Progress in Checking the Code used to Condition the Operating Models for the South African hake resource to be used in OMP-2010 testing

Carryn L de Moor and Rebecca A Rademeyer

April 2010

The code used to condition these Operating Models, the results for which have been reported in documents such as MCM/2010/FEB/SWG-DEM/05, has been subjected to careful independent checking by the first author.

This checking has focused on aspects leading to the computation of the likelihood upon which this conditioning is based. It has not (as yet) extended to the specification of outputs and all aspects of the computation of projections which are also part of the software in question. This checking is now about 90% complete.

This checking process has not yet revealed anything that would result in the computation of the likelihood (and hence the consequent results) by this software differing from what was intended.

However it has revealed a few omissions and typos in the documentation of the Operating Models, as set out in Appendix II of MCM/2010/FEB/SWG-DEM/05. The consequent corrections are shown in track changes in an updated version of that Appendix which is attached hereto.

APPENDIX II – Gender-disaggregated, Age-Structured Production Model fitting to Age-Length Keys

The model used is a gender-disaggregated Age-Structured Production Model (ASPM), which is fitted directly to age-length keys (ALKs) and length frequencies. The model also involves assessing the two species as two independent stocks and is fitted to species-disaggregated data as well as species-combined data. The general specifications and equations of the overall model are set out below together with some key choices in the implementation of the methodology. Details of the contributions to the log-likelihood function from the different data considered are also given. Quasi-Newton minimisation is used to minimise the total negative log-likelihood function (implemented using AD Model Builder^{TM}, Otter Research, Ltd.).

Population Dynamics

Numbers-at-age

The resource dynamics of the two populations (*M. capensis* and *M. paradoxus*) of the South African hake are modelled by the following set of equations:

Note: for ease of reading, the 'species' subscript *s* has been omitted below where not relevant.

$$N^g_{y+1,0} = R^g_{y+1} \qquad\qquad\qquad\qquad \text{(App.II.1)}$$

$$N^g_{y+1,a+1} = \left(N^g_{ya} e^{-M^g_a/2} - \sum_f C^g_{fya} \right) e^{-M^g_a/2} \qquad\qquad \text{for } 0 \le a \le m-2 \qquad \text{(App.II.2)}$$

$$N^g_{y+1,m} = \left(N^g_{y,m-1} e^{-M^g_{m-1}/2} - \sum_f C^g_{f,y,m-1} \right) e^{-M^g_{m-1}/2} + \left(N^g_{ym} e^{-M^g_m/2} - \sum_f C^g_{fym} \right) e^{-M^g_m/2} \text{ (App.II.3)}$$

where

N^g_{ya} is the number of fish of gender *g* and age *a* at the start of year *y*[1],

R^g_y is the recruitment (number of 0-year-old fish) of fish of gender *g* at the start of year *y*,

m   is the maximum age considered (taken to be a plus-group),

M^g_a denotes the natural mortality rate on fish of gender *g* and age *a*, and

C^g_{fya} is the number of hake of gender *g* and age *a* caught in year *y* by fleet *f*.

Recruitment

The number of recruits (i.e. new zero-year old fish) at the start of year *y* is assumed to be related to the corresponding **female** spawning stock size (i.e., the biomass of mature female fish) by means of the Beverton-Holt (Beverton and Holt, 1957) or a modified (generalised) form of the Ricker stock-recruitment relationship, parameterized in terms of the "steepness" of the stock-recruitment relationship, h, and the pre-exploitation equilibrium female spawning biomass, $K^{\female,sp}$, and pre-exploitation recruitment, R_0 and assuming a 50:50 sex-split at recruitment.

---

[1] In the interests of less cumbersome notation, subscripts have been separated by commas only when this is necessary for clarity.

$$R_y^g = \frac{4hR_0 B_y^{\female,sp}}{K^{\female,sp}(1-h)+(5h-1)B_y^{\female,sp}} e^{(\varsigma_y - \sigma_R^2/2)} \qquad \text{(App.II.4a)}$$

for the Beverton-Holt stock-recruitment relationship and

$$R_y^g = \alpha B_y^{\female,sp} \exp\left(-\beta\left(B_y^{\female,sp}\right)^{\gamma_R}\right) e^{(\varsigma_y - \sigma_R^2/2)} \qquad \text{(App.II.4b)}$$

with

$$\alpha = \frac{R_0 \exp\left(\beta\left(K^{\female,sp}\right)^{\gamma}\right)}{K^{\female,sp}} \qquad \text{and} \qquad \beta = \frac{\ln(5h)}{\left(K^{\female,sp}\right)^{\gamma_R}\left(1-5^{-\gamma_R}\right)}$$

for the modified Ricker relationship (for the true Ricker, $\gamma_R = 1$) where

ς_y reflects fluctuation about the expected recruitment in year y;

σ_R is the standard deviation of the log-residuals, which is input;

γ_R is a parameter of the modified Ricker relationship, which is estimated in the model fitting procedure;

$B_y^{\female,sp}$ is the female spawning biomass at the start of year y, computed as:

$$B_y^{\female,sp} = \sum_{a=1}^{m} f_a^{\female} w_a^{\female} N_{ya}^{\female} \qquad \text{(App.II.5)}$$

where

w_a^g is the begin-year mass of fish of gender g and age a;

f_a^g is the proportion of fish of gender g and age a that are mature; and

$$.R_0 = K^{\female,sp} / \left[\sum_{a=1}^{m-1} f_a^{\female} w_a^{\female} e^{-\sum_{a'=0}^{a-1} M_{a'}^g} + f_m^{\female} w_m^{\female} \frac{e^{-\sum_{a'=0}^{m-1} M_{a'}^g}}{1-e^{-M_m^g}} \right] \qquad \text{(App.II.6)}$$

For the Beverton-Holt form, h is bounded above by 0.98 to preclude high recruitment at extremely low spawning biomass, whereas for the modified Ricker form, h is bounded above by 1.5 to preclude extreme compensatory behaviour.

