The data used in the 2020 anchovy assessment

de Moor, C.L.[#] , Merkle, D.[+], and Coetzee, J.[+]
Correspondence email: *carryn.demoor@uct.ac.za*

Data to which the 2020 assessment of the South African anchovy resource is to be conditioned are listed.

The data to which the South African anchovy assessment is conditioned are not all raw data. Some data have already been subjected to 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 there is a change in personnel responsible for developing the data to be used for input to the assessment. This document serves to record the generation of the data used in the anchovy assessment carried out in 2020 from the raw data.

Information on Ageing

Anchovy age-length keys for the November surveys from 1992 to 1995 were derived by Prosch (De Oliveria 2003), and by Kerstan for 1990, 1992 to 1995 (unpublished data). These data are not used in the assessment.

Anchovy Commercial Data

Monthly anchovy catch numbers are available for 1981 to 1983 (De Oliveria pers. comm.) but no LFs are available for these months. These data are not used in the assessment.

Monthly length frequencies were constructed for the anchovy landings using the method in Appendix A. Although it is possible to split the length frequencies by area from 1987, as the assessment will be run for a single stock in a single area, length frequencies for a single area only are considered. A few 15cm fish were recorded as anchovy landings in September/October 2001, but these were excluded from the assessment as this is most likely a mistake in the reported species. Anchovy catches are mostly inshore and the possibility of landing a 15cm anchovy is small.

In some 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:

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

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

[#] MARAM (Marine Resource Assessment and Management Group), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701, South Africa.
[+] Department of Environment, Forestry and Fisheries – Branch Fisheries, Private Bag X2, Vlaeberg, 8018, South Africa.

The LFs by month and year, i.e. the number of fish in length class l in month m of year y, $N_{y,l,m}$, from 1984 to 2019 are stored in *AnchovyCatchLength.csv*, together with the observed tonnage in each month and year, $ObsT_{y,m}$.

Expected mass by length class and month is calculated as:

$$EM_{y,l,m} = 0.008157 - 0.003501 \sqrt{\frac{1}{2\pi \times 1.9304^2}} e^{-\frac{1}{2}\left(\frac{m-8.0616}{1.9304}\right)^2} \times l_{mid}^{3.0979} \times N_{y,l,m}$$

where l_{mid} is the mid-point of the length class considered and mass is in grams and length in centimetres. The monthly length-weight relationship is a time-invariant relationship calculated from commercial data collected over 1990-1996 (model viii) of de Moor and Butterworth 2015). The "raised" length frequency used in the assessment, is then calculated as follows: $\hat{N}_{y,m,l} = N_{y,m,l} \times \frac{ObsT_{y,m}}{\sum_i EM_{y,l,m}}$.

Annual data used in the assessments are taken over the months November $y-1$ to October y, except for 1984 when the sum is from January to October 1984. The monthly catch tonnage is shown in Table 1.

Recruit catch prior to the survey

Length frequencies were also calculated from 1 May to the day before the commencement of the survey using the method in Appendix A. Inspector data (which include samples for species split) are required to do this (see Appendix A), but were 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 length frequencies for 1-19 May 1985 and 1-9 June 1986 were calculated as follows: $N_{l,partmonth} = N_{l,fullmonth} \times SkipperTonnage_{partmonth}/SkipperTonnage_{fullmonth}$ 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 LFs of catch from 1 May to the day before the survey are also stored in *AnchovyCatchLength.csv*, together with the corresponding observed tonnage. The "raised" length frequency used in the assessment is calculated using the same method as described above.

The number of juveniles landed between 1 May and the day before the commencement of the survey is calculated within the model from these data using selectivity-at-length and length-at-age distributions.

November Survey Data

The time series of total biomass estimates and associated CVs from the acoustic surveys in November each year, corresponding to the standard survey area between Hondeklip Bay and Port Alfred, are given in Table 2. Daily egg production method (DEPM) estimates of anchovy adult biomass between 1984 and 1993 are also available (Table 2). For assessment purposes we assume this corresponds to spawning biomass. These data are stored in *SurveyData.xls*. The early part of this time series was updated

by de Moor et al. (2008) so that the full time series corresponds to uncapped survey biomass determined with a new target strength formula (Coetzee et al. 2008).

Length frequencies are also available for the hydroacoustic surveys and are stored in *AnchovySurveyLength.csv*.

Recruit Survey Data

The time series of recruit biomass and associated CVs from the May/June recruit surveys is given in Table 3. The recruit numbers at the time of the survey were calculated by summing the number of fish smaller than a cut-off length in the weighted length frequency (as per Method 2 of Appendix B). The average recruit weight is calculated by applying a length-weight regression to the weighted length frequency. This mean weight is then adjusted by the difference between the two biomasses (Method 1 of Appendix B). This calculated biomass and average recruit weight were calculated in a separate database, using the uncapped density per interval as input. The two biomass series are not identical due to the different methods of weighting used. A brief description of the two methods is given in Appendix B. Although not ideal, this difference has been narrowed from what has previously been used. This is a matter that needs to be addressed at some stage[1]. In the assessments, the recruit numbers are used together with the CVs on recruit biomass. These survey data are stored in *SurveyData.xls*. The early part of this time series was updated by de Moor et al. (2008) and revised again in 2011 so that the full time series corresponds to uncapped recruitment determined with a new target strength formula (Coetzee et al. 2008). Adjustment is made for years in which the survey did not cover the full area west of Cape Infanta (Appendix C).

