Optimizing chlo and sst for skipjack tuna fisheries in the banda sea and its surroundings using GAM and ECDF

Sunarwan Asuhadi[1,2*], *Mukti* Zainuddin[1], Safruddin[1], and *Musbir* Musbir[1]

[1]Faculty of Fisheries and Marine Sciences (FIPK) Hasanuddin University, Jl. Perintis Kemerdekaan, KM 10 Tamalanrea, Makassar, Indonesia 90245,
[2]National Research and Innovation Agency (BRIN), Gedung B.J. Habibie, Jl. M.H. Thamrin No. 8, Jakarta Pusat 10340, Indonesia

Abstract. This study aims to determine the optimal values of chlorophyll-a (Chlo) and Sea Surface Temperature (SST) for skipjack fishing (*Katsuwonus pelamis*) in the Banda Sea and its surroundings using the GAM model and ECDF function. The GAM model yields the formula CPUE = 33.07 + s(Chlo, edf = 7.14, Ref.df = 8.09) + s(SST, edf = 8.95, Ref.df = 8.99), with a significant value for Chlo of 0.58 mg/m³ and an SST of 28.26°C, as well as the highest CPUE value of 51.84. Due to the low performance of the GAM model, ECDF is used to determine the optimal values of Chlo and SST. The results of the Kolmogorov-Smirnov test on the ECDF value showed a p-value for Chlo of 0.0016 and SST of 4.45e-14 ($p < 0.05$). Thus, there was a significant difference between the f(t) and g(t) distributions at Chlo = 0.28 mg/m³ and SST = 30.57°C. K-Means analysis shows that the optimal value range for Chlo is 0.075 - 0.33 mg/m³ and for SST is 29.47 - 30.47°C. The optimal months are January to June as well as October, where the combination of SST and Chlo concentrations reaches levels that support the productivity of skipjack fishery. However, from July to September, lower SSTs begin to dominate, which can reduce fishery optimality, although some regions still have supportive Chlorophyll. In November and December, SSTs reach very high levels, potentially reducing productivity efficiency.

1 Introduction

The skipjack tuna fishery (*Katsuwonus pelamis*) is one of the important sectors in the fishing industry, especially in Indonesia's waters. Skipjack tuna has high economic value and is a major food source for many coastal communities. A deep understanding of the environmental factors that affect the distribution and abundance of skipjack tuna is essential to optimize catches and maintain the sustainability of these fisheries [1], [2].

* Corresponding author: suna032@brin.go.id

Chlorophyll-a (Chlo) and Sea Surface Temperature (SST) are two oceanographic parameters that are very important in determining the optimal habitat for skipjack tuna. Chlo is a key indicator of marine primary productivity, indicating the presence of phytoplankton that form the basis of the marine food chain. SST, on the other hand, affects various aspects of fish physiology and behavior, including the distribution, migration, and abundance of fish populations [3], [4].

Several studies have shown a significant relationship between Chlo concentrations and SSTs with skipjack tuna abundance. For example, research by Elepathage and Tang (2019) found that the main pelagic habitat of skipjack tuna correlates with Chlo concentrations between 0.4 - 0.7 mg/m³ and SSTs between 26 - 27°C in the Indian Ocean around Sri Lanka [5]. Another study by Hidayat et al. (2021) showed that the best habitat area for skipjack tuna in the Makassar Strait has an SST between 28.78 to 31.25°C and Chlo between 0.18 to 0.28 mg/m³ [6].

The use of statistical models such as Generalized Additive Models (GAM) and Empirical Cumulative Distribution Function (ECDF) in fisheries research has helped in identifying and predicting potential fishing zones more accurately. For example, the GAM model was used to predict the relationship between SST and Chlo with skipjack tuna abundance in various aquatic areas, suggesting that these parameters can be important indicators for detecting the optimal habitat of skipjack tuna [7], [8].

This study aims to optimize the use of Chlo and SST in skipjack tuna fisheries in the Banda Sea and its surroundings by using GAM and ECDF models. With a better understanding of these oceanographic parameters, it is hoped that it will improve the efficiency and sustainability of skipjack tuna fishery practices in the region.

2 Method

2.1 Research Area

The location of the research is in Fisheries Management Area 714, namely Tolo Bay, Banda Sea, and its surroundings as shown in Figure 1 below. The location of FMA 714 is regulated in Marine and Fisheries Regulation No. 18/PERMEN-KP/2014.

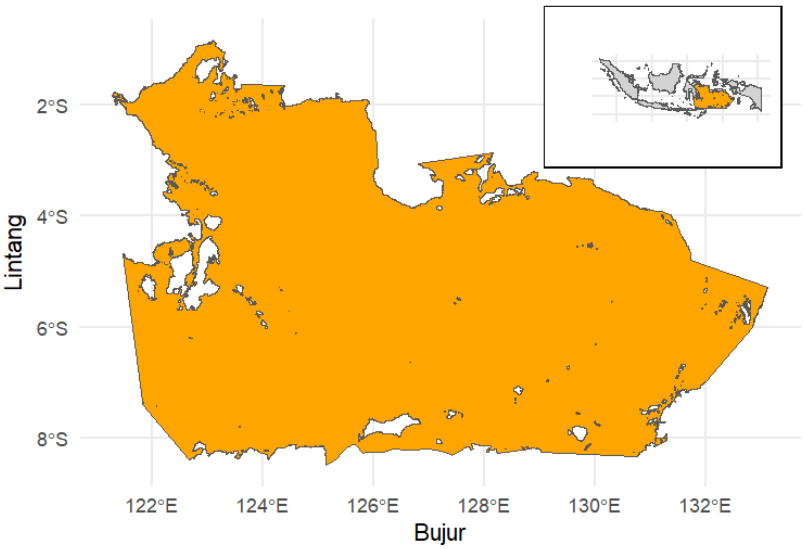


Fig. 1. Research Area.

2.2 Data Collection and Processing

This study uses data from capture fisheries logbook datasets in the Banda Sea area and its surroundings, Chlo concentration, and SST data. The dataset comes from the capture fisheries logbook of the Ministry of Marine Affairs and Fisheries of the Indonesian Republic from 2014 to 2022. The concentrations of Chlo and SST come from Aqua Modis. Furthermore, Ocean Data View (ODV) is used to process raw data on Chlo and SST which has a resolution of 4 km. Finally, all datasets are processed, modeled, and visualized using R Studio version 4.4.0.

