POLISH MARITIME RESEARCH Special Issue 2017 S2 (94) 2017 Vol. 24; pp. 140-146
10.1515/pomr-2017-0076

YELLOWFIN TUNA (THUNNUSALBACARES) FISHING GROUND FORECASTING MODEL BASED ON BAYES CLASSIFIER IN THE SOUTH CHINA SEA

ZHOU Wei-feng[1)], **Ph.D.**
LI An-zhou[1, 3)]
JI Shi-jian[1)]
QIU Yong-song[2)]
[1)] Key Laboratory of East China Sea and Oceanic Fishery Resources Exploitation and Utilization, Ministry of Agriculture, China, 200090 Shanghai, China
[2)] South China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510300, China
[3)] College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China

ABSTRACT

Using the yellowfin tuna (Thunnusalbacares,YFT)longline fishing catch data in the open South China Sea (SCS) provided by WCPFC, the optimum interpolation sea surface temperature (OISST) from CPC/NOAA and multi-satellites altimetric monthly averaged product sea surface height (SSH) released by CNES, eight alternative options based on Bayes classifier were made in this paper according to different strategies on the choice of environment factors and the levels of fishing zones to classify the YFT fishing ground in the open SCS. The classification results were compared with the actual ones for validation and analyzed to know how different plans impact on classification results and precision. The results of validation showed that the precision of the eight options were 71.4%, 75%, 70.8%, 74.4%, 66.7%, 68.5%, 57.7% and 63.7% in sequence, the first to sixth among them above 65% would meet the practical application needs basically. The alternatives which use SST and SSH simultaneously as the environmental factors have higher precision than which only use single SST environmental factor, and the consideration of adding SSH can improve the model precision to a certain extent. The options which use CPUE's mean ± standard deviation as threshold have higher precision than which use CPUE's 33.3%-quantile and 66.7%-quantile as the threshold

Keywords: Bayes classifier, South China Sea,yellowfin tuna, fishing ground forecasting

INTRODUCTION

Yellowfin tuna (*Thunnusalbacares*) is one of the economic important targets harvested primarily by means of the longline catch, generally accounting for a large proportion of the gross catch in the South China Sea tuna fisheries [1,2,3]. So far, the studies on yellowfin tuna in the South China Sea mainly concentrated on biology such as basic biology [4], population genetic structure [5], etc. Yet hardly any specific researches have been conducted on yellowfin tuna fishing grounds forecasting in the South China Sea.

Statistical methods are widely used in tuna fishing ground forecasting in other sea areas that need sufficient historical data to analyze a relationship between the oceanic environment and the fish catch and then to forecast the future condition of the fishing ground with the relationship, such as linear regression model (LRM) [6], time series analysis [7], spatial overlay analysis [8], geostatisticalanalysis [9], Bayes probability model[10,11,12,13], etc. Among them, the Bayes probability model has a solid theoretical foundation of Mathematics. It uses the historical statistical data of fish catch and the specific environmental factors to figure out the prior probability and conditional probability, and then quantitatively evaluates the category of the fishing grounds it forecasted by the posterior probability. The forecast results not only reflect the fishing experiences of the fishermen but also consider the oceanic environmental influence on fishing grounds. However, due to the variability, complexity of oceanic environment and the instability of spatial-temporal

distribution of fishery resources, a single forecast model build plan can not completely fit with any sea areas and any fish species, the differences of the interior parameter setting such as the choice of the environmental factors and the classification strategy of the fish zones in the Bayes model must bring different forecast results.

In this study, the oceanic environmental data from satellite remote sensing, and the historical fish catch data were used to build eight alternative models based on Bayesclassifier model to forecast the yellowfin tuna fishing groundin the South China Sea in 2011. The classification results ofthe eight options were compared with the actual ones for validation and how different the options impact on classification results and precision were analyzed.

MATERIALS AND METHODS

FISHERY AND ENVIRONMENTAL DATA

The fishery data provided by the Western Central Pacific Fisheries Commission (WCPFC) at a 5°×5° spatial resolution and a monthly time resolution include operation time (year/month), position (latitude/longitude), and fish catch statistical information of each fish species (hooks, catches, numbers). The latitude and longitude recorded in the data represent the latitude and longitude of the southwest corner of a 5° grid. The study area covers the South China Sea and adjacent waters (105°E-125°E and 0°-25°N) and the data within the area were extracted.

