Social Science and Humanities Journal
Everant Journal
ISSN: 2456-2653
Volume 09 Issue 01 January 2025

https://sshjournal.com/
Impact Factor: 2024: 6.576
2023: 5.731
DOI: https://doi.org/10.18535/sshj.v9i01.1437

Forecasting Sea Cucumber Catches Using the ARIMA Method

Maria Oktaviani Jehanus[1,a] | Lusi Mei Cahya Wulandari[2,b] | Agrienta Bellanov[3,c]

1,2,3 Industrial Engineering, Darma Cendika Catholic University, Jl. Dr. Ir. H. Soekarno No.201, Klampis Ngasem, Sukolilo, Surabaya, East Java 60117
[a]maria.oktaviani@student.ukdc.ac.id, [b]lusi.mei19@gmail*.com, [c] agrientabellanov@ukdc.ac.id

***Received* 10-12-2024**
***Revised* 11-12-2024**
***Accepted* 02-01-2025**
***Published* 04-01-2025**

Abstract:

Sea cucumbers are marine resources that have a significant ecological role and important economic value. UD. Matahari Jalan Sukolilo Baru II, Bulak District, Surabaya is an MSME that processes sea cucumber catches and sells them to various distributors, including trading abroad. The catch of sea cucumbers obtained by fishermen in uncertain quantities, every subsequent period. Therefore, the results of sea cucumber fishing are known to be influenced by several factors, one of which is climatic factors such as temperature, humidity and tides. The research was conducted to predict the uncertain catch of sea cucumbers with the ARIMA method to obtain effective modeling and equations. Forecasting can help determine the right period by using one of the methods that correspond to the sequence of time. The ARIMA method is an approach used in time series analysis to model and forecast data arranged in a specific order. Predict the catch of sea cucumbers by looking at the smallest error, and the catch of sea cucumbers after forecasting in the next period. The result of the selection of the best ARIMA model from the humidity variable is (1,1,1) more significant and effective for sea cucumber fishing in the short term (1.2 days) in the rainy season, with the smallest error of 271.11. The forecast results in April 2024 for the 107 period are 268.42 kg, the forecast data is close to the actual data in the previous period.

Keywords: Forecasting; ARIMA; Sea Cucumber.

Introduction:

In the Kenjeran Surabaya Tourism Area, fisheries are used as an effort to provide income for the community. Fishery products are utilized and processed through the manufacturing process into materials that are ready for public consumption. One of them, the fishing village in Sukolilo, precisely in UD. Matahari Jalan Sukolilo Baru II, Bulak District, Surabaya is an MSME that processes sea cucumber catches and sells them to various distributors, including being traded abroad. For sea cucumber processing, it is divided into 2 methods, namely, sea cucumbers are left wet, processed through boiling and sea cucumbers that are left dry are processed through drying stages and mixed with salt. Sea cucumber products are still traditionally taken using the Javanese calendar with two rainy and dry seasons. The Javanese calendar is a traditional calendar system used in Java, Indonesia. The rainy season (wetan) usually starts in October and lasts until April, while the dry season usually starts from May to September. Production management is erratic and does not have a target production schedule, so it experiences uncertain activities at work such as 4 people, 6 people, and 10 people working. The production of sea cucumber products faces many challenges, especially in the results of the catch and uncertain production scheduling. In the Madura area, the MSMEs have 40 fishermen and suppliers as well as export businesses such as Malaysia and China.

The results of sea cucumber fishing are influenced by several factors, one of which is climatic factors such as temperature, humidity and tides. In certain months, few workers work, because they follow the results of sea cucumber catches which can change. Forecasting can help determine the right period by using one of the methods that correspond to the sequence of time. Forecasting is the process of estimating the needs of goods and services such as the quantity, quality, time, and location needed to meet demand [1]. Forecasting can help determine demand needs for the future, and things that must be taken into account with rapidly changing consumer desires environmental circumstances that make it difficult for organizations to make decisions about production levels [2].

The Autoregressive Integrated Moving Average (ARIMA) is one of the many methods that can be used to minimize problems occurring. The ARIMA method is an approach used in time series analysis to model and forecast data arranged in a specific order [3]. Methods with an iterative approach are used to find the most appropriate model of all available models with different types of data. In its application, time series models can often be used easily for prediction because future estimates are based on the value of previous variables [4]. The advantages of the ARIMA method are that it is very *flexible* (follows data patterns), is very accurate in its forecasting, and is suitable for predicting many factors quickly and easily only requiring historical data [5]. This is in line with the findings of previous research [6] that the use of the ARIMA method can help forecast the production of cardinal *banggai fish*. This study aims to predict uncertain sea cucumber catches, effective modeling and equations through the ARIMA method in predicting sea cucumber catches by looking at the smallest errors, and sea cucumber catches after forecasting in the next period.

Literature Reference:

Time Series Analysis

Time series analysis is a statistical technique used to look at data taken over time to find patterns, trends, and seasonal or cyclical fluctuations [7]. The data is taken in a stationary state and the fluctuation is close to the average value of constant variance to form a time series model. The Box-Cox Transformation, which is written as follows, is an alternative to stationing non-stationary data.

$$T(Z_t) = \frac{z_t^{\lambda}-1}{\lambda} \quad (1)$$

Description:

$T(Z_t)$: The transformation function of the time series value *Zt*. It is the result obtained from applying the transformation to the original time series data *Zt*.

