SCRS/2017/066 Collect. Vol. Sci. Pap. ICCAT, 74(2): 391-403 (2017)

TESTING ROBUSTNESS OF CPUE STANDARDIZATION USING SIMULATED DATA: FINDINGS OF INITIAL BLIND TRIALS

Francesca C. Forrestal[1], C. Phillip Goodyear[2], Michael Schirripa[3], Elizabeth Babcock[1], Matt Lauretta[3] and Rishi Sharma[3]

SUMMARY

Increased climate variation has caused changes in the distribution, migratory patterns, and susceptibility to various fishing gears for highly migratory marine fish. These changes become especially problematic when they manifest themselves through the fishery dependent indices of abundance (such as catch-per-unit-effort, or CPUE) used to assess the status of fish stocks. The use of simulated data sets with known values of underlying population trends have been recommended by ICCAT to test the robustness of CPUE standardization methods. A longline CPUE data simulator (LLSIM) was developed to meet this requirement to simulate known data for testing a variety of hypotheses. 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 from 1986-2015 with three different underlying population structures. The resulting time series were provided to stock assessment scientists to determine if the underlying population trends could be detected with different methods of CPUE standardization incorporating environmental data.

RÉSUMÉ

*La variation climatique croissante a entraîné des changements de la distribution, des schémas migratoires et de la sensibilité à plusieurs engins de pêche de poissons marins grands migrateurs. Ces changements ont commencé à poser particulièrement problème lorsqu'ils se manifestent tout au long des indices d'abondance dépendants des pêcheries (tels que la prise par unité d'effort, ou CPUE) utilisés pour évaluer l'état des stocks de poissons. L'utilisation de jeux de données simulées avec des valeurs connues des tendances sous-jacentes de la population a été recommandée par l'ICCAT afin de tester la solidité des méthodes de standardisation de la CPUE. Un simulateur de données de CPUE palangrière (LLSIM) a été mis au point pour remplir l'exigence de simuler des données connues servant à tester plusieurs hypothèses. Les données d'effort de la flottille palangrière pélagique des États-Unis ont été associées à un modèle de qualité du volume de l'habitat pondéré pour le makaire bleu (*Makaira nigricans*) afin de calculer une série temporelle simulée de capture de makaire bleu de 1986-2015 avec trois structures de population sous-jacentes. Les séries temporelles obtenues ont été fournies aux scientifiques réalisant des évaluations des stocks afin de déterminer si les tendances des populations sous-jacentes pourraient être détectées avec différentes méthodes de standardisation de la CPUE incorporant des données environnementales.*

RESUMEN

El incremento de las variaciones climáticas ha generado cambios en la distribución, en los patrones migratorios y en la susceptibilidad a los diferentes artes pesqueros para los peces marinos altamente migratorios. Estos cambios son especialmente problemáticos cuando se manifiestan mediante índices de abundancia dependientes de la pesquería (como captura por unidad de esfuerzo o CPUE) utilizados para evaluar el estado de los stocks de peces. ICCAT ha recomendado el uso de conjuntos de datos simulados con valores conocidos de tendencias de población subyacentes para probar la robustez de los métodos de estandarización de la CPUE.

[1] CIMAS-University of Miami, 4600 Rickenbacker Cwy. Miami, FL 33149. fforrestal@ miami.edu
[2] 214 North Lakeshore Drive, Niceville, Florida 32578, USA
[3] NOAA Fisheries, Southeast Fisheries Center, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, FL, 33149-1099, USA.

*Se desarrolló un simulador de datos de palangre de CPUE (LLSIM) para cumplir este requisito de simular datos conocidos para probar diversas hipótesis. Los datos de esfuerzo de la flota de palangre pelágico estadounidense junto un modelo de idoneidad de volumen del hábitat ponderado para la aguja azul (*Makaira nigricans*) se utilizaron para derivar una serie temporal simulada de capturas de aguja azul de 1986 a 2015 con tres estructuras de población subyacente. La serie temporal resultante se facilitó a los científicos de evaluación de stock para que determinasen si las tendencias de población subyacentes podrían detectarse con métodos diferentes de estandarización de CPUE que incorporan datos medioambientales.*

KEYWORDS

Catch/effort, Longline, GLM, Simulation, Stock assessment, Environmental effects

1. Introduction

It is now a generally accepted fact that variation in the planet's climate and its effects on the world's oceans is increasing. For marine fish, specifically those of the highly migratory nature, this increased climate variation has led to changes in distribution, productivity, migratory patterns, and susceptibility to various fishing gears. Many of these changes have created situations where there is no historic analog to the current climatic and biological conditions. These changes become especially problematic when they manifest themselves through the fishery dependent indices of abundance (such as catch-per-unit-effort, or CPUE) used to assess the status of the stocks, such as is done routinely by the International Commission for the Conservation of Atlantic Tuna (ICCAT). The inclusion of oceanographic data, i.e. sea surface temperature and dissolved oxygen, into standardization methods for CPUE has been proposed as way to help understand and account for this variability.

Previously, the ICCAT Working Group on Methods recommended the use of simulated data sets with known values of underlying population trends to test the robustness of CPUE standardization methods (Anon. 2016; 2004). A longline CPUE data simulator (LLSIM) was developed to meet this requirement to simulate known data for testing a variety of hypotheses (Goodyear, 2016). The simulator uses data describing the physical environment within the model region to predict fish abundances using habitat suitability modeling (HSM). This approach is commonly used for predicting habitat quality from habitat suitability indices (HSI) 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 representative fishing effort to sample the fish populating the HSM throughout the year, producing simulated catch per unit effort data.

