The refined ORCS approach: a catch-based method for estimating stock status and catch limits for data-poor fish stocks

Christopher M. Free[1]*, Olaf P. Jensen[1], John Wiedenmann[2], Jonathan J. Deroba[3]

[1] *Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA*

[2] *Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, NJ, USA*

[3] *NOAA Fisheries, Northeast Fisheries Science Center, Woods Hole, MA, USA*

*** Corresponding author:** Department of Marine and Coastal Sciences, Rutgers University, 71 Dudley Road, New Brunswick, NJ 08901, USA. Email: cfree@marine.rutgers.edu

[1] **ORCS =** only reliable catch stocks; **TOA =** table of attributes; **RAMLDB** = RAM Legacy Stock Assessment Database; **BCT** = boosted classification trees

Highlights

- The ORCS approach to data-poor catch limit estimation has yet to be fully evaluated
- We evaluated and refined the ORCS approach by applying it to 193 data-rich stocks
- The original approach is a poor predictor of stock status and should not be used
- The refined approach performs better than the original and other data-poor methods
- With conservative catch scalars, the refined approach can estimate useful OFLs

ABSTRACT

The ‘Only Reliable Catch Stocks’ (ORCS) Working Group approach to data-poor fisheries stock status and catch limit estimation has been used by U.S. fisheries managers but has yet to be fully evaluated. The ORCS approach estimates stock status using a fourteen question ‘Table of Attributes’ and the overfishing limit by multiplying a historical catch statistic by a scalar based on the estimated status. We evaluated the performance of the approach by applying it to 193 stocks with data-rich stock assessments and comparing its predictions of stock status with the assessment model estimates. The approach classified all but three stocks as fully exploited indicating that it is a poor predictor of status and should not be used by managers. We refined the original ORCS approach by: (1) developing a more predictive model of stock status using boosted classification trees and (2) identifying the historical catch statistics and scalars that best estimate overfishing limits using assessment model data. The refined ORCS approach correctly classified 74% of all stocks and 62% of overexploited stocks in a training dataset and 74% of all stocks and 50% of overexploited stocks in an independent test dataset. The refined approach performed better than other widely used catch-only methods. However, the overfishing limits estimated by the refined approach would further deplete overexploited stocks without the use of conservative catch scalars to buffer against classification uncertainty. Conservative catch scalars can reduce the probability of overfishing below 50%, the U.S. legal maximum, but with concomitant increases in the probability and magnitude of underfishing. The refined ORCS approach may therefore be useful when other methods are not possible or appropriate and some risk of underfishing is acceptable.

Keywords: stock assessment, data-limited fisheries, only reliable catch stocks, catch-only

methods, RAM Legacy Stock Assessment Database

1. INTRODUCTION

The majority of global fish stocks lack adequate data for estimating sustainable fishing levels using conventional stock assessment methods. In developing countries, only 5-20% of fish stocks are assessed and this fraction increases to only 10-50% in developed countries (Costello et al., 2012). In the United States, 30% of stocks are managed using conventional 'data-rich' assessment methods, while the remaining 11% and 59% of stocks are managed using 'data-moderate' and 'data-poor' methods, respectively (Newman et al., 2015). Data-rich stock assessment methods combine (1) total catch; (2) an index of relative abundance; and (3) other biological information to assess stock status and estimate sustainable fishing levels (Walters and Martell, 2004). Data-poor and data-moderate methods generally utilize only one and two of these data types, respectively, with total catch information often being the only data type available. Thus, data-poor methods are often synonymous with catch-only methods.

In 2006, the U.S. Magnuson-Stevens Fishery Conservation and Management Act was amended to require scientifically-derived annual catch limits (ACLs) that prevent overfishing for all federally managed fish stocks, including data-limited stocks (DOC, 2007). This mandate stimulated the revival of old data-limited methods (Gulland, 1971; Restrepo et al., 1998), development of new data-limited methods (MacCall, 2009; Dick and MacCall, 2011; Cope, 2013; Cope et al., 2013), and evaluation of the relative performance of these methods (Wetzel and Punt, 2011; Wiedenmann et al., 2013; Carruthers et al., 2014). In 2011, the 'Only Reliable Catch Stocks' (ORCS) Working Group (Berkson et al., 2011) convened to evaluate catch-only methods for ACL estimation and recommended the following hierarchy for determining ACLs for ORCS: (1) depletion-based stock reduction analysis (DB-SRA; Dick and MacCall, 2011) when a complete time series of annual catches is available (i.e., from the start of fishing to the present); (2) depletion-corrected average catch (DCAC; MacCall, 2009) when the stock exhibits low natural mortality rates (≤ 0.20 yr^{-1}); and (3) the new ORCS Working Group approach (hereafter called the 'ORCS approach') when neither DB-SRA or DCAC are possible or appropriate (Berkson et al., 2011; later modified by SAFMC, 2012, 2013).

The ORCS approach was designed to provide an ecological basis for the Restrepo et al. (1998) scalar approach. In both methods, the overfishing limit (OFL; the catch at F_{MSY}) is calculated by multiplying an expert-defined historical catch statistic (e.g., mean catch over the previous 10 years or median catch over the whole time series) by a scalar also based on expert judgment. In the ORCS approach, the choice of scalar is determined by stock status (i.e., under, fully, or overexploited), which is estimated as the mean score of fourteen stock- and fishery-related attributes (the 'Table of Attributes' or TOA; **Table 1**). The ORCS approach allows for considerable flexibility in its implementation, as scientists and managers can exercise expert judgement to: (1) estimate status using an arithmetic, geometric, or weighted mean of the Table of Attributes scores; (2) modify the Table of Attributes' estimate of status or the thresholds used to delineate status; and/or (3) choose appropriate catch statistics and scalars. While this flexibility and reliance on expert judgement could improve performance, it is necessary to adopt a more specific, albeit less inclusive, definition of the ORCS approach to validate the method and demonstrate its transferability.

The ORCS approach is widely applicable, but the ability of the Table of Attributes to correctly predict stock status has not been evaluated and the performance of only a limited range of potential catch statistics and scalars has been tested. In the only explicit evaluation of the ORCS approach to date, Wiedenmann et al. (2013) used management strategy evaluation to show that the default scalars used to estimate the OFL are too conservative for under (scalar=0.5) and fully (scalar=1.0) exploited stocks and too generous for overexploited (scalar=2.0) stocks when stock status is correctly classified. They also show that catch limits are unsustainable when stocks are incorrectly classified into less-depleted categories (e.g., an overexploited stock incorrectly classified as fully exploited). Evaluations of scalar-based methods similar to the ORCS approach have also been shown to result in overfishing, especially for already depleted stocks and stocks whose statuses have been incorrectly classified (Carruthers et al. 2014; ICES 2014, 2015, 2017). The sensitivity of management outcomes to status classification decisions makes the validation and refinement of the ORCS Table of Attributes' ability to estimate status necessary before the ORCS approach can be used to set catch-limits more widely.

The goals of the present study are to evaluate and refine the ORCS approach to data-poor catch limit estimation using stocks with data-rich stock assessments. We evaluate the original approach by applying it to data-rich stocks and comparing its predictions of status with the assessment model estimates. We refine the ORCS approach by: (1) developing a more predictive model of stock status that uses boosted classification trees to weight attributes by their relative importance, incorporate interactions between attributes, and account for non-linearity in attribute behavior; and (2) empirically identifying the best status-specific historical catch statistics and scalars for estimating overfishing limits using assessment model data. Finally, we evaluate the ability of the refined ORCS approach to estimate overfishing limits and compare the ability of the refined approach to estimate stock status to six other catch-only assessment methods.

2. METHODS

2.1 Stock selection

We evaluated the ORCS approach to data-poor catch limit estimation by applying it to data-rich stocks with stock assessments based on underlying population dynamics models (generally statistical catch-at-age models, virtual population analyses, and production models) in the RAM Legacy Stock Assessment Database (RAMLDB v.2.95; Ricard et al., 2012). We used only stocks with assessments that estimate B_{MSY} internal to the model or estimate standard proxies used by the management agency (e.g., spawning potential ratio proxies common in the U.S. or B_0 proxies common in Australia). We excluded stocks whose assessments are considered particularly unreliable (n=2; 2002 Atlantic croaker and 2005 Atlantic herring). The resulting 193 stocks include underexploited (n=68), fully exploited (n=95), and overexploited (n=30) stocks representing a variety of taxa, geographic locations, and management agencies (**Figure 1**). The RAMLDB does not include the most up-to-date assessment for every stock. Therefore, data-rich statuses and answers to the Table of Attributes questions reflect the terminal year of the assessment in the RAMLDB.

