SCRS/1995/088

Col.Vol.Sci.Pap. ICCAT, 45 (2) : 158-161 (1996)

NOTES ON THE POISSON ERROR ASSUMPTION MADE TO ESTIMATE RELATIVE ABUNDANCE OF WEST ATLANTIC BLUEFIN TUNA

Dong, Q., V.R. Restrepo

University of Miami, RSMAS / CUFER, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA

SUMMARY

Until last year, relative abundance estimates for west Atlantic bluefin tuna were based on catch rate (CPUE) data and were typically computed from Generalized Linear Models (GLMs) assuming that the CPUE observations were log-normally or delta-lognormally distributed. For last year's SCRS, this assumption changed as some abundance indices were estimated by modeling catches rather than catch rates assuming that the catches were Poisson-distributed. In this paper, an empirical examination was undertaken of this latter assumption for two data sets: U.S. rod and reel small and large bluefin. The variance-mean relationship of the catch observations is compared against the relationship that would be expected from several types of distributions. These comparisons suggest that the Poisson assumption may be reasonable for the large bluefin data set, but not for the small bluefin one. In the latter case, it seems more reasonable to assume that catches follow a negative binomial distribution. For comparison, GEMs were run for both data sets assuming either negative binomial or poisson-distributed catches.

RESUMÉ

Jusqu'à l'an dernier, les estimations de l'abondance relative du thon rouge de l'Atlantique Ouest se basaient sur les données du taux de capture (CPUE) et étaient typiquement calculées par les modèles linéaires généralisés (GLM), en supposant que les observations de CPUE étaient distribuées de façon lognormale ou delta-lognormale. Ce postulat a changé à la dernière réunion du SCRS, quelques indices d'abondance ayant été estimés en modélisant les prises plutôt que le taux de capture, en supposant que les prises avaient une distribution de type Poisson. Le présent document entreprend un examen empirique de cette dernière hypothèse pour deux jeux de données : petit thon rouge et grand thon rouge de la canne/moulinet des Etats-Unis. Le rapport variance-moyenne des observations de la capture est comparé au rapport que l'on attendrait de plusieurs types de distribution. Ces comparaisons suggèrent que l'hypothèse de Poisson peut s'avérer raisonnable pour le jeu sur le grand thon rouge, mais pas pour le petit thon rouge. Dans ce dernier cas, il semble plus raisonnable de supposer que les prises suivent une distribution binomiale négative. Pour les besoins de la comparaison, des GLM ont été exécutés pour les deux jeux de données, en supposant une distribution des prises, soit binomiale négative, soit de type Poisson.

RESUMEN

Hasta el año pasado, las estimaciones de la abundancia relativa del atún rojo del Atlántico oeste se basaban en los datos de tasa de captura (CPUE) y se solían calcular por medio de Modelos Lineales Generalizados (GLM), suponiendo que las observaciones de CPUE tenían una distribución logarítmica normal o logarítmica normal delta. En el SCRS, el año pasado, este supuesto cambió ya que algunos índices de abundancia se estimaron modelando las capturas, en lugar de las tasas de captura, suponiendo que las capturas presentaban una distribución Poisson. En este documento se hace un examen empírico de este último supuesto para dos conjuntos de datos: caña y carrete estadounidense para atún rojo grande y pequeño. Se compara la relación varianza/media de las observaciones respecto a captura con la relación que se esperaría encontrar en varios tipos de distribuciones. Estas comparaciones sugieren que el supuesto Poisson podría resultar razonable en el caso del conjunto de datos de atún rojo grande, pero no para el del atún rojo pequeño. En este último caso, parece mas razonable suponer que las capturas siguen una distribución binómica negativa. Con el fin compararlos, se hicieron ensayos con GLM para ambos conjuntos de datos, suponiendo capturas con distribución binómica negativa o Poisson.

