

# Choosing the observational likelihood in state-space stock assessment models

Christoffer Moesgaard Albertsen, Anders Nielsen, and Uffe Høgsbro Thygesen

**Abstract:** Data used in stock assessment models result from combinations of biological, ecological, fishery, and sampling processes. Since different types of errors propagate through these processes, it can be difficult to identify a particular family of distributions for modelling errors on observations a priori. By implementing several observational likelihoods, modelling both numbers- and proportions-at-age, in an age-based state-space stock assessment model, we compare the model fit for each choice of likelihood along with the implications for spawning stock biomass and mean fishing mortality. We propose using AIC intervals based on fitting the full observational model for comparing different observational likelihoods. Using data from four stocks, we show that the model fit is improved by modelling the correlation of observations within years. However, the best choice of observational likelihood differs for different stocks, and the choice is important for the short-term conclusions drawn from the assessment model; in particular, the choice can influence total allowable catch advice based on reference points.

**Résumé :** Les données utilisées dans les modèles d'évaluation des stocks découlent de combinaisons de processus biologiques, écologiques et associés à la pêche et l'échantillonnage. Puisque différents types d'erreurs sont propagés par ces processus, il peut être difficile de cerner une famille précise de distributions pour modéliser les erreurs sur les observations a priori. En appliquant plusieurs probabilités observationnelles et en modélisant tant les nombres que les proportions selon l'âge, dans un modèle d'évaluation du stock d'espace d'états basé sur l'âge, nous comparons la fidélité du modèle pour chaque choix de probabilité et ce qu'il signifie en ce qui concerne la biomasse de géniteurs du stock et la mortalité par pêche moyenne. Nous proposons d'utiliser des intervalles de CIA basés sur le calage du modèle observationnel intégral pour comparer différentes probabilités observationnelles. En utilisant des données de quatre stocks, nous démontrons que la modélisation de la corrélation des observations au sein de mêmes années améliore la fidélité du modèle. Cependant, le meilleur choix de probabilité observationnelle varie selon le stock, et ce choix est important pour les conclusions à court terme tirées du modèle d'évaluation; le choix peut, en particulier, influencer les avis concernant le total autorisé des captures basés sur des points de référence. [Traduit par la Rédaction]

## Introduction

Stock assessment models often use aggregated and uncertain data such as surveys and landings-at-age, which rely on age classification of effectively few individuals (Aanes and Pennington 2003). Commercial fishing and scientific surveys sample from populations that vary according to, for example, season, sex, age, and region. From this catch, samples are weighed and measured to estimate the length distribution, mass-at-length, and total catch in numbers. Additional subsamples are taken to age-classify individuals for estimating proportions-at-age, either directly or through an age-length key. The samples consist of many individuals from few hauls (Aanes and Pennington 2003), which may lead to underestimated uncertainties of estimates if ignored. Finally, all this information is aggregated to numbers-at-age for each year. This aggregation may be via models including, for example, spatial location, season, gear, and length effects. Even though the stock population growth process at this level of aggregation is well described (each year the fish age by 1 year, some die of natural causes, and others die from fishing), aggregating the different sources of uncertainty makes it difficult to find the optimal (or true) distributions of the observations a priori.

Age-based stock assessment models can be divided into two classes depending on the way they utilize the data. Either the data

can be modelled as numbers-at-age or as proportions-at-age along with total mass or numbers. Most currently used age-based stock assessment models exclusively consider either numbers- or proportions-at-age and only one or few observational likelihoods (ICES 2010a). When modelling numbers-at-age, the normal distribution, parameterized to avoid too much probability on negative observations, has been used (Gudmundsson 1994; Fryer 2002) along with the lognormal distribution (Cook 2013; Nielsen and Berg 2014) and its multivariate extension (Myers and Cadigan 1995). Although recommended over the lognormal by Cadigan and Myers (2001), the gamma distribution is infrequently used to model numbers-at-age in assessment models (ICES 2010a).

The multinomial distribution has been popular when modelling proportions-at-age (Fournier and Archibald 1982; Methot and Wetzel 2013; Williams and Shertzer 2015). Based on the age classification sampling, it is an intuitive choice; however, when using the true number of data-generating samples, the variances of the modelled proportions are often too small and the correlation structure too restrictive (Crone and Sampson 1998; Aanes and Pennington 2003; Francis 2014). Efforts have been made to increase the variance by estimating an effective sample size (McAllister and Ianelli 1997; Francis 2011; Hulson et al. 2011, 2012). Nonetheless, the effective sample size must be estimated by iterative optimization (McAllister and Ianelli 1997; Francis 2011; Maunder 2011), since the

Received 18 November 2015. Accepted 19 September 2016.

C.M. Albertsen, A. Nielsen, and U.H. Thygesen. Technical University of Denmark, National Institute of Aquatic Resources, Charlottenlund Castle, DK-2920 Charlottenlund, Denmark.

**Corresponding author:** Christoffer Moesgaard Albertsen (email: [cmoe@aqua.dtu.dk](mailto:cmoe@aqua.dtu.dk)).

Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from [RightsLink](http://RightsLink.com).

**Table 1.** Overview of the observational models used in the case studies and various properties.

Model	Distribution	Class	Allows 0	Baranov	Est. params.	Est. cor.
$M_1$	lognormal	UN@ A	No	Median	1 a f*	No
$M_2$	Gamma	UN@ A	Some	Mean	1 a f	No
$M_3$	Generalized gamma	UN@ A	Some	Location	2 a f	No
$M_4$	Normal	UN@ A	Yes	Mean	1 a f	No
$M_5$	Left truncated normal	UN@ A	Yes	Location	1 a f	No
$M_6$	log-Student's t	UN@ A	No	Location	2 a f	No
$M_7$	Multivariate lognormal	MN@ A	No	Median	1 a f + 1 f†	Yes
$M_8$	Additive logistic normal	P@ AwN	No	Location	1 a f + 1 f	Yes
$M_9$	Multiplicative logistic normal	P@ AwN	No	Location	1 a f + 1 f	Yes
$M_{10}$	Dirichlet	P@ AwN	No	Mean	1 f	No
$M_{11}$	Additive logisitic normal	P@ AwM	No	Location	1 a f + 1 f	Yes
$M_{12}$	Multiplicative logistic normal	P@ AwM	No	Location	1 a f + 1 f	Yes
$M_{13}$	Dirichlet	P@ AwM	No	Mean	1 f	No

**Note:** Model properties are as follows: if zero observations are allowed; whether the Baranov catch equation determines the mean, median, or location; the number of estimated observational parameters per age (a) and fleet (f); and whether a correlation parameter is estimated. The models are divided into model classes: univariate numbers-at-age (UN@ A), multivariate numbers-at-age (MN@ A), proportions-at-age with lognormal total numbers (P@ AwN), and proportions-at-age with lognormal total mass (P@ AwM).

\*Should be read as follows: One per age per fleet.

†Should be read as follows: One per age per fleet and one additional per fleet.

multinomial distribution is improper when used for continuous data (Francis 2014). Hence, the multinomial distribution will not be considered here. To avoid iterative estimation of the effective sample size, it has been suggested to replace the multinomial with the Dirichlet distribution (Williams and Quinn 1998; Francis 2014) in which the variance is only determined by parameters.

While the Dirichlet distribution is an improvement over the multinomial distribution, they both have a very restrictive variance-covariance structure that only allows negative correlations, which may not be appropriate (Francis 2014). Therefore, distributions based on transformations of multivariate normals, such as the additive logistic normal (Francis 2014) and the multiplicative logistic normal (Cadigan 2016), have recently been proposed for proportions-at-age in stock assessment models.

Although several authors have compared different proportions-at-age models (Maunder 2011; Francis 2014), not much effort has been given to compare different observational likelihoods for numbers-at-age data (Cadigan and Myers 2001) and even less has been given to compare between the proportions- and numbers-at-age. Using the R package Template Model Builder (Kristensen et al. 2016), we implement 13 observational likelihoods, including both numbers- and proportions-at-age models, in an age-based state-space stock assessment model. Using assessment data from four European stocks, we compare the model fit for each choice of likelihood along with the implications for key outputs such as spawning stock biomass (SSB) and mean fishing mortality ( $\bar{F}$ ).

## Methods

We implemented age-based state-space stock assessment models (Nielsen and Berg 2014) with 13 different observational likelihoods (Table 1) for four different European stocks. For simplicity, the same observational likelihood was used for both commercial catch data and survey indices. While the process model was kept unchanged for each stock, we compared the goodness-of-fit of the observational likelihoods by Akaike information criterion (AIC). We considered models for numbers-at-age and proportions-at-age combined with total catch. We considered seven different distributions for numbers-at-age. When using data in the form of a total and proportions-at-age, we followed the widespread convention of modelling total catch as univariate lognormal, but considered two alternatives where the total was either in numbers or biomass. These two alternatives for total catch were crossed with three alternative distributions for the proportions. The observational likelihoods implemented cover frequently used distributions in fisheries stock assessments and close extensions.

