Response to the review panel report for the 2017 International Stock Assessment workshop: Hake

A. Ross-Gillespie and D.S. Butterworth[1]

Email: mlland028@myuct.ac.za

Note: comments on progress are inserted in *red italics* underneath each recommendation.

A. Hake

A.1 (*). Is the new selectivity model adequate/appropriate

The current formulation of selectivity should be used in the Reference Set of models. However, robustness tests should be developed to examine (a) the consequences of fishery selectivity changing in the future, and (b) an operating model in which at least one fleet or one survey has a near asymptotic selectivity pattern for *M. capensis*. The Panel does not, however, expect that OMP performance will be very sensitive to these factors because (a) the proposed OMPs use only abundance index data and not fishery age- or length-composition data, and (b) the selectivity for several of the surveys is close to asymptotic already.
The extent of the 'cryptic biomass' modelled should be reported for each survey.

A robustness test changing future selectivity has not been explored yet. A robustness test where the M. capensis trawl selectivities are forced to be asymptotic has been conducted (RT25 in MARAM/IWS/2018/Hake/P6b). The effect was to lower the M. capensis spawning biomass relative B_{MSY} in the short-term (although values were similar to the RC after 25 years of projection) and to slightly raise the biomass trajectory relative to K^{sp}. The negative log-likelihood was, however, substantially worse for this robustness test, with 90 points lost in the fit to commercial catch-at-length data and just over 7 points lost in the fit to the GLM CPUE data.

The cryptic survey biomasses have not been included in the main OMP documentation, but the estimates for the RC OM are reported in Table 1 in this document. These cryptic components are generally fairly small compared to the survey estimates. Plots of the RC survey selectivity-at-age functions have been reproduced from MARAM/IWS/Hake/P1 for ease of reference purposes (Figure 1).

A.2 Are the stock-recruitment models used adequate / appropriate, including the extent of annual variation abundance about these relationships?

The Panel explored evidence that the stock-recruitment relationship for *M. capensis* is domeshaped, and hence that reducing spawning biomass will increase expected recruitment from current levels, and agreed that the spawning biomass depletion and subsequent recovery rate implied by the historical catch is sufficient support for a dome-shaped relationship over an asymptotic one. The Panel noted that the steepness parameter was hitting its upper bound of 1.5. However, increasing this bound led to an estimate of steepness that did not differ much from 1.5.

No action requested, although the upper limit on h has now been increased to 2. The M. capensis h estimate for the RC is still at the upper boundary, but the decision was made to keep the upper boundary at 2 as this was considered more biologically realistic (see last paragraph of the first page of MARAM/IWS/2018/Hake/P3).

[1] Marine Resource Assessment and Management Group, Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701.

A.3 (*). Could the shrinkage procedure used for estimation of recent recruitments be improved?

The procedure could be improved when conducting projections by generating values for the recruitments that are currently shrunk to better capture uncertainty (i.e., implicitly using the full S_R, similar to the approach applied for South Coast Rock Lobster). However, this is not a high priority unless the species is assessed to be highly depleted. The impact of recent recruitment estimates on the evaluation of an OMP for hake will be limited because performance statistics are based on multi-decade projections.

No further action has been taken.

A.4. Currently OMP testing projections are initiated from MPLE estimates. How might taking estimation uncertainty into account best be achieved?

The Reference Set of trials on which OMP-2014 was based captured a wide range of biomass and depletion values, such that allowing for parameter uncertainty is unlikely to increase the range already considered in projections. Moreover, it is currently not possible to invert the Hessian matrix, which would be needed to initiate a procedure that could be used to generate parameter vectors. As for item A.3, the multi-decadal nature of the forecasts limits the impact of this uncertainty on the results of the OMP testing.

No further action has been taken.

A.5. Was the basis for the previous Reference Set (RS) selection adequate, and if not how should future selections be made?

