

Are assessed stocks a representative sample of fished species in the US?

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 3, 2016

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 (finfishes and marine invertebrates) with the goal of balancing both the long-term sustainability of human impacts with maximizing harvest (or value) from fishing. 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. The value of these two metrics can be plotted on a 2-dimensional "stock status" plot, and fisheries agencies are increasingly committed to maintaining fished populations in the "healthy" quadrant (i.e., fishing rate below a target level, and population abundance above a target level; 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 calculate either harvest rate or population abundance (Berkson and Thorson 2015, Newman Berkson Sautoni 2015). For example, depletion-corrected average catch (DCAC; MacCall 2009) is used to calculate an annual fishing limit for many stocks, but is not used to calculate population abundance. DCAC (and similar methods) can therefore be used to help "end overfishing" but is not otherwise informative about characteristics of the fished population.

However, conservationists and ecologists will often be more interested in estimating population abundance (or abundance relative to equilibrium conditions) than estimating an overfishing limit (citations). 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/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 call 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 much smaller percentage of fished species have a stock assessment (in our sense). This relative dearth of stock assessment models presumably arises because developing a stock assessment is time-consuming and requires extensive financial resources from NMFS and other interested parties (Geremont ref).

We argue that stock assessments (population models for estimating absolute

abundance for fished species) are important for many applied and theoretical questions regarding marine ecosystems. 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 (among others):

1. Unassessed stocks may receive less attention from the public or fisheries managers regarding management actions. Therefore, if a given taxon is less likely to be assessed, it might also be less likely to receive rapid or public attention when management changes are warranted.

2. Output from stock assessments has often been used in meta-analysis models to understand ecological characteristics of marine fishes in general (Myers ref, Thorson et al. Fish and Fisheries 2015, other ideas...). Therefore, any systematic bias in which stocks are assessed could also bias our ecological understanding of marine fishes.

3. Stock assessments often require updates periodically (e.g., Pacific hake has been re-assessed annually from X through 2016), and agency resources might be fully expended while assessing a small fraction of stocks. However, it is currently unknown whether the rate of assessing new species for the first time is increasing or decreasing. If this rate is decreasing, this could indicate the need for additional public resources for stock assessment.

We therefore seek to provide the first-ever quantitative analysis of which marine species are likely to have undergone a stock assessment. To do so, we combine two databases representing all fished marine species in the United States: a database of all landed species from 195X to 201X, and a database of management and stock assessment attributes for all US fishes 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 does not have an assessment by 2015 as a "censored" observation (i.e., it might eventually have a peer-reviewed assessment) We then apply a censored time-to-event model to answer the following questions: (1) what economic or biological characteristics are associated with a high or low annual probability of being assessed for the first time? (2) which US region has assessed stocks relatively quickly or slowly? (3) are there certain taxa (e.g., invertebrates or sharks) that are substantially faster or slower-assessed given these biological and economic attributes? and (4) is the rate of stock assessment accelerating or decelerating over time. [Add one sentence high-line conclusion – I know some people don't do this, but I think its a reasonable way to hook interest]

2 Methods

2.1 Management database

The year of first stock assessment was determined primarily by interviews with regional stock assessment scientists, or occasionally by searching through archived assessments. In reality numerous types of stock assessments exist, 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 required a stock assessment in some year to be based on a single-species population model which provided time series estimates of some form of abundance (e.g. total biomass, spawning biomass) and/or fishing mortality (or exploitation rate) as well as some type of benchmark for these time series estimates. Benchmarks could include 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 population models. Exceptions were occasionally made (i.e. definitions of an assessment were more forgiving) for some invertebrate species like squids due to short lifespans and difficulties of aging individuals.

Several variables were considered as explanatory factors affecting the year in which a stock was first assessed. Regional differences across the United States may arise. Populations were classified into four regions, which were treated as categorical random effects: Alaska; West Coast; Northeast Coast (including the mid-Atlantic Coast); and Southeast Coast (including the South Atlantic Coast and Gulf of Mexico). The habitat typically occupied by the population was also considered as a categorical random effect. Habitat types from FishBase (ref) or SeaLifeBase (ref) were aggregated into four (five?) categories (benthic; demersal; benthopelagic(?); pelagic; and reef-associated), and 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. Fisheries that developed earlier may be more likely to be assessed earlier than others. The year of fishery development, defined as the year in which landings first reached 25 percent of the maximum landings across the full time series (Sethi et al. 2011? PNAS), provides historical information about the fishery for each population as was used as a numerical predictor. 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 ex-vessel price in the year of first assessment as separate numerical predictors, drawn from the National Marine Fisheries Service landings database. These state-level landings and ex-vessel price data were linked to corresponding biological stock units according to areas of stock distribution

as defined in assessments.

