

## Article

# Distribution Characteristics of *Trichiurus japonicus* and Their Relationships with Environmental Factors in the East China Sea and South-Central Yellow Sea

Xinyu Shi <sup>1,2</sup>, Zhanhui Lu <sup>2,3</sup>, Zhongming Wang <sup>2,3</sup>, Jianxiong Li <sup>1,2</sup>, Xin Gao <sup>1,2</sup>, Zhuang Kong <sup>1,2</sup> and Wenbin Zhu <sup>1,3,\*</sup>

- <sup>1</sup> Marine and Fishery Institute, Zhejiang Ocean University, Zhoushan 316021, China; shixinyu@zjou.edu.cn (X.S.); lijianxiong@zjou.zjou.edu.cn (J.L.); gaoxin@zjou.zjou.edu.cn (X.G.); kongzhuang@zjou.edu.cn (Z.K.)
- <sup>2</sup> Zhejiang Marine Fisheries Research Institute, Zhoushan 316021, China; lzh4286@zjou.edu.cn (Z.L.); wangzhongming@zjou.edu.cn (Z.W.)
- <sup>3</sup> Key Laboratory of Sustainable Utilization of Technology Research for Fishery Resource of Zhejiang Province, Zhoushan 316021, China
- \* Correspondence: wenbinzhu@zjou.edu.cn

**Abstract:** The largehead hairtail (*Trichiurus japonicus*) is the most productive fish caught in China. In order to understand the seasonal distribution of *T. japonicus* in the East China Sea and the central and southern parts of the Yellow Sea, three species distribution models were used in this study, namely the random-forest model, K-nearest-neighbor algorithm, and gradient-ascending decision-tree model, based on the data of trawling surveys in the East China Sea and central and southern parts of the Yellow Sea from 2008 to 2009. Combined with a variance inflation factor and cross-check, a distribution model of *T. japonicus* was screened and constructed to analyze the influence of environmental factors on the distribution of *T. japonicus* in the East China Sea and central and southern parts of the Yellow Sea. The results showed that the random-forest model had the advantages of fitting effect and prediction ability among the three models. The analysis of this model showed that the water depth, bottom water temperature, and surface salinity had a great influence on the habitat distribution of *T. japonicus*. The relative resources of *T. japonicus* increased with the increase of bottom water temperature, reached the maximum at 23.8 °C, and first increased and then decreased with the increase of water depth and surface salinity, reaching the maximum when water depth is 72 m and surface salinity is 31.2‰. This study also used the random-forest model to predict the spatial distribution of *T. japonicus* in the central and southern waters of the East China Sea and south-central Yellow Sea from 2008 to 2009, and the results showed that the predicted results were close to the actual situation. The research results can provide a reference for the exploitation and protection of *T. japonicus* resources in the East China Sea and the south-central Yellow Sea.

**Keywords:** East China Sea; south-central Yellow Sea; *T. japonicus*; species distribution model; spatial distribution; environmental factor

**Key Contribution:** In this study, we used a random-forest model to analyze the effects of environmental factors on the distribution of *T. japonicus* in the East China Sea and the south-central Yellow Sea and made distribution predictions. The results of this study can help to comprehensively understand the distribution of *T. japonicus* fisheries in the East China Sea and the south-central Yellow Sea and provide valuable theoretical support for their rational development and utilization.



**Citation:** Shi, X.; Lu, Z.; Wang, Z.; Li, J.; Gao, X.; Kong, Z.; Zhu, W. Distribution Characteristics of *Trichiurus japonicus* and Their Relationships with Environmental Factors in the East China Sea and South-Central Yellow Sea. *Fishes* **2024**, *9*, 439. <https://doi.org/10.3390/fishes9110439>

Academic Editors: Heng Zhang, Yunzhi Yan, Luoliang Xu and Ying Xue

Received: 14 September 2024

Revised: 9 October 2024

Accepted: 12 October 2024

Published: 29 October 2024



**Copyright:** © 2024 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/>).

## 1. Introduction

The largehead hairtail (*Trichiurus japonicus*) (Temminck and Schlegel, 1844) belongs to the order Perciformes, family Trichiuridae, and genus *Trichiurus*. It is a warm-temperate

species that typically forms schools near the seafloor. The catch of this species exceeded 1 million tons in 1995 [1], making it one of the few marine species in China with over one million tons in landing. Currently, the main fishing methods in the East China Sea and Yellow Sea include bottom trawling and seine netting. *T. japonicus* resources in the East China Sea have been exploited since the 1950s, with catches often ranking first among various species since the late 1950s. Consequently, *T. japonicus* is a key species in domestic fisheries research and management. Many resource management systems in the East China Sea, including fishing bans and protected areas, are based on research findings related to *T. japonicus* resources, primarily focusing on conserving traditional economic fish species, with *T. japonicus* being the primary target [2].

One of the hot topics in fishery ecology is the spatial distribution characteristics of species and their relationships with environmental factors [3]. The spatial distribution of fish populations is influenced by a variety of control factors, both external and internal, of which the external control, also known as environmental control, includes hydrological conditions, substrate types, etc., and is generally considered to be the main factor affecting the spatial distribution of fish populations [4]. On the other hand, population size, age structure, fish condition, diversity, behavior, etc., are internal control factors that can also regulate the spatial distribution of fish populations through density-dependent, age-dependent habitat preference, migration ability differences, etc. [5]. The adaptability and limitation of fish to marine environments are one of the key factors determining their migration, distribution, and movement, and the study of the influence of environmental factors on the spatial distribution of fish populations is of great reference value for fishery analysis, fishing ground exploration, and rational use of fishery resources [6].

A species distribution model (SDM) is a mathematical model that uses environmental data to predict the spatial distribution of species according to their survival conditions and has become one of the important methods in the application of conservation biology and ecology [7]. Widely used species distribution models in fisheries include generalized additive models and generalized linear models [8,9], with relatively fewer applications of machine learning methods. As automation and intelligence advance, machine learning algorithms increasingly predict fish abundance and distribution [10], identify populations [11], standardize catch per unit effort (CPUE) [12], and explore relationships between fishery resources and environmental factors [13,14], showing distinct advantages.

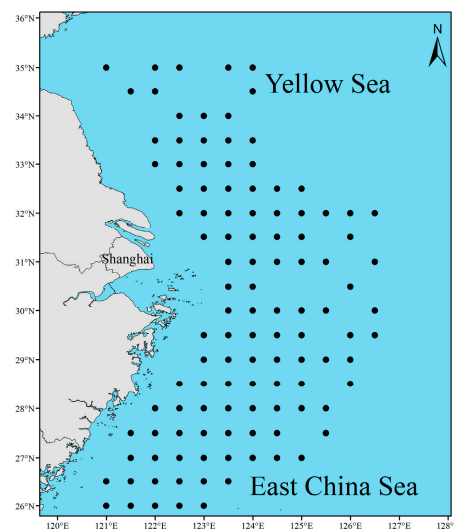
For instance, Chen [15] developed a forecasting model for Indian Ocean yellowfin tuna fisheries using a random-forest model, enhancing the forecasting capabilities of distant offshore fisheries. Hou [16] researched the modeling and forecasting of South Pacific yellowfin tuna fisheries using six ensemble learning models, improving the accuracy of their predictions. Gao [17] constructed a forecasting model for mackerel in the East and Yellow Seas, employing gradient boosting decision trees and playing a crucial role in managing and protecting mackerel resources. Song [18] built a forecasting model for bigeye tuna in the Atlantic tropical waters using K-nearest-neighbors and gradient-boosting decision trees, enhancing the accuracy of their model predictions. Currently, research utilizing species distribution models to examine the habitat distribution of ribbonfish remains scarce.

