Evaluation of Biological Sampling Protocols for At-Sea Groundfish Observers in Alaska

Steven J. Barbeaux, Sarah Gaichas, James N. Ianelli, and Martin W. Dorn

Evaluation of Biological Sampling Protocols for At-Sea Groundfish Observers in Alaska

Steven J. Barbeaux, Sarah Gaichas, James N. Ianelli, and Martin W. Dorn

ABSTRACT: In 1999 the North Pacific Groundfish Observer Program (NPGOP) changed sampling protocols in order to optimize available observer resources. As the NPGOP modifies their program to meet an increasing number of objectives, it is important to ensure that new protocols continue to meet previously defined objectives. In this study, we evaluate how the changes in sampling protocols have affected data critical to stock assessments. We explore how changes to the length and otolith sampling protocols effect estimates of length distributions, mean length, and catch-at-age distributions in the eastern Bering Sea walleye pollock *Theragra chalcogramma* fishery. We also investigate the spatial distribution of observer length and otoliths samples in relation to total catch. We found that the modified protocols employed by NPGOP in 1999 did not significantly reduce the precision of estimates based on collected data, but did improve the spatial distribution of length and otolith sampling in this fishery. Given these results, coupled with an overall reduction in the time necessary to complete these sampling tasks, we conclude that the modification to observer sampling protocols in 1999 improved data collection in the NPGOP.

INTRODUCTION

Accurate estimates of the biological attributes of catch, including species, size, weight, and age composition, are essential elements in the management of living marine resources. The sampling methods employed in the collection of biological data affect the data's precision, which in turn affect the reliability of assessments conducted using these data. In general, precision of estimates can be increased by increasing the number of samples taken, or by increasing individual sample size, or both (Cochran 1977). In almost all cases, increasing the number of samples taken or increasing sample size increases the cost of data collection. In the North Pacific, government-trained fisheries observers are employed to collect biological data from commercial groundfish catches. Although the daily costs of placing an observer on a fishing vessel in the North Pacific ($450 to $2000 per day) is independent of the amount of data collected, there are many competing data collection needs and the time an observer has available for sampling is limited and fixed (Marine Resource Assessment Group Americas, Inc. 2003, National Marine Fisheries Service 2000, Terry *in press*).

In 1999, the North Pacific Groundfish Observer Program (NPGOP) implemented new sampling protocols for the selection of fish for biological samples. These modified protocols reduced the number of fish being sampled per fishing unit (trawl haul-back, long-line or pot set), but increased the number of fishing units sampled. The NPGOP implemented this change in order to optimize available observer resources, while maintaining or increasing the precision of estimates based on these biological data. These changes to sampling protocols were developed by NPGOP staff based on the analytical and statistical review of observer sampling procedures conducted by Versar, Inc. (Volstad et al. 1997).

The Marine Resource Assessment Group Americas, Inc. (MRAG) assessed observer workload and time budgets in studies for the NPGOP (MRAG 2003). As the NPGOP uses this information to refine their program to meet an increasing number of objectives, it is important to ensure that new protocols continue to meet previously defined objectives. In this study, we evaluate how the changes in sampling protocols have affected data critical to stock assessments.

Three types of analyses were used in this evaluation. The first analysis examined precision in the length sampling. The second analysis compared catch-at-age estimates employing methods described by Kimura (1989) and modified by Dorn (1992) using data collected under both sampling protocols. We

employed the methods of Versar, Inc. (Volstad et al. 1997) where the precision of the estimates based on sampling theory is used. For the third analysis, we conducted a spatial assessment of the 1998 and 2000 distributions of length and otolith samples relative to the catch of walleye pollock *Theragra chalcogramma* in the eastern Bering Sea (EBS) by the catcher-processor and mothership (CP) sector.

MATERIALS AND METHODS

Data collection

Paired sampling study

Paired sampling was conducted on 32 observed cruises on 19 CP vessels participating in the 1999 EBS walleye pollock fishery. The CP sector of the EBS pollock fishery was chosen as the test fishery because the greatest amount of biological data is collected each year from this fishery. Because of the large amount of data available, differences due to changes in sampling protocols could be more accurately measured. In addition, catcher–processor vessels and motherships participating in the EBS pollock fishery are required to carry 2 observers, allowing us to employ both sampling protocols on a vessel concurrently.

Of the 32 observed cruises, 16 were conducted during the pollock A-season (20 January to 31 March) and 16 were conducted during the pollock B-season (20 July to 31 October). Two observers were assigned to each of the 32 cruises. On each vessel the 2 observers worked separate 12-hour shifts; one observer was assigned to collect data using the original sampling protocol (OP) and the other to collect data using the modified sampling protocol (MP).

Spatial analysis study

We employed ARCGIS software (ESRI ArcMap v8.3 Geostatical Analyst Module. Environmental Systems Research Institute) to analyze the spatial distribution of length and otolith collections from the 1998 OP and 2000 MP Bering Sea CP fleet pollock fishery. We plotted fishing locations by Alaska Department of Fish and Game (ADF&G) management area; measurements were taken and otoliths were collected. In general, ADF&G management areas are 30 latitude by 60 longitude cells, except near shore where they are confined by state waters and the shoreline.

Original protocol

The OP used to select fish for the length and otolith sample collections (Table 1) is described in the 1998 NPGOP observer sampling manual (NPGOP 1997). Each observer collected approximately 150 walleye pollock per day for length samples. Selecting fish for the length samples from several hauls per day was preferred, but observers were allowed to select all of the fish for each day's length samples from one haul, if necessary. Fish did not need to be selected from species composition samples. Ob-

Table 1. Comparison of Original Protocol and Modified Protocol sampling on pollock fishing vessels.

Sample Type		Original Protocol	Modified Protocol
Length–Frequency	Vessel	All observers are assigned to collect length–frequency samples.	All observers are assigned to collect length–frequency samples.
	Haul	Determined haphazardly by the observer, with fish selected from multiple hauls preferred, but not required.	Based on vessel type; for catcher–processor vessels with 2 observers all hauls are sampled, for smaller vessels with one observer, the hauls sampled depend on a random sample table.
	Fish	150 fish measured per day.	20 fish measured per haul for every haul selected.
Otolith	Vessel	The observer program assigns otolith collection to a subset of observers based on expected vessel assignment.	All observers assigned to collect otolith samples.
	Haul	Determined by the observers, but must be a subset of the length–frequency hauls sampled.	Every 5th haul sampled for length–frequency.
	Fish	A maximum of 20 fish samples, the collections were length stratified with 5 fish selected per sex and centimeter category.	Random sample of 2 fish from the length–frequency sample.

servers were asked to select fish in a way that did not introduce size bias.

Observers used a length-stratified collection method for selecting fish for the otolith samples. Prior to 1999, not all observers were required to collect otolith samples. The observer program assigned otolith collections to a proportion of the observers based on their planned vessel assignment. Due to changes in assignments after training or briefing, some vessels were inadvertently missed. All of the cruise and vessel combinations in the 1999 study had otolith collections assigned. Otolith collections were limited to 5 otolith pairs per sex and centimeter category for an observer's entire cruise. Observers selected a maximum of 20 pollock per day (10 male and 10 female) from their length samples for otolith samples. Observers kept a running tally of the length categories and sexes of fish selected for otolith samples throughout their cruise to determine which sizes and sexes were not collected. Observers were instructed not to search outside their length samples for their otolith sample fish. Observers measured, sexed, weighed, and extracted otoliths from walleye pollock using the standard NPGOP methodology (NPGOP 1997).

Modified protocol

For the MP, each observer was instructed to collect 20 walleye pollock from every haul sampled for species composition using either a random or stratified sampling system. Because of variability in the configuration of the sampling stations on the commercial fishing vessels, observers were allowed to determine the selection methodology which best suited their sampling situation. Observers were instructed on how to implement systematic, spatial, and temporal sampling frames for selection of fish for length and otolith samples. Observers measured, weighed, and determined the sex of all selected fish. For every fifth haul, 2 of these fish were randomly selected and otolith samples collected (NPGOP 1998).

Otolith age reading

Since the MP otoliths were used for the 2000 Bering Sea walleye pollock stock assessment, they were aged by staff from the Alaska Fisheries Science Center's Age and Growth Program in 2000 using the revised criteria described in Hollowed et al. (1995). The OP otoliths were stored and analyzed for age determination in 2003 using the same criteria and same age readers. Unlike the MP otoliths, vials for some of the OP otoliths were not completely sealed and ap-

proximately 20% dried during storage. An analysis of percent agreement between readers from the OP and MP found that the percent agreement was similar for both collection protocols—suggesting that the different storage treatments had minimal effect (Short, J. Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle WA 98115, personal communication).

Data analysis

Length–frequency sampling comparison

For the analysis of the length–frequency collections, we used diversity indices to measure the "breadth" of the length–frequency distributions. Large, consistent differences in diversity indices would indicate differences created by the change in sampling protocols regardless of the total number of fish measured. Consistently smaller index values would indicate a less diverse or less even length–frequency distribution. We also compared the precision of the 2 sampling methodologies using the methods employed by Volstad et al. (1997) in their analysis of biological sampling.

We defined the "breadth" of the length–frequency distributions as the measure of diversity or evenness of the length–frequency distribution. We employed the Simpson diversity index (SDI)

$$SDI_k = 1 - \sum_{l=1\min}^{l_{\max}} \theta_l ,$$

where θ_l is the proportion of the length–frequency sample for vessel k contained in the length l (Thompson et al. 2001).

Volstad et al. (1997) evaluated the efficiency of sampling strategies for collecting biological data for target species by estimating the CV of the mean length in fleetwide catches of pollock and yellowfin sole. They then assessed the effects on the CV of changes in the fraction of hauls sampled and the number of fish sampled from each haul.

We estimated mean length based on a 2-stage cluster sampling design with a haul as the primary unit, and the fish sampled as the secondary unit. We assumed that data on length and sex collected from individual fish were from simple random samples taken from n randomly selected hauls for each cruise. The population of interest is the pollock catch from the 32 cruises. Our analysis compares the estimates obtained from the 2 sampling protocols.

