Assessing catch rate dynamics of Yellowfin Tuna in Sri Lanka's longline fishery: Linear and non-linear approaches

By Sisira Haputhantri* and Gayan Jayasinghe

National Aquatic Resources Research and Development Agency (NARA), Crow Island, Colombo 15, Sri Lanka.

*Corresponding author

Abstract

Yellowfin tuna (*Thunnus albacares*) plays a central role in Sri Lanka's longline fishery, contributing a higher share of both catch and economic return. However, catch rates are highly variable, reflecting the combined influence of fishing practices and possible changes in stock abundance. This study examined catch per unit effort (CPUE) dynamics using port sampling data collected between 2013 and 2023. Explanatory variables included temporal (month), operational (vessel length, days fished, number of hooks, and gear operation time), and spatial (fishing area) variables. Multiple linear regression (MLR) and generalized additive models (GAMs) were employed to describe both linear and non-linear relationships, while a Random Forest model was utilized to assess variable importance. The analysis showed that CPUE varied significantly with vessel length and fishing effort, indicating that larger vessels operating longer trips with more hooks tended to achieve higher catch rates. Both linear and non-linear models indicated that fishing effort and operational characteristics accounted for a significant portion of the observed variability in CPUE. These findings highlight the value of integrating statistical and machine learning approaches to better understand catch rate dynamics in the yellowfin tuna longline fishery of Sri Lanka.

Key words: Indian Ocean, Sri Lanka, yellowfin tuna, longline, CPUE

Introduction

The yellowfin tuna (*Thunnus albacares*) is one of the most valuable tuna species exploited in the Indian Ocean, contributing significantly to Sri Lanka's fisheries sector through both domestic consumption and export earnings (World Bank, 2022; Jayaweera, 2022). Among the various gears employed, the longline fishery has become increasingly important for targeting tuna species, in particular, yellowfin tuna. Understanding catch rate dynamics in this fishery is crucial for evaluating fishing performance, identifying changes in resource availability, and informing evidence-based management decisions (Yeh and Chang, 2010).

Catch Per Unit Effort (CPUE) is commonly used as an indicator of relative abundance in fisheries (Maunder and Punt, 2004). However, CPUE values are not solely determined by fish stock abundance; they are also influenced by factors such as vessel type, trip duration, number of fishing days, gear configuration, fishing grounds, and seasonal conditions (Hoyle et al., 2017; Haputhantri et al., 2023). These operational and environmental factors can mask underlying trends, making it necessary to explore modeling approaches that can separate their effects and identify the main drivers of catch rates. This study examines yellowfin tuna CPUE in Sri Lanka's longline fishery from 2013 to 2023 using both linear and non-linear methods. Multiple Linear Regression (MLR) and Generalized Additive Models (GAM) are employed alongside selected machine learning techniques to analyze the relationship between CPUE and variables related to fishing effort and practices. By comparing the performance of linear and non-linear models, this work provides insights into the factors affecting catch rate dynamics and evaluates the potential of advanced modeling frameworks for fisheries data analysis. The findings contribute to a better understanding of fishery operations and provide valuable guidance for management and regional stock assessment efforts.

Materials and methods

Catch and effort data from the Sri Lankan longline fishery targeting yellowfin tuna from 2013 to 2023 were used for analysis. The data comes from port sampling records maintained in the PELAGOS database of the National Aquatic Resources Research and Development Agency

(NARA), Sri Lanka. To ensure consistency and reliability, landings from vessels using mixed gear combinations and records of partial unloads were excluded from the analysis.

The dependent variable was the catch per unit effort (CPUE, expressed in kg per boat per trip). Predictor variables considered were vessel length, the number of hooks deployed, days fished, gear operation time (day/night/both), IOTC 5° area, and month. All the statistical analyses were carried out using R software version 4.1.3 (R Development Core Team, 2025).

Multiple Linear Regression (MLR)

MLR is a parametric approach that models the linear relationship between a continuous response variable (CPUE) and explanatory variables such as boat length, number of hooks, and trip duration. It assumes additive and linear effects of predictors on the response.

An MLR model was applied to assess the relationship between CPUE and explanatory variables. Before modeling, CPUE was log-transformed. Model selection was carried out using stepwise procedures based on Akaike Information Criterion (AIC). Multicollinearity was evaluated using the Variance Inflation Factor (VIF), and residual diagnostics were performed to assess model assumptions (normality, homoscedasticity, independence).

Generalized Additive Model (GAM)

GAM is an extension of linear models that allows for nonlinear relationships between the response variable and predictors through smooth functions. It is particularly useful for capturing seasonal or effort-related nonlinear trends in CPUE.

A Generalized Additive Model (GAM) was applied to examine the influence of vessel, gear, temporal, and spatial factors on CPUE. The model used a Gamma distribution with a log link to account for the right-skewed nature of CPUE data. The response variable was CPUE (kg/boat/trip). Explanatory variables included:

Continuous variables: Vessel length, number of hooks, and days fished.

Categorical variables: Fishing time (day, night, both) and IOTC 5° area.

