Exploratory analyses of the Agulhas sole assessment

JP Glazer and DS Butterworth

Summary

A number of different production model approaches are considered for assessment of the Agulhas sole resource, Stationary models, either with observation error only estimation, or adding annual process error, exhibit systematic patterns in residuals. The best approach seems to be to postulate a non-stationary situation, with some change in the dynamics of the resource (or equivalently in the fishery catchability *q*) in the new century. Some initial comments are made as regards the next steps needed to advance this approach further.

Introduction

A sole Task Team was convened in 2017 to address improvements that could be made to the assessments of the Agulhas sole resource during 2018. Currently the baseline assessment comprises an observation error Schaefer model with catches and commercial standardized CPUE over the period 2000-2016 as input. Methods were developed in order to fit the sharp decline in CPUE over the period 2009-2013. These consisted of two equally plausible hypotheses: i) that the decline in CPUE was a result of a decrease in the productivity of the resource and ii) that the decline in CPUE was a result of a decrease in catchability of the resource (Glazer and Fairweather, 2017). Results from the models for each of these hypotheses contribute to advising on the annual TAC for sole.

The first meeting of the sole Task Team took place in February 2018 and one of the recommendations made was to investigate the possibility of extending the CPUE series back to the early-mid 1980s. This arose from analyses conducted by Yemane (2017) which suggested slight over-exploitation of the resource over a long period during the 1980's and 1990's and as a result current depletion was estimated to 10% of its pristine (pre-exploitation) level, which would likely result in recommendations to close the fishery. Yemane (2017) included both observation and process error in his model, the latter of which was applied to total biomass and his method assumes process errors to be both random and independent.

Updated inputs to the exploratory analyses

The following data were included in the updated analyses:

- sole catches (1920-2016),
- nominal CPUE index (1986[1]-2016)
- autumn survey index (utilizing "old" gear),
- autumn survey index (utilizing "new" gear),
- spring survey index (utilizing "old" gear), and

- [1] Fairweather and Glazer (2018) report nominal CPUE for the period 1985-2016, but the 1985 catches only accounted for 36% of the landings for that year (Fairweather, *pers commn*); hence the exclusion of 1985 from the analyses.

- spring survey index (utilizing “new” gear).

The catch data are reported in Table 1 and the indices of abundance are reported in Table 2.

The assessment model including observation error only

The dynamic Schaefer model is of the form:

$$B_{y+1} = B_y + rB_y[1 - \frac{B_y}{K}] - C_y \quad (1)$$

where:

B_y is the biomass estimated in year *y*,
r is the intrinsic rate of population growth,
K is pristine biomass (which is assumed to reflect the biomass at the start of the catch time series in 1920), and
C_y is the annual catch over the period 1920-2016.

The likelihood is calculated assuming that the abundance indices are log-normally distributed about their expected values:

$$I_y^i = q_i B_y e^{\varepsilon_y^i} \quad (2)$$

where I_y^i is the abundance index for index *i* and year *y*, $q_i B_y$ is the corresponding model estimate (q_i being the estimated catchability coefficients for each index of abundance), and ε_y^i the observation error for each index, $\sim N(0, \sigma_i^2)$, in year *y*.

The contribution of each abundance index to the negative log-likelihood function (after the removal of constants) is given by:

$$-\ell n L_i = n_i \ell n(\hat{\sigma}_i) + \frac{n_i}{2} \quad (3)$$

where n_i is the number of annual data values for index *i*.

The assessment model including both observation and process error

The dynamic Schaefer model to take account of both observation and process error considered here is as follows:

$$B_{y+1} = B_y + rB_y\left[1 - \frac{B_y}{K}\right]\{1 + \xi_y\} - C_y \quad (4)$$

where:

B_y is the biomass estimated in year *y*, with the starting biomass B_{1920} assumed to pristine,

r is an estimable parameter (the intrinsic rate of population growth),
K is an estimable parameter (pristine biomass),
ξ_y is the process error in year *y*, and
C_y is the annual catch.

