Improving Estimates of the State of Global Fisheries Depends on Better Data

2021-06-18

Title Page

Title 1: Improving Estimates of the State of Global Fisheries Depends on Better Data

Title 2: Status of Global Unassessed Fisheries will Remain Highly Uncertain without Better Data

Running Title: Unassessed Fisheries

Authors: Daniel Ovando[1], Ray Hilborn[1], Cole Monnahan[2], Merrill Rudd[3], Rishi Sharma[4], James Thorson[2], Yannick Rousseau[5], Yimin Ye[4]

Affiliations:

[1]University of Washington, School of Aquatic and Fishery Sciences 1122 NE Boat St, Box 355020 Seattle, WA 98195-5020 Seattle, WA, USA 98103

[2]NOAA Fisheries Alaska Fisheries Science Center, Seattle, WA, USA

[3]Scaleability LLC

[4]Food and Agriculture Organization of the United Nations, Fisheries and Aquaculture,Rome, IT

[5]University of Tasmania, Institute for Marine and Antarctic Studies,Hobart, TAS, AUS

Authorship: DO, RH, CM,MR, RS, JT designed model structure and ran analyses. RS, YR, and YE supplied data. All authors contributed to writing of the manuscript

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.

Abstract

Implementation of the United Nations Sustainable Development Goals requires assessments of the global state of fish populations. While we have reliable estimates of stock status for fish populations accounting for approximately half of recent global catch, our knowledge of the state of the majority of the world's 'unassessed' fish stocks remains highly uncertain. Numerous publications have produced estimates of the global status of these unassessed fisheries, but limited quantity and quality of data along with methodological differences have produced counterintuitive and conflicting results. Here, we show that despite numerous efforts, our understanding of the status of global fish stocks remains incomplete, even when new sources of broadly available data are added. Estimates of fish populations based primarily on catch histories alone on average performed 29% better than a random guess. But, on average these methods assigned fisheries to the wrong FAO status category 57% of the time. Effective improvement in estimates of the state of the world's exploited fish populations depends more on expanded collection of new information and efficient use of existing data than development of new modeling methods.

Introduction

The United Nations Sustainable Development Goal 14 (SDG 14), focusing on "Life under water," calls for the global community to "Conserve and sustainably use the oceans, seas and marine resources for sustainable development." Marine fisheries are one of the largest anthropogenic impacts in the oceans aside from the effects of climate change. Fisheries are also a critical source of economic prosperity, cultural identity, and food security around the globe. As such, meeting the SDG 14 targets depends in part on our ability to effectively measure the status of global marine fish populations and fisheries. While our understanding of the state of world fisheries has improved over the last decade, the majority of the world's fish populations, making up roughly 50% of marine landings (though only a few percent of the total number of fisheries in the world), lack formal statistical assessments (stock assessments) of their population size relative to reference points (FAO 2020; Hilborn *et al.* 2020). This is a major impediment to ensuring the sustainable development of the world's oceans. In this paper, we consider why the assessment of global fisheries remains a challenge, and chart a path towards a better understanding of the state of the world's marine resources.

What do we currently know about the state of fished marine species around the world? The Food and Agriculture Organization of the United Nations' (FAO) is the custodian agency of the SDG 14 Indicator on fisheries sustainability, and the FAO's State of World Fisheries and Aquaculture (SOFIA) report is the most widely used primary source for tracking the global state of fisheries. The latest SOFIA report, covering 70% of the landings of all fisheries in the world, estimates that as of 2017 59.6% of marine fish stocks are maximally sustainably fished (at or near targets for sustainable food production), 6.2% are underfished, and 34.2% are overfished (FAO 2020). Where possible, these statements about the status of individual fish stocks are made on the basis of formal stock assessments summarized in the RAM Legacy Stock Assessment Database (RLSADB) (Ricard *et al.* 2012). The "assessed" fisheries in RLSADB represent roughly 50% of global capture as of 2020 (Hilborn *et al.* 2020), and represent our best estimates of the state of assessed fish populations around the globe.

That leaves roughly 50% of global landings, and the majority of global fisheries, as currently "unassessed." While these unassessed stocks are generally individually smaller and less economically valuable than the fish populations in the assessed category, collectively they are a vital source of food, employment, cultural value, and ecosystem services around the world. The SOFIA report includes unassessed stocks and bases its estimates for these fisheries on data-limited methods or qualitative expert opinion for each region where these stocks are distributed (FAO 2020). While these methods combined with local knowledge can provide good insight as to the general status of unassessed fish stocks, the SOFIA assessment was designed in the 1970s based on the then available data and methods. With the surge in data availability, such as global assessments of management strength (Melnychuk *et al.* 2017), trawl footprints (Amoroso *et al.* 2018), and fishing effort (Rousseau *et al.* 2019), the SOFIA global assessment methods for unassessed fish populations need to be updated to meet the demand for tracking progress of global fish populations towards the SDG goals. In addition, the SOFIA process assigns unassessed fisheries to broad categories of stock status, while many stakeholders seek information on more specific values such as biomass relative to the biomass that would on average result in maximized yield over the long-term.

Numerous studies in recent years have put forward versions of "data-limited" models that have attempted to provide numerical estimates of the global status of unassessed fish stocks lacking the data or capacity needed for formal stock assessment (Pauly 2007; Thorson *et al.* 2012b; Costello *et al.* 2012, 2016; Rosenberg *et al.* 2018). Due to data limitations, all of these global assessment efforts have used forms of "catch-only" data-limited models (Free *et al.* (2020) and references therein). These models seek to infer the state of a fished population, for example biomass (B) relative to the biomass at maximum sustainable yield (B_{MSY}, the ratio B/B_{MSY} being a common measure of stock status), from characteristics of a fishery's catch history.

However, Free *et al.* (2020) demonstrated that the types of catch-only models used in these global assessment efforts can often produce both imprecise and biased estimates of current stock status in terms of B/B_{MSY}. While each of these prior efforts at assessing global fisheries using catch-only methods made important advances in our understanding of the global oceans, none have proven to be a consistently reliable means of estimating the state of the individual unassessed fisheries around the world. The types of global assessment efforts share a common underlying assumption; that globally available catch histories contain meaningful information on the state of the fished populations that they came from, and that this information can be revealed by the right model structure. Under this assumption, we can obtain better estimates of the state of unassessed fish populations simply by improving the catch-only models applied to the catch histories from these fisheries. However, further evaluation of the predictions made by these models (Pons *et al.* 2020; Free *et al.* 2020; Bouch *et al.* 2021) as well as works such as Pauly *et al.* (2013) and Branch *et al.* (2011) call into question this very idea.

In this paper we ask, how much can estimates from catch-only assessments be improved by moving beyond catch-only and augmenting catch histories with additional sources of broadly available fisheries data? We answer this question by aggregating a set of broadly distributed datasets that might be of use to global fishery assessment. We then used our new assessment method, sraplus, to evaluate the performance status estimates derived from different combinations of these broadly available data. We compared these results with estimates from the increasingly utilized (e.g. Palomares *et al.* (2020)) default settings of the catch-only version of CMSY described in Froese *et al.* (2017). We show that improving our understanding of the world's unassessed fisheries depends on a redoubled effort at global data collection and synthesis, not on incremental improvements to assessment methods based solely on catch histories.

Methods

The basis of this analysis is a new stock assessment software package we call sraplus (https://github.com/DanOvando/sraplus). sraplus is an extension of stochastic stock reduction analysis (SRA) (Kimura *et al.* 1984; Walters *et al.* 2006), which allows users to combine a biomass dynamics model with a variety of data sources (e.g. priors on recent stock status, or an index of abundance) in order to produce estimates of the state of a fishery over time. The key goal of sraplus is not substantial improvements in model fitting methods per say, but providing the ability to easily incorporate multiple kinds of fishery data potentially used in SRA-style analyses in a statistically rigorous manner. We paired various types of data with fisheries in RLSADB, and then used sraplus to generate predictions of B/B_{MSY} in the most recent year of each fishery across different combinations of data types. We then compared the predictions generated by sraplus using different types of data, along with predictions for the same fisheries generated by CMSY, to the reported values in RLSADB.

All analysis were conducted in the R programming language (R Core Team 2019). Model fitting was conducted using Rcpp (Eddelbuettel and François 2011) and Stan (Carpenter *et al.* 2017) implemented through Template Model Builder (Kristensen *et al.* 2016) by the tmbstan package (Monnahan and Kristensen 2018). The sraplus package is publicly available at github.com/danovando/sraplus, and all materials needed to fully reproduce this manuscript are available at github.com/DanOvando/assessing-global-fisheries. Here we describe the structure of the population model underpinning sraplus, the estimation models used, and the construction of priors used in this paper.

Data Sources

At its most "data-limited," sraplus can work as a catch-only model in the manner of Froese *et al.* (2017). These catch histories can then be augmented with prior information on stock status, derived from expert opinion, or using built-in prior generating models based on increasingly available sources of global fishery data: the Fisheries Management Index (FMI) (Melnychuk *et al.* 2017) and swept area ratio (SAR) (Amoroso *et al.* 2018) databases. The FMI database provides self-reported scores reflecting the strength of fisheries management in many fisheries around the world, which sraplus uses to construct informative priors on stock status, e.g. placing a higher probability that fisheries with high FMI scores have better stock status than those with low FMI scores. SAR is a measure of the intensity of trawl fishing within a particular area, which sraplus uses to construct an informative prior on the magnitude of fishing mortality. While fishery-independent surveys are becoming increasingly available (Maureaud *et al.* 2020), they are not yet sufficiently distributed or accessible to serve as a foundation for global assessments of unassessed fisheries. As an alternative, Rousseau *et al.* (2019) present a global reconstruction of fishing effort for countries around the world. We pair these effort reconstructions (both nominal effort and effective effort assuming an annual rate of efficiency increase of 2.6%) with the FAO's catch histories to construct a catch-per-unit-effort (CPUE) index, which we pass to sraplus along with catch histories and other available data to provide an estimate of stock status. In this way sraplus allows us to move from catch-only models to more conventional surplus production models in the manner of Winker *et al.* (2018) (Table.1).

Population Model

The core of sraplus is a Pella-Tomlinson (Pella and Tomlinson 1969) production model constructed in the manner of Winker *et al.* (2018). While models of these kinds abstract away many important details of fish biology and fleet behavior, they are the highest resolution model that the data types evaluated here will support.

The population growth equation is

$$f(x) = \begin{cases} B_{t+1} = \left(B_t + B_t \frac{r}{m-1}\left(1-\left(\frac{B_t}{K}\right)^{m-1}\right) - c_t\right)p_t, & \text{if } B_t > 0.25 \times K. \\ B_{t+1} = \left(B_t + \frac{B_t}{0.25\times K}\left(B_t \frac{r}{m-1}\left(1-\left(\frac{B_t}{K}\right)^{m-1}\right) - c_t\right)\right)p_t, & \text{otherwise.} \end{cases} \tag{1}$$

Where B_t is biomass at time t, K is carrying capacity ,r is the intrinsic growth rate, m is the scaling parameter that allows for the ratio of B_{MSY}/K to shift. When $m = 2$, B_{MSY} / $K = 0.5$. Lower values of m shift the production function left, higher values right. The shape parameter m is usually not reliably estimable given available data for surplus production models, however, Thorson *et al.* (2012b) provides estimates of the ratio of B_{MSY} to K for many fish taxa. For each stock we fix the shape parameter based on the distributions reported in Thorson *et al.* (2012b) for the genus of the species in question. We chose to fix the shape parameter at the mean stock-appropriate values from Thorson *et al.* (2012b) rather than estimating the shape parameter with an informative prior since there is so little information regarding the shape in the data considered. Attempts to estimate the shape parameter with priors from Thorson *et al.* (2012b) frequently resulted in poor model performance. $\boldsymbol{c}$ is a vector of catches, and $\boldsymbol{p}$ is vector of process errors. Growth rates can become unrealistically large when the population reaches low sizes under the Pella-Tomlinson model. We dealt with this problem by following the methods described in Winker *et al.* (2018) to reduce the production of the population when it falls below a threshold of 25% of carrying capacity.

We allow for process error p_t (in the manner of the stochastic stock reduction analysis (SRA) suggested by Walters *et al.* (2006)). Process error p_t is assumed to be log-normally distributed, such that

$$p_t \sim e^{N(-\sigma^2_{proc}/2, \sigma_{proc})} \tag{2}$$

where N is the normal distribution.

