Fisheries and Oceans Canada — Pêches et Océans Canada

Science — Sciences

C S A S — **S C C S**

Canadian Science Advisory Secretariat — **Secrétariat canadien de consultation scientifique**

Research Document 2010/101 — **Document de recherche 2010/101**

Herring Ageing: Re-ageing of Atlantic Herring Otoliths and the Development of a Revised Catch at Age

Détermination de l'âge du hareng : Réévaluation de l'âge otolithique du hareng et modification des données sur les captures selon l'âge

G.D. Melvin, M.J. Power, and D. Knox

Population Ecology Section
Department of Fisheries and Oceans
Biological Station
531 Brandy Cove Road
St. Andrews, N.B.
Canada E5B 2L9

ISSN 1499-3848 (Printed / Imprimé)
ISSN 1919-5044 (Online / En ligne)

Canada

TABLES OF CONTENTS

Correct citation for this publication:
La présente publication doit être citée comme suit :

Melvin, G.D., M.J. Power, and D. Knox. 2010. Herring Ageing: Re-ageing of Atlantic Herring Otoliths and the Development of a Revised Catch at Age. DFO Can. Sci. Advis. Sec. Res. Doc. 2010/101: vi + 25 p.

ABSTRACT

Age based analytical assessment models rely on the accurate and consistent estimation of ages to track year-classes or cohorts through the fishery. Biased or inconsistent ageing between readers/institutes can affect the model outputs. Several otolith exchanges conducted between 2002 and 2006 identified significant inconsistencies in assigned ages among readers from regional and international research institutes undertaking production herring ageing. To discern these differences a workshop was held in 2006 to examine individual reader practices and procedures, and, to establish standard protocols to improve the precision between readers. A internal DFO review recommended that two studies be initiated, involving multiple readers from several institutes, to validate the herring otolith ages: Bomb radiocarbon assay of otoliths (n=96) from the 1962 year-class over several years, and tracking of the 1983 year-class (n=1787) as it progressed through the fishery from 1985 to 1994. The results of these studies validated the ageing method and confirmed the inconsistencies in ages among readers and institutes observed in previous exchanges.

Simulation studies showed that the observed differences between readers could have a serious impact on the VPA output. This led to the suspension in 2006 of the 4WX herring assessment until the ageing issue could be resolved. Furthermore, for the 4WX herring stock it was recommendation that all otoliths from 1999 to 2005 be re-aged. Over the next couple of years new quality control measures, aging protocols, mounting media, and acceptance criteria for inter-reader comparisons were established. Subsets of the 1999 to 2005 otoliths were re-aged, as well as the 2006 to 2009 otoliths which had never been read, following the above procedures. For all years the acceptance criteria of 80% agreement, a CV <5% and no bias was met. The ages were then used to generate an age-length for each year and a revised catch at age created for the 4WX herring stock. This report provides a chronological overview of the exchanges, workshops, analyses, and results that led to the revised catch-at age.

RÉSUMÉ

Les modèles d'évaluation analytique fondés sur l'âge reposent sur des estimations cohérentes et exactes des âges, permettant de suivre les classes d'âge ou cohortes au sein du stock exploité par la pêche. Des données sur l'âge biaisées ou contradictoires selon les personnes ou les instituts dont elles émanent peuvent influer sur les résultats d'un modèle. Plusieurs échanges d'otolithes effectués entre 2002 et 2006 ont mis en évidence des incohérences importantes dans les données d'âge provenant de divers spécialistes d'instituts de recherche régionaux et internationaux qui procèdent à des déterminations d'âge parmi la production de harengs. Un atelier a été organisé en 2006 dans le but de discerner ces différences, d'examiner les procédures et méthodes de chacun des spécialistes en détermination de l'âge et d'établir des protocoles standards pour améliorer la précision des résultats. Dans le cadre d'un examen interne, le MPO a recommandé la réalisation de deux études, faisant appel à de multiples spécialistes de la détermination de l'âge qui œuvrent dans plusieurs instituts, pour valider les âges otolithiques du hareng, soit une analyse de la teneur en carbone nucléaire (n = 96) des otolithes de la classe d'âge 1962 sur plusieurs années et un suivi de l'évolution de la classe d'âge 1983 (n = 1787) parmi le stock exploité par la pêche de 1985 à 1994. Les résultats de ces études ont validé la méthode de détermination de l'âge et confirmé les incohérences dans les âges établis par les divers spécialistes et instituts qui avaient été observées lors des échanges précédents.

Des études en simulation ont montré que les différences observées dans les âges établis par divers spécialistes pourraient avoir des incidences importantes sur les résultats de l'APV. C'est ce qui a mené en 2006 à suspendre l'évaluation du stock de hareng de 4WX jusqu'à ce que la question de la détermination de l'âge soit résolue. Également, il a été recommandé au sujet du hareng de 4WX de procéder à une relecture des otolithes de 1999 à 2005 pour en déterminer l'âge. Dans les deux années qui ont suivi, de nouveaux protocoles de détermination de l'âge, mesures de contrôle de la qualité, médias de montage et critères d'acceptation pour les comparaisons entre spécialistes ont été adoptés. Des sous-ensembles de la série d'otolithes de 1999 à 2005 ont été soumis à une autre lecture de détermination de l'âge, tandis que des otolithes de 2006 à 2009 ont fait l'objet d'une première lecture, selon les procédures susmentionnées. Pour toutes les années considérées, le critère d'acceptation - 80 % de concordance, CV < 5 % et absence de biais - a été respecté. Les âges obtenus ont ensuite servi à établir une relation âge-longueur pour chaque année et à partir de là des données modifiées sur les captures selon l'âge concernant le stock de hareng de 4WX ont été produites. Le présent rapport donne un aperçu chronologique des échanges, ateliers, analyses et résultats ayant abouti à la modification des données sur les captures selon l'âge.

INTRODUCTION

Age based analytical assessment models rely on the accurate and consistent estimation of ages to track year-classes or cohorts through the fishery. A major source of uncertainty with the aged based approach is intra and inter-reader consistency with time. Quality control is imperative if consistent and comparable ages are to be obtained. Procedures must be in place to ensure that experienced readers over time do not drift toward older or younger ages, and that new readers follow established protocols to produce results similar to the primary reader. As with any subjective interpretation random error is expected but bias or drift is unwanted and can be misleading in the tracking of trends. Major differences in ageing can also have a significant impact on VPA estimates of fishing mortality and stock biomass (Melvin and Campana, 2010).

Multiple otolith exchanges have occurred over the past 8 years, between national and international research institutes involved in herring stock assessments, to investigate potential inconsistencies in assigned herring ages for several stocks from the western Atlantic. The first herring otolith exchange, since the early 1980's (Cleary et al., 1982), occurred in late 2002 in preparation for the inaugural Canada/United States Transboundary Resource Advisory Committee (TRAC) meeting held in March of 2003. The results of this exchange indicated that different and sometimes biased ages were occurring amongst the research laboratories involved with herring assessments. Unfortunately, the impact of these differences on the age based analytical assessment was unknown, but was thought to be minimal on the output results. However, given the high abundance estimates and low fishing mortality of the Gulf of Maine herring stock it was felt important to evaluate how stock perception may differ (Overholtz et al., 2005).

The 2006 TRAC also recommended that the ageing problem be investigated further, yet because the differences were observed mostly with older fish, and the Gulf of Maine herring assessment used a 6+ age group, the inconsistencies were considered unlikely to seriously affect the assessment results. Canadian participants did however raise questions regarding the Canadian stock (4WX) assessment. Later in 2006 several sensitivity analyses were conducted using the results of recent (post 2002) otolith exchanges between national and international research centres to determine the potential impact of the ageing inconsistencies on the 4WX herring assessment. The conclusions reached from these investigations were that, within the bounds of observed differences in ageing between the laboratories, the identified inconsistencies in the ageing of herring may have a serious impact on the age based assessment models used by Canada and the United States (Melvin and Power, 2006; Melvin and Power, 2007). The use of an age based assessment model for 4WX herring was therefore suspended until the issue could be resolved.

The purpose of this working paper is to provide a general chronological overview of the more recent (since 2002) otolith exchanges that have occurred, the decisions made based on the results of these exchanges, and their implication on the 4WX herring assessment. This is followed by a summary of the re-ageing of otoliths from 1999 to 2005 and the revisions to the 4WX herring catch at age (CAA) for the same period. The most recent years, 2006 to 2009, were added to the catch at age using newly aged otoliths.

OTOLITH EXCHANGES

To investigate the repeatability and comparability of assigned herring ages, three otolith exchanges of approximately 200 otoliths were undertaken between 2002 and 2006 with four scientific research institutes involved in herring stock assessments. The institutes involved included the Northeast Fisheries Science Center (NEFSC), the Maine Department of Marine Resources (MDMR), the St. Andrews Biological Station (SABS) and the Gulf Fisheries Centre (GFC), although not all readers were involved in all exchanges.

The initial investigation into inter-laboratory variation began as part of the 2003 Transboundary Resource Assessment Committee (TRAC) process for the Gulf of Maine herring assessment. During the fall of 2002, 215 otoliths collected from the Bay of Fundy were exchanged to explore potential differences or biases between institutes/readers. Four otolith readers (NEFSC, MDMR and two from SABS) independently read the selected otoliths (Overholtz et al., 2005). The information provided to each reader included sample date (for edge delineation) and fish length. The results of these comparisons clearly illustrated a significant difference in mean length at age, the assigned age and the maximum age amongst the readers. The degree of difference was dependent upon the specific reader to reader comparison. Overall, the primary 4WX otolith reader was found to be statistically biased toward younger ages compared to all other readers and had the poorest percent agreement with any reader beyond age 4. The results of the comparisons for the percent agreement and test of symmetry are presented in Table 1. Based on these results, the 2003 TRAC recommenced that an ageing workshop should be held between the two countries to resolve the differences before the next TRAC.

