ICES J. mar. Sci., 48: 303–316. 1991

Accuracy and precision of age and hatch date estimates from otolith microstructure examination

Steven E. Campana and Erlend Moksness

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

Known-age herring and cod larvae, reared in an outdoor mesocosm and the laboratory, respectively, were distributed to 18 study participants representing 10 different countries as part of an international otolith microstructure study. All details of rearing and sampling protocol were withheld from study participants, who were asked to estimate the age of multiple samples (aged 1–65 days) based only on the otolith microstructure. On average, age differences among samples were accurately estimated, while absolute ages (hatch dates) were underestimated. Otoliths from herring larvae less than 14 days old were the most difficult to interpret. Estimates of hatch date and measurements of the hatch check diameter differed significantly among readers, implying that differences in skill and/or increment interpretation among researchers could result in misleading differences among both collaborators and published results. The experience of the investigator and the degree of preparation of the larger otoliths were among the sources of the inter-investigator differences in ageing accuracy. With the possible exception of SEM, equipment type was not correlated with either accuracy or precision. The coefficient of variation (C.V.) of the age estimates decreased with larval age to a level of about 10–15%, and did not differ significantly among age readers. This study confirmed the utility of otolith microstructure examination for high-resolution age determination of young fish under quasi-natural conditions. Although the accuracy and precision levels characteristic of the herring otolith microstructure are probably somewhat lower than those of most other species, resolution of ± 1 day may be beyond the capabilities of the technique for a number of species.

Steven E. Campana: Marine Fish Division, Bedford Institute of Oceanography, P.O. Box 1006, Dartmouth, Nova Scotia, Canada B2Y 4A2. Erlend Moksness: Institute of Marine Research, Flødevigen Marine Research Station, N–4817 His, Norway.

Introduction

How accurate are the estimates of age of field-collected fish based on otolith microstructure? Perhaps more importantly, how accurate are the estimates of hatch date, growth and mortality rate which are derived from such data? In light of the growing popularity of otolith microstructure examination for ageing wild, young-of-the-year fish (for reviews, see Campana and Neilson, 1985; Jones, 1986), the ICES Recruitment Processes Working Group recommended that the above questions be addressed in an international otolith microstructure study (Anon., 1989). The study involved the distribution of large numbers of known-age, herring (*Clupea harengus*) larvae, reared in a mesocosm under natural conditions, to 18 study participants representing 10 different countries. Smaller numbers of known-age, laboratory-reared cod (*Gadus morhua*) larvae were also distributed. In the absence of any information on age, rearing conditions, or sampling protocol, study participants were asked to estimate larval age through examination of the otolith microstructure. Through distribution of a broad range of ages, the intent was to induce the same level of uncertainty concerning the sample ages and hatch dates as might have been felt if examining field samples of unknown age.

Age determinations of wild, young-of-the-year fish have rarely, if ever, been reported with an associated estimate of accuracy. Estimates of precision are much more readily available. The distinction between accuracy and precision is important, since accurate estimates need not be precise, and vice versa. Accuracy refers to the proximity of the estimate to the "true" value, while precision refers to the reproducibility of the individual measurements on a given structure (Wilson, 1987). Thus, a mean age can be accurate (close to the truth) while the individual observations are imprecise (vary widely). Conversely, and this is often the case in ageing studies, age estimates can be precise (highly reproducible, either within or among readers) but not necessarily accurate. The precision index used in this paper differs somewhat from that defined above, in that it measures reproducibility of age estimation for fish of a given age rather than reproducibility within a given otolith. Assuming that the rate of increment formation does not vary among fish, such a precision index should be more useful for assessment of ageing error than that based on the formal definition,

1054–3139/91/030303 + 14 $03.00/0

since it accounts for variability in both otolith preparation and increment counting/interpretation. The looser definition of precision will be used throughout the remainder of this paper. Note, however, that none of the precision indices can be used as proxies for tests of accuracy; the latter require an independent and absolute means of age determination. For instance, accuracy has not been demonstrated if age estimates from sagittal and lapillar otoliths concur.

Since the determination of ageing accuracy mandates the use of known-age fish, experiments designed to validate the frequency of daily increment formation would appear to be ideal vehicles for assessing accuracy. Yet, for several reasons, the type of validation experiment most often reported in the literature is not appropriate. First of all, validation experiments often serve as a training set for refining interpretational skills during the initial stages of a study; it would be unusual if a set of known-age samples were to be examined only after completion of the examination of the field samples. Second, most increment validation experiments are constrained by logistical realities, and thus are conducted in the laboratory or under otherwise unnatural conditions. Increment appearance differs markedly between laboratory and natural conditions, and between environments promoting different growth rates (Bergstad, 1984; Campana and Neilson, 1985). Thus, the interpretation of daily increments in laboratory-reared fish may differ substantially from that of wild fish. Third, validation experiments seldom span the entire age range characteristic of the field samples (Beamish and McFarlane, 1983). Indeed, the rate of increment formation through the first 1–2 weeks of larval life is often taken to represent that of the entire larval and juvenile stage. And, finally, a validation experiment is almost invariably designed and conducted under the auspices of the same investigator who conducts the field study. Despite any blind labelling and coding of the prepared samples, the age reader is almost always aware of the experiment design, sampling frequency and age range, as well as the general growth conditions of the rearing enclosure. While the impact of this knowledge is difficult to quantify, its influence on subconscious age expectations might well be substantial. Taken together, the above suggests that estimates of age determination accuracy derived from increment validation experiments are probably optimistic.

At present, there are no published comparisons of either accuracy or precision among independent investigators using otolith microstructure examination to age fish reared outside the laboratory, and very few reliable measures of age determination accuracy by individual investigators (exceptions include the studies of Tsukamoto and Kajihara (1987) and Secor and Dean (1989)). Such comparisons and measures are mandatory if the reliability of published age, growth, and mortality estimates is to be assessed. The study described in this paper was designed to compare measures of accuracy and precision, both within and among investigators, with a view towards answering the following questions:

1. Is otolith microstructure examination sufficiently accurate and precise to determine the age, hatch date, growth rate, and mortality rate of young fish under field conditions?
2. Do accuracy and precision vary sufficiently among investigators that comparison among individual results is difficult? Are collaborative studies at risk?
3. Are enhanced levels of accuracy and precision associated with any particular suite of techniques and skills? And, if so, would these same skills benefit others in this field?

Materials and methods

Herring

Herring eggs were artificially fertilized at sea on 28 February 1990. They were transported to the Flødevigen Marine Research Station, Norway on 1 March and incubated in circulating sea water at temperatures between 6.8 and 7.3°C and salinities between 33.6 and 34.5‰. Peak hatching occurred in the period 20–25 March. Newly hatched herring larvae from the mornings of 20–22 March were collected and transferred to 40-l cylinders for a period of 1 day. On 23 March, approximately 52 000 of these larvae were released into a large outdoor mesocosm (2000 m^3 volume, 600 m^2 surface area, 4 m maximum depth; see also Moksness, 1982), where they were reared for 65 days using a resident zooplankton community for food. Yolk sacs were completely resorbed between the ages of 7 and 15 days. The temperature in the enclosure increased from 6.8 to 15.0°C through the course of the experiment, while salinity and oxygen saturation remained roughly stable at 33‰ and 120%, respectively. Larvae were sampled at daily intervals (Wespestad and Moksness, 1990) and preserved in 95% ethanol prior to distribution to study participants. All participants received a minimum of four samples, each consisting of about 20 larvae of the same age. Samples were randomly allocated among participants, and none contained any information associated with date or age of collection. Since the technicians in the laboratory of one of the authors (Moksness) were familiar with the sampling protocol, data from that laboratory were not included in this analysis.

