Available online at www.sciencedirect.com

SCIENCE DIRECT®

Fisheries Research 80 (2006) 312–321

FISHERIES RESEARCH

www.elsevier.com/locate/fishres

# Precision and bias in the age determination of blue whiting, *Micromesistius poutassou* (Risso, 1810), within and between age-readers

G.R. Power[a,*], P.A. King[a], C.J. Kelly[b], D. McGrath[a], E. Mullins[b], Ole Gullaksen[c]

[a] *Commercial Fisheries Research Group, Life Sciences Department, Galway–Mayo Institute of Technology, Galway City, Ireland*
[b] *Fisheries Science Services, Marine Institute, Parkmore, Galway City, Ireland*
[c] *Institute of Marine Research, Bergen, Norway*

Received 1 September 2005; received in revised form 8 March 2006; accepted 14 March 2006

**Abstract**

Age based assessment forms the basis of management advice for the heavily exploited combined stock of blue whiting (*Micromesistius poutassou*) in the Northeast Atlantic. However, little historic attention has been given to the reliability of age estimates gathered from several nations involved in this assessment. Using recognised graphical and statistical approaches, bias and precision was investigated for experienced age-readers of blue whiting. Significant linear bias was found to exist between age-readers, with ages differing, on average, by 1 year for important year classes. Indications are that spawning checks and split rings affect the interpretation of annuli for some age-readers. An experience gradient became evident during the analysis; more experienced age-readers had greater levels of precision. Within reader precision was found to be higher than between reader precision; however, within reader bias was also evident with two out of three age-readers systematically revising ages downwards when re-ageing otoliths. Results indicate that differences exist on a limited international level for the age determination of blue whiting. This issue now needs to be addressed by the scientific community.


*Keywords:* Blue whiting; Age determination; Otoliths; Precision; Bias; Ageing error

## 1. Introduction

The blue whiting, *Micromesistius poutassou* (Risso, 1810), is a widely distributed species in the Northeast Atlantic, inhabiting waters from the Northwest coast of Africa to the Barents Sea (Fig. 1). For assessment purposes, blue whiting in this large geographical area is treated as a single large stock (ICES, 2004). This combined stock boasts relatively few year classes (Fig. 2), presently supports annual commercial landings of over 2,500,000 tonnes and is exploited by several nations including, Norway, Russia, U.K., Ireland, Spain, Germany, Denmark and Faroe Islands (ICES, 2004).

The current management of blue whiting in the Northeast Atlantic largely depends upon the results of annual age-based assessments. These assessments give fishery managers valuable information on the commercial exploitation, age composition and stock recruitment dynamics of blue whiting. Age estimation of blue whiting by intact ‘reading’ of sagittal otoliths is the most commonly accepted method in use today by fisheries scientists in the northern hemisphere. In light of the current uncertainties on the productivity and exploitation of blue whiting, it is of great importance that fishery scientists closely monitor related age reading criteria and procedures, while endeavouring to provide sound scientific advice on this species to fishery managers. Age determination errors can cause serious biases in the estimation of vital parameters such as growth, mortality and recruitment and subsequently affect yield and production estimates (Tyler et al., 1989). Surprisingly for blue whiting, there is

* Corresponding author. Tel.: +353 91752481; fax: +353 91758412.
*E-mail address:* nivagp@hotmail.com (G.R. Power).

0165-7836/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2006.03.031

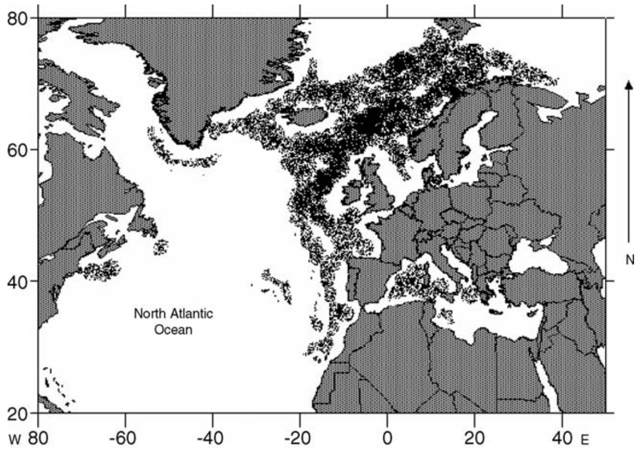


Fig. 1. Generalised distribution of blue whiting in the Northeast Atlantic and Mediterranean (adjusted from Zilanov, 1968).

no published information available on the investigation of such errors.

In age determination studies, the term "precision" is used to describe "agreement" or variability between readings of the same specimen by the same or different age-readers. The term "accuracy" is reserved to describe a comparison of ages generated by age-readers with the "true" age for specimens of known age (Kimura and Lyons, 1991). Only by mark-release-recapture studies, or use of known-age fish, can all age classes in a population be validated and accuracy proven (Beamish and McFarlane, 1983). Chronometric analyses of disequilibria in the naturally occurring uranium decay series nuclides of $^{226}$Ra and $^{210}$Pb may also be used to radiometrically validate the age determinations of longer-lived species (Bennett et al., 1982). However, in the absence of a known-age reference collection, ageing consistency is the best that can be achieved (Campana et al., 1995). Bias can be both relative, "the systematic over or under estimation of age compared to the modal age" or absolute, "the systematic over or under estimation of age compared to the true age" (Eltink et al., 2000). This paper addresses relative bias only.

