





ORIGINAL ARTICLE

A history and evaluation of catch-only stock assessment models

Daniel Ovando¹  | Christopher M. Free^{2,3}  | Olaf P. Jensen⁴  | Ray Hilborn¹ 

¹School of Aquatic and Fishery Sciences and Center for Sustaining Seafood, University of Washington, Seattle, Washington, USA

²Bren School of Environmental Science and Management, University of California, Santa Barbara, California, USA

³Marine Science Institute, University of California, Santa Barbara, California, USA

⁴Center for Limnology, University of Wisconsin - Madison, Madison, Wisconsin, USA

Correspondence

Daniel Ovando, School of Aquatic and Fishery Sciences and Center for Sustaining Seafood, University of Washington, Seattle, Washington, USA.
Email: danovan@uw.edu

Abstract

Understanding the status of fish stocks is a critical step in ensuring the ecological and economic sustainability of marine ecosystems. However, at least half of global catch and a vast majority of global fisheries lack formal stock assessments, largely due to a lack of sufficient data. Catch data, loosely referring to any catch records be it inclusive of discards or not, are the only type of fishery data available across a wide range of fisheries at a global scale. This has given rise to a long list of so-called “catch-only” models, intended to estimate aspects of stock status based primarily on characteristics of a fishery’s catch history. In this paper, we review the history, performance and potential of “catch-only” models to estimate stock biomass status. While individual catch-only models often report good performance, repeated efforts to examine the performance of these models have consistently found them to be imprecise and biased when applied to new data-limited fisheries. We demonstrate that a large reason for this is the simple lack of information on stock status contained in the shape of a catch history alone. Off-the-shelf use of catch-only models can lead to poor and biased estimates of stock status, potentially hindering efforts at effective management.

KEYWORDS

catch-only models, data-limited assessment, global fisheries, stock assessment, stock reduction analysis

1 | INTRODUCTION

Understanding the status of fish stocks is a prerequisite for an effective management system. Given the expense and complexity of surveying the oceans, for many fisheries, commercial catch data are the only source of information available to the public. Catch is often defined as the sum of both fisheries landings and discards, but throughout this paper, we use it as a broad term referring to any catch data available to fisheries scientists and managers, be it inclusive of discards or not. The Food and Agriculture Organization (FAO) of the United Nations maintains a global database of fisheries landings that includes over 20,000 individual catch histories by FAO statistical region, country and taxon. In contrast, the RAM Legacy Stock Assessment Database (RLSADB, www.ramlegacy.org), which includes most of the publicly available stock assessments conducted

around the world, contains information on just over 1,300 stocks. There is a large gap then between what is being caught and what is assessed in a formal way, but there is an alluring intuitive connection between the history of catches and the state of a fishery. A fish stock can sustain much lower catches when it is severely depleted (e.g. when biomass B much less than the biomass that could produce maximum sustainable yield B_{MSY}) than when it is more abundant (closer to B/B_{MSY} of one). The expectation that trends in catches should provide information about stock status, combined with a lack of survey data typically needed for formal stock assessments, has given rise to numerous efforts to assess the status of a stock based largely or exclusively on catch histories, so-called “catch-only models.” However, despite this long history of development and use, catch-only models have yet to provide a robust means of estimating either regional or stock-specific status, usually defined as biomass or

fishing mortality rates relative to a reference point (Free et al., 2020; Ovando et al., 2021). In this paper, we explore the history of catch-only stock assessment methods, summarize evidence as to their performance and demonstrate why catch-only models have limited potential as a means of stock assessment.

2 | A HISTORY OF CATCH-ONLY MODELS

To sustainably manage fisheries to produce food, employment and profits, we need to understand the status of fish stocks, especially so we can identify when they have been depleted to a level where the potential yield is reduced, requiring management actions to rebuild them. We also want to identify which stocks have been so reduced in abundance that they are of conservation concern and should be included in national or international endangered species listings such as CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) or the IUCN (International Union for the Conservation of Nature) Red List. In the 20th century, many fisheries management agencies began tracking the changing abundance of many stocks through various methods including scientific surveys, tagging programmes and commercial trends in catch per unit effort (CPUE). These abundance trend data are now available for over 1,300 fish stocks constituting roughly 50% of global marine fish landings in the RAM Legacy Stock Assessment Database (Hilborn et al., 2020; Ricard et al., 2012).

Unfortunately, we do not have abundance trend data for thousands of other fish stocks that are vital to many people's food security and incomes. Instead, catch histories are the most commonly available data type for fisheries across the globe. Data on catch have been recorded by various agencies since the 19th century, and commercial records have been used to reconstruct the historical catch for some fisheries for up to five centuries (Cushing, 1982). For example, the near extinction of the great whale species has been reconstructed from logbooks of whaling vessels that document the rise and fall of catch, and especially the decline of catch per unit effort, over centuries of whaling (Bockstoce & Botkin, 1983). Similarly, Cushing (1982) used processor and sales records between Norway and Sweden to reconstruct the catch of herring and to document changes in both the abundance and distribution of the stock.

Caddy and Gulland (1983) used catch histories to identify four types of fishery patterns including (i) steady state, (ii) cyclic fisheries, (iii) irregular stocks and (iv) spasmodic stocks all of which were based on examining trends in landings. They recognized that differences in patterns of catch have two fundamental causes, changes in the environment or changes in fishing pressure, and that either could cause a stock to behave in a way other than "steady state."

By the mid-1990s, the Food and Agricultural Organization of the United Nations (FAO) had compiled data on global landings going back to 1950. These landings data are summarized primarily by major FAO region, country of landing and taxonomic group—often not only described as an individual species, but also frequently described in larger taxonomic aggregates, with "marine fish not elsewhere included (nei)" representing the coarsest aggregate. We will

| | |
|--|-----|
| 1. INTRODUCTION | 616 |
| 2. A HISTORY OF CATCH-ONLY MODELS | 617 |
| 3. A TAXONOMY OF MODERN CATCH-ONLY MODELS | 618 |
| 3.1 Empirical approaches | 618 |
| 3.2 Mechanistic approaches | 619 |
| 3.3 Ensemble approaches | 619 |
| 4. EVALUATIONS OF CATCH-ONLY MODEL PERFORMANCE | 619 |
| 5. WHAT CAN CATCHES TELL US? | 621 |
| 5.1 Inference from Catch Histories | 621 |
| 5.2 Prediction from Catch Histories | 623 |
| 6. CONCLUSIONS | 627 |
| REFERENCES | 628 |

use the term "stock" to refer to these categories (region, country and taxonomy) recognizing that, in many if not most cases, they are coarse aggregates of many biologically independent populations.

In 1996, Grainger and Garcia (1996) used the FAO landings data to estimate when fisheries catch for individual stocks peaked, which they interpreted as an indicator of full exploitation. They state: "Through a preliminary analysis of trends, globally and by oceans, we attempt to demonstrate that these extended time series can be very useful in interpreting developments in the world's fisheries and so help in assessing the present situation as well as for planning and policy-making for the future." Part of their goal was to provide estimates of global fisheries potential, but, more relevant to our purpose here, they used the ratio between the maximum historical catch and the most recent catch to indicate whether there was further potential for expansion of catch, or whether the stock was fully or even overexploited and provide an estimate of overall stock status for each FAO region (Table 6 in Grainger & Garcia, 1996).

But, Grainger and Garcia were certainly circumspect regarding what the catch pattern shows. They wrote: "The difference between peak and current landings must be interpreted with caution. Peaks in a smoothed production probably give an indication of the average long-term yield that the species assemblage in a given area may be able to produce sustainably in the future, with proper management. However, in the case of demersal stocks sensitive to regime shifts on a decadal scale, peak harvests resulting from transient favourable environmental situations bear little relation to the average long-term yield." These concerns are much better understood today where we now understand that catches are known to decline for many reasons other than overfishing, perhaps the most common being shifts in productivity, which have been estimated to impact over 70% of fish stocks (Vert-pre et al., 2013). For example, changes in recruitment, and subsequently eventual fishable biomass, are often much better explained by factors other than spawning biomass (Cury et al., 2014; Szuwalski et al., 2015) This is closely followed by catches declining

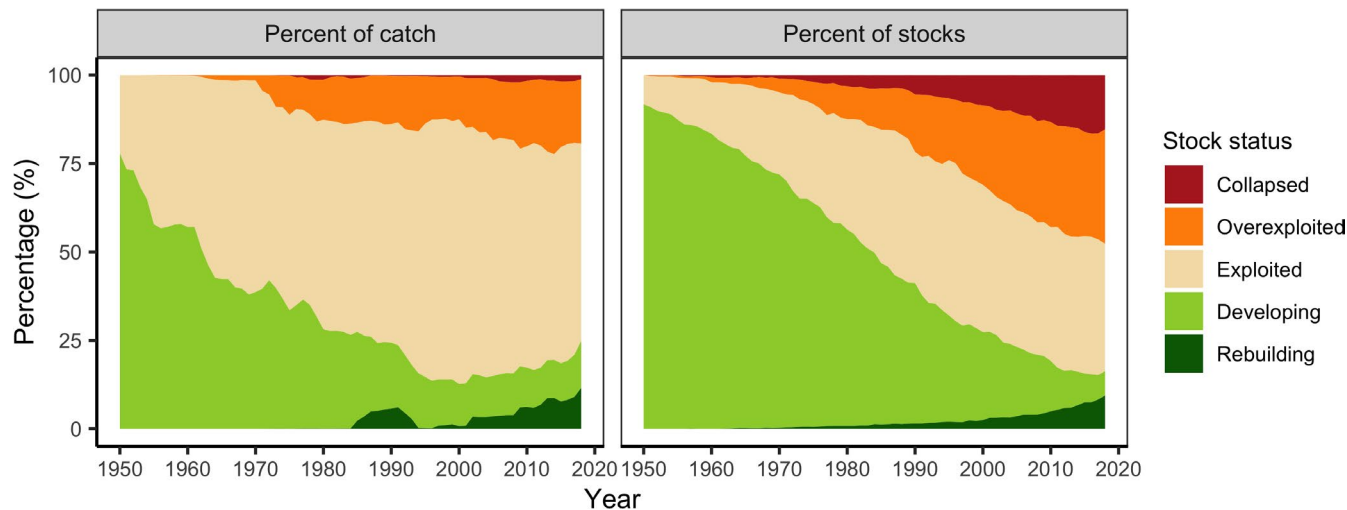


FIGURE 1 An example of estimation of stock status from catch-only data. From the Sea Around Us Project (SAUP), <http://www.seaaroundus.org/data/#/global/stock-status>

due to the implementation of fisheries regulations explicitly designed to reduce catch, which has now been observed in many coastal and high seas fisheries (Hilborn et al., 2020; Pons et al., 2018).

