



## Consequences of error in natural mortality and its estimation in stock assessment models

André E. Punt<sup>a,b,\*</sup>, Claudio Castillo-Jordán<sup>a,c</sup>, Owen S. Hamel<sup>c</sup>, Jason M. Cope<sup>c</sup>, Mark N. Maunder<sup>d,e</sup>, James N. Ianelli<sup>f</sup>

<sup>a</sup> School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, 98195, USA

<sup>b</sup> CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart, TAS, Australia

<sup>c</sup> Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, 2725 Montlake Boulevard East, Seattle, WA, 98112, USA

<sup>d</sup> Inter-American Tropical Tuna Commission, 8901 La Jolla Shores Dr., La Jolla, CA, 92037-1508, USA

<sup>e</sup> Center for the Advancement of Population Assessment, Methodology, La Jolla, CA, USA

<sup>f</sup> Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, Seattle, WA, USA

### ARTICLE INFO

Handled by Steven X. Cadrin

**Keywords:**

Fisheries management  
Natural mortality  
Parameter estimation  
Stock assessment

### ABSTRACT

Natural mortality ( $M$ ) is often considered to be one of the most important parameters in a fish stock assessment and affects productivity estimates for the population. However, it is also among the most difficult parameters to estimate using commonly available data. The magnitude of error (both bias and variance) when estimating this parameter can be substantial and can be affected by ignoring its variation over time, space, age, and length. In this study we explore the implications of errors in  $M$  on estimation and management performance using simulations and illustrative examples. The error in management reference points such as  $F_{35\%}$  and  $F_{MSY}$  is related directly to the error associated with  $M$ . Estimates of biomass are expected to be positively biased when  $M$  is overestimated and *vice versa*. Use of feedback management strategies reduces the impact of errors in  $M$ , but performance in meeting management objectives is compromised when  $M$  is in error. Estimating  $M$  was found to perform better than pre-specifying  $M$  in closed-loop simulations. Also, we found that the consequences of setting  $M$  to an incorrect value were reduced if stock-recruitment steepness was estimated. Based on our study and a review of related work, we advocate estimating  $M$  within an assessment, ideally with a prior for  $M$  tailored to the stock concerned.

### 1. Introduction

Many stock assessments are based on age- or size-structured population dynamics models fitted to available data on trends in abundance and on the age- and/or size-composition of the commercial/recreational and survey catches. These models include terms for the rate of death due to causes other than fishing, namely, natural mortality ( $M$ ). The basic equations for age- and size-structured models highlight the centrality of  $M$ :

$$N_{y+1,a+1} = N_{y,a} e^{-(M_{y,a} + F_{y,a})}; \quad N_{y+1,l} = \sum_l X_{l,a} N_{y,l} e^{-(M_{y,l} + F_{y,l})} \quad (1)$$

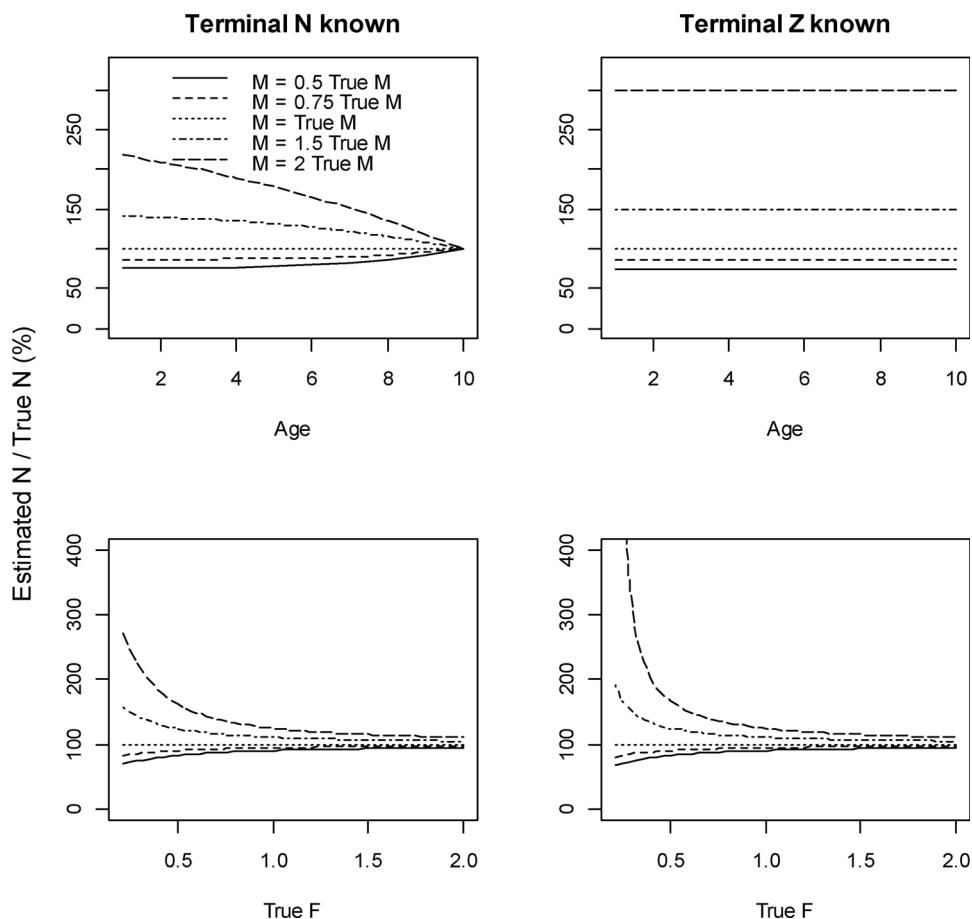
where  $N$  is numbers-at-age ( $a$ )/-size ( $l$ ),  $M$  is natural mortality by age/

size and year ( $y$ ),  $F$  is fishing mortality by age/size and year, and  $X$  is the size-transition matrix. Eq. 1 is necessarily a simplification of reality because  $M$  likely depends on other factors such as sex, stage and location, some of which are included in the population models on which stock assessments are based. This holds also for models such as delay-difference models (Deriso, 1980; Schntute, 1985) and per-recruit methods for estimating reference points.

$M$  is an important parameter in population models because it is a key determinant of productivity. Data to directly estimate  $M$  are rare, making it difficult to estimate, even in cases where it may be reasonably assumed to be independent of age and time (Beverton and Holt, 1957; Ricker, 1975; Vetter, 1988). In reality  $M$  is a function of age and sex, and likely changes over time due to a host of physiological and ecological

\* Corresponding author.

E-mail address: [aepunt@uw.edu](mailto:aepunt@uw.edu) (A.E. Punt).



**Fig. 1.** Relationship between the errors in the estimates of numbers-at-age (upper panels) and recruits (lower panels) for various levels of error in the estimate of natural mortality. The upper panels are based on an (age-independent) fishery mortality rate for all ages of  $0.3\text{yr}^{-1}$  while the lower panels show results for a range of (age-independent) fishery mortality rates.

factors such as predator numbers, food availability, disease, and environmental effects. These process factors can also be related to density-dependence (e.g., Fournier, 1983; Punt, 1996; Powers, 2014).

The difficulties associated with estimating  $M$  are compounded by the fact that age-specific  $M$  will be confounded with (a) declines in fish availability (or selectivity) with increasing age (Butterworth and Punt, 1990; Thompson, 1994), (b) the steepness of the stock-recruitment relationship, and (c) the growth rate. More generally, the confounding between time-specific trends in recruitment, age- (and time-) specific trends in  $M$ , time-specific trends in growth, and age- (and time-) specific trends in selectivity lead to an underdetermined problem (more parameters than data points) (Butterworth and Punt, 1990). This fact leads to strong assumptions and simplified models unless additional data can be brought to bear for parameter estimation. Common among the assumptions is that  $M$  is a constant, or at least separable into age and time components, and that selectivity is also separable, and ideally related functionally among ages and over time.

Model misspecification can arise if the assumptions made so that  $M$  is identifiable are wrong. Model-misspecification can lead to several problems for stock assessment, and hence the provision of management advice. These include: (a) biased estimates of the quantities on which management could be based, (b) incorrect estimates of precision, (c) retrospective patterns in outputs of the assessment (e.g., Brooks and Legault, 2016; Legault, 2009; Hurtado Ferro et al., 2015; Szwalski et al., 2018), (d) increased Type I and Type II errors associated with attempting to detect environmental or other drivers of biological and fishing processes (e.g., Lapointe and Peterman, 1991), and (e) an increased likelihood that model selection methods will select the

incorrect model (Jiao et al., 2012).

Several approaches have been proposed to estimate natural mortality  $M$  (see reviews by Vetter, 1988; Cope, this volume; Maunder et al., this volume). The approaches can be divided into those that attempt to estimate  $M$  (or  $M$ -at-age/size) outside the context of a stock assessment and those that attempt to estimate  $M$  (possibly by age, size, and time) within the context of a stock assessment. All of these approaches are, however, subject to error.

That stock assessment results are sensitive to the value assumed for  $M$  has been a concern since the development of the earliest methods of stock assessment based on age-structured models. For example, Ulltang (1977) explored the implications of errors in  $M$  on the results of Pope's (1972) cohort analysis method, while Lapointe et al. (1989, 1992) examined the impact of errors in estimated  $M$  on the results of virtual population analysis (VPA)-based assessments. Other authors who examined the impact of errors in  $M$  on the results of VPA-based assessments, include Sims (1984); Hilden (1988); Aksland (1994) and Mertz and Myers (1997). The earliest evaluations of the impact of error in  $M$  on estimates from methods such as VPA could be conducted analytically (see, for example, Fig. 1, which shows the sensitivity of the estimates of  $N$ -at-age based on VPA back-calculation). These early analyses highlighted the fact that errors in  $M$  have the largest impact for lightly fished populations.

Although the conclusions regarding the impact of errors in  $M$  for VPA-based approaches are informative, most of the stock assessments used for management purposes are now based on the integrated analysis approach, which separates the development of the model of the population dynamics from that of the model of how the observed data relate

to the predictions of the population dynamics model (Maunder and Punt, 2013). Stock assessments based on the integrated analysis approach typically represent uncertainty based on statistical likelihood (including Bayesian) methods (Magnusson et al., 2013). This assessment method often includes sensitivity analyses on assumptions of the models, as well as how the data are weighted (Maunder and Punt, 2013; Siddeek et al., 2017). Interestingly, some of the impacts of using an incorrect value of  $M$  appear to differ between VPA-based approaches (for which catch-at-age data are not included in the objective function minimized) and integrated analysis approaches (e.g., Deroba and Schueller, 2013). In this study, we focus on issues related to practices of specifying  $M$  using modern integrated methods.

The impact of  $M$  arises from both estimation of stock-status (e.g., current biomass and fishing mortality) in the stock assessment model and their analogous reference points (e.g., the biomass and fishing mortality corresponding to maximum sustainable yield [ $B_{MSY}$  and  $F_{MSY}$ ]; or the harvest control rule used for management). In general, a higher value of  $M$  corresponds to higher  $F_{MSY}$  and lower  $B_{MSY}$  as a fraction of the unexploited biomass ( $B_{MSY}/B_0$ ). Estimation in stock assessment models generally relies on two sources of information, indices of relative abundance and composition data, and  $M$  will influence interpretation of both of these types of information (Maunder and Piner, 2015). Information on absolute abundance from indices of relative abundance arises from measuring the impact of catch on the index as modified by the production function, which depends on  $M$ . Therefore, biased values of  $M$  will impact the information contained in the index. A higher value for  $M$  will mean that the population has to be larger to produce the same relative decline given the catch. Information on absolute abundance from composition data arises through following relative abundance of cohorts through time, and estimating total mortality ( $Z$ ) consistent with known catch levels. A higher value for  $M$  for the same  $Z$  will correspond to a lower fishing mortality ( $F$ ) given the catch, so over-estimation of  $M$  will result in a higher biomass. This illustrates that  $M$  can impact stock assessment estimates and consequent management advice. Biased  $M$  estimates may follow different patterns due to many interacting processes and other parameters being estimated simultaneously.

We first provide a brief summary of methods for estimating  $M$  (see Maunder et al., this volume, for details on likely levels of error when estimating  $M$ ). Next we describe what is known about the impact of errors in  $M$  on quantities used for management and management performance. We then outline the basis and details of a simulation study with a focus on identifying the impacts of misspecification of  $M$  (the operating model details are provided in Appendix A). We conclude with some general recommendations and issues needing further work.

## 2. Methods for estimating natural mortality and how well they perform

### 2.1. Empirical and theory based methods

There are many empirical and theory-based methods (Kenchington [2014], for example, outlines 29 published methods for estimating  $M$ ). Most of the methods depend on maximum observed age (e.g., Hoenig, 1983; Then et al., 2015), the parameters of the (von Bertalanffy) growth curve (e.g., Pauly, 1980), and/or age-at-maturity in various combinations. Some methods (e.g., Gunderson, 1980; Gunderson and Dygert, 1988) depend on measures of reproductive output. These methods are generally derived from a theoretical basis and use empirical estimates for establishing rules of thumb (or explicit formulae). Their performance lacks explicit tests per se, but prediction errors can be large (e.g., Pascual and Iribane, 1993; Kenchington, 2014).

### 2.2. Tagging methods

Tag-recapture studies follow individuals through time, and the rate of recaptured fish provides insights on natural mortality given mixing

assumptions for tagged fish relative to the population. The basis for estimating survival rates in most tagging methods is the 'Brownie model' (originally summarized by Brownie et al., 1985). Given information on reporting rates of recaptured fish (e.g., McDermott et al. 2005), this method allows estimation of natural and fishing mortality. The basic methods for estimating  $M$  (and other quantities) using tag-recapture data have been extended to make use of data on fishing effort, account for mixing rates of tagged animals, simultaneously estimate reporting and tag-loss rates, allow for different age-specific mortality rates (and probability of recapture), and account for spatial structure. Sibert (1984) and Hilborn (1990) developed multi-area models in which tagging data were used to estimate movement and mortality rates. Other examples extend from their approach and make use of tagging data combined with more conventional stock assessment data (e.g., Maunder, 1998, 2001; Hampton and Fournier, 2001; Goethel et al., 2011). These advances represent one of the best approaches towards understanding natural mortality rates.

### 2.3. Catch-curve methods

Several approaches for estimating  $M$  arise from catch-curve methods (e.g., regression of log-numbers on age). The classic approaches (e.g., Chapman and Robson, 1960) can estimate  $M$  by applying catch curve methods when fishing mortality is known to be negligible. Dunn et al. (2002) tested various estimators for catch-curve analyses using simulated data where the underlying assumptions are violated, and found all estimators to be biased and subject to imprecision. Another approach is to estimate a time-series of total mortality ( $Z$ ) based on catch-curve methods and then regress  $Z$  on effort  $E$  (where the intercept would be an estimate of  $M$ ) (Paloheimo, 1980; Paloheimo and Chen, 1996). Fournier and Archibald (1982) considered incorporating effort data in a population model that estimates total mortality  $Z$  and integrating the relationship  $Z = M + qE$ , where  $q$  is the catchability coefficient, within the model.

