Inaccuracies in routinely collected Atlantic herring (*Clupea harengus*) maturity data and correction using a gonadosomatic index model

LINDSAY R. MCPHERSON[1], KONSTANTINOS GANIAS[2] AND C. TARA MARSHALL[1]

[1]University of Aberdeen, Zoology Building, Tillydrone Avenue, Aberdeen, AB24 2TZ, [2]School of Biology, Aristotle University of Thessaloniki, 54 124 Thessaloniki, Greece

Macroscopic maturity staging data are widely used to distinguish between reproductive and non-reproductive individuals. The implicit assumption is that these data are accurate. The accuracy of macroscopic maturity staging of North Sea herring (Clupea harengus) has not been checked since the macroscopic scale was produced in 1961. The aim of this study was to assess the accuracy of macroscopic maturity staging of female North Sea herring by comparison to histological staging and the gonadosomatic index (GSI). Ovary samples were collected during the North Sea Herring Acoustic Survey in 2006 on-board FRV 'Scotia' (Scotland) and in 2007 on-board FRV 'Scotia' and RV 'Johan Hjort' (Norway). Commercial samples were also collected by Marine Scotland, Aberdeen in both years. The maturity staging error was relatively low in 2006 (21% error) but was much higher on-board FRV 'Scotia' (57%) and RV 'Johan Hjort' (47%) in 2007. There was estimated to be a 27% under-estimation of the spawning stock biomass (SSB) in 2007 due to the differences in the proportion mature but no change in SSB estimates in 2006. GSI cut-off scores, estimated by means of multinomial regression models were successfully able to separate immature females from both mature-active and recovering females; however, there was some overlap between the mature-active and recovering individuals. We conclude that an effective and low-cost means of reducing error in herring maturity studies is the combined use of a four-point macroscopic maturity scale with routinely collected GSI data, the latter acting to validate and fine tune macroscopic staging.

Keywords: Atlantic herring, GSI, maturity scale, operator error, fisheries management

Submitted 8 April 2010; accepted 9 January 2011; first published online 14 March 2011

INTRODUCTION

The maturation process is most often described through routine macroscopic assessment of the gonad using a maturity scale based on the appearance of the gonad when observed with the naked eye (e.g. Table 1). Macroscopic staging is the quickest method for assigning maturity status and is therefore routinely used, particularly at sea. High accuracy of macroscopic maturity staging is critical for drawing correct inferences about the magnitude of the spawning stock biomass (SSB), size and age at maturity and length of the spawning season (Carrassón & Bau, 2003). However, it can be highly subjective and involves intensive training (West, 1990) which may or may not be given to personnel working in the field. Due to these constraints, the reliability of macroscopic staging has been questioned by a number of studies (McQuinn, 1989; West, 1990; Saborido-Rey & Junquera, 1998; Bromley, 2003; Hunter & Macewicz, 2003; La Mesa *et al.*, 2003). Very few studies have tested the accuracy of macroscopic staging against relatively unambiguous microscopic assessments of maturity stage. Without this verification,

macroscopic staging should be regarded as potentially unreliable (Bromley, 2003). The few studies that have been carried out comparing histological and macroscopic assessments have often found that the error, or percentage of fish incorrectly staged, is high (Table 2). Systematic errors between virgin and adult fish are of particular concern as they can lead to consistent over- or under-estimation of the spawning stock biomass (SSB).

Histology is the most accurate method for determining the maturity stage of fish (West, 1990). Histological staging involves studying structures within oocytes or testes and so gives an unambiguous interpretation of maturity status (West, 1990; Vitale *et al.*, 2006). Bowers & Holliday (1961) used histology to modify the Lea macroscopic maturity staging scale for Atlantic herring (*Clupea harengus*) (Lea, 1910). This was to ensure that macroscopic stages accurately reflected changes in gonad development as described by microscopic evaluations. Very little histological work has been undertaken on Atlantic herring since this study was carried out in 1961. Furthermore, no studies have been carried out to estimate the incongruities between macroscopic and microscopic staging techniques. This is most likely due to the labour-intensive, expensive and time consuming nature of histology (West, 1990; Tomkiewicz *et al.*, 2003).

Due to the impracticalities and costs of histological maturity staging, it has been suggested by Vitale *et al.* (2006) that

Corresponding author:
L.R. McPherson
Email: lindsay.mcpherson@abdn.ac.uk

Table 1. British and Norwegian scales for macroscopic maturity staging Atlantic herring (*Clupea harengus*). These macroscopic scales are both based on the staging scale from Bowers & Holliday (1961).

Maturity stage	British macroscopic description	Norwegian macroscopic description
I, virgin	Gonads very small—threadlike; 2–3 mm broad; ovaries wine red	Juvenile phase. Gonads thread-like, completely transparent and colourless. Difficult to determine sex
II, virgin with small gonads	The height of ovaries is about 3–8 mm; eggs not visible to the naked eye but can be seen with a magnifying glass; ovaries bright red colour	Gonads are somewhat larger in volume, sex is easier to determine. The gonads continue to be transparent and colourless with a hint of colour
III, maturing	Gonads occupy about half of the ventral cavity; breadth of the sexual organs is between 1 and 2 cm; eggs are small but can be distinguished with the naked eye; the ovaries are organs	Gonads opaque but developed in volume. Distinct veins. Ovaries have yellow/white eggs in lamellae and can occupy half of the body cavity or more
IV, maturing	Gonads are almost as long as the body cavity; eggs larger than in 3, varying in size and opaque; ovaries orange or pale yellow in colour	Gonads larger in volume. Distinct veins. Ovaries yellowish or white, can occupy 2/3 or more of the body cavity depending on the condition of the fish. The eggs can be seen distinctly and feel like grain. The eggs in the front part of the gonad are beginning to become transparent
V, maturing	Gonads fill the body cavity; eggs are large and round; some are transparent; ovaries are yellowish; eggs do not flow	Ovaries fill the entire body cavity. Most of the eggs are transparent
VI, spawning	Ripe gonads; eggs transparent; eggs flow freely	Running gonads. A light pressure on the abdomen causes the eggs to run
VII, spent	Gonads baggy and bloodshot; ovaries are empty and only contain a few residual eggs; body cavity may contain bloody fluid. At this stage there can be difficulty in deciding sex; if the gonads are spread out it is easier to view the leading edge—sharp for male and rounded for female	Gonads loose, contain remaining eggs
VIII, recovering	Ovaries are firm and larger than virgin herring in immature 2. Eggs are not visible to the naked eye. The walls of the gonads are striated vertically and blood vessels are prominent. Gonads are wine-red in colour (this stage passes into III)	Gonads are small. Eggs are not visible. Difficult to distinguish from stage 2/3

