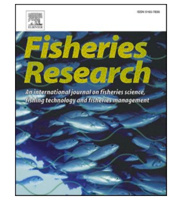
Fisheries Research 275 (2024) 107010

Contents lists available at ScienceDirect

Fisheries Research

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



Good practices for surplus production models

A. Kokkalis [a,*], C.W. Berg [a], M.S. Kapur [b], H. Winker [c], N.S. Jacobsen [a], M.H. Taylor [d], M. Ichinokawa [e], M. Miyagawa [e], W. Medeiros-Leal [f], J.R. Nielsen [a], T.K. Mildenberger [a]

[a] National Institute of Aquatic Resources (DTU Aqua), Henrik Dams Allé, Building 202, 2800 Kgs. Lyngby
[b] School of Aquatic and Fishery Sciences, 1122 NE Boat St, University of Washington, Seattle, WA, 98115, USA
[c] Department of Aquatic Resources, Institute of Marine Research, Swedish University of Agricultural Sciences, 45330, Sweden
[d] Thünen-Institute of Sea Fisheries, Bremerhaven, Germany
[e] Fisheries Stock Assessment Center, Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Yokohama, Japan
[f] Institute of Marine Sciences – OKEANOS, University of the Azores, Horta, Portugal

A R T I C L E I N F O

Keywords:
SPiCT
JABBA
Data-moderate stock assessment
Fisheries management
Science-based advice

A B S T R A C T

Surplus production models (SPMs) have a long history in fisheries ecology and are important for the assessment and management of marine species and stocks. However, the implementation, application, and usage of these models vary across regions and case studies. Good practice guidelines can streamline modelling workflows, inform the acceptance or rejection of an assessment, and facilitate the derivation of management advice from accepted assessments. This paper discusses current practices in the application of SPMs and proposes good practice guidelines for their use in stock assessment. We complement our recommendations with results from a simulation study examining the performance of an age-based operating model and a SPM assessment model under 60 scenarios with various assumptions regarding data quantity, quality, and model priors. We provide specific good practice guidelines for two widely used state-space SPMs: SPiCT and JABBA. Finally, we discuss current limitations and suggest avenues for future developments for SPMs.

1. Introduction

Seventy years after their development and first application (Schaefer, 1957, 1954), surplus production models (SPMs) still play an important role for the assessment of fish stocks and, thus, fisheries management advice for their sustainable exploitation. SPMs are used to provide official conservation and management advice by at least five international fisheries advice bodies: International Council for the Exploration of the Sea (ICES), International Commission for the Conservation of Atlantic Tunas (ICCAT), Indian Ocean Tuna Commission (IOTC), Northwest Atlantic Fisheries Organisation (NAFO), and the General Fisheries Commission for the Mediterranean (GFCM). In fact, more than half of the stocks within ICCAT were assessed with SPMs in 2022 (Cousido-Rocha et al., 2022). SPMs have additionally been implemented for several data-limited domestic stocks, such as in the United States (e.g., Kapur et al., 2019), Japan (e.g., Chiba et al., 2023), or South Africa (Winker et al., 2020b). SPMs have been applied to demersal flatfish species such as megrim *Lepidorhombus spp.* (ICES, 2021a) and brill *Scophthalmus rhombus* (ICES, 2023a), pelagic species such as albacore *Thunnus alalunga* (ICCAT, 2017a) and swordfish *Xiphias gladius* (ICCAT, 2017b), elasmobranchs such as blue shark *Prionace glauca* (ISC, 2017) and thornback ray *Raja clavata* (ICES, 2023a), invertebrates such as Norway lobster *Nephrops norvegicus* (González Herraiz et al., 2023) and tiger prawn *Penaeus monodon* (Zhou et al., 2009), and marine mammals such as fin whales *Balaenoptera physalus* (Moore and Barlow, 2011). SPMs have also been applied to global data sets investigating how the productivity of fish populations varies between taxonomic groups (Thorson et al., 2012) or used to couple environmental conditions with population dynamics (Free et al., 2020; Fréon, 1986).

SPMs are based on the ecological theory of density-dependent population growth with exponential population growth at low abundance (numbers or biomass) and low population growth when abundance is close to the stock carrying capacity (K). For most implementations, biomass is aggregated across ages, sizes, and sexes. SPMs rely on the following assumptions: (i) the modelled biomass belongs to a closed population, where there is no immigration or emigration, (ii) only the part of the population that is vulnerable to fishing fleet is modelled, i.e.,

* Corresponding author.
E-mail address: alko@aqua.dtu.dk (A. Kokkalis).

https://doi.org/10.1016/j.fishres.2024.107010
Received 31 May 2023; Received in revised form 29 March 2024; Accepted 29 March 2024
Available online 6 April 2024

the exploitable biomass, (iii) there are no lag effects, and (iv) typically estimated variables are constant throughout the extent of the available data. The generalised SPM includes a parameter that determines the shape of the production curve (n) and can be expressed in terms of biomass (B) by the following differential equation (Pella and Tomlinson, 1969):

$$\frac{dB_t}{dt} = \frac{r}{n-1}B_t\left(1-\left[\frac{B_t}{K}\right]^{n-1}\right) - F_tB_t \tag{1}$$

where r is the intrinsic population growth rate and F is the instantaneous fishing mortality rate. Surplus production is the amount of the biomass that can be sustainably harvested while maintaining population size. The relationship between the surplus production and stock biomass is described by the production curve, where the maximum corresponds to the maximum sustainable yield (MSY). Two common assumptions for the shape parameter, which determines the skewness of the production curve, are: the Schaefer model that assumes a symmetrical production curve with MSY corresponding to the biomass at half of carrying capacity K ($n = 2$; Schaefer, 1954) and the Fox model that has an asymmetrical production curve with MSY at 37% of carrying capacity K ($n \to 1$; Fox, 1970).

SPMs typically require a time series of fishery removals and a relative measure of stock abundance (either as catch-per-unit-effort or from a survey index). There is a long history of estimating parameters using surplus production models and estimation methods have evolved considerably since their first introduction. Early approaches relied on the assumption that the stock is at equilibrium wherein the relationship between the index and effort is linear. This facilitates parameter estimation but must be avoided as these methods often over-estimate surplus production and F_{MSY} (Hilborn and Walters, 1992), especially when the data are from a period where the stock has been declining.

Non-equilibrium methods were developed with the advent of computers, including observation-error and process-error estimators and more recently state-space formulations that estimate both types of uncertainty. Several studies compared these methods with varying results (e.g., Pedersen and Berg, 2017; Polacheck et al., 1993; Punt, 2003; Williams and Prager, 2002). Observation-error-only models seem to perform better than process-error models, while both are out-performed by state-space models in most cases. We recommend using non-equilibrium methods and especially state-space methods that are readily available as well-documented, user-friendly, free, and open-source software implementations.

Like in any model, the process equation (Eq. 1) is a simplified version of the real world that neglects many aspects affecting the population, such as ecosystem and environmental effects or population structure. A useful model ideally captures important characteristics of the population and can, therefore, be used to do short-term forecasts and support scientific advice. Process variability is modelled by adding an error term to the deterministic equation of surplus production (Eq. 1), most often as lognormally distributed deviations or as a Wiener process for continuous time models like the Stochastic surplus Production model in Continuous Time (SPiCT; Pedersen and Berg, 2017). SPiCT is using the Fletcher reparameterization of Eq. 1 (Fletcher, 1978) with parameters $m = rK/n^{n/(n-1)}$, which is the deterministic maximum sustainable yield, and the purely numerical $\gamma = n^{n/(n-1)}/(n-1)$. Adding the stochastic part leads to the stochastic differential equation:

$$dB_t = \left(\gamma m\frac{B_t}{K} - \gamma m\left[\frac{B_t}{K}\right]^n - F_tB_t\right)dt + \sigma_B B_t dW_t \tag{2}$$

where σ_B is the standard deviation of the biomass process noise and W_t is Brownian motion. The underlying assumption is that the deterministic model is correct except for random deviations due to unmodelled processes. This assumption can be tested by evaluating the process residuals.

The unobserved processes are estimated using the observation equations describing the biomass index, I_t, and catch, C_t:

$$\log(I_t) = \log(q_tB_t) + e_t \tag{3}$$

$$\log(C_t) = \log\left(\int_t^{t+\Delta t} F_sB_s ds\right) + \varepsilon_t \tag{4}$$

where q_t is the catchability parameter of the abundance index and $e_t \sim N(0, \sigma_I^2)$, $\epsilon_t \sim N(0, \sigma_C^2)$ are independent normal deviates; σ_I and σ_C are the standard deviations of the index and catch observation errors, respectively. The catch is observed over a time interval Δt. A time-varying catchability q_t is presented in the above formulation, but catchability is often assumed to be constant. The quantities are modelled in log-space ensuring strictly positive observations and transforming the multiplicative error structure to additive, which leads to a more stable fitting problem (Pedersen and Berg, 2017). The independence of the observations implies that any dependence between responses is only due to the unobserved states (Aeberhard et al., 2018).

Various SPM implementations relax some of the baseline assumptions, e.g., SPMs can include time-varying model parameters (Chang et al., 2020; Mildenberger et al., 2020). In theory, every assessment model involves the concept of surplus production (Schnute and Richards, 2002). However, many other widely-used population dynamics and stock assessment models such as the state-space assessment model (SAM; Berg and Nielsen, 2016; Nielsen and Berg, 2014) or Stock Synthesis (Methot and Wetzel, 2013) model the age structure of the population and separately represent the processes of recruitment, somatic growth, maturation, and natural mortality. SPMs, on the other hand, only track the aggregated exploitable stock abundance over time (Pedersen and Berg, 2017) and the population dynamics processes are lumped together into a single production function (Eq. 1), while ignoring maturation and time lag effects of recruitment into the fishery (Aalto et al., 2015). The spawning stock biomass (SSB) is not estimated in the model. It can be approximated by the exploitable biomass if the fishery selectivity ogive is similar to the maturity ogive and for higher biomass levels where the ratio of exploitable biomass to spawning stock biomass is constant (Winker et al., 2020a).

While the model simplifications do not allow for inference about cohort length or age-specific processes, SPMs are valuable for the assessment of data-moderate stocks where assessment methods that model the population structure cannot be applied (Ludwig and Walters, 1989, 1985). Here, we use the term "data-moderate" to describe stocks that lack sufficient age or length information but have time series of catch and one or more abundance indices; we use the term "data-rich" to describe stocks that can be assessed with age- or length-based methods.

The modelled biomass B is the part of the biomass that is affected by the catch (Eq. 1) and therefore should be the exploitable biomass. Nevertheless, the assessment can be based on additional assumptions, e.g., that the fisheries selectivity ogive is very similar to the maturity ogive, i.e., B can be interpreted as the SSB.

SPMs require only catch time series and an index of abundance, based either on a fishery-independent (scientific) survey or on a fishery-dependent catch per unit of effort (CPUE) time series. The influential and difficult to estimate natural mortality rate does not need to be explicitly defined or estimated (Maunder et al., 2023; Punt et al., 2021). Here, we focus on biomass-dynamic models, and do not cover other types of production models like age-structured production models (ASPM; Hilborn, 1990), or extensions to SPMs that directly incorporate additional information (e.g., life-history parameters, selectivity), like Just Another Bayesian Biomass Assessment-select (JABBA-Select; Winker et al., 2020a), because these models rely on different underlying assumptions and an extended set of stock parameters.