Following similar methods to Grainger and Garcia, (Grainger & Garcia, 1996), Froese (Froese & Kesner-Reyes, 2002) attempted to assess the status of global fisheries using the ratio between the current catch and the maximum catch to determine stock status. Stocks were considered collapsed if the catch has peaked and recent catches were less than 10% of the peak, overfished if recent catches were between 10% and 50% of the peak, and fully exploited if recent catches were greater than 50% of the peak. This broad method has been used extensively by the Sea Around Us Project to judge the status of global and regional fisheries (Pauly, 2007) (Figure 1).

In 2006, Worm et al. (2006) used these catch-ratio heuristics to identify stocks classified on this basis as “collapsed,” calculated the proportion of stocks collapsed for each year and projected a trend forward. Their projection of the proportion of fisheries categorized as collapsed reached 100% by 2048 (Worm et al., 2006), a claim that was intensively criticized by examination of the estimates of individual stocks where abundance was known (Hilborn, 2007; Murawski et al., 2007). As a result of the critique, a joint group of the original authors and critics evaluated trends in changes in abundance, rather than catch, and found only a small decline in the overall abundance of stocks with abundance estimates from data-rich stock assessment models (Worm et al., 2009). This group developed a global database of stock abundance called the RAM Legacy Stock Assessment Database (Ricard et al., 2012), which now contains estimates of the trends in abundance of stocks representing about 50% of global catch. The most recent global assessment of these data suggests that assessed fish stocks are, on average, increasing in abundance and above management targets (Hilborn et al., 2020).

Nevertheless, the remaining 50% of global fish landings reported to FAO come from stocks that are not formally assessed, and these stocks tend to come from countries where fisheries are particularly

important to food security and employment—largely in the tropical world. Thus, there has been an ongoing interest in using the reported landings data to infer the status of the unassessed stocks.

3 | A TAXONOMY OF MODERN CATCH-ONLY MODELS

Early catch-only models were based on simple heuristics comparing current landings to historic maxima. The importance of understanding the status of unassessed fisheries has led to a rapid proliferation of the number and types of catch-only assessment models built on Garcia and Grainger's (2005) initial efforts. In general, modern catch-only models utilize either an empirical, mechanistic or ensemble approach to estimate stock status from catch histories (Free et al., 2020). Empirical methods use statistical models trained on assessed stocks to derive associations between stock status and catch time series, often with auxiliary information about the biology of the exploited species or characteristics of the exploiting fishery. In contrast, mechanistic methods postulate an underlying population dynamics model, sometimes coupled within an underlying effort dynamics model, to explain changes in catch through a combination of changes in both stock abundance and fishing effort. Finally, ensemble methods use statistical models to combine and leverage the strengths of individual catch-only models, including predictions from both empirical and mechanistic methods. We provide a brief overview of the methods included within each approach and their relative strengths and weaknesses.

3.1 | Empirical approaches

All existing empirical catch-only assessment models estimate stock biomass status (e.g. B/B_{MSY} or depletion) by training statistical models on assessed stocks with data-rich status estimates in the RAM Legacy Stock Assessment Database (Ricard et al., 2012). These

methods employ a range of regression techniques and predictor variables including (i) panel regression models that use properties of the catch time series and characteristics of the stock (Costello et al., 2012; Thorson et al., 2012); (ii) a boosted regression tree model that uses properties of the catch time series (Zhou et al., 2017); and (iii) a boosted regression tree model that uses characteristics of the fishery including attributes of the catch history (Free et al., 2017). In general, the developers of these methods validated performance through cross-validation conducted during model fitting and/or through application to a withheld test dataset. Despite these efforts, challenges remain in truly validating the performance of empirical catch-only models. First, empirical methods are difficult to simulation test—the gold standard in model validation—due to their reliance on real-world relationships that cannot be reliably simulated. For example, surprising environmental responses, shifting management regulations, changing costs and prices and market disruptions from natural disasters, civil wars or global pandemics all impact abundance (Branch et al., 2011) but are challenging to simulate. Second, empirical methods are trained on data-rich stocks, whose dynamics may systematically differ from the data-poor stocks for which these methods are intended to assess, making their performance on data-poor stocks challenging to evaluate (Free et al., 2017, 2020).

3.2 | Mechanistic approaches

Existing mechanistic catch-only models use various implementations of stochastic stock reduction analysis (SRA) (Kimura et al., 1984; Kimura & Tagart, 1982; Walters et al., 2006) to describe the distribution of population trajectories and characteristics that could have resulted in an observed catch time series. Generally speaking, SRAs reconstruct historical abundance and exploitation rates by simulating biomass trajectories that could produce an observed catch time series given priors on the levels of depletion in the initial and final years of the catch time series, and often on population dynamics parameters such as carrying capacity, K , or intrinsic growth rate, r , in simple biomass dynamics models (Pella & Tomlinson, 1969; Schaefer, 1954). Existing mechanistic catch-only assessment models can be divided into those that estimate only a population dynamics model (Martell and Froese, 2013; Froese et al., 2017; Zhou et al., 2018; Ovando et al., 2021) and those that estimate a population dynamics model coupled with an effort dynamics model (Thorson et al., 2013; Vasconcellos & Cochrane, 2005). They also differ in the specific algorithms and assumptions used to perform the SRA.

Mechanistic approaches are attractive because (i) they can more easily be validated (or refuted) through simulation testing; (ii) they can estimate time series of biomass and fishing mortality and MSY-based reference points (unlike empirical approaches which generally estimate only one outcomes at a time, such as B/B_{MSY}); and (iii) they can estimate demographic parameters (i.e. r and K) that can be used to drive operating models for projecting fisheries outcomes under alternative management scenarios (e.g. Costello et al. (2016) and related works). However, the utility of these methods can be limited in that

they are highly sensitive to the choice of priors for key parameters (Bouch et al., 2021), the default priors of these methods can lead to inaccurate and biased estimates of stock status (Free et al., 2020; Ovando et al., 2021), and as we will show here, the catch time series does little to update status predictions beyond the priors. Despite these serious challenges, SRA methods, especially the CMSY family of methods (a shorthand for "Catch" and "Maximum Sustainable Yield," Froese et al., 2017; Martell & Froese, 2013), have been used to provide tactical advice for large commercial fisheries (e.g. Barman et al. (2020)) and strategic advice for regional (Froese et al., 2018; Smith et al., 2021) and global (Palomares et al., 2020) fisheries management.

3.3 | Ensemble approaches

Ensemble models, which attempt to combine the strengths of individual catch-only models to generate better predictions of stock status, have emerged as the most accurate and least biased catch-only predictors of biomass stock status (Anderson et al., 2017; Free et al., 2020). Existing ensemble approaches have experimented with various formulations ranging from weighted averages to statistical models based on traditional regression techniques to machine learning models that are themselves ensemble models (leading to the term *super ensembles*) (Anderson et al., 2017). Ensemble approaches have been shown to outperform individual assessments when used appropriately. However, ensembles require a method for aggregating results, whether simply taking means of predictions, or weighting individual models based on their perceived performance for the task at hand. The latter can perform better but requires some measure of the performance of individual models, often relative to a historical empirical baseline or based on simulation (Anderson et al., 2017). Furthermore, existing ensemble models only predict biomass stock status in the final year of the catch time series (i.e. they do not predict time series, reference points or effort status) and cannot be used to drive population dynamics in fisheries operating models.

4 | EVALUATIONS OF CATCH-ONLY MODEL PERFORMANCE

The use of catch-only models in data-poor stock assessments and fisheries management depends on quantitative evidence demonstrating their ability to infer and predict aspects of stock status, and their subsequent utility to management. By inference, we refer to the ability of catch-only models to estimate often unobservable parameters of a stock, such as MSY, based on data, in the manner of any statistical stock assessment model. By predictive ability, we refer to the ability of a model to make accurate predictions of observations outside of those on which it was trained (e.g. in the manner of (Costello et al., 2012)).

A thorough evaluation of catch-only model performance must therefore demonstrate that a method can provide accurate estimates and predictions of stock status and useful management

advice under a range of potential conditions (e.g. species type, fisheries management context, and length and uncertainty of the catch time series). These evaluations can be achieved through testing on assessed fisheries with data-rich estimates of stock status or testing on simulated fisheries with known stock status. Simulation testing can offer true “known” values to compare model predictions against, greater sample sizes, the ability to compare performance across more conditions and an opportunity to evaluate the consequences of using advice from catch-only models in management decisions. However, the validity of simulation testing also depends on the degree to which the dynamics of the simulation model match the “true” dynamics of the systems in which catch-only models are intended to be applied. Empirical validation can allow comparison to a wider range of potentially more realistic models, but have the shortcoming of comparison against model estimates, not true known values.

Rigorous performance evaluations should also reflect the conditions under which a method is likely to be applied. For example, if the performance of a catch-only model depends on the priors used to parameterize the method, then testing should reflect an expert's ability to accurately set priors for a new fishery under the conditions they are likely to face in the real world. It has been rare for either internal evaluations (testing conducted by method developers) or external evaluations (testing conducted by peers) to include all of these idealized components (i.e. simulation testing, empirical testing, robust cross-validation across a range of applicable scenarios and management strategy evaluation). Thus, we synthesize the performance of catch-only models across a combination of several studies, acknowledging that a full review of the testing methods used to evaluate every catch-only model is beyond the scope of this paper.

