

Article

Spatial Prediction and Environmental Response of Skipjack Tuna Resources from the Perspective of Geographic Similarity: A Case Study of Purse Seine Fisheries in the Western and Central Pacific

Shuyang Feng¹, Xiaoming Yang^{1,2,3,4,5,*}, Menghao Li¹, Zhoujia Hua¹, Siquan Tian^{1,2,3,4,5}
and Jiangfeng Zhu^{1,2,3,4,5}

¹ College of Marine Living Resource Sciences and Management, Shanghai Ocean University, Shanghai 201306, China; m230200740@st.shou.edu.cn (S.F.); m230200732@st.shou.edu.cn (Z.H.); sqtian@shou.edu.cn (S.T.); jfzhu@shou.edu.cn (J.Z.)

² National Engineering Research Center for Oceanic Fisheries, Shanghai Ocean University, Shanghai 201306, China

³ Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai 201306, China

⁴ Key Laboratory of Oceanic Fisheries Exploration, Ministry of Agriculture and Rural Affairs, Ministry of Agriculture and Rural Affairs, Shanghai 201306, China

⁵ Scientific Observing and Experimental Station of Oceanic Fishery Resources, Ministry of Agriculture and Rural Affairs, Shanghai 201306, China

* Correspondence: xmyang@shou.edu.cn

Abstract

Skipjack tuna constitutes a crucial fishery resource in the Western and Central Pacific Ocean (WCPO) purse seine fishery, with high economic value and exploitation potential. It also serves as an essential subject for studying the interaction between fishery resource dynamics and marine ecosystems, as its resource abundance is significantly influenced by marine environmental factors. Skipjack tuna can be categorized into unassociated schools and associated schools, with the latter being predominant. Overfishing of the associated schools can adversely affect population health and the ecological environment. In-depth exploration of the spatial distribution responses of these two fish schools to environmental variables is significant for the rational development and utilization of tuna resources and for enhancing the sustainability of fishery resources. In sparsely sampled and complex marine environments, geographic similarity methods effectively predict tuna resources by quantifying local fishing ground environmental similarities. This study introduces geographical similarity theory. This study focused on $1^{\circ} \times 1^{\circ}$ fishery data (2004–2021) released by the Western and Central Pacific Fisheries Commission (WCPFC) combined with relevant marine environmental data. We employed Geographical Convergent Cross Mapping (GCCM) to explore significant environmental factors influencing catch and variations in causal intensity and employed a Geographically Optimal Similarity (GOS) model to predict the spatial distribution of catch for the two types of tuna schools. The research findings indicate that the following: (1) Sea surface temperature (SST), sea surface salinity (SSS), and net primary productivity (NPP) are key factors in GCCM model analysis, significantly influencing the catch of two fish schools. (2) The GOS model exhibits higher prediction accuracy and stability compared to the Generalized Additive Model (GAM) and the Basic Configuration Similarity (BCS) model. R^2 values reaching 0.656 and 0.649 for the two types of schools, respectively, suggest that the geographical similarity method has certain applicability and application potential in the spatial prediction of fishery resources. (3) Uncertainty analysis revealed more stable predictions for unassociated schools, with 72.65% of the results falling within the low-uncertainty range (0.00–0.25), compared to



Academic Editor: Alen Soldo

Received: 19 May 2025

Revised: 2 July 2025

Accepted: 8 July 2025

Published: 29 July 2025

Citation: Feng, S.; Yang, X.; Li, M.; Hua, Z.; Tian, S.; Zhu, J. Spatial Prediction and Environmental Response of Skipjack Tuna Resources from the Perspective of Geographic Similarity: A Case Study of Purse Seine Fisheries in the Western and Central Pacific. *J. Mar. Sci. Eng.* **2025**, *13*, 1444. <https://doi.org/10.3390/jmse13081444>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

52.65% for associated schools. This study, based on geographical similarity theory, elucidates differential spatial responses of distinct schools to environmental factors and provides a novel approach for fishing ground prediction. It also provides a scientific basis for the dynamic assessment and rational exploitation and utilization of skipjack tuna resources in the Pacific Ocean.

Keywords: geographical similarity; *Katsuwonus pelamis*; GCCM; environmental response; Western and Central Pacific Ocean

1. Introduction

Skipjack tuna (*Katsuwonus pelamis*) is central to the global tuna fishery. In the Western and Central Pacific Ocean (WCPO), skipjack catches account for over 65% of the world's total production, making it the primary fishing ground for this species, with purse seine fishing being the dominant harvesting method [1,2]. In practical operations and resource studies, skipjack tuna are often classified according to their schooling characteristics, mainly into two categories. One is the unassociated schools (UNA), which refers to tuna schools that migrate and forage autonomously in the natural environment without attaching to other structures. The other is the associated schools (ASS), which usually forms aggregations around objects such as natural floating debris, artificial fish aggregating devices (FADs), or cetaceans. Current research demonstrates that oceanic environmental variables significantly influence skipjack catch rates and spatial distribution [3]. Therefore, advancing the understanding of skipjack–environment relationships and developing predictive models using these variables is critically important for fisheries management. However, despite the existence of various methods for exploring the relationships between fishery resources and the environment, there is still a lack of in-depth investigation into the differences in environmental responses among different types of fish schools at the spatial scale. In particular, when facing complex environmental backgrounds, how to enhance the local adaptability and mechanistic interpretability of models remains a critical issue that urgently needs to be addressed.

Currently, scholars have conducted extensive research on the relationship between skipjack tuna and the marine environment. Commonly used methods mainly include the Generalized Additive Model (GAM) [4–6] based on statistical theory and the Habitat Suitability Index (HSI) model [7], as well as machine learning models such as the Maximum Entropy Model (MaxEnt) [8,9] and the BP neural network model [10,11]. However, these statistical and machine learning-based approaches are typically applicable at global scales, with model development contingent upon comprehensive observational coverage of environmental variable spaces, where predictive accuracy heavily depends on the quality and quantity of available data [12,13]. Furthermore, while machine learning models can capture complex patterns, they may struggle with interpretability and extrapolation, especially in ecologically dynamic systems. Moreover, most models assume that the influence of environmental variables on skipjack tuna distribution remains consistent across the entire study area, failing to account for spatial heterogeneity [13]. In reality, over large-scale spatial domains, the relationship between fishing vessel operations and environmental factors often exhibits significant spatial non-stationarity [14]. Neglecting this characteristic may lead to prediction biases in the models at local scales, making it difficult to accurately capture the local adaptability of the resources. It is not feasible to design large-scale spatial studies and conduct controlled experiments to reveal causal relationships.

Traditional spatial statistical methods are primarily employed to identify correlations among variables. However, they do not directly incorporate the concept of causality, even

though many terms with implicit causal meanings are frequently used in their descriptions. It is crucial to clarify that correlation does not necessarily imply causation. Even when a causal relationship exists, the correlation between the two may be influenced by factors such as selection bias and survivorship bias. Given the unique characteristics of spatial data, including non-randomness, non-replicability, and synchronicity, research progress in spatial causal inference has been relatively limited [15]. There is currently significant controversy regarding whether existing causal inference frameworks, such as the Structural Equation Model (SEM), Structural Causal Model (SCM), causal graphs, Granger causality, and the Convergent Cross Mapping (CCM) model proposed based on dynamic system theory, can effectively overcome the mirror effect and be applicable to spatial data [16]. Meanwhile, although global spatial autocorrelation models and spatially stratified heterogeneity models have advantages in simulating spatial processes, they have not yet been widely applied in the field of spatial causal inference [15]. In addition, some time-series-based causal inference methods have limited capabilities in capturing causal relationships and may overlook some important causal links [17]. In summary, the identification of spatial causal relationships between fishery resources and environmental factors is still at an exploratory stage. Particularly in the context of large-scale, nonlinear, and multivariate coupling, there is a lack of effective analytical frameworks. The GCCM model, as a method for revealing potential causal connections, exhibits greater applicability and advantages in handling large-scale complex spatial systems, thereby effectively avoiding fallacies that may arise from spatial statistical models. Meanwhile, the GOS model also demonstrates superior predictive capabilities and environmental response identification accuracy at local spatial scales and under complex environmental conditions.

Compared with traditional causal analysis models, the GCCM model demonstrates certain advantages in handling nonlinear relationships and variables that are difficult to disentangle. Based on spatial cross-sectional data, this method can, to a certain extent, mitigate the spatial mirroring effect and capture bidirectional, asymmetric potential causal relationships. Particularly, without relying on time-series data, it holds promise for identifying global causal structures that are difficult to observe using some time-based models, thereby detecting weak or moderate causal relationships [18]. However, the identification results of the GCCM model may also be subject to interference from sampling density and spatial noise, leading to reduced model stability. Geographical similarity theory aims to quantify environmental similarity between known and unknown locations, allowing reliable estimation based on a limited set of representative observations. This approach allows for the integrated consideration of multiple environmental factors, particularly in complex scenarios where identifying geographically similar attributes facilitates more accurate inference of resource conditions in unsampled areas [19]. Currently, geographical similarity theory has been widely applied in various fields including pedology [20–22], agronomy [23], and risk assessment [19], while demonstrating potential value in fisheries research. For migratory fish species like skipjack tuna, which are highly sensitive to the environment, geographic similarity methods can effectively predict the spatial distribution of skipjack tuna resources by quantifying the local environmental similarity of fishing grounds.

This study systematically accounts for linear correlations among multiple environmental factors. Targeting the two distinct schools in purse seine skipjack tuna fisheries—associated and unassociated schools—the GCCM model was employed to identify the key environmental factors influencing the catch of these two schools. On this basis, the GOS model, which is based on the principle of geographic similarity, was applied to generate separate spatial distribution predictions for these two schools. The study also incorporated GAM and the BCS model as controls, conducting systematic compar-

isons in terms of goodness-of-fit and predictive stability. This work not only validates the applicability of geographical similarity methods in fisheries science but also offers novel perspectives for sustainable utilization and management of skipjack tuna resources in complex marine ecosystems.

2. Data and Method

2.1. Data Sources

2.1.1. Fishery Data

The fishery data were obtained from the official website of the Western and Central Pacific Fisheries Commission (WCPFC) (<https://www.wcpfc.int/>, accessed on 20 May 2024), consisting of monthly purse seine fishery production data with a spatial resolution of 1° × 1°. The dataset includes information such as year, month, fishing location coordinates (longitude and latitude), fishing days, number of fish caught, and weight. This study selected data spanning the period from 2004 to 2021, and focused on skipjack tuna fishery data within the spatial range of 140° E–165° W and 10° S–10° N.

