An alternative method for estimating bycatch from the U.S. shrimp trawl fishery in the Gulf of Mexico, 1972–1995

Mauricio Ortiz

Christopher M. Legault

Nelson M. Ehrhardt

Division of Marine Biology and Fisheries
Rosenstiel School of Marine and Atmospheric Sciences
University of Miami
4600 Rickenbacker Causeway
Miami, Florida 33149
Present address (for M. Ortiz): Miami Laboratory
Southeast Fisheries Science Center
National Marine Fisheries Service, NOAA
75 Virginia Beach Dr., Miami, Florida 33149
E-mail address (for M. Ortiz): mauricio.ortiz@noaa.gov

Abstract.—Finfish bycatch taken by the U.S. Gulf of Mexico shrimp fishery is an important issue in the management of fisheries resources given the perceived high mortality of the different fish stocks taken as bycatch in the region. Bycatch data are characterized by a high number of low catches, a few high catches, and depending on the finfish species, a significant proportion of observations with zero bycatch. An evaluation of the current general linear model for generating bycatch estimates indicates that the bycatch data do not conform to the assumptions of this model because bycatch estimates depend upon choices within the model that can significantly change the results of the model. These choices include the constant value added to catch-per-unit-of-effort (CPUE) values prior to the logarithmic transformation (to avoid undefined logarithms with zero CPUEs) and the standard time-unit selection for calculating CPUE values from catch in numbers and variable tow times. Currently a value of one is added to observed CPUE, and a constant time unit of one hour has been used; however, these choices are somewhat arbitrary.

An alternative approach to model bycatch data is to use a delta distribution that has two components. Component one models the proportion of zeros, and component two models the positive catches. In our study, we applied the delta lognormal model to estimate finfish bycatch in the shrimp fishery. This model avoids the problems of 1) the addition of a constant positive value to log-transformed CPUEs, and 2) the selection of a standard time unit for CPUE calculations. Bycatch estimates determined with the current general linear model were compared with those determined with the delta lognormal model for Atlantic croaker (*Micropogonias undulatus*), red snapper (*Lutjanus campechanus*), Spanish mackerel (*Scomberomorus maculatus*), and all finfish from 1972 through 1995. Analysis and evaluation of the performance of the delta lognormal model indicated that this model fits the bycatch database better than the current general linear model.

Manuscript accepted 10 January 2000.
Fish. Bull. 98:583–599 (2000).

In recent years shrimp bycatch has become one of the most important issues in fishery management in the southeastern United States, including the U.S. Gulf of Mexico. In 1990, the U.S. Congress requested a 3-year research program to assess the impact of bycatch by the shrimp fishery on federally managed fishery resources along the south Atlantic and the U.S. Gulf of Mexico coasts (Public Law 101-627, sec110c[1]). As a result, the National Marine Fisheries Service (NMFS) created the Cooperative Shrimp Bycatch Characterization Project (NOAA[1]), a four-year program which focused on 1) characterizing onboard shrimp trawl bycatch, 2) developing and testing bycatch reduction devices (BRDs), and 3) evaluating alternative bycatch management options. Among the major objectives identified in this project were those of updating and expanding temporal and spatial bycatch estimates (offshore and inshore waters) (NOAA[1]).

Since 1987, the NMFS has provided bycatch estimates for several finfish species in the Gulf of Mexico by using a catch-per-unit-of-effort (CPUE) method, where bycatch CPUEs are estimated following a general linear approach (Nichols et al.[2]). Briefly, a bycatch CPUE rate is estimated for each fish species by year (1972–95), area, season, and depth-zone stratum. These bycatch CPUEs are multiplied by an estimated annual shrimping effort within the stratum, and the total annual bycatch is the sum of the bycatch for each stratum. To estimate bycatch, the sample unit is defined as the number of fish of a given species caught each net-hour during a tow. The current general linear model was evaluated by considering two main topics: 1) the assumptions entailed with using the model and the theoretical basis for generating the estimates, and 2) the appropriateness of the available data to the configuration and analysis of the model. More specifically, we examined the matrix structure used in the general linear model, the logarithm usage in the general linear model, and the standardization of effort in the CPUE's in the current general linear model.

[1] NOAA. 1995. Cooperative research program addressing finfish bycatch in the Gulf of Mexico and South Atlantic shrimp fisheries: a report to congress. National Marine Fisheries Service Southeast Regional Office, 9721 Executive Center Dr. N., St. Petersburg, FL 33702.

[2] Nichols S., A. Shah, G. J. Pellegrin, and K. Mullin. 1987. Estimates of annual shrimp fleet bycatch for thirteen finfish species in the offshore waters of the Gulf of Mexico. Report to the Gulf of Mexico Fishery Management Council. The Commons at Rivergate, 3018 U.S. Highway 301 N., Tampa, FL 33619.

Procedure for estimating bycatch with the general linear model

In the bycatch estimation procedure with CPUE, it is assumed that the estimated annual shrimping effort is known and no variance is associated with this value. Therefore, we restricted our evaluation to the general linear model method to estimate the bycatch rates (CPUE) within each stratum. The general linear model is defined for each bycatch species by Nichols et al. (1987[2]) as

$$\text{Log}_{10}\,(CPUE + 1)_{ijklm} = mean + dataset_i + year_j + season_k + area_l + depth_m + \varepsilon_{ijklm}\,, \quad (1)$$

where CPUE = the catch in numbers per trawl for each hour of shrimp fishing;
mean = the overall mean;
dataset (*i*) = a fixed effect term differentiating commercial shrimp fishing from research trawls; and
the terms *year (j)*, *season (k)*, *area (l)*, and *depth (m)* = also fixed-effect terms characterizing the spatiotemporal variability of shrimp bycatch.

This model assumes that the error terms are random, independent, and normally distributed, with equal variance throughout. Predicted catch per trawl net for each hour of shrimp fishing is then estimated for each stratum for the commercial shrimp fishery as

$$CPUE = 10^{(\hat{Y}+1.1513\times RMS)} - 1, \quad (2)$$

where $\hat{Y}$ = the general linear model predicted $\log_{10}$ *(CPUE+1)*; and
RMS = the residual mean square from the general linear model.

The *RMS* term is required to estimate the arithmetic mean from the geometric mean of the assumed lognormal distribution. The constant 1.1513 is a correction factor for estimations derived with log base 10 instead of the natural log.

The predicted CPUE in each stratum is then multiplied by the estimated shrimping effort in the corresponding stratum. CPUEs are estimated for each trawl net. An average of two trawl nets per commercial shrimp vessel for the 1972–95 time series is assumed owing to the lack of information on number of nets per boat for each stratum (cell in the matrix configuration) or other grouping category. Total annual bycatch estimates for a given species are then simply the sum of the commercial bycatch (i=1) in all strata for that year (j) as

$$Bycatch_j = 2\times\sum_{klm} CPUE_{1jklm}\times f_{jklm}, \quad (3)$$

where f_{jklm} = the estimated total shrimping effort (hours of fishing) for year j, area k, season l, and depth zone m.

The general linear model estimates an approximate variance for the arithmetic mean CPUE for each cell as

$$\left(10^{2(\hat{Y}+1.15\times RMS)+2.3(S^2_{\hat{Y}}+2.65\times RMS^2/rdf)}\right)\left(10^{2.3(S^2_{\hat{Y}}+2.65\times RMS^2/rdf)}-1\right),$$

where $\hat{Y}$ and *RMS* = the predicted $\log_{10}$ *(CPUE+1)* and the residual mean square respectively;
$S^2_{\hat{Y}}$ = the estimate of variance of the predicted $\log_{10}$ *(CPUE+1)* for the cell; and
rdf = the residual degrees of freedom.

No variance estimates for the estimated shrimping effort are included in this model; thus effort is considered as if it were known exactly (Nichols et al.[2]).

The database for estimating shrimp bycatch CPUEs was derived from information collected in several projects. The current database comprises two types of data sources: 1) direct measurements of finfish catch by observers onboard of commercial shrimp vessels, and 2) catch rates from research surveys. Direct observations came from four main programs: the Sea Turtle Incidental Catch and Mortality Project (Henwood and Stuntz, 1987), the Excluder Trawl Device Evaluation Project (Henwood and Stuntz, 1987), the Shrimp Fleet Discards Project (Pellegrin, 1982), and the Cooperative Shrimp Bycatch Characterization Project (NOAA[1]). Direct observations were discontinuous in time and space; in particular, no onboard commercial vessels observations occurred between 1982 and 1991. Research observations came primarily from two annual trawling projects: the Fall Groundfish Surveys and the Summer SEAMAP Program. With over 22,000 tows documented from 1972 through 1995, research observations were the main source of the bycatch database. Research observations were restricted to tow surveys with the RV *Oregon II* equipped with a standard 40-ft shrimp trawl (Nichols et al.[3]).

For estimating bycatch, the U.S. Gulf of Mexico was divided into four geographic areas, two depth zones, and three seasons. Area 1 covered the Texas coastline, area 2 covered the Louisiana coast, area 3 covered the Alabama and Mississippi coasts, and area 4 covered to the Florida West Coast and the Lower Florida Keys. Two depth strata were defined by using the 10-fathom depth as the divider of inshore and offshore regions. Temporal variability of shrimp bycatch was taken into account by including three seasons: 1) January–April, 2) May–August, and 3) September–December.

Annual estimates of bycatch for the finfish category (i.e. all fish species, in weight units instead of numbers of fish), and for three fish species (Atlantic croaker, Spanish mack-

[3] Nichols S., A. Shah, G. J. Pellegrin and K. Mullin. 1990. Updated estimates of shrimp fleet bycatch in the offshore waters of the US Gulf of Mexico 1972–1989. Report to the Gulf of Mexico Fishery Management Council. The Commons at Rivergate 3018 U.S. Highway 301 N., Tampa, FL 33619.

Table 1

Distribution of number of cells and observations per cell for the general linear model and the modified models. The 3-area model refers to a reduced number of levels in the area factor of the general linear model (from 4 to 3) by combining areas 2 and 3 into a single area (see text for description of each area). The 2-season model refers also to a reduced number of levels in the season factor of the general linear model where season 1 is from September to April and season 2 is from May to August. The no-depth-zone model refers to the general linear model without the depth zone factor. The combined model refers to a model of 3 areas, 2 seasons, and no depth zone. The year and dset (data set) refers to a general linear model with only these two factors (i.e. excluding season, area, and depth-zone factors). Percent coverage refers to the proportion of cells in the matrix that have tow observations, both by type of data (commercial, research, and combined) as well as the number of positive bycatch tows with Spanish mackerel.