### 2.4. Estimating $M$ within standard integrated assessment models

Age- and size-structured stock assessments typically include  $M$  as a parameter. Thus, in principle at least,  $M$  can be estimated rather than being set based on one of the methods outlined above. Fournier and Archibald (1982) developed an early integrated assessment framework and identified conditions where  $M$  could reliably be estimated. Subsequently, most methods generally pre-specified  $M$  "since reliability of estimates for that parameter seem especially poor" (Deriso et al., 1985). Schnute and Richards (1995) noted that it is possible to estimate  $M$  within the context of a standard stock assessment (i.e., one that does not make use of data on consumption due to predators), and the practice has become more common in recent years. All of the major stock assessment software packages (Stock Synthesis (Methot and Wetzel, 2013); GADGET (Begley, 2014); MULTIFAN-CL (Fournier et al., 1998); CASAL (Bull et al., 2012); and SAM (Nielsen and Berg, 2014; Berg and Nielsen, 2016)) allow  $M$  to be treated as an estimable parameter. Typically, these stock assessment models are configured to have  $M$  estimated as a single parameter that is constant over age, sex, stage, time, etc. Even when data and assumptions prevent reliable estimates of  $M$ , standard practice is to construct likelihood profiles for  $M$  and/or examine the sensitivity of the results to alternative values of  $M$  (e.g., MacCall, 2013). For management purposes, sensitivity to the value for  $M$  has formed the "axis of uncertainty" in derived model quantities for several stock assessments (e.g., those for Pacific halibut *Hippoglossus stenolepis* (Stewart and Martell, 2015), cowcod *Sebastodes levis* (Dick and He, 2019), aurora rockfish *Sebastodes aurora* (Hamel et al., 2013), and petrale sole *Eopsetta jordani* (Haltuch et al., 2013)). When estimating  $M$ , the development of a defensible prior (e.g., based on the approach of Hamel (2015) or Cope, this volume) is critical. If the data and assessment model configuration inform the estimate of  $M$ , the estimate (and its uncertainty) may depart

from the prior distribution but the prior distribution may be highly influential on the results for assessments with uninformative data. Values of  $M$  also affect fitting indices of relative abundance and composition data. In some cases, alternative values of  $M$  can improve fits by compensating for other model misspecification issues. To help diagnose such cases, Lee et al. (2011) argue that unrealistic estimates of  $M$  can identify model misspecification (see Piner et al., 2011, for a description of this concept).

Although  $M$  may be estimable in principle, the precision with which it can be estimated depends on the quality and contrast of the available data (Lee et al., 2011). Situations in which estimation is likely to be most successful are those in which fishing mortality is low (e.g., arrowtooth flounder *Atheresthes stomias* in Alaska; Wilderbuer and Turnock, 2009), when fishery or survey selectivity are asymptotic, or those in which data collection started soon after the fishery. In relation to the latter situation, Punt et al. (2001) estimated  $M$  (by sex) for blue grenadier (*Macruronus novaezealandiae*) off southeast Australia given age-composition data from virtually the start of the target fishery, which provided information on the unfished population age-structure. Other situations in which estimates of  $M$  are likely to be reliable include when survey data are available for age/size-classes that are not available to the fishery, because (a) the fishery selectivity is such that the youngest age-classes are not vulnerable to the fishery (e.g., for North Pacific crab stocks; e.g., Stockhausen, 2019; Szuwalski, 2019), or the fishery was closed (e.g., Cook, 2004; Ianelli et al., 2015). In contrast, simulation studies (e.g., Lee et al., 2011; Maunder and Wong, 2011) show that  $M$  can be estimated with only moderate bias and imprecision, even when sex- or stage-structured, if composition data are available from fisheries and/or surveys (as long as model specification or data assumptions are correct).

Age-specific natural mortality can be estimated as discrete changes in  $M$  that may coincide with the onset of sexual maturity (e.g., eastern Bering Sea snow and Tanner crab *Chionoecetes opilio* and *C. bairdi*; Stockhausen, 2019; Szuwalski, 2019). In these cases, a piecewise linear function of age (e.g., Punt et al., 2014) may be used. In other cases, a continuous functional form, perhaps allowing for  $M$  to be higher at both lower and higher ages, may be most appropriate (e.g., the Chen and Watanabe (1989), Lorenzen (1996) and Siler (1979) functions). Modelling  $M$  that varies over time is typically done as either a random walk or as discrete changes (see Jiao et al. (2012) for an age-structured statistical catch-at-age model that allows for a variety of formulations for age- and time-variation in natural mortality). The random effects state-space model SAM (Nielsen and Berg, 2014; Berg and Nielsen, 2016) and some Bayesian models (e.g., as shown by Lewy and Nielsen, 2003; and Aanes et al., 2007) allow  $M$  to change over time and age.

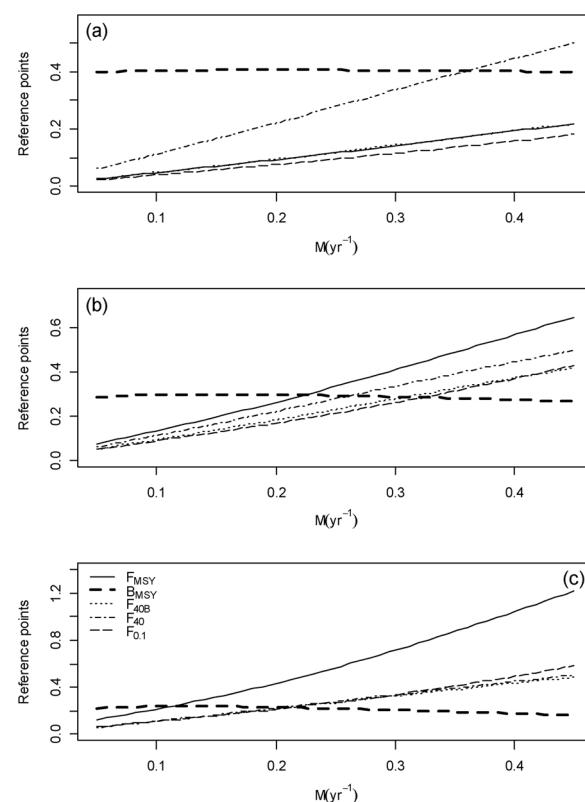
## 2.5. Estimating $M$ by including predation in stock assessments

There are many ways in which predation mortality can be included in stock assessments (Plagányi, 2007). The simplest way to jointly model multiple stocks, and hence the interactions among them, is the approach of using correlated processes to describe their respective population trends (with implications for shared mortality impacts; Albertsen et al., 2017). A more explicit formulation to include predation mortality in a population model is to treat  $M$  as the sum of a residual (or basal) rate of natural mortality and to add estimated mortality due to predation mortality, i.e.:

$$M_{y,a}^i = M_a^{i,residual} + \sum_j M_{y,a}^{i,j} \quad (2)$$

where  $M_{y,a}^i$  is the rate of natural mortality for animals of age  $a$  and species  $i$  during year  $y$ ,  $M_a^{i,residual}$  is the residual natural mortality for animals of species  $i$  and age  $a$ , and  $M_{y,a}^{i,j}$  is the rate of natural mortality for animals of species  $i$  and age  $a$  during year  $y$  due to predator  $j$ .

Extended single-species models represent  $M_{y,a}^i$  as the product of the



**Fig. 2.** Relationship between five reference points and the assumed value for  $M$ . Results are shown in (a) – (c) for steepness values of 0.4, 0.75, and 0.95 respectively.

(pre-specified) number of predators of species  $j$  during year  $y$ ,  $E_y^j$ , and a ‘catchability’ coefficient,  $q_a^{ij}$  i.e.  $M_{y,a}^{i,j} = q_a^{ij} E_y^j$  (Livingston and Methot, 1998; Hollowed et al., 2000). More complete trophic interactions within stock assessments were first represented using extended multi-species VPA (e.g., MSVPA; Gislason and Helgason, 1985; Sparre, 1991; Magnusson, 1995), where the index  $E_y^j$  above is replaced by abundance from a simultaneously modelled population of predators. Several integrated multispecies stock assessment methods (e.g., Jurado-Molina et al., 2005, 2006; Kinzey and Punt, 2009; Van Kirk et al., 2010, 2015; Curti et al., 2013; Tsehay et al., 2014; Ross-Gillespie, 2016; Holsman et al., 2016; Trioulet et al., 2019) have been developed in recent years constructed around the predation models on which MSVPA is based and generalizations thereof. These methods make use of the stomach content data for parameter estimation and differ in terms of how predation mortality is modelled and parameterized. In general, these latter methods adopt a more statistical approach to adding diet data and in theory, can perform well in providing both time- and age-specific estimates of natural mortality.

## 3. Implications of errors in estimates of $M$

This section first reviews past studies related to how errors in  $M$  impact the quantities on which management advice is based (i.e., reference points), the estimation of biomass and recruitment, and the performance of management strategies. It then (section 4) describes and reports on a simulation study that examines specific questions about the impact of error in  $M$  on estimation and management performance. As noted in the introduction, the focus here is on methods based on (or related to) integrated analysis assessments that use age- or size-structured population dynamics models.

### 3.1. Per-recruit analysis and reference points

Per-recruit analyses (spawning biomass-per-recruit and yield-per-recruit) are used to estimate reference points (e.g.,  $F_{0.1}$ ,  $F_{\text{MAX}}$ ,  $F_x$  [the fishing mortality rate at which spawning biomass-per-recruit is x% of that in an unfished state]), while per-recruit analysis in combination with a stock-recruitment relationship determines reference points such as  $B_{\text{MSY}}$ ,  $F_{\text{MSY}}$ ,  $F_{\text{crash}}$  (the lowest fishing mortality rate at which the population would be zero in steady-state), and  $F_{x\text{B}}$  (the fishing mortality rate at which spawning biomass is x% of that in an unfished state) (see Sissenwine and Shepherd (1987) for the basic approach). There is a very direct and clear relationship between error in  $M$  and errors in the reference points ( $F_{\text{MSY}}$ ,  $F_{0.1}$ ,  $F_{40}$  and  $F_{40\text{B}}$ ; Fig. 2<sup>1</sup>), suggesting that values for fishing mortality-related reference points will be over-estimated if  $M$  is overestimated and vice versa, a result consistent with those of simulation studies such as that of He et al. (2011) and the analytical analysis by Clark (1999). The rate of increase in the fishing mortality reference points with  $M$  depends on steepness (particularly when the stock-recruitment relationship has the Beverton-Holt form), with  $F_{\text{MSY}}$  increasing very rapidly with increasing  $M$  for a steepness of 0.95 but much less so for a steepness of 0.4. The value of  $B_{\text{MSY}}$  depends on steepness, but is not impacted substantially by the value for  $M$  if steepness is correctly specified.

The analyses of Fig. 2 and Supplementary Fig. 3 are based on the assumption that the other parameters involved in per-recruit analysis and the calculation of yield functions (growth, selectivity, and the form and parameters of the stock-recruitment relationship) are known. Clark (1999) notes that the impact of error in  $M$  is compounded in relation to the target harvest rate because an overestimate of  $M$  will also lead to an overestimate of the slope of the stock-recruitment relationship (although this conclusion is not supported by the results of the simulation study of this paper; section 4.4).

### 3.2. Estimation of biomass

Several studies have examined the ability of integrated analysis stock assessments to estimate quantities of management importance given uncertainty in  $M$  (Table 1). Wetzel and Punt (2011) evaluated the performance of the Stock Synthesis (an integrated analysis framework; Methot and Wetzel, 2013) for scenarios in which  $M$  was known and in which it was uncertain, but with an expected value (over simulations) equal to the true value. The impact of a lack of data (in particular survey data) was found to have a much larger impact on estimation performance than uncertainty in  $M$ . However, the cases in which  $M$  was uncertain led to greater inter-simulation variability in estimates of management-related quantities, in particular current relative biomass.

He et al. (2011) explored the impact of misspecification of  $M$  and the interaction between misspecification of  $M$  and whether selectivity is logistic or dome-shaped. This trade-off (often referred to as ‘hide them’ and ‘kill them’ hypotheses) was highlighted by Thompson (1994) and illustrated by the assessment of black rockfish *Sebastodes melanops* (Cope et al., 2016). As expected, He et al. (2011) found a strong interaction between the estimate of  $M$  and selectivity, and concluded that  $M$  was estimable if independent of age (and time) and when selectivity was logistic, but that estimation performance was very poor when  $M$  for young and old fish was high and selectivity was dome-shaped. He et al. (2011) found that depletion (current biomass relative to unfished biomass) could be estimated well even when natural  $M$  was misspecified. In contrast, the parameters of the stock-recruitment relationship, those governing age-specific selectivity, catchability and consequently overfishing levels were biased in many cases.

Deroba and Schueller (2013) compared a statistical catch-at-age

analysis (SCAA) method with pre-specified constant  $M$ , age-specific  $M$ , and time-varying  $M$  in the estimation method. Unlike Lapointe et al. (1989), the results were robust to trends in fishing mortality (perhaps due to differences between the VPA- and SCAA-based approaches). Including age-specific  $M$  did not improve the ability to estimate trends in biomass and recruitment in contrast to the results of Ulltang (1977) and He et al. (2011), although this may be due to fact that the assumed constant value of  $M$  was correct on average. Allowing for time-variation in  $M$  improved estimation performance when  $M$  was highly variable over time, but with larger uncertainty in the estimates of biomass and recruitment. The impact of errors in  $M$  for the long-lived species were found to be more substantial for estimation of spawning biomass, with the opposite conclusion for recruitment.

Johnson et al. (2015) compared the performances of a single age- and time-invariant  $M$ , specified at one of the three levels (high, medium, and low) or estimated in terms of estimating spawning biomass, fishing mortality and the outcomes of a harvest control rule. The operating model included an age-invariant but time-varying  $M$ . Fixing  $M$  higher than the average true value led to the poorest performance. The most robust approach to specifying time-invariant  $M$  in an estimation model in cases where  $M$  was thought to vary across time was to estimate  $M$ . Their results were qualitatively insensitive to trends in fishing mortality, confirming the results of He et al. (2011).

Fu and Quinn (2000) found that estimating  $M$  (even when it was quite poorly estimated) was better than fixing it at an incorrect value in terms of estimating fishing mortality, recruitment, and spawning stock biomass irrespective of how catchability was modelled.

There has been little simulation testing of multispecies stock assessments. Kinzey (2010) found that it was difficult to reliably estimate the parameters of complex (multi-parameter) predation functions using the data for Aleutian Islands Pacific cod, walleye pollock and Atka mackerel. This result supports the decision in most multispecies assessment methods to assume either a Holling Type I (linear increase with prey density) or a Holling Type II (decelerating increase with prey density) (Holling, 1959) feeding functional relationship.

The multi-species method of Van Kirk et al. (2015) performed reasonably in self-tests, but model misspecification in survey selectivity curves combined with uninformative diet data led to a degradation in estimation performance (estimates of median age-1 abundance and  $M_0$  respectively up to 40 % and 20 % different from the true values).

Trijoulet et al. (2019) examined the impact of the amount of stomach content data on estimation performance and subsequently tested their estimation method, which confirmed that ignoring time-varying predation mortality led to large biases in parameters and model outputs (Trijoulet et al., 2020).

### 3.3. Evaluation of management strategies

Management Strategy Evaluation (MSE; Butterworth and Punt, 1999; Bunnefeld et al., 2011; Punt et al., 2016) is the state-of-the-art method for evaluating the performance of management strategies in terms of the ability to achieve management goals. An MSE consists of several components of which two are the operating models (the models that represent the system to be managed) and the management strategies being evaluated (management strategies are often combinations of data collection methods, estimation methods, and harvest control rules, although the latter two components can be combined). Stock assessments can form the basis for either of these components. The range of uncertainties considered in operating models are usually much broader than the range conventionally considered in stock assessments (Punt et al., 2016). Many MSEs allow for uncertainty in  $M$  during the process of setting the values for the parameters of the operating model (i.e., the conditioning process). However, relatively few MSE (or closed loop simulation) studies have specifically focused on uncertainty due to  $M$ . The few such cases include:

<sup>1</sup> Results for the Ricker stock-recruitment relationship are shown in Supplementary Fig. 3

**Table 1**

Simulation studies that have explored estimation performance based on integrated population model approaches for estimating  $M$  within the context of single-species assessment methods, the quantities of management interest considered and the key conclusions related to estimation of  $M$  and other quantities of management interest.