Based on the trawl survey data in the central and southern waters of the East China Sea and the Yellow Sea from 2008 to 2009, this study used random-forest model, K-proximity algorithm, and gradient-lifting decision tree to analyze the distribution characteristics of *T. japonicus* and their relationships with environmental factors and then compared and analyzed the fitting effect and prediction ability of the models. The habitat index was used to predict the distribution of *T. japonicus* in the East China Sea and the south-central Yellow Sea, so as to provide a basis for the rational utilization and scientific conservation of its resources and provide a reference for fishery policy management.

## 2. Materials and Methods

### 2.1. Data Sources

The samples of belt fish in this study were collected from the fixed bottom trawl survey of the national science and technology support program “Investigation and Assessment of important fishery Resources in the Main fishing grounds of the East China Sea” conducted in May (spring), August (summer), and December (autumn) of 2008 and February (winter) of 2009 in the waters of the East China Sea and the south-central Yellow Sea. The sea area covered 121–126.5° E and 26–35° N (Figure 1), with 119 stations. The survey ship used a 6 × 80 m target net; the width of the network port was 48 m, the mesh size of the bag net was 30 mm, the towing speed of the survey ship was 2.0 kn, and the towing time of each station was 1 h. The relative catch  $Y$  (g/h) was obtained by using the trawling time of 1 h and trawling speed of 2 kn.



**Figure 1.** Survey stations.

Sample processing and environmental factor measurements adhered to the “Ocean Survey Standards” [19]. Environmental data were collected using a shipborne synchronized CTD instrument, which measured sea water depth (SWD), sea surface temperature (SST), sea bottom temperature (SBT), sea surface salinity (SSS), and sea bottom salinity (SBS).

### 2.2. Model Construction

Random forest (RF), proposed by Breiman [20], is an ensemble-learning method based on the classification and regression tree algorithm. This approach improves the predictive performance of models by combining multiple decision trees. Specifically, random forest achieves this through several steps: first, it randomly extracts multiple samples from the original data set, known as bootstrap samples; next, it models decision trees for each bootstrap sample; finally, it aggregates the predictions from each decision tree, arriving at the final prediction through voting or averaging. This method exhibits high tolerance to noise and outliers, achieves high classification accuracy and predictive precision, shows a lower probability of overfitting, and possesses strong generalization capabilities [21,22].

The expression for a random forest is as follows:

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m), I = \begin{cases} 1, X \in R_m \\ 0, x \notin R_m \end{cases} \quad (1)$$

In the formula,  $x$  is the independent variable,  $y$  is the dependent variable,  $R_m$  is the feature space partition unit, and  $c_m$  is the fixed output value on each unit.

The K-nearest-neighbor (KNN) algorithm serves as a widely adopted classification method. The steps for classification are straightforward: First, compute the distance

between an object, whose category is unknown, and every sample in the training set. Next, select the K-most-similar (nearest) samples within the feature space. Then, determine which category most of these K samples belong to. Finally, if the majority of samples fall into a specific category, classify the object into that category as well [23]. The fundamental concept behind the KNN algorithm is clear: if most of the K-nearest samples reside in a particular category, the sample should be assigned to that category too [24]. The KNN algorithm can facilitate both regression and classification by evaluating distances between various feature values. For any N-dimensional input vector, which correlates to a point in the feature space, the outcome is the category label or predicted value associated with that feature vector. While the concept remains simple and intuitive, the algorithm boasts significant maturity and stability [25–27].

The expression for KNN is as follows:

$$d(x, y_i) = \|x - y_i\|_2 = \left( \sum_{k=1}^n |x_k - y_{ik}|^2 \right)^{1/2} \quad (2)$$

In the formula,  $x$  is the sample to be classified,  $y$  is a known category sample,  $n$  represents the data dimension, and  $i, k$  is the sample number.

A gradient boosting decision tree (GBDT) is an enhanced ensemble learning model based on the classification and regression trees (CART) algorithm [28]. It is one of the important algorithms in the field of machine learning. This model combines multiple weak classifiers into strong classifiers by iterating continuously. In each iteration, based on the previous iteration, the loss function is calculated to obtain the pseudo-residual and the iteration is obtained, and then, a new decision tree is constructed. Then, all the generated decision trees are weighted and fused according to the weight of the decision tree through gradient descent [29]. The model can deal with nonlinear relations effectively, has good generalization performance and accuracy in many prediction studies, and can identify and correct errors in the modeling process. However, GBDT is sensitive to outliers, and in multiple iterations, GBDT models will try to fit outliers, which may lead to overfitting. Therefore, when applying this model, hyperparameter tuning should be performed on the imported data to obtain the optimal solution of parameters and reduce the risk of overfitting [30].

The expression for GBDT is as follows:

$$F_i(x) = \sum_{t=1}^m r^{(t)} f^{(t)}(x_i) \quad (3)$$

In the formula,  $F_i(x)$  represents the final prediction result for the  $i$ -th sample, and  $F_i(x)$  is the predicted value of the  $i$ -th sample in the  $t$ -th tree for the observed sample  $x$ .

### 2.3. Factor Screening and Model Fitting

$\ln(Y+1)$  was obtained by natural logarithm conversion of the relative resource amount ( $Y$ ) of *T. japonicus* as the response variable and SWD, SST, SBT, SSS, and SBS were selected as the explanatory variables. A significant correlation between two or more explanatory variables in a multicollinearity representation model can negatively affect the final result. In order to avoid such influence, a variance inflation factor (VIF) [31] is used in this study to test the multicollinearity of the above five factors and screen out the factors to be added to the model. In general,  $\sqrt{VIF} < 2$  indicates that there is no multicollinearity, and explanatory variables that exceed the threshold need to be removed.

$$VIF = \frac{1}{1 - R^2} \quad (4)$$

In the formula,  $R^2$  represents the goodness of fit of the relationship between the independent variable and other independent variables through a simple linear regression model.

#### 2.4. Evaluation of Model Prediction Ability

We compared the fitting effects and predictive capabilities of three models to select the optimal one. We analyzed the relationship between the distribution characteristics of *T. japonicus* in the East China Sea and south-central Yellow Sea and environmental factors. Subsequently, we predicted their distribution.

