

Which fish stocks are assessed: an analysis and forecast of stock assessment in the United States

Philipp Neubauer, Dragonfly Data Science, Wellington, New Zealand

James T. Thorson, NOAA Northwest Fisheries Science Center, Seattle, USA

Michael C. Melnychuk, School of Aquatic and Fisheries Science,
University of Washington, Seattle, USA

Richard Methot, NOAA Northwest Fisheries Science Center, Seattle, USA

Kristan Blackhart, ECS Federal, Inc., in support of NOAA Fisheries,
Office of Science and Technology, USA

May 18, 2017

Fisheries management is most effective when based on scientific estimates of sustainable fishing rates. While some simple approaches allow estimation of harvest limits, more data-intensive stock assessments are generally required to evaluate the stock's biomass relative to target levels and fishing rates relative to sustainable levels. To date, only a subset of landed species have stock assessments. Here we evaluate factors influencing the probability that a previously-unassessed stock in the United States will become the subject of a stock assessment. Using a statistical model based on time-to-event analysis and 569 coastal marine fish and invertebrate stocks landed in commercial fisheries, we quantify the impact of region, habitat, life-history, and economic factors on the annual probability of being assessed. Although the majority of landings come from assessed stocks in all regions, less than half of the regionally-landed species currently have been assessed. Nevertheless, we find that the overall rate at which new stocks are assessed has been increasing since 1960, the year of the first recorded stock assessment. Our time-to-event model identifies landed tonnage and ex-vessel price as the domi-

nant factors determining the rate of new assessments, with greater landings and higher prices leading to higher probabilities of being assessed. Placement of stocks within federal fishery management plans was identified as adding momentum toward stocks being assessed. We also find that after controlling for landings and price, there has been a consistent bias towards assessing larger-bodied species. A number of vulnerable groups such as rockfishes (Scorpaeniformes) and groundsharks (Carcharhiniformes) have a relatively high annual probability of being assessed after controlling for their relatively small tonnage and low price. Given the characteristics of species that are currently unassessed, our model suggests that the number of assessed stocks will increase more slowly in future decades, as the landed tonnage and price for the remaining unassessed stocks are relatively low, and may therefore confer less incentive for implementing costly stock assessments.

1. Introduction

Fisheries scientists have measured human impacts on populations of finfishes and invertebrates for over 100 years with the goal of balancing the value derived from fishing with the long-term sustainability of populations ([32]). This is principally achieved by estimating two measures of human impact: (1) fishing rate, i.e., the instantaneous mortality or annual fraction of the population that is harvested relative to an estimated target level, and (2) the population abundance, i.e., spawning biomass or reproductive output relative to an estimated target level. Together these measures reflect the "stock status" of an assessed population, and fisheries agencies are increasingly committed to maintaining fished populations at fishing rates below and population abundances above limit levels that are defined based on biological and economic considerations ([22]). Where these quantities are most completely measured, global studies find that the most effective management occurs ([38, 21]).

The National Oceanic and Atmospheric Administrations' National Marine Fisheries Service (NMFS, the agency in charge of federal fisheries management in the United States (US)) is committed to "end overfishing" for all marine species within regional fisheries management plans (with exceptions granted in certain circumstances; [22]). In the US, overfishing is defined as any stock having annual harvest rate or quantity above limit levels, generally set at the level that would produce maximum sustainable yield. A target (or limit) harvest quantity can in theory be calculated by combining a target (limit) harvest rate with an estimate of current population abundance. However, the vast majority of overfishing limits are currently estimated using methods that do not individually estimate either harvest rate or population abundance ([1, 25]). For example, catch-only methods (COMs), such as depletion-corrected average catch (DCAC; [18]), are used to estimate an annual fishing limit for many data-limited stocks, but are not capable of estimating population abundance. COMs can therefore be used to help "end overfishing", but are not otherwise informative about the status of a fished population.

Conservationists and ecologists will often be more interested in estimating population abundance (or abundance relative to equilibrium conditions) than estimating an overfishing limit (e.g., [16]). Estimating abundance generally requires applying a population model to available harvest data and an index of population depletion (either an index proportional to population abundance, or average body size or age data). Estimated abundance is then compared to a biological reference point, or benchmark, for assessing current status relative to target levels. In the following, we consider this pairing of model-estimated abundance with estimated reference points as "stock assessments", although we acknowledge that other authors have used the term "stock assessment" more broadly to also include methods that estimate overfishing limits but not population abundance. Although NMFS has estimated overfishing limits for the vast majority of fishes in US fisheries management plans, a smaller percentage of fished species has a stock assessment under our more restrictive definition. The dearth of stock assessments arises because developing a stock assessment typically requires more input data and time, and therefore requires extensive financial resources ([10]).

Stock assessments are important for many applied and theoretical questions regarding marine ecosystems. In particular, managing and monitoring rebuilding of overfished stocks to a target biomass level, rather than simply managing annual fishery removals, is possible only by estimating population abundance relative to target levels using stock assessment models. However, there is little previous research regarding which fished species are more or less likely to receive sufficient attention to develop a stock assessment. Understanding which species are more or less likely to be assessed could be useful for the following three reasons:

1. Conservation concerns and, conversely, missed opportunities for increased exploitation will be undocumented, so will receive less attention from the public and fishery managers.
2. Output from stock assessments has often been used in meta-analyses to understand ecological characteristics of marine fishes in general ([24, 34, 17]). Therefore, any systematic bias in which particular stocks are assessed will also bias our ecological understanding of marine fishes.

3. Stock assessments often require periodic updates (e.g., Pacific hake has been re-assessed annually from 1982 through 2016; [e.g., 14]), and agency resources might be fully expended while assessing a small fraction of possible stocks. If the rate of assessing new species is decreasing, this could indicate the need for additional public resources for stock assessment or improved strategies for prioritizing which stocks to assess ([23]).

In this paper, we provide a quantitative analysis of which marine species landed by commercial fisheries are likely to have undergone a stock assessment using a statistical population dynamics model coupled with estimated reference points. We combine two databases representing fished coastal marine species in the continental US and Alaska: a database of landed tonnage and value by species from 1950 to 2013, and a database of management and stock assessment attributes for US fishes and invertebrates drawn from peer-reviewed stock assessments. We record the year that each stock landed in the continental US and Alaska (whether caught in US federal or state jurisdictions) first had a stock assessment, and we treat any stock that did not have an assessment by 2013 as a "censored" observation (i.e., it might eventually have an assessment). We then apply a censored time-to-event model to answer the following questions: (1) What economic and biological characteristics are associated with a high or low annual probability of being assessed for the first time?; (2) how has the rate of assessing stocks differed among four US regions (Northeast, Southeast, Alaska, and US West Coast) and with federal vs non-federal management authority?; (3) are there certain taxa (e.g., invertebrates, sharks, flatfishes, etc.) that are assessed substantially faster or slower after accounting for biological and economic attributes?; and (4) is the rate of stock assessment accelerating or decelerating over time? We show that landed tonnage and ex-vessel price are the main drivers of increasing rates of stock assessments, but larger fish and some taxa of conservation concern defy these trends and are more or less likely to be assessed.

2. Methods

2.1. Operational definition of US stock assessments

Many types of stock assessments are applied in the US, with varying levels of model complexity and input data requirements. Assessments for any given stock also tend to change over time, typically becoming more complex as warranted by available data. For consistency across US regions, we defined a stock assessment in this study as:

(A) a single-species model of density-dependent population dynamics (e.g., including some combination of individual growth, recruitment, or aggregate surplus production); where

(B) model parameters were estimated by fitting to abundance index and/or age or length compositional data;

(C) the model provided time series estimates of population abundance (e.g. total biomass, spawning biomass) and/or exploitation rates (e.g., fishing mortality or harvest fractions); and

(D) management benchmarks corresponding to these time series estimates were estimated within the assessment or were otherwise explicitly stated, where benchmarks included target reference points, reference points based on maximum sustainable yield (MSY) or its proxies, or initial population abundance; ratios of the time series and their corresponding reference points provide a relative index of stock status.

Age-structured models, delay-difference models, biomass dynamics models, and surplus production models all qualified as assessment models ([7]) as long as they also explicitly reported management benchmarks. We recognize that COMs such as stock-reduction analyses (SRAs) are often used to estimate overfishing limits for stocks in the absence of a population-dynamics model fitted to data ([18, 8]). However, stock-reduction analyses did not qualify as stock assessments under this definition because they typically are not fitted to abundance-index or compositional data.

2.2. Defining the set of landed stocks

The set of stocks for this analysis included all landed species of fish and invertebrates in US marine waters, and is not restricted to stocks listed in federal fishery management plans (FMPs; stocks that NFMS has jurisdiction over).

We excluded some species based on practical considerations. Highly migratory species often have population boundaries that substantially exceed the jurisdiction of any single nation, and also are often difficult to assign to any one of the regions that we define for later analysis. We therefore excluded species that are typically assessed by Regional Fisheries Management Organisations, including tuna, billfish, and oceanic sharks (noting that some are managed by both RMFOs and federal FMPs). We also excluded salmon and shad from our analysis because assessments for these anadromous species are often conducted at a fine spatial resolution which might otherwise either numerically dominate the other landed marine species or conflict with the typical spatial resolution for marine stock assessments. We include mollusks, crustaceans and echinoderms, but exclude corals, sponges, and other benthic invertebrates as the latter have limited use as seafood. Finally, we exclude stocks landed or assessed in the US Pacific Islands and the Caribbean, which have not yet been added to our stock assessment database.

After excluding the above species, we used the US National Oceanic and Atmospheric Administration (NOAA) landings database ([26]) to identify landed stocks. This database provides annual landings for both assessed and unassessed species by state from 1950-2013. We aggregated state landings into four bio-geographic regions, defined as: Alaska (i.e., the Eastern Bering Sea, Gulf of Alaska, and Aleutian Islands); US West Coast (i.e., the marine waters of Oregon, Washington, and California); Northeast Coast (including the mid-Atlantic Coast); and Southeast Coast (including the South Atlantic Coast and Gulf of Mexico). Assignments of states to regions were generally unambiguous, except for differentiating stocks in Northeast and Southeast regions. For assigning state landings into Northeast vs. Southeast regions, we generally treated all states north of North

Carolina (i.e., north of Cape Hatteras) as the Northeast region, and all other east coast states as the Southeast region. However, we made exceptions as several assessed stocks on the US east coast straddle our regional boundary; we assigned the assessed stock and its associated state landings to the region with the greatest average landings.

Landings of each species in each state were either assigned to an assessed stock or used to define an unassessed "stock". For assessed stocks, we used areas of distribution as defined in assessments to link their state landings. We considered occasional low-volume landings in nearby states to also belong to an assessed stock, because straying occurs and because fleets may catch fish in waters within a stock's area of distribution but land fish in nearby states outside of that distribution. For example, the area of distribution for the South Atlantic/Gulf of Mexico finetooth shark stock comprises waters from North Carolina southward, but occasional landings in Virginia and Maryland were considered to also pertain to the assessed stock. However, if an assessment indicated that landings from only certain state(s) were considered for the assessment, we did not link landings in other nearby states to the assessed stock. Landings from states that were not linked to assessed stocks allowed us to define unassessed stocks. These state landings were pooled within each region to define the unassessed stock; thus, a maximum of one unassessed stock per region was defined for each species. Landings that were not resolved to species but only to higher taxonomic levels (e.g., "Scallops") were excluded from our analysis. However, for some assessed stocks that were landed only as stock complexes prior to their first assessment, we inspected assessment documents for species-specific landings prior to the year of first assessment, and manually added these stocks to our final dataset to have as complete a record of assessed stocks as possible.