Note that the process error is applied to the surplus production component of the biomass equation rather than to total biomass as per Yemane (2017).

The likelihood is calculated assuming that the abundance indices are log-normally distributed about their expected values:

$$I_y = q_y B_y e^{\varepsilon_y} \tag{5}$$

where I_y is the abundance index for year *y*, qB_y is the corresponding model estimate, and ε_y is the observation error, $\sim N(0, \sigma_{cpue}^2)$, in year *y*.

The contribution of the abundance indices to the negative log-likelihood function (after the removal of constants) is given by:

$$-\ell nL = n\ell n\left(\hat{\sigma}_{cpue}\right) + \frac{n}{2} \tag{6}$$

The contribution of the process errors to the negative log-likelihood function is given by:

$$-\ell nL = \sum_y 0.5(\frac{\xi_y}{\sigma})^2 \tag{7}$$

where σ the associated standard deviation for ξ_y.

Additional analyses conducted

In order to get a better understanding of the model a number of analyses were undertaken related to *r* (intrinsic growth rate) and the process errors ξ_y. These were as follows:

Observation error model:

i. *r*=0, 0.2 and 0.4 respectively,
ii. *r*=0.2 over the period 1920-1999 and *r*=0 from 2000-2016, and
iii. *r*=0.4 over the period 1920-1999 and *r*=0 from 2000-2016.

Observation plus process error model:

iv. *r*=0.2 and σ=0.4, and
v. *r*=0.4 and σ=0.2.

Results

Table 3 reports the estimable parameter *K*, recent biomasses relative to *K*, the contribution of each index (including the process errors ξ_y where applicable) to the model fit and the total negative log-likelihood for the models described above. A comparison of the observation error model fits

($-\ell nL_{total}$) shows that the best fit is achieved for *r*=0 (which clearly does not make biological sense) and as *r* increases so the model fit deteriorates. The model of Yemane (2017) estimated a median *r* value of 0.2, which is still considered to be on the "low" side for sole (an *r* of 0.4 would probably be considered more plausible), but his estimate is heavily influenced by the prior chosen for *r*. Including process error in the model improves the model fit (compare $-\ell nL_{total}$ for *r*=0.2 *vs* *r*=0.2, σ=0.4, for example) and an even better fit is achieved for the observation error model when *r* is fixed over the period 1920-1999 and set at zero post-1999.

Figure 1a plots fits to the CPUE indices for *r*=0.2 scenarios and Figure 1b plots the same, but for *r*=0.4 scenarios. The models including process error show a drop in estimated CPUE in 2000. The corresponding residuals are plotted in Figures 2a and 2b and these are clearly non-random.

Figure3 plots the trends in process error (ξ_y) and these show clear systematic patterns (inconsistent with the Yemane (2017) model approach which assumes the process errors to be randomly distributed).

Discussion

The key inferences from these analyses are:

- Certainly the stationary observation error model is unacceptable, leading to systematic misfits to abundance index data (see Figures 1 and 2).
- Adding process error to the model improves the fit to the CPUE data, but the associated process errors show systematic patterns rather than the randomness assumed by the associated models.
- Allowing for non-stationarity in the form of a change in the value of the *r* parameter post-1999 can maintain a good fit to the CPUE data, and provides an improved overall (penalised) log-likelihood, though the *r* value estimated to apply from 2000 onwards is unrealistically low.

Broadly speaking, this suggests a non-stationary situation, with some change in the dynamics of the resource (or equivalently in the fishery catchability, *q*, in the new century). This is compatible with the approach the DWG has used in recent years (e.g. as in Glazer and Fairweather, 2017) to provide management advice for this resource.

On the positive side, the framework developed here (specifically the final model allowing for some change in the dynamics and using an observation error estimator) provides an improvement to that used previously by the DWG, and can provide a basis to take such analyses further forward.

A difficulty, however, is that further computations with such models (not reported in detail here) indicate that the abundance data reflect insufficient contrast to allow a reliable estimate of the *r* parameter that applies in the 1900s (this is a manifestation of what is known as the "one way trip" effect). Some further research (e.g. the use of literature surveys to obtain information on similar species) will be needed to decide upon an appropriate range of values to consider in further analyses.

