The impact on the inclusion of marine subsurface variables on habitat modelling of swordfish in the Indian Ocean

Wei Tang, Xuefang Wang*, Feng Wu, Jiangfeng Zhu

College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China

Correspondence: xfwang@shou.edu.cn

Abstract: Species distribution models (SDM) have emerged as an important technique for defining and forecasting species habitats. In the maritime domain, species distribution models have historically been constructed primarily as two-dimensional species occurrence and environmental data, with a lack of understanding of the real ecological environments of species. Our capacity to examine subsurface marine characteristics has grown thanks to developments in models and technology, but their integration into SDM is still somewhat restricted. We explored the impact of adding subsurface environmental factors in SDM on species habitat suitability to define habitat variation of adult swordfish (*Xiphias gladius*) in the Indian Ocean using environmental data from various depth layers in the research region. Data from the 2017–2019 Chinese Indian Ocean tuna longline fishery observer were used to generate the species data. In order to determine the best depth for simulating swordfish habitat, we utilized MaxEnt to investigate the impacts of subsurface factors and combine several indicators to evaluate the model's effectiveness. The significance of environmental elements was quickly assessed, and the connection between habitat appropriateness and CPUE was skimmed over. The outcomes demonstrated that the model's effectiveness was increased by include subsurface environmental factors, displaying a clear increasing trend and stabilizing after reaching the ideal level (300 m). Although the adult swordfish's distribution range for habitat suitability shrank little as depth increased, its core habitat suitability value increased dramatically. Among all environmental factors, T and DO were closely related to adult swordfish habitat distribution. In contrast, there was no significant correlation between habitat suitability and CPUE. Incorporating three-dimensional environmental considerations into the model yields more realistic results than model results that consider habitat as two-dimensional. Develop explicit three-dimensional methods to better represent the distribution of marine species.

Keywords: Maximum entropy model, adult swordfish, Indian Ocean, subsurface variables, habitat simulation

1 Introduction

Species Distribution Model (SDM), a model commonly used in theoretical and applied research in ecology and biogeography (Peterson et al., 2015; Peterson and Soberón, 2012), has grown in recent years with the increase in computer processing power and the availability of environmental datasets and species distribution data, applications have gradually grown in recent years in many areas (Hortal et al., 2012), the most common of which are determining suitable habitat for species (Guisan and Thuiller, 2005), predicting the effects of future climate change on species distributions (Sahlean et al., 2014), assessing the invasive potential of exotic species invasion potential (Jiménez-Valverde et al., 2011) and conservation planning (Hu et al., 2020), and species distribution models applied to marine species are becoming increasingly popular (Robinson et al., 2017).

However, the vast majority of current SDM implementations treat species occurrence and environmental data as two-dimensional, an approach that is a reasonable fit for species in shallow waters (Bentlage et al., 2013). However, for marine species inhabiting true 3D environments, given the lack of information on the 3D occurrence of marine species (Webb et al., 2010), the available environmental data layers (Assis et al., 2017a), and the limited understanding of the ecology and behavior of many organisms (Bentlage et al., 2013), most SDM applications often ignore or struggle to address the three-dimensionality of marine species habitats (Melo-Merino et al., 2020).

As a result, Duffy and Chown advocate the development of explicit 3D methods that extend SDM into 3D space (Duffy and Chown, 2017). To date, several approaches have also emerged to consider the marine 3D environment, such as a modified SDM approach that approximates the representation of the true 3D habitat of a species by stacking 2D SDM at different depths layer by layer (Bentlage et al., 2013); or treating the multi-depth layer as a single three-dimensional layer and then filtering relevant variables from all environmental information for modelling (Alabia et al., 2016); then there is the use of a combination of both approaches, dividing the multi-depth layer into a smaller number of composite depth layers, which are then stacked (Venegas-Li et al., 2017; Pérez-Costas et al., 2019).

Pelagic species are functional groups of fish that typically inhabit oceanic regions, and many previous studies have described their habitat use only through surface environmental variables, spatial variables (longitude, latitude) and some typical depth-related variables (e.g., MLD and BAH (ocean bathymetry, BAH), Chang et al., 2012). As these species develop, however, they can use deeper water layers as habitat space, so using only two-dimensional models to describe their habitat may not be sufficient (Dambach and Rödder, 2011). For example, Igarashi (Igarashi et al., 2018) and Brodie (Brodie et al., 2018) et al. tested model performance by adding subsurface environmental variables to studies of Pacific neon flying squid and four species of fish from the U.S. coast and found significant improvements in model performance.

