

Update of the hake species split models in the light of more recent survey data and a revision of the large/medium/small size classification.

by

J.D. Gaylard and M.O. Bergh

OLRAC

Ocean and Land Resource Assessment Consultants

Suite 4, Silvermine House

Steenberg Office Park

Tokai 7945

November 2009

Summary

At the request of the Demersal Working Group, we have conducted a revision of the hake species split models using up-to-date survey data. Despite the additional data and some slight changes in methodology, the results do not indicate a substantive departure from the prevailing predicted aggregate proportions of *M.capensis* in the offshore catch.

However, skipper estimates of size composition appear to be both biased and highly variable estimators of the true size composition when compared to OROP observer data, leading to a resulting under-estimate of the proportion of *M.capensis* of approximately 2 to 3% (of total hake catch) on aggregate, when applied to a typical space/size distribution of commercial catch.

We present an alternative species split model, based on a revised size-classification, as a candidate mechanism for correction of this bias.

The authors are also in the process of testing the survey-based models on SADSTIA observer data. The results of this part of the study will be presented in a separate document.

1. Introduction

Gaylard and Bergh (2003, 2004a) used survey data to develop a depth and size based algorithm for splitting hake catches into the component species *M. paradoxus* and *M. capensis*. In the absence of species information in the commercial catch and effort logbooks, this algorithm has been used since 2004 to produce species-specific catch and CPUE data for input into the stock assessments. Gaylard & Bergh (2007) tested the algorithm using observer data collected by CapFish on behalf of SADSTIA, and found reasonable agreement at an aggregated level.

However, before the current study, the models have not been re-trained to updated survey data since 2004, and so, at the request of the Demersal Working Group, we conduct this exercise using data from Demersal survey cruises up to Autumn 2009.

Separate models are applied to the “large”, “medium” and “small” components of the hake catch as recorded in the skipper logs. The fit of the models thus requires some assumption of the distribution of lengths of hake corresponding to each of these size categories. The assumption used in the development of the existing formulae has been one of mutually exclusive categories with the following ranges:

Large: 58 cm +; Medium: 43 cm to 57 cm; Small: 21 cm to 42 cm

which represents a compromise between the classifications assumed to be in use at the time by I&J and Sea-Harvest, and as reported by Leslie, Rose and Scholte (1998). A second component of the study addresses the concern that there may be bias in the skipper estimates of large, medium and small fish in the hake catch and that the assumed size ranges may be inappropriate.

A third component of the study, which is still work in progress, is to use the SADSTIA observer data to further evaluate the accuracy of the models. Results from this component will be tabled in a separate document.

2. Data

The following data have been used in the study:

- A. Length frequencies per station for the 2 hake species for each of the Demersal Survey Cruises from Summer 1984 to Autumn 2009, referred to hereafter as **SURVEY data**.
- B. Observer Length frequencies per trawl for the 2 hake species together with corresponding vessel, time, depth and location data collected by CapFish on behalf of SADSTIA for the species split study from November 2005 to August 2009, referred to hereafter as **SADSTIA data**.
- C. Observer length frequencies per trawl for the 2 hake species, together with corresponding vessel, time, depth and location from the Ocean Research Observer Program, referred to hereafter as **OROP data**.
- D. Commercial trawl by trawl skipper log data from 2000 to 2008 including vessel, position, depth, date, time, hake catch and breakdown into size categories, referred to hereafter as **SKIPPER data**.
- E. A reference table mapping vessel number (as in the SKIPPER data) to vessel name (as in the OROP data.)

3. Methods A. Testing the SMALL, MEDIUM, LARGE classification.

The principle method here is the comparison of length frequencies recorded by OROP observers with the large/ medium/ small breakdown recorded by skippers for what may reasonable be assumed to be the same trawl.

The matching of SKIPPER trawl records with OPOP records is not a trivial exercise, largely due to the fact there is no common trawl identifier key linking these two data sources. Thus any basis for matching records must be based on comparing vessel name, location, date and time. Of course it is very unlikely that exact

agreement would occur between skipper and observer, particularly with regard to location and time, so some allowance for error must be made.

For the purposes of this study, we have assumed that a skipper record and an observer record apply to the same trawl if:

- The vessel name agrees.
- The date and time of hauling agrees to within 30 minutes
- The location agrees to within 30 miles
- The depth agrees to within 20metres.

OROP Length frequencies have been converted to proportions by mass per 1cm length class, assuming a cubic mass/length relationship.

These Masses have been summed over the traditional “large”, “medium” and “small” length ranges (the “OLD” classification in Table 1) to produce estimates of mass proportion in each of these categories.

These observer estimates have then been directly compared with the corresponding skipper estimates from the matched trawls.

4. Results and Discussion A.

17383 records were made available to us from the OROP data, over the period 19 July 2002 to 31 Dec 2008

Of these, 14667 matched on vessel name, date and hour.

Of these, 10402 agree on depth within 20m and position within 30 miles.

Of these, 6732 contained (from the SKIPPER data) a break-down of hake catch into large, medium and small categories. It is these trawls that constitute the results below.

Figure 1 illustrates the large-scale variance of observer-estimated proportions vs skipper-estimated proportions for each size category, where the observer proportions are according to the “OLD” classification.