As the NOAA landings database does not differentiate between wild-caught and farmed (aquaculture) landings, we obtained information about aquaculture landings from the NOAA Fisheries of the US report ([27]). Most aquaculture landings over the past five years were from salmon, oysters, clams and mussels. To test the sensitivity of our results to the inclusion of landings from the

latter three species groups (salmon are already excluded), we repeated all analyses without these groups (i.e., by discarding a total of 21 stocks from the landings database). As the results were nearly identical to those found when including these stocks, we chose to maintain these stocks within our dataset.

2.3. Defining year of first stock assessment

Given the set of assessed stocks in each of the four US regions, we determined the year of first assessment for each of them. Identifying assessed stocks and their year of first assessment was accomplished by a combination of interviews with regional stock assessment scientists and literature reviews of archived assessments (Appendix A). For quality control, we compared our assignments of first assessment year with the NOAA Species Information System (SIS) database to ensure consistency for federally managed species (Appendix B). The SIS database does not contain information about when a stock was first assessed for the entire period considered here, so comparisons were restricted to the most recent SIS classification. These comparisons generally showed consistency among datasets, with categories of “Levels of Stock Assessment Models” in SIS aligning with our assignments of first stock-assessment year (as defined by criteria A-D above). Of the 211 stocks for which we assigned a year of first assessment, there were 28 discrepancies with SIS classifications which resulted from violation of criteria A-D (see Appendix B). These stocks were previously assessed using population models, but are either currently assessed with less complex methods or have had the most recent assessment rejected. For our analyses, we continue to consider these stocks as “assessed” and use the year that they were first assessed by a population dynamics model as their year of first stock assessment.

2.4. Explanatory variables

Several variables were considered as explanatory factors affecting the year in which a stock was first assessed. Region and habitat typically occupied by the population were each treated as categorical

random effects. Habitat types from FishBase [12] or SeaLifeBase [28] were compiled in R using rfishbase [2] and aggregated into six categories: deep sea (>200m; bathy-pelagic or bathy-demersal); benthic; demersal; benthopelagic; pelagic; and reef-associated. Maximum body length of the species was also assigned to each population and used as a numerical predictor, drawing from FishBase and SeaLifeBase. The catch quantity and ex-vessel price of the population together determine landed value of the population; more valuable populations may be more likely to be assessed. We considered maximum annual landings prior to the first assessment and mean ex-vessel price (US\$·kg⁻¹) prior to the first assessment as separate numerical predictors, drawn from the NOAA landings database. Note that the database starts in 1950 and these values may therefore be biased for stocks for which maximum landings occurred prior to 1950 or had substantial landings from foreign fleets, or for which the mean ex-vessel price has changed substantially since 1950. However, given that we consider time-to-assessment from 1960 (see next section), we have at least 10 years of data for stocks landed prior to 1950. The full dataset is provided in the supplementary material.

To evaluate whether the presence of particular stocks in a federal FMP changed the time-to-assessment, we repeated our analysis with this factor included as a binary fixed effect in our model detailed below. However, given the potential collinearity between this factor and other explanatory variables (i.e., the same factors that drive stock assessments may also drive inclusion in FMPs), we used our model without this factor to evaluate effects of other covariates.

2.5. Time-to-event model

To assess which factors drive the overall rate of assessments and the time from first recorded landings to a full stock assessment, we applied a time-to-event model. These models account for censored data (i.e., species that are landed but not yet assessed) while modeling time-to-assessment within a parametric framework. The first stock assessment (as defined by our criteria above) occurred in 1960, and we therefore used 1960 as the first possible assessment year for stocks that

were first landed prior to 1960. We thus assume, based on the first recorded assessment, that the necessary technology (models, computers to fit models, stock status reporting requirements, etc.) was not available prior to 1960 to conduct a full stock assessment as we have defined. Thus $T = \min(Y_a - Y_l, Y_a - 1960)$, where Y_a is the year of first assessment for assessed stocks or 2013, the last year in our database, for unassessed stocks, and Y_l is the year of first landings in the NOAA database (starting in 1950). In other words, we defined time-to-event (T) as the time between 1960 (or first recorded landings if post-1960) and the first full stock assessment or last record in our database (2013) for unassessed stocks.

The Weibull distribution is often used as a flexible model that has several desirable properties for this type of analysis, and one can easily check whether the Weibull model is appropriate for the data at hand (see Figure C.1). The shape parameter of the Weibull density can be interpreted in terms of the rate of events occurring. A shape parameter >1 suggests an increasing rate of events, whereas a shape parameter <1 indicates a decreasing rate. This allows us to directly estimate the change in assessment rates over time.

A further desirable property is that the estimated regression coefficients can be interpreted both in terms of the ratio of event rates as well as time-to-event probabilities. For example, one can interpret a model coefficient as decreasing or increasing the likelihood of an event occurring at any particular time relative to the baseline (this is usually called the hazard ratio interpretation). Coefficients are estimated for explanatory variables and thus indicate the level and direction of influence of the variable on the base rate of assessment. A coefficient can also be transformed to allow a time-to-event interpretation, where time-to-event parameters represent a multiplicative increase or decrease in the expected time until an event occurs. For example, in a hypothetical scenario, the median time-to-assessment of a demersal stock may be 0.5 times that of a pelagic stock, suggesting that it takes twice as long for pelagic stocks to get assessed. Such acceleration factors are just transformations of the parameters obtained for the event rate interpretation - the

two interpretations are easily exchangeable in the Weibull model.

We thus model time-to-assessment as Weibull-distributed with shape parameter τ and rate λ :

$$T \sim Weibull(\tau, \lambda) \quad (1)$$

The connection between the event rate and the time-to-event interpretations is made explicit by writing the Weibull density as a function of the product of the rate $r(t)$ at which assessments occur, and the probability $A(t)$ of the assessment not occurring prior to time t .

$$f(t) = A(t) \times r(t) \quad (2)$$

$$= \exp(-\lambda t^\tau) \times \lambda \tau t^{\tau-1}, \quad (3)$$

where $A(t) = 1 - P(T \leq t) = 1 - F(t)$, with $F(t) = \exp(-\lambda t^\tau)$ the Weibull distribution function.

We modeled the scale λ of the Weibull distribution as a linear combination of covariates and categorical random effects via a log-link function:

$$\log(\lambda_{i,r,h,c,o,f}) = \beta X_i + \alpha_r + \gamma_h + \kappa_c + \omega_o + \zeta_f, \quad (4)$$

where β is a row-vector of regression coefficients, and X_i is a vector of continuous covariates as well as the binary FMP (fixed) effect. Continuous covariates were taken as the (base 10) logarithms of mean ex-vessel price, maximum landings, their interaction (i.e., mean ex-vessel price \times maximum landings) and species maximum length, all standardised by twice the standard deviation of the variable to allow comparison with the binary predictor for presence in FMPs ([13]). Categorical variables other than presence in FMPs, α (region), γ (habitat), κ (class), ω (order), and ζ_f (family) were all treated as random effects. The model was fit within a Bayesian framework, using Markov Chain Monte Carlo (MCMC) as implemented in the JAGS package. MCMC was run using three

chains of 210 000 iterations each, keeping every 100th iteration, with 10 000 iterations for each chain discarded as burn-in. This provided 6 000 samples from the posterior distribution for each parameter.

The variance of each random effect was given a half-Cauchy prior with a scale of $\Theta = 100$, regression coefficients had vague normal priors with a precision of $1/\sigma^2 = 1e^{-5}$, and τ was estimated using a gamma distribution prior with parameters $a = b = 1e^{-5}$.

3. Results

The number of landed marine stocks in the US (excluding salmons, shads, some benthic invertebrates, and highly migratory species) increased steadily from the 1960s into the 1990s (Figure 1a). During this period, the number of landed stocks in Alaska, West Coast, and Southeast regions approximately doubled, while the number of landed stocks in the Northeast increased more slowly (but was already relatively high at the start of this period). Most of the newly-landed stocks were unassessed throughout this period; by 1996 (the year of the reauthorization of the Sustainable Fisheries Act that required rebuilding of overfished stocks, thus required biomass limits to be estimated), less than 30 stocks in each of the four regions were assessed according to the definition of assessment used here. As a proportion of all landed stocks, however, the trend in assessed stocks has steadily increased in all regions since the 1970s or 1980s (Figure 1b). Currently, the proportion of landed stocks in our dataset that are assessed ranges from 32% of 193 stocks landed in the Southeast to 51% of 83 stocks landed in Alaska. In terms of regional landings (using only landings that are resolved to species level), the assessment of stocks with high landings in each region between the 1970s and 2000s lead to rapidly increasing proportions of total landed tonnage being comprised of assessed populations. By 1996, >93% of landings in Alaska, Northeast, and Southeast regions were comprised of assessed stocks, and in the West Coast this proportion has increased rapidly from 50% in 1996 to >77% currently (Figure 1c).

