Estimating maturity from size-at-age data: Are real-world fisheries datasets up to the task?

Henry F. Wootton[1]* (ORCID: 0000-0001-6506-0248), John Morrongiello[1] (ORCID: 0000-0002-9608-4151), Asta Audzijonyte[2] (ORCID: 0000-0002-9919-9376)

[1] School of Biosciences, The University of Melbourne, Melbourne, Victoria, Australia.

[2] IMAS, University of Tasmania, Hobart, Tasmania, Australia.

* hwootton1@student.unimelb.edu.au

Acknowledgements

Abstract

The size and age at which individuals mature is rapidly changing due to plastic and evolved responses to fisheries harvest and global warming. Understanding the nature of these changes is essential because maturity schedules are critical in determining population demography and ultimately, the economic value and viability of fisheries. Detecting maturity changes is, however, practically difficult and costly. A recently proposed biphasic growth modelling likelihood profiling method offers great potential as it can statistically estimate age-at-maturity from population-level size-at-age data, using the change-point in growth that occurs at maturity. Yet, the performance of the method on typical marine fisheries datasets remains untested. Here, we assessed the suitability of 12 North Sea and Australian species' datasets for the likelihood profiling approach. The majority of the fisheries datasets were unsuitable as they had too small sample sizes or too large size-at-age variation. Further, datasets that did satisfy data requirements generally showed no correlation between empirical and model-derived maturity estimates. To understand why the biphasic approach had low performance we explored its sensitivity using simulated datasets. We found that method performance for marine fisheries datasets is likely to be low because of: 1) truncated age structures due to intensive fishing, 2) an under-representation of young individuals in datasets due to common fisheries-sampling protocols, and 3) large intrapopulation variability in growth curves. To improve our ability to detect maturation changes from population level size-at-age data we need to improve data collection protocols for fisheries monitoring.

Key Words

Biphasic growth model, Lester model likelihood profiling, statistical maturity estimates, fisheries-induced evolution, maturity changes, simulations, life history.

Introduction

Phenotypic change is a prevalent response to human interference in wild populations (Hendry et al. 2008; Sih et al. 2011). In the sea, these changes are often attributed to plastic and evolutionary responses to harvest (Kuparinen and Festa-Bianchet 2017; Law 2007) and warming (Cheung et al. 2013; Crozier and Hutchings 2014). Rapid changes in age and size at maturation are particularly common (e.g. Sharpe and Hendry, 2008; Audzijonyte et al. 2013) and have been observed in many fisheries stocks (e.g. Audzijonyte et al. 2016). Such changes could be driven by evolutionary response to increased mortality and size-selectivity from fishing (Jorgensen et al. 2007), or by increasing temperatures through a range of mechanisms behind the temperature-size rule (Atkinson et al. 2006; Audzijonyte et al. 2019; Crozier and Hutchings 2014). Changes in maturation size and age have important implications for stocks and ecosystem dynamics because altered maturation schedules have obvious implications for demography, carrying capacities and economic yields (Eikeset et al. 2013; Heino et al. 2013; Law and Grey 1989). For example, shifts to earlier maturity reduce average body size in fish populations through trade-offs between growth and reproduction (Enberg et al. 2012). Smaller average body sizes can, in turn, cause reductions in *per-capita* fecundity (Roff 1983) and carrying capacities (Audzijonyte et al. 2013b). Life history data are also fundamental inputs to fisheries stock assessments (Methot and

Wetzel 2013); small errors in parameter estimates have been shown to significantly affect estimates of biomass, age structure and thus future protected catches (Whitten et al. 2013).

Regular spatio-temporal monitoring of maturity status is not available for many commercially important fish stocks. This is because in sexually monomorphic and indeterminately growing fish the transition to maturation is not readily observable from live animals (e.g. compared to rutting or nesting behaviours), so maturation status can only be confirmed by specific and invasive internal sampling. Innovative statistical approaches to estimate age and size at maturity from a population's average somatic growth trajectory (based on size-at-age data) offer great potential to rectify these data deficiencies (reviewed in Wilson et al. 2018). Such techniques generally rely on the assumption that somatic growth decreases as individuals begin to invest in reproduction prior to maturity (e.g. Rijnsdorp and Storbeck 1995) and that statistical methods can be used to detect this change in growth rate.

While the theory underpinning statistical approaches to estimate maturity is sound, finding an appropriate technique to successfully estimate maturity from size-at-age data has proven difficult. Recently, Honsey et. al. (2017) presented a novel application of the Lester (2004) biphasic growth-modelling technique to statistically estimate maturity status from size-at-age datasets, called Lester model likelihood profiling (LMLP). The method is based on joint maximum likelihood estimates of age at maturity and four other life-history parameters using length-at-age data and assumes that data can be approximated by the continuous biphasic growth model, with linear juvenile growth and von-Bertalanffy type post-maturation growth (see Wilson et al. 2018 for a review of different biphasic growth models). The likelihood intervals for the

maturity age are generated by finding the maximum likelihood of all growth parameters for a given age at maturity and then profiling across all plausible values of maturation age to find the confidence ranges of maturity age estimate. The LMLP method could accurately reproduce empirically observed maturity in walleye (*Sander vitreus*) gillnet scientific survey datasets from individual years across multiple lakes in North America; the method also worked on other datasets, such as a marine fish (haddock -*Melanogrammus aeglefinus*) or an amphibian (seal salamander -*Desmognathus monticola*) (Honsey et al. 2017). Further, Wilson et al. (2018) showed that LMLP had low bias and good precision, with the performance comparable to Bayesian methods. These findings are exciting, because if the method is generalisable, then LMLP could be routinely used to estimate maturity characteristics for subsequent statistical or stock assessment procedures (Methot and Wetzel 2013). However, the LMLP method is sensitive to high variation in size-at-age, small sample sizes, young age at maturity, uneven sample distribution across ages within a stock, and violations of biphasic growth model assumptions (Honsey et al. 2017). The applicability of LMLP has also not yet been tested for standard fishery datasets derived from fishery dependent (e.g. catch data) or independent (e.g. scientific trawls) sources. The main challenge with these datasets is that the youngest aged individuals in a population may not be available due to catches selectively excluding smaller individuals, or because high fishing mortalities have resulted in demographic truncation and the loss of the oldest aged individuals. Furthermore, fishery-dependent and independent datasets may not truly represent the underlying population because for example, depending on the gear selectivity in a fishery-dependent sample, only the fastest growing individuals may be vulnerable to capture in the age groups that are not yet fully recruited to fisheries (Morrongiello and Thresher 2015). Honsey et al. (2017) did not account for any size selectivity effects because it was not considered

important for the walleye dataset; the authors suggested that the importance of these biases should be considered on a species-by-species basis (same applies to Wilson et al. 2018). Together, these sample biases could affect the LMLP methods' ability to characterise population lifetime-growth and hinder maturity estimation.

In this study, we evaluated the applicability and performance of Honsey et al.'s (2017) LMLP method to estimate maturity using a range of size-at-age datasets from commercially important fish species in the North Sea (Europe), a region with prolonged and intensive fishing history (ICES 2018) and from south-east Australia, where fishing mortality is lower (Patterson et al. 2016). Such fishery-derived or fishery-independent datasets are commonly available around the world and span the decadal time-frames required to detect trends in maturity characteristics (Morrongiello et al. 2012), if the LMLP method can be shown to perform well. First, we evaluated the quality of commonly available datasets by testing how many of the North Sea and Australian species satisfied the data requirement (variation around size-at-age and sample sizes) for the method to perform (Honsey et al. 2017). Second, we focused on the fishery independent North Sea datasets where empirical maturation estimates were available. For these datasets we tested the performance of LMLP and showed that the method performed poorly. Empirical estimates of maturity were not available for the Australian datasets. Third, we used simulated fisheries-independent size-at-age datasets to investigate the minimal conditions that could result in adequate LMLP performance, and explored how age truncation or big variation in size-at-age affect maturity estimates. We outline conditions under which the method is unlikely to perform well and make recommendations for improved age-length data collection and life-history estimation.

Methods

Fisheries data

We used empirical age-length datasets from seven North Sea (Europe) and five south-east Australian species to assess the performance and generality of the LMLP method. North Sea datasets, collected from fishery independent trawl surveys, were downloaded from the ICES DATRAS (http://www.ices.dk/marine-data/dataset-collections/Pages/default.aspx) data portal on 11/5/2018. Australian datasets, collected from fishery dependent surveys of commercial trawl catch, were received from the Australian Fisheries Management Authority on the 29/5/2017. Further information on sampling methodologies to generate data from both regions can be found in the Supplement. Data from both the North Sea and Australia were used to investigate the general suitability of the LMLP method for use in fishery contexts (i.e for maturity estimation with datasets derived from fishery dependent and independent sampling). LMLP performance was then investigated using datasets from the North Sea alone as these data also contained empirical maturity information. Our 12 species were selected to represent high value fisheries and diverse life-history strategies, and included: Atlantic herring (*Clupea harengus*), Atlantic cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), plaice (*Pleuronectes platessa*), saithe (*Pollachius virens*), Atlantic mackerel (*Scomber scombrus*) and sprat (*Sprattus sprattus*) from the North Sea, and pink ling (*Genypterus blacodes*), orange roughy (*Hoplostethus atlanticus*), blue grenadier (*Macruronus novazelandiae*), jackass morwong (*Nemadactylus macropterus*) and tiger flathead (*Platycephalus richardsoni*) from south-east Australia (see Supplementary Table 1 for a list of life history parameters for each species).

Precision investigation

Initially we assessed whether length-at-age datasets from both the North Sea and south-east Australia were suitable to estimate maturity using the Lester model likelihood profiling (LMLP) method presented in Honsey et al. (2017). Honsey et al. (2017) explored the performance of LMLP and demonstrated that for populations with age at maturity = 5yrs, the method performs poorly on datasets with precision lower than 6 and sample sizes (n) smaller than 100. 'Precision' is defined as the mean inverse coefficient of variation (CV, which is standard deviation in length divided by mean length in the age group) in length at each age across ages, weighted by sample size-at-age within the dataset:

$$\hat{P} = \frac{\sum_{i=k}^{j}\left(\frac{1}{CV_i}*n_i\right)}{\sum_{i=k}^{j}n_i} \tag{1}$$

where $\hat{P}$ is precision, CV is the coefficient of variation at age i, k is the first age and n is weight (sample size) at age i across ages k through j. Note, that age i corresponds to discrete ages classes (1, 2, 3, etc.) and is different to the continuous time (t) used in the biphasic growth model (below). For our purposes, we chose a conservative average precision level of $\geq$10 as indicative of good performance, because some of our species had an age at maturity lower than 5.

We calculated average precision and average n across datasets, subsampled across species, sex, year of capture (Year), fishing zone of capture (Zone: see (ICES 2015) and (Morrongiello and Thresher 2015) for descriptions of zone designations), annual quarter of capture (Quarter) and month of capture (Month; only available for Australian datasets), and all relevant combinations of these (i.e. Zone by Year, Zone by Sex, Year by Sex and so on, see Supplementary Table 2). Such subsampling reduces variation in length-at-age in the samples and should therefore increase

precision. Data points with missing relevant factorial information were removed prior to each $\hat{P}$ calculation; for example, if zone information was missing, the datapoints were removed from analyses that assessed precision in different zones separately, but not in analyses where zones were combined.

All data analyses were done using R 3.5.1 (R Development Core Team 2013) using the RStudio (RStudio Team 2015) interface (version 1.1.463).

Validation of age at maturity estimates

Honsey model

The Honsey et al. (2017) method is based on the Lester (2004) biphasic growth model which assumes linear growth for juveniles and asymptotic von Bertalanffy type growth after maturation. Juvenile (linear) and adult (asymptotic) growth are defined by four parameters (l_0, h, T, and g) which are estimated (including error σ^2) from length-at-age datasets using a maximum-likelihood approach (see; Honsey et al. 2017).

For length at time t (l_t), growth is defined by:

$$l_t = l_0 + ht, \quad t \le T \ for \ juveniles, \qquad (2)$$
$$l_t = l_\infty \left(1 - e^{-k(t-t_0)}\right), t > T \ for \ adults, \qquad (3)$$

where

$$t_1 = -\frac{l_0}{h},$$
$$l_\infty = \frac{3h}{g},$$
$$k = \ln(1 + \frac{g}{3}),$$

and $t_0 = T + \ln(1 - \frac{g\,(T-t_1)}{3})/\ln(1 + \frac{g}{3})$ as derived in Lester et al. (2004)

Here l_0 is the theoretical length at age 0 (mm), h is the net rate of energy acquisition expressed as somatic growth rate (mm/yr), T is the last immature age (yr; Lester age at maturity parameter), l_∞ is asymptotic length (mm), k is the von Bertalanffy growth coefficient or asymptotic growth rate (per year), t_0 is the von Bertalanffy hypothetical age at length 0 (yr), t_1 is the Lester hypothetical age at length 0 (yr) and g is the cost to somatic growth of maturity (expressed in equivalent energetic units and is constant for adults).

This formulation of the Lester (2004) model assumes that metabolism scales with body size in a two-thirds power relationship across taxa. The estimated parameter T can be interpreted as average age at maturity in the dataset (Honsey et al. 2017). The estimate of T (we hereafter refer to T as LMLP estimate) can then be compared to empirical assessments of maturity (hereafter referred to as 'empirical estimate of maturity' or EEM), such as the age at which 50% of individuals are mature age in a population as derived from logistic regression (Chen and Paloheimo 1994) using directly observed individual maturation status (mature or not) (as in Honsey et al. 2017).

Accurate growth curve estimates require unbiased sample from a population, but due to selectivity of sampling gears, only the fastest or slowest growing individuals might, for example, be collected in size groups that are not fully recruited to fisheries. Ideally, growth curve or maturity estimates therefore should simultaneously estimate or account for gear selectivity. In practice however, this would require even larger sample sizes and may not be feasible for many

species, and most growth estimation methods do not account for this potential bias (see Wilson et al. 2018 for a review or growth models). The Honsey et al. (2017) method did not consider the effect of gear selectivity on the samples used to derive growth curves in the analyses, because in their analyses "the potential impacts of gill net size selectivity on model estimates are likely relatively small for organisms such as walleye across the ages included in the simulations". The subsequent follow-up study by Wilson et al. (2018) also did not account for the gear selectivity effect. Because the goal of our study is to test the LMLP method presented by Honsey et al. (2017) and explored by Wilson et al. (2018) we follow the protocols provided in these two studies. In the simulation study (below) we show the method still cannot reliably estimate accurate maturity age even when non-selective sampling is applied.

