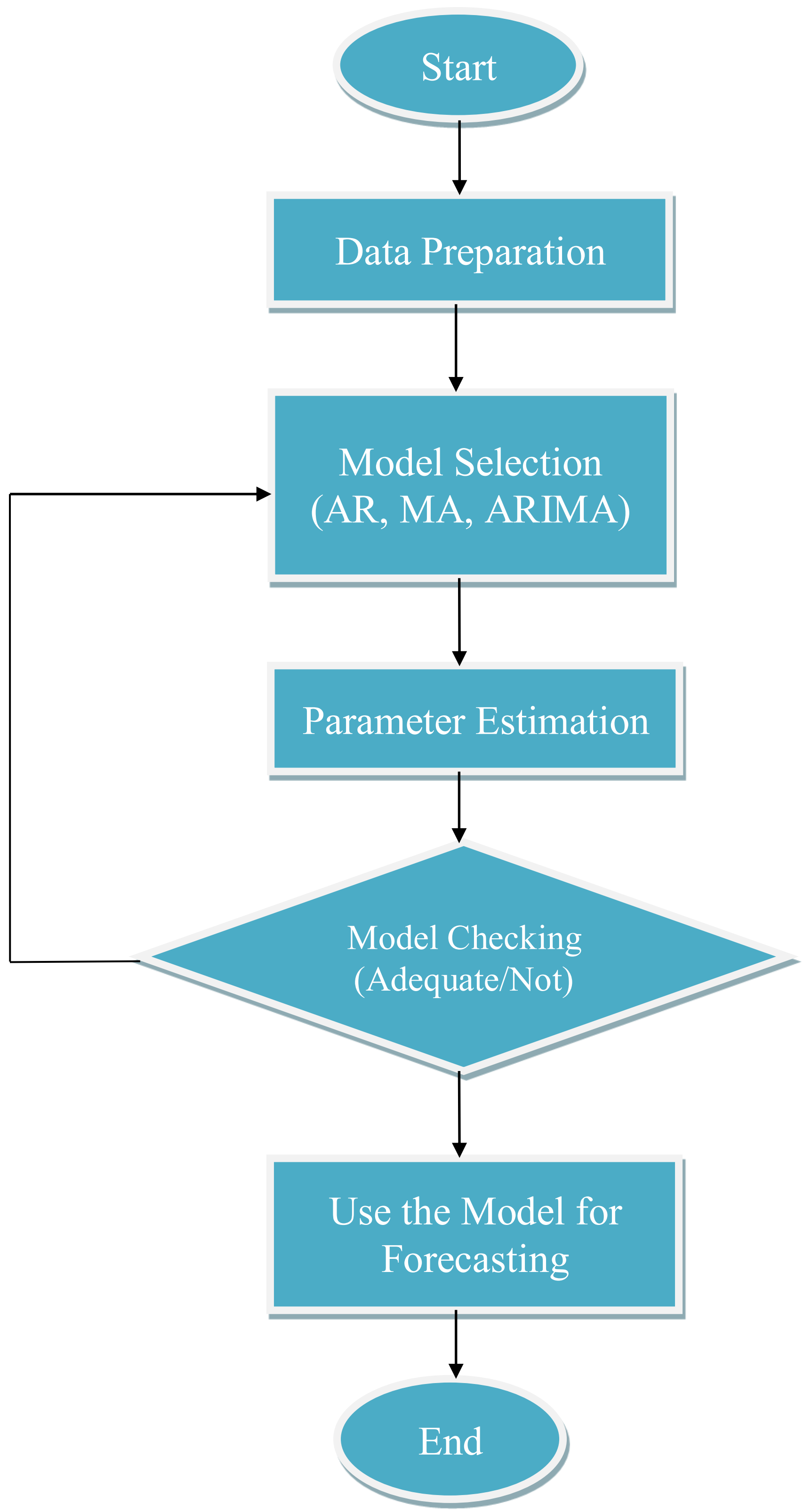


Figure 4 : *Flowchart of Box-Jenkins Method*

3.2 Program and Software

3.2.1 Microsoft Office Excel

Microsoft Office Excel is one of the tools that help to sort, modify data and plot graph. It also can be used to solve calculations and tabulate data.

3.2.2 IBM SPSS Statistics 21

Comprehensive and flexible statistical analysis and data management can be done by using SPSS Statistics 21. SPSS can be used to take any data from almost any type of file and can be used to produce charts, plots of distributions and trends, descriptive statistics, conduct complex statistical analyses and reports.

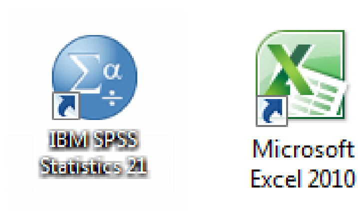


Figure 5 : *Software used in this study*

3.3 Key Project Milestones

The planned schedules for Final Year Project II are as follows.

Submission of Progress Report	Week 7
Pre-SEDX	Week 10
Draft Report/Dissertation/Technical Paper	Week 11/12/13
Viva	Week 14
Project Dissertation (Hard Bound)	Week 15

Key mile stone

Normal event

NO.	DETAILS / WEEK	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Analysis of Dulang, PMO wave height data															
2	Submission of Progress Report															
3	Analysis of Erb West, SBO wave height data															
4	Analysis of Tukau, SKO wave height data															
5	Pre-SEDEX															
6	Submission of Draft Report															
7	Submission of Dissertation (Soft Bound)															
8	Submission of Technical Paper															
9	Viva															
10	Submission of Project Dissertation (Hard Bound)															

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Data Analysed and Validation

In this study, there are 3 different regions in Malaysia represented by 3 different platforms location are used. The 3 platforms are Dulang, Peninsular Malaysia region, Tukau, Sarawak region and Erb West, Sabah region. These 3 platforms have their own function and all measurements are done in these platforms in order to get the significant wave height (metocean data) that shows every region's metocean condition.

The following table shows the details of the 3 platforms.

Table 4 : *Details of Platform*

DETAILS	DULANG (PMO)	TUKAU (SKO)	ERB WEST (SBO)
Platform type	Fixed Steel Jacket	Fixed Steel Jacket	Fixed Steel Jacket
Latitude	5°48'0'' N	4°24'0'' N	6°21'0'' N
Longitude	104°9'0'' E	113°45'0'' E	115°39'0'' E
Water Depth	62.508m	40.000m	24.000m

4.2. Time Series of Wave Height

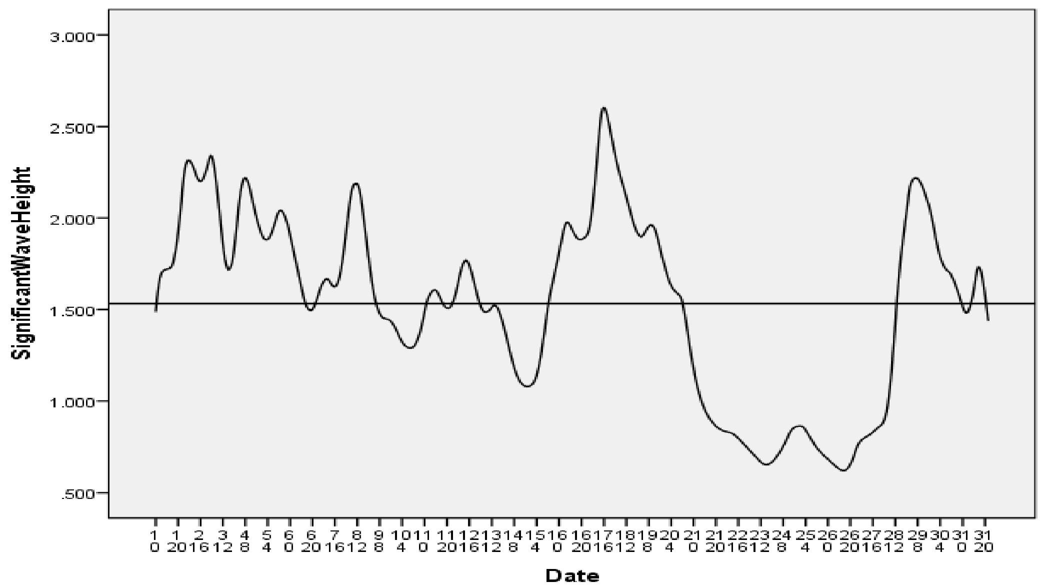


Figure 6 : *Time Series of Significant Wave Height*

Figure 6 is the example of time series for wave height in January 2003 at Dulang platform modeled by using SPSS software. The straight line in the graph represents the mean of the time series.

4.3 ARIMA (*p, d, q)* model

By looking at the BIC value, the most suitable ARIMA model for January 2003 wave height time series is ARIMA (2,0,1). ARIMA (2,0,1) has been selected to forecast for January 2003 wave height times series because it has the lowest value of BIC compared to other models produced.

Table 5 : *Model Description*

Model Description

			Model Type
Model ID	SignificantWaveHeightCorrected	Model_1	ARIMA(2,0,1)(0,0,0)

Table 6 : *Model Summary*

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	1.000	.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
R-squared	1.000	.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
RMSE	.004	.	.004	.004	.004	.004	.004	.004	.004	.004	.004
MAPE	.169	.	.169	.169	.169	.169	.169	.169	.169	.169	.169
MaxAPE	3.942	.	3.942	3.942	3.942	3.942	3.942	3.942	3.942	3.942	3.942
MAE	.003	.	.003	.003	.003	.003	.003	.003	.003	.003	.003
MaxAE	.061	.	.061	.061	.061	.061	.061	.061	.061	.061	.061
Normalized BIC	-10.974	.	-10.974	-10.974	-10.974	-10.974	-10.974	-10.974	-10.974	-10.974	-10.974

4.3.1 Comparison of BIC values

The best model will be selected based on BIC value. The smaller value of BIC indicates better fit of model. For each month data, there are 6 ARIMA models used for forecasting. The models and the BIC value are summarised in Table 7 for each location.

Table 7 : *Summary of BIC values for Dulang*

	Lowest value of BIC

Month	BIC (2,0,1)	(2,0,2)	(2,0,3)	(2,1,1)	(2,1,2)	(2,1,3)
January	-10.974	-11.098	-11.165	-11.332	-11.322	-11.314
February	-9.542	-9.572	-9.573	-12.066	-12.055	-12.045
March	-11.047	-11.224	-11.333	-12.145	-12.135	-12.126
April	-12.067	-12.073	-12.036	-13.021	-13.024	-13.015
May	-10.474	-10.688	-10.743	-11.279	-11.269	-11.267
June	-7.249	-7.245	-7.252	-12.227	-12.239	-12.23
July	-8.583	-8.599	-8.598	-12.029	-12.019	-12.011
August	-10.246	-10.26	-10.278	-12.28	-12.299	-12.289
September	-11.457	-11.491	-11.521	-12.291	-12.28	-12.273
October	-9.443	-9.45	-9.442	-12.275	-12.293	-12.283
November	-11.379	-11.59	-11.744	-12.449	-12.443	-12.437
December	-9.565	-9.581	-9.576	-11.79	-11.33	-11.772

Table 8 : *Summary of BIC values for Tukau*

	Lowest value of BIC

	BIC					
Month	(2,0,1)	(2,0,2)	(2,0,3)	(2,1,1)	(2,1,2)	(2,1,3)
January	-9.658	-9.687	-9.704	-11.841	-11.831	-11.821
February	-9.574	-9.577	-9.605	-11.686	-11.675	-11.663
March	-11.037	-11.187	-11.336	-11.888	-11.889	-11.883
April	-10.303	-10.329	-10.336	-12.135	-12.139	-12.129
May	-8.943	-8.963	-8.96	-10.862	-10.879	-10.874
June	-10.13	-10.135	-10.127	-12.111	-12.106	-12.097
July	-8.391	-8.402	-8.41	-12.509	-12.553	-12.543
August	-9.353	-9.364	-9.369	-12.632	-12.642	-12.62
September	-11.197	-11.363	-11.448	-12.294	-12.335	-12.359
October	-8.449	-8.446	-8.442	-12.63	-12.672	-12.661
November	-11.447	-11.46	-11.463	-12.529	-12.53	-12.52
December	-10.763	-10.806	-10.793	-11.583	-11.579	-11.569

Table 9 : *Summary of BIC values for Erb West*

	BIC					
Month	(2,0,1)	(2,0,2)	(2,0,3)	(2,1,1)	(2,1,2)	(2,1,3)
January	-10.906	-10.912	-10.914	-10.958	-10.972	-10.965
February	-10.042	-10.04	-10.061	-11.13	-11.13	-11.128
March	-10.957	-11.063	-11.135	-11.72	-11.711	-11.701
April	-9.467	-9.493	-9.502	-11.941	-11.94	-11.936
May	-8.728	-8.755	-8.755	-11.126	-11.151	-11.142
June	-9.524	-9.541	-9.543	-11.683	-11.683	-11.676
July	-8.402	-8.424	-8.379	-10.352	-10.35	-10.345
August	-10.484	-10.522	-10.523	-11.651	-11.643	-11.633
September	-11.651	-11.743	-11.78	-11.827	-11.824	-11.816
October	-9.881	-9.901	-9.886	-10.546	-10.552	-10.542
November	-9.893	-9.791	-9.804	-10.822	-10.819	-10.812
December	-8.613	-8.615	-8.61	-10.916	-10.911	-10.903

4.4 Forecast Result

By using ARIMA model, the future wave height data can be forecasted according to the desired period of time which is 7 days. The forecast result is set to fall between 95% confidence interval.

