# Time Series Analysis On Mackerel (Scombridae) Landings in Malaysia

## Husna Afzan Zakaria[1], Mohd Saifullah Rusiman[2, *], Abdul Wahab Abdullah[3], Muhammad Ammar Shafi[4]

[1,2]Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, UTHM Pagoh, 84600 Muar, Johor, MALAYSIA

[3]FRI Kg Acheh, Kompleks Perikanan Kg Acheh, 32000 Sitiawan, Perak, MALAYSIA

[4]Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, MALAYSIA

*Corresponding Author : saifulah@uthm.edu.my

**Abstract**: Mackerel fish is one type of pelagic fish that live in the surface of the ocean. It is also have benefits in terms of protein which also has high demand in Asian and others countries and helps gaining profits in fisheries industries. This study aims to predict mackerel landings in Malaysia in one year advance which is 2018. The data of 132 monthly of mackerel landings from year 2007 until 2017 is used to make a prediction of mackerels landing by using four methods which are Seasonal Autoregressive Integrated Moving Average (SARIMA) method, Multiplicative Holt-Winters, Additive Holt-Winters Method and Simple Exponential Smoothing method. The aim is to compare the performance among four methods by measuring the accuracy of each method. The result shows that Additive Holt-Winters method is the best method used to forecast mackerel landings in 2018 with the lowest value of Mean Absolute Percentage Error (MAPE) and Mean Square Error (MSE). In conclusion, the potential result from this study could be used by fish farmers in their annual planning of supplying fish in Malaysia.

## 1. Introduction

Pelagic fish are found near in the surface of the ocean or in the middle depths of the ocean. Then, most of the pelagic fish is known as one type of species that exhibit behaviors that can increase their catchability. Pelagic fishes occupy the surface of the waters in the ocean [1]. Economically, pelagic fish

is important. Pelagic fish also commercially and traditionally accumulating the fishing gear that contributed towards the total marine landings in Malaysia [2]. Mackerel is one type of small pelagic fish that has high demand due to its main source of protein [3].

Mackerel landings has obtained high demand in Asian trade which import and export to others countries. Therefore, time series analysis is used to make a prediction of mackerel landings. The ability of fishery industries in Malaysia also is determined. Mackerel also contribute in shortfall supply of canning from national countries during fishing seasons [4]. There are researchers that used Autoregressive Integrated Moving Average (ARIMA) and Holt-Winters method to be apply in time series that exhibit seasonality [4].

Forecasting the fish price have a long history of study, for example, Hudson & Capps [5] forecasted the hard blue crabs price using time series method and econometric method. Vukina & Anderson [6] applied the State-Space method to forecast the price of salmon and the result showed that four out of five times series follow random walk and cyclic behaviour was indicated in time series of prices. Park [7] forecasted the Walleye Pollock landings in Korea using seasonal autoregressive integrated moving average (SARIMA) and the result showed the fluctuations of the monthly catches of Walleye Pollock.

In order to make a time series analysis, the forecast accuracy was used to measure the performance of method used by obtaining the lowest value of error. The lowest value of error will show that model tend to be good [8]. Autoregressive Integrated Moving Average (ARIMA), Multiplicative Holt–winters, Additive Holt-Winters and Simple Exponential Smoothing Methods will be used. Moreover, the comparison will be made and the methods with lowest values of Mean Absolute Percentage Error (MAPE) and Mean Square Error (MSE) will be selected as the best model.

## 2. Materials and Methods

### 2.1 Materials

The data for this study were obtained from the Fisheries Department of Malaysia. The data that involved were the monthly landings of Mackerel in Malaysia from January 2007 until December 2017. The method that will be used to forecast the monthly landings of Mackerel were Seasonal Autoregressive Integrated Moving Average (SARIMA) method, Multiplicative Holt-Winters, Additive Holt-Winters Method and Simple Exponential Smoothing method. The criterion that will be used as forecast accuracy to measure the performance among the four method were the MSE and MAPE values.