LMLP validation

Subsampled datasets from four North Sea species satisfied the precision and sample size requirements for LMLP analyses. These data were from Atlantic herring, saithe, sprat and Atlantic mackerel. We estimated age at maturity of these accepted North Sea datasets using the LMLP method (using code developed in Honsey et al. (2017) which uses the R package 'boot' (Canty and Ripley 2017)) and compared these estimates to the empirical-estimates of maturity (EEM, see below). We do not formally assess LMLP performance for Australian species as time series of empirically derived maturity estimates do not exist for them. Australian data was used to assess the general suitability of fisheries data for LMLP.

North Sea LMLP analyses were conducted on all unique combinations of factors outlined above where precision was satisfied, e.g. Sex by Quarter (see Fig. 1 and Results for the specific factor combinations), and age at maturity was estimated (Equations 2 and 3) for each year where sufficient data was available. Following the recommendations of Honsey et al. 2017, model fits for each group of data were visually inspected and retained only if the likelihood surface contained a single peak and had confidence intervals around LMLP estimates of ≤ 2 years. Next, for each dataset that was used for LMLP analyses, we derived EEM estimates by applying a logistic regression function on maturation status data (see ICES (2015) for the methodology used to determine maturation status) to estimate the age at which 50% of individuals were mature (traditionally called *A50*) .

Here, the logit of the probability p of individual b being mature M (pM_b) was modelled as a function of age a_b;

$$logit(pM_b) = \beta_0 + \beta_1 a_b \qquad (4)$$

where β_0 and β_1 are estimated regression coefficients.

For each dataset, sample (population level) EEM estimates were extracted as age where 50% of individuals were estimated to be mature. In some instances, non-convergence or negative EEM estimates occurred and these data were removed from subsequent analyses. Correlation between LMLP and EEM estimates within subsampled datasets were then compared using standardised major axis regression using the R package 'smatr' (Warton et al. 2012). We apply the Huber's M estimation, which is robust to outliers (Warton et al. 2012). Standardised major axis regression assumes error in both variables and allows for the meaningful comparison of two estimates. It

also downplays the effects of outliers and thus results in a more robust assessment of performance. We also assessed LMLP performance by counting how many LMLP estimates were within ±1 year of EEM estimates for species with average empirical maturity >3 years (within ±0.5 year for species with average empirical maturity ≤3 years). Finally, we looked at whether the 95% CI's for LMLP estimates included EEM estimates for each stock dataset.

To ensure that our investigation of the performance of the LMLP method was exhaustive, we also ran an additional set of targeted analyses. These included:

1. Data selection was modified to reduce the chance of erroneous age estimation (replicated techniques employed in Honsey et al. (2017));

2. LMLP and EEM were compared after subtracting 1 year from EEM estimates to match differences in interpretation between the two metrics (see Supplement for further explanation;

3. LMLP estimates were judged against LMLP method error characterisation described in Honsey et al. (2017) which allows for empirical dataset characteristics (combinations of data amount, LMLP model-estimated g and precision) to be judged against minimum requirements generated from simulated size-at-age datasets. Individual datasets that failed error characterisation were excluded from subsequent comparison of LMLP against EEM.

None of these additional analyses significantly improved the performance of LMLP analyses, so we do not describe them further in the main text (see Supplement for full details).

Performance of LMLP method in simulated datasets

The LMLP method showed poor performance for the empirical datasets analysed above (see Results). We therefore explored its behaviour using simulated datasets based on characteristics

of two well studied Atlantic cod stocks with contrasting life-histories – the late maturing Greenland coast cod stock and the early maturing Skagerrak coast stock.

Population length-at-age data was simulated using a modified biphasic growth model published in Wilson et al. (2018). The simulation model uses von Bertallanfy growth curve parameters (l_∞, k and t_0), age at maturity and a vector of possible ages for a species as defined by its longevity. It then calculates Lester (2004) biphasic growth model parameters g, h and $t1$ to estimate size-at-age distributions (see supplied code in Electronic Supplement for details). For length at time t (l_t), growth is defined by Equations (2) and (3) used in LMLP analyses above. We parameterised the growth model using life history parameters from Fishbase. Specifically, Greenland cod parameters were l_∞ = 1540mm, k = 0.06, t_0 = -2.88 (Rätz et al. 1999) and age at maturity = 8 (Jónsson 1959), whilst for the Skagerrak stock they were and l_∞ = 1160mm, k = 0.208, t_0 = 0.18 and age at maturity = 2 (Froese and Sampang 2013). For simplicity, we ignored potential differences in growth between sexes. A coefficient of variation (CV) was applied to add variation around each modelled 'average' size across age classes. The baseline CV for the Greenland stock (CV = 0.156) was calculated using size-at-age variation in the North Sea Atlantic cod data from the ICES database (DATRAS database, analysed by year and zone, then averaged). For the Skagerrak stock the baseline CV (CV = 0.237) was taken from Olsen et al. (2009) estimates for 0-year old cod. Coefficients of variation were assumed to be constant across ages. Next, natural mortality was simulated using a decaying age-specific survival function to generate more realistic size-at-age distributions in simulated populations. Our mortality curve followed the approach in Wilson et al. (2018), where the probability of survival at age i (pS_i) is given by:

$$pS_i = pS_{i-1} * \frac{(\frac{g}{1.18}-1)}{-1}, \text{ for ages} \geq 2 \ (pS_i = 1 \text{ for } i = 1) \quad (5)$$

where g is the cost to somatic growth of maturity (as above in Equation (3)).

We explored a range of relevant values for CV in size-at-age, g, age at maturity, sample sizes and fisheries selectivity to assess whether certain assumptions about growth or sampling could explain low LMLP method performance. The CV values for both stocks were varied from 0.05 to 0.45 (at 5 steps each 1 unit value apart) while the values for cost of maturity to growth, g was varied from 0.05 to 0.8 (5 steps each 0.1875 units apart) in the Skagerrak stock alone, because varying g in Greenland stock simulations produced unrealistic growth trajectories. Age at maturity ranged from 5 to 9 years (from the baseline of 8) for the Greenland stock (5 steps 1 year apart) and 1 or 2 years for Skagerrak stock (baseline was 2). We simulated the impacts of fisheries on population age structure using different assumptions about fisheries gear selectivity. For that we altered the age based midpoint of the selectivity curve (age at which 50% of individuals are removed by fisheries) from 4 to 12 years (5 steps 2 years apart) for the Greenland stock, and 2 to 6 years (5 steps 1 year apart) for the Skagerrak stock. The probability of selection at age i (pF_i) is given by:

$$pF_i = \frac{1}{1 + e^{-\frac{age_i - sm}{st}}} \quad (6)$$

where sm is age at 50% selectivity, and st is the steepness of the selectivity curve. We use a steepness of 2 to reflect common selectivity patterns in fisheries (e.g. Blaber et al. (2005) and see Supplementary Figure 1 for an example of our selectivity curves), which means that if 50%

selectivity is at age 8, then ~25% and ~75% selectivity occurs at ages 6 and 10. Selectivity was not considered an important source of selection bias in Honsey et al. (2017).

After applying the fisheries mortality to the population, the remaining individuals were sampled randomly (proportionally to their abundance in the population) to be used in the LMLP maturity estimation. We also explored the impact of excluding the youngest individuals by removing from the sample individuals smaller than 100 to 500mm (5 steps each 100mm apart) for the slower growing Greenland stock and from 200 to 400mm (5 steps each 50mm apart) for the Skagerrak stock. We also varied the overall sample size n from 50 to 250 for both stocks (5 steps each 50 sample sizes apart) to explore how sample size affects method performance. For each combination of parameters (26 for each stock) we produced 10 replicate 'samples' of 250 individuals for the LMLP analyses (and fewer where effect of sample size was explored).

Low quality LMLP model fits (i.e. those with more than one likelihood peak and CI's > 2 years) were not excluded from these analyses because we wanted to explore raw LMLP method performance across all fits to data and equal sample sizes. LMLP method performance was assessed in two ways: first we counted how many of the 10 replicate LMLP estimates gave age of maturity values that were within ±1 year of the actual simulated values for the Greenland stock (maturing at 8 years) and within ±0.5 year for the Skagerrak stock (maturing at 2 years), and second whether 95% CI's for LMLP estimates included the simulated maturity values for each stock. The R code used to perform simulations is available as a supplement and on Github (link provided upon acceptance of manuscript).

Results

Precision analyses in different fisheries stocks

Only two species from the North Sea (Atlantic herring; $\hat{P} = 10$ and Atlantic mackerel; $\hat{P} = 11$) and two species from SE Australia (orange roughy; $\hat{P} = 12$ and blue grenadier; $\hat{P} = 10$) showed adequate precision using their complete datasets. When the full datasets for species were subsampled into factor combinations (e.g. by Zone, by Sex etc.), 46% or 74 of the resulting 161 North Sea datasets had average precision greater than 10 (Supplementary Table 2). Average precision values were much higher for the SE Australian species after subsampling, where 94% (165 out of 175 resultant datasets) had average precision >10 (Supplementary Table 2, Fig. 1).
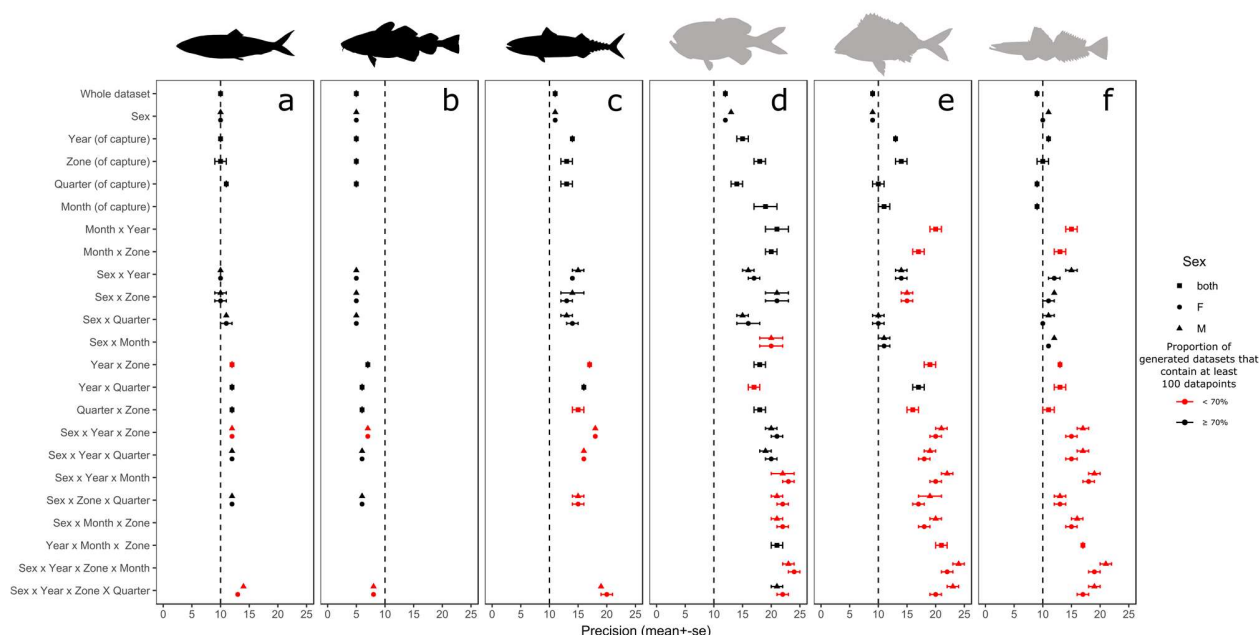
Figure. 1.Precision (*mean* and *se*) of length-at-age estimates in six selected example North Sea (black silhouettes) and south-east Australian (grey silhouettes) stocks for different levels of subsampling (represented by rows on the y-axis). Dashed lines represent a precision of 10, above which LMLP analyses are expected to perform well. Black points represent subsamples where ≥70% of data contain at least 100 individuals and thus, are suitable for LMLP analysis (red points show subsamples were <70% contain 100 individuals). Panel 'a' - Atlantic herring, 'b' - Atlantic cod, 'c' - Atlantic mackerel, 'd' - orange roughy, 'e' - jackass morwong and 'f' - tiger flathead (See Table S1 for all details).

Subsampling data by Sex, Year, Zone and Month showed the highest performance for SE Australian species (Sex, Year, Zone and Quarter for North Sea species) with a mean precision of 22 (12 for North Sea), compared to the mean precision of 9 when using original non-subsampled data per species (Supplementary Table 2, Fig. 1). Sample sizes necessarily decreased as subsampling became more complex, often falling below the minimum '100 individual' requirement (from Honsey et al. 2017) (Supplementary Table 2, Fig. 1). Nonetheless, sufficient precision and sample sizes were available for many subsamples, suggesting that fisheries datasets could satisfy the requirements for the LMLP method (Supplementary Table 2). For

example, in Atlantic herring across 100 generated Sex x Year x Quarter datasets, the mean precision in females was 11.93 (*se* = 0.38) and mean sample size was 761.21 (*se* = 56.14) while for males it was 11.88 (*se* = 0.26) and 702.7 (*se* = 49.72) (Supplementary Table 2). For orange roughy females, mean precision and mean sample size across 29 Sex x Year x Zone x Quarter datasets was 22.11 (*se* = 0.86) and 262.38 (*se* = 38.49), respectively.

