A Record of the Generation of Data to be Used in the Anchovy Assessment

Cunningham, C.L.[#] and van der Westhuisen, J.J.[*]

The data to which the South African anchovy assessment is tuned are not raw data. Some of the data have already been subject to a number of analyses and refinements. These associated calculations are often done "behind the scenes" and their details are seldom recorded. This lack of record can result in a discontinuity in the method used to calculate data for subsequent assessments, particularly if assumptions made in the calculations are not documented and/or a new person becomes responsible for developing the data to be used for input to the assessment. This document serves as a first step to record the generation from the raw data of the data to be used in the anchovy assessment to be carried out this year, and will at a later stage be updated similarly for the sardine assessment and any further data used in the development and testing of the pelagic OMP.

Anchovy catch data

Monthly Raised Length Frequencies (RLFs)

Monthly raised length frequencies were constructed for the anchovy landings using the method in Appendix A. From 1987-2006, RLFs are available by Western (west of Cape Agulhas), Southern (Cape Agulhas to Cape St. Francis) and Eastern (east of Cape St. Francis) areas. Between 1984 and 1986, 31 pelagic catch positions were recorded outside of the Western area. With the exclusion of one catch position (5103 in pool area 20, see Figure A.2) in the Southern area (only just outside the Western area) in 1986, all the recorded positions were east of East London. The only boats that could possibly have fished in those areas during the 1980s would have been small bait boats targeting sardine only. The landings recorded outside of the Western area were from boats that were all equipped with power blocks and suction pumps and they would not have been able to land fish of the quantities in question on the east coast. These boats were bigger than the bait boats and it is therefore highly unlikely that these catch positions are correct. Apart from one digit the four number positions are all the same as positions on the west coast, leading us to believe that these were punch errors. The one throw in 5103 in 1986 was most probably sardine, not anchovy, and thus all anchovy catch from 1984 to 1986 was assumed to occur in the Western area.

[#] MARAM (Marine Resource Assessment and Management Group), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701, South Africa. Email: c.l.cunningham@telkomsa.net, dll@maths.uct.ac.za.

[*] Marine and Coastal Management, Private Bag X2, Rogge Bay, 8012, South Africa.

In 7 months no length frequencies were available although there were landings. In these cases the length frequencies of former months were used to estimate a raised length frequency as follows:

$$RLF_{y,missing,l} = RLF_{y,previous,l} \times Tonnage_{y,missing} / Tonnage_{y,previous}$$

The "former" month used in this estimation is listed in the below table.

Year	Month for which length frequency was missing	Tonnage landed in missing month	Area in which landings occurred	Month from which length frequency was used	Tonnage landed in this used month
1984	October	22 878t	Western	July 1984	18 193t
1984	November	7 281t	Western	July 1984	18 193t
1990	August	215t	Western	July 1990	558t
1990	September	34t	Western	July 1990	558t
1993	December	64t	Western	November 1993	7t
1996	November	18t	Western	October 1996	21t
2005	December	27t	Western	November 2005	1950t

The RLFs by month from 1984 to 1986 and by area from 1987 to 2006 are stored in *Anchovy RLFs with Cut-Offs.xls*.

Splitting Juvenile and Adult Catch

The following cut-off lengths (Cunningham and Butterworth 2007) were applied to each month and area to calculate the number of juveniles and adults from 1984 to 2006:

Month	Cut-off length
January	7cm
February	8cm
March	9cm
April	9.5cm
May	10cm
June	10.5cm
July	10.5cm
August	10.5cm
September	10.5cm
October	10.5cm
November	5cm
December	6cm

Monthly anchovy catch numbers are available for 1981 to 1983 (De Oliveria pers comm.) but no RLFs are available for these months and so a cut-off length cannot be applied to split the catches between juveniles and adults. For these 4 years, therefore, the adult catch is assumed to consist of all landings between November and March and the juvenile catch is assumed to consist of all landings between April and October. This is the method used for all years prior to the introduction of the above cut-off lengths (Cunningham and Butterworth 2004).