2.1.2. Environmental Data

This study’s selected marine environmental variables are presented in Table 1, including Mixed Layer Depth (MLD), Sea Level Anomaly (SLA), and Net Primary Production (NPP). Environmental factors at 0–200 m depth include the following: water temperature at 0, 5, 50, 55, 100, 105, and 200 m depth (denoted as SST, T5, T50, T55, T100, T105, and T200, respectively); salinity at 0, 5, 50, 55, 100, 105, and 200m depth (denoted as SSS, S5, S50, S55, S100, S105, and S200, respectively); east–west current velocity at 5, 55, and 105m depth (denoted as U5, U55, and U105, respectively); and north–south current velocity (denoted as V5, V55, and V105, respectively).

Table 1. List of environmental data and their sources.

Parameters	Unit	Spatial Resolution (Latitude × Longitude)	Temporal Resolution	Data Source
SLA	m	0.333° × 0.333°	monthly	http://marine.copernicus.eu/ accessed on 20 May 2024
MLD	m	0.333° × 0.333°	monthly	http://www.science.oregonstate.edu/ accessed on 20 May 2024
CHL	mg/m ³	0.333° × 0.333°	monthly	
NPP	mg/m ² /day	0.333° × 0.333°	monthly	
SST, T50, T100, T150, T200	°C	1° × 1°	monthly	http://www.argo.org.cn/ accessed on 20 May 2024
SSS, S50, S100, S150, S200	PSU	1° × 1°	monthly	
U5, U55, U105	m/s	0.333° × 1°	monthly	https://cfs.ncep.noaa.gov/ accessed on 20 May 2024
V5, V55, V105	m/s	0.333° × 1°	monthly	
T5, T55, T105, T155	°C	1° × 1°	monthly	
S5, S55, S105, S155	PSU	1° × 1°	monthly	

Occupying mainly the peripheries of warm pools, skipjack tuna demonstrate a marked foraging inclination near zones with salinity gradients and undertake seasonal migrations within their physiological tolerance range [24,25]. From the perspective of ocean dynamics, SLA serves as a crucial dynamic indicator, effectively characterizing both water mass movement patterns and nutrient transport processes [26]. MLD modulates key parameters in the mixed layer including temperature and dissolved oxygen concentration, thereby indirectly influencing the habitat suitability of skipjack tuna [27]. Furthermore, NPP provides fundamental data for assessing the productive potential of marine ecosystems, while current velocity governs the trophic resource transport efficiency [7,28,29].

2.2. Models and Methods

The workflow of this study is demonstrated in Figure 1.

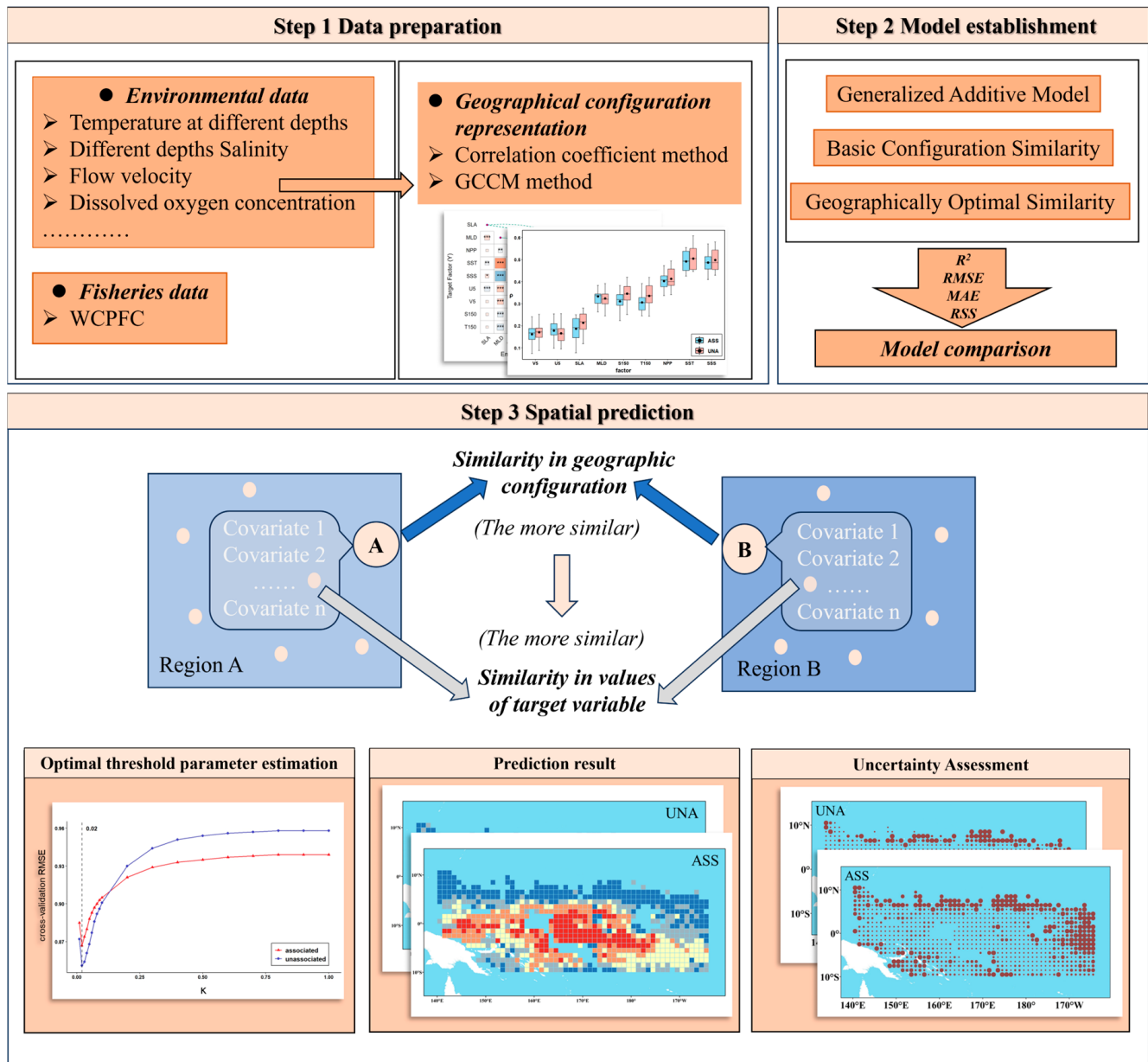


Figure 1. An illustration of the workflow implemented for this study.

2.2.1. Data Preprocessing

Based on the research objectives, the data were processed as follows: (1) data of unassociated schools and associated schools within the study period and area, including year, month, fishing days, and catch information, were extracted; (2) on this basis, records with missing fishing days data were removed, spatiotemporal matching between environmental factors and fishery data was performed, and then the cumulative catch per fishing grid unit was calculated.

2.2.2. Selection of Environmental Factors

When multicollinearity exists among environmental variables, it may cause model overfitting and reduce generalization capability [30]. Therefore, variance inflation factor (VIF) was employed to test collinearity among marine environmental variables, which

examines the independence among factors to enhance the regression results. Variables with a VIF < 10 were retained. In this study, the diagnosis of collinearity among environmental variables was performed using IBM SPSS Statistics 26 software, with detailed results presented in Table 2.

Table 2. Variance inflation factor between environmental variables. (UNA represents unassociated schools; ASS represents associated schools.)

Environment Variables	VIF	
	UNA	ASS
SLA	1.068	1.078
MLD	3.429	3.883
NPP	1.821	1.966
SST	3.593	4.011
SSS	8.742	8.993
U5	2.980	3.047
V5	1.954	2.153
T150	6.623	6.545
S150	8.419	8.317

2.2.3. GAM

Serving as a flexible alternative, the Generalized Additive Model (GAM) broadens generalized linear models through nonparametric techniques [31], employing unspecified smooth functions to replace traditional linear covariate functions. This approach effectively captures nonlinear relationships between environmental factors and catch rates, making it currently one of the most widely used models for assessing marine environment–fisheries relationships [5,32]. GAM was compared with the GOS model to measure the applicability of the GOS model in this study. Its fundamental formula is expressed as follows:

$$g(\mu) = \beta + s_1(X_1) + s_2(X_2) + \cdots + s_i(X_i) + \varepsilon_i \quad (1)$$

In the given formula, $g(\mu)$ is the link function; $s_1, s_2, \dots, s_i$ are smooth functions of environmental variables. $X_1, X_2, \dots, X_i$ represent environmental variables, μ denotes target variable (catch), β is the intercept term, and ε_i represents the error term. GAM was implemented based on R's mgcv Packages (version 4.3.2).

2.2.4. Geographic Similarity and GOS

Zhu et al. proposed a spatial prediction framework where “The more similar geographic configurations of two points (areas), the more similar the values (processes) of the target variable at these two points (areas)” [33]. Based on this theory, the Basic Configuration Similarity (BCS) model employs full-sample information to make weighted predictions for unknown locations. Although it is applicable under small-sample conditions, it is susceptible to interference from redundant information in big data scenarios, which can affect the efficiency and accuracy of inference. To overcome the aforementioned limitations, Song et al. proposed the Geographically Optimal Similarity (GOS) model [20], which improves the accuracy and robustness of spatial inference by selecting observation points with the highest environmental similarity to the target location and extracting high-quality similarity information. The specific steps are as follows:

- (1) Geographical Configuration Representation

The selected environmental factors are utilized as explanatory variables, constructing environmental vectors at observation points to characterize habitat conditions, with the computational formula expressed as follows:

$$e = \{e_i\} \quad (i = 1, 2, \dots, p), \quad (2)$$

where e_i is the value of the explanatory variable X_i at a given location.

(2) Similarity Assessment

The similarity is evaluated through geospatial attributes between unknown observation locations v_β ($\beta = 1, \dots, n$) and known observation locations u_α ($\alpha = 1, \dots, m$). The formula for similarity calculation is given as follows:

$$S(u_\alpha, v_\beta) = P\{E_i(e_i(u_\alpha), e_i(v_\beta))\} \quad (3)$$

The function E_i , associated with the i -th covariate, evaluates the degree of similarity between location u_α and v_β . By comparing the covariate-scale similarities for all variables, the function p quantifies the overall similarity between u_α and v_β . At a certain observation point, e_i ($i = 1, \dots, p$) denotes the value taken by the explanatory variable X_i ($i = 1, \dots, p$). When X_i is continuous, E_i is defined accordingly:

$$E_i(u_\alpha, v_\beta) = \exp\left(-\frac{(e_i(u_\alpha) - e_i(v_\beta))^2}{2(\sigma^2 / \delta(u_\alpha, v_\beta))^2}\right) \quad (4)$$

σ is defined as the pooled standard deviation for the explanatory variable X_i , aggregated over both known and unknown points. Meanwhile, $\sigma(u, v)$ measures the Root Mean Square Error (RMSE) comparing X_i at all unknown locations v_β with its value at u_α , an observed location.

