Received: 1 October, 2011

A comparison of methods for prediction of Integrated Habitat Index of *Thunnus albacares* in the Indian Ocean – general linear model and quantile regression model considerations

Liming Song [a,b,*],Yaping Wu [b] , Tom Nishida[c]

(*aThe Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306，China*

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

cNational Research Institute of Far Seas Fisheries (NRIFSF), Fisheries Research Agency, Japan 5-7-1, Orido, Shimizu City, Shizuoka, Japan 424-8633)

Abstract: There are many methods to study the tuna spatial distribution , and it is important to know the habitat of the fish species for better conservation and management of marine ecosystems. Based on the survey data collected by *Huayuanyu No. 18*, the vertical profile data of temperature, salinity, chlorophyll-a concentration, dissolved oxygen concentration and the catch rate data of yellowfin tuna (*Thunnus albacares*) were applied to develop the "Integrated Habitat Index (*IHI*)" models by the quantile regression method and general linear model (GLM). We used the statistical test, Wilcoxon test, residual analysis to test the results from the two kinds of models. The results showed that, the quantile regression method could be better than general linear method to study the pelagic species spatial distribution; yellowfin tuna's frequently swimming depth was from 80 to 200 m in the survey area; the main environmental variables which influence the distribution of yellowfin tuna in specific depth stratum were different; the weighted average of temperature and dissolved oxygen concentration effected to the spatial distribution of yellowfin tuna significantly.

Key words: *Thunnus albacares*; GLM; quantile regression; Indian Ocean

---

*Corresponding author: Liming Song, College of Marine Sciences, Shanghai Ocean University, 999 Huchenghuan Road, Lingangxincheng Shanghai 201306, China. Tel.: +86 21 61900311; fax: +86 21 61900301. E-mail address: lmsong@shou.edu.cn

1. Introduction

There are many methods to study organisms' habitats (environment), *e.g.*, general linear model (Clark *et al.*, 1999; Labonne *et al.*, 2003), general linear additive model (Swartzman *et al.*, 1992; Maravelias, 1999), multiple linear regression model (Beamish and Lowartz, 1996), regression tree model (Norcross *et al.*, 1997; Turgeon and Rodriguez, 2005), logistic regression model (Norcross *et al.*, 1999; Guay *et al.*, 2000;Turgeon and Rodríguez, 2005), habitat suitability index (HSI) model (Brown *et al.*,2000;Cade and Noon,2003; William and Maughan,2004), and quantile regression model (Terrell *et al.*, 1996; Dunham *et al.*, 2002; Eastwood *et al.*, 2003; Wang, 2006; Feng *et al.*, 2007; Song *et al.*, 2007; Song and Zhou, 2010). Some of the studies include combinations of these models and geographic information systems (GIS) to construct fish habitat maps (Riou *et al.*, 2001; Stoner *et al.*, 2001). General linear model (GLM) is the most common method in study fish abundance (Allen and Punsly, 1984). Shono *et al* (2002) applied the GLM to study the resource status of yellowfin tuna (*Thunnus albacares*) in the Indian Ocean, including six explanatory variables: year, month, area, hooks between the floats (HBF), sea surface temperature (SST), Southern Oscillation index and the interaction among them. Nishida *et al.* (2003) used general linear model- habitat based model (GLM habitat HBM) to analyze the CPUE standardization of yellowfin tuna, results showed that, the CPUE that was standardized by GLM/HBM was consistent with the CPUE estimated by GLM.

Although general linear model have their advantages, they also have some disadvantages (Song and Zhou, 2010). In general linear model, the CPUE was defined as a linear combination between dependent variable and explanatory variables. These variables are usually continuous or discrete, and the continuous variables usually were classified as discrete variables, which possibly changed the original character of the variables. Thereby, the uncertainty of the results was increased. In statistics, quantile regression is a type of regression analysis, (1) quantile regression can be used to estimate the median, rather than ordinary least squares regression to estimate the mean, and will be more robust in response to large outliers. Quantile regression can be seen as a natural analogue in regression analysis to the practice of using different measures of central tendency and statistical dispersion to obtain a more comprehensive and robust analysis (Koenker, 2005); (2) In ecology, quantile regression has been proposed and used as a way to discover more useful predictive relationships between variables in cases where there is no relationship or only a weak relationship between the means of such variables. The need for and success of quantile regression in ecology has been attributed to the complexity of interactions between different factors leading to data with unequal variation of one variable for different ranges of another variable (Cade and Noon, 2003).

General linear model as a common method, it was applied extensively in CPUE standardization and habitat investigation, but there is no study on the performance comparison about quantile regression and general linear models for prediction of the integrated habitat index of tuna species. In this study, by using the quantile regression method and general linear model, the Integrated Habitat Indices (*IHI*) models were developed for yellowfin tuna in the Indian Ocean based on the data collected by *Huayuanyu No. 18 (HYY18)* in 2005. The *IHI* models were used to evaluate responses of yellowfin tuna to environmental variables and to predict their spatial distribution. The models were validated with the environmental variables and fishery data collected by *Huayuanyu No. 19 (HYY19).* The objectives of this study are to (1) determine an optimal method to study Integrated Habitat Indices (*IHI*) of yellowfin tuna; (2) define key habitat variables to influence the spatial distribution of yellowfin tuna; and (3) develop forecasting models for yellowfin tuna spatial distribution based on the key habitat variables.

2. Materials and methods

2.1 Materials

2.1.1 Survey vessels and fishing gear

Data were collected from operations on two longliners, *Huayuanyu 18(HYY18)* and *Huayuanyu 19(HYY19)*, in 2005. Two vessels' specifications are same. They were equipped with super spool and chill sea water equipment. Two vessels' length over all, mould breadth, mould depth, gross tonnage, net tonnage and main engine power is 26.12 m, 6.05 m, 2.70 m, 150.00 t, 45.00 t and 407.00 kW, respectively.

The configurations of the gear used are shown in Table 1 of Song and Zhou (2010). The maximum diameter of the branch lines was 5mm. Two configurations of fishing gear were used, conventional and experimental gears. Their configurations and designs are shown in Song *et al.* (2009). Fishing parameters, *e.g.* fishing time, total number of hooks per set, bait, and deploying method *etc.* are shown in Song *et al.* (2009) and Song and Zhou (2010).

2.1.2 Survey duration and area

The survey duration was from September 15 to December 12 in 2005. Fishing took place mainly between about 1°N and 10°N and between about 62°E and 70°E (Fig. 1). The data collection locations were shown in Fig.1. During the survey, two fishing vessels targeted bigeye tuna (*Thunnus obesus*), and the bycatch included yellowfin tuna, swordfish (*Xiphias gladius*), albacore (*Thunnus alalunga*) and billfishes (*Istiophoridae*).