Temporal variable: Month.

Smooth terms (s ( )) were used for boat length, number of hooks, and days fished to allow for potential nonlinear relationships. Model fit was evaluated using adjusted R^2, deviance explained, and residual diagnostics.

Random Forest (RF) regression model

RF is a non-parametric, ensemble machine learning method that builds many decision trees and averages their predictions. It can model complex, nonlinear interactions among predictors without requiring assumptions about the underlying data distribution.

An RF regression model was applied to examine the influence of fishing effort and vessel-related variables on the CPUE. The dataset included the following predictor variables: vessel length, fishing time (day, night, both), IOTC area, number of hooks, and days fished. Before model fitting, the dataset was screened for missing values, and incomplete records were excluded to ensure consistency.

The Random Forest model was implemented using the randomForest package in R with 500 trees (ntree = 500) and default tuning parameters. Variable importance was assessed using two measures: (i) percentage increase in mean squared error (%IncMSE), which reflects the predictive accuracy loss when a variable is permuted, and (ii) increase in node purity (IncNodePurity), which indicates the contribution of each variable to reducing residual variance.

Partial Dependence Plots (PDPs) were generated for the most influential predictors to visualize their marginal effects on CPUE while averaging over the effects of other variables. This allowed for the interpretation of non-linear relationships between CPUE and key fishing effort indicators.

Results and discussion

Data visualization

Box plots of continuous predictors (vessel length, number of hooks, and days fished) show clear patterns in CPUE (Figure 1). Larger vessels, higher numbers of hooks, and longer fishing durations were generally associated with higher catch rates, though variability and outliers were also evident.

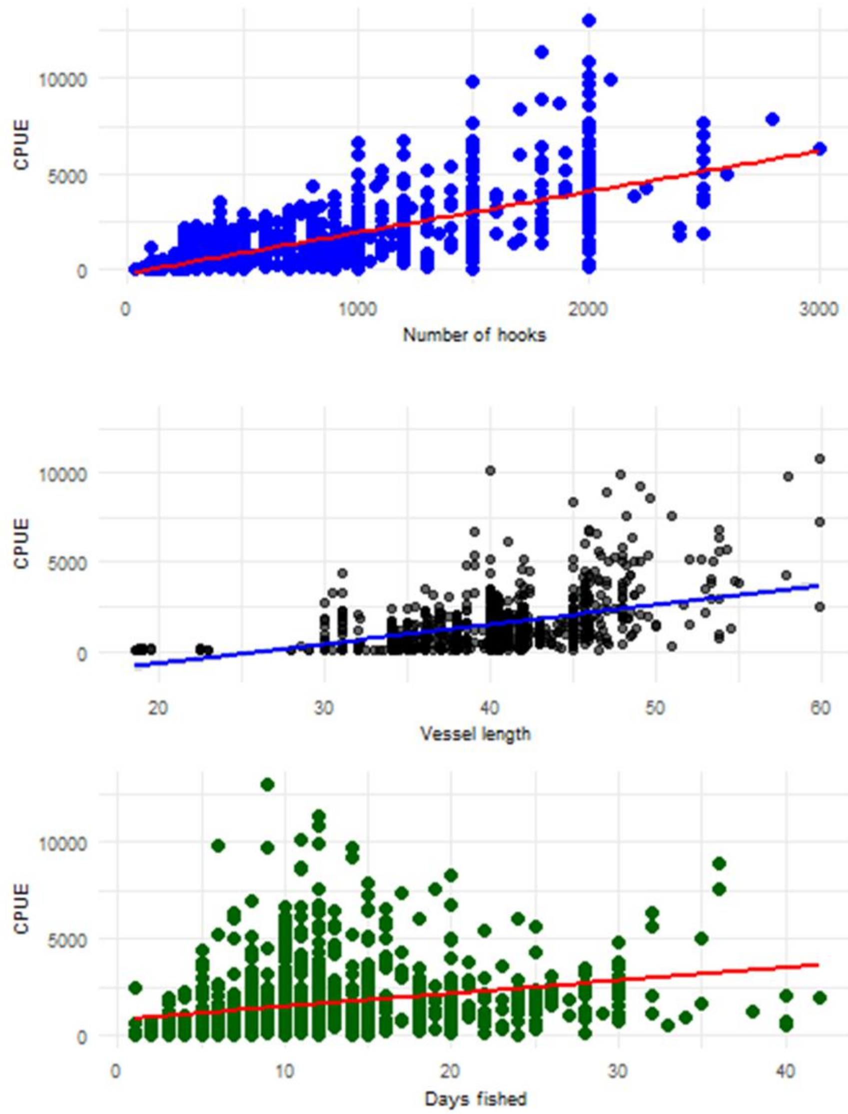

Figure 1. Box plot for the distribution of continuous predictor variables used in the CPUE analysis

Box plots of categorical predictors highlight differences in CPUE across vessel type, fishing time, and fishing area (Figure 2). Night operations tended to yield higher catches than day sets, while variation was also observed across IOTC fishing areas, suggesting spatial and temporal influences on yellowfin CPUE.

