Sample Article for Oxford University Press Journals

Lisa Chong

Leibniz Centre for Tropical Marine Research (ZMT), Bremen, Germany lisa.chong8594@gmail.com

Second Author

Institute or Organization, Department, City, State,

Zip Code, Country e-mail address

November 26, 2018

1. Abstract

Length data are quickly and easily collected and relatively economical, and thus are often the primary data source collected in data-limited fisheries. As there are several length-based methods currently being applied, a performance evaluation of length-based methods is instrumental to sort out conflicting information to capture the degree of uncertainty and to determine their robustness and sensitivities relative to each other. In the paper presented here, we conducted a simulation-estimation analysis to comparatively analyse the following methods: Length-Based Spawning Potential Ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), Length-Based Risk Analysis, Thompson and Bell (TB), and Length-Based Reference Points (LBRP). The analysis involved testing how the methods performed under conditions of varying longevity, exploitation status, and recruitment types. In similar studies, the operating models used were identical to the assessment models, whereas in this study, we used an alternative structure operating model, an individual based model, to avoid problems involving the design of the assessment model.

This study highlights the importance of capturing uncertainty in estimating stock status and population dynamics to test the strengths and weaknesses of different length-based assessment methods.

Keywords: word1; word2; word3; and word4

1. Introduction

Measuring uncertainty in stock assessment is integral in giving advice for fish stocks as it allows fisheries managers to weigh the benefits and losses of different strategies (Rosenberg and Restrepo, 1994). Testing these stock assessment models in various fisheries scenarios gives understanding of their behaviour. This allows stock assessment scientists and managers to examine issues associated with data collection and availability, model misspecifications and stochasticity in population dynamics. In a data-rich stock, more data and information are available that can help eliminate implausible scenarios. For data-limited stocks, this is often not the case and thus conducting a stock assessment is significantly more challenging in these fisheries.

Fisheries are considered data-limited if the available scientific information (typically catches and/or length compositions) is inadequate for determining current stock status with respect to meaningful reference points (Richards and Maguire, 1998; Pilling *et al.*, 2008; Dowling *et al.*, 2015). In data-limited fisheries, length data is often the primary data type collected as it is relatively economical and easy to collect (Pilling *et al.*, 2008; Hordyk *et al.*, 2015a; Mildenberger *et al.*, 2017). As a result, length-based methodologies have been developed and applied to estimate biological stock characteristics, fisheries performance, and stock status (Pauly andMorgan, 1987; Beverton and Holt, 1993; Pilling *et al.*, 2008; Mildenberger *et al.*, 2017).Prominent length-based methods include the Length-based Thompson-and-Bell model (TB; Thompson and Bell, 1934), Length-Based Reference Points (LBRP) (Froese, 2004; Cope and Punt, 2009), Length-Based Risk Analysis (Ault *et al.*, 1998, 2008, 2018), Length-Based Spawning Potential Ratio (LBSPR) and Length-based Integrated Mixed Effects (LIME).

TB is a yield per recruit model that evaluates the stock’s status relative to reference levels and the impact of a certain management control measure (Mildenberger *et al.*, 2017). LBRP is an extension of Froese’s (2004) metrics based on well-established relationships between fisheries management and life history theory, and further explores how these metrics are related to fishing mortality, spawning biomass and current reference points based on spawning biomass (Cope and Punt, 2009). Ault et al. (1998, 2008, 2018) developed the Length-based Risk Analysis, assuming fishing mortality influences mean length of the catch, deriving the Spawning Potential Ratio (SPR) as a reference point amongst other derived quantities. SPR is defined as the proportion of the unfished reproductive potential left ay any given level of fishing pressure (Hordyk *et al.*, 2015b). The SPR equals 100% in an unexploited stock, and 0% in a stock with no spawning (e.g. all mature fish have been removed or all female fish have been caught). LBSPR uses length composition data to derive SPR from Beverton-Holt life history invariants, including the ratio of natural mortality and the von Bertalanffy growth coefficient (*M/K*). Finally, LIME accounts for time-varying recruitment and fishing mortality and derives population parameters associated with an age-structured model, including SPR. Each of these length-based methods have been applied to assess and manage fish stocks.

The above-mentioned methods are all being used by different authors and in different contexts, but a comprehensive performance evaluation is needed to evaluate which methods perform best in various circumstances. Therefore, the primary objective is to find the best methods for each configuration of data scenario of target stocks analysed. Most of these data-limited length-based methods are being used in management and have been tested given different fisheries scenarios or against another method (e.g. LIME vs LBSPR, Rudd and Thorson, 2018). However, many of these methods have either never been tested thoroughly, only tested against one other method, or the testing was done by the author of the method often via matching the operating and estimation model. A fundamental next step in the stock assessment process is to determine conflicting information and capture the degree of uncertainty in assessment work (Cadrin and Dickey-Collas, 2015). Misinterpreting error and uncertainty in model outputs could lead to misleading assessment interpretations and misinformed decisions. Therefore, performance evaluations of interpreting length-based methods are essential to identify practical decisions in data-limited fisheries.

Here we analyse the five length-based methods through a simulation-estimation analysis to test their performance with scenarios differing in fish longevity, exploitation level and recruitment type. The results are expected to reveal the strengths and weaknesses of each method with reference to how well they capture the stocks’ status and estimate the parameters, which could help managers and stock assessment scientists determine which methods to apply. Evaluating the performance of these methods will promote further development of data-limited approaches that will be able to better capture the fishery status and understand discrepancies in the performance of the methods (Cadrin and Dickey-Collas, 2015).

1. Methods

We conducted a simulation-estimation analysis where the operating model was an individual-based population model (IBM) (Cao *et al.*, 2016) that simulated population dynamics and generated length composition data. Many operating models are identical to the assessment models, and assume all dynamic processes are fully understood. Therefore, using an alternative structure operating model can help avoid this problem and identify misleading assumptions that may be implicit in the design of an assessment model (Cao *et al.*, 2016). The “true” input parameters for this study were used in the operating models and then assumed known in the length-based models used as estimation models. This simulation loop allows us to compare how far the outputs of the assessment models are from the “true values and investigate the sensitivities of the models.

* 1. *Operating model*

The stock dynamics were simulated using the “fishdynr” R package (Taylor, 2017), which can be used to describe fish growth, mortality and recruitment, and contains several models for simulating stock or population dynamics and management. The function “virtualPop” creates an IBM of a fish stock with certain life history traits subjected to a fishing fleet with specific selectivity characteristics. Information about the modelling approach for growth, mortality, and recruitment are outlined by Taylor and Mildenberger (2017). A burn-in period with no fishing activity of 10 years was simulated in all iterations and scenarios as it takes about 5-10 years for the IBM to reach equilibrium. Twenty-five additional years were simulated in the IBM, but only one year of monthly data with 200 individuals per month was extracted for the length frequency data at the end of the simulation period (year 34). This reflects a one-year field phase to record landings and is common in tropical artisanal fisheries (Tesfaye *et al.*, 2016; Herrón *et al.*, 2018; Tuda, 2018). For each scenario, 300 iterations were simulated, and thus 300 length frequency data sets generated. All operating models assumed von Bertalanffy growth and logistic- type selectivity and maturity. Equations for the operating model are listed in Table 1.

