DOI: https://doi.org/10.1016/j.fishres.2018.09.025

This is the author's pre-print version of the paper

FULL PAPER available at the Publisher website:

https://www.sciencedirect.com/science/article/pii/S0165783618302662?via%3Dihub

Testing robustness of CPUE standardization and inclusion of environmental variables with simulated longline catch datasets

Francesca C. Forrestal[a]*, Michael Schirripa[b], C. Phillip Goodyear[c], Haritz Arrizabalaga[d], Elizabeth A. Babcock[e], Rui Coelho[f], Walter Ingram[g], Matthew Lauretta[b], Mauricio Ortiz[h], Rishi Sharma[i], John Walter[b]

[a] Cooperative Institute of Marine and Atmospheric Science, 4600 Rickenbacker Cswy, Miami, FL 33149, USA [b] Miami Laboratory, Southeast Fisheries Science Center, National Marine Fisheries Service, 75 Virginia Beach Drive, Miami, FL 33149-1099, USA
[c] 1214 North Lakeshore Drive, Niceville, FL 32578, USA
[d] AZTI Tecnalia, Marine Research Division, Herrera Kaia Portualdea, z/g 20110, Pasaia, Gipuzkoa, Spain
[e] Rosenstiel School of Marine & Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA
[f] Instituto Português do Mar e da Atmosfera (IPMA I.P.), Avenida 5 de Outubro s/n, Olhão, 8700-305 Portugal
[g] NOAA Fisheries, Southeast Fisheries Science Center, Mississippi Laboratories, 3209 Frederic Street, Pascagoula, MS 39567, USA
[h] ICCAT SECRETARIAT - Corazón de María, 8. 28002 Madrid, SPAIN
[i] NOAA Fisheries, Northwest Fisheries Science Center, 1201 Lloyd Boulevard, Suite 1100, Portland, OR 97232

Abstract

Environmental variability changes the distribution, migratory patterns, and susceptibility to various fishing gears for highly migratory marine fish. These changes become especially problematic when they affect the indices of abundance (such as those based on catch-per-unit-effort: CPUE) used to assess the status of fish stocks. The use of simulated CPUE data sets with known values of underlying population trends has been recommended by ICCAT (International Commission for the Conservation of Atlantic Tunas) to test the robustness of CPUE standardization methods. A longline CPUE data simulator was developed to meet this objective and simulate fisheries data from a population with distinct habitat preferences. The simulation was used to test several statistical hypotheses regarding best practices for index standardization aimed at accurate estimation of population trends. Effort data from the US pelagic longline fleet was paired with a volume-weighted habitat suitability model for blue marlin (*Makaira nigricans*) to derive a simulated time series of blue marlin catch and effort from 1986-2015 with four different underlying population trends. The simulated CPUE data were provided to stock assessment scientists to determine if the underlying population abundance trend could accurately be detected with different methods of CPUE standardization that did or did not incorporate environmental data. While the analysts' approach to the data and the modeling structure differed, the underlying population trends were captured, some more successfully than others. In general, the inclusion of environmental and habitat variables aided the standardization process. However, differences in approaches highlight the importance of how explanatory variables are categorized and the criteria for including those variables. A set of lessons learned from this study was developed as recommendations for best practices for CPUE standardization.

Keywords: Catch/effort, Longline, Statistical models, Simulation, Stock assessment, Environmental effects
**Corresponding author*: fforrestal@miami.edu

1. Introduction

Indices of abundance derived from fishery-dependent time series of catch per unit effort (CPUE) are often an integral part of the stock assessment process. Thus, there is a need to understand the processes that might lead to biases in the indices. Nominal CPUE values are often not proportional to the abundance of the stock being assessed (Campbell, 2015, 2016; Maunder et al., 2006; Maunder and Punt, 2004). Variations in CPUE can be the result of changes in the abundance of the fish stock, shifts in movement patterns, environmental and climatic changes as well as changes in fishing strategy over time (Bigelow et al., 1999). Use of CPUE to track abundance is based on the assumption that catch (C) is related to the effort (E), the abundance (N) and the catchability (q):

$$C = qEN$$

The use of the CPUE (C/E) as an index of abundance (N) thus depends on the assumption that catchability is constant or that changes in catchability can be modeled and removed from the index. Changes in catchability can be related to any changes to the fishing gear, species targeting and fishing methods. Additionally, the spatial extent of the fish population or the fishery may shift over time, influencing the fraction of the stock that is available to each fleet. Habitat suitability, such as dissolved oxygen concentration and water temperatures in the pelagic environment, can affect fish availability or catchability (e.g., by altering fish behavior). Incorporation of environmental covariates into index standardization might address some of these issues, but this is not routinely done. Best practices for incorporating environmental variables in CPUE standardization have not been defined, which adds uncertainty in choosing standardization methods aimed at minimizing CPUE bias.

A species distribution model (SDM) and longline simulator (LLSIM) were developed to test methods of CPUE standardization, amongst other goals. This paper uses simulated longline catch data sets with known values of underlying population trends to test the robustness of CPUE standardization methods. A species distribution model for Atlantic blue marlin (*Makaira nigricans*) was developed using pop-up satellite archival tag (PSAT) data paired with detailed data describing the physical environment within the model region (Figure 1) to predict fish abundances using habitat suitability modeling (Goodyear et al., 2017; Goodyear, 2016). This approach is commonly used for predicting habitat quality from habitat suitability indices based on ecological niche theory (Hirzel and Lay, 2008). Applications to billfish species include the identification of potential new fishing grounds (Chang et. al., 2012, 2013), and forecasts of the effects of climate change (Robinson et al., 2015). This approach is paired with fishing fleet dynamics, using historical effort distribution and gear configurations of the US pelagic longline fishery. Fleet catchability was defined to be gear-specific, while spatial effort allocation mimicked observed longline fishing locations. The simulated fleet was used to sample the blue marlin populating the SDM throughout the year, producing simulated catch per unit effort data based on the interactions between fishing effort and habitat suitability (i.e., fish availability) as well as gear configuration (gear efficiency) (Forrestal, et al., in press). The historical effort and gear configurations of the US longline fleet as adapted for use in the longline simulator are extensively discussed in Forrestal et al. (In press). Four distinct population trends were simulated for blue marlin (steady, increasing, decreasing, and fluctuating) to produce simulated catch datasets. These datasets were provided to eight stock assessment scientists with expertise in standardizing CPUE indices who used methods of their choice to standardize the indices. The goals of this work are to determine how well different standardization methods currently in use capture population trends and if the inclusion of environmental and habitat data aids in the standardization process.

