Performance evaluation of data-limited length-based stock assessment methods

Lisa Chong

1. Leibniz Centre for Tropical Marine Research (ZMT), Fahrenheitstr. 6, Bremen, 28359 Germany

lisa.chong8594@gmail.com

Tobias K. Mildenberger

2. National Institute of Aquatic Resources, Technical University of Denmark, Kemitorvet, Building 202, 2800 Kgs. Lyngby, Denmark

tobm@aqua.dtu.dk

Merrill B. Rudd

3. School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA

4. Fishery Resource Analysis and Monitoring Division, National Oceanic and Atmospheric Administration, National Marine Fisheries Service,

Northwest Fisheries Science Center, 2725 Montlake Boulevard East, Seattle, WA 98112-2097, USA

merrillrudd@gmail.com

Marc H. Taylor

5. Marine Living Resources, Thünen Institute of Sea Fisheries, Herwigstr. 31, Bremerhaven, 27572, Germany

marc.taylor@thuenen.de

Jason M. Cope

3. School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA

4. Fishery Resource Analysis and Monitoring Division, National Oceanic and Atmospheric Administration, National Marine Fisheries Service,

Northwest Fisheries Science Center, 2725 Montlake Boulevard East, Seattle, WA 98112-2097, USA

jason.cope@noaa.gov

Trevor A. Branch

3. School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA

tbranch@uw.edu

Matthias Wolff

1. Leibniz Centre for Tropical Marine Research (ZMT), Fahrenheitstr. 6, Bremen, 28359 Germany

matthias.wolff@leibniz-zmt.de

Moritz Stäbler

1. Leibniz Centre for Tropical Marine Research (ZMT), Fahrenheitstr. 6, Bremen, 28359 Germany

moritz.staebler@leibniz-zmt.de

12/03/19

1. Abstract

Performance evaluation of the multitude of data-limited length-based methods is instrumental to determine and quantify their precision and bias under various assumptions/scenarios and thus providing guidance about model applicability and limitations to stock assessment scientists and managers. A simulation-estimation analysis was conducted using an individual based operating model to compare the performance of four length-based stock assessment methods, length-based Thompson and Bell (TB), length-based spawning potential ratio (LBSPR), length-based integrated mixed effects (LIME), and length-based risk analysis (LBRA), under varying life history, exploitation status, and recruitment error scenarios. We found that all methods have difficulties when assessing short-lived species. When the stocks are severely overexploited, the methods are less accurate in estimating the degree of recruitment overfishing while the methods present inconsistencies in determining growth overfishing when the stocks are underexploited. While having additional recruitment error does decrease precision, it does not necessarily increase bias. Furthermore, we found that there is a higher precision in estimation of relative reference points (i.e. spawning potential ratio and F/FMSY) than absolute ones (e.g. FMSY). Across all scenarios, TB and LBSPR were most the consistent and accurate assessment methods. This study highlights the importance of quantifying the accuracy of stock assessment methods and testing methods in different scenarios to determine their strengths and weaknesses, which will lead to better utilization of these methods.

Keywords: simulation-estimation analysis; data-limited fishery; length-based assessment; spawning potential ratio; MSY

1. Introduction

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 extensively. Prominent length-based methods include the length-based Thompson-and-Bell model (TB), length-based spawning potential ratio (LBSPR), length-based integrated mixed effects (LIME), and length-based risk analysis (LBRA).

TB is one of the oldest length-based methods that is a yield per recruit model and evaluates the stock’s status relative to fishing and selectivity reference levels (Mildenberger *et al.*, 2017). LBSPR is a prominent length-based model that assesses stock status by comparing the length composition data to the expected unfished length structure (Hordyk *et al.*, 2015b). LIME accounts for time-varying recruitment and fishing mortality, relaxes the equilibrium assumption and derives population parameters associated with an age-structured model (Rudd and Thorson, 2018). LBRA uses the mean length of the catch, one of the most common metabolic-based indicators 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), to calculate reference points that address sustainability risks (Ault *et al.*, 1998, 2008, 2018).

These methods derive the spawning potential ratio (SPR) and F/FMSY. SPR is defined as the proportion of the unfished reproductive potential left at 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). 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). F/FMSY is a ratio-based reference point of current fishing mortality to the level that would generate maximum sustainable yield (FMSY). While the use of either SPR and F/FMSY is still utilized in many assessments around the world, the common use of SPR is to indicate recruitment overfishing and F/FSMY is to indicate growth overfishing.

The above-mentioned methods are used by different authors and in different contexts, however a comprehensive performance evaluation is necessary to evaluate which methods perform best in various circumstances. Thus, a primary objective is to find the best methods for each configuration of fisheries give an understanding of their behaviour and allow 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, giving a better understanding of uncertainties. However, this is often not the case for data-limited stocks and conducting a stock assessment is significantly more challenging in these fisheries. Therefore, a performance evaluation of the multitude of data-limited length-based methods is instrumental to determine and quantify their precision and bias under various assumptions/scenarios.

Here we analyse four 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 understand the applicability of these methods. 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) and thus provide guidance about model applicability and limitations to stock assessment scientists and managers.

1. Methods

We conducted a simulation-estimation analysis using an individual-based population model (IBM) as the operating model, which simulated population dynamics and generated length composition data. The “true” input parameters for this study were used in the operating models, and then assumed known in the length-based models. This simulation loop allows us to compare how far the outputs of the assessment models are from the “true” stock status estimates and investigate the sensitivities of the models.

* 1. Operating model

The stock dynamics were simulated using the “fishdynr” R package (Taylor, 2017), which 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, selectivity, and recruitment are outlined by Taylor and Mildenberger (2017). Functions and equations for the population dynamics used in the operating model are listed in Appendix 1.

Seven scenarios were simulated based on variations in life history, fishing exploitation level, and recruitment: (1) the base model comprised of a medium-lived species (18 years), an exploitation rate at the target level of SPR (SPR ≈ 40%), and constant recruitment with no recruitment variability. From this base model, we varied one of the three characteristics – life history, current exploitation status, and recruitment. We tested a (2) short-lived (4 years) species, a (3) longer-lived (26 years) species, (4) a state of overexploitation, (5) a state of underexploitation, (6) constant recruitment with stochastic error (*σR* = 0.4537; Thorson, in press), and (7) autocorrelated recruitment with the associated autocorrelated error (*σR* = 0.737 and = 0.43; Thorson *et al.*, 2014b). The input values of the operating models are listed in Table 1.

For each scenario, 300 iterations, i.e. 300 length frequency data sets, were simulated, and a burn-in period with no fishing activity of 10 years was simulated 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. This reflects a one-year field phase to record landing, which is common in tropical artisanal fisheries (Tesfaye et al., 2016; Herrón et al., 2018; Tuda, 2018). All operating models assumed von Bertalanffy growth and logistic-type selectivity and maturity. The overview of this study is depicted in Figure 2. Three life histories 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). Examples of each life history are shown in Figure1 as length frequency graphs for one iteration. All simulations and analyses were conducted using the statistical programming language R (R Core Team, 2018).

* 1. Estimation models

The estimation models refer to the length-based methods that derive estimates of stock status from simulated data. The input life history values, *L∞*, *K*, *M*, and , were not estimated as the objective of this paper focused on the sensitivities of the methods, and therefore the same values that were used in the operating models were applied to each of the methods as inputs. Additionally, the selectivity values and , were calculated using the Length-Converted Catch Curve (LCCC) from TropFishR and used as the selectivity inputs for all estimation models. Four length-based methods were analysed: (1) TB, (2), LBSPR, (3) LIME, and (4) LBRA. 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). The inputs, assumptions, and outputs of all the methods, including the LCCC, are listed in Appendix 2.

3.3 Performance measures

The performance of the EM’s under seven different scenarios was compared to the simulated “truth” in which the “true” SPR and F/FMSY values were based on life history and obtained by pushing the IBM 100 years forward. We measured performance of the estimation models based on relative error of each simulation replicate of SPR and F/FMSY, calculating bias as the median relative error (MRE) and precision as the median absolute relative error (MARE):

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 and give indication to accuracy. We would interpret a model as more accurate than another if the bias and precision values are closer to zero.

1. Results

The results of the four methods (TB, LBSPR, LIME, and LBRA) are depicted as violin plots in Figure 3 for SPR and Figure 4 for F/FMSY and in bias/precision tables in Table 2 for SPR and Table 3 for F/FMSY. The fishing mortality, FMSY, SPRMSY were also analysed and the results are in Appendices 3-8. It should be noted again that the medium-lived scenario is the base model comprised of target level exploitation and no error in recruitment.