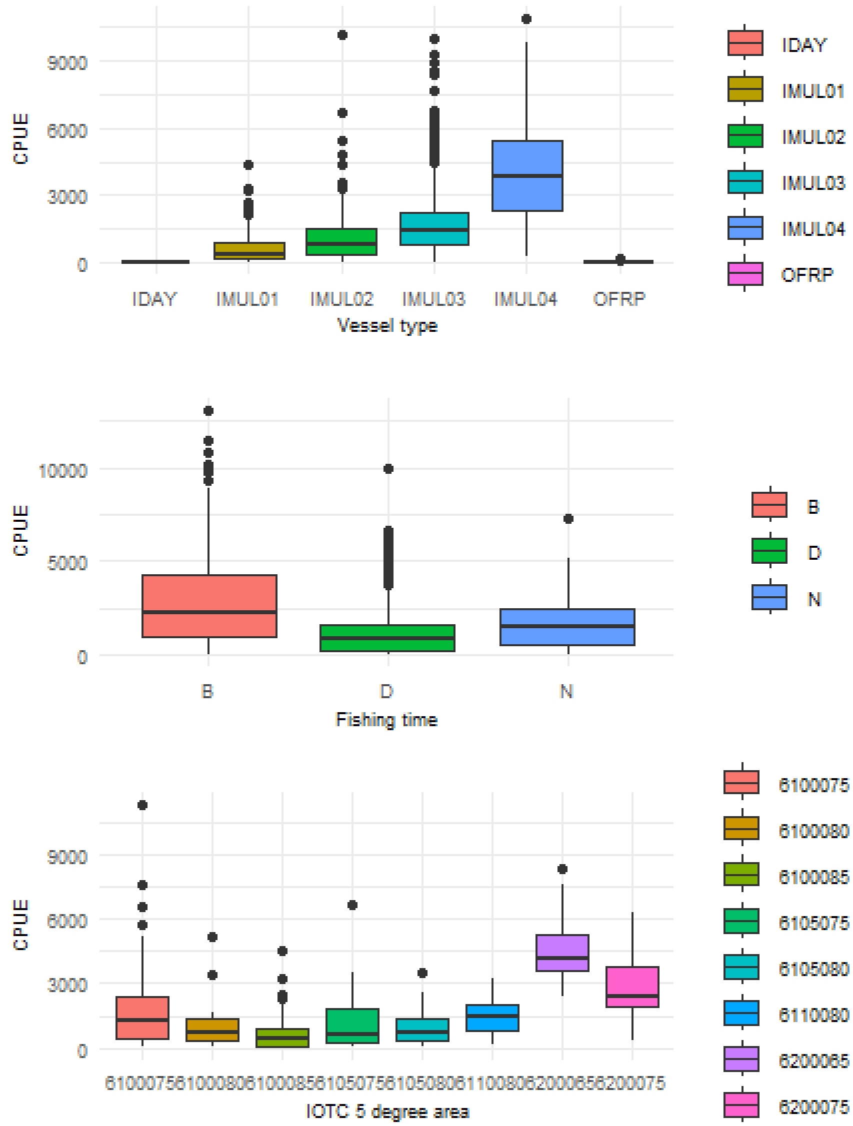

Figure 2. Box plot for the distribution of categorical predictor variables used in the CPUE analysis

MLR model

Table 1 presents the regression coefficients of the final MLR model developed to examine the factors influencing yellowfin tuna CPUE in the Sri Lankan longline fishery. The model explained about 52% of the total variance in CPUE (adjusted $R^2 = 0.489$, $p < 0.001$). Vessel length, number of hooks deployed, and days fished were all positively and significantly associated with CPUE, indicating that larger vessels operating longer trips with more hooks tended to achieve higher catch rates. Among the spatial factors, fishing in IOTC area 6100085 was associated with a significantly lower CPUE compared to the reference area, while other areas showed negative but non-significant effects. Month also had a significant negative effect, suggesting seasonal variability in catch rates.

The diagnostic plots of the final MLR confirmed that residuals were approximately normally distributed, heteroscedasticity was minimal, and variance inflation factors (VIF) were below 5, indicating no strong multicollinearity (Figure 3).

Table 1. Regression coefficients of the final MLR model used for CPUE analysis

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	4.8981898	0.6615898	7.404	1.69e-12	***
Boat length	0.0440782	0.0125689	3.507	0.000531	***
Number of hooks	0.0007057	0.0001510	4.674	4.67e-06	***
Days fished	0.0369238	0.0080423	4.591	6.76e-06	***
IOTC.area6100065	-0.0655334	0.4436013	-0.148	0.882666	
IOTC.area6100075	-0.5263315	0.3876322	-1.358	0.175657	
IOTC.area6100080	-0.6607307	0.4596187	-1.438	0.151716	
IOTC.area6100085	-0.8457109	0.4009569	-2.109	0.035845	*
IOTC.area6105075	-0.7780837	0.4095525	-1.900	0.058520	.
IOTC.area6105080	-0.7616405	0.4213074	-1.808	0.071750	.
IOTC.area6110080	-0.4445284	0.4391323	-1.012	0.312307	
IOTC.area6110085	-0.6009549	0.4479714	-1.342	0.180885	
IOTC.area6115085	-0.7170109	0.4841168	-1.481	0.139754	
IOTC.area6200060	-0.1352358	0.4418702	-0.306	0.759800	
IOTC.area6200065	-0.1362781	0.3956430	-0.344	0.730778	
IOTC.area6200075	-0.3196707	0.3895400	-0.821	0.412577	
IOTC.area6205065	-0.2215876	0.4722268	-0.469	0.639277	
IOTC.area6205075	-0.5079135	0.4211430	-1.206	0.228859	
Month	-0.0405854	0.0169952	-2.388	0.017625	*