As a globally distributed and highly migratory large pelagic fish, swordfish (*Xiphias gladius*) are an important target and by-catch species for longline fleets (Abecassis et al., 2012). Due to the biological demands of feeding and reproduction, swordfish move in response to changes in their biotic and abiotic environment (Humston et al., 2000). This mobility, in turn, will determine the species' spatial and temporal scale responses to environmental conditions, including preferences for real habitats (Pickens et al., 2021). In previous studies of swordfish habitat distribution, the vast majority of swordfish habitats have been treated as two-dimensional. Therefore, it is important to incorporate the third-dimensional attributes of the real habitat environment of swordfish into the model. In addition to the complexity of three-dimensional marine habitats, other aspects, such as the different environmental ecological niches of juvenile and adult individuals, should be considered when applying SDM (Dambach and Rödder, 2011). For swordfish, there are differences in temperature tolerance (Poisson and Fauvel, 2009), spawning migrations (Arocha, 2007), etc. between sexes and sizes of individuals. A finer division should be made in the study.

Among the many approaches to SDM, the maximum entropy model(MaxEnt, Phillips et al., 2006) has been one of the most widely used modeling techniques for predicting species distributions from partial data (Melo-Merino et al., 2020).MaxEnt is designed to work with presence-only data, requiring relatively little information (Phillips et al., 2006), has robust prediction accuracy (Elith et al., 2006; Graham et al., 2008) even with small samples (Hernandez et al., 2006; Wisz et al., 2008), and provides, through user-friendly interfaces (e.g., functions and regularization multipliers) in model construction great flexibility, allowing the user to adapt the model to specific needs and available information (Merow et al., 2013).

In this study, we attempted to characterize the true three-dimensional preferences of adult Indian Ocean swordfish for their habitat, with environmental characteristics at depth being explicitly considered in the model. Starting from the surface environment and adding environmental variables from different depth layers downwards in a stepwise manner to constitute a new composite environmental layer, this approach was attempted through a swordfish habitat study, using MaxEnt, with a view to (1) assessing the impact of different maximum depths of the composite environmental layer on modeling swordfish habitat, (2) assessing the appropriate depth of the composite environmental layer for inclusion in the MaxEnt model, and (3) comparing the three-dimensional approach with the results of the two-dimensional approach. Finally, it is hoped that this will provide SDM with experience in developing a 3D spatial approach.

2 Materials and methods

2.1 Data sources

2.1.1 Species occurrence data

In this study, raw data from 2441 swordfish catches recorded in the Chinese Indian Ocean tuna longline fishery observer data from 2017 to 2019, including year of operation, month, fishing location, sex, and lower jaw-to-fork length, were used in this study. Given the range of occurrence data, we selected maritime space from 30°N to 40°S and 30°E to 80°E as the study area (Fig. 1).

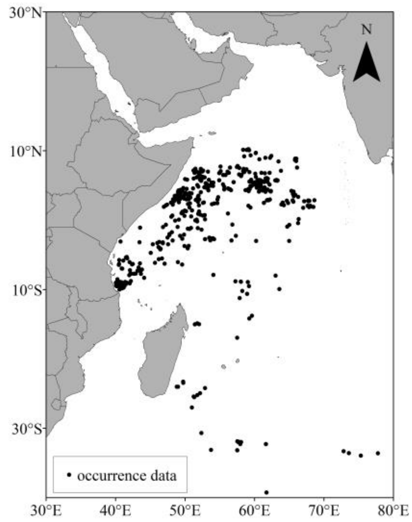

Figure 1 Study area and species occurrence data (small black dots)

2.1.2 Environmental data

Marine environmental data were obtained from the COPERNICUS website (http://marine.copernicus.eu/), an EU project that provides data products for various marine applications to the scientific community. Based on previous experience with swordfish studies (Chang et al., 2012; Evans et al., 2014; Brodie et al., 2018; Su et al., 2020), we selected seven variables from different datasets on the COPERNICUS website that were downloaded as netCDF files: temperature, salinity, dissolved oxygen, total plant primary production, northward sea water velocity, eastward sea water velocity, and mixed layer depth (see Table 1 for details).

Table 1 Environmental variables used in the maximum entropy model

Variable name	Acronym	Unit	Raw resolution	Depth range
sea water temperature	T	°C	0.25°×0.25°	surface to 500
sea water salinity	S	‰	0.25°×0.25°	surface to 500
net primary production	NPP	$mgC\cdot m^{-2}\cdot day^{-1}$	0.25°×0.25°	surface to 500
moole concentration of dissolved oxygen	DO	$mmol\cdot m^{-3}$	0.25°×0.25°	surface to 500
northward sea water velocity	NV	$m\cdot s^{-1}$	0.083° × 0.083°	surface

eastward sea water velocity	EV	$m\cdot s^{-1}$	0.083° × 0.083°	surface
ocean mixed layer	MLD	m	0.083° × 0.083°	surface

2.2 Data pre-processing

2.2.1 Species occurrence data screening

According to Poisson and Fauvel (Poisson and Fauvel, 2009), the first maturity size (L_{50}) for Indian Ocean swordfish females and males is 170 cm and 120 cm (lower jaw-to-fork length, LJFL), respectively. Given that catch sex records are incomplete in the original fishery data. Therefore, for catch information with sex records, adults were selected according to LJFL of 170 cm and 120 cm, respectively; for information without clear sex records, categories were classified according to female swordfish first maturity size (L_{50}), and catch records with LJFL ≥ 170 cm were retained. Also, 902 catch records were finally retained considering the sample size in different months and the coupling with environmental data on the time scale (Fig. 1).