In the medium-lived and longer-lived scenarios on average, TB, LBSPR, and LIME overestimated SPR while LBRA on average underestimated SPR. LIME performed the best in the medium and longer-lived species with bias less than 5% and precision less than 10%, while LBRA was the most biased (bias of -0.159 in medium-lived; bias of -0.215 in longer-lived) and imprecise (precision of 0.159 in medium-lived; precision of 0.215 in longer-lived). On the contrary for the medium-lived, TB, LBSPR, and LIME underestimated F/FMSY while LBRA overestimated F/FMSY. In this case, TB was the least biased (-0.091) and most precise (0.114) and LIME was less accurate (bias of -0.334, precision of 0.334). For the longer-lived, TB and LBSPR were negatively biased, although TB was the closest to the truth with a bias of -0.022, while LIME and LBRA were positively biased. LBRA was the most biased and imprecise (bias and precision of 0.343) out of the four methods in the longer-lived scenario. All the methods were positively biased in the short-lived scenario in estimating SPR, although TB estimated SPR to be almost zero (bias of 0.016 and precision of 0.043) while LBSPR was the furthest from the truth (bias of 0.188, precision of 0.187). The performance of F/FMSY varied between methods where TB and LBSPR were positively biased and LIME and LBRA were negatively biased. Out of the four, LBRA was the least biased (-0.066) and most precise (0.098) while LIME was the most biased (-0.684) and least precise (0.098).

When looking at different exploitation levels, the performance of the methods varied between estimating SPR and F/FMSY. LBSPR and LIME overestimated SPR when the stocks were underexploited with LIME performing the worst (bias and precision of 0.424) while TB and LBRA underestimated SPR with TB performing the best (bias of -0.023 and precision of 0.066). In estimating F/FMSY, the methods generally were over 30% more biased and imprecise where LBSPR was the least biased (-0.022) and most precise (0.145) and LBRA was the most biased and imprecise (1.484). When the stocks were overexploited, TB, LBSPR, and LIME were positively biased in estimating SPR with LIME being the most biased and imprecise (0.560) and LBSPR being the least biased (0.035) and most precise (0.044). TB and LIME on average underestimated F/FMSY with TB closest to the truth (bias of -0.018 and precision of 0.075) and LBSPR and LBRA on average overestimated F/FMSY with LBRA furthest from the truth (bias and precision of 0.505).

The four methods were generally more imprecise in estimating SPR and F/FMSY as more recruitment error was included (i.e. stochastic error then autocorrelated pattern and error). In the stochastic scenario, all the methods on average underestimated SPR with all having less than 5% bias and TB performing the best and LBRA performing the worst (bias of -0.377 and precision of 0.377), although TB, LBSPR, and LIME in terms of bias were close to the zero. All but LIME overestimated F/FMSY with LBSPR being the least biased (bias of 0.105), TB being the most precise (precision of 0.197) and LBRA furthest away from the truth (bias and precision of 0.867). When autocorrelated recruitment was included, TB, LBSPR, and LIME were positively biased in estimating SPR with LBSPR performing the best (bias of 0.076 and precision of 0.146), while LBRA was negatively biased (-0.315) and most inaccurate. On the contrary for F/FMSY, TB, LBSPR, and LIME were negatively biased with LBSPR closest to the truth (bias of -0.089) and LIME furthest from the truth (bias of -0.644) and LBRA was positively biased.

TB and LBSPR were less than 30% biased and precise across scenarios. TB was the least biased and imprecise in the short-lived, underexploited and stochastic scenarios out of the four methods in estimating SPR. When estimating F/FMSY, TB performed the best in the base and overexploited scenarios. LBSPR was the closest to the truth in estimating SPR in the overexploited and autocorrelated scenarios while it was the closest in estimating F/FMSY in the underexploited, stochastic, and autocorrelated scenarios. There were few cases where LIME was the least biased and most precise when estimating SPR (base and longer-lived), however it also performed the worst in estimating SPR (under- and overexploited) and F/FMSY (short-lived, base and autocorrelated) in some scenarios. Lastly, LBRA, although it did perform well in the short-lived scenario when estimating F/FMSY, was the furthest from the truth in many scenarios when estimating SPR (base, longer-lived, stochastic and autocorrelated) and F/FMSY (longer-lived, underexploited, overexploited and stochastic).

1. Discussion

In this study, we used simulation-estimation analysis to compare the performance of different length-based stock assessment methods under different life history traits, exploitation regimes, and recruitment errors and types. Although the analysis does not focus around the performance of absolute reference points (i.e. fishing mortality, FMSY, and SPRMSY) and only the relative ones (SPR and F/FMSY), some aspects of the performance of the absolute reference points are discussed as some of them have influence on F/FMSY. We found that all these methods had higher precision in estimating relative references points than absolute ones, meaning that these methods are best in indicating stock status through relative references points and using absolute ones (e.g. exploitation rate) will lead to imprecise estimation of stock status and should be used with caution.

It is evident in this study that various life histories have an impact on the performance of the methods. Overall, all four methods had difficulties in assessing short-lived stocks, especially in the absolute values. Usually stock assessment methods do not perform well with short-lived species because the main problem with their assessment 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 contrary, individuals were then dispersed across more length bins for the longer-lived as evident in Figure 1. 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 performance of the longer-lived species in this study does not necessarily reflect this information, however this is evident in the overexploited stocks as the model specified low SPR levels.

While the performance of different exploitation scenarios varied between reference points, it is evident that stocks that are either under- or overexploited are more difficult to assess than those that are exploited and around the target exploitation level. Specifically, we found that when the stocks are severely overexploited, the methods are less accurate in estimating SPR, the degree of recruitment overfishing. When the stocks are severely underexploited, the methods present inconsistencies in determining F/FMSY, the degree of growth overfishing. In the case of calculating F/FMSY, the error stems from the calculation of fishing mortality as seen in Appendices 3 and 4. When the “true” values of SPR or fishing mortality are low, the four methods have difficulties assessing these estimates. 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 increasing recruitment error were also evident in all the methods evaluated where their precision decreased in the stochastic and autocorrelated scenarios. Although the precision decreased, the bias in most of the methods decreased from implementing no recruitment error (base model). Many stock assessment models, including 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). Thus, it is important to consider stochasticity in recruitment when conducting stock assessments, and including additional recruitment error does decrease precision, but it does not necessarily increase bias. Since these methods were relatively unbiased, a recommendation could be to use these methods initially despite the underlying uncertainties about the recruitment error and type. For fisheries that are known to be non-equilibrium conditions or in later assessments, it is possible to run these four methods at the same time to see if they give consistent results to determine what the prevailing condition (low or high recruitment variability) is.

Overall, TB was the most consistent in its performance across the seven scenarios. TB can calculate estimates at specified intervals of fishing mortality, which is why the calculation of the current estimates (current fishing mortality and SPR) and reference levels (FMSY and SPRMSY) were unbiased and precise. While a common practice of using TB is to use the length-based Jones’ cohort analysis beforehand, however this approach adds error from the cohort analysis to TB and therefore it is recommended to instead calculate fishing mortality (calculated from the LCCC) per length class using prior information about the selectivity. It should be noted however that TB estimates FMAX, which is the fishing mortality at which yield per recruit is maximized. There is a common practice of linking FMAX to FMSY which is based on the assumption that recruitment is independent of spawner stock size for fishing mortalities between 0 and FMAX (Reynolds, 2001). This assumption is usually invalid for most species, however in this study, the assumption behind the IBM matches this practice, and thus FMAX was used in place of FMSY.

Rudd and Thorson (Rudd and Thorson, 2018) also found that LBSPR performed better when the stocks were in equilibrium and when the operating model matched LBSPR’s assumptions; this result is supported here given that the IBM assumptions included constant recruitment (except in the autocorrelated scenario) and fishing. LBSPR was also relatively robust in this study as natural mortality was fixed (M/k input), which increases precision in the calculation of F/M and thus SPR and F/FMSY. Additionally, LBSPR was found to be accurate when the underlying selectivity is asymptotic (Hordyk *et al.*, 2015b; Rudd and Thorson, 2018; Pons *et al.*, 2019). While LBSPR calculated the relative reference points well, it however did not calculate FMSY and SPRMSY accurately even though the bias was not high, and the precision was low. This may be the case as the optimization process uses the LBSPR model rather than looking across a range of fishing mortalities like TB does. Like TB, LBSPR was relatively accurate and consistent across scenarios.

While we found that LIME performed better in the medium- and longer-lived scenarios, Rudd and Thorson (2018) found that LIME performed best in the short-lived scenario and worst in the longer-lived scenario. This could be the case because different operating models were used between the two studies. As explained previously, LIME decreases in accuracy when either the stocks are severely overexploited or when analysing long-lived species with low SPR. Additionally out of the four scenarios, LIME is highly biased and imprecise when estimating F/FMSY,especially in the short-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. In this study, we only violated the equilibrium assumption in one of the scenarios, and thus the performance of LIME was not fully explored. However, this study proves that LIME can still give relatively unbiased and precise answer in several scenarios.

LBRA on average consistently underestimated SPR while it overestimated F/FMSY, and thus was the most biased and least precise method out of the four. 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 and overestimation of F/FMSY. This was especially evident in the medium-lived and longer-lived scenarios (Figures 3 and 4) where more length classes were cut off (Figure 1). It should be noted however that despite its overall performance, LBRA can give better estimates with short-lived species, almost matching in performance with TB even in estimating the absolute relative points. Although assessments with short-lived species are still more difficult in all length-based methods in this study, TB and LBRA could be used to give more accurate estimates.