This paper describes the application of a realistic fishing fleet, one based off the US longline fishery, to several simulated blue marlin populations. The resulting simulated catch datasets were provided to three stock assessment scientists with expertise in standardizing CPUE indices. The goals of this work are to determine how well current standardization methods capture population trends and if the inclusion of environmental data aids in the standardization process.

2. Material and methods

2.1 Simulation model

A longline CPUE data simulator (LLSIM) was developed to meet the earlier requirement to simulate known data for testing a variety of hypotheses (Goodyear 2006). The core element of the simulation is the catch on a single hook of a longline set. Each hook has a depth probability distribution and attributes of latitude, longitude, year, month, time of day, and position along the longline. Each of these attributes is associated with the individual 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 35S to 55N latitude and 95W to 20E longitude, exclusive of major land masses. This area is broken down into 7067 cells; each cell is 1 degree of latitude by 1 degree of longitude. All spatial distributions of input and output variables reference an array of cell identities. Each longitude-latitude cell is also divided into 46 depth strata of unequal bins, corresponding to the environmental depth data.

2.2 Environmental Data

Application of the HSM approach to predict the spatial distribution of a species requires quantitative data about the physical environmental variables that are important determinants of its habitat. Temperature and dissolved oxygen concentrations are major factors shaping the pelagic marine environment. Temperature is perhaps the major feature of the pelagic ocean and is the environmental variable most frequently employed in habitat standardizations. Dissolved oxygen is an important variable to include because at low levels it becomes a critical factor limiting habitat suitability. Environmental data were obtained though the Community Earth System Model (CESM1), which is a global ocean-sea-ice coupled model coupled to a biogeochemistry model as known BEC (Biogeochemical Elemental Cycle). The model covers the global ocean with a latitudinal and longitudinal resolution of 1.0° and 60 vertical layers with the bottom level at 5500 m.

2.3 Species distribution

Species data required for simulations are defined in two steps. The first defines the average population number alive during the year and month by species (and sex-age grouping if considered). The second step defines the relative densities of the population by latitude, longitude, year, month and depth (these densities are computed so that the sum of the products of the relative density x volume for each latitude, longitude, and depth = 1.0). The products of the two vectors give the actual densities relative to each hook for the simulation.

Three population trends were used in this study, a constant population of 500,000 individual blue marlin, a decreasing population, and an increasing population (**Figure 1**).

2.4 Longline sets

The effort data used in the LLSIM was based off the US longline pelagic fishery logbooks. While the US commercial fishery has been operating since the 1960s, logbook data is only available from 1986 onwards. The pelagic longline fleet targets swordfish, yellowfin tuna and bigeye depending on the gear configuration. To account for the differing gear configurations, the logbook data was sorted by hook type, numbers of light sticks deployed, bait type and the numbers of hooks between floats. The resulting gear combinations yielded 128 discrete gear types.

To ensure data confidentially, sets were removed when less than three vessels fished in the same region. Set locations were transformed into the cell ID used within the LLSIM grid. From there, set locations were randomly jittered by adding or subtracting 1-5 from the cell ID value. Sets were discarded that occurred on land after the jittering process. A large sample of longline sets was creating using bootstrapping and 737,000 sets were randomly selected, replicating the original amount of usable longline sets.

2.5 Data Analysis

The longline simulator outputs a catch by set file with column headings typically observed in logbook data. For this exercise, the variables included with the number of blue marlin caught were: total number of hooks, hook type, bait type, light sticks, month, year and latitude and longitude. The light sticks were binned values corresponding to no light sticks reported, zero light sticks deployed, 1-500 and 501-1500 light sticks. The sea surface temperature and the dissolved oxygen at the surface for the location, month and year from 1986-2012 were also supplied from the outputs of the CESM.

Three simulated catch datasets corresponded to the three population trends were distributed to three analysts with extensive knowledge and experience developing standardized indices of abundance from CPUE data. The analysts developed their own unique approach to the data without consultation with the authors. The details of each analyst's approach are included below. The populations 2 and 3 are not included in this analysis. These populations have been archived however the naming sequence has been retained for continuity in the study.

2.5.1 Analyst 1

Analyst 1 used a delta lognormal approach **(Table 1)** to standardize CPUE, using the Lo *et al.* method to calculate the standard errors. 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.

The CPUE of blue marlin was calculated as catch per thousand hooks. The 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 **Figure 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). Final model structure for all models listed in **Table 2.**

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 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 observations. The final dataset included 96.5% of the total observations.

2.5.2 Analyst 2

Analyst 2 used a negative binomial GLMM to standardize the CPUE, which was calculated as catch per thousand hooks. The latitude and longitude were grouped into four areas (SE, NE, SW, NW) and months were grouped into quarters. This analyst used four different models: a full model which 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 correlated to sea surface temperature. SST was treated at a continuous variable. The final two models were termed Simple **(Table 1)** and just 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.

The models were run consecutively in R using MASS, nlme and lme4 packages. Interaction effects were not used, just main effects which were fixed. Deviance was used as the main model selection criteria along with ANOVA and F tests. Final model structure for all models listed in
Table 3.

