Biomass limit reference points are sensitive to estimation method, time-series length and stock development

van Deurs, Mikael; Brooks, Mollie Elizabeth; Lindegren, Martin ; Henriksen, Ole; Rindorf, Anna

Published in:
Fish and Fisheries

Link to article, DOI:
10.1111/faf.12503

Publication date:
2021

Document Version
Peer reviewed version

Citation (APA):
van Deurs, M., Brooks, M. E., Lindegren, M., Henriksen, O., & Rindorf, A. (2021). Biomass limit reference points are sensitive to estimation method, time-series length and stock development. *Fish and Fisheries*, *22*(1), 18-30. https://doi.org/10.1111/faf.12503

Title: Biomass limit reference points are sensitive to estimation method, time-series length, and stock development

Running title: Biomass limit reference points

Authors: Mikael van Deurs[†¶], Mollie E. Brooks[†¶], Martin Lindegren[†], Ole Henriksen[†], Anna Rindorf[†]

[¶] Mikael van Deurs and Mollie E. Brooks should be considered joint first author

[†]Technical University of Denmark, Institute for Aquatic Resources, Kemitorvet 1, 2800 Lyngby, Denmark

Corresponding author: Mikael van Deurs, Technical University of Denmark, Institute for Aquatic Resources, Kemitorvet 1, 2800 Lyngby, Denmark. E-mail: mvd@aqua.dtu.dk)

Running title: Biomass limit reference points

Abstract: Biomass limit reference points are widely used in fisheries management and defines the biomass threshold below which stock productivity (i.e. recruitment) is likely to be impaired. Scientifically sound and transparent methods for estimating biomass thresholds are therefore needed together with ways of quantifying uncertainties. The main focus of the study was placed on two methods currently applied to several small-bodied pelagic species in the Northeast Atlantic. These methods have not formerly been described in the scientific literature and are in the present study being compared to some already described methods, of which, one is broadly applied outside the Northeast Atlantic. Using a combination of data simulations and data from 51 small-bodied pelagic fish stocks, we analyzed the sensitivity of estimated biomass thresholds to (i) the choice of method, (ii) time-series length, and (iii) stock development (e.g. rebuilding or declining). It was demonstrated that estimated biomass thresholds are associated with considerable uncertainty not previously quantified. Furthermore, the level of the estimated threshold and the amount of uncertainty depended on choice of method, time-series length, and stock development trends. Hence, this study contributes to improving the quality of future biomass limit reference points by providing guidance regarding choice of method and how to demonstrate stock-specific uncertainties.

Key words: Fisheries management, stock-recruitment, statistical uncertainty, small pelagics, RAM Legacy, sustainable exploitation

Table of content:

1. Introduction

2. Materials and Methods

2.1. Methods for estimating biomass thresholds

2.2. Simulation study

2.3. Data-driven study

2.3.1. Stock assessment data

2.3.2. Comparisons of methods

2.3.3. Effects of time-series length and stock development

2.3.4. Analysis of case-study stocks

3. Results

3.1. Simulation study

3.2. Data-driven analysis

3.2.1. Comparison of methods

3.2.2. Effects of time-series length and stock development

3.2.3. Case study stocks

4. Discussion

1. Introduction

Fisheries management worldwide uses a variety of biomass limit reference points to ensure a sustainable exploitation of fish stocks (Foley et al. 2015). Biomass limit reference points based on stock-recruitment (SR) relationships originate from the expectation that when the biomass of mature and reproducing individuals (i.e. spawners) falls below a certain biomass threshold (BT), recruitment and stock productivity is impaired and the risk of stock collapse increases (e.g. Sissenwine & Shepherd, 1987; Mace, 1994; Nakatsuka et al., 2017). This assumption is supported by the observation that many stocks, which have experienced low spawning stock biomass also suffered from poor recruitment (Myers et al., 1994; Barrowman & Myers, 1999).

BTs based on SR relationships, in one or the other way, are particularly important for management of small-bodied pelagic species. Maximum sustainable yield of these species is often attained using either the so-called "escapement-strategy", where the management objective is to maintain a fishing pressure resulting in a low probability that spawning stock biomass will fall to a level where recruitment is impaired (i.e. below BT), or harvest control rules, where so-called "biomass triggers" guide shifts in fishing pressure to avoid impaired recruitment (Deroba & Bence, 2008; Gjøsæter et al., 2014, Pikitch, 2015). While it is often assumed that maintaining a fishing pressure capable of providing F_{MSY} (i.e. the fishing mortality resulting in maximum sustainable yield) does not place recruitment at risk, this is not necessarily the case for small-bodied pelagic species. It has, for example, been shown that F_{MSY} is either above or close to the fishing mortality leading to a 5% risk of impaired recruitment in three out of four clupeids stocks in the North Sea and Baltic Sea (Rindorf et al., 2017).

The SR relationship is often weak, and for the majority of stocks, there is either no apparent "best" relationship or the parameters of the selected functional SR model is highly time varying (Ricard et al., 2012, Subbey et al., 2014; Szuwalski et al., 2015; Szuwalski et al.,

2019; Zhou, 2007). Because of this ambiguity about the SR relationship, several simplified nonparametric (or semi-parametric) methods for estimating BTs have been suggested.

Presently, in particular, two types of approaches are applied to stocks in the Northeast Atlantic region by the International Council for the Exploration of the Sea (ICES): (i) a nonparametric approach where BT is equal to "the lowest spawning stock biomass where large recruitment is observed" (with no strict default definition of "large recruitment") and (ii) a simple parametric approach, where BT is equal to the spawning stock biomass at the breakpoint in a segmented regression (i.e. a hockey stick model) fitted to the SR data (ICES, 2017). These methods have not formerly been described in the scientific literature, but documentation can be found in ICES (2017).

In contrast to the above methods, alternative semi-parametric approaches, relying on assumptions about the underlying functional SR relationship, have previously been suggested in the literature. Examples of such approaches are those where BT is equal to the spawning stock biomass producing 50% of maximum recruitment predicted by a Ricker or Beverton-Holt model (Ricker, 1954; Myers et al., 1994; Beverton & Holt, 1957).

In the case of the approaches used by ICES in the Northeast Atlantic for small pelagics, simple assumptions about the SR relationship are used to derive BT, in relation to which, a target fishing mortality is set to ensure that BT is avoided with a high probability (ICES, 2014). However, in other parts of the world, models of the SR relationship is used directly in the estimation of the target fishing mortality, while the biomass limit reference points are set based on a certain percentage of the expected stock biomass when fishing is absent (i.e. virgin biomass or B_0). Lower biomass limits are commonly defined as 0.20 times B_0 and target limits as 0.40 - 0.60 times B_0 (AFMA, 2007; Pikitch et al., 2012). The consistency between this approach and the approaches used for Northeast Atlantic stocks is unknown.

Besides choice of method, numerous other aspects may influence BT estimates. These could be the length of the time-series used to estimate BT. Time-series length varies between assessments and increases over time as more data is being added. Furthermore, the time-series used when assigning reference points are occasionally shortened if regime-shifts have taken place, or if quality issues are raised in relation to historical landings. Finally, stocks that are depleting or rebuilding (i.e. demonstrating declining or increasing spawning biomass over time) may pose additional challenges (i.e. Szuwalski et al., 2019), since the range of observed biomasses expands as time-series length increases, potentially changing the perceived SR relationship over time. Currently, we know of no published systematic study evaluating the confounding effects and biases of BTs arising from these factors. Hence, experts confronted with the task of assigning BTs to stocks are left with little guidance.

To overcome this issue and provide guidance for the estimation and use of BTs, the objective of the present study was to assess BT uncertainty. The sensitivity of BT estimates to (i) choice of method, (ii) time-series length, and (iii) stock development was analyzed and new R code to estimate the bias and precision of the BT estimates was developed. The first part of the study used simulated SR data. However, to demonstrate that conclusions derived from simulations are relevant in a practical setting, we supplemented these with analyses of real SR data sets from 51 small-bodied pelagic fish stocks.

2. Materials and Methods

2.1. Methods for estimating biomass thresholds

The main focus of the study was placed on the methods currently applied to several small-bodied pelagic species in the Northeast Atlantic, which has not formerly been described in the scientific literature. However, for comparison, we also included previously described

methods; including one method, which is broadly applied outside the Northeast Atlantic (Pikitch et al., 2012; AFMA, 2007).

Altogether we considered six different methods: two nonparametric methods based on percentiles (ICES, 2017), one parametric approach based on a simple hockey stick SR relationship (ICES, 2017), two parametric methods based on estimated SR curves (Myers et al., 1994; Beverton & Holt, 1957), and one semi-parametric approach based on a fixed proportion of B_0 (Myers, 1999).