Estimation Model

All of our estimates are Bayesian in nature. sraplus can be run in two forms: either as a stock reduction analysis (SRA, Walters *et al.* 2006), or fit to an index of abundance (fishery dependent or independent) using Hamiltonian Monte Carlo with the No-U-Turn sampler (Hoffman and Gelman 2011). Unless there is an abundance index to fit to, the model runs as an SRA. A stock reduction analysis works by specifying prior distributions on population parameters and, critically, the recent state of the fishery. sraplus allows users to specify the most recent status in units of depletion, B/B_{MSY}, F, or F/F_{MSY}. We then sample from the prior distributions of the population model parameters and apply those to the production model, along with the catch history. Any run that results in the collapse of the population (catch greater than biomass in any time step) is immediately rejected. The remaining viable draws from the prior distributions are sampled in proportion to the supplied prior on recent stock status. All stock reduction analysis runs in our paper sampled 2,000 draws of the prior-predictive distribution from a total of $1e^6$ candidate draws. For our main sets of results (everything excluding the value of information analysis), the estimated parameters are r (with a prior distribution drawn from Fishlife, Thorson (2020)), K, σ_{proc}, and B_0 (initial depletion, B/K). q is also estimated when needed. σ_{proc} (process error) is estimated indirectly through the parameter γ, the ratio of process (*proc*) to observation (*obs*) error (σ_{obs}). See Table.S1 for prior distributions for each of these parameters.

When an index of abundance is available the model estimates the posterior probability distributions of the estimated and transformed parameters using Hamiltonian Monte Carlo implemented in Stan (Stan Development Team 2018) accessed through the `tmbstan` interface (Monnahan and Kristensen 2018). By default the model uses 2000 draws with a 1000 step warm-up and one chain. Any detailed fit for an individual fishery would likely use more draws and chains, but we verified that this sampling routine produced an acceptable tradeoff of speed and convergence criteria. The model fits to a direct estimate of abundance (e.g. a fishery independent survey or a standardized catch-per-unit-effort index), the likelihood is

$$log(a_t) \sim N(f_{pt}(r, K, m, B_0, \boldsymbol{p}, \boldsymbol{c}) \times q, \sigma_{obs})$$

where a_t is the observed abundance index and f_{pt} is the Pella-Tomlinson production model (Equation.(1)). When an effort index is available, sraplus constructs an index of abundance based on the catch and effort data. Rousseau *et al.* (2019) measure an index of abundance as catch divided by their effort index, either nominal or effective (assuming the 2.6% annualized technology creep). Treating this raw effort data as the denominator in the CPUE calculation assumes that every increase in fishing effort translates to a commensurate increase in fishing mortality. When effort increases dramatically above historic levels, this can create a CPUE index that decreases faster than the true population. This is due to the fact that in reality the marginal fishing mortality produced by increasing units of effort decreases as effort approaches infinity (since the realized fishing mortality rate must be between between 0% and 100%). To accommodate this, we generate a catch per effective harvest rate index of abundance, as

$$cpue_t = \frac{c_t}{(1 - e^{-F_t})}$$

$$F_t = q_t E_t$$

Where q_t can has a technology creep component τ

$$q_t = q_{t-1} \times (1 + \tau)$$

We then fit to the index of abundance per

$$log(cpue_t) \sim N(f_{pt}(r, K, m, B_0, \boldsymbol{p}, \boldsymbol{c}), \sigma_{obs})$$

CMSY

In addition to the results from sraplus, we include a set of results produced by the default settings of the CMSY method (Froese *et al.* 2017). For computational efficiency, we used a ported version of the CMSY model available at https://github.com/DanOvando/portedcmsy. The only modification made is to convert the underlying population model to C++ for faster computation. For each stock we used all the default options and priors provided and generated by CMSY, in the same manner as Palomares *et al.* (2020), except for resilience, which was pulled from the vulnerability scores from FishBase accessed through `rfishbase` (Boettiger *et al.* 2012). Vulnerability scores greater than 66 were scored as low resilience, between 33 and 66 medium resilience, and lower than 33 high resilience.

Priors

Prior Predictive Tuning

In the absence of any data to fit to, sraplus works by assuming that we know current stock status, and then finds feasible parameters to satisfy that belief given a catch history, life history priors, and model structure. This creates a problem for the Bayesian nature of our analysis. Consider a production model with two parameters, a growth rate r and a carrying capacity K. Once we specify prior distributions on r and K, and then apply these distributions to our model (the shape of the production function along with the catch histories), we have implicitly provided a prior on the status of the stock in all time periods, since each unique combination of r and K together with the model and the catch history produces a deterministic stock status in each time step. Doing so places two priors on recent stock status: one implicit prior through the population parameter priors, and one explicit through the users perception of recent stock status, creating a problem termed Borel's Paradox (See Poole and Raftery (2000) and references therein for a discussion of Borel's Paradox in a fisheries context).

This may seem like an academic concern, and indeed in our experience when the data are sufficiently informative the Bayesian version of our model subject to Borel's paradox produces effectively identical results to those produce by the same model fit by maximum likelihood. However, Borel's Paradox poses a particular problem when there are no data to fit to(i.e. when the model is simply filtering through prior distributions in the manner of a traditional SRA) due to the fact that there are more parameter combinations that allow for a fishery to be relatively unexploited than for a fishery to be close to collapse (but never actually collapsed, i.e. predicted biomass less than observed catch). In this context Borel's Paradox causes the posterior distribution of stock status to be positively biased relative to the supplied prior (although combined with other modeling choices can result in a net negative bias in stock status, Free *et al.* (2020)). This process can also make it easy for users to accidentally supply very informative priors on stock status, without realizing that choices relating to population biology priors that may appear independent of stock status are in fact dictating the posterior distributions of stock status resulting from the SRA algorithm.

We use an approximate solution to this problem here, similar in spirit to Bayesian melding (Poole and Raftery 2000). Our solution amounts to a two-step sampling-importance-resampling (SIR) algorithm. We first run the standard SRA algorithm as described in the Estimation Model section of the methods. We then break the resulting draws into bins based on terminal stock status, and calculate the mean probability density p (defined by the prior distributions of estimated parameters) of each bin.

$$p(bin_i) = \frac{1}{N_i} \sum_{n=1}^{N_i} p(b_{n,i})$$

We then divide the mean probability density of bin i evenly among each of the draws within that bin n

$$p(n_i) = \frac{p(bin_i)}{N_i}$$

And we then perform a second SIR algorithm but now sampling each observation n_i in proportion to $p(n_i)$.

The net result of this is that it allows users to place an explicit prior on stock status, and then adjust their priors on life history parameters to reflect this prior. While the range of possible life history values supplied still influences stock status under this approach, this prior predictive tuning process makes the resulting priors more consistent with explicit priors on recent stock status supplied by the user. Users can turn this functionality off and instead base priors on stock status primarily on life history. See Supplementary Information for a detailed explanation of this problem and our solution.

Priors Informed by Outside Data

Along with allowing users to supply their own priors, the sraplus package contains three built-in methods for converting information on stock status from additional outside data into a form usable as a stock status prior by sraplus. We paired data on catch histories, swept area ratio, and Fisheries Management Index with estimates of stock status from the RLSADB. We then trained a regression of the general form $log(status) \sim N(variable, \sigma)$ for each of these three data types. Given values of these variables for a new fishery, sraplus uses the fitted model to generate posterior predictive distributions of stock status based on these data, which can then be used as priors on stock status by sraplus for new fisheries. For example, given data on SAR or FMI scores, together with a catch history, sraplus uses these regressions to convert those SAR and FMI values into priors on B/B_{MSY} or F/F_{MSY} in the most recent year of the fishery usable by sraplus (See Supplementary Information). All prior regression models were tested by out-of-sample predictive power, and where competing models were considered the final model was chosen by leave-on-out validation (Vehtari *et al.* 2017). The final models are intended as a reasonably robust means of translating available data (catch histories, FMI, and SAR values) into a form usable by sraplus. For all results presented in this paper we used these data to provide priors on F/F_{MSY}, as we found clearer predictive relationships and subsequent model performance between catch, FMI, and SAR values and F/F_{MSY} than we did for B/B_{MSY}.

Assessing Performance

Simulation testing can be preferable in many ways to comparison against model outputs. However it is not possible to simulation test the value of information derived from empirical relationships between variables such as fisheries management strength and fishery outcomes, which our study depends on. As such we assess model performance through comparison to best available estimates of stock status available in RLSADB. We based this test on 393 stocks from RLSADB, covering 19 broad taxonomic groups, with estimates of B/B_{MSY} and greater than 25 years of continuous catch history. B/B_{MSY} values from RLSADB are themselves estimates, not data, but they are the best available information on global stock status. We then paired the catch histories for these RLSADB stocks with regional-level SAR, FMI, and effort data. Our methods approximated a regional-level assessment exercise, where data beyond catch histories are available at regional levels, but not for specific fisheries. We also estimated B/B_{MSY} values of our candidate RLSADB stocks by using sraplus to fit to an abundance index drawn directly from RLSADB itself "RLSADB Index," as a measure of the ability of models like sraplus if given perfect information

We then fit a range of models utilizing different combinations FMI, SAR, and effort data, along with CMSY, and a set of runs fit to the RLSADB Index (Table.2). We assessed model performance using three metrics: median percent error (MPE, a measure of bias), median absolute percent error (MAPE, a measure of accuracy), and classification accuracy. Classification accuracy was calculated as the proportion of times that use of a given combination of data resulted in a stock being classified into the correct FAO status classification (one of underfished, maximally sustainably fished, and overfished). This experiment structure allows us to assess the ability of catch-only models to estimate the B/B_{MSY} values reported in RLSADB, and evaluate how much these estimates can be improved by augmenting the catch histories with additional sources of information. Our benchmark model is a simple "Guess." Guess assigns each stock a random B/B_{MSY} of 1.6, 1, or 0.4, corresponding to the middle value of the FAO status bins of underfished ($B/B\ MSY \ \geq 1.2$), maximally sustainably fished ($0.8 \geq B/B\ MSY \ < 1.2$), overfished ($0 \geq B/B\ MSY \ < 0.8$). We performed a matching analysis measuring performance in estimating F/F_{MSY} as well, with results presented in the SI.

Value of Information Calculations

We performed a value of information assessment to determine what types of data may be most beneficial to acquire at a global scale if we are to improve our knowledge of the state of global fisheries. The value of information analysis was performed by using sraplus to generate estimates of B/B_{MSY} for stocks in RLSADB, and comparing the estimated values to the values reported in RLSADB. There are too many combinations for us to run the full expansion of possible parameter states. To resolve this we generated fits for 3000 combinations of a RLSADB stock and available data. For any one draw, we randomly sampled a RLSADB stock and a list of available data and data quality. For example, we might sample stock A with information on recent fishing mortality rates for the first iteration, and stock A again for the second iteration but now with information on recent fishing mortality rates and a recent index of abundance. The result is a set of model performance estimates where the characteristics of the stock and the data made available to the model are randomized. We then measured the value of information as the average reduction in root mean squared error (RMSE) in B/B_{MSY} over the most recent five years of the fishery (in order to evaluate the ability of the model to capture recent trends as well as the most recent value), resulting from use of different kinds of data.

We considered the value of information of having data on: the most recent B/B_{MSY}, treating initial state of the population as one of unfished ($B_0 = 1$), known (B_0 taken from stock assessment), or estimated based on initial shape of the catch history, F/F_{MSY} values in the most recent year, over the last five years, and over the complete time series, and an abundance index spanning the complete, most recent half, or most recent quarter of the time series. We considered the value of information of a longer time series of F/F_{MSY} than B/B_{MSY} to consider the potential of augmenting catch-only models with additional data that can inform F/F_{MSY}, such as length-composition data, that might be available over the history of the fishery.

Case Study

We also ran a case study demonstrating how different kinds of data led to different conclusions about stock status. We selected 26 stocks for which we have stock specific FMI and SAR scores. We then paired effort data at the resolution of year, country, and FAO statistical area from Rousseau *et al.* (2019) to each stock. As a benchmark, we first estimated stock status for these case study fisheries using the default settings of the CMSY (Froese *et al.* 2017) method, as this has become one of the most widely used catch-only models currently available. We then used sraplus to generate estimate of stock status based solely on catch heuristics, taking into account the prior-predictive tuning implemented by default in sraplus. We also used stock-specific data on SAR and FMI to generate priors on F/F_{MSY} for each of the stocks, which were then passed to sraplus. Lastly, we used the reconstructed effort data (Rousseau *et al.* 2019) to create an index of abundance for each stock, and estimated stock status by fitting to this index while using priors on fishing mortality rates informed by each stock's FMI and SAR values. This case study represents a more "localized" assessment, where external data sources (FMI and SAR) are available at the stock level, rather than at the regional level. We summarized the performance of each model fit in the case study based on root mean squared error (RMSE).

Results

Case Study

Nearly all of the fisheries used in this case study have F/F_{MSY} values less than one, and most have B/B_{MSY} values greater than one. Both the catch heuristic implemented in sraplus and CMSY, using no stock-specific information but the catch history (which defines the prior on K and stock depletion, and hence the estimate of B/B_{MSY}) and the estimated resilience of the target species used to provide a prior on r (based on Fishlife, Thorson (2020)), badly miss this trend, predicting instead that nearly all the fisheries in this group are currently overfished and experiencing overfishing (Fig.1). Overall, models informed by the SAR and FMI values performed the best in terms of B/B_{MSY} (RMSE = 0.89). The ranking of performance by data sources used was the same for F/F_{MSY}, though the range of values was much higher, with the SAR and FMI based result producing an RMSE of 0.39 while the catch-only case studies produced F/F_{MSY} RMSE values greater than 1 (Fig.1). However, this improved performance of the SAR and FMI data is to be expected as the prior-generating models were trained on data from RLSADB, and therefore this is likely an optimistic assessment of the performance of these prior-generating models when applied to new fisheries. Fitting to an index of abundance created by regional effort data in fact produces a slightly poorer fit in B/B_{MSY} and F/F_{MSY}. This is likely because a number of stocks in this case study set have very low catches (relative to historic highs) but high biomass values, creating a mismatch between the CPUE trends based on the effort and catch data and actual biomass. However none of the models fit to the four case-study data combinations were able to explain much of the variation in B/B_{MSY}, though the models performed better in some cases for predicting F/F_{MSY}.

