Variability and precision in the aging of halibut otoliths

Steven R. Hare

Abstract

The utility of "quality control" second readings of halibut otoliths is considered. It is recommended that 10% of halibut otoliths continue to be re-aged as a standard method of ensuring consistency in aging over time. However, given the importance of the age data for the annual stock assessment, and the relatively minor differences between initial and second readings, higher priority should be given to enlarging the sample of first readings in time for the annual stock assessment.

Introduction

The annual stock assessment is highly reliant upon the age data collected both on the grid survey and from the commercial fishery. Since 2003 all otolith aging has been done using the break and burn procedure, superseding surface reading. The IPHC aging protocol is thoroughly described in Forsberg (2001). Sampling protocol has remained basically unchanged since 1990 for market sample (MS) otoliths and 1997 for the stock assessment survey (GS, for General Series) otoliths. The goal is to obtain a sample of 2000 otoliths per regulatory area from both the fishery and the survey (Quinn et al. 1983, Clark et al. 2000). In 2003, the goal for MS otoliths was dropped to 1500 and in 2A the goal has been 1000 MS otoliths since around 1995.

Every year, the collected otoliths are aged as rapidly as possible in order to have estimated age compositions ready for the annual stock assessment. The general deadline has been around Oct. 15th in order to allow a full month of assessment modeling prior to the IPHC Interim Meeting when the staff quota recommendations are presented. This deadline has been somewhat problematic for the aging staff, particularly as concerns the GS otoliths since many of the samples do not arrive until after the survey season is completed, usually in late August. The rush to complete initial reading of the GS and MS otoliths also has the effect of delaying quality control second readings until the following year. The aging unit sets a goal of re-aging 10% of both sets of otoliths. Beginning in 2006, a third reading was done if the first two readings differed by more than one year. A summary of the number of otoliths aged in support of the stock assessment is given in Table 1.

Re-aging of the otoliths, by a second reader, is considered an important component of quality control. Accommodating an "in-season" re-aging schedule would require a) another reader in the age lab; b) a delayed deadline for passing the aging data to the assessment team; c) reducing either the market or survey sample sizes by up to 10%. At present it is not likely that a new age reader will be brought on staff. The regulatory area goal of 2000 otoliths for survey age samples and 1500 for market samples have been long established and lowering those targets should be a measure of last resort. Before any such consideration it would be instructive to review how variable recent aging has been for break and burn otoliths and, in particular, to document how different "resolved" age distributions are from the initial distributions used in the stock assessment. This review also provides an opportunity to examine whether aging variability is greater among certain regulatory areas, and how it compares with previous examinations of this subject.

Variability of initial age readings

In general the stated collection goals have been met in most areas for both GS and MS otoliths (Table 1). Where the actual collections have fallen short, the greatest shortages have been in Area 4D, particularly the GS otoliths in 2007.

The precision of the age samples by regulatory area for the initial readings of otoliths is usefully described by the coefficient of variation (CV) of the proportion at age. Plots of CVs by area, age, sex (GS otoliths only) for 2006 and 2007 are illustrated in Figs. 1 and 2 for GS otoliths and Figs. 3 and 4 for MS otoliths. There are general trends in the CVs for both sample types and in all regulatory areas. CVs are around 10 percent for the most common ages in the samples, ages 8 to around 15. CVs increase sharply at the younger and older ages. The MS CVs are based on smaller sample sizes, however as the fish cannot be sexed the CVs are generally a bit lower than the (sexed) survey CVs. CVs for older fish in Area 4 are generally lower than for Area 2 and 3 fish, reflecting the greater proportion of older fish caught in both the survey and commercial fishery for Area 4.

There is no standard CV target for age proportions. As one of the core components of the stock assessment, there is a great deal of reliance on the age data. Reference to a target CV of 10% for the most common age groups is made in Quinn and Deriso (1999). That is achieved at the current level of sampling. Previous Monte Carlo simulations have suggested that a target of less that 2000 otoliths per regulatory area might be supportable (Clark and Vienneau 1990), however that analysis was based on unsexed otoliths. The current standard of 2000 seems reasonable to maintain given that proportions at age are now required for both sexes – and a reduction to a target of 1500 has already been set for market sample otoliths. In summary, age proportion CVs are at an appropriate level and any reduction in sample size will have the immediate effect of increasing the CVs.

Variability between first and second readings

For 2006 and 2007 combined, a total of 5619 otoliths were re-aged by a second reader (this total does not include otoliths from Areas 4C, 4E, or the closed area). This represents slightly more than 10% of the 55,916 otoliths read a first time. Second readings increased for both GS (1477 to 1518) and MS (1192 to 1432) otoliths between 2006 and 2007.

Three common statistical measures were used to determine the amount of variation between the two sets of readings: coefficient of variation (CV), absolute percent error (APE), and percent agreement (PA) within one year (+/- 1 yr). Formulas and citations are listed in Appendix 1. Age-specific results are plotted in Fig. 5 (GS otoliths) and Fig. 6 (MS otoliths). Annual summaries of APE and CV are listed in Table 2. All summaries and statistics are based on ages 7-25 (initial reading) as sample sizes are often very small outside this age range. This restriction eliminates about 5% of the double readings.

Averaged across 2006 and 2007, CV was 3.83% and 3.79% for GS and MS otoliths, respectively. APE averaged 2.71% (GS otoliths) and 2.68% (MS otoliths). PA (+/- 1 yr) for double readings was 88.6% (GS otoliths) and 87.0% (MS otoliths). The annual average CVs and APEs increased from 2006 to 2007 for both GS and MS otoliths, while the PAs declined by about 3% for both otoliths samples. The standards, however, were still well above those established for halibut aging and listed in Forsberg (2001), i.e., that CVs be below 7.0 and PA be above 75%. These standards are also generally met when examined across ages. There is little trend in either CVs or APEs with increasing age, though PA starts to decline somewhat sharply after age 20.

There appears to be little evidence of systematic differences between the first and second readings of an otolith. Panel **d** in Figs. 5 and 6 shows all double readings and the spread of points around the agreed age line appears to be symmetric. A closer, area by area, illustration of aging differences for both GS and MS otoliths is given in Figs. 7-10. These figures also list sample size, APE and CV by area and otolith source. Most all plots show a near normal distribution around an aging difference of zero, with a more or less equal number of areas showing a greater number of positive age differences and negative age differences. Area 4D appears to be the most skewed and variable among the areas, with 2B the second most variable.

When it comes to impact on the assessment, the bottom line is how different a “resolved” age distribution is compared to an initial age distribution. In other words, is it better to obtain as complete an initial set of age readings as possible, or is there evidence that the initial set may be in error such that they need quality control prior to use in the assessment. The desire to have as large an age sample as possible was stated earlier. A second bit of evidence is to examine the initial and the resolved age distributions. To iterate, the resolved distribution is comprised of initial readings, second readings (for readings where the 1st and second reading differ by one year) and third readings (where the 1st and 2nd readings differed by more than one year). Plots of the initial and resolved age compositions are given for GS otoliths in Figs. 11 and 12, and for MS otoliths in Figs. 13 and 14. There is no proper statistical test to compare whether the initial and resolved age compositions differ, however close examination does not suggest much difference.

Conclusions

This review was conducted to examine how much variability exists in aging of halibut and to determine whether it is preferable to have a larger set of initial readings, or a smaller set of resolved readings, available for the annual stock assessment. I conclude that the assessment would benefit more from the larger set of initial readings. This conclusion is based on the relatively small variability in double readings, the lack of bias between first and second readings, and the consistent similarity between the initial and resolved age distributions. The APEs and CVs shown here for halibut compare very favorably to those published for groundfish aged at the Alaska Fishery Science Center[1]. Averaged across 22 groundfish species, APE and CV were 4.2 (sd=2.2) and 6.0 (sd=3.1), respectively – compared to values of around 2.7 and 3.8 for halibut. Further, previous studies on halibut have similarly found little increase in the precision of estimated age composition by double readings (Smith 1988, Clark and Vienneau 1991).

The importance of double readings is still a very important aspect of aging. Ensuring that multiple age readers are consistent in their aging methods, both among themselves and over time, is critical to maintaining the consistency of the age readings. This is accomplished both through double readings as well as ongoing testing against a reference set of known age (or agreed upon) otoliths. The increase in both APE and CV between 2006 and 2007 is worth noting, particularly if the trend should continue. The argument presented here is that the verification process should take place at a time other than during the rush to complete the initial age readings in time for the annual stock assessment.

---

[1] http://www.afsc.noaa.gov/refm/age/Stats/Precision_Stats.htm

Appendix

Statistics used to measure variability in double readings of otoliths.