In similar studies, the operating models used are identical to the assessment models, and assume all dynamic processes are fully understood. Therefore, using an alternative structured 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). Additionally, population-based methods (age- or length) make assumptions regarding length error at age to create length distribution within each age class (Cao *et al.*, 2016). This study has proven the merits of using an alternative operating model structure for simulation-estimation analyses and could be helpful for management strategy evaluations (MSE’s).

It is commonly known that life history inputs have a significant impact on the performance of length-based methods (Hordyk *et al.*, 2015b; Kokkalis *et al.*, 2016; Taylor and Mildenberger, 2017; Rudd and Thorson, 2018; Pons *et al.*, 2019), however this was not addressed in this study as the objective was to focus on the sensitivities of the methods themselves, ignoring issues associated with wrong inputs. Few suggestions have been made to improve estimation of life history parameters, such as using local literature, a database (i.e. FishLife; Thorson *et al.*, 2017) or a methodology that utilizes the data (ELEFAN; Pauly and Morgan, 1987; Taylor and Mildenberger, 2017). However, issues have been addressed with each of these suggestions and therefore either a Bayesian approach or an ensemble model (Anderson *et al.*, 2017; Rudd *et al.*, 2019) would be worth investigating to improve these growth and mortality estimates and combine these approaches. Additionally, an MSE could be conducted to determine how one strategy of collecting growth and mortality information or a combination of strategies would affect stock assessments in the long run.

One of the main assumptions of length-based methods is that recruitment is constant, however it is evident that this assumption is violated in many fisheries as the stock-recruit relationship is one of the most common uncertain population dynamics processes (Maunder, 2012). While there have been several studies that investigated this uncertainty (Isaac, 1990; Thorson *et al.*, 2014; Rudd and Thorson, 2018). Understanding recruitment is essential especially in tropical fisheries as its influence on stock assessment is not well known and seasonality plays a role. Some fisheries have different recruitment patterns (e.g. pulsed, autocorrelated), seasonal variation and modes (uni- vs. bimodal; Isaac, 1990). While this study lightly addressed the effects of including recruitment variability and many studies involving MSE’s investigate different levels of recruitment error, further studies should investigate how this may affect data collection and its quality in data-poor/-limited fisheries.

There are many recommendations given by many authors about the ideal sample size (Erzini, 1990; Gerritsent and McGrath, 2007; Hordyk *et al.*, 2015b), however the sample size is ultimately dependent on life history and the fishery selectivity. 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-limited areas as only a limited amount of data is usually collected. Ono et al. (2015) found that infrequent sampling over a longer period was more informative than frequent surveys covering a shorter period. Additionally, in data-limited areas, fisheries also may 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. Therefore, the sampling amount and time period are significant especially in data-limited 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/-limited fisheries are often strongly biased due to gear selectivity. The data generation of our simulation study, based on the assumption of asymptotic selectivity, is thus most often not reflecting conditions of tropical multi-gear 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. A study by Pons et al. (2019) investigated the influence of different gear selectivity from multiple fleets on length composition data of scombrids in the Atlantic Ocean, however it would still be worth investigating gear selectivity influences in length data from tropical fisheries and assess the sensitivities of these length-based methods.

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/-limited 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, e.g. an ensemble model (Anderson *et al.*, 2017). 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. 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 in data-limited fisheries to provide a cost-effective starting point to later begin long-term processes of data collection, assessment, and management.

1. Acknowledgements
2. References

Anderson, S. C., Cooper, A. B., Jensen, O. P., Minto, C., Thorson, J. T., Walsh, J. C., Afflerbach, J., Dickey-Collas, M., Kleisner, K. M., Longo, C., Osio, G. C., Ovando, D., Mosqueira, I., Rosenberg, A. A., and Selig, E.R*.* 2017. Improving estimates of population status and trend with superensemble models. Fish and Fisheries, 18: 732–741.

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., and Bryan, D. R*.* 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.

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.

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

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.

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). An annoted and illustrated catalogue of the grouper, rockcod, hind, coral grouper, and lyretail species known to date. *In* FAO Species Catalogue, p. 125. Rome.

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.

Isaac, V. J. 1990. The accuracy of some length-based methods for fish population studies. ICLARM contribution. International Center for Living Aquatic Resources Management, Manila, Philippines. 81 pp.

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.

Kokkalis, A., Eikeset, A. M., Thygesen, U. H., Steingrund, P., and Andersen, K. H. 2016. Estimating uncertainty of data limited stock assessments. ICES Journal of Marine Science: Journal du Conseil: fsw145.

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. 2012. Evaluating the stock–recruitment relationship and management reference points: Application to summer flounder (*Paralichthys dentatus*) in the U.S. mid-Atlantic. Fisheries Research, 125–126: 20–26.

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., McGilliard, C. R., Monnahan, C. C., Szuwalski, C. S., Valero, J. L., Vert-Pre, K. A., Whitten, A. R., and Punt, A. E*.* 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.

Pons, M., Kell, L., Rudd, M. B., Cope, J. M., Lucena Frédou, F., and Handling editor: Emory Anderson. 2019. Performance of length-based data-limited methods in a multifleet context: application to small tunas, mackerels, and bonitos in the Atlantic Ocean. ICES Journal of Marine Science. https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsz004/5307953.

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

Reynolds, J. D. (Ed). 2001. Conservation of exploited species. Conservation biology. Cambridge University Press, Cambridge, U.K.; New York. 524 pp.

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.

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.

Rudd, M. B., Thorson, J. T., Sagarese, S. R., and Handling editor: Anna Kuparinen. 2019. Ensemble models for data-poor assessment: accounting for uncertainty in life-history information. ICES Journal of Marine Science. https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsz012/5315645.

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

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.

Thorson, J. T. in press. Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model.

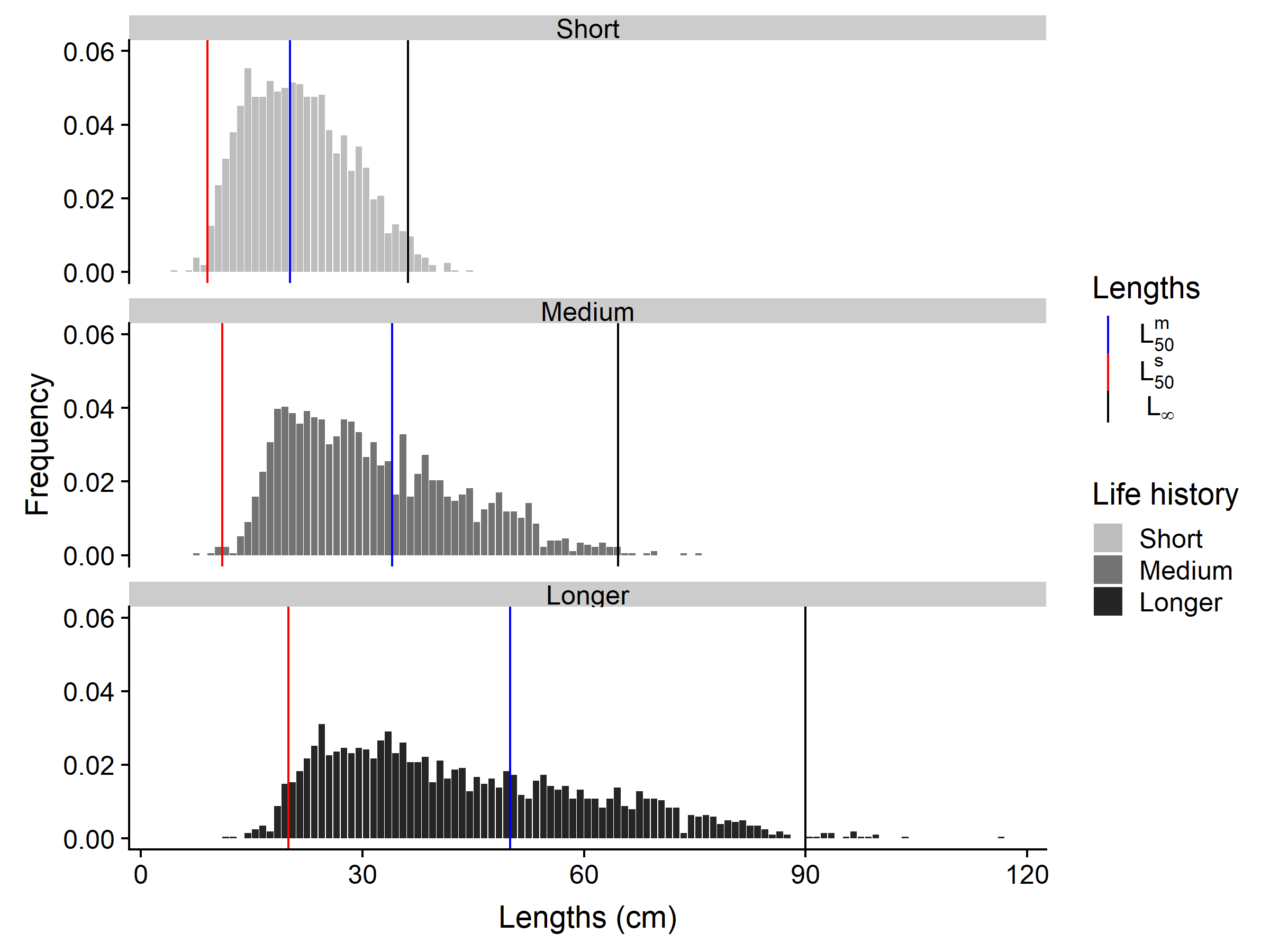
Thorson, J. T., Jensen, O. P., and Zipkin, E. F. 2014. How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. Canadian Journal of Fisheries and Aquatic Sciences, 71: 973–983.

