Fisheries Research 229 (2020) 105617

Contents lists available at ScienceDirect

Fisheries Research

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

Essential features of the next-generation integrated fisheries stock assessment package: A perspective

André E. Punt[a,b,*], Alistair Dunn[c], Bjarki Þór Elvarsson[d], John Hampton[e], Simon D. Hoyle[f], Mark N. Maunder[g,h], Richard D. Methot[i], Anders Nielsen[j]

[a] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98115, United States
[b] CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart, TAS, Australia
[c] Ocean Environmental Ltd. Wellington, New Zealand
[d] Marine and Freshwater Research Institute, Skúlagata 4, 101 Reykjavík, Iceland
[e] Oceanic Fisheries Programme, The Pacific Community, B.P. D5, 98848 Noumea, New Caledonia
[f] NIWA, 217 Akersten St, Port Nelson 7010, New Zealand
[g] Inter-American Tropical Tuna Commission, United States
[h] Center for the Advancement of Population, Assessment Methodology (CAPAM), United States
[i] NOAA Fisheries, Northwest Fisheries Science Center, Seattle, WA, United States
[j] Technical University of Denmark (DTU-Aqua). Kemitorvet, Building 201, 2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Handled by George A. Rose

Keywords:
Fisheries assessment
Integrated analysis
Simulation
Spatial dynamics
State-space
Tagging

ABSTRACT

Integrated analysis (or integrated population modelling) methods have become the preferred approach for conducting stock assessments, and providing the basis for management advice for fish and invertebrate stocks since the publication of a seminal paper by Fournier and Archibald in 1982. Methods to assess fish stocks based on single-species, single-area, age-structured models are now standard, with the major debates associated with these models related to data choice, model configuration assumptions, and data weighting. However, the current generation of stock assessment packages is not addressing all of the needs of stock assessment analysts and managers. A major challenge for any next-generation stock assessment package is the set of extensions needed to assess stocks that do not satisfy the 'well-mixed single-stock' paradigm. In addition, the next-generation stock assessment package needs to: (a) be able to capture age and size/stage dynamics simultaneously yet computationally efficiently, (b) scale from data-rich to data-poor, (c) include some multi-species capability, and (d) more appropriately deal with temporal variation (e.g., random effects and state-space models). In relation to data, there is a need to ensure that the next-generation stock assessment package better handles tagging data (age-size/stage models may help in this regard), in particular, to be able to use close-kin mark-recapture data. Efficient methods are needed to share parameter priors among stocks (satisfying the promise of the 'Robin Hood' paradigm). The next-generation stock assessment package needs to have associated appropriate training programs and documentation. Adoption of such a package will be facilitated by a data entry system that is well-documented, does not require specification of inputs that will not be used in an application, has an expert system to configure default settings based on best practices, and has associated code to automatically produce diagnostic statistics. Some technical challenges that have plagued stock assessment for decades warrant continued attention (at the theoretical and applied level) such as automatic data weighting and tuning, how to handle spatial and stock structure, improved coding to facilitate application of state-of-the-art methods for quantifying uncertainty, and adoption of true state-space formulations to allow more parameters to be treated as random effects. Future needs for features cannot be anticipated, so the key design consideration for the next-generation stock assessment package is to be flexible and modifiable to meet the requirements of analysts and users.

1. Introduction

Fish stock assessment methods have evolved over time due to (a) improvements in computational approaches (e.g., automatic differentiation and faster computers), (b) advances in methods for fitting models to data, and (c) the need to not only provide best estimates of

* Corresponding author at: School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98115, USA.
E-mail address: aepunt@uw.edu (A.E. Punt).

https://doi.org/10.1016/j.fishres.2020.105617
Received 3 February 2020; Received in revised form 11 April 2020; Accepted 27 April 2020
Available online 14 May 2020

Table 1

Comparison of the most commonly applied packages in terms of whether they include the main issues raised in this paper. The issues are divided into those for which methods to address them are well understood within the stock assessment community, those for which work has been undertaken but best practices are still lacking, and those for which developmental work is still being undertaken.

(a) The focus packages of this paper.

	Stock Synthesis	CASAL	MULTIFAN	GADGET	SAM	Casal2
Issue is fairly well understand and best practices are understood						
Scalable from data-rich to data-poor	Yes	Yes	Yes	Yes	No	Yes
Age-length dynamics	No[1]	No[1]	No	Yes	No	No[1]
Estimate data weights (index, composition and tagging)	Yes[2]	Yes[2]	Yes[3]	Yes[4]	Yes	Yes[5]
State-space formulation	No[6]	No[6]	No	No	Yes	No[6]
Generate expected values of data	Yes	Yes	Yes	Yes[7]	Yes	Yes
Reference point calculation	Yes[8]	Yes[8]	Yes[8]	Yes[7]	Yes[9]	No
Projections	Yes[8]	Yes[8]	Yes[8]	Yes[7]	Yes	Yes
Several alternative models exist but the field has yet to identify best practices						
Spatial structure	Yes	Yes	Yes	Yes	No	Yes
New issue to most assessment analysts; methods under development						
Multiple stocks	No	Yes	Implicitly	Yes	No[10]	Yes
Close-kin genetics	No	No	No	No	No	No
Multispecies relationships	No	No	No	Yes	No	Yes

(b) Other integrated analysis packages that are applied commonly

	BAM	ASAP	A4a
Scalable from data-rich to data-poor	Yes	Yes	Yes
Age-length dynamics	No	No	No
Estimate data weights (index, composition and tagging)	No	No	Yes[11]
State-space formulation	No	No	No
Generate expected values of data	Yes	Yes	Yes
Reference point calculation	Yes[9]	Yes[9]	Yes[9]
Projections	Yes	No	Yes
Spatial structure	No	No	No
Multiple stocks	No	No	No
Close-kin genetics	No	No	No
Multispecies relationships	No	No	No

1: Approximately using the morph feature (SS) or growth paths (CASAL); 2: not tagging data; 3. Not yet for composition data. 4. Taylor et al. (2007) (github.com/hafro/rgadget); 5. for normal and lognormal likelihoods in index and composition data in CASAL and Casal2, and Dirichlet in Casal2; 6: Can be implemented using Bayesian methods; 7 Elavarsson et al. (2018); 8: but more flexibility needed; 9: currently on a separate branch (https://github.com/fishfollower/SAM/tree/reference_points); 10: Yes, but as separate package "multi_SAM" (https://github.com/calbertsen/multi_SAM);
11: catch-at-age data only.

model parameters and outputs but also to quantify the uncertainties associated with the estimates. In the mid-1980s, the most common methods for stock assessment were surplus production models and Virtual Population Analysis (VPA). Surplus production models were fitted either as “process error estimators” (which allowed the estimation to be performed as a linear regression; Schnute, 1977) or as “observation error estimators” (which assumed that the population dynamics were deterministic, and the differences between the model predictions and the data were due to observation error, e.g. Butterworth and Andrew, 1984). Age-structured methods of stock assessment were primarily conducted using *ad hoc* tuned VPA (e.g., Laurec and Shepherd, 1983; Pope and Shepherd, 1985) or separable VPA (Pope and Shepherd, 1982). Uncertainty, when it was considered, was quantified using sensitivity tests to alternative assumptions, asymptotic approaches, likelihood profiles (e.g., Schnute, 1977) or using bootstrapping approaches such as those of Butterworth and Andrew (1984) in relation to surplus production models and Butterworth et al. (1990) for Virtual Population Analysis.

Fournier and Archibald (1982) provided the first example of a practical integrated (statistical catch-at-age) analysis in a rigorous statistical framework (although it built on earlier ideas, including those of Doubleday (1976) and Paloheimo, 1980). Fournier and Archibald (1982) illustrated how the model of the population dynamics could be separated from that of the observation process. They based their example application on a separable model for fishing mortality, following Doubleday (1976), and a linear relationship (with error) between fully-selected fishing mortality and fishing effort. Fournier and Archibald (1982) introduced several of the features now common in integrated analysis assessments, in particular, (a) fitting to the total catch and the proportion of the catch by age separately, (b) adding a penalty on the deviations in recruitment about the stock-recruitment relationship to the objective function, (c) allowing for ageing error when fitting to catch proportion-at-age data, and (d) formulating the objective function as a weighted sum of likelihood contributions (unlike such extensions of the approach such as CAGEAN [Catch AGE Analysis]; Deriso et al. (1985), which were based on generalized sum of squares). From a features point of view, the basic approach applied by Fournier and Archibald (1982) has been extended in several key ways (Maunder and Punt, 2013; Punt et al., 2013), including (a) fitting to the data by fleet, (b) fitting to more sources of data such as tagging data, (c) fitting to conditional age-at-length data and length-frequency data rather than catch proportion-at-age data (which is often a derived data source), (d) allowing for spatial structure (e.g., Quinn et al., 1990), and (e) conducting the assessment within the Bayesian framework to make use of prior information.

The need for time-varying features, in addition to the typical allowance for variation in annual recruitment, was apparent early during the development of the integrated analysis paradigm. The earliest versions of the stock assessment package Stock Synthesis, which were developed for northern anchovy *Engraulis mordax* (Methot, 1986, 1989) followed the approach of Fournier and Archibald (1982), and included features such as time-varying availability, and links between the values for biological parameters and environmental variables. A relatively recent development involves treating the random deviations in parameters about their expected values as random effects and maximizing the marginal likelihood rather than a penalized likelihood (e.g., Maunder and Deriso, 2003; Trenkel and Skaug, 2005; Skaug and Fournier, 2006; Nielsen and Berg, 2014; Berg and Nielsen, 2016; Cadigan, 2016). However, approaches based on the Kalman filter (Gudmundsson, 1994) and Bayesian approaches that effectively achieved the same goal (e.g., Punt and Hilborn, 1997) were proposed much earlier but were not often adopted for actual stock assessments.

