WORKSHOP TO COMPARE THE INDICATORS FOR CFP AND MSFD D3 MANAGEMENT OBJECTIVES THROUGH SIMULATIONS (WKSIMULD3)

VOLUME 6 | ISSUE 4

ICES SCIENTIFIC REPORTS

RAPPORTS
SCIENTIFIQUES DU CIEM

ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA
CIEM CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H.C. Andersens Boulevard 44-46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

ISSN number: 2618-1371

ICES Scientific Reports

Volume 6 | Issue 4

WORKSHOP TO COMPARE THE INDICATORS FOR CFP AND MSFD D3 MANAGEMENT OBJECTIVES THROUGH SIMULATIONS (WKSIMULD3)

Recommended format for purpose of citation:

ICES. 2024. Workshop to compare the indicators for CFP and MSFD D3 management objectives through simulations (WKSIMULD3).
ICES Scientific Reports. 6:4. 165 pp. https://doi.org/10.17895/ices.pub.25266475

Editors

Anna Rindorf • Giuseppe Scarcella

Authors

Enrico Armelloni • Valerio Bartolino • Jose Maria Bellido • Elisabeth Bolund • Victoria Campon-Linares • Gema Canal • Ilaria Coscia • Elizabeth Duskey • Tomaso Fortibuoni • Eva Garnacho • Patracia Gonçalves • Chris Griffiths • Daniel Howell • Nis Jacobsen • Susana Junquera • Peter Kidd • Bernhard Kuehn • Stefano Moro • Natasha Murphy • Nikolaus Probst • Anna Rindorf • Owen Rowe • Giuseppe Scarcella • Sonia Seixas • Jessica Tengvall • Amina Tifoura • Ching Maria Villanueva • Håkan Wennhage • Josephine Woronoff • Nuria Zaragoza • Tomas Zolubas • Barbara Zorica

ICES
CIEM
International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

Contents

i Executive summary

The WKSIMULD3 meeting provided a platform for experts from the EU member states to meet and progress the assessment methodology on Criteria 3 of Descriptor 3 upon request by EC (DGENV). WKSIMULD3 is the third of a series of three workshops (WKD3C3SCOPE and WKD3C3THRESHOLDS being the first two) to identify operational indicators for MSFD D3C3.

The workshop was organised as a series of presentations with intermittent discussions. The group agreed on a number of stocks to run simulations to explore the relationships between indicators of population traits/dynamics and healthy population structure. The participants found that for demersal stocks in the North Sea, MSE simulations showed decreasing values of D3C3 indicators with increasing fishing mortality. However, the use of age-structure indicators for these stocks is likely to lead to frequent false positives due to the high interannual variability.

Age-based simulations were implemented also for pelagic and short-lived stocks, such as mackerel, sprat, and sandeel. The distribution of the threshold values implemented for the stocks did not show a clear response to the fishing regime. As a result, the use of age-structure indicators for these stocks is likely to lead to frequent false positives/negatives. The simulations aimed to validate the population size structure and calculate indicators under different fishing scenarios.

In the Mediterranean Sea, length-based simulations were conducted for two hake stocks and one sole stock. The length indicators obtained were compared to historical data to assess the responsiveness of the indicators to exploitation. These results show the behaviour of the length based and of the age-based indicators calculated on observed data. Overall, the indicator status can be well below the average value at Fmsy. Additionally, an exploratory run tested a harvest control rule driven by age-based indicator targets. The outputs can provide directions on how to choose a threshold that can ensure the stock to be in GES.

Overall, the workshop findings highlighted the complexity of evaluating indicators for CFP and MSFD D3 management objectives, particularly in relation to the responsiveness of the indicators to fishing pressure and environmental variation. The group emphasized the need for robust and peer-reviewed models to ensure reliable results when evaluating indicators for fish stocks. Additionally, it underscores the challenges associated with using age- and length-based indicators for different species and the importance of considering environmental and recruitment variability in simulations.

In conclusion, the series of the 3 workshops contributed valuable insights into the complexities of assessing and comparing indicators for MSFD D3 management objectives, providing important considerations for future assessments and management strategies.

ii Expert group information

Expert group name	WKSIMULD3 - workshop to compare indicators for CFP and MSFD D3 management objectives through simulations
Expert group cycle	Annual
Year cycle started	2023
Reporting year in cycle	1/1
Chair(s)	Anna Rindorf, Denmark,
	Giuseppe Scarcella, Italy
Meeting venue(s) and dates	13-17 November 2023, ICES/online, 33 participants

1 Terms of Reference

The workshop to compare indicators for CFP and MSFD D3 management objectives through simulations (WKSIMULD3), chaired by Anna Rindorf, Denmark and Giuseppe Scarcella, Italy, will meet in ICES headquarters, Denmark, from 13-17 November to:

a) Building on from preparatory work, discuss and agree on the stocks to run simulations

b) Run simulations to explore the relationships between indicators of population traits/dynamics and healthy population structure and infer cases where management under CFP objectives alone may be insufficient.

c) Compare results with recent MSE simulations

d) For the stocks identified at c), rank potential management options in decreasing order of expected effectiveness.

2 Workshop Approach

2.1 Text Workshop background

The Marine Strategy Framework Directive aims to protect the marine environment across Europe and to achieve Good Environmental Status (GES) by 2020. Assessments of the state of fish and shellfish population are required under both the CFP and the MSFD Descriptor 3: Commercial fish and shellfish (D3).

The assessment of stock status under the CFP uses the well-established indicators fishing mortality rate (F) and spawning stock biomass (SSB). These have also been adopted for use under the MSFD (criteria D3C1 and D3C2 of Commission Decision (EU) 2017/848) to ensure that a single stock assessment can serve the purposes of both the CFP and the MSFD. Under the MSFD, a third criterion (D3C3) is included in order to achieve good environmental status, the age and size distribution of individuals in a population, D3C3, defined as: *'Populations are within safe biological limits, exhibiting a population age and size distribution that is indicative of a healthy stock'*).

ICES has previously advised on possible approaches to assessing D3C3 and made proposals for suitable indicators (ICES 2016, ICES 2017) but common indicator and threshold values have yet to be agreed upon. This previous ICES advice investigated the indicators number or biomass of old fish, proportion of large or old spawners and the 95th percentile of length of individuals in the population. The indicator mean length in the catch was considered to reflect fisheries selectivity and therefore not directly related to D3C3 while the indicator proportion of mature fish in the stock was considered to be highly impacted by recruitment.

Since then, additional suggestions include indicators for CFP management like size selectivity of fisheries ('Lopt', STECF 2020) and the age-based selectivity indicator for juvenile fish (Vasilakopoulos et al, 2020, ICES WKLIFE). While these indicators have a clear link to the objectives of the CFP, it remains unclear whether they are appropriately placed under D3C3 as they are linked directly to fishing pressure rather than stock status and health.

The recently published MSFD guidance (European Commission, 2022) highlighted the need to develop D3C3 indicators with threshold values that are compatible with the threshold values of D3C1 and D3C2 to ensure simultaneous assessment of GES and to expand the focus of D3C3 beyond size and age to include aspects of recruitment, individual growth, condition and natural mortality.

DGENV requested ICES to:

1. Define characteristics of a *'healthy population structure'* for species with different life history traits and identify relevant indicators for these characteristics.
2. Identify thresholds of *'healthy population structure'* indicators and for species with different life history characteristics.
3. Explore the relationship between population traits/dynamics and healthy population structure' indicators and thresholds through simulations and infer cases where management in the context of CFP objectives -and equally of MSFD D3C1 and D3C2- alone may be insufficient and additional management measures should be envisaged. In such cases, and
4. depending for example on the characteristics and exploitation patterns of the populations concerned, suggest a set of management options, ranked in decreasing order of expected effectiveness.

5. Advise indicators and thresholds most suitable for D3C3 assessment for species with different life history characteristics, giving preference to indicators that are derived from easily collected data (*e.g.* data routinely collected under the DCF).
6. Prepare a framework for comprehensively assessing D3 criteria for commercially-exploited fish and shellfish populations (= stocks), including data-limited stocks

To answer this request, ICES will organize 3 workshops. WKD3C3SCOPE, WKD3C3THRESHOLDS and WKD3SIMUL. For details on dates and terms of reference, please see the ICES webpage: https://www.ices.dk/community/groups/Pages/default.aspx

2.2 Setting of WKSIMULD3 in relation to WKD3C3SCOPE and WKD3C3THRESHOLDS

The three workshops on D3C3 in this series each cover separate aspects. The first workshop WKD3C3Scope identified characteristics of healthy populations, indicators to measure these health characteristics and criteria to select among indicators at the second workshop WKD3C3THRESHOLDS (18-21 September 2023). At WKD3C3THRESHOLDS, as many as possible of the suggested D3C3 indicators were calculated, validated and evaluated for a selection of stocks representing different life-histories, data availability and MSFD (sub)regions. Subsequently, methods for setting thresholds for these indicators were suggested together with consequences for the stock if health indicators fall below the threshold. The consistency and complementarity with D3C1(FMSY) and D3C2 (MSYBtrigger) is evaluated and a framework for the comprehensive assessment of D3 stocks is to be drafted in WKSIMULD3. The framework will include data requirements to assess D3, recommended indicator(s) for the assessment of D3C3 that are compatible with D3C1 and D3C2 and methods to set thresholds and reference levels. The third workshop WKSIMULD explores the relationships between indicators of population traits/dynamics and healthy population structure through simulation. The workshop infers cases where management under CFP objectives alone may be insufficient and rank potential management options in decreasing order of expected effectiveness to remedy adverse effects on or of stock health.

2.3 Background of D3C3 among D3C1 and D3C2

The indicator and assessment threshold of fishing pressure (F) in D3C1 and abundance (SSB) in D3C2 are already well established within the ICES stock assessment framework. Both have been embedded into the MSFD from the start and can be considered as operational for stocks with analytical stock assessments.

The purpose of D3C3 is to assess the health of the population structure within an exploited stock. As discussed during WKD3C3SCOPE, a healthy stock should be abundant, productive and contain a wide age and size distribution, taking the maximum age/size of the species into account. The age distribution reflects mortality and thereby expresses the chance of survival to high age and large sizes. In accordance with estimations of the benefits of old spawners in stocks of rockfishes (Berkley et al. 2004), a population can be divided in components (Figure 2.3.1), each of which reflects a different aspect of stock productivity and is sensitive to environmental conditions and mortality, including the intensity of fishing pressure (Table 2.3.1). The proportion of old and large individuals reflects the survival from juvenile to a large adult allowing individuals to have different roles within the stock (e.g. with regards to spawning, migration) and the ecosystem (e.g. predation, biomass transport).

