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

Philipp Neubauer, Dragonfly Data Science, Wellington, NZ
James T. Thorsen, NOAA Northwest Fisheries Science Center, Seattle
Michael C. Melnychuk, School of Aquatic and Fisheries Science,
University of Washington, Seattle

November 24, 2016

Target journals

1. Fish & Fisheries

2. ICES JMS

3. Journal of Applied Ecology (should be easy to write for this, but maybe have less of a fisheries audience)

4. Ecological Applications (ditto JAPPL)

Abstract

At a minimum, fisheries management requires estimating harvested quantities or fishing rates so that the fishery can sustainably derive value from the resource base. Beyond this, many management agencies also define objectives regarding population status (e.g., biomass relative to biological targets). The different mandates of estimating harvest rates versus population status are often achieved using different types of models. For example, agencies in the United States have a legislative mandate to "end overfishing" by setting harvest limits below biological targets, and have estimated these targets for many species using models that do not estimate population status. By contrast, estimating population status generally requires a "stock assessment" a statistical model of population dynamics to estimate population size relative to harvest. Stock assessments are often more costly and resource-intensive than models that only estimate harvest limits, and to date only a subset of landed species have stock assessments. Here we quantitatively explore the 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 600 coastal marine fish and invertebrate stocks, 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 a stock assessment. Landed tonnage is thus the dominant factor determining

the rate of new assessments. However, we also find that a high ex-vessel price of landed fish leads to earlier stock assessment. The overall rate at which new stocks are assessed has been increasing since the 1950s, and 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 slowly in future decades, as the landed tonnage and price for the remaining unassessed stocks makes it unlikely that current resources are sufficient to generate stock assessments for these species.

1 Introduction

It is often said, "what gets measured, gets managed". For over XXXX years (historical citation), fisheries scientists have measured human impacts on ocean populations of finfishes and marine invertebrates with the goal of balancing the value derived from fishing with the long-term sustainability of populations. In particular, stock assessments generally involve 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 target levels and population abundances above target levels (citation).

Recently, the National Marine Fisheries Service (NMFS, the agency in charge of science supporting fisheries management in the United States) has committed to "end overfishing" for all marine species within regional fisheries management plans (with exceptions granted in a few circumstances; citation). In the US, overfishing is defined as any stock having annual harvest above limit levels. A target (or limit) harvest can in theory be calculated by combining a target harvest rate with a target 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 (Berkson and Thorson 2015; Newman, Berkson, and Suatoni 2015). For example, depletion-corrected average catch (Alec D MacCall 2009) is used to estimate an annual fishing limit for many stocks, but is not used to estimate population abundance. DCAC and similar methods 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 (citations : BAUM). 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 size or age data); in these cases, the population model essentially treats the historical fishing as a "depletion experiment" to estimate the fraction of the population that is fished every year. In the following, we consider these population models "stock assessments", although we acknowledge that other authors have used the term

"stock assessment" more broadly to also include methods for estimating overfishing limits (e.g., DCAC). Although NMFS has estimated overfishing limits for the vast majority of fishes in US fisheries management plans, a smaller percentage of fished species have a stock assessment as defined. The dearth of assessed species presumably arises because developing a stock assessment is time-consuming, typically has greater data input needs, and requires extensive financial resources from NMFS and other interested parties (Geremont ref).

Stock assessments (population models for estimating absolute abundance of fished species and associated biological reference points) are important for many applied and theoretical questions regarding marine ecosystems. In particular, managing population stability (a characteristic of population size), rather than simply managing annual fishery removals, is possible only by estimating population abundance 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. Unassessed stocks may receive less attention from the public or fisheries managers when management changes are warranted.

2. Output from stock assessments has often been used in meta-analyses to understand ecological characteristics of marine fishes in general (Myers, Bowen, and Barrowman 1999; Thorson, Cope, et al. 2015). Therefore, any systematic bias in which 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 X through 2016; e.g., Helser et al. 2006), 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.

In this paper, we seek to provide a quantitative analysis of which marine species are likely to have undergone a stock assessment using a statistical population dynamics model. To do so, we combine two databases representing fished coastal marine species in the continental United States and Alaska: a database of landed tonnage and value by species from 195X to 201X, and a database of management and stock assessment attributes for US fishes and invertebrates with peer-reviewed stock assessments. For each landed stock in each of four US regions (northeast, southeast, northwest, and southwest), we record the year that it first had a stock assessment; we treat any stock that did not have an assessment by 2015 as a "censored"

observation (i.e., it might eventually have a peer-reviewed assessment). We 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 US regions?; (3) are there certain taxa (e.g., invertebrates or sharks) that are assessed substantially faster or slower-given these biological and economic attributes?; and (4) is the rate of stock assessment accelerating or decelerating over time? We show that ex-vessel price and landed tonnage are the main drivers of increasing rates of stock assessments, but some taxa defy these trends and are more likely to be assessed.

2   Methods

2.1   Year of first stock assessment

The year of first stock assessment for marine stocks previously assessed in the United States was determined primarily by interviews with regional stock assessment scientists, and supplemented by literature reviews of archived 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 populations, we defined a stock assessment as a single-species model of density-dependent population dynamics (e.g., including some combination of individual growth, recruitment, or aggregate surplus production) with the following criteria: (i) model parameters were estimated by fitting to abundance index and/or age or length compositional data; (ii) 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 (iii) management benchmarks corresponding to these time series estimates were estimated within the assessment or were otherwise explicitly stated. 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. We recognize that stock-reduction analyses (SRAs) are often used to estimate overfishing limits for stocks in the absence of a population-dynamics model fitted to data. 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.

Assignments of first assessment year were compared with the NOAA Species Information System (SIS) database to ensure consistency. The SIS database does not contain information about when a stock was first assessed, so comparisons were restricted to recent years, assuming that stocks which qualified as assessed in a previous year continue to qualify as presently assessed. These comparisons generally showed consistency among datasets, with categories of Levels of Stock Assessment Models in SIS aligning with our assignments. Of the nearly 200 stocks for which we assigned a year of first assessment, there were seven discrepancies with SIS classifications which resulted from violation of the abovementioned assumption. These stocks were previously assessed using populations models, but currently are assessed with less complex methods. For analyses, we maintain our assignments of when these stocks were first assessed.

