SCRS/2017/057 Collect. Vol. Sci. Pap. ICCAT, 74(4): 1664-1674 (2017)

EVALUATION OF ENVIRONMENTAL CONDITIONS AS PREDICTORS FOR MAKO SHARK CPUE USING GENERALIZED LINEAR MIXED MODELING AND QUANTILE REGRESSION

Halie OFarrell[1,2] and Elizabeth Babcock[1,3]

SUMMARY

*Environmental conditions were evaluated for their influence on catch per unit effort (CPUE) of shortfin mako (*Isurus oxyrinchus*). Catch rates of shortfin mako were calculated from the US pelagic longline observer program (1992-2016) using a generalized linear mixed model (GLMM) with a delta-lognormal approach. The GLMM analysis included consideration of the following environmental variables as predictor variables: sea surface height, sea surface temperature, and bathymetry. The addition of environmental predictor variables resulted in an index that spans 2003-2012. The final index was used to predict average catch per unit effort (CPUE) based on environmental conditions. The two portions of the delta-lognormal approach retained different suites of variables with sea surface temperature and bathymetry retained to predict proportion of positive sets while bathymetry was retained to predict the CPUE of positive catches. Quantile regression was also performed to evaluate whether environmental variables can predict spatial areas with high CPUE. As with the delta approach, environmental data were used to predict conditions that favor high CPUE. Maps generated from both the approaches will later be used for determining mako shark habitat for a spatial management strategy evaluation.*

RÉSUMÉ

*Les conditions environnementales ont été évaluées afin de déterminer leur influence sur les captures par unité d'effort (CPUE) du requin-taupe bleu (*Isurus oxyrinchus*). Les taux de capture du requin-taupe bleu ont été calculés à partir du programme d'observateurs à bord de palangriers pélagiques des États-Unis (1992-2016) en utilisant un modèle mixte linéaire généralisé (GLMM) avec une approche delta-lognormale. L'analyse de la GLMM prenait en compte les variables environnementales suivantes comme variables de prédiction : hauteur de la surface de la mer, température de la surface de la mer et bathymétrie. L'ajout de variables de prédiction environnementale a entraîné un indice qui s'étend de 2003 à 2012. L'indice final a été utilisé pour prédire la capture par unité d'effort (CPUE) moyenne basée sur des conditions environnementales. Les deux parties de l'approche delta-lognormale ont retenu différentes suites de variables, la température de surface de la mer et la bathymétrie étant retenues pour prédire la proportion d'opérations positives tandis que la bathymétrie a été retenue pour prédire la CPUE de prises positives. Une régression quantile a également été effectuée afin d'évaluer si les variables environnementales peuvent prévoir des zones avec une CPUE élevée. Comme avec l'approche delta, des données environnementales ont été utilisées pour prédire les conditions qui favorisent une CPUE élevée. Les cartes obtenues des deux approches seront utilisées ultérieurement pour déterminer l'habitat du requin-taupe bleu aux fins de l'évaluation d'une stratégie de gestion spatiale.*

RESUMEN

*Se evaluaron las condiciones medioambientales para determinar su influencia en la captura por unidad de esfuerzo (CPUE) del marrajo dientuso (*Isurus oxyrinchus*). Se calcularon las tasas de captura del marrajo dientuso a partir del programa de observadores de palangre pelágico de Estados Unidos (1992-2016) utilizando un modelo lineal generalizado mixto*

[1] Department of Marine Biology and Ecology, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149. USA
[2] hofarrell@rsmas.miami.edu
[3] ebabcock@rsmas.miami.edu

(GLMM) con un enfoque delta-lognormal. El análisis GLMM incluía la consideración de las siguientes variables medioambientales como variables de predicción: altura de la superficie del mar, temperatura de la superficie del mar y batimetría. Añadir las variables de predicción medioambientales tuvo como resultado un índice que abarca desde 2003 hasta 2012. El índice final se utilizó para predecir la CPUE media basándose en las condiciones medioambientales. Las dos partes del enfoque delta lognormal retenían diferentes conjuntos de variables y la temperatura de la superficie del mar y la batimetría se retuvieron para predecir la proporción de lances positivos, mientras que la batimetría se retuvo para predecir la CPUE de capturas positivas. Se llevó a cabo también una regresión por cuantiles para evaluar si las variables medioambientales pueden predecir áreas con una CPUE elevada. Al igual que con el enfoque delta, se utilizaron los datos medioambientales para predecir las condiciones que favorecen una CPUE elevada. Los mapas generados a partir de ambos enfoques se utilizarán posteriormente para determinar el hábitat del marrajo dientuso para una evaluación de la estrategia de ordenación espacial.