Cod

Cod were reared in two cohorts in the laboratory at the St Andrews Biological Station, Canada, as part of another study. Eggs were derived from naturally spawning broodstock reared at the station. Larvae were transferred to the

Table 1. Equipment and methods applied by each investigator during the examination of herring and cod otoliths. The equipment is that used for increment counts, not that used for measurement. Equipment codes are as follows: Micro = microscope, Video = video or image analysis system, Photo = photograph, SEM = scanning electron microscopy. Preparation refers to the presence (Y) or absence (N) of otolith polishing. The experience category was self-coded.

Invest.	Herring ages (*n*)						Number of cod	Preparation	Equipment	Magnif.	Experience
1	9 (3)	15 (5)	40 (5)				5	Y	Micro	1000	High
2	6 (10)	18 (8)	37 (10)	54 (4)			5	N	Micro	1000	High
3	27 (3)	35 (4)	47 (4)				8	Y	Micro	1250	High
4	12 (5)	21 (5)	32 (5)	43 (5)			6	N	Video	800	High
5	9 (5)	18 (5)					0	N	Micro	300	Low
6	9 (5)	18 (5)	33 (5)				0	N	Photo	1000	Low
7	15 (1)	24 (4)	57 (9)				0	Y	SEM	4000	High
8	12 (10)	21 (10)	32 (10)	37 (10)			6	N	Photo	1250	Medium
9	12 (9)	21 (13)	32 (16)	54 (16)			0	N	Video	1000	Low
10	3 (13)	12 (11)	18 (8)	27 (18)	37 (19)	47 (11)	5	Y	Micro	1500	Low
11	5 (10)	12 (10)	21 (10)	29 (10)			1	Y	Micro	1575	Medium
12	29 (7)	43 (5)					6	Y	Micro	1000	High
13							6	Y	Video		High
14	6 (12)	15 (5)	29 (11)	47 (12)			6	Y	Micro	1250	Medium
15	21 (15)	35 (10)	47 (9)				5	N	Micro	1250	Medium
16	7 (5)	16 (5)	29 (20)	35 (35)	55 (8)		0	N	Video	1000	Medium
17	6 (5)	18 (7)	27 (6)	35 (6)			0	Y	Video	1000	High
18	6 (8)	18 (8)	30 (8)	35 (7)			0	N	Micro	1575	High

rearing tanks on the day of hatch, where they were reared under a 12:12 photoperiod at a salinity of 31‰. Cohort no. 1 was stocked at a density of 12 larvae $\times l^{-1}$ in 40-l flow-through tanks and reared at 7–9°C. Cohort no. 2 was stocked at a density of 17 larvae $\times l^{-1}$ in 30-l static-flow tanks and reared at 6°C. While growth rates were not monitored in either of the two cohorts, the higher temperatures to which Cohort no. 1 was exposed probably resulted in more rapid growth than in Cohort no. 2. Both cohorts were fed rotifers once daily, supplemented by *Artemia* after 7 days. Larvae in Cohort no. 1 were sampled after 12 days, videotaped alive for subsequent length measurements, decapitated, and the heads/otoliths preserved in 95% ethanol. Larvae in Cohort no. 2 were sampled after 22 days and then videotaped and preserved as described above. Samples of both cohorts were distributed without further preparation to all study participants.

Otolith preparation

With the exception of the samples described below, all larvae were distributed intact to study participants. No guidelines were offered concerning mounting, sample preparation, means of examination or even the otolith pair (e.g. sagittae or lapilli) to be used; complete flexibility was given to all participants. The range in equipment, means of examination, magnification, presence/absence of otolith preparation, and self-described experience with otolith microstructure preparations are presented in Table 1.

To determine the influence of standardized sample preparation on the accuracy and precision of increment counts, some samples of both herring and cod were mounted and polished prior to distribution to study participants. Sagittae from a 36-day-old herring sample and lapilli and/or sagittae from a 22-day-old cod sample were mounted individually in cyanoacrylate glue on standard microscope slides. Each otolith was then polished with 3 μm metallurgical lapping film to a plane just above that of the nucleus. Two slides of each species were then mailed to each participant.

Results

Herring reared in the mesocosm grew at an average rate of 0.27 mm d^{-1} (Fig. 1), which is comparable to that expected of wild larvae (Lough *et al.*, 1982). The microstructure of the sagittae appeared to be indistinguishable from that of wild larvae, in terms of both increment width and increment appearance. Thus, we felt that the otoliths of the mesocosm-reared herring were a realistic proxy for those of wild fish.

Mesocosm-reared cod larvae were not available for this experiment. In keeping with the findings of other studies (Bergstad, 1984; Campana, 1989), the microstructure of the lab-reared cod otoliths was less clear than that of wild larvae. However, on the basis of the increment widths, the growth rate of Cohort no. 1 appeared to be comparable to that observed in wild cod (Campana and Hurley, 1989).

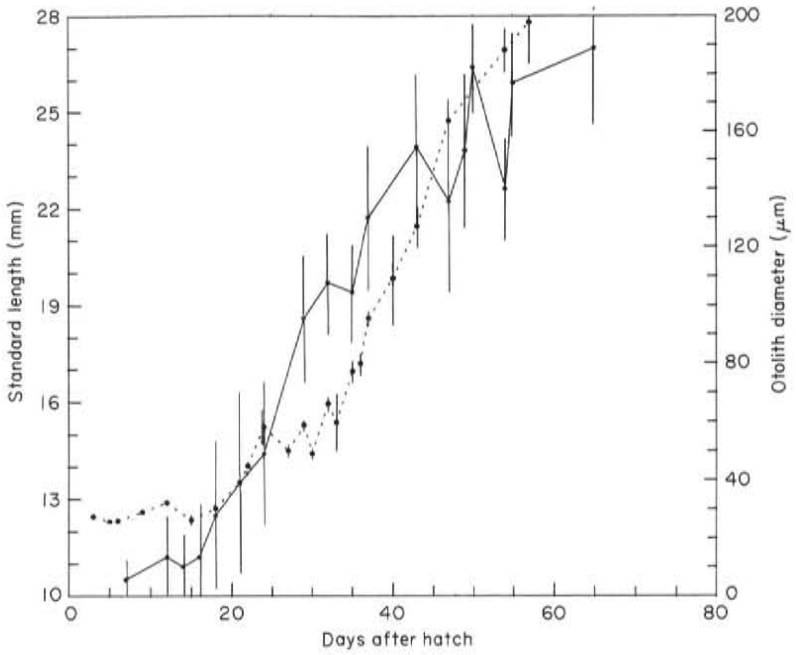

Figure 1. Growth of herring larvae reared in a mesocosm for 65 days. Standard lengths (SL) are presented in terms of mean ± 1 s.d. Error bars around the otolith diameter points are 1 s.e. —●— = SL, --●-- = Otolith diameter.