Performance of Regional Fishery Assessments

We first considered the performance of sraplus when the model was given a perfect index of abundance from RLSADB (RLSADB Index) for each stock in RLSADB it was tasked with assessing. Assessment models in RLSADB are typically much more structurally complex than the simple biomass dynamics used in sraplus, and so this "perfect information" test tells us how much of a penalty in model performance we are likely to pay due to model misspecification alone. The sraplus estimates of B/B_{MSY} resulting from fitting directly to the abundance indices from RLSADB were relatively accurate and unbiased at a macro level (MPE 14%,MAPE 29%, accuracy = 69%, Table.3, Fig.2-4). This exercise tells us that given sufficiently high quality index of abundance, a surplus production model such as sraplus is reasonably capable of reproducing the global state of fisheries as understood from formally assessed fisheries.

We next assessed the ability of FMI, SAR, and effort data to improve estimates of global stock status beyond those derived from catch-only methods. Many of the datasets used produced similar levels of bias as the RLSADB data (e.g. FMI, SAR, Nominal CPUE fits), though notable the "Guess" method actually performed near the top in terms of bias. However, this is somewhat an artifact of the data. The status of most stocks in RLSADB is relatively good, with recent B/B_{MSY} values generally near one. As the mean value of the "Guess" method is one, on average the "Guess" model is an unbiased but imprecise measure of stock status in RLSADB (Table.3).

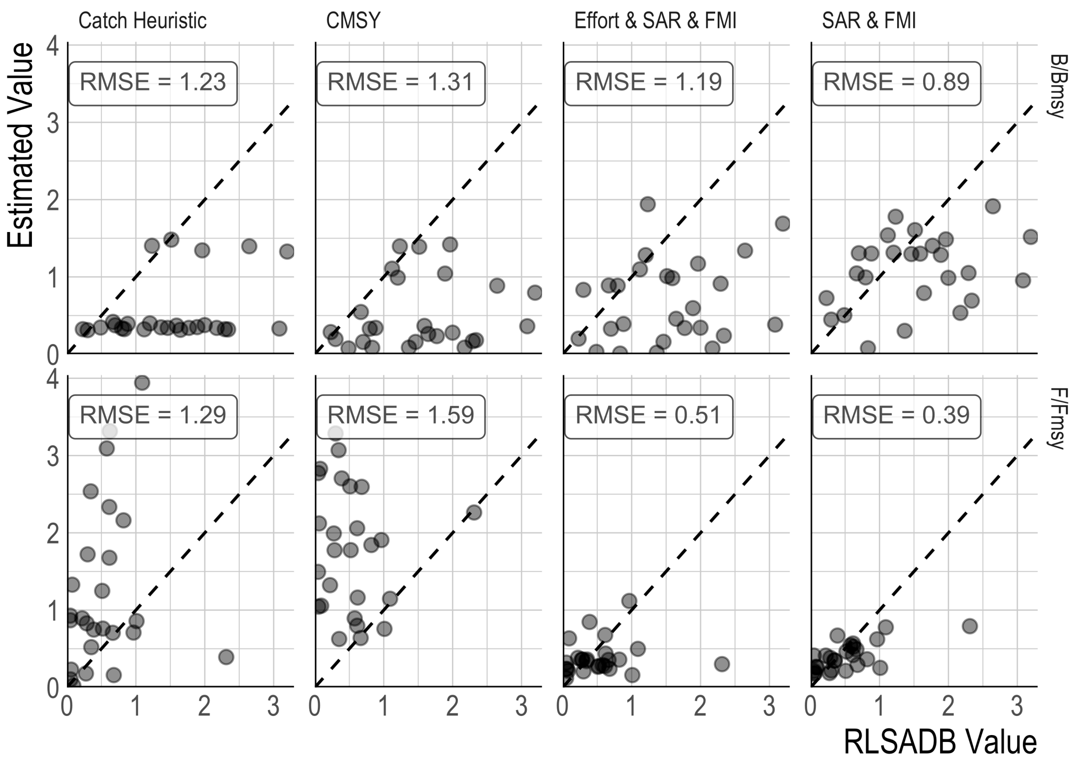

Figure 1: RLSADB values of B/B_{MSY} and F/F_{MSY} (x-axes) for case study fisheries plotted against estimated values (y-axes) using CMSY (Froese *et al.* 2017), catch heuristics, priors informed by stock-specific Fisheries Management Index (FMI) and swept area ratio (SAR) scores, and an abundance index based on reconstructed effort (Effort) trends assuming a rate of technological increase of 2.6%. Each point is a stock in the RLSADB. Black dashed line shows the 1:1 relationship.

Focusing on MAPE (our measure of error rather than bias) and classification accuracy, the error of the models jumps dramatically as soon as data other than the RLSADB abundance indices are used, to a minimum value of 47% and a maximum of 68%. The mean accuracy of the sraplus models across all non-RLSADB data fits was 43%. Note that there are only three bins in the FAO stock status classifications, and as such our "Guess" model has a mean accuracy of 34%. This means that the accuracy of our models designed as a proxy for a global assessment process were across all non-RLSADB index data fits 25% more accurate than a random guess, certainly an improvement, but on average assigned fisheries to the wrong FAO status bin 57% of the time.

Looking geographically we found a similar pattern of a rapid decrease in performance for models besides those fit to the RLSADB Index. Across the models, performance was not consistent in space: use of different data performed best or worst for different FAO regions. Models fit to nominal CPUE data substantially overestimate stock status in the Mediterranean, while models based on data using effective CPUE perform better in that region (but worse in others) (Fig.3). We find similarly inconsistent performance for both bias (Fig.2) and accuracy (Fig.4). Overall, while some data sources performed slightly better than others by some metrics in some places, no models using any non-RLSADB index data were able to capture the overall state or geographic distribution of stock status represented in RLSADB in a consistent manner. Performance in estimating F/F_{MSY} was similarly variable and poor, with the exception that the default settings of CMSY performed much more consistently poorly in terms of F/F_{MSY} than B/B_{MSY} (due to systemic overestimating of F/F_{MSY}, see SI).

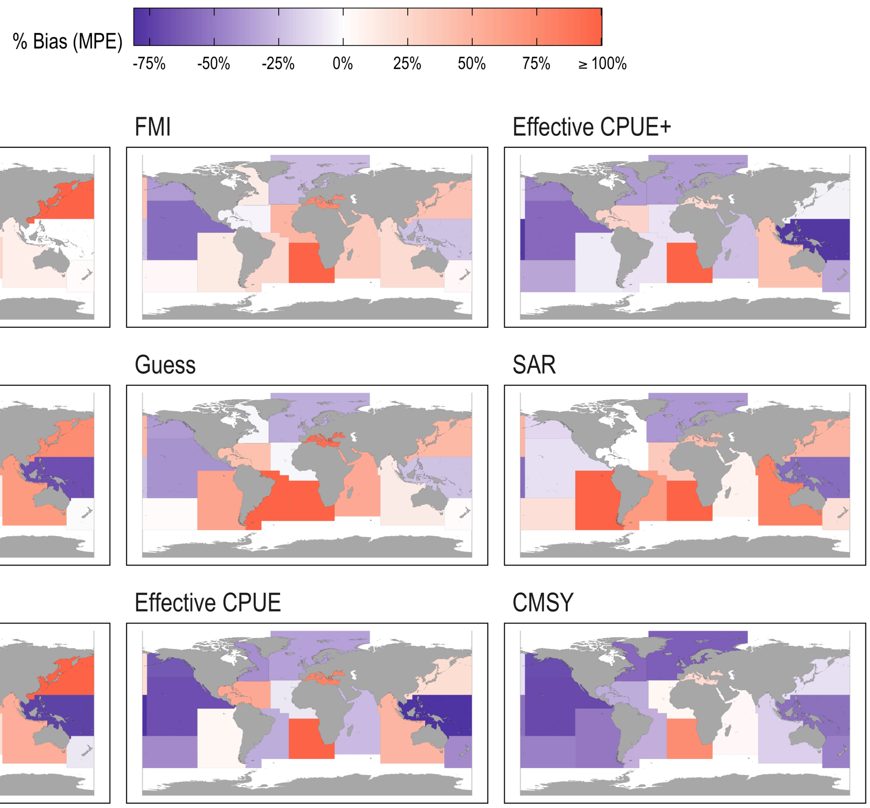

Figure 2: Median percent error (MPE, predicted relative to observed) in most recent B/B_{MSY} by FAO statistical area from different data sources. Data source panels are ordered in ascending (starting from top left) mean MPE at the FAO region level. RLSADB Index refers to catch and abundance index drawn from RLSADB. Effective CPUE refers to an index of abundance based on reconstructed effort data. Effective CPUE+ uses CPUE along with Fisheries Management Index (FMI) and/or swept area ratio (SAR) data. For both CPUE series 'nominal' assumes a 0% technology creep, for 'effective' a 2.6% technology creep is assumed. FMI uses FMI scores to develop a prior on recent fishing mortality rates, SAR does the same but based on swept area ratio. CMSY uses the methods from Froese et al. 2017 (Froese *et al.* 2017). Guess assigns a random recent B/B_{MSY} of 0.4,1, or 1.6.

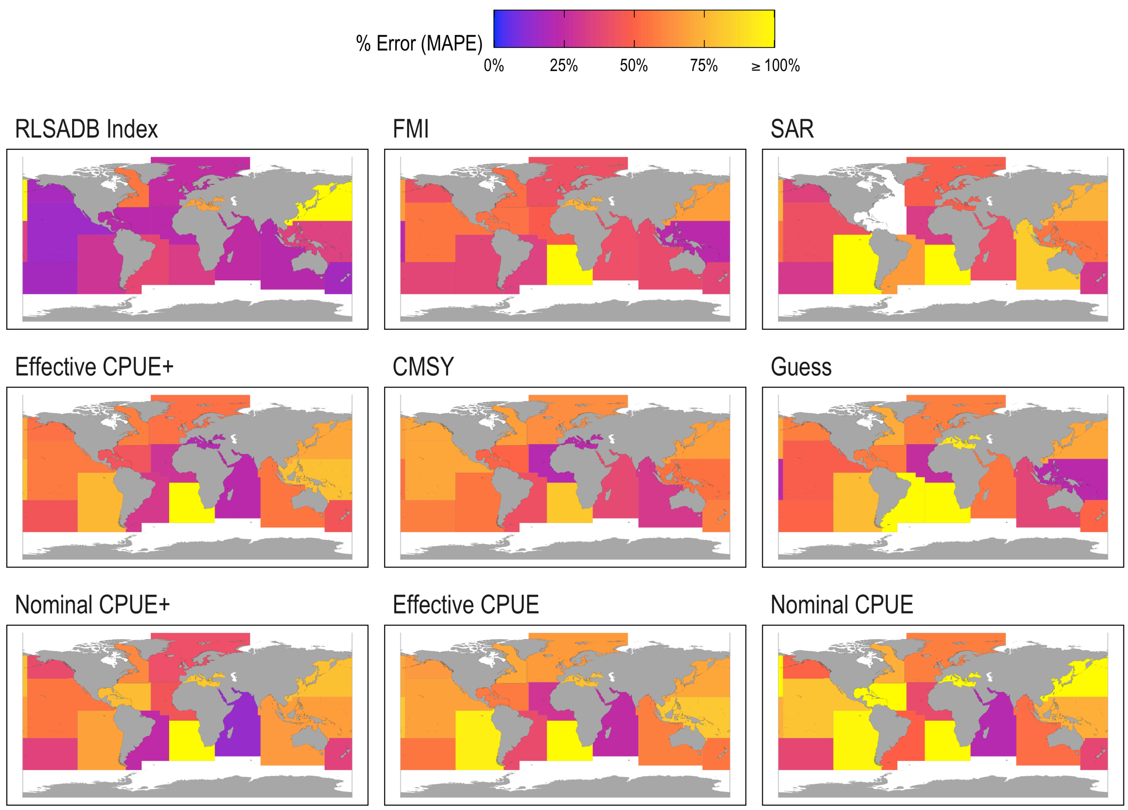

Figure 3: Median absolute percent error (MAPE) in most recent B/B_{MSY} by FAO statistical area from different data sources. Data source panels are ordered in descending (starting from top left) mean MAPE at the FAO region level. RLSADB Index refers to catch and abundance index drawn from RLSADB. Effective CPUE refers to an index of abundance based on reconstructed effort data. Effective CPUE+ uses CPUE along with Fisheries Management Index (FMI) and/or swept area ratio (SAR) data. For both CPUE series 'nominal' assumes a 0% technology creep, for 'effective' a 2.6% technology creep is assumed. FMI uses FMI scores to develop a prior on recent fishing mortality rates, SAR does the same but based on swept area ratio. CMSY uses the methods from Froese et al. 2017 (Froese *et al.* 2017). Guess assigns a random recent B/B_{MSY} of 0.4,1, or 1.6.

