Model selection for selectivity in fisheries stock assessments

Andre E. Punt, F. Hurtado-Ferro, F. and A.R. Whitten

SEDAR39-RD-09

September 2014

G Model
FISH-3658; No. of Pages 11

ARTICLE IN PRESS

Fisheries Research xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Fisheries Research

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

Model selection for selectivity in fisheries stock assessments

André E. Punt*, Felipe Hurtado-Ferro, Athol R. Whitten

School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195, United States

ARTICLE INFO

Article history:
Received 11 April 2013
Received in revised form 14 June 2013
Accepted 18 June 2013

Keywords:
AIC
Asymptotic selectivity
Dome-shaped selectivity
Information criteria model selection
Simulation

ABSTRACT

The choice of how to model selectivity differs among approaches to fisheries stock assessment. VPA tends to make only weak assumptions regarding (age-specific) selectivity (asymptotic selectivity and temporal stability of selectivity for the most recent years). In contrast, selectivity is more parametric in "integrated" methods, and can be age-, length-, and age- and length-based. The use of parametric selectivity functions tends to reduce estimation variation because fewer parameters have to be estimated, but incorrect choices for the functional form for selectivity can lead to bias. This paper illustrates effects of poor choices for selectivity on the outcomes of stock assessments, outlines methods for evaluating whether a particular choice for selectivity is appropriate using residual diagnostics, and summarizes current ways to select among alternative functional forms for selectivity. This paper also provides a synthesis of the results of past simulation studies which have explored the ability to correctly parameterize selectivity.

1. Introduction

A variety of methods have been developed to provide quantitative scientific advice to enable fisheries to be managed. These methods can be broadly categorized into yield-per-recruit methods, surplus production methods, and age- or length-structured assessment methods. The first and third of the methods rely on estimates of "selectivity" for their application. "Selectivity" has traditionally been defined in relation to fishing gear selectivity, i.e., the probability of capturing an animal given it encounters the gear, usually as a function of length. However, yield-per-recruit and age- or length-structured assessment methods treat "selectivity" for a "fleet" (see Section 4 for an operational definition of fleet) more generally, defining it (often implicitly) as the probability of a fleet capturing an animal of a given size (or age) relative to the probability of that fleet capturing an animal of a different size (or age), usually the size (or age) at which the probability of capture is highest. Thus, "selectivity" as defined in stock assessments, and in this paper, generally includes both the concepts of gear selectivity and availability (i.e., the probability that an animal is the area/at the correct time which would allow it to be captured). There are exceptions to this general practice. For example, Pribac et al. (2005) set the selectivity patterns for the fishing gears used to capture gummy shark, *Mustelus antarcticus*, based on experimental results (Kirkwood and Walker, 1986), and consequently explicitly estimate the length-specific probability of animals of each age being in the area of the fishery (i.e., "availability" and "selectivity" are treated differently).

Several methods have been developed to estimate gear selectivity, including comparing length compositions from side-by-side fishing using different mesh or hook sizes (e.g., Kirkwood and Walker, 1986; Millar, 1992; Millar and Holst, 1997), using underbag studies (e.g., Somerton and Otto, 1999), fitting models to tagging data (e.g., Anganuzzi et al., 1994), and comparing estimated length-frequencies from acoustic surveys with those from trawl catches (e.g., Somerton et al., 2011). Perez Comas and Pikitch (1994) conducted a meta-analysis of several studies to relate length-at-50%-retention to mesh size. However, all of these studies were focused on estimating gear selectivity rather than selectivity as it pertains to stock assessments. Comparisons between selectivity patterns estimated within stock assessments and those based on gear selectivity studies often find marked differences, owing to the impacts of availability (e.g., Turnock and Rugulo, 2011). Some assessments, particularly those for stocks for which there are no data on the age or length composition of the landings, assume that selectivity is known (e.g., Hilborn, 1990; Punt and Walker, 1998; Punt et al., 2000; Cope, 2013). In such cases, selectivity may be fixed based on the results of gear selectivity experiments, but most assessments treat selectivity as estimable within the stock assessment.

Age-structured stock assessments can be divided into two major categories: (1) those which assume catches-at-age are available for all of the years considered in the assessment and are known with negligible error compared to the data on relative abundance ("VPA-type" methods) and (2) those which allow for years for which catch-at-age data are missing and also allow for error in the catch-at-age data ("integrated" methods). Assumptions regarding selectivity are inherent in VPA-type methods: rates of change of selectivity with age for the oldest ages and with time for the most

* Corresponding author. Tel.: +1 206 221 6319; fax: +1 206 685 7471.
E-mail addresses: aepunt@uw.edu, aepunt@u.washington.edu (A.E. Punt).

0165-7836/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fishres.2013.06.003

recent years are assumed known (Laurec and Shepherd, 1983; Pope and Stokes, 1989; Shepherd, 1999). However, this paper focuses on how selectivity is treated within integrated assessment methods (Maunder and Punt, 2013; Punt et al., 2013). These methods are more amenable to approaches for evaluating model fit and applying model selection methods because they are generally based on maximum likelihood (or Bayesian) estimation methods. Integrated assessment modelling platforms (e.g., Stock Synthesis, Methot and Wetzell, 2013; CASAL, Bull et al., 2012; A-SCALA, Maunder and Watters, 2003; MULTIFAN-CL, Fournier et al., 1998) are the methods of choice for conducting the stock assessments on which management advice is based for fish and invertebrate stocks off the west coast of the US and Canada, including Alaska, as well as off New Zealand, Australia and South Africa. Integrated assessment methods are also being used increasingly to guide management advice for fish stocks off the eastern coast of North America and in Europe. Unlike VPA-type methods, integrated methods usually assume selectivity is related to age or length according to a parametric functional relationship, though this is not always the case (see Ianelli, in this issue). This is because parametric selectivity patterns have fewer parameters which should lead to more precise (but perhaps mis-specified) assessment outcomes.

Many choices need to be made when conducting assessments based on age- and length-structured population dynamics models under the integrated analysis paradigm (see Ianelli, in this issue for a review). For example: the analyst needs to decide whether selectivity is a function of age, length, or both age and length; whether selectivity is parametric or non-parametric; which of the many functional forms for selectivity should be assumed for each fleet (in particular whether selectivity should be asymptotic or dome-shaped); and whether selectivity is time-varying (and how time-varying selectivity is modelled). Stock Synthesis currently includes the ability to implement all of these choices. Sampson and Scott (2011) and Sampson (in this issue) outline how the overall selectivity of the fishery can be dome-shaped and time-varying, even though the selectivity for each individual fleet is asymptotic. In addition, a key decision when conducting stock assessments is how many fleets are to be modelled.

This paper reviews issues related to estimating selectivity and choosing appropriate selectivity models. It first explores the conditions under which selectivity is estimable. It then illustrates the sensitivity of assessment outcomes for two case studies, and outlines the principles related to choosing how many fleets to include in an assessment. Next, methods for selecting among functional forms for selectivity using regression diagnostics and model selection methods are presented. This is followed by a summary of the conclusions related to model selection which can be drawn from simulation studies. Finally, recommendations for further research on this topic are given.

2. Can selectivity be estimated even in principle?

It is well-known that selectivity, natural mortality, and trends in recruitment are confounded in catch-at-age (or catch-at-length) data (Butterworth and Punt, 1990; Thompson, 1994; Clark, 1999). Fig. 1 illustrates the theoretical catch age-composition which arises when natural mortality is constant ($M = 0.2\,\mathrm{yr}^{-1}$), recruitment is constant, fully-selecting fishing mortality is $0.2\,\mathrm{yr}^{-1}$, and selectivity is knife-edged at age 4 and then declines exponentially with age. Fits to these catch-at-age data in which each of natural mortality, selectivity, and recruitment are in turn assumed to change exponentially with age/time are shown. Selectivity for ages 0–3 is correctly assumed to be zero for all three fits. All three sets of model assumptions fit the catch-at-age data very well (essentially perfectly), but the estimates of fully-selected fishing mortality for each of these assumptions are markedly different: $0.2\,\mathrm{yr}^{-1}$, $0.29\,\mathrm{yr}^{-1}$, and $0.15\,\mathrm{yr}^{-1}$ respectively.

