SCRS/2017/065 Collect. Vol. Sci. Pap. ICCAT, 74(2): 379-390 (2017)

LONGLINE DATA SIMULATION: A PARADIGM FOR IMPROVING CPUE STANDARDIZATION

C. Phillip Goodyear[1] Michael Schirripa[2] and Francesca Forrestal[3]

SUMMARY

The computer program LLSIM was developed to simulate longline catch data to quantify how well alternative methods estimate abundance from longline CPUE by comparing results to known true values from the simulation. It employs a Monte-Carlo algorithm with a probability of capture computed for each hook on each simulated set. The probability value is computed from the population size using time-varying three dimensional distributions of hooks and species relative densities. The species distributions, locations, and depth distributions of hooks are modeled separately and integrated by the simulator. The method decomposes catchability at the hook into an essential gear coefficient (k) and a habitat effect (H) that varies in time and 3-D space. The approach provides a basis for estimating a habitat coefficient (w) that quantifies the effect of habitat quality on catchability on each set with potential as a covariate for CPUE standardizations. Simulations can be used to compare alternative approaches to data imbalance and other routine complications of standardization modeling with experimentally designed test data. Simulations can also be designed to test performance of alternatives for individual assessments.

RÉSUMÉ

Le programme informatique LLSIM a été mis au point pour simuler des données de capture palangrière dans le but de quantifier la mesure dans laquelle les méthodes alternatives estiment l'abondance à partir de la CPUE palangrière en comparant les résultats de la simulation obtenus avec des valeurs réelles connues. Celui-ci emploie un algorithme de Monte Carlo avec une probabilité de capture calculée pour chaque hameçon de chaque opération simulée. La valeur de probabilité est calculée à partir de la taille de la population au moyen de distributions d'hameçons en trois dimensions variant dans le temps et de densités relatives à l'espèce. Les distributions des espèces, les emplacements et les distributions des profondeurs des hameçons sont modélisés séparément et intégrés dans le simulateur. La méthode décompose la capturabilité par hameçon en un coefficient d'engin essentiel (k) et un effet de l'habitat (h) qui varie dans le temps et dans un espace en trois dimensions. L'approche fournit une base pour estimer un coefficient de l'habitat (w) qui quantifie l'effet de la qualité de l'habitat sur la capturabilité de chaque opération, pouvant éventuellement servir de covariable pour les standardisations de la CPUE. Les simulations peuvent être utilisées pour comparer des méthodes alternatives de déséquilibre de données et d'autres complications routinières de la modélisation de la standardisation avec des données de test conçues de manière expérimentale. Les simulations peuvent également être conçues pour tester les performances des alternatives pour des évaluations individuelles.

RESUMEN

El programa informático LLSIM fue desarrollado para simular datos de captura de palangre con el fin de cuantificar cuán bien estiman los métodos alternativos la abundancia a partir de la CPUE de palangre, comparando los resultados de la simulación con valores reales conocidos. Emplea un algoritmo de Monte-Carlo con una probabilidad de captura calculada para cada anzuelo en cada operación simulada. El valor de la probabilidad se calcula a partir del tamaño de la población utilizando distribuciones de anzuelos tridimensionales que varían

[1] 1214 North Lakeshore Drive, Niceville, Florida 32578, USA. phil_goodyear@msn.com

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

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

en el tiempo y densidades relativas de las especies. Las distribuciones de las especies, su localización y las distribuciones de los anzuelos en profundidad son modeladas por separado e integradas por el simulador. El método descompone la capturabilidad por anzuelo en un coeficiente esencial de arte (k) y en un efecto de hábitat (H) que varía en el tiempo y en el espacio en 3D. El enfoque proporciona una base para estimar un coeficiente de hábitat (w) que cuantifique el efecto de la calidad del hábitat en la capturabilidad en cada operación de pesca con potencial como covariable para las estandarizaciones de CPUE. Las simulaciones pueden utilizarse para comparar enfoques alternativos de desequilibrio en los datos y otras complicaciones rutinarias de la modelación de la estandarización con datos de pruebas diseñadas de forma experimental. Las simulaciones pueden diseñarse también para probar el funcionamiento de alternativas para evaluaciones individuales.

