Fisheries Research 261 (2023) 106642

Contents lists available at ScienceDirect

Fisheries Research

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

Those who fail to learn from history are condemned to repeat it: A perspective on current stock assessment good practices and the consequences of not following them

André E. Punt [a,b,*]

[a] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98185-5020, USA
[b] CSIRO Oceans and Atmosphere, GPO Box 1538, Hobart, TAS 7001, Australia

ARTICLE INFO

Keywords:
Fisheries assessment
Integrated analysis
Management
Uncertainty

ABSTRACT

The ideal stock assessment would be able to estimate all of the key parameters related to population processes within a framework that assigns appropriate weight to the data, fits the data adequately, and captures all sources of uncertainty related to estimation, including model uncertainty, process uncertainty, and observation uncertainty. The aim of good practice guidelines is to avoid the pitfalls of earlier analysis methods, and consequently provide assessments that reflect objective scientific information on which management decisions can be based. This paper outlines a framework for the component of a stock assessment related to fitting population dynamics models to monitoring data to support decision making, which follows from what would be considered good (but not necessarily best) practice in the field. The paper identifies current good and best practices related to selecting a model structure, parameterizing growth, recruitment, natural mortality and the stock-recruitment relationship, as well as how to select among model configurations based on diagnostics and weight data and priors within assessments based on the existing literature, including past Center for the Advancement of Population Assessment Methodology (CAPAM) workshop reports and the results of simulation studies that explored the performances of different ways to configure stock assessments.

1. Introduction

Fisheries stock assessments form the quantitative basis to support decision making. They are used primarily to provide estimates of stock status (biomass and fishing mortality relative to reference points), estimates of catch limits / effort that will achieve management goals, often based on harvest control rules, and to form the basis for projections and management strategy evaluations (MSE). The 'ideal' stock assessment is based on a model of the population dynamics that adequately matches the reality of the system being modelled, capturing just the correct number of processes to enable accurate and precise outcomes to be created. The ideal stock assessment would be able to estimate all of the key parameters related to natural mortality, growth, recruitment, selectivity and movement, within a framework that assigns appropriate weight to the data, fits the data adequately and captures all sources of uncertainty related to estimation, including model uncertainty, process uncertainty, and observation uncertainty. Reality naturally seldom matches this ideal, and the aim of best (or good) practices is for stock assessments[1] to come as close as possible to the ideal given the limitations of system understanding, data quality and quantity, and computational limitations.

The earliest quantitative model-based stock assessments were based on simple production models and virtual population analysis (see the guides for conducting these methods and other early methods of stock assessment by Gulland, 1969, 1983, 1988 and Pauly, 1984). These methods have a limited number of options so the number of types of best practices are quite limited. For example, for production models the decisions to make relate to the form of the production function (e.g., Schaefer, Fox or Pella-Tomlinson), whether the errors are assumed to be observation or process, how to quantify uncertainty, and how to weight alternative indices of abundance. For virtual population analysis, the key decision is related to the "tuning algorithm" used to set the terminal

* Corresponding author at: School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98185-5020, USA.
E-mail address: aepunt@uw.edu.

[1] 'Good practices' are defined here as decisions that can made and implemented using existing stock assessment packages while 'best practices' will require future research on model development and implementation.

https://doi.org/10.1016/j.fishres.2023.106642
Received 2 November 2022; Received in revised form 2 January 2023; Accepted 27 January 2023
Available online 1 February 2023
0165-7836/

numbers-at-age for each cohort (equivalent to the terminal fishing mortalities) (e.g., Pope and Shepherd, 1985).

Contemporary methods of stock assessment are based on the 'integrated analysis' paradigm, in which the model of the population dynamics is developed separately from that of the model that relates the data to the population dynamics model (the observation model). Stock assessments based on this paradigm have increased in sophistication since the original idea was outlined by Fournier and Archibald (1982) and several major stock assessment software packages now exist that implement various forms of integrated analysis assessment (e.g., a4a: Jardim et al., 2015; Stock Synthesis: Methot and Wetzell, 2013; CASAL [C++ Algorithm Stock Assessment Library]: Bull et al., 2012; MULTIFAN-CL [MULTIple length Frequency ANanlyis - Catch at Length]: Fournier et al., 1998; GADGET [Globally applicable Area Disaggregated General Ecosystem Toolbox]: Begley, 2014; SAM [State-space Assessment Model]: Nielsen and Berg, 2014; Berg and Nielsen, 2016; and WHAM [Woods Hole Assessment Model]; Miller and Stock, 2020; Stock and Miller, 2021) These packages contain myriads of options for the model of the population dynamics and that of the observation process, meaning that given a single data set, the final assessment may depend on the philosophy of the analyst. However, while professional judgement is always part of statistical modelling, the expectation is that analysts make use of what is generally considered best (or at least good) practice.

The technical and statistical sophistication of the methods used to construct and fit integrated stock assessments means that given a data set, different analysts may construct quite different assessments. Ralston et al. (2011) and Punt et al. (2018) highlight that the variation in biomass trajectories among assessments of the same stock conducted in different years can be attributed not only to additional data points for existing data sources and the availability of new data sources, but also to the choices made by the analysts and the groups tasked to review draft assessments. As will be noted below, these choices relate to many components of the population dynamics model used for assessment purposes, the data and how they are weighted, and most recently the statistical paradigm used for parameter estimation, with methods that treat model parameters as random effects becoming increasing common but underutilized (e.g., Nielsen and Berg, 2014; Thorson, 2019).

The importance of best practices (or at least good practices) cannot be overemphasized given the consequences of inappropriate decisions when conducting stock assessments on the ability to achieve management goals as well as the credibility of the scientific process. The workshops conducted as part of Center for the Advancement of Population Assessment Methodology (CAPAM) [Selectivity: Maunder et al. (2014); Growth: Maunder et al. (2016); Data weighting: Maunder et al. (2017), Diagnostics: Maunder et al., In press; Recruitment: Sharma et al. (2019); Spatial structure: Cadrin et al. (2020); Natural mortality: Hamel et al. (2023); Next generation stock assessment packages: Hoyle et al. (2020)] have established the basis for best practices for stock assessment. Several jurisdictions have been developed Terms of Reference for conducting stock assessments, which outline the requirements for stock assessment reports, and to assist with the review process (e.g., PFMC, 2020) and some 'acceptable' practices have been developed (e.g., PFMC, 2021). However, to date, no synthesis of good or best practices for how to conduct contemporary 'statistical' methods has been conducted, although several texts on stock assessment (e.g., Hilborn and Walters, 1992; Quinn and Deriso, 1999; Haddon, 2011) provide suggestions, which the best practices of this paper build on.

There is consequently a need to synthesize good and best practices for conducting stock assessments, which likely depend on the aims of the assessment (e.g., to provide unbiased estimates of current and historical biomass, versus to provide precise estimates of biomass for use in harvest control rules). Table 1 summarizes the factors considered in a stock assessment that analysts need to take into account when constructing an assessment based on the perspective of the author. There are many other decisions that need to be made when conducting stock assessments, but in the experience of the author the items in Table 1 are the decisions that are most commonly influential in terms of inferences about stock status and levels of sustainable catch.

Table 1
The questions that must be answered when conducting a stock assessment that are most influential in terms of results for a baseline standard assessment.

Basic structure
How is the spatial / stock structure of the assessment selected?
How are the number of sexes, age- and length-classes selected?
What is the time period considered by the model, including the projection period?
What is the model time-step?
How are the fisheries and surveys selected and then aggregated for analysis?
Is the stock in quasi-equilibrium at the start of the modelled period?
Biological parameters
How is natural mortality modelled (functional form and age-, sex-, and time-varying?)?
How is growth modelled (functional form, empirical vs. parametric, and age-, sex-, and time-varying?)?
How are the growth and natural mortality parameters set (estimated, with priors, or based on auxiliary analyses)
Stock and recruitment
Which functional form and which parameters estimated?
Which parameters are estimated and which are pre-specified based on auxiliary information?
Are recruitment deviations treated as random effects or is penalized likelihood applied?
How is account taken of the lognormal bias-correction factor?
Fishery and survey Selectivity
Is selectivity a function of age, size or both? How flexible is the relationship?
Does selectivity vary over time and/or between sexes?
Is selectivity asymptotic for some or all of the fisheries and surveys?
Diagnostics
Which diagnostics to apply?
Can diagnostics identify model mis-specification and help with model weighting?
Data weighting
How to specify the initial weights for each data source / prior?
How to update the weights given model fit?
General issues
Are recruitment deviations treated as random effects or is penalized likelihood applied?

This paper therefore aims to synthesize previous work on good practices related to fitting integrated population dynamics models to monitoring data based on the work of CAPAM and other research, with a view to providing a guide to conducting stock assessments. It illustrates the some of consequences of not following best practices using a set of simulations based on a simple assessment (see Supplementary Appendix A for a mathematical description of the operating model on which the simulations are based and the various estimation methods that are tested[2]). The final section synthesizes the rest of the paper in the form of a 'recipe' that should lead to 'good' assessments, with a clear focus on assessments for populations for which there are minimally reliable data on removals, an index of abundance, and some measure of population/catch composition. The paper will distinguish between "good" and "best" practices given the practical and computational constraints when conducting assessments. The quality of an assessment depends on the available data. However, developing good/best practices for data collection is beyond the scope of this paper – and in most cases assessments are required irrespective of the data available. The recipe is focused on conducting stock assessments, but most of good and best practices also apply to the development of the operating models for management strategy evaluation (Punt et al., 2016).

[2] The simulation study is designed so that a self-test leads to unbiased and fairly precise estimates of management- related quantities (Fig. S.5; Table S.5).

2. Population models, spatial structure and process error

2.1. Population models

There are many ways to conduct a stock assessment that aims to estimate biomass and how it has changed over time in response to removals from the population and environmental effects. The most common of these are biomass dynamics (aka surplus production) models, age-structured models and size-structured models. Most model-based assessments are conducted using age-structured models, with size-structured models used primarily for hard-to-age species such as prawns, crabs and rock lobsters. Biomass dynamics models can be used to conduct stock assessments for species for which the only data are a time-series of catches and one or more indices of abundance (e.g., Winker et al., 2018). However, they are not used for the world's primary fisheries, but might be for non-target and byproduct species.

Age-structured models can be used to predict the population size-composition usually under the assumption that the distribution of size-at-age is not impacted by fishing (e.g., Methot and Wetzell, 2013). Consequently, unlike size-structured models, most age-structured models ignore the impact of size-selective mortality (Dichmont et al., 2016). The disadvantage of using age-structured catch-at-size methods is that the modeled size-at-age distributions do not change over time. This is potentially consequential for stocks[3] that are managed using a minimum legal size or for which selectivity is close to knife-edged (as is the case for many crab stocks) and experience very high fishing mortalities for legal individuals, such that size-specific fishing mortality will change the size-at-age distributions in the population. Age-size models combine the benefits of age-structured and size-structured models within a single framework but at the cost of (substantially) increased computational costs. Deriso and Parma (1988) outlined a full age-size-structured population dynamics model and describe the likelihood function that could be used to estimate its parameters for Pacific halibut, *Hippoglossus stenolepis*, and Quinn et al. (1998) extended this approach by discretising the size distributions. Gilbert et al. (2006) created an age- and size-structured model for New Zealand snapper, *Pagrus auratus*, which allowed growth to be a function of both age and size and to vary over time, and Allen Akselrud et al. (2017) developed an age- and size-structured model and applied it to data for Pacific cod *Gadus macrocephalus* in the Eastern Bering Sea. Best practice would therefore be to use an age-size model, but these assessment methods are still in development and a key component of the next generation of stock assessment methods (Punt et al., 2020) so good practice is to base assessments on age- or size-structured models.

Most stock assessments involve an annual time-step. However, some species, such as short-lived coastal pelagic species and prawns that grow quickly during the year require much shorter time-steps (e.g., quarters for South African anchovy: De Moor et al., 2011; weekly for prawns in northern Australia: Dichmont et al., 2003).

The initial conditions for the model need to be specified. The ideal is to treat the numbers-at-age or -size at the start of the first year as estimable parameters (as is the case for the assessment of red king crab *Paralithodes camtschaticus* in Bristol Bay, Alaska, Zheng et al., 2021), but good practice is to compute the initial conditions by calculating numbers-at-age or -at-sizes under the assumption that the stock was in equilibrium given an estimated fishing mortality (which can be set to zero for populations for which catches are available since the start of the fishery) and then adding recruitment deviations to the resulting numbers-at-age thereby estimating the sizes of the cohorts that entered the population before the start of the modelled period.