KEYWORDS

Catch/effort, commercial fishing, long lining, pelagic fisheries, shark fisheries, by-catch, logbooks, observer programs, environmental factors, mako sharks

1. Introduction

The objective of this paper is to evaluate how environmental conditions influence mako shark catch per unit effort (CPUE) and determine if they can be used to predict CPUE in time and space, especially high values of CPUE. A generalized linear mixed modeling (GLMM) approach is used to predict mean CPUE, similar to the standardization methods used to generate abundance indices for past assessments (Cortés 2013; 2007). However, this study has a particular interest in determining what conditions favor particularly high CPUE, where much of the fishing mortality occurs. The use of quantile regression to focus on the upper tail of the CPUE distribution, rather than the mean, will reveal what environmental conditions predict higher CPUE. Quantile regression is a method appropriate for abundance data with non-linear, non-symmetric, heterogenous scatter in response to a gradient (Fukunaga et al. 2016, Anderson 2008, Cade and Noon 2003, Koenker and Bassett 1978). It addresses the scatter by allowing the slopes to differ for the different parts of the abundance distribution (Koenker and Bassett 1978). It has been used in ecology particularly to test environmental predictor variables (Fornaroli et al. 2016, Fukunaga et al. 2016, Anderson 2008, Cade and Noon 2003). Application of quantile regression to the upper extreme allows the independent environmental variables to be viewed more easily as boundaries to suitable habitat (Fornaroli et al. 2016). This is the first step in a larger body of work aimed at determining mako habitat and distribution for a management strategy evaluation.

2. Methods

Environmental conditions were considered within the GLMM and quantile regression as explanatory variables. The GLMM predicts mean CPUE while the quantile regression predicts CPUE at the defined percentile. As this is a preliminary study, only sea surface temperature, sea surface height, and bathymetry were considered. Future studies will incorporate distance from front and other oceanographic variables. Catch and effort data was obtained from the US pelagic longline observer program (1992-2016) (Brown and Minnett 1999) while weekly sea surface temperature composites (2003-2016) and daily sea surface height (1992-2012) (Ducet et al 2000) were downloaded from the NOAA CoastWatch satellite database and bathymetry was downloaded from the Scripps Institute of Oceanography Geodesy satellite database (Smith and Sandwell 1997). The catch dataset was reduced to those events in time for which environmental variables were available so there would be no null values. This resulted in a dataset spanning 2003 to 2012. All of the following analyses were conducted in R version 3.1.2 using the MASS (Venables and Ripley 2002) and quantreg (Koenker 2015) libraries.

Mean catch rates were modeled using a GLMM with a delta-lognormal approach. In this approach the proportion of positive catch assumes a binomial error distribution while the mean catch rate for positive observations assumes a lognormal error distribution so they are considered separately then combined (Maunder and Punt 2004, Ortiz and Arocha 2004). The predicted mean CPUE is then calculated as the probability of presence from the binomial model multiplied by the mean CPUE when present from the lognormal model. The variables year, fishing area, quarter, hooks between floats, and the use of lights were initially considered in all models along with all combinations of possible environmental variables. Interactions were not considered. All environmental variables were converted to factors (factor levels provided in **Table 1**). After viewing diagnostic results, models excluding variables that explained less than 1% of the total deviance were considered. The Bayesian information criteria (BIC) was used to select the final binomial and lognormal models because it tended to prefer simpler models and more complex models had unreasonably high coefficients of variation.

The quantile regression method analyzes patterns in the user-specified quantile rather than the mean. For this analysis, the interest was focused on particularly high values of CPUE; the 99^{th} percentile was analyzed looking at different combinations of the environmental variables as numerical variables, including the interactions between all possible pairs in addition to quadratic terms for each variable, with the response variable being (log CPUE + 0.1). Gear variables use of light and hooks between floats were also considered as it is assumed that gear differences will bias catch rates. All variables except use of light were considered to be continuous and not converted to the factors used in the GLMM. Explanatory variables and terms were retained based on their significance, calculated using the default settings of the summary function in the quantreg R library. This method is based on a t statistic, with the standard error calculated using a Huber sandwich estimate of the standard error (Koenker 2015).

Comparing the results of the two methods will give an indication of how environmental conditions affect both mean catch rates and particularly high catch rates.

3. Results

In the GLMM analysis of the observer data, as determined by the BIC, variables retained for the proportion of positive sets were year, area, quarter, use of light, hooks between floats, sea surface temperature, and bathymetry (**Table 2**). Residuals for the proportion of positive catch are shown in **Figure 1**. Model coefficients for bathymetry indicate a negative effect on the proportion of positive sets except for bottom depths shallower than 800m (**Figure 2a**). Overall with decreasing bottom depth there is a “U” shape with the greatest coefficient magnitude at the 2400m-3200m bin (Figure 2a). Mako catches are least likely in waters 2400m-3200m and most likely in waters less than 800m (Figure 2a). Model coefficients for sea surface temperature indicate a positive effect on the proportion of positive catch except for the 25-30 degree bin and a higher probability of having a positive catch in colder waters (**Figure 2b**).

The variables retained by the BIC for the lognormal model of positive catch were year, area, quarter, hooks between floats, and bathymetry (**Table 2**). Residuals for the positive catch model have increasing variance with higher predicted values indicating that the assumed error distribution may not be the most appropriate (**Figure 3**). Model coefficients for bathymetry indicate that, using the deepest category as a reference level, all shallower levels have a positive effect on CPUE, and the CPUE tends to increase as the bottom depth gets shallower, with the highest CPUE in bottom depths shallower than 800m (**Figure 4**).