2.5.3 Analyst 3

Prior to developing a model to standardize the CPUE trend, this analyst ran an exploratory script and observed a gear effect immediately. The areas were split into the spatial zones used by the Southeast Fishery Science Center. From there, a delta gamma model was selected with all variables as factors for all populations. The model structure was the same for the model that contained environmental data, which were treated as continuous variables. 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. The final models were selected using AIC and deviance explained. The models were run in R using the packages lsmeans and glmmADMB.

2.5.4 Analysis of final trends

Standardized trends from the three analysts and the true population trends were normalized to the mean to examine differences in the time series. The normalized, modeled CPUE trends were compared to the normalized, true CPUE values using a linear regression and root-mean-square-errors (RMSE) estimated using the values predicted from the regression.

3. Results

Population 1

Analyst 1 noted a clear, declining trend in CPUE with respect to hooks between floats and used hook between floats as a numerical variable for population 1. Due to the large sample size, AIC and p values did not eliminate any variables from the models. The SST factor explained a large fraction of the deviance in probability of presence (binomial), but had no influence on the CPUE when caught (lognormal). Hooks between floats was also important for probability of presence, but less important for positive CPUE.

The trends obtained by analyst 1 exhibited a drop in the population in 2002 that did not occur in the true population trend for both models (**Figure 3)**, however the model including sea surface temperature and dissolved oxygen had a lower RMSE value than the model that did not include the environmental variables (**Table 5** and **Figure 6**). Analyst 1 did not use hook type in the models as hook type did not explain more than 5% of the deviance observed.

Analyst 2 used the environmental data in a simple model with only year, quarter and area as factors and a full model with all possible variables. The simple model with environmental data had the drop observed in 2003 the same as analyst 1, however the addition of the environmental data smoothed the trend out (**Figure 3)**. Both versions of the full model had a very close agreement to the true population trend time series. The full model with the environmental variables had the lowest RMSE, followed by the simple model with the environmental variables (**Table 5** and **Figure 6**).

Both time series obtained by analyst 3 fluctuated around the true population trend (**Figure 3**) and had a RMSE value lower than both analyst 1's models and lower than analyst 2's simple model with no environmental variables (**Table 5** and **Figure 6**). Unlike the other analysts, the RMSE value for the model containing the environmental variables was higher than that of the model lacking the environment.

Population 4

Analyst 1 treated hooks between floats as a factor for population 4 as the relationship between HBF and CPUE was less clear to the analyst.

The time series obtained from all three analysts not fit the true population trend in the earliest years (1986-1993), which corresponded to the highest CPUE values (**Figure 4**). In later years, the modeled trends followed the true population trend, however in the most recent years, analyst 1 and 2 underestimated the true population. The model with the environmental data had a better fit than the one without for analyst 1. Analyst 2's time series all converged along the same trend in 2001 while the simple model with year, area, quarter and the environmental variables had the best fit with respect to the regression and RMSE values (**Table 5** and **Figure 7)**.

The two models of analyst 3 had similar patterns in the standardized trends as the other 2 analysts, however these models did not have the same overestimation the true population values at the start of the time series. The models were both able to converge on the true population completely after 2011. The RMSE values were both lower than any of the fits obtained by analysts 1 and 2. The model containing the environmental variables had a higher RMSE value.

Population 5

Population 5 had the largest discrepancy between modeled values to the true population values (**Figure 5**). Analyst 1 and 2's models in the earliest years underestimated the abundance while in the later years, the models overestimated the values. Analyst 1's model with the environmental data was closer to the true trend and this was reflected in the lower RMSE value. Analyst 2's four models all performed similarly and converged beginning in 2001. The full model with sea surface temperature had the best performance with respect to the RMSE value (**Table 5 and Figure 8**). Analyst 3 had the opposite pattern than analyst 1 and 2 with respect to model fits. Both model values in the beginning of the time series matched quite well to the true population values but in later years, discrepancies between the true trend and the models occurred. As was the case with this analyst's models of the population 1 and 4, the model containing environmental variables had a higher RMSE.

Discussion

While the analysts' approach to the data and the modeling structure differed, the underlying population trends were captured, some more successfully than others. However, differences in approaches highlight the importance of how variables are grouped and the criteria for including variables and factors. All the analysts used unique area combinations for the spatial structure of their models. Interestingly the models with the simplest combination of four quadrants had the lowest error across all populations (analyst 2).

Analyst 1 noted that hook types did not explain a large amount of deviance (<5%), leading that analyst to exclude it from the final model structure. Hook type was also excluded from the simple model of analyst 2, leading to similar underestimation of population 1 trend. The switch from an overestimation of the population trend to an underestimation in those 2 models coincided with a switch in the gear configuration of the fleet from J hooks to circle hooks. This underestimation of the population may reflect a lower catchability however this is difficult to determine from this exercise.

The inclusion of sea surface temperature into the simple model of analyst 2 corrected the underestimation trend in population 1 whereas the inclusion of the environmental variables of analyst 1 did not correct the modeled trend. It is difficult then to provide clear-cut recommendations for the addition of environmental data or all available variables into the final model structure. The choice, and method, of final selection criteria for variables also varied across analysts. Deviance explained is often an accepted level for variable inclusion (Ortiz and Arocha, 2004), however, as evidenced by this exercise, cannot be used in isolation.

