Fisheries forecasting, physical approach comparison between regression and autoregressive integrated moving average (ARIMA)

T E Ahmad1, A Rais1, D R Azhari1, A A Rosalia1 and R Y F Hutapea2

1Marine Information System, Indonesia University of Education, Dr. Setiabudi Avenue No.229, Isola, Bandung, 40154, Indonesia.
2Dumai Marine and Fisheries Polytechnic, Jl. Wan Amir No. 1, Kota Dumai, Riau, Indonesia

Corresponding author: taufiqejazahmad@upi.edu

Abstract. Empowerment of fishery resources requires analysis of forecasting results as an effort to maintain sustainability and human needs. Forecasting is an approach to predicting based on past facts, which is expected to be used as a decision support system. However, there are problems with the accuracy. In this study, we compare the regression method with ARIMA to find out which method can be used as the most appropriate choice in fisheries forecasting with physical benchmarks (seasonal and climate). We conduct a systematic literature study on various studies with the theme of fisheries forecasting. The search focuses on studies with the main criteria in the form of an explicit discussion of the basic forecasting methods and literature on marine physical influences. Then define a search method by combining fishery stocks or landings against forecasting with marine physical features optionally using PRISMA. The results show that the ARIMA models have a better fair value and accuracy than the regression because the ARIMA model captures the history of data autocorrelation and extrapolates it to the forecasting framework that will be carried out. Hence, it is most suitable for use with additional marine physical variability.

Keywords: ARIMA; Forecasting; PRISMA; Regression

1. Introduction
Empowerment of fisheries resources is all activities that make fisheries a resource, including management or capture activities [1]. The empowerment of fishery resources is based on market needs, which leads to processing and consumption needs [2]. In fisheries resources, the influence of environment and management are the main factors in the availability of fishery resources [3,4].

The empowerment of fishery resources is increasing rapidly in line with the growth of the human population globally or regionally [2]. Ideally, every resource empowerment will coexist with good management to support sustainability [4,5]. There are many benefits to conducting management that considers ecosystems and sustainability a minimum standard orientation [3]. Business perspective, management is an attempt to increase the period of operational period of fishing operations while at the same time ensuring the availability of resources as the primary raw material. Ecologically, this is a step to avoid scarcity, leading to an imbalance of nature [4–6].

Apart from the benefits provided, fisheries management must be supported by various decision support systems. One of which is an estimate to forecast future availability, this is aimed at increasing the carrying capacity to produce the right decisions. However, the inaccuracy of the forecasting carried out is commonplace. This is a representation of the seriousness of data collection and processing to the commitment to consider ecosystems and sustainability [4].

Fishery resources come from an environment with specific physical characteristics. These physical factors play a major role as variables that affect the existence of a fishery resource [5]. Thus marine
physical variables can be measured and projected to support forecasting. In addition, high-accuracy forecasting could be the main answer to increase the carrying capacity for accurate decisions. Apart from the accuracy of data sources that adequately represent reality to the commitment to use forecasting results, various studies have shown how this forecasting has always experienced significant development [6,7]. Starting from the simplest to the most advanced forecasting process where physical variables support fish living ecosystems in an open environment are included to represent the state of reality in the calculations used [7].

2. Methods
For a comprehensive understanding, we systematically conducted literature studies and literature reviews on various studies of fisheries forecasting. This research focuses on studies with the main criteria explicitly discussing the basic forecasting methods and literature on marine physical influences.
Then defined the search method by optionally combined fishery stocks or fish landing with the forecasts with marine physical features. With these advantages coupled with wide acceptance, we used PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [8]. The literature search was carried out on Google Scholar (figure 1), with a publication range between 2000 to 2020 with keyword criteria:

1. Fisheries AND (stock OR catch)
2. Statistic AND forecasting AND (accuracy OR assessment)
3. Forecasting AND (physical OR seasonal OR climate)
4. (Fisheries OR catch OR stock) AND (forecasting OR physical OR seasonal OR climate)

Fisheries forecasting techniques

In general, forecasting in capture fisheries estimated the amount of fishery availability in the environment for the future [9]. The calculation assumptions were taken from the number of catches to various environmental variables [4]. In general, the forecasting could be divided into two-stage: First we must Mapping the data used, any differences in characteristics and variability will result in a different representation. And then we need to determine the forecasting method, the suitability of the variability and character of the data held with the method used is the key to the expected accuracy of the data.

After the initialised two stages, each will entered each processing with different processing techniques. In forecasting, how confident or accurate the forecasting has been a game-changer [10]. As is well known, marine monitoring is the most expensive monitoring activity [8, 11, 12]. This lack was the main factor in raising the value of assumptions in various forecasting methods. This lack was not strange and familiar, considering that fish was still an unmeasurable resource, so forecasting methods would continue to develop to meet needs and reduce uncertainty.