References

Coetzee JC, Merkle D and Butterworth DS. 2020. Adjustment to the November 2019 acoustic biomass survey data to account for incomplete coverage of the standard survey area. FISHERIES/2020/JAN/SWG-PEL/01.

Coetzee JC and Merkle D. 2019. Incorporating length frequencies from the SAPFIA vessel-based observer program into the small pelagic catch allocation program. DEA Branch Fisheries document FISHERIES/2019/NOV/SWG-PEL/26.

Coetzee JC and Merkle D. 2011. Revised method of weighting November sardine survey length frequencies. DAFF: Branch Fisheries document FISHERIES/2011/SWG-PEL/24.

Coetzee JC, Merkle D, de Moor CL, Twatwa NM, Barange M and Butterworth DS. 2008. Refined estimates of South African pelagic fish biomass from hydro-acoustic surveys: quantifying the effects of target strength, signal attenuation and receiver saturation. *African Journal of Marine Science* 30:205-217.

Coetzee JC, Geja Y, van der Lingen C and Merkle D. 2018. Deriving cut-off lengths for estimating recruitment indices from hydro-acoustic surveys. DAFF: Brach Fisheries document FISHERIES/2018/MAY/SWG-PEL/09.

de Moor CL and Butterworth DS. 2015. A new length-weight relationship for South African anchovy. DAFF: Branch Fisheries document FISHERIES/2015/JUN/SWG-PEL/26.

de Moor CL, Butterworth DS and Coetzee JC. 2008. Revised estimates of abundance of South African sardine and anchovy from acoustic surveys adjusting for echosounder saturation in earlier surveys and attenuation effects for sardine. *African Journal of Marine Science* 30:219-232.

---

[1] An alternative method of deriving the WLF (converting the acoustic biomass to numbers at length) has been implemented for sardine in November surveys. This does away with the weighting of each trawl LF by the acoustic density assigned to it and rather uses a L/W regression to calculate numbers at length based on proportion at length from the pooled stratum sample LFs (Coetzee and Merkle 2011). This adjustment has not yet been implemented for anchovy in either the November or recruit surveys and will result in revised number of recruits and recruit mean weight. The extent of such change is unknown at this stage.

de Moor CL and Butterworth DS. 2011. Extrapolation of recruit numbers to Cape Infanta in the years for which the survey only reached Cape Agulhas. DAFF: Brach Fisheries document FISHERIES/2011/AUG/SWG-PEL/42.

de Moor CL, Merkle D, Coetzee J and van der Lingen CD. 2019. The data used in the 2019 sardine assessment. DEA: Branch Fisheries Document FISHERIES/2019/SWG-PEL/22.

De Oliveira JAA. 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. The monthly anchovy commercial catch tonnage (in thousands of tons) landed.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1984	4.743	16.698	30.497	49.975	69.031	53.193	18.193	0.000	0.000	22.878	7.281	0.000
1985	2.284	11.296	13.005	65.782	74.245	83.354	20.954	0.938	0.784	0.000	0.000	0.000
1986	0.694	13.890	30.537	69.041	47.203	64.662	37.707	28.787	7.068	0.000	0.000	0.000
1987	61.838	114.134	139.674	48.084	14.925	50.851	78.632	67.886	24.357	0.000	0.000	0.000
1988	23.643	108.117	60.589	11.366	50.259	74.417	60.868	70.459	38.843	71.175	3.008	0.000
1989	10.582	62.255	24.864	59.553	83.796	39.623	13.747	0.000	0.000	0.000	0.000	0.000
1990	1.485	0.649	13.561	38.471	37.069	59.514	0.559	0.216	0.035	0.000	0.000	0.000
1991	0.502	2.923	29.246	36.466	22.908	51.769	6.106	1.091	0.025	0.009	0.000	0.000
1992	28.370	11.826	13.471	68.664	58.777	35.221	44.465	57.044	26.366	4.836	0.002	0.002
1993	18.851	20.643	17.703	43.292	14.101	1.196	10.878	67.182	38.890	3.052	0.015	0.065
1994	2.156	19.528	22.211	19.193	42.200	17.431	0.279	30.109	2.830	0.003	0.000	0.000
1995	0.089	0.406	20.440	22.748	13.473	35.621	32.335	39.432	1.873	10.396	1.617	0.000
1996	3.737	5.570	2.779	3.853	10.078	14.578	0.242	0.001	0.002	0.021	0.019	0.000
1997	0.026	0.014	0.000	0.022	1.196	0.758	20.434	10.927	23.355	3.653	0.000	0.000
1998	0.005	0.691	1.132	18.537	23.155	42.516	12.263	3.776	4.407	1.383	0.000	0.000
1999	0.000	0.000	3.548	8.490	19.660	28.584	20.120	33.472	51.708	13.273	1.037	0.000
2000	0.000	0.026	32.352	30.594	41.620	16.153	50.979	55.191	34.363	5.196	0.795	0.022
2001	0.949	8.547	8.285	34.781	34.083	45.297	10.296	30.812	51.150	60.380	2.735	0.195
2002	1.030	6.296	0.179	22.147	6.965	48.724	48.337	35.085	44.024	0.048	0.302	0.308
2003	0.019	0.041	3.524	16.636	23.991	78.245	48.150	16.801	40.602	27.158	3.479	0.232
2004	0.215	1.313	1.447	18.697	39.884	20.901	65.994	23.176	16.266	0.734	1.249	0.219
2005	0.000	2.018	27.919	51.448	56.259	21.245	42.187	27.097	43.257	9.318	1.951	0.028
2006	0.670	2.317	3.459	6.943	7.536	31.404	35.579	20.658	22.852	2.182	0.492	0.093
2007	0.009	0.145	2.252	17.700	58.403	31.053	34.540	37.773	43.707	25.136	0.681	1.691
2008	0.099	2.649	9.632	23.948	34.912	21.353	26.454	59.198	28.840	49.958	8.826	0.779
2009	1.747	11.461	9.782	13.018	18.797	8.903	38.577	35.513	27.714	6.081	2.873	0.000
2010	0.962	6.405	17.317	26.661	15.218	39.444	65.833	40.072	4.962	0.079	0.110	0.000
2011	0.037	3.177	3.136	16.316	24.467	17.060	40.931	14.733	0.016	0.005	0.000	0.000
2012	8.795	19.050	43.246	36.001	57.935	32.047	43.947	21.848	31.944	10.669	1.820	0.000
2013	3.268	0.073	0.025	23.381	17.990	12.717	3.384	3.530	7.995	3.132	3.295	0.012
2014	0.030	26.265	45.775	48.831	43.119	3.936	21.955	19.855	21.961	7.592	1.167	0.012
2015	0.015	4.478	47.497	54.216	68.953	28.559	29.077	4.308	0.224	0.138	0.937	0.018
2016	0.063	5.446	5.140	51.077	44.622	64.863	41.486	22.310	22.977	3.343	0.001	0.000
2017	0.046	0.167	4.448	47.376	46.408	11.362	76.731	22.674	4.997	1.169	0.057	1.125
2018	0.002	16.344	28.523	33.738	48.608	12.944	39.138	61.431	8.427	3.348	0.744	0.119
2019	0.125	6.633	17.620	23.436	33.308	20.543	18.989	15.327	27.672	0.775	0.006	0.009

