Technical note

Comments on incongruous formulations in the SAM (state-space assessment model) model and consequences for fish stock assessment

M. Aldrin[a,*], S. Aanes[a], S. Subbey[b]

[a] Norwegian Computing Center, P.O. Box 114 Blindern, N-0314 Oslo, Norway
[b] Institute of Marine Research, P.O. Box 1870 Nordnes, N-5817 Bergen, Norway

ARTICLE INFO

ABSTRACT

The stock assessment model SAM (state-space assessment model) is an important tool for analysing fishery data and is used for abundance estimation of several different fish stocks. We demonstrate incongruous formulations in the equations that relate the catch data to fish abundance, and in the fishery and natural mortality rates. We present example results to show how these formulations introduce bias, and compound the uncertainty in model estimates. Finally, we provide corrections to the system equations to remedy the drawback.

1. Introduction

The stock assessment model SAM (state-space assessment model) presented in Nielsen and Berg (2014) and in Berg and Nielsen (2016) has become a popular tool for fisheries science in recent years. Various configurations of the model are used by many ICES working groups to assess fish stocks (e.g., ICES, 2017a,b, 2018). This widespread is partly due to the user-friendly and flexible implementation of SAM (https://github.com/fishfollower/SAM) within the Template Model Builder (TMB) framework (Kristensen et al., 2016).

We consider the model structure of SAM as it is described in Nielsen and Berg (2014) and Berg and Nielsen (2016), even though other variants exist. We argue that the total mortality is mis-specified in the catch equation, i.e. in the equation where the number of fish caught is computed as a proportion of the number of fish that have died from one year to the next. Furthermore, we demonstrate that the model formulation implies that the expected values of catch estimates are biased.

2. Models

2.1. A basic model

As a starting point we will assume that:

1. The fishing mortality rate per age group is constant during a year.
2. The natural mortality rate per age group is constant during a year.
3. There is no migration of fish into or out from the area under consideration.

These assumptions are widely used in stock assessment models, either explicitly stated or implicitly given by the models used. If the first two of the above assumptions are not met, a year may be divided into shorter time intervals, though this is not our focus here.

The assumptions above imply that the rate of change of number of fish at age a during year y is proportional to the current number of fish at that age, i.e.

$$dN_{a,t}/dt = -(F_{a,y} + M_{a,y})N_{a,t}, \qquad (1)$$

where $F_{a,y}$ is the fishing mortality rate at age a during year y, $M_{a,y}$ is the corresponding natural mortality rate and $N_{a,t}$ is the number of fish at age a at a time point t within year y. A solution to this differential equation gives the relationship between the number of fish $N_{a,y}$ at age a at the beginning of year y, and the number of fish in the next age group one year later as

$$N_{a+1,y+1} = \exp(-F_{a,y} - M_{a,y})N_{a,y}. \qquad (2)$$

Furthermore, the total catch of fish at age a during year y is then given by

$$C_{a,y} = [F_{a,y}/(F_{a,y} + M_{a,y})](1 - \exp(-F_{a,y} - M_{a,y}))N_{a,y}, \qquad (3)$$

where $(1 - \exp(-F_{a,y} - M_{a,y}))N_{a,y} = N_{a,y} - N_{a+1,y+1}$ is the number of fish that have died from one year to the next and $[F_{a,y}/(F_{a,y} + M_{a,y})]$ is the proportion of these that have died due to fishing activities.

Above, $M_{a,y}$ was defined as the (total) natural mortality rate. If we assume that the natural mortality rate is stochastic, it may be convenient to instead write it as $M_{a,y} + m_{a,y}$, where $M_{a,y}$ now is the *expected* mortality rate (at this scale), and $m_{a,y} \sim N(0, \sigma_m^2)$. Then Eq. (2) is

extended to

$$N_{a+1,y+1} = \exp(-F_{a,y} - M_{a,y} - m_{a,y})N_{a,y}$$
$$= \exp(-F_{a,y} - M_{a,y})N_{a,y}\exp(-m_{a,y}). \tag{4}$$

The total mortality rate is now $F_{a,y} + M_{a,y} + m_{a,y}$, and the number of fish that have died from one year to the next is $N_{a,y} - N_{a+1,y+1} = (1 - \exp(-F_{a,y} - M_{a,y} - m_{a,y}))N_{a,y}$, so the resulting catch equation must be changed to

$$C_{a,y} = [F_{a,y}/(F_{a,y} + M_{a,y} + m_{a,y})](1 - \exp(-F_{a,y} - M_{a,y} - m_{a,y}))N_{a,y}. \tag{5}$$

The restriction $(M_{a,y} + m_{a,y}) > = 0$ is required to prevent negative natural mortality rates. If $m_{a,y}$ instead represents, or at least includes, migration, it is enough to use the restriction $F_{a,y} + M_{a,y} + m_{a,y} > = 0$. Note that if $F_{a,y} + M_{a,y} + m_{a,y} = 0$, Eq. (5) should be replaced by $C_{a,y} = F_{a,y}N_{a,y}$ since $\lim_{x\to 0}(1 - \exp(-x))/x = 1$, where $x = F + M + m$. While this may be an unlikely result in a final estimate of a model, it may still be useful to include this adjustment in the computer code to prevent numerical problems during the iterative optimisation of the likelihood.