We begin with an examination of the methods used to internally validate CMSY (Froese et al., 2017), arguably the most widely used model for assessing the status of fisheries using only catch data. When testing the performance of CMSY on data-rich stocks, the developers manually updated the default priors for 54 (34%) of the test stocks to ensure that they contained the true parameter values. They argued that this represents a scenario in which expert opinion has “not made gross errors in setting broad prior biomass ranges.” However, this likely represents an optimistic view of the ability of experts to set priors in a new fishery and does not reflect the common use of these models. For example, default priors, not expert opinion, were used to assess 1,320 global stocks in Palomares et al. (2020), with the authors justifying the use of CMSY based on its performance under the idealized conditions it was tested on. The developers also assessed the performance of CMSY on a set of 48 simulated stocks. However, the operating model used to generate the simulated test dataset was identical in structure to the underlying SRA model. A more rigorous testing approach would test the simplified population dynamics algorithms underlying CMSY and other catch-only models against more realistic age-structured models, which now have widespread software support (e.g. *FLR*; (Kell et al., 2007)).

As another example, Zhou et al. (2018) tested OCOM, another SRA approach, on assessed stocks that contributed to the

parameterization of their prior on final year depletion. Best practices in predictive modelling would have the model tested on a set of data fundamentally isolated from its development (Kuhn & Johnson, 2013). The testing regime in Zhou et al. (2018) also employed a common “leave-one-out” strategy, in which the model is fit to all but one stock, and then used to predict the status of the omitted stock, with this process repeated across each of the stocks. This can be an appropriate test if the goal of a model is, for example, to provide an assessment for one unassessed stock in an otherwise heavily assessed region. However, to the extent that stock status exhibits correlation across stocks, whether due to common management regimes, life histories or environmental shocks, it is not clear that a model trained in one region will be able to predict as well in a new region with different dynamics.

External evaluations, which often employ a more thorough set of testing procedures than those employed in the original publication describing a given catch-only model, generally show catch-only models to be much less accurate and more biased than reported by internal evaluations (Anderson et al., 2017; Bouch et al., 2021; Free et al., 2020; Ovando et al., 2021; Pons et al., 2020). Rosenberg et al. (2014) measured the accuracy and bias of four catch-only models (Catch-MSY, COMSIR, SSCOM and mPRM) across 5,760 stocks simulated with an age-structured operating model (i.e. different from the production models commonly used in the SRA methods) and found them to be imprecise and biased predictors of B/B_{MSY} . Anderson et al. (2017) found similar results when testing the same methods on data-rich stocks in the RAM Legacy Database. Weir (2017) showed that the same four methods were poor classifiers of IUCN threat status (i.e. vulnerable, endangered or critically endangered) when applied to the Rosenberg simulated stocks. Free et al. (2020) evaluated seven additional methods on both RAM stocks and the Rosenberg simulated stocks and found all eleven to be imprecise and biased predictors of B/B_{MSY} and categorical status (i.e. under exploited, fully exploited or overfished). Pons et al. (2020) tested the performance of both Catch-MSY (Martell and Froese, 2012) and CMSY (Froese et al., 2017) on a new set of 2,700 simulated stocks and still found both to be imprecise and biased, especially when the stocks were lightly exploited. Bouch et al. (2021) applied CMSY (Froese et al., 2017) to 17 ICES stocks and found it to overestimate relative fishing mortality and underestimate relative stock status, especially for stocks showing signs of recent recovery. Sharma et al. (2021) applied CMSY and sraplus (Ovando et al., 2021) to 48 ICES stocks and found both to generate biased predictions of stock status when using default priors. Ovando et al. (2021) similarly found that catch-only models frequently produced imprecise and biased estimates of stock status, misclassifying the overall state of a fishery 57% of the time.

Developing a robust set of tests for a predictive model, particularly for a problem as complex as predicting the stock status of a fishery, is a difficult task. It is understandable that different groups will develop different testing regimes based on their anticipated use of the model, and the suite of models and data at their disposal. However, the potentially optimistic testing regimes used in the initial

justifications for many existing catch-only models may explain the sharp drop in performance exhibited by these models when confronting external evaluations.

Estimating stock status is only the first step in the fisheries management process and fully judging the utility of catch-only models in data-poor fisheries management requires tracking the impact of their status determinations on harvest control rules, biomass status and fisheries outcomes such as catch, profits and employment. A well-designed harvest control rule can help overcome known biases in an assessment method. Such testing can be achieved through management strategy evaluation (MSE), which simulates the resource, fishing fleet and management decisions in a closed-loop system (Punt et al., 2016) and has been underutilized in either internal or external evaluations of contemporary catch-only models (but see (Harford & Carruthers, 2017)). Walsh et al. (2018) used a MSE model to evaluate whether an ensemble model, the most accurate catch-only approach to estimating stock status (Anderson et al., 2017; Free et al., 2020), can reliably inform fisheries harvest control rules. They found that the large inaccuracies in even this most reliable of catch-only models require cautious harvest control rules to prevent overfishing and improve B/B_{MSY} status. While successful, cautious harvest policies result in considerable foregone sustainable yields (Walsh et al., 2018). Such trade-offs are likely larger with even less accurate status determinations.

5 | WHAT CAN CATCHES TELL US?

While the extensions to the basic catch heuristics first employed by Froese, Garcia and Grainger (Froese & Kesner-Reyes, 2002; Garcia & Grainger, 2005) described above may—under the right circumstances—provide performance gains, the general consensus of numerous external evaluations is that catch-only models provide imprecise and biased estimates of stock status when applied to new fisheries. In this section, we consider why this might be, by breaking down the ability of catch-only models to infer information on stock status and the ability of catch-only models to predict stock status, two related but critically different tasks.

Drawing from Donoho (2017), this distinction relates to the “two cultures” outlined by Breiman (2001), prediction and inference. Inference is concerned with estimating the parameters of a hypothesized data-generating process. This is the realm where most fisheries, and indeed statistical, models operate. A classic example of an inferential task would be to estimate the growth rate and carrying capacity, and by extension stock status, of a fishery based on an assumption of logistic growth and given an index of abundance and a time series of catches. Inferential models are often judged by the strength of the evidence in the data for a particular model structure, for example the posterior distribution of a parameter of interest (Tredennick et al., 2021). Critically, inference allows us to estimate the value of unobserved parameters (e.g. stock status relative to reference points) based on assumptions of a data-generating process linked to observed data (e.g. time series of catch and an abundance index).

Purely predictive models can be agnostic as to the underlying data-generating model. They are concerned only with the accuracy of predictions made by the model and not with the strength of evidence for particular parameters of the model. While more interpretable models such as linear regressions can be used for prediction, increasingly “black-box” models such as random forests, boosted regression trees and other machine learning methods are employed for predictive modelling. In these cases, the model seeks to leverage correlations between an outcome of interest and candidate variables to make accurate predictions. Predictive models of this type cannot estimate the value of unobserved parameters on their own; they require training on a dataset with “known” values. In the fisheries context, a predictive model might seek to predict stock status as a function of a catch history-based off relationships observed between catch histories and stock status values for a subset of fisheries with “known” stock status.

5.1 | Inference from Catch Histories

The broad intuition behind catch-only models is relatively simple: Collapsed fisheries produce zero or little catch, and fisheries collapse after overexploitation. Therefore, a sharp rise in catches followed by a sudden and sustained crash in catches likely tells us that the stock was overexploited and collapsed. While this intuition seems sound and would certainly apply to some overexploited open-access fisheries, what can we actually infer from simply observing the catch history of a stock? To answer this, consider a simple and common equation for catch in a fishery:

$$C_t = q_t E_t B_t \quad (1)$$

Where t is time, C is catch, q is the catchability coefficient for fishing effort E , and B is fishable biomass. If all we observe is C , we cannot clearly estimate B using conventional statistical without making numerous other assumptions: that is we have one equation and three unknowns (though see discussions of Takens' theorem, as in (Thorson et al., 2013), for considerations of ways in which information on biomass may be embedded in catches). Given only a catch history, to equate changes in catch with changes in biomass, we must assume that both catchability q and effort E are constant over the same time period. While this may occur in some fisheries, it is an unlikely scenario. This is why formal fisheries assessments are instead based on catch-per-unit-effort data and focus largely on separating catchability from biomass (Hilborn & Walters, 1992).

It is clear then, that in the absence of other data, through most approaches, we cannot infer anything about trends in biomass from catch alone without invoking strong assumptions about parameters such as catchability and effort over time. What though about the shape of the catch history? Surely, conditional on a model, certain catch histories imply different fishery states? This is the core idea on which most catch-only models are based. To test this idea, we ran a series of stock reduction analysis (SRA) models to evaluate the

