

Article

Relationship between *Engraulis japonicus* Resources and Environmental Factors Based on Multi-Model Comparison in Offshore Waters of Southern Zhejiang, China

Wen Ma ¹, Chunxia Gao ^{1,2}, Wei Tang ¹, Song Qin ³, Jin Ma ^{1,2} and Jing Zhao ^{4,*}

- ¹ College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China; m200200694@st.shou.edu.cn (W.M.); cxgao@shou.edu.cn (C.G.); m200210753@st.shou.edu.cn (W.T.); jma@shou.edu.cn (J.M.)
² Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai 201306, China
³ Zhejiang Mariculture Research Institute, Wenzhou 325005, China; qinsong218@gmail.com
⁴ Institute of Marine Science, Shanghai Ocean University, Shanghai 201306, China
* Correspondence: jzhao@shou.edu.cn

Abstract: In order to accurately explore the relationship between the density of *Engraulis japonicus* and environmental factors, five types of models, including Tweedie-Generalized Additive Model (GAM), two-stage GAM, Ad hoc-GAM, and Generalized Additive Mixing Model (GAMM), were compared based on the survey data in offshore waters of southern Zhejiang, China from 2015 to 2021 in this study. The results showed the best goodness of fit for two-stage GAM when processing the data of *E. japonicus* resource density. The deviance explained of GAM1 and GAM2 were 19.9 and 53.8%, respectively. According to this study, water temperature and salinity are important environmental factors affecting the distribution of *E. japonicus*, which are also closely related to latitude. In general, the resource density of *E. japonicus* decreases gradually with the increase in water temperature. When the salinity was between 26 ppt and 34 ppt, the resource density was higher. Also, there were some differences in the spatial distribution of *E. japonicus* in different seasons. The relationship between the resource density of *E. japonicus* and environmental factors was analyzed through various models to provide a scientific basis for the conservation management of *E. japonicus* in offshore waters of southern Zhejiang, China.

Keywords: spatio-temporal distribution; Generalized Additive Model; zero values; cross-validation



Citation: Ma, W.; Gao, C.; Tang, W.; Qin, S.; Ma, J.; Zhao, J. Relationship between *Engraulis japonicus* Resources and Environmental Factors Based on Multi-Model Comparison in Offshore Waters of Southern Zhejiang, China. *J. Mar. Sci. Eng.* **2022**, *10*, 657. <https://doi.org/10.3390/jmse10050657>

Academic Editors: Ming-An Lee, Meng-Hsien Chen and Yu-San Han

Received: 27 April 2022

Accepted: 10 May 2022

Published: 12 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 relationship between fishery resources and the marine environment is very complex, with non-linear and non-additive relationships [1]. Therefore, it is important to select appropriate methods for quantitative analysis of the relationship between fishery resources and the marine environment. Previous studies [2–4] indicated that changes in the stock spatial structure might cause due to changes in environmental conditions. Thus, understanding the relationship between species distribution and habitat can provide necessary information for predicting the impact of climate change and formulating relevant management strategies [5].

Species Distribution Models (SDMs) are effective tools to study the relationship between research objects and habitat environmental factors as well as to explore the spatio-temporal distribution of fish [6]. SDMs can be divided into two types [7]: (1) “only presence” models that ignore the missing values and zero values, such as ecological niche factor analysis (ENFA) [8] or maximum entropy model (MaxEnt) [9], which can find habitats similar to species record sites based on the environmental conditions, and (2) “presence-absence” models, which need to record resources and efforts in the survey, such as Generalized

Additive Model (GAM), Generalized Linear Model (GLM), or Regression Trees Analyses. Many scholars use the presence-absence models to explore the relationship between fishery resource density and environmental factors [10–13]. As a kind of SDM, GAM can deal with the non-linear relationship between response variables and explanatory variables, such as spatial, temporal, and environmental variables [14], which comes with good stability and more flexibility for exploring the relationship between fishery resource density and environmental factors [15].

In the actual sampling process, due to sampling methods, selection of fishing gear, and small stock size, there may be a large number of zero values in fishery resource data causing certain difficulties in estimation of the model with log-normal distribution error [16]. Bouska et al. [17] showed that model selection was the main uncertainty factor in the establishment of SDMs. Therefore, selecting an appropriate SDM based on the actual survey data can effectively improve the accuracy and reliability of the model. In relevant modeling research in the fishery field, several methods are usually used to deal with a large number of zero values, such as Ad hoc-GAM, Generalized Additive Mixed Models (GAMM), two-stage GAM, and Tweedie-GAM. Ad hoc-GAM applies the model to data with zero values by adding a constant (the link function is a natural logarithm) [18]. GAMMs are very suitable for processing time-series data and have good performance in processing fishery data [4]. Two-stage GAM is a widely used data survey with a large number of zero values that ensures good results [3,10,19]. To better process data with a large number of zero values, Tweedie [20] also proposed a special distribution law called Tweedie distribution, which is suitable for processing non-negative data with a large number of zero values [16].

The offshore waters of southern Zhejiang are located in the East China Sea. Under the influence of water mass, such as coastal waters of Fujian and Zhejiang as well as Taiwan Warm Current, there are adequate nutrients and bait organisms supporting abundant fish resources [21]. However, after the 1990s, due to overfishing and water pollution, the main economic fish in the offshore waters rapidly declined, and fish species, such as *Engraulis japonicus*, gradually became the fishing target [21,22]. *E. japonicus* is a pelagic migratory fish with a strong clustering pattern. It is widely distributed in the East China Sea and Yellow Sea of China, and is the prey of many higher trophic level species [23], thus playing an important role in the ecosystem. In recent years, under the influence of water pollution and overfishing, the sampling zero value of *E. japonicus* frequently appeared in the resource survey and monitoring stations. Therefore, how to select an appropriate model method to explore the distribution mechanism of *E. japonicus* resources in this area has become an urgent scientific problem to be solved. Based on the independent survey data of offshore fisheries in southern Zhejiang from 2015 to 2021, this study explored the goodness of fit and prediction performance of Tweedie-GAM, two-stage GAM, Ad-hoc GAM, and GAMM in processing large amounts of zero value data and analyzed the relationship between *E. japonicus* and environmental factors. This gave a further understanding of the distribution pattern and the latest dynamics of *E. japonicus* resources that enhanced our understanding of the ecological mechanism of species distribution, which acted as the research reference for the conservation management and sustainable utilization of fishery resources.

2. Materials and Methods

2.1. Data Sources

From November 2015 to February 2021, a seasonal survey of fishery resources was conducted in the offshore waters of southern Zhejiang. Since no *E. japonicus* was found in the survey in summer (August) every year, we only used survey data in spring (May), autumn (November), and winter (February). The survey area ranged from 120.93° E to 122.95° E, 27.21° N to 28.97° N (Figure 1). The field survey was conducted during the day. *E. japonicus* mainly lives in the middle and lower to bottom waters during the day [24]. So, we used the bottom trawl for fishing. The total length of gear was 95 m. The fishing gear

was 40 m wide and 7.5 m high. The length of the bottom and floating substrates was 80 m. The mesh of the net capsule was 2 cm, and the towing speed was 2–4 kn. The operation time of each survey station was about 1 h. At each survey station, the water quality analyzer WTW-Multi 3430 was used to collect environmental data, such as water temperature and salinity. The collection, determination, and analysis of water quality samples were carried out in accordance with the Specifications for Oceanographic Survey (GB/T 12763) [25] and Specification for Marine Monitoring (GB 17378) [26]. According to the total catch and proportion of each station, the catch data were standardized according to trawl time (1 h) and trawl speed (3 kn).

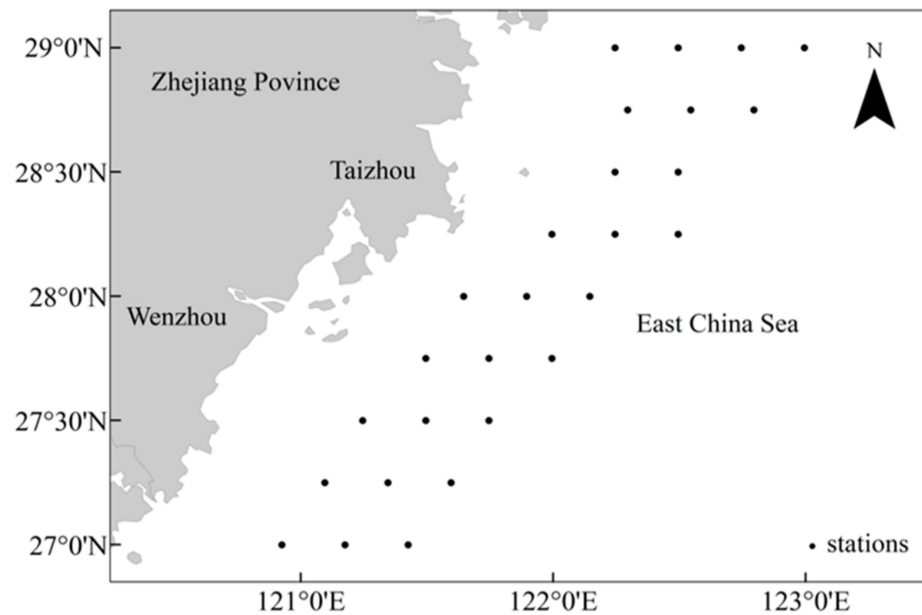


Figure 1. Distribution of sampling stations in offshore waters of southern Zhejiang.