WKD3C3SCOPE and WKD3C3THRESHOLDS considered that productivity could be assessed within D3C3 by recruitment (R) and growth. Time series on R are available outputs from analytical stock assessments and could be assessed against B_{lim} (Figure 2.3.2). Growth indicators can be estimated from weight at age data from stock assessments.

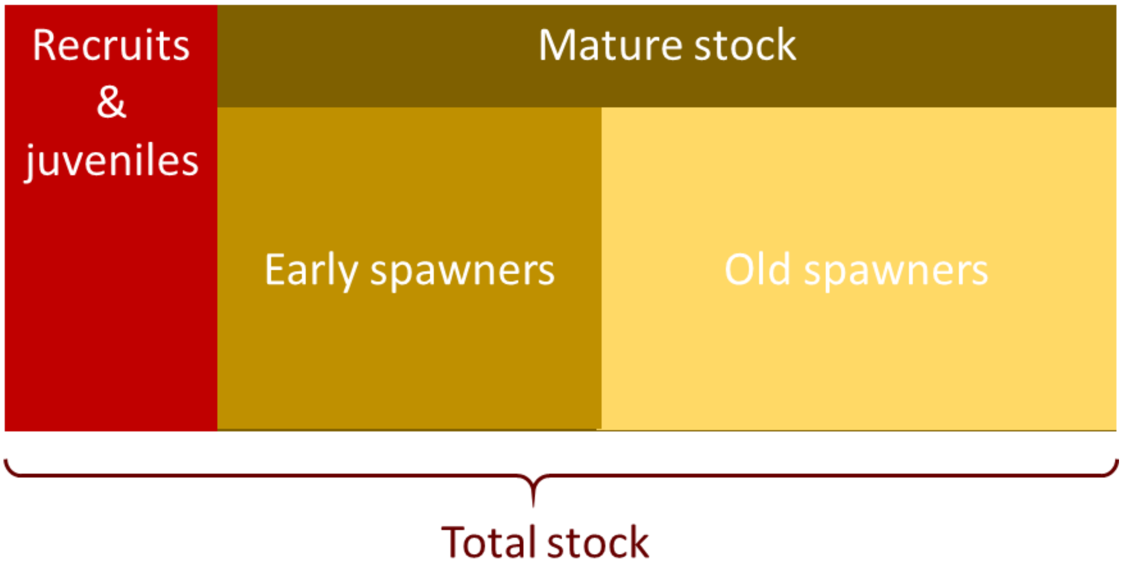


Figure 2.3.1. A fished stock can be divided into different sub-components i.e. recruits/juveniles and spawners. The latter can be further divided into early spawner and old spawners. The assessment of a healthy population structure can account for all these stock components, as each component reflects a different aspect of stock health. However, for short-lived species and one-time spawners (e.g. salmon), the distinction between early and old spawners may be obsolete.

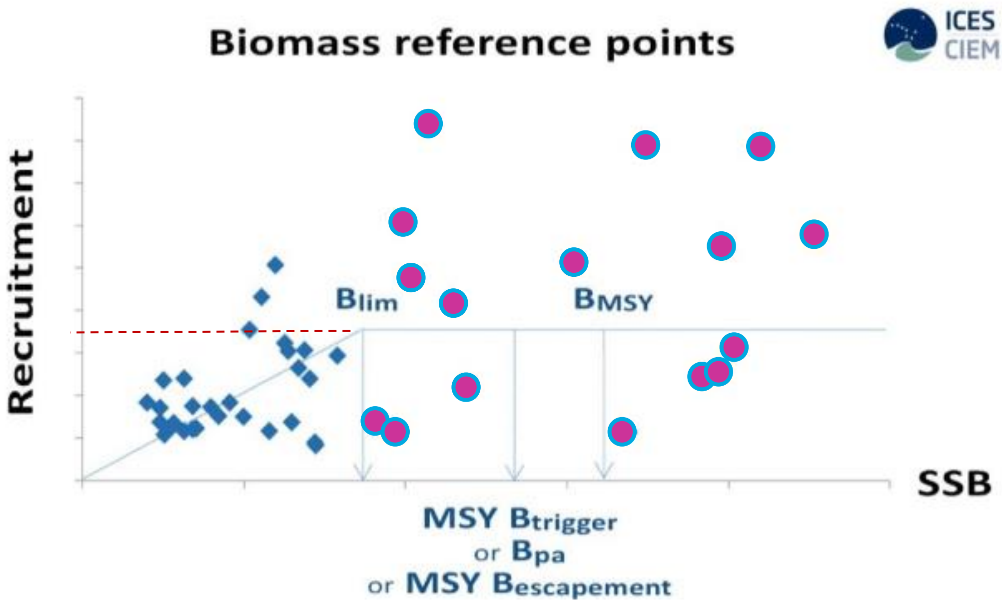


Figure 2.3.2. A spawner-recruit-relationship can be used to identify recruitment (R) values which occur above Blim (purple dots). From the distribution of these R-values the 5th-percentile could be defined as reference point. Modified from ICES Advice on fishing opportunities 2023 (https://doi.org/10.17895/ices.advice.22240624).

The range and width of age/size distribution should be addressed by age-based or size-based indicators. WKD3C3SCOPE; WKD3C3THRESHOLD and WKSIMULD3 analysed various indicators and other indicators have also been proposed (see Shin et al., 2005; Probst et al., 2013; Froese et al., 2016; ICES, 2016a; ICES, 2016b for examples; Probst et al., 2021). However, for all indicators the definition of assessment thresholds has been the central problem. The most critical point when defining thresholds is centred around the question of how stock health should be defined as fishing cannot be conducted without impact and hence any form of fishing will ultimately influence the stocks age and size composition.

For stocks which are assessed regularly by analytical stock assessments, indicators are available for each aspect concerning the health of the stocks age and size structure. All D3 indicators should be assessed as six-year averages of the MSFD assessment period (Probst et al., 2021).

2.4 Data requirements to assess D3

The participants at WKD3C3THRESHOLDS agreed that health of the stock as assessed under D3C3 can only be evaluated if information on recruitment, weight at age, size/age distribution and/or fisheries selection pattern of the stock is available. If these data are not available, D3C3 cannot be assessed and Member States should be encouraged to remedy through increased data collection. For stocks with age based assessments, the necessary data are all input to or output from the assessment. D3C3 assessments based on length distributions alone appear to provide more noisy results than assessments based on age data.

2.5 Indicator(s) for the assessment of D3C3 that are compatible with D3C1 and D3C2 and recommended by WKD3C3THRESHOLDS

WKSIMULD3 followed the indicators recommended by WKD3C3THRESHOLDS for stocks with age based assessments but added the indicator SSB/R though this was considered too responsive to recruitment:

1. ABI_{MSY}
2. Proportion of old spawners by biomass (POS)
3. Average age of spawners (ASA)
4. SSB/R
5. Recruitment from stock assessment output (R)
6. Average weight at age anomaly from stock assessment input (ASW)
7. Fjuv/Fapical

For stocks with length based assessments or length based survey catch rates, the indicators considered were:

1. Recruitment index from e.g. catch at lengths below recruitment length (R)
2. The 90^{th} percentile of length of individuals larger than the length at recruitment (L90R)
3. Where recruitment length is not available, the 90^{th} percentile of length of individuals (L90)

2.6 Data used for historical evaluations

2.6.1 North East Atlantic

The Northeast Atlantic dataset used to assess the proposed indicators for D3C3 consisted of 81 stocks spanning 26 species. The dataset is formed of stock objects of the 'FLStock' class as defined within the FLR framework (Kell et al. 2007) and was collated as part of Workshop on ICES reference points (WKREF1; ICES, 2022) and updated in the work of Griffiths et al. (2023). During the WKD3C3THRESHOLDS meeting, the stock objects for five short lived stocks of small pelagics (spr.27.3a4, san.sa.1r, san.sa.2r, san.sa.3r and san.sa.4) were also updated.

All 81 stock objects contain the input and/or output from each stock's respective stock assessment (including all available information on stock status), and have a final assessment year among the years 2019-2022. All of these stock assessments are classified as Category 1 by ICES, meaning they involve the application of age-based analytical stock assessment methods. If available, all stock objects also contain a vector of stock-specific reference points including F_{MSY}, B_{MSY}, MSY $B_{trigger}$ and B_{lim}. F_{MSY} and B_{lim} are directly used in the calculation of the indicator ABI_{MSY}.

A graphical summary for this dataset is provided in Figure 2.6.1, including the number stocks per species, area and assessment method. The full dataset is freely available via ICES as part of WKREF1 or WKD3C3Thresholds, or can be accessed directly on GitHub at https://github.com/cagriffiths/ABIs-fish.

During some of the analyses in KD3C3THRESHOLDS, stocks were aggregated by exploitation level and life span. Stocks were split into three exploitation level groups of $F/F_{MSY} < 1$ (medium; n = 12), $F/F_{MSY} >= 1$ and < 2 (high; n = 42) and $F/F_{MSY} >=2$ (very high; n = 18), as well as three life span groups of short (i.e., sandeel, sprat, sardine, and herring), long (i.e., anglerfish, redfish, tusk and ling) and medium (all other stocks). Stocks were assigned to exploitation level groups based on the average F/F_{MSY} over the whole stock assessment time series, whereby the length of that time series differed across stocks.