2.2 Defining un-assessed stocks

Not all species landed in US fisheries are assessed using a population-dynamics model, and some are only assessed in limited sub-regions. Our year-of-first assessment database therefore contained few records of unassessed stocks. Failing to include unassessed stocks would not allow us to make general inference about the propensity of stocks to be assessed, so we complement this year-of-first assessment database with data from NOAA's landings database. The NOAA landings database provides annual landings by species and state, whereas areas of stock distribution as defined in stock assessments comprise a wide array of spatial designations including state, regional, and national levels. To align databases, we identified the US states included within the area of distribution for all stocks in our dataset. Species with landings in states outside of these areas of assessed stocks were considered as unassessed in those states. These assignments of unassessed species:state combinations were pooled into species:region combinations so that an unassessed stock represented a single species in a particular region that was not already represented by the area of distribution of an assessed stock. We distinguish four regions, defined as: Alaska (i.e., the Eastern Bering Sea, Gulf of Alaska, and Aleutian Islands); US West Coast (i.e., the federal 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). These regions do not include the portion of US federally-managed waters in the Caribbean or central Pacific, because the year-of-first assessment database has not been expanded

to these regions.

This definition of "unassessed stocks" is difficult to apply in the case of highly migratory species, which often span multiple regions or cross national jurisdiction. Further, the year of first assessment of highly migratory species that are assessed is likely influenced by other countries, unlike coastal stocks of the continental US and Alaska. We therefore excluded species that are typically assessed by Regional Fisheries Management Organisations, including tuna, billfish, and oceanic shark species. We also excluded salmon species from our analysis because our year-of-first assessment database has not been expanded to cover salmon species.

2.3   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 a categorical random effect. Habitat types from FishBase (ref) or SeaLifeBase (ref) were aggregated into six categories: deep sea (¿200m; bathy-pelagic or bathy-demersal); benthic; demersal; benthopelagic; pelagic; and reef-associated. Adjustments were made to the default global species-level classifications for some populations to better reflect local habitat usage. 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 thw first assessment as separate numerical predictors, drawn from the NOAA landings database. For unassessed stocks, the maximum annual landings throughout the time series and the ex-vessel price in the final year of the time series were used as values for these predictor variables.

2.4   Time-to-event model

To assess which factors drive the over-all rate of assessments and the time from first recorded landings to a full stock assessment in US stocks, we applied a time-to-event model. Also called termed survival models in the medical and ecological literature, 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. We defined time to event as the time between first recorded landings and a full stock assessment. The first stock assessment (as defined by our criteria above) occured in 1960, and we therefore used 1960 as the first possible assessment

eyar for stocks that were first landed prior to 1960. We thus assume, based on the first recorded assessment, that the technology (models, computers to fit models etc.) was not available prior to 1960 to conduct a full stock assessment.

The Weibull distribution is often used as a flexible model that has several desireabledesirable properties for this type of analysis, and one can easily check whether the Weibull model is appropriate for the data at hand (see Figure A.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 an decreasing rate. This allows us to directly estimate the change in assessment rates over time. Weibull

A further desirable property is that the model coefficient estimates 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). Thus, the rate of events may be modified by a particular variable. The model A coefficient can also be interpreted as the in terms of time-to-event parameters, which scale the time to the an event, scaled by an acceleration factor. For example, in a hyothetical 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 parmetersparameters 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. The connection between the event rate and the time-to-event interpretations can be is made explicit by writing the Weibull density as a function of the product of the probability of the assessment occurring at time t and the probability of the assessment not occurring prior to time t. Thus, $P(T = t)$ depends on the probability of an assessment not having occurred prior to time t, $(A(t) = 1 - P(T \leq t) = 1 - F(t))$, where $F(t) = \exp(-\lambda t^{\tau})$ is the Weibull distribution function, and the rate $r(t)$ with which assessments occur at time t.

$$T \sim Weibull(\tau, \lambda) \tag{1}$$

$$= A(t) \times r(t) \tag{2}$$

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

We modeled the scale λ of the Weibull distribution as a function of covariates and categorical variables:

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

where β is a row-vector of regression coefficients, and X_i is a vector of continuous covariates. Continuous covariates were taken as the logarithms of mean ex-vessel price per kg, maximum landings, their interaction (i.e., mean ex-vessel price per kg $\times$ maximum landings) and species maximum length, standardised for the analysis. Categorical variables α, γ, κ, ω, and ζ_f were region, habitat, class, order, and family, respectively, all treated as specific random effects. The model was estiamated 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 6000 samples from the posterior distribution for each parameter.

All random effects were given half-Cauchy priors 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 populations in the United States (excluding salmons and highly migratory species) increased steadily from the 1950s into the 1990s (Figure 1a). During this period, the number of landed populations in Alaska, West Coast, and Southeast regions approximately doubled, while the number of landed populations 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 the enactment of the Sustainable Fisheries Act in 1996, less than 30 populations in each of the four regions were assessed. As a proportion of all landed populations, 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 that are assessed ranges from 12% in the Southeast to 33% in Alaska. In terms of regional landings, the assessment of large stocks in each region between the 1970s and 2000s lead to rapidly increasing proportions of total landed tonnage being comprised of assessed populations. By 1996, >85% of landings in Alaska, Northeast, and Southeast regions were comprised of assessed stocks, and in the West Coast this proportion has increased rapidly from 45% in 1996 to >74% currently (Figure 1c).
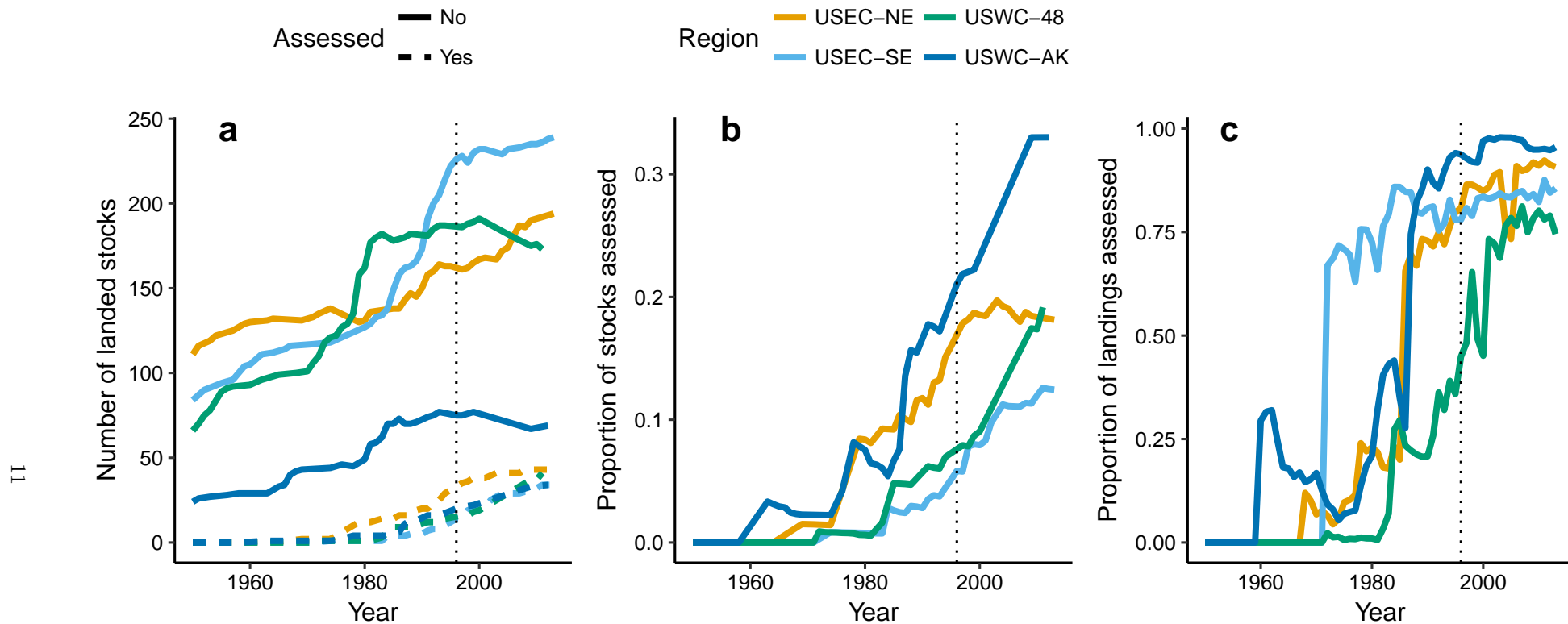
Figure 1: Timeline of a) the number of stocks landed by region and assessment status, b) proportion of landed stocks that are assessed, and c) the proportion of landed tonnage derived from assessed stocks. The dotted vertical line marks the enactment of the Sustainable Fisheries Act of 1996.