2.2 Evaluation of the ORCS Table of Attributes

We estimated stock status using the expanded Table of Attributes developed by SAFMC (2012) with a few modifications to increase clarity and objectivity in the scoring process (**Table 1**; **Supplementary Appendix A.1**). We scored: *TOA #1 Status of assessed stocks in fishery* using U.S. Fisheries Management Plans and their foreign analogs to identify groups of stocks managed together and references from management agencies to determine the status of these stocks; *TOA #2 Refuge availability*, *#3 Behavior affecting capture*, *#4 Morphology affecting capture*, and *#11 Habitat loss* using information on the distribution, biology, and habitat of the taxa in FishBase (Froese and Pauly, 2016); *TOA #5 Discard rate*, *#6 Targeting intensity*, *#7 M compared to dominant species*, *#8 Occurrence in catch,* and *#14 Proportion of population protected* using information in the stock assessment documents; *TOA #9 Value* using ex-vessel price data from the Sea Around Us Project (Pauly and Zeller, 2015); and *TOA #10 Recent trend in catch*, *#12 effort*, and *#13 abundance index* using time series in the RAMLDB. Other technical sources (i.e., government reports or websites, peer-reviewed scientific papers, technical reports) were used when an attribute could not be scored using the principal reference. In some cases, attributes could not be scored due to a lack of data or applicability and were given an 'NA' value. Detailed information on the scoring process is available in **Supplementary Appendices A.2** and **A.3** and the scores and their justifications are available in **Supplementary Appendix B.** Estimated stock status was determined from the mean of the Table of Attributes scores with the following classifications provided by the original method: underexploited (<1.5), fully exploited (1.5–2.5), and overexploited (>2.5). This simplification of the broadly flexible ORCS approach is necessary for testing and validating the performance of the method on such a diverse and global array of stocks.

The ORCS approach has been thought to estimate both stock status (i.e., lightly, moderately, and heavily exploited; Berkson et al. 2011) and the risk of overexploitation (i.e., low, moderate, and high risk of overexploitation; SAFMC, 2012, 2013). Consequently, we evaluated the performance of the original approach using linear regression to assess the correlation between predicted status (mean Table of Attributes score) and the assessment's most recent estimates of (1) B/B_{MSY} as a proxy for stock status and (2) F/F_{MSY}

as a proxy for overexploitation risk. We also assessed the ability of the original approach to correctly classify stock status using both percentage agreement (accuracy) and Cohen's kappa. Cohen's kappa measures inter-rate agreement between categorical items and is more robust than simple percentage agreement because it takes into account the probability of agreement occurring by chance (Cohen, 1968). This metric was preferred given the volume and ease of identifying fully exploited stocks compared to the paucity and difficulty of identifying overexploited stocks. If the method misclassifies most overexploited stocks but correctly classifies most fully exploited stocks, then it would still earn a high accuracy percentage, but it's kappa value would be appropriately penalized. Although there are no definitive rules for interpreting Cohen's kappa, general guidelines suggest that values >0.70 are 'excellent', 0.4-0.7 are 'good', 0.2-0.4 are 'fair', and <0.2 are 'poor' (Landis and Koch, 1977; Fleiss, 1981).

2.3 Refinement of the ORCS Table of Attributes

We refined the ORCS Table of Attributes using boosted classification trees (BCT) to weight attributes by their relative importance, incorporate interactions between attributes, and account for non-linearity in attribute behavior. Boosted classification trees combine classification and machine learning and offer predictive power superior to other modeling methods (Elith et al., 2008). Boosted classification trees can also accommodate missing values (i.e., NA scores) by imputing values from surrogate variables, which allowed the use of all scored stocks. The BCT analysis was performed using the *caret* (Kuhn, 2016) and *gbm* (Ridgeway, 2016) packages in R v.3.3.2 (R Core Team, 2016).

We trained the BCT model to estimate categorical status (i.e., under, fully, or overexploited) rather than continuous status (i.e., B/B_{MSY}) because (1) the ORCS approach was designed to use status categories and (2) stock assessment models exhibit more uncertainty in estimates of B/B_{MSY} than in more general status classifications. We trained the BCT model to estimate stock status rather than risk of overexploitation because (1) stock status is a more widely used metric and can be easily compared to other assessment methods and (2) F/F_{MSY} is an unsatisfying proxy for overexploitation risk because it can change rapidly and even sustained F/F_{MSY} values greater than 1.0 may not be "risky" over the short-term if B/B_{MSY} is high ($\gg$1.0). The BCT model attempts to determine stock status – whether a stock is under (B/B_{MSY}>1.5), fully (B/B_{MSY}=0.5–1.5), or overexploited

(B/B_{MSY}<0.5) – from the TOA scores with a few modifications (**Table 1**): (1) we removed *TOA #2 Refuge availability* and *#4 Morphology affecting capture* because they lacked contrast (i.e., 97.9% and 100% of stocks were assigned scores of 3–>75% of habitat accessible and 2–Average susceptibility, respectively); (2) we used continuous rather than categorical price values for *TOA #9 Value* because these values are readily available to managers and continuous variables can increase predictive performance; and (3) we used all three categories for *TOA #10 Recent trend in catch* (i.e., 1=increasing, 2=stable, and 3=decreasing rather than the originally proposed options of 1.5=increasing/stable and 3=decreasing) because boosted classification trees can account for interactions between catch, effort, and abundance index trends.

We randomly divided the TOA scores into training (80% of data, n=155 stocks) and test (20% of data, n=38 stocks) datasets with stratification by stock status to ensure that both the test and training datasets included the same proportion of under, over, and fully exploited stocks. The training dataset was used to fit the BCT model, while the test dataset was used to provide an independent evaluation of the BCT model's predictive capacity. A grid search for the BCT model parameters that maximize Cohen's kappa using repeated 10-fold cross validation on the training dataset found the following optimal parameters: learning rate=0.001, interaction depth=2, number of trees=3000, and bag fraction=0.8 with multinomial error. Detailed information on model fitting is available in **Supplementary Appendix A.4.**

We evaluated the predictive performance of the BCT model by calculating the percentage agreement and Cohen's kappa for both the training and test datasets. For comparison, we evaluated the performance of six other catch-only methods for estimating status on stocks in the test dataset: SSP-2002 (Froese and Kesner-Reyes, 2002) and SSP-2013 (Kleisner et al., 2013), which estimate development status (e.g., undeveloped, developing, fully exploited), and CMSY (Martell and Froese, 2013), COM-SIR (Vasconcellos and Cochrane, 2005), SSCOM (Thorson et al., 2013), and mPRM (Costello et al., 2012), which estimate B/B_{MSY} (**Table 2**). The latter four methods were applied using the *datalimited* package in R (Anderson, 2016) based on the methods described in Rosenberg et al. (2014) and Anderson et al. (2017). Detailed information on implementing the alternative catch-only methods is available in **Supplementary Appendix A.5**.

2.4 Refinement of the historical catch statistics and scalars

The second step of the ORCS approach is to estimate the OFL as a factor of some historical catch statistic based on stock status; however, the original approach offers no formal recommendations on the choice of catch statistic and recommends simple catch scalars (i.e., 2.0, 1.0, 0.5 for under, fully, and overexploited stocks, respectively).

We identified the best status-specific historical catch statistics and scalars by comparing the most recent OFL (U_{MSY} x total biomass) to 24 historical catch statistics for the 105 stocks in the RAMLDB with the necessary information (i.e., U_{MSY}, total biomass time series, and catch/landings time series). The 24 historical catch statistics represent eight metrics (IQR, Winsorized, and arithmetic mean; 10th, 25th, 50th, 75th, and 90th percentiles) proposed in the original ORCS approach over three time periods (10 yr, 20 yr, whole time series). We used linear regression to assess the correlation between the OFL and each catch statistic and Akaike's Information Criterion (AIC) to rank the catch statistics within each status category. The best status-specific catch statistics were selected based on AIC ranking.

We calculated the ratio of the best status-specific catch statistic to the OFL for each stock based on its data-rich status estimate. We then calculated the 10th to 50th percentile of the observed ratios in each status category to evaluate as potential status-specific scalars for estimating the OFL. If stock status is correctly identified, the 50th percentile scalars should promote a 50% probability of overfishing (i.e., catch > OFL) in a given year, the U.S. legal maximum (DOC, 2016). Scalars more conservative than the median may be useful for buffering against classification uncertainty. Detailed information on calculating the OFL and the best status-specific historical catch statistics and scalars is available in **Supplementary Appendices A.3 and A.6**.