KEYWORDS

Longline, Catchability, Gear coefficient, Habitat coefficient,
Stock assessment, Statistics, GLM, Data simulation, Population modeling

1. Introduction

In 2003 Myers and Worm made international headline news with the publication of their article "Rapid worldwide depletion of predatory fish communities" in the journal Nature (Meyers and Worm, 2003). Their study popularized the notion that fishing had reduced all of the world's major tuna and billfish populations by more than 90%, based largely on interpretation of longline catch data. The publication sparked intense rebuttal and debate in the most prestigious scientific journals (e.g., Walters 2003, Worm *et al.*, 2006, Sibert *et al.*, 2006, Wilberg and Miller 2007, Murawski *et al.*, 2007, Worm *et al.*, 2007). The initial disagreement centered on misinterpretation of data because of the changing overlap of species and longline hooks with time that imposes an intractable statistical complexity (Walters, 2003; Ward and Myers, 2005; Bigelow and Maunder, 2007). There is general agreement that the original analyses were misleading (e.g., Worm *et al.*, 2009), but methodologies to perform accurate analyses with longline data have proven difficult to confirm (Maunder and Punt 2004, Maunder *et al.*, 2006) The issue is all the more important because stock assessments of most highly migratory species including nearly all billfishes worldwide rely on longline data to quantify stock abundance.

There is no clear, superior, objective method for quantifying abundance from longline catch rates, and no accepted best practice (Campbell 2015). This information is a fundamental requirement of stock assessment and the outcome can still be swayed as much by the bias of the participants as by the data. One reason is that controlled experimentation is impossible. The problem is also growing in complexity because changes in the climate are shifting the distributions of species habitats which can violate stationarity assumptions of common statistical methods. Establishing best scientific methods requires comparing how well alternatives can measure truth, which is never known; hence, the dilemma.

There is no real-world solution to the problem. Data simulation is a workable approach, but requires the simulations to be based on sufficient reality that they capture the important features of real-world variability. This research initiates such an experiment. The project has two major components. The first involves the creation of simulated longline catch data. The second will use those data to identify methods which can reliably estimate population abundance trends from data. The project originated as an international effort coordinated by the ICCAT Working Group on Stock Assessment Methods (Anon 2016).

The focal point of this research involves the computer model (LLSIM) to simulate the longline data that can be used as a set of knowns for experiments to test alternative methods to analyze such data (Goodyear 2006a,b). LLSIM requires information about the distributions of species and of longline hooks in time and three-dimensional space. The current simulator is the result of an evolution of approaches (Goodyear 2003, 2006a,b), that requires each distribution to be specified as a data input (Goodyear 2017). Each component has been another research study in its own right. A new species distribution model (SDM) was developed to predict time-varying, 3-D, day and nighttime relative distributions of pelagic species. It uses thermal utilization patterns from PSAT tagging, published oxygen requirements, and the time-varying distribution of these variables in oceanographic data (Goodyear 2016). The SDM was parameterized with statistics for blue marlin. Personnel from the National Atlantic Oceanographic and Meteorological Laboratory (AOML) applied the Earth System Model to provide oceanographic data monthly from 1946 to 2012 at a resolution of 1º latitude and 1º longitude in 46 layers from the surface to a depth of almost 2 km. Realistic longline fishing effort data required by LLSIM were derived from an analysis of data compiled by the US Longline Logbook and Observer program (Forrestal *et al.* 2017).

Together these models and data provide predictions of the distributions of blue marlin and longline hooks at an unprecedented resolution in time and space. Data are also being assembled to extend the SDM to swordfish, and to include fishing characteristics of other fleets in the simulations. With the development of these simulation tools and supporting data files, the project has begun study of trial simulated data (Forrestal *et al.* 2017). Technical aspects of the simulator are described in the working draft of the LLSIM Users Manual (Goodyear 2017).

The approach adopted for the simulations required careful attention to the factors involved in the partitioning of the notion embodied in the term "catchability". The smallest unit of data for recorded for longline fisheries is the catch on an individual set. The effort is usually based on the number of hooks fished so the catch per unit effort (CPUE) metric for the set is catch per hook (commonly catch/1000 hooks). By definition catchability (q) is the fraction of a fish stock which is caught by a defined unit of fishing effort (Ricker 1975). This definition looses meaning at the level of an individual longline set where the relative abundance of the species is the predominant determinant of the fraction of the stock that is caught. This peculiarity is a confounding feature of the term which takes on different properties depending on how it is applied. Here, the primary concern is related to estimating abundance for stock assessment modeling, but the term may take on different definitions at different points within the same stock assessment because it aggregates data at different scales of resolution.

2. The catchability coefficient (q)

A central tenant of management is to maintain populations at some level of abundance, often with the fishery as the primary or only source of information. Efforts to monitor population abundance from catch proceed from the obvious notion that at some level, catch and CPUE will decline as population abundance falls. The classical approach models this process as a linear effect via a catchability coefficient (q) defined solely from observed catch and effort.