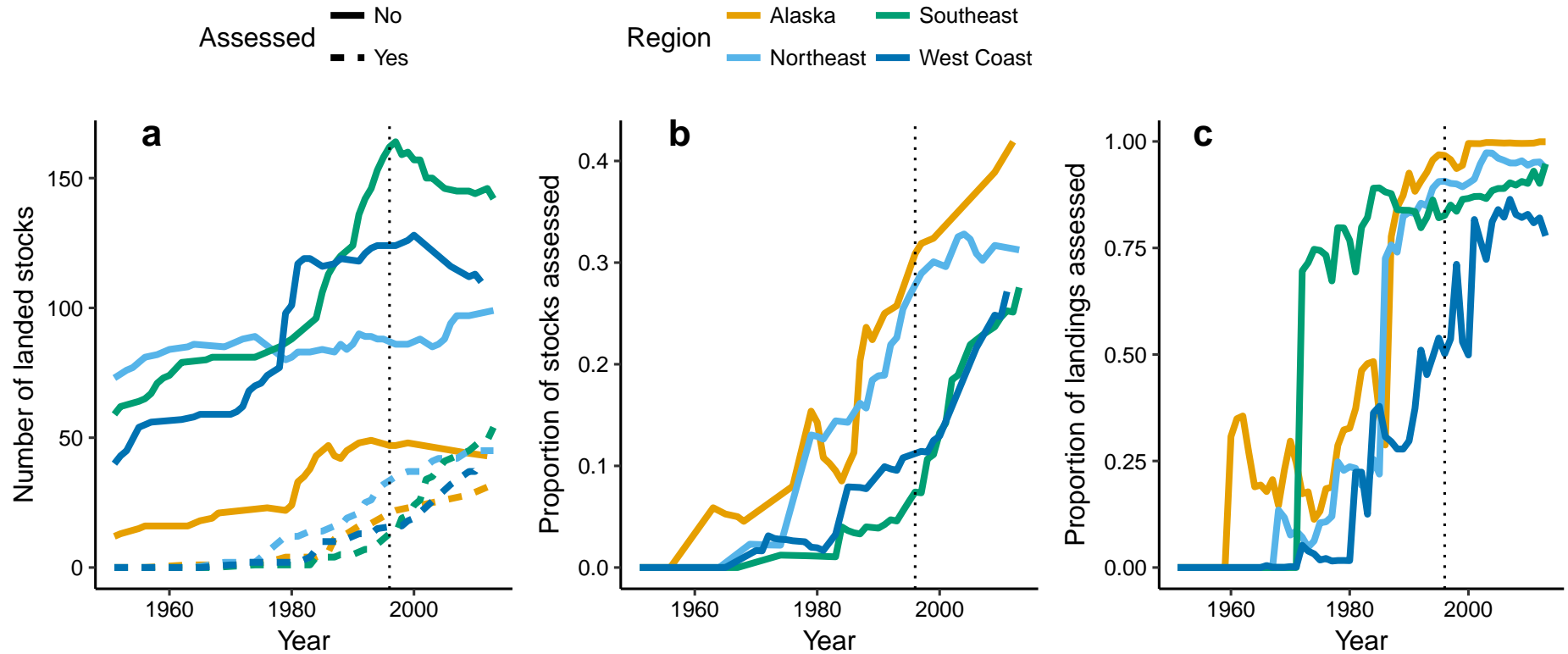


Figure 1: Timeline of a) the number of stocks landed by region and assessment status, b) proportion of landed stocks that are assessed, and c) the proportion of landed tonnage derived from assessed stocks. Values in (a) and (b) are based solely on NOAA landings data resolved to species level, and exclude landings of higher taxonomic groupings (e.g., stock complexes). Some assessed stocks appear only as stock complexes prior to their first assessment; we manually added these stocks to our final dataset, but they lack a complete time-series of species-specific landings so do not appear in (a) or (b). The dotted vertical line marks the reauthorization of the Sustainable Fisheries Act in 1996 that required rebuilding of overfished stocks, and required biomass limits to be estimated.

The majority of landed stocks were fish species (Figure 2a), with Perciformes, Pleuronectiformes and Scorpaeniformes dominating both the number of assessed and unassessed stocks. Among invertebrate taxa, decapod species were the most commonly landed and also most commonly assessed. Demersal species represent a higher proportion of landed populations than species associated with other habitat types (Figure 2b), and also accounted for the highest number and proportion of stock assessments.

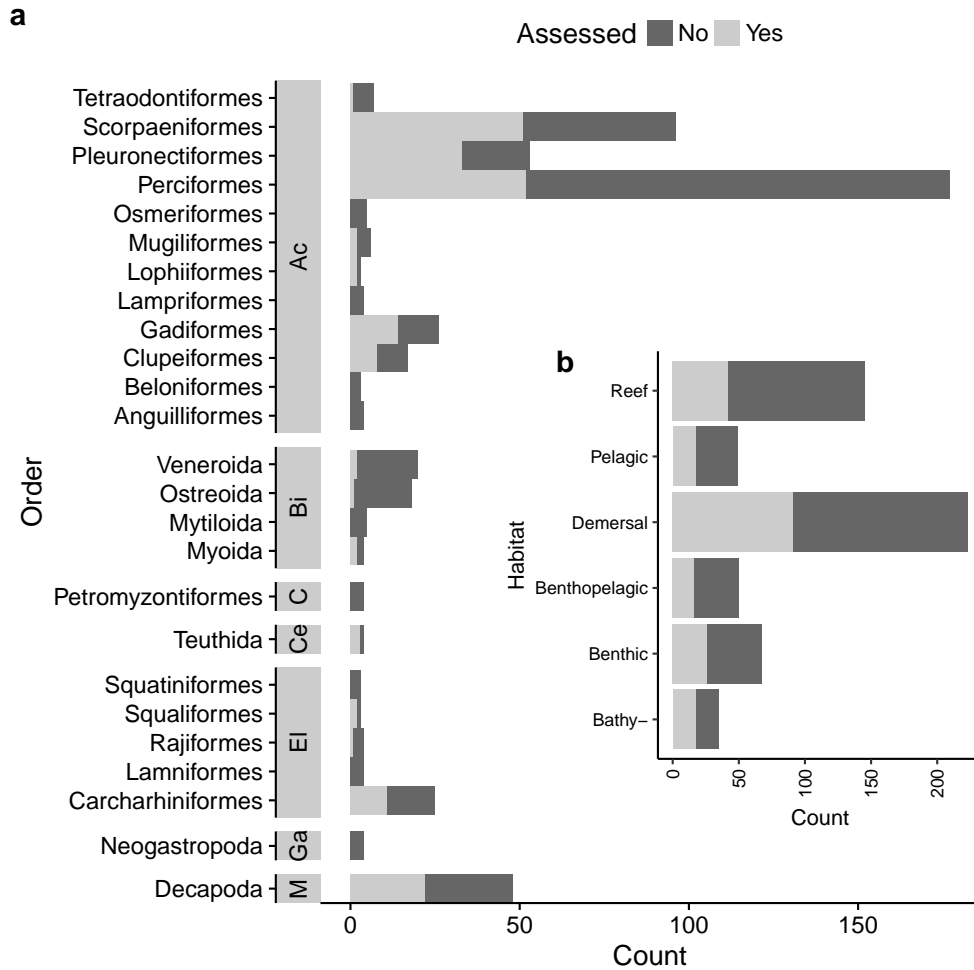


Figure 2: Assessment status at time of last known status (censoring time) a) by taxonomic order and sorted by class and b) by habitat type. In a), classes are abbreviated as Ac: Actinopterygii, M: Malacostraca, Bi: Bivalvia, El: Elasmobranchii, C: Cephalaspidomorphi, Ga: Gastropoda, and Ce: Cephalopoda. Only orders with more than three stocks are shown.

Our time-to-event model effectively disentangled the biological and fishery characteristics that

explaining differences in annual probability of first assessment among stocks (see Appendix Figures C.1 and C.2 for model diagnostics). Among the numerical covariates considered (Figure 3, Table D.1), maximum annual landings and ex-vessel price both had positive and strongly significant impacts on annual assessment probabilities. The effect of landings on assessment probability therefore explains how each region has a large proportion of landed tonnage derived from assessed populations (Fig. 1c), but a smaller proportion of landed stocks being assessed (Fig. 1b). The interaction between price and landings was negative, suggesting that price is more influential when landings are small, and that the landed tonnage drives assessments for species with a low per-kg price. The effect size (per two standard deviations) of maximum body length was smaller than that for price or landings, but was nevertheless greater than zero, suggesting that larger-bodied species have been preferentially assessed.

Fifty-nine percent of 293 stocks in our analysis within federal FMPs had assessments meeting our criteria, compared with 13% of 276 stocks that were not in a FMP. In our time-to-event model, being within a FMP lead to more rapid assessments, but also lowered the effect size of other continuous covariates (Figure D.3), whereas an effect on categorical variables was not noticeable.

Among explanatory random effects, taxonomic factors (order and class) explained a larger portion of residual variance than either habitat or region factors (Figure D.4). This is reflected in the probability of prior assessment in any given year after first being landed (Figures 3 and D.6, Table D.1), for which sea cucumbers (Holothuroidea), squids (Cephalopoda) and sea urchins (Echinoidea) had a slightly higher probability of prior assessment than bony fishes (Actinopterygii) or other taxonomic classes. Rockfishes (Scorpaeniformes), groundsharks (Carcharhiniformes), and flatfishes (Pleuronectiformes) had the highest probabilities of prior assessment among taxonomic orders, each having a higher assessment probability relative to the average of their taxonomic classes (Figure 4, Table D.1). Gadids (Gadidae) also had a high assessment probability relative to the average for bony fishes, while oysters (Ostreoida) had a low probability, even relative to the already-low

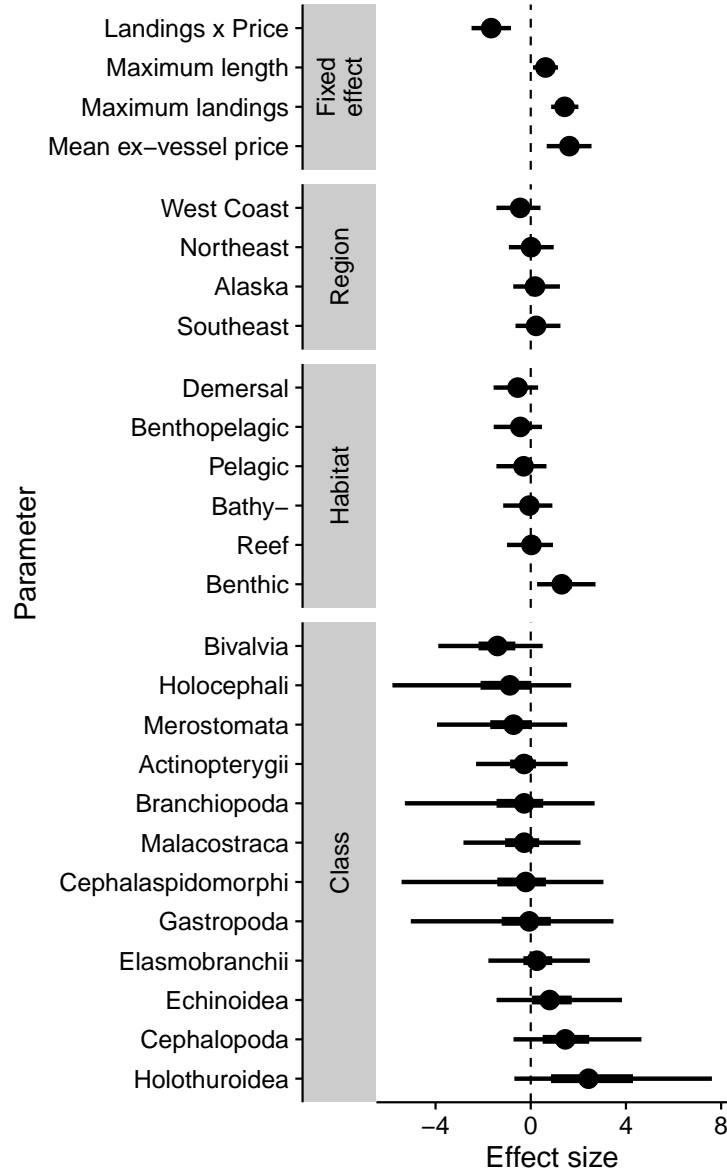


Figure 3: Summaries of estimated posterior distributions for fixed effects, regional random effects, habitat random effects, and taxonomic class random effects in the time-to-event model. Circles show posterior medians, thick bars show inter-quartile ranges of the posteriors, and thin lines show 95% confidence intervals.

average for bivalves. Habitat and regional effects were generally smaller than taxonomic effects. After controlling for other factors, benthic species had a higher probability of assessment than species from other habitats.