Table 2. Anchovy total biomass (in tons) as far as Port Alfred and associated CV from the November acoustic survey and anchovy spawner biomass (in tons) and associated CV determined by the DEPM.

Area	Hydroacoustic survey		DEPM	
Year	Total Biomass	CV	Spawner Biomass	CV
1984	1553813	0.282	1100000	0.450
1985	1366294	0.211	616000	0.400
1986	2568625	0.172	2001000	0.350
1987	2108771	0.157	1606000	0.300
1988	1607060	0.222	1679000	0.350
1989	751529	0.167	421000	0.350
1990	651711	0.183	723000	0.580
1991	2327834	0.159	2913000	0.350
1992	2088025	0.161	3600000	0.310
1993	916359	0.209	770000	0.340
1994	617276	0.159		
1995	601271	0.217		
1996	162048	0.410		
1997	1482633	0.267		
1998	1229132	0.217		
1999	2052156	0.156		
2000	4653779	0.125		
2001	6720287	0.107		
2002	3867649	0.154		
2003	3563232	0.236		
2004	2037011	0.131		
2005	3077001	0.144		
2006	2106273	0.136		
2007	2505655	0.157		
2008	3598790	0.120		
2009	3792547	0.136		
2010	2077035	0.144		
2011	754124	0.204		
2012	3187964	0.116		
2013	3819666	0.102		
2014	2970760	0.137		
2015	1944258	0.157		
2016	1733040	0.227		
2017	1568398	0.196		
2018	1559620	0.235		
2019	843745[2]	0.204[3]		

---

[2] The survey ended off the Sundays River mouth in Algoa Bay with an estimate of anchovy biomass of 826 264t. This value was extrapolated to the standard survey area up to Port Alfred based on the average proportion of biomass in Stratum E up to the Sundays River mouth to that for the full Stratum E, excluding 2016 as an outlier (see Coetzee et al. 2020).

[3] This was adjusted from the estimated CV of 0.202 up to the Sundays River mouth by adding the CV of the average proportion of the biomass in Stratum E up to the Sundays River mouth, excluding 2016 as an outlier, to account for the extrapolation to the standard survey area up to Port Alfred (see Coetzee et al. 2020).

Table 3. The date of the commencement of the annual recruit survey (taken to be the date on which the first acoustic interval is logged and not the date on which the first station was occupied) and the cut-off lengths used to estimate juvenile anchovy from the recruit surveys, anchovy recruitment (in thousands of tons and numbers in billions) from Hondeklip Bay to Cape Infanta and associated CV from the recruitment acoustic survey, and the mean recruit weight (in grams).