2.5 Evaluation of the refined ORCS approach

We evaluated ten potential refinements of the original ORCS approach. The first approach (the 'weighted 50th percentile scalar' approach) uses the BCT model to estimate the probability a stock is in each status category. It then estimates the OFL as the probability weighted average of the OFLs for each status category using the best status-specific catch statistics and 50th percentile scalars. The second approach (the 'unweighted 50th percentile scalar' approach) uses the BCT model to identify the most likely status category, then estimates the OFL using the best catch statistic and 50th percentile scalar for the category. The remaining eight approaches use the 45th-10th percentile scalars in the unweighted framework to examine the tradeoffs associated with using scalars more conservative than the median. We used the unweighted framework because preliminary analysis showed that the unweighted framework was superior to the weighted framework (**Table 4**; **Figure 6**). We evaluated the performance of these approaches by applying them to the 97 stocks ($n_{training}$=79, n_{test}=18) in the RAMLDB with the necessary information (i.e., B/B_{MSY}, U_{MSY}, total biomass time series, catch/landings time series) and calculated the percentage of stocks for which the predicted OFL exceeded the data-rich OFL estimate to use as a measure of the probability of overfishing. We also assessed the correlation between the OFLs predicted by the ORCS approach and those estimated by the data-rich assessments using linear regression.

3. RESULTS

3.1 Evaluation of the ORCS Table of Attributes

Although most attributes exhibited good variation in scores, a few were dominated by a single score category (*TOA #2, #4*), omitted an entire score category (*TOA #3*), or underutilized a score category (*TOA #11, #14*) (**Figure 2A**). The original approach classified all but three stocks as fully exploited (**Figure 2B**). Although the approach correctly classified the U.S. Mid-Atlantic weakfish stock as overexploited (B/B_{MSY}=0.131 in 2008), it incorrectly classified the fully exploited U.S. Gulf of Maine haddock (B/B_{MSY}=0.585 in 2011) and New Zealand bluenose (B/B_{MSY}=0.658 in 2011) stocks as overexploited. In fact, there was no correlation between the statuses predicted by the ORCS approach and those estimated by the data-rich assessment models (**Figure 2C**), and a Cohen's kappa value of 0.0001 indicates 'poor' classification accuracy. There was a weak correlation between the overexploitation risks predicted by the ORCS approach and those estimated by the data-rich assessment models (**Figure 2D; Supp. Figure** 1)

3.2 Refinement of the ORCS Table of Attributes

The BCT model correctly classified 74% of stocks in the training dataset and yielded a Cohen's kappa of 0.56 indicating 'good' classification accuracy (**Figure 3A**). The model performed better on fully exploited stocks (89% correct) than either underexploited (58% correct) or overexploited (62% correct) stocks. The BCT model also correctly classified 74% of stocks in the independent test dataset and yielded a Cohen's kappa of 0.56 indicating 'good' classification accuracy (**Figure 3B**). The model still performed better for fully exploited stocks (79% correct) than underexploited (77% correct) or overexploited (50% correct) stocks in the test dataset. The nearly equivalent performance of the BCT model on the training and test datasets suggests that the model is not overfit, which is consistent with the flat model tuning curves (**Supplementary Appendix A.4**). In 60% of misclassifications, the correct classification was the second most probable status identified by the model and only one misclassification (U.S. S. Pacific Coast gopher rockfish – no remarkable scores to explain this outcome) was so egregious as to classify an underexploited stock as overexploited or vice versa (**Figure 3B**). The BCT model was a better predictor of stock status, in terms of both accuracy and Cohen's kappa, than the other six catch-only methods that we evaluated (**Table 2; Supp. Tables 1 & 2**).

The BCT model identified seven attributes that each contribute more than 5% of the total predictive power (percents indicate relative influence of an attribute on the classification of a stock): *TOA #9 Value* (33.5%), *#1 Status of assess stocks in fishery* (13.1%), *#6 Targeting intensity* (12.3%), *#5 Discard rate* (8.8%), *#8 Occurrence in catch* (8.5%), *#7 M compared to dominant species* (8.0%), and *#3 Behavior affecting capture* (7.3%; **Figure 4A**). The attribute marginal effects, the effect of each attribute when the other attributes are held constant, suggest that stocks are more likely to be: (1) underexploited if there is a low rate of overexploitation of other stocks in the fishery, the taxon is worth less than US$1.00 per pound, and the taxon does not exhibit any aggregation behavior; (2) fully exploited if the stock is occasionally or actively targeted, the taxon exhibits aggregation behavior, and the taxon is worth more than US$2.00 per pound; and (3) overexploited if there is a high rate of overexploitation of other stocks in the fishery, the taxon is worth more than US$1.00 per pound, and the taxon occurs sporadically in the catch (**Figure 4B; Supp. Figure 2**).

3.3 Refinement of the historical catch statistics and scalars

The 90th percentile catch over the whole time series was most highly correlated with the OFL for underexploited stocks and longer timeframe metrics generally performed better than shorter timeframe metrics (**Table 3; Supp. Table 3**). The 25th percentile catch over the previous 10 years performed best for fully exploited stocks with more central and shorter timeframe metrics generally performing better than higher percentile and longer timeframe metrics (**Table 3; Supp. Table 3**). The mean catch of the previous 20 years performed best for overexploited stocks but this correlation was driven by a single strong leverage point (S. Labrador/E. Newfoundland Atlantic cod, whose 20-year mean exceeded the current OFL by more than 5 times, considerably more than the other overexploited stocks) and may be spurious. The 10th percentile catch over the whole time series provided the second best correlation and is more appropriate for overexploited stocks whose catch limits must be significantly reduced to allow rebuilding under U.S. law (**Table 3; Supp. Table 3**). The median scalars for relating the best catch statistic to the OFL were 1.90, 2.16, and 1.56 for under, fully, and overexploited stocks, respectively (**Table 3**). Scalars more conservative than the median are provided in **Table 3.**

3.4 Evaluation of the refined ORCS approaches

The OFLs predicted by the ORCS approach and estimated by the data-rich assessment models were significantly correlated in all ten potential refined ORCS approaches (**Table 4**; **Figure 5A-D**). The 'weighted 50th percentile scalar' approach resulted in the underutilization (i.e., predicted OFL less than data-rich OFL) of 63% of underexploited stocks and overfishing (i.e., predicted OFL exceeds data-rich OFL) of 73% and 91% of fully and overexploited stocks, respectively (**Figure 5E**). The 'unweighted 50th percentile scalar' approach performed better, resulting in the underutilization of 54% of underexploited stocks and overfishing of 56% and 45% of fully and overexploited stocks, respectively (**Figure 5F**). The more conservative 'unweighted 45th–10th percentile scalar' approaches reduced the overfishing of overexploited stocks but increased the underexploitation of under and fully exploited stocks (**Table 4; Figure 5G-H**). The 'unweighted 40th percentile scalars' are the largest scalars to reduce the probability of overfishing below 50%, the U.S. legal maximum (DOC, 2016), in all three status categories (**Table 4**).

4. DISCUSSION

Before being implemented, new stock assessment methods should be evaluated to validate their usefulness and transferability. Although the fully-flexible version of the original ORCS approach may produce useful status and catch limit estimates, it is challenging to validate because of its subjectivity. Therefore, we adopted a more specific, albeit less inclusive, definition of the ORCS approach for evaluation and refinement. Our results show that this interpretation of the ORCS approach is a poor predictor of stock status and should not be used for management decisions. The approach is heavily biased towards moderate classifications and classified all but three data-rich stocks as fully exploited. This result is not surprising given that all 20 stocks in the U.S. Southeast scored using the interpretation of the ORCS approach evaluated here were also classified as fully exploited, despite expert opinion that the stocks ranged from under to overexploited (SAFMC, 2012, 2013). The bias of the approach towards moderate classifications likely arises from: (1) an overrepresentation of moderate scores (in *TOA #4* notably and *#10, #12, #14* additionally) and (2) inappropriately wide threshold values for delineating status categories (1.75 and 2.25 might perform better). Furthermore, the non-linearity in the relative influence and marginal effects of the TOA attributes highlights the necessity of a weighting scheme. Although the original ORCS approach suggests that these adjustments can be made through expert judgement (Berkson et al. 2011), the refined ORCS approach presents an objective, transferable, and effective alternative.