	Matrix structure of general linear model								
	Research			Commercial			Total		
Scenario	No. cells of matrix	Cells with tows	Cells with Spanish	No. cells of matrix	Cells with tows	Cells with Spanish	No. cells of matrix	Cells with tows	Cells with Spanish
General linear model	576	274	175	576	181	77	1152	455	252
3 areas	432	176	129	432	148	68	864	324	197
2 seasons	384	236	153	384	152	69	768	388	222
No depth zone	288	143	110	288	112	61	576	255	171
Combined	144	80	66	144	71	41	288	151	107
Year dset only	24	24	24	24	15	10	48	39	34
	Percentage coverage								
General linear model		47.6%	30.4%		31.4%	13.4%		39.5%	21.9%
3 areas		40.7%	29.9%		34.3%	15.7%		37.5%	22.8%
2 seasons		61.5%	39.8%		39.6%	18.0%		50.5%	28.9%
No depth zone		49.7%	38.2%		38.9%	21.2%		44.3%	29.7%
Combined		55.6%	45.8%		49.3%	28.5%		52.4%	37.2%
Year dset only		100.0%	100.0%		62.5%	41.7%		81.3%	70.8%

erel, and red snapper) were used to compare results of the sensitivity analysis. Atlantic croaker is a species commonly caught as bycatch, found in about 61% of tows. Red snapper and Spanish mackerel are important commercial and recreational fisheries in the U.S. Gulf of Mexico; the directed fishery management actions for these fisheries are influenced by the level of bycatch in the shrimp trawl fishery. Red snapper is caught as bycatch in the shrimp fishery in about 28% of tows, whereas Spanish mackerel is less commonly caught in only 5% of tows.

To evaluate the general linear model, we first describe some characteristics of the database that are important regarding assumptions entailed with the use of the model and analysis of the model , then we present a sensitivity analysis on the model structure and parameters. Three main analyses were performed: 1) analysis of the general linear model matrix structure, 2) analysis of the logarithmic scaling of the observed CPUE values, and 3) analysis of the standard tow time unit used to calculate observed CPUE values.

Evaluation of the general linear model

The present general linear model configuration for estimating bycatch created a matrix of 1152 cells, comprising data for 24 years, 2 datasets, 4 areas, 3 seasons, and 2 depth zones. Although there were a relatively large number of observations in the database (26,380), the percentage of cells in the matrix that had observations was only 39%, only 4160 tows (16%) were from commercial vessels during normal shrimp fishing operations, and the remaining 22,220 (84%) tows were from research vessels. In addition, the number of observations per cell varied largely, from 1 to 466 within research tows, and from 1 to 181 within the commercial tows.

Given the unbalanced distribution of observations per cell, we investigated the effects of the matrix structure on the general linear model. Our approach was to fit the model to scenarios with a reduced number of levels within factors or a reduced number of factors in the model (or to scenarios with both). For example, we combined seasons 1 and 3 to reduce the season factor to two levels or we eliminated the season factor from the model. Table 1 describes all the scenarios evaluated. Correspondingly, shrimping effort was adjusted to the new general linear model matrix by adding the annual shrimping effort within the modified strata, and annual bycatch was estimated as the product of the predicted CPUEs and the shrimping effort for each cell. Defining the percentage of coverage as the number of cells with observations divided by the total number of cells

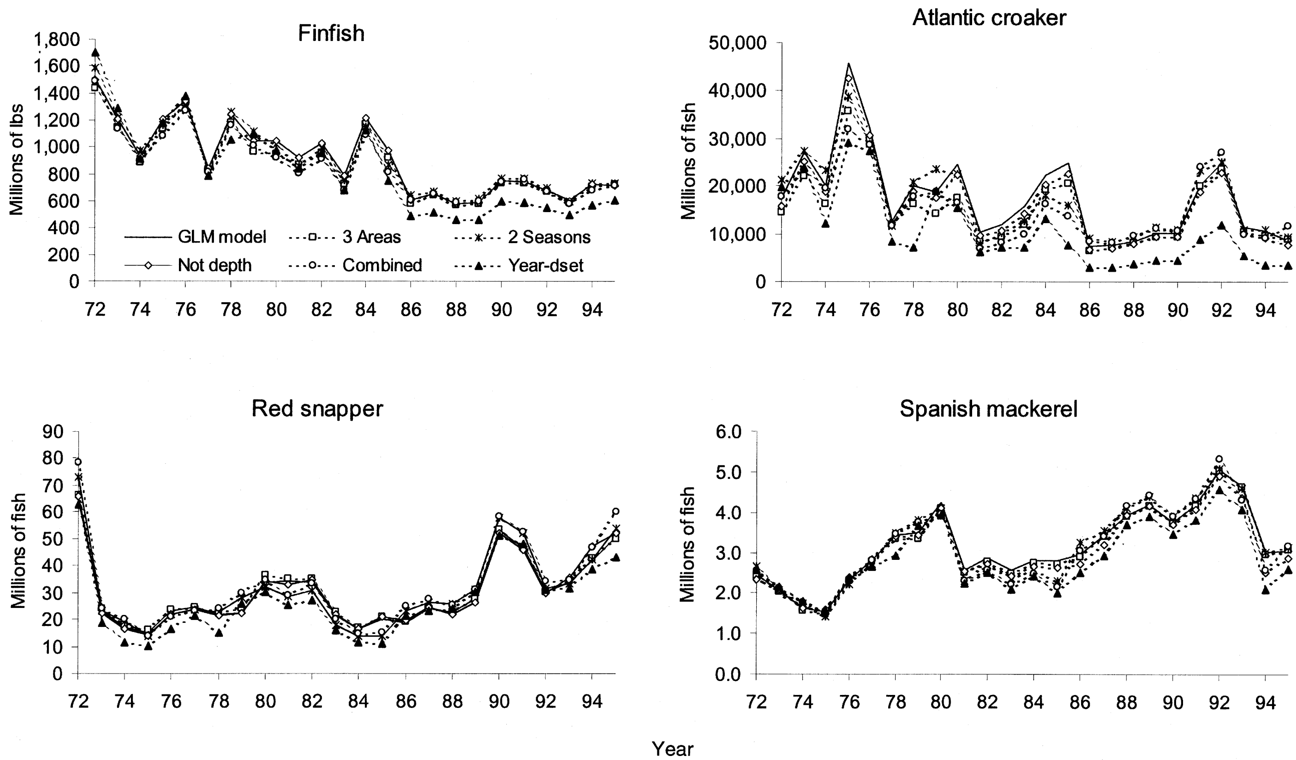

Figure 1

Estimated total bycatch with the general linear model and modified models of the general linear model for the finfish group, Atlantic croaker, red snapper, and Spanish mackerel. (See Table 1 for descriptions of the modified models.)

in the matrix, we found that this value increased from 39% in the base scenario of the current general linear model to 81%, in what was defined as the "minimum model" where only the factors year and dataset were included.

Overall, the results showed that total bycatch estimates did not vary substantially, although the assumed model was radically modified (Fig. 1). These results suggest that season, area, and depth zone are factors that do not significantly contribute to the explanation of the observed variability in the data. Although the F-values from the ANOVA tables were highly significant (P<0.05) for each factor in all general linear model matrix scenarios, this significance may be a response to the large number of degrees of freedom. Alternatively, it is possible that the structure of the general linear model does not reflect all the main factors that account for bycatch variability among years, except for dataset source. Indeed, interactions between the main factors may also be important. Given the limited data coverage, however, the inclusion of other factors or interactions among factors in the general linear model is clearly not advisable.

In summary, the simple model with year and dataset as factors produced similar estimates of bycatch in relation to the complex model, including season, area, and depth zone factors. In particular, for species that are not common as shrimp bycatch, a simple model avoids empty cells and highly unbalanced input matrix designs.

Use of logarithms in the general linear model

One of the assumptions in the linear regression model is that the error within the matrix cells should follow a normal distribution and have a constant equal variance. In the bycatch dataset, the CPUE variance increases as the mean CPUE increases, indicating a constant coefficient of variation. This condition suggests a logarithmic transformation of mean CPUE values. To avoid the problem of undefined logarithms for zero catches, a constant value c of 1 was added to all observed CPUE (Eq. 1) in the model. Then the linearization procedure was carried out on the log base 10 of the modified CPUE. This c value was then subtracted in the back transformation of the predicted means (Eq. 2). No particular explanation for the choice of 1 in the current general linear model has been given.

Thus, we considered the effects of using different c values in the general linear model. Three different c values where used: 10, 0.5, and the smallest positive CPUE-value for each species (i.e. 0.0178 for finfish, 0.0779 for Atlantic croaker, 0.0685 for red snapper, and 0.0685 for Spanish mackerel). The results showed that annual bycatch estimates vary dramatically depending upon the c value used in the algorithm (Fig. 2). Although the magnitudes varied with changes in the c value, the trends were the same for each species. However, the direction of change was not the same among spe-