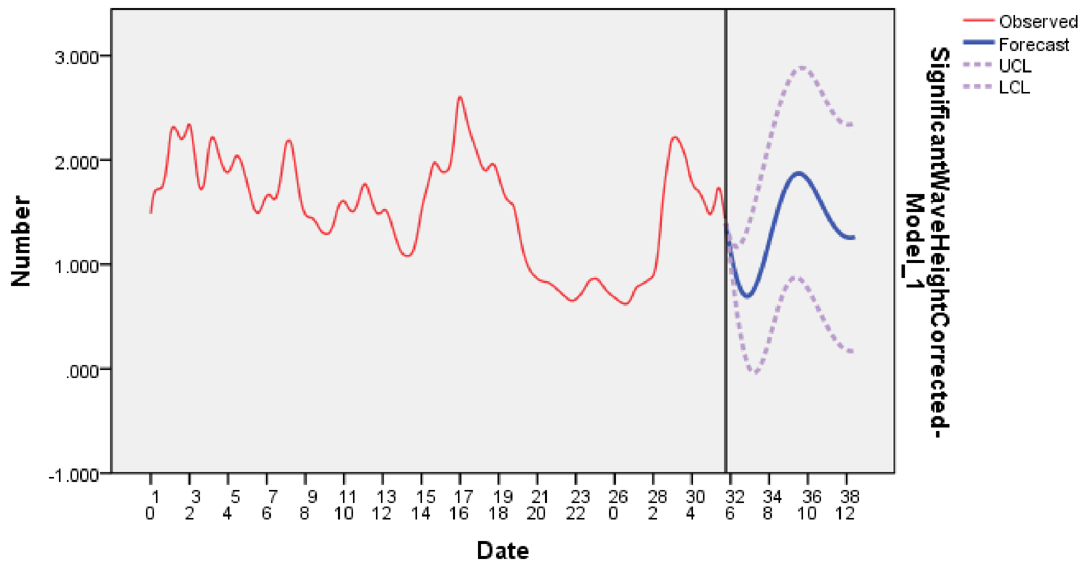


Figure 7 : *7 days forecasting using ARIMA (2,0,1) model for Dulang Platform data (Jan 2003)*

4.5 Comparison between Forecasted Wave Height and Actual Wave Height

The accuracy of the model is measured by making a comparison between the forecasted wave height and the actual wave height.

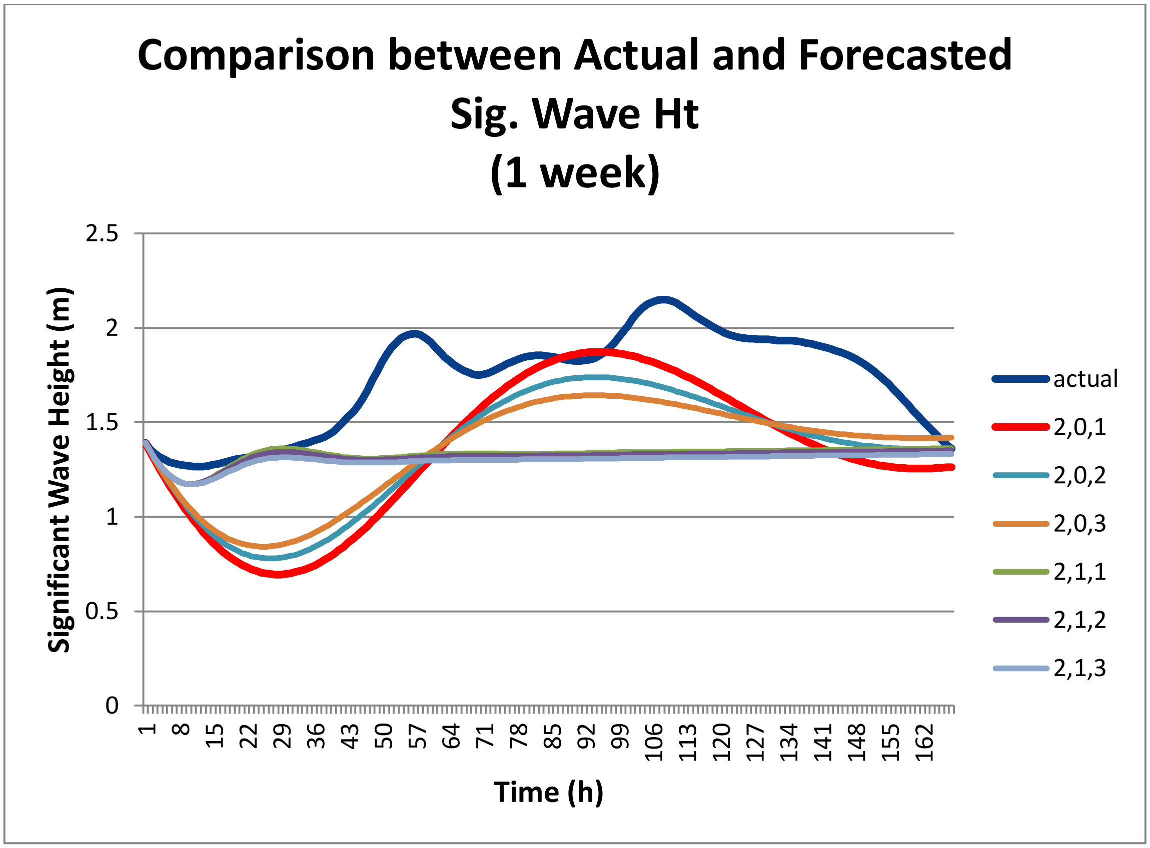


Figure 8 : *Using different ARIMA models for Dulang January 2003 time series (hourly data) to forecast from 1/2/2003 -7/2/2003 (1 week)*

There are 6 forecasted wave height data from different ARIMA (*p*,*d*,*q*) model. Based on the BIC value, for January 2003 Dulang Platform data, the best fit model is ARIMA (2,0,1). Based on the plotted graphs, the red colour graph which is ARIMA (2,0,1) model has the smallest difference with the actual wave height data. The accuracy of the model for January 2003 dropped with time. However, to perform long-term forecasting, the data need to be updated regularly as the statistical properties of the upcoming wave height will vary along with time.

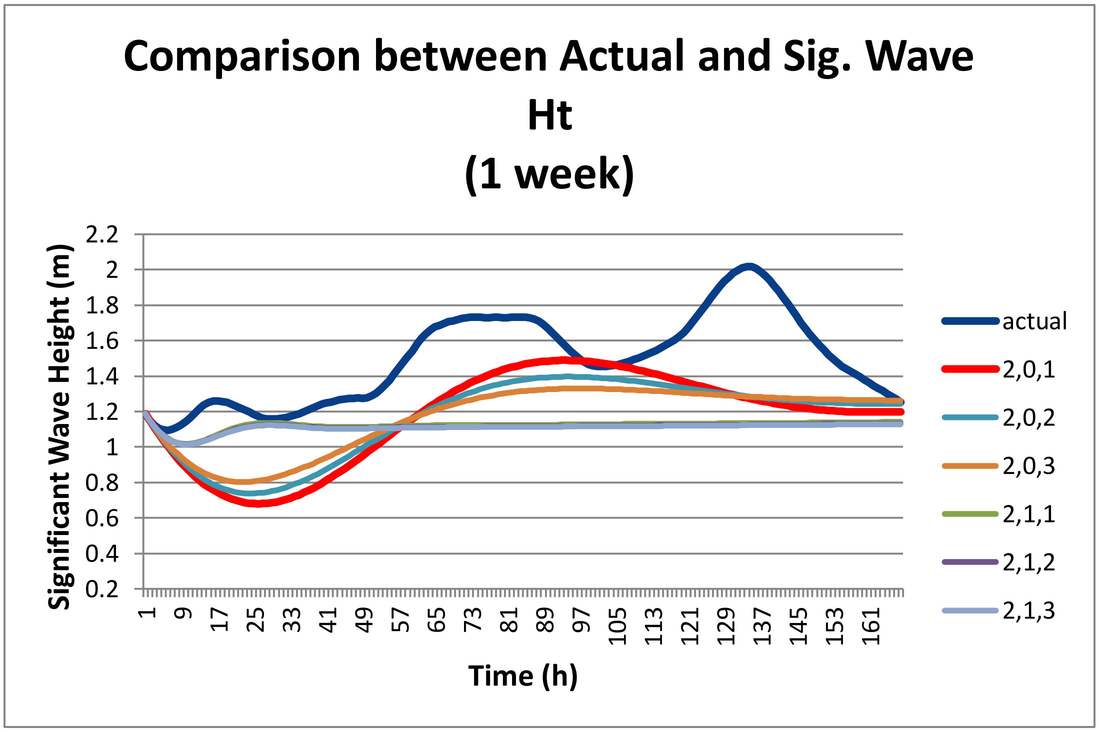


Figure 9 : *Using different ARIMA models for Tukau January 2003 time series (hourly data) to forecast from 1/2/2003 -7/2/2003 (1 week)*

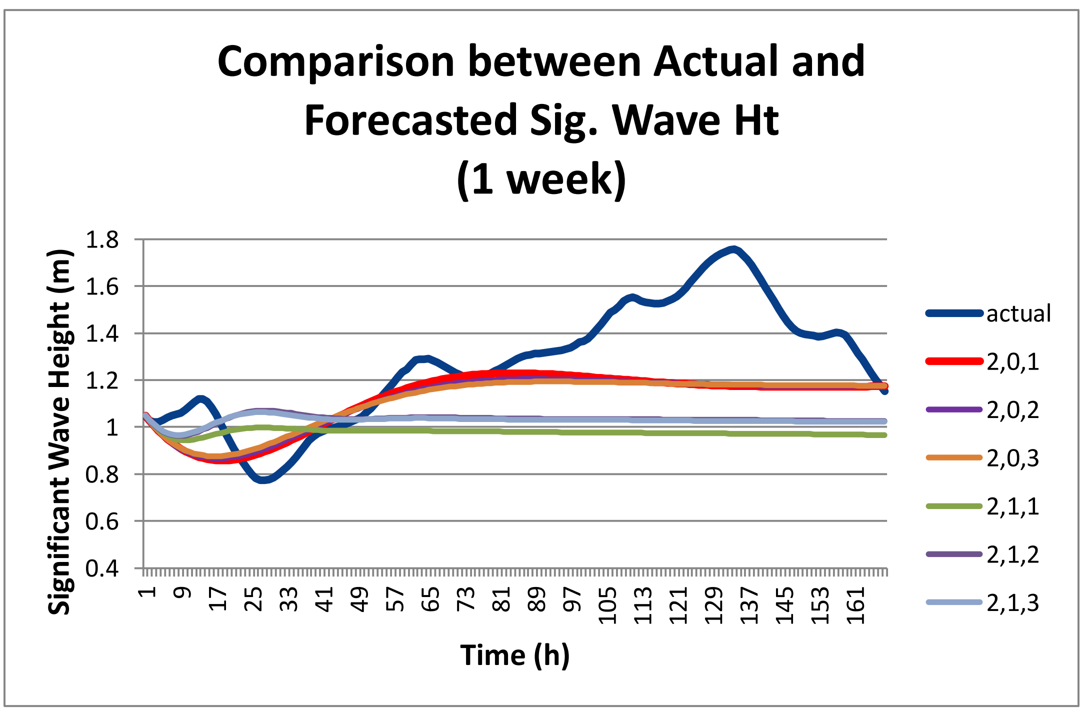


Figure 10 : *Using different ARIMA models for Erb West January 2003 time series (hourly data) to forecast from 1/2/2003 -7/2/2003 (1 week)*

4.6 ARIMA (*p*,*d*,*q*) parameters

ARIMA (*p*,*d*,*q*) parameters that have been used for forecasting for three different regions are summarised according to month in 2003. Table 10 shows the summary of the parameters.