<u>Total catch and catches-at-age</u>

The fleet-disaggregated catch by mass, in year y is given by:

$$C_{fy} = \sum_g \sum_{a=0}^{m} w_{a+1/2}^g \, C_{fya}^g = \sum_g \sum_{a=0}^{m} w_{a+1/2}^g \, N_{ya}^g e^{-M_a^g/2} F_{fy} \tilde{S}_{fya}^g \qquad \text{(App.II.7)}$$

where

C_{fya}^g is the catch-at-age, i.e. the number of fish of gender g and age a, caught in year y by fleet f;

F_{fy} is the fishing mortality of a fully selected age class, for fleet f in year y (independent of g) ;

$$\tilde{S}_{fya}^g = \tilde{w}_{fy,a+1/2}^g \Big/ w_{a+1/2}^g \qquad \text{(App.II.8)}$$

$\tilde{S}_{fya}^g$ is the effective commercial selectivity of gender g at age a for fleet f and year y; with

$$\widetilde{w}^g_{fy,a+1/2} = \sum_l S^g_{fyl} w^g_l P^g_{a+1/2,l}$$ (App.II.9)

$\widetilde{w}^g_{fy,a+1/2}$ is the selectivity-weighted mid-year weight-at-age a of gender g for fleet f and year y;

w^g_l is the weight of fish of gender g and length l;

$w^g_{a+1/2}$ is the mid-year weight of fish of gender g and age a, at median length for that age;

S^g_{fyl} is the commercial selectivity of gender g at length l for year y, and fleet f;

$P^g_{a+1/2,l}$ is the mid-year proportion of fish of age a and gender g that fall in the length group l (i.e.,
$$\sum_l P^g_{a+1/2,l} = 1 \text{ for all ages } a).$$

The matrix P is calculated under the assumption that length-at-age is log-normally distributed about a mean given by the von Bertalanffy equation, i.e.:

$$l_a \sim N\left[\ln(l_\infty (1 - e^{-\kappa(a-t_0)})) ; \left(\frac{\theta_a}{l_\infty (1 - e^{-\kappa(a-t_0)})} \right)^2 \right]$$ (App.II.10)

where θ_a is the standard deviation of length-at-age a, which is estimated directly in the model fitting for age 0, and for ages 1 and above a linear relationship applies, with species and gender-specific α and β estimated in the model fitting procedure. A penalty is added so that θ_a is increasing with age.

Exploitable and survey biomasses

The model estimate of the mid-year exploitable ("available") component of biomass for each species and fleet is calculated by converting the numbers-at-age into mid-year mass-at-age and applying natural and fishing mortality for half the year:

$$B^{ex}_{fy} = \sum_g \sum_{a=0}^m w^g_{a+1/2} \widetilde{S}^g_{fya} N^g_{ya} e^{-M^g_a/2} \left(1 - \sum_f S^g_{fya} F_{fy} / 2 \right)$$ (App.II.11)

The model estimate of the survey biomass at the start of the year (summer) is given by:

$$B^{surv}_y = \sum_g \sum_{a=0}^{m_s} w^g_a \widetilde{S}^{g,sum}_a N^g_{ya}$$ (App.II.12)

and in mid-year (winter):

$$B^{surv}_y = \sum_g \sum_{a=0}^m w^g_{a+1/2} \widetilde{S}^{g,win}_a N^g_{ya} e^{-M^g_a/2} \left(1 - \sum_f S^g_{fya} F_{fy} / 2 \right)$$ (App.II.13)

where

$\widetilde{S}^{g,sum/win}_a$ is the effective survey selectivity of gender g for age a, converted from survey selectivity-at-length and selectivity-weighted weight-at-age in the same manner as for the commercial selectivity (eqns App.II.8 and App.II.9), taking account of the being-year ($\widetilde{w}^{g,sum}_{y,a}$ from $P^g_{a,l}$) or mid-year ($\widetilde{w}^{g,win}_{y,a+1/2}$ from $P^g_{a+1/2,l}$) nature of the surveys

Note that both the spring and autumn surveys are taken to correspond to winter (mid-year).

It is assumed that the resource is at the deterministic equilibrium that corresponds to an absence of harvesting at the start of the initial year considered, i.e., $B_1^{g,sp} = K^{g,sp}$, and year y=1 corresponds to 1917 when catches are taken to commence.

<u>MSY and related quantities</u>

The equilibrium catch for a fully selected fishing proportion F^* is calculated as:

$$C\left(F^*\right) = \sum_g \sum_a w_{a+1/2}^g \tilde{S}_a^g F^* N_a^g\left(F^*\right) e^{-\left(\left(M_a^g + S_a^g F^*\right)/2\right)} \tag{App.II.14}$$

where

S_a^g and $\tilde{S}_a^g$ are average selectivities and effective selectivities across all fleets, for the most recent five years;

$$S_a^g = \frac{\displaystyle\sum_{y=2005}^{2009} \sum_f S_{fya}^g F_{fy}}{\max\left(\displaystyle\sum_{y=2005}^{2009} \sum_f S_{fya}^g F_{fy}\right)} \tag{App.II.15}$$

$$\tilde{S}_a^g = \frac{\displaystyle\sum_{y=2005}^{2009} \sum_f \tilde{S}_{fya}^g F_{fy}}{\max\left(\displaystyle\sum_{y=2005}^{2009} \sum_f \tilde{S}_{fya}^g F_{fy}\right)} \tag{App.II.16}$$

where the maximum is taken over genders and ages; and with

$$N_a^g\left(F^*\right) = \begin{cases} R_1\left(F^*\right) & \text{for } a = 1 \\ N_{a-1}^g\left(F^*\right) e^{-M_{a-1}^g}\left(1 - S_{a-1}^g F^*\right) & \text{for } 1 < a < m \\ \dfrac{N_{m-1}^g\left(F^*\right) e^{-M_{m-1}^g}\left(1 - S_{m-1}^g F^*\right)}{\left(1 - e^{-M_m^g}\left(1 - S_m^g F^*\right)\right)} & \text{for } a = m \end{cases} \tag{App.II.17}$$

where

$$R_1\left(F^*\right) = \frac{\alpha B^{\female,sp}\left(F^*\right)}{\beta + B^{\female,sp}\left(F^*\right)} \tag{App.II.18}$$

for a Beverton-Holt stock–recruitment relationship.