The Reference Set can be reduced in size by eliminating the factor for natural mortality (see item A.7 for additional details). Consideration should be given to including a scenario in which the model projection starts in 1978 rather than 1917 in the Reference Set, as this scenario would exclude assumptions related to earlier years that are highly uncertain. If this scenario is added, the Reference Set will consist of 12 trials (four ways to handle the uncertainty of the split of the historical catches between *M. paradoxus* and *M. capensis*, and three options for the stock-recruitment relationship).

The predation mortality-at-age vectors have been assumed for the RS OMs and some robustness tests have been conducted testing sensitivity to these vectors. A model starting in 1978 was attempted, but it was not possible in the time available to develop it to a standard acceptable for the RS. More details on the progress that has been achieved can be seen in the Appendix of MARAM/IWS/2018/Hake/P6b.

A.6. Was the previous set of robustness tests adequate?

During its consideration of other questions, the Panel identified additional robustness tests (see items A.1, A.8, A.9, A.12 and A.13).

See relevant sections.

A.7. Should a change be made to use of *M*-at-age estimates from the hake predation model – an average over last 2-3 decades, or time varying by year since commencement of fishery?

The *M*-at-age vectors in the current Reference Set were selected semi-arbitrarily. In contrast, the estimates of *M*-at-age in MARAM/IWS/2017/Hake/P5 are based on an approach that explicitly accounts for time-varying predation due to hake. The Panel therefore **recommends** that future assessments and operating models be based on the average *M*-at-age by species from this predation model. The average should be based on the years from 1984 to the present, when data to inform year-class strength were available. The predation model is still a work-in-progress, but its use is still preferable to a somewhat arbitrary selection of *M*-at-age values.

This has been implemented for the RS OMs used in testing OMP2018.

A.8 (*). Should a penalty function on the survey *q*'s be included (e.g. to restrict to values below 1).
The Panel notes that several of the estimated survey *q*'s exceed 1 for the Reference Case operating model. This is unexpected because the surveys do not cover the full spatial distribution of *M. paradoxus*. Moreover, the surveys are for one coast only whereas the model predictions apply to the combined abundance on both coasts. Both of these factors would lead one to expect that the estimated survey *q* values would be below 1. The apparent discrepancy is greatest for the west coast summer survey and the south coast spring survey for *M. paradoxus*. The Panel considered possible reasons for the unexpected high survey *q*'s, including model mis-specification and herding, but there is no clear support for any of the reasons considered. Thus, the Panel **recommends**:

- the operating model should be fitted restricting survey $q < 1$ and examining which data sets are fitted appreciably worse than before, suggesting a conflict between the model with survey $q < 1$ and those data; and

- the model with survey *q*'s restricted in this way should be considered a robustness test.

See robustness test RT13 of MARAM/IWS/2018/Hake/P6b. Restricting the survey qs to be less than one has the effect of reducing the estimates of both species relative to B_{MSY}, although the M. paradoxus estimate of B^{sp} relative to K^{sp} is more optimistic. The negative log-likelihood for this robustness test is worse than the RC by 6.66 points, but this deterioration is fairly evenly spread between worse fits to the CPUE data, survey abundances and commercial catches-at-length, which are countered by a better fit to the survey catches-at-length. Thus, there is no one data source that stands out as a driving force behind the q>1 estimates.

A.9 (*). Should a penalty function on von Bertalanffy L_∞ values be included?
The parameterization of the von Bertalanffy growth curve should be changed from L_5, ln(κ) and t_0 to L_1, L_5 and ln(κ), which should improve convergence. In addition, a robustness test should be conducted in which a lower bound is imposed on κ.

This has not been implemented yet.

A.10. How best should (differing?) *q* values for surveys be estimated given gear changes and sometime use of industry vessels?
The Panel agreed with the analysts that the data for industry vessels should be downweighted to reflect lack of knowledge of survey *q* (except for the one vessel, *Andromeda*, which had conducted sufficient surveys to allow a separate *q* to be estimated). In addition, operating model projections should be undertaken in which the survey *q* for each future year is generated from a distribution that reflects uncertainty regarding possible *q*'s for industry vessels acting as survey vessels.