Although our model suggests an increasing rate of assessments over time (posterior median for τ : 2.62), all regions show a relatively slow projected increase in the predicted proportion of assessed

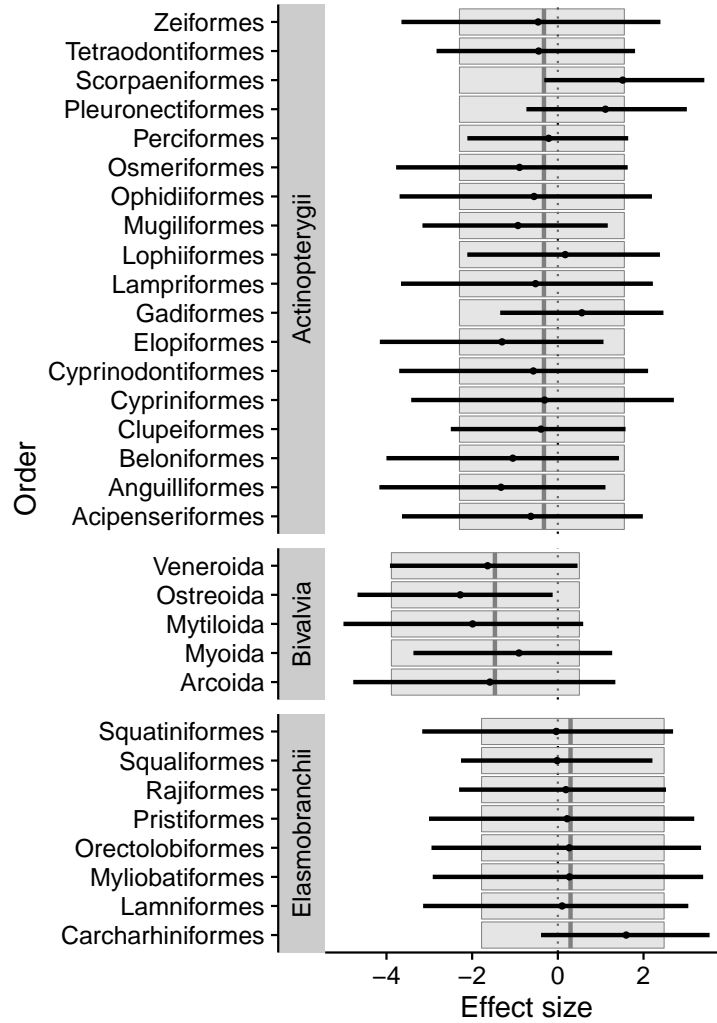


Figure 4: Summaries of estimated posterior distributions for random effects of orders within classes. For classes containing multiple nested orders in the dataset, grey lines show posterior means and coloured boxes show 95% confidence intervals of class effects. Order effects are shown as relative to the class effect within which they are nested, with points showing posterior means and black lines showing 95% confidence intervals.

populations over the next three decades (Figure 5) compared to the rapid increases in the observed proportions of assessed populations over the last 35 years (Fig. 2b). These projections rely on the values of maximum landings and ex-vessel price for all unassessed stocks, and were calculated from the final year of available time series data used for model fitting (usually 2013). The slow predicted increase occurs because stocks with a high assessment probability have typically been assessed, so that remaining stocks have low landings and prices, or other characteristics associated with low

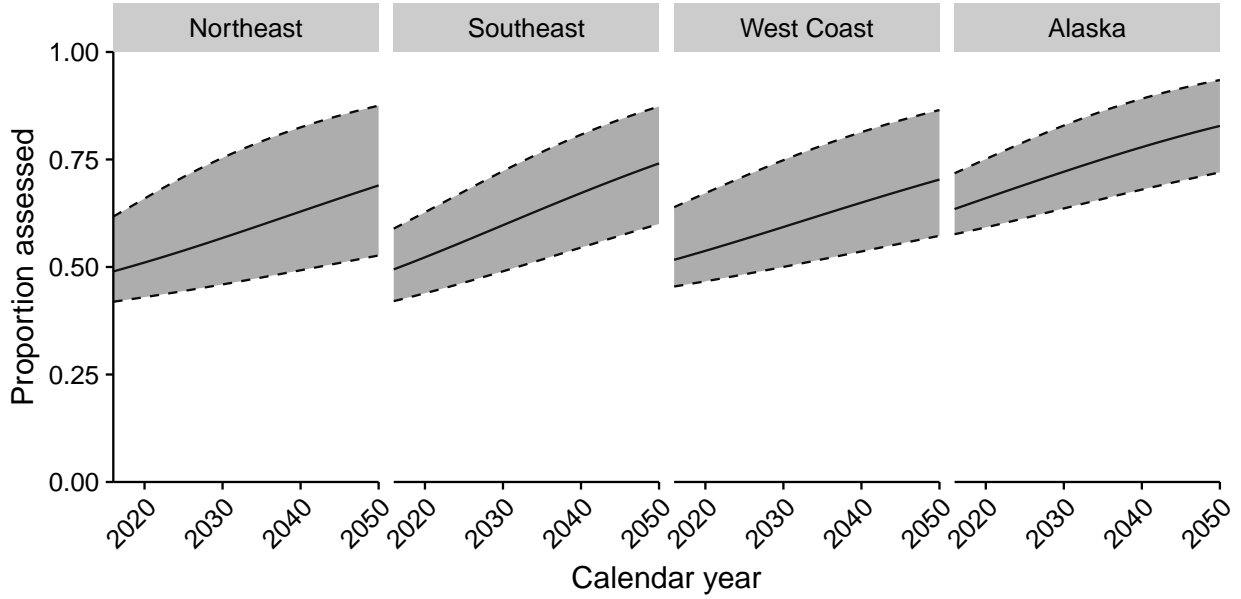


Figure 5: Projected proportion of stocks assessed by region and calendar year, based on assessment probabilities of stocks within each region over the projected year range.

probability of assessment.

4. Discussion

We introduced this study with a common phrase from business management which equally pertains to natural resource management, "what gets measured, gets managed". NMFS currently estimates annual catch limits (ACLs) for the vast majority of fishes in federal fisheries management plans, and has established accountability measures that are triggered whenever recorded annual harvest exceeds the ACLs ([22]). Similarly, state-based management agencies also monitor catches and catch-per-unit-effort for many species, and implement management actions when this is deemed necessary. Thus, NMFS and other agencies both measure and manage annual harvest for the majority of US fishes. However, different methods are used for setting ACLs. Catch-only models (COMs), used to estimate ACLs for the majority of stocks in most federal US management regions ([1]), do not estimate population size relative to management targets ([8, 37]). In some cases, it may be possible to rebuild or maintain fish and invertebrate stocks at levels of sustainable

harvest using COMs, without using a stock-assessment model as defined by our criteria ([e.g., 35, 36]). Specifically, COMs can be used to develop a harvest plan with fishing at a proportion of the estimated ACL, which is expected to have a pre-specified probability of maintaining population abundance near management targets ([36]). Nevertheless, we have excluded COMs from our definition of “stock assessments”. We see two main benefits to measuring population abundance for marine fishes beyond simply estimating ACLs:

1. COMs generally involve managing a fishery to target a constant annual harvest, which is chosen to perform adequately on average: some stocks may be overfished, others may be underfished, but the average stock is fished at target fishing rates. However, even these target rates may induce unsustainable fishing mortality if population abundance is low. In contrast, stock assessments are likely to perform adequately over time for each individual stock: some decades may be overfished, others may be underfished, but on average over time each individual stock is fished sustainably. Both approaches are expected to perform well on average, but stock assessments improve the expected performance for each stock individually. This advantage is important for both conservationists and fishers who do not wish to see any given individual stock overfished, irrespective of whether or not the average stock is overfished. Not only does overfishing pose a conservation challenge for depleted stocks, but may impose stricter fishing limits for other stocks as a result of bycatch limits for the depleted stocks ([15, 19]).

2. The ability of stock assessments to inform harvest plans based on updated data has been repeatedly shown to improve management outcomes ([e.g., 3]). For example, managing with a harvest control rule in which fishing mortality targets are updated based on stock assessment estimates of population abundance can substantially decrease variability in abundance and fishery catches (assuming the assessment does not change fundamentally over time), even relative to cases where a COM estimates sustainable fishing mortality rates perfectly ([33]). Updating harvest plans based on new data can also prevent cases in which COMs over-estimate a sustainable fishing

rate, which would otherwise collapse the fishery ([36]). We therefore see benefits both to ocean conservation and to fishing industries by continuing to transition from management based on COMs to management based on stock assessments with associated biological reference points.

There are many differences in quality and complexity among stock assessments. NMFS categorizes assessments using six "levels", and high-level assessments are distinguished by having more or higher-quality data assimilated, using a model that allows for greater attention to biological mechanisms and realism. We have ignored these subtler distinctions here, and have instead used a single cutoff criterion, which essentially falls between statistical population models and COMs. In general, our classification of unassessed stocks aligned with SIS categories 0–2, while our assessed stocks aligned with SIS categories 3–5 (Appendix B). We suggest that future research could build on our model to include annual probabilities of transitioning among multiple categories (e.g., among all six NMFS assessment levels). This future analysis would allow greater detail regarding historical changes over time in the average quality of stock assessments, and may show alternative patterns among regions dependent on the level of assessment complexity considered. NMFS has already embarked upon the task of defining and compiling records of different assessment types and qualities, so this research will soon be feasible for stocks in US-federal jurisdiction.

Given our operational definition of stock assessment, maximum landings and ex-vessel price were particularly strong predictors of the year in which stocks were first assessed. The product of landings and price is a rough measure of the gross economic value to commercial fisheries derived from fish and invertebrate stocks. Fisheries managers and scientists must choose among several candidate species in a given region and devote stock assessment time and resources towards only a subset of these. Our results suggest that fisheries managers prioritize stocks with high commercial value, and this result is consistent with previous research showing that fishery development is also driven primarily by landed tonnage and ex-vessel prices of fished species ([31]). We were unable to identify a variable proportional to recreational value that was consistently measured across all regions, so

we cannot estimate the potential impact of economic value for recreational fishing on assessment probabilities. However, we acknowledge that recreational value and take will likely have an impact on both population dynamics and management in some regions, and especially for state-managed nearshore stocks.

The presence of stocks within FMPs lead to accelerated time-to-assessment, while effects associated with landings and price were smaller when presence in an FMP was included in our analysis. This trade-off is likely driven by the strong dependence of presence in FMPs on landings and price. Indeed, logistic regression of the presence in a FMP against the same linear predictors used in our time-to-event model (without the FMP factor) shows similarly strong effects of landings and price, and qualitatively similar results to our time-to-event analysis (not shown). It thus appears that landings and price are primary drivers on the path towards assessment, and that placing stocks within FMPs increases the momentum towards a full assessment for these stocks.

We suspect that the positive effect of body size on time-to-assessment may be due to the large number and diversity of small-bodies stocks, especially invertebrates, that are landed in low numbers and at a relatively low ex-vessel price. Since these stocks do not have a consistent price or landings, and are taxonomically diverse, it may be more parsimonious within the model to attribute the lack of an assessment for these stocks to body length as opposed to taxonomic random effects.

