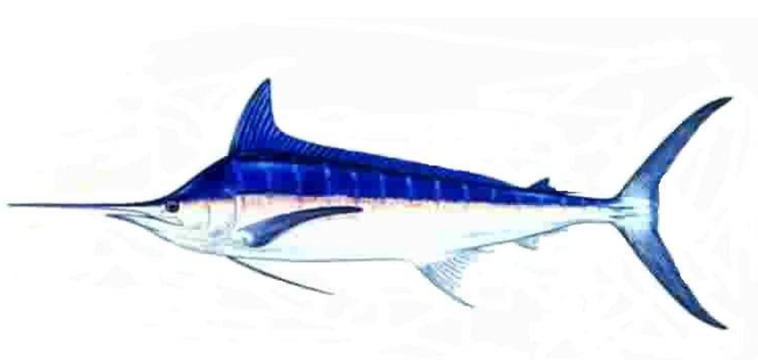
ISC/06/MARWG&SWOWG-2/07

Comparison of dELTA GLM and statistical habitat-based models (statHBS) to estimate standardized CPUE for striped marlin[1]

Keith Bigelow
Pacific Islands Fisheries Science Center
National Marine Fisheries Service, NOAA
Honolulu, Hawaii



[1]Working document submitted to the ISC Marlin Working Group Workshop, November 8-15, 2006, Shimizu, Shizuoka, Japan. Document not to be cited without authors' written permission.

COMPARISON OF DELTA GLM AND STATISTICAL HABITAT-BASED MODELS (statHBS) TO ESTIMATE STANDARDIZED CPUE FOR STRIPED MARLIN[1]

Keith A. Bigelow[2]

1. INTRODUCTION

The ISC Marlin working group met during 15 to 21 November, 2005 and developed standardized CPUE time-series for striped marlin in the north Pacific Ocean. This document extends the CPUE analysis of the Japanese distant-water longline fishery by applying two approaches: 1) delta Generalized Linear models (GLM) and 2) statistical habitat-based models (statHBS).

2. METHODS

Dataset

Two datasets from the Striped Marlin Stock Assessment Workshop (November 2005) were used in the effort and catch-per-unit-effort (CPUE) standardization. Japanese distant-water longline data were aggregated by month and 5° of latitude and longitude resolution at the National Research Institute of Far Seas Fisheries. The region of interest was the north Pacific with a five-area stratification (Fig. 1).

Dataset 1 was compiled to compare model results for the period from 1975 to 2004 when gear configuration information (number of hooks between floats of the longline gear, HBF) was available. Variables in the model included year, month, each individual HBF category, area, total number of hooks set and the catch of striped marlin. Dataset 2 was compiled from 1952 to 2004 to reflect the entire historical time-series of the fishery. Gear configuration prior to 1975 was assumed to be 5 HBF.

Delta-GLM CPUE standardization methods

Standardized striped marlin CPUE has been estimated previously by Generalized Linear Model (GLM) approaches (Yokawa and Clark 2005). An alternative delta-GLM (Lo et al. 1992) was applied which separately estimates the proportion of positive striped marlin catches assuming a binomial error distribution, and the mean catch rate of positive catches by assuming a different error distribution such as lognormal, gamma or inverse gaussian. The standardized index is the product of these model estimated components.

[1] PIFSC Working Paper WP-06-014. Issued 8 November 2006.
[2] NOAA Fisheries, Pacific Islands Fisheries Science Center, 2570 Dole Street, Honolulu, HI 96822, USA
Email: Keith.Bigelow@noaa.gov

The formulation of the delta GLM for both dataset 1 (n=132,851 strata) and 2 (n=162,281 strata) was:

$\text{CPUE}_{ijkl} \sim \text{Year}_i+\text{month}_j+\text{HBF}_k+\text{area}_l+e$,

where CPUE_{ijkl} is the catch in number per 1,000 hooks in year *i*, month *j*, gear *k*, area *l* and *e* represents the random error term. No interaction terms were considered. Analyses were done using the R statistical computer software (R version 2.2.0), and a delta-GLM procedure obtained from E.J. Dick (NOAA Fisheries). Selection of the error distribution for the positive catch rates was based on the Akaike's Information Criterion (AIC) after Dick (2004). Given the preferred error distribution, a step-wise regression procedure was used to determine the set of explanatory variables. The difference in deviance between two consecutive models was evaluated by AIC and deviance analysis tables are presented for the data series, including the deviance for the proportion of positive observations and for positive catch rates.

statistical habitat-based (statHBS) standardization methods

The methodology for the calculation of statHBS indices followed Maunder et al. (In Press) and a copy of the manuscript is provided to the ISC Marlin and Swordfish working groups as background information.