Table 10 : *Summary of ARIMA (p,d,q) paramaters for Dulang, Tukau and Erb West*

Month	Dulang			Tukau			Erb West		
	ARIMA Parameters								
	AR(*p*)	I(*d*)	MA(*q*)	AR(*p*)	I(*d*)	MA(*q*)	AR(*p*)	I(*d*)	MA(*q*)
January	2	0	1	2	0	1	2	0	1
February	2	0	1	2	0	1	2	0	2
March	2	0	1	2	0	1	2	0	1
April	2	0	3	2	0	1	2	0	1
May	2	0	1	2	0	1	2	0	1
June	2	0	2	2	0	3	2	0	1
July	2	0	1	2	0	1	2	0	3
August	2	0	1	2	0	1	2	0	1
September	2	0	1	2	0	1	2	0	1
October	2	0	3	2	0	3	2	0	1
November	2	0	1	2	0	1	2	0	2
December	2	0	1	2	0	1	2	0	3

The order of Autoregressive is constant for all regions which the value is 2. Besides that, no differencing is needed for all regions since the wave time series is weakly stationary. Moving Average order ranges from 1 to 3 for all regions and months. Generally, the parameters of the estimates (*p*,*d* *q*) was found to be consistent for the three locations.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

From the overall analysis of the project, there are several main points that can be concluded. ARIMA (p,d,q) parameters model varies insignificantly for different month during monsoon and non-monsoon.

Besides that, the parameters of the estimates (p,d,q) was found to be consistent for the three regions. The operation areas can be represented by one ARIMA model. In addition, it was found that the waves time series was weakly stationary and hence no differencing is needed. This ARIMA model supported the theory that South China Sea is not a fully developed sea.

ARIMA model is reliable for short-term forecast and can be used for operational by Petronas Carigali Sdn. Bhd. ARIMA (p,d,q) model also can be used for long-term forecasting future environmental load such as wave by regularly updating the data since the accuracy of the model dropped with the increasing time.

There are some recommendations can be taken into account in order to improve the findings of this research. More engineering inputs such as wind data, bathymetry data and others are needed to increase the accuracy of the model. The ARIMA model generated is solely depends on the statistical data.

Finally, more data from other locations are needed in order to have more precise analysis for each region.

REFERENCES

[1] L.Abdullah, "ARIMA Model for Gold Bullion Coin Selling Prices Forecasting," *International Journal of Advances in Applied Sciences,* vol. 1, (4), pp. 153-158, 2012.

[2] M. F. Mohd Akhir, N. Z. Zakaria, and F.Tangang, "Intermonsoon Variation of Physical Characteristics and Current Circulation along the East Coast of Peninsular Malaysia," *International Journal of Oceanography, 2014*, pp. 1-9. doi: 10.1155/2014/527587, 2014.

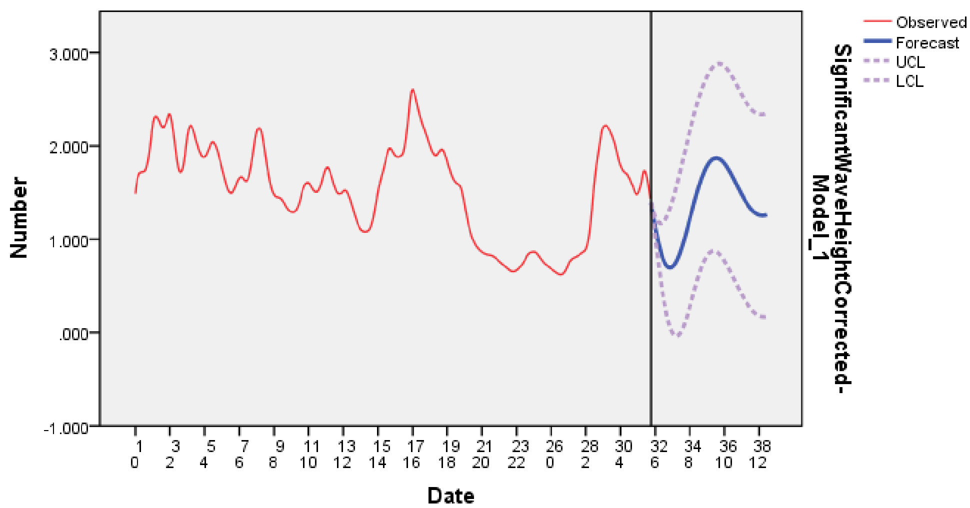
[3] F. Mohd Shafri, "*Time Series Modelling of Measured Wind & Wave Time Histories in Malaysia, Sabah and Sarawak Basin Using Autoregressive, Integrated Moving Average (ARIMA) Method,"* Final Year Project Thesis, Civil Engineering, Universiti Teknologi Petronas, Malaysia, 2011.

[4] E.P. Box. George, and M. Jenkins. Gwilym, "*Time Series Analysis: Forecasting and Control,"* John Wiley & Sons, Inc., Canada, 2008.

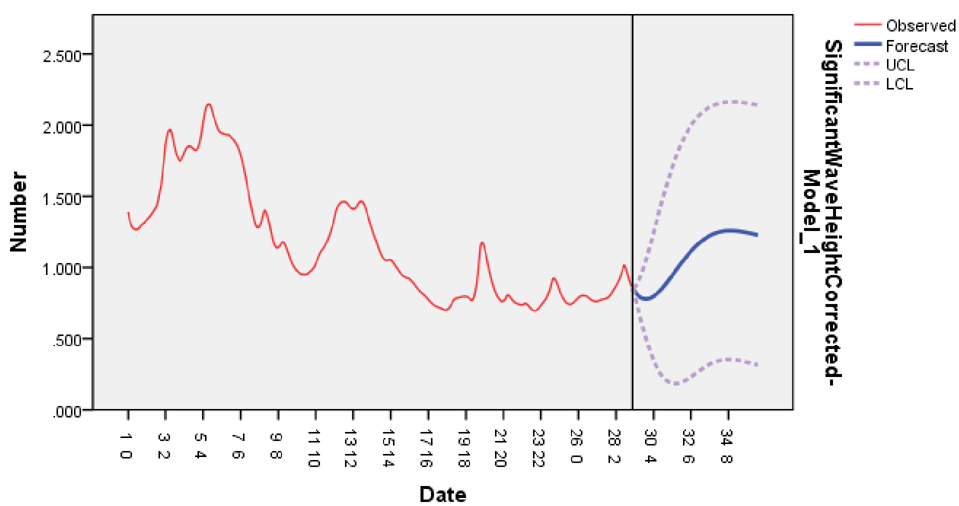
[5] J.C. Paul, Md.S. Hoque and M. Morshedur Rahman, "Selection of Best ARIMA Model for Forecasting Average Daily Share Price Index of Pharmaceutical Companies in Bangladesh: A Case Study on Square Pharmaceutical Ltd," *Global Journal of Management and Business Research Finance*, vol. 13, (3), 2013.

APPENDICES

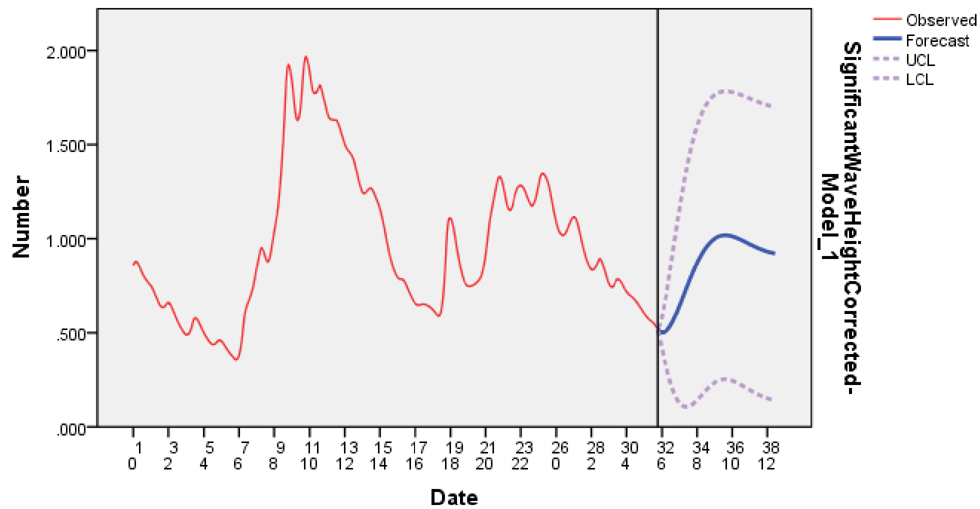
Dulang



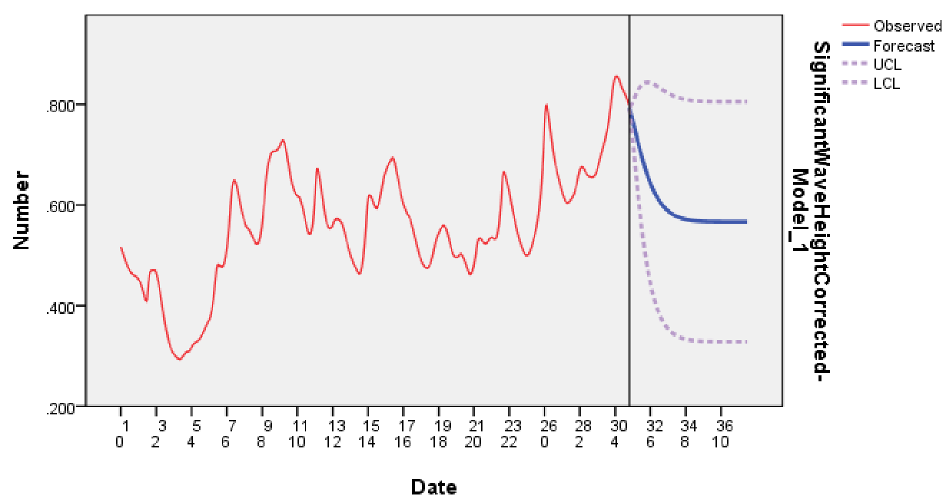
Using ARIMA (2, 0, 1) for January 2003 time series (hourly data) to forecast from 1/2/2003 -7/2/2003 (1 week)



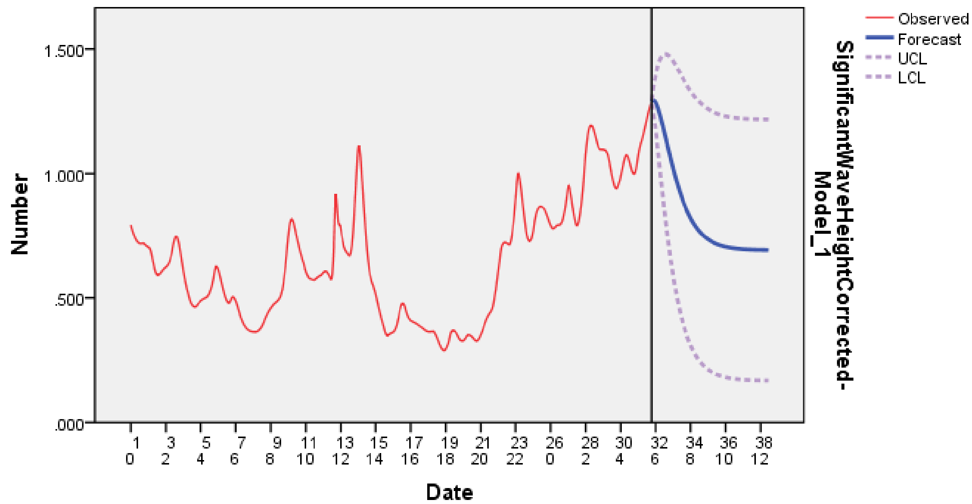
Using ARIMA (2, 0, 1) for February 2003 time series (hourly data) to forecast from 1/3/2003 -7/3/2003 (1 week)