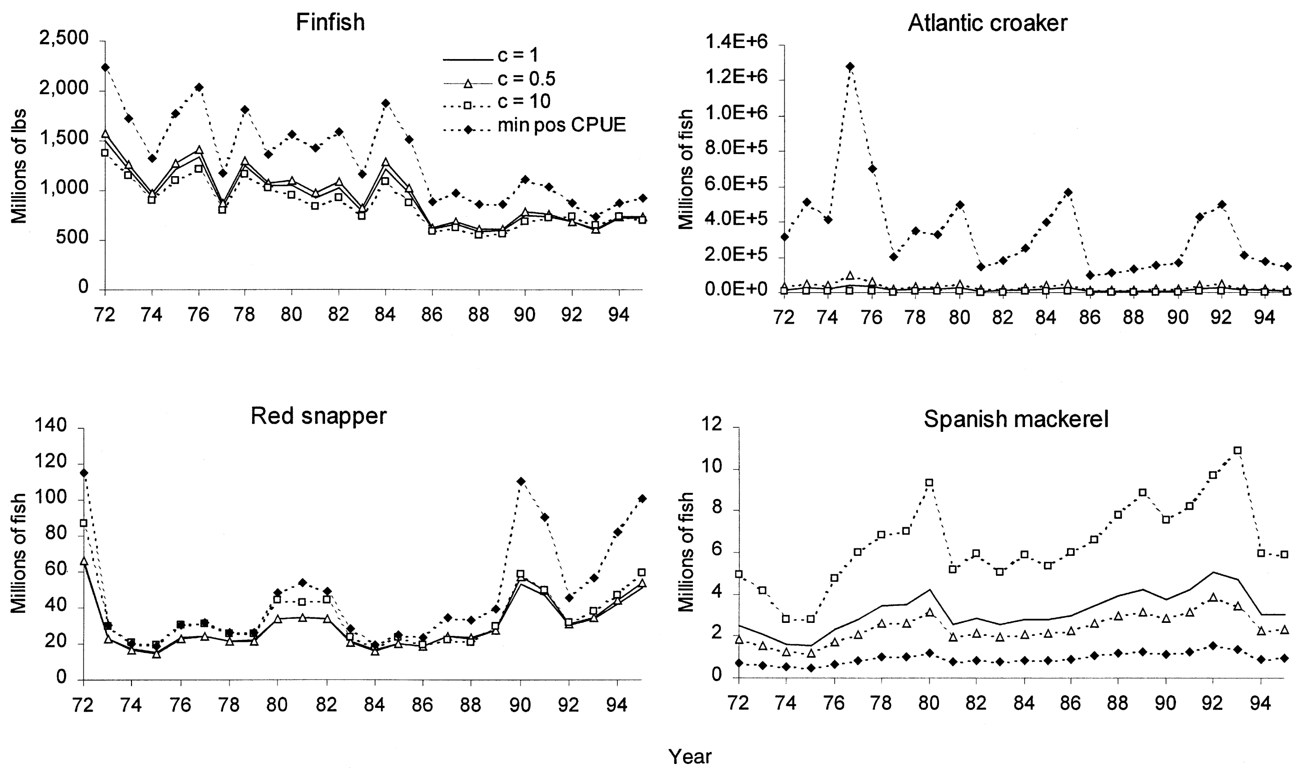

Figure 2

Estimated total with the general linear model with different c-values used in the logarithmic transformation of bycatch CPUE. Base scenario (c=1).

cies. For Spanish mackerel and red snapper, using c=10 increased the estimates of bycatch (100% and 15%, respectively). In contrast, bycatch estimates decreased for Atlantic croaker and finfish (75% and 6%, respectively). When the c was the smallest positive value of the data, annual estimates increased on average 47% for red snapper, 43% for finfish, and 1694% for Atlantic croaker, whereas bycatch estimates decreased on average 70% for Spanish mackerel.

These results show that the general linear model is highly sensitive to the logarithmic c value added to the observed CPUE values. Although it is known that logarithm transformations are affected by the selection of a c value, the large variations in magnitude of estimates for bycatch species should at least merit a review and analysis of the criteria for choosing an appropriate c value. In a review of logarithmic transformations, Berry (1987) suggested choosing a c that normalizes the log-transformed data. He specified an additive function of the skewness and the kurtosis of the data, where skewness and kurtosis are defined as

$$g_1(c) = \sum (y - \bar{y})^3 / (n\hat{\sigma}^3) \text{ and}$$

$$g_2(c) = \sum (y - \bar{y})^4 / (n\hat{\sigma}^4) - 3$$

respectively,

where $\bar{y}$ = the predicted means;
y = the observations; and
$\hat{\sigma}$ = the estimated standard deviation within the defined strata.

When the observations are normally distributed, then the g_1 function has a mean of zero, and the function g_2 has a mean equal to –6/(d+2), where d is the number of degrees of freedom of the error. The additive function of skewness and kurtosis is then defined as

$$g_0(c) = \left|g_{1(c)}\right| + \left|g_2(c) + 6/(d+2)\right|.$$

Thus, the c value that minimizes $g_0(c)$ will make the residuals closer to a sample that follows a normal distribution. Using CPUE values for Spanish mackerel, we evaluated several c values ranging from 1.0E – 8 up to 1.0E + 3. We did not find a minimum solution for $g_0(c)$, but rather an asymptotic behavior with c values less than 0.05, indicating that it is not possible to normalize the Spanish mackerel bycatch data by using a logarithm transformation. Therefore, there is not an objective criterion for selecting a particular c value, and as shown before, even relatively small changes of the c value could cause significant variation of the annual bycatch estimates. Furthermore, independent of the method used to select the c constant in

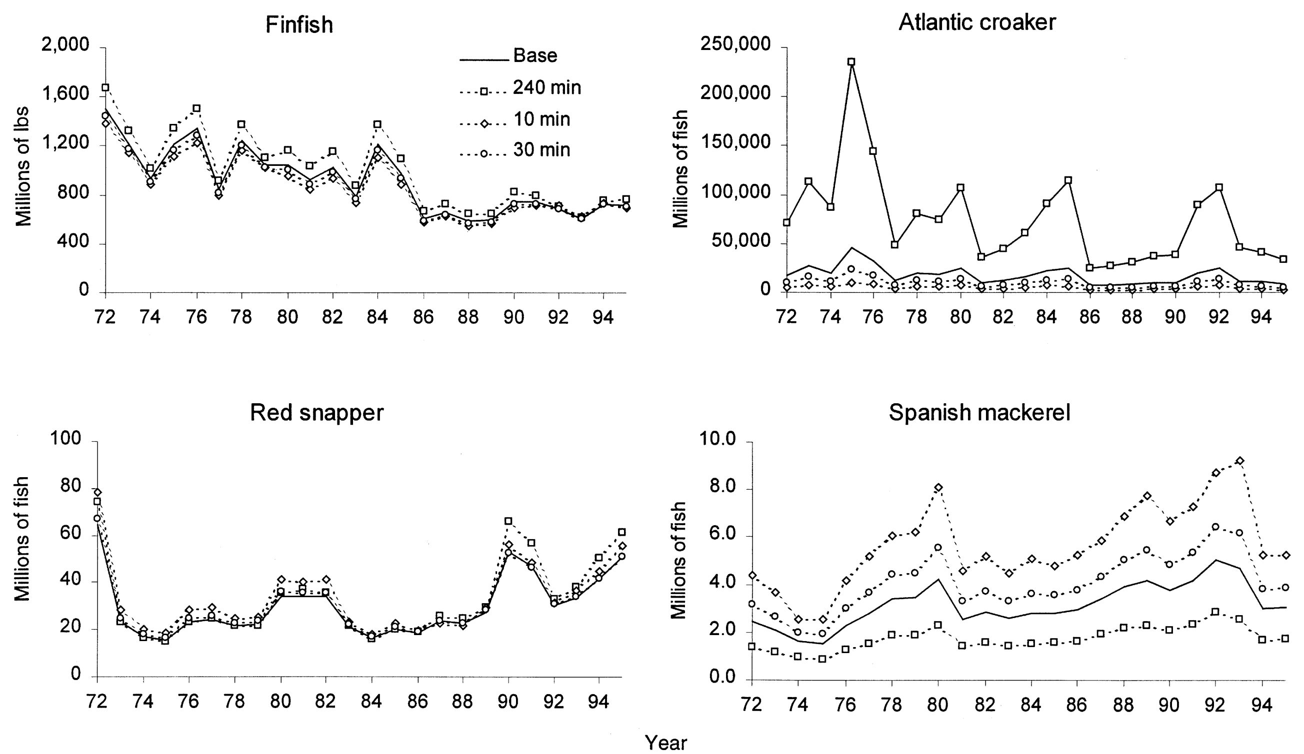

Figure 3

Estimated total bycatch with the general linear model when using different tow-time standard units. The current general linear model uses a one-hour tow time (base scenario).

the logarithmic transformation of the CPUE, the *c* values must be selected for each species independently. Therefore, the same *c* value might not be appropriate for different bycatch species, and if new bycatch data are added, then the *c* value must be re-evaluated.

Standardizing effort in the general linear model

The general linear model predicts bycatch CPUE by cell in units of number of fish caught in one shrimp trawl net per hour. Because actual observations of bycatch are the number of fish caught in a shrimp net during a tow and because tow times are variable, observations are converted to a standard unit of one hour tow time. This standardization procedure implies a direct linear relation between number of fish caught and tow time for all observations (i.e. if 10 fish were caught in a 30-min tow, the CPUE would be 20 fish per hour). However, the average tow time and the tow time distribution from commercial observations are considerably different from those from research observations. Most of the commercial tows range from 1 to 7 hours and have a mode of approximately 4 hours; a few tows are over 12 hours. In contrast, research tows are predominantly of 10-minute duration (73%), and the rest last 1 hour or less.

Given these differences in fishing and sampling time-effort between commercial and research observations, we estimated total bycatch by using different time units to convert the observed catch to CPUE values. We selected 10-, 30-, and 240-minute time units instead of the currently used one-hour unit. These were chosen on the basis of the most frequent tow time for research observations (10 min), the mean tow time of research observations (30 min), and the mode tow time for commercial observations (240 min). The predicted CPUEs were then multiplied by the shrimping effort per cell in the modified time units. Shrimping effort was given in 24-hour-day fishing effort. Therefore, if the predicted CPUE units were 0.5 hour (30 min), the 24-hour shrimping effort would be multiplied by 2. The *c* value was 1.0 for all these calculations.

Modifying the time unit for calculating CPUE values also had an effect on the annual estimates of shrimp bycatch from the general linear model (Fig. 3). Similar to the results of the evaluation of the logarithmic constant, the changes of estimated bycatch were different for each species and varied in the direction of the change. For example, for finfish and Atlantic croaker, a time unit of 10 minutes decreased estimated annual bycatch (5% and 68% on average, respectively). By contrast, red snapper and Spanish mackerel estimated bycatch increased with the 10 minute unit (12% and 78%, respectively). With the commercial mean tow time (240 min), bycatch estimates of Atlantic croaker increased on average 300%, and 10% for finfish. For red snapper, estimates changed only in the most recent years (1990–95) by 20%. In contrast, the estimated bycatch of Spanish mackerel was reduced by 44% on average.

