Habitat suitability models of Indian mackerel (*Rastrelliger kanagurta*) in Indonesia Fishery Management Area 713 using remote sensing data and machine learning technique

[1,3]Muhammad Agus, [2]Pramaditya Wicaksono, [2]Nurul Khakhim, [3]Mohamad Natsir, [3]Suparman Sasmita, [4]Aris Budiarto, [3]Ari Kuncoro, [3]Baihaqi, [3]Zainal Wassahua, [3]Helman N. Yusuf, [3]Sofiyan M. Permana

[1] Master of Science Program in Remote Sensing, Faculty of Geography, Universitas Gadjah Mada, 55281, Yogyakarta, Indonesia; [2] Department of Geographic Information Science, Faculty of Geography, Universitas Gadjah Mada, 55281 Yogyakarta, Indonesia; [3] National Research and Innovation Agency Republic of Indonesia (BRIN), 16951, Cibinong, Indonesia; [4] Ministry of Marine Affairs and Fisheries Republic of Indonesia (MMAF), 10110, Jakarta, Indonesia. Corresponding author: M. Agus, muha305@brin.go.id

Abstract. Indian mackerel (*Rastrelliger kanagurta*) is an important pelagic species within Indonesia fishery management area 713. Increasing fishing pressure is considered to be the primary cause of the species' population decline. To support the sustainable utilization of Indian mackerel, this study integrates monthly presence data from January 2018 to December 2022 with remotely sensed environmental variables to predict the species' habitat distribution using random forest algorithm. The environmental variables utilized include chlorophyll-a (Chl), sea surface height (SSH), sea surface salinity (SSS), and sea surface temperature (SST). The results indicate that the random forest algorithm performed satisfactorily, as evidenced by the overall accuracy of 0.944±0.026, true skill statistic (TSS) of 0.889±0.051, and area under the curve (AUC) of the receiver operating characteristic of 0.982±0.019. Chl, SSH, and SSS emerge as the most influential predictors affecting the spatial distribution of the species. The highest potential habitats (HSI > 0.8) were observed along the coast of South Kalimantan to Kangean Island, with smaller patches identified in the waters of West Sulawesi and South Sulawesi. This study accentuates the effectiveness of remote sensing data and machine learning techniques in determining the habitat of pelagic fish. Moreover, the findings of the study can aid decision-making related to sustainable fishery management in the region.

Key Words: ocean color, random forest, Scombridae, small pelagic fish, species distribution modeling.

Introduction. As an archipelagic nation with extensive marine territory, Indonesia holds bountiful marine fishery resources. The nation is the second largest marine fishery producer in the world, with an average annual production of approximately 6.57 million tons from 2017 to 2020 (FAO 2022). Among marine fishery products, small pelagic fish contribute 34% to Indonesia's total marine fisheries production (Jaya et al 2022). One of the most productive areas for small pelagic fish within Indonesian waters is the Indonesia Fishery Management Area (IFMA) 713. The key species commonly captured in this region is the Indian mackerel (*Rastrelliger kanagurta*) (Panggabean et al 2023).

R. kanagurta is a small pelagic fish from the Scombridae family, widely distributed in the western Pacific to the Indian Ocean (Arrafi et al 2016). This species inhabits coastal tropical waters, forages in schools, and migrates seasonally at depths of 20-90 m (White et al 2013; Yusop et al 2021). It has an important economic value and is a source of income for coastal communities, particularly in Indonesia. However, some studies have reported that the *R. kanagurta* population has been declining due to increased fishing effort, mainly in the Java Sea and the Makassar Strait (Saputra & Taufani 2021; Asni et al 2024). Therefore, identifying the key habitat of *R. kanagurta* is considerably important in the context of formulating sustainable fishery management for the species.

Habitat suitability of marine fish can be identified using correlative modeling, known as species distribution modeling (SDM) (Robinson et al 2017; Franklin 2023). SDM is an empirical method that correlates the presence or abundance data with environmental variables using statistical or machine learning algorithms to predict species spatial distribution (Elith & Leathwick 2009; Miller 2010; Sillero et al 2021). SDM can be categorized into two modeling methods: presence-only and presence-absence methods. The presence-only method relies only on the presence data of a species, while the presence-absence method needs both presence and absence data (Mateo et al 2010). Some studies have shown that the presence-absence method outperforms the presence-only method in terms of accuracy and reliability of the model (Franklin 2010; Barbet-Massin et al 2012). Therefore, it has been increasingly used to map the habitat distribution of pelagic fish (Zhao et al 2021; Yang et al 2024; Yati et al 2024).

Supervised machine learning algorithms can be used to develop habitat suitability models using the presence-absence approach (Zhang & Li 2017); one of the most widely used is random forest (RF), which is a machine learning algorithm that is built on an ensemble of several decision trees to produce more accurate predictions (Breiman 2001). This algorithm uses the bootstrap aggregating (bagging) method, which randomly samples data from an input dataset/training data and then constructs decision trees using the classification and regression tree (CART) method. RF was able to overcome prediction instability and overfitting that generally occur in single decision tree-based machine learning (Luan et al 2020). RF has been widely used in various fields involving data analysis to solve classification and regression problems (Rodriguez-Galiano et al 2012; Wicaksono et al 2019; Luan et al 2020). Previous studies have used RF in modeling and mapping *R. kanagurta* habitats within the South China Sea (Tan & Mustapha 2023). However, the robustness of habitat models remains challenging because it is influenced by several factors, such as the unique physical properties of the marine environment, biological characteristics of species, selection of modeling methods, and inclusion of ecologically relevant predictors (Dambach & Rödder 2011; Melo-Merino et al 2020).

Remote sensing technology has been extensively employed in monitoring the marine and coastal ecosystems (Amani et al 2022b). It offers several key advantages compared to in situ measurement, such as synoptic coverage of ocean features, rapid data acquisition, and the capacity for continuous observation (Santos 2000; Klemas 2013). Marine environmental parameters such as chlorophyll-a (Chl) and sea surface temperature (SST) were considered to be strongly associated with the abundance and spatial distribution of *R. kanagurta* (Nurdin et al 2017; Kamaruzzaman et al 2021). Other factors, including sea surface height (SSH) and sea surface salinity (SSS), have been reported to affect the distribution of the species (Yusop et al 2021; Akter et al 2024). However, the influence of these environmental parameters on the spatiotemporal distribution of *R. kanagurta* is not fully understood, particularly in the IFMA 713.

Therefore, this study aimed to predict the habitat distribution of *R. kanagurta* by developing habitat suitability models using the RF algorithm and to investigate the relationship between its spatial distribution and marine environmental parameters in IFMA 713. The results of this study will benefit fisheries management practice for stakeholders in the region. In addition, the study offers a cost-effective approach for monitoring small pelagic fish distribution in Indonesia.

Material and Method

Description of the study sites. The study was conducted in the IFMA 713. This area is located between 1°1’10” - 8°29’47” S and 114°25’41” - 122°40’56” E, encompassing the Makassar Strait, the Bali Sea, the Flores Sea, and the Gulf of Bone (Figure 1). IFMA 713 is passed by the Indonesian Throughflow (ITF), which transports water masses from the Pacific Ocean to the Indian Ocean (Gordon et al 2019). The region is heavily influenced by the Asian-Australian monsoon (Zainuddin et al 2023). In the southeast monsoon

(June-August), the waters' primary productivity in the region is increased due to the confluence of water masses from the Pacific Ocean, the Java Sea, and the Flores Sea (Atmadipoera & Widyastuti 2014). Meanwhile, in the northwest monsoon (December-February) with high rainfall, some areas experience an increasing nutrient concentration carried by runoff from rivers in Kalimantan and Sulawesi. Therefore, it is one of the most productive fisheries areas, especially for small pelagic fish such as mackerels (*Rastrelliger* spp.), scads (*Decapterus* spp.), selar (*Selar* spp.), and sardines (*Sardinella* spp.) (Apriansyah et al 2024).

