Pragmatic approaches to modeling recruitment in fisheries stock assessment: A perspective

Elizabeth N. Brooks [1]

National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northeast Fisheries Science Center, Population Dynamics Branch, 166 Water Street, Woods Hole, MA 02536, USA

ARTICLE INFO

ABSTRACT

Our tasks in performing stock assessment are to estimate what is in the population now, what is likely to be there in the near-future, and how we should best extract yield without imperiling future reproduction or yield. There are many considerations for modeling recruitment in a stock assessment, and I walk through the typical sequence of decisions an analyst must make when setting up their stock assessment. I begin with the deceptively simple decision of the first modeled age, progress through the treatment of fecundity, and into the important decision of parameterizing the stock recruit relationship (SRR). Subsequent assessment steps such as projections and estimation of reference points build on these earlier decisions. The consequences of many of these decisions are amplified or obviated, depending on whether or not a null SRR model (mean or median with deviations) is implemented. Examples of expanding or attenuating consequences are provided within each decision. The potential for misspecification related to these decisions is noted, and advice to reduce that misspecification as well as suggested diagnostics are offered. As a high level summary, I recommend fitting the null SRR, using weight as a surrogate for fecundity and conducting sensitivity analysis if there are data suggesting significant departure from isometry. If estimating a non-null SRR, move away from the steepness parameterization (return to the original α,β parameterization), and do not begin at age 0. Lastly, exercise caution when incorporating covariates –they may be hitchhikers posing as drivers.

1. Introduction

There are many different angles from which to study recruitment, where the type of model applied is geared toward the question being asked. Theoretical studies may try to elucidate behavior of recruitment time series under different structural functions, e.g. studying stable versus cyclic dynamics and the conditions that lead to one or the other outcome. Improving our understanding of basic biological processes such as stock productivity, survival, predation, and growth, and temporal changes therein, is fundamental to thinking about the importance of these processes and how they could influence or be incorporated in population dynamics. A broader ecological context may examine exogenous drivers such as climate that are thought to be influencing the past trends in biological processes or instigating new patterns. Each of these perspectives lays important groundwork for how one should think about recruitment in a stock assessment model, and can provide some insight (or hypotheses) on why we end up with the estimated number of recruits in a given year; and although the perspectives and approaches may differ, they don't always lead to different numeric estimates.

The primary focus of this work is how to model recruitment for stock assessment and management advice. Pragmatically speaking, our tasks in performing stock assessment are to estimate what is in the population now, what is likely to be there in the near-future, and how we should best extract yield without imperiling future reproduction or yield. In the context of modeling recruitment, we need to estimate recruitment events that reproduce observed catch and match survey trends and age composition. We also need to forecast recruitment from the assessment model's terminal year estimates in order to advise on the future stock trajectory and catch advice. Ideally, future catch advice and spawning biomass trajectories aren't too dependent on projected recruitment. However, lags between data availability, assessment meetings, and quota setting often mean that terminal year recruitment estimates will contribute to those quantities to some extent, depending on selectivity and maturity patterns. Also, in the context of longer term projections for rebuilding, those forecasted recruits (also called "paper fish" because they are a model product and not based on observations) may contribute

to a larger extent. Reference points for determining stock status and the sustainability of future catch (and rebuilding) depend on recruitment to determine the absolute scale–this is quite a heavy demand for such a challenging quantity to model.

In what follows, I consider several key decisions that an analyst must make when conducting a stock assessment, highlighting potential mis-specifications and a rationale for proposed pragmatic approaches. This is primarily germane to age structured assessment models, because recruits and the form of their dependence on spawners are explicitly defined in the model equations by a stock recruit relationship (SRR). I note that even models that estimate a time series mean (or median) recruitment and annual deviations from this mean can be considered a 'null SRR model' (and I retain this terminology below). Simpler assessment models have implicit assumptions that are not as obvious but are just as easy (if not easier) to violate (e.g., no lag between spawning and recruitment, all ages have the same weight and selectivity and contribute equally to surplus production), but they are not the focus of this work. I conclude by discussing these decisions in a broader context for other assessment approaches and data environments, and offer counterpoints as well as recommendations for future research and the next generation assessment models.

2. Methods- decision pathway

There are many considerations for modeling recruitment, and I walk through the typical sequence of decisions an analyst must make when setting up their stock assessment. I begin with the deceptively simple decision of the first modeled age, progress through the treatment of fecundity, and into the important decision of how to model recruitment. Subsequent assessment steps such as estimation of reference points and forecasting follow from these earlier decisions. The consequences of many of these decisions are amplified or obviated, depending on whether or not a null SRR model is implemented. Examples of expanding or attenuating consequences are provided within each decision. A poll sent to stock assessment scientists to characterize current practice is summarized and referenced in the appropriate sections below (see table and figures in Supplementary Material).

2.1. Assessment related decisions

2.1.1. Age at recruitment

One of the first decisions in setting up an age based stock assessment is often one the analyst doesn't even have to make, because the software package makes it for them: the age at recruitment. Many assessment software packages have implicitly coded the first age at either 0 (Stock Synthesis versions earlier than 3.30, Methot and Wetzel, 2013) or 1 (ASAP – Legault and Restrepo, 1999; WHAM – Stock and Miller, 2021; BAM – Williams and Shertzer, 2015; AMAK – Ianelli and Fournier, 1998), although some allow this specification (SAM – Nielsen and Berg, 2014; a4a – Jardim et al., 2015; FLR – Kell et al., 2007; Stock Synthesis versions 3.30 and later; Methot *pers. Comm.*) (Fig. S1). From the early days of modeling commercial fish stocks, this decision was driven by the age that fish 'recruited' to the fishery, because that was when data were first available. Beverton and Holt (1957) described the main factor between stocks with early and late recruitment age as being dependent on whether nursery grounds overlapped with areas of exploitation. The terminology 'partial recruitment' still persists for describing patterns of selectivity, and is likely a holdover of this earlier context. For the purpose of this work, I focus not on 'fishery recruitment' but on 'biological recruitment', and define recruits as the number of fish that survive the density dependent window that begins at spawning. Density dependence is assumed to be nonexistent or negligible beyond this window, although in reality this is likely to be a gradual transition (Lorenzen and Camp, 2019).

Specifying the age of recruitment in an assessment model should be motivated by both data availability and biological processes, and there

are consequences of that decision for the estimability of a stock recruit relationship. As noted above, density-dependence is assumed negligible beyond the age of recruitment, so at a minimum one should not specify the age of recruitment where it is suspected that survival is still density-dependent. The length of the density-dependent window rarely aligns with model time steps (often calendar year), but 'rounding up' from the end of that window to the end of the year by density-independent mortality does not impact the estimated non-null SRR resiliency parameter (as long as the fish remains immature in that 'round up' period; Brooks and Powers, 2007). However, 'rounding down' by truncating some portion of the density dependent window would yield biased SRR parameters (see Appendix A).

Two additional considerations for recruitment age are the earliest age for which data exist, and the maturity ogive. One should aim to use the earliest age with data, provided it is not already a mature age. The stipulation for the age to be immature is due to the fact that you cannot estimate a stock recruitment relationship if any of the 'recruits' are mature because you would then have the same fish on both sides of the SR equation. If data are only available for mature fish, then the null SRR should be used.

It is also important to avoid (or minimize to the extent possible) specifying that recruitment occurs at an age at which fish are already appearing in the fishery catch (either as landings or discards). Doing so essentially writes those fish off because it reduces the perceived productivity of the stock, and therefore will reduce the estimated yield that can be taken (Brooks and Powers, 2007). This is true for non-null SRR because the inclusion of fishing mortality in the recruitment window reduces the estimate of stock resilience (from which Fmsy is estimated) as well as the estimate of unexploited equilibrium recruitment (Brooks and Powers, 2007). For the null SRR, the estimate of mean recruitment would also be reduced. This recommendation is made with stocks that live more than a year or two in mind; stocks with very short life-spans may be available to surveys and fisheries very close to 'age zero'. For these stocks, if the model time step is monthly (or seasonally) rather than annual, it may be possible to specify a recruitment age in months that reduces the amount of fish that get written off by being caught in the recruitment window (but see 2.1.3 for discussion of this with respect to a null SRR). A seasonal model might also be appropriate for longer-lived stocks if fish are available to the fishery before age 1. To deal with pre-recruits in the fishery catch, one could *try* to estimate fishing mortality from catch that occurs within the recruitment window, but this is complicated by the assumed units of fecundity (typically weight) versus units of fish caught (numbers). If your fecundity units are numbers of eggs, then it may be possible to estimate fishing mortality (F) directly; if your fecundity unit is weight, then a relative fecundity scalar is implicit in the F, and it is not clear how you would scale bycatch estimates.

Now, if it is best to have a shorter window for recruitment (to avoid overlap with maturity and fishery selectivity), is it possible to have the window be too short? Modeling recruitment to begin at age zero seems problematic for several reasons. First, it is rare to have precise observations at age zero. Second, the x-axis in a stock recruitment plot is typically spawning biomass, which is assumed to be a proxy for eggs, which are age 0. Thus, when recruitment occurs at age zero, it would seem that you are specifying the same quantity for both the x and y axis, albeit in different units. And arguing from first principles, stock recruitment functions are derived from a differential equation describing the change in number of eggs (R(0)) from time t = 0 to the end of the recruitment window t_r:

$$\frac{dR(t)}{dt} = -[M + F + AR(t) + BR(0)]R(t) \quad 0 \leq t \leq t_r \tag{1}$$

In (1), M and F are the instantaneous natural and fishing mortality rates, respectively, and A and B define density-dependent mortality rates. Clearly, if the recruitment window begins and ends at t = 0, then the definite integral of *dR/dt* from 0 to 0 is R(0), i.e there is no change when there is no elapsed time. These points are technical, and one might

argue (with some hand waving) that the model actually estimates numbers at a later age (where data exist) and the recruitment parameters are just rescaled by the density-independent mortality between age 0 and that later age. That argument seems tenuous (see Appendix A), likely to mischaracterize uncertainty, and necessary only because of a constraint in the model code. Even if it is possible to specify smaller time intervals within the year and therefore the fraction of year when age 0 fish appear, there must explicitly be an 'N' corresponding to that first age slot in the model of numbers at age at the beginning of the year (age 0), even though that age is then 'born again' some Δt of the way through the year. It is unconvincing that sub-annual time steps solves the problem with specifying age 0 when you have a non-null SRR. Moreover, if the month of spawning within the year is not January 1, there needs to be careful coordination of age assignment conventions so that the offsets align between age readers and assessment modelers. Because of the lack of a relationship between recruits and spawners with the null SRR, the above misspecifications with age 0 can be avoided, but even then it would seem unwise to begin your model at age 0 if you have no or limited data for that age (see Fig. S2 for age modeled by SRR type in current assessments).

Pragmatic recommendations for age at recruitment can be summarized as follows: avoid modeling recruitment at age 0 and minimize to the extent possible specifying that recruitment occurs at an age at which fish are already appearing in the fishery catch; specify an age ≥ 1 that is immature (or adjust model time step to accommodate very short lived stocks); coordinate age assignment conventions between age readers and modelers so that 'biological age' and 'model age' align. If there are data at several fully immature ages, evaluations of model precision, stability (Mohn's rho), cross-validation and model skill (Kell et al., 2021), and forecast accuracy (retrospective forecasting, Brooks and Legault, 2016) may help identify the better age to specify for recruitment. Considerations for future model development are to generalize the first model age and allow the user to specify age at recruitment.

2.1.2. Units of fecundity

The weight of a fish is the most common surrogate for egg production (Fig. S3), and arguments for its use are found in Beverton-Holt (1957). But convenience isn't always the best rationale for convention. A number of concerns have been raised about the use of weight for fecundity, with hyperallometry being one of the most frequent (Marshall et al., 2021; Barneche et al., 2018;). With hyperallometry, fecundity is assumed to be more than proportional with increases in weight, such that *fecundity = weight^c*, where $c > 1$. Concerns have been raised that using weight in place of fecundity will bias stock productivity, resilience, and reference points (Marshall et al., 2021). But, fecundity data are not widely available (Tomkiewicz et al., 2003), can vary both spatially and temporally (Marteinsdottir and Begg, 2002), and are harder and more costly to collect. The diversity of reproductive strategies (McBride et al., 2015) makes it clear that presence of eggs does not guarantee spawning will occur in a given year. Empirical studies (Trippel, 1998), modeling work (Murawski et al., 2001), and synoptic reviews (Fitzhugh et al., 2012; Hixon et al., 2014) suggest that spawning success (irrespective of weight or fecundity unit) and egg viability *may* improve with each successive spawning season (but see Morgan et al., 2007). This is not a one-dimensional problem by any means, and striving for greater realism presents challenges with additional nuances for which data are sparse.

