



Original Article

Accounting for correlated observations in an age-based state-space stock assessment model

Casper W. Berg* and Anders Nielsen

National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund Castle, Charlottenlund 2920, Denmark

*Corresponding author: e-mail: cbe@aqu.dtu.dk

Berg, C. W., and Nielsen, A. Accounting for correlated observations in an age-based state-space stock assessment model. – ICES Journal of Marine Science, 73: 1788–1797.

Received 18 November 2015; revised 29 February 2016; accepted 3 March 2016.

Fish stock assessment models often rely on size- or age-specific observations that are assumed to be statistically independent of each other. In reality, these observations are not raw observations, but rather they are estimates from a catch-standardization model or similar summary statistics based on observations from many fishing hauls and subsamples of the size and age composition of the data. Although aggregation mitigates the strong intra-haul correlation between sizes/ages that is usually found in haul-by-haul data, violations of the independence assumption can have a large impact on the results and specifically on reported confidence bounds. A state-space assessment model that allows for correlations between age groups within years in the observation model for catches and surveys is presented and applied to data on several North Sea fish stocks using various correlation structures. In all cases the independence assumption is rejected. Less fluctuating estimates of the fishing mortality is obtained due to a reduced process error. The improved model does not suffer from correlated residuals unlike the independent model, and the variance of forecasts is decreased.

Keywords: age-based correlation structure, correlated observations, data weighting, SAM, state-space model, stock assessment, template model builder.

Introduction

Statistical analysis of catch data from commercial and scientific fishing vessels for stock assessment requires assigning appropriate relative weights to each data source, and the results may depend heavily on these weights (Francis and Hilborn, 2011). The analysis typically requires the optimization of a likelihood function, and the appropriate data weighting is obtained asymptotically when the likelihood function of the model is correctly specified, particularly it is important that any assumptions of independence are valid.

Integrated assessment models (Maunder and Punt, 2013) allow optimal weights to be estimated through the optimization of a joint likelihood for all the observed data. Although formal methods exist for validating that the likelihood corresponds to the way data are distributed, specifying an appropriate likelihood function for fisheries data is inherently difficult (Francis, 2014). However, if the likelihood is seriously misspecified, the output of the model and hence the stock assessment is essentially useless, so performing the model validation to ensure that the likelihood is correctly specified is important.

The catch data typically contain two types of information: abundance data (total numbers or biomass) and composition data (length or age frequencies), which can have highly varying precision due to differences in the sampling design between fleets, years, etc. A common approach is to use a multinomial likelihood for the composition data (e.g. Stock Synthesis (Methot and Wetzel, 2013)), but this distribution fails to account for the overdispersion and correlations that are usually found in compositional catch data (e.g. Pennington and Volstad, 1994; Hrafnkelsson and Stefánsson, 2004; Francis, 2014). These issues can be dealt with by pretending that the observed frequencies were obtained from a reduced sample, whose size is often referred to as the effective sample size (Thorson, 2014). However, the effective sample sizes are often specified outside of the assessment model, and so it does not fit well with the self-weighting integrated assessment approach. In addition, the correlation structure found in real data does typically not match the implied negative correlations of the multinomial model (Francis, 2014). Hence, the challenge remains to find a suitable distribution that describes the variation in catch data. Francis (2014) suggested

the logistic-normal distribution to replace the multinomial distribution for composition data. An alternative, which we will consider in this paper, is to use a lognormal likelihood for the catch-at-age data, i.e. the product of composition and abundance data rather than fitting compositions and abundances separately. A disadvantage of this approach is that information about actual sample sizes in each haul are not utilized, since we are considering an aggregation of all samples. On the other hand, the sample sizes are difficult to utilize directly, because data are not well described by a simple distribution such as the multinomial, and often sample sizes are not even readily available—only the catch-at-age data. In the situation where the total catch weight is known with high precision but the composition is more uncertain, separate abundance and composition likelihoods with different associated variances might be still preferable. Nevertheless, a clear advantage of analysing catch-at-age data using a lognormal likelihood is that both negative and positive correlations of any magnitude are easily accounted for. When applied in a state-space modelling framework, it also has the desirable properties (Francis, 2014) of being self-weighting (weights are estimated within the model as variance parameters) and various parsimonious correlation structures exist for this distribution.

The common approach is to ignore the problem of correlated errors (Francis and Hilborn, 2011), although some authors have addressed the issue (e.g. Myers and Cadigan, 1995; McDonald *et al.*, 2001; Berg *et al.*, 2014). However, such correlations should be expected (Francis, 2014).

Berg *et al.* (2014) estimated the between-age correlations in survey index catch-at-age estimates using bootstrap methodology and haul-by-haul survey data and found a general pattern of increasing positive correlations with age (Figure 1). These correlations were

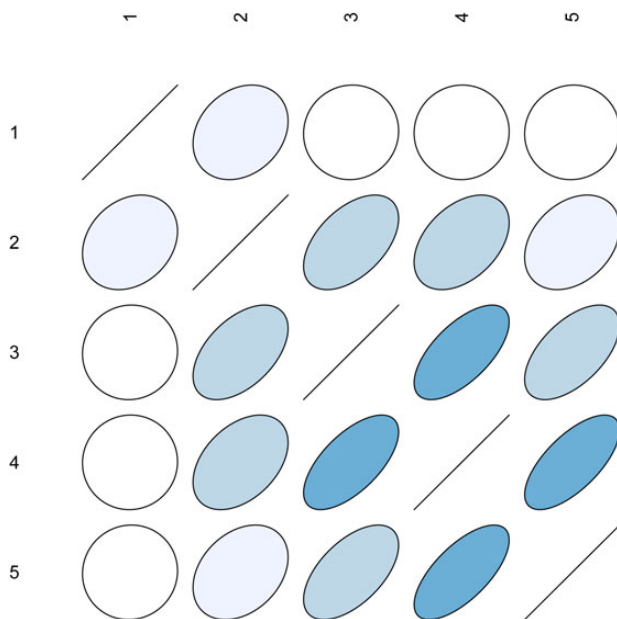


Figure 1. The correlation matrix between ages for the survey index for North Sea Autumn Spawning Herring in quarter 1 in 2005 (recreated from Berg *et al.* (2014)). Each ellipse represents the level curve of a bivariate normal distribution with the corresponding correlation. Hence, the sign of a correlation correspond to the sign of the slope of the major ellipse axis. Increasingly darker shading is used for increasingly larger absolute correlations, while uncorrelated pairs of ages are depicted as circles with no shading.

subsequently treated as known input to a stock assessment model and improved the results compared with assuming independent survey catch-at-age. By utilizing the extra information in haul-by-haul data, this method has the advantage of not introducing any additional parameters to estimate in the stock assessment model, not imposing any particular structure on the correlation structure, and finally that the correlation structure can change between years. However, the bootstrap approach is fairly computationally expensive because the index standardization procedure has to be repeated for each bootstrap sample and the requirement of haul-by-haul data restricts its applicability. Another approach was taken by Myers and Cadigan (1995), where the correlation structure was estimated within the assessment model using catch-at-age only rather than haul-by-haul data and using a simple correlation structure (compound symmetry) with only one additional parameter to estimate. They found that survey data are usually positively correlated among ages, and failing to account for this can greatly increase bias and variance if the correlations are large. Here we will follow the same approach as Myers and Cadigan (1995) of using a time-constant parameterized correlation structure, but using a modern state-space formulation of the assessment model and considering several different parameterizations inspired by the empirical correlation structures found by Berg *et al.* (2014).