2.2. Selection of Model Explanatory Variables

Given the obvious seasonal migration of *E. japonicus* during spawning, feeding, and winter migration [27], longitude, latitude, and offshore distance were selected as the influencing factors of the spatial distribution of *E. japonicus*. The offshore distance was the shortest distance from the survey station to the shore, which was calculated in the Sp package in R [28]. Water temperature, salinity, and water depth are closely related to the resource distribution of *E. Japonicus*, as it is a pelagic fish species [27,29]. A total of 7 factors (surface water temperature, surface salinity, water depth, month, longitude, latitude, and offshore distance) were selected to explore the relationship between *E. japonicus* resource distribution and environmental influencing factors. Before modeling, variance inflation factor (VIF) was used to test the multicollinearity of variables while excluding highly correlated explanatory variables with a VIF value greater than 5 [3,4].

2.3. Model Theory

In this study, five types of models (two-stage GAM, Tweedie-GAM, GAMM, and Ad hoc-GAM) were used to explore the relationship between the presence probability, resource density of *E. japonicus*, and environmental factors. In addition, the goodness of fit of the model to the data was measured by deviance explained (the proportion of the null deviance explained by the model).

GAM can fit the non-linear relationship between response variables and explanatory variables, and the expression is as follows:

$$Y = \alpha + \sum_{j=1}^P f_j(X_j) + \varepsilon \quad (1)$$

where Y denotes the resource density of *E. japonicus* (g/h); α denotes the intercept of the fitting model; refers to the smoothing function; a spline smoothing function was used in this study; is an independent variable; residual $\varepsilon = \sigma^2$ and $E(\varepsilon) = 0$.

2.3.1. Two-Stage GAM

The model has two stages: the first stage of the GAM estimates the presence probability (P) of *E. japonicus* with a binomial error distribution, and the second stage of the GAM estimates the log transformation abundance of the species a with a Gaussian error distribution [30]. The formula is given as follows:

$$\text{Logit}(P) = \text{month} + s(\text{Lat}) + s(\text{Lon}) + s(\text{depth}) + s(\text{Dis}) + s(T) + s(S) \quad (2)$$

The second stage GAM (GAM2), estimates the log-transformed *E. japonicus* resource density using the identity link function [3,10]. The two-stage GAM model formula is as follows:

$$\text{GAM1:Logit}(P) = \text{month} + s(\text{Lat}) + s(\text{Lon}) + s(\text{depth}) + s(\text{Dis}) + s(T) + s(S) + \varepsilon \quad (3)$$

$$\text{GAM2:ln}(\text{density}) = \text{month} + s(\text{Lat}) + s(\text{Lon}) + s(\text{depth}) + s(\text{Dis}) + s(T) + s(S) + \varepsilon \quad (4)$$

Combined with the results of GAM1 and GAM2, the final log-transformed *E. japonicus* density was estimated [19].

$$\ln(y) = \ln(P) + \ln(\text{density}) \quad (5)$$

where *month* denotes the month; *Lat* denotes latitude; *Lon* denotes longitude; *T* denotes water temperature; *S* denotes salinity; *depth* denotes water depth; *Dis* denotes offshore distance; density denotes *E. japonicus* resource density; P denotes the occurrence probability of *E. japonicus*; ε denotes a random error.

2.3.2. Tweedie GAM

Tweedie distribution was first proposed by a British statistician in 1984 [20], which is a special probability distribution in exponential dispersion distribution. It is usually expressed as $Tw_p(\theta, \varphi)$ and is determined by variance function $V(\mu) = \mu^p$, where θ is a standard parameter; φ is a dispersion parameter; $p \in (-\infty, 0) \cup [1, +\infty)$. Tweedie distribution includes several common important distributions. When $p = 0, 1, 2, 3$, it corresponds to a normal distribution, Poisson distribution, Gamma distribution, and inverse Gaussian distribution. When $1 < p < 2$, corresponding $Tw_p(\theta, \varphi)$ is a composite distribution between Poisson distribution and Gamma distribution [20]. The probability density equation is as follows [16]:

$$f(y : \theta, \varphi, p) = a(y : \varphi, p) \exp \left\{ -\frac{1}{2\sigma^2} d(y : \theta, p) \right\} \quad (6)$$

where θ is the location parameter; φ is the diffusion parameter; p is the energy parameter; and $d(y : \theta, p)$ is the unit deviation.

The Tweedie distribution was used to establish the relationship between *E. japonicus* and environmental factors. The Tweedie-GAM expression is as follows:

$$\text{density} = \text{month} + s(\text{Lat}) + s(\text{Lon}) + s(\text{depth}) + s(\text{Dis}) + s(T) + s(S) \quad (7)$$

where *month* denotes month; *Lat* denotes latitude; *Lon* denotes longitude; *T* denotes water temperature; *S* denotes salinity; *depth* denotes water depth; *Dis* denotes offshore distance; density denotes the *E. japonicus* resource density; ε denotes a random error.

2.3.3. GAMM

GAMM is an extension of GAM, including fixed and random effects, which is very suitable for processing time series and autocorrelation data [31]. In the southern sea area of Zhejiang, the distribution of *E. japonicus* changes with season [27]. In this study, season (*month*) was used as a random error term of GAMM. GAMM expression is as follows:

$$\text{GAMM: density} = s(\text{Lat}) + s(\text{Lon}) + s(\text{depth}) + s(\text{Dis}) + s(T) + s(S) + \varepsilon \quad (8)$$

$$\text{random} = \text{month} \quad (9)$$

where *month* denotes month; *Lat* denotes latitude; *Lon* denotes longitude; *T* denotes water temperature; *S* denotes salinity; *depth* denotes water depth; *Dis* denotes offshore distance; density denotes the *E. japonicus* resource density; ε denotes a random error.

2.3.4. Ad Hoc-GAM

Ad hoc-GAM refers to the addition of a constant *c* to the resource density so that the model can process data containing zero values. According to Tian et al. [18], constant *c* is usually 1, but some relevant studies [32] have stated that a constant selection of 10% of the average resource density can reduce the error. Therefore, 1 and 10% of the average resource density were selected as the constant *c* to compare which effect was better (named as Ad + 1 GAM and Ad hoc-mean GAM). After the addition of the constant, we carried out a logarithmicization process for the dependent variable, and the expression is as follows:

$$\ln(\text{density} + C) = \text{month} + s(\text{Lat}) + s(\text{Lon}) + s(\text{depth}) + s(\text{Dis}) + s(T) + s(S) + \varepsilon \quad (10)$$

where *month* denotes month; *Lat* denotes latitude; *Lon* denotes longitude; *T* denotes water temperature; *S* denotes salinity; *depth* denotes water depth; *Dis* denotes offshore distance; density denotes the *E. japonicus* resource density; ε denotes a random error.

2.4. Model Selection

Akaike Information Criterion (AIC) can be used to measure the goodness of fit of multiple models [33]. The smaller the AIC value, the better the fitting degree of the model. In this study, the GAM between environmental factors and resource density was established by permutation and a combination of variable factors after screening. The model with the smallest AIC value in each method was considered the optimal model.

The calculation method of AIC is as follows:

$$\text{AIC} = 2k - 2 \ln L \quad (11)$$

where *k* is the number of parameters, and *L* is the likelihood function.

2.5. Cross-Validation

In this study, 80% of the data were randomly selected as the training set and the remaining 20% as the test set. The above process was run 1000 times to verify the prediction performance of different models. During each cross-validation, a linear relationship between the model predicted value and the actual observed value was built, and the root mean square error (RMSE) and mean absolute error (MAE) between the predicted and observed values were calculated. Closer the value to zero, the better the model goodness of fit [34,35]. The linear relationship between the predicted and observed values can be expressed as follows:

$$\ln Y = a + b * \ln y \quad (12)$$

where *y* denotes the predicted value of the model; *Y* denotes the actual observed value of the model. When *a* = 0 and *b* = 1, it means that the predicted resource density and the actual observed resource density (i.e., test data) have a similar spatial pattern. The model

has a good prediction performance [19]. R^2 was used to express the prediction effect of the model. When R^2 is closer to 1, the prediction effect of the model becomes better [3].