The majority of landed populations 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.
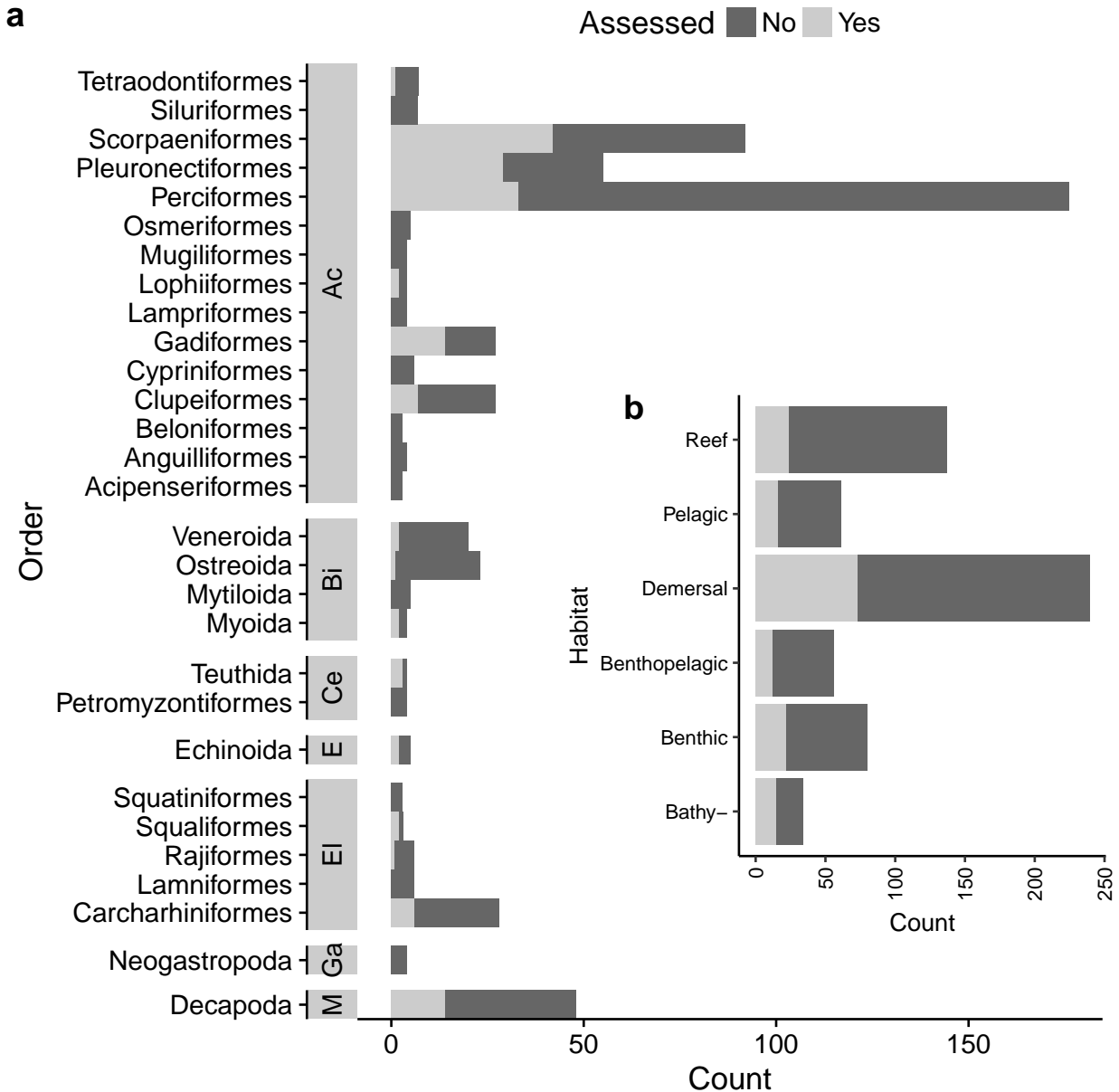

Figure 2: Assessment status at time of last known status (censoring time) a) by taxonomic order and sorted by class and b) by habitat type. In a), classes are abbreviated as Ac: Actinopterygii, Bi: Bivalvia, M: Malacostraca, E: Echinoidea, El: Elasmobranchii, Ce: Cephalaspidomorphi, Ga: Gastropoda, Ce: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

We now turn to our time-to-event model to interpret which biological and fishery characteristics explain differences in annual probability of first assessment among stocks (see Appendix Figures A.1 and A.2 for model diagnostics). Among the numerical covariates considered (Figure 3, Table A.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. 2c), but a smaller proportion of landed stocks being assessed (Fig. 2b). The interaction between price and landings was negative, sugesting that price is influencial when landings are small, and vise-versa, the landed tonnage drives assessments for species that only fetch a low price. The effect size (per standard deviation) of maximum body length was smaller than that for price and landings, but was nevertheless different from zero, suggesting that larger species have been preferencially assessed.