2.2. SAM – the state-space assessment model

Here we present those parts of the SAM that are relevant for our comments to follow. We consider the model structure described in the paper (Nielsen and Berg, 2014), and use a similar notation as in that paper. Unlike in Nielsen and Berg (2014), in the discussion to follow, we reformulate the model using the original scale, i.e. the number, rather than the logarithm, of number of fish.

The total true catch of fish at age a during year y is given by Eq. (3) above (though not explicitly stated), whereas the recursive equations for the number of fish at ages are given by

$$N_{a+1,y+1} = \exp(-F_{a,y} - M_{a,y})N_{a,y}\exp(\eta_{a+1,y+1}) \quad \text{for } 1 < a < A, \tag{6}$$

$$\eta_{a+1,y+1} \sim N(0, \sigma_S^2) \quad \text{for } a > 1. \tag{7}$$

Nielsen and Berg use both the terms "process error" and "survival process" (Berg and Nielsen, 2016) for the stochastic term $\eta_{a,y}$. Since there is a one to one relationship between survival and mortality, and since Eq. (6) is the same as our Eq. (4) with $-\eta_{a+1,y+1} = m_{a,y}$ and $\sigma_S^2 = \sigma_m^2$, the (total) natural mortality rate is $M_{a,y} + m_{a,y} = M_{a,y} - \eta_{a+1,y+1}$. Therefore, the catch equation Eq. (3) used by Nielsen and Berg should be replaced by Eq. (5). In addition, the restriction $(M_{a,y} + m_{a,y}) > = 0$ should be included to prevent negative natural mortality rates.

We have investigated the consequences of using the wrong catch equation Eq. (3) when the stochastic term $m_{a,y}$ is included in the recursive equations for the number of fish, i.e., Eq. (4) or Eq. (6), by comparing the fishing mortality rate F^{SAM} computed by Eq. (6) with the true fishing mortality rate F^{true}. We have done this for various values of F^{true} and σ_m for $M = M_{a,y} = 0.2\,\mathrm{yr}^{-1}$. The bias $E(F^{true} - F^{SAM})$ is negligible for small and moderate values of F^{true} and σ_m, but F^{SAM} tends to be lower than F^{true} if both F^{true} and σ_m are high (Table 1). The corresponding standard error $sd(F^{true} - F^{SAM})$ is also negligible when F^{true} and/or σ_m are small, but is significant even for moderate values of F^{true} and σ_m (Table 2). The results are similar when $M = 0.1\,\mathrm{yr}^{-1}$ or $M = 0.3\,\mathrm{yr}^{-1}$. In their preferred model D for a case study on North Sea cod in Nielsen and Berg (2014), σ_m is estimated to be 0.10 (given in their Supplementary Data). In this case, σ_m is so small that including the terms $\eta_{a+1,y+1}$ probably has little practical significance. However, in cases where σ_m is larger, we believe this may introduce some undesirable bias and compound the uncertainty.

Table 1
Introduced bias $E(F^{true} - F^{SAM})$ in the estimate of the true fishing mortality F^{true} for $M = M_{a,y} = 0.2\,\mathrm{yr}^{-1}$ when using catch equation Eq. (3) without the "process error", while the "process error" is included in the recursive equations Eq. (4) or Eq. (6).

F^{true}	σ_m			
	0.1	0.2	0.3	0.4
0.1	0.00	0.00	0.00	0.00
0.5	0.00	0.00	−0.01	−0.02
1	0.00	−0.02	−0.04	−0.07
1.5	−0.01	−0.04	−0.10	−0.20

Table 2
Introduced standard error $sd(F^{true} - F^{SAM})$ in the estimate of the true fishing mortality F^{true} for $M = M_{a,y} = 0.2\,\mathrm{yr}^{-1}$ when using catch equation Eq. (3) without the "process error", while the "process error" is included in the recursive equations Eq. (4) or Eq. (6).

