A review of integrated analysis in fisheries stock assessment

Mark N. Maunder and Andre A. Punt

SEDAR39-RD-06

September 2014

Fisheries Research 142 (2013) 61–74

Contents lists available at SciVerse ScienceDirect

Fisheries Research

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

A review of integrated analysis in fisheries stock assessment

Mark N. Maunder[a,*], André E. Punt[b]

[a] Inter-American Tropical Tuna Commission, 8604 La Jolla Shores Drive, La Jolla, CA 92037-1508, USA
[b] School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA

ARTICLE INFO

Keywords:
Bayesian
Data assimilation
Fisheries stock assessment
Integrated analysis
Maximum likelihood
Meta-analysis
Multivariate nonlinear regression

ABSTRACT

Limited data, and the requirement to provide science-based advice for exploited populations, have led to the development of statistical methods that combine several sources of information into a single analysis. This approach, "integrated analysis", was first formulated by Fournier and Archibald in 1982. Contemporary use of integrated analysis involves using all available data, in as raw a form as appropriate, in a single analysis. Analyses that were traditionally carried out independently are now conducted simultaneously through likelihood functions that include multiple data sources. For example, the traditional analysis of converting catch-at-length data into catch-at-age data for use in an age-structured population dynamics models can be avoided by including the basic data used in this conversion, length-frequency and conditional age-at-length data, in the likelihood function. This allows for consistency in assumptions and permits the uncertainty associated with both data sources to be propagated to final model outputs, such as catch limits under harvest control rules. The development of the AD Model Builder software has greatly facilitated the use of integrated analyses, and there are now several general stock assessment models (e.g., Stock Synthesis) that allow many data types and model assumptions to be analyzed simultaneously. In this paper, we define integrated analysis, describe its history and development, give several examples, and describe the advantages of and problems with integrated analysis.

1. Introduction

The requirement to provide science-based advice for exploited populations in data-poor situations has forced analysts to be creative in how to use data sources efficiently. It has also led to the explicit representation and estimation of uncertainty so that the uncertainty can be taken into consideration in management decision making (Patterson et al., 2001). Initial attempts to achieve these goals were generally ad hoc, but substantial progress has been made in the field of fisheries stock assessment over the last 30 years (Quinn, 2003), driven by improvements in computers and computational algorithms (e.g. Gelman et al., 1995; Punt and Hilborn, 1997; Fournier et al., 2012). It is no longer necessary to have closed form solutions for parameter and variance estimates. Rather, computers can be used to estimate parameters for complex non-linear models, and Monte Carlo simulations or other methods to determine the level of uncertainty associated with these estimates (Maunder et al., 2009). Simulation analysis can be used to determine how well different models and estimation techniques perform when their assumptions are violated.

Integrated analysis (IA), the focus of this review, combines several sources of data into a single analysis. IA is essentially the construction of a joint likelihood for the observed data and can be used in any likelihood-based framework. IA, as used in fisheries stock assessment, was first formulated by Fournier and Archibald (1982). Contemporary use of IA involves using all available data, in as raw a form as appropriate, in a single analysis. Analyses that were traditionally carried out independently are now conducted simultaneously through joint likelihood functions (Maunder, 2004). Integrated analysis is referred to by various names depending on the field of research (weather forecasting, oceanography, hydrology) and even within the same field of research (Draper et al., 1993), i.e. data assimilation (Dorigo et al., 2007; Ghil and Malanotte-Rizzoli, 1991; Hobbs and Ogle, 2011; Zobitz et al., 2011), model-data melding, inverse modeling, automatic calibration, multivariate nonlinear regression (Porporato and Ridolfi, 2001), or ensemble estimation. IA methods similar to those used in fisheries stock assessment are now becoming popular in wildlife research (Besbeas et al., 2003; Buckland et al., 2007; Lebreton, 2005; Schaub et al., 2007; Schaub and Abadi, 2010). Integrated analysis differs from the past practice of analyzing each data type separately and comparing the results.

Most early (prior to 1990) and many contemporary stock assessment methods often used a two-step approach: (1) the raw data were summarized using one analysis, possibly as values for

---

* Corresponding author. Tel.: +1 858 546 7027; fax: +1 858 546 7133.
E-mail address: mmaunder@iattc.org (M.N. Maunder).

0165-7836/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fishres.2012.07.025

parameters that were then assumed known in subsequent analysis, and (2) the summarized data (or parameter values) from the first analysis were then used in the assessment model. For example, catch-per-unit-of-effort (CPUE) data can be standardized to account for factors other than total abundance (e.g. fishing method, season, or vessel size), using a general linear model (GLM) (e.g., Gavaris, 1980), and the standardized CPUE, represented by the year effect from the GLM, can be used as an index of relative abundance when conducting a stock assessment (Maunder and Punt, 2004). Link (1999) described this approach using the phrase "do statistics on the statistics". Similarly, estimates of growth parameters from the analysis of tagging data can be used in an age-structured stock assessment model that is fit to other data sources. However, this approach is subject to several concerns (Maunder, 1998):

(1) *Information may be lost when data are summarized*. For example, age-structured tagging data contain information on fishery selectivity, but this information is lost if the tagging data are used to estimate total absolute abundance, which is subsequently used in a stock assessment.
(2) *The two analyses may be logically inconsistent.* For example, the tagging model of Anganuzzi et al. (1994) used initial selectivity estimates from a prior catch-at-age analysis to estimate other parameters that were then used in the same catch-at-age analysis to re-estimate the selectivity parameters.
(3) *Difficulty in determining the appropriate likelihood function*. The likelihood function for summarized data may be more difficult to determine than that for the raw data. For example, Xiao (1998) suggested that it is difficult to represent the error structure of the ratio of catch divided by effort, and Maunder and Starr (2003) described the complications of using the year effect from a GLM-based standardized CPUE as an index of relative abundance. If the raw data are included in the likelihood function, the frequently encountered heterogeneity with hierarchical structure and clustered sampling can be explicitly modeled and evaluated.
(4) *Difficulty in fully accounting for uncertainty*. It is difficult to adequately represent uncertainty of summarized data when they are used in a likelihood function or as values for parameters. This is particularly the case if there is high correlation among the data/parameters, or the shape of the sampling distribution for the summarized data/parameters is unusual. The representation of correlation among data points/parameters often requires the use of joint distributions that are harder to represent than the marginal distributions that are commonly used in the likelihood functions on which stock assessments are based. The situation of under-represented uncertainty can be even more severe if parameter values are assumed known in the stock assessment. However, it should be noted that structural uncertainty in the assessment model is often substantial and may be more important to account for than estimation uncertainty conditioned on one model.
(5) *Reduced diagnostic ability*. Residuals (differences between observed data and the values predicted by the model) are often used to evaluate the fit of a model. If summarized data are used in the likelihood function, it may not be possible to determine if a lack of fit is due to how the data are summarized or how the summarized data can be mimicked by the population model. The use of raw data eliminates the summarization step allowing a more thorough evaluation of the cause of a lack of fit and provides more information to develop an appropriate solution to improve the fit.

IA forms the basis for several general stock assessment programs (e.g. Stock Synthesis (Methot and Wetzel, 2013), MULTFAN-CL (Fournier et al., 1998; Hampton and Fournier, 2001), A-SCALA (Maunder and Watters, 2003b), CASAL (Bull et al., 2005), Coleraine (Hilborn et al., 2000)) that are applied globally.

Despite the intuitive advantages of IA, it is not a panacea for poor quality data or model structure uncertainty in stock assessments. There are several disadvantages, mostly related to model misspecification, the complexity of the resulting models, and the associated, often considerable, computational requirements (e.g. the use of remotely sensed environmental information). Consequently, in some situations, the traditional two-step approach remains a better approach. Data analysis is an integral part of producing good, reliable stock assessments so there is value in analyzing each component data set before conducting an IA of all data simultaneously. Rather than deciding whether to apply IA or not, the issue is often how to determine the appropriate degree of data aggregation.

This paper reviews IA as it is used in fisheries stock assessment. We start by giving a brief history of IA in fisheries, then describe several examples of IA, highlight the advantages and disadvantages of IA, and finally draw some general conclusions.

2. History of fisheries stock assessment and integrated analysis

Table 1 describes the major historical stock assessment methods and the data they require (further details of these methods can be found in most fisheries texts, e.g., Hilborn and Walters, 1992; Quinn and Deriso, 1999; Haddon, 2001). These methods are generally based on one data source, and hence do not use all available data sources. For example, virtual population analysis, VPA, was often preferred when a long and continuous time-series of catch-at-age data was available. Subsequently, analysts have used multiple assessment methods and compared the results when there is more than one data source. For example, ICCAT (2000) compared the results from VPA, which uses catch-at-age data, with those from a surplus production model, which uses catch and effort data. Pella and Tomlinson (1969) used results from yield-per-recruit analysis, which uses biological information, to motivate the development and parameterization of the Pella–Tomlinson surplus production model for which the B_{MSY} does not have to occur at half of carrying capacity. Some of the methods in Table 1 are used to generate input for the other methods. For example, VPA can be used to calculate stock and recruitment data for stock-recruitment analysis and tagging growth analysis can be used to estimate growth parameters that are used in yield-per-recruit analysis. Despite some obvious connections among these methods, it has only been in the last 20 years that researchers have started regularly combining these data sources in analyses.

The use of multiple data sets simultaneously in fisheries assessment was developed in the context of catch-at-age analysis (Fournier and Archibald, 1982; Deriso et al., 1985; Gavaris, 1988; Methot, 1990a). Original VPA methods (Gulland, 1965; Pope, 1972) assumed that the catches-at-age were measured with negligible error, and did not use the information gained by analyzing multiple cohorts simultaneously. Doubleday (1976) developed a method that considered error in the catch-at-age data, and which fitted catch-at-age data sets for multiple cohorts simultaneously. By separating fishing mortality into age and year components, Doubleday (1976) reduced the number of free parameters so that there were fewer parameters than data points (under-determined). He also established that catch-at-age data alone are insufficient to estimate stock biomass reliably because biomass and fishing mortality are negatively correlated. In some early catch-at-age methods (e.g. separable VPA; Pope and Shepherd, 1982) the number of free parameters exceeded the number of data points unless additional constraints were added. Therefore, moderate amounts of

Table 1
The main types of analyses used historically and the data they require.

Catch-at-age data – Cohort analysis (Pope, 1972), and **Virtual Population Analysis** (Gulland, 1965), use a time series of catch-age-data to follow the catch taken from a cohort as it ages over time. The initial numbers in a cohort can be calculated as the sum of the catch from the cohort in conjunction with an estimate of the rate of natural mortality. Annual total population size is then calculated by summing over cohorts. There are several limitations to these methods, and it is well known that they are unreliable unless auxiliary information is used (Pope and Stokes, 1989; Rivard, 1989). **ADAPT-VPA** (Gavaris, 1988), XSA (Shepherd, 1999), and tuned methods (Laurec and Shepherd, 1983) are the most common ways to include auxiliary information into a catch-at-age analysis. The output (recruitment and biomass) is often used as input for stock-recruitment analysis (see below).

