NOT TO BE CITED WITHOUT PRIOR
REFERENCE TO THE AUTHOR(S)

Northwest Atlantic Fisheries Organization

Serial No. N5843 **NAFO SCR Doc. 10/052**

NAFO/ICES *PANDALUS* ASSESSMENT GROUP—OCTOBER 2010

Calculating length- and sex-specific biomass in the West Greenland trawl survey

by

Michael C. S. Kingsley and Nikoline Ziemer

Pinngortitaleriffik, Greenland Institute of Natural Resources
Box 570, DK-3900 Nuuk, Greenland

Abstract

The West Greenland research trawl survey provides biomass and demographic information on the stock of northern shrimp for use in the annual assessment. Cod-end samples that are weighed, and the individual shrimps in which are then sexed and measured, provide information on numbers at length, by sex, in the total stock. Applying a weight-length curve to the data on numbers allows sex- and length-specific stock biomasses to be calculated. Weight-length curves are usually got by weighing individual shrimps. But unless the weight-length curve is appropriate to the relationship between the weights of the cod-end samples and their length compositions, discrepancies arise between the sum of sex- or length-specific biomasses and the total survey biomass estimate. It is proposed that a suitable weight-length curve for use in partitioning the stock biomass can be calculated directly from the cod-end sample data, without the need to refer to a separate data set of individual weights. In a test analysis of 2009 data this method produced only a small discrepancy between the sum of sex- or length-specific biomasses and the total survey biomass estimate.

Introduction

The West Greenland stock of northern shrimp *Pandalus borealis* is assessed annually. A source of information for the assessment is an annual stratified semi-systematic trawl survey that covers the offshore shrimp fishing grounds on the West Greenland shelf out to 600 m depth, and includes Disko Bay, considered an 'inshore' area. The catch weights in the survey—carried out with a 20-mm liner in the cod-end—provide an estimate of the total biomass; the sexing and measuring of a sample of the shrimps caught at each station also provides estimates of the sex and length distribution of the catch. The conversion of these distributions by numbers into the corresponding weight distributions, so that biomasses of females and of males, and an estimate of the 'fishable' biomass—at least 17 mm CPL—can be calculated, is not always without problems. With the methods in use now, it often happens that there are discrepancies of a few percent, sometimes up to nearly 10%, between the total survey biomass and the sum of calculated component biomasses. NAFO Scientific Council recommended in 2009 that '*improvements in the estimation of weight-length relationships, and their use in estimating sex-specific biomasses, should be investigated*'. This document presents the methods now in use, makes some suggestions as to the reason why discrepancies in weight occur, and suggests some options for avoiding them.

Estimation of Parameters

When the trawl catch is taken on board, a sample is taken from the cod-end to be measured. It is sorted, and among other measurements, the weight of *P. borealis* is recorded. All the *P. borealis* in the sample are measured (the 'Length Sample'). Some are selected for individual weighing (the 'Length-Weight sample'). The rest of the catch is also sorted and weighed (Wieland et al. 2004)

The catch weight of shrimps, and the haul swept area, are used to estimate the overall biomass. Together with the lengths measured in the Length Sample and the sample weight of commercial shrimps, they are used to estimate numbers in the stock by length class and sex. Together with a weight-length curve, the length- and sex-specific numbers can be grossed up ('raised') to length- or sex-specific biomasses.

Methods

Survey strata are considered here as independent sampling problems. Survey totals are got by summing stratum results. 'Length class' can be generalised to include sex or sex-length class.

There are two methods readily available for calculating stratum biomass from station catches. From the catch and swept area at a station, the single-station estimate of stratum biomass for station *s* in stratum *t* is

$$\hat{B}_{ts} = A_t \frac{Catch_{ts}}{Sw.Area_{ts}}$$

where A_t is the stratum area. These single-station estimates of density can be simply averaged, or a mean weighted by swept area could be used. An unweighted analysis estimate is:

$$\hat{B}_t = \frac{A_t}{K_t} \sum_s \frac{Catch_{ts}}{Sw.Area_{ts}}$$

where *t* is the stratum and *s* is the station, of which there are K_t in stratum *t*. Shrimp density does not vary much within a haul's distance (Kingsley et al. 2002) and so it is statistically preferable to use the unweighted estimate of stratum biomass; it also makes the calculations simpler throughout.