Robustness test RT19 (MARAM/IWS/2018/Hake/P6b) down-weights the negative log-likelihood contributions from abundance estimates arising from surveys conducted by industry vessels by a factor of 10. This did not have a notable impact on assessment results. MARAM/IWS/2018/Hake/BG5 estimates the ratio between industry and research vessel q's,as well as the associated uncertainty (median estimate of 0.797 with se of 0.178). A run of the RC OM estimating a separate q for the industry vessel surveys yielded a ratio of industry to research q of 0.80, i.e. comparable to that of BG5. Robustness test RT2 (MARAM/IWS/2018/Hake/P6a) simulates a situation where all future surveys are conducted by industry vessels and takes the uncertainty about the value of q into account. Simulations suggest that should this scenario occur in reality, the OMP will perform adequately provided the OMP is re-tuned accordingly.

A.11. Need the ageing error matrices used be reconsidered?
The existing set of ageing error matrices should be adequate for the current OMP revision unless data from a new age-reader are included in the assessment, in which case the ageing error matrices should be updated using all available double-read information.

No further action required.

A.12. For surveys might changing abundance estimation from the current random stratified to a geostatistical approach constitute an improvement?

The CVs of the estimates of abundance from the geostatistical approach tend to be lower than for standard design-based methods, as there is often spatial auto-correlation in density.

The DAFF Working Groups should establish a standard set of diagnostics to examine when reviewing survey results from the geostatistical approach. For example, the causes for large differences in point estimates from the geostatistical and design-based approaches should be understood, the residuals should appear random in time and space, and anisotropy estimates should be consistent with estimates based on simple analyses of the raw data.

In principle, the geostatistical approach should be applied to the survey size-composition data. However, this is not a high priority at present.

Operating models that are fitted to the geostatistical estimates of biomass (to 500m) should be included in the robustness tests. These estimates should be based on analysing data for each year separately, rather than relying on temporal as well as spatial auto-correlation. This is to avoid further complexity in the assessment methodology. Otherwise it will be necessary to modify assessment models to allow for temporal autocorrelation in survey indices by adopting a multivariate likelihood function if abundance estimates are based on the geostatistical approach.

This has not yet been done.

A.13. How best might results from the extension of surveys into deeper water be taken into account?

A robustness test should be developed that includes the data (index and size-composition) from the region from 500-750m as a separate time-series, with the index derived from a design-based analysis of the data. It seems unlikely that the results will be very sensitive to including these data, so conducting this robustness test should be considered low priority if it requires considerable recoding of the operating model.

This has not yet been done.

A.14. Should attempts be made to allow for some explicit movement, either as the basis for an updated RC or as a robustness test?

Implementing the movement model as an operating model requires the development of approaches to allocate future catches spatially. However, this is a substantial exercise and given that the movement model is not final yet, the development of even a robustness test should be deferred to the next time the OMP for hake is reviewed.

This has not been pursued for the development of OMP-2018.

A.15. (*). Is there any way of independently checking the *M. paradoxus / M. capensis* biomass ratio implied by the assessments?

Evaluation of this question is not straightforward because it is not clear that the *M. paradoxus* /*M. capensis* ratio is implausible given that only spawning stock biomass values are being provided in assessment reports. The Panel offers the following ways to explore this issue:

- Compare estimates of exploitable (or total) biomass in addition to spawning biomass.
- Compare estimates of biomass at size between predators and prey of suitable size.
- Examination of the ratio of catches to biomass over time, in the context of catch-rate and stock trends, could also potentially give information on absolute biomass estimates. However care would need to be taken that appropriate fishable biomasses are used in this analysis.
- Explore the *M. paradoxus* /*M. capensis* ratio using various measures of biomass from the model for hake that includes predation explicitly.
- Compare estimates of density for *M. paradoxus* and *M. capensis* from surveys.

- Evaluate the sensitivity of the *M. paradoxus* /*M. capensis* ratio in models in which catchability is restricted to be less than 1 for all surveys.