For each sampling protocol we estimated the mean length ($\overline{X}_k$) by sex for the 32 cruises (Cochran 1977, Volstad et al. 1997) and tested each for homogeneity using a 2-tailed paired Student's *t*-test (Zar 1999) at

an $\alpha=0.05$. We then calculated an estimator for the variance of using the formula:

$$v\{\bar{X}_k\} = \left\{\frac{1-f_{1k}}{M_k^2 n_k}\sum_{i=1}^{n_k}\frac{M_{ik}^2(\bar{x}_{ik}-\bar{x}_K)^2}{(n_k-1)}\right\} + \frac{1}{n_k N_k \bar{M}_k^2}\sum_{i=1}^{n_k}M_{ik}^2(1-f_{2ik})\frac{s_{2ik}^2}{m_{ik}},$$

where N_k is the total number of hauls for vessel k, and n_k is the number of hauls sampled for vessel k, f_{2ik} is the fraction of fish caught that were measured in haul i for vessel k, f_{1k} is the fraction of hauls sampled for vessel k, and where

$$s_{2ik}^2 = \sum_{j=1}^{m_{jk}}\frac{(x_{ijk}-\bar{x}_{ik})^2}{(m_{ik}-1)}$$

is an estimate of the within-haul variance of the individual lengths based on the measurements of m_{ik} fish in the subsample from haul i (Cochran 1977, Volstad et al. 1997). The fraction of hauls sampled f_{1k} was derived from the 1998 season for the OP and the 2000 season for the MP. Two vessels in the study did not fish for pollock in 2000, f_{1k} for these vessels was calculated as the mean fraction of hauls sampled in 2000 for all other vessels from the study (f_{1k}=0.96).

The CVs were calculated for each of the cruises and curves fitted to the OP and MP CVs using the formula:

$$CV = \alpha n_k^\beta,$$

where n_k is the number of hauls sampled for vessel k, α is the Y-axis intercept, and β determines the slope of the curve.

Otolith and length–frequency sampling comparison: Age–length keys

We estimated the catch-at-age composition using the methods described by Kimura (1989), modified by Dorn (1992), and employed by Ianelli et al. (2001) for both OP and MP data. Length-stratified age data were used to construct age–length keys for each sampling system, stratum, and sex. These keys were then applied to the catch and length–frequency data for each sampling system for the 32 cruises. The stratum-specific age composition estimates were then weighted by the total pollock catch for the cruises within each stratum to arrive at an overall age composition for the 32 cruises for each sampling system. The 3 strata for this analysis were the same as that used for the 1999 stock assessment: (1) International North Pacific Fisheries Commission (INPFC) Area 51 from January–June, (2) INPFC Area 51 from July–December, and (3) INPFC Area 52 from January–December. In the species composition samples, we expanded the weight of pollock out to the official total catch for our total catch estimates for each cruise. Because there was a large discrepancy in the total number of pollock measured between the 2 methods (165,000 for the OP versus 65,000 for the MP), the OP lengths were randomly sampled to reduce the number to equal that of the MP collections (Table 2). This catch-at-age CV estimate both illustrates the potential differences in catch-at-age estimates due to collection methods, and minimizes the differences of catch-at-age estimates due to the disparity in the number of lengths collected.

We used the Vartot error index of estimated age composition as described by Kimura (1977) and Lai (1987) to compare the variance in the age–length key data collected under the 2 protocols. The proportion of fish at the ith age class (p_i) and variance of p_i are estimated as

$$p_i = \sum_{j=1}^{L}l_j q_{ij},$$

$$Var(p_i) = \sum_{j=1}^{L}\left[\frac{l_j^2 q_{ij}(1-q_{ij})}{n_j} + \frac{l_j(q_{ij}-p_i)^2}{N}\right],$$

and

$$Vartot = \sum_{i=1}^{A}Var(p_i).$$

where l_j is the proportion of fish that fall into the jth length stratum, N is the total length sample size, n_j is

Table 2. Number of aged and measured fish in the 32 cruises used to estimate fishery age composition.

	Number Aged			Number Measured			Number Weighed		
	Males	Females	Total	Males	Females	Total	Males	Females	Total
OP	486	514	1,000	36,068	32,663	68,731	1,431	1,503	2,934
MP	531	460	991	32,530	29,609	62,139	459	524	983

the size of the age subsample in the *j*th length stratum, q_{ij} is the proportion of n_j fish classified in to the *i*th age group, *A* is the number of age groups, and *L* is the number of length strata.

Spatial analysis

To compare 1998 and 2000, we estimated the total number of pollock caught in each ADF&G management area. Since all CP vessels fishing pollock in the Bering Sea required 100% observer coverage, all catch by this sector of the fleet was accounted for and fishing locations for every haul were obtained. Not all hauls were sampled for composition; therefore we needed to estimate the total numbers of pollock caught from the catch weight. We obtained average kilograms of pollock caught per ton of catch and average weight of pollock from observer species composition data for 1998 and 2000. These numbers were then applied to the total catch in tons of pollock per ADF&G management area per year to obtain an estimate of the total number of pollock caught in each ADF&G management area per year.

A Chi-squared homogeneity test was used to test proportionality of pollock length–frequency and otolith sampling in comparison to total numbers of pollock caught (Zar 1999). The larger the Chi-squared value, the further the sample numbers are from that expected if the number of pollock sampled are proportional to the number of pollock caught in each cell. Since variation from exact proportionality is expected even in the best sampling circumstances for this analysis, we deemed an α of 0.001 as significant ($\chi_{0.001,1}$=10.828). An overall measure of homogeneity—and therefore proportionality—can be obtained through summing the cell Chi-squared values for each year. The F-statistic was then used to test whether there was a significant difference in proportionality between sampling in 1998 and 2000.

RESULTS

Length–frequency sampling analysis

The analysis of the pollock mean length from 32 cruises conducted in 1999 revealed no measurable difference between mean length estimates obtained from OP and MP length–frequency collections. Length–frequency distributions were similar in all 32 cruises (Figure 1). The SDI revealed no consistent differences between the length–frequency distributions collected through the OP and MP (Figure 2). The mean SDI for female pollock by cruise was 0.944 in the OP and 0.943 in the MP. The mean SDI for male pollock by cruise was 0.934 in the OP and 0.933 in the MP. The SDI indicates that diversity was higher in the cruises occurring in the summer and fall of 1999 than in the earlier cruises. In general, the difference in diversity was higher between seasons than between sampling methods.

Mean length for female pollock ranged from 37.74 to 49.08 cm (an average mean of 44.10 cm and variance of the means of 6.16 cm) for the OP and 37.59 to 48.99 cm (an average mean of 44.10 cm and variance of the means of 6.12 cm) for the MP. The mean length for male pollock ranged from 37.65 to 46.60 cm (an average mean of 42.96 and variance of the means of 4.05 cm) for the OP and 37.67 to 46.23 cm (an average mean of 43.08 cm and variance of the means of 3.76 cm) for the MP. The 2-tailed Student's *t*-test for female and male pollock mean lengths suggests that the mean lengths of pollock are not significantly different for the 2 sampling systems ($P=0.96$ and $P=0.33$ for female and male pollock respectively).