### 2.2 Methods

#### 2.2.1 Seasonal Autoregressive Integrated Moving Average (SARIMA)

Before using ARIMA model, the data should fulfill the normality assumptions. Transformation is used to stabilize the variance when the data show different variation at different levels of series. The equation of $ARIMA(p, d, q)$ can be written as described in Eq. 1 - Eq. 5 [7],

$$\text{Square root } w_t = \sqrt{y_t} \qquad \qquad Eq.\ 1$$
$$\text{Cube root } w_t = \sqrt[3]{y_t} \qquad \qquad Eq.\ 2$$
$$\text{Logarithm } w_t = \log(y_t) \qquad \qquad Eq.\ 3$$

where;

$\quad\quad y_t$ = actual value

Then, differencing is used to stabilize the mean.
$$y_t = y_{t-1} + e_t \qquad \qquad Eq.\ 4$$
$$y_t'' = y_t' - y_{t-1}' \qquad \qquad Eq.\ 5$$
where;

$\quad\quad y_t$ = actual value

$y_{t-1}$ = previous value
$y_t'$ = First differencing
$y_t''$ = Second differencing
$e_t$ = error

Table 1 shows the identification model of ARIMA models by referring to Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plot of the data [8].

**Table 1: Identification model of ARIMA models**

| ACFs | PACFs | Model |
|---|---|---|
| Decay to zero with exponential pattern | Cuts off after lag p | AR ($p$) |
| Cuts off after lag q | Decay to zero with exponential pattern | MA ($q$) |
| Decay to zero with exponential pattern | Decay to zero with exponential pattern | ARMA ($p,q$) |
| Cuts off after lag q | Cuts off after lag p | AR ($p$) or M ($q$) |

After that, determination of optimal model parameters is a crucial step in model selection. The model with the lowest values of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) will be selected [8] as stated in Eq. 6 and Eq. 7.

$$AIC(p) = n \ln\left(\frac{\hat{\sigma}_e^2}{n}\right) + 2p \qquad Eq.\ 6$$

$$BIC(p) = n \ln\left(\frac{\hat{\sigma}_e^2}{n}\right) + p + p \ln(n) \qquad Eq.\ 7$$

where;

$n$ = number of effective observations which used to fit the model
$p$ = number of parameters in the model
$\hat{\sigma}_e^2$ = sum of sample squared residuals

Seasonal difference followed first difference [8] as in Eq. 8,

$$(1 - B)(1 - B^m)y_t \qquad Eq.\ 8$$

where;

$B$ = backward shift operator
$B^m$ = backward shift operator of monthly
$y_t$ = actual value

General Box-Jenkins model which allocated seasonality as in Eq. 9,

$$\emptyset_p(B)\nabla^d y_t = \theta_q(B)\varepsilon_t \qquad Eq.\ 9$$

where;

$$\nabla(B) = 1 - B \qquad Eq.\ 10$$
$$\emptyset_p(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p \qquad Eq.\ 11$$
$$\theta_q = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \qquad Eq.\ 12$$
$$B^j y_t = y_{t-h} \qquad Eq.\ 13$$

## 2.2.2 Additive Holt-Winter Method

Holt-Winters is used when the data are exhibit trend and seasonal variation [6]. Below are the procedure of Additive-Holt-Winters model [9].

The least squares trend line is fitted as in Eq. 14,

$$\hat{y}_t = l_0 + b_0 t \qquad \qquad Eq.\ 14$$

where;

$l_0$= y-intercept
$b_0$= slope
$t$= time

Initial seasonal factors is defined as in Eq. 15,

$$S_t = y_t - \hat{y}_t \qquad \qquad Eq.\ 15$$

where;

$y_t$= actual value
$\hat{y}_t$= trend line

Average of seasonal values is shown as in Eq. 16,

$$\bar{S}_t = \frac{S_t + S_{t+1} + S_{t+2} + S_{t+3}}{4} \qquad Eq.\ 16$$

where;

$S_t$= seasonal factors

Point forecast from time 0 is defined as in Eq. 17,

$$\hat{y}_{T+p} = l_T + pb_T + sn_{T+P-1} \qquad Eq.\ 17$$

where;

$l_T$= level
$b_t$= trend
$sn_T$= seasonal

A seasonal equation is stated as in Eq. 18- Eq. 21,

$$\text{Level } L_t = \propto (y_t - S_{t-s}) + (1-\propto)(L_{t-1} + b_{t-1}) \qquad Eq.\ 18$$
$$\text{Trend } b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \qquad Eq.\ 19$$
$$\text{Seasonal } S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-s} \qquad Eq.\ 20$$
$$\text{Forecast } F_{t+k} = L_t + kb_t + S_{t+k-s} \qquad Eq.\ 21$$

where;
$S_{t-s}$= seasonal of seasonal before,
$L_{t-1}$= level before,
$L_t$= level,
$b_{t-1}$= trend before,
$k$= value of forecast
s = length of seasonal cycle,$0 \leq \propto \leq 1, 0 \leq \beta \leq 1, 0 \leq \gamma \leq 1$

## 2.2.3 Multiplicative Holt-Winters Method

The procedure below are for Multiplicative Holt-Winters method [9].