If the number of shrimps in class *l* in the Length Sample, of weight $Samp.Wt_{ts}$, from station *s* in stratum *t* is n_{lts}, the corresponding single-station estimate of the number of shrimps in the class in the stratum is

$$\hat{N}_{lts} = A_t \frac{n_{lts}}{Samp.Wt_{ts}} \cdot \frac{Catch_{ts}}{Sw.Area_{ts}}$$

Using the unweighted analysis, the stratum estimate from many stations would be

$$\hat{N}_{lt} = \frac{A_t}{K_t} \sum_s \left(\frac{n_{lts}}{Samp.Wt_{ts}} \cdot \frac{Catch_{ts}}{Sw.Area_{ts}} \right)$$

where the divisor K_t, the number of stations, includes those with no catch[1]. The coefficient $\frac{Catch_{ts}}{Samp.Wt_{ts} \cdot Sw.Area_{ts}}$ common to all lengths counted in a Length Sample can be called the 'raising factor' for a station.

[1] I assume that a Length Sample of shrimps is taken from every catch that contains any commercial shrimps.

Given stratum areas, catches and haul swept areas, estimates of numbers depend on length frequencies in the Length Samples and on the Length-Sample weights. Given these estimates of numbers, and if estimates of individual weight at length are available, length-class biomass by stratum is given by

$$\hat{W}_{lt} = w(l)\frac{A_t}{K_t}\sum_s\left(\frac{n_{lts}}{Samp.Wt_{ts}}\cdot\frac{Catch_{ts}}{Sw.Area_{ts}}\right)$$

and for the survey as a whole by

$$\hat{W}_l = \sum_t w(l)\frac{A_t}{K_t}\sum_s\left(\frac{n_{lts}}{Samp.Wt_{ts}}\cdot\frac{Catch_{ts}}{Sw.Area_{ts}}\right)$$

It is therefore also possible to calculate

$$\hat{W}_l = w(l)\hat{N}_l$$

The weight-length function usually used at present is a very simple power function:

$$w(l) = a\cdot l^b$$

where *a* can be called the coefficient and *b* can be called the exponent. In terms of overall survey results for length- or sex-specific biomasses the coefficient governs the overall sum of class biomasses and the exponent governs their ratio. Increasing the coefficient will increase the sum of class biomasses. Increasing the exponent will estimate greater biomass of large shrimps relative to the biomass of small ones.

If, for all Length Samples $\sum_l n_{lts}w(l) = Samp.Wt_{ts}$, i.e. if the weight of every Length Sample were to answer exactly to the number and size of the shrimps that composed it and to a common function for weight at length, the sum of length-class biomasses would equal the stratum total biomass calculated from catches and swept areas. This doesn't happen. Length Samples deviate from the weight predicted by the number and size of their shrimps and a standard weight-length curve. For one thing, there are errors in data collection; for another, there is no reason to suppose that a single weight-length relationship is valid over the entire size range for the whole survey area; for a third, some Length Samples comprise bigger shrimps, some smaller, and the very simple weight-length functions used as averages over the whole size range probably do not apply exactly to smaller segments of the size range. In general, there will exist no simple parametric weight-length relationship that will ensure that the equation holds for all the Length Samples. Furthermore, it is common practice to use a weight-length function that has little connection with the survey data being analysed—for example, one that is based on data collected several years earlier. As a result, calculated component biomasses—for example, those of the different sex classes, or the fishable and non-fishable biomass—commonly do not add up to the total survey biomass estimate.

This is because the three pieces of information that presently go into the calculation of component biomass are incompatible. They are: the weights of the Length Samples; the length frequencies in the Length Samples; and the weight-length curve calculated from the, or a, or several, Length-Weight Sample(s). There might be several options for resolving the inconsistency.

Option 1: Replace the Length-Sample Weights.

With a weight-length relationship available, and the measurement of every shrimp in the Length Sample, it would be easy to replace the recorded weight of each Length Sample by a weight synthesised from the estimated weights of the shrimps composing it. This synthesized weight could then be used instead of the recorded weight in weighting up length classes when calculating class-specific biomasses. The result would be exact agreement between the sum of class biomasses and the survey total.

The statistical model underlying this option would be that both parameters of the weight-length function were reliable, and that the counts and length measurements in the Length Samples were also reliable, but that the Length-Sample weights were unreliable. Possible reasons might be differences in the time allowed for draining off excess water, errors in the allowance made for *Pandalus* shrimps that could not be measured, or for shrimps of other species, &c. Estimates of class-specific and total *numbers* would differ from those calculated using the weighed Length-Sample weights. The total estimated survey biomass would be divided between large and small shrimps as indicated by the slope of the weight-length function.

Table: Results of replacing weighed Length-Sample weights by estimates derived from weight-length curves.