Certain taxonomic classes, or orders within classes, stood out as being more likely to have undergone a stock assessment after controlling for landings, ex-vessel price, and other factors. Elasmobranchs, and in particular groundsharks (Carchariniiformes), had relatively high rates of stock assessment when controlling for other variables. This likely results from increasing conservation interest in recent decades for shark species both in the US and worldwide ([11]). This high assessment rate after accounting for maximum landings may also result in part from the high discard rates of small coastal shark species often caught as bycatch in shrimp trawl or other fisheries ([6]). Due to bycatch, our database values for shark landings may be smaller than true harvest, thus resulting in

a compensatory increase in the estimated assessment rate for this taxon. A similar effect may drive time-to-assessment probabilities of popular recreational species, for which the true harvest mortality is probably substantially higher than those reported in the NOAA landings database. Such under-reported total landings may explain higher-than-average assessment probabilities found for flatfishes (Pleuronectiformes), which are often estuarine or nearshore species. Among bony fishes, rockfishes and greenlings (Scorpaeniformes) also had high rates of assessment, likely due to the number of Pacific rockfishes included, which have been a topic of conservation concern in Alaska and the US West Coast ([4, 5, 24]). While cephalopod abundance is commonly estimated using catch-per-unit-effort indices or survey abundance indices ([9]) rather than stock assessments, in the US most landed cephalopods are assessed (all are squid species). This may result from defined units of assessment having coast-wide distributions rather than assuming a more disaggregated stock structure in which only some of the stocks would be assessed.

Results from our model could be used to evaluate and control for systematic differences between assessed and unassessed US stocks in other analyses. These differences are important because meta-analysis of assessed stocks is widely used to understand management performance and biological characteristics of marine fishes in general ([34]). To account for systematic differences between assessed and unassessed stocks, authors could use our model within a "propensity score matching" or propensity score weighting framework ([29, 20]). For example, pairwise comparisons (or matching) between assessed and unassessed stocks should involve stocks with similar likelihoods of being assessed. Similarly, calculated propensity scores can be used as predictor variables in regressions involving variables of interest to control for the non-random assessment probabilities among analyzed stocks. If analysts find systematic differences in management outcomes or biological characteristics between assessed and unassessed stocks (e.g., systematic differences in recruitment compensation), then the relationship between the propensity of assessment and the variable of interest can be used to improve predictions for unassessed stocks.

Fish and invertebrate stocks in the US are reaching saturation with respect to the rate of first assessment. Even though most stocks in all regions are as yet unassessed (Figure 1b), the predicted rate of increase in assessed stocks over the next few decades is slower than the rate observed over the last few decades because the stocks most likely to be assessed have already been assessed. Furthermore, fisheries management agencies must weigh the benefit of attempting to obtain full assessments for new stocks against the need to update assessments of previously-assessed stocks. Since the latter have higher commercial value, and updating assessments is often more straightforward than collecting data and developing new assessments for previously-unassessed species, the percentage of assessed stocks is unlikely to attain 100%. Rather, at some point funding and resource limitations will likely lead to a steady state where no new stocks will be assessed and priority is given to updating existing assessments for stocks of high commercial, recreational and ecological value or concern.

It is not necessarily the case that stock assessments are required for effectively managing fish and invertebrate stocks, as harvest control rules or in-season adjustments to fishing effort can be based on fishery-independent survey indices or fishery-dependent catch-per-unit effort indices rather than on stock status estimates from assessments. However, a logical leap from "what gets measured, gets managed" to "what is better measured, is better managed" suggests the value of better estimating stock status through the use of stock assessments (population models fitted statistically to abundance index or age/length composition data, where estimated biomass is compared with explicit target levels). Further improvements in management performance given current resources could also be attained by improved methods for prioritizing which stocks to assess ([23]).

5. Acknowledgements

We thank Jeannette Banobi for compiling year of first stock assessment records in the management attributes database, and Nicole Baker for conducting online searches to identify additional assess-

ments. We thank Cody Szuwalski for productive discussions while planning out the data set and analysis, and Chantal Wetzel for helpful comments on the manuscript. We are grateful to fisheries scientists and managers who participated in interviews to characterize management attributes of US stocks including the year of first stock assessment. These individuals include: Ingrid Spies, Grant Thompson (Alaska Fisheries Science Center); Scott Kelley, Tim Baker, Gretchen Bishop, Dave Harris, Tim Sands, Mark Stichert, Joe Stratman, Jeff Wadle (Alaska Department of Fish and Game); Edward Poulsen (Alaska Bering Sea Crabbers Association); John Devore, Mike Burner, Kerry Griffin (Pacific Fishery Management Council); Paul Crone (Southwest Fisheries Science Center); Steven Hare (International Pacific Halibut Commission); Dan Ayres, Heather Reed, Bob Sizemore, Mike Ulrich, Lorna Wargo (Washington Department of Fish and Wildlife); Kelly Corbett, Scott Groth, Robert Hannah (Oregon Department of Fish and Wildlife); Kristine Barsky, Peter Kalvass (California Department of Fish and Game), Toni Chute, Ruth Haas-Castro, Dvora Hart, Joseph Idoine, Larry Jacobsen, Chris Legault, Loretta O'Brien, Paul Rago, Gary Shepherd, Fred Serchuk, Katherine Sosebee (Northeast Fisheries Science Center); Andy Applegate, Phillip Haring, Chris Kellog, Tom Nies, Lori Steele (New England Fishery Management Council); Robert Beale, Toni Kerns, Nichola Meserve, Brad Spear, Christopher Vonderweidt, Michael Waine (Atlantic States Marine Fisheries Commission); Jason Didden, Tom Hoff (Mid-Atlantic Marine Fisheries Commission); Derek Orner (National Marine Fisheries Service); John Carlson, Shannon Cass-Calay, Nancy Cummings, Enric Cortes, Rick Hart, Erik Williams (Southeast Fisheries Science Center); Sue Gerhart, Peter Hood, Jack McGovern, Andy Strelcheck (NOAA Southeast Regional Office); Gregg Waugh (South Atlantic Marine Fisheries Commission); Tom Matthews, and Bill Sharp (Florida Fish and Wildlife Conservation Commission). M.C.M. was supported in this work by the Walton Family Foundation and by a NSERC Banting Fellowship.

References

- [1] Jim Berkson and James T. Thorson. “The determination of data-poor catch limits in the United States: is there a better way?” In: *ICES Journal of Marine Science: Journal du Conseil* 72.1 (Jan. 2015), pp. 237–242.
- [2] Carl Boettiger, Duncan Temple Lang, and Peter Wainwright. “rfishbase: exploring, manipulating and visualizing FishBase data from R”. In: *Journal of Fish Biology* (2012).
- [3] Thomas R. Carruthers et al. “Evaluating methods for setting catch limits in data-limited fisheries”. In: *Fisheries Research* 153 (May 2014), pp. 48–68.
- [4] W. G Clark. “The effect of recruitment variability on the choice of a target level of spawning biomass per recruit”. In: *Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations*. Ed. by G. Kruse et al. Fairbanks, AK: University of Alaska, Alaska Sea Grant Report 93-02, 1993, pp. 233–246.
- [5] Flaxen Conway and Wesley Shaw. “Socioeconomic Lessons Learned from the Response to the Federally-Declared West Coast Groundfish Disaster”. In: *Fisheries* 33.6 (June 2008), pp. 269–277.
- [6] Enric Corts. *Stock assessment of small coastal sharks in the U.S. Atlantic and Gulf of Mexico*. Sustainable Fisheries Division Contribution SFD-01/02-152. National Marine Fisheries Service, Southeast Fisheries Science Center, 2002, p. 133. URL: http://www.nmfs.noaa.gov/sfa/hms/species/sharks/documents/scs_assessment_rev2.pdf.
- [7] Catherine M. Dichmont et al. “A review of stock assessment packages in the United States”. In: *Fisheries Research* 183 (2016), pp. 447–460.
- [8] E. J. Dick and A. D MacCall. “Depletion-Based Stock Reduction Analysis: A catch-based method for determining sustainable yields for data-poor fish stocks”. In: *Fisheries Research* 110.2 (2011), pp. 331–341.

- [9] Zo A. Doubleday et al. “Global proliferation of cephalopods”. In: *Current Biology* 26.10 (2016), R406–R407.
- [10] N. A. Dowling et al. “From low- to high-value fisheries: Is it possible to quantify the trade-off between management cost, risk and catch?” In: *Marine Policy* 40 (2013), pp. 41–52. ISSN: 0308-597X. DOI: 10.1016/j.marpol.2012.12.009. URL: <http://www.sciencedirect.com/science/article/pii/S0308597X12002503> (visited on 12/08/2016).
- [11] Nicholas K Dulvy et al. “Extinction risk and conservation of the worlds sharks and rays”. In: *eLife* 3 (2014), e00590. ISSN: 2050-084X. DOI: 10.7554/eLife.00590.
- [12] Rainer Froese, Daniel Pauly, et al. *FishBase*. 2012.
- [13] Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.
- [14] Thomas E Helser et al. “Stock assessment of Pacific hake (whiting) in US and Canadian waters in 2006”. In: *Pacific Fishery Management Council* 7700 (2006), pp. 97220–1384.
- [15] RAY Hilborn et al. “Defining Trade-Offs among Conservation, Profitability, and Food Security in the California Current Bottom-Trawl Fishery”. In: *Conservation Biology* 26.2 (2012), pp. 257–268. ISSN: 1523-1739.
- [16] Jeffrey A. Hutchings et al. “Trends in the abundance of marine fishes”. In: *Canadian Journal of Fisheries and Aquatic Sciences* 67.8 (2010), pp. 1205–1210. (Visited on 12/08/2016).
- [17] Olaf P Jensen, Trevor A Branch, and Ray Hilborn. “Marine fisheries as ecological experiments”. In: *Theoretical Ecology* 5.1 (2012), pp. 3–22.
- [18] Alec D MacCall. “Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations”. In: *ICES Journal of Marine Science: Journal du Conseil* 66.10 (2009), pp. 2267–2271.

- [19] Michael C. Melnychuk, Jeannette A. Banobi, and Ray Hilborn. “Effects of Management Tactics on Meeting Conservation Objectives for Western North American Groundfish Fisheries”. In: *PLOS ONE* 8.2 (2013), pp. 1–15. DOI: 10.1371/journal.pone.0056684. URL: <http://dx.doi.org/10.1371/journal.pone.0056684>.
- [20] Michael C Melnychuk et al. “Can catch share fisheries better track management targets?” In: *Fish and Fisheries* 13.3 (2012), pp. 267–290.
- [21] Michael C Melnychuk et al. “Fisheries management impacts on target species status”. In: *Proceedings of the National Academy of Sciences* 114.1 (2017), pp. 178–183.
- [22] R. D. Methot et al. “Implementing a science-based system for preventing overfishing and guiding sustainable fisheries in the United States”. In: *ICES Journal of Marine Science: Journal du Conseil* 71.2 (2014), pp. 183–194.
- [23] Richard D. Methot. *Prioritizing fish stock assessments*. NOAA Tech. Memo. Washington, D.C.: NMFS, NOAA, US Department of Commerce, 2015, p. 31. URL: https://www.st.nmfs.noaa.gov/Assets/stock/documents/PrioritizingFishStockAssessments_FinalWeb.pdf.
- [24] R. A. Myers, K. G. Bowen, and N. J. Barrowman. “Maximum reproductive rate of fish at low population sizes”. In: *Canadian Journal of Fisheries and Aquatic Sciences* 56.12 (1999), pp. 2404–2419.
- [25] David Newman, Jim Berkson, and Lisa Suatoni. “Current methods for setting catch limits for data-limited fish stocks in the United States”. In: *Fisheries Research* 164 (2015), pp. 86–93.
- [26] Fisheries Statistics Division NOAA Fisheries. *Annual Commercial Landing Statistics*. <http://www.st.nmfs.noaa.gov/commercial-fisheries/commercial-landings/annual-landings/index>.