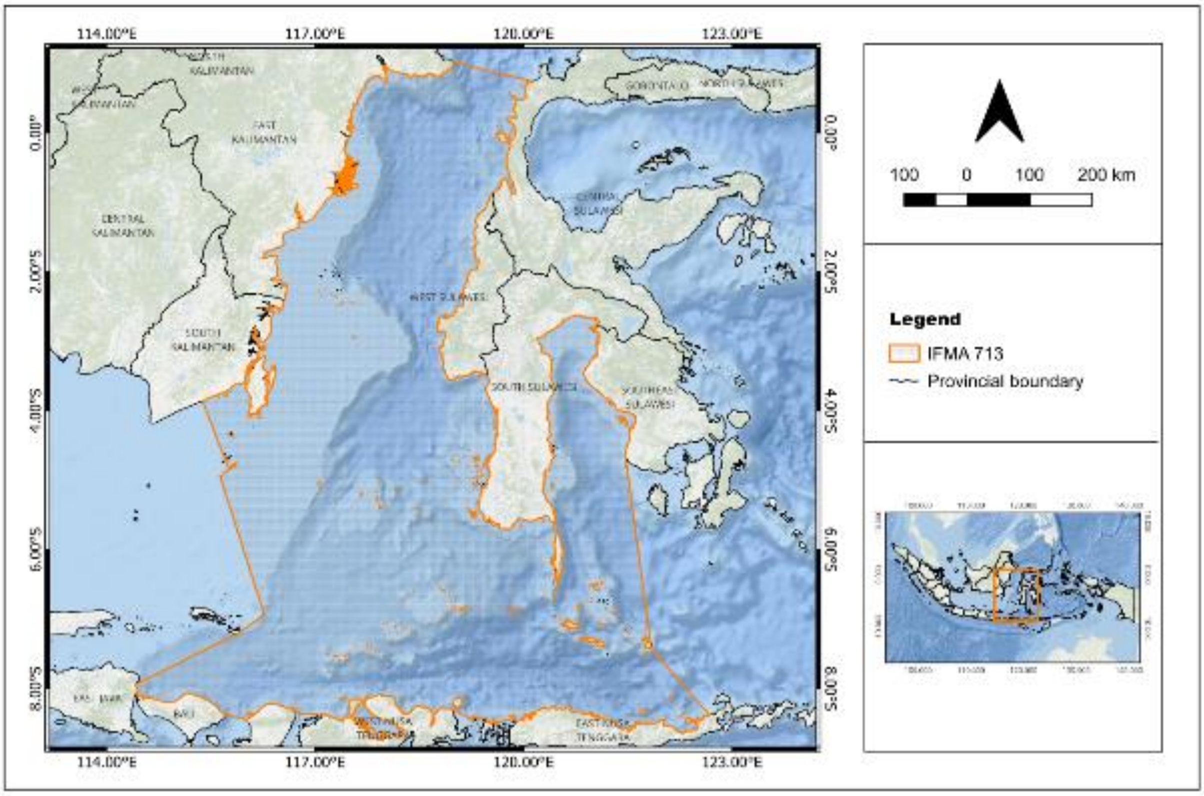

Figure 1. Map of the study area.

Fishery data. Daily occurrences of *R. kanagurta* from January 2018 to December 2022 in IFMA 713 were extracted from fishing logbook dataset retrieved from the Ministry of Maritime Affairs and Fisheries of Indonesia. The data contains the geographic coordinates of the fishing location, fishing time, catch weight, and the type of fishing gear. For modeling purposes, we used positive catch data from mixed fishing gear, such as purse seine and gillnet. The occurrence data were then compiled into a monthly dataset and resampled following the spatial extent of the environmental variables.

Environmental data from remote sensing. The present study utilized remotely sensed environmental data from 2018 to 2022, i.e., Chl, SSH, SSS, and SST, which coincided with the *R. kanagurta* occurrence record. Chl and SST were retrieved from the Aqua MODIS level 3 imagery (https://oceancolor.gsfc.nasa.gov/). Meanwhile, the sea surface height (SSH) and sea surface salinity (SSS) were obtained from the Copernicus Marine Environmental Monitoring Service (CMEMS) (https://marine.copernicus.eu/). These data have different spatial resolutions (Table 1). Hence, Chl and SST were downscaled following the spatial resolution of SSH and SSS using QGIS 3.40.4. Subsequently, each of the environmental predictors was averaged into monthly multi-year data using the median composite technique.

Table 1
List of environmental predictors obtained from remote sensing

Environmental variables	*Unit*	*Spatial resolution*	*Temporal resolution*	*Source*
Chlorophyll-a	mg m^{-3}	4 km	Monthly	Aqua MODIS
Sea surface salinity	psu	0.083°	Monthly	CMEMS
Sea surface height	m	0.083°	Monthly	CMEMS
Sea surface temperature	°C	4 km	Monthly	Aqua MODIS

Environmental profiling and pseudo-absence data generation*.* Dataset obtained from fishing logbooks generally only recorded the presence of a species without information regarding its absence. Under such conditions, pseudo-absence (background point) can be generated in the study area to represent the species absence data (Barbet-Massin et al 2012; Iturbide et al 2015). Before sampling the pseudo-absence data, we clustered the study area based on the values of the environmental parameters using the k-means algorithm. Following the clustering analysis, pseudo-absence data were then randomly sampled in the clustered area where *R. kanagurta* has been recorded (Phillips et al 2009). The number of pseudo-absence samples taken is equal to the number of presence data, as recommended by Barbet-Massin et al (2012). Pseudo-absence data were generated using the "flexsdm" package (Velazco et al 2022).

Multicollinearity test*.* Prior to modeling, it is necessary to assess multicollinearity among the predictor variables. Multicollinearity is a condition in which predictor variables are highly correlated in multiple regression modeling (Allen 1997). It can reduce the predictive capacity and interpretability of a model (Dormann et al 2013). Multicollinearity can be assessed using a pairs plot and the variance inflation factor (VIF) (Shrestha 2020). VIF identifies how much variance increases in regression coefficients as a result of multicollinearity among predictor variables (Yati et al 2024). The VIF was calculated using the following formula:

$$\mathrm{VIF} = \frac{1}{1 - R^2} \tag{1}$$

where R^2 is a coefficient determination of the targeted variable on the remaining predictor variables. Correlation coefficient | r | > 0.8 and VIF ≥ 10 are indications of multicollinearity (Shrestha 2020).
1
Model fitting and evaluation. The development of habitat suitability models for *R. kanagurta* involves the construction of a spatial dataset consisting of environmental data as predictor variables and species occurrence data (presence/pseudo-absence) as the target variables. This data set was then split into two parts: 70% training data and 30% test data. The training data were used as input to construct models using the RF algorithm. To maintain the generalization ability of the models and reduce the risk of overfitting, the training process was combined with 10-fold cross-validation method. Although RF is internally validated, the cross-validation method facilitates the automatic selection of the optimal hyperparameters of the models. The remaining 30% test data were used as external validation to evaluate the models' performance using a confusion matrix (Table 2).

Table 2
Confusion matrix used to evaluate the predictive accuracy of the Indian mackerel habitat suitability models

		Observed		
		Presence	*Absence*	*Sum*
Predicted	Presence	TP	FP	TP + FP
	Absence	FA	TA	FA + TA
	Sum	TP + FA	FP + TA	N = TP + FP + FA+ TA

Note: TP = true presence, FP = false presence, TA = true absence, FA = false absence, and N = total number of observations.