Examination of CVs from the 2 sampling methods revealed that the OP CVs of the mean length were on average lower than the MP CVs per cruise (female OP mean CV=0.0029, female MP mean CV=0.0041, male OP mean CV=0.0024, male MP mean CV=0.0035). Since observers on the 32 cruises worked in 12-hour shifts and the MP called for fewer measurements per haul, fewer lengths were collected on each cruise for the MP, therefore CVs are lower for the OP due primarily to the higher number of pollock measured per cruise (Figures 3a–d). When the CVs were plotted against the number of fish measured and curves fitted, there was no substantial difference between the OP and the MP curves (Figures 3a, 3b) where data were available. Fitting curves to the CVs plotted against number of hauls sampled resulted in the OP having a lower Y-axis intercept and a steeper slope than the MP (female OP CV=$0.0193n_k^{-0.4865}$, female MP CV=$0.0225n_k^{-0.3851}$, male OP CV=$0.0132n_k^{-0.4321}$, male MP CV=$0.0205n_k^{-0.4045}$; Figures 3c, 3d). This resulted in a difference in the CVs of 0.0018 for females and 0.0014 for males at 100 sampled hauls between the 2 protocols. The difference in CVs is decreased when the MP curve is adjusted for the increase in the proportion of hauls sampled in the MP. In 2000, 2.4 times as many hauls were sampled than in 1998. Figures 3c and 3d illustrate the change in curves when the curves are adjusted by the increase in number of sampled hauls expected under the MP. At the adjusted sampling level there is a difference in CV of 0.0007 for females and 0.0004 for males; this difference decreases further as the total number of hauls sampled increases and the CVs of the mean length approach zero.
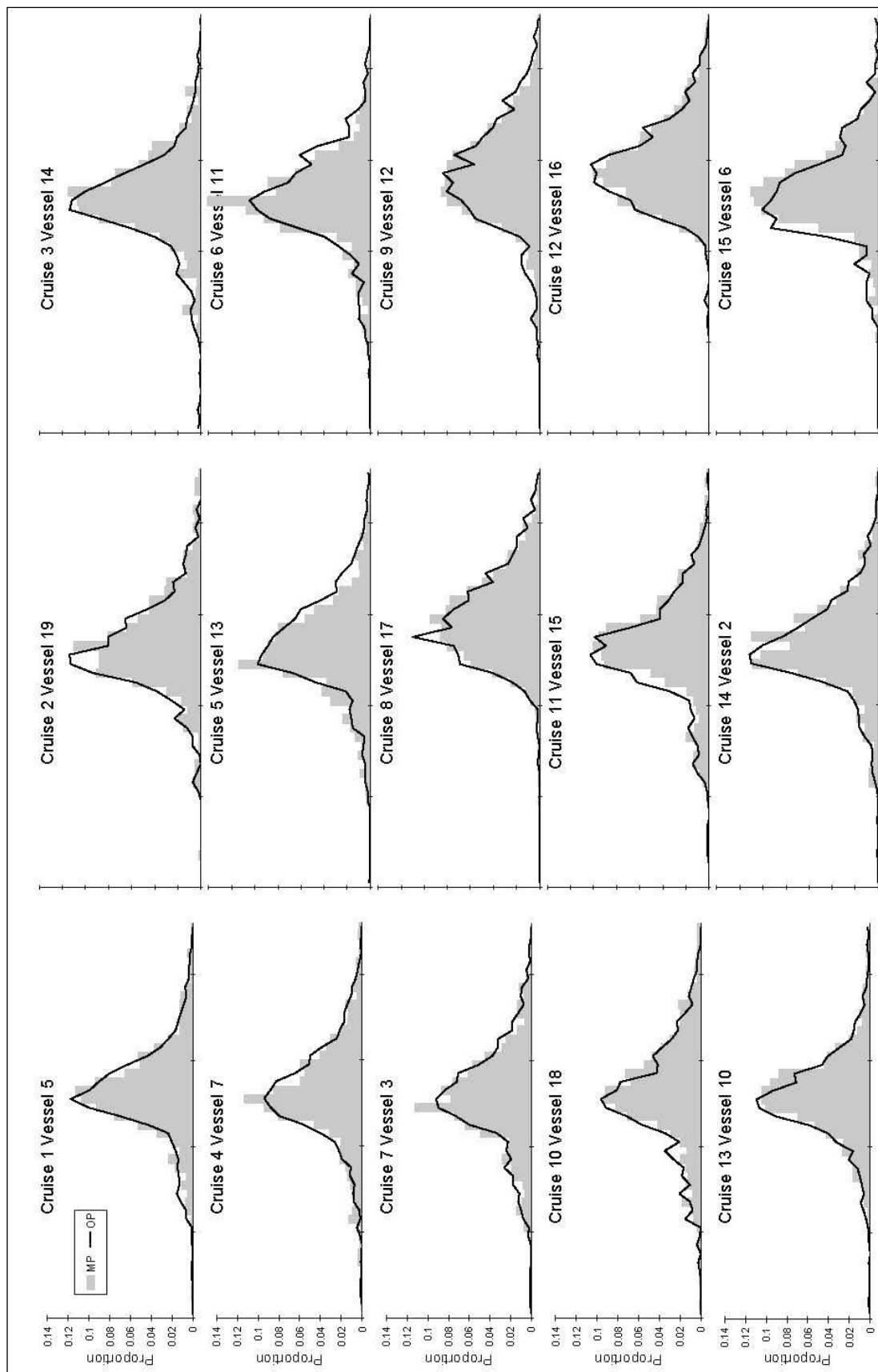
Figure 1. Female pollock proportion at length for the original protocol (OP) and modified protocol (MP) replicate length frequency samples.
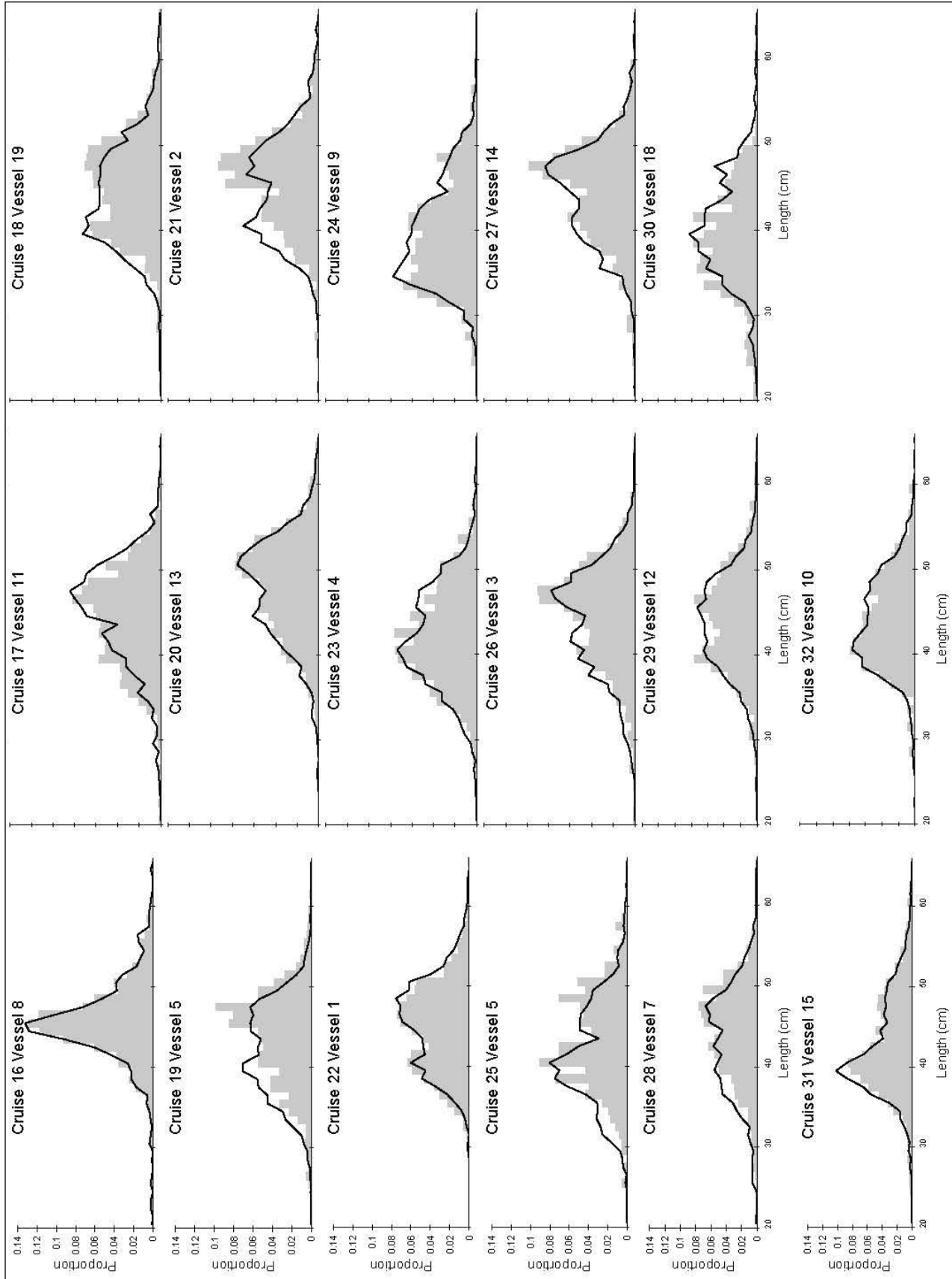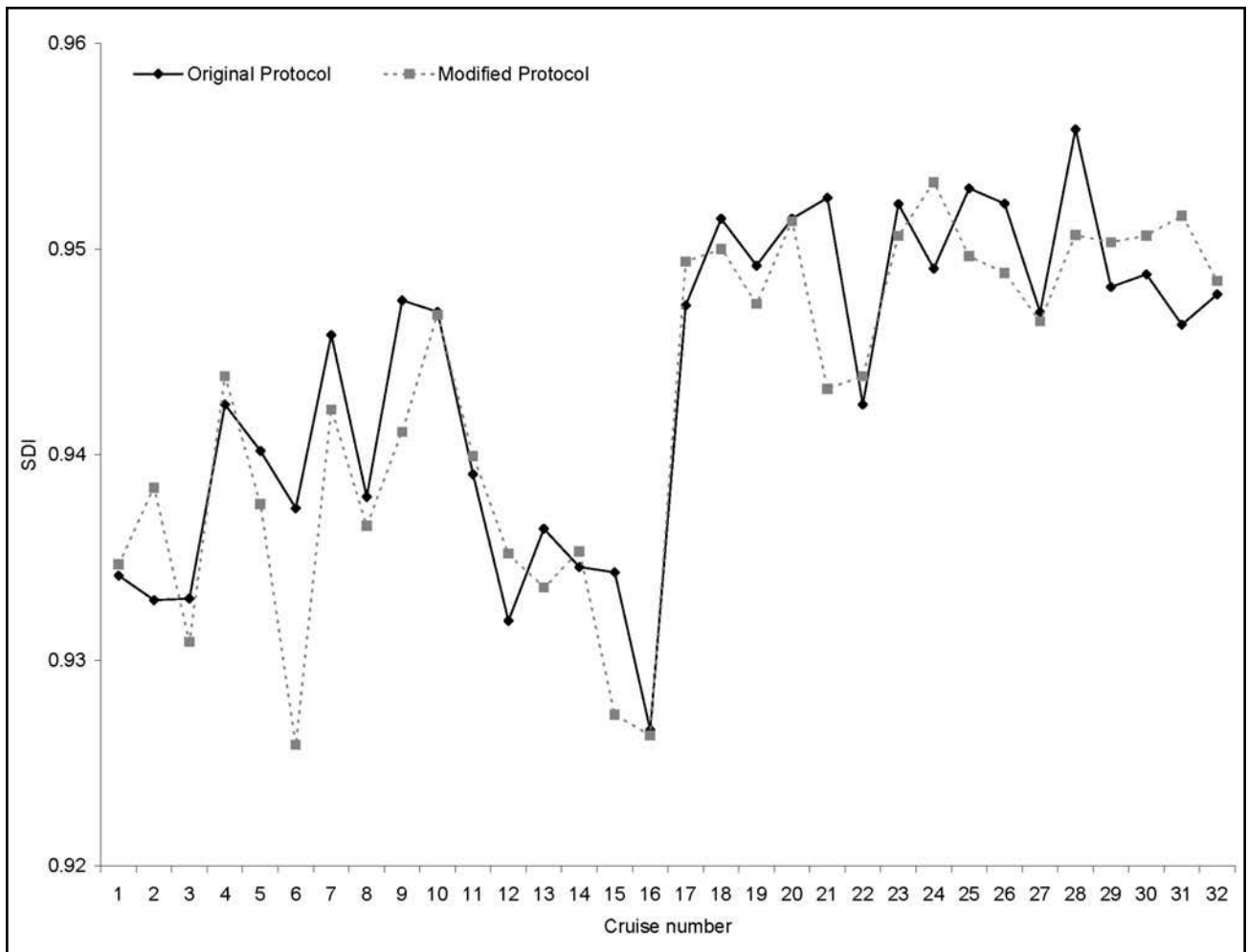
Figure 1. Continued.

Figure 2. Comparison of the Simpson diversity index for the original protocol and modified protocol length frequency collections by cruise and vessel

Otolith and length–frequency sampling comparisons

Review of the catch-at-age distribution created from the age–length keys reveals that the 2 sampling methods produce similar bimodal distributions (Table 3; Figure 4). The OP distribution is wider with more pollock in the younger (1 to 6 years) and older (10 to 15+ years) age groups and fewer in the 7- to 9-year-old age groups. Most of the age group distributions are within 2 standard deviations of the other sampling method. The catch estimates of the 1-, 2-, 8-, and 15+-year-old age groups are not within 2 standard deviations. For the 1- and 15+-year-old age groups, the difference in catch estimates may be a function of the low sample numbers. In the MP age–length key, the 1-year-old age group was based on a single 1-year-old fish, whereas the OP had no otoliths samples aged at 1-year-old.

Similarly, the 15+-year-old age group estimate was based on only 3 fish aged 15 to 17 years old from the OP otolith samples, while there were no fish aged greater than 13 years old from the selected MP otolith samples (Table 4).

The Vartot analysis results reveal that the variance estimate in less prevalent age groups was similar for both protocols, but the MP had lower variance in the more prevalent age groups (Figure 5) resulting in slightly lower overall variance (Vartot=0.00063 for the MP and Vartot=0.00086 for the OP).