In preparation for the 2006 herring TRAC, a Canada/United States workshop was held in Booth Bay Harbour to review the original 2002 otoliths and to explore ageing methods and protocols. As a result of this workshop and the reader's confidence in their approach to ageing, it was decided to undertake another exchange with 200 otoliths from the 2004/2005 Gulf of Maine herring fishery. These otoliths were selected from each month of the year and fish total length was provided. Participants included readers from the NEFSC, MDMR, and SABS. Unfortunately, the results were extremely surprising and represented a serious deterioration in the agreement between the 1st and 2nd exchange. All readers differed significantly from one another for mean age and the ages were biased toward older or younger fish depending on the reader comparison. Again, there was very poor agreement for otoliths beyond age 4 and a general pattern consistent with the first exchange (Table 2) was observed. The results were presented to the 2006 TRAC which concluded that the ageing inconsistencies between readers could have an impact on the assessment output and recommended the aggregating of older ages (Age 6+).

Given the low VPA estimates of abundance for the Canadian 4WX herring stock and the discrepancy between the VPA and acoustic biomass estimate, concern was expressed over the potential impact of the observed ageing inconsistencies on the 2006 analytical assessment results. In May of 2006, 190 otoliths covering a broad size distribution from the 4WX stock complex were selected from the 2002 fishing year and exchanged for reading. Participants included three readers from the two DFO laboratories (SABS and GFC). Again, the results of these comparisons were consistent with the Can/USA study and indicated a significant and biased ageing pattern amongst the otolith readers (Table 3). The mean ages of the otoliths were significantly different ($P<0.05$), ranged from 4.78 to 5.47 with the maximum age from 11 to 14 years old. In essence, the primary 4WX reader (DFO 1) had poor agreement with the other two readers and on average consistently aged fish younger. Even the comparison of the primary reader (DFO 1) with the originally assigned ages (DFO1_O) indicated a significant bias toward younger fish (Melvin and Power, 2006, 2007).

Initially, the impact of ageing error was thought to be minimal; however, given the low stock biomass and high fishing mortality estimates being reported for 4WX herring, it was important to evaluate the impact of observed ageing error on the perception of stock status. In 2006, several sensitivity analyses using the results of recent (post 2002) otolith exchanges between national and international research centres were undertaken to determine the potential impact of the ageing inconsistencies on the 4WX herring stock assessment. The results from these analyses indicated that, within the bounds of observed differences in ageing between the laboratories, the identified inconsistencies in the ageing of herring could have a profound impact on the age based assessment models used in both Canada and the United States (Melvin and Power, 2006; Melvin and Power, 2007).

Considering the degree of variability in age comparisons, the potential unknown temporal extent and uncertainty in the ages used to develop the 4WX herring catch at age and the age disaggregated index of abundance, the assessment using the age-based ADAPT (VPA) model was suspended until the issue could be resolved. The third framework assessment meeting to investigate assessment models scheduled for the spring of 2007 was also postponed. These postponements lead to an independent review by Campana (2007) who recommended two age validation studies.

AGE VALIDATION STUDIES

The Campana review (2006) suggested that two studies be undertaken to validate herring ages: First, a bomb radiocarbon study of a year-class from the early 1960's over multiple years and second, the tracking of a dominant year-class as it moved through the fishery. Otoliths from the 1962 (*n*=96) and 1983 (*n*=1987) year-classes, respectively, were used for the studies conducted during the summer and fall of 2007. A second Canada/USA workshop was held in January of 2008 to discuss the ageing issues, review the findings of these studies, and to make recommendations on how to proceed from here.

Bomb Radiocarbon Ageing Study

Bomb radiocarbon assay as an age validation tool is based on the increase in atmospheric and ocean radiocarbon during the 1950’s and 1960’s. The method utilizes the rapid changes in C^{14} due to atomic bomb testing and the uptake by the otoliths to estimate age (Figure 1). It is therefore important to have otoliths collected annually and covering this period to calibrate the concentrations. The St. Andrews Biological Station (SABS) is fortunate to have a large and historical collection of herring otoliths dating back to the late 1950's. Ten otoliths per year covering the period 1963 to1972 from the 1962 year-class were selected at random from the collection. Otolith selection was based on the age recorded in the database and consequently may not reflect the correct age of the fish. Unfortunately, no age 2 otoliths from 1964 were available for the analysis and only six age 9 otoliths were found from 1971. An additional ten age 1 otoliths from1963 were substituted for a total of 96.

The bomb radiocarbon study was comprised of two components or phases. The initial phase of the study involved the age determinations of the otoliths by the participating institutes. Participants included NEFSC (1 reader), MDMR (1 reader), SABS (2 readers), and GFC (1 reader). Only month of capture was provided to assist the readers with edge assignment. Readers were instructed to prepare and read otoliths according to standard laboratory practices.

The results of inter-reader comparisons and the originally assigned ages (database) for mean age are presented in Table 4 and the percent agreement/test of symmetry are presented in Table 5. In summary the otolith exchange and ageing of herring otoliths for use in the bomb radiocarbon age validation study showed:

- A disconnect between DFO database and current readers (Figure 2).
- For age 4 and older, there was very poor agreement amongst readers and with the database.
- None of the between reader comparisons met the criteria of 80% agreement and 5% CV as a guideline for acceptability,
- The general ageing pattern was similar to previous exchanges, i.e. the primary 4WX reader on average assigned the youngest ages, MDMR/NMFS middle range of ages, and DFO2/DFO3 the oldest.
- There is serious concern about the impact that historical and current ages may have on the analytical assessment.
- These results confirm the need for an age validation study.

The second phase of the BRC study involved the actual bomb radiocarbon assay. Once the otoliths were read, the trays were sent to BIO for processing, C^{14} assay and data analysis. An overview of the methodology, processing and the preliminary analysis was presented in Melvin and Campana (2010). In essence, this approach examines the change in C^{14} during a period of rapid build up in body hard parts (such as otoliths) to estimate age. The critical period for the method is the 1950's and 1960's when atmospheric nuclear testing was a common practice. The accuracy of the method was determined to be within 6 months of the true age.

Several procedures were utilized to validate the relationship between ΔC^{14} and to calibrate the methodology. The insert in Fig. 1 compares the reported ΔC^{14} for the Northwest Atlantic Ocean from 1940 to 2000 with the observed ΔC^{14} of herring of a known age from the 1962 year-class. The trajectory of herring ΔC^{14} is almost identical to that of the water (Figure 2), thereby confirming the use of the method for age validation. A quadratic regression was then used to estimate the year of formation from the radiocarbon of herring otolith core (Figure 1).

The initial examination of the reader assigned ages for otoliths used in the BRC study showed bias between the original database ages and some readers, as well as between readers (Figure 3). Otoliths used for BRC assay were selected based on the database age and for their first year of life being 1962, consequently assuming a correct database age, the ΔC^{14} for the year of formation should be constant through time. The results of this comparison clearly illustrate that the majority of otoliths were originally under aged by approximately 1-2 years after age 4 relative to the BRC estimates (Figure 4). In a few samples, the estimated year of formation was older than the BRC. Examination of the ages by reader also showed that the differences were reader dependent and that the majority of otoliths beyond 4 were aged younger than the BRC estimate (i.e. bomb age). The largest differences occurred for the 4WX primary reader and the smallest for the GFC reader.

In summary, the BRC study results are generally consistent with previous otolith exchanges in the context of the relative ageing practices. DFO Reader 1 on average aged fish younger than the USA readers and the USA readers aged fish younger than DFO Reader 2 and DFO Reader 3. Overall, based on the BRC data, there has been a general tendency to underage herring beyond age 5. The original database ages appear to shift approximately 1 year from the true age at about age 6, DFO readers 2 and 3 show a good correspondence until age 9-10, DFO1 starts to diverge around age 6 and the USA readers around ages 8-9.

Dominant Year-Class Tracking Study (DYCT)

The second age validation study undertook to track a dominant year-class through time and involved a much larger number of otoliths (*n*=1787). The purpose of the study was to investigate the individual reader's ability to document a known year-class as it progressed through the fishery/samples.

A representative sample of approximately 200 otoliths per year for years when a dominant year-class was observed in the fishery and the in the samples was selected for this analysis. Although several dominant year-classes were available, the very strong 1983 YC, which should be easily tracked was chosen for the study. Based on the ages in the database, otoliths trays were selected at random from each year between 1984 and 1994; a period of 11 years. Only purse seine samples collected in September were used for this analysis in order to minimize gear selectivity and variation in interpretation of the outer edge. No samples were weighted for catch in the analysis. The number of otoliths selected for reading was 1787 from 1985 (Age 2) to 1994 (Age 10) (Table 6). The otoliths were classified by each reader as "good" - readable, "bad"- some uncertainty about the age (best guess), and "ugly" – unreadable (Table 7). Based on the assigned ages in the database, the 1983 dominant year-class was clearly visible in every year of the 10 years select for the study (Figure 5). It should be noted that although the dominant year-class was traceable, the individual ages may not be correct and there may be some blending with other ages.

The study participants (i.e., readers) were slightly different from those in the BRC exchange. For the DYCT study there were 2 readers from the National Marine Fisheries Service, 1 reader from the Maine Department of Marine Resources, 1 reader from DFO St. Andrews Biological Station, and 1 reader from the DFO Gulf Fisheries Centre. The same readers from each institute were involved in both studies except there was an additional reader from NMFS and the secondary reader from SABS did not participate in the DYCT study.

