Malaysian Journal of Mathematical Sciences 2(2): 41-54 (2008)

Arima and Integrated Arfima Models for Forecasting Annual Demersal and Pelagic Marine Fish Production in Malaysia

[1]Mahendran Shitan, [2]Pauline Mah Jin Wee, [3]Lim Ying Chin [4]and Lim Ying Siew

[1]Department of Mathematics,
Faculty of Science, Universiti Putra Malaysia
and
Statistics and Applied Mathematics Laboratory,
Institute for Mathematical Research,
Universiti Putra Malaysia
[2,3,4] International Education Centre (INTEC),
UiTM Section 17 Campus, Shah Alam

ABSTRACT

The seas surrounding Malaysia provide a rich source of marine fisheries. The fisheries industry is an important economic earner and the total marine fish production has increased drastically from 242,900 metric tonnes in 1970 to around a million metric tonnes in the year 2000. Since fisheries resources are renewable, proper management issues should be taken to manage these fisheries resources. From the management point of view, fish forecasting is a very important tool for fisheries managers and scientists to enable them to decide on sustainable management issues. Time series models have been used to forecast various phenomena in many fields. In a previous research by Mahendran Shitan et. al. (2004), the maximum likelihood and bootstrap method were used to forecast the total Malaysian marine fish production. Marine fish can be sub-classified as demersal marine fish and pelagic marine fish and it would be interesting to forecast the individual composition of these categories. Therefore, in this research we fit time series models to forecast the demersal and pelagic marine fish production using ARIMA and integrated ARFIMA models and make predictions of each category. Our results indicate that the ARIMA models appear to be the better models and the forecasted amounts for the year 2011 are approximately 373,370 and 666,460 metric tonnes for the demersal and pelagic marine fish, respectively.

INTRODUCTION

Malaysia is surrounded by the Straits of Malacca, the Straits of Johore, the Sulu Sea and the South China Sea which is an extensive fishing ground. In 1984, Malaysia declared its Exclusive Economic Zone (EEZ) up to 200 nautical miles (Mohd. Mazlan, 2000). With such a large marine habitat, the fisheries sector in Malaysia plays a significant role in supporting the country's economic growth through provision of employment and providing source of much needed protein to the population. Over recent years, the per capita fish consumption has risen from 39.1 kg in 1995 to 49.0 kg in 2000 (Annual Fisheries Statistic, 2001) and the projection for per capita consumption in the year 2010 is at 56 kg (SERI, 2002). This figure is relatively high as compared to the average world consumption of fish per person of between 16 kg to 20 kg.

In 2003, the total fish production amounts to 1,483,958 tonnes valued at RM5.22 billion (US$ 1.36 billion). This contributed to about 1.37% of Gross Domestic Product (GDP) and provided direct employment to 89,433 fishers and 21,114 fish aquaculturists (Annual Fisheries Statistic, 2003). The Malaysian fisheries sector is divided into capture fisheries (marine and inland) and aquaculture. The marine capture fisheries cover a total area of 547,200 km^2 and categorized into coastal fisheries and deep-sea fisheries. In 2003, the coastal fisheries and deep-sea fisheries contributed about 1,084,802 tonnes (73.1%) and 198,453 tonnes (13.4%) respectively, to the total marine landings. There is still potential for further development in the deep-sea fisheries which is projected to contribute 430,000 tonnes in 2010 (Department of Fisheries Malaysia, 2005).

The marine capture fisheries can be further classified into demersal fisheries and pelagic fisheries. The demersal fish refers to the fishes that live, sink or lie near the bottom of the sea floor and feeding on benthic organisms. The important demersal fisheries are crimson jobfish (*Pristipomoides filamentosus*), humpback red snapper (*Lutjanus gibbus*), malabar blood snapper (*Lutjanus malabaricus*), trout sweetlip (*Plectorhinchus pictus*) and pickhandle barracuda (*Sphyraena jello*). The main method of catching demersal fish is the trawls while the other methods are fish traps and push net.

