

CATCH ESTIMATION UTILIZING SIMULATION MODELING AND BOOTSTRAPPING METHODOLOGY

Brown, C. A.

*U.S. Department of Commerce, National Oceanographic and Atmospheric Administration, National Marine
Fisheries Service, Southeast Fisheries Science Center, 75 Virginia Beach Drive, Miami, Florida 33149, USA*

SUMMARY

A new catch estimation methodology was developed for survey data from the handline and rod and reel fisheries off the northeast U.S. Previously, total catch of a species was expressed as the product of: Mean Est. of Fleet Size X Mean Est. of Trips per Boat X Mean Est. of Catch per Trip. A more appropriate model may be to express total catch as the sum of the catches from all trips:

$$C_w = \sum_{b=1}^F \sum_{t=1}^E C_{bt}$$

- where C_w = The total catch for time period w (usually a week long)
 F = The estimated fleet size
 E = The effort expended by boat b during the time period (in numbers of trips)
 C_{bt} = The catch for boat b on trip t .

However, these parameters are not known, as only random surveys are conducted. A computer model was developed in order to simulate the trips and catches. The parameters F , E , and C_{bt} are estimated by "bootstrapping" (sampling with replacement) the appropriate data from the survey. A simulation is run 1000 times, permitting the estimation of non-parametric 95 percent confidence limits. Results obtained agreed closely with estimates calculated using the previous method, but also provided information about the uncertainty of the estimates.

RESUME

Une nouvelle méthode d'estimation des prises a été développée pour la prospection des données des pêcheries à la ligne à main et à la canne et moulinet au large du nord-est des Etats-Unis. Antérieurement, la prise globale d'une espèce était exprimée comme produit de: moyenne estimée de la taille de la flottille x moyenne estimée des sorties par bateau x moyenne estimée de la prise par sortie. Un modèle plus approprié peut être d'exprimer la prise globale comme la somme des prises de toutes les sorties:

$$C_w = \sum_{b=1}^F \sum_{t=1}^E C_{bt}$$

- où C_w = La prise globale de la période spatiale w (normalement d'une semaine de durée)
 F = La grandeur estimée de la flottille
 E = L'effort fourni par bateau b durant la période spatiale (en nombre de sorties)
 C_{bt} = La prise par bateau b de sorties t .

Ces paramètres sont néanmoins inconnus, étant donné qu'uniquement des prospections au hasard sont menées à bien. Un modèle informatique a été développé pour simuler les sorties et les prises. Les paramètres F , E et C_{bt} sont estimés en effectuant des bootstrap (échantillonnage avec substitution) des données appropriées de la prospection. Une simulation est exécutée 1000 fois, permettant d'estimer des limites de confiance non paramétriques de 95%. Les résultats obtenus concordaient quasiment avec les estimations calculées en utilisant la méthode antérieure, mais fournissaient également des informations sur l'incertitude des estimations.

RESUMEN

Se desarrolló una nueva metodología de estimación de la captura para obtener datos de prospecciones, basándose en las pesquerías de línea y de caña-carrete frente al nordeste de Estados Unidos. Anteriormente, la captura total de una especie se expresaba como el producto de: Est. Media del Tamaño de la Flota x Est. Media de Viajes por Barco x Est. Media de Captura por Viaje. Un modelo más apropiado podría ser el de expresar la captura total como la suma de las capturas de todos los viajes:

$$C_w = \sum_{b=1}^F \sum_{t=1}^E C_{bt}$$

- donde: C_w = Captura total en un período de tiempo w (generalmente, una semana de duración).
 F = Tamaño estimado de la flota.
 E = Esfuerzo aplicado por el barco b durante el período temporal (en número de viajes).
 C_{bt} = Captura obtenida por el barco b en el viaje t .