2.1. Regression

Regression is a statistical method divided into two: regression, which was commonly used for forecasting, and regression, which is commonly used to determine the relationship between variables [13]. In general, linear regression is divided into six simple linear regression, multiple linear, logistic, ordinal, multinominal, and discriminant analyses [14]. However, linear regression is an approach that could only be used to analyse one variable.

In fisheries forecasting, techniques with linear regression are quite rare. This was because the number of variables that can be used is not more than one. This is contrary to the many environmental variables that could affect fishery resources. Which is highly recommended to build knowledge of any factors affecting fishery resources and include these variables in the calculations [15–18]. Several variables can be used, and the most suitable regression method for using more than one variable is multiple linear regression (MLR) [20,21]. MLR is a statistical method for predicting the main and explanatory variables by modelling the linearity of the relationship between the x (independent) and y (bound) variables [19]. With the basic MLR formula:

\[
\beta = \beta_0 + \beta_1x_1 + \cdots + \beta_mx_m + \varepsilon
\]  

(1)

With matrix equations:

\[
\beta = (X^TX)^{-1}X^Ty \text{ where } \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{12} & \cdots & X_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nm} \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}
\]

(2)

2.2. Auto-regressive integrated moving average

In various recent literature, many forecasting models estimated the number of fish catch per unit and fish stocks, such as regression, ARIMA model, fuzzy method, and Artificial Neural Networks (ANN). ARIMA was a forecasting approach that was most in demand by various researchers, generally in analysing scholastic time series modelling. This method was developed and can be used to answer the
need to increase the accuracy of fisheries forecasting with its statistical advantages that could provide systematic searches at every stage, including (identification, estimation, and inspection). In this paper, the author only discussed the ARIMA model, which was used to perform forecasting quickly. It was intended to meet the accuracy needs that were built on the ARIMA model itself. With the ARIMA model formula as follows [20]:

\[
(1 - n_1 B^p)(1 - N_1 B^p)(1 - B^d)(1 - B^q)X_t = (1 - u_1 B^q)(1 - U_1 B^q)e_t
\]  

(3)

Where:
- \(X_t\) = the value at time \(t\)
- \(B^p\) = backwards shift operator for which \(B^p X = X_{t=p}\)
- \(n_1, N_1, u_1, and U_1\) = arithmetic coefficients
- \(e_t\) = error term at time \(t\)

Furthermore, it was referred to as:

\[
ARIMA(p, d, q)(P, D, Q)^S
\]  

(4)

\(p\) = order of autoregressive term (AR term)
\(d\) = degree of differencing involved to achieve stationarity (1 term)
\(q\) = order of moving average term (MA term)
\(S\) = seasonality (number of periods per season)
\(P, D, Q\) = seasonal terms corresponding to \(p, d, q\) respectively

The auto-regressive integrated moving average was a model built assuming that the combination of time series was linear and stationary. This linearity refers to the error value [21]. This means that this model could not tolerate linear and stationary changes in variables in periodic variations in the time-series observations. This limitation then becomes the ARIMA model, which was generally per season [22–25] and was based on other parameters that had variable determinations in them.

This model is widely relied on in fisheries forecasting due to the capability to forecast a species with different behaviour and biological characteristics. The ARIMA model had an immense superiority compared to the forecasting method with regression [21]. The ARIMA model had undergone various developments according to needs, including the accuracy improvisation proposed by Ani Shabri [26]. The ARIMA combined empirical mode decomposition with clusterisation. Then ARIMA model was applied to capture fisheries forecasting with data objects in the form of the number of catches landed. The results which reduce the amount of error in the root mean square error (RMSE), mean absolute error (MAE) and \(r^2\).

Then Aquino et al. [21] combined ARIMA and ANN models to predict fish landing volumes with improved forecasting quality. Improvisation of the ARIMA model was still fundamental shortcomings that cannot tolerate the presence of non-linear and non-stationary variables. Which are certainly not following the reality of fish habitats that play a significant role in changes in landing quantity and fish stock quantity in nature.

