Results of the age reading exercise for Tobis in the North Sea. March 2015

Readers:

- Tom Svolgaard DTU
- Nina Fuglsang DTU
- Susanne Hansen DTU
- Lisbet Solbakken IMR
- Hildegunn Mjanger IMR
- Inger Henriksen IMR

Samples:

- 98 images
- 2011, 2012 and 2013 Survey samples
- Length classes: 7.5 – 23.5 cm
- Months: April, May, June and August

Methods:

Images of single otoliths (left or right), immersed in water were taken on a black background under reflected light using a standard magnification. These were uploaded to WebGR and a calibration exercise "Tobis Internal Exercise" was made. Initially the exercise was set up as an internal exercise to calibrate readings between the 2 laboratories at DTU. The image set was later made available to the 3 readers from IMR and another reader from DTU. Readers were given clear instructions and an image example showing which axis to annotate. They were asked to annotate the centre point and the start of each translucent zone and give a final estimation of age. The exercise was run as a blind test so the readers could not see the annotations of the other readers. Readers were provided with information on the capture date, area and total length (TL).

Analysis:

Age data:

The Guus Eltink spreadsheet (Eltink, A.T.G.W. 2000) was used to analyse the agreement between readers by means of traditional procedures:

- average % Agreement (nmodal age/ntotal*100)
- coefficient of variation (CV) (Standard deviation/average*100)
- bias plots

Measurement data:

WebGR provides a measure of distance between the annotations made by the readers and thus provides a measure of growth increment width. The "alldistances" dataset from WebGR was used to establish growth curves for each fish for each reader which were then compared by Linear Mixed Effects Models (LMEM's).

Results:

Age data:

The average agreement between readers is 88.7% with a coefficient of variation of 7.5%. For ages 0, 1 and 2 the average agreement between readers is above 90% and CV below 10% and there is clearly no problem with the identification of the first winter ring, see Figure 1. From age 3 to 6 the average agreement decreases to between 75 - 83% and the CV increases from 7 – 17%. The disagreements arise at the edge and this is most likely due to difficulties in interpreting the otolith edge as the otoliths increase in size, see Figure 2. Some of the larger otoliths have an extra point which can also lead to differences as to where the readers are placing their annotations, see Figure 3.