Previous ageing studies have identified many useful statistical indices for the identification and analysis of bias and precision. Kimura et al. (1979) measured the precision between age-readers for yellowtail rockfish (*Sebastes flavidus*) by random effects analysis of variance, while Beamish and Fournier (1981) proposed the index of 'average percent error' (APE) over the more commonly utilized 'percent agreement' in historical literature. Beamish and Fournier (1981) also proposed that the index of APE was independent of the age of particular species, therefore enabling age determination precision comparisons between species of different longevity. Chang (1982) later modified the index of APE to average coefficient of variation (CV), and this has been the widely adapted test of the reproducibility of age readings in both past and recent studies (ICES, 1994; Campana, 2001). Baker and Timmons (1991) later used an index called the 'ageing error' (AE), an index modified from Sharp and Bernard's (1988) sampling standard error. Campana and Moksness (1991) utilized a combination of CV and standard deviation while investigating the precision of herring age estimates, while work by Hoenig et al. (1995) focused on analyses using chi-squared tests of symmetry. In a later paper, Campana et al. (1995) outlined a statistical and graphical method to determine the consistency of age determinations and recommended the combination of the index of CV and age bias plots for future investigations. This combined approach has had wide application in recent ageing studies (Fossen et al., 2003; Stransky et al., 2005) and has formed the analytical basis for many quality control laboratory investigations and ageing workshops (Eltink et al., 2000; Walsh and Burnett, 2002). Other recent studies have proposed the use of ageing error matrices to identify ageing error (Heifetz et al., 1998) incorporating the use of fish of known age, however such reference collections are not available in many ageing studies.

This study sets out to determine the consistency of age determinations both within reader and between readers for blue whiting using a slightly modified version of the established combined graphical and statistical approach as proposed by Campana et al. (1995).

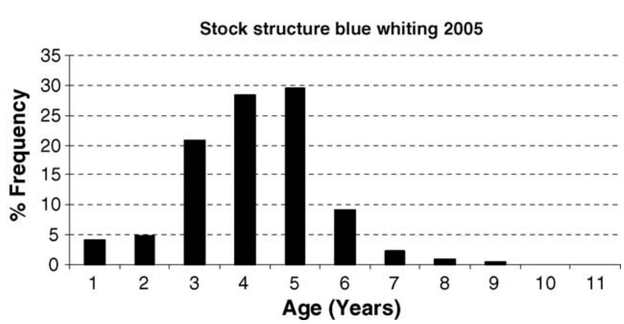


Fig. 2. Age frequency distribution of blue whiting in the spawning area to the west and northwest of Ireland during March–April 2005, from combined ICES fishing areas VIIb, VIIc, VIIj, VIIk, VIa, VIb, Vb1, Vb2. Adapted from ICES (2005).

## 2. Materials and methods

### 2.1. Study material

Five independent samples of 'dry stored' sagittal otoliths of blue whiting were chosen for comparison for the present study. This was undertaken in order to increase the sample variability and thus avoid any possible bias that may result from the 'sample effect' of re-ageing the same sample several times. Samples of otoliths were taken from fish, caught by both Irish and Norwegian commercial fishing vessels, during the 2003 and 2004 spring fishery to the West and Northwest of Ireland (ICES fishing areas VIIk, VIIc, VIb and VIc). Sample sizes ranged from 180 to 300 otoliths and were composed of otoliths taken from fish of commercial origin. Younger age groups were, therefore, mostly absent, with samples displaying similar age structures. That said however, one stratified

Table 1
Participant age-readers and laboratories in the present study

| Name | Institute |
|---|---|
| Gavin Power | Galway–Mayo Institute of Technology (GMIT) Ireland/Marine Institute Ireland. |
| Eugene Mullins | Marine Institute, Ireland. |
| Ole Gullaksen | Institute of Marine Research, Norway. |

random sample of otoliths, including both younger and older age groups, was chosen to facilitate a more detailed analysis between two specific age-readers. This simultaneously formed part of a domestic quality control exercise within the host institute. The remaining four samples of otoliths used in this study were randomly chosen. These five independent samples were labelled as samples S1–S5.

### 2.2. Ageing procedure

Three laboratories were chosen for the experiment, one academic research laboratory and two international governmental fishery laboratories. Institute and participant age-reader details are given in Table 1. The method of age estimation employed, i.e. microscope magnification and light intensity levels, was standardised as far as was possible between laboratories. Otoliths were immersed in fresh water, soaked for 24 h, examined stereoscopically (×6 magnification) using reflected light and read whole. Otoliths were viewed against a black background to aid the contrast between opaque summer growth and translucent winter growth zones. Age was determined by counting the number of recognisable winter hyaline zones (Jenson, 1965) along the longitudinal axis from the nucleus to the edge of the posterior tip, on the concave side of the otolith (Fig. 3). The age-readers involved in this study were all considered to be experienced in ageing blue whiting and are henceforth referred to as age-readers A–C.

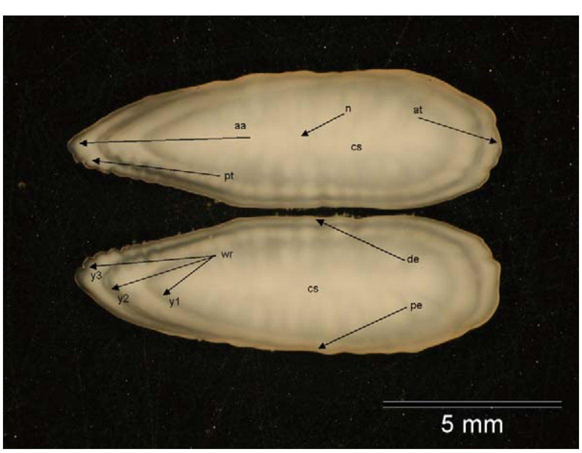


Fig. 3. Stereoscopic view of a pair of sagittal otoliths (×8 magnification) from a 3-year-old blue whiting. Otoliths were immersed in water and examined under reflected light. The convex sides of the otoliths are not visible in the image. n, nucleus; aa, axis of ageing; cs, concave or distal side; at, anterior tip; pt, posterior tip; pe, proximal or ventral edge; de, dorsal edge; wr, winter rings or hyaline annuli.