Because of the low data requirements, SPMs are often applied to data-moderate fish stocks. To estimate parameters that define the shape (n), height (MSY) and width (K) of the production curve and therefore be able to reliably estimate reference points, it is necessary to have observations (catch and index) from periods where the stock was highly and mildly exploited (Kai and Yokoi, 2019) such that the stock's response in terms of surplus production at various stock sizes can be inferred. The term 'contrast' describes how much of the production curve has been observed, e.g., low contrast indicates that the available data only cover a limited part of the production curve. When the contrast is low, there is insufficient information in the data, hindering reliable estimation of model parameters. A common way to estimate model parameters under these conditions is to use auxiliary information to define prior distributions for key model parameters. For that reason, most generalised SPMs are either fully Bayesian or have the option to specify prior-like penalties in a maximum likelihood framework.

The Johannesburg declaration on sustainable development identifies as first action for sustainable fisheries to "maintain or restore stocks to levels that can produce the maximum sustainable yield with the aim of achieving these goals for depleted stocks" (UN, 2003). This translates to management using target MSY reference points. An advantage of SPMs over other common stock assessment methods is that such target reference points are clearly defined within the model and derived from parameters estimated by the model. The uncertainty around the estimated reference points is also quantified. These reference points imply the existence of a fishing mortality F_{MSY} that on the long term will bring the stock to a biomass B_{MSY} at which the maximum yield can be extracted from the population.

Focusing mainly on data-rich stock assessment methods, Punt (2023) provides good practice guidelines for stock assessment and recommends age-structured population models. This paper extends those guidelines to situations where age or length measurements are unavailable or only sporadically sampled, but reliable time-series of catch and fishing effort data are available. Besides low commercial importance, a lack of reliable and regularly collected age and length data is still one of the main reasons why the majority of global exploited stocks remains unassessed (Blasco et al., 2020; Costello et al., 2012; Hilborn et al., 2020; Neubauer et al., 2013). We propose good practice guidelines for SPMs regarding: (1) Input data and data preparation, (2) Model configuration and fitting procedures, (3) Model performance diagnostics and model acceptance, (4) Choice of model, and sensitivity analysis, and (5) Management advice (Fig. 1). The guidelines synthesise lessons learnt and recommendations from several recent method development and assessment working groups (e.g., ICES, 2023b, 2021a, 2020a), build upon existing recommendations for SPMs (Cousido-Rocha et al., 2022), and are further supported with additional insights from a simulation study based on 60 scenarios used to explore effects of process variation, index precision, time-series length, missing data, and model priors. The simulation study is based on an age-based operating model parameterised according to the life-history parameters of the haddock (*Melanogrammus aeglefinus,* L. 1758) stock in the Celtic Sea (ICES, 2019; Supplementary Fig S1), one of the most important demersal species in the area that has an age-based assessment. SPiCT is used as the assessment model (See Supplementary Section S1 for a detailed description of the simulation-estimation framework and the scenarios including a baseline assessment in Fig S2 and the effect of observation and process uncertainty to results in Figs S3–4). We provide comprehensive good practice guidelines for SPMs in general, with a specific emphasis on the two most used SPMs in the above listed fisheries advice organisations: SPiCT

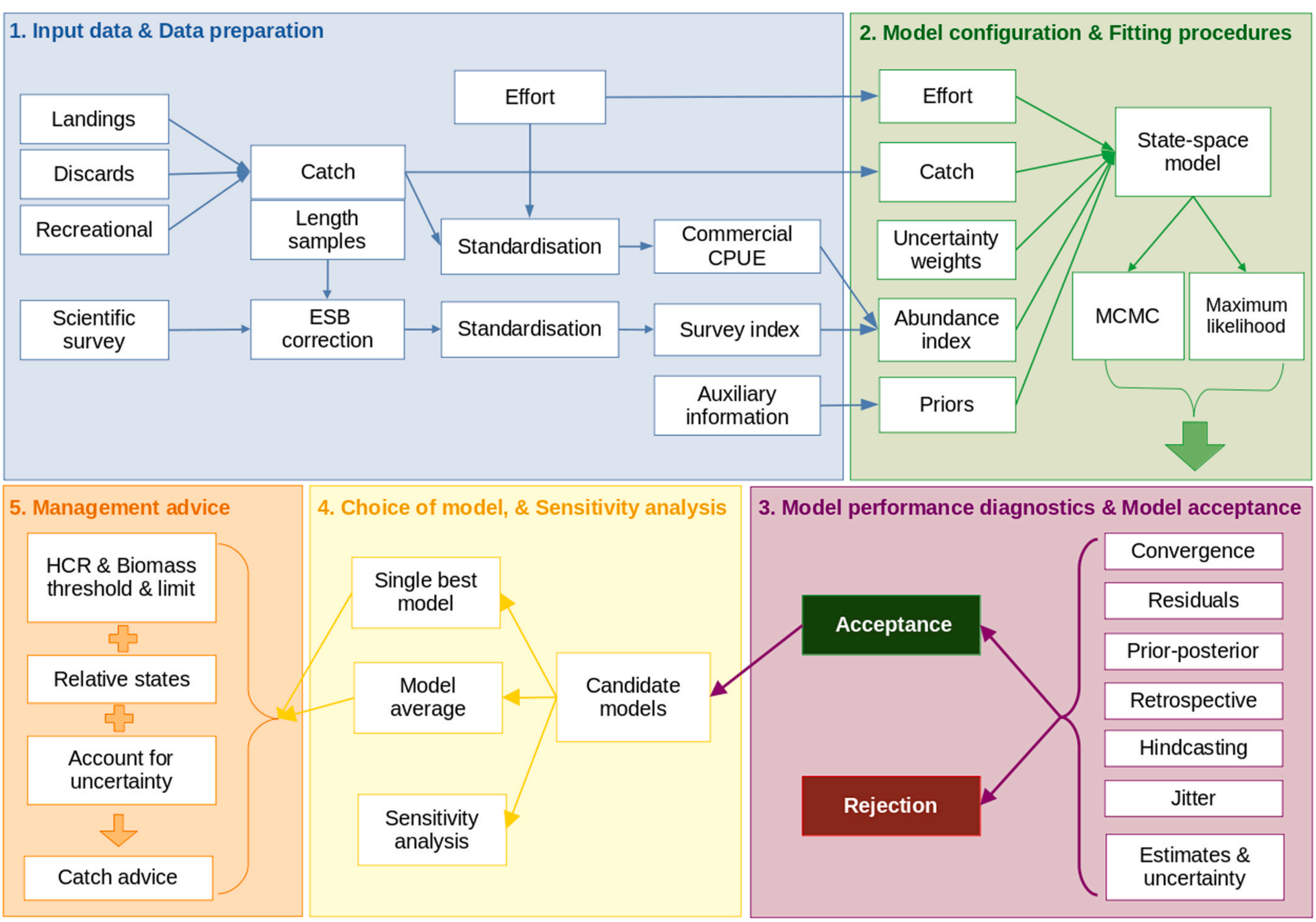


Fig. 1. Flowchart of good practices for surplus production models. MCMC refers to Markov Chain Monte Carlo. ESB is the exploitable stock biomass. CPUE is the catch per unit of effort.

(Pedersen and Berg, 2017) and JABBA (Winker et al., 2018). Finally, we discuss shortcomings of SPMs based on literature and simulation studies, and outline areas of future model research.

2. Input data and data preparation

The minimum data requirements for SPMs are a time series of catch and an index of abundance (Hilborn and Walters, 1992). Some SPMs make use of an optional effort time series (Pedersen and Berg, 2017) for the purpose of informing the underlying fishing mortality process. Finally, for Bayesian methods, prior distributions are required for all estimated model parameters.

2.1. Catch

The commercial catch time series provides important information about the absolute scale of the fish stock and the total removals from the population resulting from fishing pressure. The catch time series has to be as complete as possible and should represent either landings or total catch (landings and discards), but not a combination of both (i.e., some years with and others without discard information). Catch should be used as total weight (or total numbers) caught within the same time interval used in the model (Eq. 4). Catches are typically annual, but in the case of continuous-time models like SPiCT, shorter time intervals (e.g., quarters) can be used if more disaggregated data is available (e.g., Mildenberger et al., 2020).

Recreational fisheries can constitute a substantial fraction of or be even higher than commercial catches. Appropriate quantification of discards and recreational harvest is important to avoid estimation bias, particularly if the discarding or recreational fishing mortality rate varies over time, which can result in sub-optimal management of the stock (Omori et al., 2016). Reliable recreational fishery data is often scarce (van Gemert et al., 2022) and reconstruction of total removals can be difficult.

A common problem for many stocks is uncertain catch information during the early part of the time series due to missing historical discard rates, a lack of species disaggregated information, or misreporting according to species allocation of landings. Although it can be difficult to reconstruct historical catches (landings and discards), rough estimates (with an additional time-series of uncertainty included in the model; see section about variance estimates and uncertainty scaling below) can still be useful for SPMs as they may contain additional information and increase the contrast in the data. When historical catches constitute the highest observed catches, it adds to the contrast as there are observations from a wider range of exploitation levels. State-space SPMs can account for observation errors, such that time-periods with more uncertain catch observations can be included in the model without giving them undue leverage on model estimates (Pedersen et al., 2011).

2.2. Abundance index

Indices of abundance constitute the other important type of input data for SPMs. The main assumption is that they are proportional to the exploitable part of the stock biomass and provide information on the relative stock size over time rather than the absolute population scale. SPMs can operate using absolute biomass estimates if such time series are available (e.g., Antarctic blue whales; Hamabe et al., 2023). The catchability parameter does not need to be estimated and is used as a constant $q = 1$ in Eq. 3, when absolute abundance is available.

Abundance indices can be derived from one or more fishery-independent scientific surveys or from commercial catch per unit effort (CPUE) series, the latter ideally from the part of the fishing fleet that is responsible for the majority of the catches. Abundance indices are assumed to be snapshots at given points in time (Eq. 3), such as a single month during the fishing year. Thus, in a SPM with sub-annual time steps (e.g., SPiCT), the exact timing of the observations must be specified, or the indices must be assigned to their closest model time steps. Commercial CPUEs, which are based on annual catch and effort observations, should be assigned either to the peak of fishing activity within the year (if such information is available) or to the middle of the year in a SPM with sub-annual time steps.

The standardisation of CPUE indices plays an important role and good practices should be followed for any assessment method that uses CPUE, as unstandardised or wrongly standardised CPUE indices can introduce bias to the SPM estimates (Grüss et al., 2023; Maunder and Punt, 2004). The following is a set of considerations to inform good practice when standardising CPUE for use in SPMs; for a comprehensive review of CPUE standardisation methods, we refer to Hoyle et al. (2024).

Standardisation of indices can incorporate different types of information that are not otherwise available for data-moderate stocks. We describe how various data types can be used; in the absence of specific data, assumptions need to be made and sensitivity analyses should be performed to evaluate the impacts of each assumption. Assessors should consider that factors such as targeting, spatial patterns, or technological creep that may bias the perception in abundance trends when unaccounted for (Maunder and Punt, 2004; Nielsen, 2015). The standardization of commercial CPUE should account for the spatial distribution of the stock either on a fine (Thorson et al., 2016) or in a coarse spatial scale (Grüss et al., 2019). Typically, a fishery will have greater temporal coverage, but will be spatially focused on areas of certain life stages and likely not cover the full stock distribution area. Conversely, scientific surveys may have better spatial coverage of the stock distribution, but less temporal coverage and proportion of stock caught influencing certainty of density and abundance estimates (e.g., Nielsen, 2015; Rufener et al., 2021). Multiple surveys that each cover only a subset of the stock area (e.g., country specific surveys) should not be used as independent indices in the model, rather they should be combined into a single stock index, if possible. However, several independent representative indices should not be combined. For example, brill in the North Sea is assessed using SPiCT and includes three indices, the first two represent first and second semester indices and combine several national beam trawl surveys, while the third index is based on the first quarter North Sea International Bottom Trawl Survey (NS-IBTS Q1; ICES, 2023a).