Software packages implementing integrated analysis assessments have proliferated (see a recent summary of available packages and options in Dichmont et al., 2016). These packages (early examples include MULTIFAN-CL [Fournier et al., 1998] and Coleraine [Hilborn et al., 2000]) tend to have some common features, including that they treat what amount to random effects as "penalized deviations", they are coded using a framework (such as ADMB [Auto-Differentiation Model Builder; Fournier et al., 2012, or TMB [Template Model Builder; Kristensen et al., 2016]) that applies automatic differentiation methods to speed up parameter estimation, they are able to quantify parameter uncertainty, and they provide some (usually fairly *ad hoc*) basis for weighting data sources.

Five stock assessment packages (our 'focus packages') have seen extensive use across multiple jurisdictions in recent years: (a) Stock Synthesis (Methot and Weltzel, 2013), (b) CASAL [C++ Algorithm Stock Assessment Library] (Bull et al., 2012), (c) MULTIFAN-CL [MULTIple length Frequency ANanlyis - Catch at Length] (Fournier et al., 1998) (d) GADGET [Globally applicable Area Disaggregated General Ecosystem Toolbox] (Begley, 2014), and (e) SAM [State-space Assessment Model](Nielsen and Berg, 2014) [See Table 1 for a summary of features as they pertain to the issues raised in this paper]. These packages are under continual development to include features needed by assessment analysts so that they can tailor their stock assessments to the available data and the needs of management (ICES, 2013; Lynch et al., 2018). Each package has evolved based on the needs and interests of analysts and managers to whom the developers are beholden and hence has different strengths and weaknesses.

There will always be a need for two different types of assessment platforms: 1) those for routine use and 2) those for development purposes, where the results of the development process would be integrated into the general model once it is well-tested and determined to be useful. The existence of the five focus packages demonstrates the need and usefulness of such packages. However, development of methods is essential and bespoke models may be best for that. Similarly, bespoke models will be needed for those unique cases where a general model is not available or practical.

Even the current generation of stock assessment packages is not addressing all of the needs of stock assessment analysts and there is consequently a need to not only extend current packages but more fundamentally develop a next-generation stock assessment package that will integrate the best features of existing packages and be ready for the developments in population dynamics modelling and assessment that we expect to see in the next 5–15 years.

The key design consideration for the next-generation stock assessment package is that it is flexible and can be modified to meet the needs of analysts and users. For example, Casal2 (Doonan et al., 2016), which has not yet been widely used, was designed to succeed CASAL but using a next-generation approach, with a modular design and following modern programming standards. Casal2 was not designed with a particular model structure in mind but rather the realization that it is necessary to be able to add to the package without having to redesign it.

This paper outlines several categories of features that a next-generation stock assessment package should include (see Table 2 for a summary of these features). As a convention a single 'next-generation stock assessment package' is described, but multiple packages may well be developed. The recommendations reflect the experience of the authors who have been conducting stock assessments for 30+ years in multiple jurisdictions as well as the discussions during the 4–8 November 2019 Center for Advancement of Population Assessment Methodology (CAPAM) workshop on next generation stock assessment models.

Table 2
Key features of the next-generation stock assessment packages. Entries indicated by an asterisk are discussed in detail in this paper.

Basic structure
Age-, size/stage- and spatial structure*
Stock-structure*
Multiple fisheries and surveys
Flexible parameterization of the initial conditions
Multiple time-steps within a year
Biological parameters (estimated, pre-specified, age and/or size structured, and time-varying)
Flexible parameterization of growth
Flexible parameterization of natural mortality
Flexible parameterization of fecundity
Flexible parameterization of movement / dispersal
Stock and recruitment
Multiple functional forms, including non-parametric
Stock specific stock-recruitment relationship
Selectivity, retention, and catchability
A function of age, size or both
Multiple function forms, including dome-shaped and asymptotic
Use multiple selectivity patterns for the same fishery
Data
Index data
Composition data (length-frequency, weight-frequency, conditional age-at-length), including ageing error
Tagging data
Data on stock mixtures
Genetics data (close-kin-related)*
Multivariate priors and likelihoods*
Management
Dynamic reference points (e.g. Dynamic B_0*, Fishery Impact Plots)
Able to act as the operating model for MSE*
Ability to calculate reference points and conduct projections*
General issues
As few "tuning" parameters as possible (i.e. weights determined/ estimated from the data)*
Ability to represent uncertainty using asymptotic, bootstrap, profile and Bayesian methods
Straightforward to allow time-varying parameters to be modelled as random effects*
Fairly straightforward to add new assumptions within the general framework
All parameters can be "mirrored" among stocks, fleets, etc and linked to environmental variables
Scalable from data-poor to data-rich*
Ability to generate the expected values of the data to facilitate simulation testing*

2. Scaling from data-poor to data-rich

Stock assessments, particularly in jurisdictions such as the U.S. that require projections of catch targets and limits for most (if not all) managed stocks, are now required for a broader range of stocks than ever before. Many stocks for which assessments are needed are small and of low value, with few data on which to base stock assessments with many parameters. The features identified in this paper will generally lead to even more complex population dynamics models than those on which conventional stock assessments are based. Thus, the next-generation stock assessment package should be capable of conducting data-poor (e.g. catch or effort data only, length compositions with no catch), data-moderate (e.g., catch and one of index, age-composition or length-composition data, or only an abundance index), and data-rich (e.g., catch and multiple data sets) stock assessments within the same framework. An example of this is the hierarchy of assessment methods built on Stock Synthesis: (a) Simple Stock Synthesis (Cope, 2013) is a catch-only configuration of Stock Synthesis that uses Bayesian methods to construct probability distributions for key model outputs, including time-trajectories of biomass and fishing mortality, (b) Extended Simple Stock Synthesis (XSSS) (Cope et al., 2015a, b; Wetzel and Punt, 2016) extends Simple Stock Synthesis by allowing the prior distributions to be updated using data on time-series of relative abundance, and (c) Stock Synthesis, which includes the full array of options for conducting assessments of stocks. All use the full version of Stock Synthesis as the computational engine and are linked through the use of Bayesian methods, which allows a spectrum between parameters fully informed by data and by priors (although few fully Bayesian assessments have been conducted using Stock Synthesis; Monnahan et al., 2019). As such, other approaches that can be fitted using the Bayesian paradigm, such as CASAL or Casal2 could also be implemented to cover data-poor to data-rich situations.

General models have several important advantages over current purpose-developed data-poor assessment methods: (1) the assumptions of model structure and parameter values are explicit and therefore transparent; (2) the assessment structure can be extended naturally to accommodate additional data once they become available; (3) testing of alternative assumptions is facilitated; and (4) all the features of a general model such as diagnostics, methods for quantifying uncertainty, data weighting, simulation testing, etc., will be available in data-poor situations. Most, if not all, of the current model-based data-poor assessment methods should be special cases of the next-generation stock assessment package, at least approximately. Therefore, unless simulation tests show that an existing data-poor or data-moderate method is superior to the same, or an approximate, application in a general model[1], these methods probably should not be used for stock assessment except in cases with some specific requirements (e.g., clarity for stakeholders or efficient application for MSE).

Another aspect related to sharing of information among assessments involves informative priors for important parameters. Natural mortality rates and spawner-recruitment steepness are notoriously difficult to estimate within assessment models (e.g., Conn et al., 2010; Lee et al., 2011, 2012) or directly from data external to the assessment. The evolution of integrated analysis, and indeed stock assessment science in general, has tended to treat these parameters as fixed, such that the assessment is conditional on those fixed values. A better approach may be to take the shared information across many stocks and create informative priors for these parameters (Thorson, 2019). However, the input to the assessment should preferably be such that users select priors based on available relevant and reliable information rather than simply accepting default priors (e.g., from a general meta-analysis).

3. Stocks and spatial structure

Traditional stock assessments have assumed that the stock assessment is being applied to a single homogeneous population (aka 'stock'). However, this assumption is often violated due to low rates of mixing across a stock's range and the use of convenient management boundaries to define stocks. There are several reasons for including spatial structure in stock assessments (see Table 1 of Punt (2019a)). In particular, it is common for different areas (spatial strata) within the area defined by a stock to exhibit different trends in abundance indices or catch age-composition (that cannot be explained by differences in gear selectivity). Moreover, many simulation studies (summarized in Punt [2019a]) have explored the benefits of using spatially-structured stock assessments in terms of improved estimates of biomass and fishing mortality as well as management reference points. Punt (2019a) identified five classes of stock structure archetype and Punt (2019b) categorized the multi-stock cases into (I) a single stock is found in the region being modelled, which contains multiple areas, and (II) more than one stock occurs in the modelled region and those stocks are found in multiple areas. The latter set of cases includes when there are multiple spawning grounds and a shared feeding area, amongst others.

In general, including spatial structure in assessments when the available data reject the assumption of a homogeneous stock leads to less bias for quantities of management interest, albeit at the cost of poorer precision. Equally importantly, ignoring spatial (and stock) structure when managing fisheries can lead to unintended overexploitation of local populations (e.g., Fu and Fanning, 2004; Ying et al., 2011). There is also the benefit of sharing information among stocks as described above, which can be more effective when multiple stocks are modelled simultaneously. Unfortunately, spatially-structured assessments are rare, because of data and software limitations, problems with parameter estimation, institutional inertia, and challenging policy implications (Cadrin et al., in press). However, much research is being conducted on how to include stock and spatial structure into stock assessments (see the papers in volume 74(11) [2017] of the *Canadian Journal of Fisheries and Aquatic Sciences;* Berger et al., 2017a), but much of this work is developmental and has yet to directly impact stock assessment practice.

The following section distinguishes 'stocks' from 'spatial structure within a stock' as the model features needed to address these two concepts differ.

3.1. Stock structure

The stock concept in fisheries is challenging because there is no unique definition of a stock, and terminology is vague (Cadrin et al., 2014; International Whaling Commission, 2014; Cadrin, 2020). Here we define a stock as a group of animals that is demographically independent of groups in other areas (or at least the rates of movement among stocks are sufficiently small to be demographically inconsequential). However, as noted by Cadrin (2020) other definitions of 'stocks' exist, including 'phenotypic stocks' and 'harvest stocks' or simply defining stocks based on national or regional political boundaries.

