Investigating different methods for obtaining hake survey probability distributions for length, and an inspection the sub-sample for which biologicals were measured

A. Ross-Gillespie, D.S. Butterworth, T.P. Fairweather and R.A. Rademeyer

1 Introduction

The hake cannibalism and inter-species predation model that is currently being developed by Ross-Gillespie and Butterworth (2014) has thus far failed to reflect both a biologically feasible estimate of daily ration and the proportion of hake in the diet of hake predators indicated by the diet data obtained during surveys. Andre Punt made a suggestion to investigate the sampling strategy used to obtain the diet data, in order to ascertain whether this strategy might be giving rise to biases in the estimates for the population as a whole in terms of both the length distributions and the proportions of hake in the diet of hake predators. This has led to an examination of the raw survey catch-at-length data and the methods used to analyse these data, and a few suggestions have been made for alternative approaches to weighting the length probability distributions from individual trawls in order to obtain aggregated distributions for each stratum.

This document gives the details of the survey sampling procedures, the equations for the methodology currently in place for calculating probability distributions for length, as well as equations for suggested alternative methods for weighting. Figures illustrating the impact that these alternative methods have on the probability distributions for length are provided.

A further aspect of this document is an inspection of the sub-sample of the survey catch for which biological parameters are measured, in order to assess the sampling strategy for obtaining biological information and potential biases that might arise. The effect of weighting the diet data by stratum population estimates is also illustrated.

2 Survey sampling procedure

Each cruise consists of a series of individual trawls conducted at different stations. The sampling procedure followed for each trawl is as follows:

1. If the total catch at a station is small, all hake are measured.
2. If the catch is too large it will be sub-sampled:

    (a) Since there will be an inclination to pick up larger fish first, all the hake are sorted into bins of two size categories: large and small (these terms are relative within a catch). For some surveys, there are a further two categories for sex, i.e large male, small male, large female, small female. All the bins in each category are weighed individually and noted against the relevant category, but only the total is captured in the database.

(b) Bins are selected at random from the sorted bins to ensure all fish sizes are measured. These are referred to as the 'observed' measures. Generally this means that all bins containing larger hake are measured, and only some of the bins containing smaller hake are measured.

3. Thus for every category of hake in the catch there will be a catch weight (the sum of the weight of all the bins) and a sample or 'observed' weight (the sum of the weight of the bins measured).

3   A note on nomenclature

In the documentation of the hake cannibalism and inter-species predation model, the superscripts s_p and l_p are used for species and length of hake predators, while the subscripts s and l are used for hake prey. In this document, in the interest of compatibility with documentation of the predation model, the superscripts s_p and l_p have been used for species and length of hake caught in the survey, as these hake represent the predators in the predation model. Note that even though the nomenclature used in this document differs a bit to what has previously been used to describe the method for obtaining probability distributions for length, the methodology is the same unless specified otherwise.

4   Obtaining probability distributions for length from the sample

For years in which there is no sex information available, there are three possible categories for the hake samples: large (L), small (S) and all hake not sorted by size (A).

Let $n_{y,d,t}^{s_p,l_p,i}$ be the number of hake of species s_p in length-group l_p that were measured ('observed') in category i for trawl t in depth stratum d, where $i \in \{L, S, A\}$. The total numbers estimated to have been caught in category i are obtained by scaling the measured numbers up by a ratio of $W_{y,d,t}^{s_p,i}/\widetilde{W}_{y,d,t}^{s_p,i}$, where $W_{y,d,t}^{s_p,i}$ is the total weight of the catch of category i for trawl t in depth stratum d, and $\widetilde{W}_{y,d,t}^{s_p,i}$ is the corresponding weight of the sub-sample of category i that was measured. The total estimated number of hake of length group l_p caught in trawl t and depth stratum d is then:

$$C_{y,d,t}^{s_p,l_p} = n_{y,d,t,l}^{s_p,L} \frac{W_{y,d,t}^{s_p,L}}{\widetilde{W}_{y,d,t}^{s_p,L}} + n_{y,d,t,l}^{s_p,S} \frac{W_{y,d,t}^{s_p,S}}{\widetilde{W}_{y,d,t}^{s_p,S}} + n_{y,d,t,l}^{s_p,A} \frac{W_{y,d,t}^{s_p,A}}{\widetilde{W}_{y,d,t}^{s_p,A}} \tag{4.1}$$

The proportion of hake in trawl t and depth stratum d that are of length group l_p is given by:

$$p_{y,d,t}^{s_p,l_p} = \frac{C_{y,d,t}^{s_p,l_p}}{\sum_{l_p} C_{y,d,t}^{s_p,l_p}} \tag{4.2}$$

The probability distribution for length for depth stratum d is obtained from a weighted average of the probability distributions for length from the individual trawls in that depth stratum:

$$p_{y,d}^{s_p,l_p} = \frac{\sum_t \alpha_{y,d,t}^{s_p} p_{y,d,t}^{s_p,l_p}}{\sum_t \alpha_{y,d,t}^{s_p}} \tag{4.3}$$

Currently, the weighting factor α_t is defined as follows:

$$\alpha_{y,d,t}^{s_p} = \begin{cases} 100 & \text{for } n_{y,d,t}^{s_p} > 100 \\ n_{y,d,t}^{s_p} & \text{for } n_{y,d,t}^{s_p} \leq 100 \end{cases} \tag{4.4}$$

where $n_{y,d,t}^{s_p} = \sum_{i=L,S,A} \sum_l n_{y,d,t}^{s_p,l_p,i}$ is the total number of hake measured in trawl t in depth stratum d. This approach is a consequence of the assumption that the hake population is distributed homogeneously across depth stratum d and each trawl represents a random sample from that depth stratum. In line with this assumption, all trawls should receive the same weighting unless the sample size is very small. In reality, however, the hake population is not homogeneously distributed, so that it could be argued that the density of the hake population at the location of each trawl should be taken into account. In other words:

$$p_{y,d}^{s_p,l_p} = \frac{\sum_t \widetilde{\alpha}_{y,d,t}^{s_p} p_{y,d,t}^{s_p,l_p}}{\sum_t \widetilde{\alpha}_{y,d,t}^{s_p}} \tag{4.5}$$

where $\widetilde{\alpha}_{y,d,t}^{s_p} = \alpha_{y,d,t}^{s_p} \beta_{y,d,t}^{s_p}$, with $\alpha_{y,d,t}^{s_p}$ being as in Equation 4.4 above, and $\beta_{y,d,t}^{s_p}$ being a measure of density for trawl t, calculated by dividing the total number of hake of species s_p estimated to have been caught in the trawl by the area swept by the trawl, i.e.:

$$\beta_{y,d,t}^{s_p} = \frac{\sum_l C_{y,d,t}^{s_p,l_p}}{A_{y,d,t}} \tag{4.6}$$

This document explores the impact of weighting without density (Equation 4.3) compared to weighting with density (Equation 4.5) on the resulting probability distributions for length. The document further investigates the effect of the choice of 100 as the sample size above which sampling variability is assumed to no longer dominate (Equation 4.4). Given

$$\alpha_{y,d,t}^{s_p} = \begin{cases} X & \text{for } n_{y,d,t}^{s_p} > X \\ n_{y,d,t}^{s_p} & \text{for } n_{y,d,t}^{s_p} \le X \end{cases} \tag{4.7}$$

values $X \in \{50, 100, 200\}$ are explored.

Results are presented on both stratum-disaggregated and stratum-aggregated levels. In order to obtain a stratum-aggregated probability distribution for length, the probability distributions for length from each stratum are weighted according to the estimated population size (in numbers) in that depth stratum, i.e.

$$p_y^{s_p,l_p} = \frac{\sum_d N_{y,d}^{s_p} p_{y,d}^{s_p,l_p}}{\sum_d N_{y,d}^{s_p}} \tag{4.8}$$

For results shown in this document, the survey estimated population size $N_{y,d}^{s_p}$ has been calculated in two ways, which, along with taking density-weighting into account has resulted in the Rademeyer hake model being re-run for three different cases.

Reference Case (RC): Density-weigthing is not taken into account, and $N_{y,d}^{s_p} = B_{y,d}^{s_p,surv}/\bar{W}_{y,d}^{s_p}$, where $B_{y,d}^{s_p,surv}$ is the survey biomass estimate for stratum d in year y and $\bar{W}_{y,d}^{s_p}$ is the mean weight of hake for stratum d in year y given by $\bar{W}_{y,d}^{s_p} = \sum_{l_p} p_{y,d}^{s_p,l_p} \alpha(l_p)^\beta$. The parameters α and β are the weight-length parameters for hake. This is the method that is currently in use.

Sensitivity 1 (Sen1): Density-weighting is not taken into account, and $N_{y,d}^{s_p}$ is calculated from the survey samples directly by averaging the densities for each trawl across stratum d to obtain an average density for the stratum (i.e. $\sum_t \beta_{y,d,t}^{s_p}/\sum_t 1$), and multiplying this average by the total area of the stratum.

Sensitivity 2 (Sen 2): Density-weighting is taken into account, and and $N_{y,d}^{s_p}$ is calculated from the survey samples directly by averaging the densities for each trawl across stratum d to obtain an average density for the stratum (i.e. $\sum_t \beta_{y,d,t}^{s_p} / \sum_t 1$), and multiplying this average by the total area of the stratum.