Zero observations must be modelled properly (Maunder and Punt, 2004) and different assumptions of technological creep should be explored (Eigaard et al., 2014). The targeting behaviour of mixed fisheries fleets must be considered (Winker et al., 2013) as well as alternative error distributions and model formulations for the standardisation procedure. CPUE models should not include terms that smooth over time to avoid introducing autocorrelation in the index estimates, as these will render the resultant time series to be non-independent through time, breaking an important assumption of the SPM. The uncertainty of the indices should be quantified and used in the SPM as relative weights if uncertainties change over time, and/or used to inform the model about observation variance parameters, e.g., through priors.

A main assumption of SPMs is that the modelled biomass is the exploitable or vulnerable biomass. To satisfy this assumption, the abundance indices should represent the exploitable part of the stock corresponding to the subset of the stock that is exploited by (or vulnerable to) the overall fisheries fleet. For commercial CPUEs, this is per definition the case, but for fisheries independent indices some steps need to be taken during standardisation to satisfy this assumption, otherwise additional uncertainty or bias can be introduced to the model estimates. It is common that scientific surveys capture a larger part of the population due to difference in selectivity compared to the commercial fleet; often the smallest individuals are only caught during scientific surveys. Excluding the non-exploitable part of the survey catches can be done in different ways, depending on available information. Expert knowledge can be used to exclude all individuals below a certain size that is assumed not to be caught by the commercial fleet. If size (or age) distributions from the survey and the commercial catches are available, a simple correction factor can be calculated based on the ratio of the two distributions and used to adjust the catches from scientific survey so that they correspond to the vulnerable part of the population (cf. case study in Pedersen and Berg, 2017). We suggest keeping such correction factors constant through time. If there are major time trends in the selectivity, we caution against the use of SPMs as this will introduce time varying MSY reference points as well.

Finally, the abundance index can be defined in a different way when the modelled biomass is the SSB instead of exploitable biomass or when the stock is exploited by many different gears with different selectivities. The former case occurs when the selectivity of the commercial fleet can be assumed to be similar to the maturity ogive and an SSB index should be used, calculated solely from mature individuals caught in the survey. In practice, a known or assumed length at 50% maturity (L_{50}) can be used as cut-off. For the latter case, there is no clear definition of exploitable biomass for stocks that are harvested by multiple gears with different selectivity ogives. Additional assumptions are required about the abundance index and consecutively about the assessment in general. Since catches are a result of the combined selectivity of all commercial fishing gears, the index should ideally approximate the selectivity profile of the major fleets. When a single fleet is responsible for the majority of the catches, a CPUE index that corresponds to the selectivity of that fleet can be used. A cut-off length can be used to calculate exploitable biomass index, or the ratio of proportions at length or age can be used if length or age distributions are available from the survey and the fishery (Pedersen and Berg, 2017).

2.3. *Effort*

Another potentially informative data type, which is less frequently used in SPMs, is total fishing effort that can provide important information regarding the trend in fishing mortality. An effort time series is informative if it corresponds to a substantial part of the fishing fleet that catches the stock, and if the catchability of the fleet representing the effort does not change significantly, e.g., by extensive technological fleet developments. Therefore, effort data are more easily included in models that have an unobserved process for fishing mortality (e.g., SPiCT). The main motivation for using observations of total effort rather than deriving commercial CPUE from them, is that observation errors on effort and catch should be independent, whereas errors on catch and CPUE are not independent because the error in catches affects both. Good effort information is not always easy to obtain for several reasons and the following caveats should be considered. Effort information is not always routinely monitored like commercial catches and information collected from scientific surveys. It is even more difficult to collate effort information further back in time. Further, it is not straightforward to harmonise effort from diverse fishing fleets and fishing gears, e.g., effort from static gillnets and from active trawls, and to select fleets that target the species or that are responsible for substantial catches as bycatch (e.g., Nielsen, 2015; Nielsen et al., 2006; Ulrich et al., 2012). Furthermore, effort information is considered sensitive economic information and there are issues of sharing and combining effort information especially for stocks that are shared between different political entities that jointly exploit the stock.

2.4. *Length of time series*

The time series length of available input data is important, and a general guidance can be derived from the number of observations relative to model parameters. For example, SPiCT and JABBA have 8 and 5 model parameters using one abundance index and default settings, respectively. The simulation study shows that at least 15 years of data are required for the time-series to be informative enough and have contrast to estimate all model parameters. A shorter time series can lead to inaccurate and imprecise estimates (Supplementary Figs S5 and S6). Increasing time series length reduces uncertainty in all model estimates and derived states, but only in the case of increasing contrast in the data (Supplementary Figs S5-S10). The results of the simulation study confirm findings by Hilborn and Walters (1992) on the importance of the contrast in the data: Overall the more contrast the better, low contrast leads to non-convergence, biased estimates, and high uncertainty according to the simulation study (Supplementary Figs S9 and S10). Periods of low fishing pressure and high stock biomass are of particular importance, which can likely be attributed to the fact that these periods carry information to estimate K. Lack of information due to data quantity (e.g., short time series length) or data quality (e.g., lack of contrast) demands more informative model priors. One-way historical exploitation patterns can lead to biased and uncertain estimates (Supplementary Figs S11 and S12).

2.5. *Variance estimates and uncertainty scaling*

The input data time series discussed above are typically results of monitoring and data collection as well as modelling and simulation efforts. It is common that these time-series are varying in precision throughout their duration. Several factors can affect the relative uncertainty of these time series. For example, early catch statistics could be more uncertain compared to more recent ones as argued above, e.g., due to improving data collection, or scientific surveys could be disrupted by weather conditions or global pandemics that reduce sampling in some of the years. The observation errors for abundance indices can be formulated based on annual standard error estimates from the index standardisation model, both in terms of inter-annual and absolute precision. Accordingly, the relative uncertainty can be determined qualitatively or quantitatively, respectively. An additional time series of uncertainty can be included in the model for each of the observation time series (catch, index, effort). For each observation, the corresponding uncertainty acts as a multiplier to the overall observation standard deviation, thus acting as scaling of the estimated uncertainty of that observation. In SPiCT, the error structure in Eqs. 3 and 4, is effectively $e_t \sim N(0, \psi_t \quad \sigma_I^2)$, $\epsilon_t \sim N(0, \phi_t \quad \sigma_C^2)$, where ψ and ϕ are the uncertainty scaling vectors; the ψ and ϕ vectors need to be standardised to have an average of one to avoid having an effect on the estimate of the overall observation error and changing the interpretation of the estimated variance parameters and their prior distributions. Additional uncertainty time series can be used in a different way. In JABBA, annual variance estimates for the abundance index can be specified as additional input time-series and are treated as minimum observation error as they are added to the overall fixed and estimable variances to calculate the total variance of each observation (Winker et al., 2018).

Qualitative scaling requires an assumption on the relative weights for the different uncertainty periods. To illustrate, we consider a situation where there is knowledge about changes in the catch monitoring process of a stock where the sampling was much lower in the first part of the time series compared to the latter part. A reasonable assumption is that catch observations are more uncertain in the first part and a scaling time series can be used informing the model that the earlier observations are, e.g., twice as uncertain as the rest. Of course, the level of uncertainty difference is unknown and the effect of that assumption to the results should be tested with a sensitivity analysis, where alternative levels are selected. In addition to relative uncertainties, absolute estimates of observation variance can be very useful for SPMs, because time-series are often too short to reliably separate process error from observation error in the model.

3. Model configuration and fitting procedures

The implementation and fitting procedures of SPMs have evolved from simple models assuming equilibrium conditions and perfect observations, to models that estimate observation or process errors, and, finally, to complex state-space models with seasonal processes and multivariate prior distributions (Cousido-Rocha et al., 2022; Hilborn and Walters, 1992). Good practice is to use state-space models, i.e., models that include a set of functions that describe the unobserved stochastic processes over time based on a model of the system's dynamics and a set of functions that relate the noisy observations to the unobserved quantities. Thus, the surplus production of state-space SPMs does not only depend on the biomass (density dependence, Eq. 1), but also on a random component (process error) that describes stochastic variability around the deterministic biomass process equation (Eq. 2). Most observation and process error estimators are nested within the more complex state-space SPMs (Polacheck et al., 1993). While most modern SPMs include an unobserved process for the production function, to the knowledge of the authors, only SPiCT also assumes an unobserved process for fishing mortality (F). This enables the estimation of the uncertainty associated with catch observations, and the prediction of F at any time independent of the catch observations, including periods with abundance index observations but without catch observations, and prediction of missing catch observations. The unobserved process for F increases the model complexity, requiring more parameters to be estimated from limited data and difficulty to define the underlying F process (for example, SPiCT assumes a random walk). Another aspect that varies between different SPM implementations is whether the models are discrete models with annual time steps (e.g., Winker et al., 2018), or are formulated in continuous time (e.g., Pedersen and Berg, 2017; Prager, 1994). Sub-annual model time steps make it possible to include seasonal model processes for F and B when sampling occurs at irregular intervals, such as combining yearly and quarterly data (Mildenberger et al., 2020).

The estimation of some model parameters can be problematic. Two common examples are the estimates of the intrinsic growth rate r and carrying capacity K that are known to be correlated (Fletcher, 1978), and the shape of the production curve (n), which is notoriously difficult to estimate (see discussions in Maunder, 2003; Prager, 2002). The parameters and posterior distributions for these state-space models cannot be estimated analytically due to the large number of parameters (fixed parameters and random effects) and nonlinearity of the surplus production function (Eq. 1). Thus, parameters are typically estimated by either maximising the log-likelihood function (e.g., SPiCT) or applying a general-purpose Markov chain Monte Carlo method (MCMC; e.g., JABBA). While MCMC allows including priors directly, optional priors can be included in the former by multiplying them with the likelihood function to obtain posterior distributions. The maximum of the posterior distributions defines the Bayesian maximum a posteriori parameter estimates. While current modelling frameworks and fitting procedures for SPMs have reached a high level of sophistication, there are still some considerations to account for (Best and Punt, 2020), and ultimately, any model is only as robust as the data that went into it.

3.1. Prior distributions

A priori knowledge about model parameters can be included in a model as prior probability distributions (here referred to as priors). For true Bayesian models such as JABBA, all model parameters require a prior. On the other hand, state-space models that are fit with maximum likelihood such as SPiCT, priors for some of the model parameters can be added as likelihood penalties. These priors are optional but might be important for ensuring model convergence or addressing residual issues of conflicting abundance indices. Ideally, priors are specified based on empirical information about the stock under study or, alternatively, from closely related stocks or species with similar life histories or domain knowledge. This information can be obtained from large databases such as FishBase (Froese and Pauly, 2023) and by means of R packages such as *rfishbase* (Boettiger et al., 2012) and *FishLife* (Thorson et al., 2023). Alternative sources for defining priors are meta-analyses (e.g., Thorson, 2020; Thorson et al., 2012), auxiliary information, or expert knowledge. In absence of auxiliary information, informative priors should be avoided. Truly uninformative priors do not exist, and sensitivity analysis is required to evaluate differences between prior choices. Thorson and Cope (2017) warn against the use of seemingly uninformative uniform priors on the logarithm of the population scale (the carrying capacity in SPMs) and showed that the chosen upper bound of these uniform priors can be very influential to the estimates. The mean of any prior corresponding to quantities on log scale should be bias corrected by subtracting the variance of the mean divided by two (Finney, 1941). In general, priors for stock-specific model parameters (K, MSY) should be avoided in SPMs with maximum likelihood estimation and uninformative (wide) priors should be preferred for Bayesian SPMs. Such parameters are rarely transferrable between stocks or species and making good guesses about their magnitude is not straightforward. Finally, informative priors have to be selected with caution as they can lead to biased estimates, as is illustrated from the results of the simulation study (Supplementary Figs S13-,16). It is therefore recommended to always perform sensitivity analyses of all priors used in a production model to avoid a situation where key model output is driven by one or more priors, or at least be aware about it.