Seven scenarios were simulated based on variations in fish longevity, fishing exploitation levels, and recruitment: (1) The base model was comprised of a medium-lived species (18 years in this study), an exploitation rate at the target level of SPR a population at target level SPR (SPR ≈ 40%), and Beverton-Holt recruitment (Beverton and Holt, 1993) with moderate statistical error (*σR* = 0.5). From this base model, a we conducted several sensitivities that changed one of the three parameters - fish longevity, current exploitation status, and recruitment: (2) a short-lived (4 years) and (3) longer-lived (26 years) species; (4) a state of overexploitation (SPR ≈ 20%) and (5) a state of underexploitation (SPR ≈ 70%) in terms of SPR; and (6) equilibrium (no recruitment error) and (7) variable and autocorrelated recruitment (autocorrelated error). The fish longevity traits were simulated based on the following three fish stocks: *Siganus sutor* for the short-lived (Hicks and McClanahan, 2012), *Lutjanus guttatus* for the medium-lived (Bystrom, 2016), and *Epinephelus morio* for the longer-lived (Heemstra and Randall, 1993). All simulations and analyses were conducted using the statistical programming language R (R Core Team, 2018).

The length frequency distributions of each fish longevity scenario for one iteration are shown in Figure 1, with the *L∞*, , and values visualised. The input values of the seven operating models are listed in Table 2. The same life history values, *L∞*, *K*, *M*, and , that were used in the operating models were applied to each of the five length-based methods. The overview of this study is depicted in Figures 2 and 3.

* 1. *Estimation models*

The estimation models refer to the length-based methods that derive estimates of stock status from stimulated data. Five length-based methods were analysed using the length data created from the operating models: (1) TB, (2) LBRP, (3) Length-based Risk Analysis, (4) LBSPR and (5) LIME. These methods are contained within the R packages TropFishR (Mildenberger *et al.*, 2017), LBSPR (Hordyk *et al.*, 2015b), LIME (Rudd and Thorson, 2018), and fishmethods (Nelson, 2017).

* + 1. *TropFishR – TB*

The TropFishR routine consists of three models: Length-Converted Catch Curve (LCCC), Cohort Analysis and TB. The results of the TB are the final step in the entire routine. This routine requires several mortality and selectivity parameters: fishing mortality for each length bin (termed F- at-length-array), and the fishing mortality of the plus group, which comprises the largest length class (termed terminal F) as inputs. The LCCC provides a terminal F value, then the Cohort Analysis calculates the F-at-length-array, and TB ultimately gives the estimates of yield-per-recruit under scaled F levels with a known pattern of F-at-age. The selectivity values, and , from the LCCC were also applied to the applicable methods to keep the selectivity inputs consistent. Yield per recruit (YPR) models evaluate fishing reference points in relation to fisheries reference levels and help determine control measures, such as changing fishing effort or regulation gear types (Sparre *et al.*, 1998). These methods also estimate expected yields and biomass per recruit from a cohort subjected to varying levels of fishing and length at first capture (*Lc*). TB requires as inputs length composition, the von Bertalanffy growth parameters (*L∞*, *K*, *t0*), F-at-length-array, exploitation rate (*E*), natural mortality (*M*), terminal F, and length at first capture (in this case *Lc* = due to knife-edge selectivity). The assumptions of this model are that (i) the fishery is in equilibrium; (ii) growth is described by the von Bertalanffy function, (iii) natural mortality is constant; and (iv) the inputs are from the Cohort Analysis. TB gives an output of *F*/*F*0.1. *F*0.1 is the fishing rate at which the slope of the yield per recruit curve has decreased to 10% of its initial value and indicates a level of exploitation at which a further increase would result in an increase in yield per recruit and decrease in spawning stock per recruit (ICES, 2014). *F*/*F*0.1 is a ratio-based indicator, where *F* > *F*0.1 (*F*/*F*0.1 above 1.0) is considered overfishing, *F* = *F*0.1 (*F*/*F*0.1 = 1.0) is considered fully exploited, and *F*/*F*0.1 < 0.5 is underexploited.

* + 1. *Length-Based Reference Points (LBRP)*

Froese (2004) discussed well-established relationships between fisheries management and life history theory. These relationships are based on three simple ideas about length compositions: (i) Pmat is the percentage of mature individuals in the catch (i.e. length at 50% maturity). Mature individuals are defined as those who have spawned at least once, meaning that the target is to let all fish spawn before being caught, i.e. catching no juveniles, to maintain spawning stocks; (ii) Popt is the percentage of fish caught at optimum length, which is the length when the highest yield occurs, and thus the goal is to catch harvested fish within ±10% of optimum length. (iii) Pmega is the percentage of large, mature fish (called mega-spawners) in the catch (optimum length plus 10%). The target is to catch zero mega-spawners as these fish play an important role in long-term survival of a population (Froese, 2004). Cope and Punt (2009) further elaborated on the concepts of Froese (2004) and explored how these reference points relate to fishing mortality, spawning stock biomass (SSB) and current reference points based on SSB. The authors added a new measure, *Pobj*, which is used to distinguish between fishing patterns and construct a decision tree found in Cope and Punt’s study (2009). This decision tree describes how the value of *Pobj* gives information about selectivity characteristics of the fisheries and SSB that is used to develop stock status indicators. Given catch data and life history, managers can use the decision tree to also examine population trends and implement harvest control rules. LBRP is thus an indicator-based method that evaluates whether a stock’s SSB is at or above a specified reference point. It requires as input the proportion of catch in each length class, , *L∞*, and the length at which a stock’s cohort provides the highest yield (*Lopt*), which can be calculated using the von Bertalanffy function and length-at-weight relationship. LBRP relies on the assumption that catch length composition is representative of the length composition of the exploited stock. We only used Pobj for the analysis and comparison.