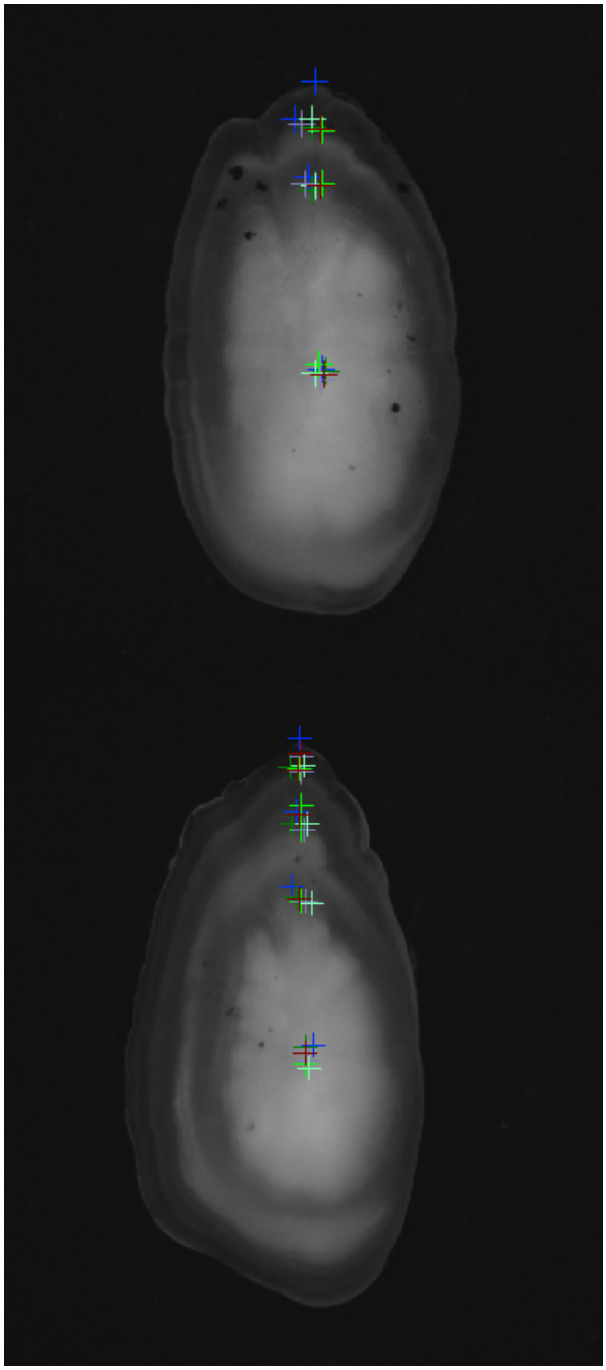


Figure 1. Image 2011_J208_33 where there is 100% agreement between readers, estimating age = 2. Tom (Pale green), Nina (fluorescent green), Lisbet (army green), Hildegunn (purple), Susanne (red) and Inger (blue).

Figure 2. Image 2011_J176_38 where there is 50 % agreement between readers. Tom (Pale green) age=3, Nina (fluorescent green) age=4, Lisbet (army green) age=3, Hildegunn (purple) age=4, Susanne (red) age=4, Inger (blue) age=3

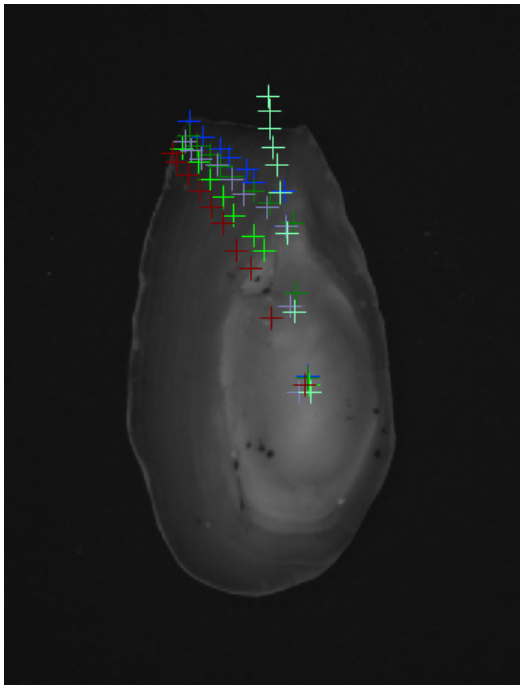


Figure 3. Image 2011_J463_33 where there is only 50% agreement between readers. Tom (Pale green) age=8, Nina (fluorescent green) age=7, Lisbet (army green) age=9, Hildegunn (purple) age=9, Susanne (red) age=9, Inger (blue) age=7

Table 1 shows the age composition based on the estimated ages and the following tables are based on the age data alone. Tables 2, 3 and 4 show the % Agreement, CV and relative bias based on modal age for each of the readers. Table 5 shows the results of the Inter reader and reader against modal age bias tests and indicate that there is a possibility of bias between Lisbet and modal age and a certainty of bias between Inger and modal age. Figure 4 shows the age bias plots for each reader and these confirm the results of the bias tests. Figure 5 shows that the largest proportion of CV is within the 0-10% range and thus overall the CV is low.

Table 1. Age composition (based on estimated ages) and the number of otoliths read

Age	Susanne	Tom	Nina	Lisbet	Hildegunn	Inger	Total
0	1	1	3	3	3	3	**14**
1	31	31	31	31	32	35	**191**
2	32	30	27	30	28	29	**176**
3	8	11	10	13	13	10	**65**
4	12	14	17	12	12	14	**81**
5	6	2	3	2	1	2	**16**
6	3	4	5	3	4	3	**22**
7	1	1	1	-	-	1	**4**
8	-	1	-	-	-	-	**1**
9	1	-	-	1	1	-	**3**
Total	**95**	**95**	**97**	**95**	**94**	**97**	**573**

Table 2. % agreement for each of the readers by modal age

Modal age	Susanne	Tom	Nina	Lisbet	Hildegunn	Inger	All
0	100%	100%	100%	100%	100%	100%	**100%**
1	97%	94%	97%	100%	100%	97%	**97%**
2	100%	93%	86%	100%	97%	83%	**93%**
3	73%	82%	64%	90%	100%	45%	**75%**
4	75%	81%	88%	75%	80%	69%	**78%**
5	100%	100%	100%	100%	100%	0%	**83%**
6	60%	80%	100%	75%	100%	60%	**79%**
7	-	-	-	-	-	-	**-**
8	-	-	-	-	-	-	**-**
9	100%	0%	0%	100%	100%	0%	**50%**
Total							**88.8%**
Weighted average	89.5%	88.4%	87.6%	93.7%	95.7%	78.4%	