Conceptually, it is only possible to estimate the shape of the production curve if a stock had remained notably below B_{MSY} over several years of the time series. The simulation study confirmed that the shape parameter is a parameter that is difficult to estimate for SPMs even with sufficient contrast and a long time series. However, a wide prior (here: SD = 2–10) can help the model to converge without causing substantial bias even if misspecified (Supplementary Figs S13-,16). At the same time, a misspecified n prior can also cause biased estimates if a tighter standard deviation is used (here: SD $<$ 2). When the misspecified n prior has a mean larger than the true n, the bias was towards the precautionary side (median relative error $< +/-$ 21%), i.e., overestimating F/F_{MSY} and underestimating B/B_{MSY}, and is less precautionary (median relative error between -34% and 56%, for the two quantities respectively) when the mean of the n prior is smaller than the true n (Supplementary Figs S13-,16).

3.1.1. Priors from meta-analyses

Meta-analyses can be useful for defining priors for some SPM parameters. Model parameters that largely depend on life-history parameters of a species, such as the intrinsic growth rate (r) or the shape of the production curve (n) are generally better candidates for borrowing information from other studies and systems, than model parameters specifying stock-specific quantities such as the carrying capacity K (and maximum productivity m for SPiCT). Two meta-analyses are available that suggest taxonomic prior values for r (Thorson, 2020) and n (Thorson et al., 2012). While r and n are likely to be more comparable between different stocks, it should be noted that stock- and fishery-specific factors, such as the fleet selectivity and the exploitation history can imply different shapes of the production curve and thus affecting n and r (e.g., Szuwalski, 2019). In addition, before borrowing model parameters from the literature, it should be confirmed that the values are biologically reasonable, and that the parameterisation of the meta-analysis model used to derive prior distributions is consistent with the stock-assessment model. Moreover, model parameter estimates of n, r and K are often correlated and consequently are the stock's reference points (MSY, F_{MSY}, B_{MSY}); this fact has to be accounted for when borrowing information between assessments or from meta-analyses. Keeping all other parameters constant, the reference points depend on the shape of the production curve, with smaller n implying larger F_{MSY} and MSY and smaller B_{MSY} (Fig. 2 A). The same is true the other way round: keeping all reference points constant, the parameter r and K (and the maximum surplus production parameter m of SPiCT) depend on the shape of the production curve, with smaller n implying smaller r and larger K (Supplementary Fig S17). In the context of using meta studies to inform priors, all parameter values of the model used for the meta-analysis should be considered when defining a prior for a model parameter. We recommend borrowing parameters between equivalent models, e.g., use meta studies based on the Schaefer model to construct priors for a

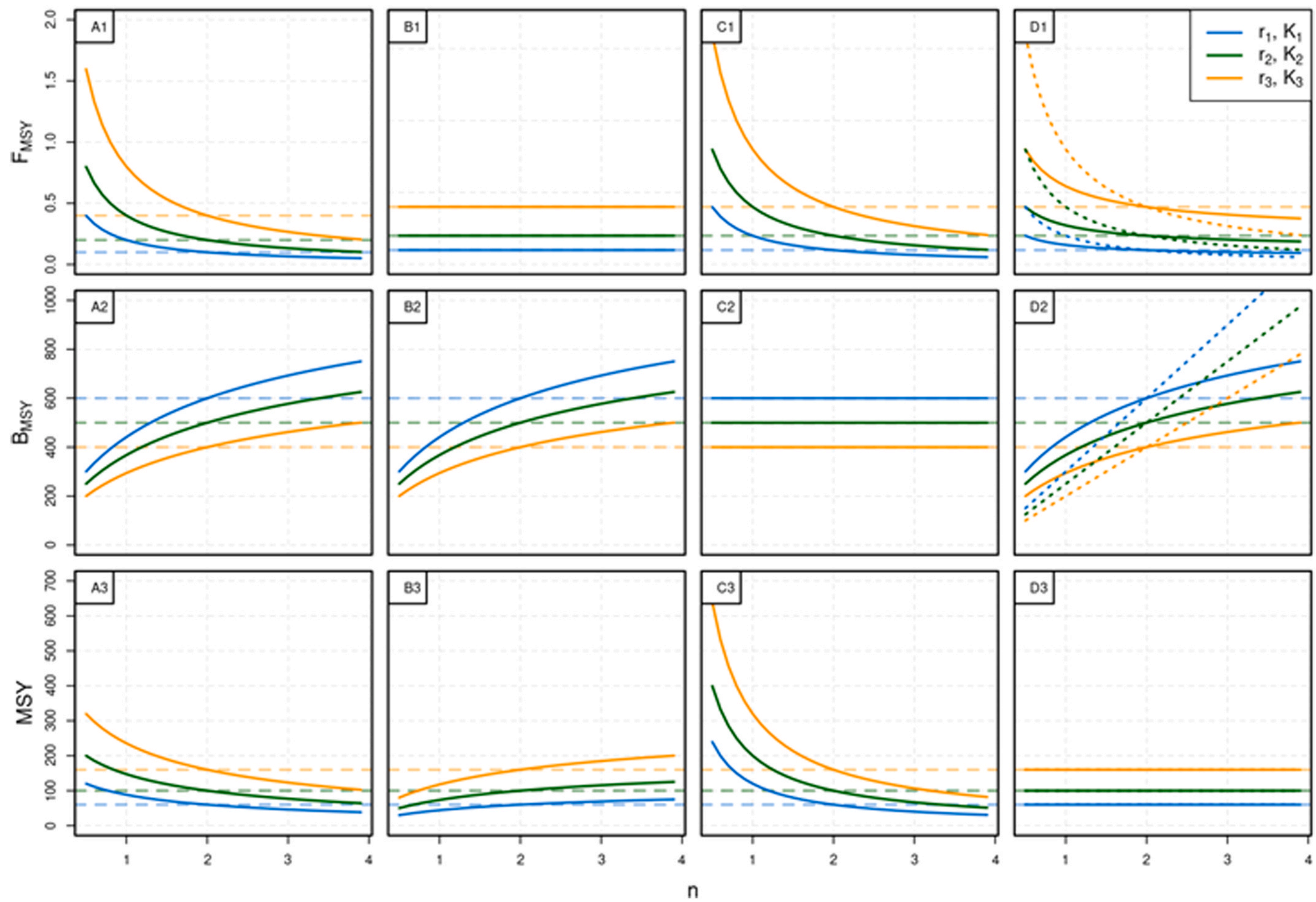


Fig. 2. The relationship between the shape of the production curve (n) and the biological reference points based on maximum sustainable yield (F_{MSY}, B_{MSY}, and MSY) for three different combinations of the intrinsic growth rate (r) and carrying capacity (K) estimated with a Schaefer production model (n = 2). The first column (A1-3) does not use any adjustment, the second column (B1-3) adjust r so that F_{MSY} is independent on n, the third column (C1-3) adjusts K so that B_{MSY} is independent on n, and the fourth column (D1-3) adjusts r (solid lines) or K (dotted lines) so that MSY is independent of n. The equations for the adjustments are listed in Table 1.

Schaefer model. Nevertheless, in many cases parameter estimates will not be available from the same model and transferring parameter values comes with some caveats and needs to be scrutinised. Borrowing parameters between models with different production curves will lead to different reference points as illustrated with the following example. Setting a prior for the intrinsic growth, r, with mean equal to 0.4 for a Fox production model ($n = 1$) based on a meta-analysis that used a Schaefer production model ($n = 2$) implies that F_{MSY} is twice as large as in the meta-analysis (0.4 instead of 0.2). To ensure that the F_{MSY} is kept the same, a prior mean of 0.2 should be used for the Fox model. In other words, F_{MSY} is transferable between models, but the intrinsic growth rate is not. The relationship between model parameters (r and K) and reference points can be approximated as equations that imply constant reference points for any $n > 1$ (Table 1). However, due to the correlation between r and K only one of the reference points can be kept constant and holding one constant might strengthen the effect of n to the others (Fig. 2B-D).

Not accounting for these relationships between parameters might have unintended consequences for the model fit and lead to biased reference points. In addition, it should be noted that different model implementations (such as a continuous time implementation versus one in discrete time), will lead to different parameter estimates for the same data. These differences are most pronounced for stocks with fast dynamics (high values of r), so extra care is advised for such stocks when using priors generated from other model implementations.

Another model parameter that is a suitable candidate for a prior from meta studies is the biomass process error. The biomass variability likely depends on the life history parameters of a species and their susceptibility to changes of environmental and ecological factors.

Table 1
Equations to adjust r or K to calculate reference points (F_{MSY}, B_{MSY}, MSY) for any $n > 1$. Where x is the n parameter of the meta study, y is the n parameter value of the assessment model (or mean of the prior), $\zeta = (1/n)^{(1/(n-1))}$, and $\gamma = n^{(n/(n-1))}/(n-1)$.

Reference point	Adjust r	Adjust K
F_{MSY}	$r_{n=y} = r_{n=x} n/x$	No effect
B_{MSY}	No effect	$K_{n=y} = K_{n=x} \quad \zeta_{n=x}/\zeta_{n=y}$
MSY	$r_{n=y} = r_{n=x} \quad \gamma_{n=y}/\gamma_{n=x}$	$K_{n=y} = K_{n=x} \quad \gamma_{n=y}/\gamma_{n=x}$

3.1.2. Priors from stock-specific (auxiliary) information

Separation of observation and process error is challenging and requires a large amount of data (De Valpine and Hilborn, 2005). In many cases the available data are insufficient and estimating simultaneously process and observation error parameters proves problematic. We suggest that observation errors of SPMs should be compared to those estimated by raw data to check if these estimates are reasonable. SPMs observation errors are not expected to be substantially lower than the raw data estimates (cf. Maunder and Punt, 2004). On the other hand, larger SPM estimates are plausible, e.g., when a survey index that does not sample the whole population distribution will potentially be noisier than estimated just by raw data. Circumventing this issue often requires including informative priors on each of the error terms based on available data or other stock assessments. For instance, a prior for the abundance index uncertainty could be based on the estimated coefficient of variation (CV) (or standard deviation (SD) on log scale) of the model used for the standardisation of the abundance index or commercial CPUE.

It is common to have an initial period where only catch information is available; in such situations, it may be necessary to provide additional information about the initial depletion level of the population to achieve model convergence. For instance, the use of this prior increased the convergence rate from 65% to up to 92% in the simulation study (Supplementary Fig S18). In SPiCT and JABBA, the prior on initial depletion is specified as the ratio of year-one biomass to carrying capacity ($B_{y=1}/K$). The choice of this prior should ideally be informed by knowledge of the fishery in the period preceding the start of the included time series. If there was no established fishery before the start of available data, it is reasonable to assume that the stock was close to carrying capacity, i.e., a prior for initial depletion with mean value close to one (e.g., 0.9) can be used. On the other hand, if there is knowledge of high catches an initial depletion prior with a lower mean should be used. We suggest choosing initial depletion priors with caution and to perform sensitivity analysis to investigate the effects of alternative choices, both on model estimates and on the short-term forecast; vague, uninformative priors should be preferred. The choice of the initial depletion prior can be very influential on the accuracy and precision of the estimated reference points and absolute states when the historical abundance index is missing. Relative states, on the other hand, seem to be more robust to misspecified initial depletion level priors when the historical abundance index is missing according to the simulation study (Supplementary Figures S18–19). However, when historical catches are missing, the initial depletion level prior can have unintended negative effects even on the relative states if misspecified and should, thus, only be used with caution (Supplementary Figures S20–21). When historical information is available, a misspecified prior for the initial depletion level with a standard deviation of 0.5 or larger did not affect the results negatively apart from an extreme depletion prior with mean 0.01 (Supplementary Figures S22–23).