5 Obtaining sex-disaggregated proportions

In certain years, a sub-sample of the measured catch was also sexed. Let $n_{g,y,d,t}^{s_p,l_p,i}$ be number of hake of species s_p, gender g and length group l_p measured in category i ($i \in \{L, S, A\}$) of trawl t in depth stratum d. Then the total number of hake of gender g and length group l_p estimated to have been caught in trawl t and depth stratum d is:

$$C_{g,y,d,t}^{s_p,l_p} = n_{y,g,d,t}^{s_p,L} \frac{W_{g,y,d,t}^{s_p,L}}{\widetilde{W}_{g,y,d,t}^{s_p,L}} + n_{y,g,d,t}^{s_p,S} \frac{W_{g,y,d,t}^{s_p,S}}{\widetilde{W}_{g,y,d,t}^{s_p,S}} + n_{y,g,d,t}^{s_p,A} \frac{W_{g,y,d,t}^{s_p,A}}{\widetilde{W}_{g,y,d,t}^{s_p,A}} \tag{5.1}$$

$W_{g,y,d,t}^{s_p,i}$ is the total weight of hake of species s_p and gender g caught in category i for trawl t in depth stratum d, and $\widetilde{W}_{g,y,d,t}^{s_p,i}$ is the corresponding weight of the sub-sample of category i that was measured. The total number of hake of gender g of length group l_p estimated to have been caught in stratum d is:

$$C_{g,y,d}^{s_p,l_p} = \sum_t C_{g,y,d,t}^{s_p,l_p} \tag{5.2}$$

Currently, the proportion of the total sexed catch in length group l_p that is of gender g is calculated by:

$$q_{g,y,d}^{s_p,l_p} = \frac{C_{g,y,d}^{s_p,l_p}}{\sum_{g'} C_{g',d,l}} \tag{5.3}$$

Strictly speaking, however, the $q_{g,y,d}^{s_p,l_p}$ values should be weighted as in Equation 4.3. For the results presented in this document, the following approach was taken:

Define a trawl-specific $q_{g,y,d,t}^{s_p,l_p}$:

$$q_{g,y,d,t}^{s_p,l_p} = \frac{C_{g,y,d,t}^{s_p,l_p}}{\sum_{g'} C_{g',y,d,t}^{s_p,l_p}} \tag{5.4}$$

The $q_{g,y,d}^{s_p,l_p}$ for stratum d is then given by:

$$q_{g,y,d}^{s_p,l_p} = \frac{\sum_t \widetilde{\alpha}_{y,d,t}^{s_p,\text{sexed}} q_{g,y,d,t}^{s_p,l_p}}{\sum_t \widetilde{\alpha}_{y,d,t}^{s_p,\text{sexed}}} \tag{5.5}$$

where $\widetilde{\alpha}_{y,d,t}^{s_p,\text{sexed}} = \alpha_{y,d,t}^{s_p,\text{sexed}} \beta_{y,d,t}^{s_p}$, similar to equation 4.5, except that

$$\alpha_{y,d,t}^{s_p,\text{sexed}} = \begin{cases} 100 & \text{for } n_{y,d,t}^{s_p,\text{sexed}} > 100 \\ n_{y,d,t}^{s_p,\text{sexed}} & \text{for } n_{y,d,t}^{s_p,\text{sexed}} \leq 100 \end{cases} \tag{5.6}$$

where $n_{y,d,t}^{s_p,\text{sexed}} = \sum_{i=L,S,A} \sum_g \sum_l n_{g,y,d,t}^{s_p,l_p,i}$ is the total number of hake that were sexed and measured in trawl t in depth stratum d. $\beta_{y,d,t}^{s_p}$ is a measure of density for trawl t same as before.

This proportion $q_{g,y,d}^{s_p,l_p}$ is then used to split the proportion-at-length probability distribution $p_{y,d}^{s_p,l_p}$ from Equation 4.3 (or 4.5) into male and female proportions as follows (taken from Rademeyer 2011):

1. The proportions-at-length are grouped into 2cm length classes.

2. For all length classes $< 21cm$, the hake are assumed to be juvenile and the proportions-at-length unsexed (U).

3. For length classes $\geq 21cm$:

   - If there is sex-information for both of the two 1 cm length classes to group (i.e. if $\sum_g q_{g,y,d}^{s_p,l_p} = 1$ and $\sum_g q_{y,g,d}^{s_p,l_p+1} = 1$), then the sex-information is used directly:

$$p_{g,y,d}^{s_p,l_p} = \begin{cases} 0 & \text{for } g = U \\ q_{g,y,d}^{s_p,l_p} p_{y,d}^{s_p,l_p} + q_{y,g,d}^{s_p,l_p+1} p_{y,d}^{s_p,l_p+1} & \text{for } g \in \{M, F\} \end{cases} \tag{5.7}$$

   - If there is sex-information for only one of the two 1 cm length classes to group (i.e. if $\sum_g q_{g,y,d}^{s_p,l_p} = 1$ or $\sum_g q_{y,g,d}^{s_p,l_p+1} = 1$), then the sex-information from the one length class is used for both:

$$p_{g,y,d}^{s_p,l_p} = \begin{cases} 0 & \text{for } g = U \\ q_{g,y,d}^{s_p,l_p}\left(p_{y,d}^{s_p,l_p} + p_{y,d}^{s_p,l_p+1}\right) \text{ or } q_{y,g,d}^{s_p,l_p+1}\left(p_{y,d}^{s_p,l_p} + p_{y,d}^{s_p,l_p+1}\right) & \text{for } g \in \{M, F\} \end{cases} \tag{5.8}$$

   - If there is no sex-information for either of the two 1 cm length classes to group (i.e. $\sum_g q_{g,y,d}^{s_p,l_p} = 0$ and $\sum_g q_{y,g,d}^{s_p,l_p+1} = 0$), then the proportion for the resulting 2 cm length class is assumed to be unsexed:

$$p_{g,y,d}^{s_p,l_p} = \begin{cases} p_{y,d}^{s_p,l_p} + p_{y,d}^{s_p,l_p+1} & \text{for } g = U \\ 0 & \text{for } g \in \{M, F\} \end{cases} \tag{5.9}$$

4. Once the stratum-aggregated $p_{g,y}^{s_p,l_p}$ has been computed (see Equation 5.10), then for each $2cm$ length class of $p_{g,y}^{s_p,l_p}$ greater than $21cm$, the male to female ratio for that length class is used to split the unsexed proportion into males and females. If there is no sex information for the length class, then the average of the length classes immediately before and after is used to allocate the unsexed proportion.

The stratum-aggregated $p_{g,y}^{s_p,l_p}$, is calculated in a similar manner to Equation 4.8:

$$p_{g,y}^{s_p,l_p} = \frac{\sum_d N_{y,d}^{s_p} p_{g,y,d}^{s_p,l_p}}{\sum_d N_{y,d}^{s_p}} \tag{5.10}$$

Here, $N_{y,d}^{s_p}$ is also calculated in two ways according to the methods described in Sen1 and Sen2.

6   Biological sampling procedure

In earlier years, the instructions were to collect five specimens per length class for biological sampling (primary target was otolith samples). These samples could be collected across the depth strata. Since smaller hake are encountered more easily (relative abundance and availability to the survey) in waters $< 300m$, the focus of biological sampling at greater depths is generally on larger hake in order to fulfil the sampling quota. In more recent years, the minimum number of samples required per length class has been increased, and the quota per length bin is reset for the second half of the cruise.

These are

This document presents the length distributions from the total survey sample plotted against the sub-sample for which biologicals were measured, showing how the sampling strategy outlined above could lead to biases. With this in mind, the question then is whether weighting by a measure of stratum density or abundance will substantially change the general trends implied by the diet data, in particular trends for the proportion of hake in the diet of hake predators.

The hake cannibalism and inter-species predation model incorporates the diet data into the negative log likelihood as follows:

$$-lnL+ = -\sum_y \left(h_{y,bio}^{s_p l_p} ln \hat{p}_{y,prop}^{s_p l_p} + (n_{y,bio}^{s_p l_p} - h_{y,bio}^{s_p l_p}) ln(1 - \hat{p}_{y,prop}^{s_p l_p}) \right) \tag{6.1}$$

where

$n_{y,bio}^{s_p l_p}$ is the observed number of hake predators of species s_p and length l_p with non-empty stomachs in year y,

$h_{y,bio}^{s_p l_p}$ is the observed number of hake predators of species s_p and length l_p with more than 50% hake prey (by weight) in the stomach content in year y, and

$\hat{p}_{y,prop}^{s_p l_p}$ is the model-predicted proportion of hake in the diet of hake predators of species s_p and length l_p in year y.

When stratum-weighting is not taken into account, $n_{y,bio}^{s_p l_p}$ and $h_{y,bio}^{s_p l_p}$ are simply the sum over the observations from the individual strata [1] , i.e. $n_{y,bio}^{s_p l_p} = \sum_d n_{y,d,bio}^{s_p l_p}$ and $h_{y,bio}^{s_p l_p} = \sum_d h_{y,d,bio}^{s_p l_p}$. Let $\widetilde{n}_{y,bio}^{s_p l_p}$ and $\widetilde{h}_{y,bio}^{s_p l_p}$ denote the stratum-weighted alternatives to $n_{y,bio}^{s_p l_p}$ and $h_{y,bio}^{s_p l_p}$. Two possible approaches for incorporating stratum weights have been considered.

Appraoch A

The first step is to calculate a stratum-weighted proportion of hake in diet from the biological samples.

$$\widetilde{p}_{y,bio}^{sp,lp} = \frac{\sum_d N_d^{s_p} h_{y,d,bio}^{s_p l_p}}{\sum_d N_d^{s_p} n_{y,d,bio}^{s_p l_p}} \tag{6.2}$$

where $N_d^{s_p}$ is the weighting factor for depth stratum d, given by the survey estimated population size (in numbers) for predators of species s_p in depth stratum d (as for Equation 4.8). Define $\widetilde{n}_y^{s_p l_p}$ and $\widetilde{h}_y^{s_p l_p}$ as follows:

$$\widetilde{n}_{y,bio}^{s_p l_p} = n_{y,bio}^{s_p l_p} = \sum_d n_{y,d,bio}^{s_p l_p} \tag{6.3}$$

$$\widetilde{h}_{y,bio}^{s_p l_p} = \widetilde{p}_y^{sp,lp} \widetilde{n}_{y,bio}^{s_p l_p} \tag{6.4}$$

These stratum-weighted $\widetilde{n}_y^{s_p l_p}$ and $\widetilde{h}_y^{s_p l_p}$ then replace $n_y^{s_p l_p} = \sum_d n_{y,d}^{s_p l_p}$ and $h_y^{s_p l_p} = \sum_d h_{y,d}^{s_p l_p}$ in the likelihood Equation 6.1.

---

[1]Note that observations from individual trawls were weighted according to Equation 4.5, but this weighting has very little impact given the relatively small sizes of the biological samples from each trawl.

Approach B

The second approach is identical to the first, except that Equation 6.2 changes to:

$$\widetilde{p}_{y,bio}^{sp,lp} = \frac{\sum_d N_d^{s_p l_p} h_{y,d}^{s_p l_p}}{\sum_d N_d^{s_p l_p} n_{y,d}^{s_p l_p}} \tag{6.5}$$

Here the weighting factor $N_d^{s_p l_p}$ is the survey estimated numbers per length for stratum d, given by

$$N_d^{s_p l_p} = p_{y,d}^{s_p,l_p} N_d^{s_p} \tag{6.6}$$

where $N_d^{s_p}$ is as for Equation 6.2 and $p_{y,d}^{s_p,l_p}$ is the survey probability-at-length distribution for depth stratum d (Equation 4.5).

A summary of the parameter symbols used in this document is given in Table 1.

7 Results

Note that for all the results presented in this document, samples taken at depths greater than 500m have been excluded, since the 501-1000m depth stratum has not been sampled consistently throughout the sampling period. Further, in order to help reduce the number of plots, *M. paradoxus* results for the 0-100m depth stratum and *M. capensis* results for the 401-500m depth stratum have not been shown here, since the respective sample sizes from these depth strata are relatively small.