**Table 2.** Overview of the data sources used in the case study.

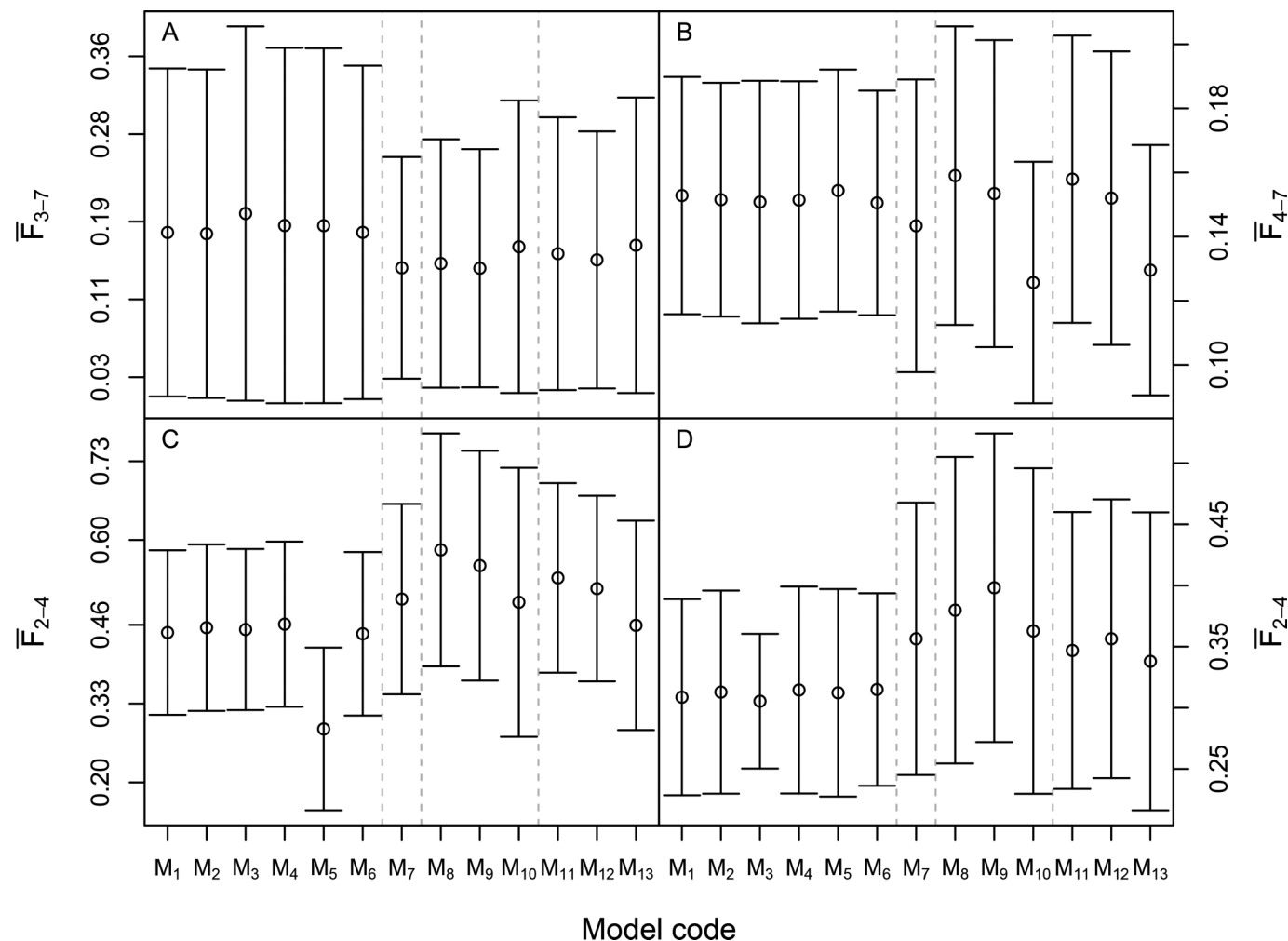
Fleet	First year	Last year	First age	Last age	Years with missing
<b>Blue whiting</b> (process model parameters = 7)					
Commercial	1981	2013	1	10	
Survey Q1	2004	2014	3	8	2010
<b>Northeast Arctic haddock</b> (process model parameters = 30)					
Commercial	1950	2013	3	11	
Survey Q4	1991	2013	3	7	
Survey Q1	1992	2013	3	7	1992–1995, 2000–2002, 2004
Survey Q1	1992	2013	3	8	1994, 1995, 2001
Survey Q3	2004	2013	3	8	2005
<b>North Sea cod</b> (process model parameters = 11)					
Commercial	1950	2011	1	7	
Survey Q1	1983	2012	1	5	
<b>Northern Shelf haddock</b> (process model parameters = 33)					
Commercial	1963	2011	0	8	
Survey Q3	1977	1991	0	6	
Survey Q3	1992	2011	0	6	
Survey Q3	1982	1997	0	6	
Survey Q3	1998	2011	0	6	
Survey Q1	1982	2011	0	4	

**Note:** Q1–Q4 indicates at which quarter of the year the survey is conducted.

## Process model

The processes described in the state-space model involved the true unobserved numbers-at-age in the stock and the true unobserved fishing mortality (see Nielsen and Berg 2014 or Appendix A for details). Following Nielsen and Berg (2014), the logarithm of the fishing mortality was assumed to follow a multivariate Gaussian random walk, where the correlation had an AR(1) structure (model D in Nielsen and Berg 2014). The true population numbers-at-age were assumed to follow an exponential decay model where the natural mortality is known. The model included recruitment to the first age group. The error terms for the true numbers-at-age were assumed to follow a lognormal distribution without correlation. All variance, correlation, and stock-recruitment parameters were estimated. Quantities such as mass-at-age and maturity were assumed to be known. The process model was related to the observations through the Baranov catch equation for catch data and through an assumption of proportionality to abundance-at-age for surveys. The proportionality constants were estimated. We denoted the calculated catch (or survey index) by  $\hat{C}_{a,f,y}$ .

**Fig. 1.** Last year fishing mortalities with 95% confidence intervals for models  $M_1$  to  $M_{13}$  (Table 1) in the case studies: blue whiting (A), Northeast Arctic haddock (B), North Sea cod (C), and Northern Shelf haddock (D). Vertical dashed grey lines separate the models into model classes (Table 1). Subscripts to  $\bar{F}$  indicates the ages the mean is over. All ages are weighed equally in the mean.