The statHBS model is an extension of the deterministic habitat-based framework (Hinton and Nakano 1996) and represents a modeling approach whereby catch rates are standardized by estimating effective longline effort from the vertical distribution of hooks, species-specific habitat preferences and the vertical, horizontal, and temporal distribution of environmental conditions. The parameters of the model are estimated by fitting to the observed catch. This is accomplished by minimizing the negative log-likelihood. A lognormal likelihood function is used for the striped marlin example as:

$$-\ln L(\boldsymbol{\theta} \mid \tilde{\mathbf{C}}) = \sum_i \ln[\sigma] + \left[\frac{\left(\ln[\tilde{C}_i + \delta] - \ln[C_i + \delta]\right)^2}{2\sigma^2} \right]$$

where $\tilde{C}_i$ is the observed catch and δ is a small constant (1.0) to avoid computational problems when the observed or predicted catch is zero. The standard deviation, σ, of the likelihood function is estimated as a parameter in the model.

For individual observations (*i*) from an effort (*E*) series *j*, an estimate of catch (*C*) in year *y* is obtained as $\hat{C}_{i,j,y} = E_{i,j,y} q_j B_y$, where *q* is overall catchability and *B* is abundance. Year effects ($\theta_y = qB_y$) are estimated because both *q* and *B* are unknown. The negative log-likelihood is minimized by simultaneously estimating various parameters with the function minimizer in AD Model Builder.

A model was fit to striped marlin in the north Pacific from 1975 to 2004 (dataset 1). The number of observations was slightly less (n=131,994) than the delta-GLM because no environmental observations were available for several strata encompassing coastal areas. The statHBS model estimated a year effect and habitat preferences within the water column for each of the 5 areas. For each area, the model had 15 ambient temperatures at 2°C intervals from 3.5 to 33.5°C and 15 temperature gradients at 0.03 $^{\circ}C*m^{-1}$ from -0.40 to 0.05 $^{\circ}C*m^{-1}$. No priors were used for the habitat preferences. Estimated temperature parameters from the 1975 to 2004 application were applied to the entire time-series (dataset 2) in a deterministic manner to estimate standardized CPUE from 1952 to 2004. The vertical distribution of hooks within each HBF configuration was estimated from longline characteristics and catenary geometry (K. Yokawa, pers. comm., National Research Institute of Far Seas Fisheries, Shimizu, Japan; Bigelow et al. 2002).

Environmental covariates of ambient temperature and thermocline gradient were obtained for the statHBS model. Temperature at discrete depths was obtained from the Global Ocean Data Assimilation System (GODAS) developed at the National Centers for Environmental Prediction (http://cfs.ncep.noaa.gov/cfs/godas/). The model has 10 and 31 vertical layers in the upper 100 and 1000 m; respectively, and a spatio-temporal resolution of 1/3º latitude and 1° longitude by one month (1980−2005). A cubic smoothing spline (smooth.spline) in R was implemented for each temperature profile to predict temperature and gradient (1^{st} derivative) for each meter of the profile. Mean temperature and gradient were then estimated for each 40-m depth category. Temperature data from 1952 to 1979 was obtained from an alternative OGCM (SODA analysis, Carton et al. 2000a, 2000b, http://apdrc.soest.hawaii.edu).

3. RESULTS AND DISCUSSION

Delta-GLM CPUE standardization methods

A delta-GLM application may be appropriate for striped marlin in the north Pacific as the percentage of 5º-month strata with zero catch (1975−2005) for areas 1 to 5 was 23%, 70%, 15%, 28% and 12%, respectively. An annual time-series indicates that the percentage of 5º-month strata with zero striped marlin catch declined from 1952 to 1975 perhaps due to the spatial expansion of the fishery. From 1975 to 2004 the percentage of zero catch strata ranged from 35% to 45%.

Details of the striped marlin model results are given in Tables 1−2. AIC results indicate that a lognormal distribution is preferred over gamma and inverse gaussian to describe the positive catches (Table 1).

Comparison of year effects between models show moderate differences between nominal and standardized delta-GLM CPUE trends for 1975−2004 (Figure 3) and 1952−2004 (Figure 4). Nominal CPUE is relatively stable from 1952 to 1960, increases sharply from the early 1960s to early 1970s perhaps due to directed targeting and then has moderate interannual variability to 2004. Nominal CPUE in 2004 is one of the lowest annual indices. Increases in standardized CPUE from the early 1960s to early 1970s were not as large as nominal trends. Additionally, the trend in standardized CPUE from 1975 to 2004 is much more optimistic then the trend in nominal CPUE.