Based on the confusion matrix, we calculated overall accuracy (Eq. 2), sensitivity/true positive rate (Eq. 3), specificity (Eq. 4), false positive rate (Eq. 5)55, true skill statistics (Eq. 6), and area under the curve of the receiver operating characteristic (Eq. 7). These metrics assessed the performance of the models from different perspectives (Franklin 2010; Guisan et al 2017). The overall accuracy (OA) measures the capacity of the models to classify accurately the presence and absence classes. Sensitivity is the percentage accuracy of the model in predicting the presence class. On the other hand, specificity is the percentage accuracy of the model in predicting the absence class. The false positive rate (FPR) is the percentage of model errors in predicting the absence class. True skill statistic (TSS) measures model performance by considering sensitivity and specificity. The TSS value is between -1 and 1, where 1 indicates perfect prediction, while a TSS value of 0 can be interpreted as no agreement between prediction and observation (Allouche et al 2006). The area under the curve of the receiver operating characteristic (AUC) is an independent threshold metric used to assess the model's discrimination capacity. AUC value is obtained by calculating the area under the receiver operating characteristic curve – a plot between sensitivity and FPR at various threshold values (Guisan et al 2017). AUC value is in the range of 0 to 1, where 0.5 is equivalent to random guess and 1 is categorized as excellent performance. The following are the formulas for those metrics:

$$OA = \frac{TP + TA}{N} \quad (2)$$

$$\text{Sensitivity (TPR)} = \frac{TP}{TP + FA} \quad (3)$$

$$\text{Specificity} = \frac{TA}{TA + FP} \quad (4)$$

$$FPR = 1 - \text{Specificity} \quad (5)$$

$$TSS = \text{Sensitivity} + \text{Specificity} - 1 \quad (6)$$

$$AUC = \int_0^1 TPR(FPR^{-1}(x))\,dx \quad (7)$$

All analysis were performed using the R programming language version 4.4.1 (R Core Team 2024). The "terra" package (Hijmans 2024) was used to construct the spatial dataset. The "randomForest" (Liaw & Wiener 2002) and "caret" (Kuhn & Max 2008) were used for modeling. We used "tidyverse" and "ggplot2" packages (Wickham 2016; Wickham et al 2019) for data analysis and visualization.

Results. Figure 2 illustrates the temporal variability of the environmental parameters in the study area. The highest average Chl concentration of 0.53 mg m^{-3} was observed in January, and the lowest of 0.34 mg m^{-3} was recorded in October. The average SSH reached its maximum value at 0.7 m in January, while its minimum at 0.56 m was recorded in August. The average SSS peaked at 33.78 psu in September and dropped to the lowest value of 31.72 psu in April. The highest SST average was recorded at 30.63°C in November, while the lowest SST of 28.78°C occurred in August. The analysis revealed skewed distributions and potential outliers in the environmental parameters, suggesting a non-parametric approach for modeling the habitat suitability of *R. kanagurta* in IFMA 713.

The pairs plot used to determine the correlation between predictor variables is shown in Figure 3. As can be seen, there was a negative correlation between Chl and SSH with $|r| = 0.51$, as well as the correlation between Chl and SSS with $|r| = 0.59$. SSS and SST also had a negative correlation with $|r| = 0.7$. Conversely, a positive correlation was found between SSH and SST with $|r| = 0.58$. However, the analysis showed no correlation exceeding 0.8. Similarly, the VIF test (Table 3) revealed no value greater than 10. As the tests indicated the absence of multicollinearity, all predictor variables were used to develop habitat suitability models for *R. kanagurta*.

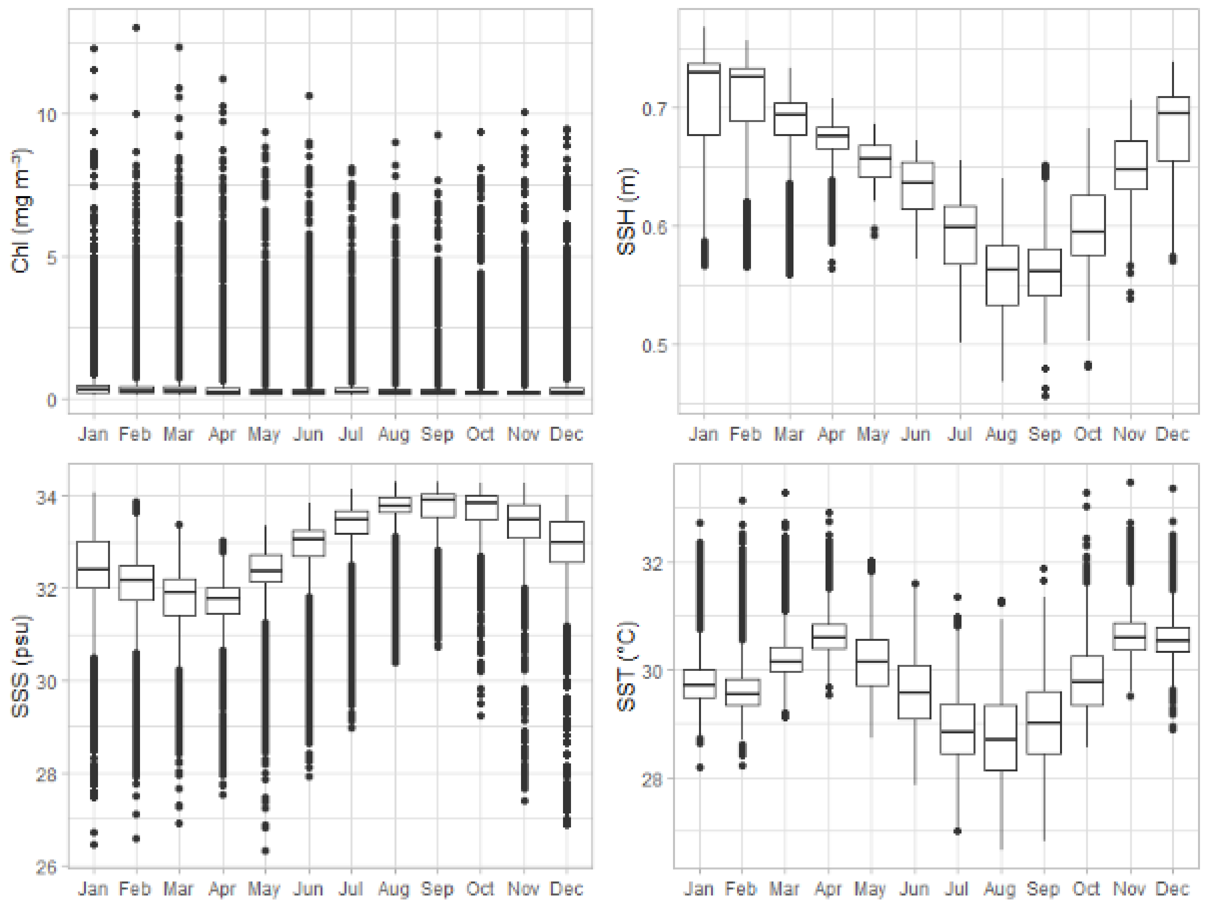

Figure 2. Monthly variability of the environmental parameters in IFMA 713 during 2018- 2022.

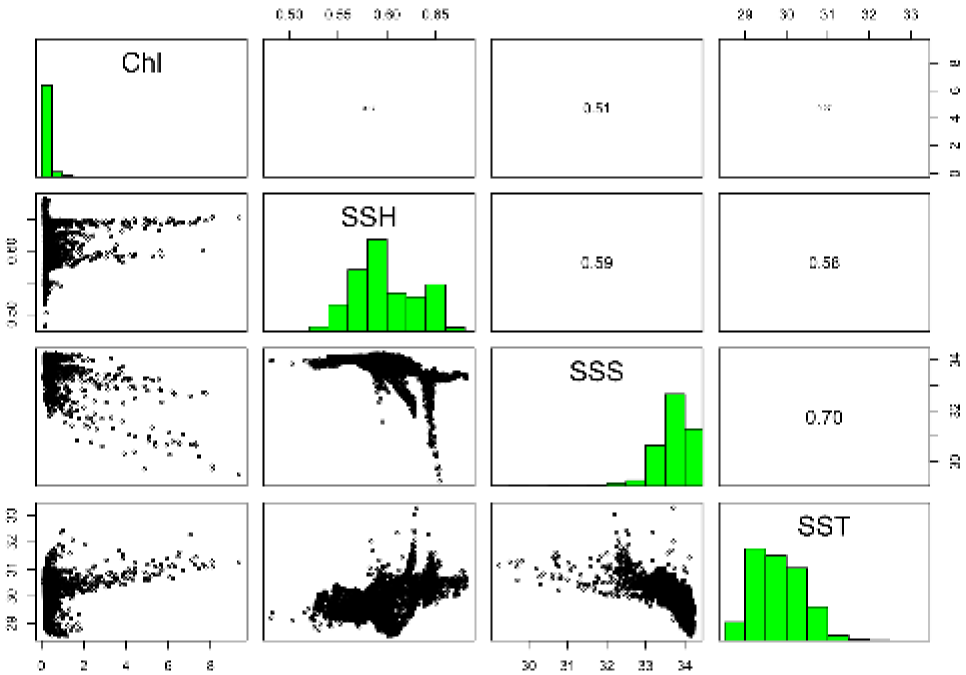

Figure 3. Pairs plot shows the correlation among predictor variables (Chl = chlorophyll-a, SSH = sea surface height, SSS = sea surface salinity, SST = sea surface temperature).