3.2. Model performance diagnostics and model acceptance

All mathematical models are a simplified version of the real world and, therefore, are constructed on a set of underlying assumptions. The above-mentioned assumptions of SPMs must be carefully considered during all stages of model application and use, from the data preparation, model fitting, and scrutinising the results. Violation of the underlying assumptions can give misleading, biased, and/or meaningless results. By extending the model it is possible to relax or completely remove some of these assumptions. For instance, SPiCT can estimate time-varying productivity and/or time-varying carrying capacity (Mildenberger et al., 2020), and JABBA-Select can model selectivity changes and, thus, fleet specific estimates of exploitable stock biomass and reference points (Winker et al., 2020a), or using a Pella-Tomlinson hockey-stick composite model for surplus production, where the surplus production is equal to the Pella-Tomlinson model above a threshold of B/K and decreases linearly to zero when it is below that threshold (Winker et al., 2018).

Diagnostic tools are an important part of all modern implementations of assessment methods. Assessments should be checked for misspecifications and evaluated according to four criteria: model convergence, fit to the data, model consistency, and prediction skill (Carvalho et al., 2021). Similar sets of tests are now used around the world to scrutinise model fits before using them to provide management advice (e.g., in ICES areas: ICES, 2023c; in the United States: Karp et al., 2022). We consider the following seven key points as a good practices checklist supporting the acceptance or rejection of an assessment based on a state-space (Bayesian) SPM, such as SPiCT or JABBA. Any deviation from this list should be investigated before basing scientific advice on the SPM assessment.

3.2.1. Optimisation convergence

The process of estimating model parameters and random effects involves an optimisation procedure, which should converge to a well-defined optimum in case of maximum likelihood estimation or equilibrium distributions for MCMC fitting. It is crucial that all parameters, derived variables, and their variances should be numeric and finite, ensuring the validity of the estimation process. Failure to reach convergence could indicate mismatches between the available data and the model specification and additional scrutiny of the underlying assumptions is necessary. Additional or different assumptions can alleviate the issue, e.g., using a more informative prior. As always, restrictive priors should come from auxiliary information and sensitivity analysis should investigate the effects of choosing different priors to model estimates, derived quantities, and short-term projections.

3.2.2. Residuals

In state-space SPMs, two types of residuals are usually considered: residuals of the observation processes for the catches and abundance indices (referred to as observation residuals), and residuals of the biomass process and, if utilised, the fishing mortality process (referred to as process residuals). As the state estimates (and hence naive residuals) are correlated in time (Trijoulet et al., 2023), the calculation and testing of observation residuals from state-space models demands specific treatment depending on the estimation method used: for models using TMB (e.g., SPiCT) there is the one-step-ahead (OSA) method, for MCMC models (e.g., JABBA) the "one sample from the posterior" method, and for models using the Kalman filter, residuals should be based on state-estimates from running the filter forward, but without the backward smoothing step (Thygesen et al., 2017). Autocorrelation in the index observation residuals can arise if there are conflicts in the signals between multiple indices. Variance priors may be used to resolve this issue. Process residuals of the model should also be independent and normally distributed. Deviations indicate disagreement between model and available data, e.g., we expect problematic process residuals when we have time-varying processes and use a model that assumes them constant through time. In cases where the SPM models fishing mortality (F) as an unobserved process, the residuals of this process are less important than the biomass process residuals and more likely to be violated as management measures or fishers' behaviour could lead to abrupt changes in fishing effort and therefore fishing mortality, thus, violating the assumption of commonly assumed F processes, such as a random walk (e.g., Pedersen and Berg, 2017).

3.2.3. Prior – posterior distributions

While any prior should be carefully selected and based on the best a priori knowledge, it is important to evaluate the prior and posterior distributions. A posterior distribution for a given parameter (or derived variable) that is identical to the prior distribution indicates that there is no information about this parameter in the data and the prior determines the posterior. It is recommended to conduct sensitivity tests, i.e., contrasting results from alternative prior distributions. On the other hand, if the posterior distribution is far from the prior, this is not necessarily problematic, but may highlight that the *a priori* assumptions about parameter values oppose the information in available data. In a case where the standard deviation of the prior distribution has a big effect on the model estimates, it is suggested to conduct a sensitivity analysis using different standard deviation values for the prior distribution; the sensitivity analysis is even more important when the SD was arbitrarily determined.

3.2.4. Retrospective analysis

A retrospective analysis consists of a series of assessments with decreasing length of time-series and the investigation of final-year estimates. Retrospective biases are any "systematic inconsistency among a series of estimates of population size, or related assessment variables, based on increasing periods of data" (Mohn, 1999) and can reveal conflicts in observations or assumption violations. The retrospective trajectories of the relative fishing mortality (F/F_{MSY}) and relative biomass (B/B_{MSY}) should be inside the credible intervals of the base run (i.e., the run that includes the full length of the time-series) and there should not be any strong retrospective patterns, i.e., no tendency of consistent under- or overestimation in successive assessments. Although it is difficult to give a specific threshold for Mohn's ρ, Hurtado-Ferro et al. (2015) suggest rule-of-thumb thresholds depending on life-history characteristics of the stock; these thresholds are used by ICES (ICES, 2020b).

3.2.5. Hindcasting analysis

The hindcast follows a similar procedure to the retrospective analysis, where recent data are removed, and the model is refitted using the remaining data (Carvalho et al., 2021). Instead of unobserved quantities that are model-dependent, in the hindcast, observations are compared to their estimated counterparts in a model-free validation (Kell et al., 2016). In SPiCT and JABBA the hindcast is implemented so that the index observations are sequentially removed, whereas catch observations are included in the fitting procedure (cf. Kell et al., 2021). The ability of the model to predict the abundance index is quantified with the mean absolute scaled error (MASE), which measures how well the SPM predicts the index compared to a naïve forecast, i.e., that the index in the following year will be equal to the last observation (Hyndman and Koehler, 2006). Values higher than 1 indicate that the model is performing worse than the naïve predictor.

3.2.6. Jitter analysis

During maximum likelihood optimisation, initial parameter values should not affect the parameter estimates, provided there are no local optima in the likelihood function. A simple test, referred to here as jitter analysis, compares parameter estimates for a set of initial parameter values (Carvalho et al., 2021). When the jitter analysis is based on random initial values, non-converged runs can be neglected, as it is possible that a set of parameters is not appropriate as a starting point for the optimisation. Large deviations in the parameter estimates indicate that multiple local optima exist; it should be ensured that the global optimum is used, or unrealistic parameter values are avoided using vague priors.

3.2.7. Biologically meaningful parameter estimates and uncertainty

Estimated parameters and derived quantities have to be within the ranges that are expected according to our understanding of the life-history of the stock, i.e., which are biologically meaningful and not too uncertain. For instance, very skewed production curves, e.g., $B_{\mathrm{MSY}}/K < 0.5$, would imply infinite population growth at small stock size and are biologically difficult to explain. We caution against models that have very skewed production curves; we argue for models that have $n > 1$ in line with a more precautionary approach. This is also considering the two potential sources of positive biases in SPMs: (1) Ignoring the lag effects between reproduction and recruitment can lead to overly optimistic estimates of rebuilding times for depleted stocks under highly skewed instant surplus production expectations and (2) the typical smaller size at selectivity than maturation can lead to a disproportionally stronger decrease in spawning biomass than exploitable biomass (Winker et al., 2020a). On the other hand, it is important to note that SPMs model the exploitable stock biomass and, therefore, $B = 0$ does not imply that SSB or total stock biomass (TSB) or SSB are equal to zero. Thus, it is possible that a part of the SSB (or TSB) is not vulnerable due to, e.g., gear selectivity or spatial extent of the fishery, which theoretically could lead to large surplus production even when the exploitable stock biomass is very low. High parameter uncertainty, could indicate a lack of contrast in the input data or violation of assumptions, e.g., using constant parameters for a population that has changed productivity over time. If the 95% credible intervals for the estimated stock status, B/B_{MSY} and F/F_{MSY}, span more than one order of magnitude, the assessment is likely too uncertain to be used as basis for scientific advice. Further, a population in the stochastic Pella-Tomlinson model cannot persist with a large biomass process error and the upper limit of process error depends on the intrinsic growth rate r and the shape parameter n (Bordet and Rivest, 2014), therefore, large process error estimates require extra scrutiny. Alternative or additional assumptions can help reducing the estimated uncertainty; it is of course still crucial that best available expert judgement is used to formulate such assumptions and sensitivity of the results to alternative models should be checked.

3.3. Choice of model, and sensitivity analysis

Model configurations and performance diagnostics play an important role in the acceptance and evaluation of assessments with SPMs and the selection of the best fitting model configuration. In some cases, alternative model configurations pass the model diagnostics and provide meaningful and equally likely results. While from an ecological and scientific point of view these alternative models are relevant and interesting, fisheries management usually requires a single best model (e.g., ICES, 2022a). In cases where multiple competing models pass all diagnostic tests but imply differences in estimated stock status and lead to very different catch advice, it is often difficult to choose a single best model. When all input data are the same between competing models, information criteria can be used to objectively choose the best model (e.g., Akaike Information Criterion, Akaike, 1998); "same data" here includes the input time series of observation and used priors. If management objectives prioritise the precautionary approach, the model leading to the most precautionary advice could be chosen. Otherwise, approaches that use multiple models as a range of uncertainty (e.g., in Japan stock assessment using SPMs used multiple base case models; Chiba et al., 2023) or ensemble modelling approaches (Anderson et al., 2017) can be used. In most cases, advice based on an assessment model will be more robust than alternative trend-based empirical rules for giving advice on catch quotas (Jardim et al., 2015), and especially compared to catch-only methods (Ovando et al., 2022). In addition to scenarios with different prior assumptions, alternative scenarios should be considered that explore different options, such as competing assumptions regarding the qualitative relative uncertainty weights, or different fixed parameter values, e.g., exploring the Schaefer or Fox models if the shape of the production curve is not freely estimated.

4. Management advice

4.1. Biological reference points

Management is based on biological reference points that act as targets, thresholds, or limits. SPMs directly define target reference points that aim at maximising the yield while not compromising the long-term survival of the stock, relating to catch (MSY), fishing pressure (F_{MSY}) and exploitable stock biomass (B_{MSY}). In state-space implementations of SPMs, where observation and process errors are estimated, it is recommended to use stochastic versions of these reference points that include correction factors that depend on estimated error of the biomass process (σ_B). The stochastic F_{MSY} will be lower than the deterministic one and thus becomes more precautionary with increasing process error, but only when $n > 1$ (Pedersen and Berg, 2017). Therefore, it is recommended to use stochastic reference points when considering models with shape parameter $n > 1$ and deterministic ones otherwise. Although maximising the yield is traditionally the aim of fisheries management, more precautionary targets might be considered, e.g., in Australia the biomass that maximises the economic yield (B_{MEY}) is used as the target reference point and is approximated as $1.2B_{\mathrm{MSY}}$ (DAWR, 2018).