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Multiple R-squared: 0.521, Adjusted R-squared: 0.489

```
F-statistic: 16.31,  p-value: < 2.2e-16
```

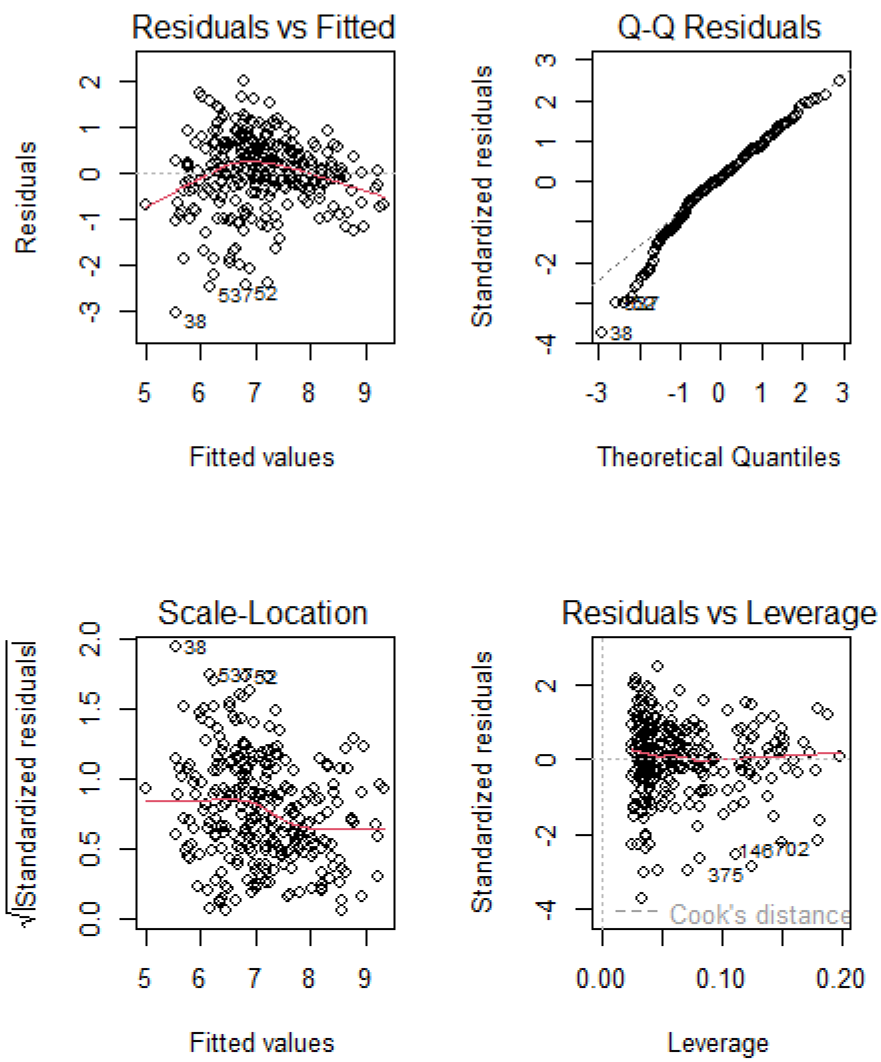

Figure 3. Diagnostic plots of the MLR model: (a) Residuals vs Fitted; (b) Normal Q–Q plot; (c) Scale–Location and (d) Residuals vs Leverage

GAM model

The fitted GAM explained 53% of the deviance in CPUE, with an adjusted R2 of 0.567**,** indicating a moderate to good model fit (Table 2). The parametric component of the GAM showed that several categorical factors had a significant influence on the catch rates. Among these, month exhibited a marked seasonal pattern, with certain months contributing significantly higher or lower CPUE compared to the reference month, reflecting the seasonal availability of yellowfin tuna. The IOTC area variable was also significant, indicating that spatial differences in fishing grounds strongly affect catch rates.

Among the parametric terms, fishing time did not show a statistically significant effect on CPUE. Both daytime ($p = 0.35004$) and night-time operations ($p = 0.22272$) were not significantly different from the reference category. However, the lower p-value for nighttime fishing may indicate a weak trend towards higher CPUE compared with the reference. The absence of a statistically significant effect of fishing time (day/night) on CPUE may partly reflect limitations of the port-sampling data. In the current dataset, fishing time is reported at the trip level rather than recorded set-by-set at sea, which could introduce misclassification or averaging effects and doubtful real differences in catch rates between day and night operations.