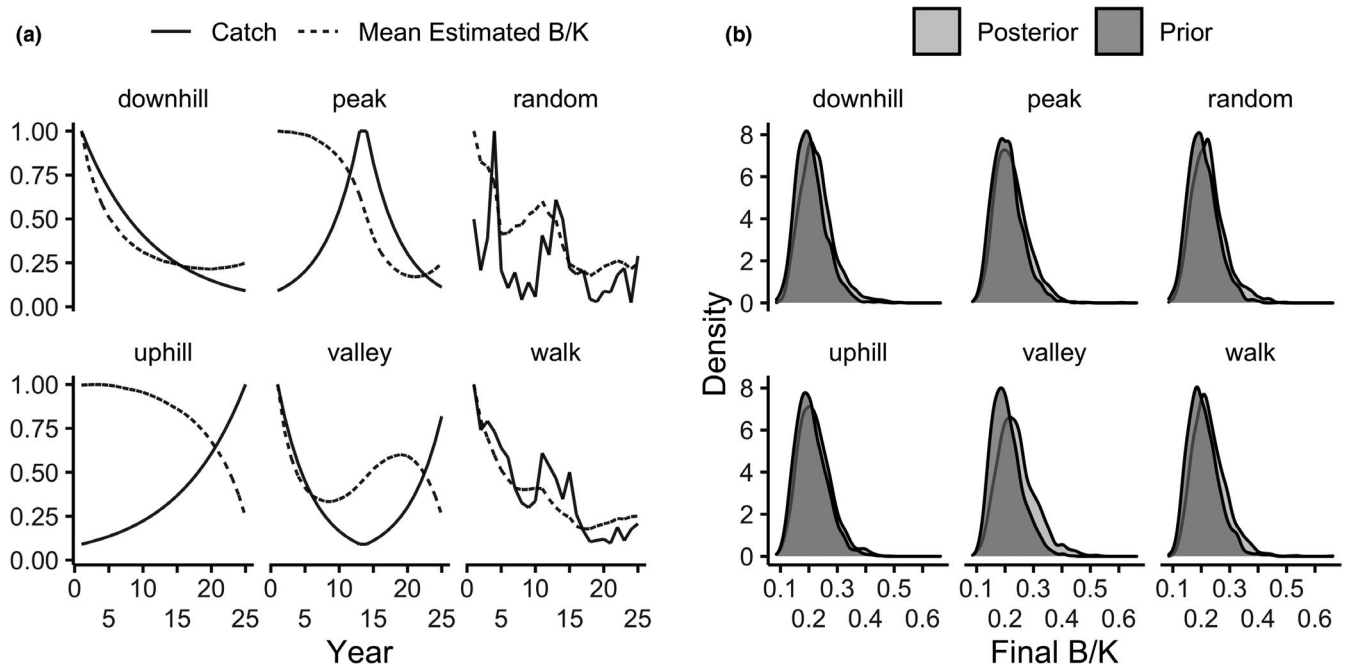


FIGURE 2 Simulated case studies demonstrating informational content of catch histories alone. Key parameters are growth rate and carrying capacity (K). The prior for the growth rate is log-normally distributed with a mean obtained from FishLife and a coefficient of variation (CV) of 20%. The prior for K is log-normally distributed with a mean equal to five times the maximum catch and CV of 100%. The prior on stock status (B/K) in the final year is log-normally distributed with mean of 0.2 and CV of 25%, mean of 1 and CV of 20%. (a) Shows simulated catch histories (solid line) and the resulting estimates of B/K (biomass relative to carrying capacity, dashed line) given each simulated catch history. (b) Shows the prior (dark grey) and posterior (light grey) distributions of B/K in the final year. Priors on initial and final B/K and life history are identical across all simulations

impacts of the shape of the catch history on the stock status estimated by each algorithm.

The stock reduction algorithm used is *sraplus*, described in Ovando et al. (2021). The model is supplied with a prior on the growth rate r (log-normal with mean = 0.35 and CV = 20%) and the carrying capacity K (log-normal mean = 5 times max catch, CV = 100%, capturing a range similar to the version of CMSY used in (Anderson et al., 2017)) for a Pella-Tomlinson (Pella & Tomlinson, 1969) surplus production model, as well as priors on B relative to K in the initial (log-normal mean = 1, CV = 25%) and final years of the catch history (log-normal mean = 0.2, CV = 25%). We chose these relatively informative priors to reflect ranges used by default for analogous parameters in many common catch-only models (Table S1). We then generated a series of catch histories that have the same average lifetime catch but vary starkly in their dynamics: a constant increase, a constant decrease, a peak then a decline, a decline then a recovery, a random walk and random values (Figure 2a). We then used *sraplus* to perform a standard stochastic stock reduction analysis on these artificial catch histories, using the same priors on growth rate, carrying capacity, and initial and final B/K for each simulated catch history. In each iteration of the stock reduction analysis, the algorithm randomly selects a value for the growth rate, carrying capacity and initial and final B/K from the supplied prior distributions. The Pella-Tomlinson model is then run with the selected prior values on growth rate and carrying capacity, initialized at the sampled initial B/K , using the artificial catch history in question. Any combination of

prior draws that results in biomass values less than or equal to zero is immediately rejected. The remaining “viable” draws from the stock reduction analysis are then sampled in proportion to the likelihood of the projected final B/K relative to the prior distribution of the final B/K . These general methods match the steps laid out in Walters et al. (2006) barring the difference in the operating model used and reflect the general steps in CMSY (omitting a prior on intermediate B/K and strict bounds on B/K).

We then examined the distribution of B/K predicted by the catch-only model relative to the explicit priors supplied on stock status. Note that the supplied B/K prior in *sraplus* is log-normally distributed, so the model has support for a continuous range of B/K values greater than zero. This means that, in theory, the post-model-pre-data B/K values generated by *sraplus* are allowed to vary greatly from their inputs. Any stock reduction analysis done in this manner will produce a positive bias in final B/K estimates, as there are always more ways for a stock to be less depleted than more depleted. Consider a prior on final B/K with nearly all its probability density concentrated near zero, a stock reduction analysis model will assign a low probability to an individual draw from the stock reduction algorithm that produced a final B/K near one, at the tail end of the prior distribution. However, suppose that our prior on carrying capacity was much greater than the average catch volume, nearly all draws from the stock reduction algorithm will have a final B/K near one, since catches will be much lower than carrying capacity. The final distribution of B/K produced by *sraplus* will then be pulled close

to one, since even though the prior assigns a low probability to any individual draw with B/K equal to approximately one, the model has many more opportunities to select these high values.

By default, srplus employs a prior-predictive tuning procedure to correct this behaviour (see Ovando et al. (2021)). But, we have turned that feature off for this analysis to isolate the impact of the catch history shape itself on estimates of stock status from catch-only models. A common Bayesian model diagnostic is to compare the prior and posterior distributions, to see how much any priors in the model were updated by the data the model was confronted with. The premise behind catch-only models is that the catch history on its own can tell us something new about stock status that was not encoded in the priors themselves. However, comparing the prior and posterior of the most recent B/K estimate for these simulated catch histories, we see that the prior and posterior are essentially identical despite the priors being confronted with starkly different catch history shapes. Conditional on our life history and stock status priors, and the model choice,¹ in most cases, the shape of the catch history alone does not contain meaningful information about stock status in the final year of the fishery that was not already encoded in our priors, although there may be edge cases where this will not be true. For example, if we know a given fish species exists, but catches for that species have been zero in every time step, barring massive process error, any stock reduction will not be able to find many parameters values that results in final B/K values much less than one. But for more typical catch history forms, like those used in our example, conditional on the priors, the shape of the catch history itself does not meaningfully update our prior beliefs about final stock status (Figure 2b).

This result may appear somewhat confusing to readers who have previously encountered catch-only models that produced posterior distributions of final B/K values that appeared very different than their supplied priors. In our above examples, we use sufficiently diffuse priors on key life history parameters (growth rate and carrying capacity) such that they do not constrain the model. However, this is not always the case in catch-only models. Consider a fishery with an average annual lifetime catch of 10,000 MT (metric tons), suppose then that we supply a prior on carrying capacity K that is uniform [0, 11,000], and a prior on final B/K that is uniform [0,1]. Clearly, given average annual catches of 10,000 MT and a prior on K that at most barely exceeds that level, the only possible solution for a standard model is that the stock is at very low levels. Therefore, despite having a very diffuse explicit prior on final B/K , the posterior will show a very tight distribution at a very low stock status. See Supporting Information for additional analyses demonstrating these phenomena across a range of more diffuse and more informative priors for both high and low stock status. These scenarios collectively demonstrate that stock status in the most recent time step is positively correlated with the diffusion of the priors in stock reduction algorithms of this form.

¹Alternative model choices, such as assuming that fishing effort must stay constant or is strongly autocorrelated, could change this.

The problem here is that when we provide a model with priors on, in the case of a simple Schaefer model (Schaefer, 1954), growth rate r , carrying capacity K , and initial depletion, conditional on the model structure and the catch history, we have provided an implicit prior on stock status. When we then place a second *explicit* prior on final stock status, we create a problem known as Borel's Paradox (Poole & Raftery, 2000). Borel's Paradox suggests that when we supply an implicit and explicit prior on final stocks status, the posterior will reflect the joint combinations of all these layered priors, creating the appearance of learning. This means that certainly, if we provide sufficiently informative and accurate priors, a catch-only model may provide good results. But, the performance is entirely dependent on the priors and the validity of the priors and the chosen model, not on information gleaned from the shape of the catch history alone. This means that the choice of priors on life history has a substantial influence on the outcomes of catch-only models.

In summary then, catch histories alone do not allow us to infer stock status. The "information" about stock status comes almost exclusively from specific assumptions such as relatively constant effort, or priors on r and K and final stock status. Given specific priors, different catch histories can imply different states of a fishery, but this updating is dependent on the supplied r and K priors, not intrinsic information derived from the catch history that is preserved as you modify the r and K priors.

5.2 | Prediction from Catch Histories

This simple simulation exercise illustrates that the ability of catch-only models to infer recent stock status based on the shape of a catch history is limited. However, on its own, this does not necessarily mean that, as a measure of stock status, catch-only models would perform badly as a predictive model in the real world, suppose that every stock in the world followed Schaefer dynamics (Schaefer, 1954) and had the same r , K and initial depletion. If we knew the value of these parameters, or could provide reasonable priors on their distribution, applying that knowledge to the catch history of each stock through a catch-only model would provide the correct stock status for every fishery. The question then is how robust any selected priors are, which, as we show above is where the actual information about stock status in catch-only models is coming from.

Many catch-only models have built in models or heuristics for generating priors. For example, CMSY (Froese et al., 2017) has a series of internal heuristics that create priors on initial, intermediate and final B/K based on the characteristics of the catch history. Costello et al. (2012) also produced estimates of stock status based on empirically derived relationships between catch histories and observed stock status. Both of these methods share a common trait of having observed, either informally or statistically, relationships between catch histories and stock status. The question then is to what extent do such prior-generating mechanisms produce valid

TABLE 1 Status of fish stocks (measured as B/B_{MSY}) from different regions from known assessments or expert knowledge, and from three published catch-only models

| Region | B/B_{MSY} | Catch-only assessment method | | |
|----------------------------------|-------------------|--|---|---------------------------------|
| | | Panel regression (Costello et al., 2012) | Ensemble model (Rosenberg et al., 2018) | Catch-MSY (Froese et al., 2012) |
| NE Atlantic | 1.30 ^b | 0.91 | 0.97 | 0.24 |
| NE Pacific | 1.78 ^b | 1.18 | 1.02 | 0.31 |
| WC Pacific (China and Indonesia) | Low ^a | 0.89 | 1.06 | 1.06 |
| Eastern Indian | Low ^a | 0.94 | 1.04 | 1.04 |
| Mediterranean and Black Sea | 0.56 ^b | 0.88 | 0.96 | 0.96 |

^aFrom (Melnichuk et al., 2017). ^bfrom RAM Legacy Stock Assessment Database.