### 2.3. Age determination comparisons

A schematic diagram of otolith comparisons is given in Fig. 4. For the analysis of inter-reader precision and bias, samples S1–S3 were firstly analysed by all three age-readers. Each sample was subsequently sent to a second age-reader allowing three independent estimates of inter-reader error. For the analyses of intra-reader precision and bias, two age-readers were asked to re-age 'in house' samples of otoliths, for which previous age estimations existed. The age-reader at the academic laboratory carried out two independent age determinations on a sample that had previously been circulated between fisheries laboratories. This was undertaken as an effort saving exercise only, as results from this comparison would have no affect on any previous inter-reader compar-

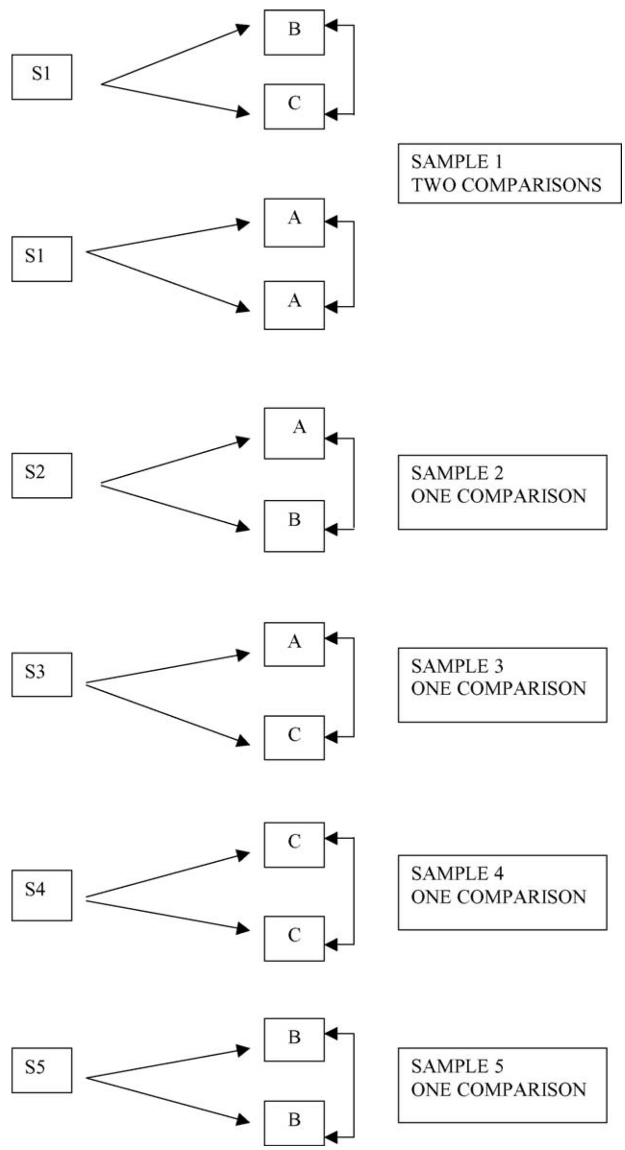


Fig. 4. A schematic representation of sample analysis for the comparisons of age estimates between and within age-readers A–C, during the present study.

isons carried out using this same sample. All age readings were carried out without reference to previous age determination data.

### 2.4. Data analysis

Replicated age determination data was analysed for precision and bias using a combination of age frequency tables, age difference plots, age bias plots, CV plots, as well as parametric and nonparametric statistical analysis. Age frequency tables form a central component of many paired age comparisons but are not particularly well suited to the detection of age differences between age-readers (Campana et al., 1995). A more useful graphical tool is the use of age-difference plots where the difference between two age-readers is plotted as a function of one of the sets of ages (Campana et al., 1995).

For the analysis of bias, simple linear regression was carried out on pair-wise comparisons and was performed using[1] GRAPHPAD PRISM Version 4.03 for Windows (2005). A slope other than one suggests inconsistency in the interpretation of annuli by one of the age-readers, an intercept other than zero suggests systematic differences between two age-readers, for example due to different interpretation of the first annulus (Campana et al., 1995). The regression slope and intercept were tested for significant differences from one and zero, respectively, using an *F*-test for slopes ($\alpha = 0.05$). Use of the nonparametric One Sample Sign test of the median is proposed here to identify potential ageing bias, rather than the commonly utilized Paired *t*-test. The *t*-test operates under the assumption of normally distributed paired differences, a characteristic not always observed in datasets subject to significant bias. The One Sample Sign test analyses the paired differences allowing the identification of both positive and negative bias for comparative age readings ($\alpha = 0.05$), and was carried out in[2] MINITAB® release 14 Statistical Software for Windows. Age bias plots (Campana et al., 1995), a useful graphical tool to identify both linear and nonlinear bias, were applied to the paired data of inter-reader analysis. Age bias plots were constructed by plotting the mean (±95% confidence intervals) of ages assigned by one age-reader, for all fish estimated as belonging to a specific age category by a second age-reader. Within such plots, results were interpreted through reference to a 1:1 equivalence line.

For the analysis of precision, both percent agreement and CV data were computed for comparative age determinations. The CV is a commonly used statistical index of precision, is expressed as the ratio of the standard deviation to the mean, and can be averaged across many fish. A nonparametric statistical test for asymmetry, the Wilcoxon Two Sample Test (Conover, 1980), can be applied to paired age data to test for imprecision or significant differences in age distribution, via test of the median ($\alpha = 0.05$). This test has been widely applied to test for bias in various published ageing studies (Campana et al., 1995; Bergstad et al., 1998; Peltonen et al., 2002) by analysis of the ratio of positive and negative paired differences only. However, in this study the Wilcoxon Two Sample Test was used as a test for precision by the inclusion of tied ranked values in rank sum analysis. This method allows the testing of paired age estimates for differences in median values, thus compares overall sample distribution rather than just the analysis of paired differences between samples. Data analysis, using the Wilcoxon test, was carried out through an online package provided by a web accessible server of the[3] Department of Humanities, University of Amsterdam. Lastly, plots of both the CV and percentage agreement at age were plotted for all between reader comparisons. The CV was firstly calculated for all individual paired age estimates for all fish. CV values were then averaged across fish that had been assigned within age classes by one age-reader. This resulted in a mean CV value at age relative to one age-reader only, but was suitable for relative comparisons in a CV plot. The percentage agreement at age was calculated in a similar manner relative to all fish assigned specific ages by one age-reader. CV plots were constructed to graphically analyse any imprecision identified between age-readers. These plots can identify the fish age at which differences between age-readers develop and imprecision becomes problematic. For a more elaborate description of age difference plots, age bias plots and CV plots, see Campana et al. (1995).