A total of 544 herring larvae and 59 cod larvae were aged by 18 different investigators as part of this study (Table 1). Sagittal otoliths were used almost exclusively in the herring examinations, while both the lapilli and sagittae were used in the case of cod. One half of the investigators polished the herring otoliths prior to examination. Age determinations were made most frequently directly through the microscope, although video and image analysis systems were also commonly used (Table 1). The latter two systems were applied more often for measurement purposes. Scanning electron microscopy (SEM) and photomicrography were used relatively infrequently. Most investigators employed magnifications of 1000–1250 ×. The range of experience levels among the investigators varied from the novice level to those with long-term experience (Table 1).

Accuracy of the age estimates

Herring

In the majority of the herring otoliths, daily increment counts underestimated the number of days elapsed since hatch. On average, there were 9.0 fewer increments visible in the otolith than would be expected of daily increment formation from the time of hatch. However, there was a broad distribution around the mean value, with the discrepancy between age and increment count ranging between −9 and +28 (Fig. 2). At least part of this variability was due to the effect of larval age; both the mean and the variance of the discrepancy increased with larval age up until an age of about 14 days (Fig. 3). At ages above 14 days, the mean discrepancy between age and increment count appeared to stabilize at about 10.1. However, even among older larvae, it was clear that individual hatch date back-calculations could vary widely.

A significant portion of the variability evident in Figures 2 and 3 was associated with differences among individual investigators. A box and whisker plot of the age-increment discrepancies (restricted to larvae greater than 14 days old in order to eliminate the age effect described above) indicated that both the mean and variance varied widely among individual investigators (Fig. 4). Analysis of variance indicated that the among-investigator differences were significant (Kruskal-Wallis, $p < 0.01$). Thus, at least some of the participants in the study used different criteria and/or methods to age their otoliths than did the others.

The results presented above indicate that herring larva age estimates were not always comparable among investigators. Such a finding could complicate interpretations of published age and hatch date distributions. However, as long as a given investigator's methods (bias) were consistent across all samples, the estimates of elapsed time between samples would not necessarily be biased. In other words, even if a given investigator underestimated true age by 100 days in all larvae, rate calculations (such as

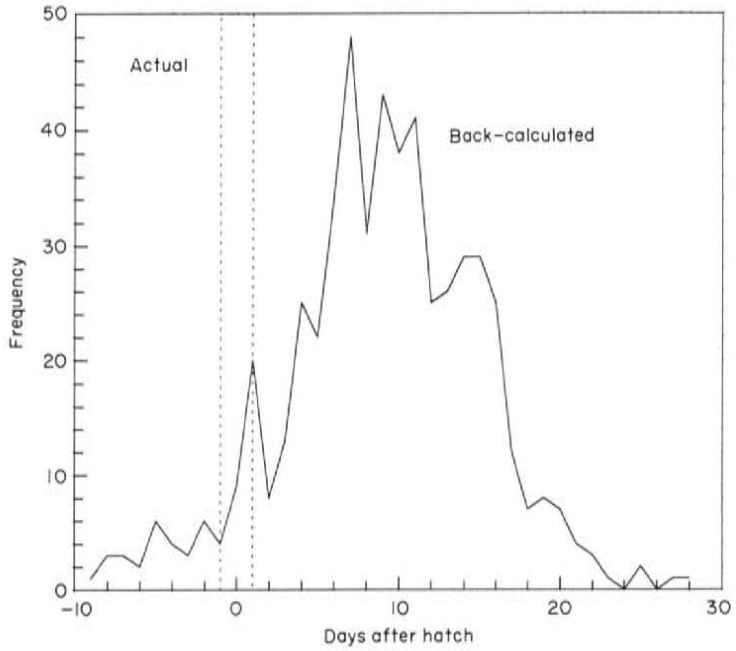

Figure 2. Frequency histogram of hatch dates back-calculated from 544 otoliths of known-age herring larvae. Herring were pooled across all investigators and samples for this figure. The actual hatching period extended over a 3-day period.

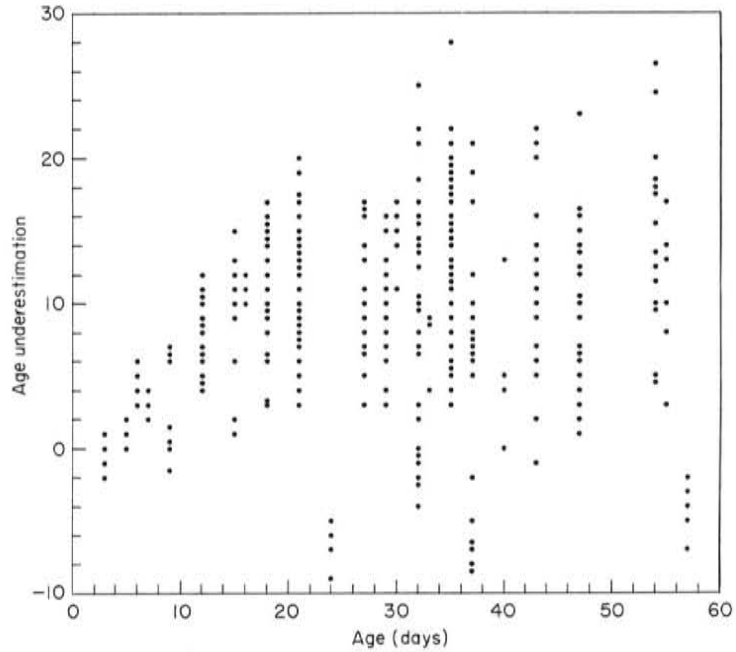

Figure 3. Age underestimation (age minus otolith increment count) as a function of age in larval herring, pooled across all investigators and samples. Both the mean and variance of the age-increment discrepancy appear to increase until at least age 14.

growth and mortality) based on elapsed time between samples could still be accurate. To test this hypothesis, apparent increment formation rates were calculated for each investigator based on the regression of daily

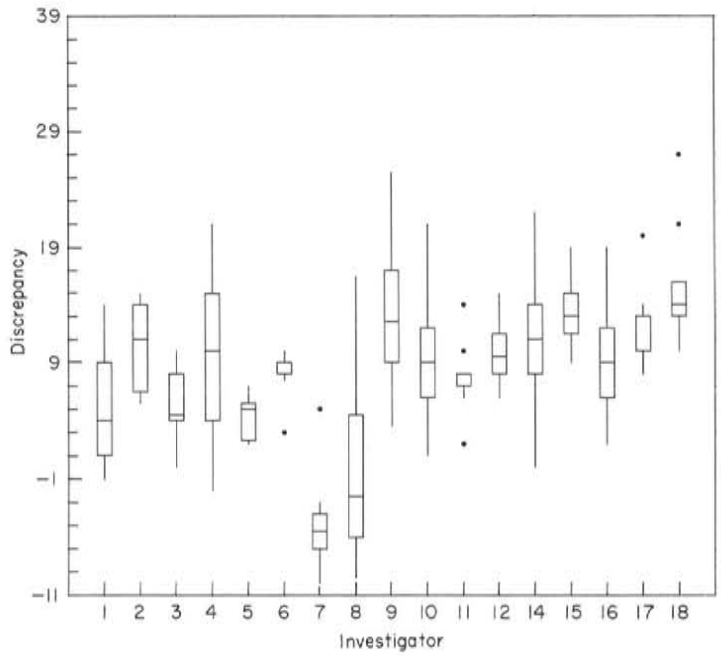

Figure 4. Box and whisker plot of the differences in the herring age-increment count discrepancy (age underestimation) among investigators. Each box extends over the central 50% of the data, with the horizontal line corresponding to the median. The whiskers extend to the outer limits of the data, except for widely divergent points which are marked individually. In order to eliminate the age effect evident in Figure 3, the analysis was restricted to larvae >14 days old. The among-investigator differences were highly significant.