The calculation equation of RMSE is as follows [34]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (13)$$

The calculation equation of MAE is as follows [35]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (14)$$

where n denotes the observation frequency; O_i denotes the i th observed value; P_i denotes the i th predicted value.

2.6. Comparison of Models

Based on the results of cross-validation, the prediction effects of the 5 models were compared to select the optimal model for processing zero value data.

All statistical analyses were carried out in R (V3.6.0), and the model was implemented through the “mgcv” package. The distribution of *E. japonicus* resource density and stations were drawn in Arcmap 10.8.

3. Results

3.1. Zero Value Ratio of *E. japonicus*

From 2015 to 2021, the sampling zero-value of *E. japonicus* accounted for the largest proportion (90.5%) in autumn and the smallest proportion (46.1%) in spring. In spring, the zero value ratio of *E. japonicus* was mainly concentrated in 40–60%, with an average of $44.6 \pm 23.7\%$. In autumn, the zero value ratio of *E. japonicus* was mainly concentrated within 75–100%, with an average of $90.6 \pm 12.7\%$. In winter, the zero value ratio of *E. japonicus* was mainly concentrated within 60–80%, with an average of $74.6 \pm 17.7\%$. The sampling zero value ratio in three seasons was ranked from highest to lowest as follows: autumn > winter > spring (Figure 2).

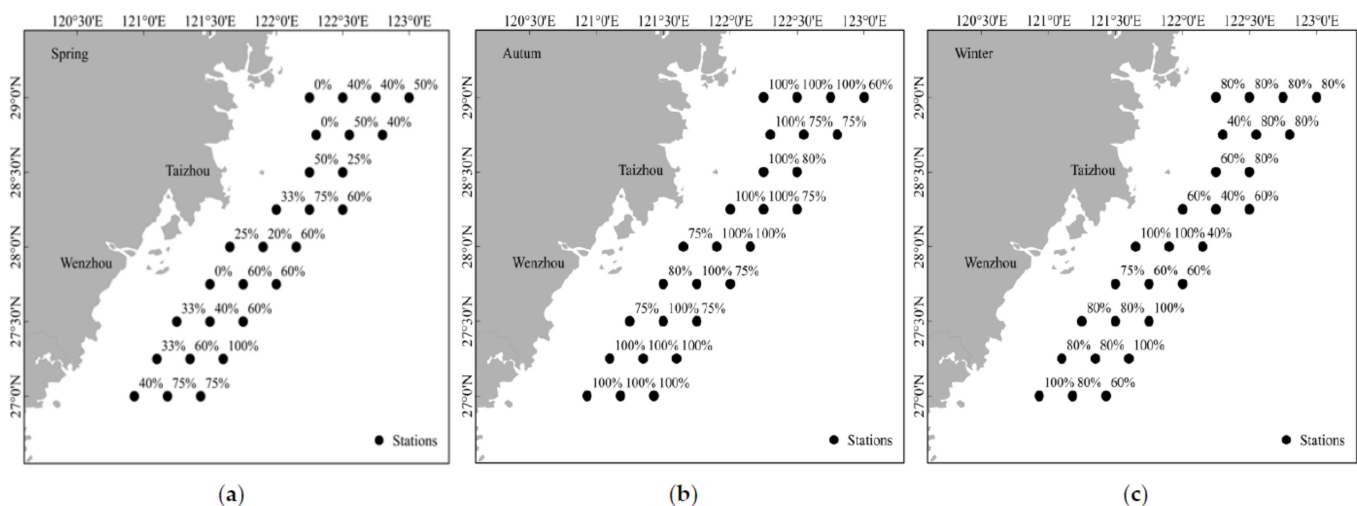


Figure 2. Zero value ratio of *E. japonicus* in offshore waters of southern Zhejiang in spring, autumn, and winter from 2015 to 2021: (a) Spring; (b) Autumn; and (c) Winter.

3.2. Spatial and Temporal Distribution of *E. japonicus* Resource Density

The distribution of *E. japonicus* resources in offshore southern Zhejiang showed obvious spatial differences in different seasons (Figure 3). The distribution characteristics of resource density were opposite to the zero value ratio of *E. japonicus* (spring > winter

> autumn). In spring, the *E. japonicus* in offshore waters was higher than that of the open waters, while it was the opposite in autumn, presenting a distribution pattern of higher resource density in open waters than in offshore waters. In winter, there was a significant north-south difference in the study waters, and *E. japonicus* mainly concentrated in the northern waters of 28° N.

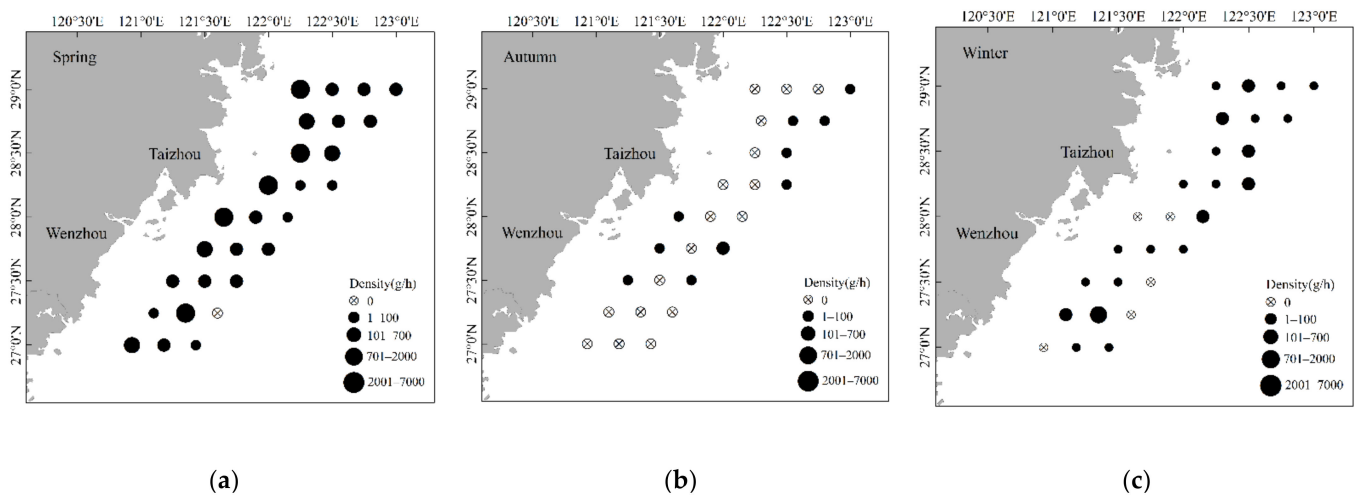


Figure 3. Distribution of *E. japonicus* in offshore waters of southern Zhejiang in spring, autumn, and winter from 2015 to 2021: (a) Spring; (b) Autumn; and (c) Winter.

3.3. Results of Different Models

VIF test results showed that VIFs of water temperature, water depth, salinity, and offshore distance were all less than 5. The VIFs of longitude and latitude were greater than 5, and the VIF of longitude was the largest. After removing the longitude, the collinearity test was conducted again for the influencing factors, and VIFs were less than 5 (Table 1). Therefore, six influencing factors, including month, water temperature, salinity, water depth, offshore distance, and latitude, were adopted to establish the model.

Table 1. Collinearity test for predictor variables.

Factor	T	S	Depth	Dis	Lat	Lon
VIF	1.08	1.56	3.76	4.14	31.8	38.8
	1.08	1.56	2.78	2.18	1.17	-

Note: the “-” denotes removing this factor.

In terms of two-stage GAM, GAM1 consisting of month, latitude, water temperature, salinity, and water depth was the optimal model at this stage, with a deviance explained of 19.9%. Latitude and water temperature were significantly correlated with the occurrence probability of *E. japonicus* ($p < 0.05$, Table 2). GAM2 consisting of month, latitude, water temperature, salinity, and water depth, was the optimal model at this stage, with a deviance explained of 53.8%. Water temperature was significantly correlated with the occurrence probability of *E. japonicus* ($p < 0.05$, Table 2).

The optimal variable combination of Tweedie-GAM was month, water temperature, and salinity, and the model deviance explained was 46.7%, among which salinity was found to be a significant influencing factor ($p < 0.001$, Table 2).

All explanatory variables were included in the optimal GAMM, and the deviance explained of the model was 73.2%. All five influencing factors were significantly correlated with the resource density of *E. japonicus* ($p < 0.001$, Table 2).

