Fisheries Research 192 (2017) 52–65

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Some insights into data weighting in integrated stock assessments

André E. Punt[a,b,*]

[a] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195-5020, USA
[b] CSIRO Oceans and Atmosphere Flagship, Castray Esplanade, Hobart, Tasmania, Australia

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form
11 December 2015
Accepted 14 December 2015
Available online 2 January 2016

Keywords:
Age-structured stock assessment methods
Conditional age-at-length data
Data weighting
Length-composition data
Stock Synthesis
Simulation
Spatial structure

ABSTRACT

The results of fishery stock assessments based on the integrated analysis paradigm can be sensitive to the values for the factors used to weight each of the data types included in the objective function minimized to obtain the estimates of the parameters of the model. These assessments generally include relative abundance index data, length-composition information and conditional age-at-length data, and algorithms have been developed to select weighting factors for each of these data types. This paper introduces methods for weighting conditional age-at-length data that extend an approach developed by Francis (2011) to weight age- and length-composition data. Simulation based on single-zone and two-zone operating models are used to compare five tuning methods that are constructed as combinations of methods to weight each data type. The single-zone operating models allow evaluation of the tuning methods in terms of their ability to provide unbiased estimates of management-related quantities and the correct data weights in the absence of model mis-specification, while the two-zone operating models allow the impacts of model mis-specification on the performance of tuning methods to be explored. The results of assessments are sensitive to data weighting, but the choice of method for data weighting is most consequential when there is model mis-specification. Overall, the results indicate that arithmetic averaging of effective sampling sample sizes from the McAllister and Ianelli (1997) approach is inferior to other methods, and the new method for computing effective sample sizes for conditional age-at-length data seems most appropriate.

1. Introduction

Increasingly, the scientific advice given to fishery managers is based on the results of integrated population models (Newman et al., 2014). Integrated population models separate the development of the model of the population dynamics from that of the relationship between the data and the model, and allowance can be made for error in both the population dynamics and the observations (Maunder and Punt, 2013). Integrated population models (or integrated analyses) have been used in fisheries for decades, the earliest examples of the method in fisheries being Doubleday (1976), Fournier and Archibald (1982) and Deriso et al. (1985). Use of integrated analysis has been common in fisheries because there are often many data types (e.g., age-and-growth information, catch-rates, survey indices of abundance, catch-at-age data), each of which can provide information about some, but not all, of the parameters or processes that govern the dynamics of exploited fish and invertebrate populations.

Many integrated analysis assessments are now conducted using one of three packages that include generalized estimation frameworks (Stock Synthesis (Methot and Wetzel, 2013), MULTIFAN-CL (Fournier et al., 1998; Hampton and Fournier, 2001), and CASAL (Bull et al., 2005)), although several other, but less generally-applicable, packages have been developed.

One of the key advantages of integrated analysis is the ability to use multiple data sources to estimate the current abundance, trend in abundance, and productivity of populations. However, it is not uncommon for data sources to be in conflict with each other to some extent. Thus, each data type (and each data point within each data type) needs to be assigned a weight. In principle, this weight should relate to the deviation between the data point and its expected value. However, it is not straightforward to objectively select weights, and history reveals that data weighting is influential (e.g., Richards, 1991).

The selection of weights for compositional data (length-composition data, age-composition data, and conditional age-at-length data) is perhaps the most challenging aspect of selecting weights for data (although selecting the extent of variation in pro-

* Corresponding author at: School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195-5020, USA. Fax: +1 206 685 7471.
E-mail address: aepunt@uw.edu

http://dx.doi.org/10.1016/j.fishres.2015.12.006
0165-7836/

Table 1
Notation and derived symbols.

Symbol	Description	Derivation
(a) Length-composition data		
$p_{y,L}$	Observed proportion of animals in length-class L during year y	Input data
$\hat{p}_{y,L}$	Predicted proportion of animals in length-class L during year y	Model prediction
N_y	Input effective sample size for the length-composition data for year y	Input data
n_y	Number of years with length-composition data	Input data
$\bar{L}_L$	Mid-point of length-class L	Input data
E_y	McAllister-Ianelli effective sample size for the length data for year y	$\sum_L \hat{p}_{y,L}(1-\hat{p}_{y,L})/\sum_L (p_{y,L}-\hat{p}_{y,L})^2$
$\bar{L}_y$	Observed mean length of the catch during year y	$\sum_L \bar{L}_L p_{y,L}$
$\hat{\bar{L}}_y$	Predicted mean length of the catch during year y	$\sum_L \bar{L}_L \hat{p}_{y,L}$
$SE(\hat{\bar{L}}_y)$	Predicted standard error of the mean length of the catch for year y	$\sqrt{\sum_L \hat{p}_{y,L}(\bar{L}_L-\hat{\bar{L}}_y)^2}/\sqrt{N_y}$
(b) Conditional age-at-length data		
$p_{y,L,a}$	Observed proportion of animals in length-class L during year y that are of age a	Input data
$\hat{p}_{y,L,a}$	Predicted proportion of animals in length-class L during year y that are of age a	Model prediction
$N_{y,L}$	Input effective sample size for the conditional age-at-length data for year y and length-class L	Input data
$n_{y,L}$	Number of combinations of years and length-classes with conditional age-at-length data	Input data
$w_{y,L}$	Proportion of the conditional age-at-length data for year y that is in length-class L	$N_{y,L}/\sum_{L'} N_{y,L'}$
$E_{y,L}$	McAllister-Ianelli effective sample size for the conditional age-at-length data for year y and length-class L	$\sum_a \hat{p}_{y,L,a}(1-\hat{p}_{y,L,a})/\sum_a (p_{y,L,a}-\hat{p}_{y,L,a})^2$
$\bar{a}_{y,L}$	Observed mean age of the catch for year y and length-class L[a]	$\sum_a (a+0.5)p_{y,L,a}$
$\hat{\bar{a}}_{y,L}$	Predicted mean age of the catch for year y and length-class L[a]	$\sum_a (a+0.5)\hat{p}_{y,L,a}$
$SE(\hat{\bar{a}}_{y,L})$	Predicted standard error of the mean age of the catch for year y and length-class L[a]	$\sqrt{\sum_a \hat{p}_{y,L,a}\left((a+0.5)-\hat{\bar{a}}_{y,L}\right)^2}/\sqrt{N_{y,L}}$
$\bar{a}_y$	Observed mean age of the age-length key for year y	$\sum_L w_{y,L}\bar{a}_{y,L}$
$\hat{\bar{a}}_y$	Predicted mean age of the age-length key for year y	$\sum_L w_{y,L}\hat{\bar{a}}_{y,L}$
$SE(\hat{\bar{a}}_y)$	Predicted standard error of the mean age of the catch for year y	$\sqrt{\sum_L (w_{y,L})^2 SE(\hat{\bar{a}}_{y,L})^2}$

[a] The +0.5 is introduced to account the fact that fisheries occur throughout the year.

cess error is often a close second). Until recently, the approach for weighting compositional data was often to apply the following iterative approach:

(a) the values for the weights for the composition data (usually a factor that multiplies some input or initial effective sample size e.g., the number of fish, hauls or even trips sampled) are set;
(b) the population dynamics model is fitted to the data;
(c) the method of McAllister and Ianelli (1997) is used to calculate an overdispersion factor for the composition data, and the residual variances for the indices of abundance are set to the mean square errors;
(d) the values for the weights (the extent of overdispersion for the computational data and the residual variances for the indices of abundance) are replaced by the calculated overdispersion factors (often separately by year) and mean square errors respectively;
(e) steps (b–d) are applied until convergence occurs.

This approach can be criticized for several reasons, including that it may not converge to a sensible result (e.g., resulting in weights of zero or unrealistically high weights for some data types or years within series), but particularly because it fails to account for positive correlation in residuals between adjacent age- or length-classes (Francis, 2011). The likelihood function for compositional data, usually the multinomial, assumes that residuals should be negatively correlated and there should be no "runs" of positive or negative residuals. However, it is commonly the case that there are "runs" of residuals (e.g., Whitten and Punt, 2014). Basing the weighting for the compositional data on the method of McAllister and Ianelli (1997), which assumes that residuals are independent, may lead to over-weighting (compared to other data sources in the assessment) of the compositional data. In principle, a likelihood function could be selected that allows for "runs" of Pearson residuals (such as the multivariate normal). However, it is more common, and computationally easier, to downweight the compositional data using the approach outlined above.

Francis (2011) provided an alternative way to weight age- and length-composition data that accounts for the positive correlation between the residuals (and generally leads to lower weights for such data). However, Francis (2011) did not provide a way to weight conditional age-at-length data. This data type (essentially an age-length key) provides key information on growth as well as year-class strength in integrated stock assessments. Unfortunately, "Francis weighting", as originally conceived, was developed for compositional data types that are vectors of numbers whereas an annual age-length key is a matrix.

This paper first introduces a set of five candidate tuning methods (combinations of methods to weight each of the three key data types included in an integrated analysis assessment: abundance index, length-frequency data and conditional age-at-length data). The methods include two new ways to weight conditional age-at-length data. It then uses simulation to evaluate the tuning methods when they are applied within the context of a spatially-aggregated multi-fleet (trawl and non-trawl) integrated analysis estimation method. The simulations are based on the ideal situation in which the operating model is essentially identical to the estimation method, as well as on cases in which the estimation method is mis-specified owing to un-modeled spatial structure in exploitation rates. The analyses are based on simulations that have been developed to explore the performance of spatially-structured stock assessments for pink ling (*Genypterus blacodes*) off southeast Australia (Punt et al., 2015, 2017), but the results are relevant to stocks with similar life histories and data types.