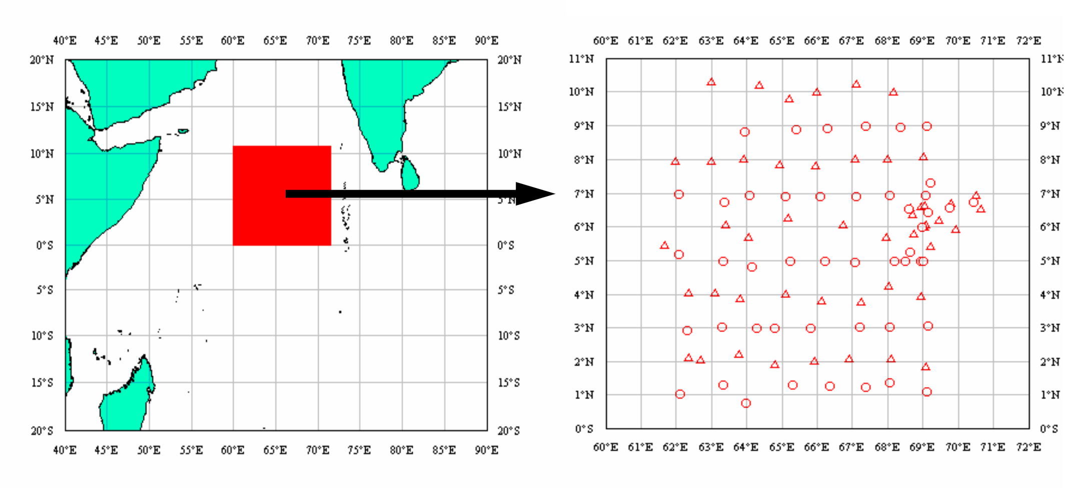


Fig. 1 *Huayuanyu 18* and *Huayuanyu 19* measured positions in 2005
(○*Huayuanyu 18*； △ *Huayuanyu 19*)

2.1.3 Instrumentation and measurement

The environmental sampling instruments included Submersible Data Logger (SDL), XR-620

(RBR Ltd., Ottawa, Ontario, Canada) and Temperature Depth Recorders (TDR), TDR-2050 (RBR Ltd.) (14 in total), Conductivity Temperature Depth Recorder (CTD), SBE37SM (SeaBird Co., Bellevue, Washington, USA) and three dimension (3D) Aquadopp Current Profile (ADCP), Aquadopp-2000 (NORTECK Co., Vangkroken, Norway) which has the same sampling sites as those of XR-620. The measurement ranges of environmental variables and the precision of the data were shown in Song *et al.* (2009). The water temperature, salinity, dissolved oxygen and chlorophyll-a were measured at sampling sites. Considering the accuracies of data from various instruments and requirements of the study, the data of depth and temperature were processed to one effective decimal place, salinity, dissolved oxygen, and catch rate to two decimal places, and chlorophyll-a and 3D current to three decimal places.

The hook depth measuring method by TDRs and the environmental variables measuring method are shown in Song and Zhou (2010). The following information was collected: deployment position and time, course and speed, line shooter speed, number of hooks between floats (HBF), time interval between deploying fore and after branch lines, number of hooks, time of retrieving lines, hook position codes at which fish were caught, number of hooked tuna per fishing operation, and positions at which yellowfin tuna were hooked.

2.2 Data processing

The catch rate of yellowfin tuna for station i, $CPUE_i$, was calculated as follows:

$$CPUE_i = \frac{U_i}{f_i} \times 1000 \tag{1}$$

where i denotes station, U_i is the number of yellowfin tuna hooked at station i, f_i is the number of hooks deployed at station i, and i=1,2,3, ..., 48.

The data for the yellowfin tuna were grouped into depth strata. The data were assigned to seven depth strata of 40 m each (80-120 m, 80-120 m, … 320-360 m).

The catch rate, $CPUE_{ij}$, of *HYY18* at sampling station i in depth stratum j was calculated by:

$$CPUE_{ij} = \frac{N_{ij}}{H_{ij}} \times 1000 \tag{2}$$

where H_{ij} is the number of hooks deployed by *HYY18* at station i in depth stratum j, and N_{ij} is the number of yellowfin tuna caught by *HYY18* at station i in depth stratum j. N_{ij} was calculated as:

$$N_{ij} = \frac{N_j}{N} \times N_i \tag{3}$$

where N_j is the number of yellowfin tuna caught by *HYY18 and HYY19* in depth stratum j during the survey, following Song *et al.* (2008; 2009), N is the number of yellowfin tuna caught by

HYY18 and HYY19 during the survey. In 2005, the data on depth of capture were collected for 299 of the 516 yellowfin tuna (56.8%). N_i is the number of yellowfin tuna caught by *HYY18* at sampling station *i*, and i = 1, 2, 3, ……, 48.

The catch rate of yellowfin tuna in depth stratum *j*, $CPUE_j$, was clculated as follows:

$$CPUE_j = \frac{N_j}{f_j} \times 1000 \tag{4}$$

where *j* denotes stratum, f_j is the number of hooks deployed at stratum *j*, and j=1,2,3, ..., 7.

The data processing procedures to analyze the catch rate data at the specific depth stratum and at each sampling station are shown in Song and Zhou (2010). The overall values of $CPUE_j$ for 2005 are shown in Fig. 2.

The weighted average value of environmental variables based on the catch rate of different depth stratum at sampling station *i* was calculated, following Song *et al.* (2007) , and Song and Zhou (2010), as:

$$ENV_i = \sum (CPUE_j ENV_{ij}) / \sum CPUE_j \tag{5}$$

where ENV_i was the weighted average environmental variable of whole water bin, the environmental variable includes temperature (T_i), salinity (S_i), chlorophyll-a concentration (Ch_i), dissolved oxygen concentration (DO_i), horizontal current (HC_i), and vertical current (WC_i) at sampling station *i* from *HYY 18*, and ENV_{ij} was the value of the above environmental variables at sampling station *i* in depth stratum *j* (*i.e.*, 80～120 m, 120～160 m, . . ., 320～360 m). T_{ij}, S_{ij}, Ch_{ij}, and DO_{ij} were the arithmetic means measured with the XR- 620 at sampling station *i* in depth stratum *j*. HC_{ij} and WC_{ij} were the arithmetic means measured with the Aquadopp-2000 at sampling station *i* in depth stratum *j*.

The shear of horizontal current component (denoted as ξ) was estimated by integrating the original data measured with the Aquadopp-2000 from the near-surface to the largest predicted hook depth (z) at each sampling station *i* (Bigelow *et al.*, 2006). The coefficient ξ was used to study the potential *CPUE* of sampling station *i*.

$$\xi = \log \left\{ \frac{\sum_{j=1}^{7} \left[\left(\frac{\beta_{j+1} - \beta_j}{\tau_{j+1} - \tau_j} \right)^2 + \left(\frac{\delta_{j+1} - \delta_j}{\tau_{j+1} - \tau_j} \right)^2 \right]^{\frac{1}{2}} (\tau_{j+1} - \tau_j)}{\sum_{j=1}^{7} (\tau_{j+1} - \tau_j)} \right\} \tag{6}$$

where ξ was the shear of horizontal current component, β_j was the East-Westward component of current in the *j* depth stratum, δ_j was the North-Southward component of current in the *j* depth stratum, τ_j was the depth of *j* depth stratum.

Thermocline intensity (TI_i) (°C m[-1]) was calculated by temperature profile measured by XR-620 at station i as:

$$TI_i = \frac{T_u - T_b}{D_b - D_u} \tag{7}$$

where T_u, T_b, D_u, and D_b was thermocline's upper temperature (°C), bottom temperature (°C), upper depth (m) and bottom depth (m).