The resulting monthly catch numbers of juveniles and adults, summed over all areas, are stored in *Anchovy Commercial Catch.xls*. The annual juvenile and adult anchovy catch for year *y* is calculated as the sum over all months from November *y*-1 to October *y*. The annual juvenile and adult anchovy catch data are given in Table 1 and stored in *Anchovy Commercial Catch.xls*.

Catch Weight

The data available for these calculations include the number of fish in length class *l* in month *m* in area *a*, $N_{l,m,a}$, (used above) and the observed tonnage in month *m* in area *a*, $ObsT_{m,a}$ from 1984 to 2006. These data are recorded in *Anchovy RLFs with Cut-Offs.xls*. The length-weight relationship used is (Lynne Shannon pers comm. Using 1990-1996 data):

$mass = 0.00750 \times L_c^{3.110}$

Expected mass by length class, area and month is calculated as: $EM_{l,m,a} = 0.0075 \times l_{mid}^{3.110} \times N_{l,m,a}$

Where l_{mid} is the mid-point of the length class considered.

Adjusted mass by length class, area and month is calculated as: $AM_{l,m,a} = \frac{EM_{l,m,a}}{\sum_l EM_{l,m,a}} \times ObsT_{m,a}$

Average monthly adjusted mass by length class, area and month is calculated as:

$$\overline{AM}_{l,m,a} = \frac{AM_{l,m,a}}{N_{l,m,a}} = \frac{\frac{EM_{l,m,a}}{\sum_l EM_{l,m,a}} \times ObsT_{m,a}}{N_{l,m,a}}$$

Average juvenile mass by month for the total area is calculated as: $M_m^{juv} = \frac{\sum_a \sum_{l<cutoff} \overline{AM}_{l,m,a} \times N_{l,m,a}}{\sum_a \sum_{l<cutoff} N_{l,m,a}}$

Average adult mass by month for the total area is calculated as: $M_m^{ad} = \frac{\sum_a \sum_{l\geq cutoff} \overline{AM}_{l,m,a} \times N_{l,m,a}}{\sum_a \sum_{l\geq cutoff} N_{l,m,a}}$

A check is performed on the calculations such that: $M_m^{juv} \times \sum_a \sum_{l<cutoff} N_{l,m,a} + M_m^{ad} \times \sum_a \sum_{l\geq cutoff} N_{l,m,a} = \sum_a ObsT_{m,a}$.

The above calculations and average juvenile and adult anchovy catch mass by month are stored in *Anchovy RLFs with Cut-Offs.xls*.

The annual average juvenile and anchovy catch mass are calculated using a weighted average: $\frac{\sum_m M_m^{juv} \times N_m^{juv}}{\sum_m N_m^{juv}}$ and $\frac{\sum_m M_m^{ad} \times N_m^{ad}}{\sum_m N_m^{ad}}$, where N_m^{juv} and N_m^{ad} are the monthly juvenile and adult catch-at-age reported in Table 1. These sums are taken over the months November *y*-1 to October *y*, except for 1984 when the sum is from January to October 1984. The annual values are given in Table 1 and stored in *Anchovy Commercial Catch.xls*.

Between 1981 and 1983 there were no data to calculate catch weights-at-age as above and thus annual catch weight-at-age for juveniles between 1981 and 1983 and for adults between 1982 and 1983 was taken from De Oliveria (2003).

Juvenile catch prior to the survey

RLFs were also calculated from the first of the month in which the annual recruit survey took place to the day before the commencement of the survey using the method in Appendix A. Inspector data (which includes samples for species split) is required to do this (see Appendix A), but was not available in 1985 and 1986. Daily skippers' estimates of tonnage landed were, however, available for these years. Although the total tonnage landed in May 1985 and June 1986 was estimated by the skippers to be different to that arising from the source data, it was assumed that the proportion of catch taken before the survey compared to the whole month was the same between the skippers' estimates and the source data. Thus RLFs for 1-19 May 1985 and 1-9 June 1986 were calculated as follows: $N_{l,\,partmonth,a} = N_{l,\,fullmonth,a} \times SkipperT_{partmonth} \,/\, SkipperT_{fullmonth}$, using the data in the below table.