The estimated CPUE should be independent of the time unit used (because it is a constant factor for all observations). However, the differences seen in our study in estimated bycatch were due to the presence of zero CPUE values. By dividing by different time units, the relative distance between the groups of zero CPUE values and the positive CPUE values is changed; as a result, estimators of the central tendency for these data will vary. Although the end results of the time-unit and *c* value choices are similar (biased estimates), their mathematical origin is different. The time-unit choice is a multiplier of the positive catch data (zero catch /any time unit=zero CPUE), whereas the *c* value choice adds the *c* value to all CPUE data. Although a change in time unit could be exactly matched by the appropriate change on the *c* value, addition of more data, with the same time unit, would require recalculating the appropriate *c* value.

Procedure for estimating bycatch with the delta lognormal model

Delta models have been used to analyze fisheries data, in particular when there is a predominant group of zero observations. These models have been used to obtain estimates of abundance for highly aggregated organisms, such as planktonic samples (Pennington, 1983), in the analysis of catch-per-unit-of-effort data for the development of CPUE indices (Lo et al., 1992; Cooke and Lankester[4]), as well as in the analysis of ground trawl surveys to estimate total or relative abundance (Pennington, 1996; Stefánsson, 1996). The main advantage of delta models is that they allow for an explicit and finite probability of zero catch. In a delta model, the estimated values are the product of two independent components: the probability of nonzero observations, and the probability of effective density if there is a positive observation. In the case of fishery surveys, the nonzero probability can be analogous to the probability of encountering a fish aggregation, whereas the probability within the positive observations would correspond to the estimated density of a given fish aggregation (Cooke and Lankester[4]).

Delta models are multivariate distributions with a nonzero probability mass at the origin (Shimizu, 1988). Stefánsson (1996) presented a mathematical model based on a generalized delta lognormal model for analyzing groundfish survey data. This model defines the cumulative density function of abundance at a given sampling station as

$$F_i(\omega) = P[Y_i \leq \omega] = (1 - p_i) + p_i G_i(\omega),$$

where G_i = a continuous cumulative density function describing the distribution of positive values in a station I; and

p_i = the probability of finding fish in that station.

[4] Cooke, J. G., and K. Lankester. 1995. Consideration of statistical models for catch-effort indices for use in tunning VPA's. ICCAT Collect. Vol. Sci. Pap. 45(2):125–131.

If p_i is constant and G_i is a lognormal distribution within a stratum, the function is the delta lognormal model. If p_i is set to one (i.e. excluding zero values), and G_i is set to a gamma or other exponential function with a parameterized mean, this model becomes a generalized linear model (GLiM, Stefánsson, 1996). The advantage of this formulation is that each component in the delta model can be expressed in terms of a GLiM (McCullagh and Nelder, 1989). Thus, the choice of a particular density function in each of the delta model components can be related to other measured variables, such as tow times, location effects, and seasonal or year effects, through assumptions on distribution.

Bycatch data derived from observers in the Gulf of Mexico shrimp trawl fishery typically have a high proportion of zero bycatches and a skewed distribution of the positive bycatch CPUE rates, with a large number of low bycatches and very few large bycatches. The large catches most likely reflect the spatial-temporal distribution characteristics of fish stocks rather than are outliers of the data. This type of distribution is far from normal, and commonly used transformations are unable to make the data comply with the normal assumptions with the classical regression models. Furthermore, in the case of so-called "non-frequent bycatch species," the proportion of zero observations is markedly increased (above 95%); this significantly biases and reduces the efficiency of statistical estimators of central tendency and overestimates the variance (Pennington, 1996).

The delta lognormal model was used in our study to generate annual bycatch estimates for all finfish combined, as well as for three specific finfish species: Atlantic croaker, red snapper, and Spanish mackerel in the U.S. Gulf of Mexico shrimp trawl fishery. Briefly, bycatch CPUE rates of a given fish species in a given cell were estimated as the product of two components: 1) the proportion of tows with positive catch and 2) the mean catch rate if at least one fish was caught. Bycatch per cell is then the product of the estimated CPUE and the corresponding shrimping effort for that particular cell. Total annual bycatch is then the sum over all strata within a year for the commercial component, as in the general linear model (see Eq. 3).

Each component of the delta lognormal model, the proportion of positive tows and the mean bycatch rate, was estimated by following a general linear model approach with the procedure GENMOD in the SAS statistical software package (SAS Institute Inc., 1993). General linear models consist of three elements: 1) the *random* component which defines the error structure of the model, 2) the *systematic* component which defines a set of explanatory variables $x_1, x_2, \ldots, x_q$, and 3) the *link* function which defines the relation between the random and the systematic components (McCullagh and Nelder, 1989). We described the delta lognormal model for estimating shrimp bycatch on the basis of the assumptions entailed with each component of the model. To compare models, the same explanatory variables used in the current general linear model were used with the delta lognormal model.

Proportion of positive tows

The proportion of positive tows for a particular fish species was estimated after classifying each tow as either 0 (no fish caught) or 1 (at least one fish caught). For the shrimp bycatch data, the model assumes that the data are independent results from n successive trials of a Bernoulli-type random variable with a probability p of catching a given fish species. In this case, it is assumed that the frequency distribution of observed zero and positive tows in each cell follows a binomial distribution. The error term is assumed to be constant and independent among the cells. The binomial distribution is then defined in terms of the proportion (y) of positive tows *(r)* to total tows *(n)* per cell, and the probability density function *f(y)* and associated variance *Var (y)* function are given by

$$\text{for } y = r/n,\ f(r) = \binom{n}{r}\mu^r(1-\mu)^{n-r} \text{ where } r = 1, 1, 2, \ldots\ldots, n$$

$$Var(y) = \mu(1-\mu)/n$$

where μ = the mean of y. The response variables y_i are independent for i=1,2,…, n tow trials.

The systematic component defines the set of explanatory variables $x_1, x_2, \ldots, x_q$ which produce a linear predictor η given by

$$\eta = \sum_{j=1}^{q} x_j \beta_j + \beta_0.$$

For the shrimp bycatch data, the linear predictor is a linear function of the fixed explanatory variables dataset, year, season, area, and depth zone, such that

$$\eta = \beta_0 + \beta_1 \cdot dataset + \beta_2 \cdot year + \beta_3 \cdot season + \beta_4 \cdot area + \beta_5 \cdot depth\ zone,$$

where the β_j are parameters to be estimated.

The link function that relates the linear predictor η to the expected value μ of observations y in each cell of the model must be a monotonic differentiable function g such that

$$g(\mu_1) = \eta.$$

In this case, the logit or logistic function expresses the relationship between the assumed binomial error distribution of μ and the given linear function of explanatory variables η, as

$$\eta = \log\left[\frac{\mu}{(1-\mu)}\right].$$

The GENMOD algorithm uses maximum-likelihood estimates for assumed binomial distributions, which are unbiased to a first order of approximation (McCullagh and Nelder, 1989)

Mean bycatch rate

In this section only positive tows were considered. The delta lognormal model assumed that for a given species the number of fish caught as bycatch relates to fixed variables: data source (commercial or research), year, season, area, and depth zone. The mean bycatch CPUE given a nonzero catch was also estimated following a generalized linear model approach. In this case, the random component for the estimated CPUE was assumed to follow a lognormal error distribution within cells. The probability density function is given by the normal function

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)},$$

where μ = $E[y]$; and
σ^2 = $Var(y)$ with a logarithmic link function.

This specification is mathematically equivalent to defining the random component as lognormal with the identity as the link function.

The systematic component is defined as

$$\text{Log}(CPUE) = \beta_0 + \beta_1 \cdot dataset + \beta_2 \cdot year + \beta_3 \cdot season + \beta_4 \cdot area + \beta_5 \cdot depth,$$

where *CPUE* = the catch rate in numbers of fish per net hour for nonzero catches;
β_0 = the overall mean;
dataset = a fixed effect differentiating data sources from commercial shrimp fishing from those in research trawls, the terms *year*, *season, area*, and *depth* are also fixed effects; and
the β_j = parameters to be estimated.

The link function between the random and systematic components is the identity function:

$$\eta = \mu.$$

Estimation of bycatch

The overall model is then referred to as the delta lognormal model. This model generates the estimated proportion of positives tows ($\hat{p}_{ijklm}$) and the mean bycatch rate ($\hat{CPUE}_{ijklm}$) for a given species. Estimates of bycatch are calculated as the product of the proportion of positives tows ($\hat{p}_{ijklm}$) multiplied by the mean bycatch rate ($\hat{CPUE}_{ijklm}$) multiplied by the shrimping effort (f_{jklm}) multiplied by the two nets (assumed) per boat. Shrimping effort data are the same as those used in the current general linear model. Annual estimates of bycatch are simply the sum of bycatch per cell over the season, area and depth zone strata, for the commercial sector (i=1).

$$Bycatch_j = 2 \times \sum_{klm} \hat{p}_{1jklm} \times \hat{CPUE}_{1jklm} \times f_{jklm}.$$

Evaluation of the delta lognormal model

Before comparing the annual bycatch estimates of total finfish, Atlantic croaker, red snapper, and Spanish mackerel from the general linear model (Nichols[5]) and the delta lognormal model, the delta lognormal model was evaluated and assessed. Because there is not yet a formal strategy for model verification, acceptance of a particular model should not be based exclusively on "goodness of fit" scores (McCullagh and Nelder, 1989).

In general, model assessments can be classified into two main groups. The first group checks for systematic departure from the underlying model, testing for additional factors, factor interactions, or covariates that could explain a significant proportion of the residual model variation. The second group involves evaluation of particular or isolated points in the data. McCullagh and Nelder (1989) and O'Brien and Kell[6] have described six specific tests for evaluating generalized linear models: 1) assessment of the scale of independent variables; 2) assessment of the link function adequacy; 3) assessment of variance function adequacy; 4) investigation of systematic departure from the assumed model; 5) investigation of outliers; and, 6) investigation of omitted predictor variables. Most of these analyses are based on the behavior of the model residuals, either as graphical or informal tests, rather than an exact statistical test. For the delta lognormal model, only tests evaluating systematic departure from the assumed model were performed on each of the model components (i.e. an estimation of the proportion of positive tows and the estimation of bycatch rates) separately.

