First Sea Turtle Informal Intersessional Working Group Meeting (STIIWG01)
Review of CMM 2018-04 (Sea Turtles)
8 and 10 April 2026 (Pohnpei)| 10 AM-2PM
Virtual Meeting

HABITAT DISTRIBUTION MODELING OF THE LOGGERHEAD TURTLE (*CARETTA CARETTA*) IN THE WESTERN AND CENTRAL PACIFIC OCEAN USING INTEGRATED MULTI SOURCE FISHERIES DATA

STIIWG01-2026-IP-32[1]
9 April 2026

---

[1] Reference 32

HABITAT DISTRIBUTION MODELING OF THE LOGGERHEAD TURTLE (*CARETTA CARETTA*) IN THE WESTERN AND CENTRAL PACIFIC OCEAN USING INTEGRATED MULTI SOURCE FISHERIES DATA

Lianhuan Xie[1], Wei Tang[1], Xuefang Wang[1, 2, 3, 4]

(1. *College of Marine Living Resource Sciences and Management, Shanghai Ocean University, Shanghai 201306, China; 2. National Distant-water Fisheries Engineering Research Center, Shanghai 201306, China; 3. The key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai 201306, China; 4. Scientific Observing and Experimental Station of Oceanic Fishery Resources, Ministry of Agriculture and Rural Affairs, Shanghai 201306, China*)

Email: 17851979910@163.com (Lianhuan Xie); xfwang@shou.edu.cn (Xuefang Wang)

Abstract: As a species listed as Vulnerable on the IUCN Red List, understanding the spatiotemporal habitat patterns of the loggerhead sea turtle (*Caretta caretta*) in pelagic regions is essential for effective population conservation. Using bycatch sea turtle records from the Western and Central Pacific Fisheries Commission Regional Observer Programme, this research applied an optimized Maximum Entropy framework to integrate a suite of environmental predictors, including sea surface temperature (SST), dissolved oxygen concentration (DO), sea surface salinity (SSS), mixed layer depth (MLD), chlorophyll a concentration (CHL), distance to shore (DTS), and eddy kinetic energy (EKE) derived from zonal and meridional current velocities. Predictor screening and parameter tuning were conducted using variance inflation factors (VIF) and the corrected Akaike information criterion (AICc), and three models were developed: a model using longline data only (LL), a model using purse seine data only (PS), and a multi source integrated model combining longline and purse seine data (L+S). Results indicated that SST, DO, and EKE were the dominant environmental drivers of loggerhead distribution. The LL model identified hotspot areas primarily in the northern and western sectors, whereas the PS model emphasized hotspots near the equatorial band. The L+S model achieved the best predictive performance (mean area under the receiver operating characteristic curve > 0.75 with a concentrated distribution; overall accuracy at a moderate to high level) and revealed broader, more connected hotspot regions extending into the southern and eastern sectors. These findings suggest that integrating multi source fisheries data can effectively reduce sampling bias associated with a single gear type and improve model comprehensiveness and robustness. Overall, the research supports the use of multi source fisheries data integration for more reliable habitat suitability assessment, thereby providing a scientific basis for spatial conservation of loggerhead sea turtles and sustainable fisheries management in the region.

Key words: *Caretta caretta*; optimized MaxEnt; multi source fisheries data; habitat hotspots; Western and Central Pacific Ocean

The loggerhead turtles (*Caretta caretta*), also known as red sea turtles, are distributed across subtropical and temperate regions worldwide (Wallace *et al.*, 2010), and their survival and conservation status are jointly affected by a range of anthropogenic and natural threats (Lewison *et al.*, 2014). As an integral part of coastal and offshore ecosystems, loggerhead sea turtles play a crucial role in promoting nutrient cycling and supporting fisheries ecosystem services (Mariani *et al.*, 2023). In recent years, multiple pressures, including habitat degradation, fisheries bycatch, pollution, and climate change, have altered its distribution patterns across spatial and temporal scales, leading to reduced use intensity or spatiotemporal shifts in some historically important areas (Patel *et al.*, 2021). The species is currently listed as Vulnerable on the Red List of the International Union for Conservation of Nature (IUCN, 2023). Spatial identification of habitat suitability not only helps define conservation priority areas, but also provides a scientific basis for fisheries spatial management and long-term monitoring under climate change scenarios. Therefore, conducting systematic habitat suitability modeling is of substantial practical significance.

Current research on loggerhead turtle habitat has been focused primarily on coastal areas, including nesting beaches, bays, and nearshore shallow waters (Robinson *et al.*, 2023). These areas represent critically important habitats and breeding grounds throughout the life cycle of loggerhead turtles. Oceanic habitats likewise play a vital role in their foraging, migration, and connectivity among different nesting sites, while oceanic corridors and offshore foraging patches profoundly influence their migratory routes and energy replenishment (Shimada *et al.*, 2020). Therefore, focusing solely on nearshore areas may underestimate the ecological functions and risk distribution of loggerhead sea turtles on an oceanic scale. Conducting systematic studies of habitat suitability at the oceanic scale will help fill the research gap regarding loggerhead turtles' use of offshore habitats and provide essential spatial information to support cross regional conservation and fisheries management.

When conducting research at the oceanic scale, issues related to data sources and their inherent biases are unavoidable (Hughes *et al.*, 2021).In oceanic regions, species occurrence data are derived primarily from observer records collected by offshore fisheries, particularly those from tuna longline and purse seine operations. These two fishing sectors differ systematically in gear deployment and fishing-ground selection, and their fishing effort also varies in terms of spatial coverage, intensity, and temporal continuity. As a result, the observed occurrence records exhibit marked spatiotemporal heterogeneity and are subject to sampling bias. Therefore, reliance on a single fishery data source may introduce systematic bias and thereby affect inference in habitat suitability models. In current habitat suitability assessments and resource evaluation studies, an increasing number of researchers have adopted a multi-source data integration strategy to improve model interpretability and predictive accuracy (Pacifici *et al.*, 2017). This approach has become relatively well established in tuna stock assessment research. Hoyle et al.(2024)noted that different data sources do not reflect exactly the same aspects of population information, and therefore require standardized modeling and appropriate integration that account for differences in gear type, fishing practice, spatiotemporal distribution, and observation processes in order to improve the reliability of assessment results. By jointly using CPUE indices and species composition data derived from different fishing sectors, such as longline and purse seine fisheries, their approach more comprehensively captured changes in stock abundance and population structure through explicit characterization of differences among gears in spatial coverage, fishing depth, and selectivity. Lin et al.(2023)jointly used longline and purse seine fishery data in combination with oceanographic environmental variables to model the spatiotemporal distribution and abundance dynamics of bigeye tuna in the Pacific within a neural network framework.The research showed that data from different fishery sources can jointly enhance the characterization of bigeye tuna distribution patterns and their environmental responses, while also revealing distinct distributional features reflected by longline and purse seine fisheries. This suggests that incorporating information derived from different observation systems or fishing practices into a unified modeling framework can improve both the representation of underlying ecological processes and the predictive performance of the model. However, in oceanic sea turtle habitat modeling, systematic comparisons of how different fishery data sources and their integration influence model outcomes remain limited. It is therefore necessary to compare results derived from single source fishery data with those obtained from integrated datasets within a unified modeling framework, in order to more accurately assess the spatial patterns of potential suitable habitat for loggerhead turtles and their ecological response characteristics.

The Maximum Entropy Model (MaxEnt), which is based on the principle of maximum entropy and is suitable for presence-only data modeling (Phillips *et al.*, 2006), has been widely used in marine species distribution modeling. MaxEnt requires only species presence locations and environmental background data. It controls model complexity and reduces overfitting through regularization and feature classes, while also providing estimates of relative variable contributions and response curves that facilitate interpretation of driving factors (Kong *et al*., 2011). Zampollo et al.(2022) used MaxEnt to model loggerhead turtle habitats during the wintering and nesting seasons in the central and southern Adriatic Sea and the northeastern Ionian Sea within the Mediterranean region. Their results indicated that the central and southern Adriatic Sea represents an important year round foraging area with a relatively broad ecological niche, whereas suitable habitats in the Ionian Sea were more fragmented and concentrated in nearshore waters, being driven mainly by environmental factors such as water depth, distance from shore, and sea surface temperature. Fujisaki et al.(2020), based on satellite tracking locations and by classifying individual behavioral states into “migration” and “foraging” applied ecological niche models including MaxEnt to predict suitable foraging habitats and potential hotspot areas for different sea turtle species in the Gulf of Mexico. Their research characterized sea turtle distribution patterns and their relationships with surface environmental conditions at the basin scale, thereby providing an interpretable spatial basis for identifying key marine areas and assessing overlap risks with fisheries activities.