2.1 Standard definition

Catchability is generally used to denote a constant of proportionality (=q) that satisfies the relation between fishing mortality (F) and fishing effort (E), where catch (C) can be determined in biomass (B) or numbers (N). For simplicity, and since hooks catch units of individual fish rather than biomass we restrict notation to N. Following the general notation (Ricker 1975) we have:

$$C = FN, \qquad 1$$

where:

$$F = qE, \qquad 2$$

$$C = qEN, \qquad 3$$

and therefore,

$$q = \frac{C}{EN}\ , \qquad 4$$

or, q = the average proportion of a stock that is taken by each unit of fishing effort.

Of particular interest here is that the customary definition nowhere includes specific reference to fish density. Yet in actuality, it is the change in density that causes the equal units of effort to catch different numbers of fish at different times or places. The term was coined before there was a need to explicitly consider spatial variability in catch rates. It is now clearly understood that catchability combines a gear effect with stock availability (e.g. Zhou *et al.* 2014). In practice, catchability has been used as a catchall term to describe scalars at different resolutions and for different purposes depending on circumstances. Often nuances about usage are left to the insight of the reader. As a matter of principle, even for a single gear, catchability is a continuous variable that is only a constant in special circumstances, i.e., where habitat effects are stationary. Stratification of longline data by season and area attempts to exploit that property. However, complex within and between strata behaviors of the values of q computed from pooled catch data complicate the reliability of this approach (e.g. Wilberg *et al.* 2010).

2.2 Functional definition

We partition the catchability coefficient (q_h) into two factors 1) a gear effect, *k*, and 2) a habitat effect, *H*, due variations in density caused by variations in features of the environment:

$$q_h = k\overline{H}_h, \qquad 5$$

where:
q_h = the catchability coefficient for the hook,
k = gear coefficient, and
$\overline{H}_h$ = the average habitat around the hook.

Since catch is an integer resulting from a series of probabilistic encounters, it is:

$$C_h \cong k\overline{H}_h N. \qquad 6$$

Because of gravity and other effects, depths fished by each longline hook vary by the position of the hook between floats (e.g., Rice *et al.* 2007). In LLSIM, this feature is accommodated by different depth distribution probabilities for each hook position. For convenience, we define a habitat coefficient, *w*, *to* represent the average of $\overline{H}_h$, so the average habitat around all of the hooks during the set is:

$$w = \overline{\overline{H}_h}, \quad \text{and} \qquad 7$$

$$q_s = kw. \qquad 8$$

where:
q_s = the catchability coefficient for the set, and
w = the habitat coefficient.

Again, since catch is an integer resulting from a series of probabilistic encounters, it is:

$$C_s \cong kwEN \qquad 9$$

where:
E is the number of hooks deployed during the set.

2.2.1. The essential gear effect (*k*).

The essential gear effect (*k*) is a constant that accounts for features affecting catch by the gear that do not include the availability of fish in the vicinity of a hook. Within limits, these may be thought of as those factors that affect catch rates operate independently of fish density. Features of longlines that affect *k* include such things as hook type, bait, lightsticks, etc., - essentially everything that does not affect the density of fish in the water surrounding the hooks.

2.2.2. About the habitat effect (*H*)

Both the habitat coefficient (*w*) and the habitat variable on which it depends (*H*) are continuous variables, proportional to the fish density but independent of abundance. The habitat coefficient is the part of the catchability of a longline set for a unique longline configuration that varies in time and space. As reflected by the use of $\overline{H}$ in equations 5-7, the habitat around a hook varies continuously from initial deployment to retrieval. The habitat itself is vertically stratified. In addition, most pelagic species exhibit a vertical diel migratory behavior so the hooks set during hours of daylight fish in a different suite of species densities than those set during the night. In LLSIM, *H* values are input from a separate species distribution model (see Goodyear 2016 for a discussion of the basic design using Atlantic blue marlin). The SDM uses oceanographic data and species habitat preferences to partition the population to a resolution of 1º latitude and 1º longitude and to compute the relative densities in 46 depth layers during hours of daylight and darkness. LLSIM uses these data to compute $\overline{H}$ surrounding each hook on a simulated longline set based on *H* values for the year and month, latitude and longitude, the depth distribution of the hook, and the time of day. Hooks are apportioned to the either the daytime or nighttime species depth distribution using a proportion specified for the set. The effective fishing time is assumed constant for each hook and is independent of the duration of the set. This approach is intended to capture major features of the variability of species habitat utilization but omits many possibly important factors. Examples include effects deriving from lunar phase, crepuscular transitioning, differential feeding schedules, etc. The potential impacts of some of these omissions can be studied by experimentally manipulating input data with the existing model; others may require extensions to the model itself. However imperfect, the existing model provides the means to study performance of standardization methods to many common issues of standardization modeling. Additionally, both $\bar{H}$ in the vicinity of longline sets and values of w estimated for those sets have potential application as covariates for such analyses. The distinction between the two is that $\bar{H}$ can be defined for any convenient local spatial stratification while w incorporates gear specific hook-depth probabilities to estimate $\bar{H}$ around the hooks (Equation 7). LLSIM can provide simulated data to evaluate the merits of these possibilities.