Study	Species/stock	Type of assessment method (model-type; estimation framework)	Quantities	Key conclusions
Deroba and Scheueller (2013)	Generic long-lived and short-lived species	Age-structured (PML)	SSB, Rec	The scale and trends among years in bias was generally similar when specifying a constant $M$ and an age-varying $M$ . Annually varying $M$ in the EM generally reduced bias when $M$ was really time-varying Bias was more marked for SSB than Rec for the long-lived species, but the opposite was true for short-lived species.
Fu and Quinn (2000)	Pink shrimp in Kachemak Bay	Size-structured (sum of squares)	$M$ , SSB, Rec, growth parameters	Mean absolute deviation was lower when $M$ was time-varying and $M$ was estimated / selectivity was time-varying and was estimated.
He et al. (2011)	Widow rockfish	Age-structured (PML)	SSB depletion, steepness, selectivity, $q$	Misspecification of age-specific $M$ had a negligible effect on estimation performance for SSB depletion but greatly impacted estimation of $SSB_0$ , steepness, selectivity and $q$ . $M$ and selectivity could not both be adequately estimated if $M$ was higher in old fish and selectivity was dome-shaped.
Jiao et al. (2012)	Atlantic weakfish	Age-structured (Bayesian)	$F$ , SSB, Recr, reference points	The probability of selecting the correct model (using DIC) was adequate when $M$ was constant or time-varying but not when it was age-specific. Relative errors calculated from the "best model" selected based on DIC were low and close to those for the true model.
Johnson et al. (2015)	cod-like, flatfish-like, and sardine-like	Age-structured (PML)	SSB, $F$ , $M$ , $R_0$ , $q$	The most robust approach to specifying $M$ when it is thought to be time-varying was to estimate it. The least robust approach for most scenarios examined was to fix $M$ at a high value (large positive bias), suggesting that the consequences of misspecifying $M$ are asymmetric.
Kanaiwa et al. (2008)	American lobster	Size-structured (PML)	Rec, biomass and abundance by size-range, exploitation rate	Temporal trends in $M$ and biased estimates of growth parameters led to the largest relative errors.
Kienzle et al. (2016)	Brown tiger prawns in Morton Bay	Delay-difference (maximum likelihood)	$M$ , $q$ , growth parameters, initial biomass	It is possible to estimate all parameters adequately ( $M$ was estimated with less than 22% discrepancy).
Lee et al. (2011)	Groundfish stocks	Age-structured (PML)	$M$	Median relative error in estimating $M$ ranged between -0.306 to 0.117.
Sippel et al. (2017)	Groundfish stocks	Age-structured (PML)	$M$	The production model diagnostic can be used to assess when $M$ is estimable (i.e., the assessment model converged).
Maunder and Wong (2011)	US mid-Atlantic summer flounder	Age-structured (PML)	$M$	It was possible to estimate both female and male $M$ with moderate precision, but with some bias depending on the true value of $M$ , depending on the model assumptions and the type of data available. The estimates of $M$ were more precise than those from relationships with life history parameter.
Szuwalski and Punt (2012)	Eastern Bering Sea snow crab	Size-structured (PML)	MMB, OFL, FOFL, $B_{MSY}$ , $F_{35\%}$ , Rec, F	All management quantities were negatively biased, except MMB and $B_{MSY}$ , which were positively biased when $M$ was negatively biased. The opposite trend was apparent when $M$ was positively biased.
Schnute and Richards (1995)	Generic fish	Age-structured (maximum likelihood)	$M$ , $q$ , stock-recruitment parameters, variance parameters	Increased fishing contrast improved estimates of some model parameters, but not $M$ . $M$ might be estimable along with other parameters when catch-age data are available back to the first year of the fishery.
Vincent et al. (2017)	Lake Erie walleye	Age-structured (PML)	Abundance, Rec, movement, $M$	Integrated tag-recapture models were able to simultaneously estimate movement rates, $M$ , and tag reporting rates, though accuracy and precision of model estimates decreased with greater model complexity and fewer tags released.
Wang and Liu (2006)	Generic species	Age-structured (sum of squares)	$M$	The ability to estimate $M$ depended on recruitment trends. $M$ was biased when recruitment had a CV of 10%. Estimation of $M$ was best for a one-way trip fishery, while a recovering fishery had the worst performance.
Wetzel and Punt (2011)	West Coast Groundfish	Age-structured (PML)	Depletion, $R_0$ , $SSB_0$ , $F_{40\%}$ , Yield	The estimation method provided negatively biased estimates of current SSB relative to the unfished level (final depletion) when only recent survey indices were available. Estimation of quantities of management interest improved substantially even when only minimal-length-composition data from the survey were available. However, the

(continued on next page)

**Table 1** (continued)

Study	Species/stock	Type of assessment method (model-type; estimation framework)	Quantities	Key conclusions
				estimates of some quantities (final depletion and $SSB_0$ ) remained biased (either positively or negatively) even in the scenarios with the most data.

Abbreviations: PML Penalized maximum likelihood;  $SSB$  spawning biomass ( $SSB_0$  is unfished  $SSB$ ); Rec Recruitment ( $R_0$  is unfished recruitment);  $F$  Fishing mortality ( $F_{x\%}$  is the fishing mortality corresponding to a reduction of spawning biomass-per-recruit to  $x\%$ );  $q$  catchability.

**Table 2**

The operating model scenarios. “(x,y)” relates to changes in the historical fully-selected fishing mortality (representing the ratio of  $F$  to  $F_{MSY}$ ). A dash indicates that the specifications are the same as those for the reference operating model.

Reference	M	(x,y)
	Constant	2, 0.5
A	Age-specific	–
B	Random	–
C	Increasing	–
D	Decreasing	–
E	–	1.25, 0.5
F	–	0.5, 0.5

- [Forrest et al. \(2018\)](#) evaluated the performance of threshold management strategies for Hecate Strait Pacific cod and Hecate Strait Rock Sole (*Lepidopsetta* spp.). The simulations allowed for compensatory natural mortality for Pacific cod and alternative levels of natural mortality for rock sole.
- [Punt et al. \(2013\)](#) examined the estimation performance of a size-structured stock assessment and its performance when combined with a harvest control rule for southern rock lobster *Jasus edwadssii*. They found that increasing  $M$  in the operating model led to positively biased estimates of exploitable biomass. This positive bias not only impacted estimates during the projection period, but also for the years before the management strategy was first applied, suggesting that retrospective analysis may have detected the level of change in  $M$  considered. However, even given the biases due to time-varying  $M$ , the management strategy reduced catches and did not have as severe an impact as increasing catchability (the assessment relied on CPUE as an index of abundance).
- [Hordyk et al. \(2019\)](#) conducted closed-loop simulations based on four species, arrowtooth flounder, Pacific hake *Merluccius productus*, Pacific ocean perch *Sebastodes alutus*, and silver warehou *Seriola punctata* to explore the implications of six types of misspecification, including misspecification of  $M$ . The total allowable catches (TACs) on which the projections were based were determined using a stock reduction analysis and a harvest control rule that aimed to set fishing mortality equal to  $F_{MSY}$ . The results confirmed the expectation that risks are asymmetric, with an over-estimate of  $M$  leading eventually to lower catches than when  $M$  was underestimated. Recommended catches in the short-term with higher  $M$  were overestimated but due to the stocks collapsing over the longer-term, the bias in  $M$  led to the most asymmetric outcomes of the factors considered. They showed that for their study, the impact of bias in  $M$  depended on life-history, with the impact less for the longer-lived Pacific ocean perch.

Many data-poor management strategies require an estimate of  $M$  (see, for example, [Table 2](#) of Carruthers et al. [2014]). [Carruthers et al. \(2014\)](#) evaluated the performances of the several data-limited management strategies, along with several strategies that did not rely of estimates of  $M$  using a forerunner of the Data-Limited Methods toolkit (DLMtool). The operating models on which the simulations were based allowed for bias in the estimate of  $M$  with a CV of 0.5, along with bias (and imprecision) in the other inputs on which management strategies are based. [Sun et al. \(2018\)](#) evaluated data-limited management

strategies using DLMtool where the values for  $M$  were set using the maximum age and growth approaches of [Then et al. \(2015\)](#). Performance differed among the approaches used to estimate  $M$  (and other inputs to the data-limited management strategies), with an approach based on ELEFAN and the age-length approach of [Morgan \(1987\)](#) combined with the growth approach of [Then et al. \(2015\)](#) performing best.

#### 4. Simulation study

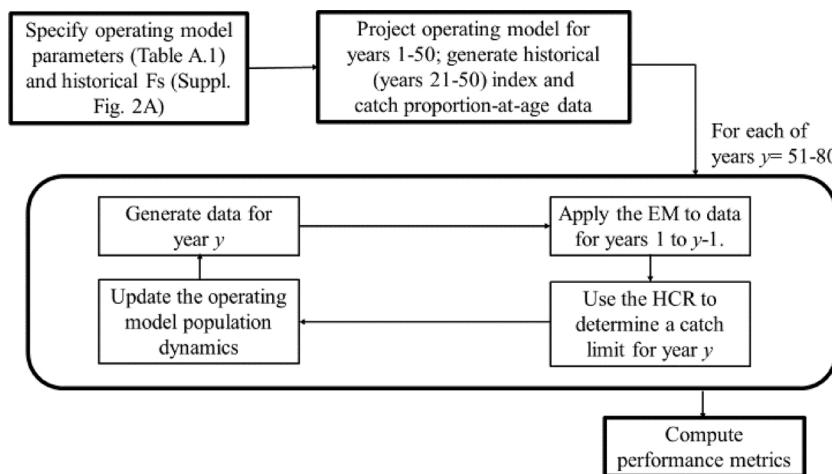
We conducted a limited simulation study to further explore aspects of how treatment of  $M$  affects assessments and management performance. To focus the study, a simple operating model was used to isolate and identify the primary driving factors in terms of both estimation and management performance.

##### 4.1. Operating model

The operating model was configured as a single-sex age-structured model parameterized to mimic a cod-like species (Appendix A). Fishing mortality  $F$  was modelled to have a “separable” time-invariant (logistic) age-based selectivity pattern. Over time,  $F$  was specified to be negligible for 10 years and then to increase linearly to  $xF_{MSY}$  in year 35 and drop linearly to  $yF_{MSY}$  in year ‘50’ (see Supplementary Fig. 2a for three scenarios regarding  $x$  and  $y$ ). The data available for estimation purposes included an abundance index (fishery CPUE—so related to the selectivity pattern for the fishery, lognormal with  $CV = 0.2$ ) and catch age-composition data (Dirichlet with an effective sample size of 100). Catches were assumed to be available for all years but the data for assessment purposes (index and composition data; Section 4 of Appendix A) were only available from year ‘21’ (so that the age-composition data do not reflect an unfished stock). The selection of 50 % errors in  $M$  is illustrative but consistent with the levels of error common when empirical methods are used to set  $M$  ([Maunder et al., 2020](#)).

##### 4.2. Estimation method and management strategies

The estimation method was structured to be identical to the operating model, except that  $M$  was specified to be independent of age and time, and the model was fitted using a penalized likelihood framework in which the annual recruitment deviates were penalized to follow a normal distribution. Growth, weight- and fecundity-at-age were assumed known (Supplementary Fig. 1), as were the CV of CPUE, the effective sample size for the age-composition data (assumed to be multinomial) and the standard deviation of the random fluctuations in recruitment. The (potentially) estimable parameters were  $M$  (assumed to be age- and time-invariant; see below for further details on the scenarios regarding  $M$  in the estimation method), initial recruitment size ( $R_0$ ), the steepness of the stock-recruitment relationship, the annual fishing mortality rates, the deviations in recruitment about the stock-recruitment relationship, the parameters of the logistic selectivity function, and catchability. The initial values for the parameters of the estimation method were selected randomly from ranges about the true values. Two variants of the estimation method were considered, one in which  $M$  was pre-specified (equal to the true value; 50 % larger and

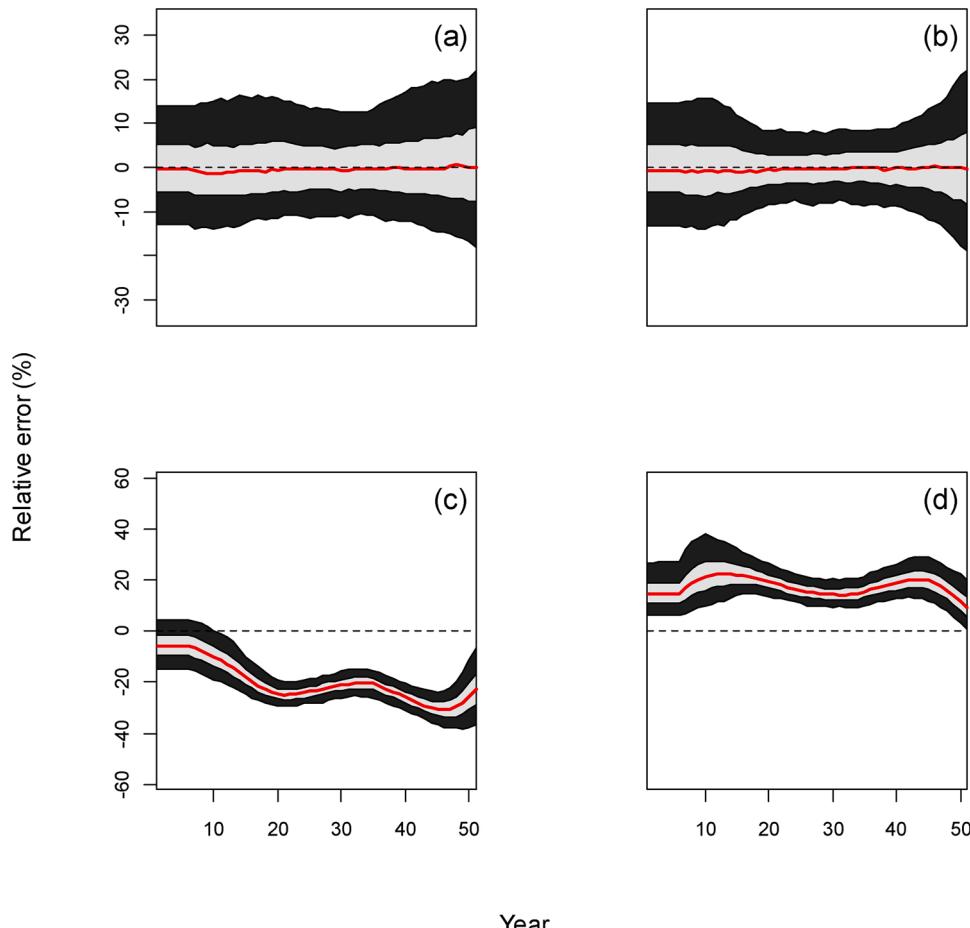


**Fig. 3.** Flowchart of the evaluation of management performance ('EM' denotes 'estimation method' and 'HCR' denotes 'harvest control rule').