Table 3. Co-efficient of Variation (CV) for each of the readers by modal age

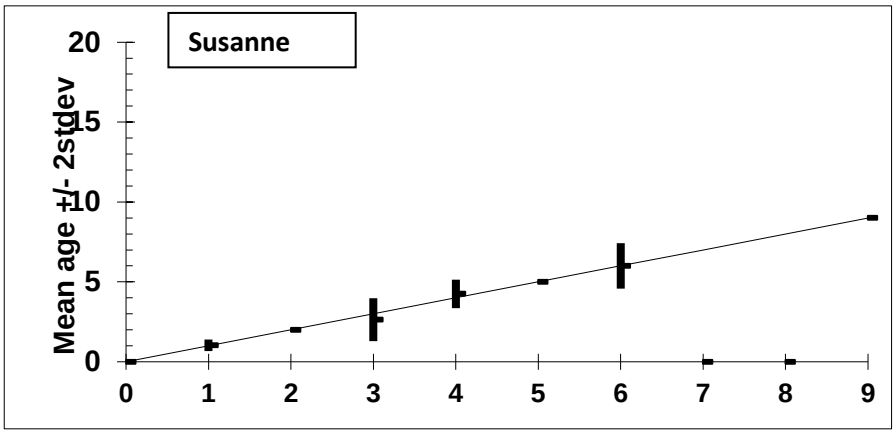
Modal age	Susanne	Tom	Nina	Lisbet	Hildegunn	Inger	All
0	-	-	-	-	-	-	**-**
1	17.4%	23.5%	17.4%	0.0%	0.0%	17.4%	**3.6%**
2	0.0%	13.4%	18.0%	0.0%	9.4%	21.0%	**8.4%**
3	25.6%	14.9%	19.0%	10.9%	0.0%	26.6%	**17.0%**
4	10.5%	11.2%	8.3%	11.9%	10.9%	13.0%	**8.1%**
5	-	-	-	-	-	-	**-**
6	11.8%	7.2%	0.0%	8.7%	0.0%	9.8%	**7.0%**
Total							**7.5%**
Weighted average	11.0%	15.7%	14.4%	3.5%	4.7%	17.5%	

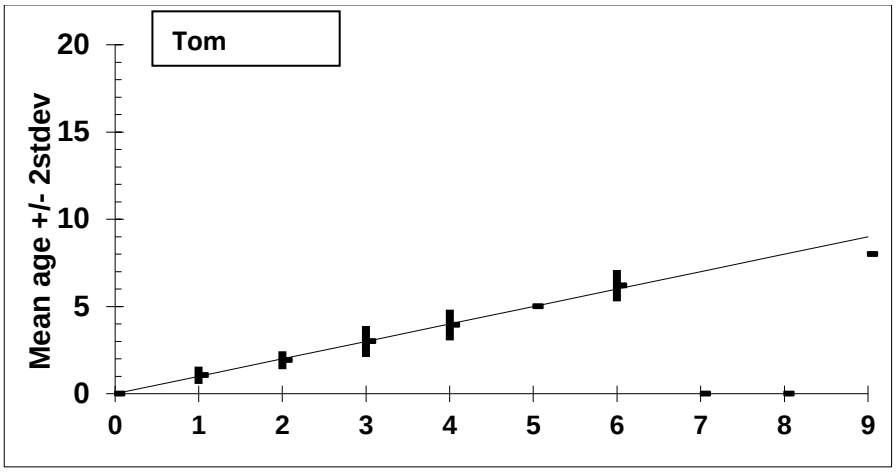
Table 4. Relative Bias for each of the readers by modal age