2. Methods

2.1. Weighting/tuning methods

Tuning of the standard errors for the logarithms of the indices of abundance involves setting these standard errors to:

$$\tilde{\sigma} = \sqrt{\frac{1}{n}\sum_{y} \ell n(I_y/\hat{I}_y)^2} \tag{1}$$

where I_y is the observed index of abundance for year y, $\hat{I}_y$ is the model-estimate of the index of abundance for year y, and n is the number of index values. Eq. (1) is applied separately for each index of abundance.

There are several approaches to calculating weights (factors to multiply the input effective sample sizes by) for compositional data. Until recently, weights have frequently been calculated by applying the method of McAllister and Ianelli (1997) to calculate an effective sample size for each year with length-composition data (E_y in Table 1) or an effective sample size for each combination of year and length-class with conditional age-at-length data ($E_{y,L}$ in Table 1). The formulae for E_y and $E_{y,L}$ in Table 1 arise by taking inverse of the ratio of the variance of the residuals of the fit of the model to the data (observed less predicted proportions) to the expected variance under the assumption that the data were multinomial.

The values for E_y and $E_{y,L}$ can then be expressed relative to the initial effective sample sizes (i.e., E_y/N_y and $E_{y,L}/N_{y,L}$) as measures of over-dispersion. The initial effective sample sizes can then be multiplied by a constant, computed either as the arithmetic or harmonic mean of the year- and year-and-length-class-specific estimates of overdispersion (Eq. (1.A), (1.B), (2.A) and (2.B) in Table 2).

Francis (2011) provided an alternative approach to computing effective sample sizes for length- and age-composition data based on the extent to which the observed and model-predicted mean lengths and ages differ, taking account of the variance expected given the model-predicted length and age distributions and the input effective sample sizes (Eq. (1.C) in Table 2). Eq. (1.C) arises because the variance (among years) of the differences between the observed and model-predicted mean lengths and ages will be larger than expected given the standard errors of the means expected from the sample sizes and the model-predicted distributions of length and age frequency if numbers by age and length are correlated. It sets the weight for time-series of length- or age-compositions so that the expected variance of the differences matches the actual variance of the Pearson residuals for mean length or age.

Table 2
The tuning methods considered in this paper for the length and conditional age-at-length data.

Data type/tuning method	Equation	
Length-composition data		
McAllister-Ianelli-1	$\frac{1}{n_y}\sum_{y}(E_y/N_y)$	(1.A)
McAllister-Ianelli-2	$\left\{\frac{1}{n_y}\sum_{y}(E_y/N_y)^{-1}\right\}^{-1}$	(1.B)
Francis	$1/\underset{y}{var}\left\{(\bar{L}_y-\hat{\bar{L}}_y)/SE(\hat{\bar{L}}_y)\right\}$	(1.C)
Conditional age-at-length data		
McAllister-Ianelli-1	$\frac{1}{n_{y,L}}\sum_{y}\sum_{L}(E_{y,L}/N_{y,L})$	(2.A)
McAllister-Ianelli-2	$\left\{\frac{1}{n_{y,L}}\sum_{y}(E_{y,L}/N_{y,L})^{-1}\right\}^{-1}$	(2.B)
Francis-B	$1/\underset{y,L}{var}\left\{(\bar{a}_{y,L}-\hat{\bar{a}}_{y,L})/SE(\hat{\bar{a}}_{y,L})\right\}$	(2.C)
Francis-A	$1/\underset{y}{var}\left\{(\bar{a}_y-\hat{\bar{a}}_y)/SE(\hat{\bar{a}}_y)\right\}$	(2.D)

Francis (2011) did not provide ways to tune the initial effective sample sizes for conditional age-at-length data even though this data type is now widely used in integrated analysis stock assessments to enable growth to be estimated within the assessment (e.g., Punt et al., 2006). However, it is possible to extend the method of Francis (2011) to apply to conditional age-at-length data. The first way to achieve this goal (denoted Francis-B[1] in Table 2) is to apply the Francis tuning method calculating a mean age for each combination of year and length-class in the age-length key (Eq. (2.C) in Table 2) while a second way (denoted Francis-A in Table 2) is to compute the mean age of the age-length key for each year (Eq. (2.D) in Table 2). Application of the Francis-A approach is based on setting the proportion of animals in each length-class based on the observed proportions (see the definitions of $\bar{a}_y$ and $\hat{\bar{a}}_y$ in Table 1). In principle, the Francis-A approach could be applied weighting the mean ages by length-class by the model-predicted proportions by length-class. However, that would make an 'observation' ($\bar{a}_y$) depend on a model-estimate (the predicted proportions by length-class). One (in principle) advantage of the Francis-A method over the Francis-B method is that it accounts for correlations in the difference between the observed and model-predicted mean ages by length-class in the age-length key among length-classes.

2.2. Overview of the simulation procedure

The simulations are based on generating data sets from two operating models. The operating models are both based on pink ling off southern Australia. The primary difference between the two operating models is the number of spatial areas (zones) represented (one or two). There is a non-trawl fleet and a trawl fleet in each zone. The operating model covers a 43-year period (nominally '1970 "to '2012"), and when there are two zones they contain the same biomass when they were at unfished equilibrium (hence the 0.5 in Eq. (2). The populations in each zone in the two-zone operating model are independent, except for recruitment of age-0 animals, which depends on the total (over zones) spawning biomass.

[1] The naming convention for these methods was selected to match that in the r4ss package (Taylor et al., 2014).

Given a Beverton-Holt stock-recruitment relationship, the recruitment (at age-0) to zone z at the start of year y, R_y^z, is given by:

$$R_y^z = 0.5\frac{4hR_0\tilde{S}_y/\tilde{S}_0}{(1-h)+(5h-1)\tilde{S}_y/\tilde{S}_0}e^{\varepsilon_y-\sigma_R^2/2}; \quad \varepsilon_y \sim \mathrm{N}(0;\sigma_R^2) \tag{2}$$

where h is the "steepness" of the stock-recruitment relationship (Francis, 1992), R_0 is the (total over zones) unfished equilibrium recruitment, $\tilde{S}_y$ is total (over zones) spawning biomass, $\tilde{S}_0$ is the unfished total spawning biomass, and σ_R is the standard deviation among recruitment deviations in log space. Spawning biomass is defined as:

$$\tilde{S}_y = \sum_z\sum_a O_a^z N_{y,a}^{\mathrm{fem},z} \tag{3}$$

where O_a^z is the product of maturity-at-age and weight-at-age (see Methot and Wetzel (2013) for details of how O_a^z is calculated), and $N_{y,a}^{\mathrm{fem},z}$ is the number of females of age a in zone z at the start of year y.

The value for h is set to 0.75 and that for σ_R to 0.7 (Whitten and Punt, 2014). Selectivity is a logistic function of length for the non-trawl fleet and a double normal function of length for the trawl fleet (Fig. 1a). Growth, maturation, and natural mortality are assumed to be the same in each zone in the operating models that consider two zones.

Fig. 1b shows the time-trend in fishing mortality for the single-area operating model while Fig. 1c and d, shows the time-trends in fishing mortality by gear and zone for the two-zone operating model in which fishery mortality varies spatially. The maximum level of fishing mortality is assumed to be the same spatially, while the fully-selected fishing mortality for the non-trawl fleet is assumed to be half that for the trawl fleet (Fig. 1b–d). The fishing mortality by year, gear and zone is the value in Fig. 1b–d multiplied by a scalar selected so that total (over zones) spawning biomass is 40% of its unfished level in 2012 (the last year of the simulated period).

Catch, catch-rate and length-composition data are assumed to be available for all years and fleets, while conditional age-at-length data are assumed to be available for a random sample of half of the combinations of year and fleet. The catch-rate data for each zone (or the only zone for a single-zone operating model) are log-normally distributed with a coefficient of variation of 0.1. The observed catch length-composition data for a year-fleet-zone combination is a Dirichlet sample from the true catch length-composition, with an effective sample size of 100. This is lower than the actual sample sizes for pink ling, but the length-composition data for pink ling are known to be over-dispersed, and these data are considerably down-weighted to reflect this when assessments of this stock are conducted (Whitten and Punt, 2014). Thus, there should be no correlation between the length and age residuals for the case in which there is only one zone. However, two zones will lead to correlations between the length and age residuals given the model is mis-specified.

The length-composition data are assumed to be unsexed, as is the case in reality for pink ling and most fish stocks. The age-length keys are based on a sample size of 500 (by sex). Given a year-gear-zone, the age data are assumed to be a simple random sample from the catches by age and length, as is the intent of the sampling program used in reality to collect age data for this stock. The age-estimates are not subject to age-reading error.

2.3. The estimation method

The estimation method is based on Stock Synthesis (Methot, 1990; Methot and Wetzel, 2013). It ignores spatial differences in population structure and abundance (i.e., the NSWA approach of Punt et al., 2016a, 2015, 2017 is applied), and combines the data spatially as follows when the operating model contains two zones:

- The catch data are summed over zones.
- The catch-rate data are aggregated across zones, defining the catch-rate for year y as total catch for year y divided by the total effort for year y (i.e., the catch-rate data by zone are weighted by the effort by zone).
- The annual trawl and non-trawl catch length-composition data by fleet are pooled over zones, weighting the length-composition data for each zone by the annual catch by the zone.
- The annual age-length keys (i.e., the conditional age-at-length data) are summed over zones (without catch weighting). This reflects how data have been aggregated in actual assessments for pink ling (Whitten and Punt, 2014).

