

Forecasting Modelling of Cockles in Malaysia by Using Time Series Analysis

Hani Nabihah Aziz¹, Mohd Saifullah Rusiman^{1*}, Siti Noor Asyikin Mohd Razali¹, Abdul Wahab Abdullah², Nur Amira Azmi¹

¹Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Educational Hub, 84600 Pagoh, Muar, Johor, Malaysia

²Fisheries Research Institute (FRI) Kampung Acheh, 32000 Sitiawan, Perak, Malaysia

*Corresponding author E-mail: saifulah@uthm.edu.my

Abstract

Cockle farmed in Malaysia are from *Anadara* genes and *Arcidae* family which known as blood cockle. Normally, it was found in the farmed around mangrove estuary areas in the muddy and sandy shores. This study aims to predict the production of cockle to ensure sure the cockle supplies are synchronised with the demand. Then, based on the demand, the prediction result could be used to make decision either to import or export the cockle. The data were taken from the Department of Fisheries Malaysia (DFM) and it has cyclic pattern data. There are two methods used in this study which are Holt-Linear method and Auto regressive moving average (ARMA). In determining the best fitted model between the two methods, the mean square error (MSE) values will be compared and the lowest value of MSE will assign as the best model. Result shows that ARMA(1,1) is the best model compared to Holt-Linear. Therefore, ARMA(1,1) model will be used to forecast the production of cockle in Malaysia.

Keywords: Auto regressive moving average (ARMA); Holt-Linear; Mean Square Error (MSE)

1. Introduction

The cockles farmed in Malaysia and Thailand are bivalves from genes *Anadara*, family *Arcidae*. The blood cockle or the scientific name *Anadara granosa* are not original cockles but as arc cockles [1]. Cockles have a medium sized clam-liked shell round with radiating ridges. The characteristics of the shell equivalve are thick, solid, ovate, strongly inflated and slightly longer than high [2]. The range of yearly capture production that reported by FAO's Yearbook of Fishery Statistics in 1999 for China, Taiwan and Korea Republic are around 1415mt in 1995 to 6503mt in 1999. Meanwhile, the ranges of yearly aquaculture production for China, Korea Republic, Thailand and Malaysia in 1999 are around 252233mt in 1995 to 315811mt.

In 1973, Malaysia exported seed for culture in Satun Province which is a part of Southern Thailand. The state of Perak started culture for *A.granosa*. It has been developed to be the most important and organise agriculture to Malaysia. *A.granosa* are naturally distributed from Kedah to Johor but the seed also been found in shelters coastal areas in Pahang [3]. The suitable ammonia level for the cockle to bred should be less than 0.25 part per million. Adequate food supply and stoking density will help in cockle growth. In a population, the mean wet weight of *A.granosa* in a sample was negatively correlated with density [4]. In future, the cockle culture will be endangered by human activities since the breeding grounds of the cockle are treated by pollution. The disposals of cockle shell also increased due to cockle trade that continuously growing. The cockle trade also produces a generation of abundant waste shell. There are quite considerable studies were carried out to study the time series in the world [5, 6, 7, 8, 9].

The main purpose of this study is to predict the production of cockle so that the Department of Fisheries Malaysia (DFM) can come out with a plan to increase the production of cockle if the cockle supply is less than current demand or to export the cockle if the stocks of cockle are more than current demand. Here, the prediction of cockle production could be used to solve the cockle problem.

ARMA are the combination of autoregressive and moving average. The general ARMA model was described in the 1951 thesis of Peter Whittle, hypothesis testing in time series analysis, and it was popularized in the 1970 by Box and Jenkins [10, 11]. The difference between ARMA and ARIMA (autoregressive integrated moving average) is the ARMA model used when the data are found as stationary. When the data are not stationary, the differencing is needed to do before ARIMA model can be used. The comparison of models aims to find which model is better in predicting the production of cockles. There are two forecasting methods will be used in this study which are auto regressive moving average (ARMA) model and Holt-Linear method. The best fitted model with lowest MSE will be used to predict the future production of cockle for 2014 onward. The output of the model is generated by using Minitab and Microsoft Excel.

2. Material and Method

The data are collected from Department of Fisheries Malaysia (DFM). There are 108 samples of data based on monthly production of cockle from 2005 until 2013 with no missing data.

