Available online at www.sciencedirect.com

SCIENCE DIRECT®

Fisheries Research 75 (2005) 175–192

FISHERIES RESEARCH

www.elsevier.com/locate/fishres

Analysis of logbook accuracy for blue marlin (*Makaira nigricans*) in the Hawaii-based longline fishery with a generalized additive model and commercial sales data

William A. Walsh[a,*], Russell Y. Ito[b], Kurt E. Kawamoto[b], Marti McCracken[b]

[a] *Joint Institute for Marine and Atmospheric Research, University of Hawaii, 2570 Dole Street, Honolulu, HI 96822-2396, USA*
[b] *Pacific Islands Fisheries Science Center, NOAA Fisheries, 2570 Dole Street, Honolulu, HI 96822-2396, USA*

Received 16 June 2004; received in revised form 3 November 2004; accepted 20 November 2004

Abstract

Blue marlin, *Makaira nigricans*, catch (number per set) in the Hawaii-based longline fishery from March 1994 through June 2002 was analyzed by integrated use of observer reports, commercial logbooks, and auction sales records. The objective was to provide a corrected catch history for this species in this fishery during this period. The study was conducted because official statistics compiled from the logbooks are known to be biased by billfish (Istiophoridae) misidentifications. The initial step entailed fitting a generalized additive model (GAM) of blue marlin catch to environmental and operational data gathered by fishery observers during 8397 longline sets deployed by commercial vessels. The GAM included nine significant predictors and explained 41.1% of the deviance of observed blue marlin catches. The GAM coefficients were then applied to the corresponding predictors in the logbook reports from unobserved sets to estimate catches in evaluations of the accuracy of data from unobserved sets ($N = 87\,277$ longline sets on 8437 trips; 95.4% of unobserved effort). This was done by regressing the logbook catch data on the predictions, using the residuals to identify trips with systematic misidentifications, and then checking their logbooks against sales records from the public fish auction in Honolulu. The large majority of the misidentifications consisted of striped marlin, *Tetrapturus audax*, reported as blue marlin, with lesser numbers of shortbill spearfish, *T. angustirostris*, logged as blue marlin, and blue marlin logged as either striped marlin or black marlin, *M. indica*. An estimate obtained by use of the GAM and observer data indicated that the nominal catch of blue marlin was inflated by 29.4%. The 95% prediction limits about the GAM-generated estimate (34 201–41 507 blue marlin) did not include the catch total from the logbooks (48 911 blue marlin). The corrections also refined understanding of the distribution of blue marlin by reducing the impression that large numbers of blue marlin are sometimes caught north of Hawaii in the autumn and early winter months. There was no evidence of widespread underreporting of marlins. We conclude that this study significantly improved the accuracy of logbook data for blue marlin and should also contribute to improved understanding of the ecology and distribution of blue marlin. We infer that self-reporting could yield accurate marlins catch data if species identifications were improved because there was no apparent

* Corresponding author. Tel.: +1 808 983 5346; fax: +1 808 983 2902.
E-mail address: William.Walsh@noaa.gov (W.A. Walsh).

0165-7836/$ – see front matter
doi:10.1016/j.fishres.2004.11.007

underreporting problem. Finally, we recommend that logbook data accuracy receive serious attention in the context of stock assessments.

Keywords: Blue marlin; Generalized additive model; Corrected catch history; Fishery observer data; Fish auction data; Logbook accuracy

1. Introduction

The Hawaii-based longline fishery takes five istiophorid billfishes (blue marlin (*Makaira nigricans*), black marlin (*M. indica*), striped marlin (*Tetrapturus audax*), shortbill spearfish (*T. angustirostris*), and sailfish (*Istiophorus platypterus*)), most often as incidental catch in tuna-targeted fishing (Ito and Machado, 2001) or until recent years in mixed species assemblages targeted en masse (He et al., 1997). These species typically comprise ca. 5–10% of the total annual catch of this fishery (Ito and Machado, 2001), are of moderate commercial value, and are ecologically important epipelagic predators. Blue marlin is also a very highly prized recreational species. In addition to such general importance, however, the first four of these species represent a major monitoring challenge for the Pacific Islands Fisheries Science Center (PIFSC) of NOAA Fisheries. Misidentifications caused by several superficial similarities (e.g., similar dorsal and ventral coloration in immature blue and full-grown adult striped marlin; maximum body length and mass of black marlin approach those of blue marlin) (Nakamura, 2001) are known to be present in the federally mandated daily longline logbook reports (Dollar, 1992). Although these billfishes are generally reported with reasonable accuracy in the aggregate, the fractions reported for the individual species are at times inaccurate (Walsh, 2000), which is reflected by the official fishery statistics compiled from the logbooks and which may also affect population estimates generated by stock assessment models that use the official data.

The objective of this paper was to correct the blue marlin catch for the Hawaii-based longline fishery from March 1994 through June 2002, which necessarily entailed examination of the catch reports for the other istiophorid species as well. This objective was met by fitting a generalized additive model (GAM) of blue marlin catch to environmental and operational data gathered by fishery observers (Walsh and Kleiber, 2001) and then applying its coefficients fishery-wide to the logbook reports from unobserved trips to generate predictions as comparison standards (Walsh et al., 2002). Walsh et al. (2002) employed these methods to estimate underreporting and characterize reporting biases with blue shark, *Prionace glauca*, a predominant bycatch species in this fishery, which is sometimes caught in very high numbers. In this study, GAM-generated results were used to assess the accuracy of billfish identifications, which required the capability to detect bias at much lower levels of catch and with much smaller discrepancies between reported and predicted values than had previously been the case. We attained this capability by using an independent data source, sales records from the public fish auction that serves as the principal outlet for the landings of this fleet, to verify analytical results. As such, this paper demonstrates use of these methods when the major factor affecting data quality is the accuracy of identifications among species that are superficially similar but not particularly numerous.

Results presented herein include a GAM of blue marlin catch in the Hawaii-based longline fishery from March 1994 through June 2002, a corrected catch history for this period obtained by application of the GAM to logbook data, overall and annual estimates of reporting error, and identification of sources of reporting error. As such, this paper presents an integrated use of observer, logbook, and commercial sales data with a species of economic, ecological, and recreational importance.

2. Materials and methods

2.1. GAM development

The premise underlying this study was that a data set of known quality (fishery observer records) could be used to develop a statistical model that could in turn be used as a comparison standard for blue marlin catch on unobserved longline sets. The specific intention was to model blue marlin catch in terms of a relatively small suite of variables present in both the observer and logbook reports, as in Walsh and Kleiber (2001) and Walsh et al. (2002), which therefore represented a basis for comparison.

A GAM of blue marlin catch was developed by forward selection using a cubic splines algorithm (Walsh et al., 2002). Detailed discussions of GAM theory and methodology are presented by Hastie (1992), Venables and Ripley (1994), and Schimek and Turlach (2000). The GAM can be expressed as:

$$\log(\mu) = \sum_{j=1}^{p} S_j(x_j, d_j), \quad (1)$$

where μ was the conditional mean catch for the set of predictors (x_1, x_2, ..., x_p), S_j an unspecified smooth function, and d_j was the degrees of freedom of the smoother. While the model was additive, the non-parametric form of the functions S_j made it flexible. As the degrees of freedom in a GAM increase, the function S_j gains more flexibility and can become 'rougher', which allows more hills and valleys or other complex shapes to be exhibited. Because the GAM fitting procedure does not accommodate missing predictor values, it was necessary to delete 6.5% of the observed sets that had one or more missing predictor value(s) and an additional 0.4% of the sets with extreme predictor values that were highly influential. One trip was deleted because the observer was incapacitated, and an additional 2.1% of the sets were deleted on the basis of comparisons to fish auction sales records (see Section 3, below). This left 8397 sets, or 90.9% of the observer data, to fit the GAM. The underlying probability distribution of catch per set was assumed to be an overdispersed Poisson distribution; logarithms were the link function.

The forward selection procedure involved evaluating each candidate predictor at each fitting stage and entering into the model the predictor that minimized the Akaike Information Criterion (AIC) and residual deviance. Sequential F-tests were used to ensure that all terms were significant, with a $P < 0.05$ entry criterion. Standard residuals and partial residuals plots were used in model evaluation. The fit of the model was depicted by a time series plot of the monthly mean catch rates reported by the observers with the corresponding monthly mean GAM-corrected catch rates. The effects of individual predictors were depicted by traces with a standardized ordinate. GAM development and all other statistical procedures described herein were conducted in S-PLUS Version 6.1.2 (Insightful Corp., 2002).

2.2. Data sources

2.2.1. Model development data

The Hawaii Longline Observer Program (HLLOP) was established in March 1994 to monitor interactions between the longline fishery and sea turtles (DiNardo, 1993). In addition to monitoring protected species interactions, the observers also record species-specific tallies of the catch and a large suite of environmental (e.g., sea state and weather conditions) and operational (e.g., geographic position, number of hooks, set and haul times, and type of bait) details from each set (i.e., gear deployment) (Pacific Islands Regional Office, 2003). In this study, eight operational variables (the date of fishing, hook numbers, latitude, longitude, vessel length (ft), begin-set time, catches of bigeye tuna (*Thunnus obesus*), and yellowfin tuna (*T. albacares*)) were evaluated as candidate predictors, along with the lunar phase and weekly mean sea surface temperature (°C) measured by the advanced very high resolution radiometer mounted aboard a satellite operated by the National Environmental Satellite, Data, and Information Service (NESDIS) of NOAA. Each predictor was allotted 30 non-linear degrees of freedom.