No obstante, se desconocen estos parámetros, ya que sólo se llevan a cabo prospecciones aleatorias. Se desarrolló un modelo por ordenador para simular los viajes y las capturas. Los parámetros F , E , y C_{bt} se estiman mediante ensayos iterativos de reajuste a partir de submuestras ("bootstrap") (muestreo con sustituciones) de los datos pertinentes de la prospección. Se efectúan 1.000 ensayos de una simulación, lo que permite la estimación de límites de confianza no paramétricos del 95%. Los resultados obtenidos concordaron estrechamente con las estimaciones calculadas aplicando el método anterior, pero también suministraron información acerca de la incertidumbre de las estimaciones.

Introduction

Catch estimates for the handline and rod and reel fisheries off the northeast U.S. have typically been developed from large pelagic survey data. Traditionally, total catch (of a species for a given area) was expressed using a multiplicative equation incorporating fleet size, trips per boat, and catch per trip:

$$C_w = F * \bar{T} * \bar{c} \quad (1)$$

where C_w = The total catch for time period w (usually a month long)

F = The estimated fleet size

\bar{T} = The mean number of trips per boat during the time period

\bar{c} = The mean catch per trip during the time period.

Problems were encountered in computing confidence limits around the catch estimates by this method. Observed values around each mean typically followed non-normal distributions. Ignoring the assumptions of normality and applying techniques such as the delta method resulted in unrealistic confidence limit estimates.

An alternative catch estimation model was developed to simulate the fishing activity and resulting catch using "bootstrapping" methodology (Efron 1979) to incorporate the degree of uncertainty reflected in the data. A bootstrap sample is drawn randomly with replacement from an ordinary sample. Bootstrap inference assumes that: 1) the known population is an ordinary sample from some unknown population and 2) the bootstrap sample is drawn from the ordinary sample in exactly the same way that the ordinary sample was drawn from the unknown population (Lunneborg 1987). In this instance, we are assuming that the empirical data are sampled randomly from the fishery; therefore inferences can be made by randomly drawing from these samples during computer simulation. As a result, an entire distribution of values that reflects the uncertainty in the data is obtained. The median of this distribution is usually close to the mean estimated from the original sample.

A goal of this approach was to define, given the observed data, a range within which 95% of the possible catch estimates would fall. In order to incorporate the empirical data as measures of fleet size, effort, and catch per trip, an alternative equation defining catch for a given area as the sum of the catches from every fishing trip made is:

$$C_w = \sum_{b=1}^F \sum_{t=1}^E C_{bt} \quad (2)$$

where C_w = The total catch for time period w (usually a week long)

F = The estimated fleet size

E = The effort expended by boat b during the time period (in numbers of trips)

C_{bt} = The catch for boat b on trip t .

This model allows for the estimation of confidence limits through simulation. The parameters F , E , and C_{bt} can be estimated from the appropriate survey data.

Methods

Each year from June through October, a survey is conducted to collect catch, effort and size data on pelagic recreational fisheries from Virginia to Massachusetts. There are two main elements to this survey: 1) dockside intercept interviews with boat captains returning from fishing trips and 2) telephone interviews with captains known to participate in the fishery. The data components collected during the survey and used in catch estimation for each species are: 1) names of captains and their boats, 2) the number of trips for each boat by time and area strata and 3) the catch (in number of fish of each species) for each trip by time and area strata.

Fleet Size Estimate (F)

Telephone sampling is used to obtain names of individuals/boats which fish from each area for large pelagics. The number of previously known boats (those appearing on the phone list, or "marked") was multiplied by the ratio of boats named ("captures") to previously known boats named ("recaptures") to calculate a simple Petersen estimate of fleet size:

$$F = \frac{C}{I} * M \quad (3)$$

where C = the number of boats named by captains (captures)

I = the number of those boats also appearing on the list of boats known to participate in the fishery (recaptures)

M = total number of boats on the known list (marked)