Pelagic fish refers to those fish that spend most of their life swimming in the water column (seas, oceans or open waters which associated with the surface or middle depths of a water body) with little contact with or dependency on the bottom of the sea floor. Many pelagic fish feed on plankton. The important pelagic fish are short mackerel (*Rastrelliger brachysoma*), common dolphinfish (*Coryphaena hippurus*), kawakawa (*Euthynnus affinis*), black marlin (*Makaira indica*), frigate tuna (*Auxis thazard thazard*) and Indo-Pacific sailfish (*Istiophorus platypterus*) (Bachok et al, 2004). The purse seine is the principal method of catching pelagic fish. Other types of nets employed are the drift gill net, fishing stakes and lift net.

Time series models have been used to forecast various phenomena in many fields like environmental, agriculture, economics, tourism, meteorology, etc including fisheries. Previous research on fisheries has been done to show and describe the importance of the application of fish forecasting in fisheries management. Stergiou et al (1996 and 1997) has shown the comparison between various regression, univariate, multivariate time series techniques to model and forecast the monthly and annual fisheries catches. Hae-Hoon Park (1998) analyzed and predicted fisheries landings in Korea. Pierce and Boyle (2003) also used time series models to model interannual variation squid abundance in Scottish waters. In Malaysia, Mahendran Shitan et al (2004) used an ARIMA (0, 1, 1) model and bootstrap estimation to forecast the annual total marine fish production.

Since marine fish can be sub-classified as demersal marine fish and pelagic marine fish, it would be interesting to forecast the individual composition of these categories as different fishing methods are employed for the categories. Therefore, in this research we consider fitting time series models, namely the ARIMA and integrated ARFIMA models.

The data set and the methodology of this research are discussed in Section 2 followed by the results in Section 3. The conclusions are contained in Section 4.

METHODOLOGY

The Data Set

The data used in this study was obtained from the Food and Agriculture Organization of the United Nations (FAO) website for the fisheries. In particular, we concentrated on the annual total marine fish production in Malaysia from year 1961 to 2001. The production (catch) of the total marine fish refers to the weight of the fish at the time of removal from the water, or in other words, live weight in metric tonnes. The time series plots of the annual demersal and pelagic marine fish in Malaysia are shown in Figures 1 and 2 respectively.

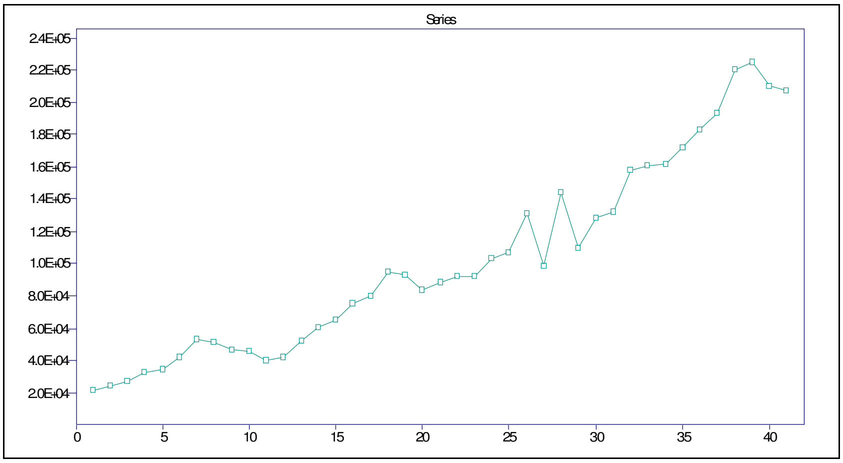

Figure 1: Plot of the annual demersal marine fish production in Malaysia from year 1961 to 2001