Analyst 1 was the only to use interaction effects and these were set to random effects. As the catch data was patchy across space and time, there were several combinations that did not contain data, necessitating the use of random treatment effects rather than fixed. R cannot converge when large amounts of missing interaction terms are present, however this is not the case with other statistical language software such as SAS. Future studies will include the use of SAS and whether the inclusion of fixed interaction effects improve the estimation of population trends. In terms of the underlying model used to develop the catch datasets, the random effects are gear catchability and habitat suitability for blue marlin. The environmental variables DO and temperature can be viewed as random but were treated as fixed effects by all the analysts.

The addition of environmental variables did improve the estimates of the true population trends across populations for some analysts and models. The inclusion of these variables in the cases of analyst 1 for population 4 and all the populations for analyst 3 resulted in a higher RMSE values and did not follow the true population values as well as the models that did not contain the environmental variables. The use of environmental variables is thought to be a good predictor of density of a species in the vicinity of the set and/or hook. However, suitable transformation of the data may be necessary. The values at the surface can only be correlated with the values at depth that influence species distributions and subsequently habitat suitability. Future studies will take advantage of the CESM data outputs at the actual depths where both the blue marlin and hooks occupy.

The use of environmental variables did increase those model's CVs as compared to the models run without the environmental variables. This data is not presented here as it was not obtained from all analysts however there does not appear to be a relationship between how well a model fit the true population values and the CV values of the model. This is a relationship that will be further explored in future studies.

The models applied to the increasing population, population 5, did not estimate the true population trend as accurately as populations 1 (flat) and 4 (decreasing). This may be explained by the behavior in the gear and the fish species. The gears in the US longline fleet have been fishing deeper and deeper as their techniques to find targeted fish have been refined. Blue marlin are not the target of this fishery and as such, deeper gear would result in a declining CPUE for this species. Population 5 models an increasing population but this may be masked or confounded by the decrease in catchability.

Conclusions/future work

Work with the simulated datasets is still in the early stages and there remains much to be explored. Further applications are explained in depth in Goodyear (in press) presented at WGSAM 2017.

This in-house study lead the authors to conclude two important points. First, analysts need to be given very specific instructions in order that the results between them can be considered relative to each other. An example of this was that only some of the analysts constructed GLMs both with and without consideration of the environmental data. More insights could likely have been gleaned had it been made clear that each of the analysts should have taken this approach so that comparison of GLMs with and without the use of the environmental data could be compared. Second, while the blind study was extremely informative to address the question "are we currently doing it correctly?" it lacked a systematic approach that could better lead to conclusions regarding best practices. This is likely better accomplished by the authors conducting the GLM analysis themselves. Future experiments will use both methods to address these two different objectives.

Acknowledgements

C.P. Goodyear's contribution to this research was supported by The Billfish Foundation. F. C. Forrestal's contribution was supported the NOAA Southeast Fisheries Science Center via the University of Miami Cooperative Institute of Marine Science. 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 for gathering and compiling the US longline effort data.

References

Anon. 2004. Report of the 2003 meeting of the ICCAT Working Group on Assessment Methods. Collective Volume Scientific Papers, ICCAT, 56(1): 75-105.

Anon. 2016. Report of the 2015 meeting of the ICCAT Working Group on Assessment Methods. Collective Volume Scientific Papers, ICCAT, 71(6): 2742-2780.

Chang, Y. J., Sun, C. L., Chen, Y., Yeh, S. Z., and DiNardo, G. 2012. Habitat suitability analysis and identification of potential fishing grounds for swordfish, *Xiphias gladius*, in the South Atlantic Ocean. International Journal of Remote Sensing 33(23) 7523-7541.

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

Dunne, J.P., John, J.G., Shevliakova, E., Stouffer, R.J., Krasting, J.P., Malyshev, S.L., Milly, P.C.D., Sentman, L.T., Adcroft, A.J., Cooke, W. and Dunne, K.A., 2013. GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part II: Carbon System Formulation and Baseline Simulation Characteristics. *Journal of Climate*, *26*(7), 2247-2267.

Gent, P.R., Danabasoglu, G., Donner, L.J., Holland, M.M., Hunke, E.C., Jayne, S.R., Lawrence, D.M., Neale, R.B., Rasch, P.J., Vertenstein, M., Worley, P.H., Yang, Z-L., and Zhang, M. 2011: The Community Climate System Model Version 4. J. Climate, 24, 4973–4991.

Goodyear, C. P. 2006. Performance diagnostics for the longline CPUE simulator. Collective Volume Scientific Papers, ICCAT, 59(2): 615-626.

Goodyear, C. P., and Bigelow, K. A. 2010. Preliminary explorations of CPUE standardization of the U.S. longline observer sailfish data using stathbs. Collective Volume Scientific Papers, ICCAT, 65(5): 1713-1724.

Goodyear, C. P., and Bigelow, K. A. 2012. Preliminary analyses of simulated longline Atlantic blue marlin CPUE with HBS and generalized linear models. Collective Volume Scientific Papers, ICCAT, 68(4): 1510-1523.

Goodyear, C. P. 2016. Habitat volume: modeling the time-varying density distribution of Atlantic blue marlin in the context of stock assessment. (Unpublished manuscript)

Goodyear C.P., Schirripa M., and Forrestal F., (in press) Longline data simulation: a paradigm for improving CPUE standardization. Document SCRS/2017/065 (in press). 12p.

Hirzel, A. H., and Lay, G. L. 2008. Habitat suitability modelling and niche theory. Journal of Applied Ecology 45:1372–138.