In the quantile regression, the gear variables hooks between floats and use of lights were never found to be significant. All three environmental variables tended to be significant when included as a linear term while only bathymetry tended to be significant as a quadratic term and the only significant interaction was that between sea surface temperature and bathymetry. Model exploration showed that when analyzed separately sea surface temperature and sea surface height tended to be significant. However, possibly because sea surface height and sea surface temperature are moderately correlated (**Table 3**), all models that included both variables found that only one was significant. For this reason, a “final” model was not selected but rather several models were selected for discussion; one including appropriate sea surface temperature and bathymetry terms (Model A) and one including appropriate sea surface height and bathymetry terms (Model B). Both models are summarized in **Table 4**.

Model A indicates that predicted log CPUE has an inverse relationship with sea surface temperature. At any given depth, the predicted 99th percentile of log CPUE is higher for colder sea surface temperatures (**Figure 5**). Model B indicated that predicted log CPUE also has an inverse relationship with sea surface height, with the predicted log CPUE greater at lower heights at any given depth (**Figure 6**). Bathymetry appears to be more complicated with the highest predicted log CPUE at the deepest bottom depths and moderately high predictions at the shallowest bottom depths for any given sea surface temperature (**Figure 5**) or any given sea surface height (**Figure 6**).

Because some of the tested environmental variables were retained in the GLMM but show some promise for predicting high catches, a *post hoc* analysis was performed visually inspecting the annual CPUE trend and the North Atlantic Oscillation (NAO) (http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml). The NAO is associated with height, temperature, and pressure anomalies that could be influencing mako abundance and catch. The predicted mean log CPUE and the median NAO index appear to follow similar trends except in 2007 and 2012 (**Figure 7**).

4. Discussion

The two portions of the GLMM approach retained different environmental variables; proportion positive retained sea surface temperature and bathymetry while the positive catches retained only bathymetry. The difference in environmental variables retained in the two parts of the GLMM model indicate that sea surface temperature plays a role in whether mako will be caught at all but not necessarily how many will be caught. Sea surface height was never retained indicating that it may not play a role at all. Even when retained in the BIC best models, the environmental variables explained less than 3% of the deviance in the binomial and lognormal GLMM. On the other hand, the quantile regression indicates that all three environmental variables may be important for predicting high values of CPUE. This difference between methods could indicate that sea surface temperature and sea surface height are more influential on particularly high catch but not mean catch. It is also possible that the traditional GLMM method is not picking up the environmental signals. Sea surface temperature is known to vary with the season and latitude. This data covers most of the western-north Atlantic over several years and the local sea surface temperature signal could be hidden by the variable quarter or the large spatial scale of the data. Sea surface height contours are considered to be indicative of frontal position. It is possible that another frontal descriptor, such as distance from front, would be more informative. The gear variables, which are assumed to influence catchability, were not significant in the quantile regression models but were in the GLMM indicating that gear may matter less for large catches than for presence/absence.

Sea surface temperature and height could be influencing mako distribution and density. The results of both methods indicate that sharks may be found in higher densities at certain environmental conditions (colder waters and lower heights) therefore increasing the proportion of positive catch and the number caught at locations with these conditions. Bathymetry could be influencing mako density and/or catchability. The GLMM shows higher probability of positive catch and higher catches in shallow waters, again indicating that there are possibly higher shark densities there. However, low catches in deeper waters could also be caused by a decrease in catchability, if sharks are present but occupy a larger volume in deeper waters. However, the quantile regression method shows high CPUE in very deep and shallow waters indicating high densities for both or possibly high catchability.

This work was intended to be an exploratory study to test methods for ultimately determining mako habitat preference. This will be used to distribute individuals spatially in an individual-based model (IBM) based on environmental conditions at each location. The more traditional method of including environmental variables in a GLMM does not appear to tell the whole story. Sea surface height was not retained in either model and variables that were retained explained very little deviance. The quantile regression of the 99^{th} percentile of CPUE indicates that these variables are significant and can potentially predict high catch.

For the purposes of determining habitat preference and distribution, as opposed to standardized catch rates over time, quantile regression may be a better method. High CPUE can be interpreted as a result of high shark density indicating a preference for the conditions found at that particular set. Mako sharks are highly migratory, long-lived sharks that are not known to aggregate. Catches are typically zero so the data is zero-inflated. When there is a positive catch, most of the time only 1 or 2 makos are caught per set. If we assume that CPUE and catch are representative of density, then sets with low CPUE (which are the overwhelming majority in this dataset) and ultimately the mean CPUE, tell us very little about where we find makos. Quantile regression looking at the upper extreme eliminates the issue of large numbers of zeros and avoids making habitat preference assumptions based on very low average CPUE. There is also more freedom with the variables themselves because there is no assumption of a normal distribution and the error distribution does not need to be specified. This method may provide the tools necessary to determine which environmental conditions makos actually inhabit.

The quantile regression method seems promising after this study. However, more analysis needs to be carried out to define mako habitat preference. This study only looked at the 99th percentile, which was chosen arbitrarily; another quantile may be more informative. In this study a "final model" was never selected and retained variables were determined using p-values. More diagnostics need to be performed and a selection method needs to be solidified in order to eventually select a final model to be employed within the spatial IBM. Furthermore, the p-value of a variable is influenced by its location in the model in relation to other variables (Sokal and Rohlf 2011) and alone may not be the most reliable for final variable selection.