Material and methods

The stock assessment model used here is an extension of the state-space assessment model (SAM) by Nielsen and Berg (2014), which is used as basis for management advice for several stocks monitored by the International Council for the Exploration of the Sea. State-space models present a general framework for analysing dynamical systems, where the quantities of interest (the state) are not observed directly, but rather through indirect measurements with noise, which are related to the states via the observation equations. States are connected in time through the state (or system) equations and are also subject to random perturbations known as process noise. An important feature of state-space models is the quantification of the random variability in both the observations and the system equations. This is expressed through observation and process variances, respectively, which can be estimated using maximum likelihood techniques and give the appropriate weighting to each data source in the estimation of the unobserved state. The unobserved state vector consists of the log-transformed numbers-at-age $\log N_1, \dots, \log N_A$ and fishing mortalities $\log F_1, \dots, \log F_A$. The oldest age groups may share fishing mortality, which is indicated by using maximum index A^* instead of A , which is the oldest age group in the assessment. The state equations for $\log N_i$ are as in Nielsen and Berg (2014):

$$\log N_{1,y} = \log N_{1,y-1} + \varepsilon_y^{(R)}, \quad (1)$$

$$\log N_{a,y} = \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \varepsilon_{a,y}^{(S)} \text{ for } 2 \leq a < A, \quad (2)$$

$$\log N_{A,y} = \log(e^{\log N_{A-1,y-1} - F_{A-1,y-1} - M_{A-1,y-1}} + e^{\log N_{A,y-1} - F_{A,y-1} - M_{A,y-1}}) + \varepsilon_{A,y}^{(S)}, \quad (3)$$

where $M_{a,y}$ is the natural mortality at age a in year y , which is assumed to be known *a priori*, and F is the total fishing mortality. Although more

elaborate functions can be used, a simple random walk model is chosen for the recruitment process (eq. 1) as in [Berg et al. \(2014\)](#). The process errors on $\log N$ are assumed to be independent normal with (at least) separate variance parameters for recruitment and survival: $\varepsilon^{(R)} \sim N(0, \sigma_R^2)$ and $\varepsilon^{(S)} \sim N(0, \sigma_S^2)$.

The fishing mortality vector $F_y = \{F_{1,y}, \dots, F_{A^*,y}\}$ is assumed to follow a correlated random walk with an AR(1) correlation structure as in [Nielsen and Berg \(2014\)](#):

$$\log F_y = \log F_{y-1} + \varepsilon_y^{(F)}, \quad (4)$$

where $\varepsilon_y^{(F)} \sim N(\mathbf{0}, \Sigma^{(F)})$. The covariance matrix $\Sigma^{(F)}$ is constructed via the vector of process standard deviations σ_F and the correlation matrix $\mathbf{R}^{(F)}$ as $\Sigma^{(F)} = \text{diag}(\sigma_F) \mathbf{R}^{(F)} \text{diag}(\sigma_F)$. The elements in $\mathbf{R}^{(F)}$ are defined as $\mathbf{R}_{a,\tilde{a}}^{(F)} = \rho^{|a-\tilde{a}|}$, where $-1 < \rho < 1$, which is known as AR(1) correlation structure.

When observations are assumed multivariate normal rather than independent, which is where we depart from [Nielsen and Berg \(2014\)](#), we obtain the following observation equations for total catches C and survey indices I :

$$\log C_y = \log \left(\frac{F_y}{Z_y} (1 - e^{-Z_y}) N_y \right) + \varepsilon_y^{(C)}, \quad (5)$$

$$\log I_y^{(s)} = \log \left(Q^{(s)} e^{-Z_y(D^{(s)}/365)} N_y \right) + \varepsilon_y^{(I^{(s)})}, \quad (6)$$

where $\varepsilon_y^{(I^{(s)})} \sim N(\mathbf{0}, \Sigma_y^{(I^{(s)})})$ and $\varepsilon_y^{(C)} \sim N(\mathbf{0}, \Sigma_y^{(C)})$ express the possible covariances between age groups within years, $Z_{a,y} = M_{a,y} + F_{a,y}$ is the total mortality rate, $D^{(s)}$ is the number of days into the year where the survey s is conducted, and $Q_a^{(s)}$ are catchability parameters. Note that $Q^{(s)}$, $D^{(s)}$, and all quantities with year subscript but without age subscript are vectors containing all age groups at once, and multiplication and division operations are elementwise. As with the fishing mortalities, we choose to parameterize the observation covariance matrices via the vectors of process standard deviations and correlation matrices. However, to allow for a correlation structure that resembles the ones found by [Berg et al. \(2014\)](#) (Figure 1), we consider the following parameterization of the correlation matrix

$$\mathbf{R}_{a,\tilde{a}} = 0.5^{|d_a - d_{\tilde{a}}|} \quad 1 \leq a, \tilde{a} \leq A_f, \quad (7)$$

where A_f is the number of age groups for fleet f , and $d_1 = 0$ and $d_2 \dots d_{A_f}$ are parameters to be estimated with the constraints that $d_a \leq d_{\tilde{a}}$ for all $a < \tilde{a}$. Equation (7) corresponds to an AR(1) structure on an irregular lattice [IRAR(1)], where the lattice is defined by the d 's. If all neighbouring distances given by $|d_a - d_{a-1}|$ can be assumed equal, then the regular AR(1) structure is obtained. For distances going to infinity, a correlation of zero is obtained between the corresponding age groups, and the zero limit gives a correlation of one. The constraints on the d 's are obtained by estimating $\log \Delta d_a = \log(d_a - d_{a-1})$ rather than the d 's themselves. This parameterization can, with few model parameters, capture the pattern of gradually increasing positive correlations with older age groups seen in surveys ([Berg et al., 2014](#)) by having decreasing neighbouring distances with age. The base constant of $\rho = 0.5$ in the IRAR(1)-structure is arbitrary, but to avoid over-parameterization one must fix either one of the distance parameters or ρ .