proxies should be used in an attempt to make macroscopic staging more accurate. One potential proxy of maturity stage may be the gonadosomatic index (GSI), which is the weight of the gonad expressed as a percentage of the total weight or somatic weight of the fish. GSI is known to increase throughout gonadal development (Potoschi *et al.*, 1999) and so a model using GSI may be able to accurately describe maturity stages. As whole weight and gonad weight are quick and easy to measure this method would be less laborious than histological staging. This method has been employed by McQuinn (1989) who created a GSI model which was able to successfully separate spring- and autumn-spawning north-western Atlantic herring. However, the use of GSI as a proxy has been criticized, in part because GSI can be correlated with fish size in some species (DeVlamming *et al.*, 1982; Erickson *et al.*, 1985; West, 1990). However, if GSI is not related to fish size it could potentially be used to correct macroscopic maturity staging if the accuracy of the staging is not satisfactory.

The aim of this study was to assess the accuracy of macroscopic maturity staging of female North Sea herring. Macroscopic maturity stages from research vessels and commercial catches were compared to histological maturity stages to quantify the percentage of incorrectly staged fish,

Table 2. Studies of error in routinely used macroscopic maturity staging methods for different species. Macroscopic maturity staging was compared to histological maturity staging for all studies.

Species	Nature of error	N	Reference
Walleye pollock (*Theragra chalcogramma*)	25% of macroscopic maturity stages were incorrect as compared to histological assessment	126	Williams (2007)
Kattegat cod (*Gadus morhua*)	Macroscopic staging error estimated to have caused a 21–35% overestimation of the female spawning stock biomass	668	Vitale *et al.* (2006)
Baltic cod (*Gadus morhua*)	80% of females were assigned incorrect maturity stage through macroscopic determination	139	Tomkiewicz *et al.* (2003)
Kingfish (*Scomberomorus commerson*)	53% of fish were found to be incorrectly staged macroscopically	236	Claereboudt *et al.* (2005)
Icefish (*Chionodraco hamatus*)	Discrepancies between macroscopic and histological readings that could affect estimates of the proportion of mature fish	362	La Mesa *et al.* (2003)

N, sample size.

termed here as the total error. The percentage agreement between macroscopic and microscopic assessments after chance agreement has been removed was then calculated. Finally, GSI cut-off points that could be used to separate immature and mature-active fish and mature-inactive and mature-active fish were derived by means of multinomial regression models and their use was validated using fish that had been histologically staged. The potential for systematic over- or under-estimation of SSB due to incorrect maturity staging and the use of GSI to correct any errors was then considered.

MATERIALS AND METHODS

Sampling at sea

Female North Sea herring were collected during the North Sea Herring Acoustic Survey (HERAS) in 2006 and 2007. Samples were collected on-board FRV 'Scotia' (UK) from 1–21 July 2006 and 29 June–19 July 2007 and RV 'Johan Hjort' (Norway) from 21 June–19 July in 2007. All fish obtained at sea were sampled immediately after capture. Mid-water trawl nets were used and, in some cases on the FRV 'Scotia', a multi-sampler was employed which has three cod-ends that can be opened and closed remotely and independently. On-board FRV 'Scotia', for each fish total wet weight, gutted weight (gonad, stomach and viscera removed) (to the nearest 1g) and gonad wet weight (to the nearest 0.1g) were measured. For each fish sampled on-board RV 'Johan Hjort' total wet weight (to the nearest 1g) was measured. On-board both ships, the trained technician sexed each fish and categorized them into macroscopic maturity stages using two eight-point scales based on the gross appearance of the gonads (Table 1). The scales of the two nations are very similar and should be interchangeable. It should be noted that the technician that assigned maturity stages on-board FRV 'Scotia' in 2006 was different from the two technicians that assigned maturity stages on-board FRV 'Scotia' and from market samples in 2007. It is normal to have several technicians assigning maturity stages within and across surveys and so this can be considered as a random source of background error.

In 2006 up to five ovary samples of each maturity stage were taken for histological analysis. In 2007 up to 14 ovary samples of each maturity stage were taken (Table 3). A small section (between 0.5 and 1.5 cm²) was removed from the central portion of each gonad and was immediately fixed in 40% formalin. It was not possible to obtain samples from all maturity stages due to restrictions in time and location during surveys. However, the stages collected are representative of the stages that are routinely sampled on-board FRV 'Scotia' and RV 'Johan Hjort' during HERAS.

Market sampling

As not all maturity stages were observed during the survey cruises, it was necessary to sample fish from commercial fishing vessels. Commercial samples were obtained by the Marine Scotland, Aberdeen in August 2006 and 2007 from day-trips. These fish were caught by commercial trawl and placed on ice until sampling. All fish were sampled within 24 hours of capture at the Marine Scotland. Biological information was obtained in the same way as on FRV 'Scotia' and, again, a trained technician assigned sex and maturity stage as part of routine sampling. Gonad samples were removed and fixed using the same method as on-board the vessels. It was not possible to collect market samples from Norway (see Table 3 for market sample sizes for each maturity stage and year).

Histological slide preparation and staging

All samples were processed in the Histology and Electron Microscopy Core Facility of the Institute of Medical Sciences (IMS), Aberdeen University. Altogether, 107 wax histological slides were created (Table 3). Gonad samples were transferred to 70% ethanol and left overnight in 95% ethanol before dissecting to block size to prevent tissue damage from trimming. After trimming, samples were

Table 3. Sample sizes by macroscopic maturity stage on the three cruises (FRV 'Scotia' 2006, FRV 'Scotia' 2007 and RV 'Johan Hjort' 2007) and from the Scottish Herring Acoustic Survey (HERAS) data. Samples were staged and sexed by experienced technicians on-board. Samples for histological analysis were chosen at random. Market samples were obtained from Marine Scotland, Aberdeen. The full survey data from the Scottish HERAS was also obtained from Marine Scotland.

| Macroscopic maturity stage | Number of samples analysed | | | |
| | Histological samples | | | Survey data |
	FRV 'Scotia' 2006	FRV 'Scotia' 2007	RV 'Johan Hjort' 2007	HERAS ('Scotia') 2006–2008
I	0	0	0	11
II	4	8	8	108
III	5	14	14	2259
IV	5	7	8	192
V	4 (m)	7 (m)	0	0
VI	5 (m)	0	0	0
VII	0	0	0	0
VIII	5	14	0	637
Total	27	50	30	3207

(m), market samples.