Overall, the results of this study were consistent with the BRC study in that all readers generally under aged the otoliths relative to the original ages contained within the database (Table 8). As in previous studies, DFO Reader 1 had the youngest mean age (4.6) and DFO3 the oldest (5.3) with the USA agers in the middle. Examination of the database ages and the assigned ages clearly illustrated that with the exception of DFO Reader 3, most readers were missing the older age fish from age 8 and beyond (Figure 6??). Comparison of the ages using Bowkers test of symmetry indicated that all readers were statistically biased with respect to the database ages and amongst themselves (Table 9). In all cases the bias was towards under ageing relative to the original ages in the database; the degree being reader dependent.

The result of the individual comparisons with the database ages are presented in Figure 9. For every reader there is significant under ageing compared with the database the degree of which varies amongst the readers. The most extreme under ageing appears to occur for DFO Reader 1 where the divergence begins after age 4 compared to ages 5 to 6 for the NMFS and MDMR, and age 8 for DFO Reader 3. It is important to note that a direct comparison with the BRC study cannot be made although the trends follow a similar pattern. The database readers for this study (1985 – 1994) are not the same as those for the BRC (1963 -1971). Three different otolith readers contributed to the database otolith ages in the DYC study (Melvin and Power, 2007).

Tracking the dominant year-class as it progressed through the fishery was examined by attempting to follow the 1983 YC through the years. Figure 5 clearly shows that the DYC is track able and identifiable in the samples used for the study over a 10 year period. However, when the results from individual readers were examined, only DFO3 was considered to have identified the 1983 year-class as dominant throughout the time series. Comparison of the original ages with Reader 1, the primary reader for 4WX herring, illustrates the serious inconsistencies with the ageing (Figure 6). The 1983 year-class, although present, all but disappears in significance by age 5 in 1988. For all readers except Reader 3, the strong year class fades into the background between ages 5 and 6 and is completely absent in many of the later years for ages 8-10. In essence, only one of the 5 readers was able to track the strongest year-class in the recent history of the fishery.

The results of the DYC tracking study are as follows:

- The observed ageing patterns were similar to BRC. DFO Reader 1 assigned ages that were among the youngest, DFO Reader 3 the oldest, and USA readers in the middle.
- There was a consistent disconnect and biased difference between the database (DB) and all readers. All readers were found to on average age the otoliths younger than the DB.
- The reader closest to the DB ages was DFO Reader 3 and furthest away was DFO Reader 1.
- Generally there was poor percent agreement between the readers (56% or less) in most comparisons.
- The agreement in age between readers appears to deteriorate after age 4.
- NMFS readers showed smallest bias between readers for the otoliths read, however, biased readings were observed between MDMR and NMFS1.
- No reader reported otoliths as old as in the DB.
- Large differences in mean length at age were observed between readers.
- The dominant year-class could not be tracked by any of the readers except DFO Reader 3.

Workshop Summary:

Upon completion of the BRC and DYCT studies a workshop was convened with all participants to discuss the results and to make recommendations for improvement. The main conclusion, based on the studies presented, was that major inconsistencies were occurring with herring ageing amongst the readers and with the historical database. The degree of difference varied depending upon the reader. It was further concluded that each institute must examine the extent of these inconsistencies, identify their potential impact, and determine a course of action to overcome this problem. Unfortunately, the current 4VWX otolith reader demonstrated the poorest percent agreement, biased ageing, and consistently on average under aged the test otoliths relative to the other readers, the BRC assays, and the database. The implication or impact, of this under ageing on the 4WX herring VPA has been examined by several investigations over the past couple of years (Melvin and Power, 2006; Melvin and Power, 2007). The results from these studies indicate that under ageing generally leads to an over-estimate of fishing mortality and an under estimate of biomass; the amount dependent upon the severity of the under ageing.

A number of recommendations evolved from the 2008 ageing workshop that were designed to improve the ageing of herring. These included a new mounting media for the otoliths, a zoom rather than step focus microscopes, the absence of length information during the reading process, a reference collection for quality control and a few new ageing protocols. Although these practices will improve the ageing in future years, they will not do anything to correct the past. It was therefore concluded that a significant number of the 4WX herring otoliths would have to be re-read following the new protocols and quality control procedures. At a bare minimum it was recommended that the otoliths be re-read back to 1999 and preferably back to the late 1980's. It was also stressed that not all otoliths previously read need to be re-read and for this exercise a sub-sample would be sufficient. Once the otoliths have been re-read a new catch at age and age-disaggregated index of abundance will need to be constructed for input into a VPA.

Ageing of 1999-2009 Otoliths:

Based on the results and recommendations of the 2008 Ageing Workshop it was decided that a sub-sample (~1000) of herring otoliths from each year between 1999 and 2005 would be re-aged following the newly established protocols. Un-aged otoliths from 2006 to 2009 sample years would also be included in the production ageing to bring the dataset up to date now that the ageing problem was resolved. Approximately 1000 otoliths per year from 2006 to 2009 were also aged following the standard protocols.

Re-ageing, which was given a high priority, began in about June of 2008 with a major effort to complete the task as soon as possible. Quality control procedures were established in an attempt to avoid past problems. During the initial period, comparisons were undertaken with previously aged/validated otoliths to ensure consistency. As well, regular random testing of approximately 10% of the aged otoliths were examined by an alternate reader. The criteria for acceptance during production ageing of good quality otoliths was set at 80% agreement, <5% CV, and no bias. Initial comparisons of newly aged DYCT otoliths were positive and did not show any signs of deviation. Unfortunately, in November of 2008, the first comparison of newly aged otoliths from 2007 did not meet the criteria when the primary reader was compared with the alternate reader at SABS.

Another 280 otoliths from 2007 were aged for comparison between the readers. Each otolith was classified as good, bad or ugly for reading purposes and the comparison was conducted on three groupings: 1) all otoliths read, 2) good otoliths and 3) good/bad. Unfortunately, none of the groupings met this level (Table 10). Comparing only the good otoliths, which only went to a maximum age 6, improved the results but still did not meet the established criteria. Agreement between readers, as in the past, was found to deteriorate after age 4-5.

Several approaches were tried to improve the precision of the results including additional training, an internal workshop with several DFO readers and the re-ageing of the 2007 otoliths. In the end a final exchange was undertaken with the external reader (Reader 3) for a random selection of approximately 200 otoliths from 2007. The results were obtained in June of 2009 and confirmed that the primary, not the alternate reader, did not meet the standards established at the 2008 workshop (Table 10). In July of 2009 it was decided to replace the primary reader with a new reader and to implement quality control procedures to check each year's otoliths as they were completed. A random selection of approximately 100 of the 1000 otoliths, selected from each year being re-aged, were sent to the external reader and the ages were only used for the catch at age if the new or the secondary reader met the acceptance protocols.

Given an almost two year delay in trying to improve the accuracy of the ages and the urgency to get the assessment back on track, 3 readers were employed in the final ageing of 1000 otoliths per year between 1999 and 2009: the external expert from DFO Gulf Region, the secondary reader from DFO-SABS (Reader 2) and the new reader from SABS. Whenever possible, otoliths aged by Reader 2 were included in the age-length key as all earlier comparisons with the external expert reader met the acceptability criteria. Each reader was assigned several complete years for ageing Quality control measures were implemented for both the SABS secondary reader and the new SABS reader. The following summarizes the re-ageing and the observed results of the comparisons with the external expert. Note that the re-aged otoliths were considered acceptable if the otoliths considered "good" by both readers met the >80% agreement, <5% CV and no bias. The use of “good” otoliths for the acceptance criteria was based on the logic that only these otoliths would be included in a test or reference collection used for production ageing evaluations.

1999 Otoliths:

Otoliths were aged by the new herring reader. The results for the random selected comparison otoliths were poor when all otoliths were evaluated but met the criteria when only the "good" otoliths were examined (Table 11). Based on the results for the "good" otoliths the ages from this year were accepted for the production of an age-length key. The total number of ages available to generate an age-length key for 1999 was 1197.

2000 Otoliths:

To save time, otoliths from a previous comparison involving this year were used to evaluate the acceptability of ages by the external (reader 3) and the alternate SABS reader. Overall, 325 otoliths from 2000 had been previously aged by both readers. The comparison showed no bias and good agreement (88.3%) with a low CV. In total, 2148 ages were available to generate the catch at age for 2000 (Table 11).

2001 Otoliths:

Comparison of all 100 otoliths read by the New reader and the external did not meet the criteria for acceptance (Table 11). However, when only the otoliths identified as "good" by both readers were compared the results improved to meet the acceptance level with 80% percent agreement, a CV of 2.97% and no bias. Overall, 818 otoliths were re-read. Combining previously read Reader 2 otoliths with the re-read otoliths produced 3116 otoliths to develop an age length key for the year.

2002 Otoliths:

All readers noted that the annuli from the 2002 fishery otolith collections were extremely difficult to identify. Otoliths from this year had the lowest percent (35%) rated as "good". The poor quality was caused by otoliths with multiple checks and the deterioration of the mounting media. Major differences were observed for otoliths assigned ages older than 5. It was also the only year where the new reader ages did not meet the criteria for acceptance in both the percent agreement and CV (Table 11). As such, these ages could not be used to create an age-length key. Given this was one of the last years subjected to a comparison, insufficient time was available to do a complete re-read and meet the timeframe for the assessment meeting. As an alternative, a number of previously aged otoliths by Reader 2 from 2002 were combined with the new Reader ages (<age 6) to generate a catch at age from 1386 otoliths.

2003 Otoliths:

The secondary reader (Reader 2) for SABS aged 705 otoliths from the 2003 fishery. Like most comparisons, when the entire sub-sample of 100 otoliths sent to the external reader were analyzed the acceptance criteria were not met. However, no bias was observed toward older or younger ages indicating random error. When those otoliths identified as "good" were compared the standards were met and the entire set of 705 otoliths was included in the age length key to develop the catch at age matrix. Combining the re-aged otoliths with previously aged otoliths by the same reader meant that 2278 otoliths were available to generate the 2003 age length key.