There are a variety of ways to overcome the confounding effect evident in Fig. 1, each of which involves making assumptions to reduce the number of estimated parameters. Typical assumptions include that natural mortality is independent of age and/or time, and that selectivity follows some pre-specified functional form for a subset of ages.[1] In principle, a model which estimates annual recruitment, age-specific selectivity, and age- and time-independent natural mortality will be estimable if there are at least four years of catch-at-age data (Butterworth and Punt, 1990). However, as shown by Pope and Shepherd (1982), stochasticity will lead to very poor estimates in this situation. Rather, it is necessary to have data in addition to catches-at-age even when selectivity is parameterized using a few parameters, to enable reliable inferences to be drawn. These additional data take the form of time-series of relative abundance, absolute indices of abundance, or of fishing mortality (Butterworth and Punt, 1990).

3. Examples of the sensitivity of assessment results to choices related to selectivity

The examples in this paper are based on assessments for the northern subpopulation of Pacific sardine, *Sardinops sagax*, and for Australia's eastern stock of pink ling, *Genypterus blacodes*.

The northern subpopulation of Pacific sardine currently supports the second largest federally managed fishery off the U.S. west coast. Pacific sardine has, at times, been the most abundant fish species in the California Current system. The management decisions for this stock are based on a harvest control rule that uses an estimate of the biomass of the age 1+ component of the population at the start of the current fishing season. The assessment for Pacific sardine has changed over time, and results are shown in this paper for the stock assessment configuration on which the 2011 assessment (Hill et al., 2011) was based. This assessment configuration pooled the sexes, modelled ages 0–15, with age 15 being a plus-group, and divided the year into two seasons. It started the assessment in 1993, and included three fleets (Mexico–California, season 1 (MexCal1); Mexico–California, season 2 (MexCal2); Pacific Northwest (PNW)), each with a different selectivity pattern (Fig. 2a). Selectivity was assumed to change in 1999 for the two MexCal fleets. The 2011 assessment for Pacific sardine included data on length-frequency and age-at-length (both for each fleet), as well as indices of abundance from aerial surveys, acoustic-trawl surveys, and the daily egg production method (Hill et al., 2011).

Pink ling stocks form the basis for major fisheries off the coasts of Australia and New Zealand. Pink ling off south-eastern Australia have been divided into two stocks (eastern and western) for assessment and management purposes because of differences among areas in size- and age-compositions, as well as in trends in catch rates (Whitten et al., in press). Assessments of the eastern stock of pink ling have been conducted using integrated analyses for several years and each assessment has considered multiple model configurations. One configuration during the 2012 assessment pooled data across "zones" (10, 20 and 30) (Fig. 2b) while another configuration considered each of the three zones as separate fleets. Selectivity was estimated separately for the trawl and non-trawl (primarily auto-longline) fisheries in both configurations. The data available for assessment purposes are the catches by fleet and zone, length-frequencies, and age-at-length data for these fleets (for some years),

[1] MacCall and Teo (2013) show that it is possible to treat selectivity for a subset of years and ages as free parameters, but impose a functional form on selectivity for other ages when applying integrated analysis.

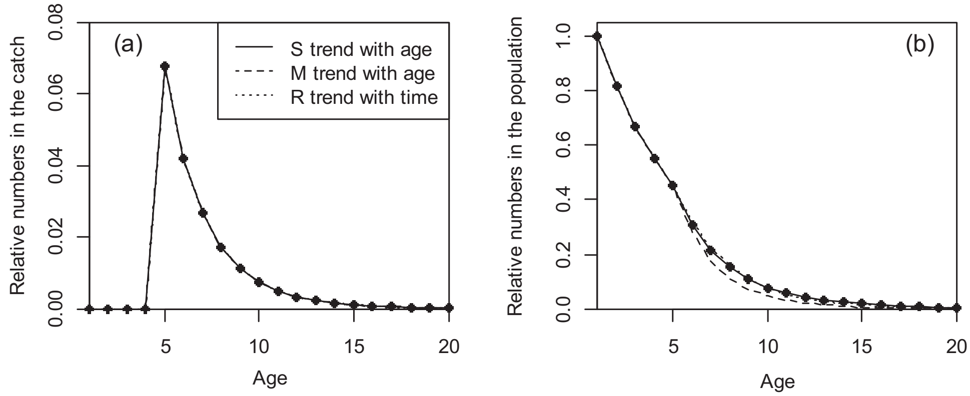

Fig. 1. Theoretical catch age-composition data with fits which allow for trends in selectivity with age, in natural mortality with age, and in recruitment over time (a), and the associated estimated numbers-at-age (b). The predictions in (a) are almost identical so the three lines of predicted catch proportions at age cannot be seen. Similarly in (b) the predictions from the models in which selectivity changes with age and recruitment changes over time are indistinguishable from the true proportions at age.

Fig. 2. Map of the west coast of North America showing the fleets on which the 2011 stock assessment for Pacific sardine was based (a), and the zones on which current assessment for pink ling are based (b).

Please cite this article in press as: Punt, A.E., et al., Model selection for selectivity in fisheries stock assessments. Fish. Res. (2013), http://dx.doi.org/10.1016/j.fishres.2013.06.003

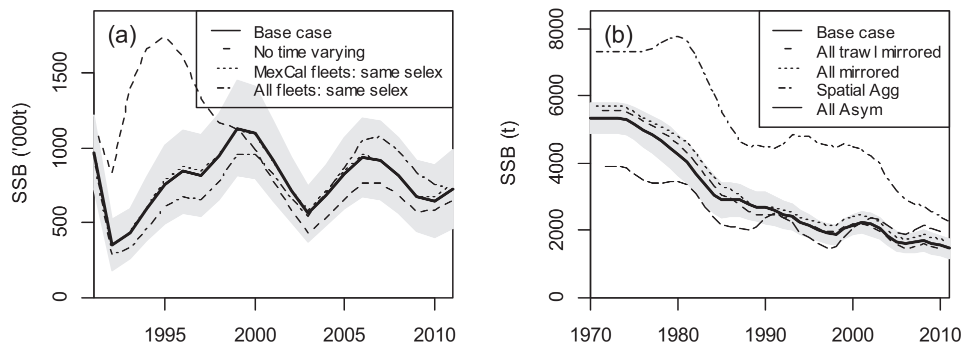

Fig. 3. Time-trajectories for spawning stock biomass for Pacific sardine (a) and pink ling (b). The solid line is a base-case assessment, the shaded area denotes asymptotic 95% confidence intervals, and the remaining lines are assessment configurations which differ in terms of specifications for length-based selectivity.

catch-rate indices for the trawl and non-trawl fisheries aggregated over zone as well as catch-rate indices for the trawl fishery in zones 20 and 30 separately. Results (abundance indices and length-frequency data) from a fishery-independent survey (Graham, 1999) are available for zone 10.

Selectivity is assumed to be a function of length (rather than age) for both assessments, and the results of the assessments for both species are sensitive to how selectivity is modelled (Table 1). Fig. 3 shows time-trajectories of estimates of spawning stock biomass for the base-case assessments, when fleets are assumed to have the same selectivity and when the number of fleets is changed (ling only). In general, the time-trajectories of biomass lie within the 95% confidence intervals for the base-case assessment (although it should be noted that the trajectories of biomass are not independent). However, ignoring fleet structure or assuming that all fleets have asymptotic (logistic) selectivity (ling) and ignoring time-variation in selectivity (sardine) have marked impacts on the results of the assessment. The consequences of different choices for fleet structure for pink ling (treating each zone as a separate fleet or pooling the data into trawl and non-trawl fleet categories [without weighting the data spatially]) are substantial; a difference in Recommended Biological Catch of 650 t was calculated between models under the agreed harvest control rule.

Table 1
Specifications of the base-case and sensitivity models. The specifications for the sensitivity tests reflect changes from the base-case analysis.

