Spatio-temporal Trends of Sailfish (*Istiophorus platypterus*) Catch Rates in the Sri Lankan Tuna Longline Fishery

S.S. Gunasekara, R.P.P.K. Jayasinghe*

* corresponding author (prabathj@nara.ac.lk)

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

Abstract

Sailfish (*Istiophorus platypterus*) is an apex predator inhabiting pelagic waters above thermocline are often caught as by-catch in tuna longline fisheries. Spatial and temporal trends of sailfish catch rates of tuna longline fishery of Sri Lanka in relation to environmental variables were investigated using general linear models (GLM), generalized additive models (GAM) and machine learning algorithms such as Random Forest (RF) and Gradient Boosted Trees (GRF). This analysis utilise logbook data consists of 17625 non-zero catch fishing operations from January 2016 to December 2019 in the northern and central Indian ocean. Spatiotemporal variables included latitude, longitude, month and year of fishing operation. Environmental variables such as sea surface temperature, sea surface salinity, sea surface height, chlorophyll-*a* concentration, eddy kinetic energy, euphotic depth, and depth of mixed layer downloaded from Copernicus Marine Environmental Monitoring Service (CMEMS). A linear regression model with seven environmental variables showed a very low explanatory power (10-fold cross-validation: R^2=0.0016, RMSE=24.05), suggesting that environmental variables may not have a simple linear relationship with catch rates. The GRF model improved prediction accuracy compared to the linear model (R^2 = 10.1%), with sea surface salinity and eddy kinetic energy identified as dominant predictors. The Random Forest model significantly improved the explanation of variability in catch rate (R^2 = 15.3%), highlighting sea surface salinity, eddy kinetic energy, and chlorophyll-*a* as the most influential variables. Cross-validated GAM model of combined spatiotemporal and environmental predicators offered the highest explanatory power, accounting for 19% of the variability (R^2 = 0.19). The study highlights that spatio-temporal and environmental factors are essential drivers of sailfish catch rates, vital information for the sustainable management of the incidental sailfish fishery.

Keywords: Sailfish, Spatio-temporal trends, Tuna longline fishery, Environmental variables, Machine learning

Introduction

Sailfish (*Istiophorus platypterus*) is an apex predator inhabiting pelagic waters above thermocline. They are highly migratory species and have cosmopolitan distribution of tropical waters close to coasts and islands in depths from 0 to 200 m (Nakamura., 1985; FAO-FIGIS., 2001). Sailfish are often caught as by-catch in tuna longline and gillnet fisheries and one of main target species in sport fishery (Hoolihan 2003., Ndegwa & Herrera., 2011).

In Indian ocean, four countries are responsible for 75% of the total Indo-Pacific sailfish catches mainly from north-west Indian Ocean: Iran (30%), Pakistan (18%), India (18%), and Sri Lanka (9%). These nations primarily use gillnets, while India also employs troll lines and Sri Lanka uses longlines (IOTC., 2025). The 2022 stock assessment by the Indian Ocean Tuna Commission (IOTC) reports a 54% exploitation level for sailfish. There is limited reliable information on the catches of this species and no information on the stock structure or growth and mortality in the Indian Ocean (IOTC., 2025).

Methodology

Data Preparation

For the analysis, logbook data from 17,625 non-zero catch operations of sailfish were used. These records, covering the 2016-2019 period, were obtained from longline fishing logbooks provided by the Department of Fisheries and Aquatic Resources in Sri Lanka. Catch per unit effort was calculated fish per 100 hooks.

The environmental parameters for this study, including temperature, salinity, and currents, were sourced from the Global Ocean Physics Analysis and Forecast dataset (product GLOBAL_ANALYSISFORECAST_PHY_001_024) provided by the Copernicus Marine Service (CMEMS, 2025). Chlorophyll-a and euphotic depth data were obtained from multi-satellite merge data product of the Globcolour project (Maritorena et al., 2010).

Eddy kinetic energy (EKE) was derived from the zonal (u) and meridional (v) surface current velocity components using the standard formulation

$$EKE=0.5\times(u^2+v^2)$$

Environmental and fisheries data were compiled into a single dataset of 0.25 degree resolution. The response variable was log-transformed catch per unit effort (log-CPUE), while environmental predictors included sea surface salinity (sal0), sea surface height (ssh), mixed layer depth (mld), chlorophyll-a concentration (ssc), euphotic depth (zsd), sea surface temperature (temp0), and eddy kinetic energy (eke). Observations containing missing values were removed prior to analysis. All variables were standardized where appropriate to ensure comparability across models.