2.2.2 CPUE calculation

Catch per unit effort (CPUE) is used in many fisheries studies as a common indicator of abundance of fishery resources and the spatial distribution of their habitats (Maunder and Punt, 2004) (Igarashi et al., 2018; Su et al., 2020). However, it is unclear whether there is always a strong linear association between predicted habitat suitability or probability of occurrence and species abundance (Vanderwal et al. 2009; Dallas and Hastings 2018). Therefore, this study calculates CPUE with the results of the final model to make a brief observation. For longline fisheries, fishing effort is typically expressed as the number of hooks placed (Wu et al., 2021). CPUE is calculated as follows：

$$CPUE = \frac{U_{catch}}{f_{hooks}} \times 1000$$

where U_{catch} is the cumulative catch in tails (ind) for the location and f_{hooks} is the total number of hooks cast in hooks for the location.

2.2.3 Environmental data processing

To avoid multiple correlations among environmental variables, the variance inflation factor (VIF) was used to measure the multicollinearity among environmental factors. The results were calculated by SPSS 20.0 and the VIF values of each factor were less than 10 (Table 2).

Table 2 Correlation coefficients for surface environmental factors

Pearson'r	**DO**	**EV**	**MLD**	**NPP**	**NV**	**S**	**T**	**VIF**
DO	1.000							5.370
EV	-0.070	1.000						1.185
MLD	0.217	0.217	1.000					1.906

NPP	0.259	0.086	-0.152	1.000				1.188
NV	0.005	-0.164	-0.190	0.059	1.000			1.058
S	-0.050	0.245	0.092	-0.065	-0.040	1.000		1.080
T	-0.872	0.000	-0.480	-0.205	0.053	-0.010	1.000	6.735

For MaxEnt, the inputs to the model must be of the same resolution in order to be used for matching. ArcMap 10.4 software was used to convert the netCDF format files to TIFF format raster layers, which were later converted to ASCII format files with a spatial resolution of 0.25° × 0.25° and the same boundaries. The processed environmental data were used to filter the species occurrence data to reduce sampling bias in the species occurrence data and to keep the species occurrence at the same resolution as the environmental data.

2.3 Analysis options

Using the software MAXENT 3.4.1 (https://biodiversityinformatics.amnh.org/open_source/ maxent), 75% of the occurrences in the swordfish sample were randomly set as the training set, and the remaining 25% of occurrences were used as the test set, and the calculation was repeated 50 times to eliminate randomness, and the results were output in logistic format. The interrelationship between environmental factors and habitat suitability of swordfish was analyzed by combining the contribution rate with the Jackknife method. The rest of the settings were kept as default. In this case, only the surface model environmental variables were chosen, and then the corresponding depth environmental variables were added in 100 m intervals for modeling until the maximum depth of 500 m was reached.

A combination of methods was used for comparison in order to assess the model efficacy across depth layers. The magnitude of the Area Under Curve (AUC) of the Receiver Operating Characteristic Curve (ROC) was used as an indicator to assess MaxEnt prediction accuracy (Phillips et al., 2006). In a conservative approach, the habitat suitability value at Equal test Sensitivity and Specificity (ESS) is used as the threshold to measure model sensitivity (Liu et al., 2013). Omission error and Commission error of the models are assessed using true skill statistics (TSS; Allouche et al., 2006).The AIC (Akaike information criterion; Akaike, 1974) and BIC (Bayesian information criteria; Schwarz, 1978) were calculated for each model using ENMTools (http://purl.oclc.org/enmtools/) values to synthetically assess model complexity and fit in place of the regularization parameters in the MaxEnt model setup with feature classes to select the optimal model (Phillips and Dudík, 2008).

3 Results

3.1 Assessment of models based on evaluation indicators

The overall trend in AUC was observed to show an increasing trend in model efficacy with the addition of subsurface environmental variables up to 200 m; after 200 m it tended to a steady state. Observed from the general level of AUC, the model efficacy was best at 300 m (Fig. 2). Based on the overall trend observed for Sensitivity and TSS, the model's ability to assess omission and misclassification errors tends to increase with the addition of subsurface environmental variables until 200 m; it tends to a steady state after 200 m. At a general level, the model's ability to assess both types of errors was strongest at 300 m (Fig. 3). Based on the AIC and BIC metrics, the complexity and fit of the model reach the optimal value at 300 m (Fig. 4). Thus, considering all metrics together, the model's integrated capability shows a clear upward trend with the addition of subsurface environmental variables until 200 m; it reaches an optimal level at 300 m.