Model run	Specifications
Pacific sardine	
Base-case model	MexCal1, MexCal2 and PNW fleets; selectivity for the MexCal fleets changes in 1999
No time varying	Selectivity for the MexCal fleets does not change in 1999
MexCal fleets; same selex	MexCal2 selectivity mirrors MexCal1; selectivity for the MexCal fleets changes in 1999
All fleets; same selex	MexCal2 and PNW selectivities mirror MexCal1; selectivity for the MexCal fleets changes in 1999
Pink ling	
Base-case model	Trawl and non-trawl fleets in each of zones 10, 20 and 30. Selectivity for trawl fleets in zones 10 and 20 assumed to change in 2001 and 2006; selectivity for zone 30 changes in 2006 only. Dome-shaped double normal selectivity functions apply: parameters describing the "peak," "ascending width," and selectivity of the first length-class are time-varying
All trawl mirrored	Selectivity for the trawl fleets are the same
All mirrored Spatial Agg	The data are pooled over zones 10, 20 and 30 and there is one trawl and one non-trawl fleet
All Asym	Selectivity for all fleets is assumed to be asymptotic

4. Choosing fleets

Considerable effort has been dedicated to selecting between alternative functional forms for selectivity patterns for the fleets included in an integrated assessment (see Sections 5 and 6). However, far less attention has been paid to the selection of how many (and which) fleets should be modelled. Fleets can be defined as groups of vessels operating in an area over a time-period (e.g., one of the fleets in the Pacific sardine assessment is defined by vessels which made catches during the first half of the year off southern California). It is not necessary for a vessel to be in only one fleet. For example, a vessel may be in several fleets depending on the species it is targeting, the gear it is using, etc. Ideally, the selectivity patterns for a fleet should not change over time so that the number of parameters may remain small. Based on the authors' observations, the factors which determine the number of fleets included in an assessment include: (a) computational demands associated with coding and running models, (b) model stability, (c) data availability, and (d) the extent to which there are spatial trends in data sources.

Computational and programming demands were a major determinant of the number of fleets that could be modelled in stock assessments before the advent of generalized stock assessment packages (such as Stock Synthesis, CASAL, and MULTIFAN-CL) in which the number of fleets is an input and it is possible to assume that the selectivity patterns for several fleets are the same. The availability of sufficient information to estimate selectivity can be a major constraint on how many fleets are modelled. For example, the assessment of pink ling is based on three non-trawl fleets. However, there are several non-trawl gears (e.g., hook and autolongline) which are pooled together owing to a lack of sufficient length-frequency data to allow reliable estimation of selectivity for each of the non-trawl gear types. Similarly, selectivity for the non-trawl catches in zones 10 and 30 is assumed to be the same ("mirrored" in Stock Synthesis parlance) to those for zone 20 owing to lack of data.

It is increasingly common to treat spatial structure in a stock by dividing the range over which the stock is distributed into spatial strata (such as the zones in Fig. 2b) and treating the data for each stratum as though coming from a different fleet (e.g., Cope and Punt, 2011; Berger et al., 2012; Hurtado-Ferro et al., in this issue). This is clearly an approximation to reality: It implies that the population is fully mixed throughout its range, and that differences in catch and survey age- and length-composition data among areas are due to differences in selectivity. However, such differences could also arise when gear selectivity is constant across areas, but the proportion of the population by age or length differs among spatial strata. Cope and Punt (2011) compared this “areas as fleets” approach with assessing each area separately and pooling data spatially by means of simulation. They found the “areas as fleets” approach often performed poorest of the approaches considered when it didn't perform best. The poor performance of the “areas as fleets” approach may have been related to not allowing the selectivity patterns for fleets to be dome-shaped when the assessment method was applied (even when the true selectivity pattern was asymptotic).

Fleets can be selected by looking for spatial patterns in catch-rates (e.g., Cope and Punt, 2009) or length–frequency distributions (e.g., Lennert-Cody et al., 2010). Lennert-Cody et al. (2013) extended these earlier methods by identifying spatial areas by applying multivariate regression trees so that data on catch-rate and catch length-compositions can be used simultaneously.

Finally, one guideline which should be adopted more generally when conducting assessments is to define the maximum number of possible fleets when applying the stock assessment, but using regression diagnostics and model selection methods to identify how many fleets should be modelled as having different selectivity patterns. This is because unless data are entered in fleet-disaggregated form, it is unlikely that analysts will test for the possibility that different gear types/areas differ in terms of selectivity. Having many fleets in an assessment, but assuming that selectivity for several (or all) of them is the same does not lead to a more complicated model (in terms of number of estimated parameters), but does allow an evaluation of whether data from those fleets are in conflict or not. Pooling data across possible fleets means such conflicts could not be detected.

5. Selecting among selectivity patterns using model fit diagnostics

Assessment configurations which lead to systematic patterns in residuals about model fits to data are mis-specified, which implies there are assumptions underlying those assessment configurations that are violated. A thorough examination of residual patterns is therefore a key first step when conducting any stock assessment. Examination of whether a fit is mis-specified is made easier by the availability of software such as the r4ss package (Taylor et al., 2011); r4ss has been developed for Stock Synthesis, and automatically produces fit diagnostics.

Fig. 4 shows observed and model-predicted “aggregated” length-frequency distributions for pink ling (Punt and Taylor, 2012) when selectivity is assumed to be same for the trawl and non-trawl fleets in each of zones 10, 20 and 30. The aggregation in Fig. 4 is achieved by averaging the observed and model-predicted length-frequencies over time to be able to identify general patterns. There is clear evidence for model-misspecification in Fig. 4 (e.g., for Trawl 10B and Trawl 30B). This was a key factor, along with differences in trends in catch-rate spatially, which motivated the development of assessment configurations which separated zones 10, 20 and 30. Analyses in which each zone was treated as a separate fleet eliminated this model mis-specification (Punt and Taylor, 2012).

Butterworth and Rademeyer (2008) compared assessment configurations for Gulf of Maine cod, *Gadus morhua*, in which commercial selectivity is asymptotic and in which it is dome-shaped using residual diagnostics (bubble plots of the standardized residuals for the fits to the catch-at-age data) as well as using Information Theoretic approaches (see below). They argued that the residual patterns were inconsistent with the assumption of asymptotic selectivity because most of the commercial catch-at-age residuals for the older ages were negative (i.e., fewer older fish were observed than was consistent with the asymptotic selectivity assumption).

Although most studies have been based on model selection criteria such as AIC, BIC, and DIC (see below), Maunder and Harley (2011) note that these approaches will not perform well at selecting between selectivity functions with a separate parameter for each age and year (where the change in these parameters among years and ages is constrained by a smoothness penalty) and selectivity functions which are parametric functions of a small number of parameters. This is related to the fact that most assessment models are not fitted as state-space models (Gudmundsson and Gunnlaugsson, 2012 being a notable exception). Maunder and Harley (2011) propose using hold-out cross validation to select the size of the smoothness penalty (high values implying only small changes in selectivity among ages/years and lower values allowing selectivity to change markedly over time and among ages). This method involves selecting a random subset of the catch-at-age data, fitting the model without those data, and computing the out-of-sample error. The size of smoothness penalty is selected to minimize the out-of-sample error. Although the method has been applied for illustrative purposes to data for bigeye tuna, *Thunnus obesus*, in the eastern Pacific Ocean, it has not been adopted broadly, nor has it been tested using simulation.

Thompson and Lauth (2012) have developed an approach for determining the extent to which random effects (in this case, changes over time in selectivity parameters) should be penalized, when these changes are assumed to be normally distributed, and the penalty on the random effects is parameterized in terms of the standard deviation of the random effects. This approach involves three steps:

(1) Fit the stock assessment model with minimal (or no) constraints on the random effects and calculate the standard deviation of the resulting estimates of the random effects (denoted σ_s).
(2) Fit the assessment iteratively in which the standard deviation of the random effects for iteration i is set to the standard deviation of the estimates of the random effects for iteration $i-1$ until the results converge. $\tilde{\sigma}_s$ is the value for this standard deviation at the end of the iterative procedure.
(3) Set the standard deviation of the random effects for use in the actual assessment to $\sqrt{\sigma_s^2 - \tilde{\sigma}_s(\sigma_s - \tilde{\sigma}_s)}$.

Thompson and Lauth (2012) note that this estimator was developed in the context of a univariate linear model and its properties for non-linear stock assessment models are unknown (see Lee et al., in this issue-a, in this issue-b).

Posterior predictive checks (Gelman et al., 2004) can be used to evaluate the ability of Bayesian models to fit data. However, few assessments are based on Bayesian methods, and none have formally tested model fit using such checks. In contrast, posterior predictive checks have been used to evaluate simpler Bayesian assessment methods (e.g., Zhou et al., 2011).

In principle, the evaluation of residual patterns could be automated. For example, Polacheck et al. (1998) developed a set of rules for evaluating whether the residuals about indices are mis-specified. However, this approach has neither been tested nor expanded to residuals related to fits to age- and length-compositions.