As a general principal, improvements in the SDM with better data or methods will improve all aspects of the study, including estimation of *H*-related variables that may be used as covariates in standardizations. One feature not yet considered involves density-dependent process involved in the spatial distributions. These considerations relate to density-dependency population processes that may enhance species utilization of marginal habitat at high population densities and not to density-dependent behaviors of the fishery as discussed in 3.3. Density-dependent features of the population could be accommodated via the SDM should this aspect of the problem warrant investigation. Though possibly important in many situations, this is a problem of secondary importance when compared to effects of gross data pooling that induce violations of statistical stationarity and other assumptions needed to estimate abundance from CPUE. Density-dependent dispersal is not presently modeled.

3. Simulations

The term "simulation" is somewhat ambiguous. Simulations can be separated into two categories: models of convenience, and models of the processes that lead to observable phenomena (e.g., Punt 2003; Phair 2014; Uusitalo *et al.* 2015; Muhling *et al.* 2016). Models of convenience here refer to models or equations selected because they can be fitted to data. Simulations are often modeling exercises to quantify effects of imprecision and inaccuracy of a fitted equation or set of equations. The simulation evaluates the effect of parameter or measurement error by imposing an error structure on the data or fitted terms and then repeats the analysis many times to study the effect of uncertainty on the prediction of interest. This is perhaps the most common form of simulation applied in many analyses in stock assessment research. This process is ***not*** the method employed here**.**

3.1. The LLSIM algorithm

LLSIM uses a mechanistic approach to simulation in which the predicted event is an inescapable consequence of some fundamental mechanism(s), i.e., where the nature of the relationship is specified in terms of the processes that are thought to give rise to the data - in the case of LLSIM the mechanism is simply a probability calculated from the proximity of a hook with an individual of a vulnerable species. Both the depth of a hook and a fish's position are probabilistic phenomena and the interaction is treated as a probabilistic event. The probability of a catch is estimated for each depth fished by each hook on each longline set. The hooks on a set are partitioned into those that fish primarily during hours of darkness and those that fish primarily during daylight. The catch on each hook fished is treated as a probabilistic event where the probability, *p,* is given by:

$$P = kN\left(\sum_{z=1}^{Z} f_z H_{txyz}\right) \qquad 10$$

where:
f_z = Time (fractions of set duration) fished by the hook in depth layer z, and
H_{txyz} = habitat value for the time (month, year and time of day), at latitude x, longitude y, and depth z.

As with the previous versions of the model, LLSIM can simultaneously model several "species" groupings (Goodyear 2006a, 2017). This feature allows the model to partition a single species into different life stages and/or sexes that exhibit different distributions as well as separate species or even aggregations of different species. This latter feature allows LLSIM to simulate datasets that include species targeting effects and significant bycatch. If more than a single species is simulated, the probabilities for the species are summed and each species competes for the opportunity to be caught. The computed probabilities are entered into a Monte-Carlo to determine the outcome of each fished hook. Essentially, the catch on each hook predicted by the algorithm corresponds to Equation 6, and the total for the set corresponds to Equation 9. The algorithm is described in more detail in the LLSIM User's Manual (Goodyear 2017).

3.2 Calibration

The scale of catch in simulated data is an important consideration. The values of *H* from the SDM should be reasonably constrained and are not amenable to arbitrary adjustment. Similarly, population numbers are likely to be based on information that has at least some touchstone with reality. That leaves the gear coefficient, *k*, as the unknown free parameter that contains the scalar information relating the gear and population size. Inspection of equation 10 reveals the possibility of deterministically predicted probabilities greater than unity which is clearly impossible. Adjusting the value of *k* to get reasonable agreement in catch numbers between simulated and real data assures ranges of simulated data will be relevant to real-world problems. Example considerations would include features such as the frequency of zero catches and avoidance of unintended gear saturation effects.