Performance of LMLP method in North Sea datasets

The performance of the LMLP method was assessed for 16 subsampled datasets derived from four species with adequate precision, sample size and available empirical maturity estimates. These were Atlantic herring, saithe and sprat datasets that were subsampled by Sex, Year and Quarter, and Atlantic mackerel datasets subsampled by Year and Quarter. LMLP estimates performed poorly against EEM estimates in all but one of the 16 datasets (Table 1, Fig. 2, Supplementary Figures 2, 3, 4 and 5). The exception was Atlantic mackerel from Quarter 1 (Table 1, Fig. 2 row D). Across species, LMLP estimates fell within the specified qualitative ranges from EEM estimates (±1 year for species with average empirical maturity >3 years and ±0.5 year for species with average maturity ≤3 years) in 57% of cases (Fig. 2, Supplementary Figures 2, 3, 4 and 5). This underperformance was also mirrored when the LMLP method was characterised by testing the instances where LMLP CI's contained the EEM estimate across all stock datasets. This occurred in only 22% of cases (Fig. 2, Supplementary Figures 2, 3, 4 and 5).
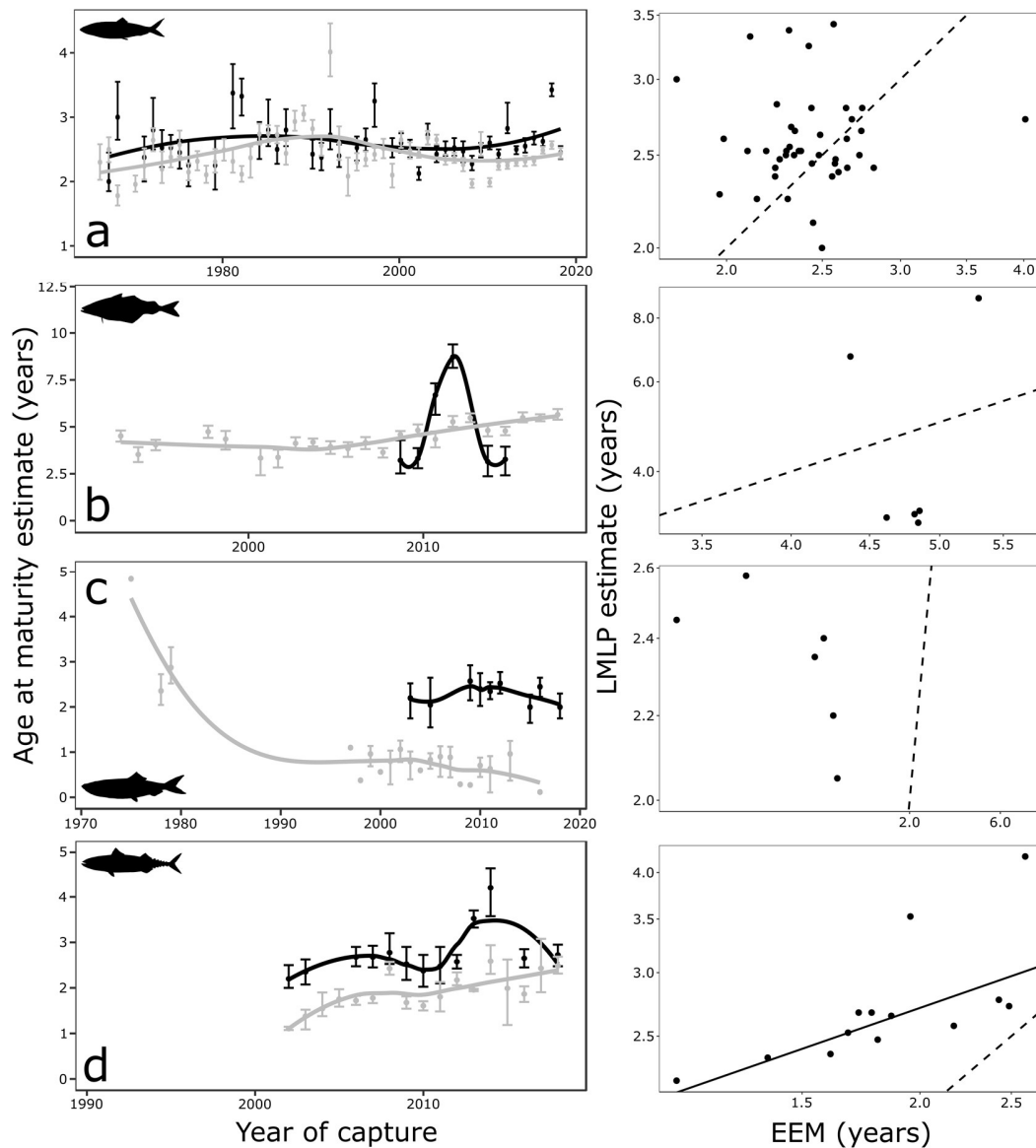
Figure 2. Comparison of statistical estimates of maturity (LMLP; *black lines showing LOESS smoothed trend*) against empirical estimates of maturity (EEM; *grey lines LOESS smoothed*) for four example datasets: row 'a' Atlantic herring females from Quarter 1, 'b' - Saithe males from Quarter 1, 'c' - Sprat males from Quarter 1, 'd' - Atlantic mackerel both sexes from Quarter 1. These datasets were selected because they represent a range of LMLP method performance across species and sexes (see Supplement for full representation of each species). Very wide error bars or those spanning 0 are not shown to improve clarity. The right column shows standardised major axis regression analyses of LMLP and EEM, where dashed lines represent a 1:1 relationship and the solid black line in row d shows the regression line for the significant correlation. Differences in temporal replication of LMLP and EEM arise from instances where estimates of maturity were not possible due to poor model fits to data in LMLP (i.e. CI's > 2 years or multiple likelihood peaks) or non-convergence in EEM (see methods). Note a log-transformation in axes scale in the right column of plots.

Table 1. Results of standardised major axis regression analyses comparing statistical estimates of age at maturity (LMLP) (biphasic approach) to empirical estimates (EEM) (logistic regression applied to raw maturity at age data) across Years for each Species, Sex (excluding Atlantic mackerel) and Quarter (within Year) combination. Parameter estimates for the regression intercept and slope with accompanying 95% confidence intervals are also reported. Perfect match between LMLP and EEM estimates would result in an R^2 value of 1 where slope is 1 and the intercept is 0.

Species	Sex	Quarter	Intercept	Slope	R^2	P-value
Atlantic herring	F	1	0.059 (-0.075, 0.059)	0.910 (0.626, 1.322)	<0.001	0.974
	F	2	0.862 (0.228, 1.495)	-1.206 (-3.786, -0.384)	0.035	0.721
	F	3	0.863 (0.600, 1.126)	-1.744 (-2.831, -1.075)	0.002	0.831
	F	4	-0.807 (-2.928, 1.314)	4.679 (1.263, 17.337)	0.508	0.287
	M	1	0.171 (0.088, 0.255)	0.584 (0.418, 0.831)	0.022	0.353
	M	2	0.703 (0.315, 1.091)	-0.776 (-2.366, -0.255)	0.001	0.943
	M	3	0.919 (0.636, 1.203)	-2.123 (-3.368, -1.339)	0.001	0.878
	M	4	-0.023 (-0.564, 0.517)	1.619 (0.547, 4.792)	0.137	0.540
Saithe	F	1	-0.503 (-2.056, 1.050)	1.941 (0.741, 5.081)	0.428	0.158
	M	1	5.409 (-3.163, 13.980)	-7.123 (-27.119, -1.871)	0.056	0.652
	M	3	-6.067 (-59.293, 47.160)	11.048 (0.685, 178.069)	0.202	0.703
Sprat	F	1	0.369 (0.328, 0.410)	-0.065 (-0.154, -0.028)	0.036	0.654
	F	3	0.177 (-0.185, 0.538)	0.754 (0.288, 1.972)	0.375	0.272
	M	1	0.328 (0.276, 0.381)	-0.125 (-0.300, -0.052)	0.443	0.149
Atlantic mackerel	-	1	0.310 (0.250, 0.370)	0.409 (0.247, 0.676)	0.464	0.010
	-	3	0.426 (0.357, 0.494)	0.480 (0.272, 0.847)	0.115	0.144

Assessing LMLP method performance with simulated datasets

Performance of the LMLP method was explored in simulated datasets for two Atlantic cod stocks with different life-history characteristics under a range of values for CV, g, age at maturity and selectivity parameters. Populations simulated when the CV was 0.15 for Greenland cod and when the CV was 0.25 for Skagerrak cod represent baseline parameters. The LMLP method successfully estimated maturity for the Greenland stock under the baseline scenario, where 9 out of 10 LMLP estimates were within ±1 year of the true value (Fig. 3 panel A). However, LMLP analyses did not successfully estimate maturity for the earlier-maturing

Skagerrak cod stock for baseline simulations (1 out of 10 LMLP estimates were within ±0.5 year of the true value) (Table 2, Fig. 3). LMLP analyses generally performed better across all simulations in the Greenland stock even when life-history parameters and sampling regimes were varied, with maturity estimates falling within ±one year of the true value in 72% of cases (Table 2, Fig. 3). In contrast, the method generally performed poorly for the early maturing Skagerrak stock, where maturity estimates fell within ±0.5 year of the true value in only 28% of cases. Notably, LMLP estimates from each of the 10 replicate samples for each parameter combination were variable, especially for high values of CV and small sample sizes (Table 2). The performance of LMLP analysis was most sensitive to increases in the minimum size limit of fish in the sample, where maturity estimates fell were outside the specified confidence bounds (±1 year for Greenland and ±0.5 year for Skagerrak simulations) in 70% of simulations across the two stocks for the largest minimum size limit (Table 2, Fig. 3 panel D). We also characterised the LMLP method's performance by testing whether the 95% CI's of LMLP estimates included the simulated maturity value (Table 2), but method performance outcomes for different parameter sets were similar to those outlined above.
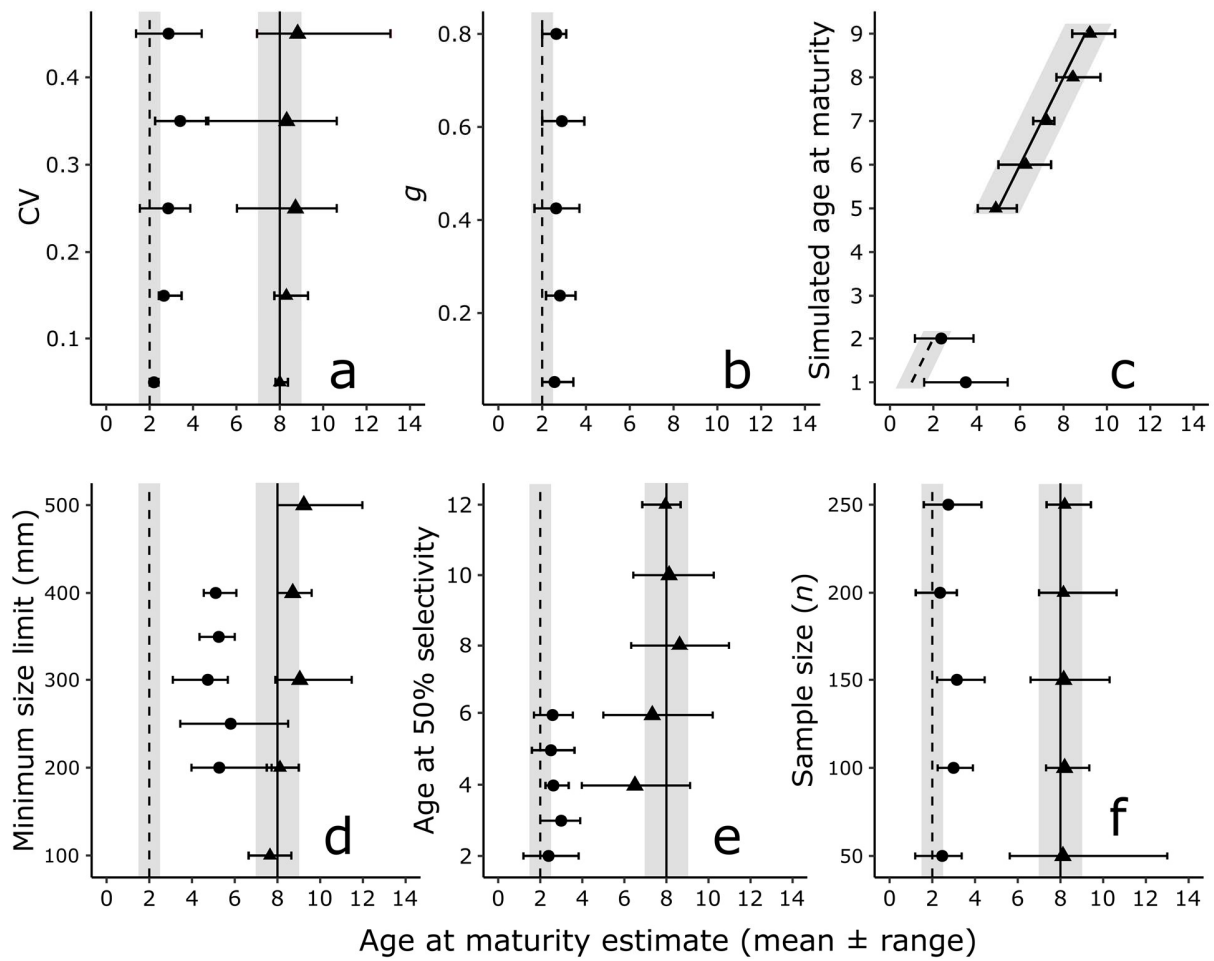
Figure 3. Statistical age at maturity estimates from 10 replicates of data sets for Greenland and Skagerrak cod stocks simulated using different life-history parameters and sampling regimes. Each panel represents a different parameter (panel 'a' - CV, 'b' - g, 'c' – age at maturity input, 'd' - minimum fisheries size limit (mm), 'e' – age at 50% selectivity and 'f' - sample size (n)). Dashed lines represent age at maturity inputs for the Skagerrak stock and solid lines are for the Greenland stock. Points represent mean values (squares show Skagerrak and triangles show Greenland stocks) and bars represent ranges (min and max values) for LMLP estimates across 10 replicate simulated populations. Grey shaded areas are the ranges of agreement for LMLP estimates against true maturity in simulations (±1 year for Greenland and ±0.5 year for Skagerrak stocks).

Table 2. Results of simulation analyses where LMLP estimates are compared to age at maturity model inputs in simulated Greenland and Skagerrak datasets. Age at maturity model inputs were 8 years for the Greenland stock and 2 years for the Skagerrak stock across simulations except where specifically varied (i.e. in Age at maturity input columns). Parameter values are represented in rows alongside LMLP ranges of maturity age (min and max values in years). Proportions of estimated 95% confidence intervals that contained the true maturity age and the proportion of estimates that fell within ±1 year (Greenland) or within ±0.5 year (Skagerrak) of the true maturity age are also shown (based on 10 replicate maturity estimates for each parameter combination). Parameter values with an asterisk represent baseline values for each stock.