An alternative method for identifying model misspecification is to conduct assessments for a range of fixed values for the parameter which determines the scale of the population (e.g., $\log R_0$ for Stock Synthesis) and plotting the negative log-likelihood (less the negative log-likelihood at the MLE) for the compositional data for each fleet against the value for this parameter (see Fig. 5 for a pink ling example; Lee et al., in this issue-a, in this issue-b; Wang

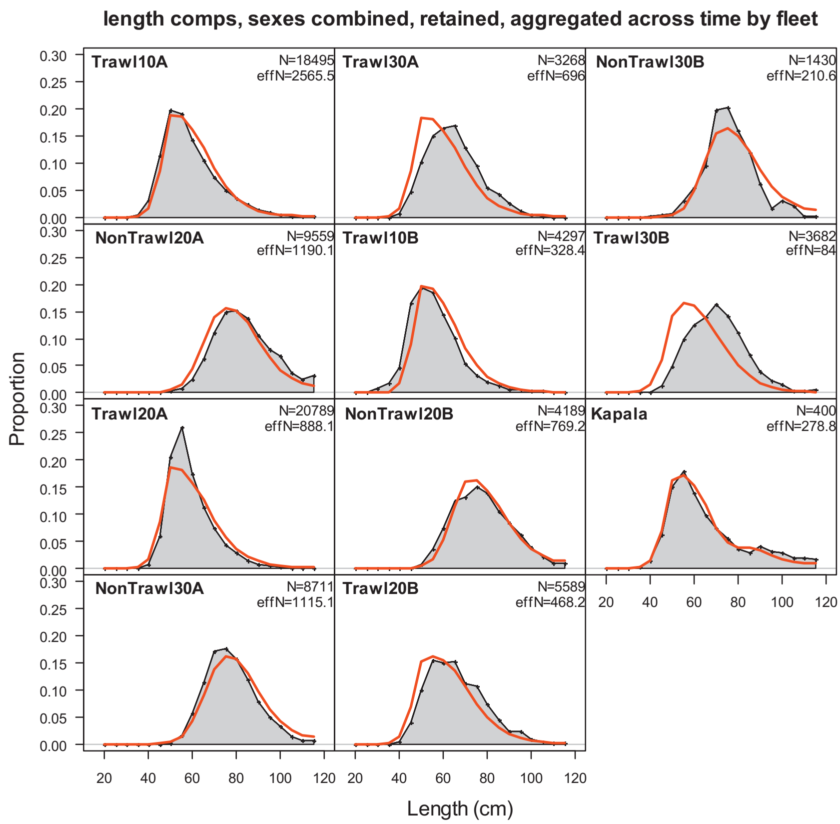

Fig. 4. Fits (solid lines) to the aggregated length-frequency data for pink ling when no account is taken of zone-specific selectivity ("A" = onboard, "B" = port).

et al., in this issue). Large changes in the difference in negative log-likelihood with small changes in the scaling parameter are then indicative of model mis-specification. As expected, the composition data for some of the fleets (e.g., 2, 3, 13 for the spatially disaggregated assessment and 4 and 5 for the spatially aggregated assessment) indicate lower unfished biomass, while those for other fleets (e.g., 5, and 8 for the spatially disaggregated assessment and 1 and 3 for the spatially aggregated assessment) indicate higher unfished biomass. The likelihood profiles for fleets 2 and 13 for the spatially aggregated assessment and fleet 1 for the spatially-aggregated assessment change fast with $\log R_0$. The fact that fleet 13 (the fishery-independent survey off New South Wales) appears to be so informative regarding R_0 is of some concern because there are only two length–frequency samples for this fleet and no age-composition data. In contrast, the change in negative log-likelihood with $\log R_0$ for fleet 2 is less surprising because there are substantial age- and length-composition data for fleet 2 and its selection pattern is assumed to be asymptotic, as is that for the non-trawl fishery. However, the results in Fig. 5a suggest that future assessments should explore sensitivity for this fleet to greater extent than has been the case in the past.

6. Model selection, AIC, DIC and all that

Model selection (e.g., Burnham and Anderson, 1998; Hilborn and Mangel, 1997) is an integral part of modern statistical modelling. The aim of model selection is to identify the best assessment configuration (or to weight each assessment configuration). Likelihood ratio tests, *F*-tests, AIC (Akaike, 1973; Burnham and Anderson, 1998), and the Bayes Information Criterion (BIC, Schwarz, 1978),

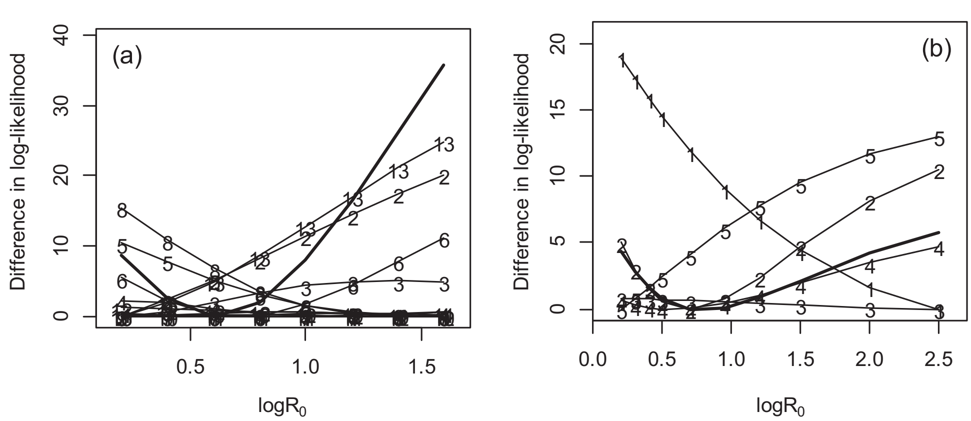

Fig. 5. Differences between the negative log-likelihoods for the compositional data (length and age) and the lowest negative log-likelihood by fleet for the spatially disaggregated (13 fleet) and -aggregated (5 fleet) assessments. The thick solid line is the sum over fleets and the numbers denote the individual fleets.

Please cite this article in press as: Punt, A.E., et al., Model selection for selectivity in fisheries stock assessments. Fish. Res. (2013), http://dx.doi.org/10.1016/j.fishres.2013.06.003

have been used to select among assessment configurations, including the form and shape of selectivity patterns for assessments based on maximum likelihood and penalized maximum likelihood. For Bayesian assessments, model selection has been based on the Bayes factors (Aitkin, 1991) and the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). Likelihood ratio tests and *F*-tests can only be applied to nested models whereas the other statistics can be applied to non-nested as well as nested models. However, the data on which each assessment configuration is based needs to be the same[2] when applying any of these model selection methods. This means, for example, that they cannot be used to compare the spatially disaggregated and spatially aggregated assessments for pink ling. This is unfortunate given the results for pink ling are more sensitive to whether or not the data are aggregated than to the choice of selectivity pattern given a particular choice of fleets (Fig. 3b). Splitting the data into many fleets and mirroring selectivity among fleets is one way to evaluate different approaches to aggregating data.

Butterworth et al. (2003) used AIC to compare among assessment configurations in which selectivity is constant over time and in which it is changing over time as a time-series (with a penalty on the 2nd derivative of the change in the parameters which determine time-varying selectivity). In that study, time-varying selectivity was strongly supported over static selectivity. However, the lack of an objective way to select the values for the parameters which determine the extent to which selectivity may vary among blocks of years, and the size of weight imposed on smooth changes in selectivity was arbitrary, along with the fact that treating what amount to random effects parameters as penalized fixed effects formally invalidated the use of AIC in this case.

Butterworth and Rademeyer (2008) used AIC to select among assessment configurations in which commercial selectivity is asymptotic versus those in which it is dome-shaped for Gulf of Maine cod, and found that AIC strongly supported dome-shaped selectivity. The difference in AIC was very substantial in this case, and the impacts of basing an assessment on dome-shaped versus asymptotic selectivity were profound in terms of whether the stock was assessed to be overfished or subject to overfishing (Butterworth and Rademeyer, in this issue).

Bogaards et al. (2009) applied a Bayesian state-space model to plaice, *Pleuronectes platessa*, in the North Sea in which the western and central North Sea were modelled separately, with mixing between them. Nine assessment configurations were compared using DIC. The best assessment configuration assumed that selectivity was the same in the two regions, but that recruitment and measurement error variation differed among regions, even though the weight of evidence in favour of the more complicated model was relatively small (ΔDIC = 1.9). Most comparisons using AIC or DIC have involved comparing asymptotic selectivity versus dome-shaped selectivity. However, Aarts and Poos (2009) compared selectivity patterns based on splines (static and time-varying) using AIC for North Sea plaice.