The prediction ability of each model was tested by the 50-fold cross-validation method. The total data set was randomly and equally divided into 5 sub-data sets. Each time, 4 sub-data sets were randomly selected as the training set, and the other one was used as the validation set for the accuracy evaluation of the model prediction. The calculation was repeated 100 times, and the average effect was taken for the accuracy evaluation of each model. According to the mean squared error (MSE) and coefficient of determination ( $R^2$ ) obtained, the prediction ability of each model was determined.

The MSE is the ratio of the square sum of the deviation between the predicted value and the true value and the number of observations  $n$ , which can reflect the degree of dispersion of the data set [32]. The smaller the MSE value, the higher the accuracy of the model prediction and the more accurate the description of the test data.  $R^2$  is the proportion of the sum of squares caused by the independent variable  $X$  in the total sum of squares of the dependent variable  $Y$  [33], which can be used to evaluate the fitting degree of the prediction model. The closer  $R^2$  is to 1, the higher the reference value of the model, which can well describe the trend and rule of the data set. The closer  $R^2$  is to 0, the lower the reference value of the model, and the trend and rule of the data set cannot be well described [34].

The formulas for calculating the MSE and  $R^2$  are as follows:

$$MSE(y, p) = \frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2 \quad (5)$$

$$R^2(y, p) = 1 - \frac{\sum_{i=0}^n (y_i - p_i)^2}{\sum_{i=0}^n (y_i - \bar{y})^2} \quad (6)$$

In the formula,  $y$  represents the original value,  $p$  stands for the predicted value, and  $n$  denotes the sample size.

#### 2.5. Mapping Habitat Distribution Prediction

The habitat suitability index (HSI) was initially proposed in the 1980s [35] and is primarily utilized for assessing habitat quality, providing a more comprehensive depiction of the adaptation process of marine organisms to their environment. Currently, it has gained widespread application in the fields of biological spatial distribution and fishing-ground forecasting [36–38]. In this study, after conducting comparisons, we selected the species distribution model with superior predictive performance. Subsequently, HSI values were calculated for each station and the ArcGIS 10.2 software's spatial analysis module was employed to generate habitat distribution maps for *T. japonicus* during different seasons using Kriging interpolation based on an exponential semi-variance function [39].

### 3. Results and Analysis

#### 3.1. Impact Factor Screening

Five factors (SWD, SST, SBT, SSS, and SBS) were tested with multicollinearity using  $V_{IF}$ , and their values were 1.78, 1.32, 1.51, 1.53 and 1.27, respectively. The results show that there is no multicollinearity between the factors and they all can be added to the model.

#### 3.2. Model Performance Evaluation

As can be seen from Table 1, after model fitting, the MSE of the random-forest model is 0.348, which is smaller than values for the KNN and GBDT models, and the  $R^2$  is 0.919, which is higher and closer to one than the values for the KNN and GBDT models. Therefore, this model has the best fitting effect. The mean values of the MSE and  $R^2$  of 100 model

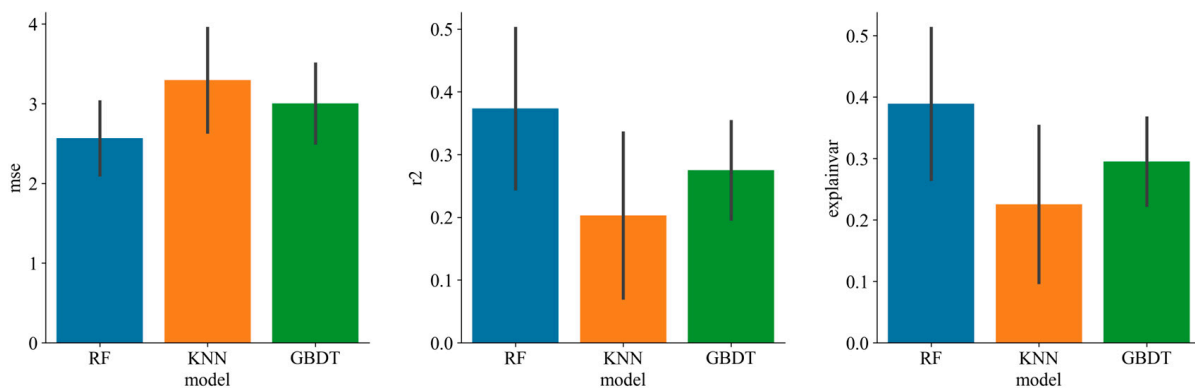
predictions and observations were obtained by cross-validation. The results show that the MSE of the random-forest model is  $2.566 \pm 1.734$ , is smaller than the values for the KNN and GBDT models, and its  $R^2$  is  $0.373 \pm 0.563$ , which is higher than the values for the KNN and GBDT models. The difference between the prediction results of random-forest model and the observed values is smaller, so the prediction ability is the best. Therefore, the random-forest model is superior to KNN and GBDT in all aspects, so the random-forest model is adopted for follow-up research.

**Table 1.** Cross-validation comparison between three models.

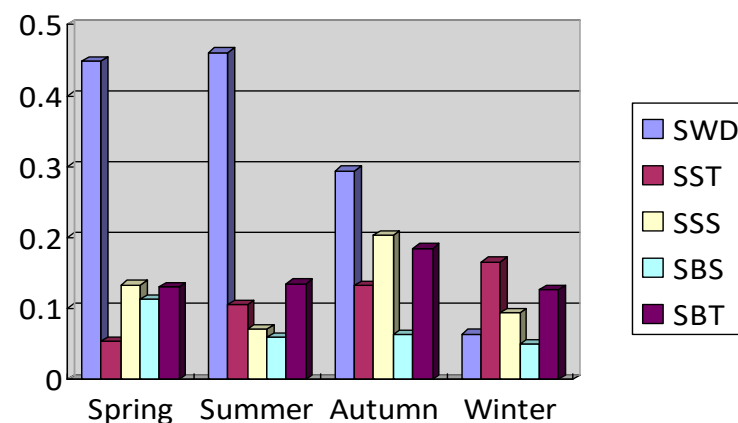
Inspection Method	Statistical Parameters	RF	KNN	GBDT
Model fitting	MSE	0.348	2.120	2.445
	$R^2$	0.919	0.506	0.431
Cross validation	MSE	$2.566 \pm 1.734$	$3.295 \pm 2.161$	$3.004 \pm 1.264$
	$R^2$	$0.373 \pm 0.563$	$0.203 \pm 0.385$	$0.275 \pm 0.255$

### 3.3. Importance Ranking of Impact Factors

In RF, the contribution rate of a feature is usually calculated based on the number of node splits of the feature in the decision tree and the information gain obtained by splitting. The random-forest model was constructed, and the input variables were SWD, SST, SBT, SSS, and SBS, and the output variable was resource density. The results show that in the random-forest model, the contribution rates of each impact factor to resource density in different months are shown in Figures 2 and 3.