Many papers have outlined aspects related to a generic model with multiple stocks (see the summary in Punt [2019a]) and these could form the basis for specifying how stock-structure is included in the next-generation stock assessment package. Among the simplest ways to jointly model multiple stocks is the approach of modelling multiple stocks with stock-correlated processes errors (see Albertsen et al., 2017 for the approach using multiple species) while among the most complicated are the models used for assessments of whale stocks by the Scientific Committee of the International Whaling Commission (e.g., Punt et al., 2014; Ross-Gillespie et al., 2014, 2015).

GADGET, CASAL and Casal2 allow for multiple stocks, and the morph concept in Stock Synthesis is essentially a stock concept. MULTIFAN-CL estimates recruitment by area, and the "morph" option in Stock Synthesis allows recruits to join a morph at birth (and never change morphs), with each morph distributed across spatial areas at birth and each independently moving among areas as it ages. However, Stock Synthesis does not allow for morph-specific density-dependence. CASAL and Casal2 have a similar process, referred to as growth-paths. A key challenge for multi-stock models is determining the degree to which density-dependence acts locally or globally, particularly the stock-recruitment relationship.

3.2. Spatial structure

A common assessment assumption treats spatially-segregated

[1] For example, Winker et al. (2020) use simulation to compare the 'JABBA-Select' Bayesian biomass dynamics method and an age-structured production model (ASPM) approach and found that JABBA-Select performed as well as the ASPM in terms of point estimation and outperformed it in terms of quantifying uncertainty.

fisheries as having fleet-specific selectivity relative to the total stock, with the hope that this can mimic the effects of spatial population structure. While this 'fleets-as-areas' approach (Berger et al., 2012; Waterhouse et al., 2014; Hurtado-Ferro et al., 2014) is often able to provide reasonable fits to available data, it has also been shown to lead to biased estimates of biomass and fishing mortality (e.g., Punt et al., 2016a). This is particularly concerning when indices of abundance are available for multiple areas but show different trends that are not explained by the gear selectivity. One solution is to use a spatial model, as is the case with MULTIFAN-CL (Fournier et al., 1998; Hampton and Fournier, 2001). However, this extension leads to more complex population dynamics models, requiring additional parameters relating to movement, the spatial allocation of recruitment and, potentially, spatial variation in biological parameters (in particular growth rates). Therefore, for such models, there is a need to include data that are informative about these processes. Growth is particularly problematic because it can be genetic (based on the area of birth), environmental (based on the current locations), or both, and is difficult to model in standard age-structured models.

3.3. Next-generation features

Including spatial (and stock) structure into an existing population dynamics model is complicated, and it will be critical for the developers of any next-generation stock assessment package to include this in the fundamental design. Inclusion of movement among spatial areas introduces another dimension of flexibility that will be confounded with natural mortality, growth, selectivity, and recruitment (Cadrin et al., in press). The next-generation stock assessment package will need to provide tools to diagnose and control this confounding to provide accurate assessments.

There are several barriers to the implementation of spatial structure in stock assessments (Table 1 of Berger et al., 2017b), some of which relate to lack of data, others to how to designate spatial strata and stocks and communication of results to stakeholders, and yet others to how best to incorporate spatial structure in assessments to achieve the lowest bias and highest precision, as well as adequate forecasting ability. The main technical needs of a next-generation stock assessment package in relation to stock and spatial structure (some of which are included in the focus packages above while others require research to identify best practices) are the abilities to:

- track numbers-at-age (or size/stage) by stock and spatial area;
- allow for density-dependence at the spatial area level as well as at the total stock level (high local density for one of several stocks in an area may lead to density-dependence due to competition for resources);
- allow each data set to apply to multiple stocks (e.g., Punt et al. [2000] assumed that spatial CPUE indices for school shark *Galeorhinus galeus* applied to all of the stocks in the area concerned; Ross-Gillespie (2016) took this concept further and conducted assessments of the two Cape hakes *Merluccius capensis* and *M. paradoxus* in which the CPUE indices applied to two species);
- allow for nesting of spatial scales such that a population model can appropriately utilize data types collected at fine scale and coarse spatial resolutions;
- allow for multiple movement types including advection, diffusion, and that movement responds to environmental drivers;
- account for multiple hypotheses regarding movement, including age- and sex-specific processes as well as density-dependent and time varying movement;
- integrate tagging data into the assessment model – tag-recapture data can be included in an assessment using the 'release-conditioned' formulation (i.e., the probabilities of recapture in different regions are computed as proportions of the total number of animals originally tagged and released in each source region which allows estimation of both mortality and movement rates; e.g., Hilborn, 1990; Maunder, 1998; 2001a,b; Hampton and Fournier, 2001; Hannesson et al., 2008; Goethel et al., 2011) or the 'recapture-conditioned' formulation which is robust to tag loss rate but only allows movement rates to be estimated (e.g., McGarvey and Feenstra, 2002; McGarvey et al., 2003) or using both formations simultaneously (e.g., IWC, in press) – Vincent et al. (2020) found that estimation using a release-conditioned model was usually less biased and more precise than using a recapture-conditioned model; and
- fit to data on stock mixing, including information from tagging (e.g., Hilborn, 1990) and other biological markers such as parasites (e.g., De Moor et al., 2017) and otolith microchemistry (e.g., Kerr et al., 2017), as already implemented in Stock Synthesis.

Multi-stock models have the potential to assist in cases where there are multiple stocks, some data-rich and others data-poor. For example, Punt et al. (2011) developed a 'Robin Hood' approach to stock assessment that linked assessments for stocks in terms of the trajectories of fleet-specific fishing mortality expressed relative to its mean. Pribac et al. (2005) conducted joint assessments of three stocks of gummy shark *Mustelus antarcticus* in which the parameters that determine productivity and availability are the same among stocks. This allowed the information for the data-rich stock of gummy shark in Bass Strait to inform the assessments of the more data-poor stocks off Tasmania and South Australia, but also allowed the uncertainty associated with those parameters to be propagated to the outputs of the models for the data-poorer regions. This would not be the case had the values for the parameters for the Tasmanian and South Australian stocks been set to the estimates for the stock in Bass Strait.

A major constraint on incorporating spatial structure in stock assessment is the lack of data of sufficient quality and quantity for each model region, and to inform estimates of potentially time-varying movement among regions. Major technical challenges with multi-stock spatial models that will require research prior to specification include: (a) how to model how stock integrity is maintained, (b) the development of ways to define management reference points when, for example, animals move and growth changes among areas, and (c) how to model fine-scale spatial structure such as local depletion (Maury and Gascuel, 2001), or isolation by distance, a feature observed in a range of species, including Pacific cod, *Gadus macrocephalus* (Cunningham et al., 2009).

Generalizations to include multiple stocks in an assessment provide an initial way to allow for multiple species. However, including multiple species in an assessment package could require account to be taken of trophic interactions (although see Albertsen et al., 2017). GADGET and Casal2 already allow for predation among species, and several multispecies methods based on the integrated analysis paradigm have been developed (e.g., Jurado-Molina et al., 2005, 2006; Kinzey and Punt, 2009; Van Kirk et al., 2010; Curti et al., 2013; Plaganyi et al., 2014; Holsman et al., 2016; Trijoulet et al., 2020). However, these methods tend to be for 'research' purposes and have rarely been adopted for use in assessments on which tactical management decisions are made. Multiple species also complicate how the data are input to the assessment and the calculation of likelihood components, some of which may be for individual species and some may be for data combined across more than one species (e.g., Ross-Gillespie, 2016).

The MICE [Models of Intermediate Complexity for Ecosystem] approach (Plaganyi et al., 2014) is an alternative framework for providing management advice given predation, competition, and environmental drivers of population dynamic processes. This approach can scale from data-poor to data-rich by developing population dynamics models that are tailored to each species and stock in the system rather than forcing all species to be governed by the same dynamics equations. Given the very large number of ways that species may interact, it may be appropriate for the next-generation stock assessment package initially not to include multispecies interactions beyond predation and fishery technical interactions, but allow flexibility for extending the model to include other relationships in the future.

4. Age- and size/stage-population models

The first integrated analysis assessments (e.g., Fournier and Archibald, 1982; Quinn et al., 1985) fitted only to index and age-composition data, and this is still the case for stock assessments applied to stocks for which there is extensive age-composition data (e.g., ASAP; Legault and Restrepo, 1998; Miller and Legault, 2015; SAM). However, it is common for a key data source to be size (or stage)-composition data for less data-rich fisheries (and fisheries for tunas for which no or limited ageing data are usually available). Assessments for hard-to-age species such as crabs, rock lobsters and prawns are often based on length-structured population dynamics (Punt et al., 2013). CASAL and GADGET have the ability to represent the population dynamics using a length- instead of an age-structured model, which allows length-dependent mortality to modelled., They can be set up as purely length-based models.

The ability to fit to size-composition data is a key advantage of integrated methods over alternative age-based assessment methods as VPA. Size-composition (usually length-frequency) data are often available for fleets for which age-composition data are not available and the sample sizes for size-compositions are often much larger than for age-compositions. In some cases, age-length keys are not available for all years, and sharing age-length keys among years to convert lengths into ages can lead to substantial bias. Most of the major stock assessment packages (Stock Synthesis, CASAL, Casal2, MULTIFAN-CL) are based on age-structured population dynamics models and the size-structure of the catches (or surveys) is predicted from the underlying population and catch age-structure. This involves assuming that the probability distribution for size-at-age takes some pre-specified form (e.g., normal, log-normal, or gamma), setting the mean size-at-age based on an estimated growth curve, and assuming that the standard deviation of size-at-age is a function of age or expected size-at-age.