Among the smooth predictors, boat length ($edf=2.213$, $F=5.098$, $p= 0.00257$) number of hooks ($edf=1.734$, $F=10.605$, $p<0.001$) and days fished ($edf=4.432$, $F=9.642$, $p<0.001$) showed significant non-linear relationships with CPUE. These results indicate that the response variable is strongly influenced by non-linear relationships with boat length, the number of hooks, and days fished.

The GAM smooth plots illustrate how boat length**,** number of hooks**,** and days fished each influence the response variable (CPUE), after controlling for other factors (Figure 4). For boat length, the predicted effect increases steadily from approximately 20 Feet to around 50 Feet, indicating that larger vessels tend to achieve higher predicted catch or CPUE. Beyond 50 Feet, the effect levels off, suggesting little additional benefit from even longer vessels. Wider confidence intervals at the extremes reflect fewer data points.

For the number of hooks, the effect increases gently and almost linearly across the observed range, implying that using more hooks per operation is positively associated with the response. In contrast, days fished shows a non-linear pattern: the response rises sharply during the first few days, peaks around 10 days, and then levels off, with a slight decline beyond 25–30 days and increasing uncertainty. Overall, these results highlight how the model captures non-linear relationships and suggest thresholds where increasing effort no longer improves outcomes, valuable information for optimizing fishing practices and management decisions.

Table 2. Parametric and smooth term estimates from the Generalized Additive Model (GAM) for Yellowfin tuna CPUE

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.25838	0.32177	25.665	< 2e-16 ***
Month	-0.03815	0.01424	-2.679	0.00784 **
Fishing.timeD	-0.22478	0.24011	-0.936	0.35004
Fishing.timeN	0.13627	0.11149	1.222	0.22272
IOTC.area6100065	-0.26556	0.36911	-0.719	0.47249
IOTC.area6100075	-0.64983	0.33317	-1.950	0.05219 .
IOTC.area6100080	-0.94877	0.38804	-2.445	0.01514 *
IOTC.area6100085	-0.71628	0.35266	-2.031	0.04326 *
IOTC.area6105075	-0.95120	0.35165	-2.705	0.00728 **
IOTC.area6105080	-1.06993	0.36281	-2.949	0.00347 **
IOTC.area6110080	-0.65220	0.38142	-1.710	0.08846 .
IOTC.area6110085	-0.80911	0.38966	-2.076	0.03882 *
IOTC.area6115085	-0.79240	0.42598	-1.860	0.06398 .
IOTC.area6200060	-0.17779	0.36964	-0.481	0.63093
IOTC.area6200065	-0.28646	0.32839	-0.872	0.38382
IOTC.area6200075	-0.59945	0.32783	-1.829	0.06860 .
IOTC.area6205065	-0.47108	0.39306	-1.199	0.23180
IOTC.area6205075	-0.77535	0.35244	-2.200	0.02868 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

	edf	Ref.df	F	p-value
s(Boat.length)	2.213	2.878	5.098	0.00257 **
s(Number.of.hooks)	1.734	2.186	10.605	2.22e-05 ***
s(Days.fished)	4.432	5.445	9.642	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) = 0.567 Deviance explained = 53%
GCV = 0.5757 Scale est. = 0.4629