(3) Estimation of Optimal Similarity Threshold

By determining the optimal similarity threshold, observed locations with the highest similarity to unknown locations are selected. A cross-validation-based prediction error evaluation approach was utilized to identify it. Training and test sets were generated in equal proportions, and similarity measures were computed and evaluated for each pair of training–test observations. This procedure was repeated for 10 iterations to ensure result accuracy and stability. By using a series of optional percentage thresholds, candidate percentage thresholds in $(0, 1]$ were systematically evaluated to identify high-similarity reference locations. Ultimately, by calculating the average RMSE value over multiple experiments, the final threshold was selected by minimizing the mean RMSE across iterations.

(4) Spatial prediction and uncertainty analysis

The dataset was partitioned into training and test sets at a 7:3 ratio [34]. The optimal similarity threshold derived from training data was applied to extract high-similarity subsets from the training set. Spatial prediction was then performed using Equation (5) [20], which involves computing similarity metrics between unknown target locations and known reference locations, and estimating values at unknown locations by weighted aggregation of observations from known locations based on optimal similarity. The mathematical formulation is given by the following:

$$\hat{Z}(v_\beta) = \frac{\sum_{\alpha=1}^m S_\lambda(u_\alpha, v_\beta) Z_\lambda(u_\alpha)}{\sum_{\alpha=1}^m S_\lambda(u_\alpha, v_\beta)} \quad (5)$$

$\hat{Z}(v_\beta)$ denotes the estimated value at the unknown site v_β , while $Z_\lambda(u_\alpha)$ refers to the observed value at location u_α , which exhibits the highest similarity.

The GOS model reveals an inverse relationship between prediction uncertainty and the degree of similarity between unknown and observed locations [20]. To evaluate the extent and influence of model uncertainty, an assessment is performed based on the prediction outputs. The corresponding calculation is given below:

$$\Delta(v_\beta) = 1 - Q(S_\lambda(u, v_\beta), \zeta) \quad (6)$$

where $\Delta(v_\beta)$ is the uncertainty at the unknown location v_β , Q is the quantile operator, and ζ is the confidence interval used to determine the similarity value that critically affects the prediction results. The values of ζ used in this study were 0.9, 0.95, 0.99, 0.995, 0.999, and 1 for assessing the reliability to ensure the stability of the model predictions.

Both the BCS model and the GOS model were implemented through the R's geosimilarity Packages (version 4.3.2).

2.2.5. Geographical Convergent Cross Mapping

As linear correlation and similar traditional statistical methods fall short in identifying intricate nonlinear relationships, this research employed the Geographical Convergent Cross Mapping (GCCM) model as an exploratory tool to uncover potential causal association structures among spatial variables. which identifies causal relationships between variables through state space reconstruction and Convergent Cross Mapping prediction. It is primarily used to determine causal relationships between two sets of cross-sectional spatial data. The model works by classifying and combining the observed values of different variables to construct a spatial dynamic system. It then further examines whether these different variables originate from the same spatial dynamic system, thereby identifying the potential causal relationships among spatial variables [18].

GCCM replaces the lags in traditional convergent cross-mapping (CCM) models, which rely on time-series observations, with spatial lags (measurements of a specific unit and its neighborhood) [35]. For two spatial variables X and Y defined on the same set of spatial units—organized as regular grids (raster data) or irregular polygons (vector data)—their values and spatial lag terms can be regarded as observational functions of the values read from each spatial unit. Following the generalized embedding theorem [36–38], the phase space is reconstructed using observational functions, transforming spatial variables X and Y into shadow manifolds M_x and M_y , to perform spatial prediction on the mapping relationship between shadow manifolds and infer causality.

Thus, for a given x , the value of Y can be predicted based on the nearest neighbors identified from M_x , and this nearest neighbor-based prediction approach is defined as cross-mapping prediction. The formula is as follows:

$$\hat{Y}_s|M_x = \sum_{i=1}^{L+1} (w_{si}Y_{si}|M_x) \quad (7)$$

where s represents the spatial unit of the Y value to be predicted; $\hat{Y}_s$ is the prediction result; L is the number of dimensions of the embedding; si is the spatial unit used in the prediction; Y_{si} is the observation at si and at the same time the first component in M_y , denoted as $\psi(y, si)$, and furthermore, $\psi(y, si)$ is determined by its one-to-one mapping point $\Psi(x, si)$, which is one of the $L + 1$ nearest neighbors of the focal state in M_x one of the $L+1$ nearest neighbors of $\psi(x, s)$; and w_{si} is the defined correlation weight.

The performance of cross-mapping prediction is evaluated using Pearson's correlation coefficient calculated between the actual observed values and their predicted counterparts. The formula is given below:

$$\rho = \frac{Cov(Y, \hat{Y})}{\sqrt{Var(Y)Var(\hat{Y})}} \quad (8)$$

where $Cov()$ represents the covariance of the two variables and $Var()$ represents the variance.

Sugihara et al. [16] proposed using the convergence of the correlation coefficient ρ to infer causal relationships. For the GCCM model, convergence implies that ρ increases with the size of the embedding dimension L and achieves statistical significance when L is at its maximum value [16,39]. The GCCM model can automatically adjust the L parameter to find the maximum ρ . In this study, the significance level was set at 0.05, with a 95% confidence interval. The GCCM model used in this research was implemented through the GCCM function package in R (version 4.3.2).

2.2.6. Evaluation Metrics

This study selected Residual Sum of Squares (RSS), R-Squared (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) as the criteria for evaluating model performance. Among these, RSS, RMSE, and MAE values closer to 0 and R^2 values closer to 1 indicate better model fitting performance.

The corresponding formulas for the evaluation metrics are as follows:

$$RSS = \sum_{j=1}^n (\hat{Z}_j - Z_j)^2 \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (\hat{Z}_j - Z_j)^2} \quad (10)$$

$$MAE = \frac{\sum_{j=1}^n |\hat{Z}_j - Z_j|}{n} \quad (11)$$

$$R^2 = \frac{\sum_{j=1}^n (Z_j - \bar{Z})^2}{\sum_{j=1}^n (\hat{Z}_j - \bar{\hat{Z}})^2} \quad (12)$$

where $\hat{Z}_j$ and Z_j are the predicted and observed values for the j test position, respectively. n is the number of samples.

3. Results

3.1. Environmental Factor Correlation

In this study, the Pearson's correlation coefficient method was employed to check for data redundancy among environmental factors that were filtered using the VIF approach. Figure 2 and Table 3 present the correlation matrix results obtained using the Pearson's correlation coefficient method, covering correlations among environmental factors and between environmental factors and the catch of the two schools. The results show that all correlation coefficients between environmental factors and catch have absolute values less than 0.5, indicating generally weak to moderate correlations. Meanwhile, the correlations between most environmental variables are also relatively weak, indicating low redundancy among the variables.

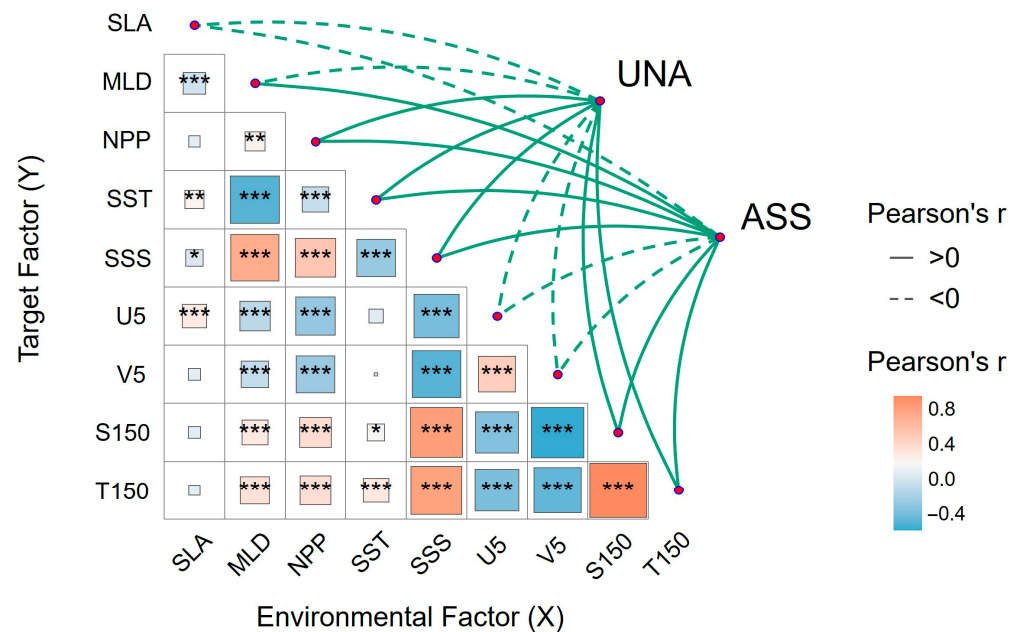


Figure 2. Environmental factor correlation diagram. (* represents 90% confidence interval; ** represents 95% confidence interval; *** represents 99% confidence interval. Font size represents degree of relevance). (UNA represents unassociated schools; ASS represents associated schools).

Table 3. Correlation of environmental variables with catch of different fish schools. (UNA represents unassociated schools; ASS represents associated schools). (* represents 90% confidence interval).

Environment Variables	Correlation	
	UNA	ASS
SLA	−0.16 *	−0.17 *
MLD	−0.07	0.11 *
NPP	0.35 *	0.30 *
SST	0.29 *	0.15 *
SSS	0.11 *	0.27 *
U5	−0.42 *	−0.48 *
V5	−0.12 *	−0.23 *
S150	−0.14 *	0.23 *
T150	0.30 *	0.34 *

Figure 3 presents the results of the analysis based on the GCCM model. The findings are as follows:

- (1) Overall, the two types of fish populations exhibit a high degree of consistency in their response patterns to environmental factors, with no significant differences observed in the types of major influencing factors and their causal strengths.
- (2) Among all environmental factors, SST, SSS, and NPP demonstrate strong causal relationships with catch. Their ρ values are significantly higher than those of other factors, with the highest values exceeding 0.5. This indicates that these three environmental factors are crucial in influencing the catch of both fish populations.
- (3) Additionally, the ρ values for S150 (salinity at 150 m depth) and T150 (temperature at 150 m depth) are also relatively high, reaching above 0.4 in some cases, showing a certain degree of causal association. The ρ values for MLD and SLA are relatively low but still reflect indirect effects on nutrient redistribution. Finally, U5 and V5 exhibited the lowest ρ values, demonstrating weak influence of surface currents on catch.