A final consideration is the first year considered in the model. This can be the first year with removals, the first year with reliable estimates of removals, the first year for which recruitment is informed by data, selected with the aim of providing management advice, or selected to be after all the things that might bias the assessment ended (e.g., large oceanic regime shifts, changes in fishing technology, erroneous or incorrect estimates of removals, and changes in management regimes).

[3] The term "stock" is used in this paper in the sense of a management unit while the underlying biological units will be referred to as "populations". In some cases, "populations" are "stocks", but this is not always the case.

2.2. Spatial and stock structure

A critical part of conducting a stock assessment is how to deal with spatial and stock structure. All fish and invertebrate populations exhibit spatial structure to some extent (Berger et al., 2017), which can be reflected as spatial variation in demographic parameters as well as in fishing mortality (Cadrin et al., 2020). The consequences of mis-specification of spatial and stock structure include bias in estimates of management-related quantities and an inability to achieve management goals. Many studies have shown that accounting for space when conducting assessments reduces these problems to some extent, although often at the expense of reduced precision of model outputs (Punt, 2019, and references therein).

Accounting for spatial structure can take two main forms apart from ignoring it altogether. The 'areas as fleets' approach (e.g., Berger et al., 2012; Hurtado-Ferro et al., 2014; Waterhouse et al., 2014) approximates the spatial distribution by size and age caused by movement or differences in exploitation rates or biological differences (e.g., recruitment) using selectivity. This approach is very widely adopted given that the underlying population dynamics model does not explicitly model spatial structure and hence there are fewer additional estimable parameters and computational demands than for spatial models. However, the 'areas as fleets' approach can lead to severe bias in estimates of management-related quantities (Punt et al., 2015; Supplementary Table S.8). The alternative (a true spatial model) is to base the assessment on a population dynamics model that explicitly allows for spatial structure. This involves dividing the population into a (small) number of spatial cells (but see the method of Dunn et al., 2015) and modelling movement among the cells.

The key decisions to make when conducting a spatial assessment are (a) how many populations and subpopulations to include in the population dynamics model, (b) how recruitment and movement are to be modelled, and (c) the number of fleets and areas. The first (and often most important) step when conducting a spatial stock assessment is identify a conceptual model (or set of conceptual models) that characterize the system (how many populations and subpopulations and where they are likely to be found by sex, age, size, etc). This involves first selecting a definition for a 'population' and perhaps a 'subpopulation'. Punt (2019) suggested defining a 'population'[4] as a biological unit that does not exchange individuals with other biological units, and 'subpopulations' as biological units that exchange individuals sufficiently often that their demographics are linked, such that dispersal from other subpopulations would constitute an appreciable proportion of the rate of recovery of a subpopulation that is depleted by fishing. However, this distinction between populations and subpopulations does not deal with the case in which there is exchange between biological units but of a magnitude that is too small to impact the population dynamics appreciably. In such cases, ignoring exchange is likely an appropriate (and simpler) practice. Other definitions of the terms "stock" and "population" exist (e.g., Goethel and Berger, 2017) and there would be value in the community reaching agreement of terminology.

The following are potential stock structure 'archetypes': (a) a single population found in multiple areas, (b) a single population found in multiple areas but with no post-settlement movement, (c) multiple populations that are located in the modelled region with movement

[4] Punt (2019) used the terms "stock" and "sub-stock" for "population" and "subpopulation" as used here,.

among areas but no dispersal (permanent movement among populations), (d) multiple populations/ subpopulations located in the modelled region, with movement among areas and dispersal among populations/ subpopulations, (e) multiple populations/ subpopulations located in the modelled region, with movement among areas and dispersal among populations/ subpopulations and natal homing, and (f) multiple populations/ subpopulations located in the modelled region, but with no movement among areas or dispersal among them. Model configurations that explicitly include spatial structure should explore local and global density-dependence and allow for both annual and spatial deviations in recruitment about the stock-recruitment relationship (unless data show that the latter is inconsequential).

Punt (2019) recommends starting with a large number of potential areas and reducing the number of areas based on data availability and model selection methods. Age- / length-frequency data for common gear types and tagging data can be used to identify areas that differ spatially in terms of population structure or biological characteristics (e.g., Lennert-Cody et al., 2010, 2013; Maunder et al., 2022). In general, ecological boundaries among the spatial areas in an assessment are preferable. Unfortunately, in many cases there is not a clear boundary between populations (e.g., in situations with clinal changes in genetic structure). Consequently, often boundaries based on management jurisdictions are the only possibility given data availability. However, this may be appropriate given these areas also usually define how management regulations will be implemented.

There are many more possible options when constructing a spatially-structured rather than a spatially-aggregated population dynamics model, so it is best to start with multiple models and remove models that (a) cannot be supported by the data, (b) have convergence or confounded parameter problems, (c) are unable to the fit the available data, (d) lead to model mis-specification / retrospective patterns, or (e) lead to estimates that are inconsistent with conceptual basis for the model (e.g., estimates of movement that decrease rather than – as expected – increase with age). Biological parameters (growth, natural mortality, fecundity) may differ among populations and subpopulations, but it would seem appropriate to place a prior on the extent to which these parameters differ among populations to ensure that the information for one population provides some information / bounds for the dynamics of the other populations.

One of the major challenges associated with spatial models is how to parameterize the rates of movement and dispersal (and how they differ among ages and sexes). In principle, such rates can be estimated using changes over time in age- and size-compositions (e.g., McGilliard et al., 2015), but generally estimation is based on integrating tagging data into the assessment by modelling multiple tagged populations, along with the actual population (e.g., Hilborn, 1990; Maunder, 2001; Goethel et al., 2019; Vincent et al., 2020). In principle, movement rates can be estimated outside of the assessment but the common problems of the inability to propagate uncertainty and the potential lack of consistency of assumptions remains.

Spatial variation in population demography is a major challenge when applying spatial models. This is particularly the case for growth when size-composition data are a major source of information on abundance and trends because assuming (incorrectly) that growth is spatially-invariant will lead to differential estimates of population depletion even if this not the case in reality. Another concern with spatial variation in growth is how to model mean growth when animals move among areas.

2.3. Modelling process error

Process error, random natural variation about the expectation for a model quantity, is a key source of uncertainty for any stock assessment. Process error is usually included in stock assessments in the form of variation in recruitment about the stock-recruitment relationship. However, almost every process in a stock assessment (e.g., growth, recruitment, natural mortality, movement, and selectivity) is potentially subject to process error. Traditionally, process error has been included in stock assessments using "penalized likelihood" in which the deviations in a process about its expectation are treated as fixed effects parameters, with the model objective function extended to include a penalty in the form of a normal density with mean zero and a pre-specified standard error. Some authors have attempted to estimate the process error standard error (e.g., Maunder and Deriso, 2003) within the penalized likelihood framework, but any non-zero estimate is convergence to a local minimum. Fig. S.5 shows that it is possible to estimate the extent of variance in recruitment, particularly given reliable data on age-composition when the model is represented in state-space formulation. The ability to estimate the variance in recruitment will depend on how imprecise (and biased) the age data are, and the amount to which ageing error can be quantified.

Process error variances can be estimated within the Bayesian paradigm given a prior distribution for these parameters, and within the frequentist paradigm by formulating the assessment as state-space model, which involves integrating out the random effects and maximizing a marginal likelihood. Nevertheless, the estimates of process error variances can be quite imprecise (and even biased), as is clear from the simulations in Fig S.5.

The ability to implement state-space stock assessments has increased considerably with the availability of Template Model Builder (Kristensen et al., 2016), and two stock assessment packages (SAM and WHAM) are formulated as state-space models. There are, however, constraints to applying state-space models including computational cost, the fact that attempting to estimate too many random effects processes can often lead to degenerative solutions (i.e., estimated zero variance for some sources of process error). Nevertheless, and notwithstanding the technical challenges associated with state-space models (in Bayesian or frequentist modes), their use remains best practice.

3. The key processes: growth, selectivity, natural mortality and recruitment

3.1. Growth

Growth affects conventional age- and size-structured stock assessments in multiple ways (Francis, 2016): (a) to convert catches from weight to numbers, (b) to convert from numbers-at-age and -at-size to weight-at-age and -at-size, (c) to convert length-based selectivity to age-based selectivity, and (d) to compute expected length-compositions. Error when modelling growth can consequently lead to bias in estimates of management-related quantities, most noticeably when the only data available for assessment purposes are size-composition data, owing to the confounding between selectivity, growth and fishing mortality (Maunder et al., 2016).

Many assessments model growth using parametric relationships. In the case of age-structured assessments, length-at-age is usually modelled using the von Bertalanffy growth equation or a generalization thereof such as the Schnute growth model (Schnute, 1981), with weight-at-age computed from length-at-age according to an allometric length-weight relationship, accounting for the effects of variation about the growth curve (Methot and Wetzell, 2013). In size-structured models, a size-transition matrix, where the expected growth increment is modelled parametrically, is used to represent the probability of animals growing from one size-class to another. Some assessment methods (such as AMAK and ASAP: Anon, 2015; Legault and Restrepo, 1998) do not model growth using a parametric form but instead specify weight-at-age (often by year) under the assumption that sampling for age-composition is sufficiently precise to enable the assumption that weight-at-age is known exactly (little ageing or sampling error) to be justified. This "empirical weight-at-age" approach can also be included in assessment frameworks such as Stock Synthesis.

Growth models should consider not only the expected growth increment but also how individual variation in growth is captured. It is often assumed that variation in growth can be approximated by a coefficient of variation, which is usually assumed to be independent of age and expected length. This approach is consistent with the observation that variation in growth is most evident for asymptotic size (Lorenzen, 2016), but this should be tested in applications. In general, all of the parameters of the growth curve should be estimated, and the von Bertalanffy growth curve should be not be adopted automatically. Rather, alternative forms such as the general Schnute growth model and curves that allow for changes in growth rate with reproduction (e.g., Minte--Vera et al., 2016) should be examined. Growth can be modelled non-parametrically (e.g., using a spline) but this is not common for statistical catch-at-age assessments. Age-length models (e.g., Allen Akselrud et al., 2017; Begley, 2014) can be used to capture individual variation in growth and hence the effect of fishing mortality on length-at-age (i.e., 'the Rosa Lee phenomenon'). Taylor and Methot (2013) provide a computationally efficient way to allow for the Rosa Lee phenomenon by dividing the population into 'platoons', each of which has its own growth curve, but this feature has been used rarely in actual assessments.

The data available to estimate growth can be divided into direct data, i.e., information from tagging on growth increments (commonly included in assessments for hard-to-age species such as prawns, crabs and rock lobsters; Punt et al., 1997, 2013), and on size-at-age, usually included in age-structured assessments in the form of conditional at-age-length data. The use of conditional-at-age data leads to an accurate characterization of growth because the effect of selectivity is accounted for when computing the model-predicted distribution of the age-composition of each size-class (Piner et al., 2016). In principle, data on the size-composition of catches or surveys provide information on growth, but size-composition distributions reflect the combined effects of growth, recruitment, fishing mortality rate and selectivity, requiring that selectivity be known to enable growth to be estimated accurately.

There is strong evidence that growth varies by sex and over time (Stawitz et al., 2015; Thorson and Mine-Vera, 2016) and assessments should by default test for time-variation in growth and for sex-differences in growth rates. Time-variation in growth can be modelled within a stock assessment by assuming that the parameters of the growth curve differ among years and cohorts (Methot and Wetzell, 2013) or that the growth rate by cohort depends on year-class strength (Punt et al., 2001). Modelling time-variation in growth using random effects for deviations in parameters from their expectations makes forecasting challenging, highlighting the value of identifying environmental covariates that explain interannual variation in growth increments or parameters (e.g., Punt et al., 2021a).

3.2. Selectivity

Stewart and Martell (2014) define selectivity as length- or age-based probabilities used to link observed composition data to model predictions about population abundance-at-age/-size. Selectivity is a combination of two processes: availability (i.e., the probability that a fish of a specific age or size is in the same vicinity at the same time as gear deployment) and contact (or gear) selectivity (i.e., the relative probability that a fish of specific age or size is caught given it is available to the gear) (Privitera-Johnson et al., 2022). Selectivity is perhaps the population process with the most options among which to choose and for which analysts differ the most. The three main decisions that need to be made when conducting assessments relate to whether selectivity is length- or age-based (or both), the functional form for selectivity, and whether (and how) allowance is made for time-variation in selectivity.