3.2.1 RevK

Catch on a single set is the smallest unit of measurement in real data. A hook typically encounters many different densities of fish as it sinks and rises in the water column during a set and so has an average during the set, $\bar{H}_h$. The average of $\bar{H}_h$ over all hooks on a set is *w*. For any aggregation of longline sets there is an average habitat effect, $\bar{w}$. During a simulation LLSIM computes $\bar{H}_h$, and *w*, for each simulated set using *H* predicted from the SDM and the hook depth distributions for the gear configuration fished. When the input data are designed to represent an actual fishery, the LLSIM predictions ($\hat{w}$) are estimates of the real *w* values. These data can then be used to estimate a value of *k* for the gear that will result in a catch equal to the observed catch for any non-trivial vector of the population during the simulated period. Catch is generally proportional to the value of *k*. So for a given population vector and gear, the simplest approach is to use LLSIM predicted catches from a trial k_g to estimate a new value based on the ratio of true and simulated catch. The computer program, RevK, automates this task. This process will assign to k_g all of the variability in *q* not removed by *H*. When the k_g are estimated using this approach the differences among gears are actually estimated from the observational data rather than imposed as variables in the simulation. For example, factors such as lightsticks, bait type, hook type, etc. will reflect patterns in the CPUE observations rather than functional relations imposed on the values of *k*.

3.2.2 Echoes of real information in observed longline CPUE trends

Real data impose some constraints on the dynamic range over which a population can fluctuate. For example since *k* is a constant, the dynamic range of simulated cpue ($\bar{H}$ constant) is predominantly controlled by changes in the abundance postulated in the simulation. Using RevK to estimate *k* for a gear using such a hypothetical abundance time series can result in simulated CPUE values remarkably inconsistent with real data and possibly impossible; yet the values may agree on average, nonetheless. When this happens the effect should increases as the real and postulated abundances depart from one another. This phenomenon may or may not be important depending on how the simulated data are to be used. The issue should be closely monitored whenever analyses of simulated data are extrapolated to conclusions about an actual fishery, but otherwise may still be informative about the relative performance of alternative standardization methods.

3.3 Effects of data aggregation: the emergence of density-dependent catchability (q) and other objectionable complications to data analysis

In real data, longline cpue is most commonly expressed in units of catch per hook, conceptually equivalent to Equations 6 and 9. *H* is unknown, and the only avenue to estimating *q* from data requires aggregating catch and effort over expanses of time and space by solving Equation 4 in which *H* will have had great variability. This would not be a concern if the stratum aggregated $\bar{H}$ is constant. However, because of the nature of the longline data, broad seasonal and geographic stratifications are needed to serve as surrogates for *H*. Within those large strata, fishermen can choose to fish during times and within areas where *H* is larger (fish denser) and hence *q* is higher than the stratum as a whole. This leads to within-strata density-dependence and lessens the prospect for stationary *q* for the strata actually used in statistical analyses. A similar effect accompanies many other features that negate the randomness or balance of spatial data. The analytical complications of these phenomena have received considerable attention in the fisheries literature (e.g., Walters 2003, Wilberg *et al.* 2010).

Variability with depth can be an even more intractable problem because hooks can encounter widely disparate ranges of *H*. For example hooks with the same longline gear set in the highly stratified tropical Atlantic encounter a dramatically different depth distribution of fish compared to another area, even with the same surface temperature. The catches on longlines can also be locally depensatory because of gear saturation, i.e., the proportion of fish caught on a unit of longline (the hook) declines with increasing density of fish. Once a fish is caught on a hook, no further increase in density will increase the catch on that hook.

The effects of these phenomena are captured by the LLSIM algorithm. They will emerge in the simulated data whether intended or not. Users are cautioned to avoid undue extrapolations of results from analyses of simulated data to other situations. Results can vary substantially from one set of assumptions to another, both in the simulated and real data, unbeknownst to the analyst. The extent to which such phenomena emerge is strongly dependent on the conditions assumed, and may not be apparent. This is a particular concern when fishing effort is derived from an actual fishery because an analyst can construct the effort data without needing to fully understand issues that may confound subsequent data analysis and lead to possibly unwarranted conclusions.

3.4 Accuracy, error and variability

The current research concerns the performance of methods in terms of accuracy (Walther and Moore 2005). In real data, there is hook by hook, set by set variability in real catch. There is also observational error in the reporting of that catch, the associated effort, its location, etc. These features are part of statistical variability called "error" included in every CPUE standardization. Error quantifies variability about the statistical fit rather than the accuracy of estimates. This fact inevitably leads to confusion. Anyone with experience with the stock assessment process can attest there is a strong tendency to accept results with low residual error as being accurate. Additionally, the error estimates may be used to weight standardized CPUE time series in the assessment models. The error essentially describes "goodness of fit" rather than accuracy. Analyses of simulated data can clarify the validity of this convention.

There is presently no additional random source of error added to the LLSIM simulations. Except for the Monte Carlo outcome of the catch/no catch probability on an individual longline hook, all of the variability in the simulation of CPUE is deterministic, i.e., no parameter uncertainty is considered. There is huge variability in simulated CPUE, independent of population abundance, which is a consequence of time-space variability because of how the hooks are deployed in the fish's probability distribution (the H). This variability is deterministic, but statistical models will consider it to be random error. Other sources of variation can easily be added to the LLSIM data, post-simulation, to study issues related to observational error or bias.