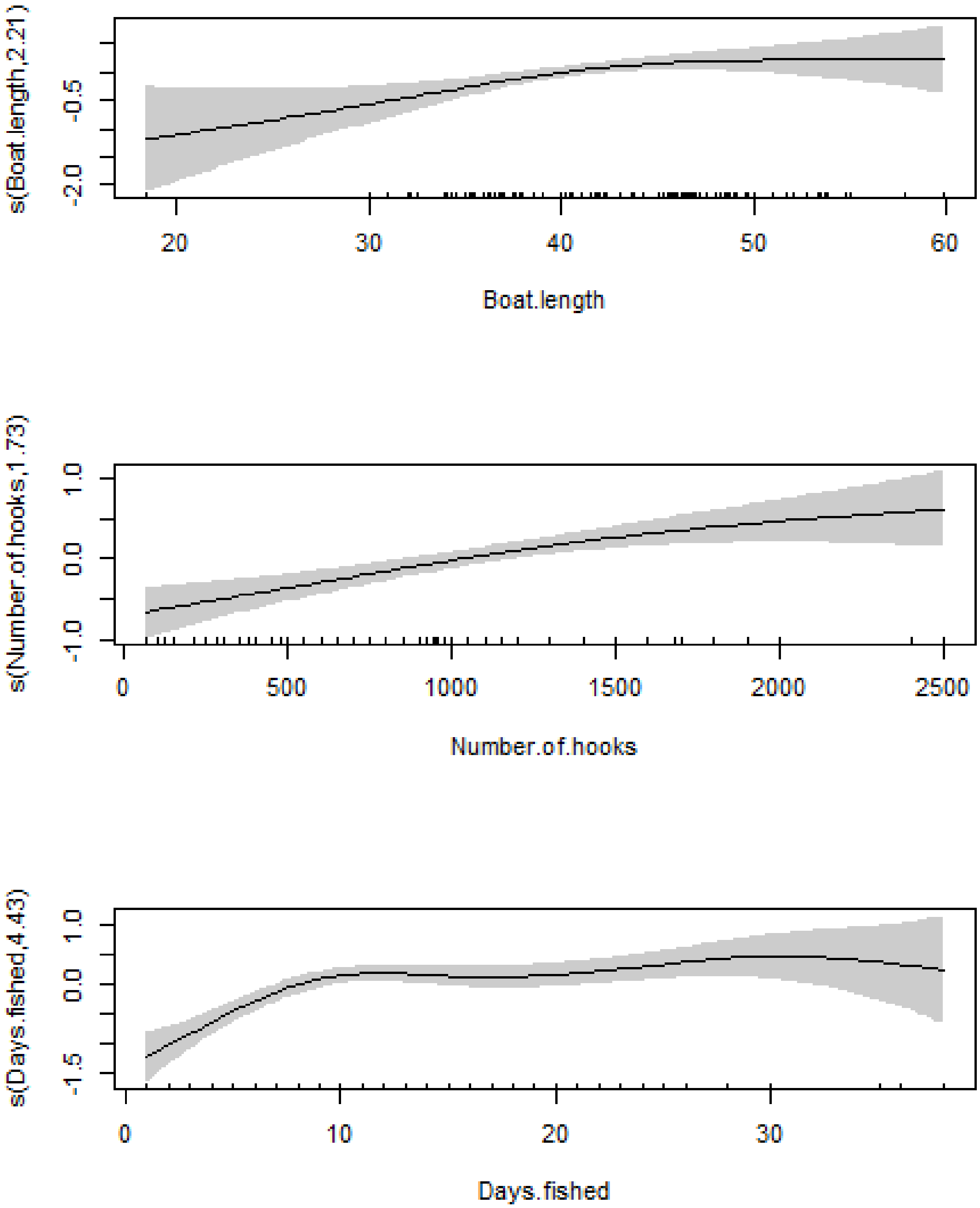

Figure 4. Partial smooth effects of model predictors on Yellowfin tuna CPUE

Random Forest (RF) regression model

The Random Forest (RF) model was applied to evaluate the relative importance of predictor variables in explaining the variation in CPUE. The variable importance plot (Figure 5) indicates that number of hooks was the most influential predictor, followed by boat length and IOTC area, while days fished, and fishing time contributed less to the model performance.