The maximum of $C\left(F^*\right)$ is then found by searching over F^* to give F_{MSY}^*, with the associated female spawning biomass given by

$$B_{MSY}^{\female,sp} = \sum_a f_a^\female w_a^\female N_a^\female\left(F_{\text{MSY}}^*\right)$$

The likelihood function

The model is fit to CPUE and survey abundance indices, commercial and survey length frequencies, survey age-length keys, as well as to the stock-recruitment curve to estimate model parameters. Contributions by each of these to the negative of the log-likelihood (-ℓnL) are as follows[2].

CPUE relative abundance data

The likelihood is calculated by assuming that the observed abundance index (here CPUE) is log-normally distributed about its expected value:

$$I_y^i = \hat{I}_y^i e^{\varepsilon_y^i} \quad \text{or} \quad \varepsilon_y^i = \ell n\left(I_y^i\right) - \ell n\left(\hat{I}_y^i\right) \qquad \text{(App.II.19)}$$

where

I_y^i is the abundance index for year y and series i (which corresponds to a specified species and fleet)

$\hat{I}_y^i = \hat{q}^i \hat{B}_{fy}^{ex}$ is the corresponding model estimate, where $\hat{B}_{fy}^{ex}$ is the model estimate of exploitable resource biomass, given by equation App.II.11,

$\hat{q}^i$ is the constant of proportionality for abundance series i, and

ε_y^i from $N\left(0,\left(\sigma_y^i\right)^2\right)$.

In cases where the CPUE series are based upon species-aggregated catches (as available pre-1978), the corresponding model estimate is derived by assuming two types of fishing zones: z1) an "*M. capensis* only zone", corresponding to shallow water and z2) a "mixed zone" (Fig. App.II.1).

The total catch of hake of both species (*BS*) by fleet f in year y ($C_{BS,fy}$) can be written as:

$$C_{BS,fy} = C_{C,fy}^{z1} + C_{C,fy}^{z2} + C_{P,fy} \qquad \text{(App.II.20)}$$

where

$C_{C,fy}^{z1}$ is the *M. capensis* catch by fleet f in year y in the *M. capensis* only zone (z1),

$C_{C,fy}^{z2}$ is the *M. capensis* catch by fleet f in year y in the mixed zone (z2), and

$C_{P,fy}$ is the *M. paradoxus* catch by fleet f in year y in the mixed zone.

Catch rate is assumed to be proportional to exploitable biomass. Furthermore, let γ be the proportion of the *M. capensis* exploitable biomass in the mixed zone ($\gamma = B_{C,fy}^{ex,z2} \big/ B_{C,fy}^{ex}$) (assumed to be constant throughout the period for simplicity) and ψ_{fy} be the proportion of the effort of fleet f in the mixed zone in year y ($\psi_{fy} = E_{fy}^{z2} \big/ E_{fy}$), so that:

$$C_{C,fy}^{z1} = q_C^{i,z1} B_{Cfy}^{ex,z1} E_{fy}^{z1} = q_C^{i,z1}\left(1-\gamma\right)B_{C,fy}^{ex}\left(1-\psi_{fy}\right)E_{fy} \qquad \text{(App.II.21)}$$

$$C_{C,fy}^{z2} = q_C^{i,z2} B_{C,fy}^{ex,z2} E_{fy}^{z2} = q_C^{i,z2}\gamma B_{C,fy}^{ex}\psi_{fy} E_{fy} \quad \text{and} \qquad \text{(App.II.22)}$$

$$C_{P,fy} = q_P^i B_{P,fy}^{ex} E_{fy}^{z2} = q_P^i B_{P,fy}^{ex}\psi_{fy} E_{fy} \qquad \text{(App.II.23)}$$

where

---

[2] Strictly it is a penalised log-likelihood which is maximised in the fitting process, as some contributions that would correspond to priors in a Bayesian estimation process are added.

$E_{fy} = E_{fy}^{z1} + E_{fy}^{z2}$ is the total effort of fleet f, corresponding to combined-species CPUE series i which consists of the effort in the *M. capensis* only zone (E_{fy}^{z1}) and the effort in the mixed zone (E_{fy}^{z2}), and

$q_C^{i,zj}$ is the catchability for *M. capensis* (*C*) for abundance series i, and zone zj, and

q_P^i is the catchability for *M. paradoxus* (*P*) for abundance series i.

It follows that:

$$C_{C,fy} = B_{C,fy}^{ex} E_{fy} \left[q_C^{i,z1}(1-\gamma)(1-\psi_{fy}) + q_C^{i,z2}\gamma\psi_{fy} \right] \qquad \text{(App.II.24)}$$

$$C_{P,fy} = B_{P,fy}^{ex} E_{fy} q_P^i \psi_{fy} \qquad \text{(App.II.25)}$$

From solving equations App.II.24 and App.II.25:

$$\psi_{fy} = \frac{q_C^{i,z1}(1-\gamma)}{\left\{ \dfrac{C_{C,fy} B_{P,fy}^{ex} q_P^i}{B_{C,fy}^{ex} C_{P,fy}} - q_C^{i,z2}\gamma + q_C^{i,z1}(1-\gamma) \right\}} \qquad \text{(App.II.26)}$$

and:

$$\hat{I}_y^i = \frac{C_{fy}}{E_{fy}} = \frac{C_{fy} B_{P,fy}^{ex} q_P^i \psi_{fy}}{C_{P,fy}} \qquad \text{(App.II.27)}$$

Zone 1 (z1): *M. capensis* only	Zone 2 (z2): Mixed zone
M. capensis: biomass (B_C^{z1}), catch(C_C^{z1})	*M. capensis*: biomass (B_C^{z2}), catch(C_C^{z2}) *M. paradoxus*: biomass (B_P), catch(C_P)
Effort in zone 1 (E^{z1})	Effort in zone 2 (E^{z2})

Fig. App.II.1: Diagrammatic representation of the two theoretical fishing zones.