Parameter	Greenland stock (Baseline parameters; age at maturity = 8, g = 0.18, h = 95.23, t1 = 0.25 and CV = 0.156)				Skagerrak stock (Baseline parameters; age at maturity = 2, g = 0.69, h = 268.21, t1 = 0.64 and CV = 0.237)			
	Value	LMLP range	Pr. LMLP 95% CI's that contain true value	Pr. LMLP within ±1 year of true value	Value	LMLP range	Pr. LMLP 95% CI's that contain true value	Pr. LMLP within ±0.5 year of true value
CV	0.05	7.8-8.4	0.9	1.0	0.05	2.0-2.4	0.2	1.0
	0.15 *	7.7-9.3	0.9	0.9	0.15	2.4-3.5	0.2	0.3
	0.25	6.0-10.6	0.8	0.6	0.25 *	1.5-3.9	0.3	0.1
	0.35	4.6-10.6	0.9	0.5	0.35	2.2-4.4	0.1	0.1
	0.45	6.9-13.1	0.9	0.6	0.45	1.4-4.4	0.3	0.3
g					0.050	2.0-3.4	0.3	0.5
					0.237	2.2-3.5	0.3	0.2
					0.425	1.6-3.7	0.4	0.4
					0.613	2.0-3.9	0.1	0.1
					0.800	2.0-3.1	0.4	0.2
Age at maturity input (years)	5	4.0-5.8	1	1.0	1	1.6-5.4	0	0.0
	6	5.0-7.4	0.8	0.8	2	1.1-3.8	0.3	0.5
	7	6.6-7.6	1	1.0				
	8	7.7-9.7	0.7	0.7				
	9	8.4-10.4	0.9	0.9				
Minimum size (mm)	100	6.6-8.6	0.9	0.9	200	4.0-7.7	0.2	0.0
	200	7.5-9.0	0.8	1.0	250	3.4-8.5	0.4	0.0
	300	7.9-11.5	0.6	0.5	300	3.1-5.7	0.0	0.0
	400	8.0-9.6	1.0	0.7	350	4.3-6.0	0.0	0.0
	500	8.0-12.0	0.8	0.6	400	4.5-6.0	0.0	0.0
50% sigmoidal fisheries selectivity	4	4.0-9.1	0.5	0.2	2	1.2-3.8	0.5	0.5
	6	5.0-10.2	0.7	0.6	3	2.0-3.9	0.6	0.4
	8	6.3-11.0	0.8	0.6	4	2.2-3.3	0.3	0.6
	10	6.4-10.2	0.7	0.6	5	1.6-3.6	0.3	0.5
	12	6.8-8.7	0.9	0.9	6	1.7-3.5	0.4	0.5
Sample size (n)	50	5.6-13	0.7	0.4	50	1.2-3.4	0.9	0.3
	100	7.3-9.3	1	0.8	100	2.2-3.9	0.2	0.1
	150	6.6-10.3	0.7	0.7	150	2.2-4.4	0.4	0.2
	200	7.0-10.6	0.6	0.8	200	1.2-3.1	0.4	0.4
	250	7.3-9.4	0.8	0.8	250	1.6-4.3	0.5	0.5

Discussion

In this study we show that, despite obvious appeal, the statistical estimation of maturity using the current formulation of the LMLP method seems unlikely to produce reliable estimates of maturity in many real-world fisheries datasets. Consistent with Honsey et al. 2017, we find that the method works well when applied to high quality datasets of species with specific life-history characteristics, including late maturation ($\gtrless$ 5 years), small variation around length-at-age (CV $\lessgtr$ 0.2), high continuity of datapoints across ages (e.g. minimum size limit $\leq$ 200 mm in Greenland simulations) and good representation of older individuals (50% fisheries selectivity midpoint $\gtrless$ age at maturity). Fisheries datasets, however, typically do not satisfy these requirements where there can be: large variation in length-at-age as seen in many geographically widespread species, age truncation due to intensive harvest, and an absence of the smallest individuals in a population due to fishery-dependent sampling.

The estimation of maturity from auxiliary data sources is an important ambition for the future of fisheries research. If successful, statistical estimates of life-history traits could represent relatively cheap sources of crucial data that could be used in current management contexts and to answer ecological and evolutionary questions across historical growth records. However, such techniques are yet to be successfully adopted in any applied examples that we know of.
A range of approaches have been proposed to statistically estimate age at maturity from either individual or population level growth data. Early methods employed segmented linear regression to individual growth increments (Rijnsdorp and Storbeck 1995), whilst more recently biphasic growth models have been applied to individual-level weight at age (Brunel et al. 2013; Mollet et

al. 2010) and population level size-at-age data (Minte-Vera et al. 2016) (techniques reviewed in; Honsey et al. 2017; Wilson et al. 2018). The failure to adopt these methods in standard fisheries research is surprising given their theoretical basis, validation against some real data (usually snapshots of high-quality data) and obvious application. Possible explanations for the lack of adoption could be based in the complexity of fitting many of these models, correlations between estimated parameters, demanding data requirements or simply low agreement with empirical estimates of maturity (Wilson et al. 2018).

When compared to earlier techniques, the likelihood profiling approach presented in Honsey et al. (2017) is attractive as it is based upon a popular biphasic growth modelling framework (the Lester (2004) model) and is also the first technique to use common, population level length-at-age data. We identify three possible reasons why LMLP failed to successfully estimate age at maturity in our study.

First, a history of intense harvest means that the North Sea stocks analysed here are age-truncated (ICES 2018). For example, only 12 of 120,000 Atlantic cod individuals were older than 12 years, despite this species living to at least 23 years (Beverton and Holt 1959). Likewise, only 9% of Atlantic herring were older than six years despite the species living for up to 15 years in the North Sea (Corten 2002). This truncation is in stark contrast to the Walleye datasets used to validate the LMLP method in Honsey et al. (2017) that comprised many old and large individuals, an even spread of individuals across ages, and adequate precision. The simulated datasets that were used to generate error characterisation for LMLP estimates (Honsey et al. 2017) also included many older individuals and were intended to replicate fisheries-independent

sampling. Age truncation in intensively harvested stocks is likely to be a key problem for LMLP methods, because growth trajectories cannot be properly estimated from such data and estimates of maturity become inaccurate.

Many commercially important stocks are harvested at high rates and show similar age truncation to the North Sea stocks (e.g. Berkeley et al. 2004; Hsieh et al. 2010). The datasets used to validate LMLP analyses in Honsey et al. (2017) could thus be considered atypical for large-scale fisheries. This highlights a potential paradox with obtaining the datasets required for maximal LMLP method performance in exploited stocks: whilst the heavily fished stocks generally have the largest size-at-age data collections and are most likely to display temporal trends in maturity ages (Audzijonyte et al. 2013a), their populations often lack the full age distribution needed to adequately estimate growth curves. Moreover, the restricted age distribution observed in our study is not caused by fisheries based sampling, as North Sea datasets were derived from fishery independent surveys. Therefore, the observed age truncation is a real phenomenon which cannot be "improved" other than by reducing fishing mortality.

Truncation of younger ages also affected estimation of age at maturity, as removing juveniles from simulated datasets also impacted the LMLP method's performance. This situation can arise when datasets are derived from catches with minimum size regulations (e.g. fishery-dependent surveys) as is the case for Australian datasets analysed here. This situation could be improved with better sampling, and at least occasional fisheries independent surveys (further simulations are needed to assess the minimum amount of surveys needed to adequately characterise early growth). Unfortunately, time series of empirically derived maturity estimates do not exist for the

Australian data and so we cannot formally assess the LMLP method's performance here. However, qualitative comparisons with the few snapshot empirical estimates of maturity for Australian species suggested poor LMLP method performance (results not shown). Datasets generated from sampling of commercial catches, such as in our Australian example, are subject to many sources of selection (e.g. selection on growth rate and/or fisher behaviour) which can result in further data bias and failure to estimate accurate population level growth curves. Given that LMLP performed poorly on fishery independent datasets that we assume to be unbiased, LMLP is even less likely to work in fishery-dependent contexts. Nevertheless, the LMLP method could potentially be useful for less intensively harvested stocks and may hold promise for maturity assessment of new fisheries.

The second reason why the LMLP method may not have performed well is that biphasic approaches are sensitive to the duration of the juvenile growth phase (Honsey et al. 2017). A relatively young age at maturity could mean that not enough data is available to adequately characterize juvenile growth when size data are derived annually. This scenario is likely commonplace as age data within fishery datasets is usually estimated from annual growth increments deposited in the otoliths of individual fish (Morrongiello et al. 2012). Indeed, our simulations of an early maturing stock from the Norwegian Skagerrak coast showed weak LMLP technique performance regardless of parameter values and sampling regime. This finding is in agreement with simulated early-maturing populations provided in Honsey et al. (2017). Paradoxically however, the only North Sea dataset where the LMLP method gave a similar result to the empirical age at maturity estimate was the early maturing (approx. 2 years) Atlantic mackerel in Quarter 1, whereas LMLP estimates for the later maturing saithe were generally bad.

It is possible that the mackerel result is due to chance given the number of statistical comparisons made in our study.

Third, modelling approaches to estimate maturity from growth data rely on accurate estimation of age and size; these characteristics are often themselves prone to measurement error (Campana 2001). Measurement error is likely to result in decreased precision of size-at-age datasets and in turn affect LMLP estimate accuracy. Moreover, the empirical estimates of maturity used to validate the LMLP method here are treated as true values, yet they necessarily have associated errors and biases. Empirical maturity estimation is usually described by macroscopic inspection of gonad tissue which is prone to error (Flores et al. 2019) and logistic regression models commonly used to characterise maturity are also sensitive to issues such as low sample size. Any biases in sampling used to assess population level maturity (e.g. capturing only the fastest growing individuals at around the maturation age) will even further increase this error. LMLP and EEM also have slightly different interpretations of maturity, where LMLP attempts to estimate the onset of investment in reproduction and EEM represents a population average based on visual inspection of gonads (Honsey et al. 2017). We attempted to address the impact of this difference by regressing LMLP against EEM-1 year (as suggested in: Honsey et al. 2017), but performance was not improved. Moreover, given the assumption that any differences between LMLP and EEM are consistent within a species, our analyses would still have identified correlations between the two, if present. Generally, the empirical maturity estimates of the North Sea species used in our analyses are considered robust and used as time-varying parameters in age based stock assessment methods (ICES 2019). Nevertheless, the issue of regressing statistical estimates of maturity against empirical measures, each with their own potential sources

of error, poses a general statistical problem that cannot be addressed in this study or when dealing with real world fisheries datasets. Ideally, statistical techniques should be validated against datasets of the highest quality but with reasonably high variability in size-at-age and size at maturity; unfortunately for most large fisheries stocks such datasets are impossible to obtain.

Finally, it remains unclear to what degree selectivity associated with the fisheries-dependent or independent sampling is likely to affect statistical estimates of maturity. If some age groups are not sampled randomly (i.e. fast or slow growing individuals are more likely to be caught), this may bias inferred growth curves compared to the real population sample. All sampling is inherently selective which may bias empirical estimates of maturity or growth away from 'truth' (e.g. Bettoli and Miranda (2001)) or at least disagree with datasets collected with differing methodologies (e.g. scientific and commercial catch data). While these issues should be explored further, it is reasonable to assume that, as long as sampling is done in a similar way, LMLP should still be able to detect temporal trends, even if consistently biased from the true population value. This was not, however, the case in our study.

Future directions

Stock assessment models can be sensitive to changes in life history traits such as maturity, with even small changes causing shifts to reference points in population estimation (Methot and Wetzel 2013). Given that age at maturity is known to respond to both warming (e.g. Crozier and Hutchings 2014) and fisheries induced mortality (e.g. Law 2007), failure to account for these changes can increase assessment error and ultimately erode our ability to manage harvests

sustainably (Laugen et al. 2014). This is likely to be the case both for assessments that assume maturity as a function of age or as a function of length, as in both cases the growth and productivity of the stock will change as maturation changes (Laugen et al. 2014). For many harvested species, however, maturity is not routinely assessed and successful statistical techniques to estimate maturity are critically needed to improve management of fisheries stocks and contribute to 'evolutionarily enlightened management strategies' (Ashley et al. 2003; Jorgensen et al. 2007).

Further exploration of Bayesian approaches to estimate statistical maturity from growth data is warranted (Wilson et al. 2018). Whilst performance of Bayesian and likelihood profiling (as applied in this manuscript) have been shown to perform similarly on simulated datasets, the possibility of including prior information on maturity may improve the performance of Bayesian approaches (Wilson et al. 2018). However, our study shows that performance should be tested upon 'real' datasets, as simulations typically assume perfect spatial and temporal mixing and are likely to underestimate the amounts of variability seen in real fishery datasets. The use of integrated models to estimate maturity should also be investigated as a promising technique (Maunder and Punt 2013). Here, growth, maturity and selectivity could be estimated on the same joint likelihood distribution which could alleviate some of the issues around fisheries selection and data quality as discussed above. Establishing minimum data quality requirements of statistical maturity estimates is a critical step for use in fisheries management. We also need to better explore cases, stocks and assessment methods that are most sensitive to small changes in maturation age and size.

One important step in aiming to improve statistical maturity estimates should consider back calculated individual length-at-age data, where an individual fish's previous length is inferred from measured growth increments in biological hard parts such as otoliths (Francis 1990). Here, individual growth curves would be combined by year of capture or cohort with the inclusion of random terms into statistical methods, such as LMLP. This is different to the approach we investigated in this manuscript, where growth is inferred from population averages. The inclusion of individual growth would align the LMLP method with earlier approaches to estimate age at maturity that use the less common weight-at-age data format (Brunel et al. 2013; Mollet et al. 2010). Individual growth estimates may also allow enough data for LMLP to be estimated for cohorts rather than by year of capture, which has been the norm (see: Wilson et al. (2018)). Analysing by cohort groups individuals into shared environmental experiences and would likely improve growth and maturity characterisation through for example, reductions in variation around life-histories. It should be noted, however, that Mollet and Brunel's techniques mentioned above also had issues when tested against simulated and fisheries data (reviewed in: Honsey et al. 2017), and no examples yet exist of application in fisheries management. Nonetheless, the estimation of individual growth improves the temporal extent and quality of existing fishery datasets as multiple size-at-age records become available for each individual. This in turn allows for datasets to be subsetted with greater complexity, whilst still maintaining minimum data requirements, which should increase precision (as evidenced by our precision analysis above) and thus drive greater LMLP method performance. Additionally, such otolith based data will provide access to early growth records through back calculation from older catches and will aid estimation of juvenile growth, especially for datasets where juvenile data is lacking. However, LMLP method performance would still need to be assessed as back-

calculation provides another potential source of error (Campana 2001). Notably, the use of individually derived data will not help in cases where older ages are simply missing due to intensive fishing, and the issue of no maturity data for stocks with which to validate LMLP still remains.

Acknowledgements

Conflict of interest

The authors declare that they have no conflict of interest.

Funding

Data availability

Data from the North Sea is publicly available from Datras (http://www.ices.dk/marine-

data/dataset-collections/Pages/default.aspx) and South East Australian data is available from the

Australian Fisheries Management Authority (AFMA) upon request.