Selectivity is often modelled as a function of length rather than of age, and it is tempting to assume that age-specific selectivity can be computed from length-specific selectivity according to $S_a^f = S_{L_a}^f$ where L_a is the expected length of an animal of age a. Unfortunately, this estimator will be biased when there is variation in length-at-age. The correct estimator of the expected selectivity of an animal of age a is:

$$S_a^f = \int S_L^d \, P(L|a) \, dL \tag{1}$$

where S_L^f is selectivity as a function of length, and $P(L|a)$ is the probability of an animal of age a being of length L (Methot and Wetzell, 2013). The selection between age- and length-based selectivity depends on the situation. For most cases, contact selectivity will usually be related to length-based processes so when contact selectivity is primary cause for non-uniform selection, length-based selectivity is the most appropriate decision. This is also the case when the population dynamics model is length- rather than age-structured (Punt et al., 2013). Table S5 and Fig. S5 show that length-based selectivity is difficult to estimate using only age-composition data. Age-based selectivity is an appropriate formulation when the only source of composition data is age data (no length-composition or conditional age-at-length data) or when availability is age-based (e.g., due to ontogenetic shifts in distribution). An advantage of age-based selectivity is that there is no need to model growth explicitly and hence estimate $P(L|a)$ though there is a need to account for weight- (or biomass-) at-age in most cases.

In principle, the selectivity for each age-class can be treated as an estimable parameter, but in most cases selectivity-at-age (or –at-length) is governed by a parametric (or semi-parametric) equation. Unfortunately, the functional form assumed for selectivity may often be incorrect or very inflexible (Lee et al., 2014). For example, it is common to assume that the selectivity is asymptotic (the oldest or largest fish are fully selected by the fishery), but both theory and empirical evidence indicate that fish movement and availability likely lead to some doming in most cases (Sampson et al., 2011; Sampson, 2014; Waterhouse et al., 2014), and that misspecification can substantially influence assessment results (Ichinokawa et al., 2014; Wang et al., 2014; Privitera-Johnson et al., 2022). However, incorrectly assuming that selectivity is dome-shaped for all fleets can create a cryptic biomass (Crone et al., 2013), and result in biased estimates of management-related quantities. Conversely, incorrectly assuming asymptotic selectivity will force the assessment towards a result that has higher total mortality and greater stock depletion and hence biased estimates of management-related quantities (Fig. S7).

Crone et al. (2013) note that it is common practice in stock assessments to assume asymptotic selectivity for at least one fishery or survey to stabilize parameter estimation. This is because the declining limbs of the selectivity patterns when there is dome-shaped selectivity for all fisheries and surveys are inherently confounded with natural mortality and this confounding will often increase the uncertainty in abundance estimates. Assuming that selectivity for one fishery or survey is asymptotic will lead to more pessimistic estimates (and poorer fits to data) than dome-shaped selectivity.

Privitera-Johnson et al. (2022) compared asymptotic (logistic) and dome-shaped (double-normal) and flexible (spline-based) selectivity patterns using simulations for three species. They also explored the case in which the selectivity pattern is chosen using AIC. The results suggest that using AIC to select among selectivity forms is not robust, including when model misspecification is absent. The use of double normal selectivity was found to be most robust to uncertainty in the true form of selectivity. However, the double normal form performed poorly if M was estimated along with the other model parameters. Similarly, use of flexible parametric methods, such as splines, performed adequately with informative data, but poorly when the catch series exhibited low contrast and age-composition data were not available from the start of the fishery. This suggests that the best practices for selectivity will depend on knowledge of the likely information content of the data.

The flexibility of the double-normal selectivity pattern is often sufficient to mimic the wide range of single-peaked shapes that may be expected from a single fishing gear type. The degree of stiffness in fishery selectivity seeks to strike a balance between two goals. One goal is to be highly flexible so that the modelled selectivity will be highly informed by the fishery size- and age- composition data, causing the modelled removals to mimic the data, as in VPA. The other goal is to obtain direct inference about the population from which the fishery is taking fish. Here a stiff (few parameter) selectivity retains more degrees of freedom such that the fishery size- and age- composition data are a sample of the population as filtered by that stiff selectivity. Integrated analysis models seek to accomplish both goals. In an ideal, data-rich setting the best configuration should be to use a very stiff and experimentally verified selectivity for the survey and a very flexible selectivity for the fishery.

Generally, fishery selectivity patterns should be chosen that have the fewest parameters, and allow an acceptable fit to the available composition data. As surveys are designed to at least use the same fishing gear throughout, a good reason to use more complex patterns than logistic or double-normal would be required for those. If a fishery has fairly homogenous gear, stable fishing behaviour, and stable areas of fishing, a similar argument applies there as well. In the case of a fishery with mixed gear types, an opportunity exists to use a less restricted pattern shape, as provided by the age-based random walk.

Allowing for some time-variation in selectivity either where selectivity is treated a random effect (e.g., Nielsen and Berg, 2014; Berg and Nielsen, 2016) or by superimposing autocorrelated random deviations in age (or size) and year on an underlying parametric form (e.g., Xu et al., 2019) is appropriate, especially when the extent of variation and autocorrelation are estimated. Care should be taken when adopting this approach given it can lead to complex (and perhaps unrealistic) patterns selected to match changes in age-/size-composition, perhaps due to combining data for multiple gear-types into a single fleet.

3.3. Natural mortality

Of the key parameters determining the population dynamics of marine species, natural mortality (*M*) is the least well known because direct data on deaths due to natural causes are seldom available (exceptions are species such as bivalves that leave behind articulated values; Doering et al., 2021; methods based on acoustic tagging, etc. Maunder et al., 2023). Estimation of *M* is made even more challenging because trends in *M*, recruitment and selectivity tend to be confounded (Butterworth and Punt, 1990; Thompson, 1994). For example, assuming that selectivity is asymptotic when it is actually dome-shaped will lead to an over-estimate of *M* using a catch curve-based estimator of *M*.

M has traditionally been estimated 'outside' of the assessment using methods based on maximum age, life history theory, and relationships between "well-known" (those found in the literature) values for *M* and covariates, use of tagging data and catch curve analysis (Punt et al., 2021b; Maunder et al., 2023). However, estimates of *M* from these 'indirect' methods can be substantially in error (Kenchington, 2014 indicates errors of 50–200%), and best practice is to estimate *M* (usually a constant independent age, sex and time) within an assessment with a prior based on information sources that are not included in the assessment. The estimates of *M* from stock assessments will be most reliable when the data are from an area closed to fishing and immigration and emigration (so the only source of mortality is natural), during fishing moratoria, or if age-composition data are available from when the population was essentially unfished (e.g., Punt et al., 2001 for the Australian stock of blue grenadier, *Macruronus novaezelandiae*), but these are rare situations. Estimates of *M* in the self-tests were quite precise especially when a prior on *M* was provided (Table S.5). However, estimates of *M* from actual stock assessments can be substantially in error if the assessment is mis-specified. Consequently, almost all estimates of *M* should be considered as being potentially grossly in error.

If *M* is to be set outside of the assessment model using an 'indirect' method, care needs to be taken to check that the pre-specified value is consistent with the remaining data sources (e.g., see Haddon, 2017) for a case where the pre-specified value for *M* is not consistent with the estimate from the data) and that it is based on the most appropriate method. Hamel (2015) provides an approach that uses multiple data types (maximum age, relationship with the von Bertalanffy *K* parameter, and relationship with the gonadosomatic index) to create a distribution for *M*, although a distribution based only the maximum age appears to be adequate in most cases (Hamel and Cope, 2022). It will be infeasible to estimate *M* in most data-limited situations, and in such cases, it is best practice to provide assessment results and management advice for a range of approaches for estimating *M* indirectly given the error associated with all indirect methods for estimating M (Cope and Hamel, 2022)

Although most assessments assume *M* is independent of age, time, size, density, and sex, this is unlikely to be the case in actuality. Age-specific natural mortality can be estimated as discrete changes in *M* that may coincide with the onset of sexual maturity (e.g., eastern Bering Sea snow and Tanner crab *Chionoecetes opilio* and *C. bairdi*; Stockhausen, 2019; Szuwalski, 2019). However, best practice is to assume that *M* is a continuous function of age (or size) and model it using, for example, the Lorenzen (1996) or Siler (1979) curves. The Siler formulation allows for senescence but has more parameters than the Lorenzen formulation and has consequently only used rarely in practice (e.g., Punt et al., 2014a).

It would be ideal to model time-varying *M* by estimating it from the monitoring data within an assessment, using for example an assessment model formulated as a state-space model (e.g. Berg and Nielsen, 2014; Nielsen and Berg, 2016; Stock and Miller, 2021) instead of prespecifying it from trends in consumption (e.g., Dorn and Barnes, 2022), indices of starvation (e.g., Regular et al., 2022) or from data on disease prevalence (e.g., Trochta et al., 2022) as these indices only reflect a single source of natural mortality. Information from these sources could, however, be included in an assessment in the form of data on natural mortality, sensu the approach used in Stock Synthesis for indices of recruitment or to justify the need for time-varying natural mortality. However, given that natural mortality is often confounded with other parameters, the estimates of *M* estimated internally within the assessment may be unreliable (or even implausible), supporting, in that situation, the use of externally derived values for natural mortality.

Modelling *M* that varies over time is typically done as either a random walk or as discrete changes (see Jiao et al. (2012) for an age-structured statistical catch-at-age model that allows for a variety of formulations for age- and time-variation in natural mortality). The random effects state-space models SAM and WHAM allow *M* to change over time and age. These methods estimate the extent to which *M* varies over time and age, such that if the data do not support time- and age-varying *M*, the extent of variation in *M* is estimated to be near zero. Use of a state-space formulation for the population dynamics model and estimating time- and age-varying *M* is therefore best practice.

In principle, time- (and age-) varying *M* can be estimated using multispecies models such as multispecies Virtual Population Analysis (Magnusson, 1995) or multispecies statistical catch-at-age analysis (e.g., Jurado-Molina et al., 2005, 2006; Van Kirk et al., 2010, 2015; Holsman et al., 2016). However, multispecies models do not estimate the effect of disease and starvation, and the results of multispecies assessment models are often not robust to model structure. The relative sizes of *M*-at-age from multispecies models can, however, be used in single-species models, e.g., as the basis for priors.

3.4. Recruitment

How recruitment is modelled in a stock assessment impacts not only the estimates of the population dynamics and hence biomass, but also reference points and measures of sustainable yield. Thus, stock assessments based on age- and size-structured population dynamics models must estimate annual recruitments as well as the relationship between spawning biomass[5] and subsequent recruitment. However, the relationship between spawning biomass and recruitment is often difficult to detect, particularly when the range of spawning biomasses for which reliable estimates of recruitment are available does not include points at both low and high stock size. Meta-analyses (e.g., Szuwalski et al., 2015) have found that the data for few stocks strongly support the effect of spawning biomass on recruitment and that many stocks are better characterized as being driven by the environment and display "regime shifts" in which the mean expected recruitment changes significantly over time.

Recruitment is variously defined as the number of animals entering the fishable population, the number of age-0 animals and the number of animals entering the first size-class in a size-structured population dynamics model. In the first two of the cases, care need to be taken to correctly account for the lag between spawning and recruitment when defining the stock-recruitment relationship. In the third case, stock-recruitment relationships are seldom modelled because the numbers entering the first size-class will often consist of the animals from spawnings in multiple years. Although slightly more computationally intensive, it is best to model the population from age-0 to reduce errors associated with mis-specifying the lag between spawning and offspring entering the population.

Recruitment during year y, R_y, is modelled using the generic form:

$$R_y = f(S_{y-L})e^{-\varepsilon_{y-L}-b_{y-L}\sigma_R^2/2} \quad \varepsilon_y \sim N(0;\sigma_R^2) \tag{2}$$

where $f()$, is the stock-recruitment relationship, L is the time-lag between spawning and when recruits enter the model, S_y is the spawning biomass during year y, ε_y is the recruitment residual for year y, σ_R is the extent of variation in recruitment about the stock-recruitment relationship, and b_y is a factor to ensure that expectation of recruitment at a given level of spawning biomass equals the value of the deterministic component of Eq. 2 (Methot and Taylor, 2011). The value for b_y can be particularly consequential for years for which recruitment deviations are estimated but there is little data on the associated deviations. It should be set to 1 for all years if the stock assessment is based on a state-space model (frequentist or Bayesian).