Year	Date of commencement of survey	Time of the recruit survey after 1 May	Cut-off length (cm) for anchovy juveniles in the survey	Biomass (Method 1 of App B)	Biomass (Method 2 of App B)	CV*	Mean Weight	Numbers*
1985[4]	20-May	0.613	<10.50	348.547	344.245	0.276	4.176	83.454
1986[5 6]	10-Jun	1.300	<10.50	617.558	643.489	0.184	4.433	142.622
1987[4 7]	20-Jul	2.613	<11.00	676.738	716.663	0.167	5.438	127.408
1988[4 6]	27-Jun[8]	1.867	<11.50	561.451	585.053	0.164	4.352	132.089
1989[4 6]	08-Jun[9]	1.233	<10.50	161.470	168.976	0.205	4.874	33.915
1990[4 6]	22-Jun	1.700	<10.50	169.555	176.746	0.225	3.316	52.355
1991	07-May	0.194	<10.50	519.845	521.418	0.151	4.577	113.584
1992	13-May	0.387	<10.50	427.933	438.584	0.161	4.568	93.681
1993	21-May	0.645	<10.50	448.144	445.794	0.266	3.895	115.058
1994	05-May	0.129	<9.50	129.890	135.023	0.184	4.251	30.554
1995	10-Jun	1.300	<10.50	391.859	391.749	0.179	3.548	110.439
1996	05-Jun	1.133	<10.50	72.199	72.077	0.220	2.802	25.771
1997	17-May	0.516	<10.00	402.596	402.624	0.186	4.463	90.210
1998	20-May	0.613	<10.50	451.514	451.211	0.150	3.307	136.518
1999	10-May	0.290	<10.00	813.098	812.242	0.158	4.081	199.228
2000	15-May	0.452	<9.50	2477.589	2474.927	0.168	3.966	624.675
2001	05-May	0.129	<9.00	2027.740	1946.112	0.135	3.233	627.200
2002	05-May	0.129	<11.00	1541.803	1543.397	0.115	2.963	520.413
2003	14-May	0.419	<10.00	1391.468	1396.638	0.189	3.234	430.308
2004	08-May	0.226	<11.00	1060.548	1058.653	0.219	4.445	238.569
2005	13-May	0.387	<9.0	463.928	476.756	0.261	2.789	166.355
2006	19-May	0.581	<9.0	238.080	240.850	0.160	2.080	114.480
2007	18 May	0.548	<9.0	1317.486	1309.951	0.197	2.736	481.582
2008	21 May	0.645	<10.0	1513.123	1507.693	0.202	2.639	573.429
2009	15 May	0.452	<9.0	872.713	844.721	0.231	2.867	304.407
2010	27 May	0.839	<9.0	937.196	979.991	0.309	3.119	300.448
2011[10]	27 May	0.839	<9.0	266.606	265.204	0.286	2.608	100.632
2012	16 June	1.500	<8.5	466.718	470.502	0.174	3.610	129.297
2013	23 May	0.710	<9.0	977.968	985.668	0.198	3.011	324.797
2014	10 May	0.290	<9.5	602.686	595.586	0.336	3.218	187.265
2015	22 May	0.677	<9.0	797.316	793.189	0.182	3.681	216.587
2016	8 Jun	1.233	<9.0	327.898	348.192	0.206	3.541	92.613
2017	12 Jun	1.367	<9.0	2556.226	2563.199	0.138	3.079	830.201
2018	-	-	-	No survey				
2019	7 Jun[11]	1.200	<9.0	597.670	589.529	0.247	2.211	270.322

[4] The 1985 survey began in Mossel Bay and ended in Namibia. This survey included a single stratum from east of Danger Point to Mossel Bay. This full area is included in the survey estimate and thus the estimate is higher than that which would correspond to a survey area west of Cape Infanta only. It is thus not strictly comparable with the rest of the time series, but given the low anchovy recruit densities observed between Cape Infanta and Mossel Bay, it is considered acceptable for this assessment.

[5] The time series of recruitment biomass and numbers between 1985 and 1997 were updated to account for saturation of the acoustic signal in previously used echosounders (de Moor et al. 2008 and subsequently updated in 2011 after correcting an error in the way interval densities were split between adults and recruits). This early part of the time series was revised using a regression of available recruit data from 1998 to 2006. The time series of recruitment from 2005 onwards has recently been revised following a revision in the cut-off lengths (Coetzee et al. 2018). The regression analysis based on 1997-2006 data should, therefore, be updated. The extent of the impact of two year's data on this regression is unknown at this stage. However, given current time constraints and because the (potentially small) correction would only apply to the early part of the time series, this update has been delayed.

[6] Biomass and numbers west of Cape Infanta were estimated by taking that observed east of Danger Point and increasing by 2% (App C).

[7] Biomass and numbers west of Cape Infanta were estimated by taking that observed east of Cape Agulhas and increasing by 2% (App C).

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

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

[10] Biomass and numbers west of Cape Infanta were estimated by taking that observed east of Cape Agulhas and increasing by 1% (App C).

[11] The first station was on the 5th June 2019, although the first acoustic interval was logged on the 7th June 2015.

Appendix A: Pelagic sample allocation

The sample allocation method is the process whereby a length frequency (LF) 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 DAFF network. Length frequency data from the Fisheries observer program (2001 to 2011), SAPFIA-funded observer program (2017-2018) and Industry canneries (2011 to 2012 and presently ongoing) are included in the calculation. 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 LFs for anchovy (*Engraulis encrasicolus)* and sardine (*Sardinops sagax)* though LFs 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 per set and a pelagic 10' by 10' block position. Every landing is sampled for its species composition and weighed. The fisheries inspectors started the task of scale monitor and hence the catch sheet is referred to as the inspector's form. Normally this function is contracted to a private entity. Skipper data are available on Sybase from 1984 onwards but inspector data were obtained only from 1987. DAFF field station personnel collect data sheets and enter the information on Sybase.

- Field station samples: DAFF field station personnel collect random samples at the major pelagic fishing harbors for species composition and length frequency (Capricorn Fishing Monitoring was contracted from 2002 until 2005 to man St. Helena Bay and Gansbaai). Samples of industrial fish such as 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 partially-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 was terminated in 2011. An industry-funded observer program began in 2017 and is ongoing. Observer samples for 2017 and 2018 have been incorporated into the data thus far (2019 data are not yet incorporated) (Coetzee and Merkle 2019). Onboard biological sampling was started only in 2001. Observer sampling results reflect 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.

- Industry cannery samples: As part of the quality process at the canneries, fish are also measured and weighed according to DAFF field station specification.