F^{true}	σ_m			
	0.1	0.2	0.3	0.4
0.1	0.00	0.01	0.02	0.02
0.5	0.03	0.06	0.09	0.12
1	0.06	0.13	0.22	0.36
1.5	0.11	0.25	0.46	0.74

The main data used by Nielsen and Berg (2014) are catch data and survey indexes, though other data sources can also be included. The comments to follow, deal with how catch data are handled in SAM. Let $\hat{C}_{a,y}$ denote a point estimate of the catch at age a in year t, and whose true value, $C_{a,y}$, is unknown. In Nielsen and Berg (2014), these are related in a set of observation equations as

$$\hat{C}_{a,y} = C_{a,y}\exp(\varepsilon_{a,y}^C), \tag{8}$$

$$\varepsilon_{a,y}^C \sim N(0, \sigma_{C,a}^2). \tag{9}$$

This implies that the expected value of the catch estimate $\hat{C}_{a,y}$ is a factor $\exp(\sigma_{c,a}^2/2)$ higher than the true catch. Furthermore, in Nielsen and Berg (2014), the estimated values of $\sigma_{C,a}$ are 0.71 for $a = 1$, 0.20 for $a = 2$, and 0.09 for $a > = 3$. These estimates of $\sigma_{c,a}$ lead respectively, to upward biases of 29%, 2% and 0.4%, i.e. high for $a = 1$, but negligible for higher ages.

Nielsen and Berg (2014) assume a random walk process for the logarithms of the fishing mortality rates, which on the original scale may be written as

$$F_{a+1,y+1} = F_{a,y}\exp(\varepsilon_{a,y}^F), \tag{10}$$

$$\varepsilon_{a,y}^F \sim N(0, \sigma_F^2). \tag{11}$$

This implies that the fishing mortality is expected to increase by a factor $\exp(\sigma_F^2/2)$ from one year to the next. In Nielsen and Berg (2014), the estimated value of σ_F is 0.11, which implies an annual increase of 0.6%. It must be mentioned that this assumption is both odd and of ambiguous relevance in this case.

3. Conclusions

We have revealed some incongruous model formulations in SAM and partly illustrated their consequences for numerical estimates. We suggest that if the stochastic terms $m_{a,y}$ (or $\eta_{a+1,y+1}$) are kept in the recursive equations for the numbers of fish, $m_{a,y}$ should be included in the catch equation as well, i.e., Eq. (3) should be replaced by Eq. (5).
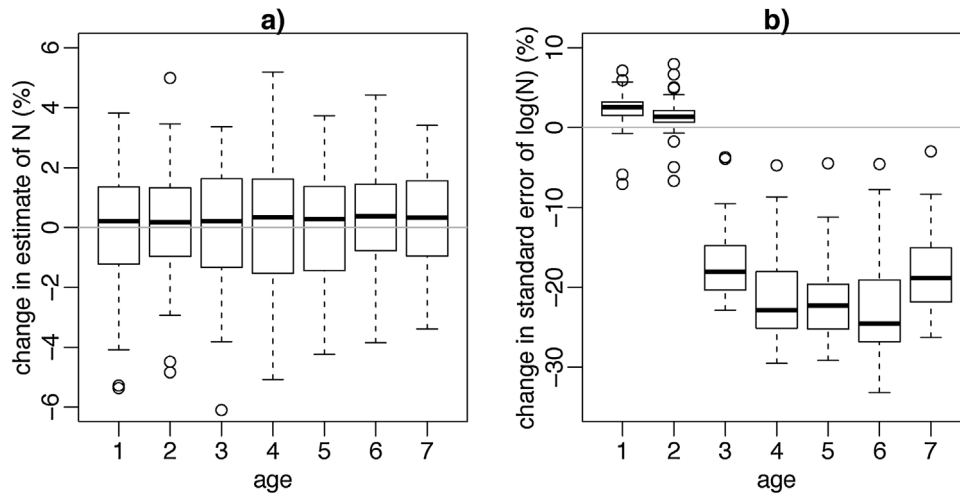
Fig. 1. Age-wise box plots of changes in estimated abundances and their estimated standard errors when correcting the catch equation.

Table 3

Fixed values of natural mortality M_a and estimated values of other quantities from the seven fish stock assessments with status "active" from https://www. stockassesment.org per 1. August 2018. When there exists multiple analyses of the same data set, the "final" assessment is chosen. When values varies over years and ages ($F_{a,y}$) or over ages (other quantities), the minimum and maximum values are given. The minimum values of $F_{a,y}$ are always from the lowest age.