The refined ORCS approach, which accounts for attribute importance, interactions, and non-linearity, is a better predictor of stock status than both the original ORCS approach and alternative catch-only methods. The refined approach correctly classified 73% (kappa=0.55, good) of the 37 stocks in the test dataset with a catch time series. In comparison, CMSY, which performed second best and also performed better than COM-SIR, SSCOM, and mPRM in Rosenberg et al.'s (2014) evaluation of these four methods, classified only 41% (kappa=0.15, poor) of these 37 data-rich test stocks correctly. The refined ORCS approach also outperformed SSP-2002 and SSP-2013, which have been shown to be poor and inherently pessimistic predictors of stock status (Branch et al., 2011; Carruthers et al., 2012), as well as mPRM, whose developers state that it should not be used to assess the status of individual stocks (Costello et al., 2012). Catch-based methods represent a class of widely used, but still controversial (Pauly et al., 2013), approaches to estimate status and the refined ORCS approach may be a useful alternative for estimating the status of data-poor stocks.

The refined ORCS approach also identifies catch statistics and scalars that estimate catch limits that prevent overfishing in accordance with U.S. legal mandates, suggesting that it can be used when data-moderate methods such as DB-SRA and DCAC are not possible or appropriate. Although the refined approach misclassifies many overexploited stocks, conservative catch scalars successfully buffer against classification uncertainty. The 40th percentile scalars produce the highest catches while reducing the probability of overfishing in all three status categories below 50%, the U.S. legal maximum (DOC, 2016);

however, they also estimate OFLs more than five times the data-rich OFL for some stocks. More conservative catch scalars will further reduce the probability and magnitude of overfishing but will result in concomitant increases in the probability and magnitude of underfishing. Managers must therefore determine which catch scalars are most consistent with their risk policies. We provide a web tool for managers to implement the approach here: https://cfree.shinyapps.io/refined_orcs_approach/

The evaluation of the ORCS approach using data-rich stock assessments, while necessary because the ability of the approach to predict stock status cannot be evaluated through traditional simulation testing (Wiedenmann et al. 2013; Carruthers et al. 2014), is somewhat problematic given the uncertainty in even the most sophisticated assessment models (Brooks and Deroba, 2015). For example, assessment model reference points (i.e., B/B_{MSY}, OFL) used to evaluate the performance of the ORCS approach and assessment model output (i.e., biomass and effort time series) used to score the ORCS Table of Attributes could both be incorrect. However, we took measures to eliminate the more uncertain assessments and we only used stock assessment output in the scoring of *TOA #12 Recent trend in effort* and *#13 Recent trend in abundance index*, which were both unimportant predictors of stock status. Furthermore, we trained the BCT model used in the refined approach to estimate categorical status (i.e., under, fully, or overexploited) rather than continuous status (i.e., B/B_{MSY}) because stock assessment models are generally more certain in status classifications than in precise B/B_{MSY} values. Finally, the ability of the refined ORCS approach to reproduce the conclusions of possibly incorrect but presumably better data-rich status determinations is still useful, especially given the recent success of data-rich assessment and management in rebuilding fisheries (Worm et al., 2009; Hilborn and Ovando, 2014).

The refinement of the ORCS approach through testing against data-rich stocks could also be problematic given the differences in the dynamics of data-poor and data-rich fisheries. Assessed (data-rich) fisheries generally target larger, slower growing, and higher trophic level species (Pinsky et al., 2011) and are higher volume, more valuable, and in better condition (Costello et al., 2012) than their unassessed (data-poor) counterparts. Consequently, it is possible that informative predictors of data-poor fisheries status could be uninformative or even trend opposite for data-rich fisheries. For example, in well-managed fisheries, decreasing catch could be the result of responsive management (Murawski et al., 2007) and increasing effort could indicate the sustainable development of a new or rebuilt fishery. Furthermore, the generally healthy status of data-rich stocks results in only a small sample of overexploited stocks (30 of 193 stocks, 15.5%) available for model training and testing. Thus, the model may have performed poorly at classifying overexploited stocks because of the limited number of overexploited stocks in the dataset.

The dynamics of the most important predictors of stock status in the BCT model are consistent with other studies and are likely conserved across data-poor and -rich fisheries. For example, the importance of ex-vessel price is not surprising given that fishery development is frequently driven by profits (Sethi et al., 2010). The importance of assessed stock status also makes intuitive sense (i.e., a stock in a generally well- or poorly-managed fishery is also likely to be well- or poorly-managed, respectively) and is similar to the region effect, which has been shown to be useful in discriminating stock status (Ricard et al., 2012; Thorson et al., 2012). The significant increase in overexploitation risk resulting from aggregation behavior is supported by emerging evidence that schooling, fast-lived fish may actually be more vulnerable to collapse than solitary, long-lived taxa due to high harvest rates lagging behind rapid changes in environment and productivity (Pinsky et al., 2011). The decrease in overexploitation risk with increasing occurrence in the catch opposes the predictions of the original Table of Attributes and suggests that rarity in the catch is indicative of a depleted stock rather than a lightly exploited one. Finally, recent trends in catch and effort, the attributes most likely to be confounding between data-poor and -rich fisheries, exert little predictive influence, increasing the likelihood that the refined ORCS approach is as predictive for data-poor stocks as it is for data-rich ones.

The refined ORCS approach also provides important guidance on the choice of historical catch statistics and scalars. Longer timeframe, higher percentile catch statistics perform best for underexploited stocks with light exploitation histories. Moderate timeframe, more central catch statistics perform best for fully exploited stocks where recent management has been effective in sustaining abundance and yield. Longer timeframe, lower percentile catch statistics perform best for overexploited stocks where recent catches have resulted in depletion. To consistently achieve a relatively low risk of overfishing, the catch scalars used to scale the historical catch statistic to the overfishing limit will have to be conservative to buffer against substantial classification uncertainty. This conclusion is especially true for data-poor stocks with uncertainty in their catch time series, such as the rarely caught snapper-grouper species in the U.S. Southeast which suffer from misidentification problems (SAFMC, 2013; Berkson et al., 2011). Although conservative scalars will effectively protect overexploited stocks, they will also result in forgone yield from under and fully exploited stocks.

The refined ORCS approach represents one step towards Berkson et al.'s (2011) recommendations for testing and improving the original ORCS approach but could under additional refinement and evaluation. The predictive performance of the approach could be improved by identifying new predictive attributes. For example, life history characteristics such as age at maturity, maximum age, maximum length, and trophic level and fishery characteristics such as time since development and exploitation history have all been shown to be useful in discriminating stock status (Sethi et al., 2010; Pinsky et al., 2011; Costello et al., 2012; Thorson et al., 2012; Neubauer et al., 2013) and could be incorporated into the refined TOA and BCT model. Furthermore, the performance of the status-specific historical catch statistics and scalars used in the refined approach should be tested through management strategy evaluation, such as in Wiedenmann et. al (2013), to determine whether they actually promote sustainable fishing levels. The development of simple data-limited decision support tools has been a central focus of recent fisheries management (Berkson and Thorson, 2015) and the refined ORCS approach provides an additional tool for managers faced with the legal mandates and data limitations of contemporary fisheries management.

ACKNOWLEDGEMENTS

This manuscript was greatly improved by the thoughtful reviews of Steve Cadrin and André Punt. We are also grateful to Emma Fuller, Doug Zemeckis, Kyle Shertzer, and Marcus Drymon for answering questions about U.S. fisheries, Jim Berkson for providing feedback on preliminary methods and results, Sean Anderson for help with the *datalimited* R package, and Michael Roswell and the members of the Jensen and Pinsky labs for their helpful feedback on the manuscript. This work was supported by a NMFS-Sea Grant Population Dynamics Fellowship to CMF and NJ Sea Grant Omnibus Grant to JW and OPJ. Thus, this publication (NJSG-XX-XXX) is also the result of research sponsored by the New Jersey Sea Grant Consortium (NJSGC) with funds from the National Oceanic and Atmospheric Administration (NOAA) Office of Sea Grant, U.S. Department of Commerce, under NOAA grant numbers NA13OAR4170114 and NA14OAR4170085 and the NJSGC.

The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the NJSGC or the U.S. Department of Commerce.