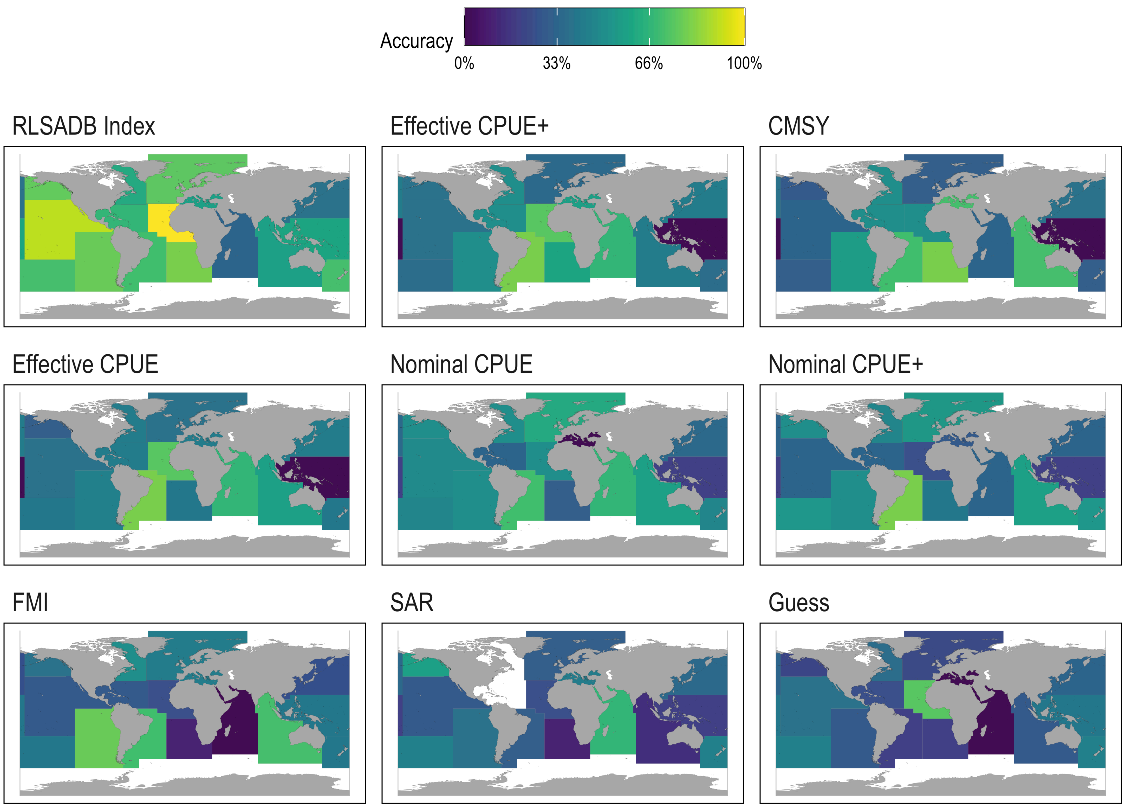

Figure 4: Mean classification accuracy (assignment to FAO stock status category) by FAO statistical area arising from different data sources. Data source panels are ordered in descending (starting from top left) mean accuracy at the FAO region level RLSADB Index refers to catch and abundance index drawn from RLSADB. Effective CPUE refers to an index of abundance based on reconstructed effort data. Effective CPUE+ uses CPUE along with Fisheries Management Index (FMI) and/or swept area ratio (SAR) data. For both CPUE series 'nominal' assumes a 0% technology creep, for 'effective' a 2.6% technology creep is assumed. FMI uses FMI scores to develop a prior on recent fishing mortality rates, SAR does the same but based on swept area ratio. CMSY uses the methods from Froese et al. 2017 (Froese *et al.* 2017). Guess assigns a random recent B/B_{MSY} of 0.4,1, or 1.6.

Value of Information

Having access to estimates of F/F_{MSY} reduced model error in proportion to the number of years for which F/F_{MSY} values are available. Interestingly though, having access to only an accurate estimate of F/F_{MSY} in the most recent year was extremely informative, reducing error on average by 15%, on par with an estimate of recent B/B_{MSY} itself (Fig.5). While having access to complete index of abundance, such as a fishery independent survey, was on average able to reduce error relative to a baseline catch-only heuristic, using only the most recent quarter of the available abundance index actually increased error on average, due to the lack of historical context for recent trends in abundance.

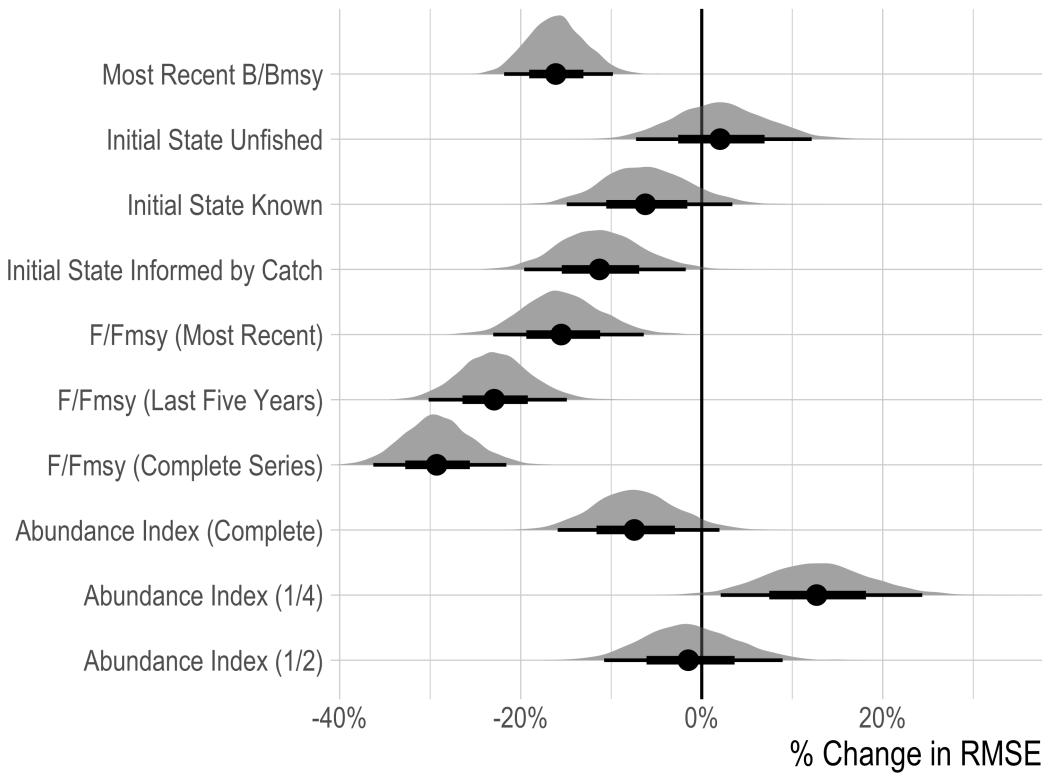

Figure 5: Posterior probability distributions of estimated effect of different data types on root mean squared error (RMSE) of B/B_{MSY} in the most recent 5 years of data available for each model fit. Distribution is full posterior probability distribution. Point is median, thicker black section inner 66th quantile of the posterior, the thinner black line the 95th. Change is relative to the mean performance of a catch-only heuristic model.

Discussion

Global-level assessments of fish populations are critical for guiding management agendas for the world's oceans, and tracking indicators such as the United Nations Sustainable Development Goals. The hope of efforts using catch-only models to estimate the status of global unassessed fisheries is fundamentally that we can learn something meaningful about the state of a fished population simply by knowing something about its catch history and life history. While in some cases the addition of globally available data such a FMI scores, SAR values, or effort reconstructions, provided value above and beyond catch histories alone (Fig.1), at the global level models fit using each of the available datasets, besides the RLSADB-derived indices, generally produced biased and imprecise estimates of fish stock status (Table.3). The simplicity and low data requirements of catch-only models are understandably appealing to many users, but our results highlight that without high-quality local data these methods can provide highly imprecise and biased results. Broad and uncritical application of these methods that frequently result in incorrect classification of stock status can be detrimental to both ecosystems and livelihoods.

Our claim is that achieving substantial progress in assessing the state of global fisheries will require improved data and capacity building. This claim cannot be definitively proven; it is possible that some yet-to-be discovered model will provide a reliable means of dramatically improving the accuracy of estimates produced by catch-based models using current broadly-available datasets. Basic logic tells us though that such a model would have to depend on strong assumptions, new data, or empirical estimation. Catch is a function of catchability, effort, and biomass, so given just catch we have one equation and three unknowns. Structurally separating out biomass (and subsequently estimating reference points etc.) requires then either data or strong assumptions on these other variables across a sufficiently representative timespan. Alternatively, empirical relationships could be developed linking attributes of catch histories to observed estimates of stock status (e.g. Costello *et al.* 2012). However, these methods require that these empirical relationships both exist and are reliably preserved in fisheries that lack formal assessments. Numerous groups made up of highly skilled scientists have attempted to resolve these fundamental challenges in various ways, but external efforts to validate the performance of these methods have consistently found them to perform generally poorly when applied in bulk (Pons *et al.* 2020; Free *et al.* 2020; Bouch *et al.* 2021). This is not intended as an indictment of those past efforts, but, we believe, a reflection on the limited amount of broadly applicable predictive power that exists within catch histories alone. Methods tested in those papers have been demonstrated to perform relatively well when given specific data or priors informative to specific fisheries. We revisited this task in this paper, and found that even when augmented with other broadly available datasets, we were unable to produce consistently accurate estimates of stock status around the globe. For these reasons we believe that improvements to our understanding of the state of global unassessed fisheries will come from targeted use of existing but underutilized data, and expanding the collection of high-priority data around the globe, rather than the development of new modeling methods alone.

What quality of assessment is needed and what constitutes a meaningful improvement in assessment quality depends on the needs of those using the assessment outputs. It may be that for particular regions, species, or uses the results presented here or in other past global analyses are sufficiently accurate. Catch-only based estimates of measures such as MSY are likely to be much more robust than reference points such as B/B_{MSY} (Martell and Froese 2013). Where catch-only methods are used for estimates of B/B_{MSY} or F/F_{MSY}, it is critical that users are aware of the broadly demonstrated poor performance of these methods in the absence of highly reliable fishery-specific information. In some instances using the data presented here did provide some improvement over use of catch-only style assessment methods; the difficulty comes in attempting to apply data types uniformly across the globe. While it is unreasonable to expect models based solely on global-scale data to be able to perform as well as detailed stock assessments reported in RLSADB, or that data-limited methods would perform well for every individual stock, our hope would be that a data-limited approach based on globally available data sources would be able to correctly capture general patterns in stock status in time and space. The overall poor performance of the models tested here, in terms of estimating both B/B_{MSY} and F/F_{MSY}, shows that improvements in estimates of global stock status depend on improvements in the quality and use of data themselves.

Our fits to the RLSADB data provide a useful diagnostic of the degree to which model mispecification might explain the poor performance of catch-only models. Catch-only models generally employ some form of biomass dynamics model, such as the Schaefer surplus production model (Schaefer 1954). These models are generally much simpler than the typically age-structured population models underpinning the assessments in the RLSADB. While we found that fitting the Pella-Tomlinson model employed in sraplus to an index of abundance drawn directly from the RLSADB provided much improved and reasonably accurate estimates of stock status relative to catch-only approaches (Table.3), even with the index of abundance sraplus still had an average MAPE of 29%. This discrepancy may provide a measure of relatively insurmountable error resulting from model misspecification that inclusion of additional data may not be able to overcome. For example, many assessments in the RLSADB calculate B/B_{MSY} based on spawning stock biomass. However, surplus production models condition the size and state of the population on the catch, which is by definition a measure of the vulnerable biomass. Vulnerable biomass may not reliably track spawning biomass in highly selective fisheries, potentially resulting in model biases as the surplus production model is not capable of separately tracking spawning and vulnerable biomass.

We chose to test the performance of methods against estimated values reported in RLSADB. A reasonable critique of this choice is that unassessed stocks, on which these methods would actually be used, are likely to have vastly different dynamics than the heavily managed fish populations represented in RLSADB, in addition to the errors in the RLSADB estimates themselves. For this reality to change our results though would require that the methods tested here have massively lower bias and higher accuracy for unassessed fisheries than RLSADB stocks. Free *et al.* (2020)'s simulation based results of these same types of models suggests that this is unlikely to be true. However, while many assessed fisheries have mechanisms to actively manage catches, catch histories in unassessed and unmanaged fisheries may be more likely to reflect underlying stock status than those in relatively well managed fisheries, though any predictive relationship will still depend on a host of factors such as changes in effort. This could be empirically tested by considering the performance of catch-only models on data from earlier in the history of fisheries in the RLSADB, though we did not do so here as we do not have historic estimates of FMI or SAR values.