2.1. ARMA model

Firstly, the aim of this study is to determine the stationary of the data. A property of stationary of the data is, it is not depending on the time when the series are observed. In general, a stationary time series will have no predictable pattern in the long term. The stationary data will be determined by referring to the ACF plot. If the data are slowly lies down, it can be said that the data are stationary [10]. Secondly is plotting a Box-Cox plot to check whether the data have constant variance. A constant variance means that the value of λ is equal to 1. If the value of λ does not include the value of 1, then the variance is not stable and an appropriated transformation to dependent variable should be applied.

The third step is to estimate appropriate values for p and q where p represents the autoregressive model and q refer to moving average model. ARMA are the combination of autoregressive and moving average. If autocorrelation function (ACF) shows sinusoidal decay and partial autocorrelation function become zero after lag p , it is pure AR process. If ACF become zero after lag q , PACF has sinusoidal decay and it will has MA process. If both have sinusoidal and become zero after lag p and q respectively it become ARMA [10].

The fourth step is diagnostic checking to look whether the model used is fitted to data. Inadequacies of the model are checked by considering the autocorrelations of the residual series to obtain the fitted model. ACF plot of residuals will be apply for diagnostic checking. If the entire plots are lies below the 5% of significant level, the model that has been used is valid.

The fifth step aim to calculate the prediction values from 2003 to 2013 and to calculate the MSE value by using the formula as shown in (1). Then, the prediction value for 2014 onward by using the same formula.

$$\text{AR}(p) \text{ model : } X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t$$

$$\text{MA}(q) \text{ model : } X_t = \mu + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

$$\text{ARMA}(p, q) \text{ model : } X_t = c + \varepsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (1)$$

where ϕ_i is coefficient parameters, θ_i is coefficient parameters, c is constant, μ is expectation of X_t , ε_t is white noise error term and ε_{t-i} is white noise error terms [11].

Accuracy of the models was determined by mean square error (MSE) as in (2). Forecasting error measurements were based on the difference between the model and the observed values [11].

$$\text{MSE} = \frac{\text{SSE}}{n}, = \sum_{i=1}^{12} (y_i - \hat{y}_i)^2 \quad (2)$$

where SSE is sum square of error, y_i is actual value, \hat{y}_i is predicted value from model and n is the number of data.

2.2. Holt-Linear method

Holt-Linear is extended from simple exponential smoothing method. Holt-Winters method is used because the data series contains both trend and seasonal variation. Besides that, this method is effective in forecasting trend time series pattern. The smoothing parameter selected will give a different forecast. In this study, the fitted value is calculated, then it will be used with the actual value to obtain the MSE value. The formula used for this method are in (3).

$$\begin{aligned} \hat{y}_{t+h|t} &= \ell_t + h b_t \\ \ell_t &= \alpha y_t + (1-\alpha) (\ell_{t-1} + b_{t-1}) \\ b_t &= \beta * (\ell_t - \ell_{t-1}) + (1-\beta *) b_{t-1} \end{aligned} \quad (3)$$

where ℓ_t is the level at period t , b_t is trend at period t , α is level smoothing coefficient, β is trend smoothing coefficient, y_t is actual value at period t and $\hat{y}_{t+h|t}$ is forecast value of a period ahead [12].

3. Results and Discussion

3.1. ARMA model

Figure 1 shows the monthly production of cockle from January 2005 to December 2013. A cyclic trend exist when the data exhibit rises and falls that are not fixed period. The highest production of cockle is in October each year. Refer to the Figure 2, since the value of λ is 0.0 which indicates that the variance is not stable, \ln will be used to transform the data.