- [27] Fisheries Statistics Division NOAA Fisheries. *Fisheries of the United States*. [http://www.st.nmfs.noaa.gov/Assets/commercial/fus/fus15/documents/03_\"Aquaculture2015.pdf](http://www.st.nmfs.noaa.gov/Assets/commercial/fus/fus15/documents/03_\).
- [28] MLD Palomares and D Pauly. “SeaLifeBase”. In: *World Wide Web Electronic Publication*, www.sealifebase.org, version 2.10 (2016) (2010).
- [29] PAUL R. Rosenbaum and DONALD B. Rubin. “The central role of the propensity score in observational studies for causal effects”. In: *Biometrika* 70.1 (1983), pp. 41–55. DOI: 10.1093/biomet/70.1.41. eprint: <http://biomet.oxfordjournals.org/content/70/1/41.full.pdf+html>. URL: <http://biomet.oxfordjournals.org/content/70/1/41.abstract>.
- [30] NOAA National Marine Fisheries Service. *Species Information System Public Portal*. <https://www.st.nmfs.noaa.gov/sisPortal/sisPortalMain.jsp>.
- [31] S. A. Sethi, T. A. Branch, and R. Watson. “Global fishery development patterns are driven by profit but not trophic level”. In: *Proceedings of the National Academy of Sciences* 107.27 (June 2010), pp. 12163–12167.
- [32] Tim D. Smith. *Scaling Fisheries: The Science of Measuring the Effects of Fishing, 1855-1955*. 1st ed. Cambridge, UK: Cambridge University Press, Aug. 2007. ISBN: 0-521-03896-0.
- [33] James T. Thorson, Olaf P. Jensen, and Ray Hilborn. “Probability of stochastic depletion: an easily interpreted diagnostic for stock assessment modelling and fisheries management”. In: *ICES Journal of Marine Science: Journal du Conseil* 72.2 (2015), pp. 428–435.
- [34] James T. Thorson et al. “Giants’ shoulders 15 years later: lessons, challenges and guidelines in fisheries meta-analysis”. In: *Fish and Fisheries* 16.2 (2015), pp. 342–361.
- [35] Chantell R. Wetzel and Andr E. Punt. “Model performance for the determination of appropriate harvest levels in the case of data-poor stocks”. In: *Fisheries Research* 110.2 (2011), pp. 342–355.

- [36] Chantell R. Wetzel and Andr E. Punt. “Performance of a fisheries catch-at-age model (Stock Synthesis) in data-limited situations”. In: *Marine and Freshwater Research* 62.8 (2011), pp. 927–936.
- [37] John Wiedenmann, Michael J. Wilberg, and Thomas J. Miller. “An Evaluation of Harvest Control Rules for Data-Poor Fisheries”. In: *North American Journal of Fisheries Management* 33.4 (2013), pp. 845–860.
- [38] B. Worm et al. “Rebuilding Global Fisheries”. In: *Science* 325.5940 (July 30, 2009), pp. 578–585. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1173146. URL: <http://www.sciencemag.org/cgi/doi/10.1126/science.1173146>.

A. Assignment of year of first stock assessment

We used a three-step process to classify stocks as assessed or unassessed and, for assessed stocks, to identify the year of first assessment.

1. Management attribute database and interviews: As part of a larger data collection effort to characterize management attributes at the stock level, we conducted a survey which included the question "Year of first stock assessment", with a description of what qualifies as an assessment (as outlined in the main text). To answer this question, we first looked through historical stock assessments archived on fisheries agency websites to identify the first assessment that met our defined criteria (as not all published assessments would meet our relatively strict criteria). Websites accessed included those of US Fisheries Management Councils, NMFS Science Centers, and state-level fisheries management agencies. To confirm our findings, or if we were unable to answer the question from archived assessments, we conducted interviews with fisheries scientists and managers familiar with one or more stocks in each region. Individuals who participated in interviews for management attributes for US stocks are listed in our Acknowledgments. These surveys were conducted for 196 US stocks, of which 165 were found to have a stock assessment as defined. The intention of these surveys was to cover primarily well-studied stocks and stocks important to regional fisheries, not to draw a random sample from the NOAA landings data. Therefore, the proportion of these stocks meeting the criteria of having a stock assessment is high compared to the proportion of stocks from the NOAA landings database that are assessed.
2. Comparison with SIS database: We compared our list of assessed stocks with the list of assessed stocks from the US Species Information System (SIS) database. We considered SIS stocks with assessment level 3 or greater (see Appendix B). We found 31 stocks in the SIS database that were not previously included in our list of assessed stocks in (1). Most of these stocks were first assessed after our period of data collection in (1), 2012–2015, which is why many of these were not originally included in our list of stocks in (1). We added these 31 stocks along with their year of first assessment to our dataset.
3. Remaining stocks in NOAA landings database: To ensure that we had not overlooked any US marine stocks with a stock assessment (including stocks managed by state agencies), we systematically searched online for a stock assessment for each species in the NOAA landings database. These searches comprised two types of species: some species in the NOAA landings database were not previously included in our collection of stock management attribute data in (1) or (2); and some species in the NOAA landings database were previously included in (1) or (2) but also had landings recorded in states outside of the defined areas of distribution of those stocks, and thus other assessments of the same species may have been available, for other areas. For both types of species, we searched for any stock assessment corresponding to species \times state recorded landings that were unaccounted for in (1) or (2). Online searches consisted of: (i) going through assessment archives on the websites of US Fisheries Management Councils, NMFS Science Centers, and state-level fisheries management agencies; and (ii) using the Google search engine with search terms ("Latin name of species" "stock assessment" "Region") to identify available assessments either in the primary literature or that may not have been otherwise available through agency websites. In addition to confirming many of the added assessments in (2), we found 15 stock assessments meeting our criteria for assessment which were not previously accounted for. Many of these were stocks managed by

state agencies so did not appear in the list in (2). These 15 stocks were added to our dataset along with their year of first assessment.

In total, after steps 1–3 we generated a list of 242 US stocks, including 211 stocks with stock assessments that qualified under our defined criteria. The remaining 31 stocks did not meet our assessment criteria. They were excluded from our dataset, although after merging our dataset of assessed stocks with the NOAA landings database, those stocks would later re-enter as unassessed stocks.

B. Validation of assessment classifications

In this appendix, we compare our assessment classifications with those of NOAA’s Species Information System (SIS) database.

Our classification system consisted of whether or not a stock has had an assessment conducted, and subsequently, the year in which the first assessment occurred. Our definition of a stock assessment required the use of a population dynamics model fit to fishery landings data, coupled with some benchmark with which to compare model-estimated time series of abundance or fishing mortality (described further in the main text). These biological reference point benchmarks may have been estimated within the same assessment model or specified externally, and allow for comparing estimates of current abundance or fishing mortality relative to target levels. To determine the year of first stock assessment, we reviewed historical assessments on the websites of US Fishery Management Councils and NMFS Science Centers, and also sought the input of fishery scientists and managers within each region (Appendix A).

The SIS database uses a 6-level categorization of assessments ([30]), as follows:

0. Although some data may have been collected on this species, these data have not been examined beyond simple time series plots or tabulations of catch.
1. Either:
 - a time series of a (potentially imprecise) abundance index calculated as raw or standardized CPUE in commercial, recreational, or survey vessel data, or
 - onetime estimation of absolute abundance made on the basis of tagging results, a depletion study, or some form of calibrated survey.
2. Simple equilibrium models applied to life history information; for example, yield per recruit or spawner per recruit functions based on mortality, growth, and maturity schedules; catch curve analysis; survival analysis; or length-based cohort analysis.
3. Equilibrium and non-equilibrium production models aggregated both spatially and over age and size; for example, the Schaefer model and the Pella-Tomlinson model.
4. Size, stage, or age structured models such as cohort analysis and untuned and tuned VPA analyses, age-structured production models, CAGEAN, stock synthesis, size or age-structured Bayesian models, modified DeLury methods, and size or age-based mark-recapture models.
5. Assessment models incorporating ecosystem considerations and spatial and seasonal analyses in addition to Levels 3 or 4. Ecosystem considerations include one or more of the following:
 - a one or more time-varying parameters, either estimated as constrained series, or driven by environmental variables,
 - multiple target species as state variables in the model, or
 - living components of the ecosystem other than the target species included as state variables in the model.”

We expect our classification of ”assessed” to align with levels 3–5 in the SIS database, and our classification of ”unassessed” to align with levels 0–2. This is generally what we found, and the discrepancies are described below. These comparisons can be followed in our final dataset provided (supplementary dataset), which list for each stock in our dataset the corresponding stock from the SIS database. In summary, the comparison shows:

I) Of the 211 stocks in our final dataset with a year of first stock assessment assignment, 189 have a corresponding stock in the SIS database. Of these 189 overlapping stocks, 161 (85%) had an assessment level of 3 or greater assigned in the SIS database. The 28 stocks that we classified as assessed but that have a current assessment level less than 3 in the SIS database fall into three groups. First, the most recent assessment has been rejected during review for 7 stocks that would have otherwise had a SIS assessment level of 3 or greater:

1. Atlantic cod - Georges Bank
2. Hogfish - Carolinas
3. Blacknose shark - Gulf of Mexico
4. Blacktip shark - Atlantic
5. Bonnethead - Atlantic and Gulf of Mexico
6. Caribbean spiny lobster - Southern Atlantic Coast / Gulf of Mexico
7. Atlantic halibut - Northwestern Atlantic Coast

Because qualifying assessments for these stocks have been accepted and used for management in the past, we treat these as assessed stocks in our analysis and maintain their year of first assessment.