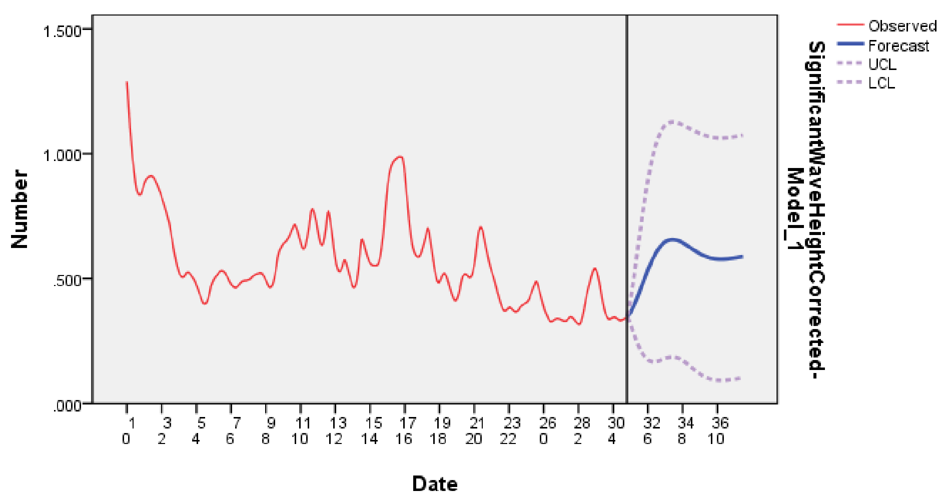
Using ARIMA (2, 0, 1) for March 2003 time series (hourly data) to forecast from 1/4/2003 -7/4/2003 (1 week)



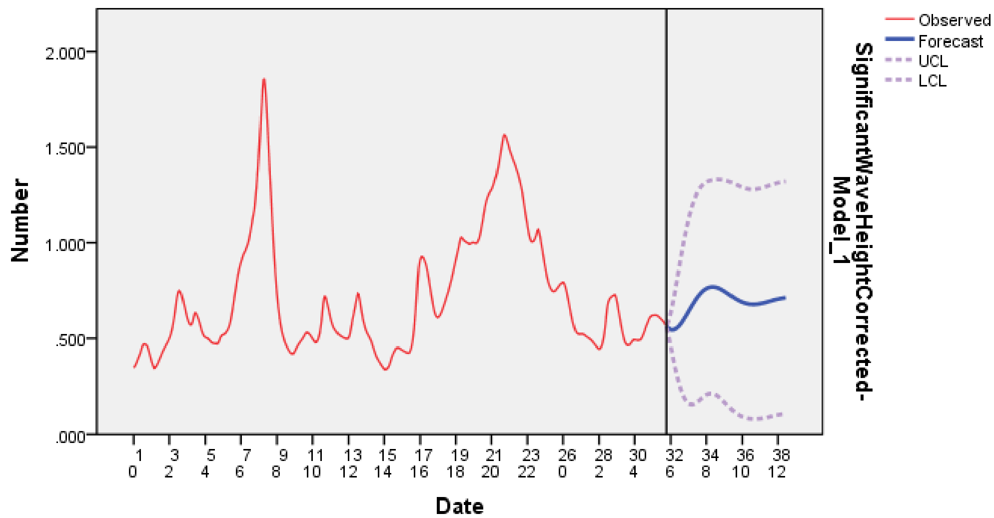
Using ARIMA (2, 0, 3) for April 2003 time series (hourly data) to forecast from 1/5/2003 -7/5/2003 (1 week)



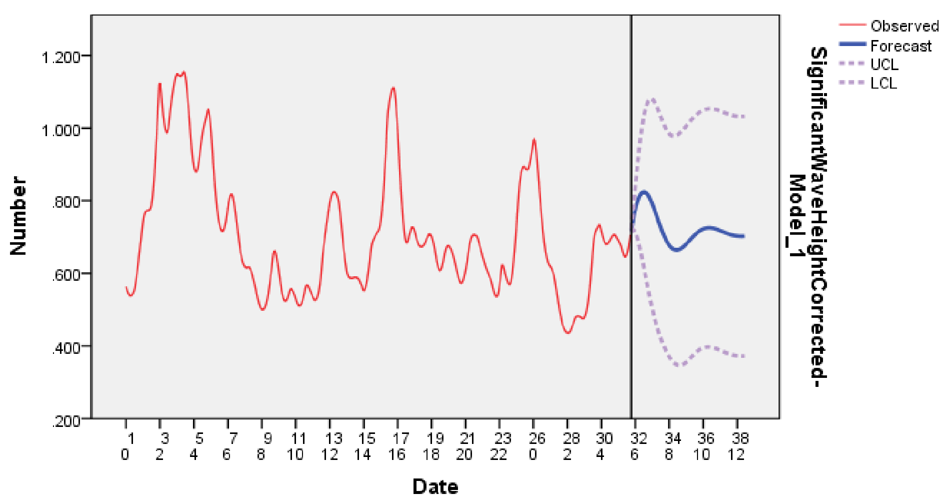
Using ARIMA (2, 0, 1) for May 2003 time series (hourly data) to forecast from 1/6/2003 -7/6/2003 (1 week)



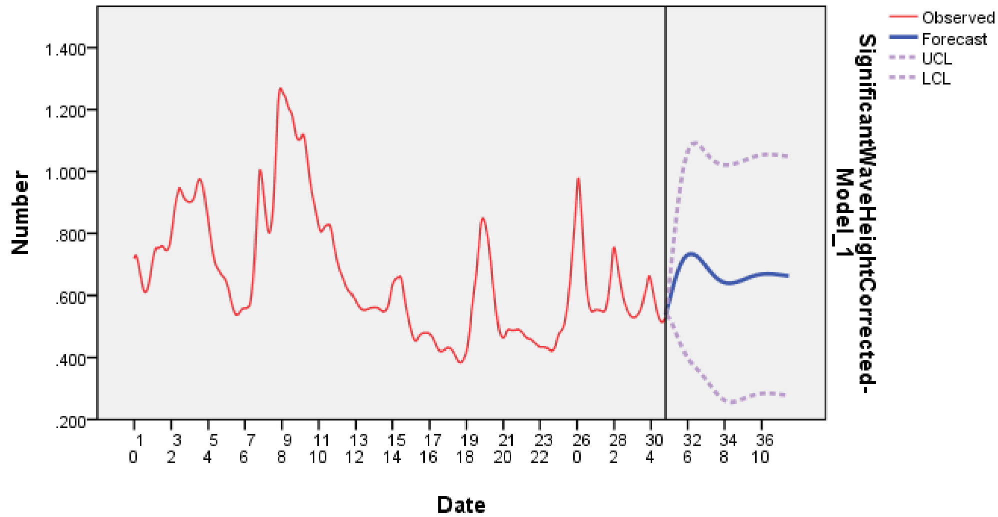
Using ARIMA (2, 0, 2) for June 2003 time series (hourly data) to forecast from 1/7/2003 -7/7/2003 (1 week)



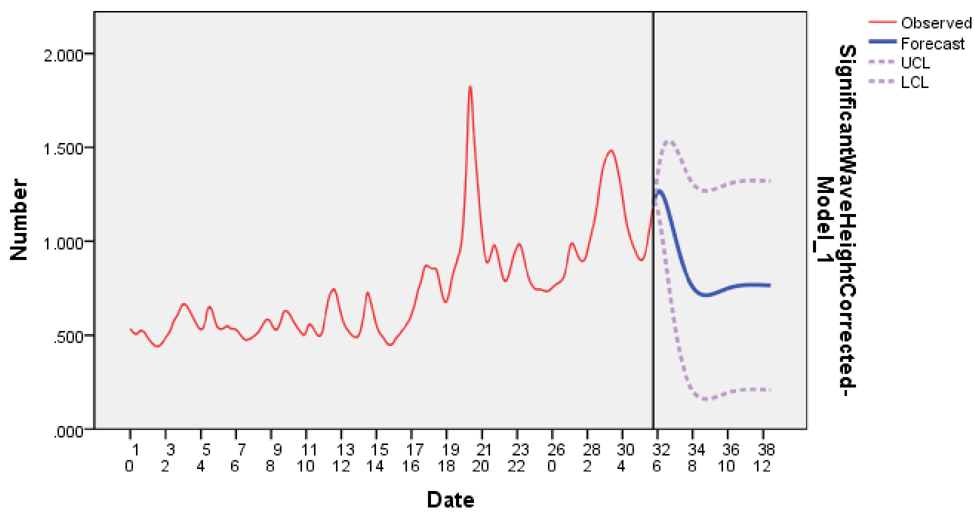
Using ARIMA (2, 0, 1) for July 2003 time series (hourly data) to forecast from 1/8/2003 -7/8/2003 (1 week)



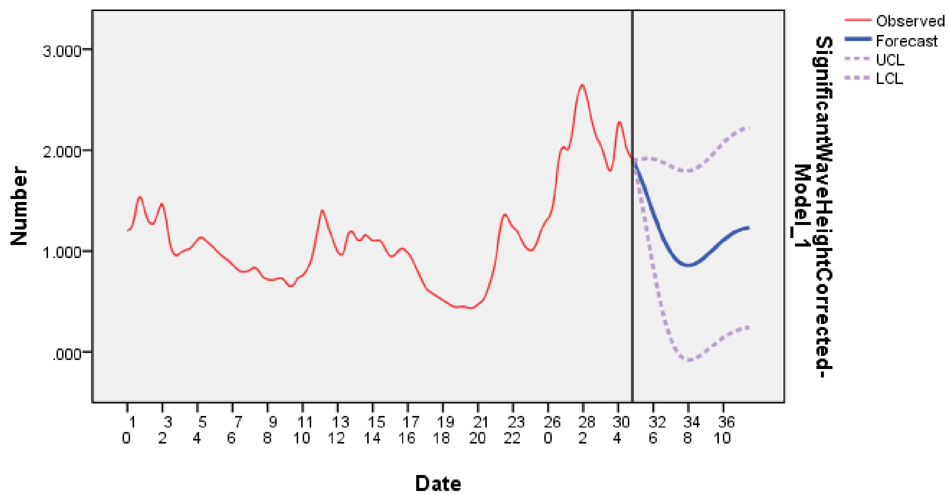
Using ARIMA (2, 0, 1) for August 2003 time series (hourly data) to forecast from 1/9/2003 -7/9/2003 (1 week)



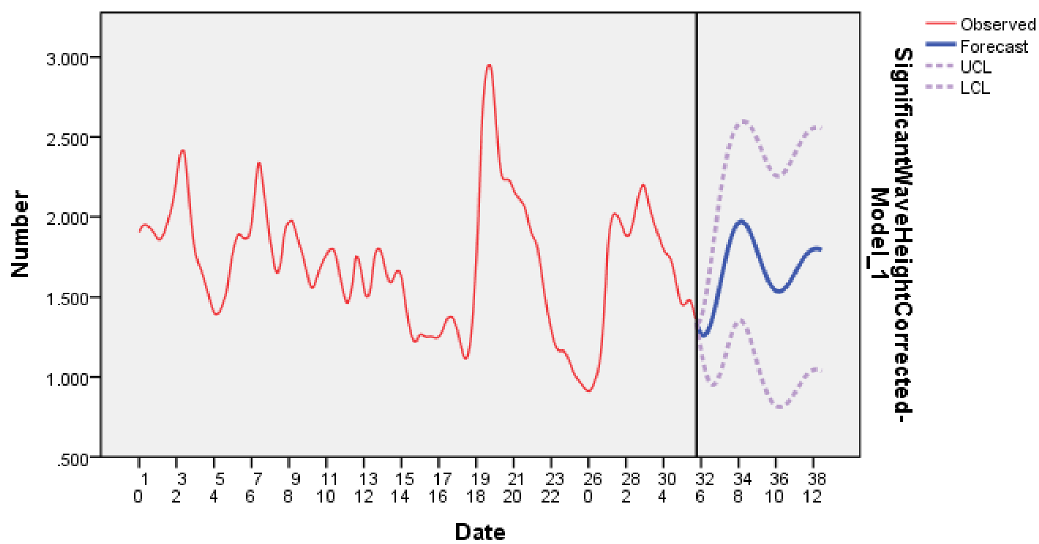
Using ARIMA (2, 0, 1) for September 2003 time series (hourly data) to forecast from 1/10/2003 -7/10/2003 (1 week)