	Numbers (× 10^9)			Biomass %[1]	
	Males	Females	Shrimps < 14 mm CL	Female	Fishable
Weighed weights	43.9263	12.8119	5.3412		
Wt-Len from Length Samples	43.7042	12.8074	5.3204	37.51	91.90
Wt-Len from Wt-Len samples	43.4153	12.6828	5.2996	37.69	92.01
Standard Wt-Len function	45.6127	13.2368	5.5994	38.09	92.25
Std fun., corrected	44.1242	12.8048	5.4167	38.09	92.25

[1] replacing the weighed Length-Sample weights by those calculated from the weight-length curve in use ensures that the discrepancy in *total* stock biomass is zero.

This option is unattractive because numbers would calculated by replacing certain data pertaining to the current survey—i.e. the weighed Length-Sample weights—by data from a different source—i.e. the given weight-length function. On the other hand, it is the logical corollary of using an exogenous weight-length curve as correct, and it does guarantee exact agreement between the survey estimate of total biomass and the sum of class biomasses, however the classes are defined.

Option 2. Replace the weight-length relationship.

There are several possibilities for replacing the weight-length function. All leave untouched the Length-Sample data, and therefore the estimates of class-specific numbers.

a) Present practice for correcting a discrepancy between the overall survey biomass estimate and the sum of class biomasses tends to be the applying a correction factor to the class biomasses so that they add up right. The statistical model underlying this procedure is that the exponent of the weight-length function being used is correct, but that the coefficient isn't—i.e. replacing the weight-length function with another one with the same exponent but a different coefficient. If the posterior correction factor applied to the class biomasses had been applied as a prior correction factor to the weight-length coefficient, the sums would have come out right in the first place.

Under this procedure, the biomass ratios of males and females, or fishable and non-fishable components, are still governed by the exponent of the given weight-length function, which is assumed reliable even though the coefficient is up for replacement.

b) A weight-length function can be calculated each year from the Length-Weight sample, usually comprising one to two thousand weight-length pairs. The overall disadvantage of doing so is that the shrimps composing the Length-Weight sample are not selected or treated in the same way as the shrimps in the Length Samples and their weights at length might not be exactly comparable. Furthermore, the sampling of shrimps for the Length-Weight samples is undescribed.

Deficiencies in present practice in applying this method are that the weight-length functions are commonly calculated using Sigmaplot, leading to the reporting of the coefficient to 1 significant digit, which is unsatisfactory; and that weight-length functions are calculated with no weighting applied to the shrimps in the sample.

c) The Length Samples, with their weights and length frequencies, provide a means of calculating a weight-length curve independent of having to weigh and measure individual shrimps, provided, which appears to be the case, that the length frequency distribution varies from one Length Sample to another. Table 1 is extracted from 2009 Length Sample length distributions, and it can be seen that the relative frequencies of different lengths differ from sample to sample. It is this between-sample variation in relative frequencies that makes it possible to consider calculating a weight-length curve from this set of data alone.

Table1: Example of length-class frequencies extracted from 2009 Length Sample data.

	CPL class (mm)														
Station	19	19.5	20	20.5	21	21.5	22	22.5	23	23.5	24	24.5	25	25.5	26
104	19	16	18	22	23	16	18	24	9	11	7	5	5	1	
105	3	4	6	13	5	12	6	6	8	7	2	3	1	1	
106	4	7	7	6	4	2	3	7	4	2	3	1	1		
107	3	3	6	2		2	2	4	4	3	5	4			
108	33	36	28	37	9	17	16	11	10	13	17	4	5		
109	1	3	3	1	7	3	8	5	10	23	33	31	28	17	11
11	2	3	5	2	1	4	2		2	1	1			1	
110	3	3	2	7	5	5	6	12	11	19	20	27	33	18	12
111			2	2	3	3	4	9	3	5	14	10	8	6	5
112	25	34	52	38	42	37	23	19	13	10	11	9	4	2	
113	22	25	26	24	12	14	5	4	3	4	3	1	1	2	1
114	15	18	10	9	8	7	9	5	4	5	3	1	1	1	
115	12	10	9	9	6	7	1	3	7	6	6	3	3		
116	23	26	31	20	23	13	8	6	10	11	10	4	3		1
117	39	55	59	65	59	58	43	33	17	13	26	13	6	5	3

The annual Length-Weight sample consists of 1 to 2 thousand individuals. The Length Sample data, with the associated sample weights, comprises some 200 weighed samples and measurements of a few tens of thousands of shrimps—a larger collection of lengths, but fewer weights. The exponential weight-length curve usually fitted requires only 2 parameters to be estimated, a coefficient and an exponent:

$$w(l,a,b) = a \cdot l^b$$

and fitting these parameters to the Length-Sample data presents no problems.

$$Samp.Wt_{ts} = a\sum_l n_{lts} l^b + \varepsilon_{ts}$$

and it is then a simple matter of adjusting the values of a and b to minimise an appropriately weighted sum of squares of the errors ε[2]. This is readily done with, for example, Excel Solver or PROC NLIN in SAS. If the objective is to fit a weight-length relationship such that the sum of class biomasses equals the survey biomass, i.e.