Obtain initial values for the level $l_0$,the growth rate $b_0$ and the seasonal factors $sn_{-3}, sn_{-2}, sn_{-1}$ and $sn_0$ (fitting least square trend line) as in Eq. 22,

$$\hat{y}_t = l_0 + b_0 t \qquad\qquad Eq.\ 22$$

where;

$l_0$ = y-intercept,
$b_0$ = slope,
$t$ = time

Initial seasonal factors described in Eq. 23 – Eq. 27:

$$\hat{y}_t = l_0 + b_0 t \qquad\qquad Eq.\ 23$$
$$S_t = \frac{y_t}{\hat{y}_t} \qquad\qquad Eq.\ 24$$
$$\bar{S}_t = \frac{S_t + S_{t+1} + S_{t+2} + S_{t+3}}{4} \qquad\qquad Eq.\ 25$$
$$CF = \frac{L}{\sum_{i=1}^{L} \bar{S}_{[i]}} \qquad\qquad Eq.\ 26$$
$$sn_{1-L} = \bar{S}_{[i]}(CF) \qquad\qquad Eq.\ 27$$

where;

$S_t$ = seasonal factors,
$\bar{S}_{[i]}$ = average of seasonal factor,
$CF$ = cumulative Frequency

Point forecast from time 0 stated in Eq. 28,

$$\hat{y}_{T+p} = l_T + p b_T + sn_{T+P-1} \qquad\qquad Eq.\ 28$$

where;

$l_T$ = level,
$b_t$ = trend,
$sn_T$ = seasonal

Update the estimates of factors as in Eq. 29 – Eq. 32:

$$\text{Level } L_t = \propto \left(\frac{y_t}{S_{t-s}}\right) + (1-\propto)(L_{t-1} + b_{t-1}) \qquad Eq.\ 29$$
$$\text{Trend } b_t = \beta \left(\frac{L_t}{L_{t-1}}\right) + (1-\beta)b_{t-1} \qquad Eq.\ 30$$
$$\text{Seasonal } S_t = \gamma \left(\frac{y_t}{L_t}\right) + (1-\gamma)S_{t-s} \qquad Eq.\ 31$$
$$\text{Forecast } F_{t+k} = (L_t + k b_t)S_{t+k-s} \qquad Eq.\ 32$$

where;

$\propto, \beta, \gamma$ = predetermined values,
$S_{t-s}$ = seasonal of seasonal before,
$L_{t-1}$ = level before,
$L_t$ = level,
$b_{t-1}$ = trend before,
$k$ = value of forecast
$s$ = length of seasonal cycle, $0 \leq \propto \leq 1, 0 \leq \beta \leq 1, 0 \leq \gamma \leq 1$.

## 2.2.4 Simple Exponential Smoothing Method

Compute the initial estimate of mean of the series at time period $t = 0$ as in Eq. 33 [4].

$$l_0 = \bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} \qquad\qquad Eq.\ 33$$

where $\bar{y}$ is mean of data.

Compute the updated estimate by using the smoothing equation as in Eq. 34.

$$l_T = \propto y_T + (1-\propto)l_{T-1} \qquad Eq.\ 34$$

where
$\propto$ = value between 0 and 1.
$y_T$= actual value,
$l_{T-1}$= level before estimate

Point forecast made at time $T$ for $y_{T+p}$ as in Eq. 35,

$$\hat{y}_{T+p}(T) = l_T \quad (p = 1,2,3, \dots) \qquad Eq.\ 35$$

### 2.3 Measurement Error

#### 2.3.1 Mean Square Error (RMSE)

The root mean square error (RMSE) is the square root of the mean of the square of all error. The formula of RMSE can be written as described in Eq. 36 [7].

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \qquad Eq.\ 36$$

where $y_t$ is the actual value of period $t$, $\hat{y}_t$ is the forecasted value of period $t$, $n$ is the total number of $n$ periods.