2. Material and methods

2.1 Species distribution model

The simulated population model is defined in two steps. The first input is the population abundance in each year and month of the time series (here equal to September 1986 to December 2015). The second input is the relative population density per one-degree latitude and longitude and water depth gradient defined by the SDM (Goodyear et al., 2017; Goodyear, 2016) based on the species habitat preferences for each model time-step. The densities were normalized so that the sum of the products of the relative density x volume over each latitude, longitude, and depth = 1.0. The SDM provided the average distribution of the entire population by month and year during hours of daylight and nighttime to account for diel vertical redistribution. The method accounts for temporal changes in the location and volume of the habitat associated with seasonal and longer-term changes in the environment. For example, it directly estimates the vertical density distributions in areas affected by the oxygen minimum zones (Stramma et al., 2012). The SDM uses published blue marlin oxygen tolerance information (Brill, 1994), coupled with temperature utilization and day-night ΔT patterns from PSAT-tagged blue marlin to predict the species distribution from the detailed environmental data (Goodyear et al., 2017; Goodyear, 2016).

Four population trends were used in this study, a constant population of 500,000 individuals, a decreasing population with a 70% reduction over 29 years, an increasing population by 70% over 29 years and a population that fluctuated around 500,000 individuals over the time period (Fig. 2-4). The declining pattern is roughly equivalent to the values estimated in the most recent assessment (Anon, 2012) and the increasing population is its mirror image.

2.2 Environmental Data

Modeling the spatial distribution of a species requires quantitative data about the physical environmental variables that determine its habitat. Temperature and to a lesser extent dissolved oxygen concentration influence blue marlin habitat use (Block et al., 1992). Environmental data were obtained though the Community Earth System Model (CESM1), which is a global ocean-sea-ice model coupled to a biogeochemistry model BEC (Biogeochemical Elemental Cycle) (Danabasoglu et al., 2012; Long et al., 2013). The model covers the global ocean with a latitudinal and longitudinal resolution of 1.0° and 60 vertical layers with the bottom level at 5,500 m. Annual data outputs from CESM were available through 2012. Mean values from the final year were used to parameterize the species distribution model for 2013-2015.

2.3 Longline simulation model

The core element of the longline simulator is the catch on a single hook of a longline set. The catch is a probabilistic event and is simulated for each hook of each set. The X-Y spatial structure of the simulator is from 35^0S to 55^0N latitude and 95^0W to 20^0E longitude, exclusive of major land masses. This area is broken down into 7,067 cells; each cell is 1 degree of latitude by 1 degree of longitude. Each longitude-latitude cell is also divided into 46 depth strata of unequal size, corresponding to the environmental depth data. Conceptual details are presented in Goodyear et al. (2017) and Forrestal et al. (in press), but fundamentally involve the integration of population size, an essential gear coefficient (k) and a habitat coefficient (w) for each set. The habitat coefficient integrates the hook-depth probabilities at depth for each hook on a simulated set with the species relative density at the latitude and longitude of the set in each of the 46 depth layers apportioned by the proportion of the set that fishes at that depth in hours, separated between daylight and darkness.

2.4 Data Analysis

The longline simulator outputs a catch by set file with column headings typically observed in pelagic longline fishery logbook data. For this exercise, the variables included with the number of blue marlin caught were: total number of hooks, hook type, bait type, number of light sticks, hooks between floats (HBF), month, year and latitude and longitude (Table 1). Hook type had four levels: circle hooks, J hook, a combination of circle and J hooks and unknown hook type. Bait type used was artificial, live, dead or unknown. The light sticks were binned values corresponding to unknown light sticks reported, zero light sticks deployed, 1-500 and 501-1500 light sticks. Hooks between floats numbered between 2 through 6. These variables are referred to as the gear variables and include those that are traditionally used for CPUE standardizations. The sea surface temperature (SST) and the dissolved oxygen (DO) at the surface for the location, month and year from 1986-2012 were also supplied from the outputs of the CESM and are referred to as the environmental variables. While the SST and DO were available from the model by depth, only the surface data were included to mimic the type of data available for CPUE standardization. All simulated fishing sets were included in the final data set, including those that did not catch blue marlin.

Four simulated catch datasets corresponding to the alternative population trends were distributed to eight analysts across several ICCAT contracting or cooperating countries (i.e., CPCs). These analysts have extensive knowledge and experience developing standardized indices of abundance from fisheries-dependant CPUE data. The work was carried out in a blind-study approach, the analysts were not aware of the true population trends or the species being simulated in the dataset. The analysts developed their own approach to the data without consultation with the authors or the other analysts (Table 2). Some analysts provided more than one standardized index for each population due to their personal preference. The details of each analyst's approach are summarized below. Analysts 1-3 did not have access to population 4 as this dataset was developed later in the study.

2.4.1 Analyst 1

Analyst 1 used a delta lognormal approach in R to standardize CPUE Factors were included if they explained at least 5% of the variance. Any two-way interactions that explained at least 5% of the variance were included as random effects, using the *glmer* function in the lme4 library for R (Bates et al., 2015).

The CPUE of blue marlin was calculated as catch per thousand hooks. The potential explanatory variables were year (1986-2015), hooks between floats (either as a number, centered by subtracting the mean or as a factor), area (the 9 ICCAT areas for billfish; ICCAT, 2016, Online Supplementary Fig.1), season (months 1-3, 4-6, 7-9, 10-12), bait type (5 levels), hook type (4 levels) and light sticks (4 levels). Sea surface temperature and DO were not available for all years, so they were only used in alternate runs ending in 2012. Both variables were coded as factors (SST <15,15-20,20-25,25-30, DO <4.5,4.5-5, >5) (Table 3).

The gear variables were not evenly distributed in time and there were many combinations of variables that did not exist. Therefore, some factors were combined or eliminated before running the models. Data from the South Atlantic (ICCAT billfish areas 96 and 97; Online Supplementary Fig. 1) was excluded since there were very few observations, with none in recent years. Hook types 2 and 5 and bait type 1 and 3 were excluded due to low numbers of observations. The final dataset included 96.5% of the total observations for all populations. The trend in CPUE was calculated as the probability of presence (calculated as the inverse logit of the year effect in the binomial model) times the mean CPUE when present (calculated by converting the year effect in the model from normal to lognormal). The Lo et al. (1992) method was used to calculate the standard errors.