INTRODUCTION

Generalized Linear Models (GLMs) have been used by ICCAT's Standing Committee on Research and Statistics (SCRS) to develop standardized indices of annual relative abundance of west Atlantic bluefin tuna for years (e.g. Brown and Turner 1989, Brown and Turner 1990, Tuner et al. 1992, Cramer and Turner1994, Brown *in press*). In GLM applications, an assumption must be made regarding the statistical distribution of the dependent variable. This choice of assumption could affect the accuracy and precision of relative abundance estimates. A lognormal distribution (or some variant of it) of catch rates (CPUE) has often been assumed in the past. Last year, following the recommendations of NRC (1994), the SCRS decided to change the way in which analyses were conducted for the 'base case' west Atlantic bluefin tuna assessment: Instead of assuming lognormally-distributed catch rates, Poisson distributed catches were assumed (SCRS 1995). We are not fully aware of all the statistical considerations that led to the basic change in GLM assumptions from lognormally-distributed catch rates to Poisson-distributed catches, other than (1) because the number of bluefin caught in a given trip or set is integer, it is intuitively appealing to assume a Poisson distribution for the catches, and (2) modeling catches as a Poisson process circumvents the problem of dealing with zero catches when modeling catch rates as a lognormal process. This study takes a very basic look at the catch data for bluefin in the U.S. rod and reel fishery to investigate if it is reasonable to assume that the catches follow a Poisson distribution.

DATA AND METHODS

The catch data of large and small bluefin tuna has been collected through interviews of rod and reel and handline fishermen off northeast United States. Turner et al. (1992) describe details of the collecting agencies, methods, regions and variability of the data of large bluefin tuna. Brown and Browder (1994) provide a similar description of data for small bluefin tuna. Each record in the data file examined by us contains information on the number caught, year, month, fishing area, boat type (private or charter), interview method (dockside or phone), fishing method, target of fishing, the number of lines used, fishing hours, sea surface temperature, interview date, and landing location for each fishing trip. Brown and Huang (*in press*) used year, target, boat type, interview type, month, area, and sea surface temperature as main effects in the GLM model with the Poisson distribution for catches. For large bluefin tuna, Brown (*in press*) used year, target, month and area in the GLM models with the assumption of Poisson and gamma-distributed catches. These are the same main factors adopted last year by the SCRS (1995).

Many statistical inferences and computations from GLMs are based on the likelihood functions. For some exponential-type distributions, likelihood theory has been used to break the variance into two components:

$$V(y) = V(u)\, s,$$

where s is a dispersion parameter and $V(u)$ is known as the "variance function" (McCullagh and Nelder 1989, see Table 1). Upon specifying a distribution type for a GLM analysis, one makes a choice about a variance function which, in turn, affects the parameters estimated in the analysis. It seems prudent to check the consistency between the observations used for the analyses and the assumptions made during the analyses.

Our main objective was to address the question: Is the assumption of Poisson-distributed catches consistent with the observations? Our approach was simply to compare the observed variance-mean scatterplots against the expected functional pattern of various distributions. Pearson's χ^2 statistic was used to measure the Goodness of fit between the data and the variance-mean expectation for a given parametric distribution. We did not attempt to address the question of what main effects should be included in a GLM, for a given error assumption for the catches. Thus, we used the main effects identified in Brown and Huang (*in press*) and Brown (*in press*) to define strata in which we could compute the mean and variance of the observed catch data on a per-trip basis: year, boat type, interview type, month and area for small bluefin tuna; year, target, month and area for large bluefin tuna.

We evaluated four distributions - normal, Poisson, gamma, and negative binomial - and acknowledge that there may be other distributions worthy of consideration. The variances of the negative binomial and gamma distributions include a parameter different from the mean, which we estimated with the SAS REG Procedure (SAS 1988)

The negative binomial distribution seems to be consistent with the actual catch observations from both data sets (see Table 2 and next section). For comparison, we ran a GLM with both the negative binomial and Poisson distribution assumptions to produce abundance indices. The SAS program using the GENMOD procedure with a Poisson distribution was provided by C. Brown (NMFS-SEFSC, pers. comm.) and the macro for use with the negative binomial distribution was written by Hilbe (1994a and 1994b).

RESULTS AND DISCUSSION

For large bluefin tuna, catch variances and means calculated from the survey data suggest that a negative binomial distribution or a Poisson distribution are similarly adequate choices (Figure 1, Table 2). For small bluefin tuna, the negative binomial distribution is a much better choice (Figure 2, Table 2). Overall, the negative binomial distribution seems to be consistent with the observed catch data for both large and small bluefin. This is not surprising, as the negative binomial distribution is identical to the Poisson when its additional parameter (see Table 1) becomes 0. Thus, one would expect the negative binomial to fit at least as well, if not better, than the Poisson. The gamma distribution which also has an added parameter, fits well the small bluefin catch data. It does not fit well the majority of the large bluefin catches, but this may be largely due to the influence of one data point with very high mean catches (Figure 1).