## 3. Results

### 3.1. Between reader bias

In Table 2, age frequency matrices are presented for each of the pair-wise age comparisons for samples S1–S3, showing the matched pairs of age determinations. Age difference plots, which highlight major systematic differences between two sets of age readings, have been presented in Fig. 5. Within these plots there is evidence of some linear differences, or an underlying bias, in the comparative age estimates of age-readers A and C, and B and C. Differences were less evident in comparisons between age-readers A and B.

Simple linear regression was carried out on each of the pair wise age comparisons and is summarised in Table 3. Results of regression analysis, and *F*-test of slopes, indicate that highly significant inconsistencies exist in comparisons between age-readers A and C and B and C, with significant inconsistency observed between age-readers A and B. Furthermore, significant systematic bias was observed in com-

[1] GRAPHPAD ANALYSIS software, homepage http://www.graphpad.com/Accessed April 7, 2005.

[2] MINITAB 14 statistical software for windows, homepage http://www.minitab.com/Accessed August 5, 2005.

[3] Department of Humanities, University of Amsterdam, Institute of Phonetic Sciences http://fonsg3.let.uva.nl/Service/Statistics.html Accessed February 16, 2005.

Table 2
Age frequency tables illustrating paired comparisons of age estimates from age-readers A–C for samples S1–S3

| Age (years) estimated by | Age (years) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Total |
| Reader A (S2) | Estimated by reader B (S2) | | | | | | | | | | | | | |
| 0 | 25 | | | | | | | | | | | | | 25 |
| 1 | | 14 | 2 | | | | | | | | | | | 16 |
| 2 | | 1 | 20 | | | | | | | | | | | 21 |
| 3 | | | 5 | 33 | | 1 | 1 | | | | | | | 40 |
| 4 | | | | 7 | 23 | 6 | | 1 | | | | | | 37 |
| 5 | | | | | 10 | 10 | 5 | | | | | | | 25 |
| 6 | | | | 1 | 1 | 4 | 14 | 2 | 2 | | | | | 24 |
| 7 | | | | | | 1 | 6 | 8 | 6 | 1 | | | | 22 |
| 8 | | | | | | | 1 | 6 | 13 | 2 | | | | 22 |
| 9 | | | | | | | | | 1 | 6 | 1 | | | 8 |
| 10 | | | | | | | | 1 | | 2 | 2 | 1 | 1 | 7 |
| 11 | | | | | | | | | | | 3 | | | 3 |
| 12 | | | | | | | | | | | | | 1 | 1 |
| Total | 25 | 15 | 27 | 41 | 34 | 22 | 27 | 18 | 22 | 11 | 6 | 1 | 2 | 251 |
| Reader A (S3) | Estimated by reader C (S3) | | | | | | | | | | | | | |
| 1 | | 32 | | | | | | | | | | | | 32 |
| 2 | | 3 | 14 | 1 | | | | | | | | | | 18 |
| 3 | | | 22 | 88 | 5 | | | | | | | | | 115 |
| 4 | | 1 | 2 | 34 | 49 | 2 | | | | | | | | 88 |
| 5 | | | 1 | 8 | 18 | 12 | 1 | | | | | | | 40 |
| 6 | | | | | 6 | 11 | 5 | | | | | | | 22 |
| 7 | | | | | | 5 | 7 | 11 | 2 | | | | | 25 |
| 8 | | | | | | | 4 | 3 | | | | | | 7 |
| 9 | | | | | | 1 | | | 1 | | | | | 2 |
| 10 | | | | | | | | | | 1 | | | | 1 |
| Total | | 36 | 39 | 131 | 78 | 31 | 17 | 14 | 3 | 1 | | | | 350 |
| Reader C (S1) | Estimated by reader B (S1) | | | | | | | | | | | | | |
| 1 | | | | | | | | | | | | | | |
| 2 | | | 32 | 17 | 1 | | | | | | | | | 50 |
| 3 | | | 2 | 93 | 22 | 2 | | | | | | | | 119 |
| 4 | | | | 2 | 55 | 12 | 3 | 2 | | | | | | 74 |
| 5 | | | | | 7 | 9 | 5 | 5 | 1 | | | | | 27 |
| 6 | | | | | | 2 | 9 | 4 | 1 | | | | | 16 |
| 7 | | | | | | | 1 | 1 | 6 | 1 | | | | 9 |
| 8 | | | | | | | | 2 | | 2 | | | | 4 |
| 9 | | | | | | | | | | | | | | |
| Total | | | 34 | 112 | 85 | 25 | 18 | 14 | 8 | 3 | | | | 299 |

parisons between all three age-readers, as intercepts differed significantly from a value of zero (Table 3).

Results from the nonparametric One Sample Sign test identified significant negative bias in comparisons between age-readers A and C and B and C, demonstrating under-ageing by age-reader C in relation to age-readers A and B. No obvious bias was observed between age-readers A and B using this test (Table 3). Age bias plots have been presented in Fig. 6, and fail to identify significant under-ageing or over-ageing in comparisons between age-readers A and B. However, negative bias is again apparent in graphical comparisons between age-readers A and C and B and C (Fig. 6), and assumes an almost linear trend. This systematic under-ageing of fish by age-reader C is most apparent in fish of age four and older (Fig. 6).

### 3.2. *Within reader bias*

Regression analysis identified age-reader A as ageing otoliths inconsistently. Furthermore, systematic differences in age determination were identified within all three age-readers, by test of the regression intercept (Table 3). This systematic difference, or bias, was found to be significant from results of the nonparametric One Sample Sign test, with age-readers A and C revising ages downwards and age-reader B revising ages upwards when re-ageing otoliths (Table 3).