The data used to fit the GAM were gathered during the 100-month period between March 1994 and June 2002, with 9242 sets deployed on 774 observed trips. The observer coverage rate through 1999 was approximately 5%, followed by increases to 11% in 2000 and 23% in 2001.

2.2.2. Fishery monitoring data

Federally mandated commercial logbook reports have been collected and archived in their original and electronic forms at the PIFSC since November 1990. A report (i.e., one logbook page) is required for each longline set, which normally corresponds to one fishing day. The reports provide species-specific tallies of the catch (i.e., kept fish + released fish) and several operational parameters that are also recorded by the observers (date of fishing, latitude, longitude, hook numbers, and catches of bigeye tuna and yellowfin tuna).

The fleet deployed 91 452 sets on 8473 unobserved trips during the study period, equivalent to 10 974 sets on 1017 trips per year. Of the total, 3.6% of the sets had one or more missing predictor value(s), which precluded application of the GAM coefficients (see below). It was necessary to delete another 1.0% of the sets because their predictor ranges exceeded those in the observer data. The resulting sample size for the initial application of the GAM was 87 277 sets, equivalent to 95.4% of the unobserved longline effort.

2.2.3. Verification data

Sales records (i.e., numbers and pounds sold, and dollar value by species) from the public fish auction conducted by the United Fishing Agency Ltd. (UFA), Honolulu, have been provided electronically to the Hawaii Division of Aquatic Resources (HDAR) since January 2000. HDAR, in turn, provides these data to the PIFSC. Before 2000, PIFSC or HDAR personnel attended the auction twice weekly (out of six auction working days) to gather data. During the transition from directly recorded to electronically provided records, paired tests documented close agreement in species identifications, numbers and weights of fish, and dollar values (Walsh, unpublished data). Sales records were used to verify species identifications when errors were suspected in either the observer or logbook data sets, in which case the numbers of fish logged as kept on any particular trip were compared to the numbers sold. The UFA data were considered definitive because their personnel are very experienced, because price differences among species demand careful identifications, because the buyers are also present and represent a second check, and because identification of the fish is much easier on display than at sea. When auction sales records were not available, possible misidentifications were checked by comparing logbook records to monthly sales receipts submitted by fishermen to HDAR. These were, however, sometimes less useful than UFA records because many captains in this fleet change vessels frequently, so their receipts may represent the yields of multiple trips.

2.3. GAM application, error identification, and error correction

The fitted GAM coefficients were applied to the unobserved logbook sets with the ‘predict.gam’ function in S-PLUS, followed by checks against the archived original logbook forms or auction sales data to detect likely misidentifications. The process was initiated by transforming reported and predicted blue marlin catches per set to $\log_e(X+1)$, regressing the transformed logbook catches on the corresponding GAM predictions, and then using the studentized residuals (SR) (Cook and Weisberg, 1982; Draper and Smith, 1981; Hoaglin et al., 1983) in a multi-stage data editing procedure analogous to that in Walsh et al. (2002).

Two of the editing stages were based upon objective statistical criteria. The first entailed identifying those trips with two or more “large” SR (i.e., $SR \geq |2|$). Because the SR approximates the *t*-distribution, this represented a conservative ($P \approx 0.0025$) examination standard. The second examination criterion was the occurrence on any trip of even one set with an $SR \geq |3|$, which would represent an even more rigorous standard ($P \approx 0.0015$). The expectation in the latter case was that such large absolute values might reflect recording, transcription, or similar errors.

The objective statistical criteria were then followed by five additional subjective, expertise-based examinations. This stage was limited in scope by the availability of sales data before 2000. The first of these entailed assessing all trips since 2000 undertaken by vessels with five or more trips with systematic discrepancies from UFA data. The second, used in the absence of sales data, was predicated upon the circumstantial evidence of vessel history and seasonality. Trips with three or more large SR from October through March by vessels with three or more previously identified systematic discrepancies were also judged to be in error. This criterion was established because preliminary results had indicated that trips in the summer and early autumn months with large positive SR generally reflected high blue marlin catches, whereas those with multiple large positive SR from October through March generally reflected misidentifications. The third of these criteria, predicated upon the known rarity of black marlin in this fishery, entailed checks on trips with three or more black marlin and any large SR not previously evaluated. This was followed by checks on trips by vessels that had logged 50 or more black marlin during the study period and exhibited any within-vessel patterns of misidentification. Finally, trips with combined totals for blue marlin, striped marlin, and shortbill spearfish that represented shortfalls ≥ 10 relative to the GAM-predicted totals for blue marlin were also checked to detect possible under- or non-reporting.

In cases of apparent systematic misidentifications, logbook trip totals for blue marlin and other billfishes were corrected in proportion to sales records (if available). Many trips were characterized by complete misidentifications (e.g., all billfish logged as blue marlin but all sales reported as striped marlin), but when this was not the case, marlin numbers were changed for selected sets to yield the estimated correct totals. It was not possible to ascertain whether the corrected sets were the actual source of error under such circumstances, but any associated inaccuracy was expected to be minor because most corrections involved substituting smaller numbers (usually zeroes) for questionable values (i.e., multiple blue marlin per set). When sales records were not available, trip totals were corrected in proportion to the mean fleet-wide ratio of striped marlin:blue marlin caught from October through March (3.5:1) after deleting results from those vessels previously determined to be associated with patterns of systematic misidentifications. Also, when numbers of fish kept exceeded those sold, as when some of the catch is discarded, traded, consumed aboard the vessel, or remains unsold, the corrections were intended to approximate blue marlin as accurately as possible, so the shortfall was usually reflected by the corrections applied to the other species.

The GAM coefficients were then applied to this corrected data set. The regression of logbook catches per set on GAM predictions was assessed in terms of its coefficient of determination, parameter estimates, *F*-test, variance estimate, and residuals plots. The predictions were summed to obtain overall and annual catch estimates. Reporting error was estimated by use of bootstrap methods to generate 95% prediction intervals (95% PI) about the overall and annual catch estimates as described in Walsh et al. (2002) and McCracken (2004). The intent of the bootstrap was to mimic the error structure of the original data using non-parametric resampling of the standardized Pearson residuals from an undersmoothed curve and then adding these residuals to the fits of an oversmoothed curve to generate new catches. The GAM was then refitted to the generated data and total catches recomputed. This process was replicated 999 times. Prediction intervals were then computed based on the standardized prediction errors of the generated totals. It should be noted that the annual PI were individual rather than simultaneous.

Temporal trends were presented as the monthly mean catches per set and catch totals computed from the uncorrected logbook data, corrected logbook data, and GAM predictions. The effects of the data corrections were depicted by plotting catches per set on 1° squares after deleting those with less than three permits fishing during each period in question to conform with data confidentiality requirements.

3. Results

3.1. Observer coverage and fishing effort

Observer effort (Table 1) underwent major changes in both allocation among trip types and coverage rates during the study period. Swordfish, *Xiphias gladius*, trips were the focus of observer coverage in 1994 because these were expected to have the highest sea turtle interaction rates. In 1995, however, observer allocation was revised to approximate fleet-wide effort more closely. Coverage levels and allocation among trip types then remained roughly constant through 1999, with an average of 36% of the active vessels carrying an observer at least once per year. An increase in coverage that began in 2000 permitted observer coverage aboard 55% of the active vessels that year, 93% in 2001, and 92% in the first half of 2002.

Fleet-wide fishing effort also changed markedly from 1994 to 2002. Swordfish-directed effort decreased considerably in 1994–1995 in response to declining catch rates (Ito and Coan, 1999). Tuna-targeted effort dominated fleet-wide activity thereafter, comprising 64% of all sets from 1995 to 2000. The remaining swordfish-targeted activity was closed in 2000 to minimize longline interactions with sea turtles (National Marine Fisheries Service, 2003). Consequently, nearly all fishing targeted tunas in 2001 and 2002.

3.1.1. Nominal catch statistics

The most important aspect of the uncorrected billfish catch statistics for this study (Table 2) was that logbook reports apparently listed excess blue marlin. The blue marlin catches reported by observers were less than those from the logbooks on observed sets, which in turn were less than those from the unob-

Table 1
Summary of effort in the Hawaii-based longline fishery from March 1994 through June 2002

Year	Vessels	Trips	Sets	Set types SF	Set types MS	Set types T	Vessels	Trips	Sets	Set types SF	Set types MS	Set types T
	Observer (all data)						Logbook (all data)					
1994–2002	143	774	9242	6.4	15.3	78.4	173	8473	91 452	10.5	22.0	67.5
	Observer (data used in GAM development)						Logbook (data used in GAM application)					
1994–2002	143	724	8397	6.8	16.6	76.6	173	8417	87 277	10.4	22.2	67.5
1994	45	47	484	48.1	17.4	34.5	121	803	7757	40.0	12.4	47.6
1995	43	47	526	13.1	36.1	50.8	110	1079	11 045	15.0	24.5	60.4
1996	47	52	617	9.4	43.1	47.5	104	1047	10 929	9.6	31.4	59.1
1997	33	37	461	11.5	46.2	42.3	105	1083	11 319	8.8	26.3	64.9
1998	40	47	542	11.3	32.7	56.0	115	1092	11 776	9.3	26.7	64.0
1999	36	39	430	12.8	28.6	58.6	120	1098	12 260	5.9	26.4	67.7
2000	69	107	1233	2.8	20.0	77.2	124	988	9622	4.9	27.3	67.8
2001	94	213	2297	0.4	4.0	95.6	101	803	8573	0.0	3.2	96.8
2002	89	156	1807	0.0	0.0	100.0	97	424	3996	0.0	0.4	99.6

Entries are numbers of active vessels, trips and sets, and percentages of effort by set type (SF, swordfish; MS, mixed species; T, tuna).

served sets. This pattern, observed in all trip types, differed from the expected, which is observer ≥ logbook (observed) ≥ logbook (unobserved) (Walsh, 2000).