Assessment packages have been extended to make use of conditional age-at-length data (aka age-length keys) to estimate the relationship between the mean and standard deviation (or CV) of length-at-age and age (Methot, 2000; Hoyle and Maunder, 2006; Punt et al., 2006; Methot and Wetzell, 2013). Integrating the estimation of growth within the stock assessment appropriately accounts for size-based selectivity and size-based sampling, which are both common features of many stocks.

A key assumption of this approach to including size-composition data in a stock assessment is that the distribution of size-at-age does not change over time due to fishing-induced effects on growth or size-selectivity. While such effects are likely small when fishing mortality is low, they could be substantial (and lead to biased assessment outcomes) when fishing mortality is high (e.g., Taylor and Methot, 2013). The ideal solution to this problem is to use a model that simultaneously keeps track of both age and size dynamics (age-size structured models; e.g. Allen Akselrud et al., 2017), but of the five major packages only GADGET currently has this ability. One key reason for this is the computational demands associated with keeping track of both age and size structure.

Taylor and Methot (2013) introduced the “platoon” concept to Stock Synthesis and McGarvey et al. (2007) developed the growth slice approach, both of which approximately allow for fishing-induced modification of size distributions. An additional reason for tracking both age- and size-dynamics is that tagging data are the key source of information on movement, growth and fishing mortality for some stocks, but it is necessary to allocate tags (for which length-at-tagging but not age-at-tagging are usually known) to age, often based on a deterministic growth curve outside the model (e.g., Stock Synthesis and the Integrated Tagging and Catch-at-Age Analysis [ITCAAN]; Maunder et al., 2001a) or inside the stock assessment (MULTIFAN-CL and bespoke model of Punt et al., 2000), which allows the internally estimated growth curve to determine the age of the releases. In contrast, calculation of the probability that a tagged animal is of a given age is straightforward if the model is age- and length-structured.

Although the computational and memory constraints associated with keeping track of age and size/stage simultaneously should not be ignored, it is now possible to include age- and size/stage-dynamics in population dynamics models coded using TMB and ADMB (e.g., Allen Akselrud et al., 2017) because of their efficient use of automatic differentiation. The next-generation stock assessment package should include age and size dynamics, but be configured so that it can be run as purely an age- or size-structured model, to avoid long run times for simple models.

An Achilles heel of contemporary assessment methods that fit to size-composition data is their temporal resolution. While animals retain the same integer age throughout a year or season, they grow at various rates during the year, so the time at which animals are sampled needs to be aligned with the appropriate time resolution along the continuous growth axis. Mismatch of length modes is thus a source of process variation, but most contemporary assessment methods downweight this lack of fit as if it was due to observation error. New algorithms to compare observed to expected size compositions could address this shortcoming. Alternatively, the process variation should be modelled (e.g., finer temporal scale and variation in growth rates). Further complications are multiple recruitment events within a year and temporal changes in growth rates. This can be accomplished to a degree within GADGET and Stock Synthesis by defining the relevant processes (e.g., growth and migration) as time-varying.

5. Data weighting and random effects

Data weighting is a key component of stock assessments. It relates firstly to the choice of which data sets to include in an assessment and, given this, the weight to assign to each data set and each point within each data set. The weights assigned to catch and index observations have generally related to the variances of the catches and indices (often negligible for catches), while the weights assigned to age-, length- and weight-composition data and to conditional age-at-length data are the ‘effective’ sample sizes (as if the data were collected according to simple random sampling from the underlying population). The data for each year are assumed to be independent, an assumption that is rarely correct for catch-rate-based indices when these are based on the results of a ‘standardization’ method (but see Butterworth and Punt (1992) for an exception; this issue is automatically dealt with when the catch-rate standardisation is integrated into the assessment model [e.g., Maunder, 2001b]).

The earliest integrated assessments pre-specified these weights in an *ad hoc* manner (e.g. “50″ for indices of abundance (corresponding to an assumed residual standard deviation of 0.1), and “200″ for effective sample size for length- and age-compositions; e.g., Fournier and Archibald, 1982), and this is still the case for some assessments (e.g., Zheng and Siddeek, 2019). However, methods have been developed to provide assessments with ‘input’ effective sample sizes[2] and to revise these using iterative methods (e.g., McAllister and Ianelli, 1997; Francis, 2011, 2013). More recently, methods have been developed that allow the effective sample size to be estimated within the assessment (Maunder, 2011; Thorson et al., 2017a). However, the new methods do not adequately deal with correlations in residuals caused by the sampling process and model misspecification (Francis, 2014). The next-generation stock assessment package should include methods for

[2] The input sample sizes (e.g., the standard errors from CPUE standardization methods and the number of fish measured for composition data) often substantially over-estimate the true effective sample size owing to the pseudoreplication inherent in fisheries data collection systems.

estimating the residual variances for the index data and, as for Stock Synthesis and Casal2, estimate Dirichlet sample sizes (and other measures of the information content of compositional data) as part of the model fitting process, in addition to being able to specify weights directly. Similarly, for assessments that use tagging data, the weight assigned these data should either be based on tuning methods (e.g., Punt et al., 2017) or estimated as part of the model fitting process by choosing a distribution for the recaptures, which includes an estimable variance parameter (e.g., the negative binomial).

A more serious problem relates to how to handle what should be random effects. Many of the parameters in contemporary stock assessments are (or should be) treated as random effects. The variable that is most commonly, by its nature, a random effect is recruitment (or if there is a stock-recruitment relationship, the deviations in recruitment about the stock-recruitment relationship). A penalty is placed on recruitment (or recruitment deviations) in many stock assessment packages (including Stock Synthesis, CASAL, and MULTIFAN-CL) when parameter estimation is based on maximum likelihood estimation. The size of the penalty depends on the value set for σ_R (the standard deviation in recruitment / recruitment deviations). However, the estimate of this parameter is biased in a penalized maximum likelihood framework because the estimate corresponds either to a local minimum of the objective function or because the likelihood is degenerative towards zero (Maunder and Deriso, 2003).

Methods such as that of Methot and Taylor (2011) have been developed to tune the value of σ_R. Tuning algorithms involve selecting an initial value for the parameter of interest, fitting the model given that specified value for the parameter, and applying a tuning algorithm to update the value for the parameter given the fit of the model to data (Thorson, 2019). However, tuning algorithms may not converge to the true values. Full integration across the recruitment deviates is computationally intensive (Maunder and Deriso, 2003), and it was not until the availability of methods based on the Laplace approximation for computing marginal likelihoods that it became common to both treat recruitment / recruitment deviations as random effects and estimate σ_R. This approach is integral to SAM. Bayesian methods automatically do the integration (e.g., Coleraine; Hilborn et al., 2000), but are also computationally intensive (Punt and Hilborn, 1997; Maunder and Deriso, 2003).

Although the most common quantity that should be treated as a random effect is recruitment, other quantities that are candidates for treatment as random effects are natural mortality, growth, catchability, and selectivity (all of which are likely to vary over time and among ages/sizes). SAM allows for temporal variation in recruitment, survival, and fishery selectivity. Xu et al. (2019) extended Stock Synthesis to include semi-parametric selectivity based on a parametric component and an autocorrelated nonparametric component consisting of deviations from the parametric component, the parameters of which can be estimated as random effects. However, almost every parameter in a population dynamics model (e.g., growth increment, fecundity, fishery and survey catchability) could potentially be treated as a random effect. Inclusion of many random effects in an assessment can substantially increase the run time (by several orders of magnitude). Miller et al (2016) implemented random effects in a way that naturally allows for inclusion of environmental factors to inform these temporal effects.

Inclusion of temporal variation if model parameters will explain some of the unmodelled process variation that is typically assumed to be observation error and thus will typically increase the estimated weight on the different data sets, particularly the composition data. Maunder et al. (2017) argue that the observation model should represent the sampling error and that the model should be specified to account for all the parameter temporal variation and remove model misspecification. Random effects can also be used to model correlation in residuals and may be a solution to the issues with modelling composition data.

Consideration should be given to coding the next-generation stock assessment package using TMB, because TMB provides a way to conduct the calculations more effectively than ever before.

6. Close-kin mark-recapture

Tagging data can be included in integrated analysis stock assessments. These data provide information on fishing mortality, natural mortality, movement and abundance. Sibert (1984) and Hilborn (1990) developed the framework on which multi-area stock assessments in which tagging data are used to estimate movement are based. This framework involves modelling each group of tag releases as a tagged population where ‘recruitment’ occurs at tagging, and the ‘mortality’ of tags occurs due to fishing and natural mortality, tag-related mortality, and tag loss. The likelihood function then includes the conventional data used for stock assessment as well as a component for the tag recaptures.

Close-kin mark-recapture (CKMR) is an approach that integrates genetic methods of population estimation and population dynamic models to estimate abundance and potentially a range of demographic parameters such as fecundity-at-age and natural mortality-at-age. In principle, close-kin genetics can also inform stock structure and movement. The approach uses genetic markers to identify animals that are related (e.g., parent-offspring pairs; half sibling pairs, and perhaps other relationships; Skaug, 2001; Bravington et al., 2016a). The data are analysed within the general framework of mark-recapture data, but the analysis is not subject to many of the problems typically associated with conventional tagging data such as tag-loss, tagging-induced mortality and tag reporting. However, in common with all tagging programs, the sampling scheme needs to be well-designed and sufficient animals need to be genotyped for precise estimation of parameters.

The approach has moved from being a research topic, where applicability was constrained by technical challenges associated with genotyping, to one that has been applied to some highly contentious fisheries, including those for southern bluefin tuna (*Thunnus maccoyii*) (Bravington et al., 2016b), school shark (*Galeorhinus galeus*), white sharks (*Carcharodon carcharias*) off Australia, and Northern Australian river sharks (Davies et al., 2015). The estimates of absolute abundance from CKMR can be more precise than those from typical stock assessments (CV = ~0.17 for southern bluefin tuna; Bravington et al., 2016b) and even the estimates of survival are remarkably precise (CV = ~0.03 for southern bluefin tuna; Bravington et al., 2016b). The application of CKMR methods to southern bluefin tuna substantially altered the estimates of the size of the population. The CKMR approach is being considered for application to several other species, in particular tunas, for which it remains challenging to obtain indices of abundance, except using commercial catch-per-unit effort data.