Linear Regression

As a baseline, a multiple linear regression model was fitted with log-CPUE as the dependent variable and all environmental predictors as covariates. This model provided an interpretable framework to identify linear associations between fish abundance and environmental drivers. Model diagnostics were checked to assess assumptions of linearity, homoscedasticity, and residual normality.

Generalized Additive Model (GAM)

To account for potential non-linear relationships between environmental variables and log-CPUE, a Generalized Additive Model (GAM) was applied. Each predictor was fitted with a smoothing spline function. Model performance was evaluated using 10-fold cross-validation, and smooth functions were inspected to identify ecological response curves. GAMs allowed flexible fitting while retaining interpretability in terms of environmental gradients.

Random Forest (RF)

A Random Forest (RF) regression model was implemented to capture complex interactions and non-linear effects. The model was trained with 10-fold cross-validation to reduce overfitting. Predictor importance was quantified based on the reduction in mean squared error, allowing identification of the most influential environmental drivers of CPUE variability. RF provided a robust, non-parametric benchmark against which more structured models (e.g., GAM) could be compared.

Gradient Boosted Trees (GRF)

A Gradient Boosted Regression Trees (GRF) model was applied to further enhance predictive performance. The model was trained using stochastic gradient boosting with 10-fold cross-validation. Hyperparameter tuning was conducted over tree depth, number of trees, learning rate (shrinkage), and minimum node size. A smaller learning rate (0.01–0.05) with a larger number of trees (up to 3000) was explored to balance bias–variance trade-offs. Variable importance was extracted to highlight key predictors. The GRF approach provided a powerful ensemble method capable of modeling non-linear, high-order interactions while maintaining predictive accuracy.

Model Comparison

All models (Linear Regression, GAM, RF, and GRF) were compared using root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) obtained through cross-validation. This allowed assessment of predictive performance across parametric, semi-parametric, and machine learning approaches.

Results

Model Performance

A total of 17,625 non-zero catch operations of sailfish were analyzed in relation to spatiotemporal and environmental predictors. Model performances varied substantially depending on the algorithm applied (Table 1).

The linear model using seven environmental predictors explained a negligible proportion of the variance (10-fold cross-validation: $R^2 = 0.0016$, RMSE = 24.05). This indicates that the relationship between catch rates and environmental conditions is not adequately captured by a simple linear form. GAM model incorporating smooth functions of predictors improved model performance substantially. The GAM explained 19% of the variability in catch rates ($R^2 = 0.19$, RMSE = 0.846), demonstrating that nonlinear effects and interactions between spatiotemporal and environmental variables are important. The GRF model achieved a moderate explanatory power of $R^2 = 10.1\%$. Among the predictors, sea surface salinity (sal0) and eddy kinetic energy (eke) emerged as dominant drivers, followed by mixed layer depth. However, GRF performance was lower than RF and GAM. The RF model provided a stronger predictive performance than GRF, with R^2 = 15.3% and a lower RMSE. Variable importance analysis revealed sea surface salinity, eddy kinetic energy, and chlorophyll-a concentration as the most influential predictors. Overall, GAM offered the highest explanatory power (R^2 = 19%), followed by RF (15.3%), GRF (10.1%), and linear regression (0.16%) (Figure 1).