The uncorrected billfish catch statistics exhibited four other noteworthy features. The first was that blue marlin, striped marlin, and shortbill spearfish catches and CPUE were lowest on swordfish trips. Second, blue marlin catches per set were similar on tuna- and mixed species-directed trips, but CPUE in the latter category was approximately double that in the former. This reflected that mixed species-directed fishing from 1994 to 1996 employed techniques (means: 849 hooks; 5.7 hooks per float) similar to those for swordfish (means: 828 hooks; 4.1 hooks per float), but in

Table 2
Summary of billfish catches (uncorrected) from March 1994 through June 2002 by the Hawaii-based longline fishery

Species	All set types: Source	Catches	Percentage[a] of total	Catch/set	CPUE	Tuna sets: Catch/set	CPUE	Mixed sets: Catch/set	CPUE	Swordfish sets: Catch/set	CPUE
Blue marlin	Observer	3768	0.9	0.41	0.30	0.43	0.27	0.44	0.52	0.11	0.13
	Logbook (O)	4544	1.5	0.49	0.37	0.50	0.32	0.58	0.69	0.16	0.19
	Logbook (U)	46 762	1.6	0.51	0.46	0.49	0.35	0.70	0.86	0.27	0.31
Black marlin	Observer	67	<0.1	0.01	0.01	0.01	<0.01	0.01	0.02	<0.01	<0.01
	Logbook (O)	260	0.1	0.03	0.02	0.03	0.02	0.02	0.03	0.01	0.02
	Logbook (U)	5620	0.2	0.06	0.05	0.07	0.04	0.06	0.07	0.03	0.03
Shortbill spearfish	Observer	7401	1.7	0.80	0.48	0.96	0.55	0.27	0.29	0.10	0.13
	Logbook (O)	6544	2.1	0.71	0.42	0.86	0.48	0.23	0.25	0.05	0.07
	Logbook (U)	68 161	2.3	0.75	0.47	1.02	0.60	0.25	0.26	0.05	0.06
Striped marlin	Observer	11 207	2.6	1.21	0.81	1.32	0.76	0.97	1.13	0.50	0.63
	Logbook (O)	9739	3.1	1.05	0.70	1.16	0.67	0.76	0.89	0.42	0.54
	Logbook (U)	107 259	3.7	1.17	0.88	1.39	0.87	0.86	1.04	0.44	0.56

"Observer" and "Logbook (O)" refer to corresponding paired results from observed sets ($N = 9242$). "Logbook (U)" refers to data from unobserved sets ($N = 91\,452$). Catches and catch per unit effort (CPUE; i.e., fish per 1000 hooks) are presented pooled and by set types.

[a] Total catch (observer reports): 429 725 fish; total catch (logbooks on observed sets): 310 719 fish; total catch (logbooks on unobserved sets): 2 934 617 fish.

different locales, as indicated by mean latitudes and sea surface temperatures (mixed species: 26°N, 23.4°C; swordfish: 30.3°N, 20.4°C). In contrast, higher mean numbers of hooks per set were deployed when targeting tunas (1698), which reduced nominal CPUE. Third, the higher mean catch rates for striped marlin and shortbill spearfish in logbooks from unobserved rather than observed trips reflected disproportional observer coverage relative to fleet-wide effort. Specifically, there was relatively low observer coverage of tuna-directed fishing in 1995 when 20.5% of the nominal striped marlin catch was harvested. Finally, black marlin were apparently also overreported, and probably to a far greater extent proportionally than blue marlin.

3.2. Blue marlin catch on observed sets

A preliminary data evaluation that utilized UFA data revealed marlin misidentification problems in observer reports from at least 2.1%, in logbooks from at least 3.2%, and in both observer and logbook reports from at least 1.2% of the observed trips. For example, in the former case, the observers reported 134 kept blue marlin and 118 kept striped marlin, whereas the UFA data listed 232 striped marlin and 32 blue marlin sold, respectively. Among the 16 observers involved, half made the errors on their first or second trip and two others did so on their third trip after previously reporting only 3 and 10 marlins, respectively. There were no instances of repeated problems with marlin identifications on the part of individual observers. In contrast, logbook reports from two vessels exhibited patterns of misidentifications twice and those from two others did so three times. The GAM was fitted after deleting trips with identifiable patterns of misidentifications, and the relationship between logbook (Y) and observer (X) catch records was expressed by:

$$\log_e(Y+1) = 0.0668 + 0.7901 \log_e(X+1) + \in (R^2 = 0.512; N = 8397) \qquad (2)$$

3.3. GAM of blue marlin catch

The GAM (Table 3) included nine significant predictors and explained 41.1% of the deviance of blue marlin catch. The first three entries (Fig. 1) yielded 84% of its explanatory power; the first six accounted for 95%. A pronounced seasonality was apparent (Fig. 1a), with the annual maximum each year in summer and early autumn. Begin-set time (Fig. 1b) was essentially a proxy for trip types; 99% of all sets deployed between 05:00 and 10:00 h targeted tunas, whereas 98% of the swordfish-targeted sets, with their low blue marlin catches, were deployed between 16:00 and 21:00 h. There was a generally positive relationship between catch and sea surface temperature from about 16 to 30 °C, although the curvature in the predictor trace was more pronounced below than above 23 °C (Fig. 1c). Blue marlin catches also increased with longitude proceeding westward from low and highly variable levels east of Hawaii. Latitudinal effects consisted primar-

Table 3
Analysis of deviance of blue marlin catches per set

Predictor variable	ΔAIC	ΔResidual deviance	d.f.	F_{enter}	P
Date of fishing	2199.48	2257.38	28.9	13.336	0
Begin-set time	823.33	879.79	29.0	8.062	0
SST (°C)	600.65	661.50	28.8	3.773	0
Longitude	143.14	202.64	28.8	2.921	2.8×10^{-7}
Latitude	100.09	159.85	28.8	4.968	0
Hooks/set	57.54	115.97	28.1	4.284	0
Vessel length	27.69	87.77	29.0	2.557	8.7×10^{-6}
Yellowfin tuna/set	15.98	77.01	28.5	2.705	2.5×10^{-6}
Bigeye tuna/set	4.46	63.97	28.8	1.911	0.002

The reductions in the Akaike Information Criterion (AIC) and residual deviance, degrees of freedom, F-test and associated significance are presented for each term.
Null deviance = 10 958.46; d.f. = 8396.
Residual deviance = 6452.58; d.f. = 8128.32.
Pseudo-R^2 = (10 958.46−6452.58)/10 958.46 = 0.411.

ily of declining catches approaching the northern and southern limits of the fishery. Relatively low catches were associated with the low hook numbers deployed on swordfish sets. Blue marlin catch also varied significantly in relation to vessel lengths from 40 to 94 ft, and catches of yellowfin tuna, *Thunnus albacares*, from 0 to 96 per set, and bigeye tuna, *T. obesus*, from 0 to 98 per set. The effects of the last three predictors were essentially fluctuations with no clear pattern, which probably reflected overfitting.

The predictors were characterized by several significant correlations. The strongest were the inverse relationships between latitude and sea surface temperature ($r = -0.796$; $P = 0$) and hook numbers and begin-set time ($r = -0.669$; $P = 0$).