Monthly sea surface temperature (SST) data were compiled from the optimally-interpolated (OI) at 1° spatial resolution, generated by the Climate Prediction Center(CPC), NOAA. Monthly multi-satellites (Topex/Poseidon, JASON-1, Jason-2, Envisat, ERS-1, ERS-2and Cryosat-2) merged sea surface height (SSH) data at 0.25° spatial resolution was downloaded from the Satellite Oceanic Data Center, the Centre National d'EtudesSpatiales(CNES). There was a difference among the environmental data and fishery data described above, so an operation of the environmental data resampling to a 5°×5° spatial resolution for matching up with the fishery data was conducted. Here the historical catch data from 2000 to 2010 were thrown into build models to forecast the potential fishing ground distributions in the year of 2011.And we conducted a validation for the forecast results by the real ones in 2011.

METHODS

Compute of catch per unit of effort

Catch per unit of effort (CPUE) is a value that can be used to represent fishery resource abundance in a statistical unit[14]. Here it was the quantitative index of fishing grounds. The equation of CPUE in every 5°×5° fishing zone grid was as follow:

$$CPUE_{(i,j)} = \frac{N_{\mathrm{fish}(i,j)} \times 1000}{N_{\mathrm{hook}(i,j)}} \quad (1)$$

where, $CPUE_{(i,j)}$, $N_{\mathrm{fish}(i,j)}$and $N_{\mathrm{hook}(i,j)}$ are the CPUE, fish catch number and fish hook number of the fishing zone grid at the i-th longitude and the j-th latitude, respectively.

Forecast model building

Bayes classifier model was used as the model to forecast and classify the fishing grounds in the South China Sea. There are several practices of scientific literature [11, 12, 13] can be referred to for the detail of the model framework and here it would not be repeated. In fact, how to choose the environmental factors and the classification strategy of the fishing zones are the two keys to the model. So, eight alternative options were designed according to different combinations of the two settings as follows (Fig. 1):

Option 1: use SST only as the environmental factor, take its real value as the model parameter and divide fishing zones into two classes by the average of the historical CPUEs;

Option 2: use SST and SSH as the environmentalfactors, take their first principal component as the model parameter and divide fishing zones into two classes by the average of the historical CPUEs;

Option 3: use SST only as the environmental factor like the option 1 and divide fishing zones into two classes by the median of the historical CPUEs;

Option 4: use SST and SSH as the environmental factorslike the option 2 and divide fishing zones into two classes by the median of the historical CPUEs;

Option 5: use SST only as the environmental factor like the option 1 and divide fishing zones into three classes by the average±standard deviation of the historical CPUEs;

Option 6: use SST and SSH as the environmental factors like the option 2 and divide fishing zones into three classes by the average±standard deviation of the historical CPUEs;

Option 7: use SST only as the environmental factor like the option 1 and divide fishing zones into three classes by the 33.3%-quantile and 66.7%-quantile of the historical CPUEs;

Option 8: use SST and SSH as the environmental factorslike the option 2 and divide fishing zones into three classes by the 33.3%-quantile and 66.7%-quantile of the historical CPUEs.

Principal component analysis

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components [16]. Bayes classifier model must be built based on anassumption that the every environmental factor independently impacts on thefishing ground. If there are a relationship among the factors, the classification accuracy of the model would be effected[15]. Therefore, it is necessary to extract the first principal component of SST and SSH before the option 2, 4, 6 and 8 were executed. The specific steps were as follows:

(1) Standardize the variables that need to be analyzed to solve the problem of different kinds of data. Here z-score standardization was adopted and the equation was as follow:

$$y_i = \frac{x_i - \bar{x}}{s} \quad (2)$$

where, $y_i$was the standardized new sequence, and sare the average and the standard deviation of the sequence , respectively;
(2) Compute correlation matrix R with the new sequence y_i ;
(3) Solve the characteristic equation $|\lambda E - R| = 0$, obtain the eigenvalue λ_i,the eigenvectorliand then sort λiin descending order;
(4) Compute the contribution rates of the principal components and the accumulative contribution rates;
(5) Obtain the principal components.
The steps listed above have been realized by programming in MATLAB software.