Our results do not imply that the kinds of broadly available data presented here are not valuable under the right conditions. The FMI and SAR based priors are an improvement over catch-only models in applicable situations (i.e. those that sufficiently resemble the data on which the regressions were trained, Fig.1). Effort data such as those reconstructed by Rousseau *et al.* (2019) can help distinguish between regions with similar catch histories but different large-scale effort trajectories, and may be quite useful as indices of abundance for areas with good knowledge of rates of evolution of fishing technology and a broadly selective fishing fleet. Despite not adding a great deal in terms of performance at the global scale, swept-area-ratio was the strongest predictor of F/F_{MSY} of any of the datasets we explore on an individual stock basis, with a Bayesian R^2 value of 0.43 (see SI). But, we must simultaneously consider data quality and resolution: applying one SAR value to all stocks in a region, even if that SAR value can provide valuable information for a subset of fisheries, causes inaccurate estimates of stock status when applied too broadly. Our analysis does not show that the data considered here are without value, but that attempting to indiscriminately apply these data to all areas of the globe results in meaningfully incorrect estimates of stock status for regions whose nature does not match the assumptions needed to apply these data sources. We found that performance of different data sources varied widely both within and among regions (Fig.1, Fig.2-4). Some of this variation is likely simply due to low sample sizes of assessed stocks reported in the RLSADB in some regions. But, other explanations for differences in performance among stocks and regions may be an interesting area for future research. For example, it may be that stocks in some regions are more suitable to being represented by surplus production models than others.

Our value of information analysis also shows the high utility of having access to even a recent snapshot of

F/F_{MSY} (Fig.5). Swept area ratios, Fisheries Management Index scores, or other similar metrics can be used to construct fishery-specific priors on fishing mortality rates, though care must be taken in applying them at the appropriate spatial resolution. Another avenue would be to prioritize the development of a global repository for length and age composition data. Given appropriate conditions, these length measurements can be used to estimate local fishing mortality rates (Hordyk *et al.* 2016; Rudd and Thorson 2017; Prince and Hordyk 2019). While length-based assessments come with a host of assumptions and potential pitfalls, properly implemented in some fisheries with appropriate life histories these methods may provide an overlooked source of information on fisheries at a global scale, at least as an improvement over relying on catch-only or regional proxies alone. Such a database could be used to construct stock or stock complex specific priors on fishing mortality for particular regions around the globe, which could meaningfully improve our understanding of global fisheries, particularly when paired with catch data and where possible indices of abundance (Thorson and Cope 2015; Rudd and Thorson 2017).

We must also prioritize collection and curation of fish population survey data worldwide. Repositories of fishery-independent survey data would be immensely beneficial, such as those maintained by FishStat (www.FishStats.org). Recent research confirms that there are bottom trawl data to support analysis of biomass-trends since 2001 and potentially earlier in many regions (Maureaud *et al.* 2020), and survey data are available for more stocks than have stock assessments. Effort reconstructions such as those utilized here may help create fishery-dependent abundance indices in some instances, and going forward datasets such as those compiled by Global Fishing Watch (https://globalfishingwatch.org/) in combination with the reconstruction approaches of Rousseau *et al.* (2019) might allow us to construct and use timeseries of fishing effort specific to particular areas, fleets, and species complexes. However, our value of information exercise indicates that we may have to wait many years for new surveys to provide substantial improvements in status estimates (Fig.5).

Expanded training of fisheries scientists around the globe is another critical need. Even were we to dramatically expand the amount and types of data available for global assessment, individual fisheries and regions will need to make informed decisions about which sources of data may be applicable and which not, and to critically evaluate the results of any model based on local expertise. This is why stock assessments even in data-rich fisheries are not an automated process; the real challenge is often not in fitting a model to data but in understanding how best to use the data and the quality and limitations of the model used. Empowering a global network of fisheries scientists through training and peer-support would help local experts make the most of available data, ensure the reliability of newly collected data, and improve the interpretation of assessment results. We may also need to acknowledge that in some fisheries reliable estimates of stock status relative to MSY based reference points are simply not possible (or perhaps desirable), and instead rely on more precautionary or empirical management management measures such as spatial closures, size restrictions, and indicator based harvest strategies (ideally tested through management strategy evaluation) (Dowling *et al.* 2015; Fulton *et al.* 2016; Prince and Hordyk 2019).

The coming decades are a critical time for the future of fisheries and ocean health. Achieving the United Nations Sustainable Development Goal 14 for the conservation and sustainable use of the world's oceans depends on our ability to effectively assess the status of fish stocks around the world. The RAM Legacy Stock Assessment Database combined with the FAO's expert elicitation of status for select stocks have dramatically improved our understanding of global fisheries in recent years. However, this process still leaves a substantial number of fisheries and proportion of global catch unassessed. Numerous catch-based data-limited approaches have attempted to fill that gap, and while these efforts have advanced our knowledge and interest in unassessed fisheries, none have yet been able to provide a solution to this problem which has proven to be unbiased and sufficiently precise at a global or regional level.

The lack of strong information on stock status within catch histories alone means that differences in models and assumptions between catch-based assessment efforts can produce starkly contrasting conclusions on global stock status, leading to debates that are inconclusive as they are inherently driven by assumptions. The FAO is leading efforts to increase technical capacity and monitoring and evaluation infrastructure to improve fisheries management in places with limited data. Such projects stand to provide a better picture of fishery status at global and local scales, furthering our ability to meet the UN SDG targets. Our results emphasize the urgency and rationale for building the infrastructure and capacity that can lead to better marine resource management globally (Costello *et al.* 2020) Achieving meaningful improvements in the assessment and management of global unassessed fisheries will depend on expanded collection of targeted data types, active management, and local capacity building.

Data Availability Statement

All data and materials needed to reproduce our results are publicly available or queried by code available at https://github.com/DanOvando/assessing-global-fisheries.

Acknowledgements

Funding for this work was provided by the Food and Agriculture Organization of the United Nations (FAO). We thank participants of the FAO "Methods for Global Assessment" workshop held in Rome, Italy February 2019 for helpful feedback on this project. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of the FAO, NOAA or the Department of Commerce.

References

Amoroso, R.O., Pitcher, C.R., Rijnsdorp, A.D., et al. (2018) Bottom trawl fishing footprints on the world's continental shelves. *Proceedings of the National Academy of Sciences*, 201802379.

Boettiger, C., Temple Lang, D. and Wainwright, P. (2012) Rfishbase: Exploring, manipulating and visualizing FishBase data from r. *Journal of Fish Biology*.

Bouch, P., Minto, C. and 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**, 264–276.

Branch, T.A., Jensen, O.P., Ricard, D., Ye, Y. and Hilborn, R. (2011) Contrasting Global Trends in Marine Fishery Status Obtained from Catches and from Stock Assessments. *Conservation Biology* **25**, 777–786.

Carpenter, B., Gelman, A., Hoffman, M.D., et al. (2017) *Stan* : A Probabilistic Programming Language. *Journal of Statistical Software* **76**.

Cope, J.M., Thorson, J.T., Wetzel, C.R. and DeVore, J. (2015) Evaluating a prior on relative stock status using simplified age-structured models. *Fisheries Research* **171**, 101–109.

Costello, C., Cao, L., Gelcich, S., et al. (2020) The future of food from the sea. *Nature*, 1–6.

Costello, C., Ovando, D., Clavelle, T., et al. (2016) Global fishery prospects under contrasting management regimes. *Proceedings of the National Academy of Sciences* **113**, 5125–5129.

Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O. and Lester, S.E. (2012) Status and Solutions for the World's Unassessed Fisheries. *Science* **338**, 517–520.

Dowling, N.A., Dichmont, C.M., Haddon, M., Smith, D.C., Smith, A.D.M. and Sainsbury, K. (2015) Guidelines for developing formal harvest strategies for data-poor species and fisheries. *Fisheries Research* **171**, 130–140.

Eddelbuettel, D. and François, R. (2011) Rcpp: Seamless R and C++ integration. *Journal of Statistical Software* **40**, 1–18.

FAO (2020) *State Of World Fisheries And Aquaculture 2020: Sustainability in action.* FOOD & AGRICULTURE ORG, S.l.

Free, C.M., Jensen, O.P., Anderson, S.C., et al. (2020) Blood from a stone: Performance of catch-only methods in estimating stock biomass status. *Fisheries Research* **223**, 105452.

Froese, R., Demirel, N., Coro, G., Kleisner, K.M. and Winker, H. (2017) Estimating fisheries reference points from catch and resilience. *Fish and Fisheries* **18**, 506–526.

Fulton, E.A., Punt, A.E., Dichmont, C.M., et al. (2016) Developing risk equivalent data-rich and data-limited harvest strategies. *Fisheries Research* **183**, 574–587.

Hilborn, R., Amoroso, R.O., Anderson, C.M., et al. (2020) Effective fisheries management instrumental in improving fish stock status. *Proceedings of the National Academy of Sciences*.

Hoffman, M.D. and Gelman, A. (2011) The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *arXiv:1111.4246 [cs, stat]*.

Hordyk, A., Ono, K., Prince, J.D. and Walters, C.J. (2016) A simple length-structured model based on life history ratios and incorporating size-dependent selectivity: Application to spawning potential ratios for data-poor stocks. *Canadian Journal of Fisheries and Aquatic Sciences*.

Kimura, D.K., Balsiger, J.W. and Ito, D.H. (1984) Generalized stock reduction analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **41**, 1325–1333.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H. and Bell, B.M. (2016) **TMB** : Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software* **70**.

Martell, S. and Froese, R. (2013) A simple method for estimating MSY from catch and resilience. *Fish and Fisheries* **14**, 504–514.

Maureaud, A., Frelat, R., P'ecuchet, L., et al. (2020) Are we ready to track climate-driven shifts in marine species across international boundaries? - A global survey of scientific bottom trawl data. Ecology.

Melnychuk, M.C., Peterson, E., Elliott, M. and Hilborn, R. (2017) Fisheries management impacts on target species status. *Proceedings of the National Academy of Sciences* **114**, 178–183.

Monnahan, C. and Kristensen, K. (2018) No-u-turn sampling for fast bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan r packages. *PloS one* **13**.

Osio, G.C., Orio, A. and Millar, C.P. (2015) Assessing the vulnerability of Mediterranean demersal stocks and predicting exploitation status of un-assessed stocks. *Fisheries Research* **171**, 110–121.

Palomares, M.L.D., Froese, R., Derrick, B., et al. (2020) Fishery biomass trends of exploited fish populations in marine ecoregions, climatic zones and ocean basins. *Estuarine, Coastal and Shelf Science*, 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**, 290–295.

Pauly, D., Hilborn, R. and Branch, T.A. (2013) Fisheries: Does catch reflect abundance? *Nature* **494**, 303–306.

Pella, J.J. and Tomlinson, P.K. (1969) A generalized stock production model. *Inter-American Tropical Tuna Commission Bulletin* **13**, 416–497.

Pons, M., Cope, J.M. and 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*.

Poole, D. and Raftery, A.E. (2000) Inference for Deterministic Simulation Models: The Bayesian Melding Approach. *Journal of the American Statistical Association* **95**, 1244–1255.

Prince, J. and Hordyk, A. (2019) What to do when you have almost nothing: A simple quantitative prescription for managing extremely data-poor fisheries. *Fish and Fisheries* **20**, 224–238.

R Core Team (2019) R: A Language and Environment for Statistical Computing.

Ricard, D., Minto, C., Jensen, O.P. and 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**, 380–398.

Rosenberg, A.A., Kleisner, K.M., Afflerbach, J., et al. (2018) Applying a New Ensemble Approach to Estimating Stock Status of Marine Fisheries around the World. *Conservation Letters* **11**, e12363.

Rousseau, Y., Watson, R.A., Blanchard, J.L. and Fulton, E.A. (2019) Evolution of global marine fishing fleets and the response of fished resources. *Proceedings of the National Academy of Sciences*, 201820344.

Rudd, M.B. and Thorson, J.T. (2017) Accounting for variable recruitment and fishing mortality in length-based stock assessments for data-limited fisheries. *Canadian Journal of Fisheries and Aquatic Sciences*, 1–17.

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**, 23–56.

Stan Development Team (2018) {{RStan{}: The {}R{} interface to {}Stan{}{}.

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**, 237–251.

Thorson, J.T., Branch, T.A. and Jensen, O.P. (2012a) Using model-based inference to evaluate global fisheries status from landings, location, and life history data. *Canadian Journal of Fisheries and Aquatic Sciences* **69**, 645–655.

Thorson, J.T. and Cope, J.M. (2015) Catch curve stock-reduction analysis: An alternative solution to the catch equations. *Fisheries Research* **171**, 33–41.

