How Should The Sardine Length Frequencies Be Weighted?

C.L. de Moor* and D.S. Butterworth*

Length frequencies are calculated for the annual sardine November surveys using samples from trawls undertaken during the spawner biomass acoustic surveys and using a method of calculation that was developed by Roel and probably also Hampton (Coetzee pers. comm.). These length frequencies are used in estimating the acoustic biomass and the proportion-at-age at the time of the survey.

This document briefly considers the current method used to estimate the survey length frequency and develops some alternative methods. The implications of the alternative methods will be discussed in a future document.

Method

The method currently used to calculate the length frequency of the spawner biomass survey is detailed in Appendix A. In essence, the length frequencies from the individual trawls are weighted by the acoustic weighting associated by that trawl as a proportion of the total of the acoustic weighting of all trawls. A new method, termed "alternative method" for the sake of this document, has been developed with the following in mind (see Appendix B):

i) the logic behind calculating the weightings (equations A.2 to A.4) in the current method is unclear;

ii) the individual trawls and their proportions-at-length are weighted using the acoustic density in the manner assumed to underlie the current method; and

iii) a method with logical steps was required, in which the weighting factor applied to each trawl was clear and easily modified given changed assumptions.

Using the alternative method, a comparison is made of the November survey proportions-at-age (see Appendix C) calculated using:

I) the length frequency for the full survey area together with the age-length-key calculated using samples from the full survey area;

II) the length frequency together with the age-length-key using data up to Port Alfred only; and

III) the length frequency using data up to Port Alfred only together with the age-length-key calculated using samples for the full survey area.

Option III) is what is currently being used, while option II) would appear to be the best choice for use in assessments to correspond with the use of survey biomass up to Port Alfred only.

* MARAM (Marine Resource Assessment and Management Group), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701, South Africa. Email: c.l.demoor@telkomsa.net, doug.butterworth@uct.ac.za.

The standard deviation and CVs of the estimated proportions-at-age and proportions-at-length are calculated from 1000 bootstrapped iterations of option II) above (see Appendix D). This is done only for years in which survey age-length keys were available (1993, 1994, 1996, 2001, 2002, 2003, 2004, 2006), since the age-length-key using commercial data up to Port Alfred only was not easily obtainable.

Results

Comparing the Alternative Method to the Current Method

The current and alternative methods both result in the same total proportion-at-length for the survey (results not shown). However, the numbers-at-length (or the "length frequency") differ by a constant per stratum between the two methods. Comparing equations (A.4) and (B.8), it is not possible to find an analytical solution to this constant per stratum. Substituting in, one can see that from equation (A.4) we have:

$$N_{sl} = \frac{T_{sl} \cdot B_s}{\sum_j Z_{sj}} = \left(\frac{B_s}{\sum_j Z_{sj}} \right) \sum_j \left(T_{sjl} \cdot \frac{Z_{sj}}{W_{sj}} \right)$$

while equation (B.8) gives:

$$N_{sl} = p_{sl} \cdot N_s = \left(\sum_j p_{sjl}^{trawl} \cdot \frac{N_{sj}}{N_s} \right) \cdot N_s = \sum_j p_{sjl}^{trawl} N_{sj} =$$

$$= \sum_j \left(\frac{T_{sjl}}{\sum_l T_{sjl}} \cdot k_s \frac{Z_{sj}}{\overline{W}_{sj}} \right) = \sum_j \left(\frac{T_{sjl}}{\sum_l T_{sjl}} \cdot \frac{B_s}{\sum_j Z_{sj}} \cdot Z_{sj} \cdot \frac{T_{sj}}{W_{sj}} \right) = \left(\frac{B_s}{\sum_j Z_{sj}} \right) \sum_j \left(T_{sjl} \cdot \frac{Z_{sj}}{W_{sj}} \cdot \frac{T_{sj}}{\sum_l T_{sjl}} \right)$$

Thus, the difference arises from the inclusion of $p_{sjl}^{trawl} = \frac{T_{sjl}}{\sum_l T_{sjl}}$ in the alternative method.

