Fishery Landing Forecasting Using Wavelet-Based Autoregressive Integrated Moving Average Models

Ani Shabri¹ and Ruaidah Samsudin²

¹Department of Science Mathematic, Faculty of Science, Universiti Teknologi Malaysia (UTM), 81310 Johor, Malaysia
²Department of Software Engineering, Faculty of Computing, Universiti Teknologi Malaysia (UTM), 81310 Johor, Malaysia

Correspondence should be addressed to Ani Shabri; ani_sabri@hotmail.com

Received 26 August 2014; Revised 7 November 2014; Accepted 11 December 2014

Academic Editor: Erol Egrioglu

Copyright © 2015 A. Shabri and R. Samsudin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The accuracy of the wavelet-ARIMA (WA) model in monthly fishery landing forecasting is investigated in the study. In the first part of the study, the discrete wallet transform (DWT) is used to decompose fishery landing time series data. Then ARIMA, as a powerful forecasting tool, is implemented to predict each wavelet transform subseries components independently. Finally, the prediction results of the modeled subseries components are summed to formulate an ensemble forecast for the original fishery landing series. To assess the effectiveness of this model, monthly fishery landing recorded data from East Johor and Pahang states of Peninsular Malaysia have been used as a case study. The result of the study shows that the proposed model was found to provide more accurate fishery landing series forecasts than the individual ARIMA model.

1. Introduction

Fishing is one of the most important industries in Malaysia. For many years, the fisheries sector in Malaysia makes a significant contribution to the national economy in terms of income, foreign exchange, and employment. Besides that, it also plays a significant role as a major supplier of animal protein for the local citizen consumptions. In order to ensure that the local demand can be catered without highly depending on imported fish, the authority has kept track of the annual total fishery production and took necessary actions to increase or maintain the level of production while at the same time maintaining a sustainable ecology. To achieve this aim, it is necessary to forecast uncontrollable events, such as possible abundance or biomass changes [1]. However, a proper selection of models for forecasting fishery landing is one of the major research efforts over the past few decades.

Traditional statistical methods such as linear regression, autoregressive, moving average, and autoregressive integrated moving average (ARIMA) models have been applied to forecast the landings and catch per unit effort of many fish and invertebrate [2–5]. For modelling fisheries sciences time series data, ARIMA model has been popular and widely chosen [1, 2, 6–9]. The ARIMA model is the standard parametric forecasting model for statistical time series analysis since the 1970s. The ARIMA model is a linear combination of time-lagged variables and error terms. The popularity of the ARIMA model is due to their statistical properties, such as the well-known Box-Jenkins methodology, forecasting capabilities, and richness of information regarding time-related changes. Although ARIMA models have been proven to be effective in many decision support applications, they still have certain shortcomings. They are basically linear models assuming that data are stationary and have a limited ability to capture nonstationarities and nonlinearities in series data [10, 11].

Due to the limitations of the traditional statistical, another approach that has been used for dealing with nonstationary and nonlinear characteristic of a time series is employed, the decomposition approach. Forecasting using a decomposition method is often more useful in providing forecasts and information regarding the component of a time series than trying to predict a single time series [12]. In the last decade, wavelet transforms have become
a common tool for analyzing variations, periodicities, and trends in time series [13–16]. Recently, new hybrid models based on wavelet transform have been proposed in time series forecasting. The corresponding empirical results demonstrated that the hybrid wavelet transform with other model outperform individual forecasting model in many cases [16–23]. Wavelet transforms provide useful decompositions of the original time series; therefore, wavelet-transformed data improve ability of a forecasting model by capturing useful information on various resolution levels. However, existing literatures regarding fishery landing forecasting have not adopted wavelet transform processes, and this study will be filling this gap.