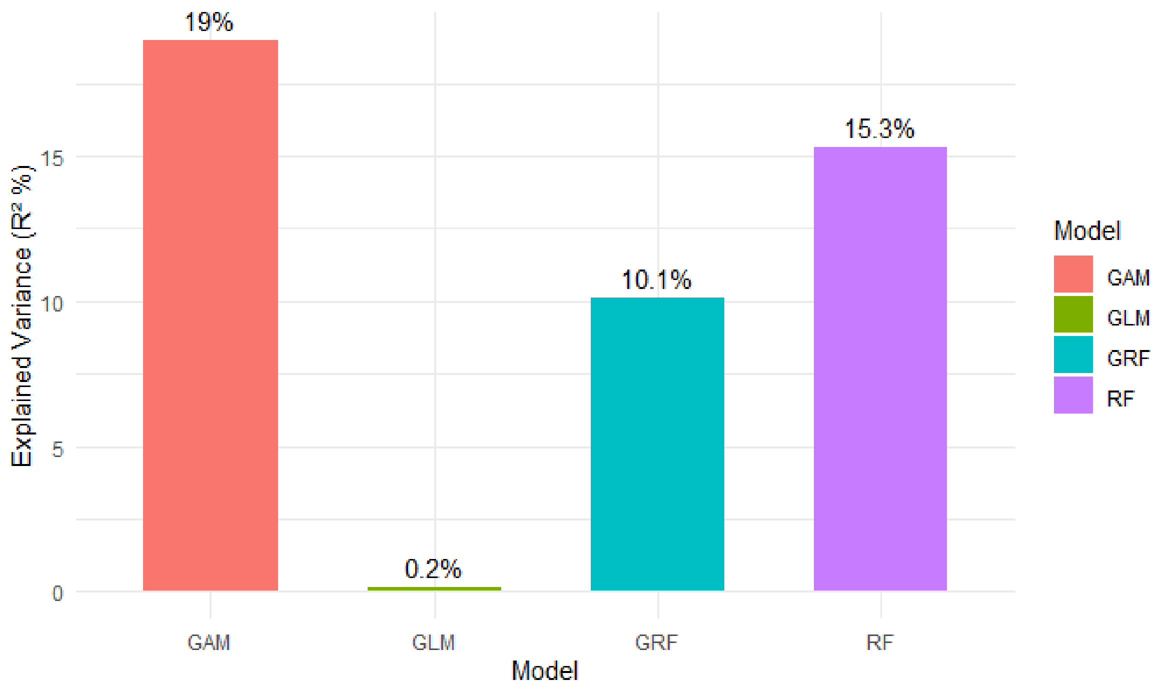

Figure 1: Model comparison for Sailfish CPUE

Across machine learning models (RF and GRF), salinity and eddy kinetic energy consistently appeared as dominant predictors of sailfish catch rates. Chlorophyll-a concentration also played a strong role in RF, while GAM highlighted nonlinear responses to both physical (e.g., mixed layer depth, SSH) and biological (chlorophyll-a) drivers.

Table 1. Cross-validated performance of statistical and machine learning models for predicting log-transformed sailfish CPUE.

Model	R^2 (%)	RMSE	Key Predictors
Linear regression (GLM)	0.16	24.05	None (poor fit)
GAM	**19.0**	0.847	Salinity, MLD, SSH, Chlorophyll-a
Gradient Boosted Trees (GRF)	10.1	~0.81	Salinity, EKE, MLD
Random Forest (RF)	15.3	~0.81	Salinity, EKE, Chlorophyll-a

Sailfish catch rates were heterogeneously distributed across the northern and central Indian Ocean fishing grounds exploited by the Sri Lankan longline fleet. High CPUE areas were consistently observed along in the Arabian Sea and EEZ of Sri Lanka and Bay of Bengal, indicating the importance of the Bay of Bengal as a seasonal foraging ground. Catch rates were relatively lower in the equatorial Indian ocean and southern latitudes (0–20°S). These patterns align with oceanographic drivers such as mesoscale eddies, sea surface height anomalies, and chlorophyll concentrations, which were also identified as key predictors in RF and GRF models.