Average Percent Error (APE, Beamish and Fournier 1981) is computed as:

$$APE_j = 100 \times \frac{1}{R} \sum_{i=1}^{R} \frac{\left| x_{ij} - x_j \right|}{\overline{x}_j}$$

$$APE = \frac{\sum_{j=1}^{n} APE_j}{n}$$

Where x_{ij} is the *i*th age determination of the *j*th fish, $\overline{x}_j$ is the mean age of the readings and R is the number of readings. APE can be averaged across fish (by area, age, etc.) to provide an average percent error.

Coefficient of Variation (CV, Kimura and Lyons 1991) is computed as:

$$CV_j = 100 \times \frac{\sqrt{\sum_{i=1}^{R} \frac{\left(x_{ij} - x_j\right)^2}{R-1}}}{\overline{x}_j}$$

$$CV = \frac{\sum_{j=1}^{n} CV_j}{n}$$

Where the components have the same definitions as those listed for APE. Similarly, CV can be averaged across fish to provide an average coefficient of variation.

Percent Agreement (+/- 1 yr) (PA) is defined as:

$$PA = \frac{\sum \left| n_{diff} \leq 1 \right|}{n}$$

Where n_{diff} is the difference in age determination between the first and second readings. PA can be determined for any set of fish.

Reference

Beamish, R.J. and Fournier, D.A. 1981. A method for comparing the precision of a set of age determinations. Can. J. Fish. Aquat. Sci. 38: 982-983.

Clark, W.G. and B.A. Vienneau. 1991. Evaluation of otolith sample size and double reading. IPHC Report of Commission Activities 1991: 229-230. Int. Pac. Halibut Comm., Seattle (mimeo).

Clark, W.G., Vienneau, B.A., Blood, C.L., and Forsberg, J.E. 2000. A review of IPHC catch sampling for age and size composition from 1935 through 1999, including estimates for the years 1963-1990. Int. Pac. Halibut Comm. Tech. Rep. 42.

Forsberg, J.E.. 2006. Aging manual for Pacific halibut: procedures and methods used at the International Pacific Halibut Commission (IPHC). Int. Pac. Halibut Comm. Tech. Rep. 46.

Kimura, D.K. and Lyons, J.L. 1991. Between-reader bias and variability in the age-determination process. Fish. Bull. 89: 53-60.

Quinn, T.J. II, Best, E.A., Bijsterveld, L. and MacGregor, I.R. 1983. Sampling Pacific halibut (*Hippoglossus stenolepis*) landings for age composition: history, evaluation, and estimation. Int. Pac. Halibut Comm. Sci. Rep. 68.

Quinn, T.J. II and Deriso, R.B. 1999. Quantitative Fish dynamics. Oxford University Press, New York. 542 pp.

Smith, P.J. 1988. Optimal two phase sampling for estimating the exploitable biomass of halibut accounting for nonsampling error. IPHC 1988 Stock assessment Document II: 1-10. Int. Pac. Halibut Comm., Seattle (mimeo).

Table 1. Age readings done in support of the annual stock assessment. BB1 is number of initial break and burn readings, BB2 is number of second break and burn readings and QC is the number of third readings ("qualith control") done. QC readings are done when BB1 and BB2 differ by more than 1 year.

Survey samples (GS)						
	2006			2007		
Area	BB1	BB2	QC	BB1	BB2	QC
2A	745	73	17	1098	109	31
2B	2035	212	44	2193	216	82
2C	2168	222	36	2151	210	61
3A	2056	204	41	2459	247	70
3B	2048	205	44	2584	256	64
4A	1704	174	32	2164	212	57
4B	2220	283	30	2043	186	45
4D	1125	104	35	812	82	28
Total	14101	1477	279	15504	1518	438

Market samples (MS)						
	2006			2007		
Area	BB1	BB2	QC	BB1	BB2	QC
2A	1441	140	30	1003	98	19
2B	1827	164	29	1772	166	54
2C	1800	160	20	1583	160	27
3A	2535	236	36	2385	322	72
3B	1682	153	24	1305	162	36
4A	1424	130	20	1957	187	31
4B	1412	128	15	1504	150	31
4D	835	81	22	1846	187	41
Total	12956	1192	196	13355	1432	311

Table 2. Summary of precision in double readings of halibut otoliths. APE is absolute precision error and CV is coefficient of variation, PA is Percent Agreemnt plus or minus 1 year.

Sample	Measure	2006	2007	Combined
Survey	APE	2.21	3.17	2.71
Survey	CV	3.13	4.49	3.83
Survey	PA	90.2	87.1	88.6
Market	APE	2.46	2.86	2.68
Market	CV	3.47	4.05	3.79
Market	PA	88.6	85.7	87.0

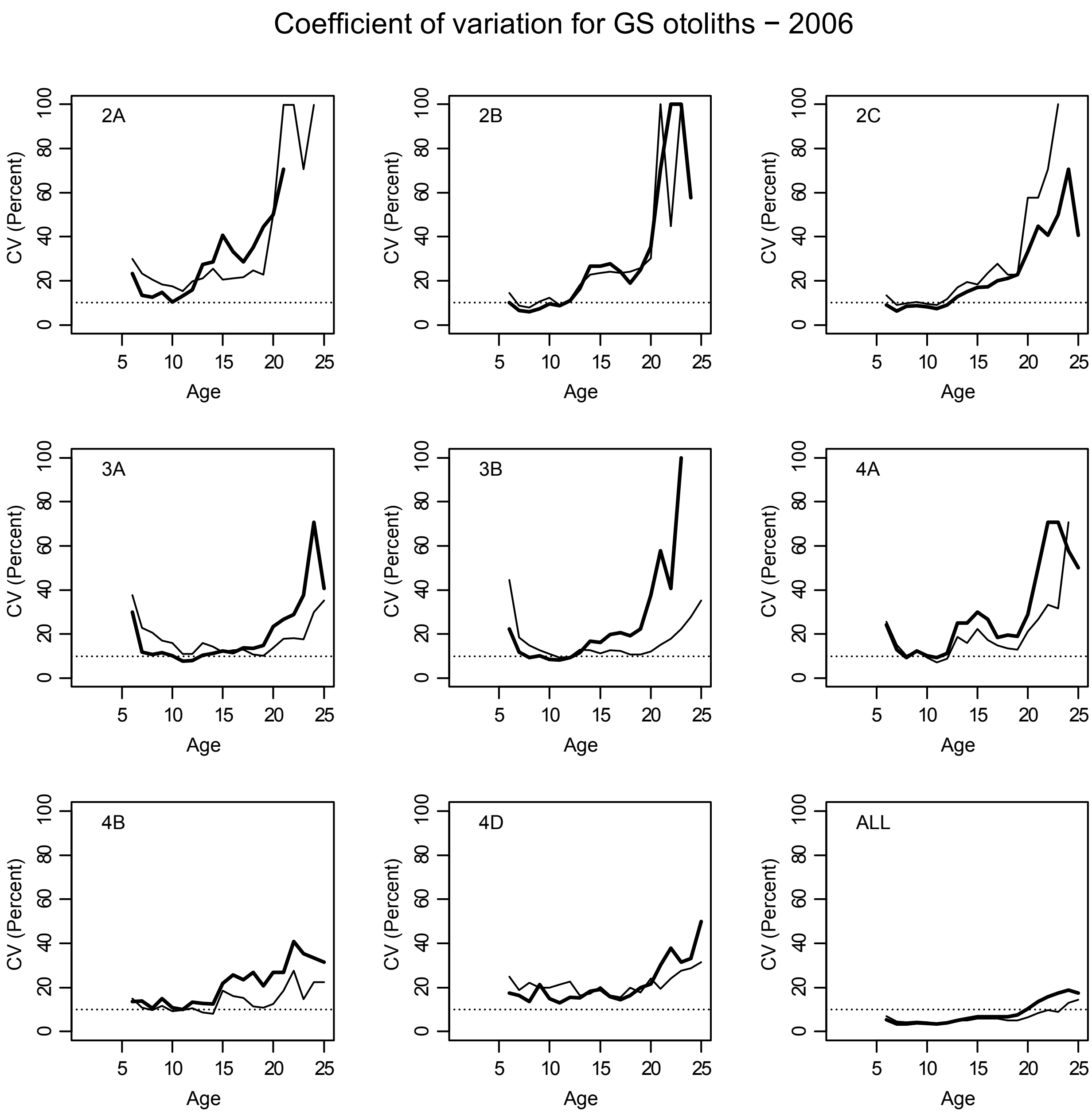

Figure 1. Coefficient of variation for initial readings of halibut, ages 6-25 for General Series (GS) otoliths collected on the 2006 grid survey. Females are indicated by the thick line, males by the thin line and the dashed line indicates a CV of 10%.

