Fisheries Research 158 (2014) 74–83

Contents lists available at ScienceDirect

Fisheries Research

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

A comparison of parametric, semi-parametric, and non-parametric approaches to selectivity in age-structured assessment models

James T. Thorson*, Ian G. Taylor

Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Boulevard East, Seattle, WA 98112, United States

ARTICLE INFO

Article history:
Received 9 May 2013
Received in revised form 4 October 2013
Accepted 7 October 2013
Available online 2 November 2013

Keywords:
Selectivity at age
Semi-parametric
Non-parametric
Random effect
Integrated assessment model
Cross-validation

ABSTRACT

Integrated assessment models frequently track population abundance at age, and hence account for fishery removals using a function representing fishery selectivity at age. However, fishery selectivity may have an unusual shape that does not match any parametric function. For this reason, previous research has developed flexible 'non-parametric' models for selectivity that specify a penalty on changes in selectivity as a function of age. In this study, we describe an alternative 'semi-parametric' approach to selectivity, which specifies a penalty on differences between estimated selectivity at age and a pre-specified parametric model whose parameters are freely estimated, while also using cross-validation to select the magnitude of penalty in both semi- and non-parametric models. We then compare parametric, semi-parametric, and non-parametric models using simulated data and evaluate the bias and precision of estimated depletion and fishing intensity. Results show that semi- and non-parametric models result in little decrease in precision relative to the parametric model when the parametric model matches the true data-generating process, but that the semi- and non-parametric models have less bias and greater precision when the parametric function is misspecified. As expected, the semi-parametric model reverts to its pre-specified parametric form when age-composition sample size is low but performs similarly to the non-parametric model when sample size is high. Overall, results indicate few disadvantages to using the non-parametric model given the range of simulation scenarios explored here, and that the semi-parametric model provides a selectivity specification that is intermediate between parametric and non-parametric forms.

Published by Elsevier B.V.

1. Introduction

Stock assessment models are designed to integrate a variety of data types to give an estimate of the productivity and both current and historical status of fish stocks. Models are then used by fishery managers to select among alternative management actions that balance tradeoffs between present and future risks and various stakeholder interests (Hilborn and Walters, 1992). Modern 'integrated' assessment models (Maunder and Punt, 2013) use a process model to project historical biomass given proposed parameters, and a measurement model to estimate the probability that available data would have occurred given the model and those parameters. This likelihood is used to estimate parameters and their precision, although some parameters (e.g., data weights or model variances) may be tuned externally to improve other measures of model fit, and residual errors between estimates and data can be assessed to evaluate model goodness-of-fit.

Assessments frequently partition stock biomass and abundance into different groups, defined by differences in age, length, sex, location, and growth characteristics (among others). This allows models to estimate different demographic rates for different groups (e.g., increasing fecundity by age). However, it also leads to additional complexity, whereby estimates of exposure to and intensity of fishing must be estimated for each group. The most common type of structure remains age structure, and selectivity at age is necessary in most modern assessment models. Assessment models may also specify selectivity as a function of length, because measurements of length are often logistically easier, less expensive, and more precise than measurements of age. In addition, changes in growth over time may have less effect on selectivity at length than selectivity at age. However, age-structured models typically convert length-based selectivity into selectivity at age using the expected distribution of age at length (Methot and Wetzel, 2013).

'Population-level' selectivity integrates many different processes and scales (Millar and Fryer, 1999). At the smallest spatial scale, fish near a given sampling or fishing gear will be subject to

* Corresponding author. Tel.: +1 206 302 1772.
E-mail addresses: James.Thorson@noaa.gov (J.T. Thorson), Ian.Taylor@noaa.gov (I.G. Taylor).

0165-7836/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.fishres.2013.10.002

contact selection, wherein older fish might be more or less susceptible to fishing gear. At a larger spatial scale, a portion of the population may not be available to fishing or survey gear (e.g., younger fish might be high in the water column and hence escape a bottom trawl), and unavailable fish will decrease population-level selectivity for their age. Finally, large-scale spatial differences in fishing intensity can cause the relative proportion of abundance in fished and unfished regions to change as a function of age, thus causing the shape of population-level selectivity to differ from contact selectivity (Sampson and Scott, 2011).

Population-level selectivity may therefore have many possible shapes, but can be broadly classified as monotonically decreasing, dome-shaped, or monotonically increasing (Bence et al., 1993), where monotonically decreasing fleets have maximum selection for juveniles, dome-shaped fleets for intermediate age, and monotonically increasing fleets for adults. Parametric models exist for these categories of selectivity, and have many or few parameters depending upon the complexity with which selectivity is approximated. Misspecifying selectivity as dome-shaped when it is asymptotic can be confounded with other model misspecifications and will lead to biased estimates of status (Taylor and Methot, 2013), as will misspecifiying selectivity in general (Punt et al., in press). More seriously, aggregate selectivity for a fishery may be unusually shaped (e.g., bimodal in the sablefish hook-and-line fishery, Stewart et al., 2011a,b) due to spatial fishing patterns (Sampson and Scott, 2012), and there are few techniques for estimating selectivity in a flexible and generic manner.