The Western and Central Pacific is not only an important foraging ground and migratory corridor for loggerhead turtles, but also a major operational area for offshore tuna fisheries, including longline and purse seine fisheries (Li *et al*., 2025). The spatial coupling between bycatch pressure and habitat use in this region therefore requires urgent clarification. Accordingly, this research applies MaxEnt to systematically model habitat suitability for loggerhead turtles in the region, with the aim of providing a spatial decision making basis for identifying high risk areas of fisheries bycatch and delineating priority conservation areas. At the same time, the research examines the influence of data source bias on model outcomes by comparing habitat suitability patterns derived from longline only data, purse seine only data, and integrated datasets from both fisheries. It further evaluates the extent to which multi source fishery data improve spatial coverage completeness, hotspot identification consistency, and predictive stability, thereby offering methodological guidance for distribution modeling of other oceanic bycatch species.

1 Materials and Methods

1.1 Research Area Overview

The research area encompassed the western and central Pacific Ocean (22.5°S~42.5°N, 132.5°E~147.5°W), spanning tropical, subtropical, and temperate climatic zones and exhibiting a pronounced latitudinal environmental gradient (Fig. 1). In the open ocean, primary productivity is spatially regulated by processes such as fronts, eddies, and topographic features, leading to the formation of seasonal high-biomass patches that in turn support the offshore foraging and migratory behavior of sea turtles, marine mammals, and other large marine species (Martinetto *et al*., 2020).

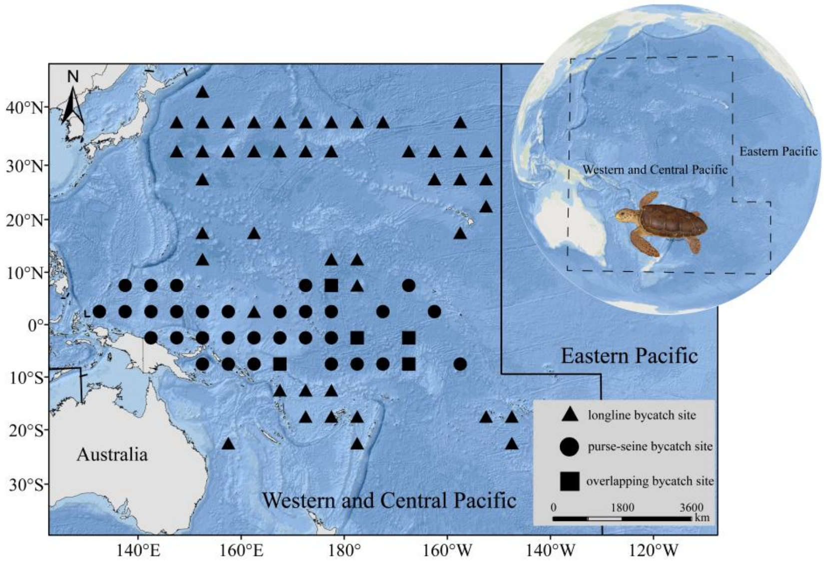

Fig. 1 Distribution of loggerhead turtle bycatch recorded by fishery observers during 2015 to 2019

1.2 Sea Turtle Bycatch and Ocean Environmental Data Sources

1.2.1 Species Occurrence Data Based on Fishery Observer Records

The species occurrence data used in this research were obtained from the Regional Observer Programme (ROP) database of the Western and Central Pacific Fisheries Commission (WCPFC). The original database covers the period from 2013 to 2023 and includes information on the spatial locations (longitude and latitude) of bycaught sea turtles, species identity, number of sets, number of hooks, and survival status. From this database, loggerhead turtle bycatch records from 2015 to

2019 were selected as the species occurrence data for modeling.

1.2.2 Ocean Environmental Data

Studies conducted in other marine regions have shown that the spatiotemporal distribution of loggerhead turtles is primarily influenced by oceanographic environmental factors such as sea surface temperature (SST), sea surface salinity (SSS), dissolved oxygen concentration (DO), mixed layer depth (MLD), chlorophyll-a concentration (CHL), distance to shore (DTS), and eddy kinetic energy (EKE) (Duquesne *et al.*, 2025; Maglietta *et al.*, 2024; Polovina *et al.*, 2001). Therefore, these seven oceanographic variables were selected in the present research for habitat modeling. Ocean environmental data for the research area were downloaded from the Copernicus Marine Environment Monitoring Service (Copernicus Marine Environment Monitoring Service, CMEMS, https://marine.copernicus.eu/) (Table 1). Among these variables, SST, SSS, DO, MLD, CHL, eastward sea water velocity (u), and northward sea water velocity (v) had a monthly temporal resolution and a spatial resolution of 0.25°×0.25°. DTS was calculated using global coastline vector data obtained from the rnaturalearth package in R 4.4.2, with the spatial resolution likewise set to 0.25°×0.25°. The EKE variable used in model construction was derived from eastward and northward sea water velocities, and the specific formula is as follows:

$$\mathrm{EKE}=\frac{1}{2}((u-\bar{u})^2+(v-\bar{v})^2)$$

Here, $\bar{u}$ and $\bar{v}$ represent the global means of the entire u and v arrays. Each was calculated as a single mean value after flattening the corresponding array.

Tab. 1 Abbreviations, means, ranges of values and units of each environmental variable used in this research

Environmental variable	Abbreviation	Mean	Range of values	Unit
sea surface temperature	SST	21.30	0.59-30.24	℃
sea surface salinity	SSS	34.52	31.19-36.58	
dissolved oxygen concentration	DO	232.35	196.36-343.26	$\mathrm{mmol/m^3}$
mixed layer depth	MLD	42.44	8.05-264.15	m
chlorophyll-a concentration	CHL	0.21	0.04-4.17	$\mathrm{mg/m^3}$
eastward sea water velocity	u	0.001	-0.99-1.22	m/s
northward sea water velocity	v	0.008	-1.62-1.38	m/s
eddy kinetic energy	EKE	0.02	0.00-0.46	$\mathrm{m^2/s^2}$
distance to shore	DTS	748.56	1.00-2653.00	km

1.3 Research Methods

1.3.1 Data Preprocessing

All data layers used to construct the MaxEnt model were required to have the same spatial resolution to ensure proper alignment and matching. In this research, the original species occurrence data had a spatial resolution of 5°×5° and an annual temporal resolution. Accordingly, the environmental variables were processed to the same spatiotemporal scale to ensure consistency between species occurrence information and the environmental background fields in both space and time. Although annual averaging and the 5°×5° spatial scale facilitated matching with observer data and improved interannual comparability, they may also have reduced the model's ability to capture responses to environmental variability at monthly and seasonal scales. In ArcMap 10.8, the raster calculator was used to average monthly environmental data across 12 months to obtain annual mean values. The processed environmental layers were then resampled to 5° × 5° using the nearest-neighbor method and exported as ASCII files for subsequent analysis. Because nearest-neighbor resampling directly preserves original pixel values without applying smoothing interpolation, it can better maintain the original value range and physical meaning of environmental variables, thereby avoiding additional numerical bias introduced by interpolation-based reconstruction (Arun *et al.*, 2014). The longitude and latitude data of loggerhead turtle bycatch locations were organized by year and fishing gear and exported as CSV files for the period 2015-2019. Gear categories included longline only data (LL), purse seine only data (PS), and the combined dataset from both gears (L+S). The processed data were subsequently used to build MaxEnt models grouped by year and data source, with the aim of quantitatively evaluating the effects of different data sources (LL, PS and L+S) on model predictive performance.