Catch and effort data – Surplus production models (Schaefer, 1954; Pella and Tomlinson, 1969) use a time series of catch and effort data to reconstruct the biomass trajectory of the population. These methods traditionally do not model age-structure, but model the changes in biomass as a simple function of biomass in the year before.

Stock size and recruitment data – Stock-recruitment analysis (Ricker, 1954; Beverton and Holt, 1957) attempts to estimate the relationship between stock and recruitment and requires estimates of both stock size and the amount of recruitment for a series of years.

Yield-per-recruit analysis (Beverton and Holt, 1957) uses information on natural mortality, growth rates, and fishery age-specific selectivity to determine optimal harvest rates and the optimal age at entry into a fishery based on an age-structured population dynamics model. The estimates from yield-per-recruit analysis ignore any stock-recruitment relationship and the resulting harvest rates are often compared to harvest rate estimates from, for example, virtual population analysis. Absolute estimates of biomass cannot be obtained from yield-per-recruit analysis.

Tagging data

Numbers released and recaptured – Petersen tagging analysis (Petersen, 1896) uses the number of tag releases, the number of individuals sampled, and the number of recoveries to estimate population size.

Length at release, length at recapture, and time at liberty – Tagging growth analysis (Fabens, 1965) uses growth increments calculated from tagging data to estimate growth rates. The growth estimates are often used as inputs into yield-per-recruit analysis.

auxiliary information, such as effort data or an assumption about the stock-recruitment relationship or selectivity-at-age, were needed for the problem to be under-determined. This additional information can also increase precision and reduce bias in the resulting model estimates (Deriso et al., 1985). Fishing effort is the most common auxiliary data source included in catch-at-age analyses, and is used to index fishing mortality. Paloheimo (1980) assumed that fishing mortality is effort multiplied by a constant parameter. However, Fournier and Archibald (1982) considered a deterministic relationship between fishing mortality and fishing effort unrealistic because catchability fluctuates over time, so they included process error in the effort-fishing mortality relationship. Doubleday (1976) assumed, effectively, that the error associated with the catch-at-age data was normal. Fournier and Archibald (1982) suggested that a multinomial likelihood function was more appropriate because it is based on sampling properties.

Many catch-at-age analyses now integrate diverse auxiliary information (e.g., Fournier and Archibald, 1982; Deriso et al., 1985; Gavaris, 1988; Methot and Wetzel, 2013), such as catch, effort, the age or length compositions from the fishery and any surveys, and abundance indices. Additional research has been directed at developing appropriate objective and likelihood functions (Maunder, 2011), including the form of the objective function (e.g. Crone and Sampson, 1998), weighting factors (e.g., McDonald et al., 2001; Deriso et al., 2007), and how to make the estimation robust to outliers (e.g., Fournier et al., 1998).

As noted above, data included in assessments are often preprocessed; the raw data are first summarized in a separate analysis, often in the form of parameter estimates, and then used when fitting the population dynamics. For example, CPUE data are standardized (e.g., Gavaris, 1980), length-increment data from tagging experiments are used to calculate growth-transition matrices for length-structured models (Punt et al., 1997), and catch-at-length data are converted into catch-at-age data (Fournier et al., 1990). Recently, Bayesian posteriors (see below) or likelihood profiles from the analysis used to preprocess the raw data have been used as priors in the procedure used to fit population dynamics models to propagate uncertainty (Punt and Hilborn, 1997; McAllister and Kirkwood, 1998). Unfortunately, to fully use the information, it is necessary to have a multivariate prior when the raw data pertain to multiple parameters, which can be difficult to construct. In many applications therefore, only the marginal priors or likelihood profiles are used, and correlation among parameters is ignored. It is also difficult to represent the shape of a joint prior using easily-available distributional forms. Fitting to the raw data overcomes these difficulties (Maunder, 2003a). Even when data are not included in their rawest form, it is becoming standard practice to fit data that have been disaggregated to their component forms. For example, it is now common to fit age-structured models to length-frequency and conditional age-at-length data rather than to age-composition data constructed by multiplying length-frequencies and age-length keys (e.g., Punt et al., 2006).

Further development of methods that integrate analyses has only occurred to a substantial extent during the last 15 years. One of the first involved the inclusion of length-frequency data in stock assessment models. Originally, catch-at-length data would be converted into catch-at-age data (e.g. MULTIFAN, Fournier et al., 1990) and the catch-at-age data would be used as input in a catch-at-age analysis. Fournier et al. (1998) combined these processes into one analysis (MULTIFAN-CL). Similar analyses were conducted to integrate tagging data (Richards, 1991; Maunder, 1998, 2001b; Hampton and Fournier, 2001; see review in Goethel et al., 2011), environmental correlations (see summary in Maunder and Deriso, 2011), stock-recruitment relationships (Fournier and Archibald, 1982; Smith and Punt, 1998), and CPUE standardization (Xiao, 1998; Maunder, 2001a). Bentley et al. (2001) integrated estimating the growth transition matrix using growth increment data obtained from tagging into a fully length-based population dynamics model. Several general fisheries stock assessment models integrate multiple types of data. General integrated models have also been developed for marine mammals (e.g., Breen et al., 2003) and for including fine-scale environmental data into spatial fish population models (Senina et al., 2008). Integrated analysis has also been used when fitting multi-species models (Jurado-Molina et al., 2005; Kinzey and Punt, 2009).

Bayesian methods have become popular in stock assessment modeling (Punt and Hilborn, 1997; McAllister and Kirkwood, 1998). In addition to transferring uncertainty from one analysis to another, Bayesian methods can use information that is not specific to the stock being assessed, such as that from other stocks and species, or from expert opinion. This information is usually included as prior probability distributions for the model parameters. Bayesian estimation has been used with a variety of assessment models, including the equilibrium Schaefer model (Hoenig et al., 1994), the dynamic Schaefer model (McAllister and Kirkwood, 1998), the Pella–Tomlinson model (Maunder et al., 2000), VPA (Walters and Punt, 1994), stock reduction analysis (McAllister et al., 1994), delay difference models (Meyer and Millar, 1999), statistical catch-at-length models (Punt and Kennedy, 1997), and statistical catch-at-age models (McAllister and Ianelli, 1997;

Maunder and Starr, 2001; Parma, 2001). Methods used to generate prior distributions include demographic models (Maunder, 2003b; McAllister et al., 2001), meta-analyses (Hilborn and Liermann, 1998; Myers et al., 1999), and expert judgment (IWC, 1995). Prior distributions, at least for some parameters, can be assumed to be "uninformative" (Punt and Hilborn, 1997). Structural uncertainty can also be accounted for within the Bayesian paradigm (Sainsbury, 1988; Patterson, 1999; Parma, 2001; Brandon and Wade, 2006; see Hoeting et al., 1999 for a description of Bayesian model averaging).

3. Types of integrated analysis

In this review we focus solely on integrated analyses that use statistical approaches to estimate parameters. First, we discuss methods used to integrate multiple data sources, including prior probability distributions and meta-analysis, into assessments. Then, we describe several different examples of integrated analysis.

3.1. Integrating data

'Statistical' stock assessments have three components: (1) the population dynamics model, including how process error is modeled; (2) the data and the associated observation model, including the likelihood function, which represents the sampling process; and (3) the estimation algorithm. Bayesian and related methods also include priors or penalties. The population dynamics model reconstructs population size and structure (e.g., age structure), the observation model relates to how the data pertain to the model predictions, including the sampling distribution for the data, and the estimation algorithm is used to provide point estimates, confidence intervals, samples from posterior distributions, etc. Additional information is included in a 'statistical' assessment through additional terms in the objective (or likelihood) function, but may also require changes to the population dynamics model to ensure that it is able to predict model quantities that can be linked to the available information. Maximum likelihood and Bayesian methods have become the standard in model fitting because they allow statistically rigorous weighting of data sets, and there are standard methods for quantifying uncertainty. Specifically, several data sources are included in the analysis by taking the product of the likelihoods (or equivalently summing the negative log likelihoods) for each data set (under the assumption that each data source is independent).

The three components can be demonstrated using the simple example of fitting a surplus production model to an annual index of relative abundance (I_y), using maximum likelihood. First, a surplus production model (e.g., the Pella–Tomlinson (1969) model) is used to represent the dynamics of the population and to calculate annual biomass (B_y) given the model parameters, $\underline{\theta}$. Second, the sampling distribution, which is used to define the likelihood function, for the index of abundance for year y would be a distribution with mean (or median) given by the product of biomass and the catchability coefficient q, while, the distribution itself could one of several common forms (e.g. normal, log-normal, gamma); for simplicity, we select normal with standard deviation σ. The observation model (and hence likelihood function) is therefore:

$$L(I|\underline{\theta}) = \prod_y \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(I_y - qB_y)^2}{2\sigma^2}\right) \tag{1}$$

The quantities that are independent of the free parameters of the model (σ is often assumed known) can be left out of the likelihood function because they provide no information about the parameters of the model, although such constants cannot be ignored if σ is estimated, uncertainty is estimated, or if information theoretic methods are to be used to compare alternative functional forms for the error distribution (e.g. normal vs. log-normal). Maximizing the likelihood is then equivalent to minimizing the following quantity (note the relationship with least squares):

$$\sum_y (I_y - qB_y)^2 \tag{2}$$

The values for the parameters are then estimated using some optimization algorithm (e.g., Fournier et al., 2012).

The above example shows how multiple data points from the one data set can be integrated into an assessment. The same approach can be used to integrate multiple data sets. For example, if an estimate of absolute abundance for 1995 was available from tagging data (T_{1995}), and assumed to be normally distributed with variance ν^2, the combined negative log-likelihood (less constants independent of the model parameters) would be (note that, unlike Eq. (2), the variance terms are included in Eq. (3) because they are needed to appropriately weight the different likelihood components):

$$-\ell nL(I, T_{1995}|\underline{\theta}) = \frac{(T_{1995} - B_{1995})^2}{2\nu^2} + \sum_y \frac{(I_y - qB_y)^2}{2\sigma^2} \tag{3}$$

Thus, two analyses that were traditionally conducted separately can be integrated. The analyses are combined by ensuring that the parameters and model structure representing the same processes are identical and summing the negative log-likelihoods (assuming independence) for each data set. For example, the negative log-likelihood when combining two analyses with data sets D_1 and D_2 and parameters $\underline{\theta}_1$ and $\underline{\theta}_2$ is:

$$-\ln L(D_1, D_2|\underline{\theta}) = -\ln L(D_1|\underline{\theta}_1) - \ln L(D_2|\underline{\theta}_2) \tag{4}$$

and the parameters are estimated by:

$$\min_{\underline{\theta}}[-\ln L(D_1, D_2|\underline{\theta})] \quad \underline{\theta} \in R^m \tag{5}$$

where $\underline{\theta} = \underline{\theta}_1 \cup \underline{\theta}_2$ and has length m.