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*.cs*v* contains the species composition data and *lfreq.csv* contains the length frequency data.

Data handling and evaluation

DEFF 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 validation 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 remain representative of what was recorded by the scale monitors. From 2015 the entering of data into Access was stopped because this data in now obtained from Industry and used to verify the catches.

- 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

RLFdata.accdb (where the LFs are generated)

This program has links to the following data sets that is required for the computation.

1) Cannery data.accdb (sardine length frequency)
2) RSAP reference.accdb (data set with vessel detail and pool positions)
3) Capfish.accdb (if data from 2001 to 2011 and from 2017 to 2018 are included)

4) Pel Catch (skipper data sets)

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 from 1999 onwards 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 length frequency is estimated using the 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 length frequencies 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, the border between area 23 and 24 was shifted slightly west to 24 degrees east (Tsitsikama), although this made little difference in practice since catches between 23 and 24 were small.

Although the LFs 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 LF 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.

- From 2012 the cannery length frequencies are included.

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 LF 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 LF average was used, but it was felt that it is better to allocate the LF from the adjacent month. The LFs 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 the species composition to individual throws. Observer sampling is ideally suited for this approach, because every throw is sampled, but greater sampling coverage and matched skipper throws are required.

Sampling coverage

Optimum sample size and sampling coverage can be determined only by 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 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 coverage is attainable.

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. In earlier years field stations were well manned and more samples were taken than presently. The Observer program was introduced in 2000 and this improved sampling and as a result field station sampling was reduced. However, when the program was terminated in 2011 the field station coverage was not adjusted to earlier levels. This poor coverage can be seen see as the low values on the graph in Figure A.1. The inclusion of the industry-funded observer samples in 2017 and 2018 have assisted coverage in those years (Coetzee and Merkle 2019). The fisheries cannery data (used since 2012) is necessary to maintain directed sardine sampling levels.
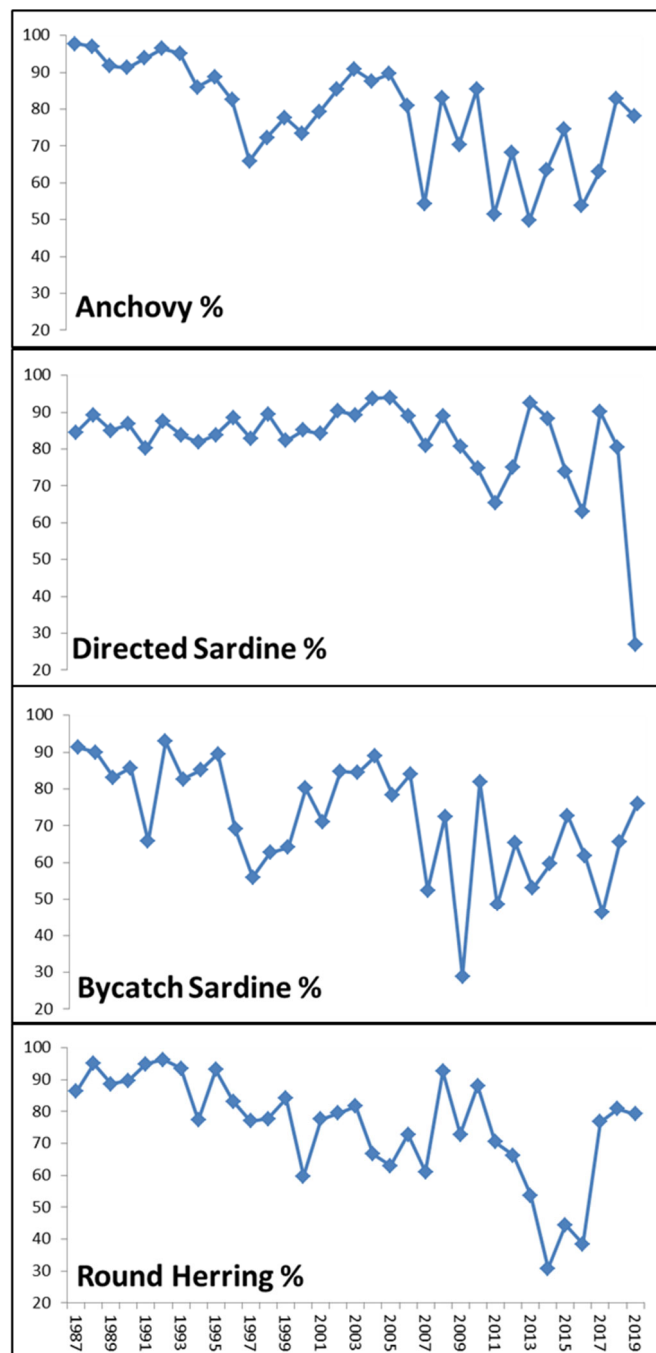