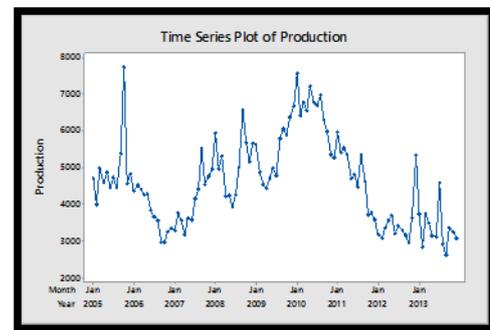


Fig. 1: Time series plot of cockle production in Malaysia

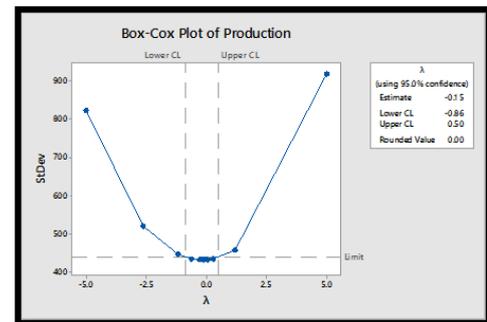


Fig. 2: Box cox plot for production

After the transformation of data completed, the ACF plot need to plot again to ensure whether the data are stationary. The ACF plot in Figure 3 shows that the plot is lies down exponentially fast, which confirmed that the series is stationary. Since the data is stationary, therefore the differencing process should not apply. Meanwhile, the PACF plot in Figure 4 shows that there is spike at lag one, so it will be AR(1). To ensure the best fitted model for Box-Jenkins method, trial and error methods are apply.

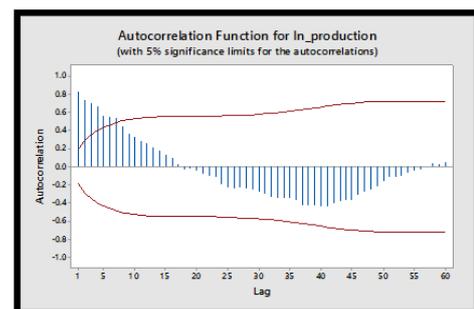


Fig. 3: ACF plot for $\ln_{\text{production}}$ of cockle

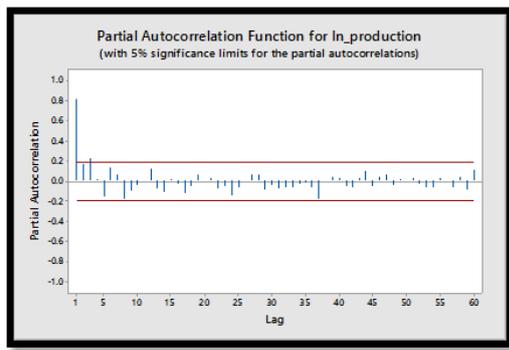


Fig. 4: PACF plot for $ln_production$ of cockle

In selecting the appropriated model, all the models need to fit a condition with original p -value must lowest than 0.05 and for p -value (Ljung-box) must be exceed 0.05. Table 1 shows the value of MSE, p -value (original) and p -value for Ljung-Box of both models. There are total 15 models that have been used in this trial and error model but only one model that follow all the conditions and have lowest value of MSE. The model that appropriate for this study is ARMA (1,1).

Table 1: Box-Jenkins model using Trial and Error

Model	MSE	p -value	p -value (Ljung-Box)			
AR(1,0)	0.02176	0.000	0.002	0.005	0.031	0.052
AR(1,1)	0.01954	0.000	0.071	0.144	0.194	0.222

3.2. Holt-Linear method

Holt-Linear is extended from simple exponential smoothing method and quite effective in forecasting trend. In this study, the smoothing parameter values which are α and β will be obtained from solver in Microsoft Excel. For this reasearch, the optimal value for α and β are 0.67 and 0.49 respectively. Lowest MSE will indicated that the predicted value is close to the actual value. The model is shown as in (4).

$$\begin{aligned} \hat{y}_{t+h|t} &= \ell_t + h b_t \\ \ell_t &= 0.67 y_t + 0.33 (\ell_{t-1} + b_{t-1}) \\ b_t &= 0.49 (\ell_t - \ell_{t-1}) + 0.51 b_{t-1} \end{aligned} \quad (4)$$

Based on Table 2, it shows that the MSE value for both models that have huge difference. The better model with the lowest value of MSE will be used to predict the production of cockle in Malaysia for 2014 onward. Table 2 clearly present that ARMA have the lowest value of MSE and therefore ARMA (1,1) will be adopt for the predictive modelling of cockle.

Table 2: MSE value of ARIMA and Holt-Linear Model

Time Series Model	MSE
ARMA model	0.01954
Holt-Linear model	93.85221

Table 3 shows the forecasting of production of cockle from January 2014 to December 2015 by using ARMA (1,1) model. Here, the production of cockle for two years period are decreasing and it has more than 3000 tan of cockle per month along the two years period.