Finally, we consider free unconstrained parameterization of \mathbf{R} (via its Cholesky factor \mathbf{L} , such that $\mathbf{R} = \mathbf{L}\mathbf{L}^T$), which has

$A_f(A_f - 1)/2$ free parameters per fleet. The five following models are investigated:

- (i) All observations are independent ($\mathbf{R} = \mathbf{I}$). No parameters are related to correlation.
- (ii) Regular lattice AR(1) observation correlation structure for all fleets. One correlation parameter per fleet, although surveys may share parameters.
- (iii) Irregular lattice AR(1) observation correlation structure for all fleets. Between 2 and $A_f - 1$ parameters per lattice, and lattice parameters are allowed to be shared among fleets.
- (iv) Unconstrained observation correlation structure for commercial catches and irregular lattice AR(1) observation correlation structure for all surveys.
- (v) Unconstrained observation correlation structure for all fleets. $A_f(A_f - 1)/2$ parameters per fleet

Note that the models using AR(1) and IRAR(1) correlation structures are allowed to share parameters both among and within fleets (the latter only applicable for IRAR). These are chosen by performing exploratory runs where no parameters are shared to identify parameters with similar estimates. Since all parameters interact, these configurations may change between models for the same fleet, e.g. the IRAR(1) structures for the surveys may differ between models 3 and 4. The chosen configurations are included in the Supplementary material. The final model selection is carried out using AICc ([Hurvich and Tsai, 1989](#)), which is a small-sample-size bias-corrected version of AIC. Although the standard AICc is derived for univariate Gaussian linear models, we use it anyway as recommended by [Burnham and Anderson \(2002, p. 378\)](#), since we do not know a more exact correction term. The consequences of using the final model over Model 1 are illustrated by comparing point estimates of $\beta = (\log \bar{F}, \log \text{SSB})$ and, as a measure of forecasting performance, the total variability of the corresponding β -covariance matrices in the final data year and for a 3-year projection. The projection is obtained simply by running the model for 3 additional years without any data. Technically this is implemented by adding a single observation of catches of the youngest age group in the projection year. That observation is set to the same as the last observed value. The sensitivity of this procedure was tested by doubling the single observation of recruits, and it was confirmed that it had negligible influence on the results.

The total variability of a covariance matrix is calculated as the sum of conditional variances maximized over all permutations ([Mustonen, 1997](#)). For the stock status β this amounts to $\text{mvar}(\Sigma_\beta) = \max \{ \text{Var}(\log \bar{F}) + \text{Var}(\log \text{SSB} | \log \bar{F}), \text{Var}(\log \text{SSB}) + \text{Var}(\log \bar{F} | \log \text{SSB}) \}$. The conditional variances are computed as the squared diagonal elements of the Choleski decomposition of Σ_β . To ensure the stability of the parameter estimation, wide bounds were put on the elements of \mathbf{L} and the correlation coefficient ρ of the F random walk.

Residuals

For state-space models, diagnostics should be based on the one-step ahead (OSA) prediction errors, also known as recursive residuals ([Harvey, 1990](#)). The OSA residual for an observation is calculated by conditioning on previous data points (and fixed effect parameters). These residuals differ from those where one has conditioned on all the data, which should have a mean of zero, but not be serially independent unlike OSA residuals ([Harvey, 1990](#)). Let $Y_1 \dots Y_N$ be the combined vector of scalar observations sorted

Table 1. Model selection criteria.

Stock	Model	Neg. log Lik	No. par	AICc
Herring	1	739.95	16	1512.56
	2	685.98	19	1410.87
	3	654.21	23	1355.77
	4	586.68	65	1314.32
	5	572.70	86	1337.06
Whiting	1	367.01	17	769.03
	2	333.89	19	707.03
	3	310.67	19	660.59
	4	261.03	57	647.77
	5	234.68	78	648.01
Haddock	1	490.35	17	1015.66
	2	471.62	19	982.43
	3	451.43	21	946.30
	4	405.68	57	936.29
	5	381.47	78	939.06
Turbot	1	1030.40	27	2116.80
	2	985.41	30	2033.29
	3	984.77	32	2036.36
	4	925.79	76	2020.12
	5	892.17	114	2051.53

“Neg. logLik” is the negative log-likelihood and “No. par” is the number of parameters (excluding random effects). Best AICc values are shown in bold face.

by time, fleet, and age, then the residual r_i associated with the i th observation is given by

$$r_i = \frac{Y_i - \hat{Y}_{ij-1}}{\text{sd}(Y_{ij-1})}, \tag{8}$$

where $\hat{Y}_{ij-1} = E(Y_i|Y_{i-1}, Y_1)$ is the OSA prediction of the observation Y_i given $\{Y_1, \dots, Y_{i-1}\}$, and $\text{sd}(Y_{ij-1})$ is the standard deviation of this prediction (Harvey, 1990). Note that the residuals reported in Nielsen and Berg (2014) and Berg et al. (2014) were of the second kind due to computational convenience when using the AD Model Builder software (Fournier et al., 2012), where the posterior distributions of $Y_{i|N}$ are readily available, but the one-step conditionals Y_{ij-1} are not. The model used here is implemented in Template Model Builder (TMB, Kristensen et al. (2016)), which has provision for calculating OSA residuals (see the TMB documentation of the oneStepPredict function for further details). Also note that the interpretation of the residuals depends on the ordering of observations within a year, although they should be standard normal distributed regardless of the ordering. A weak Gaussian prior on the state vector in the first time-step is needed for the OSA residual calculations to work in the very beginning of the