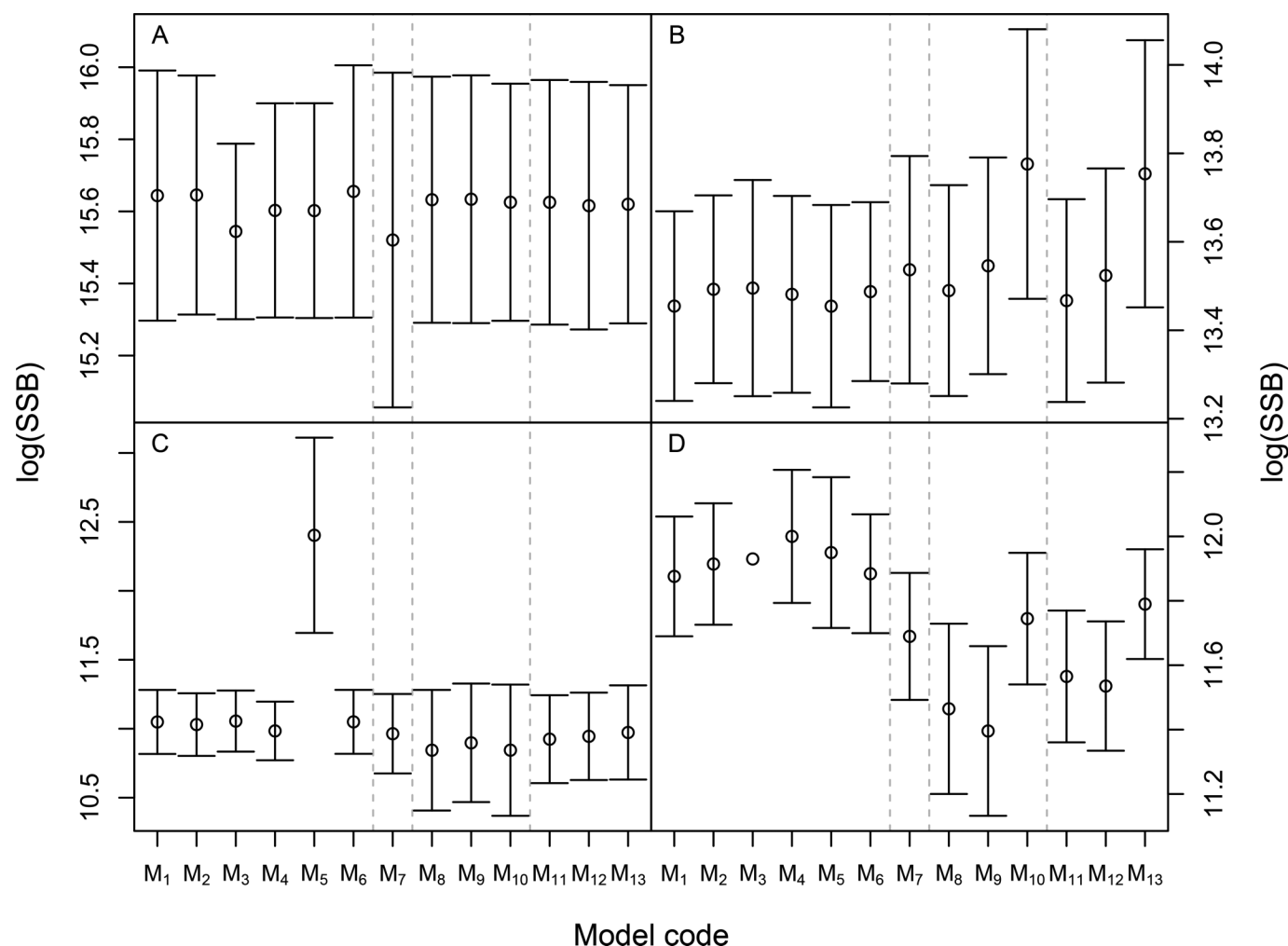


### Observational models

Our model  $M_1$  was the lognormal distribution with its usual parameterization. The median was determined by  $\hat{C}_{a,f,y}$ , while a scale parameter was estimated for each age and fleet. The model  $M_2$  was the gamma distribution parameterized to have constant coefficient of variation (Cadigan and Myers 2001). The mean was determined by  $\hat{C}_{a,f,y}$ , while a coefficient of variation (CV) was estimated for each age and fleet. The generalized gamma distribution was included as model  $M_3$  with the parameterization of Prentice (1974). This parameterization was preferred over the Stacy (1962) parameterization, as it both extends it and is numerically more stable when reducing to the lognormal distribution (Prentice 1974; Farewell and Prentice 1977). The log-location parameter was determined by  $\log(\hat{C}_{a,f,y})$ , while a shape and scale parameter was estimated for each age and fleet. The models  $M_4$  and  $M_5$  were the normal and truncated normal (with left truncation at zero), respectively. Both were parameterized based on the mean determined by  $\hat{C}_{a,f,y}$  and separate CV parameters for each age and fleet (which applied to the untruncated values for the truncated normal). The Student's  $t$  distribution on log-scale was our model  $M_6$ . The distribution was parameterized with a log-location parameter determined by  $\log(\hat{C}_{a,f,y})$  along with log-scale and log-degrees-of-freedom parameters estimated separately for each age and fleet.

Model  $M_7$  was the multivariate lognormal with its usual parameterization. The marginal medians were determined by  $\hat{C}_{a,f,y}$ , while a one-parameter AR(1) structure was used for the correlation between ages on logarithmic scale (Pinheiro and Bates 2000; Francis 2014). Separate correlation parameters were estimated for each fleet along with scale parameters estimated for each age and fleet. Models  $M_8$  and  $M_9$  were the additive logistic normal and the multiplicative logistic normal (Aitchison 2003), respectively, with lognormal total numbers. For both models, the location parameters were determined by  $\hat{C}_{a,f,y}$ , while the scale matrices were parameterized as  $M_7$ . For the lognormal distributions, the medians were determined by  $\sum_a \hat{C}_{a,f,y}$ , while a separate scale parameter was estimated for each fleet. The Dirichlet distribution with lognormal total numbers was our model  $M_{10}$ . The Dirichlet distribution was parameterized with concentration parameters proportional to  $(\hat{C}_{1,f,y}, \dots, \hat{C}_{A,f,y})^T$ . A proportionality parameter was estimated for each fleet. The lognormal distributions for total numbers were parameterized as  $M_8$  and  $M_9$ . Finally, models  $M_{11}$ ,  $M_{12}$ , and  $M_{13}$  were the additive logistic normal, multiplicative logistic normal, and Dirichlet with log-normal total mass parameterized as  $M_8$ ,  $M_9$ , and  $M_{10}$ , respectively. All estimated observational parameters were assumed to be constant over years. The densities and further details can be seen in Appendix A.

Fig. 2. Last year logarithmic spawning stock biomass ( $\log(\text{SSB})$ ) for models  $M_1$  to  $M_{13}$  (Table 1) in the case studies: blue whiting (A), Northeast Arctic haddock (B), North Sea cod (C), and Northern Shelf haddock (D). Vertical dashed grey lines separate the models into model classes (Table 1).



### Comparing by AIC

To compare the different observational models, we employed the AIC (Akaike 1974). However, the AIC applies to comparison between specific models, whereas each observational model represents an entire family of models, differing in assumed relationships among parameters for different age groups. These families include “full models” where each age group and fleet are assigned independent parameters, a “minimal model” where all age groups share common parameters, as well as a range of models between these two extremes. A standard application of the AIC would require that the optimal model in each family is identified, a task that would involve estimation of parameters in thousands of models.

To avoid this step, which is computationally very demanding and tangential to our purpose, we chose to identify an AIC interval that characterized each model family. This AIC interval gave an upper and a lower bound on the optimal AIC within that family. The upper bound of the interval was attained by the AIC for the full model. The lower bound of the interval was calculated as the AIC that would hypothetically be obtained by the smallest possible nested submodel if the negative log-likelihood would not increase compared with the full model. The difference between the upper and lower bound is thus twice the difference in the number of parameters between the full model and the minimal model.

A model family was considered clearly superior to another if the upper bound of its AIC interval was below the lower bound of the other model family’s interval (i.e., the other model family had a

higher interval). Clearly inferior model families could be discarded. To compare the remaining model families, it would be possible to narrow the AIC intervals through testing within each model family, but for simplicity we base the comparison on the full model in each family.

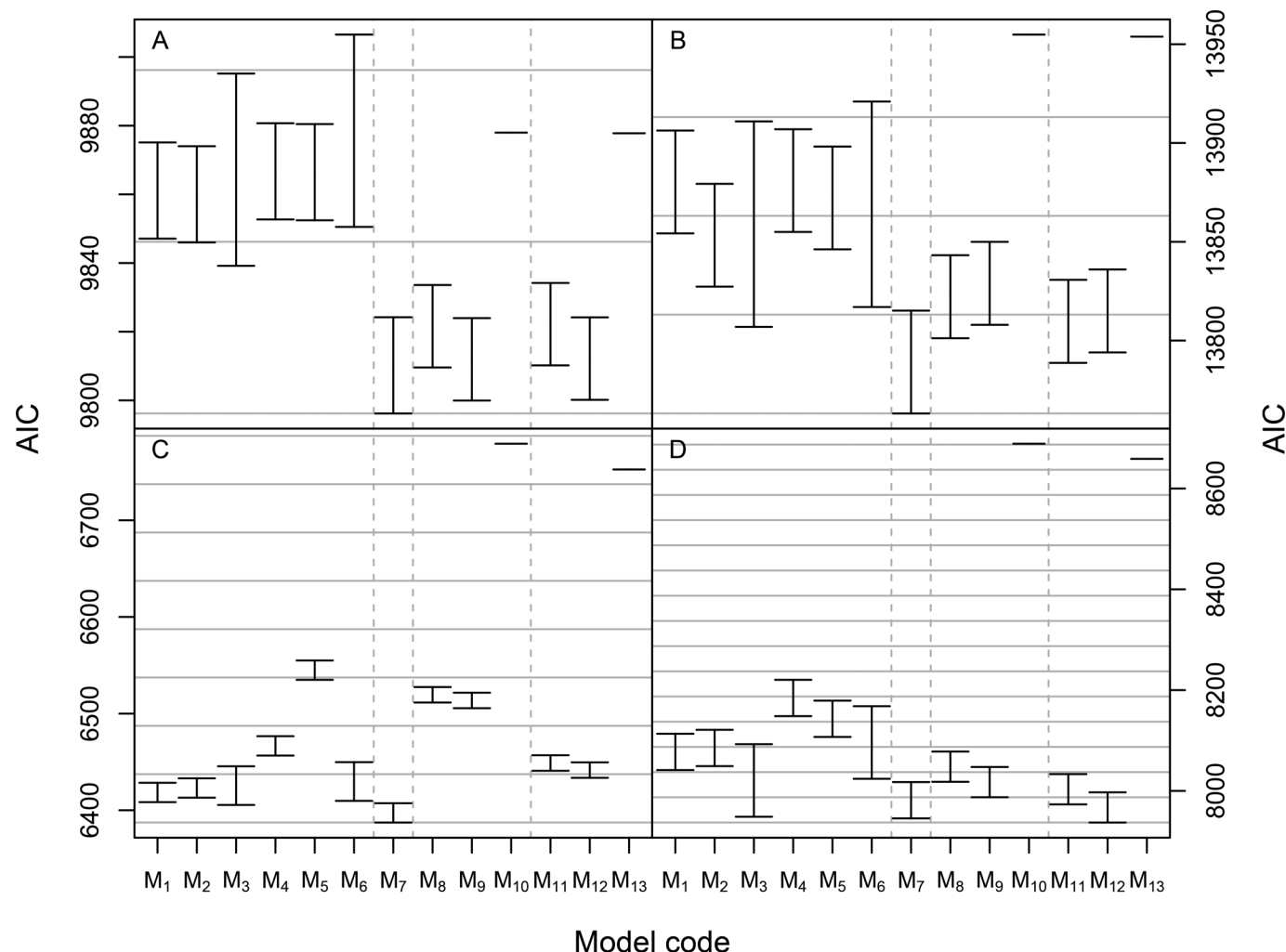
Using AIC to compare the models required that the models were defined on the same data, which was not the case when we compared between numbers-at-age models and proportions-at-age models. The proportions-at-age data were, however, a one-to-one transformation of the numbers-at-age data. Thus, using a standard transformation of densities we derived the log-likelihood for the numbers-at-age data that is consistent with our specified distributions based on proportions and totals (Appendix B). Using the transformed likelihood in the AIC calculation allowed for valid comparisons of models using numbers-at-age directly versus those using totals and proportions-at-age. Note that a similar transformation was required so that models that used total mass could be compared with models that used total numbers with the proportions.