Additional to target reference points, fisheries management relies on threshold and limit reference points (e.g., ICES, 2021b). These are set to prevent the stock declining to levels where the population is at reduced reproductive capacity or at risk of collapsing. When the biomass of the population is below the threshold reference point ($B_{threshold}$) increased precaution is suggested, e.g., lower catch advice or advising lower fishing effort. When the biomass is below the limit reference point (B_{lim}) closure of the fishery should be considered. There are three common ways to derive biomass limit reference points for SPMs. First, biomass threshold and limit reference points can be defined as fractions of B_{MSY}. Mildenberger et al. (2022) suggest that both $B_{threshold}$ and B_{lim} should depend on a common reference point, e.g., B_{MSY}. Second, these reference points can be defined as a fraction of the carrying capacity K with commonly suggested values around 0.2 K for B_{lim} and 0.48 K for $B_{threshold}$ (DAWR, 2018; Dichmont et al., 2017; Winker et al., 2018). The third way to define B_{lim} is based on the productivity of the stock, e.g., B_{lim} is defined as the biomass where the population can sustain half of the yield compared to MSY; this corresponds to $0.3B_{MSY}$ for the Schafer model (ICES, 2013).

4.2. Harvest control rule

An accepted assessment with a surplus production model does not only indicate the current stock status, but can also be used to derive fisheries management advice, such as the total allowable catch (TAC), most often based on a short-term forecast (ICES, 2022a). These catch quotas can be based on target reference points directly, such as the predicted catch corresponding to $F = F_{MSY}$ or incorporate more complex rules with additional threshold and limit reference points and uncertainty buffers. A commonly used example of a more complex harvest control rule (HCR) is the hockey-stick rule, where the target fishing mortality is equal to F_{MSY} when the population is above a threshold reference point $B_{threshold}$, e.g., MSY $B_{trigger}$ used in ICES (ICES, 2021b), is linearly reduced from F_{MSY} to zero when the biomass in the last year (B_{last}) is below that threshold ($F_{target} = F_{MSY}\ B_{last}/B_{threshold}$), and is zero when the biomass in the last year is below a limit reference point (B_{lim}). Ideally, HCRs should be based on relative states (e.g., $B/B_{threshold}$) rather than absolute states (e.g., B) and absolute reference points; absolute estimates are typically more uncertain than their relative counterparts (Pedersen and Berg, 2017). The simulation study presented here revealed that the median relative error over all scenarios was around +/- 6.5% for the relative states and around +/- 25% for the absolute states and reference points (Supplementary Table S1). Besides a target fishing mortality rate and threshold and limit reference points in the HCR, good advisory practices demand to incorporate an uncertainty buffer (Mildenberger et al., 2022). Uncertainty buffers refer to fractiles of projected distributions of catch and B/B_{MSY} that are lower than the median estimate, or in the case of F/F_{MSY} higher than the median. Uncertainty buffers, therefore, incorporate the estimated assessment uncertainty in the projected TAC and are more precautionary. Higher biomass thresholds and uncertainty buffers can lead to the same long-term yield while leading to more precautionary fisheries management advice (Mildenberger et al., 2022). Within ICES, a hockey-stick HCR as described above is used that has $B_{threshold} = 0.5\ B_{MSY}$ and uses the 35th percentile of the predicted catch distribution, with a biomass limit reference point depending on B_{MSY} (ICES, 2022b). A lower percentile (15th) has been recommended for less productive and more sensitive species, such as elasmobranchs (ICES, 2023b). Here, we covered some generic considerations based on large simulation studies of artificial populations with very different life-history characteristics (e.g., ICES, 2023b, 2020a; Mildenberger et al., 2022). Ideally, a HCR should be selected after evaluation of performances among alternative candidate rules with management strategy evaluation (MSE) parameterised as closely as possible to the stock that is assessed (Mildenberger et al., 2022). MSE is an important tool to evaluate the performance, input data, and assumptions of SPMs and associated harvest control rules (Punt et al., 2016).

5. Recommendations for future research

We are confident that SPMs will remain an important component of the future stock assessment toolbox. Not least for needed management advice on non-targeted stocks and stocks caught as bycatch in mixed fisheries without detailed sampling protocols, or stocks in regions where the collection of age and length information is infeasible hindering the application of more complex age or length structured models. When good practice guidelines as outlined in this paper are followed, SPMs can be the basis of effective management that leads to high yields while being precautionary. Although SPMs could be used as alternatives to age- or length-based models as basis for the advice, we are not suggesting that they should replace them as SPMs are not able to detect changes in population structure, often have higher uncertainty and likely lead to more conservative advice to maintain risk equivalency to data-rich assessments (Roux et al., 2022). Although the development of SPMs and good practice guidelines has increased substantially during recent years, there are still areas for improvement regarding model functionality as well as their application. In the following, we focus on the main limitations of SPMs and outline possibilities and avenues for future research to overcome them.

5.1. Limitations regarding model functionality

Most SPMs assume time-invariant parameters and model density dependence as a simple, often symmetric, function of biomass, and ignore possible detailed information about the population, its environment, and the fishery. These assumptions are likely to be violated for many stocks and, therefore, future research should investigate the effects of violating them. Spatially explicit SPMs have been developed in the past, but they should be evaluated using simulation testing in their ability to be used for management and their performance on reaching management objectives (Thorson et al., 2017). Multispecies surplus production models that aggregate stocks within a region and can estimate multispecies reference points can be found in the literature (Mueter and Megrey, 2006), but it's an open question how these can be used to inform management decisions.

Despite frequently being employed in data-moderate situations, where all available information should be utilised, stock assessments with SPMs often neglect certain available information. This includes the non-exploitable part of the abundance index, incomplete discard and recreational catches, sporadic length information, and fishing effort information. Both data pre-processing and the specification of selectivity curves, which determine the fraction of the biomass that is exploitable, could be informed by sporadically sampled length-frequency data obtained from fishery-dependent and independent sources. While JABBA-Select allows for the inclusion of multiple fleets (Winker et al., 2020a), future research efforts could allow for these fleets to represent discarded fish to improve accuracy and precision of forecasting.

5.2. Limitations regarding application of SPMs

Although the functionality for SPMs with time-varying parameters, such as productivity, catchability, or selectivity are available (Mildenberger et al., 2020; Nesslage and Wilberg, 2019; Winker et al., 2020a), their application is the exception. This is likely due to the increased data requirements for the added flexibility, as well as the challenge of establishing priors for the time-varying process parameters. Future research should investigate and contribute to a better understanding SPMs with time-varying processes and to facilitate their application. Another approach to inform time-varying processes might be to link them to environmental or ecological covariates, such as sea surface temperature, or to abundance of a predator or prey (e.g., CLIMPROD software, Fréon et al., 1993).

As previously mentioned, SPMs often suffer from excessive uncertainty due to short input time-series with limited contrast. This uncertainty may be reduced if the parameter space of the model can be restricted. We recommend that more meta-analysis studies (such as in Thorson, 2020; Thorson et al., 2012), are undertaken to overcome issues of limited data and get insights on plausible ranges for model parameters. Using this information, we can formulate priors or parameter bounds that can help assess many stocks that have the required input data (catch, abundance index) but lack enough contrast to get an acceptable assessment and base management advice on them.

5.3. Management advice

Future research should also revisit meaningful definitions of threshold and limit reference points for HCRs based on SPMs. These reference points could be based on data-rich reference points, or target reference points defined as threshold/limit reference points. HCRs based on SPMs that use a reference period and trends rather than reference points could be useful in cases where reference points are not identifiable due to a lack of contrast in data.

Identifying the most appropriate HCR for a stock is crucial for effectively managing a stock, aiming to optimise yield while minimising the risk of overexploitation. MSE is an invaluable tool for selecting the best HCR, ideally incorporating an operating model that is capable of simulating population dynamics with a high degree of realism, based on age or length. However, in cases where detailed age or length data are lacking, and knowledge of essential life-history parameters is limited, it is difficult to set up a realistic operating model. SPMs could be used in a full simulation loop, i.e., used both as the operating and the assessment model. SPiCT and JABBA estimate the observation and process error and therefore could be used to condition operating models, i.e., parameterising the operating model to follow the stock assessment closely (Chiba et al., 2023), thus providing realistic estimates of uncertainty for management purposes.

CRediT authorship contribution statement

C.W. Berg: Conceptualization, Writing – review & editing. **M.S. Kapur:** Writing – review & editing. **A. Kokkalis:** Conceptualization, Funding acquisition, Methodology, Software, Writing – original draft, Writing – review & editing. **M. Miyagawa:** Writing – review & editing. **W. Medeiros-Leal:** Writing – review & editing. **J.R. Nielsen:** Writing – review & editing. **T.K. Mildenberger:** Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **H. Winker:** Writing – review & editing. **N.S. Jacobsen:** Writing – review & editing. **M.H. Taylor:** Writing – review & editing. **M. Ichinokawa:** Writing – review & editing.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgement

AK, CWB, TKM acknowledge funding by the EMFF project "Robust management procedures and advice for data limited stocks (RoMA)" (Ref 33113-B-20–183), which is funded by the European Maritime and Fisheries Fund and the Danish Fisheries Agency. WML was funded by FCT Ph.D. fellowship (ref. UI/BD/153596/2022). The authors would like to acknowledge the inspiring discussions during ICES WKLIFE XI and ICES benchmark assessment meetings. We would like to thank André Punt and an anonymous reviewer for insightful comments and suggestions to an earlier version of this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fishres.2024.107010.

References

Aalto, E.A., Dick, E.J., MacCall, A.D., 2015. Separating recruitment and mortality time lags for a delay-difference production model. Can. J. Fish. Aquat. Sci. 72, 161–165. https://doi.org/10.1139/cjfas-2013-0415.

Aeberhard, W.H., Flemming, J.M., Nielsen, A., 2018. Review of state-space models for fisheries science. Annu Rev. Stat. Appl. 5, 215–235. https://doi.org/10.1146/annurev-statistics-031017-100427.

Akaike, H., 1998. Information Theory and an Extension of the Maximum Likelihood Principle. In: Parzen, E., Tanabe, K., Kitagawa, G. (Eds.), Selected Papers of Hirotugu Akaike. Springer New York, New York, NY, pp. 199–213. https://doi.org/10.1007/978-1-4612-1694-0_15.

Anderson, S.C., Cooper, A.B., Jensen, O.P., Minto, C., Thorson, J.T., Walsh, J.C., Afflerbach, J., Dickey-Collas, M., Kleisner, K.M., Longo, C., Osio, G.C., Ovando, D., Mosqueira, I., Rosenberg, A.A., Selig, E.R., 2017. Improving estimates of population status and trend with superensemble models. Fish Fish 18, 732–741. https://doi.org/10.1111/faf.12200.

Berg, C.W., Nielsen, A., 2016. Accounting for correlated observations in an age-based state-space stock assessment model. ICES J. Mar. Sci. 73, 1788–1797. https://doi.org/10.1093/icesjms/fsw046.

Best, J.K., Punt, A.E., 2020. Parameterizations for Bayesian state-space surplus production models. Fish. Res 222, 105411. https://doi.org/10.1016/j.fishres.2019.105411.

Blasco, G.D., Ferraro, D.M., Cottrell, R.S., Halpern, B.S., Froehlich, H.E., 2020. Substantial gaps in the current fisheries data landscape. Front Mar. Sci. 7, 1–12. https://doi.org/10.3389/fmars.2020.612831.

Boettiger, C., Lang, D.T., Wainwright, P.C., 2012. Rfishbase: exploring, manipulating and visualizing FishBase data from R. J. Fish. Biol. 81, 2030–2039. https://doi.org/10.1111/j.1095-8649.2012.03464.x.

Bordet, C., Rivest, L.-P., 2014. A stochastic Pella Tomlinson model and its maximum sustainable yield. J. Theor. Biol. 360, 46–53. https://doi.org/10.1016/j.jtbi.2014.06.012.