Table 1 shows the bias on aggregate when the predicted proportions are summed over all 6732 matched records. Also shown in this table is a break down into the three largest companies and “other”. Note that the skipper’s estimates are substantially overstated for the small category and understated for the large category, and that this result is consistent over the 4 companies. This is illustrated in Fig.2

One way of “correcting” this bias is to adjust the boundaries between size categories in order to achieve a matching of aggregated proportions. i.e. until the observer proportions agree with the skipper proportions in Table 1. We have done this exercise, the results appearing under “NEW” in Table 2. The Medium/large boundary needs to shift from 58cm to 69cm, while the small/medium boundary needs to shift from 43 to 50cm. Of course this is an over-simplification of the situation, where the de facto distributions for the three categories overlap considerably. Fig 3 illustrates this against the backdrop of a mass-histogram of length frequencies in the matched OROP data.

5. Methods B. Updating the Species split Model.

We investigate and compare 3 variants of the species splitting model:

Model 0 is the model as per Gaylard & Bergh 2004a, which is currently in use.

Model 1 is a refit of the model using updated survey data up to and including the Autumn 2009 demersal survey cruise.

Model 2 is a refit of the model, using updated survey data as per Model 1, but adopting the revised size classification suggested above.

The Model Equation

The proportion of *M.capensis* in each trawl *i* and coast/size class component *j* is calculated by:

$$\bar{p}_{i,j} = \frac{1}{1 + e^{B_{i,j}}} \quad (1)$$

$$\text{with } B_{i,j} = \kappa_j [d - (d_j^* + \alpha_{y,j} + \beta_{L,j})] \quad (2)$$

where:

- d is the trawl depth in metres;
- κ_j is the slope parameter for and size/coast component j ;
- d_j^* is a the shift for and size/coast component j ;
- $\alpha_{y,j}$ is the year parameter for year y and size/coast component j ;
- $\beta_{L,j}$ is the alongshore parameter for alongshore category L and size/coast component j ;

Changes in methodology:

Under Model 0, alongshore and year factors were assumed to be dependent on coast, but independent of size class. In Models 1 and 2, these are assumed to be independent of both coast and size class.

Under Model 0, year factors were found to be not significant on the South Coast, and hence excluded. Under Models 1 and 2, year factors are found to be significant on both coasts.

Model 0 was fitted to the SURVEY data on a trawl by trawl basis, using ADMB software by maximising a likelihood function based on the assumption of binomially distributed errors. (See Gaylard & Bergh 2003 and 2004a for details.) Models 1 and 2 are fitted using SPSS software by binomial logistic regression on a fish by fish basis, (so that the quantity modelled is the “probability of a single hake being *M.capensis*” rather than the “expected proportion of *M.capensis* in a trawl”).

6. Results and discussion B.

Tables 3 and 4 show parameter estimates from each of the 3 Models.

Figs 4 and 5 compare the predicted proportions resulting from the three models under median conditions with respect to alongshore position and year.

Fig.6 shows, on a year by year basis, since 2004 the overall proportion of *M.capensis* in the offshore hake catch. To obtain this plot we have applied a typical distribution of catch over depth, coast and size class and integrated the model predictions over this distribution. The “typical distribution” used is the average distribution from the SKIPPER data for years 2004 to 2008. Note that the Model 0 series is constant, because no data more recent than 2004 were used to train the model. For all three models, the 2010 and 2011 points are projections using median year factors.

Fig.6. suggests that Model 0 presents no bias of consequence in the application of Model 0 as a proxy for Model 1. However, both of these appear to understate the proportion of *M.capensis* systematically by 2 to 3 % (of total hake catch) by comparison with Model 2.

As is evident from Figs 4 and 5, most of this bias stems from the “medium” size category, particularly on the South Coast in the 200m to 400m depth range. For example, at a depth of 300m, Models 0 and 1 predict that 15% of hake are *M.capensis*, while Model 2 predicts 85%.

7. Conclusions.

The update of the species splitting model by inclusion of more recent survey data does not produce any substantive departure from the prevailing aggregate proportions of *M.capensis* in the offshore catch.

However, skipper estimates of size composition of appear to be both biased and highly variable estimators of the true size composition, leading to to a resulting bias in the estimation of species splits. This can be “corrected for” by the adoption of Model 2 as an alternative splitting algorithm, but this is only one of several solutions which could be considered.

The variance in skipper estimates warrants further attention. Thought should be given to the way in which the logbook information is collected. Also, use of location information has been used recently on a trial basis to interpolate size mix information where this has been unavailable in the data used for CPUE analysis. Thought should be given to applying a similar principle more widely, or perhaps even estimating species composition directly from position.

The ideal, of course, is the direct recording of species at the point of catch.

8. References.

Gaylard J.D. and M.O. Bergh. 2003. An investigation into the procedure used to split commercial catches of hake on the South African South Coast into *Merluccius paradoxus* and *Merluccius capensis*. BEN/JAN04/SAH2b

Gaylard J.D. and M.O. Bergh .2004a. A size-dependent species splitting mechanism applied to hake catches off the South African West Coast. WG/08/04/D:H:13

Gaylard J.D. and M.O. Bergh M. 2004b. A species splitting mechanism for application to the commercial hake catch data 1978 to 2003. Marine Coastal Management Document WG/09/04/D: H: 21. 8 pp.

