PENERBIT UMT
Menjana Khazanah Ilmuwan

Journal of Mathematical Sciences and Informatics
Journal Homepage: https://journal.umt.edu.my/index.php/jmsi
eISSN: 2948-3697
DOI: http://doi.org/10.46754/jmsi.2023.12.002

FORECASTING MONTHLY FISH LANDING IN EAST COAST PENINSULAR MALAYSIA USING SARIMA-ARTIFICIAL NEURAL NETWORK

NOR AZLYN EZURA HASMIN AND NORIZAN MOHAMED*

Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia; azlynezura1982@gmail.com, norizan@umt.edu.my

**Corresponding author: norizan@umt.edu.my*

ARTICLE INFO

Article History:
Received
1 APRIL 2023
Accepted
28 AUGUST 2023
Available online
25 JANUARY 2024

Section Editor:
Binyamin Yusoff

Keywords:
Marine Fish Landing; Autoregressive Seasonal Autoregressive Integrated Moving Average; Multilayer Feed Forward Neural Network; Box-Jenkins Method; SARIMA model

ABSTRACT

Marine resources management is increasingly crucial due to the growing awareness that these resources are limited. Effective fish landing management plays a pivotal role in making informed decisions, particularly in preserving fish stocks for future food security. This paper aims to develop the best model for forecasting monthly fish landings in East Coast Peninsular Malaysia using the Box-Jenkins method and Artificial Neural Networks (ANN). The monthly dataset was divided into training data (January 2012 to December 2019) and testing data (January 2020 to December 2021). Employing the Box-Jenkins approach, the SARIMA model $(1,1,0)(0,1,1)^{12}$ with minimum AIC and BIC values was identified as the most effective. By utilising lag variables from the SARIMA model as inputs for the ANN, we developed the SARIMA-ANN model. Model accuracy was assessed using the mean absolute percentage error. The findings showed that the integration of the SARIMA–ANN model enhances accuracy in forecasting the monthly fish landing performance in East Coast Peninsular Malaysia.

INTRODUCTION

The fishing industry has played a significant role in ensuring food security in Malaysia, contributing to employment and fostering economic growth. With growing awareness of the limited nature of marine resources, effective management has become vital. Making informed decisions is crucial for managers to uphold food security sustainability, given that fish serves as the primary source of protein for many. The shortage of fish landing required agencies like the Fisheries Development Authority of Malaysia (LKIM) to maintain stocks of frozen fish held by companies to address market supply gaps. As highlighted by [1], weather-related downtimes, particularly during the monsoon season, significantly impact fishing operations in Peninsular Malaysia. It is estimated that approximately 75% of fishermen are unable to venture to sea during this period.

Therefore, the forecasting of fish landing is essential to address these challenges. As noted by [2], time series analysis of fishery landings plays a vital role in fisheries management and decision-making, as it can reveal trends and seasonality patterns in the data. With the appropriate model, predicting fish landing can empower managers to make effective decisions in ensuring a stable fish supply for the country. Time series modelling proves to be a valuable tool, widely applied in various studies, including fisheries. Previously, several researchers have implemented different models, such as discrete wavelet transform- autoregressive integrated moving average (ARIMA), nonlinear regression and ARIMA transfer function [3]. However, there is a noticeable gap in studies specifically focusing on forecasting fish landing performance in Malaysia.

Researchers have employed various methods and approaches in forecasting fish landings. For this study, an approach integrating Seasonal ARIMA (SARIMA) and artificial neural network (ANN) was selected based on the behavioural characteristics of the time series data (monthly fish landing in East Coast Peninsular Malaysia). However, the efficiency of the integrated SARIMA-ANN model in predicting the fish landing performance is a novel exploration, as researchers have not yet applied this method predicting of fish landing performance in Malaysia. Therefore, this paper proposes to investigate the integration between SARIMA and ANN, with the aim of predicting monthly fish landing data and evaluating its performance. Developing the best prediction model for monthly fish landing could greatly assist in managing fish stocks, providing valuable insights for estimating future time series of fish landing performance.

FORECASTING MODELS

Accurate prediction of fish landing values is crucial for effective fish stock management, with fish stock identification being a key component in modern fisheries management [4]. While information on fish stocks is considered vital, there remains a scarcity of case studies related to stock assessment in fisheries management. In this context, [4] presented a case study aimed at identifying stock problems and outlining future directions for fish stock management. Fisheries management necessitates comprehensive information to estimate the size of fish stocks and provide guidance on the sustainable level of catches. According to the 2020 State of World Fisheries and Aquaculture report, ineffective fisheries management in some regions has led to poor and deteriorating fish stock conditions. The uneven progress in fisheries management highlights the urgent need to replicate and adapt successful policies in these regions. The most important aspect of sustainable management is making informed decisions, requiring information and historical trends for assessing fisheries status, including baseline information on the past and current usage patterns of the area [5].