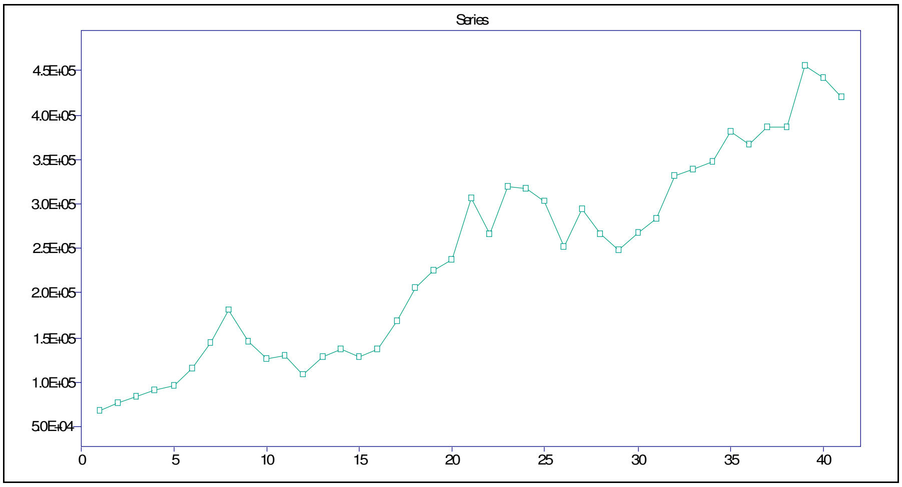

Figure 2 Plot of the annual pelagic marine fish production in Malaysia from year 1961 to 2001

The plots seem to have an approximately linear increase. The ARIMA models for each category was found by first transforming the series using the Box-Cox transformation with parameter, $\lambda = 0$. Then the transformed series was differenced at lag 1. After the mean of the series was subtracted, a model with the lowest AICC was found using the autofit menu of ITSM

2000. To fit the integrated ARFIMA model, we used the same procedure but we specified the model as fractionally integrated model before using the autofit menu to fit the model.

Time Series Modelling Procedure

For the purpose of time series modelling in this study, the first 31 observations (from year 1961 to 1991) were used to fit the ARIMA and integrated ARFIMA models while the subsequent 10 observations (from year 1992 to 2001) were kept to check the accuracy of the post sample forecast. A brief description of the time series models and definitions used in this study are as follows.

A stationary ARMA (*p*, *q*) model is defined as a sequence of random variables $\{X_t\}$, given by

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

where $\{Z_t\}$ is a sequence of uncorrelated random variables with zero mean and constant variance, denoted as $\{Z_t\} \sim WN(0,\ \sigma^2)$, (Brockwell and Davis, 2002). A process $\{X_t\}$ is called an ARIMA (*p*, *d*, *q*) process (Brockwell and Davis, 2002) if *d* is a nonnegative integer such that $(1-B)^d X_t$ is a causal ARMA (*p*, *q*) process. The ARIMA (*p*, *d*, *q*) processes satisfies the difference equation of the form $\phi^*(B) \equiv X_t \phi(B)(1-B)^d X_t = \theta(B)Z_t$, $\{Z_t\} \sim WN(0,\ \sigma^2)$, where $\phi(z)$ and $\theta(z)$ are polynomials of degrees *p* and *q* respectively, and $\phi(z) \neq 0$ for $|\,z| \leq 1$. The $\phi^*(z)$ has a zero of order *d* at $z = 1$. The process $\{X_t\}$ is stationary if and only if $d = 0$, in which case it reduces to an ARMA (*p*, *q*) process.

A long memory process (v or a fractionally integrated ARMA, ARFIMA (*p*, *d*, *q*) processes with $0 < |\,d| < 0.5$ is a stationary process with much more slowly decreasing autocorrelation function $\rho(k)$ at lag *k* as $k \to \infty$ which satisfies the property of $\rho(k) \sim Ck^{2d-1}$. The ARFIMA processes satisfy the difference equation of $(1-B)^d \phi(B)X_t = \theta(B)Z_t$, where

$\{Z_t\} \sim WN(0, \sigma^2)$, $\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$ satisfying $\phi(z) \neq 0$ and $\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q$, satisfying $\theta(z) \neq 0$ for all z such that $|z| \leq 1$, and B is the backward shift operator. The operator $(1-B)^d$ is defined by the binomial expansion of $(1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j$ with $n_0 = 1$ and $\pi_j = \prod_{0<k\leq j} \frac{k-1-d}{k}$ for $j = 0, 1, 2, ...$.

The process of model fitting was done using the computer software "ITSM 2000, version 7.0", (Brockwell et al, 2002). The criteria chosen to measure the accuracy of the forecast in this study are the mean absolute error (MAE), the root mean square error (RMSE) and the mean absolute percentage error (MAPE) are given below.

$$\text{MAE} = \frac{\sum_{i=1}^{n} |x_i - \hat{x}_i|}{n}, \quad \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{n}},$$

$$\text{MAPE} = \frac{\sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right|}{n} \times 100\%$$

where x_i and $\hat{x}_i$ are the actual observed values and the predicted values respectively while n is the number of predicted values.