3.5 Other features not included in the simulations

3.5.1 Factors affecting variability

There are at least two factors not included in the modeling that may have important effects on the catch probabilities of longline sets. 1) Within-gear variability in the gear coefficient k as might accompany different vessels or captains using what appears to be the same gear, and 2) species aggregations (like schooling) that might occur on a spatial scale smaller than the spatiotemporal grid of the simulator. Both processes would increase variability in the data, but the overall averages for the experimental strata they represent would not be affected. Either could be accommodated by adding a random multiplier on the probability calculated in Equation 10. The magnitude of the patchiness could be adjusted to match the variation in observed catch rates within appropriate strata. Alternatively differences in k between fishers due skill or other effects could be directly studied in simulations by assuming the gears employed were actually different.

3.5.2 True population time series

The time trend of the population that is used at the true value for the simulations is an input to LLSIM. There is no constraint that the values must reflect trends that may be reasonably expected of real populations. However, as with previous versions, population dynamics models using species specific biological data can be employed to predict the abundance trends that are driven by assumed variations in fishing mortality or recruitment patterns that are the outcomes of other external processes. This approach can then couple the simulated CPUE time series with detailed biological information and provide data sets that can be used to test standardization methods that employ such data, or the entire assessment process.

4. Studies

LLSIM provides the capabilities for investigating many questions related to the analysis of longline catch per unit effort data. For example, are assumptions about day/night patterns of fishing important? If so, is there a best alternative data treatment to minimize the effect for estimating the relative abundance? Similarly, does competition for hooks matter? When does the abundance of rivals for a bait matter? How does spatiotemporal variability in the density of other competing species affect abundance estimates? How important is species targeting? Does it just changing the quality of habitat fished (H), or does it fill up the hooks as well? Simulated data can also be used to directly assess error caused by stratification intrinsic to the data (as might be caused by the resolution of the data records), and stratifications required make statistical analysis possible. Such analyses may require constructing new SDM sub-model datasets which arise because of changing environmental data that possibly change species range(s). But many questions can be answered by "gaming the system" and simply moving hypothetical fishing effort to more/less productive fishing grounds. Also, the data sets can be used to study stratifications to account for gear (k) effects and interactions. The robustness of predicted abundances to these and many other features can be readily investigated from simulated data; however, the process can also help resolve the more general questions about best methods as they relate to the interactions of complex species behaviors, changing oceanography, current and historical patterns of fishing, etc.

4.1 Features in common

Data simulation has the advantage of not requiring field data as a precursor for applying a statistical or other method of analysis. The data to be analyzed are created/predicted using a model. "Truths" and data values will be known with certainty but a simplification of reality. As with real data, extrapolations of findings are always only as robust as permitted by the design of the experiments, but the experiment can be shaped to include many different situations. Describing a simulation experiment is complicated by ambiguity about what is called "data". The process is hierarchical. Data files about features of longline gears and spatiotemporal fishing patterns are used as inputs to programs that define fishing effort. Other data about fish behavior and oceanographic conditions in time and space are used to quantify fish distribution. Together these data and models are used to create the input files that LLSIM needs to simulate longline catch data. These simulated data in turn become the input data for analyses using alternative analytical and methodological treatments. Finally, the results of these analyses are the data used to form judgments about best approaches and/or reliability based on how well they replicate the temporal sequence of abundances used as input to LLSIM.

The application of simulated data to evaluate methodological issues will generally involve several common steps. Generally these will fall into three categories (**Figure 1**). First is the need to create data necessary to test the issue of concern, i.e., to challenge the methods with data relevant to the problem under consideration. For example, if the issue is related only to the verification of a particular method applied in an assessment, this step might consist of developing data sets to match environment and fishing information as closely as possible. The simulations would pair this data with randomly or systematically derived population trends. These simulated data can be tested against the assessment methodology and alternatives. Another application might examine the importance of the duration of CPUE time series, sensitivity of estimators to environmental change, missing data in the GLM stratifications, or any of innumerable alternatives. In each case the experimental design would have to be carefully crafted to simulate data specific to the problem investigated.

Once the experiment is designed, the next step is to perform the required simulations with LLSIM. Depending on the intent of the analysis this step may require manipulations of data about patterns of fishing in time and/or space, or evaluations of hypothetical changes in the environment. The "last" step in the process is to analyze the simulated CPUE data using alternative methods and data manipulations to compare estimated abundance trends to the known values. Depending upon the intent of the particular experiment, the scope of this stage of the analysis might be quite complex in order to quantify the effects of various alternative transformations and stratifications that might be employed. In each case the analyses need some measure of performance which would allow quantitative and qualitative characterization of their relative performance that would ultimately lead to a synthesis of the results of the analysis of the analyses.