Long, M.C., Lindsay, K., Peacock, S., Moore, J.K. and Doney, S.C., 2013. Twentieth-century oceanic carbon uptake and storage in CESM1 (BGC)*. Journal of Climate, *26*(18), pp.6775-6800.

Robinson, L. M., Hobday, A. J., Possingham, H. P., and Richardson, A. J. 2015. Trailing edges projected to move faster than leading edges for large pelagic fish habitats under climate change, Deep Sea Research Part II: Topical Studies in Oceanography 113: 225-234.

Schirripa, M. J., and Goodyear, C. P. 2010. CPUE standardizations of species targeting in longline fisheries: initial investigations of the utility of simulated data for method verification. Collective Volume Scientific Papers, ICCAT, 65(6): 2351-236.

Table 1. Model types used by each analyst for all populations.

	Model 1	**Model 2**	**Model 3**	**Model 4**
Analyst 1	Delta lognormal	Delta + SST+ DO		
Analyst 2	Simple (Neg. Bin.)	Full (Neg. Bin.)	Simple (NB)+SST	Full (NB)+SST
Analyst 3	Delta gamma	Delta gamma + SST +DO		

Table 2. Final model strucutre from analyst 1 for all populations.

Population	**Environment**	**Binomial**	**Lognormal**
1	no	year+hbf+area+season-1 +**year:area**+**area:season**	year+hbf+area
	yes	year+hbf+area+season+SSTfac	year+hbf+area
2	no	year+hbf+area+season-1 +**year:hbf**	year+area
	yes	year+hbf+area+season+SSTfac	year+area
3	no	year+hbf	year
	yes	year+hbf+SSTfac	year

Table 3. Final model structure from analyst 2. Model structure constant across population.

Simple	factor(year)+factor(Qtr)+factor(AreaFINAL)+offset(ln_hooks) factor(year)+factor(Qtr)+factor(AreaFINAL)+SST +offset(ln_hooks)
Full	factor(year)+factor(Qtr)+factor(AreaFINAL)+factor(gear)+factor(light)+ factor(hbf)+factor(hook)+factor(bait)+SST+offset(ln_hooks))

Table 4. Final model structure from analyst 3. The same model structure was used for all populations.

Model	**Binomial**	**Gamma**
1	factor(year)+factor(hbf)+factor(area)+factor(month)+factor(hook)+factor(bait)+factor(light)	same
2	factor(year)+factor(hbf)+factor(area)+factor(month)+factor(hook)+factor(bait)+factor(light) +SST+DO	same

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

Population 1 MSE

	Model 1	*Model 2*	*Model 3*	*Model 4*
Analyst 1	0.2880	0.2521		
Analyst 2	0.2220	0.0162	0.0160	0.0167
Analyst 3	0.0852	0.1009		

Population 4 MSE

	Model 1	*Model 2*	*Model 3*	*Model 4*
Analyst 1	0.1935	0.2710		
Analyst 2	0.3297	0.3489	0.3130	0.2954
Analyst 3	0.1014	0.1288		

Population 5 MSE

	Model 1	*Model 2*	*Model 3*	*Model 4*
Analyst 1	0.3268	0.2741		
Analyst 2	0.4224	0.4202	0.4309	0.4029
Analyst 3	0.1046	0.1456		

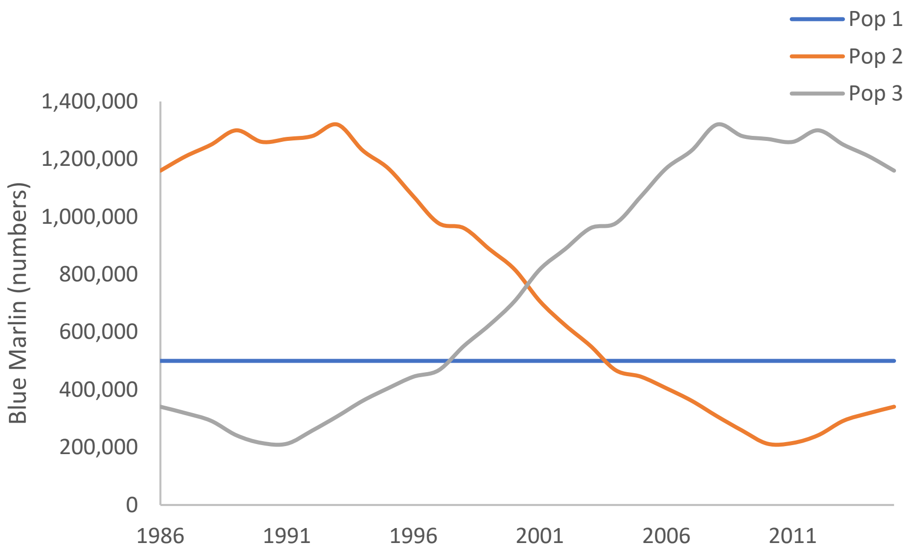

Figure 1. True population trends used in the simulated catch data.