Gaylard J.D. and M.O. Bergh. 2007. Further comparison of hake species splits from observer data with the survey-generated splitting algorithm.

Leslie R.W. B. Rose and J. Scholte. 1998. Hake grading by Irvin & Johnson and by Sea Harvest. WG/01/98/D:H:03

9. Acknowledgements.

Thanks to Tracey Fairweather (MCM) and Melanie Smith (CapFish) for their considerable efforts in providing the data in the format needed.

10. Tables and Figures.

Table 1: Comparison of Skipper and Observer estimates of the proportions of large, medium and small fish in the hake catch, aggregated over all matched OROP/SKIPPER records.

	All Companies		I&J		Sea Harvest		Atlantic		Other	
	Skipper Estimate	Observer Estimate								
Small	43.1%	23.5%	47.0%	22.6%	52.0%	29.9%	43.0%	26.1%	31.0%	19.9%
medium	44.2%	33.6%	39.0%	34.0%	36.0%	34.6%	47.0%	32.3%	57.0%	32.6%
Large	12.7%	43.0%	14.0%	43.4%	12.0%	35.5%	10.0%	41.6%	12.0%	47.5%

Table 2: Size Classifications. I&J and “Sea Harvest” classifications are those assumed to be in the use by these companies as reported in Leslie, Rose and Scholte (1998). “OLD” is the median of these two, and is the classification used by Gaylard & Bergh (2003, 2004a) to fit the species split model (Model 0 in Section 5 below) to survey data. “NEW” is the classification required to obtain agreement between observers and skippers of aggregated proportions over matched OROP/SKIPPER trawl records.

	I&J	Sea Harvest	“OLD”	“NEW”
Large	65 cm +	52 cm +	58 cm +	69 cm +
Medium	45 cm to 64 cm	41 cm to 51 cm	43 cm to 57 cm	50 cm to 68 cm
Small	21 cm to 44 cm	21 cm to 40 cm	21 cm to 42 cm	21 cm to 49 cm

Table 3. Parameters estimates from Models 0,1 and 2 for each size category on the *West Coast*.

		West coast								
		Model 0			Model 1			Model 2		
		Large	Medium	Small	Large	Medium	Small	Large	Medium	Small
	κ	-0.02784	-0.03325	-0.04722	-0.03724	-0.03707	-0.04694	-0.03759	-0.03632	-0.03938
	d^*	317.09	274.25	168.95	379.66	300.93	217.98	371.54	361.98	217.97
longshore (latitude) factors β_L	North of 29°S	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00
	29° to 30°S	-4.02	-4.02	-4.02	-30.50	-47.35	-49.81	-40.43	-39.61	-49.01
	30° to 31°S	4.81	4.81	4.81	-27.77	-43.69	-51.74	-3 1.48	-38.30	-49.21
	31° to 32°S	1.99	1.99	1.99	-21.33	-52.04	-41.81	-3 7.15	-33.85	-41.84
	32° to 33°S	5.75	5.75	5.75	-40.72	-57.25	-30.87	-5 7.35	-47.15	-33.19
	33° to 34°S	14.93	14.93	14.93	-68.62	-60.83	-28.85	-89.78	-60.22	-31.68
	34° to 35°S	34.81	34.81	34.81	-47.40	-39.53	-18.12	-71.77	-44.21	-16.14
	South of 35°S	36.27	36.27	36.27	-47.31	-58.96	-25. 60	-56.88	-59.02	-24.53
year parameters α_Y	1984	14.04	14.04	14.04	-15.91	53.20	35.54	0.69	-4.93	42.40
	1985	21.95	21.95	21.95	-5.96	76.11	31.57	33.27	30.80	37.69
	1986	13.52	13.52	13.52	17.51	71.78	3.44	43.08	19.34	19.06
	1987	8.02	8.02	8.02	-1.27	61.00	-3.03	24.68	19.88	5.51
	1988	0.5	0.5	0.5	8.34	25.49	37.50	19.67	-28.72	44.05
	1989	11.34	11.34	11.34	-23.43	52.14	31.55	33.32	10.76	39.11
	1990	32.73	32.73	32.73	2.05	2.87	1.26	44.64	-39.31	4.45
	1991	11.45	11.45	11.45	-49.43	46.96	12.79	-30.48	-8.41	21.81
	1992	21.14	21.14	21.14	-15.61	39.84	2.97	21.11	-1.67	5.36
	1993	16.31	16.31	16.31	-15.65	42.00	-9.63	15.99	-3.91	-6.03
	1994	4.84	4.84	4.84	0.00	0.00	0.00	12.35	-6.64	6.84
	1995	26.7	26.7	26.7	-13.57	48.88	-1.14	29.65	-16.17	-12.46
	1996	-6.6	-6.6	-6.6	-15.49	-5.49	-13.98	40.54	-26.57	2.01
	1997	7.22	7.22	7.22	-3.83	5.68	-0.04	18.51	11.28	-4.96
		5.25	5.25	5.25						
	1999	4.07	4.07	4.07	14.67	39.43	33.98	27.50	7.38	37.23
	2000	5.25	5.25	5.25	8.31	94.44	19.88	57.91	73.25	29.05
	2001	5.25	5.25	5.25	69.08	43.73	-4.83	113.27	17.67	2.06
	2002	21.51	21.51	21.51	13.87	1.61	-19.71	38.06	-14.40	-18.94
	2003	0	0	0	-5.62	68.69	22.69	37.04	29.58	25.41
2004	5.25	5.25	5.25	33.74	54.98	-12.66	89.63	8.18	-6.22	
2005	5.25	5.25	5.25	12.85	28.49	-17.11	58.52	-2.95	-18.14	
2006	5.25	5.25	5.25	-3.32	23.53	24.87	36.40	-14.20	26.19	
2007	5.25	5.25	5.25	-11.61	73.50	-8.69	38.13	27.54	5.75	
2008	5.25	5.25	5.25	-12.34	0.00	0.00	26.61	0.00	0.00	
2009	5.25	5.25	5.25	0.00	0.00	0.00	0.00	0.00	0.00	