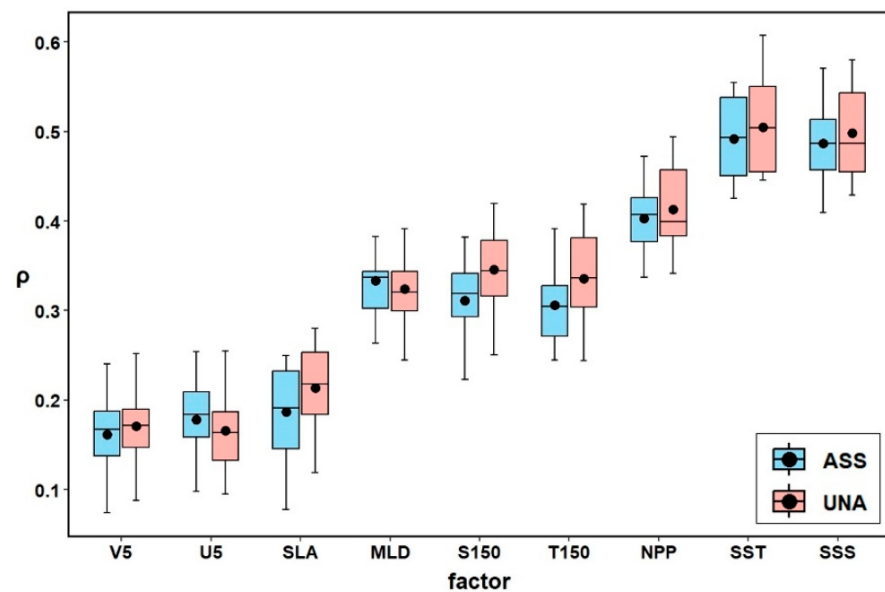


Figure 3. GCCM outputs for catch and environmental variables. (UNA represents unassociated schools; ASS represents associated schools).

Considering the causal relationship strengths and correlation coefficients between the variables and catch, all the aforementioned variables were selected as inputs for subsequent analysis.

3.2. Optimal Similarity of Geographical Configurations

The optimal percentage threshold for observations with maximum similarity to unknown locations is determined based on this configuration, with results cross-validated against RMSE for optimizing prediction performance and avoiding overfitting. As shown in Figure 4, these represent the optimal similarity results for both schools.

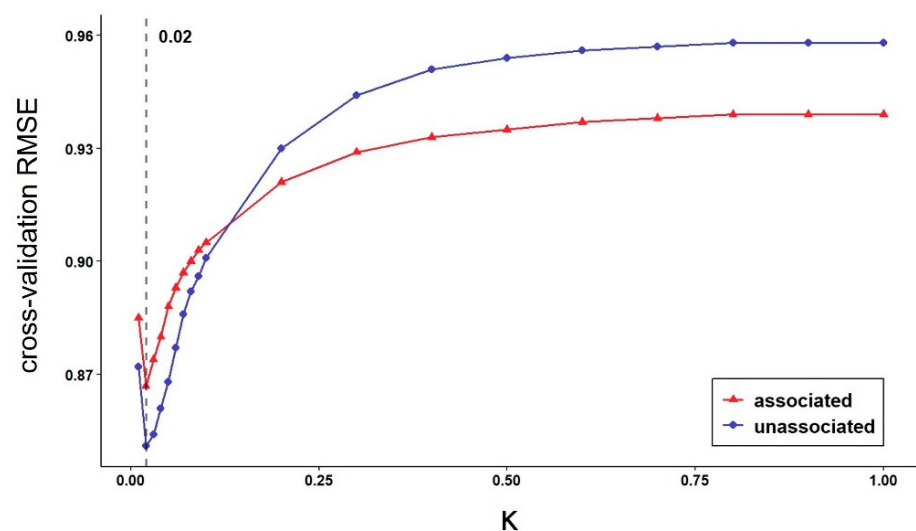


Figure 4. Optimal similarity thresholds for the geographic allocation of the two fish schools.

Figure 4 presents the optimal similarity threshold analysis results for both schools. Overall, as the percentage Kappa value of observational data used for prediction increases, the cross-validated RMSE values first decrease and then increase. This indicates that insufficient observational data leads to inadequate representativeness in similarity computation, resulting in unreliable predictions. Conversely, using more observational data for similarity computation and prediction also increases prediction error, as low-similarity observations

introduce additional errors, noise, and other confounding factors affecting catch predictions, thereby compromising accuracy. The results show that when RMSE reaches its minimum, the optimal percentage threshold is 0.02 for both unassociated and associated schools, indicating that for prediction purposes, utilizing just 2% of the known observational data to calculate similarity can achieve the optimal predictive performance. The selection of this percentage is based on a systematic cross-validation process, which effectively mitigates the risk of overfitting by minimizing prediction errors. We filtered sample points using the optimal similarity parameter and fitted them into the GOS model for catch prediction and fitting.

3.3. Model Performance Evaluation and Spatial Prediction Results

3.3.1. Model Performance Evaluation

This study used the selected environmental factors as explanatory variables, with catch of unassociated and associated skipjack schools as response variables, respectively, to construct three models: GAM, BCS, and GOS. GAM is a widely adopted statistical model in fisheries resource research, suitable for depicting nonlinear relationships between explanatory and response variables, and it offers good interpretability. However, it neglects spatial location and geographic proximity, making it difficult to effectively capture the differences in ecological processes brought about by spatial heterogeneity. Both the BCS model and the GOS model are based on the principle of geographic similarity. The BCS model uses all observed samples during prediction, which can easily be disturbed by low-similarity samples, especially when the spatial structure is complex or the sample distribution is uneven, leading to potential errors. In contrast, the GOS model, an improvement upon the BCS model, selects observations that are most geographically similar to the target location for prediction. Comparative analysis of model performance parameters (Table 4) leads to the following conclusions:

Table 4. Result of model performance evaluation.

Schools	Method	R ²	RMSE	MAE	RSS
Associated	GAM	0.572	0.893	0.658	2805.82
	BCS	0.613	0.825	0.602	2012.67
	GOS	0.649	0.793	0.554	1694.45
Unassociated	GAM	0.569	0.884	0.696	2607.15
	BCS	0.609	0.880	0.655	1908.84
	GOS	0.656	0.829	0.572	1558.43

In terms of model R-Squared (R²), for both unassociated and associated schools, both the GOS and BCS models exhibited significantly higher R² values than GAM. This indicates that incorporating geographic similarity methods better captures the spatial distribution patterns of fish schools, thereby significantly enhancing model predictive performance. In contrast, although GAM can handle nonlinear relationships, its performance in predicting complex spatial distributions remains relatively limited.

By further comparing the performance parameters of the BCS model and the GOS model, it can be observed that the GOS model outperforms the BCS model across four indicators: R-Squared (R²), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Residual Sum of Squares (RSS). Specifically, the RSS value of the GOS model is significantly lower, indicating smaller overall prediction errors and a more concentrated residual distribution. This suggests that by seeking optimal similarity among observational data, the GOS model can more accurately reflect the complex coupling relationships between fish distribution and environmental factors.

3.3.2. Prediction Results

Figures 5 and 6 present the spatial distribution predictions of catch for unassociated and associated schools in the Western and Central Pacific, generated by the GAM, BCS, and GOS models. The figures reveal significant differences in the predicted spatial ranges among the models. Using the 160° E “pocket” high seas closed areas as a boundary, the study area was divided into eastern and western regions.

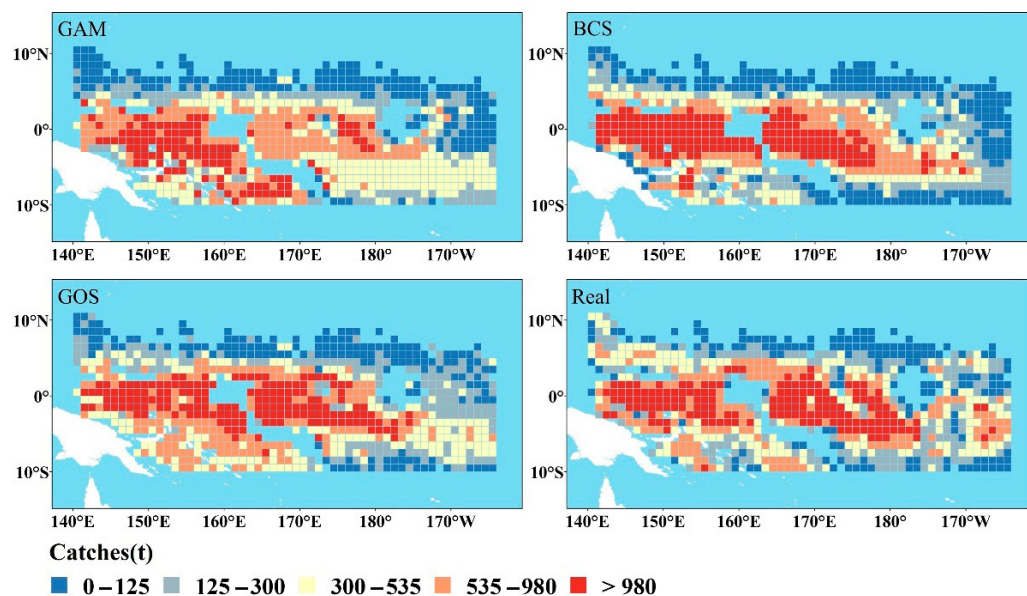


Figure 5. Distribution of catch predictions from different models for unassociated schools.

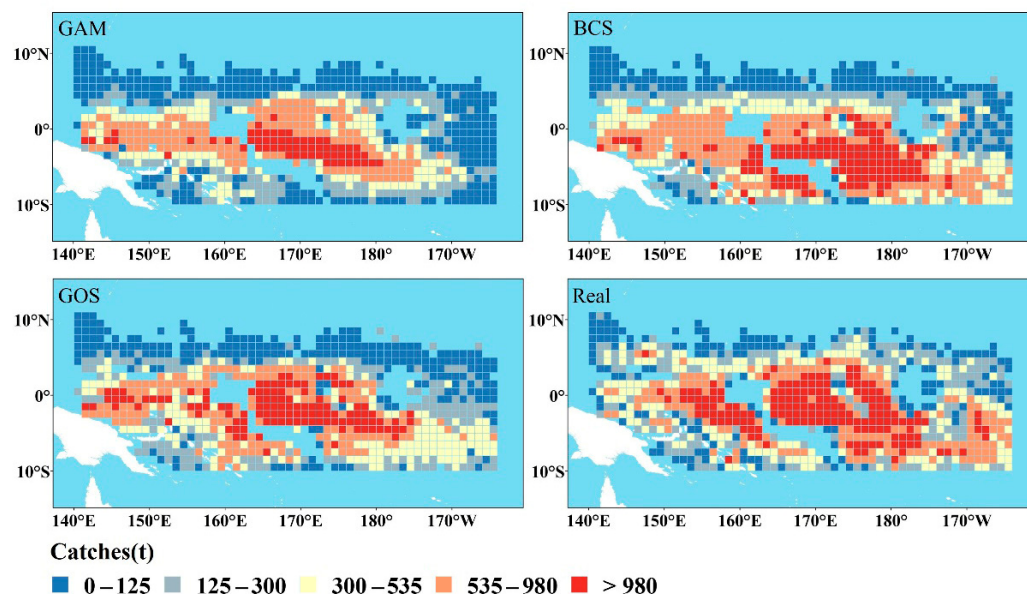


Figure 6. Distribution of catch predictions from different models for associated schools.