Figure 5 illustrates the derived effects for month, gear (HBF) and area. The area effect had the largest explanatory power, followed by month and gear configuration which had similar effects (Table 2). The area to the east of 125ºW (area 5) had the largest effect while areas to the south of 20ºN (areas 2 and 4) had relatively smaller effects. The month effect was largest during February, March, October and November, but smallest during summer. There was a negative relationship between CPUE and gear configuration as shallower gear (e.g. 5−12 HBF) had a larger CPUE effect than deeper gear. Overall, an increase of one hook between a float in gear configuration corresponded to a 2.5% decrease in catchability. The residuals for the positive catches were normally distributed with the assumption of a lognormal error distribution (Figure 6).

statistical habitat-based (statHBS) standardization methods

The effective effort from the statHBS approach provided an improved fit to the variation in striped marlin catch compared to nominal effort for both time periods (1975−2004, Nominal AIC=90,080, statHBS AIC=54,097; 1952−2004, Nominal AIC=124,547, statHBS AIC=82,049).

A comparison of year effects between models show moderate differences between nominal and standardized statHBS CPUE trends for 1975−2004 (Figure 7) and 1952−2004 (Figure 8). Increases in standardized CPUE from the early 1960s to early 1970s were not as large as the nominal trend, similar to the delta-GLM approach. Additionally, the trend in standardized CPUE from the statHBS model from 1975 to 2004 is more optimistic then the trend in nominal CPUE, but less optimistic than the delta-GLM standardized CPUE.

Fitted temperatures and thermocline gradient differed by area for striped marlin (Figure 9). The warmest temperatures (23º–28°C) occurred in area 1 with cooler temperature-at-capture (15º–23°C) for areas 2 to 5. Fitted thermocline effects indicated that striped marlin where primarily caught in the mixed layer (low gradients) in areas 1 to 3 and within the thermocline at larger gradients in areas 4 and 5 (Figure 9). The selection of a five area stratification may not represent an optimum spatial structure for a statHBS analysis as there is high oceanographic variability, especially latitudinally, within such large strata.

The estimated trend in gear catchability was similar between the delta-GLM and statHBS analyses (Figure 10). In general, catchability of deeper gear (20−25 HBF) was 30−50% of shallow gear, which appears reasonable for a marlin species. Catchability in the statHBS analysis declined more rapidly from shallow to deeper gear than the delta-GLM. Within each analysis, catchability estimates for gear of 26 HBF were much larger than other deep gear and this aspect requires further analysis. Figure 11 shows the normalized cumulative plot of residuals from the statHBS model.

Conclusions

1. The percentage of zero catch observations in the striped marlin dataset may advocate for the use of delta-GLM approaches.

2. Trends between nominal and standardized CPUE were consistent between standardization models. Standardized CPUE was more optimistic than nominal CPUE in both delta-GLM and statHBS applications, though recent annual indices estimated by the delta-GLM were more optimistic than statHBS.

3. Gear configuration trends in catchability were consistent between models. The negative relationship in catchability with deeper gear appears reasonable given our current understanding of striped marlin vulnerability to shallow and deep gear.

4. REFERENCES

CARTON, H., G. Chepurin and X. Cao. 2000a. A simple ocean data assimilation analysis of the global upper ocean 1950–95. Part I: Methodology. J. Phys. Ocn. 30:294:309.

CARTON, H., G. Chepurin and X. Cao. 2000b. A simple ocean data assimilation analysis of the global upper ocean 1950–95. Part II: Results. J. Phys. Ocn. 30:311:326.

DICK, E. 2004. Beyond 'lognormal versus gamma': discrimination among error distributions for generalized linear models. Fish. Res. 70: 351–366.

HINTON, M. and H. Nakano. 1996. Standardizing catch and effort statistics using physiological, ecological, or behavioral constraints and environmental data, with an application to blue marlin (*Makaira nigricans*) catch and effort data from Japanese longline fisheries in the Pacific. Bull. Int. Am. Trop. Tuna Comm. 21(4): 171−200.

LO, N., L. Jacobson and J. Squire. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can. J. Fish. Aquat. Sci. 49: 2515-2526.

MAUNDER, M., M. Hinton, K. Bigelow and A. Langley. In Press. Developing indices of abundance using habitat data in a statistical framework. Bull. Mar.

YOKAWA, K. and S. Clarke 2005. Standardizations of CPUE of striped marlin caught by Japanese offshore and distant water longliners in the north Pacific. ISC05/MAR-WG.

Table 1. Delta GLM comparison of distribution assumptions fit to north Pacific striped marlin.

Time-series	Error structure for positive component	AIC for positive component	AIC for binomial component
1975−2004	Lognormal	76,676	141,967
1975−2004	Gamma	89,383	141,967
1975−2004	Inverse gaussian	100,906	141,967
1952−2004	Lognormal	127,045	167,937
1952−2004	Gamma	142,660	167,937
1952−2004	Inverse gaussian	165,830	167,937

Table 2. Stepwise model results of a delta-lognormal GLM fit to north Pacific striped marlin (1975−2004).