Two species-aggregated CPUE indices are available: the ICSEAF west coast and the ICSEAF south coast series. For consistency, q's for each species _s_ (and zone) are forced to be in the same proportion:

$$q_s^{SC} = r q_s^{WC} \qquad \text{(App.II.28)}$$

~~To correct for possible negative bias in estimates of variance $\left(\sigma_y^i\right)$ and to avoid according unrealistically high precision (and so giving inappropriately high weight) to the CPUE data, lower bounds on the standard deviations of the residuals for the logarithm of the CPUE series have been enforced; for the historic ICSEAF CPUE series (separate west coast and south coast series) the lower bound is set to 0.25, and to 0.15 for the recent GLM standardised CPUE series, i.e.: $\sigma^{ICSEAF} \geq 0.25$ and $\sigma^{GLM} \geq 0.15$.~~

The contribution of the CPUE data to the negative of the log-likelihood function (after removal of constants) is then given by:

$$-\ell\text{n }L^{CPUE} = \sum_i \sum_y \left[\ell\text{n}\left(\sigma_y^i\right) + \left(\varepsilon_y^i\right)^2 / 2\left(\sigma_y^i\right)^2 \right] \qquad\qquad \text{(App.II.29)}$$

where

σ_y^i is the standard deviation of the residuals for the logarithms of index i in year y.

Homoscedasticity of residuals for CPUE series is customarily assumed[3], so that $\sigma_y^i = \sigma^i$ is estimated in the fitting procedure by its maximum likelihood value:

$$\hat{\sigma}^i = \sqrt{1/n_i \sum_y \left(\ell\text{n}(I_y^i) - \ell\text{n}(\hat{I}_y^i)\right)^2} \qquad\qquad \text{(App.II.30)}$$

where n_i is the number of data points for abundance index i.

In the application, σ^i are taken as estimable parameters in the model fitting procedure. To correct for possible negative bias in estimates of variance $\left(\sigma^i\right)$ and to avoid according unrealistically high precision (and so giving inappropriately high weight) to the CPUE data, lower bounds on the standard deviations of the residuals for the logarithm of the CPUE series have been enforced; for the historic ICSEAF CPUE series (separate west coast and south coast series) the lower bound is set to 0.25, and to 0.15 for the recent GLM-standardised CPUE series, i.e.: $\sigma^{ICSEAF} \geq 0.25$ and $\sigma^{GLM} \geq 0.15$.

In the case of the species-disaggregated CPUE series, the catchability coefficient q^i for abundance index i is estimated by its maximum likelihood value, which in the more general case of heteroscedastic residuals, is given by:

$$\ln \hat{q}^i = \frac{\sum_y \left(\ln I_y^i - \ln \hat{B}_{fy}^{ex}\right)/\left(\sigma_y^i\right)^2}{\sum_y 1/\left(\sigma_y^i\right)^2} \qquad\qquad \text{(App.II.31)}$$

In the case of the species-combined CPUE, $q_C^{WC,z1}$, $q_C^{WC,z2}$, q_P^{WC}, r and γ are directly estimated in the fitting procedure.

Survey abundance data

Data from the research surveys are treated as relative abundance indices in a similar manner to the species-disaggregated CPUE series above, with survey selectivity function $S_a^{g,sum/win}$ replacing the commercial selectivity S_{fya}^g (see equations App.II.12 and App.II.13 above, which also take account of the begin- or mid-year nature of the survey).

An estimate of sampling variance is available for most surveys and the associated σ_y^i is generally taken to be given by the corresponding survey CV. However, these estimates likely fail to include all sources of variability, and unrealistically high precision (low variance and hence high weight) could hence be accorded to these indices. The contribution of the survey data to the negative log-likelihood is of the same form as that of the CPUE abundance data (see equation App.II.29). The procedure adopted takes into account an additional variance $\left(\sigma_A\right)^2$ which is treated as another estimable parameter in the minimisation process. This procedure is carried out enforcing the constraint that $\left(\sigma_A\right)^2 > 0$, i.e. the overall variance cannot be less than its externally input component.

---

[3] There are insufficient data in any series to enable this to be tested with meaningful power.

In June 2003, the trawl gear on the *Africana* was changed and a different value for the multiplicative bias factor q is taken to apply to the surveys conducted with the new gear. Calibration experiments have been conducted between the *Africana* with the old gear (hereafter referred to as the "old *Africana*") and the *Nansen*, and between the *Africana* with the new gear ("new *Africana*") and the *Nansen*, in order to provide a basis to relate the multiplicative biases of the *Africana* with the two types of gear (q_{old} and q_{new}). A GLM analysis assuming negative binomial distributions for the catches made (Brandão *et al*., 2004) provided the following estimates:

$$\Delta \ell nq^{capensis} = -0.494 \quad \text{with } \sigma_{\Delta \ell nq^{capensis}} = 0.141 \quad \text{i.e. } \left(q^{new}/q^{old}\right)^{capensis} = 0.610 \quad \text{and}$$

$$\Delta \ell nq^{paradoxus} = -0.053 \quad \text{with } \sigma_{\Delta \ell nq^{paradoxus}} = 0.117 \quad \text{i.e. } \left(q^{new}/q^{old}\right)^{paradoxus} = 0.948$$

where

$$\ell nq^s_{new} = \ell nq^s_{old} + \Delta \ell nq^s \quad \text{with } s = capensis \text{ or } paradoxus \qquad \text{(App.II.32)}$$

No plausible explanation has yet been found for the particularly large extent to which catch efficiency for *M. capensis* is estimated to have decreased for the new research survey trawl net. It was therefore recommended (BENEFIT, 2004) that the ratio of the catchability of the new to the previous *Africana* net be below 1, but not as low as the ratio estimated from the calibration experiments. $\Delta \ell nq^{capensis}$ is therefore taken as -0.223, i.e. $\left(q^{new}/q^{old}\right)^{capensis} = 0.8$ (with $\sigma_{\Delta \ell nq^{capensis}} = 0.141$).