For the GLM analyses, the deviance and Pearson's χ^2 were much larger when a Poisson distribution was assumed than when a negative binomial distribution was, particularly for small bluefin (Table 3). Again, this is not surprising because the negative binomial is simply more flexible than the Poisson. For large bluefin tuna, the most significant difference between the indices based on the two distributional assumptions occurs in the first year (1983), with the index based on the negative binomial assumption being more than one third larger than the one based on the Poisson choice (Figure 3). For small bluefin tuna, the yearly abundance indices based on the assumption of the negative binomial distribution are generally larger than the indices based on the assumption of Poisson distribution. The most significant differences occur in the first three years (1980-1982), with up to a three-fold difference (Figure 4). Such differences may or may not have a substantial impact on VPA results, a matter which we did not investigate.

The choice of error distribution models is important. Our brief look at the data suggests that GLM models based on small bluefin tuna catches from the US rod and reel fishery assuming a Poisson distribution may not be statistically defensible. A negative binomial assumption appears to be a better choice. For catches of large bluefin in this fishery, both a negative binomial and a Poisson distribution seem to be adequate, with the negative binomial choice being only slightly better.

It seems important to analyze data sets used by the SCRS to estimate relative abundance, and check for consistency between the data and assumptions. Our analyses were limited to two of five indices used in last year's west Atlantic bluefin assessment which relied on the Poisson assumption. For a more thorough analysis, the remaining three series (US longline in the Gulf of Mexico and Japan longline in the Gulf of Mexico and in the Atlantic) should be examined for consistency between the data and assumptions.

We stress that this issue is not unique to west Atlantic bluefin tuna. For instance, albacore indices of abundance based on Japanese longline data using either the lognormal catch rate or the Poisson catch assumptions were made available to the SCRS last year. For south Atlantic albacore, production model analyses making use of either form of standardization produced very different estimates of stock status and the SCRS opted to use the catch-rate (lognormal assumption) indices in the base case assessment (SCRS 1995).

In summary, we recommend that the consistency between the data and the assumptions made during standardization be checked whenever possible, through analyses of residual patterns, simple methods as those applied here, or other more detailed forms of analysis. This may seem like an obvious thing to do, but evidently it is not done routinely enough.

ACKNOWLEDGMENTS

We are grateful to C. Brown (NMFS, Miami) who provided much assistance with the data and SAS macros and to C.M. O'Brien (MAFF, Lowestoft) for pointing out the usefulness of the variance functions. Support for this study was provided through the Cooperative Unit for Fisheries Education and Research (CUFER) by National Oceanic and Atmospheric Administration Cooperative Agreement NA90-RAH-0075.

LITERATURE CITED

Brown, C. A. *in press*. Standardized catch rates of large bluefin tuna, *Thunnus thynnus*, in the rod and reel/handline fishery off the northeast United States. Int. Comm. Conserv. Atl Tunas, Col Vol Sci. Pap. SCRS/94/133.

Brown, C. A. and Turner, S. C. 1989. Standardized catch rates of large bluefin tuna in the New England (U.S.) rod and reel/handline fishery 1982-1986. Int. Comm. Conserv. Atl Tunas, Col Vol Sci. Pap. 30(2): 302-310.

Brown, C. A. and Turner, S. C. 1990. Standardized catch rates of large bluefin tuna in the New England (U.S.) rod and reel/handline fishery. Int. Comm. Conserv. Atl Tunas, Col Vol Sci. Pap. 32(2): 248-254.

Brown, C. A. and Browder, J. A. 1994. Standardized catch rates of small bluefin tuna in the Virginia - Rhode Island (U.S.) rod and reel fishery. Int. Comm. Conserv. Atl Tunas, Col Vol Sci. Pap. 47(2): 181-191.

Brown, C. A. and Huang, H. *in press*. Standardized catch rates of small bluefin tuna, *Thunnus thynnus*, in the Virginia - Rhode Island (U.S.) rod and reel fishery. Int. Comm. Conserv. Atl Tunas, Col Vol Sci. Pap. SCRS/94/134.

Cramer, J. and Turner, S. C. 1994. Large bluefin tuna, *Thunnus thynnus*, indices of abundance from the rod and reel and handline fishery off the northeast United States. Int. Comm. Conserv. Atl Tunas, Col Vol Sci. Pap. 41(2):.