1.3.2 MaxEnt Construction and Optimization

High collinearity among variables can affect model stability and the precision of coefficient estimation. Therefore, initial variable screening was conducted to avoid this problem (Etaga *et al.*, 2021). A Pearson correlation matrix was used to examine the correlations among SST, SSS, DO, MLD, CHL, EKE, and DTS. A correlation coefficient of $|r| > 0.7$ was taken to indicate a high correlation between two variables (Dormann *et al.*, 2013). When high collinearity was detected, the variance inflation factor (VIF) was used to test for and screen multicollinearity among candidate predictors. A stepwise iterative elimination procedure was applied, in which the variable with the highest VIF exceeding the threshold (VIF > 5) was removed in each round, and the VIF values of the remaining variables were then recalculated. This process was repeated until all retained variables had VIF values below 5, and the remaining variables were subsequently used for model construction (Sillero *et al*., 2021).

In MaxEnt modeling, the selection of the regularization multiplier (RM) and feature classes (FC) is critical to predictive performance (Phillips *et al*., 2008). The RM imposes a smoothing penalty on model parameters: higher RM values generate smoother and simpler response curves, thereby reducing the risk of overfitting, whereas lower RM values allow more complex fits. FC determines the functional forms available to the model and thus influences its ability to capture nonlinear relationships and interaction effects (Warren *et al*., 2011). Therefore, robust model selection should be based on joint parameter tuning using the corrected Akaike information criterion (AICc) and the difference in AICc (ΔAICc) (Muscarella *et al*., 2014). Smaller AICc and ΔAICc values indicate a more optimal combination of FC and RM and better model performance. Accordingly, the FC and RM combination corresponding to the minimum AICc and ΔAICc values was selected for model construction. The primary purpose of AICc in this research was to compare different FC and RM combinations internally within the same year and data source, so as to identify the optimal parameter combination for a given dataset, rather than to directly compare the predictive performance among the PS, LL and L+S models. In this research, the ENMeval package in R 4.4.2 was used to set the RM range from 0.5 to 4.0, with an interval of 0.5. Candidate FC combinations included linear (L), quadratic (Q), hinge (H), product (P), and threshold (T) features. A total of 31 FC and RM combinations were evaluated to jointly tune model parameters and explore nonlinear and interactive structures. In MaxEnt, 80% of the species occurrence data were randomly selected for model training, and the remaining 20% were used for model validation. Because a single random split between training and validation datasets may cause evaluation metrics and prediction results to be affected by sampling variability, repeated random subsampling was adopted to reduce randomness and improve the stability and reproducibility of model results. Specifically, the model was run 500 times, and the mean prediction was taken as the final result.

1.3.3 MaxEnt Evaluation

The area under the receiver operating characteristic (ROC) curve (AUC) was used as an indicator of model predictive accuracy. In general, an AUC value below 0.7 is considered to indicate poor discriminative ability, a value between 0.7 and 0.9 indicates moderate to good performance, and a value greater than 0.9 is regarded as excellent discrimination (Manel *et al.*, 2001). However, these thresholds are empirical, and AUC itself is sensitive to background-point selection, sample prevalence, and the spatial extent of the research area, which may introduce bias, particularly in MaxEnt models. Therefore, in addition to AUC, this research also used accuracy and root mean square error (RMSE) as indicators of model predictive performance. Accuracy was calculated as the proportion of correctly predicted samples among the total number of test samples and was used to reflect the overall classification performance of the model. RMSE was calculated as the square root of the mean squared difference between predicted and observed values and was used to quantify the overall magnitude of prediction error, with lower values indicating closer agreement between predictions and observations.

The annual mean contribution rates of environmental variables derived from MaxEnt were visualized using the tidyverse and ggplot2 packages in R 4.4.2 to produce bubble plots for assessing variable importance, while radar charts were generated in Origin 2022 to provide an intuitive display of the key environmental factors. The distribution predictions generated by MaxEnt were exported in ASCII format and mapped in ArcMap 10.8.

2 Results and Analysis

2.1 Screening Results of Environmental Variables for the Model

According to the Pearson correlation matrix analysis, the absolute correlation coefficient between SST and DO exceeded 0.7. The results after VIF based screening are shown in Table 2. Only the 2015L+S and 2017PS models retained SST, whereas all other models excluded SST and retained DO for final model construction.

Tab. 2 Results of VIF screening for environmental variables

Model name	Environmental variable VIF value						
	CHL	DTS	EKE	MLD	SSS	SST	DO
2015PS	1.46	1.18	1.08	1.93	1.70		3.22
2016PS	1.37	1.32	1.10	2.27	1.55		3.04
2017PS	1.39	1.25	1.08	2.14	1.59	3.29	

2018PS	1.37	1.21	1.13	2.22	1.84		3.53
2019PS	1.63	1.19	1.11	2.23	1.82		3.79
2015LL	1.47	1.20	1.09	1.97	1.71		3.34
2016LL	1.38	1.33	1.11	2.29	1.55		3.09
2017LL	1.35	1.23	1.08	2.13	1.79		3.33
2018LL	1.38	1.23	1.13	2.29	1.83		3.64
2019LL	1.64	1.21	1.11	2.29	1.82		3.92
2015L+S	1.50	1.21	1.09	2.01	1.52	3.25	
2016L+S	1.38	1.32	1.10	2.27	1.56		3.06
2017L+S	1.35	1.21	1.07	2.09	1.80		3.29
2018L+S	1.38	1.21	1.13	2.24	1.85		3.56
2019L+S	1.63	1.19	1.10	2.24	1.82		3.82

2.2 Results of Optimal Model Selection

The optimal models were selected through joint parameter tuning based on AICc and ΔAICc, and the tuning results are presented in Table 3. The optimal combinations of FC and RM varied across years and data sources, indicating that interannual variation and differences in observational coverage led to different requirements for model complexity.

Tab. 3 Joint hyperparameter tuning results

Model name	FC	RM	AICc	ΔAICc
2015PS	H	2	171.26	0.00
2016PS	LH	2.5	130.00	0.00
2017PS	L	1.5	94.63	0.00
2018PS	Q	1	189.01	0.00
2019PS	LHQ	2	180.88	0.00
2015LL	LH	4	144.99	0.00
2016LL	T	0.5	241.78	0.00
2017LL	T	1.5	156.10	0.00
2018LL	H	4	183.33	0.00
2019LL	LQPT	1.5	205.07	0.00
2015L+S	LHP	1.5	357.34	0.00
2016L+S	H	1.5	346.36	0.00
2017L+S	LH	3.5	270.70	0.00
2018L+S	LHP	2.5	415.96	0.00
2019L+S	LHP	3.5	424.28	0.00

2.3 Model Performance Evaluation

According to the comprehensive evaluation of models based on different data sources from 2015 to 2019 (Fig. 2), clear interannual differences were observed among the three model types. The L+S model consistently exhibited relatively high and stable AUC values throughout 2015 to 2019, while its accuracy also remained within a comparatively narrow range, indicating that models constructed from integrated data showed better performance consistency across years. In contrast, the LL model displayed greater interannual variability, with notably higher dispersion in AUC and accuracy in 2016 and 2018, suggesting that its performance was more susceptible to fluctuations in sample structure or observational coverage in individual years. Although the PS model generally showed high AUC values across years, its advantage was more reflected in relatively high performance within specific individual years, whereas the L+S model was characterized by both relatively strong cross year performance and greater overall stability.

2.4 Evaluation of Environmental Factor Importance

The contributions of environmental variables to the three model types revealed clear differences in driver structure among models calibrated with different data sources (Fig. 3). The PS models exhibited a pronounced pattern of dominance by a few variables, with SST and DO accounting for the vast majority of contributions in most years; in the 2017 PS model, for example, the contribution of SST exceeded 90%. This indicates that models calibrated with purse seine data were highly sensitive to thermal conditions and related water mass characteristics. In contrast, the variable contribution profiles of the LL models were much more dispersed, with CHL, DO, and EKE generally showing moderate contributions and with single variable dominance occurring only rarely. The L+S models displayed a comparatively balanced contribution structure, although DO and SST still remained the major contributing variables. As shown by the radar plots (Fig. 4), different fishery data sources not only affected the spatial patterns of model predictions, but also altered the identification of key driving factors. The PS models were more likely to interpret loggerhead turtle distribution as a concentrated response to SST and DO gradients, whereas the LL models revealed a more diffuse pattern of environmental control. After integrating the two data types, the L+S models reduced the dominance of any single variable and provided a more comprehensive environmental interpretation of habitat suitability for loggerhead turtles. In contrast, MLD, SSS, and DTS consistently showed relatively low contributions across all models.