Given the necessary data on genetics, the key (but by no means trivial) modification to existing stock assessment packages is to add likelihood components for the kinship probabilities (see Bravington et al., 2016a for examples), and this has been achieved for the assessments of southern bluefin tuna and school shark. Bravington et al. (2016a) provide the probabilities of parent-offspring and half sibling matches for several scenarios (e.g., when stock structure is known and when it is not). One of the key decisions to make when including close-kin mark-recapture data into the next-generation stock assessment package is how general a framework is needed, because this could substantially complicate any package.

7. Simulation testing

7.1. Evaluation of estimation performance

Much of the knowledge regarding the performance of stock assessment methods has been obtained from simulation studies, and simulation studies have been integral to development of good practices for stock assessment (e.g., based on past CAPAM workshops; Maunder et al., 2014, 2016, 2017; Sharma et al., 2019; Cadrin et al., in press). "Self-testing" in which the population dynamics model is used to generate simulated data sets that are consistent with the estimation method, is a fundamental test of any method of stock assessment. If a stock assessment method cannot reproduce parameter values when it is correctly specified, it cannot be expected to perform adequately when its assumptions are violated. "Self-tests" are a minimum standard for performance evaluation, and simulation testing should also be conducted when assumptions are uncertain. Tests using the same estimation and generation model are also important to understand which parameters are estimable, and biases due to specific model misspecifications. Use of simulation models that are much more complex than the estimation model is, however, essential to illustrate the consequences of more realistic measures of error, but the results can be difficult to interpret.

The next-generation stock assessment package should be usable as a simulation model (where the parameter values of the population dynamics and observation model are pre-specified) to allow the expected values for the observations to be computed. Observation error can then be added to the expected values based on various sources of uncertainty, including sampling distributions that differ from those underlying the original assessment. The SS3sim framework (Anderson et al., 2014a; Johnson et al., 2019) works this way and has been used in several studies examining the performance of Stock Synthesis (e.g., Ono et al., 2015; Kuriyama et al., 2016; Stawitz et al., 2019). Some stock assessment packages, such as Stock Synthesis, GADGET, SAM, MULTIFAN-CL, CASAL, and Casal2, already include the ability to simulate data sets, but the ability to allow for unusual problems with data is limited. Thus, it may be preferable to write separate software to account for "nasty" data problems.

7.2. Management strategy evaluation

Management Strategy Evaluation (MSE; Bunnefeld et al., 2011; Punt et al., 2016b) is the state-of-the-art method for evaluating the performance of management strategies in terms of the ability to achieve management goals. An MSE consists of several components of which two are the operating models (the models that represent the system to be managed) and the management strategies being evaluated (management strategies are often combinations of data collection methods, estimation methods, and harvest control rules, although the latter two components can be combined). Stock assessments can form the basis for either of these components. However, the use of conventional stock assessment methods as operating models can stifle the range of scenarios considered in the MSE, which are usually much broader than the range conventionally considered in stock assessments (Punt et al., 2016b). The operating model and the uncertainty it represents should be conditioned on the available data to the extent possible (e.g., using bootstrap or Bayesian methods). Otherwise the reliability of the results will be questionable. The next-generation stock assessment package will have a wide range of model configuration options and the capability to include prior information to inform them, thus allowing a broader range of scenarios to be examined, and making it more applicable for use as an operating model. However, such configurations will still not cover the whole range of possible scenarios. The next-generation stock assessment package could act as the estimation model component of a management strategy evaluation (and should if the proposed management strategy is to be based on that assessment method). The ability to evaluate management strategies during an MSE would be enhanced if the stock assessment could be run in "quick mode" (e.g., by limiting the amount of output produced to the minimum needed to apply the harvest control rules).

8. Other needs for the next-generation stock assessment

8.1. Harvest control rules, projections and reference points

A key reason for conducting assessments is to develop management advice. This is often in the form of estimates of fishing mortality and biomass relative to reference points. The next-generation stock assessment package should include the ability to compute the primary biological references points used for stock status determination, including fishing mortality and biomass relative to F_{MSY} (the fishing mortality corresponding to Maximum Sustainable Yield), $F_{0.1}$, $F_{x\%}$ (the fishing mortality corresponding to a reduction in spawning biomass-per-recruit of x%), and $FB_{x\%}$ (the fishing mortality corresponding to a reduction in spawner biomass of x% from unfished levels). In addition, reference points such as B_{low} (the lowest spawning biomass ever encountered), F_{crash} (the lowest fishing mortality corresponding to equilibrium biomass of zero), and $\lambda_{\max}$ (the maximum rate of increase at low population size) should be reported. It should be possible to report point estimates of stock status relative to these reference points as well as measures of the uncertainty of the resulting ratios. Calculation of 'dynamic' reference points (such as 'dynamic B_0', which is the reference level of unfished biomass, B_0, under the prevailing environmental conditions [cf. MacCall et al., 1985]), as well their use in harvest control rules should be included in any next-generation stock assessment package. Estimates of fishery impacts, which are invaluable when determining trade-offs among fisheries (Wang et al., 2009), including during the projection period, should be part of the next-generation stock assessment package. The specification of reference points is well-developed for traditional single-stock, single-species non-spatial models, but there are challenges defining reference points for spatial models (Berger et al., 2017b) and multi-species models (Moffitt et al., 2016).

The projection capability of the next-generation stock assessment package should include deterministic projections as well as those that account for both parameter uncertainty and stochasticity due to future process variation (e.g., recruitment, growth, catchability and selectivity). In addition, allowance should be made for future changes in other parameters allowed to be time-varying. The projections should allow forecast removals to be specified in terms of catch, effort, or fishing mortality, with different specifications by fleet. Rebuilding analyses should be possible, i.e., finding levels of fishing mortality at which rebuilding to a target occurs with a specified probability by a specified time.

All of the focus packages are able to compute some reference points and conduct projections, but more flexibility is needed in all packages to capture the broad range of needs for many jurisdictions and fisheries to which the next-generation stock assessment package would be applied. This is because unlike the population models on which stock assessments are based, and the data used for parameter estimation, the reference points and harvest control rules on which management advice is often based vary substantially among jurisdictions.

Management needs vary among jurisdictions, and new reference points and management approaches will be developed in future. It should be easy to change the next generation stock assessment package to report new management-related outputs.

8.2. Parameterization

The ability to extend the model using alternative parameterizations without modifying the model code base can be beneficial to the users. Most of focus packages allow this to varying degrees. For example, GADGET allows the user to replace parameters with model formulae (functions of model variables and data), often in conjunction with parameter values that can vary according to a user-defined schedule (see "Time Variables" in Begley, 2004). This would allow the user to redefine model processes by rescaling model parameters (e.g., log transform) or setting up time-varying processes such as temperature-dependent growth. This functionality is commonly used in GADGET models, for instance to estimate initial depletion rate (Elvarsson et al., 2018) or relate environmental factors (e.g., river discharges and wind) to projected recruitment (e.g., Rincon et al., 2019).

8.3. Priors and likelihoods

Most stock assessment packages have the capability to impose priors on model parameters to include information and uncertainty regarding the parameters from other sources. Studies such as those of Thorson et al., 2017b and Hamel (2015) (priors on natural mortality) and Dorn (2002) and Thorson (2020) (priors on steepness) provide priors that could be used in stock assessments. The priors developed by Thorson et al., 2017a and Thorson (2020) are joint priors for multiple parameters. Parameters estimated from external analysis may be highly correlated (e.g., the asymptotic length and growth rate from a growth analysis) and this correlation needs to be taken into consideration in priors. The next-generation stock assessment package should consequently include the ability to specify multivariate priors.

Similarly, appropriate likelihood functions should be included in the next-generation stock assessment package, and they should continue to be refined. Multivariate likelihood functions that take the variance-covariance matrix from external analyses (e.g., CPUE and/or composition standardisation) should be included (Maunder et al., 2020). Likelihoods for additional data types should also be developed and integrated into the general stock assessment package (e.g., tagging growth increment data).

8.4. Diagnostics

An essential part of model development is diagnosing model results to ensure that the most appropriate model configuration is being used. The focus packages, as well as some bespoke assessment methods have software associated with them for the production of diagnostics (Dichmont et al., 2016). For example, the R package R4SS (Taylor et al., 2019) contains a wide variety of diagnostic statistics developed for Stock Synthesis. Nevertheless, diagnostics are underdeveloped for integrated stock assessment models and general diagnostic tools are not available, although some progress has been made (e.g., Carvalho et al., 2017). The CAPAM workshop on data conflict and weighting, likelihood functions, and process error (Maunder et al., 2017) summarized several existing and new diagnostic statistics, some of which (e.g. Punt, 2017) have now been included in the packages used routinely to evaluate model performance.

Example of diagnostics that are available for some but not all packages are the "age-structured production model' (ASPM) diagnostic and residual analyses based on one-observation ahead method. The ability to scale an assessment from data-poor to data-rich will allow the use of the ASPM diagnostic (Maunder and Piner, 2015). This diagnostic evaluates whether an ASPM can explain trends in an index of abundance with good contrast; if so, it can be concluded that a production function is apparent, and the index is providing a reasonable proxy for stock trend. The one-observation method ahead method (Thygesen et al., 2018) can compute residuals with the correct properties even when the observations are correlated (either directly assumed correlated or correlated as a consequence of random effects in the model) but is currently only available for SAM and WHAM [Woods Hole Assessment Method; Miller et al., 2016][3].

Developing appropriate diagnostic methods and the associated software should be a priority for stock assessment research, and the potential of a large user base for the next-generation stock assessment package should provide support for the investment in software that would produce new and better diagnostics. Guidelines for how to respond to the outcomes of a set diagnostics and address model misspecification are needed (Maunder et al., 2017).