The form of the relationship between spawning biomass and recruitment is conventionally assumed to be one of the Beverton-Holt, Ricker, or hockey-stick relationships, but the assumption that recruitment is independent of spawning biomass is also common, particularly for size-structured stock assessments. The conventional stock-recruitment formulation can be extended with a third parameter (e.g., Deriso, 1980; Shepherd, 1982; Punt and Cope, 2019; Liermann and Hilborn, 1997), including to allow for depensation at low stock size and to ensure than B_{MSY} (the biomass associated with Maximum Sustainable Yield, MSY) is achieved at a pre-specified fraction of unfished biomass. Special forms of the stock-recruitment relationship exist for low-productivity species (e.g., Taylor et al., 2013), but these forms are seldom used in stock assessment and usually the estimates of biomass (but not necessarily of the relationship between sustainable yield and fishing mortality) are robust to the form of stock-recruitment function given reliable estimates of recruitment.

It is desirable to include environmental covariates into the stock-recruitment relationship. This can be achieved by extending Eq. (2) to:

$$R_y = f(S_{y-L})e^{-\varepsilon_{y-L}+g(\mathbf{X})-b_{y-L}\sigma_R^2/2} \tag{3}$$

where $\mathbf{X}$ is a set of environment covariates and g is a function (usually a linear model) linking the environmental variables to the deviations in recruitment about the stock-recruitment relationship. Disadvantages of this approach are that the covariates needed to be measured without error and should be available for all years. Eq. (3) also changes the meaning of σ_R to the extent to variation in recruitment about the stock-recruitment relationship not explained by the covariates. An alternative approach (Crone et al., 2019) is to model recruitment using Eq. (2) and add a component to the likelihood function to reflect that the covariate data provide an index of the ε_y, i.e.:

$$\varepsilon_y = g(\mathbf{X}) + \eta_y \quad \eta_y \sim N(0;\sigma_\eta^2) \tag{4}$$

where η_y measures the model and unexplained variation (e.g., due to measurement error), and σ_η determines the uncertainty about the relationship between the recruitment deviations and the environmental covariates. A challenge with this approach is how to set σ_η, which determines how closely the recruitment deviations match the expectations from the covariate data, although approaches such as cross-validation may be useful here. Eq. (2) can be extended by allow the autocorrelation among the recruitment deviations to be estimated and Johnson et al. (2016) compare alternative methods for estimating the extent of auto-correlation in recruitment. Auto-correlation in recruitment can be important for projections.

The parameters of the stock-recruitment relationship can be characterized by a parameter that determines the scale of the population, usually R_0, the recruitment at unfished equilibrium, and at least one parameter that determines the shape of the stock-recruitment relationship. For the Ricker and Beverton-Holt stock-recruitment relationships, this parameter is often "steepness" (the proportion of unfished recruitment when the population is reduced to 20% of its unfished level). The parameters of the stock-recruitment relationship are usually very difficult to estimate (Lee et al., 2012; Fig. S.6), leading to the use of priors for steepness (e.g., Dorn, 2002; Thorson et al., 2019), or more generally the parameter that determines the slope of the stock-recruitment relationship at the origin. In principle, priors can be included in an assessment and the parameter then estimated, but it is also common to set the parameter determining the shape of the stock-recruitment relationship to the mean (or median) of its prior.

In principle, the parameters of Eq. (2) could change over time, including as a function of covariates or due to an abrupt change, but while some examples exist (e.g., Berger, 2019; Wayte, 2013), this is unusual (but can be consequential in terms of stock status and estimates of sustainable yield). If allowance is made for time-variation in the parameters of the stock-recruitment relationship, a parameterization based on the slope at the origin and the density-dependence parameter is preferable.

In principle, the stock-recruitment relationship can be estimated outside of the assessment by fitting the population dynamics model ignoring the stock-recruitment relationship and taking the resulting estimates of recruitment and spawning biomass and fitting a stock-recruitment relationship. However, this is not considered best practice as it leads to inconsistent estimates / assumptions between how the recruitments are estimated and the resulting stock-recruitment relationship.

4. Diagnostics and data weighting

4.1. Diagnostics

Diagnostics are applied as part of the stock assessment process to select a "best" model, select a set of models to include in an ensemble, or to be part of a scheme to weight a set of models that might be used in an ensemble. An additional aim of diagnostics is to attempt to identify those aspects of a model that may be mis-specified. Maunder et al., In press reviewed (a) which diagnostics are able to identify model mis-specification, (b) can diagnostics identify what aspect of a model is mis-specified, and (c) how can an identified mis-specification be

[5] Good practice is to use the best measure of reproductive output such as a measure of fertilized egg production rather than spawning biomass.

addressed. Overall, a model would be considered adequate for providing management advice if the optimization was successful, the model fits the data adequately (e.g., based on residual analysis), the model provides reliable estimates of trends and scale, the results of the model are consistent when updated with new data (e.g., retrospective analysis), and the model is able to make adequate future predictions (e.g., hindcasting) (Carvalho et al., 2021).

It is generally best practice to apply a range of diagnostics. The diagnostics used most commonly are:

- *Convergence diagnostics*. For stock assessments based on maximum likelihood or penalized maximum likelihood, these include checking that the final gradient of the model is small, that the Hessian matrix can be inverted, and that jitter analyses (generally) lead to the same solution. For Bayesian analysis, these include checking that the method used to sample from the posterior (usually a Markov chain Monte Carlo [MCMC] algorithm) shows no evidence for non-convergence.
- *Residual diagnostics*. Standard residual diagnostics examine whether the assumed standard deviations and effective sample sizes match the sizes of the residuals and that there is no evidence for "patterns" in the residuals (within a data set and between data sets) based on the "runs" test or generalizations thereof. Probability integrated transformed (PIT) residuals likely perform better than traditional Pearson residuals but at present few assessment packages compute these residuals (Maunder et al., In press). Similarly, one-step ahead residuals are most appropriate for correlated observations (Trijoulet et al., 2023). Examination of residuals within state-space models remains a research area while posterior predicted distributions from Bayesian analysis provide a basis for assessing model fit based on the ability to mimic the data used for model fitting.
- *Retrospective analysis*. Retrospective analysis allows a comparison of the consistency of model outputs (e.g., spawning biomass, recruitment and fishing mortality and model outputs such as MSY) as additional data are added to an assessment. The results of a retrospective analysis are often summarized using Mohn's rho (Mohn, 1999). Hurtado Ferro et al. (2015) provide guidelines for what constitutes a 'major' retrospective pattern. Several studies have explored factors that can cause a retrospective pattern in an assessment (e.g., Legault, 2009). The Rose approach (Legault, 2020) currently provides the most comprehensive way to examine possible causes for retrospective patterns and provide an integrated result based on models that address the retrospective pattern.
- *Hindcast cross-validation*. Kell et al., (2016, 2021) introduced a diagnostic based on evaluating prediction skill, defined as the ability to predict an observed quantity (index values, and summary statistics for age-compositions, length-compositions, and tagging data) using an assessment that has had some of the data removed. The results of this diagnostic are summarized using the MASE (mean absolute scaled error) statistic, with a value of 1 taken as threshold between the assessment having some predictive skill (at least for the observed data) versus effectively no predictive skill.
- *Likelihood profiling*. The likelihood component profile provides a way to identify the influence of information sources on model estimates, and a difference in the best estimates of a parameter (or derived quantity) between information sources is suggestive of data conflicts (e.g., Ichinokawa et al., 2014). The R_0 profile is most common (although profiles for natural mortality, stock-recruitment steepness and current biomass are also conducted), and is used to identify conflicting information in the data about absolute abundance. However, the performance of this diagnostic was found to be poor in the simulation study conducted by Carvalho et al. (2017).
- *Other diagnostics*. Some diagnostics (e.g., the Age-structured Production Model, ASPM, diagnostic; Maunder and Piner, 2015; Minte-Vera et al., 2017; the catch curve diagnostic; Carvalho et al., 2017) have been developed specifically for fisheries assessments. The ASPM diagnostic was developed to assess whether surplus production and observed catches alone could explain the trend in the index of abundance and hence whether the data (i.e., the indices of abundance) provide information on the scale of the population. The catch curve diagnostic was also developed to assess whether the composition data are consistent with the index data, but simulations by Carvalho et al. (2017) concluded that it performed poorly (high level of Type I error).

Overall, the ideal is to apply as many diagnostic analyses as possible, along with running sensitivity analyses to explore sensitivity even within a model that exhibits no obvious problems, recognizing that currently available diagnostics are not guaranteed to identify all problems or uncertainties. Carvalho et al. (2017) found that applying multiple diagnostics was likely to identify more problems, without a major increase in 'Type I error', i.e., incorrect reject of a correctly specified model. Few assessments apply all of the above diagnostics and the minimum set would seem to be to evaluate convergence and model fit (as summarized using residuals) and to conduct a retrospective analysis and construct likelihood profiles. The hindcast and the ASPM diagnostics can be used to better understand the "value" of the assessment (for example, is it any better than a simple AR-1 process) and its properties. Weighting of alternative model configurations using diagnostics remains a research area unfortunately.

4.2. Data weighting

Data weighting is important in contemporary integrated analysis-based assessments as these assessments use multiple data types, and conflicts among data types can have substantial impacts on management advice, quantification of uncertainty, and model selection (Maunder et al., 2017 and references therein). Maunder and Piner (2017) state that the appropriate method to deal with data conflicts depends on whether it is caused by random sampling error, process variation, observation model misspecification, or misspecification of the population dynamics model. It is good practice to routinely examine the sensitivity of the model results to data weights and this is common when assessments are conducted, but the 'base' levels of weights will determine the set of sensitivity tests conducted and consequently need to be set based on objective criteria. Francis (2011) advises that data weights be selected so that indices of abundance are able to be fitted well and that correlations are allowed for when weighting composition data thus giving "preference" to the indices over the compositions when there is conflict. However, there is no general consensus on this, although it is philosophically consistent with the ASPM diagnostic.

The weight assigned to each data source (and prior / penalty) depends on a variance parameter (usually a standard error for indices and catches, and an effective sample size for the composition and tagging data). The data weights should reflect the sampling error associated with the data (given the way the data are included in the assessment). For example, composition data are treated as independent samples from the catch/population whereas actual samples for age- and size-composition are usually based on some form of hierarchical sampling process (e.g., Francis, 2017), leading to perhaps considerable overdispersion and hence actual sample sizes grossly over-estimating effective sample sizes. The data weights will also (unintentionally perhaps) capture some of the model error (e.g., time-variation in selectivity will be reflected as lower estimated effective sample sizes – the so-called "downweighting' approach to dealing with model mis-specification).

The approaches to data weighting can be divided into (a) choice of the probability distributions for the data (traditionally log-normal for indices, multinomial for composition data, and Poisson or negative binomial for tagging data), and (b) setting of the variance parameters. Two approaches to selecting data weights have emerged: (a) 'tuning' of variance parameters, and (b) estimating variance parameters. Estimation of variance parameters (e.g., by adopting a Dirichlet or Dirichlet-multinomial distribution for composition data; Francis, 2017; Thorson et al., 2017; estimating residual variances: Nielsen and Berg, 2014) is intuitively preferable, including because the uncertainty owing to data weighting is captured within the measures of uncertainty from the assessment. However, this approach is not without its problems. For example, conflicting data can lead to some data sources being substantially upweighted and others assigned near-zero weights. Estimation of data weights is more natural within the Bayesian and state-space formulations for assessments given that they are better placed to estimate variance parameters.

Tuning methods have been applied to composition data. Until recently, the weights assigned to composition data (under the assumption of multinomial sampling) was that of McAllister and Ianelli (1997). However, that method ignores correlations in residuals among age-/size-classes and can hence over-weights composition data. Francis (2011) introduced methods that weight age- and size-composition data that allow for correlations among residuals within years (but not among years and data sets, which can occur: Thorson et al., 2017) and the use of this weighting scheme (and that developed by Punt (2017)) for conditional age-at-length data) is good practice when data weights are based on tuning. Most tuning algorithms are based on multiplying input sample sizes by a tuned constant. However, this implies that the relative pattern of effective sample sizes will be robust to the tuning process. Francis (2017) suggests that the additional variance estimated for composition data likelihoods might need to be additive rather than multiplicative (i.e., weighted by the relative sample size) to represent process variation, which may be relatively constant over time and not proportional to the sample size. However, to date few assessments are use additive weights for composition data (unlike index data where this is standard).

4.2.1. *Tentative best practice for data weighting*

The set of models considered for data weighting should all satisfy the criterion that there is no evidence for model mis-specification based on diagnostics (see Section 4.1). Then initial weights need to be set, for example, by setting the initial standard error for the indices by fitting smooth functions (e.g., splines or a loess smoother) to them and setting the initial standard errors to the resulting residual standard errors. The initial effective sample sizes for age- and size-composition data should be based on methods such as that of Stewart and Hamel (2014), which account for both the number of animals aged/measured and the number of sampling units (hauls, trips, vessels, etc). Good practice remains use of tuning methods and those outlined above, along with that of Punt et al. (2017) for tagging data, which should be applied along with re-estimation of the index residual standard errors until convergence. Care needs to be taken when applying methods such as those of Francis (2011) when the data set contains data for only a few years. Table 2 lists an approach used in Australia for setting the variance parameters in Stock Synthesis-based assessments.