2.4.2 Analyst 2

Analyst 2 used a negative binomial GLMM to standardize the catch in number, with effort taken to be an offset. The models were run consecutively in R using the MASS, nlme and lme4 packages (Pinheiro et al., 2017; Venables and Ripley, 2002). Latitude and longitude were grouped into four areas (SE, NE, SW, NW) and months were grouped into quarters. This analyst used four models including a full model that contained year, area, quarter, hook type, bait type and light sticks. This model was repeated with the inclusion of sea surface temperature. This analyst did not use dissolved oxygen as it was highly negatively correlated to sea surface temperature. SST was treated at a continuous variable. The final two models contained year, area and quarter with and without SST. An offset term of the natural log of total hooks was used in the both the simple and full model.

Interaction effects were not used for any of the models. Deviance explained was used as the main model selection criteria along with ANOVA and F tests (at the 0.05 level). The year effects were estimated from the marginal mean in R given all other factors and variables.

2.4.3 Analyst 3

Generalized linear models were run in R using the packages lsmeans and glmmADMB (Fournier et al., 2012). First, the annual CPUE observations were plotted as histograms to examine distribution shape and determine candidate models for estimating index variance. Goodness-of-fit tests (chi-squared for discrete distributions, and Kilmogorov-Smirnov for continuous distributions) were ran to evaluate the best-fit model to the observed data. The samples were assigned to spatial zones defined by the Southeast Fishery Science Center (Online Supplementary Fig. 2). From there, a delta gamma model was selected that included year, month, area, and all gear variables as factors. Model performance was assessed by model convergence and residual error distribution. The model structure was the same for the model that contained environmental data. Sea surface temperature was treated as a continuous variable, and dissolved oxygen was not used as it was found to be correlated to sea surface temperature (Table 3). The binomial model and the gamma model used all the factors with single term fixed effects. No interaction terms were used, and no observations were discarded. Temporal trends in samples sizes indicated an imbalance or temporal shift in the distribution for several factors, particularly gear, hook type, bait, hooks between floats, and area fished. This diagnostic was used as a principle tool to select factors for inclusion in the standardization model. The final model covariates were selected primarily by examining boxplots of the mean and variance of CPUE observations across model factors to examine which covariates appeared to influence CPUE and varied in sample distribution over time and secondarily, Akaike's Information Criterion (AIC) of nested models.

2.4.4 Analyst 4

This analyst was the only one to utilize a Generalized Additive Model (GAM). SAS® was used as the statistical oftware (Schlotzhauer and Littell, 1997). The GAM models were used in the delta lognormal framework to develop indices. The models applied to each population were the same and incorporated environmental variables. Smoothing splines were applied to SST, hooks, latitude, longitude, surface DO, light sticks and hooks between floats (HBF). Months, years, bait type and hook type were treated as categorical variables. The success component was modeled using a binomial distribution and the abundance component was modeled using a Poisson distribution.

2.4.5 Analyst 5

Analyst five used a delta lognormal approach implemented using Generalized Linear Mixed Models (GLMM). Analyses were conducted using the *glimmix* and *mixed* procedures from the SAS® statistical computer software (Schlotzhauer and Littell,, 1997). This analyst employed an extensive graphical exploration of the datasets, including a spatio-temporal analysis to define geographical areas and seasonality of the fishery (Online Supplementary Fig. 3). The relationship between potential factors and the nominal ln(CPUE) of the positive sets were examined using proportional boxplots. Bivariate plots were used to examine the relationships between the ln(CPUE) and the environmental variables paired with smoothing fits. The selection of the final model was based on AIC, BIC, and a χ^2 test of the difference between the [–2 log likelihood] statistic of a successive model formulations (Littell et al., 1996). Interaction effects were used, and they were assumed to be random. The model structure was constant across all four populations (Table 3) and one standardized trend was obtained for each population that contained both the gear and environmental variables (Figure -4). Relative indices for the delta model formulation were calculated as the product of the year effect least square means (LSmeans) from the binomial and the lognormal model components. The LSmeans estimates use a weighted factor of the proportional observed margins in the input data to account for the non-balance characteristics of the data. LSMeans of lognormal positive trips were bias corrected using Lo et al., (1992) algorithms.

2.4.6 Analyst 6

Analyst 6 used a Tweedie Generalized Linear Model; analyses were conducted using R and the tweedie (Dunn and Smyth, 2005, 2008), lsmeans (Lenth, 2016) and mfp (Ambler and Benner, 2015) packages. The Tweedie GLM approach does not split the response variables into success and abundance of CPUE and then apply two separate models as is the case with the delta approach used by other analysts (

Table 4). The only response variable was CPUE measured as number of blue marlin caught per 1000 hooks, which is a continuous variable with an added mass of zeros for the cases of sets with zero catches. The categorical variables included in the final model were: year, month, light, hook type, bait type and hooks between floats. The spatial variables latitude and longitude were grouped into categorical areas using regression trees, according to the method developed by Ichinokawa and Brodziak (2010). The environmental variables sea surface temperature and dissolved oxygen were used as continuous variables transformed with fractional polynomials, using the method developed by Royston and Altman (1994).

Initially, univariate models were applied for each candidate variable. Significance for inclusion were likelihood ratio tests comparing univariate models to the null model. All significant variables (5% level) were then used for a multivariate model. In the multivariate model, the final significance of each variable was analyzed using deviance tables, AIC and pseudo R^2. The final models were slightly different for each population as the area categorizations and polynomial transformations were specific to each population dataset (

Table 4). No interaction effects were used due to computational restraints. The year effects were extracted in the same manner as analyst 3.

2.4.7 Analyst 7

This analyst used a delta lognormal GLMM approach to standardize the CPUE data. The statistical software employed was R with the *glmer* function of the lme4 package (Bates et al., 2015). None of the models included environmental variables due to computational constraints and the lack of environmental data in the most recent years. Latitude and longitude were grouped into three areas, a northern region (including the Gulf of Mexico), southern and Caribbean region. Successes were modeled using a binomial distribution, and abundances using a Gaussian distribution. Variables were included in the final model if they explained 5% or more of the deviance. The models used to standardize populations 2, 3 and 4 were the same while the model applied to population 1 contained interactions between year and some of the other explanatory variables (Table 3). If interactions with year were significant, they were treated as random effects. But in most cases, interactions could not be tested due to lack of computing power. The year effect was extracted by taking the year coefficients in both models and then transforming and corrected them according to Lo et al. 1992

2.4.8 Analyst 8

Analyst 8 used a delta lognormal GLM approach. The analyses were conducted using SAS proc *glimmix* for the binomial component and SAS proc *mixed* for the lognormal component. This analyst developed eight models, a different model for each population and models with and without the environmental variables (Table 3). Latitude and longitude were grouped into the US pelagic longline logbook areas (Cramer, 1983). The Goodman (1960) exact method for calculating the variance of two independent random variables was used to obtain the variance. Two methods commonly employed to select models were used; the method of Ortiz and Arocha (2004), which uses the percent reduction in explained deviance to select factors that explain greater than a certain percentage and the method of Brown (1992), which uses the percent deviance reduction per degree of freedom. A 5% cut-off was used for all models, which is commonly used for each method. Environmental variables were originally entered as categorical and were changed to continuous (SST*SST and surface DO) due to model fitting issues. The yearly index was extracted using the SAS lsmeans statement.