2.3 GAM Modeling

The Generalized Additive Model (GAM) was used to analyze the relationship between Chlo and SST and the presence of skipjack fish.

$$Y(s) = \beta 0 + \sum f(Xi) + \varepsilon \quad (1)$$

Y(s) is the dependent variable (presence of fish), β0 is the Intercept, f(Xi) is the smoothing function for the independent variable, and ε is the Error [9], [10].

2.4 ECDF Modeling

The Empirical Cumulative Distribution Function (ECDF) is used to understand the distribution of Chlo and SST data.

$$F(x) = P(X \leq x) \quad (2)$$

F(x) is a function of cumulative distribution, and P(X ≤ x) is the probability of a random variable X being less than or equal to x [11], [12].

2.5 Uji Kolmogorov-Smirnov 2 Variables

The Kolmogorov-Smirnov test was conducted to compare the two cumulative distributions and determine if they differ significantly.

$$D = \max|F1(x) - F2(x)| \quad (3)$$

D is the test statistic, and F1(x), F2(x) is the cumulative distribution function of the two samples [13].

2.6 K-Means Clustering and Elbow Method

K-Means Clustering is used to group Chlo and SST data into clusters. Meanwhile, the elbow method is used to determine the optimal number of clusters.

$$J = \sum \min(||x_i - \mu_j||^2) \quad (4)$$

J is the sum of the squares of the distance in the cluster, x_i is the data point, and μ_j is the center point of the cluster [14].

2.7 Mapping Using R Studio

The data from the grouping and modeling results above was mapped using R Studio for visualization and interpretation [15].

3 Results And Discussion

3.1 Comparison of GAM Models with and without outliers

The results of GAM modeling show the output that distinguishes between outliers and no outliers [16]. GAM modeling with outliers yields an AIC value of 246,900.30, an intercept with a value of 125.20, a standard error of 2.16, a t-value of 57.95, and a very small p-value (<2e-16), which means that this intercept is very significant. The smooth 's(Chlo)' function has an effective degree of freedom (edf) of 4.10 and an F value of 1.44. The p-value for this is 0.21, which means that Chlo is not significant in this model at a significance level of 5% [17]. The smooth 's(SST)' function has an edf of 8.83 and an F value of 6.64. The P-value is very small (<2e-16), which means that the SST is very significant in this model [18].

The goodness of fit model is shown with an adjusted R-squared value of 0.0037, which means this model explains about 0.37% of the variability in CPUE [19]. The deviance described is 0.443%, which also shows that this model explains a small fraction of the variability in the data. The above model is then examined for its data structure, whether it has an outlier or extreme skewness that could affect the model's estimation, using the 'ggplot' function. The results of the above formula are shown in the following Figure 2. It appears that there are outliers in the Chlo and SST columns [20].

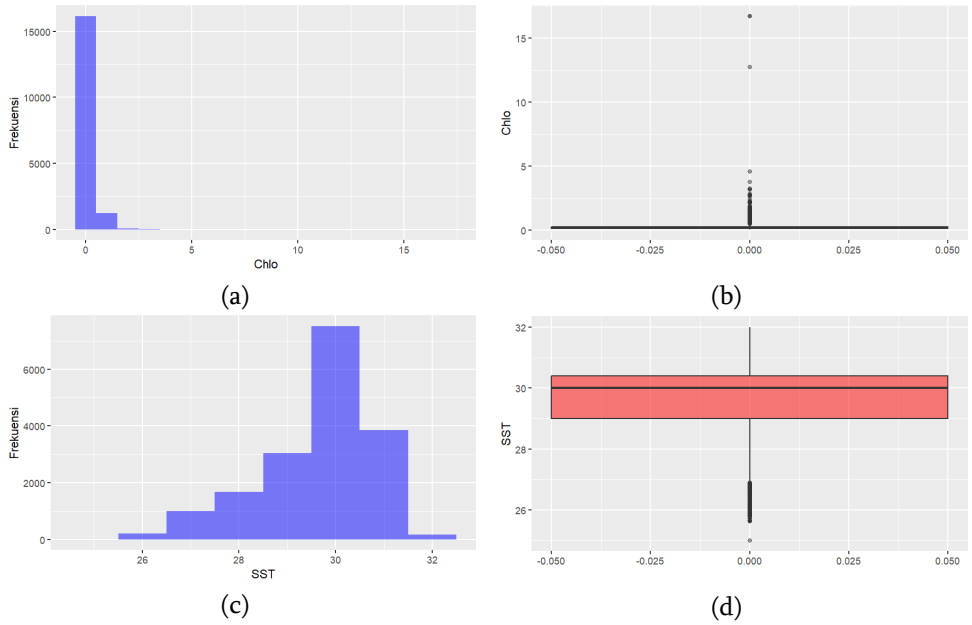


Fig. 2. *(a)* Outlier distribution of Chlo, *(b)* Outlier boxplot of Chlo, *(c)* Outlier distribution of SST, and *(d)* Outlier boxplot of SST.

Overall, the Chlo chart shows that most of the Chlo values are near zero and the frequency decreases as the Chlo values increase [21]. The boxplot shows that most of Chlo's values are near zero and some values are far from zero, indicating the existence of outliers. The distribution on the SST histogram graph shows that SST values are most common at around 30°C, with the highest event frequency reaching more than 7,000. This value indicates that the majority of SSTs in this dataset are around that number. In contrast, the lowest SST value recorded at around 25°C and the highest value at around 33°C indicates temperature variation even with a much lower frequency [22]. This distribution depicts a temperature pattern that tends to converge around 30°C, with few occurrences of extreme temperatures at both ends of the spectrum, reflecting thermal consistency in the region or period being measured.

The boxplot on the SST graph above shows the temperature distribution with details of the highest, lowest, and distribution of the data [23]. The median value of the SST is around 30°C, which is marked by a horizontal line in the box. The box includes an interquartile range (IQR) that shows that 50% of the data is between about 29°C and 31°C. The highest value (maximum) that is not considered an outlier is slightly above 31°C, while the lowest value (minimum) that is not considered an outlier is slightly below 29°C. Some data points below 28°C are considered outliers and are described as individual points under the box. This boxplot illustrates that the majority of SST data ranges from 29°C to 31°C with some lower temperatures appearing as outliers [24]. To ascertain how outliers affect the data above, a model was created that ignores outliers. Before modeling without outliers, the outlier threshold value is first simulated, from n = 1 to 7. The results of the simulation are shown in the following Figure 3.