Using ARIMA (2, 0, 3) for October 2003 time series (hourly data) to forecast from 1/11/2003 -7/11/2003 (1 week)

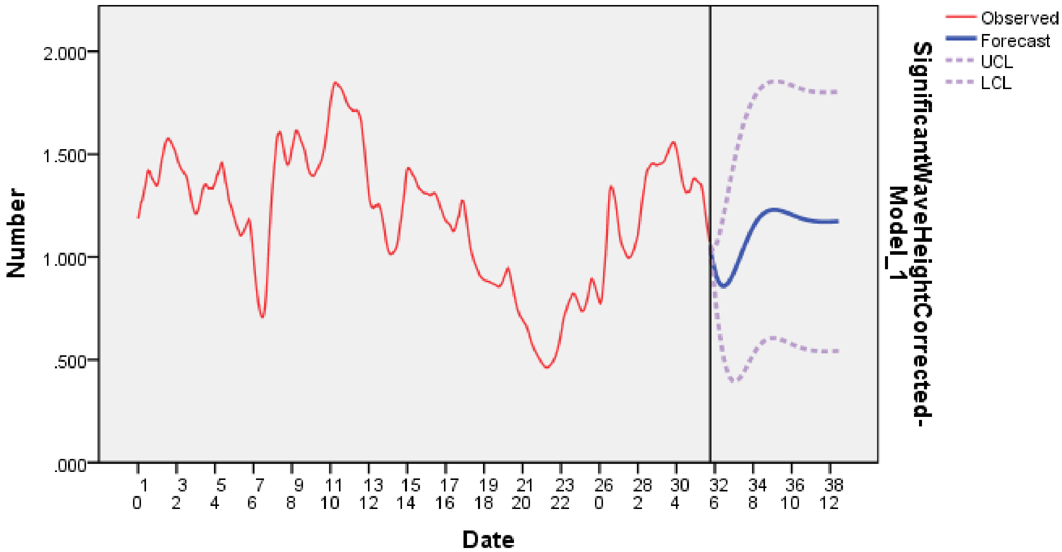


Using ARIMA (2, 0, 1) for November 2003 time series (hourly data) to forecast from 1/12/2003 -7/12/2003 (1 week)

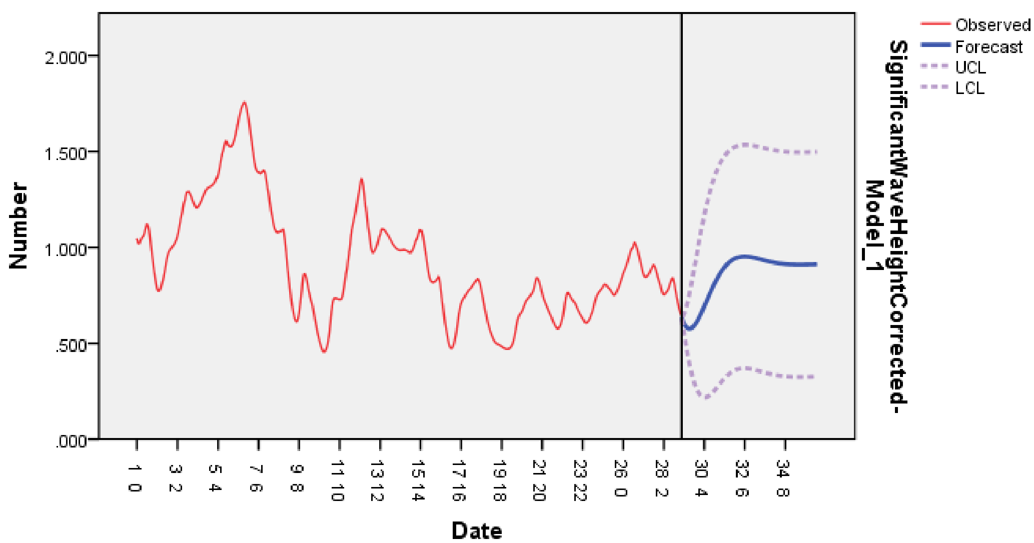


Using ARIMA (2, 0, 1) for December 2003 time series (hourly data) to forecast from 1/1/2004 -7/1/2004 (1 week)

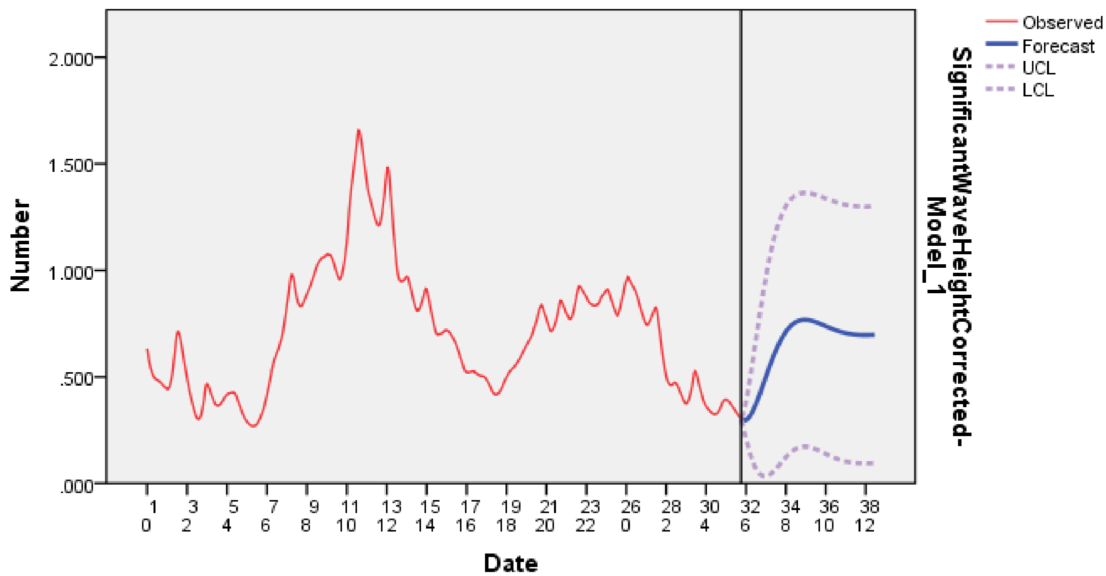
Erb West



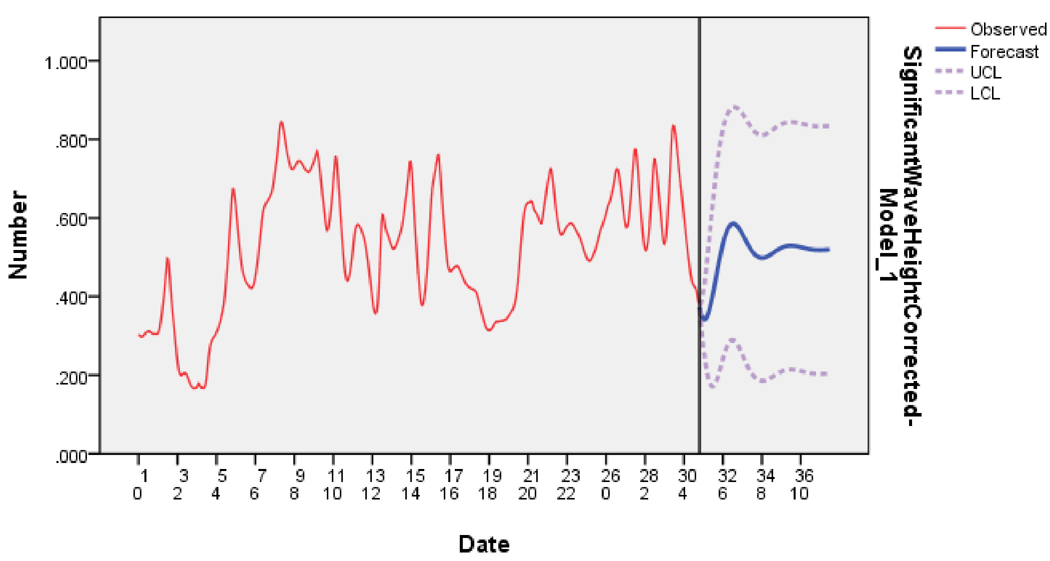
Using ARIMA (2, 0, 1) for January 2003 time series (hourly data) to forecast from 1/2/2003 -7/2/2003 (1 week)



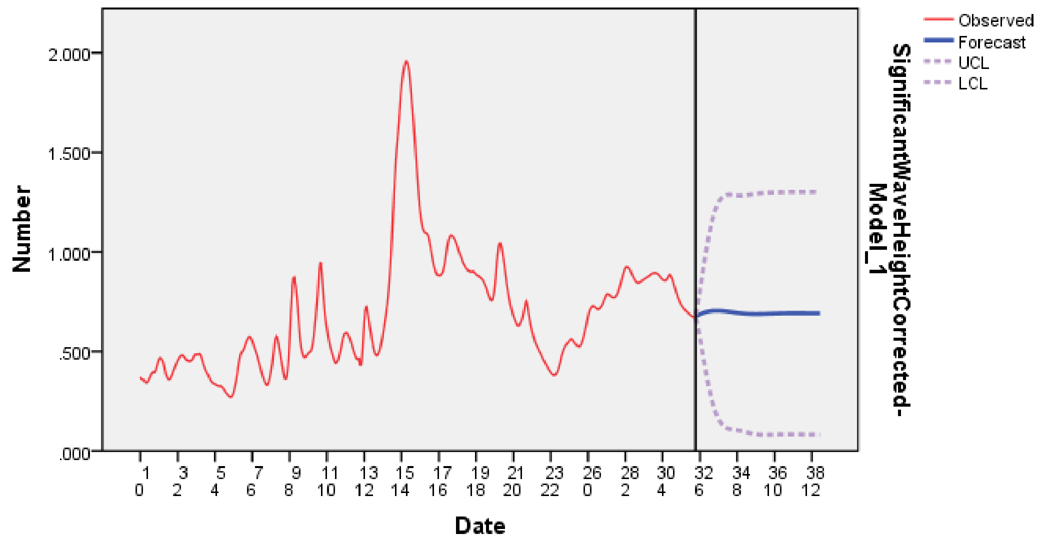
Using ARIMA (2, 0, 2) for February 2003 time series (hourly data) to forecast from 1/3/2003 -7/3/2003 (1 week)



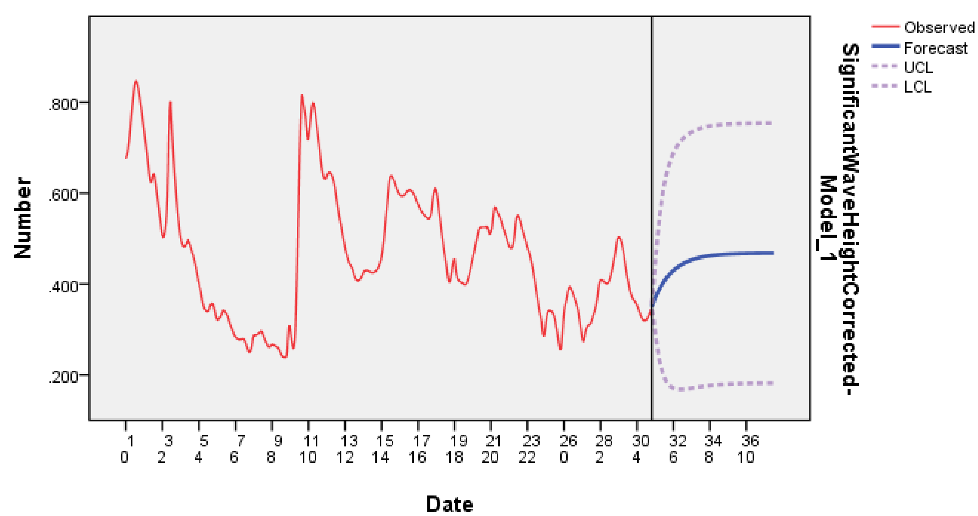
Using ARIMA (2, 0, 1) for March 2003 time series (hourly data) to forecast from 1/4/2003 -7/4/2003 (1 week)