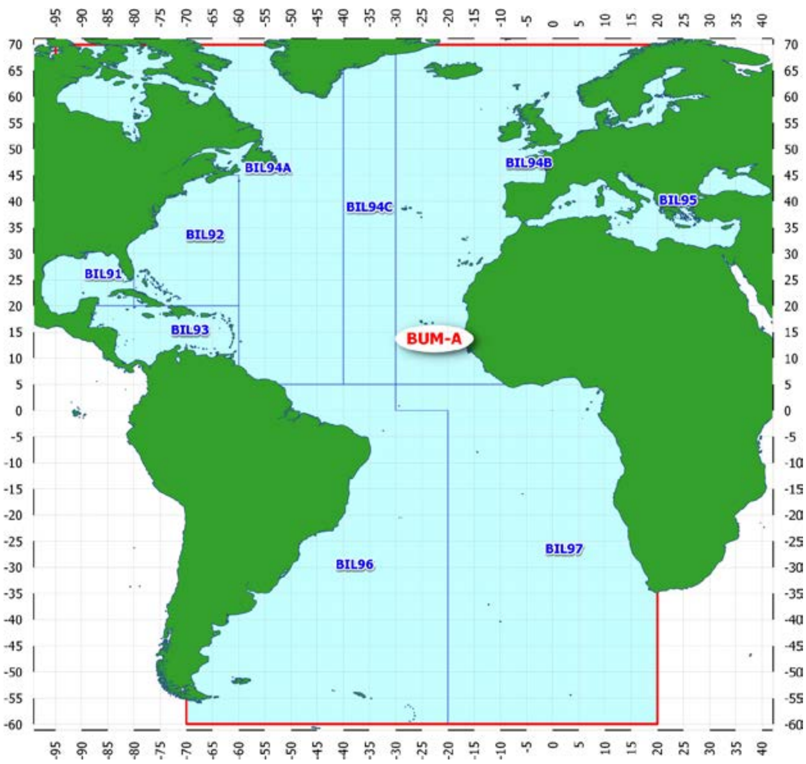

Figure 2. Areas used in standardization, from ICCAT manual.

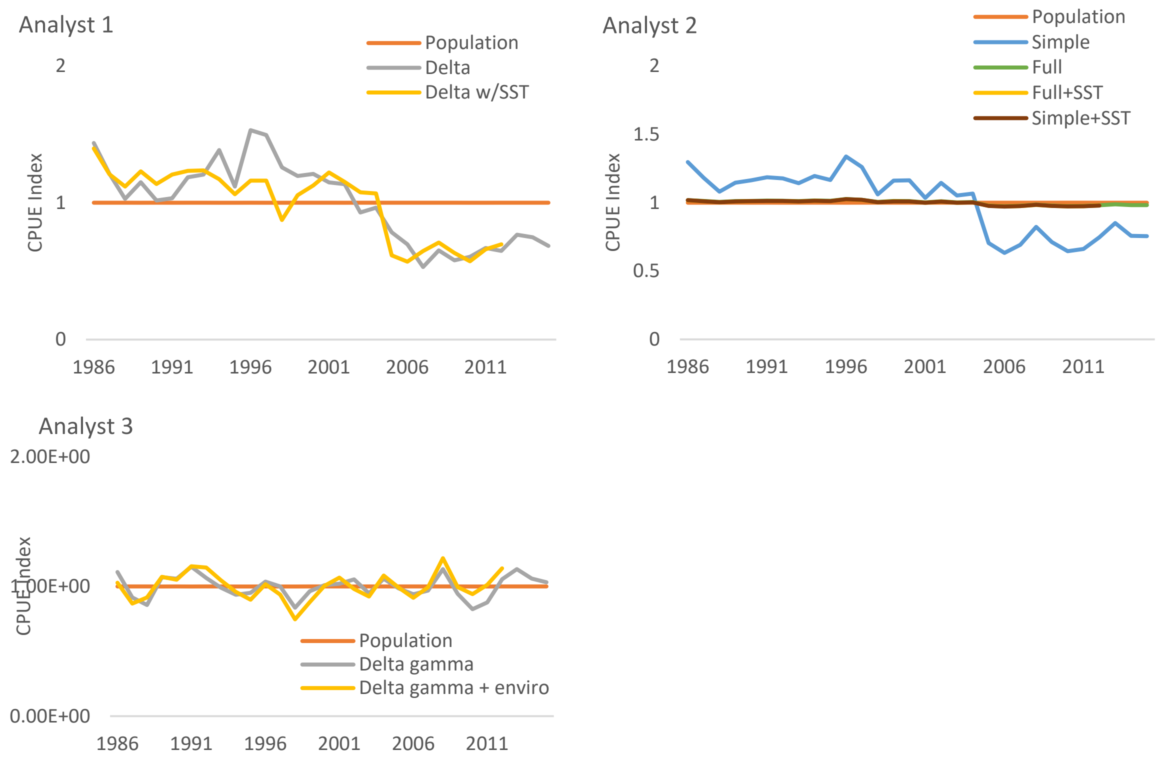

Figure 3. Standardized CPUE indices for population 1.