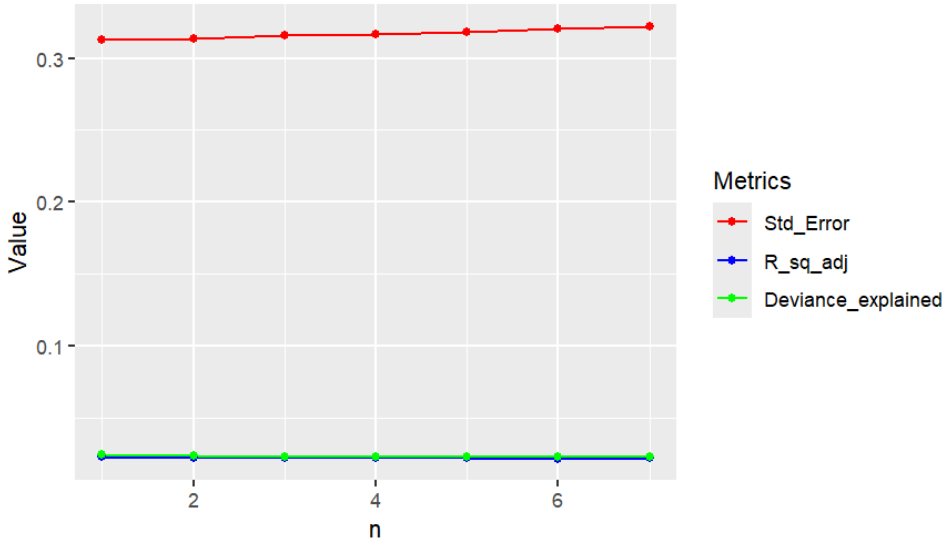


Fig. 3. GAM model outlier threshold simulation metrics.

In the model above, it appears that the addition of the value of n causes an increase in Std. Error [25], and a decrease in the value of R-sq. (adj) and Deviance explained. So the value of n = 1 is chosen which is considered the best. From the two models above, it appears that the value of the R-squared model with an outlier is 0.0037, while the R-squared without an outlier has a value of 0.023. In addition, the AIC value for the first model is 246,900.30, greater than the AIC value of the second model is 128,618.30 [26]. Thus models without outliers are better compared to models that use outliers. The difference in metrics between models with and without outliers is shown in the following Table 1.

Table 1. Comparison of model metrics with and without outliers.

Parameter	Without Outliers	With Outliers
AIC	128,618.30	246,900.30
R-sq.(adj)	0.023	0.0037
(Intercept)	33.07	125.20
Std. Error	0.31	2.16
Deviance explained	2.41%	0.443%
GCV	1,260.60	81,519
Scale est.	1,258.90	81,454
n	12,891	17,453
MSE Cho	1,080.48	15,634.70
MSE SST	39.81	9,426.06

In addition to comparing the models above, multicollinearity checks are also carried out [27]. Multicollinearity will occur if two or more predictor variables are strongly

correlated. The test was performed using the vif() function of the 'car' package to check this [28].

```
library(car)
library (carData)
vif(modelgam_cakalang_outlier)
```

The output of the multicollinearity test above is as follows.

```
          GVIF Df GVIF^(1/(2*Df))
Chlo 8.7162e+20  0           Inf
SST  8.7162e+20  0           Inf
```

The results of the multicollinearity test show that the variables Chlo, and SST have very high GVIF (Generalized Variance Inflation Factor) values (8.7162e+20 0) and infinity (Inf) for GVIF^(1/(2*Df)) [28]. This indicates a very strong multicollinearity between these variables. Multicollinearity is a condition in which several predictor variables in a regression model strongly correlate with each other. This can lead to problems in the estimation of regression coefficients and error standards, which in turn can affect the interpretation of the model [29]. The relationship between Chlo and CPUE is shown in the following figure 4.

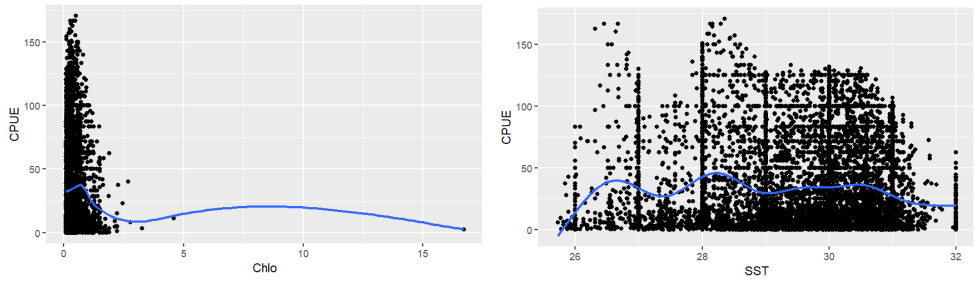


Fig. 4. Scatter plot graph between target variable (CPUE) and predictor on skipjack fishing: *(a)* Chlo, and *(b)* SST.

Graph 4.a is a scatter plot showing the relationship between Chlo (x-axis) and CPUE (y-axis). Each point represents one observation with the values Chlo and CPUE. The average Chlo is 0.25. The x-axis ranges from 0 to more than 15, and the y-axis from 0 to 150. The dense dot spread on the left of the graph shows à lot of data with low Chlo and high CPUE [30]. The blue curve shows a downward trend of CPUE with an increase in Chlo, indicating a negative correlation between the two variables [31].

Graph 4.b shows the relationship between SST (x-axis) and CPUE (y-axis). The average SST is 29.68. The x-axis ranges from 24 to 32, and the y-axis from 0 to more than 150. The black dots represent individual observations, and the blue lines indicate a general trend of increasing CPUE with an increase in SST, despite fluctuations [32], [33].