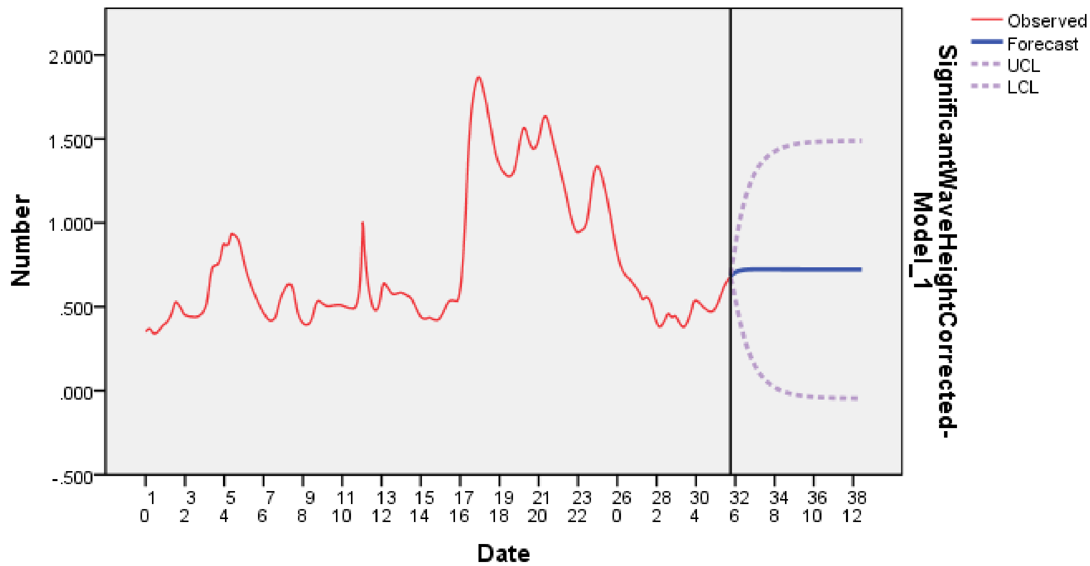
Using ARIMA (2, 0, 1) for April 2003 time series (hourly data) to forecast from 1/5/2003 -7/5/2003 (1 week)



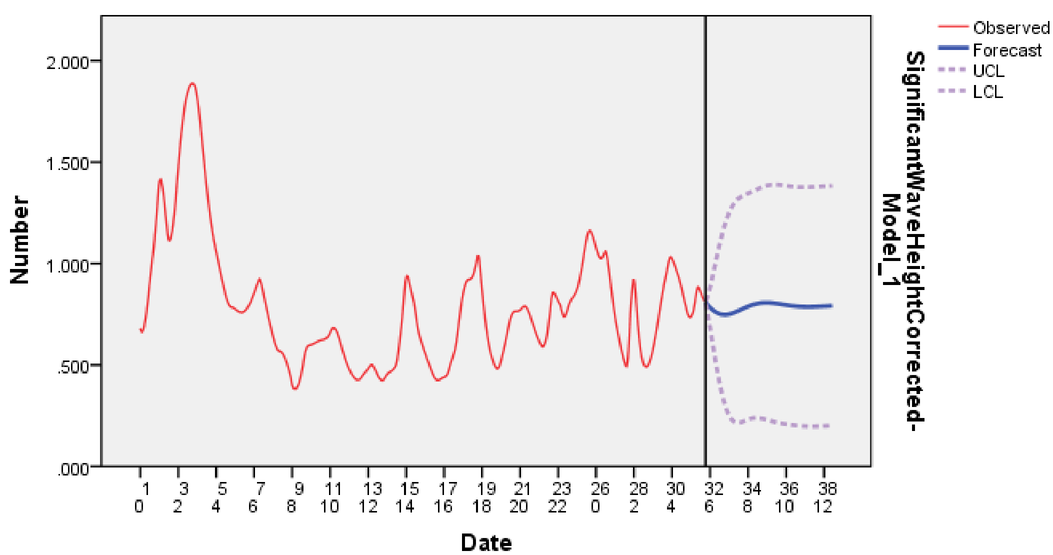
Using ARIMA (2, 0, 1) for May 2003 time series (hourly data) to forecast from 1/6/2003 -7/6/2003 (1 week)



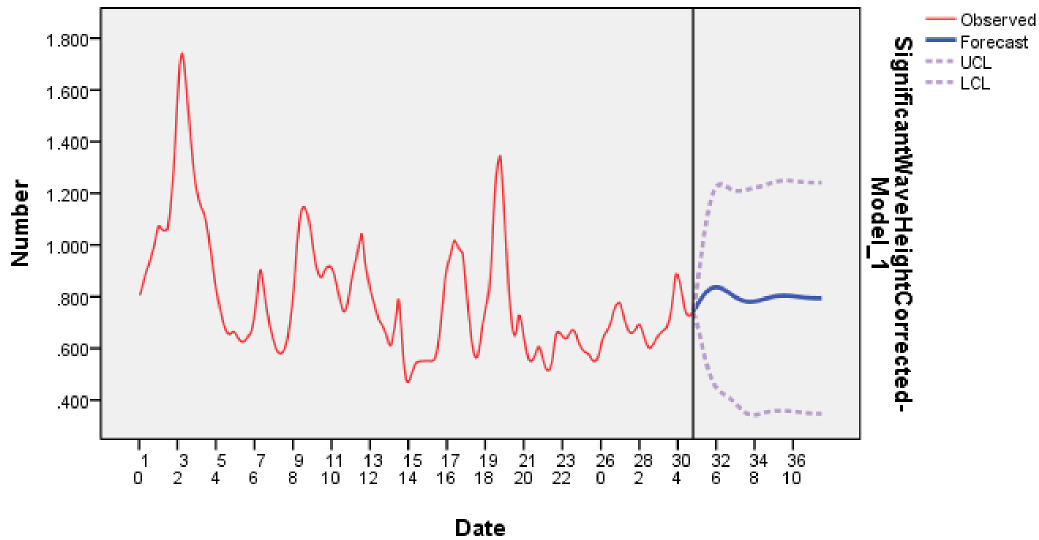
Using ARIMA (2, 0, 1) for June 2003 time series (hourly data) to forecast from 1/7/2003 -7/7/2003 (1 week)



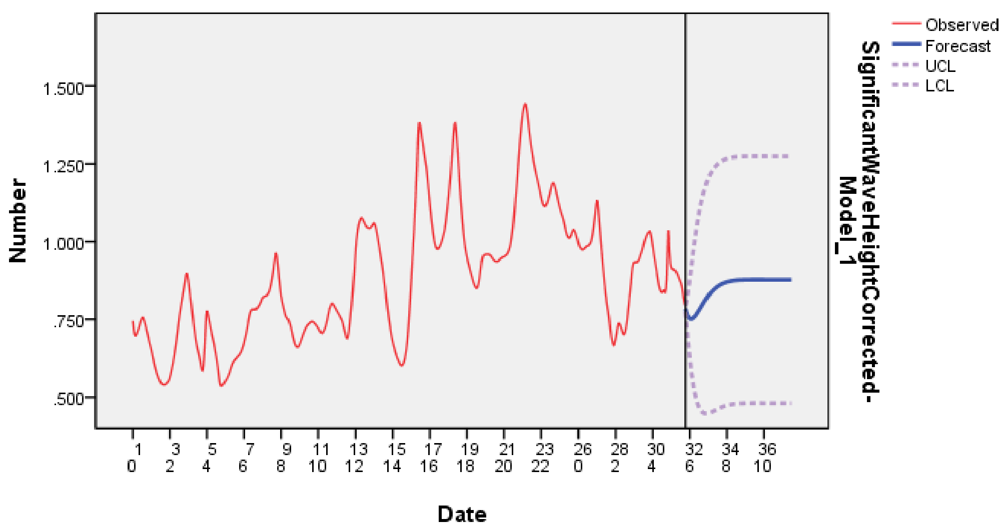
Using ARIMA (2, 0, 3) for July 2003 time series (hourly data) to forecast from 1/8/2003 -7/8/2003 (1 week)



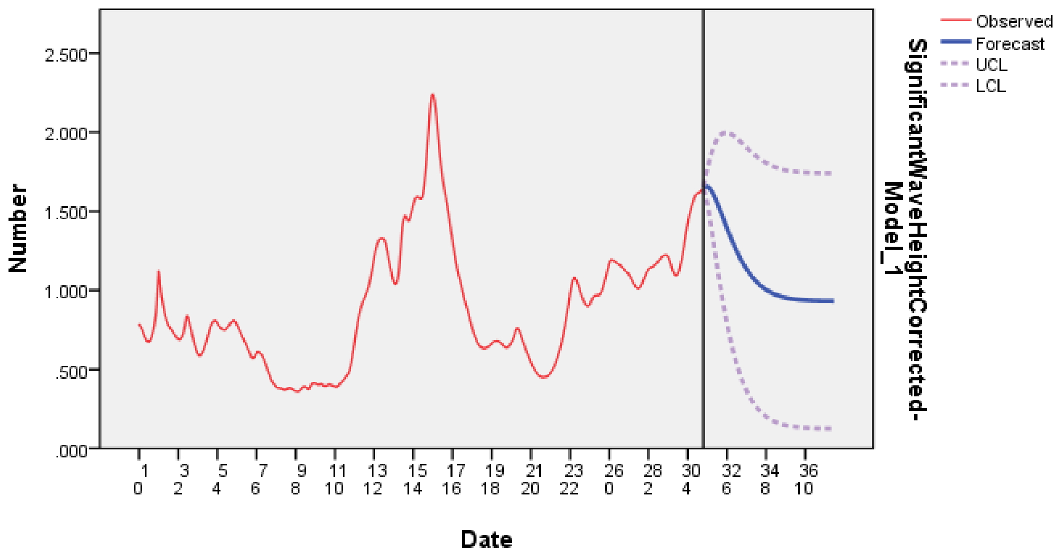
Using ARIMA (2, 0, 1) for August 2003 time series (hourly data) to forecast from 1/9/2003 -7/9/2003 (1 week)



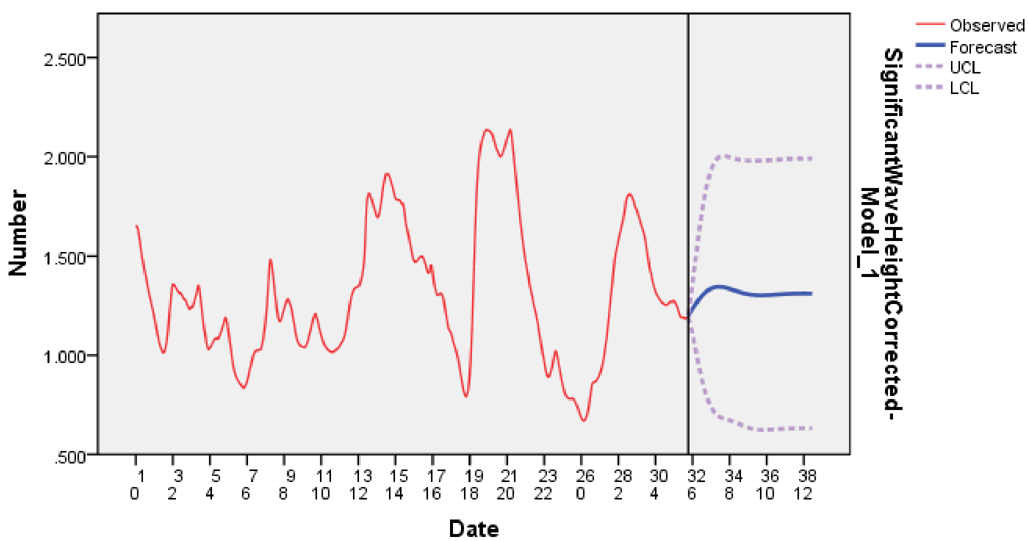
Using ARIMA (2, 0, 1) for September 2003 time series (hourly data) to forecast from 1/10/2003 -7/10/2003 (1 week)