Spatial analysis

In 1998, the Bering Sea CP pollock fishery captured approximately 1.133 billion pollock in 84 ADF&G management areas. Under the OP, in 1998 observers measured 323,490 pollock and collected 4,069 otoliths
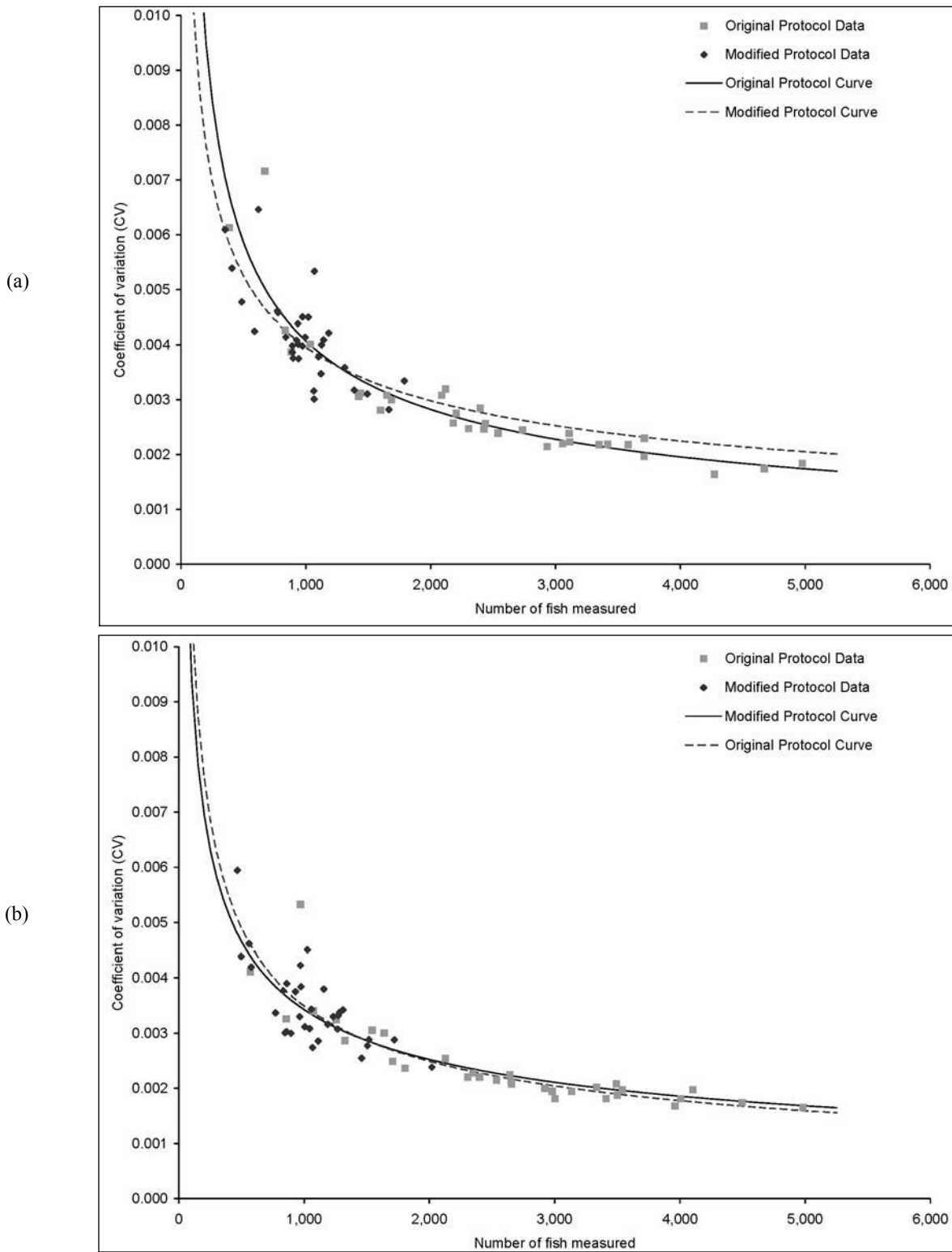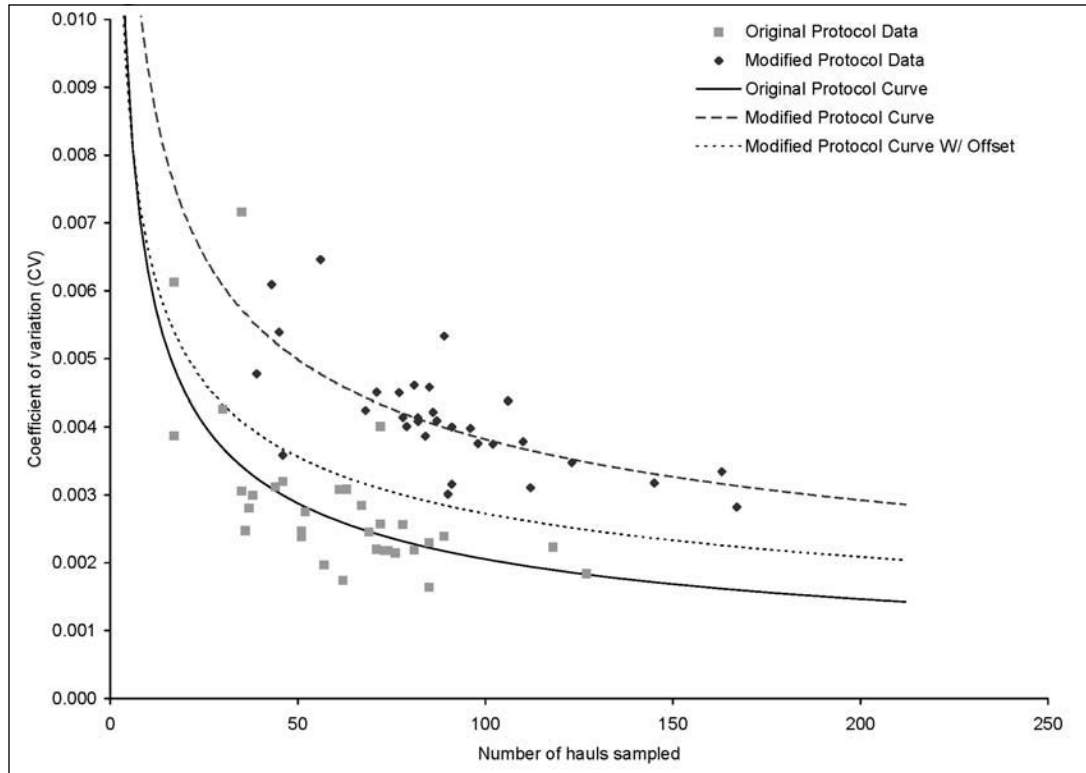
Figure 3. Coefficient of variation (CV) for the mean length estimates by cruise and vessel comparing the original protocol (OP) and the modified protocol (MP) by number of pollock measured for females (a) and males (b), and by number of hauls sampled for females (c) and males (d). Original protocol and modified protocol curves are the curves as per 1999 study. The modified protocol with offset curves in section (c) and (d) incorporate the difference in the proportion of sampled hauls using the OP and MP as observed between the 1998 and 2000 fisheries.

(c)

(d)
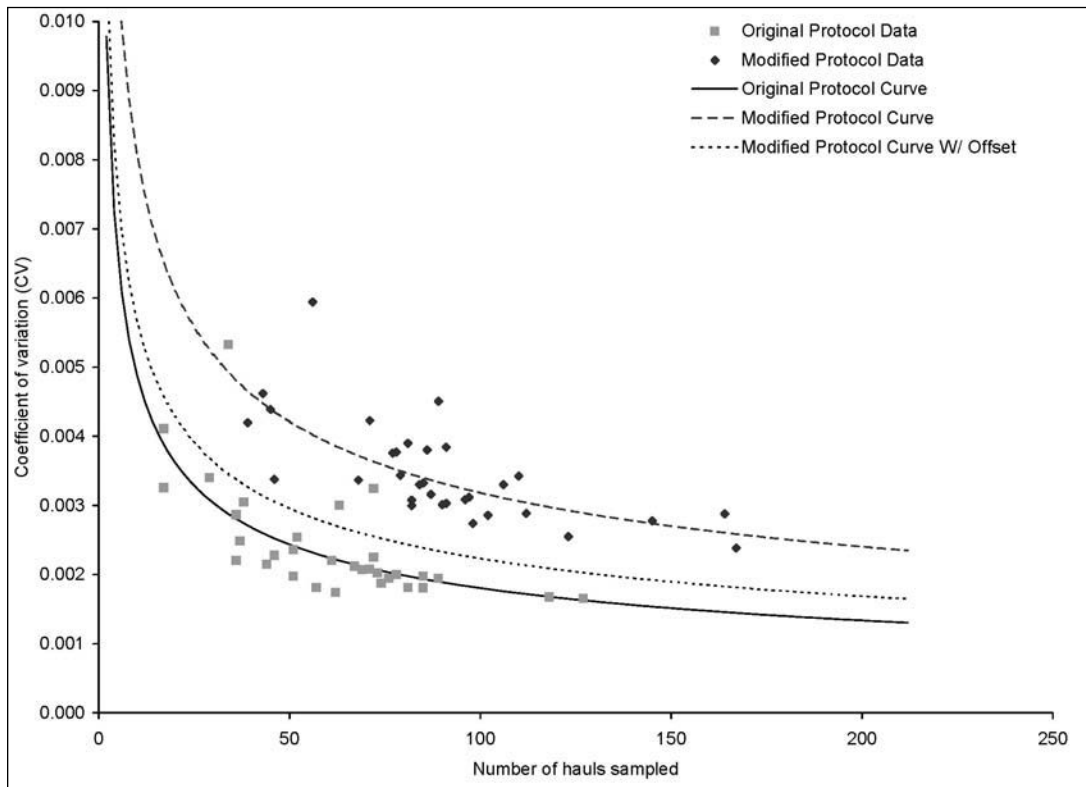
Figure 3. Continued.

Table 3. Catch-at-age in millions of pollock, standard deviation (α), and coefficients of variation (CV) for the original protocol and modified protocol for the 32 cruises.