Compute of fishing ground probability

The prior probability of fishing ground was computed based on an assumption that the more some kind of fish zone appears historically the greater the probability of the kind of fish zone is. So the formula was as follow:

$$P(h_i) = \frac{N_i}{N_{\text{total}}} \times 100\% \quad (3)$$

Where,h_i represents the event that fishing zone was defined as the i-th class, $P(h_i)$was the prior probability of the situation when h_i eventhappens without the consideration of environmental condition, N_i was the number of hieventswhich happens in the fishing zone, and N_{total} was the total number of samplesof the fishing zone historically.

The conditional probability refers to the occurrence probability of some kind of environmental condition in a situation when the fishing zone was defined as some kind of class. So the formula was as follow:

$$P(e \,/\, h_i) = \frac{M_i}{N_i} \times 100\% \quad (4)$$

where, $P(e \,/\, h_i)$ was the occurrence probability of the environmental conditione, N_i was the occurrence number of the situation whenhievent happens and M_i was the occurrence number of the environmental condition e in a situation when h_i event happens.

Finally, according to the principle of Bayes probability, the posterior probability of each fish zone can be calculated by the formulas as follow:

$$P(h_i|e) = \frac{P(e|h_i) \times P(h_i)}{\sum_i^n P(e|h_i) \times P(h_i)} \quad (5)$$

The class that the maximum of the posterior of each fish zone corresponding can be regard as the forecasted fish zone class.

Model validation

The validation of the 8 alternative options was subsequently implemented using independent sets of monthlyfishery data in 2011. The error matrix was built between the forecasted results and actual ones to calculate the generalaccuracy as follow:

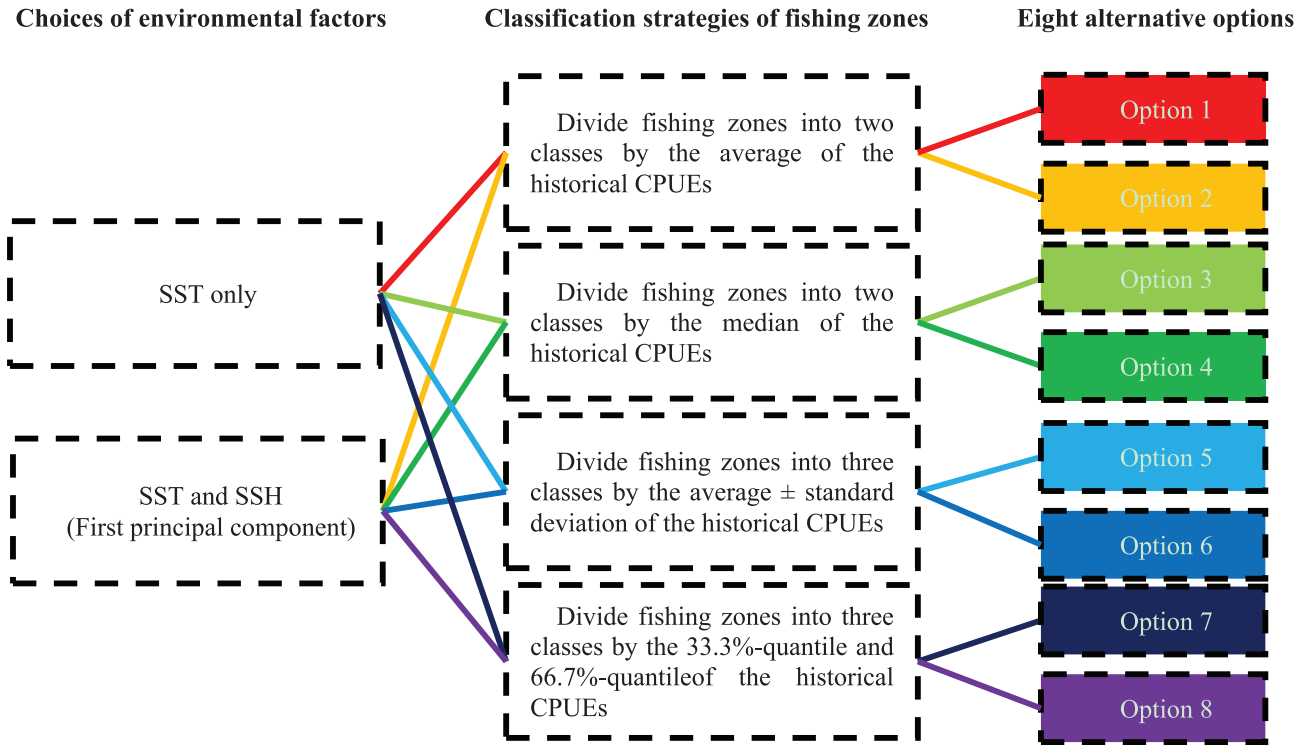

Fig. 1. Eight alternative options based on Bayes classifier model

Tab.1 Forecast accuracies of the eight options of Bayes classifier model

Options	High-CPUEs (Number/Rate)	Middle-CPUEs (Number/Rate)	Low-CPUEs (Number/Rate)	Undefined (Number/Rate)	General (Number/Rate)
1	53/84.1%	-	67/63.8%	3/1.8%	120/71.4%
2	57/90.5%	-	69/65.7%	-	126/75%
3	52/82.5%	-	67/63.8%	3/1.8%	119/70.8%
4	56/88.9%	-	69/65.7%	-	125/74.4%
5	1/20%	94/86.2%	18/36%	3/1.8%	113/66.7%
6	1/20%	100/91.7%	14/28%	-	115/68.5%
7	19/63.3%	22/43.1%	56/64.4%	3/1.8%	97/57.7%
8	19/63.3%	20/39.2%	68/78.2%	-	107/63.7%

$$p_c = \sum_{i=1}^{n} p_{ii}/p \quad (6)$$

where, p_c was the general accuracy; p_{ii} was the number of fishing zones which have been correctly classified; n was the number of classes and p was the number of total fishing zone samples[17].

RESULT

THE RESULTS OF THE EIGHT OPTIONS

In 2011, there were 168 fishing zone samples used for classification. By dividing fishing zones into two classes by the average of the CPUEs, there were 63 high-CPUEs (37.5%) and 105 low-CPUEs (62.5%). By dividing fishing zones into two classes by the median of the CPUEs, there were 63 high-CPUEs (37.5%) and 105 low-CPUEs (62.5%), too. By dividing fishing zones into three classes by the average±standard deviation of CPUEs, there were 9 high-CPUEs (5.3%), 109 middle-CPUEs (64.9%) and 50 low-CPUEs (29.8%). By dividing fishing zones into three classes the 33.3%-quantile and 66.7%-quantile of CPUEs, there were 30 high-CPUEs (17.8%), 51 middle-CPUEs (30.4%) and 87 low-CPUEs (51.8%).

The forecast accuracies of the eight alternative options of Bayes classifier model was shown in Tab1. The general accuracy of the option 1 was at 71.4%, in which high-CPUEs and low-CPUEs were at 84.1% and 63.8%, respectively. The general accuracy of the option 2 was at 75%, in which high-CPUEs and low-CPUEs were 90.5% and 65.7%, respectively. The general accuracy of the option 3 was at 70.8%, in which high-CPUEs and low-CPUEs were 82.5% and 63.8%, respectively. The general accuracy of the option 4 was at 74.4%, in which high-CPUEs and low-CPUEs were 88.9% and 65.7%, respectively. The general accuracy of the option 5 was at 66.7%, in which high-CPUEs, middle-CPUEs and low-CPUEs were 20%, 86.2% and 36%, respectively. The general accuracy of the option 6 was at 68.5%, in which high-CPUEs, middle-CPUEs and low-CPUEs were 20%, 91.7% and 28%, respectively. The general accuracy of the option 7 was at 57.7%, in which high-CPUEs, middle-CPUEs and low-CPUEs were 63.3%, 43.1% and 64.4%, respectively. The general accuracy of the option 8 was at 63.7%, in which high-CPUEs, middle-CPUEs and low-CPUEs were 63.3%, 39.2% and 78.2%, respectively. In addition, there were 3 samples (accounting for 1.8%) cannot be classified by options 1, 3, 5 and 7.