Z_t : The value of the time series data at time t. This represents the raw data collected or measured at a specific time point.

λ : The transformation parameter that determines the shape of the transformation function. λ is often used to adjust the data to make it fit better with the assumptions of statistical models, such as normality or linearity.

The Box-Cox *transform forms* for some estimated values **λ** [8].

Table 1. Transformasion *Box-Cox*

	Transformation
-1	$1/Z_t$
-0,5	$1/\sqrt{Z_t}$
0	$L_n\ Z_t$
0,5	$\sqrt{Z_t}$
1	Z_t (after transformation)

The process of *differencing* the d-order is carried out on data that is not stationary in average. This process is described as follows.

$$W_t = (1 - B)^d Z_t \quad (2)$$

Description:

W_t : Representing the transformation of the original time series Z_t after differentiation

B : The backshift operator (lag operator) where B Zt = Zt−1. This means that B shift the data Zt to the previous period

$(1 - B)^d$:The differencing operator. For d=1, this is equivalent to (1−B) Zt=Z_t−Z_{t-1}, which calculates the difference between the value at time t and the previous time t−1

1. *Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)*

A stationary process (Zt) has a constant mean and a constant variance for all times t.

$$E\,(Z_t) = E\,(Z_{t-k}) = \mu$$

$$Var\,(Z_t) = E[(Z_t - \mu)^2] = E[(Z_{t-k} - \mu)^2] = \sigma_Z^2$$

Description:

$E\,(Z_t)$: The expectation or mean of Zt, which is the expected value of the variable Zt

$Var\,(Z_t)$: How far the data Zt spreads from its mean (μ). Variance is a measure of the dispersion or variability of the data around its mean.

σ_Z^2 : The constant variance of the time series Zt, which is the same for all t and t−k, indicating that this process is stationary.

With *the covariance* between Z_t and Z_{t-k} separated by the time interval is expressed as follows

$$\gamma_k = cov\,(Z_t.Z_{t-k}) = E[(Z_t - \mu)(\,Z_{t-k} - \mu)]$$

Description:

$cov(Z_t.Z_{t-k})$: The covariance between Zt (the value of the time series at time t) and Zt and Z_{t-k} (the value of the time series at time t−k). Covariance measures the linear relationship between two variables

The Autocorrelation Function (ACF) shows the constant between and, expressed as follows

$$\rho_k = \frac{cov(Z_t.Z_{t-k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t-k})}} = \frac{\gamma_k}{\gamma_o}$$

Description:

$\sqrt{Var(Z_t)}\sqrt{Var(Z_{t-k})}$: The standard deviation of the values Zt and Z_{t-k}. Standard deviation is the square root of the variance and indicates how spread out the values are around their mean.

γ_k : This measures the covariance between Zt and Z_{t-k}, and is a metric used in autocorrelation analysis.

γ_o : This is the variance of the time series Var (Zt).

ρ_k : The autocorrelation coefficient indicates the linear relationship between the value of Zt at time t and the value of Zt and Z_{t-k} at a previous time (with lag k).

The Partial Autocorrelation Function (PACF) is used to measure *the autocorrelation* between Z_t and Z_{t-k}, where the influence of Z_{t+1}, Z_{t+2} and Z_{t+k-1} has been eliminated. The following equation shows

$$\rho_k = \frac{cov[(Z_t - \hat{z}_t)(Z_{t-k} - \hat{z}_{t+k})]}{\sqrt{Var(Z_t - \hat{z}_t)}\sqrt{Z_{t-k} - \hat{z}_{t+k})}}$$

ARIMA

The *Autoregressive Integrated Moving Average* (ARIMA) method known as *the Box-Jenkin* method was developed by George Box and Gwilym Jenkins in 1970 [9]. ARIMA *Autoregressive Integrated Moving Average* Pattern is one of the time series analysis methods (*Time series analysis*) used to make forecasts or predictions on data that show a certain pattern. The ARIMA value is calculated with the value *of the* current and previous dependent variable to make accurate short-term estimates [10].

Modified ARIMA, which combines operator estimates as an initial prediction with temperature and load data in *a multi-variable* regression process [11]. The ARIMA model groups are as follows:

1. *Autoregresif* (SE)

The Autoregressive (AR) model is regressed to the value of the previous variable, resulting in the current period data which is influenced by the previous period data. *Autoregresive* models with the order p abbreviated to AR (p) or ARIMA (p,0,0) [12].

Pattern:

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_P Z_{t-p} - a_t$$

2. *Moving Average* (MA)

The *Moving Average* (MA) model was first introduced by Slutzky in 1973, with the order q written MA (q) or ARIMA (0,0,q) and developed by Wadsworth in 1989.

Pattern:

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

Description:

Z_{t-p} : Independent variable

ϕ_P : Autoregressive parameter coefficient of order p

a_t : Residual at time t

θ_q : Moving average parameter coefficient of order q

3. *Autoregressive Integrated Moving Average* (ARIMA)

The Autoregressive Integrated Moving Average model assumes that the time series data used must be stationary, which means that the average variation of the data is constant. The Autoregressive (AR), Moving Average (MA), and Autoregressive Moving Average (ARMA) models cannot explain the meaning of the difference, and a mixed model called the Autoregressive Integrated Moving Average (ARIMA) or ARIMA (p,d,q) is used to provide a more effective explanation of differencing. This mixed model consists of a linear function of the previous stationary values, as well as the current values and the previous stationary errors.