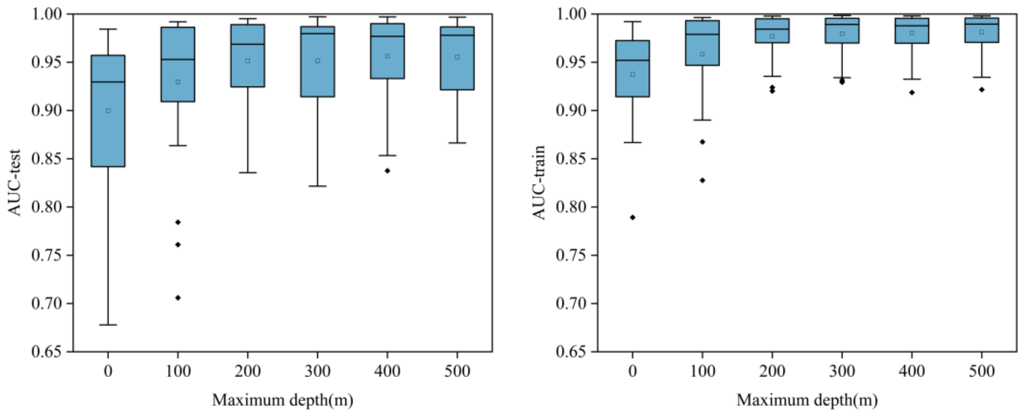

Figure 2 Comparison of AUC values for each depth model

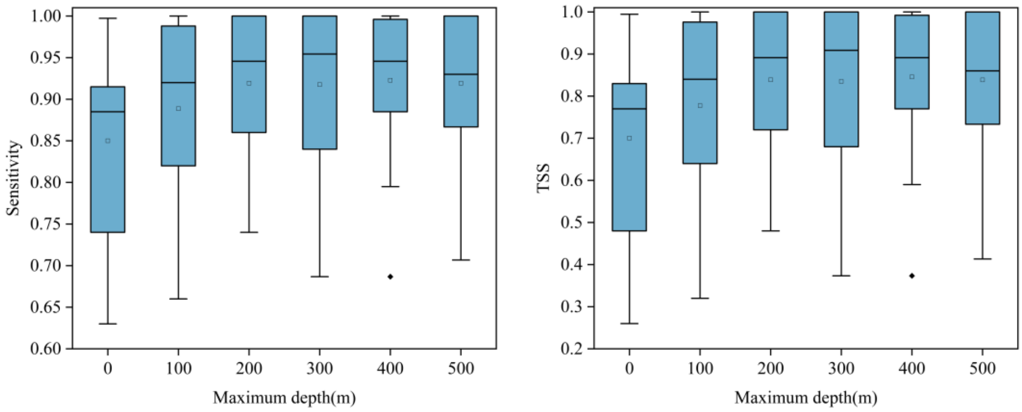

Figure 3 Comparison of Sensitivity and TSS values for each depth model

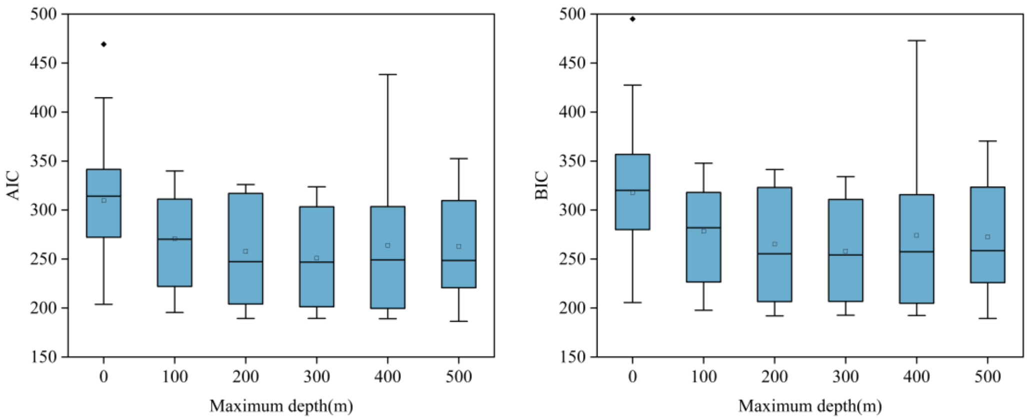

Figure 4 Comparison of AIC and BIC values for each depth model

3.2 Comparison of results of different depth layer models

We selected an excellent model result (AUC_{test} = 0.9666) and a poorly performing model result (AUC_{test} = 0.6778) from the surface results, and compared them visually with their different depth layers, respectively (Fig. 5 and 6). As can be seen from the plots, the range of habitat suitability of the different depth models for the two different months contracted to some extent with the inclusion of environmental variables at different depths at and before 200 m; after this it was largely stable (a-f in Fig. 5 and 6). The model at the optimal level (300 m) in both months, however, showed some increase in habitat suitability values for the major habitat suitability range compared to the results of the surface model.