#### 2.3.2 Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE) is a measure of prediction accuracy of a forecasting method. The formula of MAPE can be written as described in Eq. 37 [7].

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \qquad Eq.\ 37$$

where $y_t$ is the actual value of period $t$, $\hat{y}_t$ is the forecasted value of period $t$, $n$ is the total number of $n$ periods.

## 3. Results and Discussion

Figure 1 showed the monthly Scombridae landings data from January 2007 to December 2017. The pattern of the graph showed the upward trend with seasonal pattern. The landings of Scombridae was lowest from January 2007 to January 2008. This was due to fishmonger lifestyle may not be sufficient as unability to catch the Mackerel.
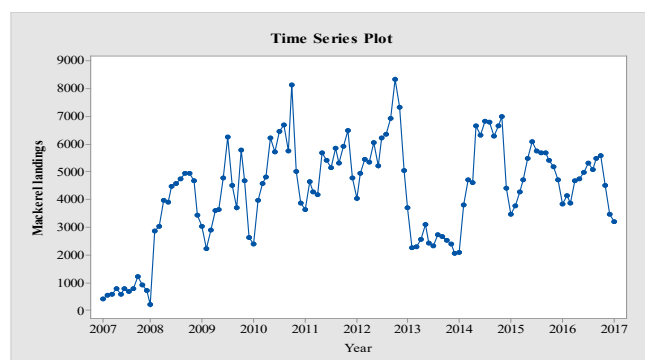


**Figure 1: Time series plot of Mackerel Landing in Malaysia**

Figure 2 showed the ACF plot before differencing, hence the data in ACF plot was slowly decreasing which mean that the data was not stationary. The data needed to be differencing as to obtain more accurate model to proceed the forecast of Mackerel landings. Meanwhile, Figure 3 showed that the data in PACF plot was cut off at lag 1.
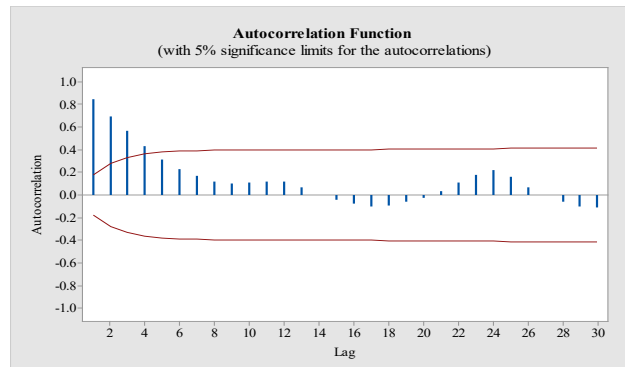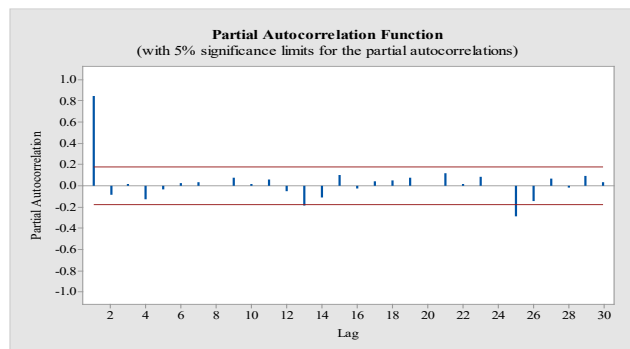


**Figure 2: ACF plot of Mackerel landing in Malaysia**



**Figure 3: PACF plot of Mackerel landing in Malaysia**

Figure 4 showed the lambda table contains an estimate of lambda (1.09) and the rounded value (1.00) which was the value used in the transformation. The lambda table also included the upper confidence interval (0.68) and lower confidence value (1.63) which were obtained by vertical lines.
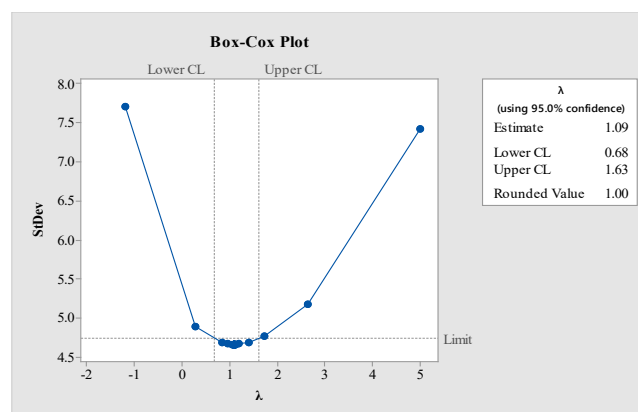