2.3. Physical forecasting extension

The development of forecasting is the development of calculation methods based on statistical science. So far, no forecasting model has really been developed for the specific purpose of forecasting with limited variables. All forecasting methods allowed the use of any variable. Its characteristics were needed to meet the needs of accuracy. The urgency of the involvement of variables that represented the actual situation presented an extension option to add a separate extension function into the appropriate forecasting method. General fishing variables such as the number of catches can be juxtaposed with variables that affect the environment.
2.3.1. Seasonal forecasting. Linearity and stationarity of variables that provided limited variability of forecasting time do reduce the percentage of actual reality. However, it should be realised that eliminating time constraints means increasing complexity and could even add objects of observation to predict the situation. As a result, the non-linear and non-stationary natural factors must be predicted to improve and provides an accuracy advantage over calculations that rely on linear and stationary variables. Considered that natural changes could not be predicted with certainty, it was likely to reduce accuracy if forecasting is based on variables that were not sufficiently representative of the actual situation. Therefore, a more accurate forecast remains based on seasonal forecasting, which did not require the variability of marine physical forecasting. However, this could be overcome by using reliable forecasting data from marine physics. Thus, a field fact could not be ignored, given the importance of using controllable and measurable variables in the forecasting process [31–33].

Seasonal forecasting was a forecasting approach with seasonal marine physical variables. Forecasting variables included temperature (sea temperature and air temperature), salinity (rain), and other factors tailored to the needs. The selected two main variables, namely temperature and salinity, are based on the impact of the growth rate on the predicted fish objects [30-31].

This forecasting was the development of an existing forecasting method, with extensive properties built from the dynamics of the ocean's physical properties. By its environment, accuracy was still influenced by the statistical method used. Thus, it was possible to compare this seasonal forecasting with the other autoregressive model [36-38]. With the presence of a field approach, this extensional method was a process that improved analytical performance. It could be helpful from a business perspective with its ability to reduce risk.

2.3.2. Climate forecasting. Climate forecasting was an external forecasting method that could provide long-term measurable information. The physical properties of the sea are divided into seasonal and climate-based, both of which were marine physical variables that had similarities and trends with their respective characteristics. The differences would be linear with time units, and information details could be adjusted to the scope of the objectives of forecasting time. Forecasting with a long time could use climate forecasting to describe the situation roughly. So that a thorough knowledge of climate change was necessary because limited knowledge would lead to wrong forecasts, especially deliberately ignoring a variable in climate [35].

Climate change could be used as a forecasting variable in marine physics due to the impacts given in the form of changes in temperature, wind speed, and direction, accumulation of CO₂ to increase in ocean acidity [36-37]. This factor was a physical variable that had a powerful influence on the availability of fish globally or in a particular aquatic environment [38]. With more measurable, linear, and stationary changes, climate forecasting is present as a solution for long-term forecasting that demands sustainability values [39].

3. Result and discussion

literature identification of studies via google scholar. Then we separated the non-English literature and sorted it by relevance, so we have found 1000 articles. In the second sorting, we screened the relevance of titles, abstracts, keywords, duplicate articles, and articles with a publication year after 2011. So that in this second selection, 443 articles were obtained for full-text reading and analysis. Publications that produce certain results but do not include analysis of the models and variables that influence and are used to obtain these results are also included in excluded articles. Only 78 publications (7.8% of all articles screened) support the analysis results in this paper.

ARIMA is a time series forecasting method requiring data to meet linearity and stationarity, which can technically predict historical data and have high accuracy in time series forecasting. However, if ARIMA is used for an extended period, it will produce flat forecasting. In contrast, linear regression is used to determine the effect of one or several variables—linear regression analyses by forming a linear equation and then using the equation to predict. However, linear regression cannot show the saturation point of the function on a variable, so there is an error in making a forecast and the possibility of
multicollinearity in the independent variables, which consequently the independent variables are unable to explain the dependent variables.

Each model of the forecasting approach in this study has its unique value. Then, is the prediction using forecasting with a correlation between variables. Despite many recent advances, fisheries science still does not recognise an effective way to identify appropriate conditioning options, so the probability distribution calculated for management purposes is not sufficiently representative of the actual probabilities in the environment [35,40].

The MLR approach has a high fitting value in fitting models because the variability used in the modelling is explicit. The MLR is inversely proportional to ARIMA, which is a dynamic and hyper-regressive approach. In contrast, the approach uses physical variability, namely seasonal and climate, not so relevant to the fitting models needed. Generally, the three methods are part of the variability that is an extension of the ARIMA model. In contrast, in terms of determining accuracy, the ARIMA approach is far superior to the MLR method because the ARIMA model captures the history of autocorrelation of data and extrapolates it to the forecasting framework that will be carried out. as a result the ARIMA approach to usually outperform other methods because other methods do not have the ability and value to measure autocorrelation and extrapolate it like the ARIMA model. The ARIMA model approach assumes that the time series is a linear combination of its past value and the current and past values of the error term, unlike the MLR method, which only takes the time series value as an independent variable entity.