References

Fairweather TP, Glazer JP. 2018. Extending the Agulhas sole Catch per Unit Effort (CPUE) time series (1983-2016). Unpublished DAFF Working Group Document, *FISHERIES/2018/MAR/SWG-DEM/12*. 6pp.

Glazer JP, Fairweather TP. 2017. An updated assessment of the Agulhas sole resource, *Austroglossus pectoralis*. Unpublished DAFF Working Group Document, *FISHERIES/2017/SEP/SWG-DEM/25*. 12pp.

Yemane D. 2017. Appraising the state of Agulhas sole (*Austroglossus pectoralis*) using state-space Bayesian surplus production implemented in STAN. Unpublished DAFF Working Group Document, *FISHERIES/2017/SEP/SWG-DEM/37*. 53pp

Table 1: Annual catches of sole.

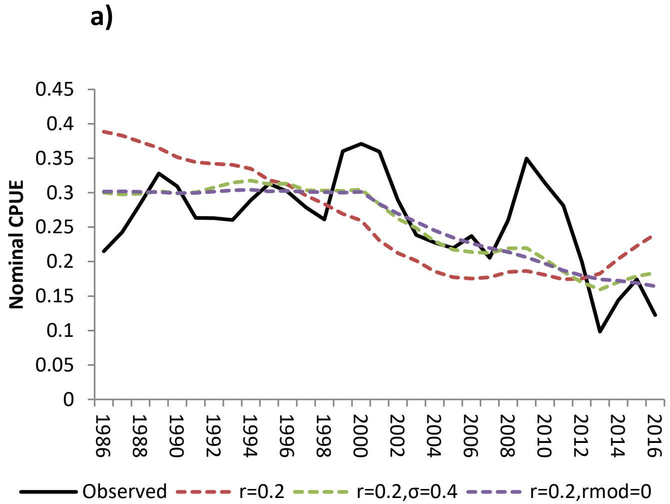
Year	Catch (t)	Year	Catch (t)	Year	Catch (t)
1920	700	1955	740	1990	808
1921	540	1956	740	1991	716
1922	560	1957	700	1992	704
1923	670	1958	700	1993	772
1924	680	1959	750	1994	938
1925	650	1960	850	1995	769
1926	820	1961	820	1996	909
1927	750	1962	800	1997	840
1928	770	1963	732	1998	859
1929	740	1964	690	1999	757
1930	780	1965	841	2000	1060
1931	680	1966	575	2001	850
1932	760	1967	520	2002	702
1933	800	1968	445	2003	754
1934	900	1969	642	2004	612
1935	1100	1970	663	2005	485
1936	1050	1971	877	2006	428
1937	1200	1972	1044	2007	331
1938	1000	1973	961	2008	448
1939	800	1974	611	2009	568
1940	650	1975	763	2010	570
1941	650	1976	1040	2011	436
1942	650	1977	500	2012	338
1943	750	1978	850	2013	127
1944	680	1979	899	2014	208
1945	675	1980	943	2015	258
1946	710	1981	1026	2016	120
1947	730	1982	817		
1948	680	1983	682		
1949	700	1984	857		
1950	710	1985	880		
1951	670	1986	796		
1952	700	1987	855		
1953	730	1988	839		
1954	750	1989	913		

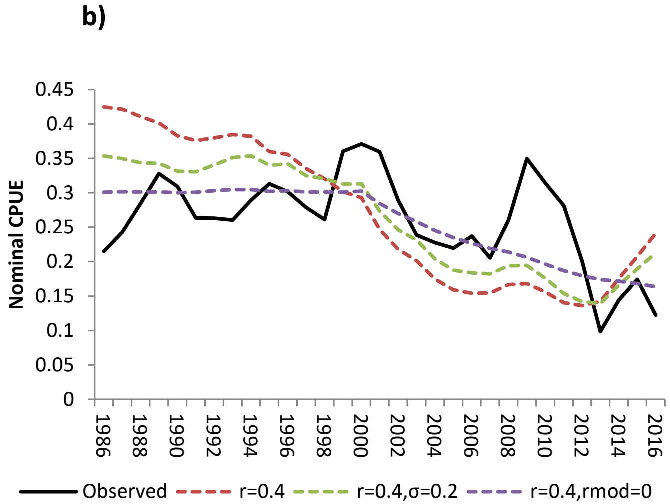
Table 2: Indices of relative abundance obtained from surveys (t) and commercial CPUE (kg/min).