The following contribution is therefore added as a penalty (or a prior in a Bayesian context) to the negative log-likelihood in the assessment:

$$-\ell nL^{q-ch} = \left(\ell nq_{new} - \ell nq_{old} - \Delta \ell nq\right)^2 \big/ 2\sigma^2_{\Delta \ell nq} \qquad \text{(App.II.33)}$$

A different length-specific selectivity is estimated for the "old *Africana*" and the "new *Africana*".

The survey's coefficients of catchability q (for the survey with the old *Africana* gear) are constrained below 1:

$$pen^q = \sum_i \left(q^i_{old} - 1\right)^2 \big/ 0.02^2 \quad \text{if} \quad q^i_{old} > 1 \qquad \text{(App.II.34)}$$

Commercial proportions at length

Commercial proportions at length cannot be disaggregated by species and gender. The model is therefore fit to the proportions at length as determined for both species and gender combined.

The catches at length are computed as:

$$C_{fyl} = \sum_s \sum_g \sum_{a=0}^m N^g_{sya} F_{sfy} S^g_{sfyl} P^g_{s,a+1/2,l} e^{-M^g_{sa}/2} \left(1 - \sum_f S^g_{sfya} F_{fy} \big/ 2\right) \qquad \text{(App.II.35)}$$

With the predicted proportions at length:

$$\widehat{p}^i_{yl} = C_{fyl} \big/ \sum_{l'} C_{fyl'} \qquad \text{(App.II.36)}$$

The contribution of the proportion at length data to the negative of the log-likelihood function when assuming an "adjusted" lognormal error distribution is given by:

$$-\ell n L^{length} = 0.1 \sum_y \sum_l \left[\ell n \left(\sigma^i_{len}/\sqrt{p^i_{yl}}\right) + p^i_{yl} \left(\ell np^i_{yl} - \ell n \, \hat{p}^i_{yl}\right)^2 / 2\left(\sigma^i_{len}\right)^2 \right] \qquad \text{(App.II.37)}$$

where

the superscript 'i' refers to a particular series of proportions at length data which reflect a specified fleet, and species (or combination thereof) and

σ_{len}^i is the standard deviation associated with the proportion at length data, which is estimated in the fitting procedure by:

$$\hat{\sigma}_{len}^i = \sqrt{\sum_y \sum_l p_{yl}^i \left(\ln p_{yl}^i - \ln \hat{p}_{yl}^i\right)^2 \bigg/ \sum_y \sum_l 1}$$ (App.II.38)

The initial 0.1 multiplicative factor is a somewhat arbitrary downweighting to allow for correlation between proportions in adjacent length groups. The coarse basis for this adjustment is the ratio of effective number of age-classes present to the number of length groups in the minimisation, under the argument that independence in variability is likely to be more closely related to the former.

Commercial proportions at length are incorporated in the likelihood function using equation App.II.37, for which the summation over length l is taken from length l_{minus} (considered as a minus group) to l_{plus} (a plus group). The length for the minus- and plus-groups are fleet specific and are chosen so that typically a few percent, but no more, of the fish sampled fall into these two groups.

Survey proportions at length

The survey proportions at length are incorporated into the negative of the log-likelihood in an analogous manner to the commercial catches-at-age, assuming an adjusted log-normal error distribution (equation App.II.~~36~~37). In this case however, data are disaggregated by species, and for some surveys further disaggregated by gender.

$p_{syl}^{g,surv} = \dfrac{C_{syl}^{g,surv}}{\sum_{l'} C_{syl'}^{g,surv}}$ is the observed proportion of fish of species s, gender g and length l from survey $surv$ in year y,

$\hat{p}_{syl}^{g,surv}$ is the expected proportion of fish of species s, gender g and length l in year y in the survey $surv$, given by:

$$\hat{p}_{syl}^{g,surv} = \frac{\sum_a S_{sl}^{g,sum} P_{sal}^g N_{sya}^g}{\sum_{l'} \sum_a S_{sl'}^{g,sum} P_{sal'}^g N_{sya}^g}$$ (App.II.39)

for begin-year (summer) surveys, or

$$\hat{p}_{syl}^{g,surv} = \frac{\sum_a S_{sl}^{g,win} P_{s,a+1/2,l}^g N_{sya}^g e^{-M_{sa}^g/2} \left(1 - \sum_f S_{sfya}^g F_{sfy} \bigg/ 2\right)}{\sum_{l'} \sum_a S_{sl'}^{g,win} P_{s,a+1/2,l'}^g N_{sya}^g e^{-M_{sa}^g/2} \left(1 - \sum_f S_{sfya}^g F_{sfy} \bigg/ 2\right)}$$ (App.II.40)

for mid-year (autumn, winter or spring) surveys.