Model factors for positive catch rates (1975−2004)	d.f.	AIC	Residual deviance	Change in deviance	% of total deviance
1			141,200		
Year	29	92,328	138,200	3,000	2.1
Year area	4	79,073	117,700	20,500	16.6
Year area month	11	77,883	116,000	1,700	17.8
Year area month gear	21	76,676	114,300	1,700	19.1
Model factors for proportion positive/total observations (1975−2004)	d.f.	AIC	Residual deviance	Change in deviance	% of total deviance
1			176,100		
Year	29	175,424	175,400	700	0.4
Year area	4	145,315	145,200	30,200	17.5
Year area month	11	142,658	142,600	2,600	19.0
Year area month gear	21	141,967	141,800	800	19.5

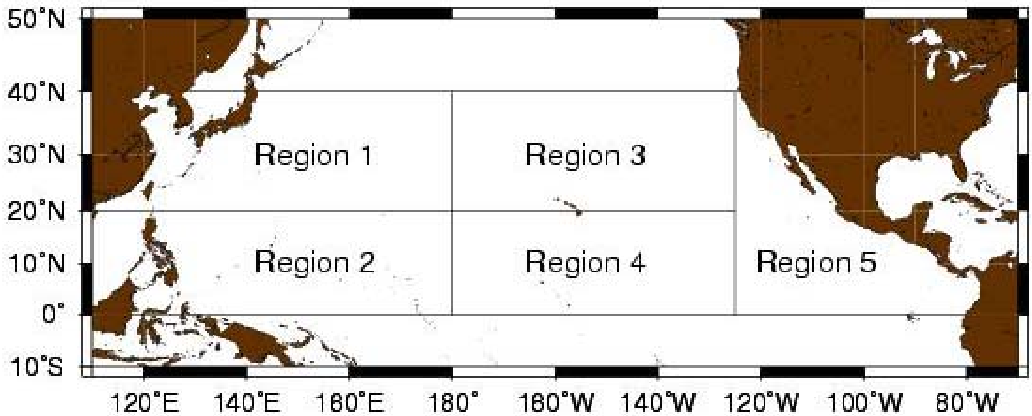


Figure 1. Area stratification used in the standardization of CPUE of striped marlin caught by Japanese distant-water longliners.

Figure 2. Time-series of the percentage of 5°-month strata with zero striped marlin catch in the north Pacific.

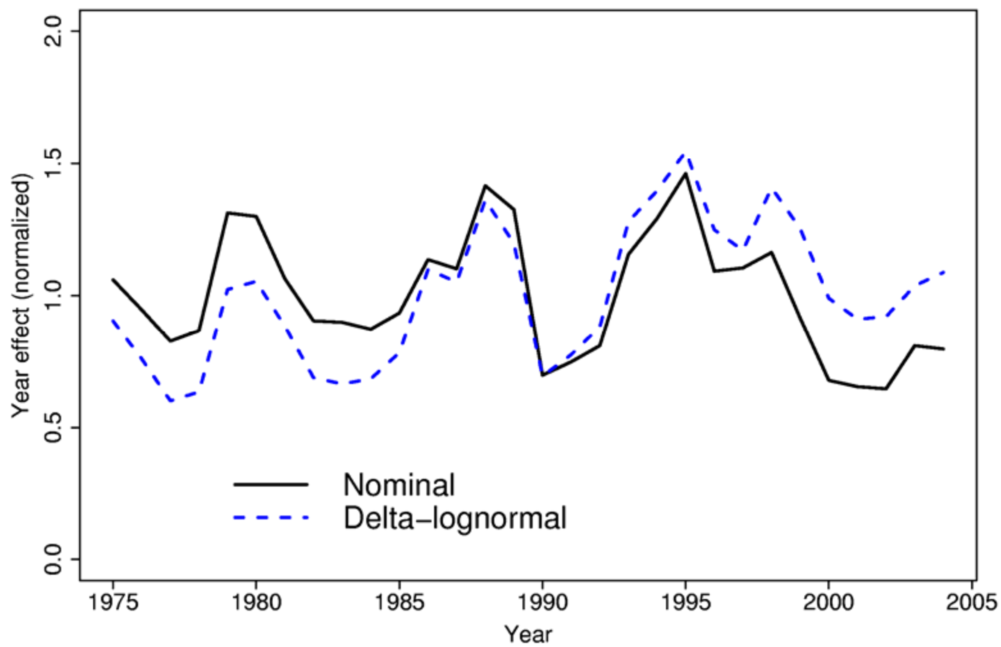


Figure 3. Comparison of north Pacific striped marlin (1975−2004) nominal CPUE (solid line) and standardized CPUE (dotted line) by a delta-lognormal Generalized Linear Model (GLM).