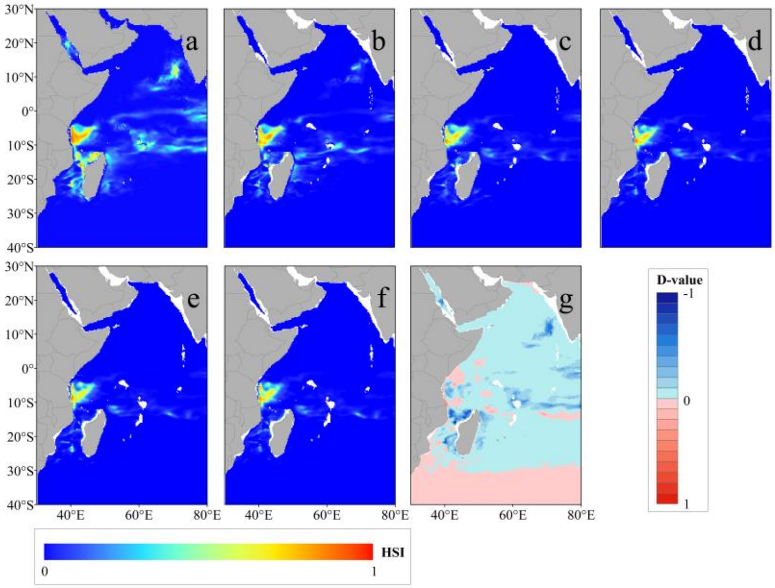

Figure 5 Model effects by depth in January 2017 (a-f are models at depths from the surface to 500 m, respectively, increasing in 100 m intervals; g is a plot of the difference in suitability values for models at 300 m minus the suitability values for models at the surface)

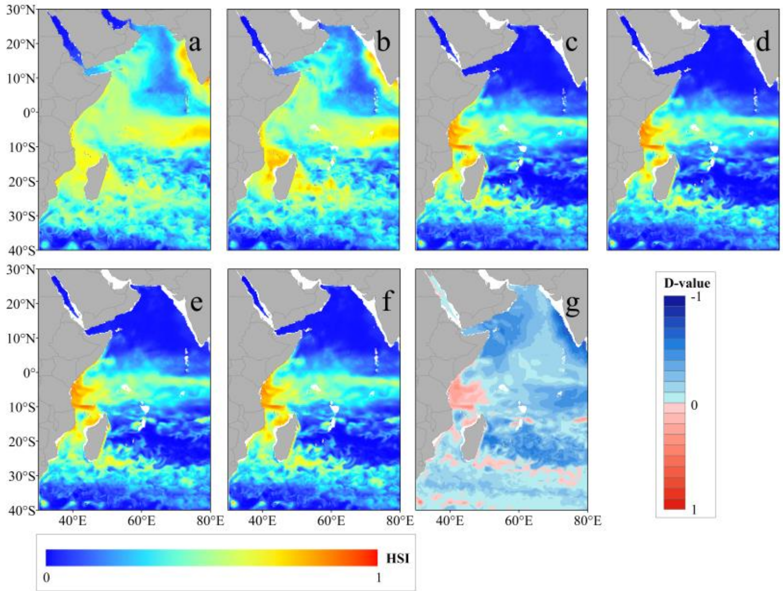

Figure 6 Model effects by depth, October 2017 (a-f are models for surface to 500 m depths, respectively, increasing at 100 m intervals; g is a plot of the difference in suitability values for models at 300 m minus the suitability values for models at the surface)

3.3 Importance of environmental factors

From the contributions of each component, it was found that, at 100 m and below, T and NPP outperformed several other categories of factors in their contribution to the model; at 200 m and above, T and DO outperformed several other categories of factors in their contribution to the model (Fig. 7). No matter the depth range, the three types of variables T, NPP, and DO were closely connected with the distribution of swordfish habitat suitability, according to the knife-cut method's findings (Fig. 8). T, DO, NPP, and MLD were highly associated with the model at its optimum level.