REFERENCES

Anderson, S.C., Afflerbach, J., Cooper, A.B., Dickey-Collas, M., Jensen, O.P., Kleisner, K.M., Longo, C., Osio, G.C., Ovando, D., Minte-Vera, C., Minto, C., Mosqueira, I., Rosenberg, A.A., Selig, E.R., Thorson, J.T., Walsh, J.C., 2016. datalimited: Stock assessment methods for data-limited fisheries. R package version 0.0.2. Available at: https://github.com/datalimited/datalimited

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

Berkson, J., Barbieri, L., Cadrin, S., Cass-Calay, S., Crone, P., Dorn, M., Friess, C., Kobayashi, D., Miller, T.J., Patrick, W.S., Pautzke, S., Ralston, S., Trianni, M., 2011. Calculating Acceptable Biological Catch for stocks that have reliable catch data only (Only Reliable Catch Stocks – ORCS). NOAA Technical Memorandum NMFS-SEFSC-616. NMFS-SEFSC, Miami, FL.

Berkson, J., Thorson, J. T., 2015. The determination of data-poor catch limits in the United States: is there a better way? ICES J. Mar. Sci. 72(1), 237-242.

Branch, T.A., Jensen, O.P., Ricard, D., Yimin, Y., Hilborn, R., 2011. Contrasting global trends in marine fishery status obtained from catches and from stock assessments. Conserv. Biol. 25(4), 777–786.

Brooks, E.N., Deroba, J.J., 2015. When "data" are not data: the pitfalls of post hoc analyses that use stock assessment model output. Can. J. Fish. Aquat. Sci. 72(4), 634-641.

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.

Carruthers, T.R., Walters, C.J., McAllister, M.K., 2012. Evaluating methods that classify fisheries stock status using only fisheries catch data. Fish. Res. 119-120, 66-79.

Cohen, J., 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychol. Bull. 70(4), 213-220.

Cope, J., Dick, E.J., MacCall, A., Monk, M., Soper, B., Wetzel, C., 2013. Data-moderate stock assessments for brown, China, copper, sharpchin, stripetail, and yellowtail rockfishes and English and rex soles in 2013. PFMC June 2013 Meeting Agenda Item F.5.a, Attachment 1.

Cope, J.M., 2013. Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for deriving overfishing limits in data-limited situations. Fish. Res. 142, 3-14.

Costello, C., Ovando, D., Hilborn, R., Gains, S.D., Deschenes, O., Lester, S.E., 2012. Status and solutions for the world's unassessed fisheries. Science 338, 517-520.

Dick, E.J., MacCall, A.D., 2011. Depletion-based stock reduction analysis: A catch-based method for determining sustainable yields for data-poor fish stocks. Fish. Res. 110, 331-341.

DOC, 2007. Magnuson-Stevens Fishery Conservation and Management Act (MSFCMA). U.S.A. Public Law 94-265. Department of Commerce (DOC), Washington, D.C.

DOC, 2016. Magnuson-Stevens Act Provisions; National Standard Guidelines; Final Rule. Federal Register 81(201): 71858-71904. Department of Commerce (DOC), Washington, D.C.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted classification trees. J. Anim. Ecol. 77(4), 802-813.

Fleiss, J.L., 1973. Statistical methods for rates and proportions. John Wiley & Sons, New York, NY.

Froese, R., Kesner-Reyes, K., 2002. Impact of Fishing on the Abundance of Marine Species. ICES CM 2002/L:12.

Froese, R., Pauly, D., 2016. FishBase. Available at: www.fishbase.org

Gulland, J.A., 1971. The Fish Resources of the Ocean. Fishing News Books: West Byfleet, UK.

Hilborn, R., Ovando, D., 2014. Reflections on the success of traditional fisheries management. ICES J. Mar. Sci. 71(5), 1040-1046.

ICES, 2014. Report of the Workshop on the Development of Quantitative Assessment Methodologies based on LIFE-history traits, exploitation characteristics, and other relevant parameters for data-limited stocks (WKLIFE IV). ICES CM 2014/ACOM:54, Lisbon, Portugal.

ICES, 2015. Report of the Fifth Workshop on the Development of Quantitative Assessment Methodologies based on Life-history Traits, Exploitation Characteristics and other Relevant Parameters for Data-limited Stocks (WKLIFE V). ICES CM 2015/ACOM:56, Lisbon, Portugal.

ICES, 2017. Report of the ICES Workshop on the Development of Quantitative Assessment Methodologies based on Life-history traits, exploitation characteristics, and other relevant parameters for stocks in categories 3–6 (WKLIFEVI). ICES CM 2016/ACOM:59, Lisbon, Portugal.

Kleisner, K., Zeller, D., Froese, R., Pauly, D., 2013. Using global catch data for inferences on the world's marine fisheries. Fish Fish. 14(3), 293-311.

Kuhn, M., 2016. caret: Classification and Regression Training. R package version 6.0-71.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33(1), 159-174.

MacCall, A.D., 2009. Depletion-corrected average catch: A simple formula for estimating sustainable yields in data-poor situations. ICES J. Mar. Sci. 66, 2267-2271.

Martell, S., Froese, R., 2013. A simple method for estimating MSY from catch and resilience. Fish Fish. 14, 504-514.

MPI, 2016. The Status of New Zealand's Fisheries 2015. Ministry for Primary Industries (MPI), Wellington, New Zealand.

Murawski, S., Methot, R., Tromble, G., 2007. Biodiversity loss in the ocean: how bad is it? Science 316(5829), 1281-1284.

Neubauer, P., Jensen, O.P., Hutchings, J.A., Baum, J.K., 2013. Resilience and recovery of overexploited marine populations. Science 340(6130), 347-349.

Newman, D., Berkson, J., Suatoni, L., 2015. Current methods for setting catch limits for data-limited fish stocks in the United States. Fish. Res. 164, 86-93.

Pauly, D., Hilborn, R., Branch, T.A., 2013. Fisheries: does catch reflect abundance? Nature 494(7437), 303-306.

Pauly D., Zeller D., 2015. Sea Around Us Concepts, Design and Data. Available at: www.seaaroundus.org

Pinsky, M.L., Jensen, O.P., Ricard, D., Palumbi, S.R., 2011. Unexpected patterns of fisheries collapse in the world's oceans. Proc. Natl. Acad. Sci. 108(20), 8317-8322.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria. Available at: http://www.R-project.org/

Restrepo, V.R., Thompson, G.G., Mace, P.M., Gabriel, W.L., Low, L.L., MacCall, A.D., Methot, R.D., Powers, J.E., Taylor, B.L., Wade, P.R., Witzig, J.F., 1998. Technical guidance on the use of precautionary approaches to implementing National Standard 1 of the Magnuson–Stevens Fishery Conservation and Management Act. NOAA Technical Memorandum, NMFS-F/ SPO-31.

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 Fish. 13(4), 380-398.

Ridgeway, G., 2016. gbm: Generalized Boosted Classification Models. R package version 2.1.1.

Rosenberg, A.A., Fogarty, M.J., Cooper, A. B., Dickey-Collas, M., Fulton, E.A., Gutiérrez, N.L., Hyde, K.J.W., Kleisner, K.M., Longo, C., Minte-Vera, C.V., Minto, C., Mosqueira, I., Osio, G.C., Ovando, D., Selig, E.R., Thorson, J.T., Ye, Y., 2014. Developing new approaches to global stock status assessment and fishery production potential of the seas. FAO Fisheries and Aquaculture Circular, Rome, Italy.

SAFMC, 2012. SSC ORCS Workshop Report. South Atlantic Fishery Management Council-Scientific & Statistical Committee (SAFMC-SSC), North Charleston, SC.

SAFMC, 2013. SSC ORCS Workshop II Report. South Atlantic Fishery Management Council-Scientific & Statistical Committee (SAFMC-SSC), North Charleston, SC.

Sethi, S.A., Branch, T.A., Watson, R., 2010. Global fishery development patterns are driven by profit but not trophic level. Proc. Natl. Acad. Sci. 107(27), 12163-12167.

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. Can. J. Fish. Aquat. Sci. 69(4), 645-655.

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. Can. J. Fish. Aquat. Sci. 70(12), 1829-1844.

Vasconcellos, M., Cochrane, K., 2005. Overview of world status of data-limited fisheries: inferences from landings statistics. In: Kruse, G.H., Gallucci, V.F., Hay, D.E., Perry, R.I., Peterman, R.M., Shirley, T.C., Spencer, P.D., Wilson, B., Woodby, D. (eds.): Fisheries Assessment and Management in Data-Limited Situations. Alaska Sea Grant College Program, University of Alaska Fairbanks, Fairbanks, AK, pp. 1–20.

Walters, C.J., Martell, S.J., 2004. Fisheries Ecology and Management. Princeton University Press, Princeton, NJ.