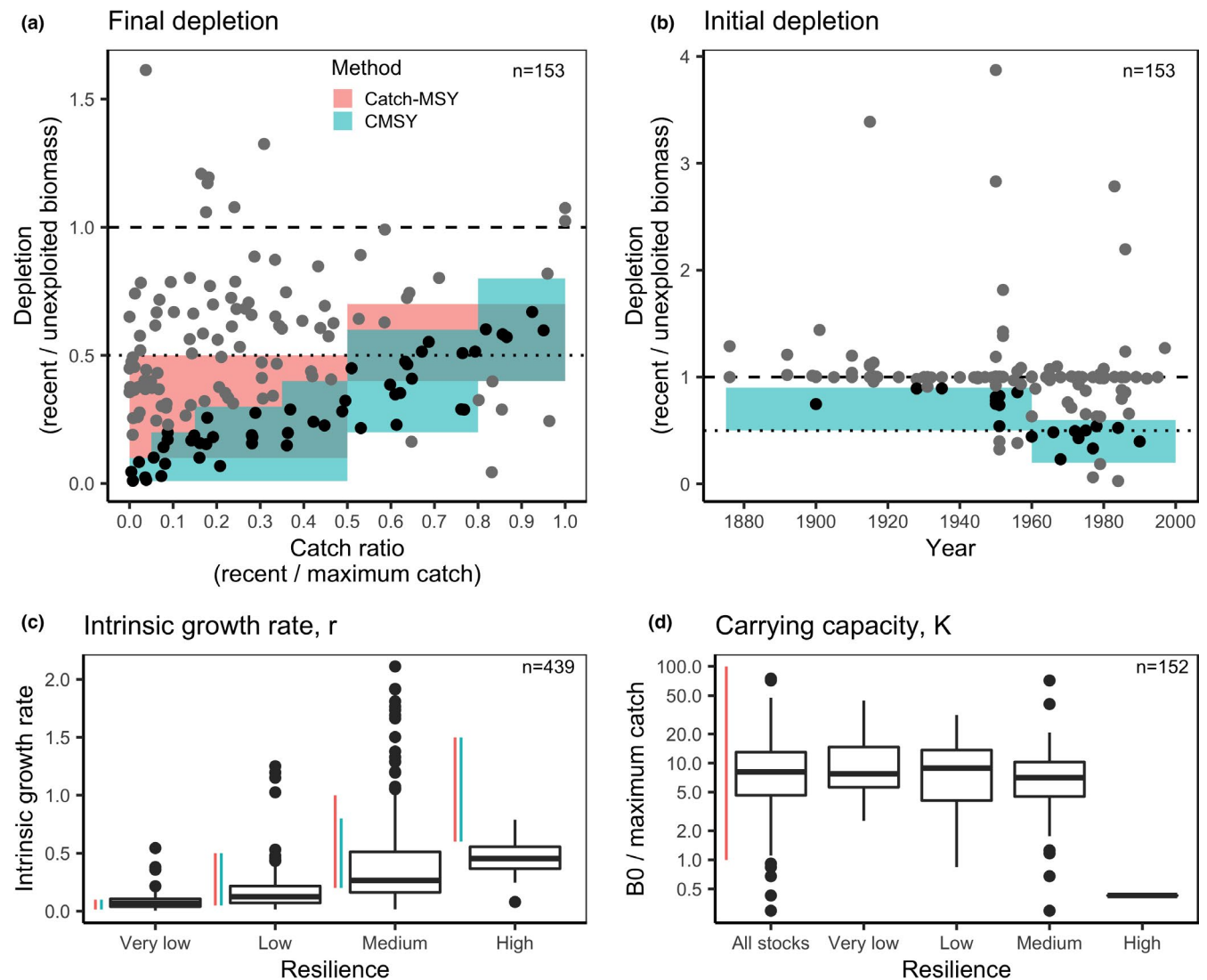


FIGURE 3 Default priors for (a) final B/K , (b) initial B/K , (c) intrinsic growth rate, r , and (d) carrying capacity, K , from Catch-MSY (Martell & Froese, 2013) and CMSY (Froese et al., 2017) compared with values derived for stocks with data-rich stock assessments. In all panels, the coloured shading indicates default priors and the black points and/or boxplots indicate values derived from data-rich stocks. In (a) and (b), black and grey points indicate stocks with B/K estimates occurring inside and outside the default priors used in CMSY respectively. In these panels, the dashed lines indicate pre-exploitation biomass ($B/K = 1$), which can be exceeded due to age-structured population dynamics and the dotted lines indicate B_{MSY} ($B/K = 0.5$). Catch-MSY sets default initial B/K priors based on the catch ratio in the final year of the catch time series and is not shown. CMSY sets default carrying capacity priors based on resilience, initial B/K and maximum catch and is not shown

predictions when applied to new fisheries? To that end, we consider here whether catch-only models show evidence of being effective predictors of stock status.

When we compare published predictions from catch-only models for regions of the world where we either have good information on stock status or general recognition of concern about stock status—the catch-only models detect only minor differences between regions that have very different statuses (Table 1). A very high proportion of stocks are assessed in the NE Pacific and NE Atlantic and are well above target reference points while numerous assessments from the Mediterranean show that stocks there are in poor shape. For the Western and Central Pacific (mostly China and Indonesia) and the Eastern Indian Ocean, we have expert opinion that stock status is poor (Melnichuk et al., 2017). All three catch-only models considered here struggle to reproduce this understanding of either the absolute or relative differences in stock status between these regions (Table 1).

Diving into these discrepancies, we examined the performance of the default priors used by Catch-MSY (Martell & Froese, 2013) and CMSY (Froese et al., 2017), arguably the two most widely used catch-only stock assessment methods, when applied to stocks with data-rich assessments in the RAM Legacy Stock Assessment Database (Ricard et al., 2012). We found that the default priors used by Catch-MSY and CMSY for final B/K , which are set based on the ratio of final to maximum catch (Table S1), correctly captured only 52% and 33% of the data-rich estimates respectively (Figure 3a). Furthermore, the default priors were highly biased for stocks with low catch ratios (final/maximum catch <0.5), which represent over three-quarters of the evaluated RAM stocks. The default priors assume that all of these stocks are below B_{MSY} when 38% of these stocks are actually above B_{MSY} (Figure 3a). The default priors for initial B/K , which are set based on the first year in the catch time series, are also poor predictors of initial B/K . These priors correctly captured only 13% of the data-rich estimates (Figure 3b) and were especially biased for catch time series beginning after 1960. The priors assume that stocks with catch time series beginning after 1960 begin over to fully exploited ($B/K = 0.2$ – 0.6 ; Table S2) when, in fact, 82% of these stocks were lightly exploited in the first year of the catch time series ($B/K > 0.6$; Figure 3b).

The default priors for intrinsic growth rate r , which are based on resilience (Table S3), are more reasonably distributed than the B/K priors but include high values not relevant to most stocks. This is problematic given that the CMSY algorithm favours high growth rates and low carrying capacities in the “tip of the triangle” of viable r - K pairs (Froese et al., 2017). This drives the CMSY algorithm to favour high, and unlikely, growth rates, which could lead to biased predictions of stock productivity and target fishing mortality rates (F_{MSY}). We surmise that this could lead to high overfishing if default values of CMSY are used to assess stocks and to guide tactical management. Finally, the default priors used by Catch-MSY for carrying capacity are reasonable but are very broad (Figure 3d). There is little empirical support (Figure 3d) for the default priors for carrying capacity used by CMSY, which are based on a complex combination of

resilience, initial B/K and maximum catch (Table S4). Furthermore, the “tip of the triangle” assumption employed by CMSY favours low carrying capacities, which contributes to CMSY’s pessimistic bias toward low stock biomass status (the majority of this bias likely comes from the pessimistic final B/K priors).

To the extent that one could a priori reliably predict life history parameters r and K for a fishery, a catch-only model would produce reliable predictions to the degree that the population can be well represented by the selected population dynamics model. However, the default prior-generating algorithms for life history parameters utilized in one of the most commonly employed catch-only models, CMSY, have clear biases that will in turn bias estimates of stock status (Figure 3). An alternative strategy then is to identify empirical predictive relationships between catch history attributes and stock status. This is the underlying philosophy of efforts such as Costello et al. (2012), Thorson et al. (2012) and Zhou et al. (2017). We revisit these approaches here to illustrate both the potential and limitations of predictive catch-only models.

We trained a machine-learning model (a boosted regression tree) to predict stock status as a function of catch history. The model treats values of B/B_{MSY} reported in the RLSADB as “known” and then fits a model using candidate characteristics of the catch history, along with life history parameters, to predict the values reported in the RLSADB. Performance of predictive models is assessed based on the accuracy of predictions made by the model on observations held out from the training process. The testing regime must then be carefully designed to reflect the intended use of the model in question. In this case, the goal was to be able to predict the status of fisheries with unknown stock status, often in completely new geographic regions, based on observed relationships in fisheries with “known” stock status. We filtered the RLSADB down to 293 stocks with values of B/B_{MSY} and at least 25 years of catch data. We then included only the values after the first 20 years of the fishery, to prevent the model from spending too much effort trying to estimate stock status in the early years of the fishery (assuming, in this case, that the purpose of the model was to provide estimates of current stock status).