* + 1. *Length-Based Risk Analysis*

Based on length composition data, various indicators can be applied to provide a proxy for stock status (Chrysafi and Kuparinen, 2016). One of the most common indicators is the mean length, which is a metabolic-based indicator that is highly correlated with population size (Ricker, 1963; Pauly and Morgan, 1987; Ehrhardt and Ault, 1992; Beverton and Holt, 1993; Jennings *et al.*, 2001; Kerr and Dickie, 2001; Ault *et al.*, 2008). The mean length is usually estimated from all fish length captured between the lengths at first capture (*Lc*) and the maximum observed length (*Lmax*), (Ault *et al.*, 2008, 2018). From the mean length, the total instantaneous mortality rate can be estimated, which negatively correlates with mean length. As an extension, Ault et al. (1998, 2008, 2018) used the mean length to calculate reference points to address sustainability risks, including fishery yields, SPR, and precautionary control rules. Therefore, the Length-Based Risk Analysis method is an indicator-based method that expands on the mean length to include a comparison of population metrics at current levels of fishing mortality against sustainability measurements. It requires as inputs length composition of catches, the von Bertalanffy growth parameters, maximum length of the largest bin class, length at first capture ( in this case), length weight relationship (*α* and *β*), maximum age, number of mature individuals (amount of individuals larger than the length-at-maturity ()), and natural mortality (*M*). The assumption of this method is that the fishery must be in an equilibrium condition (i.e. fishing mortality is constant).

* + 1. *Length-Based Spawning Potential Ratio (LBSPR)*

Work done by Hordyk et al. (2015b) demonstrated that the expected length composition of the catch of an exploited stock is determined by the ratio of natural mortality to the von Bertalanffy growth coefficient (*M/K*) and the ratio of fishing mortality to natural mortality (*F/M*). The *M/K* ratio is one of the life history ratios that seems relatively consistent between closely related stocks and can give an estimate of length-at-maturity found at the length of maximum biomass in the population, which gives another life history ratio *Lm/L∞* (Hordyk *et al.*, 2015a). These ratios are known as the Beverton-Holt life history invariants (Hordyk *et al.*, 2015b, 2015c). Hordyk et al. (2015b) also suggested that if *M/K* is known, it is possible to estimate *F/M* from size composition of catch. The authors determined that under assumptions of knife-edge selectivity at length at first capture (*Lc*) and knife-edge maturity at , the SPR is determined by the ratios *M/k*, *F/M*, *Lm/L∞* and *Lc/L∞*. The Length Based SPR model (LBSPR) is a prominent length-based model that assesses stock status assuming equilibrium conditions. LBSPR requires as input the length data (minimum one year), *M/K* ratio, the von Bertalanffy growth parameters, coefficient of variation of *L∞*, length at 50% and 95% maturity ( and), length-weight parameters, and start values for lengths at 50% and 95% selectivity ( and). LBSPR relies on several assumptions: (i) the stock is in steady state with constant recruitment; (ii) natural mortality and growth rates are constant; (iii) selectivity follows a logistic curve; (iv) growth is described by the von Bertalanffy function; (v) both sexes have the same growth curve and the sex ratio is equal unless using length composition of female fish only; and (vi) the lengths at each age are normally distributed around a mean length-at-age value.

* + 1. *Length-based Integrated Mixed Effects (LIME)*

Many length-based methods follow the equilibrium assumption that recruitment and fishing mortality arise from deterministic relationships and are constant through time. These assumptions are usually violated as recruitment and fishing mortality often vary due to stochastic ocean conditions and productive regime shifts (Vert-pre *et al.*, 2013; Thorson *et al.*, 2014; Szuwalski *et al.*, 2015; Rudd and Thorson, 2018). Therefore as an alternative, a mixed-effects model can be used to account for random variation from natural or measurement processes separately (de Valpine and Hastings, 2002; Thorson *et al.*, 2014; Rudd and Thorson, 2018). A new length-based method was thus created by Rudd and Thorson (2018) called Length-based, Integrated, Mixed-Effects (LIME), which was built upon the catch-curve stock reduction analysis model (Thorson and Cope, 2015). The minimum inputs for LIME are length data (minimum one year), length-at-age relationship, natural mortality, and length at 50% maturity (*Lm*). LIME then estimates fishing mortality, and lengths at 50% and 95% selectivity ( and). The assumptions of LIME are that (i) growth is described by von Bertalanffy function; (ii) maturity-at-length follows a logistic curve; and (iii) natural mortality is constant.

* + 1. *SPR-based methods*

The Length-Based Risk Analysis method, LBSPR, and LIME give an output of SPR. Many studies have explored the levels of SPR to be used as target and limit reference points, resulting in assessments of many species using SPR of 30% as a limit and 40% as a target reference point (Mace and Sissenwine, 1993; Clark, 2002; Hordyk *et al.*, 2015c).

* 1. Performance measures

We measured performance of the estimation models based on relative error of each simulation replicate, calculating bias as the median relative error (MRE) and precision as the median absolute relative error (MARE): (Equations 22 and 23):

(22)

(23)

where *xest* is the estimated value (calculated from the estimation models) and *xtrue* is the true value (calculated from the operating models). The bias and precision values are performance indicators that are relatively robust to outliers. Looking at the distribution of errors also helps indicate accuracy to determine how far some of the estimates could deviate from the ‘truth’ for any given assessment method. We would interpret a model as more useful than another if the bias and precision values are closer to zero, and if the range of errors are smaller.

1. Results

The results of the simulations differed among longevity, exploitation levels, and recruitment types, depicted in Table 3 and Figures 4-6.

* 1. *TB*

TB was the least biased and imprecise in the short-lived scenario, but the most biased and imprecise in the longer-lived. It estimated *F/F0.1* with as bias of -0.4140 for the short-lived and -0.5703 for longer-lived. The precision in *F/F0.1* was 0.4140 for the short-lived and 0.5703 for the longer-lived. When *F/F0.1* gave indication of overexploitation, TB improved in accuracy (from -0.4805 to -0.4305). Although in the underexploited state, TB performed the worst with bias of -0.5011 and precision of 0.5011. Given high recruitment variability and autocorrelated error increased bias (from -0.4805 to -0.5211) and decreased precision (from 0.4805 to 0.5211). Lastly, TB performed the best in the Beverton-Holt scenario (base model). Compared to the other methods, TB was the most biased and least precise method throughout the scenarios.

* 1. *LBRP*

LBRP performed the worst in the short-lived scenario with a bias of -0.2166 and precision of 0.2166 and the best in the medium-lived (base model) with a bias and precision of 0.0349. When *Pobj* gave indication of overexploitation, LBRP’s performance improved with a bias and precision of 0.0406. It became more biased and imprecise (0.0406 both) when *Pobj* gave indication of underexploitation. Lastly, looking at varying recruitment types, LBRP performed just as well in the equilibrium scenario as the autocorrelated scenario (bias and precision of 0.0333 vs 0.0321), meaning that either having no or high variability in recruitment improved the bias and precision. Overall, LBRP was the least biased and most precise method.