Table 2 gives a summary of the Figures presented in this document. It can be seen from Figures 1, 2 and 4 that different choices for X (the sample size above which sampling variability is assumed to no longer dominate) make no appreciable difference to the probability distributions. Figures 3a and 3b, as well as the stratum-aggregated plots in Figures 5a and 5b, suggest that the impact of weighting by stratum density might be more substantial.

Figures 6a, 6b, 7a and 7b superimpose the probability-at-length distributions from the biological samples on top of those from the whole survey sample. The last panel in each plot shows the stratum-aggregated proportions. While the length distributions from the biological samples are not necessarily an accurate reflection of the length distribution of the population as indicated by the survey (likely owing to the biological sampling strategy), the biological samples to seem to at least represent most length classes reasonably well when aggregated across the depth strata.

Figures 8a and 8a show the survey-estimated densities for each stratum and each cruise against the proportion of biological samples taken in each stratum. There seems to be a general trend of a relatively large proportion of biological samples coming from deeper strata where the survey estimates of the population density are small. This is particularly the case for *M. capensis* . This indicates that weighting the diet data by stratum density would be justified.

Figure 9 shows the *M. paradoxus* and *M. capensis* spawning biomass trajectories for the reference case where no density-weighting is implemented, and the estimated population size (in numbers) per stratum is calculated the

survey estimate of spawning biomass; Sen1 where there is no density-weighting but the estimated population size per stratum is calculated directly from the survey samples; Sen2 where density-weighting has been taken into account and the estimated population size is calculated directly from the survey samples. The impact of density weighting on the population trajectory seems very small.

Figure 10 shows the proportion of hake in the diet of *M. paradoxus* and *M. capensis* predators as indicated by the diet data when stratum-weighting is taken into account and when it is not. For Case A (when the samples are weighted by the stratum population estimate in terms of numbers), there is not much change to be observed for *M. paradoxus*, but the proportion is lower for *M. capensis* when stratum weighting is taken into account. The reverse is true for Case B (when the samples are weighted by the stratum population estimate in terms of numbers-at-length), where there is not much change for *M. capensis*, but the *M. paradoxus* proportion of hake in diet is lower. Both weighting approaches have the potential to help mitigate the difficulties experienced in fitting the hake cannibalism and inter-species predation model to the available data.

References

Rademeyer, R.A. 2011. The evolution of management procedures for the South African hake resource in the 2000s. PhD Thesis. University of Cape Town.

Ross-Gillespie,A. and Butterworth, D.S. Modelling cannibalism and inter-species predation for Cape hake *Merluccius capensis* and *M. paradoxus* : an update to MARAM IWS/DEC13/Ecofish/P10. Document MARAM/IWS/DEC14/Hake/P8 submitted to the International Stock Assessment Workshop, December 2014, Cape Town.

Table 1: Parameter symbols used in this document and their corresponding descriptions. In all of the below, subscript t refers to trawl and d to depth stratum.

	General survey samples
$n_{y,d,t}^{s_p,l_p,i}$	number of hake measured in category $i \in \{L, S, A\}$ that fall into length-group l_p
$n_{y,d,t}^{s_p}$	total number of hake measured in trawl t in depth stratum d
$W_{y,d,t}^{s_p,i}$	total weight of hake caught in category i for trawl t in depth stratum d
$\widetilde{W}_{y,d,t}^{s_p,i}$	total weight of the sub-sample of hake in category i that was measured
$C_{y,d,t}^{s_p,l_p}$	total number of hake of length group l_p estimated to be caught by trawl t in depth stratum d
$p_{y,d,t}^{s_p,l_p}$	estimated proportion of hake of length group l_p caught in trawl t and depth stratum d
$p_{y,d}^{s_p,l_p}$	estimated proportion of hake of length group l_p caught in depth stratum d
p_l	estimated stratum-aggregated probability distribution for length
$\alpha_{y,d,t}^{s_p}$	density independent weighting factor for trawl t and depth stratum d
X	sample size above which sampling variability is assumed to no longer dominate
$\widetilde{\alpha}_{y,d,t}^{s_p}$	density dependent weighting factor for trawl t and depth stratum d
$\beta_{y,d,t}^{s_p}$	density for trawl t, taken to be the total number of hake caught divided by the area swept by trawl t
$A_{y,d,t}$	area swept by trawl t in depth stratum d
	Sex-disaggregated samples
$n_{g,y,d,t}^{s_p,l_p,i}$	number of hake of gender g measured in category i ($i \in \{L, S, A\}$) that fall into length-group l_p
$W_{g,y,d,t}^{s_p,i}$	total weight of hake of gender g caught in category i for trawl t in depth stratum d
$\widetilde{W}_{h,y,d,t}^{s_p,i}$	total weight of the sub-sample of hake of gender g in category i that was measured
$C_{g,y,d,t}^{s_p,l_p}$	total number of hake of gender g and length group l_p estimated to be caught by trawl t in stratum d
$q_{g,y,d,t}^{s_p,l_p}$	the proportion of the total sexed catch in length group l_p and trawl t that is of gender g
$q_{g,y,d}^{s_p,l_p}$	the proportion of the total sexed catch in length group l_p and depth stratum d that is of gender g
$p_{g,y,d}^{s_p,l_p}$	estimated proportion of hake of gender g and length group l_p caught in depth stratum d
$p_{g,y}^{s_p,l_p}$	estimated stratum-aggregated probability-at-length distribution for gender g
$\alpha_{y,d,t}^{s_p,\text{sexed}}$	density independent weighting factor for sex-disaggregated samples
$\widetilde{\alpha}_{y,d,t}^{s_p,\text{sexed}}$	density dependent weighting factor for sex-disaggregated samples
n^{sexed}	total number of hake that were sexed and measured in trawl t in depth stratum d
	Biological samples
$n_{y,bio}^{s_p l_p}$	observed number of hake predators of species s_p and length l_p with non-empty stomachs in year y
$h_{y,bio}^{s_p l_p}$	observed number of predators of species s_p and length l_p with more than 50% hake prey in the stomach content in year y
$\widetilde{n}_{y,bio}^{s_p l_p}$	stratum-weighted alternative to $n_{y,bio}^{s_p l_p}$
$\widetilde{h}_{y,bio}^{s_p l_p}$	stratum-weighted alternative to $h_{y,bio}^{s_p l_p}$
$\widetilde{p}_{y,bio}^{s_p,l_p}$	stratum-weighted proportion of hake in the diet of predators of species s_p and length l_p in year y, calculated from biological samples
$\hat{p}_{y,prop}^{s_p l_p}$	model-predicted proportion of hake in the diet of predators of species s_p and length l_p in year y

Table 2: List of Figures with corresponding descriptions

Figure number	Description	Page
Figure 1a	*M. paradoxus* probability distributions for length for different X without density weighting	11
Figure 1b	*M. capensis* probability distributions for length for different X without density weighting	11
Figure 2a	*M. paradoxus* probability distributions for length for different X with density weighting	12
Figure 2b	*M. capensis* probability distributions for length for different X with density weighting	12
Figure 3a	*M. paradoxus* probability distributions for length for $X = 100$ with and without density weighting	13
Figure 3b	*M. capensis* probability distributions for length for $X = 100$ with and without density weighting	14
Figure 4a	*M. paradoxus* stratum-aggregated probability distributions for length for different X	15
Figure 4b	*M. capensis* stratum-aggregated probability distributions for length for different X	15
Figure 5a	*M. paradoxus* stratum-aggregated distributions for $X = 100$ with and without density weighting for all cruises	16
Figure 5b	*M. capensis* stratum-aggregated distributions for $X = 100$ with and without density weighting for all cruises	17
Figure 6a	Comparison of *M. paradoxus* survey probability distributions for length with length-distributions of the biological sample	18
Figure 6b	Continuation of Figure 6a	19
Figure 7a	Comparison of *M. capensis* survey probability distributions for length with length-distributions for the biological sample	20
Figure 7b	Continuation of Figure 7a	21
Figure 8a	*M. paradoxus* biological sample size in relation to the survey estimated population size per stratum	22
Figure 8b	*M. capensis* biological sample size in relation to the survey estimated population size per stratum	23
Figure 9	*M. paradoxus* and *M. capensis* biomass trajectories for when density-weighting is taken into account and when it is not	24
Figure 10	Proportion of hake in the diet of *M. paradoxus* and *M. capensis* predators for when stratum-weighting is taken into account and when it isn't	25