2004 Otoliths:

The 2004 otoliths were also re-read by the secondary SABS reader. A total of 810 otoliths were selected for ageing. As per the quality control protocol approximately 100 otoliths were selected at random for ageing by the external reader. As with the previous comparisons, the initial comparison with all otoliths did not meet the acceptance criteria, however, when only the “good” were compared the criteria were met. No bias was found for either all otoliths or the good otoliths. Overall 1030 ages were available to develop the 2004 age length key.

2005 Otoliths:

This was the first year that the new reader undertook to age on his own and the last year of re-aged otoliths. Otoliths from subsequent years (2006-2009) were deferred until the ageing problem was resolved. Upon completion of the approximately 1000 otoliths for 2005 a random sample of 100 was sent to the external expert for ageing. Unfortunately, neither the comparison of all otoliths or only the “good” otoliths met the standard set for acceptance. All ages were rejected. The new reader was then sent to Moncton for several days of additional training. Subsequently, all 2005 re-read otoliths were re-read and a new comparison undertaken. This time those otoliths identified by both the external and the new reader as “good” met the criteria and the new ages were accepted for development of the age length key.

2006-2008 Otoliths:

Otoliths from these years had not been previously aged. The majority of otoliths for these years were aged by the external expert and considered to be accurate based on previous testing and the bomb radio carbon results (Melvin and Campana, 2010).

2009 Otoliths:

The 2009 otoliths were the first to be completely mounted in the new mounting media and where a portion of the otoliths were aged by all 3 readers. The new reader was responsible for ageing the majority of the otoliths, however, several hundred were selected at random for both the external reader and the secondary reader to age. Immediately evident in the results was the fact that for both all the selected otoliths and the “good” otoliths the acceptance criteria were met (Table 11). This is likely reflective of the clarity of the new media and the experience gained by this study, especially for the new reader. Percent agreement between readers ranked in the high 80’s for the good otoliths. In total, 3929 otoliths were available for the 2009 age length key.

Catch at Age:

The re-ageing of otoliths used to develop age length keys for the period 1999 to 2005 were applied to the lengths at age to create a revised catch at age for this period. Over the period there appears to be a significant divergence from the earlier CAA (Figure 8). There also appears to be trend from 1999 to 2009 where the earlier years seem to compare reasonably well, but as time progressed the older ages virtually disappear in the original CAA. Several year-classes were trackable in the revised catch at age. In particular the 1998 year-class was apparent from age 2 in 2000 to age 7 in 2005 and the 2000 year class was trackable through to age 8 in 2009. Overall, while the numbers of older fish are not large in the revised CAA, they are present unlike in the original CAA. When comparing the old and the revised CAA it should be noted that the yearly total number of fish in the catch at age does not change from the original to the revised. The age distribution of the catch simply shifts around within the total. The CAA for the entire time series (1965-2009) is presented in Figure 9.

Summary:

The inconsistencies in ageing Atlantic herring were first noted in 2003. Following a number of regional and international exchanges concern was expressed about the implication of ageing error on the evaluation of the Gulf of Maine stock complex and the 4WX herring stock. The two key DFO readers were found to be at the opposite ends of the extremes with the primary reader for 4WX herring constantly on average under-ageing the otoliths relative to all other readers. Simulation studies indicated that within the bounds of observed differences between readers significant differences in the VPA output could occur affecting the interpretation of stock status. Based on the results of these exchanges and the simulation studies, the 4WX herring analytical assessments was suspended in 2006 until the ageing problem could be resolved resolved.

Bomb radiocarbon assay results validated that the herring otolith rings were annuli and that age interpretations showed a consistent pattern of general under-ageing by the primary 4WX herring reader. The dominant year-class tracking study showed the poor tracking of one of the largest year-classes in recent 4WX herring history by all but one reader and supported the observations of earlier exchanges. A workshop held in 2008 to discuss the study results recommended that the otoliths from 1999 to 2005 be re-aged and that a revised CAA needed to be developed from the new ages.

Several attempts were made to re-train the primary 4WX reader to meet the comparison criteria for acceptance but without success. In the end, the primary reader was removed from ageing herring and replaced with a new reader (already in training) to complete the re-ageing exercise. Quality control measures were implemented to ensure the re-aged otoliths met the standard of 80% agreement, a CV of less than 5% and no bias. A random selection of 100 aged otoliths from each year that was re-aged was sent to the external expert for comparison. Only when the reader met the acceptance criteria were the ages from a given year used to generate an age-length key.

The revised CAA shows some significant differences from the old CAA, especially for the older ages and has good correspondence in tracking several year-classes. The revised CAA, including the first time ageing of otoliths from 2006-2009 was used in the 2010 4WX herring analytical assessment (Power et al. 2010).

REFERENCES

Campana, S. 2006. Report on the source of errors affecting the ageing of the 4VWX herring stock with a recommended solution. DFO Internal Report. December 11, 2006. 9p.

Cleary, L., J.J. Hunt, J. Moores, and D. Tremblay. Herring Aging Workshop St. John's Newfoundland March-1982. CAFSAC Research Document 82/41. 10p

Melvin, G.D., and S. Campana. 2010. High resolution bomb dating for testing the accuracy of age interpretations for a short-lived pelagic fish, the Atlantic herring. Environmental Biology of Fishes. (DOI 10.1007/s10641-010-9663-4)

Melvin, G. D. and M. J. Power. 2007. Ageing inconsistencies and sensitivity analysis for 4WX herring. Canadian Science Advisory Secretariat Research Document 2007/54: 1-35.

Melvin, G.D. and M.J. Power. 2006. Expert Opinion on the Impact of Age Reading Inconsistencies on the 2006 TAC Advice for Southwest NS / Bay of Fundy (SWNS/BoF) Herring. Expert Opinion 06/06. 6p

Overholtz, W.J., Jacobson, L.J., Melvin, G.D., Cieri, M., Power, M.J., Libby, D. and Clark, K. 2004. Stock assessment of the Gulf of Maine-Georges Bank herring stock complex, 1959-2002. Northeast Fisheries Science Center, Reference Document 04-06. 300p.

Table 1. Percent agreement by age groups for the 2002 Canada/USA herring otolith exchange. Note DFO1A are the original assigned ages, and DFO1 the ages from a subsequent re-read. Chi2 values are associated with Bowker's test of all ages (2+) to investigate bias between readers.

	Percent Agreement by Age Groups						
Comparison	2+	3+	4+	5+	6+	Chi^2	P-Value
DFO1A/DFO1	7.9	4.4	0.0	0.0	0.0	197.0	0.0
DFO1A/DFO2	49.1	46.6	27.2	16.4	18.5	81.1	0.0
DFO1/DFO2	62.3	56.8	21.7	18.3	25.0	81.0	0.0
DFO1/USA1	57.5	51.1	15.4	11.9	5.0	87.3	0.0
DFO1/USA2	56.8	50.3	14.3	10.2	5.0	88.3	0.0
DFO2/USA1	75.7	71.7	54.1	53.7	46.4	11.4	0.0
DFO2/USA2	77.9	74.3	60.2	61.0	55.4	20.8	0.0
USA1/USA2	85.0	82.5	73.5	75.0	71.9	8.5	0.4

Table 2. Percent agreement by age groups for the 2006 Canada/USA herring otolith exchange. Comparisons with the ending "Ä" indicate the original ages extracted from the database. Otoliths for this study originated from the Gulf of Maine fishery. Chi2 values are associated with Bowker's test of all ages (Age 2+) to investigate bias between readers.

	Percent Agreement by Age Groups						
Comparison	2+	3+	4+	5+	6+	Chi^2	P-Value
DFO1/USA1	39.3	37.2	19.2	7.7	4.0	119.0	0.0
DFO1/USA2	50.8	59.1	50.4	27.7	32.0	66.8	0.0
DFO1/USA2A	60.3	59.7	48.0	29.2	32.0	47.6	0.0
USA1/USA2	50.0	47.6	31.9	17.8	20.3	77.3	0.0
USA1/USA2A	48.0	45.5	26.1	13.1	10.1	88.0	0.0
USA2/USA2A	72.9	71.6	63.4	49.3	46.4	16.3	0.1

Table 3. Percent agreement by age groups for the May 2006 herring otolith exchange among Canadian readers only. The “O” extension on the reader indicates the original database ages assigned the by the reader. Otoliths were from the 2002 4WX fishery. Chi2 values are associated with Bowker’s test of all ages (Age 2+) to investigate bias between readers.

	Percent Agreement by Age Groups						
Comparison	**2+**	**3+**	**4+**	**5+**	**6+**	**Chi^2**	**P-Value**
DFO1/DFO2	52.4	41.7	41.3	38.0	39.4	70.5	0.00
DFO1/DFO3	43.9	32.5	28.6	23.9	25.4	88.2	0.00
DFO1/DFO_O	69.8	57.8	56.8	53.8	50.0	21.9	0.01
DFO2/DFO3	53.2	44.1	43.9	40.7	35.4	56.5	0.00
DFO2/DFO_O	73.1	63.1	61.5	55.7	51.0	14.5	0.34
DFO3/DFO_O	51.7	35.4	37.1	30.0	22.6	48.1	0.13

Table 4. Summary of mean reader age t-test comparisons for the bomb radio Carbon (BRC) otolith exchange and the original database ages (DB). Mean ages are presented along the diagonal, t-value above diagonal and significance below the diagonal (* = P<0.05).

	Reader					
	DB	DFO-1	DFO-2	DFO-3	MDMR	NMFS
DB	5.25	1.332	-0.633	-0.752	0.050	0.907
DFO-1	-	4.73	-1.976	-2.110	-1.348	-0.390
DFO-2	-	*	5.53	-0.114	0.707	1.534
DFO-3	-	*	-	5.58	0.832	1.659
MDMR	-	-	-	-	5.23	0.899
NMFS	-	-	-	-	-	4.87

Table 5. Bomb radio carbon otolith comparison percent agreement and test of symmetry for the 5 readers and the original database (DB). The number aged is presented along the diagonal, the Chi^2 value and level of significance above diagonal and the percent agreement below the diagonal.