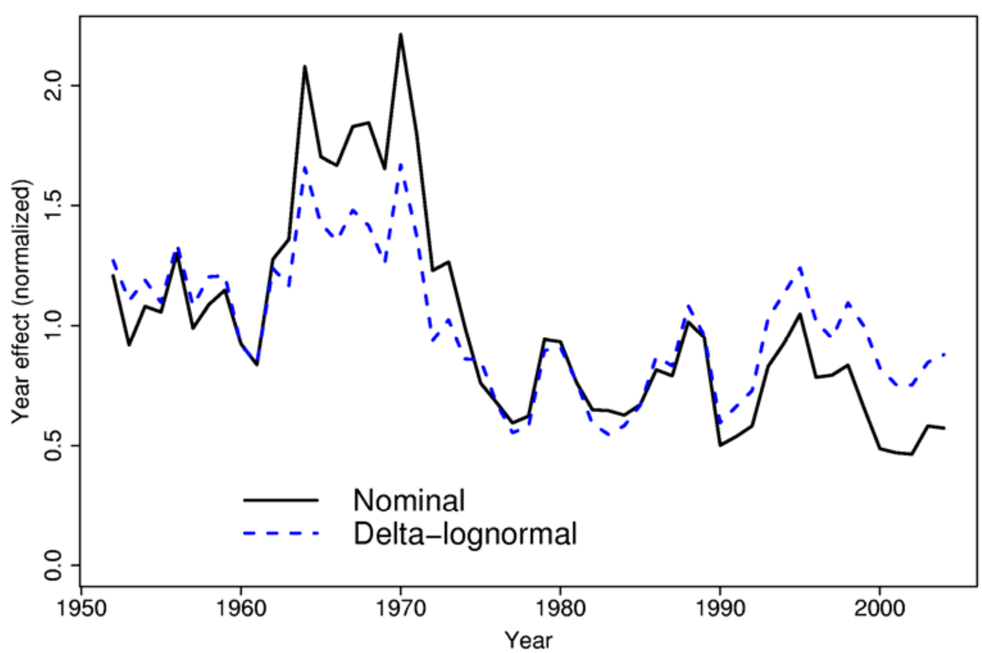


Figure 4. Comparison of north Pacific striped marlin (1952−2004) nominal CPUE (solid line) and standardized CPUE (dotted line) by a delta-lognormal Generalized Linear Model (GLM).

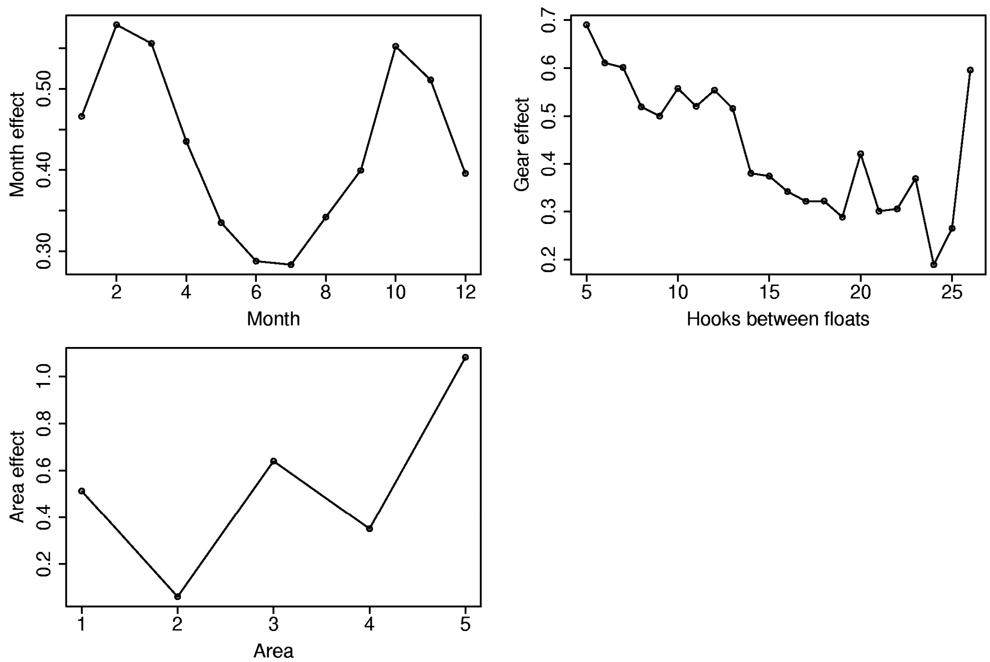


Figure 5. Delta-lognormal GLM (1975−2004) derived effects of month, gear (hooks between floats) and area on north Pacific striped marlin CPUE.