Comparing the proportion-at-age using data from different parts of the survey area

Tables 1a to 1k show the proportion-at-age calculated using the three options of combining the length frequency and age-length-key. There is no difference in the proportions-at-age calculated using methods I) and III) and the difference between methods I) and II) is minor.

Discussion

Although the past 'mismatch' between the survey area used for the length frequency and age-length-key used to calculate the November survey proportions-at-age (option III)) has not proved to result in a large difference in the results had data up to Port Alfred only been used, we recommend that for self-consistency, the November survey proportions-at-age be calculated using the length frequency together with the age-length-key using data up to Port Alfred only.

Way Forward

In order to test the effect of weighting the trawls independently of the acoustic weighting, two further methods will be tested:

i) an "equal weighting" method in which all trawls are weighted equally. In this case, the weighting factor becomes $w_{sj} = \frac{1}{n_s}$ where n_s is the number of trawls in the stratum and therefore equation (B.7) changes to:

$$p_{sl} = \sum_j p_{sjl}^{trawl} \cdot \frac{1}{n_s}$$

ii) a "down weighted" method in which small trawl sub-samples are downweighted. Although scientists aim to sample 100 fish from each trawl (Coetzee pers comm.), the actual number of fish sub-sampled can vary considerably (Figure 1). As 100 fish is currently used as a 'cut-off' when sampling, the weighting factor remains $w_{sj} = \frac{1}{n_s}$ if $T_{s,j} > 100$, but is $w_{sj} = \frac{T_{s,j}}{100 * n_s}$ if $T_{s,j} \leq 100$.

The median and standard deviations around the proportions-at-age and proportions-at-length from these two further methods will be compared to the alternative method to investigate the effect the chosen method of weighting has on the median proportions, and in particular the standard deviations.

Acknowledgements

We thank Janet Coetzee for providing the current method used to weight the length frequencies and for comments on earlier parts of this work.

References

Cunningham, CL and Butterworth, DS 2007. Base Case Assessment of the South African Sardine Resource. MCM Document MCM/2007/SEP/SWG-PEL/06. 30pp.

Table 1a. The proportions-at-age for 1993 calculated using the alternative method, together with the bootstrapped averages, medians, and CVs[1].

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.09	0.46	0.23	0.17	0.05	0.01	0.00	0.00	0.00
II) Port Alfred	0.00	0.09	0.46	0.23	0.17	0.05	0.01	0.00	0.00	0.00
III) "Mismatch"	0.00	0.09	0.46	0.23	0.17	0.05	0.01	0.00	0.00	0.00
Bootstrapping										
Average	0	0.13	0.56	0.16	0.11	0.04	0.00	0.00	0.00	0.00
Median	0	0.11	0.50	0.20	0.14	0.04	0.00	0.00	0.00	0.00
SD	0	0.042	0.140	0.087	0.062	0.024	0.003	0.000	0.00	0.00
CV	N/A	0.33	0.25	0.55	0.55	0.65	0.70	N/A	N/A	N/A

[1] The bootstrapped CV is calculated as (Boostrapped Average)/(Bootstrapped SD).

Table 1b. The proportions-at-age for 1994 calculated using the alternative method, together with the bootstrapped averages, medians, and CVs.

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.02	0.26	0.19	0.25	0.15	0.09	0.02	0.01	0.00
II) Port Alfred	0.00	0.02	0.26	0.19	0.25	0.15	0.09	0.02	0.01	0.00
III) "Mismatch"	0.00	0.02	0.26	0.19	0.25	0.15	0.09	0.02	0.01	0.00
Bootstrapping										
Average	0	0.01	0.14	0.09	0.27	0.27	0.16	0.05	0.00	0.00
Median	0	0.01	0.11	0.08	0.29	0.27	0.16	0.05	0.00	0.00
SD	0	0.011	0.126	0.079	0.038	0.094	0.058	0.022	0.003	0.002
CV	N/A	0.96	0.91	0.86	0.14	0.35	0.36	0.43	1.02	0.99

Table 1c. The proportions-at-age for 1996 calculated using the alternative method, together with the bootstrapped averages, medians, and CVs.