Year of first assessment assignments 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. The only exceptions... [squids, couple others].

2.2 Landings data

2.3 Data preparation

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 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.

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 1). The rate parameter of the Weibull can be interpreted in terms of the rate of events occurring. A slope greater than one suggests an increasing rate of events, whereas a slope of less than one indicates a decreasing rate. This allows us to directly estimate the change in assessment rates over time.

A further desirable property is the model can be interpreted in terms of the ratio of event rates as well as time-to-event. 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 can also be interpreted in terms of time-to-event parameters, which scale the time to the event by an acceleration factor. For example, 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 interpretation are therefore 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 made explicit by writing the Weibull density as a function of the product of the probability of an event occurring at time t and the probability of an event not occurring prior to time t . Thus, $P(T = t)$ depends on the probability of that event not having occurred prior to time t and the

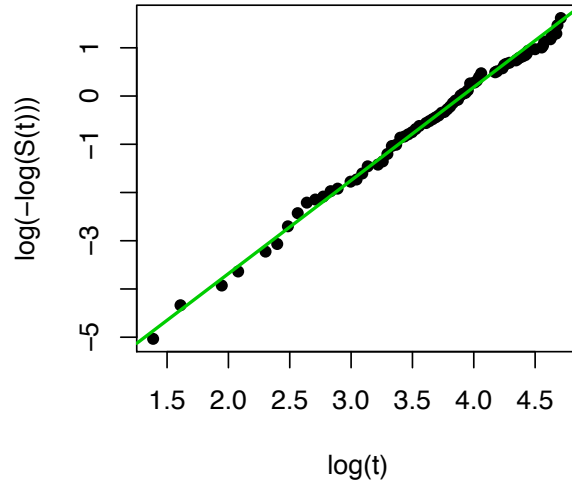


Figure 1: Appropriateness of the Weibull event-time model for the time-to-assessment dataset. If the Weibull applies, the time from fishery development to assessment should fall on a line with slope p (the Weibull shape parameter) between $\log(-\log(\hat{S}(t)))$, where $\hat{S}(t)$ is the non-parametric Kaplan-Meier estimate of survival at time t , and the log of t . Here, p evaluates to 1.93, suggesting an increasing assessment rate with increasing time t .

