SCRS/2018/016 Collect. Vol. Sci. Pap. ICCAT, 75(5): 912-915 (2018)

COMPARISON OF LOGBOOK DATA TO OBSERVER DATA USING A LONGLINE SIMULATOR WITH BLUE MARLIN AS AN EXAMPLE

F. Forrestal,[1] M. Schirripa,[2] C.P. Goodyear[3]

SUMMARY

Logbook and observer data based off the US pelagic longline fishery were used to in a longline simulator (LLSIM) to simulate catch datasets from a known population of blue marlin. The blue marlin population was simulated from a species distribution model that was built using PSAT data and environmental data obtained from the Community Earth Systems Model (CESM). The catch datasets derived from LLSIM were standardized using a delta lognormal approach with and without sea surface temperature. Bait type was omitted from the model using observer data but all over factors were retained in both the logbook and observer data. The logbook data had a better model fit to the true population (RMSE=0.13) than the observer data (RMSE=0.28). This pattern was repeated with the models containing SST, however these models had a worse fit to the true population (RMSE = 0.16 and RMSE = 0.37. Future studies will utilize model selection to optimize the model fits to the true population.

RÉSUMÉ

Les données des carnets de pêche et des observateurs de la pêcherie palangrière pélagique des Etats-Unis ont été utilisées dans un simulateur palangrier (LLSIM) pour simuler des jeux de données de capture d'une population connue de makaire bleu. La population de makaire bleu a été simulée à partir d'un modèle de distribution des espèces qui a été construit à l'aide de données PSAT et de données environnementales obtenues à partir du Community Earth Systems Model (CESM). Les jeux de données de capture obtenus de LLSIM ont été standardisés en utilisant une approche delta log-normale avec et sans température de surface de la mer. Le type d'appât a été omis du modèle qui utilisait les données d'observateurs, mais tous les autres facteurs ont été conservés dans les données du carnet de pêche et de l'observateur. Les données du carnet de pêche avaient un meilleur ajustement du modèle à la population réelle (RMSE = 0,13) que les données de l'observateur (RMSE = 0,28). Ce schéma a été répété avec les modèles contenant SST, mais ces modèles ont eu un ajustement plus mauvais à la population réelle (RMSE = 0,16 et RMSE = 0,37. Les études futures utiliseront la sélection du modèle pour optimiser les ajustements du modèle à la population réelle.

RESUMEN

Se utilizaron los datos de los cuadernos de pesca y de los observadores de la pesquería de palangre pelágico de Estados Unidos en un simulador de palangre (LLSIM) para simular conjuntos de datos de captura de una población conocida de aguja azul. La población de agua azul fue simulada a partir de un modelo de distribución de especies que fue construido utilizando datos de PSAT y datos medioambientales obtenidos a partir del Community Earth Systems Model (CESM). Los conjuntos de datos de captura derivados del LLSIM fueron estandarizados utilizando un enfoque delta lognormal con y sin temperatura de la superficie del mar. El tipo de cebo se omitió en el modelo que utiliza datos de observadores, pero todos los demás factores se mantuvieron tanto en los datos de los cuadernos de pesca como en los de observadores. Los datos de los cuadernos de pesca lograban un mejor ajuste del modelo a la población verdadera (RMSE=0,13) que los datos de observadores (RMSE=0,28). Este patrón se repitió con los modelos que contenían SST, sin embargo, estos modelos tenían peor ajuste a la población verdadera (RMSE = 0,16 y RMSE = 0,37). En estudios futuros se utilizará la selección del modelo para optimizar los ajustes del modelo a la población verdadera.

KEYWORDS
Blue Marlin, Longline, Stock assessment,
CPUE Standardization, GLM, Data simulation

[1] CIMAS/RSMAS University of Miami, 4600 Rickenbacker Cwy. Miami, FL 33149. fforrestal@rsmas.miami.edu

[2] NOAA Fisheries, Southeast Fisheries Center, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, FL, 33149-1099, USA. Michael.Schirripa@noaa.gov

[3] 1214 North Lakeshore Drive, Niceville, Florida 32578, USA. phil_goodyear@msn.com

1. Introduction

Observer data can provide high quality information on how a fishery operates and its catch, either targeted or incidental. However, due to logistical constraints, it is not possible to have 100% observer coverage of a given fleet. Logbooks provide information on a fishery for every trip and set but often do not contain all the information that is recorded by observers. The coverage level of observers within a fishery has been recommended to be set at 20% to provide an informative level of detail, particularly the species of interest is caught as bycatch in the fishery, such as blue marlin in the US longline fishery. This low level of coverage in comparison to the true extent of a fishery may not be able to provide an accurate index of abundance using catch per unit effort for these non-target species Torres-Irineo *et al.*, 2014). To test this, two catch datasets were simulated by the longline simulator (LLSIM), one derived from the logbook data and other from the observer data.

2. Methods

2.1 Blue marlin species distribution model

The species distribution model (SDM) in this study is a detailed model of the four-dimensional distribution of blue marlin (Goodyear 2016). Oceanographic data and species habitat preferences are used to distribute the population in time and space. The current implementation partitions the Atlantic from 35 S to 55 N latitude at a spatial resolution of 1º latitude and 1º longitude with 46 depth layers. The oceanographic data were monthly values from the Earth System Model from 1956 to 2012 which matched the spatial resolution of the SDM. The oceanographic data were provided by colleagues at the US National Atlantic Oceanographic and Meteorological Laboratory (AOML). A constant population of blue marlin was used with the species distribution model as the known, underlying population. Blue marlin abundance was set at 500,000 individual fish/year from 1986-2015.