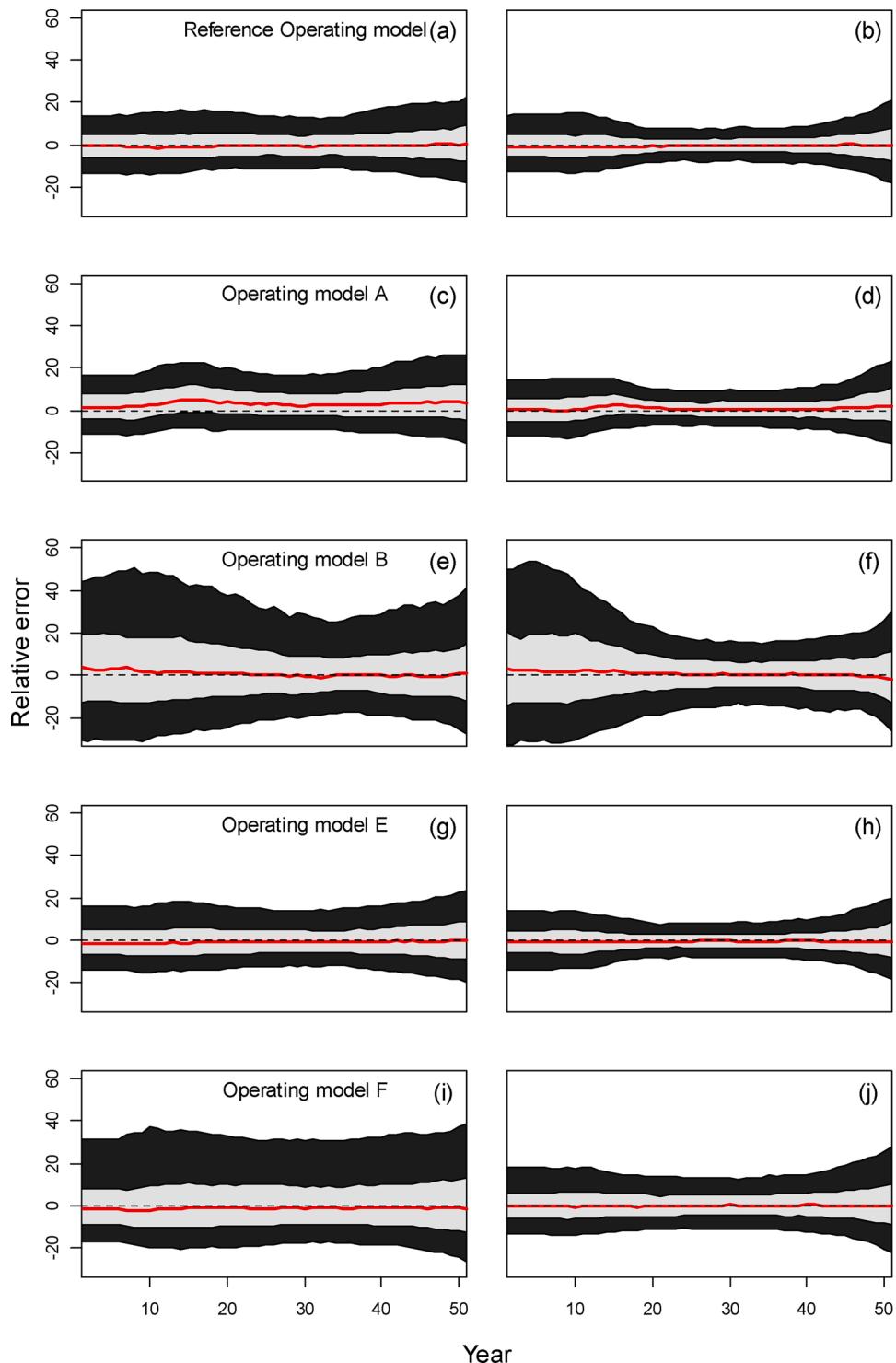
smaller than the true value) and one in which age- and time-invariant  $M$  was estimated.

The evaluation of management performance involved a 30-year projection period (nominally years '51'-'80'), with assessments and updates to catch limits conducted annually (see Fig. 3 for a flowchart of the feedback process). The harvest control rule was a threshold management strategy where fishing mortality was set to a target level above the threshold and declined linearly with biomass below the threshold level

to zero when the biomass was 20 % of the estimated unfished level (Supplementary Fig. 4). The two variants of the harvest control rule differ in terms of the target reference point ( $F_{40B}$ , the fishing mortality rate that reduces spawning biomass to 40 % of its unfished level, and  $F_{MSY}$ ).



**Fig. 4.** Distributions of relative error (i.e. (true-estimated values)/true value)) in spawning biomass (black 90 % percentiles, gray 50 % percentiles, red line median) for the four estimation methods ( $M$ -estimated, a;  $M = 0.2\text{yr}^{-1}$ , b;  $M = 0.1\text{yr}^{-1}$ , c;  $M = 0.3\text{yr}^{-1}$ , d) for the reference operating model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).



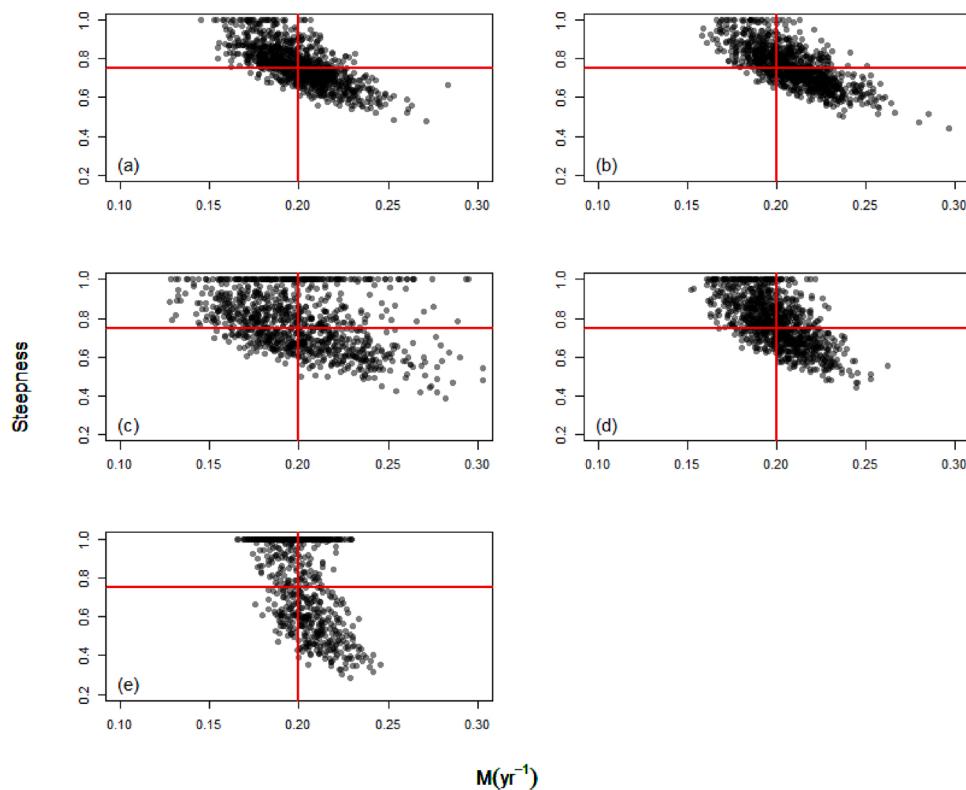
**Fig. 5.** Distributions of relative error in spawning biomass (black 90 % percentiles, gray 50 % percentiles, red line median) for  $M$ -estimated (left column) and  $M = 0.2 \text{yr}^{-1}$  (right column). The rows show results for the reference operating model (a,b), and operating models A, B, E and F (c-j). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

#### 4.3. Scenarios and performance metrics

The analyses were based 200 replicates, sufficient to detect the key major trends. The scenarios (Table 2) considered a reference scenario where most of the population dynamics and observation model assumptions underlying the management strategy were correct, age-dependence in  $M$  (the “true”  $M$  in this case is the weighted average  $M$  where the weights are the product of numbers-at-age and fecundity-at-

age in the unfished state), time-variation in  $M$ , and differences in the time-trajectory of historical fishing mortality (and hence population depletion at the start of the first year that the management strategy is applied; Supplementary Fig. 5).

Estimation performance was quantified in terms of relative error, i.e.  $(E-T)/T$ , where  $E$  is the estimated value of some quantity, and  $T$  is its true value. Estimation performance was evaluated based on applying the estimation method to the historical (years 1–50) data as well as each



**Fig. 6.** Estimates of steepness vs estimates of  $M$  for the reference operating model (a), and operating models A, B, E and F (b-e). The red lines represent the operating model values for each parameter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

fifth year during the projection period. The performance metrics used to measure management performance were:

- Probability of the spawning biomass at the end of the 30-year projection period (year '80') exceeding  $B_{MSY}$ .
- Probability that once during the 30-year projection period, the stock is rendered overfished (i.e., spawning biomass less than 0.5  $B_{MSY}$ ).
- Probability of overfishing each year (i.e., probability the annual fully-selected fishing mortality exceeds that associated with MSY).
- Median (over simulations) of the average catch over the first five years of the projection period.
- Median (over simulations) of the average catch over the last ten years of the projection period.

#### 4.4. Results: estimation of biomass (historical period)

Estimating  $M$  and setting  $M$  equal to  $0.2\text{yr}^{-1}$  (Figs. 4a,b) led to unbiased<sup>2</sup> estimates of spawning biomass (see Supplementary Fig. 6 for detailed results for the reference operating model). This is expected because the model on which the estimation method was based was correctly specified. Knowing that  $M = 0.2\text{yr}^{-1}$  led to less inter-simulation variability than estimating  $M$ , particularly between years '15' and '40'. Pre-specifying  $M$  to values other than  $0.2\text{yr}^{-1}$  led, as expected from previous studies, to bias (Figs. 4c,d). The level of bias depended on whether  $M$  was under-estimated (Fig. 4c) or over-estimated (Fig. 4d). These results were robust to replacing the stock-recruitment relationship by the Ricker form (Supplementary Fig. 7).

Fig. 5 (Supplementary Fig. 8 for the Ricker stock-recruitment relationship) summarizes the distributions of relative error for spawning stock biomass for the  $M$ -estimated and  $M = 0.2\text{yr}^{-1}$  estimation methods

for the reference operating model, the operating models with age- and time-varying natural mortality, and the operating models that have different historical time-trajectories of fishing mortality hence that of historical biomass (A, B, E and F). Allowing for age-dependence in  $M$  (Fig. 5c,d) led to some positive bias, which was larger when  $M$  was estimated. This is unsurprising because fishing mortality will remove the older animals and reduce the average value of  $M$  for the remaining mature animals. This led to  $M = 0.2\text{yr}^{-1}$  (slightly) exceeding true  $M$ . Time-variation in  $M$  led to unbiased estimates of spawning biomass, but there was considerably more inter-simulation variance (Figs. 5e,f).

The results were robust to reducing the maximum historical fishing mortality from 2 to  $1.25 F_{MSY}$  (compare Figs. 5h,g with Figs. 5a,b). However, reducing the maximum historical fishing mortality from 2 to  $0.5 F_{MSY}$ , led to substantially greater inter-simulation variation, particularly when  $M$  was estimated (Figs. 5i,j). This was probably a consequence of lesser contrast in the data (contrast Supplementary Figs. 5a and 5e).

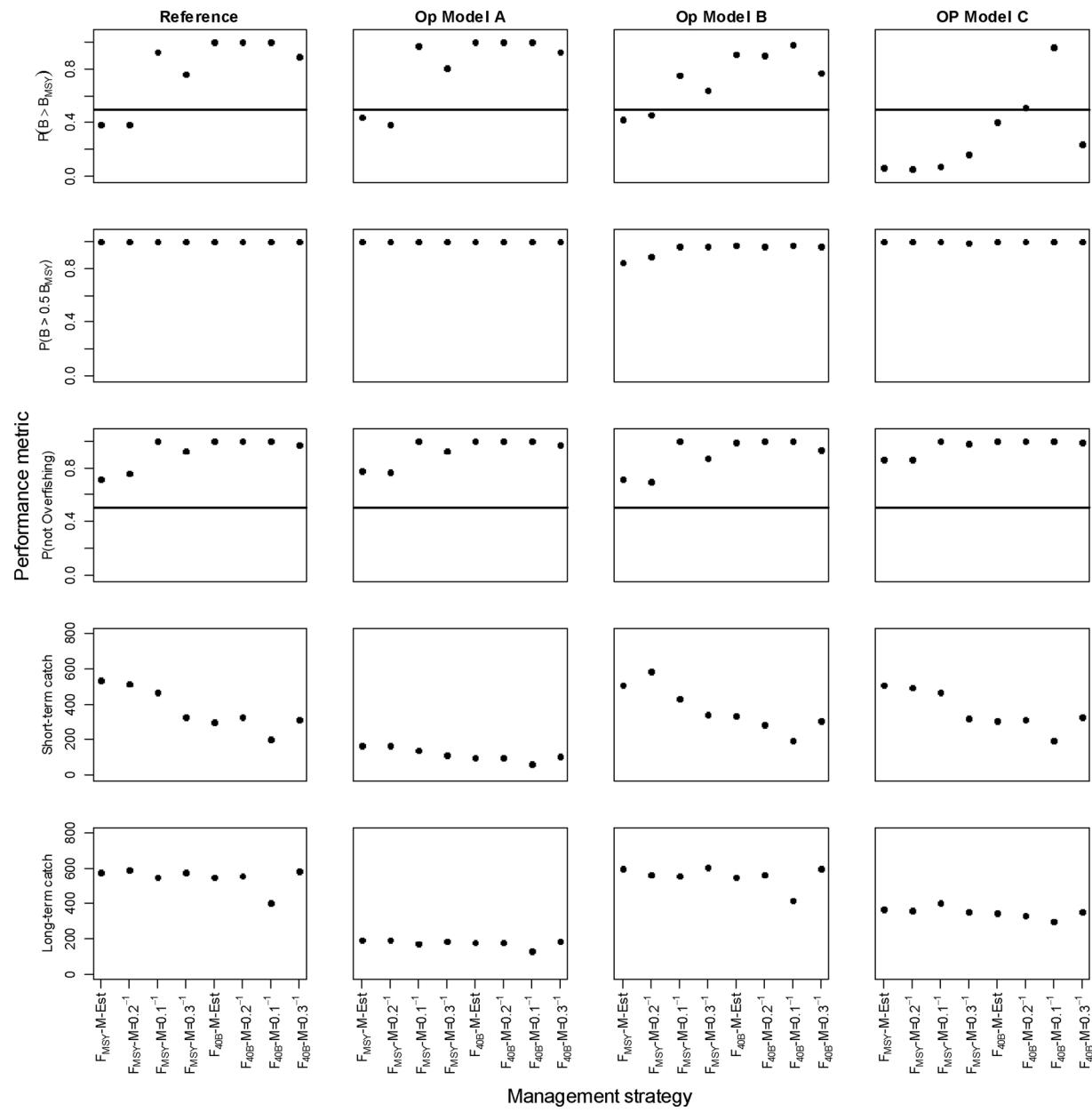
The estimates of steepness were negatively correlated with the estimate of  $M$  (Fig. 6). The extent of correlation was lower when the true value of  $M$  varied over time (Fig. 6c) and was smaller when the maximum historical fishing mortality was  $0.5 F_{MSY}$ . The proportion of estimates of steepness equal (or very close) to 1 was much higher for the  $0.5 F_{MSY}$  operating model than the other operating models, likely again due to less informative data series.

#### 4.5. Results: management performance

##### 4.5.1. Reference operating model

The first column of Fig. 7 shows the values for the five performance metrics for the reference operating model. Only two of the management strategies ( $F_{MSY}\cdot M\text{-Est}$  and  $F_{MSY}\cdot M = 0.2\text{yr}^{-1}$ ) led to a greater than 0.5 probability that the final spawning biomass was less than  $B_{MSY}$ . Over-estimating  $M$  led to a higher probability of being below  $B_{MSY}$  for this operating model. The management strategies based on a target fishing

<sup>2</sup> Taken here to be a difference between the median of the simulations and the true value.