Thorson, J. T., Munch, S. B., Cope, J. M., and Gao, J. 2017. Predicting life history parameters for all fishes worldwide. Ecological Applications, 27: 2262–2276.

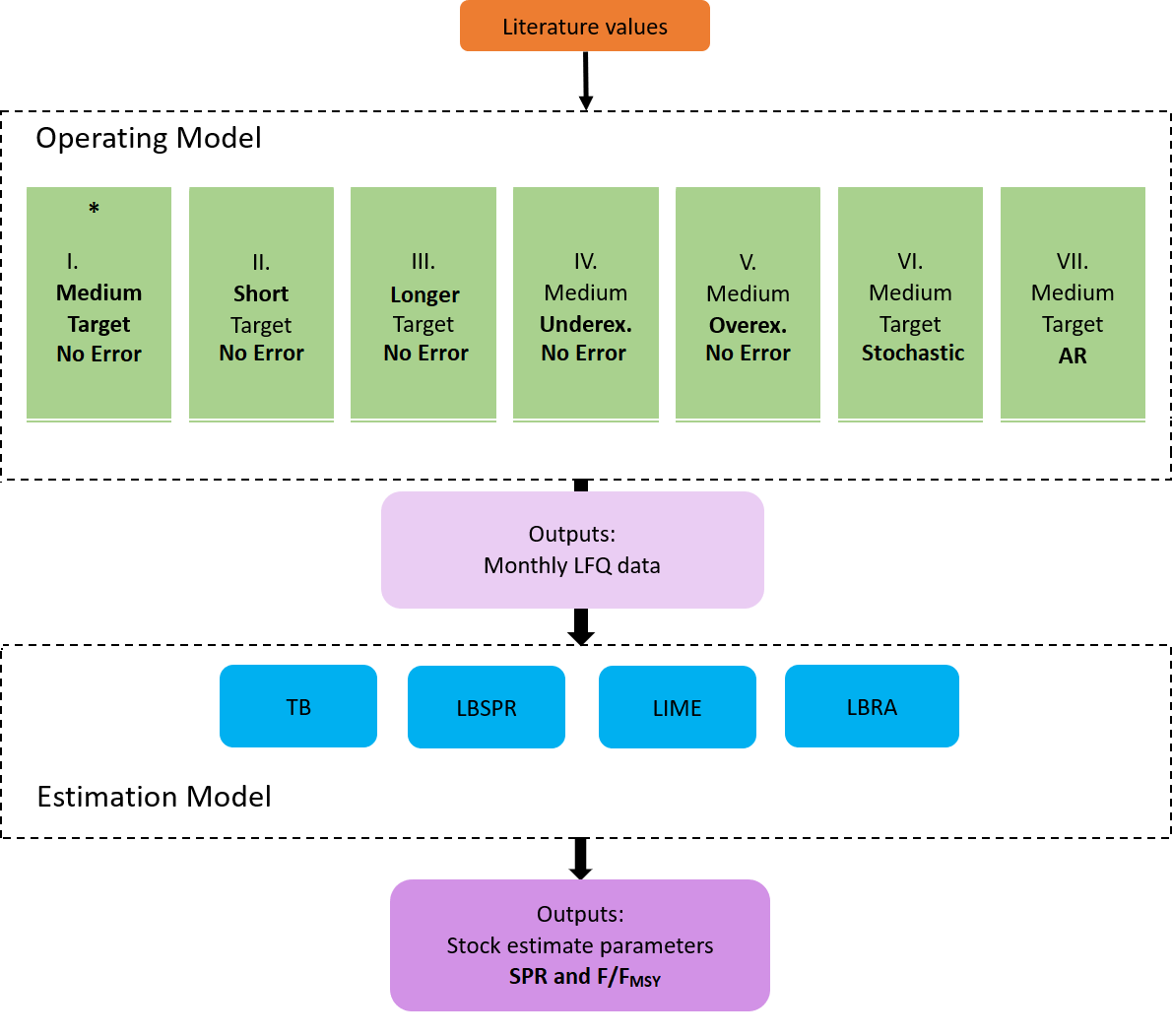
1. Figures and Tables

**Table 1.** Inputs values in generation of the operating models and length frequency data.The base model is comprised on the medium-lived species at target exploitation level with no error recruitment. From the base model, a change in life history (short- and longer-lived), exploitation level (under- and overexploited), and recruitment type (stochastic and autocorrelated) was made.

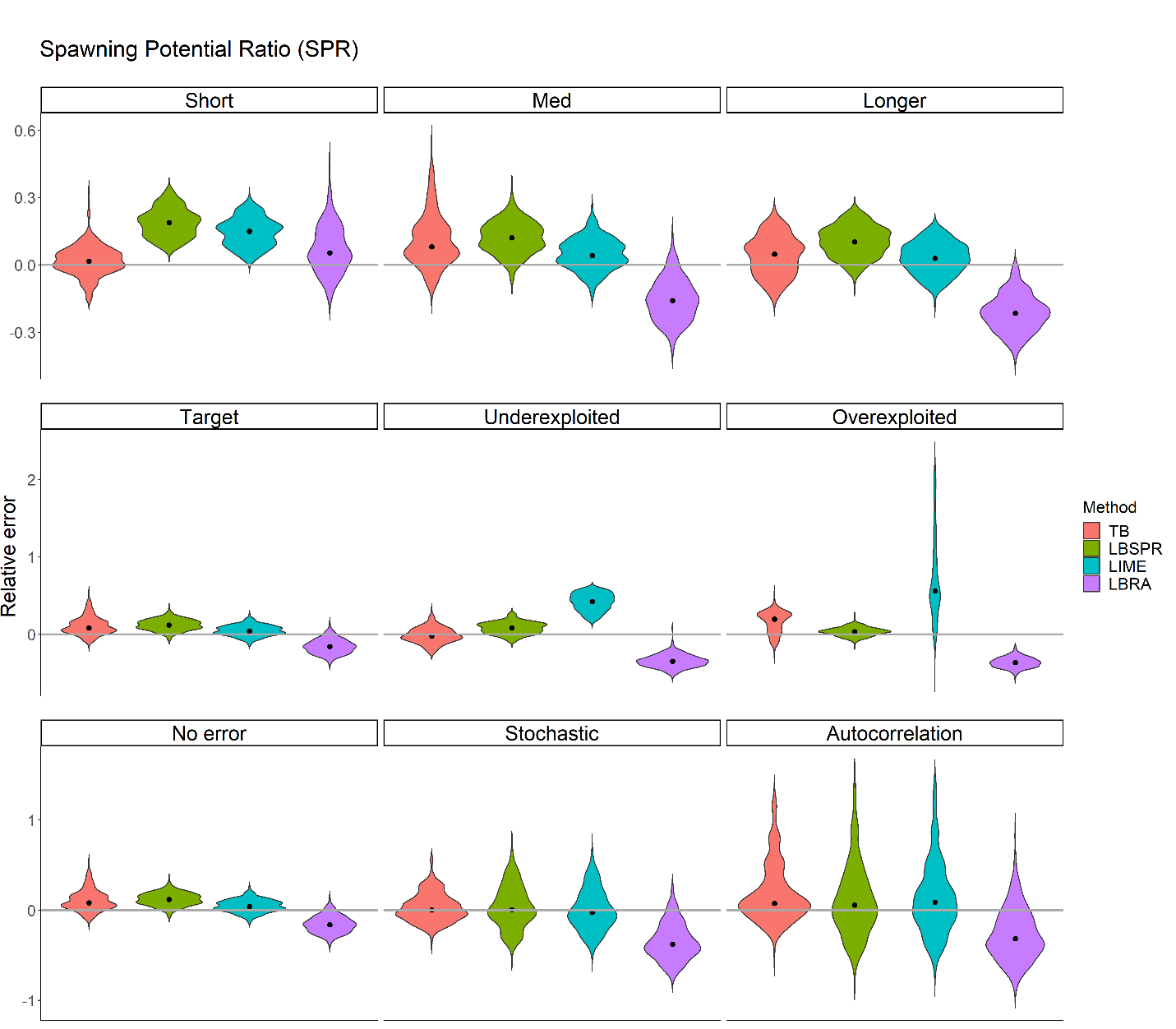
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | Life history | | | Exploitation level | | Recruitment | |
| Description | Symbol | | Base | Short | Longer | Underex. | Overex. | Stochastic | AR |
| Asymptotic length (cm) | |  | 64.6 | 36.2 | 90 | 64.6 | 64.6 | 64.6 | 64.6 |
| Growth coefficient | |  | 0.21 | 0.87 | 0.13 | 0.21 | 0.21 | 0.21 | 0.21 |
| Age at length = 0 | |  | -0.01 | -0.01 | -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| Natural mortality | |  | 0.32 | 0.9 | 0.18 | 0.32 | 0.32 | 0.32 | 0.32 |
| Fishing mortality | |  | 0.13 | 0.45 | 0.08 | 0.06 | 0.28 | 0.13 | 0.13 |
| Theoretical maximum age | |  | 18 | 4 | 26 | 18 | 18 | 18 | 18 |
| Length at 50% maturity | |  | 34 | 20.2 | 50 | 34 | 34 | 34 | 34 |
| Width of maturity ogive (cm) | |  | 6.8 | 4.04 | 10 | 6.8 | 6.8 | 6.8 | 6.8 |
| Recruitment standard deviation | |  | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.4537 | 0.737 |
| Length at 50% selectivity | |  | 11 | 9 | 20 | 11 | 11 | 11 | 11 |
| Width of selectivity ogive (cm) | | wqs | 2.2 | 1.8 | 4 | 2.2 | 2.2 | 2.2 | 2.2 |
| Bin size | | 2 | 2 | 1 | 3 | 2 | 2 | 2 | 2 |