In this study, we introduce wavelet transform and ARIMA to construct a novel fishery landing forecasting methodology. In this methodology, the original fishery landing series is decomposed into several subseries using wavelet transform by Mallat algorithm. Secondly, the tendencies of these subseries are then modeled and forecasted using ARIMA. Finally, the forecasted value of the proposed model can be obtained by summing the forecasted value of each subseries. In order to evaluate the performance of the proposed approach, the monthly fishery landing series in East Johor and Pahang of Peninsular Malaysia were used as the illustrative example and its prediction performance was compared with some popular individual ARIMA model.

2. Methodology

2.1. ARIMA Model. The ARIMA models were introduced by Box and Jenkins [24] and have dominated many areas of time series forecasting. Box-Jenkins models used ARIMA($p, d, q$) × ($P, D, Q$)$_s$ models composed of the nonseasonal part and seasonal part which are represented by the following way:

$$\phi(B) \Phi(B^s) (1 - B)^d (1 - B^s)^D y_t = \theta(B) \Theta(B^s) a_t, \quad (1)$$

where $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$, $\Phi(B^s) = 1 - \sum_{i=1}^{P} \Phi_i (B^s)^i$, $\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$, $\Theta(B^s) = 1 - \sum_{i=1}^{Q} \Theta_i (B^s)^i$ are the polynomials in $B$ of degree $p$, $q$, $P$, and $Q$, respectively. $p$, $d$, $q$, $P$, $D$, and $Q$ are integers, $p$ and $q$ are the order of nonseasonal autoregressive and moving average, and $P$ and $Q$ are the order of seasonal autoregressive and moving average, respectively. $d$ is the number of regular difference, $D$ is the order of seasonal differences, and $a_t$ is the random error.

The Box-Jenkins methodology includes four iterative steps: identification, estimation, diagnostic checking, and forecasting. Figure I shows the process of ARIMA modelling.

In the identification step, data transformation is often used to make the time series stationary. The autocorrelation (ACF) and partial autocorrelation function (PACF) are used to determine whether or not the series is stationary and as the basic tools in order to identify the appropriate ARIMA model. Once the tentative model is identified, the parameters of the model are estimated straightforward. The last step in model building is the diagnostic checking of model adequacy. Adequacy of the model was performed by examining the ACF of residual and through diagnostic checks of residual using Ljung-Box test. The process is repeated several times until a satisfactory model is finally selected. The forecasting model was then used to compute the fitted values and forecasts values.

2.2. Wavelet Transform. Wavelet transformations (WT) provide useful decomposition of the original time series by capturing useful information on various decomposition levels. WTs can be divided in two categories: continuous wavelet transforms (CWT) and discrete wavelet transforms (DWT). For the time series $x(t)$, the CWT of the time series $x(t)$ with respect to a mother wavelet $\psi(t)$ is defined as

$$W(\omega, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) \psi^* \left( \frac{t - \omega}{s} \right) dt, \quad (2)$$

where ($*$) corresponds to the conjugate complex function, $t$ stands for a time, $\omega$ stands for time step, and $s \in [0, \infty]$ for the wavelet scale. The CWT is not often used for forecasting due to its computationally complex and time requirements to compute. Instead, successive wavelet is often discrete in forecasting applications to simply the numeric solutions. DWT requires less computation time and is simpler to implement. DWT can be defined as

$$W_{mn} = 2^{-m/2} \sum_{n=0}^{N-1} \psi \left( 2^{-m} t - n \right) x(t), \quad (3)$$

where $W_{mn}$ is the wavelet coefficient for the discrete wavelet at scale $s = 2^m$ and $\tau = 2^m n$. According to Mallat’s theory, the original discrete time series $x(t)$ can be decomposed into a series of linearity independent approximation and detail signals by using the inverse DWT. The inverse DWT is given by Mallat [25]:

$$x(t) = T + \sum_{m=1}^{M} \sum_{n=0}^{2^M-m-1} W_{mn} 2^{-m/2} \psi \left( 2^{-m} t - n \right) \quad (4)$$
or in a simple format as
\[ x(t) = A_M(t) + \sum_{m=1}^{M} D_m(t), \] (5)
where \( A_M(t) \) is called approximation subseries or residual term at level \( M \) and \( D_m(t) (m = 1, 2, \ldots, M) \) are detail subseries which can capture small features of interpretational value in the data.