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.15	0.34	0.28	0.14	0.07	0.01	0.00	0.00	0.00
II) Port Alfred	0.00	0.15	0.34	0.28	0.14	0.07	0.01	0.00	0.00	0.00
III) "Mismatch"	0.00	0.15	0.34	0.28	0.14	0.07	0.01	0.00	0.00	0.00
Bootstrapping										
Average	0	0.15	0.36	0.26	0.14	0.07	0.01	0.01	0	0
Median	0	0.14	0.36	0.26	0.14	0.07	0.01	0.01	0	0
SD	0	0.078	0.080	0.048	0.045	0.024	0.006	0.010	0	0
CV	N/A	0.52	0.22	0.19	0.32	0.33	0.49	1.10	N/A	N/A

Table 1d. The proportions-at-age for 2000, using an age-length-key from November commercial data, calculated using the alternative method, together with the bootstrapped averages, medians, and CVs.

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.17	0.44	0.14	0.13	0.08	0.03	0.00	0.00	0.00
II) Port Alfred										
III) "Mismatch"	0.00	0.17	0.44	0.14	0.13	0.08	0.03	0.00	0.00	0.00

Table 1e. The proportions-at-age for 2001 calculated using the alternative method, together with the bootstrapped averages, medians, and CVs.

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.53	0.34	0.04	0.05	0.03	0.01	0.00	0.00	0.00
II) Port Alfred	0.00	0.53	0.34	0.04	0.05	0.03	0.01	0.00	0.00	0.00
III) "Mismatch"	0.00	0.53	0.34	0.04	0.05	0.03	0.01	0.00	0.00	0.00
Bootstrapping										
Average	0	0.63	0.28	0.02	0.04	0.02	0.01	0.00	0.00	0
Median	0	0.64	0.28	0.02	0.03	0.02	0.01	0.00	0.00	0
SD	0	0.109	0.060	0.017	0.026	0.013	0.004	0.002	0.001	0
CV	N/A	0.17	0.21	0.68	0.69	0.77	0.72	0.70	0.71	N/A

Table 1f. The proportions-at-age for 2002 calculated using the alternative method, together with the bootstrapped averages, medians, and CVs.

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.13	0.28	0.23	0.19	0.10	0.05	0.01	0.00	0.00
II) Port Alfred	0.00	0.13	0.28	0.23	0.19	0.10	0.05	0.01	0.00	0.00
III) "Mismatch"	0.00	0.13	0.28	0.23	0.19	0.10	0.05	0.01	0.00	0.00
Bootstrapping										
Average	0	0.19	0.38	0.17	0.14	0.08	0.03	0.01	0.00	0.00
Median	0	0.18	0.39	0.17	0.13	0.07	0.03	0.01	0.00	0.00
SD	0	0.050	0.099	0.042	0.046	0.029	0.013	0.004	0.001	0.001
CV	N/A	0.27	0.26	0.25	0.34	0.37	0.40	0.39	0.50	0.54

Table 1g. The proportions-at-age for 2002, using an age-length-key from November commercial data, calculated using the alternative method, together with the bootstrapped averages, medians, and CVs.

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.27	0.27	0.22	0.12	0.10	0.02	0.01	0.00	0.00
II) Port Alfred										
III) "Mismatch"	0.00	0.27	0.27	0.22	0.12	0.10	0.02	0.01	0.00	0.00

Table 1h. The proportions-at-age for 2003 calculated using the alternative method, together with the bootstrapped averages, medians, and CVs.