RESULTS

The results are discussed in this section.

Demersal Marine Fish Production

The annual forecast from 1992 to 2001 of the demersal marine fish production using ARIMA (1, 1, 0) model and integrated ARFIMA (0, –0.4154, 2) model are shown in Table 1 and Table 2 respectively. The graphs of the predicted values and the actual values together with their 95% forecast intervals are shown in Figure 3 and Figure 4.

TABLE 1: The annual forecast of the demersal marine fish production in Malaysia from 1992 to 2001 using the ARIMA (1, 1, 0) model

Year	Actual Production	Forecast	95% Forecast Interval
1992	158104	142130	(109500, 184470)
1993	160768	150500	(110450, 205060)
1994	161652	160220	(110820, 231650)
1995	171786	170240	(112540, 257540)
1996	182626	181020	(114660, 285770)
1997	192840	192420	(117300, 315650)
1998	219905	204570	(120320, 347810)
1999	225296	217470	(123700, 382340)
2000	210428	231190	(127410, 419510)
2001	206886	245780	(131440, 459570)

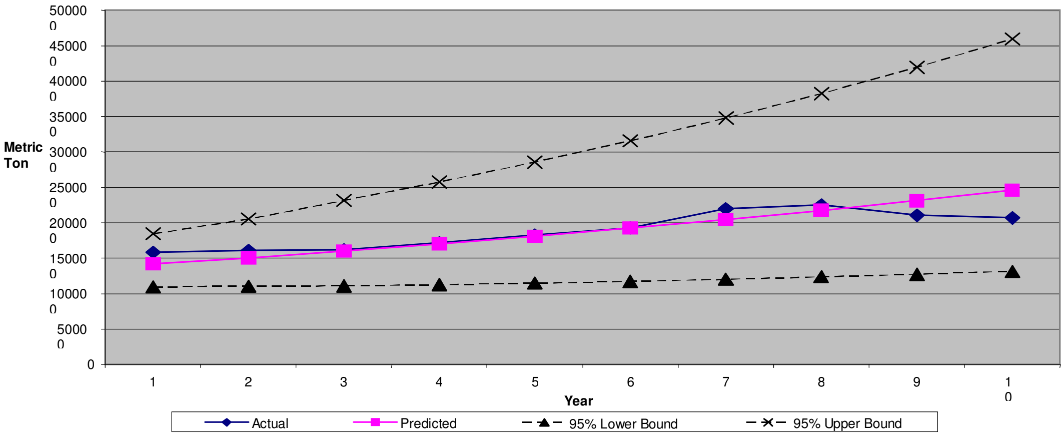

Figure 3: The annual demersal marine fish production values with 10 predicted values of the ARIMA (1, 1, 0) model and the actual values from 1992 to 2001

TABLE 2: The annual forecast of the demersal marine fish production in Malaysia from 1992 to 2001 using the integrated ARFIMA (0, –0.4154, 2) model

Year	Actual Production	Forecast	95% Forecast Interval
1992	158104	141210	(106630, 187010)
1993	160768	155340	(112160, 215140)
1994	161652	167770	(110160, 255520)
1995	171786	180230	(113510, 286170)
1996	182626	193070	(118220, 315290)
1997	192840	206460	(123750, 344470)
1998	219905	220530	(129900, 374390)
1999	225296	235360	(136620, 405460)
2000	210428	251030	(143880, 437980)
2001	206886	267620	(151690, 472160)