The two nonparametric methods defined BT as the lowest observed spawning stock biomass producing a “large” recruitment (type 1 in the ICES guidelines (ICES, 2017)). Here, we used two definitions of “large”: recruitment > 50th percentile and recruitment > 80th percentile (respectively referred to as P0.5 and P0.8 hereafter).

In the parametric hockey stick method (HS) (type 2 in ICES (2017)), BT was defined as the breakpoint of a segmented regression. The breakpoint was estimated using the grid-search method recommended in Borrowman & Myers (2000). This method is referred to as HS in the following.

The two parametric methods, which involved model fitting of Beverton-Holt (Beverton & Holt, 1957) and Ricker (Ricker, 1954) relationships, relied on the assumption that BT is equal to a certain percentage of the spawning stock biomass that produces maximum recruitment as predicted by the SR model, with the added requirement that the BT cannot be larger than the maximum observed spawning stock biomass (Myers et al., 1994). Myers et al. (1994) used 50% of the maximum predicted recruitment, but initial investigations (in preparation for the present study) showed that 50% resulted in BTs systematically lower than the minimum spawning stock biomass of the simulated datasets for the Ricker model. Therefore, to ensure that all BTs were targeting the same point of the SR curve, rather than using the (arbitrary)

choice of 50% of maximum recruitment, we used the percentage that maximized the probability of identifying the breakpoint from a hockey stick. The resulting percentages were 83% and 51% for the Ricker and Beverton-Holt, respectively. These two methods are hereafter referred to as RK83 and BH51, respectively. Beverton-Holt and Ricker curves were fitted using maximum likelihood estimation via Template Model Builder (TMB) assuming log-normal distributed residuals (Kristensen et al., 2015).

In relation to the semi-parametric approach where BT was equal to a fixed proportion of B_0, we used 20% as the fixed proportion and used the maximum observed spawning biomass (B_{max}) as a simply proxy for B_0. We will hereafter refer to this approach as $0.2B_{max}$.

Summary and illustrations of methods are provided in figure 1.

2.2. Simulation study

In order to simulate SR data with a common underlying threshold, we simulated from a hockey stick relationship (i.e. simulated threshold = breakpoint of hockey stick). The hockey stick in all our simulations had a breakpoint and asymptote at (100,100), corresponding to a simulated threshold of 100. In the default simulations, data sets were simulated from a hockey stick with n=10, 20, 30, or 40 spawning stock biomasses (i.e. mimicking time-series lengths of 10, 20, 30, or 40 years) drawn from a uniform distribution ranging from 40 to 200% of the breakpoint. For each spawning stock biomass, recruitment was simulated from a log-normal distribution around the hockey stick. A recruitment coefficient of variation (CV) of 0.63 was used for the default simulations, corresponding to the median CV in the real stock data described in the subsequent section. In the same way, we also based the before-mentioned spawning stock range on the real stock data (to which we fitted hockey stick models to derive the median range). The default spawning stock biomass range from 40 to

200% of the breakpoint corresponded to a steepness of 0.4. Steepness is defined as the predicted recruitment (in this case based on a hockey stick) at 20% of B_{max} divided by the predicted recruitment at B_{max} (e.g. Mace & Doonan, 1988; Myers et al., 1999; Mangel et al., 2013). A total of one thousand replicated SR data sets were simulated for each sample size (n=10, 20, 30, and 40). It was not considered meaningful to apply the $0.2B_{max}$ approach to the simulated SR data sets because B_{max} is tightly linked to the range of spawning biomasses used in the simulations.

The sensitivity of the results to each simulation setting were carefully investigated. Firstly, the effect of the recruitment CV, was tested by re-running the simulations using the 5 and 95 percentiles of the CVs found in the real stock data (CV = 0.18 and 1.23, respectively). Illustrations of what the simulated data looks like under different assumptions about the recruitment CV can be found in the supplementary materials.

The sensitivity of the results to erroneous assumptions regarding the underlying SR processes was investigated by simulating SR data from different underlying curves (i.e. not just a strict hockey stick assumption as used in the default simulations). In order to do this, we fitted Ricker and Beverton-Holt curves to simulated hockey stick data with low variance (CV = 0.01) and a large sample size (n = 1000) with the same range of spawning stock biomasses as in the default simulations. The rest of the calculations were performed as described above for HS, except this time we used the fitted Beverton-Holt and Ricker curves to simulate recruitment, instead of the hockey stick.

Lastly, we were re-running simulations with spawning stock range from 40 to 300% of the breakpoint and 40 to 400% of the break point, which effectively increase steepness to 0.6 and 0.8, respectively (since B_{max} increase).

Potential effects of stock development (rebuilding or declining) were investigated by sorting each of the 1000 replicate SR data sets (n=40) from highest to lowest spawning stock biomass. We then drew sub-samples of n=20, 25, 30, and 35 from either end of the sorted SR data (i.e. from the high end when imitating declining stocks and from the low end when imitating rebuilding stocks). This approach ensured that depleting stocks initially had no or few data below the breakpoint, whereas rebuilding stocks initially had relatively few data above the breakpoint. BT was calculated using P0.5, P0.8, HS, RK83, and BH51 for each of one thousand replicate sub-sampled data sets of each sample size.

All code used to simulate data and estimate BTs and the associated uncertainty is available at https://github.com/mebrooks/StockRecruit.

2.3. Data-driven study

2.3.1. Stock assessment data

We compiled time-series of spawning stock biomass and recruitment of 51 small-bodied pelagic stocks with more than 10 years of data from the RAM (Ransom A. Myers) Legacy Stock Assessment Database (Version 4.44-assessment-only. Released 2018-12-22. Accessed [Date accessed 2017-18-08]. Retrieved from DOI:10.5281/zenodo.2542919) (Ricard et al., 2012) and the ICES database (extracted using ICES tools: https://github.com/ices-tools-prod/icesSAG, accessed 2018-09-06 using the getSAG function in recruitment (recruitment package: icesSAG)) (ICES, 2018). In order to avoid stock data from the Ram Legacy data base originating from stock assessment models with build-in SR models, only stocks used in Britten et al. (2016) were included. The compilation of stocks had a global coverage and the following families were represented: Clupeidae, Scombridae, Gadidae, Hexagrammidae, Ammodytidae, Engraulidae, Osmeridae, and Carangidae. The average number of years in

each time-series was 35, but seven stocks had less than 20 years of data. The longest time-series was 79 years. If the same stock was recorded in both the RAM and ICES database, the ICES data was selected. A detailed overview of the stocks is provided in the supplementary material.

2.3.2. Comparisons of methods

To allow comparisons of BT estimates across stocks, we expressed BT relative to the geometric mean spawning stock biomass of each stock.

The precision of BTs resulting from the HS method was assessed by bootstrapping the CV of the breakpoint (CV_{HS}) (i.e. estimating the uncertainty of the breakpoint estimate). The bootstrap procedure was organized in three steps: (i) re-sample paired observations with replacement 1000 times from the SR time-series to obtain the same number of observations as in the original dataset, (ii) estimate BT for each of the re-sampled data sets, and (iii) calculate the CV_{HS} across the re-sampled data sets. The R-code for this is available in the function bootBlim() at https://github.com/mebrooks/StockRecruit.

In the data-driven part of the study, we only focused on BT approaches that are currently applied in fish stock assessment and management (i.e. P0.5, P0.8, HS, and $0.2B_{max}$). It was not considered meaningful to include RH83 and BH51, mainly since these were adapted specifically to the simulated data; without applied cases available to guide the choice of realistic percentages (i.e. different from 83% and 51%).

2.3.3. Effects of time-series length and stock development

In order to test the sensitivity of BTs to time-series length and stock development in real data, we performed the following three steps: (i) from each stock a subset comprised of the first 10 years (n=10) of stock-recruitment pairs was selected (i.e. representing a historic perspective when only 10 years of data were available and the first assessment of the stock was carried out). (ii) HS, P0.5, and P0.8 BTs were estimated, and (iii) data subsets were sequentially prolonged by adding one year (n=11, 12, …, 40). These three steps were repeated for each n. BT was scaled to the geometric mean spawning stock biomass of the whole time-series. Subsequently, we calculated the slope of BT as a function of n (hereafter referred to as BT-slope). A positive BT-slope indicates that BT increases with increasing time-series length. To determine if the relationship between time-series length and BT was influenced by stock development, we also calculated the slope of the spawning stock biomass as a function of time (S-slope). A positive S-slope indicates rebuilding stock and negative slope indicates declining stock. All slopes were calculated using linear regression. S-slopes were plotted against BT-slopes to reveal any relationships.