* 1. *Length-Based Risk Analysis*

Length-Based Risk Analysis was the least biased (-0.1108) and most precise (0.1820) in the short-lived scenario. It performed the worst in the medium-lived with a bias of -0.3049 and precision of 0.3059. While Length-Based Risk Analysis was the most biased (-0.3576) and imprecise (0.3601) when SPR gave indication of underexploitation, it was the least biased (-0.2700) and most precise (0.3601) when SPR gave indication of overexploitation. In terms of recruitment type, when there is no recruitment variability, i.e. the equilibrium scenario, Length-Based Risk Analysis was the most biased (-0.3642). When there is high recruitment variability and autocorrelated error, the estimation of SPR was the most imprecise (0.3661) Comparatively to the other SPR-based methods, the Length-Based Risk Analysis consistently on average underestimated SPR and performed the worst out of the three other SPR-based methods.

* 1. *LBSPR*

In comparing life history scenarios, LBSPR was the most biased (0.1398) and the least precise (0.2469) in the short-lived scenario. It was the least biased (0.0871) in the medium-lived but most precise in the longer-lived (0.1538). Ultimately, LBSPR performed the best in the longer-lived scenario, being the most precise (0.1538 vs. 0.1895 base model) and it was not significantly more biased (0.0873 vs. 0.0871) than the medium-lived. LBSPR was the most biased (0.1027) when SPR gave indication of overexploitation and the most imprecise (0.1895) when SPR gave indication of target level exploitation (base model). It was the least biased (0.0560) and most precise (0.1493) when SPR gave indication of underexploitation. LBSPR was less biased in the autocorrelated scenario (0.0297) than the Beverton-Holt (base model; 0.0871). However, LBSPR was significantly less precise (0.2504) in the autocorrelated scenario than the Beverton-Holt, and therefore performed the worst in the autocorrelated scenario. It performed the best in the equilibrium scenario with a bias of 0.0711 and precision of 0.0729.

* 1. *LIME*

LIME performed the worst in the longer-lived with a bias of -0.1904 and precision of 0. 2426 and performed the best in the medium-lived with a bias of 0.0554 and precision of -0.2004. SPR estimated by LIME was the most biased (-0.2545) and the least precise (0.2628) when SPR gave indication of overexploitation. It performed best when SPR gave indication of target level exploitation with a bias of 0.0554 and precision of 0.2004. LIME was the most biased in the equilibrium scenario (-0.0874) and the most imprecise in the autocorrelated scenario (0.2152). Although, LIME was the most unbiased (0.0554) in the Beverton-Holt scenario, it was the most precise in the equilibrium scenario (0.1533).

* 1. *Performance of length-based methods and across scenarios*

The performance of the methods was affected by fish longevity, although the results were quite variable. LBRP and LBSPR performed the worst in the short-lived while TB and LIME performed the worst in the longer-lived. The Length-Based Risk Analysis performed the best in the short-lived scenario, and LBRP performed best in the medium- and longer-lived scenarios. When excluding the performance of LBRP (discussed later in the Discussion section), then LIME was the least biased and LBSPR was the most precise in the medium-lived scenario, and LBSPR performed the best in the longer-lived scenario. In terms of changing exploitation levels, the underexploited scenario performed the worst on average (TB, LBRP, and Length-Based Risk Analysis). LBRP and LBSPR performed the best and most consistently through all the exploitation scenarios as evident in the performance statistics. Under conditions of changing recruitment differences between methods were the most evident. All but LIME performed the best in the equilibrium scenario, most notably with all the precision increasing. LBRP and LBSPR again performed the best in the equilibrium scenario. The methods performed the worst in the autocorrelated scenario with decreased precision and increased bias. LBRP and LIME performed the best in the autocorrelated scenario.

1. Discussion

In this study, we used simulation analysis to compare performance of different length-based stock assessment methods under different life history traits and exploitation regimes of fish. Overall, LBRP performed the best according to the performance indicators. However, there was no error included with the estimation of , and thus the *Pmat,* the proportion of individuals in the catch that is mature, was the same between the operating and estimation models, ultimately leading to the small relative errors in *Pobj,* the indicator calculated by LBRP that gives information about selectivity characteristics and SSB. While this method has the least input requirements and is especially helpful in truly data-poor assessments, the reference points do not have a control rule determining how to change the stock status. For example, if the *Pobj* value gives indication that the fish are small and immature, it does not tell how much fishing pressure needs to be reduced to prevent overfishing. On the contrary, LBRP does give indication size limits that could be used to avoid catching smaller fish, and other control rules besides effort controls could be used with this method. LBRP may have been the least biased and most precise, but it still needs a control rule to link the results to a management action.

The next best performing method was LBSPR as it performed the most consistently across the varying scenarios. Rudd and Thorson (2018) also found that LBSPR performed better when the stocks were in equilibrium and when the operating model matched LBSPR’s assumption; this result is supported here given that the IBM requires that the stocks simulated have constant fishing. An advantage of LBSPR is that it recalculates the selectivity curve, meaning that the method is more likely to give accurate estimates of the exploitation level based on the length frequency distribution, possibly explaining why LBSPR outperformed the other methods through the scenarios. On the contrary, this study and Rudd and Thorson (2018) found that LIME performed better in the autocorrelated scenario, which matches the results in Hordyk et al. (2015b) where with high recruitment variability (*σR*= 0.6 or 0.9), LBSPR had imprecise and biased results.

LIME is the only length-based model evaluated that does not require the equilibrium assumption. The mixed effects nature of LIME allows consideration of varying recruitment and fishing mortality over time, avoiding the need of the equilibrium assumption. Although this study did not look at fishing rate variability, recruitment variability was considered as in Rudd and Thorson’s (2018) study, and both studies found that LIME performed better in the autocorrelated scenario. While this study found that LIME performed better in the medium-lived scenario, Rudd and Thorson (2018) found that LIME performed best in the short-lived scenario. This could be the case as different operating models were used between the two studies. Otherwise, this study also found that LIME performed worst in the longer-lived scenario. LIME also performs better when there are multiple years of length composition data or more time steps as LIME was designed to capture variability in fishing and recruitment through multiple time steps. It came close to LBSPR in terms of performance throughout the scenarios.

The Length-Based Risk Analysis method implements the long-known idea that mean length is an important parameter to estimate stock status and fishing pressure (Nadon *et al.*, 2015). Out of the three SPR-based methods, the Length-Based Risk Analysis method performed the worst, which was the most biased and imprecise in most cases across scenarios. In a majority of the scenarios, the Length-Based Risk Analysis method consistently underestimated SPR. As the Lmax could not be bigger than *L∞* because the method uses a ‘truncated model’ of Ehrhardt and Ault (1992), some of the length classes were truncated, and truncations lead to overestimation of Z (Then *et al.*, 2015), leading to underestimation of SPR. The major difference between the Length-Based Risk Analysis, LBSPR, and LIME is in how the SPR was calculated, where the Length-Based Risk Analysis method does not consider selectivity-at-age when calculating the total mortality (used to calculate SPR). This method also can give negative *F* estimates, meaning either that *M* is wrong or there are biased estimates of *Z* and selectivity in cases when we know *M* to be true. This problem is evident in the TropFishR routine as well.