increment count on true age for all larvae examined by each investigator.[1] The mean apparent increment formation rate across all investigators was 0.95 (s.e. = 0.08), which was not significantly different from the value of 1.0 which would be expected of accurate age determination. The majority of the study participants were reasonably accurate in their estimation of the average age difference among samples (Fig. 5). Investigators with less than 25 larvae in total and/or larvae less than 20 days old were less accurate (and less precise) in their estimates of age differences among samples. The relative accuracy of the age differential estimate was particularly sensitive to the range of ages present in the samples; age differences of less than 20 days tended to be inaccurately estimated (Fig. 6). However, the absolute counting errors were relatively stable across the sample age range; most of the apparent increment formation rates corresponded to estimated age differentials which differed by less than 5 days from the differential between the minimum and maximum sample ages.

Cod

The accuracy of cod larva age estimation was negatively correlated with the relative growth rate of the cohort, probably through the influence of growth on the formation of narrow, unresolvable increments (Campana *et al.*, 1987; Jones and Brothers, 1987). Mean increment counts of the faster-growing, 12-day-old cohort (Cohort no. 1) were similar to, although significantly less than, the true age (Table 2). On the other hand, the ages of the slower-growing, 22-day-old larvae (Cohort no. 2) were significantly and substantially underestimated by almost everyone (Table 2). Sample sizes were too low to test for differences among investigators.

[1]While we wished to predict age in this regression, age was the fixed variable in the study design. Therefore, the regression of age on increment count would not have been appropriate.

Precision of the herring age estimates

Ageing precision within each sample of each investigator was assessed through two calculations: with the standard deviation and with the coefficient of variation (C.V.) (defined as the standard deviation divided by the mean) (Chang, 1982). As mentioned earlier, such a measure of precision includes both reader effects and any intrinsic variability which may exist between otoliths of different fish of the same age. The within-sample standard deviation increased with the age of the sample, suggesting that the counting error was cumulative (Fig. 7). The presence of significant differences in standard deviation among investigators (ANOVA of larvae >14 days old, $p < 0.05$) did not appear to be associated with the age effect evident in Figure 7. On the other hand, relative precision (as

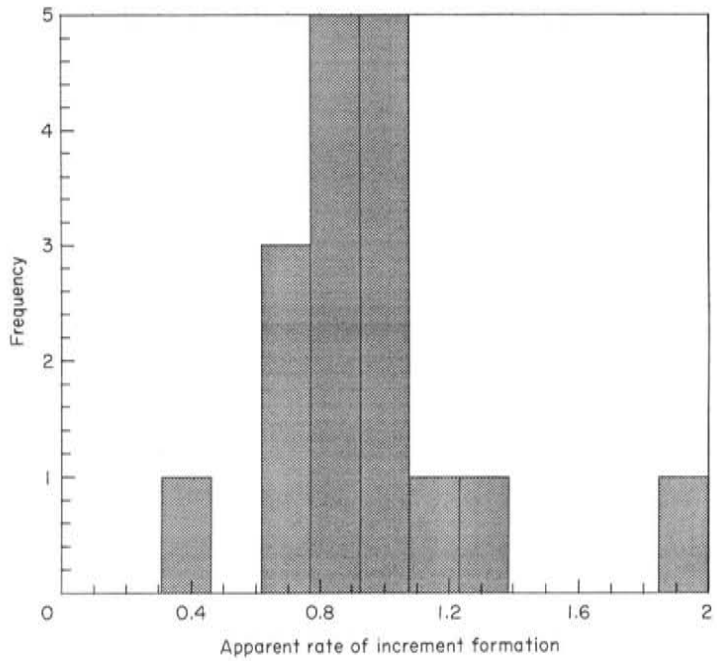

Figure 5. Frequency histogram of the apparent increment formation rate in herring for each investigator. The apparent rate was calculated as the regression slope of increment count on larval age for all larvae examined by a given investigator. An apparent rate of 1.0 would indicate that the investigator had, on average, accurately estimated the age differences among the samples.

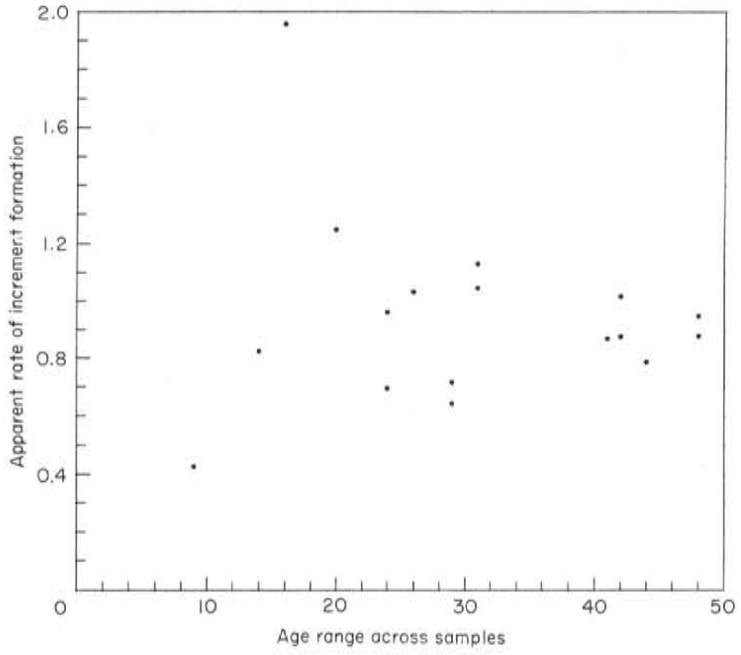

Figure 6. Estimation accuracy of the apparent rate of daily increment formation in herring as a function of the range in age across the samples. An apparent rate of 1.0 would indicate that the investigator had, on average, accurately estimated the age differences among the samples. Age differences tended to be accurately estimated when the range of sample ages spanned at least 20 days.

Table 2. Summary statistics of the daily increment counts of the laboratory-reared cod otoliths. Cohort no. 1 grew more quickly than did Cohort no. 2.