probability of the event occurring at time t . Thus,

$$P(T = t) = Weibull(\tau, \lambda) \tag{1}$$

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

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

3 Results

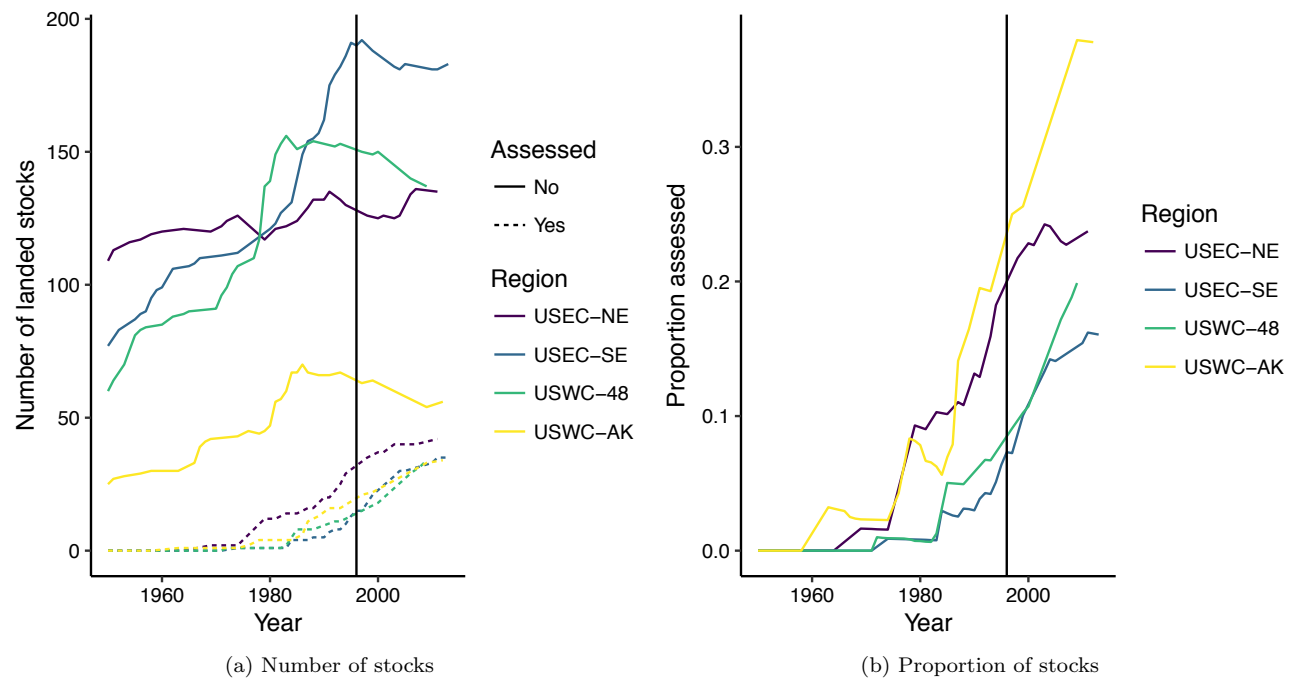


Figure 2: Timeline of a) the number of stocks landed by region and assessment status and b) proportion of landed stocks that are assessed. The vertical line marks the enactment of the Sustainable Fisheries Act of 1996

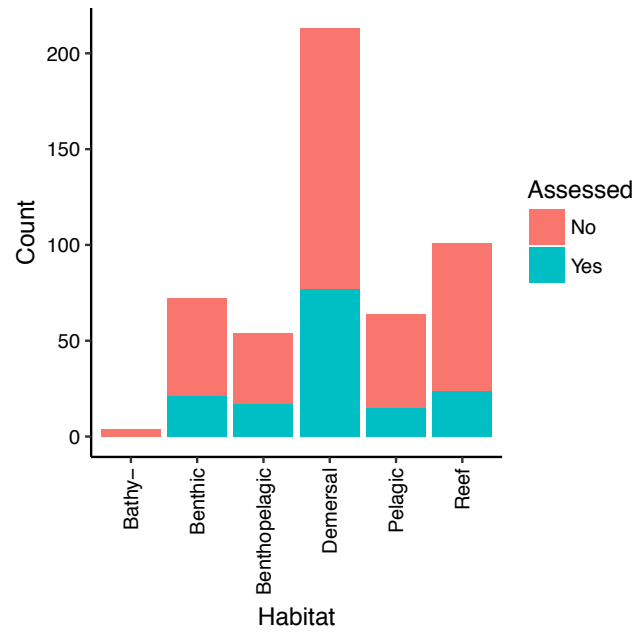


Figure 3: Assessment status at time of last known status (censoring time) by habitat