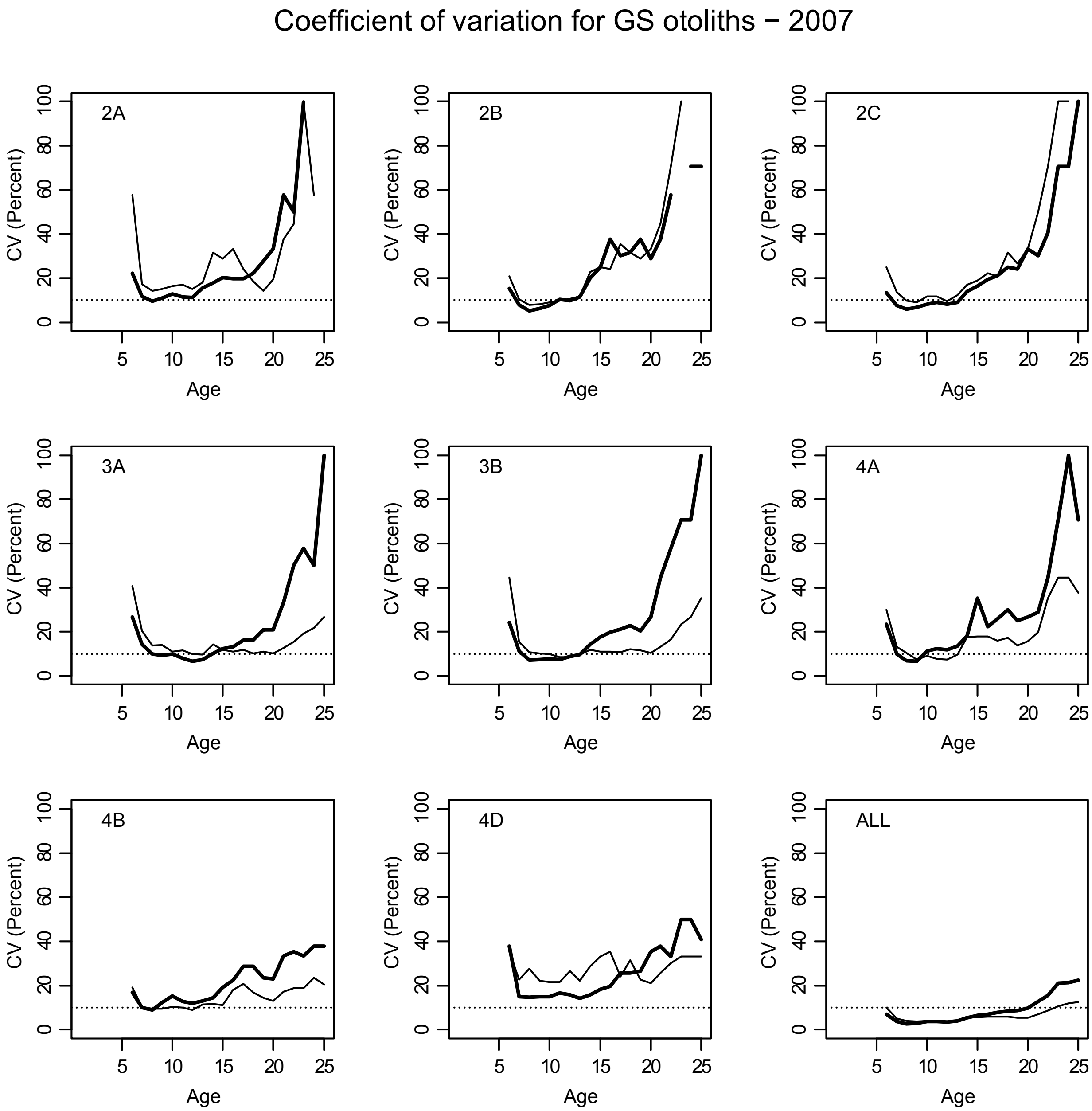

Figure 2. Coefficient of variation for initial readings of halibut, ages 6-25 for General Series (GS) otoliths collected on the 2007 grid survey. Females are indicated by the thick line, males by the thin line and the dashed line indicates a CV of 10%.

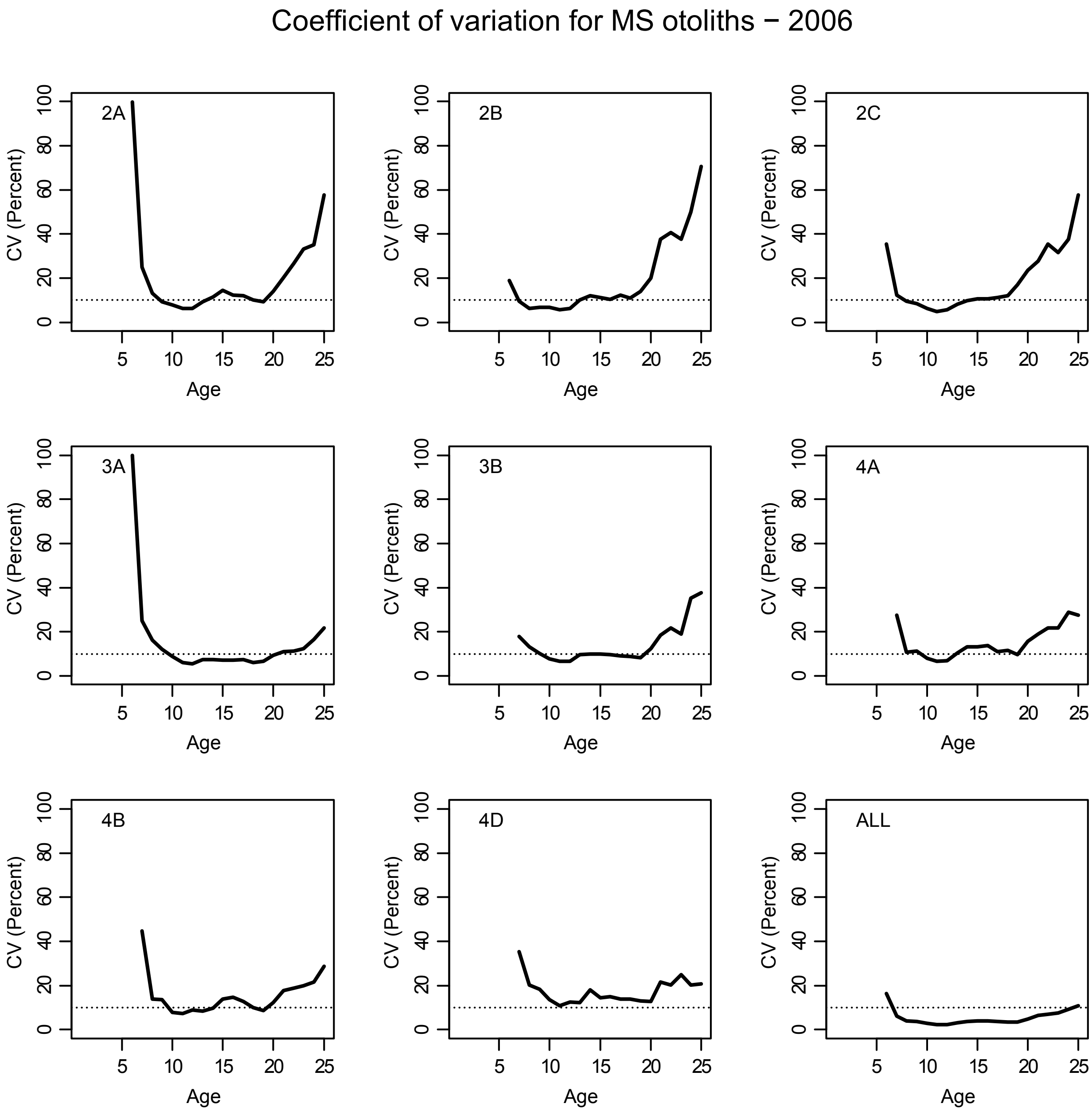

Figure 3. Coefficient of variation for initial readings of halibut, ages 6-25, for Market Sample (MS) otoliths collected during the 2006 commercial fishery. MS otoliths are taken from unsexed fish; the dashed line indicates a CV of 10%.