**Figure 1.** Length frequency distribution graphs**.** For each life history scenario (short, medium, longer), , , and L∞ are visualised (in blue, red, and black respectively) over the length frequency graphs for one iteration.



***Figure 2.*** Simulation study methodology diagram. There are seven operating model setups. Scenarios differ in life history with (1) medium-lived (base model\*), (2) short-lived, (3), and longer-lived stocks. All fish longevity simulations were run with constant recruitment and exploitation at target level (SPR ~ 0.4). From the base model, the exploitation scenario, (4) underexploited and (5) overexploited) or the recruitment scenario, (6) stochastic and (7) autocorrelation, changes; 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 (*L∞, K, M,* ) from the operating models and the length frequency data were then used as input for the simulated assessment with the four length-based estimation models: (1) length-based Thompson and Bell (TB), (2) length-based spawning potential ratio (LBSPR), (3) length-based integrated mixed effects (LIME), and (4) length-based risk analysis (LBRA).

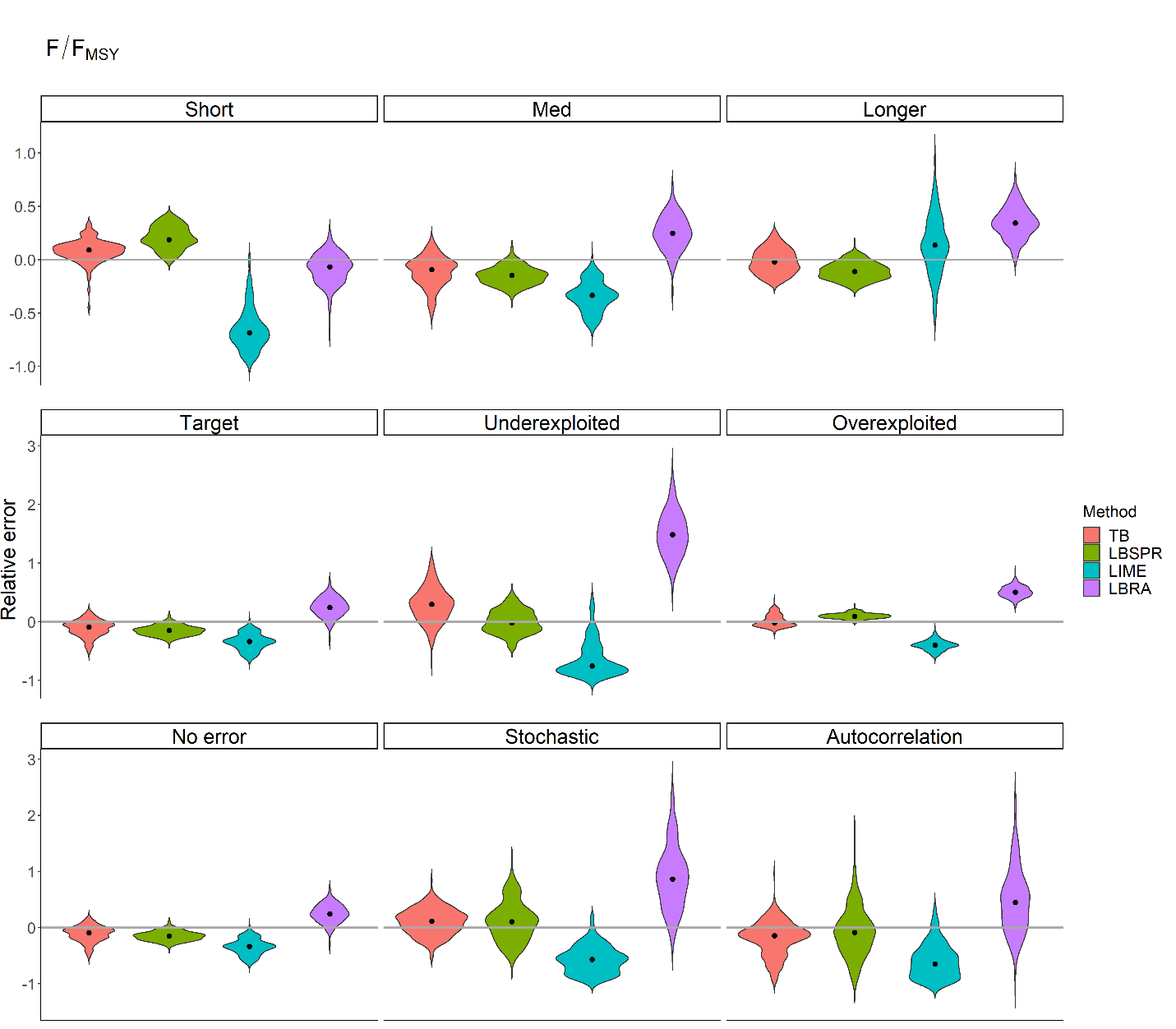


**Figure 3.** Violin plots of relative error for SPR for 300 iterations per scenario with 200 individuals per month for one year across three life histories (short-, medium-, and longer-lived), three exploitation levels (target, under-, and overexploited), and three recruitments (no error, stochastic error, and autocorrelation pattern and error). Four methods (Length-based Thompson and Bell (TB), Length-Based Spawning potential ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), and Length-Based Risk Analysis (LBRA)) were analysed. The grey horizontal line is the zero relative error line, and the black dot is the median relative error indicating bias. Each plot has a different y-axis range with a smoother tail.

**Table 2.** Bias (MRE) and precision (MARE) table of SPR from length-based Thompson and Bell (TB), Length-Based Spawning potential ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), and Length-based Risk Analysis (LBRA) performance across life histories, exploitation levels, and recruitment types. The lightest red colour indicates bias/precision less than 5%, and the darkest red colour indicates bias/precision greater than 30%.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Life history | | | |  | Exploitation level | | | |  | Recruitment | | |
|  | Short | Medium\* | Longer | |  | Target\* | Under-exploited | | Over-exploited |  | No Error\* | Stochastic | Auto-correlation |
| Bias (MRE) | | | |  | | | |  | | | | | |
| TB | 0.016 | 0.081 | 0.048 | |  | 0.081 | -0.023 | | 0.197 |  | 0.081 | -0.003 | 0.076 |
| LBSPR | 0.188 | 0.121 | 0.102 | |  | 0.121 | 0.081 | | 0.035 |  | 0.121 | -0.007 | 0.056 |
| LIME | 0.149 | 0.042 | 0.029 | |  | 0.042 | 0.424 | | 0.560 |  | 0.042 | -0.023 | 0.089 |
| LBRA | 0.054 | -0.159 | -0.215 | |  | -0.159 | -0.350 | | -0.364 |  | -0.159 | -0.377 | -0.315 |
| Precision (MARE) | | | |  | | | |  | | | | | |
| TB | 0.043 | 0.084 | 0.073 | |  | 0.084 | 0.066 | | 0.197 |  | 0.084 | 0.099 | 0.146 |
| LBSPR | 0.187 | 0.121 | 0.102 | |  | 0.121 | 0.082 | | 0.044 |  | 0.121 | 0.150 | 0.218 |
| LIME | 0.149 | 0.062 | 0.057 | |  | 0.062 | 0.424 | | 0.560 |  | 0.062 | 0.163 | 0.226 |
| LBRA | 0.074 | 0.159 | 0.215 | |  | 0.159 | 0.350 | | 0.364 |  | 0.159 | 0.377 | 0.337 |
|  |  |  |  | |  |  |  | |  |  |  |  |  |

\* These are components of the base model, and thus are of a single scenario.