2.4.9 Analysis of standardized trends

Standardized trends from the eight analysts and the true population trends were normalized to the mean to examine differences among the time series. The normalized, modeled CPUE trends were regressed to the normalized, underlying population trends. Root mean square errors (RMSEs) were estimated using residuals between the population trend and the standardized CPUE to quantify the accuracy of each standardization. Further examination of model fits were estimated using the median absolute relative error (Ono et al., 2015, Online Supplementary Table 1). The average RMSE for all analysts within populations for models with and without environmental variables were compared with a t-test or Mann-Whitney *U*. The mean standardized trends with and without environmental covariates were plotted using ggplot2 and Hmisc packages (Wickham, 2009; Harrell, 2017).

3. Results

3.1 Population 1

Population 1 led to the lowest average RMSE of the four populations examined for the model types that included only gear variables and those with environmental variables added (Table 5). The models that contained environmental variables had lower RMSE for all the analysts that examined both model types. However, there was no difference between the models that used the environmental models and those that did not (two-sample t (12) =1.49, p=0.16, Table 5). Two general patterns emerged from examining the standardized CPUEs in comparison to the population trends: (1) standardized CPUEs that fluctuated around the true population and (2) an overestimation of population size in the start of the time series and an underestimation beginning in 2002. The five models that underestimated the true values after 2002 did not include hook type in their final model. The exception to this trend was analyst 5 who did include hook type in the final model structure. This analyst was also the only one to use a GAM approach.

The trends obtained by analysts 1, 2, 4, 7 and 8 exhibited a drop in population size in 2002 that did not occur in the true population trend (Figure). Analyst 1 noted that hook type was not used in the final model as it did not explain more than 5% of the deviance observed. Analyst 2 used the environmental data in a model with only year, quarter and area (SE, NE, SW, NW) as factors and a full model with all possible variables (models environment 1 and 2 respectively, Figure). The simpler model with environmental data had the drop observed in 2002. However, adding the environmental data smoothed the trend out even though hook type was not included. Both versions of the complete model (Gear 2 and Environment 2) had a very close agreement to the true population trend time series.

Both time series obtained by analyst 3 fluctuated around the true population trend as did analyst 6's time series. However, the error was lower for analyst 6. This pattern was also observed in three of analyst 2's models although those standardized trends did not fluctuate around the true population. The RMSE for those three models were the lowest across all models and populations.

Analyst 5's standardized time series also fluctuated around the true population. However, starting in 2012, the standardized trend greatly overestimated the true population size. This analyst utilized SAS and incorporated the environmental variables into the final model. The environmental data points did not extend past 2012. Analysts that used these variables truncated the standardized CPUE at 2012 to account for the shorter time series. This was either discovered through an initial exploration of the data or, if R was used as the statistical software, the software automatically excluded records with data, in this case, environmental data. However, analyst 5 used SAS which runs with years that contain missing data but uses the average value of the missing variable; this resulted in predictions for these years diverging from the true values. There are estimated values from the model including environmental effects in the 2013 and 2014, but they are highly uncertain. This occurred with all models across the four populations for analyst 5. For comparison purposes to other analysts, the model residuals used in the RMSE analysis were from 1986-2012.

3.2 Population 2

The population 2 dataset contained a declining population trend and all the analysts were able to capture the decline. In general, the standardized CPUE overestimated the true population size in the earliest years of the dataset. However, in the most recent years, the analysts either accurately estimated or underestimated the true population size. As was observed in population 1, the models with the environmental variables had a non-significant lower average RMSE than those models that did not incorporate the environmental covariates (Mann-Whitney U=18.0, n_1=6, n_2=8, p=0.49, Table 5). However, whether environmental variables reduced RSME varied by analyst. Models including the environmental variables had a higher RMSE for analysts 1 and 3, but not for analysts 2 and 8 (Table 5).

Analyst 1 treated hooks between floats as a factor for population 2 as the relationship between HBF and CPUE was not as clear for in population 1. Analyst 8's binomial gear model only contained year and area.

The time series obtained from analysts 1, 2, 4, 5 and 8 did not match the true population trend in the earliest years (1986-1993), which corresponded to the highest CPUE values (Figure). In later years, the modeled trends converged on the true population trend for analysts 3, 6 and 7. Analysts 1, 2, 4, and 8 underestimated the true population size in the most recent years. The time series from analyst 5 followed the true population trend before the extreme values began in 2013.

3.3 Population 3

Population 3, which had an increasing population size, had the largest discrepancy between modeled values and the true population values as measured by the RMSE (Table 5). As with populations 1 and 2, the environmental models had a lower error than the gear models, but again the difference was not significant (two-sample t (12) =0.87, p=0.40, Table 5).

The model produced by analysts 1, 2, 4 and 7 overestimated the population size in the earliest years and underestimated in the later years (Figure). The environment models for analysts 3, 6 and 8 all had very similar patterns, closely following the true population trends from 1986 to 2002 and then exhibiting a spike of overestimation in 2008 and again in 2012. The gear models for analysts 1, 2, 7 and 8 underestimated the true population size starting in 2004; the inclusion of environmental variables corrected the underestimation in analyst 8's model, but not for analysts 1 and 2. An examination of the mean standardized trends shows an overall overestimation of the earliest years population for both the gear and environmental models and an underestimation of both models beginning in 2004. However, the environmental models track closer to the true population trend (Figure).

3.4 Population 4

There are results from five analysts for population 4 as opposed to eight for the other populations. This is the result of this dataset being distributed to the analysts later in the study. This dataset represents a fluctuating population with two occurrences of population decline and resurgence. For this population, the gear models had a lower mean RMSE than the environment models, although this was not significant (two-sample t (4) =-0.135, p=0.89, Table 5).

Analyst 6 and 7 were able to track the true population's fluctuations quite well (Figure) while analysts 4 and 8 overestimated population size in the first year and then underestimated population size starting in 2005. Analyst 5 was able to capture the initial population trend quite well before a similar underestimation of the population starting in 2005. The two mean model trends were quite similar from 1986 until 1995, with the environmental model tracking closer to the true population trend from 1995 to 2005. After 2005, both models underestimated the true population with very similar observed patterns (Figure).