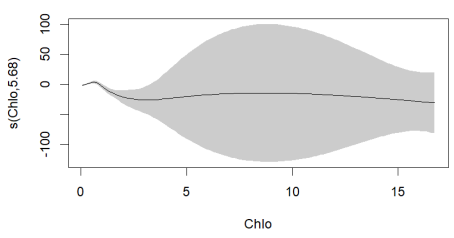
The relationship between each predictor variable and the target variable is expressed in the GAM model. In this case, the target variable is CPUE. The relationship

is visualized in a scatter plot chart. To obtain such a relationship, a comparison was made between the single model (CPUE ~ s(Chlo) and CPUE ~ s(SST)) and the combined model (CPUE ~ s(Chlo) + s(SST)). More data points are accumulated in the intermediate value range of the SST data, and fewer at the edges [34]. A comparison of all GAM models is shown in the following Table 2.

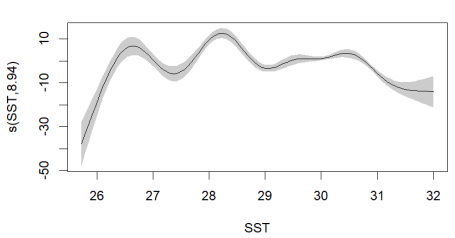
Table 1. Comparison of GAM models between single and combined predictors.

Parameter	**Model 1 (CPUE ~ s(Chlo))**	**Model 2 (CPUE ~ s(SST))**	**Model 3 (CPUE ~ s(Chlo) + s(SST))**
Intercept Estimate	33.07	33.07	33.07
Std. Error	0.32	0.31	0.31
t value	104.80	105.60	105.80
edf	5.68	8.94	(s(Chlo)) = 7.14, (s(SST)) = 8.95
Ref.df	6.69	8.99	(s(Chlo)) = 8.09, (s(SST)) = 8.99
F	6.92	27.69	(s(Chlo)) = 7.49, (s(SST)) = 28.95
p-value	<2e-16	<2e-16	(s(Chlo)) = <2e-16, (s(SST)) = <2e-16
R-sq.(adj)	0.0035	0.019	0.023
Deviance Explained	0.39%	1.92%	2.41%
GCV	1,284.60	1,265.60	1,260.60
Scale est.	1,283.90	1,264.60	1,258.90
n	12,891	12,891	12,891
AIC	128,861.40	128,668.90	128,618.30

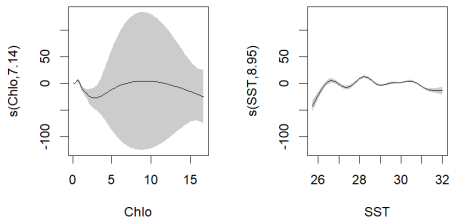
Based on comparative analysis, model 3 (CPUE ~ s(Chlo) + s(SST)) shows the best performance among the three models. The model has the lowest Std. Error (0.31) and the highest t-value (105.80), indicating more precise and significant parameter estimation [34]. Model 3 also had the highest F-statistic values for s(SST) (28.95) and s(Chlo) (7.49), as well as higher R-sq.(adj) (0.023) and deviance explained (2.41%) values, indicating a better ability to explain data variations. The lowest GCV value (1,260.60) and the lowest scale est. (1,258.90) confirm better prediction performance. With the lowest AIC (128,618.30), model 3 balances data complexity and relevance well, so it can be concluded that model 3 is the most accurate and significant for CPUE data analysis compared to model 1 and model 2. The three models above are visualized in the following Figure 5.



(a)



(b)



(c)

Fig. 5. The 'edf' chart of a single model on skipjack fishing: *(a)* Chlo, *(b)* SST, and *(c)* combined model.

Figure 5.a shows a CPUE prediction model based on the variable s(Chlo). The flat black curve and consistent confidence interval suggest that variations in Chlo concentrations do not significantly affect CPUE, indicating Chlo is not a strong predictor [34]. Figure 5.b shows a CPUE prediction model based on the variable s(SST). The black curve shows significant fluctuations in CPUE as the SST changes, indicating a non-linear relationship between SST and CPUE. Wider confidence intervals in some parts indicate higher uncertainty in model estimates. Figure 5.c shows a CPUE prediction model based on s(Chlo) and s(SST), with two separate panels. The left panel shows a horizontal relationship between CPUE and Chlo with a wide confidence interval, signaling a weak influence of Chlo. The right panel shows the relationship between CPUE and SST with small fluctuations and narrow confidence intervals, showing a more significant and complex SST influence [35]. Overall, SST had a stronger and more predictable influence on CPUE than Chlo.

According to the results of the GAM modeling, the performance of the model is statistically low, it appears that the deviation value described < 50%, and the Std. Error > 0.05 [34]. The causes can be a lack of data, an unrepresentative sample, an overly simple model, high multicollinearity between the predictor variables, and missing data. The GAM model yields the formula CPUE=33.07+s(Chlo,edf=7.14,Ref.df=8.09)+s(SST,edf=8.95,Ref.df=8.99) with a significant value for Chlo of 0.58 mg/m3 and SST of 28.26 °C, and the highest CPUE value of 51.84. Due to the low performance of the GAM model, the determination of the optimal values of Chlo and SST was carried out by the Empirical Cumulative Distribution Function (ECDF) Model. ECDF has advantages such as a focus on empirical distribution, freedom from distribution assumptions, outlier detection capability, ease of implementation, and relevance in identifying the relationship between CPUE and oceanographic conditions [36].

ECDF graphs are used to observe the data distribution of each parameter. The ECDF diagram identifies the values that produce the maximum difference between the two functions, showing a significant relationship between Chlo and SST with CPUE in skipjacks [5]. The following figure shows the ECDF plots for the variables Chlo and SST, which illustrate the distribution of values in the data. The lines on the graph represent the ECDF of the Chlo and SST data. The plot shows that the maximum value for Chlo is 0.280 mg/m^3, while for SST it is 30.57°C. However, these values do not directly indicate

that the values of Chlo and SST are significant for skipjack fishing in the Banda Sea and its surroundings. To assess the significance of these two parameters, a two-variable Kolmogorov-Smirnov test (CPUE ~ Chlo and CPUE ~ SST) was carried out to test the suitability between the data sample and the predetermined distribution [6].

ECDF graphs are used to observe the data distribution of each parameter. The ECDF diagram identifies the values that result in the maximum difference between the two functions. The significant relationship between Chlo and SST with CPUE in skipjack is shown in Figure 6.