2.3.4. Analysis of case-study stocks

Four stocks were selected as our case-studies. The selected stocks all had at least 38 years of data, only minor temporal trends in spawning stock biomass, and represented different types of SR relationships based on visual inspection: (i) Peruvian anchoveta (*Engraulis ringens*, Engraulidae) from North Central Peru (approaching an ever-increasing SR relationship, recruitment CV = 0.35 and CV_{HS} = 0.52), (ii) Pacific herring from the central coast (*Clupea pallasii*, Clupeidae) (occasional very large recruitment, recruitment CV = 0.98 and CV_{HS} = 0.47), (iii) Chub mackerel from Tsushima strait (*Scomber japonicas*, Scombridae) (hockey stick or Ricker-like SR relationship, recruitment CV = 0.3, CV_{HS} = 0.20), and (iv) Pacific

herring from the prince Rupert district *C. pallasii* (no clear stock-recruitment relationship, recruitment CV = 0.66, CV_{HS} = 0.49).

For each stock we evaluated estimation uncertainty and examined the effect of increasing the time-series length. This was done using the following three steps: (i) n pairs of spawning stock biomass and recruitment were randomly sampled with replacement 100 times, (ii) BT was estimated for each of the one hundred datasets (using either P0.5, P0.8, or HS), and (iii) the above steps were repeated with n=10, 11, 12, …, 38 (i.e. representing different time-series lengths).

3. Results

3.1. Simulation study

Estimation bias (estimated BT relative to the simulated biomass threshold) and precision (width of confidence intervals) were affected by time-series length and choice of method. In general, the nonparametric methods produced more precise BT estimates compared to the parametric methods (Fig. 2), but both P0.5 and P0.8 were negatively biased with the bias being greatest for P0.5 (Fig. 2). Precision increased with increasing time-series length, but due to the bias of P0.5, the confidence intervals of long time-series did not include the simulated threshold. P0.8 produced slightly less precise estimates of BT and, in spite of being consistently negatively biased, retained the simulated threshold within the confidence interval at all time-series lengths (Fig. 2).

Two of the parametric methods (HS and BH51) resulted in BTs that were unbiased regardless of time-series length and recruitment CV. The third parametric method, RK83, was positively biased, but bias decreased as time-series length increased. Confidence intervals around BH51

estimates were very large. Among the parametric methods, HS was the most precise and unbiased method, with lower confidence limits similar to those derived using the nonparametric methods but larger upper confidence intervals.

The precision of the different methods was highly dependent on the variation in recruitment around the SR curve. When the recruitment CV was low, HS was superior to all other methods, having no bias and high precision (Fig. 2 left panel). As CV increased, the nonparametric methods became increasingly biased, but retained approximately the same precision, whereas the precision of the parametric methods decreased substantially. While HS remained unbiased at all levels of recruitment CV, the HS confidence intervals ranged from -50 to +100% at the highest recruitment CV (Fig. 2). BH51 always had lower precision than the remaining parametric methods while RK83 either had lower precision (Fig. 2 left panel) or positive bias (Fig. 2 middle and right panels). Both the nonparametric (P0.05 and P0.8) and the parametric methods (HS, BH51, and RK83) were relatively insensitive to the assumption made about the shape of the underlying SR relationship and steepness, except that precision declined for all methods when steepness increased from 0.4 to 0.8 (see supplementary material).

When simulating a depleting stock, P0.5 and P0.8 decreased rapidly as time-series length increased (Fig. 3), passing from positive to negative bias. HS changed from positive bias to unbiased for longer time-series. For rebuilding stocks, P0.5 and P0.8 showed consistent negative bias of around 50% and 25% respectively, while HS was relatively unbiased. There was no substantial change in the bias of rebuilding stocks with time-series length.

3.2. Data-driven analysis

3.2.1. Comparison of methods

When comparing different methods (P0.5, P0.8, HS, and $0.2B_{max}$) several patterns emerged. BTs from P0.5 were always equal to or lower than those produced by HS (Fig. 4a). For the vast majority of stocks, P0.5, P0.8, and $0.2B_{max}$ produced BTs equal to or less than the geometric mean spawning stock, whereas, HS produced BTs above the geometric mean spawning stock for 42% of the stocks. For stocks where HS produced BTs that were less than half of the geometric mean spawning stock biomass, agreement between HS and P0.5 was relatively high (Fig. 4a). In contrast, there was no agreement between P0.5 and HS for stocks where HS produced BTs higher than the geometric mean spawning stock biomass. When CV_{HS} was above 0.3 (i.e. breakpoint estimate from HS was highly uncertain), BTs from P0.5 were equal to or less than half the geometric mean spawning stock biomass. The agreement between P0.8 and HS was in general higher when CV_{HS} was below 0.3 (Fig. 4b), while the agreement between P0.8 and P0.5 was high for some stocks and low for others (Fig. 4c). $0.2B_{max}$ were generally showing poor agreement with all other methods and the relationship between BTs from the nonparametric methods (P0.5 and P0.8) and $0.2B_{max}$ tended to be negative (i.e. the higher the BT from P0.5 the lower the BT from $0.2B_{max}$) (Fig. 4d-f).

BTs more than twice the geometric mean spawning stock biomass were observed for 0%, 4%, 8%, and 4% of the stocks, when using P0.05, P0.8, HS, and $0.2B_{max}$, respectively. The overall patterns depicted in figure 4 were similar to those seen in the simulated data (see supplementary material). We also examined effects of the recruitment CV (i.e variation in recruitment when above the hockey stick breakpoint), but no clear patterns emerged (see supplementary material). With regards to steepness, the median BT from HS was equal to 33% of B_{max} and 20% of the BTs from HS were > 60% of B_{max} (i.e. low steepness), whereas 31% were below 20% of B_{max} (i.e. high steepness).

3.2.2. Effects of time-series length and stock development

The stocks represented a range of different developmental trajectories, from depleting (negative S-slopes) to rebuilding (positive S-slopes) (Fig. 5). When using P0.5, P0.8, and HS, negative S-slopes were mainly associated with negative BT-slopes, although, there were instances where negative S-slopes were associated with positive BT-slopes when using HS. In contrast, when using $0.2B_{max}$, negative S-slopes resulted mainly in neutral BT-slopes. Positive S-slopes yielded mainly positive BT-slopes for HS, P0.8, and $0.2B_{max}$, whereas, for P0.5, positive S-slopes was mainly associated with neutral BT-slopes (i.e. slopes close to zero).

3.2.3. Case study stocks

P0.5 and P0.8 decreased with increasing time-series length for all four stocks, and concurrently the precision decreased (i.e. confidence intervals narrowed, Fig. 6). The anchovy stock, which appeared to have a continuously increasing SR relationship, showed low precision (i.e. large standard deviations) when applying HS, higher precision for P0.8, and very high precision for P0.5 (Fig. 6). The discrepancy between P0.5 and HS was largest for this stock. For the remaining stocks, the precision of P0.8 and HS was more similar. The precision of P0.5 varied considerably among stocks, being lowest for the two herring stocks (Fig. 6). P0.5 always produced the smallest BTs, whereas, HS produced higher BTs than P0.8 for some stocks, but not for others (Fig. 6). In general, the precision of HS was relatively low and improved only slightly with longer time-series, which was in particular pronounced for the stock where the SR relationship was ever-increasing (anchovy).

4. Discussion

Biomass limit reference points, building on the idea that recruitment is impaired when stock biomass drops below a certain critical biomass threshold (BT), are an essential part of fisheries management and knowledge of the uncertainties associated with a given BT estimate are key to providing robust management advice (Haltuch et al., 2008). This study evaluated several methods for estimating BT for small-bodied pelagic species, primarily focusing on three methods applied in the North Atlantic (HS, P0.5, and P0.8). However, for comparison, a number of alternative methods used in other parts of the world, or described in the literature, were included as well (BH51, RK83, and $0.2B_{max}$).

First of all, the present study demonstrated that estimates of BT are associated with considerable uncertainty. Precision (i.e. measured as deviation from the simulated threshold) and bias (i.e. as measured by width of confidence intervals) depended on the choice of method, year-to-year variation in recruitment, time-series length, and stock development. In general, the simulations and the data-driven studies were consistent regarding the overall conclusions. Furthermore, four case-studies illustrated the degree of inter-stock variability, supporting the notion that stock-specific decisions may always be needed (Katsukawa, 2004).

When variation in recruitment was low, the hockey stick approach (HS) was both unbiased and precise regardless of time-series length. HS generally provided unbiased BT estimates in the simulations, and was relatively insensitive to trends in stock development. For depleting stocks, HS decreased as stock size was depleted and approached the simulated threshold, leading to the elimination of an initial positive bias but not resulting in any negative bias. A similar trend was seen in real data when the time-series length was less than 20 years.

However, precision tended to be relatively low, when variation in recruitment was low, which was also evident from the four case studies.