Pattern:

$$\Phi_p(B)\, D^d Z_t = \mu + \theta_p(B)\, a_t$$

Φ_p : Coeffisien parameter autoregressive ke-p

θ_p : The coefficient of the moving average parameter to q

B : operator backshift

D : differencing

μ : Constant

a_t : remainder at the the moment

p : autoregressive drift

d : differencing process rate

q : degree moving average parameter autoregressive ke-p

Result and Discussion:

1. Data Collection

This data collection was obtained from the results of UD observations. The data used in this study consists of primary and secondary data. Primary data was collected directly from UD. Matahari Sukolilo, providing sea cucumber catch data, and from BMKG (Meteorology, Climatology, and Geophysics Agency), supplying temperature and weather data. The catch data is essential for analyzing catch volume variations based on environmental factors, while BMKG's data helps assess the impact of weather on catch results.

Secondary data was obtained from various sources, including previous studies, books, and journals, offering theoretical and methodological context. By integrating specific primary data with broader secondary sources, this research ensures a comprehensive approach to forecasting sea cucumber catches, enhancing model accuracy and relevance.

The sun produced by sea cucumbers on Surabaya. This observation and data collection was carried out in April 2023. Data collection obtained for 5 months from sea cucumber catches from November 2023 – March 2024. Data collection for environmental factors consisting of temperature, humidity, and tides of seawater from November 2023 – March 2024. Figure 1 explains the sea cucumber catch and Figure 2,3 and 4 explains environmental factors.

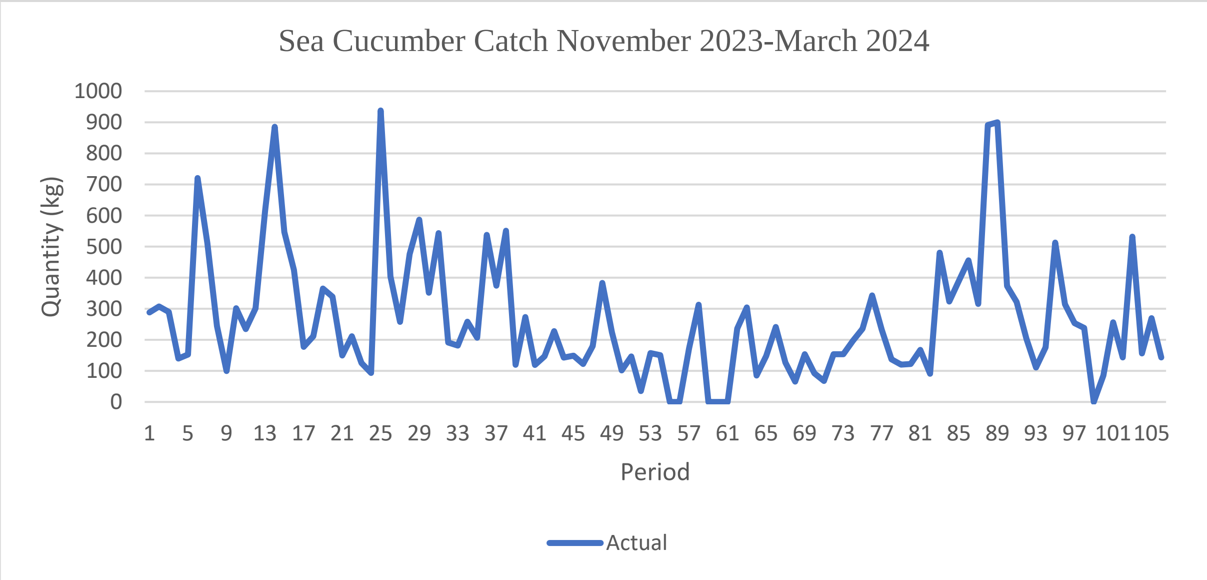

Figure 1. Sea Cucumber Catch in November 2023-March 2024

Seawater temperature affects the metabolism and distribution of fish and significant temperature changes can affect the overall health of the fish. Air humidity affects rainfall and evaporation patterns, negatively impacting salinity and food availability in fish habitats. Temperature, humidity, and tides of seawater are usually recorded with daily and monthly periods. The following is data on temperature, humidity, and tides of sea water from November 2023 to March 2024.

