The Estimation of Uncapped Acoustic Survey Biomass from Capped Data

C.L. Cunningham,* D.S. Butterworth* and J. Coetzee#

Introduction

Management of the South African sardine and anchovy resources is critically dependent on estimates of recruitment and spawner biomass obtained from hydroacoustic surveys. These surveys commenced in 1984, but in 1997 new equipment (an EK500 echo sounder) replaced the older EK400 echo sounder.

The introduction of the EK500 echo sounder revealed a saturation problem with the EK400 echo sounder, particularly for sardine (Coetzee 2003). Survey estimates since June 1997 were initially 'capped' at -29dB for assessment purposes, to maintain a comparable time-series from 1984.

Although this provided a temporary solution to the difference in estimates between the EK400 and EK500 echo sounders, it has been recommended that the survey estimates of abundance be based on the EK500 technology (BENEFIT 2001). Although absolute measures of biomass output from the anchovy assessment are scaled by separate egg survey estimates of abundance, the sardine assessment outputs are scaled purely by the hydroacoustic survey estimates. Management of these resources becomes more efficient (i.e., greater catch can be allowed for the same perceived risk) if bias in such absolute estimates can be reduced, particularly by accounting for this saturation problem and by the use of improved target strength estimates.

In dense fish schools, fish at the top of the school absorb most of the energy from acoustic echo signals such that fish lower down in the school are insonified with less energy, resulting in the acoustic echo signals no longer being proportional to the fish density. This signal attenuation affects mostly sardine and has recently been quantified (Coetzee *et al*. 2002).

In this document, a regression analysis of the capped (no attenuation in the case of sardine) densities on the uncapped (with attenuation in the case of sardine) densities is carried out. This analysis includes densities calculated with the new target strength expression, whilst previous analyses used the old target strength expressions (Cunningham *et al*. 2006). In addition, the calibration of uncapped densities from capped densities in years prior to 1998 is also performed.

* 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.

Methods

Data

The data available to determine the calibration relationships required for anchovy were the capped, new target strength estimates of density per interval (a short segment of a survey transect line) and the uncapped, new target strength estimates of this density for the recruit and spawner biomass surveys from November 1997 to May 2006. For sardine, attenuation was taken into account in the uncapped, not the capped estimates, and surveys from May 1998 to May 2006 were considered. For ease of reading, these densities are simply referred to as capped and uncapped densities below.

Regression of the capped data on the uncapped data

Both the uncapped and capped survey estimates of biomass are subject to survey sampling error. However, the uncapped estimate is free from the further error caused by the saturation problem resulting from the EK400 echo sounder. In addition, for sardine, the uncapped estimate is also free from further error due to attenuation at large densities. Thus the uncapped data are the only set of 'true' observed data. (Note that in the context of estimating the capped/uncapped calibration factors, the survey sampling error does not impact the 'trueness' of the uncapped data; this is because the comparison being made is between capped and uncapped estimates (with attenuation for sardine) for the same survey track in the same stratum, so that the resultant abundance estimates by stratum are subject to identical survey sampling errors (i.e., the pair of estimates are exactly correlated in this respect).)

With the ultimate purpose of this exercise being to calibrate capped estimates of density in earlier years to uncapped estimates of density, the model and method chosen for regression need to be able to accurately calibrate the densities of earlier years without considerable bias. The regressions were performed on a logit transformation of slope (capped/uncapped) against uncapped densities, thereby ensuring that during the process of generating data to compute the variance associated with the calibrated uncapped densities, the sampled slope would remain within the realistic range of [0,1]. Thus the 1997/8 to 2006 slopes (capped/uncapped density) per interval were regressed against the uncapped estimates of density per interval. These regressions were performed separately for the recruit and spawner biomass surveys, and for sardine and anchovy separately.

A mixture model was used so that the observed interval data where slope is 1 could be treated separately than when observed slope < 1. The mixture model thus consisted of a portion which fitted a model to estimate the probability that the observed slope was 1, and a portion which fitted a model for slope to the uncapped data in cases where slope < 1.

Considering the data, we first attempted to fit a negative exponential to the relation between the probability that the observed slope = 1 ($prob_{y,i}^{slope=1}$) and the uncapped density. A binomial error distribution was assumed. Thus the negative log-likelihood which was minimised to estimate the model parameters was:

$$-\ln L = \begin{cases} \sum_y \sum_i \left(prob_{y,i}^{slope=1}\right)^1 \left(1 - prob_{y,i}^{slope=1}\right)^0 & if \quad u_{y,i} = c_{y,i} \\ \sum_y \sum_i \left(prob_{y,i}^{slope=1}\right)^0 \left(1 - prob_{y,i}^{slope=1}\right)^1 & if \quad u_{y,i} \neq c_{y,i} \end{cases}$$

where $u_{y,i}$ denotes the observed uncapped density in interval i of year y and $c_{y,i}$ the corresponding capped density. Models which consisted of various combinations of constant slope, straight lines, and negative exponential curves were explored, with the inflection points for switching between the different curves fixed in some instances and estimated in others. Using AIC to compare between models, and looking also to apply the same model to all four sets of data, the following model (consisting of three straight lines followed by a constant) was chosen:

$$prob_{y,i}^{slope=1} = \begin{cases} m^* u_{y,i} + b & if \ u_{y,i} \leq u^* \\ m^{**}\left(u_{y,i} - u^*\right) + m^* u^* + b & if \ u^* \leq u_{y,i} \leq u^{**} \\ \frac{m^{**}\left(u^{**} - u^*\right) + m^* u^* + b - p}{u^{**} - u^{***}}\left(u_{y,i} - u^{***}\right) + p & if \ u^{**} \leq u_{y,i} \leq u^{***} \\ p & if \ u_{y,i} \geq u^{***} \end{cases} \quad (1)$$

where $\theta = \{ m^*, m^{**}, b, p, u^*, u^{**}, u^{***} \}$ are estimated parameters. Although a slightly better model fit in terms of AIC was obtained when u^{***} was fixed, this seemed an inappropriate basis for choice in this instance as the fixed value selected was arbitrary.