Cohort	Age (days)	Mean count	Age-increment discrepancy	s.d.	n	95% confidence interval	
						Lower	Upper
1	12	10.0	2.0	3.1	34	8.9	11.1
2	22	13.2	8.8	6.0	23	10.6	15.7

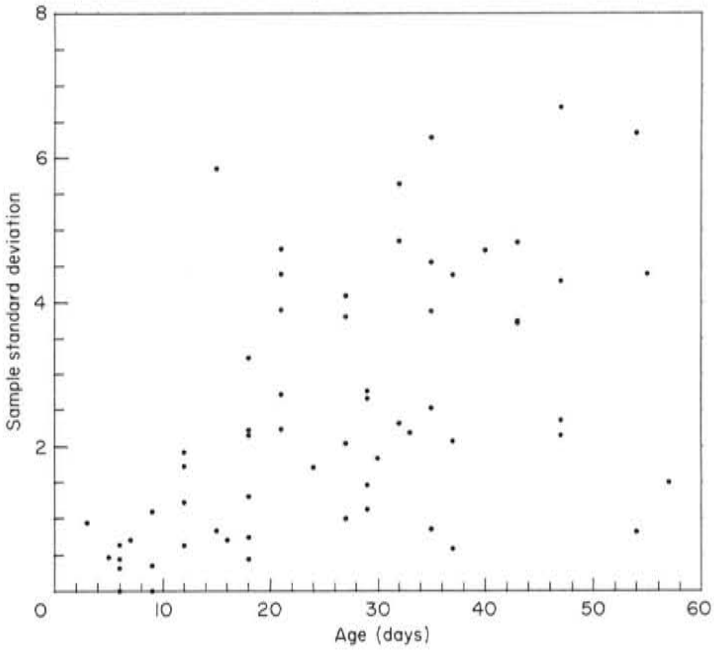

Figure 7. Counting variability as a function of herring sample age. Variability was calculated as the within-sample standard deviation of the daily increment count.

measured with the C.V.) decreased with sample age, stabilizing at a level of about 10–15% (Fig. 8). Differences in relative precision among investigators were not significant, either across all ages or when restricted to larvae >14 days old (ANOVA, $p>0.2$). All investigators had mean C.V.s less than 40%, with the overall mean being 22% (Fig. 9). Note that the 3-day variation in true hatch date within each sample would contribute less than 5% to the C.V. of an average 20-day-old sample.

Variability in the interpretation or measurement of the hatch check

In many fish otoliths, the hatch check (or an analogous structure) serves as a temporal benchmark, marking the point at which increment counts are initiated or ended along an otolith radius. The hatch check in herring otoliths is relatively well-defined; however, given the narrow widths of the increments in the vicinity of the hatch check, any ambiguity in its definition could well introduce significant and consistent differences in increment count between investigators. To test if differences in the interpretation of the hatch check could have contributed to the age estimation differences noted above, measurements of the diameter of the hatch check were compared among investigators. All hatch check measurements were included in the analysis, since hatch check diameter did not vary significantly with age ($p>0.2$).

Measurements of hatch check diameter varied significantly among investigators (ANOVA, $p<0.05$). The mean diameter reported by individual investigators varied by as much as 3.5 μm (15%) from the overall mean (Fig. 10). With increment widths around the nucleus approaching the resolution limit of light microscopy (<0.3 μm) (as noted by Campana *et al.* (1987), and as observed in this

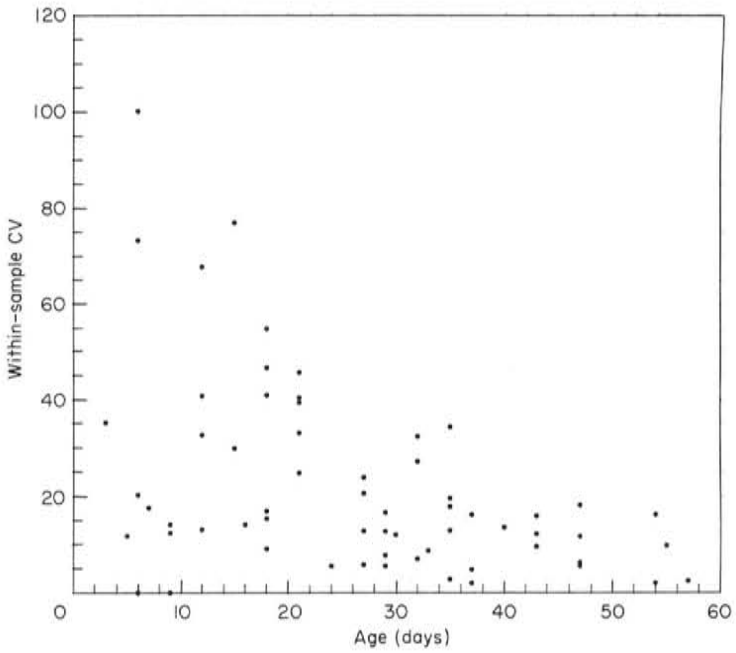

Figure 8. Relative precision of the within-sample herring age estimates as a function of sample age. Precision was defined as the standard deviation divided by the mean (C.V.) for each sample, expressed as a percentage.

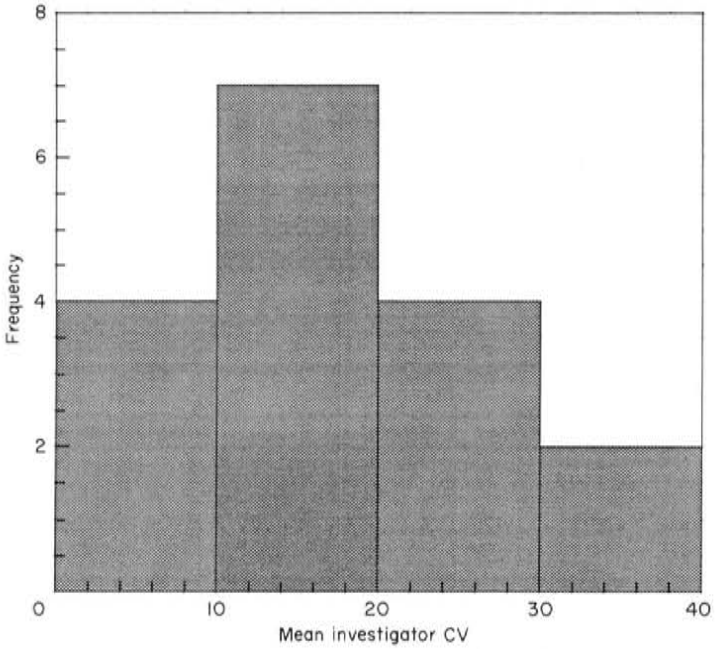

Figure 9. Frequency histogram of the mean investigator C.V.s (relative precision). C.V. values are presented as percentages.

study by the investigator using SEM), bias of such a magnitude in the determination of the hatch check could conceivably have introduced age differences of 6 or more days among investigators. Of course, it is equally possible that the measurement differences were due solely to measurement or calibration error, in which case the hatch check may have been defined in the same way by all study participants. There are no implications for ageing variability

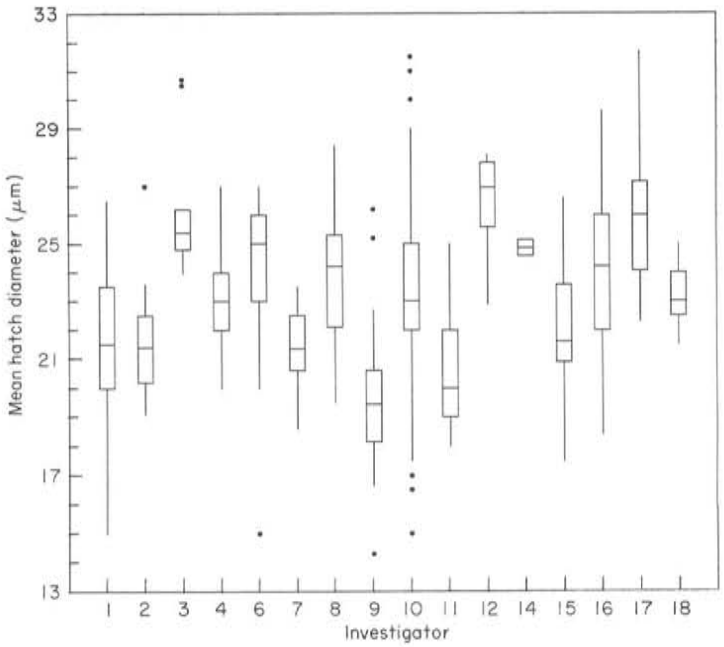

Figure 10. Box and whisker plots of the mean herring hatch check diameter by investigator. The differences, which were significant, could have been due either to differences in the definition of the hatch check or to systematic measurement error.