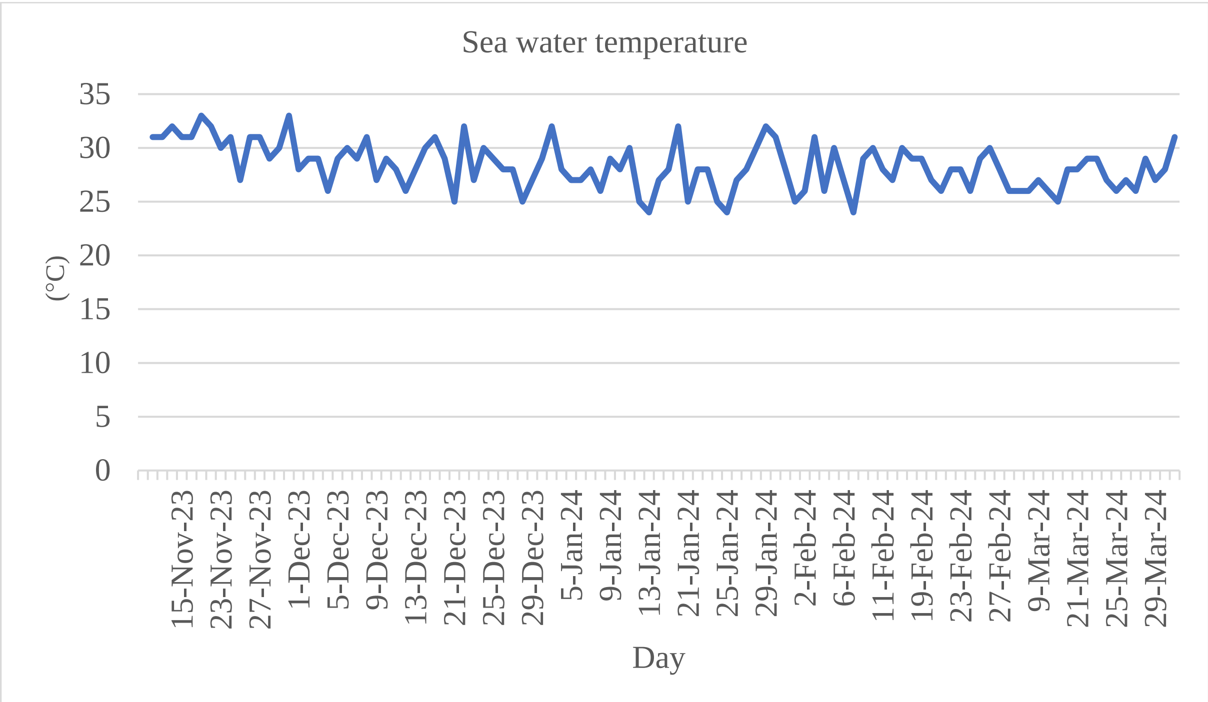

Figure 2. Sea water Temperature in November 2023-March 2024

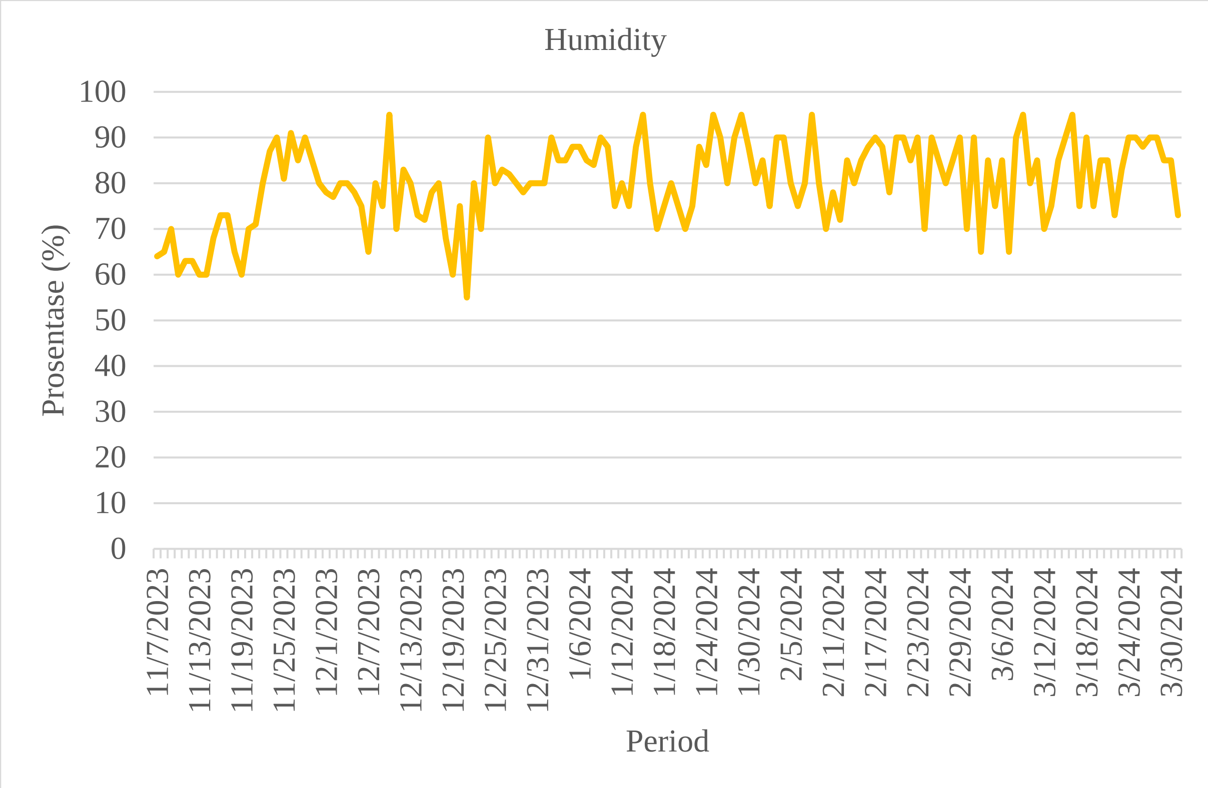

Figure 3. Humidity for November 2023-March 2024

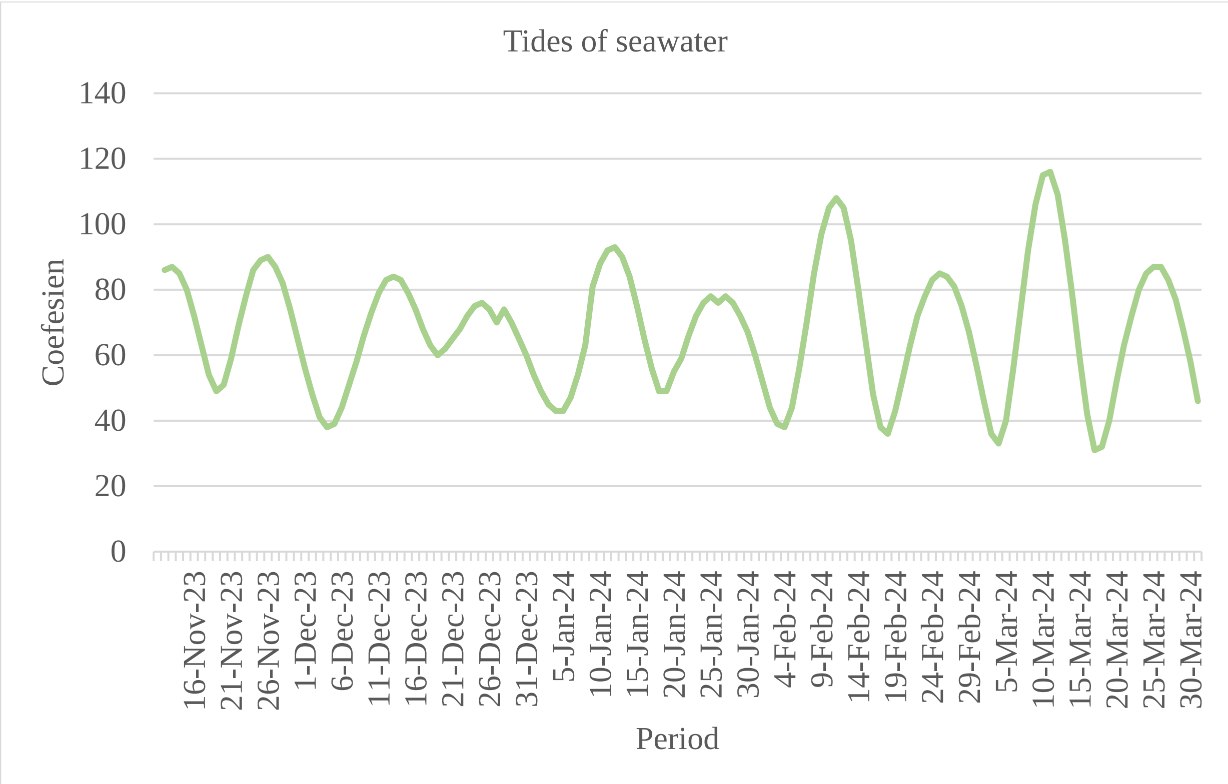

Figure 4. Sea Tides for November 2023-March 2024

2. Determination of ARIMA

The determination of the ARIMA model was obtained through a series of stationary inputs in the rainy season pattern in which the output of sea cucumbers is affected by temperature, humidity, and sea tides. The stationery in *the variance* of the input variable is found from the stationarity in *the mean* through *the autocorrelation function* (ACF) of the plot and *the partial autocorrelation function* (PACF) of the plot as follows.