This has not yet been explored further apart from the first point – exploitable and total biomass estimates were examined during the RS conditioning phase, and the resulting ratios were deemed biologically plausible by the Demersal Working Group.

A.16. (*). How should the different Reference Set OMs be weighted in reporting performance statistics?

The choice of an OMP can depend on how the trials within the Reference Set are weighted (MARAM/IWS/2017/Hake/P9). Thus, work to evaluate weighting schemes could be highly consequential. However, identification of the best approach for weighting is still an area of research globally, as well as in South Africa. The Panel has the following **recommendations** in regard to weighting of trials within the Reference Set:

- Use of AIC weighting is **not recommended** because this type of weighting relies on assumptions that are unlikely to be valid, such as that all data are independent.
- If Multidimensional Scaling methods such as those in MARAM/IWS/2017/Hake/P9 are adopted, the discrepancy metric should account for the absolute as well as the relative scale of biomass and be based on historical and not future trends.
- Consider methods for model weighting based on predictive performance because this relates most directly to the reliability of projections.

The Panel notes that several groups are exploring ensemble approaches for integrating the results of multiple models (e.g., Robert Thorpe and Mike Spence at CEFAS, UK). The South African analysts should consult with these groups, and the Panel has provided email addresses to begin this cooperation. Overall, it may be prudent to base the selection of OMP18 on equal weighting, but to nevertheless examine how sensitive the final selection would have been to alternative weighting methods, with a view towards adopting a new weighting method for a future OMP revision.

The RS OMs have received equal weighting in the evaluation of the performance of OMP-2018.Constraints on time given deadlines for a recommendation precluded immediate investigation of alternative weightings.

A.17. Should slope as well as target approaches be used in the OMP's fundamental HCR? Should HCRs that react more rapidly to the most recent data be explored further (this is a particular concern of industry)?

It is ideal to consider a wide range of OMP variants to allow the best OMP to be identified. Slope-based HCRs may be capable of responding more rapidly to recent data. However, such HCRs may "follow noise" and lead to higher TAC variance. The lag between data being collected and changes in TAC being applied may also lead to poor OMP performance if slope estimates change rapidly. These trade-offs would need to explored in the analyses.

This has not been explored for OMP-2018, which uses the same HCR as OMP-2014. This was again a consequence of time constraints, and the fact that no obvious concern about the form of the existing HCR arose in discussions, which accorded higher priority to other issues.

A.18. Should HCRs that investigate the use of some recruitment index (probably from younger fish in survey) be explored further?

This option should be assigned low priority for this OMP revision due to the low likelihood that use of a recruitment index will improve performance, and also because the work to modify (and test) the additions to the operating model may be substantial. The Panel notes that if this approach is followed in future, it may be more robust to base OMPs on consecutive years with good (or poor) recruitment rather than basing management action on single year classes observed at small sizes.

This has not been explored for OMP-2018.

A.19. Other recommendations

A.19.1. (H) Explore the reasons for the inability to obtain a positive definite Hessian matrix when fitting the hake model. Ways to enhance the likelihood of obtaining such a Hessian matrix could include (a) replacing Pope's approximate by the "Hybrid method", (b) imposing soft (rather than hard) bounds on the parameters, and (c) setting the values of parameters that are clearly equal to their bounds (e.g., the residual standard deviations for the ICSEAF catchrate series) rather than trying to estimate them. Experience with other models that have initially failed to provide a positive definite Hessian matrix suggests an approach of initially estimating only a few parameters (e.g., R_0 and the recruitment deviations), fixing the remaining parameters, and checking for convergence (i.e., here a positive definite Hessian matrix). Repeating this procedure over an increasing number of estimated parameters is a good way to identify the parameters that create such problems in the model.

Some work went into coding the Hybrid method, but this did not seem to immediately fix the "nan" issue that occurred at times. Given time pressures it was not pursued further. While a positive definite Hessian has not been achieved yet, all the RS OMs were jittered to ensure that a global minimum had been achieved, and none of the final OMs encountered the "nan" issue which can hamper the minimisation process. The importance of further work in this area is however acknowledged.