In all scenarios, the TB consistently underestimated *F/F0.1* on average. One major challenge in the TB is the routine that needs to be done before obtaining an estimate of *F/F0.1*; this routine again includes the LCCC and Jones’ Cohort Analysis. If a step in either LCCC or the Cohort Analysis was incorrectly done, this would interfere with the estimates given by TB, and finding the mistake is more difficult. The LCCC also can give negative *F* estimates, which is impossible, and these negative estimates will lead to more unusual outputs from the Cohort Analysis and the TB. Additionally, all three methods’ errors and uncertainty need to be taken into consideration. This technicality is reflected in the results of *F/F0.1* where in a majority of the scenarios and life history tools, TB performed the worst.

It is evident in this study that fish longevity has an impact on the performance of the life history tools and length-based methods, but the extent of the impact varied from method to method. Usually stock assessment methods do not perform well with short-lived species, especially the LCCC. The main problem with assessing short-lived species is that the annual time step is not enough to learn about their dynamics to the extent that can be with longer-lived species. Additionally, for every month in the length frequency distribution, there were 200 individuals for all scenarios, meaning that the distribution of individuals to each length class for the short-lived were more concentrated in certain length classes than the medium- or longer-lived.

On the other hand, the individuals were then dispersed across more length bins for the longer-lived. It is known that longer-lived species have lower SPR levels as there is a relationship between longevity and sensitivity of SPR to exploitation pressures (Nadon *et al.*, 2015). Thus, the longer-lived species have their spawning biomass represented by older individuals and their numbers easily become reduced even with a low fishing rate. Rudd and Thorson (2018) stated that short-lived stocks’ length data have distinct cohorts and the longer-lived stocks have less distinct cohorts, as with increasing length blurs the cohort as the fish ages. The medium-lived species seem the best suited for these length-based methods, and one must consider the uncertainties and problems when assessing either short- or longer-lived species.

The changes in exploitation levels did not have the expected larger influence on the estimates. There is conforming evidence in our study that most methods performed worse in the underexploited scenario, which is not a big concern as giving a precautionary signal from a method is important as well. In a future study, the influence of fishing patterns (e.g. two-way trip where there is a linear increase in *F* up to *F > FMSY* followed by a linear decrease of *F* down to *F < FMSY* and fishing down where is a continuous increase in *F* to an *F > FMSY*(Ono *et al.*, 2015)) would be of interest to investigate as fishing mortality is usually not constant.

The impacts of changing recruitment types were evident in all the methods evaluated. All methods had smaller error ranges, and worse performance when recruitments were autocorrelated. Many stock assessment models, especially length-based methods, assume equilibrium. However, this assumption is typically violated as recruitment variation changes the age structure of a population with time (Haddon, 2001). It should be noted that in all scenarios, the fishing mortality was also in equilibrium, meaning that the equilibrium recruitment scenario was also a true equilibrium scenario. The methods did not perform well in the Beverton-Holt recruitment as in the equilibrium scenario because of the added recruitment error (*σR* = 0.5), but the methods performed better than in the autocorrelated scenario as it still followed its respective pattern. While the methods did perform worse in the autocorrelated scenario, it may be beneficial to use methods (e.g., LIME) that do take into consideration recruitment variability. Since the performance of LBRP and LBSPR were relatively consistent throughout the scenarios, a recommendation could be to use those methods initially as they would give relatively good estimates despite the underlying uncertainties. For fisheries that are known to be in non-equilibrium conditions or in later assessments, LIME would be more advantageous to use as it takes fishing and recruitment variability into consideration. Another recommendation is to run LBRP, LBSPR, and LIME at the same time to see if they give consistent results to determine what the prevailing condition (low or high recruitment variability) is.

While there are many recommendations given by many authors about the ideal sample size (Erzini, 1990; Gerritsent and McGrath, 2007; Hordyk *et al.*, 2015b), the sample size is ultimately dependent on life history and the fishery selectivity. For example, there are some cases where stock assessment scientists will only analyse length frequency data if the average sample size is below 100 individuals per month. It is possible to have much more than 100 samples per month, but it is also possible to only have an effective sample size of 100. In data-poor areas, fisheries use total catch and might not take number of individuals of a certain species into account. This scenario could be simulated in a future study where a certain tonnage is given, and then the number of individuals would depend on tonnage rather than giving a set number to the IBM. Rudd and Thorson (2018) stated that the accuracy in stock status estimates is more likely improved with more data types rather than increasing sample size. Adding more data types is generally not realistic in data-poor areas as only a limited amount of data is usually collected. Ono et al. (2015) mentioned that infrequent sampling over a longer period was more informative than frequent surveys covering a shorter period. Therefore, the sampling amount and time period are significant in data-poor areas. To implement these concerns, future analyses could look at the significance of sampling schemes and investigate how sampling time periods affect the quality of the stock assessments.

It should also be considered that length frequency data obtained from tropical data poor fisheries are often strongly biased due to gear selectivity. The data generation of our simulation study, based on the assumption of logistic-like selectivity, is thus most often not reflecting conditions of tropical multigear fisheries and the size composition of the catch may rather reflect a mix of sizes due to a mix of gears used for the fishing. Ideally, the selection characteristics of the gear(s) should be known prior to any length frequency analysis and (if possible) catch length frequency data should first be reconstructed based on the selectivity features of the gear.

The study gives insight on which length-based methods and tools worked the best in certain scenarios commonly seen in fisheries. Collecting time series of historical catch, catch per unit of effort trends, or age structure information is challenging in data-limited fisheries, and thus scientists resort to size composition data (Maunder and Punt, 2013). However, many population and fishery processes contribute to the shape of size composition data, leading to more uncertainty in stock assessment estimates when that data is available without auxiliary information. Additionally, many population dynamic processes are not well understood in these areas, and stock assessment methods that can capture these dynamics are required. Simulation testing, such as a performance evaluation, is even more essential in data-poor areas as stock assessment scientists and manager need to understand the degree of error in estimating stock status and population dynamics. Evaluation of the length-based methods and tools in different scenarios thus helps determine which methods and tools work best in certain situations. Understanding how different dynamics affect these methods and tools are also essential in the development of new approaches as it highlights the weaknesses of the current methods.