There is a lot of work being done to advance our understanding about reproductive ecology, and for some stocks there is an abundance of data on eggs (e.g. egg count surveys for mackerel and horse mackerel in the Northeast Atlantic, red snapper in the Southwest Atlantic). However, data to quantify the above mentioned reproductive aspects are quite limited (if they exist) for most species, especially if estimates spanning the time series of the assessment model (and possible changes through time, Fig. S4) are needed. If only limited temporal data about one aspect

are available, it is hard to believe that using it in place of weight would create fewer problems; sensitivity analyses could identify which of the reproductive processes has the most important effect (Miller et al., 2018; Brooks, 2013).

The case for staying the course and continuing to use weight as a proxy for eggs would seem to be the default, because of greater availability of such data. However, even if fecundity is isometric and weight is therefore an appropriate proxy, the variability in observed weight at age can stray from the true population weights for reasons that are beyond the analyst's control. For example, stock weights that are used to calculate SSB in the model can be derived from catch weights (typically a mid-year timing is assumed for catch weights, and a model based approach is used to regress those weights to the timing within the year when spawning is assumed to occur). If the timing of catch changes such that more occurs earlier or later in the year over the time series, that can impact the "mid-year" assumption. Changes in processing (gutting) and market preference over time can also lead to discrepancies. Alternatively, stock weights can be calculated from sampling on scientific surveys. The design of fishery-independent surveys is to sample at roughly the same time of the year, however logistics related to weather and mechanical failure can cause shifts in survey timing. And while the survey attempts to adhere to a fairly consistent schedule, fish are under no such obligations, and important activities like spawning may occur due to temperature or other natural cues that may vary substantially from year to year. This could impact measurements of weight at age if your sample in one year includes primarily developing individuals and the next year includes more spent individuals (Wuenschel et al., 2018). In such a case, observed weight at age would appear to have decreased (e.g., Trippel and Neil, 2004, estimate of 25% loss in mass pre- vs post-spawning haddock).

Modelling with a non-null SRR is impacted directly by the choice of fecundity unit because of the functional reliance on SSB, but even the null SRR is impacted by the fecundity unit because SPR proxies use 'fecundity at age' for the calculation. However, regardless of the unit, the annual estimates of recruitment are identical for the null SRR (including in the projections). For the non-null SRR, recruitment estimates are identical in the assessment years but differ in the projections because those estimates are predicted from the mean fitted curve given the estimated SSB in those years (which differ due to the units) (Fig. 1). Comparing predicted recruitment between a null and non-null SRR, for a given fecundity unit, results in nearly identical recruitment for almost the entire time series, with differences only noticeable at the very end where the non-null SRR estimates show some shrinkage to the mean—the meaningful differences are in the projections, where recruitment estimates can be quite different.

An advantage of using eggs rather than weight for non-null SRR is that one can interpret the parameters of the SRR in terms of instantaneous mortality rates, whereas when weight is the fecundity unit, those parameters have an implicit scaling by average fecundity that obscures the mortality rate. The ability to interpret mortality directly could be of use if one had catch estimates of pre-recruited fish and wanted to attempt to estimate associated fishing mortality (although realistically, one would probably need a strong hypothesis about the rate of density-independent mortality of pre-recruits in order to estimate fishing mortality).

Pragmatic recommendations for units of fecundity are as follows: continue using weight at age because many stocks have nothing else. Try to eliminate noise in weight at age due to factors other than growth (shifts in landings pattern over years, or timing of spawning relative to sampling, e.g.). Consider modeling stock weight inside the model so that uncertainty in the process carries into stock weights in the projections (Nielsen et al.;, *this issue*). If limited direct fecundity data are available, consider a sensitivity analysis using that in place of weight at age to gauge the potential magnitude of change in assessment results–these results could form the basis of requests for continued funding to collect more of that data (and standardize its interpretation)
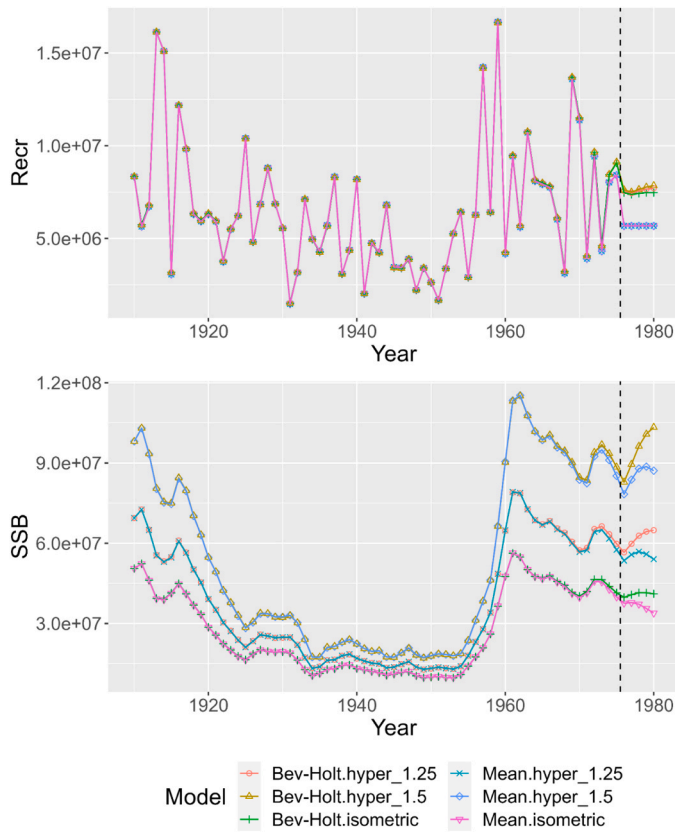
Fig. 1. Annual estimates of recruitment (top) and spawning stock biomass (SSB) (bottom) for two forms of stock recruit relationship (Mean or Beverton-Holt), and three possibilities for the exponent scaling weight to fecundity, *fecundity = weight^c*: isometric (c=1), or hyperallometric (c=1.25 or c=1.5). The vertical dashed line separates assessment years from projection years.

and to explore ways to extend the time series further back in time if possible. If direct fecundity data exist, by all means use it, but as noted when using weight, be attentive to any misleading patterns due to sampling rather than biology, and continue monitoring for potential temporal change.

2.1.3. Functional form for SRR

The two most popular functional shapes of non-null SRR are distinguished by the pattern of recruitment as spawning stock increases – either recruitment asymptotes (Beverton and Holt, 1957), or recruitment declines after peaking at some intermediate spawning stock abundance (Ricker, 1975). The primary alternatives in current practice are the null SRR with mean deviations (correlated or uncorrelated), and a hockey stick where a SSB hinge point is imposed such that predicted recruitment declines to the origin when SSB is less than that hinge (Fig. S5; but see Albertsen and Trijoulet, 2020, for additional parametric and non-parametric forms including the 'compensatory mortality property' model of Cadigan, 2013). The hockey stick SRR (Barrowman and Myers, 2000) was used in 3 out of 88 responses, and a recent study found it had the poorest identifiability in model selection testing (Trijoulet et al., 2022), and it is not described further here. Biological mechanisms are invoked as justification for one or the other shape, with the likelihood of cannibalism offered as justification for selecting the Ricker overcompensatory SRR. Biological mechanisms are also the principal argument against the null SRR, because to some it violates the primacy of biology in describing population dynamics. But our inability to detect a signal from noisy data with low contrast and no replicates should not be taken as an affront to our biological sensibility.

If an analyst is contemplating fitting a SRR other than the null model,

another seemingly innocuous decision is how to parameterize the SRR – but often, this is also pre-determined by the software package being used (Fig. S6). Currently, the most common SRR parameterization is in terms of steepness (h) and unexploited recruitment (R0), the other alternative being the original α and β parameters:

$$R = \frac{4hR_0 S}{\varphi_{F=0}R_0(1-h) + (5h-1)S}(\text{Beverton} - \text{Holt}) \quad (1a)$$

$$R = \frac{\alpha S}{1 + \beta S}(\text{Beverton} - \text{Holt}) \quad (1b)$$

$$R = \frac{1}{\varphi_{F=0}}S(5h)^{5\left/4\left[1-\frac{S}{R_0\varphi_{F=0}}\right]\right.}(\text{Ricker}) \quad (1c)$$

$$R = \alpha S e^{-\beta S}(\text{Ricker}) \quad (1d)$$

The derivation of steepness (Mace and Doonan, 1988) was very clever in that it resulted in a unit-less value reflecting stock productivity as well as vulnerability to exploitation, and even more useful was that it could be compared across stocks for inference about being more or less resilient to exploitation. Further analytical work demonstrated that steepness could be transformed into maximum lifetime reproductive rate ($\widehat{\alpha}$, also unitless and comparable across stocks; Myers et al., 1999). Comparability and intuitive interpretability are strong selling points and likely fueled enthusiasm for coding SRR in terms of steepness rather than α and β.

When model choices are hardwired in the code, thereby removing the choice from the user, a potential consequence is that the user is unaware of embedded assumptions (unless one carefully reads the code line by line – asssuming it is available and well commented). Mace and Doonan (1988) explicitly noted the assumption of a population in equilibrium, and to translate from α and β one also needed to define that steady state in terms of unexploited recruitment (R0), which by definition is achieved by unexploited spawners (S0) divided by unexploited spawners per recruit ($\varphi_{F=0}$) which is calculated as

$$\varphi_{F=0} = \sum_{a=r}^{A-1}\mu_a f_a \prod_{j=r}^{a-1}\exp(-M_j) \quad + \mu_A f_A \prod_{j=r}^{A-1}\exp(-M_j)\frac{1}{1-\exp(-M_A)} \quad (2)$$

where A is the plus group, M is natural mortality, μ is maturity, and *f* is fecundity. The replacement line from the origin to the unexploited equilibrium (S0, R0) therefore has slope $1/\varphi_{F=0}$. This is a place where misspecification can creep in, as highlighted in Miller and Brooks (2021). A number of data rich stocks are able to estimate annual maturity (or stanzas of maturity ogives for blocks of years) as well as annual growth equations or empirical weights at age (WAA), and in some cases even annual natural mortality (see Fig. S4). The existence of this rich data challenges the lack of a 'y' subscript (for year) in the equation for $\varphi_{F=0}$ above. What are the implications? First, if the assessment software uses the steepness parameterization, then assessment model users should know what is specified for the unexploited replacement line in the software they are using (is it calculated from values in the first year of observations? The most recent year? A mean of all observed values at age?). Second, the user should be aware that if any biological parameters are changing by year (maturity, fecundity, natural mortality), then the assumption of equilibrium state used to derive steepness has been violated (Fig. S7 characterizes frequency of time-varying biological parameters by SRR).

From the perspective of model estimation, the consequence of this violation is that you will get a different estimate of steepness and R0,S0 for each combination of age-specific biological parameters (because each combination produces a different replacement line), yet the uncertainty (and possible arbitrariness) associated with default values used in the model is probably not recognized by many practitioners. For

example, Miller and Brooks (2021) show an illustration where the point estimate of steepness could vary from 0.47 to 0.69 for the Beverton-Holt (or 0.44–0.77 for the Ricker), while spawning biomass and yield at MSY each varied by a factor of about 2.5, depending on the year from which biological parameter estimates were taken. Rather than leaving this range of uncertainty buried by a default model specification, Miller and Brooks (2021) recommend to estimate the SRR in terms of α and β, which does not rely on the replacement line (Eqs. 1b and 1d versus 1a and 1c), and then to transparently derive reference points from the fitted SRR so that the assumed biological values are explicitly stated. Subsequently, the α and β parameters can be converted into associated parameters of steepness for comparison between stocks.

There are also consequences of variable biological parameters for two common diagnostics: R0 profiling and dynamic B0 (Lee et al., 2014; Berger, 2019). As Miller and Brooks (2021) note, there are inconsistencies with the dynamic B0 approach, and for both diagnostics one would expect results to vary depending on which replacement line was used (because each yields a different R0,B0).