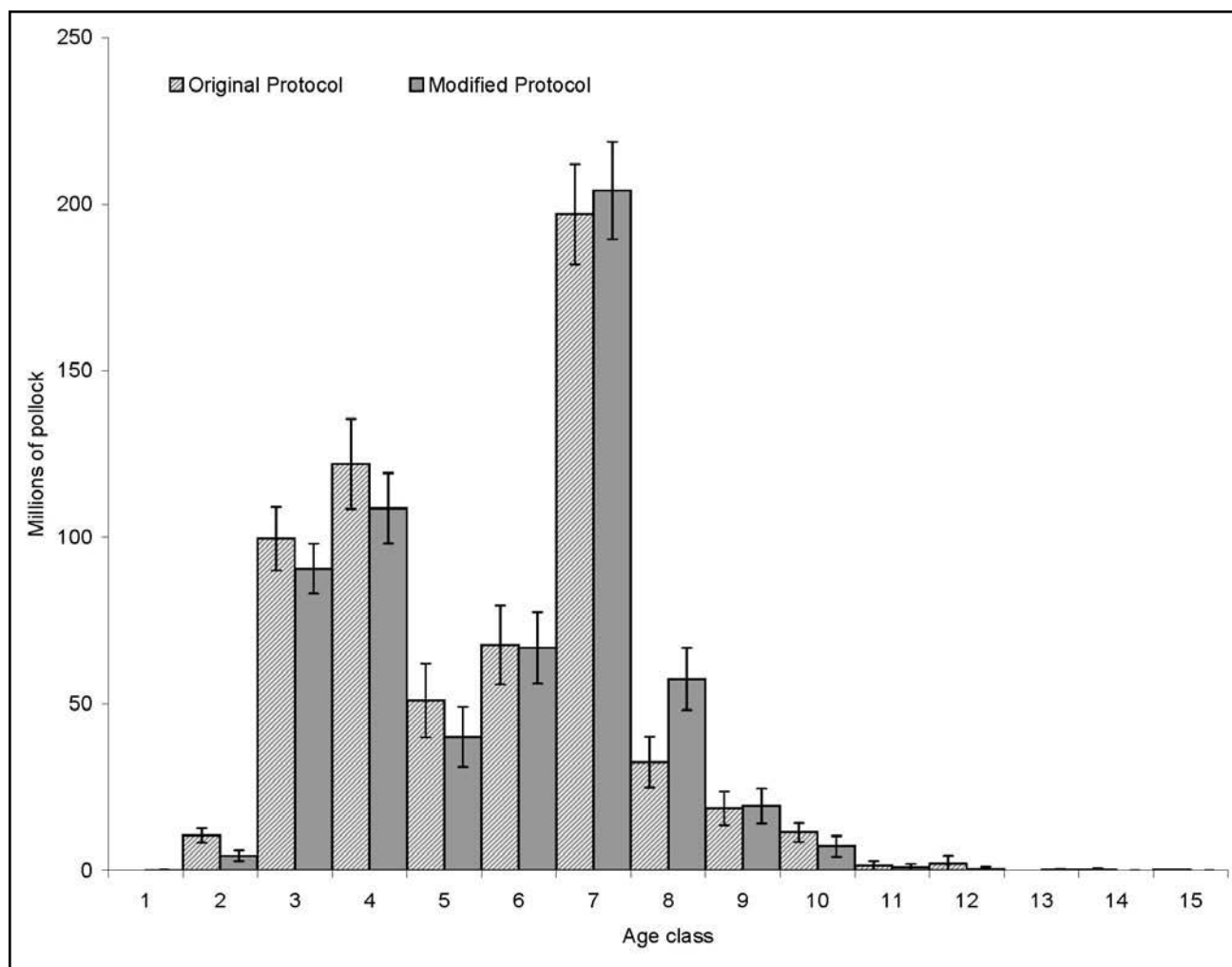| Age	Original Protocol			Modified Protocol		
	Catch-at-Age	α	CV	Catch-at-Age	α	CV
1	0.00			0.07	0.03	0.41
2	10.54	1.09	0.10	4.42	0.83	0.19
3	99.53	4.77	0.05	90.56	3.72	0.04
4	121.99	6.77	0.06	108.67	5.32	0.05
5	51.00	5.53	0.11	40.07	4.46	0.11
6	67.68	5.94	0.09	66.76	5.35	0.08
7	196.94	7.53	0.04	204.04	7.28	0.04
8	32.53	3.80	0.12	57.44	4.67	0.08
9	18.64	2.51	0.13	19.40	2.61	0.13
10	11.44	1.43	0.13	7.28	1.55	0.21
11	1.43	0.71	0.50	0.98	0.49	0.50
12	2.07	1.21	0.58	0.38	0.38	1.00
13	0.00			0.11	0.17	1.54
14	0.28	0.16	0.56	0.00		
15	0.12	0.05	0.43	0.00		

Figure 4. Catch-at-age calculated through methods described by Kimura (1989), modified by Dorn (1992), and employed by Ianelli et al. (2001) for both original protocol and modified protocol data.

Table 4. Number of aged pollock used to create age–length keys by age group and sampling protocol: original protocol (OP) and modified protocol (MP).

Age Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15+
OP	0	44	164	155	65	93	286	73	50	52	7	6	0	2	3
MP	1	10	132	157	63	108	350	103	43	20	3	2	1	0	0

samples. In 2000, the Bering Sea CP pollock fishery captured approximately 989 million pollock in 72 ADF&G management areas. Using the MP, observers in 2000 measured 190,120 pollock and collected 2,618 otolith samples. The proportion of hauls sampled for length–frequency increased from 0.40 in 1998 to 0.96 in 2000 and the proportion of hauls sampled for otoliths increased from 0.03 in 1998 to 0.15 in 2000. The distribution of these proportions was also more even within 2000 compared to 1998 (Figures 6 and 7).

Although sampling in 2000 resulted in fewer pollock measurements, comparisons of the spatial distributions of 1998 and 2000 sampling reveal that the 2000 MP produced a pollock length–frequency collection that was significantly more proportional to the pollock total catch than the 1998 OP, $\chi^2_{1998}=37,996$ and $\chi^2_{2000}=3,651$ (F=7.7 and $F_{0.05(2),87,72}=1.46$). In 1998, 16 of the 87 (18.4%) ADF&G management area cells had pollock length–frequency sampling proportional to pollock catch ($\chi<10.828$), while in 2000, 24 of the 72
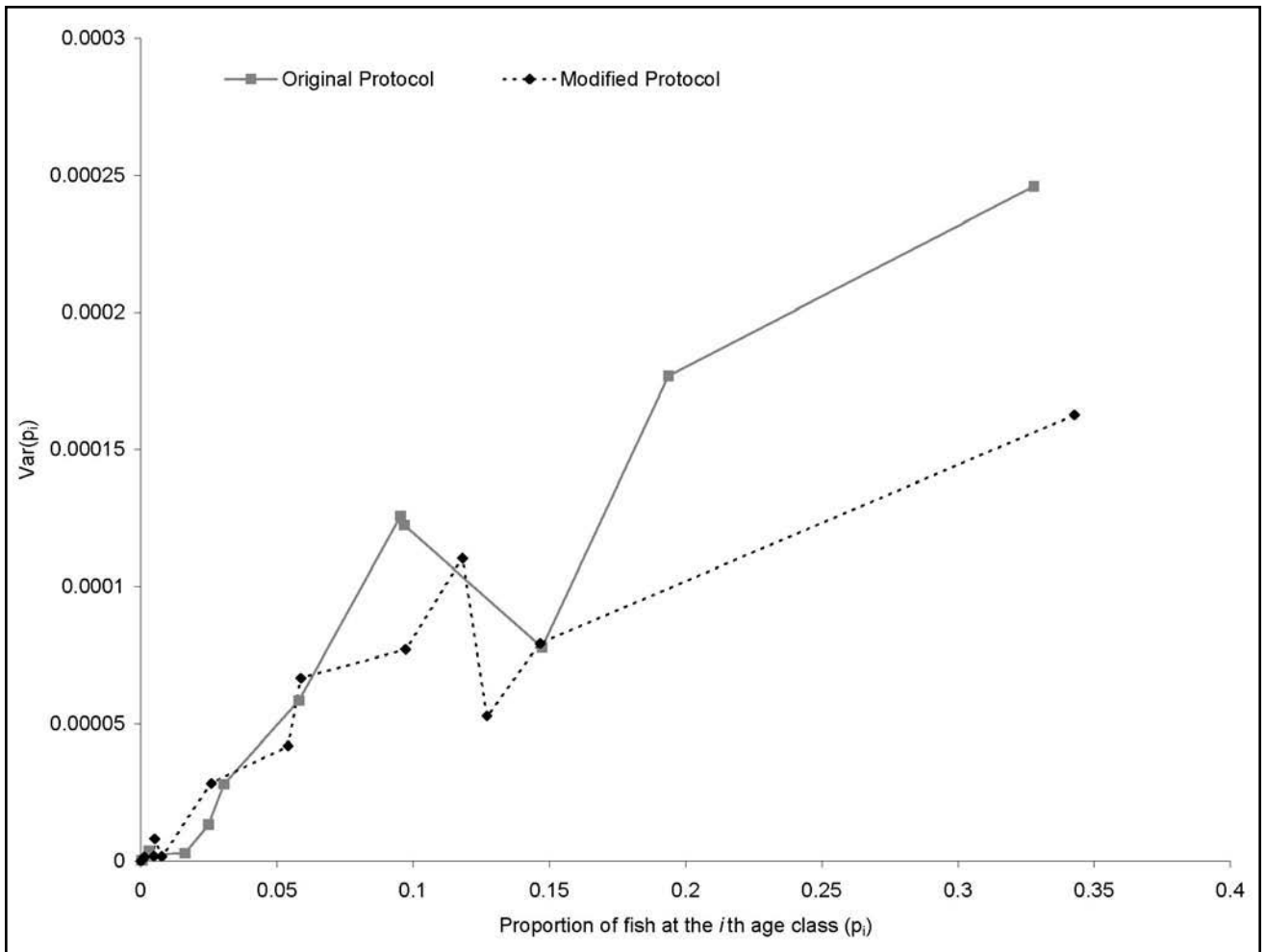