Table 4. Parameters estimates from Models 0,1 and 2 for each size category on the South Coast.

		South coast								
		Model 0			Model 1			Model 2		
		Large	Medium	Small	Large	Medium	Small	Large	Medium	Small
κ		-0.02085	-0.03786	-0.09074	-0.03457	-0.05547	-0.15184	-0.03813	-0.04083	-0.10559
d*		386.85	257.29	181.62	346.75	238.82	206.64	389.05	291.24	200.59
longshore (longitude) factors βL	West of 21°E	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00
	21°to 22°E	18.92	18.92	18.92	60.40	44.52	11.04	59 .06	65.61	15.35
	22°to 23°E	-20.74	-20.74	-20.74	-107.35	-65.27	-3 1.23	-76.89	-94.03	-29.31
	23°to 24°E	-33.63	-33.63	-33.63	-102.50	-81.32	-3 0.82	-55.28	-101.12	-31.94
	24°to 25°E	-34	-34	-34	-72.34	-74.22	-36.81	1.14	-90.00	-31.96
	25°to 26°E	-11.64	-11.64	-11.64	15.10	24.78	-23.9 1	35.76	34.49	-19.44
	East of 26°E	44.51	44.51	44.51	44.10	-8.17	-22.50	83.01	11.49	-10.75
year parameters αY	1986				88.72	31.47	-21.40	14.40	66.73	-9.54
	1987				44.63	32.34	-29.00	-22.76	51.97	-14.27
	1988				66.22	46.35	-13.28	34.78	57.35	-4.07
	1989				406.96	63.06	-24.05	296.02	380.37	-6.48
	1990				28.50	66.32	-21.82	322.57	65.43	-10.15
	1991				68.98	202.05	41.78	53.53	83.67	52.35
	1992				66.43	69.58	-21.71	-18.47	79.30	-3.10
	1993				78.08	28.07	-41.73	-3.21	62.79	-29.38
	1994				42.03	48.27	-30.41	-33.49	67.33	-17.88
	1995				29.59	60.06	-39.47	-19.33	47.98	-21.47
	1996				-35.72	21.34	-26.58	-90.84	1.33	-13.81
	1997				-33.04	22.47	-33.48	-78.25	2.76	-20.64
	1998									
	1999				8.79	8.97	-30.56	-39.93	18.35	-18.38
	2000				12.09	13.72	-17.38	427.23	25.48	-4.98
	2001				74.09	79.25	-25.22	-40.68	109.08	-7.32
	2002									
	2003				-46.28	13.49	-35.70	-73.82	-7.56	-22.65
	2004				16.73	29.50	-14.46	-52.96	43.17	-7.26
	2005				-35.95	29.52	-23.38	-122.34	24.24	-16.40
2006				18.67	12.64	-33.46	8.91	13.16	-23.44	
2007				12.18	34.67	-24.76	-33.59	43.02	-12.28	
2008				62.53	48.34	-26.02	-20.95	73.99	-13.65	
2009				0.00	0.00	0.00	0.00	0.00	0.00	

Figure 1. Scatter plots of observer estimates (pL1, pM1, pS1) vs skipper estimates (pL_skip, pM_skip, pS_skip) of the proportion of large/medium/small fish in the hake catch. Each point represents a trawl for which observer records have been matched to the skipper's log.