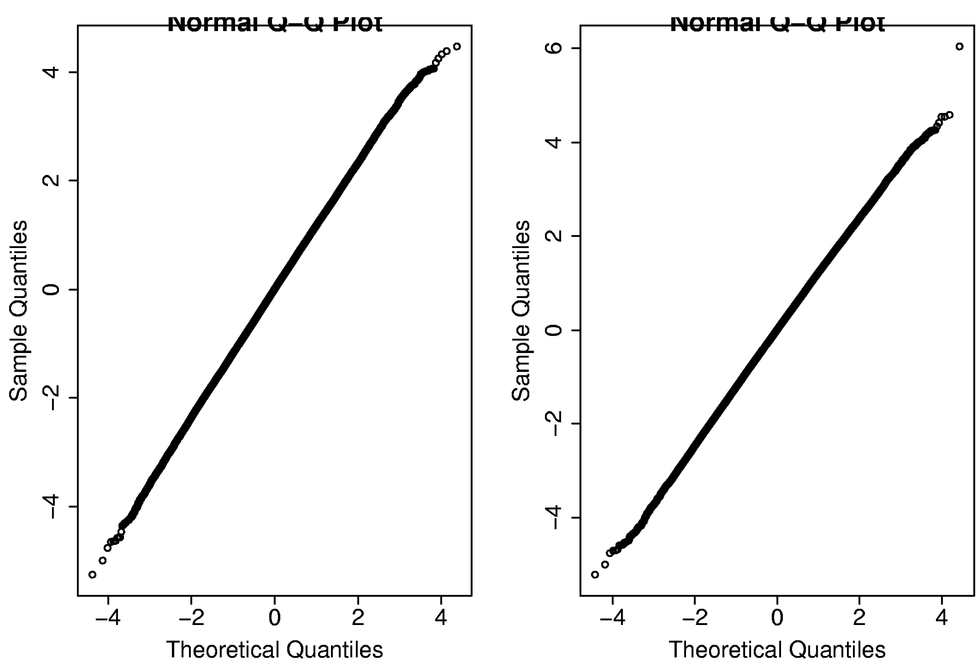


Figure 6. Cumulative normalized deviance residuals or qq-plot of the positive catches for striped marlin in the north Pacific assuming a lognormal error distribution for 1975 to 2004 (left) and 1952 to 2004 (right).

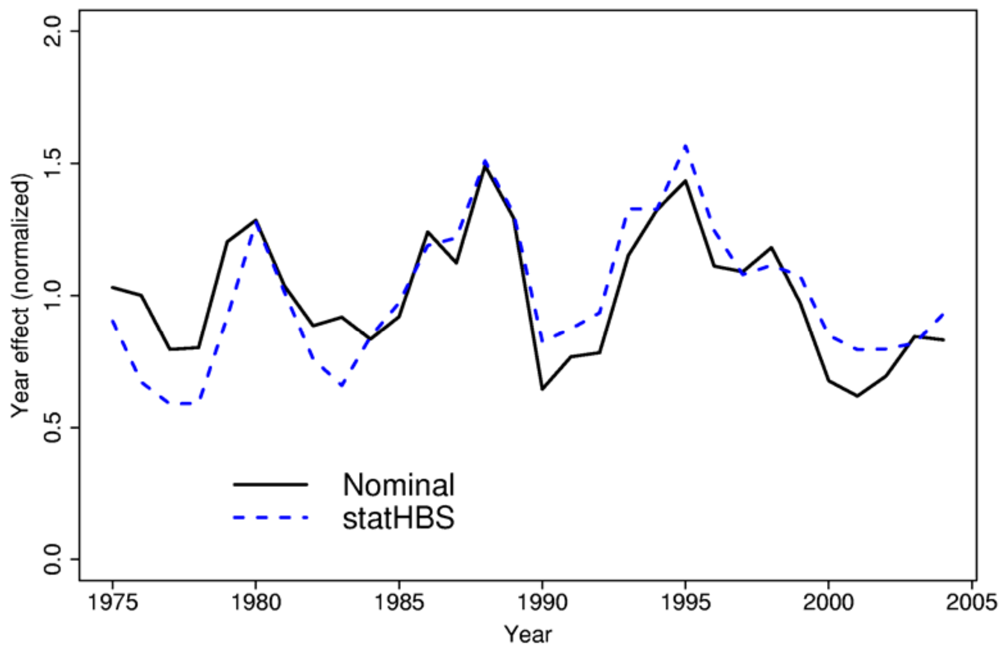


Figure 7. Comparison of north Pacific striped marlin (1975−2004) nominal CPUE (solid line) and standardized CPUE (dotted line) by a statistical habitat-based model (statHBS).