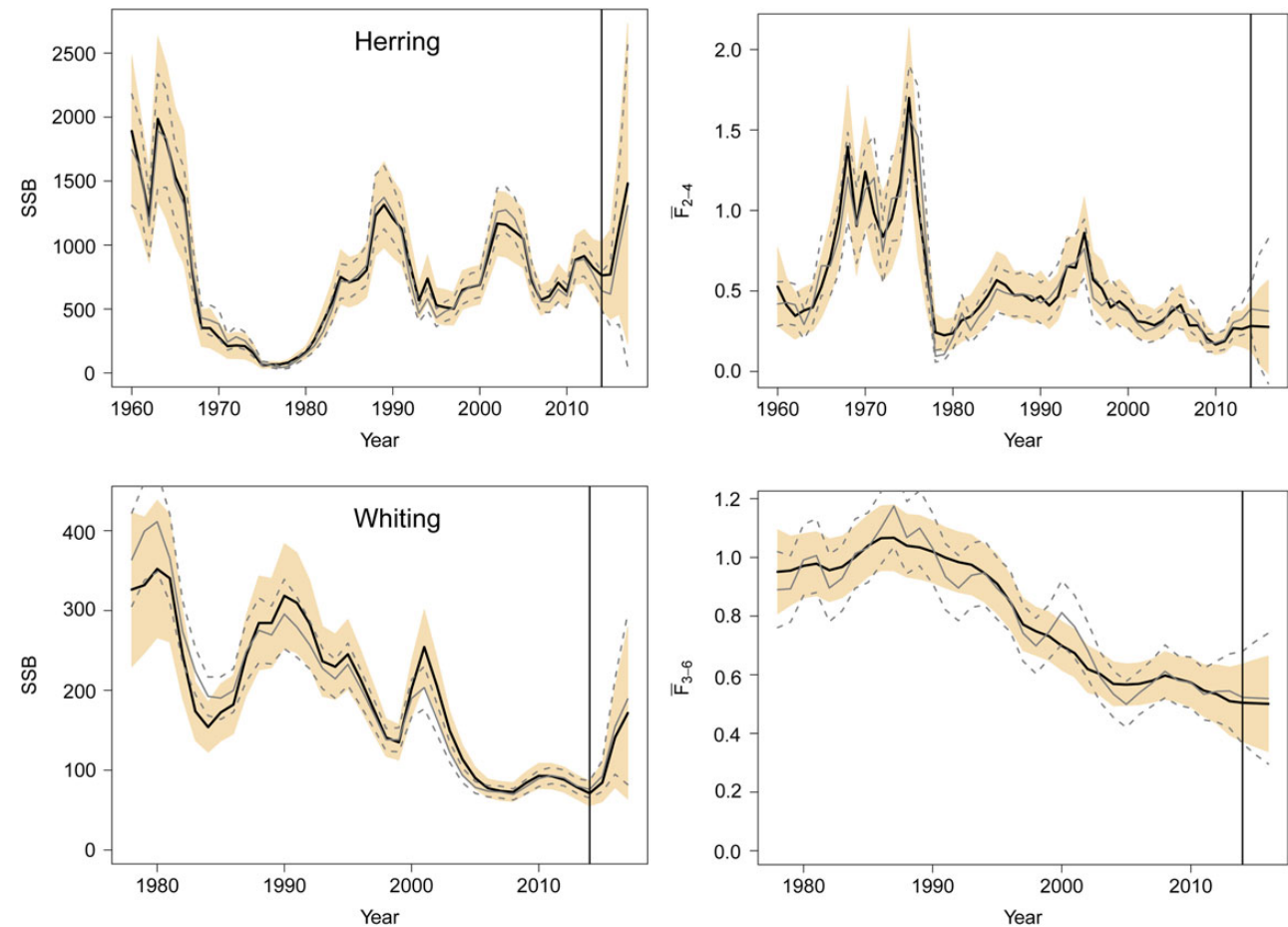


Figure 2. Estimated spawning-stock biomasses (left column) and average fishing mortalities herring, whiting, haddock, and turbot (from top to bottom row). Light solid lines represent results from models assuming independent observations (Model 1), whereas black lines come from the model with the best AICc (Model 4). The light dashed lines and shaded areas represent 95% marginal confidence intervals calculated from Model 1 and the best model, respectively. Vertical lines indicate the start of the projection period.

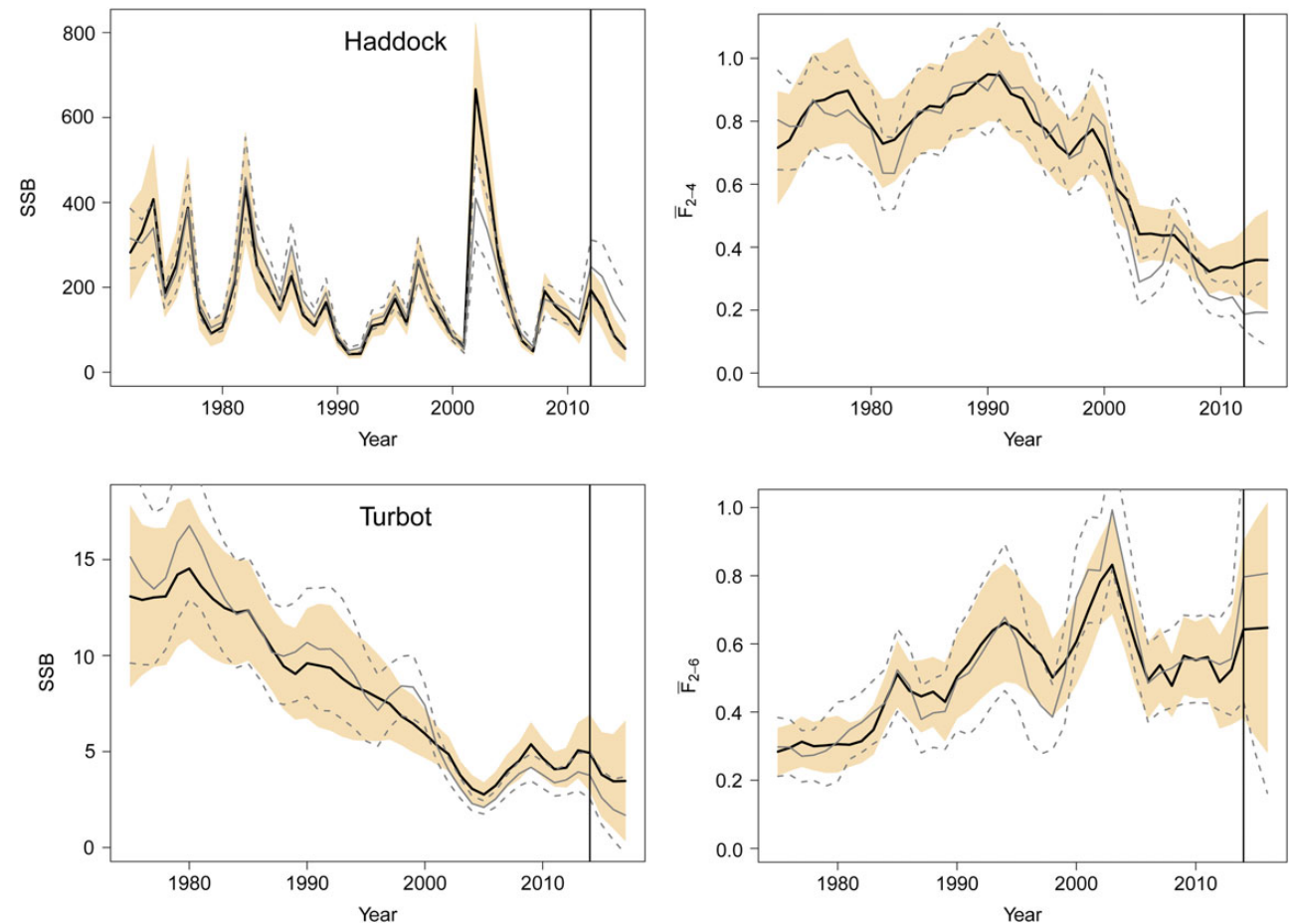


Figure 2. Continued.