Table 3. Two-way analysis of variance of the effect of both reader experience and the presence/absence of otolith preparation on the herring age-increment discrepancy. The analysis was restricted to larvae ≥20 days old.

Source of variation	Sum of squares	d.f.	Mean square	F-ratio	Significance
Main effects	571.40	3	190.46	5.06	0.002
Experience	557.24	2	278.62	7.40	0.001
Preparation	0.26	1	0.26	0.01	0.935
Interaction					
Exper. × prep.	257.59	2	128.79	3.42	0.034
Residual	13 932.60	370	37.65		
Total (corrected)	14 761.58	375			

if measurement error was responsible for the measurement differences among investigators. Unfortunately, it was not possible to differentiate between interpretation error and measurement error with these data.

Sources of variability in age estimation

There were four factors which could have influenced the accuracy and precision of individual investigators: the presence/absence of otolith polishing, the type of viewing equipment, the magnification used for viewing, and reader experience. None of the four factors was significantly correlated with any of the others ($p > 0.08$), indicating that no one technique was most associated with experienced investigators. However, certain approaches were more accurate than others, as indicated below.

Of the four main factors examined, otolith polishing and reader experience appeared to have the greatest effect on the accuracy of age determination. Using only herring larvae of a size at which otolith preparation might be expected to be useful (larvae >20 days old, based on Figure 1 and Campana *et al.* (1987)), a 2-way ANOVA indicated that accuracy improved with experience, and that novices who polished their otoliths provided more accurate ages than novices who did not (Table 3). The effect of experience remained significant, but that of otolith preparation did not, when younger larvae (<20 days old) were included in the analysis. When the ANOVA was repeated without the single-investigator

SEM observations, experience and the experience–preparation interaction terms were again significant ($p<0.01$); however, in this analysis, otolith preparation was associated with improved accuracy of both the experienced and the novice investigators. There was no significant effect of either polishing or experience, alone or in combination, on precision (C.V.) (2-way ANOVA, $p>0.16$, $n=39$), although there was a suggestion that precision improved with the level of experience.

The effect of standardized otolith polishing on ageing accuracy was tested in a pairwise comparison of increment counts from polished and unpolished otoliths from the 22-day-old cod cohort (Cohort no. 2). Mean counts by investigators who examined both sample types were significantly higher in the polished samples than in the unpolished samples (mean difference = 4.25; paired *t*-test, $p<0.05$, $n=9$). However, no significant differences were observed between the polished herring otolith sample and unpolished samples of comparable age (30–40 days). The absence of a significant effect in the herring otoliths was at least in part due to low sample size; of the nine investigators who received and examined both the polished sample and unpolished samples of comparable age, all but three routinely polished all of their samples. Thus, the effect of polishing on herring otolith increment counts could not be adequately tested while controlling for inter-investigator differences.

The effect of equipment on age estimation accuracy was significant (ANOVA for larvae >14 days old, $p<0.000$). Scheffe's test indicated that the age-increment count discrepancy was significantly less with the SEM (mean discrepancy = −4.07 (s.e. = 0.93)), and significantly greater with a video system or microscope (mean discrepancy = 11.65 (s.e. = 0.40) and 11.16 (s.e. = 0.30), respectively). However, these results may be misleading. Only one investigator used SEM; thus, equipment effects could not be distinguished from investigator effects in the case of SEM. Further, all of the samples examined with SEM were at least 15 days old. Since the SEM was the only equipment which was associated with increment counts which were greater than the true age of the larvae, it is possible that the SEM counts inadvertently included subdaily increments. SEM examination of larvae <15 days old would be required to reject this possibility. With respect to the photographic results (mean discrepancy = 2.80 (s.e. = 1.17)), the apparent accuracy was an artifact of two widely-divergent investigator means: one stable at a value of 9, and another which became increasingly and sharply negative with the age of the sample. Thus, the accuracy of photography-derived increment counts remains open to question. No differences in accuracy were apparent between video systems and direct microscopy. Furthermore, precision did not vary significantly among any of the equipment types (ANOVA, $p>0.2$).

Neither precision (mean C.V.), hatch-date estimation accuracy, equipment, experience, nor otolith preparation was significantly correlated with the accuracy of the apparent rate of increment formation (the average age difference among samples). However, the variance of the apparent rate was least in the case of the most experienced investigators and those who prepared their otoliths.

We could find no evidence of a learning curve which could have improved the accuracy of individual investigators through the course of this experiment. Since the youngest herring larva samples were the first to be distributed, it is conceivable that a learning curve could have contributed to the observed increase in the age-increment discrepancy with age. If such were true, one might expect to see a trend in increment count within a sample as the investigator examined increasing numbers of larvae, despite the fact that all larvae within a given sample were of the same age. To test for such a trend, the increment count within a sample was first standardized by subtracting the mean increment count of the sample from each individual count within the sample. There was no suggestion of a trend in the standardized increment count within the first sample examined by each investigator, pooled across investigators (regression of pooled standardized counts on larval ID number for all first samples for which at least five larvae were examined, $n=13$, $p=0.35$). Similarly, no trend was evident when the standardized counts from the first two samples of each investigator were pooled ($p=0.55$). Thus, if a learning curve was present, its effects were too subtle to be detected with these data.

Discussion

Is otolith microstructure examination sufficiently accurate and precise as a technique to determine the age, hatch date, growth rate, and mortality rate of larvae under field conditions? The answer appears to depend on the specific application. Estimates of rate processes such as growth and mortality, based on age differences among samples, are likely to be more accurately determined than are simple estimates of age or hatch date. The mean estimated rate of time(= apparent rate of increment formation = 0.95), based on the herring otolith microstructure, was very similar to the value of 1.0 expected of accurate rate calculations. In addition, the mean rate (with a C.V. of 34%) tended to be more precisely estimated than was the hatch date (with a C.V. of 55%). The accuracy of rate calculations relative to age estimates was almost certainly due to the fact that investigator-specific bias is effectively eliminated when subtracting one age estimate from another.