The nonparametric methods defining “large” recruitment as greater than the 50th percentile (P0.5) or the 80th percentile (P0.8) provided relatively precise estimates of the simulated threshold in both the simulation study and the four case-studies, in particular for P0.5. Considering that P0.5 was by far the most precise of the methods tested, using this methods and adding a fixed percentage as a buffer may seem as an appealing approach. However, these nonparametric methods were also prone to systematic underestimation of the simulated threshold (in the simulation study) and when analyzing real data, P0.5 yielded BTs that were, on average, roughly half of those from HS. Lastly, when recruitment variation was increased in the simulation study, estimated BT decreased further. This pattern was, however, mainly confined to the simulation study and did not show up in the analyses of real stock data (see supplementary materials).

The percentage providing results comparable with HS for Beverton-Holt (51%, BH51) was virtually identical to the 50% suggested by Myers et al. (1994). However, BH51 was also the most imprecise of the methods. Presumably because the maximum recruitment was poorly estimated by the Beverton-Holt curve. In relation to the Ricker approach, the percentage of maximum recruitment providing results comparable with HS was 83% (RK83), demonstrating that using 50%, as suggested by Myers et al. (1994), would have resulted in a systematic underestimation of the simulated threshold. Hence, the method relies heavily on the ability to identify an appropriate percentage. Another limitation of this methods is the low precision in cases where recruitment variation is high (similar to HS and BH51) (see also Szuwalski et al., 2019). Lastly, BH51 and RK83 is based on relatively strict assumptions (that cannot be easily tested) about the underlying functional SR relationship.

Besides the fact that $0.2B_{max}$ was based on an arbitrary choice (in this case 20% of B_{max}) it is also hinged on the assumption that the "virgin" (or unfished) stock biomass is known. Furthermore, the present study revealed a very low agreement with other methods analyzed in the present study, which is worrying (i.e. as seen in Fig. 4). Small pelagic fishes generally exhibit increases in recruitment over a greater range of spawning biomass than gadids and pleuronectid and hence attain a lower proportion of the maximum recruitment on average at 20% of B_{max} (i.e. lower steepness (Mace and Doonan, 1988; Myers et al., 1999)). This may explain some of the systematic differences between this methods and those considering both spawning stock biomass and recruitment.

One approach to selecting appropriate biomass limit reference points, could be to combine several methods and consider agreements/disagreements among methods. In general there was surprisingly little agreement between methods (i.e. as seen in Fig. 4). There was, however, relatively high consistency between P0.5 and HS when the HS-estimated BTs were less than the geometric mean spawning stock biomass. However, for stocks where the HS-estimated BTs were greater than the geometric mean spawning stock biomass, the BTs from the different methods were unrelated. P0.8 and HS produced similar ranges of BTs when averaging across stocks, but at the level of the individual stock, P0.8 and HS yielded very different results. This was also evident from the case studies, where P0.8 provided the highest value for the Prince Rupert district herring stock, whereas, for the anchovy stock, the highest value was derived from HS.

To gain acceptance and credibility of agreed biomass limit reference points, not only among scientists, but also among managers and end-users (Caddy & Seijo, 2005; Rice, 2005), robust and transparent approaches to estimate, evaluate, and explain the associated uncertainty of are needed. Our assessment and comparison of BT methods serve to formally illustrate several previously unstudied and unquantified sources of biases associated with the

estimation of BTs. In addition, it shows that our ability to identify a simulated threshold is rather limited based on the current set of methods applied by fisheries scientists in the northeast Atlantic (ICES, 2017). The primary limitation and source of uncertainty arises from one of the most well studied, yet still insufficiently understood process in fisheries science, namely our ability to reliably represent and characterize SR relationships (Szuwalski et al., 2015). The approaches tested within this study perform well (i.e. in terms of precision and accuracy) when recruitment variation was low. However, this is seldom the case for marine fish stocks that typically demonstrate highly variable and potentially non-stationary SR relationships (Lindegren & Eero 2013; Beggs et al., 2014). Such high variability is particularly prevalent in short-lived, fast growing and early maturing species of small pelagic fish (e.g. MacKenzie & Köster, 2004; Lindegren & Checkley, 2013; Szuwalski et al., 2019). To overcome these challenges, a better understanding of the sources of variation in recruitment is needed to appropriately account for these underlying mechanisms in SR models and ultimately in relation to biomass limit reference points.

One of the most frequently invoked sources of variability in stock abundances and/or recruitment, besides density-dependence, is the underlying variability in climate or system productivity (i.e. carrying capacity) that ultimately determines investment in reproduction, fecundity, early-life survival and consequently recruitment success (e.g. Brunel & Boucher 2006; MacKenzie et al., 2012; Sparrevohn et al., 2013; Bartolino et al., 2014; Hobday et al., 2016; Tommasi et al., 2017; Payne et al., 2019). Although meta-analyses of environmental correlates and predictors of recruitment show a common tendency for published recruitment-environment correlations to fail when verified with new data (Myers, 1998), more recent examples show that environmental predictors of recruitment of small pelagic fish may hold upon retesting (Lindegren & Checkley, 2013; Lindegren et al., 2017). Hence, identifying and introducing ecologically underpinned predictors of recruitment variability into methods for

assigning reference points could be one way to proceed. In addition to the potential consideration and inclusion of environmental predictors, new techniques to assign reference points are needed. Recent developments of nonparametric models allowing for estimations of reference points, while accounting for model uncertainty seem promising in this regard (e.g. Cadigan, 2012) and should be formally tested and considered as part of the toolbox needed for fisheries scientist and managers to estimate and decide on robust BT within fisheries management.

The estimation of SR relationships and the spawning stock biomass at which recruitment is impaired is a key uncertainty in estimation of not only biomass limit reference points, but also F_{MSY} (Cadigan, 2012). The two most influential factors are the natural mortality the stock experiences and the steepness of the stock recruitment relationship (i.e. the proportion of the maximum recruitment attained at $0.2B_{max}$ (Mace, 1994; Forrest et al., 2010; Zhou et al., 2012)). Steepness is easily estimated from HS and hence the code in the supplementary can be used to provide a rapid check of the validity of any assumptions of steepness. In the stocks investigated here, the median HS of $0.33B_{max}$ corresponded to a median steepness of 0.55, which is low compared to previous analyses using Ricker models for clupeids and engraulids and scombrids (0.71 and 0.62) but in line with values for scombrids (0.52) (Myers et al., 1994).

Choice of method was the most influential factor in estimating BT. Without clear guidelines for when to choose each of the methods, this introduces inconsistencies, hidden subjectivity and lack of transparency in stock management. Secondly, time-series length was found to be important. Since time-series, displaying clear signs of a regime-shift, are occasionally shortened in real life assessment settings when updating reference points (i.e. including only the most recent part of the time-series), knowing the sources of uncertainty introduced by doing so is essential. Finally, stocks that are depleting or rebuilding (i.e. demonstrating

declining or increasing spawning biomass across years) may pose additional challenges, since the range of observed biomasses expands as time-series length increases, potentially changing the perceived SR relationship over time.

When recruitment variation is low, the HS method seem promising. However, as also discussed above, variation in recruitment can be substantial when it comes to small-bodied pelagic fish stocks. Hence, it may be useful to consider other methods in addition to HS. However, further studies will be needed before we can formulate quantitative guidelines as to which combination of methods to use and how and when to use them. Furthermore, as the methods differed in their probability of over- or under-estimating the simulated threshold, the decision on which method is preferable also depends on the risk-averseness of the decision makers and the approach taken by scientists to incorporate the increase in *reported* uncertainty (Rice, 2005). Lastly, we recommend that in addition to clearly stating the methods used, the amount of data used, trends in the stock development, and the uncertainty of the estimated BT, should be clearly reported as well.

Acknowledgement:

Rodrigo Wiff, David C.M. Miller, and one anonymous reviewer provided constructive critiques on drafts of this manuscript. The study was partly supported by the European Maritime and Fisheries Foundation and the Ministry of Environment and Food of Denmark (grant ID: 33113-B-19-150). AR & MB also received funding from the PROBYFISH project (EASME/EMFF/2017/022). Thanks to A. Kokkalis for helpful discussion.

Data availability:

Important R-code are available at https://github.com/mebrooks/StockRecruit.

Stock-recruitment data were downloaded from the open access Ram Legacy Stock Assessment Database (Version 4.44-assessment-only. Released 2018-12-22) (Ricard et al., 2012) and the ICES database (extracted using ICES tools: https://github.com/ices-tools-prod/icesSAG) (ICES, 2018).

References

AFMA, Australian Fisheries Management Authority. (2007). Commonwealth Fisheries Harvest Strategy: Policy and Guidelines. Department of Agriculture, Fisheries, and Forestry.