### Case study

We implemented the models for four different data sets used for assessments (Table 2). The blue whiting (*Micromesistius poutassou*) data set was the basis of the 2014 ICES advice (ICES 2014a) for Subareas I–IX and XIV; the Northeast Arctic haddock (*Melanogrammus aeglefinus*) data was used for the 2014 ICES advice (ICES 2014b) for



**Fig. 3.** AIC intervals for models  $M_1$  to  $M_{13}$  (Table 1) in the case studies: blue whiting (A), Northeast Arctic haddock (B), North Sea cod (C), and Northern Shelf haddock (D). The horizontal grey lines indicate AIC differences of 50 starting at the lowest lower bound of the models. Vertical dashed grey lines separate the models into model classes (Table 1).



Subarea IV (North Sea) and Division IIIa West (Skagerrak); the North Sea cod (*Gadus morhua*) data was obtained from the 2012 ICES advice (ICES 2012) for Subarea IV (North Sea) and Divisions VII (Eastern Channel) and IIIa West (Skagerrak); and the Northern Shelf haddock data was obtained from the ICES advice for Subarea IV (North Sea) and Division IIIa West (Skagerrak) in 2012 (ICES 2012). A Beverton–Holt curve was assumed for the relationship between stock and recruitment for North Sea cod, whereas a random walk was assumed for the other stocks. Two of the data sets had missing data. For the blue whiting data set a whole year was missing for the survey, whereas for the Northeast Arctic haddock data set, values were missing for at most three ages per year. The proportions-at-age models above could not easily handle years with missing age observations. Hence, to give a fair comparison of the univariate and multivariate models, we treated years with missing age observations as if it was missing all ages. Process model parameters were assumed equal between ages in the same way as they were for the advice.

## Results

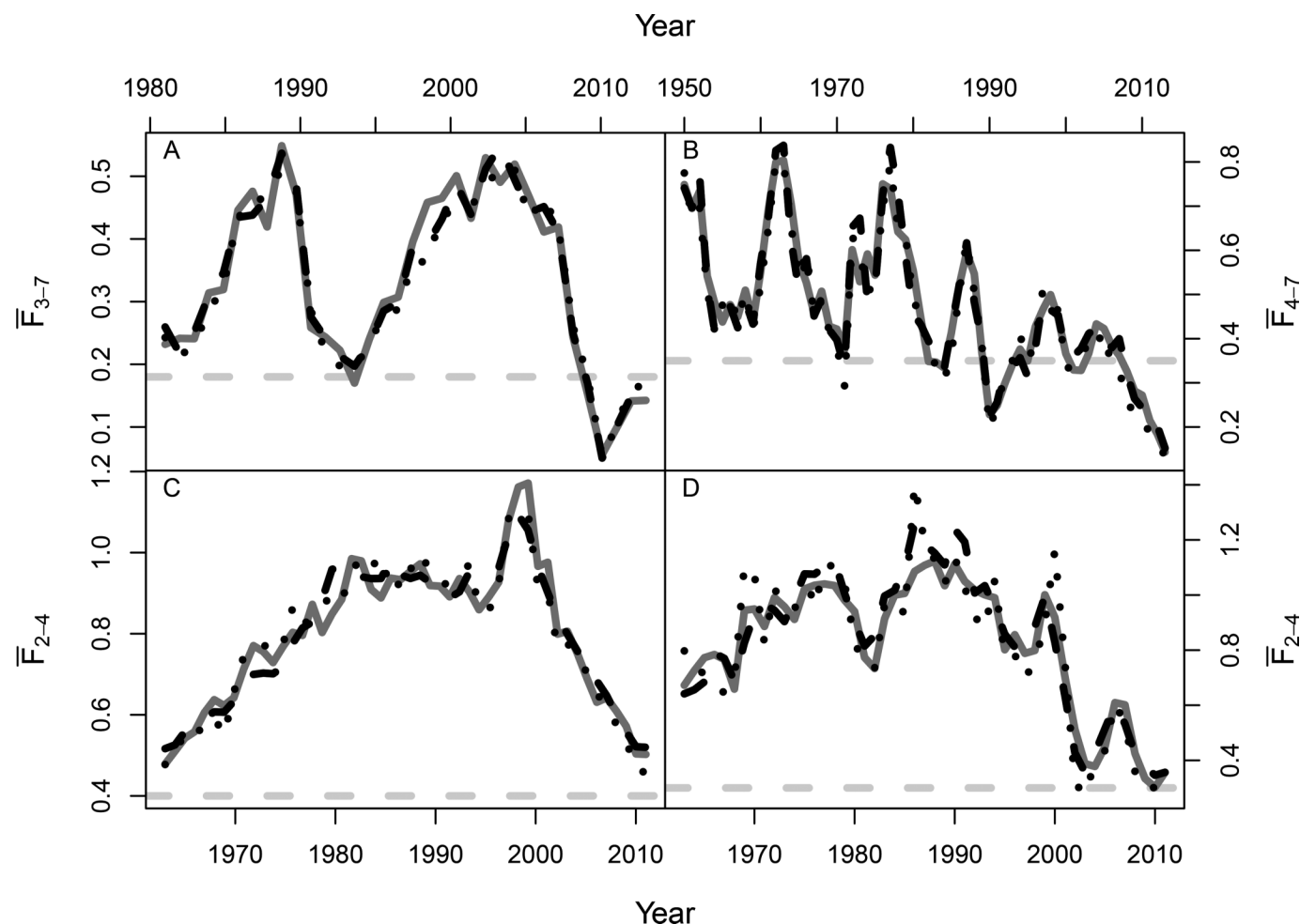
In all four case studies we found that the estimated mean fishing mortality (Fig. 1), spawning stock log-biomass (Fig. 2), and their standard errors in the final year differed among models. In particular, we found that for North Sea cod, the highest fishing mor-

tality was two times the lowest fishing mortality, and the widest confidence interval was 1.7 times the narrowest. For Northern Shelf haddock, the confidence interval of the estimated final year SSB for model  $M_4$ , which had the highest estimate, did not overlap with the confidence interval for model  $M_9$ , which had the lowest estimate.

We found that the models including correlation parameters obtained better fit to the data for the full models than models without correlation parameters within each model class (Fig. 3); the AIC for the full model (upper bound of the interval) was lower for the multivariate lognormal than for the univariate numbers-at-age, and similarly the logistic normals had better model fits than the Dirichlet distribution, which provided one of the highest AIC intervals of all models in all case studies. In the North Sea cod and Northern Shelf haddock cases, the lower bounds of the AIC intervals for the Dirichlet distribution were clearly separated from the upper AIC bounds of all other models, with an AIC difference of more than 190.