Probability distributions for length for different X WITHOUT density weighting

(a) *M. paradoxus*

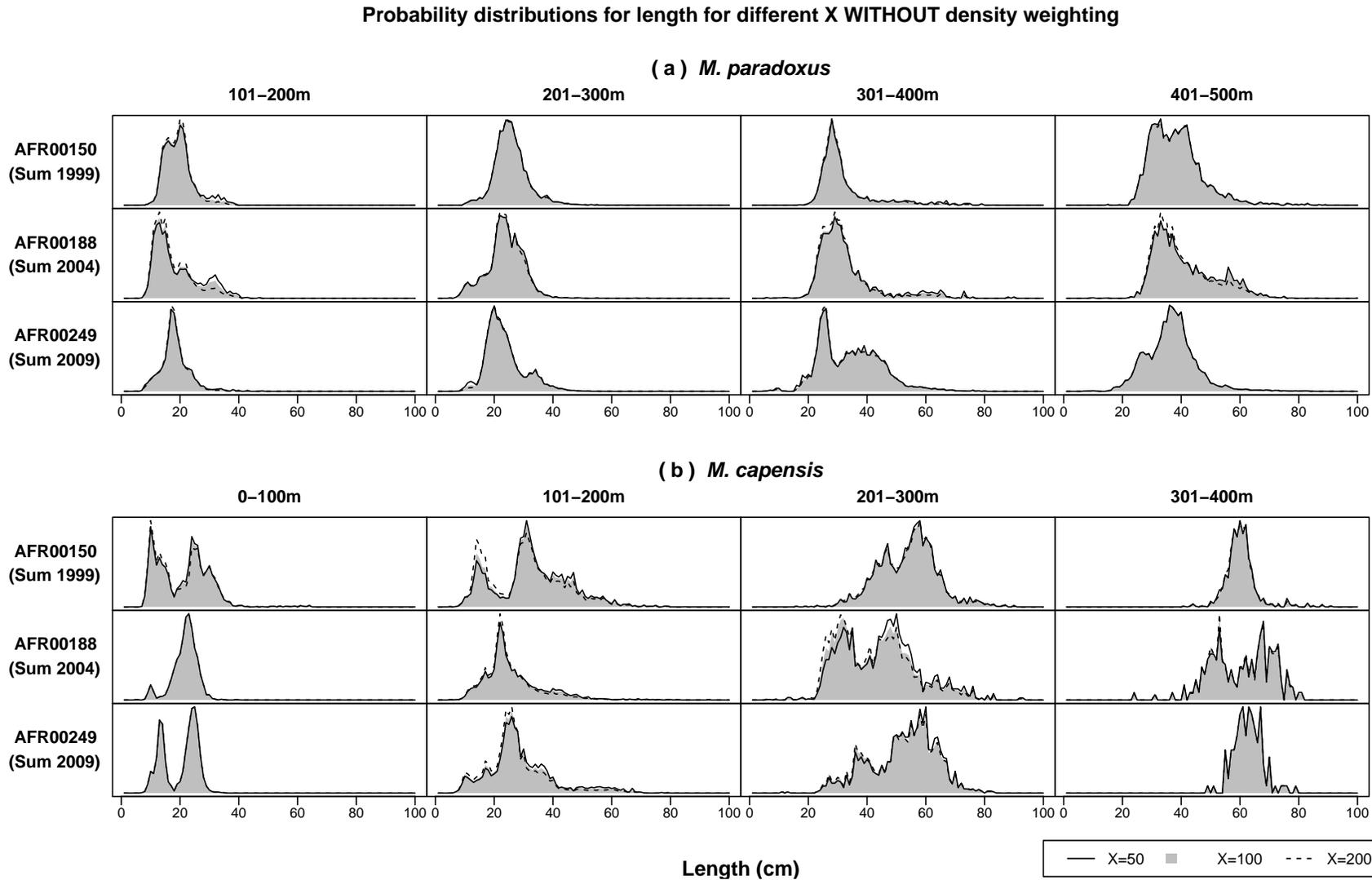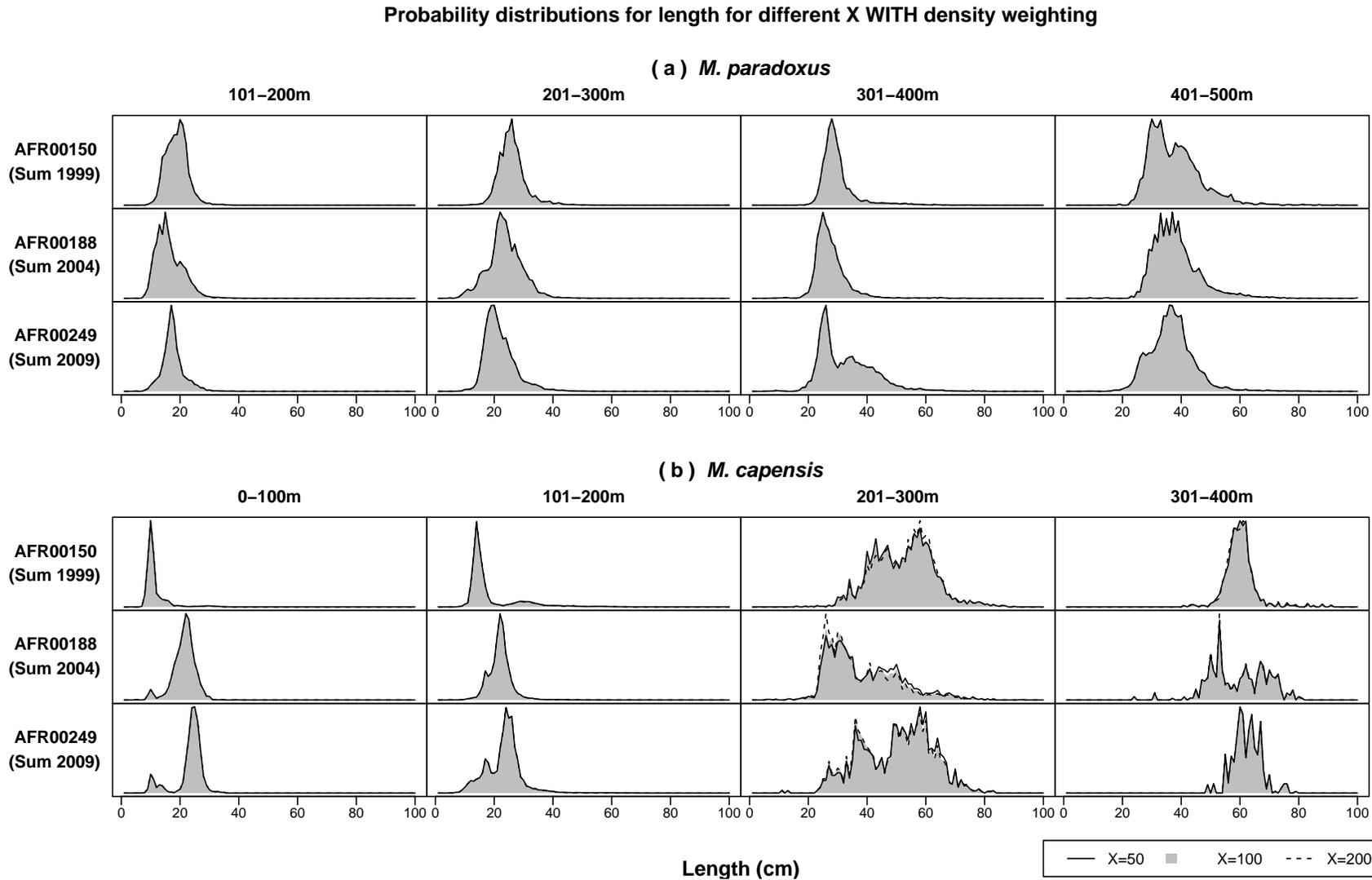**(b) *M. capensis***

Length (cm)

Figure 1: *M. paradoxus* and *M. capensis* probability distributions for length shown for a selection of three cruises and for three different values of X, the sample size above which sampling variability is assumed to no longer dominate. The results here are for the case when density weighting is not taken into account (Equation 4.3). Note that for the sake of clarity, the vertical axes are not all to the same scale.

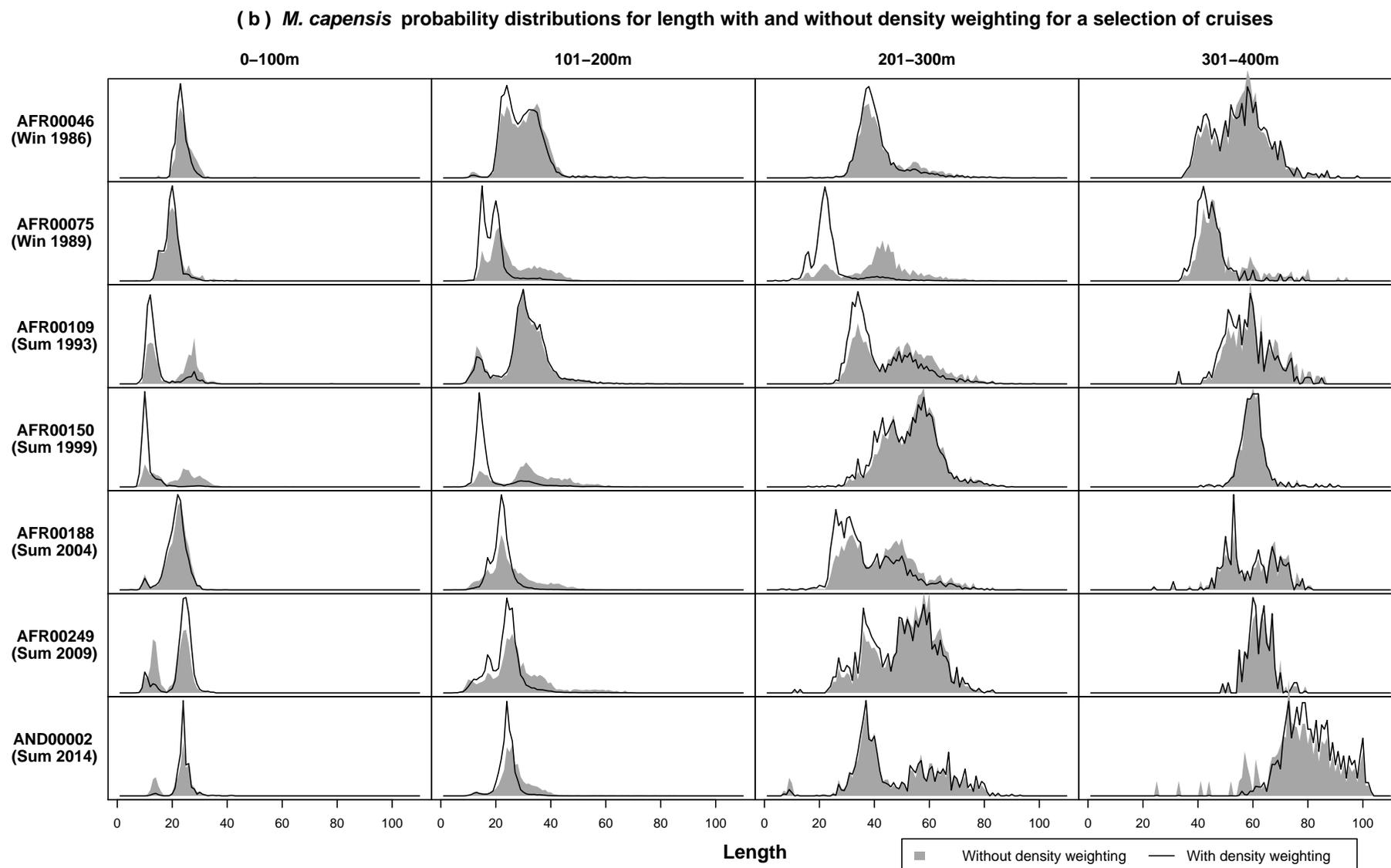**Figure 2:** *M. paradoxus* and *M. capensis* probability distributions for length shown for a selection of three cruises and for three different values of X, the sample size above which sampling variability is assumed to no longer dominate. The results here are for the case when density weighting is taken into account (Equation 4.5). Note that for the sake of clarity, the vertical axes are not all to the same scale.