Even though the differences in outcomes are not marked among the ways of treating selectivity in Fig. 3, application of AIC provides "definitive" evidence (ΔAIC > 400–500) in favour of the base-case over the simpler models for both sardine and pink ling. It is not straightforward to apply BIC to most stock assessments because what constitutes a data point is not entirely clear.

The reliability of methods such as AIC and DIC depends on how the various data sources have been weighted. In particular, "overweighting" the data (i.e., underestimating the standard deviations for the survey indices, ignoring correlations among data when these correlations exist, assuming likelihood functions which do not allow for "outlier" observations when there are "outliers," or overestimating the effective sample sizes for the compositional data) will lead to more complicated models being selected over simpler alternatives even when this is not justified. Many assessments include setting the input standard deviations for the survey indices so that standard deviations of the residuals match the standard deviations of the survey indices assumed when fitting the model. However, survey standard deviations are sometimes set equal to the sampling standard errors. A more difficult problem pertains to the effective sample sizes for the compositional data. Setting the effective sample sizes to the number of fish measured will almost always lead to overestimation of the information content of the compositional data. This is because fish often aggregate by size or age so that the lengths of fish within a set or haul are not independent, and some of the error between the observed and model-predicted age/length compositions is attributable to violations of assumptions (e.g., that selectivity is static when it is not).

Many assessments, particularly those based on Stock Synthesis, have set the effective sample sizes for the age- and length-composition data using the approach of McAllister and Ianelli (1997). However, that approach relies on the assumption that the residuals are independent among age- or length-classes (which is seldom the case; see, for example, Fig. 6). Francis (2011) recommends down-weighting age and length data to ensure that the model fits abundance indices adequately, and outlines an approach based on defining the effective sample sizes in terms of variation in the expected mean age (or length) of the catch. Unfortunately, this approach may fail when selectivity is modelled as a random walk because the model predictions of the mean age (length) will match the observed value very closely and the effective sample sizes could become very large. If the compositional data and other sources of information are in conflict, allowing for time-varying selectivity and using Francis's method for defining effective sample sizes could lead to the model converging to the fit which mimics the compositional data better.

Maunder (2011) provides a summary of the likelihood functions used in age- and length-structured integrated stock assessments, and illustrates how the effective sample size (or the equivalent residual standard deviation) is estimated. Maunder (2011) evaluated six of the methods for estimating effective sample sizes (iterative multinomial likelihood; normal approximation, using binomial variance; lognormal likelihood with variance proportional to the inverse of the proportion; Dirichlet likelihood; and a multivariate normal approximation). He indicated a preference for the approach based on the lognormal likelihood with variance proportional to the inverse of the proportion (although see Legault, in this issue), which was originally developed by Punt and Kennedy (1997) and has been used extensively for assessments in South Africa (e.g., Rademeyer et al., 2008). Unfortunately, the impact of the choice of the likelihood function on model selection has yet to be evaluated thoroughly, e.g. by way of a simulation-based study.

Most stock assessments treat the deviations about the stock-recruitment function as estimable parameters subject to a penalty (where the size of the penalty is "tuned," e.g. Hill et al., 2011). However, this is an approximation to treating these parameters as random effects. The parameters which are used to model random walks in selectivity (or catchability) should similarly be treated as random effects, but are not. Model selection methods such as AIC and BIC rely on a count of the number of parameters, but this number is not straightforward to compute when random effects are treated as penalized parameters. In principle, this problem could be overcome by adopting a state-space formulation for the assessment model and estimating all of the variance parameters (see Schnute, 1994 for an early example of a state-space formulation of a fisheries stock assessment). Gudmundsson and Gunnlaugsson (2012) fitted

[2] Being "the same" in this context includes that the raw data are aggregated to fleet the same way.

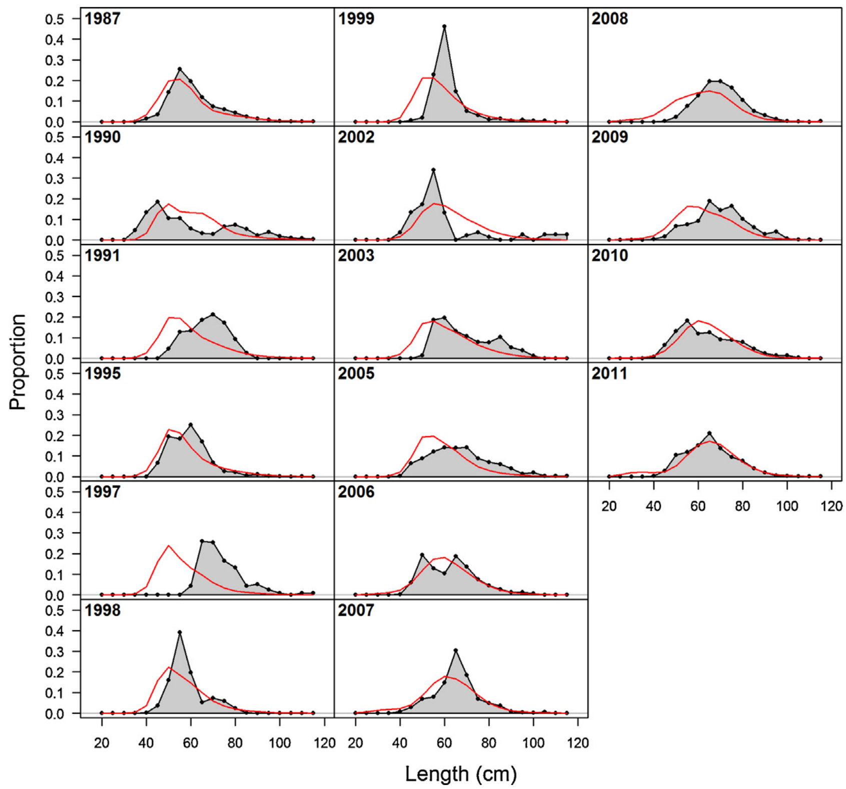

Fig. 6. Fits (solid lines) to the length–frequency data for pink ling caught by the trawl fishery in zone 20. The fits are based on the assessment configuration in which selectivity is the same for all trawl fleets, as is selectivity for all non-trawl fleets.

an age-structured population dynamics model within the context of the Kalman Filter (Harvey, 1989) and compared assessment configurations with time-varying and static selectivity for Icelandic cod and pollock, *Pollachius virens*, using likelihood ratio tests. Time-varying selectivity was strongly supported for both stocks, but more so for pollock. Adopting a fully Bayesian formulation for an assessment in which the variances of the random effects were treated as estimable parameters and assigned prior distributions would also overcome the problem of treating random effects as penalized parameters, albeit at the expense of additional computational demands, and potentially problems with convergence. The choice between maximum likelihood and Bayesian formulations of state-space models is largely one of preference. For example, Bayesian formulations would be most appropriate when there is prior information or the estimates of the random effects (e.g. recruitments and selectivity patterns for individual years) are of interest. However, model selection for Bayesian assessment using DIC can be complicated. For example, the deviance corresponding to the posterior mean needs to be determined to compute DIC. However, this deviance can be much larger than the deviances for the points in posterior given the non-linearities inherent in population dynamics models. In addition, Bayesian model selection will be impacted by the prior chosen for the parameters.

Although most assessments have attempted to select a best set of specifications related to selectivity patterns, it is possible to apply model averaging (Burnham and Anderson, 1998; Hoeting et al., 1999) to obtain results that integrate over multiple assumptions related to selectivity. For example, Brodziak and Piner (2010) apply model averaging based on BIC-weights to account for the uncertainty related to the form of the stock–recruitment relationship. Care must be taken when selecting alternative models to avoid, for example, including essentially the same model in the model average multiple times. This might occur when many variants of one type of model (e.g., a model with dome-shaped selectivity) and only a few variants of a different type of model (e.g. a model with asymptotic selectivity) are included in an analysis. Furthermore, the analyst must carefully select the prior probabilities for each alternative model if Bayesian model averaging is applied.

7. What can simulation studies tell us

Simulation studies provide the ability to understand the properties of statistical estimation methods, including stock assessment methods. In principle, simulation studies provide an understanding of the generic behaviour of methods. However, the results of simulation studies will depend on the values for the model used to generate data (the operating model) so all simulation studies are to some extent "case-specific."