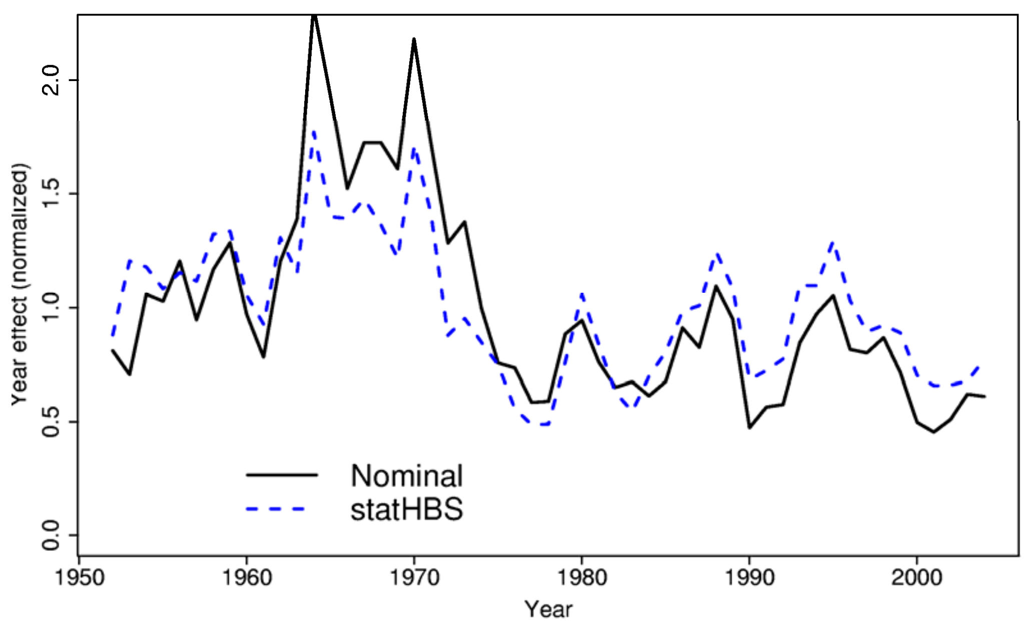


Figure 8. Comparison of north Pacific striped marlin (1952−2004) nominal CPUE (solid line) and standardized CPUE (dotted line) by a statistical habitat-based model (statHBS).

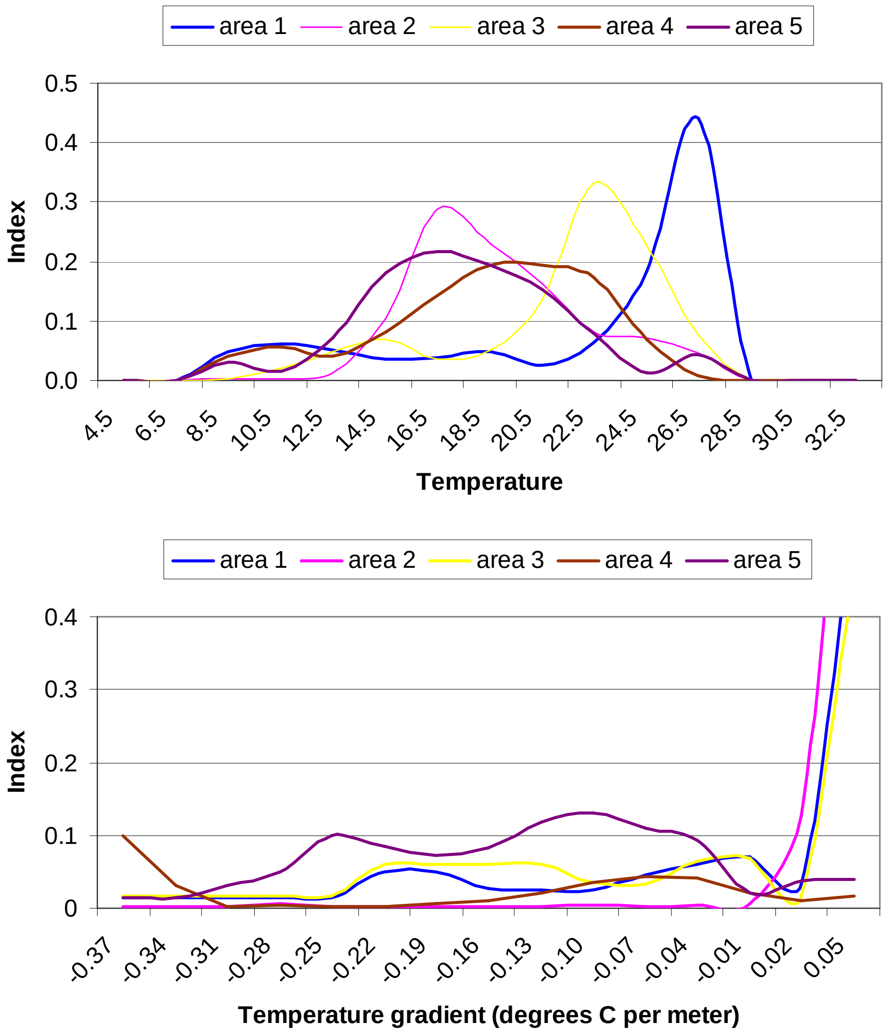


Figure 9. Ambient temperature (top) and thermocline gradient (bottom) fitted relationships for striped marlin estimated in a statistical habitat-based model (statHBS) for five areas in the north Pacific.