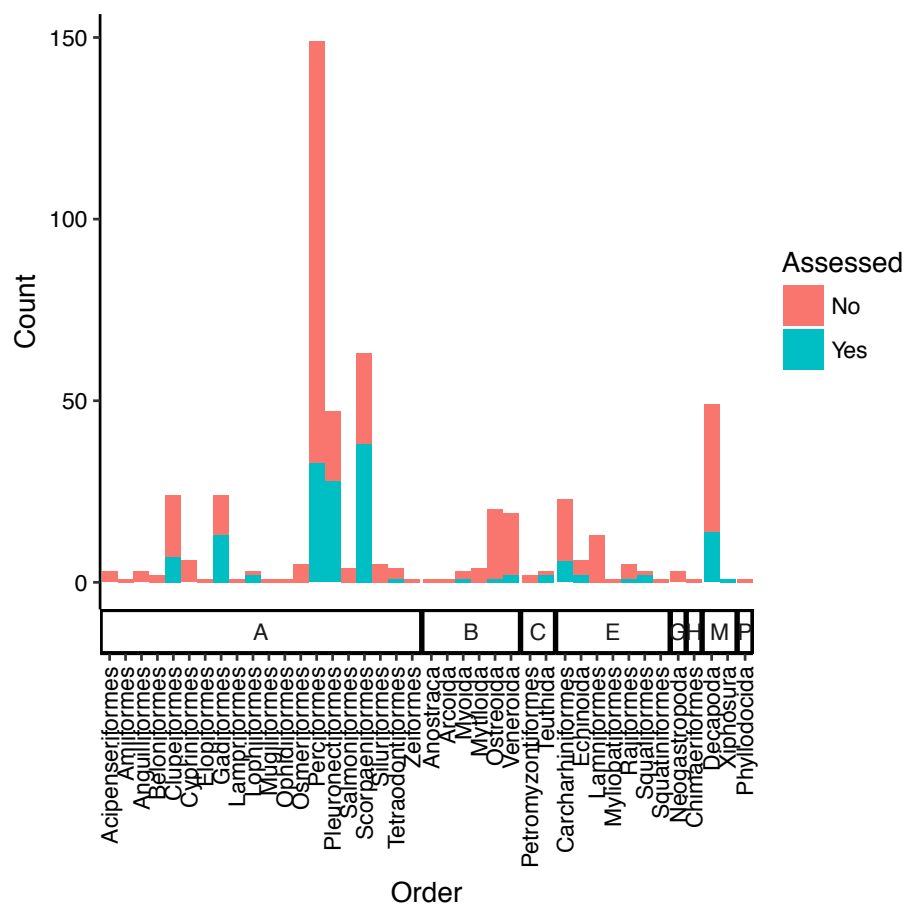


Figure 4: Assessment status at time of last known status (censoring time) by Order and sorted by Class. Classes are abbreviated as A: Actinopterygii, B: Bivalvia, M: Malacostraca, E: Echinoidea, P: Merostomata, G: Elasmobranchii, C: Polychaeta, H: Gastropoda, A: Cephalopoda, B: Cephalaspidomorphi, M: Holocephali, E: Branchiopoda

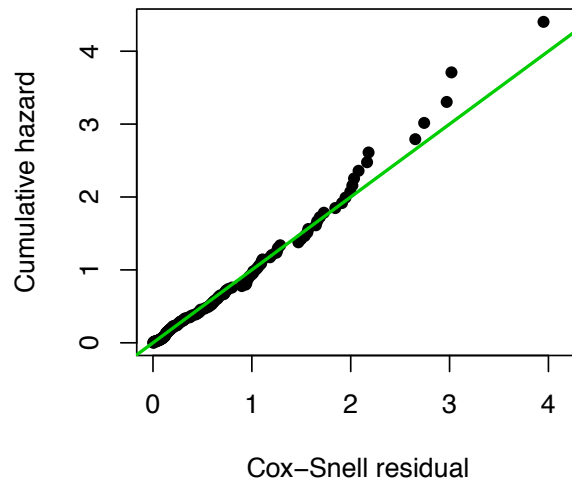


Figure 5: Model fit of the Weibull survival model, based on Cox-Snell residuals calculated at the posterior mean of the linear predictor. For a perfect fit all data points (solid points) would lie on the $y=x$ line.

Table 1: Posterior mean and $P(\beta > 1)$ for model parameters. Parameters can be interpreted as the ration of rates (rate effect, i.e., rates at which stocks with different characteristics are assessed) or as multiplicative acceleration factors (time effect, i.e., $\alpha=0.5$ suggests a stock with these characteristics is assessed twice as fast as the average stock).

Parameter	Rate effect (β)	Time effect ($\alpha = \beta\tau^{-1}$)	$P(\beta > 1)$
Length	1.06	1.00	0.52
Maximum landings	1.78	0.79	0.99
Mean price per kg	5.39	0.49	1.00