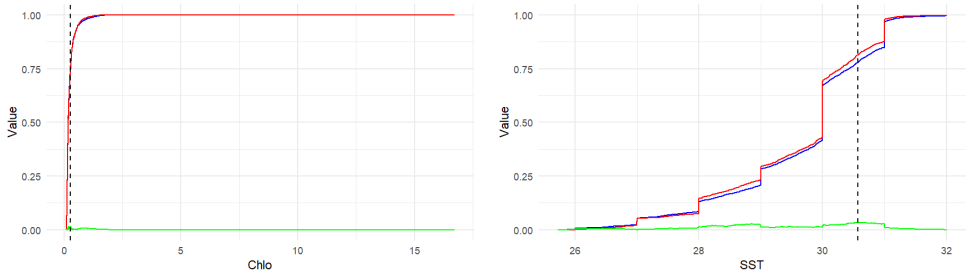
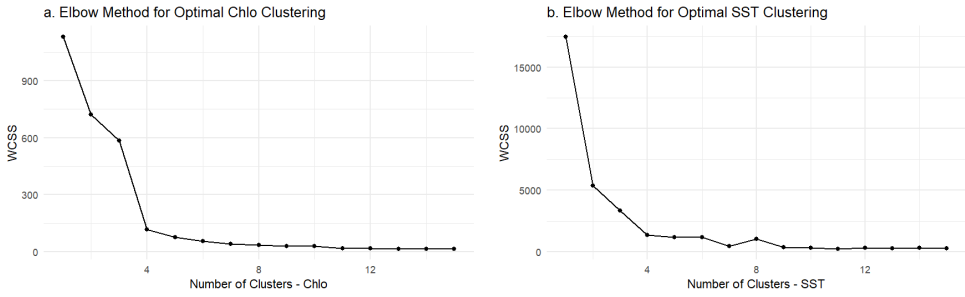


Fig. 6. *(a)* Relationship between Chlo and ECDF, and *(b)* SST and ECDF

Graph 6.a shows the Empirical Cumulative Distribution Function (ECDF) functions for three different calculations (f_CakalangChlo(t), g_CakalangChlo(t), and D_CakalangChlo(t)) for Chlo. The X-axis shows the Chlo value from 0 to about 15, and the Y-axis shows the ECDF value from 0 to 1. The red line (f_CakalangChlo(t)) increases sharply near Chlo = 0 and reaches a plateau at a value of 1, indicating the concentration of data in this range [37]. The horizontal green line at the value of 0 represents g_CakalangChlo(t), and the vertically dotted black line around Chlo = 2 indicates D_CakalangChlo(t). This graph shows that the significant ECDF Chlo value for skipjack fishing in WPP 714 is 0.28. Graph 6.b shows the ECDF function for three different datasets (f_CakalangSST(t), g_CakalangSST(t), and D_CakalangSST(t)) against SST. The X-axis shows the SST value from about 24 to 32, and the Y-axis shows the ECDF value from 0 to 1. The red line represents f_CakalangSST(t), the blue line represents g_CakalangSST(t), and the green line represents D_CakalangSST(t). All lines show the cumulative distribution along the SST value range. The vertical dotted line around SST = 30 indicates a specific threshold point. This graph illustrates the cumulative distribution of SST values in the dataset [38].

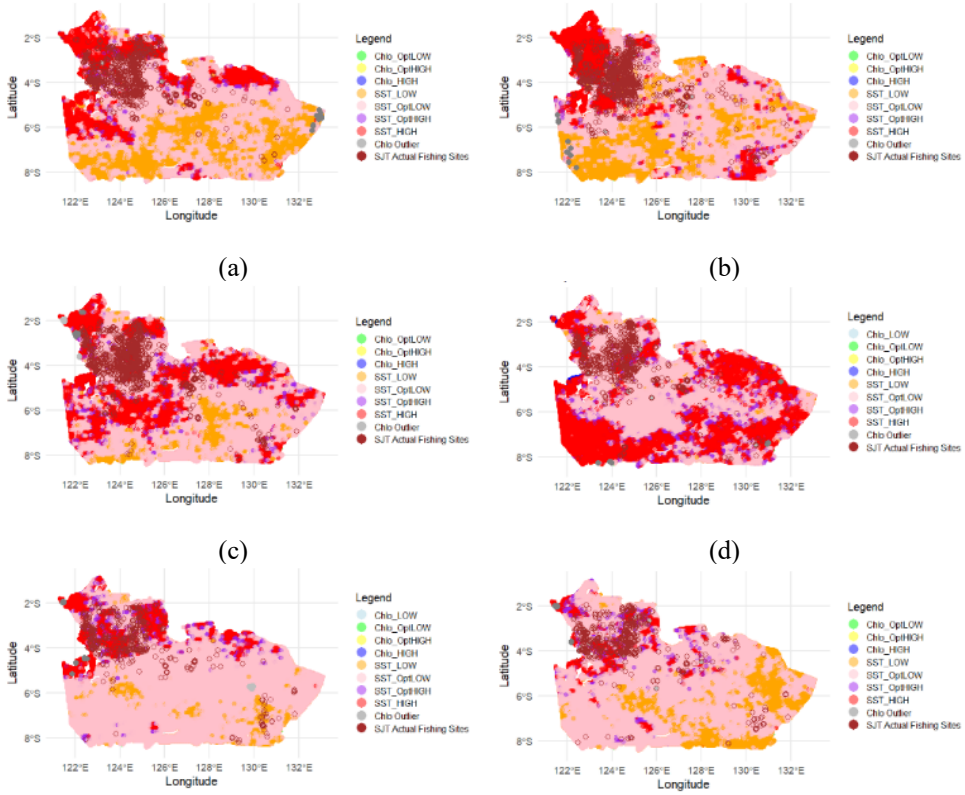
The results of the Kolmogorov-Smirnov test showed that the p-value for Chlo was 0.0016 and the SST was 4.45e-14 ($p < 0.05$), so these values showed a very significant difference between the distribution of f(t) and g(t) at Chlo = 0.28 mg/m^3 and SST = 30.57°C. Thus, the significant value for Chlo is 0.28 mg/m^3, and the SST is 30.57°C [39]. To determine the optimal value of Chlo and SST, the elbow method is used as shown in Figure 7 below. In the image below, it appears that the elbow points for Chlo and SST are in cluster 4 respectively. These 'elbow' points are then incorporated into the 'kmeans' function equation [40].