Figure A.1. Sampling coverage obtained on a first level pool-area/week including 2000-2011 and 2017 and 2018 observer data.
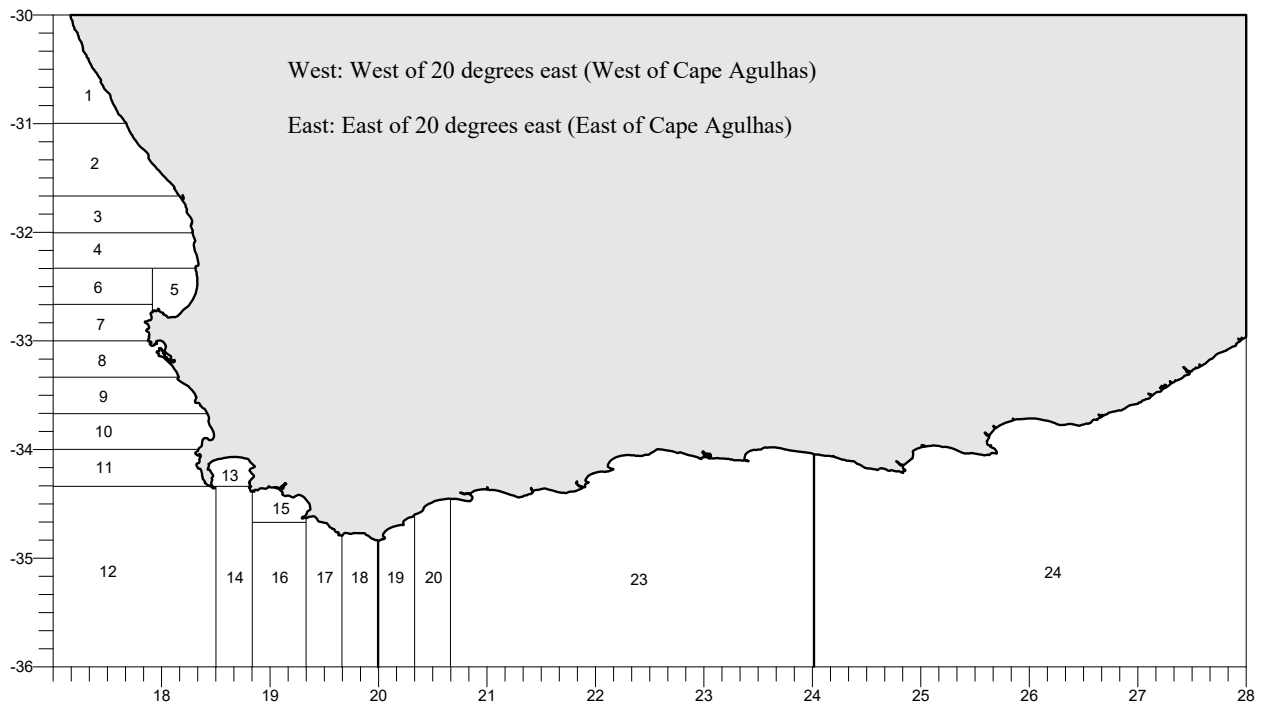
Figure A.2. The pool areas that are used for sample allocation.

Appendix B: Methods Used to Calculate Recruit Biomass

Two different methods have been used to derive recruit-only biomass. The first, used since the start of the time series, converts the total recruit and adult biomass into a weighted length frequency and then through the use of a L/W regression and a cut-off length for recruits, splits the biomass into a recruit and adult portion. The second method, devised as a method to estimate CVs of the recruit-only biomass, is based on apportioning the acoustic density between recruits and adults for each interval, based on the length- specific target strength and a cut-off length (derived independently from maturity at length information). The second method incorporates some of the first method to then convert this recruit-only biomass into numbers at length and to calculate mean recruit weight. The recruit biomasses differ between the methods due to the differences in the way the densities are weighted, but it is assumed that method 2 gives a more accurate estimation of the biomass as it is unaffected by variability in length at weight.

Method 1

This method, designed by Ian Hampton and Beatriz Roel, has been used since the start of the time series and calculates recruit biomass, number of recruits (less than a certain cut-off length) and a recruit mean weight:

1) The acoustic biomass per stratum (of adults and recruits) is calculated using the Jolly and Hampton method (i.e., each interval is weighted by interval length and a mean density per transect is calculated. Each transect is again weighted by its length to get a mean density per stratum).

2) Each acoustic interval has been linked to a particular grid reference (trawl sample) which was used to scale the acoustic energy to density. The trawl sample has a length frequency (LF) and associated length frequency mass (LFMASS). This LF and LFMASS include both adults and recruits as it is impossible at this stage (at sea) to know what the cut-off length for a recruit is. The LFMASS is the total weight of the LF sample (the combined weight of all fish of a particular species measured for the LF distribution).

3) For each interval, the acoustic density is multiplied by the interval length. This weighted interval density is then summed over all intervals for each grid reference, per stratum and per species to give an acoustic weighting to each grid reference, $w_{GR}(grid, stratum, species)$.

4) The weighted grid reference is then summed over all grid references for each stratum and species to give a weighted grid reference per stratum for each species, $w_{GR}(stratum, species)$.

5) For each length class of each grid reference, calculate a Trawl WF (trawl weighting factor) $=w_{GR}(grid, stratum, species)/LFMASS$. This converts the acoustic weighting (in terms of mass) into a factor in terms of numbers.

6) The length frequency (LF) is then weighted by this Trawl WF and summed for each length class to give a weighting to each length class (Lgroup) for each stratum for each species sum(number*trawl WF), $WLF(Lgroup, stratum, species)$.

7) $WLF(Lgroup, stratum, species)$ is then scaled to the biomass of the stratum: $BLF(Lgroup, stratum, species) = [WLF(Lgroup, stratum, species) * Biomass(stratum, species)]/\sum w_{GR}(stratum, species)$.

8) BLF is then summed across all strata for each species to give a final length frequency per species for the survey (this is done separately up to Cape Infanta and for the whole survey).

9) For each species an age/length matrix is then generated using a cut-off length for recruits.