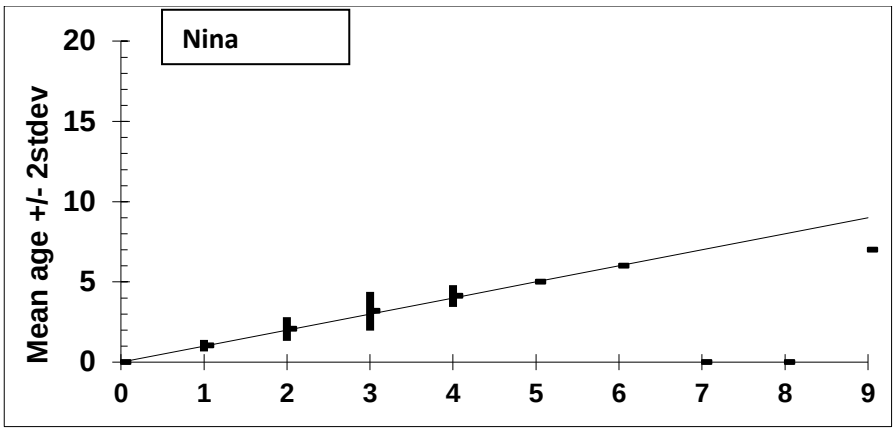
Modal age	Susanne	Tom	Nina	Lisbet	Hildegunn	Inger	All
0	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
1	0.03	0.06	0.03	0.00	0.00	0.03	**0.03**
2	0.00	-0.07	0.07	0.00	-0.03	-0.17	**-0.03**
3	-0.36	0.00	0.18	-0.10	0.00	-0.18	**-0.08**
4	0.25	-0.06	0.13	-0.25	-0.20	-0.31	**-0.07**
5	0.00	0.00	0.00	0.00	0.00	-1.00	**-0.17**
6	0.00	0.20	0.00	-0.25	0.00	-0.40	**-0.07**
7	-	-	-	-	-	-	**-**
8	-	-	-	-	-	-	**-**
9	0.00	-1.00	-2.00	0.00	0.00	-2.00	**-0.83**
Total							**-0.04**
Weighted average	0.01	-0.01	0.05	-0.06	-0.04	-0.16	

Table 5. Inter reader bias test and reader against modal age bias test

	Susanne	Tom	Nina	Lisbet	Hildegunn	Inger
Susanne						
Tom	-					
Nina	-	-				
Lisbet	*	-	*			
Hildegunn	-	-	-	-		
Inger	* *	-	* *	-	*	
Modal age	—	—	—	*	—	* *