	Days for which catch is required	Catch for the month (tons)	Skipper estimated catch for the month (tons)	Skipper estimated catch prior to the survey (tons)
May 1985	1-19th	74245	77174	48396
June 1986	1-9th	64662	68189	10338

The cut-off length method described on page 2 was applied to calculate the number of juveniles landed in the month prior to the commencement of the survey. The associated average juvenile catch weight was also calculated using the method detailed on page 3. The total juvenile catch prior to the survey was then summed over all months from November *y*-1 to the day prior to the commencement of the survey. The average juvenile mass in this catch was calculated as a weighted average, taking the number of juveniles caught in each month into account. These data are given in Table 2 and are available together with the necessary calculations in *Anchovy RLFs with Cut-Offs.xls* and *Survey Data.xls*.

Summary

This document provides a record of all catch data to be used in the anchovy assessment this year. Survey indices of abundance are provided in Cunningham *et al*. (2007). This document will be updated with similar data for the sardine assessment at the next PWG meeting.

References

Cunningham, C.L. and Butterworth, D.S. 2004. Base Case Bayesian Assessment of the South African Anchovy Resource. MCM document WG/PEL/APR04/01. 19pp.

Cunningham, C.L. and Butterworth, D.S. 2007. Proposed Cut-Off Lengths to Split Recruits and Adults for Anchovy Commercial Landings. MCM document MCM/2007/FEB/SWG-PEL/08. 15pp.

Cunningham, C.L., Butterworth, D.S. and Coetzee, J. 2007. The Estimation of Uncapped Acoustic Survey Biomass from Capped Data. MCM document MCM/2007/MAY/SWG-PEL/08 14pp.

De Oliveira, J.A.A. 2003. The Development and Implementation of a Joint Management Procedure for the South African Pilchard and Anchovy Resources. PhD Thesis, University of Cape Town, South Africa.

Table 1. Annual juvenile and adult anchovy catch (in numbers) and mean catch weight (in kilograms). Annual data for year y consists of data from November y-1 to October y, as described in the text above.

	Annual anchovy catch		**Annual anchovy catch weight**	
Year	**0 year olds**	**1 year olds**	**0 year olds**	**1 year olds**
1981	178633	113316	0.0079	-
1982	199079	107082	0.0058	0.0108
1983	164121	27425	0.0070	0.0106
1984	29987537	9416485	0.0057	0.0102
1985	32687599	8544017	0.0057	0.0111
1986	50114319	6250229	0.0045	0.0116
1987	28038404	34024541	0.0065	0.0123
1988	48450985	21236966	0.0057	0.0138
1989	19000666	14283375	0.0064	0.0123
1990	32169066	1117853	0.0043	0.0120
1991	24742109	1474539	0.0055	0.0100
1992	59420844	7873901	0.0043	0.0122
1993	31856839	9228806	0.0041	0.0115
1994	21611587	5469886	0.0044	0.0113
1995	40036305	1631826	0.0040	0.0093
1996	6141948	1417886	0.0048	0.0093
1997	12014815	60026	0.0050	0.0130
1998	21877746	763655	0.0045	0.0111
1999	35061348	428159	0.0050	0.0110
2000	45940811	2839358	0.0051	0.0114
2001	55658108	2651615	0.0047	0.0096
2002	43361634	3339933	0.0042	0.0104
2003	62090898	1167115	0.0039	0.0117
2004	39136380	1604959	0.0045	0.0090
2005	32838364	8917360	0.0058	0.0105
2006	29487772	1330591	0.0041	0.0109

Table 2. Juvenile anchovy catch (in numbers) and mean catch weight (in kilograms) from 1 November y-1 to the day before the annual recruit survey in year y.