4.2 Branches of Investigations

Stock assessments routinely employ population trends derived with GLM methods. There are two somewhat separable classes of inquiry that can be aided by simulated data 1) those related to specific realizations of catch and effort statistics encountered during stock assessments, and 2) generic evaluations intended to elucidate strengths and weaknesses of alternative approaches to maximizing accuracy.

4.2.1 Confirmation/estimation of specific trends

Here, the question is related to the estimation of abundance from an actual time series of catch and effort for a particular fish stock. In some cases GLM analyses are problematic because different but reasonable assumptions produce disparate estimates of population trends from the same set of data. The times series of abundance is so important to the stock assessment that estimation usually cannot be simply omitted. The issues revolve around the selection of variables, stratification, data transformations, and similar concerns. In many cases, observation quality varies in important ways between the early and latter part of the time series. Also it is common that dissimilar data must be somehow aggregated into strata to accommodate missing information, and/or other basic assumptions that must be compromised to even apply a statistical model.

Simulations designed around the constraints posed by the real data could add information about the strengths and weakness of estimated time series in ways that traditional sensitivity analyses cannot. In this paradigm it is important for the distributions of the species and fishing gears to be modeled as accurately as possible. These methods could also include analysis of the SDM and/or gear sub-model predictions of habitat coefficients or similar variables calculated for each set. The accuracy of methods of analysis can then be directly examined using alternative time series with different vectors of population abundance. In this situation, the experimental design might use a predesigned matrix of alternative decision variables, or be based on the subjective judgments of analysts skilled in stock assessment methods using their standard approaches. In the latter case, extrapolations of the findings might be enhanced by having the analyses of the simulated data completed without knowledge of the true values. If simulated data are employed as a part of the assessment process, special attention should be given to uncertainty surrounding the SDM and gear models used to produce the LLSIM input. The identification of such limitations might be enhanced with sensitivity analyses of the whole process to SDM and gear model error which would ideally be included in the assessment process.

4.2.2 Investigational

The other class of studies involves situational manipulations of the environment or fishery to learn about how different methods and assumptions lead to robust estimations in different circumstances. The scope of possible questions that could be investigated with simulated data is beyond the reasonable grasp of a single study. The basic list would begin with the question: is the GLM the best framework for analysis? Other topics include: what features cause the greatest risk of error when using a GLM to quantify abundance from longline data, and, how to avoid them. These larger topics are themselves aggregations of more specific issues. Many applications of LLSIM will be more or less sensitive to error in the SDM or gear models depending on the issue studied. This factor should be a central part of study planning. The availability of the simulation technology will permit analyses of problems that cannot practically be accomplished with real data. However, the designs of experiments to examine these issues may often require careful planning, ideally well thought out and completed before the studies commence.

4.3 Provisional hypothesis

One important result of the research into the methods used in the simulations was the separation and quantification of the effect of habitat H and the definition of the habitat coefficient, w, as a part of catchability coefficient q_s associated with each set. If a suitable SDM exists for the species that is the subject of a CPUE standardization, then $\bar{H}$ in the vicinity of a set is estimable. If in addition, suitable information about the hook depths for the longline involved is available, then w is also estimable. Either variable should be superior to other treatments of data to represent habitat effects in a GLM. Although precision would be dependent on the scale of the spatial specificity of the CPUE observation, either would be a continuous random variable available for each set. Consequently $\bar{H}$ and w would be variables of choice in longline standardizations. This hypothesis can be tested using the LLSIM simulation environment, and should be included as an option during evaluations of alternative standardization methodologies.

Acknowledgements

CPG's contribution to this study was supported by The Billfish Foundation. FCF's contribution was supported by 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 oceanographic data.

References

Anon. 2016. Report of the 2015 meeting of the ICCAT Working Group on Assessment Methods. Collective Volume Scientific Papers, ICCAT, 78(8):2249-2303.

Bigelow, K.A., Maunder, M.N. 2007. Does habitat or depth influence catch rates of pelagic species? Can. J. Fish. Aquat. Sci. 64, 1581–1594.

Campbell, R.A. 2015. Constructing stock abundance indices from catch and effort data: some nuts and bolts. Fisheries Research 161:109-130.

Forestall, 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. ICCAT SCRS/2017-??.

Goodyear, C.P. 2003. Tests of the robustness of habitat-standardized abundance indices using simulated blue marlin catch-effort data. Marine and Freshwater Research 2003(54):369-381.

