Standardization of Catch Per Unit Effort for Chilean Jack Mackerel (*Trachurus murphyi*) from Chinese Trawl Fleet on the High Seas in the Southeast Pacific Ocean (2001-2010)

Gang Li, Xiaorong Zou, Yinqi ZHOU

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

Abstract

Generalized linear model (GLM) and generalized additive model (GAM) are commonly used to evaluate impacts of environmental variables on fisheries catch per unit fishing effort (CPUE) and to develop standardized CPUE which is often used as a relative index in fisheries stock abundance. The Standardized CPUE for Chilean jack mackerel (*Trachurus murphyi*) caught by Chinese fishing fleet in the Southeast Pacific Ocean from 2001 to 2010 were derived by means of GLM and GAM approaches. Nine variables, Year, Month, Vessel, La Niña and El Niño events (ELE), Longitude, Latitude, SST, SSTA and Nino3.4 index were used to build GLM and GAM models. The first five variables were significant in GLM analysis which account for 27.34% of the variance in nominal CPUE. In stepwise GAM analysis, all the nine factors were significant and explained 30.60% of the variance, and Month, Year and Vessel were the main effects on CPUE. High CPUEs were observed during April to September, at sea surface temperatures 12-16 ℃ and sea surface temperature anomaly 0.2-1.0℃. CPUE was higher significantly in normal years than it in La Niña and El Niño years. Declined Chilean jack mackerel abundance was detected from 2001 to 2010 with a rise in 2007.

1. Instruction

Chilean jack mackerel, *Trachurus murphyi*, is a pelagic fish and widely distrbuted in the southern Pacific Ocean, from the west Latin American coast to New Zealand (Elizarov et al., 1993; Cubillos ea al., 2008). Chinese factory trawlers caught this species on the high seas off Chile waters from 2000, and average annual catch was around 100 thousand tons with dramatically fluctuant during 2001-2010. In 2010, the total catch was only 63.61 thousand tons.

The data of Catch per fishing unit (CPUE), is often used as a relative index of fisheries stock abundance (Hilborn and Walters, 2001; Nishida and Chen, 2004), and it can be influenced by many factors such as fishing capacity (e.g. characteristics of fleet, fishing gear and equipment), environmental factors (e.g. sea water temperature at varied depth, sea surface temperature, dissolved oxygen concentration, marine currents and lunar phases), and spatial and temporal factors (e.g. latitude, longitude, year and month). The CPUE needs to be standardized so that the impact of these factors can be removed or minimized and to got the better performance of CPUE reflecting the changes in population abundance only (Maunder and Punt, 2004). GLM is the most frequently used (Nishida and Chen, 2004; Maunder and Langley, 2004), although the use of GAM has increased substantially over the last decade (Venables and Dichmont, 2004). Chambers and Hastie (1997) suggested that GAMs are capable of dealing with non-linearity by incorporating terms non-parametrically into the model; while in GLMs, the response variables were described by the linear combination of predicted values (Quinn and Keough, 2002).

The purpose of this study is to obtain standardized CPUE which can be used for Chilean jack mackerel stock assessment. We used GLM and GAM methods to standardize the nominal Chilean jack mackerel catch rate of the Chinese trawl fishery operating in the Southeast Pacific Ocean, and evaluated impacts of a list of temporal, spatial, environmental and fisheries operational variables on the catch rate.

2. Materials and Methods

2.1 Database

Catch and effort data were mainly collected from the logbook of the Chinese trawl fishery for Chilean jack mackerel. Logbook data were available from 2001 to 2010 on a daily basis and incorporated name of trawler, duration of each trawling, start and end location (latitude and longitude) of each trawling, weight of fish caught per trawling. The start location of fishing operation was used as the trawling position. The uncompleted data records (lack of information about location and duration of trawling, and fishing years of a trawler less than 5 years) was deleted, so catch and effort data of seven trawlers were selected, and the information about theses vessels were listed in Table 1. Effort was measured as haul times in hours, and CPUE was expressed as ton per hour. The data were compiled by year, month, vessel name, latitude and longitude (0.5°spatial resolution). Therefore, monthly catch rate by vessel in the same 0.5° ×0.5° cell was calculated as the ratio of the recorded catch to total hours fished.