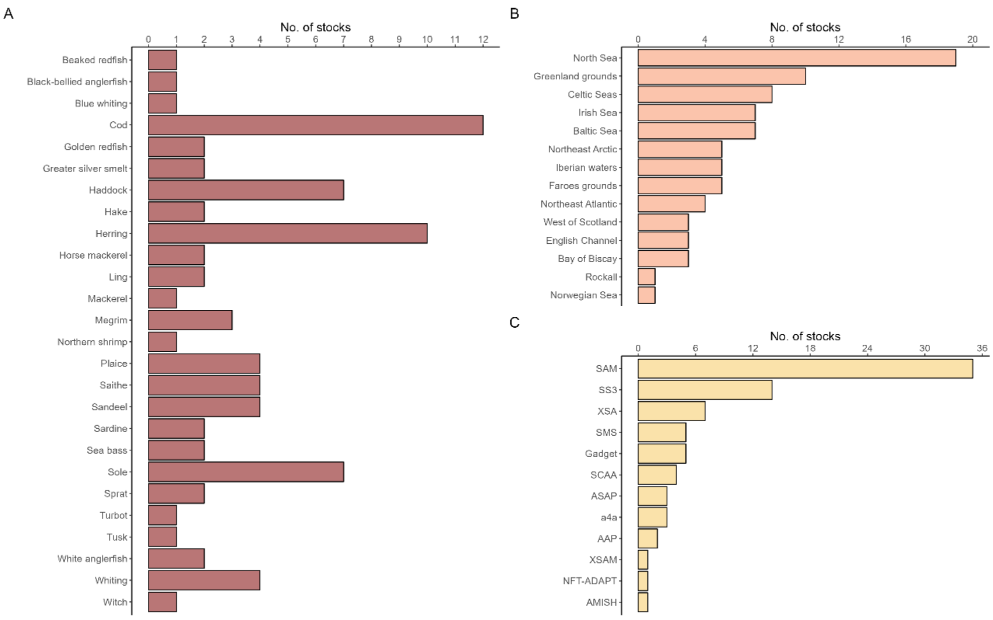


Figure 2.6.1. Graphical of the Northeast Atlantic dataset. The dataset contains 26 species (A) from 14 different areas (B), whereby area is a broad characterisation based on ICES areas and ecoregions. All 81 stocks are assessed using age-structured analytical assessment frameworks (C).

2.6.2 Mediterranean

For the Mediterranean Sea, data were derived from the latest official scientific advice reports of STECF, available online (ANNEX II - stecf.jrc.ec.europa.eu/reports/medbs). The assessments for these stocks were carried out by the two stock assessment working groups for demersal species in the Mediterranean Sea (STECF, 2022a; 2022b). The stocks available in the dataset have an accepted analytical stock assessment and were provided in the form of FLStock, a class of the FLR library (Kell et al., 2007). For those stocks for which biomass reference points have been calculated, an additional slot was added to the FLStock object, named "benchmark", which contains the information on F_{MSY} and B_{MSY}. In total, 28 stocks were provided, with 12 of them accompanied by their biomass and fishing mortality reference points.

Stocks having an age structured stock assessment and mostly falling in geographical sub-areas (GSAs) covered by surveys under the responsibility of the Italian government were selected to produce length-based indicators (Table 3.3.1). TA, TB, and TC files storing the MEDITS survey data (Bertrand *et al.*, 2002) for the selected stocks were made available to the Italian experts involved in the working group for producing elaborations needed. The data included timeseries of length-based indicators.

Table 2.6.1. Stocks inspected for length-based indicators. GSA is the mediterranean geographic sub area.

Species	GSA	Stock ID	Length threshold for recruits (mm)
ARS	9; 10; 11	ARS9_10_11	14.8
HKE	8; 9; 10; 11	HKE8_9_10_11	150
HKE	17; 18	HKE17_18	150
MUT	9	MUT9	105
NEP	9	NEP9	20

3 Methodology to identify thresholds

3.1 Methods to identify thresholds suggested by WKD3C3THRESHOLDS

WKD3C3THRESHOLDS produced a decision tree to identify the most appropriate threshold estimation approach (Fig. 3.1.1). The suggested threshold definition methods were:

1. Estimate the indicator level at which there is a clear decrease productivity at low indicator levels (similar principle to Blim)
2. Estimate 5, 10, 50, 90, and 95 percentiles of the indicator from MSE simulations of fishing according to the MSY approach (decision on percentiles pending).
3. Estimate 5, 10, 50, 90 or 95 percentiles of the indicator using data from a minimum of 3 generations years with near sustainable exploitation available (decision on percentiles pending).
4. Estimate the trend of the indicator using historic data (mostly from unsustainable fishing)

To these suggestions, WKSIMULD3 added the options

5. Estimate 5, 10, 50, 90, and 95 percentiles of the indicator from MSE simulations of no fishing (decision on percentiles pending).
6. Estimate 5 and 95 percentiles of the indicator from MSE simulations of fishing at $F_{P0.5}$.
7. Estimate 95 percentiles of the indicator using data periods of unsustainable exploitation available (two levels of unsustainable exploitation examined, 1.5*F_{MSY} and 2*F_{MSY}).
8. Estimate median or average threshold across MSEs investigated in WKSIMULD3

There was no evidence to suggest that specific levels of the indicators led to a clear decrease in productivity or a clear breakpoint in the simulated relationships with fishing pressure (see section 6). In the absence of evidence to suggest a threshold linked to observed harm, a percentage of 5 percent was agreed to be the most relevant for options 2 and 3 as this would mean that there was no inherent conflict between D3C1, D3C2 and D3C3 thresholds.

Following this, the implications of each of the 8 methods to derive thresholds were investigated by comparing the thresholds to the historical development of the indicator along with the historical development in F, SSB and R. The simulated indicators from the MSE were first investigated for signs of clear effects on stock status or productivity. Then, the magnitude of the change in the indicator as a result of changing fishing mortality was compared to the natural variation between years when fishing at a constant fishing pressure either with a harvest control rule decreasing F at low stock size or without this reduction.

Even when fishing according to an MSY management approach, the indicators are expected to vary between as a result of variation in e.g. productivity of the stock. As a result, the inspections also considered whether the natural variability when fishing according to the MSY approach was so large that a change due to increased fishing would not be detected, in which case the indicator was thought not to be useable for management decisions (approximately equal number of false positives, true positive, false negatives and true negatives).

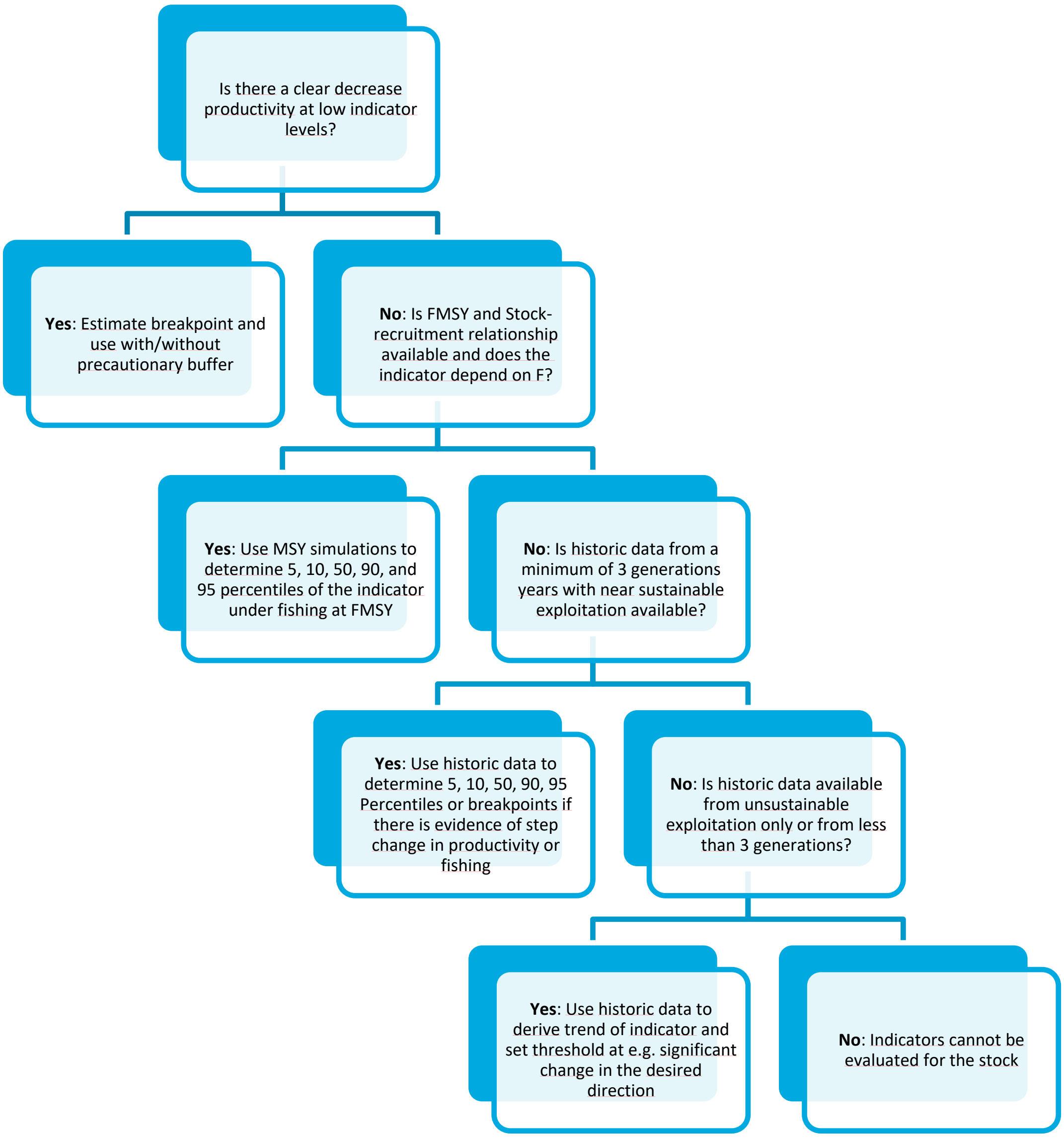


Fig. 3.1.1. Decision tree to choose a threshold setting method from WKD3C3THRESHOLDS.

3.2 Examining the use of Receiver operating characteristic (ROC) curves

Receiver operating characteristic (ROC) curves are a graphical method used to assess and compare the performance of binary classification models. Within fisheries science, an indicator of stock status and its discriminatory threshold can be considered a binary classification model. The combination of the indicator and threshold classifies stock status at each interval as either positive or negative (i.e. the stock is within safe biological limits or is impaired), depending on whether the indicator value is above or below the threshold. ROC curves enable us to easily validate these predictions against observations, providing that we have observations of true stock status to compare to. However, knowing true stock status is rare and instead, a reliable, well-established indicator of stock status must be used (see WKLIFE XII section 5.2 for an example where spatial indicators are validated against SSB/MSY $B_{trigger}$ outputs from data-rich stock assessments; ICES, 2023).