Although a number of different error structures were investigated when testing regressions of capped densities against uncapped densities, the absolute error structure (i.e. errors with a distribution N(0,σ^2) added to the equations below) was generally found to be adequate for regressing slope against uncapped density. The models tested included:

i) a linear model: $slope_{y,i} = m_1$

ii) a "2-line (cst large *u*)" model: $slope_{y,i} = \begin{cases} m_1\left(u_{y,i} - u_1\right) + b_1 & 0 < u_{y,i} \leq u_1 \\ b_1 & u_{y,i} > u_1 \end{cases}$

iii) a "2-line (sloped)" model: $slope_{y,i} = \begin{cases} m_1\left(u_{y,i} - u_1\right) + b_1 & 0 < u_{y,i} \leq u_1 \\ m_2\left(u_{y,i} - u_1\right) + b_1 & u_{y,i} > u_1 \end{cases}$

iv) a "2-line (cst small *u*)" model: $slope_{y,i} = \begin{cases} b_1 & 0 < u_{y,i} \leq u_1 \\ m_2\left(u_{y,i} - u_1\right) + b_1 & u_{y,i} > u_1 \end{cases}$

v) a "Beverton-Holt type" (or BH type) model: $slope_{y,i} = \frac{\alpha}{1 + \beta u_{y,i}}$

vi) a "Beverton-Holt (adjusted) type" (or BH-adjusted type) model: $slope_{y,i} = \begin{cases} \frac{\alpha}{1 + \beta u_{y,i}} & 0 < u_{y,i} \leq u_1 \\ \frac{\alpha}{1 + \beta u_1} & u_{y,i} > u_1 \end{cases}$

Model iv) was needed only for the anchovy May survey. Initial results indicated that the absolute variance around the fitted relation could change above a certain uncapped density. Thus two different error models were tested:

a) constant variance: $\sigma_{y,i} = \sigma$

b) changing variance: $\sigma_{y,i} = \begin{cases} \sigma & if\ u_{y,i} \le u_2 \\ \delta\sigma & if\ u_{y,i} > u_2 \end{cases}$

The parameters estimated are thus a subset of $\phi = \{m_1, m_2, b_1, u_1, \alpha, \beta, \sigma, \delta\}$ (dependent on the model i)-vi) and variance formulation a) or b) chosen). The inflection point for change in variance, u_2, was fixed based on a grid trial for the best fit. A normal likelihood was used to fit the model predicted logit slope to the observed data where slope < 1:

$$-\ln L = \sum_y \sum_i 0.5\ln\left(\sigma_{y,i}^2\right) + 0.5\left(\ln\left(\frac{c_{y,i}/u_{y,i}}{1-c_{y,i}/u_{y,i}}\right) - slope_{y,i}\right)^2 \Big/ \sigma_{y,i}^2 \text{ , for } c_{y,i} < u_{y,i} \text{ .} \quad (2)$$

Calibration of uncapped data from capped data

The mixture model chosen above gives an expected capped density, $c_{y,i}$, from a given uncapped density as:

$$c_{y,i} = f\left(\theta, \phi, u_{y,i}\right) = prob_{y,i}^{slope=1} \times u_{y,i} + \left(1 - prob_{y,i}^{slope=1}\right) \times u_{y,i} \times \frac{e^{slope_{y,i}}}{1+e^{slope_{y,i}}} \text{ .}$$

This equation can be inverted to solve non-linearly for $u_{y,i}$, i.e.

$$u_{y,i} = g\left(\theta, \phi, c_{y,i}\right) \quad (3)$$

given the capped interval densities in early years.

Results and Discussion

The model fits to the probability that observed slope is 1 are given in Figure 1, with the estimated model parameters given in Table 1.

Table 2 lists the AIC values for combinations of models i) to vi) and error structures a) and b) defined above. In all cases the changing error option resulted in better fits to the data according to AIC compared to the unchanging absolute error option.

The "2-Line (sloped)" model had the lowest AIC value for the sardine November survey, but the fitted model resulted in decreasing capped density at large uncapped density which is unrealistic for the relationship between uncapped densities from capped densities. The "BH-adjusted type" model (which by construction avoids this problem) was thus chosen, having a good fit to the data (Figure 2) and the second lowest AIC value. For the sardine May survey, the "BH type" model resulted in a good fit to the data. The inclusion of the extra parameter in the “BH-adjusted type” model did not result in a substantial improvement in the fit and thus the AIC value for the “BH type” model was best.

For anchovy May and November surveys the “2-Line (sloped)” model had the lowest AIC value. However, the fitted slope was (marginally) positive for lower uncapped densities for the May survey and positive for higher uncapped densities for the November survey. Both such behaviours seem unrealistic, even though (weakly) supported by the data. For the anchovy November survey, the “BH-adjusted type” model was thus chosen, having the second lowest AIC value and a good fit to the data (Figure 4). For the anchovy May survey, although the “2-Line (cst small u)” model was the second best fit to the data according to AIC, it was decided to use the “BH type” model (which had a similar AIC value) to maintain consistency with the other three cases.

The model fits to the observed logit slope and the consequent “regression” of capped against uncapped densities per interval are shown in Figures 2 to 5. The maximum likelihood estimates of the parameters are given in Table 3. The standardised residuals do not suggest any obvious model misspecification (results not shown).