The fit of the model (Fig. 2) is depicted by the monthly mean blue marlin catch reported by the observers and the corresponding corrected values. The GAM-corrected values were greater than the observer-reported catches in mid-1995, mid-1996, and very noticeably in mid-1997. The latter case reflected a low coverage rate (two trips; 1.1% of deployed sets) and possibly two other factors. First, sea surface temperature was in the top quartile on 15 of the 16 sets. Second, although both trips were categorized as tuna-directed, the numbers of hooks per float (4–5) and the ratio of yellowfin tuna:bigeye tuna caught (22:1) indicated that these vessels were fishing near the surface, targeting the former species. Thereafter, the GAM-corrected monthly means generally tracked the mean catches per

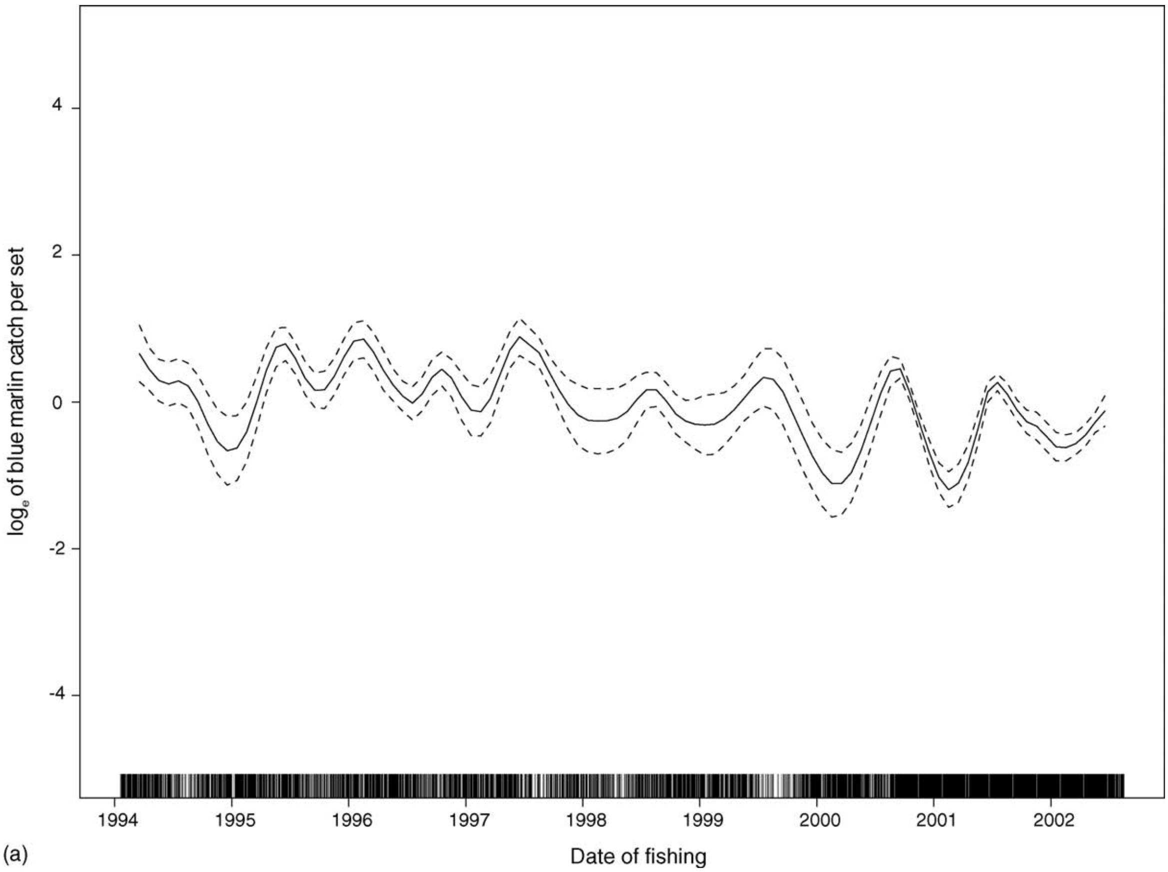

Fig. 1. (a) The relationship between the date of fishing and the natural logarithm of blue marlin catch as estimated by a nine-variable GAM. (b) The relationship between begin-set time and the natural logarithm of blue marlin catch as estimated by a nine-variable GAM. (c) The relationship between sea surface temperature and the natural logarithm of blue marlin catch as estimated by a nine-variable GAM.

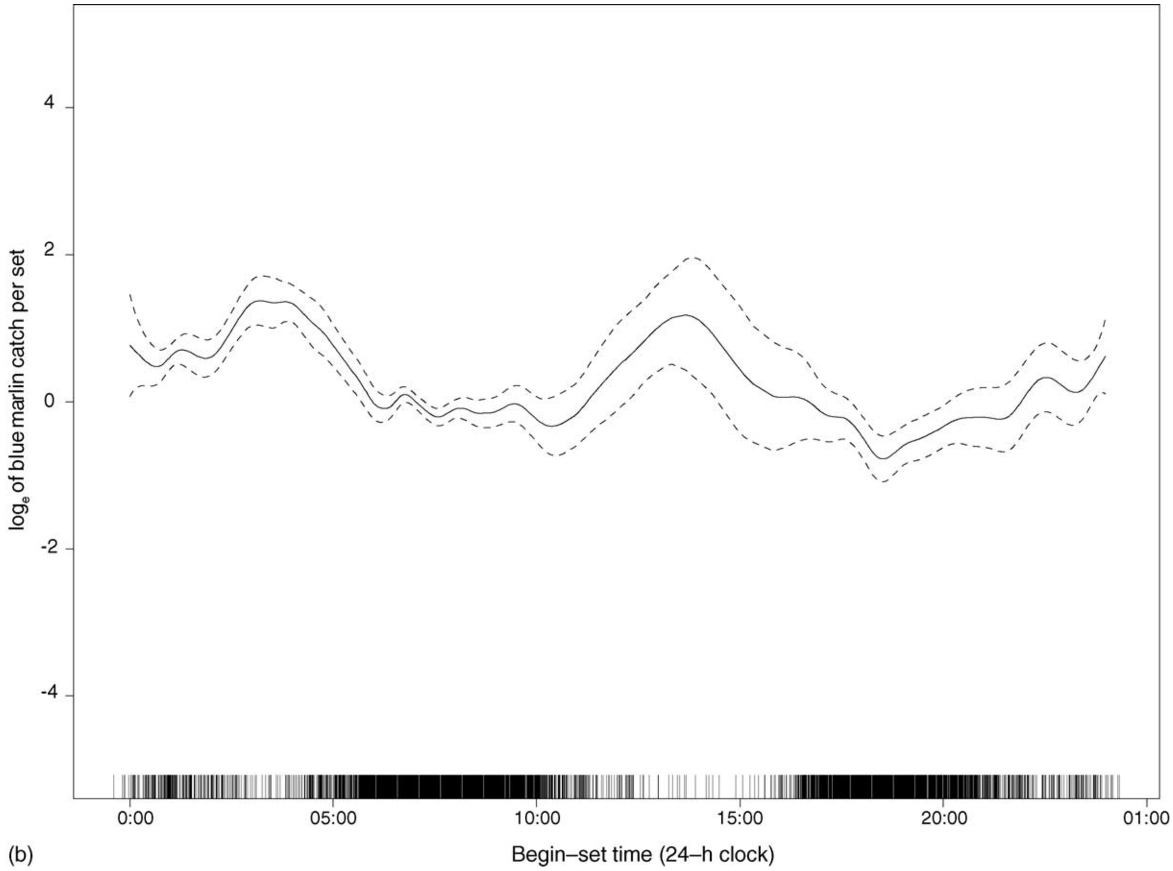

Fig. 1. (*Continued*)

set from the observers closely, which indicated that the model fit well and that use of the GAM coefficients to generate predictions for evaluation of unobserved trips was justifiable.

3.4. Application of the GAM to blue marlin catch from unobserved trips

The initial application of the GAM coefficients to the unobserved sets was summarized by:

$$\log_e(Y+1) = 0.0693 + 0.6265\log_e(X+1) + \in \quad (3)$$

where Y represents the logbook value for catch and X represents the GAM prediction ($F_{1,87\,275} = 13\,530$; $P = 0$; $R^2 = 0.134$; $s^2{}_{y*x} = 0.2051$). There were 5361 large SR, corresponding to 6.1% of the unobserved sets. The preponderance (96.5%) was positive, reflecting higher reported than predicted catches, such as occurred when other species were misidentified as blue marlin.

Detailed evaluation (Table 4) of 541 trips, with 45.5% of the large SR, revealed the species composition and magnitude of the misidentifications. The checks that concentrated on seemingly anomalous blue marlin catches primarily detected misidentifications of striped marlin and some shortbill spearfish and reduced the estimated total for kept blue marlin by 19.9%. Of these corrections, 96% resulted from direct checks against auction data; only 4% were predicated upon the circumstantial evidence of seasonality and vessel history. In contrast, evaluations of seemingly anomalous black marlin reporting revealed two distinct error patterns while reducing the blue marlin total slightly. The first detected excess blue and black marlin with short-