Wetzel, C.R., Punt, A.E., 2015. Evaluating the performance of data-moderate and catch-only assessment methods for US west coast groundfish. Fish. Res. 171, 170-187.

Wiedenmann, J., Wilberg, M.J., Miller, T.J., 2013. An evaluation of harvest control rules for data-poor fisheries. N. Am. J. Fish. Manag. 33(4), 845-860.

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., 2009. Rebuilding global fisheries. Science 325(5940), 578-585.

Figure 1. Demographics of the 193 data-rich stocks scored using the ORCS approach by: **(A)** taxonomic group; **(B)** managing country or multinational body; **(C)** U.S. assessment agency (U.S. stocks only; n=99, 51.3% of scored stocks); **(D)** assessment year; **(E)** stock status (B/B_{MSY}); and **(F)** fishery size (average annual catch over the most recent 5 years).

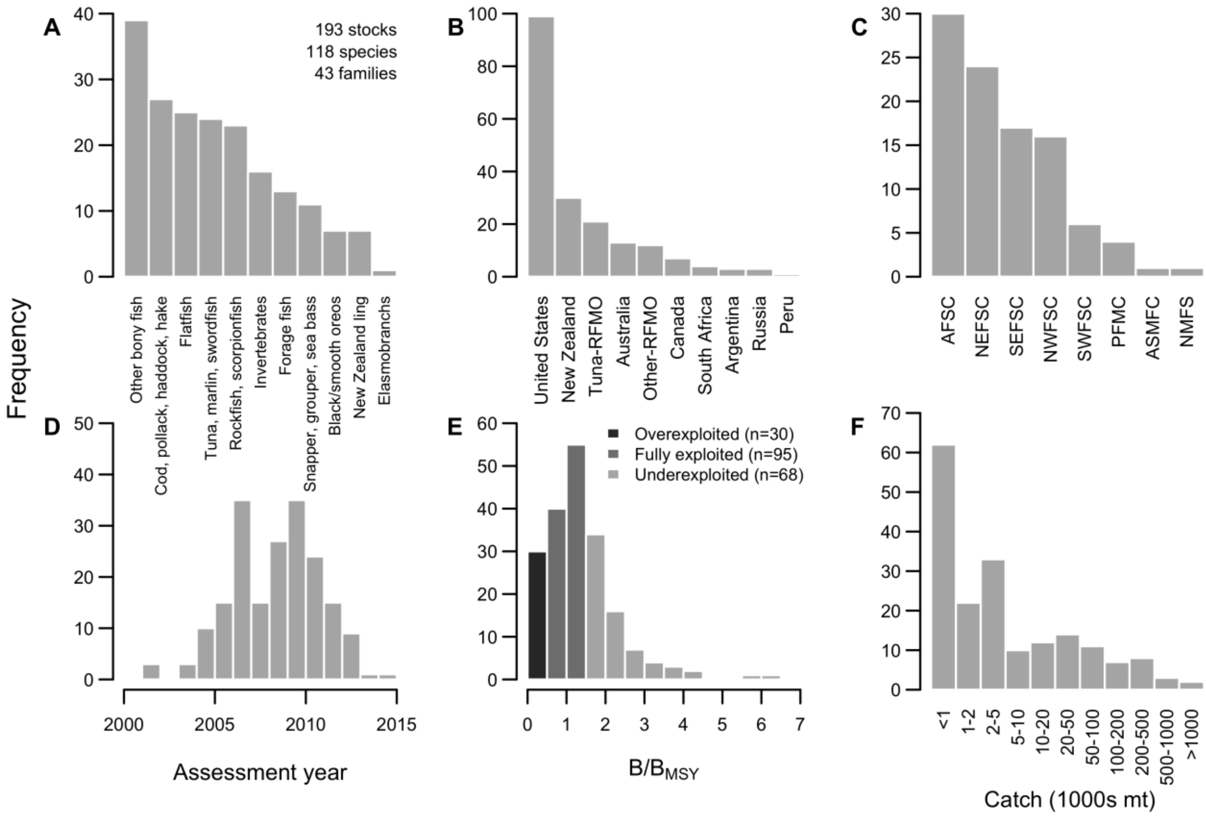

Figure 2. The distribution of **(A)** attribute scores and **(B)** overall scores for the 193 data-rich stocks scored using the original ORCS approach and **(C)** comparison of statuses and **(D)** risks predicted by the ORCS approach and estimated by data-rich assessment models. In **(A)**, bars show the proportion of scores represented in each TOA attribute. For *TOA #10*, scores of 1 and 2 (hatched) are reassigned scores of 1.5 and only count towards the overall score if effort is stable (*TOA #12*, score=2). In some cases, attributes could not be scored due to a lack of data or applicability and were given an 'NA' value (grey shading). In **(B)**, vertical lines indicate the threshold values (1.5 & 2.5) that separate under, fully, and overexploited stocks. In **(C)** and **(D)**, the black lines indicate linear regressions fit to the data and the gray shading indicates the confidence intervals for the regressions.

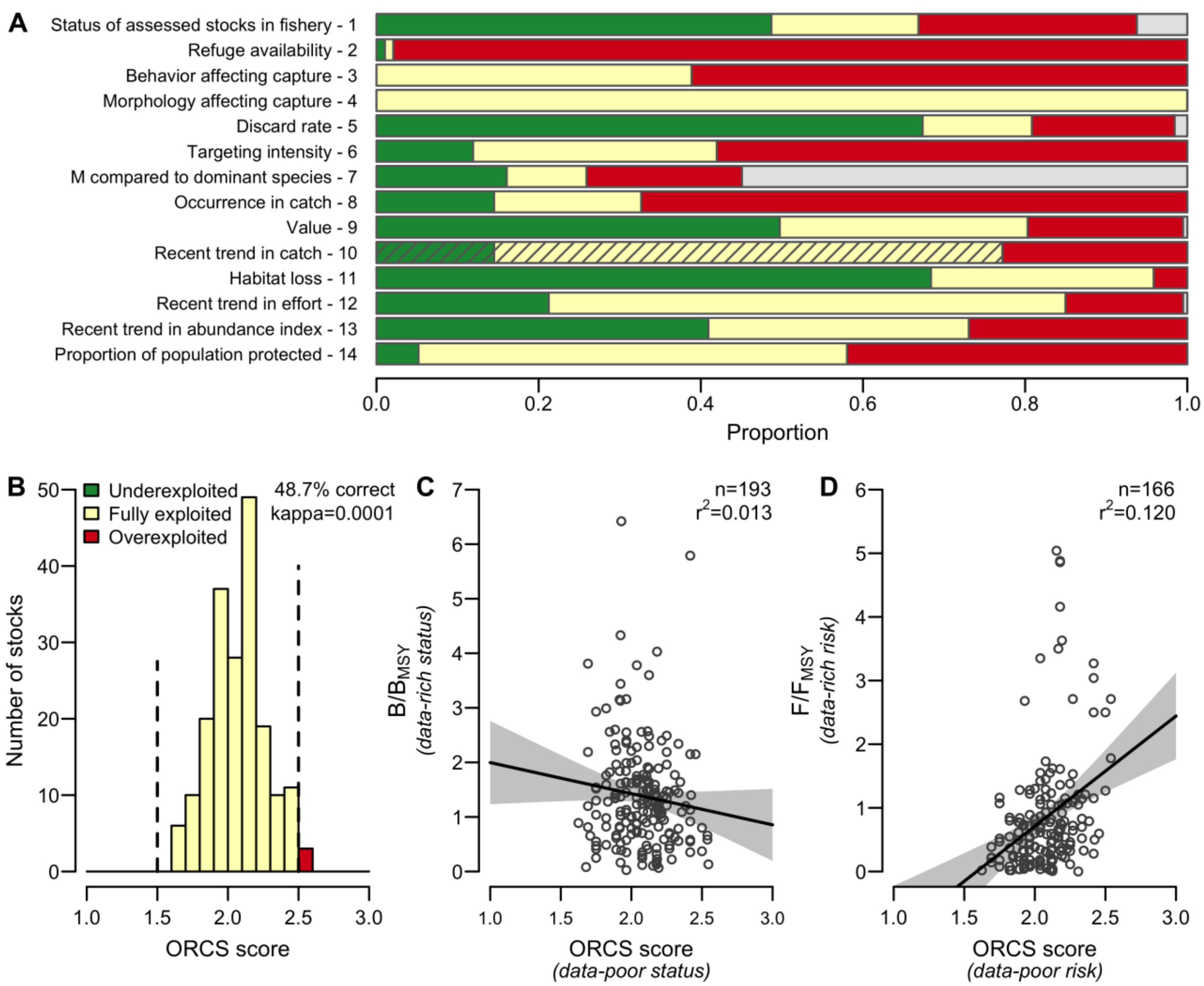

Figure 3. The performance of the boosted classification tree (BCT) model on the **(A)** training (n=155 stocks, 80% of data) and **(B)** test datasets (n=38 stocks, 20% of data). In **(A)**, bars show the proportion of status predictions for each status category. Percentages indicate the proportion of correct classifications in each category (overall accuracy=74% and Cohen's kappa=0.56). In **(B)**, bars show the probability that a stock is in each status category, where the highest probability category is the BCT model's prediction of stock status; stocks are grouped and sorted by B/B_{MSY} from the data-rich assessment model. Percentages indicate the proportion of correct classifications in each category; stars mark incorrectly classified stocks with colors indicating the direction of the misclassification (overall accuracy=74% and Cohen's kappa=0.56).