Year	Date of commencement of survey	Time of the recruit survey after 1 May	Juvenile catch prior to the survey	Mean juvenile catch weight prior to the survey
1985	20-May	0.613	14446081	0.0058
1986	10-Jun	1.300	21077845	0.0074
1987	20-Jul	2.613	13610181	0.0067
1988	27-Jun[1]	1.867	12445201	0.0054
1989	08-Jun[2]	1.233	12420888	0.0069
1990	22-Jun	1.700	31131308	0.0043
1991	07-May	0.194	12327687	0.0054
1992	13-May	0.387	12865144	0.0039
1993	21-May	0.645	1211617	0.0058
1994	05-May	0.129	4234179	0.0041
1995	10-Jun	1.300	12511225	0.0044
1996	05-Jun	1.133	4051491	0.0050
1997	17-May	0.516	166349	0.0065
1998	20-May	0.613	6083460	0.0051
1999	10-May	0.290	1843042	0.0052
2000	15-May	0.452	8120212	0.0061
2001	05-May	0.129	5802894	0.0058
2002	05-May	0.129	1620008	0.0062
2003	14-May	0.419	3066935	0.0049
2004	08-May	0.226	3870663	0.0056
2005	13-May	0.387	4292109	0.0064
2006	19-May	0.581	907536	0.0051

[1] The first station was on 27th June 1988, although the first acoustic interval was only logged after midnight, i.e. on 28th June 1988.

[2] The first station was on 8th June 1989, although the first acoustic interval was only logged after midnight, i.e. on 9th June 1989.

Appendix A: Pelagic sample allocation

The sample allocation method is the process whereby a length frequency is allocated to every commercial landing, enabling the transformation of the catch to its raised length frequency (RLF). The commercial catch data and field station length frequency data are entered and stored on a Sybase database on the MCM network and the calculations are performed in Access.

Species

For the assessments which serve as the operating models to test Operational Management Procedures it is necessary to calculate RLFs for anchovy (*Engraulis encrasicolus)* and sardine (*Sardinops sagax)* though RLFs for round herring (*Etrumeus whiteheadii)* and horse mackerel (*Trachurus trachurus capensis)* are also generated for every run.

Data sources

- Commercial catch: The skipper completes a skipper form for every trip and records the estimated catch and the geographic position of individual throws. The scale monitor contract was awarded to Nosipho Consultants in 2002. They sample every landing for its species composition and tonnage landed. Prior to 2002 this was the task of the fisheries inspector and hence the catch sheet is referred to as the inspector's form. Skipper data are available on Sybase from 1984 onwards but inspector data were obtained only from 1987. MCM field station personnel collect data sheets and enter the information on Sybase.
- Field station samples: MCM field station personnel collect random samples at the major pelagic fishing harbors for species composition and length frequency (Capricorn fishing was contracted from 2002 until 2005 to man St. Helena Bay and Gansbaai). Samples of industrial fish like anchovy and round herring are obtained from the top of the hold before the vessel discharges. For this reason industrial samples are obtained mainly from the last throw of the trip. Offloading further damages the already semi-decomposed fish and one cannot sample from the conveyer belt because it would be impossible to weigh those fish. Directed sardine catch, on the other hand, is kept in a very good condition onboard on ice and good quality samples are easily obtained from the conveyor belt, whilst the vessel is discharging. Unfortunately it is seldom possible to establish which throw is being sampled. Field station data are available on Sybase from 1984 onwards. Ports sampled over the period include Lamberts Bay, Laaiplek, St. Helena Bay, Saldanha, Cape Town, Hout Bay, Kalk Bay, Hermanus, Gansbaai, Mossel Bay and Port Elizabeth.
- Observer samples: The observer program started in 1999 but onboard biological sampling was only started in 2001. Observer sampling is an improvement on the field station data because samples are obtained from a known throw, all throws are sampled and the fish is always in a good

condition. Unfortunately the length frequency samples have to be taken ashore for weighing and this gives rise to room for error. The data are stored in an Access database called CAPFISH.