Hilbe, J. 1994a. Log Negative Binomial Regression Using the GENMOD Procedure SAS/STAT Software. Proceedings of SUGI 19, SAS Institute.

Hilbe, J. 1994b. Log Negative Binomial Regression as a Generalized Linear Model, Technical Report 26, Graduate College Committee on Statistics, Arizona State University, Tempe, AZ 85287.

McCullagh, P. and Nelder, J. A. 1989. Generalized Linear Models. Chapman and Hall, London.

NRC (National Research Council). 1994. An assessment of Atlantic bluefin tuna. National Academy Press, Washington D.C. 148 pp.

SAS Institute Inc. 1988. SAS/STAT® User's Guide, Release 6.03. SAS Institute. Cary, NC.

SAS Institute Inc. 1993. SAS® Technical Report P-243, SAS/STAT® Software: The GENMOD Procedure, Release 6.09. SAS Institute. Cary, NC.

SCRS. 1995. Report for the biennial period 1994-95. Part I (1994). Intl. Comm. Cons. Atl. Tunas, 283 pp.

Turner, S. C., Cramer, J. and Brown, C. A. 1992. Indices of abundance large bluefin tuna, *Thunnus thynnus*, from the rod and reel/handline fishery off the northeast United States. Int. Comm. Conserv. Atl Tunas, Col Vol Sci. Pap. 39(2): 758-777.

TABLE 1. Variance components of four error distributions.

Distribution	Variance Function	Dispersion
Normal	$V_i(u_i) \equiv 1$	s^2
Poisson	$V_i(u_i) \equiv u_i$	1
Gamma	$V_i(u_i) \equiv u_i^2$	$1/v$
Neg. bin.	$V_i(u_i) \equiv u_i + ku_i^2$;	1

TABLE 2. Goodness of fit of bluefin catch data to four distribution models. Pearson's χ^2 statistic is used to measure the discrepancy between the data and the expected variance-mean relationship. A smaller χ^2 value (in boldface) indicates smaller discrepancy and better fit.

Distribution	Large BFT	Small BFT
	χ^2	χ^2
Normal	71.245	13515.7
Poisson	2.768	28440.2
Gamma	112.856	2261.1
Neg. bin.	**2.727**	**1401.0**

TABLE 3. Summary fit statistics from GLM fits to bluefin catches with two assumption models. A lower Pearson's χ^2 (in boldface) is indicative of a better fit.

Distribution	Statistic	Large BFT	Small BFT
Poisson	Deviance	4839.3	28557.4
	χ^2	18107.3	39728.1
Neg. bin.	Deviance	2433.8	5238.7
	χ^2	**12273.5**	**5647.3**

Figure 1. Variances and means of the number of catches of large bluefin tuna

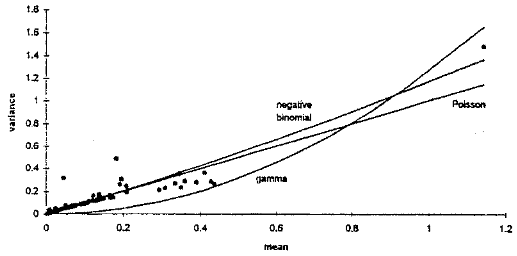

Figure 2. Variances and means of the number of catches of small bluefin tuna

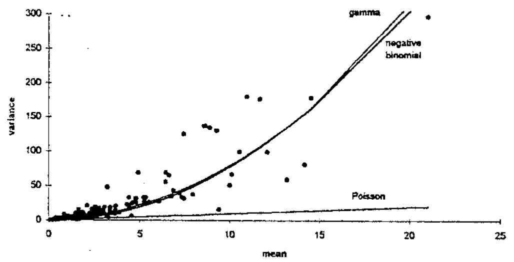

Figure 3. Relative abundance indices for large bluefin tuna

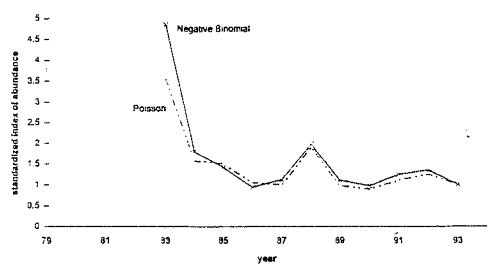

Figure 4. Relative abundance indices for small bluefin tuna