Year	Autumn "old" index	Autumn "new" index	Spring "old" index	Spring "new" index	Nominal CPUE index
1986			3747.15		0.22
1987			3419.56		0.24
1988	2753.38				0.28
1989					0.33
1990					0.31
1991	3894.63				0.26
1992	4942.32				0.26
1993	4802.34				0.26
1994	3088.35				0.29
1995	4733.55				0.31
1996	4568.96				0.30
1997	2114.68				0.28
1998					0.26
1999	3547.29				0.36
2000					0.37
2001			5129.56		0.36
2002					0.29
2003	3338.29			1504.33	0.24
2004		1291.01		1255.07	0.23
2005		603.49			0.22
2006	2272.59		3172.65		0.24
2007		811.74		134.78	0.21
2008		300.11		1070.75	0.26
2009		1636.68			0.35
2010	1419.25				0.31
2011		707.03			0.28
2012					0.20
2013					0.10
2014		1003.73			0.14
2015		1281.24			0.17
2016		496.21			0.12

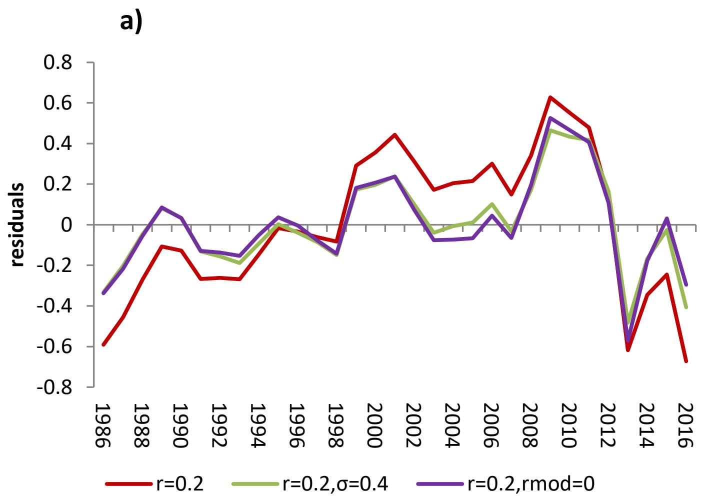
Table 3: Results from the analyses conducted for (i) observation (denoted "Obs") and (ii) observation plus process error models (denoted "Obs+proc"). Intrinsic growth, *r*, is a fixed input and *K* (pristine biomass in tons) is an estimable parameter. Depletion levels in 2016 and 2017 are also reported, as are the $-\ell n L_i$ contributions for each of the indices of abundance and the process error (where applicable).