Table 2. Change in process error parameters.

	$\sigma_F / \tilde{\sigma}_F$	$\sigma_R / \tilde{\sigma}_R$	$\sigma_S / \tilde{\sigma}_S$	$\rho / \tilde{\rho}$
Herring	0.77	0.90	0.51	1.01
Whiting	0.54	1.08	0.86	1.35
Haddock	0.65	1.12	0.63	1.00
Turbot	0.51	0.97	1.01	1.27

Estimates from Model 4 are divided by the corresponding estimate from Model 1, hence σ -ratios values less than one imply reduced process error in Model 4. For ρ -ratios it is reversed—values greater than one imply reduced variance.

dataset, but if the variances are chosen sufficiently large this has no bearing on the final results.

Case studies

The data used in this study consist of total catches and survey indices by age for haddock, herring, turbot, and whiting in the North Sea. These cases were chosen to facilitate comparison with the results of Berg et al. (2014) (herring and whiting), and because the numbers of age groups for these stocks are rather high (9, 10, 10, and 9 for haddock, herring, turbot, and whiting, respectively), such that one would expect to gain something from including correlations between age groups. Data were obtained from ICES assessment reports (ICES, 2014a,b), except survey indices for herring and whiting, which were calculated using the methodology described in Berg et al. (2014). The full source codes and also the

datasets are available online at www.stockassessment.org under the names “Haddock-corrObs”, “Herring-corrObs”, “Turbot-corrObs”, and “Whiting-corrObs”. The precise configuration of the observation covariance structure for each case is included in Supplementary material, while the additional assessment configurations, which are identical for all five models, can be found at www.stockassessment.org.

Simulation study

A small simulation study is performed to ensure correct implementation of the model and the ability to differentiate between datasets simulated with independent and correlated observations using AICc. The simulations are set up to resemble the North Sea whiting case. The efficiency of the model to reconstruct the true N 's and F 's is also cross-tabulated. This is quantified using root mean squared errors (RMSE) between true and estimated values of N and F . Details about the simulation study can be found the online Supplementary materials.

Results

The simulation study showed that the model is able to reconstruct the correct parameters and states from data. When data were truly independent, Model 1 was correctly selected in all 20 cases by the AICc criterion, and Model 5 was also correctly selected in 20 out of 20 cases when observations was truly correlated. The correct models were as expected the most effective at reconstructing the

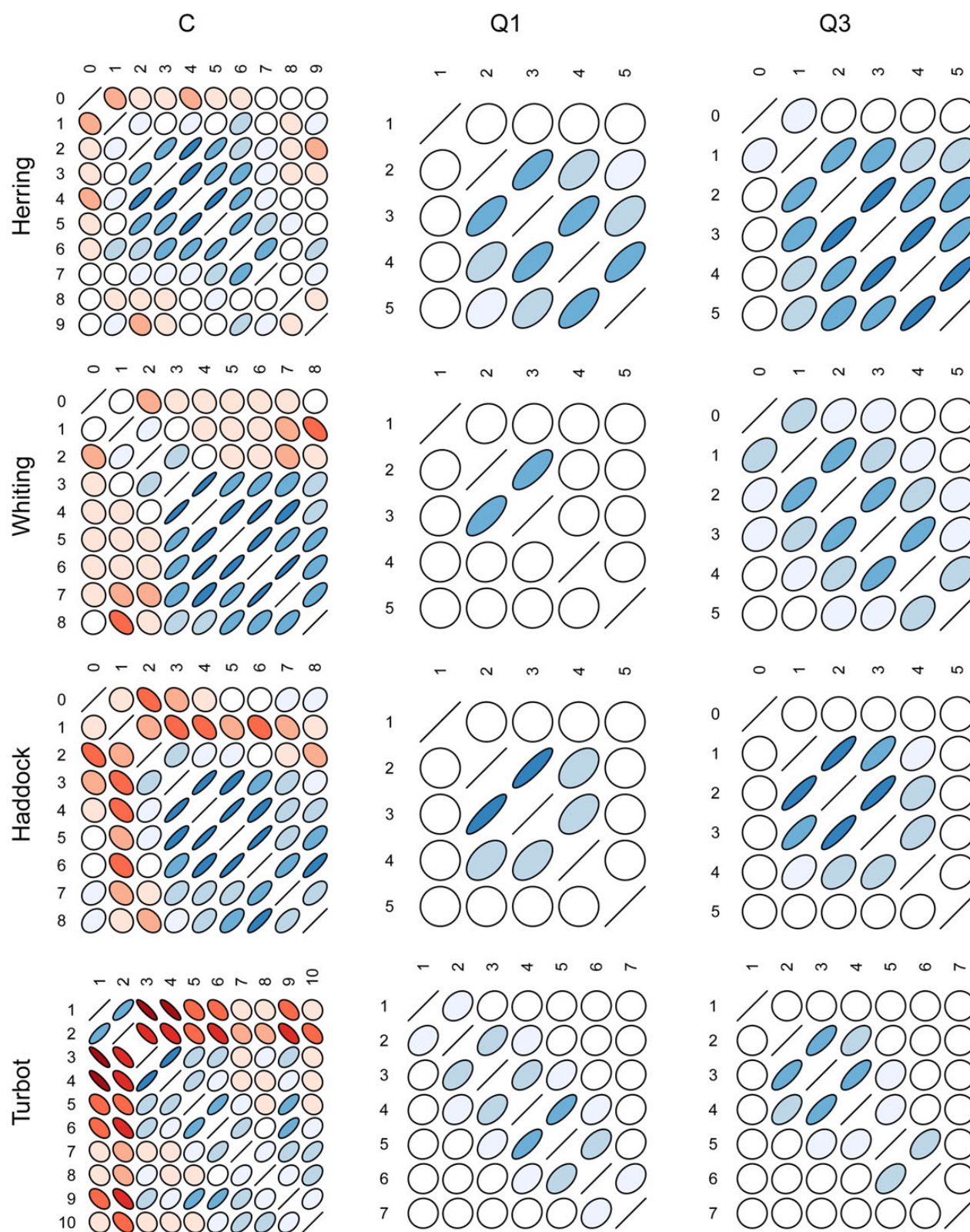


Figure 3. Estimated observation correlation matrices for herring, whiting, haddock, and turbot (from the top to bottom). All estimates are from the final models (Model 4). The columns represent from left to right: commercial catches, survey 1, and survey 2.

true states. However, wrongly assuming that observations are independent had worse consequences (22% RMSE increase) than estimating correlations in the independent case (10% RMSE increase).