3.2. Data weighting

Although it is straightforward in principle to include multiple data sources in an analysis, a major challenge is how to weight each data source (Francis, 2011). For the most widely used distributions in IA, the weighting factors are the sampling standard deviations (σ in Eqs. (1) and (2)) or the effective sample sizes for binomial and multinomial sampling distributions. The effective sample size for a data set (e.g., catch-at-age) may be much smaller than the actual sample size (e.g., number of fish measured) because sampling is not truly random or the model ignores key sources of process uncertainty such as inter-annual variation in selectivity (Francis, 2011). Analysts often examine the sensitivity of the outcomes of an assessment to each data source by removing them one at a time from the analysis or changing the (pre-specified) weights. Deriso et al. (1985) found that their catch-at-age model applied to Pacific halibut was robust to moderate changes to the weighting factors, but this is not always the case (Crone and Sampson, 1998). If the different data sets are consistent, there should be little sensitivity to changes in the weighting factors, except, perhaps, reduced precision or if one data set contains the only information for a certain parameter. Patterns in the residuals should be examined to identify violations of the model assumptions that might indicate that the model is misspecified. Model misspecification can be in the population dynamics model, the observation model, or the parameter values assumed known. If the reason that two data sets are contradictory cannot be determined, Schnute and Hilborn (1993) suggest using a likelihood function that is the average of those for each data source so that the posterior distributions (or likelihood profiles) for some of the parameters will be bimodal, rather than unimodal. The approach of Schnute and Hilborn (1993) can be generalized to assign probabilities to each data source being correct, i.e. $L = pL_1 + (1-p)L_2$ where p is the probability that data set 1 is correct, L_1 is the likelihood for data set 1 and L_2 is the likelihood for data set 2.

Pre-specifying the factors which weight data can interact with model selection. For example, setting the effective sample sizes for the age composition too large can result in model selection criteria selecting unnecessarily complex selectivity functions. The values for some of the weighting factors (such as σ in Eqs. (1) and (2)) can be treated as estimable parameters (Deriso et al., 2007) while the effective sample sizes for age or length composition data are often "tuned" by comparing the assumed input sample sizes with the variance of the Pearson residuals about those data. However, when an assessment includes contradictory data, this can lead to the parameter estimation process "choosing" one data source (e.g. because the variance of the likelihood function is estimated to be low) over another (e.g. because the variance of the likelihood function is estimated to be high), usually a series which is relatively long and exhibits the low inter-annual variation.

The variance of the residuals between the observed and predicted values should be compared to evaluate the statistical assumptions on which the observation model is based. If the variance of the residuals differs substantially from that which is assumed, the weighting factors in the objective functions are not likely appropriate.

3.2.1. *Outliers*

Data from natural populations often have more extreme values than would be expected from standard statistical theory (Fournier et al., 1990). In simple applications, these extreme values can be removed from the analysis to determine their influence on the results. Unfortunately, this is often not feasible due to the complexity of IA, and the large amount of data typically used. Therefore, automatic methods are needed to reduce the influence of extreme values, such as modifying the likelihood function. This can be achieved by assuming that the data are contaminated and come from one of two distributions, one that describes the majority of the data and a heavier tailed distribution that describes the outliers (Fournier et al., 1990). A simple approach is to assume that the contamination distribution occurs with probability $1-p$ and is uniformly distributed between two bounds (Maunder, 2001c).

$$L_{robust}(D|\underset{\sim}{\theta}) = pL(D|\underset{\sim}{\theta}) + (1-p)\psi \tag{6}$$

Eq. (6) is similar to the mixture model of Francis (1988), where ψ is a constant that ensures that the sampling distribution integrates to one. Also note the similarity with the previously mentioned method of Schnute and Hilborn (1993) to deal with contradictory data. Fournier et al.'s (1990) robust likelihood function based on the normal distribution simply adds a constant 0.01, which ensures that the influence of observations reduces rapidly as their distance from the predicted value becomes greater than about three standard deviations:

$$L_{robust}(I|\underset{\sim}{\theta}) = \prod_y \frac{1}{\sqrt{2\pi}\sigma}\left[\exp\left(-\frac{(I_y - qB_y)^2}{2\sigma^2}\right) + 0.01\right] \tag{7}$$

Simulation analysis has shown that robust likelihood functions perform only slightly worse than their non-robust equivalents when there are no outliers, but perform substantially better when outliers are present (Chen et al., 2000).

3.3. *Integrating prior information*

"Most stock assessments use no quantitative information derived from previous experience on other fish stocks, and often fix parameter values at some 'best' estimates rather than admit uncertainty in their value" (Hilborn and Liermann, 1998). However, it is straightforward to include 'experience' into a stock assessment. For example, the decision about which model structure is most appropriate (e.g., a Ricker vs. a Beverton–Holt stock recruitment curve) is usually based on previous experience. Prior information can be integrated in IA in a manner similar to how data are included (Hilborn and Mangel, 1997, Chapter 9). The difference is that integration of data comprises individual data points and a likelihood function, whereas prior information usually involves a probability distribution for a parameter of the population dynamics or observation model. The prior can be the result of a compilation of estimates for many stocks of the same or similar species, or simply based on expert judgment. For example, a prior distribution for a parameter, θ, could be normally distributed with mean μ and standard deviation σ:

$$P(\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right) \tag{8}$$

The objective function would then be based on the product of the likelihood and prior distributions (*sensu* a Bayesian analysis), i.e. $L(D|\underset{\sim}{\theta})P(\underset{\sim}{\theta})$. The negative log-likelihood objective function when there is an index of abundance and a normal prior on catchability (with mean μ_q and standard deviation σ_q) would be (ignoring constants):

$$-\ell n[L(D|\underset{\sim}{\theta})P(\underset{\sim}{\theta})] = \frac{(q-\mu_q)^2}{2\sigma_q^2} + \sum_y \frac{(I_y - qB_y)^2}{2\sigma^2} \tag{9}$$

The parameter estimates that minimize Eq. (9), which would be maximum likelihood estimates (MLEs) if there was no prior information, are variously referred to as posterior modal (e.g. Smith and Punt, 1998; Maunder and Starr, 2001), maximum a posteriori (Parma, 2001), highest posterior density (Schnute, 1994), or penalized likelihood estimates. Priors are usually associated with Bayesian methods (see Punt and Hilborn, 1997), but it is uncommon for IA to be based on full Bayesian integration owing to computational demands. Rather, in practice Eq. (9) is often treated as a negative log-likelihood and standard frequentist methods are applied.

A recent development is the use of approximate likelihoods to transfer information from one analysis to the next (Besbeas et al., 2003). Information from a previous analysis is summarized in the form of an approximate likelihood and this approximate likelihood is then included in the next analysis. The log-likelihood for the first analysis (based on data $\mathbf{Y}_A$) is approximated by the logarithm of a multivariate normal distribution for the parameters, based on the maximum likelihood estimates for the parameters, $\hat{\underset{\sim}{\theta}}$, and the estimated dispersion matrix, $\hat{\Sigma}$. This approximate likelihood is then combined with the likelihood of the remaining data set ($\mathbf{Y}_B$) in a second analysis, which also includes any additional parameters $\underset{\sim}{\phi}$, i.e. the negative log likelihood is:

$$-\ell nL(\mathbf{Y}|\underset{\sim}{\theta},\underset{\sim}{\phi}) = -\ell nL(\mathbf{Y}_B|\underset{\sim}{\theta},\underset{\sim}{\phi}) + 0.5(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})'\hat{\Sigma}^{-1}(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}) \tag{10}$$

This approximation to the joint likelihood is analogous to using data-based priors in a Bayesian analysis, where the prior is a multivariate normal approximation of the posterior from the first analysis.

3.4. *Meta-analyses*

Biological data are often highly imprecise, and it is often impractical to improve the precision for a single stock by collecting additional data. However, data from multiple stocks can be used in a single analysis to increase the precision of the estimates if there are similarities among stocks of the same or similar species (Myers and Mertz, 1998). These methods assume that any bias caused by sharing information among stocks is small in comparison to the increase in precision. In addition, they can be used to create prior distributions. There are several meta-analysis methods, ranging from combining independent estimates of parameters from multiple stocks to the use of Bayesian hierarchical methods (Davidian and Giltinan, 1995).

A naive meta-analysis can be carried out as follows: (1) estimate the parameters for each available data set; and (2) calculate summary statistics of the parameter estimates (i.e. mean and variance). This two-stage method will provide adequate results when the variability in parameter values among data sets is small, the sample size for each data set is large, and there are many data sets. However, many applications will have large random errors or small sample size for at least some of the data sets. The problems with this method are: (1) data sets that contain little information have the same weight as data sets that contain large amounts of information because only point estimates are used (i.e. no account is taken of the precision in the estimates, Liermann and Hilborn, 1997); (2) variability from observation error is not differentiated from variability among data sets, so the distribution may be broader than the true distribution (Liermann and Hilborn, 1997); and (3) the estimates for each data set do not use information from the other data sets (i.e. the two-stage method does not shrink the individual estimates toward the mean). In contrast, hierarchical meta-analysis assumes that the values for a parameter for different stocks are drawn from a common distribution and accounts for the precision of the available data. Meta-analysis of stock-recruitment data is the most common application in fisheries (e.g. McAllister et al., 1994; Liermann and Hilborn, 1997; Myers and Mertz, 1998) although other applications have addressed, for example, whether CPUE is linearly proportional to abundance (Harley et al., 2001) and estimated length-specific catchability (Harley and Myers, 2001).

In principle, meta-analyses could be conducted in which full age-structured models for each stock are analyzed together (e.g., Punt et al., 2011) because this allows the uncertainty from estimating recruitment and biomass to be automatically included in the meta-analysis. Jiao et al. (2011) used this approach to share parameters of a surplus production model among shark stocks.

4. Examples of IA

4.1. Length frequency and catch-at-age analysis

MULTIFAN (Fournier et al., 1990) is a likelihood-based method to estimate catch-at-age from catch-at-length data. It predicts the length-frequency data as the sum of normal distributions; one distribution for each age class, which represents the length distribution for that age class. Each normal distribution is calculated from three parameters; the mean length, the standard deviation of the lengths about the mean length, and the proportion that that age comprises of the total length-frequency data set. The parameters are estimated to make the predicted length-frequencies most consistent with the observations, as quantified by a likelihood function. The mean length-at-age is often represented by the von Bertalanffy growth curve and the standard deviation of length-at-age by a linear relationship with mean length. In MULTIFAN, the proportions for each age class are estimated as free parameters. The estimated catches-at-age, which are represented by age proportions, can then be used in a catch-at-age analysis.