Comparison of the GLMM with the NAO index, a synergy of environmental anomalies, indicates that despite the low deviance explained and lack of variable retention, these environmental conditions could be impacting catch rates. The possible correlation may be caused by changes in distribution of either the mako population or the fishing fleet, or some other mechanism that affects catchability. It is unlikely that mako shark abundance would change rapidly enough to track the NAO. However, the NAO may be relevant to determining shark habitat distributions over time.

References

Anderson, M.J., 2008. Animal-sediment relationships re-visited: Characterising species' distributions along an environmental gradient using canonical analysis and quantile regression splines. Journal of Experimental Marine Biology and Ecology, 366(1-2):16-27.

Brown, O.B. and P.J. Minnett. 1999. MODIS Infrared Sea Surface Temperature Algorithm, Version 2.0. Algorithm Theoretical Basis Document, ATBD25. University of Miami, Miami, FL. Under Contract Number NAS5-31361.

Cade, B.S. and B.R. Noon. 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment, 1(8):412-420.

Cortés, E. 2013. Standardized catch rates of mako sharks from the U.S. pelagic longline logbook and observer programs using a generalized linear mixed model. Collect. Vol. Sci. Pap. ICCAT, 69(4): 1578-1590.

Cortés, E. 2007. Standardized catch rates for blue shark and shortfin mako from the US pelagic longling logbook program. Collect. Vol. Sci. Pap. ICCAT, 60(2): 617-628.

Ducet, N., P.-Y. Le Traon, and G. Reverdin. 2000. Global high resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. J. Geophys. Res., 105:19477-19498.

Fornaroli, R. et al. 2016. Optimal flow for brown trout: Habitat-prey optimization. Science of the Total Environment, 566-567:1568-1578.

Fukunaga, A., R.K. Kosaki, D. Wagner, and C. Kane. 2016. Structure of mesophotic reef fish assemblages in the northwestern Hawaiian Islands. Plos One, 11(7).

Koenker, R. 2015. Quantile Regression in R: a vignette. www.R-project.org.

Koenker, R. and B. Gilbert. 1978. Regression quantiles. Econometrica, 46(1): 33-50.

Maunder M.N., and A.E. Punt. 2004. Standardizing catch and effort data: as review of recent approaches. Fisheries Research, 70(2-3):141-159.

Ortiz, M. and F. Arocha. 2004. Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery. Fisheries Research, 70(2-3):275-297.

Sokal, R.R. and F.J. Rohlf. 2011. Biometry. Fourth Edition. W.H. Freeman.

Smith, W.H.F. and D.T. Sandwell 1997. Global seafloor topography from satellite altimetry and ship depth soundings. Science, 277:1957-1962.

Venables, W. N. and B.D. Ripley. 2002. Modern Applied Statistics with S. Fourth edition. Springer.

Table 1. Factor levels used for hooks between floats, sea surface temperature, sea surface height, and bathymetry predictor variables used in the generalized linear mixed model approach to predict catch rates.

Variable	*Number of Levels*	*Levels*
Hooks Between Floats	5	<3
		4
		5
		6
		>7
Sea Surface Temperature (°C)	5	<15°
		15°-20°
		20°-25°
		25°-30°
		30°-35°
Sea Surface Height (m)	4	<0m
		0m-0.3m
		0.3m-0.6m
		>0.6m
Bathymetry (m)	7	0m-800m
		800m-1600m
		1600m-2400m
		2400m-3200m
		3200m-4000m
		4000m-4800m
		>4800m

Table 2. Analysis of variance table for factors retained in the model of proportion of positive sets and positive catch of shortfin mako sharks for U.S. pelagic longline observer data.

Proportion Positive Sets	*Df*	*Deviance*	*Residual Df*	*Residual Deviance*	*p-value*	*Percent Deviance*
NULL			6442	8930.5		
Year	10	1935.29	6432	6995.2	0	21.7
Fishing Area	7	866.25	6425	6129.0	0	9.7
Quarter	3	32.60	6422	6096.4	<0.0001	0.4
Use of Light	1	83.20	6421	6013.2	0	0.9
Hooks Between Floats	4	11.80	6417	6001.4	<0.02	0.1
Sea Surface Temperature	4	60.93	6413	5940.4	0	0.7
Bathymetry	6	155.62	6407	5784.8	0	1.7
Positive Catch						
NULL			1528	1707.95		
Year	10	866.66	1518	841.29	0	50.7
Fishing Area	7	84.48	1511	756.81	0	4.9
Quarter	3	33.93	1508	722.88	<0.0001	2.0
Hooks Between Floats	4	30.59	1504	692.29	<0.0001	1.8
Bathymetry	6	33.70	1498	658.59	<0.0001	2.0

Table 3. Correlation matrix for the environmental predictor variables sea surface temperature, sea surface height, and bathymetry.

	Sea Surface Temperature	*Sea Surface Height*	*Bathymetry/1000m*
Sea Surface Temperature	1		
Sea Surface Height	0.443	1	
Bathymetry/1000m	-0.0282	-0.389	1

Table 4. Summary table for variables retained in the quantile regression models of proportion of positive sets and positive catch of shortfin mako sharks for U.S. pelagic longline observer data.