References

Ashley MV, Willson MF, Pergams ORW, O'Dowd DJ, Gende SM, Brown JS (2003) Evolutionarily enlightened management. Biol Conserv 111:115-123. doi:10.1016/s0006-3207(02)00279-3

Atkinson D, Morley SA, Hughes RN (2006) From cells to colonies: at what levels of body organization does the 'temperature-size rule' apply? Evolution & Development 8:202-214. doi:10.1111/j.1525-142X.2006.00090.x

Audzijonyte A, Barneche DR, Baudron AR, Belmaker J, Clark TD, Marshall CT, Morrongiello JR, van Rijn I (2019) Is oxygen limitation in warming waters a valid mechanism to explain decreased body sizes in aquatic ectotherms? Global Ecol Biogeogr 28:64-77. doi:10.1111/geb.12847

Audzijonyte A, Kuparinen A, Fulton EA (2013a) How fast is fisheries-induced evolution? Quantitative analysis of modelling and empirical studies. Evolutionary Applications 6:585-595. doi:10.1111/eva.12044

Audzijonyte A, Kuparinen A, Gorton R, Fulton EA (2013b) Ecological consequences of body size decline in harvested fish species: positive feedback loops in trophic interactions amplify human impact. Biol Lett 9. doi:10.1098/rsbl.2012.1103

Berkeley SA, Hixon MA, Larson RJ, Love MS (2004) Fisheries Sustainability via protection of age structure and spatial distribution of fish populations. Fisheries 29:23-32. doi:10.1577/1548-8446(2004)29[23:fsvpoa]2.0.co;2

Bettoli PW, Miranda LE (2001) Cautionary note about estimating mean length at age with subsampled data. N Am J Fish Manage 21:425-428. doi:10.1577/1548-8675(2001)021<0425:cnaeml>2.0.co;2

Beverton RJH, Holt SJ (1959) A review of the life-spans and mortality rates of fish in nature, and their relationship to growth and other physiological characteristics. CIBA Foundation Colloquia on Ageing 54:142-180

Blaber SJM, Dichmont CM, Buckworth RC, Badrudin, Sumiono B, Nurhakim S, Iskandar B, Fegan B, Ramm DC, Salini JP (2005) Shared stocks of snappers (Lutjanidae) in Australia and Indonesia: Integrating biology, population dynamics and socio-economics to examine management scenarios. Rev Fish Biol Fish 15:111-127. doi:10.1007/s11160-005-3887-y

Brunel T, Ernande B, Mollet FM, Rijnsdorp AD (2013) Estimating age at maturation and energy-based life-history traits from individual growth trajectories with nonlinear mixed-effects models. Oecologia 172:631-643. doi:10.1007/s00442-012-2527-1

Campana SE (2001) Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. J Fish Biol 59:197-242. doi:10.1006/jfbi.2001.1668

Chen Y, Paloheimo JE (1994) Estimating fish length and age at 50-percent maturity using a logistic type model. Aquat Sci 56:206-219. doi:10.1007/bf00879965

Cheung WWL, Sarmiento JL, Dunne J, Froelicher TL, Lam VWY, Palomares MLD, Watson R, Pauly D (2013) Shrinking of fishes exacerbates impacts of global ocean changes on marine ecosystems. Nature Climate Change 3:254-258. doi:10.1038/nclimate1691

Corten A (2002) The role of "conservatism" in herring migrations. Rev Fish Biol Fish 11:339-361

Crozier LG, Hutchings JA (2014) Plastic and evolutionary responses to climate change in fish. Evolutionary Applications 7:68-87. doi:10.1111/eva.12135

Eikeset AM, Richter A, Dunlop ES, Dieckmann U, Stenseth NC (2013) Economic repercussions of fisheries-induced evolution. Proceedings of the National Academy of Sciences of the United States of America 110:12259-12264. doi:10.1073/pnas.1212593110

Enberg K, Jorgensen C, Dunlop ES, Varpe O, Boukal DS, Baulier L, Eliassen S, Heino M (2012) Fishing-induced evolution of growth: concepts, mechanisms and the empirical evidence. Mar Ecol-Evol Persp 33:1-25. doi:10.1111/j.1439-0485.2011.00460.x

Flores A, Wiff R, Ganias K, Marshall CT (2019) Accuracy of gonadosomatic index in maturity classification and estimation of maturity ogive. Fisheries Research 210:50-62. doi:10.1016/j.fishres.2018.10.009

Francis R (1990) Back-calculation of fish length - a critical-review. J Fish Biol 36:883-902. doi:10.1111/j.1095-8649.1990.tb05636.x

Froese R, Sampang A (2013) Potential Indicators and Reference Points for Good Environmental Status of Commercially Exploited Marine Fishes and Invertebrates in the German EEZ. World Wide Web electronic publication

Heino M, Baulier L, Boukal DS, Ernande B, Johnston FD, Mollet FM, Pardoe H, Therkildsen NO, Uusi-Heikkila S, Vainikka A, Arlinghaus R, Dankel DJ, Dunlop ES, Eikeset AM, Enberg K, Engelhard GH, Jorgensen C, Laugen AT, Matsumura S, Nussle S, Urbach D, Whitlock R, Rijnsdorp AD, Dieckmann U (2013) Can fisheries-induced evolution shift reference points for fisheries management? ICES J Mar Sci 70:707-721. doi:10.1093/icesjms/fst077

Hendry AP, Farrugia TJ, Kinnison MT (2008) Human influences on rates of phenotypic change in wild animal populations. Mol Ecol 17:20-29. doi:10.1111/j.1365-294X.2007.03428.x

Honsey AE, Staples DF, Venturelli PA (2017) Accurate estimates of age at maturity from the growth trajectories of fishes and other ectotherms. Ecol Appl 27:182-192. doi:10.1002/eap.1421

Hsieh CH, Yamauchi A, Nakazawa T, Wang WF (2010) Fishing effects on age and spatial structures undermine population stability of fishes. Aquat Sci 72:165-178. doi:10.1007/s00027-009-0122-2

ICES (2015) Manual for the International Bottom Trawl Surveys. Series of ICES Survey Protocols SISP 10 - IBTS IX. 86 pp

ICES (2018) Greater North Sea Ecoregion – Fisheries overview. doi:10.17895/ices.pub.4647

ICES (2019) Working group on the assessment of demersal stocks in the North Sea and Skagerrak (WGNSSK). vol 1:7. ICES Scientific Reports. 1271 pp

Jónsson J (1959) On the spawning stocks of cod in east Greenlandic and Icelandic waters. vol 103. ICES.

Jorgensen C, Enberg K, Dunlop ES, Arlinghaus R, Boukal DS, Brander K, Ernande B, Gardmark A, Johnston F, Matsumura S, Pardoe H, Raab K, Silva A, Vainikka A, Dieckmann U, Heino M, Rijnsdorp AD (2007) Ecology - Managing evolving fish stocks. Science 318:1247-1248. doi:10.1126/science.1148089

Kuparinen A, Festa-Bianchet M (2017) Harvest-induced evolution: insights from aquatic and terrestrial systems. Philosophical Transactions of the Royal Society B-Biological Sciences 372. doi:10.1098/rstb.2016.0036

Laugen AT, Engelhard GH, Whitlock R, Arlinghaus R, Dankel DJ, Dunlop ES, Eikeset AM, Enberg K, Jorgensen C, Matsumura S, Nussle S, Urbach D, Baulier L, Boukal DS, Ernande B, Johnston FD, Mollet F, Pardoe H, Therkildsen NO, Uusi-Heikkila S, Vainikka A, Heino M, Rijnsdorp AD, Dieckmann U (2014) Evolutionary impact assessment: accounting for evolutionary consequences of fishing in an ecosystem approach to fisheries management. Fish Fish 15:65-96. doi:10.1111/faf.12007

Law R (2007) Fisheries-induced evolution: present status and future directions. Mar Ecol Prog Ser 335:271-277. doi:10.3354/meps335271

Law R, Grey DR (1989) Evolution of yields from populations with age-specific cropping. Evol Ecol 3:343-359. doi:10.1007/bf02285264

Maunder MN, Punt AE (2013) A review of integrated analysis in fisheries stock assessment. Fisheries Research 142:61-74. doi:10.1016/j.fishres.2012.07.025

Methot RD, Wetzel CR (2013) Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142:86-99. doi:10.1016/j.fishres.2012.10.012

Minte-Vera CV, Maunder MN, Casselman JM, Campana SE (2016) Growth functions that incorporate the cost of reproduction. Fisheries Research 180:31-44. doi:10.1016/j.fishres.2015.10.023

Mollet FM, Ernande B, Brunel T, Rijnsdorp AD (2010) Multiple growth-correlated life history traits estimated simultaneously in individuals. Oikos 119:10-26. doi:10.1111/j.1600-0706.2009.17746.x

Morrongiello JR, Thresher RE (2015) A statistical framework to explore ontogenetic growth variation among individuals and populations: a marine fish example. Ecol Monogr 85:93-115. doi:10.1890/13-2355.1

Morrongiello JR, Thresher RE, Smith DC (2012) Aquatic biochronologies and climate change. Nature Climate Change 2:849-857. doi:10.1038/nclimate1616

Patterson H, Noriega R, Georgeson L, Stobutzki I, Curtotti R (2016) Fishery status reports 2016, Australian Bureau of Agricultural and Resource Economics and Sciences. Canberra. CC BY 3.0.

R Development Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rätz H-J, Stein M, Lloret J (1999) Variation in growth and recruitment of Atlantic cod (Gadus morhua) off Greenland during the second half of the twentieth century. J Northwest Atl Fish Sci 25:161-170. doi:10.2960/J.v25.a15

Rijnsdorp AD, Storbeck F (1995) Determining the onset of sexual maturity from otoliths of individual female North Sea plaice, Pleuronectes platessa L. Recent Developments in Fish Otolith Research, vol 19. In: Secor D, Dean JM,Campana S (ed) Recent

developments in fish otolithresearch. University of South Carolina Press, Columbia,SC, p 581–598.

Roff DA (1983) An allocation model of growth and reproduction in fish. Can J Fish Aquat Sci 40:1395-1404. doi:10.1139/f83-161

RStudio Team (2015) RStudio: Integrated Development for R. RStudio, Inc., Boston, MA

Sih A, Ferrari MCO, Harris DJ (2011) Evolution and behavioural responses to human-induced rapid environmental change. Evolutionary Applications 4:367-387. doi:10.1111/j.1752-4571.2010.00166.x

Whitten AR, Klaer NL, Tuck GN, Day RW (2013) Accounting for cohort-specific variable growth in fisheries stock assessments: A case study from south-eastern Australia. Fisheries Research 142:27-36. doi:10.1016/j.fishres.2012.06.021

Wilson KL, Honsey AE, Moe B, Venturelli P (2018) Growing the biphasic framework: Techniques and recommendations for fitting emerging growth models'. Methods in Ecology and Evolution 9:822-833. doi:10.1111/2041-210x.12931

Estimating maturity from size-at-age data: Are real-world fisheries datasets up to the task?

Supplementary material

Reviews in Fish Biology and Fisheries

Henry F. Wootton[1]*, John Morrongiello[1], Asta Audzijonyte[2]

[1] School of Biosciences, The University of Melbourne, Melbourne, Victoria, Australia.

[2] IMAS, University of Tasmania, Hobart, Tasmania, Australia.

* hwootton1@student.unimelb.edu

1

Supplementary methods

Summary of sample collection methodologies

Datasets used in this manuscript were from the North Sea and south-east Australia. North Sea data come from fishery independent surveys while Australian data derive from fishery catch records.

North Sea datasets were downloaded from the DATRAS database (http://www.ices.dk/marine-data/dataset-collections/Pages/default.aspx) on the 11/5/2018. Specimens contributing to this dataset are collected in the International Bottom Trawl Survey (IBTS) in the North Sea which is contributed to by multiple European countries and overseen by the International Council for the Exploration of the Sea (ICES) (ICES 2015). These surveys have been conducted since the beginning of the 1960's and methodologies have varied slightly over time where for example, quarter of the year sampled, sweep lengths and fishing gears have changed (ICES 2015). Trawls have occurred each year and are stratified into survey grids of approximately 30 x 30 nautical miles and multiple species and data types are recorded (e.g. recruitment indices, monitoring of commercial stocks and collection of hydrographical and environmental information) (ICES 2015). Trawl 'shots' are standardised (e.g. standard speed is 4 knots and 30 minutes in length) and the entire catch is sorted and sampled when possible (large catches can be subsampled) (ICES 2015). Catches are sorted into species for measurement of length, sex, age and maturity (ICES 2015).

35   Australian datasets were obtained from the Australian Fisheries Management Authority with

36   permission on the 29/5/2017. These data are part of a routine monitoring programme run by

37   Australian authorities since 1941 on the Southern and Eastern Scalefish and Shark Fishery

38   (Knuckey et al. 1999). This fishery covers a large geographical area spanning the south-east

39   coast of Australia from the state of Queensland through to South Australia. Fleets operate out of

40   multiple ports, target multiple species and mainly use trawl gears (Knuckey et al. 1999).

41   Fisheries observers regularly obtain specimens from commercial operations at ports to provide

42   age estimates for stock assessment modelling (Knuckey et al. 1999).

43 Supplementary Table 1. Life-history parameters for North Sea and Australian species analysed in the manuscript. Von Bertalanffy
44 growth parameters for each species are taken (as averages) from Fishbase (www.Fishbase.org). Linf is the asymptotic length, K is the
45 rate at which growth approaches asymptotic length (of dimension time^{-1}) and Ø is a growth performance index equal to $\log_{10} K + 2/3$
46 $\log_{10}$ Linf. Maturity parameters are shown from studies on North Sea or Australian stocks unless specified. Lengths represent total
47 length and sexes are combined unless indicated otherwise (i.e. standard length (sl)).