Species	Area	$F_{a,y}$	M_a	σ_m	$\sigma_{C,a}$	Implied bias in catch estimates (%)
Cod	North Sea	0.06–1.18	0.2–1.1	0.12	0.06–0.59	0–19
Cod	Iceland and East Greenland	0.00–2.37	0.2	0.17	0.82	40
Cod	Faroe Plateau Ecosystem	0.03–1.54	0.2	0.29	0.25	3
Saithe	North Sea	0.06–0.89	0.2	0.17	0.23	3
Saithe	Faroe Plateau Ecosystem	0.01–1.10	0.2	0.27	0.41	9
Sole	North Sea	0.03–0.61	0.1	0.18	0.29–0.63	4–22
Herring	Baltic Sea	0.02–0.30	0.15	0.16–0.21	0.14–0.16	1–1

Furthermore, we suggest that the observation equations for the catch are modified to give unbiased catch estimates. This is easily done by replacing Eq. (9) with

$$\varepsilon_{a,y}^C \sim N(-\sigma_{C,a}^2/2, \ \sigma_{C,a}^2). \tag{12}$$

Similarly, the expectations of the stochastic term $\varepsilon_{a,y}^F$ could be changed from 0 to $-\sigma_F^2/2$ to prevent implicit assumptions of positive increasing fishing mortality.

Our suggested modifications of the SAM model are easy to implement by altering the TMB code published by Nielsen and Berg (2014). When we corrected the catch equation, and refitted their Model D on the North Sea cod data used in their paper, the AIC (Akaike's Information Criterion, Akaike, 1974) was improved by 11.4, i.e. a significant improvement. The corresponding estimates of the abundances $N_{a,y}$ were changed slightly, at most 6% downwards or upwards, but their average was almost unchanged (panel a), Fig. 1). However, there was a clear change in the corresponding estimated standard errors (panel b), Fig. 1). These increased slightly for the two first age group, but decreased significantly for the remaining ages. Recall that the estimate of σ_m was rather small for these data, so the consequences of correcting the catch equation will probably be more prominent for data sets with larger values of σ_m.

When we in addition changed the expectations of the stochastic terms $\varepsilon_{a,y}^C$ and $\varepsilon_{a,y}^F$ from 0 to minus half of their variances, the estimates of the abundances $N_{a,y}$ were increased with in average 5% for $a = 1$, 1.4% for $a = 2$ and 0.6% for higher ages, compared to the model where only the catch equation was corrected. The corresponding standard errors were almost identical. There was a slight (0.6) increase in the AIC, so this adjustment did not improve the fit. The moderate increase in the AIC value should be taken with the following caveat. First, we have assumed that the catch estimates were unbiased. The literature, however, indicates strong bias in the historically long catch data (see Patterson, 1998). The focus therefore, should be on the change in computed abundance levels (and possible effects on harvest rules), induced by varying the expectations of the stochastic terms.

The North Sea cod data has been used mainly for the sake of comparison with results in Nielsen and Berg. It should, however, be borne in mind that different data sets may show different sensitivity to the biases that result from the original model formulation. Table 3 gives an overview over relevant quantities from SAM analyses of six other fish stocks, as well as from an updated SAM analysis of the North Sea cod data. The standard deviation σ_m of the "process error" is higher for the six other fish stocks than for the North Sea cod, indicating that the problem with the catch equation often may be more severe than in our example. Furthermore, the implied bias in the catch estimates are between 9 and 40% for at least some ages in three of these six stocks, and removing this implied bias will probably affect the abundance estimates.

Acknowledgements

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Berg, C., Nielsen, A., 2016. Accounting for correlated observations in an age-based state-space stock assessment model. ICES J. Mar. Sci. 73, 1788–1797. https://doi.org/10.1093/icesjms/fsw046.

ICES, 2017a. Report of the Arctic Fisheries Working Group (AFWG), 19–25 April 2017. Technical Report ICES CM 2017/ACOM:06. ICES, Copenhagen, Denmark 493 pp.

ICES, 2017b. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 30 August–5 September 2017. Technical Report ICES CM 2017/ACOM:23. ICES, Copenhagen, Denmark 994 pp.

ICES, 2018. Report of the Working Group on Assessment of Demersal Stocks in the North Sea and Skagerrak (2017), 26 April–5 May 2017. Technical Report ICES HQ. ICES CM 2017/ACOM:21. ICES, Copenhagen, Denmark 1077 pp.

Kristensen, K., Nielsen, A., Berg, C., Skaug, H., Bell, B., 2016. TMB: automatic differentiation and Laplace approximation. J. Stat. Softw. 70, 1–21. https://doi.org/10.18637/jss.v070.i05.

Nielsen, M., Berg, C., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. 158, 96–101. https://doi.org/10.1016/j.fishres.2014.01.014.

Patterson, K., 1998. Assessing fish stocks when catches are misreported: model, simulation tests, and application to cod, haddock, and whiting in the ICES area. ICES J. Mar. Sci. 55, 878–891. https://doi.org/10.1006/jmsc.1998.0351.