In addition, environmental data, monthly sea surface temperature (SST), monthly sea surface temperature anomaly (SSTA) and Nino 3.4 index were also assembled into the data records. SST data (9km resolution) and SSTA data (1° resolution) were obtained from the Physical Oceanography DACC of NASA (http://poet.jpl.nasa.gov), and Nino 3.4 index was obtained from Climate Prediction Center of NOAA (http://www.cpc.ncep.noaa.gov). El Niño and La Niña events (ELE) was expressed as 1 and -1 respectively, while in normal years, the value of monthly ELE is 0. SST data were averaged and SSTA data were interpolated to 0.5° grid cells to match the resolution of the catch and effort data.

2.2 Generalized linear model and generalized additive model

Generalized linear models (GLMs), formally introduced by Nelder and Wedderburn (1972) and developed by McCullough and Nelder (1989), are the most frequently used to standardize the catch and effort data (Nishida and Chen, 2004; Maunder and Langley, 2004). Its key assumption is that the relationship between the some function of the expected value of the response variable and the explanatory variables is linear (Maunder and Pount, 2004):

$$g(u_i) = X_i^T \beta$$

Where g is the differentiable and monotonic link function, $\mu_i = E(Y_i)$, X_i is the vector that specifies the explanatory variables for the ith value of the response variable, β is a vector of parameters, and Y_i the ith response.

Generalized additive models (GAMs; Hastie et al., 2001) are extensions of GLMs (Maunder and Pount, 2004; Venables and Dichmont, 2004). The assumption for GAMs is that the response variable is related to smooth additive functions of the explanatory variables (Guisan et al., 2002). GAM can be expressed as:

$$g(u_i) = \mu + \sum_{i=1}^{P} f_i(X_i)$$

where f_i is the smoother function, loess or spline smoother is common used (Maunder and Pount, 2004; Venables and Dichmont, 2004).

Temporal (Year and Month), spatial (Longitude and Latitude), fishing technological (Vessel code) and environmental (SST, SSTA and Nino3.4 index) factors are the explanatory variables, the natural logarithm of the Chilean jack mackerel catch-rate +0.1 is the response variable in the GLM and GAM analysis. Constant 0.1 was added so as to avoid taking the logarithm of zero (Campbell, 2004; Su et al., 2008). Year, Month, Vessel code and ELE are categorical variables in GLM models and factors in GAM models, and the other five are continuous variables in GLM models and in GAM models (covariates). All of the analyses are based on the assumption that catch-rate is normal distribution after log-transformation, and plots of frequency and quantile–quantile plots will be used to examine this assumption. Therefore, identity link function was used with the Gauss distribution for the response variable ln(CPUE+1). The full GLM can be written as:

ln(CPUE+1) ~ Year + Month + Vessel + Longitude + Latitude + ELE +SST
+ SSTA + Nino3.4index.

The full GAM is:

ln(CPUE+1) ~ Year + Month + Vessel + ELE + s(Longitude) + s(Latitude) + s(SST) + s(SSTA)+ s(Nino3.4index);
where s(x) is the spline smoother function of the covariate x.

Type III sums of squares were used to test the significance of each model parameters (Damalas et al., 2007). Stepwise GLM and GAM was used to the explanatory variables, which were added to GLM and GAM models one by one. Models were built by adding in new variables and seeing how much they improved the fit, and non-significant variables were dropped. Alternative model structures (different choices for the explanatory variables) were compared using Akaike Information Criterion (AIC), the pseudo-coefficient of determination (pseudo-R2) and the adjusted pseudo-R2 (Su et al., 2008). The optimal GLM or GAM model was considered as the one with the smallest AIC, the biggest pseudo-R2 and adjusted pseudo-R2:

$$\text{Pseudo - R}^2 = 1 - \frac{\text{Residual deviance}}{\text{null deviance}};$$

$$\text{Adjusted pseudo - R}^2 = 1 - \frac{\text{Residual deviance/degree of freedom}}{\text{null deviance/degree of freedom}}.$$

3. Results

In total, there were 7011 records to be analyzed in GLM and GAM. The histogram with density line (Fig. 1) and Q-Q polt (Fig.2) for log-transformed catch rate showed that ln(CPUE+0.1) tended to follow normal distribution (μ=1.63, σ=1.14).