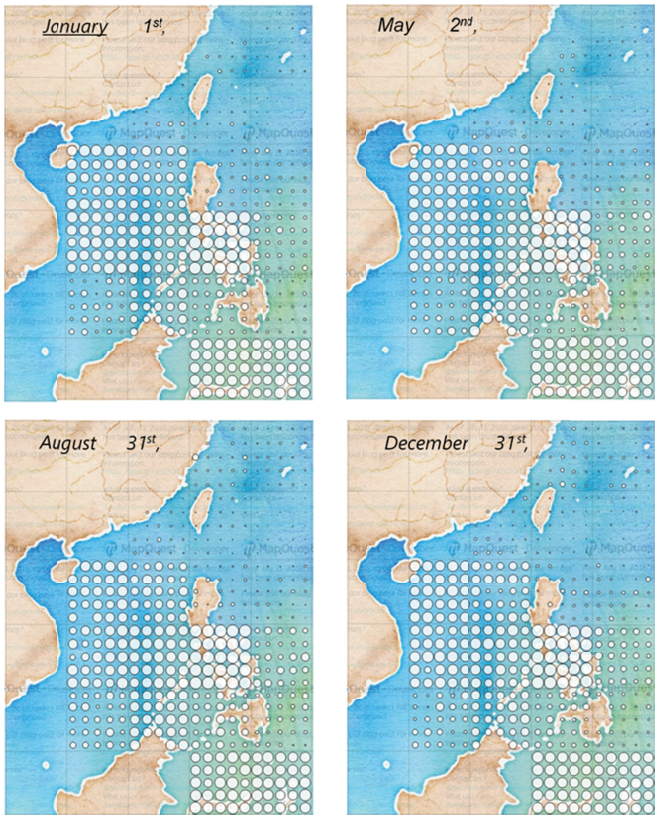

Fig.2The predicting results of the Option2 of the Bayesian classifier model

INFORMATION SYSTEM IMPLEMENTATION OF THE BAYESIAN CLASSIFIER MODEL BASED ON THE FIRST PRINCIPAL COMPONENT OF SST AND SSH

The results from Tab 1 implies that Bayes classifier model of Option 2 has the highest general accuracy. So we choose the Option 2 as the scheme of marine environment factors' selection and fishing zones' classification to build the forecasting model based on Bayes Classifier rule. And the tuna fishing ground forecasting information service system for the open South China Sea has been set up to realize the model's computing and share the forecasting result. This information system adopts three-tier architecture. And it includes four sectors such as data acquisition, fishing ground forecasting model, forecasting information publishing service and custom interface [18]. The system uses Java as the development language coded in Eclipse. The system calls the forecasting model implemented by Matlab script to computing every fishing zone's probability of high-CPUE. The forecasting results are updated and published by GeoServer, an open source server for sharing geospatial data. Fig 2 shows the forecasting results of different times in the open South China Sea about the every fishing zone's probability of high-CPUE using this information system.

DISCUSSION

The optionswere classified as the Group 1, which used SST only as the environmental factor and took its real value as the model parameter, have favorable general accuracy except the option 7. The fishing ground distribution of yellowfin tuna is related to SST, which has been studied frequently and deeply both at home and abroad [19, 20]. Fan and Zhou, etc. used SST remote sensing data to build forecast models based on Bayes theory in the Pacific Ocean and the Indian Ocean and the returned accuracy is at 65%~70%[11,13]. It means that the single environmental factor SST can be used as the model forecast parameter. The options 2, 4, 6 and 8, which were regarded as the Group 2, used SST and SSH as the environmental factors and took their first principal component as the model parameter.The Group 2 has higher accuracies compared with the Group 1. It explains that the factor SSH doesn't weaken the contribution of SST. It can be inferred,the models that added SSH factor can improve the general accuracies to a certain extent and the SSH factor have a certain influence on the fishing ground distribution of yellowfin tuna. The research of Wang et al.[21]shows that SSH affected the CPUE distribution of the Central West yellowfin tuna fishing grounds significantly. The annual CPUE is high in the areas where the SSH values are high and it is favorable to conduct fishing operations in the areas. The oceanic environmental factors are not independent, and there is a certain relationship among various factors[22]. Using the first principal component of the two factors instead of themselves as input parameters makes the forecast more accurate relatively.