Table 2. Model selection results of the five optimal models.

Model		Optimal Model	Degrees of Freedom	<i>p</i>	AIC	Deviance Explained
Two-stage GAM	GAM 1	latitude	1.001	0.014 *	375.90	19.9%
		temperature	2.076	0.009 **		
		salinity	1.108	0.141		
		depth	2.749	0.074		
		month	-	-		
	GAM 2	latitude	1.000	0.185	445.35	53.8%
		temperature	7.890	0.037 *		
		salinity	6.282	0.136		
		distance	5.646	0.229		
		month	-	-		
Tweedie-GAM		temperature	1.480	0.112	19,870.02	46.7%
		salinity	7.903	<0.001 ***		
		month	-	-		
GAMM		latitude	7.997	<0.001 ***	202,817.3	73.2%
		temperature	8.978	<0.001 ***		
		salinity	8.998	<0.001 ***		
		distance	8.997	<0.001 ***		
		depth	8.997	<0.001 ***		
		month	-	-		
Ad hoc-GAM	Ad + 1 GAM	latitude	1.000	0.009 **	1641.45	30%
		temperature	2.098	0.030 *		
		salinity	7.293	0.104		
		month	-	-		
	Ad hoc-mean GAM	temperature	3.621	0.045 *	1120.89	29.6%
		salinity	7.452	0.057		
		latitude	1.000	0.033 *		
		depth	1.000	0.092		
		month	-	-		

Note: * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$.

The variable combination of optimal Ad + 1 GAM included all factors except water depth and offshore distance, and the deviance explained of the model was 30.0%. Latitude and water temperature were significantly correlated with the resource density of *E. japonicus* ($p < 0.01$ and $p < 0.05$, respectively; Table 2). For Ad hoc-mean GAM, the optimal variable combination was month, water temperature, salinity, latitude, and water depth, with a deviance explained of 29.6%. Water temperature and latitude were significantly correlated with the resource density of *E. japonicus* ($p < 0.05$, Table 2).

3.4. Relationship between *E. japonicus* Resource Density and Environmental Factors

There was a non-linear relationship between water temperature and resource density as well as occurrence probability (Figures 4a, 5a, 6a, 7a and 8a). In GAM1 (Figure 4a), Ad + 1 GAM (Figure 7a) and Ad hoc-mean GAM (Figure 8a) showed a negative correlation between water temperature and occurrence probability as well as resource density of *E. japonicus*. In GAM2, the relationship between water temperature and resource density was non-linear, with multiple peaks, showing an overall negative correlation (Figure 5a). Compared to other models, the relationship between water temperature and *E. japonicus* resource density was significantly different in GAMM. The resource density of *E. japonicus* was the lowest at 16.5 °C and the highest at 25 °C (Figure 6a).

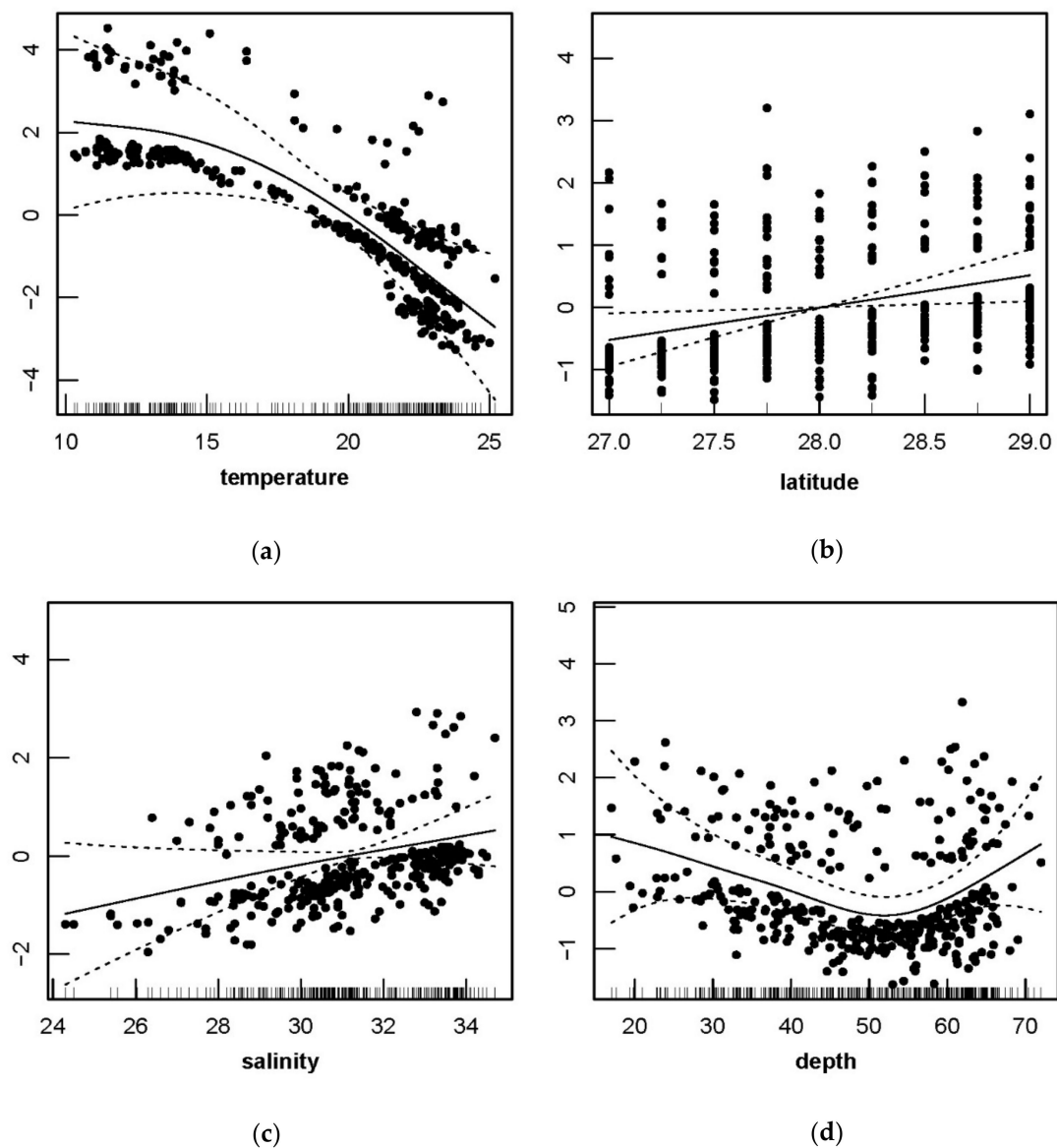


Figure 4. Response curves for variables of GAM 1 in two-stage GAM: (a) effect of temperature; (b) effect of latitude; (c) effect of salinity; and (d) effect of depth. Dashed lines show 95% confidence intervals. The dots represent the residuals.

When the salinity ranged from 24 ppt to 34.5 ppt, the resource density of *E. japonicus* increased significantly with the increase in salinity (Figures 6b and 9b). When salinity reached 26 ppt, it showed a multi-wave non-linear relationship, which was at a high level.

According to GAM1 (Figure 4b), Ad + 1 GAM (Figure 7b), and Ad hoc-mean GAM (Figure 8b), there was a positive linear correlation between latitude and resource density as well as the occurrence probability of *E. japonicus*. However, in GAMM (Figure 6d), the relationship between latitude and *E. japonicus* became very complicated.