10) The proportion in each length class is multiplied by BLF to get the total number of 0-year olds (recruits) and the total number of 1-year olds (adults). This is again done separately as far as Cape Infanta and for the whole survey. The

number of fish in each length class is then multiplied by a length weight regression to get an estimated weight (in grams) for each length class, where $w = 0.00924 \times Lgroup^{3.046}$ for anchovy[12] and $w = 0.0096 \times Lgroup^{3.075}$ for sardine.

11) The numbers and weights are then summed across all length classes for each species to give total number of 0-year-olds, $N_{tot,0}$, and 1-year-olds, $N_{tot,1}$, and total weight of 0-year-olds, $W_{tot,0}$, and 1-year-olds, $W_{tot,1}$.

12) The mean weight of 0-year-olds and 1-year-olds is then calculated by $MW_a = \left(w_{tot,a}/1000000\right)/N_{tot,a}$. The calculated biomass is then $B_{calc} = MW_0 \times N_{tot,0} + MW_1 \times N_{tot,1}$ and should be close to the acoustic biomass, $B_{acoustic}$. B_{calc} and $B_{acoustic}$ are not always identical because in some years the fish are heavier/lighter than that predicted by the length weight regression. The mean weight of recruits and 1–year-olds is weighted by the ratio of the calculated to actual acoustic biomass to get a corrected mean weight: $CMW_a = MW_a \times B_{acoustic}/B_{calc}$.

Method 2

This method was devised to map recruit-only density rather than the density of combined adults and recruits. It therefore allows for the calculation of a recruit-only biomass and CV, independent of method 1 above. In summary the density in each interval is multiplied by the proportion of recruits in that interval to get a recruit only density. The proportion of recruits in each interval is obtained by calculating the proportion of acoustic energy backscattered by recruits only, based on the length frequency that each interval has been assigned and a cut-off length:

1) For each trawl (grid) the acoustic back scattering for each length class is calculated for each species and multiplied by the number of fish in that length class (basically applying the species specific target strength relationship to the length class (L_t)):

$$BS = \begin{cases} 10^{0.1 \times -21.12} \times L_t^{-12.15/10} \times N & if\ Sp = 1 \\ 10^{0.1 \times -13.21} \times L_t^{-14.9/10} \times N & if\ Sp = 2\ or\ 5 \\ 10^{0.1 \times -7.75} \times L_t^{-15.44/10} \times N & if\ Sp = 3\ or\ 4 \end{cases}$$

where Sp 1 = anchovy, Sp 2 = sardine, Sp 3 = horse mackerel, Sp 4 = mackerel and Sp 5 = round herring.

2) The backscattering (BS) is summed for each species for each trawl to give a total backscatter for each grid, BS_{tot}.

3) The backscattering due to recruits, BS_{rec}, is then calculating by summing BS for only the length classes less than the cut-off length for each species for each trawl. The cut-off length is obtained from the length at 20% maturity, derived from samples collected during the survey (Coetzee et al. 2018).

4) The proportion of recruits in each trawl is then calculated by BS_{rec}/BS_{tot}.

5) This proportion is then multiplied by the original interval density (of recruits and adults) to obtain the recruit only density (for all years).

6) This recruit only density is used in the regressions of capped to uncapped data in order to estimate (using the Jolly and Hampton weighting procedure) the uncapped recruit only biomass prior to 1997 together with a CV.

7) The steps outlined in Method 1 from 1-8 are then repeated to derive a weighted length frequency and the number of recruits is calculated by summing the WLF for all length classes smaller than the cut-off length.

---

[12] This L/W regression differs from a more recent relationship used by de Moor 2020 ($W = 0.0079 \times Lgroup^{3.0979}$). This should be updated in the future, but has no impact on the number of recruits or recruit biomass as currently estimated and only affects the recruit mean weight and proportion of recruits as calculated under method 1, neither of which are used in the assessment.

Appendix C. Extrapolation of recruit numbers to Cape Infanta in the years for which the survey only reached Cape Agulhas (identical to Appendix C of de Moor et al. (2019))

The time series of sardine and anchovy recruits used in assessments and to set the final anchovy TAC and sardine TAB in June each year are based on the surveyed area west of Cape Infanta. In some years, the survey did not cover the full area eastwards to Cape Infanta. Between 1985 and 1990, the survey was regularly terminated at Cape Agulhas. In 2011, although the survey was planned to extend eastwards of Cape Infanta, the survey only covered the area west of Cape Agulhas as damage to the survey vessel resulted in an early end to the survey. This Appendix uses the same approach as in de Moor and Butterworth (2011) to extrapolate the survey estimate of recruit numbers and associated variance in years when only the area up to Cape Agulhas was surveyed, but based now on an updated and extended time series of data.

Method

Table C.1 lists the recruit numbers west of Cape Agulhas (strata A-E in recent years) and between Cape Agulhas and Cape Infanta (stratum F in recent years) for years for which these data can be extracted. The ratio of recruits between Cape Agulhas and Cape Infanta to those west of Cape Agulhas is computed as r_y. The recruits between Cape Agulhas and Cape Infanta can then be estimated using the average of these proportions, $\bar{r}$, i.e. $N_{stratumF} = \bar{r}N_{strataA-E}$.

The variance of the extrapolated estimate west of Cape Infanta is then estimated as follows:

$$var(N_{stratumA-F}) = var\big((1+\bar{r})N_{strataA-E}\big) = var(1+\bar{r})(N_{strataA-E})^2 + (1+\bar{r})^2 var(N_{strataA-E})$$

where

$$var(1+\bar{r}) = var(\bar{r}) = \frac{1}{n-1}\sum_y (r_y - \bar{r})^2$$

and

$var(N_{strataA-E}) = (N_{strataA-E}/CV(N_{strataA-E}))^2$. Although the CV is calculated from the biomass rather than the number of recruits west of Cape Agulhas, this is assumed to be an adequate assumption for these purposes.