2.5 Comparison of Prediction Results from the Three Models

MaxEnt predicted the spatial patterns of occurrence probability for loggerhead turtles in the western and central Pacific during 2015 to 2019, and substantial differences were observed among models calibrated with different data sources (Fig. 5). Models based on longline data generally exhibited a broad pattern of moderate habitat suitability across most years. In contrast, models based on purse seine data were concentrated in the equatorial zone, showing pronounced distribution hotspots extending along the equator, with relatively stable interannual patterns. Models integrating longline and purse-seine data displayed a mixed pattern, typically amplifying occurrence probabilities in the southern and eastern parts of the research area and forming an obliquely oriented high probability band. Compared with the LL models, the L+S models produced a more clearly differentiated habitat suitability pattern, while also exhibiting greater interannual variability than the PS models.

The multi year mean predictions of the three model types also showed marked differences (Fig. 6). Actual species occurrence locations were overlaid to evaluate the accuracy of model predictions. In the LL model, habitat suitability across the western and central Pacific remained relatively homogeneous, and the model failed to clearly distinguish between areas of high and low suitability (Fig. 6-a). In the PS model, the hotspot region shifted toward the equatorial band, extending from the western Pacific into the central Pacific, and effectively captured loggerhead turtle habitat near the equator (Fig. 6-b). The L+S model exhibited a mixed pattern, with the high-probability band expanding obliquely toward the southern and eastern regions. Its hotspot areas were broader and more connected, and the occurrence points showed the widest spatial coverage, extending along a gradient from northern to southern areas. This pattern highlights that integrating data sources improved the comprehensiveness of the model and its fit to actual species occurrence records (Fig. 6-c).

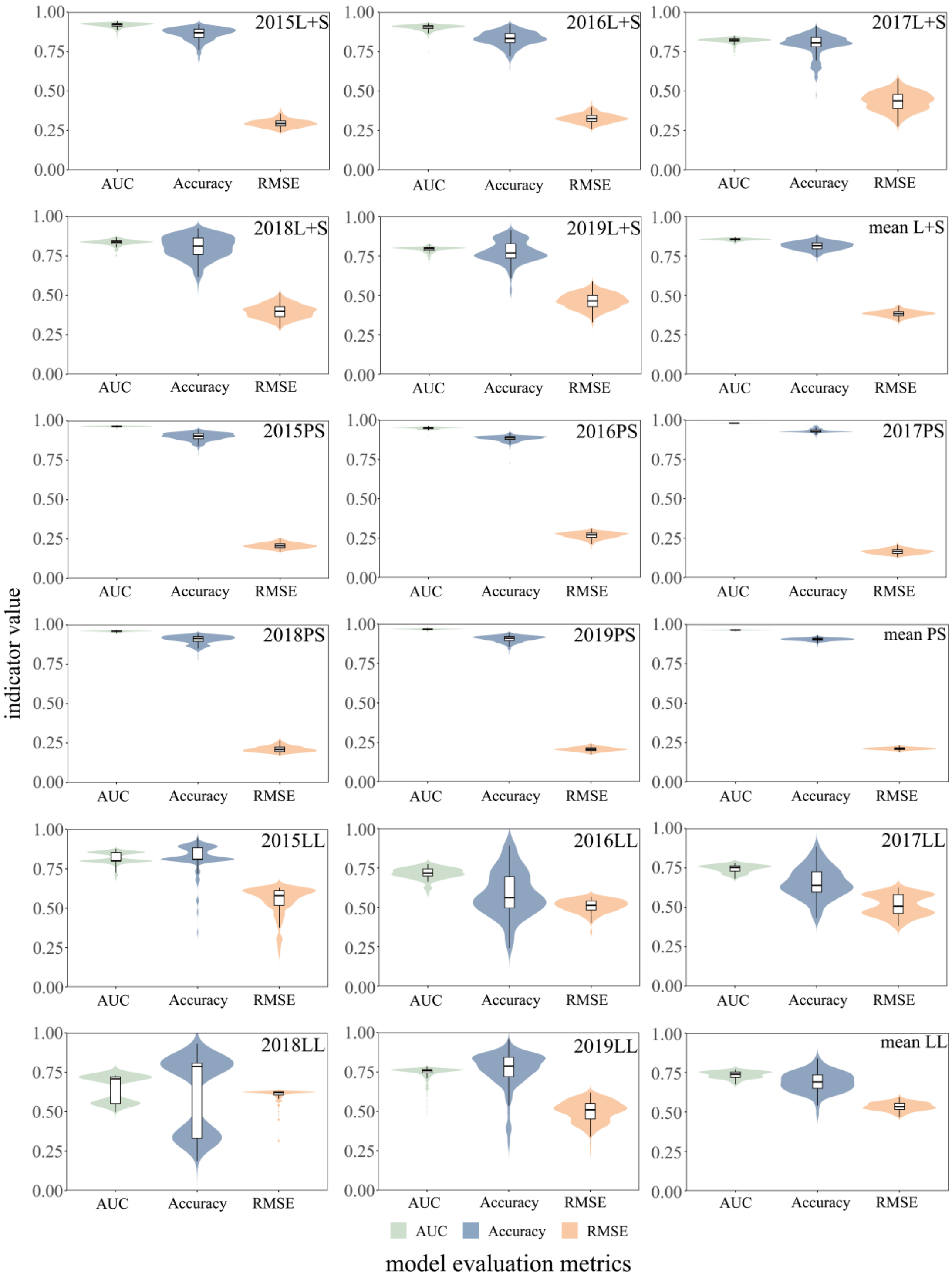

Fig. 2 Comprehensive evaluation of model performance for *Caretta caretta* models based on different data sources across 2015-2019