A.19.2. (H) Whether the non-linear minimizer is converging to the true minimum of the objective function can be investigated using jittering (i.e., randomly perturbing the initial values of the parameters and re-fitting the model). Jittering should be a standard part of South African assessments based on complex models.

Jittering was conducted for all the RS OMs (at least five different starting positions), but not the robustness tests.

A.19.3. (H) There are some large changes (since 2013) to the results for the scenarios that form the Reference Set of trials (MARAM/IWS/2017/Hake/P4). While both data and model specifications differ between the new and old scenarios, some of the changes appear unexpectedly large and any such changes should be checked for possible convergence problems. In general, no results should be presented that have not had convergence verified.

Comparisons with 2013 have not yet been explored. This was mainly because the difference in results following a number of changes to the assessment over the last year proved to be large and consequential, so that effort was focussed on checking that these changes were justified.

A.19.4. (L) Consider applying the method developed by Methot and Taylor (2011) for specifying year-specific bias-correction factors for the stock-recruitment residuals once the assessment is able to provide a positive definite Hessian matrix.

This has not yet been explored.

Table 1: Cryptic biomasses estimated in the RC OM are reported. "Bsurv" is the total survey biomass, "Btot" is the survey exploitable biomass when the selectivity functions are adjusted to be asymptotically flat at 1, and Bcrit=Btot-Bsurv.

		M. paradoxus				*M. capensis*			
	Year	Bsurv	Btot	Bcrit	Bcrit/Bsurv	Bsurv	Btot	Bcrit	Bcrit/Bsurv
WC sum	1985	183.96	197.64	13.68	0.069	139.13	149.39	10.26	0.069
	1986	179.32	192.87	13.55	0.070	127.61	137.01	9.40	0.069
	1987	159.91	172.54	12.64	0.073	121.81	130.58	8.77	0.067
	1988	160.30	171.89	11.60	0.067	124.71	133.44	8.73	0.065
	1990	199.52	214.41	14.89	0.069	136.20	145.69	9.49	0.065
	1991	205.01	220.60	15.59	0.071	137.87	147.63	9.77	0.066
	1992	206.66	222.34	15.68	0.071	135.41	145.18	9.77	0.067
	1993	205.51	221.22	15.72	0.071	130.02	139.45	9.43	0.068
	1994	200.46	215.72	15.26	0.071	128.55	137.84	9.28	0.067
	1995	211.67	226.87	15.19	0.067	124.40	133.42	9.02	0.068
	1996	228.23	244.89	16.66	0.068	117.07	125.60	8.53	0.068
	1997	228.27	245.33	17.07	0.070	110.48	118.46	7.99	0.067
	1999	184.90	199.28	14.38	0.072	100.14	107.35	7.20	0.067
	2002	164.50	175.68	11.18	0.064	84.54	90.43	5.89	0.065
	2003	172.80	184.69	11.89	0.064	81.78	87.45	5.68	0.065
	2004	156.07	168.94	12.87	0.076	91.28	96.75	5.47	0.057
	2005	157.09	168.98	11.89	0.070	96.86	102.61	5.75	0.056
	2006	185.32	197.11	11.78	0.060	96.89	103.49	6.60	0.064
	2007	194.61	209.88	15.27	0.073	111.41	118.26	6.85	0.058
	2008	186.42	202.07	15.65	0.077	115.97	123.22	7.25	0.059
	2009	185.23	200.53	15.31	0.076	118.77	126.36	7.58	0.060
	2010	204.62	219.88	15.27	0.069	105.18	112.90	7.72	0.068
	2011	195.73	212.89	17.15	0.081	111.97	119.28	7.31	0.061
	2012	189.42	205.84	16.42	0.080	108.54	115.58	7.04	0.061
	2013	185.02	200.98	15.96	0.079	109.07	116.03	6.96	0.060
	2014	194.12	209.82	15.70	0.075	117.37	124.69	7.33	0.059
	2015	211.49	228.08	16.59	0.073	126.99	134.86	7.86	0.058
	2016	225.86	243.96	18.10	0.074	138.01	146.60	8.59	0.059
	2017	223.27	242.22	18.95	0.078	144.16	153.26	9.10	0.059
	AVE	**192.60**	**207.49**	**14.89**	**0.072**	**116.97**	**124.92**	**7.95**	**0.064**
WC win	1985	209.34	215.67	6.33	0.029	116.37	126.03	9.67	0.077
	1986	196.57	202.75	6.18	0.030	105.05	113.89	8.84	0.078
	1987	175.96	181.56	5.60	0.031	100.09	108.22	8.13	0.075
	1988	187.85	193.18	5.33	0.028	105.31	113.32	8.01	0.071
	1989	214.70	220.76	6.05	0.027	113.39	121.87	8.47	0.070
	1990	230.04	237.00	6.96	0.029	120.35	129.49	9.14	0.071
	AVE	**202.41**	**208.49**	**6.08**	**0.029**	**110.10**	**118.80**	**8.71**	**0.073**
SC spring	1986	37.81	46.67	8.86	0.190	191.00	199.10	8.10	0.041
	1987	30.88	38.29	7.41	0.193	178.60	186.28	7.67	0.041
	2001	32.84	39.51	6.68	0.169	115.63	120.23	4.60	0.038
	2003	44.81	52.64	7.82	0.149	92.78	96.31	3.52	0.037
	2004	37.08	43.76	6.68	0.153	93.48	96.72	3.24	0.033
	2006	43.56	52.81	9.25	0.175	116.99	120.35	3.36	0.028
	2007	52.57	62.22	9.66	0.155	119.77	123.42	3.65	0.030
	2008	47.80	56.77	8.97	0.158	133.73	137.88	4.15	0.030
	AVE	**40.92**	**49.08**	**8.16**	**0.166**	**130.25**	**135.04**	**4.79**	**0.035**