Table 3

Variance inflation factor of predictor variables across different months

Month	*Variance inflation factor (VIF)*			
	Chl	*SSH*	*SSS*	*SST*
January	2.269	2.401	3.752	1.230
February	2.047	2.226	2.897	1.474
March	1.610	1.187	1.643	1.232
April	1.660	1.053	1.657	1.153
May	1.751	1.162	1.755	1.263
June	1.486	1.190	1.764	1.183
July	1.260	1.362	1.486	1.251
August	1.148	1.400	1.384	1.369
September	1.219	1.542	2.381	2.194
October	1.447	1.741	2.976	2.188
November	1.695	1.470	2.154	1.105
December	2.146	1.670	3.264	1.242

***Model performance and influence of environmental predictors**.* The performance of habitat suitability models from January to December is summarized in Table 4. All models obtained OA > 0.8, TSS > 0.7, and AUC > 0.9. These results suggest that the models performed well and can be further used to analyze the *R. kanagurta* habitat in the study area.

Table 4

Monthly model performance across different metrics

Model	*OA*	*Sensitivity*	*Specificity*	*FPR*	*TSS*	*AUC*
January	0.931	0.980	0.882	0.118	0.862	0.991
February	0.933	0.983	0.883	0.117	0.866	0.967
March	0.931	0.985	0.877	0.123	0.862	0.987
April	0.962	0.981	0.942	0.058	0.923	0.980
May	0.939	0.938	0.939	0.061	0.877	0.991
June	0.950	0.900	1.000	0.000	0.900	0.997
July	0.947	0.936	0.957	0.894	0.894	0.997
August	0.958	0.958	0.958	0.042	0.917	0.995
September	0.987	1.000	0.974	0.026	0.974	0.999
October	0.979	1.000	0.958	0.042	0.958	0.988
November	0.924	0.923	0.925	0.075	0.848	0.958
December	0.892	0.946	0.838	0.162	0.784	0.939

In RF, the relative importance of predictor variables can be calculated using variable importance measures (Boulesteix et al 2012). The variable importance measures (VIM) for classification can be performed in two forms: permutation importance (mean decrease accuracy) and gini importance (mean decrease gini) (Goldstein et al 2011). While gini importance is computationally efficient, permutation importance is considered more reliable (Altmann et al 2010; Boulesteix et al 2012). Therefore, we use permutation importance to estimate the relative contribution of predictor variables in the constructed habitat suitability models for *R. kanagurta*.

The importance score of predictor variables in the models exhibited monthly variation (Figure 4). Chl, SSH, and SSS emerge as the most influential predictors in determining the spatial distribution of *R. kanagurta*. These variables alternately obtained the highest importance from January to December. While SST did not exhibit a significant contribution, its exclusion as a predictor variable led to deterioration of the model's performance.

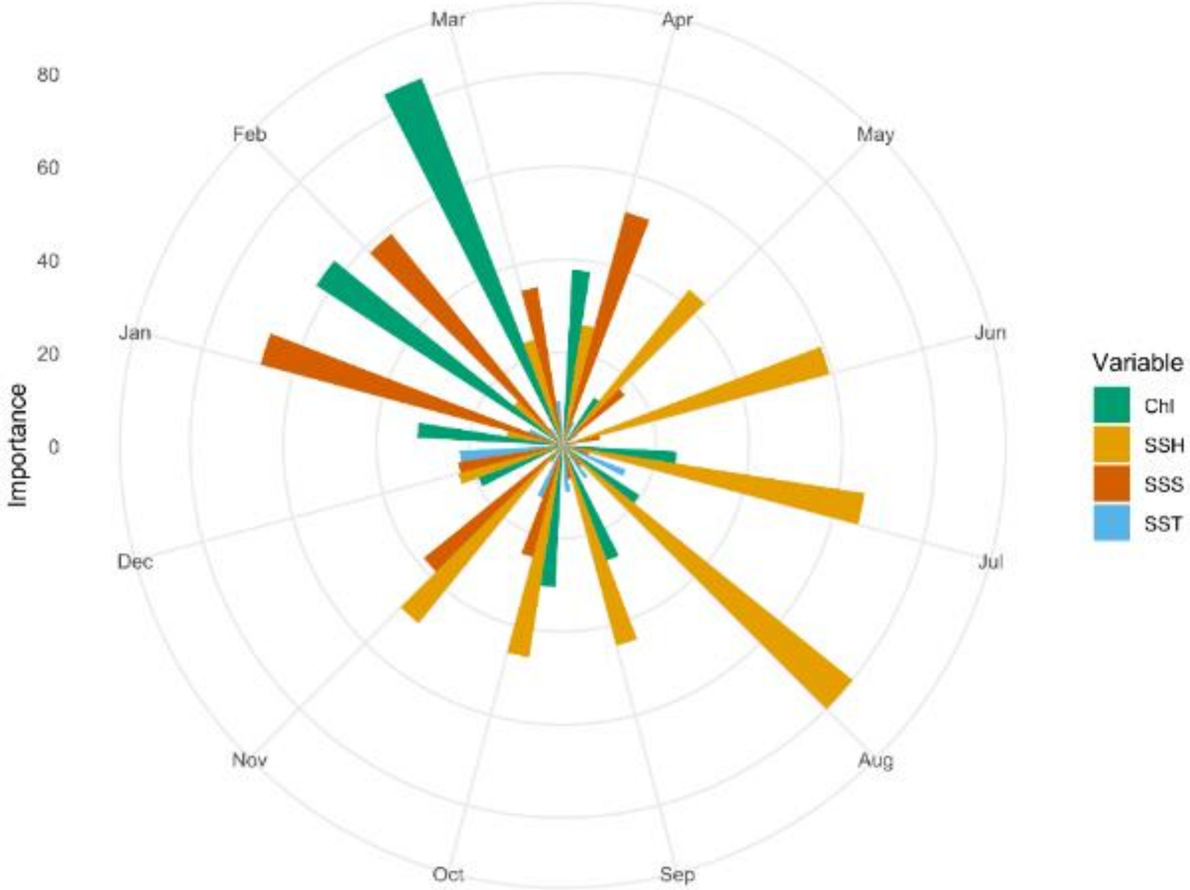

Figure 4. A polar plot illustrates the relative contribution of each environmental variable on the habitat distribution of *Rastrelliger kanagurta* in IFMA 713 from January to December.

The implementation of RF in the "randomForest" package adopts the partial plot feature (Liaw & Wiener 2002). This feature allows the measurement of the marginal effect of predictor variables on the target variable. The partial plot of the habitat models is shown in Figure 5. This graph shows the non-linear relationship between environmental predictors and the occurrence probability of the *R. kanagurta* and illustrates how environmental gradients affect the likelihood of the species' presence in the study area.

In January, when SSS and Chl were identified as the most influential variables, the occurrence probability of the species was approximately 70 to 90% for SSS values of 31.9 to 32.6 psu and Chl concentration of 0.15 to 0.38 mg m^{-3}. During February and March, in which the model outputs indicated Chl as the variable with the highest importance score, followed by SSS, the likelihood of the *R. kanagurta* present ranged from 70 to 93% under Chl of 0.12 to 0.3 mg m^{-3} and SSS value of 32.12 to 33.82 psu. In May and June, with SSH and SSS as the most influential variables, the probability of the species occurrence was approximately 80 to 90% at SSH levels of 0.62 to 0.69 m and SSS values of 32.64 to 33.77 psu. From July to October, when SSH and Chl were recorded as the primary and the second important predictors, the likelihood of the species present was around 75 to 90% for SSH level 0.54 to 0.67 m and Chl concentration at 0.13 to 0.22 mg m^{-3}. In November, with SSH and SSS emerged as the most important variables, the occurrence probability of the species was approximately 93 to 99% at SSH 0.66 to 0.7 m and SSS value of 33.69 to 34.21 psu. By December, when all predictors made almost the same contribution, the presence probability of *R. kanagurta* was approximately 65 to 77% under SSH level 0.71 to 0.74 m, SSS value of 32.69 to 33.07, Chl concentration of 0.15 to 0.25 mg m^{-3}, and SST value of 29.65 to 30.19°C.