Age-length keys

Under the assumption that fish are sampled randomly with respect to age within each length-class, the contribution to the negative log-likelihood for the ALK data (ignoring constants) is:

$$-\ln L^{ALK} = -w \sum_i \sum_l \sum_a \left[A_{i,l,a}^{obs} \ln\left(\hat{A}_{i,l,a}\right) - A_{i,l,a}^{obs} \ln\left(A_{i,l,a}^{obs}\right)\right]$$ (App.II.41)

where

w is a downweighting factor to allow for overdispersion in these data compared to the expectation for a multinomial distribution with independent data; for the moment this weight factor is set to 0.01,

$A_{i,a,l}^{obs}$ is the observed number of fish of age a that fall in the length class l, for ALK i (a specific combination of survey, reader, year, species and gender),

$\hat{A}_{i,a,l}$ is the model estimate of $A_{i,a,l}^{obs}$, computed as:

$$\hat{A}_{i,a,l} = W_{i,l} \frac{C_{i,a,l}\pi_{a,l}^{i}}{\sum\limits_{a'} C_{i,a,l}\pi_{a',l}^{i}} \qquad\qquad \text{(App.II.42)}$$

where

$W_{i,l}$ is the number of fish in length class l that were aged for ALK i,

$C_{i,a,l}$ is the predicted catch-at-length for age a and ALK i:

$$C_{i,a,l} = S_{sl}^{g,sum} N_{sya}^{g} \qquad \text{for begin-year surveys,}$$

$$C_{i,a,l} = S_{sl}^{g,sum} N_{sya}^{g} e^{-M_{sa}^{g}/2}\left(1 - \sum_{f} S_{sfya}^{g} F_{sfy}\Big/2\right) \qquad \text{for mid-year surveys, and}$$

$$C_{i,a,l} = N_{sya}^{g} S_{sfyl}^{g} e^{-M_{sa}^{g}/2}\left(1 - \sum_{f} S_{sfya}^{g} F_{fy}\Big/2\right) \qquad \text{for commercial ALKs.}$$

$$\pi_{a,l}^{i} = \sum_{a} Q^{r}\left(a'|a\right) P_{sal}^{g} \quad \text{is the ALK for age } a \text{ and length } l \text{ after accounting for age-reading error,}$$

with $Q^{r}\left(a'|a\right)$, the age-reading error matrix for reader r, representing the probability of an animal of true age a being aged to be that age or some other age a'. For commercial ALKs and ALKs from middle of the year surveys, $P_{s,a+1/2,l}^{g}$ replaces P_{sal}^{g} in the computation of $\pi_{a,l}^{i}$ above.

Age-reading error matrices have been computed for each reader and for each species in Rademeyer (2009).

When multiple readers age the same fish, these data are considered to be independent information in the model fitting.

Stock-recruitment function residuals

The stock-recruitment residuals are assumed to be log-normally distributed. Thus, the contribution of the recruitment residuals to the negative of the log-likelihood function is given by:

$$-\ell n L^{SR} = \sum_{s}\left[\sum_{y=y1}^{y2} \varsigma_{sy}{}^{2}\Big/2\sigma_{R}^{2} + \left(\sum_{y=y1}^{y2}\varsigma_{sy}\right)^{2}\Big/0.01^{2}\right] \qquad \text{(App.II.43)}$$

where

ς_{sy} is the recruitment residual for species s, and year y, which is assumed to be log-normally distributed with standard deviation σ_{R} and which is estimated for year $y1$ to $y2$ (see equation App.II.4) (estimating the stock-recruitment residuals is made possible by the availability of catch-at-age data, which give some indication of the age-structure of the population); and

12

σ_R is the standard deviation of the log-residuals, which is input.

The stock-recruitment residuals are estimated for years 1985 to 2006, with recruitment for other years being set deterministically (i.e. exactly as given by the estimated stock-recruitment curve) as there is insufficient catch-at-age information to allow reliable residual estimation for earlier years. A limit on the recent recruitment fluctuations is set by having the σ_R (which measures the extent of variability in recruitment – see equation – App.II.43) decreasing linearly from 0.45 in 2004 to 0.1 in 2009, effectively forcing recruitment over the last years to lie closer to the stock-recruitment relationship curve.

The second term on the right hand side is introduced to force the average of the residuals estimated over the period from *y1* to *y2* to be close to zero, for reasons elaborated in the main text.

Model parameters

Estimable parameters

The primary parameters estimated are the species-specific female virgin spawning biomass $\left(K_s^{\female\, sp} \right)$ and "steepness" of the stock-recruitment relationship (h_s). The standard deviations σ^i for the CPUE series residuals (the species-combined as well as the GLM-standardised series) as well as the additional variance $\left(\sigma_A^i \right)^2$ for each survey abundance series are treated as estimable parameters in the minimisation process. Similarly, in the case of the species-combined CPUE, $q_C^{WC,z1}$, $q_C^{WC,z2}$, q_P^{WC} , r and γ are directly estimated in the fitting procedure.

The species- and gender-specific von Bertalanffy growth curve parameters (L_5, κ and t_0) are estimated directly in the model fitting process, as well as θ_0 , θ_1 and θ_{14}, values used to compute the standard deviation of the length-at-age *a*.

The following parameters are also estimated in the model fits undertaken (if not specifically indicated as fixed).

Natural mortality:

Natural mortality (M_{sa}^g) is assumed to be age-specific and is calculated using the following functional form:

$$M_{sa}^{\female} = \begin{cases} M_{s2}^{\female} & \text{for} \quad a \leq 1 \\ \alpha_s^M + \dfrac{\beta_s^M}{a+1} & \text{for} \quad 2 \leq a \leq 5 \\ M_{s5}^{\female} & \text{for} \quad a > 5 \end{cases} \qquad (\text{App.II.44})$$

and

$$M_{sa}^{males} = \upsilon^s M_{sa}^{females} \qquad\qquad (\text{App.II.45})$$

M_{s0} and M_{s1} are set equal to M_{s2} $(= \alpha_s^M + \beta_s^M / 3)$ as there are no data (hake of ages younger than 2 are rare in catch and survey data) which would allow independent estimation of M_{s0} and M_{s1}.