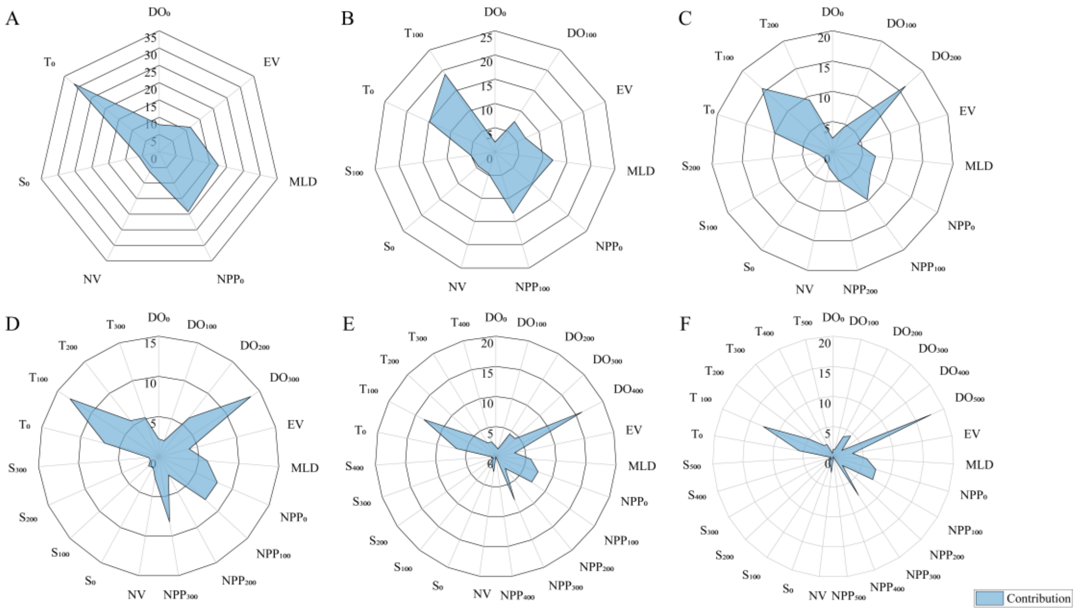

Figure 7 Average contribution of environmental factors across depth layers (A-F are modeled for surface to 500 m depth, respectively, increasing at 100 m intervals)

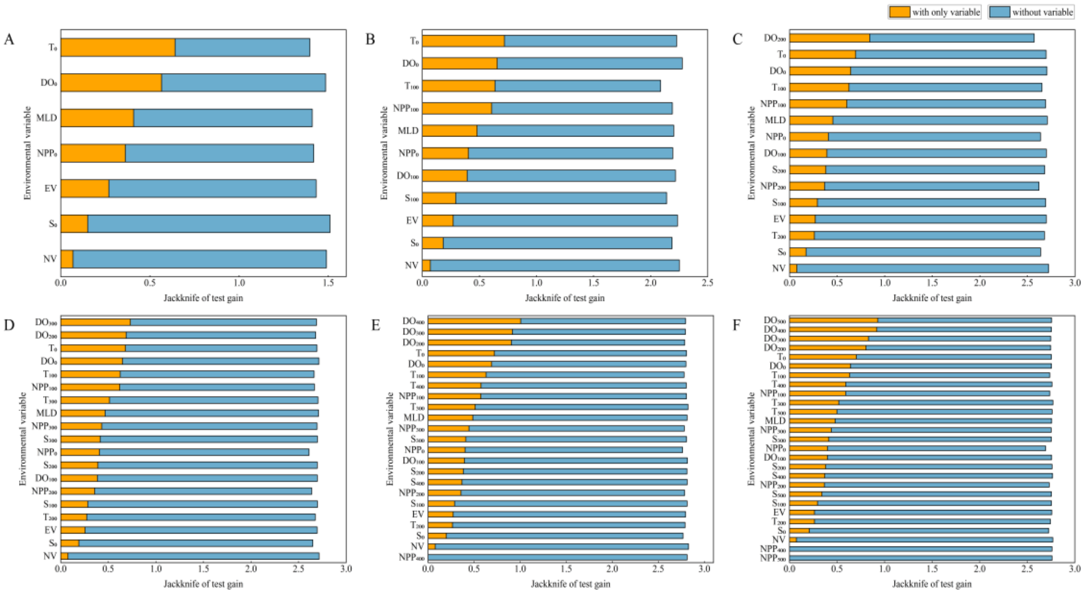

Figure 8 Results of the Jackknife method of averaging environmental factors across depth layers (A-F are modeled for surface to 500 m depth, respectively, increasing at 100 m intervals)

3.4 Relationship between HSI and CPUE

The linear relationship between swordfish CPUE (ln(CPUE)) and HSI for the different depth models is evident in the graph (Fig. 9). Regression analysis showed that the HSI correlations for the different depth models were all statistically significant at the 0.05 level, but there was not a strong correlation.