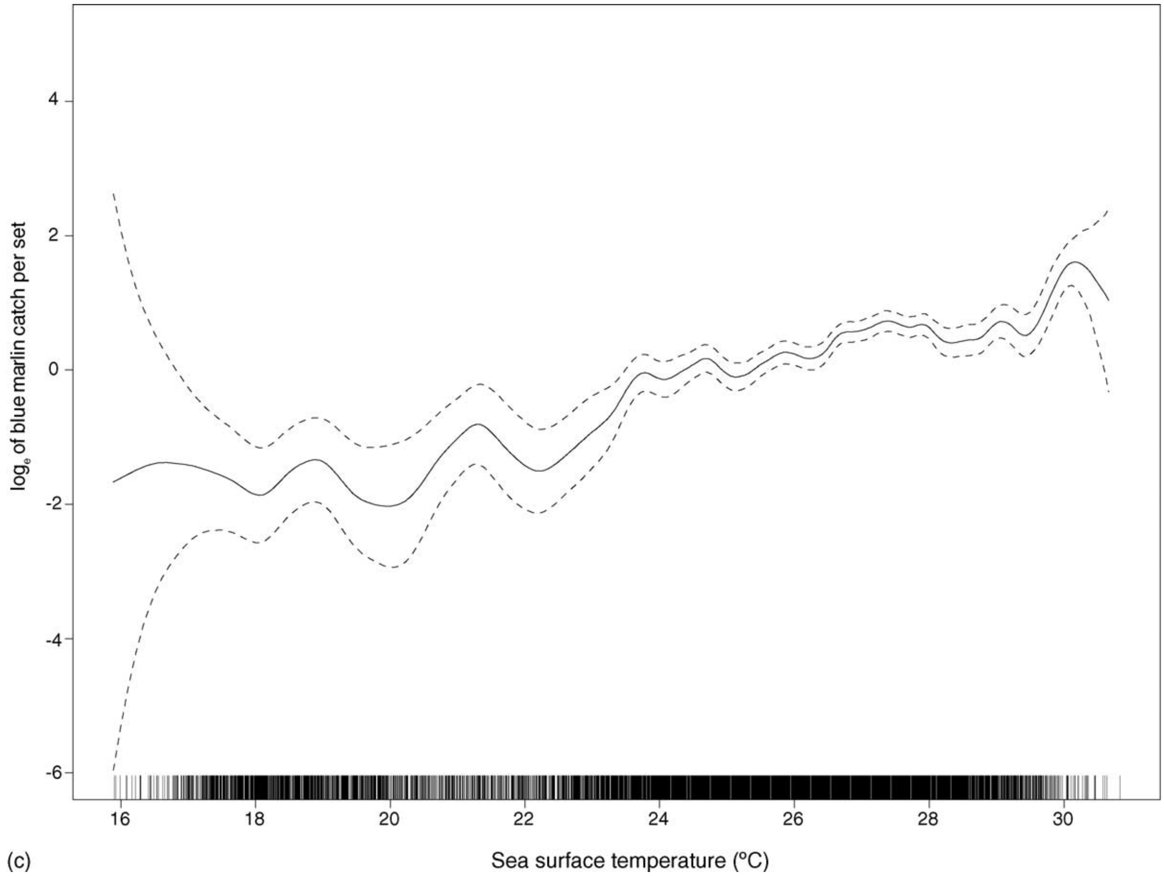

Fig. 1. (*Continued*).

Table 4
Summary of sequential logbook data editing ($N = 87\,277$ longline sets)

Apparent MisIDs	Trips	Vessels	Sets	Large SR	Blue marlin		Striped marlin		Shortbill spearfish		Black marlin		ΔBlue marlin (%)
					Log	Sold	Log	Sold	Log	Sold	Log	Sold	
Blue marlin	391	81	4541	2118	11 442	2742	3532	11 860	3181	3736	249	42	−19.9
Black marlin	43	17	458	49	328	276	505	830	301	443	442	1	−0.1
Blue, striped, and black marlins	10	8	117	12	46	111	50	94	52	45	102	0	0.2

"Apparent MisIDs" refers to the species that caused identification problems. Entries associated with each species include numbers of unobserved trips, vessels, sets, and large SR, along with kept fish as reported in logbooks, the corresponding numbers sold, and the percent change from the nominal initial total. See below for initial data.

Blue marlin caught (nominal): 45 043; sets with large SR: 5361; trips with large SR: 2161; vessels with large SR: 151.

Blue marlin kept (nominal): 43 747; blue marlin kept from sets with large SR: 23 363; trips with ≥2 large SR: 1106; vessels with ≥100 large SR: 12.

Blue marlin released (nominal): 1293.

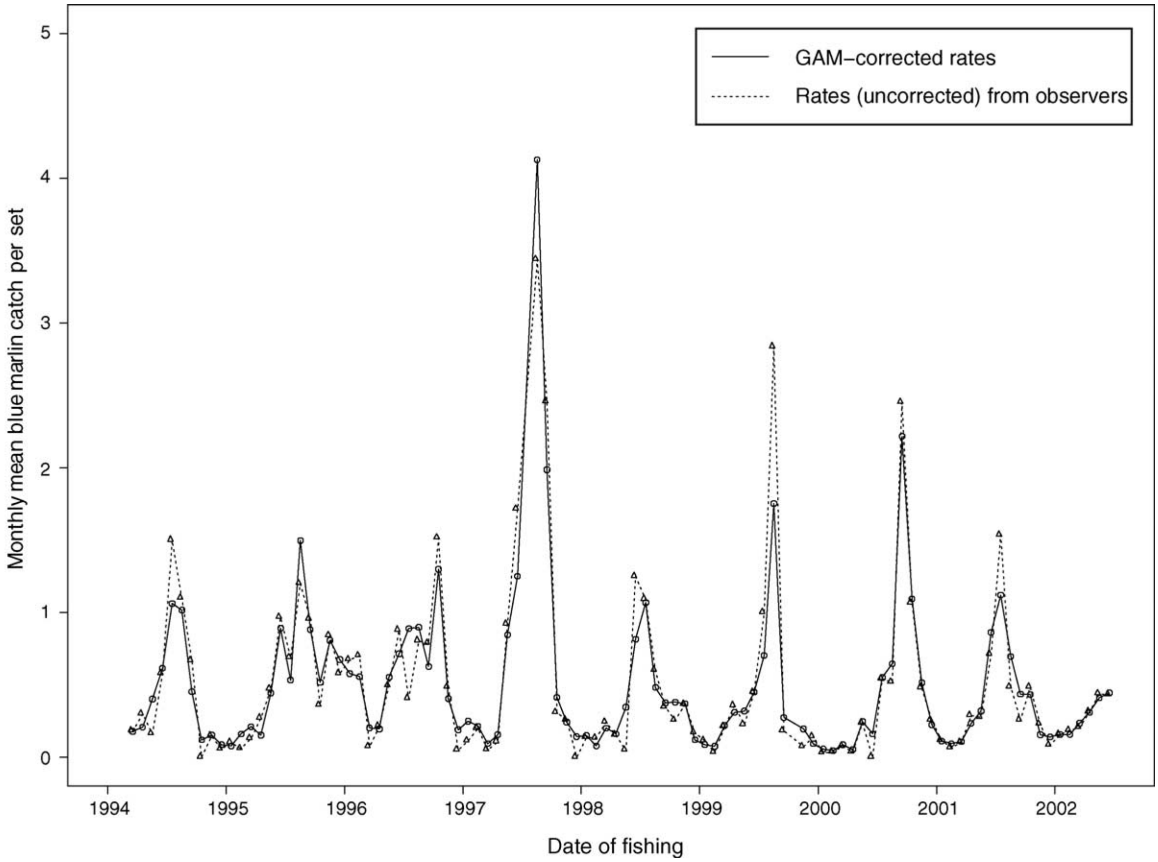

Fig. 2. GAM-corrected (solid trace with open circles) and uncorrected (dotted trace with open triangles) from March 1994 through June 2002.

ages of striped marlin and shortbill spearfish, which reflected the requirement for large SR predicated upon blue marlin reporting. The second detected underreporting of blue marlin, striped marlin, and shortbill spearfish, which reflected the decision to seek any patterns of misidentifications, including those involving striped marlin or shortbill spearfish logged as black marlin. Finally, results generated when all three marlin species were checked were unexpected because the sums of blue marlin, striped marlin, and shortbill spearfish on the identified sets exhibited shortfalls against the GAM predictions, but the kept and sold totals tallied exactly when misidentifications as black marlin were considered. An additional 97 trips with 1007 sets (4.9% of the large SR) were checked and accepted as accurate (blue marlin: 2250 kept, 2212 sold; striped marlin: 797 kept, 830 sold). Most (79%) of these trips were undertaken from April through September. Thus, editing reduced the estimate of kept blue marlin by 19.8%. If the edited value were considered accurate, the nominal total would be inflated by 24.6%.

The misidentifications were contagiously distributed within the fleet; with 39% of the large SR associated with only 7% of the permitted vessels. This indicated that the logbooks from these vessels were habitually in error. Moreover, all of these error-prone vessels ranked in the top 40% of the fleet in total marlins (i.e., of all species) caught, and 3 were in the top 10, including the highest total. Thus, logbook data accuracy was adversely affected by the combination of a strong tendency toward error combined with large catches of marlins subject to this tendency.