For North Sea cod, the multivariate lognormal achieved the lowest AIC for the full model. The AIC interval for this distribution (upper bound: 6407.32) only barely overlapped with the intervals for the generalized gamma (lower bound: 6405.52). Hence, among the observational likelihoods we considered, the multivariate lognormal was the most appropriate for the North Sea cod data and

**Fig. 4.** Estimated mean fishing mortality,  $\bar{F}$ , for multivariate lognormal (solid grey line), multiplicative logistic normal with lognormal mass (dashed black line), and Dirichlet with lognormal mass (dotted black line) in the case studies: blue whiting (A), Northeast Arctic haddock (B), North Sea cod (C), and Northern Shelf haddock (D). Horizontal dashed grey lines show the management plan reference point. Subscripts to  $\bar{F}$  indicate the ages the mean is over. All ages are weighed equally in the mean.



this particular process model. The multivariate lognormal also had the lowest AIC for the full model for Northeast Arctic haddock, whereas it was the multiplicative logistic normal with total mass for Northern Shelf haddock and the additive logistic normal with total numbers for blue whiting. However, in these cases the AIC intervals of the multivariate lognormal, the additive logistic normals, and the multiplicative logistic normals all overlapped. For the two haddock cases, the AIC interval of the generalized gamma also overlapped with the AIC interval of the model with the lowest AIC interval upper bound. We further found that the AIC intervals for proportions-at-age models using total mass overlapped with the corresponding model using total numbers-at-age, except for the North Sea cod where the total mass models had lower AIC intervals. In addition, the intervals for the additive and multiplicative logistic normals overlapped for all four data sets.

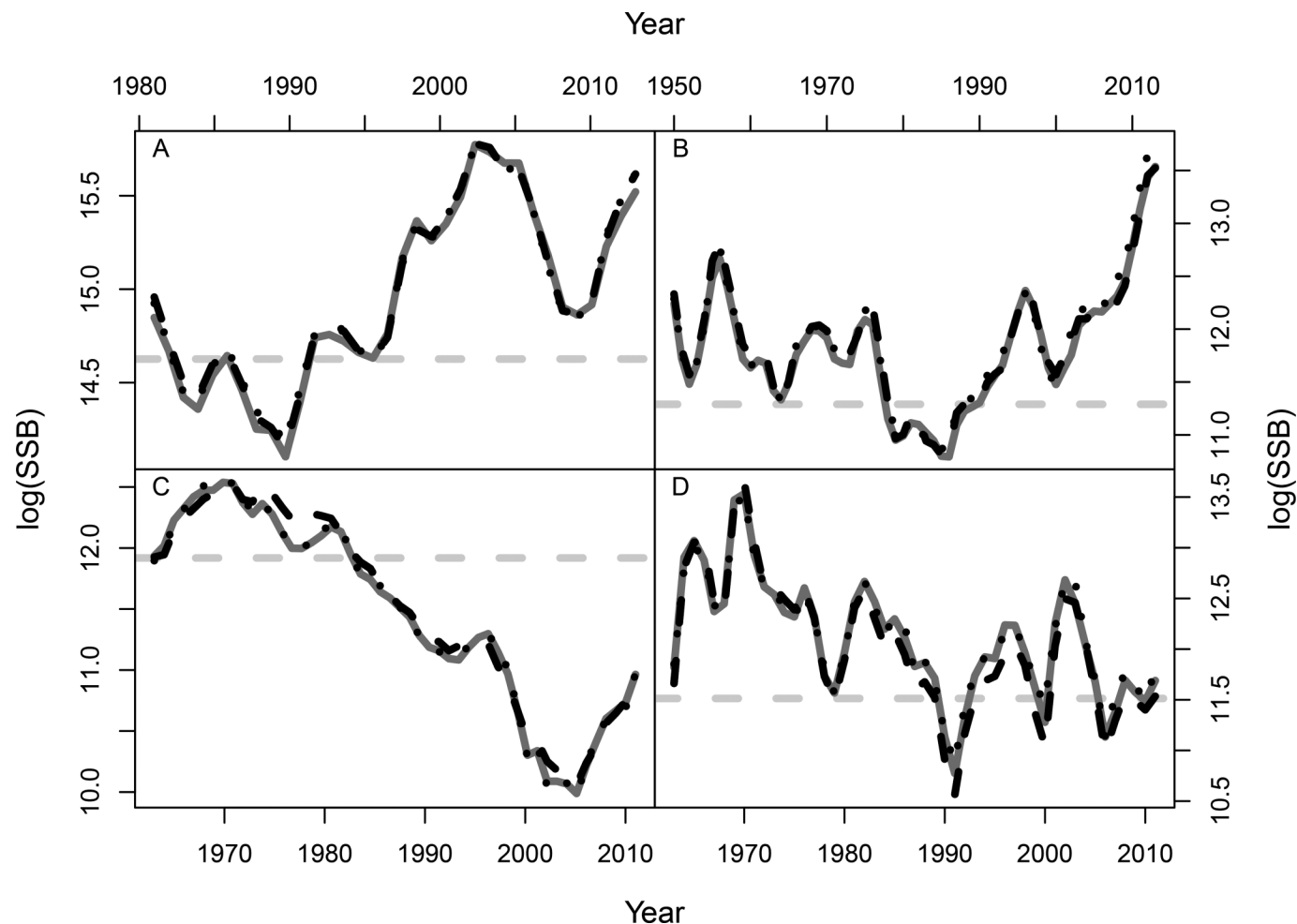
Overall, the trends in estimated fishing mortality (Fig. 4) and SSB (Fig. 5) were similar among the models (see also online supplementary material<sup>1</sup>). However, there were noticeable differences in single years. For blue whiting, the estimated fishing mortality and SSB for multivariate lognormal, multiplicative logistic-normal, and Dirichlet distribution followed each other closely. The largest difference in mean fishing mortality between

the multiplicative logistic-normal and the multivariate lognormal was 0.07 (16%), and the difference in SSB was up to 12%. In the North Sea cod case, the multivariate lognormal and the logistic normals had larger differences in fishing mortality and SSB. The largest difference in mortality was 0.12 (11.2%), while the SSB differed as much as 23%. The resulting confidence intervals also differed among the models. For North Sea cod the standard errors of the estimated mean fishing mortality were up to 76.6% larger for the Dirichlet model, which had the highest AIC, compared with the multivariate lognormal, which had the lowest AIC. Although the trends were similar to the other models, the SSB was estimated to be 3.8 to 14.2 times higher for the left-truncated normal than for the other models (Fig. S31<sup>1</sup>). Likewise, the mean fishing mortality was estimated to be lower. For both North Sea cod (Figs. S34, S35, S37, S38<sup>1</sup>) and Northern Shelf haddock (Figs. S47, S48, S50, S51<sup>1</sup>), the logistic normals provided less volatile estimated time series of fishing mortality and SSB than other models. For these models, the CVs for commercial catch were estimated to be higher than for the other data sets (Table S5<sup>1</sup>).

For blue whiting, Northeast Arctic haddock, and Northern Shelf haddock, estimated CVs were similar for the two logistic normals (Table S5<sup>1</sup>). The CVs were estimated to be between 0.09 and 0.37.

<sup>1</sup>Supplementary data are available with the article through the journal Web site at <http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2015-0532>.

**Fig. 5.** Natural logarithm of estimated spawning stock biomass,  $\log(SSB)$ , for multivariate lognormal (solid grey line), multiplicative logistic normal with lognormal mass (dashed black line), and Dirichlet with lognormal mass (dotted black line) in the case studies: blue whiting (A), Northeast Arctic haddock (B), North Sea cod (C), and Northern Shelf haddock (D). Horizontal dashed grey lines show the management plan reference point.



For North Sea cod, the logistic normals with total numbers had higher CVs (0.32 for commercial catch; 0.27 for survey) than the logistic normals with total mass (0.15 for commercial catch; 0.20 for survey). For the multivariate lognormal, the estimated marginal CVs were estimated to be between 0.01 and 1.86 (Tables S1–S4<sup>1</sup>). The CVs were typically estimated to be higher for the first and last ages.

## Discussion

When modelling highly aggregated stock assessment data, the optimal observational likelihood to use can not be known a priori. We provide an objective method for limiting the number of candidate models. By fitting each model with the largest possible number of parameters, upper and lower bounds for the lowest attainable AIC can be calculated. After discarding models with AIC intervals that does not intersect with the lowest interval, the remaining AIC intervals can be narrowed by combining parameters until a single observational likelihood is left. Once a final stock assessment model is found, model validation tools such as residuals and retrospective analysis should be used. AIC was used to determine the optimal observational likelihood, since the AIC estimates the Kullback–Liebler divergence between the candidate model and the true data-generating system (Akaike 1974), that is, the information lost by using the candidate model instead of the true data-generating system (Burnham and Anderson 2002). The

Kullback–Liebler divergence can not be used directly because the calculation requires full knowledge of the true data-generating system. Although the AIC was used here, other criteria allowing multivariate data could be used instead.

Choosing the best possible observational likelihood for the data are vital for the short-term management and conservation of fish stocks (Figs. 1 and Fig. 2). In 2012, the ICES advice for Northern Shelf haddock (ICES 2012), based on extended survivors analysis (XSA; Shepherd 1999), suggested a 15% increase in the total allowable catch. The suggested increase was based on the difference between the estimated mean fishing mortality in the last year and the management reference point. XSA can be seen as a special case of the univariate lognormal (ICES 2010b) by fixing parameters in a suitable way. Here we saw that the multivariate lognormal and multiplicative logistic normal were more suitable for the Northern Shelf haddock data than the univariate lognormal, which in turn provides a better fit than the XSA. The estimated mean fishing mortality from both the multivariate lognormal and the multiplicative logistic normal are above the reference point. Hence, using one of these models would have suggested a 19% decrease of the total allowable catch to get the fishing mortality at a sustainable level (Fig. 4D). Thus, the choice of observational likelihood in an assessment model can have a substantial effect on the advice brought forward to the fisheries management system.