Among explanatory random effects, taxonomy factors (order and class) explain a larger portion of residual variance than either habitat or region factors (Figure A.3). This is reflected in the probability of prior assessment in any given year after first being landed (Figures 3 and A.4,, Table A.1), for which octopii and squids (Cephalopods) and sharks (Elasmobranchs) have a slightly higher probability of prior assessment than bony-fishes (Actinopterygii) or other taxonomic classes. Groundsharks (Carcharhiniformes), rockfishes (Scorpaeniformes), and flatfishes (Pleuronectiformes) have the highest probability of prior assessment among taxonomic orders, each having a higher assessment probability relative to the average of their taxonomic classes (Figure 4, Table A.1). Gadids (Gadidae) also have a relatively high assessment probability relative to the average for bony fishes. 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 (in particular demersal species), and assessment probabilities were greatest for stocks in Alaska.

Finally, we use our model to forecast annual assessment probabilities for all unassessed stocks in each region, and use this to calculate the expected proportion of assessed stocks through 2050 (Figure 5). These projections rely on the values of maximum landings and ex-vessel price for all un-assessed stocks, and were calcualted from the final year of available time series data used for model fitting (usually 2013). Alaska is predicted to have the highest proportion of assessed populations throughout the 25-year projection window. Notably, all regions show a relatively slow increase in the predicted proportion of assessed populations compared to the rapid increases in the observed proportions of assessed populations over the last 35 years
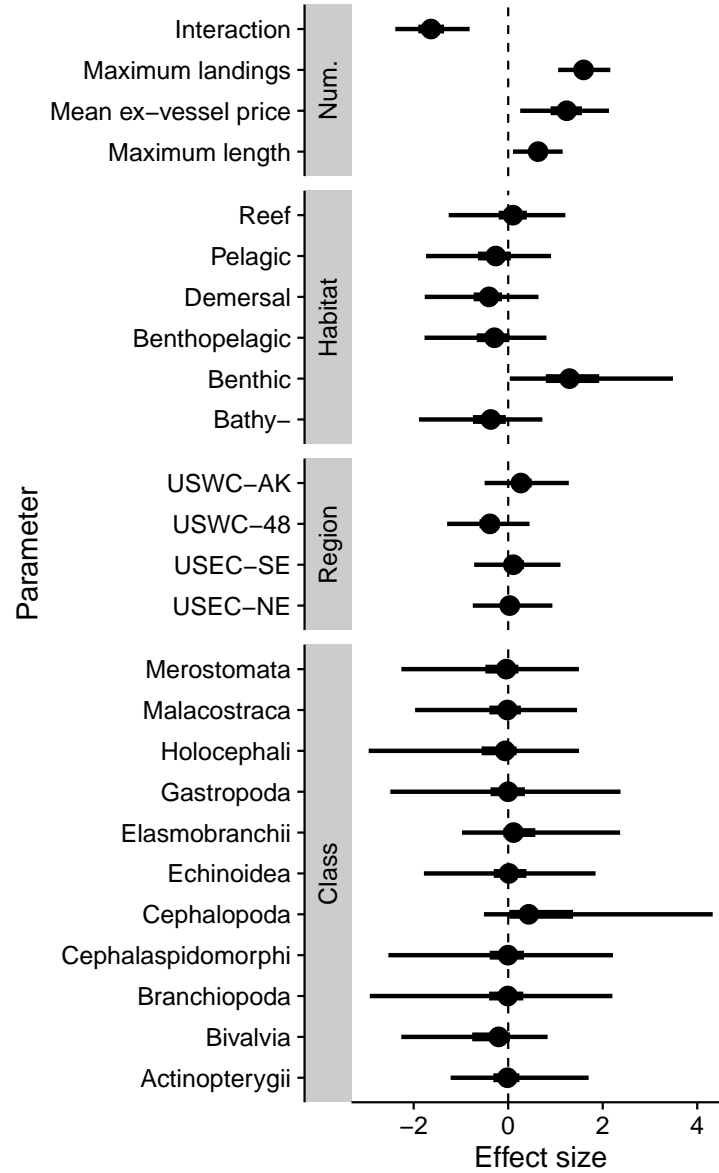
Figure 3: Summaries of estimated posterior distributions for numeric (Num.) covariates in the model, regional random effects, habitat random effects, and taxonomic class random effects. Circles show posterior medians, thick bars show inter-quartile ranges of the posteriors, and thin lines show 95% confidence intervals.