**Figure 2.** Performance comparison of the three machine learning methods.



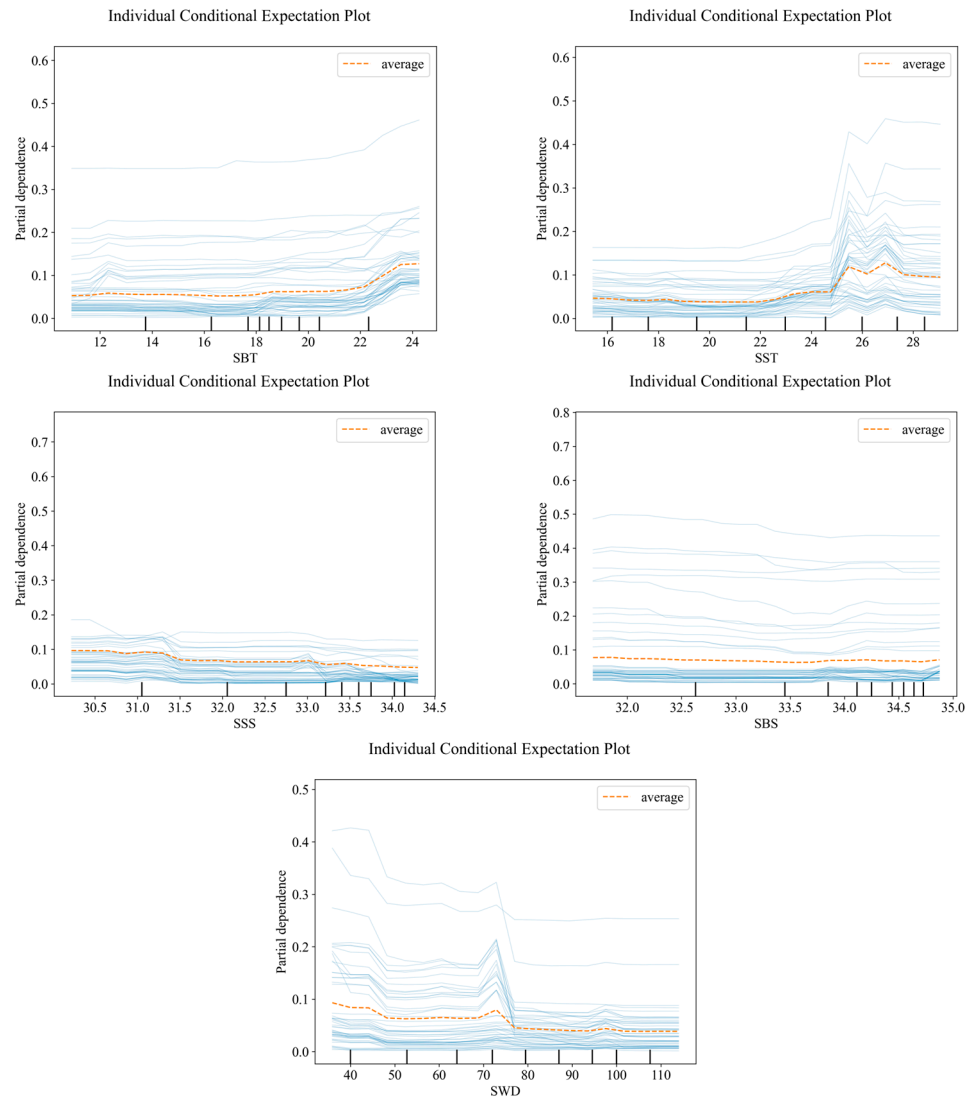
**Figure 3.** Importance ranking of factors affecting the density distribution of *T. japonicus* in the study area.



The results show that SWD is the most important in May (spring), followed by SSS, SBT, SBS, and SST; SWD is the most important in August (summer), followed by SBT, SST, SSS, and SBS; SWD is the most important in November (autumn), followed by SSS, SBT, SST, and SBS; and SST is the most important in February (winter), followed by SBT, SSS, SWD, and SBS. It can be seen that among the five factors, SWD, SBT, and SSS are relatively more important.

### 3.4. Relationship between the *T. japonicus* Distribution and Explanatory Variables

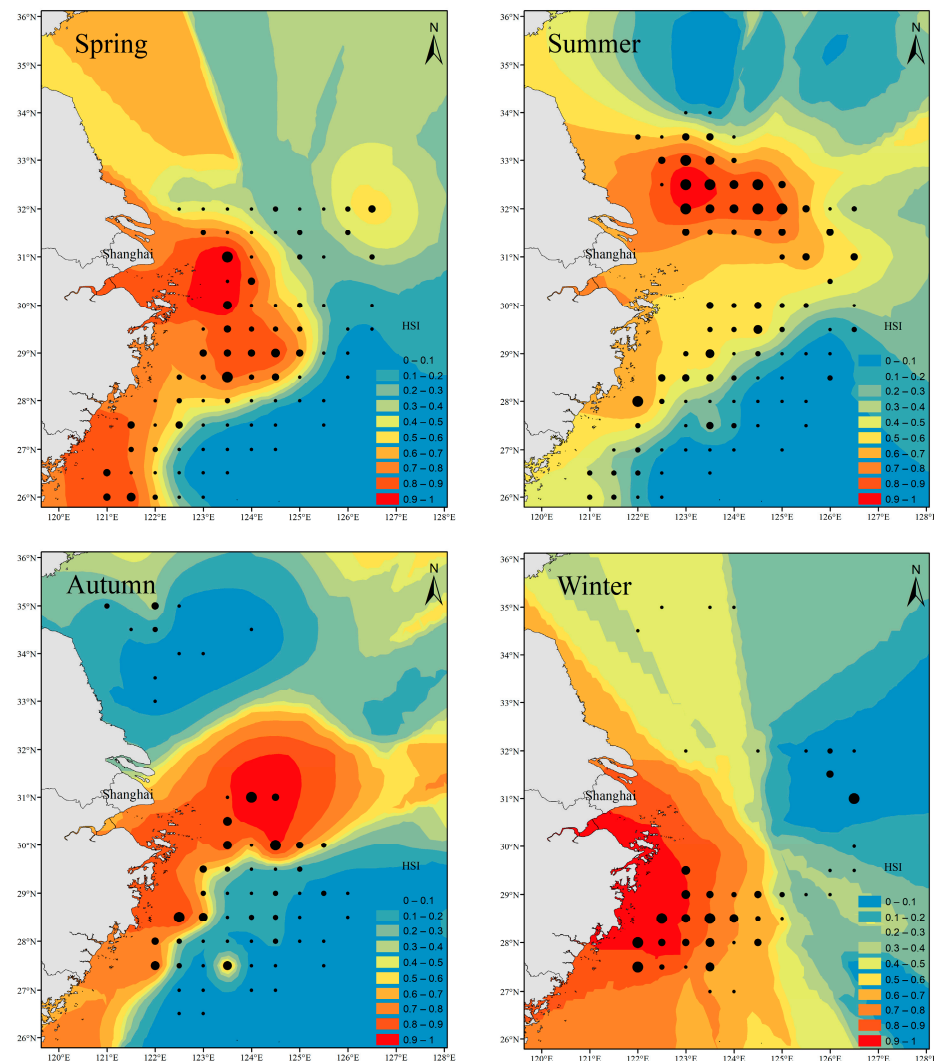
The influences of various factors on the relative resources for *T. japonicus* are shown in Figure 4. The relative resources of belt fish increased slowly when the SST was less than 24.8 °C, fluctuated after 24.8 °C, and became stable after 27 °C. The relative resources of *T. japonicus* increased slowly when the SBT was less than 22.2 °C and the increase rate increased after 22.2 °C and reached the maximum at 23.8 °C, showing an overall increasing trend. The relative resources of *T. japonicus* increased when the SSS was less than 31.2%, reached the maximum at 31.2%, and decreased when the SSS was more than 31.2%. The relative resource amount of *T. japonicus* showed a higher level when the SBS was less than 33.3%, and a lower level when the SBS was more than 33.3%, showing a decreasing trend in general. The relative resources of belt fish increased when the SWD was less than 72 m, reached the maximum at 72 m, and decreased after 72 m.