Reader	Reader					
	DB	DFO-1	DFO-2	DFO-3	MDMR	NMFS
DB	96	31.0*	27.26*	27.6	21.5	28.4
DFO-1	45.8	95	37.2**	38.3**	20.4	16.6
DFO-2	41.1	45.3	96	18.7	20.8	29
DFO-3	46.9	42.1	63.5	96	28.1	27.5
MDMR	40.6	44.2	40.6	46.9	96	20.6
NMFS	41.5	56.4	52.1	48.9	44.7	94

Table 6. Number of otoliths and age distribution for yearly combinations of trays.

Period	Otoliths	Ages	Period	Otoliths	Ages
1984-1992	1794	1-9	1985-1992	1602	2-9
1984-1993	1979	1-10	1985-1993	1787	2-10
1984-1994	2172	1-11	1985-1994	1980	2-11

Table 7. Summary of how each reader classified the herring otoliths and the number read by each reader for the dominant year-class tracking study.

Category	DB	DFO1	DFO3	MDMR	NMFS1	NMFS2
Good	1787	1777	1481	1197	797	535
Bad	0	2	215	563	634	398
Ugly	8	16	99	35	365	863
# Read	1787	1779	1697	1761	1449	1148
Percent read	99.55	99.11	94.54	98.11	80.72	63.96

Table 8. Summary of mean, minimum, and maximum age by source for the DYC study.

	DB	DFO1	DFO3	MDMR	NMFS1	NMFS2
Number	1787	1779	1697	1761	1449	1148
Min Age	2	2	2	2	2	2
Max Age	14	12	12	12	10	11
Mean Age	5.62	4.60	5.30	4.86	4.73	4.29

Table 9. Summary of the number of otoliths aged, the percent agreement and the Bowker's test of symmetry for the readers and the database (DB) ages from the dominant year-class tracking study. Above the diagonal equals the number of otoliths read in common and below the diagonal the percent agreement and significance (P<0.01).

	Reader					
Reader	DB	DFO1	DFO3	MDMR	NMFS1	NMFS2
DB		1775	1694	1757	1445	1143
DFO1	42.8**		1689	1753	1446	1132
DFO3	67.4**	46.4**		1675	1411	1111
MDMR	48.5**	51.8**	52.7**		1414	1145
NMFS1	54.1**	56.2**	55.6**	55.7**		1062
NMFS2	40.9**	55.2**	44.4**	51.1**	65.8**	

Table 10. Comparison of primary reader (Primary) assigned ages with the second DFO reader (Reader 2), the database, and the external expert reader (Reader 3).

Readers		Type	Number	% Agreement	CV %	Bias	Significance
Primary	Reader 2	All	280	52.9	9.85	51.51	P<0.01
Primary	Reader 2	Good	103	73.8	6.41	9.10	ns
Primary	Reader 2	Good/Poor	225	59.1	8.52	33.14	P<0.01
Primary	Database		196	48.0	8.81	41.50	P<0.01
Primary	Reader 3	Good	187	51.3	7.32	11.48	ns
Primary	Primary		207	74.4	5.49	20.56	P<0.05

Table 11. Summary of the comparisons for the random selected otoliths from each year reaged between 1999 and 2005, and the newly aged otoliths from 2009. The brown indicates comparisons that did not meet the established acceptance criteria while the green indicates those that were acceptable.

Year	Readers		Type	Number Otoliths	% Agreement	CV %	Chi Square	Significance
1999	New	Reader 3	All	88	58.0	8.28	23.0	ns
	New	Reader 3	Good	51	80.4	2.58	3.2	ns
2000	Reader 2	Reader 3	Good	325	88.3	1.45	18.4	ns
2001	New	Reader 3	All	82	65.9	5.75	20.2	ns
	New	Reader 3	Good	50	80.0	2.97	10.0	ns
2002	New	Reader 3	All	89	52.8	10.31	34.7	P<0.05
	New	Reader 3	Good	48	68.8	9.22	15.0	ns
2003	Reader 2	Reader3	All	82	62.2	7.35	20.0	ns
	Reader 2	Reader3	Good	45	80.0	2.98	6.3	ns
2004	Reader 2	Reader3	All	97	67.0	5.49	16.2	ns
	Reader 2	Reader3	Good	81	82.7	4.41	8.7	ns
2005	New	Reader 3	All	94	62.8	6.61	19.7	ns
	New	Reader 3	Good	80	73.8	6.13	14.3	ns
	New	Reader 3	All	88	69.3	6.33	11.2	ns
	New	Reader 3	Good	57	84.2	2.28	9.0	ns
2009	New	Reader 2	All	221	81.9	3.12	20.3	ns
	New	Reader 2	Good	188	88.3	2.96	15.5	ns
	New	Reader 3	All	217	81.8	3.18	18.2	ns
	New	Reader 3	Good	179	89.9	3.11	10.0	ns
	Reader 2	Reader 3	All	216	82.5	3.31	29.1	P<0.01
	Reader 2	Reader 3	Good	179	87.2	2.64	19.5	ns

Table 12. Number of re-aged otoliths and the total number of acceptable otoliths used in the construction of a revised CAA. Note the otoliths between 2006 and 2009 were not aged until after the ageing issues were resolved.

	Year										
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number Re-aged	931	725	818	1054	705	810	748	-	-	-	-
Total for ALK	1197	2148	3116	1386	2278	1030	748	1270	988	811	3929

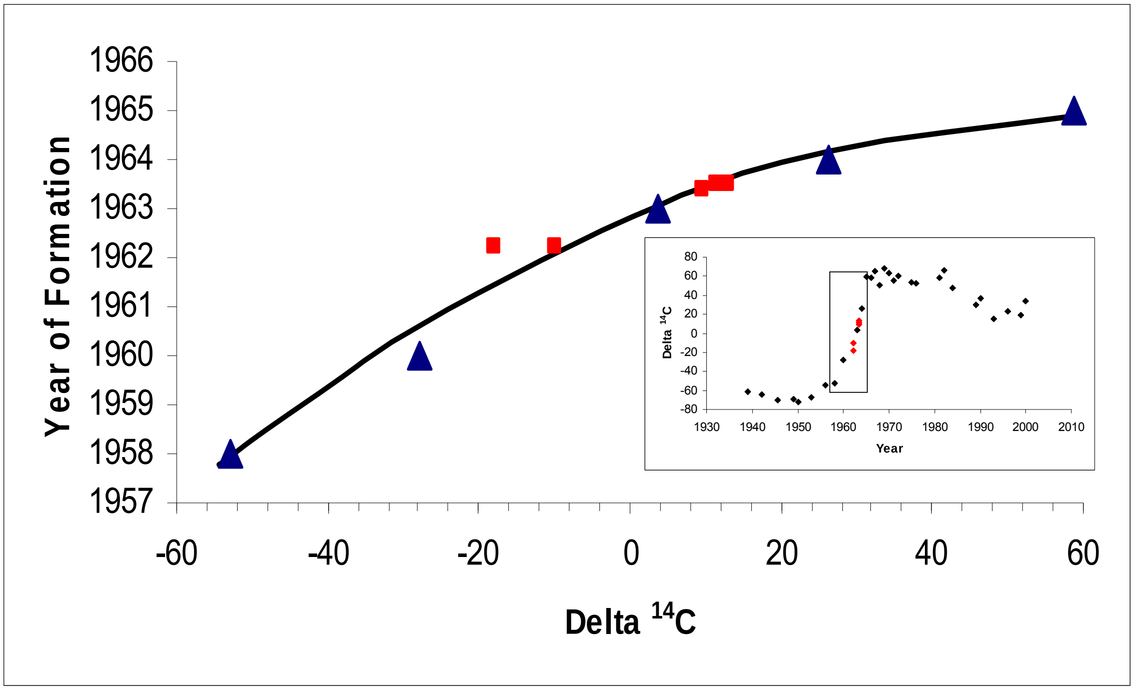

Figure 1. Relationship of $\Delta^{14}C$ observed in young herring otoliths of known age (red) and the year of core formation relative to the Northwest Atlantic chronology (black triangles) for the years 1958 to 1965. A quadratic equation was fitted to the data to describe the relationship. The insert shows the bomb radiocarbon ($\Delta^{14}C$) reference chronology for the Northwest Atlantic (black symbols) with the rectangle high-lighting the region where the main figure was derived.