With the delta lognormal model, an evaluation of the proportion of positive tows was restricted to a graphical analysis of the frequency distribution of positive tows of observed and predicted data. This restriction was warranted because most of the tests suggested for assessing model adequacy are uninformative for binomial data (McCullagh and Nelder, 1989; O'Brien and Kell[6]). Figure 4 shows the standardized frequency distributions of proportion of positive tows per cell for the combined finfish category, Atlantic croaker, red snapper, and Spanish mackerel. Each plot shows the observed and the predicted proportions estimated by the binomial distribution of the delta lognormal model. The predicted frequencies fitted closely those observed in all four cases. The assumed binomial distribution is able to predict appropriately the proportion of positive tows in a broad range (from the combined finfish category case where almost all tows were positive [97%] to the case of Spanish mackerel where only 5% of the tows were positive).

The suitability of the delta lognormal general linear model component for the positive tows was evaluated by the following graphical tests: 1) adequacy of the link function, 2) adequacy of the variance function, and 3) systematic departure from the assumed model.

[5] Nichols, S. 1996. Estimates of annual shrimp fleet bycatch in the offshore waters of the Gulf of Mexico. Personal commun. NMFS Pascagoula laboratory. 3209 Frederic St. Pascagoula, MS 39567.

[6] O'Brien, C. M., and L. T. Kell. 1996. The use of generalized linear models for the modelling of catch-effort series. I. Theory. ICCAT Collect. Vol. Sci. Pap. 46(4):476–482.

By plotting the adjusted dependent variable (log CPUE) we were able to assess the link function against the estimated linear predictor ($\hat{\eta}$). A linear configuration is expected for normal, assumed Poisson or gamma error distributions. In our case, the delta lognormal model assumed a normal error distribution for log CPUE of positive catch. Figure 5A shows the plots of the linear predictor (lp-logcp) against the adjusted dependent variable (log CPUE) for red snapper. In the case of high density of points as in Figure 5A, locally weighted regression smoothing procedures (i.e LOESS smoothing) have been suggested for showing the trend of the response variable (McCullagh and Nelder, 1989).

Adequacy of the variance or assumed error distribution function was evaluated by using a plot of residuals against fitted values. The spread of residuals is expected to be approximately constant and independent of the fitted values, confirming the adequacy of the assumed error distribution in the model. Figure 5B shows the plots of residuals (R-logcpu) against the fitted values (P-logcpu) for red snapper. The residuals are evenly distributed about the zero line and are without any apparent trend with respect to the fitted values. Likewise, a plot of residuals versus the normalized cumulative residuals (QQ plot) can be used to assess the variance function adequacy. A linear relationship is expected for residuals from a normal error distribution.

A plot of standardized residuals (rs-logcp) against fitted values (log CPUE) was used to identify possible trends or curvatures that would suggest a departure from the assumed model (Fig. 5C). The null pattern of this plot is a linear configuration of the standardized residuals (O'Brien and Kell[6]). In conclusion, assessments of each of the delta lognormal model components did confirm the model choices and assumptions for the finfish group and the fish species examined (similar plots were created for finfish, Spanish mackerel, and Atlantic croaker but are not presented here for briefness).

As shown before, bycatch estimates from the current general linear model depend upon the standard time unit chosen to convert catches in numbers to CPUE values. Similarly, the same tow time evaluation with the delta lognormal model was performed as with the general linear model. CPUE values were calculated by using 10-, 30- and 240-min tow times, and concurrently, shrimping effort unit, given in hours, were multiplied by a scale factor to make the time unit compatible with the modified CPUE values. With the delta lognormal model, the annual bycatch estimates were exactly the same, independent of the time unit used to calculate the CPUE values, further demonstrating the benefits of using a model that separates the zero catch observations from the positive catch. In addition, delta models do not require adding a constant value to logarithmic transformed values because the estimated density component is restricted to positive catch only, thus avoiding the uncertainty in selecting a *c* value to log transform CPUE values as required in the general linear model.

Results and discussion

Because the bycatch database complied with the delta lognormal model specifications, a stepwise analysis of deviance was performed to assess the importance of the factors selected in the delta model. Table 2 gives the percent change in deviance as each factor is added to the binomial fitted proportion of the zero versus positive tows component of the delta lognormal model. The deviance explained by the model is equivalent to the r^2 concept in linear models (McCullagh and Nelder, 1989; Stefánsson, 1996). Tests of significance were based on the χ^2 statistic for the binomial distribution of the proportion of positive tows (McCullagh and Nelder, 1989). Overall, the delta lognormal model, with all factors, explained between 55% and 75% of the total deviance for the finfish group and the three fish species. However, as expected, the percentage of deviance explained by each factor differed for each species. For example, the *dataset* factor appeared to be unimportant in estimating the proportion of positive tows for red snapper and Atlantic croaker. Instead, area and season factors were more important for red snapper, and area and depth zone for Atlantic croaker.

Table 3 shows the lognormal component of the delta model r^2 values, sum of squares error or residual deviance, residual degrees of freedom, and the P values. Similarly to the proportion of positive tows, a stepwise analysis of the r^2 shows that dataset, year, season, area, and depth zone are significant factors in explaining the overall variability of the model. An exception is the depth zone factor in estimating bycatch CPUE rates for red snapper. The delta lognormal estimated density model explained from 17% (Atlantic croaker) to 36% (Spanish mackerel) of the total variation, indicating that a significant portion of the bycatch CPUE variability is still unexplained by the model.

The annual shrimp bycatch estimates for the four species groups in the U.S. Gulf of Mexico differed in several aspects between the delta lognormal model and the current general linear model. Results varied for the finfish group and the fish species analyzed. Differences were found both in the absolute magnitude of bycatch estimates and in the trend over the time series 1972–95. For the total finfish bycatch, the delta lognormal model estimated an average of 795 million lbs. for the period 1972–95, or 14% lower than the equivalent general linear model estimate of 916

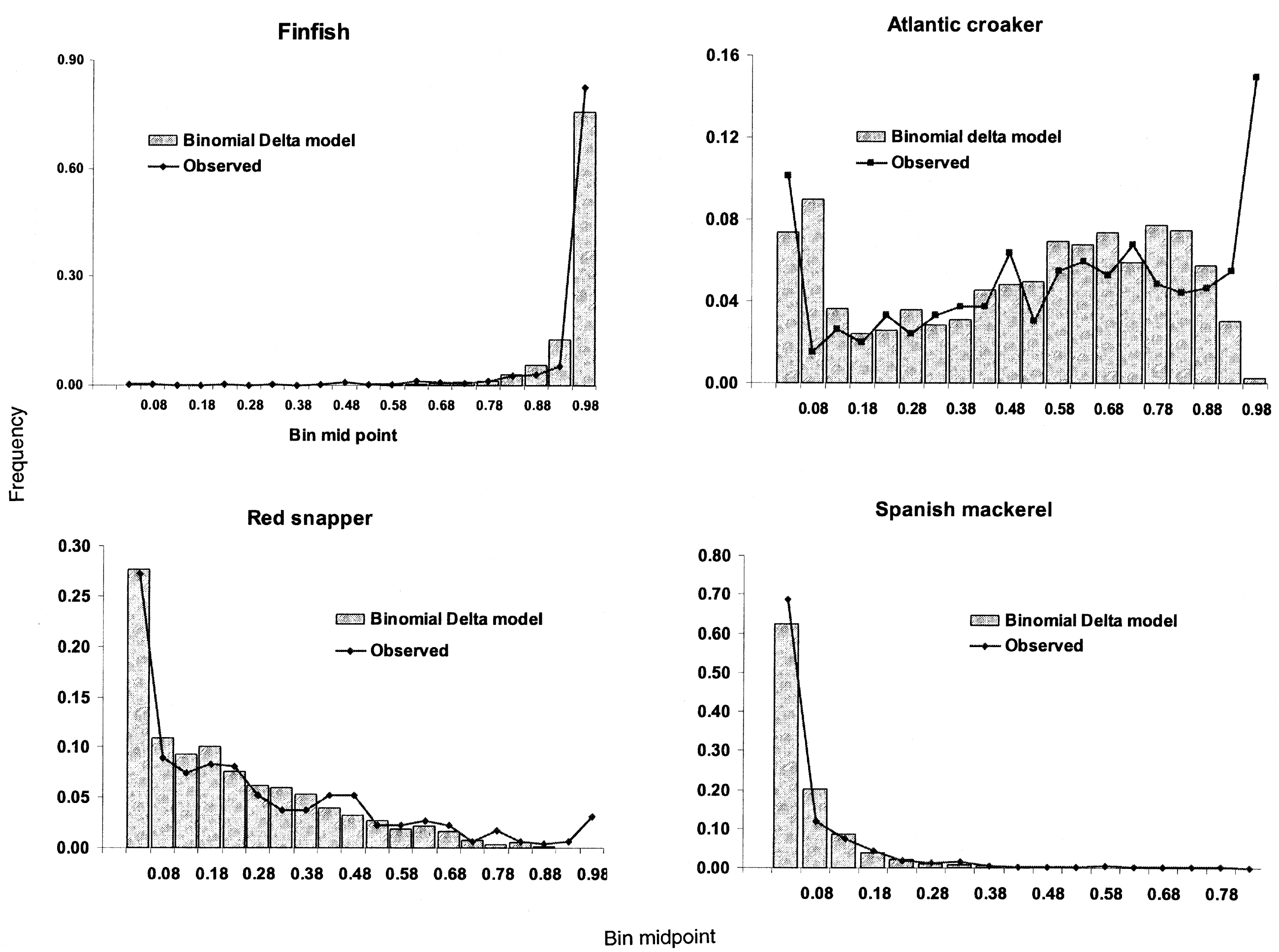

Figure 4

Standardized frequency distribution of the proportion of positive tows from the bycatch data 1972–95 and the proportions estimated by the binomial-based delta lognormal model component.