(a) *M. paradoxus* **probability distributions for length with and without density weighting for a selection of cruises**

Figure 3a: *M. paradoxus* probability distributions for length for each of four depth strata are shown for a selection of cruises for $X = 100$. The grey shaded areas indicate the probability distributions when density weighting is not taken into account (Equation 4.3), while the black lines show the probability distributions for when this weighting is taken into account (Equation 4.5). Note that for the sake of clarity, the vertical axes are not all to the same scale.
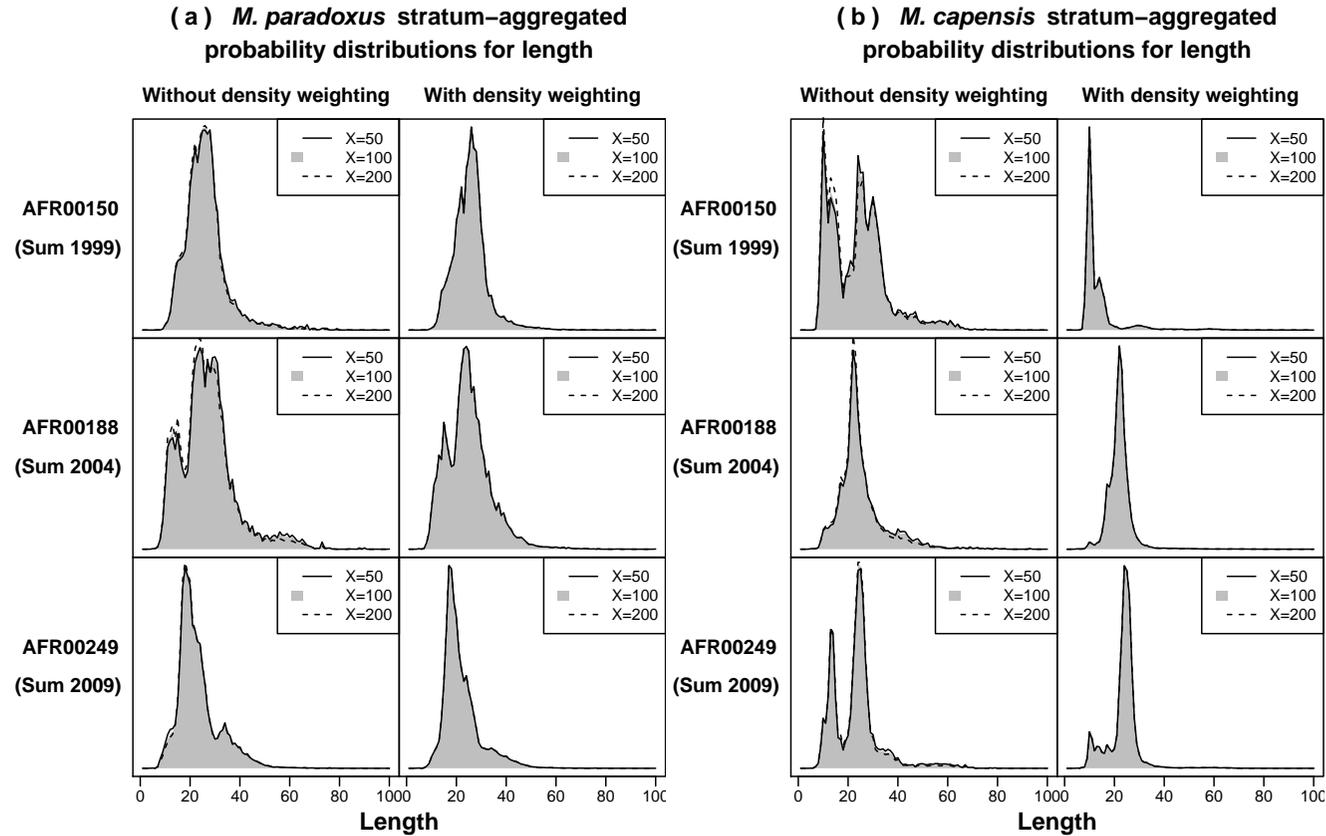
(b) *M. capensis* probability distributions for length with and without density weighting for a selection of cruises

Figure 3b: *M. capensis* probability distributions for length for each of four depth strata are shown for a selection of cruises for $X = 100$. The grey shaded areas indicate the probability distributions when density weighting is not taken into account (Equation 4.3), while the black lines show the probability distributions for when this weighting is taken into account (Equation 4.5). Note that for the sake of clarity, the vertical axes are not all to the same scale.

(a) *M. paradoxus* stratum–aggregated probability distributions for length

(b) *M. capensis* stratum–aggregated probability distributions for length

Figure 4: *M. paradoxus* and *M. capensis* stratum-aggregated probability distributions for length are shown for a selection of three cruises and for three different values of X, the sample size above which sampling variability is assumed to no longer dominate. The results are shown for both the case where density weighting is not taken into account and where it is. Note that for the sake of clarity, the vertical axes are not all to the same scale. Note also that the stratum-aggregated proportions were calculated using the methodology of Sen 1.
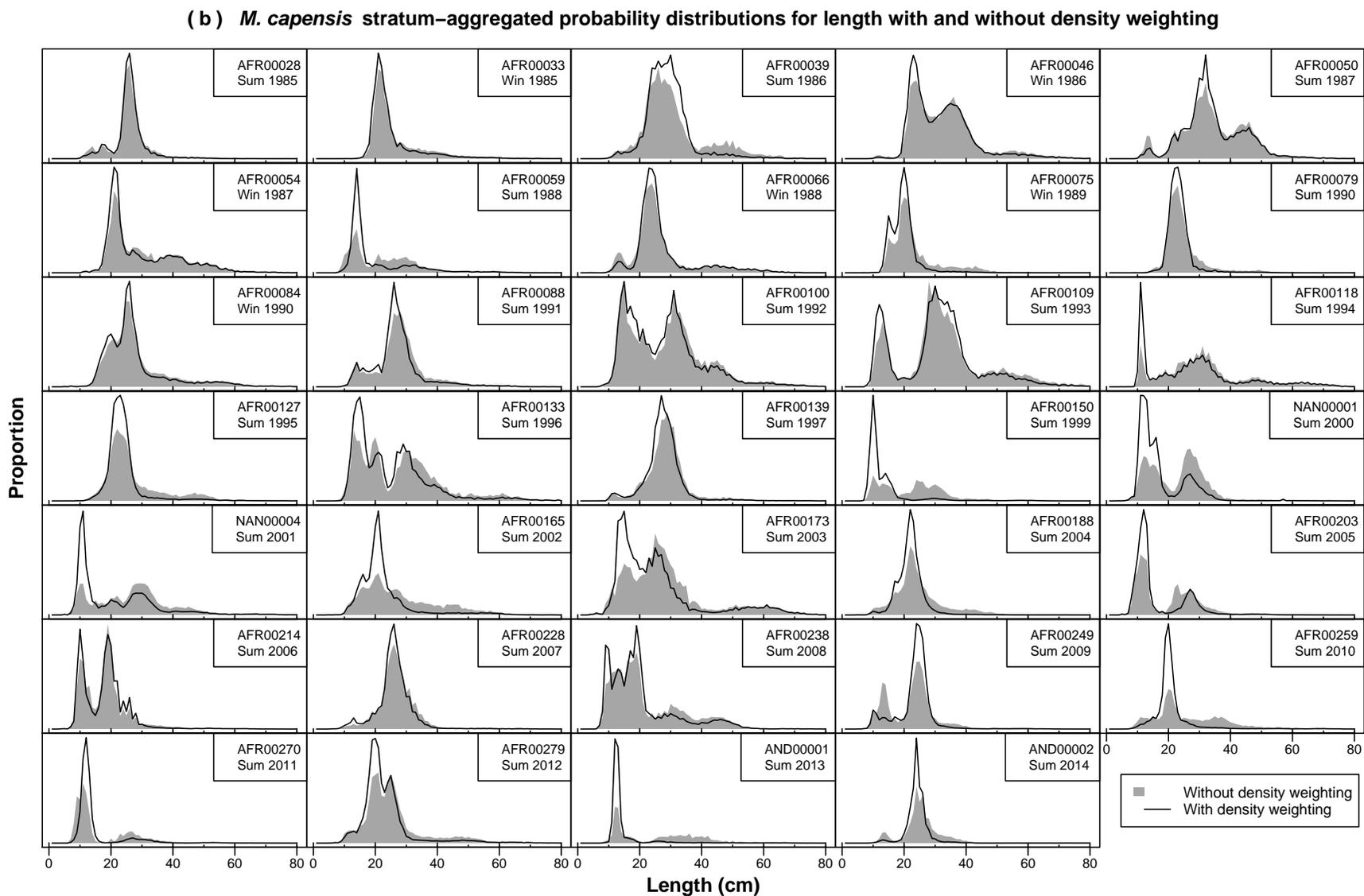
(a) *M. paradoxus* **stratum–aggregated probability distributions for length with and without density weighting**