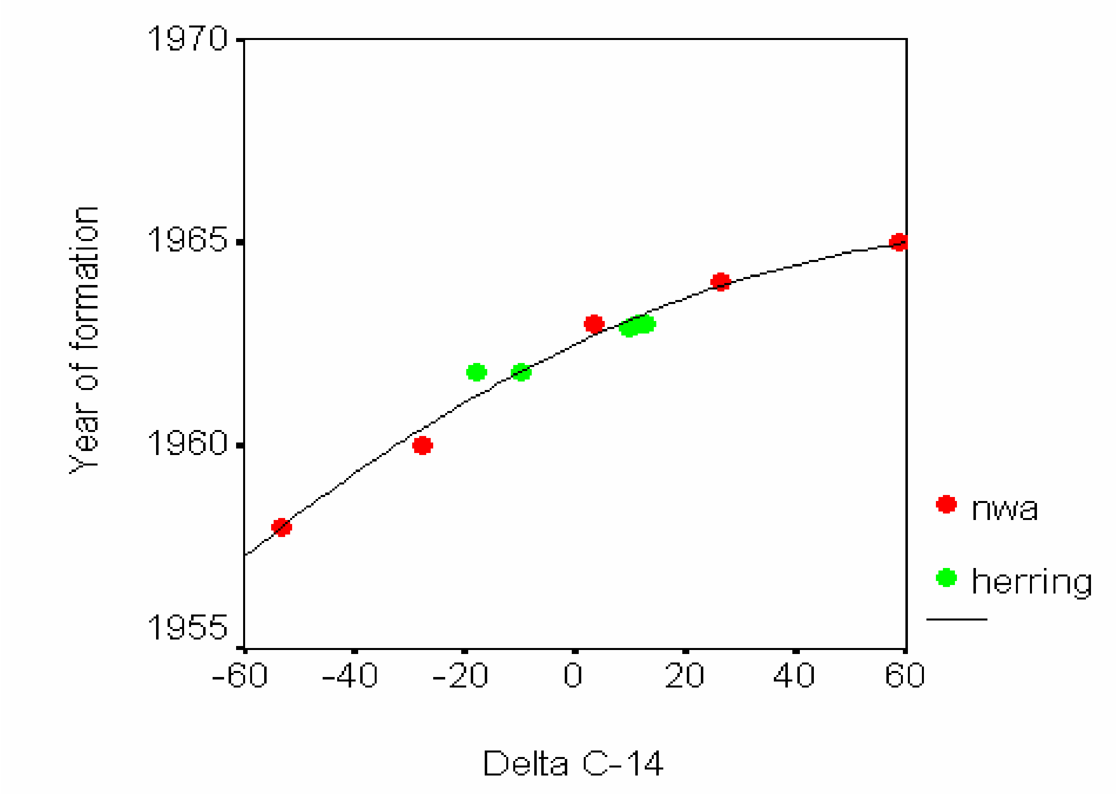

Figure 2. Relationship between year of otolith formation and Delta C^{14} for the Northwest Atlantic and herring of known age from the 1962 year class.

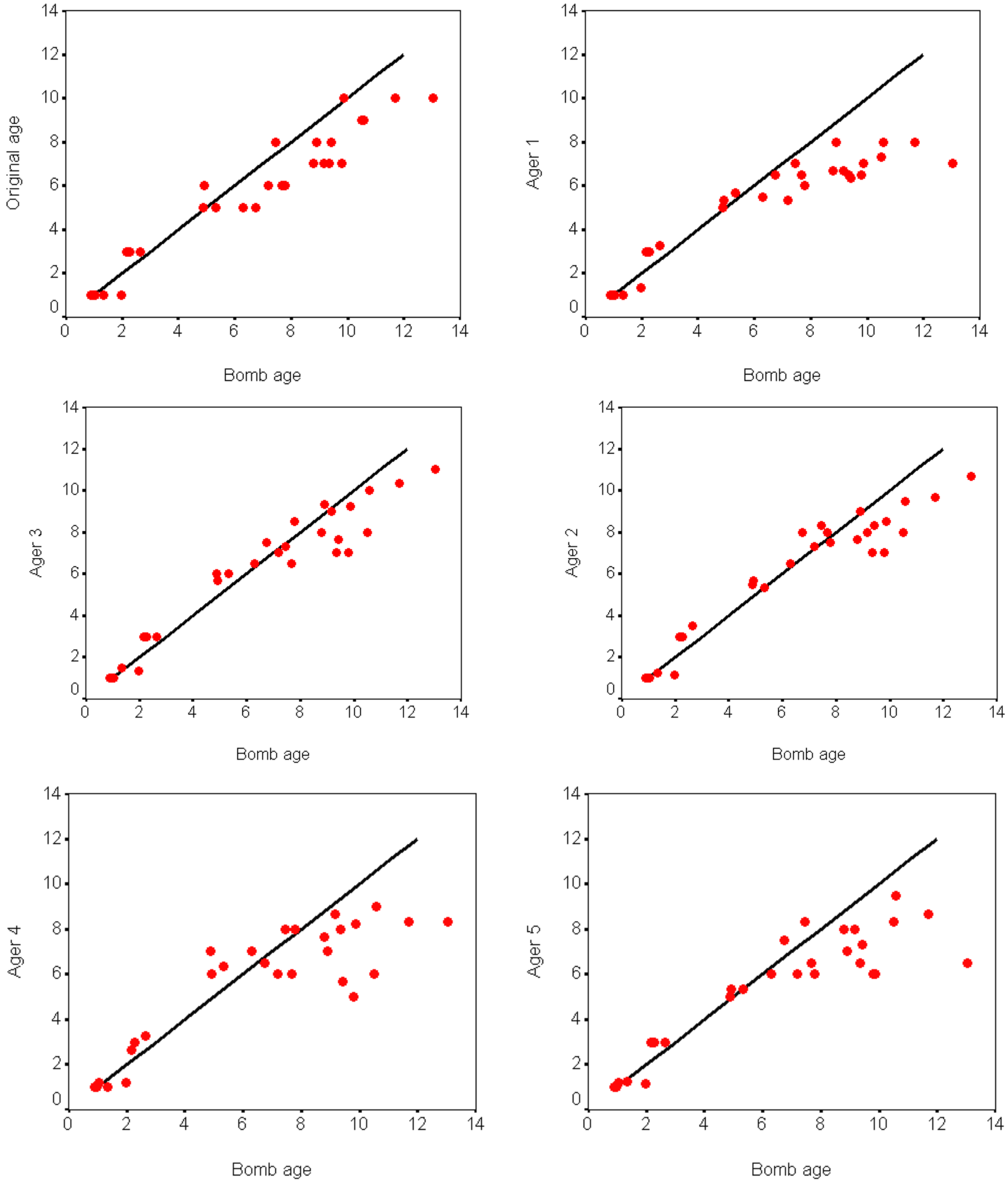

Figure 3. Comparison of bomb radiocarbon estimated age with reader age for all readers involved in the study including the original assigned age from the database (DB).

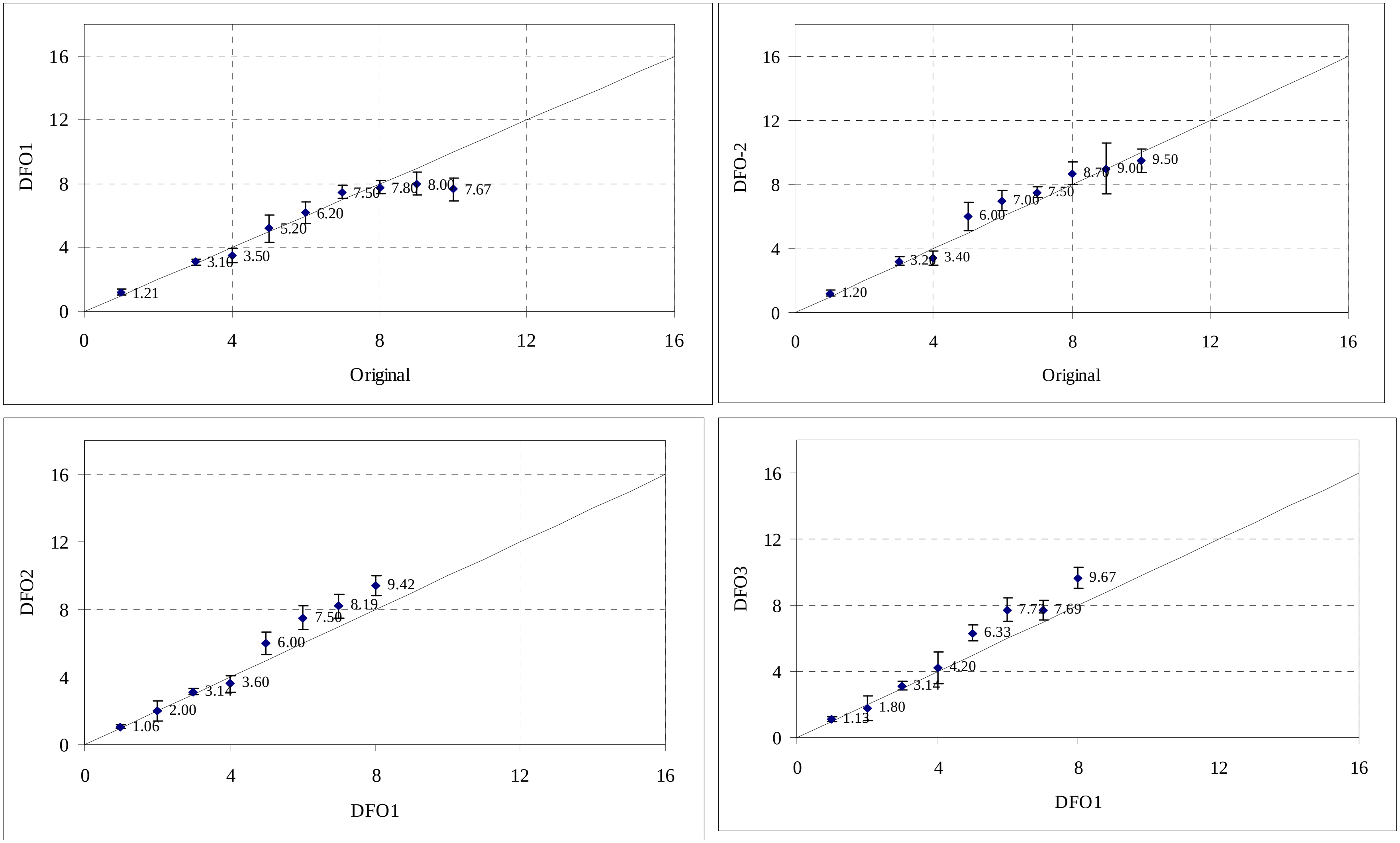

Figure 4. Comparison of reader assigned ages for otoliths (1963-1971) used for the bomb radiocarbon assay study.