Second, 20 stocks currently have a SIS assessment level less than 3 despite having had a qualifying assessment in the past:

1. Silver hake - Gulf of Maine / Northern Georges Bank
2. Silver hake - Southern Georges Bank / Mid-Atlantic
3. Goosefish - Gulf of Maine / Northern Georges Bank
4. Goosefish - Southern Georges Bank / Mid-Atlantic
5. Tilefish - Mid-Atlantic Coast
6. Bluefish - Atlantic Coast
7. Atlantic croaker - Mid-Atlantic Coast
8. Atlantic mackerel - Gulf of Maine / Cape Hatteras
9. Scup - Atlantic Coast
10. Butterfish - Gulf of Maine / Cape Hatteras
11. Yellowtail flounder - Georges Bank
12. Winter flounder - Gulf of Maine
13. Ocean quahog - Atlantic Coast
14. Atlantic surfclam - Mid-Atlantic Coast
15. Opalescent inshore squid - Pacific Coast

16. Longfin inshore squid - Georges Bank / Cape Hatteras
17. Northern shortfin squid - Northwestern Atlantic Coast
18. Spiny dogfish - Atlantic Coast
19. Red deepsea crab - Northwestern Atlantic
20. Horseshoe crab - Atlantic

The reasons for these discrepancies can be attributed to a population model (e.g. VPA) coupled with biological reference points previously used in the stock assessment process to inform fisheries management, but currently assessments rely on simpler models with lower assessment level categories in the SIS database (e.g. index-based methods or simpler equilibrium models).

Third, one additional stock that was assessed in 2016 for the first time is not in a federal FMP and not included in SIS as having had an assessment: White seabass - Pacific Coast.

II) Of the 358 stocks in our final dataset that were considered unassessed (based on species in the NOAA landings-by-state database that did not pair with assessed stocks) 143 have a corresponding stock in the SIS database. Of these 143 overlapping stocks, 138 (97%) had an assessment level of 2 or less assigned in the SIS database, and an additional 3 (2%) had a recent assessment rejected that would have been a level 4 assessment had it not been rejected. The two stocks that we classified as unassessed but that have an assessment level of 3 or 4 in the SIS database are:

1. Yelloweye rockfish - Bering Sea / Aleutian Islands; Yelloweye rockfish - Gulf of Alaska
2. Stone crabs (*Menippe* spp.) - Gulf of Mexico

Stock assessments for these stocks are based on area-swept biomass estimates from surveys, other index-based methods, or do not appear to be coupled with biological reference points to meet our definition of a full assessment. To be consistent with other stocks in our analysis, we consider these stocks as unassessed.

C. Model fit

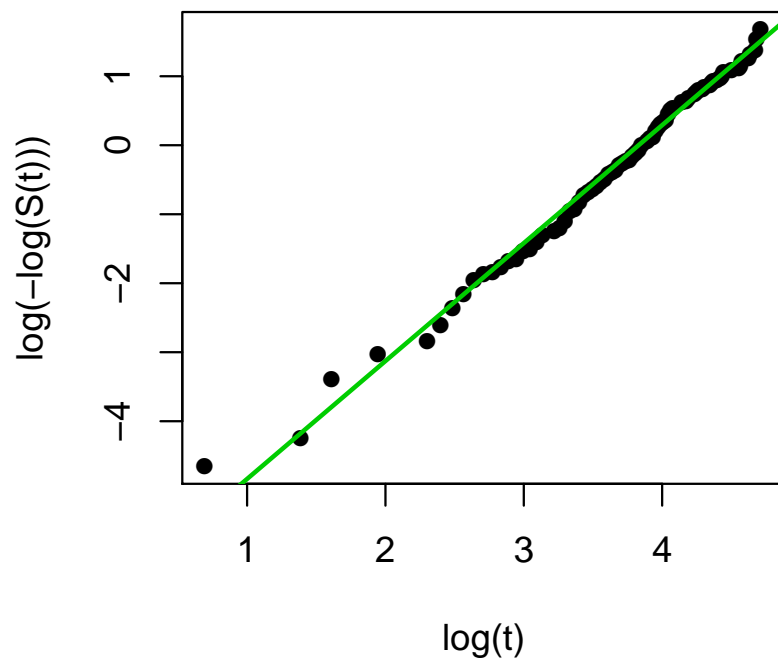


Figure C.1: Appropriateness of the Weibull event-time model for the time-to-assessment dataset. If the Weibull applies, the time from first landings (or from first quantitative stock assessment in 1960 if a stock was landed before 1960) to the year of first assessment should fall on a line with slope τ (the Weibull shape parameter) between $\log(-\log(\hat{S}(t)))$, where $\hat{S}(t)$ is the non-parametric Kaplan-Meier estimate of survival at time t , and the log of t . Here, τ evaluates to 1.71 (slope of the green line), suggesting an increasing assessment rate with increasing time t .

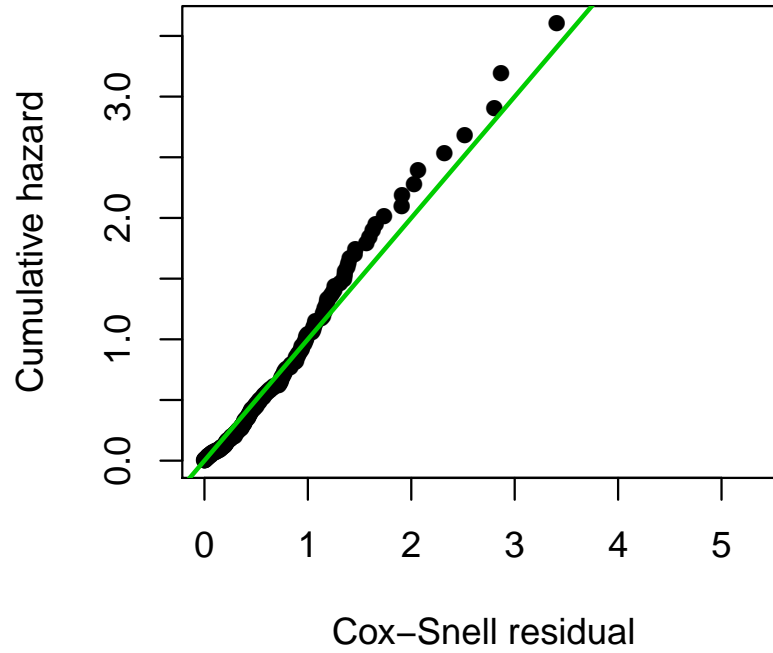


Figure C.2: Model fit of the Weibull survival model, based on Cox-Snell residuals calculated at the posterior mean of the linear predictor. For a perfect fit all data points would lie on the $y=x$ (green) line.

D. Model estimates and predictions

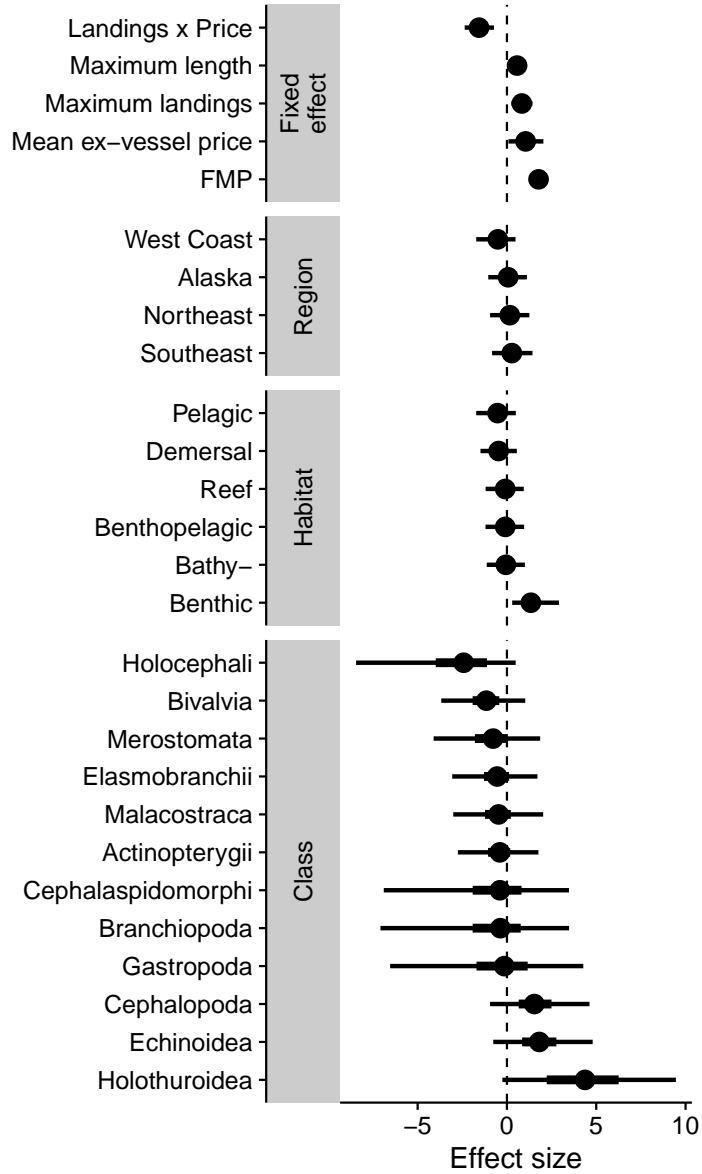


Figure D.3: Summaries of estimated posterior distributions for fixed effects, regional random effects, habitat random effects, and taxonomic class random effects in the time-to-event model with fishery management plan (FMP) variable. Circles show posterior medians, thick bars show inter-quartile ranges of the posteriors, and thin lines show 95% confidence intervals.

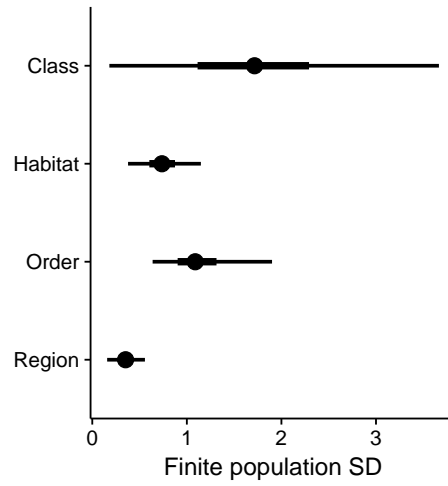


Figure D.4: Comparison of finite population standard deviation (i.e., variance attributed to each variable) for random effects in the Weibull survival model. Circles show posterior medians, thick bars show inter-quartile ranges of the posteriors, and thin lines show 95% confidence intervals.