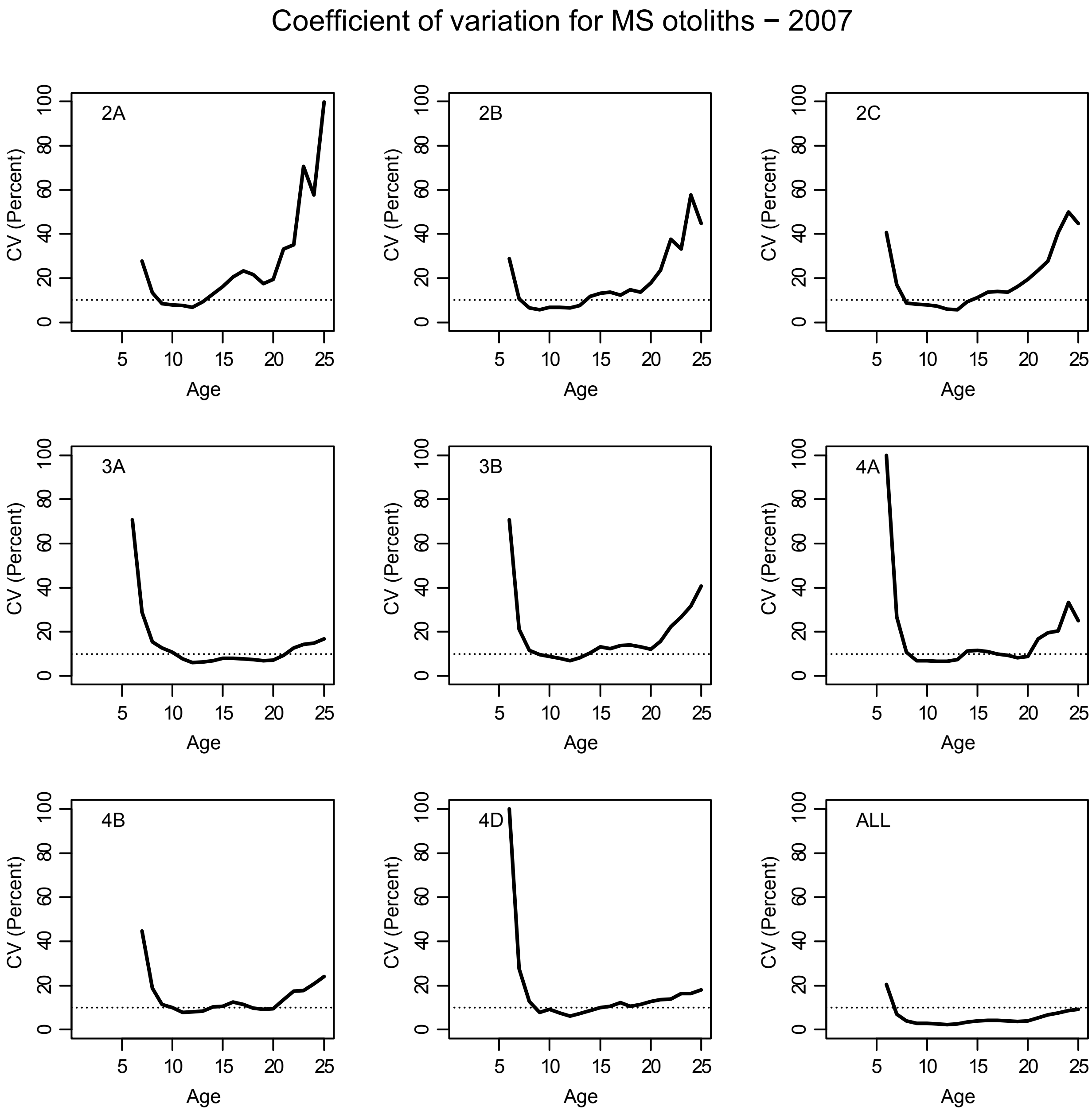

Figure 4. Coefficient of variation for initial readings of halibut, ages 6-25, for Market Sample (MS) otoliths collected during the 2007 commercial fishery. MS otoliths are taken from unsexed fish; the dashed line indicates a CV of 10%.

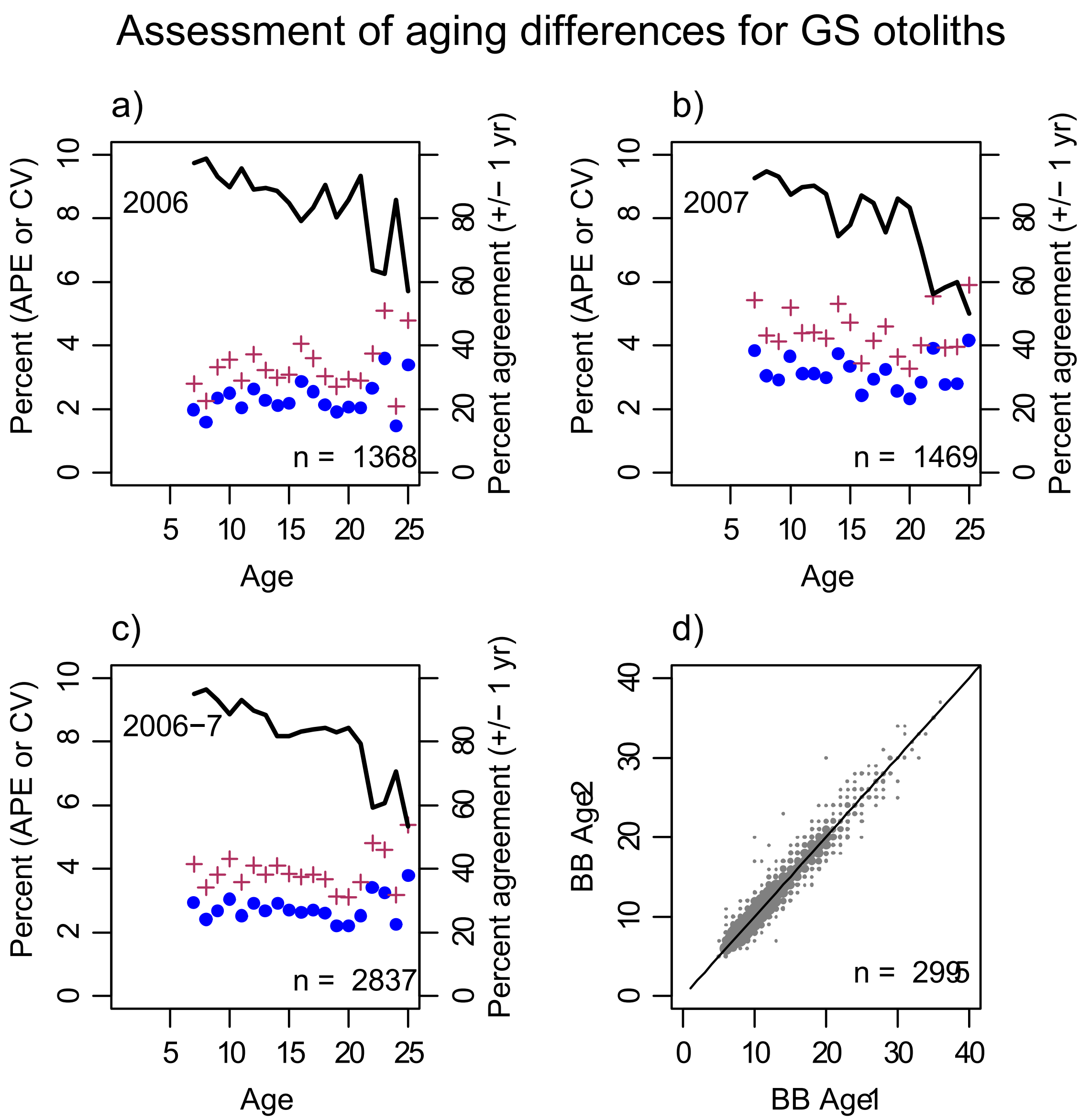

Figure 5. Summary of variability in double readings of GS otoliths in 2006 and 2007. For panels a, b, and c, the dots represent Average Percent Error (APE), the plus signs indicate Coefficient of Variation (CV) and the solid lines represent percent agreement (+/- 1 year) between the two readings. APE, CV and +/- 1 yr agreement are shown for ages 6-25. Panel d shows all double readings; dot size is proportional to sample size.