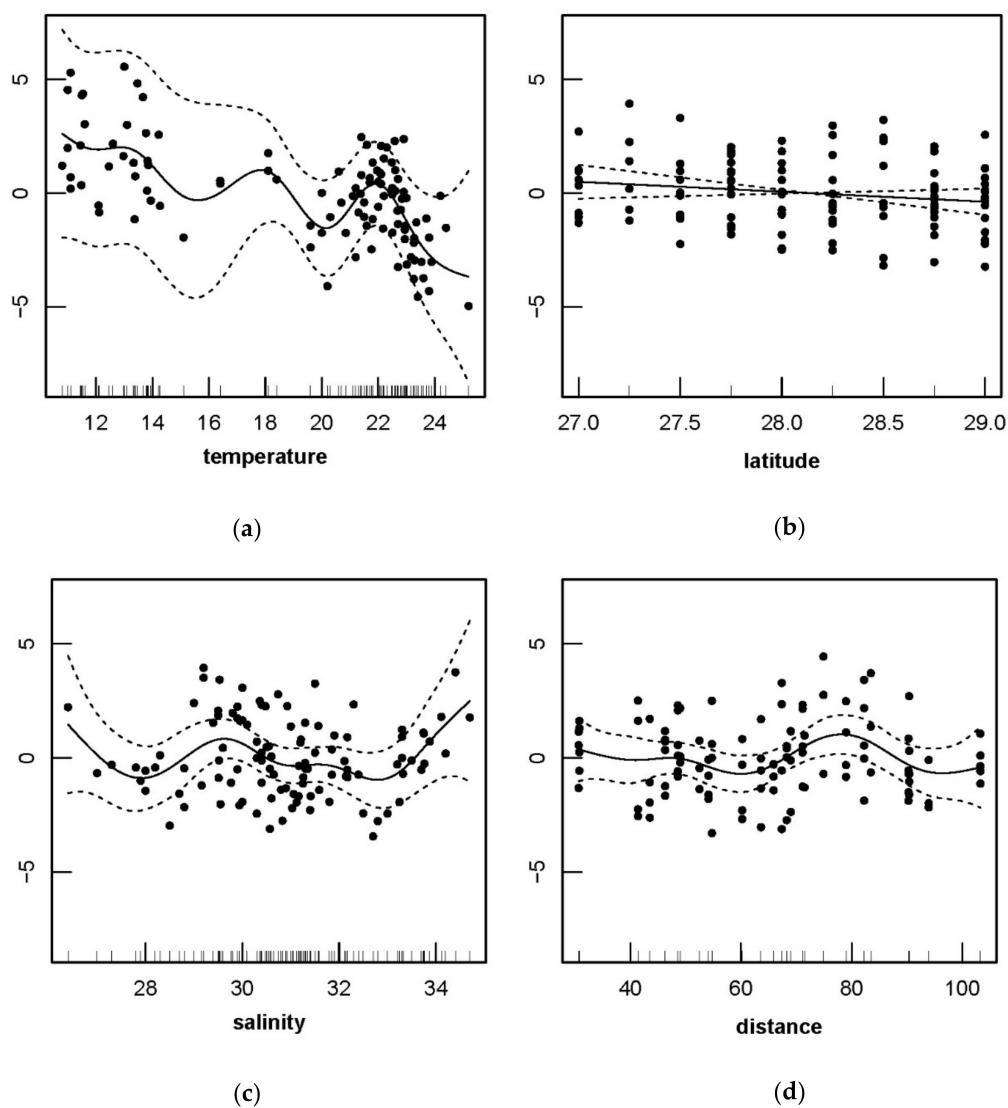


Figure 5. Response curves for variables of GAM 2 in two-stage GAM: (a) effect of temperature; (b) effect of latitude; (c) effect of salinity; and (d) effect of distance. Dashed lines show 95% confidence intervals. The dots represent the residuals.

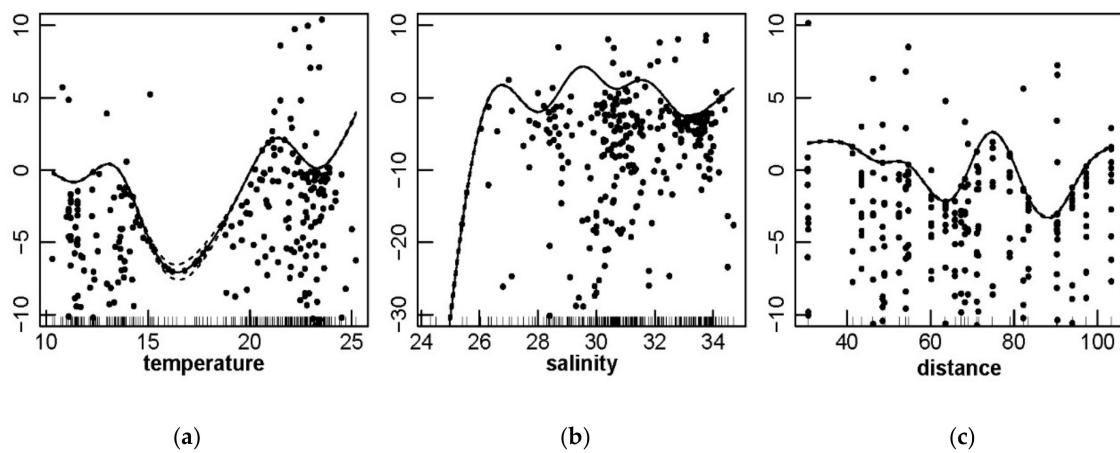


Figure 6. Cont.

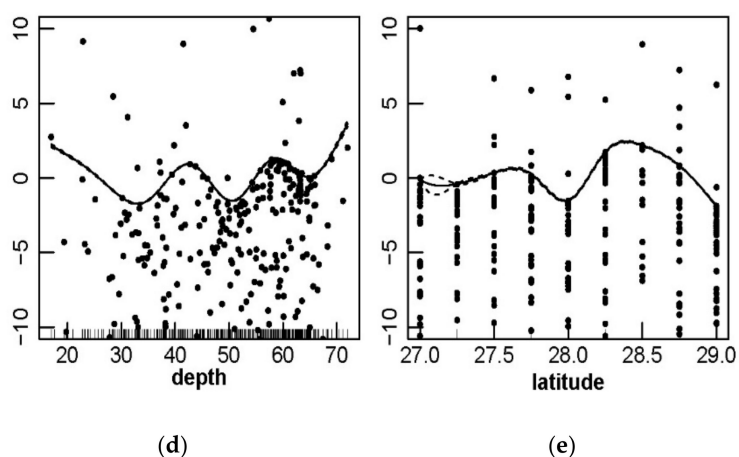


Figure 6. Response curves for variables of GAMM: (a) effect of temperature; (b) effect of salinity; (c) effect of distance; (d) effect of depth; and (e) effect of latitude. Dashed lines show 95% confidence intervals. The dots represent the residuals.

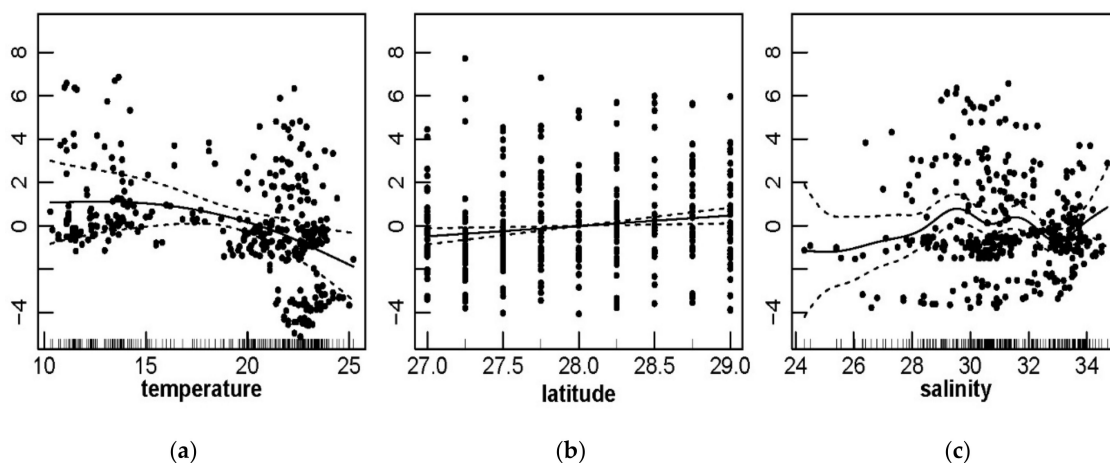


Figure 7. Response curves for variables of Ad + 1 GAM: (a) effect of temperature; (b) effect of latitude and (c) effect of salinity. Dashed lines show 95% confidence intervals. The dots represent the residuals.

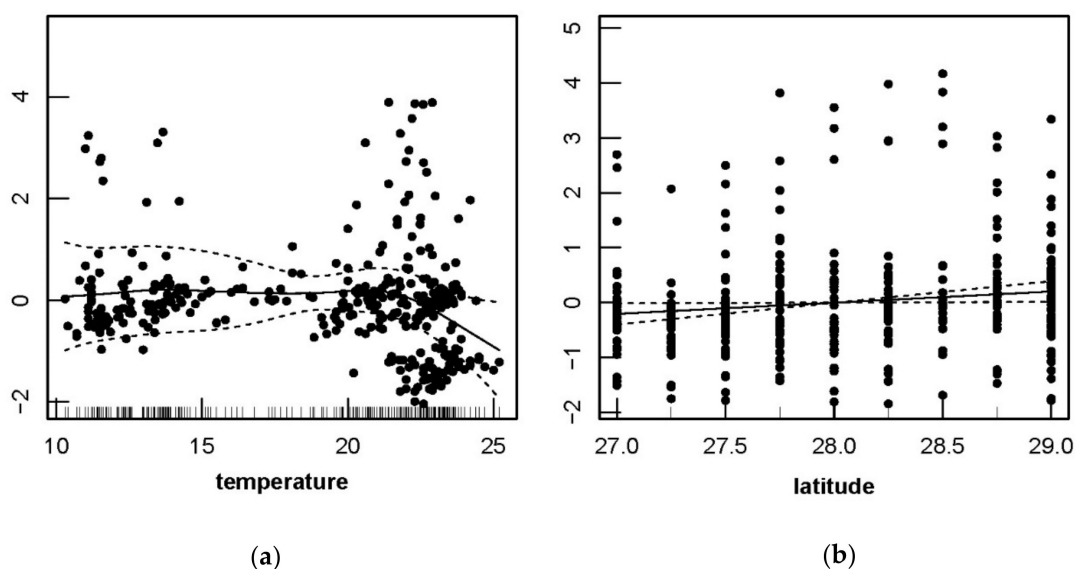


Figure 8. Cont.