Type III sums of squares analysis and F test revealed that five of nine main effects, Year, Month, Vessel, ELE and Latitude were significantly different from zero at α=0, while the other four variables were not (Table 2). Thus, these five factors were selected as the final explanatory variables for the GLM.

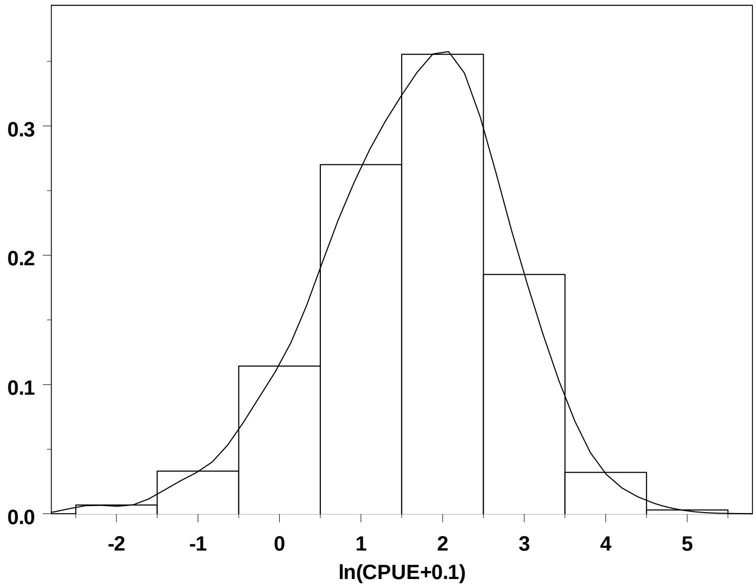

Fig.1 Frequency and fitted density line for ln(CPUE+0.1)

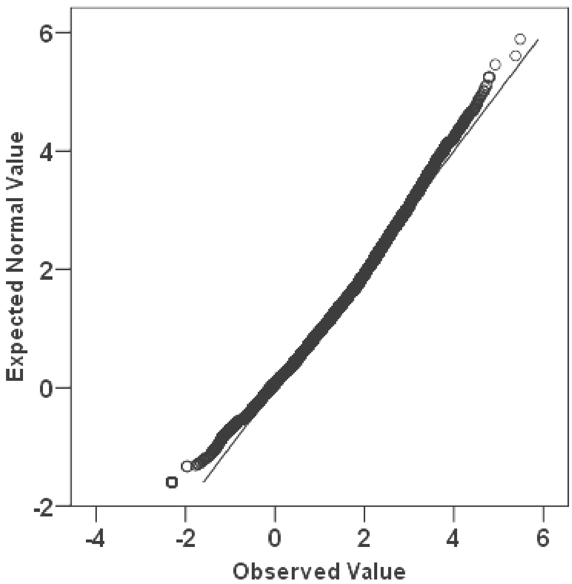

Fig.2. Normal Q-Q polt for ln(CPUE+0.1)

The GLM model with five main effects explained 27.34% deviance of the total (Table 3). The variable with the highest contribution to the explained deviance was Month (16.58%), followed by Vessel (6.01%) and Year (3.33%), while the explained deviance of ELE and Latitude was less than one percent. Most of the explained deviance came from the temporal effects and Vessel. The GLM identified that the three categorical factors, Year, Month and Vessel, were the most influential and explained 25.91% of the total variance in the Chilean jack mackerel CPUE. GLM effects of the predictor variables on ln(CPUE+0.1) are illustrated in Fig. 3.

Table 1 Information of the 7 trawlers, the parameters of Length, Beam et al reflected different fishing performance among these trawlers to a certain degree, and code was an index to categorize the fishing efficiency for the 7 trawlers, which be used as a variable in generalized linear model and generalized additive model.