Thorson, J.T., Cope, J.M., Branch, T.A. and Jensen, O.P. (2012b) Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. *Canadian Journal of Fisheries and Aquatic Sciences* **69**, 1556–1568.

Vehtari, A., Gelman, A. and Gabry, J. (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* **27**, 1413–1432.

Walters, C.J., Martell, S.J.D. and Korman, J. (2006) A stochastic approach to stock reduction analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **63**, 212–223.

Winker, H., Carvalho, F. and Kapur, M. (2018) JABBA: Just Another Bayesian Biomass Assessment. *Fisheries Research* **204**, 275–288.

Tables

Table 1: Data sources included across model fits.

Data Source	Short Name	Data Use	Caveats
Catch data (FAO 2020)	catches	Priors on stock status, scaling of population size, exploitation history	Heuristics or regressions used to translate shape of catch history into priors on stock status
Fisheries Management Index (Melnychuk *et al.* 2017)	FMI	Priors on most recent F/F_{MSY} values	Priors produced by regression trained on data from RAM Legacy Stock Assessment Database
Swept Area Ratio (Amoroso *et al.* 2018)	SAR	Priors on most recent F/F_{MSY} values	Priors produced by regression trained on data from RAM Legacy Stock Assessment Database
Reconstructed effort data (Rousseau *et al.* 2019) |	effort |	Combined with catch data to create an index of abundance |	Total reconstructed effort across all sectors. Assumed rate of technology creep reported in individual sections

Table 2: Data sources used for terminal stock status estimate

Data Name	Description
RLSADB Index	Fit to abundance index from RLSADB
SAR	Prior on terminal F/Fmsy set by regional swept area ratio
FMI	Prior on terminal F/Fmsy set by regional fisheries management index scores
Efective CPUE	Fit to CPUE index created from RLSADB catch and regional effort index. 2.6% technology creep

Data Name	Description
Effective CPUE+	Fit to CPUE index created from RLSADB catch and regional effort index with priors informed by SAR and FMI. 2.6% tech. creep
Nominal CPUE	Fit to CPUE index created from RLSADB catch and regional effort index. 0% tech. creep
Nominal CPUE+	Fit to CPUE index created from RLSADB catch and regional effort index with priors informed by SAR and FMI. 0% tech. creep
Guess	Priors on terminal B/Bmsy randomly sampled from 0.4,1.0,1.6

Table 3: Global performance statistics in the most recent year available of models using different sources of data. MPE = median percent error (bias), MAPE = median absolute percent error (error), Accuracy = percent of times that stocks were classified to the correct FAO status bin (underfished, maximally sustainably fished, overfished). Performance is judged relative to B/B_{MSY} reported values in RAM Legacy Stock Assessment Database.

Data Used	MPE	MAPE	Accuracy
RLSADB Index	0.14	0.29	0.69
FMI	-0.09	0.47	0.42
SAR	-0.04	0.50	0.38
Effective CPUE+	-0.30	0.52	0.43
Nominal CPUE+	-0.01	0.52	0.46
Guess	-0.08	0.54	0.34
CMSY	-0.54	0.60	0.41
Nominal CPUE	0.05	0.63	0.48
Effective CPUE	-0.36	0.68	0.41

Figure Legends

Figure 1: RLSADB values of B/B_{MSY} and F/F_{MSY} (x-axes) for case study fisheries plotted against estimated values (y-axes) using CMSY (Froese *et al.* 2017), priors informed by stock-specific Fisheries Management Index (FMI) and swept area ratio (SAR) scores, and an abundance index based on reconstructed effort trends assuming a rate of technological increase of 2.6%. Each point is a stock. Black dashed line shows the 1:1 relationship.

Figure 2: Median percent error (MPE, predicted relative to observed) in most recent B/B_{MSY} by FAO statistical area from different data sources. RLSADB Index refers to catch and abundance index drawn from RLSADB. Effective CPUE refers to an index of abundance based on reconstructed effort data. Effective CPUE+ uses CPUE along with Fisheries Management Index (FMI) and/or swept area ratio (SAR) data. For both CPUE series 'nominal' assumes a 0% technology creep, for 'effective' a 2.6% technology creep is assumed. FMI uses FMI scores to develop a prior on recent fishing mortality rates, SAR does the same but based on swept area ratio. CMSY uses the methods from Froese et al. 2017 (Froese *et al.* 2017). Guess assigns a random recent B/B_{MSY} of 0.4,1, or 1.6. Panels ordered in ascending (starting from top left) mean MPE at the FAO region level.

Figure 3: Median absolute percent error (MAPE) in most recent B/B_{MSY} by FAO statistical area from different data sources. RLSADB Index refers to catch and abundance index drawn from RLSADB. Effective CPUE refers to an index of abundance based on reconstructed effort data. Effective CPUE+ uses CPUE along with Fisheries Management Index (FMI) and/or swept area ratio (SAR) data. For both CPUE series 'nominal' assumes a 0% technology creep, for 'effective' a 2.6% technology creep is assumed. FMI uses FMI scores to develop a prior on recent fishing mortality rates, SAR does the same but based on swept area ratio. CMSY uses the methods from Froese et al. 2017 (Froese *et al.* 2017). Guess assigns a random recent B/B_{MSY} of 0.4,1, or 1.6. Panels ordered in descending (starting from top left) mean MAPE at the FAO region level

Figure 4: Mean classification accuracy (assignment to FAO stock status category) by FAO statistical area arising from different data sources. RLSADB Index refers to catch and abundance index drawn from RLSADB. Effective CPUE refers to an index of abundance based on reconstructed effort data. Effective CPUE+ uses CPUE along with Fisheries Management Index (FMI) and/or swept area ratio (SAR) data. For both CPUE series 'nominal' assumes a 0% technology creep, for 'effective' a 2.6% technology creep is assumed. FMI uses FMI scores to develop a prior on recent fishing mortality rates, SAR does the same but based on swept area ratio. CMSY uses the methods from Froese et al. 2017 (Froese *et al.* 2017). Guess assigns a

random recent B/B_{MSY} of 0.4,1, or 1.6. Panels ordered in descending (starting from top left) mean accuracy at the FAO region level.

Figure 5: Posterior probability distributions of estimated effect of different data types on root mean squared error (RMSE) of B/B_{MSY} in the most recent 5 years of data available for each model fit. Distribution is full posterior probability distribution. Point is median, thicker black section inner 66th quantile of the posterior, the thinner black line the 95th. Change is relative to the mean performance of a catch-only heuristic model.

Supporting Information

Population Model

The core of our model is a Pella-Tomlinson (Pella and Tomlinson 1969) production model in the manner of (Winker *et al.* 2018). While models of these kinds abstract away many important details of fish biology and fleet behavior, they are the highest resolution model that the potential data evaluated here will support.

Table S1: Name, abbreviations, and priors distribution for parameters potentially estimated by sraplus in this manuscript. LN refers to log normal, where the mean is reported on the unit scale.

Parameter	Abbreviation	Default Prior
Carrying Capacity	K	Prior predictive tuning
Growth rate	r	Thorson, 2020 (Thorson 2020) updated by prior predictive tuning
Shape parameter	m	Drawn from Thorson et al. 2012 (Thorson *et al.* (2012b))
Catchability	q	$LN(1e^{-3}, 1)$
Observation Error	σ_{obs}	$LN(.05, 1)$
Ratio of process to observation error	γ	$LN(.5, 0.25)$
Initial State	B_0	Posterior probability dist. of catch-based regressions

Table S2: SSBMSY to SSB ratios from Thorson et al. (2012) used in the paper. Taxa not within the groups assigned at the genus level by Thorson et al. (2012) are assigned the ratio reported for 'Other'

Taxanomic Group	SSBMSY/SSB0	SSBMSY/SSB0 SD
Pleuronectiformes	0.395	0.119
Gadiformes	0.439	0.122
Perciformes	0.353	0.114
Clupeiformes	0.261	0.097
Scorpaeniformes	0.463	0.122
Other	0.405	0.120

Prior Predictive Tuning

Our prior predictive tuning regime is similar in spirit to Bayesian melding (Poole and Raftery 2000). Our solution amounts to a two-step sample-importance-resampling (SIR) algorithm. We first run the standard SIR algorithm as described above. We then break the resulting draws into bins based on terminal stock status, and calculate the mean sampling probability of each bin. The net result of this is that it allows users to place explicit prior on stock status, and then adjust their priors on life history parameters to reflect this prior, rather than creating a complicated and biased prior on stock status based on a mixture of explicit and implicit priors.

The SRA algorithm works in two steps. First, the algorithm rejects any draws that resulted in the collapse of the population (biomass less than catch in a given timestep). From there a standard SRA would sample from the priors in proportion to the stated prior on recent stock status. If the bulk of the prior on terminal stock status was concentrated at 50% of K, combinations of r and K that produce terminal stock status near 50% of K are sampled proportionally more frequently. However, lower values of terminal stock status have fewer candidate values of r and K, since it becomes harder and harder to find viable pairs that come close to but do not crash the population at any time step. Conversely, in the absence of constraints higher values of stock status have infinite combinations of plausible r and K combinations: since under this model the population cannot be greater than carrying capacity, as for example K approaches infinity terminal stock status asymptotes at close to 100% of K. The net result of this is that even though individual combinations of r and K that produce higher stock status than the mean of the prior on recent stock status individually have lower probability of being sampled, there are many more opportunities for the lower-probability events that produce higher stock status to be sampled. As a result, the post-model-pre-data prior on terminal depletion will always be higher under this method than the supplied prior on stock status.

The net result of our correction is a post-model-pre-data distribution of life history parameters that produce a distribution of recent stock status that roughly matches the supplied prior on recent stock status. In effect, this process answers the question "given the model, what combinations of parameters produce my prior on recent stock status." This is only an approximate solution, but it helps ensure that the post-model-pre-data distribution of stock status much more closely matches the stated prior on recent stock status, and reduced the positive bias resulting from use of the raw SRA algorithm (Fig.S1, Fig.S2).

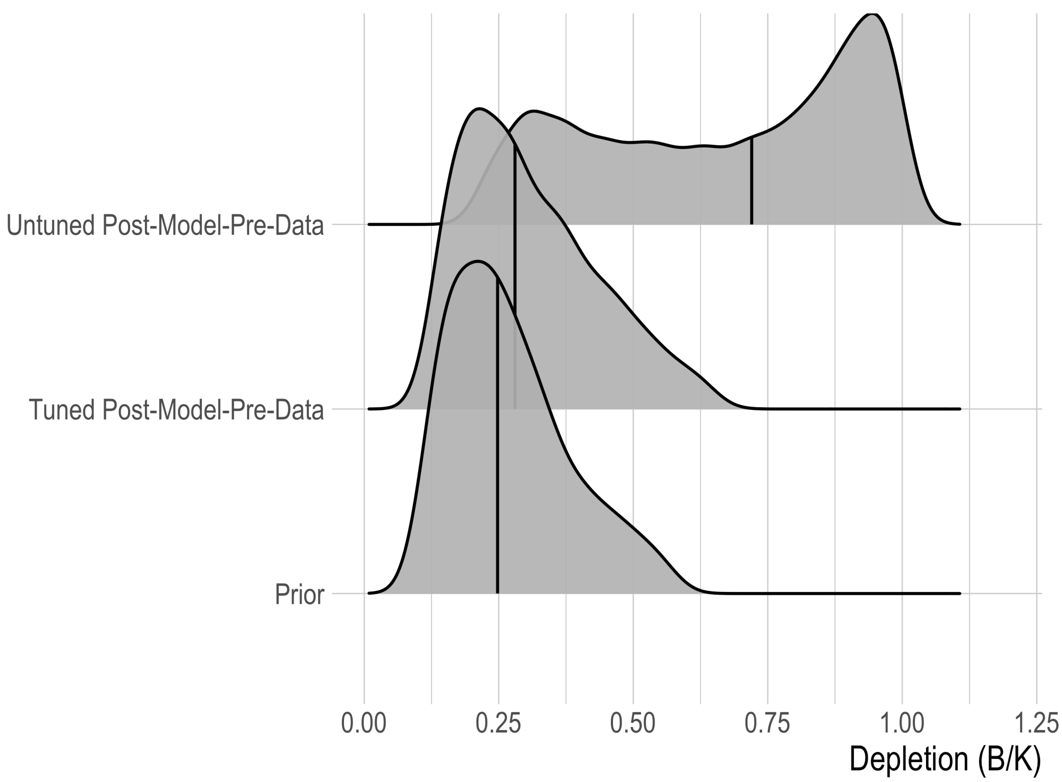

Figure S1: Post-model-pre-data distribution of depletion (biomass relative to carrying capacity) from raw SRA algorithm (untuned, top row), from SRA algorithm with approximate tuning applied (tuned, middle row), compared to the supplised prior on depletion (bottom row). Black vertical line indicates median value.