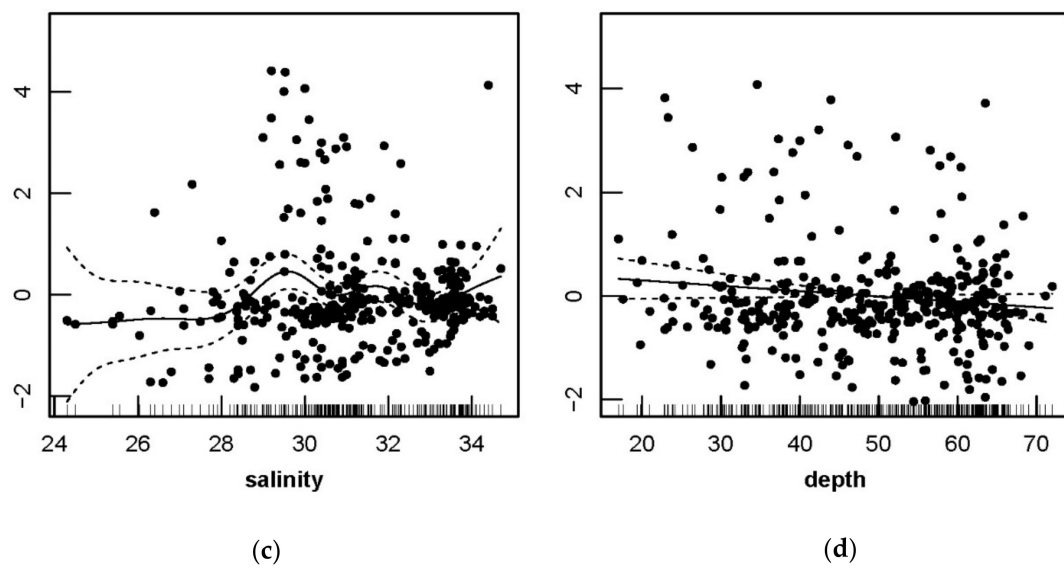


Figure 8. Response curves for variables of Ad hoc-mean GAM: (a) effect of temperature; (b) effect of latitude; (c) effect of salinity; and (d) effect of depth. Dashed lines show 95% confidence intervals. The dots represent the residuals.

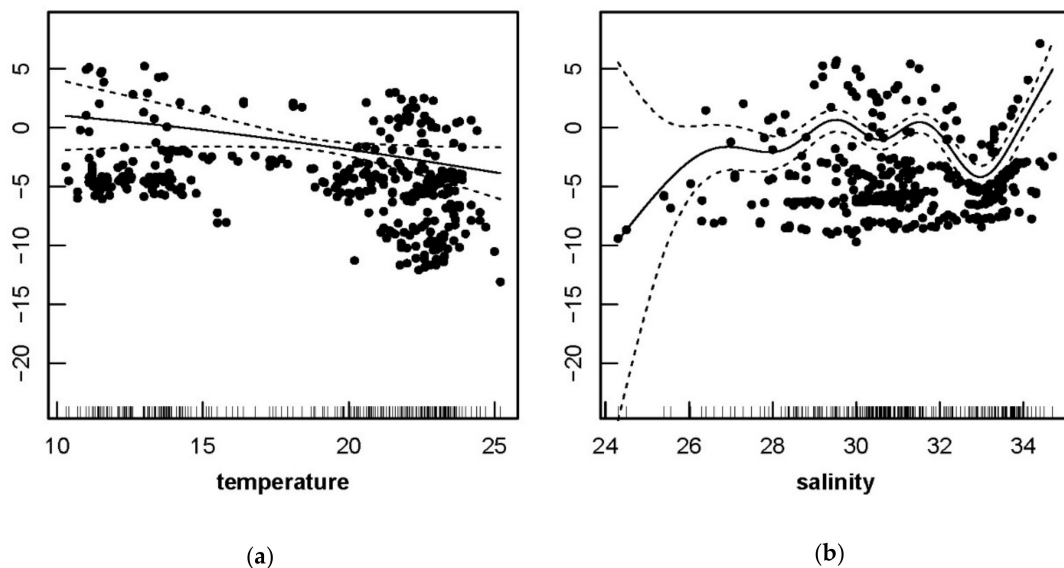


Figure 9. Response curves for variables of Tweedie-GAM: (a) effect of temperature; and (b) effect of salinity. Dashed lines show 95% confidence intervals. The dots represent the residuals.

3.5. Prediction Performance of the Model

The results of 1000 cross-validation showed that the R^2 of Ad + 1 GAM fitting curve was the largest, and that of GAMM was the smallest among the five models. Two-stage GAM had the smallest RMSE and MAE and a relatively large R^2 of 0.18 (Table 3). Therefore, comprehensively considering MAE, RMSE, and R^2 , two-stage GAM should be selected as the optimal model to predict the resource density of *E. japonicus*.

Table 3. 1000 cross validation results of different models.

Model	RMSE	MAE	R ²
Two-stage GAM	1324	274	0.18
Tweedie-GAM	1725	389	0.14
GAMM	1697	361	0.10
Ad + 1 GAM	1709	381	0.24
Ad hoc-mean GAM	1816	467	0.17

4. Discussion

4.1. Comparison between Different Models

Due to the decline in fishery resources, the patch aggregation of fish, and the selection of fishing gear for survey [36], it is impossible to effectively capture the target species in the fishery resource survey, thus resulting in many zero values in the survey data. In this study, zero value data accounted for 70% of the total data, which did not follow the general data distribution pattern (positively skewed distribution). Five models commonly used to deal with zero value data were used to establish the relationship between the resource density data of *E. japonicus* and environmental factors. It was found that two-stage GAM had more advantages in processing the data of *E. japonicus* resources. There is a certain difference in deviance explained of GAM1 and GAM2. The deviance explained of GAM1 was 19.2%, which was relatively lower, while that of GAM2 was 53.4%, and the model goodness of fit was good. However, there might be other important factors affecting the presence of *E. japonicus*. *E. japonicus* has obvious clustering characteristics, and waters with higher resource density represent the suitable living environment for this species to some extent, which makes the deviance explained of GAM2 relatively higher. In future studies, if biological factors (e.g., bait organisms) [37] and species interactions (e.g., predation and prey) [38] can be included in the model as explanatory variables, it will be conducive to model goodness of fit. The results of 1000 cross-validation showed that R² of the fitting curve of Ad + 1 GAM predicted and measured which values were the largest. While two-stage GAM had the smallest RMSE and MAE, the R² was relatively large, indicating that it had a better effect on the processing data of *E. japonicus* resource data in the waters of southern Zhejiang. Although two-stage GAM is considered more suitable for processing the density data of *E. japonicus* resources as compared to other models, its prediction effect is not particularly good, which may be due to fewer influencing factors discussed in this study and the formation of fishing ground associated with the spatio-temporal distribution structure of environmental factors. It is difficult to dynamically parameterize such spatio-temporal correlations [39]. Therefore, the survey frequency of *E. japonicus* migration during flood season and the collection of marine environmental data can be increased in future studies to better match the data collection with the living habitats of *E. japonicus* and improve the prediction ability of the model.

4.2. Relationship between *E. japonicus* Resource Density and Environmental Factors

The current study analyzed the relationship between different environmental factors and *E. japonicus* resources using different models. Although the goodness of fits and prediction performance of each model was different in comparative analysis, except for the Tweedie-GAM excluding the latitude, water temperature, salinity, and latitude, all existed in these models. Therefore, water temperature, salinity, and latitude may play an important role in the distribution of *E. japonicus* in southern Zhejiang.