Figure 5. Variability by age group for original protocol and modified protocol age-length key estimates.
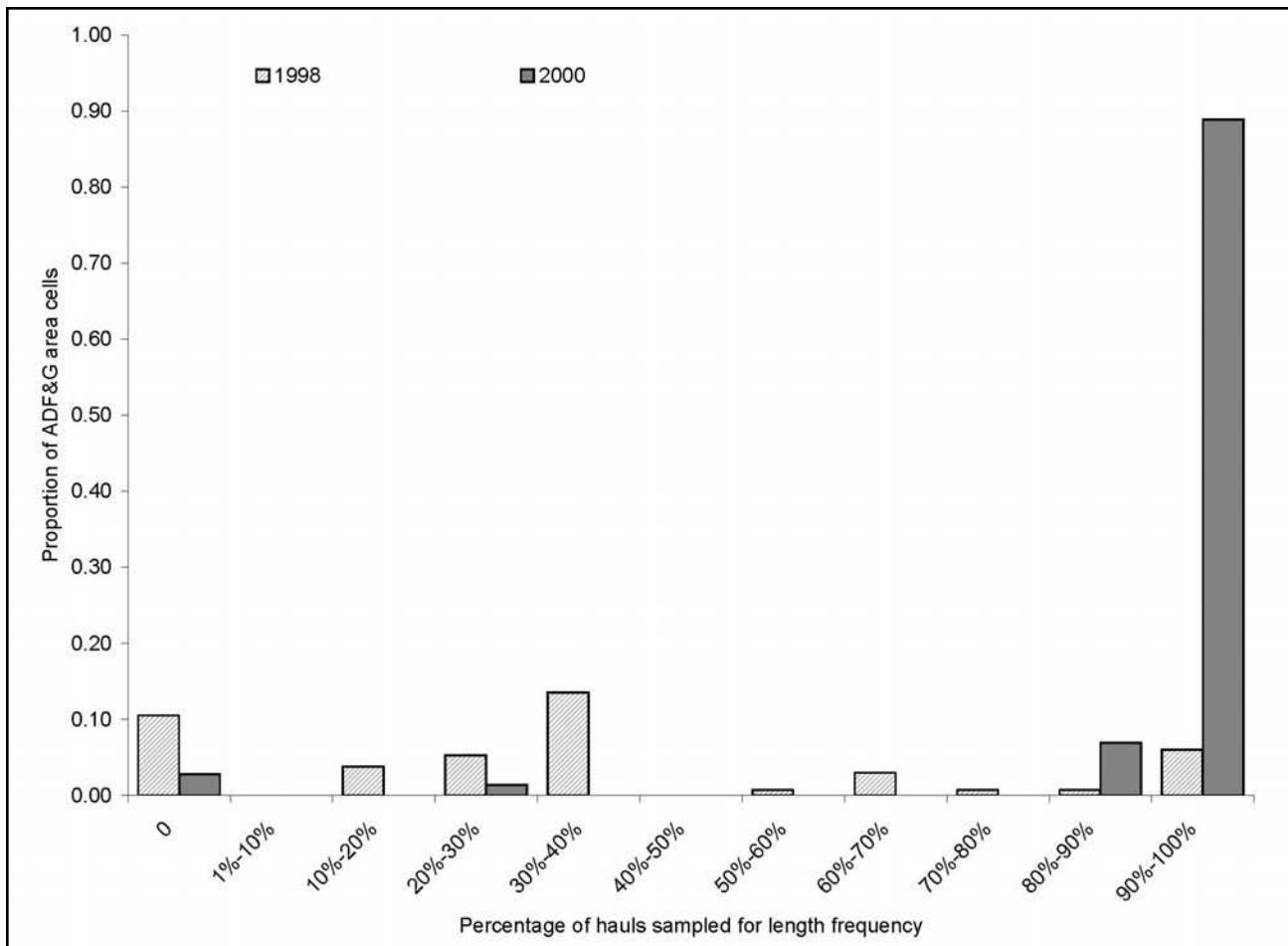
Figure 6. Frequencies of the percentage of hauls sampled for length frequency between 1998 and 2000 as a function of the proportion of ADF&G management area cells.

(33.3%) ADF&G management area cells had pollock length–frequency sampling proportional to pollock catch (Figures 8, 9). In addition, there was a larger proportion of fished ADF&G management area cells sampled for length–frequency in 2000 (97.2%) than in 1998 (83.9%).

Comparisons of the spatial distributions of 1998 and 2000 otolith sampling reveal that the 2000 MP resulted in a pollock otolith collection that was significantly more proportional to the pollock total catch than the 1998 OP, $\chi^2_{1998}=2{,}948$ and $\chi^2_{2000}=317$ (F=9.3 and $F_{0.05(2),87,72}=1.46$). In 1998, 55 of the 87 (63.2%) ADF&G management area cells had pollock otolith sampling proportional to pollock catch ($\chi<10.828$), while in 2000, 61 of the 72 (84.7%) ADF&G management area cells had pollock otolith sampling proportional to pollock catch (Figures 10, 11). In addition, there was a larger proportion of fished ADF&G

management area cells sampled for otoliths in 2000 (79.2%) than in 1998 (52.9%).

DISCUSSION

The NPGOP has been active in monitoring North Pacific fisheries since 1973 and is now one of the largest programs in the world, collecting over 35,000 sea days of data annually. In the North Pacific, the observer program has become the backbone of the fisheries management system, providing scientists and managers with data essential for monitoring the health of the resources under their stewardship. One challenge faced by the observer program is being responsive to needs of fisheries managers while maintaining consistency in the data collected. The modifications made in observer sampling in 1999 were necessary to address the evolving needs of managers and scientists. This
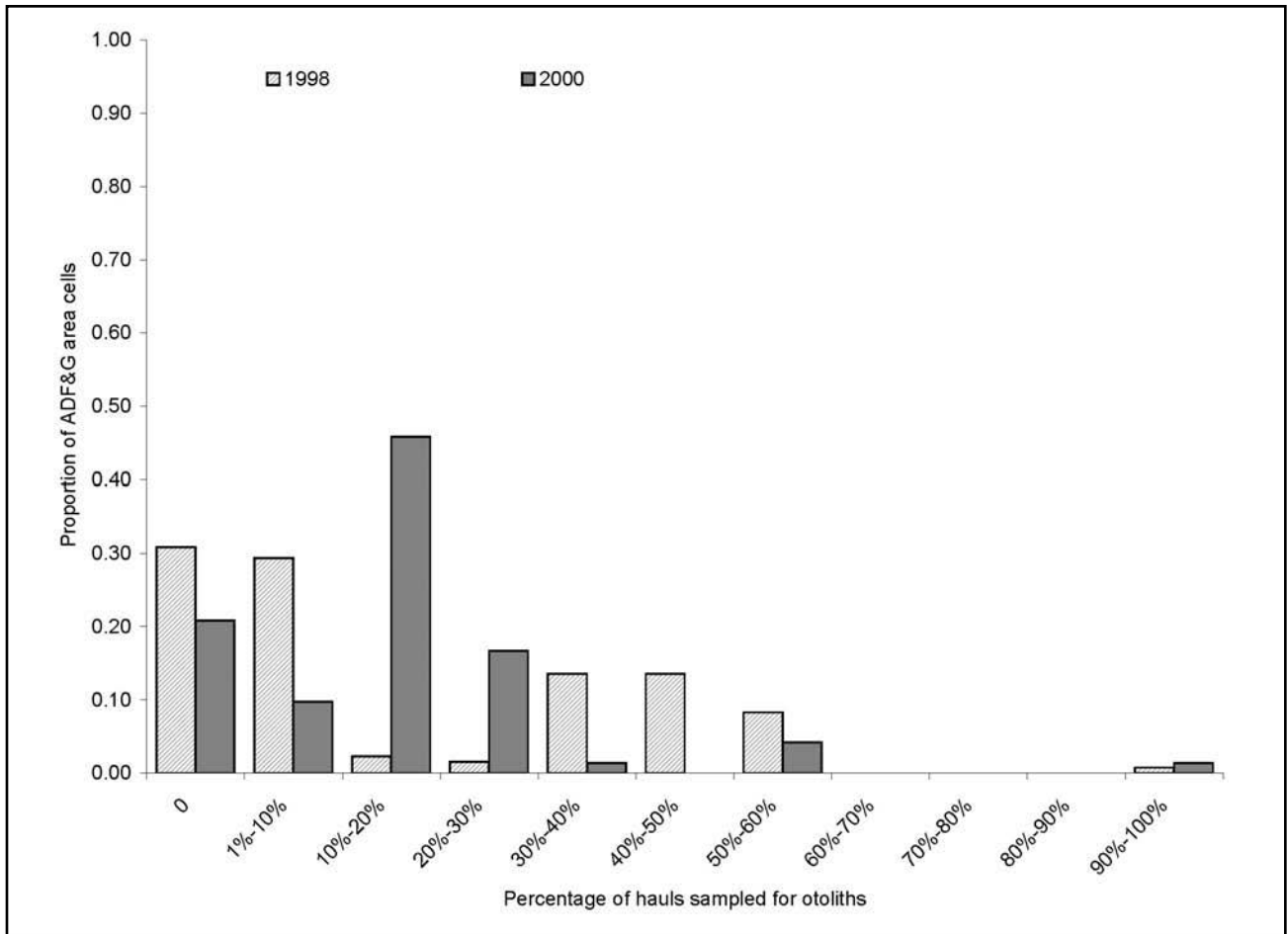
Figure 7. Frequencies of the percentage of hauls sampled for otoliths between 1998 and 2000 as a function of the proportion of ADF&G management area cells.

analysis was conducted to find out if modifications to the sampling methods resulted in inconsistencies between data collected using the OP and the MP. Our analysis has shown that the OP and MP sampling methods result in similar length–frequency distributions. Further, the diversity of length–frequency distributions and the estimate of mean length for the 32 cruises are not significantly different for the 2 sampling methods. In addition, the precision of the mean length estimates are almost identical (mean CV within 0.0012) for the OP and MP. Therefore, the application of the length–frequency data from the 2 sampling methods should produce similar results.