Table 4. Histological scale for maturity staging of Atlantic herring (*Clupea harengus*). This scale is based on work by Bowers & Holliday (1961), Landry & McQuinn (1988), Tyler & Sumpter (1996) and Koya *et al.* (2003).

Maturity stage	Histological description
I, virgin	Circular cells of approximately $10–14 \mu m$ with a large nucleus
II, virgin with small gonads	Appearance of cortical alveoli in the cytoplasm of the oocyte
III, maturing	Primary yolk globule stage: appearance of yolk globules in the peripheral cytoplasm. These appear as small bright pink structures with haematoxylin and eosin staining
IV, maturing	Secondary yolk globule stage: the yolk globules fill the entire cytoplasm right up to the nucleus and cortical alveoli are only in the oocyte periphery
V, maturing	Tertiary yolk globule stage: fusion of yolk globules into yolk masses occurs. Migratory nucleus stage: movement of the nucleus towards the periphery of the cytoplasm where it is disintegrated
VI, spawning	Ovulated eggs: follicle no longer surrounding the cell
VII, spent	Ruptured, empty follicles contract and cast off granulosa cells
VIII, recovering	Cortical alveoli present in cytoplasm surrounding the nucleus. This is differentiated from the virgin cortical alveoli stage by a thick, convoluted tunica or the presence of post-ovulatory follicles

embedded in wax, sectioned at $5 \mu m$ and stained with haematoxylin and eosin. Histological maturity staging was carried out based on work by Bowers & Holliday (1961), Landry & McQuinn (1988), Tyler & Sumpter (1996) and Koya *et al.* (2003) (see Table 4 and Figure 1 for histological staging scheme).

Analysis of error

The percentage of fish incorrectly staged for each maturity stage was calculated and this was termed the stage-specific error. Total error is the percentage of fish incorrectly staged across all maturity stages. Cohen's (1960) coefficient of agreement kappa (κ) was also calculated for each vessel in each year in order to determine the proportion of agreement once chance agreement had been taken into consideration. It should be noted that κ gives the proportion *agreement* and this is, therefore, the opposite of the proportion *error*. κ is given by:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \qquad (1)$$

where P_o is the observed proportion of agreement and P_e is the expected proportion of agreement by chance. P_e is given by:

$$P_e = \sum_{i=n}^{n} P_{a,i} - P_{b,i} \qquad (2)$$

where $P_{a,i}$ is the proportion of the sample macroscopically attributed to category i and $P_{b,i}$ is the proportion of the sample attributed to category i histologically. A κ value of 1 indicates complete agreement whereas a κ value of -1 indicates complete disagreement. A κ value of 0 indicates that all agreement is due to chance.

Re-estimation of spawning stock biomass

The SSB is calculated as the estimated total number of fish (N) multiplied by the mean weight at age (W_a) and the proportion of fish mature at age (M_a). Data on N, W_a and M_a estimated from the HERAS surveys in July 2007 were taken from Table 2.3.1.2. of the 2008 Herring Assessment Working Group report (ICES, 2008). Data on the macroscopic maturing staging error was not available by age. Therefore the M_a was then recalculated for each age using the total error observed between immature and mature fish (including stage VIII) in 2007, i.e. the error applied was the same for every age-class. It should be noted that this assumes that the maturity staging error is equivalent for females and males. The SSB was calculated with the published M_a and then again with the re-estimated M_a to estimate the difference in the SSB caused by maturity staging errors. The SSB was not re-calculated for 2006 as there were no macroscopic maturity staging errors that would lead to incorrect estimation of M_a in that year (see Results section).

Validation of GSI

The GSI (%) was estimated as the gonad weight (W_{go}) divided by the gutted weight (W_{gt}), and multiplied by 100. In order to assess if GSI could be used as a predictor of maturity we first needed to validate its use, i.e. test whether GSI is affected by fish size. To do so, the nature of the relationship between the W_{gt} and W_{go} for each maturity stage was analysed. In order to have a larger sample size for this, the dataset containing all females that were routinely sampled on FRV 'Scotia' during HERAS in 2006 and 2007 was used (N = 3207; Table 3). Analysis using the HERAS dataset only took into account stages I, II, III, IV and VIII since all other stages (i.e. V, VI and VII) were not recorded. For this analysis, immature fish (stages I and II) were amalgamated into one class as there were relatively few samples in these stages and both stages are assumed not to have potential to mature in the SSB estimation.

The relationship between W_{gt} and W_{go} in the i_{th} maturity stage is described by the formula (DeVlamming *et al.*, 1982):

$$W_{go} = a_i W_{gt}^{b_i} \qquad (3)$$

For the GSI to be valid, the slope b_i of the relationships should not be different among maturity stages and isometry should be apparent (i.e. $b_i = 1$) (DeVlamming *et al.*, 1982; Erickson *et al.*, 1985; Somarakis *et al.*, 2004). We estimated a_i and b_i on a logarithmic scale in order to use linear modelling techniques:

$$\log(W_{go}) = \log a_i + b_i \log(W_{gt}) \qquad (4)$$

In order to test for a difference between slopes, an analysis of covariance (ANCOVA) was carried out using the HERAS dataset. This ANCOVA had $\log W_{gt}$ as the dependent variable
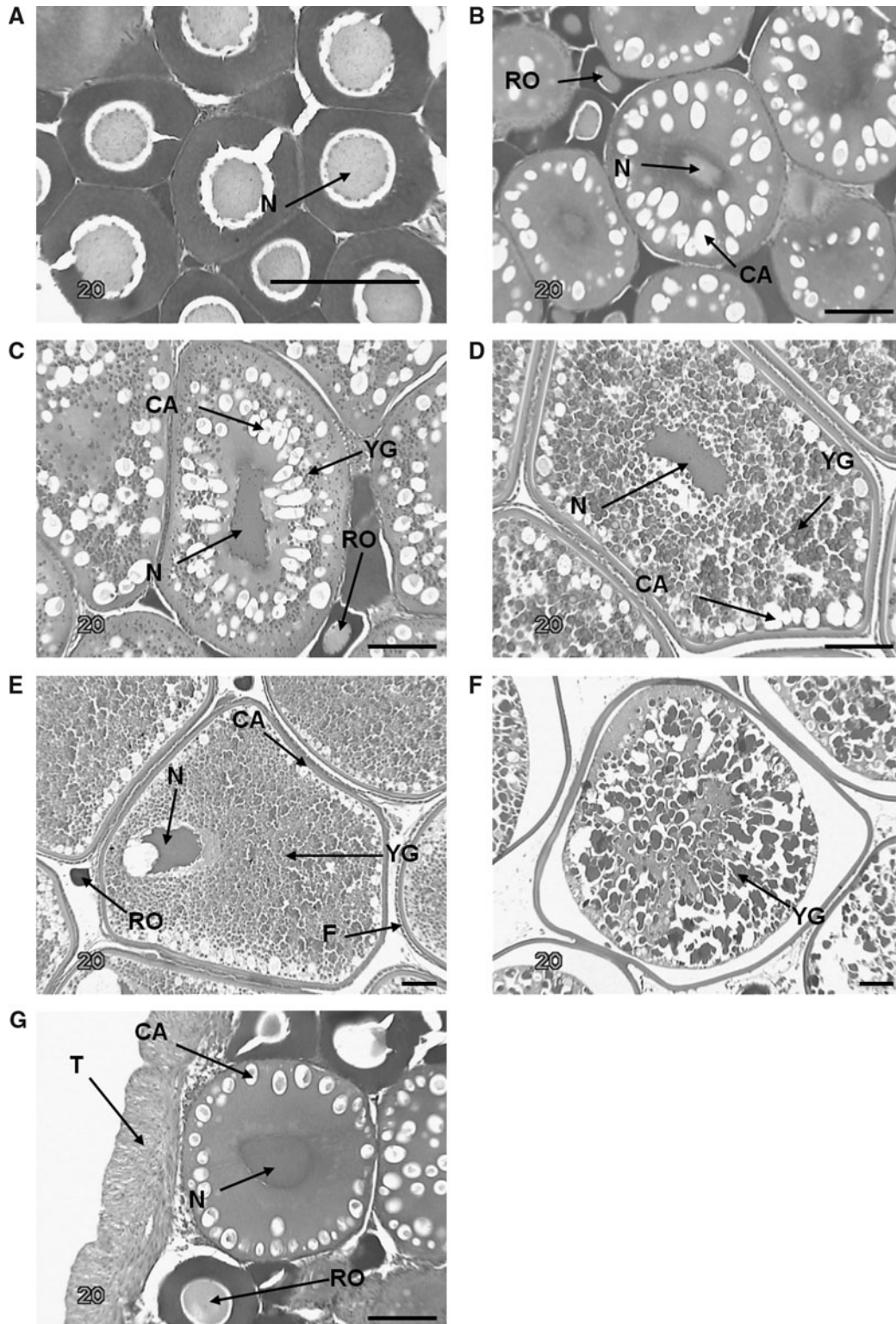
Fig. 1. Micrographs of sections (5 μm) of Atlantic herring (*Clupea harengus*) ovaries at different stages of development. Stain, haematoxylin and eosin. (A) peri-nucleolus stage (stage I); (B) cortical alveolus stage for virgin herring (stage II); (C) primary yolk globule stage (stage III); (D) secondary yolk globule stage (stage IV); (E) migratory nucleus stage (stage V); (F) ovulated egg (stage VI); (G) cortical alveolus stage for recovering herring (stage VIII). N, nucleus; CA, cortical alveoli; RO, resting oocyte; YG, yolk granules; F, follicle; T, tunica. Scale bars = 100 μm (B, C, D, E, F & G); 20 μm (A).

and $\log W_{go}$, maturity stage and the interaction between $\log W_{go}$ and maturity stage as the explanatory variables. If the interaction between $\log W_{go}$ and maturity stage was significant, then the slopes between different maturity stages were significantly different. In order to compare the slopes of all

maturity stages, the ANCOVA was carried out multiple times with different maturity stages as the base comparison to ensure that comparisons between all stages were made.

We then tested for isometry by comparing coefficient b_i in each maturity stage with unity by means of a *t*-test. In the case

of isometry the GSI would be equal to a_i and thus independent of female weight. As maturity stage IV had the lowest stage-specific error (see Results section), we tested for isometry in the above relationship (Equation 4) mainly using this stage.

Estimation of GSI cut-off scores

Multinomial logistic regression (MLR) was subsequently used to distinguish between immature (stages I and II), mature-active (stages III and IV) and recovering (stage VIII) classes through calculating multiple cut-off GSI scores between the different maturity classes again using the HERAS dataset. This kind of analysis was judged as preferable to using separate binary logistic regression models because in the MLR models, the likelihood function utilizes the data involving all categorical variables in a single function. Mature-active fish were designated as the reference category and each of the other categories (immature and recovering) were compared with this baseline.

The predicted conditional probability of being in the y_i maturity class using GSI as the single predictor variable is given by the formula:

$$P(Y = y_j | GSI) = \frac{\exp(\alpha_j + \beta_j * GSI)}{\sum_{h=1}^{J} \exp(\alpha_h + \beta_h * GSI)} \quad (5)$$

$$j = 1, 2 \ldots J$$

where α_j and β_j are the intercept and the slope respectively corresponding to GSI. The parameters corresponding to the reference category are equal to zero.