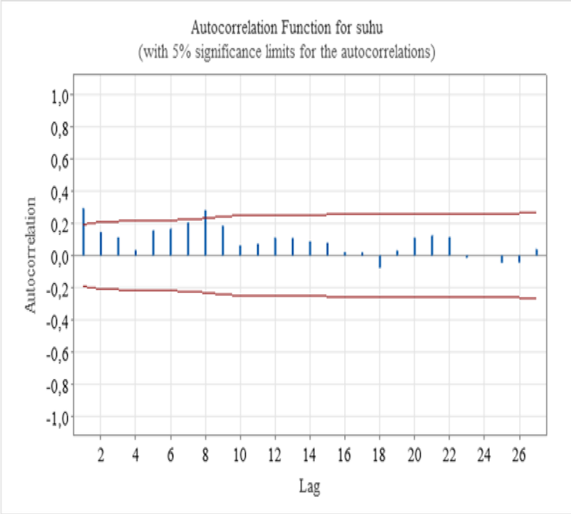

Figure 5. ACF Temperature Input Series

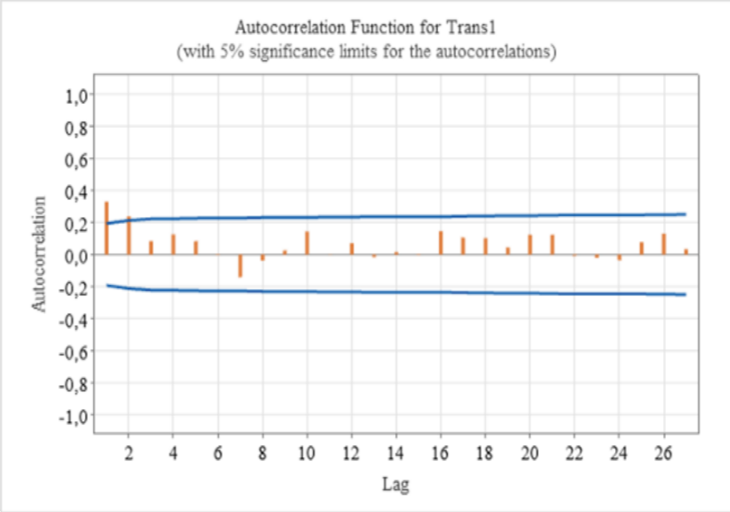

Figure 6. ACF Humidity Input Series

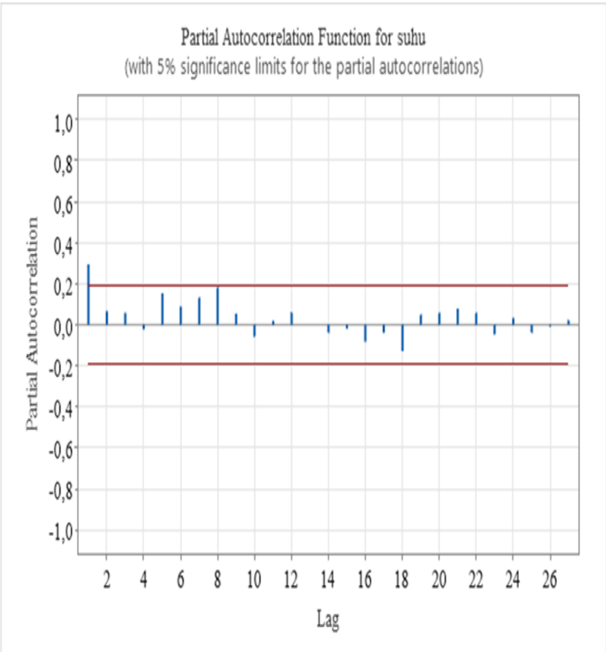

Figure 7. PACF Temperature Input Series

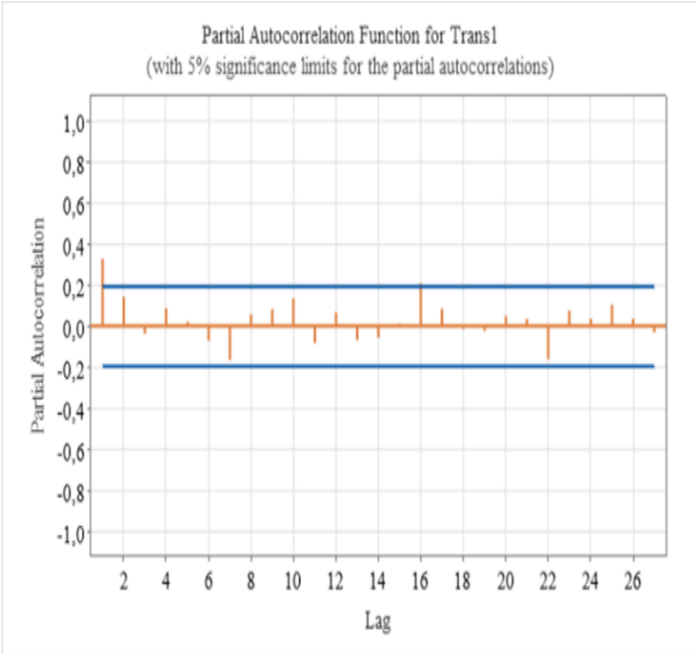

Figure 8. PACF Humidity Input Series

Based on figures 5-6 which explain the autocorrelation function (ACF) plot and partial autocorrelation function (PACF) of each factor that affects the catch of sea cucumbers that have experienced stationary data. In Figure 7-8, the humidity input series has undergone transformation, which means that the data has been transformed and the lag value does not exceed the average. At the ACF and PACF stages, the tidal variables of seawater cannot be done because the tidal data of seawater is periodic and has a clear seasonal pattern. While ARIMA can handle some simple seasonal shapes, SARIMA models or other specialized models are better suited for complex seasonal patterns. The tides of seawater are highly dependent on cycles and other astronomical factors, so the standard ARIMA model cannot fully capture the pattern. This explanation is in line with the findings from [13].

Model Formation:

The formation of the model is obtained through the parameter values of a statistical model with available data called parameter estimation. The goal is to find the parameter values that are most likely to produce the observed data. Based on the ACF and PACF stationary input series, the provisional model is expected to be the input variable for temperature and humidity data in the rainy season. The following is a table of the approximate models of temperature and humidity.

Table 2. Model Parameter Estimation

Variable	Model	*Type*	*Coef*	*SE Coef*	T-*Value*	P-*Value*
Temperature	1,0,1	AR 1	0,9998	0,0005	1729,35	0
		MA 1	0,9187	0,0362	25,35	0
Humidity	1,1,1	AR	1	-0,004	0,116	0,971
		MA	1	0,7388	0,0778	0
	1,1,0	AR	-0,5412	0,0718	-7,53	0

An evaluation process to assess the suitability and performance of statistical models that have been known as diagnostic models. The purpose of model diagnosis is to ensure that the resulting model is among the best and provides accurate and reliable results. Residual analysis in 2 ways, namely: the results of the ACF and PACF plots of the residuals which show that the residuals must be random and not correlated, Here are the diagnostic results of the model from temperature, and humidity

Table 3. Model Diagnostics

Variable	Model	DF	SS	MS	AIC	BIC
Temperature	1,0,1	139	575,708	4,14179	575,207	584,053
Humidity	1,1,1	139	575,708	4,14179	205,199	214,085
	1,1,0	139	9894,93	71,1866	606,506	612,416

The best model is seen from the results of the residual and lend-box test data on each variable that has a model. To choose the best model, you can see from the lower AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) numbers, the model that has low AIC and BIC numbers out of the 3 known models is the model of the humidity variable (1,1,1).