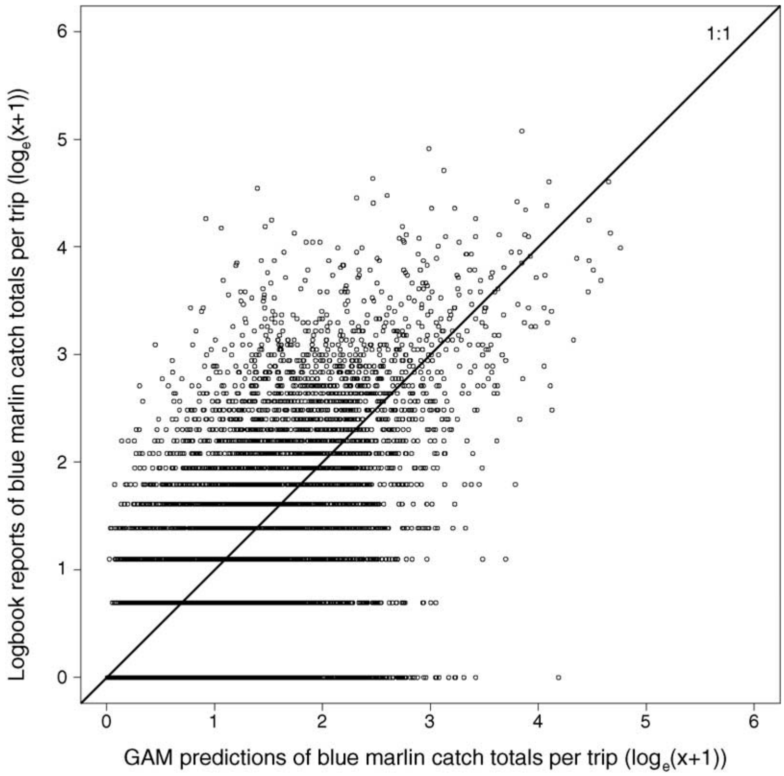

Fig. 3. Relationship between blue marlin catches (as trip totals) as reported in logbooks and as predicted by a GAM after transformation to $\log_e(X+1)$.

The GAM coefficients were then reapplied to the corrected data and the results summarized by

$$\log_e(Y+1) = 0.0369 + 0.6205\log_e(X+1) + \in \quad (4)$$

where Y represents the logbook value (catch) and X represents the GAM prediction. The corrections yielded improvements in the test statistic, coefficient of determination, and variance estimate ($F_{1,87\,275} = 16\,430$; $P = 0$; $R^2 = 0.161$; $s^2{}_{y*x} = 0.1624$), but the regression parameters indicated that bias engendered by misidentifications remained. Presentation of the corrected catches and predictions as trip totals (Fig. 3) revealed a generally direct relationship between the two, but pronounced curvature near the origin confirmed the presence of bias, which was not surprising given the conservative editing procedures.

A reporting summary (Table 5) for both observed and unobserved sets, which collectively comprised 95.0% of all effort during the study period, yielded an estimate of apparent logbook reporting error of 29.4%, equivalent to an average of 1335 fish misidentified as blue marlin per year. The GAM-generated 95% PI for the total catch (34 201–41 507 blue marlin) did not include the uncorrected overall total from the logbook reports (48 911 blue marlin). The uncorrected annual logbook totals exceeded the estimate every year and were not included within the individual 95% PIs in 6 of the 9 years. The 1994 total was included within the 95% PI by only two fish.

Fig. 4 depicts temporal patterns in blue marlin catches throughout the study period. The monthly mean GAM predictions and the mean corrected blue marlin catches per set (Fig. 4a) were very highly correlated ($r = 0.888$; d.f. = 98; $P = 0$). These traces did not track closely in August 1997, when the GAM predictions were considerably greater than the logbook values, but this reflected the aforementioned low observer coverage and gear configuration. As

Table 5
Summary of blue marlin reporting from March 1994 through June 2002 in the Hawaii-based longline fishery (N = 8397 observed plus 87 277 unobserved longline sets)

Year	Observer	GAM	Estimate (GAM + obs)	Logbook (obs)	Logbook (unobs)	Logbook (total)	Δ	95% PI	Apparent logbook error (%)
1994–2002	3208	34 580	37 788	3868	45 043	48 911	11 123	34 201–41 507	29.4
1994	142	2677	2819	200	3772	3972	1153	2070–3974	40.9
1995	265	5229	5494	345	8443	8788	3294	4342–7144	60.0
1996	348	5288	5636	449	6173	6622	986	4325–6523	17.5
1997	249	7821	8070	281	7962	8243	173	5994–10 170	2.1
1998	191	3415	3606	219	5030	5249	1643	2364–4414	45.6
1999	99	2917	3016	190	4729	4919	1903	1909–4246	63.1
2000	647	2864	3511	600	2927	3527	16	2968–4309	0.5
2001	770	3297	4067	992	4599	5591	1524	3613–4464	37.5
2002	497	1070	1567	592	1408	2000	433	1337–1702	27.6

Entries are numbers of blue marlin as reported by observers, as estimated by the GAM on unobserved sets, the sums of these values, as reported in logbooks on observed and unobserved sets, their sums, the differences between the estimated and logbook totals, the 95% PIs, and the apparent logbook error.

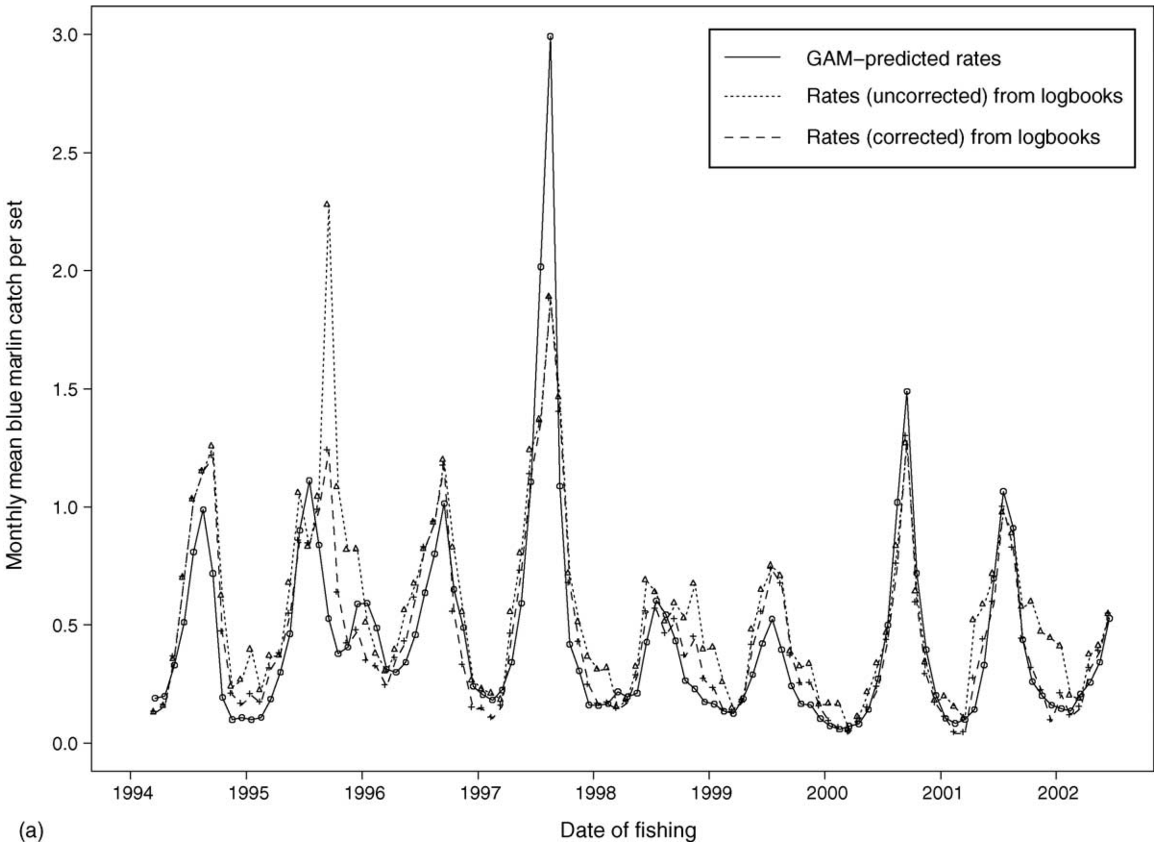

Fig. 4. (a) Monthly mean blue marlin catch as reported in logbooks (uncorrected: dotted trace with open triangles; corrected: dashed trace with crosses) and as predicted by a GAM (solid trace with open circles). (b) Monthly blue marlin catch totals as reported in logbooks (uncorrected: dotted trace with open triangles; corrected: dashed trace with crosses) and as predicted by a GAM (solid trace with open circles).