### 3.3. *Between reader precision*

The nonparametric Wilcoxon Two Sample test for precision, found similar results to that of the One Sample Sign test,

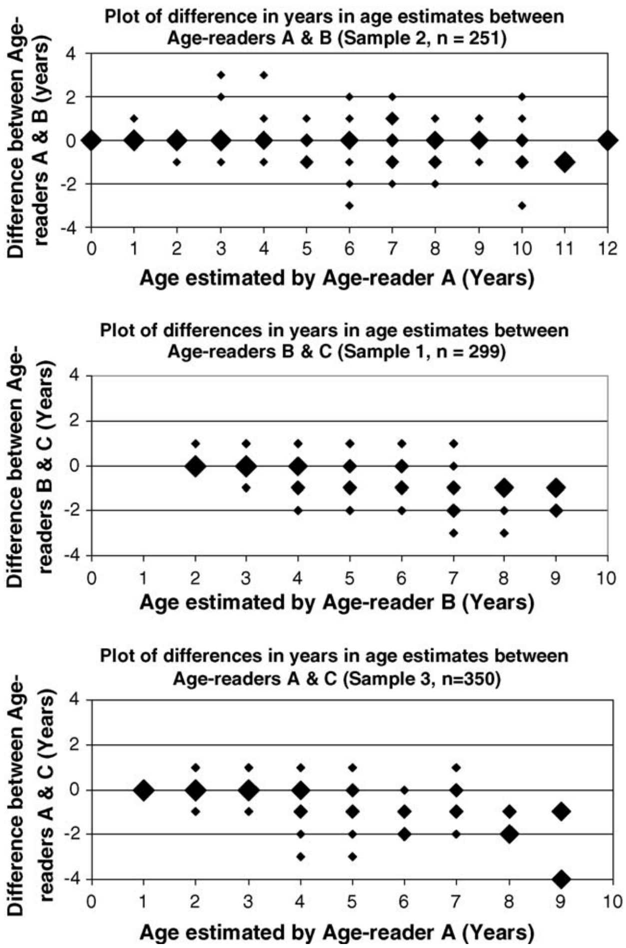


Fig. 5. Age difference plots for each of the three pairwise age comparisons in Table 2. Plotted points represent one or more observations and are weighted by percentage frequency of observations ((◆) = 0–25%; (◆) = 26–50%; (◆) = 51–75%; (◆) = 76–100%). (Top): no distinct bias visible; (middle and bottom): negative bias visible.

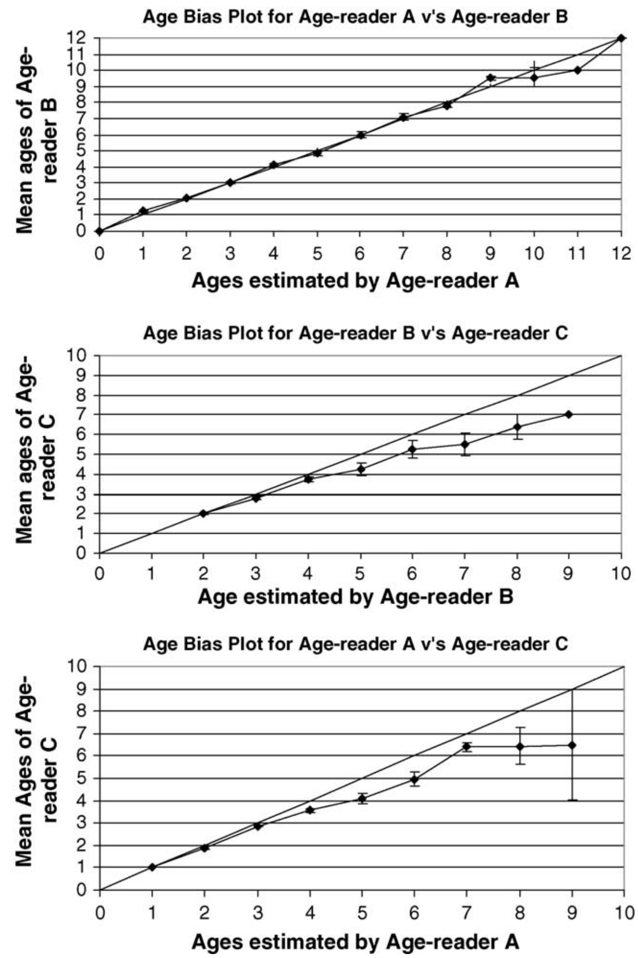


Fig. 6. Age bias plots for each of the three paired age comparisons presented in Table 2. Error bars represent the 95% confidence interval about the average age assigned by one age-reader for all fish assigned a certain age by another age-reader. The 1:1, or zero difference line (solid line) is indicated.

highlighting significant imprecision in comparisons between age-readers A and C and B and C (Table 3). Percentage agreement and CV have been calculated for precision analysis, for all age comparisons, and are presented in Table 4. Between reader precision was observed to be quite poor using these two indices of precision. The mean percent agreement between age-readers was calculated at 63.2%. Analysis of CV indicated a similar pattern of imprecision with mean CV computed at 7.6% between readers. Lastly, plots of mean CV and percentage agreement have been formulated to graphically identify imprecision and bias at fish age, in comparisons between age-readers (Fig. 7). Results graphically highlight both imprecision and bias in comparative age determinations between age readers A and C and B and C.

### 3.4. Within reader precision

Within-reader imprecision was not identified as significant by the nonparametric Wilcoxon Two Sample test method (Table 3). Index of precision analysis, namely percentage agreement and CV, identified higher levels of precision within age-readers than between age-readers (Table 4). Mean percentage agreement was calculated at 79.7% for within age-readers, compared to 63.2% for between age-readers. Similarly, the average CV for within age-readers (2.9%) was found to be much lower than that for between age-readers (7.6%).

## 4. Discussion

Repeated age determinations of a sample of fish are generally conducted for one of two reasons, to determine if there are systematic differences in age estimates between one or more, age-readers, methodologies or laboratories; or to estimate the precision 'reproducibility' of age estimates (Campana et al., 1995). The purpose of the present study was to investigate both parameters for the age determination of blue whiting.