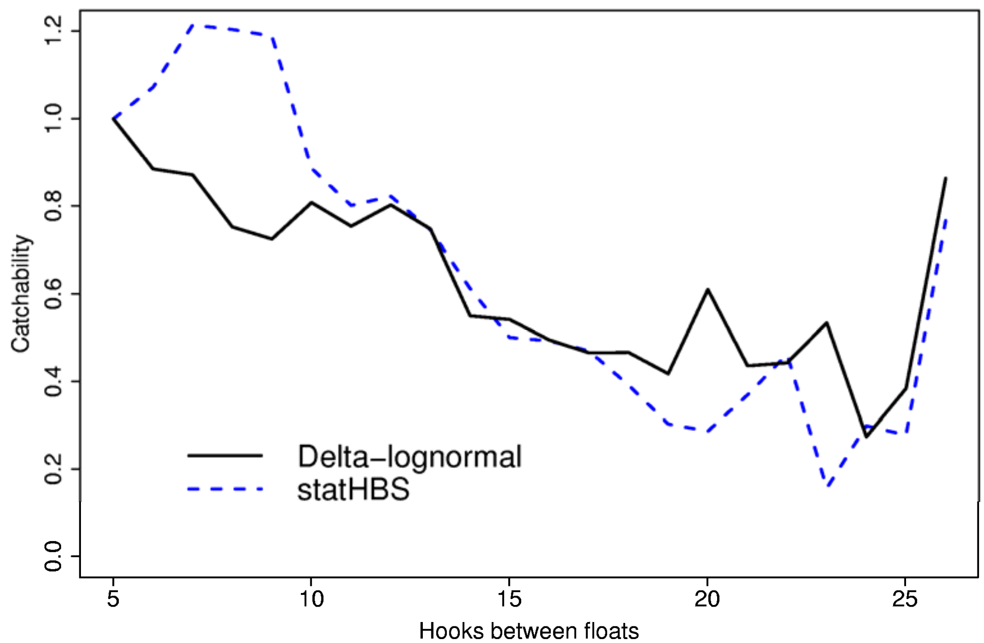


Figure 10. Gear configuration trends in catchability estimated by a delta-lognormal GLM and a statistical habitat-based model (statHBS) for north Pacific striped marlin (1975−2004).

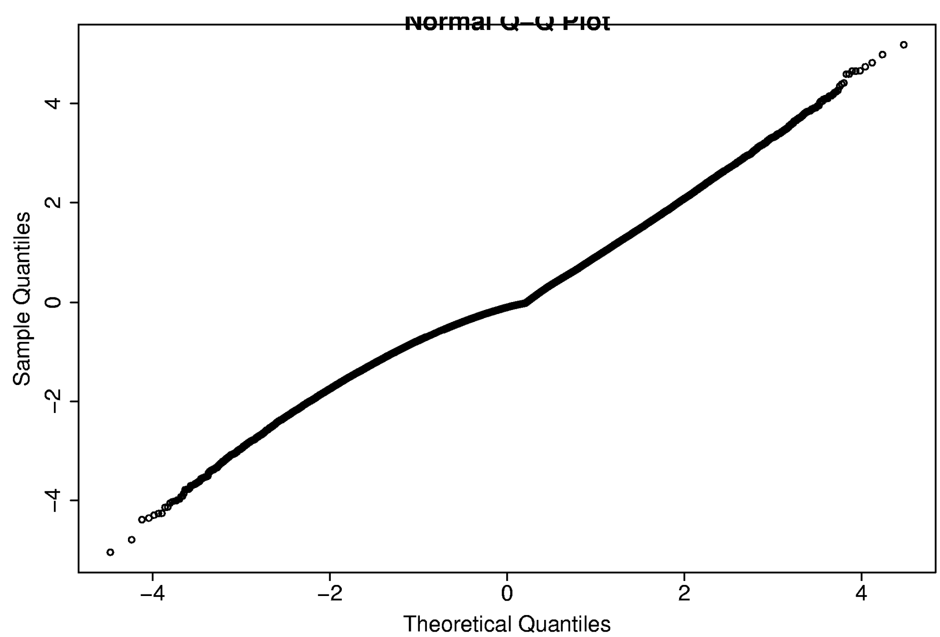


Figure 11. Cumulative normalized deviance residuals or qq-plot for a statHBS model applied to north Pacific striped marlin (1975−2004).