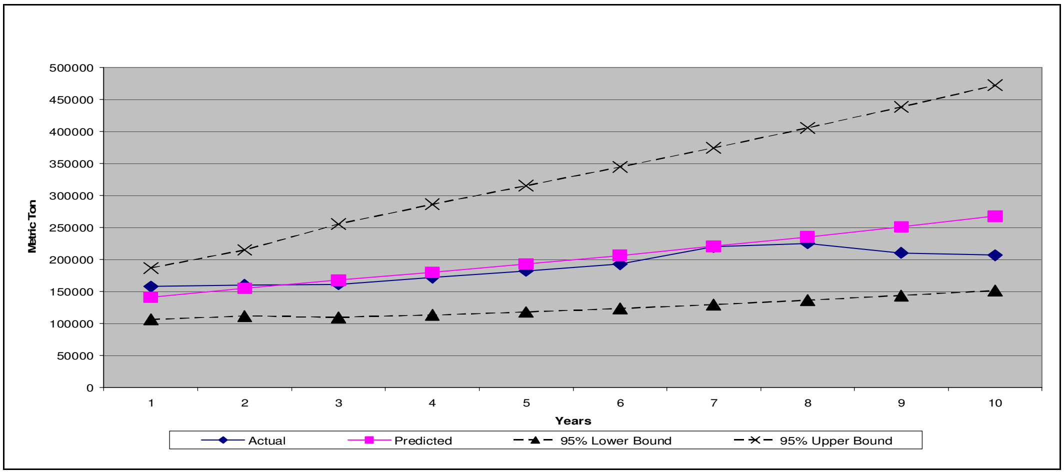

Figure 4: The annual demersal marine fish production values with 10 predicted values of the integrated ARFIMA (0, –0.4154, 2) model and the actual values from 1992 to 2001

The ARIMA model is given as $X_t = -0.3617 X_{t-1} + Z_t$ where $\{Z_t\} \sim WN(0, 0.1770)$ and the integrated ARFIMA model is given as $(1-B)^{-0.4154} X_t = Z_t + 0.002180 Z_{t-1} + 0.4837 Z_{t-2}$ where $\{Z_t\} \sim WN(0, 0.012557)$.

From Table 3, the MAE, MAPE and RMSE values of the ARIMA (1, 1, 0) model are lower compared to those of the integrated ARFIMA (0, –0.4154, 2) model. Therefore, ARIMA (1, 1, 0) model appears to be a better model.

TABLE 3: The MAE, MAPE and RMSE values of the ARIMA (1, 1, 0) model and the integrated ARFIMA (0, –0.4154, 2) model

Model	MAE	RMSE	MAPE
ARIMA (1, 1, 0)	9505.25	14742.26	4.87%
ARFIMA (0, –0.4154, 2)	14414.42	22651.34	7.41%

Pelagic Marine Fish Production

The annual forecast from 1992 to 2001 of the pelagic marine fish production using ARIMA (1, 1, 0) model and integrated ARFIMA (2, –0.5, 2) model are shown in Table 4 and Table 5 respectively. The graphs of the predicted values and the actual values together with their 95% forecast intervals are shown in Figure 5 and Figure 6. The ARIMA model is given as $X_t = Z_t - 0.1649 Z_{t-1}$ where $\{Z_t\} \sim WN(0, 0.29290)$,
while the integrated ARFIMA model is given as $(1-B)^{-0.5}(X_t - 1.817 X_{t-1} + 0.9791 X_{t-2}) = Z_t - 1.877 Z_{t-1} + Z_{t-2}$
where $\{Z_t\} \sim WN(0, 0.029217)$.

TABLE 4: The annual forecast of the pelagic marine fish production in Malaysia from 1992 to 2001 using the ARIMA (1, 1, 0) model

Year	Actual Production	Forecast	95% Forecast Interval
1992	331572	298210	(228980, 388360)
1993	339498	312950	(214870, 455810)
1994	348500	328420	(207000, 521070)
1995	381270	344660	(202150, 589650)
1996	367511	361700	(199110, 657050)
1997	386712	379580	(197330, 730180)
1998	386589	398350	(196460, 807700)
1999	456880	418040	(196320, 890180)
2000	442066	438710	(196760, 978170)
2001	421127	460400	(197700, 1072200)