Relatively few simulation studies have focused on estimation of selectivity because the parameters which define selectivity are generally considered to be "nuisance" parameters compared to assessment outputs such as estimates of biomass and of fishery exploitation rates. Although many simulation studies have explored the performance of assessment methods, results are seldom reported in terms of the ability to estimate selectivity. Moreover, few studies have considered cases in which selectivity is mis-specified. The ability to estimate selectivity is generally good (estimates are accurate and precise – at least compared to parameters such as natural mortality and the parameters of the stock–recruitment relationship) when simulation studies report how well the parameters of selectivity functions are estimated. Most simulation studies generate age/length composition data from the survey or fishery catch in a way which matches the distributions assumed when fitting the model (e.g., Bence et al., 1993; Sampson and Yin, 1998; Taylor et al., 2005; Radomski et al., 2005) so patterns of residuals such as those in Figs. 4 and 6 are highly unlikely to be observed in simulated data sets. In contrast, Hurtado-Ferro et al. (in this issue) generate length and age-at-length data accounting for spatial structure in the population and fishery, as well as the nature of the way catches are sampled. Unlike most previous studies, the effective sample sizes inferred from the data generated by Hurtado-Ferro et al. (in this issue) are markedly smaller than the number of fish sampled (see Crone and Sampson, 1998 for an example of the relationship between effective sample size and numbers of fish measured). Concerns aside, some useful insights have been gained from simulation studies that consider the estimation of selectivity.

Kimura (1990) conducted the first simulation study to investigate the consequences of violation of selectivity assumptions on the ability to estimate quantities of management interest using integrated analysis. He found the performance of the estimation method depended on the form of the selectivity pattern. Sampson (1994) extended Kimura's analyses by considering more factors that could impact estimation performance, as well as a range of assessment methods. He found that Stock Synthesis, CAGEAN (Deriso et al., 1985), and the multiplicative catch-at-age model of Shepherd and Nicholson (1991) were all sensitive to undetected changes in selectivity.

Fu and Quinn (2000) conducted simulations based on length-structured operating and estimation models. They found, as expected, that estimating time-varying selectivity when selectivity was specified to vary over time, led to better estimation performance than assuming selectivity was static and instead estimating time-varying natural mortality.

Yin and Sampson (2004) explored the impact of a number of factors on the ability of Stock Synthesis to estimate terminal biomass, terminal fishing mortality, and relative biomass in the terminal year of an assessment, amongst other management-related quantities. Whether selectivity was dome-shaped or asymptotic was one of those factors. However, selectivity was generally less important in terms of size of bias and CV than other factors such as the number of years and precision of data, the rate of change of fishing mortality, and the true value for natural mortality (except when the quantity of interest was relative biomass in the terminal year).

Radomski et al. (2005) evaluated the performance of a statistical catch-at-age analysis method using an operating model in which age-specific selectivity was static, changed in blocks, or was related to fishing mortality ("nonadditive" in Radomski et al., 2005). The results of the simulations suggested that allowing fishery selectivity to change over time when conducting assessments performed about as well as assuming static selectivity when selectivity was indeed static. However, allowing for time-varying selectivity performed much better than assuming static selectivity when selectivity really changed over time. Unfortunately, in common with most simulation studies, Radomski et al. (2005) did not consider assessment configurations in which model selection methods were used to select between static and time-varying selectivity, even though this best reflects common practice. The mean relative errors for spawning biomass showed a retrospective pattern when an assessment model in which selectivity was assumed to be time-invariant was fitted to data generated from a model in which selectivity differed among time periods (Fig. 6 of Radomski et al., 2005), which suggests retrospective patterns may indicate that selection patterns are time-varying. Martell and Stewart (in this issue) have found retrospective patterns for Pacific halibut, *Hippoglossus stenolepis*, can be eliminated by allowing for time-varying selectivity, giving further support to this hypothesis.

Helu et al. (2000) evaluated the performance of AIC and BIC to select among asymptotic and dome-shaped selectivity. The performance of both model selection criteria was generally excellent (probability of selecting the correct model > 0.95) for most cases, but AIC occasionally performed poorly (e.g. probability of selecting the correct model < 0.7). Correctly selecting the form of the selectivity function was, not surprisingly, shown to improve estimates of recent year biomass.

Linton and Bence (2011) compared the ability to estimate population biomass and exploitation rate when selectivity is time-varying. They considered four parameterizations of selectivity (three based on allowing subsets of the parameters of a double-logistic selectivity function to change over time as random walks, and the fourth based on the approach in Butterworth et al., 2003). Linton and Bence (2011) also evaluated the performance of estimators which selected among the four parameterizations using (a) the RMSE between the observed and predicted catches-at-age, (b) the size of any retrospective bias using the DR statistic of Mohn (1999), and (c) DIC. As expected, performance was best when the estimation method matched the method used to generate the data. The DR-based model selection method performed better than the RMSE and DIC methods, and achieved median relative errors and median absolute relative errors similar to the estimation methods which mimicked the operating model.

8. Discussion and recommendations

The general process for conducting model selection and model evaluation is well-known. However, model selection, and the results of stock assessments, can be impacted by data-weighting (Richards, 1991; Francis, 2011). Each assessment should therefore clearly document its "weighting philosophy." For example, the "weighting philosophy" for the pink ling assessment of Punt and Taylor (2012) and of Whitten et al. (in press) was: (a) the model should fit the trends in the abundance indices as well as possible, and (b) the effective sample sizes and CVs assigned to the data should match the variation implied by the residuals. This philosophy was implemented by:

(a) conducting the initial model selection analyses (selecting whether growth and selectivity change over time, and how and whether selectivity differs among fleets) while imposing high weight (an average CV of 0.1) on the abundance indices;
(b) modifying the years for which recruitment and growth deviations are estimated based on application of model selection methods (and also considering the precision of the resulting estimates of the year-specific parameters which determine time-varying growth and recruitment);
(c) modifying the effective sample sizes for the age- and length data (generally using the approach of McAllister and Ianelli, 1997); and
(d) adjusting (increasing) the CVs for the catch-rate data so that the variation of the residuals matches the assumed residual standard deviation.

All assessment configurations were checked for model misspecification using graphical summaries output from r4ss during each of these steps.

This paper shows that many methods are available to select how selectivity should be modelled. However, it is also clear that the field has not yet stabilized, and there is no accepted standard practice. This is partially because the various studies which have compared methods have been incomplete. The following key areas should be considered for future research:

(a) Time-varying selectivity is increasingly being included in stock assessments. However, no simulation studies have yet compared the relative benefits of time-blocking selectivity versus allowing the parameters of selectivity patterns to change over time. The authors of this paper anticipate that allowing parameters to change over time as, for example, a random walk, will generally perform best (except when a major change occurred in a fishery, which may be better modelled by splitting the fishery into multiple fleets at that point anyway). However, modelling selectivity as a random walk will likely only be adopted more broadly if methods are developed to more rigorously select the parameters which control the extent to which selectivity can change over time – the methods of Thompson and Lauth (2012) and Maunder and Harley (2011) are good steps in this direction, but have yet to be evaluated using simulation.

(b) Time-varying selectivity will be adopted more broadly if assessment models can be represented as state-space models so that model selection methods can be applied, and the parameter which determines the extent to which selectivity changes over time can be estimated. Some work along these lines has been conducted, but progress is currently limited owing to lack of efficient computational algorithms.

(c) Simulation studies have explored the impact of making the wrong assumption about the shape of the selectivity patterns included in stock assessments. However, those studies have been limited because: (i) they commonly generate data in a way which is consistent with the assumed likelihood function; (ii) they seldom allow for the overdispersal which arises from the way length and age data are actually collected (Hurtado-Ferro et al., in this issue being a noteworthy exception); and (iii) no attempt has been made to simulate the process of fleet selection and only a few studies (e.g., Helu et al., 2000; Linton and Bence, 2011) have attempted to integrate the model selection process into the assessment method which is being tested. Future simulation studies should attempt to cover all of these challenges that analysts face in reality.

(d) Simulation studies have to date only focused on how well selectivity can be estimated within age-structured assessment methods. However, integrated analysis methods based on length-structured population dynamics models are applied widely for species which are hard to age (Punt et al., 2013). Few studies have examined the ability to estimate selectivity within such assessments (see Szuwalski and Punt, 2012 for a recent exception). The ability to select among alternative functional forms for selectivity in length-structured models is likely to be poorer than in age-based models, but the extent to which this is the case needs to be examined.