**Fig. 7.** Performance metrics for the reference operating model and operating models A-F (columns).

mortality of  $F_{MSY}$  would be expected to have a probability of being below  $B_{MSY}$  of  $\sim 0.5$ , whereas a biomass target of  $0.4B_0$  should (and did) lead to an appreciably lower probability of  $B < B_{MSY}$  given  $F_{40B} < F_{MSY}$ . The only strategy that led to a negligible probability ( $< 0.01$ ) of the spawning biomass dropping below  $0.5B_{MSY}$  was for  $F_{MSY}\text{-}M\text{-Est}$ . The probability of overfishing was less than 0.5 for all management strategies, with  $F_{MSY}\text{-}M\text{-Est}$  leading to the highest probability of overfishing (0.287). As expected,  $F_{MSY}\text{-}M\text{-Est}$  and  $F_{MSY}\text{-}M = 0.2\text{yr}^{-1}$  led to the highest short-term catches, while  $F_{40B}\text{-}M = 0.1\text{yr}^{-1}$  led to the lowest short-term catch. The two  $M = 0.1\text{yr}^{-1}$  strategies performed quite differently depending on whether the target fishing mortality was  $F_{MSY}$  or  $F_{40B}$  in terms of short-term catch. In contrast to short-term catch, all of the management strategies with the exception of the  $F_{40B}\text{-}M = 0.1\text{yr}^{-1}$  achieved very similar long-term catches.

The estimates of spawning biomass from the  $M\text{-Est}$  and  $M = 0.2\text{yr}^{-1}$  estimation methods were unbiased and remained so over time, while the inter-simulation variability in the estimates of spawning biomass

became less biased as more data on year-class strengths accumulated (Fig. 8, upper panels). In contrast, the relative errors of spawning biomass for the  $M = 0.1\text{yr}^{-1}$  and  $M = 0.3\text{yr}^{-1}$  estimation methods illustrated bias (Fig. 8, lower panels). The relative errors for these estimation methods were also suggestive of retrospective patterns (not unexpected given that the assessment model was misspecified), with increasing negative bias over time for the  $M = 0.1\text{yr}^{-1}$  estimation method (Fig. 8c), and increasing positive bias over time for the  $M = 0.3\text{yr}^{-1}$  estimation method (Fig. 8d).

#### 4.5.2. Alternative operating models

Allowing  $M$  to be age-specific in the operating model but age-invariant in the estimation method (Fig. 7; operating model A) had little impact on the three risk measures (upper three rows of Fig. 7), and led to lower short- and long-term catches, suggesting that misspecifying average  $M$  is more important than misspecifying how  $M$  varies with age. The lower catches for the operating model with age-specific  $M$  can be

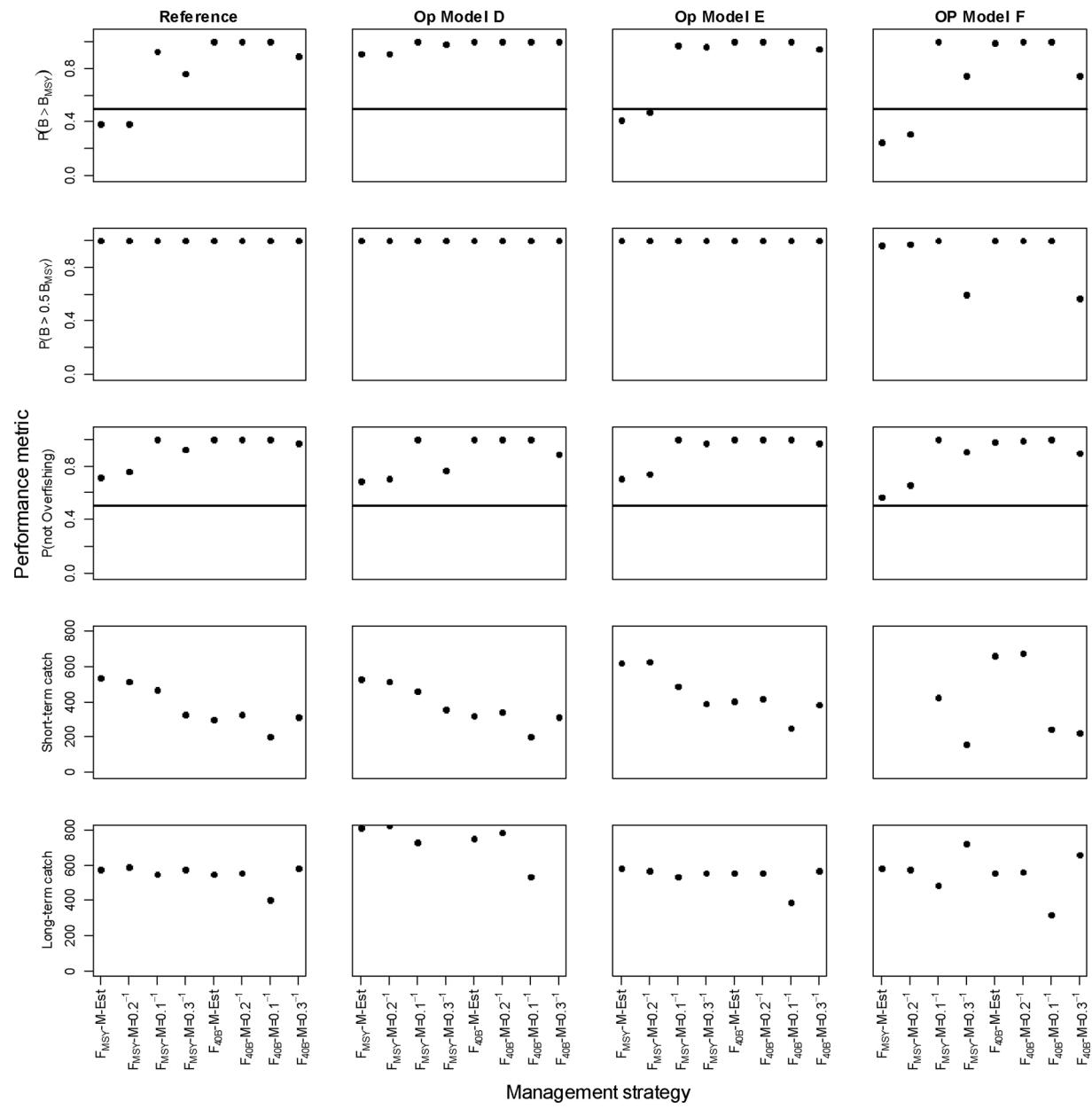


Fig. 7. (continued).

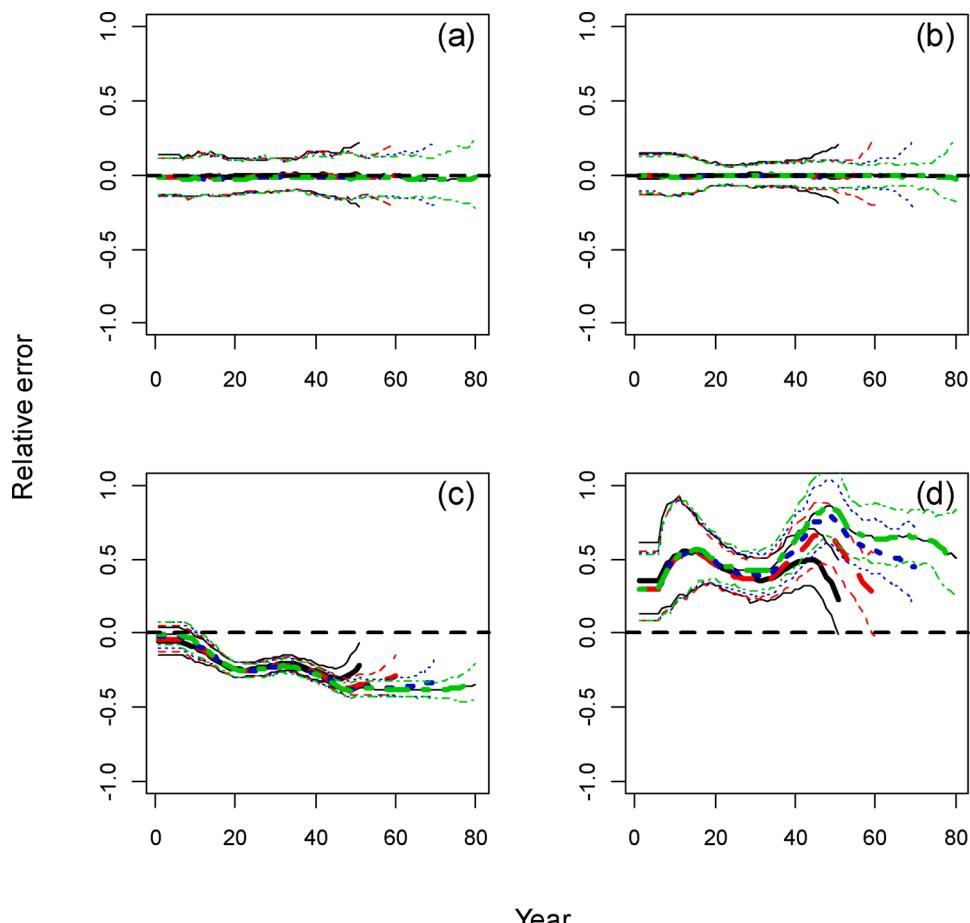
attributed to the slight difference in yield-per-recruit (and hence yield) between the two  $M$ -at-age profiles (Supplementary Fig. 9)

The probability of being below  $B_{MSY}$  and  $0.5B_{MSY}$  was higher when  $M$  was specified as being time-varying as a random walk (Fig. 7; operating model B), except for the probability of being below  $B_{MSY}$  for  $F_{MSY\text{-}M\text{-Est}}$  and  $F_{MSY\text{-}M} = 0.2\text{yr}^{-1}$ . The long- and short-term catches were similar between the reference operating model and operating model B, except for  $F_{MSY\text{-}M} = 0.2\text{yr}^{-1}$ , which led to higher short-term catches.

The impact of increasing and decreasing trends in  $M$  (operating models C and D respectively) had expected impacts. Risk was substantially higher for increasing  $M$  and lower for decreasing  $M$ . The only management strategy that had a less than 0.5 probability of overfishing when  $M$  was increasing is  $F_{40B\text{-}M} = 0.1\text{yr}^{-1}$ . The time-trajectory of spawning biomass for operating model C increased initially due to catches that were lower than replacement yield, but spawning biomass declined thereafter even given catches that declined after 10 years into the projection period (year '60') (Fig. 9; see Supplementary Fig. 10 for results for the reference operating model), and was just below the

original  $B_{MSY}$  at the end of the projection period. Given this, results show there were relatively small differences among operating models C and D for a given management strategy in terms of short-term catch, but substantially lower long-term catches for operating model C (although the reduction in long-term catch was less for  $F_{40B\text{-}M} = 0.1\text{yr}^{-1}$  than for the remaining management strategies). The opposite effect was evident for operating model D, with substantially higher long-term catches for all but management strategy  $F_{40B\text{-}M} = 0.1\text{yr}^{-1}$ . Unlike the situation for the reference operating model, all of the estimation methods exhibited time-varying bias, with the pattern in bias was a function of the change in  $M$  over time (e.g., compare the patterns of relative error for the  $M = 0.1\text{yr}^{-1}$  estimation method for operating model D with the pattern for the reference operating model; Fig. 10).

The probabilities of spawning biomass below  $B_{MSY}$  and  $0.4B_{MSY}$  were not directly comparable between the reference operating model and the operating models with less contrast in biomass due to less intensive fishing (operating models E and F) because spawning biomass started at higher proportions of  $B_{MSY}$  in year 50 for operating models E and F (see



**Fig. 8.** Distributions of relative error for the four estimation methods (a:  $M$ -Est; b:  $M = 0.2\text{yr}^{-1}$ ; c:  $M = 0.1\text{yr}^{-1}$ ; and d:  $M = 0.3\text{yr}^{-1}$ ) for the reference operating model. The solid lines denote the medians and the light lines denote 90 % intervals. The colors denote the relative errors for the assessments in the first year of the projection period and those conducted every 5 years thereafter.

Supplementary Fig. 11 for operating model F). Nevertheless, the probability of spawning biomass being below  $B_{\text{MSY}}$  and  $0.5B_{\text{MSY}}$  was appreciably higher for operating model F for management strategies  $F_{\text{MSY}}\text{-}M\text{-Est}$ ,  $F_{\text{MSY}}\text{-}M = 0.3\text{yr}^{-1}$ , and  $F_{40\text{B}}\text{-}M = 0.3\text{yr}^{-1}$ , with the probability of spawning biomass below  $0.5B_{\text{MSY}}$  almost 0.5 for  $F_{\text{MSY}}\text{-}M = 0.3\text{yr}^{-1}$  and  $F_{40\text{B}}\text{-}M = 0.3\text{yr}^{-1}$ .

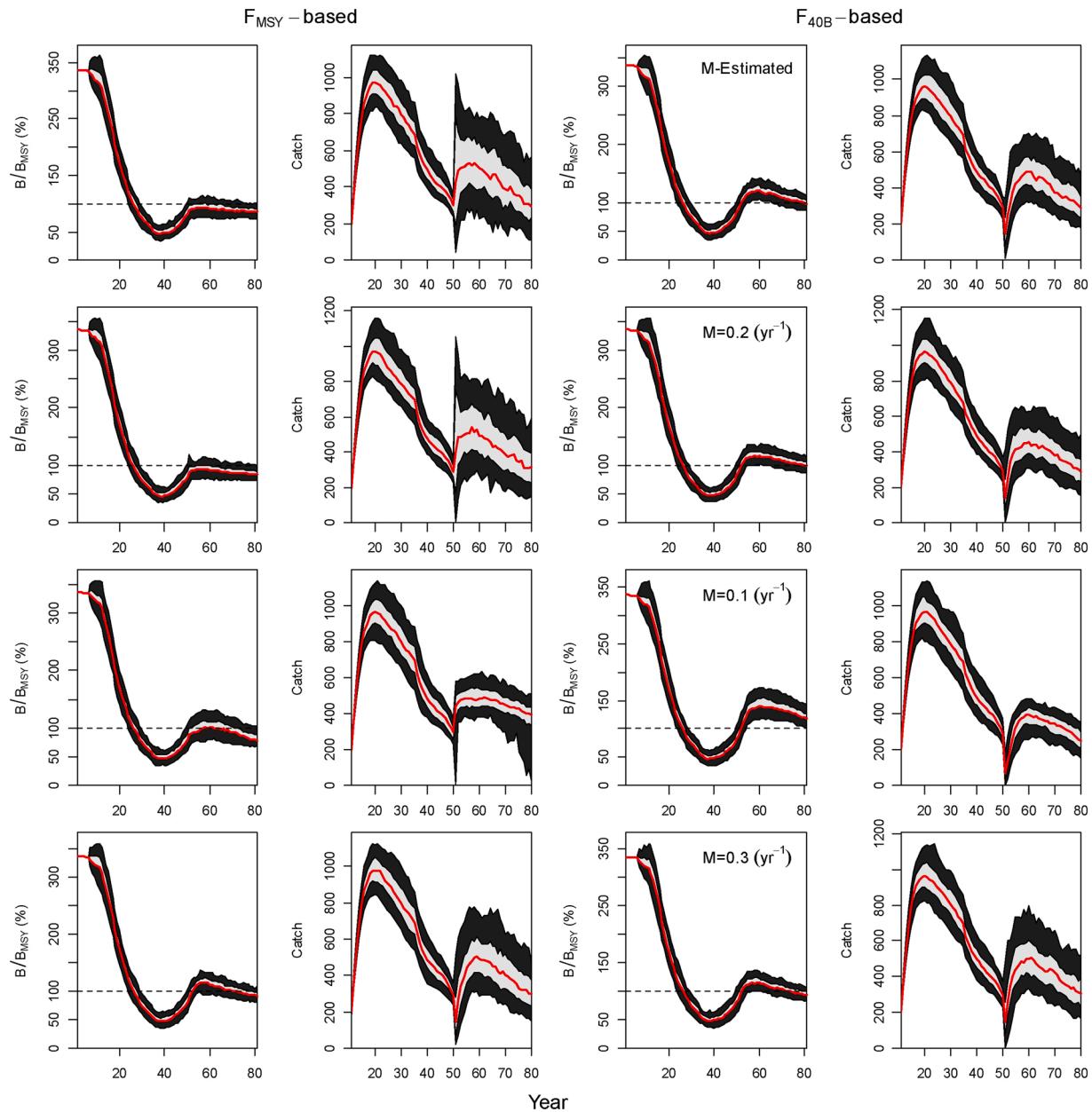
## 5. Discussion and conclusions

One result from our simulations suggested that biased estimation methods perform well within a management strategy context. Since estimating the stock-recruit parameter “steepness” was part of the management strategy, an interaction apparently compensated to some degree with misspecified  $M$  (the form and shape of the stock-recruitment relation is notoriously difficult to estimate; Conn et al., 2010; Lee et al., 2012). Fig. 11 compares results for the reference operating model of the management strategies in Figs. 7–10 and management strategies that set steepness equal to 0.65, 0.75 and 0.85. The impact of pre-specified steepness on the conservation-related performance metrics was marked, irrespective of whether  $M$  was estimated or pre-specified, with the impact greater for larger pre-specified values for steepness (reflecting a higher perceived productivity). Pre-specifying steepness eliminated the negative correlation between the estimates of  $M$  and steepness (e.g., Fig. 6), which would otherwise mitigate the impact of an erroneous value for  $M$ . The performances of  $F_{\text{MSY}}\text{-}M\text{-Est}$ ,  $F_{\text{MSY}}\text{-}M = 0.2\text{yr}^{-1}$ ,  $F_{40\text{B}}\text{-}M = 0.3\text{yr}^{-1}$ , and (particularly)  $F_{\text{MSY}}\text{-}M = 0.3\text{yr}^{-1}$  changed the most compared to the reference operating model, with almost all simulations for  $F_{\text{MSY}}\text{-}M = 0.3\text{yr}^{-1}$  and well over half of the simulations for