Probability density functions (pdfs) of the standardised residuals are given in Figure 6, together with comparisons to the pdfs of the standardised residuals for the same models, but without using a logit transformation in equation (2). In all four cases it is clear that the logit transformation provides the additional benefit of resulting in residuals that are less skew than would have been obtained had no transformation been used.

Tables 4 and 5 list the annual capped and uncapped density and biomass calculated using the calibration of equation (3). The full series is given for completeness. Figure 7 demonstrates the difference between the annual capped and uncapped biomass for the early years, with the greatest difference occurring in the sardine November survey.

Summary

This document has summarised the regression analysis performed on the capped and uncapped new target strength interval densities, taking attenuation into account with the uncapped data in the case of sardine. Two models were chosen: the “BH type model” for the May surveys and the “BH-adjusted type” model for the November surveys; both with the variance changing with uncapped density. The estimation of uncapped new target strength (with attenuation in the case of sardine) densities from capped new target strength (no attenuation in the case of sardine) densities was performed using the relationships obtained. The annual and strata CVs for the estimated uncapped biomass in early years will be finalised shortly.

Acknowledgements

Dagmar Merkle is gratefully thanked for assisting Janet in providing the capped and uncapped densities by interval.

References

BENEFIT 2001. BENEFIT Stock Assessment Workshop, Cape Town. 19-24 November 2001. BENEFIT Programme, Swakopmund, 50 pp.

Coetzee, J.C, Boyer, D., Hampton, I., Boyer, H. and A. Kreiner 2002. Correcting acoustic estimates of fish density for absorption in dense schools. *In* book of abstracts, 6th ices symposium on acoustics in fisheries and aquatic ecology, 10-14 June 2002, Montpellier, France, p104.

Coetzee, J. 2003. Influence of new technology: the capping problem. MCM Document WG/JUNE2003/PEL/WS3.

Cunningham, C.L., and Butterworth, D.S. 2006. An Update on the Calibration of Capped and Uncapped Acoustic Survey Data. MCM Document SWG/OCT2006/PEL/03. 12pp.

Table 1. Parameter estimates for the model predicted probability that observed slope is 1. The uncapped density inflection points are given in g.m^2.

Parameter		Sardine November	Sardine May	Anchovy November	Anchovy May
m^*	slope of straight line when $u \leq u^*$	-4.1	-8.4	-25.4	-26.1
m^{**}	slope of the straight line when $u^* \leq u \leq u^{**}$	-0.018	-0.043	-0.017	-0.015
b	probability-axis intercept for straight line when $u \leq u^*$	0.79	0.99	0.98	0.99
p	constant probability when $u \geq u^{**}$	0.006	0.016	0.083	0.079
u^*	1st inflection point	0.086	0.052	0.017	0.014
u^{**}	2nd inflection point	12.3	7.9	13.6	17.7
u^{***}	3rd inflection point	109.8	107.8	115.6	208.0

Table 2. AIC values for a combination of models and error structures for slope, given that observed slope < 1. The values in ***bold*** *are the lowest and those in shaded italics represent the model chosen.*

Survey / Error Model	Unchanging(a)						Changing (b)				
	Linear	2-Line (sloped)	2-Line (cst large u)	2-Line (cst small u)	BH Type	BH-adjusted Type	2-Line (sloped)	2-Line (cst large u)	2-Line (cst small u)	BH Type	BH-adjusted Type
Sardine November	2781.7	2413.0	2411.0		2415.8	2414.1	**2334.9**	2351.8		2350.9	*2342.9*
Sardine May	1915.6	1774.6	1773.3		1760.9	1762.9	1767.2	1768.3		***1751.9***	1753.9
Anchovy November	3127.5	3125.6	3125.3		3124.5	3125.5	**3112.7**	3119.7		3119.0	*3117.6*
Anchovy May	2210.3	2208.9	2210.3	2212.8	2212.3	2214.3	**2204.3**	2204.8	2206.6	*2207.3*	2207.8

Table 3. Maximum likelihood parameter estimates for the model of slope, given that observed slope < 1. The values for u_2 were fixed.

Parameter		Sardine November	Sardine May	Anchovy November	Anchovy May
α	BH type model parameter	0.93	0.93	0.92	0.94
β	BH type model parameter	0.0047	0.0044	0.0002	1.5 E-07
u_1	inflection point for change of slope	577.9		642.6	
σ	standard deviation in fit to the data	1.78	1.84	1.85	1.95
δ	Multiplicative change in standard deviation	0.53	0.58	0.55	1.16
u_2	Density at which variance changes	150	200	450	100

Table 4. The annual capped new target strength density and biomass with CV and the annual uncapped new target strength (with attenuation for sardine) density and biomass for the November spawner biomass surveys. The annual total density and biomass over all strata are given from 1984 to 1997, while the totals up to Port Alfred only are reported between 1998 and 2006. Calibrated values are given in ***bold****.*