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.46	0.17	0.16	0.13	0.06	0.02	0.00	0.01	0.00
II) Port Alfred	0.00	0.45	0.17	0.16	0.13	0.06	0.02	0.00	0.01	0.00
III) "Mismatch"	0.00	0.46	0.17	0.16	0.13	0.06	0.02	0.00	0.01	0.00
Bootstrapping										
Average	0	0.56	0.30	0.06	0.05	0.01	0.02	0.00	0.00	0
Median	0	0.58	0.31	0.05	0.04	0.01	0.01	0.00	0.00	0
SD	0	0.068	0.045	0.046	0.045	0.010	0.014	0.000	0.000	0
CV	N/A	0.12	0.15	0.75	0.86	0.89	0.95	N/A	0.79	N/A

Table 1i. The proportions-at-age for 2004 calculated using the alternative method, together with the bootstrapped averages, medians, and CVs.

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.33	0.15	0.23	0.16	0.07	0.05	0.01	0.01	0.00
II) Port Alfred	0.00	0.33	0.15	0.23	0.16	0.07	0.05	0.01	0.01	0.00
III) "Mismatch"	0.00	0.33	0.15	0.23	0.16	0.07	0.05	0.01	0.01	0.00
Bootstrapping										
Average	0	0.26	0.16	0.25	0.18	0.08	0.06	0.01	0.00	0
Median	0	0.25	0.16	0.25	0.18	0.08	0.06	0.01	0.00	0
SD	0	0.100	0.042	0.035	0.042	0.020	0.021	0.004	0.002	0
CV	N/A	0.39	0.26	0.14	0.23	0.26	0.33	0.50	0.50	N/A

Table 1j. The proportions-at-age for 2005 calculated using the alternative method, together with the bootstrapped averages, medians, and CVs.

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.28	0.23	0.13	0.16	0.11	0.06	0.02	0.01	0.00
II) Port Alfred										
III) "Mismatch"	0.00	0.28	0.23	0.13	0.16	0.11	0.06	0.02	0.01	0.00

Table 1k. The proportions-at-age for 2006, using an age-length-key from November commercial data, calculated using the alternative method, together with the bootstrapped averages, medians, and CVs.

Age	0	1	2	3	4	5	6	7	8	9
I) Full Area	0.00	0.65	0.25	0.02	0.03	0.03	0.01	0.00	0.00	0.00
II) Port Alfred	0.00	0.65	0.25	0.02	0.03	0.03	0.01	0.00	0.00	0.00
III) "Mismatch"	0.00	0.65	0.25	0.02	0.03	0.03	0.01	0.00	0.00	0.00
Bootstrapping										
Average	0	0.61	0.26	0.05	0.03	0.03	0.01	0.01	0	0
Median	0	0.64	0.22	0.03	0.03	0.03	0.01	0.00	0	0
SD	0	0.178	0.139	0.045	0.026	0.028	0.010	0.011	0	0
CV	N/A	0.29	0.53	0.99	0.79	0.81	1.06	1.64	N/A	N/A

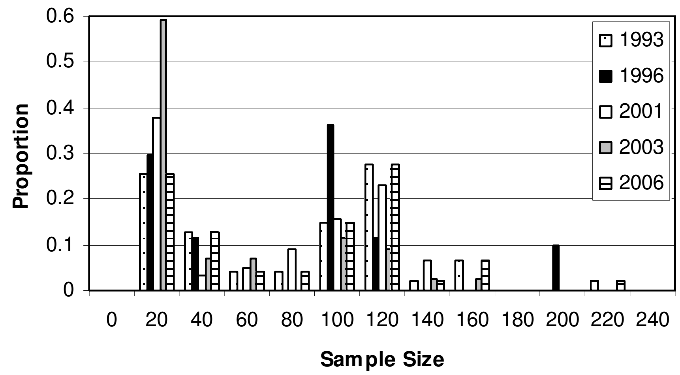

Figure 1. *The histogram of the number of fish in each trawl sub-sample for selected years.*

Appendix A: The Current Method Used to Calculate the Weighted Length Frequencies

The acoustic weighting for each trawl sample *j* in stratum *s* is given by:

$$Z_{sj} = \sum_i L_{sji} \cdot \rho_{sji} \tag{A.1}$$

where

L_{sji} is the mean interval length (nmi) for trawl sample *j* and interval (ESDU) *i* in stratum *s*, and

ρ_{sji} is the mean acoustic interval density (g.m^{-2}) for trawl sample *j* and interval (ESDU) *i* in stratum *s*.