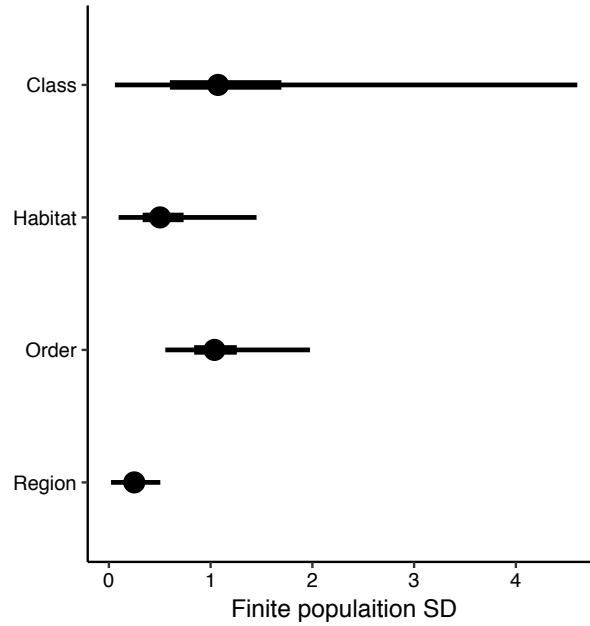


Figure 6: Comparison of finite population standard deviation for random effects in the Weibull survival model. The circle shows the posterior median, with thick bars showing the inter-quartile range of the posterior and the thin line is the 95% confidence interval

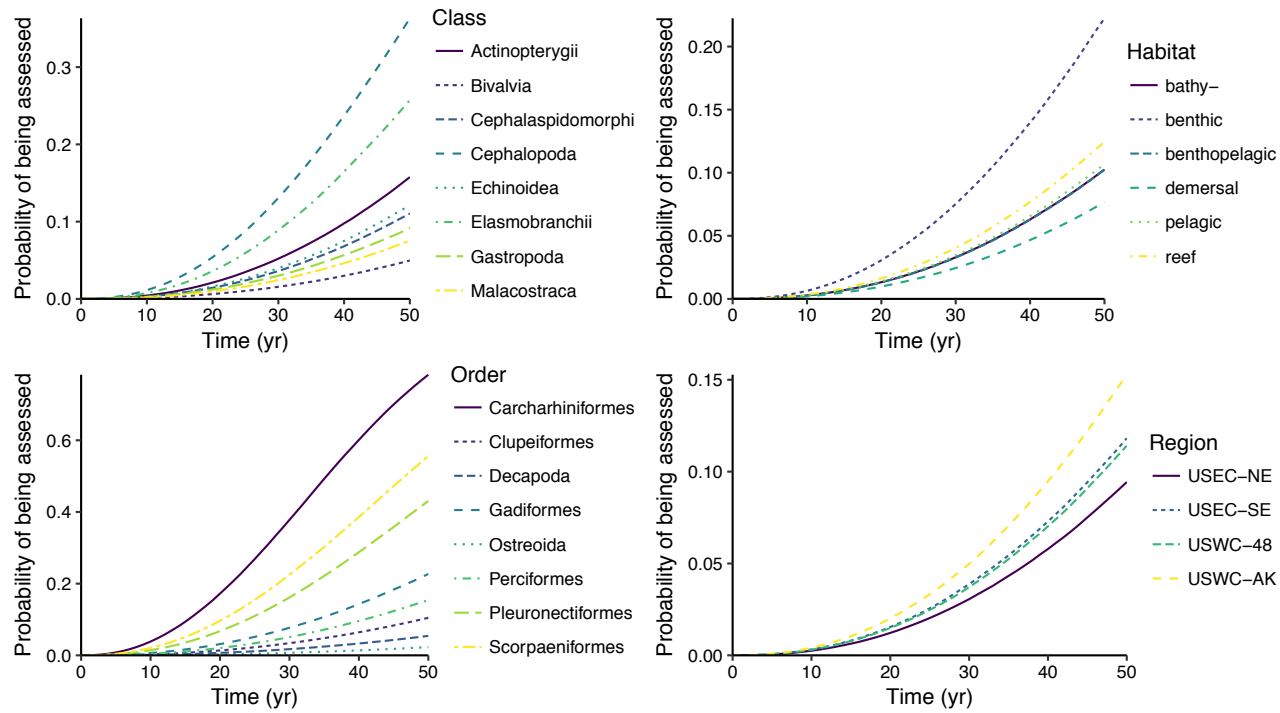


Figure 7: Marginal probability of a stock being assessed as a function of time, for stocks of various taxonomic orders, class, regions and habitats. For taxonomic variables, only the eight levels with the most stocks represented in our dataset are shown. Marginal probabilities were evaluated at the mean of continuous covariates.