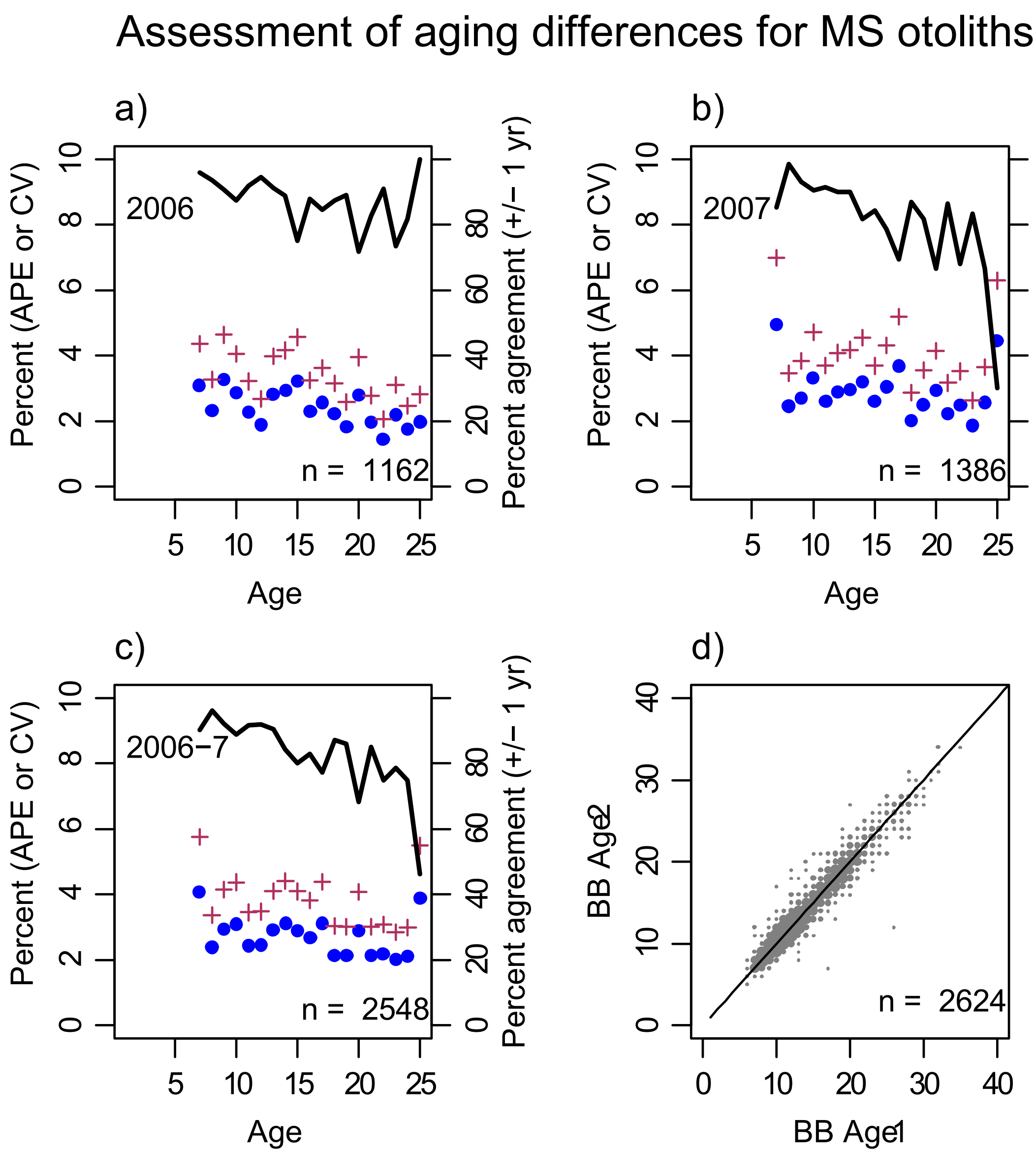

Figure 6. Summary of variability in double readings of MS otoliths in 2006 and 2007. For panels a, b, and c, the dots represent Average Percent Error (APE), the plus signs indicate Coefficient of Variation (CV) and the solid lines represent percent agreement (+/- 1 year) between the two readings. APE, CV and +/- 1 yr agreement are shown for ages 6-25. Panel d shows all double readings; dot size is proportional to sample size.

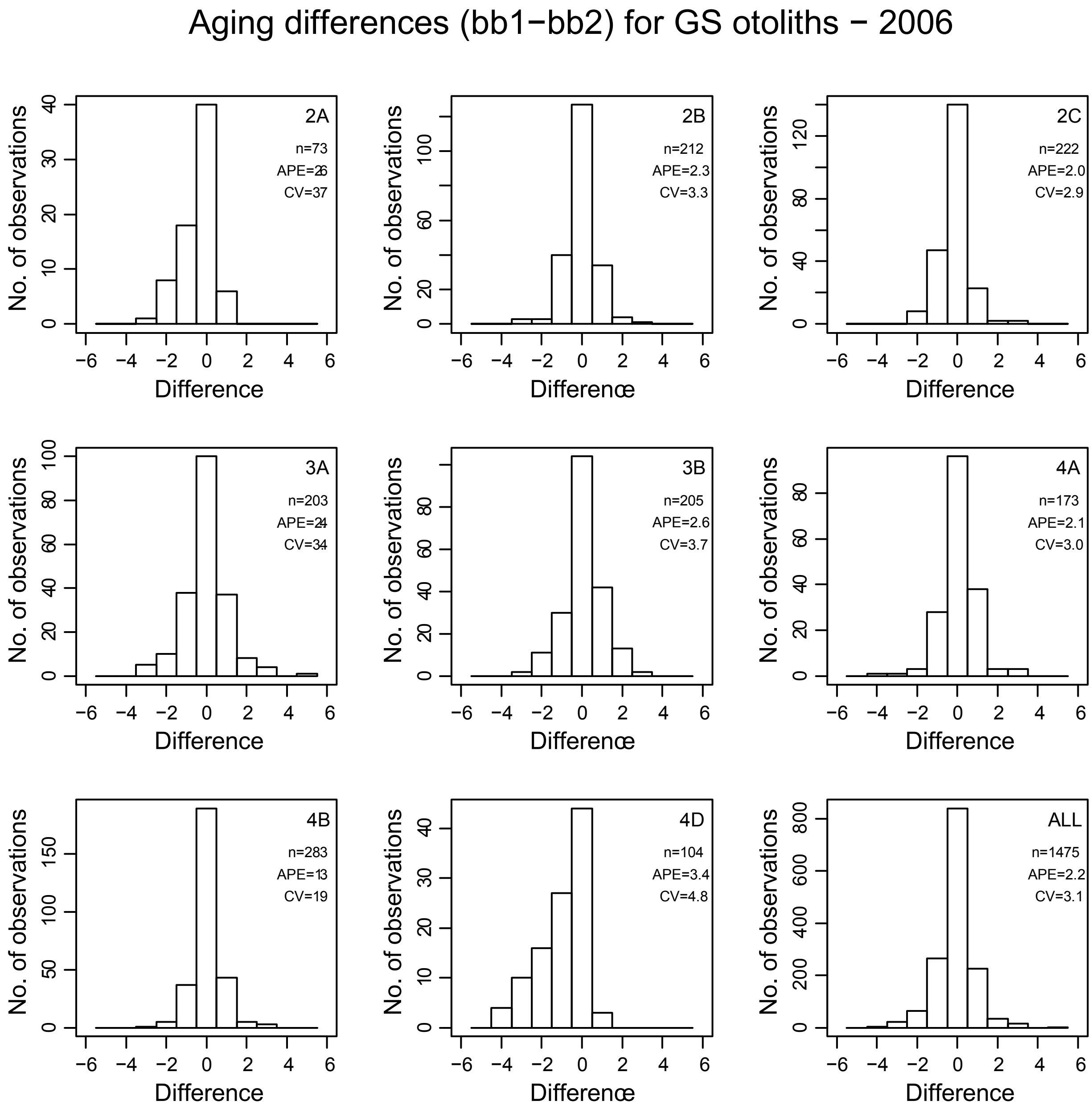

Figure 7. Distribution of age differences between first (BB1) and second (BB2) readings of GS halibut otoliths in 2006. Difference is always computed as BB1-BB2, thus a positive difference indicates the second reading was younger than the initial reading. Average Percent Error (APE) and Coefficient of Variation (CV) are shown for each regulatory area.