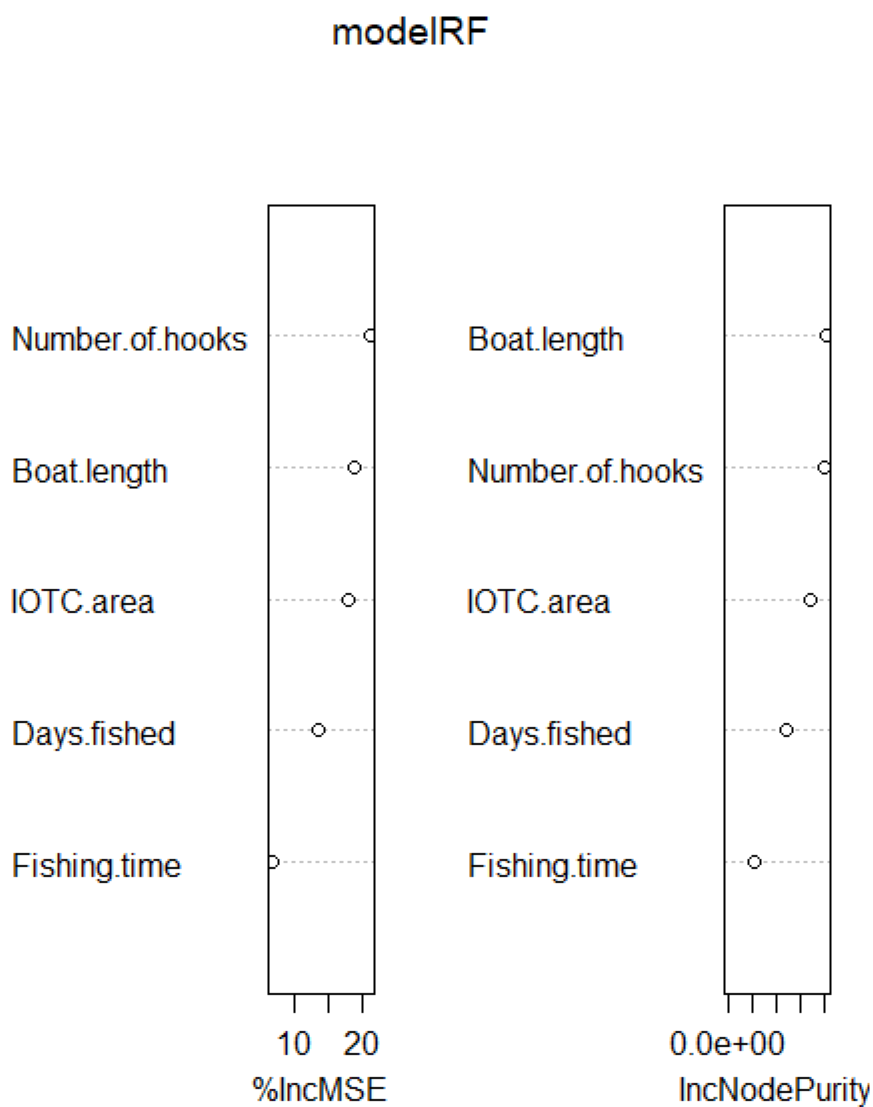

Figure 5. Random forest variable importance plot for CPUE prediction

The partial dependence plots (Figure 6) depict the marginal effects of key predictors on CPUE. For the number of hooks, a strong positive association with CPUE was evident, with catch rates increasing steadily as hook numbers rose. This pattern indicates that greater fishing effort, through the deployment of more hooks, generally enhances catch rates, although the effect appears to level off at very high hook numbers, suggesting diminishing returns at extreme effort levels.
For boat length, CPUE increased with vessel size, particularly for vessels exceeding 40 Feet in length. This finding highlights the greater operational capacity, fishing power, and resource exploitation efficiency associated with larger vessels. Finally, the IOTC area effect showed little to no variation in CPUE across areas, implying that spatial differences in fishing grounds, within the range of the available data, exert limited influence on CPUE in the model.

The partial dependence plots provide an additional insight into how individual predictors influence CPUE after accounting for the effects of other variables. The strong positive relationship between CPUE and the number of hooks reflects the direct role of fishing effort in determining catch rates. However, the apparent plateau at higher hook numbers suggests that beyond a certain threshold, additional hooks yield only marginal gains, indicating diminishing returns to effort. Similarly, the positive effect of boat length highlights the operational advantages of larger vessels, which may include greater fishing power, larger storage capacity, and the ability to access more distant or productive grounds.

In contrast, the effect of the IOTC area on CPUE was weak or absent. This lack of spatial variation may not necessarily indicate uniform catch rates across areas, but rather could reflect limitations in the port sampling data. In many cases, fishermen report aggregated catch and effort data for an entire trip rather than at the set-by-set level. This practice can obscure the true spatial distribution of fishing effort, especially when vessels operate in more than one IOTC area during a single trip. Consequently, CPUE values may represent a composite of multiple areas rather than a single location, diluting any area-specific signal. Improving spatial resolution in data collection, such as through set-by-set reporting or electronic logbooks, would help clarify the true impact of fishing grounds on CPUE.