Model A	*Coefficient Value*	*Standard Error*	*p-value*
Sea surface Temperature	-0.0827	0.0364	<0.03
Bathymetry/1000m	0.0192	0.399	>0.05
Bathymetry/1000m Squared	0.137	0.0320	<0.0001
Interaction	0.0360	0.0137	<0.01
Model B			
Sea Surface Height	-1.73	0.332	0.0
Bathymetry/1000m	1.17	0.244	0.0
Bathymetry/1000m Squared	0.241	0.0566	<0.0001

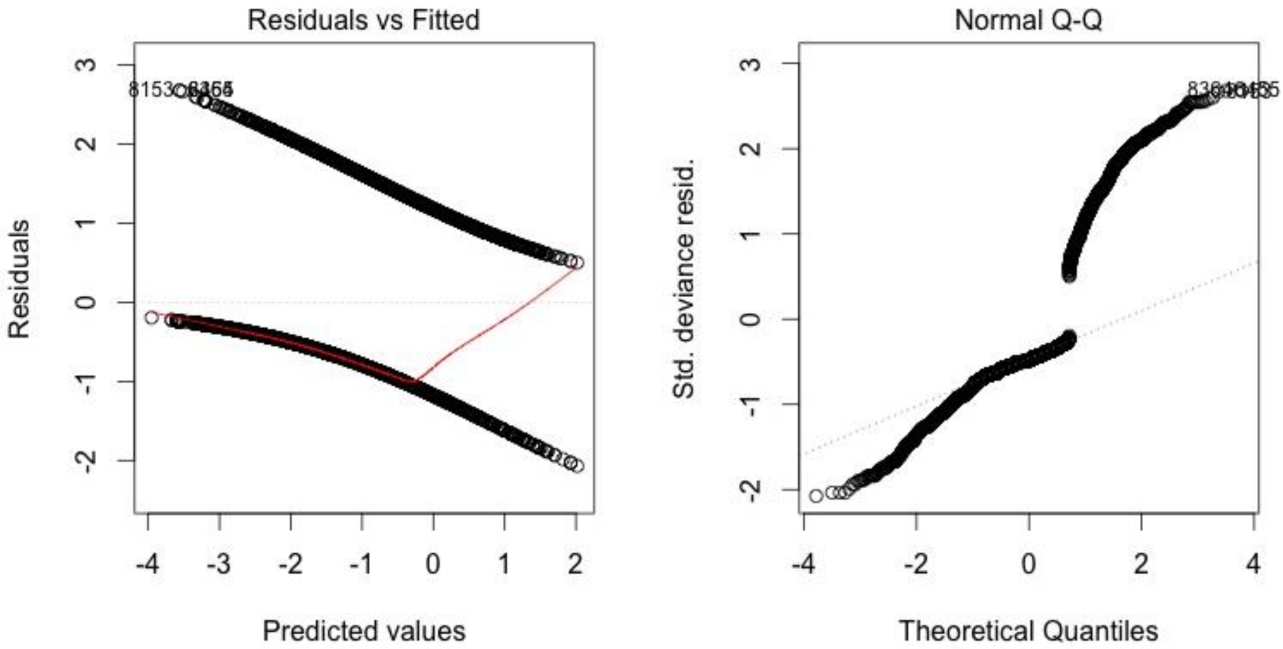

Figure 1. Proportion of positive sets binomial GLMM model diagnostics.

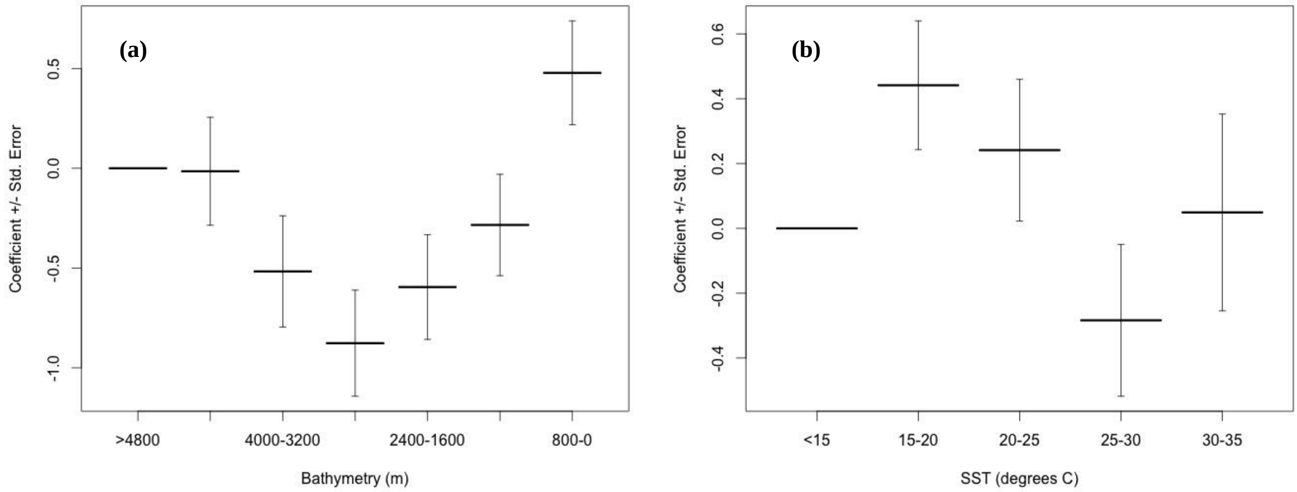

Figure 2. Coefficients +/- the standard error for the **(a)** bathymetry and the **(b)** sea surface temperature predictor variable levels determined for the proportion of positive sets binomial model of the GLMM approach.