**Figure 4.** Violin plots of relative error for F/FMSY for 300 iterations per scenario with 200 individuals per month for one year across three life histories (short-, medium-, and longer-lived), three exploitation levels (target, under-, and overexploited), and three recruitments (no error, stochastic error, and autocorrelation pattern and error). Four methods (Length-based Thompson and Bell (TB), Length-Based Spawning potential ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), and Length-Based Risk Analysis (LBRA)) were analysed. The grey horizontal line is the zero relative error line, and the black dot is the median relative error indicating bias. Each plot has a different y-axis range with a smoother tail.

**Table 3.** Bias (MRE) and precision (MARE) table of F/FMSY from length-based Thompson and Bell (TB), Length-Based Spawning potential ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), and Length-based Risk Analysis (LBRA) performance across life histories, exploitation levels, and recruitment types. The lightest red colour indicates bias/precision less than 5%, and the darkest red colour indicates bias/precision greater than 30%.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Life history | | | | |  | Exploitation level | | | |  | Recruitment | | |
|  | Short | Medium\* | | Longer | |  | Target\* | | Under-exploited | Over-exploited |  | No Error\* | Stochastic | Auto-correlation |
| Bias (MRE) | | | | |  | | |  | | | | | | |
| TB | 0.091 | | -0.091 | -0.022 | |  | -0.091 | | 0.298 | -0.018 |  | -0.091 | 0.112 | -0.142 |
| LBSPR | 0.189 | | -0.148 | -0.111 | |  | -0.148 | | -0.022 | 0.096 |  | -0.148 | 0.105 | -0.089 |
| LIME | -0.684 | | -0.334 | 0.138 | |  | -0.334 | | -0.752 | -0.400 |  | -0.334 | -0.568 | -0.644 |
| LBRA | -0.066 | | 0.247 | 0.343 | |  | 0.247 | | 1.484 | 0.505 |  | 0.247 | 0.867 | 0.448 |
| Precision (MARE) | | | | |  | | |  | | | | | | |
| TB | 0.096 | | 0.114 | 0.089 | |  | 0.114 | | 0.306 | 0.075 |  | 0.114 | 0.197 | 0.190 |
| LBSPR | 0.189 | | 0.148 | 0.111 | |  | 0.148 | | 0.145 | 0.096 |  | 0.148 | 0.223 | 0.282 |
| LIME | 0.684 | | 0.334 | 0.188 | |  | 0.334 | | 0.752 | 0.400 |  | 0.334 | 0.568 | 0.644 |
| LBRA | 0.098 | | 0.253 | 0.343 | |  | 0.253 | | 1.484 | 0.505 |  | 0.253 | 0.867 | 0.469 |

\* These are components of the base model, and thus are of a single scenario.

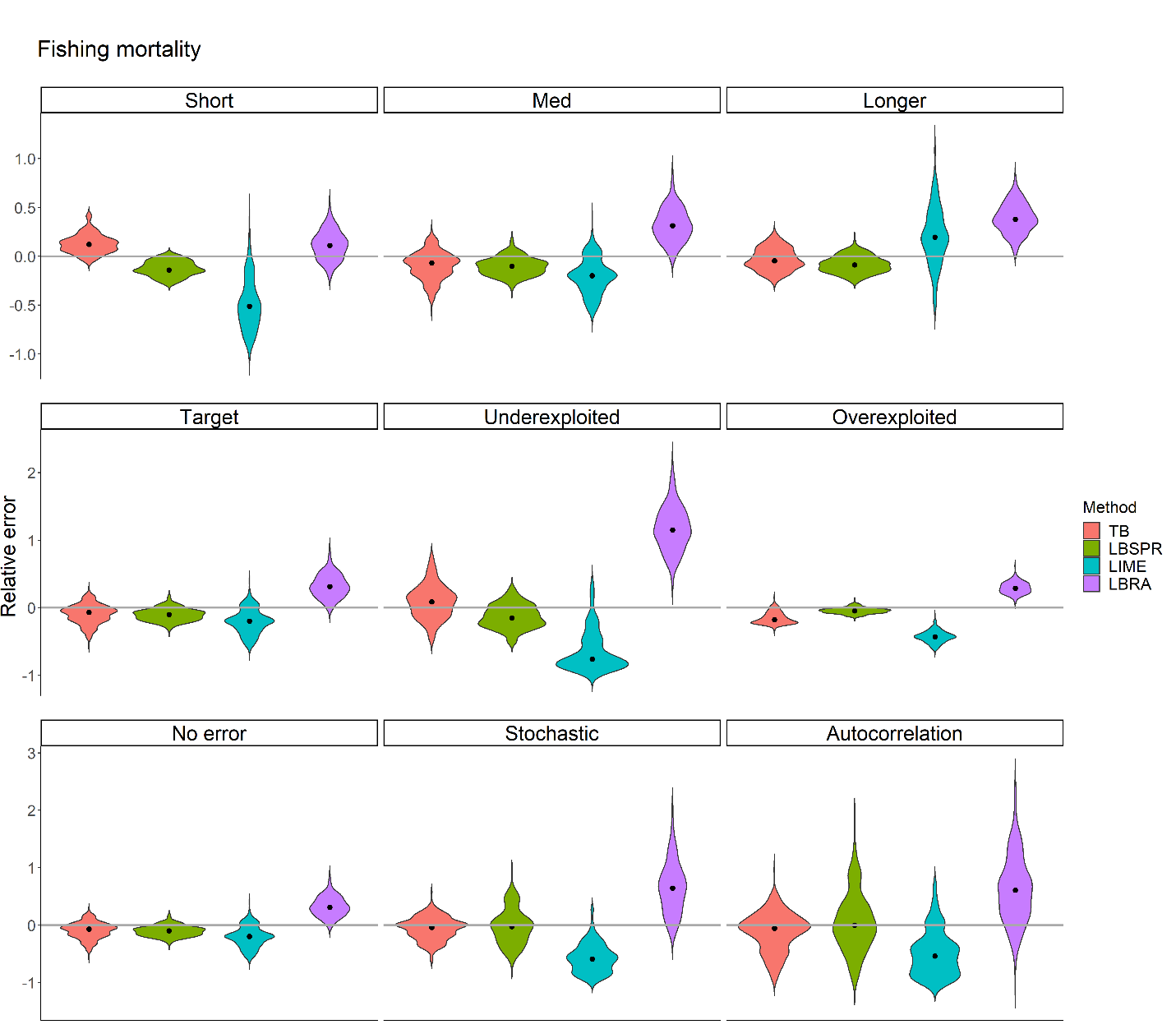
1. Appendix

**Appendix 1.** Functions and population dynamic equations used for generating stocks and length frequency data in the operating models.

|  |  |
| --- | --- |
| Function/Equation | Description |
|  | von Bertalanffy growth function |
|  | Variability in of von Bertalanffy growth function |
|  | Variability in *K* of von Bertalanffy growth function |
|  | Length-weight relationship |
|  | Selectivity/Maturity probability |
|  | Non-autocorrelated recruitment deviation |
|  | Autocorrelated recruitment deviation |
|  |  |
|  | Spawning stock biomass |
| ; beta = 1 (constant recruitment) | Beverton-Holt relationship |
|  | Total mortality |
|  | Fishing mortality |
|  | Probability of death |
| If , individual dies | Rand = random number generated |
|  | Probability of death due to either natural or fishing mortality (0 = M, 1 = F) |
|  | Unfished expected lifetime egg production |
|  | Fished expected lifetime egg production |
|  | Spawning potential ratio (SPR) |
| YF,Lc,t= FtBLc,t | Yield |

**Appendix 2.** Summary of methods. The data inputs, assumptions, and expected outputs are listed for each method including the length-converted catch curve. In the outputs, the estimates in bold are the estimated this study uses for the comparison.