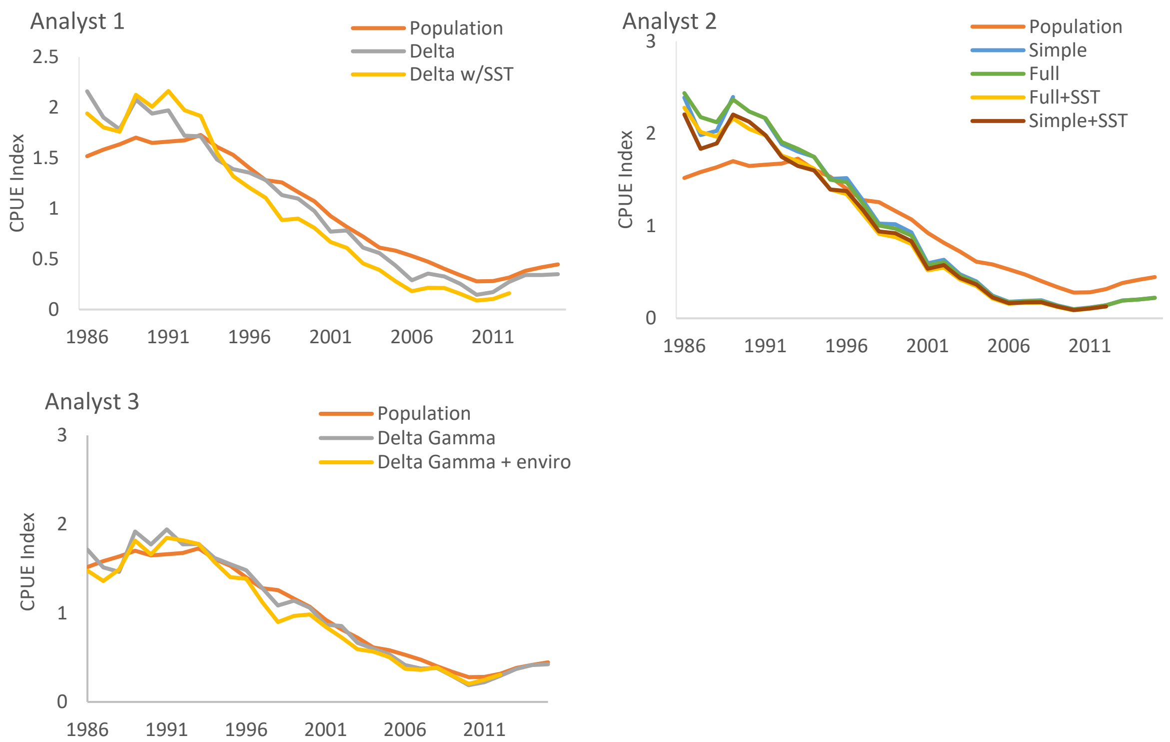

Figure 4. Standardized CPUE indices for population 4.

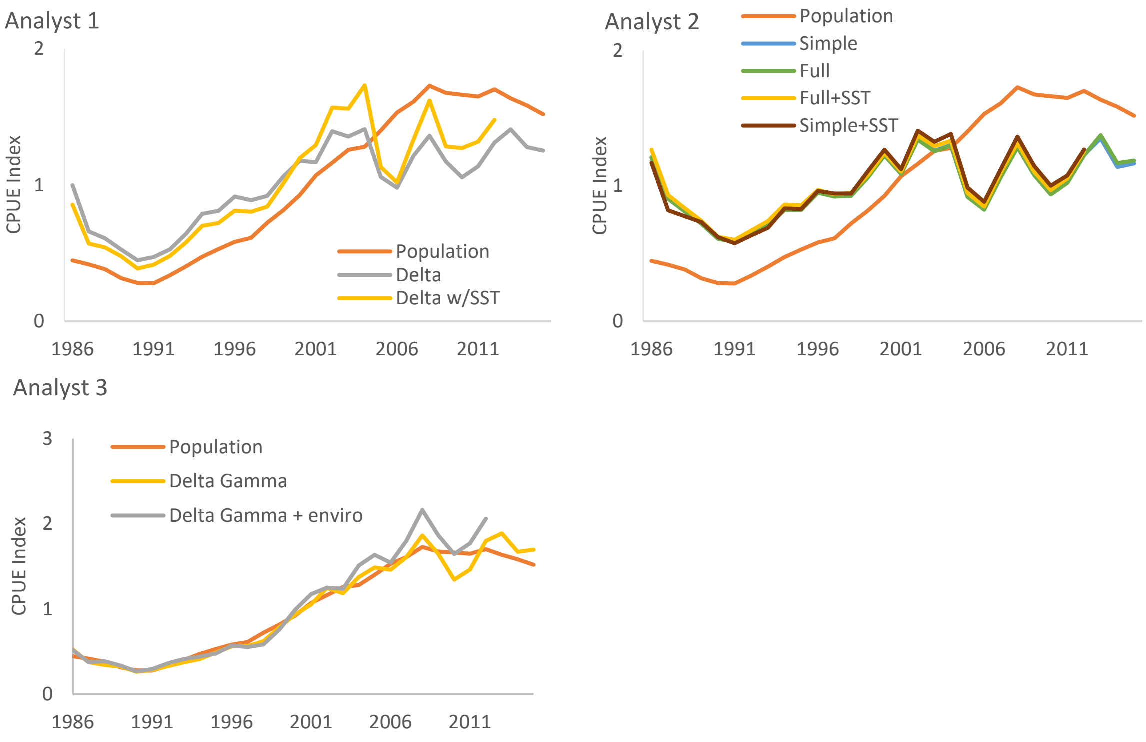

Figure 5. Standardized CPUE indices for population 5.