Figure 5a: *M. paradoxus* stratum-aggregated probability distributions for length are shown for all cruises for $X = 100$. The grey shaded areas indicate the probability distributions when density weighting is not taken into account (Equation 4.3), while the black lines show the probability distributions for when this weighting is taken into account (Equation 4.5). Note that for the sake of clarity, the vertical axes are not all to the same scale. Note also that the stratum-aggregated proportions were calculated using the methodology of Sen 1.
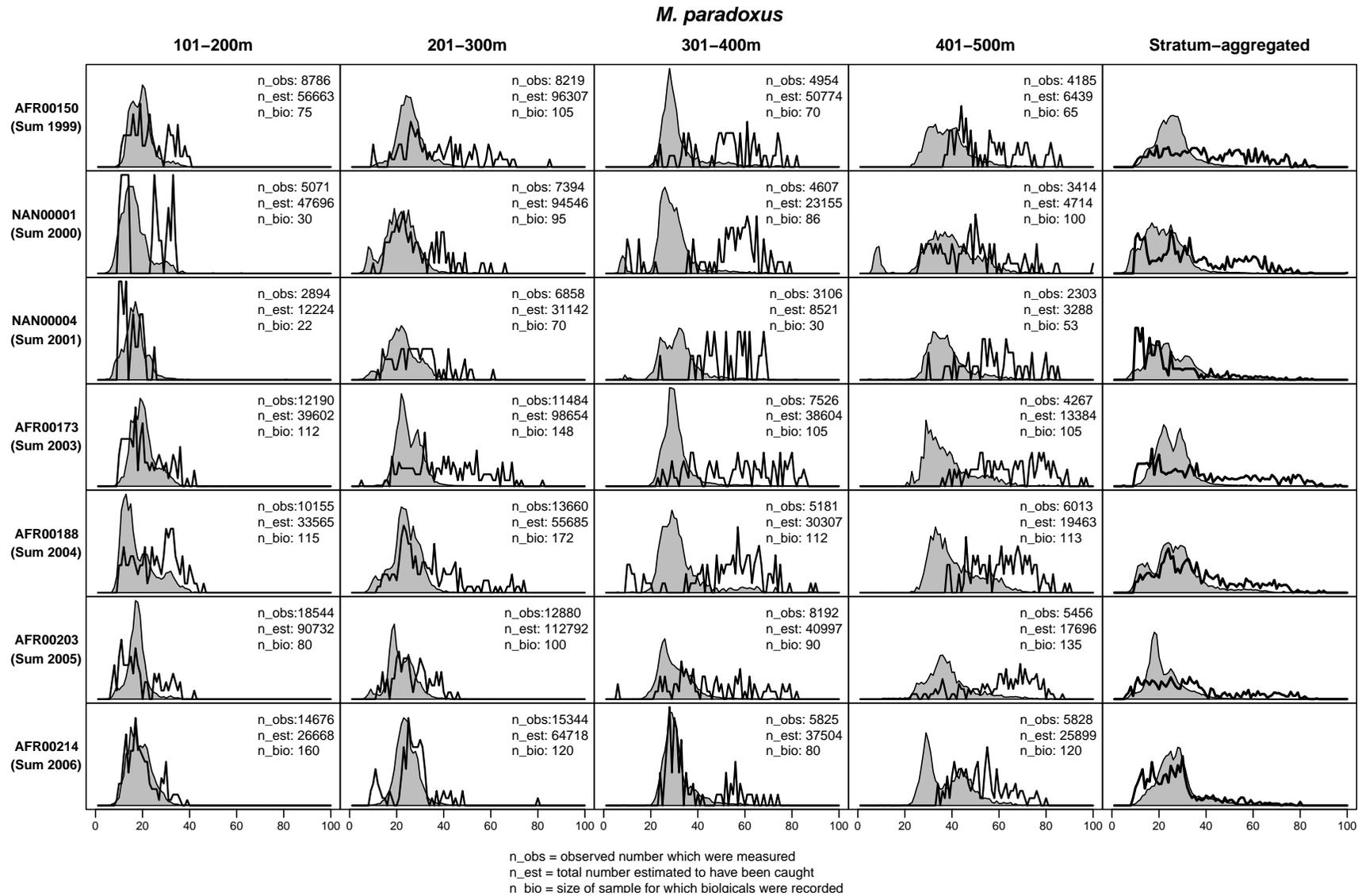
(b) *M. capensis* **stratum−aggregated probability distributions for length with and without density weighting**

Figure 5b: *M. capensis* stratum-aggregated probability distributions for length are shown for all cruises for $X = 100$. The grey shaded areas indicate the probability distributions when density weighting is not taken into account (Equation 4.3), while the black lines show the probability distributions for the this weighting is taken into account (Equation 4.5). Note that for the sake of clarity, the vertical axes are not all to the same scale. Note also that the stratum-aggregated proportions were calculated using the methodology of Sen 1.
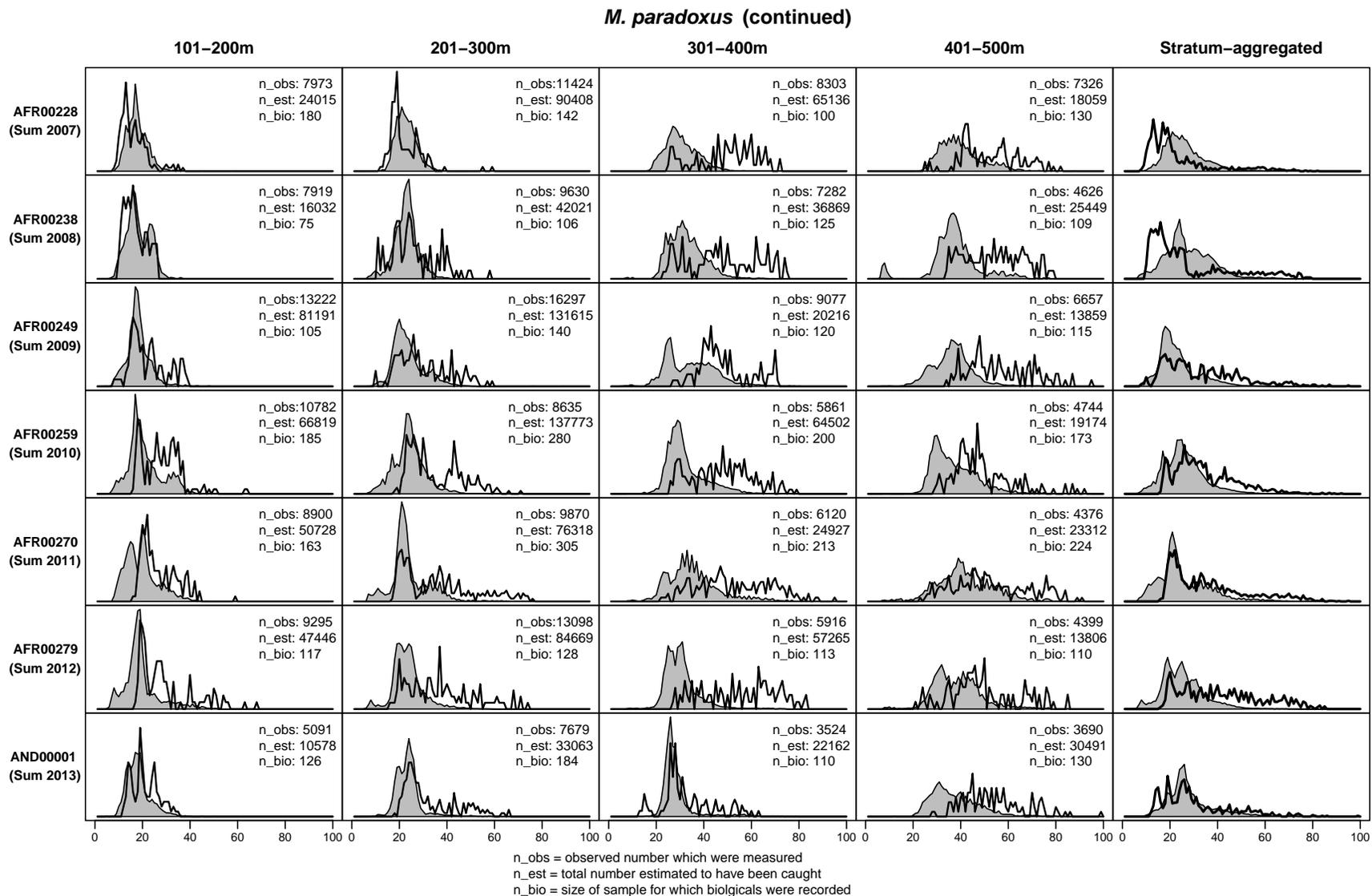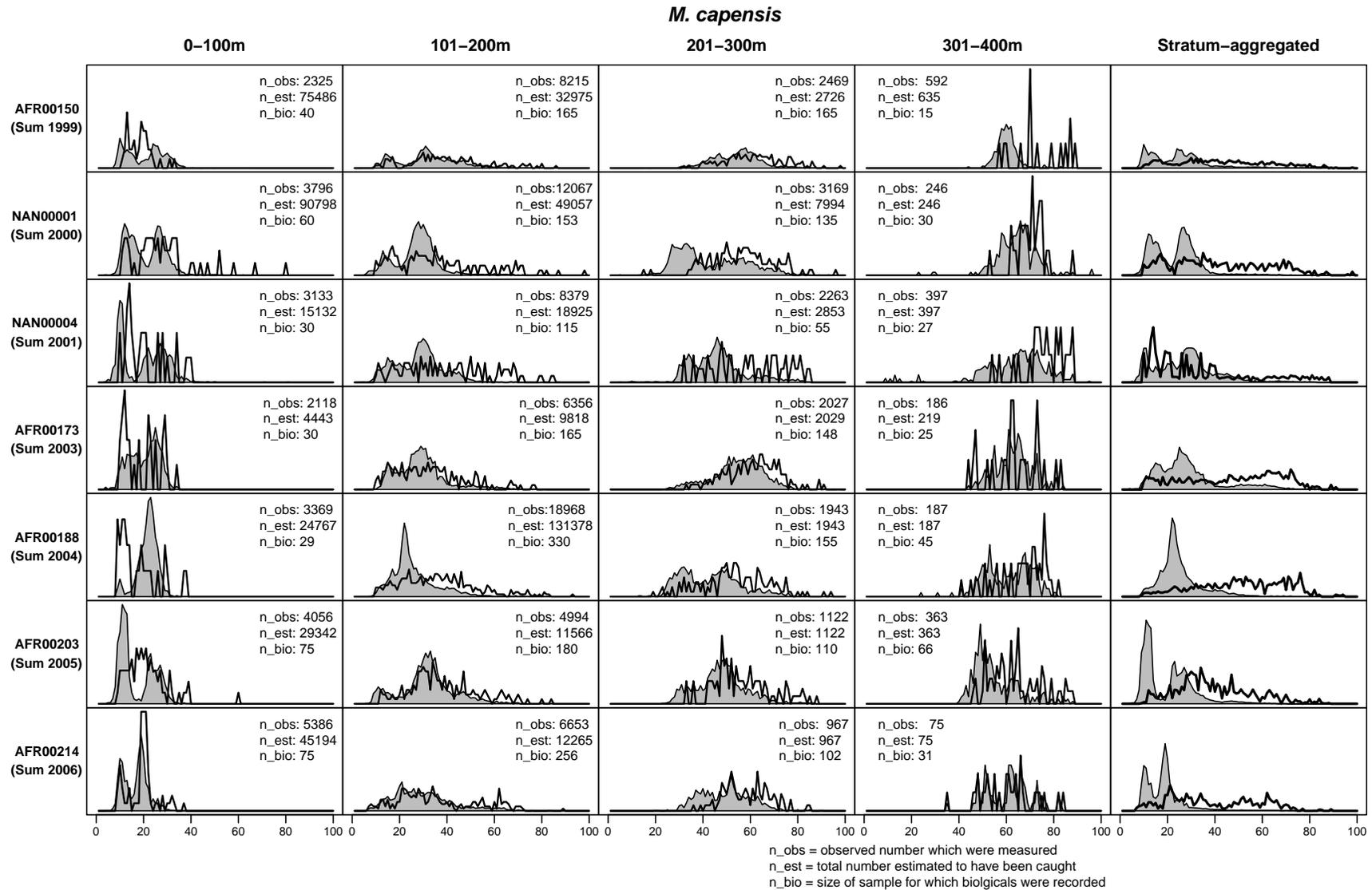
Figure 6a: Comparison of **M. paradoxus** survey probability distributions for length without density weighting for $X = 100$ (grey shaded area) with the length-distribution of the sub-sample of the survey catch for which biologicals were measured (solid black lines). The distributions indicated by the grey shaded area are the same as those for $X = 100$ in Figure 1(a). Distributions are given by depth stratum for all cruises in Figures 6a and 6b. The vertical axes here are to the same scale. Note that the stratum-aggregated proportions were calculated using the methodology of Sen 1.
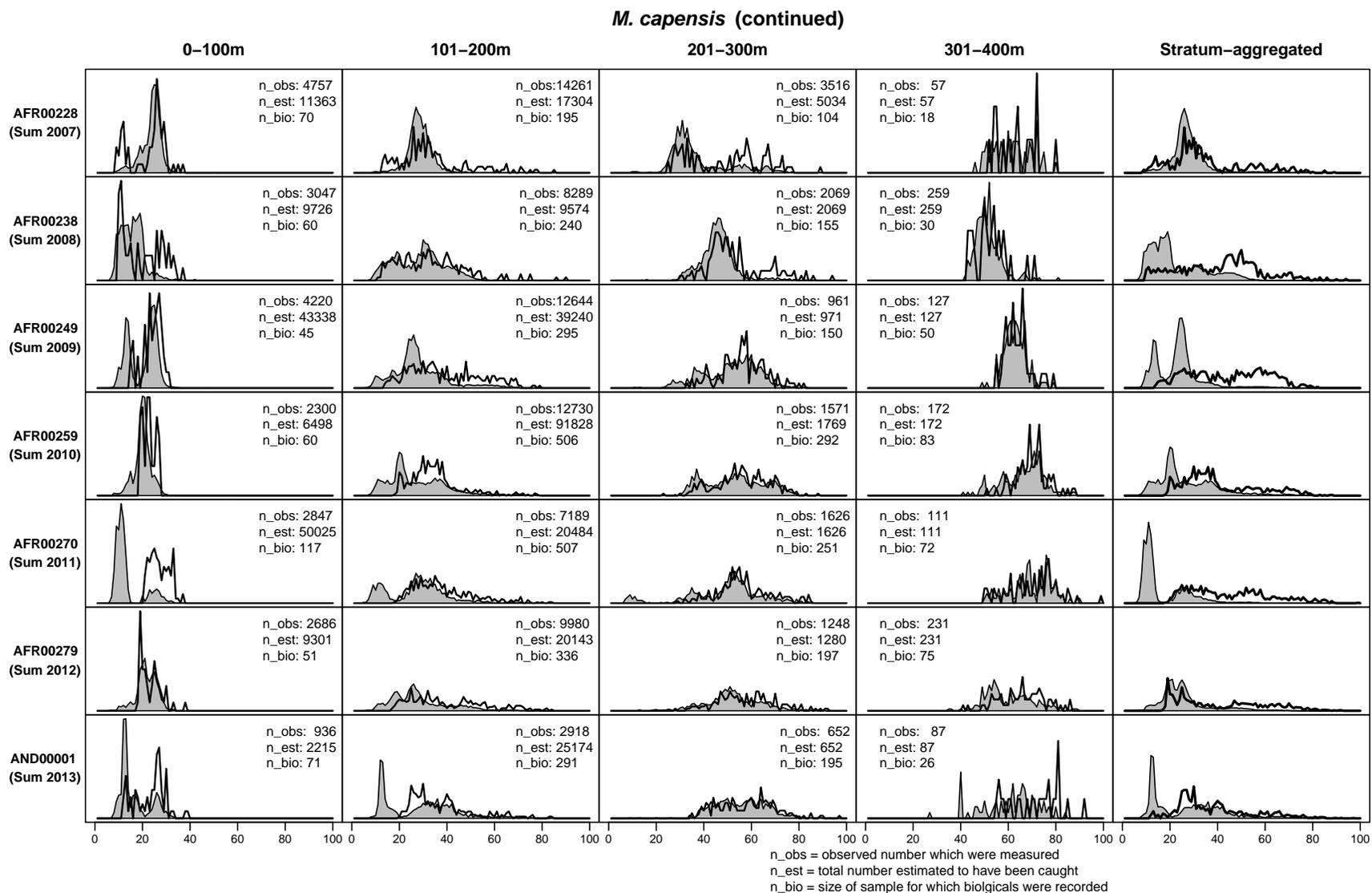
Figure 6b: Continuation of Figure 6a.