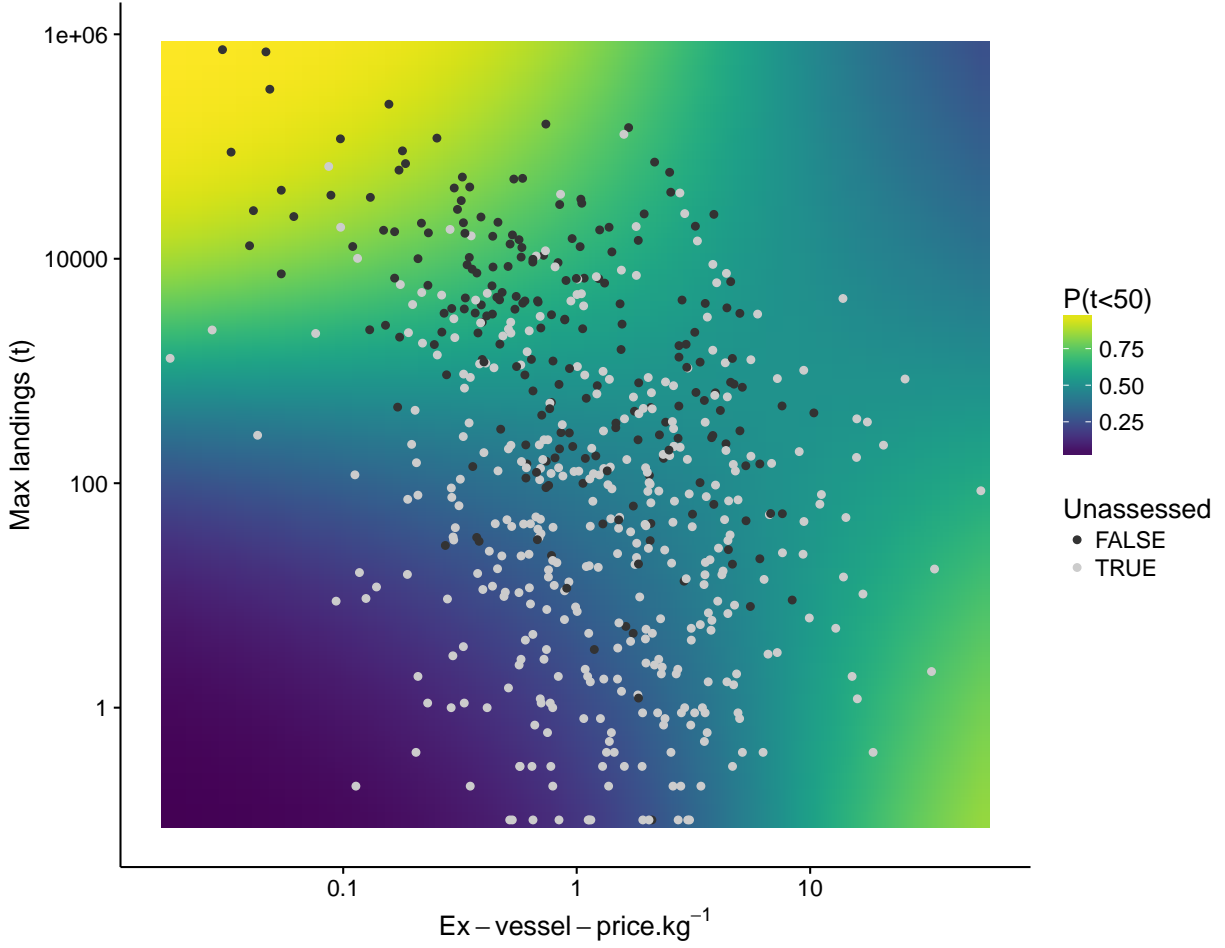


Figure D.5: Marginal probability of a stock being assessed after 50 years as a function of mean ex-vessel price (US\$.kg⁻¹) and maximum landings prior to assessment. Marginal probabilities were evaluated at the mean of remaining continuous covariates. The dataset used for analysis is overlayed with assessed stocks as dark grey points and unassessed stocks in light grey.

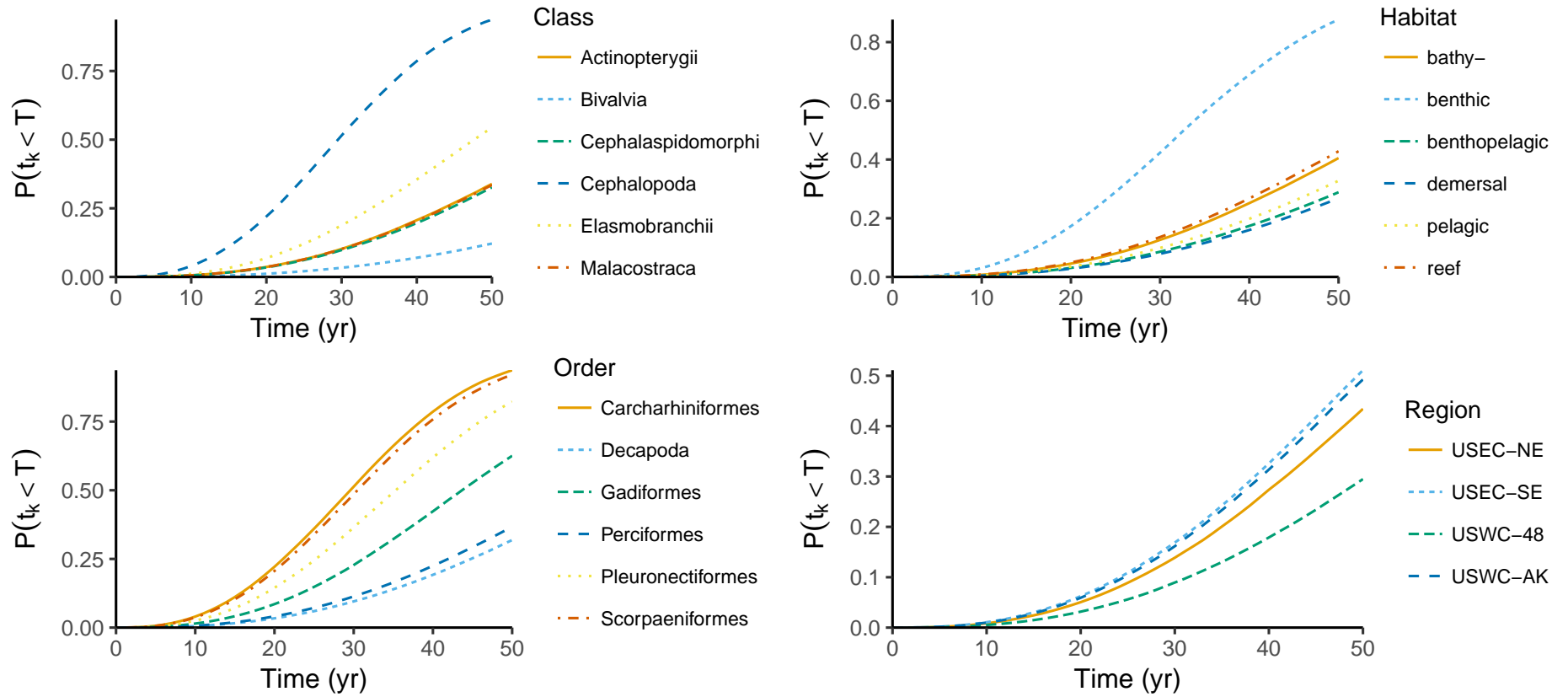


Figure D.6: Marginal probability of a stock in category k being assessed as a function of time ($P(T_k \leq t) = F_k(t) = \exp(-\lambda_k t^\tau)$), for stocks of various taxonomic orders, class, regions and habitats. For taxonomic variables, only the six levels with the most stocks represented in our dataset are shown. Marginal probabilities were evaluated at the mean of (centered) continuous covariates.

Table D.1: Posterior means of model parameters under interpretations of ratio of rates (θ) or time-to-assessment (ν), and probability $P(\theta > 1)$ that increasing parameter values or stocks in a given category have an increased likelihood of assessment compared to the baseline. Under the ratio of rates interpretation, the rate effect θ represents rates at which stocks with different characteristics are assessed relative to a baseline of 1. Under the time-to-assessment interpretation, the time effect ν is a multiplicative acceleration factor, i.e., $\nu = 0.5$ suggests a stock with these characteristics is assessed twice as fast as the average stock.

Effect	Category	Rate effect (θ)	Time effect (ν)	$P(\theta > 1)$
Region	Southeast	1.25	0.92	0.77
Region	Alaska	1.20	0.93	0.73
Region	Northeast	1.01	1.00	0.51
Region	West Coast	0.64	1.18	0.08
Habitat	Benthic	3.68	0.61	0.99
Habitat	Reef	1.03	0.99	0.53
Habitat	Bathy-	0.94	1.02	0.44
Habitat	Pelagic	0.74	1.12	0.25
Habitat	Benthopelagic	0.65	1.18	0.15
Habitat	Demersal	0.58	1.23	0.08
Class	Actinopterygii	0.76	1.11	0.35
Class	Bivalvia	0.25	1.71	0.08
Class	Branchiopoda	0.76	1.11	0.40
Class	Cephalaspidomorphi	0.81	1.08	0.43
Class	Cephalopoda	4.27	0.57	0.88
Class	Echinoidea	2.21	0.74	0.77
Class	Elasmobranchii	1.29	0.91	0.62
Class	Gastropoda	0.94	1.02	0.47
Class	Holocephali	0.42	1.40	0.26
Class	Holothuroidea	11.32	0.39	0.90
Class	Malacostraca	0.76	1.11	0.38
Class	Merostomata	0.48	1.32	0.26
Order	Acipenseriformes	0.77	1.11	0.40
Order	Anguilliformes	0.40	1.42	0.13
Order	Anostraca	0.85	1.07	0.44
Order	Arcoida	0.93	1.03	0.47
Order	Aspidochirotida	2.57	0.70	0.80
Order	Beloniformes	0.55	1.26	0.23
Order	Carcharhiniformes	3.56	0.62	0.97
Order	Chimaeriformes	0.68	1.16	0.35
Order	Clupeiformes	0.92	1.03	0.43
Order	Cypriniformes	1.02	0.99	0.50
Order	Cyprinodontiformes	0.83	1.07	0.42
Order	Decapoda	0.91	1.04	0.45
Order	Echinoida	1.55	0.85	0.67
Order	Elopiformes	0.42	1.40	0.13

Continued on next page

Table D.1 – continued from previous page

Effect	Category	Rate effect (θ)	Time effect (ν)	$P(\theta > 1)$
Order	Gadiformes	2.37	0.72	0.97
Order	Lamniformes	0.87	1.05	0.44
Order	Lampriformes	0.85	1.06	0.44
Order	Lophiiformes	1.64	0.83	0.74
Order	Mugiliformes	0.55	1.26	0.17
Order	Myliobatiformes	0.97	1.01	0.49
Order	Myoida	1.74	0.81	0.75
Order	Mytiloida	0.65	1.18	0.32
Order	Neogastropoda	0.95	1.02	0.48
Order	Ophidiiformes	0.83	1.07	0.43
Order	Orectolobiformes	0.98	1.01	0.49
Order	Osmeriformes	0.61	1.20	0.29
Order	Ostreoida	0.46	1.34	0.18
Order	Perciformes	1.08	0.97	0.57
Order	Petromyzontiformes	0.94	1.03	0.48
Order	Pleuronectiformes	4.05	0.59	1.00
Order	Pristiformes	0.95	1.02	0.47
Order	Rajiformes	0.90	1.04	0.45
Order	Scorpaeniformes	6.06	0.50	1.00
Order	Squaliformes	0.73	1.12	0.34
Order	Squatiniiformes	0.77	1.10	0.39
Order	Tetraodontiformes	0.90	1.04	0.45
Order	Teuthida	2.13	0.75	0.76
Order	Veneroida	0.86	1.06	0.42
Order	Xiphosura	0.70	1.14	0.36
Order	Zeiformes	0.90	1.04	0.46
Fixed effect	Mean ex-vessel price	5.07	0.54	1.00
Fixed effect	Maximum landings	4.14	0.58	1.00
Fixed effect	Maximum length	1.86	0.79	0.99
Fixed effect	Landings x Price	0.19	1.89	0.00