4. Discussion

The aim of this study was to examine some of the methods employed by ICCAT CPC scientists who are routinely tasked with creating indices of abundance for the fisheries they participate in and to determine if these methods were able to reliably capture the underlying population trend in the provided datasets. The results of this work highlight the wide range of standardization approaches taken as a result of each ICCAT member country conducting their own analysis. The strengths of the ICCAT approach is that it is an inclusive process that subjects the analysis to review from other national scientists and allows those that are most knowledgeable about the fisheries to conduct the analyses. However, the weakness of this approach is the use of various methodologies can lead to conflicting CPUE trends that may or may not be reflective of the true biomass. Other tuna regional fishery management organization (tRFMO; e.g., WCPFC - Western and Central Pacific Fisheries Commission) differ from the approach of having each CPC scientist produce standardized CPUE trends and instead utilize the tRFMO Secretariat or the services of other advice bodies, such as SPC (Pacific Community). This leads to consistent standardization techniques applied over different datasets and over time. However, weaknesses of this approach are that it tends to exclude member countries' scientists, and the analysts conducting the analysis may not have the same level of understanding of the fisheries as member country scientists. An effective compromise between these differing approaches may involve having the national scientists conduct their own analyses, but with generally consistent and agreed upon methods of standardization.

While the analysts' approach to the data and the modeling structure differed, most models were able to capture the underlying population trends well. However, differences in performance highlight the importance of how spatial dimensions are defined, how categorical variables are grouped, how continuous variables are modeled and, importantly, the criteria for model selection. The analysts used different area combinations for the spatial structure of their models, some grouping latitude and longitude according to the ICCAT areas for billfish, and others using the raw 1x1° latitude and longitude values. Analyst 6 utilized a regression tree approach, which led to different area groupings for each population. Analyst 2 used the spatial domain of the observations to define four areas of equal quadrants based on the magnitude of effort. The variables included in the final model also differed between analysts. Hook type was excluded from the models developed by several of the analysts. Nominal catch rates for population 1 were higher, prior to the switch from J-hooks to circle hooks in 2004 and then were systematically lower than the true population CPUE. Models that failed to include hook type often failed to re-create the true population trend. Analyst 8 conducted model selection independently for each population, noting that models did not converge when hook type was included.

The addition of environmental variables improved the accuracy of estimates of the population size across all populations with a few exceptions, such as when SAS filled in missing data with mean environmental values for analyst 5. The inclusion of these variables in the cases of analyst 1 for population 2 and all the populations for analyst 3 resulted in a higher RMSE values and these models did not follow the true population values as well as the models that did not contain the environmental variables. Environmental variables are thought to be good predictors of density of a species in the vicinity of the set and/or hook. Environmental variables that determine suitability of adjacent habitat should improve estimation of CPUE by accounting for differential availability of a species in the vicinity of the set and/or hook. However, given the linear nature of GLM models, suitable transformation of the data (continuous explanatory variables) may be necessary, such as polynomials (e.g., SST*SST^2) to mimic species' habitat preference curves. Also, the values of environmental variables at the surface may not be highly correlated with the values at depth that influence species' distributions. Future studies should take advantage of the CESM data outputs at the actual depths where blue marlin and the hooks are located.

While the use of environmental variables increased accuracy, their inclusion also increased the annual CVs compared to the models without the environmental variables (e.g. see CVs for analyst one, Online Supplementary Table 2), likely due to the added requirement of estimating a relatively imprecise relationship between catch rates and SST or DO. In theory, a strong relationship between a species density and environmentally-mediated habitat suitability may exist and is a fundamental part of the species distribution model (Goodyear et al., 2017). However, within the statistical models estimated in this exercise, this relationship is estimated from noisy CPUE data which may lead to relatively imprecise parameter estimates in the models and higher CVs as compared to not including SST or DO. Additionally, if there is insufficient contrast in the data to estimate the coefficients related to the environmental predictor variables, the estimates may be very imprecise, and possibly biased. This could be the case with fishery-dependent data where fishers may only fish in good temperature windows so the necessary contrast to estimate a CPUE-SST relationship is missing. Further improvements in the concept of habitat modeling such as occupancy modeling or use of ancillary information from tagging or tracking in the form of Bayesian priors may provide improvements in both the accuracy and precision of CPUE-based abundance indices when including environmental data.

The inclusion of the environmental variables caused a problem for the SAS-based analyses. Incomplete SST and DO values for the last two years caused the models of analyst 5 to diverge substantially from the true values. Most analysts did not, or their software packages could not, estimate the year effects for the years with the missing environmental variables. The SAS models converged, but estimates for the last two years were incorrect. This situation highlights the problem that missing data creates for CPUE standardization. Environmental data such as SST, DO, etc. are likely to be missing, due to either not being recorded, or, if assigned based on satellite oceanography, missing due to cloud cover. Hence missing data are commonplace and the model results can depend upon how the missing data are treated. It is therefore critical to examine *a priori* whether missing data exists and to decide how it is going to be treated rather than allowing software to use default settings.

The poor performance of some models implies that standard model selection criteria such as those based on either a 1 or 5% reduction in deviance per degree of freedom can often fail to select key factors, in this case, hooks between floats or hook type, that affected catchability. Hook type had a substantial impact on CPUE in the true populations. Hook type in the fishery changed as a result of regulations from J-hooks to circle hooks in 2004. This shift in hook type resulted in a substantial decrease in the nominal CPUE relative to the true populations and was manifest in all of the four populations. Unfortunately, the knife-edge change in hook type meant that the years pre- and post-2004 and hook type did not overlap, causing hook type not to be selected using deviance explained. This result illustrates model selection methods based only on reduction in deviance may be prone to error regarding factor exclusion and that analysts should err on the side of keeping factors in the models. This is particularly the case if *a priori* exploratory analyses or knowledge of the fishery indicate that the variable could affect CPUE, which is surely the case with hook type or hooks between floats. Ortiz and Arocha (2004) found that variables that explained more than 5% of total deviance were generally significant according to likelihood ratio tests, which supports the use of 5% deviance explained in model selection. However, this selection method supports models with fewer variables than the AIC and BIC, which frequently include variables that are not significant in the best models. It should be noted that model selection criteria such as AIC and BIC supported including hook type. These methods of model selection have a better theoretical basis than *ad hoc* methods such as deviance explained, so more frequent use of them is warranted (Gelman et al., 2014). Our results indicate that these more complex models were better at predicting the overall trend, supporting the use of information criteria rather than deviance explained in CPUE standardization. While including many variables in a model may result in decreased model performance such as failed convergence, requiring selection of a subset of variables, most fishery-dependent CPUE standardization data sets have very high sample sizes relative to the number of model factors so over-parameterization is rarely a concern.