Table 2
Example of how variance parameters are set for stock assessments conducted for groundfishes in southeast Australia. The algorithm is predicated on having selected the model structure including the fleets and having appropriately standardized the catch and effort and composition data for inclusion in the assessment.

Step	Description
1	Remove all compositions with “extremely small” (e.g. < 100) sample sizes to avoid plots being dominated by results for which fits should not be expected to be good.
2	Set the initial weights for survey indices to their sampling CVs and for CPUE indices to the residual standard error about a loess smoother to the indices.
3	Select the basis for initial weights for the compositions (e.g., trip numbers or hauls or the formula developed for the US West Coast by Stewart and Hamel, 2014).
4	Run the assessment and compute the weights for the composition data based on Francis (2011)
5	Adjust the effective sample sizes for the composition data based on multiplying the current effective sample sizes by the multipliers from step 4.
6	Run and the assessment and compute the weights for the conditional age-at-length data based on the algorithm in Punt (2017).
7	Adjust the effective sample sizes for the conditional age-at-length data based on multiplying the current effective sample sizes by the multipliers from steps 4 & 6.
8	Repeat steps 4–7
9	Adjust the bias ramp and update the value of σ_R using the approach of Methot and Taylor (2011).
10	Update the standard errors for the survey and CPUE index data so the assumed standard errors or coefficients of variation match the variances in the residuals about the fit to the model.
11	Repeat step 9
12	Repeats steps 4–7

5. Discussion and a tentative recipe

The ultimate aim of a fisheries stock assessment is to support management decision making. This can take the form of providing estimates of biomass in absolute terms or relative to management reference points, which can then be synthesized, for example, in the RAM Legacy database (Ricard et al., 2012). The results of syntheses and meta-analyses can be used to better understand the status of world fisheries, the factors that are more likely to lead to achievement of management objectives (e.g., Hilborn et al., 2020), and to examine questions such as whether growth varies over time (Stawitz et al., 2015) and what are the primary drivers of recruitment (e.g., Szuwalski et al., 2015). The results of stock assessments also support application of harvest control rules that determine limits on catches or fishing effort for specific fisheries and to parameterize the operating models that are used to compare the performances of alternative harvest strategies.

The major challenges related to conducting stock assessments are that the data available for assessment purposes are seldom collected in an ideal manner. For example, data are rarely available from the start of the fishery, there are rarely fishery-independent data, and sampling for fishery-dependent data relies on the operations of the fishery and may not be representative of the populations being assessed. Moreover, the nature of fisheries and their operations means that the trend of the population may be uninformative about key population dynamics processes (e.g., one-way trips in biomass and fishing mortality are less informative than biomass and fishing mortality trajectories that exhibit contrast; Hilborn, 1979; Magnusson and Hilborn, 2007). These challenges are inherent to the populations being assessed and cannot be removed through analysis – but their consequences should be documented in assessments.

Another set of challenges relates to the time, software, and computational constraints confronted by analysts conducting assessments. While the availability of stock assessment packages (see Dichmont et al., 2021 for a recent summary) reduces these challenges, it remains almost impossible, for example, to conduct a Bayesian analysis using the MCMC algorithm during an assessment review meeting (although work is being understand the speed up runtimes for Bayesian analyses, e.g. Monnahan et al., 2017, 2019), and it is not uncommon for existing stock assessment packages not to include all the features that seem pertinent for a case study. Two key consequences of this are that this paper has focused on the expectation that *good practices* and not necessarily *best practices* are the current gold standard in the field, and that regular updates to assessments and good/best practices are necessary.

5.1. *Data-rich catch-at-age and -at-length assessments – what about other assessments*

This paper has focused on good (and best) practices for single-species stock assessments for data-rich stocks. The question arises whether the same practices can be applied to data-limited species and multispecies / ecosystem models. My opinion (and that of the authors of the paper on next generation stock assessments, Punt et al., 2020) is that assessments for data-limited species should be based on the same approaches as those for data-rich stocks, recognizing that the range of uncertainty will be greater. Many stock assessments can be applied in a data-limited context using models and estimation methods applied in data-rich cases (e.g., Simple Stock Synthesis, Cope, 2013, requires only catch data, while SS_CL, Rudd et al., 2021 and LIME, Rudd and Thorson (2018) descrive stock assessment packages that can be applied using only length-composition data) and that is best practice. The use of methods that can be shown to be biased, if not statistically incorrect (e.g., methods based solely on catch data have been shown not be able to estimate stock status relative to reference points adequately; Free et al., 2020; Ovando et al., 2022) should be avoided even if they seem to be very easy to apply.

Multispecies and ecosystem models are not used for tactical management, although applications of multispecies stock assessment methods such as CEATTLE (Holsman et al., 2016; Adams et al., 2022) could be based on most of the good practices outlined in this paper. Care would need to be taken regarding how any predation functions were chosen and parameterized given that the results of multispecies models can differ markedly depending on how predation is modelled (Kinzey and Punt, 2009; Plagányi, 2007). Ecosystem models (e.g., Ecosim, Atlantis) are currently only used as the operating models for the application of MSE and are seldom fitted to data (however, this is becoming increasingly common for Ecosim models; e.g. Heymans et al., 2016; Scott et al., 2016). Nevertheless, these models are increasingly being reviewed using the same processes as the single-species assessments used for status determination, and guidelines are starting to emerge (e.g., Kaplan and Marshall, 2016). Good (and best) practices for these types of models should ultimately be based on the types of issues, and the ways they are best addressed, as outlined above.

5.2. Inside vs outside

In general, it is undesirable to estimate parameters outside of the fitting of the assessment model to the data, e.g., fitting a growth curve to length-at-age data and then assuming that the resulting growth curve is 'known'. This is because fixing parameters will lead to under-estimation of uncertainty as reflected in standard errors and confidence intervals from a single model run, thereby highlighting the value of (and need for) an ensemble approach. In addition, it is not uncommon when estimating parameters outside of an assessment to make assumptions that are not consistent with those in the assessment such as assuming that there is no length selection when fitting a growth curve and then estimating length selection in the assessment model. The most common parameters that can be estimated outside of the assessment are the weight-length relationship and the relationship between age/length and being mature because the data to determine these relationships are rarely included in the model likelihood.

Weight-at-age can be estimated outside of the assessment when the sampling program is such that weight-at-age can be well estimated and weight-at-age changes in a complex way that would be hard to quantify using a parametric model. Nevertheless, while growth can be estimated outside an assessment, it is preferrable to estimate it within the assessment because this enables consistency of assumptions between how the growth function is estimated and the population dynamics are modelled and to ensure that uncertainty when estimating growth is propagated to all of the management-related quantities (Maunder et al., 2016).

In general, the rule to apply when deciding whether to estimate a parameter within an assessment is whether the data in the assessment inform the estimate of that parameter, something that can be explored using likelihood profiles, and whether some of the parameters are almost completely confounded given the available information, as indicated by large values for off-diagonal elements of the correlation matrix for the parameters.

5.3. Modelling time-varying parameters

Most population processes in a stock assessment vary over time. However, it is desirable to assume that parameters are time-invariant in assessments based on the integrated approach because this reduces the number of estimable parameters. Moreover, it is not the case that assessments should (or can) estimate time-variation in all processes. Good practice would be to estimate deviations in recruitment about the stock-recruitment relationship, deviations in growth parameters over time, and deviations in selectivity about expected selectivity. It is less common to estimate time-variation in other parameters (such as natural mortality, the parameters of stock-recruitment relationship, and movement). However, there are an increasing number of cases where natural mortality is allowed to be time-varying (e.g., Berg and Nielsen, 2014) and changes over time in R_0 are modelled in a small number of assessments (e.g., Wayte, 2013).

Traditionally, selectivity was assumed to be time-invariant (by fleet) within assessments based on the integrated approach, and this remains the default for most assessments. However, selectivity is a function of fishing and biological processes. Consequently, it is unlikely to be homogeneous over space and time at the population level, and hence the constant selectivity assumption implied by selectivity in Eq. 1 not depending on y as well as a is often violated (Martell and Stewart, 2014). This suggests that time-varying selectivity should be assumed for most fisheries (Maunder et al., 2014). However, estimating selectivity changes over time can be difficult, and relies on having good data and a clear understanding of the fishery characteristics. Ignoring temporal changes in selectivity can produce biased estimates of management quantities (Punt et al., 2014b) and underestimate uncertainty. Maunder et al. (2014) and Punt et al. (2014b) recommended multiple diagnostics (based on residuals, profiles, retrospective patterns, values for information criteria, etc) be examined before increasing model complexity by allowing for time-varying selectivity. The issues of over-parameterization can be overcome by fitting the stock assessment as a space-space model (c.f., Nielsen and Berg, 2014; Stock and Miller, 2021). Privitera-Johnson et al. (2022) found that estimation of time-variation in selectivity did not lead to appreciable improvements in performance when the true time-variation was random without trend.

Care should be taken when estimating time-variation in parameters to remove retrospective patterns because adding a source of process error (e.g., for M, selectivity, growth) can 'resolve' apparent model-misspecification evident from retrospective analyses, even when the cause of the retrospective pattern is not unmodelled trends in the process assumed to varying over time (Szuwalski et al., 2018).

5.4. Assessments, MSE and projections

It is not the aim of this paper to provide good practices for constructing the operating models used for MSE or for calculating reference points and conducting projections. However, the process of fitting operating models to data ('conditioning') has many of the features of conducting a stock assessment, except that the aim is to find a set of model configurations that are able to mimic the available data and to 'capture the full range of uncertainty' (Punt et al., 2016). As such, unlike many assessments that aim to identify a set of parsimonious models on which to base management advice, the aim for MSE is to identify a broad range of model configurations.

5.5. A tentative good practice schema for single-species stock assessment

Conducting a stock assessment is necessarily a multi-step and often iterative exercise (see Table 3 for some specific suggestions, and Table 4 for some comparisons between good and best practices).

The first step is to evaluate what is known about the stock to be assessed within the region in which it is located as well as for other regions (for similar populations, stocks and species), with the intent to

Table 3
Tentative good practices for conducting assessments.

Basic formulation
1. Determine the stock structure hypotheses and whether there is a need to consider spatial variation in population structure.
2. Identify the maximum possible number of fleets and surveys (based on spatial, temporal, and gear considerations).
3. Apply regression tree and other methods to the composition and index data to identify the set of fleets and areas to consider in the model.

Constructing data inputs
4. Apply best practices guidelines to construct the index and composition. The use of spatio-temporal models is preferred, especially for surveys but also for fishery CPUE (Thorson et al., 2020).
5. Compute the initial weights for the data (and priors).

Spatial models, time-steps, etc
6. Base the assessment on an age- or size-structure population dynamics model with two sexes.
7. Decide whether to conduct the assessment using a spatial model or the 'areas-as-fleets' approach based on preliminary analyses of the data and basic knowledge of the population.
8. Select the number of age- and size (length)-classes. Unless there is good evidence not to do so, more classes, particularly for the population model, are to be preferred – note that the classes for the observation model can be coarser than those for the population.
9. Select the model time-step. This can be annual unless growth is fast during the year.
10. Estimate the initial age-/size-structure as deviations about an equilibrium expectation (fished or unfished)

Population processes (basic formulation)
11. Estimate growth within the model using a general model (e.g., Schnute), unless the model is to be fitted only to weight-at-age. It may be necessary to place priors on some of the parameters (particularly if the model is to fitted only to length-composition data; i.e. no conditional age-at-length data). The extent of variation in growth should be estimated from the data as well as the parameters determining expected length-at-age.
12. Estimate natural mortality for adults and adopt a Lorenzen relationship between M and age. Place a prior on adult M based on auxiliary analyses (e.g., using the approaches of Hamel, 2015 and Hamel and Cope, 2022).
13. Fit a two parameter (usually Beverton-Holt) stock-recruitment relationship with both parameters estimated (but with a prior on the parameter that determines the slope of the relationship at the origin – usually steepness). Apply a bias-correction factor based on Methot and Taylor (2011). Allow for auto-correlation in recruitment deviations temporally and perhaps spatially.

Selectivity
14. Assume that selectivity is time-invariant for surveys and allow for time-variation for fishery selectivity. Assume at least one fishery/survey has asymptotic selectivity in at least one area and assume a double normal selection pattern for the remaining fleets/ surveys (with selectivity for some fleets perhaps constrained to be asymptotic).