The estimation method estimates unfished recruitment, natural mortality (assumed to be the same for males and females), growth by sex (five parameters per sex: the parameters that govern von Bertalanffy growth and the CVs of length-at-age for ages of 1 and 20), length-specific selectivity parameters (logistic for the trawl fleet, double-normal for the non-trawl fleet, i.e., based on the correct selectivity patterns in the operating models; Fig. 1a), catchability for the CPUE indices, and annual recruitment values. Recruitment estimation accounts for a bias-ramp (Methot and Taylor, 2011) to avoid bias when estimating the deviations about the stock-recruitment relationship. The estimation methods are assumed to know the correct functional forms for the stock-recruitment relationship and the growth curve, the true value of steepness and the true value for the extent of variation about the stock-recruitment relationship. This is to ensure that the focus of the study is on the effects of data weighting, given it is well-known that it is hard to estimate steepness (Conn et al., 2010; Lee et al., 2012).

2.4. Scenarios

2.4.1. Performance statistics

The estimation method provides many outputs (Methot and Wetzel, 2013). However, the focus for this paper are the estimates of initial total (female) spawning biomass[2] (B_0), final (2012) total spawning biomass (B_{CUR}) and relative spawning biomass (B_{CUR}/B_0). Within Australia and the U.S., biomass in absolute terms directly impacts estimates of sustainable catch limits, while relative spawning biomass is used to assess stock status relative to biological reference points. The results of the simulations are summarized by relative error distributions as well as the median over simulations of the absolute relative errors (MARE). The relative error for a given quantity is the estimated value of the quantity less its true value divided by its true value and multiplied by 100, i.e., a positive value indicates an overestimate of the quantity and vice versa.

The performance of the tuning methods is also evaluated by how close the tuning method is able to correctly calculate the true overdispersion factor (the ratio of the actual sample size used to generate the data and input effective sample size), with performance summarized using a discrepancy metric: the square root of average of the squared differences between the tuned value for the overdispersion factor and the true value.

[2] Total spawning biomass is spawning biomass summed over zone.

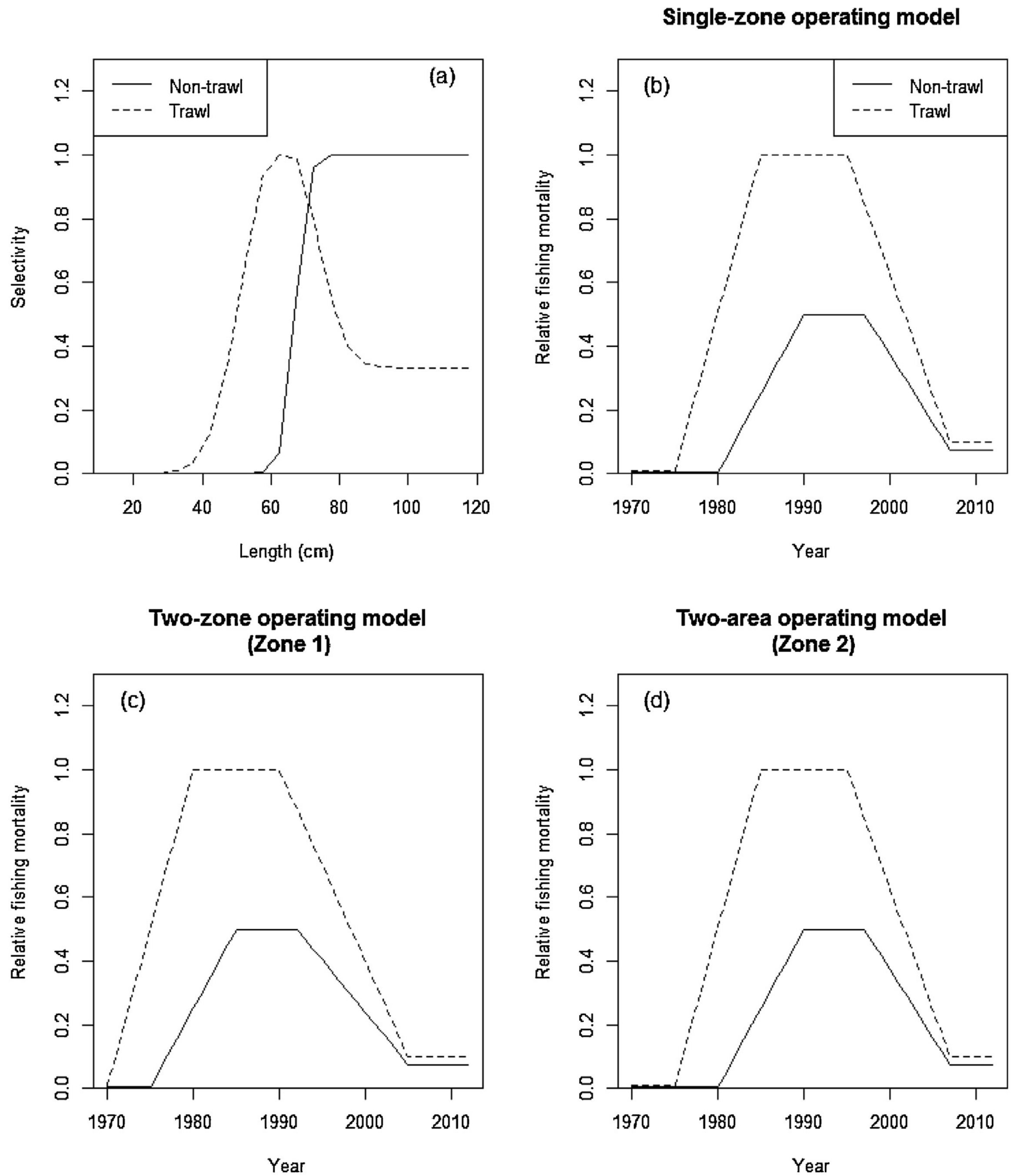

Fig. 1. Selectivity (a), fishing mortality by gear for the single-zone operating model (and the two zone-operating model with no spatial variation in fishing mortality) (b), and fishing mortality by gear and zone for the two-zone operating model with spatial variation in fishing mortality (c and d).

2.4.2. Single-zone operating model

There is no model mis-specification for the single-zone operating model so the estimation method would be expected to provide unbiased estimates of spawning biomass irrespective of the weights assigned to the various data sources. Moreover, given that the estimation model is not mis-specified, the standard deviation assigned to the CPUE indices should be 0.1 while the overdispersion factors to multiply the number of animals sized and aged should be 1; analyses are conducted to confirm this. The first set of analyses involves setting the initial ("input") effective sample sizes to the true effective sample sizes and 20% and 400% of the true effective sample sizes. Results are provided for each set of initial effective sample sizes when the effective sample sizes used when applying the estimation methods are set to the initial effective sample sizes (hence referred to as "untuned") as well as when the McAllister-Ianelli-1 and McAllister-Ianelli-2 methods (Table 2) are used to tune the weights for the length-composition and conditional age-at-length data. These two methods were selected because they are respectively the worst and best at estimating effective sample sizes for the single-zone operating model (results shown below).

Although the estimation method would be expected to provide unbiased estimates of spawning biomass irrespective of the initial effective sample sizes and the methods used to tune the residual variance for the CPUE data and effective sample sizes for the compositional data, there are aspects of the implementation (e.g., the bias-correction ramp) that may mean that some tuning methods may be better than others. The estimates of initial and final total spawning biomass and the ratio of final to initial spawning biomass are compared with the true values when each of five tuning methods are applied (the estimation method is applied six times, first with the initial effective sample sizes and then using tuned residual standard deviations and effective sample sizes based on the selected tuning method):

- McAllister-Ianelli-1: This method involves using Eq. (1) to tune the standard deviations for the CPUE indices, and using Eqs. (1.A) and (2.A) in Table 2 to respectively tune the effective sample sizes for the length-composition and conditional age-at-length data.
- McAllister-Ianelli-2: As for McAllister-Ianelli-1, except that the effective sample sizes for the length-composition and conditional age-at-length data are tuned using Eqs. (1.B) and (2.B) in Table 2.
- Francis/Francis-B: As for McAllister-Ianelli-1, except that the effective sample sizes for the length-composition and conditional age-at-length data are tuned using Eqs. (1.C) and (2.C) in Table 2.
- Francis/Francis-A: As for Francis/Francis-B, except that the effective sample sizes for the conditional age-at-length data are tuned using Eq. (2.D) in Table 2.
- Francis/Francis-B*: As for Francis/Francis-B, except that the residual standard deviations for the CPUE indices are set to its initial value of 0.1 until the last tuning step takes place. This tuning method mimics that approach actually used to conduct assessments for pink ling off southeastern Australia (Whitten and Punt, 2014).

2.4.3. Two-zone operating model

The variant of the zone-area operating model in which the trends in fishing mortality are the same for the two zones is essentially identical to the single-area operating model. In contrast to the situation for the two-zone operating model in which the trends in fishing mortality are the same for the two zones, allowing for different time-trends in fishing mortality between zones (Fig. 1c and d) leads to model mis-specification because the data from the operating model are analysed using an estimation method that assumes the population being assessed is homogenously distributed across the assessed area. The five tuning methods are evaluated in terms of which method leads to the lowest values for the MAREs for initial and final total spawning biomass and for the ratio of final to initial total spawning biomass. The correct values for the effective sample sizes cannot be determined analytically for the two-zone operating model when fishing mortality differs between the two zones, so the qualitative changes to the values for the effective sample sizes are explored for each tuning method.