One of the appealing features of the steepness parameterization was that steepness is unitless and therefore directly comparable across stocks. This has led to meta-analyses and implementations of steepness priors to aid estimation in cases where the data are not informative about the SRR. However, comparability doesn't necessarily imply interchangeability; the misspecification just discussed about time varying biological parameters can profoundly bias the estimates from which priors are derived (Miller and Brooks, 2021, specifically Table 1 and Fig. 5). Steepness is a function of density-independent recruit survival (α, scaled to account for fecundity if unit is weight) and $\varphi_{F=0}$:

$$h = \frac{\alpha\varphi_{F=0}}{4 + \alpha\varphi_{F=0}} \text{(Beverton} - \text{Holt)} \tag{3a}$$

$$h = \frac{1}{5}(\alpha\varphi_{F=0})^{4/5} \text{(Ricker)} \tag{3b}$$

Therefore, assigning two stocks with different $\varphi_{F=0}$ to have the same steepness implicitly changes the α parameter in order to balance the stock-specific $\varphi_{F=0}$, which confounds variability in the pre- and post-recruit processes (α and $\varphi_{F=0}$, respectively) (Miller and Brooks, 2021). Similarly, when simulation studies fix both M and steepness as factors that vary independently, the same confounding occurs (but see Cortes and Brooks, 2018, for a simulation where α and $\varphi_{F=0}$ were the factors varied so that hypotheses about the population process was explicit). Even if biological parameters are constant, He and Field (2019) found that accurate estimates of steepness were dependent on the degree of contrast in stock abundance (generated through exploitation patterns) and recruitment variability, and that even informative priors could produce biased estimates depending on the appropriateness of that prior.

Not all stocks have sufficient data to estimate annual biological parameters (Fig. S4, S7) – is it safe for these stocks to use the steepness parameterization? In these instances, one assumes a time invariant vector for biological parameters at age, so in principle the stationarity condition has not been violated. Whether that is true in reality, or just appears to be so due to low sampling, is a question analysts should confront. The same question could be asked for stocks with annual observations that appear to fluctuate without trend – is it so egregious to simply estimate a timeseries average for biological parameters and

assume that reflects equilibrium conditions? Perhaps not, but the uncertainty associated with the annual departures from that time series average are not conveyed in the assessment results or in the management advice unless they are directly modeled in the assessment. Again, in these situations, the α,β parameterization could be used instead, and then replacement line assumptions for deriving reference points can be addressed separately and transparently. Unfortunately (or fortunately), α and β are not amenable to priors, so if your data are not informative about these parameters then there isn't a meta-analysis to fall back on, and you should probably consider the null SRR as your starting point.

If you have observed high contrast in the data for your stock and have success fitting a SRR, yet another fly in the ointment is model selection. Multiple studies have shown that the Ricker has a greater probability of being selected by information criteria as the correct model even when it isn't (de Valpine and Hastings, 2002; Zhou, 2007; Brooks et al., 2019). Several explanations have been offered for this (Subbey et al., 2014 and references therein), but no remedy.

It is sometimes proposed that part of the difficulty with fitting SRR is due to not accounting for influential mechanisms, and thus the incorporation of indices that purportedly explain variation in the SRR is a windmill that is often tilted at. But do these indices reflect environmental drivers or merely hitchhikers (sometimes referred to as lurking variables)? Identified correlations that subsequently break down have been discussed extensively (Myers, 1998; Walters and Collie, 1988; etc.). A comprehensive review in Haltuch et al. (2019) identified stocks with a clear bottleneck in the recruitment window as the most likely cases to have success incorporating environmental indices, and Haltuch and Punt (2011) found high Type I error for accepting an environmental index when infact it spuriously coincided with stock declines from fishing. None of the survey respondents reported including an environmental driver in the SRR (Fig. S8), although two of the stocks are exploring sensitivity assessment model configurations that include an index to inform recruitment deviations (SEDAR, 2018; Sculley et al., 2018), and another stock uses it as auxiliary information to inform expectations about recent year classes (Ianelli et al., 2022). This seems consistent with Skern-Mauritzen et al. (2016), who examined 1250 stock assessments and found only 24 that incorporated an environmental index in any biological process.

The increasing adoption of state-space approaches in stock assessment (SAM, Nielsen and Berg, 2014; WHAM, Stock and Miller, 2021) provides the means for more appropriate statistical treatment of environmental indices and their mechanistic effect on recruitment (e.g., Miller et al., 2016). Even with these more advanced tools, identifying robust models is challenging. I illustrate this for an expanded suite of models originally considered in Miller et al. (2016) for incorporating a cold pool index (CPI) for southern New England yellowtail flounder (see vignette 2 at https://timjmiller.github.io/wham). The best fitting model (by AIC) from the vignette set was a Beverton-Holt with the CPI acting as a "controlling" influence (see Iles and Beverton, 1998; Xu et al., 2018; Maunder and Thorson, 2019) with AR1 process error, and with uncorrelated (iid) process error in the annual transitions between numbers at age (NAA). Performing a retrospective analysis by removing 10 years of data showed that dropping even 1 year led to selection of a different model as the best (still a Beverton-Holt, but a limiting instead of controlling mechanism for the CPI effect, which was modeled as a random walk instead of AR1; Fig. 2). When the vignette set of models was

Table 1

Proportion of recruits estimated in the terminal year of the assessment (T in **bold**), proportion of paper-fish (PF in *italics*), and combined proportion of terminal year estimate and paper fish (T + PF in ***bold italics***) that contribute to projected catch in biomass.

Projection Year	Age1	Age 2	Age 3	Age4	Age5	Age6	Age7	Age8	Age9	Prop. T estimate	Prop. PF	Prop. T+PF
T + 1	*0.01*	**0.13**	0.06	0.15	0.04	0.12	0.05	0.03	0.41	**0.13**	*0.01*	***0.14***
T + 2	*0.01*	*0.05*	**0.31**	0.08	0.13	0.03	0.09	0.04	0.27	**0.31**	*0.06*	***0.37***
T + 3	*0.01*	*0.04*	*0.12*	**0.38**	0.06	0.09	0.02	0.06	0.21	**0.38**	*0.17*	***0.55***
T + 4	*0.01*	*0.04*	*0.10*	*0.16*	**0.34**	0.05	0.07	0.02	0.21	**0.34**	*0.32*	***0.66***
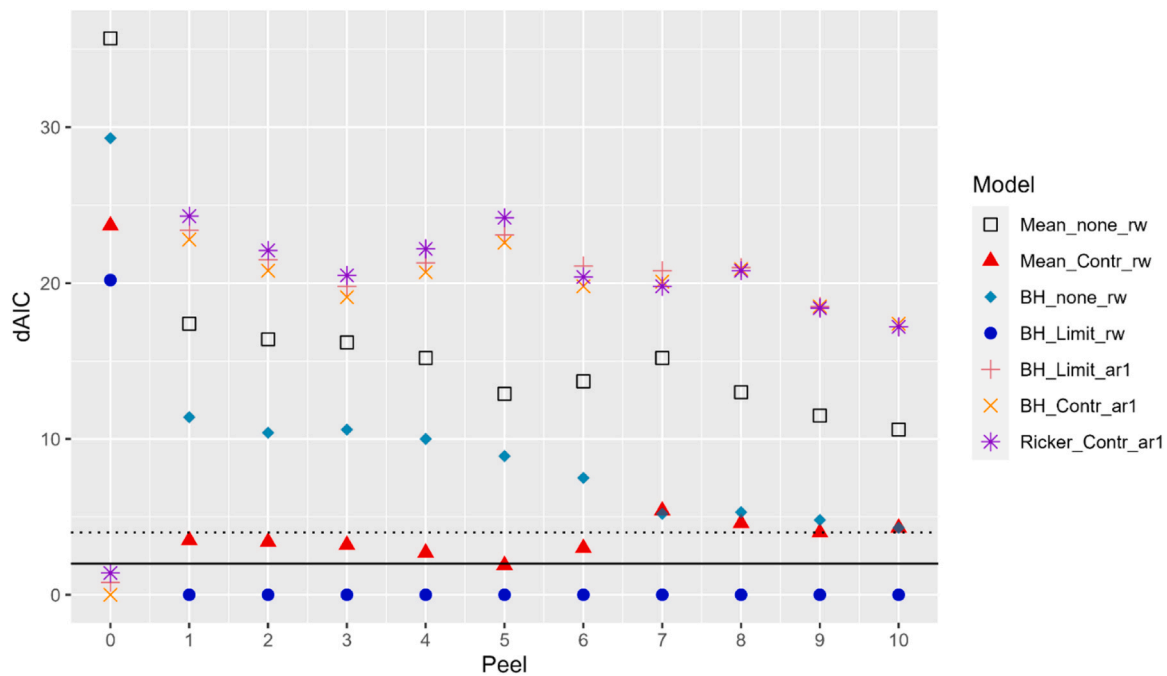
Fig. 2. Retrospective model selection based on AIC for 7 models. The difference in AIC (dAIC) is calculated from the model with the minimum AIC. Peel 0 is the full data set, and peels 1–10 successively remove one year of data. The models fit the null stock recruit relationship ("Mean"), the Beverton-Holt ("BH"), or the Ricker, where the cold pool index (CPI) either had no effect on recruitment ("none"), was controlling ("Contr"), or was limiting ("Limit"). Process error for the CPI was either a random walk ('rw') or ar1; process error in numbers at age transitions was iid. [See vignette 2 at https://timjmiller.github.io/wham].

expanded to include an AR1 process in the NAA transitions, then the best fitting model was the null SRR where the CPI acted as a controlling mechanism with a random walk for process error, and NAA transitions had an AR1 correlation in process error (this model selection held as the best for all 10 peels). Notably, the time series estimates and CVs in the assessment time period are remarkably similar regardless of the form of the SRR and the correlation among NAA transitions (Fig. 3). Differences in the projections among these models are discussed below in the forecasting decisions.

The challenges of incorporating an environmental driver mount when there are so many ways to model it—how many different models should be considered? Can any be ruled out a priori? If we try all of them, how do we avoid *post-hoc* justification for the model selected (or a significant result by chance alone)? And what is the best assumption for projecting the driver into the future? This harkens back to a memorable quote in Walters and Collie (1988):

> …*"the model worked well until next year." The amazing thing is not that the correlations breakdown with such disturbing regularity, but that apparently sensible scientists keep searching for them."*

It seems clear that climate is changing, and living beings are changing in response to it. Skepticism of spurious results does not make one a climate denier but rather a careful and dispassionate analyst. Work to identify drivers will likely continue, because as scientists we are driven to ask questions and try to improve our understanding and our ability to anticipate future conditions. But we should be mindful of the potential for hitchhikers in our quest to identify drivers. Most analysts perform retrospective analysis and often do so for several models being considered – with that work already done, it takes minimal effort to also look at model selection across those same models in each of the peels to identify stability of the model selection (see Miller et al., 2016, e.g.). This recommendation applies to assessment models without environmental covariates as well.

With all of the decisions to be made for modeling recruitment in stock assessment, there is ample opportunity for misspecification. And given the complexity of processes that occur within the recruitment

window, detecting a signal other than the null SRR can be difficult (Brooks et al., 2019). Diagnostics can help us understand whether our recruitment estimates are robust. A primary tool is retrospective analysis (Mohn, 1999; Legault, 2009; Deroba, 2014; Hurtado-Ferro, 2015). The retrospective pattern of recruitment is often very messy, and this is not surprising given terminal year model estimates are operating on the least amount of data (as demonstrated by large CVs associated with these estimates). Nevertheless, terminal year estimates can be quite influential in projections, depending on the length of the projection and the age at which those recruits become mature (enter SSB) and reach the fishery. Several diagnostics that can help frame expectations for the terminal year recruitment estimates, based on recent model performance, are shown in Fig. 4. For example, given a retrospective analysis has been performed (Fig. 4a), another way to look at the same data is to calculate how the terminal year estimate of all year classes in the retrospective peels were subsequently re-scaled as additional years of data were added to the model (Fig. 4b). This allows one to determine if there is a consistent pattern of overestimation with only 1 year of data, or whether there is a pattern of over/under estimation depending on the magnitude of the initial estimate. With more complex models, it may be possible to achieve stability in some model estimates, but we should look to see if something else is absorbing the changes as new data are added to the model (Fig. 4c, showing retrospective pattern in recruitment deviations). Lastly, hindcasting or retrospective forecasting analyses (Brooks and Legault, 2016) can help identify robustness of the chosen recruitment model (Fig. 4d, showing updated assessment estimates of recruitment and values projected from the previous assessment).