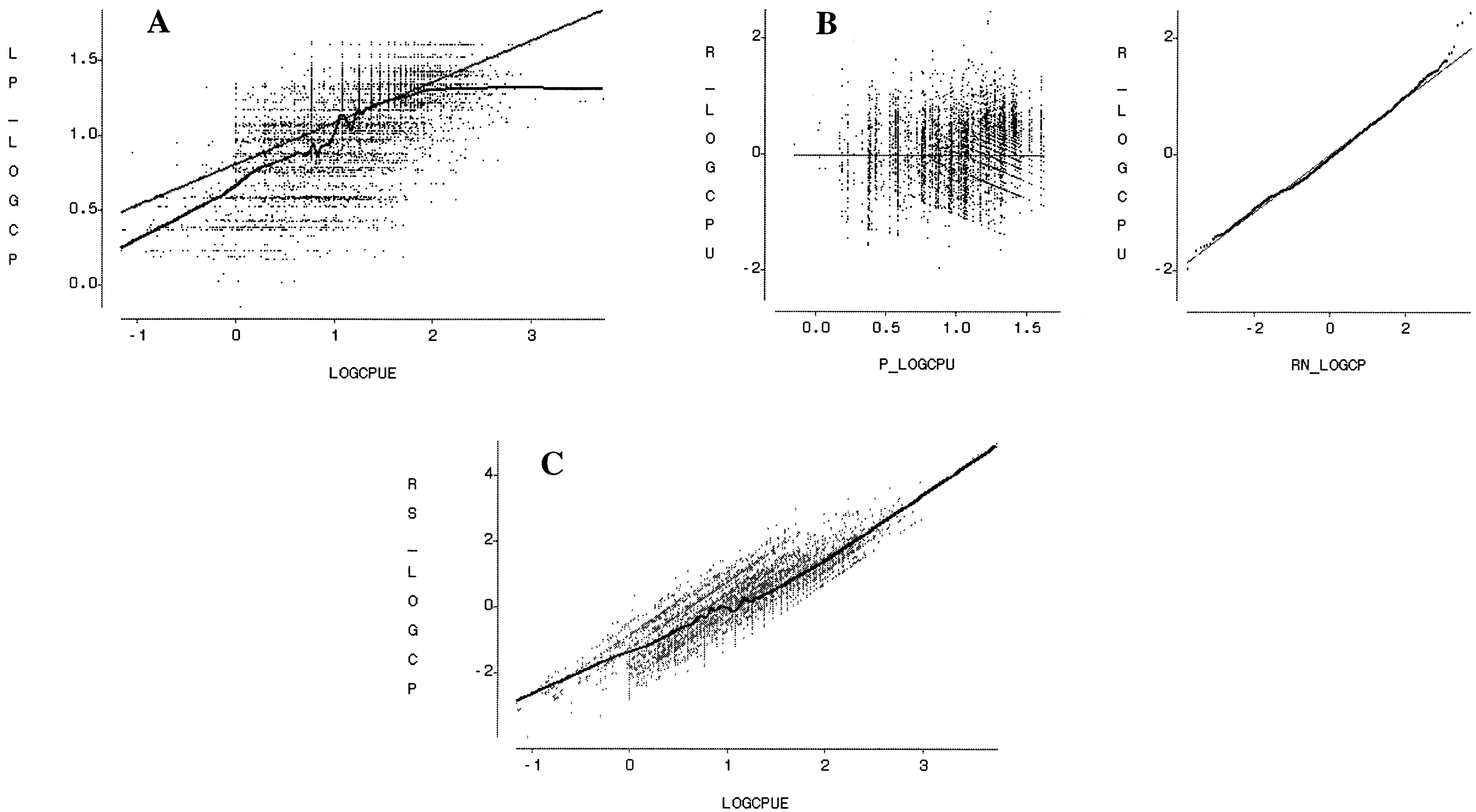

Figure 5

Evaluation plots for the positive catch component for red snapper with the delta lognormal model. (**A**) Plot of the linear predictor (lp_logcp) against the dependent variable (logCPUE). The straight line is the linear function of the variables for comparison purposes, the dark line is the LOESS smoothing procedure of the data. (**B**) Plot of the residuals (r_logcpu) against predicted values (p_logcpu) and the cumulative normalized-residuals or QQ plot. (**C**) Plot of the standardized residuals (rs_logcp) against the fitted values (logCPUE). The line represents the LOESS smoothing procedure to evaluate the data trend.

Table 2

Analysis of deviance table for different binomial-based delta lognormal models fitted to positive/total tows of each fish species and finfish category in the bycatch database for the U.S. Gulf of Mexico 1972–95. The models are fitted sequentially and the columns give the residual degrees of freedom for each model, the residual deviance, the resulting change in deviance, the percentage of total deviance change, and the *P*-value when χ^2 test was used for significance. Model 1 refers to estimating only the overall mean.

Model factors	Residual df	Residual deviance	Change in deviance	% of total deviance	*P*	Residul deviance	Change in deviance	% of total deviance	*P*
	Finfish					Atlantic croaker			
1	454	2403				8869			
Data set	453	1676	726.05	0.30	<0.001	8866	3.38	0.00	<0.0660
Data set + year	430	1113	563.93	0.23	<0.001	7890	976.05	0.11	<0.001
Data set + year + season	428	1079	33.70	0.01	<0.001	6194	1696.32	0.19	<0.001
Data set + year + season + area	425	1026	52.80	0.02	<0.001	3866	2327.89	0.26	<0.001
Data set + year + season + area + depth zone	424	1016	9.66	0.00	<0.001	3849	17.22	0.00	<0.001
	Red snapper					Spanish mackerel			
1	454	6390				2356			
Data set	453	6389	0.34	0.00	<0.5605	2086	270.00	0.11	<0.001
Data set + year	430	4935	1454.20	0.23	<0.001	1810	275.91	0.12	<0.001
Data set + year + season	428	4013	921.97	0.14	<0.001	1550	260.29	0.11	<0.001
Data set + year + season + area	425	2770	1243.33	0.19	<0.001	1468	81.92	0.03	<0.001
Data set + year + season + area + depth zone	424	1633	1136.46	0.18	<0.001	1065	402.92	0.17	<0.001

million lbs. (Table 4, Fig. 6). Total finfish bycatch estimates from the delta lognormal model were consistently lower for all years, and overall followed the same trend as the estimates from the current general linear model. The normalized plot of total finfish bycatch (i.e. year estimate minus the mean divided by the standard deviation of the time period) shows that the trends are identical between the two models up to 1990, but in 1991–95 some discrepancies were observed (Fig. 7). However, both models show a decreasing trend in the total finfish bycatch estimates from about 1,100 million lbs. (1972–84) to less than 700 million lbs. during the last 10 years (1985–95). This decline can be attributed to improvements in the selectivity of the shrimp trawl gear to retain less bycatch (i.e. introduction of TEDs and BRDs) or to an overall reduction of the trawlable fish stock biomass in the U.S. Gulf of Mexico.

For Atlantic croaker, bycatch estimates from the delta lognormal model were on average 4.5 billion fish for the 1972–95 period, 74.5% lower than estimates of 17.7 billion fish from the general linear model (Table 4, Fig. 6). Once more, the normalized plot shows a similar decreasing trend in bycatch estimates from both models in the 1972–95 period (Fig. 7). Atlantic croaker, together with longspine porgy (*Stenotomus caprinus*), are the most common finfish bycatch species in the Gulf of Mexico shrimp fishery, therefore a significant reduction in bycatch estimates of Atlantic croaker most likely correlates with a reduction in total estimated finfish bycatch.

Bycatch estimates of red snapper from the delta lognormal model were slightly greater in general from 1972 to 1982, and much lower from 1987 to 1995 compared with estimates yielded with the general linear model (Fig. 6). On average, the delta lognormal model bycatch estimates were 22.1 million fish for the years 1987–95, 40% lower than the equivalent average of 36.8 million fish estimated with the general linear model (Table 4). The normalized plot shows that since 1987, there has been an overall increasing trend in red snapper bycatch according to both the general linear model and delta model estimates, a peak in bycatch in 1990, subsequent low in 1992, and an increasing trend since then (Fig. 7). Prior to 1987, red snapper bycatch was relatively lower, with an exception of the highest bycatch peak in 1972 and some above average bycatch in 1980–82.

Delta lognormal estimates of Spanish mackerel bycatch were 97% higher on average than those from the general linear model (Fig. 6, Table 4) for the time period 1972–95. Spanish mackerel bycatch estimated by the delta lognormal model was on average 6.5 million fish, compared with 3.2 million fish estimated by the general linear model. In

Table 3

Analysis of deviance table for different lognormal-based delta models fitted to the positive bycatch CPUEs of each finfish species and the finfish category in the bycatch database for the US Gulf of Mexico 1972–95. The models were fitted sequentially and the columns give the residual degrees of freedom for each model, the residual deviance, the resulting change of deviance, the r^2 values, and the P-value when the F-test was used for significance. Model 1 refers to estimating only the overall mean.

Model factors	Residual df	Residual deviance	Change in deviance	% of total deviance	P	Residual df	Residul deviance	Change in deviance	% of total deviance	P
	Finfish					Atlantic croaker				
1	25,636	10,040.40				15,985	16,036			
Data set	25,635	8870.06	1170.34	0.12	<0.0001	15,984	15,238	798.03	0.05	<0.0001
Data set + year	25,612	7809.32	1060.74	0.22	<0.0001	15,961	13,855	1382.60	0.14	<0.0001
Data set + year + season	25,610	7685.07	124.25	0.23	<0.0001	15,959	13,790	65.09	0.14	<0.0001
Data set + year + season + area	25,607	7440.94	244.13	0.26	<0.0001	15,956	13,463	326.60	0.16	<0.0001
Data set + year + season + area + depth zone	25,606	7436.59	4.35	0.26	<0.0001	15,955	13,308	154.96	0.17	<0.0001
	Red snapper					Spanish mackerel				
1	7377	2491.32				1240	430.90			
Data set	7376	2122.94	368.38	0.148	<0.0001	1239	369.80	61.10	0.142	<0.0001
Data set + year	7353	1889.23	233.71	0.242	<0.0001	1216	329.53	40.27	0.235	<0.0001
Data set + year + season	7351	1847.03	42.20	0.259	<0.0001	1214	305.09	24.43	0.292	<0.0001
Data set + year + season + area	7348	1803.48	43.55	0.276	<0.0001	1211	288.32	16.77	0.331	<0.0001
Data set + year + season + area + depth zone	7347	1803.26	0.22	0.276	<0.0001	1210	273.54	14.78	0.352	<0.0001

our study, the delta lognormal model showed a larger year variability of bycatch with prominent peaks in 1980 and 1992. The normalized plot of Spanish mackerel bycatch illustrates that estimates from the general linear model and delta lognormal model followed similar trends from 1972 to 1981, and from 1988 to 1995 (Fig. 7). The time period from 1984 to 1987, the period of greatest oscillation in bycatch estimates for the delta lognormal model, corresponds with the years of no bycatch observations in the commercial fishery. Although delta lognormal bycatch estimates show a comparable trend to the general linear model estimates in the later years (1987–95), the magnitude of bycatch is much greater; the peak estimate of 14.4 million fish in 1993 is twice as high as the reported estimates from the general linear model (Nichols[5]).