A cut-off GSI score can be defined as the test score at which an individual is as likely to be in category j as in category $j + 1$. GSI is increasing as fish pass from immature to the mature-active state but decreases as fish pass from the mature-active to the recovering state (Figure 2). In that respect, the cut-off scores could not be estimated by splitting the predictor
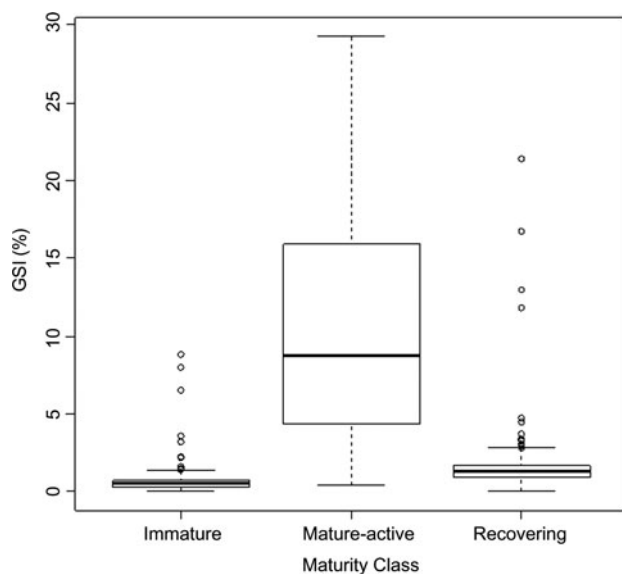

Fig. 2. Relationship between maturity status (immature, stages I–II; mature-active, stages III–IV; recovering, stage VIII) and gonadosomatic index for Atlantic herring (*Clupea harengus*) during the Scottish Herring Acoustic Survey (HERAS) in June/July 2006–2009.

(GSI) in the ordered categories that correspond to the intersection of the adjacent curves (immature and recovering and recovering with mature-active). As a consequence, GSI_1 and GSI_2 were estimated as the test scores that correspond to the intersection of the immature and recovering curves to the mature-active curve respectively, using the formula:

$$GSI_x = \frac{\alpha_x}{\beta_x} \quad (6)$$

In order to test if these GSI cut-off values were valid, the GSI of females that had been histologically assessed (i.e. with accurate maturity stages) were examined to see if the maturity stages and GSI corresponded to the estimated cut-off values. All statistical analyses were carried out using the R statistical software version 2.9.2. (R Development Core Team, 2009).

Accuracy of macroscopic assessments

Macroscopic maturity assessments differed in accuracy both between ships and years. On-board FRV 'Scotia' and from market samples in 2006 the total error was 21% (Table 5). In 2006 all stage-specific errors apart from one were due to macroscopic assessments being less mature than they were (Table 6). In 2006, the κ value was 0.78 and there was no disagreement between virgin fish (stages I and II) and adults (stages III–VIII) (Table 6). However, it should be noted that one fish that was macroscopically assessed to be female stage II was, in fact, a male. The total error was much higher on-board FRV 'Scotia' and in market samples in 2007 at 58% (Table 5). There was much confusion in assignments of maturity stage between stages VIII and III (Table 6), so fish assigned as recovering individuals were actually maturing. One individual that was classified as a repeat spawner (stage VIII) was a juvenile (stage II). The total error reduced to 49% if market samples were not included in the calculation. This was due to a total error of 86% in market sample maturity staging in 2007. The κ value calculated for all samples on-board FRV 'Scotia' and market samples was 0.22. Some errors that were found in macroscopic maturity staging on FRV 'Scotia' could be clearly seen in a simple scatterplot of gutted weight versus gonad weight from the years 2006–2009 (Figure 2).

On RV 'Johan Hjort' in 2007 there was a total error of 47% (Table 5). The largest stage-specific error on RV 'Johan Hjort' was that 63% of stage II ovaries were actually stage VIII (Tables 7). 57% of macroscopic stage III's were incorrectly staged. This stage-specific error for stage III's went in both directions with 35% actually being more mature (stage IV) and 21% actually being less mature (stage VIII) suggesting that the error on-board RV 'Johan Hjort' is non-directional (Table 7). The κ value for maturity staging on-board RV 'Johan Hjort' in 2007 was 0.37.

Re-estimation of spawning stock biomass

Firstly, it should be noted that there was no error in the distinguishing between immature and mature (both active and

Table 5. Stage-specific error, total error and Cohen's κ (kappa) for North Sea herring on all three acoustic survey cruises. % error is the percentage of fish that were incorrectly staged (stage-specific error). Total error is the total percentage of all fish of all stages that were incorrectly staged and κ (kappa) values show the proportion agreement between macroscopic and histological staging after chance agreement is removed. Market samples were obtained from Marine Scotland, Aberdeen.

Ship/year	Maturity stage (macroscopic)	% error
FRV 'Scotia' 2006 and market samples	I	-
	II	50
	III	20
	IV	20
	V	20 (m)
	VI	0 (m)
	VII	-
	VIII	40
	Total error	21
	κ	0.78
FRV 'Scotia' 2007 and market samples	I	-
	II	63
	III	64
	IV	29
	V	86 (m)
	VI	-
	VII	-
	VIII	50
	Total error	58
	κ	0.22
RV 'Johan Hjort' 2007	I	-
	II	63
	III	57
	IV	13
	V	-
	VI	-
	VII	-
	VIII	-
	Total error	47
	κ	0.37

(m), market samples; κ, Cohen's coefficient kappa; -, no samples obtained.

inactive) fish in 2006 and so the SSB was not re-calculated as there would be no change. The SSB was estimated from the July 2007 data taken from ICES (2008) to be 1.18 million tonnes with the published M_a and 1.50 million tonnes with the recalculated M_a incorporating maturing staging error. Therefore, macroscopic maturing staging error led to an under-estimation of the SSB of 27%. With data from only two years it is not possible to know if this is a random or systematic error. However, any inter-annual variation in staging error could be caused by the differences in accuracy between individual technicians.

Gonadosomatic index model

For differences in GSI between immature, mature-active and resting/recovering individuals see Figure 2. The ANCOVA results showed that there was no significant difference in the slope of the relationship between $\log W_{gt}$ and $\log W_{go}$ between stages II and III (ANCOVA, $P = 0.7$) or between stages IV and VIII (ANCOVA, $P = 0.1$). However, all other stage combinations displayed significant differences in their

Table 6. Histological versus macroscopic maturity stage determination of female North Sea herring sampled on-board FRV 'Scotia' in 2006 and 2007, from market samples assessed at Marine Scotland in 2006 and 2007. The values outside of the shaded cells represent misidentification in macroscopic evaluation.

Macroscopic maturity stage	Histological maturity stage																	
	I		II		III		IV		V		VI		VII		VIII		Total	
	2006	2007	2006	2007	2006	2007	2006	2007	2006	2007	2006	2007	2006	2007	2006	2007	2006	2007
I																	0	0
II	1		2	3												5	3	8
III					4	5	1	9									5	14
IV							4	5		2							4	7
V								6(m)	4(m)	1(m)							4	7
VI											5(m)						5	0
VII																	0	0
VIII				1	2	6(m)									3	7	5	14
Total	1	0	2	4	6	11	5	20	4	3	5	0	0	0	3	12	26	50

Table 7. Histological versus macroscopic maturity stage determination of female North Sea herring sampled on-board 'RV Johan Hjort' in 2007. The values outside of the shaded cells represent misidentification in macroscopic evaluation.