By comparing binary predictions of stock status to the true binary states of the stock, a confusion matrix can be formed summing the counts of true positives (TP), instances where the indicator correctly predicts stock status to be positive; false negatives (FN), instances incorrectly classified as negative; false positives (FP), instances incorrectly classified as positive; and true negatives (TN), instances correctly classified as negative. The confusion matrix can then be used to calculate the true positive rate (TPR = TP/(TP+FN)) and false positive rate (FPR = FP/(FP+TN)). The TPR is the proportion of positive instances correctly classified as positive whereas the FPR is the proportion of negative instances incorrectly classified as positive. An indicator and the threshold that performs well will maximise TPR while minimising FPR (for a perfect indicator TPR = 1 and FPR = 0). The performance of the indicator and current threshold level can be summarised by calculating the true skill score (TSS = TPR-FPR; TSS = 1 for a perfect indicator).

The ROC curve is formed by repeating the processes needed to calculate and plot TPR and FPR for every possible threshold level. The area under the curve (AUC) summarises the overall classification performance of the indicator (AUC = 1, perfect; AUC = 0.5, random; AUC = 0, worst) and can be used to easily compare different indicators. At some point on the ROC curve, TSS will be maximised, representing threshold level that optimises classification performance of the indicator. This optimal threshold can be derived and used accordingly.

ROC curves therefore serve a useful way to assess relationship between an indicator and true stock status, compare classification performance of different indicators, and to derive thresholds that optimise performance. WKSIMULD3 participants discussed how this approach could be of use in the workshop but were not able to solve the problem that the true state of the stock would need to be known for it to be used.

4 Tor a: Building on from preparatory work, discuss and agree on the stocks to run simulations

Constructing and running a stock specific MSE requires careful consideration of the implementation of uncertainty and variability and quality assurance through review by experts on MSEs. To ensure that the MSEs used in WKSIMULD3 were peer reviewed, it was decided to use existing MSE models that had been peer reviewed and used directly in ICES or models clearly demonstrated the ability to replicate ICES EQSIM results, as EQSIM is the default MSE for estimating F_{MSY} in ICES. The chairs contacted persons in the ICES system who had led on MSEs and received positive replies and data from simulations of the stocks Northeast Atlantic Mackerel (Thomas Brunnel, pers. comm), North Sea sprat and North Sea area 1 sandeel (Mollie Brooks, pers. Comm.). In addition, an MSE was implemented in FLBEIA for eight demersal stocks in the North Sea and Eastern Channel (cod.3a47d, ple.7d, ple.3a4, pok.3a4, sol.4, sol.7d, tur.4, had.3a4) and results validated against EQSIM results for the same stocks. The fit for haddock was considered insufficient due to the modelling of the variability in recruitment and this stock was therefore not considered further.

5 Tor b and c: Run simulations to explore the relationships between indicators of population traits/dynamics and healthy population structure and infer cases where management under CFP objectives alone may be insufficient and compare results with recent MSE simulations

5.1 Potential simulation tool SSM

5.1.1 Methods

A size spectrum model (SSM) was presented by Elizabeth Duskey. It has been used to simulate the influence of environmental variation on both age- and length-based indicators for a multispecies fishery in the Central Baltic Sea. This fishery includes Baltic cod, flounder, sprat, and herring. Cod was chosen as a focal species, as it suffers from extreme declines in stock health (Eero et al. 2020). The SSM approach treats body size (g) as the master trait determining individual physiological and behavioral rates such as metabolism, mortality, predation, and diet. Species-specific parameters such as mass conversion efficiency, density dependence, growth patterns, and feeding preference may also be included. Fishing may occur by species, gear, and/or habitat. The McKendrick von Foerster equation (McKendrick 1925 and von Foerster 1959) guides the determination of equilibrium biomass for all species. For this simulation tool, the baseline SSM structure provided by the R package **mizer** (Delius et al. 2023) was altered to include individual-level physiological and behavioral responses to fluctuations in oxygen and temperature. Benthic and pelagic habitats were separated to represent different food sources and limited overlap of species throughout the water column. Models of this kind have been used to explore the mechanisms leading to altered community dynamics in warming, hypoxic waters (Duskey 2023). For this workshop, rates of annual growth, as well as size spectra (number of individuals in logged body weight bins) were used to calculate age- and size-based indicators. Scenarios of oxygen, temperature, and fishing of important prey species for cod were explored to determine the magnitude of the resultant change in indicators.

5.1.2 Results

Age- and length-based indicators show expected declines or hump-shaped relationships with increasing fishing mortality in the model (Figure 5.1.2.1). Some respond strongly to changes in fishing mortality, including SSB/R, L90, and A90prop/A90fmsy (Figure 5.1.2.1). Note that patterns are similar across scenarios, though the magnitude may differ (Figures 5.1.2.2 and 5.1.2.3). The determination of thresholds based on any discernible peaks or overall distributions of the indicators may thus change from year to year in response to dynamic climate, and to fishing on other components of the food web. The major exception is A90prop/A90fmsy, which sees a dramatic change in response to deoxygenation (Figure 5.1.2.2). Some indicators, such as recruitment (R; calculated here as the number of individuals entering the mature size class) do not appear to be appropriate representations of the processes at work in this model, and warrant further investigation. This SSM framework can be used to explore how changes in fecundity, density dependence, and food availability affect fishes across life history gradients, elements which are often excluded from traditional models. In general, these results suggest that it is worthwhile to consider how indicators may respond to environmental variation when determining appropriate thresholds.

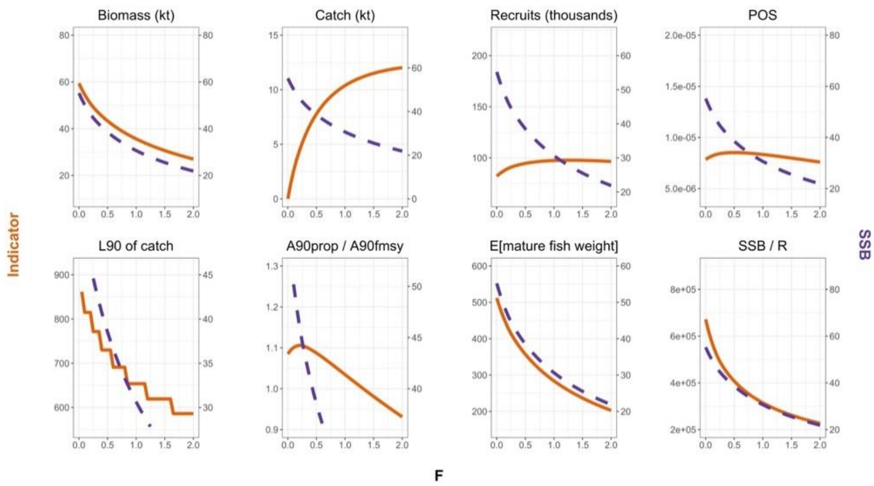


Figure 5.1.2.1. Modelled indictors for Baltic Cod at 5-year means for oxygen, temperature, and fishing mortality.

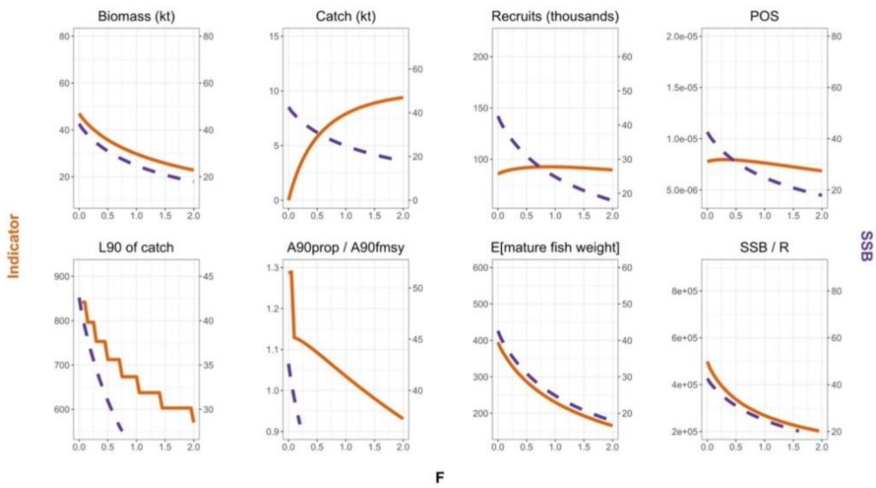


Figure 5.1.2.2. Modelled indicators for Baltic Cod with a 5% decline in benthic oxygen.

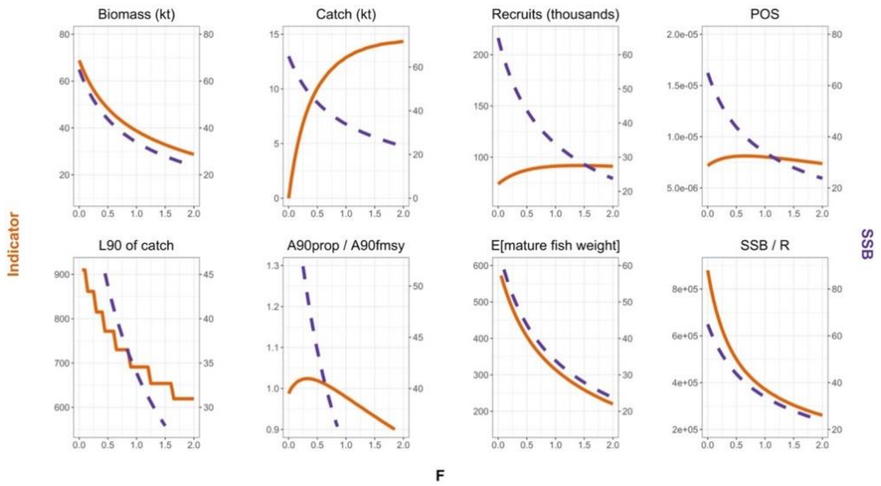


Figure 5.1.2.3. Modelled indicators for Baltic Cod with a 50% decline in fishing on sprat.

5.2 Demersal stocks in the North Sea

5.2.1 Methods

5.2.1.1 MSE-setup

The work was based on MSE-baseline simulations conducted under the SEAwise project, aiming to ensure consistency of the FLBEIA modelling framework with the assumptions of recent benchmarks and associated reference points.