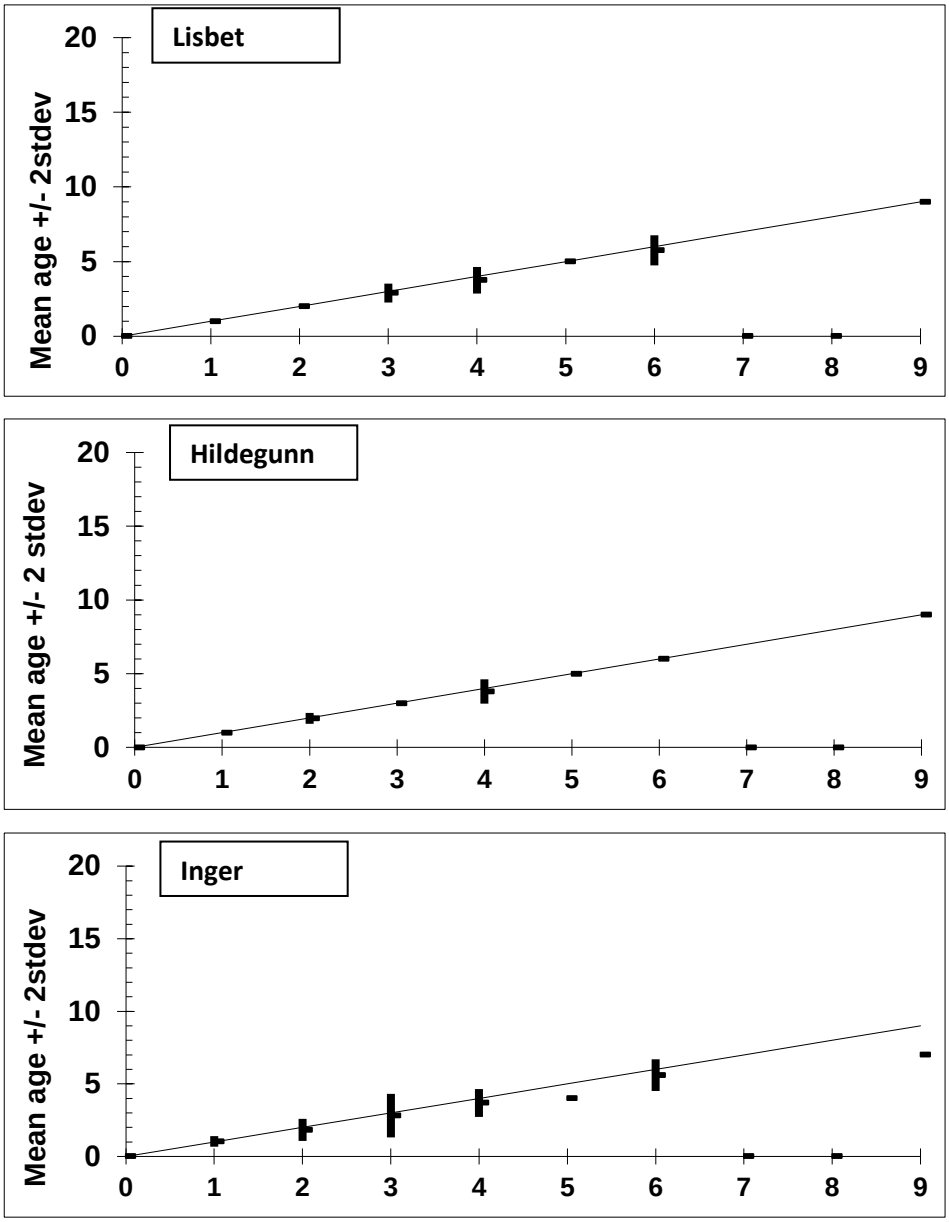


Figure 4 Age bias plots for each reader as mean age ± 2 SD on modal age

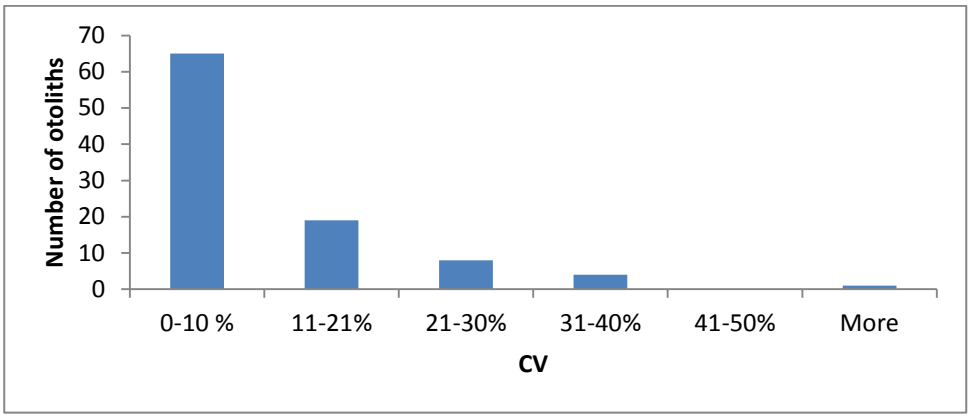


Figure 5 Histogram representing the frequency of CV's within the aged collection

Measurement data:

The analysis of the measurement data supports the results of the age data. Figure 6 shows the average distance from the centre to each winter ring for each reader. For Winter ring 1 and 2 there is no overlap between the coloured boxes which shows that there are clear differences between where the readers identify the first and second winter ring to be. Between Winter ring 2 and 3 there is a slight overlap which increases between the following Winter rings. It is this overlap which shows that the readers do not annotate the same structures for Winter rings closer to the edge. Ingers' measurements stand out from the rest as she did not mark the centre of the otolith, this also meant that her measurements could not be compared directly to the others and thus the certainty of bias which showed up in the age tests is not apparent here. The very large boxes for Hildegunn are due to some of the annotations not being been made in the correct order.