Raman *et al.* [6] conducted a study on forecasting marine fish landing in Odisha using the ARIMA model. Data for the analysis was sourced from the 'National Marine Fisheries Data Centre' of the Central Marine Fisheries Research Institute, Kochi, covering the period from 1985 to 2012. The study aimed to estimate short-term forecasting by fitting the ARIMA model in two scenarios: One accounting for intervention in the model and the other involving log-transformed data. The results revealed that the model with log-transformed data performed better with the ARIMA (2, 0, 2) $(0, 1,4)^4$ model than the intervention component model with ARIMA (0, 1, 1) $(0, 1, 1)^4$.

According to [7, 6], ARIMA stands out as the most widely utilised model for forecasting time series data. However, a notable weakness of this model lies in its assumption of linearity. Since it is uncommon for time series to contain linear components, relying solely on ARIMA may not be sufficient in modelling and forecasting. The time series analysis of marine fish landing, spanning January 2023 to December 2019, reveals a distinct seasonal pattern. Hence, replying only to ARIMA might prove inadequate to accurately forecast marine fish landings. Farhan *et al.* [8] applied the SARIMA method in forecasting container throughput at ports. This method takes into account seasonal variations present in the time series data of container ports. The effectiveness of the SARIMA model was assessed for 20 major international container ports. Evaluation metrics, including the Wilcoxon signed ranked test for bias determination (Hollander & Wolfe, 1999) and mean absolute error (MAE) and mean absolute percentage error (MAPE) for forecast accuracy, demonstrated the applicability and effectiveness of the SARIMA model in predicting container throughput at major international ports.

In an effort to aid decision-makers in setting priorities for fisheries management, [9] conducted a study employing the Box-Jenkins methodology to develop a SARIMA model for the monthly catches of two fish species over a five-year period (2007 to 2011). The stationary test, utilising the autocorrelation function (ACF), initially indicated that the internal organisation of the time series exhibited non-stationarity as it displays a hyperbolic decay pattern. Hence, appropriate differencing was implemented to transform the data into a stationary time series. The study concludes that the Box-Jenkin method is one of the most efficient and prominent approaches for forecasting time series data. Specifically, the SARIMA model was developed to predict the catches of *Trichiurus lepturus* (beltfish) and *Amblygaster leiogaster* (blue sprat) for the upcoming 5 months. The models ARIMA $(1, 1, 0)(0, 0, 1)^{12}$ and SARIMA $(0, 1, 1)(0, 0, 1)^{12}$ were identified as the best-fit models, which was confirmed by the Ljung-Box test.

METHODOLOGY

Box-Jenkin Method-SARIMA Modelling Approach

The Box-Jenkins method, developed by George E.P Box and Gwilym M. Jenkins (1970), is founded on the principle of stinginess [10]. Its objective is to discern the most optimal and reliable model capable of forecasting future values for a given time series. The autoregressive moving average (ARMA) is a stationary series that combines the autoregressive (AR) and moving average (MA) models and is effective for explaining stationary time series. The general representations of AR, MA and ARMA models are as follows:

$$\text{AR: } y_t = \mu + Ø_1 y_{t-1} + Ø_2 y_{t-2} + \ldots + Ø_p y_{t-p} + \varepsilon_t, \tag{1}$$

where y_t is the current value of the dependent variable, y_{t-p} depends on the value of the dependent or current values, $Ø_j$ is the parameter to be estimated, and ε_t is the term error.

$$\text{MA: } y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t, \tag{2}$$

where μ is the mean variable size, and θ is the estimated moving average parameter.

$$\text{ARMA: } (1-Ø_1 B-Ø_2 B^2-\ldots-Ø_p B^p)\, y_t = \mu + (1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q)\, \varepsilon_t \tag{3}$$

where $Ø_p$ is the autoregressive parameter, μ is a constant coefficient, and the values of p and q obtained through AR (p) and MA (q) contribute to developing the ARMA model.

The ARIMA model is applied to non-stationary data through a differencing process. It is categorised into non-seasonal and seasonal ARIMA models. The general expression of the non-seasonal ARIMA (p,d,q) model can be defined as:

$$Ø_p(B)(1-B)^d y_t = \theta_q(B) a_t \tag{4}$$

where d is a sequence of differences, $Ø_p(B)$ is represented as the stationary autoregression operator $Ø_p(B) = (1-Ø_1 B - \ldots . Ø_p B^p)$, while $\theta_q(B)$ is the invertible moving average defined as $\theta_q(B) = (1-\theta_1 B - \ldots . \theta_p B^p)$. The term $(1-B)^d y_t$ denotes the differencing operator.

The seasonal ARIMA is capable of modelling time series data containing seasonal periodic components, represented as SARIMA(P, D, Q), where P is the seasonal autoregression parameter, D is the seasonal differencing parameter, and Q is the seasonal moving average parameter. To address this seasonal series, Box-Jenkins recommends the following model:

$$\emptyset_p(B)\Phi_p(B)(1\text{-}B)^d(1\text{-}B^s)^D Y_t = \theta_q(B)\theta_Q(B^S)a_t \quad (5)$$

where s refers to the number of periods per season.