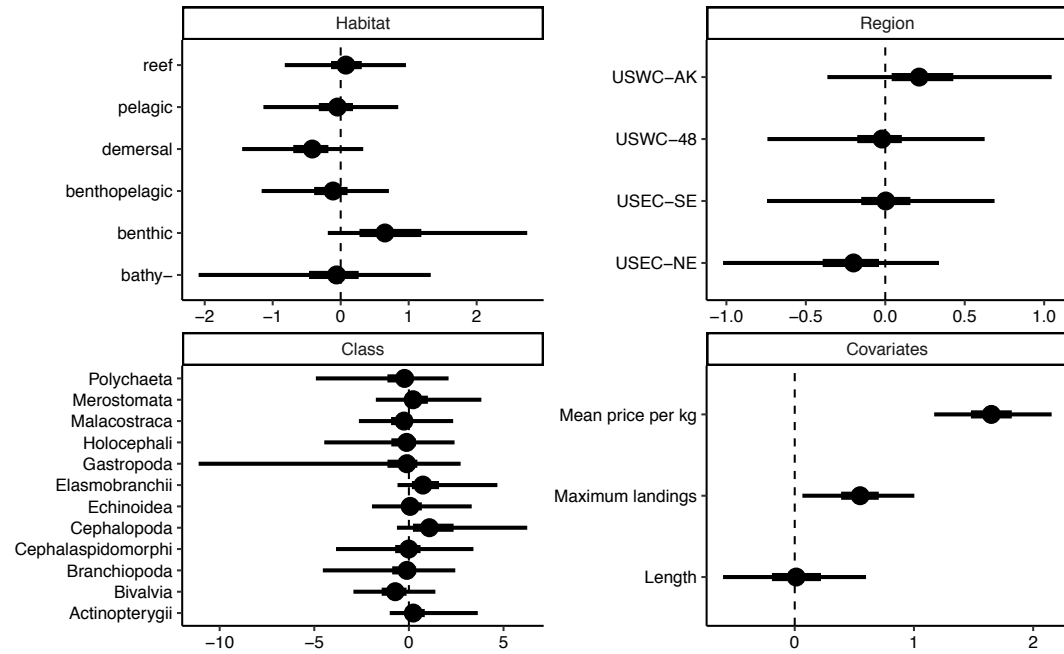


Figure 8: Summaries of estimated posterior distributions for a) continuous covariates in the model, b) habitat random effects, c) regional random effects, and d) taxonomic class random effects. The circle shows the posterior median, with thick bars showing the inter-quartile range of the posterior and the thin line is the 95% confidence interval.

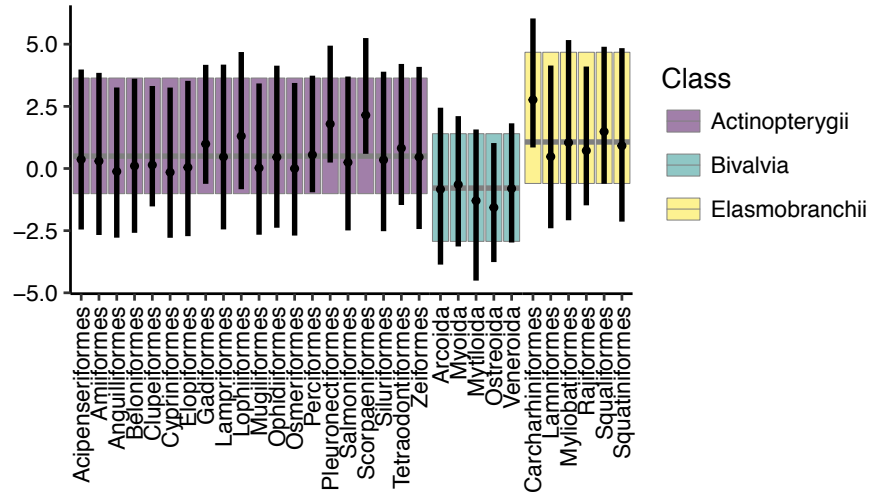


Figure 9: Summaries of estimated posterior distributions for order within class, showing a summary (grey line: posterior mean, coloured box: 95% confidence) of class effects, and order effects relative to class effects (points: posterior mean, black line: 95% confidence).