Name of Trawler	Code	Length (m)	Beam (m)	Moulded depth (m)	Gross register tonnage	Power of main engine (kw)	Hold capacity (m^3)
KAIFU	1	91.1	9.2	20	7671	5920	4000
KAICHUANG	2	84.1	9.4	15	3080	2646	1900
KAILI	3	119.6	12.2	19	7874	5296	4351
KAISHUN	4	119.6	12.2	19	7874	5296	4351
KAIXIN	5	96.7	10.2	16	4407	5152	2702
KAIYU	6	91.1	9.2	20	7671	5920	4000
FUXINHAI	7	84.1	9.4	15	3080	2646	1900

Table 2 Summary of analysis of deviance for generalized linear model fitted to the Chilean jack mackerel catch rate data of the Chinese trawl fishery during 2001-2010 in the Southeast Pacific Ocean.

Source	Sum of Squares	d. f.	Mean Square	*F*	*p*
GLM—dependent variable: ln(CPUE + 0.1)					
Year	188.84	9	20.98	22.30	0.0000
Month	714.30	11	64.93	69.00	0.0000
Vessel	492.77	6	82.13	87.27	0.0000
ELE	66.03	2	33.03	35.09	0.0000
Latitude	9.24	1	9.24	9.82	0.0017
Longitude	0.22	1	0.22	0.24	0.6252
SST	1.27	1	1.27	1.35	0.2452
SSTA	0.02	1	0.02	0.02	0.8822
Nino3.4index	0.71	1	0.73	0.75	0.3853
Residual	6565.76	6977	0.94		
Total	27661.90	7011			
Corrected toal	9041.56	7010			

R^2 =0.274 (Adjusted R^2 =0.270)

Table 3 Stepwise generalized linear model building for factors affect Chilean jack mackerel catch rate of the Chinese trawl fishery from 2001 to 2010 in the Southeast Pacific Ocean.

Added terms	d.f.	Deviance	Resid. d.f.	Resid. dev.	*p*	Cumulative explained %	AIC	Pseudo adj.R^2	Pseudo R^2
NULL			7010	9041.56					
+Year	9	300.97	7001	8740.60	0.0000	3.33	21464.27	0.032	0.033
+Month	11	1498.65	6990	7241.95	0.0000	19.90	20167.58	0.197	0.199
+Vessel	6	543.40	6984	6698.55	0.0000	25.91	19632.73	0.256	0.259
+ELE	2	81.88	6982	6616.68	0.0000	26.82	19550.50	0.265	0.268
+Latitude	1	46.99	6981	6569.69	0.0000	27.34	19502.55	0.270	0.273

Table 4 Stepwise generalized additive model building for factors affect Chilean jack mackerel catch rate of the Chinese trawl fishery from 2001 to 2010 in the Southeast Pacific Ocean.

Added terms	d.f.	Deviance	Resid. d.f.	Resid. dev.	*p*	Cumulative explained dev. %	AIC	Pseudo adj.R^2	Pseudo R^2
null	1		7010	9041.56					
+Year	9	300.97	7001	8740.60	0.0000	3.33	21464.27	0.032	0.033
+Month	11	1498.65	6990	7241.95	0.0000	19.90	20167.58	0.197	0.199
+Fleet	6	543.40	6984	6698.55	0.0000	25.91	19632.73	0.256	0.259
+ELE	2	81.88	6982	6616.68	0.0000	26.82	19550.51	0.265	0.268
+s(Longitude)	1	147.46	6978	6469.22	0.0000	28.45	19394.50	0.281	0.285
+s(Latitude)	1	76.93	6974	6392.28	0.0000	29.30	19312.62	0.289	0.293
+s(SST)	1	42.01	6970	6350.27	0.0000	29.77	19268.39	0.294	0.298
+s(SSTA)	1	8.88	6966	6341.39	0.0449	29.86	19260.58	0.294	0.299
s(Nino3.4)	1	66.56	6962	6274.83	0.0000	30.60	19188.60	0.301	0.306