MULTIFAN-CL (Fournier et al., 1998) integrated the MULTIFAN method with an age-structured model. In MULTIFAN-CL, the proportions for each age-class are calculated from an age-structured model, rather than being estimated as free parameters. This ensures that these proportions are consistent from year to year, because they are constrained by a population dynamics model. The likelihood function includes the length frequency data as well as catch and effort data (and any other data). The parameters of the age-structured model (e.g., annual recruitment and selectivity parameters), the mean lengths-at-age or the parameters of the growth curve, and the standard deviations of the lengths-at-age are estimated simultaneously by maximizing the combined likelihood. This ensures that the uncertainty associated with estimating catch-at-age is propagated into the stock assessment.

4.2. Integrating catch-at-age analysis with a stock recruitment analysis

Stock-recruitment analysis has traditionally been carried out after time-series of spawning biomass and recruitment have been estimated using another estimation procedure. Integrating the stock-recruitment relationship into the stock assessment is a clear demonstration of the benefits of IA (Ianelli, 2002). There is often inadequate information to reliably estimate recruitment for every year. Therefore, an assumption is often made that annual recruitment is distributed (usually lognormally) about a mean (e.g., Smith and Punt, 1998; Maunder and Deriso, 2003; Butterworth et al., 1990). This is essentially treating recruitment as a random effect, except historically it has been impractical to integrate out the random effects. If there is little information about recruitment strength for a particular year, it will be estimated to be the average, independent of stock size. The uncertainty associated with this estimate will be high, but this uncertainty is usually not included in a stock-recruitment analysis carried out post estimation (see Dichmont et al. (2003) for an exception).

In contrast, the recruitment for a year with few data will have no influence on the stock- recruitment relationship if the stock-recruitment analysis is integrated into the stock assessment by including a penalty on the difference between each recruitment and the value expected given the stock-recruitment relationship (e.g., Fournier and Archibald, 1982; Smith and Punt, 1998; Maunder and Watters, 2003b). In addition, the estimated recruitment for years with few data will be based on the stock–recruitment relationship, rather than average recruitment. This is particularly the case for the most recent poorly-selected cohorts for which data are usually insufficient to reliably estimate year-class strength.

4.3. Integrating growth estimation into a length-based model

The estimation of growth using tagging data can be integrated with a length-based population dynamics model so that information on growth can be obtained from both the length-frequency and tagging data. Both tagging data (e.g. Punt et al., 1997) and length frequencies (e.g. Fournier et al., 1990) have been used to estimate growth parameters, but many length-based analyses have ignored the information on growth in length-frequency data. Traditionally, tagging data are first used to estimate a size-transition matrix (Punt et al., 1997), and this matrix is then used in a length-based model that is fit to length-frequency data (Punt and Kennedy, 1997; Starr et al., 1999). Bentley et al. (2001) integrated the tagging data into a length-based stock assessment by assuming that the variation in growth rates is normally distributed around a von Bertalanffy growth curve. They also assumed that the integral over the sizes in a length-class, which represents growth into that length-class, is approximated by the average growth rate at the midpoint of the length-class to simplify the computations. An alternative, and arguably better, method is to model the probability of moving from one size class to the other size classes as a multinomial distribution (Punt et al., 2010). Other considerations include the modeling of measurement error in the release and recovery data (Maunder, 2001c), the impossibility of negative growth, and the binomial probability of not growing (e.g. to model molting in crabs).

4.4. Integrating tagging and catch-at-age analysis

Tagging studies are commonly used to estimate stock size, natural mortality, age-specific selectivity, fishing mortality, and movement. These are important parameters for stock assessment models. However, only recently has raw tagging data been incorporated into stock assessments. In a review of integrated tagging models, Goethel et al. (2011) stated: “One of the biggest breakthroughs in stock assessment modeling in recent years has been the development of “tag-integrated assessment models,” which include a tag-recapture component to the objective function.” Richards (1991) investigated sharing parameters between a tagging model and a simple stock assessment model for a stock of lingcod, *Ophiodon elongatus*, off Canada. Punt and Butterworth (1995) and Porch et al. (1995) integrated tagging data into catch-at-age analyses with spatial structure, but the tagging data were aggregated over time and age, and the model was fitted to the ratio of recaptures in the two areas. Haist (1998) used temporally disaggregated tag data and modeled the age- and sex-structure of the tagged fish based on the age- and sex-composition of the population in the modeled population at the time of tagging, but did not use the size- or age-structured information from the tag recovery data in the likelihood function. Punt et al. (2000) integrated age-structured tagging data into a stock assessment for school shark, *Galeorhinus galeus*, off southern Australia, but did not use catch-at-age data. Maunder (1998, 2001b) developed a general integrated tagging and catch-at-age analysis (ITCAAN), that explicitly included movement and fits to both age-structured tagging and catch-at-age data. Maunder (1998) applied this model to the snapper, *Pagrus auratus*, stock off the northeast coast of New Zealand. A similar method was developed by Hampton and Fournier (2001) for a catch-at-length analysis, which was applied to yellowfin tuna, *Thunnus albacares*, in the western and central Pacific Ocean. Catch-at-length analysis may be more appropriate for tagging data because it is much easier to measure the length of all the fish released than to age them. Maunder (2001b) showed that integrating tagging data into a catch-at-age analysis can improve the precision of estimates for both productivity and depletion levels. Tagging data can now be included in several general stock assessment models (e.g. Stock Synthesis; Methot and Wetzel, 2013). Recently, data from archival tags have been integrated into stock assessments (e.g., Taylor et al., 2009) and hence the multiple location recordings available from such data are used to provide additional information about movement.

Integrating tagging data into a stock assessment is relatively straightforward if there is no spatial structure, but becomes more complicated if individuals move among areas. When individuals move among areas, each tag release group must be modeled using the equations for the total population. The tagged population is assumed to have the same dynamics and parameters as the total population, with the only difference being that ‘recruitment’ to the tagged population occurs through releases, and can therefore occur at any age (e.g. McGarvey et al., 2010). Each set of releases (a release group) that can be uniquely identified in the recoveries (e.g., different area/time of release) is modeled as a separate ‘population’. The most common likelihood function used for recaptures is the multinomial, because it is suitable for modeling multiple outcomes (e.g. caught in years x, y, z, or not caught at all). There may be more variation in the recaptures than described by the multinomial, and other likelihood functions may be more appropriate. For example, Hampton and Fournier (2001) used the negative binomial. Many extensions to this approach are possible; for example, adjustments for different mixing rates of tagged with untagged fish, reporting rates, tag loss, and mortality due to tagging and handling (Hampton, 1997; Hampton and Fournier, 2001).

4.5. Integrating the standardization of CPUE data

The traditional approach to include fishery CPUE data into stock assessments involves the two-step approach outlined above. These two steps can be combined into one integrated analysis by including the raw CPUE data in the likelihood function and representing the catchability coefficient as a linear model which depends on parameters such as area and month (e.g., Xiao, 1998). The parameters of the population dynamics model and those which determine the catchability parameters are then estimated simultaneously, thereby integrating the standardization process with fitting the population dynamics model. This method can also be used to include raw survey data in an assessment.

Maunder (2001a) outlines a general method to integrate the standardization of CPUE into stock assessments, and Maunder and Langley (2004) provide an example using a statistical catch-at-age analysis. Maunder (2001a) showed, using simulation analysis, that the integrated approach leads to narrower confidence intervals that more accurately represent the uncertainty in the parameter estimates, but notes that the reasons why integrating the CPUE data improved performance were not clear, and that the generality of the results for models more complex than the production model used in that study was not certain. The most probable reason for the improved results is the automatic inclusion of the variance and covariance among year effects into the likelihood function. Integrating the CPUE standardization into the stock assessment model could be made more computationally efficient by using analytical solutions for the regression coefficients. Alternatively, the two-step approach could be modified along the lines of Eq. (10) so that information is not lost and the uncertainty is appropriately captured.

4.6. Integrating correlations with environmental data

A temperature-recruitment relationship (Francis, 1993) and a pre-recruit index were integrated into the population dynamics model for New Zealand SNA1 snapper, *Pagrus auratus* (Maunder and Starr, 2001) so that recruitment estimates could be based on these data sources as well as catch-at-age data. Inclusion of the environmental data directly into the stock assessment allows all the available data to inform the relationship (if any) between recruitment and temperature, captures the uncertainty in that relationship, and allows the relationship to inform recruitment for years for which there are insufficient data to estimate recruitment (*sensu* a prior). Maunder and Watters (2003a) showed, using simulation analysis, that this integrated approach led to parameter estimates with lower bias and variance than correlating temperature and recruitment. However, Haltuch and Punt (2011) suggest that the benefits of integrating environmental data with decadal-scale variability directly into stock assessments via the stock–recruitment relationship is outweighed by high type I error rates due to fishing-induced stock declines that coincide with directional environmental change. They do, however, suggest that the impact of type I errors can be minimized by choosing an appropriate combination of assessment method and reference points.

Maunder and Watters (2003a) describe a general method based on penalized likelihood for integrating environmental data into population dynamics models that is an improvement over that used by Maunder and Starr (2001) because it allows for the estimation of process error in the environmental relationship. A random effects framework, (e.g., Maunder and Deriso, 2003; Fournier et al., 2012) would be even more appropriate, but this is considerably more computationally demanding. Stock Synthesis (Methot and Wetzel, 2013) allows most estimable parameters to be linked to environmental covariates such as temperature, but restricts the user to a limited choice of parameterizations that are not necessarily suitable.

Which of the available environmental variables should be included in an assessment can be selected using stepwise hypothesis testing based on, for example, likelihood ratio tests or Akaike's (1973) Information Criterion, AIC (Maunder and Watters, 2003a). Density-dependence can moderate the effects of environmental covariates and it is important to correctly identify the timing of when environmental factors influence the population with respect to the timing of density-dependence processes and available data (Maunder and Deriso, 2011). The correct modeling of observation and process variability is also important for hypothesis testing. If process variability is not modeled, likelihood ratio and AIC-based tests are biased towards incorrectly including environmental covariates in the assessment (Maunder and Watters, 2003a) while incorrect sampling distribution assumptions (e.g., assumed values for the observation error variance) can influence the covariate selection process, and the weighting given to each data set can change which covariates are chosen (Deriso et al., 2007). Missing covariate data needs to be dealt with appropriately, such as by using the random effects methods described in Gimenez et al. (2009) and Maunder and Deriso (2010).

5. Advantages and disadvantages of the integrated approach

5.1. Advantages

We outlined five problems with the traditional two-step procedure to stock assessment in the introduction. These problems can be overcome by IA because it allows all of the information to be included in the assessment in as raw a form as appropriate. Equally importantly, an IA ensures consistency of assumptions among analyses since it is unlikely that the analysis is going to model the same processes differently in the same model. For example, unlike traditional assessments that use Petersen estimates of population size, integrated tagging and catch-at-age analysis (Maunder, 1998, 2001b) uses the same model for the tagged and total populations. Furthermore, the error structure can be based on the statistical properties of the sampling process, rather than on the assumed statistical properties of summarized data.