Macroscopic maturity stage	Histological maturity stage								Total
	I	II	III	IV	V	VI	VII	VIII	
I									0
II		3						5	8
III			6	5				3	14
IV			1	7					8
V									0
VI									0
VII									0
VIII									0
Total	0	3	7	12	0	0	0	8	30

slopes (ANCOVA, $P < 0.05$) (Figure 3). As the staging accuracy was highest for stage IV fish on FRV 'Scotia' (Table 5), a linear model was fitted for just stage IV samples to test for isometry (linear regression, $r^2 = 0.2$, $P < 0.05$, N = 192). The slope of the relationship between $\log W_{gt}$ and $\log W_{go}$ for stage IV fish was not significantly different from unity (*t*-test, $P = 0.44$, estimate = 0.93, SE = 0.09) suggesting that the relationship is isometric. Therefore, the use of GSI to predict maturity stage would not be biased by fish size.

The output of the MLR model for estimating GSI cut-off scores is illustrated in Figure 4. The lines in Figure 4 represent the probability of being in each maturity class at a given GSI value. The goodness of fit of the MLR model was analysed using the likelihood ratio test, $G = D_o - D_m$, where D_m

Fig. 3. Relationship between log gutted weight and log gonad weight for all stages present in the Atlantic herring (*Clupea harengus*) during the Scottish Herring Acoustic Survey (HERAS). Data are from the HERAS survey June/July 2006–2009. Lines represent simple linear models. Black circles, immature (stages I & II); black triangles, stage III; grey triangles, stage IV; grey circles, stage VIII.

1355.8 is the residual deviance and D_o 2538.8 is the deviance of the null model which includes only the intercepts. G has a large-sample Chi-square distribution with degrees of freedom equal to the difference in the number of parameters estimated. The value of G was very large (G 2538.8 – 1355.8 1187.1; df 2; P 0.0001), indicating a highly significant decrease in the deviance when GSI is included in the model. The pseudo r^2 value (r^2 1 – D_m/D_o) was 0.467 suggesting that the inclusion of GSI decreased the model's deviance by 46.7%.

Despite the fact that the overall performance of the MLR model was very good the degree of overlap between immature and recovering classes was large and their maximum fitted probability was lower than 1, only reaching ∼0.6 (Figure 4). On the other hand, both immature and recovering classes were clearly distinguishable from the mature-active class and the GSI cut-off scores were estimated to be 1.042 and 1.798 respectively.

When the cut-off values were validated using histologically staged samples, the immature fish were successfully separated from the mature-active and recovering fish as all immature fish were below the lower GSI cut-off value (Figure 5). Furthermore, 90% of recovering fish (nine of the ten fish) were within the two GSI cut-off values. 90% of mature-active fish were above the upper GSI cut-off value, and so 10% of mature-active fish would be classed by the model as recovering fish (Figure 5). No recovering fish were above the upper GSI cut-off.

DISCUSSION

The magnitude of the total error in 2006 on-board FRV 'Scotia' was low relative to other studies (Table 2). However, the magnitude of the total errors on-board FRV 'Scotia' and RV 'Johan Hjort' in 2007 were of similar levels to those found by Claereboudt *et al.* (2005) who found 47% accuracy in kingfish (*Scomberomorus commerson*) and Tomkiewicz *et al.* (2003) who detected an accuracy of 20% for Baltic cod (*Gadus morhua*) staging. The sample sizes in all of these studies are relatively small and similar to the current study, due to the expensive, time-consuming nature of histology (West, 1990). Collectively, these results suggest a lack of standardization in training and expertise between maturity readers.
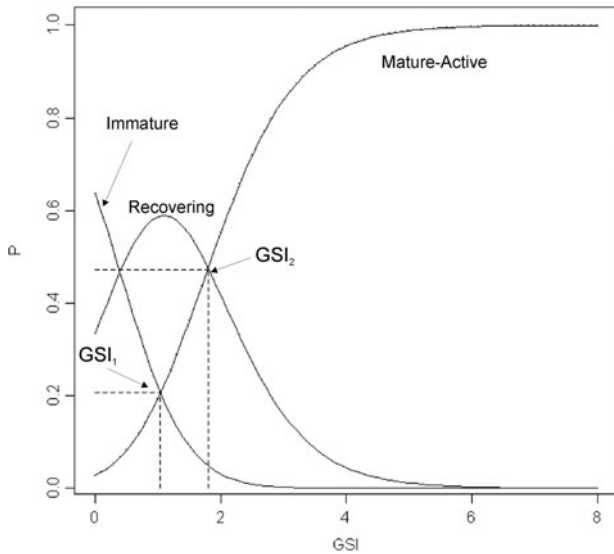
Fig. 4. Results from multinomial logistic regression for Atlantic herring (*Clupea harengus*) using the North Sea Herring Acoustic Survey dataset (N = 3207). *P*, probability of being mature.

Fig. 5. Validation of gonadosomatic index (GSI) cut-off values estimated from multinomial logistic regression for Atlantic herring (*Clupea harengus*). All samples were histologically staged. Circles, immature (stages I–II); filled squares, mature-active (stages III–IV); triangles, recovering (stage VIII). Plotted data come from samples that have been maturity staged histologically (N = 51). Solid line, lower GSI cut-off value (all fish below are predicted to be immature); dashed line, upper GSI cut-off value (all fish above are predicted to be maturing). Fish between the two GSI cut-off values are predicted to be recovering. GSI cut-off values were calculated using data from the North Sea Herring Acoustic survey (N = 3207).