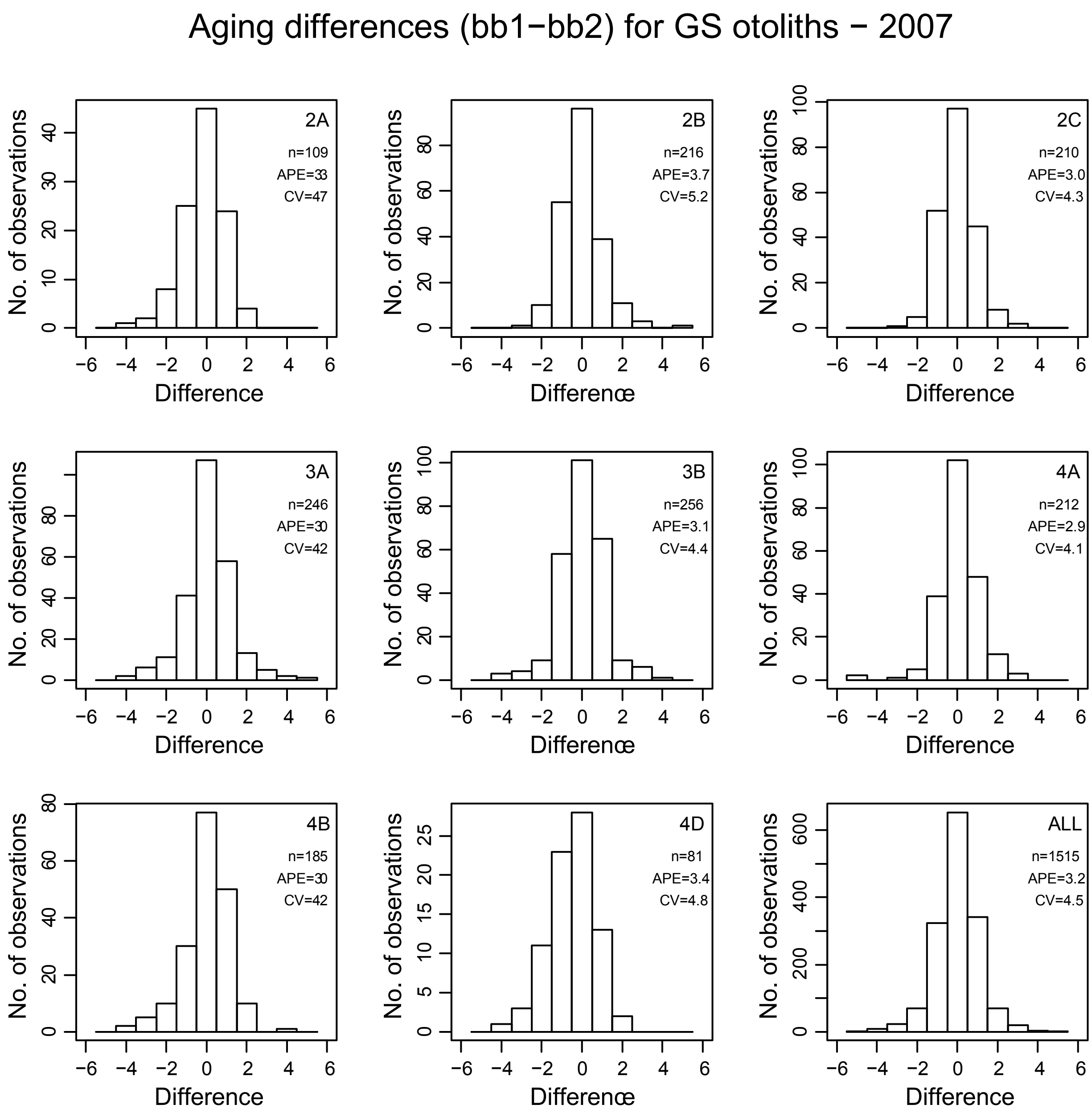

Figure 8. Distribution of age differences between first (BB1) and second (BB2) readings of GS halibut otoliths in 2007. Difference is always computed as BB1-BB2, thus a positive difference indicates the second reading was younger than the initial reading. Average Percent Error (APE) and Coefficient of Variation (CV) are shown for each regulatory area.

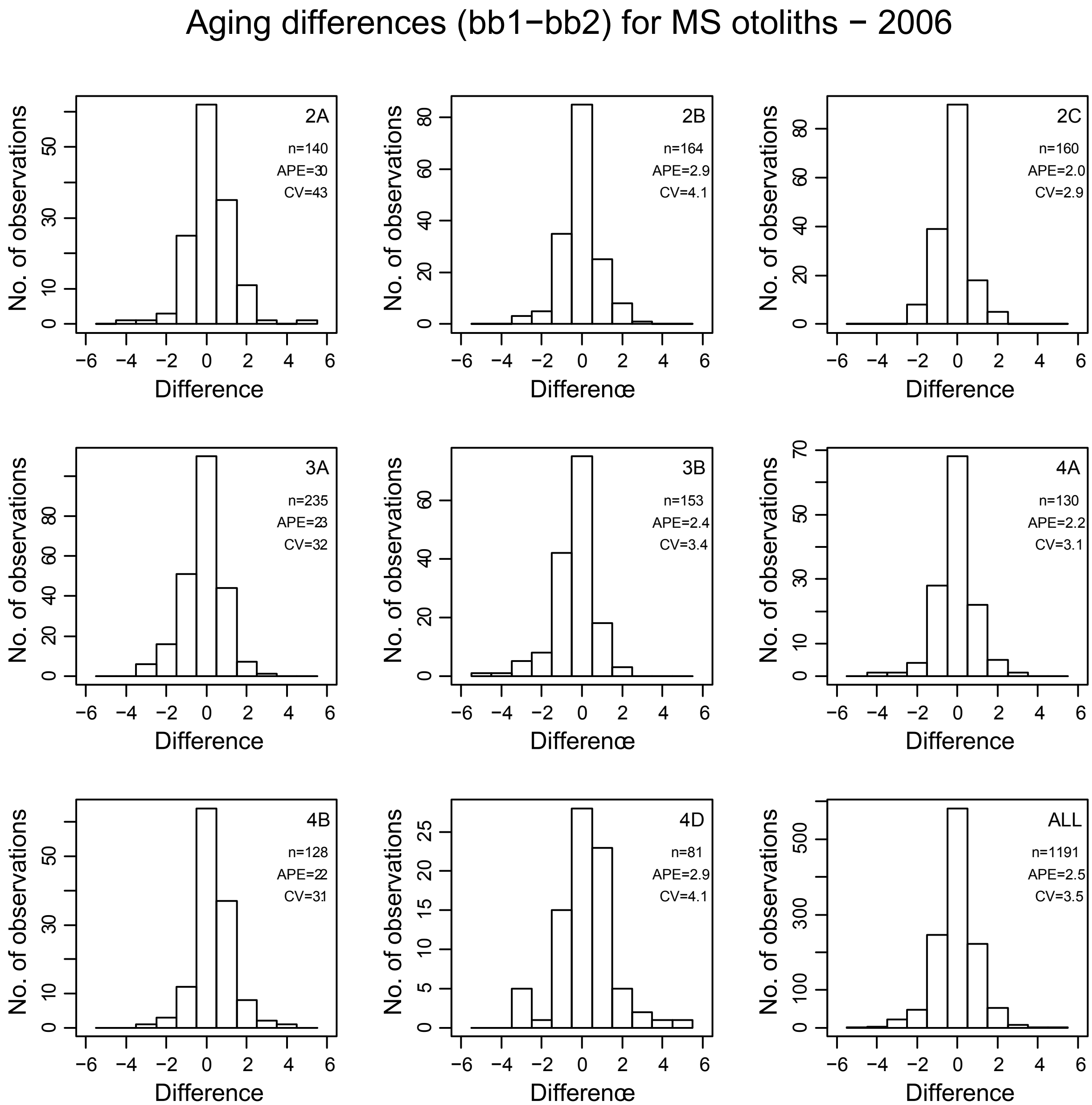

Figure 9. Distribution of age differences between first (BB1) and second (BB2) readings of MS halibut otoliths in 2006. Difference is always computed as BB1-BB2, thus a positive difference indicates the second reading was younger than the initial reading. Average Percent Error (APE) and Coefficient of Variation (CV) are shown for each regulatory area.