In Figure 5, GAM predicted that the high-catch areas of unassociated schools were mainly located in the western region (145° E–160° E, 10° S–2° N), with some high-catch areas also in the eastern region (175° E–180° W, near 0°). By contrast, the BCS model suggested a more focused distribution, primarily around the equator in the west (140° E–160° E) and in eastern waters (165° E–178° E, 7° S–2° N), covering a broader range than GAM. The GOS model’s predictions showed more concentrated high-catch areas in western waters (142° E–158° E) compared to BCS, while also capturing extensive eastern

activity reaching close to 180° . By comparing with the actual catch of unassociated schools, it can be observed that there are certain discrepancies between the predicted results of GAM and the BCS model and the distribution of the actual catch. In contrast, the predicted results of the GOS model demonstrate higher accuracy and can better reflect the actual distribution of catch.

In Figure 6, for associated schools, GAM estimated major catch zones predominantly in the eastern region (165°E – 180°W , 7°S – 0°), while fewer areas were observed in western waters. The BCS model indicated a similar eastern concentration (162°E – 175°W , 8°S – 0°). The GOS model predicted high-catch areas distributed between 165°E – 180°W and 5°S – 2°N , showing a clear northwestward shift compared to BCS predictions, and also detected additional zones in the western region (155°E – 162°E , 8°S – 5°S). Comparison with actual associated school catch data revealed certain discrepancies between the predicted and actual catch distributions for both the GAM and BCS models, while the GOS model demonstrated higher accuracy in representing actual catch distribution patterns.

In summary, the GOS model demonstrated superior accuracy in predicting the spatial distribution of catch for two schools, particularly in the eastern waters, where its predicted high-catch zones aligned more closely with actual catch distribution. In contrast, while the GAM and BCS models were able to provide predictive results to some extent, they exhibited lower precision and spatial accuracy compared to the GOS model.

3.4. Uncertainty Analysis of Prediction Results

Figure 7 displays the uncertainty distribution of GOS model predictions at a confidence level of $\zeta = 0.99$. When varying confidence levels ($\zeta = 0.9, 0.95, 0.99, 0.995, 0.999, 1$) are selected, the spatial patterns of uncertainty remain relatively consistent, although the magnitude varies with the confidence level. Overall, as ζ decreases, uncertainty values progressively increase; at $\zeta = 0.9$, uncertainty peaks, with more pronounced high-uncertainty zones and greater spatial variability; however, for $\zeta \geq 0.99$, uncertainty stabilizes and exhibits more homogeneous spatial distribution. As shown in the figure, the high-uncertainty zones of unassociated and associated schools show certain spatial distribution similarities, primarily concentrated near the study area's margins and the high seas closed areas. Comparatively, unassociated schools exhibit smaller high-uncertainty zones, whereas associated schools demonstrate more extensive high-uncertainty areas. Using the 160°E high seas closed areas as a boundary, the study area is divided into eastern and western sectors.

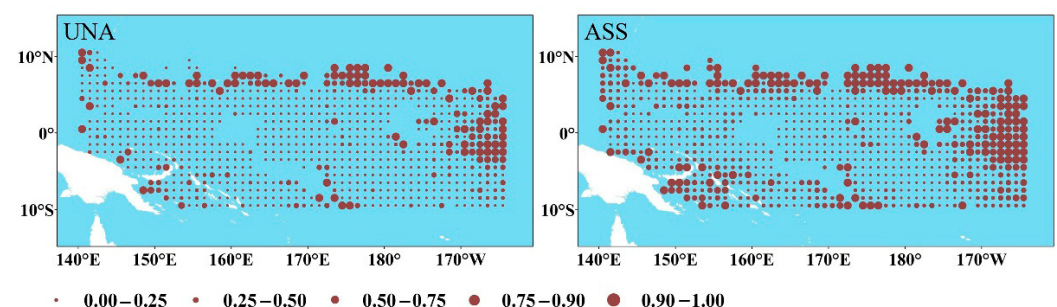


Figure 7. Uncertainty analysis of geographical-optimal similarity prediction results ($\zeta = 0.99$). (UNA represents unassociated schools; ASS represents associated schools.)

In the uncertainty distribution of unassociated schools, high-value zones are primarily concentrated in the eastern region (165°W – 170°W , 5°S – 5°N), where uncertainty is significantly higher than other areas. Additionally, areas north of 7°N generally exhibit higher uncertainty. In contrast, high-uncertainty zones in western waters are more dispersed, mainly scattered near coastal study areas and adjacent to the “pocket” high seas closed areas.

In comparison, the high-uncertainty zones of associated schools cover a larger area in the eastern waters, primarily distributed between 165° W and 172° W. In the western waters, these zones are concentrated within 150° E–158° E and 10° S–5° S. Compared with unassociated schools, the associated schools exhibit more extensive high-uncertainty zones across both eastern and western waters.

The proportions of prediction uncertainty values are shown in Table 5. Overall, the GOS model's prediction uncertainty values within the high seas closed areas were higher for associated schools than for unassociated schools. The distributions of unassociated and associated schools across uncertainty intervals differed significantly. Predictions for unassociated schools were predominantly in the low-uncertainty interval (0.00–0.25), accounting for 72.65%, whereas associated schools accounted for only 52.65% in this interval. As uncertainty increased, the proportion of associated schools gradually surpassed that of unassociated schools, particularly in the 0.25–0.50, 0.50–0.75, and 0.90–1.00 intervals, with associated schools comprising 16.75%, 8.92%, and 16.39%, respectively, all significantly higher than the corresponding proportions for unassociated schools (9.76%, 5.06%, and 9.76%). This indicates that predictions for associated schools generally exhibit higher uncertainty, whereas predictions for unassociated schools are more stable and reliable.

Table 5. Proportion of prediction uncertainty values. (UNA represents unassociated schools; ASS represents associated schools.)

Uncertainty	UNA/%	ASS/%
0.00–0.25	72.65	52.65
0.25–0.50	9.76	16.75
0.50–0.75	5.06	8.92
0.75–0.90	2.77	5.30
0.90–1.00	9.76	16.39

4. Discussion and Conclusions

4.1. Environmental Factors Affecting Purse Seine Skipjack Tuna Catch Distribution

Skipjack tuna in purse seine fisheries are schooling fish [40], whose aggregation patterns are strongly influenced by marine environmental conditions [41,42]. However, unassociated and associated schools exhibit distinct behavioral mechanisms due to differing aggregation characteristics. Currently, Fishing Aggregating Devices (FADs) have become the predominant method for catching tuna-associated schools in purse seine fisheries [43]. Furthermore, studies have shown that different school types demonstrate variations in environmental preferences and habitat selection [44]. This study analyzed the potential causal relationships and their strengths between the catch of two types of purse seine skipjack tuna populations in the Western Central Pacific Ocean and marine environmental factors using the GCCM model. The results (in Figure 3) indicate that surface environmental factors, namely SST, SSS, and NPP, exhibit strong causal strengths ($\rho > 0.5$) for both fish schools. In contrast, factors such as U5, V5, and SLA demonstrate weaker causal strengths ($\rho < 0.3$) with catch.

Combining these findings with correlation coefficient analyses, it was observed that the correlation coefficients for the three surface factors (SST, SSS, and NPP) were not particularly high (with a maximum value of 0.35), indicating only weak to moderate correlations. Conversely, U5 exhibited the highest correlation coefficient in the analysis. As indicated in Table 3, the correlation coefficient between the catch of unassociated schools and MLD failed to pass the 90% confidence interval, suggesting that at the given confidence level, the data do not support a significant relationship. This implies that the relationship between the independent and dependent variables may not be sufficiently strong, or that

the sample data are insufficient to prove such a relationship. The GCCM model is capable of comprehensively quantifying the influence intensity of various environmental factors on catch at a 95% confidence level, thereby enhancing the authenticity and credibility of the research findings.

These findings underscore the limitations of using correlation coefficients to explore the relationships between catch and environmental factors. They also demonstrate that correlation coefficients cannot reflect causality between variables and should not be used to infer causal relationships.

This study indicates that SST is likely a crucial environmental variable influencing catch, a finding that is highly consistent with existing research. As typical warm-water migratory fish, skipjack tuna are highly sensitive to changes in water temperature, with their feeding, migration, and schooling behaviors all being regulated by SST [5,45]. Variations in SST not only affect the suitability of skipjack tuna habitats but also drive adjustments in their spatial distribution patterns [46]. In the Western Central Pacific Ocean region, the spatial center of gravity of skipjack tuna catch often shifts significantly with the movement of the 29 °C isotherm along the edge of the tropical warm pool, reflecting their high responsiveness to changes in sea surface temperature [3].

Compared with SST, SSS also showed significant influence, with its ρ values exceeding 0.5, nearly equivalent to SST. Studies have shown that salinity is not only an important environmental factor representing nutrient transport and distribution in the ocean [47,48], but also an effective tracer for identifying ocean circulation pathways [49]. In addition, SSS affects MLD by regulating seawater density, playing a key role in regulating vertical heat and nutrient transport [50]. Some previous studies suggested that SSS is not sensitive enough in short-term dynamic changes and therefore is not one of the main environmental drivers. This study, based on an analysis of long-term averaged environmental data, reveals the response characteristics of skipjack tuna catch to SSS over long time scales.

In this study, the influence of NPP ranks third among the environmental factors considered. As the fundamental source of nutrients for skipjack tuna, NPP typically affects the distribution and reproduction of phytoplankton and zooplankton [51], and through the trophic cascade effect between phytoplankton and prey, indirectly affects the abundance level of skipjack tuna resources. Studies have shown that high NPP levels help improve the spawning and reproductive success of skipjack tuna, having a positive effect on population maintenance [52].