Table 4. Modeling Result

Variable	Model	Lag	*Chi-Square*	DF	P-*Value*
Temperature	1,0,1	12	19,54	10	0,034
		24	34,05	22	0,049
		36	42,36	34	0,154
		48	55,47	46	0,16
Humidity	1,1,1	12	14,79	10	0,14
		24	25,76	22	0,262
		36	40,04	34	0,22
		48	49,34	46	0,341
	1,1,0	12	23,68	11	0,014
		24	31,9	23	0,102
		36	48,66	35	0,062
		48	65,86	47	0,036

Based on Tables 2-4, the model that allows forecasting to be carried out is the model from the humidity (1,1,1) formula:

$$(1 - 0{,}2426B)(1 - B)Y_t = \mu + (1 + 0{,}991B)\epsilon t$$

$$Y_t - Y_{t-1} - 0{,}2426(Y_{t-1} - Y_{t-2}) = \mu + \epsilon t + 0{,}991_{\epsilon t-1}$$

$$Y_t = (1 + 0{,}2426)Y_{t-1} - 0{,}2426Y_{t-2} + 0{,}991_{\epsilon t-1} + \epsilon t$$

$$Y_t = 1{,}2426Y_{t-1} - 0{,}2426Y_{t-2} + 0{,}991_{\epsilon t-1} + \epsilon t$$

This model implies that the value of Y_t at a given time is positively influenced by the value at the previous time (Y_{t-1}) with a large positive coefficient (1.2426), as well as negatively affected by the value from the previous two times (Y_{t-2}) with a negative coefficient (-0.2426). In addition, the error at the previous time (ε_{t-1}) has a large positive influence on the current Y_t value, with a coefficient of 0.991.

Forecasting Results

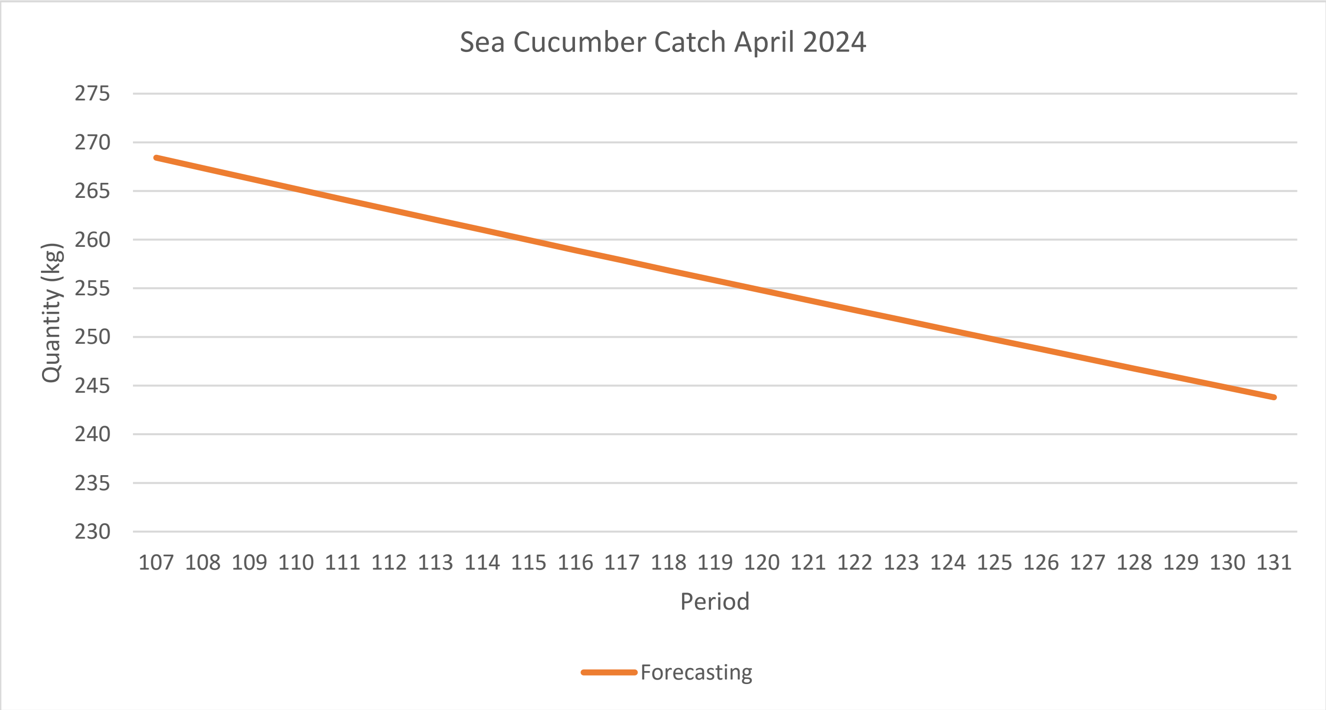

Figure 9. Forecast Result

Based on the best model, it is seen from the results of the residual and end-box test data on each variable that has a model. To choose the best model, see from the lower AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) numbers, the model that has low AIC and BIC numbers out of the 3 known models is the model of the humidity variable (1,1,1). Modelling from the following is the forecasting results of the humidity model (1,1,1) for sellers of sea cucumber products for the period of April 2024 and the Root Mean Square Error value obtained is 271.11 and is as follows.