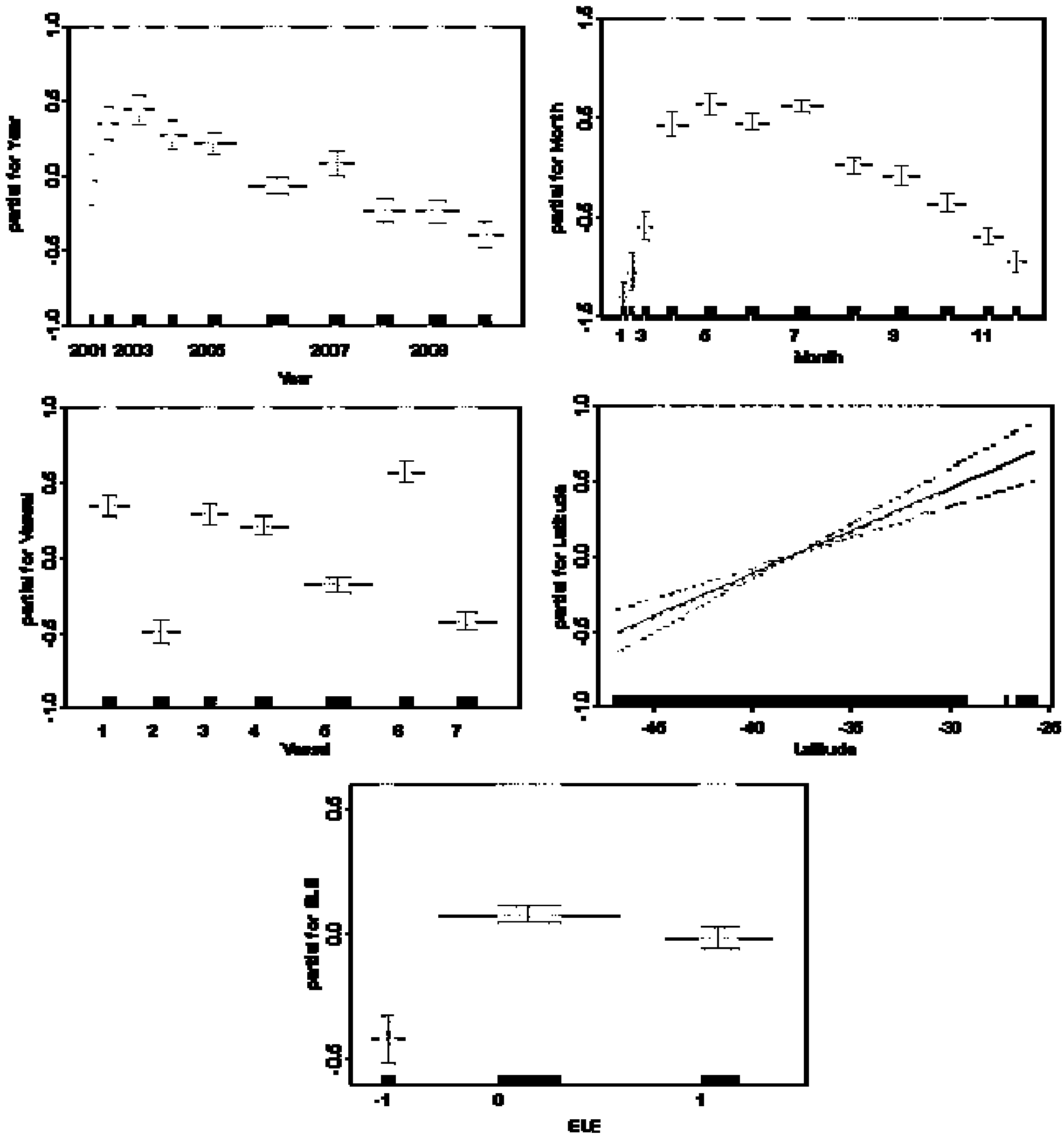

Fig.3. GLM effects of the five factors (Year, Month, Vessel, ELE and Latitude) on the log-trabsformed catch-rate of Chilean jack mackerel by Chinese trawl fishery from 2001 to 2010 in the Southeast Pacific Ocean.