	Sardine					Anchovy				
	Capped New Target Strength, No Attenuation			Uncapped New Target Strength With Attenuation		Capped New Target Strength			Uncapped New Target Strength	
Year	**Density**	**Biomass**	**CV**	**Density**	**Biomass**	**Density**	**Biomass**	**CV**	**Density**	**Biomass**
2006	Data unavailable			52.83	712552.7	Data unavailable			100.50	2106273.2
2005	42.88	394033.3	0.29	127.96	1048990.9	90.50	2439135.5	0.14	116.68	3077001.4
2004	52.12	1078175.5	0.27	130.40	2615715.3	88.14	1680796.8	0.13	101.82	2044615.1
2003	145.3	1439415.17	0.17	311.0	3564170.89	126.61	3025983.5	0.26	151.64	3563231.6
2002	161.6	1443846.77	0.18	358.0	4206250.5	132.38	3152741.3	0.14	163.04	3867649.2
2001	63.02	1130642.7	0.12	122.46	2309600.3	362.22	5425611.0	0.11	446.61	6720287.0
2000	36.61	855415.0	0.36	85.73	2292396.7	166.60	4107741.3	0.12	190.57	4653803.3
1999	85.04	883571.4	0.17	161.21	1635410.5	120.22	1723504.2	0.15	137.42	2052155.7
1998	95.94	843077.8	0.21	189.77	1607328.3	74.68	970108.6	0.21	95.74	1229132.5
1997	94.10	669297.0	0.22	**235.85**	**1436479.5**	55.83	1485779.7	0.26	**60.52**	**1616062.4**
1996	86.48	376130.8	0.21	**159.59**	**632927.0**	11.85	118989.5	0.38	**12.57**	**126320.5**
1995	45.69	556030.0	0.33	**83.95**	**1246701.1**	39.37	523739.4	0.16	**42.62**	**569432.0**
1994	34.25	395308.7	0.50	**47.44**	**466577.0**	38.71	502325.6	0.93	**41.59**	**538273.2**
1993	36.17	477061.2	0.53	**62.15**	**713150.7**	36.81	132254.4	0.38	**39.71**	**140270.8**
1992	40.01	460924.3	0.33	**110.52**	**715326.1**	127.94	2231043.0	0.60	**140.66**	**2450286.7**
1991	56.98	1691168.7	0.43	**128.86**	**3627360.5**	135.60	3347425.9	0.33	**147.90**	**3645663.1**
1990	21.98	605112.6	0.44	**29.91**	**759199.6**	30.47	352310.2	0.78	**32.58**	**375827.1**
1989	54.86	1016813.0	0.50	**103.70**	**2007255.5**	67.36	1611812.8	0.35	**72.76**	**1743498.9**
1988	9.23	85377.3	0.55	**13.27**	**158423.7**	112.77	1982190.5	0.22	**122.39**	**2155611.1**
1987	4.66	47308.1	0.43	**6.08**	**52373.6**	127.67	1176872.9	0.22	**140.08**	**1283723.0**
1986	27.45	306480.9	0.65	**57.25**	**497992.7**	529.78	5472987.8	0.73	**603.31**	**6292371.4**
1985	2.11	40078.3	0.43	**2.32**	**44481.7**	75.54	1561205.2	0.21	**81.74**	**1692415.2**
1984	0.63	187.1	0.83	**0.75**	**221.7**	35.32	13631.8	0.28	**38.08**	**14645.2**

Table 5. The annual capped new target strength density and biomass with CV and the annual uncapped new target strength (with attenuation for sardine) density and biomass for the May recruit surveys. The annual total density and biomass over all strata are given from 1984 to 1996/7, while the totals up to Cape Infanta only are reported between 1997/8 and 2006. Calibrated values are given in ***bold****.*

	Sardine					Anchovy				
	Capped New Target Strength, No Attenuation			Uncapped New Target Strength With Attenuation		Capped New Target Strength			Uncapped New Target Strength	
Year	**Density**	**Biomass**	**CV**	**Density**	**Biomass**	**Density**	**Biomass**	**CV**	**Density**	**Biomass**
2006	35.38	83557.2	0.29	50.34	140545.5	72.74	256455.5	0.15	80.49	297871.5
2005	7.12	30509.1	0.30	13.91	99656.8	155.12	292041.0	0.30	173.79	312513.4
2004	56.26	60626.4	0.32	144.64	129397.9	416.18	980944.7	0.21	476.78	1113421.8
2003	167.08	398871.2	0.17	445.52	1032278.0	428.45	1292077.5	0.21	492.79	1436165.4
2002	112.48	425214.4	0.16	226.67	716249.3	621.02	1661819.6	0.12	675.29	1855122.9
2001	88.54	308999.8	0.21	149.53	593843.6	468.13	1754743.9	0.14	552.57	1999119.9
2000	90.75	241650.1	0.23	320.22	510005.9	778.17	2318384.8	0.18	952.05	2569290.3
1999	54.77	157293.7	0.24	203.57	421852.3	371.79	729152.2	0.14	440.08	878681.8
1998	98.93	94320.2	0.34	190.75	142607.7	238.15	403627.7	0.15	288.73	466993.8
1997	124.35	205046.6	0.17	**372.91**	**383810.4**	273.7	384414.75	0.17	305.7	443128.67
1996	21.50	80530.6	0.24	**23.83**	**89511.2**	32.65	83634.3	0.21	**33.94**	**86638.67**
1995	91.90	150134.9	0.18	**128.63**	**206807.8**	279.92	440705.8	0.16	**294.44**	**461867.19**
1994	65.57	137812.7	0.22	**118.43**	**183983.8**	93.56	139295.7	0.17	**97.90**	**145337.86**
1993	57.58	81075.4	0.24	**112.01**	**112058.3**	77.87	459671.2	0.25	**81.29**	**481095.20**
1992	11.65	66546.6	0.25	**12.96**	**74304.4**	128.70	439496.6	0.16	**134.52**	**458446.75**
1991	4.59	24299.1	0.23	**4.97**	**26514.6**	197.18	503787.5	0.14	**207.06**	**528163.66**
1990	6.68	24454.6	0.54	**7.93**	**30270.8**	40.09	163346.0	0.22	**41.70**	**170080.23**
1989	14.53	51772.5	0.29	**18.86**	**65328.7**	65.31	166411.2	0.19	**68.69**	**173343.90**
1988	2.13	4921.1	0.37	**2.27**	**5193.9**	215.73	537073.3	0.15	**226.47**	**563097.25**
1987	17.37	59692.3	0.33	**24.12**	**96099.1**	286.88	685798.0	0.15	**302.84**	**722011.20**
1986	96.24	60490.5	0.43	**616.29**	**646452.0**[1]	516.22	595897.0	0.17	**544.50**	**626841.98**
1985	16.84	31023.1	0.34	**19.69**	**37619.8**	148.48	351192.2	0.25	**155.84**	**368611.30**

[1] This large uncapped biomass is questionable and due to a large uncapped density estimated for stratum F (Cape Point to Danger Point). Reasons for this are currently being investigated.