A comparison of the 4 classification strategy has been conducted. The results showed that the options dividing fishing zones into two classes by the average were more accurate than by the median, and the options dividing fishing zones into three classes by the average±standard deviation were more accurate than by the 33.3%-quantile and 66.7%-quantile. Actually, to divide fishing zones by the median or the 33.3%-quantile and 66.7%-quantile, all were based on the assumption that the number of the fishing zones belonging to each class historically were equal. However, the assumption was inconsistent with the actual situation in the oceanic fishing operation. Generally, the fishermen have rich experience so that they tend to fish operation in the areas where CPUEs are higher. In other words, the CPUE in many years should follow the normal distribution or the skew distribution. The accuracies of high-CPUEs in options 5 and 6 were only at 20% with the number of high-CPUE fishing zones was only 9, which was unable to conclude that adopting the average±standard deviation could not identify high-CPUE fishing zones accurately.

In addition, the accuracy of forecast models was also associated with the time span of historical data. The time span of data should not be too long nor too short because the data was too old to reflect current real situation by considering the changes of the fishing capacity and resources. Certainly, if the time span of data was too short, it can lead to insufficient training and effect the reliability of classification results. Therefore, an assumption that the changes in the population size of the yellowfin tuna are little and the fishing operation is on the same level of fishing capacity in the South China Sea from 2001 to 2010 was accepted by this paper. Besides, there may be some inter-monthly differences between historical fishing data and environmental data. If so, the models could be performed less successfully by using constant relationship mode directly. So it is necessary to analyze the relationships of each month separately under the condition of sufficient data. The effective factor of model accuracy are far more than those mentioned above, it is a way that we only try our best to improve the accuracy can make the model perfect gradually.

CONCLUSION

We have used historical catch data and sea surface environmental factors (SST and SSH) to forecast and classify the fishing grounds in 2011 of yellowfin tuna in the open South China Sea with eight alternative options based on Bayes classifier model according to different strategies on the choice of environment factor and classification of fishing zones, and the forecast results were validated compared with actual fishing ground distribution. The results of validation showed that the accuracies of the eight options were 71.4%, 75%, 70.8%, 74.4%, 66.7%, 68.5%, 57.7% and 63.7% in sequence, the first to sixth among them above 65% would meet the practical application needs basically. The accuracies of options 7 and 8 under 65% need further improvement in classification strategy.

ACKNOWLEDGEMENT

The study is funded by The Key Technologies R&D Program of China under Contract No. 2013BAD13B06, Project 31602206 supported by National Natural Science Foundation of China, Natural Science Foundation of Shanghai under contract NO.16ZR1444700, Project NO.2016T05 Supported by Special Scientific Research Funds for Central Non-profit Institutes(East China Sea Fisheries Research Institute),andOpening Project of Scientific Observing and Experimental Station of Fishery Remote SensingofMinistry of Agricultureof China(OFSOESFRS201505)

BIBLIOGRAPHY

1. ZHANG Peng, YANG Li, ZHANG Xufeng, et al. 2010. The present status and prospect on exploitation of tuna and squid fishery resources in South China Sea. SOUTH CHINA FISHERIES SCIENCE, 6(1):68-74.

2. MENG Xiaomeng, YE Zhenjiang, WANG Yingjun. 2007. Review on fishery and biology of yellowfin tuna (Thunnusalbacares). South China Fisheries Science, 3(4):74-80.

3. JI Shijian, ZHOU Weifeng, CHENG Tianfei, et al. 2015. On the forecast and analysis of fishing grounds in the open South China Sea. Modern Fisheries Information, 2015(2):98-105.

4. FENG Bo, LI Zhonglu, HOU Gang. 2014. Biology and distribution of thunnusobesus and thunnusalbacresin the South China Sea. OceanologiaetLimnologiaSinica, 2014(4):886-894.

5. WANGZhongduo, GUO Yusong, YAN Yunrong, et al. 2012. Population genetics of tunas in South China Sea inferred from control regions. Journal of Fisheries of China, 36(2): 191-201.