Table 1 continued

		M. paradoxus				*M. capensis*			
	Year	Bsurv	Btot	Bcrit	Bcrit/Bsurv	Bsurv	Btot	Bcrit	Bcrit/Bsurv
SC autumn	1988	44.51	52.98	8.47	0.160	151.58	158.91	7.33	0.046
	1991	63.29	75.56	12.26	0.162	156.03	161.66	5.63	0.035
	1992	63.94	75.80	11.87	0.157	167.33	173.39	6.05	0.035
	1993	64.33	75.94	11.61	0.153	174.07	181.07	7.00	0.039
	1994	63.03	74.24	11.21	0.151	174.10	182.04	7.94	0.044
	1995	63.63	74.77	11.13	0.149	165.50	173.69	8.19	0.047
	1996	72.41	85.43	13.03	0.152	154.05	161.77	7.72	0.048
	1997	74.90	88.49	13.59	0.154	143.91	150.79	6.88	0.046
	1999	61.94	72.36	10.42	0.144	130.31	136.46	6.14	0.045
	2003	53.66	63.12	9.47	0.150	93.11	96.88	3.77	0.039
	2004	57.24	65.83	8.59	0.131	79.91	83.42	3.51	0.042
	2005	49.84	57.11	7.27	0.127	81.93	85.51	3.58	0.042
	2006	50.51	59.57	9.06	0.152	99.18	102.67	3.49	0.034
	2007	66.81	77.26	10.44	0.135	98.86	102.70	3.84	0.037
	2008	68.93	79.62	10.69	0.134	111.00	115.35	4.35	0.038
	2009	66.97	77.09	10.11	0.131	120.77	126.11	5.34	0.042
	2010	64.24	75.79	11.55	0.152	139.78	145.45	5.66	0.039
	2011	76.50	87.83	11.34	0.129	121.56	128.16	6.61	0.052
	2014	67.30	76.84	9.55	0.124	109.07	115.23	6.16	0.053
	2015	72.23	82.70	10.47	0.127	113.52	119.29	5.77	0.048
	2016	80.50	92.25	11.75	0.127	123.97	129.72	5.75	0.044
	AVE	**64.13**	**74.79**	**10.66**	**0.143**	**129.03**	**134.77**	**5.75**	**0.043**