To weigh individual trawls, one needs to convert the acoustic weighting factor into a factor in terms of numbers. The trawl weighting factor is given by:

$$Q_{sj} = \frac{Z_{sj}}{W_{sj}} \tag{A.2}$$

where

W_{sj} is the total mass (kg) (also termed length frequency mass) of all the fish in the sub-sample from trawl *j* in stratum *s*.

The weighted length frequency in stratum *s* is the vector $\underline{\mathbf{T}}_s$, which has elements

$$T_{sl} = \sum_j T_{sjl} \cdot Q_{sj} \tag{A.3}$$

where

T_{sjl} is the number of fish of length *l* in the sub-sample from trawl *j* in stratum *s*.

The total number of fish of length *l* estimated by the survey to be in stratum *s* is then given by:

$$N_{sl} = \frac{T_{sl} \cdot B_s}{\sum_j Z_{sj}} \tag{A.4}$$

where

B_s is the biomass (kg) in stratum *s*.

The total number of fish of length *l* estimated by the survey is then given by:

$$N_l^{tot} = \sum_s N_{sl} \tag{A.5}$$

Appendix B: The Alternative Method to Calculate the Weighted Length Frequencies

The acoustic weighting for each trawl sample *j* in stratum *s* is given by:

$$Z_{sj} = \sum_i L_{sji} \cdot \rho_{sji} \tag{B.1}$$

where

L_{sji} is the mean interval length (nmi) for trawl sample *j* and interval (ESDU) *i* in stratum *s*, and

ρ_{sji} is the mean acoustic interval density (g.m^{-2}) for trawl sample *j* and interval (ESDU) *i* in stratum *s*.

From the total number of fish of all species sampled from a trawl, a sub-sample of a species of interest is drawn for the purposes of estimating length distribution. The total number of fish in this sub-sample from trawl *j* in stratum *s* is given by:

$$T_{sj} = \sum_l T_{sjl} \tag{B.2}$$

where

T_{sjl} is the number of fish of length *l* in the sub-sample from trawl *j* in stratum *s*.

The mean mass (kg) of an individual fish in the sub-sample from trawl *j* in stratum *s* is given by:

$$\overline{W}_{sj} = \frac{W_{sj}}{T_{sj}} \tag{B.3}$$

where

W_{sj} is the total mass (kg) of all the fish in the sub-sample from trawl *j* in stratum *s*.

The number of fish estimated by the survey to be in stratum *s* that are associated with trawl *j* is then given by:

$$N_{sj} = \frac{k_s \cdot Z_{sj}}{\overline{W}_{sj}} \tag{B.4}$$

where $k_s = \dfrac{B_s}{\sum_j Z_{sj}}$ is the conversion factor (in kgs) which relates the estimate of biomass for the stratum, B_s (NB: in kgs) to the sum of the acoustic weights.

The total number of fish estimated by the survey to be in stratum *s* is then calculated as follows:

$$N_s = \sum_j N_{sj} = k_s \sum_j \frac{Z_{sj}}{\overline{W}_{sj}} \tag{B.5}$$

The proportion of fish of length *l* in the sub-sample from trawl *j* in stratum *s* is then given by:

$$p_{sjl}^{trawl} = \frac{T_{sjl}}{\sum_{l} T_{sjl}} \tag{B.6}$$

and the proportion of all the fish of length l in stratum s is then estimated as:

$$p_{sl} = \sum_{j} p_{sjl}^{trawl} \cdot \frac{N_{sj}}{N_{s}} \tag{B.7}$$

In this context, therefore, the proportion of fish of length l in stratum s is estimated from the proportion of fish of length l in the sub-samples taken from each trawl j in the stratum, weighted by the proportion of the total number of fish estimated to be in that stratum which is associated with trawl j, i.e. $w_{sj} = \frac{N_{sj}}{N_{s}}$.