2.3 Developing quantile regression models (QRM)

The development of quantile regression models for $CPUE_{ij}$ and $CPUE_i$ is detailed in Song and Zhou (2010). In this study, the prediction model was developed based on the data measured by XR-620 of *HYY 18* at 30 sampling stations. The following six variables were considered as the independent variables in the quantile regression models of yellowfin tuna catch rate at sampling station i in depth stratum j: temperature (T_{ij}), salinity (S_{ij}), chlorophyll-a concentration (Ch_{ij}), dissolved oxygen concentration (DO_{ij}), horizontal current (HC_{ij}), and vertical current (WC_{ij}). Fifteen interaction terms of these six variables were also considered.

The full regression model for describing the relationship between the expected catch rate at sampling station i in depth stratum j, $CPUE_{QRMij}$ versus T_{ij}, S_{ij}, Ch_{ij}, DO_{ij}, HC_{ij}, WC_{ij}, and the interaction terms can be written as:

$$\begin{aligned} CPUE_{QRMij} &= C_{ij} + a_{ij}T_{ij} + b_{ij}S_{ij} + c_{ij}Ch_{ij} + d_{ij}DO_{ij} + e_{ij}HC_{ij} + f_{ij}WC_{ij} + g_{ij}TS_{ij} + h_{ij}TCh_{ij} + k_{ij}TDO_{ij} + l_{ij}THC_{ij} \\ &+ m_{ij}TWC_{ij} + n_{ij}SCh_{ij} + o_{ij}SDO_{ij} + p_{ij}SHC_{ij} + q_{ij}SWC_{ij} + r_{ij}ChDO_{ij} + s_{ij}ChHC_{ij} + t_{ij}ChWC_{ij} + u_{ij}DOHC_{ij} + v_{ij}DOWC_{ij} \\ &+ w_{ij}HCWC_{ij} + \varepsilon_{ij} \end{aligned} \tag{8}$$

where C_{ij} was the constant, TS_{ij} was the interaction of temperature and salinity, TCh_{ij} was the interaction of temperature and chlorophyll-a concentration, . . ., $HCWC_{ij}$ was the interaction of horizontal current and vertical current, and ε_{ij} was the error term at sampling station i in depth stratum j. The values of a_{ij}, b_{ij}, c_{ij}, d_{ij}, e_{ij}, f_{ij} . . . *and* w_{ij} were their corresponding parameters.

The full regression model for describing the relationship between the expected catch rate at sampling station i, $CPUE_{QRMi}$ versus weighted average T_i, S_i, Ch_i, DO_i, ξ_i, WC_i, TI_i and the interaction terms can be written as:

$$\begin{aligned} CPUE_{QRMi} &= C_i + a_iT_i + b_iS_i + c_iCh_i + d_iDO_i + e_i\xi_i + f_iWC_i + g_iTI_i + h_iTS_i + k_iTCh_i + l_iTDO_i + m_iT\xi_i \\ &+ n_iTWC_i + o_iTTI_i + p_iSCh_i + q_iSDO_i + r_iS\xi_i + s_iSWC_i + t_iSTI_i + u_iChDO_i + v_iCh\xi_i + w_iChWC_i + \\ &x_iChTI_i + y_iDO\xi_i + z_iDOWC_i + aa_iDOTI_i + ab_i\xi WC_i + ac_i\xi TI_i + ad_iWCTI_i + \varepsilon_i \end{aligned} \tag{9}$$

where C_i was the constant, TS_i was the interaction of temperature and salinity, , . . . $WCTI_i$ was the interaction of vertical current and thermocline intensity; ε_i was the error term of expected catch rate at sampling station i. The a_i, b_i, c_i, d_i, e_i, f_i … … ad_i were their corresponding parameters.

It is more appropriate to select θ values between 0.50 and 0.95 to build the upper-quantile model (Feng *et al.*, 2007). In this study, we used the statistical software Blossom to process the data, which was developed by Midcontinent Ecological Science Center (U.S.Geological Survey).

2.4 Developing general linear models (GLM)

In order to compare with the quantile regression models, the environmental variables in different depth strata which we inputted into the GLM models were same, we applied *R* project 2.13 (Chambers, 1992; Fox, 2005, 2007) to fit the linear models.

The full regression model for describing the relationship between the expected catch rate at sampling station *i* in depth stratum *j*, $CPUE_{GLMij}$ versus T_{ij}, S_{ij}, Ch_{ij}, DO_{ij}, HC_{ij}, WC_{ij}, and the interaction terms can be written as:

$$\begin{aligned}&LN(CPUE_{GLMij}+CONSTANT_{ij})=INTERCEPT_{ij}+a_{ij}'T_{ij}+b_{ij}'S_{ij}+c_{ij}'Ch_{ij}+d_{ij}'DO_{ij}+e_{ij}'HC_{ij}+f_{ij}'WC_{ij}+g_{ij}'TS_{ij}\\&+h_{ij}'TCh_{ij}+k_{ij}'TDO_{ij}+l_{ij}'THC_{ij}+m_{ij}'TWC_{ij}+n_{ij}'SCh_{ij}+o_{ij}'SDO_{ij}+p_{ij}'SHC_{ij}+q_{ij}'SWC_{ij}+r_{ij}'ChDO_{ij}+s_{ij}'ChHC_{ij}+\\&t_{ij}'ChWC_{ij}+u_{ij}'DOHC_{ij}+v_{ij}'DOWC_{ij}+w_{ij}'HCWC_{ij}+\varepsilon_{ij}' \qquad (10)\end{aligned}$$

$$\varepsilon_{ij}' \sim (0,S^2)$$

where LN is Napierian Logarithm, $CONSTANT_{ij}$ was generally 10% of total average nominal $CPUE_{ij}$. $INTERCEPT_{ij}$ and ε_{ij}' were vertical intercepts and error terms. The a_{ij}', b_{ij}', c_{ij}', d_{ij}', …… w_{ij}' were their corresponding parameters.

The full regression model for describing the relationship between the expected catch rate at sampling station *i*, $CPUE_{GLMi}$ versus weighted average T_i, S_i, Ch_i, DO_i, ξ_i, WC_i, TI_i and the interaction terms can be written as:

$$\begin{aligned}&LN(CPUE_{GLMi}+CONSTANT_i)=INTERCEPT_i+a_i'T_i+b_i'S_i+c_i'Ch_i+d_i'DO_i+e_i'\xi_i+f_i'WC_i+g_i'TI_i\\&+h_i'TS_i+k_i'TCh_i+l_i'TDO_i+m_i'T\xi_i+n_i'TWC_i+o_i'TTI_i+p_i'SCh_i+q_i'SDO_i+r_i'S\xi_i+s_i'SWC_i+t_i'STI_i\\&+u_i'ChDO_i+v_i'Ch\xi_i+w_i'ChWC_i+x_i'ChTI_i+y_i'DO\xi_i+z_i'DOWC_i+aa_i'DOTI_i+ab_i'\xi WC_i+ac_i'\xi TI_i\\&+ad_i'WCTI_i+\varepsilon_i' \qquad (11)\end{aligned}$$

$$\varepsilon_i' \sim (0,S^2)$$

where $CONSTANT_i$ was generally 10% of total average nominal $CPUE_i$. $INTERCEPT_i$ and ε_i' were vertical intercepts and error terms. The a_i', b_i', c_i', d_i', …… ad_i' were their corresponding parameters.