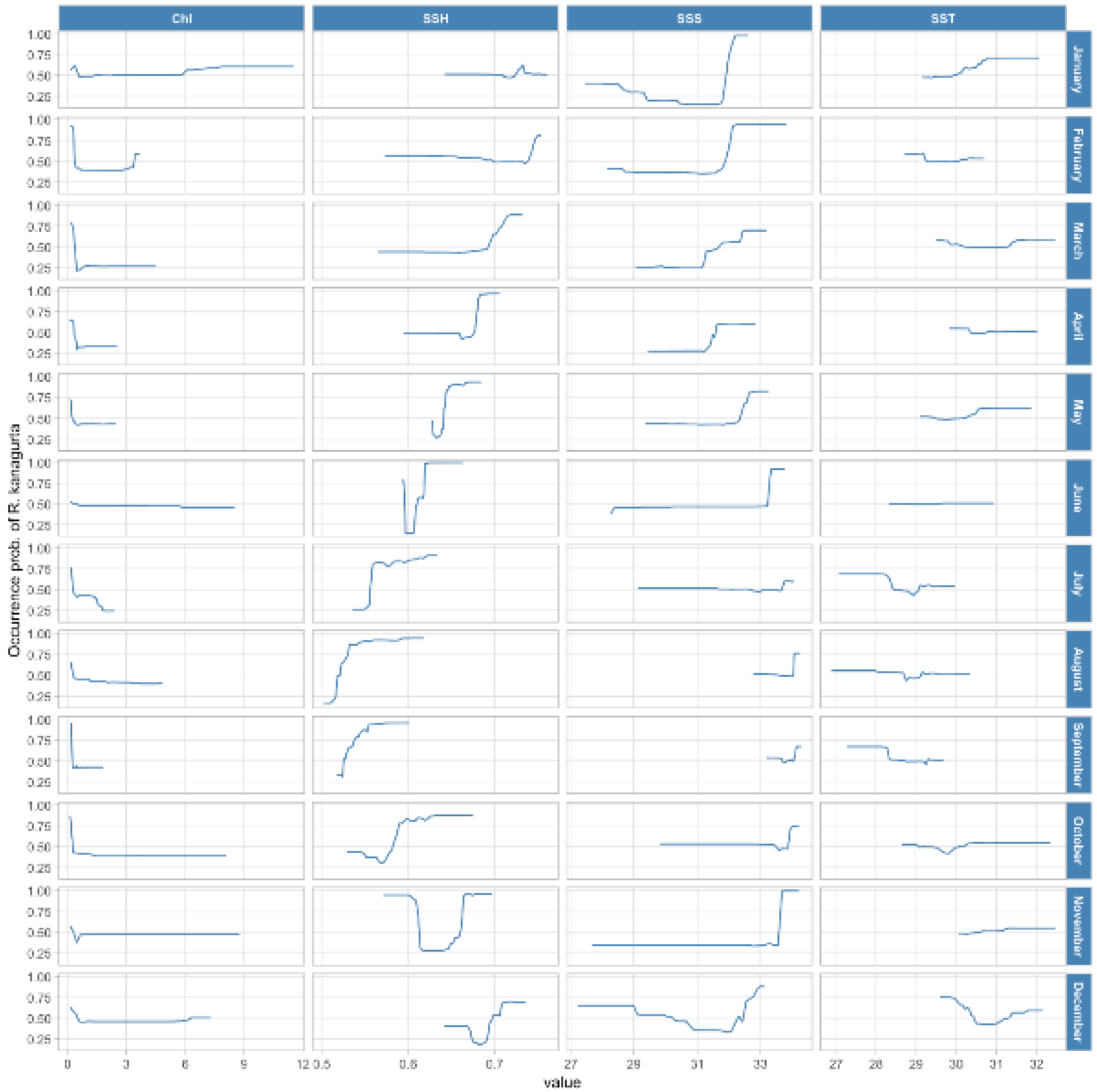

Figure 5. Partial plots show the marginal effect of environmental variables on the occurrence probability of *Rastrelliger kanagurta* in IFMA 713 from January to December.

***Predicted habitat suitability maps**.* Monthly habitat suitability models based on the random forest (RF) algorithm were used to predict the spatiotemporal distribution of *Rastrelliger kanagurta* within the IFMA 713. The predicted maps of the potential habitat from January to December are shown in Figure 6. This map illustrates the habitat suitability index (HSI) for *Rastrelliger kanagurta*, which ranges from 0 to 1, where 0 indicates unsuitable habitat and 1 represents optimally suitable habitat. From January to March, the HSI > 0.5 extended from the coast of East Kalimantan to South Kalimantan, with some patches near the coast of West Sulawesi and South Sulawesi. During April to August, the high potential areas were concentrated mainly around the coast of South Kalimantan and the Pangkajene Islands. From September to November, the HSI > 0.5 shifted slightly towards southwest, covering South Kalimantan and Kangean Island (part of IFMA 712), while scattered patches were also detected in the Gulf of Bone. In December, high-potential areas were dispersed across almost the entire region, yet the highest potential area persisted along the coast of South Kalimantan.