Barrowman, N. J., & Myers, recruitment. A. (2000). Still more spawner-recruitment curves: the hockey stick and its generalizations. Canadian Journal of Fisheries and Aquatic Sciences, 57(4), 666-676. https://doi.org/10.1139/f99-282

Bartolino, V., Margonski, P., Lindegren, M., Linderholm, H., Cardinale, M., Rayner, D., Wennhage, H., & Casini, M. (2014). Forecasting fish stock dynamics under climate change: Baltic herring as a case study. Fisheries Oceanography, 23, 258–269. https://doi.org/10.1111/fog.12060

Beggs, S. E., Cardinale, M., Gowen, R. J., & Bartolino, V. (2014). Linking cod (Gadus morhua) and climate: investigating variability in Irish Sea cod recruitment. Fish. Oceanogr, 23, 54–64. https://doi.org/10.1111/fog.12043

Beverton, R. J. H. & Holt, S.J. (1957). On the Dynamics of Exploited Fish Populations (Fisheries Investigations, Ser. 2, Vol. 19). UK Ministry of Agriculture and Fisheries, London.

Britten, G. L., Dowd, M., & Worm, B. (2016). Changing recruitment capacity in global fish stocks. Proceedings of the National Academy of Sciences, 113, 134-139. https://doi.org/10.1073/pnas.1504709112

Brunel, T., & Boucher, J. (2006). Pattern of recruitment variability in the geographical range of the exploited northeast Atlantic fish species. Journal of Sea Research, 55, 156-168. https://doi.org/10.1016/j.seares.2005.07.003

Caddy, J. F., & Seijo, J. C. (2005). This is more difficult than we thought! The responsibility of scientists, managers and stakeholders to mitigate the unsustainability of marine fisheries. Philosophical Transactions of the Royal Society B: Biological Sciences, 360, 59-75. https://doi.org/10.1098/rstb.2004.1567

Cadigan, N. G. (2012). Fitting a non-parametric stock–recruitment model in R that is useful for deriving MSY reference points and accounting for model uncertainty. ICES Journal of Marine Science, 70, 56-67. https://doi.org/10.1093/icesjms/fss183

Clausen, L. W., Rindorf, A., Van Deurs, M., Dickey-Collas, M., & Hintzen, N. T. (2018). Shifts in North Sea forage fish productivity and potential fisheries yield. Journal of Applied Ecology, 55, 1092-1101. https://doi.org/10.1111/1365-2664.13038

Deroba, J. J., & Bence, J. R. (2008). A review of harvest policies: understanding relative performance of control rules. Fisheries Research, 94, 210-223. https://doi.org/10.1016/j.fishres.2008.01.003

Foley, M. M., Martone, recruitment. G., Fox, M. D., Kappel, C. V., Mease, L. A., Erickson, A. L., ... Scarborough, C. (2015). Using ecological thresholds to inform resource

management: current options and future possibilities. Frontiers in Marine Science, 2, 95. https://doi.org/10.3389/fmars.2015.00095

Forrest, R. E., McAllister, M. K., Dorn, M. W., Martell, S. J., & Stanley, R. D. (2010). Hierarchical Bayesian estimation of recruitment parameters and reference points for Pacific rockfishes (Sebastes spp.) under alternative assumptions about the stock–recruit function. Canadian Journal of Fisheries and Aquatic Sciences, 67, 1611-1634. https://doi.org/10.1139/F10-077

Gjøsæter, H., Bogstad, B., Tjelmeland, S., & Subbey, S. (2015). A retrospective evaluation of the Barents Sea capelin management advice. Marine Biology Research, 11, 135-143. https://doi.org/10.1080/17451000.2014.928414

Haltuch, M. A., Punt, A. E., & Dorn, M. W. (2008). Evaluating alternative estimators of fishery management reference points. Fisheries Research, 94, 290-303. https://doi.org/10.1016/j.fishres.2008.01.008

Hobday, A. J., Spillman, C. M., Paige Eveson, J., & Hartog, J. R. (2016). Seasonal forecasting for decision support in marine fisheries and aquaculture. Fisheries Oceanography, 25, 45–56. https://doi.org/10.1111/fog.12083

ICES (2014). Report of the Joint ICES-MYFISH Workshop to consider the basis for FMSY ranges for all stocks (WKMSYREF3). ICES CM 2014/ACOM:64.

ICES (2017). ICES fisheries management reference points for category 1 and 2 stocks. ICES Advice Technical Guidelines, ICES Advice 2017, Book 12. DOI: 10.17895/ices.pub.3036.

ICES (2018). ICES Stock Assessment Graphs database, 2018/September. ICES, Copenhagen.

Katsukawa, T. (2004). Numerical investigation of the optimal control rule for decision-making in fisheries management. Fisheries science, 70, 123-131.

Lindegren, M., & Eero, M. (2013). Threshold-dependent climate effects limit recruitment and recovery of the Kattegat cod. Marine Ecology-Progress Series, 490, 223-232. https://doi.org/10.3354/meps10437

Lindegren, M., & D. M. Checkley, Jr. (2013). Temperature dependence of Pacific sardine (Sardinops sagax) recruitment in the California Current revisited and revised. Canadian Journal of Fisheries and Aquatic Sciences, 52, 566-577. https://doi.org/10.1139/cjfas-2012-0211

Lindegren, M., Van Deurs, M., Mackenzie, B. R., Clausen, W. L., Christensen, A., & Rindorf, A. (2017). Productivity and recovery of forage fish under climate change and fishing: North Sea sandeel as a case study. Fisheries Oceanography, 27, 212-221.https://doi.org/10.1111/fog.12246

Mace, P. M., & Doonan, I. J. (1988). A generalised bioeconomic simulation model for fish population dynamics. MAFFish, NZ Ministry of Agriculture and Fisheries.

Mace, P. M. (1994). Relationships between common biological reference points used as thresholds and targets of fisheries management strategies. Canadian Journal of Fisheries and Aquatic Sciences, 51, 110-122. https://doi.org/10.1139/f94-013

MacKenzie B. R., & Köster F. W. (2004). Fish production and climate: Sprat in the Baltic Sea. Ecology, 85, 784–794. https://doi.org/10.1890/02-0780

MacKenzie, B. R., Meier, H. E. M., Lindegren, M., Neuenfeldt, S., Eero, M., Blenckner, T., Tomczak, M.T., & Niiranen, S. (2012). Impact of climate change on fish population dynamics in the Baltic Sea: A dynamical downscaling investigation. Ambio, 41, 626-636.

Mangel, M., MacCall, A. D., Brodziak, J. K., Dick, E. J., Forrest, R. E., Pourzand, R., & Ralston, S. (2013). A Perspective on Steepness, Reference Points, and Stock Assessment. Can. J. Fish. Aquat. Sci. 70, 930-940. https://doi.org/10.1139/cjfas-2012-0372

Myers, R. A., Rosenberg, A. A., Mace, P. M., Barrowman, N., & Restrepo, V. R. (1994). In search of thresholds for recruitment overfishing. ICES Journal of Marine Science, 51, 191-205. https://doi.org/10.1006/jmsc.1994.1020

Myers, R. A. (1998). When do environment–recruitment correlations work? Reviews in Fish Biology and Fisheries, 8, 285–305.

Myers, R. A., Bowen, K. G., & Barrowman, N. J. (1999). Maximum reproductive rate of fish at low population sizes. Canadian Journal of Fisheries and Aquatic Sciences, 56, 2404-2419. https://doi.org/10.1139/f99-201

Nakatsuka, S., Ishida, Y., Fukuda, H., & Akita, T. (2017). A limit reference point to prevent recruitment overfishing of Pacific bluefin tuna. Marine Policy, 78, 107-113. https://doi.org/10.1016/j.marpol.2017.01.017

Payne, M. R., Hobday, A. J., MacKenzie, B. R., & Tommasi, D. (2019). Seasonal-to-decadal prediction of marine ecosystems: Opportunities, approaches, and applications. Frontiers in Marine Science, 6, 100. https://doi.org/10.3389/fmars.2019.00100

Pikitch, E., Boersma, P.D., Boyd, I.L., Conover, D.O., Cury, P., Essington, T., Heppell, S.S., Houde, E.D., Mangel, M., Pauly, D., Plagányi, É., Sainsbury, K., & Steneck, R.S. (2012). Little fish, big impact: Managing a crucial link in ocean food webs. Lenfest Ocean Program. Washington, DC. 108 pp. https://www.lenfestocean.org/-/media/assets/extranets/lenfest/len_little_fish_big_impact.pdf

Pikitch, E. K. (2015). Stop-loss order for forage fish fisheries. Proceedings of the National Academy of Sciences, 112, 6529-6530. https://doi.org/10.1073/pnas.1505403112

Ricard, D., Minto, C., Jensen, O. P., & Baum, J. K. (2012). Evaluating the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. Fish and Fisheries, 13, 380-398. https://doi.org/10.1111/j.1467-2979.2011.00435.x

Rice, J. C. (2005). Implementation of the ecosystem approach to fisheries management—asynchronous co-evolution at the interface between science and policy. Marine Ecology Progress Series, 300, 265-270.

Ricker, W. E. (1954). Stock and recruitment. Journal of the Fisheries Board of Canada, 11, 559-623.