The AIC and BIC values (Sakamoto et al, 1986; Moore, 2000) were used to test the model goodness-of-fit**.** At first, we input all parameters into the models, according to the *P* value to select the parameters, and the smaller of AIC and BIC value, the better of the model goodness-of-fit. Finally, the independent variables in model shall be with significant (P < 0.05). We used GLM model to get the CPUE prediction models in specific depth strata and the whole water bin.

2.5 IHI_{ij} calculation based on the GLM and QRM

Based on the regression models (Eq. 8, Eq. 10), $CPUE_{QRMij}$ and $CPUE_{GLMij}$ were estimated at sampling station *i* in depth stratum *j*. IHI_{QRMij} and IHI_{GLMij} were calculated by the following equations:

$$IHI_{QRMij} = \frac{CPUE_{QRMij}}{CPUE_{QRM\max}} \tag{12}$$

$$IHI_{GLMij} = \frac{CPUE_{GLMij}}{CPUE_{GLM\max}} \tag{13}$$

where $CPUE_{QRMmax}$ and $CPUE_{GLMmax}$ were the maximum value among all $CPUE_{QRij}$ and $CPUE_{GLMij}$ at sampling station *i* in depth stratum *j*, specifically.

2.6 $\overline{IHI}$ calculation based on the GLM and QRM at sampling station *i*

Based on the regression models (Eq.9, Eq.11), $CPUE_{QRMi}$ and $CPUE_{GLMi}$ at sampling station *i* were estimated. $\overline{IHI_{QRM}}$ and $\overline{IHI_{GLM}}$ were derived from $CPUE_{QRMi}$ and $CPUE_{GLMi}$ using the following equations:

$$\overline{IHI_{GRM}} = \frac{CPUE_{QRMi}}{CPUE_{QRMi\max}} \tag{14}$$

$$\overline{IHI_{GLM}} = \frac{CPUE_{GLMi}}{CPUE_{GLMi\max}} \tag{15}$$

where $CPUE_{QRMi\max}$ and $CPUE_{GLMi\max}$ were the maximum value among all $CPUE_{QRMi}$ and $CPUE_{GLMi}$, specifically.

2.7 The predictive power of the GLM and QRM

The predictive power of the *IHI* models were qualitatively analyzed by plotting the map of average value of predicted IHI_j and observed $CPUE_j$. The nonparametric test, two related samples Wilcoxon test (SPSS 16.0), was used to test the difference between IHI_{QRMij} and IHI_{GLMij}, $\overline{IHI_{QRM}}$ and $\overline{IHI_{GLM}}$. If there were significant difference between them, we would use other statistical test to testify the predictive power of QRM and GLM. According to the predicted $CPUE_{QRMij}$, $CPUE_{GLMij}$, $CPUE_{QRMi}$ and $CPUE_{GLMi}$, we used quantile comparison (QQ) plot of residuals between predicted $CPUE_{QRMij}$, $CPUE_{GLMij}$, $CPUE_{QRMi}$ and $CPUE_{GLMi}$ and their respective observed CPUE by *R*-project 2.13 to analyze the reliability of the models (Eq.16,17,18,19).

$$\text{Residual}_{QRMij} = CPUE_{QRMij} - CPUE_{ij} \quad (16)$$

$$\text{Residual}_{QRMi} = CPUE_{QRMi} - CPUE_{i} \quad (17)$$

$$\text{Residual}_{GLMij} = CPUE_{GLMij} - CPUE_{ij} \quad (18)$$

$$\text{Residual}_{GLMi} = CPUE_{GLMi} - CPUE_{i} \quad (19)$$

2.8 Validation of the QRM and GLM

The $\overline{IHI_{QRM}}$ and $\overline{IHI_{GLM}}$ models were validated by applying to the observed *CPUE* data from *HYY 18 and HYY 19* (operation duration and area were Sep.～Dec. 2005 and 0ºN～8ºN, 61ºE～71ºE, respectively). The nominal *CPUE* from *HYY 18* and *HYY 19* and the $\overline{IHI}$ isoline from *HYY 18* at all stations were mapped by Marine explorer 4.71.

Table 1 Estimation parameters of optimal predicting equation

	80-120		120-160		160-200		200-240	
Depth stratum (m)	QRM	GLM	QRM	GLM	QRM	GLM	QRM	GLM
	θ=0.90	*AIC=89.38* *BIC*= 99.19	θ=0.85	*AIC=78.84* *BIC=95.65*	θ=0.75	*AIC= 56.55* *BIC= 73.36*	θ=0.85	*AIC= 63.83* *BIC= 75.04*
	PMa							
C_j (constant) /Intercept	256.73	-1792.40	-4226.95	-520.08	-8333.64	21965.81	-52.03	-52.54
$a_j(T_{ij})$	-10.19	83.91	-27.20	-3.12	-12.92	-1421.34	0	2.71
$b_j\ (S_{ij})$	0	50.81	132.89	15.88	240.67	-626.60	0	-
$c_j\ (Ch_{ij})$	0	-	0	431.10	0	3873.11	1549.15	284.47
$d_j\ (DO_{ij})$	0	-100.51	0	1.08	8.80	-994.74	0	15.27
$e_j(HC_{ij})$	0	-	0	-128.05	0	-	0	7.30
$f_j\ (WC_{ij})$	0	-	5646.19	223.13	0	-	0	-
$g_j\ (TS_{ij})$	0	-2.38	0	-	0	40.56	0	-
$h_j\ (TCh_{ij})$	0	-	0	-18.45	0	-266.49	0	-
$k_j\ (TDO_{ij})$	0	-	0	-	0	-1.74	0	-0.95
$l_j\ (THC_{ij})$	0	-	0	9.04	0	-	0	-
$m_j\ (TWC_{ij})$	0	-	-319.79	-14.36	0	-	0	-
$n_j\ (SCh_{ij})$	0	-	0	-	0	-	0	-
$o_j\ (SDO_{ij})$	0	2.85	0	-	0	28.89	0	-
$p_j\ (SHC_{ij})$	0	-	0	-	0	-	0	-
$q_j\ (SWC_{ij})$	0	-	0	-	0	-	0	-
$r_j(ChDO_{ij})$	0	-	0	-	0	-	0	-
$s_j\ (ChHC_{ij})$	0	-	0	-253.64 -253.64	0	- 270.05	0	-
$t_j\ (ChWC_{ij})$	0	-	0	-	0	-	0	-
$u_j\ (DOHC_{ij})$	0	-	0	-	0	-4.39	0	-2.76
$v_j\ (DOWC_{ij})$	0	-	0	-	0	-	0	-
$w_j\ (HCWC_{ij})$	0	-	0	-	0	-	0	-

a Parameters.