Pragmatic recommendations for the functional form of SRR model: model recruitment with a null SRR as a default because it avoids a lot of potential for misspecification and many stocks lack contrast in abundance anyway. If you have a lot of contrast in stock abundance and can estimate a non-null SRR, then move from steepness to α,β parameterization when you have temporal variability in biological parameters, and be aware of model selection bias towards the Ricker. Steepness priors are prone to abuse and misspecification and should also be avoided. Adopt state space modeling approaches, where process error
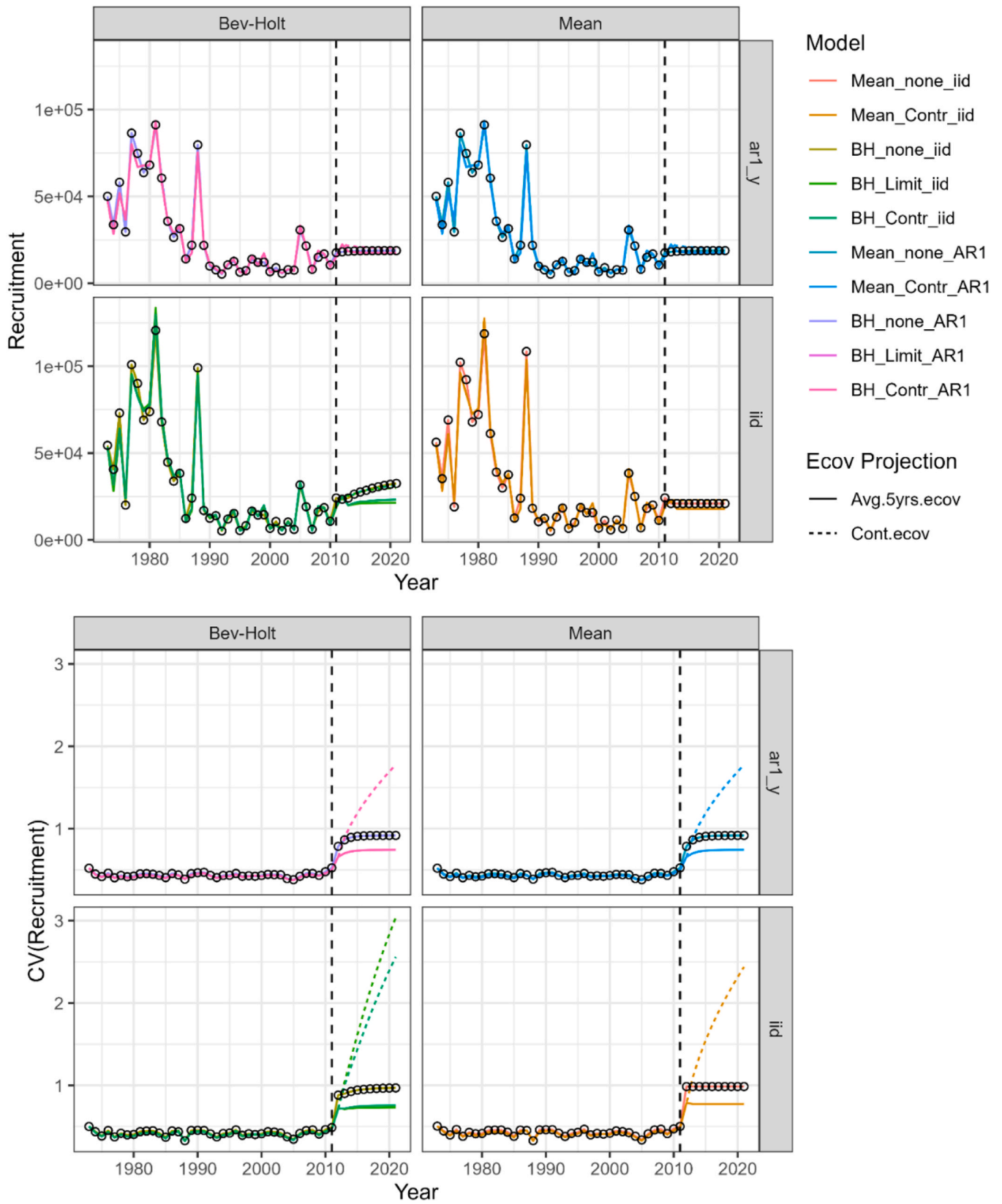
Fig. 3. Time series of recruitment estimates (top) and CVs (bottom) for southern New England yellowtail flounder. The models fit the null stock recruit relationship ("Mean") or the Beverton-Holt ("BH"), and the cold pool index (CPI) either had no effect on recruitment ("none", open circles), was controlling ("Contr"), or was limiting ("Limit"). Process error for the CPI was a random walk. Row facets indicate the process error in numbers at age transitions (iid or AR1 across years, "ar1_y"). Projections are to the right of the vertical dashed line, where the CPI was either held constant at a recent 5 year average ("Avg .5 yrs.ecov", solid lines) or the CPI process error continued ("Cont.ecov", dashed lines). [See vignette 2 at https://timjmiller.github.io/wham].
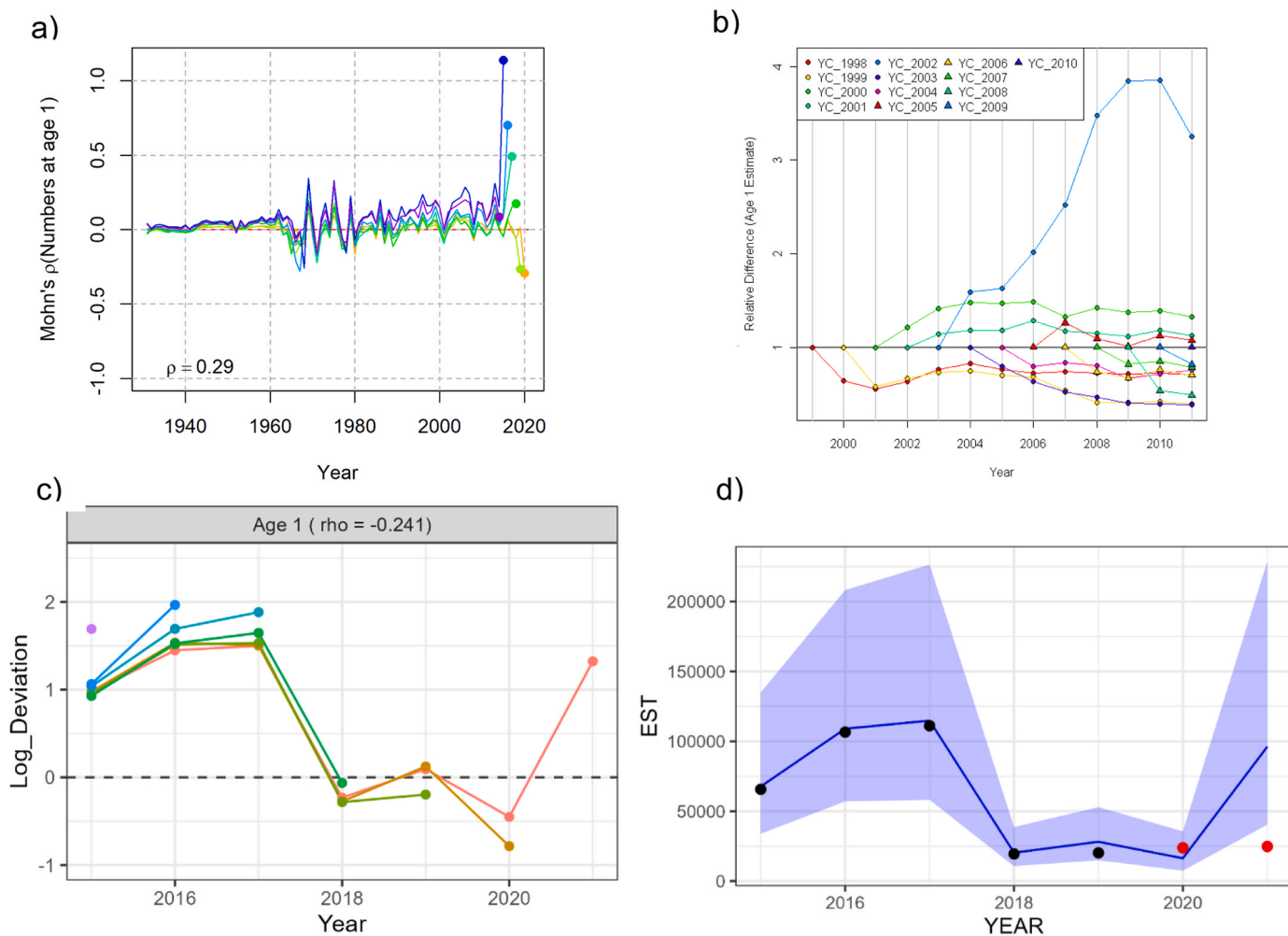
Fig. 4. Diagnostics for evaluating recruitment model robustness: (a) Retrospective analysis; (b) change in scale of recruitment from the first estimate (with only 1 year of observations on a year class) to the most recent estimate; (c) retrospective analysis of deviations in recruitment; (d) evaluating forecast accuracy from an earlier assessment (black circles are point estimates, red circles are projected recruitment) with the most recent assessment (solid blue line is median, shaded polygon is 95% CI).

and correlation in random effects are estimated and carried forward into projections. Be wary of covariate hitchhikers posing as environmental drivers. Examine robustness of model selection across retrospective peels, as well as diagnostics that relate to stability of terminal year estimates. For future model development, reduce hard-wired parameterizations about the SRR (e.g., the replacement line) so that the user is aware of the choice and can explore its consequences. Finally, allow the user to choose among SRR parameterizations.

2.2. Forecasting decisions

Projecting the stock assessment model for management advice requires many assumptions, because data are rarely available to be fit in those years. To better characterize risk, and to maintain a consistent treatment of uncertainty, decisions made while fitting the stock assessment model to data should carry forward into projections. Specifically, the unit of fecundity impacts projected SSB, which influences perception of stock biomass relative to management targets (Minte-Vera et al., 2019); it also influences projected recruitment if a non-null SRR is used (Fig. 1). Assumptions about fecundity (typically weight at age) and maturity may apply a recent average by age, or results from a fitted model that is used to extrapolate recent trends. A more consistent approach is to model this process within the stock assessment to allow the uncertainty to be propagated into the projections (Nielsen et al., this

issue). Assumptions about selectivity in the projections are more nuanced, because one could use a recent average at age, or allow the estimated process to continue, or explore the consequence of management action that is intended to reduce bycatch of recruits or to shift fully selected ages. The appropriate decision will depend on the management need, and there is value in comparing the impact of this assumption.

Similarly, the treatment of process error in recruitment (iid or autoregressive) should carry forward in projections. Johnson et al. (2016) found improved performance in projections (better statistical coverage) when correlation in recruitment deviations was included, however the penalized likelihood did a poor job at estimating the true correlation. For a given form of process error (iid or AR1), the recruitment fit within the assessment window will likely be very similar between the null and non-null SRR, however the assumed form of process error can impact both the projected trend as well as the CV of those predictions, especially for non-null SRR (Fig. 3). In the short term, autoregressive process error may be a better reflection of prevailing environmental conditions, but one should evaluate this assumption when the assessment is updated with additional years of data.