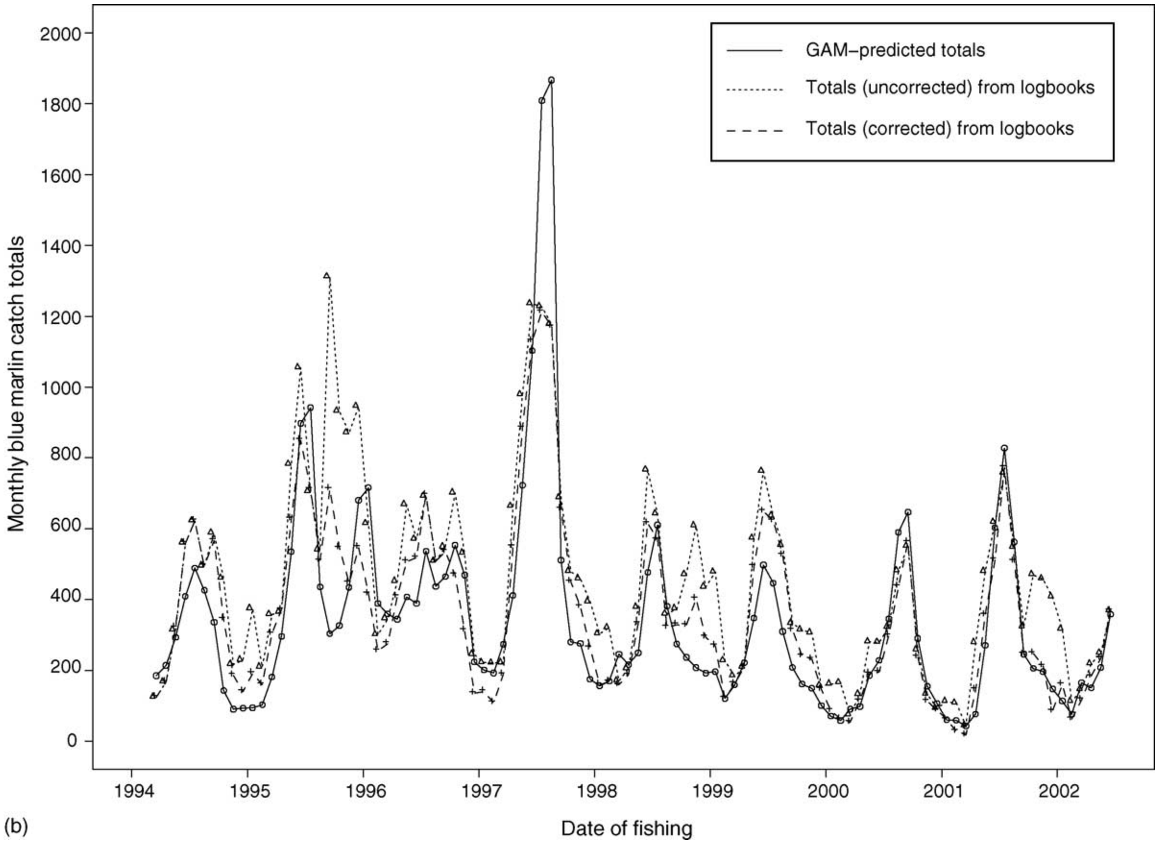

Fig. 4. (*Continued*).

expected, the uncorrected logbook data were also correlated with the GAM predictions ($r = 0.787$; d.f. = 98; $P = 0$), but the correspondence was weaker than with the corrected data. These uncorrected mean rates peaked in September 1995, caused largely by misidentifications of striped marlin on four trips, including two in succession by a single vessel that collectively inflated the uncorrected monthly mean by 1.0 blue marlin per set. The monthly catch totals (Fig. 4b) again demonstrated that the greatest effect of the corrections occurred in the autumn of 1995. The misidentifications in the fourth quarter of 2001 were equivalent to 19% of the nominal annual and 52% of the nominal quarterly catch totals, but the availability of full auction data allowed reconciliation of the sums of the GAM predictions and corrected logbook data to within 0.5%.

Evaluation of logbook data by trip types revealed an additional feature. Corrections applied to swordfish- and mixed species-trips primarily involved blue marlin and striped marlin. Tuna trips were much more likely to require corrections involving shortbill spearfish or black marlin.

3.5. *Effects of corrections on the apparent distribution of blue marlin catches*

The logbook corrections refined understanding of the distribution of blue marlin (Fig. 5). Uncorrected and corrected mean catches per set in 1° squares are presented for the fourth quarter of 2001 (Fig. 5a and b) and the last 4 months of 1995 (Fig. 5c and d), respectively. Results from 2001 indicated that high uncorrected mean catches (>5 per set) in two squares north of Hawaii were inflated. The uncorrected data from late summer and autumn of 1995 suggested that there were high blue marlin catches north of Hawaii, with four

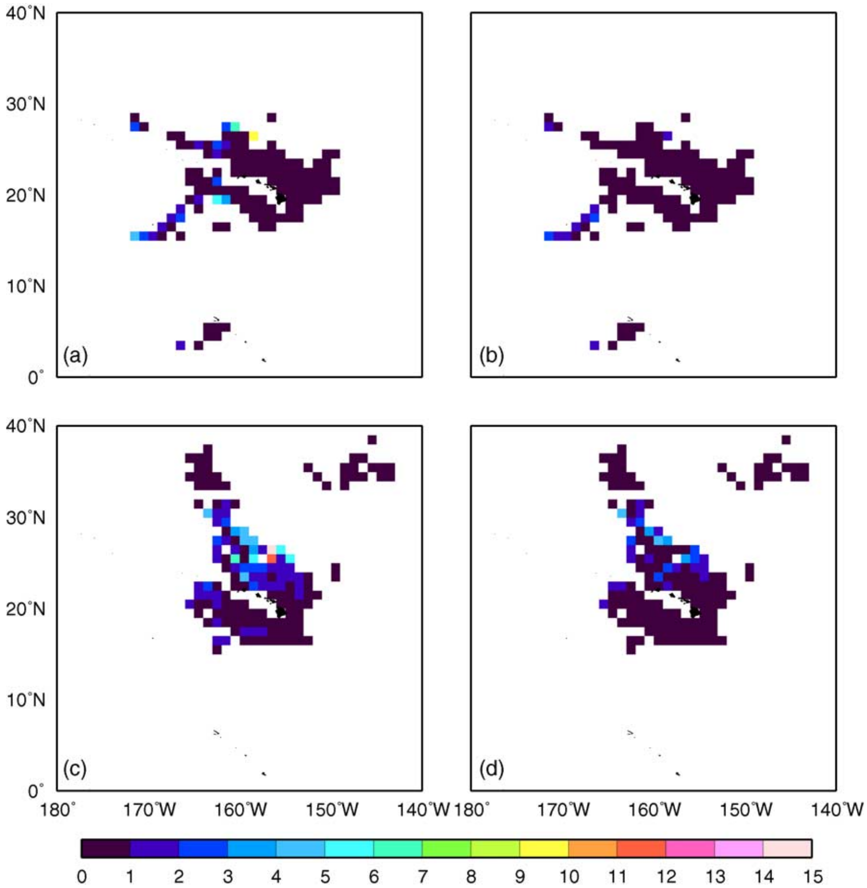

Fig. 5. (a) Uncorrected and (b) corrected blue marlin catches (≥ 3 permits per 1° square) from October to December 2001. (c) Uncorrected and (d) corrected blue marlin catches (≥ 3 permits per 1° square) from September to December 1995.

squares at 5–10 per set and two with >10 per set; the corrected data eliminated these high values and lessened this suggestion. The 1995 map depicts apparent catches to nearly 40°N, but it is likely that many were undetected misidentifications because the auction data were limited to partial coverage. More extensive auction data would probably have permitted correction of additional trips because 65% of the blue marlin reported during this period from north of 30°N was caught on sets with large SR.

3.6. Effects of corrections on the other species

The corrections applied to blue marlin necessitated substantial changes from the striped marlin and black marlin nominal catch totals and a smaller change in the shortbill spearfish total, despite involving small fractions of the fishing effort. Correction of 2.7 and 0.4% of the unobserved longline sets indicated that the striped marlin and shortbill spearfish catch totals were at least 8.2 and 1.0% greater than reported, respectively, whereas correction of only 0.4% of the sets changed the nominal black marlin total by −14.2%.

4. Discussion

The Hawaii-based longline fishery is characterized by virtually optimal circumstances for monitoring purposes. The HLLOP deploys its observers on a substantial and growing fraction of the fishing trips, gathering data that permitted development of a predictive model of blue marlin catch. The fleet is relatively small, with 97 permitted vessels active in 2002, and centrally located in Honolulu Harbor and nearby Kewalo Basin. As such, vessel movements can be closely tracked, which has contributed to a very high degree of reporting compliance (ca. 99%), defined as report submission within the allotted time, accuracy of identifications notwithstanding (W.A.E. Machado, PIFSC, personal communication). The preponderance (ca. 95%) of the landings passes through the UFA, which has provided full sales data since January 2000 with a lag of about one month (G.R. Kagami, HDAR, personal communication). We have utilized these fortuitous circumstances by an integrated process of prediction, comparison, and independent verification to improve the accuracy of logbook data for blue marlin, while identifying sources of reporting bias and estimating its magnitude. In so doing, we have demonstrated that methods previously applied to blue shark (Walsh et al., 2002), a numerous bycatch species where the primary concern was the possibility of under- or non-reporting, can be adapted for use with an economically valuable, ecologically and recreationally important, much less numerous, incidentally caught species, where the issue of primary concern is the accuracy of identifications.

The observer data that underlay the analysis met the expectation of a low systematic error rate. As such, the HLLOP training, which included visits to the UFA under the supervision of the training coordinator, examinations of preserved specimens under the supervision of a senior fisheries biologist, and additional examinations of photographs, was apparently adequate. In those cases with misidentifications in the observer data or the observer and logbook data, lack of experience was apparently the principal cause for the inaccuracies. Because there were no repeated systematic errors, it seems reasonable to infer that the debriefings also contributed to observer data quality and facilitated learning species identifications.