(a) (b)

Fig. 7. Graph of cluster values based on the elbow method: *(a)* Chlo, and *(b)* SST

According to the K-Means equation above, the optimal value range of Chlo in skipjack fishing in the Banda Sea and its surroundings is 0.075 - 0.33 mg/m^3. The optimal value range for SST is 29.47 - 30.47°C. Based on the optimal values above, the condition of the waters of the Banda Sea can be visualized on the map as seen in Figure 8 below.



(a) (b)

(c) (d)

(e) (f)

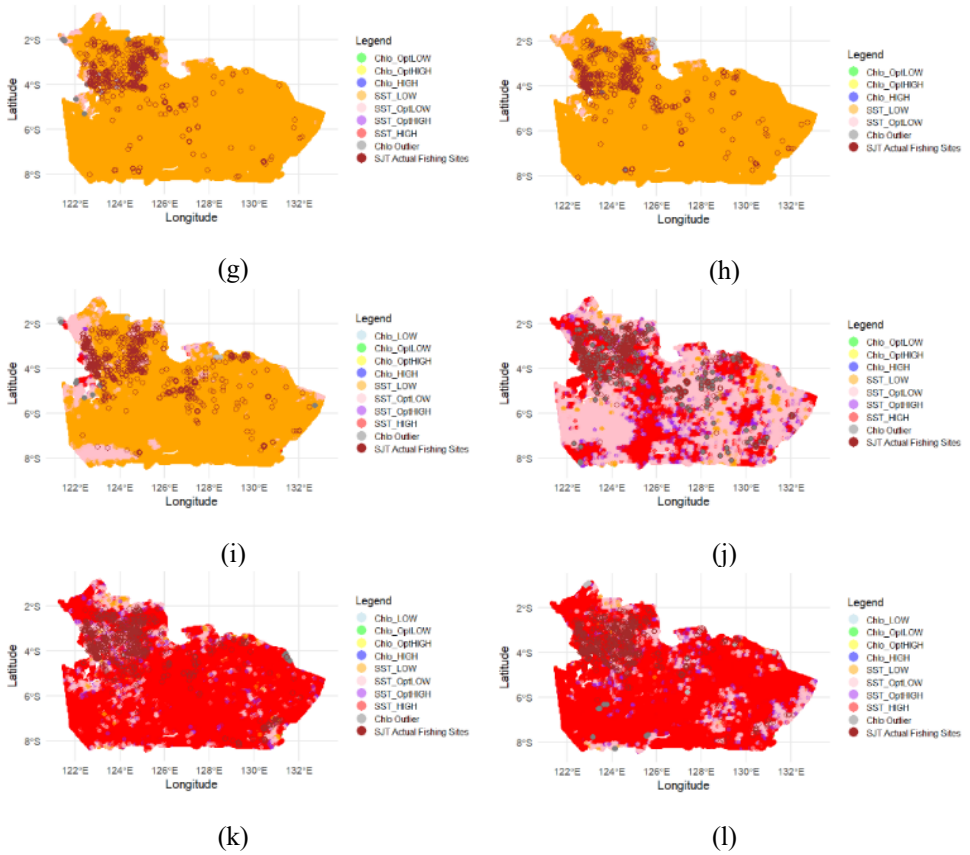


Fig. 8. Map of Banda Sea water conditions for the Potential of skipjack fishing throughout the year: *(a)* January, *(b)* February, *(c)* March, *(d)* April, *(e)* May, *(f)* June, *(g)* July, *(h)* August, *(i)* September, *(j)* October, *(k)* November, and *(l)* December.

Map above, it appears that overall, the waters of the Banda Sea show varying conditions throughout the year for fishery activities. The optimal months are January to June as well as October, where the combination of SST and Chlo concentrations reaches levels that support the productivity of skipjack fishery. However, from July to September, lower SSTs begin to dominate, which can reduce fishery optimality, although some regions still have supportive Chlorophyll. In November and December, SSTs reach very high levels, potentially reducing productivity efficiency.

4 Results And Discussion

This study evaluated the optimal value of Chlo and SST for skipjack fishing in the Banda Sea using the GAM model and ECDF function. The results of the GAM model showed that the significant value for Chlo was 0.58 mg/m³ and for SST was 28.26°C, with the highest CPUE value of 51.84. However, due to the low performance of the GAM model, the ECDF function is used to determine the optimal value more accurately. The analysis showed that the optimal value of Chlo ranged from 0.075 to 0.33 mg/m³, and SST between

29.47 to 30.47°C. The optimal months for skipjack fishing are January to June and October, while lower sea temperatures from July to September and very high temperatures from November to December tend to reduce fishing efficiency. These optimal conditions can be considered in the catch of bigeye tuna.

The results of previous studies also showed that the optimal Chlo value ranged from 0.10 to 0.15 mg/m^3 in Bone Bay and the Flores Sea [41], and 0.2 to 0.35 mg/m^3 in the waters of Ternate Island [42]. Optimal SST values range from 29.00 to 29.50°C in Bone Bay and the Flores Sea [41]. In addition, a study in 2018 showed that higher concentrations of Chlo (>0.30 mg/m^3) have the potential to increase skipjack fish production [43].

The results of this study show that the optimal values of Chlo and SST for skipjack vary in various locations, reflecting the importance of considering local oceanographic conditions in determining fishing strategies. This knowledge can be used to improve the sustainability and yield of catches, both for skipjack and bigeye tuna, through more precise management of fishing times and locations based on local environmental conditions.

5 Conclusions

The optimal values of Chlo and SST for skipjack fishing (Katsuwonus pelamis) in the Banda Sea and its surroundings can be determined using the GAM model and ECDF function. The results of the analysis showed that the optimal value of Chlo was in the range of 0.075 - 0.33 mg/m³ and SST in the range of 29.47 - 30.47°C. The optimal months for skipjack fishing are January to June as well as October when the combination of SST and Chlorophyll concentrations reaches levels that support fishery productivity. However, from July to September, lower SSTs begin to dominate, which can reduce fishery optimality. In November and December, SSTs reach very high levels, which has the potential to reduce the efficiency of fisheries productivity.