The true age and hatch date of the herring larvae and one of the cod cohorts was not accurately estimated by the majority of the investigators. However, age underestimation on the basis of increment counts is an expected phenomenon in these species, and has been observed in

other studies (Geffen, 1982; Lough *et al.*, 1982; Campana *et al.*, 1987, 1989; Moksness and Wespestad, 1989). If it is expected, the discrepancy between age and increment count is not, in itself, a problem. Where problems could occur is with respect to growth-associated variations in the magnitude of the discrepancy. Since the discrepancy is due to the presence of narrow, unresolvable increments around the nucleus (SEM measurements of increment widths in this study; Campana *et al.*, 1987; Jones and Brothers, 1987), and since the increment sequence can become increasingly compressed (and unresolved) with decreases in growth or development rate, variations in growth rate among larvae will induce misleading variations in apparent age. The differences in the age-increment discrepancy among investigators in this study were apparently due to individual biases, confounded by variations in development rate among larvae. However, the results of other studies, using slower-growing herring larvae, indicate that the discrepancy can be significantly larger than the mean of 10.1 reported here (Geffen, 1982 – >40 days for some treatments; Lough *et al.*, 1982 – 17 days; Campana *et al.*, 1987 – 15–20 days). The implication of such variations is that the addition of a constant value (such as 10.1) to the observed increment count will not necessarily result in an accurate age; the value of the constant will vary with the growth rate of the cohort immediately after hatch. Given the variability in the degree of age underestimation observed in both this and the above-cited studies, it would appear that consistent determination of larval herring ages/hatch dates with a resolution of less than a week could be very difficult.

The accuracy of the cod age determinations was somewhat equivocal, since the larvae were not mesocosm-reared, and thus the age determinations were not representative of those of wild larvae. Despite the poor visibility of the otolith increments associated with lab-rearing conditions, the age of the fast-growing cohort was estimated to within 2 days (10%) in keeping with the results of other studies (Bergstad, 1984; Campana *et al.*, 1989; Radtke, 1989). However, the increment counts underestimated true age by nearly 9 days (40%) in the older, slow-growing cohort, indicating that resolution problems can have a serious effect in this species as well as in herring. Nevertheless, on the basis of this and other studies (Bergstad, 1984; Campana *et al.*, 1989; Radtke, 1989), it would appear that age and hatch date can, in general, be determined with more accuracy in cod than in herring.

The variability in the degree of herring age underestimation among investigators has some interesting implications for both comparative and collaborative studies. With mean hatch date differences among investigators of over a week, it would appear that investigators who compare their own hatch date estimates with those of others should do so with caution. In particular, the results of this study clearly indicate that published reports of significant hatch date differences among independent studies are not necessarily due to real differences in hatch dates. By corollary, intercalibration of otolith interpretations would appear to be necessary for collaborators sharing an inventory of samples for ageing purposes. Note, however, that comparison of growth and mortality rates among investigators is likely to be more reliable than is that of hatch dates.

The levels of precision (C.V.) reported in this study were somewhat higher than those reported elsewhere, at least in part due to the fact that the former incorporated variability among otoliths of the same age, as well as that due to preparation and counting. However, the difference may also reflect the relative difficulty of ageing herring larvae compared to other species. For instance, Schultz (1990) rejected dwarf surfperch (*Micrometrus minimus*) otoliths in which the C.V. exceeded 15%, while Pitcher (1988) reported a C.V. of 5% in tropical damselfish (*Pomacentrus* spp.) otoliths. In contrast, mean C.V.s of older larvae in the current study were between 10–15%. Slower-growing, pelagic larvae may be intrinsically more difficult to age than other larvae, since Savoy and Crecco (1987) noted an inverse relationship between C.V. and age in the pelagic larvae of American shad (*Alosa sapidissima*) which was almost identical to the one presented here. While the presence of high C.V.s in young herring larvae suggested that the relative counting error was high, the relationship between age and counting standard deviation indicated that absolute counting error was still low. For instance, counting variability of ± 1 in a 10-day-old larva corresponds to a C.V. of about 9%, while the same variability in a 1-day-old larva will result in a C.V. close to 90%.

It is important to note that the precision of the age estimates was not correlated with their accuracy, either in terms of the degree of age underestimation or the apparent rate of increment formation. Although precision (or counting reproducibility) can be calculated in the absence of any knowledge of the true age of the larva, and although it serves a useful purpose in reducing variance and ensuring consistency among age readers, this study demonstrates that it cannot be used as a proxy for age estimation accuracy.

Approximately one-half of the variability in the herring age estimates was due to investigator-specific differences. Sources of the variability which were identified included the experience of the investigator with otolith microstructure preparations, the degree of preparation of larger otoliths, and the type of equipment that was used. The range of equipment used in this study reflected that generally applied in otolith studies: direct microscopy and image analysis/video systems were most popular, largely because of the convenience and reduced preparation time. While use of SEM resulted in age estimates which were closest to the true

age, the differences in SEM-derived hatch dates among larvae of different ages indicated that increment interpretation may be as subjective here as in the other equipment types. Indeed, it is clear from this study that individual interpretational skills play a significant role in this field, and that a subjective component is likely to remain part of otolith microstructure examination for the foreseeable future. Subjectivity is not as likely an explanation for the significant differences in hatch check measurements among investigators. Calibration error is the most likely source of those differences.

Based on our experience with other species, larval herring ages based on the otolith microstructure are among the more difficult to obtain with a high degree of accuracy. Accordingly, the levels of accuracy and precision reported here will probably be somewhat lower than those of most other species (although not all e.g. pollock, *Pollachius virens*; Campana, 1989). In particular, we would expect that faster-growing species, such as those without a pelagic larval stage and tropical species, could be interpreted more accurately. Nevertheless, the following implications may be applicable to more than just the species examined for this study: (a) rate calculations based on age differences among samples can be more accurately determined than can age and hatch date estimates, (b) differences in skill, equipment, and preparation among researchers can contribute to apparent age and measurement differences among samples, (c) ageing precision is not a good proxy for ageing accuracy, and (d) there are limits to the resolution which can be expected of this technique. In the case of cod and herring larvae, data resolution on a scale of less than 3 days is probably unrealistic.

Recommendations

Based on our findings, we offer the following recommendations for the conduct of otolith microstructure studies. While based on studies of herring and cod otoliths, we believe that the recommendations have a broad applicability, and may prove useful to a greater or lesser extent in studies of other species.

1. Interpretation, based on skill and experience, has a substantial effect on both accuracy and precision. Newcomers to the field may find it valuable to compare their otolith interpretations against those of other, more experienced workers. Otolith readers of all experience levels are urged to check their interpretational skills against blind-coded larvae of known age, particularly those drawn from a quasi-natural environment such as a mesocosm or an outdoor pond. Such a procedure is recommended as an adjunct to all field studies. The use of blind-coded otoliths and a randomized sample selection order will minimize any learning-, site-, or date-related effects which may exist.
2. Preparation (polishing) of large otoliths can improve accuracy, through enhanced resolution of narrow increments around the nucleus (core). It may also result in improved precision. While the size at which otolith preparation becomes important is probably species-specific, ageing accuracy improved significantly in larval herring when otolith diameters exceeded 40 μm (Campana *et al.*, 1987).
3. Collaborators are advised to exchange otoliths, or otherwise calibrate their interpretations, in order to reduce the probability of introducing investigator-specific bias into their results.
4. Use of as broad a range of ages as possible will improve both precision and the relative accuracy of rate calculations (e.g. growth and mortality) among samples.
5. Through reference to the accuracy and precision levels reported here and elsewhere, the feasibility of applications requiring high temporal resolution should be carefully evaluated prior to implementation. Studies requiring daily resolution may be beyond the capabilities of otolith microstructure studies of many species. Detection limits are particularly important in comparisons of hatch date distributions, where accuracy may limit studies of herring and some other species to the detection of hatch date differences of more than a week.
6. Light microscopy, with or without the aid of a video/image analysis system, is a cost-effective and relatively accurate means of otolith examination compared to SEM. Magnifications of $>1000\times$ are best used when increment widths are less than 2 μm. Photomicrographs may give equivalent results, but resolution can be limited by the fixed focal plane. Where unresolved increments are suspected, the additional expense and labour of SEM may be warranted.
7. Given the magnitude of the error that can otherwise result, microscopic measurements deserve accurate calibration. Video/image analysis system measurements are likely to be more precise than those made with an ocular micrometer.
8. On average, rate calculations based on age differences among samples can be expected to give more accurate results than absolute age or hatch date estimates. However, rate calculations in young larvae may be inaccurate due to the presence of incompletely resolved increments.
9. The primary source of age estimation inaccuracy in pelagic larvae such as herring appears to lie with the first-formed increments. In instances where increments are being dated, say for correspondence with a storm, accuracy can be enhanced by counting inwards from the otolith edge (since the outermost increment is of a known date, corresponding to the date of sampling).