3. Results

3.1. Single area analyses

Fig. 2 shows time-trajectories of relative errors for the estimates of spawning biomass for the single-area operating model. Nine scenarios are illustrated that explore assumptions related to the initial effective sample sizes for the length-composition and conditional age-at-length data (20% of the true values, the true values, and 400% of the true values) and the method used to tune the effective sample sizes and the residual variances for the CPUE indices (none, McAllister-Ianelli-1, and McAllister-Ianelli-2). The results for the cases in which weights are tuned are shown after the tuning method has been applied five times. As expected, all nine assessment options provide estimates that are essentially equal to the true values in median terms (Fig. 2). However, the distributions of relative error differ among the nine assessment options. The distributions are most narrow when the initial effective sample sizes for the compositional data are four times their true values and the effective sample sizes are not tuned (Fig. 2, lower left panel). The distributions of relative error get narrower toward the end of the assessment period when no tuning occurs or when the McAllister-Ianelli-2 method is applied, irrespective of the initial effective sample sizes.

The nine assessment options perform about equally well for the three quantities of management interest (Table 3). However, variants of the untuned method perform best for all three quantities, although setting the effective sample sizes (and the residual variances for the CPUE data) to their true values is not the best option for any of the quantities (see the underlined values in Table 3 for the methods that lead to the lowest MAREs).

Table 4 compares the untuned approach and the five tuning methods in terms of the MAREs for the three management quantities when the tuning methods start at the true effective sample sizes, and Fig. 3 compares the associated relative error distributions for spawning biomass. The tuning methods apart from the McAllister-Ianelli-1 tuning method lead to less between-simulation variation in relative error over time. Setting the effective sample sizes to their true values (untuned) and all of the tuning methods except the McAllister-Ianelli-1 tuning method lead to very similar values for the MAREs for the three quantities of management interest, although the Francis/Francis-B and Francis/Francis-B* methods are most consistently among the best of the tuning methods.

Table 5 compares the methods for calculating effective sample sizes for the compositional data in terms of the discrepancy metric[3] when the assessments are based on each of the five tuning methods. The McAllister-Ianelli-2 method is best in terms of the discrepancy metric in calculating effective sample sizes for the length-composition data (Table 5), and this result holds for both fleets, irrespective of how the weightings are actually set when conducting the assessment. The McAllister-Ianelli-1 method leads to the largest values for the discrepancy metric for the length-composition data, primarily because it is biased, i.e. the expected calculated effective sample size differs from the true value (Fig. 4). The Francis method is unbiased, but exhibits fairly high between-simulation variation (Fig. 4), which leads to a higher value for the discrepancy metric.

The McAllister-Ianelli-1 method is again the poorest method when performance is relative to the ability to estimate an effective sample size for conditional age-at-length data (Fig. 4; Table 5), but unlike the situation for length-composition data, the performance of this method differs between the two fleets. The McAllister-Ianelli-2 method is again best (lowest discrepancy metric) for the conditional age-at-length data, but the Francis-B method is nearly as good. The Francis-A method is unbiased, but like the Francis method for length-composition data, exhibits higher between-simulation variation (Fig. 4)

3.2. Two-zone operating model with no spatial variation in fishing mortality

The results are as expected for this operating model given its lack of major model mis-specification (there is some minor model mis-specification of error structure given the way the CPUE indices and length-composition data are constructed from the zone-specific data). Specifically, the tuning methods are about equally good, except that McAllister-Ianelli-1 is poorer than the rest (Table 4 centre panels). The MAREs are lower for the two-zone operating model

[3] The square root of average of the squared difference between the tuned value for the overdispersion factor and the true value.

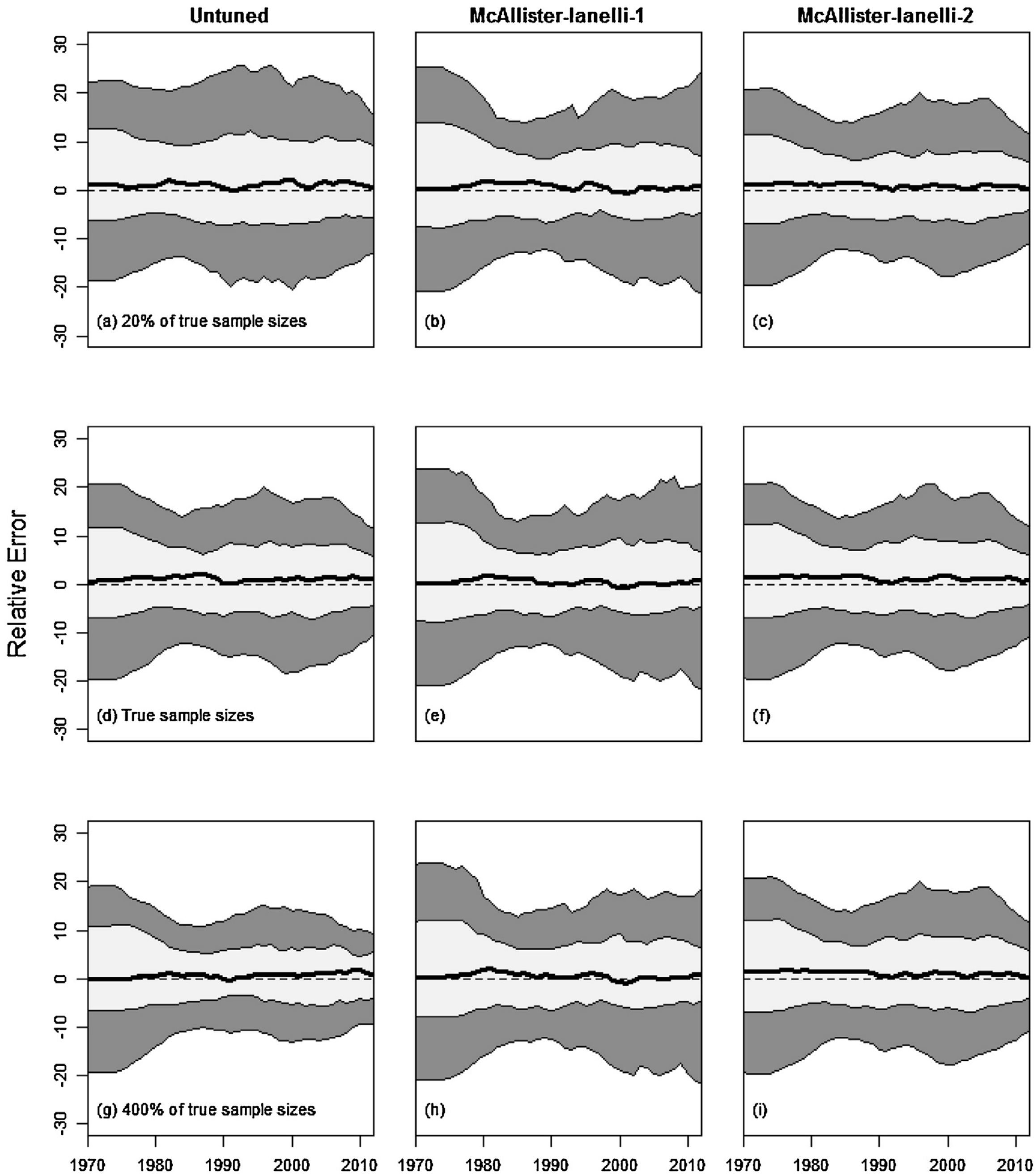

Fig. 2. Relative error distributions (median relative errors, and 50% and 90% intervals) for the time-trajectory of total spawning biomass for a single-area operating model. Results are shown for three initial sample sizes for the compositional data (rows) and when the residual variances for the CPUE data and the effective sample size are tuned using the McAllister and Ianelli-1 and McAllister and Ianelli-2 methods (columns 1–3).

Table 3
Median (over simulations) absolute relative errors for three quantities of management interest for the single-zone operating model. Results are shown for three initial sample sizes for the compositional data, and when the residual variance for the CPUE data and the effective sample sizes are tuned using the McAllister and Ianelli-1 and McAllister and Ianelli-2 methods. The scenarios that achieve the lowest median absolute relative errors by management quantoty are indicated in bold underline.

Initial input sample sizes	Initial spawning biomass	Final spawning biomass	Final to initial spawning biomass	Initial spawning biomass	Final spawning biomass	Final to initial spawning biomass	Initial spawning biomass	Final spawning biomass	Final to initial spawning biomass
	Untuned			McAllister-Ianelli-1			McAllister-Ianelli-2		
20% of the true sample sizes	**<u>9.29</u>**	7.00	**<u>3.42</u>**	11.21	6.13	6.64	9.37	5.14	5.02
True sample sizes	9.30	5.22	5.30	10.84	6.14	6.54	9.75	5.33	5.03
400% of the true sample sizes	9.80	**<u>4.48</u>**	7.91	10.21	6.00	6.36	9.53	5.22	5.03

Table 4
Median (over simulations) absolute relative errors for three quantities of management interest. The assessment methods that achieve the lowest median absolute relative errors by management quantity for each operating model are indicated in bold underline. See Section 2.4.2 and Table 2 for a description of the tuning methods.