Table 3
Results of statistical analysis for inter-reader and intra-reader precision and bias

| Statistic | Inter-reader comparisons | | |
|---|---|---|---|
| | Reader A vs. reader B ($N = 251$) | Reader A vs. reader C ($N = 350$) | Reader B vs. reader C ($N = 299$) |
| Regression | | | |
| Slope | $0.967 \pm 0.03$ | $0.80 \pm 0.02$ | $0.731 \pm 0.02$ |
| *P* value | 0.046 | 0.000 | 0.000 |
| Intercept | $0.0838 \pm 0.087$ | $0.338 \pm 0.08$ | $0.659 \pm 0.1$ |
| *P* value | 0.000 | 0.000 | 0.000 |
| 1 Sample Sign Test | | | |
| Positive ranks | 32 | 11 | 11 |
| Negative ranks | 50 | 128 | 102 |
| Ties | 169 | 211 | 186 |
| *P* value | 0.0605 | 0.001 | 0.001 |
| Wilcoxon Two Sample Test | | | |
| *P* value | 0.75 | 0.001 | 0.002 |
| | Intra-reader comparisons | | |
| | Reader A vs. reader A ($N = 299$) | Reader B vs. reader B ($N = 176$) | Reader C vs. reader C ($N = 300$) |
| Regression | | | |
| Slope | $0.83 \pm 0.02$ | $1.011 \pm 0.02$ | $1.009 \pm 0.01$ |
| *P* value | 0.000 | 0.62 | 0.46 |
| Intercept | $0.5714 \pm 0.10$ | $0.038 \pm 0.10$ | $0.04 \pm 0.05$ |
| *P* value | 0.000 | 0.011 | 0.000 |
| 1 Sample Sign Test | | | |
| Positive ranks | 27 | 17 | 5 |
| Negative ranks | 45 | 6 | 25 |
| Ties | 227 | 153 | 270 |
| *P* value | 0.0451 | 0.0347 | 0.001 |
| Wilcoxon Two Sample Test | | | |
| *P* value | 0.71 | 0.72 | 0.57 |

The slope and intercept of simple linear regression are tested for significant differences ($\alpha = 0.05$) from 1 and 0, respectively. The nonparametric '1 Sample Sign' and 'Wilcoxon Two Sample Test' of the median are used as tests for both bias and precision, respectively. Error terms are 95% confidence limits.

### 4.1. Between reader bias and precision

Significant inter-reader bias, in the age estimates of one of the three age-readers was identified, with significant inconsistency apparent in comparisons between all three age-readers. Regression analysis proved sensitive to systematic differences between age-readers but results of the nonparametric One Sample Sign test provided a more descriptive scrutiny for potential bias.

Age bias plots best described the nature of bias between age-readers (Fig. 6), describing almost linear comparative under-ageing by age-reader C for older fish. However, this does not rule out linear over-ageing by the remaining two age-readers examined in this study. Furthermore, the observed between reader bias was most obvious for fish of age four and older. This resulted in the estimated average age of fish, from important year classes, differing by over 1 year between age-readers in some cases. Since blue whiting begin to mature

Table 4
Coefficient of variation (Chang, 1982) and percent agreement are presented for both between and within-reader age comparisons

| | Reader A vs. B ($N = 251$) | Reader A vs. C ($N = 350$) | Reader B vs. C ($N = 299$) | Mean |
|---|---|---|---|---|
| Between reader comparisons | | | | |
| %Agreement | 66.93 | 60.45 | 62.20 | 63.20 |
| Mean CV (%) | 5.58 | 8.92 | 8.35 | 7.61 |
| | Reader A vs. A ($N = 299$) | Reader B vs. B ($N = 176$) | Reader C vs. C ($N = 300$) | Mean |
| Within reader comparisons | | | | |
| %Agreement | 62.27 | 86.93 | 90.00 | 79.73 |
| Mean CV (%) | 4.76 | 2.21 | 1.80 | 2.92 |

Mean %agreement and CV are also calculated across both sets of comparisons (mean CV% estimates are calculated by averageing individual CV across fish within a sample).

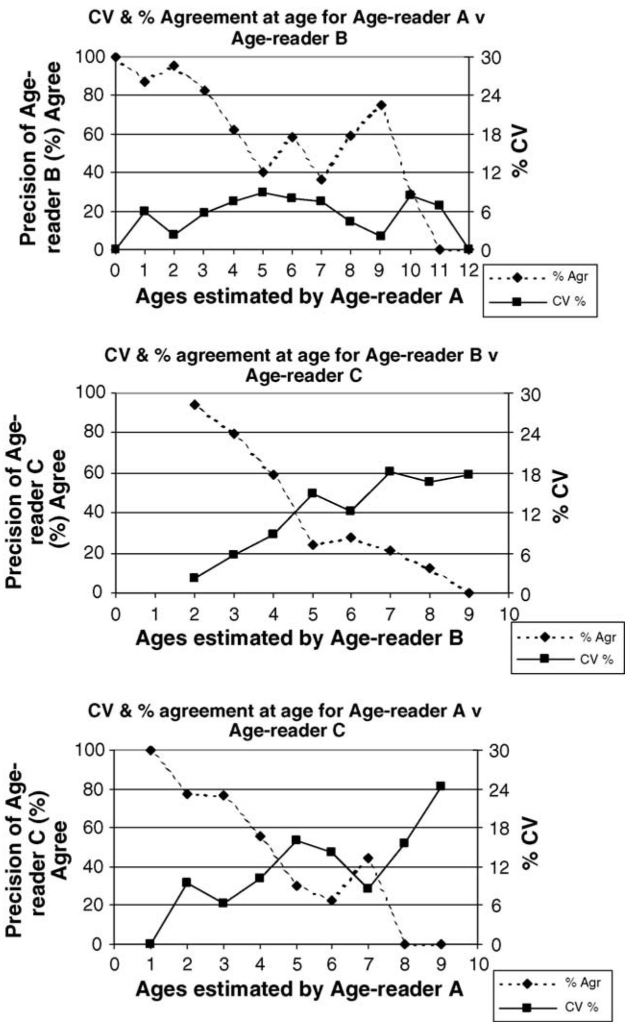


Fig. 7. Plots of percentage agreement and CV at age, for the three pair-wise age comparisons presented in Table 2. (Top): bias absent between age-readers; (middle and bottom): bias present between age-readers.

at age two and are mostly fully recruited to the spawning stock at age four (Monstad, 2004), the observed pattern of under-ageing by age-reader C may signify a bias due to misinterpretation of winter checks as spawning checks at and above the age at maturation. This phenomenon could have influenced the results of this study. Furthermore 'split rings' are often encountered in the structure of blue whiting otoliths. Split rings appear as the break-up of the winter hyaline zone of the otolith into two weak hyaline zones with an opaque zone in-between (Ø. Tangen, Personal Communication, IMR, Norway.). This anomaly can be confusing to some age-readers resulting in possible double counting of winter rings. It is most probable that this 'split ring' phenomena had an effect on the results of the present study.