Relevant studies have shown that water temperature can dominate the growth, development, and reproduction of fish [21] and affect the entire food web structure of fish by participating in the regulation of primary viability [40]. In this study, the relationship between water temperature and *E. japonicus* resources was similar, but there were slight differences. After careful consideration, 10–11 °C is considered the optimum water temperature range for *E. japonicus*. The range for optimum water temperature of *E. japonicus* is different from other relevant studies [27,41], where 8–11 °C is the optimum temperature

range for *E. japonicus*. Niu et al. [42] have reported that the optimum temperature is different due to the different temperature ranges of fish in their different life stages, and the stock size, age structure, and fishing situation also affect the distribution of *E. japonicus*.

Salinity is one of the important environmental factors for fish development and distribution. It can change the fish stock by changing the osmotic pressure of fish eggs and affecting the development of embryos [14,43,44]. However, few studies have focused on how salinity affects the presence and distribution of *E. japonicus*. In this study, the effect of salinity on the resource distribution of *E. japonicus* showed a multi-wave non-linear relationship (Figure 5c). Previous studies have shown that the salinity of offshore waters of southern Zhejiang was due to the influence of the coastal current with cold water and low salinity, and Taiwan warm current with warm water that shows higher distribution characteristics in the east and lower distribution characteristics in the west [45,46]. As a kind of migratory fish [27], there is a big difference in the spatial distribution characteristics of *E. japonicus* in different seasons, which even showed a contrary distribution pattern in this study. Therefore, the relationship between salinity and *E. japonicus* is complex and variable, which may explain the multi-wave non-linear relationship between the resource density of *E. japonicus* and salinity.

Latitude, as a spatial factor, plays an important role in the resource density of *E. japonicus* in the offshore waters of southern Zhejiang. It has an indirect effect on the distribution of *E. japonicus*, by changing other environmental factors, such as temperature and salinity [47]. In this study, GAM1 (Figure 4b) showed a positive linear correlation, while GAM2 showed a negative linear relationship (Figure 5b). Every spring, with the gradual increase in water temperature, the wintering stock located at lower latitudes gradually leaves the wintering grounds for waters at higher latitudes for breeding migration under the effect of gonad maturation. In autumn, influenced by the decreasing water temperature, *E. japonicus* began to migrate to the southern waters with higher water temperature [45], which migrates between high and low latitudes. Latitude represents water temperature to a certain extent. Thus, it can affect the resource density of *E. japonicus* stock.

4.3. Importance of Sampling

The relationship between fishery resource density and environmental factors and the spatio-temporal distribution of target species are affected by time, space, and fishing methods. *E. japonicus* is a small pelagic fish [37]. However, bottom trawl nets were used to investigate the depth of the water layer, which was not fully matched with the habitat water layer of *E. japonicus*, leading to the occurrence of a large number of zero values. In addition, spawning migration and wintering migration of *E. japonicus* are found in different seasons [42]. In this study, samples were taken in quarters with a longer time scale, thus weakening the importance of the *E. japonicus* migration process. In recent years, under the influence of overfishing, the number of resources was found to be at its lowest value [21]. Thus, *E. japonicus* was not captured in many stations, leading to zero value data.

The distribution of *E. japonicus* is closely related to spatio-temporal and hydrologic environmental factors [27,42], and the prediction of its resource distribution using fewer environmental factors may lead to some deviation. In this study, due to the selection of fixed stations with equal spacing for sampling, there is a high correlation between longitude and latitude. So, longitude was excluded in modeling, which might limit the prediction effect of the model to some extent. Li et al. [19] showed that random station sampling was superior to fixed station sampling, and the fixed station sampling might underestimate the true value of resource density. Therefore, in the follow-up studies, it is necessary to optimize the sampling scheme to obtain more accurate data and improve the fitting ability of the model.

Several models commonly used to deal with zero value data were compared to select a more suitable model to explore the relationship between the *E. japonicus* resource density data and environmental factors in the present study. At the same time, researchers thought of selecting suitable SDMs for other fish species. However, the influence of biological

factors should be properly considered in future research to have a more comprehensive understanding of the habitat change mechanism and resource density in this sea area.

Author Contributions: Data analysis was performed by W.M., W.T. and J.Z.; W.M. wrote the first draft of the manuscript; S.Q. C.G. and J.M. designed the survey; J.Z. revised the manuscript and approved it for submission. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (31902372) and the Fisheries Resource Survey of Zhejiang Province, China (158053).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Acknowledgments: The authors of this research would like to thank the teachers and students from the Research Laboratory of Quantitative Assessment and Management of Fisheries Resources and Ecosystems, Shanghai Ocean University and Zhejiang Mariculture Research Institute.

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

References

1. Yu, J.; Liu, Z.N.; Chen, P.M.; Yao, L.J. Environmental factors affecting the spatiotemporal distribution of *Decapтерus maruadsi* in the western Guangdong waters, China. *Appl. Ecol. Env. Res.* **2019**, *17*, 8485–8499. [\[CrossRef\]](#)
2. Li, B.; Cao, J.; Guan, L.; Mazur, M.; Chen, Y.; Wahle, R.A. Estimating spatial non-stationary environmental effects on the distribution of species: A case study from American lobster in the Gulf of Maine. *ICES J. Mar. Sci.* **2018**, *75*, 1473–1482. [\[CrossRef\]](#)
3. Liu, X.X.; Gao, C.X.; Zhao, J.; Tian, S.Q.; Ye, S.; Ma, J. Modeling and comparison of count data containing zero values: A case study of *Setipinna taty* in the south inshore of Zhejiang, China. *Environ. Sci. Pollut. Res.* **2021**, *28*, 46827–46837. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Ma, J.; Li, B.; Zhao, J.; Wang, X.; Hodgdon, C.T.; Tian, S.Q. Environmental influences on the spatio-temporal distribution of *Coilia nasus* in the Yangtze River estuary. *J. Appl. Ichthyol.* **2020**, *36*, 317–327. [\[CrossRef\]](#)
5. Manderson, J.; Palamara, L.; Kohut, J.; Oliver, M. Ocean observatory data are useful for regional habitat modeling of species with different vertical habitat preferences. *Mar. Ecol. Prog. Ser.* **2011**, *438*, 1–17. [\[CrossRef\]](#)
6. Mouton, A.M.; Schneider, M.; Peter, A.; Holzer, G.; Müller, R.; Goethals, P.L.; de Pauw, N. Optimisation of a fuzzy physical habitat model for spawning European grayling (*Thymallus thymallus* L.) in the Aare river (Thun, Switzerland). *Ecol. Model.* **2008**, *215*, 122–132. [\[CrossRef\]](#)
7. Václavík, T.; Meentemeyer, R.K. Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecol. Model.* **2009**, *220*, 3248–3258. [\[CrossRef\]](#)
8. Hirzel, A.H.; Hausser, J.; Chessel, D.; Perrin, N. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* **2002**, *83*, 2027–2036. [\[CrossRef\]](#)
9. Phillips, S.J.; Anderson, R.P.; Schapire, R.E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **2006**, *190*, 231–259. [\[CrossRef\]](#)
10. Chang, J.H.; Chen, Y.; Holland, D.; Grabowski, J. Estimating spatial distribution of American lobster *Homarus americanus* using habitat variables. *Mar. Ecol. Prog. Ser.* **2010**, *420*, 145–156. [\[CrossRef\]](#)
11. Hua, C.X.; Zhu, Q.C.; Shi, Y.C.; Liu, Y. Comparative analysis of CPUE standardization of Chinese Pacific saury (*Cololabis saira*) fishery based on GLM and GAM. *Acta Oceanol. Sin.* **2019**, *38*, 100–110. [\[CrossRef\]](#)
12. Zhang, Y.L.; Xu, B.D.; Ji, Y.P.; Zhang, C.L.; Ren, Y.P.; Xue, Y. Comparison of habitat models in quantifying the spatio-temporal distribution of small yellow croaker (*Larimichthys polyactis*) in Haizhou Bay, China. *Estuar. Coast. Shelf Sci.* **2021**, *261*, 107512. [\[CrossRef\]](#)
13. Zhao, J.; Cao, J.; Tian, S.Q.; Chen, Y.; Zhang, S.Y.; Wang, Z.H.; Zhou, X.J. A comparison between two GAM models in quantifying relationships of environmental variables with fish richness and diversity indices. *Aquat. Ecol.* **2014**, *48*, 297–312. [\[CrossRef\]](#)
14. Dai, L.B.; Hodgdon, C.; Tian, S.Q.; Chen, J.H.; Gao, C.X.; Han, D.Y.; Kindong, R.; Ma, Q.Y.; Wang, X.F. Comparative performance of modelling approaches for predicting fish species richness in the Yangtze River Estuary. *Reg. Stud. Mar. Sci.* **2020**, *35*, 101161. [\[CrossRef\]](#)
15. Piet, G.J. Using external information and GAMs to improve catch-at-age indices for North Sea plaice and sole. *ICES J. Mar. Sci.* **2002**, *59*, 624–632. [\[CrossRef\]](#)
16. Shono, H. Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fish. Res.* **2008**, *93*, 154–162. [\[CrossRef\]](#)
17. Bouska, K.L.; Whitledge, G.W.; Lant, C. Development and evaluation of species distribution models for fourteen native central U.S. fish species. *Hydrobiologia* **2015**, *747*, 159–176. [\[CrossRef\]](#)
18. Tian, S.Q.; Chen, X.J.; Chen, Y.; Xu, L.X.; Dai, X.J. Standardizing CPUE of *Ommastrephes bartramii* for Chinese squid-jigging fishery in Northwest Pacific Ocean. *Chin. J. Oceanol. Limnol.* **2009**, *27*, 729–739. [\[CrossRef\]](#)