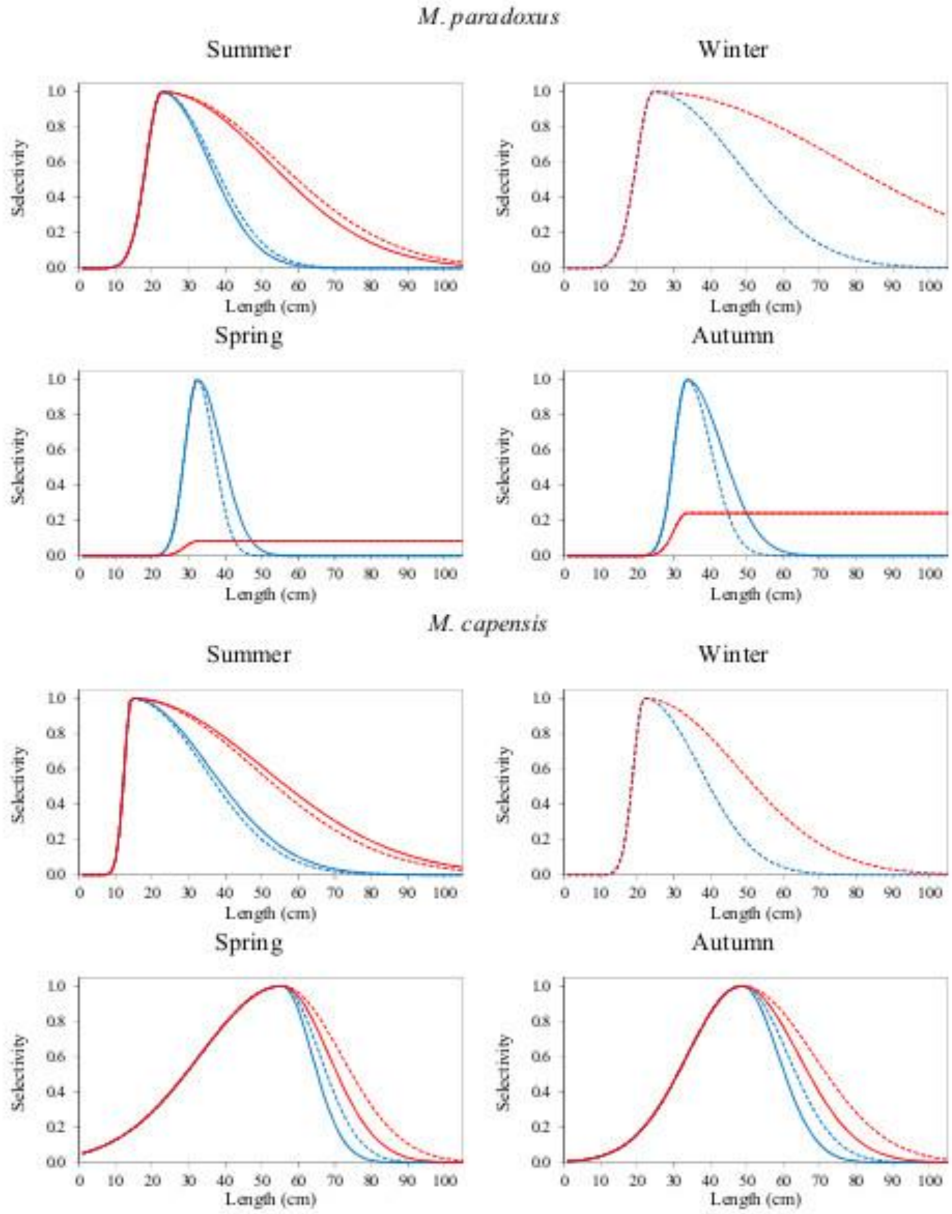

Figure 1: Survey selectivities-at-length for the Reference Case (blue curves for males, red curves for females, dashed curves for old gear and full curves for new gear).