Initial sample sizes	Single-zone operating model			Two-zone operating model (no spatial variation in fishing mortality)			Two-zone operating model (with spatial variation in fishing mortality)		
	Initial spawning biomass	Final spawning biomass	Final to initial spawning biomass	Initial spawning biomass	Final spawning biomass	Final to initial spawning biomass	Initial spawning biomass	Final spawning biomass	Final to initial spawning biomass
Untuned	**9.30**	5.22	5.30	8.88	4.36	4.64	7.14	14.73	10.27
McAllister-Ianelli-1	10.84	6.14	6.54	9.10	4.77	6.35	**6.13**	15.68	14.57
McAllister-Ianelli-2	9.75	5.33	5.03	9.42	**4.15**	4.14	7.12	12.82	9.49
Francis/Francis-B	9.35	**4.98**	4.99	9.38	4.39	**4.07**	7.38	12.72	9.19
Francis/Francis-A	9.41	5.33	**4.91**	8.93	4.46	4.14	7.48	**11.50**	**8.37**
Francis/Francis-B*	9.44	**4.98**	4.98	**8.67**	4.28	4.09	7.33	12.89	9.02

Table 5
Square root of the average of the squared differences between the true and computed values for the effective sample sizes by fleet and data type for the single-area operating model. Results are shown for various approaches to computing effective sample sizes and when the sample sizes in the assessment are tuned using various approaches. Results reflect computed effective sample sizes after five estimation-tuned steps.

Fleet/ tuning approach	Length-composition data			Conditional age-at-length data			
	McAllister-Ianelli-1	McAllister-Ianelli-2	Francis	McAllister-Ianelli-1	McAllister-Ianelli-2	Francis-B	Francis-A
Non-trawl fleet							
McAllister-Ianelli-1	0.417	0.119	0.276	0.710	0.029	0.068	0.218
McAllister-Ianelli-2	0.419	0.087	0.299	0.736	0.031	0.068	0.234
Francis/Francis-B	0.419	0.087	0.312	0.736	0.031	0.071	0.238
Francis/Francis-A	0.419	0.087	0.311	0.731	0.029	0.069	0.252
Francis/Francis-B*	0.419	0.087	0.310	0.736	0.031	0.070	0.236
Trawl fleet							
McAllister-Ianelli-1	0.357	0.106	0.252	17.446	0.066	0.088	0.237
McAllister-Ianelli-2	0.370	0.087	0.254	13.930	0.067	0.088	0.210
Francis/Francis-B	0.369	0.098	0.262	13.263	0.066	0.090	0.210
Francis/Francis-A	0.369	0.087	0.263	12.989	0.064	0.087	0.219
Francis/Francis-B*	0.370	0.087	0.263	13.395	0.067	0.090	0.211

Table 6
Square root of the average of the squared differences between the true and computed values for the effective sample sizes by fleet and data type for the two-zone operating model with spatially-invariant fishing mortality. Results are shown for various approaches to computing effective sample sizes and when the sample sizes in the assessment are tuned using various approaches. Results reflect computed effective sample sizes after five estimation-tuned steps.

Fleet/ tuning approach	Length-composition data			Conditional age-at-length data			
	McAllister-Ianelli-1	McAllister-Ianelli-2	Francis	McAllister-Ianelli-1	McAllister-Ianelli-2	Francis-B	Francis-A
Non-trawl flee							
McAllister-Ianelli-1	0.441	0.095	0.255	0.736	0.025	0.065	0.225
McAllister-Ianelli-2	0.455	0.111	0.272	0.750	0.023	0.061	0.207
Francis/Francis-B	0.459	0.097	0.286	0.749	0.023	0.067	0.207
Francis/Francis-A	0.459	0.097	0.284	0.746	0.022	0.065	0.214
Francis/Francis-B*	0.460	0.097	0.289	0.749	0.023	0.067	0.209
Trawl fleet							
McAllister-Ianelli-1	0.342	0.130	0.241	12.754	0.058	0.070	0.208
McAllister-Ianelli-2	0.357	0.089	0.248	17.374	0.060	0.072	0.196
Francis/Francis-B	0.366	0.128	0.270	15.526	0.064	0.077	0.197
Francis/Francis-A	0.367	0.129	0.269	13.806	0.064	0.077	0.199
Francis/Francis-B*	0.358	0.089	0.256	17.122	0.059	0.072	0.197

than for the single-zone operating model because there are more data for the two-zone operating model.

The qualitative ranking of the methods for tuning the length-composition and the conditional age-at-length are similar for two-zone operating model with no spatial variation in fishing mortality and the single-zone operating model (Tables 5 and 6). The ability to estimate the effective sample sizes is generally better for the McAllister-Ianelli-2 method (length-composition data) and McAllister-Ianelli-2 and Francis-B methods (conditional age-at-length data) for the two-zone operating model with no spatial variation in fishing mortality.

3.3. Two-zone operating model with spatial variation in fishing mortality

The presence of spatial variation in fishing mortality leads to model mis-specification and marked changes in the ability to estimate spawning biomass and the management quantities. The estimates of biomass are initially positively biased, then become negatively biased and become positively biased again, irrespective of the tuning method applied (Fig. 5). As expected from Fig. 5, the MAREs for final spawning biomass and for the ratio of final to initial spawning biomass are much higher for the variant of the two-zone operating model with spatial variation in fishing mortality (compare the rightmost three columns of Table 4 with the centre three columns). The McAllister-Ianelli-1 tuning method is again the worst of those considered, but unlike the situation for single-zone operating model and two-zone with no spatial variation in fishing mortality, the remaining tuning methods outperform the "untuned" approach, suggesting that tuning can lead to improved estimation performance, at least when the assessment model is mis-specified. Unfortunately, none of the tuning methods lead to unbiased estimates of spawning biomass (Fig. 5). Of the tuning methods, the Francis/Francis-A tuning method performs best in

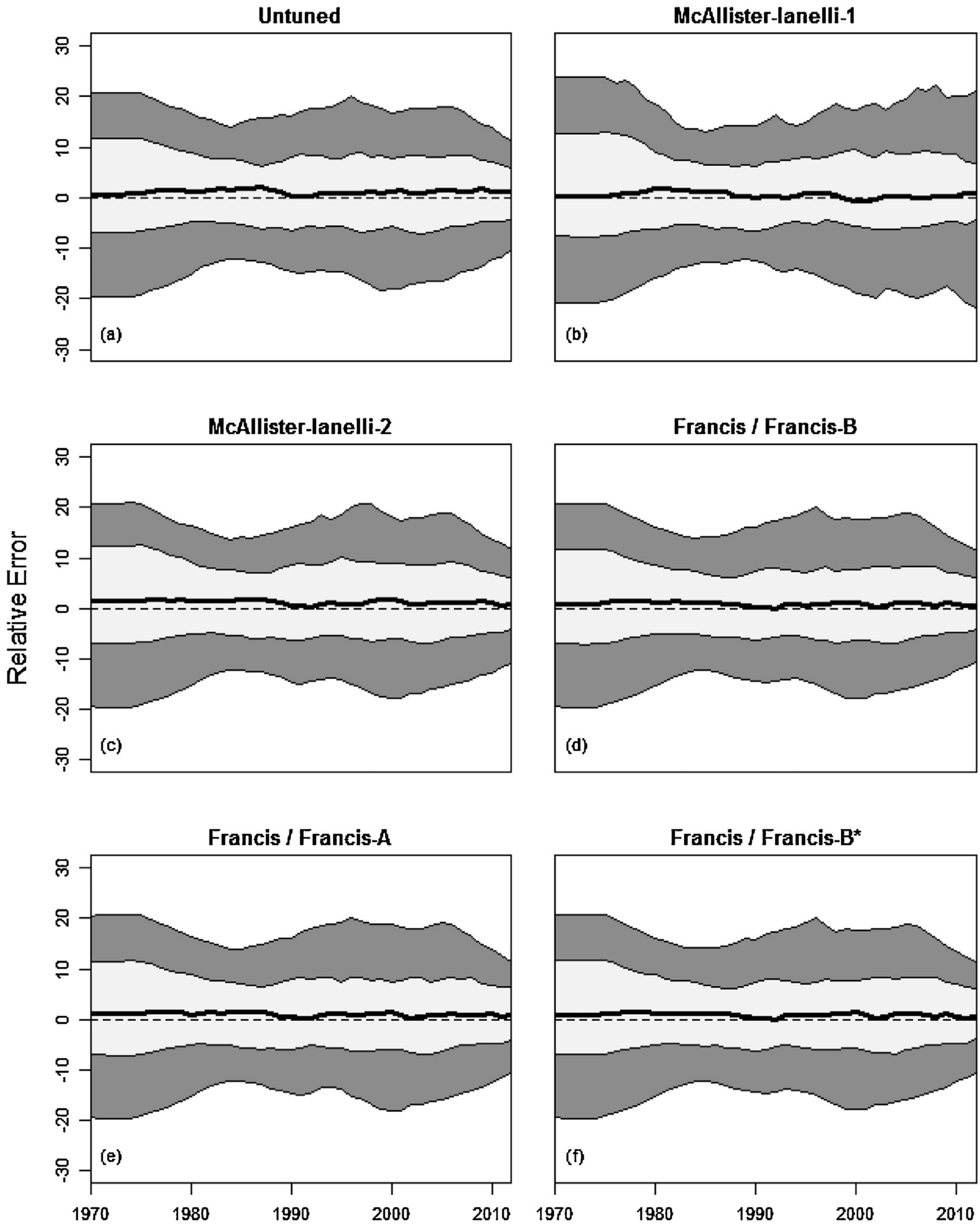

Fig. 3. Relative error distributions (median relative errors, and 50% and 90% intervals) for the time-trajectory of total spawning biomass for the single-area operating model. Results are shown for setting the effective sample sizes for the CPUE and the compositional data to the true values ("Untuned") and based on five tuning methods. The initial effective sample sizes are set equal to the true values.

terms of estimating final spawning biomass and estimating the ratio of final to initial spawning biomass (Table 4).