8.5. Coding and output

Although not a focus of this paper, the software design of the next-generation stock assessment package needs to be decided before model specification work is initiated. The coding should ensure that the final product is modular to allow for easy addition of new features (e.g., a new stock-recruitment relationship, selectivity pattern or likelihood component), and that the decision to include a new 'partition' (i.e., a new dimension to the N-matrix) can be achieved without needing to rewrite code. For example, most stock assessments compute spawning biomass by multiplying numbers-at-age by a maturity ogive. However, assessments of some crab stocks such as snow crab *Chionoecetes opilio* (e.g., Szuwalski, 2019) keep track of separate partitions for mature and immature animals, in particular because growth differs between immature and mature animals (there is a 'terminal' molt to maturity, which means that mature animals do not grow). In addition, the code should be 'efficient' in that support for complex (and likely little used) features does not unduly slow down the code.

Development of the next-generation stock assessment package will be a substantial undertaking so standard software development good practices should be implemented, including clear documentation of the code, and automated testing to ensure that new features do not lead to unintended changes in results for existing assessments. Version control needs to be implemented, along with clear rules that designate who is able to update the main branch of the code.

Stock assessments often involve many model runs to adequately explore alternative models. Long run times may limit the number of analyses that can be conducted in the (usually) limited time available to conduct an assessment. Some analyses can be run simultaneously, and simply purchasing more processors can solve the problem. However, other model runs need to be conducted consecutively. For example, developing a model requires running an assessment and then evaluating the results before deciding on the next analysis to run. Therefore, efficient code and parameter estimation are needed.

Recent advances in the availability of affordable multi-thread processors have opened many potential ways to reduce the processing time of stock assessments, including parallelizing components of the assessment (e.g., evaluating different data sets, tag cohorts, stocks), use of large numbers of threads to speed up minimisation, Bayesian estimation with MCMC using speculative chain execution (Byrd et al., 2010), and rapid evaluation of ensemble models with large numbers of alternative parameters. The potential to implement new algorithms not yet conceived of reinforces the need for the software architecture of the next generation package to be highly flexible and modular, so that new ideas can be implemented effectively.

8.6. User interaction

The decisions related to setting up a model become more arduous as models and their associated estimation methods become more complex and include more features. Moreover, less time is available to make these decisions, and less experienced analysts have to conduct assessments as the demand for more assessments increases. Training and documentation will be essential to ensure that broadest adoption of the next-generation stock assessment package. Training should range from that targeted towards novice as well as expert users given that there is a wide range of expertise in stock assessment worldwide. Contemporary general models are complex with many options that need to be specified and even the most experienced analyst can make mistakes. Therefore, the development of a user interface that includes an expert system that implements defaults based on best practices, makes numerous

[3] This model is still under development and has yet to be used for an operational stock assessment.

consistency checks, and produces warnings needs should be developed to enhance ease of use. The availability of test data sets as well as example data sets that cover a broad range of stock assessment scenarios (e.g., data-rich to data-poor; non-spatial to spatial) will help users quickly specify initial versions of assessments. Finally, use of the next-generation stock assessment package will be facilitated by a data entry system that is well-documented and does not require specification of inputs that will not be used in an application, as well as code to automatically produce diagnostics statistics.

An important component of the process of developing the next-generation stock assessment package should be agreement on a standard format for inputs and (particularly) outputs. Such a common format would enable analysts to share data files (enabling the same data to be easily analysed using multiple assessment packages) and output files. A common format for output files will enhance development and application of common diagnostic statistics. A large potential user base should encourage and justify the development of new diagnostics.

Finally, although the state of stock assessment science has improved considerably (see the outputs of the CAPAM workshop series; Maunder et al., 2014, 2016, 2017; Sharma et al., 2019; Cadrin et al., in press), more work is needed to decide on the guidelines for best practices.

8.7. Is there an almost next-generation stock assessment package

It is likely that the development and testing of the next-generation stock assessment package will take at least three years. Each of the issues identified in Table 2 is likely to be an important need for stock assessments by the time the package is available. However, the state-of-the-art for some of the issues is more stable than for others, as illustrated in Table 1, for example because previous CAPAM workshops have identified candidate best practices. Developers of the next-generation stock assessment package should recognize that some of the issues will remain in flux for several years and account for this when developing the package.

None of the focus packages nor three other integrated analysis packages that have seen use in Europe (a4a), the US northeast (ASAP), and US southeast (BAM) have all of the features required (in our opinion) of the next-generation stock assessment package (Table 1), although the key feature of any next-generation stock assessment package is its modular nature and hence ability to add new features. GADGET has the most flexibility of the packages in relation to features that can be included (in particular it already includes age-length structure, spatial structure and multiple species). However, GADGET can be very slow to run because it is not coded such that the gradients of the objective function with respect the parameters are computed automatically, nor is it able to compute the marginal likelihood over random effects. CASAL, Casal2 and Stock Synthesis also have many features of the next-generation stock assessment package, but they also cannot compute the marginal likelihood over random effects. GADGET, Stock Synthesis, CASAL and Casal2 allow for multiple stocks and, along with MULTIFAN-CL, have some capacity to include spatial effects. Of the five packages referred to earlier, SAM is the only one that can treat random effects correctly (although WHAM also treats random effects correctly and some of the other packages can implement them using Bayesian methods). Thus, while many of the pieces needed to construct the next-generation stock assessment package exist, much work remains and should be the focus of the stock assessment community for the next several years.

Finally, it should be recognized that many assessments are constrained more by a lack of appropriate data than by the package to analyse the data, and a new package will not increase the amount of data available. However, the hope is that a next-generation stock assessment package can make best use of the data and provide managers with the best possible information on which to base management decisions. It should also help determine what data should be collected to improve the assessment.

Author credit statement

AP and MM developed the concept of the paper; AP created the initial draft and structure; All authors contributed to writing and editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Malcolm Haddon, L. Richard Little, and Geoff Tuck (CSIRO), Owen Hamel and Ian Taylor (NOAA, NWFSC), Richard McGarvey (SARDI), the guest editor, and an anonymous reviewer are thanked for their comments on the draft of this paper. This publication was partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement NA15OAR4320063, Contribution No. 2020-1066.

References

Allen Akselrud, C., Punt, A.E., Cronin-Fine, L., 2017. Exploring model structurey uncertainty using a general stock assessment framework: the case of Pacific cod in the Eastern Bering Sea. Fish. Res. 193, 104–120.
Albertsen, C.M., Nielsen, A., Thygesen, U.H., 2017. Connecting single-stock assessment models through correlated survival. ICES J. Mar. Sci. 75, 235–244.
Anderson, S.C., Monnahan, C.C., Johnson, K.F., Ono, K., Valero, J.L., 2014a. ss3sim: an R package for fisheries stock assessment simulation with Stock Synthesis. PLoS One 9, e92725.
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., Goethel, D.R., Lynch, P.D., 2017a. Introduction to “Space oddity: the mission for spatial integration1. Can. J. Fish. Aquat. Sci. 74, 1693–1697.
Berger, A.M., Goethel, D.R., Lynch, P.D., Quinn, T., Mormede, S., McKenzie, J., Dunn, A., 2017b. Space oddity: the mission for spatial integration1. Can. J. Fish. Aquat. Sci. 74, 1698–1716.
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.
Bravington, M.V., Skaug, H.J., Anderson, E.C., 2016a. Close-kin mark-recapture. Stat. Sci. 31, 259–274.
Bravington, M.V., Grewe, P.M., Davies, C.R., 2016b. Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. Nat. Commn. 7, 13162.
Bunnefeld, N., Hoshino, E., Milner-Gulland, E.J., 2011. Management strategy evaluation: a powerful tool for conservation? Trends Ecol. Evol. 26, 441–447.
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., Andrew, P.A., 1984. Dynamic catch-effort models for the hake stocks in ICSEAF Divisions 1.3-2.2. Colln Scient. Pap. Int. Commn SE. Atl. Fish 11, 29–58.
Butterworth, D.S., Punt, A.E., 1992. Assessments of the East Greenland-Iceland fin whale stock. Rep. Int. Whal. Comm. 42, 671–696.
Butterworth, D.S., Hughes, G., Strumpfer, F., 1990. VPA with ad hoc tuning: implementation for disaggregated fleet data, variance estimation, and application to the Namibian stock of Cape horse mackerel. *Trachurus trachurus capensis*. S. Afr. J. Mar. Sci. 327–357.
Byrd, J.M.R., Jarvis, S.A., Bhalerao, A.H., 2010. On the parallelisation of MCMC by speculative chain execution. 2010 IEEE International Symposium on Parallel Distributed Processing, Workshops and Phd Forum (IPDPSW) 1–8. https://doi.org/10.1109/IPDPSW.2010.5470689.
Cadigan, N., 2016. A state-space stock assessment model for northern cod, including under-reporting catches and variable natural mortality rates. Can. J. Fish. Aquat. Sci. 73, 296–308.
Cadrin, S.X., 2020. Defining spatial structure for fishery stock assessment. Fish. Res. 221, 105397.
Cadrin, S.X., Kerr, L.A., Mariani, S. (Eds.), 2014. Stock Identification Methods: Applications in Fisheries Science, 2nd ed. Elsevier Academic Press.
Cadrin, S.X., Maunder, M.N., Punt, A.E., In press. Spatial Structure: Theory, estimation and application in stock assessment models. Fish. Res. 00, 00–00.
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.