IA also more appropriately captures uncertainty provided that the model is not mis-specified and key parameters are not pre-specified. For example, an integrated stock recruitment model includes the uncertainty in the stock size and recruitment estimates when estimating the relationship between stock and recruitment. However, uncertainty may not necessarily always be larger when applying IA. For example, the integrated CPUE standardization model leads to narrower and more accurate confidence intervals than a two-step procedure (Maunder, 2001a) because it automatically includes correlations among year factors.

Finally, integrating all of the data into one analysis makes it easier to investigate the sensitivity to each data set or assumption. Turning one component on or off, or down-weighting it, can be used to investigate sensitivity of the results of an assessment to different data sets and assumptions, illuminating the assumptions that are important. Changing weights also improves the ability to detect inconsistencies among data sets and hence model-misspecification. IA can therefore be used to guide future research and data collection that will differentiate among various hypotheses and improve management advice.

5.2. Disadvantages

Despite the intuitive advantages of IA, there are several problems that must be overcome: (1) computational demands; (2) convergence; (3) parameter confounding; (4) model misspecification; (5) model selection; (6) weighting among data sets; (7) level of data abstraction; and (8) transparency.

5.2.1. Computational demands

IA is more computationally intensive than traditional methods because there are usually more data and more parameters (Maunder et al., 2009). In addition, IA usually requires nonlinear parameter optimization, rather than analytical solutions. The increased computational time is not a serious problem if only a single result is needed. However, methods that take a long of time to produce a solution have several limitations:

(1) Development time is greatly increased. A model is usually run many times during the developmental phase of an assessment to check for programming errors and other problems. Therefore, small data sets must be used to test the model, but it is important that all the characteristics that are in the full data set are included in the smaller data set. We check our programs (usually programmed in AD Model Builder, Fournier et al., 2012) by repeating the calculations in other software such as Microsoft Excel or R. Excel is ideal for this because it is conceptually different from a procedural programming language, so it is unlikely that the same errors will occur in both versions of the model. However, integrated models are often so large that only parts of them can be programmed in a single Excel workbook.

(2) It is important to carry out simulation analyses to evaluate the performance of a new method and to determine how robust it is to its main assumptions. The number of assumptions that can be tested is limited because simulation analysis requires several, usually hundreds, of simulations using different data sets. For some applications of IA, simulation tests may not even be feasible.

(3) It is important for any application to test the sensitivity of the results to all of the important assumptions made in the analysis. However, if the model takes a long time to run, researchers will be able to perform only a limited number of sensitivity tests.

(4) Calculation of measures of precision may be impractical or limited to being based on approximations that are unsatisfactory. Methods such as bootstrap, likelihood profile, or Bayesian integration require the likelihood be evaluated many times, which may be infeasible, and the researcher will be forced to use normal approximations based on the inverse of the Hessian matrix.

(5) The inclusion of parameter uncertainty in forward projections may be impractical or limited to approximations that are unsatisfactory. Similar to the calculation of measures of precision, the methods used to include parameter uncertainty in forward projections require that the model to be evaluated many times. Many applications based on IA do not include parameter uncertainty, and only recruitment variation (e.g. Maunder and Watters, 2001). Others use approximations using the normal distribution based on the Hessian matrix (e.g., Punt et al., 2000; Maunder et al., 2006). MacCall (2013) presents a delta method approach to include uncertainty due to parameters that are typically fixed in stock assessments.

5.2.2. Convergence

A further computational issue is that efficient nonlinear function optimizers or Bayesian integration techniques are needed to implement IA. The automatic differentiation method that supplies machine precision derivatives to the function minimizer has shown promise for these types of analyses, providing efficient and more stable solutions compared to methods that use numerical derivatives (Schnute et al., 1998). The most useful method for Bayesian integration has been Markov chain Monte Carlo (MCMC, Punt and Hilborn, 1997). The AD Model Builder software (Fournier et al., 2012), which includes automatic differentiation and MCMC integration, has been very useful for IA and has driven much of the recent advances in fisheries stock assessment. However, due to their complex nature, it is difficult to determine whether reasonable estimates have been obtained for integrated models. Even, if an accurate optimizer is used, it is often not clear whether a local, rather than the global solution, has been found. Using multiple starting points can provide support that the global solution has been found if all starting points result in the same answer. However, it is often difficult to determine which parameter values should be changed and by how much when selecting alternative starting values when there are many parameters. In addition, this does not guarantee that the global solution has been found. More complex estimation methods, such as genetic algorithms or simulated annealing, could be used but are slow.

Often the estimation procedure cannot find a feasible solution. Reparameterization can sometimes help convergence by removing elongation or curvature in the likelihood surface (Seber and Wild, 1989). However, it is often difficult to determine which parameters should be reparameterized in complicated models. A good start is to make sure that all component models of the integrated model have good convergence behavior. David Fournier introduced the concept of model sculpting in the analysis of integrated models. Model sculpting is the procedure of directing the solution by using constraints and estimation phases (Fournier et al., 2012). The estimation procedure is divided into a number of phases. In each phase, the algorithm optimizes the objective function by estimating a subset of the parameters. The subset of parameters is increased in each sequential phase until the last phase in when all parameters are estimated. It is usually best to estimate the ‘major parameters’ (e.g., those that scale the biomass) in the first few phases and the ‘minor parameters’ or those for which good guesses can be made (e.g. selectivity) in the final few phases. The method of phases is equivalent to methods that find good starting points for the optimization process. Constraints are put on the model in the first few phases to guide the estimation procedure to the parameter space that is biologically or intuitively reasonable. These constraints might then be removed in the last few phases of the algorithm. The goal is to prevent the estimation scheme from getting stuck in a part of the parameter space that is not feasible. An example of a constraint is to make the average fishing mortality equal to 0.1 in the initial phases. Note that this is different from a prior, which is typically used for a single parameter and kept for the entire estimation procedure. The constraint, which may affect all parameters simultaneously, is removed in the final phases.

Finally, programming ability is very important when developing integrated models. The computational problems faced in IA are similar to those faced when computers were first used for quantitative fisheries analysis. Due to the large amounts of data and the complex models, there are many calculations to perform and many values to store. Programs must be coded so that memory usage is minimized and that redundant calculations are avoided. Matrix algebra available in many programming environments can be much more efficient than using loops (Fournier et al., 2012). Parallel processing could be used to reduce the runtime if the model can be parallelized.

5.2.3. Parameter confounding

Confounding of parameters becomes a common problem as models become more complex. Confounding occurs when two parameters are highly correlated, and there is insufficient information in the data to separate the effects of each parameter. Often, only a combination of the parameters, rather than the individual parameters, can be estimated. For example, the total combined recruitment of multiple interacting subpopulations is often well determined in an integrated tagging and catch-at-age analysis, but the individual subpopulation recruitments are not because they are confounded with movement (Maunder, 2001b). Confounding between parameters can lead to unrealistic parameter estimates, but this confounding can be differentiated from model misspecification (discussed below) by the fact that confidence intervals for confounded parameters should be large. Covariance estimates, joint confidence intervals, or joint posteriors can be used to identify which parameters are confounded (i.e. they will show high correlation). Confounding of parameters may cause problems for the non-linear function optimizer, so that a solution cannot be found.

Parameter confounding may not, however, be as important as suggested by the estimated extent of uncertainty of the parameter estimates. Often management quantities are not affected by parameter confounding. For example, estimates of MSY from a Schaefer production model may be reliable even if r and K are confounded because MSY is $rK/4$. The relationship is not as clear with the complex models used for IA, and likelihood profiles or Bayesian posteriors are needed to determine the effect of parameter confounding on management quantities such as MSY. As noted above, reparameterization can sometimes remove parameter confounding (Seber and Wild, 1989).

5.2.4. Model misspecification

Model misspecification and the large amounts of detailed data included in integrated analysis may interact to create apparently precise estimates that are obviously wrong. Model misspecification occurs when an influential process (including error structure) is left out of the model, a process is incorporated into the model incorrectly, or an incorrect value for a fixed parameter is used. If the model is misspecified, the parameters will be estimated in a way that compensates for the model misspecification (Piner et al., 2011), and may lead to the parameter estimates to be inconsistent with prior information.

Unfortunately, model misspecification appears to be a common problem when complex models are used with large amounts of detailed data. For example, Hampton (1991) found that estimates of area-specific natural mortality differed by implausibly large amounts when using an unconstrained tagging model. Using a two-area model for Atlantic bluefin tuna, *Thunnus thynnus*, which integrated summarized tagging data with a virtual population analysis, Punt and Butterworth (1995) estimated reporting rates for the two areas that differed by unrealistic amounts. Finally, the application of an integrated tagging and catch-at-age model to the snapper stock off the northeast coast of New Zealand indicates model misspecification because the initial fishing mortality rate was precisely estimated at unrealistically low levels, which is contradictory to prior information about historical catch being substantial before the modeling time period started (Maunder, 1998).

In principle, model misspecification can be detected by residual patterns where patterns in the residuals can be indicative of a problem with some of the specifications of the assessment. However, lack of residual patterns is not a guarantee that there is no model misspecification because, as noted above, the parameters may be estimated to compensate for misspecification. Moreover, residual patterns tend to occur when there are multiple sources of information for essentially the same model component which may be the case in data-rich situations, but is unlikely to be the case in data-poor situations. A particular form of model misspecification occurs when a parameter is set to a pre-specified value (based on auxiliary analyses or expert judgment). Setting the value of a parameter incorrectly often leads to biased assessment outcomes as well as residual patterns. However, it is often hard, if not impossible, to detect which parameters are incorrectly specified in models with many pre-specified parameters. Pre-specifying parameters will also tend to lead to under-estimation of uncertainty. It is common practice when conducting stock assessments using IA to pre-specify parameters which are poorly determined by the data, such as the rate of natural mortality, M, and the steepness of the stock–recruitment relationship, using auxiliary information of expert judgment. This should, however, be done with care because some of the key outputs from a stock assessment such as the estimate of the fishing mortality at which maximum sustainable yield is achieved are determined primarily by these parameters. Moreover, fixing these key parameters can lead to large under-estimation of uncertainty (MacCall, 2013).

Sensitivity analysis is often used to investigate the effect of model misspecification, but it is of little use to management unless probabilities can be assigned to the different sets of parameters or model structures. However, sensitivity analysis may provide information about which data should be collected to help correct the model misspecification and indicate what model development is needed.

5.2.5. *Model selection*

Model selection is an integral part of modeling (Burnham and Anderson, 1998; Hilborn and Mangel, 1997). Models should be developed so that, if practical and if it does not interfere with estimability, important components of the model can be eliminated by setting a parameter to a fixed value. Likelihood ratio tests, AIC, Bayes Information Criterion (BIC, Schwarz, 1978), and Bayes factors (Aitkin, 1991) have been used to select among integrated models (e.g. Fournier et al., 1998; Helu et al., 2000; Maunder and Watters, 2003b). Unfortunately, the application of these techniques is complicated by the large amount of data and many data types.