Data extraction from Sybase

- Catch data are extracted from Sybase as text (flat) files; *throw.csv* contains the skippers' data and *catch.csv* contains the inspectors' data.
- Field station data are extracted in the same manner; *spcomp*.c*sv* contains the species composition data and *lfreq.csv* contains the length frequency data.

Data handling and evaluation

MCM data

- Unfortunately there is no manual proof reading of all the data, except in cases where the number of throws is excessive (more than 10) and the trip duration is of an unrealistic duration (more than 3 days). Data evaluation is limited to electronic checking for noticeable mistakes.
- A duplicate dataset of *catch.csv* which is regularly updated by email is kept at Saldanha in an Access table. This means that the data are entered twice, but into separate databases and this allows for the comparison of the two data sets on a regular basis for differences and errors. It might appear unnecessary to keep two data sets, but this is the sole reason that the pelagic catch data is still representative of what was recorded by the scale monitors.
- The expected sample weights associated with the length frequency data in *lfreq.csv* are computed and samples that deviate more than 30% are flagged and checked against the raw data. If a flag results from a punch error then the data are corrected, but in the case of a sampling error the record is deleted from the data base.
- Suspect positions, for example areas outside the normal catch areas are checked against the raw data and, if necessary, corrected.

Observer data

- Limited manual proof reading of data
- Only observer trips that match the commercial data for vessel name and date are used. Mismatched dates do occur, making it very difficult to establish whether a specific vessel carried an observer on a specific date. Therefore samples from such observer trips are ignored to prevent the inclusion of poor data. Only trips that do link can be used, because the scale monitor's species composition is used to determine the target species of the length frequency sample.
- The structure of the observer length frequency table is altered to make it compatible with the Sybase dataset.

- Only observer length frequencies whose predicted sample weights fall within the set range are used. Data with possible measurement errors or wrong species names are excluded.

Access programs

1) Capfish.mdb (observer data)
2) RLFdata.mdb (where the RLFs are generated)

General program outline

- Catches are allocated to pool-area/week strata:
    1. Week: the throw date with the largest catch is used.
    2. Pool area: the existing 21 areas (see Figure A.2) are used, but in 1999 area 21 was subdivided into areas 23 and 24, to accommodate the eastward fishing expansion. The throws within each landing are examined, and the throw with the greatest mass is used as the representative throw.
    3. Assign a target species to every catch. The species with the largest mass is defined as the dominant species in the landing.
- The length frequency samples are grouped by species and target species for the pool-area/week strata and summed.
- A new catch table with additional space for the allocated length frequencies is created.
- The length frequency table is searched and a frequency based on the species, target species, week and pool area criteria are assigned to the catch table.
- In the event of catches not being represented by an appropriate sample, the pool-area/week will be expanded to include surrounding areas and weeks. Stratum expansion continues alternately by week and pool until an appropriate frequency is located.
- If no appropriate sample is found then the average sample for the month is applied. Where no sample for the month exits in the case of anchovy, the raised length frequency is estimated using the raised length frequency of a former month as detailed in the text. Where no sample for the month exists in the case of sardine, the previous month is used. Catches of each species and the length frequencies are summed by month over larger user specified areas.
- The RLFs are exported as Excel files in numbers per length group.

The user specified areas that are used are:

1. Areas 1-6: North of Cape Columbine
2. Areas 7-12: Cape Columbine to Cape Point
3. Areas 13-20: Cape Point to Cape Infanta
4. Area 23: Cape Infanta to Plettenberg Bay

5. Area 24: East of Plettenberg Bay

In 2007 three new areas were introduced because of planned changes to the OMP:

1. West: West of 20 degrees east (West of Cape Agulhas)
2. South: East of 20 degrees east and west of 24 degrees 50 minutes east (between Cape Agulhas and Cape St. Francis)
3. East: East of 24 degrees 50 minutes east (East of Cape St. Francis)

Although the RLFs are summarized according to different areas, the allocation process is still based on the original pool areas, with the exception of those cases where pool areas were split by the new borders.