Between reader precision analysis in this study was complicated by the systematic under-ageing or relative negative bias detected in one of the age-readers tested. The nonparametric Wilcoxon Two Sample test can be used as a test of precision for paired data as it includes tied values in its ranked analysis. This test identified significant imprecision in comparisons with age-reader C, which can be attributed to the bias identified through earlier analysis. Comparisons between age-readers A and B were found to be precise using this test. Analysis of the index of CV for between-reader precision (CV 5.5–8.3%) (Table 4), compared favourably with those reported in the literature for species other than blue whiting (CV 5%, Kimura and Lyons, 1991; 6–22%, Campana et al., 1995; 7.6–8.7%, Stransky et al., 2005; 5–20%, Fossen et al., 2003). However, because CV is relative to the number of age classes examined for a particular species, the results of this study may indicate that for a short lived species such as blue whiting, observed CV values of 5.5–8.3% may reflect a degree of imprecision in comparison to investigations on longer lived species.

Percentage agreement between age-readers was found to be quite poor in the present study (60–67%) given the relatively few year classes present in experimental samples. Beamish and Fournier (1981) illustrated how if 95% of age determinations between two age-readers agree within $\pm 1$ year, this can be very poor precision since most commercial samples of short-lived species are comprised of few year classes. In a study on precision for Alaskan Pollack, a species with similar stock structure to blue whiting, Kimura and Lyons (1991) reported a mean percentage agreement of 63.8% averaged across ages. These values also compared favourably with the results of this study, however, Kimura and Lyons (1991) did highlight the fact that percentage agreement can be misleading in adequately describing the extent of imprecision and that CV is much more descriptive. This was found to be the case in the present study where percentage agreement best described the prevalence of differences in paired age estimates, while the coefficient of variation best described the intensity of these observed differences.

### 4.2. Within reader bias and precision

Despite high levels of ageing experience, significant bias was identified for all three age-readers tested in the present study. The systematic under-ageing by two age-readers, identified by the One Sample Sign test, may be explained by a 'conservative' ageing approach, where age-readers may have been reluctant to risk applying older ages to fish under experimental scrutiny. Imprecision within age readers was not found to be significant in the present study through nonparametric analysis. Precision was observed to be higher within age-readers than between age-readers, when analysing CV and percentage agreement, even in the presence of identifiable bias. Age-reader A demonstrated the lowest percent agreement and highest CV, 62.27 and 4.76%, respectively, and was the less experienced of the three age readers. Age-reader C demonstrated the highest percent agreement and lowest CV, 90 and 1.8%, respectively, and was the most experienced of the age-readers tested. This observed experience gradient has been well documented in the literature, one report stating that, "new agers start at a high CV and work their way down" (Walsh and Burnett, 2002). Walsh and Burnett (2002) also stated that "within reader precision will always be better than between reader precision", findings comparable with the present study.

### 4.3. Concluding remarks

In general, precision between blue whiting age-readers tested in this study was found to be poor, with within reader precision substantially better. This suggests a systematic interpretation problem rather than a difficulty in seeing structures on the ototlith.

Comparative analysis uncovered systematic under-ageing by one age-reader. On an international scale the miss ageing of abundant age classes in short lived species can pose a serious problem to stock assessment (Walsh and Burnett, 2002). This is especially the case in blue whiting where such errors can affect recruitment estimates, which are critical for management of the stock (Eltink et al., 2000; Tyler et al., 1989; Le Cren, 1974).

The results of this study were used to inform a recent ICES international age-reading workshop for blue whiting. This workshop addressed these issues, by standardising age-reading methodologies and investigating variability in age-determination between fisheries laboratories on a larger scale. During this workshop, the presence of 'split rings' in the blue whiting otolith structure was identified as a major contributory factor to ageing error (L.W. Clausen, Personal Communication, Workshop Chairperson, DIFRES, Denmark). Although a consensus was eventually reached during the workshop, on an agreed international protocol for the age reading of this species, the development of 'split rings' still requires further investigation. Furthermore, this study has identified age-reader experience as an important factor when encountering problematic otoliths.

At present, procedures for the interpretation of 'split rings' between international age reading laboratories remain quite heterogeneous. Some laboratories follow detailed protocols to aid identification of such structures, while others rely on individual age-reader discretion and/or experience. In addition, geographical and physiological conditions may affect the formation of such 'split rings' (N. Timoshenko, Personal Communication, Atlantniro, Russia.). The authors recommend therefore that future research should be focused at understanding the mechanism of development of 'split rings' in blue whiting otoliths, the results of which should ultimately be incorporated into agreed age reading protocols. It is also recommended that, in addition to using the new protocols, laboratories responsible for ageing this species should use experienced age-readers to train new technicians.

## Acknowledgements

The authors gratefully acknowledge the financial and structural support of the Marine Institute and the Marine RTDI Measure, National Development Plan 2002–2006. Many thanks go to Dr. Maurice Clarke, Marine Institute, Ireland, for constructive advice, and to Ms. Kirsti B. Eriksen of IMR Norway for preparation of ageing material. We would also like to thank Dr. Deirdre Brophy, GMIT, Ireland, for her help and advice with data analysis. This project (Grant-aid Agreement No. PhD/01/004) was carried out with the support of the Marine Institute and the Marine RTDI Measure, Productive Sector Operational Programme, National Development Plan 2000–2006.