The delta lognormal model protocol appears to be an improved alternative procedure for estimating shrimp bycatch in the U.S. Gulf of Mexico compared with the currently used general linear model. In theory, the delta model allows for an explicit probability for zero catches, which are highly common in the bycatch data set, especially when dealing within single species cases. Myers and Pepin (1990) stated that lognormal-based estimators are sensitive to violations of model assumptions, in particular if the number of observations is below 40 or if there is no confirmation that the sample came from a true lognormal distribution (or if both situations occur). However, their arguments are restricted to the positive tows (i.e. nonzero observations); they concluded that lognormal estimators should be used only in cases where the assumed lognormal distribution can be confirmed. Following their criteria, Myers and Pepin's arguments should then be applied to the delta lognormal model (more specifically to the density estimation component that models the nonzero catches) and to the current general linear model as well (if a lognormal distribution can be assumed for all observations, Nichols et al.[2]). In the bycatch database, there are a large number of cells with low number of observations (i.e. ≤40). Restricting the database to cases where the number of observations per stratum (year, area, season, depth, and dataset) were greater than 40, we were able to use cumulative CPUE distributions more approximate to lognormal

Table 4

Bycatch estimates from the general linear model and the delta lognormal model. Percentage of change is with reference to the bycatch general linear model estimates, negative percentages refer to lower estimates. The average values are over the 24-year period.

	Finfish			Atlantic croaker			Red snapper			Spanish mackerel		
	Millions of lbs.			Millions of fish			Millions of fish			Millions of fish		
Year	GLM	Delta	% change	GLM	Delta	% change	GLM	Delta	% change	GLM	Delta	% change
1972	1501.10	1305.45	–13	17,529.94	4736.80	–73	69.49	73.27	5	2.47	5.63	128
1973	1211.12	1093.34	–10	27,161.33	9034.68	–67	23.00	28.22	23	2.08	3.62	74
1974	934.85	826.66	–12	20,205.60	5396.10	–73	16.97	19.58	15	1.62	1.65	2
1975	1209.90	1098.10	–9	45,615.42	7337.83	–84	15.23	18.01	18	1.53	2.35	54
1976	1343.26	1177.58	–12	32,140.84	7806.93	–76	23.27	30.27	30	2.32	5.63	143
1977	843.11	772.07	–8	12,793.05	4405.55	–66	24.45	25.41	4	2.80	8.22	194
1978	1248.09	1113.57	–11	20,133.40	6648.36	–67	21.62	18.40	–15	3.43	6.27	83
1979	1045.06	957.10	–8	18,851.25	5121.00	–73	22.36	22.91	2	3.48	8.64	148
1980	1045.81	925.60	–11	24,707.77	5860.63	–76	34.07	38.35	13	4.24	16.93	299
1981	922.37	787.92	–15	10,431.83	4727.13	–55	34.21	37.99	11	2.57	5.94	131
1982	1028.24	878.77	–15	11,953.52	4264.06	–64	33.77	36.31	8	2.85	6.94	144
1983	790.33	680.57	–14	15,826.07	5940.60	–62	21.18	17.97	–15	2.58	3.08	19
1984	1217.03	1043.25	–14	22,381.82	7291.54	–67	16.44	13.57	–17	2.79	6.89	147
1985	975.74	821.60	–16	24,975.37	4558.15	–82	20.15	14.68	–27	2.79	2.17	–22
1986	606.40	513.85	–15	7453.91	2134.62	–71	18.80	14.31	–24	2.95	4.11	39
1987	656.50	556.19	–15	7778.19	1281.58	–84	23.88	11.99	–50	3.42	2.33	–32
1988	582.70	498.09	–15	8601.77	1732.06	–80	22.69	11.72	–48	3.94	8.33	111
1989	594.09	507.50	–15	10,286.57	2800.51	–73	27.51	18.10	–34	4.20	8.38	100
1990	748.97	639.65	–15	10,370.38	2414.03	–77	53.17	32.35	–39	3.77	6.64	76
1991	742.31	597.00	–20	20,449.99	3775.18	–82	46.93	27.03	–42	4.19	6.97	66
1992	684.74	430.25	–37	24,818.83	5298.03	–79	30.37	17.06	–44	5.05	10.74	113
1993	605.72	517.55	–15	11,556.16	1998.04	–83	33.71	18.77	–44	4.68	14.41	208
1994	729.20	660.36	–9	10,984.66	2177.96	–80	41.98	32.32	–23	3.01	4.11	37
1995	719.92	669.17	–7	8715.51	1500.90	–83	50.87	29.94	–41	3.06	5.32	74
Average	916.11	794.63	–14	17,738.47	4510.09	–74	30.26	25.36	–14	3.16	6.47	97

density function in the case of nonzero catches (i.e. delta lognormal model) than in the case when both zero and positive catches are included (i.e current general linear model). Thus, if departures from the assumed distribution produced biased lognormal estimates, certainly the current general linear model would be more prone to these biases than the delta lognormal model.

As stated by Pennington (1991) in his response to Myers and Pepin's (1991) article, the assumed lognormal data were contaminated with data from distributions that generated extremely small values, close to zero, which in a logarithmic scale become large negative values. These large negative values then biased estimates of the mean. In the case of the bycatch database this is not a problem because the smallest positive bycatch CPUE values are in most cases greater than 0.05.

Another point to consider when comparing the delta lognormal model and the current general linear model is the variance associated with the estimated bycatch. Smith (1988) described an exact variance for the delta lognormal distribution estimates. He also pointed out that the efficiency of the delta lognormal variance is a function of the sample size, the proportion of zero observations, and the variance within the nonzero observations. The variance of bycatch estimates are, however, restricted to the variance from the general linear model or the delta lognormal model because the shrimping effort multiplier is assumed to be exactly known (Nichols et al.[2]) . Thus to compare true standard errors of bycatch estimates, one would require the variance of the shrimping effort and calculate an overall variance through a mathematical approach such as the delta method or use resampling techniques such as bootstrapping procedures. Because point estimates of bycatch are more frequently used in stock assessments of affected species rather than the confidence intervals, the present analysis focused on the point estimates of bycatch.

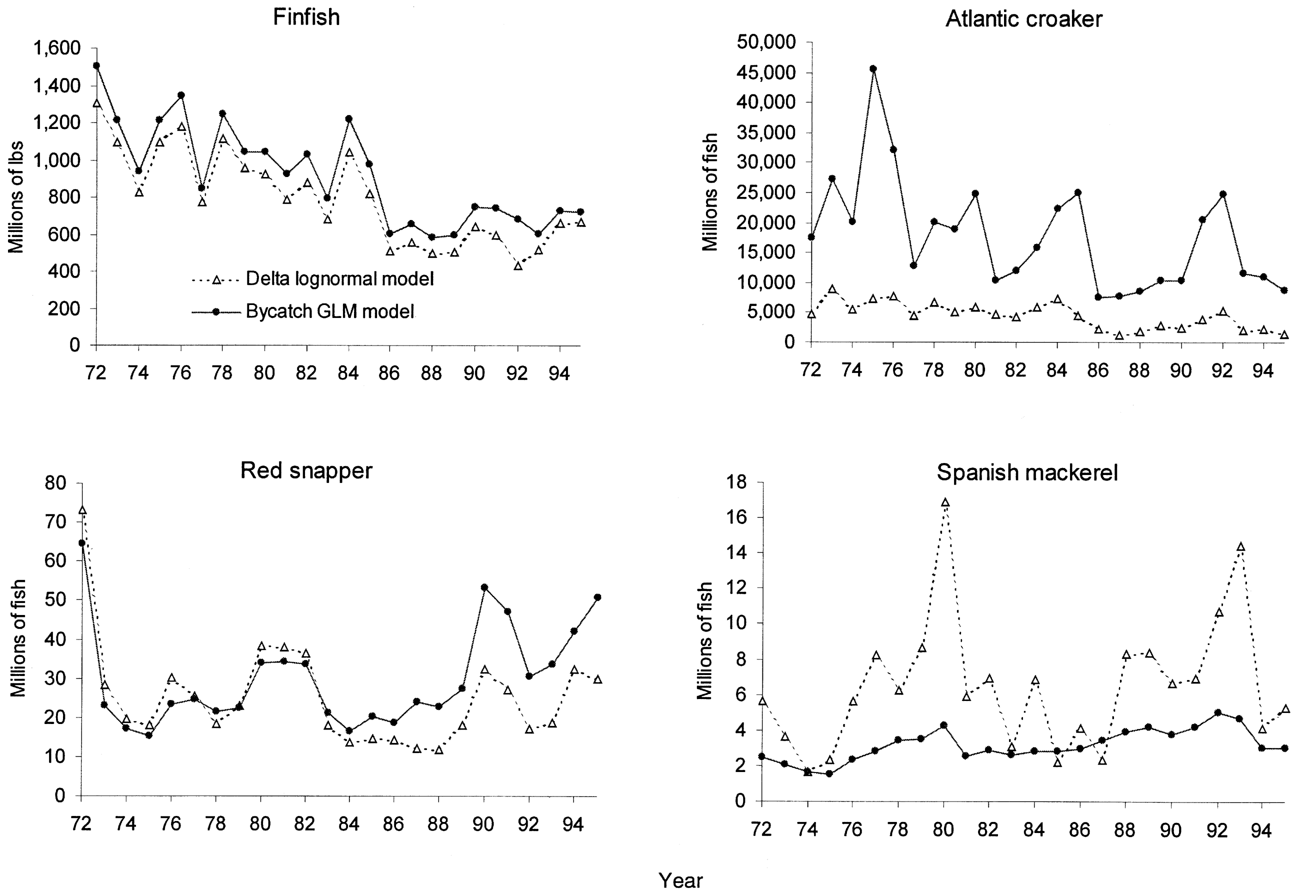

Figure 6

Estimates of total annual bycatch from the delta lognormal model and the current general linear model, for finfish (millions of pounds), Atlantic croacker, red snapper, and Spanish mackerel (millions of fish) 1972–1995.