Rindorf, A., Cardinale, M., Shephard, S., De Oliveira, J. A., Hjorleifsson, E., Kempf, A., ... & Simmonds, J. (2017). Fishing for MSY: using “pretty good yield” ranges without impairing recruitment. ICES Journal of Marine Science, 74, 525-534. https://doi.org/10.1093/icesjms/fsw111

Sissenwine, M. P., & Shepherd, J. G. (1987). An alternative perspective on recruitment overfishing and biological reference points. Canadian Journal of Fisheries and Aquatic Sciences, 44, 913-918. https://doi.org/10.1139/f87-110

Sparrevohn C., Lindegren M., & MacKenzie B. R. (2013). Climatic induced response of commercially important flatfish species during the 20th century. Fisheries Oceanography, 22, 400-408. https://doi.org/10.1111/fog.12030

Subbey, S., Devine, J. A., Schaarschmidt, U., & Nash, R. D. (2014). Modelling and forecasting stock–recruitment: current and future perspectives. ICES Journal of Marine Science, 71, 2307-2322. https://doi.org/10.1093/icesjms/fsu148

Szuwalski, C. S., Vert-Pre, K. A., Punt, A. E., Branch, T. A., & Hilborn, R. (2015). Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish and Fisheries, 16, 633-648. https://doi.org/10.1111/faf.12083

Szuwalski, C. S., Britten, G. L., Licandeo, R., Amoroso, R. O., Hilborn, R., & Walters, C. (2019). Global forage fish recruitment dynamics: A comparison of methods, time-variation, and reverse causality. Fisheries Research, 214, 56-64. https://doi.org/10.1111/faf.12083

Tommasi, D., Stock, C. A., Hobday, A. J., Methot, R., Kaplan, I. C., Eveson, J. P., ... & Pershing, A. (2017). Managing living marine resources in a dynamic environment: the role of seasonal to decadal climate forecasts. Progress in Oceanography, 152, 15-49. https://doi.org/10.1016/j.pocean.2016.12.011 Zhou, S. (2007). Discriminating alternative stock–recruitment models and evaluating uncertainty in model structure. Fisheries Research, 86, 268–279. https://doi.org/10.1016/j.fishres.2007.06.026

Zhou, S., Yin, S., Thorson, J. T., Smith, A. D., & Fuller, M. (2012). Linking fishing mortality reference points to life history traits: an empirical study. Canadian Journal of Fisheries and Aquatic Sciences, 69, 1292-1301. https://doi.org/10.1139/f2012-060

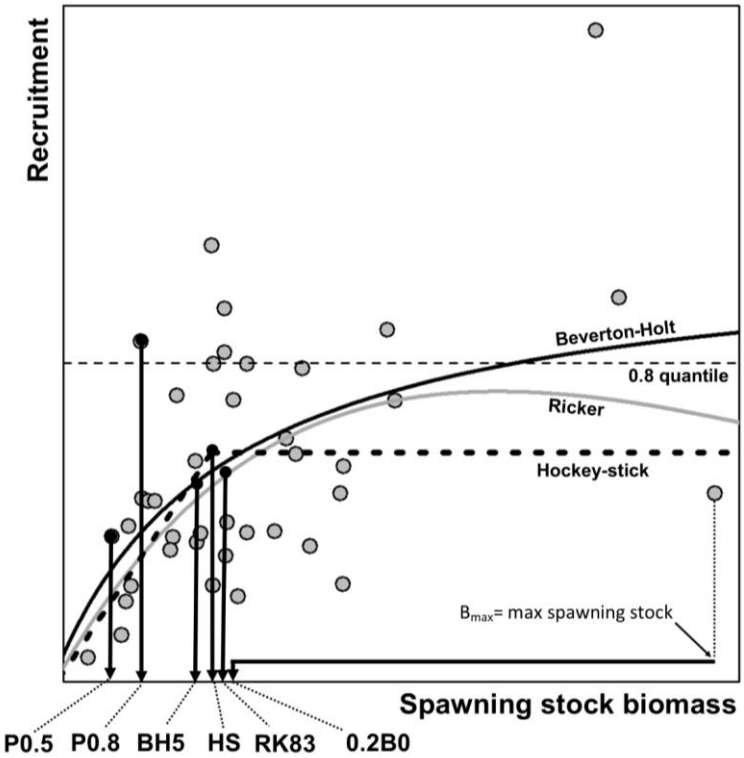

Figure 1. Illustration of different methods (P0.5, P0.8, HS, RK83, BH51, $0.2B_{max}$) for estimating biomass thresholds (BTs). Hockey stick (black dotted), Ricker (grey), and Beverton-Holt (black) curves fitted to SR data from a hypothetical stock (grey dots). The dashed horizontal line represents the 0.8 quantile of recruitment and the maximum spawning stock biomass (B_{max}) used in $0.2B_{max}$ approach is indicated by a bended arrow. The vertical arrows point to the spawning stock biomasses representing the BT derived from each of the methods.

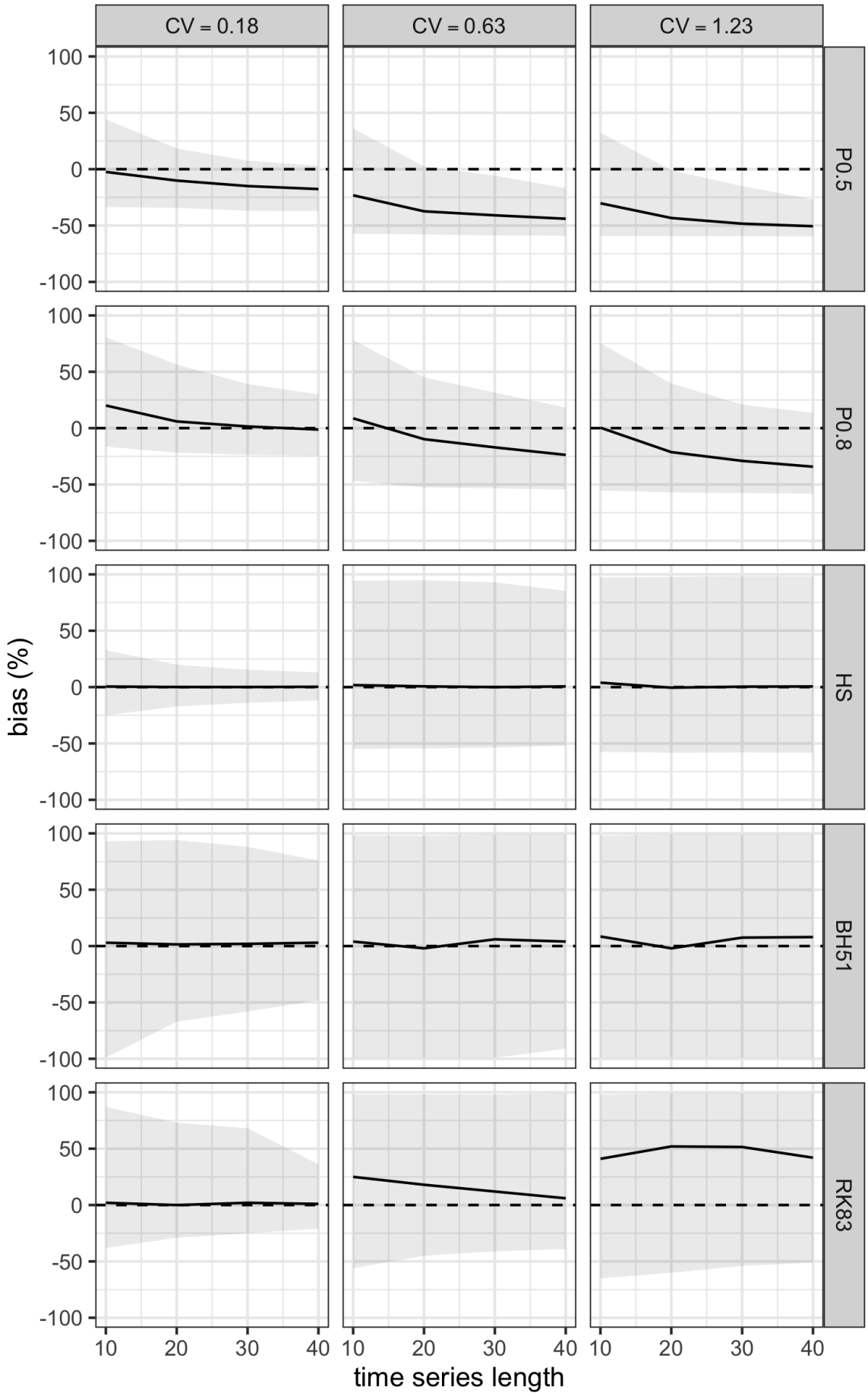
CV = 0.18
CV = 0.63
CV = 1.23
P0.5
P0.8
HS
BH51
RK83
bias (%)
100
50
0
-50
-100
10
20
30
40
time series length

Figure 2. Biomass thresholds (BTs) in simulations. Each column of panels used a different coefficient of variation (CV) in recruitment in the simulated stock-recruitment data. Horizontal dashed lines represent 0% bias from the simulated threshold (i.e. the breakpoint of the underlying hockey stick from which data were simulated). Solid lines are the medians of 1000 replicates and the grey areas encompass the 0.025 and 0.975 quantiles. The x-axis represents time-series lengths used to estimate each BT. Five different methods were used to calculate BT (from top and down: P0.5, P0.8, HS, BH51, and RK83).