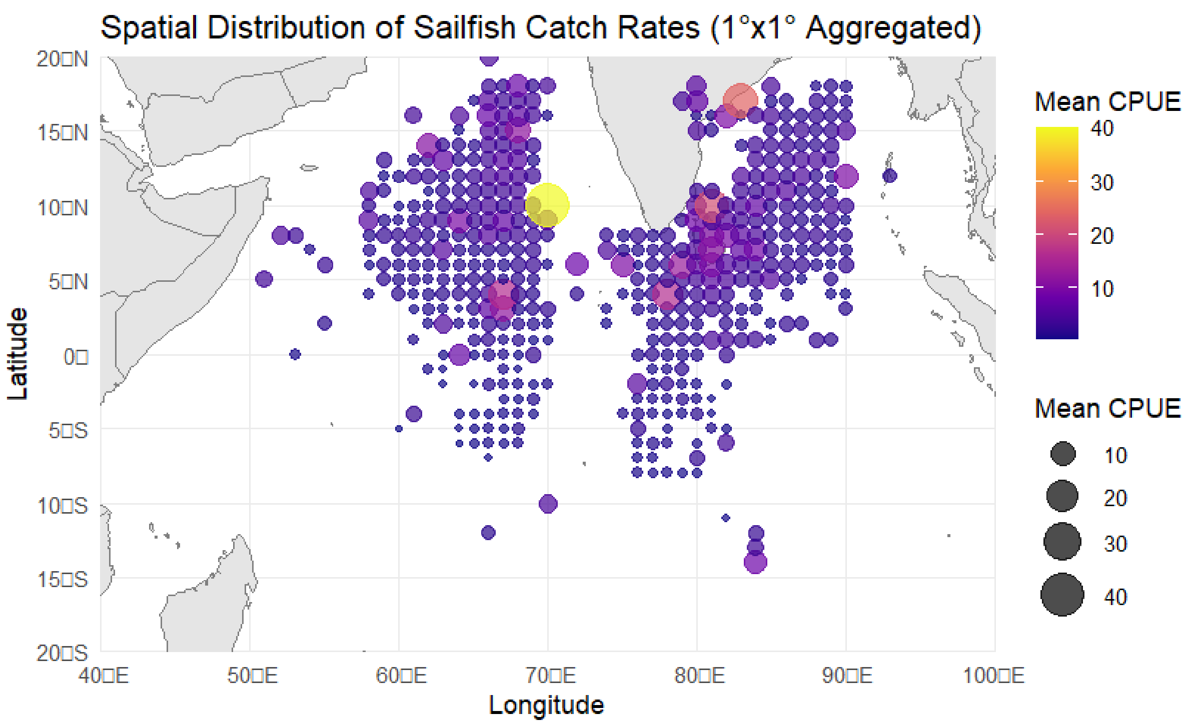

Figure 2: Spatial distribution of Sailfish Catch Rates (1°x1° Aggregated)

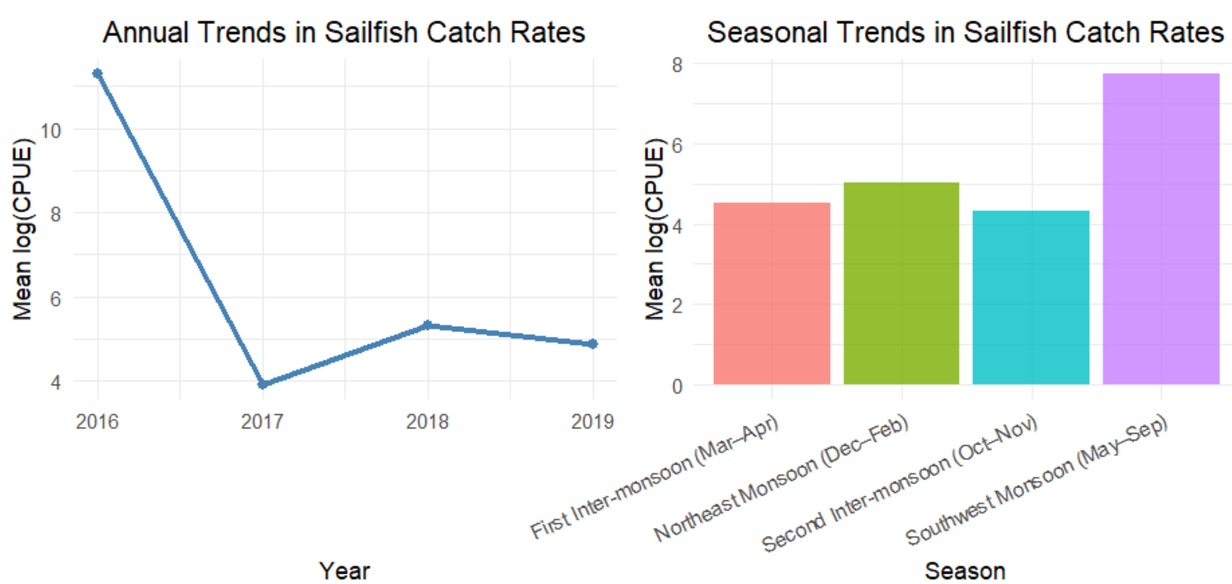

Figure 3: Seasonal and Annual Trends in Sailfish Catch Rates

Mean sailfish CPUE was highest in 2016 and dropped substantially in 2017, remaining at a lower but consistent level through 2019 (Figure 3 a). As shown in Figure 3 b, sailfish were caught at a consistently higher rate during the Southwest Monsoon (May–Sep) compared to the rest of the year, while catch rates during the other three seasons remained relatively similar and low.

Discussion

This study explains the complex spatio-temporal and environmental drivers of sailfish catch rates for the Sri Lankan tuna longline fishery in the Indian Ocean. The consistently low explanatory power of the simple linear model (R^2 = 0.16%) strongly suggests that the relationships between sailfish distribution and oceanographic variables are predominantly non-linear and likely involve complex interactions. This finding underscores the limitation of traditional linear approaches and validates the use of more sophisticated modeling techniques like GAMs and machine learning for this type of ecological data.