|  |  |  |  |
| --- | --- | --- | --- |
| Method | Inputs | Assumptions | Outputs |
| Length-Converted Catch Curve (LCCC) (not part of the comparison) | * Length frequency data (yearly catch vector) * von Bertalanffy growth function () * Natural mortality () | (i) Total mortality is constant for all length classes  (ii) Selectivity follows logistic curve (width of curve calculated from and ) | * **Length at 50% and 95% selectivity ( and )** * **Total mortality () (used to calculate F)** |
| Length-based Thompson and Bell (TB) | * Length frequency data * von Bertalanffy growth function () * Length-weight relationship ( and ) * F-at-length-array (fishing mortality for each length class; calculated based on selectivity) * Natural mortality () * Total mortality () * Length at 50% selectivity and maturity ( and ) * Width of selectivity and maturity logistic curve | (i) Stock is in equilibrium  (ii) Natural mortality is constant  (iii) Selectivity and maturity follow logistic curve | * Precautionary reference levels (F0.1, F0.5, E0.5) * Exploitation, Yield, abundance and catch across vector of fishing mortalities * Current exploitation, yield, abundance and catch * **Current F** * **SPR** * **F/FMax or F/FMSY** * **SPRMSY** |
| Length-Based Spawning potential ratio (LBSPR) | * Length frequency data * Asymptotic length () * Coefficient of variation of () * (calculated from and ) * Length-weight relationship ( and ) * Length at 50% and 95% selectivity ( and ) * Length at 50% and 95% maturity ( and ) | (i) Stock is in equilibrium  (ii) Natural mortality and growth rates are constant  (iii) Selectivity and maturity follow a logistic curve  (iii) Both sexes have the same growth curve and the sex ratio is equal  (iv) The lengths at each age are normally distributed around a mean length-at-age value. | * F/M ratio * Length at 50% and 95% selectivity ( and ) * **F/M ratio (used to calculate F)** * **SPR** * **F/FMSY** * **SPRMSY** |
| Length-based Integrated Mixed Effects (LIME) | * Length composition data * von Bertalanffy growth function () * Length-weight relationship ( and ) * Natural mortality () * Length at 50% and 95% selectivity ( and ) * Length at 50% () | (i) Natural mortality is constant  (ii) Selectivity and maturity follow a logistic curve | * (Length data only) * Recruitment * Spawning biomass * Mean length * Length at 50% and 95% selectivity ( and ) * **Current F** * **SPR** * **F/FMSY** * **SPRMSY** |
| Length-Based Risk Analysis (LBRA) | * Length composition data * von Bertalanffy growth function () * Coefficient of variation of length at age () * Length-weight relationship ( and ) * Natural mortality () * Theoretical maximum age () | (i) Average annual constant recruitment  (ii) Selectivity and maturity follow a logistic curve  (iii) The lengths at each age are normally distributed around the mean length  (iv) The observed maximum age () deviates are described by the exponential probability density function (used to calculate ) | * B/BMSY * **Total mortality () (used to calculate fishing mortality ())** * **SPR** * **F/FMSY** * **SPRMSY** |

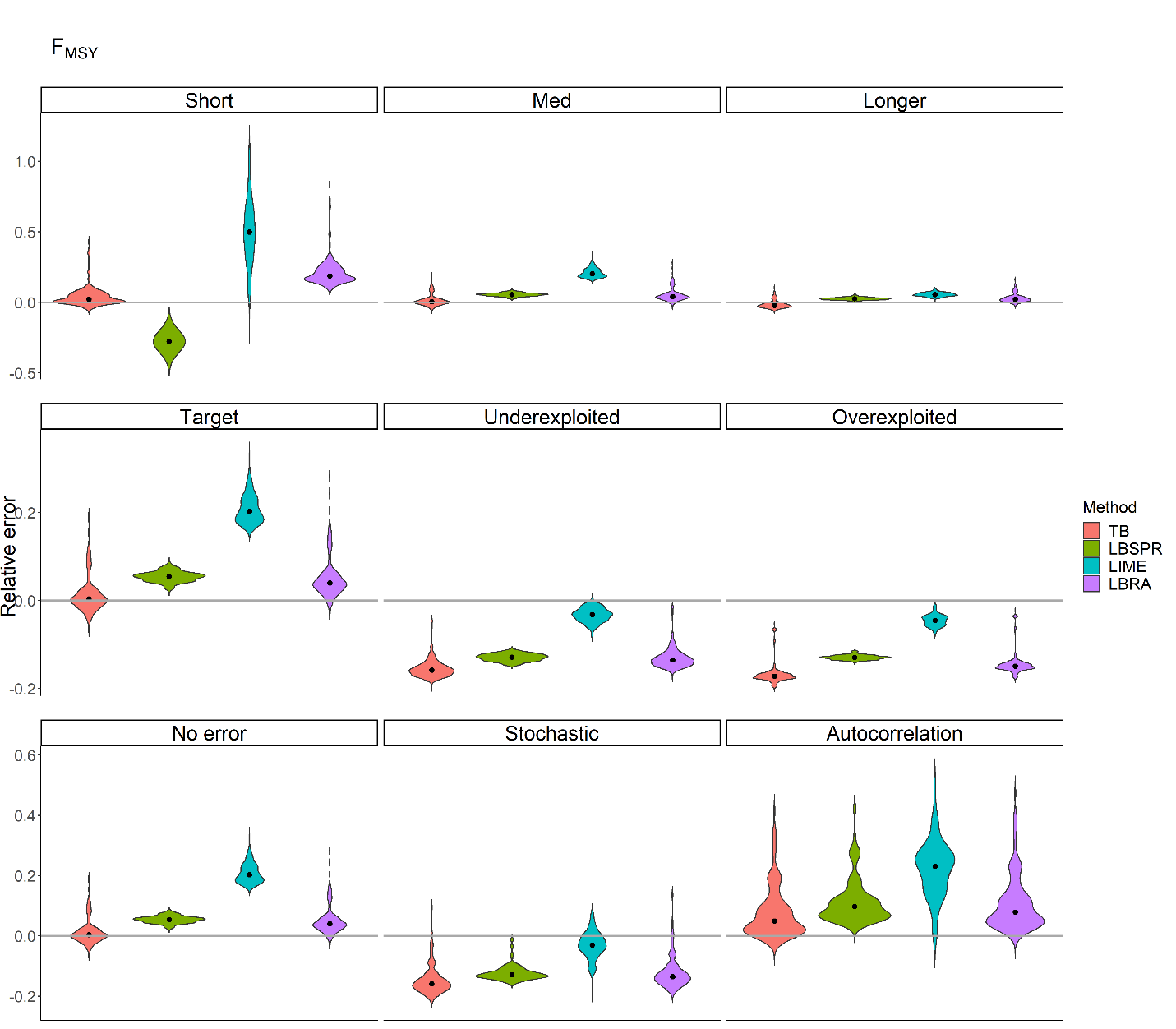


**Appendix 3.** Violin plots of relative error for fishing mortality for 300 iterations per scenario with 200 individuals per month for one year across three life histories (short-, medium-, and longer-lived), three exploitation levels (target, under-, and overexploited), and three recruitments (no error, stochastic error, and autocorrelation pattern and error). Four methods (Length-based Thompson and Bell (TB), Length-Based Spawning potential ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), and Length-Based Risk Analysis (LBRA)) were analysed. The grey horizontal line is the zero relative error line, and the black dot is the median relative error indicating bias. Each plot has a different y-axis range with a smoother tail.

**Appendix 4.** Bias (MRE) and precision (MARE) table of fishing mortality from length-based Thompson and Bell (TB), Length-Based Spawning potential ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), and Length-based Risk Analysis (LBRA) performance across life histories, exploitation levels, and recruitment types. The lightest red colour indicates bias/precision less than 5%, and the darkest red colour indicates bias/precision greater than 30%.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Life history | | | | |  | Exploitation level | | | |  | Recruitment | | |
|  | Short | Medium\* | | Longer | |  | Target\* | | Under-exploited | Over-exploited |  | No Error\* | Stochastic | Auto-correlation |
| Bias (MRE) | | | | |  | | |  | | | | | | |
| TB | 0.123 | | -0.069 | -0.045 | |  | -0.069 | | 0.088 | -0.178 |  | -0.069 | -0.038 | -0.055 |
| LBSPR | -0.140 | | -0.100 | -0.088 | |  | -0.100 | | -0.150 | -0.045 |  | -0.100 | -0.025 | 0.000 |
| LIME | -0.511 | | -0.198 | 0.194 | |  | -0.198 | | -0.762 | -0.432 |  | -0.198 | -0.592 | -0.541 |
| LBRA | 0.110 | | 0.312 | 0.379 | |  | 0.312 | | 1.152 | 0.289 |  | 0.312 | 0.642 | 0.607 |
| Precision (MARE) | | | | |  | | |  | | | | | | |
| TB | 0.123 | | 0.112 | 0.091 | |  | 0.112 | | 0.175 | 0.178 |  | 0.112 | 0.149 | 0.185 |
| LBSPR | 0.140 | | 0.100 | 0.088 | |  | 0.100 | | 0.150 | 0.045 |  | 0.100 | 0.200 | 0.300 |
| LIME | 0.511 | | 0.198 | 0.219 | |  | 0.198 | | 0.762 | 0.432 |  | 0.198 | 0.592 | 0.543 |
| LBRA | 0.126 | | 0.312 | 0.379 | |  | 0.312 | | 1.152 | 0.289 |  | 0.312 | 0.642 | 0.615 |
|  |  | |  |  | |  |  | |  |  |  |  |  |  |

\* These are components of the base model, and thus are of a single scenario.