When an environmental covariate is included in the assessment model, then it should be incorporated in the projections. If there are no observations for the covariate outside of the assessment window, then the process model assumed for fitting it can have a much greater impact on results. Returning to the vignette described in the previous section,
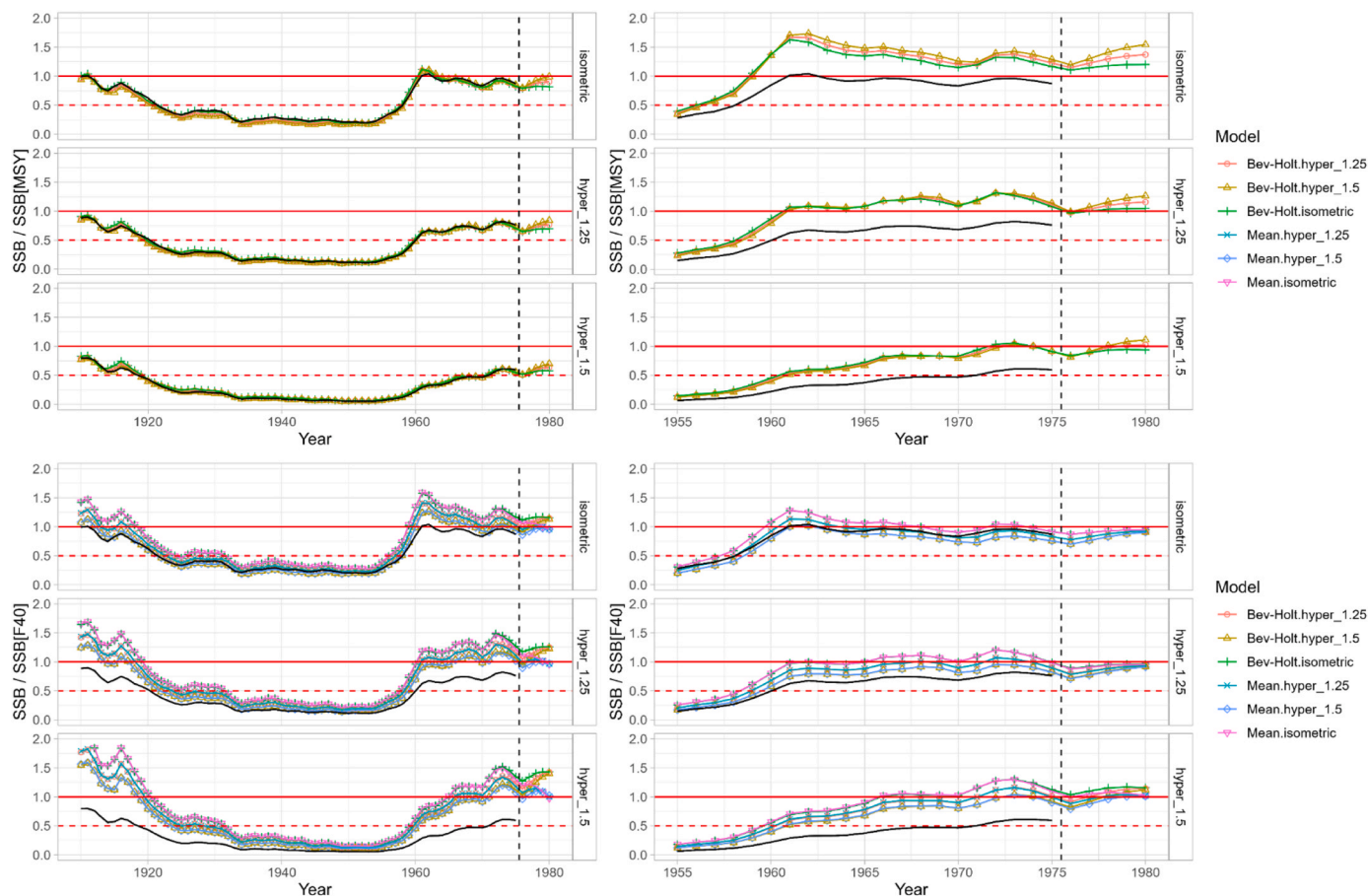
Fig. 5. Annual estimates of SSB relative to SSB[MSY] (top row) or SSB[F40] (bottom row), fitted to simulated data either from 1910 to 1975 (left column) or from 1955 to 1975 (right column) for two forms of stock recruit relationship (Mean or Beverton-Holt), and three possibilities for the exponent scaling weight to fecundity, *fecundity = weight^c*: isometric (c=1), or hyperallometric (c=1.25 or c=1.5). The vertical dashed line separates assessment years from projection years, and the true fecundity in the OM is indicated by the label on the right of each panel. The solid black line is the true operating model stock status. In the top row, only the three Beverton-Holt models estimate MSY, while in the bottom row all six models can estimate SSB associated with fishing at F40. The solid and dashed red lines indicate an example of target and limit reference points for SSB.

two alternative assumptions were made for how the environmental co-variate was treated in 10 year projections (while fishing at F40): either the random walk process was allowed to continue, or it was held constant at the most recent 5 year average of the CPI. When the random walk for the CPI was allowed to continue, the CV of recruitment dramatically increased in the projections, whereas holding the CPI at a recent 5 year average slightly reduced the CV compared to the models where the CPI had no influence because including CPI effects reduced the estimate of σ_R (Fig. 3). Minor differences in projected recruitment existed for the null SRR regardless of the form of process error in recruitment and NAA, with the AR1 process predicting slightly fewer recruits than an iid process; however, predicted recruitment and CV from the non-null SRR were much more sensitive to the process error. Taken together, it suggests that the null SRR is far less sensitive to assumptions in projections than the non-null SRR, especially with autoregressive process error, and is a sensible default.

It is common to present projection results in terms of trajectories for spawning biomass and catch. In the short term, those quantities will be primarily composed of age classes that were estimated in the assessment. To reflect uncertainty due to recruitment, it can also be informative to express the fraction of those projected quantities that is due to the terminal assessment year estimate (informed by only 1 year of data) and projected recruitment ("paper fish" that are not estimated from data and only reflect model assumptions; Table 1). Ideally, the assessment should be updated before that fraction becomes large (recognizing that the fractions will also be impacted by the accuracy of assumptions about

projected selectivity, maturity ogives, and weights at age).

Pragmatic recommendations for forecasting recruitment: maintain consistent assumptions in the assessment and projection time period with respect to biological parameters and process error structure, and evaluate the performance of those predictions when the assessment is updated. Autoregressive process error for recruitment (and potentially all NAA transitions, depending on assessment model structure) have been shown to perform better than iid in short term projections, and can be thought of as implicitly reflecting the most recent environmental conditions. The null SRR will be less sensitive than non-null SRR to assumptions about the process of environmental covariates, which supports adopting the null SRR as a robust default assessment decision. Limit the length of projections for advice so that important quantities (SSB, Catch) are not entirely reliant on paper fish.

2.3. Reference point decisions

Different management bodies have different frameworks for managing stocks, but all generally have some MSY-like underpining and a concept of a "backstop" (call it a limit or threshold, etc.) such that stocks are managed to remain above that point. I therefore refrain from specifics beyond discussing implications for MSY or proxy reference points and the associated backstop.

Many of the decision steps discussed previously have important interactions, and this is also true for the impact on reference points. Foremost of those decisions are the form of the SRR and the units of

fecundity. Reference points that correspond to Maximum Sustainable Yield (MSY) can be derived directly for non-null SRR (via replacement lines), and those should be estimated internal to the model so that the uncertainty and correlation structures are properly considered (Brooks and Deroba, 2015; Trijoulet et al., 2022). In these cases, the earlier decision about SRR parameterization when biological parameters vary is very influential. If the steepness parameterization was used, the resulting MSY reference points would be driven by the year(s) chosen for those biological parameters in the replacement line calculation, but the user (and working group members, and management bodies) would not be aware of the uncertainty and possible bias associated with that specification. The α,β parameterization recommendation in Miller and Brooks (2021) makes this decision a separate step. Also, the fecundity unit can shift the perceived importance of age classes for the reference point associated with spawning biomass (Minte-Vera et al., 2019). The rationale for choosing one suite of parameters over another may not be obvious (e.g., select values from recent years because prevailing conditions are likely to continue in the near term? Or, perhaps heavy exploitation is hypothesized to have reduced these values, so perhaps one chooses values from a period with lighter exploitation?). Similarly, if an environmental covariate is included, it will also lead to annual reference points (Xu et al., 2018), and yet another assumption is required to derive a long- term equilibrium value against which to judge current status. The repercussions of these decisions for management advice can be important (Legault and Palmer, 2016), and we should be transparent about the assumption we are making.

The null SRR presents some of the same dilemmas as the non-null SRR. While MSY reference points cannot be derived directly from replacement lines (Legault and Brooks, 2013), the lack of dependence of recruits on spawning biomass does not imply Fmax as the default reference point as some have suggested. The use of the null SRR is a recognition of low contrast in observations rather than a belief that the stock can support infinite fishing. The application of SPR proxies is an assertion that a threshold exists even if we can't estimate it. The basis of currently used proxies derives from analyses such as Clark (1991), (1993), (2002), and subsequent analytical work (e.g., Mace, 1994; Brooks et al., 2010), but only provide for "per recruit" reference points and corresponding fishing mortality but not absolute scale. A common approach is to scale those per recruit values by the estimated mean or median of the null SRR, which maintains consistency with how recruitment was modeled in the assessment period. Brooks and Legault (2015) explored several methods of scaling per recruit reference points and found that scaling by the estimated mean recruitment was fairly robust. While fitting an SRR outside of the assessment to estimated time series of recruits and SSB is sometimes attempted, the approach ignores the uncertainty, correlation, and myriad modeling decisions that produced those model estimates (Brooks and Deroba, 2015; Trijoulet et al., 2022), and recent simulated case studies in Trijoulet et al. (2022) demonstrated bias in selecting the correct SRR and the resulting reference points.

Revisiting the analyses in Section 2.1.2 (Units of fecundity) and Fig. 1, I illustrate the interaction of fecundity unit, SRR, and contrast in SSB in perceived stock status. The assessment data for that analysis were simulated from 3 operating models (OMs) that used one of the fecundity scalars previously explored (1.0, 1.25, or 1.5), and the same 6 estimating models (EMs) were fitted to each OM. The assessment fitting is repeated once with simulated data from 1910 to 1975 and a second time just with data from 1955 to 1975 using WHAM (Stock and Miller, 2021). For the longer time series, there is good contrast in SSB, and the non-null SRR is reasonably well determined but with slight positive bias in steepness (Table 2a), whereas the shorter time series has very low contrast and steepness was estimated near the upper bound of 1 for isometric and hyperallometric scaling (1.25) cases (Table 2b). The pattern of bias for estimated steepness, R0, and MSY values tends to mitigate the differences in fecundity such that SSB relative to a reference point is not as different as the values of the SSB time series on absolute scale. Consequently, stock status relative to MSY is accurate in the long time series and is indistinguishable even when the EM misspecified the fecundity scalar (Fig. 5, top left). The default SPR of 40% was used to calculate proxy reference points even though the true OM %SPR was 44.5%, 42.9%, or 41.7% for the three fecundity scalars. That introduced bias in the estimated stock status relative to SSB[F40], more so for the fecundity scalar of 1.5 (Fig. 5, bottom left). For the shorter time series, the SRR was nearly inestimable and the MSY reference points were very biased as was stock status (Fig. 5, top right). Cases with uninformative data such as these are precisely the situation where a null SRR is more pragmatic. The stock status estimates are nearly the same between the null and non-null SRR but bias in stock status is reduced when using the proxy compared to using MSY because the SRR was so poorly estimated (Fig. 5, bottom right). The main driver of the bias is primarily due to the misspecified fecundity scalar and also somewhat from the misspecification in using F40. The interaction of these three factors (fecundity unit, SRR, and contrast in SSB) can be important, and sensitivity analyses that explore alternative scalars for fecundity should carry those implications through to reference point and stock status estimates.

Pragmatic recommendations for reference points: be aware of annual replacement line if biological parameters vary and/or an environmental covariate is included in a non-null SRR, and for both null and

Table 2a

Estimation model (EM) estimates of key parameters related to SRR and reference points for data simulated from an operating model (OM) for years 1910–1975 under three scenarios of a scalar relating fecundity to weight at age. In the OM, steepness was 0.65, σ_R was 0.472, and R0 was 1e+ 7 for all cases. All EMs were fit in WHAM (Stock and Miller, 2021).