Meanwhile, we found that the temperature, salinity in the subsurface water, and MLD significantly influence the vertical distribution patterns of skipjack tuna, indicating their strong responsiveness to the thermohaline structure of the water column during habitat selection and foraging. In contrast, SLA and ocean current factors U5 and V5 exhibited relatively weaker roles in the causal contribution analysis. We posit that this may be because these dynamic factors primarily influence skipjack tuna distribution indirectly through other environmental variables. For instance, SLA is often associated with ocean eddies and circulation systems, which can create frontal zones at their edges, enhancing nutrient levels and primary productivity, thereby providing potential foraging environments for skipjack tuna [53]. Similarly, changes in ocean currents can indirectly affect the aggregation behavior of skipjack tuna on larger spatial scales by regulating key environmental conditions such as temperature, salinity, and dissolved oxygen, which in turn reshape the distribution patterns of plankton and prey resources [54]. Furthermore, there may be strong intercorrelations among multiple environmental variables, leading to relatively weaker direct effects observed in causal analyses.

4.2. Application of Geographic Similarity in Fishery Prediction

Since the 1990s, associated schools have been the primary target of purse seine fisheries. However, while the high-intensity deployment of FADs has improved fishing efficiency, it has also, to some extent, altered the physical environment of the sea surface and exacerbated the catch issue of juvenile tuna species. This has caused relatively severe negative impacts on the health of target fish populations and the pelagic marine ecosystem. In view of this, the WCPFC has gradually guided the purse seine fishery towards an operation mode focused on unassociated schools, aiming to reduce ecological risks and enhance the sustainability of fisheries resources.

To evaluate the predictive performance of the GOS model, this study conducted a comparative analysis between the GOS model and two other methods (GAM, BCS). From the perspective of the coefficient of determination (R^2), the GOS model outperformed both the GAM and BCS models for both unassociated and associated schools, demonstrating its good capability in capturing spatial structural features. Furthermore, when comparing indicators such as RSS, RMSE, and MAE, the GOS model also exhibited certain advantages in error control, reflecting its comprehensive potential in terms of accuracy and robustness. GAM, as a mature nonlinear statistical modeling method, has certain advantages in handling complex relationships. However, under data conditions with uneven sample distribution, strong noise, or significant spatial heterogeneity, its predictive accuracy may fluctuate. The BCS model, although based on the similarity principle, is highly dependent on full-sample characteristics, making it susceptible to noise interference and resulting in poor prediction stability and low computational efficiency. The advantage of the GOS model lies in its use of the cross-validation method to automatically determine the optimal similarity threshold, enabling the model to select a relatively small but representative subset from the observed data for fitting, thereby reducing reliance on full-sample modeling. This strategy helps to reduce the risks of noise interference and overfitting and, to a certain extent, improves the model's predictive error performance, especially showing good results in RMSE control [19]. In addition, the GOS model has strong interpretability. Its process of inferring fishing ground distribution based on environmental feature similarity helps to reveal the potential physical associations between environmental factors and fishing ground distribution, to some extent making up for the deficiencies in the interpretation mechanisms of traditional statistical methods or some machine learning models.

Therefore, through the evaluation of the GOS model's predictive accuracy and the comparison results with the GAM and BCS models, this study to a certain extent demonstrates the potential application value of the geographic similarity method in predicting the spatial distribution of skipjack tuna fishing grounds. Meanwhile, this method exhibits certain interpretability in characterizing the spatial relationships between environmental factors and fishing ground distribution, providing new ideas for predicting fishing ground distribution.

4.3. Analysis of Prediction Uncertainty in GOS Model for Different Fish Schools

The uncertainty values predicted by the GOS model reflect the similarity between the observed and predicted locations [22] and serve as indicators of prediction reliability. This uncertainty can be classified as estimation uncertainty, primarily stemming from prediction biases caused by environmental differences during the spatial inference process of the model, which relies on geographic similarity weights. As clearly shown in Figure 7, areas with high uncertainty values exist for both fish schools, revealing generally similar spatial patterns of uncertainty distribution between the two fish schools, though the high-value uncertainty zones are more extensive for associated schools than for unassociated schools. By quantifying the proportion of uncertainty values across different categories, it is

evident that the GOS model provides higher reliability for unassociated school predictions than for associated schools, particularly in the low-uncertainty range (0.00–0.25), where performance significantly surpasses that for associated schools. In recent years, the catch of associated schools has predominantly originated from FADs, which may be a key factor contributing to their higher prediction uncertainty.

Figures 3 and 6 further demonstrate that unassociated schools exhibit stronger responses to environmental factors than associated schools, indicating that unassociated schools rely more on natural environmental variations for habitat selection. In contrast, the deployment of FADs partially disrupts the environmental response capacity of associated schools. In terms of biological characteristics, differences in the type of purse seine used can lead to variations in the body length composition of catch [55]. Unassociated schools tend to have larger body lengths and higher ages, and they rely more on natural environmental changes to make habitat choices. In contrast, associated schools have smaller body lengths and lower ages, and they are less sensitive to environmental changes [55–57]. This difference means that the spatial distribution of associated schools is more likely to be driven by their schooling behavior, which weakens the direct impact of environmental factors on their distribution. Consequently, this affects the accuracy and stability of model predictions [58].

Overall, Figure 7 shows that some areas consistently exhibit high uncertainty values (0.75–1.00), with a portion of these high-uncertainty zones located in coastal waters. These coastal areas typically fall within the Exclusive Economic Zones (EEZs) of coastal states. Under WCPFC regulations, EEZs have specific management regimes and fisheries policies, where states hold sovereign rights over resource exploitation and utilization. Such institutional arrangements may lead to intensive human activities in these areas, particularly concentrated fishing efforts, resulting in disturbances to local ecosystems. WCPFC data also indicate [59] that these high-uncertainty zones show significant spatial overlap with areas of high fishing effort. In addition, within these regions, the proportion of fishing using FADs is relatively high. Such intensive deployment of FADs may lead to abnormal fluctuations in environmental variables, reducing environmental stability. Consequently, this undermines the performance of prediction models based on geographic similarity, manifesting as an increase in uncertainty [60–62].

Excluding areas close to land, Figure 7 shows that high uncertainty is generally prevalent at the edges of the study area. This may be closely related to the spatiotemporal variations in the “warm pool-cold tongue” system triggered by El Niño–Southern Oscillation (ENSO) events. This system exhibits significant spatial migration in the equatorial Pacific Ocean, which can easily cause drastic fluctuations in regional environmental conditions. As a result, it weakens the similarity between observation points and prediction points, thereby increasing the uncertainty in model predictions [3]. Research indicates that the “warm pool-cold tongue” system can shift up to 8000 km in the east–west direction between El Niño and La Niña phases, leading to significant shifts in the isotherms corresponding to the optimal temperature for skipjack tuna (approximately 29 °C) [50,63]. The spatial variations in these isotherms are highly consistent with the high-uncertainty values observed at the edges in this study. As the strongest air–sea interaction phenomenon globally, ENSO not only reshapes the hydrological structure of the tropical Pacific Ocean but also causes noticeable displacements in the center of gravity of skipjack tuna fishing grounds [3,64,65]. This further increases the challenges for geographic similarity-based predictions in these edge regions. In particular, during different ENSO phases, the distribution patterns of environmental variables change significantly, reducing the model’s adaptability and prediction stability in the edge areas [66].

Additionally, from a regional management perspective, high-uncertainty zones frequently occur in transitional areas between multiple EEZs or the high seas, falling under the collaborative management mandates of multilateral fisheries organizations (e.g., WCPFC). Fisheries management regimes in these areas are relatively complex, with notable disparities among nations in fishing effort, FAD regulations, data reporting, and monitoring capacity, often leading to institutional fragmentation and data inconsistencies [67–70]. Existing studies have demonstrated that institutional fragmentation undermines the coordination and enforcement capacity of Regional Fisheries Management Organizations (RFMOs) in trans-boundary resource governance, particularly in regions like the Pacific where jurisdictional boundaries are ambiguous and maritime overlaps frequently occur [67,68]. Meanwhile, disparities in national capacities for regulatory enforcement and data monitoring systems further exacerbate this challenge [69,70]. Due to their inherent complexity and limitations in data acquisition, the aforementioned management systems, FAD (Fish Aggregating Device) deployment behaviors, and national boundaries were not directly incorporated as modeling variables in this study. Such management inconsistencies may introduce environmental data discrepancies between observed and predicted locations, thereby compromising the model's accuracy in assessing environmental similarity, ultimately manifesting as elevated uncertainty in peripheral zones.

In addition to the estimation uncertainty mentioned earlier, there is also a certain degree of data uncertainty in this study. The fisheries data employed in the research are presented as $1^\circ \times 1^\circ$ grid data. However, the original sampling across different years exhibits significant spatial coverage disparities. There are instances of missing data or incomplete grids in various years, resulting in an unbalanced spatial structure of the samples and insufficient sampling density. Skipjack tuna is a highly migratory fish species, and its distribution is highly dynamic, being significantly influenced by factors such as ocean currents and temperature. This leads to a strong spatial randomness in fishing activities themselves. Its catch is easily affected by a combination of human factors (e.g., fishing routes, fishing strategies) and environmental factors in spatial terms, which may also introduce additional noise. These factors collectively contribute to data-level uncertainty, potentially affecting the reliability of model inputs and the robustness of prediction results to a certain extent.

5. Prospects

With regard to the limitations of this study and potential future research directions, the following aspects can be further expanded and improved:

For the GCCM model, while it can be used as a tool for detecting potential directional causal associations, it is not yet suitable as a standard framework for estimating causal effects because its ρ value is more akin to a dependence indicator. Interpretation of GCCM results still depends on the validity of several key assumptions, such as the existence of embedded mappings between variables and the plausibility of spatial neighborhood definitions [71,72]. However, in highly complex marine ecosystems, these prerequisites are often difficult to fully satisfy. In particular, the behavior of migratory species may introduce a large number of unobserved variables and potential confounders, posing challenges to the stability and interpretability of causal identification [16,73]. Therefore, future studies should combine sensitivity analyses, confounding variable detection methods, and other approaches to systematically assess the robustness of model inferences. In addition, more standardized causal inference frameworks such as directed acyclic graphs (DAGs) should be integrated to clarify causal relationships among variables.

The geographical similarity method: (1) Since this study conducted annual averaging of environmental and fisheries data, the geographic similarity method consequently

failed to account for the temporal variations in the marine environment and skipjack tuna resources. Subsequent research should build upon this and incorporate time-series effects to elucidate the spatiotemporal fluctuations in the responses of Pacific purse seine skipjack tuna to changes in dominant environmental factors. (2) For the setting of similarity thresholds in the geographic similarity method, further sensitivity analyses can be conducted in subsequent research to systematically evaluate the impact of different threshold settings on the model's prediction results. Although the thresholds in the current study have, to a certain extent, embodied the similarity screening mechanism, their robustness has not yet undergone rigorous testing. (3) In the future, it is also worth exploring the integration of the geographic similarity method with traditional spatial models (such as Geographically Weighted Regression (GWR) and Kriging) to further enhance its explanatory power and inferential value. For instance, attempts can be made to construct a regression Kriging method based on geographic similarity weights, or to introduce the similarity matrix into the Geographically Weighted Regression (GWR) model to replace the traditional geographic distance kernel function. This would enable local modeling of environmental configuration similarity. Such hybrid methods are expected to take into account both geographic similarity structures and spatial dependence characteristics, providing more robust and flexible tools for spatial prediction of resource distribution and mechanism identification. (4) Regarding the characterization of uncertainties, this study primarily focused on the distribution differences in spatial prediction results and has not systematically encompassed uncertainties at the levels of model parameters and structural assumptions. Therefore, in the future, it is necessary to further expand the identification and quantification of uncertainties from different sources on the basis of the existing spatial prediction uncertainties.