Conn, P.B., Williams, E.H., Shertzer, K.W., 2010. When can we reliably estimate the productivity of fish stocks? Can. J. Fish. Aquat. Sci. 67, 1–13.
Cope, J.M., 2013. Implementing a statistical catch-at-age model (Stock Synthesis)as a tool for deriving overfishing limits in data-limited situations. Fish. Res. 142, 3–14.
Cope, J., Dick, E.J., MacCall, A., Monk, M., Soper, B., Wetzel, C., 2015a. Data-moderate stock assessments for Brown, China copper, sharpchin, stripetail, and yellowtail rockfishes and English and rex soles in 2013. PacificFishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220 (298 pp.).
Cope, J.M., Thorson, J.T., Wetzel, C.R., DeVore, J., 2015b. Evaluating a prior onrelative stock status using simplified age-structured models. Fish. Res. 171, 101–109.
Cunningham, K.M., Canini, M.F., Spies, I.B., Hauser, L., 2009. Genetic isolation by distance and local fjord population structure in Pacific cod (*Gadus macrocephalus*): limited effective dispersal in the northeast Pacific Ocean. Can. J. Fish. Aquat. Sci. 66, 153–166.
Curti, K.L., Collie, J.S., Legault, C.M., Link, J.S., 2013. Evaluating the performance of a multispecies statistical catch-at-age model. Can. J. Fish. Aquat. Sci. 70, 470–484.
Davies, C., Bravington, M., Thomson, R., 2015. Advice on Close-kin Mark-recapture for Estimating Abundance of Eastern Atlantic Bluefin Tuna: a Scoping Study. ICCAT Document GBYP 07c/2015. https://www.iccat.int/GBYP/Docs/Close_Kin_Mark_Recapture_Phase_6_Scoping_Study.pdf.
De Moor, C.L., Butterworth, D.S., van der Lingen, C.D., 2017. The quantitative use of parasite data in multistock modelling of South African sardine (*Sardinops sagax*). Can. J. Fish. Aquat. Sci. 74, 1895–1903.
Deriso, R.B., Quinn, T.J., Neal, P.R., 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42, 815–824.
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.
Doonan, I., Large, K., Dunn, A., Rasmussen, S., Marsh, C., Mormede, S., 2016. Casal2: new Zealand's integrated population modelling tool. Fish. Res. 183, 498–505.
Dorn, M.W., 2002. Advice on West Coast rockfish harvest rates from Bayesian meta-analysis of stock-recruit relationships. N. Am. J. Fish Manage. 22, 280–300.
Doubleday, W.G., 1976. A least square approach to analysing catch at age data. Int. Comm. Northwest Atl. Fish. Res. Bull. 12, 69–81.
Elvarsson, B.Þ., Woods, P.J., Bjornsson, H., Lentin, J., Thordarson, G., 2018. Pushing the limits of a data challenged stock: a size- and age-structured assessment of ling (Molva molva) in Icelandic waters using Gadget. Fish Fish. 207, 95–109.
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., 2014. Replacing the multinomial in stock assessment models: a first step. Fish. Res. 151, 70–84.
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.
Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J., 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim Methods Software 27, 1–17.
Fu, C., Fanning, L.P., 2004. Spatial considerations in the management of Atlantic cod off Nova Scotia. Canada. N. Am. J. Fish. Manage. 24, 775–784.
Goethel, D.R., Quinn, I.I., T. J, Cadrin, S.X., 2011. Incorporating spatial structure in stock assessment: movement modeling in marine fish population dynamics. Rev. Fish. Sci. Aquac. 19, 119–136.
Gudmundsson, G., 1994. Time series analysis of catch-at-age observations. Appl. Statist. 43, 117–126.
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.
Hampton, J., Fournier, D.A., 2001. Aspatially disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. Mar. Freshw. Res. 52, 937–963.
Hannesson, S., Jakobsdottir, A., Begley, J., Taylor, L., Stefansson, G., 2008. On the use of tagging data in statistical multispecies multi-area models of marine populations. ICES J. Mar. Sci. 65, 1762–1772.
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., Maunder, M.N., Parma, A., Ernst, B., Payne, J., Starr, P.J., 2000. Documentation for a General Age-structured Bayesian Stock Assessment Model: Code Named Coleraine. Fisheries Research Institute, University of Washington, FRI/UW 00/01. https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/4524/0116.pdf?sequence=1&isAllowed=y.
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., 2006. Status of yellowfin tuna in the eastern pacific Ocean in 2004 and outlook for 2005. Inter-American Tropical Tuna Commission, Stock Assessment Report 6, 3–102. Available at. https://www.iattc.org/PDFFiles2/StockAssessmentReports/SAR6/SAR6-YFT-ENG.pdf.
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.
ICES, 2013. World Conference on Stock Assessment Methods (WCSAM), 15-19 July 2013. Boston, USA. ICES CM 2013/ ACOM/SCICOM:02. http://ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/ACOM%20SCICOM/WCSAM13.pdf.
International Whaling Commission, 2014. Report of the working group in stock definition. Annex I to report of the scientific committee. J. Cetacean Res. Manage. (Supplement) 15, 271–288.
International Whaling Commission. In press. Report of the Sub-Committee on In-Depth Assessments. J. Cetacean Res. Manage. (Supplement) 00, 00–00.
Johnson, K.F., Anderson, S.C., Doering, K., Monnahan, C., Stawitz, C., Taylor, I., Cunningham, C., Hicks, A., Hurtado-Ferro.F, Kuriyama.P., Licandeo, R., McGilliard, C., Murdian, M., Ono, K., Rudd, M., Szuwalski, C., Valero, J., Whitten, A., 2019. ss3sim: Fisheries Stock Assessment Simulation Testing With Stock Synthesis. R Package Version 1.03. http://cran.r-project.org/package=ss3sim.
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.
Kerr, L.A., Cadrin, S.X., Secor, D.H., Taylor, N.G., 2017. Modeling the implications of stock mixing and life history uncertainty of Atlantic bluefin tuna. Can. J. Fish. Aquat. Sci. 74, 1990–2004.
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 approximation. J. Stat. Softw. 70, 1–21.
Kuriyama, P.T., Ono, K., Hurtado-Ferro, F., Hicks, A.C., Taylor, I.G., Licandeo, R.R., Johnson, K.F., Anderson, S.C., Monnahan, C.C., Rudd, M.B., Stawitz, C.C., Valero, J.L., 2016. An empirical weight-at-age approach reduces estimation bias compared to modeling parametric growth in integrated, statistical stock assessment models when growth is time varying. Fish. Res. 180, 119–127.
Laurec, A., Shepherd, J.G., 1983. On the analysis of catch and effort data. J. Cons. Int. Explor. Mer. 41, 81–84.
Lee, H.-H., Maunder, M.N., Piner, K.R., Methot, R.D., 2011. Estimating natural mortality within a fisheries stock assessment model. An evaluation using simulation analysis based on twelve stock assessments. Fish. Res. 109, 89–94.
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.
Legault, C.M., Restrepo, V.R., 1998. A flexible forward age-structured assessment program. Collect. Vol. Sci. Pap. ICCAT 49, 246–253.
Lynch, P.D., Methot, R.D., Link, J.S. (Eds.), 2018. Implementing a Next Generation Stock Assessment Enterprise. An Update to the NOAA Fisheries Stock Assessment Improvement Plan. U.S. Dep. Commer. NOAA Tech. Memo. NMFS-F/ SPO-183. https://spo.nmfs.noaa.gov/sites/default/files/TMSPO183.pdf.
MacCall, A.D., Klingbeil, R.A., Methot, R.D., 1985. Recent increased abundance and potential productivity of Pacific mackerel (*Scomber japonicus*). CalCOFI Rep. 26, 119–129.
Maunder, M.N., 1998. Integration of Tagging and Population Dynamics Models in Fisheries Stock Assessment. PhD Thesis. University of Washington.
Maunder, M.N., et al., 2001a. 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, Pp. 123–146. Alaska Sea Grant College Program Report, AK-SG-01-02, University of Alaska Fairbanks.
Maunder, M.N., 2001b. A general framework for integrating the standardization of catch-per-unit-of-effort into stock assessment models. Can. J. Fish. Aquat. Sci. 58, 795–803.
Maunder, M.N., 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: estimating the effective sample size. Fish. Res. 109, 311–319.
Maunder, M.N., 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. Fish. Res. 72, 7–18.
Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142, 61–74.
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., Thorson, J.T., Xu, H., Oliveros-Ramos, R., Hoyle, S.D., Tremblay-Boyer, L., Lee, H.H., Kai, M., Chang, S.K., Kitakado, T., Albertsen, C.M., Minte-Vera, C.V., Lennert-Cody, C.E., Aires-da-Silva, A.M., Piner, K.R., 2020. The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models. Fish. Res. 229, 105594.
Maury, O., Gascuel, D., 2001. 'Local overfishing' and fishing tactics: theoretical considerations and applied consequences in stock assessment studied with a numerical simulator of fisheries. Aquat. Liv. Res. 14, 203–210.
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.
McGarvey, R., Feenstra, J.E., 2002. Estimating rates of fish movement from tag recoveries: conditioning by recapture. Can. J. Fish. Aquat. Sci. 59, 1054–1064.
McGarvey, R., Feenstra, J.E., Ye, Q., 2007. Modeling fish numbers dynamically by age and length: partitioning cohorts into "slices". Can. J. Fish. Aquat. Sci. 64, 1157–1173.
McGarvey, R., Fowler, A.J., Feenstra, J.E., Fleer, D.A., Jones, G.K., 2003. King George Whiting (*Sillaginodes punctata*) Fishery. Fishery Assessment to PIRSA for the Marine Scalefish Fishery Management Committee. SARDI Aquatic Sciences Publication No.