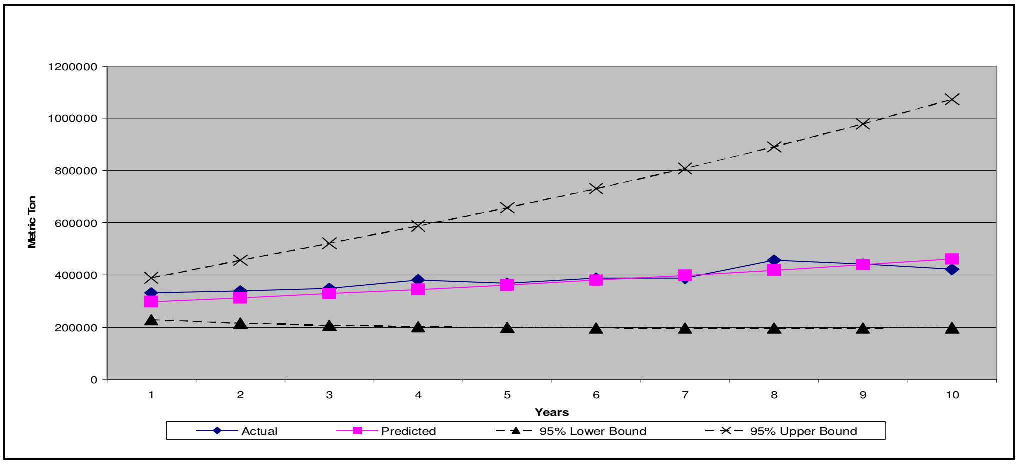

Figure 5:The annual pelagic marine fish production values with 10 predicted values of the ARIMA (1, 1, 0) model and the actual values from 1992 to 2001

TABLE 5: The annual forecast of the pelagic marine fish production in Malaysia from 1992 to 2001 using the ARFIMA (2, –0.5, 2) model

Year	Actual Production	Forecast	95% Forecast Interval
1992	331572	343440	(263700, 447280)
1993	339498	398360	(298520, 531590)
1994	348500	452200	(336250, 608140)
1995	381270	499220	(370340, 672940)
1996	367511	533310	(395410, 719300)
1997	386712	551140	(408580, 743440)
1998	386589	553690	(410430, 746940)
1999	456880	545740	(404420, 736450)
2000	442066	534090	(395370, 721480)
2001	421127	525520	(388050, 711690)

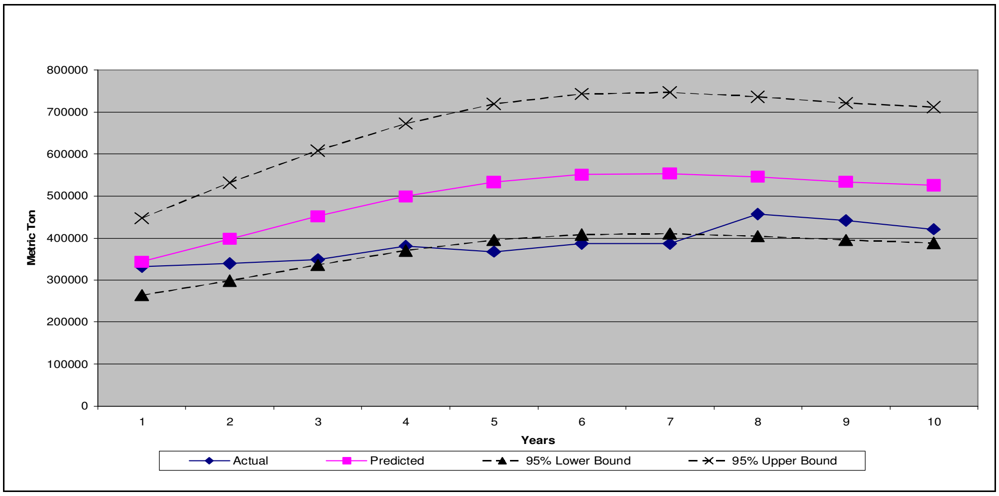

Figure 6: The annual demersal marine fish production values with 10 predicted values of the integrated ARFIMA (2, –0.5, 2) model and the actual values from 1992 to 2001

From Table 6, the MAE, MAPE and RMSE values of the ARIMA (1, 1, 0) model are lower compared to those of the integrated ARFIMA (0, –0.5, 2) model. As such, the ARIMA (1, 1, 0) model appears to be a better model.