The use of multiple assessment methods with varying assumptions and outputs is ideal for stock assessments. It is also important to keep collecting additional data required for a more comprehensive stock assessment. While assessments based on length data alone have high uncertainty, length-based methods for data-limited fisheries are important in providing an initial estimate of stock status to inform management. Thus, the assumptions and sensitivities of each of the methods must be considered when interpreting the results. The main short-coming for all these methods is that they should not be used for long-term assessments. Length-based methods should also be used carefully as length compositions are often not representative of the whole stock (Hilborn and Walters, 2015). These methods can be used for short-term assessments to provide a cost-effective starting point to later begin long-term processes of data collection, assessment, and management.

These data-poor stock assessments also have to consider uncertainty, e.g. confidence intervals, as the uncertainty at least give some understanding to the stock status. As time goes on, new assessments and information could be added to better inform the previous assessment and reduce the uncertainties. Data-limited stock assessments need to quantify uncertainty. The importance of capturing uncertainty in estimating stock status and population dynamics was highlighted in this study as there is a need to better implement these current methods and create future ones that will consider the weaknesses of the current methods.

1. References

Ault, J., Bohnsack, J., and Meester, G. 1998. A Retrospective (1979-1996) Multispecies

Assessment of Coral Reef Fish Stocks in the Florida Keys. Fishery Bulletin- National Oceanic and Atmospheric Administration, 96: 395–414.

Ault, J. S., Smith, S. G., Luo, J., Monaco, M. E., and Appeldoorn, R. S. 2008. Length-based assessment of sustainability benchmarks for coral reef fishes in Puerto Rico. Environmental Conservation, 35: 221.

Ault, J. S., Smith, S. G., Bohnsack, J. A., Luo, J., Stevens, M. H., DiNardo, G. T., Johnson, M. W., *et al.* 2018. Length-based risk analysis for assessing sustainability of data-limited tropical reef ﬁsheries: 16.

Beverton, R. J. H., and Holt, S. J. 1993. On the Dynamics of Exploited Fish Populations. Springer Netherlands, Dordrecht. http://link.springer.com/10.1007/978-94-011-2106-4 (Accessed 10 October 2018).

Bystrom, A. 2016. Análisis de características biológico-pesqueras del pargo manchado (Lutjanus guttatus (Steindachner, 1869) y tendencias socio-ecológicas de la pesca artesanal con líneas de fondo en el distrito de Bejuco, Pacífico de Costa Rica. Universidad Estatal a Distancia, San Jose, Costa Rica.

Cadrin, S. X., and Dickey-Collas, M. 2015. Stock assessment methods for sustainable fisheries. ICES Journal of Marine Science, 72: 1–6.

Cao, J., Guan, W., Truesdell, S., Chen, Y., and Tian, S. 2016. An individual-based probabilistic model for simulating fisheries population dynamics. Aquaculture and Fisheries, 1: 34–40.

Chrysafi, A., and Kuparinen, A. 2016. Assessing abundance of populations with limited data: Lessons learned from data-poor fisheries stock assessment. Environmental Reviews, 24: 25–38.

Clark, W. G. 2002. *F*35% Revisited Ten Years Later. North American Journal of Fisheries Management, 22: 251–257.

Cope, J. M., and Punt, A. E. 2009. Length-Based Reference Points for Data-Limited Situations: Applications and Restrictions. Marine and Coastal Fisheries, 1: 169–186.

de Valpine, P., and Hastings, A. 2002. Fitting Population Models Incorporating Process Noise and Observation Error. Ecological Monographs, 72: 57.

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

Ehrhardt, N. M., and Ault, J. S. 1992. Analysis of two length-based mortality models applied to bounded catch length frequencies. Transactions of the American Fisheries Society, 121: 115–122.

Erzini, K. 1990. Sample size and grouping of data for length-frequency analysis. Fisheries Research, 9: 355–366.

Froese, R. 2004. Keep it simple: three indicators to deal with overfishing. Fish and Fisheries, 5: 86–91.

Gerritsent, H. D., and McGrath, D. 2007. Precision estimates and suggested sample sizes for length- frequency data, 105: 116–120.

Haddon, M. 2001. Modelling and quantitative methods in fisheries. Chapman & Hall/CRC, Boca Raton, Fla. 406 pp.

Heemstra, P. C., and Randall, J. E. 1993. Groupers of the world (family Serranidae, subfamily Epinephelinae). *In* FAO Species Catalogue. 125 pp.

Herrón, P., Mildenberger, T. K., Díaz, J. M., and Wolff, M. 2018. Assessment of the stock status of small-scale and multi-gear fisheries resources in the tropical Eastern Pacific region. Regional Studies in Marine Science, 24: 311–323.

Hicks, C. C., and McClanahan, T. R. 2012. Assessing Gear Modifications Needed to Optimize Yields in a Heavily Exploited, Multi-Species, Seagrass and Coral Reef Fishery. PLoS ONE, 7: e36022.

Hilborn, R., and Walters, C. J. 2015. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Springer, Dordrecht. 570 pp.

Hordyk, A., Ono, K., Sainsbury, K., Loneragan, N., and Prince, J. 2015a. Some explorations of the life history ratios to describe length composition, spawning-per-recruit, and the spawning potential ratio. ICES Journal of Marine Science, 72: 204–216.

Hordyk, A., Ono, K., Valencia, S., Loneragan, N., and Prince, J. 2015b. A novel length-based empirical estimation method of spawning potential ratio (SPR), and tests of its performance, for small-scale, data-poor fisheries. ICES Journal of Marine Science, 72: 217–231.

Hordyk, A. R., Loneragan, N. R., and Prince, J. D. 2015c. An evaluation of an iterative harvest strategy for data-poor fisheries using the length-based spawning potential ratio assessment methodology. Fisheries Research, 171: 20–32.

ICES. 2014. Report of the Workshop on the Development of Quantitative Assessment Methodologies based on Life-history traits, exploitation characteristics, and other key parameters for data-limited stocks (WKLIFE V). ICES CM 2014/ACOM:54.

Jennings, S., Kaiser, M. J., and Reynolds, J. D. 2001. Marine fisheries ecology. Blackwell Science, Oxford; Malden, MA, USA. 417 pp.

Kerr, S. R., and Dickie, L. M. 2001. The biomass spectrum: a predator-prey theory of aquatic production. Complexity in ecological systems series. Columbia University Press, New York. 320 pp.

Mace, P. M., and Sissenwine, M. P. 1993. How much spawning per recruit is enough? Canadian Special Publication of Fisheries and Aquatic Science No. 120, 120: 101–118.

Maunder, M. N., and Punt, A. E. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research, 142: 61–74.

Mildenberger, T. K., Taylor, M. H., and Wolff, M. 2017. TropFishR: an R package for fisheries analysis with length-frequency data. Methods in Ecology and Evolution, 8: 1520–1527.