All models implemented in this study fall into one of two categories: either they are formulated for numbers-at-age or proportions-at-age with total catch. Most assessment tools only consider one of these categories. We have shown how to compare models between these two ways of using the stock assessment data and that models from both categories can be suitable, depending on the specific stock. Besides the choice between modelling numbers or proportions, there are other differences among the models in, for example, tail probabilities and skewness, but there are also more subtle differences. When we choose between the lognormal distribution and the gamma distribution, we also choose between whether the Baranov catch equation should model the median or the mean observed catch. Some likelihoods can be reparameterized to link the Baranov catch equation to the mean, median, or modal observed catch. Although subtle, this difference is important for prediction and interpretation of the results. These choices can be compared objectively by including them in the analysis.

Accounting for the correlations in the data are also important for more reliable stock assessment models. In all four case studies, the distributions with correlation parameters, in particular the multivariate lognormal, performed well. However, the correlation structure must be flexible enough to mimic the data, unlike the Dirichlet distribution, which performed poorly, even compared with the models assuming independence between ages. The Dirichlet distribution arises as the distribution of proportions of gamma-distributed numbers, where the gamma distributions have equal scale parameters. Stock assessment data are often believed to have constant CV (Cadigan and Myers 2001), which leads to a parameterization of the gamma distribution where the scale parameters are not equal. Therefore, the covariance structure of the Dirichlet distribution does not match the appropriate structure for the gamma distributions. This corresponds with previous findings that the additive logistic normal generally is more suitable than the Dirichlet distribution in describing the correlation structure in stock assessment data (Francis 2014). For simplicity we restricted ourselves to the simple AR(1) structure for correlations in this study. The correlation structure can easily be exchanged for other structures, such as a linear, AR(2), ARMA( $p$ ,  $q$ ), compound symmetric, or unstructured covariance matrix (Pinheiro and Bates 2000; Francis 2014), and some of these may be even more suitable than the AR(1) structure (Berg and Nielsen 2016).

We only compared frequently used distributions in fisheries stock assessments and close extensions of them, yet any conceivable distribution can be used in the framework we presented, as long as they are all comparable. Correlations among age classes could be introduced in the univariate models through copulas or multivariate extensions such as the multivariate  $t$  distribution or multivariate gamma distributions. Likewise, different ways of handling zero observations, such as zero-inflating the models, could be included. For simplicity, years with missing data were removed to compare between univariate and multivariate models. For multivariate numbers-at-age models, missing data may be handled by finding the marginal distribution of the remaining ages. We also restricted the study to use the same likelihood for both commercial catches and surveys. This can be relaxed by including combinations of models (such as Cadigan 2016), or if the survey index generation is well-understood, a suitable observational likelihood may be derived a priori. Further, the analyses were made conditional on the process models specified in the assessments from which the data was collected. A similar analysis could be made to choose the most appropriate process model conditional on the observational model, or the analyses could be combined. This may influence the specific choice of observational likelihood, as the observational model can, somewhat, compensate for misspecification in the process model and vice versa. Finally, these methods could just as well be used for length-based models, for example.

Statistical assessment modelling involves a choice of observational likelihood, and current practice is often to make this choice arbitrarily and subjectively. This applies particularly to the choice of whether the data inputs should be numbers-at-age or proportions-at-age along with total catch in numbers. Here, we have provided methods for an objective choice, by correcting the AICs so that they can be compared between these two families, and we have outlined a computationally efficient method for choosing among families of distributions by bounding the AICs, which avoids elaborate hypothesis testing within each family, and we demonstrated that the best-fitting family depends on the particular case. These results will allow stock assessment modellers to choose objectively among these representations of uncertainty on observations, thereby improving model fit and ultimately allowing more accurate assessments.

## Acknowledgements

The authors thank Casper W. Berg and three anonymous reviewers for their valuable comments to improve the presentation of this manuscript.

## References

- Aanes, S., and Pennington, M. 2003. On estimating the age composition of the commercial catch of Northeast Arctic cod from a sample of clusters. *ICES J. Mar. Sci.* **60**(2): 297–303. doi:10.1016/S1054-3139(03)00008-0.
- Aitchison, J. 2003. *The statistical analysis of compositional data*. The Blackburn Press, Caldwell, N.J.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control*, **19**(6): 716–723. doi:10.1109/TAC.1974.1100705.
- Berg, C.W., and Nielsen, A. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. *ICES J. Mar. Sci.* **73**(7): 1788–1797. doi:10.1093/icesjms/fsw046.
- Burnham, K.P., and Anderson, D.R. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. 2nd ed. Springer.
- Cadigan, N.G. 2016. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. *Can. J. Fish. Aquat. Sci.* **73**(2): 296–308. doi:10.1139/cjfas-2015-0047.
- Cadigan, N.G., and Myers, R.A. 2001. A comparison of gamma and lognormal maximum likelihood estimators in a sequential population analysis. *Can. J. Fish. Aquat. Sci.* **58**(3): 560–567. doi:10.1139/f01-003.
- Cook, R. 2013. A fish stock assessment model using survey data when estimates of catch are unreliable. *Fish. Res.* **143**: 1–11. doi:10.1016/j.fishres.2013.01.003.
- Crone, P., and Sampson, D. 1998. Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition. In *Fishery stock assessment models*. Edited by F. Funk, T.J. Quinn, II, J. Heifetz, J.N., Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Zhang. Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks. pp. 355–370.
- Farewell, V.T., and Prentice, R.L. 1977. A study of distributional shape in life testing. *Technometrics*, **19**(1): 69–75. doi:10.2307/1268257.
- Fournier, D., and Archibald, C.P. 1982. A general theory for analyzing catch at age data. *Can. J. Fish. Aquat. Sci.* **39**(8): 1195–1207. doi:10.1139/f82-157.
- Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* **68**(6): 1124–1138. doi:10.1139/f2011-025.
- Francis, R.I.C.C. 2014. Replacing the multinomial in stock assessment models: a first step. *Fish. Res.* **151**(0): 70–84. doi:10.1016/j.fishres.2013.12.015.
- Fryer, R. 2002. TSA: is it the way? In *Report of Working Group on Methods of Fish Stock Assessment*, Dec. 2001. ICES CM 2002/D:01. pp. 86–93.
- Gudmundsson, G. 1994. Time series analysis of catch-at-age observations. *J. R. Stat. Soc. Ser. C Appl. Stat.* **43**(1): 117–126. doi:10.2307/2986116.
- Hulson, P.-J.F., Hanselman, D.H., and Quinn, T.J. 2011. Effects of process and observation errors on effective sample size of fishery and survey age and length composition using variance ratio and likelihood methods. *ICES J. Mar. Sci.* **68**(7): 1548–1557. doi:10.1093/icesjms/fsr102.
- Hulson, P.-J.F., Hanselman, D.H., and Quinn, T.J. 2012. Determining effective sample size in integrated age-structured assessment models. *ICES J. Mar. Sci.* **69**(2): 281–292. doi:10.1093/icesjms/fsr189.
- ICES. 2010a. Report of the Workshop on Reviews of Recent Advances in Stock Assessment Models Worldwide: “Around the World in AD Models” (WKADSAM), 27 September – 1 October 2010, Nantes, France. ICES CM 2010/SSGSUE:10.
- ICES. 2010b. Report of the Working Group on Methods of Fish Stock Assessment (WGMG), 20–29 October 2009, Nantes, France. ICES CM 2009/RMC:12.
- ICES. 2012. Report of the ICES Advisory Committee 2012. ICES Advice 2012, Book 6.
- ICES. 2014a. Report of the ICES Advisory Committee 2014. ICES Advice 2014, Book 9.
- ICES. 2014b. Report of the ICES Advisory Committee 2014. ICES Advice 2014, Book 3.