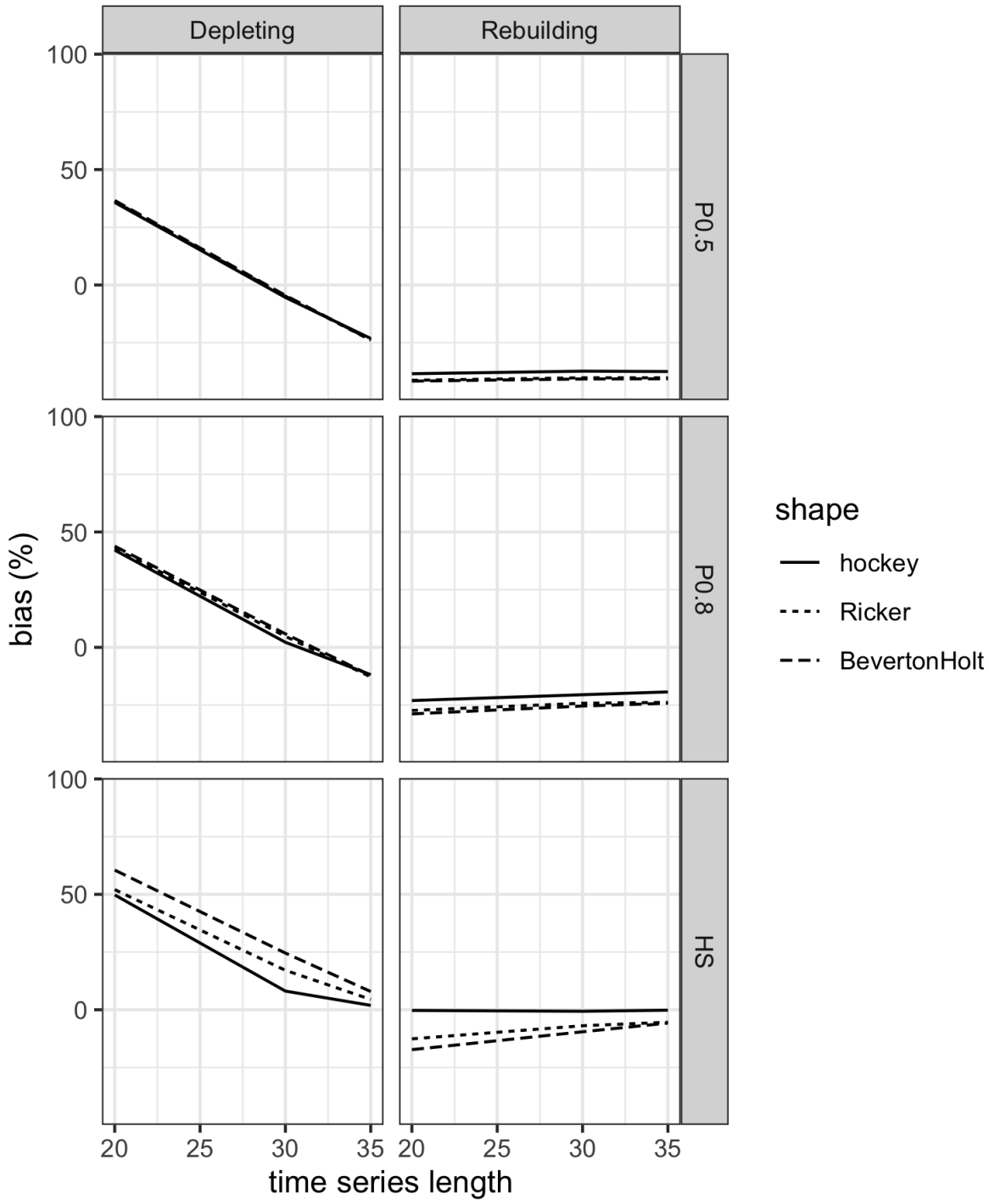

Figure 3. The effect of stock development on biomass thresholds (BTs). Median bias (relative to the simulated threshold) in each BT is shown for combinations of: (i) three different methods for estimating BT (rows of panels), (ii) different underlying stock-recruitment relationships (hockey stick (solid line), Ricker (dotted line), and Beverton-Holt (dashed

line)), (iii) different time-series lengths (x-axis), and (iv) depleting and rebuilding stocks (columns of panels).

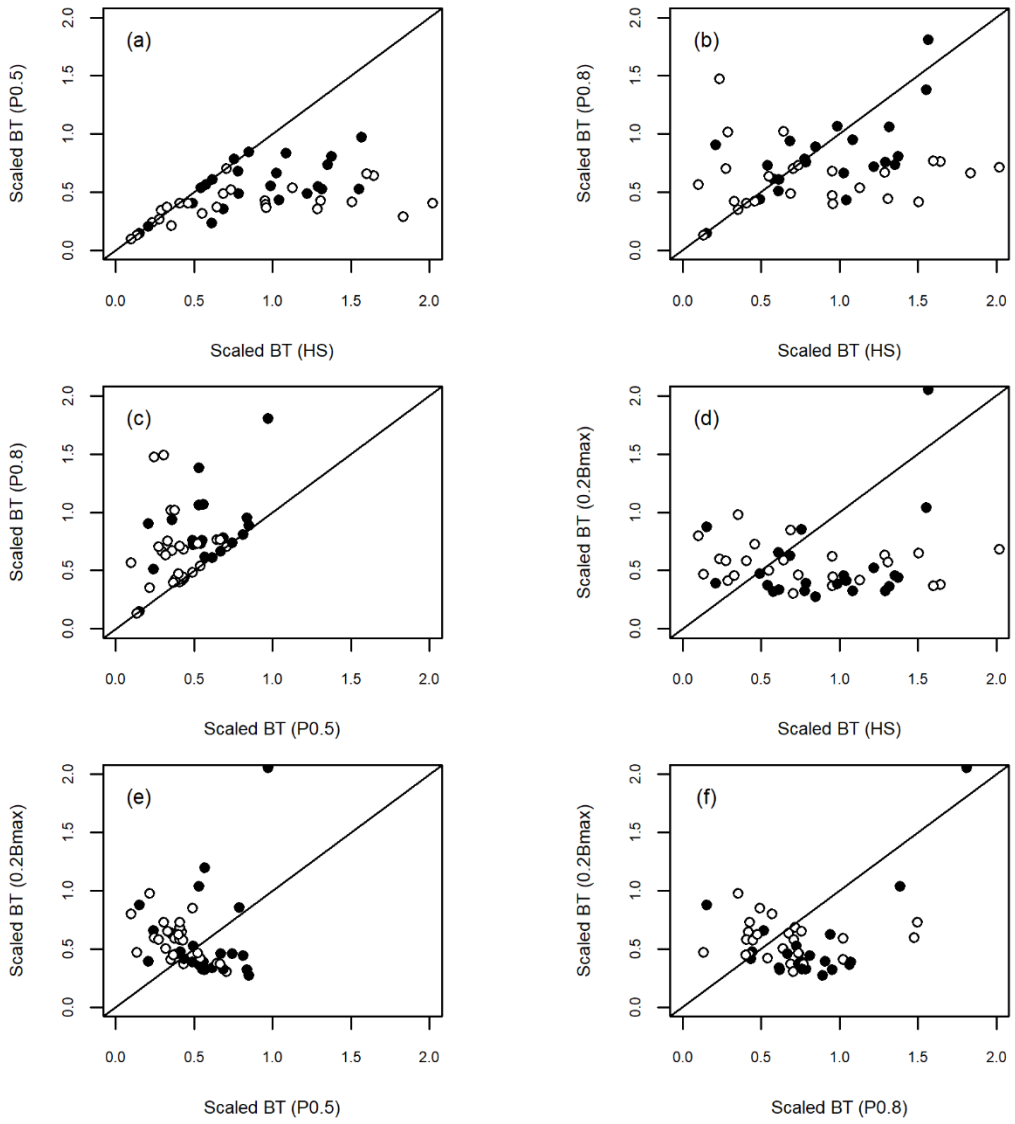

Figure 4. Comparison of biomass thresholds (BTs) of 51 small-bodied pelagic stocks (scaled to the geometric mean spawning stock biomass) and calculated using three different methods (P0.5, P0.8, HS, $0.2B_{max}$, respectively). BTs calculated using HS were divided into stocks

with relatively well defined hockey stick breakpoints (black dots, $CV_{HS} < 0.3$) and stocks with poorly defined breakpoint estimates (white dots, $CV_{HS} > 0.3$). Axes were cutoff at 2.0. Hence, outlier values twice the geometric mean spawning stock biomass were not included in the plots (mounting to 4, 5, 0, 5, 1, 3 data points in a, b, c, d, e, f, respectively).