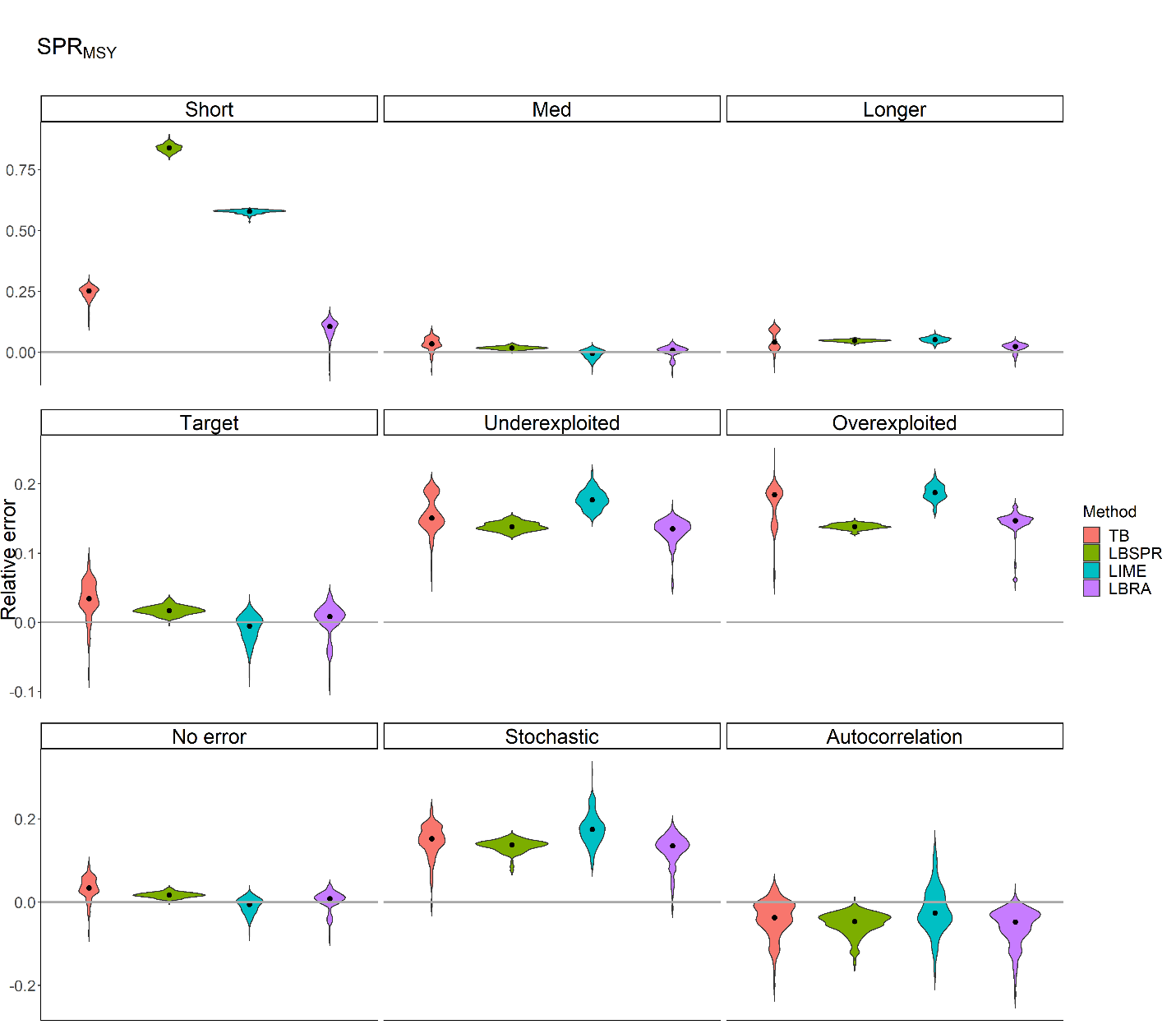


**Appendix 5.** Violin plots of relative error for FMSY for 300 iterations per scenario with 200 individuals per month for one year across three life histories (short-, medium-, and longer-lived), three exploitation levels (target, under-, and overexploited), and three recruitments (no error, stochastic error, and autocorrelation pattern and error). Four methods (Length-based Thompson and Bell (TB), Length-Based Spawning potential ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), and Length-Based Risk Analysis (LBRA)) were analysed. The grey horizontal line is the zero relative error line, and the black dot is the median relative error indicating bias. Each plot has a different y-axis range with a smoother tail.

**Appendix 6.** Bias (MRE) and precision (MARE) table of FMSY from length-based Thompson and Bell (TB), Length-Based Spawning potential ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), and Length-based Risk Analysis (LBRA) performance across life histories, exploitation levels, and recruitment types. The lightest red colour indicates bias/precision less than 5%, and the darkest red colour indicates bias/precision greater than 30%.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Life history | | | | |  | Exploitation level | | | |  | Recruitment | | |
|  | Short | Medium\* | | Longer | |  | Target\* | | Under-exploited | Over-exploited |  | No Error\* | Stochastic | Auto-correlation |
| Bias (MRE) | | | | |  | | |  | | | | | | |
| TB | 0.021 | | 0.004 | -0.022 | |  | 0.004 | | -0.158 | -0.172 |  | 0.004 | -0.159 | 0.159 |
| LBSPR | -0.277 | | 0.055 | 0.026 | |  | 0.055 | | -0.128 | -0.129 |  | 0.055 | -0.128 | 0.128 |
| LIME | 0.498 | | 0.203 | 0.055 | |  | 0.203 | | -0.032 | -0.045 |  | 0.203 | -0.030 | 0.231 |
| LBRA | 0.186 | | 0.040 | 0.023 | |  | 0.040 | | -0.135 | -0.149 |  | 0.040 | -0.135 | 0.136 |
| Precision (MARE) | | | | |  | | |  | | | | | | |
| TB | 0.027 | | 0.016 | 0.026 | |  | 0.016 | | 0.158 | 0.172 |  | 0.016 | 0.159 | 0.050 |
| LBSPR | 0.277 | | 0.055 | 0.026 | |  | 0.055 | | 0.128 | 0.129 |  | 0.055 | 0.128 | 0.097 |
| LIME | 0.498 | | 0.203 | 0.055 | |  | 0.203 | | 0.032 | 0.045 |  | 0.203 | 0.034 | 0.231 |
| LBRA | 0.186 | | 0.040 | 0.023 | |  | 0.040 | | 0.135 | 0.149 |  | 0.040 | 0.136 | 0.079 |

\* These are components of the base model, and thus are of a single scenario.



**Appendix 7.** Violin plots of relative error for SPRMSY for 300 iterations per scenario with 200 individuals per month for one year across three life histories (short-, medium-, and longer-lived), three exploitation levels (target, under-, and overexploited), and three recruitments (no error, stochastic error, and autocorrelation pattern and error). Four methods (Length-based Thompson and Bell (TB), Length-Based Spawning potential ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), and Length-Based Risk Analysis (LBRA)) were analysed. The grey horizontal line is the zero relative error line, and the black dot is the median relative error indicating bias. Each plot has a different y-axis range with a smoother tail.

**Appendix 8.** Bias (MRE) and precision (MARE) table of SPRMSY from length-based Thompson and Bell (TB), Length-Based Spawning potential ratio (LBSPR), Length-based Integrated Mixed Effects (LIME), and Length-based Risk Analysis (LBRA) performance across life histories, exploitation levels, and recruitment types. The lightest red colour indicates bias/precision less than 5%, and the darkest red colour indicates bias/precision greater than 30%.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Life history | | | | |  | Exploitation level | | | |  | Recruitment | | |
|  | Short | Medium\* | | Longer | |  | Target\* | | Under-exploited | Over-exploited |  | No Error\* | Stochastic | Auto-correlation |
| Bias (MRE) | | | | |  | | |  | | | | | | |
| TB | 0.251 | | 0.034 | 0.042 | |  | 0.034 | | 0.151 | 0.184 |  | 0.034 | 0.153 | -0.037 |
| LBSPR | 0.840 | | 0.017 | 0.047 | |  | 0.017 | | 0.138 | 0.139 |  | 0.017 | 0.138 | -0.047 |
| LIME | 0.579 | | -0.005 | 0.051 | |  | -0.005 | | 0.177 | 0.188 |  | -0.005 | 0.175 | -0.026 |
| LBRA | 0.106 | | 0.008 | 0.024 | |  | 0.008 | | 0.135 | 0.147 |  | 0.008 | 0.135 | -0.048 |
| Precision (MARE) | | | | |  | | |  | | | | | | |
| TB | 0.251 | | 0.036 | 0.045 | |  | 0.036 | | 0.151 | 0.184 |  | 0.036 | 0.153 | 0.037 |
| LBSPR | 0.840 | | 0.017 | 0.047 | |  | 0.017 | | 0.138 | 0.139 |  | 0.017 | 0.138 | 0.047 |
| LIME | 0.579 | | 0.011 | 0.051 | |  | 0.011 | | 0.177 | 0.188 |  | 0.011 | 0.175 | 0.038 |
| LBRA | 0.106 | | 0.014 | 0.025 | |  | 0.014 | | 0.135 | 0.147 |  | 0.014 | 0.135 | 0.048 |

\* These are components of the base model, and thus are of a single scenario.