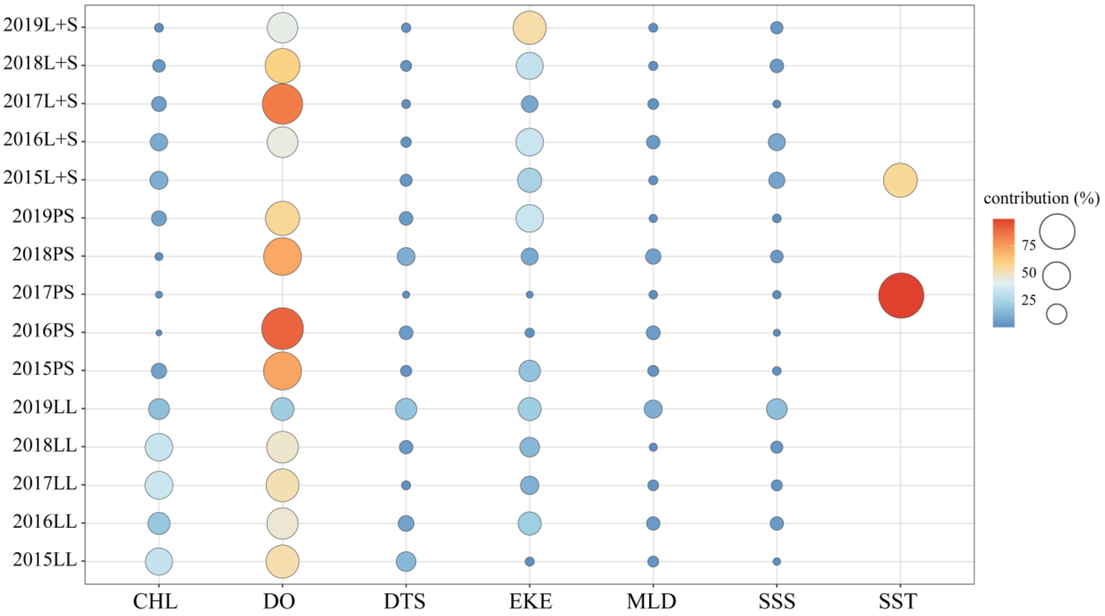

Fig. 3　Bubble plot of environmental variable contributions to model predictions

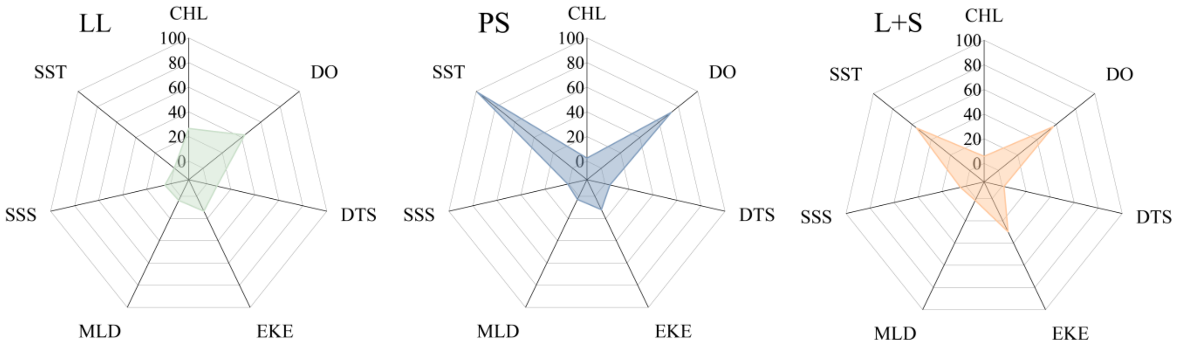

Fig. 4　Radar chart of mean environmental predictor contributions (%) to model predictions

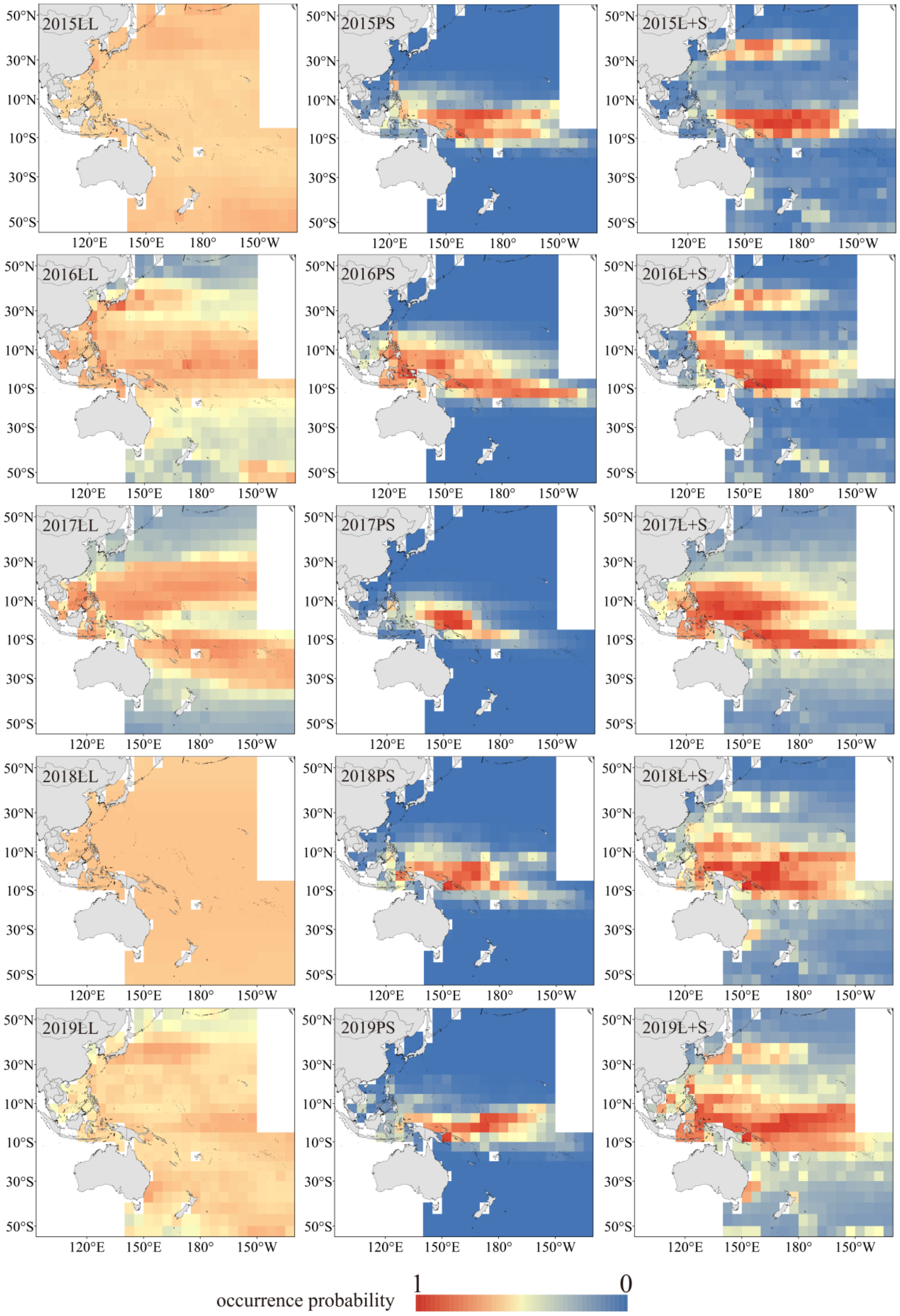

Fig. 5 Model Predicted distribution hotspots of the *Caretta caretta* in the Western and Central Pacific

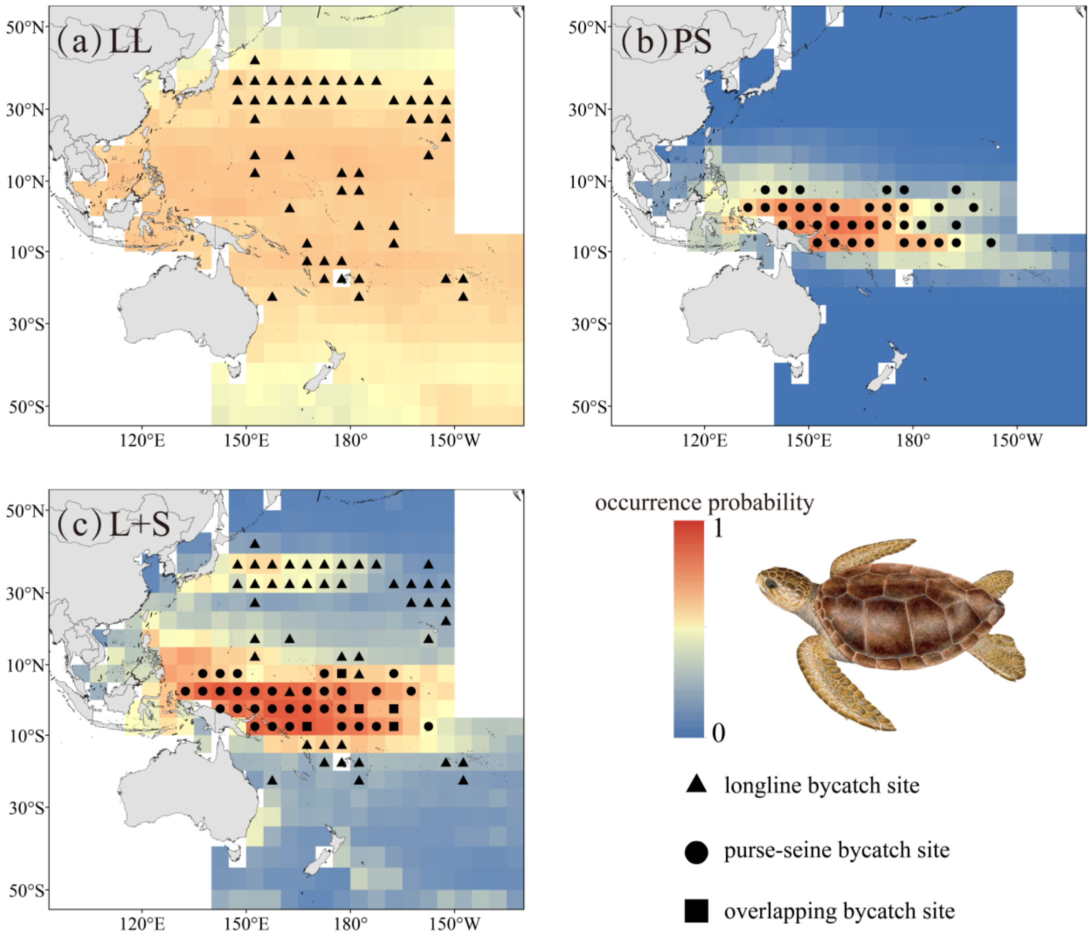

Fig. 6 Overlay comparison between the mean predicted probabilities from the three models and *Caretta caretta* occurrence records in the Western and Central Pacific Ocean