Acknowledgements

This study was only made possible through the encouragement and assistance of the members of the ICES Recruitment Processes Working Group. In particular, we would like to thank the following for donating considerable time and effort to the examination of otoliths for this study: R. Beckedorf and S. Jansen, Inst. f. Fischereiwissenschaft, Hamburg, Germany; George Bolz, Northeast Fisheries Center, Woods Hole, MA, USA; Lindsay Cargill and Mike Heath, Marine Laboratory, Aberdeen, UK; Edgar Dalley, Northwest Atlantic Fisheries Centre, St John's, NF, Canada; Peter Fossum, Institute of Marine Research, Bergen, Norway; Audrey Geffen, Port Erin Marine Laboratory, Isle of Man, UK; Ruth Harrop and John Nichols, Fisheries Laboratory, Lowestoft, UK; Gunnar Joakimsson, Institut für Meereskunde an der Universität, Kiel, Germany; Maria Karakiri, Biologische Anstalt Helgoland, Hamburg, Germany; Estelle Laberge, Jacques Gagné, and Patrick Simard, Institut Maurice Lamontagne, Mont Joli, P.Q., Canada; Françoise Lagardère, Centre de Recherche en Ecologie Marine et Aquaculture de L'Houmeau, L'Houmeau, France; Isabel Meneses, Instituto Nacional de Investigacao das Pescas, Lisboa, Portugal; Henrik Mosegaard, Uppsala University, Uppsala, Sweden; Peter Munk, Danmarks Fiskeri- og Havundersogelser, Charlottenlund, Denmark; Titt Raid, Baltic Fisheries Research Institute, Tallinn, USSR; Pedro Ré, Universidade de Lisboa, Lisboa, Portugal; Martin White, British Antarctic Survey, Cambridge, UK. The expert technical assistance of Marit Bodvin, Inger Henriksen, and Vetle Madsen at the Flødevigen Marine Research Station, and of Joanne Hamel at the Bedford Institute of Oceanography, was greatly appreciated. Chris Chambers and John Neilson kindly made the cod larvae available for our use. John Butler, John Dean, Françoise Lagardère, John Neilson, Dave Secor and an anonymous reviewer all made helpful comments on an earlier version of the manuscript.

References

Anon. 1989. Report of the Working Group on Larval Fish Ecology to the Biological Oceanography Committee of ICES. ICES CM 1989/L: 22. 46 pp.

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

Bergstad, O. A. 1984. A relationship between the number of growth increments on the otoliths and age of larval and juvenile cod. *In* The propagation of cod *Gadus morhua* L. Ed. by E. Dahl, D. S. Danielssen, E. Moksness, P. Solemdal. Flødevigen rapportser, 1: 251–272.

Campana, S. E. 1989. Otolith microstructure of three larval gadids in the Gulf of Maine, with inferences on early life history. Can. J. Zool., 67: 1401–1410.

Campana, S. E., Gagné, J. A., and Munro, J. 1987. Otolith microstructure of larval herring (*Clupea harengus*): image or reality? Can. J. Fish. aquat. Sci., 44: 1922–1929.

Campana, S. E., and Hurley, P. C. F. 1989. An age- and temperature-mediated growth model for cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) larvae in the Gulf of Maine. Can. J. Fish. aquat. Sci., 46: 603–613.

Campana, S. E., and Neilson, J. D. 1985. Microstructure of fish otoliths. Can. J. Fish. aquat. Sci., 42: 1014–1032.

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

Geffen, A. J. 1982. Otolith ring deposition in relation to growth rate in herring (*Clupea harengus*) and turbot (*Scophthalmus maximus*) larvae. Mar. Biol., 71: 317–326.

Jones, C. 1986. Determining age of larval fish with the otolith increment technique. Fish. Bull., 84: 91–104.

Jones, C. E., and Brothers, E. B. 1987. Validation of the otolith increment aging technique for striped bass, *Morone saxatilis*, larvae reared under suboptimal feeding conditions. Fish. Bull., 85: 171–178.

Lough, R. G., Pennington, M., Bolz, G. R., and Rosenberg, A. A. 1982. Age and growth of larval Atlantic herring, *Clupea harengus*, in the Gulf of Maine–Georges Bank region based on otolith growth increments. Fish. Bull., 80: 187–200.

Moksness, E. 1982. Food uptake, growth and survival of capelin larvae (*Mallotus villosus*) in an outdoor constructed basin. FiskDir. Skr. Ser. Havunders., 17: 267–285.

Moksness, E., and Wespestad, V. 1989. Ageing and back-calculating growth rates of Pacific herring, *Clupea pallasii*, larvae by reading daily otolith increments. Fish. Bull., 87: 509–513.

Pitcher, C. R. 1988. Validation of a technique for reconstructing daily patterns in the recruitment of coral reef damselfish. Coral Reefs, 7: 105–111.

Radtke, R. L. 1989. Larval fish age, growth, and body shrinkage: information available from otoliths. Can. J. Fish. aquat. Sci., 46: 1884–1894.

Savoy, T. F., and Crecco, V. A. 1987. Daily increments on the otoliths of larval American shad and their potential use in population dynamics studies. *In* Age and growth of fish, pp. 413–432. Ed. by R. C. Summerfelt and G. E. Hall. Iowa State University Press, Ames, Iowa.

Schultz, E. T. 1990. Daily otolith increments and the early life history of a vivaparous fish, *Micrometrus minimus* (Embiotocidae). Copeia, 1990: 59–67.

Secor, D. H., and Dean, J. M. 1989. Somatic growth effects on the otolith – fish size relationship in young pond-reared striped bass, *Morone saxatilis*. Can. J. Fish. aquat. Sci., 46: 113–121.

Tsukamoto, K., and Kajihara, T. 1987. Age determination of ayu with otolith. Nipp. Suis. Gakk., 53: 1985–1997.

Wespestad, V. G., and Moksness, E. 1990. Observations on growth and survival during the early life history of Pacific herring *Clupea pallasii* from Bristol Bay, Alaska, in a marine mesocosm. Fish. Bull., U.S., 88: 191–200.

Wilson, C. A. [Chairman]. 1987. Glossary Committee. *In* Age and growth of fish, pp. 527–530. Ed. by R. C. Summerfelt and G. E. Hall. Iowa State University Press, Ames, Iowa.