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

Goodyear, C.P. 2006b. Simulated Japanese longline CPUE for blue marlin and white marlin. Collective Volume Scientific Papers, ICCAT, 59(1): 211-223.

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

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.

Maunder, M.N., Hinton, M.G., Bigelow, K.A., and Langley, A.D. 2006. Developing indices of abundance using habitat data in a statistical framework. Bulletin Of Marine Science, 79(3): 545-559.

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

Muhling, B.A., Brill , R., Lamkin, J.T., Roffer, M.A., Lee, S.-K., Liu, Y., and Muller-Karger, F. 2016. Projections of future habitat use by Atlantic bluefin tuna: mechanistic vs. correlative distribution models. ICES Journal of Marine Science 74: 698-716.

Murawski, S., Methot, R. and Tromble, G. 2007. Biodiversity loss in the ocean: How bad is it? Science 316:1281-1281.

Myers, R. A. and Worm ,B. 2003. Rapid worldwide depletion of predatory fish communities. Nature 423: 280-283.

Phair, R.D. 2014. Mechanistic modeling confronts the complexity of molecular cell biology. Molecular Biology of the Cell 25: 3494–3496.

Punt A.E. 2003. The performance of a size-structured stock assessment method in the face of spatial heterogeneity in growth. Fisheries Research, 65: 391–409

Rice, P.H., Goodyear, C.P., Prince, E.D., Snodgrass, D., and Serafy, J.E. 2007. Use of catenary geometry to estimate hook depth during near-surface pelagic longline fishing: theory versus practice. North American Journal of Fisheries Management 2007;27:1148-1161

Ricker, W.E. 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. 191, 382 p.

Sibert, J., Hampton, J., Kleiber, P., and Maunder, M. 2006. Biomass, Size, and Trophic Status of Top Predators in the Pacific Ocean. Science 314:1773-1776.

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

Walther, B.A., and Moore, J.L. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography. 28: 815-829.

Ward, P. and Myers, R.A. 2005. Inferring the depth distribution of catchability for pelagic fishes and correcting for variations in the depth of longline fishing gear. Can. J. Fish. Aquat. Sci. 62: 1130–1142.

Wilberg, M. J., & Miller, T. J. 2007. Comment on" Impacts of biodiversity loss on ocean ecosystem services". Science 316(5829): 1285-1285.

Wilberg, M.J., Thorson, J.T., Linton, B.C. and Berkson, J. 2010. Incorporating time-varyingcatchability into population dynamic stock assessment models. Reviews in Fisheries Science, 18(1):7–24.

Worm, B., Barbier, E. B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., Jackson, J.B.C., Lotze, H.K., Micheli, F., Palumbi, S.R., Sala, E., Selkoe, K.A., Stachowicz, J.J., and Watson, R. 2006. Impacts of biodiversity loss on ocean ecosystem services. Science 314:787-790.

Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., Jackson, J.B.C., Lotze, H.K., Micheli, F, Palumbi SR, Sala E, Selkoe K, Stachowicz JJ, Watson, R. 2007. Response to comment on 'Impacts of biodiversity loss on ocean ecosystem services'. Science 316:1285

Worm, B., Hilborn, R., Baum, J.K., Branch, T.A., Collie, J.S., Costello, C., Fogarty, M.J., Fulton, E.A., Hutchings, J.A., Jennings, S., Jensen, O.P., Lotze, H.K., Mace, P.M., McClanahan, T.R., Minto, C., Palumbi, S.R., Parma, A., 1Ricard, D., Rosenberg, A.A., Watson, R., and Zeller, D. 2009. Rebuilding Global Fisheries. Science 325:578-585.

Uusitalo, L., Lehikoinen, A., Helle, I., and Myrberg, K. 2015. An overview of methods to evaluate uncertainty of deterministic models decision support, Environmental Modelling & Software 63: 24-31.

Zhou, S., Klaer, N. L., Daley, R. M., Zhu, Z., Fuller, M., and Smith Anthony D. M. 2014. Modelling multiple fishing gear efficiencies and abundance for aggregated populations using fishery or survey data. ICES Journal of Marine Science 71: 2436–2447.

- **Challenging the methods**
  - Duration of time series
  - Population trends
    - Magnitude
    - Time schedule of changes
    - Stability of other species
  - Environmental trends
    - Magnitude
    - Time schedule of changes
    - Spatial schedule of changes
  - Patterns of fishing
- **Run LLSIM**
- **Analyze the simulated data**
  - Select models
  - Perform analyses (twisting the knobs)
    - Transformations
    - Stratifications / Covariates to evaluate
    - Measure performance
  - **Synthesis (Analysis of the analyses)**

Figure 1. Shared features of different simulation experiments that might be conducted using LLSIM.