Burnham and Anderson (1998, pp. 96–99) showed that hypothesis testing ($\alpha = 0.05$) selects underfitted models that may be biased and have confidence intervals with low coverage, and suggested that hypothesis tests should not be used for model selection. AIC and Bayes factors usually accept additional parameters irrespective of whether they really improve the fit to the data because of the large amount of data in typical IAs. However, including additional unnecessary parameters often does not change the results (e.g., Maunder and Watters, 2003b). BIC and the sample size corrected AIC (AICc) may be more appropriate for many applications of IA because they account for the number of data points, and are therefore less likely to accept additional parameters. Unfortunately, unlike examples in text books, it is difficult to determine the number of data points to use when applying BIC to IA. For example, is a CPUE data point equivalent to a catch-at-age data point, or should the number of catch-at-age data points be the number of fish aged or, as in Helu et al. (2000), the number of age-classes?

Methods to determine the appropriateness of proposed model selection methodology are simulation tests or cross validation. Helu et al. (2000) performed some simple tests using Stock Synthesis, and found that AIC and BIC performed equally well at selecting the correct model. However, the model on which their study was based was not as complex as many integrated analyses. Based on simulation studies, Burnham and Anderson (1998) suggest that AIC performs better than BIC because BIC selects underfitted models.

Integrated models tend to be complicated, and aim to fully capture uncertainty. However, an alternative modeling philosophy is that simpler methods are better because they provide more precise estimates and are robust to several uncertainties, even if this is at the expense of a small amount of bias. For example, Ludwig and Walters (1985) showed that estimators based on age-structured models do not necessarily provide more accurate estimates than estimators based on more simple models, even if the true population is represented by an age-structured model. The result that management procedures (methods for determining management actions) based on simple estimation schemes either outperform, or are no worse than, management procedures based in complex models and estimation procedures has been shown often (e.g., Punt, 1993; De Oliveira et al., 1998; Punt and Smith, 1999). However, it should be recognized that the aim of a management procedure is to achieve management goals without necessarily trying to estimate current status accurately, while most stock assessment aim to estimate such status (and perhaps form the basis for evaluating management procedures). Thus, while there should always be a desire for parsimonious models, care needs to be taken to ensure that the aims of modeling exercises are understood.

5.2.6. *Data weighting*

It is necessary to weight each data set, but this can be a non-trivial exercise because the results from an assessment are often highly sensitive to data weights. It is important to investigate the influence of each data set. In many applications, the evaluation of the sensitivity to different data sets is achieved by leaving each data set out of the analysis in turn and comparing the results to those obtained while including the data set. However, it is often impossible to estimate all the model parameters if a data set is left out of the analysis. Therefore, the more appropriate way to explore sensitivity is to downweight data sets by increasing the variance or reducing the sample size in the corresponding likelihood component. If the variance of the residuals does not agree with the weighting assumption, this indicates that the weighting assumption is incorrect or there is some other problem with the model. Variance parameters can be treated as estimable (e.g., Deriso et al., 2007). However, Francis (2011) recommends down-weighting age and length data to ensure that the model fits the abundance indices adequately.

5.2.7. *Level of data aggregation*

One decision that is required in the development of an integrated model is what level of data aggregation should be used. Should the raw data be used, increasing the number of data points and computation time, or, to reduce the computational time, should the data be summarized by some characteristic that is not important for the analysis? For example, if trawl CPUE data are being standardized for time, area, and month, should the data be summarized by these strata or should each trawl set be used as a data point? If the data are summarized by stratum, the analysis ignores the within-strata variance, and strata with only a few data points will have the same weight as strata with many data points, unless the number of data points is used to weight the data in the likelihood function. The level of data aggregation will differ among applications, and is still an open question that should be focus for future research.

5.2.8. *Transparency*

Transparency in the stock assessment process is of key importance, particularly when the assessment is controversial and has substantial fishery management implications. In such cases, it is critical to have stakeholder buy-in for decisions. Undue complexity in the number of assertions, assumptions, and choices made in the assessment may be counterproductive. One potential problem with integrated analysis in such cases is that it may be perceived as being too complex or too-poorly described to be well understood and documented. In this context, compartmentalizing some of the analyses, such as CPUE standardization, as part of an incremental step-wise analysis can be important for a successful stock assessment process because there are fewer perceived and actual interactions and dependencies among the component analyses.

6. Integrated analysis and Stock Synthesis

Stock Synthesis (Methot and Wetzel, 2013) is an example of an integrated method. It allows simultaneous use of several data sources, such as relative and absolute indices of abundance including estimates of spawning biomass, age, length, age-length, and weight composition data, tagging data, and mean size and weight data. In addition, prior distributions can be placed on all estimable parameters. Furthermore, being based on AD Model Builder (Fournier et al., 2012), maximum posterior density estimates are available, and results can be expressed as draws from posterior distributions. Finally, Stock Synthesis also includes the ability to explore many different model configurations, such as the functional forms for selectivity and growth, and whether account should be taken of space and the impact of environmental effects on the values for parameters. Stock Synthesis has an associated R package (r4ss) which allows rapid visualization of outputs and fit diagnostics. However, although it is very flexible, Stock Synthesis is unable to implement all of the types of integrated analyses outlined above. For example, it is not possible to include raw CPUE data and integrate the standardization with the population dynamics model.

7. General considerations and future work

IA essentially involves the creation of a joint likelihood for all available observed data and attempts to include the data in as raw a form as appropriate. IA is intuitively superior to the two-stage approach because it allows the inclusion of all information from the data and the propagation of uncertainty. Nevertheless, after considering the advantages and disadvantages of IA, the obvious question is: should an IA be used? As with any situation, it depends on the circumstances. In our opinion, IA should be used when: (a) the integrated model does not take considerably longer than a two-step procedure and is within the constraints of the available computing resources; (b) uncertainty is important, but it is not easily transferred from one analysis to another; and (c) there is additional information in the data used in one analysis about parameters in the second analysis that would be lost without integration.

Several packages have been developed to implement IA for fisheries assessment. Packages such as Stock Synthesis and CASAL, which have many users, have generally been well tested (both using simulations and by applying them to data sets which have been analyzed using other assessment frameworks). Given that a general stock assessment package such as Stock Synthesis may consist of 10s of thousands of lines of code, careful validation and testing is essential. Packages can be constraining by only accommodating a subset of the types of data which could be used for assessment purposes, by often having very complicated input files to allow biological and observation processes to be specified, and by not including all possible functional forms among model components. For example, it is not straightforward within the context of Stock Synthesis to allow for a non-linear relationship between a model parameter and an environmental covariate, even though such relationships are not implausible. Software implementing an IA for a specific problem can overcome these issues (and run faster). However, testing of single-application IAs cannot be as extensive as that of well-tested and widely used packages such as Stock Synthesis, with the consequence that several single or limited application IAs (e.g. A-SCALA; Maunder and Watters, 2003b) are being replaced by packages such as Stock Synthesis.

There are several areas for additional research: (a) further development of estimation routines based on automatic differentiation and numerical integration, particularly so that process errors (e.g. temporal variability) in recruitment and selectivity can routinely be treated as random effects; (b) expansion of existing packages to more fully represent the available data; (c) development of protocols to identify when model mis-specification is occurring and identify what aspect of the model is leading to this problem; (d) development of methods to analyze multiple stocks and species simultaneously; and (e) further simulation evaluation of the performance of methods based on different levels of data aggregation. A collection of examples comparing IA with non-integrated approaches should be developed to help illustrate the types of applications that would benefit from the IA. The comprehensive use of data in IA can facilitate the determination of data needs in terms of both the type and the quantity of data. IA is directly applicable to multi-species and ecosystem models, and therefore can contribute to ecosystem based management.

Finally, access to data sets is required to most effectively use IA. There is a vast range of data types that can be included in IA. These data sets are often collected and analyzed by different researchers and, without easy access to the raw data, it is difficult to apply IA. This is especially true for meta-analysis, which involves data from different stocks or species. Databases should be constructed that contain all available data (e.g. the RAM Legacy extensive data base on stock assessment data (Ricard et al., in press) and the CPB/NCEAS Global Population Dynamics Database http://cpbnts1.bio.ic.ac.uk/gpdd). However, to use these data appropriately, the analyst should understand how the data were collected and not just blindly use them (Hampton et al., 2005).

Acknowledgements

William Bayliff, Richard Deriso, Shelton Harley, James Ianelli, Robert Olson, Ana Parma, and two anonymous reviewers and the Guest Editor, Alec MacCall, are thanked for their comments on an earlier version of this paper. AEP was partially supported for this work under NOAA grant NA10OAR4320148.

References

Aitkin, M., 1991. Posterior Bayes factors. J. R. Soc. Ser. B 53, 111–142.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), 2nd International Symposium on Information Theory. Publishing house of the Hungarian Academy of Sciences, Budapest, pp. 268–281.

Anganuzzi, A.A., Hilborn, R., Skalski, J.R., 1994. Estimation of size selectivity and movement rates from mark-recovery data. Can. J. Fish. Aquat. Sci. 51, 734–742.

Bentley, N., Breen, P.A., Starr, P.J., Kendrick, T., 2001. Assessment of the CRA3 and NSS Stock of Red Rock Lobster (Jasus edwardsii) for 2000. New Zealand Fisheries Assessment Research Document 2001/69. MAFFish Fisheries Research Centre, Wellington.

Besbeas, P., Lebreton, J.D., Morgan, B.J.T., 2003. The efficient integration of abundance and demographic data. Appl. Stat. 52, 95–102.

Beverton, R.J.H., Holt, S.J., 1957. On the Dynamics of Exploited Fish Populations, Fisheries Investigations Series 2, vol. 19. U.K. Ministry of Agriculture, London, 533 pp.

Brandon, J., Wade, P.R., 2006. Assessment of the Bering–Chukchi–Beaufort Seas stock of bowhead whales using Bayesian model averaging. J. Cet. Res. Manage. 8, 225–239.

Breen, P.A., Hilborn, R., Maunder, M.N., Kim, S.W., 2003. Effects of alternative control rules on the conflict between a fishery and a threatened sea lion (*Phocarctos hookeri*). Can. J. Fish. Aquat. Sci. 60, 527–541.

Buckland, S.T., Newman, K.B., Fernández, C., Thomas, L., Harwood, J., 2007. Embedding population dynamics models in inference. Stat. Sci. 22, 44–58.

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

Burnham, K.P., Anderson, D.R., 1998. Model Selection and Inference: A Practical Information-theoretical Approach. Springer Verlag, New York, 353 pp.

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. 9, 327–357.