Author Contributions: Conceptualization, S.F. and X.Y.; methodology, M.L.; software, Z.H.; validation, S.F., X.Y., S.T. and J.Z.; formal analysis, X.Y.; investigation, S.F.; resources, J.Z.; data curation, Z.H.; writing—original draft preparation, S.F.; writing—review and editing, S.F.; visualization, S.F.; supervision, X.Y.; project administration, X.Y.; funding acquisition, S.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Programs of China (2024YFD2400603).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Linnaeus, C. *Systema Naturae per Regna Tria Naturae: Secundum Classes, Ordines, Genera, Species, Cum Characteribus, Differentiis, Synonymis, Locis*, 10th ed.; Laurentius Salvius: Stockholm, Sweden, 1758.
2. Nauen, C.E.; Collette, B.B. *Scombrids of the World: An Annotated and Illustrated Catalogue of Tunas, Mackerels, Bonitos, and Related Species Known to Date*; FAO Fisheries Synopsis No. 125; FAO: Rome, Italy, 1983.
3. Lehodey, P.; Bertignac, M.; Hampton, J.; Lewis, A.; Picaut, J. El Niño Southern Oscillation and Tuna in the Western Pacific. *Nature* **1997**, *389*, 715–718. [[CrossRef](#)]
4. Yen, K.-W.; Su, N.-J.; Teemari, T.; Lee, M.-A.; Lu, H.-J. Predicting the catch potential of Skipjack tuna in the western and central Pacific Ocean under different climate change scenarios. *J. Mar. Sci. Tech.* **2016**, *24*, 1053–1063. (In Chinese)
5. Mugo, R.; Saitoh, S.I.; Nihira, A.; Kuroyama, T. Habitat characteristics of Skipjack tuna (*Katsuwonus pelamis*) in the western North Pacific: A remote sensing perspective. *Fish. Oceanogr.* **2010**, *19*, 382–396. [[CrossRef](#)]
6. Zhu, R.Y. Comparison of factors affecting resource abundance of different stocks of western and central Pacific Skipjack tuna. *J. Shanghai Ocean Univ.* **2025**, *34*, 403–412. (In Chinese)
7. Hsu, T.-Y.; Chang, Y.; Lee, M.-A.; Wu, R.-F.; Hsiao, S.-C. Predicting Skipjack tuna fishing grounds in the western and central Pacific Ocean based on High-Spatial-Temporal-Resolution satellite data. *Remote Sens.* **2021**, *13*, 861. [[CrossRef](#)]

8. Duque-Lazo, J.; Van Gils, H.; Groen, T.A.; Navarro-Cerrillo, R.M. Transfer ability of species distribution models: The case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. *Eco. Model.* **2016**, *320*, 62–70. [\[CrossRef\]](#)
9. Wang, W.S.; Tang, W.; Gong, Y.H. MaxEnt model-based simulation of unassociated *Skipjack tuna* habitat in the western and central Pacific Ocean. *S. China Fish. Sci.* **2023**, *19*, 11–21. (In Chinese)
10. Mugo, R.; Saitoh, S.-I. Ensemble modelling of *Skipjack tuna* (*Katsuwonus Pelamis*) habitats in the western North Pacific using satellite remotely Sensed Data; a comparative analysis using Machine-Learning models. *Remote Sens.* **2020**, *12*, 2591. [\[CrossRef\]](#)
11. Chen, Y.Y.; Chen, X.J.; Guo, L.X.; Fang, Z.; Wang, J.T. Construction and comparison of a BP neural network-based forecasting model for *Skipjack tuna* fishery in the western and central Pacific Ocean. *Guangdong Ocean Univ. J.* **2017**, *37*, 65–73. (In Chinese)
12. Huang, F.M.; Chen, B.; Mao, D.X.; Liu, L.K.; Zhang, Z.H.; Zhu, L. Landslide susceptibility prediction modeling and interpretability based on Self-Screening deep learning model. *Earth Sci.* **2023**, *48*, 1696–1710. (In Chinese)
13. Huang, F.; Zhang, J.; Zhou, C.; Wang, Y.; Huang, J.; Zhu, L. A deep learning Algorithm using a fully connected sparse autoencoder Neural Network for landslide susceptibility prediction. *Landslides* **2020**, *17*, 217–229. [\[CrossRef\]](#)
14. Ciannelli, L.; Fauchald, P.; Chan, K.S.; Agostini, V.N.; Dingsør, G.E. Spatial fisheries ecology: Recent progress and future prospects. *J. Mar. Syst.* **2008**, *71*, 223–236. [\[CrossRef\]](#)
15. Gao, B.; Wang, J.; Stein, A.; Chen, Z. Causal Inference in Spatial Statistics. *Spat. Stat.* **2022**, *50*, 100621. [\[CrossRef\]](#)
16. Sugihara, G.; May, R.; Ye, H.; Hsieh, C.; Deyle, E.; Fogarty, M.; Munch, S. Detecting Causality in Complex Ecosystems. *Science* **2012**, *338*, 496–500. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Gao, B. Temporally or Spatially? Causation Inference in Earth System Sciences. *Sci. Bull.* **2022**, *67*, 232–235. [\[CrossRef\]](#)
18. Gao, B.; Yang, J.; Chen, Z.; Sugihara, G.; Li, M.; Stein, A.; Kwan, M.-P.; Wang, J. Causal inference from cross-sectional earth system data with Geographical Convergent Cross Mapping. *Nat. Commun.* **2023**, *14*, 5875. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Xiao, Y.; Li, G.; Wei, L.; Ding, J.; Zhang, Z. Landslide susceptibility assessment using the Geographical-Optimal-Similarity model. *Appl. Sci.* **2025**, *15*, 1843. [\[CrossRef\]](#)
20. Song, Y. Geographically Optimal Similarity. *Math. Geosci.* **2022**, *55*, 295–320. [\[CrossRef\]](#)
21. Zhao, F.-H.; Huang, J.; Zhu, A.-X. Spatial prediction of groundwater level change based on the Third Law of Geography. *Int. J. Geogr. Inf. Sci.* **2023**, *37*, 2129–2149. [\[CrossRef\]](#)
22. Zhu, A.-X.; Liu, J.; Du, F.; Zhang, S.J.; Qin, C.Z.; Burt, J.; Behrens, T.; Scholten, T. Predictive soil mapping with limited sample data. *Eur. J. Soil Sci.* **2015**, *66*, 535–547. [\[CrossRef\]](#)
23. Sheng, M.; Zhang, W.; Nie, J.; Li, C.; Zhu, A.-X.; Hu, H.; Lou, W.; Deng, X.; Lyu, X.; Ren, Z.; et al. Predicting isoscapes based on an environmental similarity model for the Geographical origin of Chinese rice. *Food Chem.* **2022**, *397*, 133744. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Coletto, J.L.; Pinho, M.P.; Madureira, L.S.P. Operational oceanography applied to *Skipjack tuna* (*Katsuwonus Pelamis*) habitat monitoring and fishing in South-Western Atlantic. *Fish. Oceanogr.* **2019**, *28*, 82–93. [\[CrossRef\]](#)
25. Sund, P.N.; Blackburn, M.; Williams, F. *Tunas and their environment in the Pacific Ocean: A review* *Oceanogr. Mar. Biol. Ann. Rev.* **1981**, *19*, 443–512.
26. Ayers, J.M.; Lozier, M.S. Physical controls on the seasonal migration of the North Pacific transition zone chlorophyll front. *J. Geophys. Res.* **2010**, *115*, C05001. [\[CrossRef\]](#)
27. Ohno, Y.; Kobayashi, T.; Iwasaka, N.; Suga, T. The Mixed Layer Depth in the North Pacific as detected by the Argo floats. *Geophys. Res. Lett.* **2004**, *31*, L11306. [\[CrossRef\]](#)
28. Takahashi, W.; Kawamura, H. Detection method of the Kuroshio front using the satellite-derived chlorophyll-a images. *Remote Sens. Environ.* **2005**, *97*, 83–91. [\[CrossRef\]](#)
29. Yen, K.-W.; Lu, H.-J. Spatial–Temporal variations in primary productivity and population dynamics of *Skipjack tuna* *Katsuwonus Pelamis* in the western and central Pacific Ocean. *Fish. Sci.* **2016**, *82*, 563–571. [\[CrossRef\]](#)
30. He, Y.; Zhao, Z.; Yang, W.; Yan, H.; Wang, W.; Yao, S.; Zhang, L.; Liu, T. A unified network of information considering superimposed landslide factors sequence and pixel spatial neighbourhood for landslide susceptibility mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102508. [\[CrossRef\]](#)
31. Hastie, T.J.; Tibshirani, R. *Generalized Additive Models*; Chapman and Hall: New York, NY, USA, 2017; pp. 249–307.
32. Salazar, J.E.; Benavides, I.F.; Portilla Cabrera, C.V.; Guzmán, A.I.; Selvaraj, J.J. Generalized additive models with delayed effects and spatial autocorrelation patterns to improve the spatiotemporal prediction of the skipjack (*Katsuwonus pelamis*) distribution in the Colombian Pacific Ocean. *Reg. Stud. Mar. Sci.* **2021**, *45*, 101829. [\[CrossRef\]](#)
33. Zhu, A.-X.; Lv, G.-N.; Zhou, C.-H.; Qin, C.-Z. Geographic Similarity: Third Law of Geography? *J. Geogr.-Inf. Sci.* **2020**, *22*, 673–679. (In Chinese)
34. Khanna, K.; Martha, T.R.; Roy, P.; Kumar, K.V. Effect of time and space partitioning strategies of samples on regional landslide susceptibility modelling. *Landslides* **2021**, *18*, 2281–2294. [\[CrossRef\]](#)
35. Tian, H.; Cai, H.; Hu, L.; Qiang, Y.; Zhou, B.; Yang, M.; Lin, B. Unveiling community adaptations to extreme heat events using mobile phone location data. *J. Environ. Manag.* **2024**, *366*, 121665. [\[CrossRef\]](#) [\[PubMed\]](#)