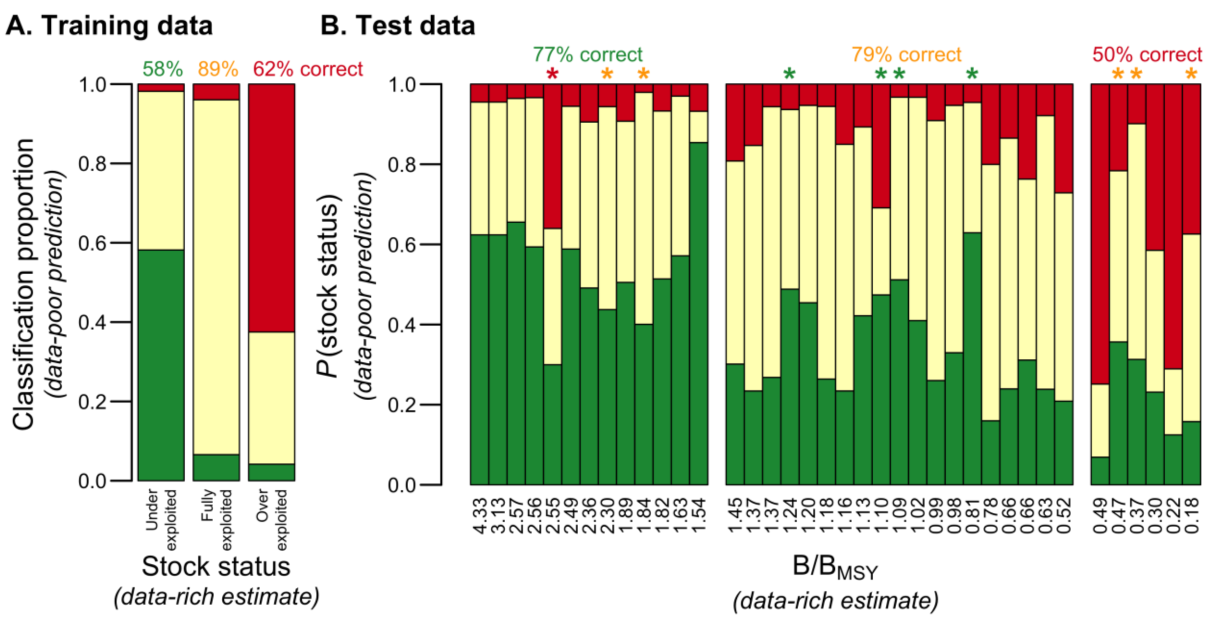

Figure 4. The **(A)** relative influence and **(B)** marginal effects of the five most important TOA attributes in the boosted classification tree model. In **(B)**, lines represent the effect of each attribute on the probability that a stock is in each status category when the other attributes are held constant. *TOA #2 Refuge availability* and *#4 Morphology affecting capture* were omitted from the model due to lack of predictive power.

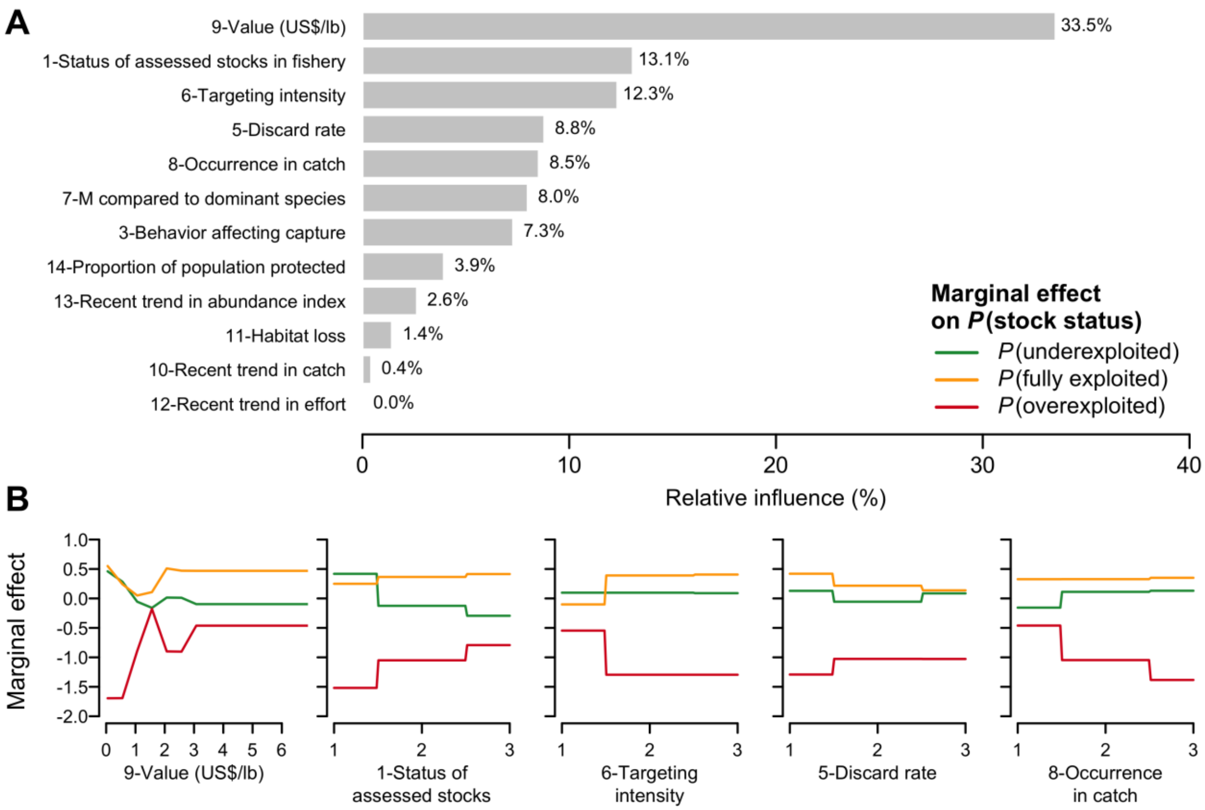

Figure 5. The **(A-D)** correlation between and **(E-H)** ratio of the overfishing limits (OFLs) predicted by the ORCS approach and estimated by data-rich assessment models for 97 stocks in four potential refined ORCS approaches. In **(A-D)**, black lines indicate linear regressions fit to the untransformed data and the gray shading indicates the confidence interval for the regression. In **(E-H)**, ratios were also calculated using the untransformed data. The dotted horizontal lines indicate perfect agreement between the ORCS predictions and the data-rich model estimates and boxplots indicate the median (heavy black line), interquartile range (IQR; box), 1.5 times the IQR (whiskers), and outliers.