To select the best-fit SARIMA model, two commonly used goodness-of-fit criteria are Akaike's Information Criterion (AIC) and Schwarz Bayesian Information (BIC). The model with the lower AIC and BIC values is considered the most adequate for the data [10]. The formulas for AIC and BIC are as follows:

$$AIC = \ln(SSE) + \frac{2k}{n} \quad (6)$$

$$BIC = \ln(SSE) + \frac{k}{n}\ln(n) \quad (7)$$

where n is the number of observations, SSE is the sum of squared error and k is the summation of non-seasonal and seasonal parameters (p+q+P+Q+s) [11].

Artificial Neural Networks

ANN are computational models inspired by the functioning of biological nervous systems, particularly the human brain, in processing information. ANN are composed of multiple layers of simple processing elements known as neurons [12]. The structure of ANN includes an input layer, hidden layers and an output layer. There are two types of ANN: Feedforward and recurrent neural network. In a feedforward ANN, information moves in only one direction from the first tier onwards until it reaches the output node. On the other hand, a recurrent neural network operates by storing the output of a particular layer and feeding it back to the input. In this study, a feedforward ANN, specifically a multilayer feedforward neural network, was selected.

Multilayer Feedforward Neural Network

A multilayer network consists of at least one hidden layer. As explained by Haykin [13], the inclusion of one or more hidden neurons serves the purpose of intervening between the input and output layers, allowing the network to extract higher-order statistics. The term "feedforward" indicates that the output from one layer of neurons is forwarded into the next layer. The inputs are fed into the units making up the input layers, and they are weighted and simultaneously transmitted to a second layer, referred to as a hidden layer. The hidden layer units will generate outputs, serving as input to another hidden layer. The overall process of the multilayer feedforward neural network is illustrated in Figure 1.

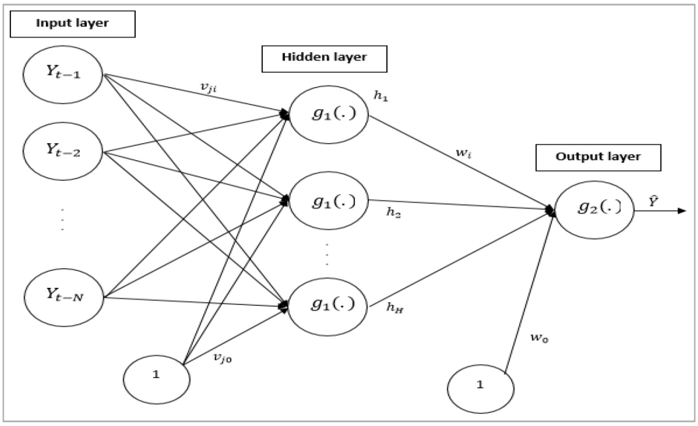

Figure 1: The architecture of the multilayer feedforward neural network model with one hidden layer, N input nodes, H hidden nodes and one output node

The connection between neurons is denoted as a weight, and each neuron consists of summing and activation functions. The value Ŷ with N input nodes, H hidden nodes and one output node is expressed as:

$$\hat{Y} = g_2(\textstyle\sum_{j=1}^{H} w_j h_{j.t} + w_0) \tag{8}$$

where the w_j represents the output weight from hidden node j to the output node, w_0 is the bias for the output node, and g_2 is an activation function. The values of the hidden nodes h_j, j =1,2,3 ...,H are given by:

$$h_{j,t} = g_1\left(\textstyle\sum_{t=1}^{N} v_{ji} y_{t,i} + v_{j0}\right) j = 1,2, \cdots H \tag{9}$$

In this equation, v_{ji} is the input weight between the input node i to hidden node j, v_{j0} is the bias for hidden node j, and $y_{t,i}$ represents the lag variables. The lag variables ($y_{t,i}$,..., $y_{t,N}$) correspond to (y_{t-i} ,..., y_{t-N}), where i = 1,2,...,*N*. The activation function g_1 may be the same as g_2 or a different function, as noted by [13].

To determine the best model, it is essential to classify the transfer function, the appropriate data transformation, the best training algorithm, the number of hidden nodes, and the combination of lag variables from the SARIMA model. The MATLAB software was used for this process.

Evaluating Model Performance

After obtaining possible SARIMA models, accuracy must be assessed using standard error metrics: Mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). These metrics are commonly used for evaluating time series performance and are defined by the following formula:

1. MAPE = $\frac{1}{n}(\sum_{t=1}^{n} \frac{|y_t - \hat{y}_t|}{y_t}) \times 100$ (10)
2. MAE = $\frac{1}{n}\sum_{t=1}^{n} |y_t - \hat{y}_t|$ (11)
3. MSE = $\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2$ (12)
4. RMSE = $\sqrt{MSE}$ (13)

Here, $\hat{y}_t$ represents the forecasted value at point t,y_t is the original data value, and n is the total number of data points. Lewis [14] (1982) devised a table of MAPE values for industrial and business data, providing a benchmark for interpreting forecasting performance, a practice adopted in various studies. This study is interpreted based on MAPE values from Lewis as shown in Table 1.