The results and discussion section of this study needs to be expanded to provide more comprehensive analysis and interpretation. The model validation was conducted by comparing the predicted results with actual data on sea cucumber catches, showing good accuracy with a low error rate. This model offers new insights into the influence of humidity on sea cucumber catch results, although it does not account for other environmental factors such as salinity and ocean currents. Incorporating these variables could enhance the model's accuracy. When compared to previous research, this model demonstrates better predictive capabilities, particularly during the rainy season. However, further research is needed for a more thorough comparison with other simulation models. With ongoing development, this model is expected to make a significant contribution to the management of marine resources and the forecasting of sea cucumber catches.

Conclusion:

This study was conducted to forecast the fluctuating and uncertain sea cucumber catch, as well as to identify the most effective predictive model using the ARIMA (AutoRegressive Integrated Moving Average) method. The main objective is to determine a model that can provide predictions with the lowest error rate, thereby enabling accurate forecasting of sea cucumber catches in the future. Given that catch results are significantly influenced by environmental factors such as humidity, the proposed model focuses on this variable as one of the key determinants in short-term predictions. Among the various models tested, ARIMA (1,1,1) proved to be the most significant and effective in predicting sea cucumber catches, particularly for short-term periods (1 to 2 days) during the rainy season. This suggests that ARIMA (1,1,1) can capture seasonal patterns and environmental changes that directly impact catch results. The ARIMA (1,1,1) model equation derived from this study is as follows:

$$Y_t = 1{,}2426Y_{t-1} - 0{,}2426Y_{t-2} + 0{,}991_{\epsilon t-1} + \epsilon t$$

with the smallest prediction error of 271.11. This indicates that the model can reduce uncertainty in forecasting with a high level of accuracy. Furthermore, the forecast for sea cucumber catches in April 2024 for the 107th period resulted in a predicted value of 268.42 kg, which closely matches the actual data from previous periods. This proximity demonstrates that the ARIMA (1,1,1) model is highly reliable for predicting future sea cucumber catches, providing valuable information for fishermen and policymakers in preparing more efficient fishing strategies.

References:

[1] Jernihati, Y., Kurniawan, G., & Zai, S. Analysis of Sales Forecasting in Raw Material Inventory Management At Sun Cafe. J. Gea. 483 Jurnal EMBA, 11(4), 483–490, 2023.

[2] Oey, E., & Ayrine, G. K. Penerapan Proses Dan Teknik Peramalan – Studi Kasus Di Manufaktur Transformer. Jurnal Manajemen Industri Dan Logistik, 2(2), 106–115, 2018.

[3] Buchori, M., & Sukmono, T. Peramalan Produksi Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) di PT. XYZ. PROZIMA (Productivity, Optimization and Manufacturing System Engineering), 2(1), 27–33, 2018.

[4] Pramujo, B., Juwono, P. T., & Soetopo, W. (2015). Pemodelan Debit Menggunakan Metode Arima Guna Menentukan Pola Operasi Waduk Selorejo. Jurnal Teknik Pengairan, 5(2), 141–148, 2015.

[5] Hutasuhut, A. H., Anggraeni, W., & Tyasnurita, R. (2014). Bahan Baku Plastik. Jurnal Teknik POMITS, 3(2), 169–174, 2014.

[6] Nur, M. I., & Puspita, A. R. Peramalan Produksi Banggai Cardinalfish Menggunakan Autoregressive Integrated Moving Average (ARIMA). Jurnal Karya Pendidikan Matematika, 8(1), 23–29, 2021.

[7] Nugroho, K. Model Analisis Prediksi Menggunakan Metode Fuzzy Time Series. Infokam, 12(1), 46–50, 2016.

[8] Wei, W.W.S. Time Series Analysis:Univariate and Multivariate Methods. New York: Pearson, 2006.

[9] Salwa, N., Tatsara, N., Amalia, R., & Zohra, A. F. Peramalan Harga Bitcoin Menggunakan Metode ARIMA (Autoregressive Integrated Moving Average). Journal of Data Analysis, 1(1), 21–31, 2018.

[10] Razak, F. A., Shitan, M., Hashim, A. H., & Abidin, I. Z. Load Forecasting Using

Time Series Models. Jurnal Kejuruteraan, 21(1), 53–62, 2009.

[11] Amjady, N. Short-Term Hourly Load Forecasting Using Time-Series Modeling With Peak Load Estimation Capability. IEEE Transactions on Power Systems, 16(4), 798–805, 2001.

[12] Djoni, H. . Penerapan Model ARIMA untuk Memprediksi Harga Saham PT. Telkom Tbk. Jurnal Ilmiah Sains, 11(1), 116–123, 2011.

[13] Pal, N., Barton, K. N., Petersen, M. R., Brus, S. R., Engwirda, D., Arbic, B. K., Roberts, A. F., Westerink, J. J., & Wirasaet, D. (2023). Barotropic Tides in MPAS-Ocean: impact of ice shelf cavities. Geoscientific Model Development, 16(4), 1297–1314. https://doi.org/10.5194/gmd-16-1297-2023.