(Fig. 2b). This occurs because stocks with a high assessment probability have typically been assessed early, so that remaining stocks have low landings, prices, or other characteristics associated with low annual assessment probability.
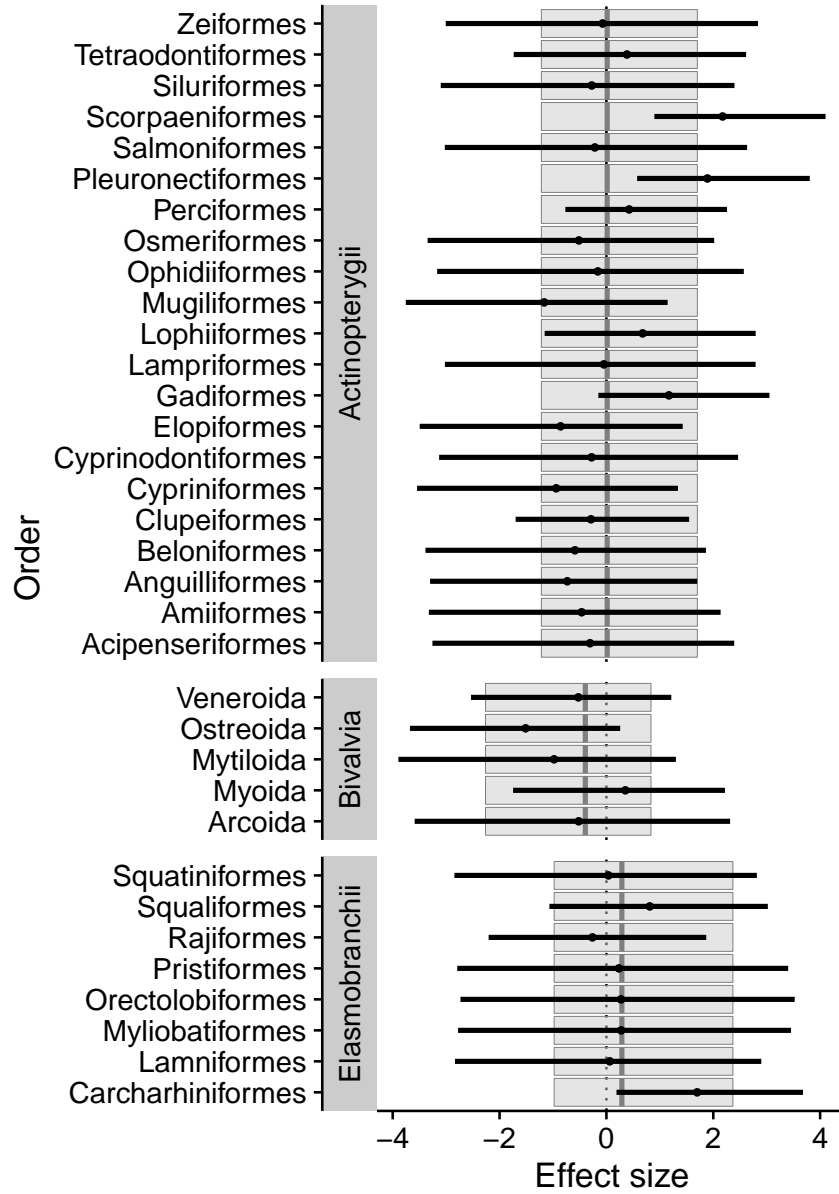
Figure 4: Summaries of estimated posterior distributions for random effects of orders within classes (Gas.: Gastropoda). For classes containing multiple nested orders in the dataset, grey lines show posterior means and coloured boxes show 95% confidence intervals of class effects. Order effects are shown as relative to the class effect within which they are nested, with points showing posterior means and black lines showing 95% confidence intervals.

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". The United States National Marine Fisheries Service (NMFS) currently estimates annual catch limits (ACLs) for the vast majority of fishes in
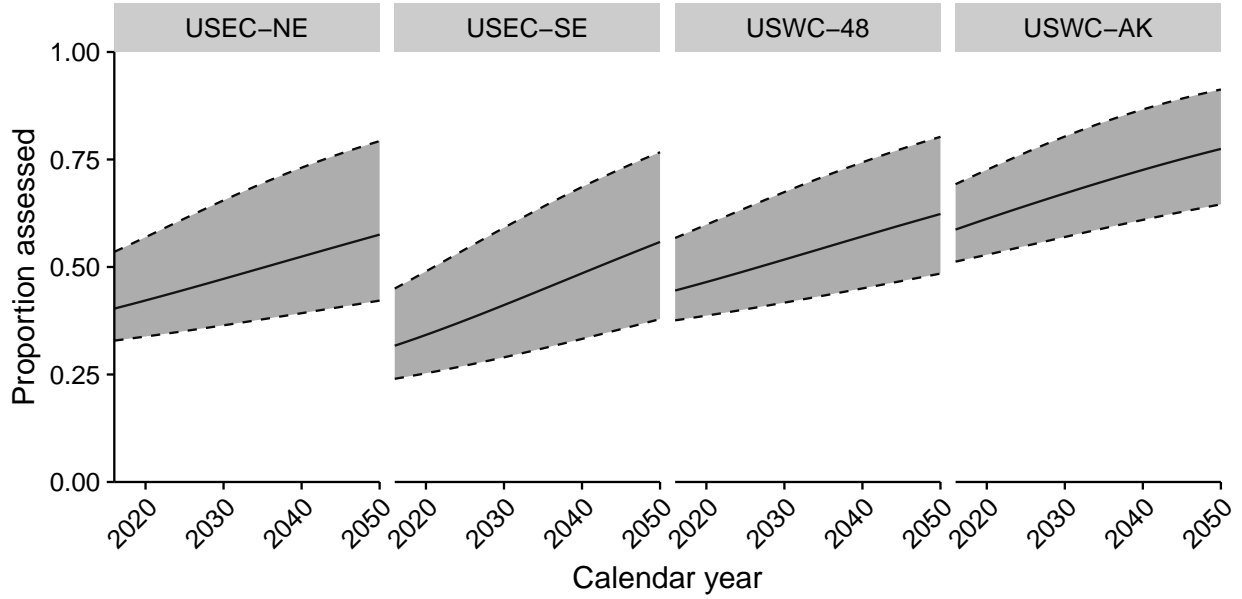
Figure 5: Projected proportion of stocks assessed by region and calendar year, based on assessment probabilities of stocks within each region over the projected year range.

federal fisheries management plans, and has established accountability measures that are triggered whenever recorded annual harvest exceeds the ACLs (Methot et al. 2014). Similarly, state-based management agencies also monitor catches for many species and implement management actions once catches or catch-per-unit effort levels reach pre-specified limits. 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. Stock-reduction analyses (SRAs) and other catch-only models (COMs), used to estimate ACLs for the majority of stocks in most federal US management regions (Berkson and Thorson 2015), do not estimate population size relative to management targets (Dick and A. D MacCall 2011; Wiedenmann, Wilberg, and Miller 2013). We have excluded SRAs and COMs from our definition of "formal stock assessments", and conclude that NMFS and other agencies are measuring population abundance only for those species that have a recent stock assessment.

In some cases, it is possible to rebuild or maintain fish and invertebrate stocks at levels of sustainable harvest without using a formal stock-assessment model, using only SRAs or COMs (e.g., Wetzel and Punt 2011a; Wetzel and Punt 2011b). 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 (Wetzel and Punt 2011b). However, we see two main

16

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 set to perform adequately on average: some stocks may be overfished, others may be underfished, but the average stock has appropriate fishing rates. In contrast, formal stock assessments are likely to perform adequately over time for all stocks: some decades may be overfished, others may be underfished, but on average each individual stock is fished sustainably over time. 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 (RAY Hilborn et al. 2012; Melnychuk et al. 2012).

2. The ability of formal stock assessments to inform harvest plans based on updated data has been repeatedly shown to improve management outcomes (e.g., Carruthers et al. 2014). 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, even relative to cases where a COM estimates sustainable fishing mortality rates perfectly (Thorson, Jensen, and Ray Hilborn 2015). Updating harvest plans based on new data can also prevent cases in which stock reduction analyses over-estimate a sustainable fishing rate, which would otherwise collapse the fishery (Wetzel and Punt 2011b). 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 formal stock assessments.