Program changes

In January 2007 four changes were made to the process above:

- The observer length frequencies were included.
- To prevent juvenile sardine frequencies from being allocated to adult sardine catches, the species was separated into directed and by catch for allocation purposes. This is applicable only when sardine is landed as a by catch with anchovy. Sardine by catch with anchovy is mainly juvenile fish whereas by catch with round herring it is mostly adult fish.
- Noticeable error in the RLF results when the field station catch composition data are used to identify the target species of the length frequency sample, and these composition data differ from those of the scale monitor. Because the field station data are not proofread, and given the inclusion of the observer length frequencies (they also need a target species to be identified), it was decided to standardize on the scale monitors species composition as the only source.
- Missing skipper data (catch area) are catered for. This occurs when the skipper fails to hand in a trip sheet. Currently this is not a major problem but it did happen in the 1980s and 1990s. Where the *catch.csv* file does not have a related record in the *throw.csv* file, the program will search for the most likely catch position, based on the catch type of the other vessels for the same date.

The first change leads to enhanced coverage, especially in the case of industrial fish, i.e. anchovy that are poorly sampled by the field stations. The last three changes were implemented to prevent errors caused by bad data or poor sampling coverage. This can typically be seen in a RLF plot as an improbable peak at a certain length group.

In March 2007 an additional change was implemented. Towards the end of the year sporadic landings can be overlooked, because it is not cost effective to continue extensive sampling. These landings are generally small but it is still necessary to allocate a size to the fish. In the past the annual RLF average was used, but it was felt that it is better to allocate the length frequency from the adjacent month. The

length frequencies are first stratified by area and species type, but where no match is found the requirements for matching area and target species are removed alternatively until a match is found.

Even though throws in multiple pool areas during a single trip do occur, only the catch area for the biggest throw is selected. This is done in order to keep continuity with the old sample allocation method. A change that could be considered would be to allocate a sample to every throw as opposed to every trip. The scale monitor samples at regular intervals and discrete throws are not sampled. However, if one assumes the species composition of the throws are uniform, then the catch per throw can be calculated, by proportionally applying individual throws to the species composition. Observer sampling is ideally suited for this method, because every throw is sampled, but greater sampling coverage and matched skipper throws are required.

Sampling coverage required

Optimum sample size and sampling coverage can only be determined using a suitable statistical study, and one can therefore only speculate on the sample size required. Logistic constraints have necessitated a random stratified sampling method, and the grouping of catches and samples on a week/pool-area basis has been adopted since electronic data processing began. Both the sampling and the raised length frequency approaches are arguably the most suitable considering the fishing strategy and the available data. The percentage coverage per stratum is readily quantified, and the first level pool-area/week coverage could possibly be used as an index of sampling coverage. 100 percent coverage is not attainable because of financial and logistic constraints, and it is more than likely unnecessary. From Figure A.1 it appears that 80 percent is attainable when the field station and observer samples are combined.

Many factors influence the relationship between the number of samples taken and the coverage obtained, but in general more samples will lead to better coverage. This partially explains the declining trend of the field station data in Figure A.1. Directed sardine samples are easily obtained but industrial fish have to be collected from the hold of the boat, a difficult and unpleasant task. The numbers of buckets to be taken at the field stations are prescribed, but when a decision has to be taken on the fish type by the field station worker, then the ice fish is favoured more often than not. Directed sardine from all areas (except Port Elizabeth) are processed at the canneries in the St. Helena Bay area and because the field station is regularly manned, good coverage was attained. Erratic sampling at Saldanha Bay, Hout Bay and Gansbaai also contributed to the decrease of industrial fish coverage. With the inclusion of observer samples however, the target percentage is reached for anchovy and juvenile sardine by catch. If 80 percent is a realistic benchmark, then one can then conclude that the sampling effort (regarding TAC species) for the time period 2001 to 2006 was adequate. It has to be stressed that this was achieved only with the inclusion of samples from the observer program.