The GAM predictors all yielded statistically significant deviance reductions (all *F*-tests; $P < 0.05$), and the effects of the major predictors (i.e., the first six) appeared consistent with the relevant literature. For example, high blue marlin landings in summer were reported from HDAR data for 1950–1963 (Strasburg, 1970) and from the official fishery statistics for 1963–1970 (Yoshida, 1974). The predictor trace of sea surface temperature suggested that blue marlin catch would exhibit a generally positive trend from approximately 23 to 30 °C, which was reminiscent of the observation by Nakamura (2001) that blue marlin are largely confined to the warmer side of the 24 °C surface isotherm. Latitudinal effects were comprehensible in light of its correlate, sea surface temperature; 67% of the observed sets were deployed between 5° and 30°N at sea surface temperatures of 24.6–30.4 °C. It is recognized that the GAM was probably overparameterized, that there was no substantive theoretical basis for the degree of smoothing chosen, and that there were significant correlations among the predictors. We chose to accept these potential shortcomings in light of the large number of degrees of freedom, an absence of spikes in the fitted curves that might have represented accommodation of only a few observations, a desire to minimize bias, and our primary objective, which was correction of the catch history rather than a detailed interpretation of the predictor effects.

The GAM application documented several important quantitative aspects of the blue marlin logbook data. First, the analysis yielded a point estimate of substantial inaccuracy in the blue marlin logbook data, with an overall 95% PI that did not approach the logbook estimate closely, and individual PIs that reflected considerable interannual variability with no clear pattern of increasing accuracy of identifications over time. The results also demonstrated that the SR did in fact detect systematic misidentifications, not simply random variation. For example, the difference between the point estimates for logbook reporting and the GAM and observer results (11 123 blue marlin) included a discrepancy of 660 blue marlins caught on observed sets. Comparison of the reduction by editing (8643 kept blue marlin) to the apparent error on unobserved sets only ($11\,123 - 660 = 10\,463$) indicated that a very large fraction (82.6%) of the apparent error on unobserved sets had been identified and corrected, despite the conservative editing criteria. Moreover, the GAM application elucidated temporal and spatial patterns in the logbook error and identified the vessels responsible. The latter was important because, as shown in 1995, even a single vessel can cause substantial inaccuracy in the statistics from a fishery of this size. Finally, the application detected no evidence of systematic under- or non-reporting of marlins. Hence, it appears that logbook inaccuracies primarily originated from mistakes of various sorts rather than deliberate misconduct.

The GAM application also revealed a greater variety of misidentifications than had been expected. Specifically, several hundred shortbill spearfish were logged as blue marlin, which was not anticipated because these species do not resemble one another closely. This observation reflected the convoluted nature of the logbook error, in which one species might be misidentified as a second, which in turn was misidentified as a third, and so forth. The consequence of such dependencies was the demonstration that blue marlin misidentifications were associated with non-trivial minimum estimates of inaccuracy for three other species.

Finally, the uncorrected logbook data presented a fundamentally distorted view of the apparent distribution of blue marlin in this fishery. Nakamura (2001) described blue marlin as the most tropical among the Indo-Pacific billfishes, but the uncorrected logbook data would suggest that blue marlin sometimes migrate to 40°N. The corrections and the large SR from north of 30°N in 1995 suggest that blue marlin may not reach that latitude, or at least not regularly. This is important because it is directly related to the thermal biology of this species.

5. Conclusions

The analysis permitted correction of substantial upward bias and improved the accuracy of logbook data for blue marlin in the Hawaii-based longline fishery from March 1994 through June 2002. In so doing, methods previously used for bycatch estimation (Walsh et al., 2002) proved adaptable for use with incidental catch data characterized by a large proportion of misidentifications. The result was a more realistic catch history.

Despite virtually optimal monitoring circumstances, the nominal logbook data were sufficiently biased as to permit spurious inferences regarding such fundamental issues as catches and distributions. This raises the possibility that serious adverse consequences could ensue if fisheries scientists revised important estimates or managers altered or imposed regulatory standards on the basis of misleading data. This finding is also important in a more conceptual sense: development of ecosystem-based management for this fishery may be hindered if logbook data distort understanding of the geographical, physiological, and population ecology of this important species.

Bias arose from various types of mistakes committed before and after (e.g., keypunch errors) logbooks were submitted to NOAA Fisheries, presumably combined with some indifference to logbook preparation and ignorance of species identifications among fishermen, but not from dishonest reporting. This suggests that meaningful improvement in logbook data accuracy should be attainable by concentrating monitoring efforts on identifications, particularly among a small, error-prone minority, rather than catch totals. It should be recognized, however, that logbook data for mixed catches of billfishes could be inaccurate under any combination of systematic misidentifications, unfavorable monitoring circumstances, or dishonest reporting. Therefore, logbook data accuracy should receive serious and critical attention in the context of stock assessments.

Acknowledgments

The authors thank Keith Bigelow, Daniel Curran, David Hamm, Robert Humphreys, Michael Musyl, Joseph O'Malley, and Jerry Wetherall for reviewing earlier drafts of this manuscript, Walter Machado for discussions concerning the Hawaii-based longline fishery, Brent Miyamoto and Karen Sender for assistance in data management and retrieval, Evan Howell for assistance with graphics, and Francine Fiust and Judith Kendig for editing and preparation of the final version. This paper was funded by Cooperative Agreement No. NA 67RJ0154 from the National Oceanic and Atmospheric Administration.

References

Cook, R.C., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman and Hall, London.

DiNardo, G.T., 1993. Statistical guidelines for a pilot observer program to estimate turtle takes in the Hawaii longline fishery. NOAA Technical Memorandum NMFS (NOAA-TM-NMFS-SWFSC-190).

Dollar, R.A., 1992. Annual report of the 1991 Western Pacific longline fishery. Honolulu Laboratory, Southwest Fisheries Science Center, NMFS-NOAA Admin. Rep. No. H-92-11, 26 pp.

Draper, N.R., Smith, H., 1981. Applied Regression Analysis, second ed. John Wiley & Sons Inc., New York.

Hastie, T.J., 1992. Generalized additive models. In: Chambers, J.M., Hastie, T.J. (Eds.), Statistical Models in S. Wadsworth &

Brooks/Cole Advanced Books and Software. Pacific Grove, CA, pp. 249–306.

He, X., Bigelow, K.A., Boggs, C.H., 1997. Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. Fisheries Res. 31, 147–158.

Insightful Corp., 2002. S-PLUS, Version 6.1.2. Insightful Corp., Seattle, WA.

Hoaglin, D.C., Mosteller, F., Tukey, J.W., 1983. Understanding Robust and Exploratory Data Analysis. John Wiley & Sons Inc., New York.

Ito, R.Y., Coan Jr., A.L., 1999. U.S. swordfish fisheries of the North Pacific Ocean. In: DiNardo, G.T. (Ed.), Proceedings of the Second International Pacific Swordfish Symposium. NOAA-TM-NMFS-SWFSC-263, U.S. Department of Commerce, pp. 19–38.

Ito, R.Y., Machado, W.A., 2001. Annual report of the Hawaii-based longline fishery. Honolulu Laboratory, Southwest Fisheries Science Center, NMFS-NOAA Admin. Rep. No. H-01-07, 55 pp.

McCracken, M.L., 2004. Modeling a very rare event to estimate sea turtle bycatch: lessons learned. NOAA-TM-NMFS-PIFSC-3, U.S. Department of Commerce, 24 pp.

Nakamura, I., 2001. Istiophoridae: Billfishes. In: The Living Marine Resources of the Western Central Pacific. vol. 6. Bony Fishes. Part 4 (Labridae to Latimeridae), Estuarine Crocodiles, Sea Turtles, Sea Snakes, and Marine Mammals. Food and Agriculture Organization of the United Nations. Rome.

National Marine Fisheries Service, 2003. Evaluating bycatch: a national approach to standardized bycatch monitoring programs. NOAA, NMFS, Silver Spring, MD, 64 pp.

Pacific Islands Regional Office, 2003. Hawaii Longline Observer Program Observer Field Manual. Pacific Islands Region, NOAA Fisheries, National Oceanic and Atmospheric Administration, United States Department of Commerce.

Schimek, M.G., Turlach, B.A., 2000. Additive and generalized additive models. In: Schimek, M.G. (Ed.), Smoothing and Regression: Approaches, Computation, and Application. John Wiley & Sons Inc., New York, pp. 277–327.

Strasburg, D.W., 1970. A report on the billfishes of the central Pacific Ocean. Bull. Marine Sci. 20, 575–604.

Venables, W.N., Ripley, B.D., 1994. Modern Applied Statistics with S-Plus. Springer-Verlag Inc., New York.

Walsh, W.A., 2000. Comparisons of fish catches reported by fishery observers and in logbooks of Hawaii-based commercial longline vessels. Honolulu Laboratory, Southwest Fisheries Science Center, NMFS-NOAA Admin. Rep. No. H-00-07, 45 pp.

Walsh, W.A., Kleiber, P., 2001. Generalized additive model and regression tree analyses of blue shark (*Prionace glauca*) catch rates by the Hawaii-based longline fishery. Fisheries Res. 53, 115–131.

Walsh, W.A., Kleiber, P., McCracken, M., 2002. Comparison of logbook reports of incidental blue shark catch rates by Hawaii-based longline vessels to fishery observer data by application of a generalized additive model. Fisheries Res. 58, 79–94.

Yoshida, H.O., 1974. Landings of billfishes in the Hawaiian longline fishery. In: Shomura, R.S., Williams, F. (Eds.), Proceedings of the International Billfish Symposium. Kailua-Kona, Hawaii, 9–12 August 1972, Part 2. Review and Contributed Papers, NOAA Technical Report NMFS SSRF-675.