Inaccuracies were found in differentiating between juvenile and adult fish, an error that will affect the SSB. A large proportion of fish were macroscopically staged as recovering (stage VIII) but were, in fact, immature (stages I–II) (Tables 6 & 7). All of these fish were sampled on-board FRV 'Scotia' in 2007 and RV 'Johan Hjort' in 2007. However, only one fish that was macroscopically staged as immature was found to be recovering after histological examination. Therefore, the inaccuracies found in this study would lead to an under-estimation of the SSB as the proportion of immature individuals would be inflated. Although there was no change in SSB estimates in 2006, maturity staging errors were estimated to cause an under-estimation of SSB of 26% in 2007. These results are in contrast to both Gerritsen & McGrath (2006) and Vitale *et al.* (2006) who observed maturity staging errors that would lead to over-estimation of the SSB for whiting (*Merlangius merlangus* L.) and cod, respectively. The extent of the over-estimation of SSB was not calculated for whiting but the Kattegat cod female spawning biomass was found to be over-estimated by up to 35% (Vitale *et al.*, 2006). While the macroscopic maturity staging errors found in this study would not lead to over-estimation of the proportion mature, and therefore potential overfishing of the stock, these errors could have implications for studies that use these data.

In order to accurately estimate the SSB, it is essential that the proportion mature is estimated correctly. Therefore, a tool that can split immature and recovering fish would be valuable, as this is where the most critical maturity staging errors lie. In the calculation of the proportion mature, maturing/spawning and recovering individuals are grouped into one class, 'mature', as they all have the potential to mature that year. Therefore, the crucial point at which the fish need to be split is between immature and maturing/recovering. The use of GSI cut-off points was successfully validated in this study with 91% fish correctly categorized. While there was some variability in the ability to separate recovering fish from maturing-active fish, the model was able to accurately assign 90% of the mature-active fish into the correct category. This has no implications now, but if recovering fish are classed
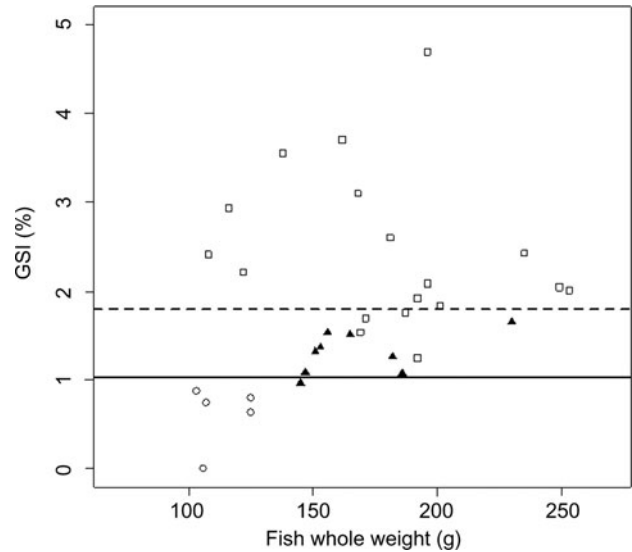
as 'skip spawners' (e.g. Kennedy *et al.*, 2010) rather than 'mature' in the calculation of the proportion mature, which is one possibility for the future, then the ability to split these two groups will become critical. This tool could be routinely and easily used to successfully correct errors in the macroscopic maturity stages between immature and recovering fish. McQuinn (1989) also found that GSI could be used successfully to crudely separate maturity stages of Atlantic herring in the north-west Atlantic when macroscopic maturity staging data are not available. Gutted weight and gonad weight are routinely collected for every fish sampled on-board FRV 'Scotia' during HERAS; however, neither are collected on-board RV 'Johan Hjort' during the same survey. It is therefore recommended that all vessels taking part in HERAS collect these data as they are not time consuming to collect and the GSI could be a valuable tool to increase the accuracy of maturity data.

The slopes of the relationship between $\log W_{gt}$ and $\log W_{go}$ for different maturity stages have previously been found to be equal for Atlantic herring (McQuinn, 1989). However, in the present study there was a significant difference between the slopes of the relationship between $\log W_{gt}$ and $\log W_{go}$ across maturity stages which casts doubt on size-independence of the GSI model. It should be noted, however, that there are clear overlaps in the maturity staging data between immature stages and stage VIII, stages III and VIII, and stages III and IV (Figure 3). This overlap may be caused by maturity staging errors as there was a high degree of error with regard to separating maturity stages III and VIII, stages III and IV, and immature stages and stage VIII. This could, therefore, be causing the differences in slope estimates. Therefore, we suggest that if the accurate maturity stages had been assigned to all fish in this analysis, all of the slopes would be likely to be

equal to the slope of the stage IV relationship, as stage IV was the most accurate macroscopic maturity stage (Table 5). We suggest that if the slope of the relationship between $\log W_{gt}$ and $\log W_{go}$ for any maturity stage is the same as it is for stage IV (i.e. equal to 1) then this may be an indication that this maturity stage has low staging error. Furthermore, we suggest that a simple scatterplot of $\log W_{gt}$ and $\log W_{go}$ can give a clear impression of where errors may lie without the need for modelling (Figure 3).

Maturity staging carried out on market samples at Marine Scotland exhibited much higher error than samples staged on-board FRV 'Scotia' by the same reader in 2007. This suggests that market samples are more difficult to stage than those sampled at sea. The market samples used in this study came from commercial vessels and had often been kept on ice for several hours before maturity assessments were carried out. In contrast, samples staged on-board FRV 'Scotia' were very seldom put on ice and are always sampled immediately upon capture. Freezing has been found to increase oocyte diameter and the weight of ovarian tissue in Atlantic cod (Klibansky & Juanes, 2007). Therefore, it may be much more difficult to assess maturity stage if fish are not fresh or have been on ice, possibly due to water uptake or freezing of the gonad tissue. This finding highlights the importance of using fresh samples for maturity staging data that will be used in SSB estimates.

The results of this study show that there are errors in the maturity staging of North Sea herring. The errors lead to under-estimation of the SSB, highlighting the unreliability of these data. The use of an 8-point scale implies a higher resolution of maturity data than can be achieved macroscopically. It is proposed that either more training is given to maturity readers or all countries involved in HERAS should switch to the four-point scale, i.e. immature, mature-active, spawning and spent/recovering. Spawning fish were absent from our analysis but this stage is easily discernible (Bowers & Holliday, 1961). Bromley (2003) has argued for the importance of using one universal key to reduce uncertainty about North Sea plaice (*Pleuronectes platessa*) maturity staging. A four-point scale would provide less fine-scale data but the data collection procedure and training of new technicians may be reduced. This may be a more cost- and time-effective means of gathering maturity staging data. Furthermore, GSI cut-off values can be used to correct macroscopic maturity staging in an objective manner. The GSI model is not only useful for validation of macroscopic maturity staging but would also be valuable in cases where there is no maturity staging data or the maturity staging data available are known to be unreliable.