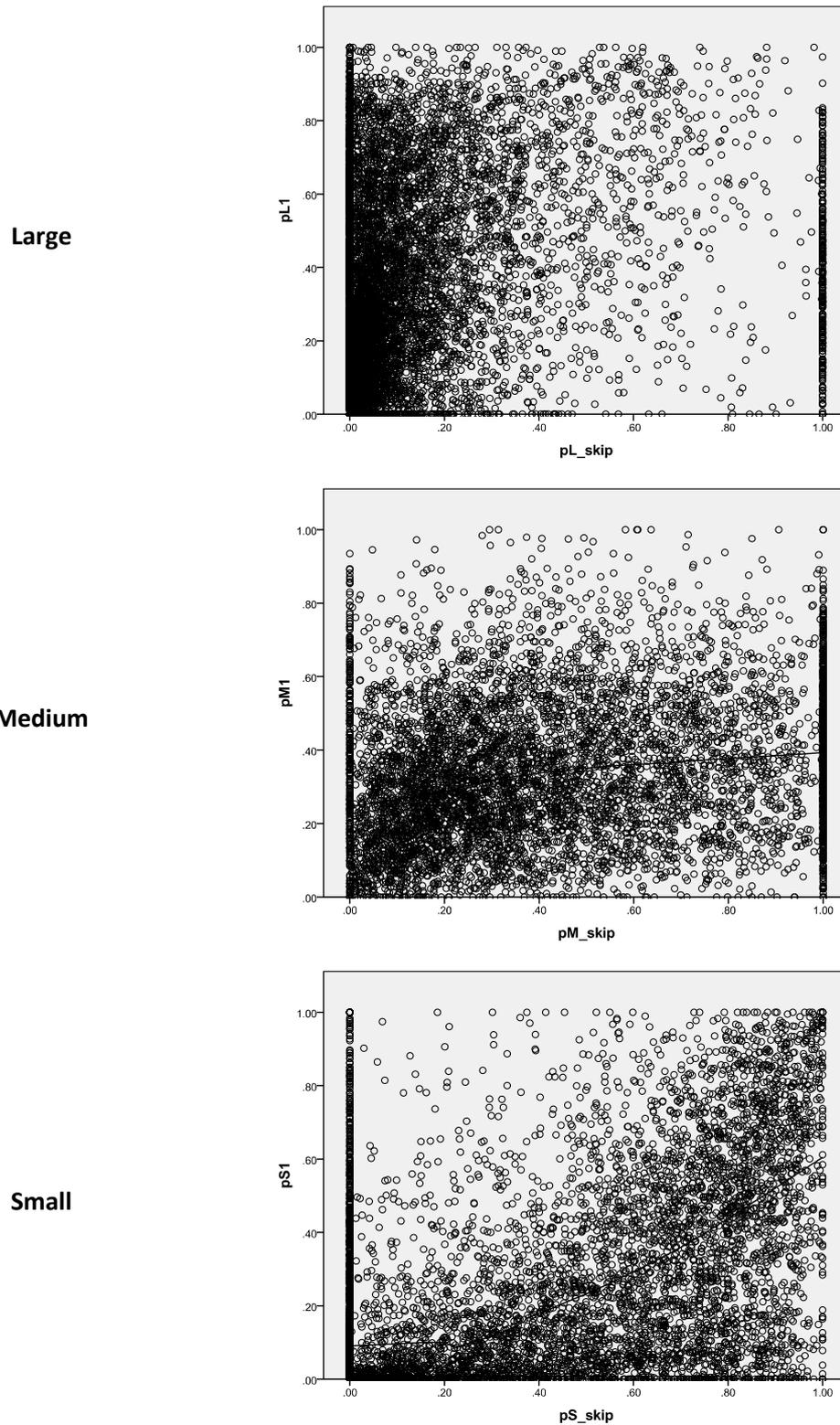


Figure 2. Comparison of Skipper Estimates (log book data) with observer estimates (OPOP data) of the proportions of large medium and small hake from trawls which have been matched between OROP and Log Book data.

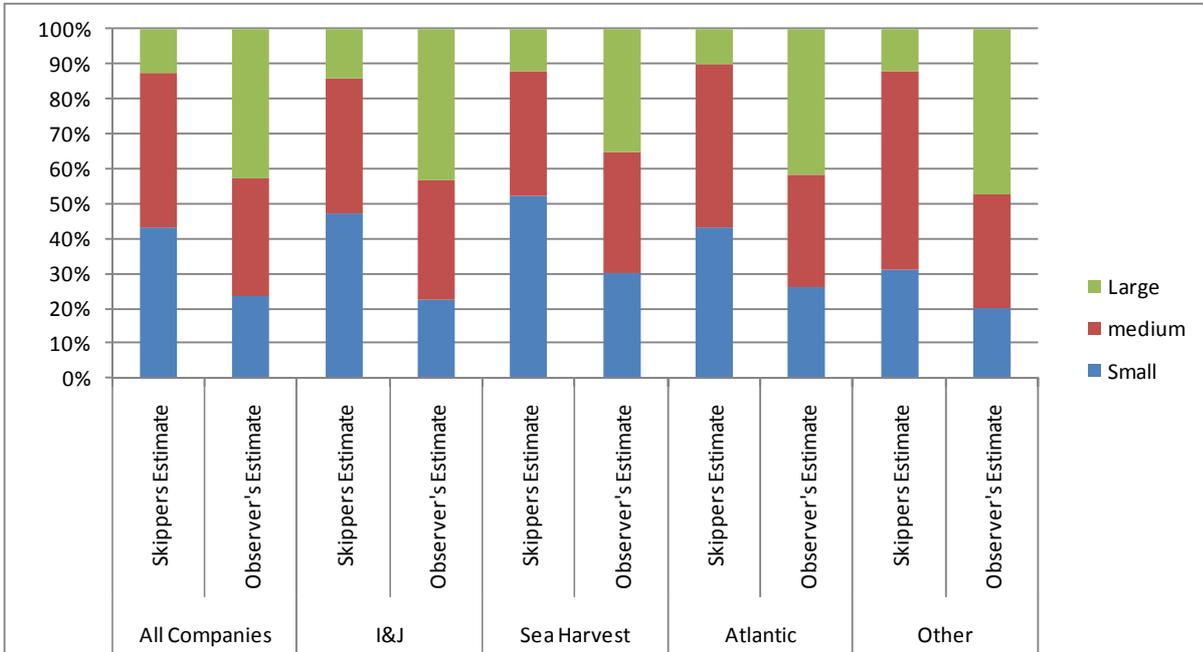


Figure 3. Proportions of catch (by mass) in 1mm length classes in the matched OROP/LogBook records. The Solid Vertical Lines show the delimiters for the "traditional" large/medium/small classification. Dashed vertical lines show the delimiters which are required to equate aggregated Observer and Skipper proportions.