Table 1: The level of forecasting based on the MAPE values (Lewis, 1982)

MAPE	Interpretation
< 10%	High accuracy forecasts
10%-20%	Good forecasting
20%-50%	Reasonable forecasting
> 50%	Inaccurate forecasts

RESULTS AND DISCUSSION

This section discusses the results of the best prediction model, comparing the single ARIMA method with the integrated SARIMA and ANN models. The dataset spans 10 years (January 2012 to December 2021) and pertains to monthly fish landing in East Coast Peninsular Malaysia, encompassing the states of Kelantan, Pahang, and Terengganu. Data was sourced from the Department of Fisheries Malaysia's website and divided into two sets: in-sample data covering January 2012 to December 2019, and out-sample data covering January 2020 until December 2021.

Seasonal Autoregressive Integrated Moving Average

The time series plot of the in-sample data, as depicted in Figure 2, suggests that the series is non-stationary. Therefore, a prerequisite for fitting the ARIMA model involves the removal of the the trend and stabilisation of variance through seasonal differencing. Figure 2 illustrates the time series plot of the first difference of fish landing time series data, indicating the stationary in mean.

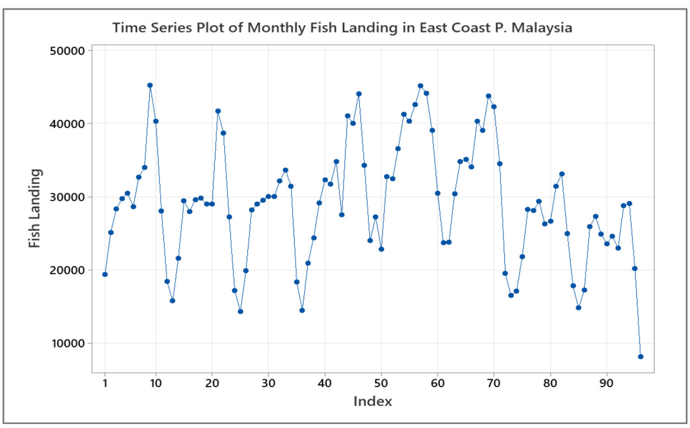

Figure 2: The time series plot of fish landing in East Coast Peninsular Malaysia

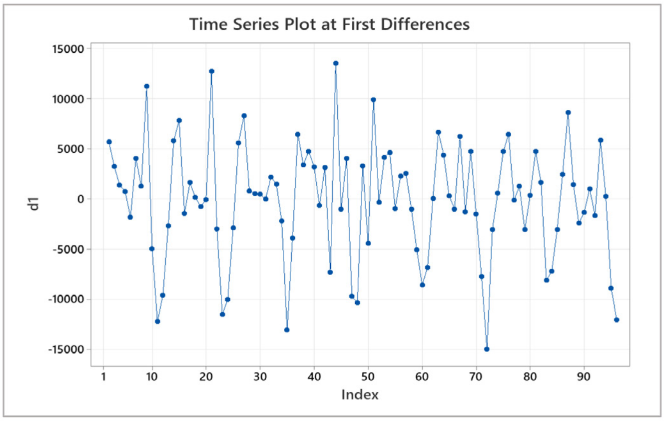

Figure 3: The plot of the first differences in fish landing on the East Coast

The plot of fish landing in East Coast Peninsular Malaysia, utilising 1st differenced data as shown in Figure 3, indicates stationarity in mean. Further exploration through the ACF and partial ACF (PACF) plots of 1st differences (Figures 4 and 5) reveals significant spikes in lags 12, indicating the presence of seasonality in the monthly fish landing in East Coast Peninsular Malaysia, with a periodicity of 12. Consequently, seasonal differencing of the 1st non-seasonal differenced data is warranted.

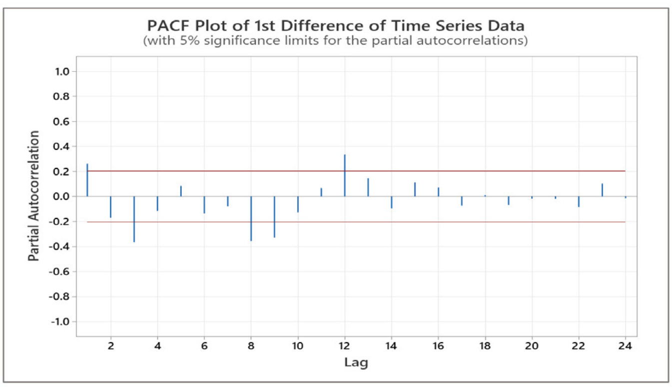

Figure 4: The PACF plot (at 1st differencing order)