It is impossible to determine the correct values for the effective sample sizes so it is not possible to compute the discrepancy metric for the tuning methods for this operating model. However, Fig. 6 shows the distributions for calculated ratios of tuned to input effective sample sizes when the Francis/Francis-A tuning method is used to conduct the assessment. Compared to the single-zone operating model and the two-zone operating model with no spatial variation in fishing mortality, the tuned residual standard variation for the trawl CPUE is lower (Figs. 4 and 6, left panel in the third row of panels). The calculated effective sample sizes from the McAllister-Ianelli-1 method are again larger than the input effective sample sizes. However, the calculated effective

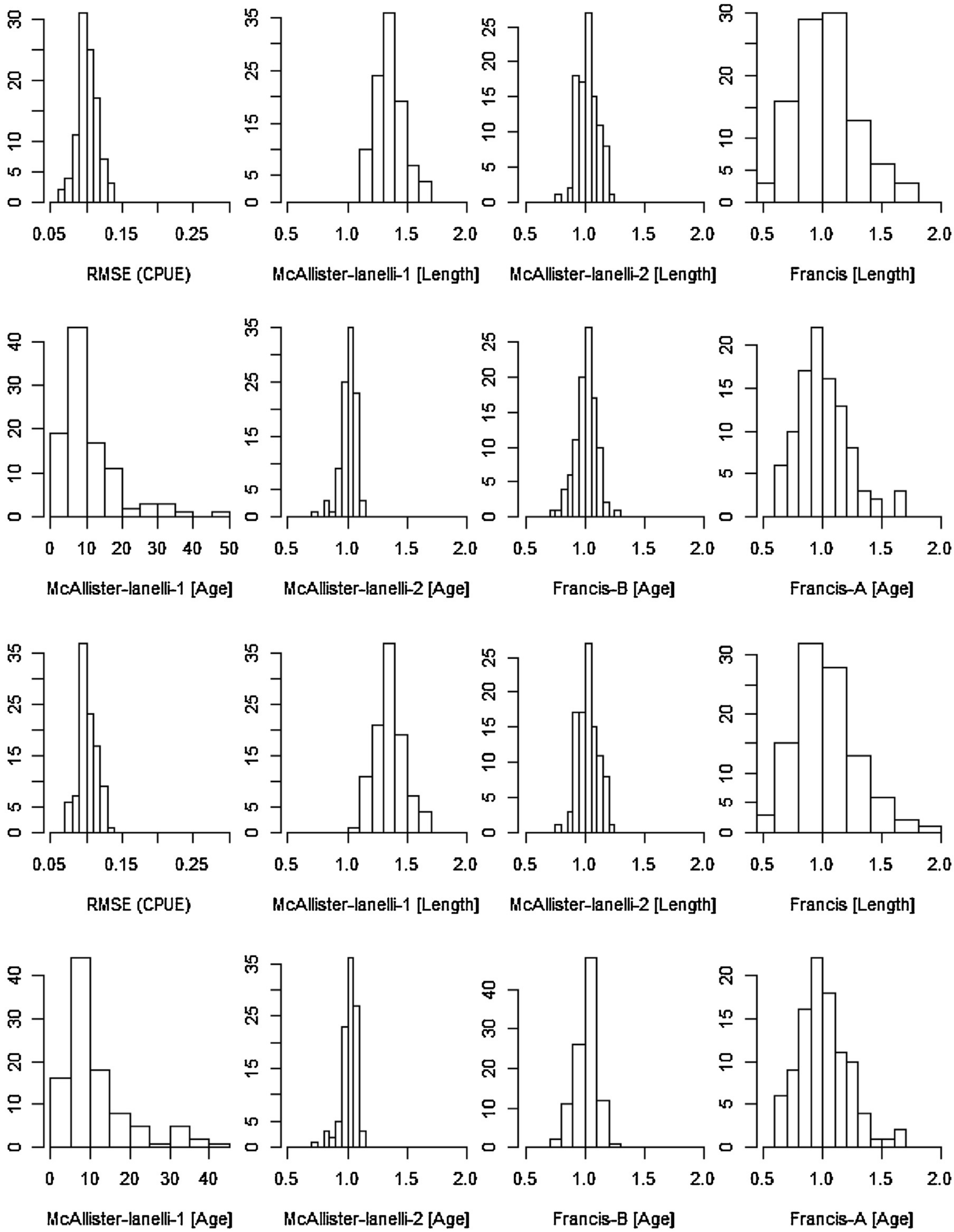

Fig. 4. Estimated standard deviations for the CPUE data (upper left panel) and the effective sample sizes for the length and conditional age-at-length data based on various tuning methods relative to the true effective sample size (other panels; see Tables 1 and 2 for details). Results are shown in the upper two rows of panels for the non-trawl fleet and in the lower two rows of panels for the trawl fleet. The operating model is the single-area model so the true values are 0.1 for the standard deviation of the CPUE data and 1 for the effective sample sizes for the length and conditional age-at-length data.

sample sizes from the other methods are on average smaller (often much smaller) than the input effective sample sizes. The calculated effective sample sizes from the McAllister-Ianelli-2 and Francis methods for length-composition data are on average 87% and 80% respectively of the corresponding input values while the average percentages for conditional age-at-length data are 82%, 66% and 23% for the non-trawl fleet and 92%, 83% and 46% for the trawl fleet respectively for the McAllister-Ianelli-2, Francis-B and Francis-A methods respectively.

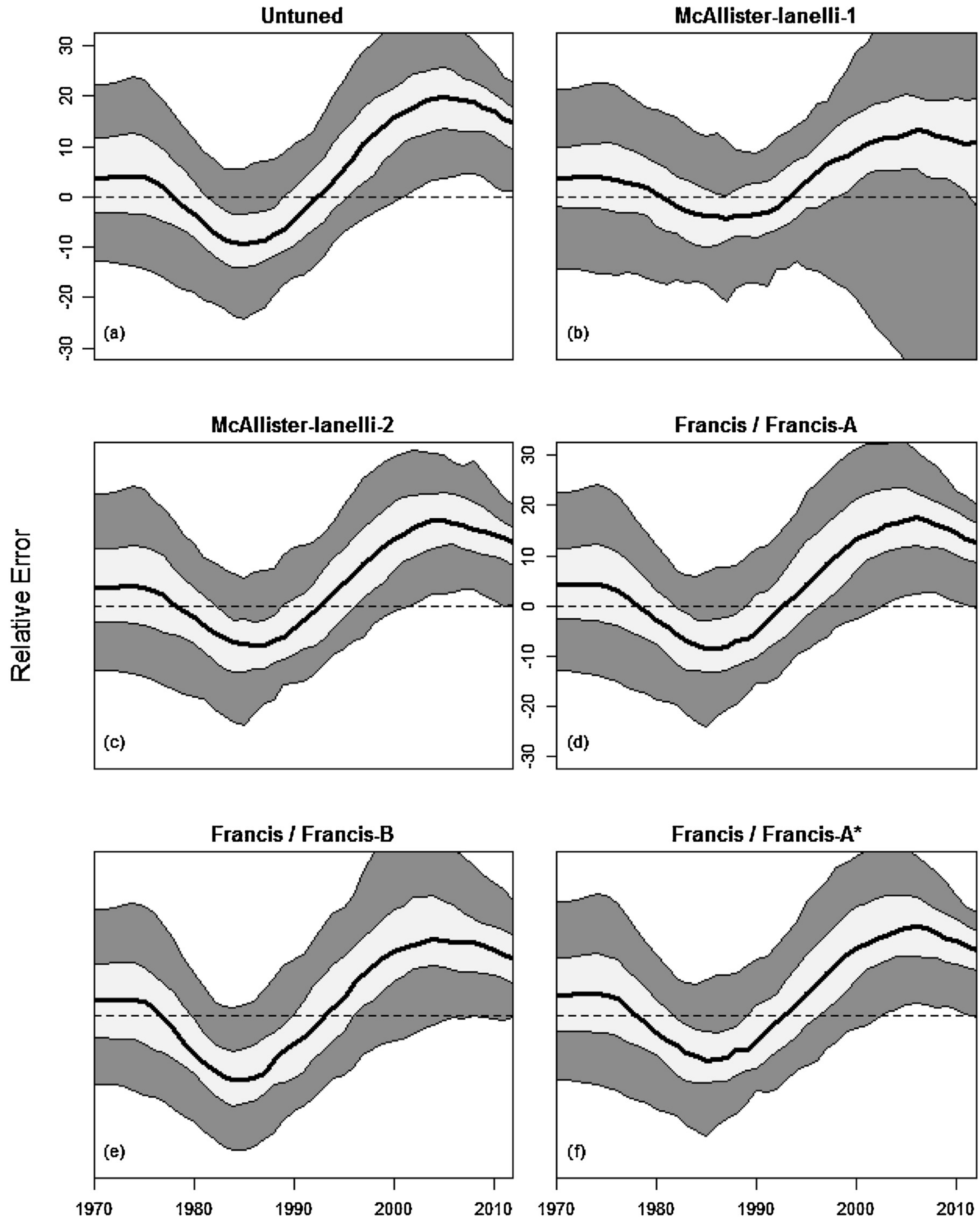

Fig. 5. Relative error distributions (median relative errors, and 50% and 90% intervals) for the time-trajectory of total spawning biomass for a two-zone operating model in which there is spatial variation in fishing mortality. Results are shown for setting the effective sample sizes for the CPUE and compositional data to the true values for two-zone operating model with no spatial variation in fishing mortality ("Untuned") and based on five tuning methods. The initial effective sample sizes are set equal to the true values for two-zone operating model with no spatial variation in fishing mortality.