**Figure 4.** Impact of environmental factors on the relative resource of *T. japonicus*. From left to right and top to bottom are SBT SST, SSS, SBS, and SWD.

### 3.5. Prediction of Habitat Distribution of *T. japonicus* in the East China Sea and South-Central Yellow Sea

The prediction performance of the three models was compared, and it was found that the RF model had the best prediction performance. The environmental data simulated by the HSI were added to the random-forest model for prediction, and the spatial distribution map was drawn. It was found that the abundance distribution of *T. japonicus* in May is high in the southwest, low in the northeast, and mainly distributed in the sea areas of 25.5–31.5° N and 119.5–124° E. In August, the abundance of *T. japonicus* is mainly concentrated in the northwest waters, mainly distributed in the waters of 30.5–33° N and 121–125° E. In November, *T. japonicus* resources were mainly concentrated in the middle of the sea area, mainly distributed in the sea areas of 29–33° N and 122–126° E. In February, the distribution of *T. japonicus* abundance showed the characteristics of being high in the southwest sea area, low in the northeast sea area, and mainly distributed in the sea areas of 27–30° N and 121–124.5° E (Figure 5). These results are consistent with the characteristics that *T. japonicus* likes to cluster in the warm environment near the bottom. In Figure 5, the predicted results are compared with the actual results, and it is found that the predicted results are close to the actual results, which shows that the predicted results have a certain accuracy.



**Figure 5.** Simulated habitat and actual survey site of *T. japonicus* in different seasons.



## 4. Discussion

### 4.1. Model Analysis

At present, there are few studies on the distribution of *T. japonicus* using a species distribution model. Zhang [40] used a GAM model to study the distribution characteristics of *T. japonicus* in the Beibu Gulf from 2006 to 2018 and the relationship between its resource density and environmental factors. The results showed that the distribution of *T. japonicus* fishing grounds in the Beibu Gulf was southwest–northeast, and the center of gravity of *T. japonicus* resources moved to southwest–northeast and south–north in summer and autumn, respectively. Chlorophyll A affected the resource density and spatial distribution of *T. japonicus*, and abnormal values of water depth and longitude and sea surface temperature affected the resource density of *T. japonicus* but did not affect its spatial distribution. Liu [41] predicted the potential distribution areas of *T. japonicus* in the coastal waters of China in 2040–2050 and 2090–2100 by using nine species distribution models such as a random-forest model based on the survey data of fishery resources from 1998 to 2000, and the prediction results showed that the distribution hotspots of *T. japonicus* tended to move to high latitudes, and the investigation stations and research simulation in that study overlapped with those in this study.

In this study, three machine learning models, the RF, KNN and GLDT models, were compared in order to choose the most suitable model to analyze the habitat distribution characteristics of *T. japonicus* in the East China Sea and the central and southern parts of Yellow Sea and their relationships with environmental factors. The results showed that the random-forest model has a good fitting effect and cross-validation result, which may be attributed to its advantages in data processing ability and algorithm. Firstly, a random-forest model introduces the concept of randomness, and randomly selects training samples and feature subsets, thus effectively enhancing the classification ability and anti-noise ability and reducing the possibility of over-fitting of random forest. At the same time, random forest can effectively deal with the situation of less data, lost features, and unbalanced data sets and has a high tolerance for outliers. In addition, the random-forest model has the characteristics of integrated learning, and the accuracy of the results can be improved by constructing many different regression trees, avoiding the weak generalization ability of a single decision tree [21,22].

The utilization of the random-forest model in the fisheries domain has been progressively increasing in recent years. In comparison to conventional species, distribution models, and other machine learning techniques, the random-forest model can effectively capture the interaction between environmental variables through constructing a randomized decision tree [42]. Furthermore, it demonstrates greater robustness against outliers and random interference [20] during regression analysis. Luan [3] employed a GLM, a GAM, and the random-forest model to evaluate the spatial distribution of Portunid crab across different seasons in Haizhou Bay in 2011. Liu [42] utilized both the random-forest model and a GAM model to analyze the relationship between krill catch per unit fishing effort and environmental factors in Antarctica. Cui [13] employed artificial-neural-network models, random-forest models, and generalized-enhanced-regression models to predict and compare habitat distributions for *Tetragnatha tetragnatha* in Haizhou Bay. All the above results show that RF has good fitting effect and forecasting ability and has certain advantages, which is similar to the results in this study. However, since random forests are composed of a large number of decision trees containing only partial feature variables, the prediction results are dependent on the mode of the output category of the decision tree. As a result, the interpretation of the results obtained from random forests is challenging. Therefore, when employing random forests for fishery forecasting, it is essential to complement this approach with other analytical techniques to gain a deeper understanding of the underlying dynamics shaping the fishery. This will help to mitigate the limitations of this model in fishery forecasting.

#### 4.2. The Influence of Environmental Factors on the Distribution of *T. japonicus*

The temporal and spatial differences of environmental factors are one of the main reasons for the temporal and spatial changes of fish resources; environmental factors influence the growth of individual fish and their age at maturity [43] and fish usually distribute along the distribution characteristics according to environmental gradients [44]. In this study, the distribution characteristics of *T. japonicus* and their relationships with environmental factors were analyzed by a random-forest model. It was found that environmental factors have different effects on the distribution of *T. japonicus* in different seasons, and the SWD, SSS and SBT are relatively important.