This provides a single estimate of fleet size for each area. For the purposes of simulation, a new fleet size estimate must be created for each run which incorporates a measure of the degree of uncertainty in the data. To accomplish this, an array variable was created for each area representing the "mark-recapture" sample:

MRSAMP(A, B)

(4)

where A = area, ranging from 1 to 7 for each state from Virginia through Massachusetts (Connecticut and Rhode Island treated as one area)

B = boats named, ranging from 1 through the total number of captures

Elements of the array are assigned a value of 1 if the boat is a recapture, otherwise each is assigned a value of 0. Thus, for an example where 100 captures were made of which 10 were recaptures, ten elements of the array would have a value of one (1) and the rest would be filled with zeros. For each run, this array is bootstrapped (Figure 1), drawing one element for each capture and adding each value (1 or 0) to the sum of recaptures to determine r (for equation 3). For each simulation run, a separate fleet size estimate can be calculated using equation 3.

Effort per Boat (E)

During the telephone interviews, captains are questioned about trips made during the previous recall period, or "wave" (usually a week long). The number of trips made for each boat form the elements for an effort array variable:

$$TRIPS(A, W, B) \quad (5)$$

where W = wave, ranging from 1 to n (n = length of season)

B = boats contacted during the wave, ranging from 1 through the total number contacted.

During simulation, an element of this array is randomly accessed once for each boat in the fleet within a given wave and area (Figure 2). This determines the number of trips to be simulated for each boat.

Catch per trip (c_{bt})

Catch data are collected during both telephone and dockside interviews. These data were assigned to the catch array variable:

$$CPT(A, W, I) \quad (6)$$

where I = interviews with catch information, ranging from 1 through the total number of trips interviewed.

During simulation, an element of this array is randomly accessed once for each trip within a given wave and area (Figure 3). This determines the catch to be simulated for each trip.

Description of Simulation Process

The modules concerning fleet size estimation, effort per boat

and catch per trip provide estimates for the parameters F, E, and c_{bt} and together form the core of the simulation model (Figure 4). For a given area and species, the fleet size (F) is derived using the (bootstrapped) mark-recapture sample, the number of trips is randomly drawn for each of those boats (E) from the effort sample for the wave, and the catch for each of those trips (c_{bt}) is drawn from the catch per trip sample for the wave and added to the catch per wave sum (C_w). Total catches can then be estimated by summing the catches for all waves contained within each time period. This process is typically repeated 1000 times. From these iterations, the median and empirical 95% confidence limits can be calculated.

Although not described in the flow charts, catches are estimated simultaneously for all species of interest. This is accomplished by adding an additional level to the catch per trip array variable (CPT) to represent species. Whenever the catch information from a trip is randomly accessed, the catch is incremented for each species caught on that trip. This approach preserves any within trip proportions of catch by species.

One potential problem occurs when the level of effort and/or catch sampling within a wave is considered to be inadequate. In these cases, samples from adjacent waves can be pooled. This is accomplished by treating all data from adjacent waves as if they were collected during the wave with inadequate sampling, expanding simultaneously to earlier and later waves until the desired number of array elements is reached. For our catch estimates, a minimum sample size of 30 has usually been used. As an example, if the effort array TRIPS contained only 20 observations for wave 4, data from waves 3 and 5 would be added to the array and treated as if they were sampled during wave 4. This procedure would be repeated until the number of observations in TRIPS for wave 4 was at least 30. This process presumes that data pooled across adjacent waves more accurately reflects the population distribution than would a small sample from within one wave.

Results and Discussion

The median estimates obtained using this methodology (equation 2) agreed closely with estimates calculated using the previous method (equation 1). Example results using both approaches are presented in Table 1. The species selected for presentation, including bluefin tuna [*Thunnus thynnus*], yellowfin tuna [*Thunnus albacares*] and blue marlin [*Makaira nigricans*] are examples of both high and low catch levels. Differences between the two types of estimates may occur if extremely high catch rate observations occur infrequently in the samples; this may be expected to elevate the results from equation 1. Differences may also result because the catch rates are calculated for a month at a time for the multiplicative approach (equation 1); this may fail to match fluctuating catch rates with the corresponding level of effort.