3 Discussion

3.1 Consistency between Environmental Drivers and Ecological Mechanisms

In this research, the optimized MaxEnt models identified SST and DO as the principal drivers of habitat suitability for loggerhead turtles. Their contributions were particularly pronounced in the PS and L+S models, where they were substantially higher than those of the other variables. In the 2017PS model, for example, the contribution of SST exceeded 90%, reflecting a strong association between the spatial distribution of loggerhead turtles and temperature gradients. This strong relationship may be closely linked to their foraging behavior and energetic efficiency. Previous studies have shown that loggerhead turtles tend to utilize waters with temperatures of 20-25℃ in order to maintain a relatively stable metabolic state and optimize energy balance (Abecassis *et al.*, 2013). The prominent role of DO may be related to the diving behavior of loggerhead turtles, as areas with higher DO may help prolong underwater residence time and reduce hypoxic stress (Hochscheid *et al*., 2003). It should be noted that DO and SST were highly correlated in this research ($|r|>0.7$), indicating strong statistical collinearity between the two variables. Under such conditions, MaxEnt is unable to clearly disentangle the independent ecological effects of DO and SST. In the surface ocean, SST and DO are often jointly regulated by water mass properties, air-sea exchange, and biogeochemical processes, and therefore commonly exhibit coordinated spatial variation (Matear *et al*., 2003). Consequently, the relatively high contribution of DO in the model does not necessarily imply that it exerts a fully independent ecological effect on the distribution of loggerhead turtles; rather, it more likely reflects that DO, together with SST, serves as an indicator of the thermal environment and water mass structure in the surface ocean.

In the L+S model of the present research, EKE also ranked prominently among the major contributing variables. This suggests that, at a broad spatial scale, especially in integrated predictions based on samples from different fishing gears, dynamic variables play an important role in shaping the spatial distribution of loggerhead turtles and identifying foraging hotspots. Mesoscale eddies and their associated dynamic processes in the ocean, such as frontal convergence, eddy induced upwelling and downwelling, and water mass transport, can substantially restructure the spatial patterns of nutrients and biomass. At different stages of the eddy life cycle, nutrients may be injected into the euphotic zone through upwelling, thereby enhancing primary productivity (McGillicuddy *et al*., 1998). At the same time, physical convergence and eddy edge transport can promote high-density aggregations of zooplankton, juvenile fish, and other prey organisms, forming prey hotspots at short spatial and temporal scales; these hotspots may provide sea turtles with energetically profitable foraging grounds (Schmid *et al*., 2020). Multiple studies have documented overlaps between eddies, prey aggregations, and hotspots of top predators, including seabirds, cetaceans, and sea turtles, in different marine regions (Arostegui *et al*., 2022). In addition, Gaube et al.(2017) reported that the spatial distribution of juvenile or migrating loggerhead turtles was significantly associated with mesoscale eddies, with individuals frequently occurring in eddy cores or along eddy margins. This pattern may reflect either foraging advantages derived from the highly productive environments generated by eddies or passive transport facilitated by eddy circulation, indicating that mesoscale dynamic processes represented by EKE may strongly influence the movement strategies and habitat use of loggerhead turtles. The annual mean EKE used in this research is more suitable for characterizing the long-term background intensity of mesoscale dynamic activity in the research area, or the average level of eddy activity in a given region over multiple years. It can therefore represent the broad scale spatial pattern of mesoscale dynamic influence, but it cannot precisely identify the immediate effects of specific eddy events in a particular year, season, or short-term period on turtle aggregation. Annual mean EKE may thus serve as a macroscopic proxy for the ecological effects of eddies, but it is still insufficient to fully capture eddy edge effects and their short term ecological processes. Accordingly, the relatively high contribution of EKE in this research should be interpreted as indicating the importance of the mesoscale dynamic background to the broad scale distribution pattern of loggerhead turtles, rather than as direct evidence that annual mean EKE has fully captured the ecological mechanisms operating at the scale of individual eddy events.

The pronounced differences in driver structure among models calibrated with different data sources do not necessarily indicate that the true ecological responses of loggerhead turtles changed fundamentally among years. Rather, they more likely reflect, at least to some extent, the sampling biases inherent in different fishery data sources. In the PS models, SST and DO consistently dominated the contribution profiles, suggesting that samples collected from purse seine fisheries primarily emphasized the explanatory power of thermal gradients in the equatorial and adjacent waters for loggerhead turtle occurrence probability. In contrast, the more dispersed contribution spectra of the LL models indicate that samples derived from longline fisheries covered a broader or more heterogeneous range of environmental conditions, allowing variables related to prey aggregation and dynamic processes, such as CHL and EKE, to also enter the explanatory framework. After integrating the two data sources, the L+S models still showed relatively high contributions from SST and DO, but the dominance of any single thermal variable was clearly reduced, while the importance of variables such as EKE increased. This suggests that multi source data integration can, to some extent, alleviate the spatial coverage bias associated with samples from a single fishing gear, thereby yielding a more balanced environmental interpretation of loggerhead turtle habitat and one that more closely reflects their integrated ecological response at a broad spatial scale.

3.2 Improvements in Model Predictions through Multi-Source Data Integration

In conventional studies of loggerhead turtle habitat, species occurrence records are often derived from catches or observer data associated with a single fishing gear. Such fishery dependent data collection is typically not based on random sampling design, but is instead influenced by the spatial and temporal distribution of fishing effort, gear selectivity, and preferences for target species, which can introduce ecological niche bias in sample sources and constrain the model's ability to characterize overall habitat patterns (Karp *et al*., 2023; Degenford *et al*., 2021). In the present research, the L+S model, constructed by integrating longline and purse-seine data, substantially expanded the extent of identified hotspots and enhanced their spatial continuity. Overlay validation with occurrence points further showed that the L+S model provided more consistent spatial coverage from the northern offshore waters to the equatorial zone and the southern western and central Pacific than models based on either single data source alone (Fig. 6).

The LL model exhibited a broad pattern of moderate habitat suitability, whereas the PS model was concentrated in the equatorial zone, showing pronounced hotspots extending along the equator, with marked interannual variation but insufficient representation of the northern offshore region. In contrast, the L+S model displayed a mixed pattern: it generally amplified occurrence probabilities in the southern and eastern parts of the research area and, overall, formed a high-probability band slanting from north to south. At the same time, the interannual variability of the L+S model was intermediate between that of the other two models: hotspot areas were relatively extensive in 2015 and 2019, but became more contracted in 2016-2017. As shown by the occurrence-point overlay in Fig. 6, the L+S model provided the broadest coverage across the north-south gradient, indicating that data integration effectively reduced the directional biases associated with single-gear data sources, with the LL model biased toward deeper northern waters and the PS model toward surface equatorial waters. In addition to expanding spatial coverage, multi-source data integration also improved model robustness by enhancing tolerance to interannual fluctuations. The L+S model maintained consistently high AUC values across all years from 2015 to 2019, while its accuracy values were also relatively concentrated, with no obvious anomalous years, indicating better predictive consistency across years. By contrast, models based on a single data source were more susceptible to differences in sample size, fishing area shifts, or observational structure in particular years. The LL model showed greater dispersion in some years, suggesting that its performance was more sensitive to interannual sample variability. Although the PS model generally achieved high AUC values, this pattern was more reflective of high responses under a specific data-source condition, and its apparent stability in interpretation remained constrained by the sampling characteristics of a single gear type. By integrating data from both longline and purse-seine fisheries, the L+S model partially balanced the random fluctuations arising from differences in sample coverage among years and gears, thereby maintaining more consistent predictive performance across years.