Residual patterns emerging from the model fits to population 2 (the decreasing population) were a possible indication of high collinearity between the year effect and at least one other estimated parameter. Direct knowledge of the fishery and proper *a priori* examination of the raw data was critical in realizing the true population trend was correlated with hooks between floats in *post hoc* analysis. As the true population declined, the average depth of hooks increased. Strong collinearity between the year effect and other parameters can lead to confounding in parameter estimates and thus an inability of the model to distinguish between the correlated trends and thus produce an accurate estimate of the true population trend. However, this association could not have been detected without knowledge of the true population trend. Thus, collinearity between factors and the year effect needs to be inferred rather than detected by a means dependent on knowledge of the true population trend.

Three analysts modeled the population with several year×factor interaction terms, which cause problems for interpretation of strict year effects (Maunder and Punt, 2004). Certain non-year interactions, such as month×area or area×season could be manifestations of the migratory behavior of blue marlin. The month factor signifies something different in a northern region than in a southern region, which is straightforward to explain. In contrast, interactions with year are harder to explain, and represent a potential confounding of the abundance signal with another model factor, such as gear changes or environment.

A common approach when year×factor interactions are significant is to model them as random effects as was done by several analysts. Unfortunately, modeling year×factor interactions as random effects can lead to several problems. First, random year×factor interactions can affect the parameter estimates for other variables. Second, it is important to plot year×factor parameter estimates and their standard errors to determine if they are actually random and not showing trends with respect to either year or the other variable in the interaction. Given the potential for serial depletion (Walters, 2003) or range shifts in populations due to climatic factors and the high probability of models finding spurious year×factor interactions, plots of the interaction terms provide critical information about patterns in these interactions. Truly random interactions would look random or would fail to reject a test of randomness. Significant interactions could exist as a single outlier year, which might not merit modeling or substantially trended interactions with year which require additional considerations as to why the population signal differs with different values of another factor. While several analysts used interaction terms, the interactions did not consistently improve the accuracy of the estimated trends. Future studies employing a factorial design to specifically compare different model types will further explore the use of interaction terms.

Several of the results point to problems in current CPUE standardization approaches. The different performance of standardization methods, and the different performance with different methods for defining geographical areas raise some concerns about the ability of models to estimate population trends. Using an adaptive area partitioning method, Analyst 6 estimated different spatial partitionings for each population, even though each population had the same model factors operating and the same spatial structure. This indicates a possible dependence between the population trend and the estimation of the model parameters other than the year effect which is intended to capture the trend. It may be possible to diagnose adverse correlation between year and other factors by examining variance inflation factors (VIFs) or by examining the covariance between 'year' and other model coefficients. High VIF or high covariance with year indicate that the model cannot separate the abundance trend from a trend in other model factors.

5. Conclusions

This study with simulated longline datasets sought to determine if standardization methods used by the ICCAT CPCs scientists can routinely capture underlying population trends from fishery-dependant CPUE data and to derive a set of 'best practices'. Overall, despite the diversity of distributional assumptions, model selection methods, software and treatments of variables, most models were able to capture the underlying population trends. The inclusive stock assessment practice utilized by ICCAT allows the scientists most familiar with the specific, regional fleet to develop standardized CPUE time series that are then used as proxies for relative abundance trends in the stock assessment models. The downside to this practice is the wide variation in methodology, which may contribute to conflicting trends for the same species, and may be an artefact of standardization methodology rather than a true difference in signal between datasets. Thus, it is important that standardization methods be reviewed carefully before indices are used in assessment, and that multiple methods be applied to the same datasets to identify whether estimated trends differ with standardization methodology.

This exercise highlights that there are several problems with some of the *status quo* approaches that warrant further exploration: unknown correlations between model factors and the year effect that can confound estimation of the population signal, the usefulness of standard model selection criterion to choose the correct models, and the dangers posed by missing data depending upon how a modeling platform deals with it.

As a result of this work, we have developed a set of lessons learned:

1) Priority of variable inclusion or exclusion should be based on a first principles knowledge of the fishery and the historical management measures that have taken place. If known changes in the fishery have occurred (e.g., changes in legal retained size, geographic distribution of fish and/fishery, changes in gear type) then these variables should be given the highest consideration for inclusion, whether or not model diagnostics support their inclusion. Alternatively, in cases where such variables cannot be accommodated in the statistical models due to technical issues, the CPUE series may have to be split and modeled as several independent time series to reflect those unaccounted changes in catchability.
2) *A priori* evaluation of model balance across factor combinations over time and plots of CPUE time series by model factors are absolutely critical to determining which model factors are important or missing. This procedure would have captured the knife-edge switch in hook types in 2004 and the missing environmental data.
3) Evaluation of multiple-collinearity of model variables with the year factor is essential. Strong collinearity with the year effect results in a GLM not being able to distinguish between inter-annual changes in abundance and those in the correlated variable.
4) Embrace divergence of the nominal CPUE from the standardized model estimate. Often, the observation is made that the standardized trend diverges from the nominal as a shortcoming against the model selected. The lack of divergence between nominal and standardized trends is often used as a *post hoc* diagnostic of model performance. In the examples within this study, the only way to have obtained the correct estimate of the true population was to depart substantially from the nominal trend.

Acknowledgements

C. P. Goodyear's contribution to this research was supported by The Billfish Foundation. F. C. Forrestal's and E. Babcock's contribution was supported by the NOAA Southeast Fisheries Science Center via the University of Miami Cooperative Institute for Marine and Atmospheric Science. R. Coelho research is supported by an *Investigador*-FCT contract (Ref: IF/00253/2014) from FCT, the Portuguese Foundation for Science and Technology. The authors wish to thank Sang-ki Lee and Yanyun Liu of NOAA Atlantic Oceanographic and Meteorological Laboratory for gathering and compiling the oceanographic data. We also wish to thank Guillermo Diaz and Allison Shideler for gathering and compiling the US longline effort data. Declarations of interest: none.

References

Ambler, G., Benner, A,. 2015. mfp: Multivariable Fractional Polynomials. R package version 1.5.2. Available at: https://CRAN.R-project.org/package=mfp

Anon., 2012. Report of the 2011 Blue Marlin stock assessment and white marlin data preparatory meeting. Col. Vol. Sci. Pap. ICCAT, ICCAT 68(4), 1273-1386. www.iccat.int.