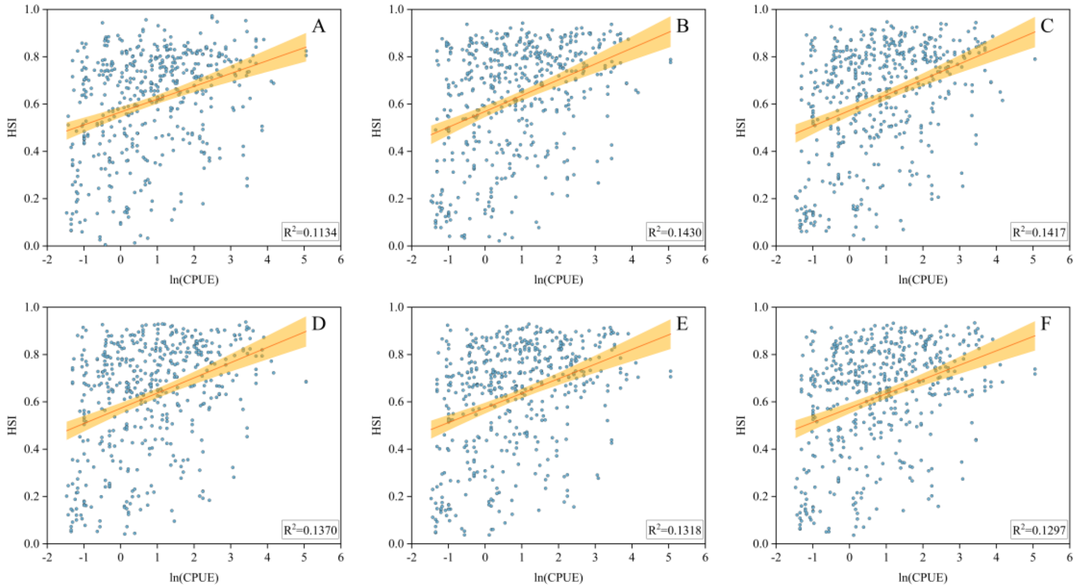

Figure 9 Comparison of CPUE and HSI for each depth layer (A-F are models for surface to 500 m depth, respectively, increasing at 100 m intervals)

4 Summary

We used a maximum entropy model, using three years of catch data from the longline fleet, by gradually adding subsurface variables to the model with the view to simulating that swordfish meet suitable realistic habitat conditions in the area. With the gradual increase in subsurface variables, the model's effectiveness showed an increasing trend and stabilized after 300 m. In the model at all depths, T, DO and NPP were strongly correlated with swordfish habitat distribution in the model. These provide some lessons for adult swordfish habitat in the Indian Ocean region. Also, the CPUE data of swordfish did not show a significant correlation with the model results. Considering the three-dimensional environment into the model gives more realistic results than considering the habitat as a two-dimensional model result. Develop an explicit three-dimensional approach to better represent the distribution of marine species.

Acknowledgements

We thank the Tuna Technical Group of the China Ocean Fisheries Association for providing the fishery scientific observer data of the Chinese fleet for this study. We are also deeply grateful for the collaboration of scientific observers and tuna longline fishing vessels.

Reference

Abecassis M, Dewar H, Hawn D, et al. Modeling swordfish daytime vertical habitat in the North Pacific Ocean from pop-up archival tags. Marine Ecology Progress Series, 2012, 452: 219-236.

Akaike H. A new look at the statistical model identification. IEEE transactions on automatic control, 1974, 19(6): 716-723.

Alabia I D, Saitoh S I, Igarashi H, et al. Ensemble squid habitat model using three-dimensional ocean data. ICES Journal of Marine Science, 2016, 73(7): 1863-1874.

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

Arocha F. Swordfish reproduction in the Atlantic Ocean: and overview. Gulf Caribb. Res. 2007, 19, 21–36.

Assis J, Tyberghein L, Bosch S, et al. Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. Global Ecology and Biogeography, 2018, 27(3): 277-284.

Bentlage B, Peterson A T, Barve N, et al. Plumbing the depths: extending ecological niche modelling and species distribution modelling in three dimensions. Global Ecology and Biogeography, 2013, 22(8): 952-961.

Brodie S, Jacox M G, Bograd S J, et al. Integrating dynamic subsurface habitat metrics into species distribution models. Frontiers in Marine Science, 2018, 5: 219.

Chang Y J, Sun C L, Chen Y, et al. Habitat suitability analysis and identification of potential fishing grounds for swordfish, *Xiphias gladius*, in the South Atlantic Ocean. International Journal of Remote Sensing, 2012, 33(23): 7523-7541.

Dallas T A, Hastings A. Habitat suitability estimated by niche models is largely unrelated to species abundance. Global Ecology and Biogeography, 2018, 27(12): 1448-1456.

Dambach J, Rödder D. Applications and future challenges in marine species distribution modeling. Aquatic Conservation: Marine and Freshwater Ecosystems, 2011, 21(1): 92-100.

Duffy G A, Chown S L. Explicitly integrating a third dimension in marine species distribution modelling. Marine ecology progress series, 2017, 564: 1-8.