- Kristensen, K., Nielsen, A., Berg, C.W., and Bell, H.S.B. 2016. TMB: Automatic Differentiation and Laplace Approximation. *J. Stat. Softw.* **70**(1): 1–21. doi:10.18637/jss.v070.i05.
- Maund, M.N. 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: estimating the effective sample size. *Fish. Res.* **109**(2–3): 311–319. doi:10.1016/j.fishres.2011.02.018.
- McAllister, M.K., and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the sampling — importance resampling algorithm. *Can. J. Fish. Aquat. Sci.* **54**(2): 284–300. doi:10.1139/f96-285.
- Methot, R.D., Jr., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fish. Res.* **142**: 86–99. doi:10.1016/j.fishres.2012.10.012.
- Myers, R.A., and Cadigan, N.G. 1995. Statistical analysis of catch-at-age data with correlated errors. *Can. J. Fish. Aquat. Sci.* **52**(6): 1265–1273. doi:10.1139/f95-123.
- Nielsen, A., and Berg, C. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fish. Res.* **158**: 96–101. doi:10.1016/j.fishres.2014.01.014.
- Pinheiro, J.C., and Bates, D.M. 2000. Mixed-effects models in S and S-Plus. Springer, New York.
- Prentice, R.L. 1974. A Log gamma model and its maximum likelihood estimation. *Biometrika*, **61**(3): 539–544. doi:10.1093/biomet/61.3.539.
- Shepherd, J.G. 1999. Extended survivors analysis: An improved method for the analysis of catch-at-age data and abundance indices. *ICES J. Mar. Sci.* **56**(5): 584–591. doi:10.1006/jmsc.1999.0498.
- Stacy, E.W. 1962. A generalization of the gamma distribution. *Ann. Math. Stat.* **33**(3): 1187–1192. [Institute of Mathematical Statistics.] doi:10.1214/aoms/1177704481.
- Williams, E.H., and Quinn, T.J. 1998. A parametric bootstrap of catch-age compositions using the Dirichlet Distribution. *In* Fishery stock assessment

- models. Edited by F. Funk, T.J. Quinn, II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Zhang. Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks. pp. 371–384.
- Williams, E.H., and Shertzer, K.W. 2015. Technical documentation of the Beaufort Assessment Model (BAM). NOAA Technical Memorandum, US Department of Commerce. doi:10.7289/V57M05W6.

## Appendix A

### Process model

The process model is identical to Nielsen and Berg (2014). For a model including age groups from 1 to  $A^+$  (where age group  $A^+$  contains all ages from  $A$  and up), the fishing mortality is modelled by a multivariate random walk, where the oldest modelled ages may be grouped together, indicated by using  $A^*$  rather than  $A^+$ . Let  $F_y = (F_{1,y}, F_{2,y}, \dots, F_{A^*,y})^T$  be a vector of age-specific fishing mortalities in year  $y$ . Then

$$\log F_y = \log F_{y-1} + \epsilon_y$$

where  $\epsilon_y \sim N(0, \Sigma)$ . The covariance matrix,  $\Sigma$ , is parameterized by an AR(1) structure,  $\Sigma_{ij} = \rho^{|i-j|} \sigma_i \sigma_j$ .

Given the fishing mortalities, the population is modelled by an exponential decay model:

$$\begin{aligned} \log N_{1,y} &= \log[R(w_{1,y-1}, \dots, w_{A^+,y-1}, p_{1,y-1}, \dots, p_{A^+,y-1}, N_{1,y-1}, \dots, N_{A^+,y-1})] + \eta_{1,y} \\ \log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y}, \quad 2 \leq a < A^+ \\ \log N_{A^+,y} &= \log[N_{A^+-1,y-1} \exp(-F_{A^+-1,y-1} - M_{A^+-1,y-1}) + N_{A^+,y-1} \exp(-F_{A^+,y-1} - M_{A^+,y-1})] + \eta_{A^+,y} \end{aligned}$$

where all error terms are assumed independent normal distributed. The natural mortalities ( $M_{a,y}$ ), the age-specific mass in stock ( $w_{a,y}$ ), and the proportion mature ( $p_{a,y}$ ) are all assumed to be known. The function  $R$  describes the relationship between recruitment and spawning population. For North Sea cod,  $R$  was modelled by a Beverton–Holt curve

$$\begin{aligned} R(w_{1,y-1}, \dots, w_{A^+,y-1}, p_{1,y-1}, \dots, p_{A^+,y-1}, N_{1,y-1}, \dots, N_{A^+,y-1}) \\ = \frac{a \cdot \text{SSB}_{y-1}}{1 + b \cdot \text{SSB}_{y-1}}, \quad a, b > 0 \end{aligned}$$

with  $\text{SSB}_y = \sum_{a=1}^{A^+} p_{a,y} w_{a,y} N_{a,y}$ . However, for the other stocks,  $R$  was modelled by a random walk

$$R(w_{1,y-1}, \dots, w_{A^+,y-1}, p_{1,y-1}, \dots, p_{A^+,y-1}, N_{1,y-1}, \dots, N_{A^+,y-1}) = N_{1,y-1}$$

### Observational models

#### Univariate numbers-at-age models

We consider six univariate observational models for numbers-at-age. Their densities are listed below for each age each year. The joint density for the vector of catches-at-age each year is the prod-

uct of the age-wise densities. Unless otherwise noted,  $x > 0$  is the observed catch (or survey index) at a given age,  $\mu > 0$  is the calculated catch at that age (or survey index) based on the Baranov catch equation (or proportional to the total abundance),  $\sigma > 0$  is a scale parameter, and  $\tau \in \mathbb{R}$  is a shape parameter.

#### Lognormal distribution

$$f_1(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[\log(x) - \log(\mu)]^2}{2\sigma^2}\right\} x^{-1}$$

The mean of the lognormal distribution is  $\mu \exp(\sigma^2/2)$ , the median is  $\mu$  and the variance is  $(\exp(\sigma^2) - 1)\exp[2\log(\mu) + \sigma^2]$ .

#### Gamma distribution

$$f_2(x; \mu, \sigma) = \frac{1}{\Gamma(\sigma) \left(\frac{\mu}{\sigma}\right)^\sigma} x^{\sigma-1} \exp(-x\sigma/\mu)$$

The gamma distribution has mean  $\mu$  and variance  $\mu^2/\sigma$ .

### Generalized gamma distribution

$$f_3(x; \mu, \sigma, \tau) = \begin{cases} \left| \tau (\tau^{-2})^{\tau-2} \exp\left(\tau^{-2} \left\{ \frac{\log(x) - \log(\mu)}{\sigma} - \exp\left[\frac{\log(x) - \log(\mu)}{\sigma}\right] \right\} \right) \right| [\sigma x \Gamma(\tau^{-2})], & \tau \neq 0 \\ (2\pi)^{-1/2} \exp\left\{-\frac{[\log(x) - \log(\mu)]^2}{2\sigma^2}\right\} (\sigma x)^{-1}, & \tau = 0 \end{cases}$$

Note that  $f_3(x; \mu, \sigma, 0) = f_1(x; \mu, \sigma)$ , and  $f_3(x; \mu, \sigma, \sigma) = f_2(x; \mu, \sigma)$  for  $\sigma > 0$  (Cox et al. 2007). The mean of the generalized gamma is

$$\frac{\mu(\tau^{-2})^{\tau-2}}{\Gamma(\tau^{-2})}(\tau^2)^{\frac{\sigma\tau+1}{\tau^2}}\Gamma\left(\frac{\sigma\tau+1}{\tau^2}\right), \quad \tau < 0$$

$$\mu \exp(\sigma^2/2), \quad \tau = 0$$

$$\frac{\mu}{\Gamma(\tau^{-2})}\tau^{\frac{2\sigma}{\tau}}\Gamma\left(\frac{\sigma\tau+1}{\tau^2}\right), \quad \tau > 0$$

and the variance is

$$-\frac{\mu^2}{[\Gamma(\tau^{-2})]^2}\left\{(\tau^{-2})^{2\tau-2}(\tau^2)^{\frac{2\sigma\tau+1}{\tau^2}}\left[\Gamma\left(\frac{\sigma\tau+1}{\tau^2}\right)\right]^2 - (\tau^2)^{\frac{2\sigma\tau+1}{\tau^2}}(\tau^{-2})^{\tau-2}\Gamma\left(\frac{2\sigma\tau+1}{\tau^2}\right)\Gamma(\tau^{-2})\right\}, \quad \tau < 0$$

$$[\exp(\sigma^2) - 1] \exp[2 \log(\mu) + \sigma^2], \quad \tau = 0$$

$$\frac{\mu^2}{[\Gamma(\tau^{-2})]^2}\tau^{\frac{4\sigma}{\tau}}\left\{\Gamma\left(\frac{2\sigma\tau+1}{\tau^2}\right)\Gamma(\tau^{-2}) - \left[\Gamma\left(\frac{\sigma\tau+1}{\tau^2}\right)\right]^2\right\}, \quad \tau > 0$$

### Normal distribution

$$f_4(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi(\mu\sigma)^2}} \exp\left[-\frac{(x - \mu)^2}{2(\mu\sigma)^2}\right]$$

For the normal distribution,  $x \in \mathbb{R}$  and  $\mu \in \mathbb{R}$ . The mean is  $\mu$  and the variance is  $\mu^2\sigma^2$ .