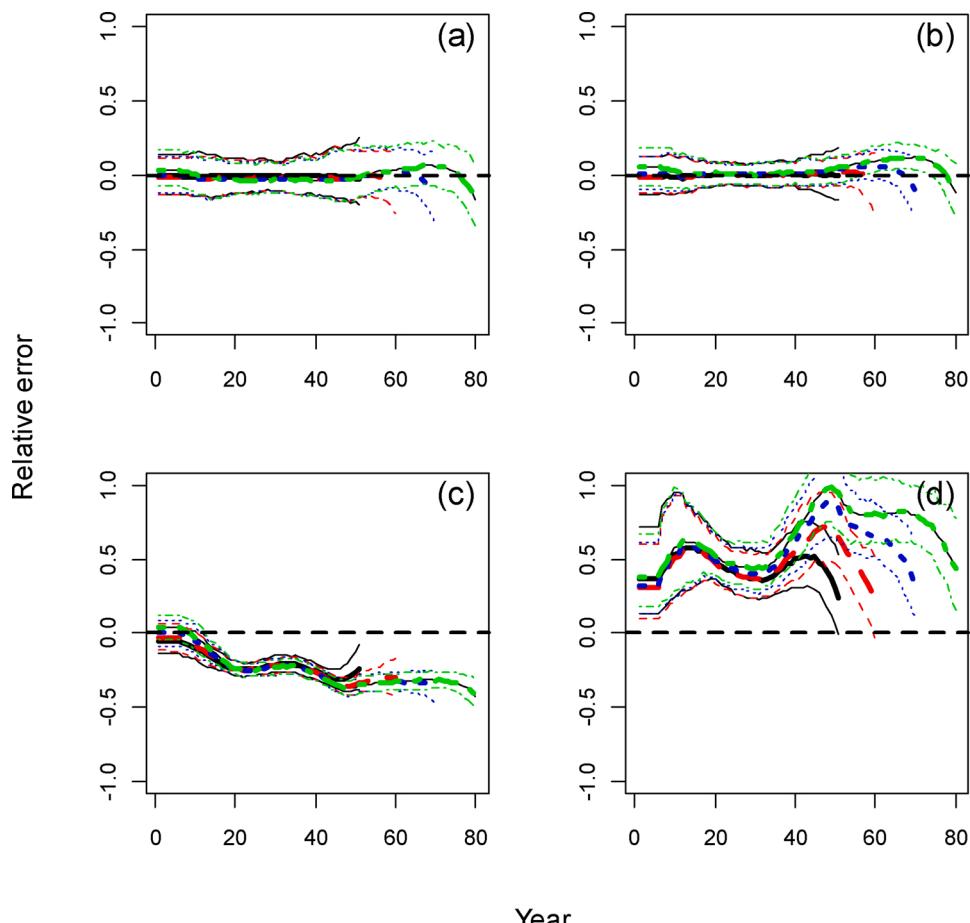
$F_{\text{MSY}}\text{-}M\text{-Est}$  and  $F_{\text{MSY}}\text{-}M = 0.2\text{yr}^{-1}$  for a fixed steepness of 0.75 and 0.85 leading to a final spawning biomass less than  $B_{\text{MSY}}$ . The probability of overfishing was ~0.5 or higher for  $F_{\text{MSY}}\text{-}M\text{-Est}$ ,  $F_{\text{MSY}}\text{-}M = 0.2\text{yr}^{-1}$  and  $F_{\text{MSY}}\text{-}M = 0.3\text{yr}^{-1}$  when steepness was pre-specified at 0.85. The patterns of relative error changed when steepness was pre-specified rather than being estimated (Supplementary Fig. 12). In particular, there was no evidence of reduced bias over time for the  $M = 0.3\text{yr}^{-1}$  estimation method (Supplementary Fig. 12d).

Natural mortality (even if assumed to be independent of age, sex, location, and time) remains hard to estimate even given all the advances in data collection methods and analytical techniques. Traditionally,  $M$  was either set based on *a priori* considerations (e.g., the conventional  $M = 0.2\text{yr}^{-1}$  assumption) or pre-specified based on empirical methods, catch-curve analysis or (much less frequently) tagging methods. These approaches all make assumptions that will be violated to some extent, with the consequence that the value of  $M$  used in stock assessments will be incorrect because the method used to estimate it is biased and imprecise. The level of bias and imprecision depends on the method of estimation. For example, Pascual and Iribane (1993) estimated prediction errors to be 10–36 % and Kenchington (2014) found such errors to be even larger (50–200 %). Hamel (2015) computed prediction errors when the results of multiple estimation methods were combined and found these to be large in general.

Although all of the methods for estimating  $M$  are likely biased and imprecise, we prefer the ‘direct’ methods based on catch-curve analysis and tagging data, with a preference for the latter. These preferred methods are not without their drawbacks such as assumptions about selectivity for catch-curve analysis and appropriate sampling design and



**Fig. 9.** Distributions (black 90 % percentiles, gray 50 % percentiles, red line median) for spawning biomass relative to  $B_{MSY}$  and catch for the eight management strategies for operating model C. The rows show results for the four estimation methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).



**Fig. 10.** Distributions of relative error for the four estimation methods (a:  $M$ -Est; b:  $M = 0.2 \text{ yr}^{-1}$ ; c:  $M = 0.1 \text{ yr}^{-1}$ ; and d:  $M = 0.3 \text{ yr}^{-1}$ ) for operating model D. The solid lines denote the medians and the light lines denote 90 % intervals. The colors denote the relative errors for the assessments in the first year of the projection period and those conducted every 5 years thereafter.

sample size for the tagging studies. It should be noted that only the empirical approaches may be viable in data-poor situations and we advocate providing assessment results and management advice for a range of these approaches in this situation (e.g., based on approach similar to that of Hamel, 2015; see also Cope, this volume).

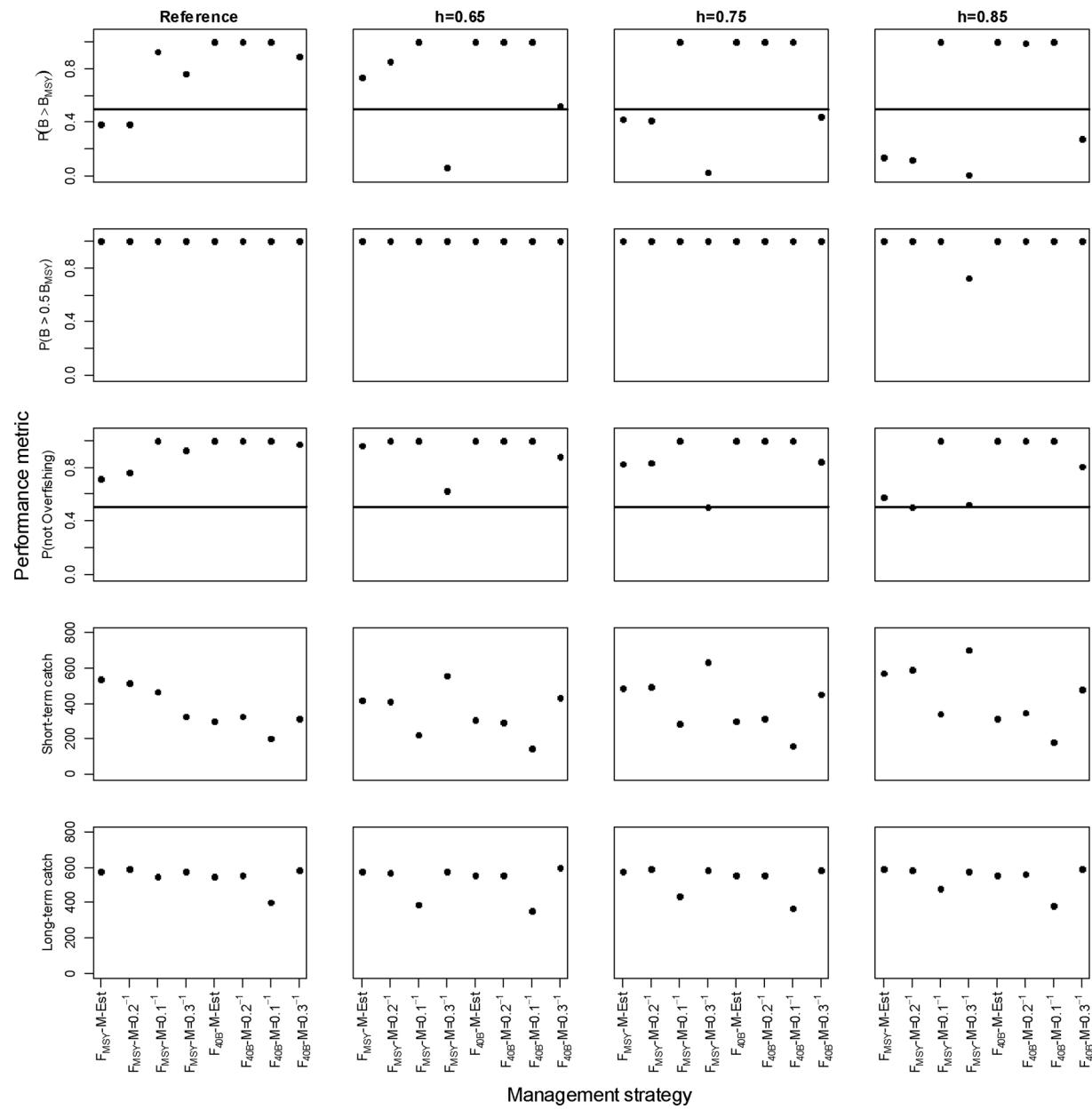
Assuming the incorrect value of  $M$  when conducting assessments impacts almost all of the quantities on which management advice is based. Specifically, the value to be assumed for the target fishing mortality rate will be positively biased if  $M$  is over-estimated, irrespective of the policy used to select the basis for the target reference point (Fig. 2). This study and several previous studies (see Table 1) have quantified the level of bias to be expected in estimates of spawning biomass and recruitment if  $M$  is misspecified. The numerical results of these types of studies depend on the range and extent of error in  $M$  (e.g., on average, due to time variation, and/or due to age/size/sex dependence), but several general patterns emerge. Specifically, estimates of spawning biomass will be positively biased when  $M$  is overestimated and *vice versa*. Whether the level of bias is robust to the time-trajectory of fishing mortality remains unclear but likely depends on life history characteristics.

Few studies have explored the impact of errors in  $M$  on the performance of management strategies (compare the number of rows in Table 1 with the studies outlined in Section 3.3), and the results are more case-specific than for estimation of biomass. The results of this paper (Section 4), which pertain to a data-rich case with no other sources of model misspecification, lead to the conclusion that feedback management can alleviate some, but not all, of the problems caused by misspecified  $M$ . However, this was only the case when a second

parameter (in this case steepness) as estimated instead of  $M$  (Fig. 11). It is seldom the case that both  $M$  and stock-recruitment steepness are estimated in stock assessments, and if they are estimated together, informative priors are usually placed on both parameters. The results of this paper support this practice.

There is much that remains to be learned about consequences of errors in  $M$  and the lack of ability to estimate it. In particular, there is notable lack of studies that have explored the impact of errors in  $M$  within the context of spatially- and size-structured models.

Overall, our general conclusion is that misspecified  $M$  leads to a decreased ability to achieve management goals. The results of the simulation study of this paper support estimating  $M$  within the assessment when possible (a trend evident in an increasing number of stock assessments; Castillo-Jordán et al., this volume). If a prior (or penalty in the context of penalized maximum likelihood) is placed on  $M$ , the estimate of  $M$  will equal the mean of the prior if the data and other model configuration characteristics are uninformative, and will be updated if there are data in the assessment that inform  $M$ . Thus, in principle at least, if the model is not misspecified (and this case can be evaluated to some extent using the  $R_0$  likelihood profile method of Wang et al., 2014), estimating  $M$  will behave like pre-specifying  $M$  when data are inadequate to estimate it but the available data will be used to estimate  $M$  when this is possible. One concern with this approach is that the estimate of  $M$  may be compensating for the effect of misspecification of a biological process other than natural mortality (a feature confirmed in principle by Szuwalski et al. [2018]). Estimating  $M$  when there are insufficient data to update the prior (or penalty) has the advantage that the uncertainty in  $M$  will be propagated into the estimated uncertainty



**Fig. 11.** Performance metrics for the reference operating model and operating models in which steepness is pre-specified.

**Table A1**  
Values for the parameters of the reference operating model.

Parameter	Value
Growth parameters	
Asymptotic size, $L_\infty$	100 cm
Growth parameter, $\kappa$	0.2 yr <sup>-1</sup>
Theoretical age at which length is zero, $t_0$	-0.1 yr <sup>-1</sup>
Length-weight $\alpha$	0.0001
Length-weight $\beta$	3
Selectivity parameters	
Age-at-50 % selectivity, $a_{50}$	5 years
Selectivity width, $\Delta$	4 year
Recruitment-related	
Unfished recruitment, $R_0$	1000
Steepness, $h$	0.75
Extent of recruitment variation, $\sigma_R$	0.6

of other model outputs. Finally, if the estimated  $M$  is unrealistic, this indicates that there is serious model misspecification (indicated by a precise estimate of  $M$ ) or correlated parameters (indicated by an imprecise estimate of  $M$ ).

#### Author credit statement

AP developed the concept of the paper; AP created the initial draft and structure; All authors contributed to writing and editing, and the experimental design of the simulations.

#### Declaration of Competing Interest

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

## Acknowledgements

Ian Taylor and Jim Hastie (NWFSC), M.S.M. Siddeek (ADF&G), and an anonymous reviewer are thanked for their comments on an earlier

version of this paper. This publication was [partially] funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement NA15OAR4320063, Contribution No. 2020-1103.