Elith J, H. Graham C, P. Anderson R, et al. Novel methods improve prediction of species' distributions from occurrence data. Ecography, 2006, 29(2): 129-151.

Evans K, Abascal F, Kolody D, et al. The horizontal and vertical dynamics of swordfish in the South Pacific Ocean. Journal of Experimental Marine Biology and Ecology, 2014, 450: 55-67.

Graham C H, Elith J, Hijmans R J, et al. The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology, 2008, 45(1): 239-247.

Guisan A, Thuiller W. Predicting species distribution: offering more than simple habitat models. Ecology letters, 2005, 8(9): 993-1009.

Hernandez P A, Graham C H, Master L L, et al. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 2006, 29(5): 773-785.

Hortal J, Lobo J M, Jiménez-Valverde A. Basic questions in biogeography and the (lack of) simplicity of species distributions: putting species distribution models in the right place. Natureza & Conservação, 2012, 10(2): 108-118.

Hu W, Wang Y, Zhang D, et al. Mapping the potential of mangrove forest restoration based on species distribution models: A case study in China. Science of The Total Environment, 2020, 748: 142321.

Humston R, Ault J S, Lutcavage M, et al. Schooling and migration of large pelagic fishes relative to environmental cues. Fisheries Oceanography, 2000, 9(2): 136-146.

Igarashi H, Saitoh S I, Ishikawa Y, et al. Identifying potential habitat distribution of the neon flying squid (*Ommastrephes bartramii*) off the eastern coast of Japan in winter. Fisheries Oceanography, 2018, 27(1): 16-27.

Jiménez-Valverde A, Peterson A T, Soberón J, et al. Use of niche models in invasive species risk assessments. Biological invasions, 2011, 13(12): 2785-2797.

Liu C, White M, Newell G. Selecting thresholds for the prediction of species occurrence with presence-only data. Journal of biogeography, 2013, 40(4): 778-789.

Maunder M N, Punt A E. Standardizing catch and effort data: a review of recent approaches. Fisheries research, 2004, 70(2-3): 141-159.

Melo-Merino S M, Reyes-Bonilla H, Lira-Noriega A. Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. Ecological Modelling, 2020, 415: 108837.

Merow C, Smith M J, Silander Jr J A. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography, 2013, 36(10): 1058-1069.

Pérez-Costas E, Guisande C, González-Vilas L, et al. NOO3D: A procedure to perform 3D species distribution models. Ecological Informatics, 2019, 54: 101008.

Peterson A T, Papeş M, Soberón J. Mechanistic and correlative models of ecological niches. 2015.

Peterson A T, Soberón J. Species distribution modeling and ecological niche modeling: getting the concepts right. Natureza & Conservação, 2012, 10(2): 102-107.

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.

Pickens B A, Carroll R, Schirripa M J, et al. A systematic review of spatial habitat associations and modeling of marine fish distribution: A guide to predictors, methods, and knowledge gaps. PloS one, 2021, 16(5): e0251818.

Poisson F, Fauvel C. Reproductive dynamics of swordfish (*Xiphias gladius*) in the southwestern Indian Ocean (Reunion Island). Part 1: oocyte development, sexual maturity and spawning . Aquatic Living Resources, 2009, 22(1): 45-58.

Robinson N M, Nelson W A, Costello M J, et al. A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. Frontiers in Marine Science, 2017, 4: 421.

Sahlean T C, Gherghel I, Papeş M, et al. Refining climate change projections for organisms with low dispersal abilities: a case study of the Caspian whip snake. PloS one, 2014, 9(3): e91994.

Schwarz G. Estimating the dimension of a model. The annals of statistics, 1978: 461-464.

Su N J, Chang C H, Hu Y T, et al. Modeling the spatial distribution of swordfish (*Xiphias gladius*) using fishery and remote sensing data: approach and resolution. Remote Sensing, 2020, 12(6): 947.

VanDerWal J, Shoo L P, Johnson C N, et al. Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. The American Naturalist, 2009, 174(2): 282-291.

Venegas-Li R, Levin N, Possingham H, et al. 3D spatial conservation prioritisation: Accounting for depth in marine environments. Methods in Ecology and Evolution, 2018, 9(3): 773-784.

Webb T J, Vanden Berghe E, O'Dor R. Biodiversity's big wet secret: the global distribution of marine biological records reveals chronic under-exploration of the deep pelagic ocean. PloS one, 2010, 5(8): e10223.

Wisz M S, Hijmans R J, Li J, et al. Effects of sample size on the performance of species distribution models. Diversity and distributions, 2008, 14(5): 763-773.

Wu X H, Liu S Y V, Wang S P, et al. Distribution patterns and relative abundance of shortfin mako shark caught by the Taiwanese large-scale longline fishery in the Indian Ocean. Regional Studies in Marine Science, 2021, 44: 101691.