The demersal stocks of the North Sea were simulated using stock assessment summary objects provided to the Working Group on Mixed Fisheries Advice (WGMIXFISH) by the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK) (ICES 2022). Stock objects are of class *FLStock,* from the *FLCore* R package (Kell et al., 2007), and include catch data and estimated stock parameters up to 2021. The evaluated stocks were age-based stocks included in the demersal mixed fishery model of the North Sea, including cod, haddock, saithe, plaice (North Sea and English Channel), sole (North Sea and English Channel) and turbot (see table 5.2.1.1 for stock codes).

Fmsy and Fp05 reference points (see definitions below) were re-estimated to check consistency with reported values and evaluate the model set-up. Two approaches were used; 1. *Eqsim* (stochastic equilibrium reference point software) of the *msy* R package (Simmonds et al., 2022), and 2. management strategy evaluations (MSE) using the *FLBEIA* (Garcia et al., 2017). Both Eqsim and MSE approaches are permitted according to the ICES guidelines for reference point estimation for category 1 and 2 stocks (ICES, 2021), although Eqsim is currently used in the most recent benchmark assessment of all stocks considered here.

Settings were maintained consistent with the most recent benchmark assessment (Table 5.2.1.1); in:

1. the year span of data for stock recruitment relationship (SRR) fitting,

2. the year span of biological parameters for resampling in forecast years,

3. the year span for resampling selectivity at age in forecast years, and

4. uncertainty in spawning stock biomass (SSB) estimation and advised fishing mortality implementation (*Fcv*) and autocorrelation (*Fphi*).

According to ICES guidelines (ICES 2021), fishing mortality associated with maximum sustainable yield (*Fmsy*) (in landings) is estimated without the use of the ICES HCR, but includes estimation and implementation errors. The maximum fishing mortality resulting in <5% probability of SSB below Blim (*Fp05* or *Fpa*) is estimated with the use of the ICES HCR. When Fp05 < Fmsy, Fp05 is used as the Fmsy reference point.

For all stocks, resampled data was used to fit SRRs to the defined models (*eqsr_fit* function, n = 2001), followed by forecasts over a range of F values (*eqsim_run* function). Forecasts were run for 200 years and the final 50 years were used for Fmsy and Fp05 estimation.

Table 5.2.1.1. Eqsim settings derived from most recent benchmark assessments.

Stock-abbreviation	Stock codes	srr_years	srr_models	bio_years	bio_const	sel_years	sel_const	Fcv	Fphi	SSBcv
COD-NS	cod.27.47d20	1998-2019	segreg3	2015-2019	FALSE	2015-2019	FALSE	0.140	0.440	0.2
HAD	had.27.46a20	2000-2020	segreg3	2011-2020	FALSE	2011-2020	FALSE	0.212	0.423	0.0
PLE-EC	ple.27.7d	1980-2021	Segreg, Bevholt	2017-2021	FALSE	2017-2021	FALSE	0.212	0.423	0.2
PLE-NS	ple.27.420	1957-2020	Segreg, Bevholt	2018-2020	FALSE	2011-2020	FALSE	0.212	0.423	0.2
POK	pok.27.3a46	1998-2017	segreg3	2008-2017	FALSE	2013-2017	FALSE	0.212	0.423	0.2
SOL-EC	sol.27.7d	1982-2020	Segreg	2017-2021	FALSE	2017-2021	FALSE	0.212	0.423	0.2
SOL-NS	sol.27.4	1958-2018	Segreg	2014-2018	FALSE	2014-2018	FALSE	0.212	0.423	0.2
TUR	tur.27.4	1981-2017	Ricker, Segreg	2013-2017	FALSE	2013-2017	FALSE	0.212	0.423	0.2

***Notes*: *srr_years* = The years from which to sample spawning stock biomass (SSB) and recruitment data pairs for estimating the stock-recruitment relationship (SRR); *srr_models* = SRR models fit to resampled data in determining best-fitting relationships ("Ricker" is Ricker, "Segreg" is segmented regression with freely-estimated breakpoint, "segreg3" is segmented regression with fixed breakpoint, "Bevholt" is Beverton-Holt); *bio.years* = The years from which to sample maturity, weights and natural mortality; *bio.const* = If TRUE, mean of the biological values from the years selected are used (usually only applies to the estimation of Fpa with HCR); *sel.years* = The years from which to sample selection patterns; *sel.const* = If TRUE mean of the selection patterns from the years selected are used (usually only applies to the estimation of Fpa with HCR); *Fcv* = Assessment error in the advisory year; *Fphi* = Autocorrelation in assessment error in the advisory year; *SSBcv* = Spawning stock biomass error in the advisory year.**

FLBEIA was used to test whether similar reference points could be estimated to those of Eqsim. FLBEIA generated output was preferred to those of Eqsim in WKD3C3-SIMUL as that the output could be easily accessed in the form of Stock objects (*FLStock*), suitable for the D3C3-indicator calculation. Eqsim-derived SRRs were also used by FLBEIA. Due to more extensive computation time of FLBEIA, only 100 iterations per SRRs were evaluated, but the subset of SRR models maintained the frequencies of the full set estimated by Eqsim (e.g. 4% Ricker (4), 46% Segreg (46), 50% Bevholt (50)). A slightly smaller range of F values were evaluated (0-0.7), and forecasts were run for 100 years, with the final 10 years used for the estimation of Fmsy and Fp05. Fmsy was estimated without the application of a specific HCR, whereas Fp05 and the associated risk of falling below Blim was estimated using forecasts with the application of the ICES HCR, respectively.

Estimation and implementation errors were not considered. A "short-cut" MSE was conducted that assumed perfect knowledge of the stock (i.e. without assessment loop) in order to improve the speed of the simulations as well as due to the varying assessment approaches used in reality across stocks. Interannual recruitment variability was introduced by the addition of log-normal random noise of standard deviation equal to the coefficient of variation (CV) of residuals for each SRR.

5.2.1.2 Reference point estimation

F_{msy} was estimated by fitting a smooth spline model to the resulting landings at F, and determining the F associated with the maximum of the yield curve. Due to the lower number of runs, the probability of SSB < Blim at each F interval was estimated via bootstrapping, whereby SSBs were resampled (1000 draws with replacement) to get a more robust estimate of risk probability. The maximum F where the 5% risk quantile is <0.05 was used to define Fp05.

5.2.1.3 Calculation of D3C3 Indicators

Based on the MSE simulations of the demersal North Sea stocks, six D3C3 indicators: recruitment (R), SSB/R, average spawner age, average spawner weight (ASW), proportion of old spawner (POS), the relative age distribution in relation to the age distribution at Fmsy (ABI_{msy}) were calculated for further use in the workshop. If not otherwise stated, the same code and definitions were used as defined in WKD3C3THRESHOLDS. For ABI_{msy} the reference age distribution at F_{msy} and resulting reference proportion of the stock above this age was calculated based directly on the estimated reference points and output of the simulations, rather than running a separate stock projection via the *FLBRP*-package as in the original code. The closest F-interval to F_{msy} was used for calculation of the reference level, which could lead to deviations (normally ABI_{msy} at F_{msy} should be 1) since the F-intervals were quite coarse (0.25 increments). For POS the splitting of the age distribution into young and old spawners was based on the historical time series only (up to the year 2021) and using this age as a reference for the indicator in the forecast period.

5.2.2 Results

5.2.2.1 Consistency of reference points

A graphical example of FLBEIA outputs in terms of landings and risk by F level is presented in Figure 5.2.2.1. A summary of fishing mortality reference points from Eqsim and FLBEIA are presented in Table 5.2.2.1. In most cases the values are similar to the existing reference points, providing confidence in the model set-up. Deviations can, in some cases, be explained by existing reference points being estimated several years ago (e.g. saithe, POK, 2019), while our estimates use current assessment outputs, which may differ somewhat in historical values (e.g. SSB and recruitment). FLBEIA-derived Fp05 estimates may be particularly affected by the low number of iterations evaluated (n = 100) and assumption of perfect knowledge (apart from recruitment in simulated short-term forecasts) within the assessment and management loop.