## Appendix A. The simulation model used to evaluate bias for per-recruit analysis and as the operating model for the MSE

### A Basic dynamics

The basic dynamics are governed by a single-sex age-structured population dynamics model:

$$N_{y+1,a} = \begin{cases} R_y & \text{if } a = 0 \\ N_{y,a-1} e^{-Z_{y,a-1}} & \text{if } 1 \leq a < x \\ N_{y,x-1} e^{-Z_{y,x-1}} + N_{y,x} e^{-Z_{y,x}} & \text{if } a = x \end{cases} \quad (\text{A.1})$$

where  $N_{y,a}$  is the number of animals of age  $a$  at the start of year  $y$ ,  $x$  is the plus-group age (age 20),  $Z_{y,a}$  is the total mortality for animals of age  $a$  during year  $y$ :

$$Z_{y,a} = M_{y,a} + F_{y,a} \quad (\text{A.2})$$

$M_{y,a}$  is the rate of natural mortality for animals of age  $a$  during year  $y$ ,  $F_{y,a}$  is the rate of fishing mortality for animals of age  $a$  during year  $y$ :

$$F_{y,a} = S_a F_y \quad (\text{A.3})$$

$S_a$  is fishery selectivity for animals of age  $a$ :

$$S_a = (1 + e^{-\ell n^{19}(a-a_{50})/\Delta})^{-1} \quad (\text{A.4})$$

$a_{50}$  is the age-50 %-selectivity,  $\Delta$  determines the width of the age-specific selectivity pattern,  $F_y$  is the fully-selected fishing mortality during year  $y$ ,  $R_y$  is recruitment during year  $y$ :

$$R_y = \frac{4hR_0(S_y/S_0)}{(1-h)+(5h-1)(S_y/S_0)} e^{\varepsilon_y - \sigma_R^2/2} \quad \varepsilon_y \sim N(0; \sigma_R^2) \quad (\text{A.5})$$

$R_0$  is the unfished recruitment,  $S_y$  is the spawning biomass during year  $y$ :

$$S_y = \sum_{a=1}^x f_a N_{y,a} \quad (\text{A.6})$$

$f_a$  is fecundity for animals of age  $a$ :

$$f_a = \begin{cases} 0 & \text{if } a < a_{\text{mat}} \\ w_a & \text{otherwise} \end{cases} \quad (\text{A.7})$$

$a_m$  is the age-at-maturity,  $w_a$  is the weight of an animal of age  $a$ :

$$w_a = \alpha(L_\infty(1 - e^{-\kappa(a-t_0)}))^\beta \quad (\text{A.8})$$

$L_\infty$  is the asymptotic size,  $\kappa$  is the growth parameter,  $t_0$  is the theoretical age at which length is zero,  $\alpha$  and  $\beta$  are the parameters of the length-weight relationship,  $h$  is the steepness of the stock-recruitment relationship, and  $\sigma_R$  is the standard deviation of the deviations about the stock-recruitment relationship.

The population has an unfished age-structured at the start of the first year (-9).

### A.2 Catches

The catches (in weight) for year  $y$ ,  $\tilde{C}_y$ , are given by:

$$\tilde{C}_y = \sum_a w_{a+1/2} C_{y,a} = \sum_a w_{a+1/2} \frac{F_{y,a}}{Z_{y,a}} N_{y,a} (1 - e^{-Z_{y,a}}) \quad (\text{A.9})$$

### A.3 Parameterization

Weight- and fecundity-at-age are plotted in Supplementary Fig. 1. The reference operating model assumes that natural mortality is a constant, independent of age and time, and equal to  $0.2\text{yr}^{-1}$ . Age-specific natural mortality is modelled using the [Siler \(1979\)](#) equation:

$$M_a = M_1 e^{-M_2 a} + M + M_4 e^{-M_3 a/x} \quad (\text{A.10})$$

where  $M_1 = 0.5$ ,  $M_2 = 0.4$ ,  $M_3 = 2$ , and  $M_4 = 0.2$ , with the parameter values selected so that the weighted (by numbers multiplied by fecundity) average value of  $M$  is  $0.2\text{yr}^{-1}$ . Future time trends in  $M$  are modelled according to:

- Monotonic change in  $M$ ,  $M_y = M e^{\delta(y-40)}$  where  $\delta$  is such that  $M$  is either  $0.25\text{yr}^{-1}$  or  $0.15\text{yr}^{-1}$  at the end of the projection period.
- Random change in  $M$ , i.e.  $\ell n M_{y+1} = \rho_M \ell n M_y + \sqrt{1 + \rho_M^2} \eta_{M,y}$ ;  $\eta_{M,y} \sim N(0; \sigma_M^2)$ , where  $\rho_M = 1/\sqrt{2}$  and  $\sigma_M = 0.2$ .

Supplementary Fig. 2b plots the assumed values for selectivity-at-age, fecundity-at-age, and weight-at-age. The reference values for  $x$  and  $y$  (section 3.1) are set to 2 and 0.5 respectively. The population is assumed to be in unfished equilibrium 10 years before the first fishery catches, with recruitment variation introduced so that the population is not in equilibrium at the start of the first year with catches.

#### A.4 Data generation

The data available for assessment purposes are the annual catches (measured without error,  $\tilde{C}_y$ ), an index of abundance:

$$I_y = q \sum_a w_{a+1/2} \frac{S_a}{Z_{y,a}} N_{y,a} (1 - e^{-Z_{y,a}}) \quad (\text{A.11})$$

where  $q$  is catchability, and a catch proportion at age data set with Dirichlet distributed errors. The CV of the CPUE data is assumed to be 0.2 while the effective sample size for the age-composition data is 100 (Table A1).

#### Appendix B. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.fishres.2020.105759>.

#### References

- Aanes, S., Engen, S., Sather, B.-E., Annes, R., 2007. Estimation of the parameters of fish stock dynamics from catch-at-age data and indices of abundance: can natural and fishing mortality be separated? *Can. J. Fish. Aquat. Sci.* 64, 1130–1142.
- Aksland, M., 1994. A general cohort analysis method. *Biometrics* 50, 917–932.
- Albertsen, C.M., Nielsen, A., Thygesen, U.H., 2017. Connecting single-stock assessment models through correlated survival. *ICES J. Mar. Sci.* 75, 235–244.
- Begley, J., 2014. Gadget User Guide. Available at <http://www.hafro.is/gadget/userguide/userguide.html>.
- Berg, C., Nielsen, A., 2016. Accounting for correlated observations in an age-based state-space stock assessment model. *ICES J. Mar. Sci.* 73, 1788–1797.
- Beverton, R.J.H., Holt, S.J., 1957. On the dynamics of exploited fish populations. U.K. Minist. Agric. Fish. Fish. Invest. Ser. 2, 19.
- 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.
- Brownie, C., Anderson, D.R., Burnham, K.P., Robson, D.S., 1985. Statistical Inference From Band-recovery Data—a Handbook, 2nd edition. U.S. Fish and Wildlife Service Resource Publication, p. 156.
- Bull, B., Dunn, A., McKenzie, A., Gilbert, D.J., Smith, M.H., Bian, R., Fu, D., 2012. CASAL (C++ Algorithmic Stock Assessment Laboratory) User Manual v2.30-2012/03/21. NIWA Technical Report 135. National Institute of Water and Atmospheric Research, Wellington, 280 p.
- Burnefeld, N., Hoshino, E., Milner-Gulland, E.J., 2011. Management strategy evaluation: a powerful tool for conservation? *Trends Ecol. Evol. (Amst.)* 26, 441–447.
- Butterworth, D.S., Punt, A.E., 1990. Some preliminary examinations of the potential information context of age-structure data from Antarctic minke whale research catches. *Rep. int. Whal. Comm.* 40, 301–315.
- Butterworth, D.S., Punt, A.E., 1999. Experiences in the evaluation and implementation of management procedures. *ICES J. Mar. Sci.* 56, 985–998.
- Carruthers, T.R., Punt, A.E., Walters, C.J., MacCall, A., McAllister, M.K., Dick, E.J., Cope, J., 2014. Evaluating methods for setting catch limits in data-limited fisheries. *Fish. Res.* 153, 48–68.
- Castillo-Jordán, C., Johnson, K.F., Hamel, O.S., Doreing, K.L., Taylor, I.G., Monnahan, C.C., Ianelli, J.N., Punt, A.E., 2020. This issue. How high and low steepness of the stock-recruitment relationship changes estimates of natural mortality and management quantities. *Fish. Res.* 00, 00–00.
- Chapman, D.G., Robson, D.S., 1960. The analysis of a catch curve. *Biometrics* 16, 354–368.
- Chen, S., Watanabe, S., 1989. Age dependence of natural mortality coefficient in fish population dynamics. *Nippon Suisan Gakkaishi* 55, 205–208.
- Clark, W.G., 1999. Effects of an erroneous natural mortality rate on a simple age-structured stock assessment. *Can. J. Fish. Aquat. Sci.* 56, 1721–1731.
- Conn, P.B., Williams, E.H., Shertzer, K.W., 2010. When can we reliably estimate the productivity of fish stocks? *Can. J. Fish. Aquat. Sci.* 67, 511–523.
- Cook, R.M., 2004. Estimation of the age-specific rate of natural mortality for Shetland sandeels. *ICES J. Mar. Sci.* 61, 159–164.
- Cope, J.M., 2020. This issue. Upgrading from M 0.2: an application-based method for accessible estimation, evaluation and uncertainty characterization of natural mortality. *Fish. Res.* 00, 00–00.
- Cope, J.M., Sampson, D., Stephens, A., Key, M., Mirick, P.P., Stachura, M., Tsou, T., Weyland, P., Berger, A., Buell, T., Councill, E., Dick, E.J., Fenske, K.H., Monk, M., Rodomsky, B.T., 2016. Assessments of California, Oregon and Washington Stocks of Black Rockfish (*Sebastodes melanops*) in 2015. Available from: <https://www.pcouncil.org/stock-assessments-star-reports-stat-reports-rebuilding-analyses-terms-of-reference/ce/groundfish-stock-assessment-documents/>.
- Curti, K.L., Collie, J.S., Legault, C.M., Link, J.S., 2013. Evaluating the performance of a multispecies statistical catch-at-age model. *Can. J. Fish. Aquat. Sci.* 70, 470–484.
- Deriso, R.B., 1980. Harvesting strategies and parameter estimation for an age-structured model. *Can. J. Fish. Aquat. Sci.* 37, 268–282.
- Deriso, R.B., Quinn II, T.J., Neal, P.R., 1985. Catch-age analysis with auxiliary information. *Can. J. Fish. Aquat. Sci.* 42, 815–824.
- Deroba, J.J., Schueller, A.M., 2013. Performance of stock assessments with misspecified age- and time-varying natural mortality. *Fish. Res.* 146, 27–40.
- Dick, E.J., He, X., 2019. Status of Cowcod (*Sebastodes levis*) in 2019. Available from: Pacific Fishery Management Council, Portland, OR. <https://www.pcouncil.org/stock-assessments-star-reports-stat-reports-rebuilding-analyses-terms-of-reference/groundfish-stock-assessment-documents/>.
- Dunn, A., Francis, R.I.C.C., Doonan, I.J., 2002. Comparison of the Chapman-Robson and regression estimators of Z from catch-curve data when non-sampling stochastic error is present. *Fish. Res.* 59, 149–159.
- Forrest, R.E., Holt, K.R., Kronlund, R., 2018. Performance of alternative harvest control rules for two Pacific groundfish stocks with uncertain natural mortality: bias, robustness and trade-offs. *Fish. Res.* 206, 259–286.
- Fournier, D., 1983. An analysis of the Hecate Strait Pacific cod fishery using an age-structured model incorporating density-dependent effects. *Can. J. Fish. Aquat. Sci.* 40, 1233–1243.
- Fournier, D., Archibald, C.P., 1982. A general theory for analyzing catch at age data. *Can. J. Fish. Aquat. Sci.* 39, 1195–1207.
- Fournier, D.A., Hampton, J., Sibert, J.R., 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. *Can. J. Fish. Aquat. Sci.* 55, 2105–2116.
- Fu, C., Quinn, T.J.I., 2000. Estimability of natural mortality and other population parameters in a length-based model: *pandalus borealis* in Kachemak Bay, Alaska. *Can. J. Fish. Aquat. Sci.* 57, 2420–2432.
- Gislason, H., Helgason, Th., 1985. Species interaction in assessment of fish stocks with special application to the North Sea. *Dana* 5, 1–44.
- Goethel, D.R., Quinn II, T.J., Cadrian, S.X., 2011. Incorporating spatial structure in stock assessment: movement modeling in marine fish population dynamics. *Rev. Fish. Sci. Aquac.* 19, 119–136.
- Gunderson, D.R., 1980. Using r-K selection theory to predict mortality. *Can. J. Fish. Aquat. Sci.* 37, 2266–2271.
- Gunderson, D.R., Dygert, P.H., 1988. Reproductive effort as a predictor of natural mortality rate. *J. Cons. Int. Explor. Mer.* 44, 200–209.
- Haltuch, M.A., Ono, K., Valero, J., 2013. Status of the U.S. Petrale Sole Resource in 2012. Available from: Pacific Fishery Management Council, Portland, OR. <https://www.pcouncil.org/stock-assessments-star-reports-stat-reports-rebuilding-analyses-terms-of-reference/ce/groundfish-stock-assessment-documents/>.