There are many differences in quality and complexity among formal stock assessments. NMFS categorizes assessments using five "tiers", and high-tier 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 SRAs. In general, our classification of not assessed aligned with the SIS categories of 0 or 1 (Table ??). We recommend future research that expands our model to include annual probabilities of transitioning among multiple categories (e.g., among six categories including unassessed and all five NMFS assessment tiers). 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 consisered.

Given our operational definition of stock assessment, maximum landings was a particularly strong predictor of the year in which stocks were first assessed, and ex-vessel price was also positively correlated with the rate of assessments. The product of landings and price is a rough measure of the gross economic value 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. If this choice reflects in part the landed value of stocks, then these results support previous findings (Sethi, Branch, and Watson 2010) suggesting that fishery development is also driven primarily by landed tonnage and ex-vessel prices of fished species.

Certain taxonomic classes, or orders within classes, stood out as being more likely to have undergone a formal stock assessment after controlling for landings, ex-vessel price, and other factors. Elasmobranchs, and in particular groundsharks (Carchariniformes), had relatively high rates of stock assessment. This likely results from greatly increasing conservation interest in recent decades for shark species both in the United States and worldwide (Dulvy et al. 2014). 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 (Corts 2002). That is, the assessment rate may be overestimated for groundsharks because landings underestimate the total catch. Among bony fishes, flatfishes (Pleuronectiformes) and scorpaenfishes such as rockfishes and greenlings (Scorpaeniformes) had high rates of assessment. Results for Scorpaeniformes seem reasonable to us, given the number of Pacific rockfishes included, which have been a topic of conservation concern in Alaska and the US West Coast (Clark 1993; Conway and Shaw 2008; Myers, Bowen, and Barrowman 1999). While cephalopod abundance is commonly estimated using catch-per-unit effort indices or survey abundance indices (Doubleday et al. 2016) rather than formal stock assessments, in the US most landed cephalopods are assessed (all are squid species). This may result from defined units of assesment having coastwide 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 (Thorson, Cope, et al. 2015). 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 (Rosenbaum and Rubin 1983; Melnychuk et al. 2012). For example, pairwise comparisons (or matching) between assessed and un-assessed 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 United States 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. However, this pattern is likely not characteristic of most countries, in part because of political mandates in the United States for quantifying the status of all fished stocks (CITATION). Most countries currently have a lower proportion of assessed stocks, and are likely still within a period of rapid increase in the proportion of assessed stocks. It is not necessarily the case that formal stock assessments are required for effectively managing fish and invertebrate stocks, as harvest control rules or in-season adjustments to fishing effort can instead 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 formal stock assessments.

5   Acknowledgements

References

Berkson, Jim and James T. Thorson (2015). "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, pp. 237–242.

Carruthers, Thomas R. et al. (2014). "Evaluating methods for setting catch limits in data-limited fisheries." In: *Fisheries Research* 153, pp. 48–68.

Clark, W. G (1993). "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, pp. 233–246.

Conway, Flaxen and Wesley Shaw (2008). "Socioeconomic Lessons Learned from the Response to the Federally-Declared West Coast Groundfish Disaster." In: *Fisheries* 33.6, pp. 269–277.

Corts, Enric (2002). *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, p. 133. URL: http://www.nmfs.noaa.gov/sfa/hms/species/sharks/documents/scs_assessment_rev2.pdf.

Dick, E. J. and A. D MacCall (2011). "Depletion-Based Stock Reduction Analysis: A catch-based method for determining sustainable yields for data-poor fish stocks." In: *Fisheries Research* 110.2, pp. 331–341.

Doubleday, Zo A. et al. (2016). "Global proliferation of cephalopods." In: *Current Biology* 26.10, R406–R407. ISSN: 0960-9822. DOI: 10.1016/j.cub.2016.04.002. URL: http://dx.doi.org/10.1016/j.cub.2016.04.002.

Dulvy, Nicholas K et al. (2014). "Extinction risk and conservation of the worlds sharks and rays." In: *eLife* 3. Ed. by Ian T Baldwin, e00590. ISSN: 2050-084X. DOI: 10.7554/eLife.00590. URL: https://dx.doi.org/10.7554/eLife.00590.

Helser, Thomas E et al. (2006). "Stock assessment of Pacific hake (whiting) in US and Canadian waters in 2006." In: *Pacific Fishery Management Council* 7700, pp. 97220–1384.

Hilborn, RAY et al. (2012). "Defining Trade-Offs among Conservation, Profitability, and Food Security in the California Current Bottom-Trawl Fishery." In: *Conservation Biology* 26.2, pp. 257–268. ISSN: 1523-1739.

MacCall, Alec D (2009). "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, pp. 2267–2271.

Melnychuk, Michael C et al. (2012). "Can catch share fisheries better track management targets?" In: *Fish and Fisheries* 13.3, pp. 267–290.

Methot, R. D. et al. (2014). "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, pp. 183–194.

Myers, R. A., K. G. Bowen, and N. J. Barrowman (1999). "Maximum reproductive rate of fish at low population sizes." In: *Canadian Journal of Fisheries and Aquatic Sciences* 56.12, pp. 2404–2419.

Newman, David, Jim Berkson, and Lisa Suatoni (2015). "Current methods for setting catch limits for data-limited fish stocks in the United States." In: *Fisheries Research* 164, pp. 86–93.

Rosenbaum, PAUL R. and DONALD B. Rubin (1983). "The central role of the propensity score in observational studies for causal effects." In: *Biometrika* 70.1, 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.

Sethi, S. A., T. A. Branch, and R. Watson (2010). "Global fishery development patterns are driven by profit but not trophic level." In: *Proceedings of the National Academy of Sciences* 107.27, pp. 12163–12167.

Thorson, James T., Jason M. Cope, et al. (2015). "Giants' shoulders 15 years later: lessons, challenges and guidelines in fisheries meta-analysis." In: *Fish and Fisheries* 16.2, pp. 342–361.

Thorson, James T., Olaf P. Jensen, and Ray Hilborn (2015). "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, pp. 428–435.

Wetzel, Chantell R. and Andr E. Punt (2011a). "Model performance for the determination of appropriate harvest levels in the case of data-poor stocks." In: *Fisheries Research* 110.2, pp. 342–355.

— (2011b). "Performance of a fisheries catch-at-age model (Stock Synthesis) in data-limited situations." In: *Marine and Freshwater Research* 62.8, pp. 927–936.