Nadon, M. O., Ault, J. S., Williams, I. D., Smith, S. G., and DiNardo, G. T. 2015. Length-Based Assessment of Coral Reef Fish Populations in the Main and Northwestern Hawaiian Islands. PLOS ONE, 10: e0133960.

Nelson, G. A. 2017. fishmethods: Fishery Science Methods and Models in R. https://CRAN.R-project.org/package=fishmethods.

Ono, K., Licandeo, R., Muradian, M. L., Cunningham, C. J., Anderson, S. C., Hurtado-Ferro, F., Johnson, K. F., *et al.* 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES Journal of Marine Science, 72: 31–43.

Pauly, D., and Morgan, G. R. (Eds). 1987. Length-based methods in fisheries research. ICLARM contribution. International Center for Living Aquatic Resources Management; Kuwait Institute for Scientific Research, Makati, Metro Manila, Philippines : Safat, Kuwait. 468 pp.

Pilling, G. M., Apostolaki, P., Failler, P., Floros, C., Large, P. A., Morales-Nin, B., Reglero, P., *et al.* 2008. Assessment and Management of Data-Poor Fisheries. *In* Advances in Fisheries Science, pp. 280–305. Ed. by A. Payne, J. Cotter, and T. Potter. Blackwell Publishing Ltd., Oxford, UK. http://doi.wiley.com/10.1002/9781444302653.ch12 (Accessed 10 October 2018).

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Richards, L. J., and Maguire, J.-J. 1998. Recent international agreements and the precautionary approach: new directions for fisheries management science, 55: 8.

Ricker, W. E. 1963. Big Effects from Small Causes: Two Examples from Fish Population Dynamics. Journal of the Fisheries Research Board of Canada, 20: 257–264.

Rosenberg, A. A., and Restrepo, V. R. 1994. Uncertainty and Risk Evaluation in Stock Assessment Advice for U.S. Marine Fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 51: 2715–2720.

Rudd, M. B., and Thorson, J. T. 2018. Accounting for variable recruitment and fishing mortality in length-based stock assessments for data-limited fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 75: 1019–1035.

Sparre, P., Ursin, E., and Venema, S. C. 1998. Introduction to tropical fish stock assessment. Pt. 1: Manual. FAO fisheries technical paper. Food and Agriculture Organization of the United Nations, Rome. 407 pp.

Szuwalski, C. S., Vert-Pre, K. A., Punt, A. E., Branch, T. A., and Hilborn, R. 2015. Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish and Fisheries, 16: 633–648.

Taylor, M. H. 2017. fishdynr: Fisheries science related population dynamics models. https://github.com/marchtaylor/fishdynr.

Taylor, M. H., and Mildenberger, T. K. 2017. Extending electronic length frequency analysis in R. Fisheries Management and Ecology, 24: 330–338.

Tesfaye, G., Wolff, M., and Taylor, M. 2016. Gear selectivity of fishery target resources in Lake Koka, Ethiopia: evaluation and management implications. Hydrobiologia, 765: 277–295.

Then, A. Y., Hoenig, J. M., Hall, N. G., Hewitt, D. A., and Handling editor: Ernesto Jardim. 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES Journal of Marine Science, 72: 82–92.

Thompson, W. F., and Bell, F. H. 1934. Biological statistics of the Pacific halibut fishery. 2. Effect of changes in intensity upon total yield and yield per unit of gear. Report Internal Fisheries (Pacific Halibut) Commission, 8: 49.

Thorson, J. T., Cope, J. M., and Patrick, W. S. 2014. Assessing the quality of life history information in publicly available databases. Ecological Applications, 24: 217–226.

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

Tuda, P. M. 2018. Assessing the State and Impacts of the Artisanal Reef Fisheries and their Management Implications in Kenyan South Coast. Universität Bremen: Biologie/Chemie, Bremen, Germany. 198 pp. https://elib.suub.uni-bremen.de/edocs/00106467-1.pdf.

Vert-pre, K. A., Amoroso, R. O., Jensen, O. P., and Hilborn, R. 2013. Frequency and intensity of productivity regime shifts in marine fish stocks. Proceedings of the National Academy of Sciences, 110: 1779–1784.