36. Wells, R.O. Differentiable manifolds. In *Differential and Complex Geometry: Origins, Abstractions and Embeddings*; Wells, J.R.O., Ed.; Springer: Berlin/Heidelberg, Germany, 2017.
37. Deyle, E.R.; Sugihara, G. Generalized theorems for nonlinear state space reconstruction. *PLoS ONE* **2011**, *6*, e18295. [[CrossRef](#)] [[PubMed](#)]
38. Schiff, S.J.; So, P.; Chang, T.; Burke, R.E.; Sauer, T. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys. Rev. E* **1996**, *54*, 6708–6724. [[CrossRef](#)] [[PubMed](#)]
39. Hassani, H.; Ghodsi, M.; Huang, X.; Silva, E.S. Is There a Causal Relationship Between Oil Prices and Tourist Arrivals? *J. Appl. Stat.* **2021**, *48*, 191–202. [[CrossRef](#)]
40. Roger, C. Relationships among yellowfin and *Skipjack tuna*, their prey-fish and plankton in the tropical western Indian Ocean. *Fish. Oceanogr.* **1994**, *3*, 133–141. [[CrossRef](#)]
41. Druon, J.-N.; Chassot, E.; Murua, H.; Lopez, J. *Skipjack tuna* availability for purse seine fisheries is driven by suitable feeding habitat dynamics in the Atlantic and Indian Oceans. *Front. Mar. Sci.* **2017**, *4*, 315. [[CrossRef](#)]
42. Trigueros-Salmeron, J.A. Spatial and seasonal variation of relative abundance of the *Skipjack tuna Katsuwonus pelamis* (Linnaeus, 1758) in the Eastern Pacific Ocean (EPO) during 1970–1995. *Fish. Res.* **2001**, *49*, 227–232. [[CrossRef](#)]
43. Forget, F.G.; Capello, M.; Filmlalter, J.D.; Govinden, R.; Soria, M.; Cowley, P.D.; Dagorn, L. Behaviour and vulnerability of target and non-target species at drifting fish aggregating devices (FADs) in the tropical tuna purse seine fishery determined by acoustic telemetry. *Can. J. Fish. Aquat. Sci.* **2015**, *72*, 1398–1405. [[CrossRef](#)]
44. Wang, W.S. Habitat Simulation of Different Bonito Populations in the Western and Central Pacific Ocean Based on the MaxEnt Model. Master's Thesis, Shanghai Ocean University, Shanghai, China, 2023. (In Chinese)
45. Lehodey, P. The pelagic ecosystem of the tropical Pacific Ocean: Dynamic spatial modelling and biological consequences of ENSO. *Prog. Oceanogr.* **2001**, *49*, 439–468. [[CrossRef](#)]
46. Wang, J.; Chen, X.; Staples, K.W.; Chen, Y. The *Skipjack tuna* fishery in the west-central Pacific Ocean: Applying neural networks to detect habitat preferences. *Fish. Sci.* **2018**, *84*, 309–321. [[CrossRef](#)]
47. Reygondeau, G.; Maury, O.; Beaugrand, G.; Fromentin, J.M.; Fonteneau, A.; Cury, P. Biogeography of Tuna and Billfish Communities. *J. Biogeogr.* **2012**, *39*, 114–129. [[CrossRef](#)]
48. Bestley, S.; Gunn, J.S.; Hindell, M.A. Plasticity in vertical behaviour of migrating juvenile southern *Bluefin tuna (Thunnus maccoyii)* in relation to oceanography of the South Indian Ocean. *Fish. Oceanogr.* **2009**, *18*, 237–254. [[CrossRef](#)]
49. Yang, C.L.; Yang, X.M.; Zhu, J.F. Response of environmental factors to the distribution of tuna purse seine bonito in the western and central Pacific Ocean during different types of ENSO. *South China Fish. Sci.* **2021**, *17*, 8–18. (In Chinese)
50. Picaut, J.; Ioualalen, M.; Menkes, C.; Delcroix, T.; McPhaden, M.J. Mechanism of the Zonal Displacements of the Pacific Warm Pool: Implications for ENSO. *Science* **1996**, *274*, 1486–1489. [[CrossRef](#)]
51. Guo, A.; Yu, W.; Chen, X.J.; Qian, W.G.; Li, Y.S. Relationship between spatio-temporal distribution of chub mackerel *Scomber japonicus* and net primary production in the coastal waters of China. *Haiyang Xuebao* **2018**, *40*, 42–52. (In Chinese)
52. Yen, K.-W.; Wang, G.; Lu, H.-J. Evaluating habitat suitability and relative abundance of skipjack (*Katsuwonus pelamis*) in the Western and Central Pacific during various El Niño events. *Ocean. Coast Manag.* **2017**, *139*, 153–160. [[CrossRef](#)]
53. Arrizabalaga, H.; Pereira, J.G.; Royer, F.; Galuardi, B.; Goñi, N.; Artetxe, I.; Arregi, I.; Lutcavage, M. Bigeye tuna (*Thunnus obesus*) Vertical Movements in the Azores Islands Determined with Pop-up Satellite Archival Tags. *Fishe. Oceanogr.* **2008**, *17*, 74–83. [[CrossRef](#)]
54. Meehl, G.A. Characteristics of Surface Current Flow Inferred from a Global Ocean Current Data Set. *J. Phys. Oceanogr.* **1982**, *12*, 538–555. [[CrossRef](#)]
55. Langley, A.; Ogura, M.; Hampton, J. Stock assessment of *Skipjack tuna* in the western and central Pacific Ocean [SKJ-1]. In Proceedings of the Sixteenth Meeting of the Standing Committee on Tuna and Billfish, Mooloolaba, QLD, Australia, 9–16 July 2003.
56. Lewis, A.D.; Williams, P.G. Overview of the western and central Pacific Ocean tuna fisheries, 2001 [GEN-1]. In Proceedings of the Fifteenth Meeting of the Standing Committee on Tuna and Billfish, Honolulu, HI, USA, 22–27 July 2002.
57. Xu, L.X.; Wang, X.F.; Zhu, G.P.; Ye, X.C.; Wang, C.L. Population Structure Analysis of *Skipjack tuna (Katsuwonus pelamis)* in Drifting Object-Associated Schools Under Purse Seine Fisheries in the Western and Central Pacific Ocean. *Chin. J. Ecol.* **2009**, *28*, 293–299. (In Chinese)
58. Fan, X.; Fan, N.; Qin, C.-Z.; Zhao, F.-H.; Zhu, L.-J.; Zhu, A.-X. Large-area soil mapping based on environmental similarity with adaptive consideration of spatial distance to samples. *Geoderma* **2023**, *439*, 116683. [[CrossRef](#)]
59. Vidal, T.; Williams, P.; Ruaia, T. Overview of Tuna Fisheries in the Western and Central Pacific Ocean, Including Economic Conditions—2023. In Proceedings of the Nineteenth Regular Session of the Scientific Committee, Koror, Palau, 16–24 August 2023.
60. Murua, J.; Itano, D.; Hall, M.; Dagorn, L.; Moreno, G.; Restrepo, V. Advances in the Use of Entanglement-Reducing Drifting Fish Aggregating Devices (Dfads) in Tuna Purse Seine Fleets. ISSF Technical Reports. 2016. Available on-

- line: <https://www.iss-foundation.org/about-issf/what-we-publish/issf-documents/issf-technical-report-2016-08-advances-in-the-use-of-entanglement-reducing-drifting-fish-aggregating-devices-in-tuna-purse-seine-fleets/> (accessed on 20 May 2024).
61. He, P.; Suuronen, P. Technologies for the marking of fishing gear to identify gear components entangled on marine animals and to reduce abandoned, lost or otherwise discarded fishing gear. *Mar. Pollut. Bull.* **2018**, *129*, 253–261. [[CrossRef](#)] [[PubMed](#)]
 62. Cillari, T.; Allegra, A.; Andaloro, F.; Gristina, M.; Milisenda, G.; Sinopoli, M. The use of echo-sounder buoys in mediterranean sea: A new technological approach for a sustainable FADs fishery. *Ocean. Coast Manag.* **2018**, *152*, 70–76. [[CrossRef](#)]
 63. McPhaden, M.J.; Zebiak, S.E.; Glantz, M.H. ENSO as an Integrating Concept in Earth Science. *Science* **2006**, *314*, 1740–1745. [[CrossRef](#)]
 64. Zhou, S.F. Impacts of the El Niño I Southern Oscillation (ENSO) phenomenon on *Skipjack tuna* purse seine fisheries in the western and central Pacific Ocean. *J. Fish. Sci. China* **2005**, *6*, 73–78. (In Chinese)
 65. Zhou, S.F.; Shen, J.H.; Fan, W. Analysis of the impacts of the ENSO phenomenon on *Skipjack tuna* purse seine fisheries in the Western and Central Pacific Ocean. *Mar. Fish.* **2004**, *3*, 167–172. (In Chinese)
 66. Delcroix, T.; Picaut, J. Zonal Displacement of the Western Equatorial Pacific “Fresh Pool”. *J. Geophys. Res.* **1998**, *103*, 1087–1098. [[CrossRef](#)]
 67. Miller, A.M.; Bush, S.R.; Van Zwieten, P.A. Sub-regionalisation of fisheries governance: The case of the western and central Pacific Ocean tuna fisheries. *Marit. Stud.* **2014**, *13*, 17. [[CrossRef](#)]
 68. Blanchard, C. Fragmentation in high seas fisheries: Preliminary reflections on a global oceans governance approach. *Mar. Policy* **2017**, *84*, 327–332. [[CrossRef](#)]
 69. Katznelson, D.; Sohns, A.; Kim, D.; Roozee, E.; Donner, W.R.; Song, A.M.; De Vries, J.R.; Temby, O.; Hickey, G.M. Examining the Presence and Effects of Coherence and Fragmentation in the Gulf of Maine Fishery Management Network. *Reg. Environ. Chang.* **2025**, *25*, 3. [[CrossRef](#)]
 70. Gilman, E.; Passfield, K.; Nakamura, K. Performance of regional fisheries management organizations: Ecosystem-based governance of bycatch and discards. *Fish Fish.* **2014**, *15*, 327–351. [[CrossRef](#)]
 71. Pearl, J. *Causality*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009.
 72. Hernan, M.A.; Robins, J.M. *Causal Inference: What If*, 1st ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2020.
 73. Runge, J.; Bathiany, S.; Bollt, E.; Camps-Valls, G.; Coumou, D.; Deyle, E.; Glymour, C.; Kretschmer, M.; Mahecha, M.D.; Muñoz-Mari, J.; et al. Inferring Causation from Time Series in Earth System Sciences. *Nat. Commun.* **2019**, *10*, 2553. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.