- [pcouncil.org/stock-assessments-star-reports-stat-reports-rebuilding-analyses-terms-of-reference/groundfish-stock-assessment-documents/](http://pcouncil.org/stock-assessments-star-reports-stat-reports-rebuilding-analyses-terms-of-reference/groundfish-stock-assessment-documents/).
- Hamel, O.S., 2015. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. Available from ICES J. Mar. Sci. 72, 62–69. <https://www.pcouncil.org/stock-assessments-star-reports-stat-reports-rebuilding-analyses-terms-of-reference/groundfish-stock-assessment-documents/>.
- Hamel, O.S., Cope, J.M., Matson, S., 2013. Stock Assessment of Aurora Rockfish in 2013. Available from Pacific Fishery Management Council, Portland, OR. <http://www.pcoinc.org/groundfish/stock-assessments/>.
- Hampton, J., Fournier, D.A., 2001. Aspatially disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. Mar. Freshw. Res. 52, 937–963.
- He, X., Ralston, S., MacCall, A.D., 2011. Interactions of age-dependent mortality and selectivity functions in age-based stock assessment models. Fish. Bull. US. 109, 198–216.
- Hilborn, R., 1990. Determination of fish movement patterns from tag recoveries using maximum likelihood estimators. Can. J. Fish. Aquat. Sci. 47, 635–643.
- Hilden, M., 1988. Errors in perception in stock and recruitment studies due to wrong choices of natural mortality rate in virtual population analysis. J. Cons. Int. Explor. Mer. 44, 123–134.
- Hoening, J.M., 1983. Empirical use of longevity data to estimate mortality rates. Fish. Bull. US. 82, 898–903.
- Holling, C.S., 1959. The components of predation as revealed by a study of small-mammal predation of the European pine sawfly. Can. Entomol. 91, 293–320.
- Hollowed, A.B., Ianelli, J.N., Livingston, P.A., 2000. Including predation mortality in stock assessments: a case study for Gulf of Alaska walleye pollock. ICES J. Mar. Sci. 57, 279–293.
- Holsman, K.K., Ianelli, J., Aydin, K., Punt, A.E., Moffit, E.A., 2016. Comparative biological reference points estimated from temperature-specific multispecies and single species stock assessment models. Deep Sea Res. II 134, 360–378.
- Hordyk, A.R., Huynh, Q.C., Carruthers, T.R., 2019. Misspecification in stock assessments: common uncertainties and asymmetric risks. Fish. Fish. Oxf. (Oxf) 20, 888–902.
- Hurtado Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R.R., McGillaird, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-pre, K.A., Whitten, A.R., Punt, A.E., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES J. Mar. Sci. 72, 99–110.
- Ianelli, J.N., Barbeaux, S.J., McKelvey, D., Honkalehto, T., 2015. 1.B. Assessment of Walleye Pollock in the Bogoslof Island Region. in: Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pac. Fish. Mgmt. Council, Anchorage, AK, pp. 233–250. www.fisheries.noaa.gov/resource/data/2015-assessment-walleye-pollock-bogoslof-island-region.
- Jiao, Y., Smith, E.P., O'Reilly, R., Orth, D.J., 2012. Modelling non-stationary natural mortality in catch-at-age models. ICES J. Mar. Sci. 69, 105–118.
- Johnson, K.F., Monnahan, C.C., McGillaird, C., Vert-pre, K., Anderson, S.C., Cunningham, C., Hurtado-Ferro, F., Licandeo, R., Muradian, M., Ono, K., Szuwalski, C., Valero, J., Whitten, A., Punt, A.E., 2015. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach ICES. J. Mar. Sci. 72, 137–150.
- Jurado-Molina, J., Livingston, P.A., Ianelli, J., 2005. Incorporating predation interactions to a statistical catch-at-age model for a predator-prey system in the eastern Bering Sea. Can. J. Fish. Aquat. Sci. 62, 1865–1873.
- Jurado-Molina, J., Gatica, C., Cubillos, L.A., 2006. Incorporating cannibalism into an age-structured model for the Chilean hake. Fish. Res. 82, 30–40.
- Kanaiwa, M., Chen, Y., Wilson, C., 2008. Evaluating a seasonal, sex-specific size-structured stock assessment model for the American lobster, *Homarus americanus*. Mar. Freshw. Res. 59, 41–56.
- Kenchington, T.J., 2014. Natural mortality estimators for information-limited fisheries. Fish. Fish. Oxf. (Oxf) 15, 533–562.
- Kienzle, M., Sterling, D., Zhou, S., Wang, Y.-G., 2016. Maximum likelihood estimation of natural mortality and quantification of temperature effects on catchability of brown tiger prawn (*Penaeus esculentus*) in Moreton Bay (Australia) using logbook data. Ecol. Model. 322, 1–9.
- Kinney, D., 2010. Multispecies Stock Assessment With Predator-prey Interactions. PhD Thesis. University of Washington.
- Kinney, D., Punt, A.E., 2009. Multispecies and single-species models of fish population dynamics: comparing parameter estimates. Nat. Res. Model. 22, 67–104.
- Lapointe, M.F., Peterman, R.M., 1991. Spurious correlations between fish recruitment and environmental factors due to errors in the natural mortality rate used in virtual population analysis (VPA). ICES J. Mar. Sci. 48, 219–228.
- Lapointe, M.F., Peterman, R.M., MacCall, A.D., 1989. Trends in fishing mortality rate along with errors in natural mortality rate can cause spurious time trends in fish stock abundances estimated by virtual population analysis (VPA). Can. J. Fish. Aquat. Sci. 46, 2129–2139.
- Lapointe, M.F., Peterman, R.M., Rothschild, B.J., 1992. Variable natural mortality rates inflate variance of recruitments estimated from virtual population analysis (VPA). Can. J. Fish. Aquat. Sci. 49, 2020–2027.
- Lee, H.-H., Maunder, M.N., Piner, K.R., Methot, R.D., 2011. Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Fish. Res. 109, 89–94.
- Lee, H.-H., Maunder, M.N., Piner, K.R., Methot, R.D., 2012. Can steepness of the stock-recruitment relationship be estimated in fishery stock assessment models? Fish. Res. 125–126, 254–261.
- Legault, C.M., 2009. Report of the Retrospective Working Group, January 14–16, 2008, Woods Hole, Massachusetts. US Dept. Commer., Northeast Fish Sci Cent Ref Doc. 09-01, 30 p. Available from: (Chair) 02543-1026. National Marine Fisheries Service, 166 Water Street, Woods Hole, MA. <http://www.nfsc.noaa.gov/nfsc/publications>.
- Lewy, P., Nielsen, A., 2003. Modelling stochastic fish stock dynamics using Markov chain Monte Carlo. ICES J. Mar. Sci. 60, 743–752.
- Livingston, P.A., Methot, R.D., 1998. Incorporation of predation into a population assessment model of eastern bering Sea walleye pollock. Fishery stock assessment models. In: Funk, F., Quinn II, T.J., Heifetz, J., Ianelli, J.N., Powers, J.E., Swweigert, J.F., Sullivan, P.J., Zhang, C.-I. (Eds.), Alaska Sea Grant College Program Report AK-SG-98-01. University of Alaska Sea, Fairbanks.
- Lorenzen, K., 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. J. Fish Biol. 49, 627–647.
- MacCall, A., 2013. Use of the delta method to evaluate the precision of assessments that fix parameter values. Fish. Res. 142, 56–60.
- Magnusson, K., 1995. An overview of the multispecies VPA—theory and applications. Rev. Fish Biol. Fish. 5, 195–212.
- Magnusson, A., Punt, A.E., Hilborn, R., 2013. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish. Fish. Oxf. (Oxf) 14, 325–342.
- Maunder, M.N., 1998. Integration of Tagging and Population Dynamics Models in Fisheries Stock Assessment. PhD Thesis. University of Washington.
- Maunder, M.N., Piner, K.R., 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. 72, 7–18.
- Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142, 61–74.
- Maunder, M.N., Wong, R.A., 2011. Approaches for estimating natural mortality: application to summer flounder (*Paralichthys dentatus*) in the U.S. Mid-Atlantic. Fish. Res. 111, 92–99.
- Maunder, M.N., et al., 2001. Integrated tagging and catch-at-age analysis (ITCAAN). In: Kruse, G.H., Bez, N., Booth, A., Dorn, M.W., Hills, S., Lipcius, R.N., Pelletier, D. (Eds.), Spatial Processes and Management of Fish Populations. University of Alaska, Fairbanks, pp. 123–146. Alaska Sea Grant College Program Report, AK-SG-01-02.
- Maunder, M.N., Lee, H.-H., Piner, K.R., Hamel, O.S., Cope, J.M., Punt, A.E., Ianelli, J.N., Methot, R.D., 2020. This issue. A review of estimation methods for natural mortality and their performance. Fish. Res. 00, 00–00.
- Mertz, G., Myers, R.A., 1997. Influence of errors in natural mortality estimates in cohort analysis. Can. J. Fish. Aquat. Sci. 54, 1608–1612.
- 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.
- Morgan, G.R., 1987. Incorporating age data into length-based stock assessment methods. Length-based Methods in Fisheries Research. Edited by D. Pauly and G.R. Morgan. International Center for Living Aquatic Resources Management, Metro Manila, Philippines and Kuwait Institute for Scientific Research, Safat, Kuwait, pp. 137–146.
- Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. 158, 96–101.
- Paloheimo, J.E., 1980. Estimating mortality rates in fish populations. Trans. Am. Fish. Soc. 109, 378–386.
- Paloheimo, J.E., Chen, Y., 1996. Estimating fishing mortality and cohort sizes. Can. J. Fish. Aquat. Sci. 53, 1572–1579.
- Pascual, M.A., Iribane, O.O., 1993. How good are empirical predictions of natural mortality? Fish. Res. 16, 17–24.
- Pauly, D., 1980. On the interrelationship between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. J. Cons. Int. Explor. Mer. 39, 175–192.
- Piner, K.R., Lee, H.-H., Maunder, M.N., Methot, R.D., 2011. A simulation-based method to determine model misspecification: examples using natural mortality and population dynamics models. Mar. Coast. Fish. 3, 336–343.
- Plagányi, E.É., 2007. Models for an Ecosystem Approach to Fisheries. FAO Fish. Tech. Pap. 477, 108pp.
- Pope, J.G., 1972. An investigation of the accuracy of Virtual Population Analysis using cohort analysis. Res. Bull. Int. Comm NW Atlant. Fish. 9, 65–74.
- Powers, J.E., 2014. Age-specific natural mortality rates in stock assessments: size-based vs. Density-dependent. ICES J. Mar. Sci. 71, 1629–1637.
- Punt, A.E., 1996. The effects of assuming that density dependence in the HITTER-FITTER models act on natural mortality rather than fecundity. Rep. int. Whal. Commun 56, 629–636.
- Punt, A.E., Smith, D.C., Thomson, R.B., Haddon, M., He, X., Lyle, J.M., 2001. Stock assessment of the blue grenadier *Macruronus novaezealandiae* resource off south-eastern Australia. Mar. Freshw. Res. 52, 701–717.
- Punt, A.E., Trinnie, F., Walker, T.I., McGarvey, R., Feenstra, J., Linnane, A., Hartmann, K., 2013. The performance of a management procedure for rock lobsters, *Jasus edwardsii*, off western Victoria, Australia in the face of non-stationary dynamics. Fish. Res. 137, 116–128.
- Punt, A.E., Hakamada, T., Bando, T., Kitakado, T., 2014. Assessment of Antarctic minke whales using statistical catch-at-age analysis. J. Cetacean Res. Manage. 14, 93–116.
- Punt, A.E., Butterworth, D.S., De Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016. Management strategy evaluation: best practices. Fish. Fish. Oxf. (Oxf) 17, 303–334.
- Ricker, W.E., 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board of Can. 191, 29–73 pp.
- Ross-Gillespie, A., 2016. Modelling Cannibalism and Inter-species Predation for the Cape Hake Species *Merluccius capensis* and *M. paradoxus*. PhD Thesis. University of Cape Town.
- Schnute, J.T., 1985. A general theory for analysis of catch and effort data. Can. J. Fish. Aquat. Sci. 42, 414–419.
- Schnute, J.T., Richards, L.J., 1995. The influence of error on population estimates from catch-age models. Can. J. Fish. Aquat. Sci. 52, 2063–2077.

- Sibert, J.R., 1984. A Two-fishery Tag Attrition Model for the Analysis of Mortality, Recruitment and Fishery Interaction. Tech. Rep. 13. Tuna and Billfish Assessment Progr. South Pacific Comm, Noumea, New Caledonia, 27 pp.
- Siddeek, M.S.M., Zheng, J., Punt, A.E., Pengilly, D., 2017. Effect of data weighting on the mature male biomass estimate for Alaskan golden king crab. *Fish. Res.* 192, 103–113.
- Siler, W., 1979. A competing-risk model for animal mortality. *Ecol.* 64, 750–757.
- Sims, S.E., 1984. An analysis of the effect of errors in the natural mortality rate on stock-size estimates using virtual population analysis (cohort analysis). *J. Cons. Int. Explor. Mer.* 41, 149–153.
- Sippel, T., Lee, H.-H., Piner, K., Teo, S.L.H., 2017. Searching for M: is there more information about natural mortality in stock assessments than we realize? *Fish. Res.* 192, 135–140.
- Sissenwine, M.P., Shepherd, J.G., 1987. An alternative perspective on recruitment overfishing and biological reference points. *Can. J. Fish. Aquat. Sci.* 44, 913–918.
- Sparre, P., 1991. Introduction to multispecies virtual population analysis. *ICES mar. Sci. Symp.* 193, 12–21.
- Stewart, I.J., Martell, S.J.D., 2015. Reconciling stock assessment paradigms to better inform fisheries management. *ICES J. Mar. Sci.* 72, 2187–2196.
- Stockhausen, W.T., 2019. 2019 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. <https://meetings.npfmc.org/CommentReview/DownloadFile?p=8ce00537-a862-4a28-ac75-a58c0eef905c.pdf&fileName=C4%203%20Tanner%20Crab%20SAFE%202019.pdf>.
- Sun, M., Zhang, C., Chen, Y., Xu, B., Xue, Y., Ren, Y., 2018. Assessing the sensitivity of data-limited methods (DLMs) to the estimation of life-history parameters from length-frequency data. *Can. J. Fish. Aquat. Sci.* 75, 1563–1572.
- Szuwalski, C., 2019. A stock assessment for eastern Bering Sea snow crab. <https://meetings.npfmc.org/CommentReview/DownloadFile?p=30a57738-eb23-40a7-9e82-ca415b36cd1.pdf&fileName=C4%201%20Snow%20Crab%20SAFE%202019.pdf>.
- Szuwalski, C., Punt, A.E., 2012. Identifying research priorities for management under uncertainty: the estimation ability of the stock assessment method used for eastern Bering Sea snow crab (*Chionoecetes opilio*). *Fish. Res.* 134–136, 82–94.
- Szuwalski, C.S., Ianelli, J.N., Punt, A.E., 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. *ICES J. Mar. Sci.* 75, 596–609.
- Then, A.Y., Hoenig, J.M., Hall, N.G., Hewitt, D.A., 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. *ICES J. Mar. Sci.* 72, 82–92.
- Thompson, G.G., 1994. Confounding of gear selectivity and the natural mortality rate in cases where the former is a nonmonotone function of age. *Can. J. Fish. Aquat. Sci.* 51, 2654–2664.
- Trijoulet, V., Fay, G., Curti, K.L., Smith, B., Miller, T.J., 2019. Performance of multispecies assessment models: insights on the influence of diet data. *ICES J. Mar. Sci.* 76, 1464–1476.
- Trijoulet, V., Fay, G., Miller, T.J., 2020. Performance of a state-space multispecies model: what are the consequences of ignoring predation and process errors in stock assessments? *J. Appl. Ecol.* 57, 121–135.
- Tsehay, I., Jones, M.L., Bence, J.R., Brenden, T.O., Madenjian, C.P., Warner, D.M., 2014. A multispecies statistical age-structured model to assess predator-prey balance: application to an intensively managed Lake Michigan pelagic fish community. *Can. J. Fish. Aquat. Sci.* 71, 627–644.
- Ulltang, Ø., 1977. Sources of errors in and limitations of virtual population analysis. *J. Cons. Int. Explor. Mer.* 37, 249–260.
- Van Kirk, K.F., Quinn, T.J., Collie, J.S., 2010. A multispecies age-structured assessment model for the Gulf of Alaska. *Can. J. Fish. Aquat. Sci.* 67, 1135–1148.
- Van Kirk, K.F., Quinn, T.J., Collie, J.S., A'Mar, Z.T., 2015. Assessing uncertainty in a multispecies age-structured assessment framework: the effects of data limitations and model assumptions. *Nat. Res. Model.* 28, 184–205.
- Vetter, E.F., 1988. Estimation of natural mortality in fish stocks: a review. *Fish. Bull. US.* 86, 25–43.
- Vincent, M.T., Brenden, T.O., Bence, J.R., 2017. Simulation testing the robustness of a multiregion, tag-integrated assessment model that exhibits natal homing and estimates natural mortality and reporting rate. *Can. J. Fish. Aquat. Sci.* 74, 1930–1949.
- Wang, Y., Liu, Q., 2006. Estimation of natural mortality using statistical analysis of fisheries catch-at-age data. *Fish. Res.* 78, 342–351.
- Wang, S.P., Maunder, M.N., Piner, K.R., Aires-da-Silva, A., Lee, H.H., 2014. Evaluation of virgin recruitment profiling as a diagnostic for selectivity curve structure in integrated stock assessment models. *Fish. Res.* 158, 158–164.
- Wetzel, C.R., Punt, A.E., 2011. Performance of a fisheries catch-at-age model (Stock Synthesis) in data-limited situations. *Fish. Res.* 62, 927–936.
- Wilderbuer, T.K., Turnock, B.J., 2009. Sex-specific natural mortality of arrowtooth flounder in Alaska: implications of a skewed sex ratio on exploitation and management. *N. Am. J. Fish. Manage.* 29, 306–322.