OM fecundity	EM fecundity	EM SRR	EM h	EM R0	EM σ_R	OM SPR[MSY]	EM SPR[MSY]	OM SSB[MSY]	EM SSB[MSY]	EM SSB{F40]
isometric	isometric	Mean	NA	NA	0.542	0.445	NA	4.83E+ 07	NA	3.54E+ 07
isometric	hyper_1.25	Mean	NA	NA	0.542	0.445	NA	4.83E+ 07	NA	5.62E+ 07
isometric	hyper_1.5	Mean	NA	NA	0.542	0.445	NA	4.83E+ 07	NA	9.09E+ 07
isometric	isometric	Bev-Holt	0.69	1.06E+ 07	0.468	0.445	0.424	4.83E+ 07	5.03E+ 07	3.56E+ 07
isometric	hyper_1.25	Bev-Holt	0.74	1.05E+ 07	0.467	0.445	0.380	4.83E+ 07	7.18E+ 07	5.65E+ 07
isometric	hyper_1.5	Bev-Holt	0.78	1.03E+ 07	0.467	0.445	0.341	4.83E+ 07	1.04E+ 08	9.15E+ 07
hyper_1.25	isometric	Mean	NA	NA	0.612	0.429	NA	7.28E+ 07	NA	2.85E+ 07
hyper_1.25	hyper_1.25	Mean	NA	NA	0.612	0.429	NA	7.28E+ 07	NA	4.52E+ 07
hyper_1.25	hyper_1.5	Mean	NA	NA	0.612	0.429	NA	7.28E+ 07	NA	7.31E+ 07
hyper_1.25	isometric	Bev-Holt	0.63	1.06E+ 07	0.460	0.429	0.454	7.28E+ 07	5.19E+ 07	2.87E+ 07
hyper_1.25	hyper_1.25	Bev-Holt	0.68	1.04E+ 07	0.458	0.429	0.409	7.28E+ 07	7.40E+ 07	4.55E+ 07
hyper_1.25	hyper_1.5	Bev-Holt	0.73	1.02E+ 07	0.458	0.429	0.369	7.28E+ 07	1.08E+ 08	7.37E+ 07
hyper_1.5	isometric	Mean	NA	NA	0.741	0.417	NA	1.13E+ 08	NA	2.19E+ 07
hyper_1.5	hyper_1.25	Mean	NA	NA	0.741	0.417	NA	1.13E+ 08	NA	3.47E+ 07
hyper_1.5	hyper_1.5	Mean	NA	NA	0.741	0.417	NA	1.13E+ 08	NA	5.62E+ 07
hyper_1.5	isometric	Bev-Holt	0.58	1.09E+ 07	0.451	0.417	0.484	1.13E+ 08	5.45E+ 07	2.20E+ 07
hyper_1.5	hyper_1.25	Bev-Holt	0.63	1.06E+ 07	0.449	0.417	0.440	1.13E+ 08	7.83E+ 07	3.50E+ 07
hyper_1.5	hyper_1.5	Bev-Holt	0.68	1.03E+ 07	0.448	0.417	0.400	1.13E+ 08	1.14E+ 08	5.67E+ 07

Table 2b

Estimation model (EM) estimates of key parameters related to SRR and reference points for data simulated from an operating model (OM) for years 1955–1975 under three scenarios of a scalar relating fecundity to weight at age. In the OM, steepness was 0.65, σ_R was 0.472, and R0 was 1e+ 7 for all cases. All EMs were fit in WHAM (Stock and Miller, 2021).

OM fecundity	EM fecundity	EM SRR	EM h	EM R0	EM σ_R	OM SPR[MSY]	EM SPR[MSY]	OM SSB[MSY]	EM SSB[MSY]	EM SSB{F40]
isometric	isometric	Mean	NA	NA	0.443	0.445	NA	4.83E+ 07	NA	4.31E+ 07
isometric	hyper_1.25	Mean	NA	NA	0.443	0.445	NA	4.83E+ 07	NA	6.85E+ 07
isometric	hyper_1.5	Mean	NA	NA	0.443	0.445	NA	4.83E+ 07	NA	1.11E+ 08
isometric	isometric	Bev-Holt	1.00	8.30E+ 06	0.443	0.445	0.303	4.83E+ 07	3.39E+ 07	4.31E+ 07
isometric	hyper_1.25	Bev-Holt	1.00	8.30E+ 06	0.443	0.445	0.261	4.83E+ 07	4.63E+ 07	6.85E+ 07
isometric	hyper_1.5	Bev-Holt	1.00	8.30E+ 06	0.443	0.445	0.226	4.83E+ 07	6.49E+ 07	1.11E+ 08
hyper_1.25	isometric	Mean	NA	NA	0.440	0.429	NA	7.28E+ 07	NA	3.39E+ 07
hyper_1.25	hyper_1.25	Mean	NA	NA	0.440	0.429	NA	7.28E+ 07	NA	5.39E+ 07
hyper_1.25	hyper_1.5	Mean	NA	NA	0.440	0.429	NA	7.28E+ 07	NA	8.72E+ 07
hyper_1.25	isometric	Bev-Holt	0.87	7.29E+ 06	0.432	0.429	0.342	7.28E+ 07	3.12E+ 07	3.40E+ 07
hyper_1.25	hyper_1.25	Bev-Holt	0.89	7.30E+ 06	0.432	0.429	0.303	7.28E+ 07	4.40E+ 07	5.41E+ 07
hyper_1.25	hyper_1.5	Bev-Holt	0.91	7.29E+ 06	0.432	0.429	0.270	7.28E+ 07	6.35E+ 07	8.75E+ 07
hyper_1.5	isometric	Mean	NA	NA	0.471	0.417	NA	1.13E+ 08	NA	2.38E+ 07
hyper_1.5	hyper_1.25	Mean	NA	NA	0.471	0.417	NA	1.13E+ 08	NA	3.78E+ 07
hyper_1.5	hyper_1.5	Mean	NA	NA	0.471	0.417	NA	1.13E+ 08	NA	6.12E+ 07
hyper_1.5	isometric	Bev-Holt	0.75	6.46E+ 06	0.403	0.417	0.398	1.13E+ 08	2.98E+ 07	2.41E+ 07
hyper_1.5	hyper_1.25	Bev-Holt	0.78	6.43E+ 06	0.402	0.417	0.358	1.13E+ 08	4.27E+ 07	3.83E+ 07
hyper_1.5	hyper_1.5	Bev-Holt	0.81	6.40E+ 06	0.401	0.417	0.324	1.13E+ 08	6.25E+ 07	6.20E+ 07

non-null be transparent about the selection of year(s) used to define the "equilibrium" biological parameters associated with MSY or SPR proxies. Parameterizing the non-null SRR with α,β instead of steepness allows separation between fitting the SRR and specification of biological parameters for calculating reference points. Be aware of the interaction of earlier modeling decisions about fecundity unit and carry sensitivity analyses through to reference point and stock status estimates.

3. Discussion

Recruitment is the age class for which data are typically the most sparse (and concomitantly most variable), and yet the way that it is modeled can have important impacts on reference points, catch advice, and rebuilding timelines. It is also an age class where we have the least ability to control directly – beyond limiting discards and protecting a portion of spawning biomass, we have no tools to dial recruitment up or down. The task of modeling recruitment has many challenges, and the pragmatic approaches herein are offered with the aim of reducing the potential for misspecification, exploring sensitivity analyses where warranted, and considering some additional diagnostics that may help evaluate robustness.

Some established approaches rely on stationarity (implicitly or explicitly), but this assumption will only become harder to justify given the changes already observed in the biological parameters of many stocks, including in the environment that they inhabit. This observation, and analysis in Miller and Brooks (2021), form the basis for the strong recommendation to move away from steepness parameterizations and priors. Additional concerns with priors derived from meta-analyses are the lack of details about assumed parameters used for the replacement lines (when those parameters vary), use of stock assessment output to estimate the SRR, and defensibility/compatibility with other assessment data of the fitted SRR for assessments that had originally been configured with the null SRR. And as an example of what could go wrong in practice, Thorson et al. (2019), detail an experience on the West Coast of the U.S. where updates to a meta-analysis for steepness of Pacific rockfishes led to large changes in the estimate of mean steepness, which affected rebuilding estimates. If the stock being assessed is data poor, it is even more unlikely that a SRR can be estimated, and for such stocks meta-analyses are often employed. I see this approach as useful for generating "if-then" scenarios based on the stocks sourced for the steepness prior and other life history assumptions, with the aim to identify and prioritize data that could help reduce the set of "if-then" scenarios. But in such contexts, careful scrutiny should be paid to the information used to derive the priors, and it should be conveyed that

results are assumption driven rather than strictly data driven.

Recommendations on the age of recruitment are both pragmatic and based on avoiding misspecifications such as having the same fish appear in both the x- and y-axes. Incidental fishing mortality that occurs on juveniles prior to the age at recruitment will reduce yield for the directed fishery on older individuals; therefore, if the age of recruitment in the model is defined so as not to include the incidental fishing mortality, it can then be accounted for in the assessment model and potentially addressed with management measures. Evaluating the magnitude of bias that results from rounding down the density-dependent window could be explored with simulated data using assessment models that allow specification of the age at recruitment.

Identifying the most appropriate unit of fecundity is a thorny issue, with so many nuances for each stock based on spawning behavior, age structure, reproductive condition, environmental state – all of which may vary over time and space. It is no wonder that we most often use weights, which are widely available, even though they have an explicit assumption of isometry. The consequences of not incorporating any of the myriad factors that could violate the isometry assumption can be explored by sensitivity analysis, where the hypotheses can be motivated from the available data on these different aspects of spawning. Continuing to collect data on fecundity (and other biological parameters) is important for evaluating whether parameters are changing over time.

The proposal to default to the null SRR is a pragmatic approach to deal with lack of contrast, model selection bias, possible misspecification due to non-stationary biological parameters, and the propagation of these issues into forecasts. It also removes the issue of age at first recruitment. Furthermore, modeling autoregressive process error in the null model produces recruitment estimates that are very similar to the non-null SRR (see discussion in Maunder and Thorson, 2019), and may mitigate differences in recruitment estimates for the most recent years. Differences may persist between the null and non-null SRR for historic data where data are sparser (especially if age composition data are lacking) and a wide variety of model assumptions are made (e.g., starting in equilibrium at unfished conditions, not estimating recruitment deviations and/or not bias correcting, fixing or tightly constraining both steepness and the variance about the SRR). Miller and Brooks (2021) recommend comparing results from assessments with the null and non-null SRR to assess the impact of these assumptions, and if results differ then to increase the assumed variance of recruitment deviations until those differences are negligible.

For the null SRR, adopting SPR proxies for reference points, scaled by mean recruitment, allows specification of both "target" and "limit" SSB,

thereby creating a backstop on the depletion of spawners. Treating recruitment as an autoregressive process (and the annual transition between ages beyond recruitment, if the model supports it), can implicitly account for unidentified factors that lead to recruitment in adjacent years being similar, and has been shown to reduce bias in short term forecasts (Johnson et al., 2016). Future research that explores performance of this approach with real rather than simulated data is strongly recommended.

The caveats about including environmental covariates, evaluating robustness through retrospective model selection, and evaluating sensitivity to projections are based on extensive literature review of real applications that don't stand the test of time, even case studies with high-resolution ocean modelling and long time series of survey observations (Plagányi-Lloyd et al., 2019). These model failures don't imply that the environment is not influential but rather reflect the recognition that a single factor is rarely the only driver in all years and we likely can't measure all influential drivers. This modeling conundrum where more than one driver could impact more than one biological process, with the balance and intensity across those processes changing through time, is a bit like trying to ascribe a retrospective pattern to just one cause – it is likely driven by more than one factor. This point is similar to the emphasis of the null SRR as an appropriate default – it doesn't reject the necessity of spawners, rather it deals with the lack of contrast in the available data. As further support for this perspective, simulations that explored multi-stage recruitment processes with different density-dependent mechanisms in each stage found that a non-null SRR fit between the first and last stage resembled the null SRR because each successive density-dependent stage reduces the contrast in scale on the final 'recruit' axis (Brooks et al., 2019).

In spite of the rationale against including environmental drivers, an argument for inclusion of a driver is to explore potential impacts to the stock under various future scenarios. In this case, by including the driver in the assessment and specifying the mechanism by which recruitment is affected, the fit and uncertainty associated with that driver can be carried forward into projections (as opposed to not including/fitting the index and then trying to explore those scenarios only in the projections). In cases where there are no observations of the driver outside of the assessment window, then hypotheses about future trends in the driver need to be explicit.

Forecasts from assessment models will always be uncertain, and it is recognized that recruitment in most cases is "essentially unpredictable" (Sharma et al., 2019). Both past performance and the age when recruits make important contributions to projected catch and SSB should be factors that limit the length of projections for management advice. Management frameworks sometimes require longer projections for rebuilding scenarios. Depending on the life history and generation time of the stock, those projections can be unrealistically long. Conveying the uncertainty and tracking progress regularly will be important so that corrections can be made along the way.