4. Conclusion

Forecasting that only relies on correlations between variables is not enough to represent fishery stocks in real terms, especially methods that rely on correlation are not suitable for forecasting time series. So, that when compared to MLR with ARIMA, the ARIMA model is very superior in terms of input variability, dynamics, and accuracy.

Therefore, it can be concluded that increased focus should be placed on the careful examination and examination of decisions made during the analysis. More attention should be paid to examining the sensitivity of models or forecasting methods to alternative assumptions such as ocean physical variability and model structure. An approach to uncertainty in forecasting fishery values must consider the sensitivity and use the method or model with the most effective approach. The use of decision tables or simulation of management procedures so that testing can be carried out correctly when it will be applied to support decisions. In addition, it is necessary to know the fair value and accuracy of the correlation approach that will be used in forecasting. Because of that it is very crucial to get optimal forecasting results with high fitting value and good accuracy.

References

[1] Villareal L V 2004 Guidelines on the collection of demographic and socio-economic information on fishing communities for use in coastal and aquatic resources management (Food & Agriculture Org)

[2] RUDDLE K, HVIDING E and JOHANNES R E 1992 Marine Resources Management in the Context of Customary Tenure Mar. Resour. Econ. 7 249–73.

[3] Cummins A 2004 The Marine Stewardship Council: A multi-stakeholder approach to sustainable fishing Corp. Soc. Responsib. Environ. Manag. 11 85–94.

[4] Gaines S D and Costello C 2013 Forecasting fisheries collapse Proc. Natl. Acad. Sci. U. S. A. 110 15859–60.

[5] Barnuevo A and Asaeda T 2018 Integrating the ecophysiology and biochemical stress indicators into the paradigm of mangrove ecology and a rehabilitation blueprint PLoS One 13 e0202227.

[6] Walsh R G, John K H and McKeen J R 1989 Comparing long-run forecasts of demand for fish and wildlife recreation Leis. Sci. 11 337–51.

[7] Rodhouse P G 2001 Managing and forecasting squid fisheries in variable environments Fish. Res. 54 3–8.
[8] Doughty C L and Cavanaugh K C 2019 Mapping Coastal Wetland Biomass from High Resolution Unmanned Aerial Vehicle (UAV) Imagery Remote Sens. 11.

[9] Sun L, Xiao H, Li S and Yang D 2009 Forecasting fish stock recruitment and planning optimal harvesting strategies by using neural network J. Comput. 4 1075–82.

[10] Patterson K, Cook R, Darby C, Gavaris S, Kell L, Lewy P, Mesnil B, Punt A, Restrepo V, Skagen D W and Stefansson G 2001 Estimating uncertainty in fish stock assessment and forecasting Fish Fish. 2 125–57.

[11] Mannino A M and Balistreri P 2018 Citizen science: a successful tool for monitoring invasive alien species (IAS) in Marine Protected Areas The case study of the Egadi Islands MPA (Tyrrhenian Sea, Italy) Biodiversity 19 42–8.

[12] LAXON S 1994 Sea ice altimeter processing scheme at the EODC Int. J. Remote Sens. 15 915–24.

[13] Wu J, Liu C, Cui W and Zhang Y 2019 Personalized Collaborative Filtering Recommendation Algorithm based on Linear Regression 2019 IEEE International Conference on Power Data Science (ICPDS) pp 139–42.

[14] WEISBERG S 2005 Applied Linear Regression ed U of Minnesota (Minnesota: John Willey & Sons)

[15] Memarzadeh M, Britten G L, Worm B and Boettiger C 2019 Rebuilding global fisheries under uncertainty Proc. Natl. Acad. Sci. 116 15985 LP – 15990

[16] Privitera-Johnson K M and Punt A E 2020 Leveraging scientific uncertainty in fisheries management for estimating among-assessment variation in overfishing limits ICES J. Mar. Sci. 77 515–26.

[17] Rosenberg A A and Restrepo V R 1994 Uncertainty and risk evaluation in stock assessment advice for U.S. marine fisheries Can. J. Fish. Aquat. Sci. 51 2715–20.

[18] Caddy J F and McGarvey R 1996 Targets or Limits for Management of Fisheries? North Am. J. Fish. Manag. 16 479–87.

[19] Zhang Z, Li Y, Li L, Li Z and Liu S 2019 Multiple Linear Regression for High Efficiency Video Intra Coding ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp 1832–6.

[20] McLeod A I and Sales P R H 1983 Algorithm AS 191: An Algorithm for Approximate Likelihood Calculation of ARMA and Seasonal ARMA Models J. R. Stat. Soc. Ser. C Applied Stat. 32 211–23.