In all the case studies the combination of IRAR(1) covariances for the surveys and unconstrained for the commercial catches (Model 4) had the best AICc (Table 1) although the purely

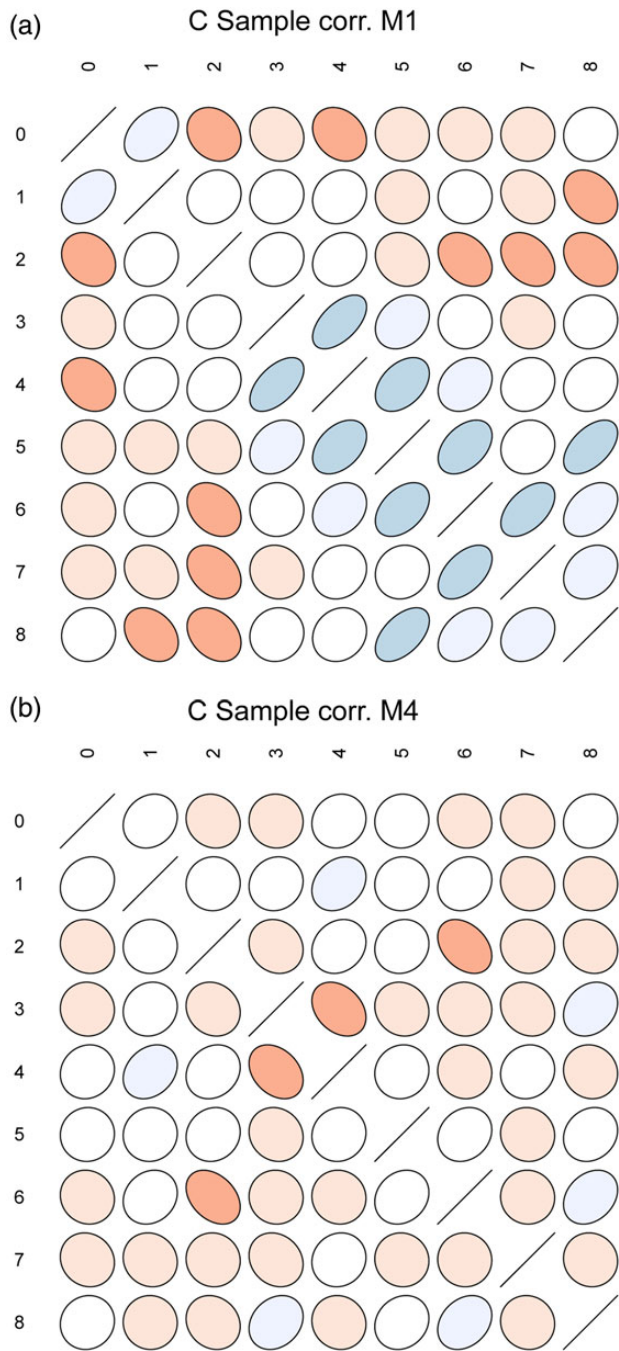


Figure 4. Whiting total catches: sample correlation between age groups of the one-step ahead residuals from Model 1 (A) and Model 4 (B). Note the resemblance between (A) and the estimated correlations in Figure 3.

unconstrained Model 5 is a close contender for whiting and haddock. There is virtually no support for the independent Model 1 compared with any of the alternative models ($\Delta\text{AICc} > 20$).

While the overall trends in F and SSB are similar between models, there are important differences in the point estimates of the stock status (\bar{F} , SSB) after the 3-year projection when changing from Model 1 to the best model—most substantially for haddock and turbot (Figures 2 and 5). The direction of change is both

towards higher abundance and lower fishing mortality (-54% SSB and $+87\%$ \bar{F} for haddock) and *vice versa* ($+106\%$ SSB and -20% \bar{F} for turbot).

The best models give smoother estimates of F compared with Model 1 (Figure 2), which must be expected when some of the observed covariation in the total catches are ascribed to the correlated error term ε_y^C rather than variability in the F -processes. This shift of variance from process error to observation error is most pronounced for whiting, where the best model is one where the F -processes are essentially reduced to a multiplicative model, that is they are completely correlated and there is effectively only one F -process and a constant selectivity (constant proportions of total F by age). This is in contrast to Model 1 for whiting, which reports substantial changes in selectivity over time to match the observed catches (Supplementary Figure S1). This effect is also evident from Table 2, where all F -process standard deviations are reduced when switching from Model 1 to Model 4. The survival process errors also reduced in all cases except in the Turbot case, where it is essentially the same. The recruitment process errors are not affected much by changing to Model 4.

The estimated correlation structures from the final models are illustrated in Figure 3, and for the surveys the pattern found is very similar to the patterns found by Berg *et al.* (2014) (compare the estimated survey correlations for herring in Figure 3 (top row, middle) with Figure 1); the youngest age group is nearly uncorrelated with the rest while fairly strong positive correlations can be found between the older age groups. This structure can be nicely captured by the IRAR(1) structure and explains why Model 4 is chosen over Model 5. For the commercial catches, on the other hand, the estimated structure is more complex and consists of a mixture of negative and positive correlations, which cannot be captured by the IRAR(1) structure. Negative correlations are predominantly found among young age groups, whereas the strong positive correlations are found among adult age groups, though with a tendency to decrease among the very oldest age groups.

Neither the residuals for Model 1 nor for the best model are without patterns, which indicate that further data and model scrutiny would be in order; however, the latter appear to have fewer years where all residuals have the same sign across ages (Supplementary Figures S5–S12). This is also evident from Figure 4, where an example (whiting, total catches) of the sample correlation of one-step ahead residuals between ages are shown for Models 1 (panel A) and 4 (panel B). The residuals from Model 1 are clearly not uncorrelated between ages as it is assumed in this model, and although the model can somewhat compensate by adjusting N and F , the patterns are similar to the estimated correlations in Model 4 (Figure 3). In contrast, the between-age sample correlations of the residuals from Model 4 are close to zero as expected. The complete set of sample correlation matrices of residuals for Model 1 and the final models can be found in Supplementary material.