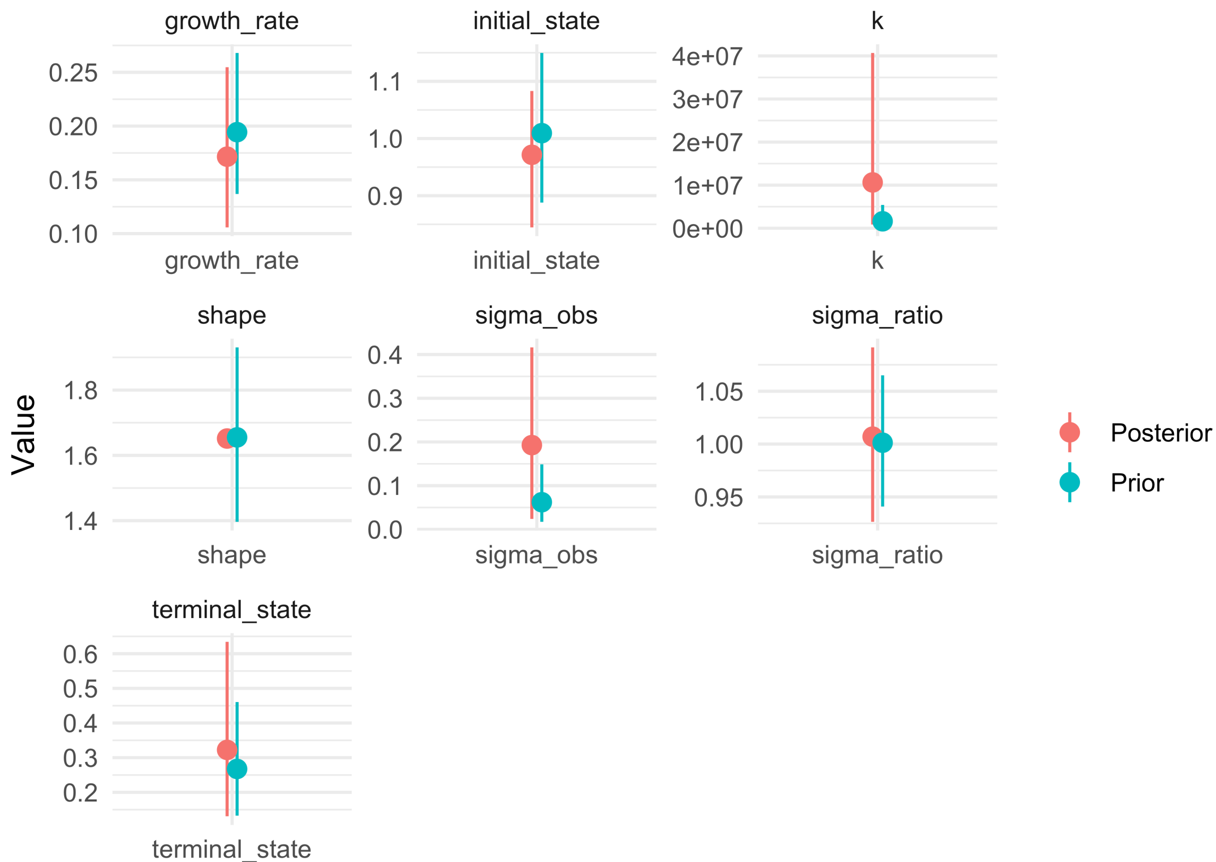

Figure S2: Prior posterior plots of fits for case study fishery

Prior Generating Regressions

Catch-Only Priors

Many of the current methods for estimating global stock status of unassessed stocks are based on predicting stock status from characteristics of the catch history (Pauly 2007; Costello *et al.* 2012, 2016; Rosenberg *et al.* 2018). While these catch-only methods have been shown to have serious shortcomings (Free *et al.* 2020), we include them as a point of reference given their ubiquity in the global assessment literature.

We used data from the RAM Legacy Stock Assessment Database to estimate a regression of stock status as a function of catch history characteristics. To facilitate the process, we first fit a spectral clustering algorithm to the scaled catch histories of fisheries in RAM, in order identify four possible clusters of catch history types within the the data. Emergent clusters show for example one built around a downward "one way trip" style catch histories, others with a boom and bust pattern, others with stable but fluctuating catches.

We then trained a classification algorithm to predict which catch cluster a given fishery would fall into based on the shape of its catch history. This algorithm was then used to assign fisheries to one of the four identified catch history types, and the catch history type was then used as a hierarchical term within our catch-based regressions (where s refers to a smoothing term). For the first regression, we restrict the data to the first year of data available for each fishery i, in order to estimate initial stock status

$$log(value_i) \sim normal(s(\frac{first(catch)}{max(catch)}|cluster_i) + s(log(lengthi)|cluster_i) + 1, \sigma)$$

For the second regression, we included data for all available years y for fishery i. The model is then used to construct a prior on fishery status in the terminal year of the data

$$log(value_{i,y}) \sim normal(s(fyear|cluster_i) + s(\frac{catch_{i,y}}{max(catch_i)}|cluster_i) + cluster_i, \sigma)$$

where *fyear* is the year of the fishery, starting from 0.

Fits for the catch-only prior-generating regression are visible in Fig.S3.

Fisheries Management Index Priors

The Fisheries Management Index (FMI), as presented in (Melnychuk *et al.* 2017), utilizes surveys filled out by regional experts to score a fishery against a set of 46 specific questions for individual species about what

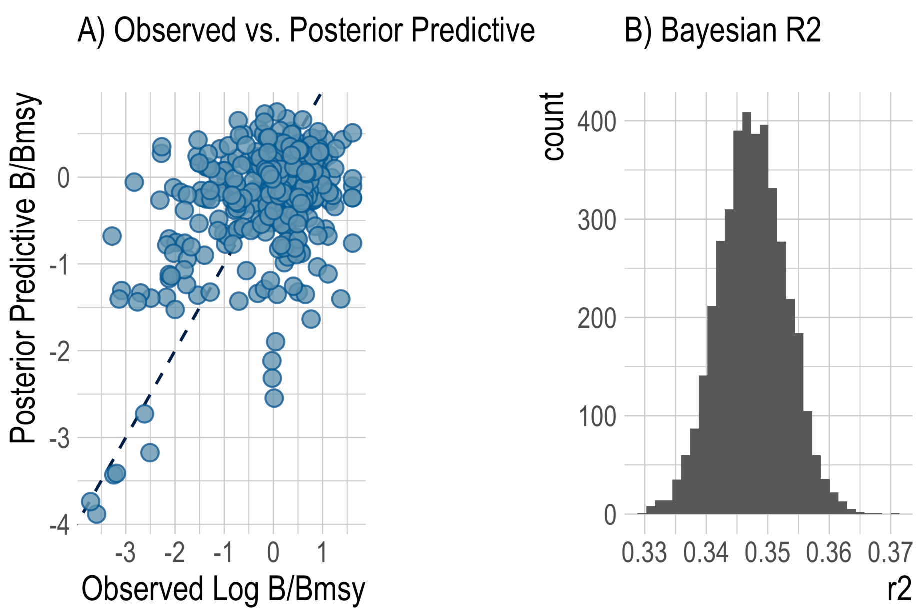

Figure S3: Observed (x-axis) vs posterior predictive (y-axis) B/B_{MSY} for regression of catch on B/B_{MSY}

elements of fisheries management were in place. These questions are then aggregated into broader categories of science, enforcement, management, and socioeconomic. The higher the score, the better the expert judges that a given metric is met in that fishery. Importantly, FMI surveys can be filled out in the absence of stock assessments. This allows us to explore how FMI values map onto stock status, and explore the ability then to use FMI scores to produce priors on stock status for unassessed fisheries (in a manner similar to (Osio *et al.* 2015) and (Cope *et al.* 2015)).

The final selected model relating FMI variable to stock status metrics was a generalized additive model (GAM) of the form

$$log(value_i) \sim N(s(research_i)+s(management_i)+s(enforcement_i)+s(socioeconomics_i)+\frac{catch_i}{max(catch)_i}+1, \sigma_{SAR})$$

Fits for the FMI prior-generating regression are visible in Fig.S4.

Swept Area Ratio Priors

(Amoroso *et al.* 2018) provides an extensive database of trawling footprints throughout the world, including both regions heavily covered by stock assessments and largely unassessed areas. This makes the trawl footprint data an ideal candidate for supporting global stock assessment efforts. As illustrated in (Amoroso *et al.* 2018), there is an evident positive relationship between the swept area ratio (SAR,the total annual area trawled divided by the total area of the region) and U/U_{MSY}. Note that SAR can be greater than 1 since the same area can be trawled multiple times in a year, e.g. if all trawl-able areas are trawled twice a year then the SAR will be 2. Also note the skewed distribution of SAR values with most concentrated well below 1 and only a handful above 1.

The final selected model relating SAR to to stock status metrics was

$$log(value_i) \sim normal(s(SAR_i) + s(\frac{catch_i}{max(catch)_i}) + 1, \sigma_{SAR})$$

Fits for the SAR prior-generating regression are visible in Fig.S5.

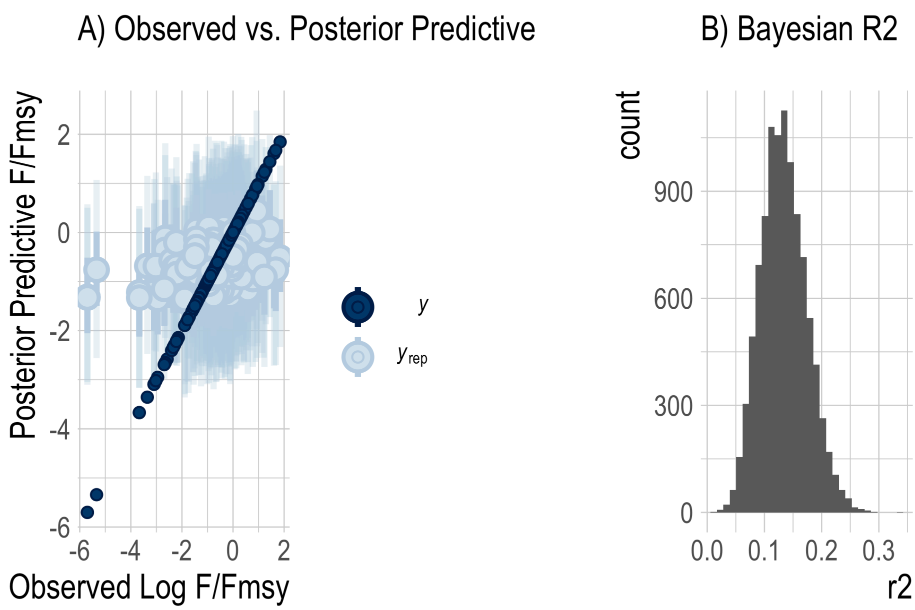

Figure S4: Observed (x-axis) vs posterior predictive (y-axis) F/F_{MSY} for regression of fisheries management index (FMI) on F/F_{MSY}

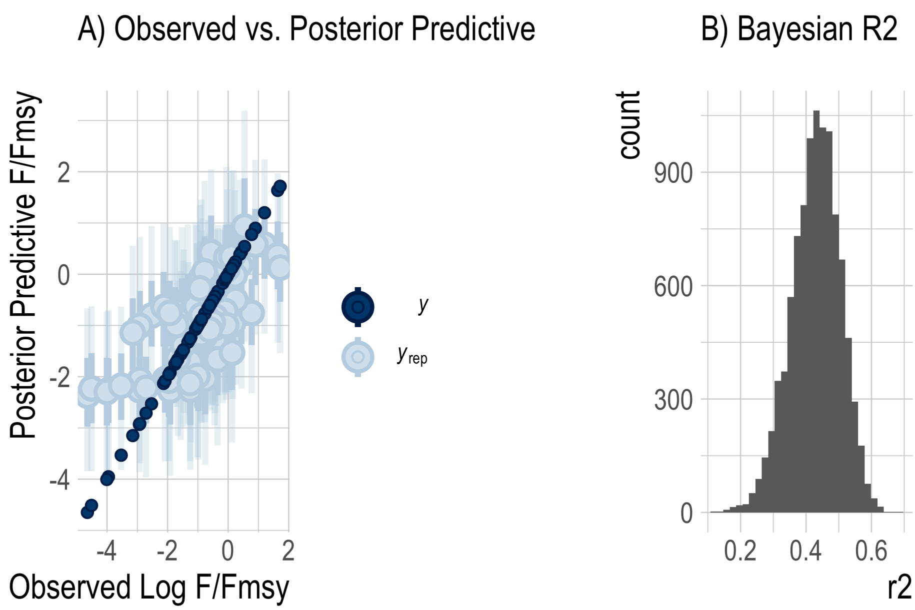

Figure S5: Observed (x-axis) vs posterior predictive (y-axis) F/F_{MSY} for regression of swept area ratio (SAR) on F/F_{MSY}

Value of Information Calculation

We performed a value-of-information (VOI) exercise we assessed performance as the root-mean-squared-error of B/B$_{\text{MSY}}$ over the most recent 5 years of the fishery, in order to evaluate the ability of the model to capture the recent trends in stock status and not just the most recent year. We evaluate the contributing of each data type to RMSE using a Gamma GLM with a log link of the form

$$rmse \sim Gamma(\boldsymbol{\beta X} + (1|stock), shape, scale)$$

Where $\boldsymbol{\beta}$ is the vector of coefficients associated with the matrix of dummy variables marking the use of different data types in the vector $\boldsymbol{X}$

F/F$_{\text{MSY}}$ Performance

Our results focused on the performance of candidate models in estimating B/B$_{\text{MSY}}$, as this reflects the broad mission of the FAO's SOFIA reports to assess the current biomass status of global fisheries. However, fishing mortality rates, specifically F/F$_{\text{MSY}}$ are also of importance to managers and commonly considered as an output of catch-only models.