Chen, Y., Breen, P.A., Andrew, N.L., 2000. Impacts of outliers and mis-specification of priors on Bayesian fisheries-stock assessment. Can. J. Fish. Aquat. Sci. 57, 2293–2305.

Crone, P.R., Sampson, D.B., 1998. Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition. In: Funk, F., Quinn II, T.J., Heifetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.J., Sullivan, P.J., Zhang, C.I. (Eds.), Fishery Stock Assessment Models. Alaska Sea Grant College Program Report No. AK-SG-98-01. University of Alaska Fairbanks, Fairbanks, Alaska, pp. 355–370.

Davidian, M., Giltinan, D.M., 1995. Nonlinear Models for Repeated Measurement Data. Chapman and Hall, London, 359 pp.

De Oliveira, J.A.A., Butterworth, D.S., Johnston, S.J., 1998. Progress and problems in the application of management procedures to South Africa's major fisheries. In: Funk, F., Quinn II, T.J., Heifetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.J., Sullivan, P.J., Zhang, C.I. (Eds.), Fishery Stock Assessment Models. Alaska Sea Grant College Program Report No. AK-SG-98-01. University of Alaska Fairbanks, Fairbanks, Alaska, pp. 513–530.

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

Deriso, R.B., Maunder, M.N., Skalski, J.R., 2007. Variance estimation in integrated assessment models and its importance for hypothesis testing. Can. J. Fish. Aquat. Sci. 64, 187–197.

Dichmont, C.M., Punt, A.E., Deng, A., Dell, Q., Venables, W., Dichmont, C.M., Punt, A.E., Deng, A., Dell, Q., Venables, W., 2003. Application of a weekly delay-difference model to commercial catch and effort data in Australia's Northern Prawn Fishery. Fish. Res. 65, 335–350.

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. Int. J. Appl. Earth Observ. Geoinf. 9, 165–193.

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

Draper, D., Gaver, D., Goel, P., Greenhouse, J., Hedges, L., Morris, C., Tucker, J., Waternaux, C., 1993. Combining Information: Statistical Issues and Opportunities for Research. Contemporary Statistics Series No. 1. American Statistical Association, Alexandria, 234 pp.

Fabens, A.J., 1965. Properties and fitting of the von Bertalanffy growth curve. Growth 29, 265–289.

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

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

Fournier, D.A., Sibert, J.R., Majkowski, J., Hampton, J., 1990. MULTIFAN a likelihood-based method for estimating growth parameters and age-composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (*Thunnus maccoyii*). Can. J. Fish. Aquat. Sci. 47, 301–317.

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. Opt. Met. SOFT 27, 233–249.

Francis, R.I.C.C., 1988. Maximum likelihood estimation of growth and growth variability from tagging data. N.Z. J. Mar. Freshwat. Res. 22, 42–51.

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

Francis, M.P., 1993. Does water temperature determine year class strength in New Zealand snapper (*Pagrus auratus*, Sparidae)? Fish. Ocean. 2, 65–72.

Gavaris, S., 1980. Use of a multiplicative model to estimate catch rate and effort from commercial data. Can. J. Fish. Aquat. Sci. 37, 2272–2275.

Gavaris, S., 1988. An adaptive framework for the estimation of population size. Can. Alt. Fish. Sci. Adv. Comm. Res. Doc. 88/29.

Gelman, A., Carlin, B.P., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall, London, 526 pp.

Ghil, M., Malanotte-Rizzoli, P., 1991. Data assimilation in meteorology and oceanography. Adv. Geophys. 33, 141–266.

Gimenez, O., Bonner, S., King, R., Parker, R.A., Brooks, S., Jamieson, L.E., Grosbois, V., Morgan, B.J.T., Thomas, L., 2009. WinBUGS for population ecologists: Bayesian modeling using Markov chain Monte Carlo methods. Environ. Ecol. Stat. 3, 883–916.

Goethel, D.R., Quinn II, T.J., Cadrin, S.X., 2011. Incorporating spatial structure in stock assessment: movement modeling in marine fish population dynamics. Rev. Fish. Sci. 19, 119–136.

Gulland, J.A., 1965. Estimation of mortality rates. Annex to Arctic Fisheries Working Group Report (meeting in Ham burg, January 1965). Internation Council for the Exploitation of the Sea, C.M. 1965, Document 3 (mimeo), Copenhagen.

Haddon, M., 2001. Modelling and Quantitative Methods in Fisheries. Chapman and Hall, New York, 450 pp.

Haist, V., 1998. Integrated catch-age mark-recapture model: application to B.C. sablefish stocks. In: Funk, F., Quinn II, T.J., Heifetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.J., Sullivan, P.J., Zhang, C.I. (Eds.), Fishery Stock Assessment Models. Alaska Sea Grant College Program Report No. AK-SG-98-01. University of Alaska Fairbanks, Fairbanks, Alaska, pp. 469–494.

Haltuch, M.A., Punt, A.E., 2011. The promises and pitfalls of including decadal-scale climate forcing of recruitment in groundfish stock assessment. Can. J. Fish. Aquat. Sci. 68, 912–926.

Hampton, J., 1991. Estimation of southern bluefin tuna *Thunnus maccoyii* natural mortality and movement rates from tagging experiments. Fish. Bull. 89, 591–610.

Hampton, J., 1997. Estimates of tag-reporting and tag-shedding rates in a large-scale tuna tagging experiment in the western Tropical Pacific Ocean. Fish. Bull. 95, 68–79.

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

Hampton, J., Sibert, J.R., Kleiber, P., Maunder, M.N., Harley, S.J., 2005. Decline of Pacific tuna populations exaggerated? Nature, E1–E2.

Harley, S.J., Myers, R.A., 2001. Hierarchical Bayesian models of length-specific catchability of research trawl surveys. Can. J. Fish. Aquat. Sci. 58, 1569–1584.

Harley, S.J., Myers, R.A., Dunn, A., 2001. Is catch-per-unit-effort proportional to abundance? Can. J. Fish. Aquat. Sci. 58, 1760–1772.

Helu, S.L., Sampson, D.B., Yin, Y., 2000. Application of statistical model selection criteria to the Stock Synthesis assessment program. Can. J. Fish. Aquat. Sci. 57, 1784–1793.

Hilborn, R., Liermann, M., 1998. Standing on the shoulders of giants: learning from experience in fisheries. Rev. Fish. Biol. Fish. 8, 273–283.

Hilborn, R., Mangel, M., 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton, 315 pp.

Hilborn, R., Maunder, M.N., Parma, A., Ernst, B., Paynes, J., Starr, P.J., 2000. Documentation for a general age-structured Bayesian stock assessment model: code named Coleraine. Fisheries Research Institute, University of Washington, Seattle, Washington (FRI/UW 00/01).

Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman and Hall, New York, 570 pp.

Hobbs, N.T., Ogle, K., 2011. Introducing data–model assimilation to students of ecology. Ecol. Appl. 21, 1537–1545.

Hoenig, J.M., Warren, W.G., Stocker, M., 1994. Bayesian and related approaches to fitting surplus production models. Can. J. Fish. Aquat. Sci. 51, 1823–1831.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging. Stat. Sci. 14, 382–417.

Ianelli, J.N., 2002. Simulation analysis testing the robustness of productivity determinations from west coast pacific ocean perch stock assessment data. N. Am. J. Fish. Manage. 22, 301–310.

IWC, 1995. Report of the Scientific Committee, Annex F. Report of the Sub-Committee on Aboriginal Subsistence Whaling. Rep. Int. Whal. Comm. 45, 142–164.

Jiao, Y., Cortés, E., Andrews, K., Guo, F., 2011. Poor-data and data-poor species stock assessment using a Bayesian hierarchical approach. Ecol. Appl. 21, 2691–2708.

Jurado-Molina, J., Livingston, P.A., Ianelli, J.N., 2005. Incorporating predation interactions in a statistical catch-at-age model for a predator–prey system in the eastern Bering Sea. Can. J. Fish. Aquat. Sci. 62, 1865–1873.

Kinzey, D., Punt, A.E., 2009. Multispecies and single-species age-structured models of fish population dynamics: comparing parameter estimates. Nat. Res. Mod. 22, 167–204.

Laurec, A., Shepherd, J.G., 1983. On the analysis of catch and effort data. J. Cons. Int. Explor. Mer. 41, 81–84.

Lebreton, J.-D., 2005. Dynamical and statistical models for exploited populations. Aust. N.Z. J. Stat. 47, 49–63.

Liermann, M., Hilborn, R., 1997. Depensation in fish stocks: a hierarchic Bayesian meta-analysis. Can. J. Fish. Aquat. Sci. 54, 1976–1984.

Link, W.A., 1999. Modelling pattern in collections of parameters. J. Wild Manage. 63, 1017–1027.

Ludwig, D., Walters, C.J., 1985. Are age-structured models appropriate for catch-effort data? Can. J. Fish. Aquat. Sci. 42, 1066–1072.

MacCall, A.D., 2013. Use of the delta method to evaluate the precision of assessments that fix parameter values. Fish. Res. 142, 56–60.

Maunder, M.N., 1998. Integration of Tagging and Population Dynamics Models in Fisheries Stock Assessment. PhD Thesis, University of Washington, 306 pp.

Maunder, M.N., 2001a. 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., 2001b. 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., Roy, C., Smith, S.J., Witherell, D. (Eds.), Spatial Processes and Management of Fish Populations, Alaska Sea Grant College Program Report No. AK-SG-01-02. University of Alaska Fairbanks, Fairbanks, Alaska, pp. 123–146.

Maunder, M.N., 2001c. Growth of skipjack tuna (*Katsuwonus pelamis*) in the eastern Pacific Ocean, as estimated from tagging data. Bull. Inter-Am. Trop. Tuna Commn. 21, 529–571.

Maunder, M.N., 2003a. Paradigm shifts in fisheries stock assessment: from integrated analysis to Bayesian analysis and back again. Nat. Res. Model. 16, 465–475.

Maunder, M.N., 2003b. Is it time to discard the Schaefer model from the stock assessment scientist's toolbox? Fish. Res. 61, 145–149.

Maunder, M.N., 2004. Population viability analysis, based on combining integrated, Bayesian, and hierarchical analyses. Acta Oecol. 26, 85–94.

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., Deriso, R.B., 2010. Dealing with missing covariate data in fishery stock assessment models. Fish. Res. 101, 80–86.

Maunder, M.N., Deriso, R.B., 2011. A state-space multistage life cycle model to evaluate population impacts in the presence of density dependence: illustrated with application to delta smelt (Hyposmesus transpacificus). Can. J. Fish. Aquat. Sci. 68, 1285–1306.

Maunder, M.N., Langley, A.D., 2004. Integrating the standardization of catch-per-unit-of-effort into stock assessment models: testing a population dynamics model and using multiple data types. Fish. Res. 70, 389–395.

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70, 141–159.