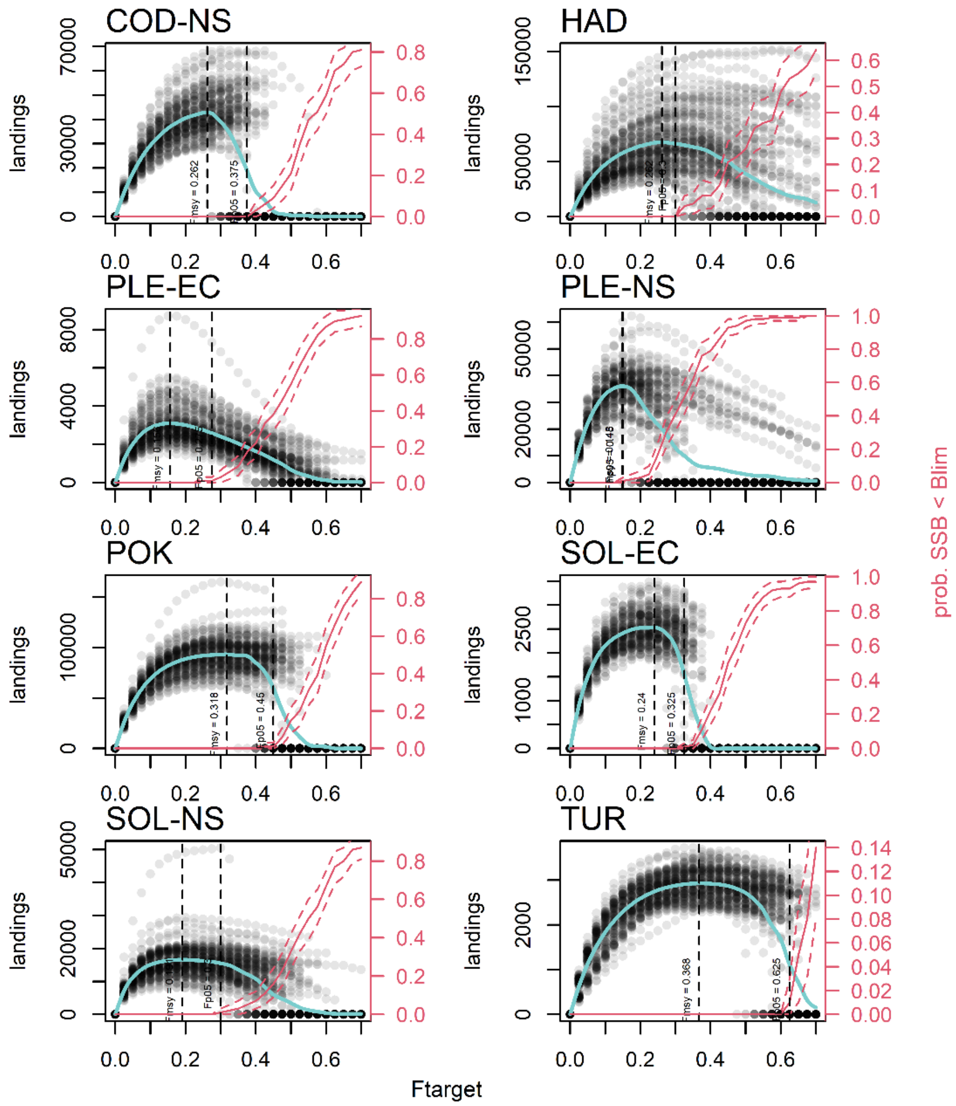


Figure 5.2.2.1. Reference point estimation with FLBEIA output for cod (COD-NS), haddock (HAD), plaice in the English Channel (PLE-EC), North Sea plaice (PLE-NS), saithe (POK), sole in the English Channel (SOL-EC), North Sea sole (SOL-NS) and turbot (TUR). Different SRR fits (n = 100), run across a range of target F values (Ftarget), with their respective long-term yields (tonnes) displayed as points. The blue line represents the spline fit to landings at F to derive F_{msy}, while the solid and dashed red lines show the median and bootstrapped confidence intervals (5% and 95% quantiles), respectively, of the probability of SSB < Blim. The maximum F where the 5% risk quantile is <0.05 is used to define Fp05.

Table 5.2.2.1. Summary of existing and re-estimated fishing mortality reference points.

Stock-abbreviation	Stock code	Fpa	EQ_Fp05	FLBEIA_Fp05	Fmsy	EQ_Fmsy	FLBEIA_Fmsy
COD-NS	cod.27.47d20	0.490	0.474	**0.375**	0.280	0.257	**0.262**
HAD	had.27.46a20	0.240	0.112	**0.300**	0.240	0.210	**0.262**
PLE-EC	ple.27.7d	0.238	0.239	**0.275**	0.156	0.161	**0.156**
PLE-NS	ple.27.420	0.182	0.182	**0.150**	0.152	0.145	**0.148**
POK	pok.27.3a46	0.576	0.446	**0.450**	0.363	0.321	**0.318**
SOL-EC	sol.27.7d	0.318	0.317	**0.325**	0.230	0.234	**0.240**
SOL-NS	sol.27.4	0.311	0.320	**0.300**	0.207	0.225	**0.191**
TUR	tur.27.4	0.856	0.766	**0.625**	0.361	0.354	**0.368**

***Notes*: Fmsy = Fishing mortality associated with maximum sustainable yield; Fp05 = Maximum fishing mortality where the risk of SSB < Blim is less than 5%; *EQ* and *FLBEIA* prefixes denote re-estimated values by Eqsim and the FLBEIA MSE, respectively, while those columns without prefixes are the existing reference points determined by the most recent benchmark assessment.**

Although some differences were found between re-estimated Fmsy and Fp05 reference points and existing values, these were generally low and explainable. For haddock, the group discussed, if the results are reliable given that the unique recruitment behaviour of haddock (intermittent recruitment pattern with occasional large spikes and certain autocorrelation) was not taken into account. In addition, the FLBEIA MSE setup differs in several aspects compared to the Eqsim method, and the lower number of iterations likely affect the robustness of the estimates. Nevertheless, the resulting estimated reference points (especially F_{msy}) were overall quite similar to those coming from Eqsim and benchmark estimates, and were deemed suitable as a basis for the calculation and discussion of D3C3-indicators.

5.2.2.2 D3C3-Indicator behaviour

MSE-simulations showed decreasing values of all D3C3-indicators with increasing fishing mortality (Fig. 5.2.2.2), except for ASW. For recruitment, the values were highly dependent on the used SRR, with the segmented-regression SRR leading to an almost binary response of the resulting recruitment under increasing F, whereas SRRs with increased influence of the Ricker and Beverton-Holt SRRs lead to more gradually declining recruitment levels and flatter yield curves under increasing F. For those stocks where the segmented regression was implemented as SRR, the value of Fp05 corresponded approximately to the level of impaired recruitment or the changepoint of a logistic-function (see overlayed smoothing function). The other indicators (SSB/R, avgAge, POS, ABI_{msy}) declined in a continuous fashion and reached a stable plateau, when the F crossed a specific threshold, likely related to the specific selectivity applied in the simulations. This threshold was close or at the level of Fp05 in some stocks (COD, POK, SOL-EC, TUR), however in others (PLE-NS, PLE-EC, SOL-NS) only began to reach a plateau under F >> Fp05 in the already unsustainable part of the yield-curve (Fig. 5.2.2.2). The level of F_{msy} was hard to track from the D3C3-indicator curves alone, as this was still in the decreasing part of the curve, with no clear change point visible.

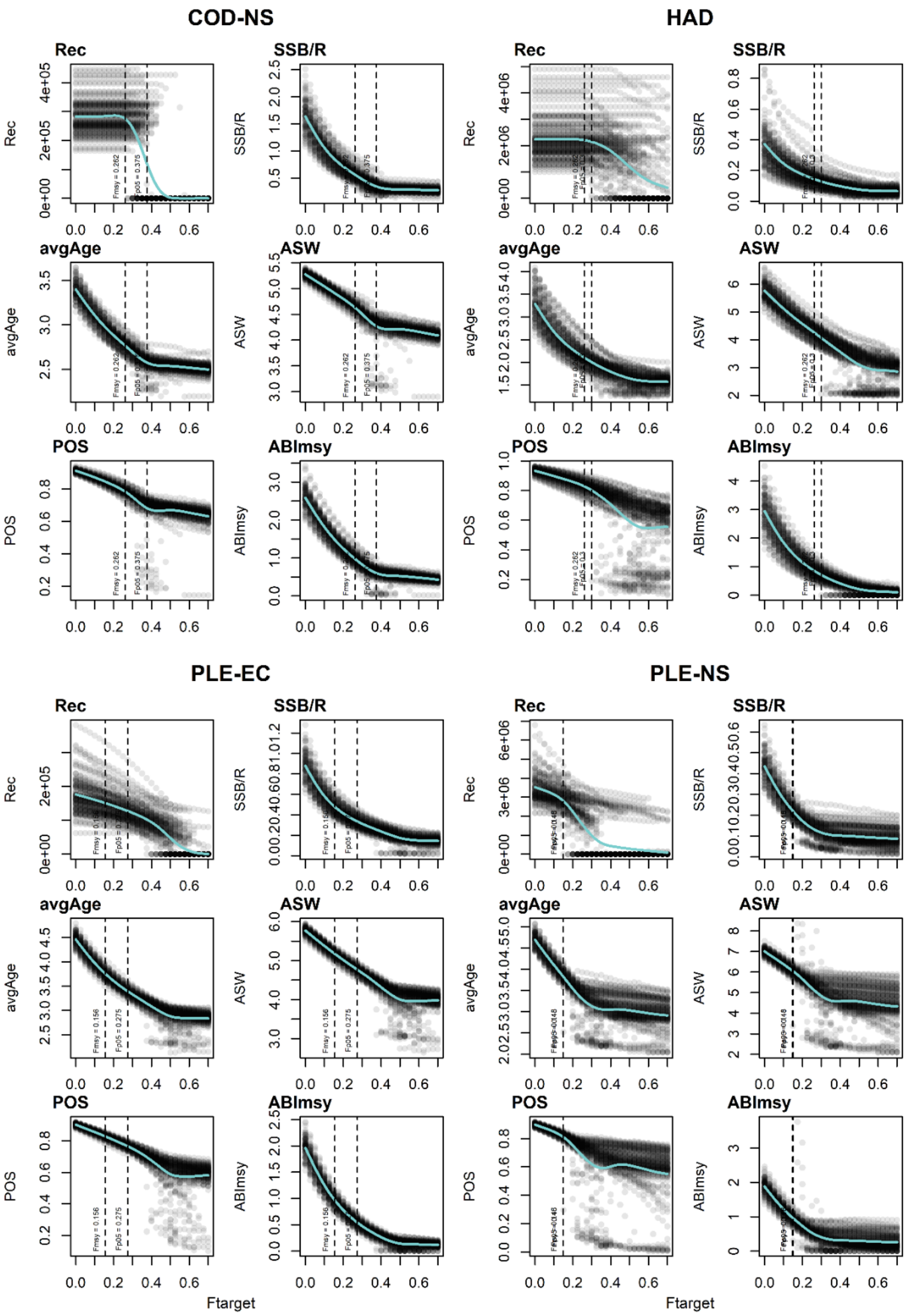


Figure 5.2.2.2: D3C3 indicators (Rec – Recruitment, SSB/R, ASA – average spawner age, ASW – average spawner weight, POS – proportion of old spawner and ABImsy) under different Ftargets for the stocks cod, haddock and plaice (North Sea and English Channel); overlayed are the estimated Fmsy and Fp05 reference points as dashed lines as well as a smoothing line (blue) for easier visualisation.