Water temperature is a very important environmental factor, which can directly affect the growth, development, reproduction, metabolism, migration, and distribution of marine organisms and other ecological processes [45]. As a warm-temperate fish clustered near the bottom, the growth and reproduction of *T. japonicus* are directly affected by water temperature, so its distribution area will also be affected by water temperature [46–48]. Wang [46,47] found that the increase of water temperature is not only beneficial to the gonad development and maturity of *T. japonicus*, but also can increase the feed supply of *T. japonicus*. The fluctuation of *T. japonicus* catch in the East China Sea is significantly related to the sea surface temperature; Yuan [48] found that the hot spots of *T. japonicus* in the East China Sea will move adaptively with the first mode change of sea surface temperature, and they are all close to the left waters of the northern branch of the Kuroshio. You [49] found that the bottom water temperature of the central fishing ground of Zhoushan fishing ground in summer flood was between 16 °C and 22 °C, and the temperature at that time might be lower than now, earlier in this research year. In addition, this research only studied *T. japonicus* in summer flood, so the result was different from this research. It was found that the relative resources of *T. japonicus* were small when the SST was less than 24.8 °C and obviously fluctuated and rose above 24.8 °C, reaching the maximum at 27 °C. The SBT is considered of low resource density when it is less than 22.2 °C, gradually increases when it is higher than 22.2 °C, and reaches the maximum at 23.8 °C, which indicates that *T. japonicus* has certain requirements on water temperature and is suitable for its survival within a certain temperature range. The relative resources of *T. japonicus* reached the maximum when the SBT was 23.8 °C, which indicated that the water temperature was suitable for *T. japonicus* survival. In the SBT range of 22.2–24.2 °C, the relative resources of *T. japonicus* are relatively large, and this temperature range is the suitable water temperature range for its life.

As one of the important environmental factors affecting the spatial and temporal distribution of marine life, salinity can affect the spatial distribution of marine life to a certain extent [50]. Previous studies have found that the fishing season of *T. japonicus* is directly affected by salinity [49,51]. You [49] found that the central fishing ground of *T. japonicus* in the Zhoushan fishing ground is located near the 34‰ isosalinity line. Zhu [51] and others found that the zonal fluctuation of the 34‰ isosalinity line was obviously related to the fishing ground in central Zhejiang during winter flood. Wang [52] found that sea surface salinity has a significant influence on the change of *T. japonicus* catch in the Zhejiang sea area, and the *T. japonicus* catch shows a linear upward trend with the increase of sea surface salinity. The above research is earlier, which may lead to a gap between those results and the results of this study. This study found that the relative resources of *T. japonicus* were higher when the SSS was less than 31.2‰, and lower after the SSS was more than 31.2‰. It is higher when the SBS is less than 33.3‰ and lower when the SBS is more than 33.3‰, indicating that salinity will affect the distribution of *T. japonicus* resources. Too high or too low salinity in seawater may affect the osmotic pressure adjustment and oxygen consumption of *T. japonicus*, thus affecting its growth. The normal growth of *T. japonicus* needs to be carried out within a certain salinity range.

Water depth affect the changes of factors such as light, pressure, and dissolved oxygen and can indirectly affect the habitat distribution of marine life and its bait [13]. Water depth is closely related to water mass movement of the fish community, processes related to

fish life history (predation and competition), and bottom sediments [53]. Some studies have found that the depth of seawater directly affects the temporal and spatial changes of hydrological factors such as temperature, salinity, and transparency, thus directly affecting the distribution of organisms and the aggregation of fish [18]. Hu [54] found that water depth is one of the main factors affecting the diversity of fish communities in a *T. japonicus* reserve in spring and autumn. Zhang [55] found that water depth is one of the main environmental factors affecting the distribution of fish in the coastal waters of the Yangtze River Estuary, and the diversity of the fish community and the distribution of the fish may affect the feeding of *T. japonicus*, thus affecting the relative resources of *T. japonicus* to some extent. The above research results are similar to this study. In this study, the relative resources of *T. japonicus* showed an upward trend when the SWD was less than 72 m, and a slow downward trend after the SWD was higher than 72 m. This is consistent with *T. japonicus*'s habit of clustering near the bottom, which shows that there is the most suitable habitat environment for *T. japonicus* in the sea area with a water depth of less than 72 m. However, this study used past data and only considered five environmental factors, which has certain limitations. More environmental factors should be added for further research in the future.

#### 4.3. Habitat Distribution Characteristics of *T. japonicus*

The habitat distribution of *T. japonicus* in the East China Sea and the central and southern parts of the Yellow Sea is characterized by high resource density in the southwest coastal and central waters and low resource density in the southeast and northern waters in spring, when *T. japonicus* is mainly distributed in the waters of 27.5–31° N and 122.5–125° E. In summer, the coastal waters in the northwest and southwest are characterized by high resource density, low resource density is found in the southeast, and *T. japonicus* is mainly distributed in the sea areas of 28–30° N, 122–124.5° E, 31.5–33.5° N, and 123–125° E. In autumn, the resource density in the southwest coast and central sea area is high, while the resource density in the north and southeast sea area is low, and *T. japonicus* is mainly distributed in the sea areas of 27.5–28.5° N, 121.5–123.5° E, 30–31° N, and 123.5–125° E. In winter, the resource density in the southwest is high, the resource density in other sea areas is low, and *T. japonicus* is mainly distributed in the sea areas of 27.5–29.5° N and 122–124.5° E. It can be found that with the increase of temperature, the distribution area of *T. japonicus* hotspots moves northward and outward, which is similar to the research conclusions of Yuan [48] and Zhu [51]. The distribution area of *T. japonicus* resource density hotspots spread to a certain extent with the seasonal changes and moved to the northern offshore, which is not only related to the rising water temperature in seasonal changes, but also related to the fact that the effective implementation of the summer fishing moratorium in the East China Sea and the central and southern parts of the Yellow Sea is beneficial to effectively replenish *T. japonicus* resources. According to the research of Yan [56], the summer fishing moratorium can protect the spawning groups and juveniles of major economic fish, reduce the fishing pressure, facilitate the cluster growth of *T. japonicus*, and distribute *T. japonicus* hotspots. Using the HSI index to predict the habitat distribution of *T. japonicus* in the East China Sea and south-central Yellow Sea in four seasons can make up for some missing data, and it also enables the effective assessment of stock [57].

## 5. Conclusions

By comparing three kinds of machine learning models, this study analyzed the habitat distribution characteristics of *T. japonicus* in the East China Sea and south-central Yellow Sea and their relationships with environmental factors. On the basis of the results and their interpretation, the following is noted: (1) The random-forest model had better fitting effect and prediction ability than the other two kinds of machine learning models; (2) Among the five environmental factors, SWD, SBT, and SSS had a great impact on the habitat distribution of belt fish. The relative resources of belt fish increased with the increase of the SBT and increased first and then decreased with the increase of the SWD and SSS;

(3) Based on the habitat index the habitat of *T. japonicus* in the East China Sea and the central and southern parts of the Yellow Sea was predicted, and the predicted results were similar to the actual survey results. The study of the habitat distribution characteristics of *T. japonicus* in the East China Sea and the south-central Yellow Sea and their relationships with environmental factors can provide some reference for the sustainable utilization and scientific management of *T. japonicus* resources. In future studies, comparative statistical methods and machine learning methods should be used to help explore models and methods that are more suitable for this species. In addition, this study only considered five environmental factors, namely SWD, SST, SBT, SSS, and SBS, and did not involve other factors, such as dissolved oxygen, chlorophyll, pH, flow rate, and mixed layer depth, which may affect the distribution of belt fish. More environmental factors and their effects on the habitat distribution of belt fish should be comprehensively analyzed. In order to further understand the relationship between the habitat distribution characteristics and environmental factors of *T. japonicus* in the East China Sea and south-central Yellow Sea and to provide reference for the protection and rational utilization of *T. japonicus* resources in the East China Sea and south-central Yellow Sea, further research is needed.