Table 1(continued) Estimation parameters of optimal predicting equation

Depth stratum(m)	240-280		280-320		320-360	
	QRM	GLM	QRM	GLM	QRM	GLM
	θ=0.90	*AIC= 63.73* *BIC= 69.34*	θ=0.80	*AIC= 73.46* *BIC= 79.07*	θ=0.90	*AIC= 54.93* *BIC= 71.74*
	PMa					
C_j （constant） /Intercept	-491.38	-14.64	5.02	-14.89	139.83	-4184.54
$a_j(T_{ij})$	36.60	-	0	-	-10.78	198.97
$b_j (S_{ij})$	0	-	0	-	0	140.18
$c_j (Ch_{ij})$	14501.80	295.15	0	277.96	0	2.73
$d_j (DO_{ij})$	0	1.313	0	1.39	0	-
$e_j(HC_{ij})$	0	-	-5.20	-	-110.51	87.52
$f_j (WC_{ij})$	0	-	206.03	-	0	-7519.96
$g_j (TS_{ij})$	0	-	0	-	0	3148.20
$h_j (TCh_{ij})$	-1082.29		0	-	0	-6.71
$k_j (TDO_{ij})$	0	-	0	-	0	-
$l_j (THC_{ij})$	0	-	0	-	8.77	-
$m_j (TWC_{ij})$	0	-	0	-	0	-4.67
$n_j (SCh_{ij})$	0	-	0	-	0	-
$o_j (SDO_{ij})$	0	-	0	-	0	-
$p_j (SHC_{ij})$	0	-	0	-	0	-
$q_j (SWC_{ij})$	0	-	0	-	0	-
$r_j(ChDO_{ij})$	0	-	0	-	0	-
$s_j (ChHC_{ij})$	0	-	0	-	0	238.01
$t_j (ChWC_{ij})$	0	-	0	-	0	-99.94
$u_j (DOHC_{ij})$	0	-	0	-	0	-
$v_j (DOWC_{ij})$	0	-	0	-	0	-
$w_j (HCWC_{ij})$	0	-	-434.82		0	-

a Parameters.

3. Results

3.1 Predictive models of the catch rate

The estimated parameters of the predictive models of the catch rate, $CPUE_{QRMij}$ and $CPUE_{GLMij}$, at sampling station i, depth stratum j were summarized in Table 1. The predictive models of the catch rate, $CPUE_{QRMij}$, were developed with a different value of quantile (θ). The value of quantile (θ) was from 0.75 to 0.90 for various depth strata. The key environmental parameters to construct the $CPUE_{QRMij}$ and $CPUE_{GLMij}$ models for various depth strata were different from depth strata (Table 1).

The optimal model of the catch rate, $CPUE_{QRMi}$ and $CPUE_{GLMi}$, at sampling station i were developed.

When θ was 0.95, the optimal $CPUE_{QRMi}$ model was derived as:

$$C\hat{P}UE_{QRMj} = -125.46 + 7.15T_j + 108.49DO_j - 6.05TDO_j \quad (16)$$

The P-values of T_i, DO_i, and TDO_i were all less than 0.05. Weighted average T_i, DO_i, and the interaction term TDO_i were identified as key variables (Eq. 16).

The optimal $CPUE_{GLMi}$ model was derived as:

$$CPUE_{GLMi} = -4184.54 + 198.97T_i + 140.18S_i + 2.736Ch_i + 87.52\xi_i - 7519.958WC_i + 3148.20TI_i - 6.71TS_i - 4.67\,T\xi_i + 238.01\,SWC_i - 99.944\,STI_i \quad (17)$$

The P-values of T_i, S_i, Ch, ξ_i, WC_i, TI_i, TS_i, $T\xi_i$, SWC_i and STI_i were all less than 0.05. Weighted average T_i, S_i, Ch, ξ_i, WC_i, TI_i and the interaction terms TS_i, $T\xi_i$, SWC_i and STI_i were identified as key variables (Eq. 17).

3.2 The predictive power of IHI_{QRM} and IHI_{GLM}

In Fig.2, the IHI_{QRMj} and observed $CPUE_j$ in the corresponding depth stratum had almost similar trend, but there were differences for the depth strata of 160～200 m and 240～280 m. IHI_{GLMj} was a little lower than IHI_{QRMj}. IHI_{GLMj} in depth stratum of 80～120 m was lower than that in the depth strata of 120～160 m and 160～200 m, which was different from IHI_{QRMj} and observed $CPUE_j$. The Fig.2 showed that the IHI were relative high in the 80～240 depth strata.

The Table.2 showed that, there were significant difference between the IHI_{QRMij} and IHI_{GLMij}, $\overline{IHI_{QRM}}$ and $\overline{IHI_{GLM}}$. Comparing Fig.3 to Fig.4, the residuals between $CPUE_{QRMi}$, $CPUE_{QRMij}$ and observed CPUE had more effective values in the red dotted line area and obeyed normal distribution, so the QRM was better than the GLM.

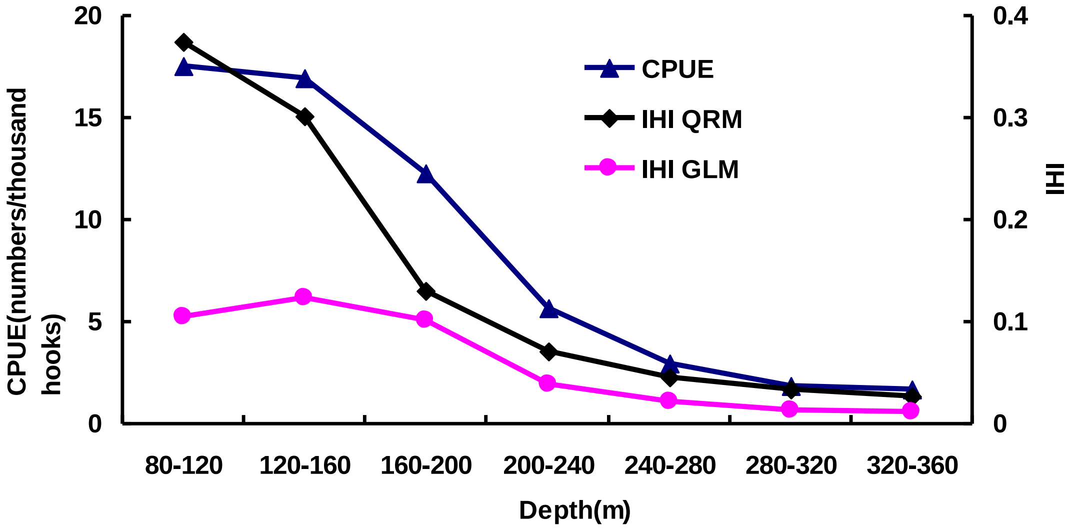


Fig.2 The arithmetic average *IHI* and the *CPUE* of yellowfin tuna in respective depth stratum.

Table 2. The results of Wilcoxon signed-rank test on the differences between IHI_{QRMij} and IHI_{GLMij}, $\overline{IHI_{QRM}}$ and $\overline{IHI_{GLM}}$

Depth stratum	*Df*	Wilcox.test	
		*Za**	*P value(two-tailed)*
80-120	29	-4.721	0.000
120-160	29	-4.573	0.000
160-200	29	-2.478	0.013
200-240	29	-4.021	0.000
240-280	29	-4.371	0.000
280-320	29	-4.456	0.000
320-360	29	-4.371	0.000
$\overline{IHI_{QRM}}$ and $\overline{IHI_{GLM}}$	29	-2.828	0.005

Za was based on the positive ranks.