[2] in principle, it is possible to fit sex-specific weight-length curves using similar methods, with a statistical model:

$$Samp.Wt_{ts} = \sum_x a_x \sum_l n_{xlts} l^{b_x} + \varepsilon_{ts}$$

where x is the sex and the parameters a_x and b_x define sex-specific weight-length curves. In practice, the distributions of female lengths do not differ enough between stations for a female weight-length curve to be satisfactorily fitted.

$$\sum_t \sum_s \left(\frac{A_t}{K_t} \cdot \frac{Catch_{ts}}{Sw.Area_{ts}} \right) = \sum_t \sum_s \left(\frac{A_t}{K_t} \cdot \frac{Catch_{ts}}{Sw.Area_{ts}} \cdot \frac{\sum_l \hat{w}(l) n_{lts}}{Samp.Wt_{ts}} \right)$$

we should weight the errors epsilon proportional to catch and to stratum area and inversely proportional to the number of stations in the stratum, to the station swept area, and to the observed Length-Sample weight[3].

These methods were tested with the 2009 data that I was able to retrieve from various files and databases. Catch, stratum and swept area data were available for 247 stations, of which I found Length-Sample data, comprising 34 803 shrimps, for 215. However, there were two stations that had catches but for which I found no Length Sample data, and in order to simplify the present enquiry, they were simply deleted from the data sets I used. The numbers of stations in their strata were adjusted accordingly. A weight-length curve was fitted to the Length-Sample data using Excel Solver and with weights as above.

I found Length-Weight data for 1768[4] individual shrimps sampled in 2009. I fitted a simple exponential weight-length curve using Excel Solver, with weighting inversely proportional to fitted individual weights but no weighting for Catch, Stratum Area, &c. (Which appears to be consistent with the usual practice.) I also fitted sex-specific curves—male and female—to the same data.

I also found parameter values for a standard weight-length curve used in previous years and apparently used also in 2009.

Parameter values estimated by the fitting methods used are highly correlated, and differences between different parameter sets are best evaluated from a plot of the weight-length relationship. Figure 1 shows the ratios of a weight-length curve fitted to the 2009 Length-Weight sample and of the standard weight-length curve to the curve fitted to the 2009 Length-Sample data. The standard curve lay well below the other two, which for their part crossed over at 18.8 mm CPL, well in the middle of the length distribution. Table 2 shows the sum of length-class weights obtained from the different methods. Fitting a weight-length curve to the Length-Sample data alone gave the closest fit to the estimated total survey biomass. However, it appears from Figure 1 that the ‘standard’ weight-length curve was inappropriate for the estimation of length- and sex-class biomasses in 2009, and weight-length curves from either the Length-Sample data or the Weight-Length data would have been better. Not only did the standard curve underestimate class biomasses in aggregate, but apparently also overestimated the proportions of female and fishable biomass relative to male and unfishable biomass.

Table 2: Parameters of exponential weight-length curves fitted to 2009 data for northern shrimp from the West Greenland trawl survey.

Fitted to:	Coefficient[1]	Exponent[1]	Total weight error (%)	Fishable Biomass	Female Biomass (%)[2]
2009 Length-Sample data	0.946259	2.8456	0.40	263.595	37.40
2009 Length-Weight Sample data	0.824921	2.8931	1.24	266.236	37.64
Not known: ‘standard’ curve	0.57837	2.9941	– 3.26	255.201	38.16
Standard curve corrected	0.59788	2.9941	0.00	263.810	38.16

[1] parameter estimates have large error correlations; to evaluate differences between parameter sets it is necessary to see plots of the curves.
[2] as a proportion of the sum of the estimated male and female weights, not of the survey total.

[3] present code of practice prescribes adjusting the size of the cod-end sample to the catch weight alone (Ziemer and Siegstad 2009). It seems that the density of stations in the stratum should also be considered, larger samples relative to catch being appropriate in strata where stations are sparse, and smaller samples where stations are thick on the ground.
[4] unusually many. Weight-length sampling of shrimps in this survey usually produces about 1000 weight-length pairs.