Kobayashi et al.(2008) used satellite tagging and remote sensing data to characterize in detail the seasonal movements and depth use of loggerhead turtles in the North Pacific. However, because the research relied on a single tracking source, its sample representativeness was limited, allowing only partial explanation of depth and spatial variation and making it difficult to capture hotspot overlap under fisheries bycatch scenarios. Dunkin et al.(2016) developed a loggerhead turtle nesting suitability model based on remote sensing and multi criteria decision analysis. That research primarily used coastal, terrestrial, and nearshore remote sensing variables to identify and evaluate nesting beaches, and achieved good performance in detecting coastal nesting areas. Nevertheless, because both the variables and samples were concentrated in coastal environments, its conclusions are difficult to extrapolate to offshore foraging corridors or to issues of spatial overlap with pelagic fisheries. In contrast, by integrating longline and purse seine data from the WCPFC Regional Observer Programme and conducting cross year comparisons, the present research enhanced the ability to detect marginal areas and offshore distribution hotspots. The suitability patterns and hotspot areas derived from multi-source data integration are therefore better suited to serve as a scientific basis for identifying high risk bycatch areas, delineating priority zones for dynamic management, optimizing the spatial and temporal regulation of fisheries, and promoting coordinated conservation across marine regions. This is particularly important in regions such as the western and central Pacific, where fishing activity is intense and loggerhead turtles undertake extensive migrations. Compared with reliance on a single fishery data source, the use of integrated datasets can reduce the influence of localized sampling bias on management decisions, thereby improving the precision and effectiveness of habitat protection and bycatch mitigation measures.

3.3 Analysis of Model Limitations and Uncertainties in Habitat Dynamics

Although the L+S model showed advantages in spatial consistency and occurrence point coverage, several sources of uncertainty and limitation may affect the robustness of the conclusions and the strength of their ecological interpretation. First, the choice of spatiotemporal resolution and temporal scale constrained the model's sensitivity to short term dynamic events. In this research, annual mean environmental layers were used and modeling was conducted on a 5° × 5° spatial grid. Such relatively coarse spatiotemporal resolution may obscure mesoscale dynamic processes, such as eddy edge effects, as well as environmental variability at monthly or seasonal scales, even though these processes are key mechanisms through which dynamic factors such as EKE drive prey aggregation and induce short-term turtle aggregation. Eddies can substantially alter surface chlorophyll distributions and trophic structure, and their ecological effects are often especially pronounced at monthly or even shorter temporal scales (Chelton, 2011). In addition, El Niño events occurred during 2015 to 2017 and may have reshaped the spatial patterns of SST and EKE fields over short time periods (Thual *et al.*, 2025; Liu *et al.*, 2024), which may partly explain the spatial fragmentation observed in the 2017 prediction results. Models based on annual mean fields may therefore underestimate the influence of events such as El Niño on interannual distribution patterns. Second, in the variable selection procedure of multiple models, highly collinear predictors were removed using VIF, and in most cases DO was retained while SST was excluded. Although this treatment helps reduce the influence of multicollinearity on the model, it inevitably results in some loss of information. Excluding SST may weaken the model's ability to capture direct temperature driven processes, including the direct effects of thermal gradients on metabolic rate, locomotion costs, and migratory pathways, as well as physiological and behavioral mechanisms through which temperature influences nesting phenology and coastal offshore spatial selection (Kobayashi, 2018). As a result, some distributional responses that are actually driven by temperature may instead be attributed to DO, which could reduce the model's ability to detect latitudinal tropical subtropical thermal gradients and introduce bias in identifying temperature driven marginal distributions or interannual migration patterns.

In this research, only surface environmental variables were used to construct the habitat suitability models, and vertical environmental heterogeneity was not incorporated. Ignoring the distributions of temperature and dissolved oxygen across different depth layers may result in the loss of important information. Features such as thermoclines, temperature gradients beneath the mixed layer, subsurface cold tongues or warm ridges, and mid water oxygen minimum zones may all influence prey aggregation, diving behavior, and vertical migration in loggerhead turtles, thereby affecting their spatial distribution (Arendt *et al*., 2012; Polovina *et al*., 2004). Future studies should therefore give fuller consideration to vertical environmental factors in order to provide a more three dimensional characterization of the key drivers influencing sea turtle migration and foraging behavior. Both the species occurrence data and environmental data used in this research were aggregated to a 5°×5° grid. Future research should move beyond annual 5°×5° analyses and adopt monthly or higher resolution data in order to better capture short term dynamic events such as El Niño and eddy edge effects. In addition, this research did not explicitly incorporate spatial autocorrelation or temporal dynamic processes at the modeling stage, making it difficult to distinguish distributional responses driven by environmental variables from structural signals arising from spatiotemporal continuity itself. Future work could therefore introduce spatiotemporal species distribution models capable of simultaneously accounting for spatial random fields and temporal dynamics. Longline and purse seine fisheries differ substantially in fishing practice, fishing grounds, operating depths, and bycatch processes, meaning that the two data sources are not fully equivalent in their observation mechanisms. In this research, the multi-source integration implemented within the MaxEnt framework mainly reflected the spatial integration of occurrence records, but did not explicitly correct for differences in bycatch efficiency or catchability between gears. Future studies should address these systematic differences more rigorously at the underlying technical level, for example by incorporating fishing effort as a covariate, in order to reduce systematic bias among fishing methods and further improve model accuracy.

4 Conclusions

Based on WCPFC Regional Observer Programme data from 2015 to 2019 and annual mean environmental fields from CMEMS, the optimized MaxEnt models developed for the western and central Pacific indicated that SST, DO, and EKE were the principal drivers of oceanic habitat suitability for loggerhead turtles. Compared with single gear models, the integrated L+S model, which combined observations from longline and purse seine fisheries, substantially improved the spatial completeness of habitat coverage and hotspot identification, revealing broader and more connected areas of high suitability and enhancing the utility of the model for bycatch risk assessment and management decision making. However, the present research was constrained by the use of annual mean data and a coarse spatiotemporal resolution of 5°×5°, uncertainties in attribution arising from the treatment of collinearity among variables, and the absence of explicit modeling of vertical environmental heterogeneity and spatiotemporal autocorrelation. As a result, model sensitivity to seasonal variation, short term eddy processes, and vertical habitat use patterns may have been weakened. To improve ecological interpretability and management relevance, future studies should incorporate monthly or higher resolution multilevel environmental profile data and adopt spatiotemporal species distribution models capable of simultaneously accounting for spatial random fields and temporal dynamics.

Acknowledgments

This research was supported by the National Key Research and Development Program (2024YFD2400605) and the 2025 Comprehensive Scientific Survey of High Seas Fishery Resources in the Central and Western Pacific (D-8025-25-5003). We also thank the Western and Central Pacific Fisheries Commission (WCPFC) Regional Observer Program for providing data support.

References

KONG WEIYAO, LI XINHAI, ZOU HONGFEI. Optimization of the Maximum Entropy Model in Species Distribution Prediction. Journal of Applied Ecology, 2019, 30(06):2116-2128.

LI GUUANGYAO, YANG SHENGLONG, Cheng Tianfei, et al. Analysis of Environmental Characteristics and Current Status of Forecasting in the Western and Central Pacific Tuna Fishing Grounds. China Agricultural Science and Technology Bulletin, 2025, 27(05):203-221.

ABECASSIS M, SENINA I, LEHODEY P, et al. A model of loggerhead sea turtle (*Caretta caretta*) habitat and movement in the oceanic North Pacific. PLoS One, 2013, 8(9): e73274.

ARENDT M D, SEGARS A L, BYRD J I, et al. Migration, distribution, and diving behavior of adult male loggerhead sea turtles (*Caretta caretta*) following dispersal from a major breeding aggregation in the Western North Atlantic. Marine Biology, 2012, 159(1): 113-125.

AROSTEGUI M C, GAUBE P, WOODWORTH-JEFCOATS P A, et al. Anticyclonic eddies aggregate pelagic predators in a subtropical gyre. Nature, 2022, 609(7927): 535-540.