Bates, D., Machler, M. Boker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Soft. 67, 1-48.

Bigelow, K.A., Boggs, C.H., He, X., 1999. Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. Fish. Oceanogr. 8, 178–198.

Block, B. A., Booth, D. T., Carey, F. G., 1992. Depth and temperature of the blue marlin, *Makaira nigricans*, observed by acoustic telemetry. Mar. Biol. 114, 175-183.

Brill, R.W., 1994. A review of temperature and oxygen tolerance studies of tunas pertinent to fisheries oceanography, movement models and stock assessments. Fish. Oceanogr. 3, 204–216.

Brown, D., 1992. A Graphical Analysis of Deviance. J Roy. Stat. Soc. Ser. C Appl. Stat. 41(1), 55-62.

Campbell, R.A., 2015. Constructing stock abundance indices from catch and effort data: Some nuts and bolts. Fish. Res. 161, 109-130.

Campbell, R.A., 2016. A new spatial framework incorporating uncertain stock and fleet dynamics for estimating fish abundance. Fish Fish. 17, 56-77.

Chang, Y. J., Sun, C. L., Chen, Y., Yeh, S. Z., DiNardo, G., 2012. Habitat suitability analysis and identification of potential fishing grounds for swordfish, *Xiphias gladius*, in the South Atlantic Ocean. International J. Remote Sens. 33, 7523-7541.

Chang, Y. J., Sun, C. L., Chen, Y., Yeh, S. Z., DiNardo, G., Su, N.-J., 2013. Modelling the impacts of environmental variation on the habitat suitability of swordfish, *Xiphias gladius*, in the equatorial Atlantic Ocean. ICES J. Mar. Sci. 70, 1000-1012.

Danabasoglu, G., Bates, S.C., Briegleb, B.P., Jayne, S.R., Jochum, M., Large, W.G., Peacock, S., Yeager, S.G., 2012. The CCSM4 ocean component. J. Clim. 25, 1361–1389.

Dunn, P.K., Smyth, G.K., 2005. Series evaluation of Tweedie exponential dispersion models. Stat. Comp. 15(4): 267-280.

Dunn, P. K., Smyth, G. K., 2008. Evaluation of Tweedie exponential dispersion models using Fourier inversion. Stat. Comput. 18(1), 73-86.

Cramer, J., 1993. Large Pelagic Logbook Newsletter. NOAA technical memorandum NMFS-SEFSC 322. https://repository.library.noaa.gov/view/noaa/8716,

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M., Nielsen, A., Sibert, J., 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw., 27, 233-249.

Forrestal, F.C., Goodyear, C.P., and Schirripa, M.J., (in press). Applications of the longline simulator (LLSIM) using US pelagic longline logbook data and Atlantic blue marlin. Fish. Res.

Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. Stat. Comput. 24, 997-1016.

Goodman, L. A., 1960. On the exact variance of products. J. Am. Stat. Assoc. 55(292), 708- 713.

Goodyear, C.P., 2016. Modeling the time-varying density distribution of highly migratory species: Atlantic blue marlin as an example. Fish. Res. 183, 469-481.

Goodyear, C.P., Schirripa. M., Forrestal, F., 2017. Longline data simulation: a paradigm for improving CPUE standardization. Col. Vol. Sci. Pap. ICCAT 74(2), 379-390. www.iccat.int.

Harrell Jr, F., E., 2017. Hmisc: Harrell Miscellaneous. R package version 4.0-3. https://CRAN.R-project.org/package=Hmisc

Hirzel, A. H., Lay, G. L., 2008. Habitat suitability modelling and niche theory. J. Appl. Ecol. 45, 1372–138.

ICCAT. 2016. ICCAT Manual. International Commission for the Conservation of Atlantic Tuna. In:ICCAT Publications [on-line]. http://www.iccat.int/en/ICCATManual.asp.

Ichinokawa, M., Brodziak, J., 2010. Using adaptive area stratification to standardize catch rates with application to North Pacific swordfish (*Xiphias gladius*). Fish. Res. 106, 249–260.

Lenth, R.V., 2016. Least-Squares Means: The R Package lsmeans. J. Stat. Softw., 69(1), 1-33.

Littell, R. C., Milliken, G., Stroup, W.W., 1996. *SAS system for mixed models*, SAS Institute, Inc., Cary, NC.

Lo, N.C.H., Jacobson, L.D., Squire, J. L., 1992. Indexes of relative abundance from fish spotter data based on delta-lognormal models. Can. J. Fish. Aquat. Sci. 49, 2515-2526.

Long, M.C., Lindsay, K., Peacock, S., Moore, J.K., Doney, S.C., 2013. Twentieth-century oceanic carbon uptake and storage in CESM1 (BGC). J. Clim. 26, 6775–6800.

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

Maunder M.N., Sibert J.R., Fonteneau A., Hampton J., Kleiber P., Harley S.J., 2006. Interpreting catch per unit effort data to assess the status of individual stocks and communities. ICES J. Mar. Sci. 63, 1373–1385.

Ono, K., Punt, A.E., Hilborn, R., 2015. Think outside the grids: An objective approach to define spatial strata for catch and effort analysis. Fish. Res. 170, 89-101.

Ortiz, M., Arocha, F., 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. Fish. Res. 70, 275-297.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2017. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-131. https://CRAN.R-project.org/package=nlme.

Robinson, L. M., Hobday, A. J., Possingham, H. P., Richardson, A. J., 2015. Trailing edges projected to move faster than leading edges for large pelagic fish habitats under climate change. Deep Sea Res. Part 2 Top. Stud. Oceanogr. 113, 225-234.

Royston, P., Altman, D.G., 1994. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. J. R. Stat. Soc. Ser. C Appl. Stat. 43(3), 429–467.

Schlotzhauer, S., Littell, R., 1997. SAS System for Elementary Statistical Analysis, Second Edition. SAS Institute, Inc. Cary, NC.

Stramma, L., Prince, E.D., Schmidtko, S., Luo, J., Hoolihan, J.P., Vesbeck, M., Wallace, D.W.R., Brandt, P., Kortzinger A., 2012. Expansion of oxygen minimum zones may reduce available habitat for tropical pelagic fishes. Nat. Clim. Change, 2(1), 33-37.

Venables, V. N., Ripley, B. D., 2002. Modern Applied Statistics with S. 4th Edition. Springer.