Results and Recommendations

The average proportion of recruits between Cape Agulhas and Cape Infanta to that west of Cape Agulhas over the years 1991 to 1995, 1998 and 1999 is 0.02 for anchovy and 0.06 for sardine (Table C.1). However, the proportion of sardine recruits between Cape Agulhas and Cape Infanta in 1999 was exceptionally high. Treating this point as an outlier results in an average of 0.004. These low proportions correspond with the previously held assumption of there being few recruits between Cape Agulhas and Cape Infanta historically. While one might argue no extrapolation for these early years is required under the perception of a regime shift in the early 2000s, to maintain consistency with the 2011 adjustment (see below), it is recommended that the anchovy and sardine recruit numbers for 1985 are multiplied by 1.02 and 1.004, respectively (see Table C.1). The CVs of these estimates have all decreased marginally, but to three decimal places they remain the same.

The average proportion of recruits between Cape Agulhas and Cape Infanta to that west of Cape Agulhas is 0.01 for anchovy and 0.52 for sardine over the years 2002 to 2019, excluding 2011 when the survey ended at Cape Agulhas and 2018 when there was no survey. It is recommended that the 2011 survey estimate of anchovy recruit numbers be increased from 99.703 billion recruits (with a CV of 0.286), reflecting the area west of Cape Agulhas, to 100.592 billion recruits (with a CV of 0.286), reflecting the area west of Cape Infanta (Table C.1). The proportion of sardine recruits between Cape Agulhas and Cape Infanta

in 2015 and 2019 was exceptionally high. Treating these points as outliers results in an average proportion of sardine recruits between Cape Agulhas and Cape Infanta of 0.03. It is recommended that the 2011 survey estimate of sardine recruit numbers be increased from 5.624 billion recruits (with a CV of 0.396), reflecting the area west of Cape Agulhas, to 5.831 billion recruits (with a CV of 0.396), reflecting the area west of Cape Infanta (Table C.1). The CVs of these estimates have decreased marginally, but to three decimal places they remain the same.

Table C.1. Recruit numbers, ratios and summary statistics estimated from surveys. The recruit numbers between Cape Agulhas and Cape Infanta, given in **bold**, have been extrapolated using the recommendations in the text. Note that numbers are quoted in billions (10^9).

	Anchovy			Sardine		
	Recruit numbers west of Cape Agulhas (strata A-E)	Recruit numbers between Cape Agulhas and Cape Infanta (stratum F)	Ratio F:A-E	Recruit numbers west of Cape Agulhas (strata A-E)	Recruit numbers between Cape Agulhas and Cape Infanta (stratum F)	Ratio F:A-E
1985	83.454	**1.984**		3.592	**0.014**	
1986	142.640	**3.311**		3.691	**0.014**	
1987	127.424	**2.958**		7.380	**0.029**	
1988	132.106	**3.067**		0.440	**0.002**	
1989	33.920	**0.787**		2.137	**0.008**	
1990	52.362	**1.216**		2.502	**0.010**	
1991	112.836	0.749	0.01	1.909	0.006	0.00
1992	87.083	6.598	0.08	5.614	0.019	0.00
1993	109.024	6.034	0.06	15.200	0.039	0.00
1994	30.401	0.153	0.01	2.642	0.013	0.00
1995	110.419	0.020	0.00	25.382	0.006	0.00
1996	25.771[13]		-	3.204[13]		-
1997	90.210[13]		-	36.856[13]		-
1998	136.458	0.060	0.00	10.616	0.100	0.01
1999	194.750	4.477	0.02	7.298	3.080	0.42
2000	624.675[13]		-	20.002[13]		-
2001	627.200[13]		-	60.065[13]		-
2002	496.720	23.693	0.05	45.786	3.367	0.07
2003	430.247	0.062	0.00	33.406	3.042	0.09
2004	236.240	2.329	0.01	4.074	0.015	0.00
2005	166.266	0.090	0.00	2.670	0.188	0.07
2006	114.263	0.217	0.00	9.463	0.043	0.00
2007	481.401	0.182	0.00	2.983	0.012	0.00
2008	568.143	5.286	0.01	4.074	0.016	0.00
2009	292.009	12.398	0.04	9.149	0.140	0.02
2010	300.007	0.441	0.00	35.117	0.453	0.01
2011	99.703	**0.889**		5.624	**0.207**	
2012	128.900	0.397	0.00	6.963	1.023	0.15
2013	324.216	0.582	0.00	12.581	0.006	0.00
2014	185.330	1.934	0.01	1.959	0.026	0.01
2015	217.010	0.159	0.00	0.731	5.527	7.56
2016	92.583	0.030	0.00	0.804	0.007	0.01
2017	819.990	10.211	0.01	6.730	0.450	0.07
2018	No survey			No survey		
2019	268.573	1.749	0.01	2.821	0.719	0.25
Average 1991-1995,98,99			0.02			0.06[14]
Average 2002-2010,2012-2017,2019			0.01			0.52[15]

---

[13] Strata E and F were combined in these years.
[14] 0.00 if 1999 is excluded.
[15] 0.05 if 2015 is excluded. 0.04 if 2015 and 2019 are excluded.