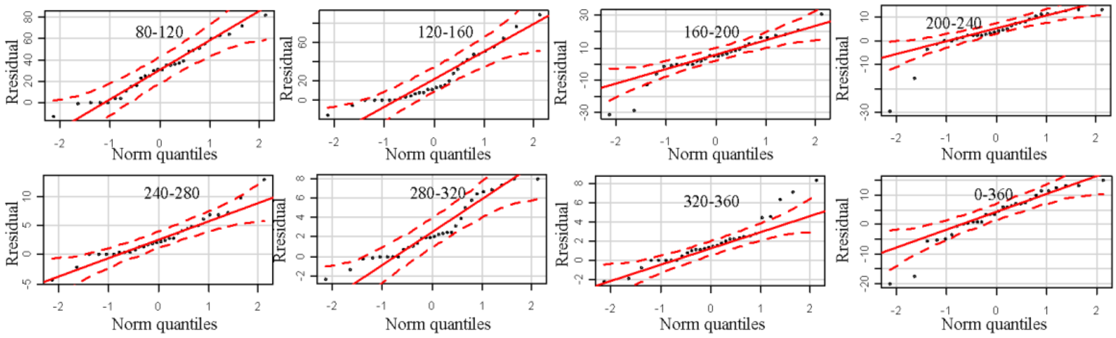


Fig.3 The quantile comparison (QQ) plot of residual between the observed *CPUE* and predicted *CPUE* from QRM of yellowfin tuna in respective depth stratum.

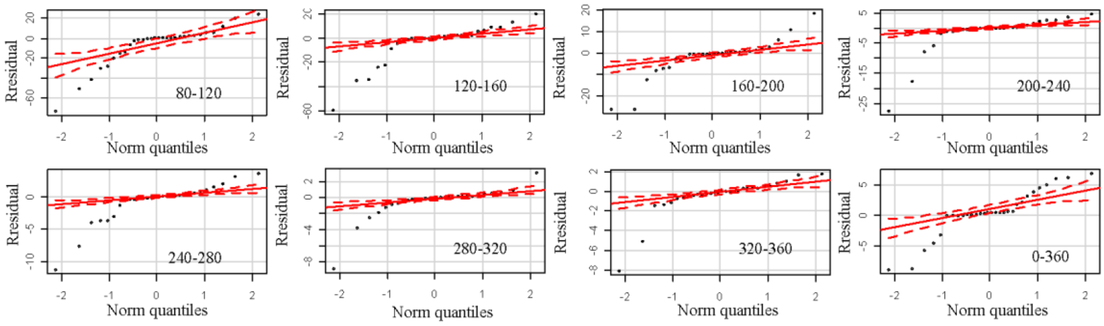
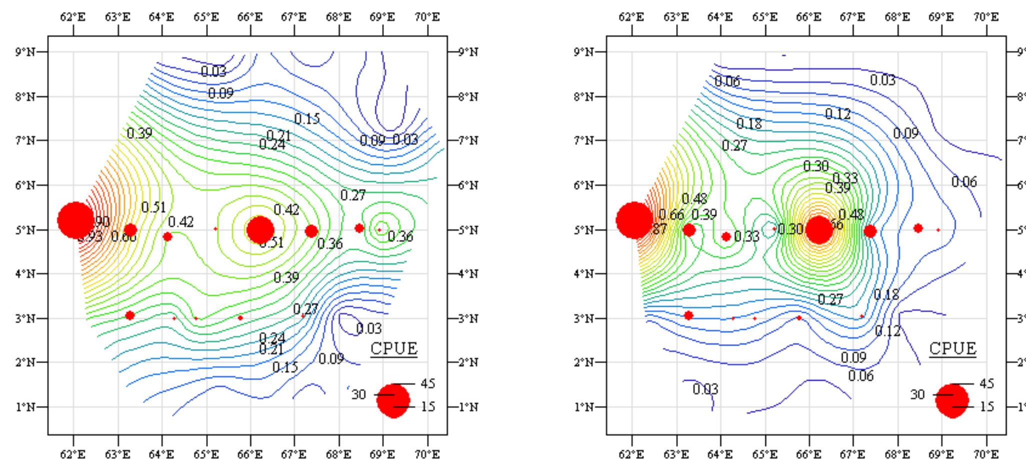


Fig.4 The quantile comparison (QQ) plot of residual between the observed *CPUE* and predicted *CPUE* from GLM of yellowfin tuna in respective depth stratum.

3.4 Validation of the GLM and QRM

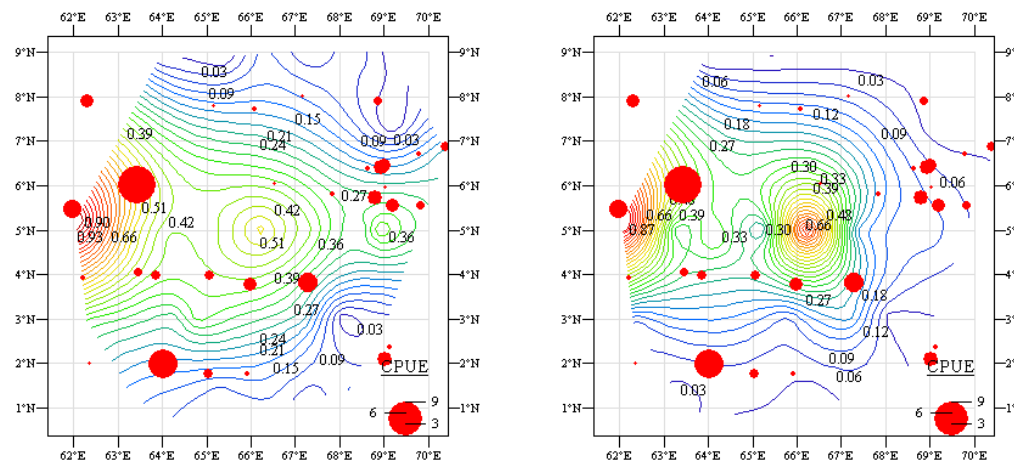
Applying the longline *CPUE* data from *HYY 18* and *HYY 19* in 2005 to validate the $\overline{IHI}$ model, in the area of 1ºN～9ºN,61ºE～71ºE, the validation result of the $\overline{IHI}$ model for the water bin was shown in Fig.5 and Fig.6. In Fig.5(a) and (b), the higher catch rate were almost distributed in the area of 3°40′N～6°20′N, 62°00′E～67°E, which was almost consistent with our prediction results. In Fig.5(a), the high catch rate distributed in the area 3°40′N～6°20′N, 62°E～69°E where the *IHI* were higher than 0.27; In Fig.5(b), the high catch rate in the area of 3°40′N～6°30′N, 62°E～67°E distributed in the area where the *IHI* were higher than 0.27, and partly in the area of 4°40′N～5°20′N, 67°30′E～69°E, the IHI was between 0.09 and 0.25.

In Fig.6(a) and (b), the high catch rate were mainly distributed in the area of 3°40′N～6°20′N, 62°00′E～67°E, the higher catch rate were almost distributed in the higher IHI area. Thus, the range of higher IHI in Fig.6(a) included more higher catch rate than the IHI in Fig.6(b).



(a) (b)

Fig.5 The catch rates of yellowfin tuna from *HYY 18* and the isolines of the yellowfin tuna's $\overline{IHI}_{QRMi}$ (a) and $\overline{IHI}_{GLMi}$ (b) predicted using the field measured environmental data of *HYY 18*.



(a) (b)

Fig.6 The catch rates of yellowfin tuna from *HYY 19* and the isolines of the yellowfin tuna's $\overline{IHI}_{QRMi}$ (a) and $\overline{IHI}_{GLMi}$ (b) predicted using the field measured environmental data of *HYY 18*.