Parameter	Atlantic herring	Atlantic cod	Haddock	Plaice	Saithe	Atlantic mackerel	Sprat	Pink ling	Orange roughy	Blue grenadier	Jackass morwong	Tiger flathead
Linf (cm)	31	110	70	54.4	111	42	14.7	136	40.1 (sl)	102	45.5	62.2
K	0.3	0.1	0.2	0.1	0.1	0.2	0.8	0.1	0.05	0.2	0.2	0.3
Ø	2.45	3.25	2.99	2.51	3.09	2.62	2.25	3.24	1.85	3.22	2.54	3.05
Age at maturity (years)	2.7	3.8	M: 2 F: 2.5	2.5	4.6	2-3	1.2 (Baltic data)	5-7	32	F: 5.4 M: 4.3	F 3 M: 3	4-5
Length at maturity (cm)	24.1	69.7	F: 31.5 M: 27.5	26.6	55.4	28 (Western stock)	8.5	60-74	28-32	F: 63.7 (sl) M: 56.8 (sl)	F: 25 M: 27	F: 36 M: 30

48 *Maturity parameters for each specific species are from as follows: Atlantic herring (Froese and Sampang 2013), Atlantic cod (ICES 1996), haddock (Hislop and Shanks 1981), plaice (age: (ICES 2017)*
49 *and length: (ICES 1996)), saithe (ICES 1996), Atlantic mackerel (age:(ICES 2005) and length: (Lockwood and Shepherd 1984)) and sprat (Froese and Sampang 2013) from the North Sea and pink ling*
50 *(age: (Caton and McLaughlin 2000) and length: (Tilzey 1994)), orange roughy (age: (Fenton et al. 1991) and length: (Mace et al. 1990)), blue grenadier (Russel and Smith 2007), jackass morwong*
51 *(age: (Smith and Wayte 2001) and length: (Jordan 1999)) and tiger flathead (Fairbridge 1951) from south-east Australi*

52
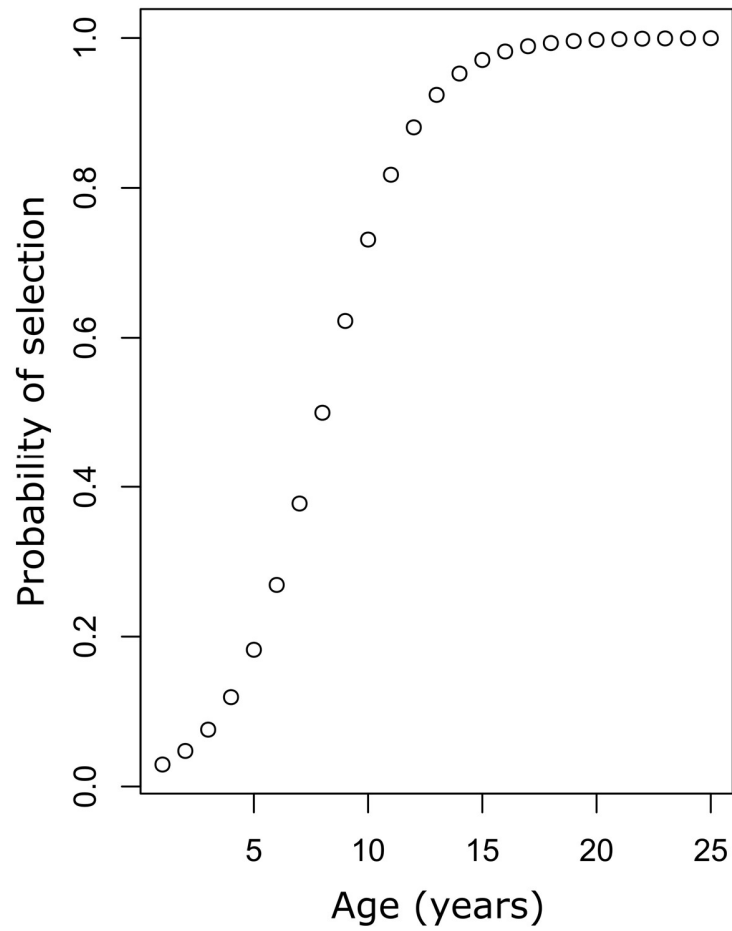
53   Supplementary Figure 1. Example of selectivity curve (equation (6) in the manuscript) applied to

54   populations in our simulation analyses where the probability of selection (fishing mortality) is

55   plotted against age. Here, the selection midpoint (sm) is 8 and the steepness (st) of the curve is

56   2.

57

58

59

60

Supplementary results

Supplementary Table 2. Full results of precision calculations in North Sea and south east Australian species. Calculation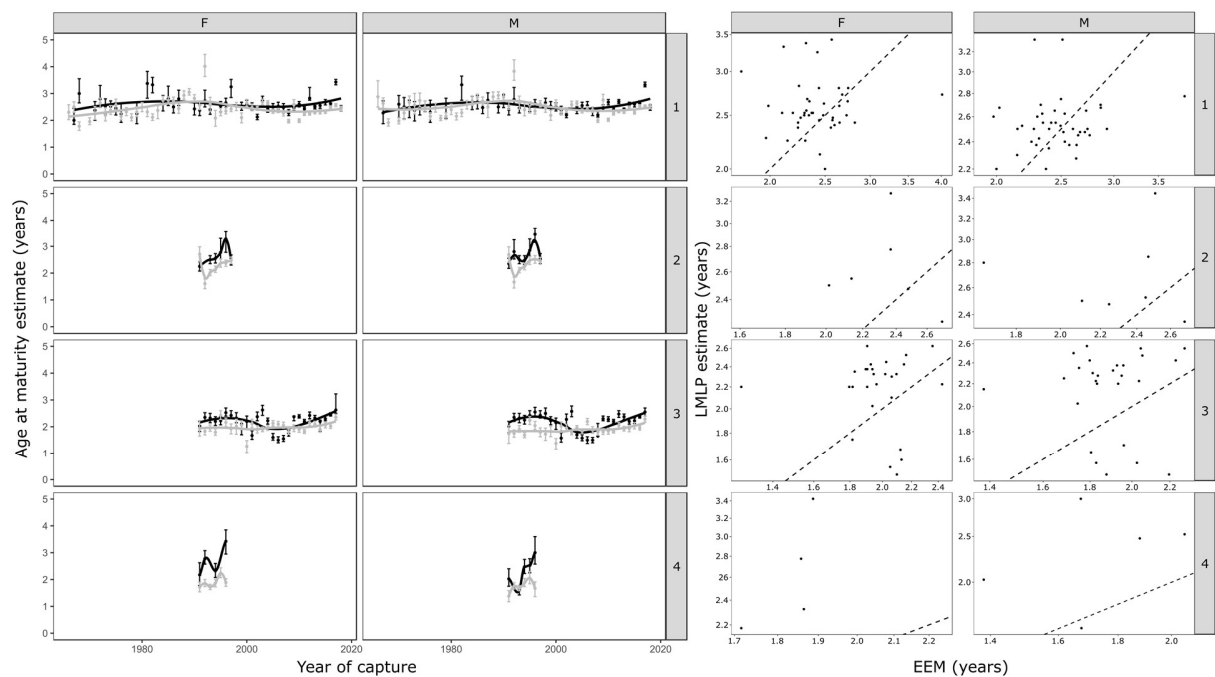s are done for the entire species datasets and for different subsampling combinations generated from the data-filtering of datasets by species and unique combinations of potential factors of interest (subsetting factor column), then divided into variables of interest (*n* and precision). Values represent means and standard errors across all datasets generated by data-filtering all available data by factor combinations for each species (rows denoted by * represent raw values as they are generated from one dataset). Values in brackets represent the number of datasets generated by each subsampling design.

Subsetting factor	Variable	Sex	Atlantic herring m	se	Atlantic cod m	se	Haddock m	se	Plaice m	se	Saithe m	se	Atlantic mackerel m	se	Sprat m	se	Pink ling m	se	Orange roughy m	se	Blue grenadier m	se	Jackass morwong m	se	Tiger flathead m	se
Species	*n* *		146391 (1)		120474 (1)		143995 (1)		74301 (1)		36409 (1)		28007 (1)		29139 (1)		16416 (1)		15249 (1)		40845 (1)		10436 (1)		13544 (1)	
	precision *		10		5		6		6		8		11		8		9		12		10		9		9	
Sex	*n* *	f	76121 (1)		60744 (1)		72354 (1)		42207 (1)		17443 (1)		14265 (1)		16302 (1)		9317 (1)		7609 (1)		20710 (1)		5801 (1)		7869 (1)	
		m	70270 (1)		59730 (1)		71641 (1)		32094 (1)		18966 (1)		13742 (1)		12837 (1)		7099 (1)		7640 (1)		20135 (1)		4635 (1)		5675 (1)	
	Precision *	f	10		5		6		6		8		11		8		9		12		11		9		10	
		m	10		5		6		6		8		11		8		10		13		12		9		11	
Year (of capture)	*n*		2762 (53)	317	2510 (48)	212	3063 (47)	264	2654 (28)	345	984 (37)	128	718 (39)	97	833 (35)	113	513 (32)	60	1386 (11)	277	1459 (28)	127	435 (24)	48	616 (22)	83
	precision		10	0	5	0	6	0	6	0	10	1	14	0	9	0	12	0	15	1	14	0	13	0	11	0
Zone (of capture)	*n*		14639 (10)	2790	12047 (10)	2729	15999 (9)	5486	7430 (10)	1559	4551 (8)	3646	3112 (9)	967	2914 (10)	699	1642 (10)	575	2178 (7)	944	4538 (9)	2842	949 (11)	399	2257 (6)	1301
	precision		10	1	5	0	6	0	6	0	8	0	13	1	9	1	15	2	18	1	13	1	14	1	10	1
Quarter (of capture)	*n*		36598 (4)	17392	30119 (4)	14707	35999 (4)	18062	18575 (4)	10645	9102 (4)	4492	7002 (4)	4530	7285 (4)	4775	4104 (4)	447	3812 (4)	2552	10211 (4)	3169	2609 (4)	328	3386 (4)	732
	precision		11	0	5	0	6	0	7	1	9	0	13	1	10	1	9	0	14	1	11	0	10	1	9	0
Month (of capture)	*n*																1368 (12)	158	1271 (12)	711	3404 (12)	1064	870 (12)	84	1129 (12)	166
	precision																10	0	19	2	11	0	11	1	9	0
Month x Year	*n*																81 (203)	5	492 (31)	98	201 (203)	17	70 (150)	4	94 (144)	5
	precision																16	1	21	2	18	0	20	1	15	1
Month x Zone	*n*																235 (70)	26	635 (24)	246	601 (24)	204	156 (670)	16	251 (54)	43
	precision																13	1	20	1	18	2	17	1	13	1
Sex x Year	*n*	f	1436 (53)	167	1266 (48)	108	1539 (47)	130	1507 (28)	201	485 (36)	61	366 (39)	49	466 (35)	65	291 (32)	33	692 (11)	122	740 (28)	70	242 (24)	32	358 (22)	48
		m	1326 (53)	151	1244 (48)	104	1524 (47)	135	1146 (28)	145	527 (36)	68	352 (39)	48	367 (35)	48	222 (32)	28	695 (11)	161	719 (28)	63	193 (24)	18	258 (22)	37
	precision	f	10	0	5	0	6	0	6	0	11	1	14	0	9	0	12	1	17	1	15	1	14	1	12	1
		m	10	0	5	0	6	0	6	0	11	1	15	1	9	0	14	1	16	1	17	1	14	1	15	1
Sex x Zone	*n*	f	7612 (10)	1420	6074 (10)	1415	8039 (9)	2764	4221 (10)	840	2180 (8)	1758	1585 (9)	489	1630 (10)	396	932 (10)	326	1087 (7)	473	2301 (9)	1443	527 (11)	232	1312 (6)	721
		m	7027 (10)	1373	5973 (10)	1317	7960 (9)	2724	3209 (10)	737	2371 (8)	1888	1527 (9)	479	1284 (10)	304	710 (10)	249	1091 (7)	474	2237 (9)	1403	421 (11)	170	946 (6)	581
	precision	f	10	1	5	0	6	0	7	0	9	1	13	1	10	1	16	4	21	2	13	1	15	1	11	1
		m	10	1	5	0	6	0	7	0	9	1	14	2	9	1	20	3	18	4	15	1	12	0		
Sex x Quarter	*n*	f	19030 (4)	9106	15186 (4)	7315	18089 (4)	9193	10552 (4)	6080	4361 (4)	2148	3566 (4)	2302	4076 (4)	2611	2329 (4)	229	1902 (4)	1248	5178 (4)	1659	1450 (4)	112	1967 (4)	430
		m	17568 (4)	8288	14933 (4)	7396	17910 (4)	8886	8024 (4)	4638	4742 (4)	2346	3436 (4)	2228	3209 (4)	2170	1775 (4)	229	1910 (4)	1305	5034 (4)	1510	1159 (4)	220	1419 (4)	314
	precision	f	11	1	5	0	6	0	8	2	9	0	14	1	10	1	10	0	16	2	11	1	10	1	10	0

		m	11	0	5	0	6	0	7	1	9	0	13	1	10	1	11	0	15	1	12	0	10	1	11	1
Sex x Month	n	f															776 (12)	96	634 (12)	363	1726 (12)	582	483 (12)	44	656 (12)	94
		m															592 (12)	65	637 (12)	350	1678 (12)	485	386 (12)	45	473 (12)	75
	precision	f															10	0	20	2	12	0	11	1	11	0
		m															11	0	20	2	13	0	11	1	12	0
Year x Zone	n		335 (437)	16	297 (405)	14	395 (365)	25	300 (248)	18	199 (183)	31	124 (225)	9	132 (220)	7	149 (110)	14	726 (21)	85	382 (107)	42	137 (76)	17	218 (62)	26
	precision		12	0	7	0	8	0	8	0	13	1	17	0	11	0	15	1	18	1	18	1	19	1	13	0
Year x Quarter	n		335 (100)	16	297 (96)	14	395 (94)	25	300 (56)	18	199 (82)	31	124 (78)	9	132 (58)	7	163 (101)	13	803 (19)	182	425 (96)	36	141 (74)	12	218 (62)	16
	precision		12	0	7	0	8	0	8	0	13	1	17	0	11	0	14	0	17	1	16	0	17	1	13	1
Quarter x Zone	n		3957 (37)	698	3256 (37)	627	4000 (36)	1052	2562 (29)	499	1300 (28)	679	903 (31)	264	1166 (25)	252	586 (28)	93	953 (16)	398	1459 (28)	600	337 (31)	65	616 (22)	179
	precision		12	0	6	0	7	0	7	1	11	1	15	1	12	1	13	1	18	1	18	4	16	1	11	1
Sex x Year x Zone	n	f	175 (436)	8	150 (405)	7	206 (352)	13	171 (247)	10	108 (161)	17	64 (222)	4	74 (220)	4	85 (109)	8	362 (21)	43	194 (107)	23	76 (76)	11	127 (62)	15
		m	161 (436)	8	148 (404)	7	201 (356)	13	129 (248)	8	112 (170)	17	63 (219)	4	59 (218)	3	68 (105)	7	364 (21)	48	190 (106)	21	63 (74)	7	93 (61)	12
	precision	f	12	0	7	0	8	0	9	0	14	1	18	0	12	0	17	1	21	1	20	1	20	1	15	1
		m	12	0	7	0	8	0	9	0	14	1	18	0	11	0	20	1	20	1	22	1	21	1	17	1
Sex x Year x Quarter	n	f	761 (100)	56	633 (96)	30	770 (94)	43	754 (56)	74	215 (81)	21	183 (78)	23	281 (58)	28	93 (100)	7	400 (19)	88	218 (95)	19	79 (73)	7	127 (62)	10
		m	703 (100)	50	622 (96)	30	762 (94)	45	573 (56)	54	234 (81)	23	176 (78)	23	221 (58)	22	71 (100)	6	402 (19)	98	210 (96)	19	64 (72)	6	93 (61)	7
	precision	f	12	0	6	0	7	0	7	0	11	0	16	0	10	0	15	1	20	1	18	0	18	1	15	1
		m	12	0	6	0	7	0	7	0	11	0	16	0	10	0	17	1	19	1	21	1	19	1	17	1
Sex x Year x Month	n	f															47 (200)	3	245 (31)	50	103 (202)	9	39 (148)	3	55 (144)	3
		m															37 (190)	3	246 (31)	50	100 (201)	9	32 (145)	2	40 (142)	3
	precision	f															18	1	23	1	20	1	20	1	18	1
		m															19	1	22	2	22	1	22	1	19	1
Sex x Zone x Quarter	n	f	2057 (37)	362	1642 (37)	318	2010 (36)	534	1455 (29)	277	646 (27)	339	460 (31)	136	652 (25)	139	333 (28)	52	476 (16)	196	740 (28)	307	187 (31)	38	358 (22)	99
		m	1899 (37)	337	1614 (37)	310	1990 (36)	519	1146 (28)	235	677 (28)	352	458 (30)	132	535 (24)	116	254 (28)	42	478 (16)	203	746 (27)	304	150 (31)	29	270 (21)	85
	precision	f	12	0	6	0	8	1	9	1	11	1	15	1	12	1	16	2	22	1	18	3	17	1	13	1
		m	12	0	6	0	8	0	8	1	11	1	15	1	12	1	17	2	21	1	18	2	19	2	13	1
Sex x Month x Zone	n	f															133 (70)	14	317 (24)	122	305 (68)	109	87 (67)	10	146 (540	24
		m															104 (68)	12	318 (24)	126	301 (67)	97	71 (65)	7	109 (52)	20
	precision	f															17	2	22	1	18	1	18	1	15	1
		m															17	1	21	1	19	1	20	1	16	1
Year x Month x Zone	n																54 (304)	3	401 (38)	58	125 (327)	10	50 (207)	3	58 (232)	3
	precision																18	1	21	1	20	0	21	1	17	0
Sex x Year x Zone x Month	n	f															31 (298)	2	200 (38)	31	64 (323)	5	29 (203)	2	34 (231)	2
		m															26 (273)	2	201 (38)	30	63 (318)	5	23 (200)	2	25 (235)	2
	precision	f															21	1	24	1	22	1	22	1	19	1
		m															21	1	23	1	23	1	24	1	21	1
Sex x Year x Zone x Quarter	n	f	99 (767)	3	79 (772)	3	113 (640)	5	105 (403)	4	58 (303)	6	41 (351)	2	47 (346)	2	44 (212)	3	262 (29)	38	91 (227)	8	42 (138)	3	59 (134)	4
		m	91 (768)	3	78 (769)	2	111 (643)	5	80 (401)	3	60 (316)	7	39 (348)	2	38 (340)	2	36 (199)	3	263 (29)	39	89 (226)	8	34 (135)	3	43 (131)	3
	precision	f	13	0	8	0	9	0	10	0	15	1	20	1	14	0	20	1	22	1	22	1	20	1	17	1
		m	14	0	8	0	9	0	9	0	15	0	19	0	13	0	20	1	21	1	23	1	23	1	19	1