Acknowledgements

We thank the Faculty of Marine Science and Fisheries, Hasanuddin University, and the dataset providers, including the Directorate of Fish Resources Management of the Ministry of Marine Affairs and Fisheries of the Indonesian Republic and Ocean Color, for their invaluable contributions that enabled in-depth and accurate analyses in our research.

References

1. R. Hidayat and M. Zainuddin, Skipjack tuna fishing ground based on oceanography satellite image data in fisheries management area (FMA) 713, Torani: JFMarSci.**3**, 2, 51–60 (2020). https://oceancolor.gsfc.nasa.gov/
2. A.R.S. Putri, M. Zainuddin, Application of remotely sensed satellite data to identify Skipjack Tuna distributions and abundance in the coastal waters of Bone Gulf, in IOP Conference Series: Earth and Environmental Science, Institute of Physics Publishing, Mar. (2019). https://doi.org/10.1088/1755-1315/241/1/012012

3. U. Tangke, F.D. Silooy, Rochmady, Z. Saing, Sea surface temperature and chlorophyll-a condition of skipjack tuna (Katsuwonus Pelamis) catching area in Ternate Island marine waters, in Journal of Physics: Conference Series, Institute of Physics Publishing, May (2020). https://doi.org/10.1088/1742-6596/1517/1/012039
4. M. Zainuddin, S. Safruddin, A. Farhum, B. Budimawan, R. Hidayat, M.B. Selamat, Satellite-based ocean color and thermal signatures defining habitat hotspots and the movement pattern for commercial skipjack tuna in Indonesia Fisheries Management Area 713, Western Tropical Pacific. Remote Sens (Basel). **15**, 5 (2023) https://doi.org/10.3390/rs15051268
5. T. Elepathage, T. Danling, Detection of pelagic habitats and abundance of skipjack tuna in relation to the environment in the Indian Ocean around Sri Lanka. Biological Oceanography at PeerJ Preprints. (2019). https://doi.org/10.7287/peerj.preprints.27663v2
6. R. Hidayat, M. Zainuddin, A. Mallawa, M.A. Mustapha, A.R.S. Putri. Mapping spatial-temporal skipjack tuna habitat as a reference for fish aggregating devices (Fads) settings in Makassar strait, Indonesia. Biodiversitas. **22**, 9, 3637–3647 (2021). https://doi.org/10.13057/biodiv/d220905
7. X.H. Wu, Y. Chang, T.Y. Liao, M.M. Ding, C.C. Ke, Real-time multi-month forecasting of skipjack tuna (*Katsuwonus pelamis*) habitat in the western and central Pacific Ocean for improved fishing efficiency and fisheries management. ICES Journal of Marine Science. **80**, 10, 2490–2503 (2023). https://doi.org/10.1093/icesjms/fsad159
8. Safruddin, R. Hidayat, S. A. Farhum, and M. Zainuddin, The use of statistical models in identifying skipjack tuna habitat characteristics during the Southeast Monsoon in the Bone Gulf, Indonesia. Biodiversitas **23**, 4, 2231–2237 (2022) https://doi.org/10.13057/biodiv/d230459
9. V.K. Yadav, S. Jahageerdar, J. Adinarayana, Comparison between different modeling techniques for assessing the role of environmental variables in predicting the catches of major pelagic fishes off India's north-west coast. (2022)
10. J.E. Salazar, I.F. Benavides, C.V. Portilla Cabrera, A I. Guzmán, J.J. Selvaraj, Generalized additive models with delayed effects and spatial autocorrelation patterns to improve the spatiotemporal prediction of the skipjack (*Katsuwonus pelamis*) distribution in the Colombian Pacific Ocean. Reg. Stud. Mar. Sci. **45**, (2021). https://doi.org/10.1016/j.rsma.2021.101829
11. H. Wibisana, B.M. Sukotjo, U. Lasminto, Analysis of correlation and mapping of chlorophyll-a concentrations and sea surface temperatures in coastal areas based on terra MODIS satellite image data. Makara Journal of Technology. **25**, 1, 30 (2021). https://doi.org/10.7454/mst.v25i1.3810
12. J.H. Kim, D.H. Lee, J.H. Kang, Associating the spatial properties of a watershed with downstream Chl-a concentration using spatial analysis and generalized additive models. Water Res. **154**, 387–401 (2019). https://doi.org/10.1016/j.watres.2019.02.010
13. R. Simard, P. L'ecuyer, Computing the Two-Sided Kolmogorov-Smirnov Distribution. (2011). http://www.jstatsoft.org/
14. P. Govender, V. Sivakumar, Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). Elsevier B.V. (2020). https://doi.org/10.1016/j.apr.2019.09.009
15. M. Lakićević, Creating maps in r (case study: national park 'Fruška Gora'). Contemporary Agriculture. **70**, 1–2, 41–45 (2021). https://doi.org/10.2478/contagri-2021-0008