Walters, C., 2003. Folly and fantasy in the analysis of spatial catch rate data. Can. J. Fish. Aquat. Sci. 60(12), 1433-1436.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Figure captions

Figure 1. Locations of simulated fishing sets for all years (1986-2015).

Figure 2. Standardized trends for population 1 for all analysts. Environment lines signify that one or two environemntal terms were incldued in the final model. Gear models contain only variables are associated with gear type and that factors or variables that are tradtionally contained in CPUE standardization models. Population is the true population trend.

Figure 3. As for Figure 2, except for population 2.

Figure 4. As for Figure 2, except for population 3.

Figure 5. As for Figure 2, except for population 4. Note results are only shown for five of the analysts.

Figure 6. Mean standardized trends for all analysts. Shading surrounding lines is the standardized error around the mean.

Tables

Table 1. Available variables to the analysts, if they were categorical or continous and the levels or range included. Latitude and longitude in one ° cells, HBF=hooks between floats, SST (°C) = sea surface temperature, DO (mg/L) = surface dissolved oxygen.

Variable	Type	Range
Year	Categorical	1986-2015
Month	Categorical	1-12
Lat.	Continuous	-30°S-53°N
Long.	Continuous	-95°W-15°E
HBF	Categorical	2-6
Hook	Categorical	1-4
Bait	Categorical	1-4
Lights	Categorical	0-3
SST	Continuous	2-31
DO	Continuous	4-8

Table 2. Model format for each analyst. The method used to select the variables within the final model structure are listed under "Criteria" (AIC=Akaike information criterion; BIC=Bayesian information criterion; LRT=Likelihood ratio test). The column "Environment" denotes if environmental variables were included in the final model, if "Both", then the analyst conducted two standardizations, one with the environmental variable and one without.

Analyst	Model	Program	Criteria	Environment
One	Delta Lognormal GLMM	R	5% deviance explained	Both
Two	Negative Binomial GLM	R	5% deviance explained	Both
Three	Delta Gamma GLM	R	First principles, AIC	Both
Four	Delta Lognormal GAM	SAS	None	Yes
Five	Delta Lognormal GLMM	SAS	AIC, BIC, χ^2	Yes
Six	Tweedie GLM	R	LRT, AIC, pseudo R^2	Yes
Seven	Delta Lognormal GLM	R	5% deviance explained	No
Eight	Delta Lognormal GLM	SAS	5% deviance explained/df	Yes

Table 3. Final model selection for analysts using the delta modeling approach. If analysts used the same final model for each population, only one model is listed for that analyst. Fixed effects are shown in plain text and random effects in bold. HBF is hooks between floats, DO is dissolved oxygen, and SST is sea surface temperature. See text for details on how each analyst defined each variable.

Analyst	**Populations**	**Presence**	**Abundance**
One	All	year+HBF+area+season+**year×area+area×season**	year+HBF+area
One	All	year+HBF+area+season+SST	year+HBF+area
Three	All	year+HBF+area+month+hook+bait+light	year+HBF+area+month+hook+bait+light
Three	All	year+HBF+area+month+hook+bait+light+SST	year+HBF+area+month+hook+bait+light+SST
Four	All	SST+hooks+lat+lon+DO+light+HBF+month+year+bait+hook	SST+hooks+lat+lon+DO+light+HBF+month+year+bait+hook
Five	All	year+area+season+HBF+hook+light+bait+STT+DO+**year×area+year×season+year×HBF+year×bait+year×light+season×hook**	year+area+season+HBF+hook+light+bait+SST+DO+**year×area+year×season+year×HBF+year×bait**
Seven	1	year+area+HBF+**year×month+year×area**	year+month+area+**year×month**
Seven	2-4	year+area+hook+HBF	year+month+area
Eight	1	year+month+bait+HBF+area	year+light+hook+HBF+area
Eight	1	year+month+bait+HBF+area+DO+SST^2	year+light+hook+HBF+area
Eight	2	year+area	year+light+hook+HBF+area
Eight	2	year+month+bait+HBF+area+DO	year+month+area
Eight	3	year+month+area	year+light+hook+HBF+area
Eight	3	year+month+light+hook+bait+area+DO+SST^2	year+month+light+hook+bait+HBF+area+DO+SST^2
Eight	4	year+month+HBF+area+bait	year+light+hook+HBF+area
Eight	4	year+mont+light+hook+bait+HBF+area+DO+SST^2	year+month+light+hook+bait+HBF+area+DO+SST^2

Table 4. Final model selection for analysts using negative binomial (Two) and Tweedie approaches (Six). All variables were fixed effects. See text for how each analyst defined each variable.

Analyst	Population	Final Model
Two (1)	All	year+quarter+area+offset(ln(hooks))
Two (1)	All	year+season+area+SST+offset(ln(hooks))
Two (2)	All	year+season+area+gear+light+HBF+hook+bait+offset(ln(hooks))
Two (2)	All	year+season+area+gear+light+HBF+hook+bait+SST+offset(ln(hooks))
Six	1	year+month+light+hook+bait+HBF+area+SST^3+SST^3*log(SST)+log(DO)+$DO^{0.5}$
Six	2	year+month+light+hook+bait+HBF+area+SST^3+SST^3*log(SST)+DO^3+DO^3*log(DO)
Six	3	year+month+light+hook+bait+HBF+area+SST^3+SST^3*log(SST)+DO^3+DO^3*log(DO)
Six	4	year+month+light+hook+bait+HBF+area+SST^3+SST^3*log(SST)+DO^{-2}+DO^{-2}*log(DO)

Table 5. Root mean square errors for model fits to the true population trends.

	Population 1		Population 2		Population 3		Population 4	
	Gear	**Enviro.**	**Gear**	**Enviro.**	**Gear**	**Enviro.**	**Gear**	**Enviro.**
Analyst 1	0.288	0.252	0.193	0.271	0.327	0.274		
Analyst 2 (1)	0.157	0.016	0.339	0.304	0.422	0.417		
Analyst 2 (2)	0.016	0.016	0.349	0.304	0.420	0.417		
Analyst 3	0.083	0.101	0.101	0.129	0.105	0.146		
Analyst 4		0.238		0.169		0.272		0.229
Analyst 5		0.284		0.204		0.499		0.323
Analyst 6		0.086		0.104		0.122		0.102
Analyst 7	0.235		0.110		0.333		0.195	
Analyst 8	0.277	0.255	0.345	0.132	0.281	0.121	0.266	0.461
Mean	0.176	0.156	0.240	0.202	0.315	0.284	0.231	0.279
SE	0.045	0.040	0.049	0.029	0.048	0.052	0.036	0.076