Recognizing that the length of forecasts should be limited presents challenges to the concept of reference points that are intended to reflect equilibrium conditions (particularly when reference points are derived from long term projections). Some analysts update reference points at each assessment update while others do not. Closed loop simulations to explore robust approaches to reference point calculations should be a high priority, and could include strategic analyses for providing advice under different environmental scenarios. Related to this, the original basis for default SPR proxies (Clark, 1991; Clark, 1993, e.g.) assumed constant biological parameters and used weights at age for fecundity. Revisiting the performance of default proxies when there is temporal variability in biological parameters and/or fecundity scaling is not isometric, for a wider range of life histories (the original analyses were based on groundfish stocks in the Northeast U.S.), could be undertaken in the same closed loop simulation just described. For the comparison of reference points in the simulation comparing isometry versus hyper-allometry, the exponent varied from 1.0 to 1.5 yet the %SPR$_{MSY}$ were

quite close (44.5–41.7%, respectively). For a given life-history, exploring when those SPR proxies diverge could help understand the potential risk of assuming isometry.

This suite of recommendations is largely informed by experience with stocks that have a well-defined spawning season, live for more than a few years, and have an identifiable ageing method. Additional characteristics informing this work are stocks with an exploitation history that predates both fisheries dependent and fisheries independent data collection, low contrast in estimated SSB time series, and inestimable SRR for fish stocks. There are plenty of exceptions to this profile, many of which require case-specific considerations. In this context, sex-changing fish, semelparous fish, and extremely short-lived fish probably need additional considerations, and experts to weigh in on those.

These recommendations also presume a reasonably data rich stock, but see Cope (this issue) for recommendations relative to data poor stocks. There is plenty of overlap with this recruitment topic and those of other keynote contributions, such as those on process variation (Nielsen et al., this issue), growth (Piner et al., this issue), stock structure (Cadrin et al., this issue), spatial models (Goethel and Berger, this issue), and diagnostics (Carvalho and Winker, this issue). Fruitful cross-fertilization of ideas would be a welcome result of this workshop.

While the above recommendations were made with age-structured models in mind, a number of these could also be applied to size-structured models. For example, many invertebrate assessments (crabs, lobsters, e.g.) use the null SRR and assume weight is proportional to fecundity. Because of sexual dimorphism in growth in some stocks, and management measures to protect egg bearing females, conditions for sperm limitation are more of a concern than they are for teleost fish (with the exception of hermaphroditic life histories).

Sensitivity analysis was recommended several times to explore the impact of assumptions on assessment results. An important aspect of sensitivity analysis is clearly conveying the risk associated with choosing one state of nature over the other. A consequence analysis, where catch advice from the base model is projected in the sensitivity model, and vice versa, is a further step that can be taken to illustrate risk to managers and other stakeholders (see, e.g. Northeast Fisheries Science Center (NEFSC) 2013). Additionally, comparing projection performance from the previous assessment (for both the base and sensitivity models) in a 'retrospective forecasting' analysis can be illuminating in terms of forecast skill and decision makers approach to risk (Brooks and Legault, 2016).

Best practice is often established by local precedent and expertise with a modeling framework, as well as common data idiosyncrasies (Li et al., 2021). It is expedient to adopt common strategies for stocks that share similarities in data characteristics and management frameworks, but it can also present a challenge for adopting new methods – the challenges of retraining, redeveloping software to produce diagnostics and reports, and explaining the basis for any impact on stock status and management advice, can be substantial. There can also be inertia in existing frameworks for conducting assessments that limit flexibility to introduce new approaches. Certainly best practice, and best implementation, can differ depending on the modelling context, management question and regional mandates. But refining methods and designing new tests and diagnostics to evaluate our practice is a common goal. With recent emphasis on moving towards open and reproducible fisheries science (Magnusson, this issue), there is an opportunity to archive model structure and data characteristics in order to synthesize and update our understanding of current practice and identify possible impediments to better practice.

CRediT authorship contribution statement

E.B. conceptualized the content of this manuscript, coded and interpreted all analyses, and wrote and edited.

Appendix A

This analysis addresses the statements that estimating a SRR at age 0 is likely to mischaracterize uncertainty and that the resulting SRR parameters are actually informed by a later age and are just scaled by density-independent mortality. Support is also provided for the statement that rounding down the recruitment (i.e. truncating a portion of density-dependent interval) creates bias in the SRR. Derivations follow from equations in Brooks and Powers (2007).

Consider a Beverton-Holt stock recruit relationship (BH SRR, hereafter), where recruitment occurs at age 1. In the interval between eggs (or spawning biomass), consider an arbitrary subdivision as in Fig. A1, such that $\Delta t0 + \Delta t1 = 1.0$. The same BH SRR can be obtained by modeling two separate BH SRR occurring in each interval, because the product of multiple BH SRR is again a BH (Beverton-Holt, 1957; Brooks and Powers, 2007). For this to hold, we must have recruitment in the first interval

$$R[y + \Delta t0] = \frac{\alpha_0 S[y]}{1 + \beta_0 S[y]} \tag{1}$$

and in the second interval,

$$R[y+1] = \frac{\alpha_1 R[y + \Delta t0]}{1 + \beta_0 R[y + \Delta t0]} \tag{2}$$

where

$$\alpha_0 = f\exp(-M_0 \Delta t0) \tag{3}$$

$$\alpha_1 = f\exp(-M_0 \Delta t1) \tag{4}$$

$$\beta_0 = fB_0 \frac{(1 - \exp(-M_0 \Delta t0))}{M_0} \tag{5}$$

and

$$\beta_1 = fB_1 \frac{(1 - \exp(-M_0 \Delta t1))}{M_0} \exp(-M_0 \Delta t0) \tag{6}$$

In (1−6), R is recruit, S is spawning biomass, α and β are the standard BH parameters, f is relative fecundity (converting biomass to numbers) and B_i is a density-dependent mortality term. The product of these two stage-specific BH SRR is

$$R[y+1] = \frac{\alpha S[y]}{1 + \beta S[y]} \tag{7}$$

where

$$\alpha = f\exp(-M_0 \Delta t0 - M_0 \Delta t1) = f\exp(-M_0) \tag{8}$$

and

$$\beta = \beta_0 + \beta_1 = fB \frac{(1 - \exp(-M_0))}{M_0} \tag{9}$$

In general, Eq. (9) will be true only if $B_0 = B_1 = B$. As the interval $\Delta t0$ gets smaller, approaching 0, there is still consistency in the SRR (including when $\Delta t0 = 0$).

Now, consider if instead of the second interval being defined by a BH SRR, it is defined to be a density-independent transition only, where that mortality is defined by an assessment scientist to be some scalar, c, of the true but unknown M_0. This would be an instance of 'rounding down', where some portion of the density-dependent window is treated as density-independent. Then we have recruitment in the first stage as

$$R[y + \Delta t0] = \frac{\alpha_0 S[y]}{1 + \beta_0 S[y]} \tag{10}$$

with α_0 and β_0 defined as in (3) and (5), and in the second stage as

$$R[y+1] = R[y+\Delta t0]\exp(-cM_0) \tag{11}$$

The product of this two-stage recruitment (one density-dependent BH, one density-independent) is

$$R[y+1] = \frac{\alpha_0 S[y]}{1+\beta_0 S[y]}\exp(-cM_0) \tag{12}$$

In order for this two-stage process to be equivalent to the original one-stage BH from age 0 to age 1, we must have

$$f\exp(-M_0\Delta t0 - cM_0\Delta t1) \equiv \alpha = f\exp(-M_0) \tag{13}$$

which is only true if $c = (1-\Delta t0)/\Delta t1$.

Additionally, we must have

$$fB_0\frac{(1-\exp(-M_0\Delta t0))}{M_0} \equiv \beta = fB\frac{(1-\exp(-M_0))}{M_0} \tag{14}$$

If $B_0 = B$, then Eq. (14) is true only if $\Delta t0 = 1$, which is not the case if you round down the recruitment interval. Lastly, as $\Delta t0$ approaches 0, the density dependent interval gets smaller and smaller, and when $\Delta t0 = 0$, then there is no density-dependence and you effectively have the null SRR with age 0 as the first age. This is the same point made in Eq. (1) of the main text.

From Brooks and Powers (2007) (see their Eq. 13, but noting that α and β are switched in their derivation), these results can be interpreted in terms of their effect on steepness (h) and unexploited recruitment (R0).

$$h = \frac{\alpha\varphi_0}{4+\alpha\varphi_0} \tag{15}$$

$$R0 = \frac{\alpha\varphi_{0-1}}{\beta\varphi_0} \tag{16}$$

Thus, so long as α is unbiased, h will be unbiased (which is true when both intervals are density-dependent, and is true in the second case when $c = (1-\Delta t0)/\Delta t1$. However, R0 depends on both α and β, and will be biased when either of those parameters are. When both intervals are density-dependent, β is unbiased when $B_0 = B_1 = B$. In the second case, β is biased and therefore, R0 would be as well.

A further consideration from the standpoint of estimation in the stock assessment model is the characterization of uncertainty. For the single stage SRR with recruitment at age 1, there is an associated σ^2 that is the variance about the SRR. What happens if you subdivide that SRR into two stages and treat the second stage as density-independent, and then fix σ^2? If the claim is that data at age 1 (or later) informs the SRR at time $\Delta t0$, then the range of R[t + $\Delta t0$] is likely to be greater than expected because the numbers at age 1 (or later) are scaled linearly by the assumed density-independent mortality rather than non-linearly if the second interval were density-dependent. The magnitude of bias in SRR parameters and recruitment deviations is hard to predict and likely to be very case specific. However, given that there are several assessment frameworks where the age of recruitment can be specified by the user, it would be valuable to simulation test this to better understand the consequences of starting a non-null SRR at age 0.
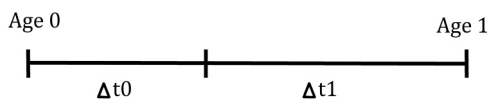

Fig. A1. Time line illustrating intervals between age 0 and age 1 where a stock recruit relationship might be subdivided and modeled as either density-dependent or density-independent.

.

Appendix A. Supporting information

References

Albertsen, C.M., Trijoulet, V., 2020. Model-based estimates of reference points in an age-based state-space stock assessment model. Fish. Res. 230, 105618.

Barneche, D.R., Robertson, D.R., White, C.R., Marshall, D.J., 2018. Fish reproductive energy output increases disproportionately with body size. Science 360, 642–645. https://doi.org/10.1126/science.aao6868.

Berger, A.M., 2019. Character of temporal variability in stock productivity inuences the utility of dynamic reference points. Fish. Res. 217, 185–197 https://doi.org/10.1016/j.fishr es.2018.11.028.

Beverton, R.J.H., Holt, S.J., 1957. On the dynamics of exploited fish populations. In: Fish and Fisheries Series No. 11, fascimile reprint, 1993. Chapman and Hall,, London.

Brooks, E.N., 2013. Effects of variable reproductive potential on reference points for fisheries management. Fish. Res. 138, 152–158. https://doi.org/10.1016/j.fishr es.2012.06.003.

Brooks, E.N., Powers, J.E., 2007. Generalized compensation in stock-recruit functions: properties and implications for management. ICES J. Mar. Sci. 64 (3), 413–424. https://doi.org/10.1093/icesj ms/fsl046.

Brooks, E.N., Deroba, J.J., 2015. When data" are not data: The pitfalls of post-hoc analyses that use stock assessment model output. Can. J. Fish. Aquat. Sci. 72 (4), 634–641. https://doi.org/10.1139/cjfas -2014-0231.

Brooks, E.N., Legault, C.M., 2016. Retrospective forecasting — evaluating performance of stock projections for New England groundfish stocks. Can. J. Fish. Aquat. Sci. 73, 935–950.

Brooks, E.N., Powers, J.E., Cortés, E., 2010. Analytical reference points for age-structured models: application to data-poor fisheries. ICES J. Mar. Sci. 67, 165–175. https://doi.org/10.1093/icesjms/fsp225.

Brooks, E.N., Thorson, J.T., Shertzer, K.W., Nash, R.D., Brodziak, J.K., Johnson, K.F., Klibansky, N., Wells, B.K., White, J., 2019. Paulik revisited: statistical framework and estimation performance f multistage recruitment functions. Fish. Res. 217, 58–70. https://doi.org/10.1016/j.fishres.2018.06.018.

Brooks, E.N. and Legault, C.M. 2015. Proxies and pragmatism: Approaches to scaling SPR-based reference points In Vieser, Jeffrey D. and Lynch, Patrick D. (eds.). Proceedings of the 12th National Stock Assessment Workshop. U.S. Dep. Commerce, NOAA Tech. Memo. F/SPO-162, 48 p.