**Figure 4: Second Box-Cox transformation**

Figure 5, 6 and 7 below showed the time series plot of Mackerel Landings with first differencing followed by ACF and PACF plot of first differencing.
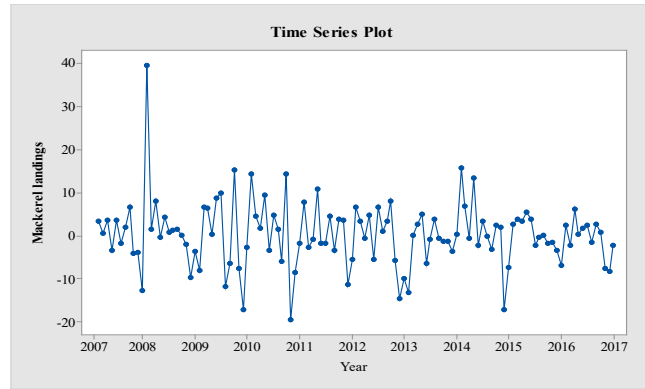
**Figure 5: Time series plot of Mackerel landing in Malaysia with first differencing**
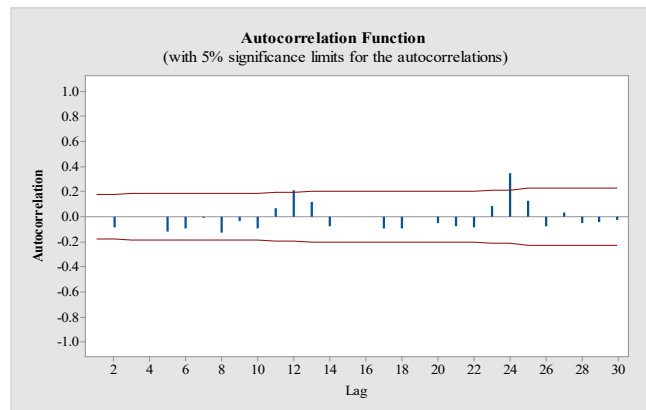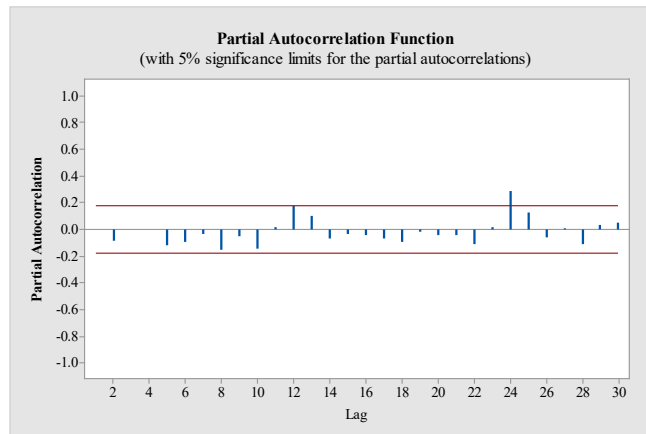


**Figure 6: ACF plot with first differencing**



**Figure 7: PACF plot with first differencing**

Table 2 showed that SARIMA $(0,1,0)(2,0,1)_{12}$ was the best model amongst others SARIMA model obtained as the value of AIC and BIC of this model was the lowest. The *p*-value of Ljung-Box in each lag was also larger than 0.05 and MSE was the lowest. Thus, this model will be used to make a prediction of Mackerel landings.

**Table 2 : Analysis of SARIMA (0,1,0)(2,0,1)**

|  | PACFs | **Model** |
|---|---|---|
| Cut off at lag 12 , SAR(1) | Cut off at lag 12, SMA(1) | $SARIMA(0,1,0)(1,0,2)_{12}$ |
| Cut off at lag 24 , SAR(2) | Cut off at lag 24, SMA(2) | $SARIMA(0,1,0)(2,0,1)_{12}$ |
|  |  | $SARIMA\ (0,1,0)(2,0,2)_{12}$ |

| Model | AIC | BIC |
|---|---|---|
| $SARIMA(0,1,0)(1,0,2)_{12}$ | 459.0968765 | 474.2468434 |
| $SARIMA(0,1,0)(2,0,1)_{12}$ | 430.9217182 | 446.0716852 |
| $SARIMA\ (0,1,0)(2,0,2)_{12}$ | 432.2695935 | 451.2070522 |

Table 3 showed the analysis of Multiplicative Holts Winter, Additive Holts Winter and Simple Exponential Smoothing method which showed the value of coefficient using optimization of Solver in Excel.