RD03/0152 77 pp.
Methot, R.D., 1986. Synthetic estimates of historical abundance and mortality for northern anchovy, *Engraulis mordax*. NMFS Southwest Fisheries Sci. Center Admin. Rep. 12 86-29.
Methot, R.D., 1989. Synthetic estimates of historical and current biomass of northern anchovy, *Engraulis mordax*. Am. Fish. Soc. Sympos. 6, 66–82.
Methot, R.D., 2000. Technical Description of the Stock Synthesis Assessment Program. NOAA Tech Memo NMFS-NWFSC-43. p. 46. .
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., Legault, C.M., 2015. Technical Details for ASAP Version 4. Northeast Fisheries Science Center Reference Document 15–17. 136 pp. .
Miller, T.J., Harem, J.A., Alade, L.A., 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern New England yellowtail flounder. Can. J. Fish. Aquat. Sci. 73, 1261–1270.
Moffitt, E.A., Punt, A.E., Holsman, K., Aydin, K.Y., Ianelli, J.N., Oritiz, I., 2016. Moving towards ecosystem-based fisheries management: options for parameterizing multi-species biological reference points. Deep Sea Res. II. 134, 350–359.
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.
Paloheimo, J.E., 1980. Estimating mortality rates in fish populations. Trans. Am. Fish. Soc. 109, 378–386.
Plaganyi, E.E., Punt, A.E., Hillary, R., Morello, E.B., Thébaud, O., Hutton, T., Pillans, R.D., Thorson, J.T., Fulton, E.A., Smith, A.D.M., Smith, F., Bayliss, P., Haywood, M., Lyne, V., Rothlisberg, P.C., 2014. Models of intermediate complexity for ecosystem assessment to support tactical management decisions in fisheries and conservation. Fish Fish 15, 1–22.
Ono, K., Licandeo, R., Muradian, M.L., Cunningham, C.J., Anderson, S.C., Hurtado-Ferro, F., Johnson, K.F., McGilliard, C.R., Monnahan, C.C., Szuwalski, C.S., Valero, J.L., Vert-Pre, K.A., Whitten, A.R., Punt, A.E., 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES J. Mar. Sci. 72, 31–43.
Pope, J.G., Shepherd, J.G., 1982. A simple method for the consistent interpretation of catch-at-age data. J. Cons. Int. Explor. Mer. 40, 176–184.
Pope, J.G., Shepherd, J.G., 1985. A comparison of the performance of various methods for tuning VPAs using effort data. ICES J. Mar. Sci. 42, 129–151.
Pribac, F., Punt, A.E., Walker, T.I., Taylor, B.L., 2005. Using length, age and tagging data in a stock assessment of a length selective fishery for gummy shark *(Mustelus antarcticus)*. J. Northw. Atl. Fish. Sci. 35, 267–290.
Punt, A.E., 2017. Some insights into data weighting in integrated stock assessments. Fish. Res. 192, 52–65.
Punt, A.E., 2019a. Spatial stock assessment methods: a viewpoint on current issues and assumptions. Fish. Res. 213, 132–143.
Punt, A.E., 2019b. Modelling recruitment in a spatial context: a review of current approaches, simulation evaluation of options, and suggestions for best practices. Press. Fish. Res 217, 140–155.
Punt, A.E., Hilborn, R., 1997. Fisheries stock assessment and decision analysis: the bayesian approach. Rev. Fish Biol. Fish. 7, 35–63.
Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016b. 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., Haddon, M., Little, L.R., Tuck, G.N., 2016a. Can a spatially-structured stock assessment address uncertainty due to closed areas? A case study based on pink ling in Australia. Fish. Res. 175, 10–23.
Punt, A.E., Hakamada, T., Bando, T., Kitakado, T., 2014. Assessment of Antarctic minke whales using statistical catch-at-age analysis. J. Cetacean Res. Manag. 14, 93–116.
Punt, A.E., Huang, T.-C., Maunder, M.N., 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES J. Mar. Sci. 70, 16–33.
Punt, A.E., Pribac, F., Walker, T.I., Taylor, B.L., Prince, J.D., 2000. Stock assessment of school shark *Galeorhinus galeus* based on a spatially-explicit population dynamics model. Mar. Freshw. Res. 51, 205–220.
Punt, A.E., Smith, D.C., Smith, A.D.M., 2011. Among-stock comparisons for improving stock assessments of data-poor stocks – the “Robin Hood” approach. ICES J. Mar. Sci. 68, 972–981.
Punt, A.E., Smith, D.C., Tuck, G.N., Methot, R.D., 2006. Including discard data in fisheries stock assessments: two case studies from south-eastern Australia. Fish. Res. 79, 239–250.
Quinn, T.J.I.I., Deriso, R.B., Neal, P.R., 1990. Migratory catch-age analysis can. J. Fish. Aquat. Sci. 47, 2315–2327.
Rincon, M.M., Corti, R., Elvarsson, B.T., Ramos, F., Ruiz, J., 2019. Granger-causality analysis of integrated-model outputs, a tool to assess external drivers in fishery. Fish. Res. 213, 42–55.
Ross-Gillespie, A., 2016. Modelling Cannibalism and Inter-species Predation for the Cape Hake Species *Merluccius capensis* and *M. Paradoxus*. PhD Dissertation. University of Cape Town viii + 195pp.
Ross-Gillespie, A., Butterworth, D.S., Findlay, K., 2014. Assessment Results for Humpback Breeding Stocks D, E1 and Oceania Following Recommendations From SC 65a. IWC Document SC/65b/SH04Rev. 28pp.. .
Ross-Gillespie, A., Butterworth, D.S., Findlay, K., 2015. Final results for the final’ base case’ three-stock DSB, BSE1 and BSO model, with sensitivity runs. J. Cetacean Res. Manage 16 (Suppl), 215–221.
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.
Schnute, J.T., 1977. Improved estimates from the Schaefer production model: theoretical consideration. J. Fish Res. Board Can. 34, 583–603.
Sibert, J.R., 1984. A two-fishery tag attrition model for the analysis of mortality, recruitment and fishery interaction. Tech. Rep. 13. Tuna and Billfish Assessment Progr.,South Pacific Comm., Noumea. New Caledonia (27 pp).
Skaug, H.J., 2001. Allele-sharing methods for estimation of population size. Biometrics 57, 750–756.
Skaug, H.J., Fournier, D.A., 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Comp. Stat. Data Anal. 51, 699–709.
Stawitz, C.C., Haltuch, M.A., Johnson, K.F., 2019. How does growth misspecification affect management advice derived from an integrated fisheries stock assessment model? Fish. Res. 213, 12–21.
Szuwalski, C., 2019. A stock assessment for eastern bering Sea snow crab. Stock Assessment and Fishery Evaluation (SAFE) Current BSAI Crab SAFE (updated Through Sept 2019). . https://meetings.npfmc.org/CommentReview/DownloadFile?p=30a57738-eb23-40a7-9e82-ca415b36cdf1.pdf&fileName=C4%201%20Snow%20Crab%20SAFE%202019.pdf.
Taylor, I.G., Methot Jr., R.D., 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fish. Res. 142, 75–85.
Taylor, I.G., Stewart, I.J., Hicks, A.C., Garrison, T.M., Punt, A.E., Wallace, J.R., Wetzel, C.R., Thorson, J.T., Takeuchi, Y., Ono, K., Monnahan, C.C., Stawitz, C.C., A’mar, Z.T., Whitten, A.R., Johnson, K.F., Emmet, R.L., Anderson, S.C., Lambert, G.I., Stachura, M.M., Cooper, A.B., Stephens, A., Klaer, N.L., McGilliard, C.R., Iwasaki, W.M., Doering, K., Havron, A.M., 2019. R4ss. https://github.com/r4ss/.
Taylor, L., Begley, J., Kupca, L., Stefanssson, G., 2007. A simple implementation of the statistical modelling framework Gadget for cod in Icelandic waters. Afr. J. Mar. Sci. 29, 224–245.
Thorson, J.T., 2019. Perspective: let’s simplify stock assessment by replacing tuning algorithms with statistics. Fish. Res. 217, 133–139.
Thorson, J.T., 2020. Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. Fish Fish. 21, 237–251.
Thorson, J.T., Johnson, K.F., Methot, R.D., Taylor, I.G., 2017a. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. 192, 84–93.
Thorson, J.T., Munch, S.B., Cope, J.M., Gao, J., 2017b. Predicting life history parameters for all fishes worldwide. Ecol. Appl. 27, 2262–2276.
Thygesen, U.H., Albertsen, C.M., Berg, C.W., Kristensen, K., Nielsen, A., 2018. Validation of ecological state-space models using the Laplace approximation. Environ. Ecol. Stat. 24, 317–339.
Trenkel, V.M., Skaug, H.J., 2005. Disentangling the effects of capture efficiency and population abundance on catch data using random effects models. ICES J. Mar. Sci. 62, 1543–1555.
Trijoulet, V., Fay, G., Miller, T.J., 2020. Performance of a state-space multispecies model: What are the consequences of ignoring predation and process errors in stock assessments? J. Appl. Ecol. 57, 121–135.
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.
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., Aires-da-Silva, A., Bayliff, W.H., 2009. Evaluating fishery impacts: application to bigeye tuna (*Thunnus obesus*) in the eastern Pacific Ocean. Fish. Res. 99, 106–111.
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.
Wetzel, C.R., Punt, A.E., 2016. Evaluating the performance of data-moderate and catch-only assessment methods for U.S. West coast groundfish. Fish. Res. 171, 170–187.
Winker, H., Carvalho, F., Thorson, J.T., Kell, L.T., Parker, D., Kapur, M., Sharma, R., Booth, A.J., Kerwath, S.E., 2020. JABBA-Select: incorporating life history and fisheries’ selectivity into surplus production models. Fish. Es. 222, 105355.
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.
Ying, Y., Chen, Y., Lin, L., Gao, T., 2011. Risks of ignoring fish population spatial structure in fisheries management. Can. J. Fish. Aquat. Sci. 68, 2101–2120.
Zheng, J., Siddeek, M.S.M., 2019. Bristol Bay red king crab stock assessment in fall 2019. Stock Assessment and Fishery Evaluation (SAFE) Current BSAI Crab SAFE (updated Through Sept 2019). . https://meetings.npfmc.org/CommentReview/DownloadFile?p=10dc1405-9525-4b04-a797-28b9898957c8.pdf&fileName=C4%202%20BBRKC%20SAFE%202019.pdf.