(e) Non-additive models have been proposed for selectivity (e.g., Radomski et al., 2005). However, these models have yet to be included in actual stock assessments, and work to identify whether violations of the assumption of additivity can be detected is yet to be conducted.

(f) The authors are not aware of any assessments that have formally evaluated whether selectivity should be a function of age or of length, and current practice seems to be one of personal preference rather than formal analysis.

Finally, some of the reasons for time-varying selectivity are that spatial structure in the population dynamics are ignored within population dynamics models which assume that the population is spatially homogeneous (Sampson, in this issue). Examinations of whether spatial models can reliability distinguish movement from selectivity, and if not, which data will allow movement to be distinguished from selectivity, are warranted.

Acknowledgements

This paper was improved by suggestions from Doug Butterworth (University of Cape Town), Mark Maunder (IATTC), Chris Francis (NIWA), and an anonymous reviewer. AEP acknowledges support from NOAA Grant NA10OAR4320148. This is JISAO contribution #2145.

References

Aarts, G., Poos, J.J., 2009. Comprehensive discard reconstruction and abundance estimation using flexible selectivity functions. ICES J. Mar. Sci. 66, 763–771.
Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: 2nd International Symposium on Information Theory, Academiai Kiado, Budapest, pp. 267–281.
Aitkin, M., 1991. Posterior Bayes factors. J. R. Statist. Soc. B 53, 111–142.
Anganuzzi, A.A., Hilborn, R., Skalski, J.R., 1994. Estimation of size selectivity and movement rates from mark-recovery data. Can. J. Fish. Aquat. Sci. 51, 734–742.
Bence, J.R., Gordoa, A., Hightower, J.E., 1993. Influence of age-selective surveys on the reliability of stock synthesis assessments. Can. J. Fish. Aquat. Sci. 50, 827–840.
Berger, A.M., Jones, M.L., Xhao, Y., Bence, J.R., 2012. Accounting for spatial population structure at scales relevant to life history improves stock assessment: the case for Lake Erie walleye *Sander vitreus*. Fish. Res. 115/116, 44–59.
Bogaards, J.A., Kraak, S.B.M., Rijnsdorp, A.D., 2009. Bayesian survey-based assessment of North Sea plaice (*Pleuronectes platessa*): extracting integrated signals from multiple surveys. ICES J. Mar. Sci. 66, 665–679.
Brodziak, J., Piner, K., 2010. Model averaging and probable status of North Pacific striped marlin, *Tetrapturus audax*. Can. J. Fish. Aquat. Sci. 67, 794–805.
Bull, B., Francis, R.I.C.C., Dunn, A., McKenzie, A., Gilbert, D.J., Smith, M.H., Bian, R., Fu, D., 2012. CASAL (C++ algorithmic stock assessment laboratory): CASAL user manual v2.30-2012/03/21. NIWA Technical Report 135.
Burnham, K.P., Anderson, D.R., 1998. Model Selection and Inference: A Practical Information-Theoretical Approach. Springer Verlag, New York, pp. 353.
Butterworth, D.S., Punt, A.E., 1990. Some preliminary examinations of the potential information content of age-structure data from Antarctic mike whale research catches. Rep. Int. Whal. Commn. 40, 301–315.
Butterworth, D.S., Rademeyer, R.A., 2008. Statistical catch-at-age analysis vs. ADAPT-VPA: the case of Gulf of Maine cod. ICES J. Mar. Sci. 65, 1717–1732.
Butterworth, D.S., Rademeyer, R.A., 2013. Fisheries management – does selectivity matter? Fish. Res. (in this issue).
Butterworth, D.S., Ianelli, J.N., Hilborn, R., 2003. A statistical model for stock assessment of southern Bluefin tuna with temporal changes in selectivity. S. Afr. J. Mar. Sci. 25, 331–361.
Clark, W.G., 1999. Effects of an erroneous natural mortality rate on a simple age-structured stock assessment. Can. J. Fish. Aquat. Sci. 56, 1721–1731.
Cope, J.M., 2013. Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for deriving overfishing limits in data-limited situations. Fish. Res. 142, 3–14.
Cope, J.M., Punt, A.E., 2009. Drawing the lines: resolving fishery management units with simple fisheries data. Can. J. Fish. Aquat. Sci. 66, 1256–1273.
Cope, J.M., Punt, A.E., 2011. Reconciling stock assessment and management scales under conditions of spatially varying catch histories. Fish. Res. 107, 22–38.
Crone, P.R., Sampson, D.B., 1998. Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition. In: Funk, F., Quinn II, T.J., Hiefetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.F., Sullivan, P.J., Zhang, C.-I. (Eds.), Fishery Stock Assessment Models, Alaska Sea Grant College Program Report No. AK-SG-98-01, Fairbanks. pp. 355–370.
Deriso, R.B., Quinn, T.J.I.I., Neal, P.R., 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42, 815–824.
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., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138.
Fu, C., Quinn II, T.J., 2000. Estimability of natural mortality and other population parameters in a length-based model: *Pandalus borealis* in Kachemak Bay, Alaska. Can. J. Fish. Aquat. Sci. 57, 2420–2432.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis. Chapman & Hall/CRC, Boca Raton, pp. 668.
Graham, K.J., 1999. Trawl fish length–weight relationships from data collected during FRV Kapala surveys. In: NSW Fisheries Research Report Series No. 2.
Gudmundsson, G., Gunnlaugsson, T., 2012. Selection and estimation of sequential catch-at-age models. Can. J. Fish. Aquat. Sci. 69, 1760–1772.
Harvey, A.C., 1989. Forecasting, Structural Time Series Models, and the Kalman Filter. Cambridge University Press, Cambridge, UK.
Hilborn, R., 1990. Estimating the parameters of full age-structured models from catch and abundance data. Bull. Int. North Pac. Fish. Commn. 50, 207–2013.
Hilborn, R., Mangel, M., 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton, pp. 315.
Hill, K.T., Crone, P.R., Lo, N.C.H., Macewicz, B.J., Dorval, E., McDaniel, J.D., Gu, Y., 2011. Assessment of the Pacific Sardine Resource in 2011 for US Management in 2012. Pacific Fishery Management Council, Portland, pp. 275.
Hoeting, J., Madigan, D., Raftery, A., Volinsky, C., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14, 382–401.
Helu, S.L., Samplson, D.B., Yin, Y., 2000. Application of statistical model selection criteria to the Stock Synthesis assessment program. Can. J. Fish. Aquat. Sci. 57, 1784–1793.
Hurtado-Ferro, F., Punt, A.E., Hill, K., 2013. Use of multiple selectivity patterns as a proxy for spatial structure. Fish. Res. (in this issue).