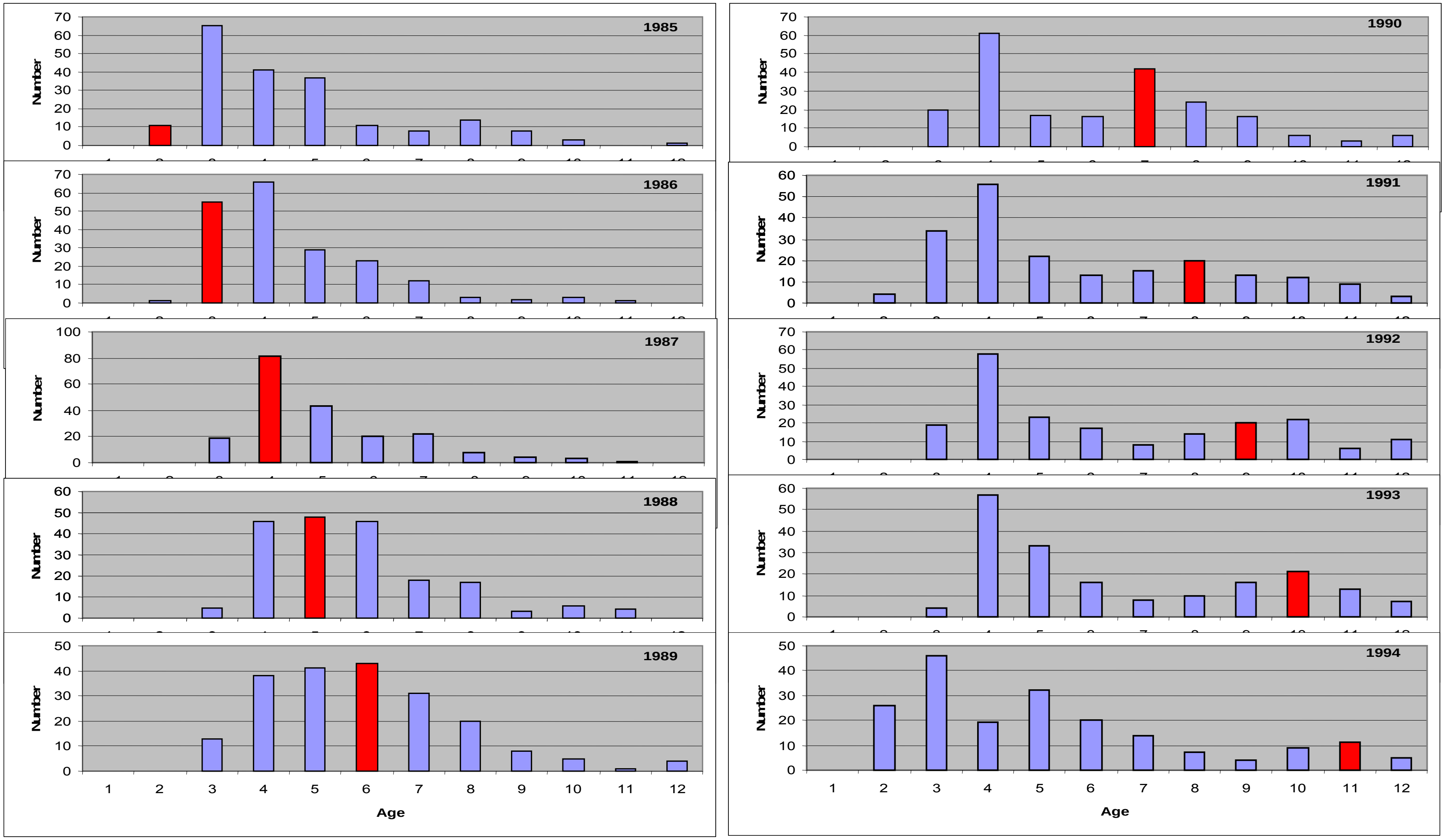

Figure 5. Age distribution of herring otoliths by year (1985 – 1994) selected for dominant year-class tracking study. The 1983 year-class is labelled in red for each year.

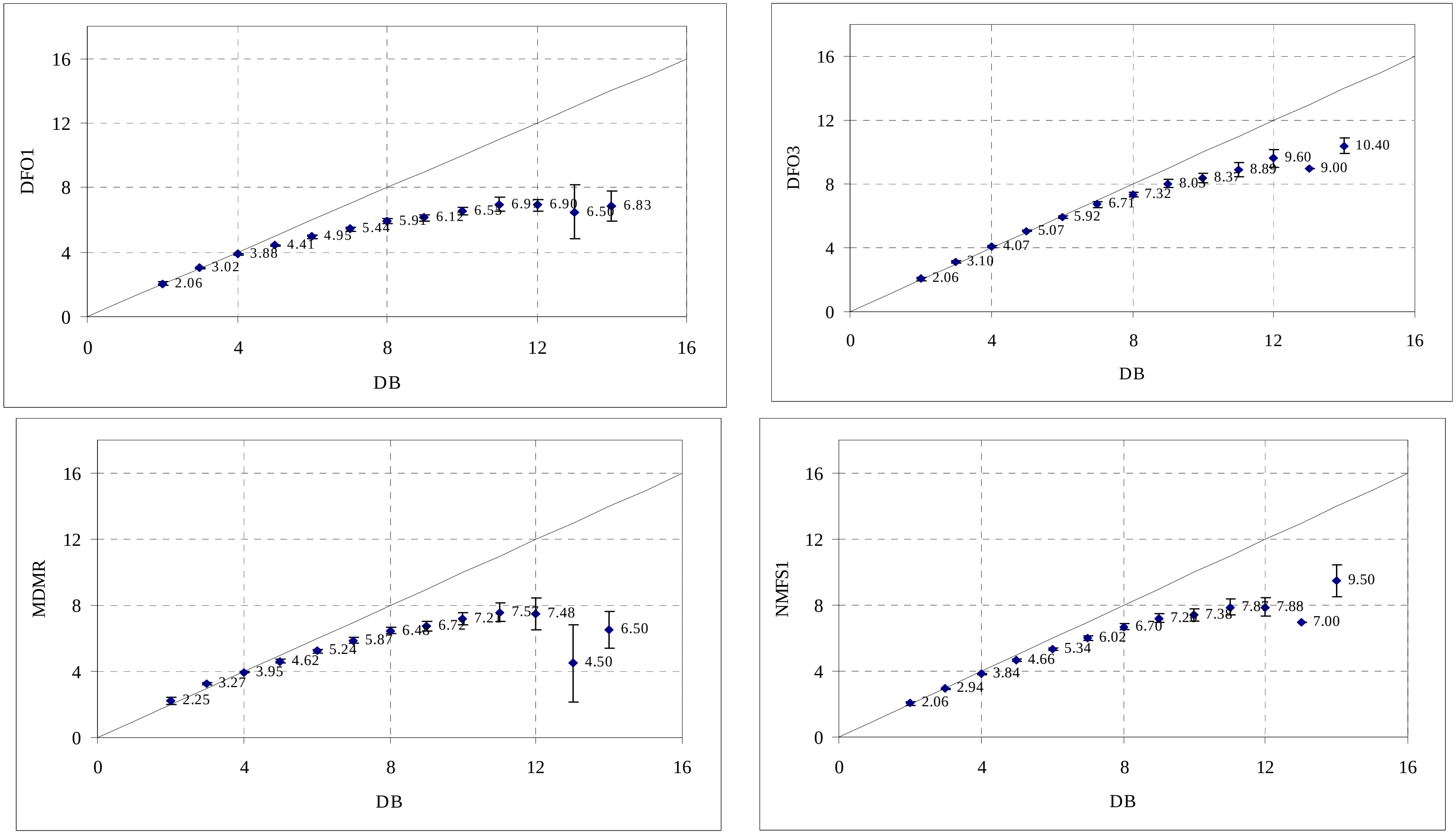

Figure 6. Comparison of reader ages with the database ages originally assigned to otoliths selected for the DYCT study.