Using ARIMA (2, 0, 1) for October 2003 time series (hourly data) to forecast from 1/11/2003 -7/11/2003 (1 week)

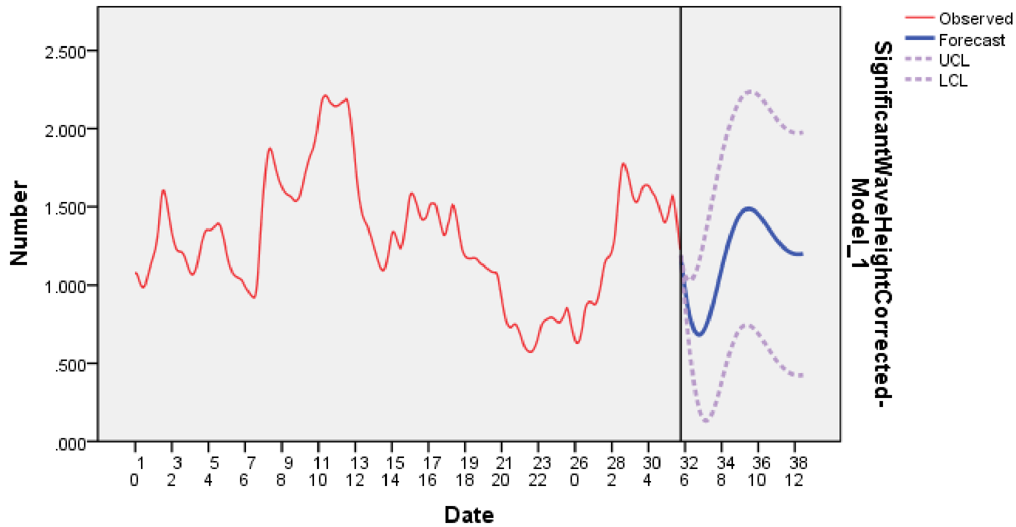


Using ARIMA (2, 0, 2) for November 2003 time series (hourly data) to forecast from 1/12/2003 -7/12/2003 (1 week)

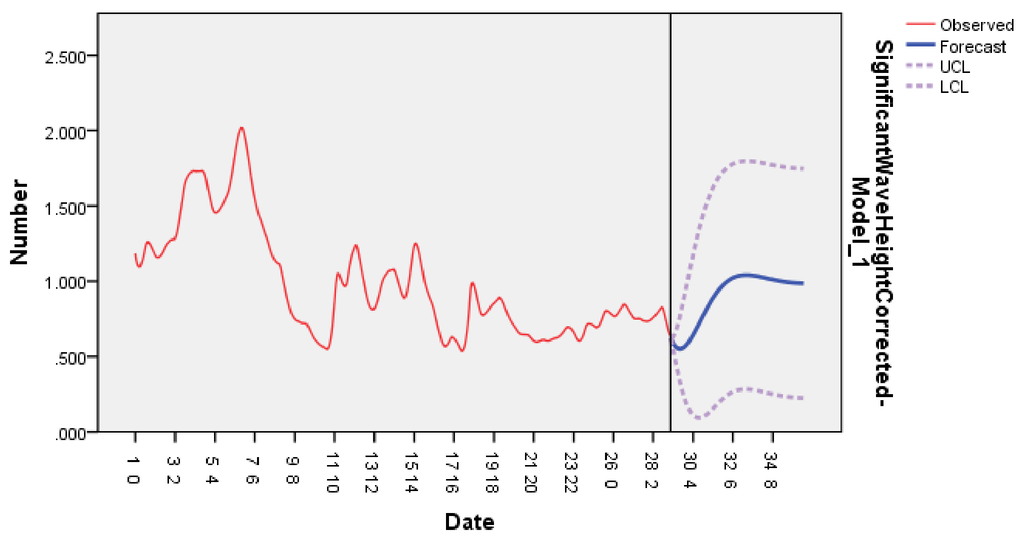


Using ARIMA (2, 0, 3) for December 2003 time series (hourly data) to forecast from 1/1/2004 -7/1/2004 (1 week)

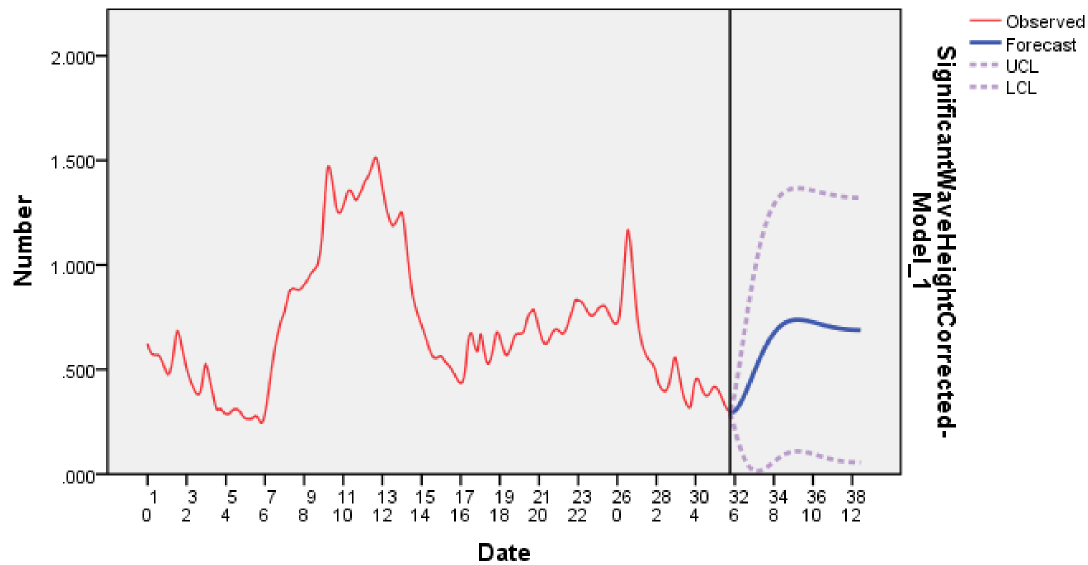
Tukau



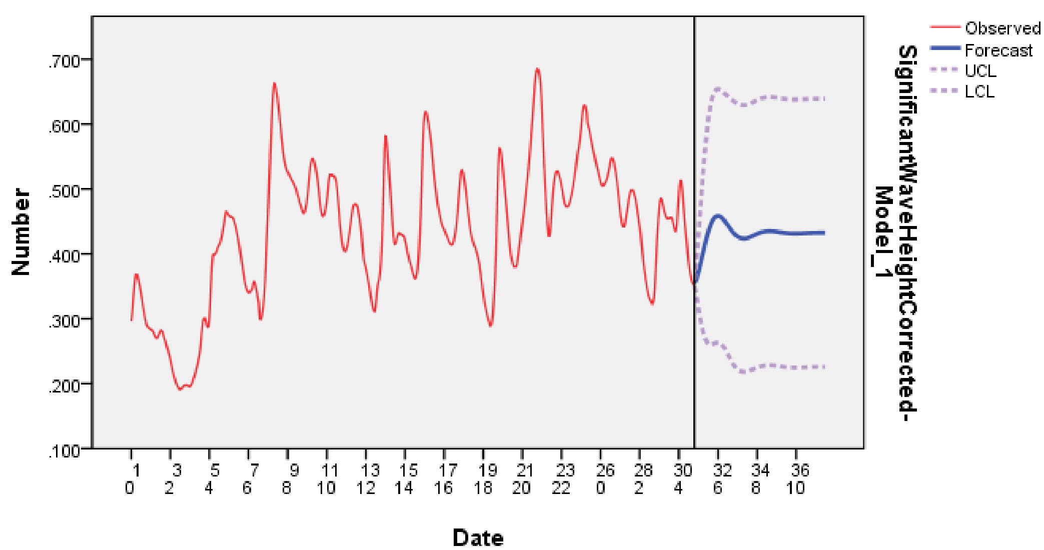
Using ARIMA (2, 0, 1) for January 2003 time series (hourly data) to forecast from 1/2/2003 -7/2/2003 (1 week)



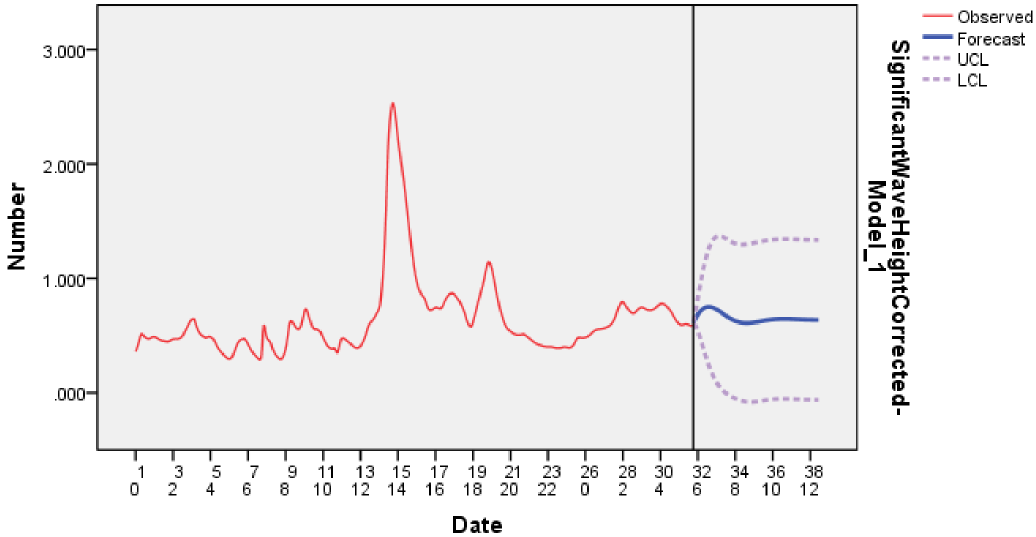
Using ARIMA (2, 0, 1) for February 2003 time series (hourly data) to forecast from 1/3/2003 -7/3/2003 (1 week)



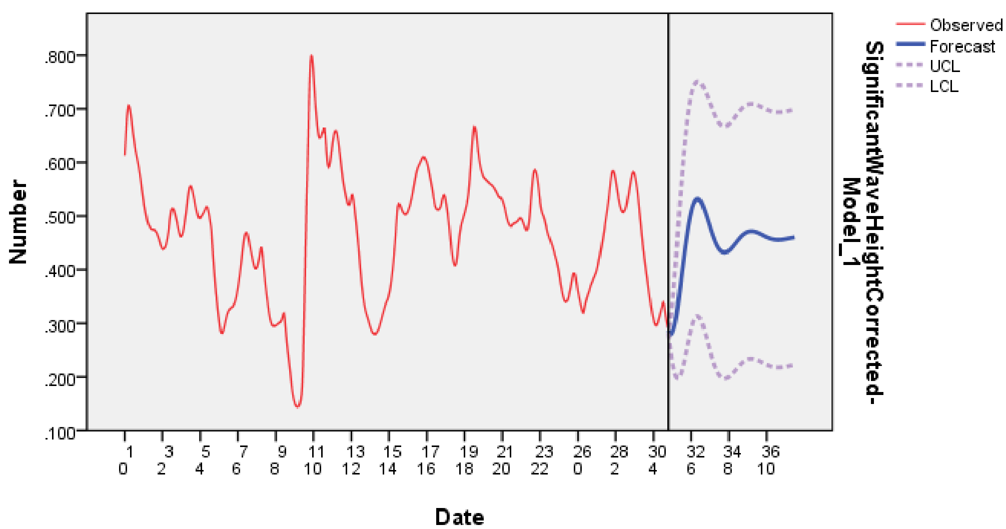
Using ARIMA (2, 0, 1) for March 2003 time series (hourly data) to forecast from 1/4/2003 -7/4/2003 (1 week)