3. Results

3.1. Study Areas and Data. To evaluate the performance of the proposed forecasting model using wavelet transform with ARIMA methodologies, this paper uses monthly fishery landing data obtained from the Annual Fisheries Statistics through the official website of Department of Fisheries Malaysia, Ministry of Agriculture and Agro-Based Industry Malaysia. This research is focusing on the marine fishery landing for two different states, East Johor and Pahang of Peninsular Malaysia (Figure 2). Whole monthly data of fishery landing in East Johor and Pahang states, covering the period from January 2001 to December 2012 with a total of 144 observations, are used, as showed in Figure 3. In this fishery landing data, the data from January 2001 to December 2011 are used as training dataset (132 observations) and the remaining data from January 2012 to December 2012 are chosen as testing dataset (12 observations).

3.2. Performance Criteria. For comparison of the forecasting performance of the proposed model, two criteria are used as accuracy measures, namely, the root mean squared error (RMSE) and mean absolute percentage error (MAPE). These criteria are calculated as follows:
\[
RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{y}_t)^2},
\] (6)
\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100,
\]
where \( y_t \) is the actual data, \( \bar{y} \) is mean of actual data and \( \hat{y}_t \) is the forecasted value of period \( t \), and \( n \) is the number of observations. Obviously the smaller the values of RMSE and MAPE, the higher the efficiency of the model.

4. Forecasting Results

4.1. Fitting ARIMA Model to the Data. Figure 3 describes the curve of monthly fishery landing series in East Johor of Peninsular Malaysia in units of tones. The data show nonlinear, nonstationary, and seasonal characteristic. The sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) for the original fishery landing series are plotted in Figure 4. In the ACF there were significant spikes present near lags 12, 24, 36, 48, and 60, and therefore the series was seasonally differenced with 12 as period. The plot of ACF and PACF after seasonal differencing is shown in Figure 5.

The ACF is damping out in sine-wave manner with significant spikes near lags 1, 7, and 12. In the PACF, there are significant spikes at lags 1, 6, 11, 12, and 13. This indicates a possible ARIMA\((p, 0, q)\times(P, 1, Q)_{12}\) model. All combinations are evaluated to determine the best model out of these candidate models. The identification of the best model for the fishery landing series is based on minimum AIC. After extensive investigation, the model finally selected was an ARIMA\((7, 0, 0)\times(0, 1, 1)_{12}\). The model can be expressed as
\[
(1 - 0.419B + 0.111B^2 - 0.007B^3 - 0.022B^4 + 0.050B^5 - 0.022B^6 - 0.178B^7)(1 - B^{12})y_t = 61.010 + (1 + 0.911B^{12})a_t.
\] (7)

Once an appropriate model is chosen, the Box-Jenkins methodology requires examining the residuals of the model to verify that the model is an adequate one for the series. For a good forecasting model, the residuals left over after fitting model should be white noise. Figure 6 displays a plot of the standardized residuals, the ACF of residuals and the \( P \) value of the Ljung-Box statistic at lags 1–20. Inspection of the time plot of standardized residual in Figure 6 shows no obvious patterns. From the residual plot of the ARIMA model, it was found that the ACF of residuals is small and lies within confidence limits which show that the residuals from the best model are white noise. Additionally, the adequacy of the model is confirmed using the Ljung-Box test. The \( P \) values of Ljung-Box test for all lags exceed 0.05 which means accepting model accuracy at 95% significance level. It
Figure 3: The monthly fishery landing series in East Johor and Pahang states from Jan. 2001 to Dec. 2012.

Figure 4: ACF and PACF of the original fishery landing series.

Figure 5: ACF and PACF of the original fishery landing series after seasonal differencing.
is clearly supported that the ARIMA\((0, 0, 1) \times (1, 1, 1)_{12}\) is the adequacy model of the fishery landing series in East Johor state.