The resulting confidence limits provide a measure of the uncertainty around the estimates which was previously unavailable. These confidence limits should be strictly interpreted as the range within which 95% of the possible catch estimates lie given the empirical data. Statistical inference regarding the prediction of the true catch level is dependent upon the degree to which the data represent a true random sample of the population.

Modifications for 1992 estimates

To provide advice on likely levels of harvest during the 1992 recreational bluefin tuna fishery, an improved sampling design was required (Turner *et al.*). This necessitated changes in the catch estimation procedure.

Management constraints required that a catch estimate be provided as rapidly as possible. The software was modified to handle effort and size category specific catch rate frequency distributions as input rather than individual interview records; this was regarded as a more rapid means of data delivery. A consequence of this approach was the loss, if any, of within-trip relationships between size category specific catch rates.

Flexibility in stratifying the catch estimation was considered desirable. The ability to stratify by interview type (phone/dockside/both), boat type (private/charter/both), day type (weekday/weekend/both) and target species was accomplished by adding additional levels to the TRIPS and CPT array variables.

Stratification by boat type required separate charter and private fleet size estimates. In general, a total fleet was first estimated using the mark-recapture samples. Charter fleets were assumed known (Turner *et al.*). Private fleet estimates were then obtained by subtraction. Thus, only the private fleet estimate contributed uncertainty to the catch estimate.

Modifications were also made because the survey was expanded to include North Carolina (NC). As described by Turner *et al.*, NC effort was measured differently from other areas; a near census of total trips during the week and the weekend was made. The number of trips was passed into the software as a point estimate. The fleet and effort estimation modules were bypassed and the catch estimation module was executed once for each trip estimated. Therefore, no measurement of uncertainty around the effort estimate was incorporated into the resulting overall uncertainty about the catch estimate.

The simulation software was also used to examine various simulated parameters in addition to catch, such as fleet size and trips. Given the same assumptions regarding random sampling, conclusions could be drawn concerning the uncertainty around these parameters as well.

Literature Cited

- Efron, B. 1979. Bootstrap methods: another look at the bootstrap. *Ann. Statist.*, 7, pp. 1-26.
- Lunneborg, C.E. 1987. Bootstrap Applications for the Behavioral Sciences, vol. 1. Dept. of Psychology, Univ. of Washington, Seattle.
- Turner, S.C., H.A. Foster, J. Witzig, C. Brown, M. Terceiro, D. Van Vorhees, W. Figley, G.P. Scott, J.A. Browder, D.J. Christensen and C.E. Porch. manuscript. Estimation and projection of the 1992 landings of bluefin tuna in the rod and reel fishery for large pelagic fish off the northeast United States. *Int. Comm. Conserv. Atl. Tunas Working Document SCRS/92/138.*

TABLE 1

Comparison of results using multiplicative approach (equation 1) and summation through computer simulation (equation 2). Lower (LCL) and upper (UCL) confidence limits are shown for the summation estimate.

Species	Year		
	1988	1989	1990
Bluefin Tuna (school size)			
Mult. Est.	11008	16896	24290
Sum. Est.	10918	17825	24541
LCL	9901	16344	22330
UCL	12148	19780	27599
Yellowfin Tuna			
Mult. Est.	21130	27766	12685
Sum. Est.	18160	25898	13303
LCL	16023	23687	12095
UCL	22285	28330	14878
Albacore Tuna			
Mult. Est.	1950	7636	8402
Sum. Est.	1912	6947	8201
LCL	1549	6140	7450
UCL	2332	8082	9129
Blue Marlin			
Mult. Est.	103	143	42
Sum. Est.	88	132	53
LCL	67	107	39
UCL	116	159	67
White Marlin			
Mult. Est.	549	529	492
Sum. Est.	569	501	494
LCL	441	436	439
UCL	867	578	554

Note: 1988 estimates do not include New Jersey catch.