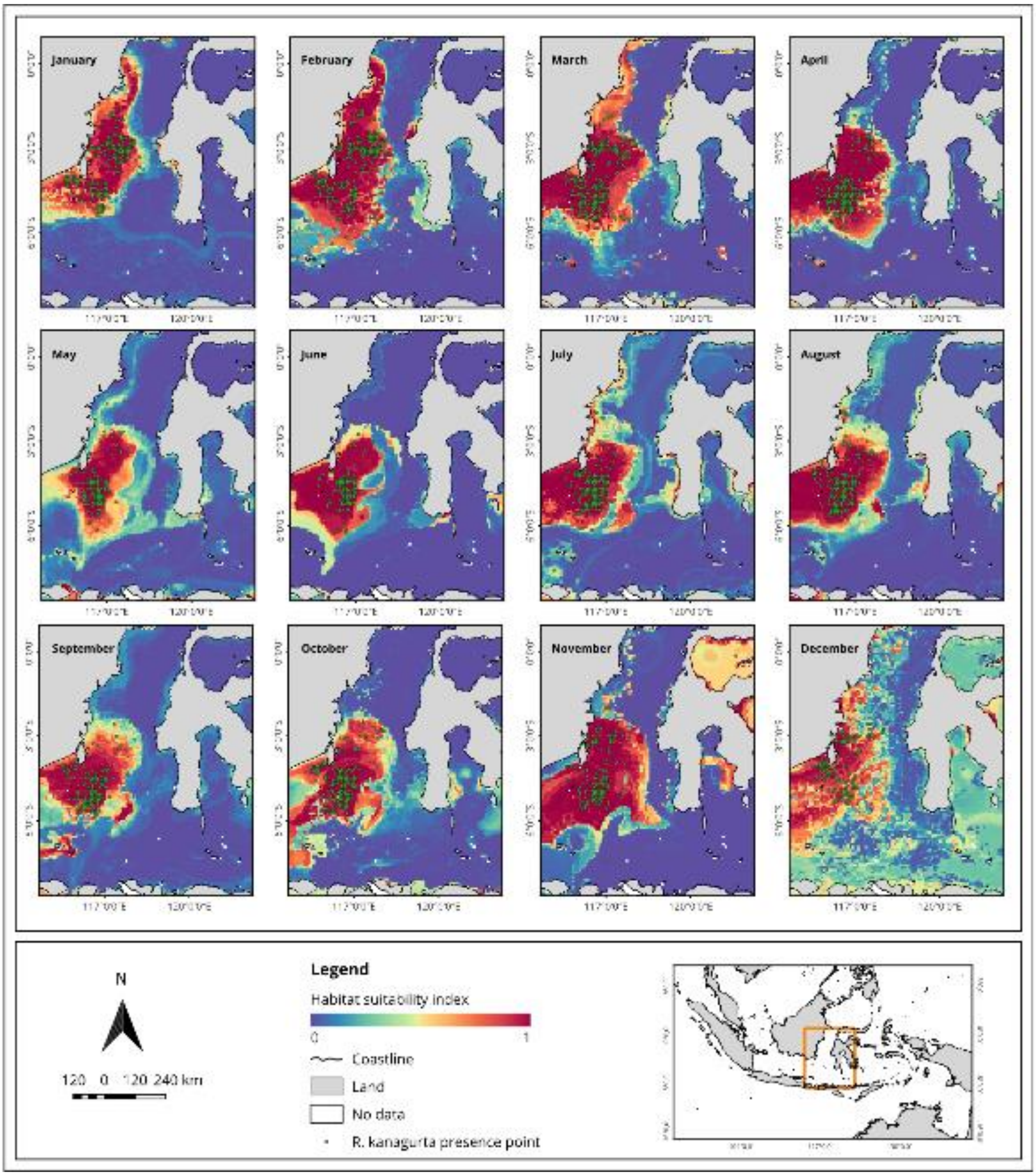

Figure 6. Spatiotemporal habitat distribution of *Rastrelliger kanagurta* in IFMA 713.

Discussion. This study presents the main results of habitat suitability modeling to investigate the spatiotemporal distribution of *R. kanagurta* and its relationship with environmental factors in the IFMA 713. By utilizing remotely sensed data and the RF algorithm, the constructed habitat model achieved an average OA of 0.944±0.026, an average TSS of 0.889±0.051, and an average AUC of 0.982±0.019. While OA is often neglected as a metric in habitat modeling due to its prevalence dependence, it provides a general overview of the thematic accuracy of the models, particularly when used on a balanced dataset (Szabó et al 2024). In contrast, TSS is not dependent on data prevalence and can gain insight into the model's predictive ability at the class level. A TSS > 0.8 indicates that the model has good predictive ability (Yati et al 2024). Meanwhile, AUC evaluates the model's performance at different thresholds between 0 and 1. An AUC > 0.9 suggests that the constructed models have excellent predictive performance (Franklin 2010).

Regarding the influence of environmental variables, our study highlights the importance of Chl, SSH, and SSS in shaping the distribution of *R. kanagurta*. Chl concentration is considered as the key information for identifying fishing grounds, as it is a proxy of phytoplankton biomass – the main diet of pelagic fish (Amani et al 2022a; Zainuddin et al 2023). Previous studies examining the habitat distribution of *R. kanagurta* in Spermonde waters (part of IFMA 713) reported that the habitat preference of this fish was at Chl 0.30 to 0.40 mg m^{-3} (Nurdin et al 2017). A study by Yeny Nadira et al (2019) also revealed that high distributions of this fish were observed at Chl 0.24 to 0.42 mg m^{-3}. It is almost the same as the findings in this study, where the higher probability of occurrence of this species is in the range of Chl 0.12 to 0.38 mg m^{-3}.

SSH reflects the dynamic height of the ocean surface, which is influenced by various factors, including wind stress, oceanic fronts, upwelling, and eddy currents (Belkin 2021). Low SSH is an indication of coastal upwelling, which contributes to increased primary productivity of the waters (Asch & Checkley 2013). According to Zainuddin et al (2023), the southeast monsoon, which generally occurs from July to August, is the main driver of upwelling in the southern part of the IFMA 713. This phenomenon is further substantiated in Figure 2, where the SSH in August is distinctly lower than in other observed months. The interplay between the Indonesian Throughflow (ITF), seasonal monsoon winds, and complex seabed topography facilitates the formation of eddies in IFMA 713 (Nuzula et al 2017). Anticyclonic eddies, which are associated with higher SSH, may concentrate the prey of pelagic fish, resulting in productive fishing grounds (Klemas 2013). Recent study by Akter et al (2024) reported that the main fishing ground of *R. kanagurta* in Maharashtra waters ranged from SSH levels of 0.3 to 0.4 m. In Malaysia's exclusive economic zone, Kamaruzzaman et al (2021) observed that the preferred SSH for the species ranged from 1.1 to 1.7 m. However, the findings of the current study diverge slightly, in which the species preferred a SSH range of 0.5 m to 0.75 m. This discrepancy may be attributed to differences in oceanographic characteristics in each region. Consequently, the species show different responses as a form of adaptation to the local environmental conditions.

SSS has a direct influence on small pelagic fish physiology. It affects the osmoregulation of the fish, which in turn impacts their growth, gonad maturity, and habitat selection (Chen 2022; Zheng et al 2022; Yang et al 2024). In IFMA 713, the fluctuation of SSS follows seasonal patterns (see Figure 2). The SSS value gradually increased during the southeast monsoon (July to August) and reached its peak during the transition monsoon (September to November). The finding of this study indicates that preferred SSS for *R. kanagurta* in IFMA 713 ranges from 31.0 to 33.07 psu. This range is slightly lower than the previous report, which identified a preferred SSS of 33 to 36.5 psu (Akter et al 2024).

Water temperature is widely acknowledged as one of the most important abiotic factors governing the distribution of small pelagic fish. It has a direct impact on the metabolism and embryo survival of the fish (Yang et al 2024). Extreme temperatures have a detrimental effect on fish growth and usually prompt fish to migrate (Sun & Chen 2014; Chen 2022; Lindmark et al 2022). According to Chen (2022), each fish species has a different temperature tolerance range, which varies due to the influence of other environmental factors or the biological status of the fish. Our findings suggest that *R. kanagurta* in IFMA 713 exhibit considerable adaptability to SST fluctuations within the range of 27 to 32.5°C. This adaptive capacity was evidenced by monthly variations of their response curve and relatively lower importance score of SST compared to other explanatory variables.

Based on the constructed habitat models, it was revealed that the habitat preference of *R. kanagurta* from January to December mainly concentrated along the coast of South Kalimantan to Kangean Island. This finding is consistent with the study conducted by Panggabean et al (2023) which identified the area as a major fishing ground in the IFMA 713. In addition, several areas with HSI > 0.5 were also identified on the coast of West Sulawesi, South Sulawesi, and Bone Gulf. Specifically in the coastal region of South Sulawesi, Nurdin et al (2015, 2017) reported the area as a high-potential fishing zone for catching *R. kanagurta*, thereby validating the findings of the current study. Furthermore, the high HSI areas (HSI > 0.8) as depicted in Figure 6, could be used as a reference for implementing dynamic fishery management practice, such as determining catch quotas and imposing temporary restrictions on certain zones to reduce pressure on spawning areas (Maxwell et al 2015; Karp et al 2025).

Conclusions. This study successfully developed habitat suitability models for the *Rastrelliger kanagurta* in the IFMA 713. Random forest proved to be an effective machine learning algorithm in developing habitat suitability models with limited training data, as indicated by TSS and AUC scores. Although the influence of environmental parameters on the distribution of *Rastrelliger kanagurta* varies throughout the year, Chl, SSH, and SSS emerge as the most significant drivers overall. The highest potential habitat of the species, which is predominantly observed around the coast of South Kalimantan and Kangean Island, offers a valuable insight in assisting sustainable fisheries management.

Acknowledgements. The authors would like to thank the Ministry of Marine Affairs and Fisheries of Indonesia for supporting the fishery logbook data. Thanks to NASA Goddard Space Flight Center for providing chlorophyll-a and sea surface temperature data. Gratitude was also conveyed to the Copernicus Marine and Environment Monitoring Service (CMEMS) for the sea surface height and sea surface salinity data.

Conflict of interest. The authors declare that there is no conflict of interest.

References

Akter S., Nakhawa A., Bhendekar S., Lal D. M., Abidi Z. J., Nayak B. B., Ramteke K., 2024 Deciphering the distribution of Indian mackerel, *Rastrelliger kanagurta* (Cuvier, 1817) along the Northwest coasts of India. Thalassas 40(3):1481-1493.