For all stocks, the total uncertainty on the stock status after a 3-year projection (see Figure 5) is smaller when using the final model compared with Model 1, while in the last data year this is reversed (except for turbot). This indicates more precise predictions from the final model, which describe the data with less process noise, while relatively more of the variation is described by the observation error component compared with the independent observation model.

Discussion

The objective of this study was to extend the SAM assessment model from assuming independent observations by age to allowing

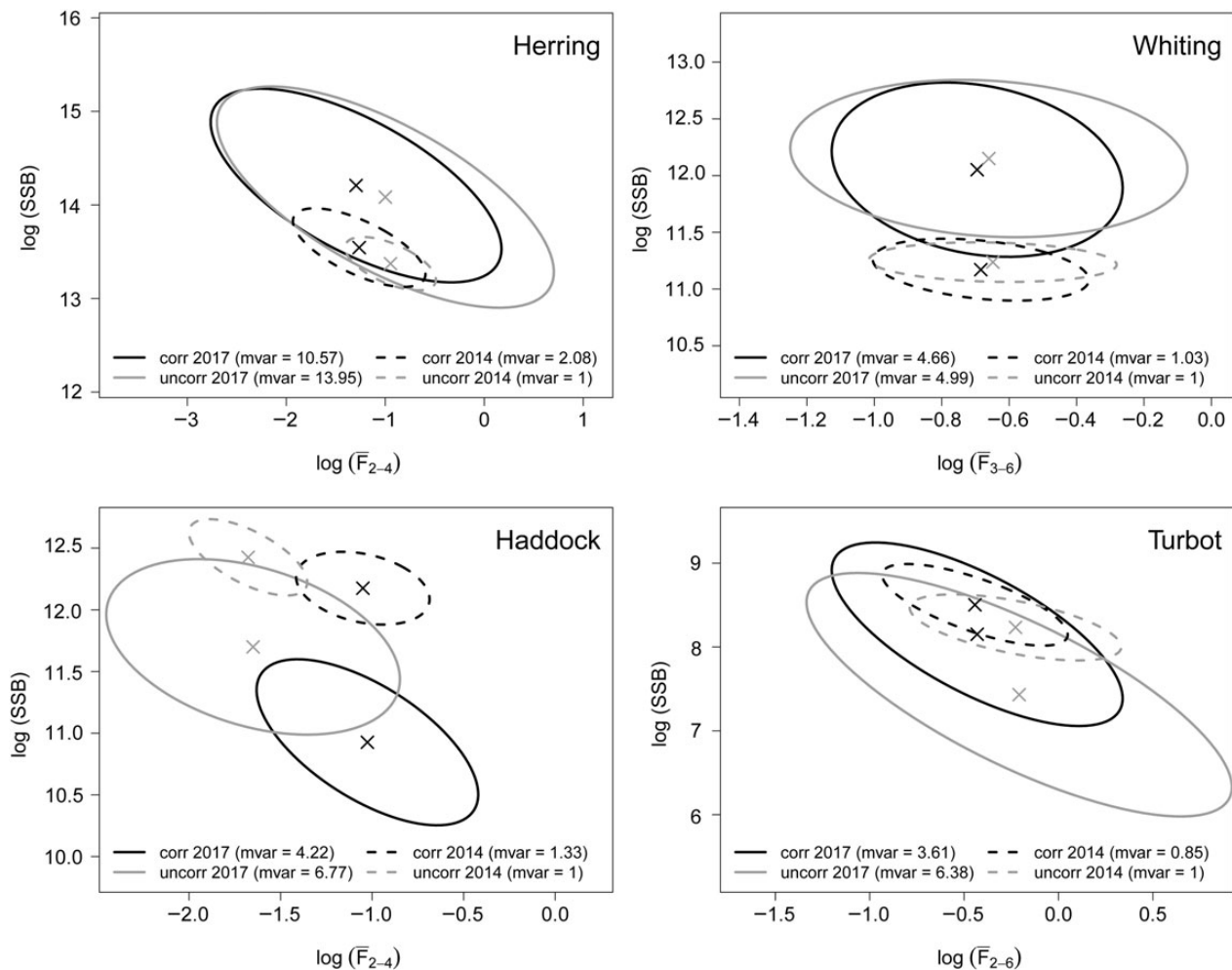


Figure 5. 95% contour ellipses for the joint distribution of $\log \bar{F}$ and $\log(\text{SSB})$ in the last data year (2014, dashed) and after a 3-year projection (2017, solid) for herring (top left), whiting (top right), haddock (bottom left), and turbot (bottom right). Light lines represent results from models assuming independent observations (Model 1), whereas black lines come from the model with the best AICc (Model 4). Note that the total uncertainties (mvar) of the projection from the best models are smaller than those from Model 1 in all cases.

correlations between ages with varying complexity of the correlation structure: AR(1), IRAR(1), unconstrained, and mixtures. In all cases, the assumption of independence was firmly rejected as was the simple AR(1) correlation structure. The IRAR(1) structure was sufficiently flexible to capture the correlation structure in the surveys. This is in line with the results found in Berg *et al.* (2014), who used haul-by-haul survey data and non-parametric methods to estimate year-specific correlation structures. For the commercial fleets, the IRAR(1) failed to account for the substantial negative correlations found between younger age classes when the unconstrained correlation structure was used.

The earliest versions of SAM assumed independence between ages in the F - and N -processes as well as in the observation equations. This is still a useful model, because it imposes very little structure on the processes involved, such that any observed patterns can be used to formulate a more structured model with improved forecasting abilities. Nielsen and Berg (2014) improved on this model by introducing an AR(1) correlation structure for the F -process, and Berg *et al.* (2014) considered correlated observation errors for the surveys, where the correlation structure was estimated with

non-parametric methods from haul-by-haul data, rather than from aggregated catch-at-age data as in this study. Earlier work (Myers and Cadigan, 1995) had already demonstrated that survey data are not independent among ages for a given year. They used a model that considered catches to be known without error and had a simple correlation structure with equal correlation between all age groups within a year (compound symmetry). This study has confirmed that substantial correlations between ages must be expected (Francis, 2014), not only for surveys, but also for the commercial fleets, and that the assessment output may be seriously affected by assumptions of independence. This occurred, for example, in the turbot case where accounting for correlations led to more than a doubling of the estimated spawning-stock biomass in the projection year. In addition, simple correlation structures such as AR(1) or compound symmetry were found to be inadequate. Our simulation study showed that there is a small increase in root-mean-squared error from estimating correlations when these are truly zero, but ignoring the correlations when they are present is more critical. A similar result was found by Myers and Cadigan (1995).