The method for estimating catch-at-age using age–length keys as described by Kimura (1989) and updated by Dorn (1992) assumes that the length–frequency samples are a simple random sample of the total catch in a stratum and that the number of samples in a length–frequency collection is proportional to the number of pollock caught in each haul. In the Bering Sea pollock fishery these assumptions may not be correct. In the OP length–frequency data collection method, length–frequency collections tended to be collected in areas where there was a greater amount of fishing; outlying areas with fewer hauls were less likely to be sampled. This resulted in many marginal areas not being sampled for length–frequencies, and thus the length distribution of these areas were not represented in catch-at-age estimations. Ianelli (2002) found that pollock in the EBS are distributed by size, and that length–frequency distributions are markedly disparate for different areas of the EBS. The length–frequency distributions from areas with lower fishing pressure could be significantly different than those areas with high fishing pressure. Underrepresentation of pollock distributions from areas with lower fishing pressure
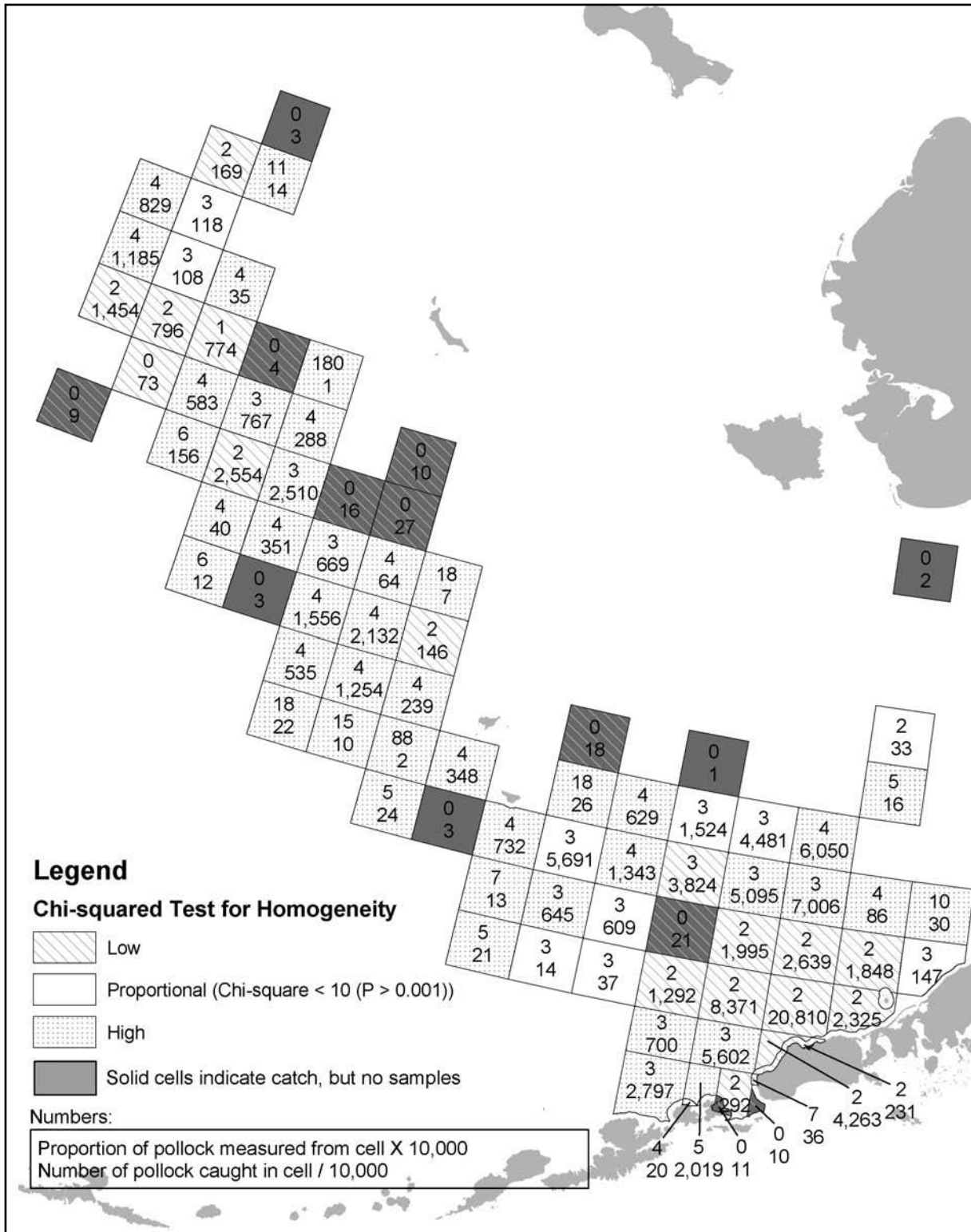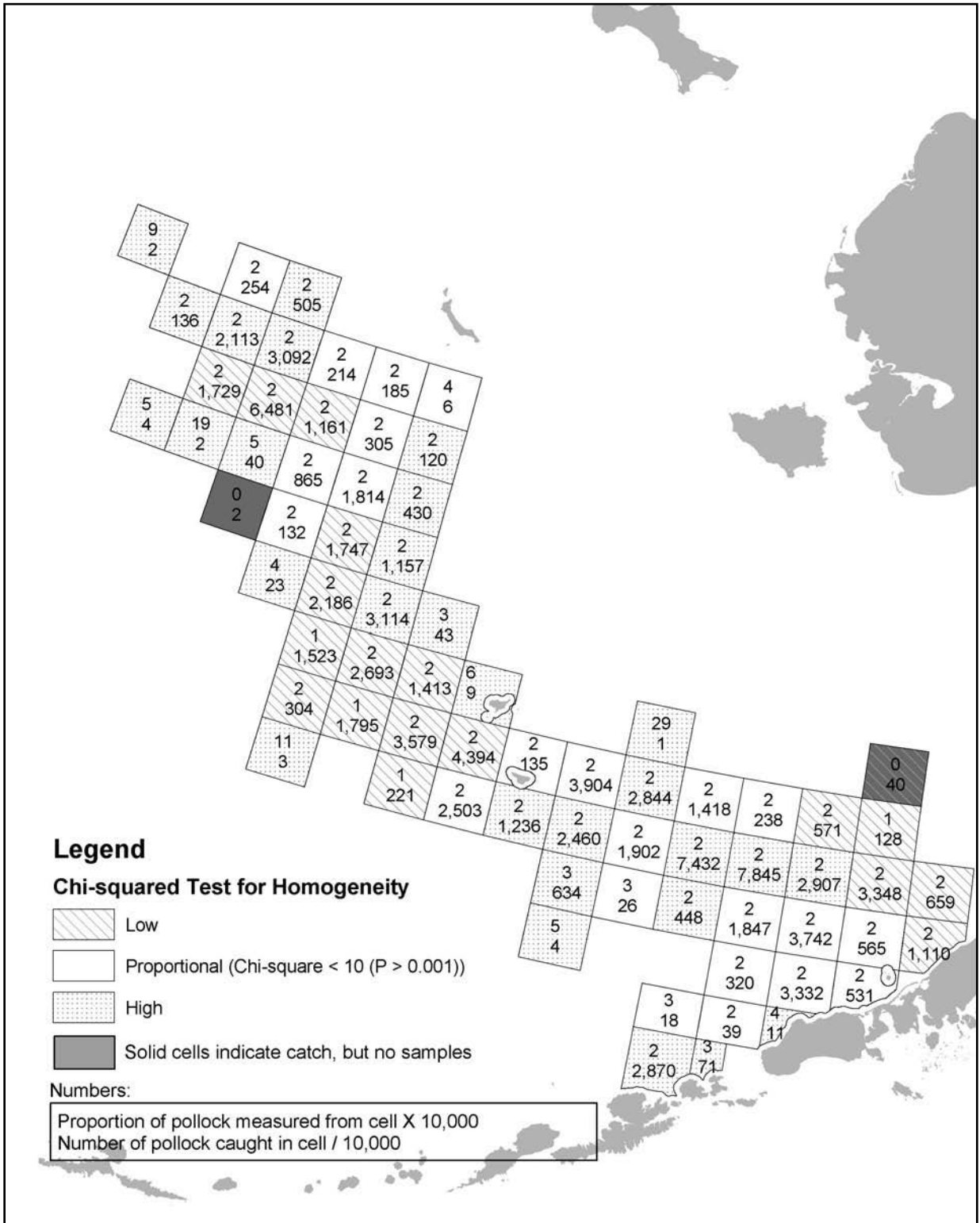
Figure 8. Chi-squared analysis for proportionality of length–frequency sampling using the original protocol for 1998 by ADF&G management area cells. Top number in each cell is the proportion of fish measured to fish caught in the cell multiplied by 10,000. The bottom number in each cell is the total number of fish caught in an area divided by 10,000. Darkened cells indicate cells with catch, but without length–frequency samples. High and Low indicate whether sampling was higher or lower than proportional for each cell.
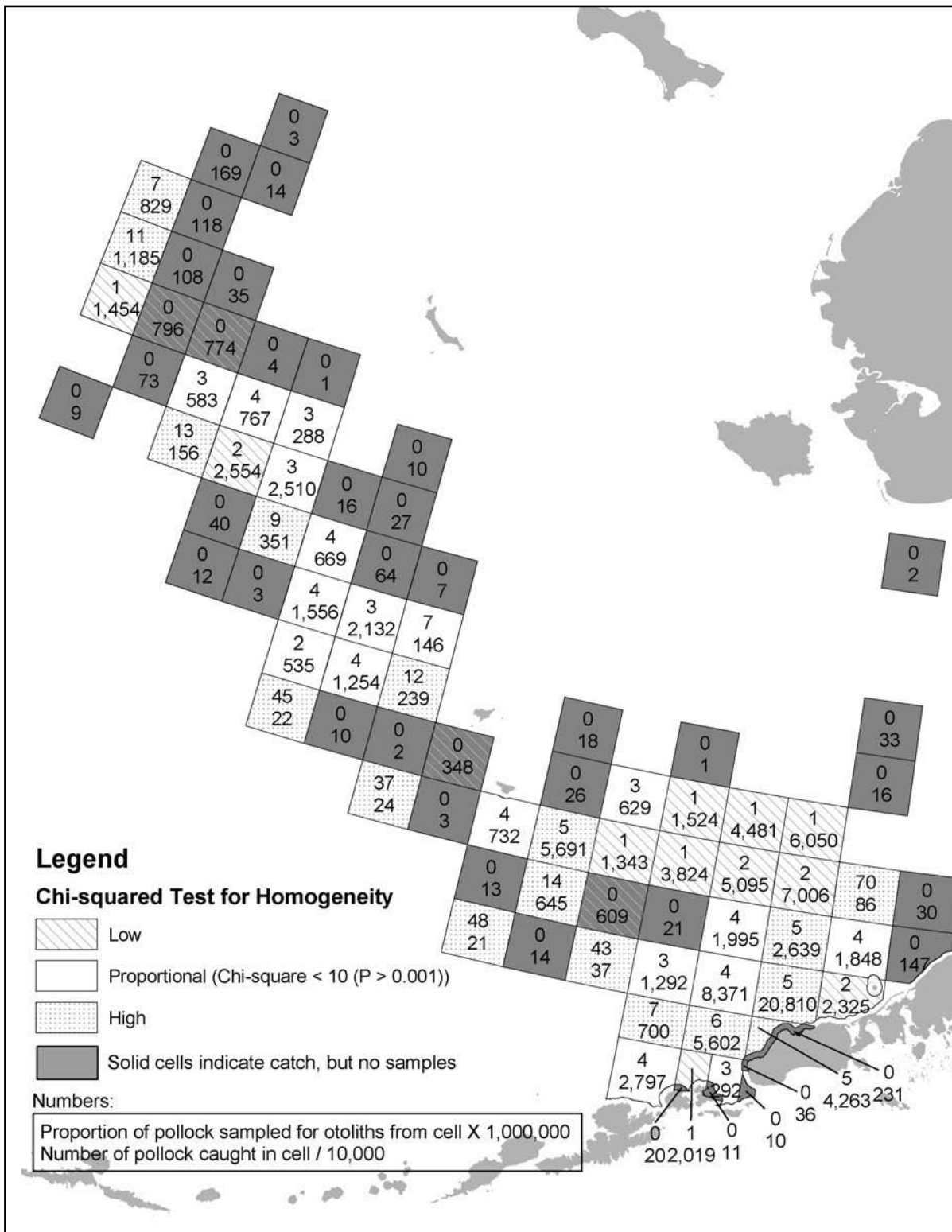
Figure 9. Chi-squared analysis for proportionality of length–frequency sampling using the modified protocol for 2000 by ADF&G management area cells. Top number in each cell is the proportion of fish measured to the total number of fish caught in the cell multiplied by 10,000. The bottom number in each cell is the total number of fish caught in an area divided by 10,000. Darkened cells indicate cells with catch, but without length–frequency samples. High and Low indicate whether sampling was higher or lower than proportional for each cell.
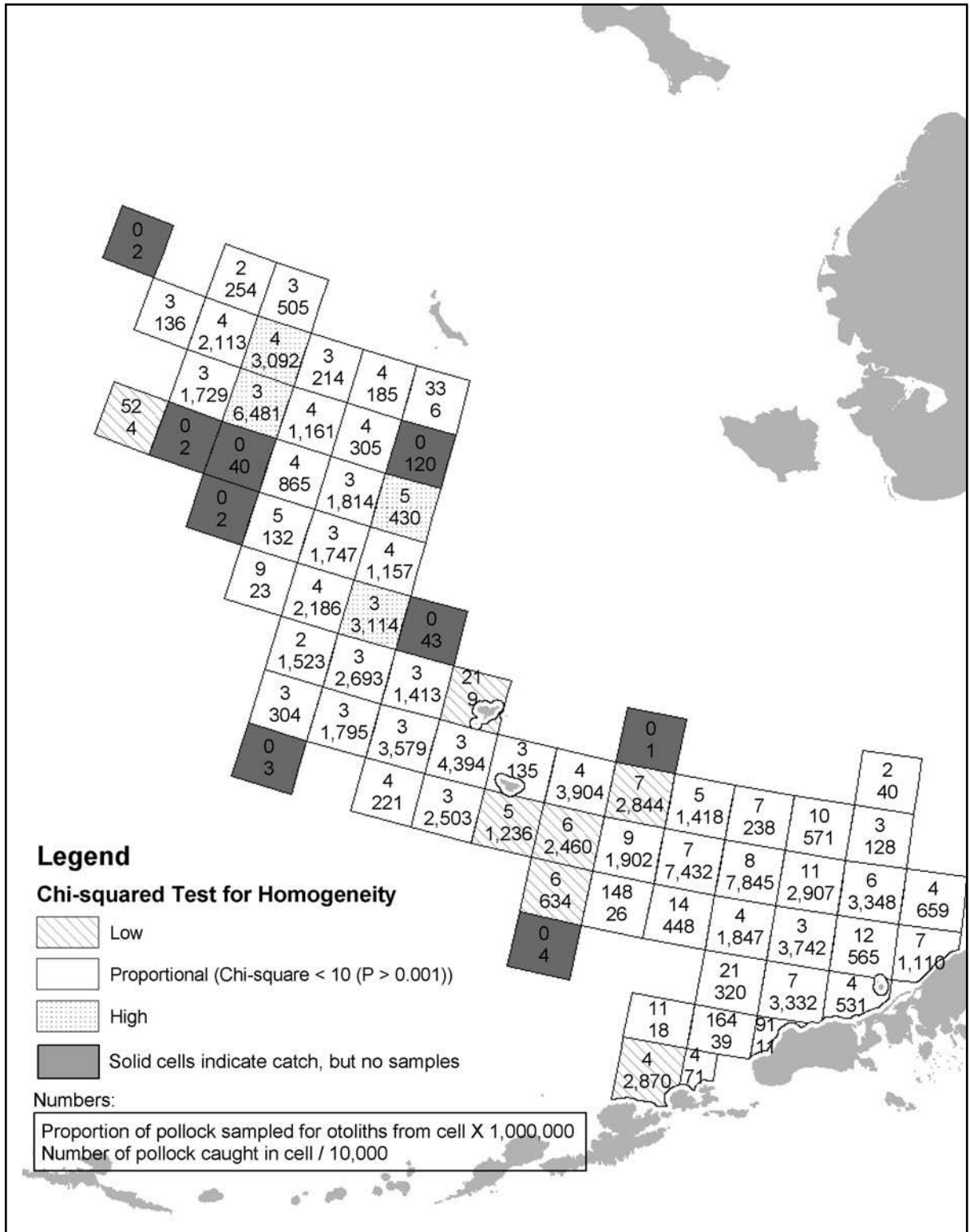
Figure 10. Chi-squared analysis for proportionality of otolith sampling using the original protocol for 1998 by ADF&G management area cells. Top number in each cell is the proportion of otoliths collected to the total number of fish caught in the cell multiplied by 1,000,000. The bottom number in each cell is the total number of fish caught in an area divided by 10,000. Darkened cells indicate cells with catch, but without otolith samples. High and Low indicate whether sampling was higher or lower than proportional for each cell.