ACKNOWLEDGEMENTS

REFERENCES

Bowers A.B. and Holliday F.G.T. (1961) Histological changes in the gonad associated with the reproductive cycle of the herring (*Clupea harengus* L.). *Marine Research Department of Agriculture and Fisheries for Scotland* 5, 1–16.

Bromley P.J. (2003) *Progress towards a common gonad grading key for estimating the maturity of North Sea plaice.* Report of the Working Group on Modern Approaches to Assess Maturity and Fecundity of Warm- and Cold-water Fish and Squids, Fisken Hav., No. 12-2003, pp 19–24, 140 pp.

Carrassón M. and Bau M. (2003) Reproduction and gonad histology of *Aidablennius sphinx* (Pisces: Blenniidae) of the Catalan Sea (north-western Mediterranean). *Scientia Marina* 67, 461–469.

Claereboudt M.R., Mcllwain J.L., Al-Oufi H.S. and Ambu-Ali A.A. (2005) Patterns of reproduction and spawning of the kingfish (*Scomberomorus commerson*, Lacépède) in the coastal waters of the Sultanate of Oman. *Fisheries Research* 73, 273–282.

Cohen J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37–46.

DeVlamming V., Grossman G. and Chapman F. (1982) On the use of the gonadosomatic index. *Comparative Biochemistry and Physiology A* 73, 31–39.

Erickson D.L., Hightower J.E. and Grossman G.D. (1985) The relative gonadal index: an alternative index for quantification of reproductive condition. *Comparative Biochemistry and Physiology A* 81, 117–120.

Gerritsen H.D. and McGrath D. (2006) Variability in the assignment of maturity stages of plaice (*Pleuronectes platessa* L.) and whiting (*Merlangius merlangus* L.) using macroscopic maturity criteria. *Fisheries Research* 77, 72–77.

Hunter J.R. and Macewicz B.J. (2003) *Improving the accuracy and precision of reproductive information used in fisheries.* Report of the Working Group on Modern Approaches to Assess Maturity and Fecundity of Warm- and Cold-water Fish and Squids, Fisken Hav., No. 12-2003 57–68, 140 pp.

ICES (2008) *Report of the planning group on herring surveys.* ICES C.M. 2008/LRC: 01, 256 pp.

Kennedy J., Skjæraasen J.E., Nash R.D.M., Thorsen A., Slotte A., Hansen T. and Kjesbu O.S. (2010) Do capital breeders like Atlantic herring (*Clupea harengus*) exhibit sensitive periods of nutritional control on ovary development and fecundity regulation? *Canadian Journal of Fisheries and Aquatic Sciences* 67, 16–27.

Klibansky N. and Juanes F. (2007) Species-specific effects of four preservative treatments on oocytes and ovarian material of Atlantic cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*) and American plaice (*Hippoglossoides platessoides*). *Fishery Bulletin* 105, 538–547.

Koya Y., Soyano K., Yamamoto K., Obana H. and Matsubara T. (2003) Oocyte development and serum profiles of vitellogenin and steroid hormone levels in captive female Pacific herring *Clupea pallasii* during their first maturational cycle. *Fishery Science* 63, 137–145.

La Mesa M., Caputo V., Rampa R. and Vacchi M. (2003) Macroscopic and histological analyses of gonads during the spawning season of *Chionodraco hamatus* (Pisces, Channichthyidae) off Terra Nova Bay, Ross Sea, Southern Ocean. *Polar Biology* 26, 621–628.

Landry J. and McQuinn I.H. (1988) *Guide to microscopic and macroscopic identification of the sexual maturity stages of the Atlantic herring* (*Clupea harengus harengus L.*). Canadian Technical Report of Fisheries and Aquatic Sciences, No. 1655, 71 pp.

Lea E. (1910) Contributions to the methodics in herring-investigations. *Publications de Circonstance* 53, 7–33.

McQuinn I.H. (1989) Identification of spring- and autumn-spawning herring (*Clupea harengus harengus*) using maturity stages assigned from gonadosomatic index model. *Canadian Journal of Fisheries and Aquatic Sciences* 46, 969–980.

Potoschi A., Renones O. and Cannizzaro L. (1999) Sexual development, maturity and reproduction of dolphinfish (*Coryphaena hippurus*) in the western and central Mediterranean. *Scientia Marina* 63, 367–372.

R Development Core Team (2009) *R: a language and environment for statistical computing*. Vienna: R Foundation for Satistical Computing. ISBN 3-900051-07-0, URL http://www.R-project.org.

Saborido-Rey F. and Junquera S. (1998) Histological assessment of variations in sexual maturity of cod (*Gadus morhua* L.) at the Flemish Cap (north-west Atlantic). *ICES Journal of Marine Research* 55, 515–521.

Somarakis S., Ganias K., Tserpes G. and Koutsikopoulos C. (2004) Ovarian allometry and the use of the gonadosomatic index: a case study in the Mediterranean sardine, *Sardina pilchardus*. *Marine Biology* 146, 181–189.

Tomkiewicz J., Tybjerg L. and Jespersen Å. (2003) Micro- and macroscopic characteristics to stage gonadal maturation of female Baltic cod. *Journal of Fish Biology* 62, 253–275.

Tyler C.R. and Sumpter J.P. (1996) Oocyte growth and development in teleosts. *Reviews in Fish Biology and Fisheries* 6, 287–318.

Vitale F., Svedäng H. and Cardinale M. (2006) Histological analysis invalidates macroscopically determined maturity ogives of the Kattegat cod (*Gadus morhua*) and suggests new proxies for estimating maturity status of individual fish. *ICES Journal of Marine Science* 63, 485–492.

West G. (1990) Methods of assessing ovarian development in fishes: a review. *Australian Journal of Marine and Freshwater Research* 41, 199–222.

and

Williams K. (2007) Evaluation of the macroscopic staging method for determining maturity of female walleye pollock *Theragra chalcogramma* in Shelikof Strait, Alaska. *Alaska Fishery Research Bulletin* 12, 252–263.

Correspondence should be addressed to:
L.R. McPherson
University of Aberdeen, Zoology Building
Tillydrone Avenue Aberdeen, AB24 2TZ
email: lindsay.mcpherson@abdn.ac.uk