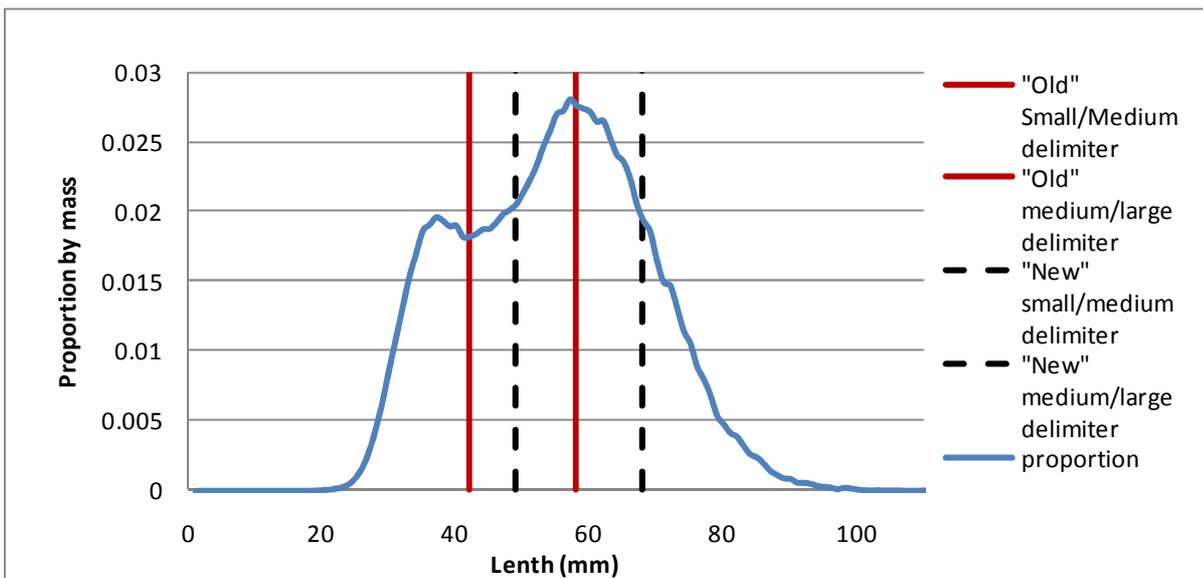
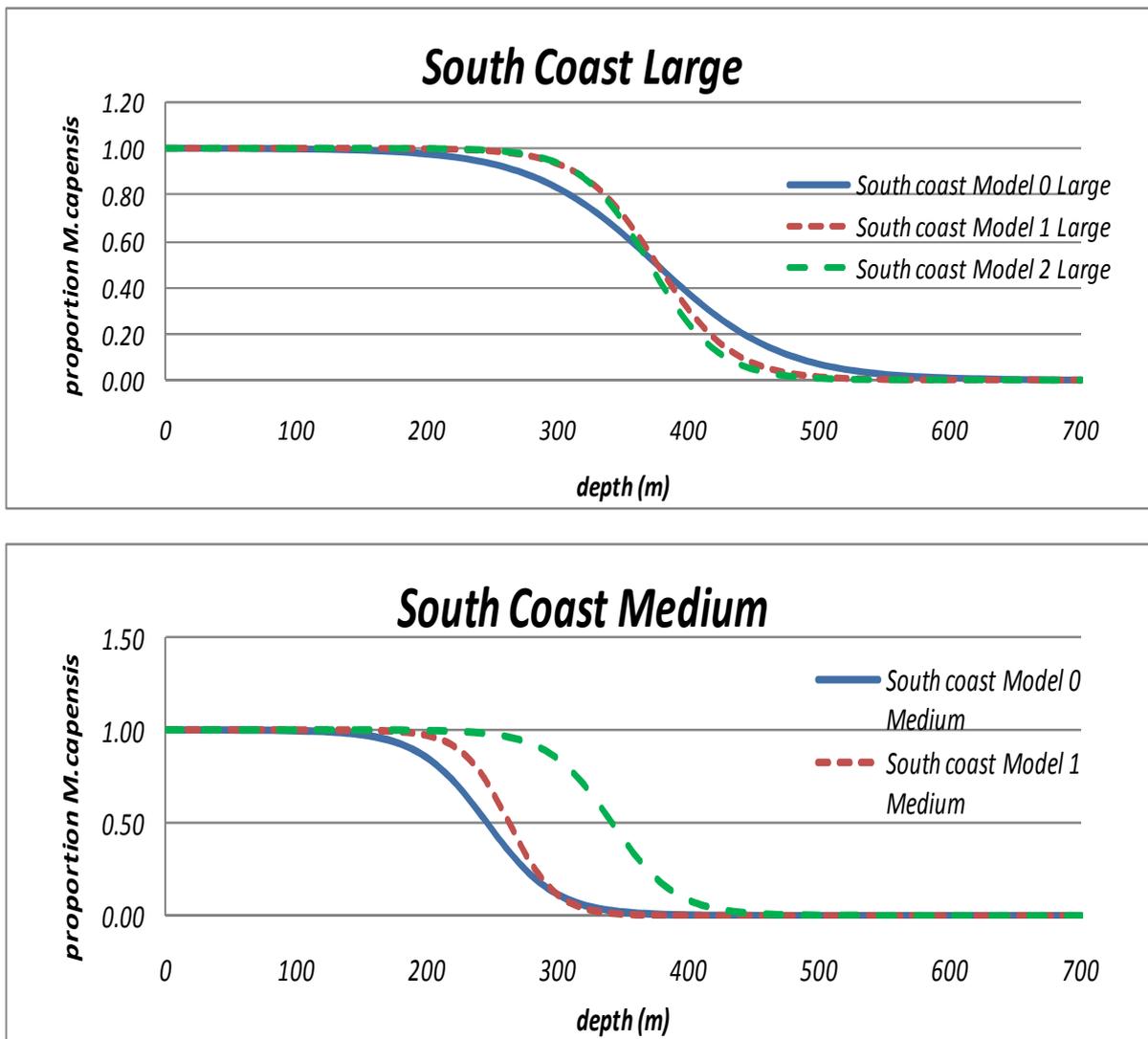