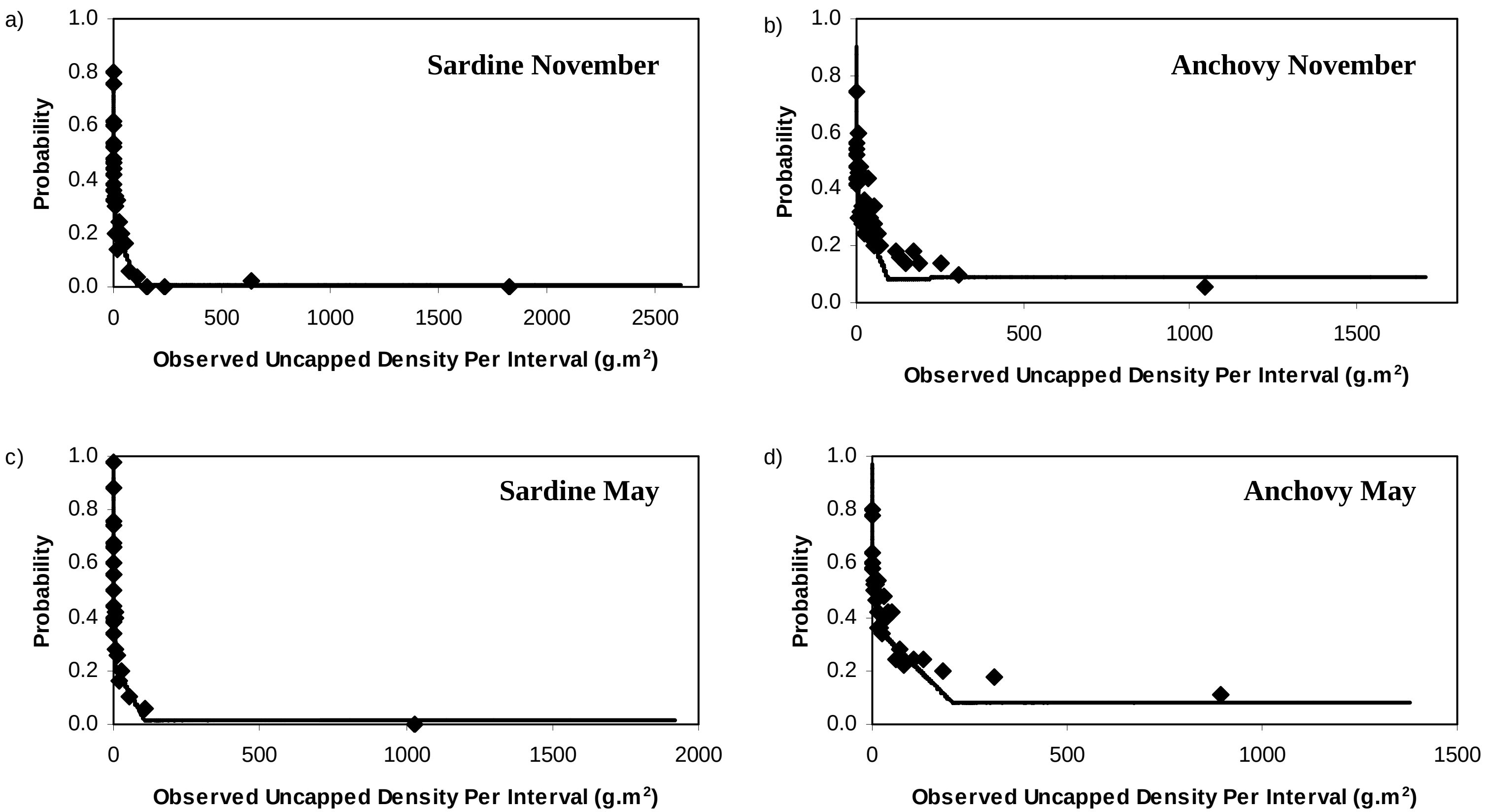

Figure 1. The model estimated probability that slope (capped / uncapped) is 1 using equation (1) for the a) sardine November spawner biomass, b) anchovy November spawner biomass, c) sardine May recruit and d) anchovy May recruit surveys. For comparative purposes the observed probability that slope is 1 is also plotted in bins of 50 data points at a time.