Figure 7a: Comparison of *M. capensis* survey probability distributions for length without density weighting for $X = 100$ (grey shaded area) with the length-distribution of the sub-sample of the survey catch for which biologicals were measured (solid black lines). The distributions indicated by the grey shaded area are the same as those for $X = 100$ in Figure 1(b). Distributions are given by depth stratum for all cruises in Figures 7a and 7b. The vertical axes here are to the same scale. Note that the stratum-aggregated proportions were calculated using the methodology of Sen 1.

M. capensis (continued)

n_obs = observed number which were measured
n_est = total number estimated to have been caught
n_bio = size of sample for which biolgicals were recorded

Figure 7b: Continuation of Figure 7a.
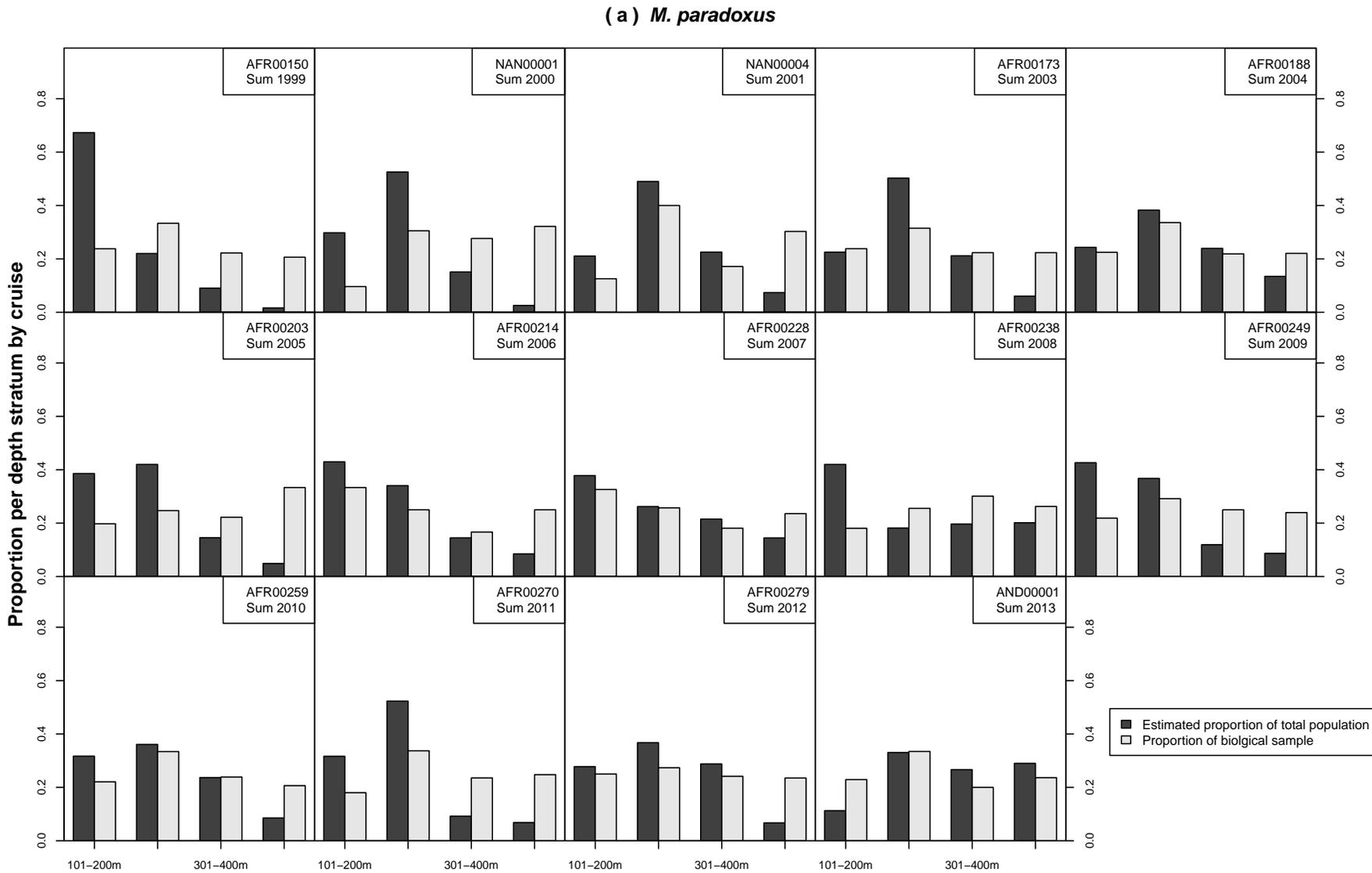
(a) *M. paradoxus*