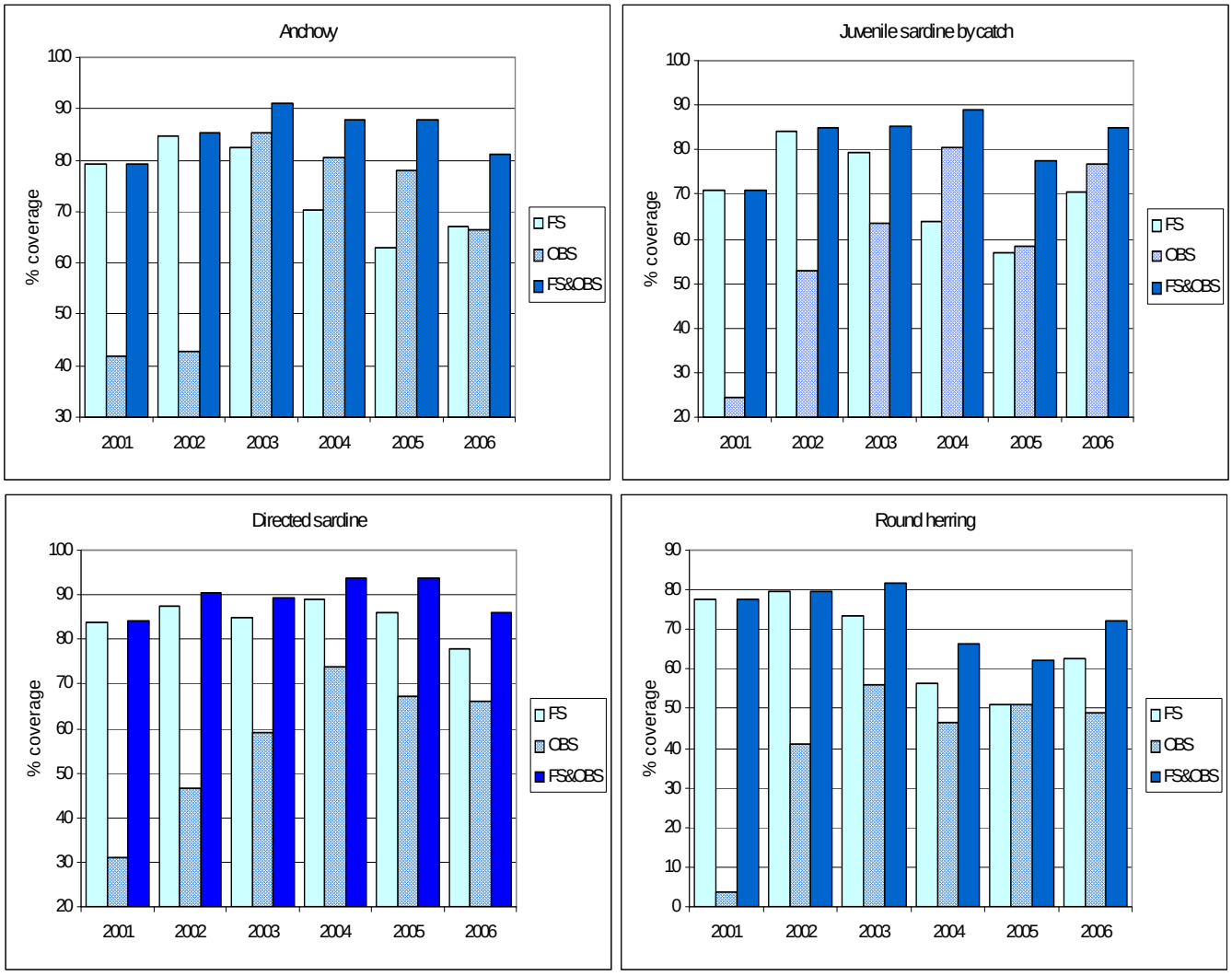

Figure A.1. Coverage obtained on a first level pool-area/week for the field stations (FS), the observers (OBS) and a combination of the two (FS&OBS).