Time-variation in parameters
15. Allow for time-variation in parameters if supported by residual diagnostics / to eliminate retrospective patterns.

Data weighting and model fitting
16. The resulting residual variances should match the assumed variance for the data, ideally select likelihood functions so that the data weights can be estimated (e.g., Dirichlet-multinomial distribution for composition data).
17. Fit the model using penalized likelihood but tune the variance parameters (seeTable 2).

Diagnostics
18. Apply all standard diagnostics.

Final steps
19. Conduct a qualitative and quantitative analysis of uncertainty, including an extensive sensitivity analysis, evaluation of likelihood profiles, etc.
20. Comment on the assessment in the context of the management problem, the plausibility of results, and the extent of uncertainty.
21. Document the data used for the assessment, and any recommendations related to model development and analyses for future assessments.

identify a set of hypotheses that can be either included in the assessment or documented as something for future work (e.g., whether M varies over time based on condition factor data, Bjornsson et al., 2022). The step will also review and document the management history, the needs of the management agency to which the assessment will be presented, and the recommendations arising from any previous assessments and their peer reviews. A key aim of this step is to identify stock structure hypotheses and to ensure that the stock assessment addresses the needs of the management body. The structure of an assessment will differ

Table 4
Some key differences between good and best practices for stock assessment.

Issue	Best practice	Good / adequate practice
Model resolution	Age-size model	Age- or size-model
Initial conditions	Estimated	Equilibrium but with variation in recruitment accounted for
Spatial structure	A spatial model fitted to tagging data	The 'areas as fleets' approach
Modelling process error	State-space formulation with all variance parameters estimated	Penalized likelihood with some data weights estimated but most tuned.
Natural mortality	A formulation that allows for high natural mortality at young and old ages	Age-specific M (estimated adult M)
Growth model	Time-varying Richards model	Von Bertalanffy growth curve
Stock-recruitment relationship	Tailored to stock in question (e.g.Taylor et al., 2013 for low fecundity species)	Beverton-Holt or Ricker
Time-variation in parameters (M, growth, selectivity)	Estimate for as many processes as is feasible given the information content of the data, and within the state-space formulation.	Dealt with in an ad hoc manner
Specifying priors	Multivariate priors (e.g. between M, selectivity and steepness)	Univariate priors
Retrospective patterns	Use the Rose / an ensemble approach	Try various plausible time-variation in parameters to resolve the problem / downweight some data sources.
Data weighting	All variance parameters estimated	Application of tuning methods

(perhaps substantially) if the aim is to examine local depletion effects, to assess whether some environmental process impacts recruitment, growth, selectivity, or natural mortality, with a view to conducting long-term projections to examine the consequences of climate change, or to support application of a harvest control rule for the entire population. Another aim of this step would be to determine whether it is likely that some population process (e.g., growth, recruitment, natural mortality, the stock-recruitment relationship, and movement) would be time-varying.[6]

The next step is to conduct a data inventory and examine how representative the data are given when, where and how they were collected (i.e., whether the data from a fishery or survey cover the range of the stock or the area to which they are assumed to be pertain). This step should include an evaluation of whether best practices are applied to compute the input data (e.g., Maunder and Punt, 2004, for catch-effort-based indices). This step, along with the earlier step, will help to select the model time-step, and whether spatial structure needs to be included explicitly in the assessment (i.e., using a spatial model) or implicitly (using the 'areas as fleets' approach). For example, evidence for different trends among fleets that operate in different areas provides support for the development of a spatial model. This step also involves computing the initial values for the variance parameters used to weight the data and the values of the parameters of the priors for the parameters related to growth, natural mortality, selectivity and recruitment.

By default, the parameters determining (temporally and spatially invariant) population parameters should be estimated but perhaps with (very) informative priors. Subsequent analyses could lead to parameters being fixed to those estimated using auxiliary information. The deviations in recruitment about the stock-recruitment relationship should

[6] These processes are always time-varying in reality, but what matters is if the effect is substantial for the population(s) being assessed.

be estimated for all years, accounting for a bias-correction factor (Methot and Taylor, 2011) to ensure the expected recruitment matches that from the assumed stock-recruitment relationship. In general, good (and best) practice for estimating any population parameter is to be guided by the data, involve trying alternative models, limiting complexity to the extent possible (more complex models should lead to clear improvements in fit), and ensuring that the resulting model outputs for the process are plausible (Francis, 2016).

The next step is to fit a model configuration (the 'naive' configuration) based on the above specifications and apply the data weighting steps in Section 4.2. Best practice would be to fit the model as a state-space model (as a frequentist or Bayesian analysis), but good practice is use of penalized maximum likelihood. If there is strong evidence that selectivity is time-varying, this first model could involve time-varying selectivity/retention, either using time-blocks based on when 'known' changes in selection and retention occurred or as random deviations about a parametric form, with the extent of variation in selectivity estimated using the approach of Xu et. (2019). The next step is to apply 'rejection' diagnostics (i.e., convergence tests, residual analysis, and retrospective analysis) and, if needed, identify a set of alternative models that could address the aims of the assessment. The set of models would be pruned by those that still fail the diagnostics, perhaps leading to an overall summary using the 'Rose' approach of Legault (2020).

The next steps of any assessment are to explore sensitivity to changing data weights, values for pre-specified parameters, and applying the 'understanding' diagnostics (e.g., hindcast evaluation, the ASPM diagnostics and plotting the results of previous assessments with those from the current base model configuration – or set of model configurations). The final step of the assessment is to quantify (to the extent possible) and document uncertainty, for example by computing the standard errors of the model outputs conditional on the model structure, conducting retrospective analyses, and examining the sensitivity of the results to changing some of the assumptions of the model. For most assessments the set of uncertainties that can be quantified will be small fraction of the true full range of uncertainty. This is most clearly the case for assessments that aim to make medium- and long-term projections, e.g., to form the basis for rebuilding strategies. Nevertheless, quantification is uncertainty is important, especially in those jurisdictions that apply buffers to catch limits based on the extent of uncertainty.

The use of the above approach would not be guaranteed to ensure that the resulting assessment outcomes are 'correct' but rather than the assessment has been conducted using current state-of-the-art good practices, and that are suitable to support management decision making as best as can be expected. The paper has focused on model fitting and there is a need for best practices for other aspects of stock assessment, such as development of data collection schemes, methods for summarizing basic monitoring data for use in assessments, the definition and calculation of reference points, the bases for forecasting, especially given the effects of climate change, how best to include stakeholders in the model building process, and how best to conduct reviews of stock assessments.

CRediT authorship contribution statement

André E. Punt conserved the structure of the paper and wrote it.

Declaration of Competing Interest

The author declares the following financial interests/personal relationships which may be considered as potential competing interests: Andre Punt reports financial support was provided by National Oceanic and Atmospheric Administration.

Acknowledgements

This publication was partially funded by the Cooperative Institute for Climate, Ocean, & Ecosystem Studies (CICOES) under NOAA Cooperative Agreement NA15OAR4320063, Contribution No. 2023-1252 Mark Maunder (IATTC), Owen Hamel (NWFSC, NOAA,NMFS) and Steve Cardin (University of Massachusetts, Dartmouth) are thanked for their comments on an initial version of this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fishres.2023.106642.

References

Pacific Fishery Management Council (PFMC), 2020, Terms of Reference for the groundfish and coastal pelagic species stock assessment review process for 2021–2022. https://www.pcouncil.org/documents/2021/01/terms-of-reference-for-the-coastal-pelagic-species-stock-assessment-review-process-for-2021–2022-december-2020.pdf.

Pacific Fishery Management Council (PFMC), 2021, in press practices guidelines for groundfish assessments in 2021 and 2021. https://www.pcouncil.org/documents/2021/02/in press-practices-and-guidelines-for-groundfish-stock-assessments-february-2021.pdf.

Adams, G.D., Holsman, K.K., Barbeaux, S.J., Dorn, M.W., Ianelli, J.N., Spies, I., Stewart, I. J., Punt, A.E., 2022. An ensemble approach to understand predation mortality for groundfish in the Gulf of Alaska. Fish. Res. 251, 106303.

Allen Akselrud, C., Punt, A.E., Cronin-Fine, L., 2017. Exploring model structure uncertainty using a general stock assessment framework: The case of Pacific cod in the Eastern Bering Sea. Fish. Res. 193, 104–120.

Anon, 2015, Assessment Model for Alaska Description of GUI and Instructions, Available from https://github.com/NMFS-toolbox/AMAK/blob/master/docs/AMAK%20Documentation.pdf.

Begley, J., 2014, Gadget User Guide. Available at http://www.hafro.is/gadget/userguide/userguide.html.

Berg, C., Nielsen, A., 2016. Accounting for correlated observations in an age- based state-space stock assessment model. ICES J. Mar. Sci. 73, 1788–1797.

Berger, A.M., 2019. Character of temporal variability in stock productivity influences the utility of dynamic reference points. Fish. Res. 217, 185–197.

Berger, A.M., Jones, M.L., Zhao, 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.

Berger, A.M., Goethel, D.R., Lynch, P.D., Quinn, T., Mormede, S., McKenzie, J., Dunn, A., 2017. Space oddity: the mission for spatial integration. Can. J. Fish. Aquat. Sci. 74, 1698–1716.

Bjornsson, B., Solmundsson, J., Woods, P.J., 2022. Natural mortality in exploited fish stocks: annual variation estimated with data form trawl surveys. ICES J. Mar. Sci. 79, 1569–1582.

Bull, B., Dunn, A., McKenzie, A., Gilbert, D.J., Smith, M.H., Bian, R., Fu, D., 2012, CASAL (C++ algorithmic stock assessment laboratory) User Manual v2.30–2012/03/21. NIWA Technical Report 135. National Institute of Water and Atmospheric Research, Wellington, 280 p.

Butterworth, D.S., Punt, A.E., 1990. Some preliminary examinations of the potential information context of age-structure data from Antarctic minke whale research catches. Rep. Int. Whal. Commn 40, 301–315.

Cadrin, S.X., Maunder, M.N., Punt, A.E., 2020. Spatial Structure: Theory, estimation and application in stock assessment models. Fish. Res. 229, 105608.

Carvalho, F., Punt, A.E., Chang, Y.-J., Maunder, M.N., Piner, K.R., 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? Fish. Res 192, 28–40.

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., PIner, K.R., Maunder, M.N., Taylor, I., Wetzel, C.R., Doering, K., Johnson, K.F., Metyhot, R.D., 2021. A cookbook for using model diagnostics in integrated stock assessments. Fish. Res. 240, 105959.

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.

Crone, P., Maunder, M., Valero, J., McDaniel, J., Semmens, B., 2013, Selectivity: theory, estimation, and application in fishery stock assessment models Workshop Series Report 1, CAPAM Workshop Series. https://swfsc.noaa.gov/publications/CR/2013/2013Crone.pdf.

Cope, J.M., Hamel, O.S., 2022. Updating from M version 0.2: An application-based methods for practical estimation, evaluation and uncertainty characterization of natural mortality. Fish. Res. 256, 106493.

Crone, P.R., Maunder, M.N., Lee, H.H., Piner, K.R., 2019. Good practices for including environmental data to inform spawner-recruit dynamics in integrated stock assessments: small pelagic species case study. Fish. Res. 217, 122–132.

De Moor, C.L., Butterworth, D.S., De Oliveira, J.A.A., 2011. Is the management procedure approach equipped to handle short-lived pelagic species with their boom and bust dynamics? The case of the South African fishery for sardine and anchovy. ICES J. Mar. Sci. 68, 2075–2085.