**Author Contributions:** Conceptualization, X.S. and W.Z.; methodology, W.Z.; writing—original draft preparation, X.S.; writing—review and editing, J.L., X.G., and Z.K.; supervision, W.Z. and Z.W.; project administration, W.Z. and Z.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the Key Technology and System Exploration of Quota Fishing, Ministry of Agriculture and Rural Affairs, Fishery Management Fund Project [Grant number: 36,2017], the National Key Research and Development Program of China (2019YFD0901505), the Zhejiang Provincial Key R&D Program project (2018C02026) and the Zhejiang Fishery Resources Survey Special Project [HYS-CZ-202314].

**Institutional Review Board Statement:** The data for this study were collected from a fixed bottom trawl survey in the East China Sea between 2008–2009 as part of a national science and technology support program. Since no actual fish were used, ethical approval and consent forms are not required, and there are no animal ethics concerns.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data will be made available on request.

**Acknowledgments:** We thank the staff of the Zhejiang Marine Fisheries Research Institute for their help and support in our experiment.

**Conflicts of Interest:** The authors declare that they have no competing interests.

## References

1. Zhang, Q.H.; Cheng, J.H.; Xu, H.X. *Fishery Resources and Their Sustainable Utilization in The East China Sea*; Fudan University Press: Shanghai, China, 2007; pp. 147–169.
2. Cheng, J.H.; Yan, L.P.; Lin, L.S. Analyses on the fishery ecological effect of summer close season in the East China Sea region. *J. Fish. Sci. China* **1999**, *4*, 81–85.
3. Luan, J.; Zhang, C.L.; Xu, B.D. Relationship between catch distribution of Portunid crab (*Charybdis bimaculata*) and environmental factors based on three species distribution models in Haizhou Bay. *J. Fish. China* **2018**, *42*, 889–901.
4. Jiang, Y.; Zhang, Y.L.; Pang, Z.W. Spatial distribution characteristics of *Sepia esculenta* in haizhou bay and adjacent waters and their relationship with environmental factors. *Acta Hydrobiol. Sin.* **2024**, *48*, 617–624.
5. Planque, B.; Loots, C.; Petitgas, P.; Lindstrøm, U.L.F.; Vaz, S. Understanding what controls the spatial distribution of fish populations using a multi-model approach. *Fish. Oceanogr.* **2011**, *20*, 1–17. [[CrossRef](#)]
6. Chen, X.J. *Fishery Resources and Fishery Oceanography*; China Ocean Press: Beijing, China, 2014; pp. 152–161.
7. Guo, Y.L.; Zhao, Z.F.; Qiao, H.J. Challenges and development trend of species distribution model. *Adv. Earth Sci.* **2020**, *35*, 1292–1305.
8. Zhu, W.B.; Zhu, H.C.; Zhang, Y.Z. Quantitative distribution of juvenile *Engraulis japonicus* and the relationship with environmental factors along the Zhejiang coast. *J. Fish. Sci. China* **2021**, *28*, 1175–1183.
9. Feng, B.; Chen, X.J.; Xu, L.X. Catch rate analysis of yellowfin tuna from longline fishery using generalized linear model in the Indian Ocean. *J. Fish. Sci. China* **2009**, *16*, 282–288.