70

71

Supplementary Figure 2. Full comparison of statistical estimates of maturity (LMLP) against
empirical estimates of maturity (EEM)(on raw maturity data) for Atlantic herring as supplement
to analyses presented in Table 1. Left two plots on each row represent LMLP estimates (black
points with black loess smoothers) through years plotted against matched EEM estimates (grey
points and smoothers) for males and females. Right two plots show standardised major axis
regression analyses where LMLP and EEM estimates are directly compared where black dashed
lines represent a 1:1 relationship. Differences in temporal replication of LMLP and EEM arise
from instances of poor model fits to data in LMLP or non-convergence in EEM for those years
which were subsequently excluded from regression analyses (see methods). Each row of plots
depicts individual Quarters (of capture, within Years) represented by numbers on the right side
of plot pairs. Note a log-transformation in axes scale in the right two columns of plots.
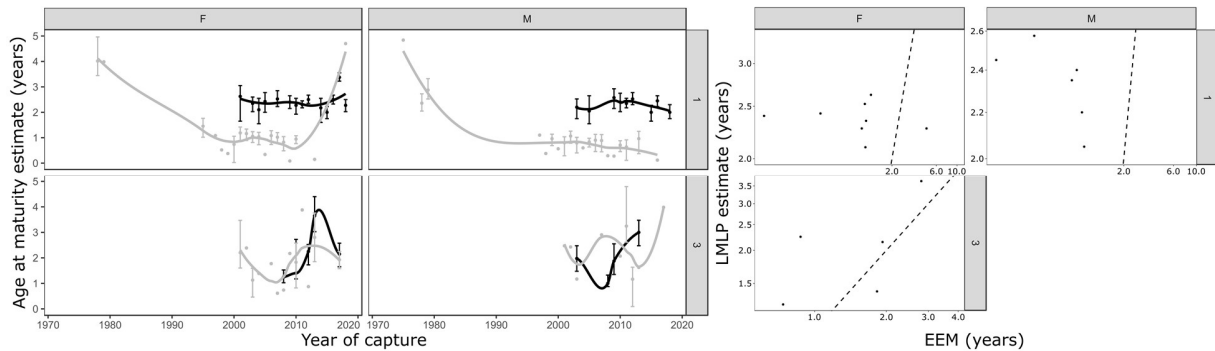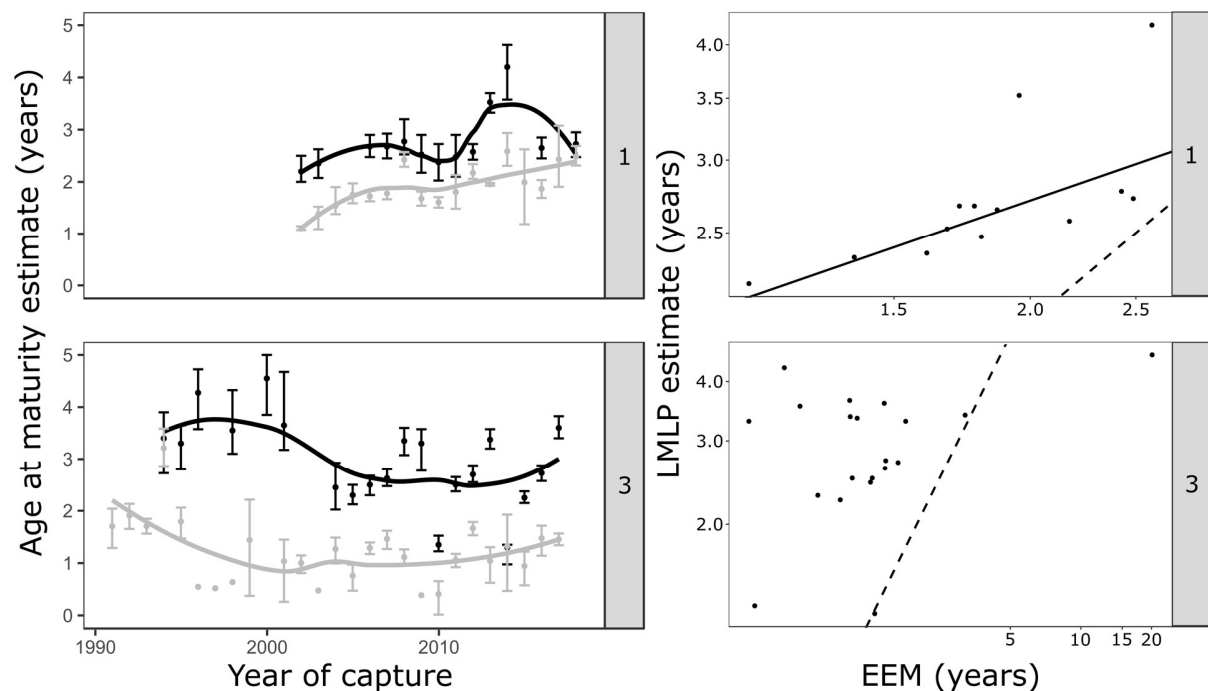
83

84
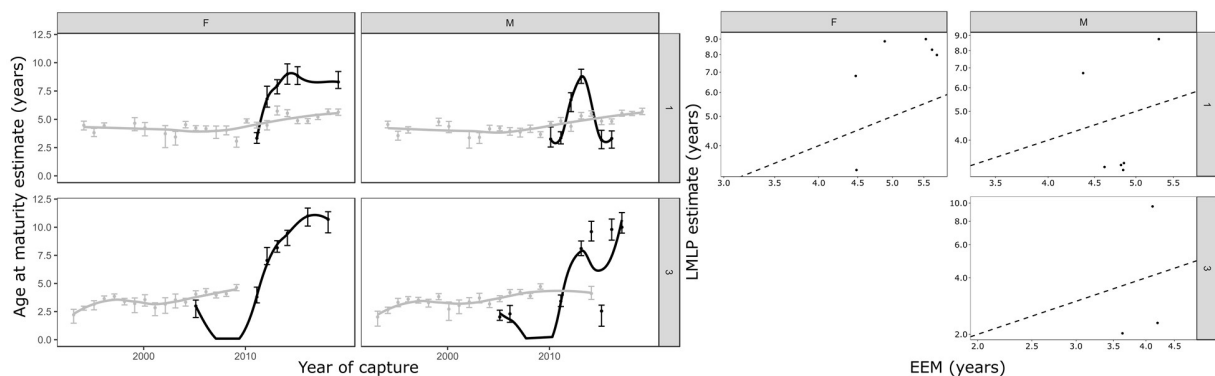
85   Supplementary Figure 3. Full comparison of statistical estimates of maturity (LMLP) against
86   empirical estimates of maturity (EEM)(on raw maturity data) for saithe as supplement to
87   analyses presented in Table 1. Note that one regression was not possible due to low replication.
88   For further details see legend of figure S1.
89

90

91   Supplementary Figure 4. Full comparison of statistical estimates of maturity (LMLP) against
92   empirical estimates of maturity (EEM)(on raw maturity data) for sprat as supplement to analyses
93   presented in Table 1. Note that one regression was not possible due to low replication. Very wide
94   error bars or those spanning 0 are not shown to improve figure clarity. For further details see
95   legend of figure S1.
96

97

98   Supplementary Figure 5. Full comparison of statistical estimates of maturity (LMLP) against
99   empirical estimates of maturity (EEM)(on raw maturity data) for Atlantic mackerel as
100   supplement to analyses presented in Table 1. Each row of plots depicts individual Quarters (of
101   capture, within Years) represented by numbers on the right side of plots. For further details see
102   legend of figure S1.

103

104

Further exploration of LMLP performance

106

107 Here, additional investigation of statistical methods to estimate maturity (LMLP) presented in

108 Honsey et al. (2017) were conducted in response to the general underperformance of the

109 technique under fisheries datasets. A number of modified analyses, as suggested by Honsey at al.

110 (2017), are presented below.

111

112 Honsey et al. (2017) data selection methodology analyses

113

114 Data selection methodologies were matched as closely as possible to those employed in Honsey

115 et al. (2017) as justified to avoid any possible erroneous age estimation in datasets and increase

116 LMLP method accuracy as a result. Three species from the North Sea were analysed and include

117 Atlantic cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*) and plaice (*Pleuronectes*

118 *platessa*). Datasets were modified to only include the Quarter (of capture) directly preceding the

119 month of otolith increment formation (3rd quarter for all species; personal communication (Cefas,

120 UK)). Secondly, unsexed individuals smaller than the smallest mature male were selected and

121 randomly allocated to male or female proportions of each respective dataset. Males were then

122 excluded, and LMLP analyses were conducted on the resulting datasets. LMLP analyses were

123 then compared to EEM analyses across years (of capture) including samples from either the

124 whole year (hereafter; EEM (whole year)) or the 3rd quarter of each year (hereafter; EEM (3rd

125 quarter)) using standardised major axis regression analyses as presented in the main manuscript

126 (see methods for specific descriptions of LMLP, EEM and standardised major axis regression

127 analyses). These comparisons were designed to assay the performance of LMLP analyses against

128 EEM analyses in a general sense (EEM (whole year)) and also more specifically (EEM (3rd

129 quarter)). LMLP and EEM estimates were uncorrelated across all analyses (across species and

130 the differing EEM analyses (LMLP x EEM (whole year) or LMLP x EEM (3rd quarter)) but

131 differing EEM (EEM (whole year) x EEM (3rd quarter)) analyses were correlated for each

132 species (Supplementary Table 2, Supplementary Figure 5).

133

134 Supplementary Table 2. Results of standardised major axis regression analyses comparing
135 statistical estimates of age at maturity (LMLP) to empirical estimates (EEM)(logistic regression
136 applied to raw maturity data) for the Quarter preceding the month of otolith increment formation
137 across Years for females in each Species. LMLP is compared to EEM estimated from both whole
138 Years of data and also the Quarter matched to that used in LMLP estimation (3rd Quarter in all
139 cases). EEM (3rd Quarter) vs EEM (whole year) regressions are included for comparative
140 purposes. Parameter estimates for the regression intercept and slope with accompanying 95%
141 confidence intervals are also reported. Perfect equivalence between estimates would result in an
142 R^2 value of 1 where slope is 1 and the intercept is 0.