In order to assess model performance when faced with stocks in an entirely new region, we then split these stocks into a training and testing datasets, where the training dataset contained all stocks that fit our criteria except those from New Zealand, Australia and South Africa ($N = 293$; 88% of stocks), and the testing dataset contained all stocks that fit our criteria from New Zealand, Australia and South Africa ($N = 40$; 12% of stocks). The data were split using these three countries to retain a testing dataset that had a reasonable sample size and was geographically isolated from the training dataset. All model tuning (Figure S2) and fitting were performed exclusively on the training dataset.

The candidate covariates for the model included various attributes of the catch history and life history traits of the species and are similar to the covariates used in other empirical catch-only models (Anderson et al., 2017; Costello et al., 2012; Thorson et al., 2012; Zhou et al., 2017). Life history traits include steepness, asymptotic size and the ratio of natural mortality to growth rate and were drawn

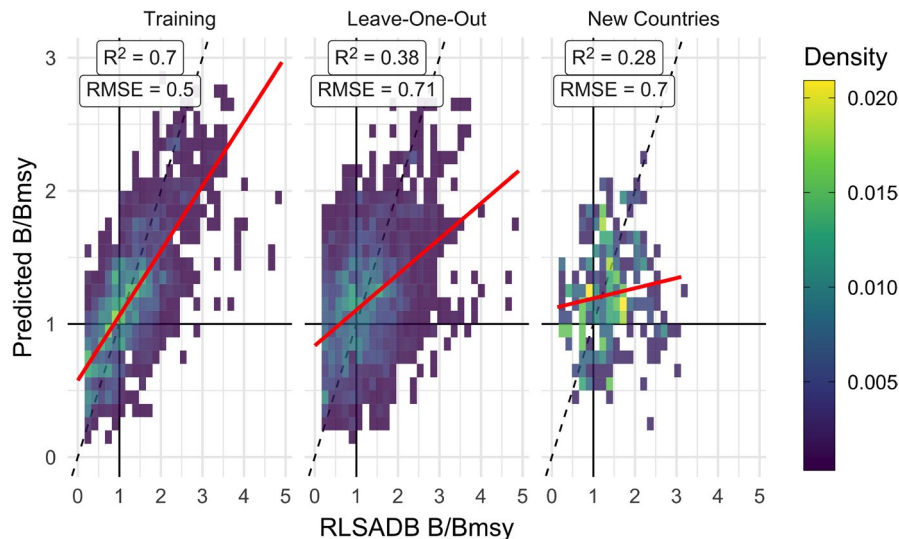


FIGURE 4 RAM Legacy Stock Assessment Database (RLSADB) reported (x-axis) and machine learning predicted (y-axis) B/B_{MSY} values. Colour represents density of individual points across these coordinates. The training dataset ($N = 293$) includes RLSADB stocks from everywhere except Australia, New Zealand and South Africa. The testing dataset includes only stocks from Australia, New Zealand and South Africa ($N = 40$). Leave-one-out analysis are the out-of-bag predicted values for the stocks in the training dataset. Test indicates R^2 and root mean squared error (RMSE) for each individual panel. The black dashed line is the 1:1 fit, and the red solid line is a linear fit between the reported and predicted values

from FishLife (Thorson, 2020). Attributes of the catch history included factors such as catch divided by the mean or maximum catch in each time step. We also used an unsupervised spectral clustering algorithm to classify each of the stocks in the RLSADB into one of four catch history shapes, with the exact shape of the catch histories clusters assigned to each cluster determined by the spectral clustering algorithm (Figure S1). This catch history classification was also included as a predictor of stock status in the model. See Table S5 for a complete list of candidate variables. These candidate predictor variables are intended to reflect the kinds of catch history traits and life history variables commonly included when constructing priors for catch-only models. In theory, if a strong predictive relationship between these types of covariates and B/B_{MSY} exists, a flexible model such as a boosted regression tree should find it.

Boosted regression trees have a number of tuning parameters that must be set outside of the model fitting process itself. We performed a grouped v -fold cross-validation routine on the training dataset to tune these parameters, splitting the training dataset into five analysis and assessment splits, each of which assigns non-overlapping sets of stocks to the analysis and assessment splits. We then fit a grid of candidate tuning parameters to the analysis splits and used them to predict the assessment splits. The set of tuning parameters with the lowest root mean squared error (RMSE) on the assessment splits was selected for use for model fitting. We then fit the model to the complete training set using the selected tuning parameters and then used the fitted model to predict the B/B_{MSY} values in the testing set (containing only stocks from New Zealand, Australia and South Africa). As a secondary test, we also conducted a leave-one-out cross-validation test on only the training data (holding the tuning parameters at their selected values), sequentially

removing one stock from the training set, fitting the model on the remaining training stocks, predicting the omitted stock, then moving on to the next stock.

Catch-only models depend on the assumption that there is a connection between stock status and catch histories. Active management of catches clearly distorts this relationship. To account for this, we ran a secondary test, repeating the above process, but restricting the model to data before 1990. As strict fisheries management was rarer in these years, this provides a test of whether a stronger predictive relationship exists between catch and stock status in years with fewer catch restrictions.

The R^2 of the model on the training dataset was 0.7, with an RMSE of 0.5. However, unlike random forests, boosted regression trees do not automatically produce out-of-bag predictions for the training data. Therefore, this relatively good fit is not a true measure of the predictive power of the model as the stocks being predicted were also included in some of the model fits. The leave-one-out analysis provides an honest assessment of the predictive power of the model on the stocks in the training dataset, producing an R^2 of 0.38 and an RMSE of 0.72. Fits for the testing dataset were slightly poorer (R^2 of 0.28 and an RMSE of 0.7), but not by much, suggesting that our tuning process did a reasonable job of selecting parameter values that prevented overfitting (Figure 4). Results for fits to the pre-1990 data were similar, see Figure S3.

In our analysis, a model using only life history and characteristics of the catch history, in the manner of Costello et al. (2012), was able to explain roughly 30% of the variance, a value higher than reported in Costello et al., (reported fits in that paper are on log-scale, and so will be worse than the reported values when converted to natural scale as evaluated here), and in line with the values in Zhou

et al. (2017). In both the leave-one-out and testing fits, the model underestimates stock status for stocks with B/B_{MSY} greater than one and overestimates it for stocks with B/B_{MSY} less than one. The RMSE of this model is high enough to frequently assign a stock to the wrong general status bin (e.g. classify as under fished when it is overfished and *vice versa*).

Clearly, further work could be done to refine performance. However, in our experience, these refinements (e.g. alternative model types and feature engineering) can produce some improvements, but rarely produce transformative increases in model performance. We suggest then that the levels of predictive power seen here and in Zhou et al. (2017) are likely close to the upper level of performance for empirical models tasked with predicting B/B_{MSY} based on estimates from the RLSADB. This predictive performance is not without use, showing some ability to differentiate between stocks that are more likely to be over-versus-under fished, but is clearly not capable of providing precise or accurate predictions for individual stocks.

While it is certainly possible that someone could come up with a better model parameterization, given that this result fits firmly in line with the types of performances seen of other catch-only models when tested in Free et al. (2020), we feel that substantial increases in out-of-sample predictive power of empirical catch-based models are simply not there to be found. This analysis demonstrates that, while in theory predictive empirical relationships could be found between catch and life history and stock status, we see no evidence that improved performance beyond the levels observed here is likely to exist given currently available data. It should also be noted that what evidence is present will degrade as soon as management measures have been put in place in response to a catch-only model, limiting the ability of predictive catch-only models to serve as a basis for continued assessment and responsive management.

6 | CONCLUSIONS

Catch histories are the most comprehensive source of global data on fisheries production, spanning both industrial and artisanal fisheries across a wide range of economic and ecological conditions (FAO, 2020), although it is widely recognized that small-scale fisheries are underrepresented in the data reported to FAO (Kelleher et al., 2012), and the quality and taxonomic resolution of the data is highly variable. These data are an important tool for helping us understand the role of global fisheries in human wellbeing and anthropogenic impacts. Catches are a function of the size of fish populations, and so contain information on the minimum size of stocks. However, based on first principles and numerous efforts at replication, it is clear that catch-only models are not a consistently reliable means of either inferring or predicting the state of fished populations.

The basic catch equation (Equation 1) shows that one cannot solve for biomass without making strong assumptions about factors such as catchability and effort. There certainly are fisheries in which both

effort and catchability have been sufficiently constant or predictable as to facilitate inference of stock status based on catches alone. However, in most fisheries, catches change for a wide number of reasons, including environmental forcing, economic incentives, technological advances and fisheries regulations (Hilborn & Branch, 2013). Each of these factors can break the relationship between stock status and catch. Our results show that the shape of the catch history itself is uninformative on the state of a fishery, conditional on life history priors and the mean volume of the catch (Figure 2).

Efforts at predicting stock status based on catch histories reveal a consistent pattern of some but limited predictive power, often with out-of-sample R^2 less than 0.35, and substantial potential for bias (Anderson et al., 2017; Bouch et al., 2021; Free et al., 2020; Ovando et al., 2021; Pons et al., 2020). These values may be higher for cases where the fishery to be estimated closely resembles those included in the training dataset, but predictions are likely to be particularly poor for fisheries that look little like the highly managed stocks on which predictive models can be trained. This is certainly the case for the other half of the world's fisheries where catch-only models are most likely to be needed or applied.

This is not to say that catch-only models are without value. The original formulation of CMSY (Martell & Froese, 2013) was designed not to estimate stock status but to estimate MSY itself. There is ample evidence and logic supporting the idea that for fisheries with catches that have ever come close to or surpassed MSY, the catch history can provide a plausible guide to the magnitude of this important reference point. In terms of stock status, to the extent that there is good reason to think that a particular stock can be represented with a surplus production model, if users are able to set reasonably accurate priors on population parameters such as r and K , then filtering these prior distributions through the lens of the population model combined with the catch history will indeed provide an accurate estimate of stock status. However, the focus then must be on understanding the critical importance of setting these life history priors accurately, and not on assuming that somehow the combination of the priors along with the shape of the catch history can provide information not encoded in the priors themselves.