The estimated number of fish of length l in stratum s is therefore

$$N_{sl} = p_{sl} \cdot N_{s} = \left(\sum_{j} p_{sjl}^{trawl} \cdot \frac{N_{sj}}{N_{s}} \right) \cdot N_{s} = \sum_{j} p_{sjl}^{trawl} N_{sj} \tag{B.8}$$

and the estimated total number of fish of length l in the complete area surveyed is given by:

$$N_{l}^{tot} = \sum_{s} N_{sl} \tag{B.9}$$

Appendix C: The Calculation of the Proportion-at-Age in the Survey, given the Length Frequency

The estimated total number of fish of age *a* in length class *l* in the complete area surveyed is given by:

$$N_{la} = \frac{n_{la}}{\sum_a n_{la}} N_l^{tot} \tag{C.1}$$

where

n_{la} is the number of otoliths in length class *l* allocated to age *a*.

The estimated proportion by age of the total number in the complete area surveyed is then given by

$$p_a = \frac{\sum_l N_{la}}{\sum_a \sum_l N_{la}}. \tag{C.2}$$

Appendix D: Bootstrapping

For $b = 1,\ldots,1000$ resample with replacement the sets of data ($Z_{sj}^b, W_{sj}^b, T_{sjl}^b \; {}_{l=4,\ldots,23}$) as used in equation (C.1), (C.2) and (C.3) above[2].

As before, the total number of fish in this sub-sample from trawl *j* in stratum *s* is given by:

$$T_{sj}^b = \sum_l T_{sjl}^b \qquad \text{(D.1)}$$

The mean mass (kg) of an individual fish in the sub-sample from trawl *j* in stratum *s* is given by:

$$\overline{W}_{sj}^b = \frac{W_{sj}^b}{T_{sj}^b} \qquad \text{(D.2)}$$

The number of fish estimated by the survey to be in stratum *s* that are associated with trawl *j* is then given by:

$$N_{sj}^b = \frac{k_s^b \cdot Z_{sj}^b}{\overline{W}_{sj}^b} \qquad \text{(D.3)}$$

where $k_s^b = \dfrac{B_s}{\sum_j Z_{sj}^b}$ is the conversion factor (in kgs) which relates the estimate of biomass for the stratum, B_s (NB: in kgs) to the sum of the acoustic weights.

The total number of fish estimated by the survey to be in stratum *s* is then calculated as follows:

$$N_s^b = \sum_j N_{sj}^b = k_s^b \sum_j \frac{Z_{sj}^b}{\overline{W}_{sj}^b} \qquad \text{(D.4)}$$

The proportion of fish of length *l* in the sub-sample from trawl *j* in stratum *s* is then given by:

$$p_{sjl}^{trawl,b} = \frac{T_{sjl}^b}{\sum_l T_{sjl}^b} \qquad \text{(D.5)}$$

and the proportion of all the fish of length *l* in stratum *s* is then estimated as:

$$p_{sl}^b = \sum_j p_{sjl}^{trawl,b} \cdot \frac{N_{sj}^b}{N_s^b} \qquad \text{(D.6)}$$

The estimated number of fish of length *l* in stratum *s* is therefore

$$N_{sl}^b = p_{sl}^b \cdot N_s^b \qquad \text{(D.7)}$$

and the estimated total number of fish of length *l* in the complete area surveyed is given by:

$$N_l^{tot,b} = \sum_s N_{sl}^b \qquad \text{(D.8)}$$

The estimated total number of fish of age *a* in length class *l* in the complete area surveyed is given by:

[2] Thus if there were originally *n* trawls in stratum *s*, then *j*=1…*n* sets of data are resampled with replacement from the trawls of stratum *s*.

$$N_{la}^{b} = \frac{n_{la}}{\sum_{a} n_{la}} N_{l}^{tot,b} \qquad \text{(D.9)}$$

and the estimated proportion by age of the total number in the complete area surveyed is then given by

$$p_{a}^{b} = \frac{\sum_{l} N_{la}^{b}}{\sum_{a} \sum_{l} N_{la}^{b}} \, . \qquad \text{(D.10)}$$