16. B.K. Kilinç, H.D. Asfha, Selection of Splines Models in the Presence of Multicollinearity and Outliers. Turkiye Klinikleri Journal of Biostatistics. **12**, 2, 183–192 (2020). https://doi.org/10.5336/biostatic.2020-74122
17. B. Wang, Z. Mao, Outlier detection based on a dynamic ensemble model: Applied to process monitoring. Information Fusion, **51**, 244–258 (2019). https://doi.org/10.1016/j.inffus.2019.02.006
18. H. Sun, Y. Cui, H. Wang, H. Liu, T. Wang, Comparison of methods for the detection of outliers and associated biomarkers in mislabeled omics data. BMC Bioinformatics. **21**, 1, (2020). https://doi.org/10.1186/s12859-020-03653-9
19. G. Manimannan, M. Salomi, R.L. Priya, R. Saranraj, Detecting outliers using R package in fitting data with linear and nonlinear regression models. International Journal of Scientific and Innovative Mathematical Research. **8**, 4 (2020). https://doi.org/10.20431/2347-3142.0804001
20. M. Awawdeh, T. Faisal, A. Bashir, A. Sheikh, Application of Outlier Detection using Re-Weighted Least Squares and R-squared for IoT Extracted Data. in *2019* Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates: IEEE, (2019), 1–6 https://doi.org/10.1109/ICASET.2019.8714261
21. A.A. Shehadeh, S.M. Alwadi, M.I. Almaharmeh, Detecting and analysing possible outliers in global stock market returns. Cogent Economics and Finance. **10**, 1 (2022). https://doi.org/10.1080/23322039.2022.2066762
22. A. Fitrianto, W.Z.A. Wan Muhamad, S. Kriswan, B. Susetyo, Comparing Outlier Detection Methods using Boxplot Generalized Extreme Studentized Deviate and Sequential Fences. Aceh International Journal of Science and Technology. **11**, 1, 38–45 (2022). https://doi.org/10.13170/aijst.11.1.23809
23. A.G. O'Carroll, E.M. Armstrong, H. Beggs, M. Bouali, K.S. Casey, G.K. Corlett, Observational needs of sea surface temperature. Frontiers Media S.A. (2019). https://doi.org/10.3389/fmars.2019.00420.
24. A. Pisano, S. Marullo, V. Artale, F. Falcini, C. Yang, F.E. Leonelli, New evidence of Mediterranean climate change and variability from Sea Surface Temperature observations. Remote Sens. (Basel). **12**, 1 (2020). https://doi.org/10.3390/RS12010132
25. J. Karch, Improving on adjusted R-squared. Collabra Psychol. **6**, 1 (2020). https://doi.org/10.1525/collabra.343
26. A. Murari, E. Peluso, F. Cianfrani, P. Gaudio, M. Lungaroni, On the use of entropy to improve model selection criteria. Entropy. **21**, 4 (2019). https://doi.org/10.3390/e21040394
27. R. Salmerón, C. García, J. García, multiColl: An R package to detect multicollinearity. arXiv: Computation. (2019). https://doi.org/10.48550/arXiv.1910.14590
28. J.H. Kim, Multicollinearity and misleading statistical results. Korean J. Anesthesiol. **72**, 6, 558–569 (2019). https://doi.org/10.4097/kja.19087
29. N.A.M.R. Senaviratna, T.M.J.A. Cooray, Diagnosing multicollinearity of logistic regression model. Asian Journal of Probability and Statistics, 1–9 (2019). https://doi.org/10.9734/ajpas/2019/v5i230132
30. R. Raidou, E. Gröller, M. Eisemann, Relaxing dense scatter plots with pixel-based mappings. IEEE Trans. Vis. Comput. Graph. **25**, 2205–2216 (2019). https://doi.org/10.1109/TVCG.2019.2903956
31. Q.V. Nguyen, M. Lin Huang, S. Simoff, Enhancing Scatter-plots with Start-plots for visualising multi-dimensional data. in Proceedings of the International Conference on

Information Visualisation, Institute of Electrical and Electronics Engineers Inc., Sep. 80–85 (2020). https://doi.org/10.1109/IV51561.2020.00023

32. F. Meng, J. Gu, L.E. Wang, Z. Qin, M. Gao, J. Chen, A quantitative model based on grey theory for sea surface temperature prediction. Front Environ Sci. **10**, (2022). https://doi.org/10.3389/fenvs.2022.1014856
33. W. Shao, T. Jiang, X. Jiang, Y. Zhang, W. Zhou, Evaluation of sea surface winds and waves retrieved from the Chinese HY-2B data. IEEE J. Sel. Top. Appl. Earth. Obs. Remote Sens. **14**, 9624–9635 (2021). https://doi.org/10.1109/JSTARS.2021.3112760
34. H. Han, C. Yang, H. Zhang, Z. Fang, B. Jiang, B. Su, Environment variables affect CPUE and spatial distribution of fishing grounds on the light falling gear fishery in the northwest Indian Ocean at different time scales. Front. Mar. Sci. **9**, (2022). https://doi.org/10.3389/fmars.2022.939334
35. Y. Feng, P. Wu, X. Tong, P. Li, R. Wang, Y. Zhou, The effects of factor generalization scales on the reproduction of dynamic urban growth. Geo-Spatial Information Science. **25**, 3, 457–475 (2022). https://doi.org/10.1080/10095020.2022.2025748
36. C. Hua, Q. Zhu, Y. Shi, Y. Liu, Comparative analysis of CPUE standardization of Chinese Pacific saury (Cololabis saira) fishery based on GLM and GAM. Acta Oceanologica Sinica, **38**, 10, 100–110 (2019). https://doi.org/10.1007/s13131-019-1486-3
37. N. Langrené, X. Warin, Fast multivariate empirical cumulative distribution function with connection to kernel density estimation. Comput Stat. Data Anal. **162**, (2021). https://doi.org/10.1016/j.csda.2021.107267
38. J. Ren, A.G. Bushmakin, P.R. Cislo, L. Abraham, J.C. Cappelleri, R.H. Dworkin, Meaningful within-patient change for clinical outcome assessments: model-based approach versus cumulative distribution functions. J. Biopharm. Stat. (2023). https://doi.org/10.1080/10543406.2023.2281575
39. Z. Cong, L. Chu, Y. Yang, J. Pei, Comprehensible Counterfactual Explanation on Kolmogorov-Smirnov Test. Proc. VLDB Endow. **14**, 1583–1596 (2021). https://doi.org/10.14778/3461535.3461546
40. Y. Hong, O. Linton, B. Mccabe, J. Sun, S. Wang, Kolmogorov-Smirnov type testing for structural breaks: A new adjusted-range based self-normalization approach. SSRN (2023). https://dx.doi.org/10.2139/ssrn.3850894
41. R. Safruddin, Hidayat, M. Zainuddin, Skipjack tuna fishing ground based on oceanography satellite image data in fisheries management area (FMA) 713. Torani: JFMarSci. 3, 2, 51–60 (2020). https://oceancolor.gsfc.nasa.gov/
42. U. Tangke, F.D. Silooy, Rochmady, Z. Saing, Sea surface temperature and chlorophyll-a condition of skipjack tuna (*Katsuwonus Pelamis*) catching area in Ternate Island marine waters, in Journal of Physics: Conference Series, Institute of Physics Publishing, May (2020). https://doi.org/10.1088/1742-6596/1517/1/012039
43. W.A.L. Ningsih, W.A. Lestariningsih, S. Heltria, M.H.I. Khaldun, Analysis of the relationship between chlorophyll-a and sea surface temperature on marine capture fisheries production in Indonesia: 2018, in IOP Conference Series: Earth and Environmental Science, IOP Publishing Ltd, Dec. (2021). https://doi.org/10.1088/1755-1315/944/1/012057