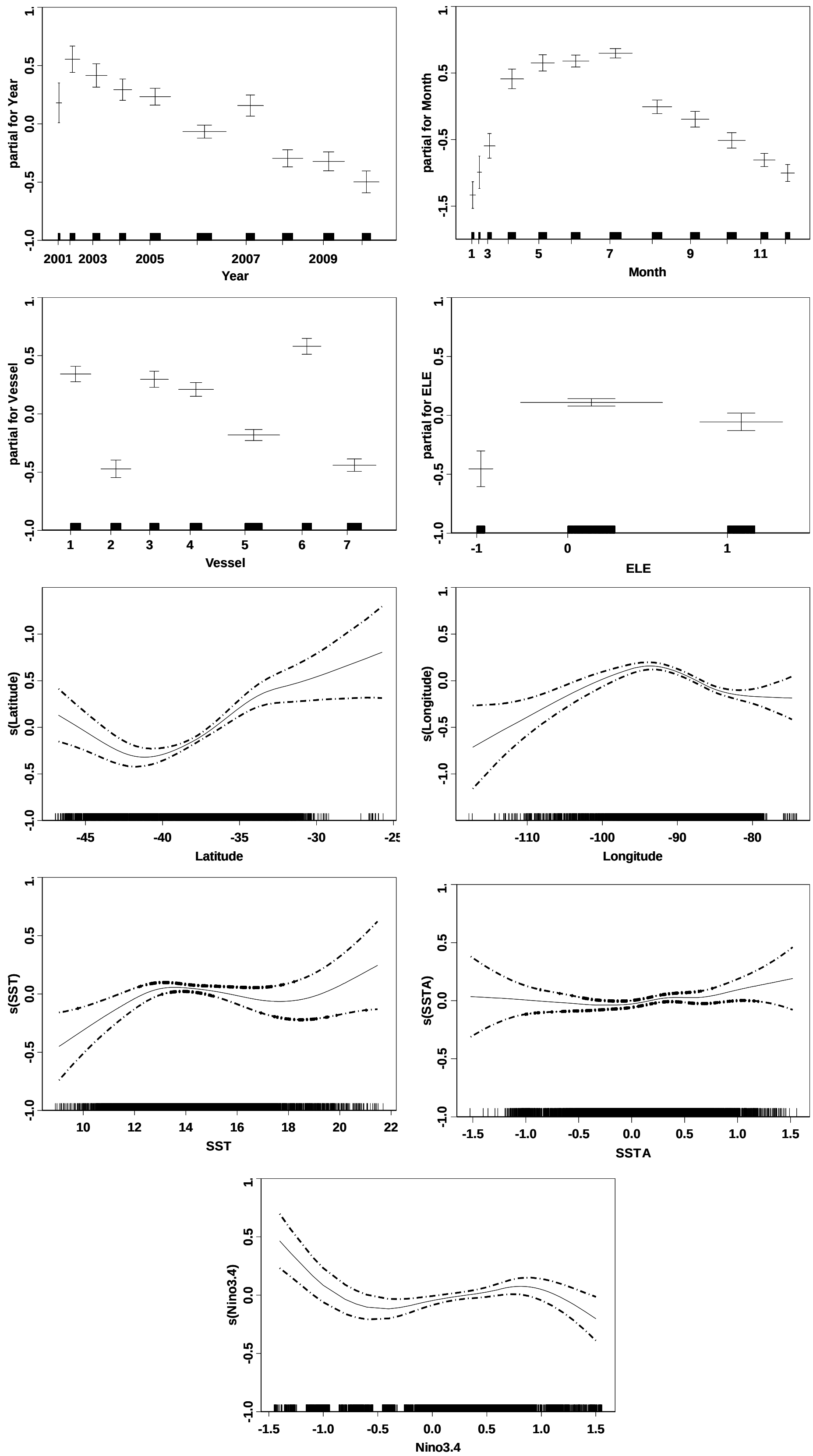

Fig.4. Impact of the main effects on the log-tranformed catch-rate of Chilean jack mackerel by Chinese trawl fishery derived from the GAM analysis during 2001–2010 in the the Southeast Pacific Ocean.