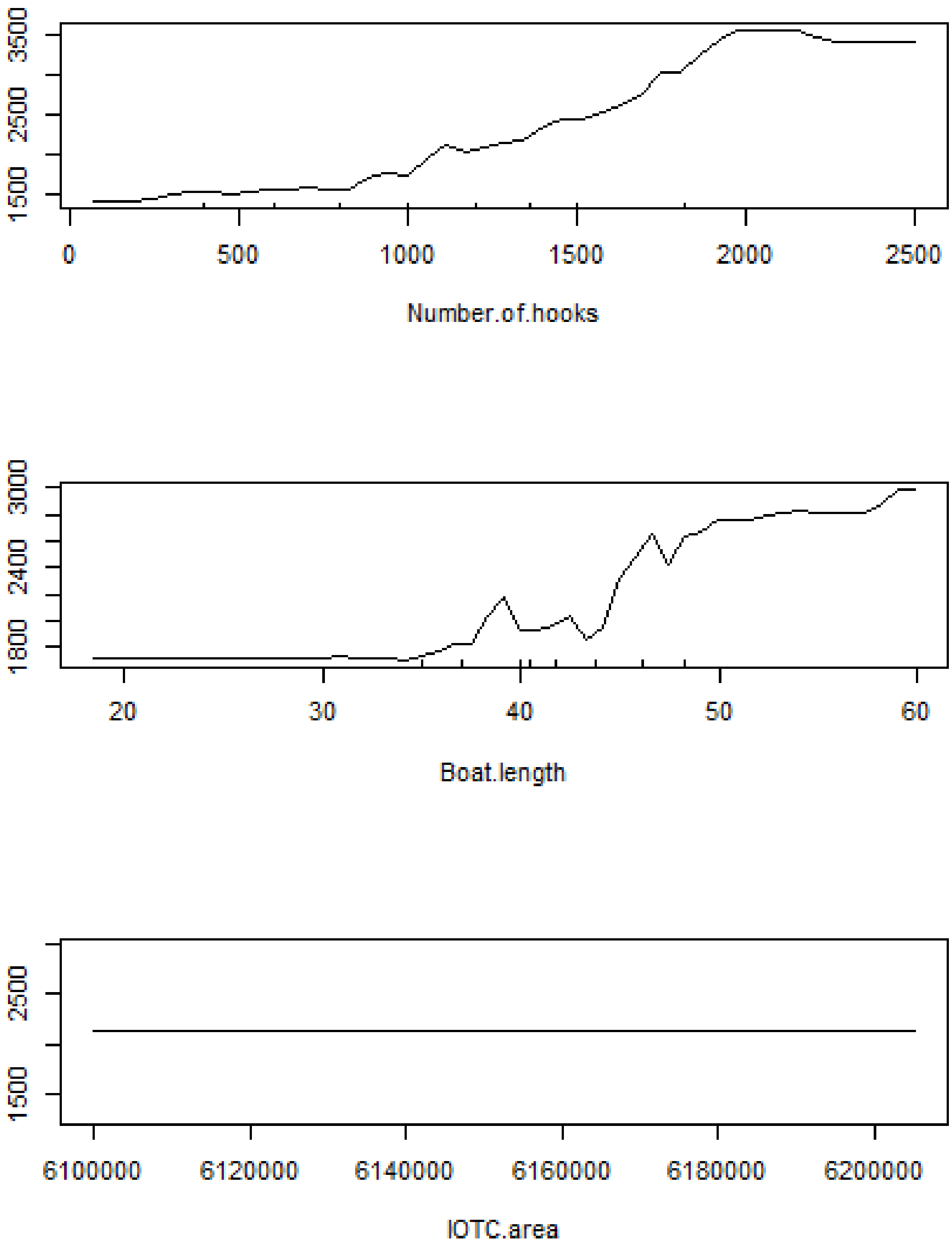

Figure 6. Partial dependence plots of key predictors in the Random Forest Model

Conclusion

The combined findings from data visualization, MLR, GAM, and Random Forest analyses show that vessel size, number of hooks deployed, and days fished are the dominant factors influencing yellowfin tuna CPUE in the Sri Lankan longline fishery. Across all approaches, larger vessels and higher hook numbers consistently produced higher catch rates, while the effect of fishing duration showed a non-linear pattern with diminishing returns beyond about 10–20 days at sea. Spatial and temporal influences were evident, with month having a significant seasonal effect and some IOTC areas producing lower CPUE, although the strength of area effects was weakened by the coarse trip-level reporting used in the current dataset. Fishing time (day vs night) showed no significant effect in either the MLR or GAM, or RF, again likely reflecting data limitations rather than the absence of a true difference.

Overall, these results highlight both the power and the limits of the available port-sampling data. They highlight the importance of improving data resolution (e.g., set-by-set logbooks or electronic reporting) to better capture spatial and temporal variation, and they point to thresholds where increasing effort yields only marginal gains. Using these findings to guide management: by adjusting vessel effort, the number of hooks used, and trip duration, and by improving data collection, Sri Lanka can create stronger, evidence-based plans to protect and sustain its yellowfin tuna stocks, even as environmental and fishery conditions change.

References

1. Haputhantri S., Jayasinghe, G. and Gunasekara, S. (2023). Accounting for spatial, temporal and operational effects in the Catch Per Unit Effort standardization of Skipjack tuna in tuna drift gillnet fishery in Sri Lanka. 2023. IOTC–2023–WPTT25–06. Twenty-fifth Session of the Indian Ocean Tuna Commission (IOTC) Tropical Tuna Working Party.
2. Hoyle, S. D., Assan, C., Chang, S. T., Fu, D., Govinden, R., Kim, D. N., Lee, S. I., Lucas, J., Matsumoto, T., Satoh, K., Yeh, Y. M., and Kitakado, T. (2017). Collaborative study of tropical tuna CPUE from multiple Indian Ocean longline fleets in 2017. Indian Ocean Tuna

Commission (IOTC). Retrieved from https://www.iotc.org/documents/collaborative-study-tropical-tuna-cpue-multiple-indian-ocean-longline-fleets-2017

3. Jayaweera, D. (2022). *Tuna fisheries in Sri Lanka: Development, management and challenges*. United Nations University Fisheries Training Programme (UNU-FTP) Final Project. Reykjavik, Iceland. Retrieved from https://www.grocentre.is/static/gro/publication/1730/document/DiliniJayaweera22prf.pdf
4. Maunder, M. N., & Punt, A. E. (2004)**.** Standardizing catch and effort data: A review of recent approaches. Fisheries Research, 70(2–3), 141–159. https://doi.org/10.1016/j.fishres.2004.08.002
5. R Development Core Team (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/
6. World Bank. (2022). Priorities for sustainably managing Sri Lanka's marine fisheries, coastal aquaculture, and the ecosystems that support them. Washington, DC: World Bank. Retrieved from https://documents1.worldbank.org/curated/en/308261634198704809/pdf/Priorities-for-Sustainably-Managing-Sri-Lanka-s-Marine-Fisheries-Coastal-Aquaculture-and-the-Ecosystems-that-Support-Them.pdf
7. Yeh, Y.M., and Chang, S.T. (2010). Catch rate standardization for yellowfin tuna (*Thunnus albacares*) in the Indian Ocean. Fisheries Research, 107(2–3), 150–157.