Deriso, R.B., 1980. Harvesting strategies and parameter estimation for an age-structured model. Can. J. Fish. Aquat. Sci. 37, 268–282.
Deriso, R.B., Parma, A.M., 1988. Dynamics of age and size for a stochastic population model. Can. J. Fish. Aquat. Sci. 45, 1054–1068.
Dichmont, C.M., Punt, A.E., Deng, A., Dell, Q., Venables, W., 2003. Application of a weekly delay-difference model to commercial catch and effort data in Australia's Northern Prawn Fishery. Fish. Res. 65, 335–350.
Dichmont, C.M., Deng, R., Punt, A.E., Brodziak, J., Chang, Y.-J., Cope, J.M., Ianelli, J.N., Legault, C.M., Methot, R.D., Porch, C.E., Prager, M.H., Shertzer, K., 2016. A review of stock assessment packages in the United States. Fish. Res. 183, 477–460.
Dichmont, C.M., Deng, R.A., Dowling, N., Punt, A.E., 2021. Collating stock assessment packages to improve stock assessments. Fish. Res 235, 105844.
Doering, K.L., Wilberg, M.J., Liang, D., Tarnowski, M., 2021. Patterns in oyster natural mortality in Chesapeake Bay, Maryland using a Bayesian model. Fish. Res 236, 105838.
Dorn, M.W., 2002. Advice on West Coast rockfish harvest rates from Bayesian meta-analysis of stock- recruit relationships. North Am. J. Fish. Manag. 22, 280–300.
Dorn, M.W., Barnes, C.L., 2022. Time-varying predation as a modifier of constant natural mortality for Gulf of Alaska walleye pollock. Fish. Res 254, 106391.
Dunn, A., Rasmussen, S., Mormede, S., 2015, Spatial Population Model User Manual. National Institute of Water & Atmospheric Research Ltd. NIWA Technical Report 138. Science Communication, NIWA, Private Bag 14901, Wellington, New Zealand, pp. 208. ftp://ftp.niwa.co.nz/software/casal/SPM.pdf.
Fournier, D., Archibald, C.P., 1982. A general theory for analysing 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., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138.
Francis, R.I.C.C., 2016. Growth in age-structured stock assessment models. Fish. Res. 180, 77–86.
Francis, R.I.C.C., 2017. Revisiting data weighting in fisheries stock assessment models. Fish. Res. 192, 5–15.
Free, C.M., Jensen, O.P., Anderson, S.C., Gutierrez, N.L., Kelisner, K.M., Long, C., Minto, C., Chato Oslo, G., Walsh, J.C., 2020. Blood from a stone: Performance of catch-only methods in estimating stock biomass status. Fish. Res. 223, 105452.
Gilbert, D.J., Davies, N.M., McKenzie, J.R., 2006. Development of an age-length structured model of the Hauraki Gulf-Bay of Plenty snapper (*Pagrus auratus*) population. Mar. Freshw. Res. 57, 553–568.
Goethel, D.R., Berger, A.M., 2017. Accounting for spatial complexities in the calculation of biological reference points: effects of misdiagnosing population structure for stock status indicators. Can. J. Fish. Aquat. Sci. 74, 1878–1894.
Goethel, D.R., Bosley, K.M., Hanselman, D.H., Berger, A.M., Deroba, J.J., Langseth, B.J., Schueller, A.M., 2019. Exploring the utility of different tag-recovery experimental designs for use in spatially explicit, tag-integrated stock assessment models. Fish. Res. 219, 105320.
Gulland, J.A., 1969. Manual of Methods for Fish Stock Assessment. Part I. Fish Population Analysis. Man. Fish. Sci. 4. FAO,, p. 154.
Gulland, J.A., 1983. Fish Stock Assessment. A Manual of Basic Methods. John Wiley and Sons,, New York, p. 223.
Gulland, J.A., 1988. Fish Population Dynamics: The Implications for Management. John Wiley and Sons,, New York, p. 422.
Haddon, M., 2011. Modelling and Quantitative Methods in Fisheries. 2nd Ed. CRC Press, Taylor & Francis Group,, Boca Raton, p. 465.
Haddon, M., 2017, Orange roughy east (Hoplostethus atlanticus) stock assessment using data to 2016, 760 – 805, in Tuck, G.N. (ed.) (2018). Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2016 and 2017. Part 2, 2017. Australian Fisheries Management Authority and CSIRO Oceans and Atmosphere, Hobart. 837p.
Hamel, O.S., 2015. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. ICES J. Mar. Sci. 72, 62–69.
Hamel, O.S., Cope, J.M., 2022. Development and considerations for application of a longevity-based prior for the natural mortality rate. Fish. Res 256, 106477.
Hamel, O.S., Ianelli, J.N., Maunder, M.N., Punt, A.E., 2023. Natural Mortality: Theory, estimation and application in fishery stock assessment models. Fish Res 261, 106638.
Heymans, J.J., Coll, M., Link, J.S., Mackinson, S., Steenbeek, Walters, C., Christensen, V., 2016. Best practice in Ecopath with Ecosim food-web models for ecosystem-based management. Ecol. Model. 331, 173–184.
Hilborn, R., 1979. Comparison of fisheries control systems that utilize catch and effort data. J. Fish. Res. Board Can. 36, 1477–1489.
Hilborn, R., 1990. Determination of fish movement patterns from tag recoveries using maximum likelihood estimators. Can. J. Fish. Aquat. Sci. 47, 635–643.
Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment: Choice Dynamics and Uncertainty. Chapman and Hall., Inc,, p. 570.
Hilborn, R., Amoroso, R.O., Anderson, C.M., Baum, J.K., Branch, T.A., Costello, C., de Moor, C.L., Faraj, A., Hively, D., Jensen, O.P., Kurota, H., Little, L.R., Mace, P., McClanahan, T., Melnychuk, M.C., Minto, C., Osio, G.C., Parma, A.M., Pons, M., Segurado, S., Szuwalski, C.S., Wilson, J.R., Ye, Y., 2020, Effective fisheries management instrumental in improving fish stock status. Proc. Nat. Acad. Sci. 117, 2218–2224.
Holsman, K.K., Ianelli, J., Aydin, K., Punt, A.E., Moffit, E.A., 2016. Comparative biological reference points estimated from temperature-specific multispecies and single species stock assessment models. Deep Sea Res. II 134, 360–378.
Hoyle, S.D., Maunder, M.N., A'mar, Z.T., 2020. Frameworks for the next generation of general stock assessment models: 2019 CAPAM workshop report. N. Z. Fish. Assess. Rep. 2020/3 9. ⟨http://www.capamresearch.org/sites/default/files/FAR-2020-39-CAPAM-Next-Generation-Stock-Assessment_Model-4115.pdf⟩.
Hurtado Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R.R., McGillaird, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-pre, K.A., Whitten, A.R., Punt, A.E., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES J. Mar. Sci. 72, 99–110.
Hurtado-Ferro, F., Punt, A.E., Hill, K.T., 2014. Use of multiple selectivity patterns as a proxy for spatial structure. Fish. Res. 158, 102–115.
Ichinokawa, M., Okamura, H., Takeuchi, Y., 2014. Data conflict caused by model mis-specification of selectivity in an integrated stock assessment model and its potential effects on stock status estimation. Fish. Res. 158, 147–157.
Jardim, E., Millar, C.P., Mosqueira, I., Scott, F., Chato Osio, G., Ferretti, M., Alzorriz, N., Orio, A., 2015. What if stock assessment is as simple as a linear model? The a4a initiative. ICES J. Mar. Sci. 72, 232–236.
Jiao, Y., Smith, E.P., O'Reilly, R., Orth, D.J., 2012. Modelling non-stationary natural mortality in catch-at-age models. ICES J. Mar. Sci. 69, 105–118.
Johnson, K., Councill, E., Thorson, J.T., Brooke, E., Methot, R.D., Punt, A.E., 2016. Can autocorrelated recruitment be estimated using integrated assessment models and how does it affect population forecasts? Fish. Res. 183, 222–232.
Jurado-Molina, J., Livingston, P.A., Ianelli, J., 2005. Incorporating predation interactions to a statistical catch-at-age model for a predator–prey system in the eastern Bering Sea. Can. J. Fish. Aquat. Sci. 62, 1865–1873.
Jurado-Molina, J., Gatica, C., Cubillos, L.A., 2006. Incorporating cannibalism into an age-structured model for the Chilean hake. Fish. Res 82, 30–40.
Kaplan, I.C., Marshall, K.N., 2016. A guinea pig's tale: learning to review end-to-end marine ecosystem models for management applications. ICES J. Mar. Sci. 73, 1715–1724.
Kell, L.T., Kimoto, A., Kitakado, T., 2016. Evaluation of the prediction skill of stock assessment using hindcasting. Fish. Res. 183, 119–127.
Kell, L.T., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., Fu, D., 2021. Validation of stock assessment methods: Is it me or my model talking? ICES J. Mar. Sci. 78, 2244–2255.
Kenchington, T.J., 2014. Natural mortality estimators for information-limited fisheries. Fish Fish 15, 533–562.
Kinzey, D., Punt, A.E., 2009. Multispecies and single-species age-structured models of fish population dynamics: Comparing parameter estimates. Nat. Res. Mod. 22, 67–104.
Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M., 2016. TMB: automatic differentiation and Laplace. Approx. J. Stat. Softw. 70, 1–21.
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.
Lee, H.-H., Piner, K.R., Methot, R.D., Maunder, M.N., 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. Fish. Res. 158, 138–146.
Legault, C.M., 2009, Report of the Retrospective Working Group, 14–16 January 2008, Woods Hole, Mass. US Department of Commerce Northeast Fisheries Science Center Reference Document 09–01.
Legault, C.M., 2020. Rose vs. Rho: a comparison of two approaches to address retrospective patterns in stock assessments. ICES J. Mar. Sci. 77, 3016–3030.
Legault, C.M., Restrepo, V.R., 1998. A flexible forward age-structured assessment program. Collect. Pap. ICCAT Vol. Sci (49), 246–253.
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.
Liermann, M., Hilborn, R., 1997. Depensation in fish stocks: a hierarchic Bayesian meta-analysis. Can. J. Fish. Aquat. Sci. 54, 1876–1984.
Lorenzen, K., 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. J. Fish. Bio 49, 627–647.
Lorenzen, K., 2016. Toward a new paradigm for growth modeling in fisheries stock assessments: Embracing plasticity and its consequences. Fish. Res 180, 4–22.
Magnusson, A., Hilborn, R., 2007. What makes fisheries data informative? Fish Fish 8, 337–358.
Magnusson, K., 1995. An overview of the multispecies VPA—theory and applications. Rev. Fish. Biol. Fish. 5, 195–212.
Martell, S.J.D., Stewart, I.J., 2014. Toward defining good practices for modelling time-varying selectivity. Fish. Res. 158, 84–95.
Maunder, M.N., 2001, Integrated tagging and catch-at-age analysis (ITCAAN). In: Kruse, G.H., Bez, N., Booth, A., Dorn, M.W., Hills, S., Lipcius, R.N., Pelletier, D. (Eds.), Spatial Processes and Management of Fish Populations. University of Alaska, Fairbanks, pp. 123–146 Alaska Sea Grant College Program Report, AK-SG-01–02.
Maunder, M.N., Deriso, R.B., 2003. Estimation of recruitment in catch-at-age models. Can. J. Fish. Aquat. Sci. 60, 1204–1216.
Maunder, M.N., Piner, K.R., 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. 72, 7–18.
Maunder, M.N., Piner, K.R., 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fish. Res. 192, 16–27.