ARUN P V, KATIYAR S K. Intelligent adaptive resampling technique for the processing of remotely sensed imagery. Annals of GIS, 2014, 20(1): 53-60.

CHELTON D B, GAUBE P, SCHLAX M G, et al. The influence of nonlinear mesoscale eddies on near-surface oceanic chlorophyll. Science, 2011, 334(6054): 328-332.

DEGENFORD J H, LIANG D, BAILEY H, et al. Using fisheries observation data to develop a predictive species distribution model for endangered sea turtles. Conservation Science and Practice, 2021, 3(2): e349.

DORMANN C F, ELITH J, BACHER S, et al. Collinearity: a review of methods to deal with it and a simulation research evaluating their performance. Ecography, 2013, 36(1): 27-46.

DUNKIN L, REIF M, ALTMAN S, et al. A spatially explicit, multi-criteria decision support model for loggerhead sea turtle nesting habitat suitability: A remote sensing-based approach. Remote Sensing, 2016, 8(7): 573.

DUQUESNE E, FOURNIER D. Climate change redefines sea turtle hotspots: Vessel strike risks and gaps in protected areas. Science advances, 2025, 11(26): eadw4495.

ETAGA H O, NDUBISI R C, OLUEBUBE N L. Effect of multicollinearity on variable selection in multiple regression. Science Journal of Applied Mathematics and Statistics, 2021, 9(6): 153.

FUJISAKI I, HART K M, BUCKLIN D, et al. Predicting multi-species foraging hotspots for marine turtles in the Gulf of Mexico. Endangered Species Research, 2020, 43: 253-266.

GAUBE P, BARCELÓC, MCGILLICUDDY JR D J, et al. The use of mesoscale eddies by juvenile loggerhead sea turtles (*Caretta caretta*) in the southwestern Atlantic. PLoS one, 2017, 12(3): e0172839.

HUGHES A C, ORR M C, MA K, et al. Sampling biases shape our view of the natural world. Ecography, 2021, 44(9): 1259-1269.

HOCHSCHEID S, BENTIVEGNA F, SPEAKMAN J R. The dual function of the lung in chelonian sea turtles: buoyancy control and oxygen storage. Journal of Experimental Marine Biology and Ecology, 2003, 297(2): 123-140.

HOYLE S D, CAMPBELL R A, DUCHARME-BARTTH N D, et al. Catch per unit effort modelling for stock assessment: A summary of good practices. Fisheries Research, 2024, 269: 106860.

International Union for Conservation of Nature (IUCN). The IUCN Red List of Threatened Species. Version 2023-1. (2023-01-01)[2026-01-16]. https://www.iucnredlist.org

KARP M A, BRODIE S, SMITH J A, et al. Projecting species distributions using fishery-dependent data. Fish and Fisheries, 2023, 24(1): 71-92.

KOBAYASHI D R, POLOVINA J J, PARKER D M, et al. Pelagic habitat characterization of loggerhead sea turtles, *Caretta caretta*, in the North Pacific Ocean (1997-2006): Insights from satellite tag tracking and remotely sensed data. Journal of Experimental Marine Biology and Ecology, 2008, 356(1-2): 96-114.

KOBAYASHI S, AOKURA N, FUJIMOTO R, et al. Incubation and water temperatures influence the performances of loggerhead sea turtle hatchlings during the dispersal phase. Scientific Reports, 2018, 8(1): 11911.

LEWISON R L, CROWDER L B, WALLACE B P, et al. Global patterns of marine mammal, seabird, and sea turtle bycatch reveal taxa-specific and cumulative megafauna hotspots. Proceedings of the National Academy of Sciences, 2014, 111(14): 5271-5276.

LIN H, WANG J, ZHU J, et al. Evaluating the impacts of environmental and fishery variability on the distribution of bigeye tuna in the Pacific Ocean. ICES Journal of Marine Science, 2023, 80(10): 2642-2656.

LIU X, ZHOU H, LIU H, et al. Characteristics and dynamics of the interannual eddy kinetic energy variation in the Western Equatorial Pacific Ocean. Progress in Oceanography, 2024, 229: 103358.

Maglietta R, Caccioppoli R, Piazzolla D, et al. Habitat suitability modeling of loggerhead sea turtles in the Central-Eastern Mediterranean Sea: a machine learning approach using satellite tracking data. Frontiers in Marine Science, 2024, 11: 1493598.

MANEL S, WILLIAMS H C, ORMEROD S J. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 2001, 38(5): 921-931.

MARIAANI G, BELLUCCI F, COCUMELLI C, et al. Dietary preferences of loggerhead sea turtles (*Caretta caretta*) in two Mediterranean feeding grounds: does prey selection change with habitat use throughout their life cycle?. Animals, 2023, 13(4): 654.

MARTINETTO P, ALEMANY D, BOTTO F, et al. Linking the scientific knowledge on marine frontal systems with ecosystem services. Ambio, 2020, 49(2): 541-556.

MATEAR R J, HIRST A C. Long-term changes in dissolved oxygen concentrations in the ocean caused by protracted global warming. Global Biogeochemical Cycles, 2003, 17(4).

MCGILLICUDDY Jr D J, ROBINSON A R, Siegel D A, et al. Influence of mesoscale eddies on new production in the Sargasso Sea. Nature, 1998, 394(6690): 263-266.

MUSCARELLA R, GALANTE P J, SOLEY-GUARDIA M, et al. ENM eval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods in Ecology and Evolution, 2014, 5(11): 1198-1205.

PACIFICI K, REICH B J, MILLER D A W, et al. Integrating multiple data sources in species distribution modeling: a framework for data fusion. Ecology, 2017, 98(3): 840-850.

PATEL S H, WINTON M V, HATCH J M, et al. Projected shifts in loggerhead sea turtle thermal habitat in the Northwest Atlantic Ocean due to climate change. Scientific Reports, 2021, 11(1): 8850.

PHILLIPS S J, ANDERSON R P, SCHAPIRE R E. Maximum entropy modeling of species geographic distributions. Ecological Modelling, 2006, 190(3/4): 231-259.

PHILLIPS S J, DUDíK M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography, 2008, 31(2): 161-175.

POLOVINA J J, HOWELL E, KOBAYASHI D R, et al. The transition zone chlorophyll front, a dynamic global feature defining migration and forage habitat for marine resources. Progress in oceanography, 2001, 49(1-4): 469-483.

POLOVINA J J, BALAZS G H, HOWELL E A, et al. Forage and migration habitat of loggerhead (*Caretta caretta*) and olive ridley (*Lepidochelys olivacea*) sea turtles in the central North Pacific Ocean. Fisheries Oceanography, 2004, 13(1): 36-51.

ROBINSON N J, AGUZZI J, ARIAAS S, et al. Global trends in sea turtle research and conservation: Using symposium abstracts to assess past biases and future opportunities. Global Ecology and Conservation, 2023, 47: e02587.

SCHMID M S, COWEN R K, ROBINSON K, et al. Prey and predator overlap at the edge of a mesoscale eddy: fine-scale, in-situ distributions to inform our understanding of oceanographic processes. Scientific Reports, 2020, 10(1): 921.

SHIMAADA T, LIMPUS C J, HAMANN M, et al. Fidelity to foraging sites after long migrations.

Journal of Animal Ecology, 2020, 89(4): 1008-1016.

SILLERO N, ARENAS-CASTRO S, ENRIQUEZ-URZELAI U, et al. Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. Ecological Modelling, 2021, 456: 109671.

THUAL S, DEWITTE B. A fresh perspective on ENSO nonlinearity: the ENSO pattern continuum metric. npj Climate and Atmospheric Science, 2025, 8(1): 352.

WALLACE B P, DIMATTEO A D, HURLEY B J, et al. Regional management units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. PLoS One, 2010, 5(12): e15465.

WARREN D L, SEIFERT S N. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecological Applications, 2011, 21(2): 335-342.

ZAMPOLLO A, ARCANGELI A, COSTANTINO M, et al. Seasonal niche and spatial distribution modelling of the loggerhead (*Caretta caretta*) in the Adriatic and Ionian seas. Aquatic Conservation: Marine and Freshwater Ecosystems, 2022, 32(7): 1141-1155.