The superior performance of the GAM (R^2 = 19%), which incorporated both spatio-temporal and environmental smoothers, indicates that a substantial portion of sailfish catchability is governed by baseline geographic and seasonal patterns. The machine learning models (RF and GBM), while slightly less explanatory, were crucial for identifying the most influential environmental variables free from linear assumptions. The consensus across these models highlights sea surface salinity (SSS) and eddy kinetic energy (EKE) as dominant predictors of sailfish CPUE.

The importance of EKE is ecologically significant, as it is a measure of mesoscale oceanographic activity, such as eddies and frontal zones. These features are known to aggregate nutrients and prey, attracting apex predators like sailfish (Kai & Marsac., 2010). The higher catch rates observed in the Arabian Sea and the Bay of Bengal (Figure 2) coincide with regions of high EKE, suggesting sailfish may forage preferentially in these dynamic environments.

The strong influence of SSS was a notable finding. Salinity can act as a tracer for water masses and oceanic fronts. Sharp salinity gradients often define boundaries between different water bodies, which can lead to prey concentration. Furthermore, the Bay of Bengal, a identified hotspot, is characterized by strong freshwater inputs from major rivers, creating pronounced salinity gradients. Sailfish may be using these fronts as cues for productive foraging grounds.

The secondary role of chlorophyll-a (a proxy for primary productivity) in the RF model suggests that sailfish distribution is also linked to bottom-up trophic processes. Higher productivity supports larger forage fish populations, which in turn support sailfish. The peak in catch rates during the Southwest Monsoon (May-Sep) aligns with the period of maximum oceanic productivity in the Arabian Sea, driven by the strong monsoon winds that cause upwelling.

The significant interannual decline in CPUE from 2016 to 2017, followed by stable but lower rates, is a concerning trend that warrants further investigation. While this study focuses on environmental correlates, this pattern could be indicative of fishery-related impacts, such as localized depletion or increased fishing pressure on the stock. This underscores the urgent need for effective management strategies to ensure the sustainability of this bycatch species.

In conclusion, our models confirm that sailfish distribution in the Indian Ocean is not random but is influenced by a combination of spatio-temporal patterns and specific oceanographic features, particularly those related to water mass characteristics (salinity) and ocean dynamics (EKE and productivity). This information is vital for developing spatial management tools. For instance, dynamic ocean management that considers real-time EKE or salinity fronts could be explored to mitigate sailfish bycatch without unduly compromising tuna catches, contributing to the ecosystem-based management of Sri Lanka's tuna longline fishery.

References

Copernicus Marine Service. (2025). Global Ocean Physics Analysis and Forecast (CMEMS product GLOBAL_ANALYSISFORECAST_PHY_001_024). E.U. Copernicus Marine Service Information. https://doi.org/10.48670/moi-00016

FAO-FIGIS, 2001. A world overview of species of interest to fisheries. Chapter: Istiophorus platypterus. Retrieved on 05 May 2005, from www.fao.org/figis/servlet/species?fid=3301. 4p. FIGIS Species Fact Sheets. Species Identification and Data Programme-SIDP, FAO-FIGIS

Hoolihan J (2003) Sailfish movement in the Arabian Gulf: a summary of tagging efforts. Mar Freshw Res 54:509–513

Indian Ocean Tuna Commission. (2025). Status summary for species of tuna and tuna-like species under the IOTC mandate, as well as other species impacted by IOTC fisheries. Retrieved August 29, 2025, from https://iotc.org/node/3379

Kai, E. T., & Marsac, F. (2010). Influence of mesoscale eddies on spatial structuring of top predators’ communities in the Mozambique Channel. *Progress in Oceanography*, *86*(1-2), 214-223.

Maritorena S., O. Hembise Fanton d Andon, A. Mangin, and D.A. Siegel. 2010. Merged Satellite Ocean Color Data Products Using a Bio-Optical Model: Characteristics, Benefits and Issues. Remote Sensing of Environment, 114, 8: 1791-1804.

Nakamura I (1985) FAO species catalogue. Billfish of the world. An annotated and illustrated catalogue of marlins, sailfishes, spearfishes, and swordfishes known to date. FAO Fish Synop 125(5) 65 p

Ndegwa S, Herrera M (2011) Kenyan Sports Fishing Sailfish Catches. IOTC–2011–WPB09–09