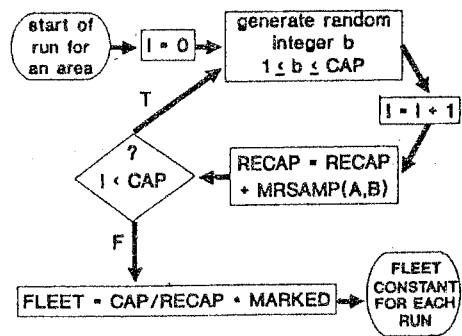


Figure 1. Flowchart for fleet estimation module. This module is executed at the start of each run for a given area. A random integer is generated between 1 and the number of captures (CAP) for the mark/recapture survey conducted in the area. This determines which element of the array MRSAMP will be drawn to increment the recapture total (RECAP). This process is repeated CAP times to generate a bootstrapped sample of the same size as the original sample. The fleet size (FLEET) is then estimated as in equation 3 and is held constant during the simulation run for the area.

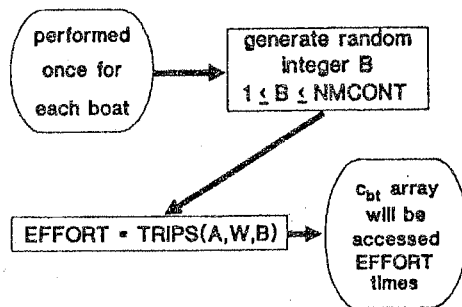


Figure 2. Flowchart for effort estimation module. This module is executed once per wave for each boat in the fleet. A random integer is generated between 1 and the number of boats contacted during the telephone interviews (NMCONT). This determines which element of the array TRIPS will be drawn to determine the number of trips to be simulated for each boat (EFFORT).

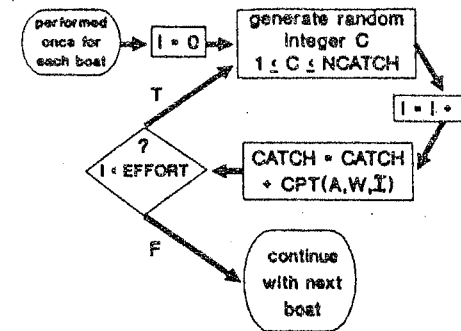


Figure 3. Flowchart for catch estimation module. This module is executed once for each trip being simulated. A random integer is generated between 1 and the number of trip interviews (NCATCH). This determines which element of the array CPT will be drawn to determine the number of fish caught during the simulated trip. This is added to the catch total (CATCH).

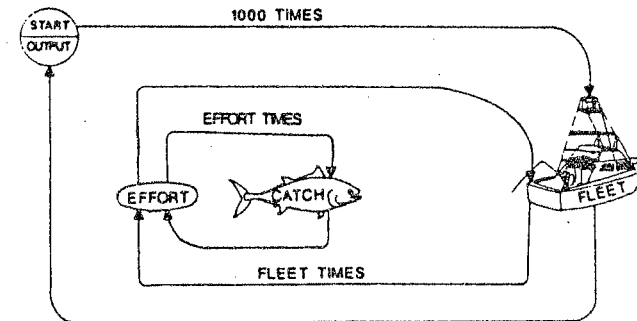


Figure 4. Basic overview of the simulation model. The outer loop represents a simulation run each time a circuit is completed. For each run, a fleet size is generated in the FLEET module. The middle loop is then completed fleet times (once for each boat). For each circuit, the effort per boat is generated in the EFFORT module. The inner loop is then completed effort times (once for each simulated trip by the boat). For each circuit, catch for the trip is generated in the CATCH module and added to the catch total. At the completion of each outer loop, the catch total is output to a data file.