2.2 Logbook and observer data

US pelagic longline observer and logbook data were formatted to be used in the longline simulator (Forrestal et al, 2017). While the data came from the same fleet, they were treated as separate fleets in the simulator to derive two separate catch datasets. Longline gear types were characterized using number of lightsticks deployed (0, 1-500, >501, and unknown), bait type (dead, live, artificial, and unknown), hook type (circle hook, J hook, mixed and unknown) and hooks between floats (HBFL 2-6). Gear types were the same between the two sources of effort. The observer database time series runs from 1992 – 2015 while the logbook data ranges from 1986 – 2015. Both the observer and logbook data were bootstrapped and jittered before being entered into the longline simulator to mask the true fishing locations and gear type to ensure data confidentiality.

2.3 CPUE Standardization

A delta lognormal model was used to standardize the logbook and observer catch datasets. For comparison purposes, the same model was used for both datasets.

Binomial Model

Pos/total=factor(year)+factor(hbf)+factor(area)+factor(month)+factor(hook)+factor(bait)+factor(light)

Abundance Model

lnCPUE=factor(year)+factor(hbf)+factor(area)+factor(month)+factor(hook)+factor(bait)+factor(light)

A second delta lognormal model was applied to both datasets that included sea surface temperature and dissolved oxygen at the location and month of the set. Sea surface temperature was treated as a factor, at 2-degree Celsius intervals. Dissolved oxygen was treated as a continuous variable.

Binomial Model

Pos/total=factor(year)+factor(hbf)+factor(area)+factor(month)+factor(hook)+factor(bait)+factor(light) +factor(SST) + DO

Abundance Model

$$lnCPUE=factor(year)+factor(hbf)+factor(area)+factor(month)+factor(hook)+factor(bait)+factor(light)+factor(SST) + DO$$

The root mean square error (RMSE) was calculated to determine the fit of the model to the true population for both the observer and logbook data.

3. Results and Discussion

The catch datasets derived from the observer data contained roughly 11 million hooks while the logbook data contained 186 million hooks, the observer data contained 6% of the effort recorded in the logbook data. Bait type was removed from the final model selection for the observer data as only one bait type was recorded in cleaned and bootstrapped data.

The standardized model using the logbook data had a better fit to the true population as compared to the observer data (RMSE = 0.13 and RMSE = 0.28). The models containing the environmental data had a similar pattern in that the logbook data had a better fit to the true population as compared to the observer data (RMSE = 0.16 and RMSE = 0.37). These values were higher, signifying a poorer fit, than the models that did not contain the environmental variables. All four models were able to capture the true population values overall (**Figure 1**).

While the data derived from the logbook data had a superior fit to the true population, the observer data was able to capture the true population trend overall despite only accounting for 6% of the total effort in the simulated fishery. The addition of habitat suitability variables may improve the fits of both models and will be examined in the future.

References

Forrestal, F.C., Goodyear, C.P., Schirripa, M., Babcock, E., Lauretta, M., and Sharma, R. 2017. Testing robustness of CPUE standardization using simulated data: findings of initial blind trials. Collect. Vol. Sci. Pap. ICCAT, 74(2): 391-403.

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

Goodyear, C.P. 2017. Simulating longline catch with LLSIM: a user's guide (version 2). pp 1-23.

Goodyear, C.P., M. Schrippa, and F. Forrestal. 2017. Longline data simulation: A paradigm for improving CPUE standardization. Collect. Vol. Sci. Pap. ICCAT, 74(2): 379-390.

Torres-Irineo, E., Amande, J. M., Gaertner, D., Delgado de Molina, A., Murua, H., Chavance, P., Lezam-Ochoa, N. (2014). Bycatch species composition over time by tuna purse-seine fishery in the eastern tropical Atlantic Ocean. *Biodiversity and Conservation*, *23*, 1157–1173. https://doi.org/10.1007/s10531-014-0655-0

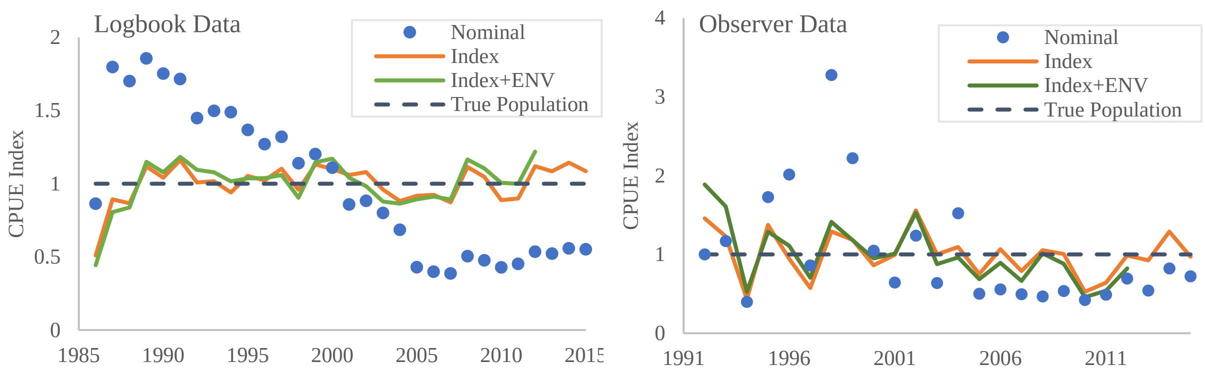

Figure 1. True population trends with standardized index and nominal CPUE for logbook data and observer data.