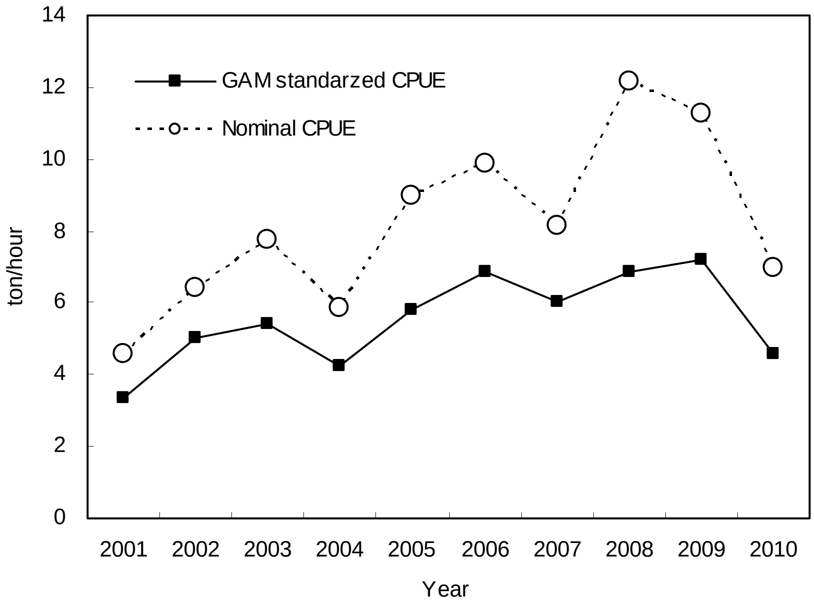

Fig.5. GAM standardized and nominal CPUE of Chilean jack mackerel

Fig. 4 and Table 4 summarized the results of GAM analysis. Stepwise GAM model building using the AIC, Pseudo R2 and Pseudo adjusted R2 concluded that all the explanatory variables were significant ($p<0.05$). With the explanatory variables added in the GAM model one by one, the residual deviance and AIC reduced significantly, in contrast, the cumulative variance explained, Pseudo R2 and Pseudo adjusted R2 were increased persistently. All these indicated the final model with the nine variables provided better fits to catch-data.

The final GAM model explained 30.60% of the variance in Chilean jack mackerel CPUE (Table 4). GAM analysis indicated that Month was the most influential effect which explained 16.58% of the deviance in Chilean jack mackerel CPUE. Vessel (6.01%), Year (3.33%) and Longitude (1.63%) were the next most important parameters, while ELE (0.91%), Latitude (0.85%), Nino3.4 index (0.74%), SST(0.46%) and SSTA (0.10%) provided minor contribution for the explained deviance .

The impacts of temporal factor Month on catch rates was the foremost factor and contributed 54.2% of the total reduction in deviance. Catch rate was low in the first three months. From April, catch rate increased stably and revealed peaks in July, followed by a monotonic decrease. The plot for factor Month indicated that April to September was the most favorable fishing season (Fig.7). The estimated Year effect showed a declining trend in catch rates from 2002 and 2010, except for a small rise in 2007. The CPUE reached the lowest value in 2010 during the ten years. The estimated Year effect in 2001 was biased resulted by small sample sizes of the catch rate data because only three trawler operated brokenly during the second half of 2011 (Fig.4). Annual variation of nominal Chilean jack mackerel CPUE compared to the GAM standardized CPUE differs obviously with the lower standardized indices (Fig.5). By and large, the annual standardized CPUE fluctuated with an increased trend from 2001 to 2009, and dropped sharply in 2010 (Fig.5).

The fishing performance predictor Vessel explained 6.0% of total deviance and the next most predominant factors in influencing CPUE. The same type trawler has the similar impact on Chilean jack mackerel catch rate (Table 1 and Fig.4). Vessel 1 and 6 was the most successful in catching Chilean jack mackerel, next was Vessel 3 and 4, and Vessel 5. Catch rate of Vessel 2 and 7 was the lowest might be connected with the smallest engine power and hold capacity.

Regarding Latitude, the GAM plot showed U-shaped curve, the lowest catch rate were found at 41°S, from which CPUE values increased northward and southward. Although catch rate at north of 31°S was still increased and very high, longitudinal trend for it was unclear because the fewer data points leads to greater uncertainty .The GAM plot for Longitude suggested catch rate reached peaks at 96°W, and decreased in easterly and westerly direction (Fig. 4).