### Truncated normal distribution

$$f_5(x; \mu, \sigma) = \frac{f_4(x; \mu, \sigma)1_{x \geq 0}(x)}{1 - \int_{-\infty}^0 f_4(y; \mu, \sigma)dy}$$

For the truncated normal distribution,  $x \geq 0$  and  $\mu \geq 0$ . The mean of the distribution is

$$\mu + \mu\sigma \frac{f_4(0; \mu, \sigma)}{1 - \int_{-\infty}^0 f_4(y; \mu, \sigma)dy}$$

and the variance is

$$\mu^2\sigma^2 - \mu^2\sigma^2 \frac{f_4(0; \mu, \sigma)^2}{\left[1 - \int_{-\infty}^0 f_4(y; \mu, \sigma)dy\right]^2}$$

### Student's *t* distribution on log-scale

$$f_6(x; \mu, \sigma, \tau) = \frac{\Gamma\left(\frac{\tau+1}{2}\right)}{\sigma x \sqrt{\tau\pi} \Gamma\left(\frac{\tau}{2}\right)} \left(1 + \frac{[\log(x) - \log(\mu)]^2/\sigma^2}{\tau}\right)^{-(\tau+1)/2}$$

For the Student's *t* distribution on log-scale,  $\tau > 0$  is the degrees of freedom. As  $\tau \rightarrow \infty$ , the distribution converges to a lognormal distribution. The distribution does not have mean and variance.

### Multivariate numbers-at-age models

We consider one multivariate observational model for numbers-at-age. In the density listed below,  $A$  is the number of ages,  $\mathbf{x} > \mathbf{0}$  is the observed vector of catches (or survey indices),  $\boldsymbol{\mu} > \mathbf{0}$  is the calculated catches (or survey indices) based on the Baranov catch

equation (or proportional to the total abundance), and  $\boldsymbol{\Sigma}$  is an  $A \times A$  symmetric positive-definite scale matrix.

### Multivariate lognormal distribution

$$f_7(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-A/2} |\boldsymbol{\Sigma}|^{-1/2} \times \exp\left\{-\frac{1}{2}[\log(\mathbf{x}) - \log(\boldsymbol{\mu})]^T \boldsymbol{\Sigma}^{-1} [\log(\mathbf{x}) - \log(\boldsymbol{\mu})]\right\} (x_1 \cdots x_A)^{-1}$$

For the multivariate lognormal distribution, the scale matrix is the covariance matrix of the logarithm of the observations. When the scale matrix is diagonal, the distribution reduces to univariate lognormals. The marginal means are  $\mu_i \exp(1/2 \Sigma_{ii})$ , and the variance-covariance is  $\mu_i \mu_j \exp\left[\frac{1}{2}(\Sigma_{ii} + \Sigma_{jj})\right] [\exp(\Sigma_{ij}) - 1]$ .

### Proportions-at-age models

We consider three multivariate observational model for proportions-at-age. In the density listed below,  $A$  is the number of ages,  $\mathbf{x} > \mathbf{0}$  with  $\sum_{i=1}^A x_i = 1$  is a vector of  $A$  observed catch proportions (or survey proportions),  $\boldsymbol{\mu} > \mathbf{0}$  with  $\sum_{i=1}^A \mu_i = 1$  is a vector of  $A$  calculated catch proportions (or survey proportions) based on the Baranov catch equation (or proportional to the total abundance), and  $\boldsymbol{\Sigma}$  is an  $A - 1 \times A - 1$  symmetric positive-definite scale matrix. A vector with subscript  $-A$  denotes the vector without the  $A$ th element.

### Additive logistic-normal distribution

$$f_8(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-(A-1)/2} |\boldsymbol{\Sigma}|^{-1/2} \times \exp\left\{-\frac{1}{2}[\alpha(\mathbf{x}) - \alpha(\boldsymbol{\mu})]^T \boldsymbol{\Sigma}^{-1} [\alpha(\mathbf{x}) - \alpha(\boldsymbol{\mu})]\right\} (x_1 \cdots x_A)^{-1}$$

Here,  $\alpha$  is the additive logratio transformation  $\alpha(\mathbf{x}) = \log\left(\frac{\mathbf{x}_{-A}}{x_A}\right)$ . The scale matrix is the covariance of the additive logratio transformed observations. Note that if a numbers-at-age vector  $\mathbf{y}$  follows a multivariate lognormal, then the proportions  $\mathbf{y}/\sum_i y_i$  follows an additive logistic-normal distribution (Aitchison 2003). The mean and variance do not have simple forms (Aitchison 2003).

### Multiplicative logistic-normal distribution

$$f_9(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-(A-1)/2} |\boldsymbol{\Sigma}|^{-1/2} \times \exp\left\{-\frac{1}{2}[\mathbf{m}(\mathbf{x}) - \mathbf{m}(\boldsymbol{\mu})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{m}(\mathbf{x}) - \mathbf{m}(\boldsymbol{\mu})]\right\} (x_1 \cdots x_A)^{-1}$$

Here,  $m$  is the multiplicative logratio transformation

$$m(\mathbf{x}) = \left[ \log\left(\frac{x_1}{1 - x_1}\right), \dots, \log\left(\frac{x_{A-1}}{1 - x_1 - \dots - x_{A-1}}\right) \right]$$

The scale matrix is the covariance of the multiplicative logratio transformed observations. The mean and variance do not have simple forms (Aitchison 2003).

#### Dirichlet distribution

$$f_{10}(\mathbf{x}; \mu, \sigma) = \frac{\Gamma\left(\sum_{i=1}^A \sigma \mu_i\right)}{\prod_{i=1}^A \Gamma(\sigma \mu_i)} \prod_{i=1}^A x_i^{\sigma \mu_i - 1}$$

For the Dirichlet distribution,  $\sigma > 0$  is a shape parameter. The marginal means of the Dirichlet distribution are  $\mu_i$ , the variances are  $\frac{\mu_i - \mu_i^2}{\sigma + 1}$ , and the covariances are  $\frac{-\mu_i \mu_j}{\sigma + 1}$ . The Dirichlet distribution is related to the gamma distribution, since if each element of a numbers-at-age vector follows a gamma distribution where the scale parameters in the usual parameterization are equal, then the vector of proportions  $\mathbf{y} / \sum_i y_i$  follows a Dirichlet distribution. Note that  $f_2$  does not have the same scale parameters for all ages, as they depend on the mean value.

#### References

- Aitchison, J. 2003. The statistical analysis of compositional data. The Blackburn Press, Caldwell, NJ.
- Cox, C., Chu, H., Schneider, M.F., and Muñoz, A. 2007. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat. Med.* **26**(23): 4352–4374. doi:10.1002/sim.2836. PMID:17342754.
- Nielsen, A., and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fish. Res.* **158**: 96–101. doi:10.1016/j.fishres.2014.01.014.

## Appendix B

### Transformation of densities for proportions-at-age models

To compare the AIC of models for numbers-at-age with models for proportions-at-age, the data must be on the same scale. We note that we can transform the numbers-at-age data to proportions-at-age with total catch in numbers by the function

$$g[(x_1, \dots, x_A)^T] = \left( \frac{x_1}{\sum_{i=1}^A x_i}, \dots, \frac{x_{A-1}}{\sum_{i=1}^A x_i}, \frac{\sum_{i=1}^A x_i}{\sum_{i=1}^A x_i} \right)^T$$

with inverse function

$$h[(y_1, \dots, y_{A-1}, y_{\text{total}})^T] = \left[ y_1 \cdot y_{\text{total}}, \dots, y_{A-1} \cdot y_{\text{total}}, \left(1 - \sum_{i=1}^{A-1} y_i\right) \cdot y_{\text{total}} \right]^T$$

Hence, if  $\mathbf{Y}$  is a random vector of proportions-at-age with total catch in numbers with density  $f$ , then the corresponding numbers-at-age is  $\mathbf{X} = h(\mathbf{Y})$ . By a change of variable

$$\begin{aligned} P(\mathbf{X} \in B) &= P[\mathbf{Y} \in g(B)] \\ &= \int_{g(B)} f(\mathbf{y}) d\mathbf{y} \\ &= \int_B f[g(\mathbf{x})] |\det(Dg)| d\mathbf{x} \end{aligned}$$

the density for the numbers-at-age data are  $f[g(\mathbf{x})] |\det(Dg)|$ . Hence, to compare the AIC between the (natively) numbers-at-age and proportions-at-age models, the log-likelihoods of the proportions-at-age models must be corrected by the logarithm of the absolute determinant of the Jacobian of  $g$ ,  $\log |\det(Dg)|$ . The entries of  $(Dg)$

are  $(Dg)_{A,j} = 1$  for all  $j$ ,  $(Dg)_{i,i} = \frac{1}{\sum_{k=1}^A x_k} - \frac{x_i}{(\sum_{k=1}^A x_k)^2}$  for all  $i < A$ , and  $(Dg)_{i,j} = \frac{-x_i}{(\sum_{k=1}^A x_k)^2}$  otherwise. If the total is in mass, where the mass-at-age is assumed to be known, then  $g$  is adjusted to

$$g[(x_1, \dots, x_A)^T] = \left( \frac{x_1}{\sum_{i=1}^A x_i}, \dots, \frac{x_{A-1}}{\sum_{i=1}^A x_i}, \sum_{i=1}^A w_i x_i \right)^T$$