As such, we repeated our performance calculations summarized in Figures @ref(fig: mpe-map)-4 but now focused on F/F$_{\text{MSY}}$. Performance was comparably poor to the B/B$_{\text{MSY}}$ based results, with the exception that the default settings of CMSY produced a consistent positive bias in F/F$_{\text{MSY}}$.

Table S3: Global performance statistics in the most recent year available of models using different sources of data. MPE = median percent error (bias), MAPE = median absolute percent error (error), Accuracy = percent of times that stocks were classified to the correct FAO status bin (underfished, maximally sustainably fished, overfished). Performance is judged relative to F/F_{MSY} values reported values in RAM Legacy Stock Assessment Database.

Data Used	MPE	MAPE	Accuracy
Effective CPUE+	0.05	0.58	0.51
Nominal CPUE+	-0.28	0.59	0.56

Data Used	MPE	MAPE	Accuracy
RLSADB Index	-0.44	0.60	0.62
SAR	0.09	0.67	0.52
Guess	0.43	0.67	0.35
FMI	0.18	0.69	0.47
Effective CPUE	-0.41	0.86	0.49
Nominal CPUE	-0.76	0.86	0.57
CMSY	1.48	1.48	0.26

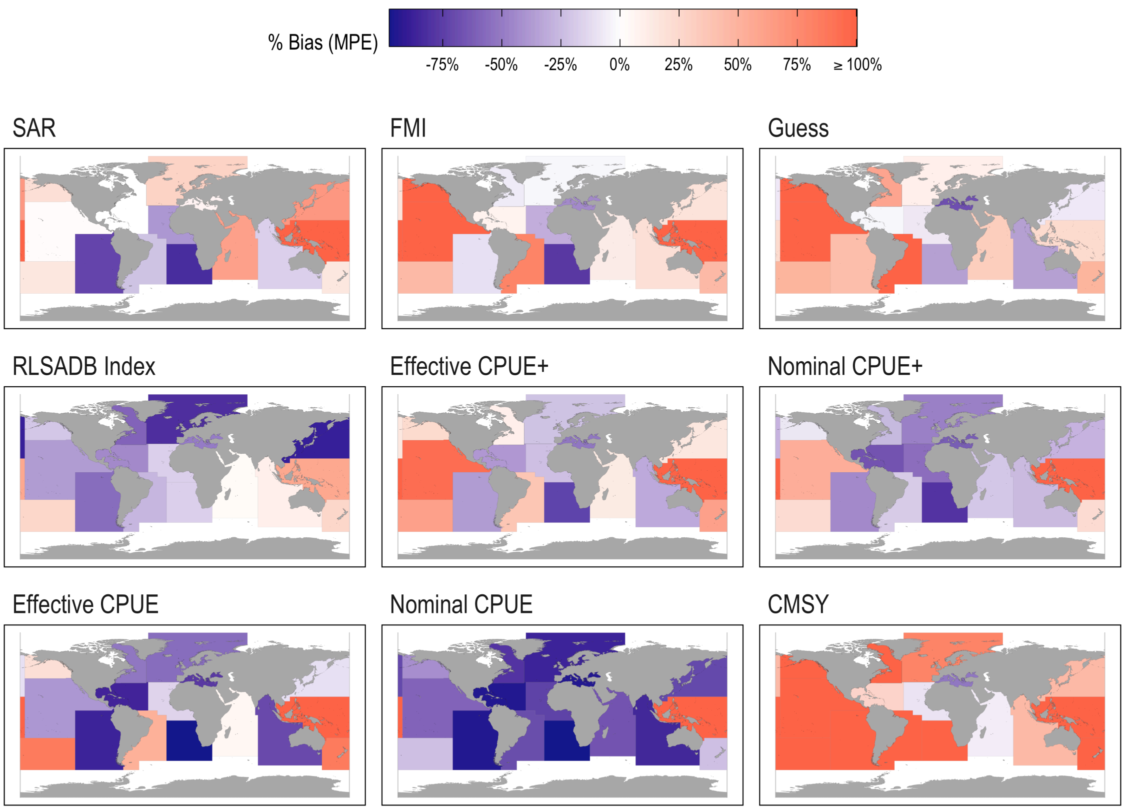

Figure S6: Median percent error (MPE, predicted relative to observed) in most recent F/F_{MSY} by FAO statistical area from different data sources. RLSADB Index refers to catch and abundance index drawn from RLSADB. Effective CPUE refers to an index of abundance based on reconstructed effort data. Effective CPUE+ uses CPUE along with Fisheries Management Index (FMI) and/or swept area ratio (SAR) data. For both CPUE series 'nominal' assumes a 0% technology creep, for 'effective' a 2.6% technology creep is assumed. FMI uses FMI scores to develop a prior on recent fishing mortality rates, SAR does the same but based on swept area ratio. CMSY uses the methods from Froese et al. 2017 (Froese *et al.* 2017). Guess assigns a random recent F/F_{MSY} of 0.4,1, or 1.6. Panels ordered in ascending (starting from top left) mean MPE at the FAO region level.

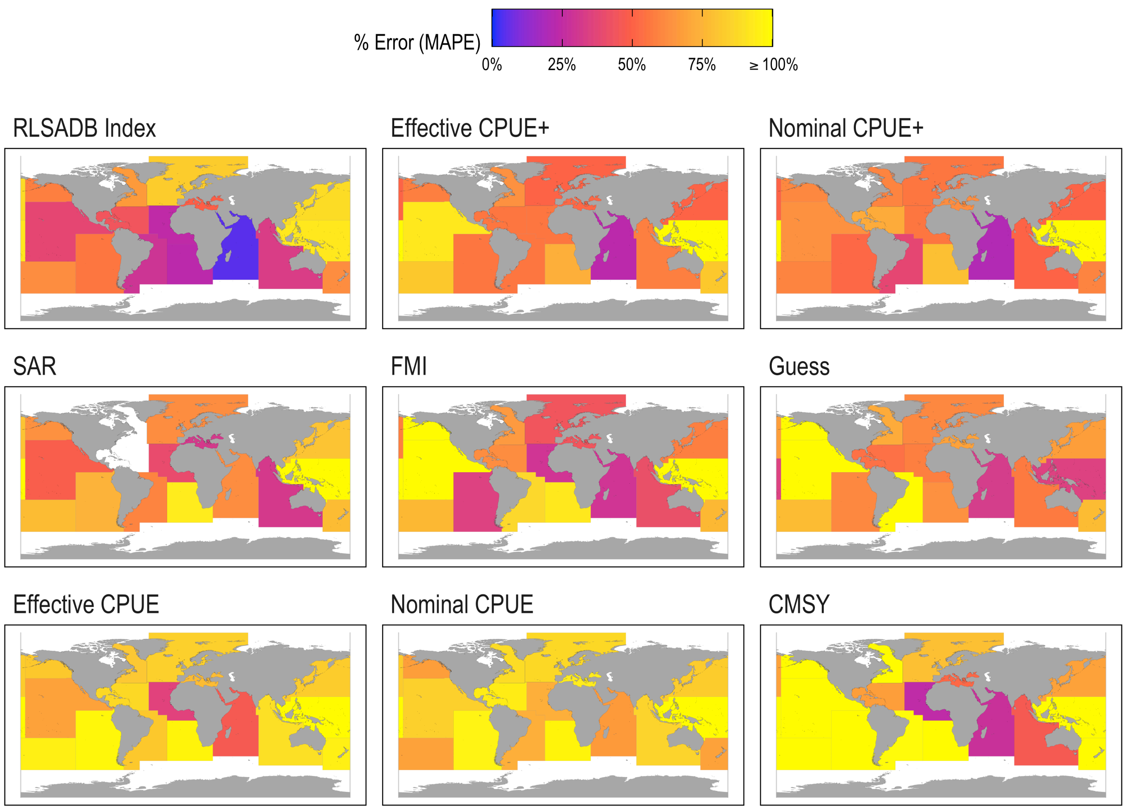

Figure S7: Median absolute percent error (MAPE) in most recent F/F_{MSY} by FAO statistical area from different data sources. RLSADB Index refers to catch and abundance index drawn from RLSADB. Effective CPUE refers to an index of abundance based on reconstructed effort data. Effective CPUE+ uses CPUE along with Fisheries Management Index (FMI) and/or swept area ratio (SAR) data. For both CPUE series 'nominal' assumes a 0% technology creep, for 'effective' a 2.6% technology creep is assumed. FMI uses FMI scores to develop a prior on recent fishing mortality rates, SAR does the same but based on swept area ratio. CMSY uses the methods from Froese et al. 2017 (Froese *et al.* 2017). Guess assigns a random recent F/F_{MSY} of 0.4,1, or 1.6. Panels ordered in descending (starting from top left) mean MAPE at the FAO region level

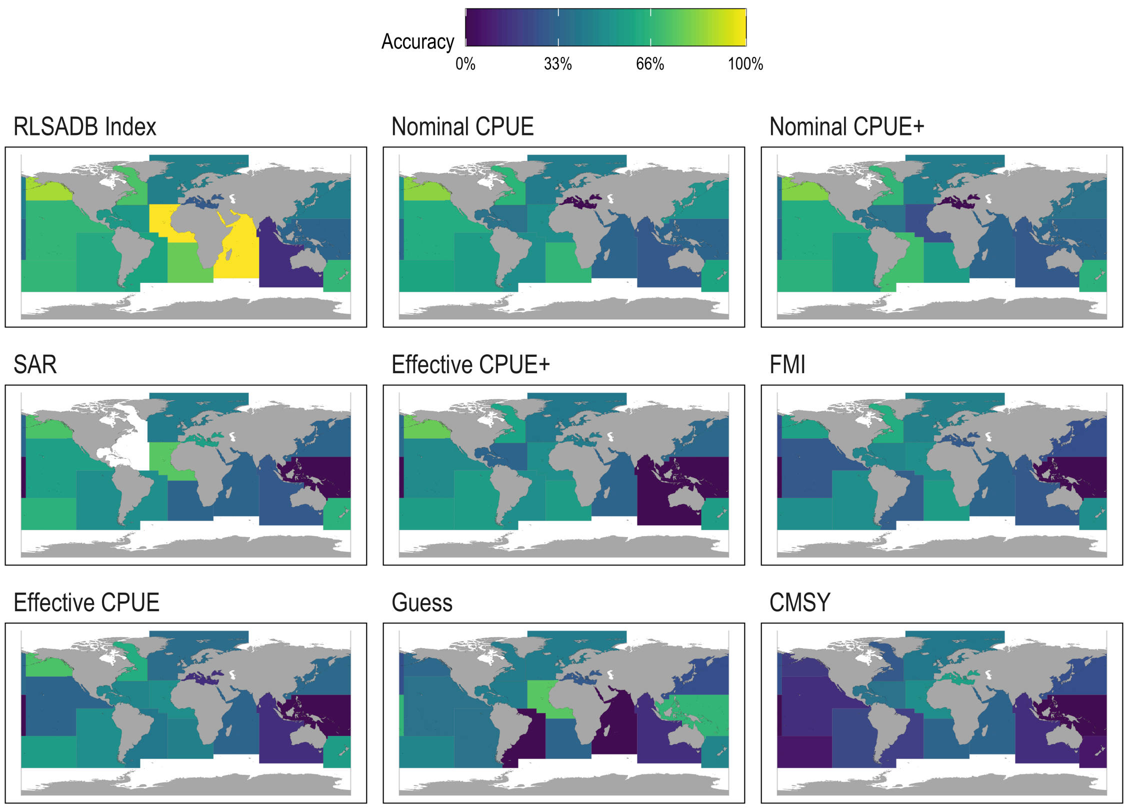

Figure S8: Mean classification accuracy (assignment to general bin of overfishing, fishing near F_{MSY}, and underfishing) by FAO statistical area arising from different data sources. RLSADB Index refers to catch and abundance index drawn from RLSADB. Effective CPUE refers to an index of abundance based on reconstructed effort data. Effective CPUE+ uses CPUE along with Fisheries Management Index (FMI) and/or swept area ratio (SAR) data. For both CPUE series 'nominal' assumes a 0% technology creep, for 'effective' a 2.6% technology creep is assumed. FMI uses FMI scores to develop a prior on recent fishing mortality rates, SAR does the same but based on swept area ratio. CMSY uses the methods from Froese et al. 2017 (Froese *et al.* 2017). Guess assigns a random recent F/F_{MSY} of 0.4,1, or 1.6. Panels ordered in descending (starting from top left) mean accuracy at the FAO region level