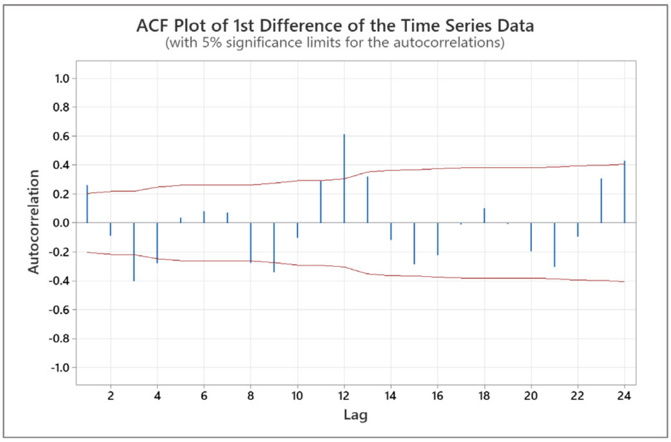

Figure 5: The ACF plot (at 1st differencing order)

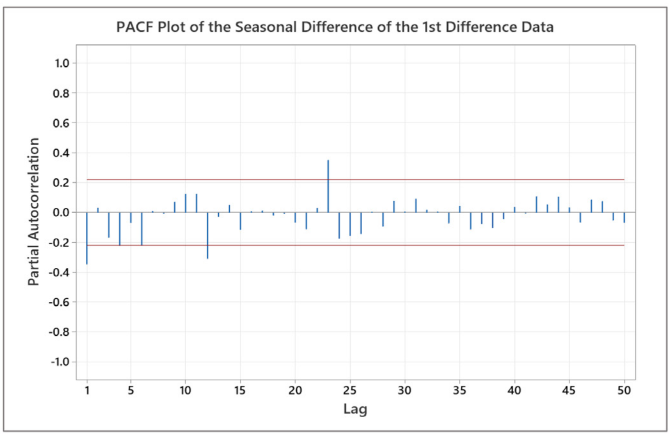

Figure 6: The PACF plot of the seasonal difference (at 1st seasonal differencing order)

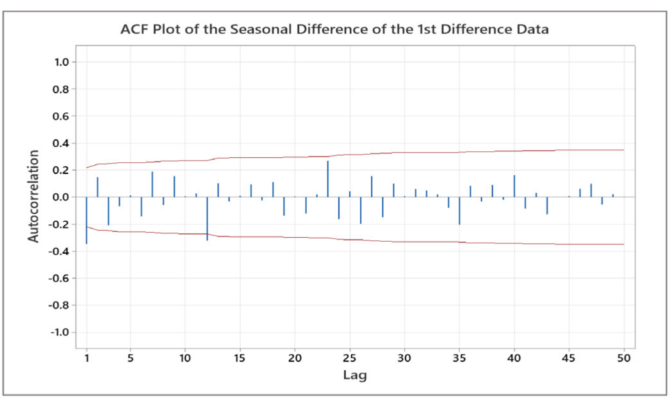

Figure 7: The ACF plot of the seasonal difference (at 1st seasonal differencing order)

Analysis of the PACF in Figure 6 reveals spikes at lag 1, lag 4 and lag 6 for the non-seasonal part. Meanwhile, the ACF in Figure 7 exhibits a cutoff at lag 1. Considering statistical simplicity and the principle that complex methods may not necessarily yield more accurate forecasts [15], it is suggested to exclude lag 4 and lag 6 in the PACF. Therefore, the proposed model for ARIMA is ARIMA (1,1,1).

The ACF and PACF analyses for seasonal differences reveal that ACF and PACF spikes at seasonal lag 12 diminish to zero for other seasonal lags, suggesting P=1 and Q=1. The SARIMA model proposed under this finding process is SARIMA $(1,1,1)(1,1,1)^{12}$. However, instead of testing only one model, a thorough examination involved the exploration of eight possible models in terms of parameter selection, as outlined in Table 2.

Table 2: The AIC and BIC values for the SARIMA models

SARIMA MODELS	AIC VALUES	BIC VALUES
$(0,1,1)\ (0,1,1)^{12}$	**20.69642**	**20.80327**
$(1,1,1)\ (1,1,1)^{12}$	20.7382	20.89847
$(1,1,1)\ (0,1,1)^{12}$	20.71787	20.85143
$(1,1,1)\ (1,1,0)^{12}$	21.01111	21.14467
$(1,1,0)\ (1,1,0)^{12}$	20.99351	21.10035
$(0,1,1)\ (1,1,1)^{12}$	20.71584	20.8494
$(1,0,0)\ (1,1,1)^{12}$	20.70262	20.80947
$(1,1,0)\ (0,1,1)^{12}$	**20.67209**	**20.77894**

According to the results of the AIC and BIC of all possible models, it shows that the best SARIMA models are SARIMA $(0,1,1)(0,1,1)^{12}$ and SARIMA $(1,1,0)\ (0,1,1)^{12}$.