19. Li, B.; Cao, J.; Chang, J.H.; Wilson, C.; Chen, Y. Evaluation of effectiveness of fixed-station sampling for monitoring American lobster settlement. *N. Am. J. Fish. Manag.* **2015**, *35*, 942–957. [[CrossRef](#)]
20. Tweedie, M.C.K. An index which distinguishes between some important exponential families. In *Statistics: Applications and new directions, Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, Calcutta, India, 16–19 December 1984*; Indian Statistical Institute: Calcutta, India; pp. 579–604.
21. Ma, W.; Qin, S.; Zhao, J. Distribution characteristics and influencing factors of fish resources in the offshore waters south of Zhejiang. *Prog. Fish. Sci.* **2021**, 1–12.
22. Zhu, W.B.; Zhu, H.C.; Wang, Y.L.; Zhang, Y.Z.; Lu, Z.H.; Cui, G.C. Heterogeneity of fork length-weight relationship for juvenile *Engraulis japonicus* based on linear mixed-effects models. *J. Appl. Ecol.* **2021**, *32*, 4532–4538.
23. Wei, S.; Jiang, W.M. Study on food web of fishes in the Yellow Sea. *Oceanol. Limnol. Sin.* **1992**, *23*, 182–192.
24. Guan, L.S.; Jin, X.S.; Wu, Q.; Shan, X.J. Statistical modelling for exploring diel vertical movements and spatial correlations of marine fish species: A supplementary tool to assess species interactions. *ICES J. Mar. Sci.* **2019**, *76*, 1776–1783. [[CrossRef](#)]
25. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China; Standardization Administration of the People's Republic of China. *National Standard (Recommended) of the People's Republic of China: Specifications for Oceanographic Survey-Part 6: Marine Biological Survey*; Standards Press of China: Beijing, China, 2008.
26. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. *The Specification for Marine Monitoring-Part 3: Sample collection, Storage and Transportation*; China Standards Press: Beijing, China, 2008.
27. Niu, M.; Jin, X.S.; Li, X.S.; Wang, J. Effects of spatio-temporal and environmental factors on distribution and abundance of wintering anchovy *Engraulis japonicus* in central and southern Yellow Sea. *Chin. J. Oceanol. Limnol.* **2014**, *32*, 565–575. [[CrossRef](#)]
28. Pebesma, E.; Bivand, R.S. S classes and methods for spatial data: The sp package. *R News* **2005**, *5*, 9–13.
29. Fujita, T.; Yamamoto, M.; Kono, N.; Tomiyama, T.; Sugimatsu, K.; Yoneda, M. Temporal variations in hatch date and early survival of Japanese anchovy (*Engraulis japonicus*) in response to environmental factors in the central Seto Inland Sea, Japan. *Fish. Oceanogr.* **2021**, *30*, 527–541. [[CrossRef](#)]
30. Meng, W.Z.; Gong, Y.H.; Wang, X.F.; Tong, J.F.; Han, D.Y.; Chen, J.H.; Wu, J.H. Influence of spatial scale selection of environmental factors on the prediction of distribution of *Coilia nasus* in Changjiang River Estuary. *Fishes* **2021**, *6*, 48. [[CrossRef](#)]
31. Baayen, H.; Vasisht, S.; Kliegl, R.; Bates, D. The cave of shadows: Addressing the human factor with generalized additive mixed models. *J. Mem. Lang.* **2017**, *94*, 206–234. [[CrossRef](#)]
32. Zhang, T.J.; Song, L.M.; Yuan, H.C.; Song, B.; Ebango Ngando, N. A comparative study on habitat models for adult bigeye tuna in the Indian Ocean based on gridded tuna longline fishery data. *Fish. Oceanogr.* **2021**, *30*, 584–607. [[CrossRef](#)]
33. Planque, B.; Bellier, E.; Lazure, P. Modelling potential spawning habitat of sardine (*Sardina pilchardus*) and anchovy (*Engraulis encrasicolus*) in the Bay of Biscay. *Fish. Oceanogr.* **2007**, *16*, 16–30. [[CrossRef](#)]
34. Stow, C.A.; Jolliffe, J.; McGillicuddy, D.J., Jr.; Doney, S.C.; Allen, J.I.; Friedrichs, M.A.M.; Rose, K.A.; Wallhead, P. Skill assessment for coupled biological/physical models of marine systems. *J. Mar. Syst.* **2009**, *76*, 4–15. [[CrossRef](#)] [[PubMed](#)]
35. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [[CrossRef](#)]
36. Maunder, M.N.; Punt, A.E. Standardizing catch and effort data: A review of recent approaches. *Fish. Res.* **2004**, *70*, 141–159. [[CrossRef](#)]
37. Ma, S.S. Relationship between distribution and hydrological conditions of the wintering anchovy in Yellow Sea and East China Sea. *J. Fish. China* **1989**, *13*, 201–206.
38. Yalcin, E.; Gurbet, R. Environmental influences on the spatio-temporal distribution of European Hake (*Merluccius merluccius*) in Izmir Bay, Aegean Sea. *Turk. J. Fish. Aquat. Sci.* **2016**, *16*, 1–14. [[CrossRef](#)]
39. Guan, W.J.; Chen, X.J.; Gao, F.; Li, G. Environmental effects on fishing efficiency of *Scomber japonicus* for Chinese large lighting purse seine fishery in the Yellow and East China Seas. *J. Fish. Sci. China* **2009**, *16*, 949–958.
40. Selleslagh, J.; Amara, R. Environmental factors structuring fish composition and assemblages in a small macrotidal estuary (eastern English Channel). *Estuar. Coast. Shelf Sci.* **2008**, *79*, 507–517. [[CrossRef](#)]
41. Niu, M.X.; Wang, J. Variation in the distribution of wintering anchovy *Engraulis japonicus* and its relationship with water temperature in the central and southern Yellow Sea. *Chin. J. Oceanol. Limnol.* **2017**, *35*, 1134–1143. [[CrossRef](#)]
42. Niu, M.X.; Wang, J.; Wu, Q.; Sun, J.Q. The relationship of stock density distribution of wintering anchovy (*Engraulis japonicus*) and environmental factors based on remote sensing in central and southern Yellow Sea. *Prog. Fish. Sci.* **2020**, *41*, 11–20.
43. Lu, Y.; Yu, J.; Lin, Z.J.; Chen, P.M. Environmental influence on the spatiotemporal variability of spawning grounds in the western Guangdong waters, South China Sea. *J. Mar. Sci. Eng.* **2020**, *8*, 607. [[CrossRef](#)]
44. Liu, X.X.; Wang, J.; Zhang, Y.L.; Yu, H.M.; Xu, B.D.; Zhang, C.L.; Ren, Y.P.; Xue, Y. Comparison between two GAMs in quantifying the spatial distribution of *Hexagrammos otakii* in Haizhou Bay, China. *Fish. Res.* **2019**, *218*, 209–217. [[CrossRef](#)]
45. Zhang, Q.H.; Cheng, J.H.; Xu, H.X.; Shen, X.Q.; Yu, G.P.; Zheng, Y.J. *Fishery Resources and their Sustainable Utilization in the East China Sea*; Fudan University: Shanghai, China, 2007.
46. Liu, X.Q.; Hou, Y.J.; Yin, B.S. Dynamic process of vertical circulation and temperature-salinity structures in coastal area of East China Sea II the structure of the water temperature and salinity. *Oceanol. Limnol. Sin.* **2004**, *35*, 497–506.
47. Zhu, W.B.; Zhu, H.C.; Zhang, Y.Z.; Wang, J.; Jiang, R.J.; Lu, Z.H.; Cui, G.C.; Dai, Q. Quantitative distribution of juvenile *Engraulis japonicus* and the relationship with environmental factors along the Zhejiang coast. *J. Fish. Sci. China* **2021**, *28*, 1175–1183.