When *M* values are estimated in the fit, a penalty is added to the total –lnL so that $M_{s2} \geq M_{s5}$:

$$pen^M = \sum_s \left(M_{s5} - M_{s2} \right)^2 / 0.01^2 \qquad\qquad \text{if } M_{s2} < M_{s5} \qquad (\text{App.II.46})$$

Stock-recruitment residuals:

Stock-recruitment residuals ς_{sy} are estimable parameters in the model fitting process. They are estimated separately for each species from 1985 to the present, and set to zero pre-1985 because there are no catch-at-length data for that period to provide the information necessary to inform estimation.

Table App.II.1 summarises the estimable parameters, excluding the selectivity parameters, and gives the bounds enforced for each estimable parameter.

Survey fishing selectivity-at-length:

The survey selectivities are estimated directly for seven pre-determined lengths for *M. paradoxus* and *M. capensis*. When the model was fitted to proportion-at-age rather than proportion-at-length, survey selectivities were estimated directly for each age (i.e. seven age classes). The lengths at which selectivity is estimated directly are survey specific (linear between the minus and plus groups) and are given in Table App.II.2. Between these lengths, selectivity is assumed to change linearly. The slope from lengths l_{minus} to $l_{minus}+1$ is assumed to continue exponentially to lower lengths to length *1*, and similarly the slope from lengths $l_{plus}-1$ to l_{plus} for *M. paradoxus* and *M. capensis* to continue for greater lengths. If the resulting slopes are positive, they are then set to zero (i.e. the selectivity cannot be increasing below the minus group and above the plus group).

For the south coast spring and autumn surveys, gender-specific selectivities are estimated for *M. paradoxus*. Furthermore, the female selectivities are scaled down by a parameter estimated for each of these two surveys to allow for the male predominance in the survey catch.

A penalty is added to the total –lnL to smooth the selectivities:

$$pen^{survS} = \sum_{i} \sum_{L=L_1+1}^{L_7-1} 3\left(S_{L-1}^i - 2S_L^i + S_{L+1}^i\right)^2 \qquad \text{(App.II.47)}$$

where *i* is a combination of survey, species and gender.

Commercial fishing selectivity-at-length:

The fishing selectivity-at-length (gender independent) for each species and fleet, S_{sfl}, is estimated in terms of a logistic curve given by:

$$S_{sfl} = \left[1 + \exp\left(-\left(l - l_{sf}^c\right)/\delta_{sf}^c\right)\right]^{-1} \qquad \text{(App.II.45)}$$

where

l_{sf}^c cm is the length-at-50% selectivity,

δ_{sf}^c cm^{-1} defines the steepness of the ascending limb of the selectivity curve.

The selectivity is sometimes modified to include a decrease in selectivity at larger lengths, as follows:

$$S_{sfl} = S_{s,f,l-1}e^{-s_{sfl}} \qquad \text{for } l > l_{slope}, \qquad \text{(App.II.46)}$$

where

s_{sfl} measures the rate of decrease in selectivity with length for fish longer than l_{slope} for the fleet concerned, and is referred to as the "selectivity slope".

l_{slope} is fixed externally from the model, values for each fleet and species are given in Table App.II.3.

Periods of fixed and changing selectivity have been assumed for the offshore trawl fleet to take account of the change in the selectivity at low ages over time in the commercial catches, likely due to the phasing out of the (illegal) use of net liners to enhance catch rates.

On the south coast, for *M. paradoxus*, the female offshore trawl selectivity (only the trawl fleet is assumed to catch *M. paradoxus* on the south coast) is scaled down by a factor taken as the average of those estimated for the south coast spring and autumn surveys. Although there is no gender information for the commercial catches, the south coast spring and autumn surveys catch a much higher proportion of male *M. paradoxus* than female (ratios of about 7:1 and 3.5:1 for spring and autumn respectively). This is assumed to reflect a difference in distribution of the two genders which would therefore affect the commercial fleet similarly.

Details of the fishing selectivities (including the number of parameters estimated) used in the assessment are shown in Table App.II.4.

Input parameters and other choice for application to hake

Age at mMaturity-at-length and at-age:

The proportion of fish of species *s*, gender *g* and length *l* that are mature is assumed to follow a logistic curve with the parameter values given below (from Fairweather and Leslie, 2008, "stage 2, >40cm" for females and Fairweather, pers. commn for males):

	l_{50} (cm)	N
M. paradoxus:		
Males	28.63	5.07
Females	42.24	4.46
M. capensis:		
Males	34.35	7.38
Females	40.80	7.51

Maturity-at-length is then converted to maturity-at-age as follows:

$$f_{sa}^{g} = \sum_{l} f_{sl}^{g} P_{a,l}^{g} \qquad \text{(App.II.47)}$$

Weight-at-length and at-age:

The weight-at-length for each species and gender is calculated from the mass-at-length function, with values of the parameters for this function listed below (from Fairweather, 2008, taking the average of the west and south coasts):

	$\perp$ (gm/cm$^{\perp}$)	$\perp$
M. paradoxus:		
Males	0.007541	2.988
Females	0.005836	3.065
M. capensis:		
Males	0.006307	3.061
Females	0.005786	3.085

Weight-at-length is then converted to weight-at-age as follows:

$$w_{a}^{g} = \sum_{l} w_{l}^{g} P_{a,l}^{g} \qquad \text{(App.II.48)}$$

for begin-year weight-at-age, and

$$w_{a+1/2}^{g} = \sum_{l} w_{l}^{g} P_{a+1/2,l}^{g} \qquad \text{(App.II.49)}$$

for mid-year weight-at-age.