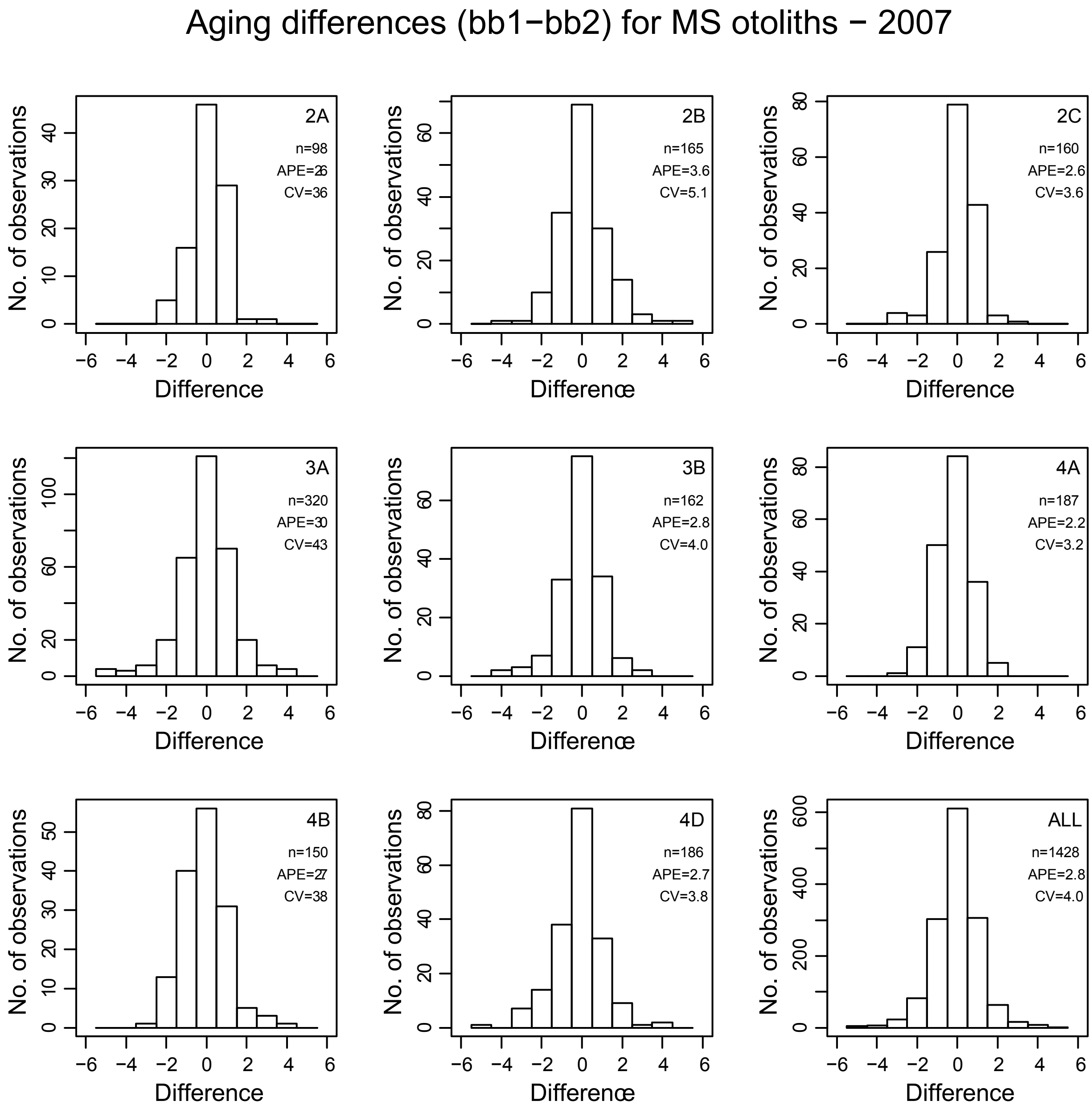

Figure 10. Distribution of age differences between first (BB1) and second (BB2) readings of MS halibut otoliths in 2007. Difference is always computed as BB1-BB2, thus a positive difference indicates the second reading was younger than the initial reading. Average Percent Error (APE) and Coefficient of Variation (CV) are shown for each regulatory area.

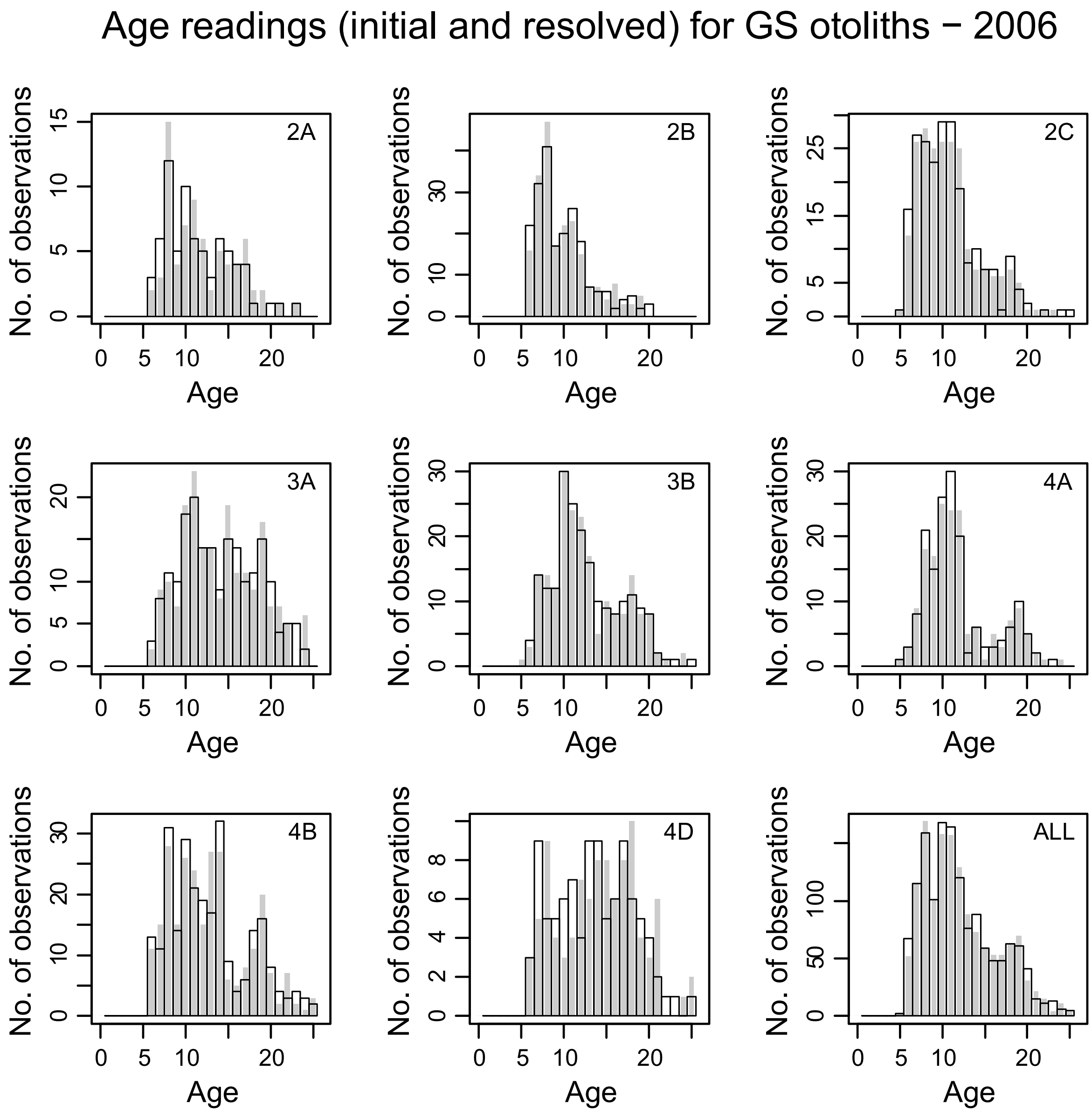

Figure 11. Estimated age distributions for initial (outline) and resolved (shaded) readings, by area, for GS otoliths in 2006.

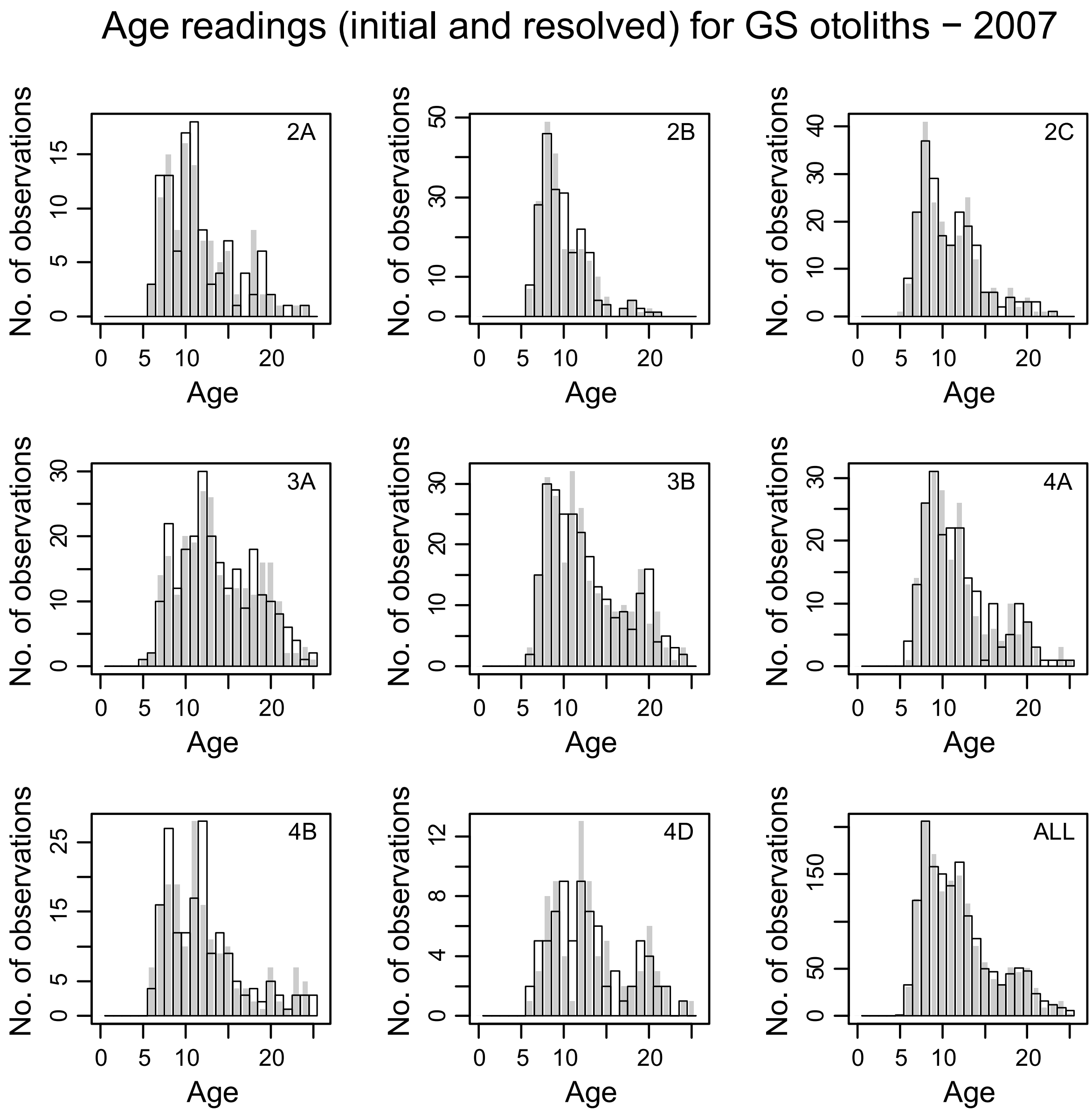

Figure 12. Estimated age distributions for initial (outline) and resolved (shaded) readings, by area, for GS otoliths in 2007.