NDP NATIONAL DEVELOPMENT PLAN    Marine Institute Foras na Mara

YOUR PLAN – YOUR FUTURE

## References

Baker, T.T., Timmons, L.S., 1991. Precision of ages estimated from five bony structures of Artic char (*Salvelinus alpinus*) from the Wood River System Alaska. Can. J. Fish. Aquat. Sci. 48, 1007–1014.

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

Beamish, R.J., McFarlane, G.A., 1983. The forgotten requirement for age validation in fisheries biology. Trans. Am. Fish. Soc. 112, 735–743.

Bennett, J.T., Boehlert, G.W., Turekian, K.K., 1982. Confirmation of longevity in *Sebastes diploproa* (Pisces: Scorpaenidae) from $^{210}Pb/^{226}Ra$ measurements in otoliths. Mar. Biol. 71, 209–215.

Bergstad, O.A., Magnusson, J.V., Magnusson, J., Hareide, N.R., Reinert, J., 1998. Intercalibration of age readings of ling (*Molva molva* L.) blue ling (*Molva dipterygia* Pennant, 1784) and tusk (*Bromse bromse* L.)—ICES. J. Mar. Sci. 55, 309–318.

Campana, S.E., 2001. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. J. Fish Biol. 59, 197–242.

Campana, S.E., Annand, M.C., McMillan, J.I., 1995. Graphical and statistical methods for determining the consistency of age determinations. Trans Am. Fish. Soc. 124, 131–138.

Campana, S.E., Moksness, E., 1991. Accuracy and precision of age and hatch date estimates from otolith microstructure examination. ICES. J. Mar. Sci. 48, 303–316.

Chang, W.Y.B., 1982. A statistical method for evaluating the reproducibility of age determination. Can. J. Fish. Aquat. Sci. 39, 1208–1210.

Conover, W.J., 1980. Practical nonparametric statistics. John Wiley & Sons Inc., Toronto, 455 pp.

Fossen, I., Albert, O.T., Nilssen, E.M., 2003. Improving the precision of ageing assessments for long rough dab by using digitised pictures and otolith measurements. Fish. Res. 60, 53–64.

Heifetz, J., Anderi, D., Maloney, N.E., Rutecki, T.L., 1998. Age validation and analysis of ageing error from marked and recaptured sablefish, *Anoplopoma fimbria*. NOAA. Fish. Bull. 97, 256–263.

Hoenig, J.M., Morgan, M.J., Brown, C.A., 1995. Analysing differences between two age determination methods by tests of symmetry. Can. J. Fish. Aquat. Sci. 52, 364–368.

ICES, 1994. Report of the Workshop on Sampling Strategies for Age and Maturity. ICES CM 1994/D:1, 67 pp.

ICES, 2004. Report of the Northern Pelagic and Blue Whiting Fisheries Working Group. Copenhagen, Denmark, April 27–May 4, 2004. ICES CM 2004/ACFM:24, 294 pp.

ICES, 2005. Report of the Northern Pelagic and Blue Whiting Fisheries Working Group. Copenhagen, Denmark, August 25–September 1, 2005. ICES CM 2006/ACFM:05, 290 pp.

Jenson, A.C., 1965. A standard terminology and notation for otolith readers. Int. Comm. Northwest Atlantic Fish. Res. Bull. 2, 5–7.

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

Kimura, D.K., Mandapat, R.R., Oxford, S.L., 1979. Method, validity, and variability in the age determination of yellowtail rockfish (*Sebastes flavidus*), using otoliths. J. Fish. Res. Board Can. 36, 377–383.

Le Cren, E.D., 1974. The effects of errors in ageing in production studies. In: Bagenal, T.B. (Ed.), The Ageing of Fish. Unwin Bros, pp. 221–224.

Monstad, T., 2004. Blue Whiting. In: Hein Rune Skjoldal (Ed.), The Norwegian Sea Ecosystem. Tapir Academic Press, Trondheim, pp. 263–288.

Peltonen, H., Raitaniemi, J., Parmanne, R., Eklund, J., Nyberg, K., Halling, F., 2002. Age determination of the baltic herring from whole otoliths and from neutral red stained otolith cross sections. ICES. J. Mar. Sci. 59, 323–332.

Risso, A., 1810. Ichthyologie de Nice, ou histoire naturelle des poissons dudepartment des Alpes Maritimes. Paris, 388 pp.

Sharp, D., Bernard, D.R., 1988. Precision of estimated ages of lake trout from five structures. N. Am. J. Fish. Manage. 8, 367–372.

Stransky, C., Guðmundsdóttir, S., Sigurosson, T., Lemvig, S., Nedreaas, K., Saborido-Rey, F., 2005. Age determination and growth of Atlantic redfish (*Sebastes marinus* and *S. mentella*): bias and precision of age readers and otolith preparation methods. ICES. J. Mar. Sci. 62, 655–670.

Tyler, A.V., Beamish, R.J., McFarlane, G.A., 1989. Implications of age determination errors to yield estimates, in: Beamish, R.J., McFarlane, G.A. (Eds.), Effects of Ocean Variability on Recruitment and an Evaluation of Parameters used in Stock Assessment Models. Can. Spec. Publ. Fish. Aquat. Sci. 108, 27–35.

Walsh, S.J., Burnett, J., 2002. The Canada–United States Yellowtail Flounder Age Reading Workshop, St. Johns, Newfoundland, November 28–30, 2000. NAFO Scientific Council Studies 35: 1–59.

Zilanov, V.K., 1968. The Biology and Exploitation of the Blue Whiting. Translation no. 2118. Murmansk Book Publishers, Marine Laboratory, Aberdeen, August 1980, 76 pp.

## *Electronic citation*

Eltink, A.T.G.W., Newton, A.W., Morgado, C., Santamaria, M.T.G., Modin, J., 2000. Guidelines and Tools for Age Reading (PDF document Version, October 2000) Internet: http://www.efan.no.