Catch-only models are also likely to work better in some systems than others. However, evaluating catch-only models does present a bit of a paradox. The only fisheries for which we are likely to have robust data-rich empirical estimates of stock status will tend to be highly managed fisheries, exactly the cases where we would expect the weakest relationships between catch and biomass due to active management and strong market forcing. This means that empirical evaluation of catch-only models may underestimate their potential performance in less managed fisheries with perhaps a clearer link between catch and biomass. However, our analysis evaluating the predictive power of an empirical catch-only model using only pre-1990 data, when fisheries management was less intense in assessed fisheries, suggests that this gap may be small (Figure S3). The only alternative to empirical testing of catch-only models is simulation testing. While simulation testing can be designed to reflect a wider range of fishery dynamics than those represented in databases such as the

RLSADB, simulated fishery dynamics can also be overly simplistic, and simulation tests often conform more to the assumed dynamics of catch-only models than reality will for the sake of computational efficiency (e.g. use of surplus production dynamics for both the operating and estimation model).

The critical need then is more robust estimates of the expected performance of catch-only models under a range of plausible circumstances, based both on confrontation with empirical observations and evaluation through simulation testing. Users can then make informed judgements as to whether the expected performance of catch-only models for their particular application are likely to be sufficient for their needs, or whether the expected imprecision and bias are simply too great for model outputs to be useful as an index of stock status. Where catch-only models are still selected for use as a source of stock status, it is critical that users are empowered with an understanding that estimates of stock status based on SRA-style catch-only models are purely a function of the accuracy of priors on stock status combined with the accuracy of required life history priors: There is no way for the shape of a catch history to “overcome” poor priors and infer the true state of a stock except by chance alone. For more empirical predictive models, the best available evidence suggests that while there is some predictive power in catch histories, it is limited and highly sensitive to the resemblance of the fisheries it is applied to compare the fisheries the models were trained on.

Our concern is that catch-only models can promise an off-the-shelf method to assess fish stocks that can and has been used without clear communication or understanding of their limitations. Despite increasingly being used as a tool informing direct management decisions, it is clear from both empirical and simulation testing that off-the-shelf use of catch-only models is not a reliable means of estimating stock status, and that a key to better global fisheries assessment and management is not wider application of catch-only models or model refinement, but rather expanded collection and curation of diverse data sources that contain meaningful information on the state of fisheries.

ACKNOWLEDGEMENTS

We thank two anonymous reviewers for their helpful revisions and comments.

CONFLICT OF INTEREST

RH receives research funding from many groups that have interests in fisheries outcomes including environmental NGOs, foundations, governments and fishing industry groups.

AUTHOR CONTRIBUTION

C.F. and D.O. conducted all analyses. All authors contributed to the conceptualization and writing of the manuscript.

DATA AVAILABILITY STATEMENT

All code and data needed to replicate this study are publicly available at <https://github.com/DanOvando/eval-coms>.

ORCID

Daniel Ovando  <https://orcid.org/0000-0003-2120-7345>

Christopher M. Free  <https://orcid.org/0000-0002-2557-8920>

Olaf P. Jensen  <https://orcid.org/0000-0001-7850-6616>

Ray Hilborn  <https://orcid.org/0000-0003-2326-2305>

REFERENCES

- Anderson, S. C., Cooper, A. B., Jensen, O. P., Minto, C., Thorson, J. T., Walsh, J. C., Afflerbach, J., Dickey-Collas, M., Kleisner, K. M., Longo, C., Osio, G. C., Ovando, D., Mosqueira, I., Rosenberg, A. A., & Selig, E. R. (2017). Improving estimates of population status and trend with superensemble models. *Fish and Fisheries*, 18(4), 732–741. <https://doi.org/10.1111/faf.12200>
- Barman, P. P., Karim, E., Khatun, M. H., Rahman, M. F., Alam, M. S., & Liu, Q. (2020). Application of CMSY to estimate biological reference points of bombay duck (*Harpadon Neherus*) from the Bay of Bengal Bangladesh. *Applied Ecology and Environmental Research*, 18(6), 8023–8034. https://doi.org/10.15666/aeer/1806_80238034
- Bockstoce, J. R., & Botkin, D. B. (1983). The historical status and reduction of the western Arctic bowhead whale (*Balaena mysticetus*) population by the pelagic whaling industry, 1848–1914 (No. 5; pp. 107–141). Report of the International Whaling Commission.
- Bouch, P., Minto, C., & Reid, D. G. (2021). Comparative performance of data-poor CMSY and data-moderate SPiCT stock assessment methods when applied to data-rich, real-world stocks. *ICES Journal of Marine Science*, 78(1), 264–276. <https://doi.org/10.1093/icesjms/fsaa220>
- Branch, T. A., Jensen, O. P., Ricard, D., Ye, Y., & Hilborn, R. (2011). Contrasting global trends in marine fishery status obtained from catches and from stock assessments. *Conservation Biology*, 25(4), 777–786. <https://doi.org/10.1111/j.1523-1739.2011.01687.x>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Caddy, J. F., & Gulland, J. A. (1983). Historical patterns of fish stocks. *Marine Policy*, 7(4), 267–278. [https://doi.org/10.1016/0308-597X\(83\)90040-4](https://doi.org/10.1016/0308-597X(83)90040-4)
- Costello, C., Ovando, D., Clavelle, T., Strauss, C. K., Hilborn, R., Melnychuk, M. C., Branch, T. A., Gaines, S. D., Szuwalski, C. S., Cabral, R. B., Rader, D. N., & Leland, A. (2016). Global fishery prospects under contrasting management regimes. *Proceedings of the National Academy of Sciences*, 113(18), 5125–5129. <https://doi.org/10.1073/pnas.1520420113>
- Costello, C., Ovando, D., Hilborn, R., Gaines, S. D., Deschenes, O., & Lester, S. E. (2012). Status and solutions for the world's unassessed fisheries. *Science*, 338(6106), 517–520. <https://doi.org/10.1126/science.1223389>
- Cury, P. M., Fromentin, J., Figuet, S., & Bonhommeau, S. (2014). Resolving Hjort's Dilemma: How Is Recruitment Related to Spawning Stock Biomass in Marine Fish? *Oceanography*, 27(4), 42–47. <http://www.jstor.org/stable/24862211>
- Cushing, D. H. (1982). *Climate and fisheries*. Academic press.
- Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), 745–766. <https://doi.org/10.1080/10618600.2017.1384734>
- FAO. (2020). State of world fisheries and aquaculture 2020: sustainability in action. Food & Agriculture org.
- Free, C. M., Jensen, O. P., Anderson, S. C., Gutierrez, N. L., Kleisner, K. M., Longo, C., Minto, C., Osio, G. C., & Walsh, J. C. (2020). Blood from a stone: Performance of catch-only methods in estimating stock biomass status. *Fisheries Research*, 223, 105452. <https://doi.org/10.1016/j.fishres.2019.105452>
- Free, C. M., Jensen, O. P., Wiedenmann, J., & Deroba, J. J. (2017). The refined ORCS approach: A catch-based method for estimating stock