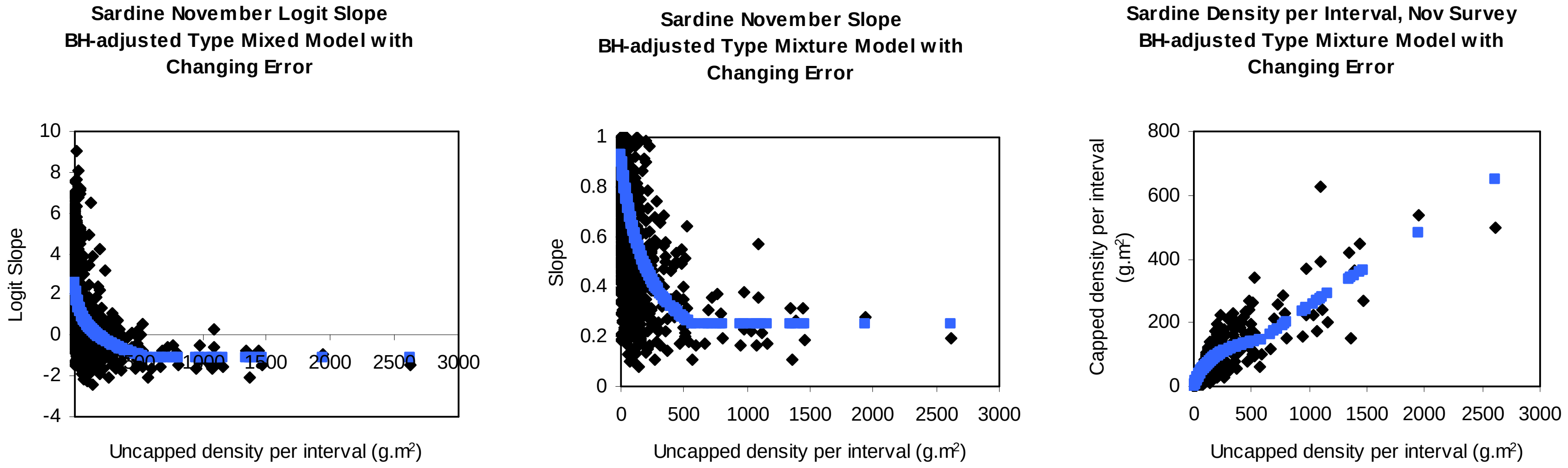

Figure 2. The model fits to the observed logit slope and the consequent slope and "regression" of capped against uncapped densities per interval for the sardine November survey.

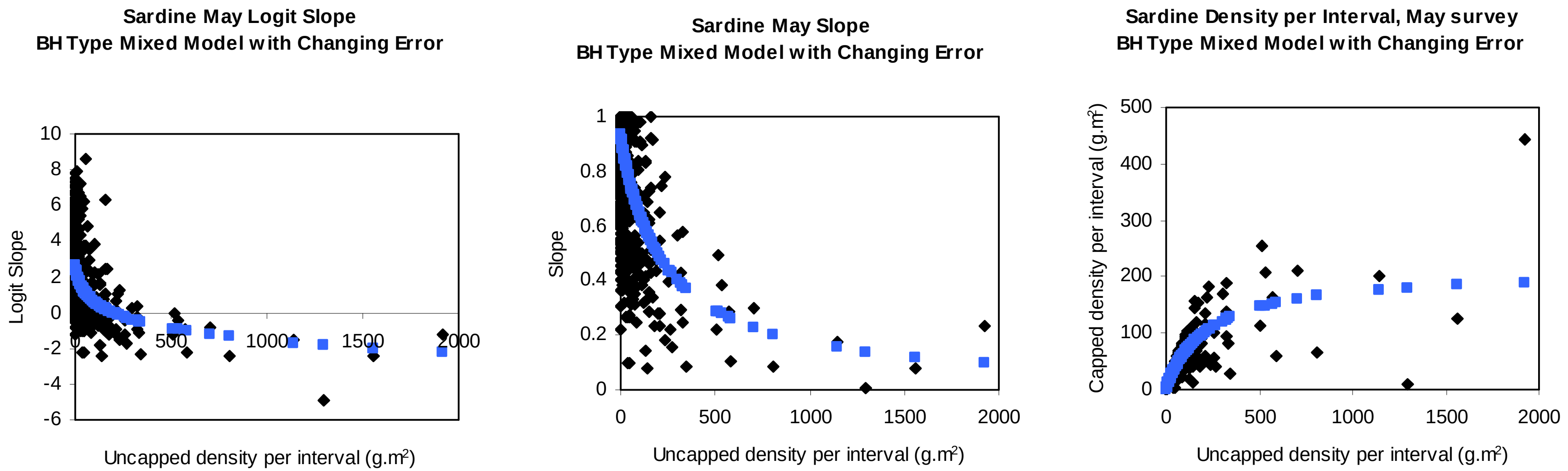

Figure 3. The model fits to the observed logit slope and the consequent slope "regression" of capped against uncapped densities per interval for the sardine May survey.