Cadigan, N.G., 2013. Fitting a non-parametric stock-recruitment model in R that is useful for deriving MSY reference points and accounting for model uncertainty. ICES J. Mar. Sci. 70, 56–67. https://doi.org/10.1093/icesjms/fss183.

Clark, W.G., 1991. Groundfish exploitation rates based on life history parameters. Can. J. Fish. Aquat. Sci. 48, 734–750.

Clark, W.G., 1993. The effect of recruitment variability on the choice of a target level of spawning biomass per recruit. In: Kruse, G., Marasco, R.J., Pautzke, C., Quinn., T.J. (Eds.), In Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations, University of Alaska. Alaska Sea Grant College Program Report, 93-02-, pp. 233–246

Clark, W.G., 2002. F35% revisited ten years later. North Am. J. Fish. Manag. 22, 251–257.

de Valpine, P., Hastings, A., 2002. Fitting population models incorporating process noise and observation error. Ecol. Monogr. 72 (1), 57–76.

Deroba, J.J., 2014. Evaluating the consequences of adjusting fish stock assessment estimates of biomass for retrospective patterns using Mohn's rho. North Am. J. Fish. Manag. 34, 380–390.

Fitzhugh, G.R., Shertzer, K.W., Kellison, G.T., Wyanski, D.M., 2012. Review of size- and age-dependence in batch spawning: implications for stock assessment of fish species exhibiting indeterminate fecundity. Fish. Bull. 110, 413–425.

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

Haltuch, M.A., Brooks, E.N., Brodziak, J., Devine, J.A., Johnson, K.F., Klibansky, N., Nash, R.D.M., Payne, M.R., Shertzer, K.W., Subbey, S., Wells, B.K., 2019. Unraveling the recruitment problem: A review of environmentally-informed forecasting and management strategy evaluation. Fish. Res. 217, 198–216. https://doi.org/10.1016/j.fishres.2018.12.016.

Hixon, M.A., Johnson, D.W., Sogard, S.M., 2014. BOFFFFs: on the importance of conserving old-growth age structure in fishery populations. ICES J. Mar. Sci. 71 (8), 2171–2185.

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., et al., 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.

Ianelli, J.N., Fournier, D.A., 1998. Alternative age-structured analyses of the NRC simulated stock assessment data. In: Restrepo, V.R. (Ed.), In Analyses of simulated data sets in support of the NRC study on stock assessment methods. NOAA Tech. Memo. NMFS-F/SPO-30, pp. 81–96.

Ianelli, J.N., Stienessen, S., Honkalehto, T., Siddon, E., and Allen-Akselrun, C. 2022. Assessment of the walleye pollock stock in the Eastern Bering Sea. In Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions for 201722. North Pacific Fishery Management Council, Anchorage, AK. Available online at: ⟨https://apps-afsc.fisheries.noaa.gov/Plan_Team/2022/EBSPollock.pdf⟩.

Jardim, E., Millar, C.P., Mosqueira, I., Scott, F., Osio, G.C., Ferretti, M., Alzorriz, N., Orio, A., 2015. What if stock assessment is as simple as a linear model? The a4a initiative. ICES Journal of Marine Science 72 (1), 232–236. https://doi.org/10.1093/icesjms/fsu050.

Johnson, K.F., Council, E., Thorson, J.T., Brooks, E.N., Methot, R.D., Punt, A.E., 2016. Can autocorrelated recruitment be estimated using integrated assessment models and how does it affect population forecasts? Fish. Res 183, 222–232.

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 (6), 2244–2255. https://doi.org/10.1093/icesjms/fsab104.

Kell, L.T., Mosqueira, I., Grosjean, P., Fromentin, J.-M., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M.A., Poos, J.J., Scott, F., Scott, R.D., 2007. FLR: an open-source framework for the evaluation and development of management strategies. ICES J. Mar. Sci. 64, 640–646.

Lee, H.-H., Piner, K.R., Methot, R.D., Maunder, M.N., 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean (https://doi. org/). Fish. Res. 158, 138–146. https://doi.org/10.1016/j.fishres.2013.12.017.

Legault, C.M. 2009. Report of the Retrospective Working Group, 14–16 January 2008, Woods Hole, Mass. US Department of Commerce Northeast Fisheries Science Center Reference Document 09–01.

Legault, C.M., Restrepo, V.R., 1999. A flexible forward age-structured assessment program. ICCAT Collect. Vol. Sci. Pap. 49 (2), 246–253. SCRS/98/058.

Legault, C.M., Brooks, E.N., 2013. Can stock – recruitment points determine which spawning potential ratio is the best proxy for maximum sustainable yield reference points? ICES J. Mar. Sci. 70 (6), 1075–1080. https://doi.org/10.1093/icesjms/fst105.

Legault, C.M., Palmer, M., 2016. In what direction should the fishing mortality target change when natural mortality increases within an assessment? Can. J. Fish. Aquat. Sci. 73, 349–357.

Lorenzen, K., Camp, E.V., 2019. Density-dependence in the life history of fishes: when is a fish recruited? Fish. Res. 217, 5–10. https://doi.org/10.1016/j.fishres.2018.09.024.

Mace, P.M., 1994. Relationships between common biological reference points used as thresholds and targets of fisheries management strategies. Can. J. Fish. Aquat. Sci. 51, 110–122.

Mace, P.M., Doonan, I.J., 1988. A generalized bioeconomic simulation model for fish population dynamics. N. Z. Fish. Assess., Res. Doc., 88/4 51.

Marshall, D.J., Bode, M., Mangel, M., Arlinghaus, R., Dick, E.J., 2021. Reproductive hyperallometry and managing the world's fisheries. PNAS 118 (34). https://doi.org/10.1073/pnas.2100695118.

Marteinsdottir, G., Begg, G.A., 2002. Essential relationships incorporating the influence of age, size, and condition on variables required for estimation of reproductive potential in Atlantic cod, Gadus morhua. Mar. Ecol. Prog. Ser. 235, 235–256.

Maunder, M.N., Thorson, J.T., 2019. Modeling temporal variation in recruitment in fisheries stock assessment: a review of theory and practice. Fish. Res. 217, 71–86. https://doi.org/10.1016/j.fishres.2018.12.014.

McBride, R.S., Somarakis, S., Fitzhugh, G.R., Albert, A., Yaragina, N.A., Wuenschel, M.J., Alonso-Fernandez, A., Basilone, G., 2015. Energy acquisition and allocation to egg production in relation to fish reproductive strategies. Fish Fish 16 (1), 23–57. https://doi.org/10.1111/faf.12043.

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.

Miller, T.J., Brooks, E.N., 2021. Steepness is a slippery slope, 2021;00 Fish Fish 1–12. https://doi.org/10.1111/faf.12534.

Miller, T.J., Hare, J.A., Alade, L.A., 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern New England yellowtail flounder. Can. J. Fish. Aquat. Sci. 73, 1261–1270. https://doi.org/10.1139/cjfas-2015-0339.

Miller, T.J., O'Brien, L., Fratantoni, P.S., 2018. Temporal and environmental variation in growth and maturity and effects on management reference points of Georges Bank Atlantic cod. Can. J. Fish. Aquat. Sci. 75, 2159–2171. https://doi.org/10.1139/cjfas-2017-0124.

Minte-Vera, C.V., Maunder, M.N., Schaefer, K.M., Aires-da-Silva, A.M., 2019. The influence of metrics for spawning output on stock assessment results and evaluation of reference points: an illustration with yellowfin tuna in the eastern Pacific Ocean. Fish. Res. 217, 35–45. https://doi.org/10.1016/j.fishres.2018.09.022.

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.

Morgan, M.J., Shelton, P.A., Brattey, J., 2007. Age composition of the spawning stock does not always influence recruitment. J. North. Atl. Fish. Sci. Vol. 38, 1–12.

Murawski, S.A., Rago, P.J., Trippel, E.A., 2001. Impacts of demographic variation in spawning characteristics on reference points for fishery management. ICES J. Mar. Sci. 58, 1002–1014.

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

Myers, R.A.M., 1998. When do environment-recruitment correlations work? Rev. Fish. Biol. Fish. 8, 285–305.

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.

Northeast Fisheries Science Center (NEFSC), 2013. 55th northeast regional stock assessmentworkshop (55th saw) assessment summary report. Northeast Fish. Sci. Cent. Ref. Doc. 2013–11, 640–845. ⟨https://repository.library.noaa.gov/view/noaa/4432⟩.

Plagányi-Lloyd, E., Haywood, M., Gorton, R., Siple, M., Deng, R., 2019. Management implications of modelling fisheries recruitment. Fish. Res. 217, 169–184. https://doi.org/10.1016/j.fishres.2019.03.007.

Ricker, W.E., 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. 191.

Sculley, M., Ijima, H., Chang, Y.-J., 2018. A base-case model in Stock Synthesis 3.30 for the 2018 north Pacificswordfish (Xiphias gladius) stock assessment. PIFSC working paper ; WP-18-005. http://doi.org/10.7289/V5/WP-PIFSC-18-005.

SEDAR, 2018. SEDAR 52 Gulf of Mexico red snapper final stock assessment report. SEDAR. North Charlest. SC 434 (Available online at). ⟨https://sedarweb.org/documents/sedar-52-gulf-of-mexico-red-snapper-final-stock-assessment-report/⟩.

Sharma, R., Porch, C.E., Babcock, E.A., Maunder, M.N., Punt, A.E., 2019. Recruitment: Theory, estimation, and application in fishery stock assessment models. Fish. Res. 217, 1–4. https://doi.org/10.1016/j.fishres.2019.03.015.

Skern-Mauritzen, M., Ottersen, G., Handegard, N.O., Huse, G., Dingsør, G.E., Stenseth, N.C., Kjesbu, O.S., 2016. Ecosystem processes are rarely included in tactical fisheries management. Fish Fish 17, 165–175.

Stock, B.C., Miller, T.J., 2021. The Woods Hole assessment model (WHAM): A general state-space assessment framework that incorporates time-and age-varying processes via random effects and links to environmental covariates. Fish. Res. 240 https://doi.org/10.1016/j.fishres.2021.105967.

Subbey, S.J.A.Devine, Schaarschmidt, U., Nash, R.D.M., 2014. Modelling and forecasting stock-recruitment: current and future perspectives, 7 ICES J. Mar. Sci. 71 (8), 2307–2322. https://doi.org/10.1093/icesjms/fsu148.

Thorson, J.T., Dorn, M.W., Hamel, O.S., 2019. Steepness for west coast rockfishes: results from a twelve-year experiment in iterative regional meta-analysis. Fish. Res. 217, 11–20. https://doi.org/10.1016/j.fishres.2018.03.014.

Tomkiewicz, J., Morgan, M.J., Burnett, J., Saborido- Rey, F., 2003. Available information for estimating reproductive potential of northwest Atlantic groundfish stocks. J. Northwest Atl. Fish. Sci. 33, 1–21.

Trijoulet, V., Berg, C.W., Miller, D.C.M., Nielsen, A., Rindorf, A., Albertsen, C.M., 2022. Turning reference points inside out: comparing MSY reference points estimated inside and outside the assessement model. ICES J. Mar. Sci. 00, 1–13. https://doi.org/10.1093/icesjms/fsac047.

Trippel, E.A., 1998. Egg size and viability and seasonal offspring production of young Atlantic cod. Trans. Am. Fish. Soc. 127, 339–359.

Trippel, E.A., Neil, S.R.E., 2004. Maternal and seasonal differences in egg sizes and spawning activity of northwest Atlantic haddock (Melanogrammus aeglefinus) in relation to body size and condition. Can. J. Fish. Aquat. Sci. 61, 2097–2110.

Walters, C.J., Collie, J.S., 1988. Is research on environmental factors useful to fisheries management? Can. J. Fish. Aquat. Sci. 45, 1848–1854.

Williams, E.H. and K.W. Shertzer, 2015. Technical documentation of the Beaufort Assessment Model (BAM). NOAA Tech. Memo. NMFS-SEFSC-671, 43 p.

Wuenschel, M.J., McElroy, D., Oliveira, K., McBride, R.S., 2018. Measuring fish condition: an evaluation of new and old metrics for three species with contrasting life histories. Can. J. Fish. Aquat. Sci. 76 (6), 886–903. https://doi.org/10.1139/cjfas-2018-0076.

Zhou, S., 2007. Discriminating alternative stock–recruitment models and evaluating uncertainty in model structure. Fish. Res 86, 268–279.