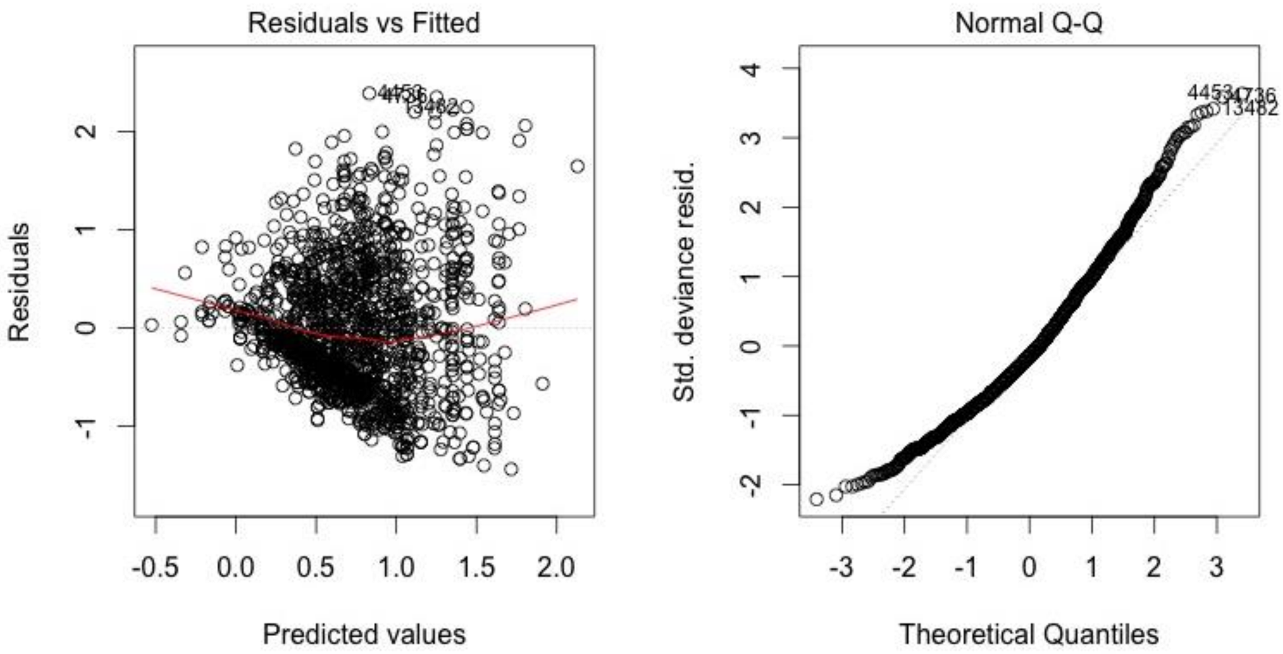

Figure 3. Positive catch lognormal GLMM model diagnostics.

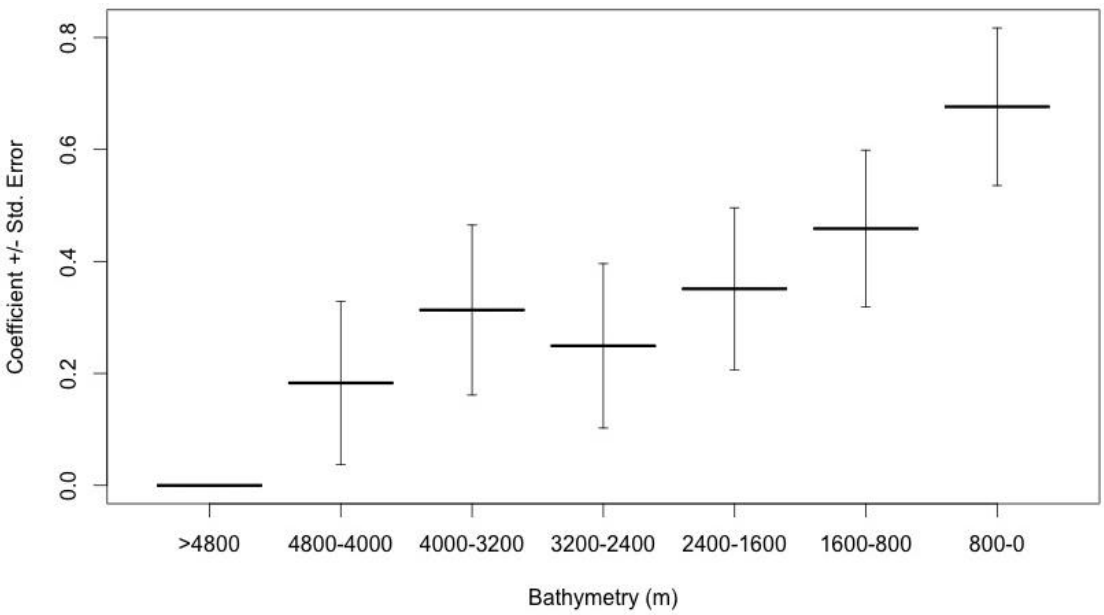

Figure 4. Coefficients +/- the standard error for the bathymetry predictor variable levels determined for the positive catch lognormal model of the GLMM approach.