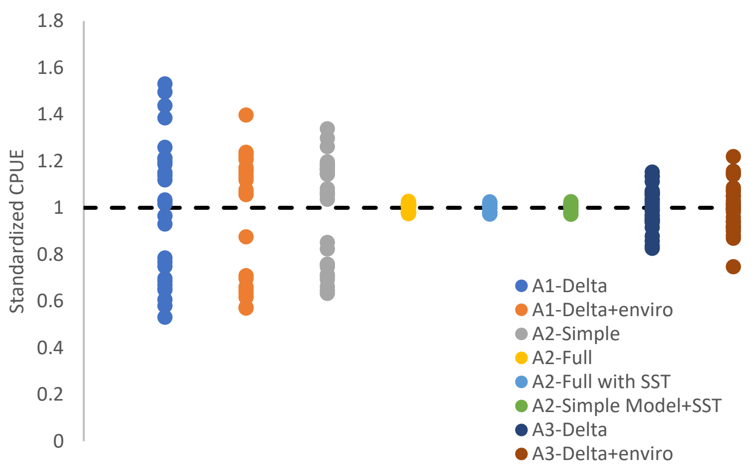

Figure 6. Standardized CPUE values for all the analysts and their methods for population 1. Line represents the true population trend, the dots are deviation from the true population trend.

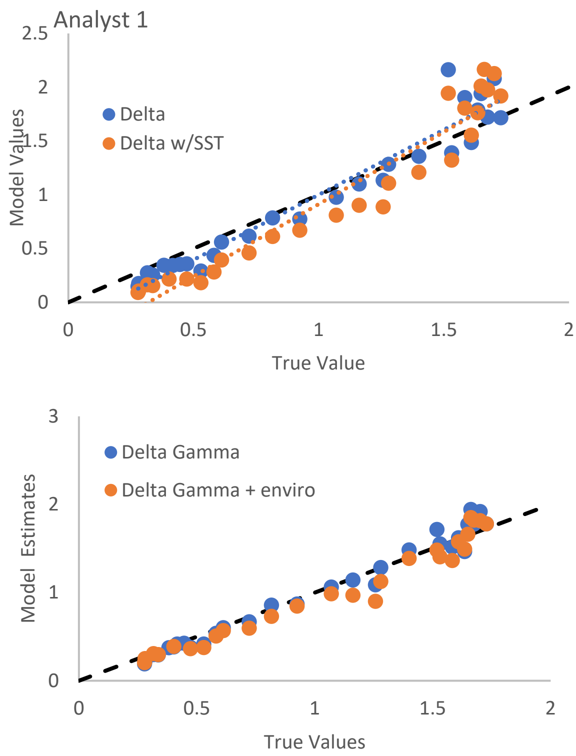

Figure 7. Linear regression comparing true population trend and modeled population values for population 4.

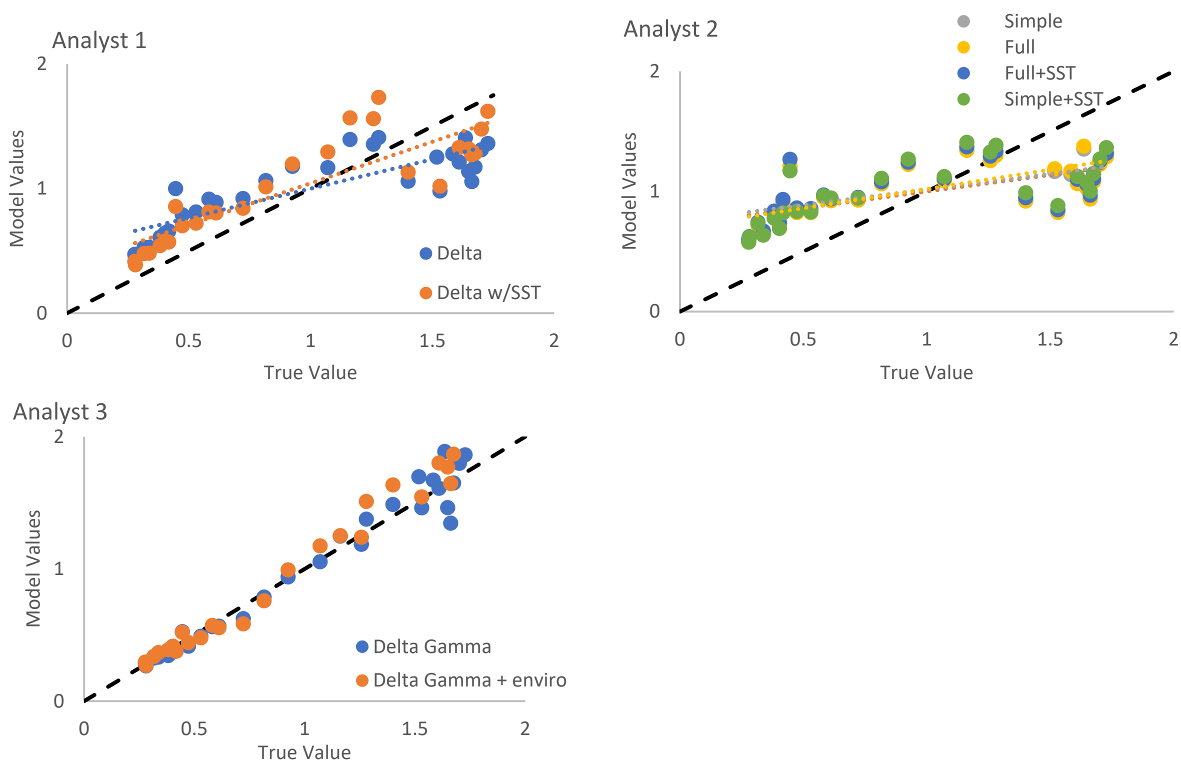

Figure 8. Linear regression comparing true population trend and modeled population values for population 5.