According to the above ARIMA model, the future fishery landing from East Johor can be obtained. RMSE and MAPE values of this model for test data set are 971.29 and 8.62\%, respectively.

For Pahang data, the fitness model generated from the data set is ARIMA\((3, 1, 1) \times (2, 0, 0)_{12}\). The equation of this model is

\[
(1 - 0.673B + 0.130B^2 + 0.040B^3) \\
\times (1 - 0.422B^{12} - 0.182B^{24})(1 - B) y_t = \left (1 + 0.973B \right ) a_t. \tag{8}
\]

Diagnostics for this model are displayed in Figure 6(b), and it appears that this model fits the data well. RMSE and MAPE values of this model for test data set are 1776.01 and 14.77\%, respectively.

4.2. Fitting Wavelet Transform-ARIMA Model to the Data. The hybrid wavelet and ARIMA model (WA) is obtained by combining two methods, discrete wavelet transform (DWT) and ARIMA model. In the WA model, the original fishery landing series was decomposed into a certain number of sub-time series components which were entered to the ARIMA model in order to improve the model accuracy. When conducting wavelet analysis, the number of decomposition levels that are appropriate for the data must be chosen. To choose the number of decomposition level, the following formula is used [17, 19]:

\[
L = \text{int} \left [ \log (N) \right ], \tag{9}
\]

where \(L\) is the level of decomposition and \(N = 144\) is the number of time series data. According to this formula, the optimal number of decomposition levels for the fishery landing series data in this study would have been two. The approximation and detail subseries of the original time series of East Johor and Pahang fishery landing series, decomposed at level 2 by the Db2 wavelet, are presented in Figures 7 and 8. Figures 8 and 9 show that the original fishery landing series is decomposed into one approximation (A2) and two detail series (D1 and D2).

In this study, we tried to investigate the effects of the used decomposition level on the model efficiency. To achieve this purpose, the time series data were decomposed into one, two, and three levels by Daubechies-2 (Db2) wavelet. Figure 9 describes the process of hybrid wavelet and ARIMA model using one-level (WA1), two-level (WA2), and three-level (WA3) wavelet decomposition for original fishery landing series, respectively. As can be seen from Figure 9, WA models can be described as the following steps.

(i) Decompose the original fishery landing time series \(y(t), t = 1, 2, \ldots, n\), which were decomposed to one, two, and three levels by DWT.

(ii) Use the ARIMA model to model each of Ds and the As. ARIMA models are then applied to forecast the future one-day values of these Ds and the As.

(iii) The forecasting value is obtained by summing up all the prediction results of each of Ds and the As component.
Figure 7: The two-level wavelet decomposition for original series of East Johor state.

Figure 8: The two-level wavelet decomposition for the original series of Pahang state.
Table 1: Performance of the four forecasting methods for monthly fishery landing data.

<table>
<thead>
<tr>
<th>Data</th>
<th>Statistics</th>
<th>ARIMA</th>
<th>WA1</th>
<th>WA2</th>
<th>WA3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>971.29</td>
<td>736.13</td>
<td>574.22</td>
<td>574.65</td>
</tr>
<tr>
<td></td>
<td>MAPE</td>
<td>8.62%</td>
<td>11.76%</td>
<td>7.34%</td>
<td>7.43%</td>
</tr>
<tr>
<td>East Johor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pahang</td>
<td>RMSE</td>
<td>1776.01</td>
<td>1367.77</td>
<td>866.14</td>
<td>774.37</td>
</tr>
<tr>
<td></td>
<td>MAPE</td>
<td>14.77%</td>
<td>11.27%</td>
<td>7.73%</td>
<td>7.00%</td>
</tr>
</tbody>
</table>