6. Zagaglia C R, Lorenzzetti J A, Stech J L. 2004. Remote sensing data and longline catches of yellowfin tuna (Thunnusalbacares) in the equatorial Atlantic. Remote Sensing of Environment, 93(1–2):267-281.

7. Georgakarakos S, Koutsoubas D, Valavanis V. 2006. Time series analysis and forecasting techniques applied on loliginid and ommastrephid landings in Greek waters. Fisheries Research. 78(1):55-71.

8. Zainuddin M, Saitoh K, Saitoh SI. 2008. Albacore (Thunnusalalunga) fishing ground in relation to oceanographic conditions in the western North Pacific Ocean using remotely sensed satellite data. Fisheries Oceanography. 17(2):61-73.

9. YANG Xiaoming, DAI Xiaojie, ZHU Guoping. 2012. Geostatistical analysis of spatial heterogeneity of yellowfin tuna (Thunnusalbacares) purse seine catch in the western Indian Ocean. ActaEcologicaSinica, 32(15): 4682-4690.

10. CUIXuesen, TANG Fenghua, ZHANG Heng, et al. 2015. The Establishment of Northwest Pacific Ommastrephesbartramii Fishing Ground Forecasting Model Based on Naive Bayes Method. Periodical of Ocean University of China, 45(2):37-43.

11. ZHOU Weifeng, FAN Wei, CUIXuesen, et al. 2012. Fishing ground forecasting of bigeye tuna in the Indian Ocean based on Bayesian probability model.Fisheries Information and Strategy, 27(3):214-218.

12. CUI Xuesen, CHEN Xuedong, FAN Wei. 2007. Development of tuna fishing grounds prediction model and system. CHINESE HIGH TECHNOLOGY LETTERS. 17(1): 100-103.

13. FAN Wei, CHEN Xuezhong, SHEN Xinqiang. 2006. Tuna fishing grounds prediction model based on Bayes probability. JOURNAL OF FISHERY SCIENCES OF CHINA. 13(3): 426-431.

14. TIAN Siquan, CHEN Xinjun. 2010. Impacts of different calculating methods for nominal CPUE on CPUE standardization. Journal of Shanghai Ocean University, 19(2):240-245.

15. WANG Guocai. 2010. Research and application of Naive Bayesian classifier[dissertation]. Chongqing JiaotongUniversity.

16. ZHU Xingyu. 2011. SPSS multivariate statistical analysis method and its application. Beijing: Tsinghua University Press.

17. ZHAO Yingshi. 2003. The principle and method of remote sensing application. Beijing: Science Press, 206-207.

18. JI Shijian, ZHOU Weifeng, XU Hongyun, et al. A WebGIS application: Tuna fishing ground forecasting information service system for the open South China Sea, 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, 2016, pp. 3628-3631.

19. Stech J L, Zagaglia C R, Lorenzzetti J A. 2004. Remote sensing data and longline catches of yellowfin tuna (Thunnusalbacares) in the equatorial Atlantic. Remote Sensing of Environment, 93(1-2):267-281.

20. Lan, Kuo-Wei, Evans, Karen, et al. 2013. Effects of climate variability on the distribution and fishing conditions of yellowfin tuna (Thunnusalbacares) in the western Indian Ocean. Climatic Change, 119(1):63-77.

21. WANG Shaoqin, XU Liuxiong, ZHU Guoping, et al. 2014. Spatial-temporal profiles of CPUE and relations to environmental factors for yellowfin tuna Thunnus albacores from purse-seine fishery in Western and Central Pacific Ocean. Journal of Dalian Fisheries University, 2014(3):303-308.

22. He R, Ke C, Timothy M, et al. 2010. Mesoscale variations of sea surface temperature and ocean color patterns at the Mid-Atlantic Bight shelfbreak. Geophysical Research Letters, 37(9): 493-533.

CONTACT WITH THE AUTHOR

ZHOU Wei-feng,Ph.D.

FacultyofEast China Sea Fisheries Research Institute,
Chinese Academy of Fishery Sciences
200090 Shanghai, China

e-mail: zhouwf@ecsf.ac.cn

CHINA