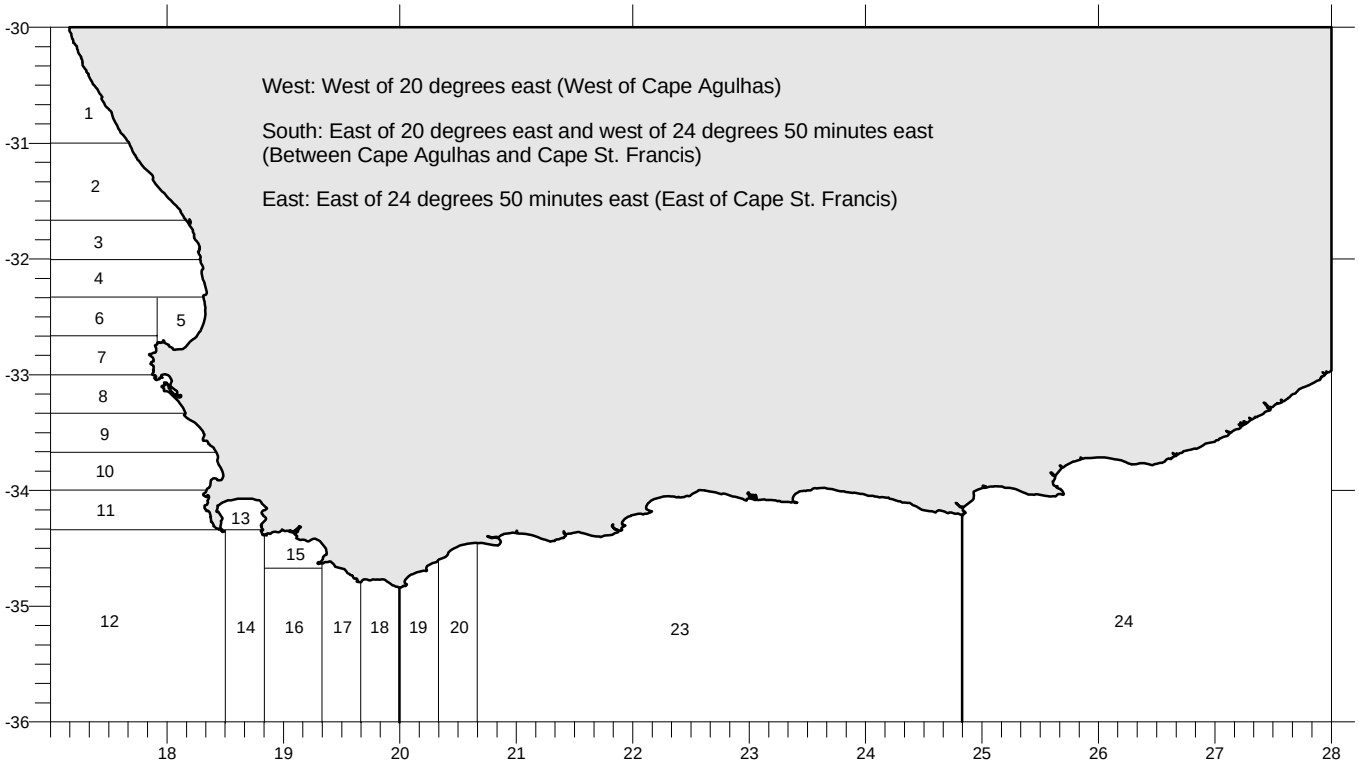

Figure A.2. The pool areas that are used for sample allocation and the 3 larger areas that are used for the OMP revision.