Figure 4. Comparison of the proportion vs depth relationships for each of three size classes on the *South Coast*, using Model 0 (the original model), Model 1 (with updated survey data) and Model 2 (with revised size definitions). In each case, the median alongshore and median year factors are used to generate the plots.



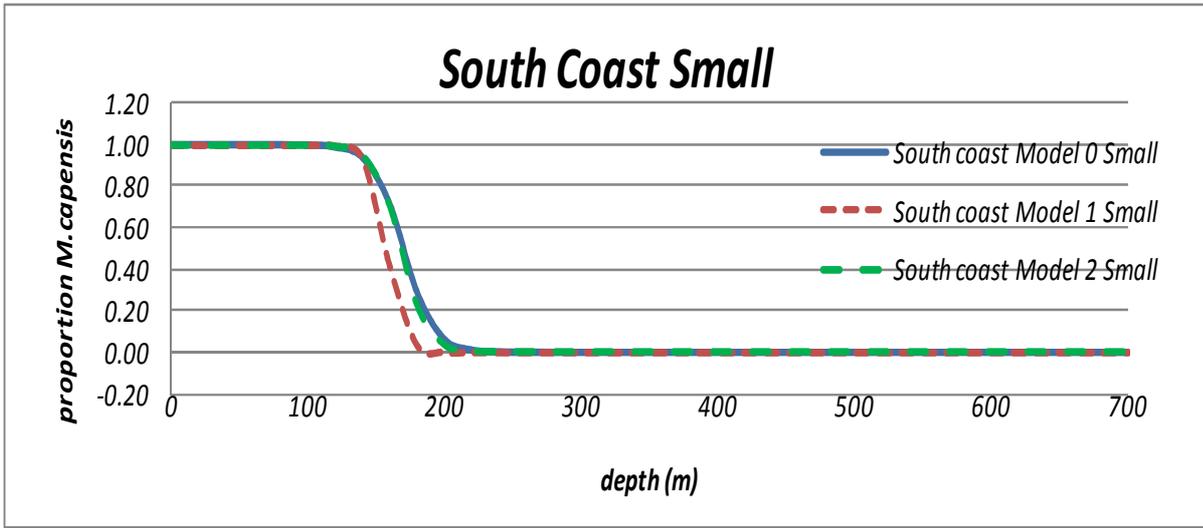
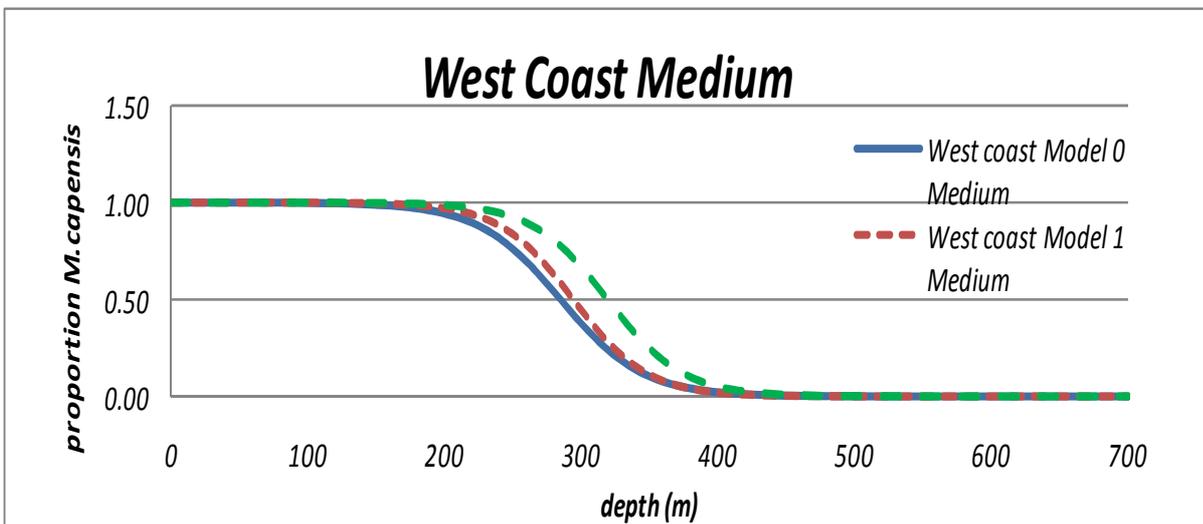
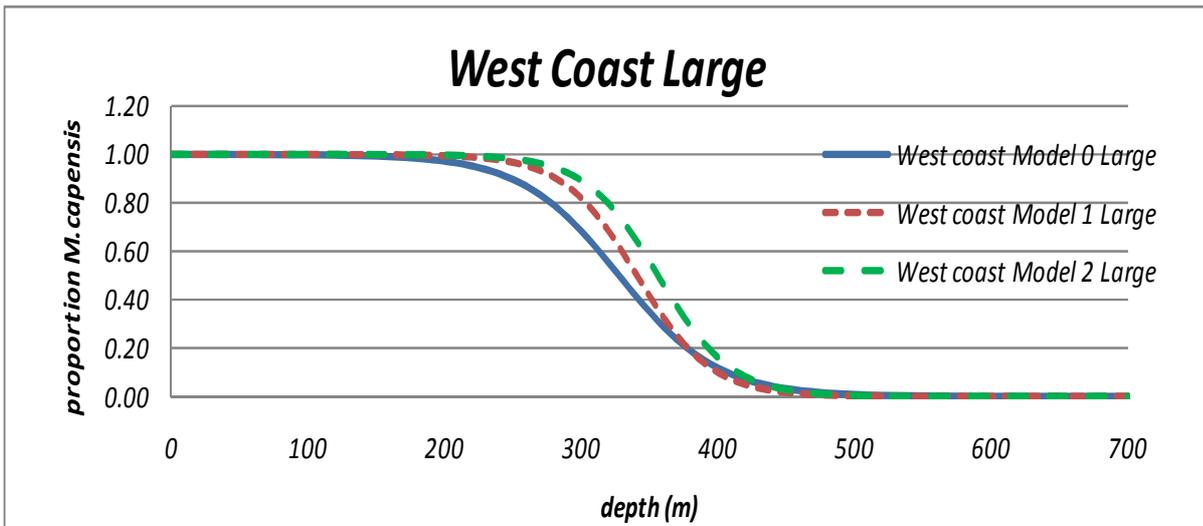