10. Li, Z.G.; Wan, R.; Ye, Z.J. Use of random forests and support vector machines to improve annual egg production estimation. *Fish. Sci.* **2017**, *83*, 1–11. [[CrossRef](#)]
11. Haralabous, J.; Georgarakos, S. Artificial neural networks as a tool for species identification of fish schools. *ICES J. Mar. Sci.* **1996**, *53*, 173–180. [[CrossRef](#)]
12. Yang, S.L.; Zhang, Y.; Zhang, H. Comparison and analysis of different model algorithms for CPUE standardization in fishery. *Trans. Chin. Soc. Agric. Eng.* **2015**, *31*, 259–264.
13. Cui, Y.H.; Liu, S.D.; Zhang, Y.L. Habitat characteristics of Octopus ocellatus and their relationship with environmental factors during spring in Haizhou Bay, China. *Chin. J. Appl. Ecol.* **2022**, *33*, 1686–1692.
14. Xu, M.Z.; Zhang, C.L.; Xue, Y. Relationship between species diversity and environmental factors in the fishery community of Shandong coastal waters. *J. Fish. China* **2022**, *46*, 1008–1017.
15. Chen, X.Z.; Fan, W.; Cui, X.S. Fishing ground forecasting of Thunnus alalung in Indian Ocean based on random forest. *Acta Oceanol. Sin.* **2013**, *35*, 158–164. (In Chinese)
16. Hou, J.; Zhou, W.F.; Fan, W. Research on fishing grounds forecasting models of albacore tuna based on ensemble learning in South Pacific. *South China Fish. Sci.* **2020**, *16*, 42–50.
17. Gao, F. *Fishing Ground Forecasting of Chub Mackerel in the East China Sea and Yellow Sea Using Boosted Regression Trees*; Shanghai Ocean University: Shanghai, China, 2016.
18. Song, L.M.; Ren, S.Y.; Zhang, M. Fishing ground forecasting of bigeye tuna (*Thunnus obesus*) in the tropical waters of Atlantic Ocean based on ensemble learning. *J. Fish. China* **2023**, *47*, 64–76.
19. General Administration of Quality Supervision; Inspection and Quarantine of the People's Republic of China; Standardization Administration of China. *Specification for Marine Survey-Part 6: Marine Biological Survey*; Standards Press of China: Beijing, China, 2008.
20. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
21. Fang, K.N.; Wu, J.B.; Zhu, J.P. A Review of Technologies on Random Forests. *Stat. Inf. Forum* **2011**, *26*, 32–38.
22. Dong, S.S.; Huang, Z.X. A brief theoretical overview of Random Forests. *J. Integr. Technol.* **2013**, *2*, 1–7.
23. Zhang, D. *Vehicle Logo Recognition Based on Convolutional Neural Network and K-Nearest Neighbor*; Xidian University: Xi'an, China, 2015.
24. Li, X.; Zhang, C.L. Analysis of LocalLDtree classification model based on K proximity algorithm. *Silicon Val.* **2013**, *6*, 33–146.
25. Ming, Y.S. Using clustering to improve the KNN-based classifiers for online anomaly network traffic identification. *J. Netw. Comput. Appl.* **2011**, *34*, 722–730.
26. Malunoud, M.; Chokri, B.A. Classification improvement of local feature vectors over the KNN algorithm. *Multimed. Tools Appl.* **2013**, *64*, 197–218.
27. Jagan, S.; Hanan, S.; Amitabh, V. A fast all nearest neighbor algorithm for applications involving large point clouds. *Comput. Graph.* **2007**, *31*, 157–174.
28. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
29. Gao, J.X.; Zhang, W.; Gao, M. Material calculation time prediction model based on gradient boosting decision trees. *Softw. Guide* **2024**, *23*, 15–20.
30. Zhu, Y.L.; Feng, X.Y.; Yan, Q.G. Spatial distribution and main controlling factors of soil organic carbon under cultivated land based on GBDT model in black soil region of Northeast China. *China Environ. Sci.* **2024**, *44*, 1407–1417. [[CrossRef](#)]
31. Kabacoff, R. R in Action: Data Analysis and Graphics with R. *Greenwich Manning Publ.* **2011**, *8*, 126–131.
32. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [[CrossRef](#)]
33. Li, J.W.; Chen, C.H.; Sun, Y. Total partial regression sum of squares method for variable selection in multiple linear regression models. *J. Math. Med.* **2007**, *20*, 126–127.
34. Xu, B.D.; Zhang, C.L.; Xue, Y. Optimization of sampling effort for a fishery-independent survey with multiple goals. *Environ. Monit. Assess.* **2015**, *187*, 252. [[CrossRef](#)]
35. Shim, J.S.; Kim, R.K.; Yoon, K.B. A basic research for the development of habitat suitability index model of Pelophylax chosonicus. *J. Korean Soc. Environ. Restor. Technol.* **2020**, *23*, 49–62.
36. Tian, S.Q.; Chen, X.J.; Chen, Y. Evaluating habitat suitability indices derived from CPUE and fishing effort data for Ommatrephes bratramii in the northwestern Pacific Ocean. *Fish. Res.* **2009**, *95*, 181–188. [[CrossRef](#)]
37. Gong, C.X.; Chen, X.J.; Gao, F. Review on habitat suitability index in fishery science. *J. Shanghai Ocean. Univ.* **2011**, *20*, 260–269.
38. Chen, F.; Li, N.; Fang, Z. Habitat distribution change pattern of Uroteuthis edulis during spring and summer in the coastal waters of Zhejiang Province. *J. Shanghai Ocean. Univ.* **2021**, *30*, 847–855.
39. Tanaka, K.; Chen, Y. Spatiotemporal variability of suitable habitat for American lobster (*Homarus americanus*) Long Island Sound. *J. Shellfish. Res.* **2015**, *34*, 531–543. [[CrossRef](#)]
40. Zhang, M.; Wang, X.H.; Cai, Y.C. Spatial aggregation and dispersion characteristics of Trichiurus haumela in the Beibu Gulf, northern South China Sea. *J. Fish. Sci. China* **2022**, *29*, 1647–1658.
41. Liu, X.Y. *Study on the Impacts of Climate Change on the Potential Suitable Habitat of Major Commercial Fish in Offshore China*; Zhejiang Ocean University: Zhoushan, China, 2022.
42. Liu, J.C.; Jia, M.X.; Feng, W.D. Spatial Temporal Distribution of Antarctic Krill (*Euphausia superba*) Resource and its Association with Environment Factors Revealed with RF and GAM Models. *Period. Ocean. Univ. China (Nat. Sci. Ed.)* **2021**, *51*, 20–29.
43. Zhang, K.; Zhang, J.; Li, J.J.; Liao, B.C. Model selection for fish growth patterns based on a Bayesian approach: A case study of five freshwater fish species. *Aquat. Living Resour.* **2020**, *33*, 17. [[CrossRef](#)]

44. Prchalova, M.; Kubecka, J.; Vašek, M. Distribution patterns of fishes in a canyon shaped reservoir. *J. Fish Biol.* **2008**, *73*, 54–78. [[CrossRef](#)]
45. Zou, Y.Y.; Xue, Y.; Ma, Q.Y. Spatial distribution of *Larimichthys polyactis* in Haizhou Bay Based on Habitat Suitability Index. *Period. Ocean. Univ. China (Nat. Sci. Ed.)* **2016**, *46*, 54–63.
46. Wang, Y.Z.; Jia, X.P.; Lin, Z.J. Responses of *Trichiurus japonicus* catches to fishing and climate variability in the East China Sea. *J. Fish. China* **2011**, *35*, 1881–1889.
47. Wang, Y.Z.; Qiu, Y.S. An analysis of interannual variations of hairtail catches in East China Sea. *South China Fish. Sci.* **2006**, *2*, 16–24.
48. Yuan, X.W.; Liu, Z.L.; Jin, Y. Inter-decadal variation of spatial aggregation of *Trichiurus japonicus* in East China Sea based on spatial autocorrelation analysis. *Chin. J. Appl. Ecol.* **2017**, *28*, 3409–3416. (In Chinese)
49. You, H.B.; Xu, R. Relationship between central fishing ground and water temperature and salinity in summer season. *Mar. Fish.* **1984**, *4*, 165–167.
50. Dai, L.B.; Chen, J.H.; Tian, S.Q. Prediction of fish species richness in the Yangtze River estuary using CART algorithm. *J. Fish. Sci. China* **2018**, *25*, 1082–1090. [[CrossRef](#)]
51. Zhu, D.K.; Yu, C.G. The relation on the environment of fishing ground with the occurrence of hairtail in winter off the middle part of Zhejiang. *J. Fish. Sci. China* **1987**, *11*, 195–203.
52. Wang, T.Z.; Han, Q.; Luo, N.J. Catches of several major demersal fish species catches inhabiting Zhejiangsea area and their relationships with main influencing factors. *Trans. Oceanol. Limnol.* **2021**, *43*, 77–85.
53. Azevedo, M.; Araújo, F.; Cruz-Filho, A.; Pessanha, A.; Silva, M.; Guedes, A. Demersal fishes in a tropical bay in southeastern Brazil: Partitioning the spatial, temporal and environmental components of ecological variation. *Estuar. Coast. Shelf Sci.* **2007**, *75*, 468–480. [[CrossRef](#)]
54. Hu, C.L.; Zhang, H.L.; Zhang, Y.Z. Fish community structure and its relationship with environment factors in the Nature Reserve of *Trichiurus japonicus*. *J. Fish. China* **2018**, *42*, 694–703.
55. Zhang, Y.Q. *Environmental Impact on the Fish Assemblage Structures Dissertation Submitted to in Adjacent Sea Area of the Yangtze River Estuary*; Graduate School of Oceanology, Chinese Academy of Sciences: Qingdao, China, 2012.
56. Yan, L.P.; Liu, Z.L.; Li, S.F. Effects of new summer close season of trawl fisheries on fishery ecology and resource enhancement in East China Sea. *Mar. Fish.* **2010**, *32*, 186–191.
57. Liao, B.; Ehsanul, K.; Kui, Z. Comparative performance of catch-based and surplus production models on evaluating largehead hairtail (*Trichiurus lepturus*) fishery in the East China Sea. *Reg. Stud. Mar. Sci.* **2021**, *48*, 102026. [[CrossRef](#)]

**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.