Wiedenmann, John, Michael J. Wilberg, and Thomas J. Miller (2013). "An Evaluation of Harvest Control Rules for Data-Poor Fisheries." In: *North American Journal of Fisheries Management* 33.4, pp. 845–860.
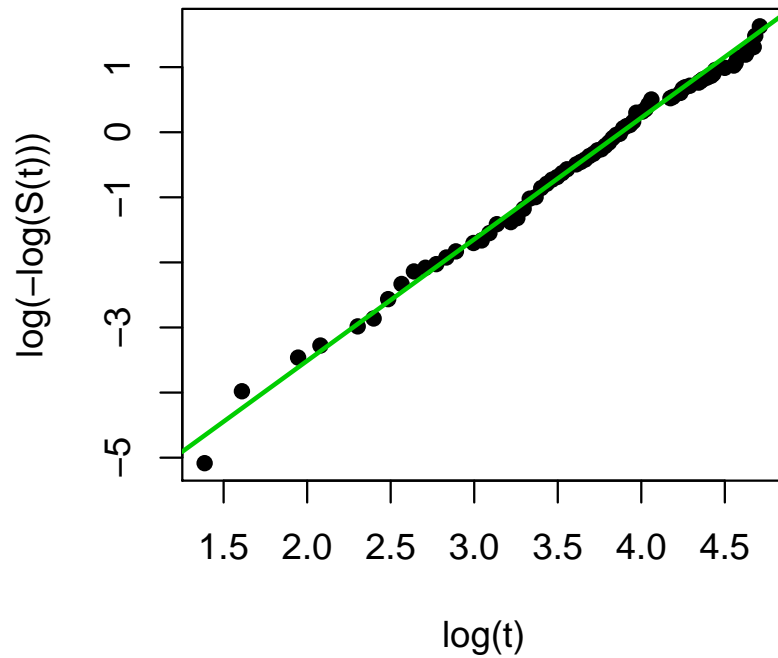
A A: Model fit

Figure A.1: Appropriateness of the Weibull event-time model for the time-to-assessment dataset. If the Weibull applies, the beginning of the landings dataset in 1950 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-Meyer estimate of survival at time t, and the log of t. Here, τ evaluates to 1.87 (slope of the green line), suggesting an increasing assessment rate with increasing time t.
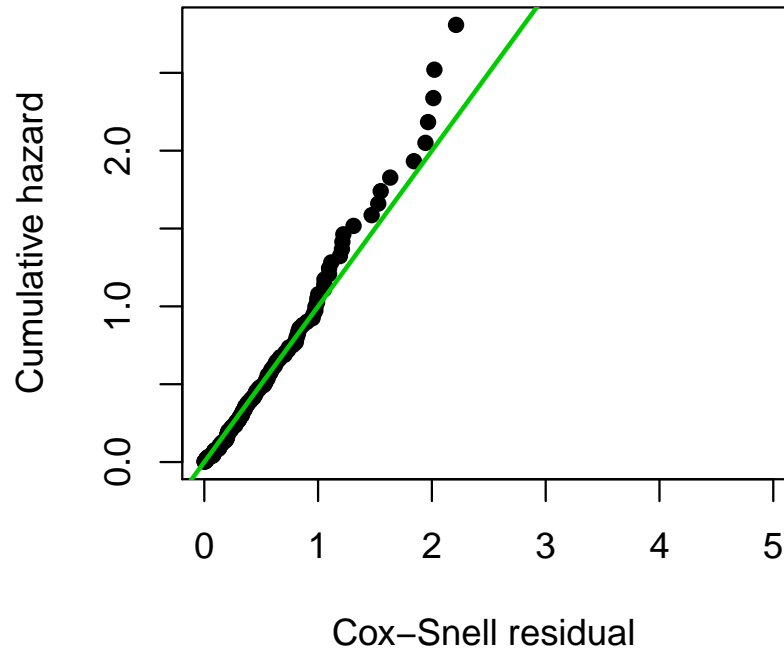
Figure A.2: Model fit of the Weibull survival model, based on Cox-Snell residuals calculated at the posterior mean of the linear predictor. For a perfect fit all data points (solid points) would lie on the y=x (green) line.
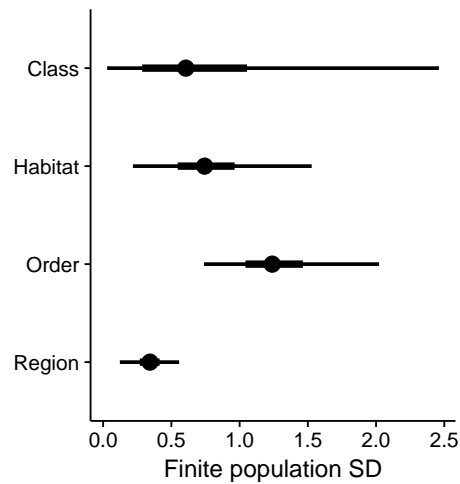

Figure A.3: Comparison of finite population standard deviation (i.e., variance attributed to each variable) for random effects in the Weibull survival model. Circles show posterior medians, thick bars show inter-quartile ranges of the posteriors, and thin lines show 95% confidence intervals.