Table 3: The significant values of the parameters for both models

Models	Parameters	Significant Parameter (P-value)
	MA 1	0.000
SARIMA $(0,1,1)\ (0,1,1)^{12}$	SMA 12	0.000
	AR 1	0.001
SARIMA $(1,1,0)\ (0,1,1)^{12}$	SMA 12	0.000

As presented in Table 3, the p-value of both models SARIMA $(0,1,1)(0,1,1)^{12}$ and SARIMA $(1,1,0)(0,1,1)^{12}$ are statistically significant since the parameter values of each model are smaller than 0.05. However, based on the performance of MAE, MSE and RMSE for the out-sample data, as shown in Table 4, SARIMA $(1,1,0)(0,1,1)^{12}$ outperforms SARIMA $(0,1,1)(0,1,1)^{12}$.

Table 4: The performance of the SARIMA models

Models	In-Sample			Out-Sample		
	MAE	**MSE**	**RMSE**	**MAE**	**MSE**	**RMSE**
SARIMA $(0,1,1)(0,1,1)^{12}$	2516.5	10791238	3285.0	3984.1	21876169	4677.1
SARIMA $(1,1,0)(0,1,1)^{12}$	2523.5	10531849	3245.2	3614.1	19152425	4376.3

Models	**MAPE In-Sample sample**	**MAPE Out-sample**
SARIMA $(0,1,1)(0,1,1)^{12}$	9.5243	22.9668
SARIMA $(1,1,0)(0,1,1)^{12}$	9.6307	20.0154

The next step in identifying the best-fit model involves considering the MAPE for both in-sample and out-sample monthly fish landing data. A lower MAPE indicates smaller differences between forecasted and actual data. The result presented in Table 4 show that the in-sample MAPE for the SARIMA $(0,1,1)(0,1,1)^{12}$ model is slightly smaller than that of the SARIMA $(1,1,0)(0,1,1)^{12}$ mode, suggesting that SARIMA $(1,1,0)(0,1,1)^{12}$ is the best fit for in-sample time-series. However, when assessing out-sample performance, SARIMA model $(1,1,0)(0,1,1)^{12}$ outperforms SARIMA $(0,1,1)(0,1,1)^{12}$, with a lower MAPE of 20.0154% as compared with 22.9668%. Consequently, the SARIMA model $(1,1,0)(0,1,1)^{12}$ is identified as the best fit for monthly fish landing data. This model can be expressed as $yt = 0.655yt\text{-}1 + 0.345yt{-}2 + yt{-}12 \quad 0.655yt{-}13 \quad 0.345yt{-}14 + a_t - 0.8606a_{t\text{-}12}$

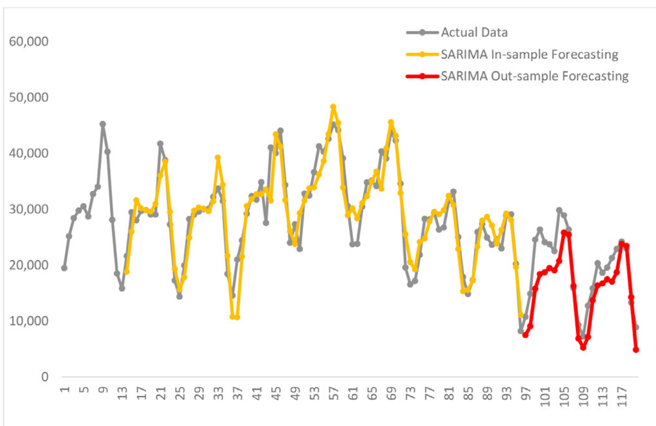

Figure 8: The performance of the SARIMA model for monthly fish landing in East Coast Peninsular Malaysia for the period of January 2019 to December 2021

As illustrated in Figure 8, the in-sample and out-sample forecast data for the SARIMA $(1,1,0)(0,1,1)^{12}$ model closely align with the actual data. The results in Table 4 further support this observation, indicating high accuracy forecasts for in-sample data and good forecasts for out-sample data.

SARIMA-ANN

The selected SARIMA model's input is supplied to the Artificial Neural Network (ANN) model, forming a basic ensemble method. The number of hidden layers is set at the default of 10 hidden nodes. The log-sigmoid function serves as the transfer function in the hidden layer, and the linear transfer function in the output layer. The Levenberg-Marquardt method is employed for the training algorithm. With SARIMA $(1,1,0)(0,1,1)^{12}$, the input variables comprise 5 lag variables: y_{t-1}, y_{t-2}, y_{t-12}, y_{t-13}, and y_{t-14}, serving as input variables for the ANN. The process begins by determining the optimal number of hidden nodes, ranging from 1 to 10. Table 5 presents the results of MAPE for both the training and testing phases for each hidden node.

Table 5: The MAPE testing and training for each number of hidden nodes

Hidden Nodes	MAPE Testing	MAPE Training
1	9.9063	20.2215
2	9.3084	20.3998
3	**8.2532**	**19.3145**
4	6.0891	19.7785
5	6.2138	23.0585
6	4.9967	25.0176
7	4.1164	20.7151
8	3.4314	23.5139
9	1.9511	37.9907
10	1.2116	40.3029

Hidden node 3, with the lowest MAPE value, is considered the optimal outcome. The next step involves determining the best combination of lag variables, ranging from one lag variable to five lag variables.