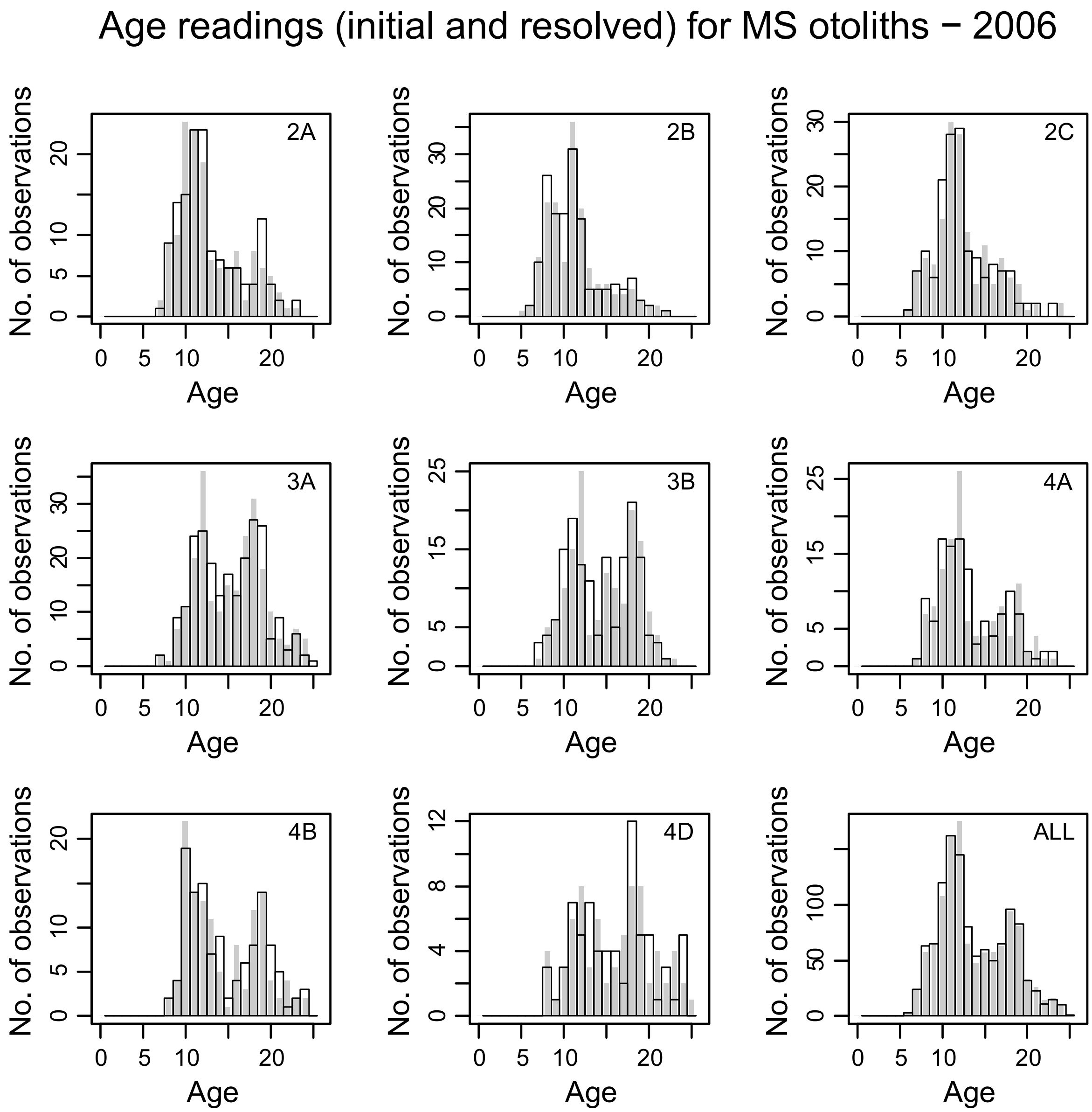

Figure 13. Estimated age distributions for initial (outline) and resolved (shaded) readings, by area, for MS otoliths in 2006.

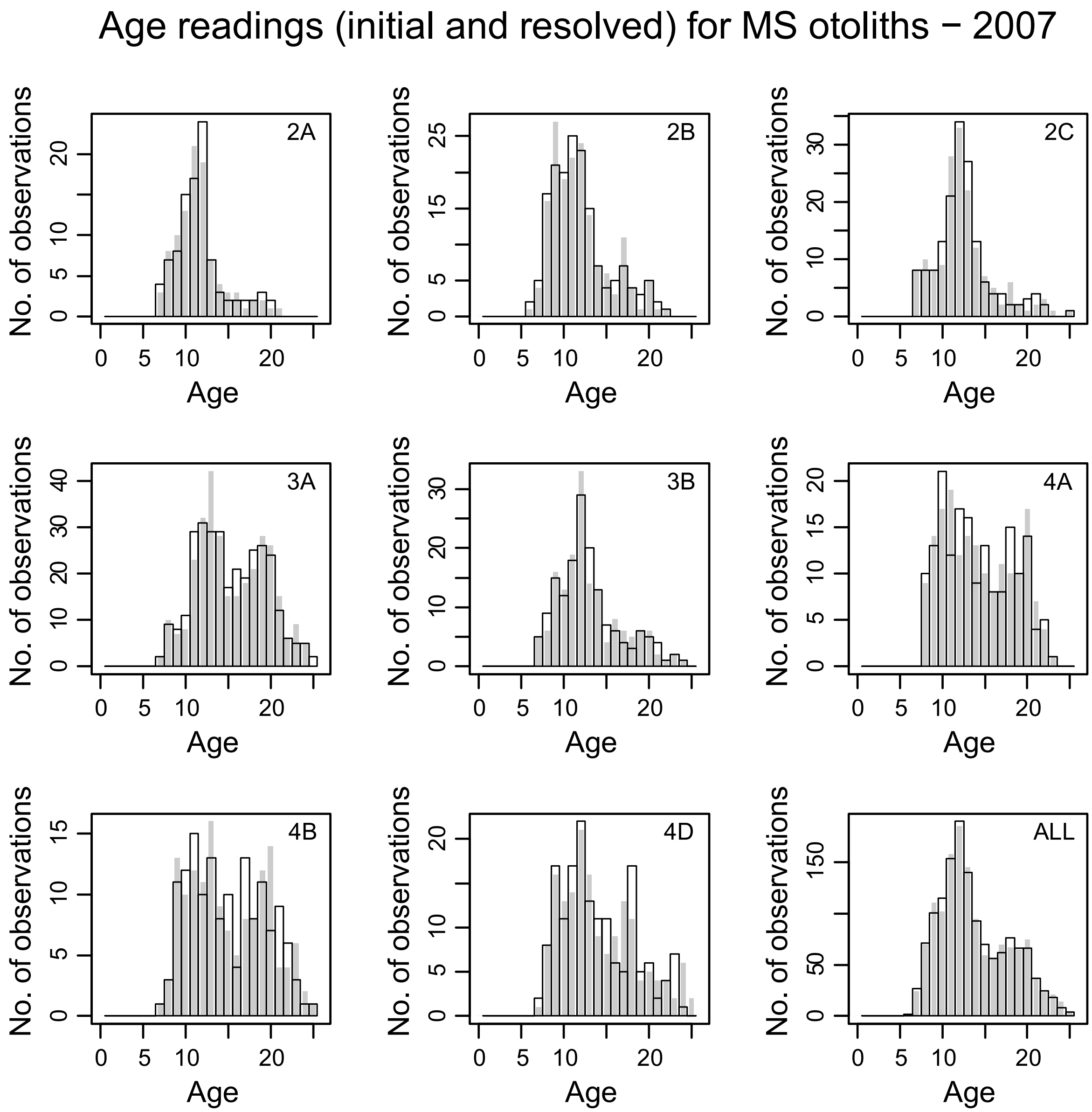

Figure 14. Estimated age distributions for initial (outline) and resolved (shaded) readings, by area, for MS otoliths in 2007.