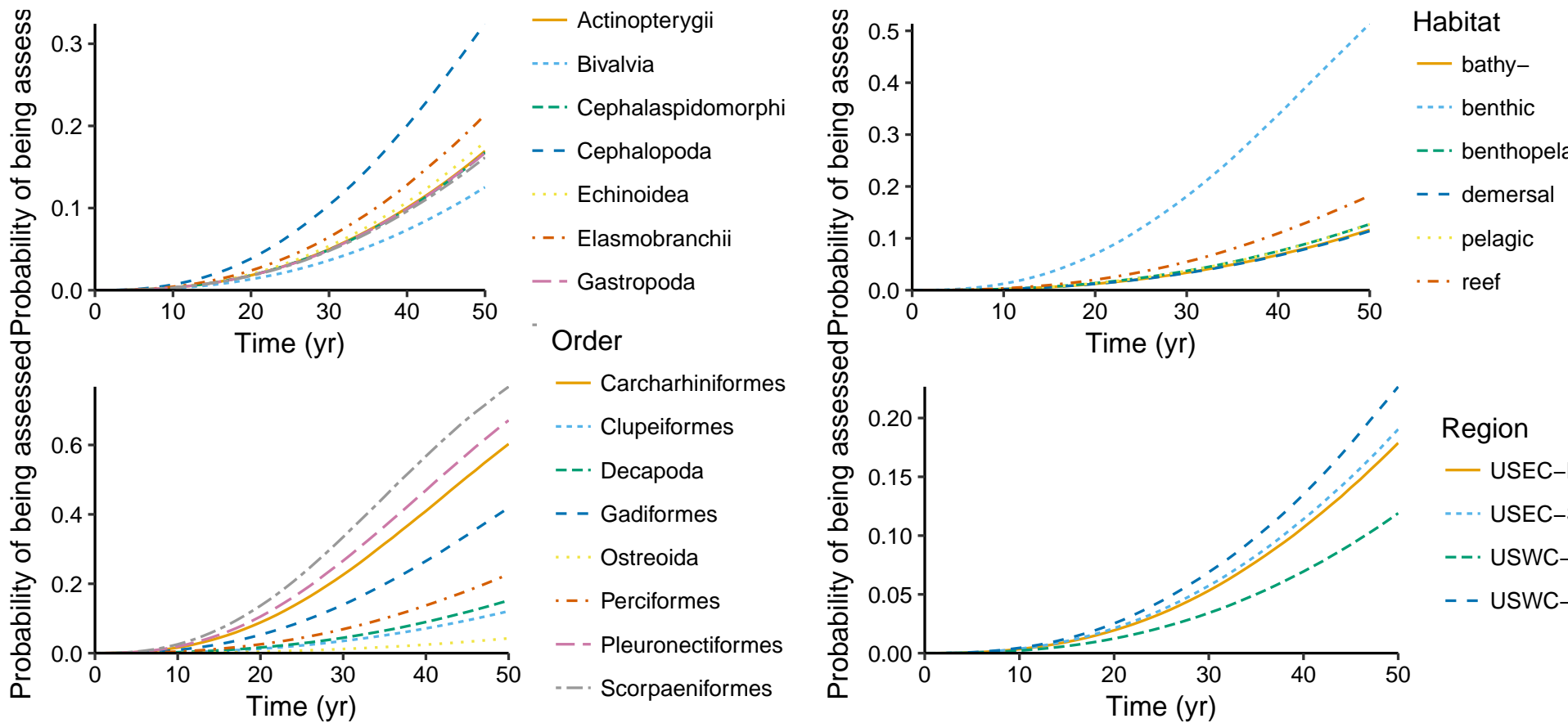Figure A.4: Marginal probability of a stock in category k being assessed as a function of time $(P(T_k \leq t) = F_k(t) = \exp(-\lambda_k t^\tau))$, for stocks of various taxonomic orders, class, regions and habitats. For taxonomic variables, only the eight levels with the most stocks represented in our dataset are shown. Marginal probabilities were evaluated at the mean of (centered) continuous covariates.

Table A.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 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	USEC-NE	1.03	0.99	0.55
Region	USEC-SE	1.12	0.96	0.64
Region	USWC-48	0.68	1.17	0.10
Region	USWC-AK	1.31	0.90	0.83
Habitat	Bathy-	0.69	1.16	0.22
Habitat	Benthic	3.66	0.60	0.98
Habitat	Benthopelagic	0.75	1.12	0.27
Habitat	Demersal	0.67	1.18	0.16
Habitat	Pelagic	0.77	1.11	0.29
Habitat	Reef	1.11	0.96	0.59
Class	Actinopterygii	0.98	1.01	0.47
Class	Bivalvia	0.82	1.08	0.30
Class	Branchiopoda	0.99	1.00	0.49
Class	Cephalaspidomorphi	0.99	1.00	0.49
Class	Cephalopoda	1.54	0.84	0.78
Class	Echinoidea	1.02	0.99	0.52
Class	Elasmobranchii	1.12	0.96	0.64
Class	Gastropoda	1.00	1.00	0.49
Class	Holocephali	0.94	1.03	0.42
Class	Malacostraca	0.98	1.01	0.47
Class	Merostomata	0.96	1.02	0.44
Order	Acipenseriformes	0.77	1.11	0.41
Order	Amiiformes	0.66	1.18	0.35
Order	Anguilliformes	0.52	1.30	0.24
Order	Anostraca	0.91	1.04	0.47
Order	Arcoida	0.92	1.03	0.47
Order	Beloniformes	0.60	1.22	0.31
Order	Carcharhiniformes	3.98	0.58	0.97
Order	Chimaeriformes	0.68	1.17	0.36
Order	Clupeiformes	0.71	1.14	0.29
Order	Cypriniformes	0.42	1.41	0.17
Order	Cyprinodontiformes	0.81	1.09	0.42
Order	Decapoda	0.92	1.03	0.45
Order	Echinoida	1.26	0.91	0.60
Order	Elopiformes	0.46	1.36	0.20
Order	Gadiformes	3.06	0.64	0.99
Order	Lamniformes	0.84	1.07	0.44
Order	Lampriformes	0.98	1.01	0.50
Order	Lophiiformes	1.93	0.77	0.80
Order	Mugiliformes	0.34	1.54	0.11
Order	Myliobatiformes	0.99	1.00	0.50
Order	Myoida	2.10	0.74	0.80
Order	Mytiloida	0.60	1.23	0.32
Order	Neogastropoda	0.96	1.02	0.49

Effect	Category	Rate effect (θ)	Time effect (ν)	$P(\theta > 1)$
Order	Ophidiiformes	0.88	1.05	0.46
Order	Orectolobiformes	0.99	1.00	0.50
Order	Osmeriformes	0.64	1.19	0.34
Order	Ostreoida	0.34	1.54	0.09
Order	Perciformes	1.43	0.87	0.80
Order	Petromyzontiformes	0.96	1.01	0.49
Order	Pleuronectiformes	6.25	0.48	1.00
Order	Pristiformes	0.95	1.02	0.49
Order	Rajiformes	0.59	1.24	0.26
Order	Salmoniformes	0.85	1.07	0.44
Order	Scorpaeniformes	8.32	0.43	1.00
Order	Siluriformes	0.79	1.10	0.42
Order	Squaliformes	1.65	0.82	0.72
Order	Squatiniformes	0.82	1.08	0.43
Order	Tetraodontiformes	1.47	0.86	0.66
Order	Teuthida	6.29	0.48	0.94
Order	Veneroida	0.87	1.05	0.43
Order	Xiphosura	0.78	1.10	0.41
Order	Zeiformes	0.95	1.02	0.48
Num.	Maximum length	1.88	0.78	0.99
Num.	Mean ex-vessel price	3.45	0.61	0.99
Num.	Maximum landings	4.93	0.53	1.00
Num.	Interaction	0.20	1.91	0.00