Ianelli, J.N., 2013. Evaluating selectivity trade-offs in groundfish assessments. Fish. Res. (in this issue).
Kimura, D.K., 1990. Approaches to age-structured separable sequential population analysis. Can. J. Fish. Aquat. Sci. 47, 2364–2374.
Kirkwood, G.P., Walker, T.I., 1986. Gill net mesh selectivities for gummy shark, *Mustelus antarcticus* Günther, taken in south-eastern Australian waters. Aust. J. Mar. Freshwater Res. 37, 689–697.
Laurec, A., Shepherd, J.G., 1983. On the analysis of catch and effort data. J. Cons. Int. Explor. Mer. 41, 81–84.
Lee, H-H., Piner, K.R., Methot, R.D., Maunder, M.N., 2013. Example of the use of likelihood profiling over a global scaling parameter to structure the population dynamics model of blue marlin in the Pacific Ocean. Fish. Res. (in this issue).
Lee, H.-H., Maunder, M.N., Aires-Da-Silva, A., Piner, K.R., 2013. Evaluation of a practical method to estimate the variance parameter of random effects for time varying selectivity. Fish. Res. (in this issue).
Legault, C.M., 2013. Simulation test of age composition error distributions. Fish. Res. (in this issue).
Lennert-Cody, C.E., Minami, M., Tomlinson, P.K., Maunder, M.N., 2010. Exploratory analysis of spatial-temporal patterns in length-frequency data: an example of distributional regression trees. Fish. Res. 102, 323–326.
Lennert-Cody, C.E., Maunder, M.N., Aires-da-Silva, A., Minami, M., 2013. Defining population spatial units: simultaneous analysis of frequency distributions and time series. Fish. Res. 139, 85–92.
Linton, B.C., Bence, J.R., 2011. Catch-at-age assessment in the face of time-varying selectivity. ICES J. Mar. Sci. 68, 611–625.
MacCall, A.D., Teo, S.L.H., 2013. A hybrid stock synthesis – virtual population analysis model of Pacific Bluefin tuna. Fish. Res. 142, 22–26.
Martell, S., Stewart, I., 2013. Best practices for modeling time-varying selectivity. Fish. Res. (in this issue).
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., Harley, S.J., 2011. Using cross validation model selection to determine the shape of nonparametric selectivity curves in fisheries stock assessment models. Fish. Res. 110, 283–288.
Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142, 61–74.
Maunder, M.N., Watters, G.M., 2003. A-SCALAL: an age-structured statistical catch-at-length analysis for assessing tuna stocks in the eastern Pacific Ocean. Inter-Am. Trop. Tuna Commn. Bull. 22, 433–582.
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., Wetzell, C.R., 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142, 86–99.
Millar, R.B., 1992. Estimating the size-selectivity of fishing gear by conditioning on the total catch. J. Am. Stat. Assoc. 87, 962–968.
Millar, R.B., Holst, R., 1997. Estimation of gillnet and hook selectivity using log-linear model. ICES J. Mar. Sci. 54, 471–477.
Mohn, R., 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56, 473–488.
Perez Comas, J.A., Pikitch, E.K., 1994. The predictive power of empirical relationships describing size selectivity, with application to gadoid fish. Fish. Res. 20, 151–164.
Pribac, F., Punt, A.E., Walker, T.I., Taylor, B.L., 2005. Using length, age and tagging data in a stock assessment of a length selective fishery for gummy shark (*Mustelus antarcticus*). J. Northwest Atl. Fish. Sci. 35, 267–290.
Pope, J.G., Shepherd, J.G., 1982. A simple method for the consistent interpretation of catch-at-age data. J. Cons. Int. Explor. Mer. 40, 176–184.
Pope, J.G., Stokes, T.K., 1989. Use of multiplicative models for separable virtual population analysis (VPA), integrated analysis, and the general VPA tuning problem. Am. Fish. Soc. Symp. 6, 92–101.
Polacheck, T., Haskard, K.A., Klaer, N., Bethlehem, A., Preece, A., 1998. An index for weighting results in catch-at-age models based on diagnostic tests for lack of fit. In: Funk, F., Quinn II, T.J., Hiefetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.F., Sullivan, P.J., Zhang, C.-I. (Eds.), Fishery Stock Assessment Models, Alaska Sea Grant College Program Report No. AK-SG-98-01, Fairbanks. pp. 137–154.
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., Taylor, B.L., 2012. Stock Assessment of pink ling (*Genypterus blacodes*) using data up to 2010. In: Tuck, G.N. (Ed.), Stock Assessment for the Southern and Eastern Scalefish and Sharkfish 2011. Part 1. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart, pp. 103–160.
Punt, A.E., Walker, T.I., 1998. Stock assessment and risk analysis for the school shark (*Galeorhinus galeus*) resource off southern Australia. Mar. Freshwater Res. 49, 719–731.
Punt, A.E., Huang, T.-C., Maunder, M.N., 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES. J. Mar. Sci. 70, 16–33.
Punt, A.E., Pribac, F., Walker, T.I., Taylor, B.L., Prince, J.D., 2000. Stock assessment of school shark *Galeorhinus galeus* based on a spatially-explicit population dynamics model. Mar. Freshwater Res. 51, 205–220.
Rademeyer, R.A., Butterworth, D.S., Plagányi, É.E., 2008. A history of recent bases for management and the development of a species-combined operational management procedure for the South African hake resource. S. Afr. J. Mar. Sci. 30, 291–310.
Radomski, P., Bence, J.R., Quinn II, T.J., 2005. Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational, kill-dominated fishery. Can. J. Fish. Aquat. Sci. 62, 436–452.
Richards, L.J., 1991. Use of contradictory data in stock assessments. Fish. Res. 11, 225–238.
Sampson, D.B., 1994. The assumption of constant selectivity and the stock assessment for widow rockfish, *Sebastes entomelas*. Fish. Bull. 91, 676–689.
Sampson, D.B., 2013. Fishery selection and its relevance to stock assessment and fishery management. Fish. Res. (in this issue).
Sampson, D.B., Scott, R.D., 2011. A spatial model for fishery age-selection at the population level. Can. J. Fish. Aquat. Sci. 68, 1077–1086.
Sampson, D.B., Yin, Y., 1998. A Monte Carlo evaluation of the Stock Synthesis assessment program. In: Funk, F., Quinn II, T.J., Hiefetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.F., Sullivan, P.J., Zhang, C.-I. (Eds.), Fishery Stock Assessment Models, Alaska Sea Grant College Program Report No. AK-SG-98-01, Fairbanks. pp. 315–318.
Schnute, J.T., 1994. A general framework for developing sequential fisheries models. Can. J. Fish. Aquat. Sci. 51, 1676–1688.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.
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, 584–591.
Shepherd, J.G., Nicholson, M.D., 1991. Multiplicative modelling of catch-at-age data, and its application to catch forecasts. J. Cons. Int. Explor. Mer. 47, 284–294.
Somerton, D.A., Otto, R.S., 1999. Net efficiency of a survey trawl for snow crab, *Chionoecetes opilio*, and Tanner crab, *C. bairdi*. Fish. Bull. 97, 617–625.
Somerton, D.A., Williams, K., von Szalay, P.G., Rose, C.S., 2011. Using acoustics to estimate the fish-length selectivity of trawl mesh. ICES J. Mar. Sci. 68, 1558–1565.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Statist. Soc. B 64, 583–639.
Szuwalski, C., Punt, A.E., 2012. Identifying research priorities for management under uncertainty: the estimation ability of the stock assessment method used for eastern Bering Sea snow crab (*Chionoecetes opilio*). Fish. Res. 134–136, 82–94.
Taylor, I.G., Stewart, I.J., Hicks, A., Garrison, T.M., Punt, A.E., Wallace, J.R., Wetzel, C.R., 2011. r4ss: R code for Stock Synthesis. R package version 1.16, http://R-Forge.R-project.org/projects/r4ss/.
Taylor, N.G., Walters, C.J., Martell, S.J.D., 2005. A new likelihood for simultaneously estimating von Bertalanffy growth parameters, gear selectivity, and natural and fishing mortality. Can. J. Fish. Aquat. Sci. 62, 215–223.
Thompson, G.G., 1994. Confounding of gear selectivity an the natural mortality rate in cases where the former is a nonmonotone function of age. Can. J. Fish. Aquat. Sci. 51, 2654–2664.
Thompson, G.G., Lauth, R.R., 2012. Assessment of the Pacific cod stock in the Eastern Bering Sea and Aleutian Islands Area. In: Plan Team for Groundfish Fisheries of the Bering Sea/Aleutian Islands (Compiler), Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands regions. North Pacific Fishery Management Council, Anchorage, AK, pp. 245–544.
Turnock, B.J., Rugulo, L.J., September 2011. Stock assessment of eastern Bering Sea snow crab. In: NPFMC (North Pacific Fishery Management Council). Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Region: 2011 Crab SAFE. North Pacific Fishery Management Council, Anchorage, pp. 37–168.
Whitten, A.R., Punt, A.E., Taylor, B.L., 2013. Pink ling (*Genypterus blacodes*) stock assessment based on data up to 2011. In: Tuck, G.N. (Ed.), Stock Assessment for the Southern and Eastern Scalefish Fishery: 2013 Part X. Australian Fisheries Management Authority and CSIRO Marine and Atmospheric Research, Hobart (in press).
Wang, S.-P., Maunder, M.N., Aires-Da-Silva, A., Piner, K.R., 2013. Virgin recruitment profiling as a diagnostic for selectivity curve mispecification in integrated stock assessment models. Fish. Res. (in this issue).
Yin, Y., Sampson, D.B., 2004. Bias and precision of estimates from an age-structured stock assessment program in relation to stock and data characteristics. N. Am. J. Fish. Manage. 24, 865–879.
Zhou, S., Punt, A.E., Deng, R., Bishop, J., 2011. Estimating multifleet catchability coefficients and natural mortality from fishery catch and effort data: comparison of Bayesian state–space and observation error models. Can. J. Fish. Aquat. Sci. 68, 1171–1181.