Figure 11. Chi-squared analysis for proportionality otolith sampling using the modified protocol for 2000 by ADF&G management area cells. Top number in the cell is the proportion of otoliths collected to the total number of fish caught in the cell multiplied by 1,000,000. The bottom number in each cell is the total number of fish caught in an area divided by 10,000. Darkened cells indicate cells with catch, but without otolith samples. High and Low indicate whether sampling was higher or lower than proportional for each cell.

and over-representation of the length–frequency distributions from areas with greater fishing pressure could bias the catch-at-age estimates.

When employing the MP, the number of fish measured in each length–frequency collection is the same for every haul and the total length–frequency collection is proportional to the total number of hauls and deliveries in a stratum. The MP provides length–frequency samples that are more proportional to catch, which results in a wider spatial distribution of samples. This should provide a more accurate representation of the distribution of pollock in the overall catch. Although the MP provides a better representation of the overall length, weight, and age distribution in a stratum than the OP, there still may be problems in assuming that these samples are simple random samples without weighting the proportion of the catch they represent. Where there are many small hauls, there may be over-sampling, and where there are few large hauls, there may be under-sampling in relation to our assumptions. These potential biases can more easily be accounted for in the MP by either increasing the number of strata in our estimations of catch at age or by applying a 2-staged sampling analysis in our catch-at-age estimations and weighting the length–frequency samples by the proportion of catch they represent.

The 1999 changes in sampling methodology hardly affected the precision of the estimates of the catch's biological characteristics, but did improve the spatial distribution of length and otolith sampling. Given that pollock distribution is not homogenous and that pollock tend to distribute spatially based on size (Ianelli 2002), improved spatial distribution of biological sampling should increase the overall accuracy of estimates. The MP employed in 1999 did not significantly reduce the precision of estimates and may have improved the accuracy of estimates. From this study, we conclude that the implementation of the MP has not adversely affected data quality for stock assessment purposes. Coupled with the increase in observer efficiency identified by MRAG, we further conclude that the MP has improved data collection in the NPGOP.

LITERATURE CITED

Cochran, W.G. 1977. Sampling Techniques, 3rd edition. New York: John Wiley and Sons.

Dorn, M.W. 1992. Detecting environmental covariates of Pacific whiting *Merluccius productus* growth using a growth-increment regression model. Fishery Bulletin 90: 260–275.

Hollowed, A. B., E. Brown, P. Livingston, B. A. Megrey and C. Wilson. 1995. Walleye pollock *in* Stock assessment and fishery evaluation report for Gulf of Alaska as projected for 1996. Prepared by the Gulf of Alaska Groundfish Plan Team, North Pacific Fishery Management Council, P.O. Box 103136, Anchorage, AK 99510.

Ianelli, J. N., T. Buckley, T. Honkalehto, N. Williamson and G. Walters. 2001. Bering Sea-Aleutian Islands walleye pollock assessment for 2002 *in* Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage, AK 99510.

Ianelli, J. N. 2002. Bering Sea walleye pollock stock structure using morphometric methods *in* A. Nishimura, editor. Papers from "Pollock Stock Structure and Identification Workshop" Yokohama, Japan 7–9 September 1999. Technical Reports of the Hokkaido National Fisheries Research Institute 5:53–58.

Kimura, D. K. 1977. Statistical assessment of age–length key. Journal of the Fisheries Research Board of Canada 31: 317–324.

Kimura, D. K. 1989. Variability in estimating catch-in-numbers-at-age and its impact on cohort analysis *in* R. J. Beamish and G. A. McFarlane, editors. Effects on ocean variability on recruitment and an evaluation of parameters used in stock assessment models. Canadian Special Publication of Fisheries and Aquatic Sciences 108:57–66.

Lai, H. L. 1987. Optimum allocation for estimating age composition using age–length key. Fishery Bulletin 85: 179–185.

MRAG (Marine Resources Assessment Group) Americas, Inc. 2003. Evaluation and analysis of current field sampling in North Pacific groundfish fisheries—task 1: biological sampling protocols final report, April 2003. North Pacific Groundfish Observer Program, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle WA 98115.

NMFS (National Marine Fisheries Service). 2000. Management control review of National Marine Fisheries Service observer programs/service delivery models. North Pacific Groundfish Observer Program, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle WA 98115.

NPGOP (North Pacific Groundfish Observer Program) 1997. 1998 observer sampling manual. North Pacific Groundfish Observer Program, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle WA 98115.

NPGOP. 1998. 1999 observer sampling manual. North Pacific Groundfish Observer Program, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle WA 98115.

Terry, J. *In Press*. Evaluation bycatch: A national approach to standardized bycatch monitoring programs, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle WA 98115.

Thompson, G., H. H. Zenger, and M.W. Dorn. 2001. Assessment of the Pacific cod stock in the Gulf of Alaska *in* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska Section 2: 2-1-2-73. North Pacific Fishery Management Council. Anchorage, AK.

Volstad, J.H., W. Richkus, S. Gaurin, and R. Easton. 1997. Analytical and statistical review of procedures for collection and analysis of commercial fishery data used for management and assessment of groundfish stocks in the U.S. Exclusive Economic Zone off Alaska, October 1997. Versar, Inc. ESM Operations 9200 Rumsey Road Columbia, MD 21045. Available: North Pacific Groundfish Observer Program, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle WA 98115.

Zar, J. H. 1999. Biostatistical Analysis, 4th Edition. Prentice Hall. Upper Saddle River, New Jersey.