4. Discussion

4.1 The predictive power of the GLM and QRM models

In general, the predictive power of the IHI_{QRM} IHI_{GLM} models were good. The trends of the IHI_{QRM} were similar to those of the $CPUE_j$ of yellowfin tuna at the respective depth strata (Fig.2). The IHI_{QRM}, IHI_{GLM} and catch rate was relatively high at the depth strata of 80～120, 120～160 m and 160～200 m, which indicated that the frequently swimming depth range for yellowfin tuna was from 80 to 200 m. Mohri and Nishida (2000) suggested that the optimal depth range of yellowfin tuna was from 80 to 120 m. Song *et al.* (2008) suggested that the depth stratum of 100.0～179.9 m was the frequently swimming depth range for yellowfin tuna. The results of Mohri and Nishida (2000) and Song *et al.* (2008) were almost consistent with the result of this study.

4.2 The reliability of IHI_{ij} and $\overline{IHI}$ predictive method

The reliability of models developed by using the quantile regression method was better than general liner method. In general, the IHI_{ij} predictive models could be used to predict the spatial distribution of yellowfin tuna. The IHI_{QRM} and IHI_{GLM} value of yellowfin tuna were higher in depth strata of 80～120 m, 120～160 m, and 160～200 m (Fig.2). In different depth strata, the key variables to predict the models were different. That is because the distributions of yellowfin tuna at different depth were different. The variables to effect the distribution of yellowfin tuna at different depth were also different (Table 1). The yellowfin tuna have obvious habitat selection, and usually were affected by the environmental variables, such as temperature, salinity and so on (Block *et al*., 1997). Antonio *et al*. (2004) found that the yellowfin tuna has apparent seasonal south-northward migration characteristic, and its migration path was relevant to the route of current moving. The distribution of yellowfin tuna have greatly influenced by the water temperature, salinity, and dissolved oxygen (Korsmeyer *et al*., 1997).

The predictive model QRM were more effective than GLM. The Wilcoxon test showed, there were significant difference between IHI_{QRMij} and IHI_{GLMij}, $\overline{IHI_{QRM}}$ and $\overline{IHI_{GLM}}$ (Table 2). The results of the QQ plot showed that, the reliability of models developed by using the QRM were better than GLM. Eastwood and Meaden (2004) mentioned that quantile regression has all the advantages of ordinary least-squares regression, and an additional advantage of being distribution-free. It can be used to estimate the effects of limiting factors, and thus provide a good means to evaluate how a species may respond to changes in its environmental variables. When the errors are not normally distributed and only part of the limiting factors is measured, it can yield several estimation results in different quantiles, and may more accurately reflect the responses of a species to habitat variables, particularly for the regression models of the upper quantile (Cade and Noon, 2003; Wu and Ma, 2006). In the areas of 1°30′N～9°N, 62°E～64°30′E, 3°40′N～6°N, 67°E～70°E, the distribution area of higher catch rate from *HYY 19* was almost consistent with the distribution area of higher *IHI* (Fig.6(a)).

The QRM were better for the study of habitat integrated index than GLM. The $\overline{IHI_{QRM}}$ prediction model included the weighted average temperature and dissolved oxygen. The $\overline{IHI_{GLM}}$ prediction model included the temperature, salinity, chlorophyll-a, the shear of horizontal current component and thermocline intensity. In the field, it is difficult to measure the environment variables. The quantile regression method included the less variables, which could reduce the cost of the survey. In addition, the results from the quantile regression models were more reliable.

4.3 Key environmental parameters to influence the spatial distribution of yellowfin tuna

The weighted average temperature and dissolved oxygen were the crucial variables to the spatial distribution of yellowfin tuna. For the $\overline{IHI_{QRM}}$ prediction model, there was the close relationship between the predicted *CPUE* and the weighted average temperature and dissolved oxygen. Temperature was the main limiting factor to effect the migration of yellowfin tuna, which limited the heart’s capacity to export dissolved oxygen, then influenced the swimming speed of yellowfin tuna (Maury *et al*., 2001). Nishida *et al*. (2001) found that the distribution of adult yellowfin tuna was influenced by the spatial and seasonal change of water temperature. Brill *et al.* (1999) suggested that the temperature, dissolved oxygen and thermocline depth influenced the spatial distribution of yellowfin tuna, and the temperature and dissolved oxygen also influenced the spawning behavior of yellowfin tuna in the Hawaii Islands waters. Schaefer (1996) reported the moving path of yellowfin tuna was influenced by the seasonal change of 18℃ sea surface temperature. Block *et al*. (1997) indicated a large number of yellowfin tuna gathered into the dissolved oxygen concentration 6.8～8.6 mg L^{-1} area by the acoustic telemetry study. We suggest that the weighted average temperature and dissolved oxygen data should be input into the model to standardize the yellowfin tuna *CPUE*.

5. Conclusion

The predictive power of IHI_{QRM} models developed in this study was good. The IHI models developed by using the quantile regression in this study could reflect the habitat selection of the yellowfin tuna more accurately, in general. We suggest the quantile regression method could be used to study the spatial distribution of pelagic fish caught by longline fishery. The optimal inhabiting depth stratum of yellowfin tuna was 80～200 m in the survey area. The *IHI* within the area defined by 4°N～9°N, 62°E～63°30′E had the largest values, and the *IHI* in 3°40′N～6°20′N, 63°30′E～69°E had larger values, the *IHI* in the other areas had smaller values. We suggest that the weighted average temperature and dissolved oxygen data should be included in the CPUE standardization to estimate the relative abundance of yellowfin tuna.

Acknowledgements

The project is funded by Ministry of Agriculture of P.R of China under Project of Fishery Exploration in High Seas in 2005 (Project No.Z05-30), Shanghai Leading Academic Discipline Project (Project No. S30702), and the National High Technology Research, Development Program of China (Project No. 2007AA092202), and the Initiation Fund for Doctors of Shanghai Ocean University (Project No. B-8202-08-0290). We thank the general manager Fang Jingmin, vice general manager Huang Fuxiong and the crews of tuna longliners, *et al.* of Guangyuang Fishery group Ltd of Guangdong province for their supporting to this project.

References

Antonio J, Alberto AB, Susana MA et al (2004) Spatial analysis of yellowfin tunas (*Thunnus albacores*) catch rate and its relation to *El Nino* and *La Nina* events in the eastern tropical Pacific. Deep-Sea Res II 51(5 / 6): 575—586.

Allen R. and Punsly R. Catch rates as indices of abundance of yellowfin tuna, *Thunnus albacares*, inthe eastern Pacific Ocean [R]. Bull. Inter-Amer. Trop. Tuna Comm, 1984, 18(4): 301–379.

Beamish FWH, Lowartz S (1996) Larval habitat of American brook lamprey.Canadian J.Fish. Aquatic Sci 53:693~700.

Bigelow KA, Musyl MK, Poisson F et al (2006) Pelagic longline gear depth and shoaling. Fish. Res 77: 173~183.

Brown SK, Buja KR, Jury SH et al (2000) Habitat suitability index models for eight fish and invertebrate species in Casco and Sheepscot Bays.Maine. North American Journal of Fisheries Management 20:408~435.

Brill RW, Block BA, Boggs CH et al （1999） Horizontal movements and depth distribution of large adult yellowfin tunas (*Thunnus albacares*) near the Hawaiian Islands, recorded using ultrasonic telemetry: implications for the physiological ecology of pelagic fishes. Marine Biology 133: 395-408.

Block BA, Keen JE, Castillo B et a1 (1997) Environmental preferences of yellowfin tunas(*Thunnus albacares*) at the northern extent of its range. Marine Biology 130: 119-132.