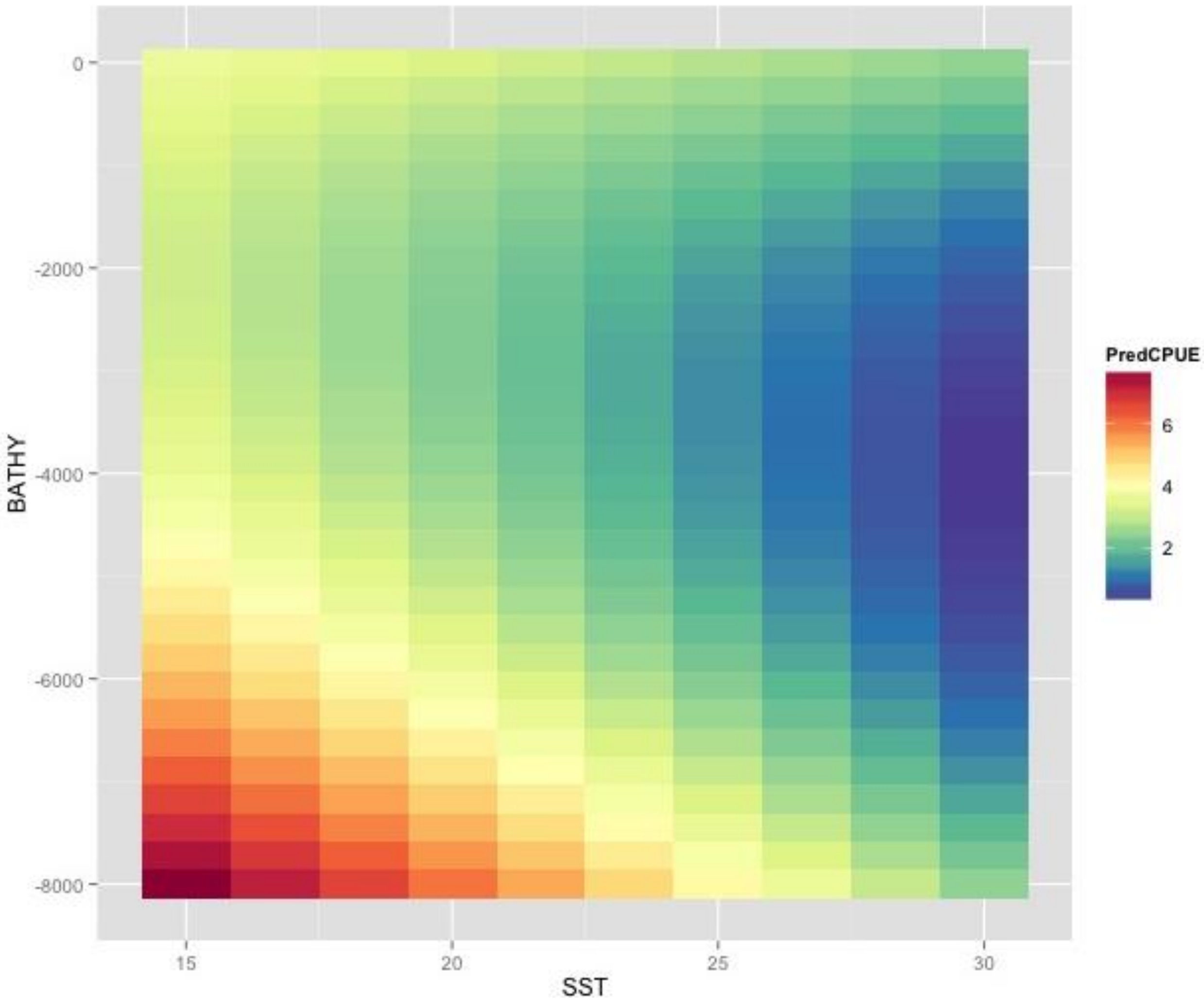

Figure 5. Heat map of predicted log CPUE generated from Model A; the quantile regression of the 99th percentile using the explanatory variables sea surface temperature, bathymetry, bathymetry squared, and an interaction term.