Table 3: Forecasting of production of cockle from January 2014 to December 2015

January 2014	3152.90
February 2014	3151.48
March 2014	3150.06
April 2014	3148.63
Mei 2014	3147.21
June 2014	3145.79
July 2014	3144.37
August 2014	3142.95

September 2014	3141.53
October 2014	3140.11
November 2014	3138.69
December 2014	3137.28
January 2015	3135.86
February 2015	3134.45
March 2015	3133.03
April 2015	3131.62
Mei 2015	3130.20
June 2015	3128.79
July 2015	3127.38
August 2015	3125.97
September 2015	3124.56
October 2015	3123.15
November 2015	3121.74
December 2015	3120.33

4. Conclusion

Based on the data production of cockle, ARMA(1,1) model are the preferable model because it suitable for forecasting any patterns of cockle if compared to the Holt-Linear model. However, the forecast values less accurate since the data only be based on the previous data and future study should consider more accurate and the latest dataset. ARMA(1,1) model become the better model by comparing the value of MSE which are 0.01954 for ARMA(1,1) and 93.85221 for Holt-Linear. Therefore, the prediction on production of cockle in Malaysia by using ARMA(1,1) is closer to the actual value. The purpose of predicting the production of cockle is to help certain authorities such as Department of Fisheries Malaysia (DFM) to handle the crisis of less stock of cockle nowadays where the certain actions are needed to increase the stock of cockle to meet the local and international demand.

Acknowledgement

The authors would like to express gratefully heartfelt thanks to the Universiti Tun Hussein Onn Malaysia and Office for Research, Innovation, Commercialization and Consultancy Management (ORICC) for the financial support under the TIER 1 research grant (H232).

References

- [1] Tookwinas S, Commercial Cockle Farming in Southern Thailand, University of California, International Center for Living Aquatic Resources Management (1985).
- [2] Poutiers JM, Bivalves (Acephala, Lamellibranchia: Pellecypoda) In: The living Marine Resources of the Western Central Pacific, Volume 1, Seaweeds, corals, bivalves and gastropods (Carpentier, K. E., Niemi, V. H. eds), FAO, Rome, Italy, (1998).
- [3] Ng FO, "Cockle Culture SAFIS Extension Manual", Southeast Asian Fisheries Development Center, Bangkok 13 (1998), pp. 22.
- [4] Broom MJ, The Biology and Culture of Marine Bivalve Molluscs of the Genus Anadara, ICLARM Studies and Reviews 12, International Center for Living Aquatic Resources Management, Manila, (1985), pp. 37.
- [5] Bako HY, Rusiman MS, Kane IL & Matias-Peralta HM (2013), Predictive modeling of pelagic fish catch in Malaysia using seasonal ARIMA models, Agriculture Forestry and Fisheries, Vol.2, No.3, pp. 136-140.
- [6] Lee MH, Rahman NHA, Suhartono, Latif MT, Nor ME, Kamisan NAB, Seasonal ARIMA for forecasting air pollution index: A case study, American Journal of Applied Sciences, Vol.9, No.4 (2012), pp. 570-578.
- [7] Nor ME, Safuan HM, Shab NFM, Asrul M, Abdullah A, Mohamad NAI, Lee, MH, Neural network versus classical time series forecasting models, AIP Conference Proceedings, Vol.1842, (2017), pp. 030027.
- [8] Khalid K, Mohamed I & Abdullah NA, An Additive Outlier Detection Procedure in Random Coefficient Autoregressive Models AIP Conference Proceedings 1682 050017, (2015).

- [9] Mohamed I, Khalid K & Yahya MS (2016), Combined Estimating Function for Random Coefficient Models with Correlated Errors Communications In Statistics—Theory and Methods, Vol.45, No.4, (2016), pp. 967-975.
- [10] Hyndman RJ & Athanasopoulos, Forecasting: Principle and practice, Otexts, (2014).
- [11] Guo YH, Shi XP & Zhang XD, A study of short term forecasting of the railway freight volume in China using ARIMA and Holt-Winters models, In 2010 8th International Conference on Supply Chain Management and Information, (2010), pp.1-6.
- [12] Veiga CPD, Veiga CRPD, Catapan A, Tortato U & Silva WVD, Demand Forecasting in Food Retail: A Comparison Between the Holt-Winters and ARIMA Models, WSEAS Transactions on Business and Economics, Vol.11 (2014), pp.608-614.