Maunder, M.N., Punt, A.E., 2004. Standardization of catch and effort data: A review of recent approaches. Fish. Res. 70, 141–159.
Maunder, M.N., Punt, A.E., Carvalho, F., Winker, H., Valero, J., Minte-Vera, C.V., Xu, H., In press, 1St workshop in improving the risk analysis for tropical tunas in the Eastern Pacific Ocean: Model diagnostics in integrated stock assessments. IATTC Document WSRSK-01.
Maunder, M.N., Crone, P.R., Valero, J.L., Semmens, B.X., 2014. Selectivity: theory, estimation, and application in fishery stock assessment models. Fish. Res. 158, 1–4.
Maunder, M.N., Crone, P.R., Punt, A.E., Valero, J.L., Semmens, B.X., 2016. Growth: theory, estimation, and application in fishery stock assessment models. Fish. Res. 180, 1–3.
Maunder, M.N., Crone, P.R., Punt, A.E., Valero, J.L., Semmens, B.X., 2017. Data conflict and weighting, likelihood functions and process error. Fish. Res. 192, 1–4.
Maunder, M.N., Xu, H., Lennert-Cody, C.E., 2022, Developing fishery definitions for the skipjack tuna stock assessment in the EPS. IATTC Document SAC-13 INF-I.
Maunder, M.N., Hamel, O.S., Lee, H.-H., Piner, K.R., Cope, J.M., Punt, A.E., Ianelli, J.N., Castillo-Jordan, C., Kapur, M, Methot, R.D., 2023. Review of estimation methods for natural mortality and their performance. Fish. Res 257, 106489.
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.
McGilliard, C.R., Punt, A.E., Methot, R.D., Hilborn, R., 2015. Accounting for marine reserves using spatial stock assessments. Can. J. Fish. Aquat. Sci. 72, 262–280.
Methot, R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. 68, 1744–1760.
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.
Miller, T.J., Stock, B.C., 2020. The Woods Hole Assessment Model (WHAM). Version 1, 0. ⟨https://timjmiller.github.io/wham⟩.
Minte-Vera, C.V., Maunder, M.N., Casselman, J.M., Campana, S.E., 2016. Growth functions that incorporate the cost of reproduction. Fish. Res. 180, 31–44.
Minte-Vera, C.V., Maunder, M.N., Aires-da-Silva, A.M., Satoh, K., Uosaki, K., 2017. Get the biology right, or use size-composition data at your own risk. Fish. Res. 192, 114–125.
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.
Monnahan, C.C., Thorson, J.T., Branch, T.A., 2017. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. Meth. Ecol. Evol. 8, 339–348.
Monnahan, C.C., Branch, T.A., Thorson, J.T., Stewart, I.J., Szuwalski, C.S., 2019. Overcoming long Bayesian run times in integrated fisheries stock assessments. ICES J. Mar. Sci. 76, 1477–1488.
Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. 158, 96–101.
Ovando, D., Free, C.M., Jensen, O.P., Hilborn, R., 2022. A history and evaluation of catch-only stock assessment models. Fish Fish 23, 616–630.
Pauly, D. 1984, Fish population dynamics in tropical waters; a manual for use with programmable calculators. ICLARM Studies and Reviews 8. International Center for Living Aquatic Resources Management. 325pp,
Piner, K.R., Lee, H.-H., Maunder, M.N., 2016. Evaluation of using random-at-length observations and an equilibrium approximation of the population age structure in fitting the von Bertalanffy growth function. Fish. Res. 180, 128–137.
Plagányi, É.E., 2007, Models for an ecosystem approach to fisheries. FAO Tech. Rep. 477. 108pp. https://www.fao.org/3/a1149e/a1149e00.html.
Pope, J.G., Shepherd, J.G., 1985. A comparison of the performance of various methods for tuning VPAs using effort data. J. Cons. Int. Explor. Mer. 42, 129–151.
Privitera-Johnson, K., Methot, R.D., Punt, A.E., 2022. Towards best practice for specifying selectivity in age-structured integrated stock assessments. Fish. Res 149, 106247.
Punt, A.E., 2017. Some insights into data weighting in integrated stock assessments. Fish. Res 192, 52–65.
Punt, A.E., 2019. Spatial stock assessment methods: A viewpoint on current issues and assumptions. Fish. Res. 213, 132–143.
Punt, A.E., Cope, J.M., 2019, Extending integrated stock assessments models to use non-depensatory three-parameter stock-recruitment relationships. Fish. Res. 217, 46-–57.
Punt, A.E., Kennedy, R.B., Frusher, S., 1997. Estimating the size-transition matrix for Tasmanian rock lobster, Jasus edwardsii. Mar. Freshw. Res. 48, 981–992.
Punt, A.E., Smith, D.C., Thomson, R.B., Haddon, M., He, X., Lyle, J.M., 2001. Stock assessment of the blue grenadier *Macruronus novaezelandiae* resource off south-eastern Australia. Mar. Freshw. Res. 52, 701–717.
Punt, A.E., Haung, 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., Hakamada, T., Bando, T., Kitakado, T., 2014a. Assessment of Antarctic minke whales using statistical catch-at-age analysis. J. Cetacean. Res. Manag. 14, 93–116.
Punt, A.E., Hurtado-Ferro, F., Whitten, A.R., 2014b. Model selection for selectivity in fisheries stock assessment. Fish. Res. 158, 124–134.
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., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016. Management strategy evaluation: best practices. Fish Fish 17, 303–334.
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., Day, J., Fay, G., Haddon, M., Klaer, N., Little, L.R., Privitera-Johnson, K.J., Thomson, R., Smith, A.D.M., Sporcic, M., Tuck, G.N., Upston, J., Wayte, S., 2018. Retrospective investigation of assessment uncertainty for fish stocks off southeast Australia. Fish. Res. 198, 117–128.
Punt, A.E., Dunn, A., Elvarsson, B.P., Hampton, J., Hoyle, S., Maunder, M.N., Methot, R. D., Nielsen, A., 2020. Essential features of the next-gen integrated assessment: A perspective. Fish. Res 229, 105617.
Punt, A.E., Dalton, M.G., Cheng, W., Hermann, A.J., Holsman, K.K., Hurst, T.P., Ianelli, J. N., Kearney, K.A., McGilliard, C.R., Pilcher, D.J., Veron, M., 2021a. Evaluating the impact of climate and demographic variation on future prospects for fish stocks: An Application for northern rock sole in Alaska. Deep Sea Res. II 189–190, 104951.
Punt, A.E., Castillo-Jordan, C., Hamel, O.S., Cope, J.M., Maunder, M.N., Ianelli, J.N., 2021b. Consequences of error in natural mortality and its estimation in stock assessment models. Fish. Res 233, 105759.
Quinn II, T.J., Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press,, p. 542.
Quinn II, T.J., Turnbull, C.T., Fu, C, C., 1998. A length-based population model for hard-to-age invertebrate populations. *In* Fishery stock assessment models. *Edited by* F. Funk, T.J. Quinn II, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Zhang. Rep. Alsk. Sea Grant Coll. Program, Univ. Alsk. –Fairbanks, Juneau No. 98–01, 531–556.
Ralston, S., Punt, A.E., Hamel, O.S., DeVore, J., Conser, R.J., 2011. An approach to quantifying scientific uncertainty in stock assessment. Fish. Bull. 109, 217–231.
Regular, P.M., Buren, A.D., Dwyer, K.S., Cadigan, N.G., Gregory, R.S., Koen-Alonso, M., Rideout, R.M., Robertson, G.J., Robertson, M.D., Stenson, G.B., Wheeland, L.J., Zhang, F., 2022. Indexing starvation mortality to assess its role in the population regulation of Northern cod. Fish. Res 247, 106180.
Ricard, D., Minto, C., Jensen, O.P., Baum, J.K., 2012. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. Fish Fish 13, 380–398.
Rudd, M.B., Thorson, J.T., 2018. Accounting for variable recruitment and fishing mortality in length-based stock assessments for data-limited fisheries. Can. J. Fish. Aquat. Sci. 75, 1019–1035.
Rudd, M.N., Cope, J.M., Wetzel, C.R., Hastie, J.R., 2021. Catch and length models in the stock synthesis framework: expanded application to data-moderate stocks. Front. Mar. Sci. 8, 663554.
Sampson, D.B., 2014. Fishery selection and its relevance to stock assessment and fishery management. Fish. Res. 158, 5–14.
Sampson, D.B., Scott, R.D., Quinn, T., 2011. A spatial model for fishery age-selection at the population level. Can. J. Fish. Aquat. Sci. 68, 1077–1086.
Schnute, J., 1981. A versatile growth model with statistically stable parameters. Can. J. Fish. Aquat. Sci. 38, 1128-–11140.
Scott, E., Serpetti, N., Steenbeek, J., Heymans, J.J., 2016. A stepwise fitting procedure for automated fitting of ecopath with ecosim models. SoftwareX 2016, 25–30.
Sharma, R., Maunder, M.N., Babcock, E., Punt, A.E., 2019. Recruitment: Theory, estimation, and application in fishery stock assessment models. Fish. Res. 217, 1–4.
Shepherd, J.G., 1982. A family of general production curves for exploited populations. Math. Biosci. 59, 77–93.
Siler, W., 1979. A competing-risk model for animal mortality. Ecol 64, 750–757.
Stawitz, C.C., Essington, T.E., Branch, T.A., Haltuch, M.A., Hollowed, A.B., Spencer, P.D., 2015. A state-space approach for detecting growth variation and application to North Pacific groundfish. Can. J. Fish. Aquat. Sci. 72, 1316–1328.
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.
Stewart, I.J., Martell, S.J.D., 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. Fish. Res. 158, 40–49.
Stock, B.C., Miller, T.J., 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fish. Res 240, 105967.
Stockhausen, W.T., 2019, 2019 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. https://meetings.npfmc.org/CommentReview/DownloadFile?p=8ce00537-a862–4a28-ac75-a58c0eef905c.pdf&fileName=C4%203%20Tanner%20Crab%20SAFE%202019.pdf.
Szuwalski, C., 2019, A stock assessment for eastern Bering Sea snow crab. https://meetings.npfmc.org/CommentReview/DownloadFile?p=30a57738-eb23–40a7–9e82-ca415b36cdf1.pdf&fileName=C4%201%20Snow%20Crab%20SAFE%202019.pdf.
Szuwalski, C.S., Vert-Pre, K.A., Punt, A.E., Branch, T.A., Hilborn, R., 2015. Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish. Fish. 16, 633–648.
Szuwalski, C.S., Ianelli, J.N., Punt, A.E., 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. ICES J. Mar. Sci. 75, 596–609.
Taylor, I.G., Methot, R.D., 2013. Hiding or dead?: A computationally efficient model of selective fisheries mortality. Fish. Res. 142, 75–85.
Taylor, I.G., Gertseva, V., Methot, R.D., Maunder, M.N., 2013. A stock–recruitment relationship based on pre-recruit survival, illustrated with application to spiny dogfish shark. Fish. Res 142, 15–21.
Thompson, G.G., 1994. Confounding of gear selectivity and the natural mortality rate in cases where the former is a nonmonotone function of age. Can. J. Fish. Aquat. Sci. 51, 2654–2664.
Thorson, J.T., 2019. Perspective: Let's simplify stock assessment by replacing tuning algorithms with statistics. Fish. Res. 217, 133–139.
Thorson, J.T., Mine-Vera, C.V., 2016. Relative magnitude of cohort, age, and year effects on size at age of exploited marine fishes. Fish. Res 180, 45–53.
Thorson, J.T., Johnson, K.F., Methot, R.D., Taylor, I.G., 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. 192, 84–93.

Thorson, J.T., Dorn, M.W., Hamel, O.S., 2019. Steepness for West Coast rockfishes: Results from a twelve-year experiment in iterative regional meta-analysis. Fish. Res. 217, 11–20.
Thorson, J.T., Maunder, M.N., Punt, A.E., 2020. The development of spatio-temporal models of fishery catch-per-unit-effort data to derive indices of relative abundance. Fish. Res 230, 105611.
Trijoulet, V., Albertsen, C.M., Kristensen, K., Legault, C.M., Miller, T.J., Nielsen, A., 2023. Model validation for compositional data in stock assessment models: Calculating residuals with correct properties. Fish. Res. 257, 106487.
Trochta, J.T., Groner, M.L., Herschberger, P.K., Branch, T.A., 2022. A novel approach for directly incorporating disease into fish stock assessment: a case study with seroprevalence data. Can. J. Fish. Aquat. Sci. 79, 611–630.
Van Kirk, K.F., Quinn, T.J., Collie, J.S., 2010. A multispecies age-structured assessment model for the Gulf of Alaska. Can. J. Fish. Aquat. Sci. 67, 1135–1148.
Van Kirk, K.F., Quinn, T.J., Collie, J.S., A'Mar, Z.T., 2015. Assessing uncertainty in a multispecies age-structured assessment framework: The effects of data limitations and model assumptions. Nat. Res. Model. 28, 184–205.
Vincent, M.T., Brenden, T.O., Bence, J.R., 2020. Parameter estimation performance of a recovery-conditioned integrated tagging catch-at-age analysis model. Fish., Res 224, 105451.
Wang, S.P., Maunder, M.N., Piner, K.R., Aires-da-Silva, A., Lee, H.H., 2014. Evaluation of virgin recruitment profiling as a diagnostic for selectivity curve structure in integrated stock assessment models. Fish. Res. 158, 158–164.
Waterhouse, L., Sampson, D.B., Maunder, M., Semmens, B.X., 2014. Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions. Fish. Res. 158, 15–25.
Wayte, S.E., 2013. Management implications of including a climate-induced recruitment shift in the stock assessment for jackass morwong (*Nemadactylus macropterus*) in south-eastern Australia. Fish. Res. 142, 47–55.
Winker, H., Carvalho, F., Kapur, M., 2018. JABBA: Just Another Bayesian Biomass Assessment. Fish. Res 204, 275–288.
Xu, H., Thorson, J.T., Methot, R.D., Taylor, I.G., 2019. A new semi-parametric method for autocorrelated age- and time-varying selectivity in age-structured assessment models. Can. J. Fish. Aquat. Sci. 76, 268–285.
Zheng, J., Siddeek, M.S.M., Palof, K.J., 2021, Bristol Bay red king crab stock assessment in fall 2021. https://meetings.npfmc.org/CommentReview/DownloadFile?p=3ada484b-7d8c-42f0–9360-abc1560aa669.pdf&fileName=2%20Bristol%20Bay%20Red%20King%20Crab%20SAFE.pdf.