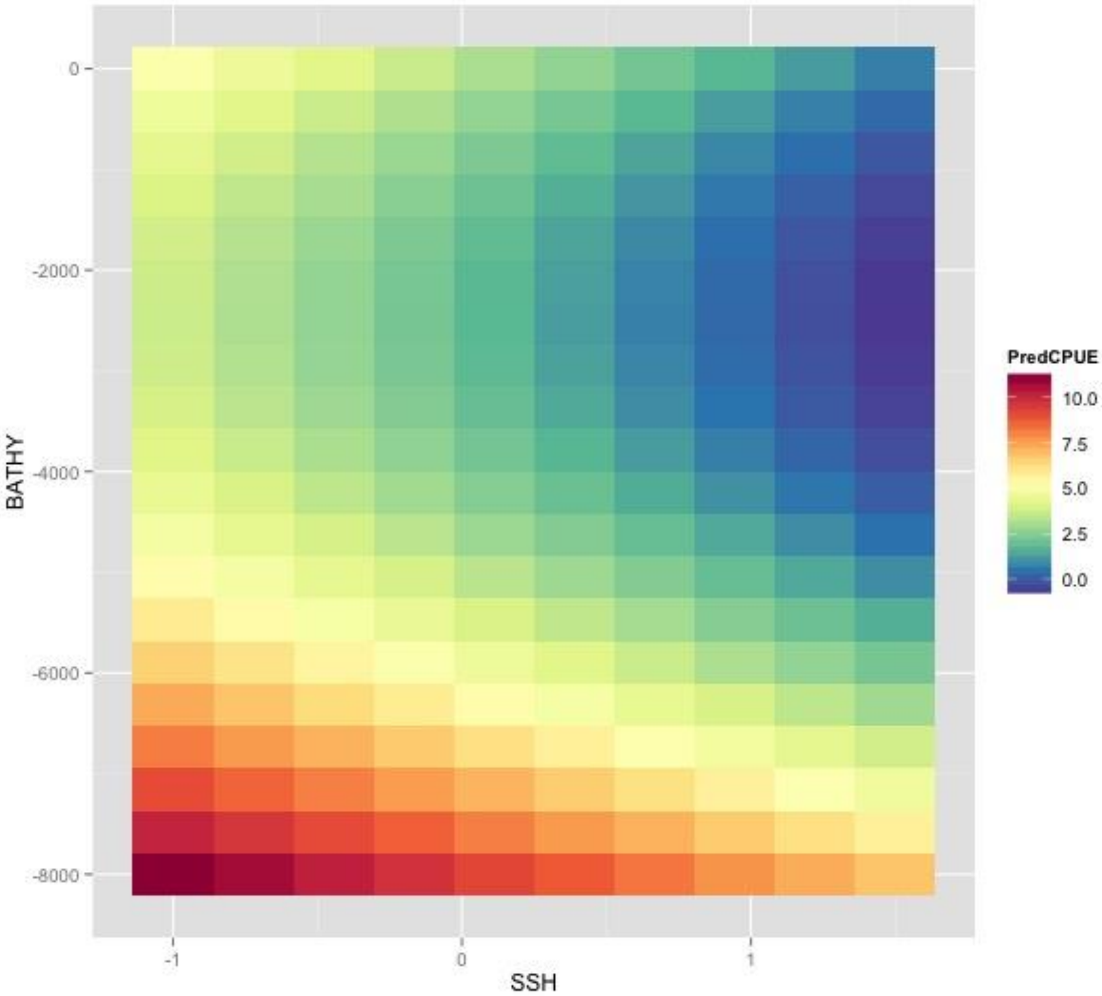

Figure 6. Heat map of predicted log CPUE generated from Model B; the quantile regression of the 99th percentile using the explanatory variables sea surface height, bathymetry and bathymetry.

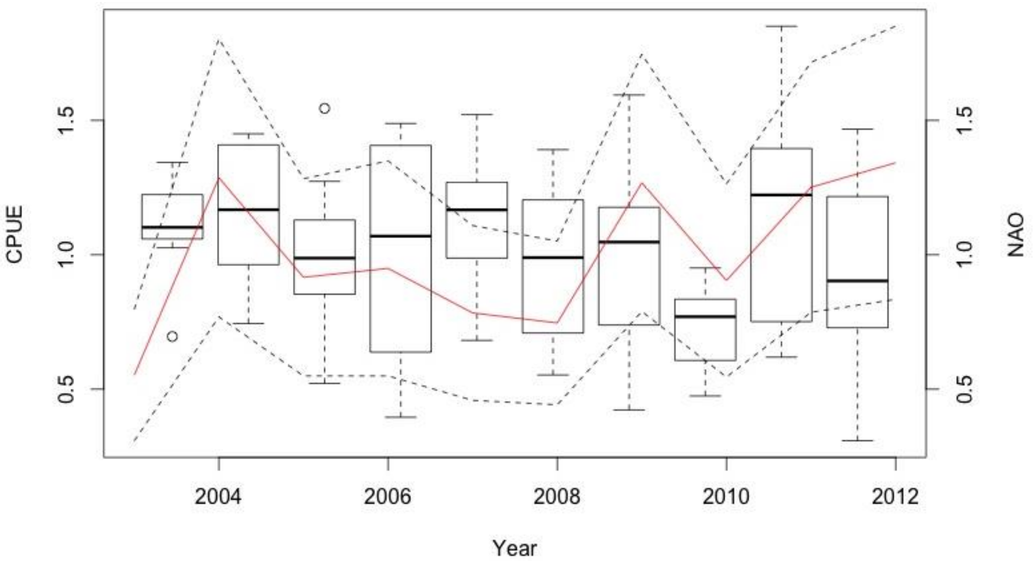

Figure 7. The annual CPUE trend (red), as predicted by the GLMM approach, overlaid with boxplots of the monthly North Atlantic Oscillation (NAO) index values for each year.