Maunder, M.N., Starr, P.J., 2001. Bayesian assessment of the SNA1 snapper (*Pagrus auratus*) stock on the north-east coast of New Zealand. N.Z. J. Mar. Freshwat. Res. 35, 87–110.

Maunder, M.N., Starr, P.J., 2003. Fitting fisheries models to standardised CPUE abundance indices. Fish. Res. 63, 43–50.

Maunder, M.N., Starr, P.J., Hilborn, R., 2000. A Bayesian analysis to estimate loss in squid catch due to the implementation of a sea lion population management plan. Mar. Mamm. Sci. 16, 413–426.

Maunder, M.N., Watters, G.M., 2001. Status of yellowfin tuna in the eastern Pacific Ocean. Inter-Am. Trop. Tuna. Comm. Stock Ass. Rep. 1, 5–86.

Maunder, M.N., Watters, G.M., 2003a. A general framework for integrating environmental time series into stock assessment models: model description, simulation testing, and examples. Fish. Bull. 101, 89–99.

Maunder, M.N., Watters, G.M., 2003b. A-SCALA: an age-structured statistical catch-at-length analysis for assessing tuna stocks in the Eastern Pacific Ocean. Inter-Am. Trop. Tuna Comm. Bull. 22 (4), 433–582.

Maunder, M.N., Harley, S.J., Hampton, J., 2006. Including parameter uncertainty in forward projections of computationally intensive statistical population dynamic models. ICES J. Mar. Sci. 63, 969–979.

Maunder, M.N., Schnute, J.T., Ianelli, J., 2009. Computers in fisheries population dynamics. In: Megrey, B.A., Moksness, E. (Eds.), Computers in Fisheries Research. Springer, pp. 337–372.

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.

McAllister, M.K., Kirkwood, G.P., 1998. Bayesian stock assessment and policy evaluation: a review and example application using the logistic model. ICES J. Mar. Sci. 55, 1031–1060.

McAllister, M.K., Pikitch, E.K., Babcock, E.A., 2001. Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding. Can. J. Fish. Aquat. Sci. 58, 1871–1890.

McAllister, M.K., Pikitch, E.K., Punt, A.E., Hilborn, R., 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 52, 2673–2687.

McDonald, A.D., Kendrick, T.H., Breen, P.A., 2001. Empirical weighting of multiple stock-abundance indices for parameter estimation and stock assessment in a multi-zone or multi-species fishery. ICES J. Mar. Sci. 58, 204–215.

McGarvey, R., Linnane, A., Feenstra, J.E., Punt, A.E., Matthews, J.M., 2010. Integrating recapture-conditioned movement estimation into spatial stock assessment: A South Australian lobster fishery application. Fish. Res. 105, 80–90.

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

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

Meyer, R., Millar, R.B., 1999. Bayesian stock assessment using state-space implementation of the delay difference model. Can. J. Fish. Aquat. Sci. 56, 37–52.

Myers, R.A., Bowen, K.G., Barrowman, N.J., 1999. Maximum reproductive rate of fish at low population sizes. Can. J. Fish. Aquat. Sci. 56, 2404–2419.

Myers, R.A., Mertz, G., 1998. The limits of exploitation: a precautionary approach. Ecol. Appl. 8, 165–169.

Paloheimo, J.E., 1980. Estimation of mortality rates in fish populations. Trans. Am. Fish. Soc. 109, 378–386.

Parma, A.M., 2001. Bayesian approaches to the analysis of uncertainty in the stock assessment of pacific halibut. Am. Fish. Soc. Symp. 24, 111–132.

Patterson, K.R., 1999. Evaluating uncertainty in harvest control law catches using Bayesian Markov chain Monte Carlo virtual population analysis with adaptive rejection sampling and including structural uncertainty. Can. J. Fish. Aquat. Sci. 56, 208–221.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., Punt, A., Restrepo, V., Skagen, D.W., Stefansson, G., 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish Fish. 2, 125–157.

Pella, J.J., Tomlinson, P.K., 1969. A generalized stock production model. Bull. Inter-Am. Trop. Tuna Comm. 13, 421–458.

Petersen, C.G.J., 1896. The yearly immigration of young plaice into the Limfiord from the German Sea. Rep. Dan. Biol. Stn. 6, 5–84.

Piner, K.R., Lee, H.-H., Maunder, M.N., 2011. A simulation-based method to determine model misspecification: examples using natural mortality and population dynamics models. Mar. Coast. Fish. 3, 336–343.

Pope, J.G., 1972. An investigation of the accuracy of virtual population analysis using Cohort analysis. Bull. Int. Comm. Northwest Atl. Fish. Res. 9, 65–74.

Pope, J.G., Shepherd, J.G., 1982. A simple method for the consistent interpretation of catch-at-age data. J. Cons. Int. Explor. Mer. 40, 176–184.

Pope, J.G., Stokes, T.K., 1989. Use of multiplicative models for separable virtual population analysis (VPA), integrated analysis, and the general VPA tuning problem. Am. Fish. Soc. Symp. 6, 92–101.

Porch, C.E., Restrepo, V.R., Turner, S.C., Scott, G.P., 1995. Virtual population analyses of Atlantic bluefin tuna incorporating movement and tagging data. Coll. Vol. Sci. Pap. ICCAT 64, 183–208.

Porporato, A., Ridolfi, L., 2001. Multivariate nonlinear prediction of river flows. J. Hydrol. 248, 109–122.

Punt, A.E., 1993. The comparative performance of production-model and ad hoc tuned VPA based feedback-control management procedures for the stock of Cape hake off the west coast of South Africa. In: Smith, S.J., Hunt, J.J., Rivard, D. (Eds.), Risk Evaluation and Biological Reference Points for Fisheries Management. Can. Spec. Publ. Fish. Aquat. Sci. 120, 283–299.

Punt, A.E., Butterworth, D.S., 1995. Use of tagging data within a VPA formalism to estimate migration rates of bluefin tuna across the North Atlantic. Coll. Vol. Sci. Pap. ICCAT 64, 166–182.

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

Punt, A.E., Smith, A.D.M., 1999. Management of long-lived marine resources: a comparison of feedback-control management procedures. Am. Fish. Soc. Symp. 23, 243–265.

Punt, A.E., Kennedy, R.B., Frusher, S.D., 1997. Estimating the size-transition matrix for Tasmanian rock lobster, *Jasus edwardsii*. Mar. Freshw. Res. 48, 981–992.

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. Freshwat. Res. 51, 205–220.

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.

Punt, A.E., Deng, R.A., Dichmont, C.M., Kompas, T., Venables, W.N., Zhou, S., Pascoe, S., Hutton, T., Kenyon, R., van der Velde, T., Kienzle, M., 2010. Integrating size-structured assessment and bio-economic management advice in Australia's Northern Prawn Fishery. ICES J. Mar. Sci. 67, 1785–1801.

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.

Quinn II, T.J., 2003. Ruminations on the development and future of population dynamics models in fisheries. Nat. Res. Model. 16, 341–392.

Quinn II, T.J., Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press, New York, 560 pp.

Ricard, D., Minto, C., Jensen, O.P., Baum, J.K., in press. Evaluating the knowledge base and status of commercially exploited marine species with the RAM legacy stock assessment database. Fish and Fish, http://dx.doi.org/10.1111/j.1467-2979.2011.00435.x

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

Ricker, W.E., 1954. Stock and recruitment. J. Fish Res. Board Can. 11, 559–623.

Rivard, D., 1989. Overview of the systematic, structural, and sampling errors in cohort analysis. Am. Fish. Soc. Symp. 6, 49–65.

Sainsbury, K.J., 1988. The ecological basis of multispecies fisheries, and management of a demersal fishery in tropical Australia. In: Gulland, J.A. (Ed.), Fish Population Dynamics. , 2nd ed. Wiley, New York, pp. 349–382.

Schaefer, M.B., 1954. Some aspects of the dynamics of populations important to the management of commercial marine fisheries. IATTC Bull. 1, 25–56.

Schaub, M., Abadi, F., 2010. Integrated population models: a novel analysis framework for deeper insights into population dynamics. J. Ornithol. 152, 227–237.

Schaub, M., Gimenez, O., Sierro, A., el ArLettaz, R., 2007. Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. Con. Biol. 21, 945–955.

Schnute, J.T., 1994. A general framework for developing sequential fisheries models. Can. J. Fish. Aquat. Sci. 51, 1676–1688.

Schnute, J.T., Hilborn, R., 1993. Analysis of contradictory data sources in fisheries stock assessment. Can. J. Fish. Aquat. Sci. 50, 1916–1923.

Schnute, J.T., Richards, L.J., Olsen, N., 1998. Statistics, software, and fish stock assessment. In: Funk, F., Quinn II, T.J., Heifetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.J., Sullivan, P.J., Zhang, C.I. (Eds.), Fishery Stock Assessment Models. Alaska Sea Grant College Program Report No. AK-SG-98-01. University of Alaska Fairbanks, Fairbanks, Alaska, pp. 171–184.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.

Seber, G.A., Wild, C.J., 1989. Nonlinear Regression. John Wiley and Sons, New York, 769 pp.

Senina, I., Sibert, J., Lehodey, P., 2008. Parameter estimation for basin-scale ecosystem-linked population models of large pelagic predators: application to skipjack tuna. Prog. Ocean. 78, 319–335.

Shepherd, J.G., 1999. Extended survivors analysis: an improved method for the analysis of catch-at-age data and abundance indices. ICES J. Mar. Sci. 56, 584–591.

Smith, A.D.M., Punt, A.E., 1998. Stock assessment of gemfish (Rexea solandri) in eastern Australia using maximum likelihood and Bayesian methods. In: Funk, F., Quinn II, T.J., Heifetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.J., Sullivan, P.J., Zhang, C.I. (Eds.), Fishery Stock Assessment Models. Alaska Sea Grant College Program Report No. AK-SG-98-01. University of Alaska Fairbanks, Fairbanks, Alaska, pp. 245–286.

Starr, P.J., Bentley, N., Maunder, M.N., 1999. Stock Assessment of Two New Zealand Rock Lobster Substocks. New Zealand Fisheries Assessment Research Document 99/34. MAFFish Fisheries Research Centre, Wellington.

Taylor, N., McAllister, M., Lawson, G., Block, B., 2009. Review and refinement of the multistock age-structured assessment tag integrated model for Atlantic bluefin tuna. Coll. Vol. Sci. Pap. ICCAT 88, 182–194.

Walters, C.J., Punt, A.E., 1994. Placing odds on sustainable catch using virtual population analysis and survey data. Can. J. Fish. Aquat. Sci. 51, 946–958.

Xiao, Y., 1998. Two simple approaches to use of production models in fish stock assessment. Fish. Res. 34, 77–86.

Zobitz, J.M., Desai, A.R., Moore, D.J.P., Chadwick, M.A., 2011. A primer for data assimilation with ecological models using Markov Chain Monte Carlo (MCMC). Oecologia 167, 599–611.