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K.R., Maunder, M.N., Taylor, I., Wetzel, C.R., Doering, K., Johnson, K.F., Methot, R.D., 2021. A cookbook for using model diagnostics in integrated stock assessments. Fish. Res 240, 105959. https://doi.org/10.1016/j.fishres.2021.105959.

Chang, Y.-J., Winker, H., Sculley, M., Hsu, J., 2020. Evaluation of the status and risk of overexploitation of the Pacific billfish stocks considering non-stationary population processes. Deep Sea Res. Part II: Top. Stud. Oceanogr. 175, 104707 https://doi.org/10.1016/j.dsr2.2019.104707.

Chiba, S., Sato, R., Morita, S., Sakai, O., Ichinokawa, M., Hamatsu, T., 2023. Stock assessment and evaluation for the northern Hokkaido stock of pointhead flounder (fiscal year 2023). Marine fisheries stock assessment and evaluation for Japanese waters (in Japanese). Tokyo, Japan.

Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., Lester, S.E., 2012. Status and Solutions for the World's Unassessed Fisheries. Science 338 (1979), 517–520. https://doi.org/10.1126/science.1223389.

Cousido-Rocha, M., Pennino, M.G., Izquierdo, F., Paz, A., Lojo, D., Tifoura, A., Zanni, M. Y., Cerviño, S., 2022. Surplus production models: a practical review of recent approaches. Rev. Fish. Biol. Fish. 32, 1085–1102. https://doi.org/10.1007/s11160-022-09731-w.

DAWR, 2018. Commonw. Fish. Harvest Strategy Policy.

De Valpine, P., Hilborn, R., 2005. State-space likelihoods for nonlinear fisheries time-series. Can. J. Fish. Aquat. Sci. 62, 1937–1952. https://doi.org/10.1139/f05-116.

Dichmont, C.M., Fulton, E.A., Gorton, R., Sporcic, M., Little, L.R., Punt, A.E., Dowling, N., Haddon, M., Klaer, N., Smith, D.C., 2017. From data rich to data-limited harvest strategies—does more data mean better management? ICES J. Mar. Sci. 74, 670–686. https://doi.org/10.1093/icesjms/fsw199.

Eigaard, O.R., Marchal, P., Gislason, H., Rijnsdorp, A.D., 2014. Technological development and fisheries management. Rev. Fish. Sci. Aquac. 22, 156–174. https://doi.org/10.1080/23308249.2014.899557.

Finney, D.J., 1941. On the distribution of a variate whose logarithm is normally distributed. Suppl. J. R. Stat. Soc. 7, 155. https://doi.org/10.2307/2983663.

Fletcher, R.I., 1978. On the restructuring of the Pella-Tomlinson system. U. S. Fish. Bull. 76, 515–534.

Fox, W.W., 1970. An exponential surplus-yield model for optimizing exploited fish populations. Trans. Am. Fish. Soc. 99, 80–88 https://doi.org/10.1577/1548-8659(1970)99<80:aesmfo>2.0.co;2.

Free, C.M., Jensen, O.P., Anderson, S.C., Gutierrez, N.L., Kleisner, K.M., Longo, C., Minto, C., Osio, G.C., Walsh, J.C., 2020. Blood from a stone: performance of catch-only methods in estimating stock biomass status. Fish. Res 223, 105452. https://doi.org/10.1016/j.fishres.2019.105452.

Fréon, P., 1986. Introduction of environmental variables into global production models. In: Wyatt, T., Larrañeta, M.G. (Eds.), Long Term Changes in Marine Fish Populations. Inst. de Investigaciones Marinas, Vigo, Spain.
Fréon, P., Mullon, C., Pichon, G., 1993. CLIMPROD: experimental interactive software for choosing and fitting surplus production models including environmental variables. Computerized Information Series: Fisheries (FAO).
Froese, R., Pauly, D., 2023. FishBase [WWW Document]. URL ⟨https://www.fishbase.org⟩.
González Herraiz, I., Vila, Y., Cardinale, M., Berg, C.W., Winker, H., Azevedo, M., Mildenberger, T.K., Kokkalis, A., Vázquez Vilamea, A.A., Morlán, R., Somavilla, R., Pennino, M.G., 2023. First Maximum Sustainable Yield advice for the Nephrops norvegicus stocks of the Northwest Iberian coast using stochastic Surplus Production model in Continuous Time (SPiCT). Front Mar. Sci. 10, 1–15. https://doi.org/10.3389/fmars.2023.1062078.
Grüss, A., McKenzie, J.R., Lindegren, M., Bian, R., Hoyle, S.D., Devine, J.A., 2023. Supporting a stock assessment with spatio-temporal models fitted to fisheries-dependent data. Fish. Res 262, 106649. https://doi.org/10.1016/j.fishres.2023.106649.
Grüss, A., Walter, J.F., Babcock, E.A., Forrestal, F.C., Thorson, J.T., Lauretta, M.V., Schirripa, M.J., 2019. Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. Fish. Res 213, 75–93. https://doi.org/10.1016/j.fishres.2019.01.008.
Hamabe, K., Matsuoka, K., Kitakado, T., 2023. Estimation of abundance and population dynamics of the Antarctic blue whale in the Antarctic Ocean south of 60°S, from 70°E to 170°W. Mar. Mamm. Sci. 39, 671–687. https://doi.org/10.1111/mms.13006.
Hilborn, R., 1990. Estimating the parameters of full age-structured models from catch and abundance data. Bull. Int. North Pac. Fish. Commn. 50, 207–213.
Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment. Quantitative Fisheries Stock Assessment. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4615-3598-0.
Hilborn, R., Amoroso, R.O., Anderson, C.M., Baum, J.K., Branch, T.A., Costello, C., De Moor, C.L., Faraj, A., Hively, D., Jensen, O.P., Kurota, H., Little, L.R., Mace, P., McClanahan, T., Melnychuk, M.C., Minto, C., Osio, G.C., Parma, A.M., Pons, M., Segurado, S., Szuwalski, C.S., Wilson, J.R., Ye, Y., 2020. Effective fisheries management instrumental in improving fish stock status. Proc. Natl. Acad. Sci. USA 117, 2218–2224. https://doi.org/10.1073/pnas.1909726116.
Hoyle, S.D., Campbell, R.A., Ducharme-Barth, N.D., Grüss, A., Moore, B.R., Thorson, J.T., Tremblay-Boyer, L., Winker, H., Zhou, S., Maunder, M.N., 2024. Catch per unit effort modelling for stock assessment: a summary of good practices. Fish. Res 269, 106860. https://doi.org/10.1016/j.fishres.2023.106860.
Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., Punt, A.E., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES J. Mar. Sci. 72, 99–110. https://doi.org/10.1093/icesjms/fsu198.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int J. Forecast 22, 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001.
ICCAT, 2017a. REPORT OF THE 2017 ICCAT ALBACORE SPECIES GROUP INTERSESSIONAL MEETING (INCLUDING ASSESSMENT OF MEDITERRANEAN ALBACORE) (No. 74(2)), Collect. Vol. Sci. Pap. Madrid, Spain.
ICCAT, 2017b. REPORT OF THE 2017 ICCAT ATLANTIC SWORDFISH STOCK ASSESSMENT SESSION, ATLANTIC SWORDFISH STOCK ASSESSMENT SESSION. Madrid, Spain.
ICES, 2013. Report of the Benchmark Workshop on Greenland Halibut Stocks (WKBUT), 26–29 November 2013, Copenhagen, Denmark. 26–29.
ICES, 2019. Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). https://doi.org/10.17895/ices.pub.5402.
ICES, 2020a. Tenth Workshop on the Development of Quantitative Assessment Methodologies based on LIFE-history traits, exploitation characteristics, and other relevant parameters for data-limited stocks (WKLIFE X). ICES Scientific Reports 2, 72 pp. https://doi.org/10.17895/ices.pub.5985.
ICES, 2020b. Workshop on Catch Forecast from Biased Assessments (WKFORBIAS; outputs from 2019 meeting), ICES Scientific Reports. https://doi.org/10.17895/ices.pub.5997.
ICES, 2021a. Benchmark Workshop on the development of MSY advice for category 3 stocks using Surplus Production Model in Continuous Time; SPiCT (WKMSYSPiCT). ICES Scientific Reports 3, 316pp.
ICES, 2021b. ICES fisheries management reference points for category 1 and 2 stocks; Technical Guidelines, Report of the ICES Advisory Committee, 2021. ICES Advice 2021, Section 16.4.3.1. https://doi.org/10.17895/ices.pub.3036.
ICES, 2022a. Advice on fishing opportunities (2022). General ICES Advice guidelines. Report. 1–11.
ICES, 2022b. ICES technical guidance for harvest control rules and stock assessments for stocks in categories 2 and 3, Report of ICES Advisory Committee, 2022. ICES Advice 2022, Section 16.4.11. https://doi.org/10.17895/ices.advice.19801564.
ICES, 2023a. Benchmark workshop 2 on the development of MSY advice using SPiCT (WKBMSYSPICT2).
ICES, 2023b. Eleventh workshop on the development of quantitative assessment methodologies based on LIFE-history traits, exploitation characteristics, and other relevant parameters for data-limited stocks (WKLIFE XI). ICES Sci. Rep. 5, 21. https://doi.org/10.17895/ices.pub.22140260.
ICES, 2023c. ICES Guidelines for Benchmarks. Version 1. ICES Guidelines and Policies - Advice Technical Guidelines. https://doi.org/10.17895/ices.pub.22316743.
ISC, 2017. Stock Assessment and Future Projections of Blue Shark in the North Pacific Ocean through 2015, International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean. Vancouver, Canada.
Jardim, E., Azevedo, M., Brites, N.M., 2015. Harvest control rules for data limited stocks using length-based reference points and survey biomass indices. Fish. Res. 171, 12–19. https://doi.org/10.1016/j.fishres.2014.11.013.
Kai, M., Yokoi, H., 2019. Performance evaluation of information criteria for estimating a shape parameter in a Bayesian state-space biomass dynamics model. Fish. Res. 219, 105326 https://doi.org/10.1016/j.fishres.2019.105326.
Kapur, M.R., Fitchett, M.D., Yau, A.J., Carvalho, F., 2019. 2018 Benchmark Stock Assessment of Main Hawaiian Islands Kona Crab. https://doi.org/10.25923/7wf2-f040.
Karp, M.A., Kuriyama, P., Blackhart, K., Brodziak, J., Carvalho, F., Curti, K., Dick, E.J., Hanselman, D., Hennen, D., Ianelli, J., Sagarese, S., Shertzer, K., Taylor, I., 2022. Common model diagnostics for fish stock assessments in the United States.
Kell, L.T., Kimoto, A., Kitakado, T., 2016. Evaluation of the prediction skill of stock assessment using hindcasting. Fish. Res. 183, 119–127. https://doi.org/10.1016/j.fishres.2016.05.017.
Kell, L.T., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., Fu, D., 2021. Validation of stock assessment methods: Is it me or my model talking? ICES J. Mar. Sci. 78, 2244–2255. https://doi.org/10.1093/icesjms/fsab104.
Ludwig, D., Walters, C.J., 1985. Are age-structured models appropriate for catch-effort data? Can. J. Fish. Aquat. Sci. 42, 1066–1072. https://doi.org/10.1139/f85-132.
Ludwig, D., Walters, C.J., 1989. A Robust method for parameter estimation from catch and effort data. Can. J. Fish. Aquat. Sci. 46, 137–144. https://doi.org/10.1139/f89-018.
Maunder, M.N., 2003. Is it time to discard the Schaefer model from the stock assessment scientist's toolbox? Fish. Res 61, 145–149. https://doi.org/10.1016/S0165-7836(02)00273-4.
Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res 70, 141–159. https://doi.org/10.1016/j.fishres.2004.08.002.
Maunder, M.N., Hamel, O.S., Lee, H.-H., Piner, K.R., Cope, J.M., Punt, A.E., Ianelli, J.N., Castillo-Jordán, C., Kapur, M.S., Methot, R.D., 2023. A review of estimation methods for natural mortality and their performance in the context of fishery stock assessment. Fish. Res 257, 106489. https://doi.org/10.1016/j.fishres.2022.106489.
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. https://doi.org/10.1016/j.fishres.2012.10.012.
Mildenberger, T.K., Berg, C.W., Pedersen, M.W., Kokkalis, A., Nielsen, J.R., 2020. Time-variant productivity in biomass dynamic models on seasonal and long-term scales. ICES J. Mar. Sci. 77, 174–187. https://doi.org/10.1093/icesjms/fsz154.
Mildenberger, T.K., Berg, C.W., Kokkalis, A., Hordyk, A.R., Wetzel, C., Jacobsen, N.S., Punt, A.E., Nielsen, J.R., 2022. Implementing the precautionary approach into fisheries management: biomass reference points and uncertainty buffers. Fish Fish 23, 73–92. https://doi.org/10.1111/faf.12599.
Mohn, R., 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56, 473–488. https://doi.org/10.1006/jmsc.1999.0481.
Moore, J.E., Barlow, J., 2011. Bayesian state-space model of fin whale abundance trends from a 1991-2008 time series of line-transect surveys in the California Current. J. Appl. Ecol. 48, 1195–1205. https://doi.org/10.1111/j.1365-2664.2011.02018.x.
Mueter, F.J., Megrey, B.A., 2006. Using multi-species surplus production models to estimate ecosystem-level maximum sustainable yields. Fish. Res 81, 189–201. https://doi.org/10.1016/j.fishres.2006.07.010.
Nesslage, G.M., Wilberg, M.J., 2019. A performance evaluation of surplus production models with time-varying intrinsic growth in dynamic ecosystems. Can. J. Fish. Aquat. Sci. 76, 2245–2255. https://doi.org/10.1139/cjfas-2018-0292.
Neubauer, P., Jensen, O.P., Hutchings, J.A., Baum, J.K., 2013. Resilience and recovery of overexploited marine populations. Science 340 (1979), 347–349. https://doi.org/10.1126/science.1230441.
Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res 158, 96–101. https://doi.org/10.1016/j.fishres.2014.01.014.
Nielsen, J.R., 2015. Methods for integrated use of fisheries research survey information in understanding marine fish population ecology and better management advice (Monograph. Dutch Dr. Sc. and PhD Thesis). Wageningen University, Wageningen, NL.
Nielsen, J.R., Sparre, P.J., Hovgrd, H., Frost, H., Tserpes, G., 2006. Effort and capacity-based fisheries management. In: Motos, L., Wilson, D. (Eds.), The Knowledge Base for Fisheries Management. Developments in Aquaculture and Fisheries Science. Elsevier, pp. 163–216. https://doi.org/10.1016/S0167-9309(06)80010-4.
Omori, K.L., Hoenig, J.M., Luehring, M.A., Baier-Lockhart, K., 2016. Effects of underestimating catch and effort on surplus production models. Fish. Res 183, 138–145. https://doi.org/10.1016/j.fishres.2016.05.021.
Ovando, D., Free, C.M., Jensen, O.P., Hilborn, R., 2022. A history and evaluation of catch-only stock assessment models. Fish Fish 23, 616–630. https://doi.org/10.1111/faf.12637.
Pedersen, M.W., Berg, C.W., 2017. A stochastic surplus production model in continuous time. Fish Fish 18, 226–243. https://doi.org/10.1111/faf.12174.
Pedersen, M.W., Berg, C.W., Thygesen, U.H., Nielsen, A., Madsen, H., 2011. Estimation methods for nonlinear state-space models in ecology. Ecol. Model. 222, 1394–1400. https://doi.org/10.1016/j.ecolmodel.2011.01.007.
Pella, J.J., Tomlinson, P.K., 1969. A generalized stock production model. Inter-Am. Trop. Tuna Comm. 13, 421–488.