Minus- and plus-groups

Because of a combination of gear selectivity and mortality, a relatively small number of fish in the smallest and largest length classes are caught. In consequence, there can be relatively larger errors (in terms of variance) associated with these data. To reduce this effect, the assessment is conducted with minus- and plus-groups obtained by summing the data over the lengths below and above l_{minus} and l_{plus} respectively. The minus- and plus-group used are given in Table App.II.5 (and plotted in Figs.I.2 and 3). Furthermore, the proportions at length data (both commercial and survey) are summed into 2cm length classes for the model fitting.

Table App.II.1: Parameters estimated in the model fitting procedure, excluding selectivity parameters, with bounds enforced.

	No of parameters	Parameters estimated	Bounds enforced
$K^{\female}$	2	$\ln(K^{\female}_{cap})$ and $\ln(K^{\female}_{para})$	(3.5; 9.0)
h	2	h_{cap} and h_{para}	(0.2; 0.98)
M_a	4 (6)*	For each species: M_2, M_5 (and υ)	M_2: (0.2; 1.0), M_5: (0.1; 0.5), υ: (0.01; 5)
Additional variance	2	$\sigma_{A,cap}$ and $\sigma_{A,para}$	(0; 0.5)
Recruitment residuals	50	$\zeta_{cap,1985-2009}$ and $\zeta_{para,1985-2009}$	(−5; 5)
σ_{CPUE}	6	1 for each series	ICSEAF: (0.25; 1), GLM (0.15; 1)
ICSEAF CPUE	5	$q_C^{WC,z1}$, $q_C^{WC,z2}$, q_P^{WC}, r and γ_R	q and r : (0,10), and γ_R (0; 1)
θ_a	12	For each species and gender: θ_0, θ_1 and θ_{14}	θ_0: (1; 100), θ_1 and θ_{14} : (1; 100)
Growth	12	For each species and gender: L_5, κ and t_0	L_5: (30; 60) κ: (0.00005; 0.2) and t_0: (-10; 0)

* if not fixed on input

Table App.II.2: Lengths (in cm) at which survey selectivity is estimated directly.

M. paradoxus	West coast summer	13	18	23	28	32	37	42	47
	West coast winter	13	18	24	29	35	40	46	51
	South coast spring	21	26	30	35	39	44	48	53
	South coast autumn	21	26	31	36	42	47	52	65
M. capensis	West coast summer	13	20	26	33	39	46	52	59
	West coast winter	13	17	21	30	40	47	54	61
	South coast spring	13	19	28	38	46	54	63	71
	South coast autumn	13	19	28	36	44	52	61	69

Table App.II.3: Length (cm) at which selectivity starts to decrease (l_{slope}) for each species and fleet.

	M. paradoxus	*M. capensis*
WC offshore trawl	40	70
SC offshore trawl	70	70
SC inshore trawl	-	55
WC longline	85	85
SC longline	-	85
SC handline	-	70

Table App.II.4: Details for the commercial selectivity-at-length for each fleet and species combination, as well as indications of what data are available.

	M. paradoxus		*M. capensis*		data available
	No of est. parameters	Comments	No of est. parameters	Comments	
1. West coast offshore					
1917-1976	0	set equal to 1989	0	set equal to 1989	
1977-1984	2	two logistic parameters estimated (same slope as 1993+)	0	differential shift compared to 1993+ as for paradoxus, slope 1/3 of inshore	species combined
1985-1992	0	linear change between 1984 and 1993 selectivity	0	linear change between 1984 and 1993 selectivity	species combined
1993-2009	3	two logistic + slope parameters estimated	0	same as SC inshore but shifted to the right by 5 cm, slope 1/3 of inshore	species combined
2. South coast offshore					
1917-1976	0	set equal to 1989	0	set equal to 1989	
1977-1984	0	differential shift compared to 1993+ as for WC (same slope as 1993+)	0	differential shift compared to 1993+ as for paradoxus, slope 1/3 of inshore	species combined
1985-1992	0	linear change between 1984 and 1993 selectivity	0	linear change between 1984 and 1993 selectivity	species combined
1993-2009	3	two logistic + slope parameters estimated	0	same as SC inshore but shifted to the right by 5 cm, slope 1/3 of inshore	species combined
	0	female downscaling factor (av. of SC spring and autumn surveys's factors)			
3. South coast inshore	-	-	3	two logistic + slope parameters estimated	*M. capensis*
4. West coast longline	3	two logistic + slope parameters estimated	0	same as South Coast longline	species combined
5. South coast longline	-	-	3	two logistic + slope parameters estimated	*M. capensis*
6. South coast handline	-	-	0	parameters taken as average of SC longline and inshore parameters	
West coast summer survey					
Africana old	7	estimated for 7 specified lengths	7	estimated for 7 specified lengths	species disaggregated
Africana new	5	same slope as old	5	same slope as old	species disaggregated
West coast winter survey					
Africana old	7	estimated for 7 specified lengths	7	estimated for 7 specified lengths	species disaggregated
South coast spring survey					
Africana old	7	estimated for 7 specified lengths	7	estimated for 7 specified lengths	species disaggregated
Africana new	5	same slope as old	5	same slope as old	species disaggregated
	1	female downscaling factor			
South coast autumn survey					
Africana old	7	estimated for 7 specified lengths	7	estimated for 7 specified lengths	species disaggregated
Africana new	5	same slope as old	5	same slope as old	species disaggregated
	1	female downscaling factor			
Total	56		49		

Table App.II.5: Minus- and plus-groups taken for the surveys and commercial proportion at length data.

SURVEY DATA

	M. paradoxus		*M. capensis*	
	Minus	Plus	Minus	Plus
West coast summer	13	47	13	59
West coast winter	13	51	13	61
South coast spring	21	53	13	71
South coast autumn	21	65	13	69

COMMERCIAL DATA

	Minus	Plus
West coast offshore, species combined	23	65
South coast offshore, species combined	27	75
South coast inshore, *M. capensis*	27	65
West coast longline, species combined	51	91
South coast longline, *M. capensis*	51	91
Both coasts offshore, species combined	25	65