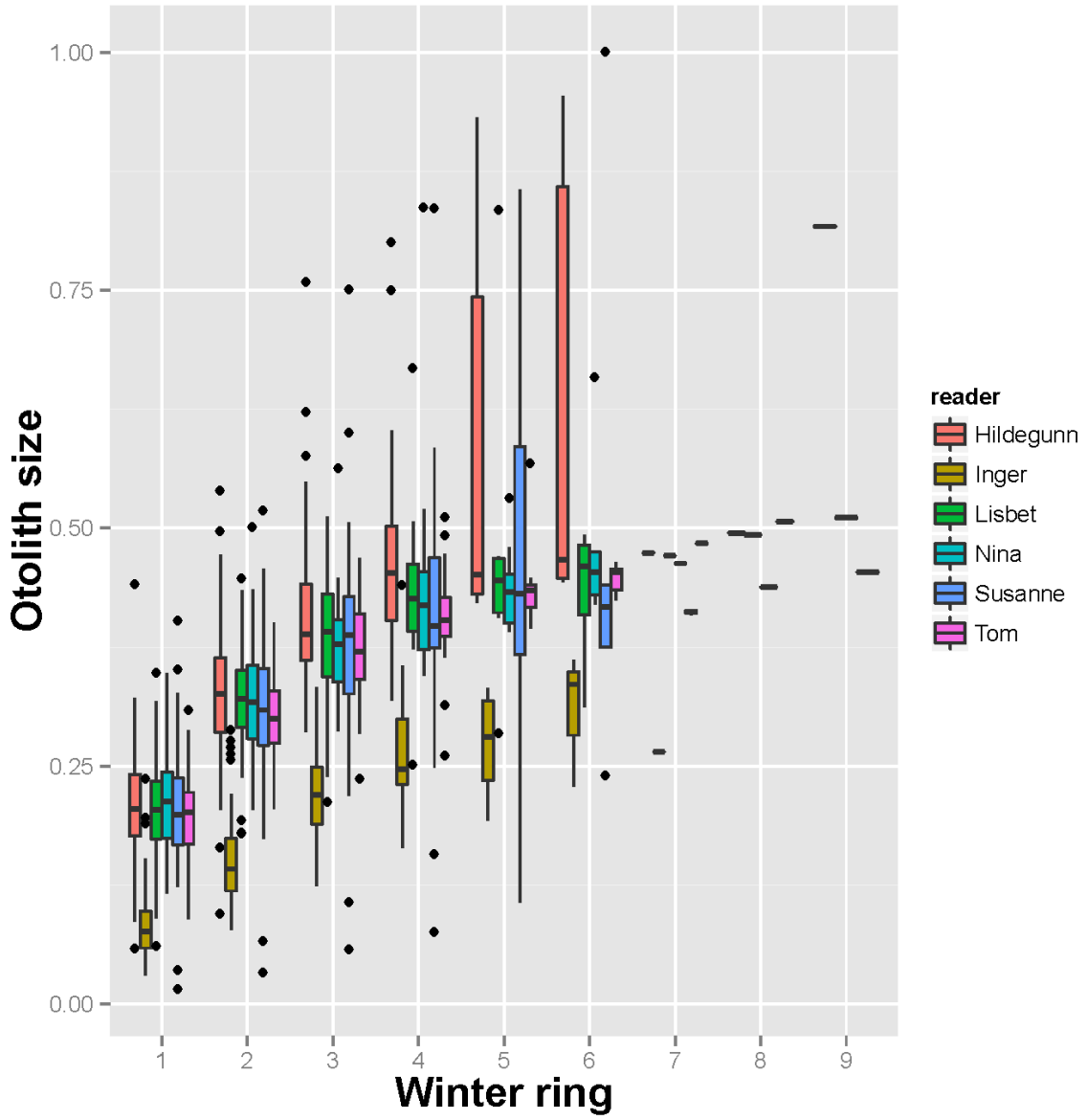


Figure 6. Plot of average distance to the centre for Winter rings 1-9 for all Readers. The boxes represent the mean, upper and lower box boundaries of the interquartile range, whiskers represent the minimum and maximum values and the dots represent the outliers.

Conclusion:

The level of agreement between readers is high and the CV is low. No consistent problems are apparent. The readers are generally reading the same structures so there is no need for huge concern when providing age data to the assessment groups. The disagreements appear with the older individuals where the edges are difficult to see clearly but also with the 3 year old fish. This may be due to the images or likely confusion as to whether to count the edge or not, based on the edge type and time of year. Creating a collection of agreed aged fish is recommended with guidelines on edge interpretation.
The Guus Eltink spreadsheet analyses are based on a modal age. Using the measurement data from WebGR allows for a more precise analysis if the annotations are made correctly. Growth curves and images can be compared to test where the ageing problems and disagreements are.
The exercise has now been made open to all participants so that they can view each other's readings and annotations.

References:

Eltink, A.T.G.W. 2000. Age reading comparisons. (MS Excel workbook version 1.0 October 2000) Internet: http://www.efan.no