Conclusions and recommendations

Analyses of the total finfish bycatch and the bycatch of Atlantic croaker, red snapper, and Spanish mackerel show that the delta lognormal model estimates differ both in magnitude and trends from those generated by the current general linear model. However, these differences are not consistent among species. In terms of absolute magnitude, they are substantially different for Atlantic croaker and Spanish mackerel over all years (1972–95), whereas for red snapper differences are greater in the most recent years of the time series (1987–95). Total finfish bycatch estimates are more similar in magnitude and trend for both models. Although the trends of bycatch in the time series from 1972 to 1995 are similar for the species examined, the absolute estimated values are highly variable. Because these estimates are included as additional catch (usually for age 0 and 1) in the stock assessments of directed fisheries, the uncertainty of the bycatch estimates will impact the results of these assessments. Further, this uncertainty will extend to management policies adopted from these assessment results for species like Spanish mackerel, king mackerel, and red snapper (Ehrhardt and Legault, 1997; Goodyear[7]).

As presented before, the general linear model estimates depend on choices about the constant added to the CPUE values prior to logarithmic transformation and on the standard time unit chosen for calculating CPUE values. These problems emerge from the noncompliance of the bycatch data with the assumptions associated with the general linear model. In particular, the observed CPUEs are not lognormally distributed owing to the significant proportion of zero observations within the data. In contrast, the delta lognormal model conforms better with the structure of the data and avoids the problems of choosing a *c* value for catches in the logarithm transformation and of selecting a standard time unit for the CPUE calculations. As

[7] Goodyear, C. P. 1995. Red snapper in U.S. waters of the Gulf of Mexico. Contribution report MIA 95/96-05, 171 p. Miami Laboratory, Southeast Fisheries Science Center, NMFS, NOAA, 75 Virginia Beach Dr., Miami, FL 33149.

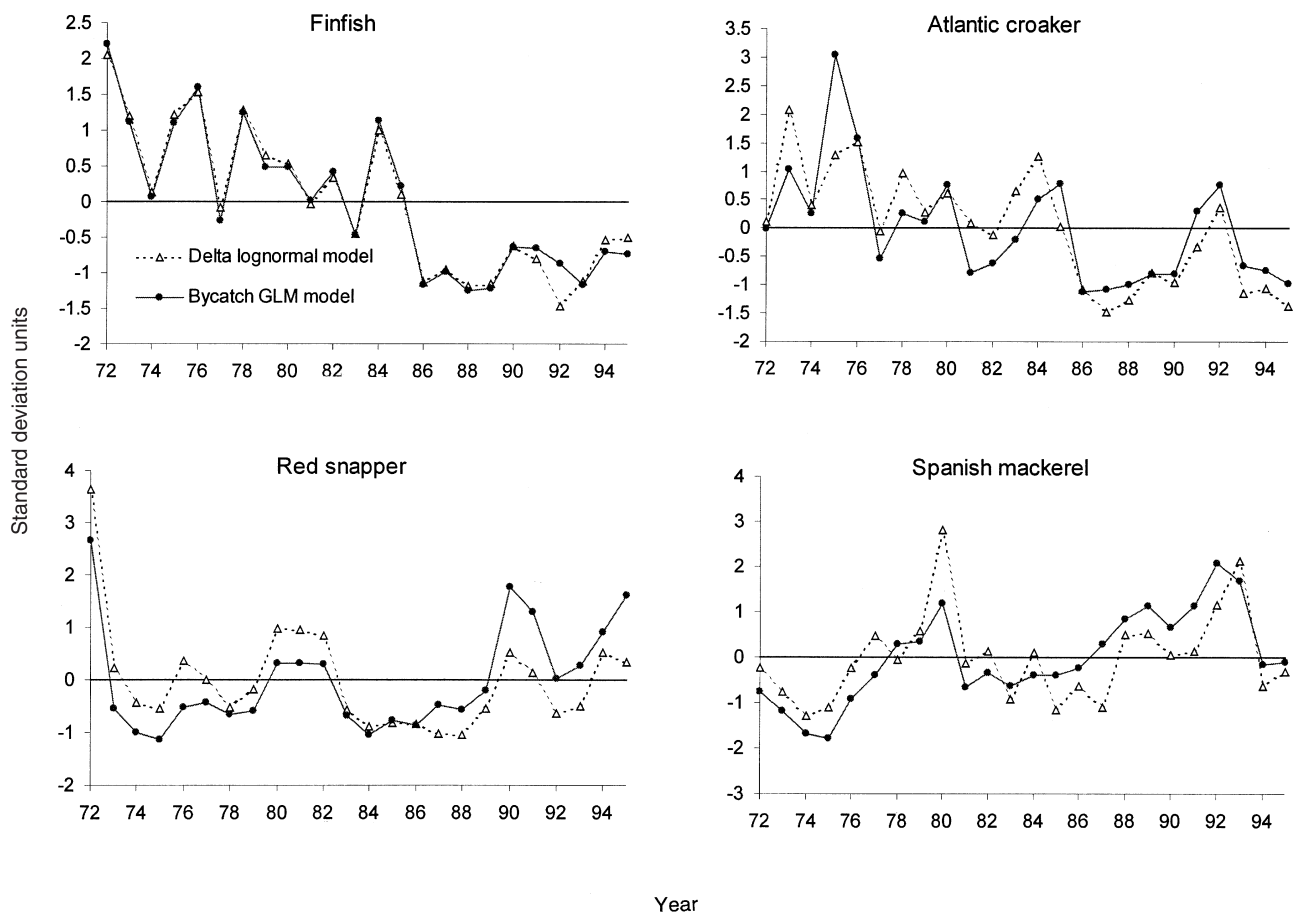

Figure 7

Normalized plots of bycatch estimates by species from the current general linear model and the delta lognormal model.

expected, both models agree better in the case of the finfish bycatch estimates where the proportion of zero CPUE values is the lowest (less than 3%).

Besides the problems related to zero observations, several other considerations must be addressed to generate annual bycatch estimates:

1 The matrix structure of area, season, year, and data source is inadequately covered by observations. This is true for any model that uses these same factors but in particular for the period 1985–90, when commercial observations were not available. It may be beneficial to limit the analysis to years, areas, and seasons where there are data from both commercial and research sources. This change, however, will require the redefinition of the objectives of the bycatch estimation procedure because the estimated annual bycatch will not be possible for the 1972–95 period.

2 Another important requirement is the standardization of the CPUE units for both the commercial and research observations. We feel that these CPUEs represent different units for each type of observation for each particular species and that a single linear relationship is not adequate. This standardization will require a thorough analysis of each fleet and additional information in order to convert the effort units from nominal to effective units for each fleet prior to bycatch estimation. It has been suggested that the more recent data obtained by the Bycatch Characterization Project (NOAA[1]) could be used for this type of analysis. As an alternative, we modified the delta lognormal model to incorporate the observed catch (i.e. numbers of fish) as the dependent variable, and we used the tow time (i.e. hours fishing) as a covariate in the systematic linear component of the delta lognormal model. With this modification, the total deviance explained by the model increased for red snapper. However, we would recommend standardizing the nominal CPUE instead of simply adding more variables to an already unbalanced matrix and avoid considering only goodness-of-fit as an indicator.

3 In the analysis of bycatch by species, it is presently assumed that estimated bycatch in number of fish belongs

to the same age class, usually the age-0 class. This may not be true for some species. Thus, bycatch estimates should take into account number of fish per age or size class.

Acknowledgments

This research was supported by a grant provided by MARFIN (NA57FF283-01); additional grants came from the Florida Sea Grant College Project (R/LR-8-37). We thank Victor Restrepo from the University of Miami RSMAS for his scientific advice, Scott Nichols from the Pascagoula laboratory of the NMFS for providing the bycatch database and algorithms, and Nancy Cummings from the NMFS Miami laboratory for her comments and suggestions. Three anonymous reviewers provided valuable comments on the final draft.

Literature cited

Berry, D. A.
1987. Logarithmic transformations in ANOVA. Biometrics 43:439–456.
Ehrhardt, N. M., and C. M. Legault.
1997. The role of uncertainty in fish stock assessment and management: a case study of the Spanish mackerel, *Scomberomorus maculatus*, in the US Gulf of Mexico. Fish. Res. 29:145–158.
Henwood, T. A., and W. E. Stuntz.
1987. Analysis of sea turtle captures and mortalities during commercial shrimp trawling. Fish. Bull. 85(4):813–17.
Lo, N. C., L. D. Jacobson, and J. L. Squire.
1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can. J. Fish. Aquat. Sci. 49:2515–26.
McCullagh, P., and J. A. Nelder.
1989. Generalized linear models. Chapman & Hall, London, 512 p.
Myers, R. A., and P. Pepin.
1990. The robustness of lognormal-based estimators of abundance. Biometrics 46:1185–1192.
1991. Replay to: on testing the robustness of lognormal based estimators. Biometrics 47:1623–1624.
Pellegrin, G. J.
1982. Fish discards from the southeastern United States shrimp fishery: fish by-catch: bonus from the sea report of a technical consultation on shrimp by-catch utilization held in Georgetown, Guyana, 27–30 October 1981. International Development Research Center (IDRC), Ottawa, 163 p.
Pennington, M.
1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics 39:281–86.
1991. On testing the robustness of lognormal-based abundance estimators. Biometrics 47:1623.
1996. Estimating the mean and variance from highly skewed marine data. Fish. Bull. 94:498–505.
SAS Institute Inc.
1993. SAS/STAT® software: the GENMOD procedure, release 6.09. SAS® Technical Report P-243. SAS Institute Inc., Cary, NC, 88 p.
Shimizu, K.
1988. Point estimation. lognormal distributions: theory and applications. Department of Information Sciences, Faculty of Science and Technology, Science University of Tokyo, Japan Edwin L. Crow, and Kunio Shimizu. Marcel Dekker, Inc., New York, NY, 387 p.
Smith, S. J.
1988. Evaluating the efficiency of the Δ-distribution mean estimator. Biometrics 44:485–493.
Stefánsson, G.
1996. Analysis of groundfish survey abundance data: combining the general linear model and delta approaches. ICES J. Mar. Sci. 53:577–88.