3.4. Do results from tuning help detect model mis-specification

The estimates of spawning biomass differ when the McAllister-Ianelli-1 and (close to best) Francis/Francis-A tuning method are used (Fig. 5; Table 4), which raises the question whether differences in biomass estimates from different tuning methods allow stock assessment errors to be detected. Fig. 7 therefore shows the estimates of the three management quantities when the Francis/Francis-A tuning method is used to tune the weights, against the difference in the estimates of these quantities between the McAllister-Ianelli-1 and Francis/Francis-A tuning methods.

There is a weak positive relationship between the absolute value of the error in estimating initial spawning biomass using the Francis/Francis-A tuning method and the absolute value of the difference in estimates from Francis/Francis-A and McAllister-Ianelli-1 tuning methods (Fig. 7). However, this relationship has

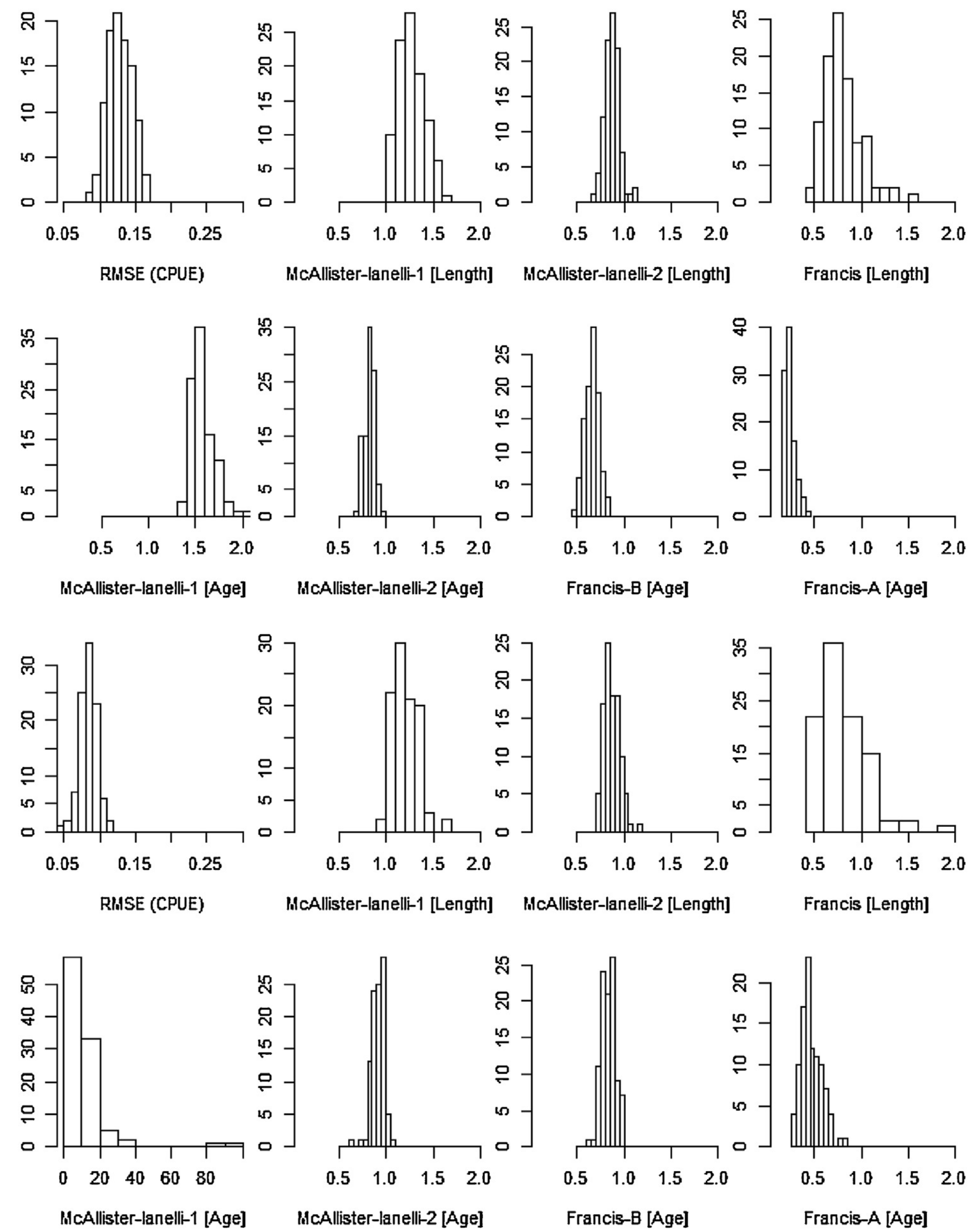

Fig. 6. Estimated standard deviations for the CPUE data (upper left panel) and the ratio of the calculated to the input effective sample sizes for the length and conditional age-at-length data based on various tuning methods (other panels; see Tables 1 and 2 for details). Results are shown in the upper two rows of panels for the non-trawl fleet and in the lower two rows of panels for the trawl fleet. The operating model is the two-zone model with spatial variation in fishing mortality.

almost no predictive value ($R^2 < 0.05$). There is no relationship apparent in Fig. 7 for final spawning biomass and the ratio of final to initial spawning biomass.

4. Discussion

There are many ways to assign weights to index and composition data, although there is less disagreement on how to assign weights to index data than to composition data. The results of stock assessments are known from actual assessments, and this study, to be sensitive to these weights in some cases, making the choice of how to weight data sources an important aspect of a stock assessment. However, as shown here, it is not uncommon for there to be considerable agreement between assessment outputs from different tuning methods.

4.1. Selection of a best method

There was no tuning method that was uniformly the best of those considered, although the McAllister-Ianelli-1 tuning method had poorer performance than the other methods irrespective of whether it was used for calculating effective sample sizes for length-composition or conditional age-at-length data. This results is consistent with results obtained by Stewart and Hamel (2014). The results suggest that there is value applying a tuning method because the quantities of management interest were estimated better when there was model mis-specification, even though applying tuning methods did not resolve the issue of model mis-specification—this could only have been adequately addressed by using a population model that had the “correct” spatial structure (Punt et al., 2015, 2017, 2016b).

The McAllister-Ianelli-2 methods consistently performed best when the assessment model was not mis-specified. This method

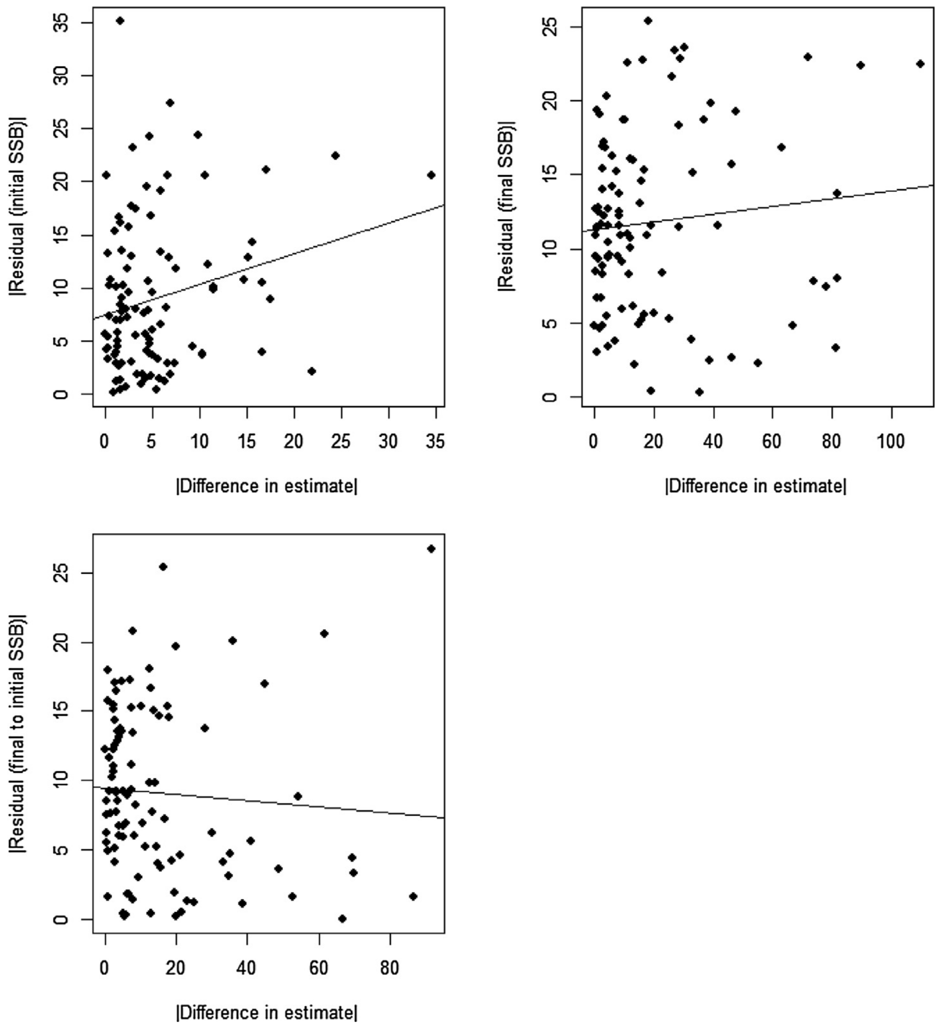

Fig. 7. Absolute relative errors for initial spawning biomass, final spawning biomass, and the ratio of final to initial spawning biomass from the Francis/Francis-A tuning method plotted against the absolute value of the difference between the estimate from the Francis/Francis-A tuning method and that from the McAllister-Ianelli-1 tuning method (expressed as percentages of the true value). The results in the figure are based on the two-zone operating model with spatially-variable fishing mortality.

led to lower (often markedly) effective sample sizes for the compositional data than the McAllister-Ianelli-1 method, and is clearly to be preferred to the McAllister-Ianneli-1 method. The comparison between the McAllister-Ianelli-2 and the Francis methods (including the new methods introduced in this paper) is less clear. The discrepancy metric suggested that the Francis methods were less precise at estimating the ratios of the effective sample sizes to the input effective sample sizes (particularly the Francis-A method). This is perhaps not unexpected because (for conditional age-at-length data) the McAllister-Ianelli-2 method is based on effective sample sizes computed for each combination of year and length-class (as is the Francis-B method) whereas the Francis-A method is based on one standardized difference between the observed and model-predicted mean age for each year. However, in the face of model-mis-specification, the Francis and Francis-A methods led to notably lower effective sample sizes (Fig. 6).