One generic approach to estimating selectivity at age is 'non-parametric' models. These models estimate selectivity for each age, while penalizing large changes in selectivity between ages. Non-parametric methods are well-established in time-series and regression models (e.g., generalized additive models) and are generally implemented using smooth- or fixed-knot splines (Thorson et al., 2013; Wood, 2006), but have not been widely tested in integrated assessment models (where freely estimating selectivity for each age is the limit of a smooth-spline with a penalty approaching zero). An alternative approach is 'semi-parametric' models (e.g., Sugeno and Munch, 2013a,b), which we define as a functional relationship (e.g., the stock–recruitment relationship) that contains an informative and explicit prior function (e.g., a Ricker curve, [Sugeno and Munch, 2013b]) but which allows available data to update this function when predicting future observations (e.g., a flexibly shaped stock-recruit curve used for predicting rebuilding). Semi-parametric functions are thus similar to non-parametric functions in allowing a flexible shape for predicting future observations, but differ by having an informative and explicit prior function (e.g., generalized additive models have an implicit prior function that is difficult to visualize or interpret). In the case of selectivity at age, a semi-parametric model would estimate selectivity for each age but penalize deviations away from an estimated parametric model for selectivity. Although semi-parametric methods are increasingly used in ecological models (Thorson et al., in press-a), they have not to our knowledge been applied to any age-structured assessment model.

Non- and semi-parametric approaches differ in several ways. Most importantly, semi-parametric models penalize selectivity toward a function that is specified explicitly (via a parametric prior), whereas non-parametric models penalize selectivity toward a function that is implicit and may be hard to visualize and interpret. For example, generalized additive models (a type of non-parametric model) have an implicit prior function, but it is difficult to specify a different prior function in the case that an analyst has prior knowledge of the function. However, semi-parametric models estimate additional parameters beyond those required for the parametric model that is specified a priori, and hence may be more computationally intensive. Finally, semi- and non-parametric approaches share a common difficulty of selecting an appropriate penalty for deviations. This penalty may be selected using Markov chain Monte Carlo (De Valpine, 2004) or maximum marginal likelihood (Bolker et al., 2013), but both methods may be slow or computationally intensive. This penalty may be selected by cross-validation techniques, but this has not been extensively tested (with the exception of Maunder and Harley, 2011).

In this study, we present semi- and non-parametric approaches to selectivity, and compare their performance with the conventional parametric approach in an age-structured assessment model. We use cross-validation to select the penalty for semi- and non-parametric approaches, and illustrate when each method performs better or worse given plausible data scenarios.

Table 1
Parameter names, symbols, value in the simulation model (varies: it depends upon the simulation configuration), and whether it is estimated in the estimation models (NA: parameter is not in the estimation models; fixed: parameter is fixed at its true value).

Parameter name	Symbol(s)	Simulation value	Estimated?
Equilibrium biomass ratio	A	0.25	*NA*
Effort dynamics rate	X	0.2	*NA*
Effort dynamics variability	σ_E	0.1	*NA*
Fishery catchability	q_E	1	*NA*
Natural mortality rate	M	0.2	Varies
Selectivity parameters	$\delta_1, \gamma_1, \delta_2, \gamma_2$	Varies	Varies
Natural logarithm of average unfished recruitment	$\ln(R_0)$	13.8	Yes
Steepness	H	0.6	Fixed
Standard deviation of recruitment variability	σ_R	0.5	Fixed
Brody growth coefficient	K	0.2	Fixed
Length at $a=0$	L_0	0.1	Fixed
Asymptotic maximum length	L_∞	1	Fixed
Weight at length coefficient	A	1	Fixed
Allometric weight at length	B	3	Fixed
Maturity at age	a_{mat}	3	Fixed
Survey age at 50% selection	δ_s	2	Yes
Survey selectivity slope	γ_s	1	Yes
Survey catchability	q_S	1	Yes
Survey variability	σ_S	0.4	Fixed
Fishery compositional sample size	n_F	Varies	Fixed
Survey compositional sample size	n_S	4	Fixed
Number of simulated years	n_{years}	25	Fixed
Maximum age	A	20	Fixed
Maximum age at which fishery or survey selectivity is estimated (older ages are fixed to selectivity at this age)	a_{max}	10	Fixed

2. Methods

We first describe a simulation model that was used to generate age-structure data, and then describe the parametric, semi-parametric, and non-parametric estimation models (parameter names, symbols, and values are summarized in Table 1, and variables in Table 2). Last, we describe the criteria that were used to evaluate model performance.

2.1. Simulation model

We simulate data for a fishery for which nominal fishing effort E follows a recently proposed effort-dynamics model (Thorson et al., in press-b):

$$\ln(E_{t+1}) \sim \text{Normal}\left(E_t\left(\frac{SB_t}{B_e \cdot SB_0}\right)^x - \frac{\sigma_E^2}{2}, \sigma_E^2\right) \quad (1)$$