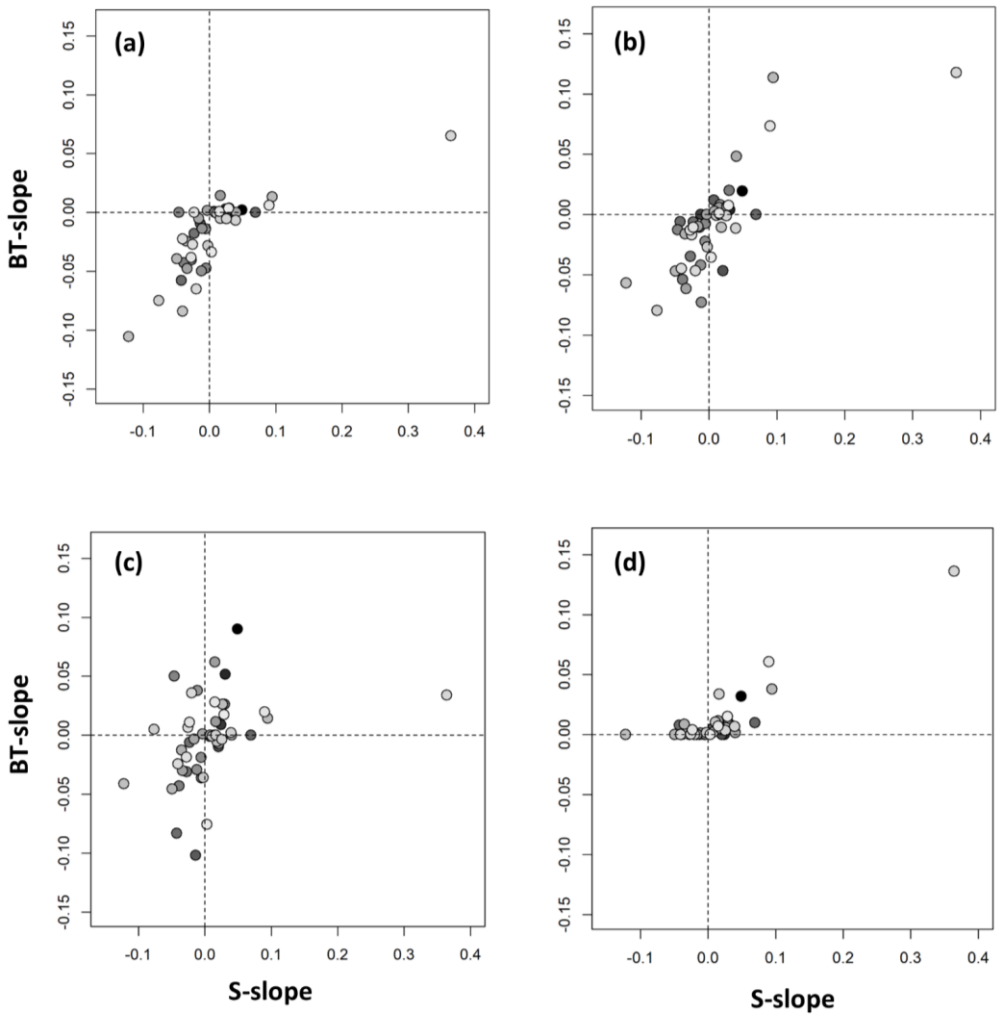

Figure 5. The effect of stock development. Two types of slopes were calculated for each of the 51 stocks: (i) the slope describing the relationship between time-series length and biomass threshold (BT-slope) and (ii) the slope of spawning stock biomass as a function time (S-

slope). Rebuilding stocks have positive S-slopes and depleting stocks have negative S-slopes. BT-slopes were derived using the following four methods: P0.5 (a), P0.8 (b), HS (c), $0.2B_{max}$ (d). The grey scale represents the uncertainty of the S-slope (dark grey is high uncertainty).

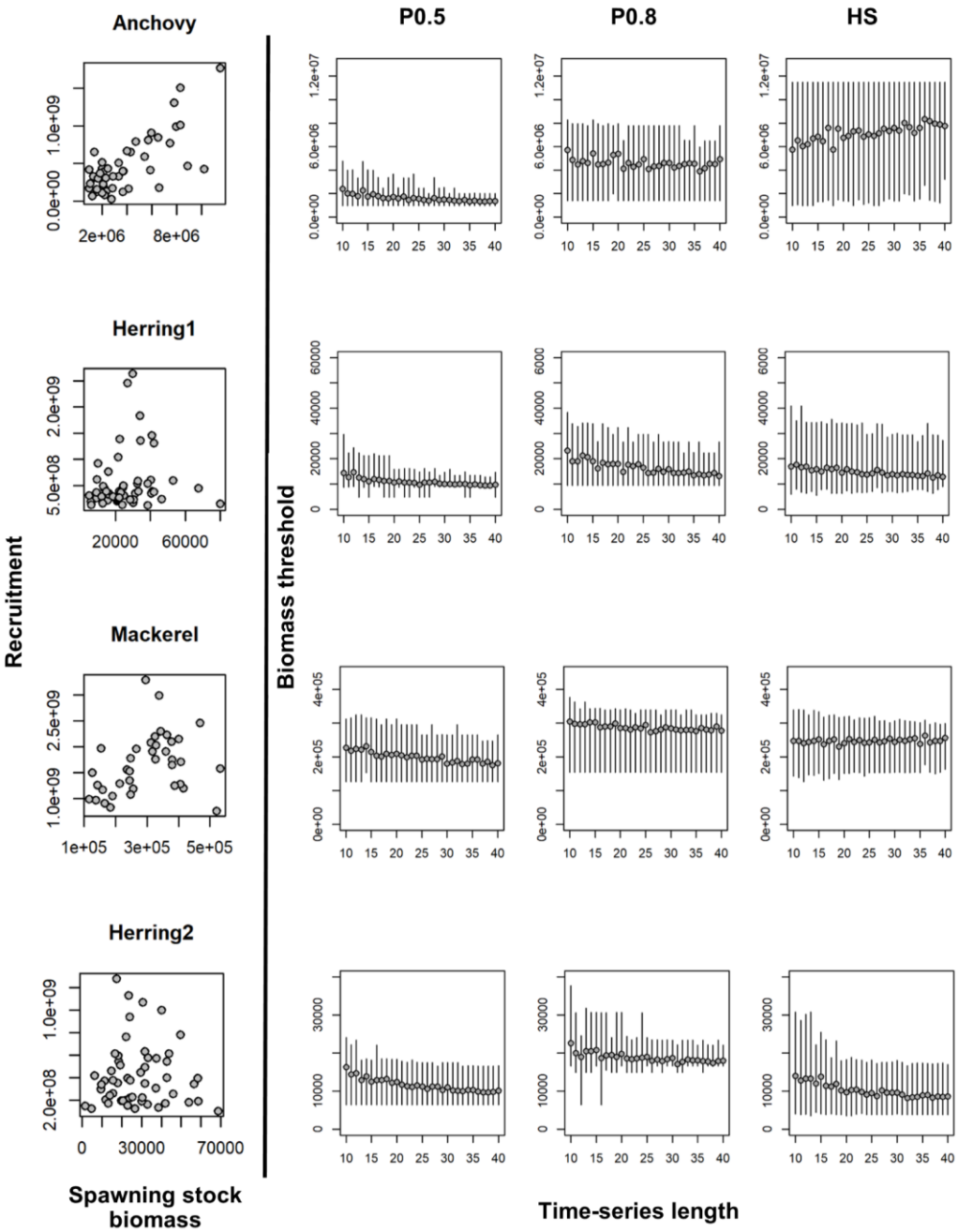

Figure 6. Four case-study stocks, representing different shapes of stock-recruitment relationships: Peruvian anchoveta from North Central Peru (top), Pacific herring from the US west coast (upper middle), chub mackerel from the Tsushima strait (lower middle), and Pacific herring from the Prince Rupert district. To the left of the vertical black line, plots of the SR relationship are shown. To the right of the vertical black line, mean biomass

thresholds (BTs), incl. 10% and 90% quantiles, are shown for three different methods (P0.5, P0.8, and HS) and as a function of time-series length. Quantiles were generated by randomly sampling with replacement n (n = 10, 11, 12, …, 40) pairs of spawning stock biomass and recruitment with replacement 100 times.