Catch rate of Chilean jack mackerel related to SST (Fig.4) fluctuated throughout the temperature range. Higher catch rate values were observed in colder water with temperatures from 12 to 16□, and 78.9% of the fishing hours were deployed during this temperature range. SSTA was the minimum statistically significant factor in GAM analysis. Catch rate showed an increased trend from the negative SSTA to positive SSTA in general with small fluctuation. Catch rate was higher with SSTA range 0.2-1.0□.

The result of GAM analysis showed that El Niño and La Niña events also impact catch rate of Chilean jack mackerel. The abundance of Chilean jack mackerel was highest in normal years, while in El Niño or La Niña years, it decreased significantly. Nino3.4 index was another variable to analyze the impact of El Niño and La Niña events on Chilean jack mackerel catch rate. Catch rate was fluctuated with Nino3.4 index and could be divided into three range corresponded with La Niña, normal and El Niño years. Range a, SSTA from -1.5 to -0.5 (La Niña yeas); range b, SSTA from -0.5 to 0.5 (normal years); range c, SSTA from 0.5 to 1.5 (El Niño years). The change trend of catch rate was very similar between the two variables in GAM analysis, except for in range a of SSTA and La Niña yeas (Fig.4),which might relate to the lower density of data points leads to greater standard error ranges or the difference between Nino3.4 index and definition of La Niña event (Nino3.4 index is less than -0.5 for three months in a row at least) .

References

Campbell, R., 2004. CPUE standardization and the construction of indices of stock abundance in a spatially varying fishery using general linear models. Fish. Res. 70, 209-227.

Chambers, J. M., Hastie, T. J., 1997. Statistical Models. London: Chapman and Hall.

Cubillos L. A., Paramo J., Ruiz P., Núñez S., Sepúlveda A., 2008.The spatial structure of the oceanic spawning of jack mackerel (*Trachurus murphyi*) off central Chile (1998–2001). Fish. Res. 90, 261-270.

Damalas, D., Megalofonou, P., and Apostolopoulou, M., 2007. Environmental, spatial, temporal and operational effects on swordfish (*Xiphias gladius*) catch rates of eastern Mediterranean Sea longline fisheries. Fish. Res., 84, 233-246.

Elizarov, A.A., Grechina, A.S., Kotenev, B.N., Kuzetsov, A.N., 1993. Peruvian jack mackerel, *Trachurus symmetricus murphyi*, in the open waters of the South Pacific. J. Ichthyol. 33, 86-104.

Hilborn, R., and Walters C. J. 2001. Quantitative fisheries stock assessment choice, dynamics and uncertainty. Boston: Kluwer Academic Publishers.

Guisan, A., Edwards Jr., T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. modell. 157, 89-100.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York.

Maunder, M.N., Langley, A. D., 2004. Integrating the standardization of catch-per-unit-of-effort into stock assessment models: testing a population dynamics model and using multiple data types. Fish. Res. 70, 389-395.

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70, 141-159.

McCullough, p., Nelder, J. A., 1989. Generalized linear models, 2nd ed. New York: Chapman and Hall.

Nelder, J.A.,Wedderburn,R.W.M., 1972. Generalised linear models. J. R. Statist. Soc. A 137, 370-384.

Nishida, T., and Chen, D. G. 2004. Incorporating spatial autocorrelation into the general linear model with an application to the yellowfin tuna (*Thunnus albacares*) longline CPUE data. Fish. Res. 70, 265-274.

Quinn, G. P., Keough, M. J., 2002. Experimental design and data analysis for biologists. Cambridge: Cambridge University Press.

Su N. J. , an Yeh, S. Z., Sun, C. L., Punt , A. E., Chen, Y, Wang, S. P., 2008. Standardizing catch and effort data of the Taiwanese distant-water longline fishery in the western and central Pacific Ocean for bigeye tuna, *Thunnus obesus*. Fish. Res. 90, 235-246.

Venables, W. N., Dichmont, C. M., 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fish. Res. 79, 319-337.