**Table 3: $\alpha, \beta, \gamma$ value by solver**

|  | MHWM | AHWM | SESM |
|---|---|---|---|
| $\alpha$ | 0.90 | 0.97 | 0.01 |
| $\beta$ | 0.02 | 0.02 | - |
| $\gamma$ | 0.98 | 0.21 | - |

Table 4 showed the summarization of all 4 methods that has been used to make a prediction of Scombridae landings in Malaysia. Then, as a result, Additive Holt-Winters Method was the best method as it have lowest value of Mean Absolute Error and Mean Square Error compared to other methods.

**Table 4: Summarization of all methods**

|  | $SARIMA$ $(0,1,0)(2,0,1)_{12}$ | Multiplicative Holt-Winters Method | Additive Holt-Winters Method | Simple Exponential Smoothing Method |
|---|---|---|---|---|
| MAPE | 0.41 | 0.12 | 0.11 | 0.13 |
| MSE | 2269323.26 | 266288.2 | 251965.1 | 308823.6 |

## 4. Conclusion

Additive Holt-Winters method was selected as the best model to predict of Mackerel landings in Malaysia. This method has the lowest value of MAPE and MSE and showed that this method has the best performance compared to the others method. Therefore, by conducting this time series analysis of Mackerel landings, all of the fishmongers in Peninsular Malaysia can make a well preparation to facilitate their work and also can increase the economics of Malaysia through import and exports of mackerel to the others countries. The information provided by the monthly retail price of Indian Mackerel was enough to help the fish farmers and the planners for future planning in Malaysia. Time series analysis of fish prices has been important tool for fisheries management and decision making.

## Acknowledgement

## References

[1]     A. J. Hobday, S. Griffiths and T. Ward, Pelagic fishes and sharks, A marine climate change impacts and adaptation report card for Australia, 2009, Online : www.oceanclimatechange. org.au.

[2]     C. P. Ean, The status of the Rastrelliger (Kembung) fishery on the west coast of Peninsular Malaysia. Norway Government Cooperative Programme, Supplement to the Report of a Workshop on the Fishery and Management of Short Mackerel (Rastrelliger spp.) on the West Coast of Peninsular Malaysia, Penang, Malaysia: Food and Agriculture Organization of the United Nations, 2000.

[3]     M. Pritchard, A. Gordon, G. Patterson and A. Gray, Utilization of Small Pelagic Fish Species in Asia. Chatham, United Kingdom: Natural Resources Institute, 1996.

[4]     J. Lloret, J. Lleonart and I. Sole, "Time series modelling of landings in Northwest Mediterranean  Sea," ICES Journal of Marine Science, vol. 57, no. 1, pp. 171-184, 2000.

[5]     H. H. Park, "Analysis and Prediction of Walleye Pollock (Theragra Chalcogramma) Landings in Korea by Time Series Analysis", Fisheries Research, vol. 38, no. 1, pp. 1-7, 1998.

[6]     C. Floros and P. Failler, "Forecasting Monthly Fisheries Prices: Method Comparison using Data from Cornwall (UK)," European Journal of Scientific Research, vol. 14, no. 2, pp. 613 - 624, 2006.

[7]     R. K. Paul, "Stochastic Modelling of Wholesale Price of Rohu in West Bengal, India," Interstat, vol. 11, pp. 1-9, 2010.

[8]     C. R. P. Da Veiga, A. Catapan, U. Tortato and W. V. Da Silva, "Demand forecasting in food retail: a comparison between the holt-winters and ARIMA models," WSEAS Transactions on on Business and Economics, vol. 11, no. 1, pp. 608-614, 2014.

[9]     C. Chatfield, "The holt-winters forecasting procedure," Applied Statistics, vol. 27, no. 1, pp. 264-279, 1978.

[10]    R. J. Hyndman and A. B. Koehler, " Another look at measures of forecast accuracy," International Journal of Forecasting, vol. 22, no. 4, pp. 679-688, 2006.