Figure 5. Comparison of the proportion depth relationships for each of three size classes on the **West Coast**, using Model 0 (the original model), Model 1 (with updated survey data) and Model 2 (with revised size definitions). In each case, the median alongshore and median year factors are used to generate the plots.



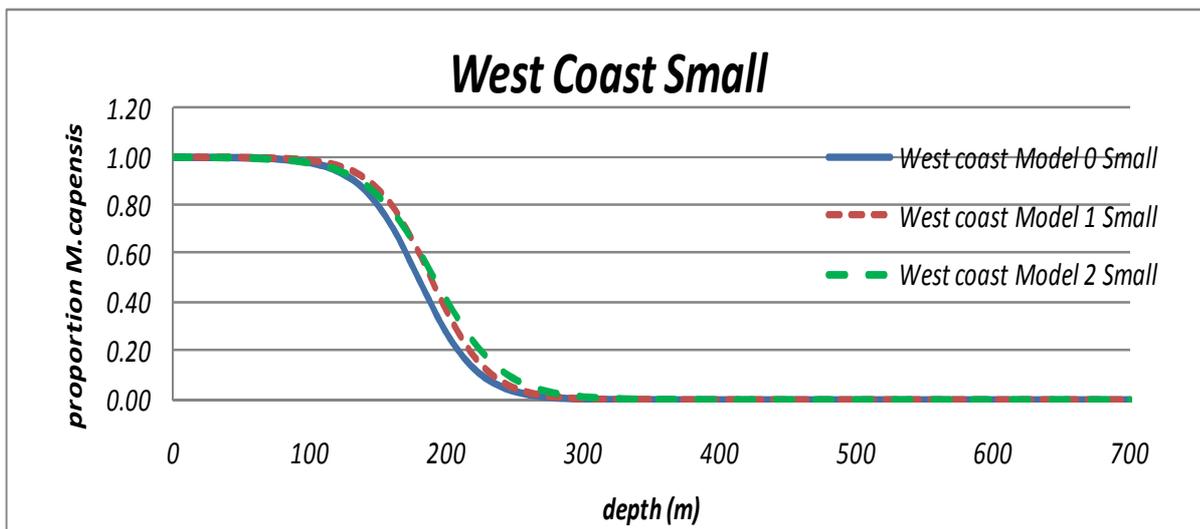


Figure 6. Comparison of the predicted overall proportion of *M. capensis* in the total offshore hake catch using Model 0 (the original model), Model 1 (with updated survey data) and Model 2 (with revised size definitions). In each case, the models are applied each size class on both coasts at 20m depth intervals then integrated over a typical space and size distribution of the commercial offshore hake catch.