Table 6 provides a summary of the best performance, and Table 7 further consolidates the results. The optimal number of input nodes is determined to be three (3), representing a combination of lag variables y_{t-12}, y_{t-1}, y_{t-13}. The architecture of these inputs is illustrated in Figure 9.

Table 6: The result of combination of input lag variables

Input Variables	MAPE Training	MAPE Testing
y_{t-1}	16.5389	29.0417
y_{t-2}	23.8731	58.4280
y_{t-12}	**15.4365**	**27.9546**
y_{t-13}	21.1023	47.7769
y_{t-14}	26.0293	66.5311
y_{t-12}, y_{t-1}	**12.2748**	**18.5385**
y_{t-12}, y_{t-2}	14.3494	26.9270
y_{t-12}, y_{t-13}	15.5949	26.9423
y_{t-12}, y_{t-14}	14.5067	25.8596
y_{t-12}, y_{t-1}, y_{t-2}	11.2863	17.8618
y_{t-12}, y_{t-1}, y_{t-13}	**10.0606**	**15.3782**
y_{t-12}, y_{t-1}, y_{t-14}	8.116	19.6086
y_{t-12}, y_{t-1}, y_{t-13}, y_{t-2}	8.4376	18.1713
y_{t-12}, y_{t-1}, y_{t-13}, y_{t-14}	8.2651	21.3415
y_{t-12}, y_{t-1}, y_{t-13}, y_{t-2}, y_{t-14}	**9.1191**	**19.4621**

Table 7: The summary of input lag variables

Input Variables	MAPE Training	MAPE Testing
y_{t-12}	15.4365	27.9546
y_{t-12}, y_{t-1}	12.2748	18.5385
$\mathbf{y_{t-12}, y_{t-1}, y_{t-13}}$	**10.7247**	**15.276**
$y_{t-12}, y_{t-1}, y_{t-13}, y_{t-2}$	8.4376	18.1713
$y_{t-12}, y_{t-1}, y_{t-13}, y_{t-2}, y_{t-14}$	9.1191	19.4621

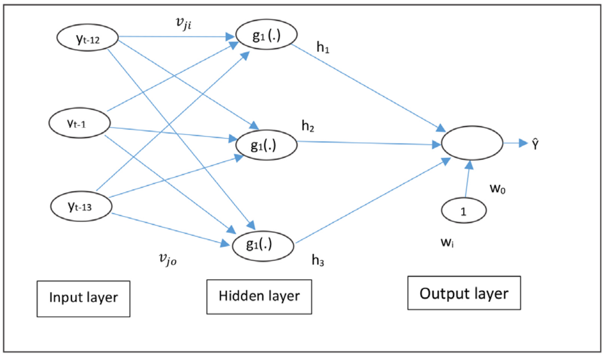

Figure 9: The architecture of the best SARIMA-ANN model with three input nodes, three hidden nodes and one output node, log-sigmoid transfer function in the hidden layers and linear transfer function in the output layer

Discussion

Figure 10 visually demonstrates that the SARIMA-ANN model for monthly fish landing in East Coast Peninsular Malaysia is highly effective as a forecasting model. Both the in-sample forecast data and out-sample forecast data closely align with the actual data. This robust performance is reflected in the in-sample and out-sample MAPE values, both of which are less than 20%.

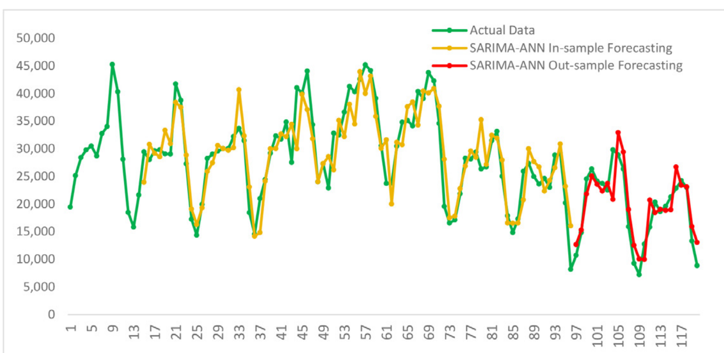

Figure 10: The performance of the SARIMA-ANN model for monthly fish landing on the East Coast in Peninsular Malaysia for the period of January 2019 to December 2021

The results in Table 8 clearly indicate that the application of the SARIMA method alone is not as effective as combining SARIMA with ANN in the SARIMA-ANN method. Statistical evidence supports the conclusion that the integration of SARIMA and ANN demonstrates superior performance compared to using SARIMA alone.