We still have independent N -process errors (equations 1–3), which is probably not optimal. Cadigan (2015) estimates in a model where process error is defined as deviations from a constant natural mortality (M) with autocorrelation in the time and age directions. High correlations in the M -process are found for northern cod. Since the M -process is governed by the environment, which is naturally autocorrelated in time and space, this is a natural extension to consider, and may in fact explain some of the same co-variation over ages that in this study is ascribed to the observation error. On the other hand, it is most likely that correlations exist in both observations as well as the F and survival processes. Further studies should investigate the identifiability of a model that include correlations in all three, and the consequences of model misspecification in this respect.

We did not consider correlated observation errors in the time direction; only correlations between ages were considered. Correlations in time could arise naturally in surveys where different survey vessels with different catchabilities are covering different periods of time. However, when aggregated data are used, it is unlikely that it is possible to estimate different unconstrained covariances over many time-blocks due to the large number of parameters required for each block. In addition, introducing correlations across time will seriously affect the estimation speed, as this breaks the sparseness of the Hessian matrix, which is heavily exploited in TMB (Kristensen *et al.*, 2016). Using haul-by-haul data to increase sample size may be the only feasible way to model such effects. While Berg *et al.* (2014) found that the correlation patterns in the IBTS surveys are rather constant over time, our use of a time-constant correlation structure should be viewed as an approximation to facilitate analysis and may not be optimal in case of large changes in fishing practices or sampling procedures.

While we considered correlation structures for age-structured data, there is no reason why the same methodology cannot be used with length-structured data as well, in fact, the independence assumption is probably even more critical for length data, since there are typically many more length groups than age groups. The IRAR(1) structure of the correlation matrix that we propose here might also be applicable for lengths, although the IRAR(1) parameterization cannot capture the gradual change from negative to positive correlations that was found for several fleets in this study. Further work is needed to find a parsimonious parameterization of the covariance matrices that fits the observed patterns. An approach similar to that of Nielsen *et al.* (2014) where covariance matrices are partitioned into parametric and unconstrained blocks might be useful considering the amount of structure observed in our estimated covariance matrices. Miller and Skalski (2006) also found strong patterns in correlation matrices for length data, but less so for age data.

This study has underlined previous findings that catch-at-age data should not be considered independent, although this is still a very common assumption. At least such assumptions should be validated through inspection of residuals. While it may not always be possible to estimate the fully unconstrained covariance matrices for all fleets, it will most likely be possible to estimate one of the simpler alternatives with fewer parameters.

The source code for the model is made publicly available and since it does not require extra data, it can easily be applied to all existing SAM assessments, or other models using the TMB software (Kristensen *et al.*, 2016). Given our results we expect that better assessments with more robust forecast will result from this, especially if the number of age or size groups is large.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

Acknowledgements

The authors thank Daniel Goethel, Chris Legault, and one anonymous reviewer for valuable inputs to this manuscript. We also thank Kasper Kristensen for his assistance with the implementation of the model in TMB, and Christoffer M. Albertsen for useful discussions and comments on a draft version of the manuscript.

References

- Berg, C. W., Nielsen, A., and Kristensen, K. 2014. Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. *Fisheries Research*, 151: 91–99.
- Burnham, K. P., and Anderson, D. R. 2002. Model selection and multi-model inference: a practical information-theoretic approach. Springer Verlag, New York. 488 pp.
- Cadigan, N. G. 2015. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. *Canadian Journal of Fisheries and Aquatic Sciences*, doi:10.1139/cjfas-2015-0047.
- Fournier, D. A., Skaug, H. J., Ancheta, J., Sibert, J., Ianelli, J., Magnusson, A., Maunder, M. N., *et al.* 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27: 233–249.
- Francis, R. C. 2014. Replacing the multinomial in stock assessment models: a first step. *Fisheries Research*, 151: 70–84.
- Francis, R. C., and Hilborn, R. 2011. Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, 68: 1124–1138.
- Harvey, A. C. 1990. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge. 555 pp.
- Hrafnkelsson, B., and Stefánsson, G. 2004. A model for categorical length data from groundfish surveys. *Canadian Journal of Fisheries and Aquatic Sciences*, 61: 1135–1142.
- Hurvich, C. M., and Tsai, C.-L. 1989. Regression and time series model selection in small samples. *Biometrika*, 76: 297–307.
- ICES. 2014a. Report of the Herring Assessment Working Group for the Area South of 62°N (HAWG), 11–20 March 2014, ICES HQ, Copenhagen, Denmark. ICES Document CM 2014/ACOM:06. 1257 pp.
- ICES. 2014b. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK), 30 April–7 May 2014, ICES HQ, Copenhagen, Denmark. ICES Document CM 2014/ACOM:13. 1493 pp.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. J., and Bell, B. 2016. TMB: automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70: 1–21.
- Maunder, M. N., and Punt, A. E. 2013. A review of integrated analysis in fisheries stock assessment. *Fisheries Research*, 142: 61–74.
- McDonald, A. D., Kendrick, T. H., and Breen, P. A. 2001. Empirical weighting of multiple stock-abundance indices for parameter estimation and stock assessment in a multi-zone or multi-species fishery. *ICES Journal of Marine Science*, 58: 204–215.
- Methot, R. D., and Wetzel, C. R. 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, 142: 86–99.
- Miller, T. J., and Skalski, J. R. 2006. Integrating design-and model-based inference to estimate length and age composition in north pacific longline catches. *Canadian Journal of Fisheries and Aquatic Sciences*, 63: 1092–1114.

- Mustonen, S. 1997. A measure for total variability in multivariate normal distribution. *Computational Statistics & Data Analysis*, 23: 321–334.
- Myers, R. A., and Cadigan, N. G. 1995. Statistical analysis of catch-at-age data with correlated errors. *Canadian Journal of Fisheries and Aquatic Sciences*, 52: 1265–1273.
- Nielsen, A., and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*, 158: 96–101.
- Nielsen, J. R., Kristensen, K., Lewy, P., and Bastardie, F. 2014. A statistical model for estimation of fish density including correlation in size, space, time and between species from research survey data. *PLoS One*, 9: e99151.
- Pennington, M., and Volstad, J. H. 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. *Biometrics*, 50: 725–732.
- Thorson, J. T. 2014. Standardizing compositional data for stock assessment. *ICES Journal of Marine Science: Journal du Conseil*, 71: 1117–1128.

Handling editor: Shijie Zhou