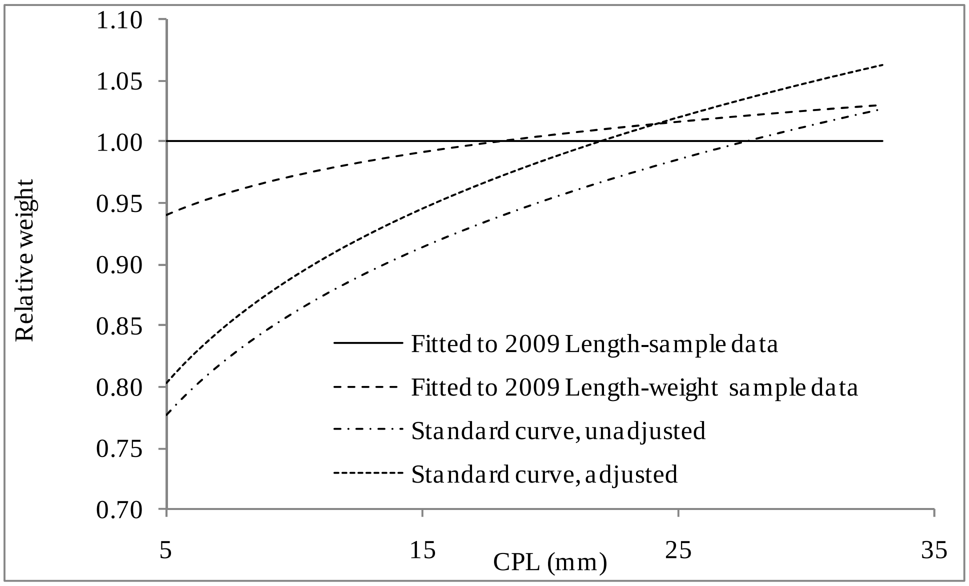

Figure 1: Ratios of length-specific shrimp weights from weight-length curves fitted to 2009 Length-Weight samples and to Length-Weight samples from some earlier years, relative to weights from a weight-length curve fitted to 2009 Length Samples.

This option is attractive from a statistical point of view because it does away with the need to refer to a separate data source. The processing—weighing, sorting, classifying and measuring—of Length Samples from the catches is a central activity in the West Greenland trawl survey directed toward evaluating the state of the stock, and much thought has gone into prescribing standard methods for carrying out the work and recording the measurements. There is therefore much to be said for making the maximum use of the data. There is statistical advantage in calculating the weight-length relationship directly from the data set to which it will be applied when class-specific biomasses are to be calculated, instead of referring to the results of a separate procedure which is likely to produce a weight-length relationship that is not quite compatible with the rest of the data. The weight-length curve fitted from the Length Samples alone will tend to fit best in the length classes where there is the greatest variation in numbers, which will tend to be where there are most shrimps in the catch, and this is what we want.

This option would also be readily applied to subdivisions of the survey, for example if we wanted to calculate male and female biomasses for a northern area and a southern area separately.

Conclusions and Recommendations

1. In connection with the reporting of survey results, there is much to be said for doing demographic analyses as simply as possible, while trying to avoid discrepancies between the sum of class biomasses and the total survey biomass.
2. For simplicity, it is suggested in this connection to fit a single weight-length curve to all sex classes combined over the whole survey and the entire length range; the advantage of fitting separate curves is small.
3. For simplicity, it is also suggested to fit the weight-length curve directly to Length-Sample data, and not to a separate set of individual weight and length measurements. By fitting the curve to the data to be partitioned, discrepancies between the sum of the partitions and the survey total will be reduced or minimised.
4. Which is not to suggest that the collecting of weight-length data on individual shrimps should be completely abandoned.

References

WIELAND, K. P. KANNEWORFF and B. BERGSTRÖM. 2004. Results of the Greenland Bottom Trawl Survey for Northern Shrimp (Pandalus borealis) off West Greenland (NAFO Subarea 1 and Division 0A), 1988-2004. *NAFO SCR Doc. 04/72;* Ser. No. N5042. 31 pp.

ZIEMER, N. and H. SIEGSTAD. 2009. Results of the Greenland bottom trawl survey for Northern Shrimp (*Pandalus borealis*) off West Greenland (NAFO Sub area 1 and Division 0A), 1988–2009. *NAFO SCR Doc. 09/067;* Ser. No. N5728. 34 pp.