For comparing the forecasting accuracy, the same testing data set is examined for three proposed forecasting models. The performance measurements of the selected forecasting models are given in Table 1. For East Johor state, it can be observed that the magnitudes of RMSE and MAPE using the proposed W2 and W3 models are almost the same and smaller than those using the W1 and ARIMA models. Table 2 shows the percentage improvement of the proposed model with ARIMA model. The improvement listed in this paper is calculated in terms of RMSE and MAPE by

\[
\text{Percentage Improvement} = \frac{E_b - E_{\text{proposed}}}{E_b} \times 100\%, \quad (10)
\]

where \(E_b\) denotes the error of basic method used as comparison, which is here the ARIMA prediction error. Table 2 shows the proposed forecasting procedures using the W1 model which was able to improve the RMSE in comparison with the single ARIMA model by about 24.21%. Forecasting precision of WA2 model has a great improvement to ARIMA where the RMSE and MAPE reduced by 40.88% and 14.86%, respectively. Moreover, WA3 also gives better forecasting than ARIMA (reduced in the RMSE of 40.84% and MAPE of 13.83%).

In Table 1, for Pahang state, these results demonstrate again that the proposed models perform better in fishery landing forecast. Also, it has been observed from Table 2 that the proposed forecasting procedures using the WA1, WA2, and WA3 models lead to 22.99%, 51.23%, and 56.40% reductions in total RMSE and 23.74%, 47.68%, and 52.62% reductions in total MAPE, respectively, in comparison with the ARIMA model alone.

By comparing the obtained results (Table 2), it can be clearly seen that, by increasing the decomposition level to 3, the proposed model’s performance increases; therefore level 3 can be considered as proper decomposition level for the data. The actual fishery landing data and forecasted values in East Johor and Pahang states for the ARIMA, WA1, WA2, and WA3 models are illustrated in Figures 10 and 11, respectively. It can be observed from Figures 10 and 11 that the forecasted values obtained from the proposed models are closer to the actual values than those obtained from the ARIMA model. Obviously, the single ARIMA model does not perform well. The forecasting accuracy of the ARIMA model is the worst among all models investigated in this paper. The
Table 2: Percentage improvement of the proposed model with ARIMA model.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>East Johor</th>
<th>Pahang</th>
</tr>
</thead>
</table>
|            | WA1        | WA2    | WA3    | WA1 | WA2 | WA3 |%
| RMSE (%)   | 24.21%     | 40.88% | 40.84% | 22.99% | 51.23% | 56.40% |%
| MAPE (%)   | −36.41%    | 14.86% | 13.83% | 23.74% | 47.68% | 52.62% |%

Figure 10: Comparison of forecast results from ARIMA, WA1, WA2, and WA3 models for fishery landing in East Johor state of Peninsular Malaysia.

Figure 11: Comparison of forecast results from ARIMA, WA1, WA2, and WA3 models for fishery landing in Pahang state of Peninsular Malaysia.

5. Conclusion

ARIMA models have been widely used in fisheries science time series forecasting problems. Unfortunately, ARIMA models are basically linear not capable of accurately forecasting the fishery landing time series, due to the fact that the series which is often highly nonstationary, nonlinearity, and seasonality. A fishery landing forecasting methodology based on wavelet transform combined with ARIMA is proposed in this study. To assess the effectiveness of this model, monthly fishery landing record data from East Johor and Pahang states of Peninsular Malaysia have been used as a case study. Empirical results indicate that the proposed model showed a great improvement in fishery landing modeling and produced better forecasts than the ARIMA models alone. ARIMA models have enhanced forecasting accuracy when the wavelet transform is applied to original fishery landing data. Thus it can be concluded that the proposed wavelet-ARIMA model may be an effective tool as a very promising methodology for complex problems such as fishery landing time series forecasting with seasonality variations and nonlinearity.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The author wish to thank the Ministry of Higher Education (MOHE) for the financial support of the work under the Fundamental Research Grant Scheme (FRGS no. 4F275). Also the authors gratefully acknowledge the critical comments and corrections of the anonymous reviews; their comments significantly improved the original paper.

References