- status and catch limits for data-poor fish stocks. *Fisheries Research*, 193, 60–70. <https://doi.org/10.1016/j.fishres.2017.03.017>
- Froese, R., Demirel, N., Coro, G., Kleisner, K. M., & Winker, H. (2017). Estimating fisheries reference points from catch and resilience. *Fish and Fisheries*, 18(3), 506–526. <https://doi.org/10.1111/faf.12190>
- Froese, R., & Kesner-Reyes, K. (2002). Impact of fishing on the abundance of marine species. *ICES CM*, 50, 12.
- Froese, R., Winker, H., Coro, G., Demirel, N., Tsikliras, A. C., Dimarchopoulou, D., Scarcella, G., Quaas, M., & Matz-Lück, N. (2018). Status and rebuilding of European fisheries. *Marine Policy*, 93, 159–170. <https://doi.org/10.1016/j.marpol.2018.04.018>
- Froese, R., Zeller, D., Kleisner, K., & Pauly, D. (2012). What catch data can tell us about the status of global fisheries. *Marine Biology*, 159(6), 1283–1292. <https://doi.org/10.1007/s00227-012-1909-6>
- Garcia, S. M., & Grainger, R. J. R. (2005). Gloom and doom? The future of marine capture fisheries. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1453), 21–46. <https://doi.org/10.1098/rstb.2004.1580>
- Grainger, R. J., & Garcia, S. M. (1996). *Chronicles of marine fishery landings (1950–1994): Trend analysis and fisheries potential*. FAO.
- Harford, W. J., & Carruthers, T. R. (2017). Interim and long-term performance of static and adaptive management procedures. *Fisheries Research*, 190, 84–94. <https://doi.org/10.1016/j.fishres.2017.02.003>
- Hilborn, R. W. (2007). Biodiversity loss in the ocean: How bad is it? *Science (New York, NY)*, 316(5829), 1281–1284.
- Hilborn, R., Amoroso, R. O., Anderson, C. M., Baum, J. K., Branch, T. A., Costello, C., de Moor, C. L., Faraj, A., Hively, D., Jensen, O. P., Kurota, H., Little, L. R., Mace, P., McClanahan, T., Melnychuk, M. C., Minto, C., Osio, G. C., Parma, A. M., Pons, M., ... Ye, Y. (2020). Effective fisheries management instrumental in improving fish stock status. *Proceedings of the National Academy of Sciences*, 117(4), 2218–2224. <https://doi.org/10.1073/pnas.1909726116>
- Hilborn, R., & Branch, T. A. (2013). Fisheries: Does catch reflect abundance? No, it is misleading. *Nature*, 494(7437), 303–306. <https://doi.org/10.1038/494303a>
- Hilborn, R., & Walters, C. J. (1992). *Quantitative fisheries stock assessment*. Springer. <https://doi.org/10.1007/978-1-4615-3598-0>
- Kell, L. T., Mosqueira, I., Grosjean, P., Fromentin, J.-M., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M. A., & Poos, J. J. (2007). FLR: An open-source framework for the evaluation and development of management strategies. *ICES Journal of Marine Science*, 64(4), 640–646. <https://doi.org/10.1093/icesjms/fsm012>
- Kelleher, K., Westlund, L., Hoshino, E., Mills, D., Willmann, R., de Graaf, G., & Brummett, R. (2012). Hidden harvest: The global contribution of capture fisheries. Worldbank; WorldFish.
- Kimura, D. K., Balsiger, J. W., & Ito, D. H. (1984). Generalized stock reduction analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 41(9), 1325–1333. <https://doi.org/10.1139/f84-162>
- Kimura, D. K., & Tagart, J. V. (1982). Stock reduction analysis, another solution to the catch equations. *Canadian Journal of Fisheries and Aquatic Sciences*, 39(11), 1467–1472. <https://doi.org/10.1139/f82-198>
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer. <https://doi.org/10.1007/978-1-4614-6849-3>
- Martell, S., & Froese, R. (2013). A simple method for estimating MSY from catch and resilience. *Fish and Fisheries*, 14(4), 504–514. <https://doi.org/10.1111/j.1467-2979.2012.00485.x>
- Melnichuk, M. C., Peterson, E., Elliott, M., & Hilborn, R. (2017). Fisheries management impacts on target species status. *Proceedings of the National Academy of Sciences*, 114(1), 178–183. <https://doi.org/10.1073/pnas.1609915114>
- Murawski, S., Methot, R., Tromble, G., Hilborn, R. W., Briggs, J. C., Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., & Folke, C. (2007). Biodiversity Loss in the Ocean: How Bad Is It? *Science*, 316(5829), 1281–1284. <https://doi.org/10.1126/science.316.5829.1281b>
- Ovando, D., Hilborn, R., Monnahan, C., Rudd, M., Sharma, R., Thorson, J. T., Rousseau, Y., & Ye, Y. (2021). Improving estimates of the state of global fisheries depends on better data. *Fish and Fisheries*, 22(6), 1377–1391. <https://doi.org/10.1111/faf.12593>
- Palomares, M. L. D., Froese, R., Derrick, B., Meeuwig, J. J., Nöel, S.-L., Tsui, G., Woroniak, J., Zeller, D., & Pauly, D. (2020). Fishery biomass trends of exploited fish populations in marine ecoregions, climatic zones and ocean basins. *Estuarine, Coastal and Shelf Science*, 243, 106896. <https://doi.org/10.1016/j.ecss.2020.106896>
- Pauly, D. (2007). The sea around us project: Documenting and communicating global fisheries impacts on marine ecosystems. *AMBIO: A Journal of the Human Environment*, 36(4), 290–295.
- Pella, J. J., & Tomlinson, P. K. (1969). A generalized stock production model. *Inter-American Tropical Tuna Commission Bulletin*, 13(3), 416–497.
- Pons, M., Cope, J. M., & Kell, L. T. (2020). Comparing performance of catch-based and length-based stock assessment methods in data-limited fisheries. *Canadian Journal of Fisheries and Aquatic Sciences*. <https://doi.org/10.1139/cjfas-2019-0276>
- Pons, M., Melnychuk, M. C., & Hilborn, R. (2018). Management effectiveness of large pelagic fisheries in the high seas. *Fish and Fisheries*, 19(2), 260–270. <https://doi.org/10.1111/faf.12253>
- Poole, D., & Raftery, A. E. (2000). Inference for deterministic simulation models: The Bayesian melding approach. *Journal of the American Statistical Association*, 95(452), 1244–1255. <https://doi.org/10.1080/01621459.2000.10474324>
- Punt, A. E., Butterworth, D. S., de Moor, C. L., Oliveira, J. A. A. D., & Haddon, M. (2016). Management strategy evaluation: Best practices. *Fish and Fisheries*, 17(2), 303–334. <https://doi.org/10.1111/faf.12104>
- Ricard, D., Minto, C., Jensen, O. P., & Baum, J. K. (2012). Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish and Fisheries*, 13(4), 380–398. <https://doi.org/10.1111/j.1467-2979.2011.00435.x>
- Rosenberg, A. A., Fogarty, M. J., Cooper, A. B., Dickey-Collas, M., Fulton, E. A., Gutiérrez, N. L., Hyde, K. J., Kleisner, K. M., Kristiansen, T., & Longo, C. (2014). Developing new approaches to global stock status assessment and fishery production potential of the seas. *FAO Fisheries and Aquaculture Circular*, 1086, 1.
- Rosenberg, A. A., Kleisner, K. M., Afflerbach, J., Anderson, S. C., Dickey-Collas, M., Cooper, A. B., Fogarty, M. J., Fulton, E. A., Gutiérrez, N. L., Hyde, K. J. W., Jardim, E., Jensen, O. P., Kristiansen, T., Longo, C., Minto-Vera, C. V., Minto, C., Mosqueira, I., Osio, G. C., Ovando, D., ... Ye, Y. (2018). Applying a new ensemble approach to estimating stock status of marine fisheries around the world. *Conservation Letters*, 11(1), e12363. <https://doi.org/10.1111/conl.12363>
- Schaefer, M. B. (1954). Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Inter-American Tropical Tuna Commission Bulletin*, 1(2), 23–56.
- Sharma, R., Winker, H., Levontin, P., Kell, L., Ovando, D., Palomares, M. L. D., Pinto, C., & Ye, Y. (2021). Assessing the potential of catch-only models to inform on the state of global fisheries and the UN's SDGs. *Sustainability*, 13(11), 6101. <https://doi.org/10.3390/su13116101>
- Smith, D. C., Haddon, M., Punt, A. E., Gardner, C., Little, L. R., Mayfield, S., O'Neill, M. F., Saunders, T., Stewart, J., Wise, B., Fulton, E. A., & Conron, S. (2021). Evaluating the potential for an increased and sustainable commercial fisheries production across multiple jurisdictions and diverse fisheries. *Marine Policy*, 124, 104353. <https://doi.org/10.1016/j.marpol.2020.104353>
- Szuwalski C. S., Vert-Pre K. A., Punt A. E., Branch T. A., Hilborn R. (2015). Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. *Fish and Fisheries*, 16(4), 633–648. <http://dx.doi.org/10.1111/faf.12083>
- Thorson, J. T. (2020). Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated

- life-history model. *Fish and Fisheries*, 21(2), 237–251. <https://doi.org/10.1111/faf.12427>
- Thorson, J. T., Branch, T. A., & Jensen, O. P. (2012). Using model-based inference to evaluate global fisheries status from landings, location, and life history data. *Canadian Journal of Fisheries and Aquatic Sciences*, 69(4), 645–655. <https://doi.org/10.1139/f2012-016>
- Thorson, J. T., Minto, C., Minte-Vera, C. V., Kleisner, K. M., & Longo, C. (2013). A new role for effort dynamics in the theory of harvested populations and data-poor stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, 70(12), 1829–1844. <https://doi.org/10.1139/cjfas-2013-0280>
- Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, 102(6), e03336. <https://doi.org/10.1002/ecy.3336>
- Vasconcellos, M., & Cochrane, K. (2005). Overview of world status of data-limited fisheries: Inferences from landings statistics. In G. H. Kruse, V. F. Gallucci, D. E. Hay, R. I. Perry, R. M. Peterman, T. C. Shirley, P. D. Spencer, & B. Wilson (Eds.), *Fisheries assessment and management in data-limited situations* (pp. 1–20). Alaska Sea Grant, University of Alaska Fairbanks. <https://doi.org/10.4027/famdl.2005.01>
- Vert-pre, K. A., Amoroso, R. O., Jensen, O. P., & Hilborn, R. (2013). Frequency and intensity of productivity regime shifts in marine fish stocks. *Proceedings of the National Academy of Sciences*, 110(5), 1779–1784. <https://doi.org/10.1073/pnas.1214879110>
- Walsh, J. C., Minto, C., Jardim, E., Anderson, S. C., Jensen, O. P., Afflerbach, J., Dickey-Collas, M., Kleisner, K. M., Longo, C., Osio, G. C., Selig, E. R., Thorson, J. T., Rudd, M. B., Papacostas, K. J., Kittinger, J. N., Rosenberg, A. A., & Cooper, A. B. (2018). Trade-offs for data-limited fisheries when using harvest strategies based on catch-only models. *Fish and Fisheries*, 19(6), 1130–1146. <https://doi.org/10.1111/faf.12316>
- Walters, C. J., Martell, S. J. D., & Korman, J. (2006). A stochastic approach to stock reduction analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 63(1), 212–223. <https://doi.org/10.1139/f05-213>
- Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., Jackson, J. B. C., Lotze, H. K., Micheli, F., Palumbi, S. R., Sala, E., Selkoe, K. A., Stachowicz, J. J., & Watson, R. (2006). Impacts of Biodiversity Loss on Ocean Ecosystem Services. *Science*, 314(5800), 787–790. <https://doi.org/10.1126/science.1132294>
- Worm, B., Hilborn, R., Baum, J. K., Branch, T. A., Collie, J. S., Costello, C., Fogarty, M. J., Fulton, E. A., Hutchings, J. A., Jennings, S., Jensen, O. P., Lotze, H. K., Mace, P. M., McClanahan, T. R., Minto, C., Palumbi, S. R., Parma, A. M., Ricard, D., Rosenberg, A. A., ... Zeller, D. (2009). Rebuilding Global Fisheries. *Science*, 325(5940), 578–585. <https://doi.org/10.1126/science.1173146>
- Zhou, S., Punt, A. E., Smith, A. D. M., Ye, Y., Haddon, M., Dichmont, C. M., & Smith, D. C. (2018). An optimized catch-only assessment method for data poor fisheries. *ICES Journal of Marine Science*, 75(3), 964–976. <https://doi.org/10.1093/icesjms/fsx226>
- Zhou, S., Punt, A. E., Ye, Y., Ellis, N., Dichmont, C. M., Haddon, M., Smith, D. C., & Smith, A. D. (2017). Estimating stock depletion level from patterns of catch history. *Fish and Fisheries*, 18(4), 742–751. <https://doi.org/10.1111/faf.12201>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Ovando, D., Free, C. M., Jensen, O. P., & Hilborn, R. (2022). A history and evaluation of catch-only stock assessment models. *Fish and Fisheries*, 23, 616–630. <https://doi.org/10.1111/faf.12637>