Allen M. P., 1997 The problem of multicollinearity. In: Understanding regression analysis. Springer US, Boston, MA, pp. 176-180.

Allouche O., Tsoar A., Kadmon R., 2006 Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43(6):1223-1232.

Altmann A., Toloşi L., Sander O., Lengauer T., 2010 Permutation importance: a corrected feature importance measure. Bioinformatics 26(10):1340-1347.

Amani M., Mehravar S., Asiyabi R. M., Moghimi A., Ghorbanian A., Ahmadi S. A., Ebrahimy H., Moghaddam S. H. A., Naboureh A., Ranjgar B., Mohseni F., Nazari M. E., Mahdavi S., Mirmazloumi S. M., Ojaghi S., Jin S., 2022a Ocean remote sensing techniques and applications: a review (part II). Water 14(21):3401.

Amani M., Moghimi A., Mirmazloumi S. M., Ranjgar B., Ghorbanian A., Ojaghi S., Ebrahimy H., Naboureh A., Nazari M. E., Mahdavi S., Moghaddam S. H. A., Asiyabi R. M., Ahmadi S. A., Mehravar S., Mohseni F., Jin S., 2022b Ocean remote sensing techniques and applications: a review (part I). Water 14(21):3400.

Apriansyah, Atmadipoera A. S., Natih N. M. N., Nugroho D., Zuraida R., Hartanto M. T., Syahdan M., 2024 Water mass exchange in triangle seas of the Java-Makassar-Flores (JMF): a modeling study. Continental Shelf Research 275:105225.

Arrafi M., Azmi Ambak M., Piah Rumeaida M., Muchlisin Z. A., 2016 Biology of Indian mackerel, *Rastrelliger kanagurta* (Cuvier, 1817) in the western waters of Aceh. Iranian Journal of Fisheries Sciences 15(3):957-972.

Asch R. G., Checkley Jr. D. M., 2013 Dynamic height: a key variable for identifying the spawning habitat of small pelagic fishes. Deep Sea Research Part I: Oceanographic Research Papers 71:79-91.

Asni A., Hasrun, Ihsan, Najamuddin, 2024 Biological aspects and population dynamics of Indian mackerel (*Rastrelliger kanagurta*) in Barru, Makassar Strait, Indonesia. Fisheries and Aquatic Sciences 27(6):392-409.

Atmadipoera A. S., Widyastuti P., 2014 A numerical modeling study on upwelling mechanism in Southern Makassar Strait. Jurnal Ilmu dan Teknologi Kelautan Tropis 6(2):355-371.

Barbet-Massin M., Jiguet F., Albert C. H., Thuiller W., 2012 Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3(2):327-338.
Belkin I. M., 2021 Remote sensing of ocean fronts in marine ecology and fisheries. Remote Sensing 13(5):883.
Boulesteix A. L., Janitza S., Kruppa J., König I. R., 2012 Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2(6):493-507.
Breiman L., 2001 Random forests. Machine Learning 45(1):5-32.
Chen X., 2022 Theory and method of fisheries forecasting. Springer Nature, Singapore, 216 pp.
Dambach J., Rödder D., 2011 Applications and future challenges in marine species distribution modeling. Aquatic Conservation: Marine and Freshwater Ecosystems 21(1):92-100.
Dormann C. F, Elith J., Bacher S., Buchmann C., Carl G., Carré G., Marquéz J. R. G., Gruber B., Lafourcade B., Leitão P. J., et al., 2013 Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36(1):27-46.
Elith J., Leathwick J. R., 2009 Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40:677-697.
FAO, 2022 The state of world fsheries and aquaculture 2022: towards blue transformation. FAO. Rome, 236 pp.
Franklin J., 2010 Mapping species distributions: spatial inference and prediction. Cambridge University Press, 320 pp.
Franklin J., 2023 Species distribution modelling supports the study of past, present and future biogeographies. Journal of Biogeography 50(9):1533-1545.
Goldstein B. A., Polley E. C., Briggs F. B. S., 2011 Random forests for genetic association studies. Statistical Applications in Genetics and Molecular Biology 10(1):32.
Gordon A. L., Napitu A., Huber B. A., Gruenburg L. K., Pujiana K., Agustiadi T., Kuswardani A., Mbay N., Setiawan A., 2019 Makassar Strait throughflow seasonal and interannual variability: an overview. Journal of Geophysical Research: Oceans 124(6):3724-3736.
Guisan A., Thuiller W., Zimmermann N. E., 2017 Habitat suitability and distribution models: with applications in R. Cambridge University Press, 462 pp.
Hijmans R. J., 2024 terra: Spatial Data Analysis. Available at: https://CRAN.R-project.org/package=terra. Accessed: May, 2025.
Iturbide M., Bedia J., Herrera S., del Hierro O., Pinto M., Gutiérrez J. M., 2015 A framework for species distribution modelling with improved pseudo-absence generation. Ecological Modelling 312:166-174.
Jaya I., Satria F., Wudianto, Nugroho D., Sadiyah L., Buchary E. A., White A. T., Franklin E. C., Courtney C. A., Green G., Green S. J., 2022 Are the working principles of fisheries management at work in Indonesia? Marine Policy 140:105047.
Kamaruzzaman Y. N., Mustapha M. A., Ghaffar M. A., 2021 Determination of fishing grounds distribution of the Indian mackerel in Malaysia’s exclusive economic zone off South China Sea using boosted regression trees model. Thalassas 37(1):147-161.
Karp M. A., Cimino M., Craig J. K., Crear D. P., Haak C., Hazen E. L., Kaplan I., Kobayashi D. R., Moustahfid H., Muhling B., Pinsky M. L., Smith L. A., Thorson J. T., Woodworth-Jefcoats P. A., 2025 Applications of species distribution modeling and future needs to support marine resource management. ICES Journal of Marine Science 82(3):fsaf024.
Klemas V., 2013 Fisheries applications of remote sensing: an overview. Fisheries Research 148:124-136.
Kuhn M., 2008 Building predictive models in R using the caret package. Journal of Statistical Software 28(5):1-26.

Liaw A., Wiener M., 2002 Classification and regression by randomForest. R News 2(3): 18-22.
Lindmark M., Audzijonyte A., Blanchard J. L., Gårdmark A., 2022 Temperature impacts on fish physiology and resource abundance lead to faster growth but smaller fish sizes and yields under warming. Global Change Biology 28(21):6239-6253.
Luan J., Zhang C., Xu B., Xue Y., Ren Y., 2020 The predictive performances of random forest models with limited sample size and different species traits. Fisheries Research 227:105534.
Mateo R. G., Croat T. B., Felicísimo Á. M., Muñoz J., 2010 Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. Diversity and Distributions 16(1):84-94.
Maxwell S. M., Hazen E. L., Lewison R. L., Dunn D. C., Bailey H., Bograd S. J., Briscoe D. K., Fossette S., Hobday A. J., Bennett M., et al., 2015 Dynamic ocean management: defining and conceptualizing real-time management of the ocean. Marine Policy 58:42-50.
Melo-Merino S. M., Reyes-Bonilla H., Lira-Noriega A., 2020 Ecological niche models and species distribution models in marine environments: a literature review and spatial analysis of evidence. Ecological Modelling 415:108837.
Miller J., 2010 Species distribution modeling. Geography Compass 4(6):490-509.
Nurdin S., Mustapha M. A., Lihan T., Ghaffar M. A., 2015 Determination of potential fishing grounds of *Rastrelliger kanagurta* using satellite remote sensing and GIS technique. Sains Malaysiana 44(2):225-232.
Nurdin S., Mustapha M. A., Lihan T., Zainuddin M., 2017 Applicability of remote sensing oceanographic data in the detection of potential fishing grounds of *Rastrelliger kanagurta* in the archipelagic waters of Spermonde, Indonesia. Fisheries Research 196:1-12.
Nuzula F., Syamsudin M. L., Yuliadi L. P. S., Purba N. P., Martono, 2017 Eddies spatial variability at Makassar Strait – Flores Sea. IOP Conference Series: Earth and Environmental Science 54(1):012079.
Panggabean D., Wiryawan B., Monintja D. R. O., Jaya I., Atmadipoera A. S., 2024 [Monthly spatial distribution mapping and estimation of small pelagic fish fishing ground in JMF Triangle]. BAWAL Widya Riset Perikanan Tangkap 15(3):147-162. [in Indonesian]
Phillips S. J., Dudík M., Elith J., Graham C. H., Lehmann A., Leathwick J., Ferrier S., 2009 Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19(1):181-197.
R Core Team, 2024 R: a language and environment for statistical computing. R Foundation for Statistical Computing. Available at: https://www.R-project.org/. Accessed: April, 2025.
Robinson N. M., Nelson W. A., Costello M. J., Sutherland J. E., Lundquist C. J., 2017 A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. Frontiers in Marine Science 4:421.
Rodriguez-Galiano V. F., Ghimire B., Rogan J., Chica-Olmo M., Rigol-Sanchez J. P., 2012 An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67:93-104.
Santos A. M. P., 2000 Fisheries oceanography using satellite and airborne remote sensing methods: a review. Fisheries Research 49(1):1-20.
Saputra S. W., Taufani W. T., 2021 Population parameters and exploitation rate of Indian mackerel (*Rastrelliger kanagurta*) on the Java's North Coast. AACL Bioflux 3(3): 1181-1189.
Shrestha N., 2020 Detecting multicollinearity in regression analysis. American Journal of Applied Mathematics and Statistics 8(2):39-42.
Sillero N., Arenas-Castro S., Enriquez-Urzelai U., Vale C. G., Sousa-Guedes D., Martínez-Freiría F., Real R., Barbosa A. M., 2021 Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. Ecological Modelling 456: 109671.