143

Species	Sex	Comparison	Intercept	Slope	R^2	P-value
Atlantic cod	f	LMLP vs EEM (whole year)	-0.355 (-0.817, 0.107)	2.023 (1.264, 3.238)	0.034	0.478
	f	LMLP vs EEM (3rd quarter)	0.015 (-0.293, 0.322)	1.345 (0.811, 2.231)	0.024	0.555
	f	EEM (whole year) vs EEM (3rd quarter)	0.223 (0.152, 0.293)	0.531 (0.394, 0.716)	0.573	<0.001
Haddock	f	LMLP vs EEM (whole year)	-0.546 (-0.968, -0.123)	3.105 (1.994, 4.835)	0.050	0.281
	f	LMLP vs EEM (3rd quarter)	0.751 (0.504, 0.998)	-1.488 (-2.618, -0.845)	0.207	0.057
	f	EEM (whole year) vs EEM (3rd quarter)	0.180 (0.131, 0.228)	0.368 (0.238, 0.569)	0.285	0.015
Plaice	f	LMLP vs EEM (whole year)	1.035 (0.754, 1.316)	-1.058 (-1.842, -0.607)	<0.001	0.956
	f	LMLP vs EEM (3rd quarter)	0.263 (0.068, 0.457)	0.655 (0.365, 1.177)	0.042	0.398
	f	EEM (whole year) vs EEM (3rd quarter)	0.079 (-0.042, 0.201)	0.777 (0.571, 1.057)	0.315	0.003

144

145

146

147
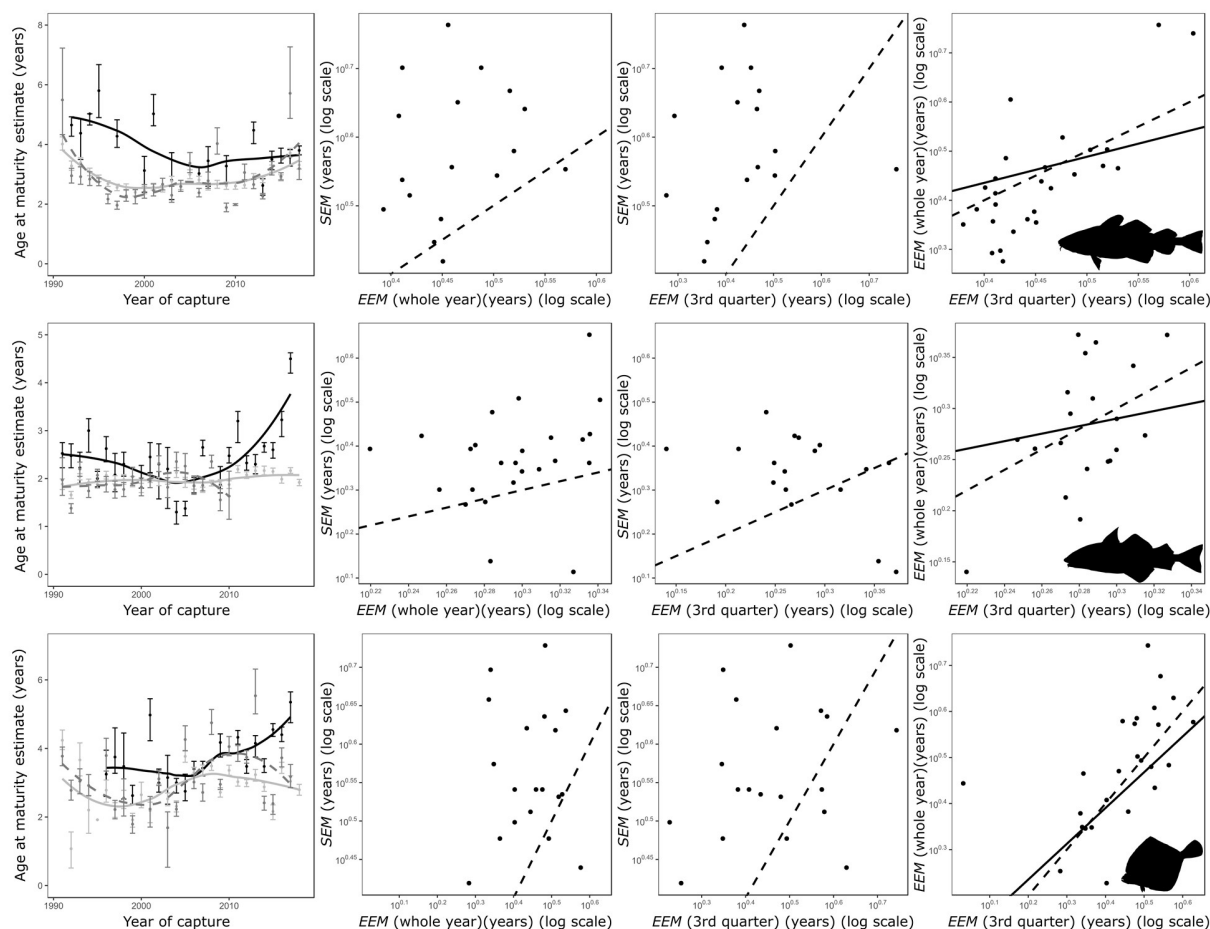
Supplementary Figure 6. Comparisons of LMLP against EEM for three North Sea species using data selection methodologies from Honsey et al. (2017). Each row of plots represents separate species; the first row - Atlantic cod, second row – haddock, third row - plaice. The first plot on each row represent LMLP estimates (black points with black loess smoothers) through years plotted against matched EEM estimates where either the whole year (light grey points with smoothers) or the third quarter (dark grey points with smoothers) of data is included (note differences in y-axis scale). The right three plots on each row show standardised major axis regression analyses where LMLP and EEM estimates are compared for each Year show (second plot - LMLP vs EEM whole year, third - LMLP vs EEM 3rd quarter and fourth plot - EEM whole year vs EEM 3rd quarter). For standardised major axis regression plots, black dashed lines represent a 1:1 relationship and solid black lines represent significant correlations between variables.

Modified analyses to encompass differences in interpretation between LMLP and EEM

12

163   LMLP and EEM were also compared after subtracting 1 year from all EEM estimates. This was

164   done because LMLP estimates predict the point at which populations start investing in

165   reproduction while logistic regressions of raw maturity data predict maturity from visual

166   inspection of ripe gonads, which occurs later than the onset of reproductive investment. Despite

167   this EEM adjustment, LMLP estimates still did not show any increase in correlations to EEM

168   estimates (Supplementary Table 3) and thus, there was no evidence that LMLP was consistently

169   biased and the performance did not improve if we corrected for the possible underestimation of

170   maturation by LMLP.

171

172 Supplementary Table 3. Results of standardised major axis regression analyses comparing
173 statistical estimates of age at maturity (LMLP)(Honsey et al. 2017 model) to adjusted empirical
174 estimates (EEM - *1*)(logistic regression applied to raw maturity data) across Years for each
175 Species, Sex (excluding Atlantic mackerel) and Quarter (within Year) combination. Parameter
176 estimates for the regression intercept and slope with accompanying 95% confidence intervals are
177 also reported. Perfect equivalence between LMLP and EEM-1 estimates would result in an R^2
178 value of 1 where slope is 1 and the intercept is 0.

Species	Sex	Quarter	Comparison	Intercept	Slope	R^2	P-value
Atlantic herring	F	1	LMLP x (EEM -1)	0.329 (0.295, 0.363)	0.527 (0.367, 0.758)	<0.001	0.901
	F	2	LMLP x (EEM -1)	0.502 (0.347, 0.657)	-0.688 (-2.168, -0.218)	0.026	0.760
	F	3	LMLP x (EEM -1)	0.354 (0.320, 0.387)	0.860 (0.530, 1.396)	0.002	0.837
	F	4	LMLP x (EEM -1)	0.590 (0.212, 0.968)	2.066 (0.555, 7.691)	0.502	0.292
	M	1	LMLP x (EEM -1)	0.343 (0.320, 0.367)	0.347 (0.244, 0.492)	0.021	0.367
	M	2	LMLP x (EEM -1)	0.472 (0.375, 0.568)	-0.444 (-1.357, -0.145)	0.002	0.928
	M	3	LMLP x (EEM -1)	0.280 (0.228, 0.331)	-1.004 (-1.593, -0.633)	0.001	0.864
	M	4	LMLP x (EEM -1)	0.457 (0.210, 0.703)	0.628 (0.208, 1.891)	0.132	0.548
Saithe	F	1	LMLP x (EEM -1)	-0.080 (-1.161, 1.000)	1.554 (0.594, 4.061)	0.432	0.156
	M	1	LMLP x (EEM -1)	3.816 (-1.931, 9.563)	-5.624 (-21.402, -1.478)	0.052	0.665
	M	3	LMLP x (EEM -1)	-3.326 (-34.580, 27.927)	8.204 (0.510, 132.032)	0.204	0.702
Sprat	F	1	LMLP x (EEM -1)	0.336 (0.229, 0.442)	-0.042 (-0.140, -0.013)	0.005	0.909
	F	3	LMLP x (EEM -1)	0.293 (-0.302, 0.889)	1.142 (0.192, 6.788)	0.897	0.208
Atlantic mackerel	-	1	LMLP x (EEM -1)	0.436 (0.414, 0.459)	0.171 (0.100, 0.294)	0.385	0.024
	-	3	LMLP x (EEM -1)	0.539 (0.452, 0.627)	0.140 (0.075, 0.263)	0.060	0.397

179

180

181

182   LMLP error characterisation

183

184   To ensure our investigation of the LMLP method was exhaustive, LMLP estimates were judged

185   against error characterisation methodologies suggested by Honsey et al. (2017). The authors

186   present error characterisation as a series of plots where individual LMLP estimates can be

187   checked against minimum data amounts, precision and model-estimated g for three levels of age

188   at maturity to ascertain confidence in LMLP estimates. Honsey et al. (2017) generated error

189   characterisation using simulated datasets with simulated fishery selection applied and found

190   increasing data quality requirements for earlier-maturing simulated populations. For analyses

191   presented here, each LMLP estimate was judged against the nearest lower contour (defined by g

192   estimated by the model) and excluded from subsequent regression analyses if it fell below the

193   contour (see; Honsey et al. 2017). Standardised major axis regression analyses were then

194   performed on all remaining LMLP estimates where comparisons to empirical estimates (EEM)

195   across all Years for each Species, Sex and Quarter (within Year). Regression analyses that were

196   not modified by error characterisation (as all analyses fell above relevant error contours) are not

197   presented here as they unmodified from those presented in Table 2 within the manuscript. Of 6

198   possible comparisons modified by the exclusion of individual LMLP estimates that failed error

199   characterisation (all other analyses were unmodified and so are equivalent to those presented in

200   Table 2), only male saithe datasets from Quarter 1 showed a significant relationship between

201   LMLP and EEM estimates (Supplementary Table 4, Supplementary Figures 1, 2, 3 and 4).

202

203   Supplementary Table 4. Standardised major axis regression analyses where statistical estimates
204   of age at maturity (LMLP) have been validated against error characterisation methods presented

205 in Honsey et al. (2017) before comparison with empirical estimates (EEM)(logistic regression
206 applied to raw maturity data) across Years for each Species, Sex (excluding Atlantic mackerel)
207 and Quarter (within Year) combination. Exclusion of LMLP estimates that fell under error
208 contours rendered some analyses unsuitable as data amounts were too low. Adjusting maturity
209 estimates to match interpretation of the LMLP parameter (analogous to methods used to generate
210 Supplementary Table 2) had no effect on significance on results and are not presented here.
211 Excluding analyses that fell below error contours for simulated data where maturity = 3 but
212 above error contours for simulated data where maturity = 5 (this only occurred in 8 saithe
213 datasets) also had no effect on the significance of standardised major axis regression analyses
214 and so these analyses are not presented. Parameter estimates for the regression intercept and
215 slope with accompanying 95% confidence intervals are also reported. Perfect equivalence
216 between LMLP and EEM estimates would result in an R^2 value of 1 where slope is 1 and the
217 intercept is 0.

218

Species	error characterisation maturity value	Sex	Quarter	Intercept	Slope	R2	P-value
Saithe	5	f	1	-0.503 (-2.056, 1.050)	1.941 (0.741, 5.081)	0.428	0.158
	5	m	1	5.409 (-3.163, 13.980)	-7.123 (-27.119, -1.871)	0.056	0.652
	3	m	1	-6.625 (-10.310, -2.940)	10.446 (6.424, 16.985)	0.997	0.036
Sprat	3	f	1	0.336 (0.172, 0.499)	1.412 (0.331, 6.028)	0.329	0.427
	3	f	3	-0.336 (-2.520, 1.849)	2.060 (0.327, 12.979)	0.883	0.222
Atlantic mackerel	3	-	3	0.569 (0.461, 0.677)	-0.834 (-1.527, -0.455)	0.014	0.695

219

220

221

References

223

224 Caton A, McLaughlin K (2000) Fishery status reports 1999. Resource assessments of Australian
225   commonwealth fisheries. Bureau of Rural Sciences, Kingston, ACT, Australia.
226 Fairbridge WS (1951) The New South Wales tiger flathead, *Neoplatycephalus macrodon*
227   (Ogilby). I. Biology and determination. Aust J Mar Freshw Res 2:117-178
228 Fenton GE, Short SA, Ritz DA (1991) Age-determination of orange roughy, *Hoplostethus*
229   *atlanticus* (Pisces, Trachichthyidea) using Pb-210-Ra-226 disequilibria. Mar Biol
230   109:197-202. doi:10.1007/bf01319387
231 Froese R, Sampang A (2013) Potential Indicators and Reference Points for Good Environmental
232   Status of Commercially Exploited Marine Fishes and Invertebrates in the German EEZ.
233   World Wide Web electronic publication
234 Hislop JRG, Shanks AM (1981) Recent investigations on the reproductive biology of the
235   haddock, *Melanogrammus aeglefinus*, of the northers North Sea and the effects on

236 fecundity of infection with the copepod parasite *Lernaeocera branchialis*. Journal Du
237   Conseil 39:244-251
238 Honsey AE, Staples DF, Venturelli PA (2017) Accurate estimates of age at maturity from the
239   growth trajectories of fishes and other ectotherms. Ecol Appl 27:182-192.
240   doi:10.1002/eap.1421
241 ICES (1996) Report of the Working Group on Assessment of Demersal Stocks in the North Sea
242   and Skagerrak. ICES HQ. ICES CM. .
243 ICES (2005) Report of the Working Group on the Assessment of Mackerel, Horse Mackerel,

244 Sardine and Anchovy (WGMHSA), 6 - 15 September 2005. Vigo, Spain. 615 pp.
245 ICES (2015) Manual for the International Bottom Trawl Surveys. Series of ICES Survey
246   Protocols SISP 10 - IBTS IX. 86 pp
247 ICES (2017) Report of the Working Group on Assessment of Demersal Stocks in the North Sea
248   and Skagerrak (2017). ICES HQ. ICES CM. 2017/ACOM:21. 1248 pp.
249 Jordan AR (1999) The life-history ecology of *Platycephalus bassensis* and *Nemadactylus*
250   *macropterus*. PhD Thesis, University of Tasmania.
251 Knuckey I, Grieve C, Smith DC (1999) Evolution of the integrated scientific monitoring
252   programme in Australia's South East Fishery. Proceedings of the International
253   Conference on Integrated Fisheries Monitoring, Sydney, Australia, February 1–5 1999.
254   FAO, Rome, Italy.
255 Lockwood SJ, Shepherd JG (1984) An assessment of the Western mackeral stock. Journal Du
256   Conseil 41:167-180
257 Mace PM, Fenaughty JM, Coburn RP, Doonan IJ (1990) Growth and productivity of orange
258   roughy (*Hoplostethus atlanticus*) on the north Chatham Rise. N Z J Mar Freshwat Res
259   24:105-119. doi:10.1080/00288330.1990.9516406
260 Russel S, Smith DC (2007) Spawning and reproduction biology of blue grenadier in south-
261   eastern Australia and the winter spawning aggregation off western Tasmania. Final report
262   to Fisheries Research and Development Corporation Project No 2000/102 Fisheries
263   Research Branch, Fisheries Victoria, Queenscliff
264 Smith ADM, Wayte SEAFMA (2001) The South east fishery 2000. South East Fishery
265   Assessment Group. (2001). Australian Fisheries Management Authority., Barton, A.C.T.
266   Australia.
267 Tilzey RDJ (1994) The South East Fishery: A scientific review with reference to quota
268   management. Commonwealth of Australia, Canberra, Australia.

269