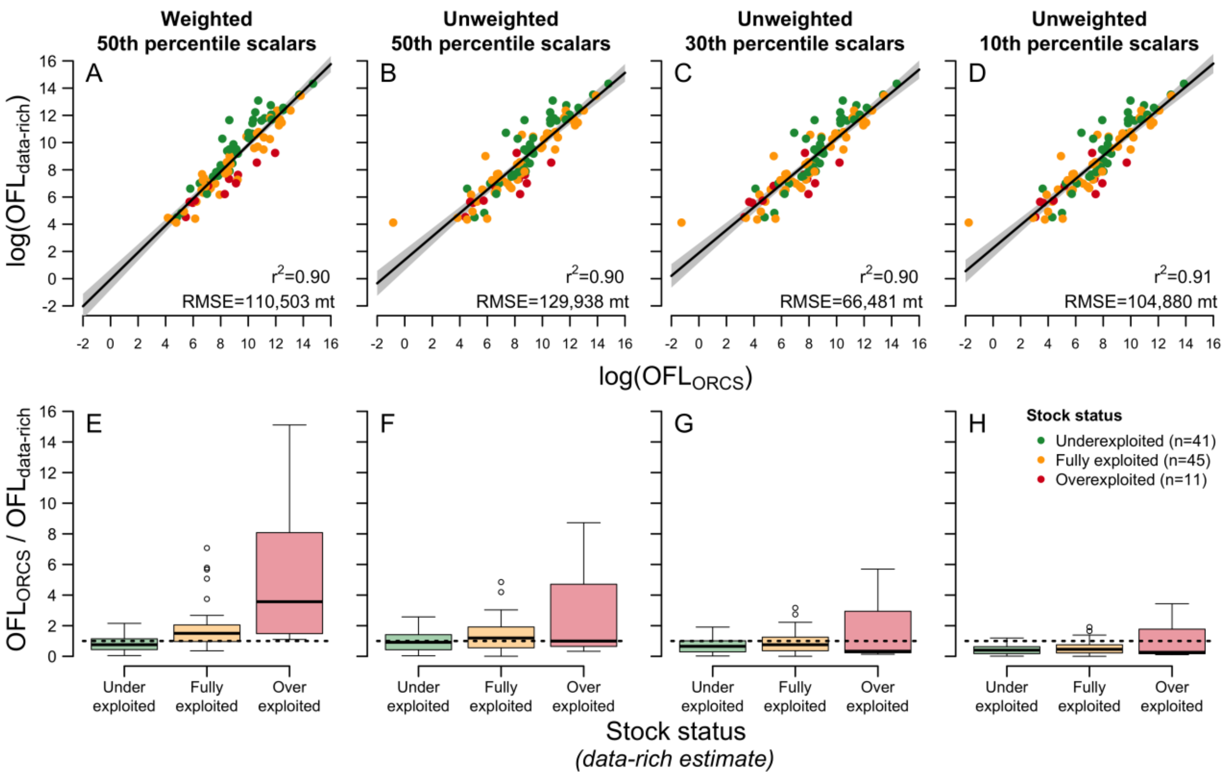

Table 1. ORCS Table of Attributes (TOA; adapted from SAFMC, 2012).

#	Attribute	Stock status[1] Underexploited (1)	Fully exploited (2)	Overexploited (3)
1	Status of assessed stocks in fishery[2]	<10% overfished	10-25% overfished	>25% overfished
2	Refuge availability *(not used in refined approach)*	<50% of habitat accessible	50-75% of habitat accessible	>75% of habitat accessible
3	Behavior affecting capture	-------------	No aggregation behavior	Exhibits aggregation behavior
4	Morphology affecting capture *(not used in refined approach)*	Low susceptibility	Average susceptibility	High susceptibility
5	Discard rate[2]	Discards <10% of catch	Discards 10-25% of catch	Discards >25% of catch
6	Targeting intensity	Not targeted	Occasionally targeted	Actively targeted
7	M compared to dominant species[3]	Higher mortality rate	Equivalent mortality rates	Lower mortality rate
8	Occurrence in catch	Sporadic *(in <10% of efforts)*	Common *(in 10-25% of efforts)*	Frequent *(in >25% of efforts)*
9	Value (US$/lb, 5-year mean) *(continuous value in refined approach)*	<$1 / lb	$1-$2.25 / lb	>$2.25 / lb
10	Recent trend in catch	Increasing last 5 years *(score=1.5 in original approach)*	Stable last 5 years *(score=1.5 in original approach)*	Decreasing last 5 years
11	Habitat loss[4]	No time in threatened habitats	Part time in threatened habitats *(full time in partially threatened habitats)*	Full time in threatened habitats
12	Recent trend in effort	Decreasing last 5 years	Stable last 5 years	Increasing last 5 years
13	Recent trend in abundance index	Increasing last 5 years	Stable last 5 years	Decreasing last 5 years
14	Proportion of population protected	Most of resource is protected *(size limits AND time/space closures)*	Some of resource is protected *(size limits OR time/space closures)*	None of resource is protected *(no size limits or time/space closures)*

[1] In the original ORCS approach, stock status is estimated as the mean of the TOA scores (<1.5=underexploited; 1.5-2.5=fully exploited; >2.5=overexploited).
[2] Replaced vague score descriptions in the original table with straightforward percentage thresholds. Note: the definition of overfishing varies by management agency (**Supplementary Appendix A.2**).
[3] Removed ambiguity of score descriptions in the original table and specified that M's must differ by >20% to be considered different. See **Supplementary Appendix A.2** for definition of dominant species.
[4] Rephrased original attribute to be conceptually simpler and easier to score. See **Supplementary Appendix A.2** for list of threatened habitats.

Table 2. Status classification performance of catch-only assessment methods applied to the 37 data-rich stocks with catch time series in the test dataset (COM-SIR and SSCOM converged for only 33 test stocks).

		37 stocks[1]		**33 stocks**[2]		
	Method	**Kappa**	**Accuracy**	**Kappa**	**Accuracy**	**Bias**
1	Refined ORCS approach	0.549	0.730	0.558	0.727	slightly optimistic
2	cMSY	0.148	0.405	0.205	0.455	pessimistic
3	SSP-2002	0.120	0.378	0.171	0.424	pessimistic
4	COM-SIR	-----	-----	0.114	0.424	optimistic
5	SSP-2013	0.041	0.270	0.052	0.303	toward extremes
6	mPRM	-0.015	0.405	-0.005	0.394	central
7	Original ORCS approach	-0.035	0.459	-0.040	0.424	central
8	SSCOM	-----	-----	-0.120	0.333	slightly central

[1] 13 underexploited, 18 fully exploited, and 6 overexploited stocks; See **Supp. Table 1** for status classifications.
[2] 12 underexploited, 15 fully exploited, and 6 overexploited stocks; See **Supp. Table 2** for status classifications.

Table 3. Best status-specific historical catch statistics and potential status-specific catch scalars for relating the best catch statistic to the overfishing limit (OFL).

				OFL scalars**								
Stock status	**Best historical catch statistic***	**r^2**	**n**	**50th**	**45th**	**40th**	**35th**	**30th**	**25th**	**20th**	**15th**	**10th**
Underexploited	90th percentile, whole time series	0.91	45	1.90	1.78	1.62	1.53	1.41	1.34	1.29	1.11	0.88
Fully exploited	25th percentile, previous 10 years	0.91	49	2.16	1.84	1.77	1.57	1.41	1.22	1.15	1.02	0.85
Overexploited	10th percentile, whole time series	0.89	11	1.56	1.53	1.49	1.00	0.52	0.51	0.50	0.45	0.41

* See **Supp. Table 3** for the ranking of potential catch statistics.
** The 50th percentile scalars should promote a 50% probability of overfishing if stock status is correctly identified. The other, more conservative scalars may be useful for buffering against classification uncertainty.

Table 4. The percentage of stocks (n=97) whose predicted OFLs exceeded the data-rich OFL estimates under potential catch scalars with unweighted status predictions (maximum observed OFL_{ORCS}:$OFL_{data\text{-}rich}$ ratio shown in parentheses).

		OFL scalars								
Stock status	**n**	**50th**	**45th**	**40th**	**35th**	**30th**	**25th**	**20th**	**15th**	**10th**
Underexploited	41	46.3% (2.6)	41.5% (2.4)	39.0% (2.2)	34.1% (2.1)	26.8% (1.9)	19.5% (1.8)	17.1% (1.7)	9.8% (1.5)	4.9% (1.2)
Fully exploited	45	55.6% (4.8)	51.1% (4.1)	44.4% (4.0)	37.8% (3.5)	35.6% (3.2)	31.1% (2.7)	26.7% (2.6)	22.2% (2.3)	17.8% (1.9)
Overexploited	11	45.5% (8.7)	45.5% (7.4)	45.5% (7.2)	45.5% (6.3)	45.5% (5.7)	36.4% (4.9)	36.4% (4.6)	36.4% (4.1)	36.4% (3.4)
r^2:		0.90	0.90	0.90	0.90	0.90	0.91	0.91	0.91	0.91
RMSE, mt:		129,938	93,462	86,589	71,454	66,481	69,664	73,673	85,452	104,880

* See **Table 3** for the status-specific catch scalars and best historical catch statistics.