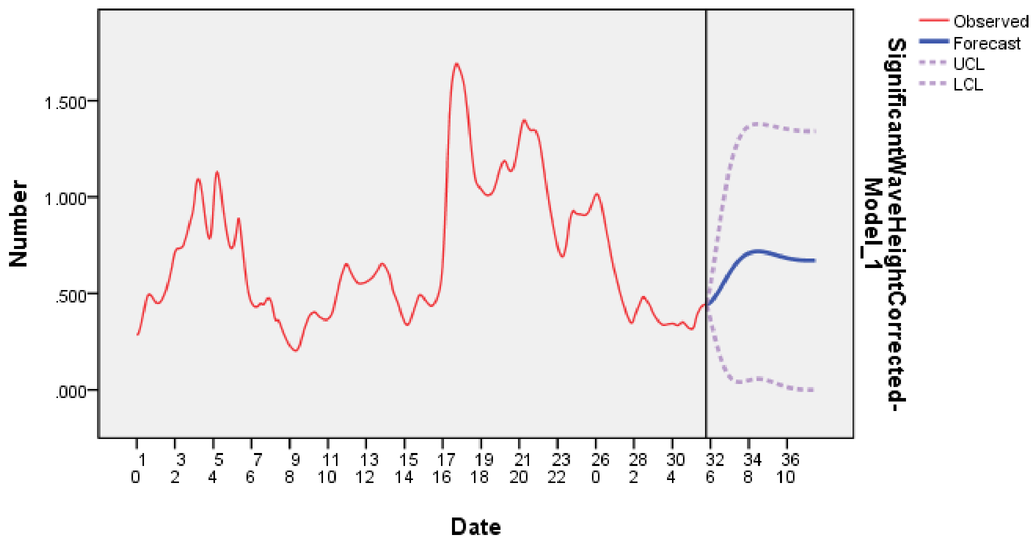
Using ARIMA (2, 0, 1) for April 2003 time series (hourly data) to forecast from 1/5/2003 -7/5/2003 (1 week)



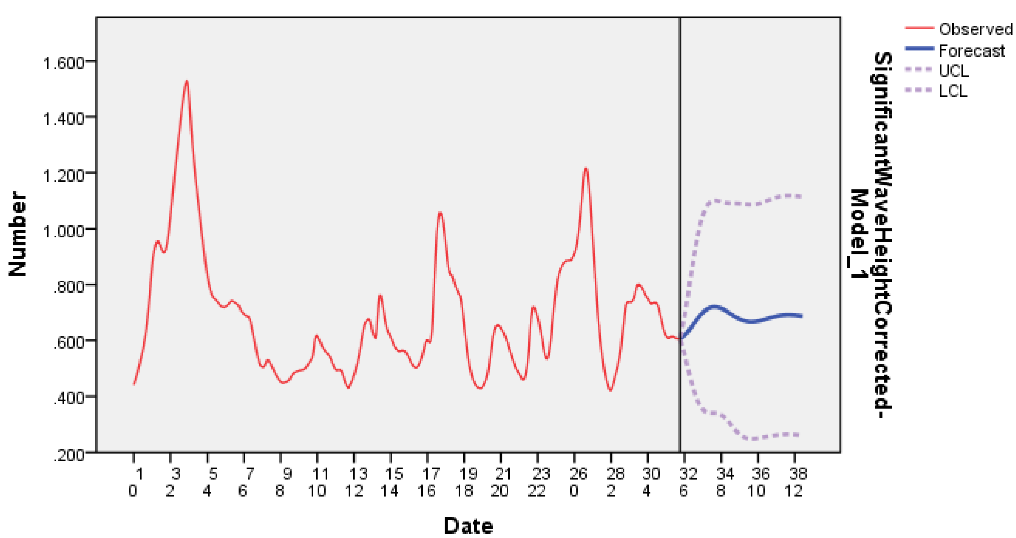
Using ARIMA (2, 0, 1) for May 2003 time series (hourly data) to forecast from 1/6/2003 -7/6/2003 (1 week)



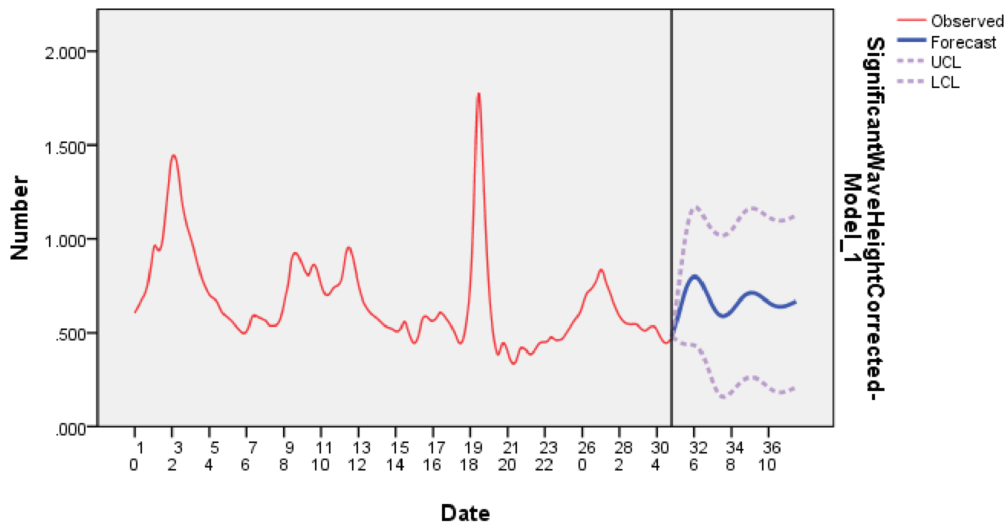
Using ARIMA (2, 0, 3) for June 2003 time series (hourly data) to forecast from 1/7/2003 -7/7/2003 (1 week)



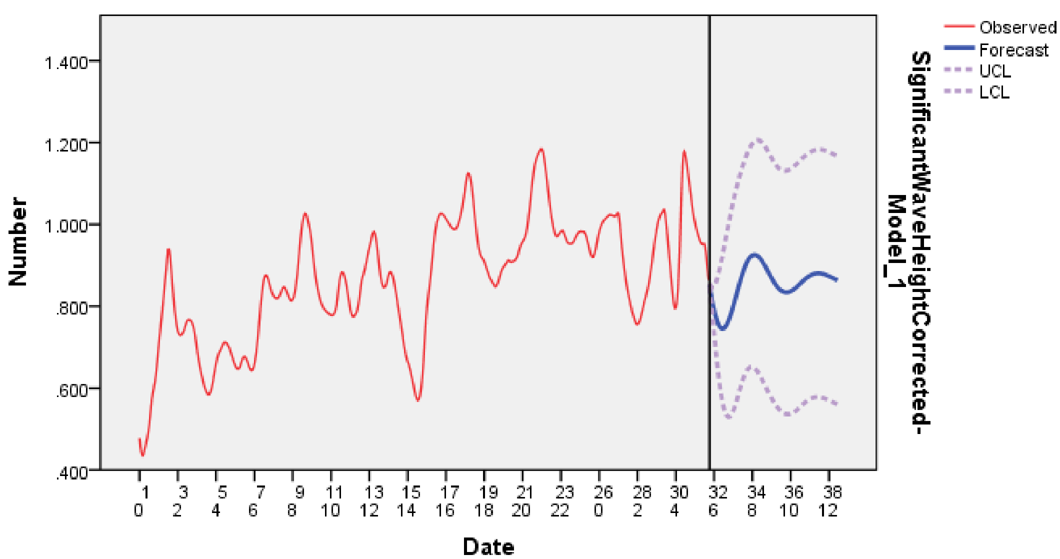
Using ARIMA (2, 0, 1) for July 2003 time series (hourly data) to forecast from 1/8/2003 -7/8/2003 (1 week)



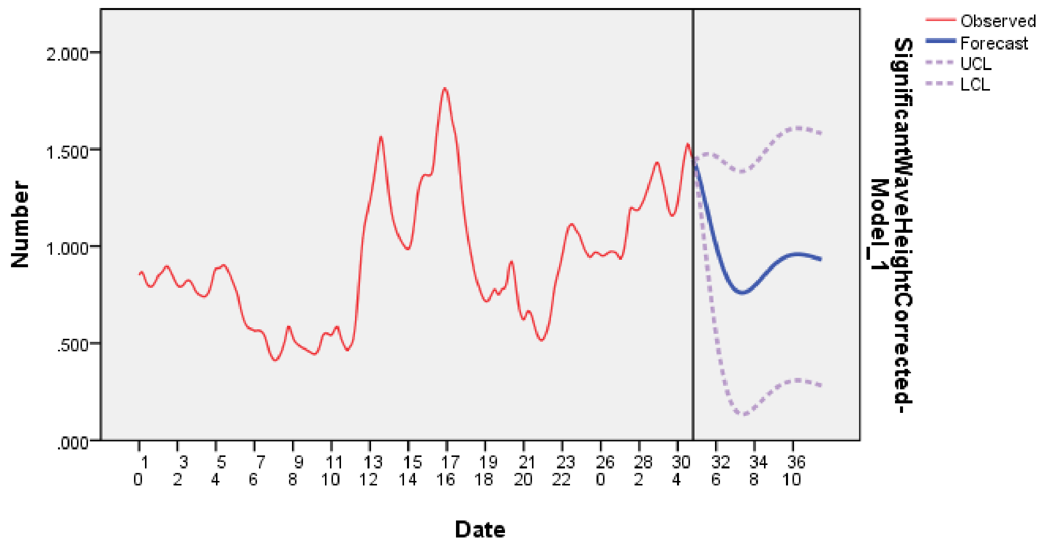
Using ARIMA (2, 0, 1) for August 2003 time series (hourly data) to forecast from 1/9/2003 -7/9/2003 (1 week)



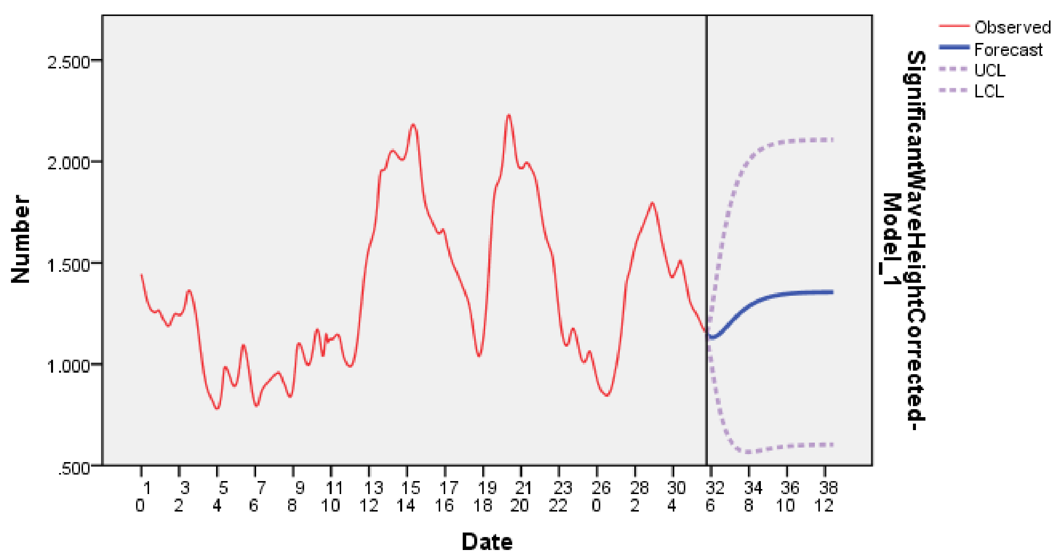
Using ARIMA (2, 0, 1) for September 2003 time series (hourly data) to forecast from 1/10/2003 -7/10/2003 (1 week)



Using ARIMA (2, 0, 3) for October 2003 time series (hourly data) to forecast from 1/11/2003 -7/11/2003 (1 week)



Using ARIMA (2, 0, 1) for November 2003 time series (hourly data) to forecast from 1/12/2003 -7/12/2003 (1 week)



Using ARIMA (2, 0, 1) for December 2003 time series (hourly data) to forecast from 1/1/2004 -7/1/2004 (1 week)