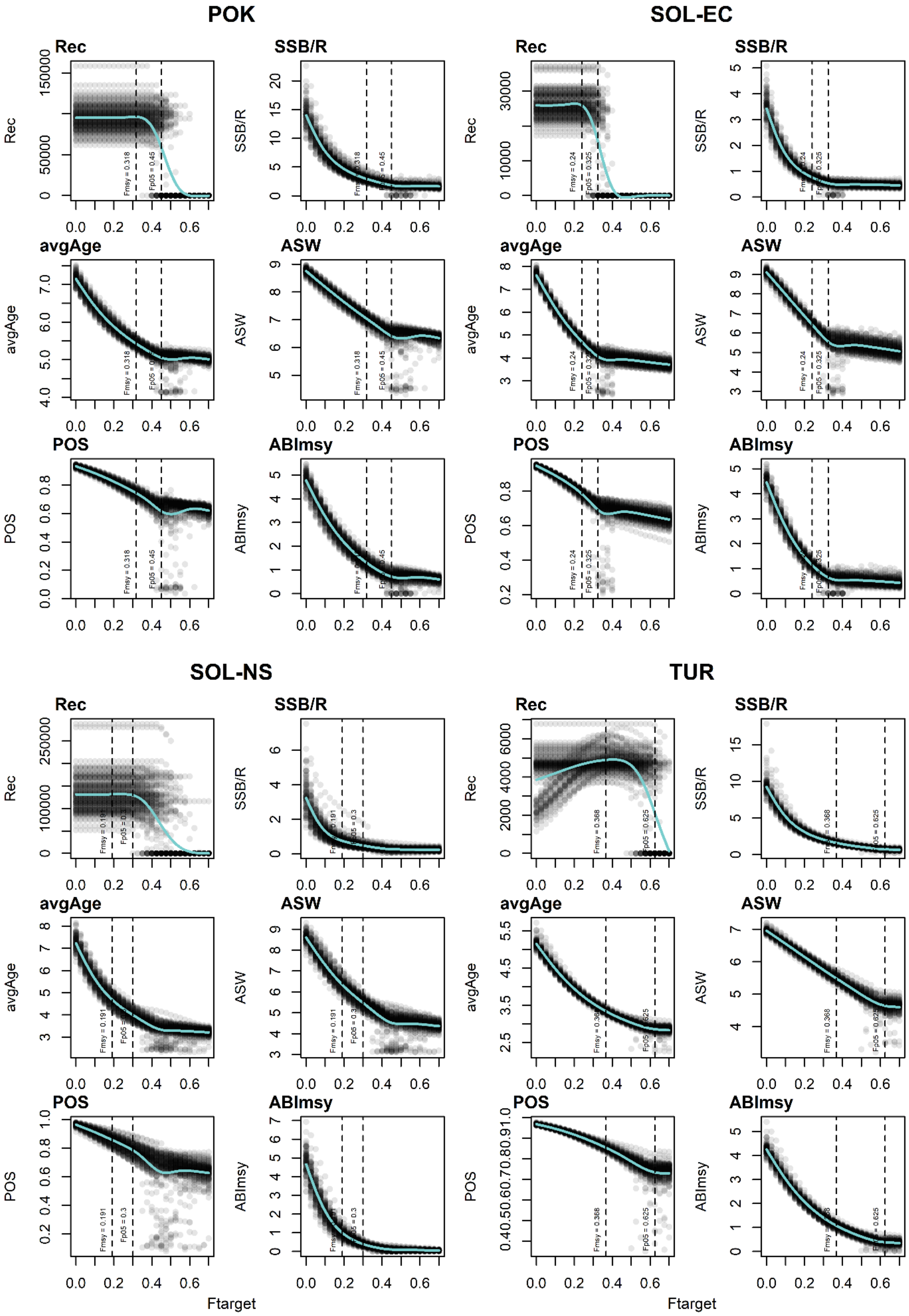


Figure 5.2.2.2 continued: D3C3 indicators (Rec – Recruitment, SSB/R, ASA – average spawner age, ASW – average spawner weight, POS – proportion of old spawner and ABImsy) under different Ftargets for the stocks saithe, sole (North Sea and English Channel) and turbot; overlayed are the estimated Fmsy and Fp05 reference points as dashed lines as well as a smoothing line (blue) for easier visualisation.

5.3 Pelagic and short lived stocks

Pelagic and short lived stocks are interesting from an indicator perspective, as the values of many age and length based indicators depend on the incoming cohort sizes (i.e., recruitment), which tends to be highly variable for these stocks. In the simulations investigated in this working group we focused on three data-rich species, which have full analytical age-based assessments and accepted management strategy evaluations. The species we had available were sandeel (area 1r), North Sea sprat, and Northeast Atlantic mackerel.

5.3.1 Methods

Data

The pelagic and short lived stocks, which had an available MSE that had been through a peer-reviewed process in ICES included three stocks; sandeel (*Ammodytes spp.*) in area 1r, North Sea sprat (*Sprattus sprattus*), and Northeast Atlantic mackerel (*Scomber scombrus*). The two former stocks have an MSE based on the SMS framework (Lewy and Vinther, 2004), and have a short life span, while mackerel have a maximum age of about 12 years and an MSE and assessment based on the SAM framework (Nielsen and Berg, 2014, p 201).

Each of the stocks had data from the most recent stock assessment and the most recent MSE to calculate the indicators ASW, POS, and R, and ABI_{MSY}. Sprat and sandeel do not have an estimate of F_{MSY} and cannot be used to calculate ABI_{MSY}, so that indicator was only calculated for mackerel.

For mackerel, an MSE with an F_{target} of 0.26 equivalent to F_{MSY}, and $B_{trigger}$ of 2.5 MT was used. Sprat used the MSE with the $B_{escapement}$ harvest control rule, and an F_{cap} of 0.69, and B_{PA} of 125000 tonnes. The sandeel MSE was available for 10 different F_{cap} values ranging from 0-0.9, but included only information to calculate average age.

Thresholds

Based on output from the sprat and mackerel MSE, we experimented with calculating lower thresholds as percentiles of three different subsets of data

1) The 5th percentile of historic data
2) The 5th percentile of historic data, where F/Fmsy < 1 (mackerel) or SSB > $MSY\ B_{trigger}$(sprat)
3) The 5th percentile of simulated data from the MSE

5.3.2 Results

Mackerel had a declining trend for all four indicators since the 1980s, but had an increase in the last decade. The MSE predicts on average a slight decline compared to the last couple of years and with a moderate amount of variation between the MSE runs (Figure 5.3.1).

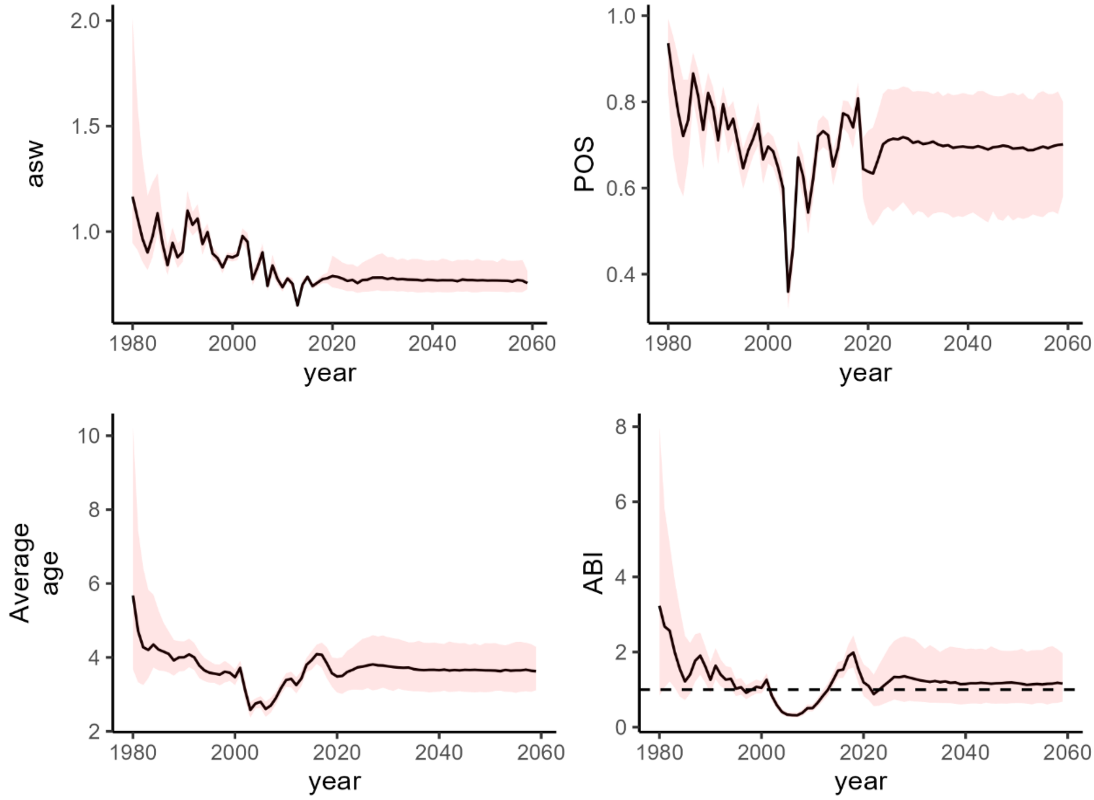


Figure 5.3.4 The four indicators for Northeast Atlantic mackerel. Black line denotes the median over the MSE runs, and the red shade indicates the 95th and 5th percentile. The dashed line on the ABI plot is equal to 1.

The mackerel MSE predicts on average the ABI value is quite close to the predicted median of 1 (Figure 5.3.1).

The sprat MSE (and assessment) has high interannual variability in the indicators POS, asa and recruitment. The variability is likely due to the indicators being dependent on recruitment pulses, as the maximum age of the stock is very short (4 years, average asa is approximately 1.25). This variability is also reflected in the uncertainty in the MSE estimation (Figure 5.3.2).