TABLE 6: The MAE, MAPE and RMSE values of ARIMA (1, 1, 0) model and integrated ARFIMA (2, –0.5, 2) model

Model	MAE	RMSE	MAPE
ARIMA (1, 1, 0)	18564.42	23885.93	4.86%
ARFIMA (2, –0.5, 2)	89582.08	107231.26	23.13%

Forecasting with Fitted Model

The forecasted amounts of the demersal and pelagic marine fish for 2005 to 2011 are shown in Table 7.

TABLE 7 The annual forecast of the demersal and pelagic marine fish in Malaysia from 2005 to 2011

Year	Demersal Marine Fish	Pelagic Marine Fish
2005	265020	505630
2006	280710	529440
2007	297150	554380
2008	314620	580500
2009	333090	607850
2010	352650	636480
2011	373370	666460

Based on our models, the prediction of the demersal and pelagic marine fish amounts for the year 2011 are approximately 373,370 and 666,460 metric tonnes respectively.

CONCLUSIONS

In this study we have fitted time series models and for the demersal marine fish production, the ARIMA model appears to have a *slightly* better forecasting performance compared to that of the ARFIMA model. However, for the pelagic marine fish production, the ARIMA model is clearly the better forecasting model, in the sense that it has a much smaller MAPE, when compared with the ARFIMA model. This may be possibly due to the fact ARFIMA model is based on a long memory process and hence demersal marine fish are less subjected to variations caused by anthropogenic activities and other processes that may occur at the surface.

This research can be extended by considering multiple time series models and also by incorporating other variables such as salinity, temperature, food availability, nutrients etc. that would affect the marine habitats.

REFERENCES

1982 - 2003. *Annual Fisheries Statistics. 1982 - 2003*. Vol. 1. Department of Fisheries Malaysia. Ministry of Agriculture and Agro-based Industry. Kuala Lumpur.

2002. Economic Briefing To The Penang State Government, Vol. 4, Issue 10, Socio-Economic & Environmental Research Institute (SERI).

Department of Fisheries Malaysia, 2005. *Malaysia National Plan Of Action For The Conservation And Management Of Shark.* Kuala Lumpur: Department of Fisheries Malaysia, Ministry of Agriculture and Agro-Based Industry Malaysia.

Bachok Z., M.I. Mansor and R.M. Noordin, 2004. Diet composition and food habits of demersal and pelagic marine fishes from Terengganu waters, east coast of Peninsular Malaysia. *NAGA, WorldFish Center Quarterly,* **27**(3 & 4), p 41-47.

Brockwell, P. J. and R.A. Davis.2002. *Introduction To Time Series And Forecasting,* 2nd Edition, Springer-Verlag, New York.

Food and Agricultural Organization of the United Nations (FAO): FAOSTAT Database Collections, Fisheries Data, http://faostat.fao.org/faostat/collections?subset+fisheries. Accessed on 15 February 2005.

Mahendran Shitan, Chin Hong Yee, R.P.P.K Jayashinghe 2004. An ARIMA (0, 1, 1) Model and Bootstrap Estimation for Forecasting the Annual Total Marine Fish Production in Malaysia. *Proceedings of 7th Asian Fisheries Forum held on 30 Nov – 4 Dec, 2004 in Penang*.

Mohd. Mazlan, J. 2000. Management of living aquatic resources in the Straits of Malacca in *Towards Sustainable Management of the Straits of Malacca*. M. Shariff, N. Gopinath, H.M. Ibrahim and R. A. Nik Mustapha (eds), p. 15-24. Malacca Straits Development Centre (MASDEC), Universiti Putra Malaysia, Serdang Malaysia.

Park, H. 1998. Analysis and prediction of walleye polloc (*Theragra chalcogramma*) landings in Korea by time series analysis. *Fisheries Research 1998,* **38**: p.1-7.

Peirce G. J. and P.R. Boyle. 2003. Empirical modeling of interannual trends in abundance of squid (*Loligo forbesi*) in Scottish waters, *Fisheries Research 2003,***59.**

Stergiou, K. I. and E.D. Christou, 1996. Modelling of forecasting annual fisheries catches: comparison of regression, univariate and multivariate time series methods, *Fisheries Research 1996*, **25**, *p. 105-138.*

Stergiou K. I. and E.D. Christou and G. Petrakis. 1997. Modelling of forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods, *Fisheries Research 1997,* **29**, p. 55-95.