Polacheck, T., Hilborn, R., Punt, A.E., 1993. Fitting surplus production models: comparing methods and measuring uncertainty. Can. J. Fish. Aquat. Sci. 50, 2597–2607. https://doi.org/10.1139/f93-284.

Prager, M.H., 1994. A suite of extensions to a nonequilibrium surplus-production model. Fish. Bull. 92, 374–389.

Prager, M.H., 2002. Comparison of logistic and generalized surplus-production models applied to swordfish, Xiphias gladius, in the north Atlantic Ocean. Fish. Res. 58, 41–57. https://doi.org/10.1016/S0165-7836(01)00358-7.

Punt, A.E., 2003. Extending production models to include process error in the population dynamics. Can. J. Fish. Aquat. Sci. 60, 1217–1228. https://doi.org/10.1139/f03-105.

Punt, A.E., 2023. Those who fail to learn from history are condemned to repeat it: a perspective on current stock assessment good practices and the consequences of not following them. Fish. Res. 261, 106642 https://doi.org/10.1016/j.fishres.2023.106642.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016. Management strategy evaluation: best practices. Fish Fish 17, 303–334. https://doi.org/10.1111/faf.12104.

Punt, A.E., Castillo-Jordán, C., Hamel, O.S., Cope, J.M., Maunder, M.N., Ianelli, J.N., 2021. Consequences of error in natural mortality and its estimation in stock assessment models. Fish. Res. 233, 105759 https://doi.org/10.1016/j.fishres.2020.105759.

Roux, M.-J., Duplisea, D.E., Hunter, K.L., Rice, J., 2022. Consistent risk management in a changing world: risk equivalence in fisheries and other human activities affecting marine resources and ecosystems. Front. Clim. 3, 1–14. https://doi.org/10.3389/fclim.2021.781559.

Rufener, M.C., Kristensen, K., Nielsen, J.R., Bastardie, F., 2021. Bridging the gap between commercial fisheries and survey data to model the spatiotemporal dynamics of marine species. Ecol. Appl. 31, 1–15. https://doi.org/10.1002/eap.2453.

Schaefer, M.B., 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Bull. Inter-Am. Trop. Tuna Comm. 1, 27–56. https://doi.org/10.1016/S0092-8240(05)80049-7.

Schaefer, M.B., 1957. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical pacific ocean. Bull. Inter-Am. Trop. Tuna Comm.

Schnute, J.T., Richards, L.J., 2002. Surplus Production Models. In: Hart, P.J.B., Reynolds, J.D. (Eds.), Handbook of Fish Biology and Fisheries, Volume 2. Blackwell Science Ltd, Oxford, UK, pp. 105–126. https://doi.org/10.1002/9780470693919.ch6.

Szuwalski, C., 2019. Comment on "Impacts of historical warming on marine fisheries production.". Science 365 (1979), 329–335. https://doi.org/10.1126/science.aax5721.

Thorson, J.T., 2020. Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. Fish Fish 21, 237–251. https://doi.org/10.1111/faf.12427.

Thorson, J.T., Cope, J.M., 2017. Uniform, uninformed or misinformed?: The lingering challenge of minimally informative priors in data-limited Bayesian stock assessments. Fish. Res 194, 164–172. https://doi.org/10.1016/j.fishres.2017.06.007.

Thorson, J.T., Jannot, J., Somers, K., 2017. Using spatio-temporal models of population growth and movement to monitor overlap between human impacts and fish populations. J. Appl. Ecol. 54, 577–587. https://doi.org/10.1111/1365-2664.12664.

Thorson, J.T., Cope, J.M., Branch, T.A., Jensen, O.P., 2012. Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. Can. J. Fish. Aquat. Sci. 69, 1556–1568. https://doi.org/10.1139/F2012-077.

Thorson, J.T., Fonner, R., Haltuch, M.A., Ono, K., Winker, H., 2016. Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data. Can. J. Fish. Aquat. Sci. 74, 1794–1807. https://doi.org/10.1139/cjfas-2015-0598.

Thorson, J.T., Maureaud, A.A., Frelat, R., Mérigot, B., Bigman, J.S., Friedman, S.T., Palomares, M.L.D., Pinsky, M.L., Price, S.A., Wainwright, P., 2023. Identifying direct and indirect associations among traits by merging phylogenetic comparative methods and structural equation models. Methods Ecol. Evol. 2023 1259–1275. https://doi.org/10.1111/2041-210X.14076.

Thygesen, U.H., Albertsen, C.M., Berg, C.W., Kristensen, K., Nielsen, A., 2017. Validation of ecological state space models using the Laplace approximation. Environ. Ecol. Stat. 24, 317–339. https://doi.org/10.1007/s10651-017-0372-4.

Trijoulet, V., Albertsen, C.M., Kristensen, K., Legault, C.M., Miller, T.J., Nielsen, A., 2023. Model validation for compositional data in stock assessment models: calculating residuals with correct properties. Fish. Res 257. https://doi.org/10.1016/j.fishres.2022.106487.

Ulrich, C., Wilson, D.C.K., Nielsen, J.R., Bastardie, F., Reeves, S.A., Andersen, B.S., Eigaard, O.R., 2012. Challenges and opportunities for fleet- and métier-based approaches for fisheries management under the European Common Fishery Policy. Ocean Coast Manag 70, 38–47. https://doi.org/10.1016/j.ocecoaman.2012.06.002.

UN, 2003. Johannesburg Declaration on Sustainable Development and Plan of Implementation of the World Summit on Sustainable Development: the final text of agreements negotiated by Governments at the World Summit on Sustainable Development, 26 August-4 September 200. UN, [New York]:. 2003.

van Gemert, R., Koemle, D., Winkler, H., Arlinghaus, R., 2022. Data-poor stock assessment of fish stocks co-exploited by commercial and recreational fisheries: applications to pike Esox lucius in the western Baltic Sea. Fish. Manag Ecol. 29, 16–28. https://doi.org/10.1111/fme.12514.

Williams, E.H., Prager, M.H., 2002. Comparison of equilibrium and nonequilibrium estimators for the generalized production model. Can. J. Fish. Aquat. Sci. 59, 1533–1552. https://doi.org/10.1139/f02-123.

Winker, H., Kerwath, S.E., Attwood, C.G., 2013. Comparison of two approaches to standardize catch-per-unit-effort for targeting behaviour in a multispecies hand-line fishery. Fish. Res 139, 118–131. https://doi.org/10.1016/j.fishres.2012.10.014.

Winker, H., Carvalho, F., Kapur, M., 2018. JABBA: just another Bayesian biomass assessment. Fish. Res 204, 275–288. https://doi.org/10.1016/j.fishres.2018.03.010.

Winker, H., Parker, D., Da Silva, C., Kerwath, S., 2020b. First comprehensive assessment of soupfin shark Galeorhinus galeus in South Africa.

Winker, H., Carvalho, F., Thorson, J.T., Kell, L.T., Parker, D., Kapur, M., Sharma, R., Booth, A.J., Kerwath, S.E., 2020a. JABBA-Select: incorporating life history and fisheries' selectivity into surplus production models. Fish. Res. 222, 105355 https://doi.org/10.1016/j.fishres.2019.105355.

Zhou, S., Punt, A.E., Deng, R., Dichmont, C.M., Ye, Y., Bishop, J., 2009. Modified hierarchical Bayesian biomass dynamics models for assessment of short-lived invertebrates: a comparison for tropical tiger prawns. Mar. Freshw. Res 60, 1298. https://doi.org/10.1071/MF09022.