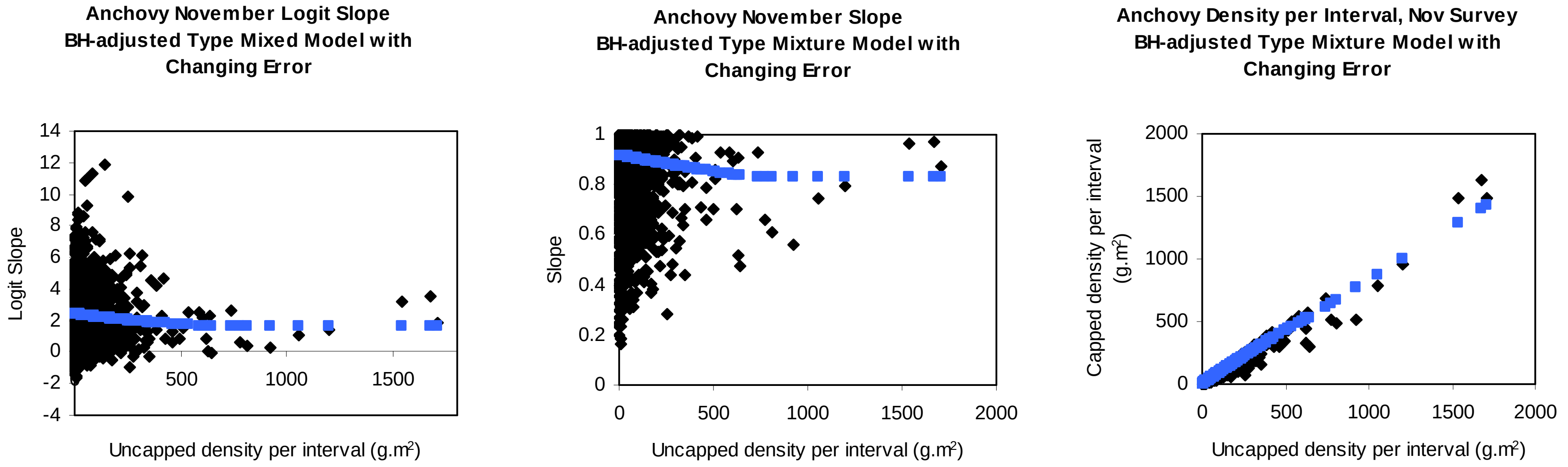

Figure 4. The model fits to the observed logit slope and the consequent slope "regression" of capped against uncapped densities per interval for the anchovy November survey.

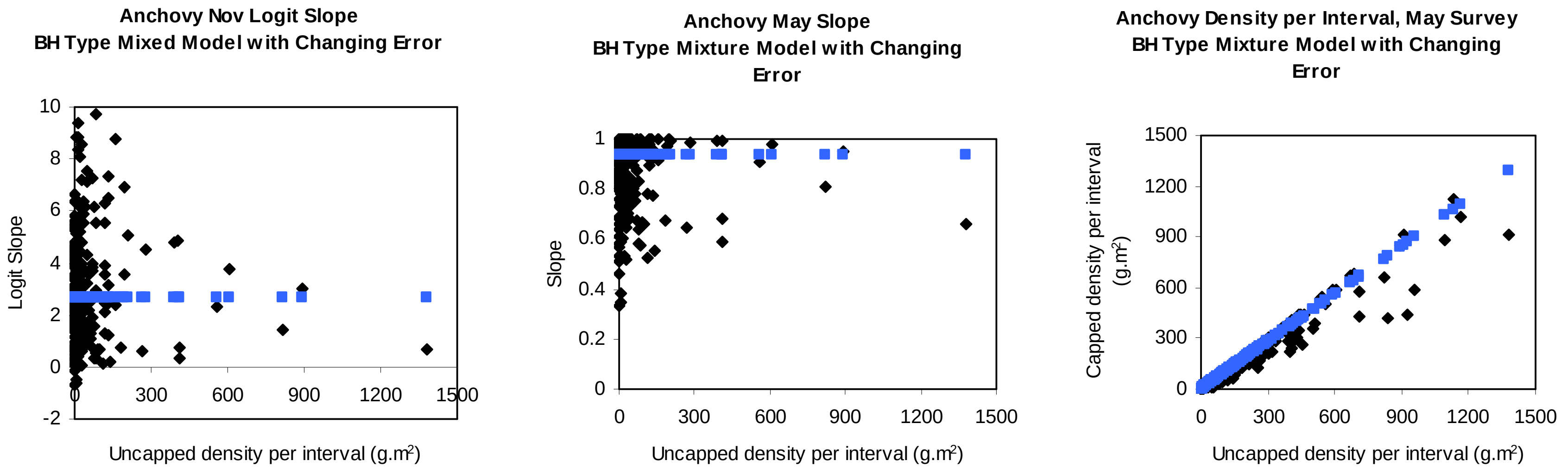

Figure 5. The model fits to the observed logit slope and the consequent slope "regression" of capped against uncapped densities per interval for the anchovy May survey.