Figure 8a: *M. paradoxus* proportion of the total biolgical sample size per depth stratum (light grey bars) plotted in contrast to the proportion of the survey estimate of population size attributed to each stratum (dark grey bars). The population estimate here is the product of the estimated density and the area of each stratum. Results are shown for all cruises for which biological data have been provided. Population estimates are not available for the 501-1000m depth stratum as this stratum was not sampled every year within the sampling period.
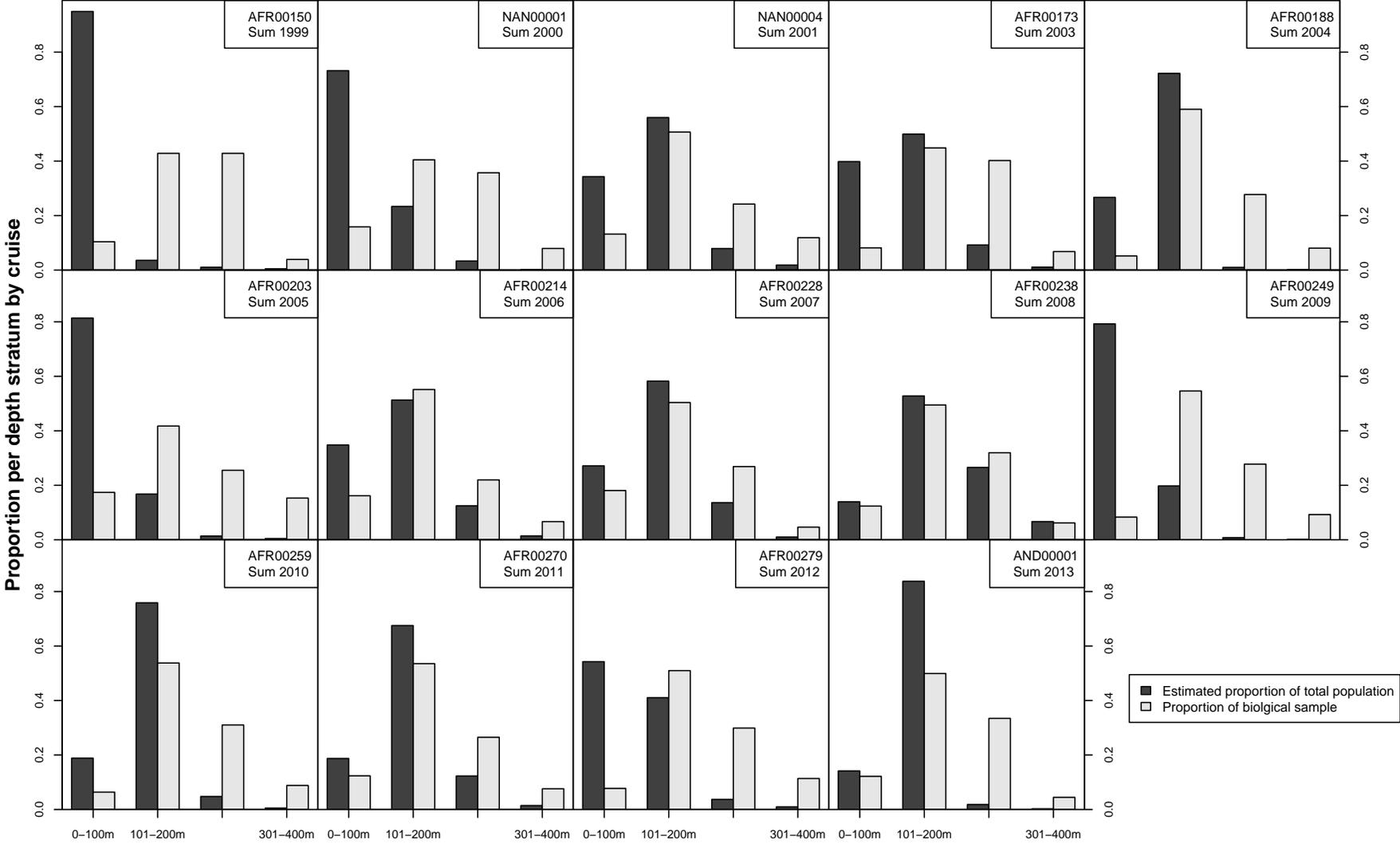
(b) *M. capensis*

Figure 8b: *M. capensis* proportion of total biolgical sample size per depth stratum (light grey bars) plotted in contrast to the proportion of the survey estimate of population size attributed to each stratum (dark grey bars). The population estimate here is the product of the estimated density and the area of each stratum. Results are shown for all cruises for which biological data have been provided.
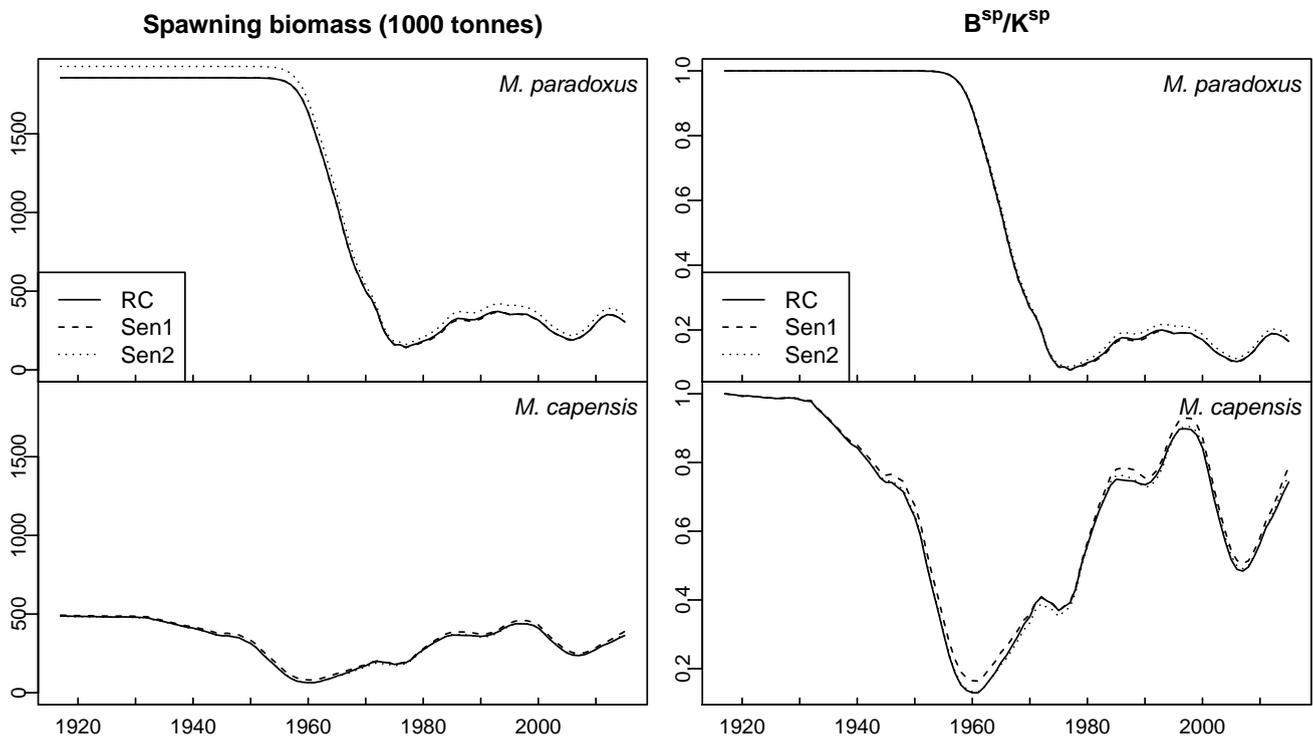
Figure 9: *M. paradoxus* and *M. capensis* spawning biomass shown in terms of absolute value and relative to pristine spawning biomass. Results are given for the three cases described in Equation 4.8 and immediately thereafter.

RC: No density-weighting; survey estimated population size ($N_{y,d}^{s_p}$ from equation 4.8) is calculated from the survey estimate of spawning biomass. This is the methodology currently in use.

Sen1: No density-weighting; survey estimated population size $N_{y,d}^{s_p}$ is calculated directly from the survey samples.

Sen2: Density-weighting; survey estimated population size $N_{y,d}^{s_p}$ is calculated directly from the survey samples.
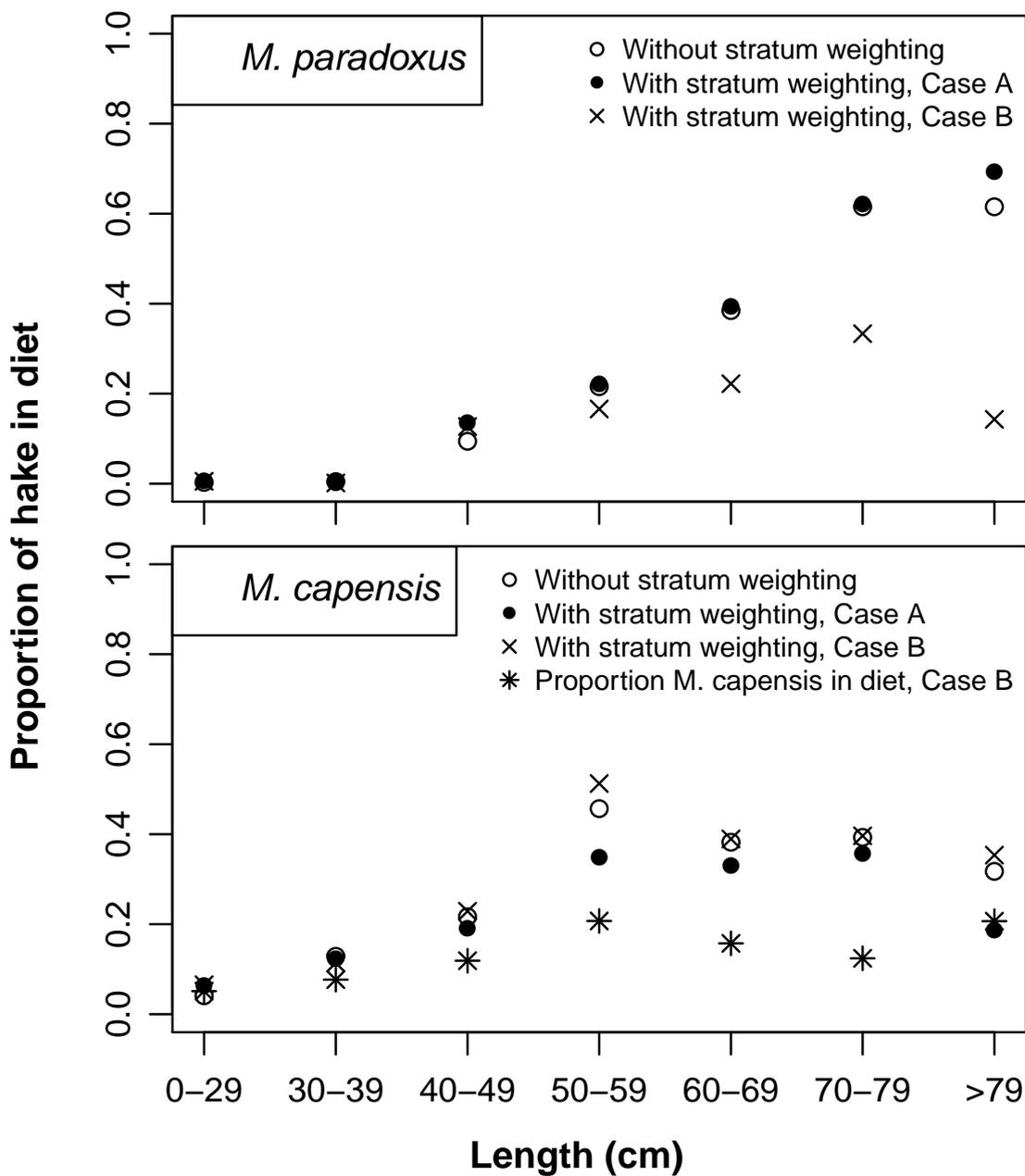
Figure 10: Proportion of hake in diet, given by $\sum_y \widetilde{h}_y^{s_p l_p} / \sum_y \widetilde{n}_y^{s_p l_p}$ for three cases: When stratum weighting is not taken into account (open circles), when samples from each stratum are weighted by the population size in terms of numbers (filled cirlces, Case A) and when the samples from each stratum are weighted by the populations size in terms of numbers-at-length (crosses, Case B).