Sun L., Chen H., 2014 Effects of water temperature and fish size on growth and bioenergetics of cobia (*Rachycentron canadum*). Aquaculture 426-427:172-180.

Szabó S, Holb I. J., Abriha-Molnár V. É., Szatmári G., Singh S. K., Abriha D., 2024 Classification assessment tool: a program to measure the uncertainty of classification models in terms of class-level metrics. Applied Soft Computing 155: 111468.

Tan M. K., Mustapha M. A., 2023 Application of the random forest algorithm for mapping potential fishing zones of *Rastrelliger kanagurta* off the east coast of peninsular Malaysia. Regional Studies in Marine Science 60:102881.

Velazco S. J. E, Rose M. B., de Andrade A. F. A., Minoli I., Franklin J., 2022 FLEXSDM: an R package for supporting a comprehensive and flexible species distribution modelling workflow. Methods in Ecology and Evolution 13(8):1661-1669.

White W. T., Last P. R., Dharmadi, Faizah R., Chodrijah U., Prisantoso B. I., Pogonoski J. J., Puckridge M., Blaber S. J. M., 2013 Market fishes of Indonesia. ACIAR Monograph No. 155, Australian Centre for International Agricultural Research, Canberra, 438 pp.

Wicaksono P., Aryaguna P. A., Lazuardi W., 2019 Benthic habitat mapping model and cross validation using machine-learning classification algorithms. Remote Sensing 11(11):1279.

Wickham H., 2016 ggplot2: elegant graphics for data analysis. Springer Cham, 260 pp.

Wickham H., Averick M., Bryan J., Chang W., McGowan L. D., François R., Grolemund G., Hayes A., Henry L., Hester J., et al., 2019 Welcome to the tidyverse. Journal of Open Source Software 4(43):1686.

Yang S., Wang L., Fei Y., Zhang S., Yu L., Zhang H., Wang F., Wu Y., Wu Z., Wang W., Shi J., Jiang K., Fan W., 2024 Spatio-temporal variability of fishing habitat suitability to tuna purse seine fleet in the Western and Central Pacific Ocean. Regional Studies in Marine Science 70:103366.

Yang Z., Chen W., Wang X., Liu B., Dong J., Deng Y., 2024 Suitable habitat of the scad fish (*Decanters* spp.) in Northern South China Sea predicted by MaxEnt model. Regional Studies in Marine Science 69:103315.

Yati E., Sadiyah L., Satria F., Alabia I. D., Sulma S., Prayogo T., Marpaung S., Harsa H., Kushardono D., Lumban-Gaol J., Budiarto A., Efendi D. S., Patmiarsih S., 2024 Spatial distribution models for the four commercial tuna in the sea of maritime continent using multi-sensor remote sensing and maximum entropy. Marine Environmental Research 198:106540.

Yeny Nadira K., Mustapha M. A., Ghaffar M. A., 2019 The Indian mackerel aggregation areas in relation to their oceanographic conditions. Sains Malaysiana 48(11):2575-2581.

Yusop S. M., Mustapha M. A., Lihan T., 2021 Determination of spatio-temporal distribution of *Rastrelliger kanagurta* using modelling techniques for optimal fishing. Journal of Coastal Conservation 25(1):15.

Zainuddin M., Safruddin S., Farhum A., Budimawan B., Hidayat R., Selamat M. B., Wiyono E. S., Ridwan M., Syamsuddin M., Ihsan Y. N., 2023 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 Sensing 15(5):1268.

Zhang J., Li S., 2017 A review of machine learning based species' distribution modelling. In: 2017 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), pp. 199-206.

Zhao H., Feng Y., Dong C., Li Z., 2021 Spatiotemporal distribution of *Decapterus maruadsi* in spring and autumn in response to environmental variation in the northern South China Sea. Regional Studies in Marine Science 45:101811.

Zheng Q., Wang Y., Chen J., Li Y., Zhao F., Liu D., Wei M., Kong D., Chen Z., Zhang Z., Ma Y., 2022 Effects of salinity on the growth, physiological characteristics, and intestinal microbiota of the echiura worm (*Urechis unicinctus*). Frontiers in Marine Science 9:912023.

Received: 18 July 2025. Accepted: 28 September 2025. Published online: 10 December 2025.
Authors:
Muhammad Agus, National Research and Innovation Agency Republic of Indonesia (BRIN), 16951, Cibinong, Indonesia, e-mail: muha305@brin.go.id; agusmohakk@gmail.com
Pramaditya Wicaksono, Department of Geographic Information Science, Faculty of Geography, Universitas Gadjah Mada, 55281 Yogyakarta, Indonesia, e-mail: prama.wicaksono@ugm.ac.id
Nurul Khakhim, Department of Geographic Information Science, Faculty of Geography, Universitas Gadjah Mada, 55281 Yogyakarta, Indonesia, e-mail: nurulk@ugm.ac.id
Mohamad Natsir, National Research and Innovation Agency Republic of Indonesia (BRIN), 16951, Cibinong, Indonesia, e-mail: moh.natsir@brin.go.id
Suparman Sasmita, National Research and Innovation Agency Republic of Indonesia (BRIN), 16951, Cibinong, Indonesia, e-mail: supa038@brin.go.id
Aris Budiarto, Ministry of Marine Affairs and Fisheries Republic of Indonesia (MMAF), 10110, Jakarta, Indonesia, e-mail: arisbudiarto@gmail.com
Ari Kuncoro, National Research and Innovation Agency Republic of Indonesia (BRIN), 16951, Cibinong, Indonesia, e-mail: arik003@brin.go.id
Baihaqi, National Research and Innovation Agency Republic of Indonesia (BRIN), 16951, Cibinong, Indonesia, e-mail: baih001@brin.go.id
Zainal Wassahua, National Research and Innovation Agency Republic of Indonesia (BRIN), 16951, Cibinong, Indonesia, e-mail:zain001@brin.go.id
Helman Nur Yusuf, National Research and Innovation Agency Republic of Indonesia (BRIN), 16951, Cibinong, Indonesia, e-mail: helm006@brin.go.id
Sofiyan Muji Permana, National Research and Innovation Agency Republic of Indonesia (BRIN), 16951, Cibinong, Indonesia, e-mail:sofi009@brin.go.id

How to cite this article:
Agus M., Wicaksono P., Khakhim N., Natsir M., Sasmita S., Budiarto A., Kuncoro A., Baihaqi, Wassahua Z., Yusuf H. N., Permana S. M., 2025 Habitat suitability models of Indian mackerel (*Rastrelliger kanagurta*) in Indonesia Fishery Management Area 713 using remote sensing data and machine learning technique. AACL Bioflux 18(6):2689-2704.