[21] Stergiou K I 1991 Short-term fisheries forecasting: comparison of smoothing, ARIMA and regression techniques J. Appl. Ichthyol. 7 193–204.

[22] Al-Gounmeein R S and Ismail M T 2020 Forecasting the exchange rate of the Jordanian dinar versus the US dollar using a Box-Jenkins seasonal arima model Int. J. Math. Comput. Sci. 15 27–40.

[23] Noureen S, Atique S, Roy V and Bayne S 2019 Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A case study of small scale agricultural load Midwest Symp. Circuits Syst. 2019-August 521–4.

[24] Raman R K, Sathianandan T V, Sharma A P and Mohanty B P 2017 Modelling and Forecasting Marine Fish Production in Odisha Using Seasonal ARIMA Model Natl. Acad. Sci. Lett. 40 393–7.

[25] Oliveira P J, Steffen J L and Cheung P 2017 Parameter Estimation of Seasonal Arima Models for Water Demand Forecasting Using the Harmony Search Algorithm Procedia Eng. 186 177–85.

[26] Maulud D and Abdulazeex A M 2020 A Review on Linear Regression Comprehensive in Machine Learning J. Appl. Sci. Technol. Trends 1 140–7.

[27] Aqueveque P E, Wiechmann E P, Herrera J and Pino E J 2015 Measurable Variables in Copper Electrowinning and Their Relevance to Predicting Process Performance IEEE Trans. Ind. Appl. 51 2607–14.
[28] Shmueli G 2010 To Explain or to Predict? Stat. Sci. 25 289–310.
[29] He L-M (Lee) and He Z-L. 2008 Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in southern California, USA Water Res. 42 2563–73.
[30] Neuheimer A B, Thresher R E, Lyle J M and Semmens J M 2011 Tolerance limit for fish growth exceeded by warming waters Nat. Clim. Chang. 1 110–3.
[31] Thresher R E, Koslow J A, Morison A K and Smith D C 2007 Depth-mediated reversal of the effects of climate change on long-term growth rates of exploited marine fish Proc. Natl. Acad. Sci. 104 7461 LP – 7465.
[32] Alves O, Balmaseda M A, Anderson D and Stockdale T 2004 Sensitivity of dynamical seasonal forecasts to ocean initial conditions Q. J. R. Meteorol. Soc. 130 647–67.
[33] Spillman C M 2011 Advances in Forecasting Coral Bleaching Conditions for Reef Management Bull. Am. Meteorol. Soc. 92 1586–91.
[34] Osborn D R and Smith J P 1989 The Performance of Periodic Autoregressive Models in Forecasting Seasonal U.K. Consumption J. Bus. Econ. Stat. 7 117–27.
[35] Broad K, Pfaff A S P and Glantz M H 2002 Effective and Equitable Dissemination of Seasonal-to-Interannual Climate Forecasts: Policy Implications from the Peruvian Fishery during El Niño 1997–98 Clim. Change 54 415–38.
[36] Metz B, Change. I P on C and III. W G 2007 Climate change 2007: mitigation of climate change: contribution of Working Group III to the Fourth assessment report of the Intergovernmental Panel on Climate Change (Cambridge: Cambridge University)
[37] Doney S C, Busch D S, Cooley S R and Kroeker K J 2020 The Impacts of Ocean Acidification on Marine Ecosystems and Reliant Human Communities Annu. Rev. Environ. Resour. 45 83–112.
[38] Hollowed A B, Barange M, Ito S, Kim S, Loeng H and Peck M A 2011 Effects of climate change on fish and fisheries: forecasting impacts, assessing ecosystem responses, and evaluating management strategies ICES J. Mar. Sci. 68 984–5.
[39] Schmidt J O, Bograd S J, Arrizabalaga H, Azvedo J L, Barbeaux S J, Barth J A, Boyer T, Brodie S, Cárdenas J J, Cross S, Druon J N, Fransson A, Hartog J, Hazen E L, Hobday A, Jacox M, Karstensen J, Kupchus S, Lopez J, Madureira L A S P, Martinelli Filho J E, Miloslavich P, Santos C P, Scales K, Speich S, Sullivan M B, Szoboszlai A, Tommasi D, Wallace D, Zador S and Zawislak P A 2019 Future Ocean Observations to Connect Climate, Fisheries and Marine Ecosystems Front. Mar. Sci. 6 1–18.
[40] Tommasi D, Stock C A, Alexander M A, Yang X, Rosati A and Vecchi G A 2017 Multi-annual Climate predictions for fisheries: An assessment of skill of sea surface temperature forecasts for large marine ecosystems Front. Mar. Sci. 4 1–13.