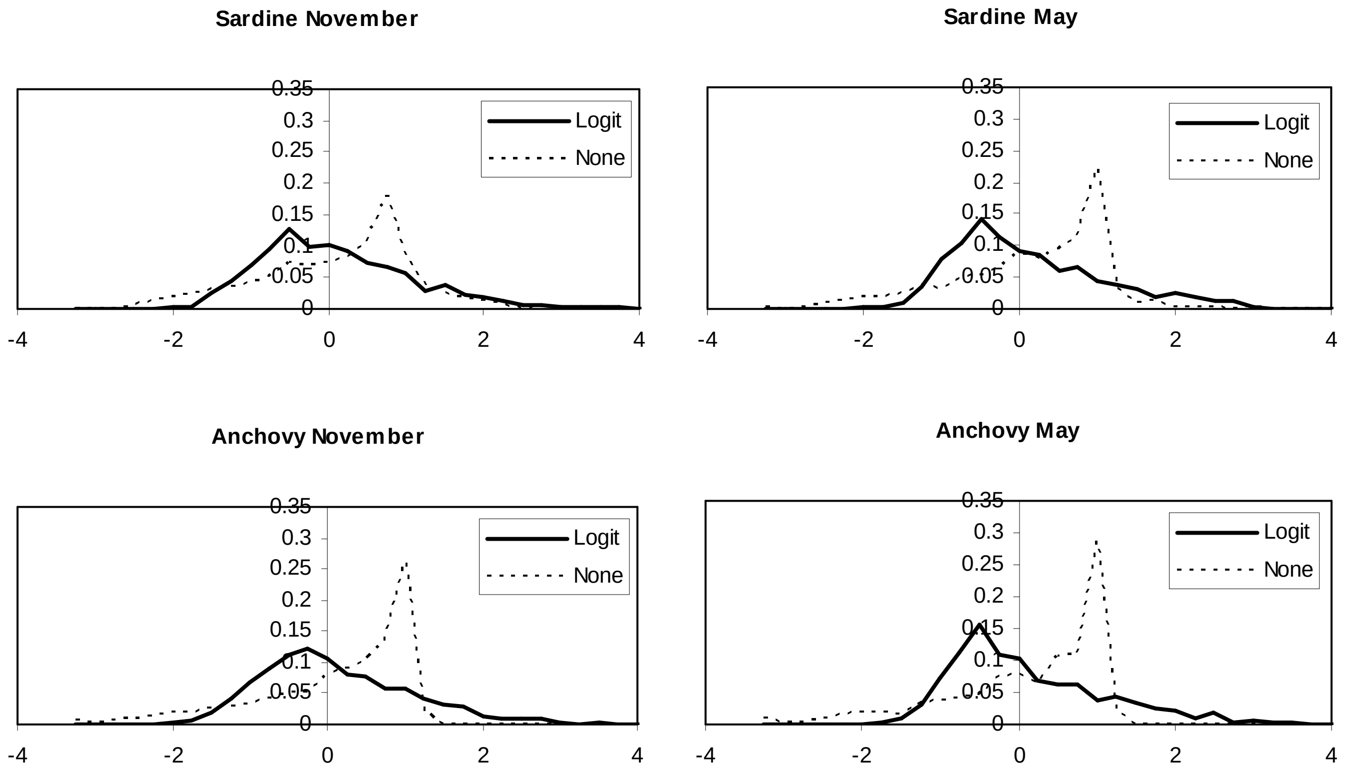

Figure 6. Probability density functions of the standardised residuals from the regression of capped data against uncapped data. Pdfs are given for both the case of fitting to the logit transform of the slope and the case of fitting directly to slope (no transformation).

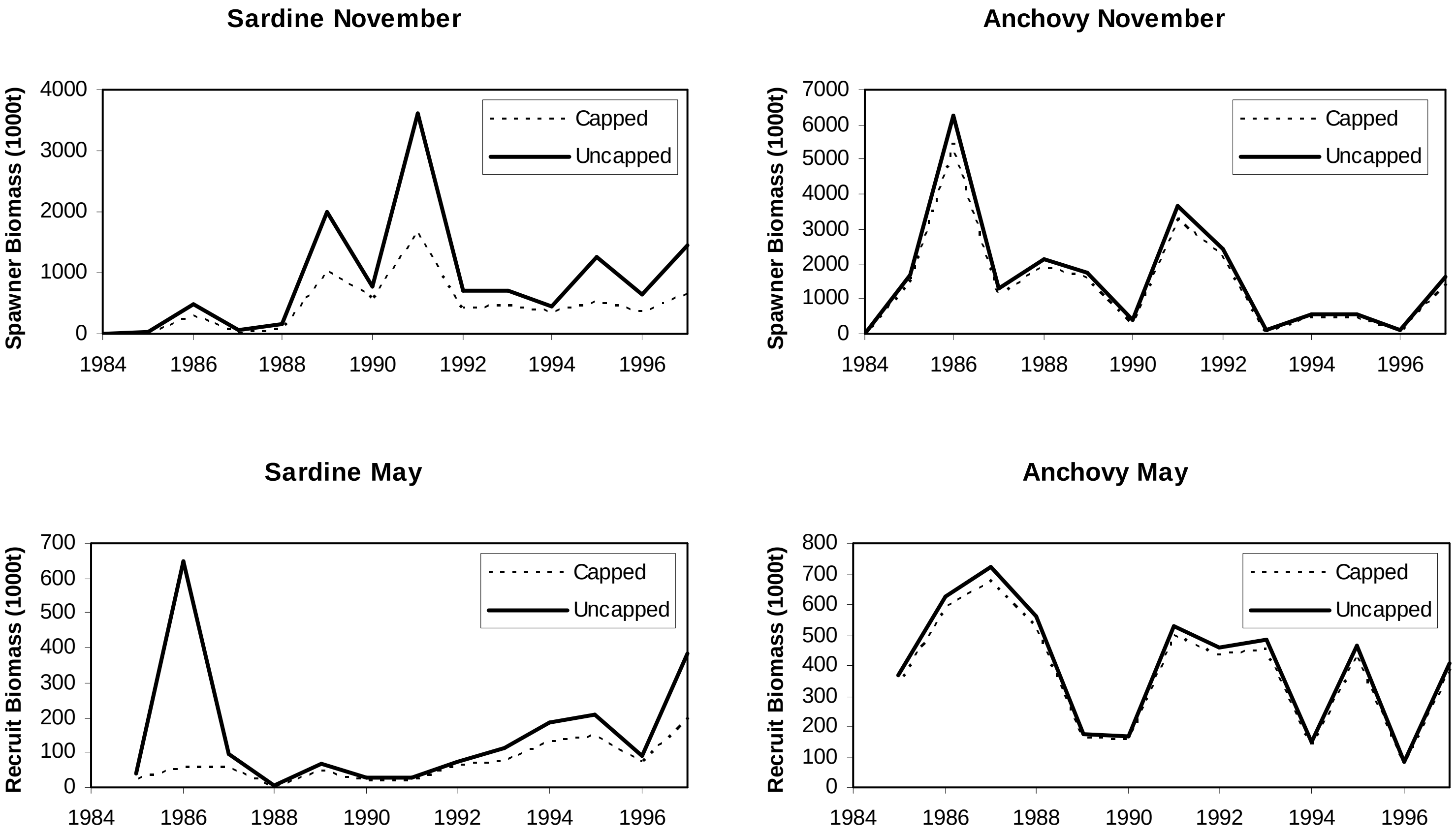

Figure 7. The capped new target strength (no attenuation for sardine) annual biomass and calibrated uncapped new target strength (with attenuation for sardine) annual biomass.