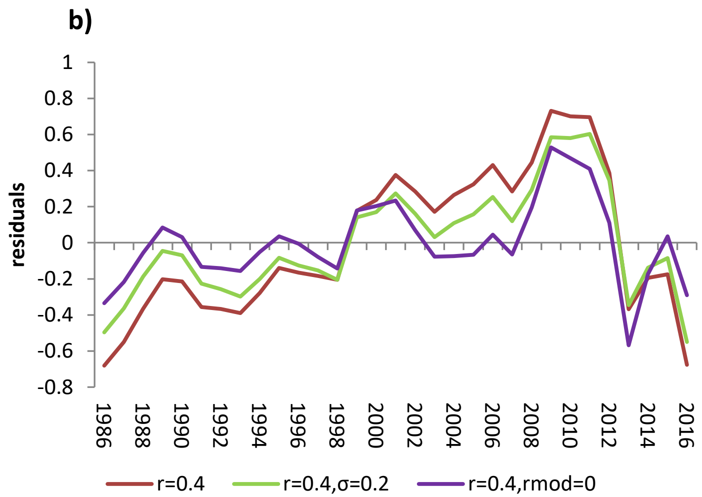
	r=0	*r*=0.2	*r*=0.2,σ=0.4	*r*=0.2,rmod=0	*r*=0.4	*r*=0.4,σ=0.2	*r*=0.4,rmod=0
Model	**Obs**	**Obs**	**Obs+proc**	**Obs**	**Obs**	**Obs+proc**	**Obs**
K	100960	14389.7	14275	23314.4	7685.67	7601.31	20137.3
B2016/K	0.31	0.27	0.20	0.42	0.26	0.23	0.48
B2017/K	0.31	0.30	0.23	0.42	0.33	0.28	0.47
-lnL CPUE	-25.58	-17.41	-31.58	-30.68	-13.11	-22.22	-30.72
-lnL old_aut	-8.05	-8.88	-9.24	-9.06	-8.65	-9.99	-9.11
-lnL new_aut	-1.60	-1.49	-1.38	-1.30	-1.25	-1.40	-1.29
-lnL old_spr	-3.36	-1.74	-4.99	-4.96	-1.21	-2.98	-5.00
-lnL new_spr	1.84	1.76	1.73	1.74	1.60	1.63	1.73
-lnL ξ	-	-	8.87	-	-	6.23	-
-lnL total	-36.76	-27.75	-36.59	-44.28	-22.62	-28.74	-44.39





Figures 1a and b: Fits to the nominal CPUE index. The left panel shows the fits for the various options for *r*=0.2 and the right panel for *r*=0.4.





Figures 2a and b: Residuals corresponding to models for *r*=0.2 (left panel) and *r*=0.4 (right panel).

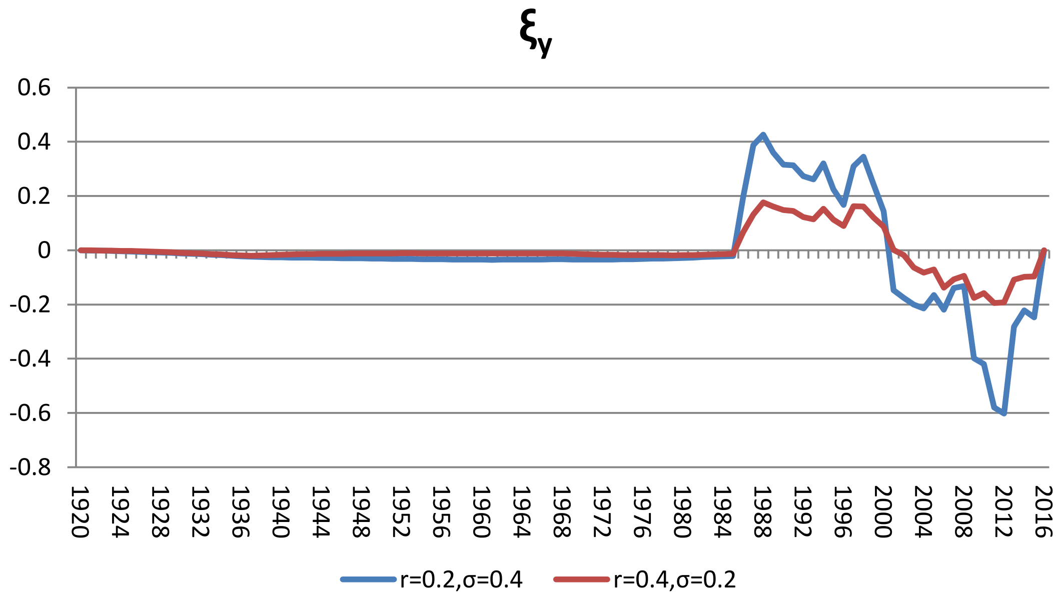


Figure 3: Trends in process error.