1. Figures and Tables



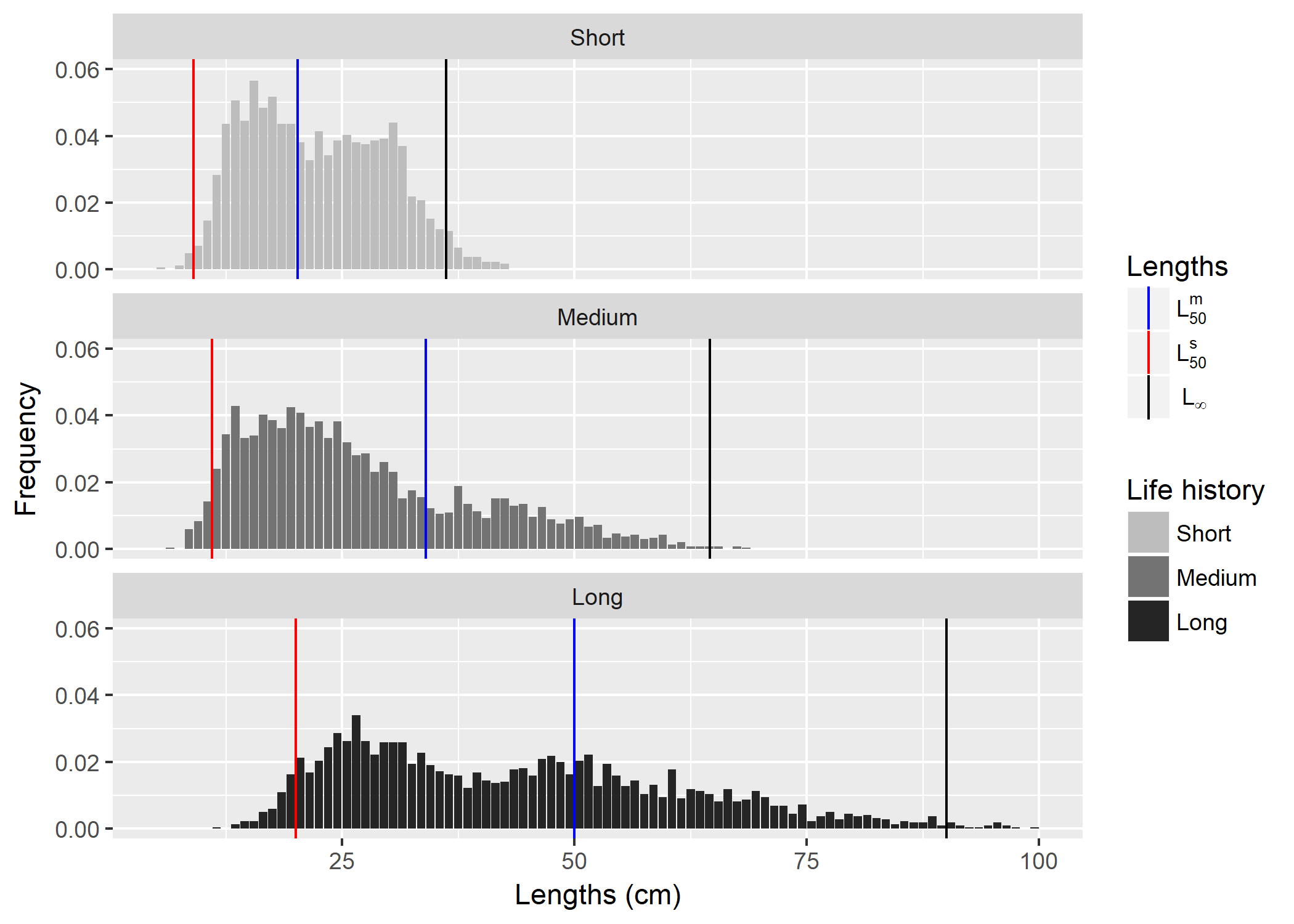


Figure 1: Length frequency distribution graphs. For each fish longevity scenario (short, medium, long), L50m, L50s, and Linf are visualised (in blue, red, and black respectively) over the length frequnecy graphs for one iteration.

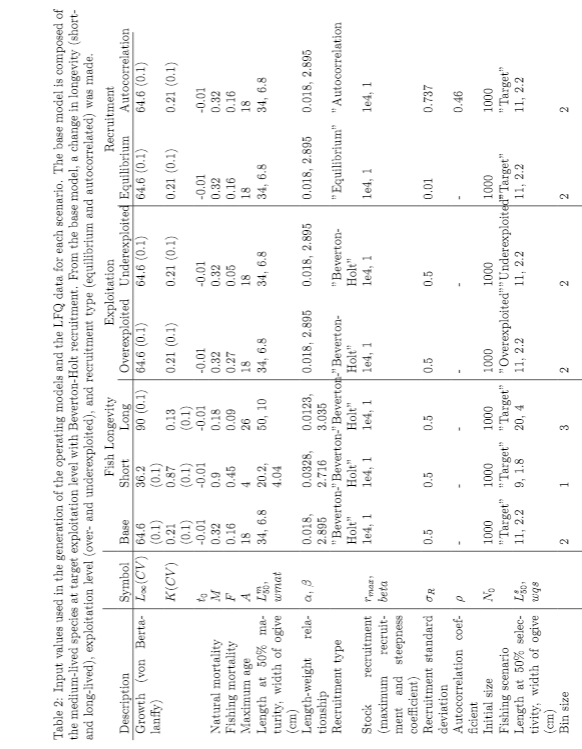


Figure 2 Simulation-estimation analysis diagram. How a real fishery is assessed is compared to how we assessed our simulated stocks in this study.

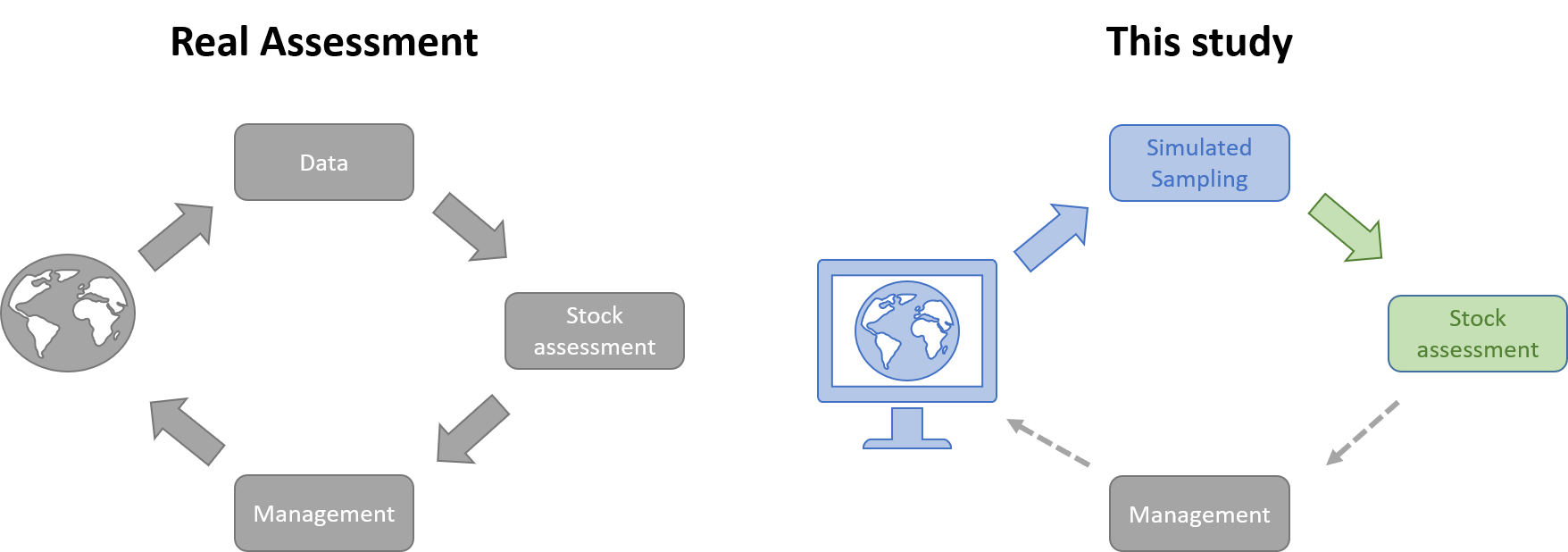


Figure 3: Simulation study methodology diagram. There are seven operating model setups. Scenarios differ in fish longevity with (1) medium-lived (base model), (2) short-lived, (3), and longer-lived stocks. All fish longevity simulations were run with Beverton-Holt recruitment and exploitation at target level (SPR = 0.4). From the base model, either the recruitment or exploitation scenario changes; (4) equilibrium and (5) autocorrelated recruitment; and (6) overexploited and (7) underexploited fishing scenarios. For each operating model, one year of monthly length frequency data and the "true" stock estimate parameters were simulated and extracted. The "true" life history values (L1, K, M, Lm50) from the operating models and the length frequency data were then used as input for the simulated assessment with the five length-based estimation models: (1) Thompson-and-Bell (TB), (2) Length-Based Reference Points (LBRP), (3) Length-Based Risk Analysis, (4) Length-Based Spawning Potential Ratio (LBSPR), and Length-based Integrated Mixed Effects (LIME).