Table 8: A comparison of the performance of SARIMA and SARIMA-ANN

Model	**In-Sample**			**Out-Sample**		
	MAE	**MSE**	**RMSE**	**MAE**	**MSE**	**RMSE**
SARIMA	2,523.5	10,531849	3,245.2	3,614.1	19,152425	4,376.3
SARIMA-ANN	2.82472	1.375353	1.17275	2.47961	9.71628	3.11709

Model	**MAPE Training**	**MAPE Testing**
SARIMA	9.6307	20.0154
SARIMA-ANN	10.7247	15.2760

CONCLUSION

The rising demand for fish sources correlates with the growing human population, exerting constant pressure on marine resources. The challenges and potential limitations in managing these resources are intricate and not easily observable or calculable, adding complexity to the situation. This study aims to develop an optimal model for Monthly Marine Fish Landing in East Coast Peninsular Malaysia using ARIMA and ANN. Two approaches, namely single SARIMA and SARIMA-ANN methods, were compared. Several possible SARIMA models were developed, and accuracy evaluations based on AIC and BIC values identified SARIMA $(1,1,0)(0,1,1)^{12}$ as the best forecasting model. To further enhance model performance, ANN was employed due to its capability to learn and model non-linear time series, enabling the model to generalize and predict unseen data. The out-of-sample MAPE results indicate that SARIMA-ANN outperformed SARIMA in forecasting fish landing in the East Coast Peninsular Malaysia.

This study aims to provide valuable suggestions to stakeholders by highlighting the SARIMA-ANN method as one of the most effective approaches for forecasting single time-series data, particularly monthly fish landing in East Coast Peninsular Malaysia. The utilization of this method can significantly contribute to enhancing managers' confidence in employing a reliable forecasting mechanism for informed decision-making in the near future. Despite the widespread recognition of the importance of fish stock management information in fisheries management, there remains a gap in the implementation of statistical techniques by responsible agencies to support decision-making on fish stock management. Inefficient management of fish stocks may result in fish shortages, especially during low landing seasons, subsequently causing an increase in fish prices. Therefore, agencies, such as LKIM, must efficiently identify the amount of fish stock to be kept frozen to stabilise the supply and prevent disruptions in the market.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

REFERENCES

[1] Yaakob, O., & Chau, Q. P. (2005). Weather downtime and its effect on fishing operation in Peninsular Malaysia. *Jurnal Teknologi*, *42*(A), 13-26. http://dx.doi.org/10.11113/jt.v42.730

[2] Koutroumanidis, T., Iliadis, L., & Sylaios, G. K. (2006). Time-series modeling of fishery landings using ARIMA models and Fuzzy Expected Intervals software. *Environmental Modelling & Software*, *21*(12), 1711-1721.

[3] Shabri, A., & Samsudin, R. (2015). Fishery landing forecasting using wavelet-based autoregressive integrated moving average models. *Mathematical Problems in Engineering*, *2015*(SI), 1-9. https://doi.org/10.1155/2015/969450

[4] Begg, G. A., Friedland, K. D., & Pearce, J. B. (1999). Stock identification and its role in stock assessment and fisheries management: an overview. *Fisheries Research*, *43*(1-3), 1-8. https://doi.org/10.1016/S0165-7836(99)00062-4

[5] Teh, L. S. L., Zeller, D., Cabanban, A., Teh, L. C. L., & Sumaila, U. R. (2007). Seasonality and historic trends in the reef fisheries of Pulau Banggi, Sabah, Malaysia. *Coral Reefs*, *26*(2), 251-263. http://dx.doi.org/10.1007/s00338-006-0182-x

[6] Raman, R., Sathianandan, T. V., Sharma A. P., & Mohanty, B. P. (2017). Modelling and forecasting marine fish production in Odisha using seasonal ARIMA model. *National Academy Science Letters*, *40*(6), 393–397. https://doi.org/10.1007/s40009-017-0581-2

[7] Anuja, A., Yadav, V. K., Bharti, V. S., Kumar, N. R. (2017). Trends in marine fish production in Tamil Nadu using regression and autoregressive integrated moving average (ARIMA) model. *Journal of Applied and Natural Science*, *9*(2), 653–657. http://dx.doi.org/10.31018/jans.v9i2.1252

[8] Farhan, J., & Ong, G. P. (2018). Forecasting seasonal container throughput at international ports using SARIMA models. *Maritime Economics & Logistics*, *20*, 131-148.

[9] Bako, H. Y., Rusiman, M. S., Kane, I. L., & Matias-Peralta, H. M. (2013). Predictive modeling of pelagic fish catch in Malaysia using seasonal ARIMA models. *Agriculture, Forestry and Fisheries*, *2*(3), 136-140.

[10] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control, 19*, 716–723. http://dx.doi.org/10.1109/TAC.1974.1100705

[11] Otu, O. A., Osuji, G. A., Opara, J., Mbachu, H. I., & Iheagwara, A. I. (2014). Application of Sarima models in modelling and forecasting Nigeria's inflation rates. *American Journal of Applied Mathematics and Statistics*, *2*(1), 16-28.

[12] Dongare, A. D., Kharde, R. R., & Kachare, A. D. (2012). Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology* (IJEIT), *2*(1), 189-194.

[13] Haykin, S. S. (2009). *Neural networks and learning machines.* New York: Prentice Hall/ Pearson.

[14] Lewis, C. D. (1982). *Industrial and business forecasting methods*. London: Butterworth.

[15] Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusion and implications. *International Journal of Forecasting*, *16*, 451-476.