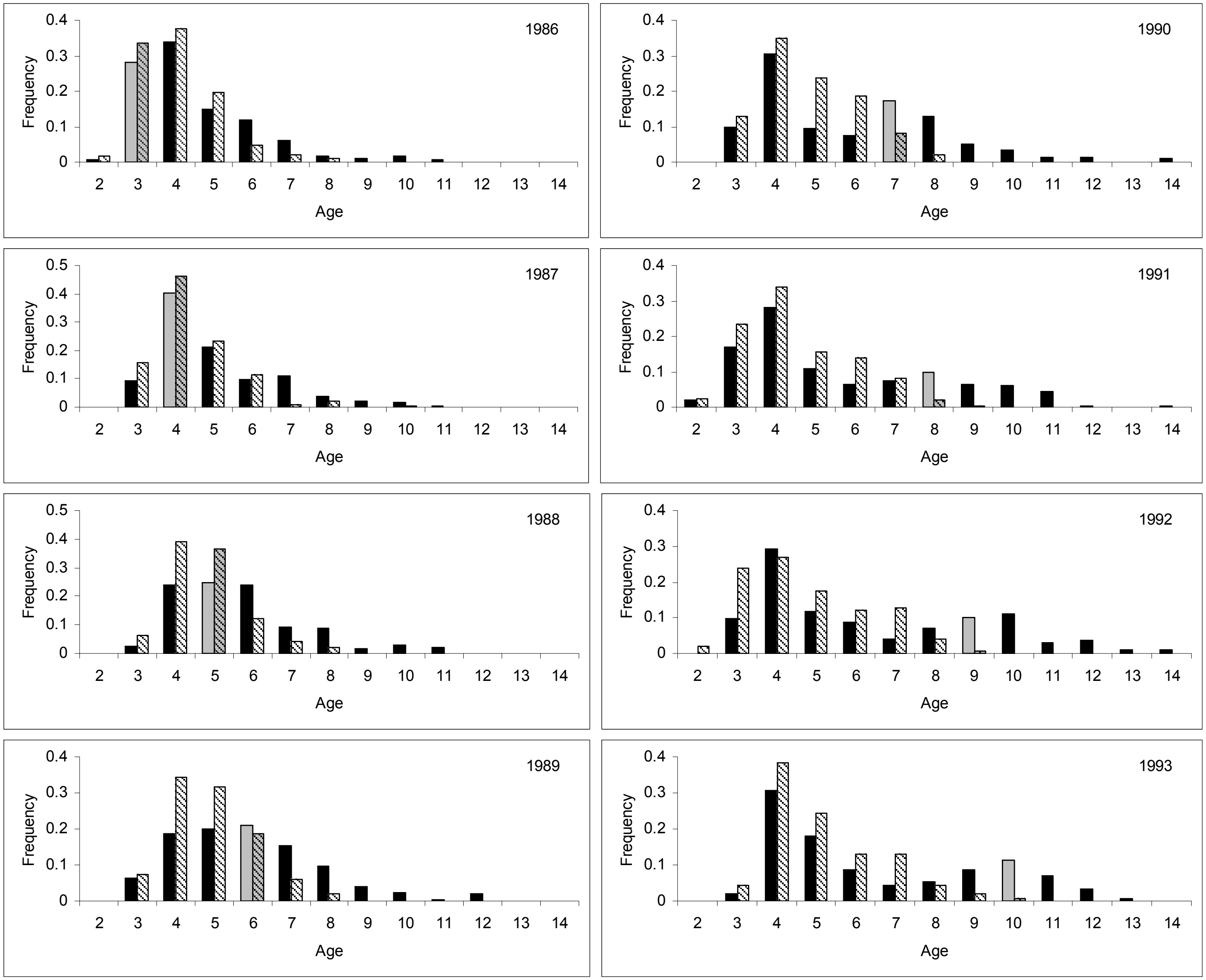

Figure 7. Distribution of DYCT ages by year for the original database age (solid), SABS Primary Reader (hatched). The 1983 year-class is presented in grey.

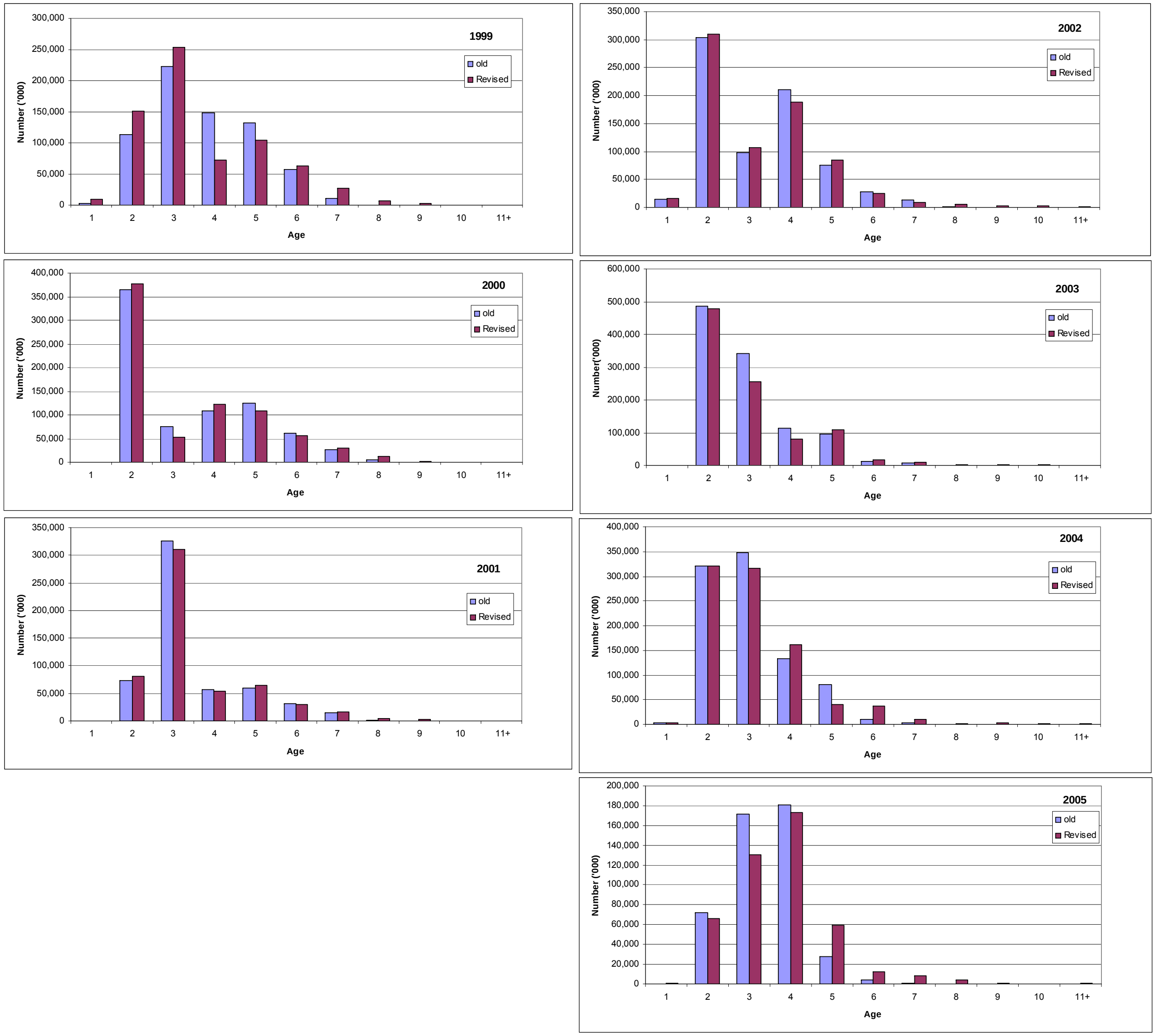

Figure 8. Comparison of the original and the revised catch at age for the period 1999-2005.

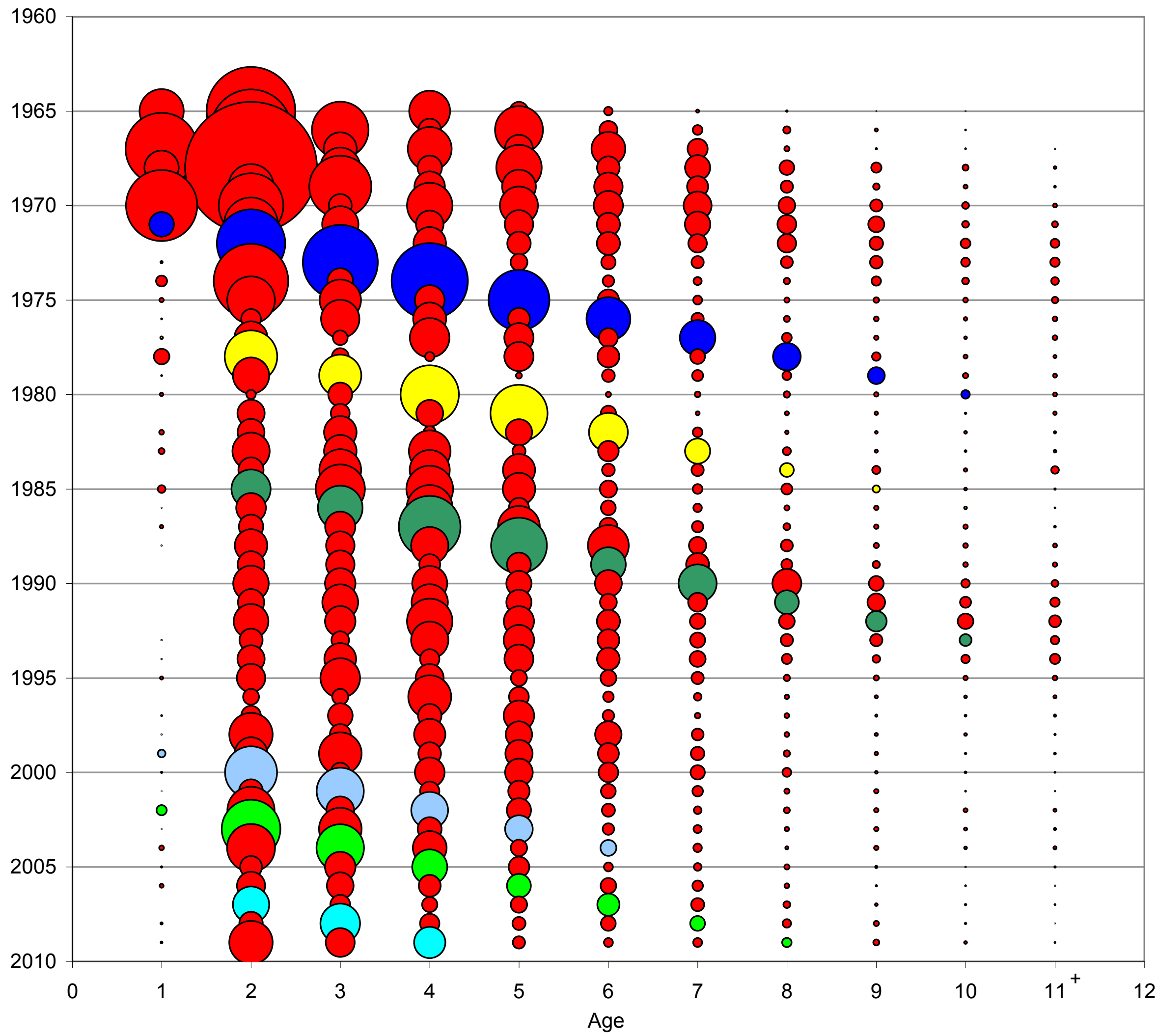

Figure 9. The 4WX herring catch at age for the entire time series from 1965 to 2009, including the 1999-2005 revisions.