Cade BS, Richards JD (2001) User Manual for BLOSSOM Statistical Software. Colorado: Midcontinent Ecological Science Center U. S. Geological Survey 1~106.

Cade BS, Noon BR (2003) A gentle introduction to quantile regression for ecologists[J]. Front Ecol. Environ 1(8): 412~420.

Chambers, J. M. (1992) Linear models. Chapter 4 of Statistical Models in S eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

Clark RD, Minello TJ, Christensen JD et al（1999） Modeling nekton habitat use in Galveston Bay, Texas: an approach to define essential fish habitat.NOAA/NOS Biogeography Program, Silver Spring, Maryland, and National Marine Fisheries Service, Galveston, Texas.

Dunham JB, Cade BS Terrell JW (2002) Influence of spatial and temporal variation on fish-habitat relationships defined by regression quantile. Trans Am Fish Soc 131: 86~98.

Eastwood PD, Meaden GJ, Carpentier A et al (2003) Estimating limits to the spatial extent and suitability of sole (*Solea solea*) nursery grounds in the Dover Strait.Journal of Sea Research 50: 151~165.

Feng B, Chen XJ, Xu LX（2007） Study on distribution of Thunnus obesus in theIndian Ocean based on habitat suitability index. J. Fish. China 31,805–812 (in Chinese, with English abstract).

Fox, J. (2005). The R Commander: A basic-statistics graphical user interface to R. Journal of Statistical

Fox, J. (2007). Extending the R Commander by “plug in” packages. R News, 7(3):46—52.

Software, 19(9):1—42.

Guay JC, Boisclair D, Rioux D et al (2000) Development and validation of numerical habitat models for juveniles of Atlantic salmon (*Salmo salar*).Fish.Aquatic Sci 57:2065~2075.

Koenker, R (2005) Quantile Regression, Cambridge University Press. ISBN 0-521-60827-9

Korsmeyer KE, Lai NC, Shadwick RE et al （1997） Heart rate and stroke volume contributions to cardiac output in swimming yellowfin tunas: response to exercise and temperature. Exp Biol 200(14):1975-1986.

Labonne J, Allouche S, Gaudin P (2003) Use of a generalised linear model to test habitat preferences: the example of *Zingel asper*, an endemic endangered percid of the River RhÔne.Freshwater Biology 48: 687~697.

Maravelias CD (1999) Habitat selection and clustering of a pelagic fish:effects of topography and bathymetry on species dynamics.Fish.Aquatic Sci 56:437~450.

Maury O, Gascuel D, Marsac F et a1 （2001）Hierarchical interpretation of nonlinear relationships linking yellowfin tunas (*Thunnus albacares*) distribution to the environment in the Atlantic Ocean[J]．Fish Aquat Sci 58(3)：458—469．

Mohri M, Nishida T (2000) Consideration on distribution of adult yellowfin tunas(*Thunnus albacares*) in the Indian Ocean based on Japanese tunas longline fisheries and survey information. IOTC Proceedings 3: 276-282.

Moore, D. S. (2000). The Basic Practice of Statistics, Second Edition. Freeman, New York.

Norcross BL, Blanchard A, Holladay BA (1999) Comparison of models for defining near shore flatfish nursery areas in Alaskan waters. Fish.Oceanogr 8:50~67.

Norcross BL, MÜter FJ, Holladay BA (1997) Habitat models for juvenile pleuronectids around Kodiak Island, Alaska.Fish. Bull 95: 504~520.

Nishida T, Mohri M, Itoh K et a1 (2001) Study of bathymetry effects on the nominal hooking rates of yellowfin tunas (*Thunnus Albacares*) and bigeye tunas (*Thunnus obesus*) exploited by the Japanese tunas longline fisheries in the Indian Ocean, IOTC Proceedings 4: 191-206.

Nishida T, Bigelow K, Mohri M, et al. Comparative study on J apanese tuna longline CPUE standardization of yellowfin tuna (*Thunnus albacares*) in the Indian Ocean based on two methods:-general linear model (GLM) and habitat-based model (HBM)/GLM combined-(1958-2001). WPTT-03-05. IOTC Proceedings no. 6, 2003 :48-69.

Riou P, Le Pape O, Rogers SI (2001) Relative contributions of different sole and plaice nurseries to the adult population in the eastern Channel: application of a combined method using generalized linear models and a geographic information system.Aq.Liv.Res 14:125~135.

Song LM, Gao PF, Zhou YQ, Zhang Y（2007）Habitat environment integration index of Thunnus obesus in the high seas of the central Atlantic Ocean based on the quantile regression. J. Fish. China 31,798–804 (in Chinese, with English abstract).

Schaefer KM (1996) Spawning time, frequency, and batch fecundity of yellowfin tunas, *Thunnus albacares*, near Clipperton Atollin the eastern Pacific Ocean . Fish Bull US 94: 98-112.

Shono H., Okamoto H., Nishida T. Standardized cpue for yellowfin tuna (*thunnus albacares*) of the Japanese longline fishery in the Indian Ocean by generalized linear models [R]. IOTC Proceedings no. 5 , 2002: 240-247.

Sakamoto, Y., Ishiguro, M., and Kitagawa G. (1986). Akaike Information Criterion Statistics. D. Reidel Publishing Company.

Song LM, Zhang Y, Xu LX et a1 (2008) Environmental preferences of longlining for yellowfin tunas(*Thunnus albacares*)in the tropical high seas of the Indian Ocean. Fisheries Oceanography 17 (4): 239 – 253.

Song LM, Zhou J, Zhou YQ, Nishida T, Jiang WX, Wang JQ (2009) Environmental preferences of bigeye tunas, *Thunnus obesus*, in the Indian Ocean: an application to a longline fishery. Environ. Biol. Fish. 85 : 153–171.

Song LM, Zhou YQ (2010) Developing an integrated habitat index for bigeye tunas (*Thunnus obesus*) in the Indian Ocean based on longline fisheries data. Fish. Res 105(2):63-74.

Stoner AW, Manderson JP, Pessutti JP (2001) Spatial explicit analysis of estuarine habitat for juvenile winter flounder: combining generalized additive models and geographic information systems.Marine Ecology Progress Series 213:253~271.

Swartzman G, Huang C, Kaluzny (1992) Spatial analysis of Bering Sea ground fish survey data using generalized additive models.Fish.Aquatic Sci 49:1366~1378.

Terrell JW, Cade BS, Carpenter J et al (1996) Modeling stream fish habitat limitation from wedge-shaped patterns of variation in standing stock. Trans Am Fish Soc 125: 104~117.

Turgeon K and Rodríguez M (2005) Predicting microhabitat selection in juvenile Atlantic salmon *Salmo salar* by the use of logistic regression and classification trees.Freshwater Biology 50:539~551.

Wang JQ (2006) Modelling the habitat suitability for bigeye tuna(*Thunnu obesus*)in the Indian Ocean based on Quantile Regression．Shanghai Fishes University。

William GL, Maughan OE (2004) Spotted bass habitat evaluation using an unweighted geometric mean to determine HSI values.Proceeding of the Oklahoma Academy of Science 65:11~18.

Wu JN, Ma W (2006) A comparison of the quantile regression and the significance weighting analyzing technic. Stat. Decis 4:4–7 (in Chinese, with English abstract).