The Francis-A method leads to the lowest effective sample sizes for the conditional age-at-length data[4]. This is unsurprising because this method accounts for the correlation in the errors in estimated age-composition between length-classes unlike the McAllister-Ianelli-1 and Francis-B methods. The Francis-A method accounts for most of the systematic errors amongst the tuning methods considered and its performance was 'best' (although not by much for the case in which the assessment model was mis-specified) and is hence preliminarily the best method.

[4] But less so than in the original software (Taylor et al., 2014) that has been developed to assist assessment authors conduct and evaluate stock assessments conducted using Stock Synthesis.

4.2. Caveats and future work

This study has not explored how to tune the residual variance for the index data in the same detail as it explored how to tune the effective sample sizes for the compositional data, but it provides little evidence of an advantage of tuning this residual variance at all steps of the iterative tuning process rather than tuning it only at the last step.

The results of this study were markedly different between the case in which there was model mis-specification and when the estimation model closely matched the operating model. This further highlights the need to carefully account for model mis-specification in simulation studies (Francis, 2012). This study generated correlated residuals by applying a spatially-aggregated stock assessment method to data generated from a two-zone operating model in which population structure differs spatially. An alternative way to allow for correlations between adjacent length-classes in the length-composition data would be generate the data using, for example, a logistic multivariate normal distribution. This approach could be used to explore within-year correlations separately from between-year correlations, which was not possible in this study.

Occasionally, the values for weights change monotonically every time the tuning method is applied so that the weights do not converge with each successful application of the tuning method (A.E. Punt, unpublished data). However, this pathological behavior was not observed in the simulations of this paper and the causes of this behavior remain unknown.

There has been limited evaluation of the performance of tuning methods for stock assessments using simulation (essentially this study and Maunder (2011), which explored alternative likelihood functions). However, selection of tuning methods is not the only way data are weighted. Another important aspect of data weighting relates to how to select the weights for each year within a time-series (see Stewart and Hamel (2014) for alternative approaches for selecting these weights) as this paper only considered the weight to be assigned to an entire data series. Similarly, the choice of the likelihood function for the index and compositional data is a form of data weighting (Quinn and Deriso, 1999; Maunder, 2011). Several alternative choices for the likelihood function are available. Some of these likelihood functions, such as the robust normal for proportions (Fournier et al., 1998) and the Punt-Kennedy likelihood function (Punt and Kennedy, 1997) are self-weighting.

This paper has focused on index and compositional data, but there are other aspects of an assessment that need weighting, including other data types such as mark-recapture data when these data are used within an assessment to estimate mortality or movement (Punt et al. 2015, 2017). In addition, most integrated assessments include priors and penalties, and how these are specified can impact the outcomes of assessments, particularly in data-poor cases. In addition, how process error is modelled, including whether the extent of process error is estimated or pre-specified can have an important impact on assessment outcomes and is part of "data weighting".

4.3. Final comment

This paper has focused on automatic methods for tuning the weights assigned to various data sources. However, data weighting in assessments almost always has a non-automated component. This non-automated aspect to data weighting includes deciding which data series (usually CPUE series) to include in the assessment, and perhaps the even more extreme selection of the type of model to use and hence whether some data sources are to be considered at all. For example, basing an assessment on a biomass dynamics model implicitly assigns many data types (e.g., length-composition data, conditional age-at-length data, tagging data) zero weight even if those data sources provide information on stock status and productivity. This highlights the need to clearly document how the model used for the assessment was chosen, as well as the basis for selecting initial weights and how those weights are modified.

Acknowledgments

Rich Little and Malcolm Haddon (CSIRO Ocean and Atmosphere) and two anonymous reviewers are thanked for their comments on an earlier version of this paper. This work was partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative agreement No. NA10OAR4320148, Contribution No. 2488.

References

Bull, B., Francis, R.I.C.C., Dunn, A., McKenzie, A., Gilbert, D.J., Smith, M.H., 2005. CA SAL (C++ algorithmic stock assessment laboratory): CA SAL user manual v2.07-2005/08/21, NIWA Technical Report 127.

Conn, P.B., Williams, E.H., Shertzer, K.W., 2010. When can we reliably estimate the productivity of fish stocks? Can. J. Fish. Aquat. Sci. 67, 511–523.

Deriso, R.B., Quinn II, T.J., Neal, P.R., 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42, 815–824.

Doubleday, W.G., 1976. A least squares approach to analyzing catch at age data. Int. Comm. Northw. Atl. Fish. Res. Bull. 12, 69–81.

Fournier, D., Archibald, C.P., 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39, 1195–1207.

Fournier, D.A., Hampton, J., Sibert, J.R., 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 55, 2105–2116.

Francis, R.I.C.C., 1992. Use of risk analysis to assess fishery management strategies: a case study using orange roughy (*Hoplosthethis atlanticus*) on the Chatham Rise, New Zealand. Can. J. Fish. Aquat. Sci. 49, 922–930.

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138.

Francis, R.I.C.C., 2012. The reliability of estimates of natural mortality from stock assessment models. Fish. Res. 119-120, 133–134.

Hampton, J., Fournier, D.A., 2001. A spatially disaggregated length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. Mar. Freshwatater Res. 52, 937–963.

Lee, H.-H., Maunder, M.N., Piner, K.R., Methot, R.D., 2012. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? Fish. Res. 125-126, 254–261.

Maunder, M.N., 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: estimating the effective sample size. Fish. Res. 109, 311–319.

Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142, 61–74.

McAllister, M.K., Ianelli, J.N., 1997. Bayesian stock assessment using catch-age data and the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 54, 284–300.

Methot, R.D., 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. Int. North. Pac. Fish. Comm. Bull. 50, 259–277.

Methot Jr., R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of esti-mated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. 68, 1744–1760.

Methot, R.D., Wetzel, C.R., 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142, 86–99.

Newman, K.B., Buckland, S.T., Morgan, B.J.T., King, R., Borchers, D.L., Cole, D.J., Beabeas, P., Gimenez, O., Thomas, L., 2014. Modeling Population Dynamics. Spinger, New York, xii+215 pp.

Punt, A.E., Kennedy, R.B., 1997. Population modelling of Tasmanian rock lobster, *Jasus edwardsii*, resources. Mar. Freshwater Res. 48, 967–980.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016a. Management strategy evaluation: best practices. Fish Fish. (in press).

Punt, A.E., Haddon, M., Little, L.R., Tuck, G.N., 2016b. Can a spatially-structured stock assessment address uncertainty due to closed areas? A case study based on pink ling in Australia. Fish. Res. 175, 10–23.

Punt, A.E., Deng, R.A., Siddeek, M.S.M., Buckworth, R.C., Vanek, V., 2017. Data weighting for tagging data in integrated size-structured models. Fish. Res. 192, 94–102.

Punt, A.E., Haddon, M., Tuck, G.N., 2015. Which assessment configurations perform best in the face of spatial heterogeneity in fishing mortality, growth and recruitment? A case study based on pink ling in Australia. Fish. Res. 168, 85–99.

Punt, A.E., Smith, D.C., Tuck, G.N., Methot, R.D., 2006. Including discard data in fisheries stock assessments: two case studies from South-eastern Australia. Fish. Res. 79, 239–250.

Quinn, T.J.I., Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press, New York, pp. 542.

Richards, L.J., 1991. Use of contradictory data sources in stock assessments. Fish. Res. 11, 225–238.

Stewart, I.J., Hamel, O.S., 2014. Bootstrapping of sample sizes for length- or age-composition data used in stock assessments. Can. J. Fish. Aquat. Sci. 71, 581–588.

Taylor, I., Stewart, I., Hicks, A., Garrison, T., Punt, A., Wallace, J., Wetzel, C., Thorson, J., Takeuchi, Y., Monnahan, C., 2014. Package r4ss. https://github.com/r4ss.

Whitten, A.E., Punt, A.E., 2014. Stock assessment of pink ling (*Genypterus blacodes*) using data up to 2012. In: Tuck, G.N. (Ed.), Stock Assessment for the Southern and Eastern Scalefish and Sharkfish Fishery 2013, Part 1, 014. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart, pp. 116–142.