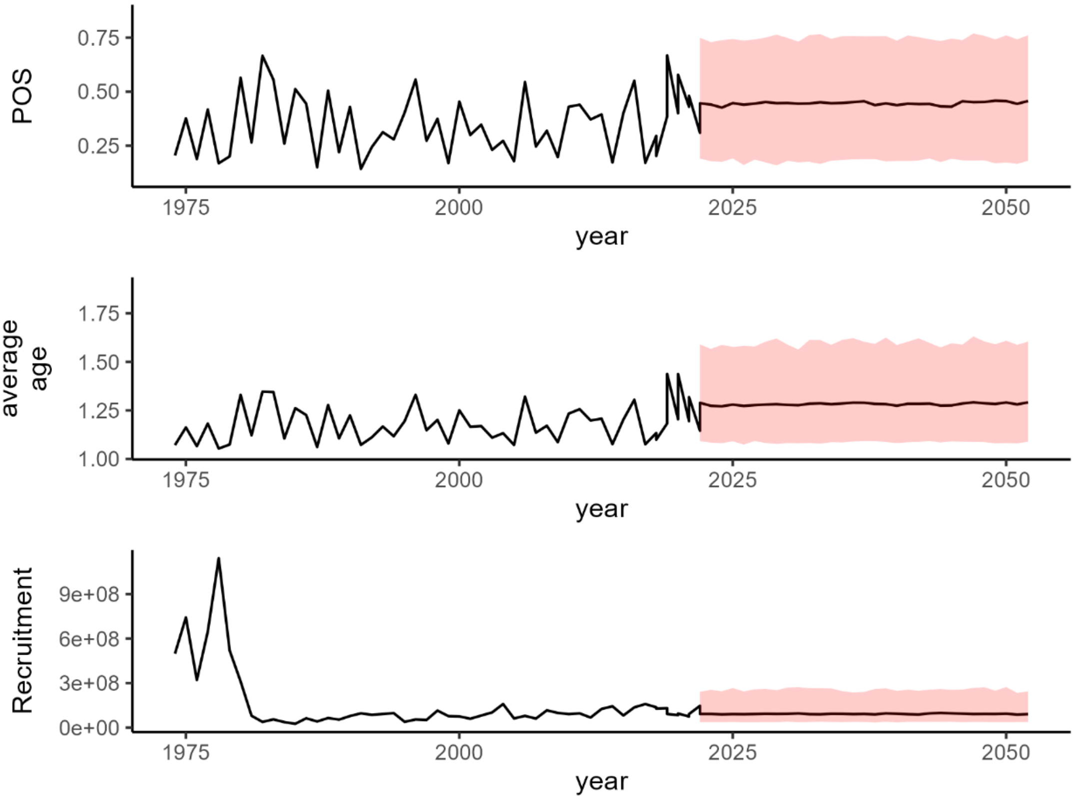


Figure 5.3.5. Three indicators for North Sea sprat, POS (proportion of old spawners), asa (average age), and recruitment. The red shaded area represents the 95th and the 5th percentile from the MSE simulations with Fcap = 0.69.

For sandeel, only the average age was available as a function of $Fcap$. Higher Fcap values led to lower average ages with a zero fishery giving an average age of approximately 1.7, and the maximum Fcap investigated (0.9) gives an asa of ≈1.3 (fig. 5.3.3).

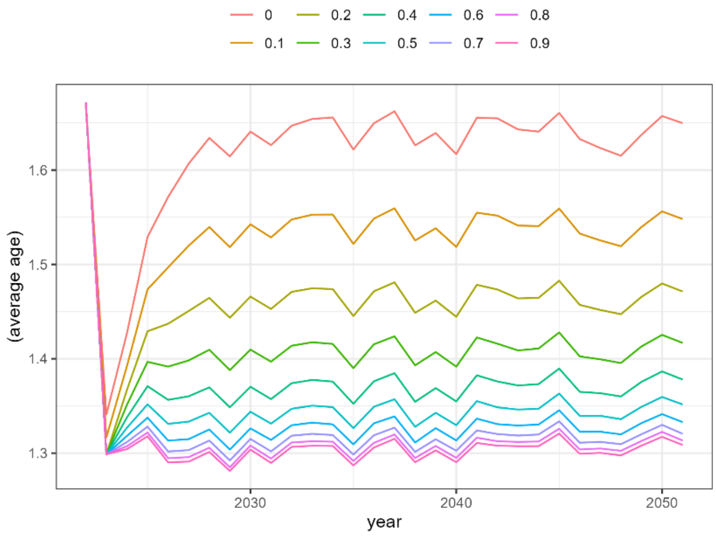


Figure 5.3.6: Simulated ASA (average age) for sandeel in the North Sea area 1r. The line color indicates different Fcap values from the MSE.

Thresholds

In the case for mackerel, the fifth percentile of the historical data only (threshold 1) is lower, while thresholds 2 and 3 are similar (Fig. 3.5.4). In the 9 years from 1991-2000 where the stock is experiencing overfishing, the stock is below the threshold value which supports these scenarios. The stock has remained above the R threshold despite continuous overfishing in the period 2000-2010 (red dots in fig. 5.3.4).

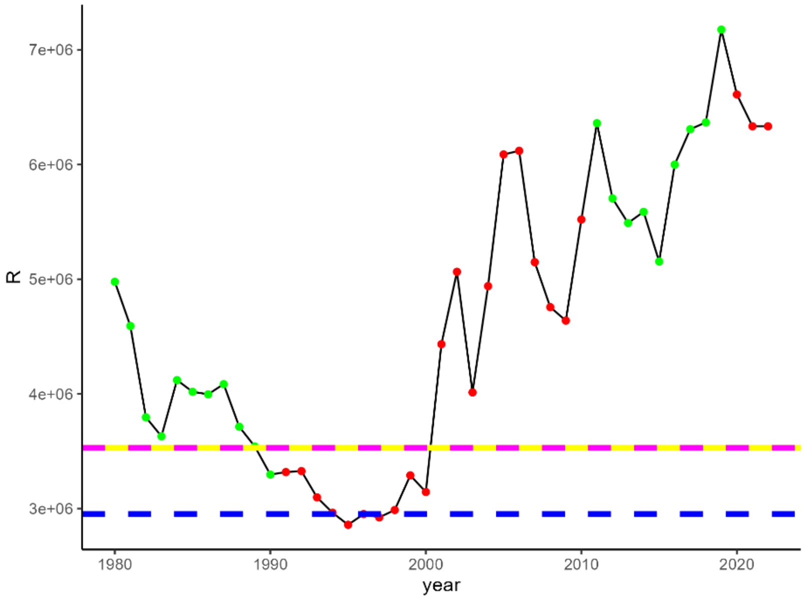


Figure 5.3.7. Mackerel recruitment thresholds with the three different methods listed above. The black line indicates the recruitment time series, with green dots indicating years where F < Fmsy, and red dots F > Fmsy. The blue line indicates threshold (1), the magenta line indicates threshold (2), and the yellow line indicate threshold (3).

The average weight of mackerel was above all the thresholds, despite a prolonged period of being overfished (Figure 5.3.5). Two years are below the historical 5th percentile, one of them has no overfishing that year and the other is occurring after many years of high fishing pressure. Five out of six data points are below the threshold from scenario 2 (historical data without overfishing).

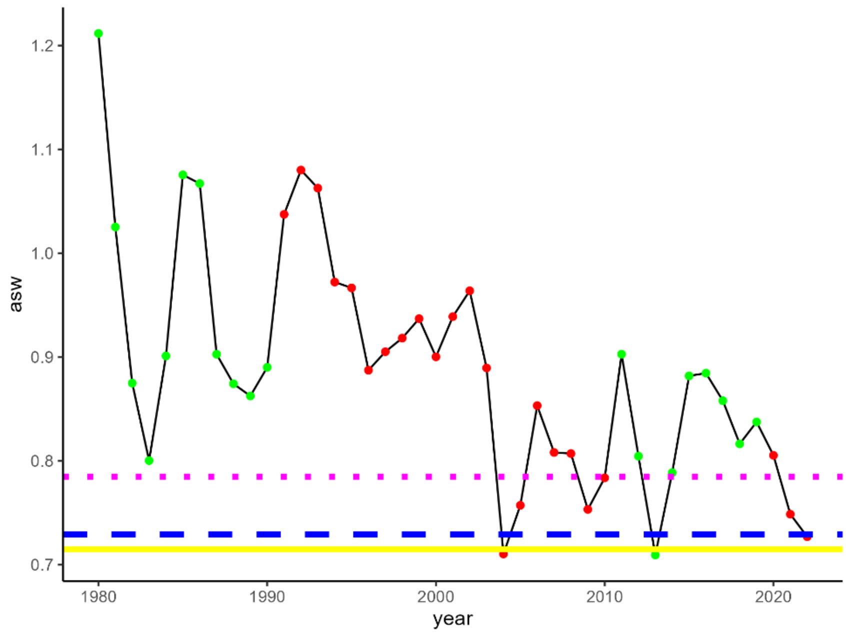


Figure 5.3.8. Mackerel asw thresholds with the three different quantiles listed above.The black line indicates the recruitment time series, with green dots indicating years where F < Fmsy, and red dots F > Fmsy. The blue line indicates threshold (1), the magenta line indicates threshold (2), and the yellow line indicate threshold (3).

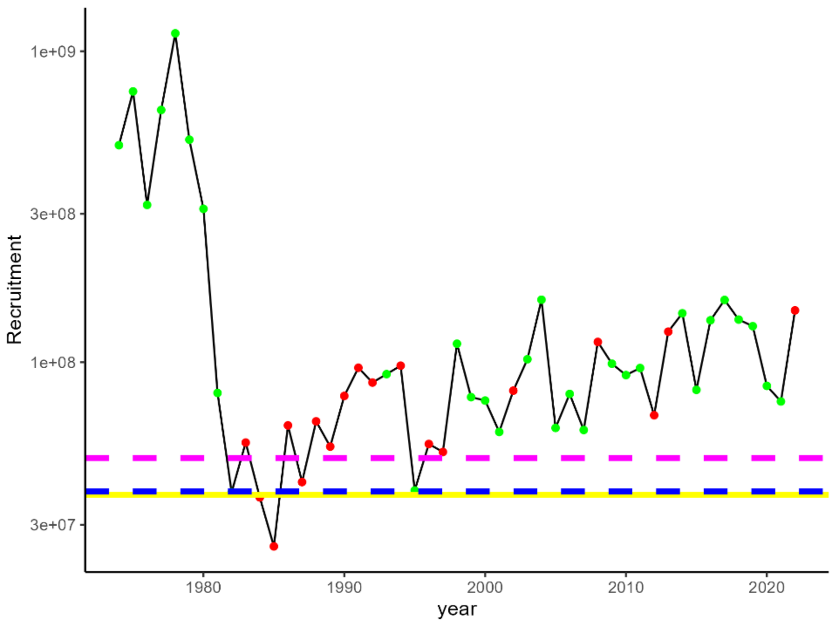


Figure 5.3.9. Thresholds for sprat recruitment. The black line indicates the recruitment time series, with green dots indicating years where F < Fmsy, and red dots F > Fmsy. The blue line indicates threshold (1), the magenta line indicates threshold (2), and the yellow line indicate threshold (3).

For sprat there was little connection between the overfished years and the indicator values. (Figure 5.3.6). Thresholds 1 and 3 were very similar, with just 2 points in the time series going below the threshold. Threshold 2 was higher but still with just 6 points in the timeseries falling below this threshold.

The distribution of the threshold values implemented here did not show a clear response to the fishing regime. Though there was an impact of F on the average of the age-structure indicators, the change was small compared to the interannual variability when fishing according to MSY principles. As a result, the use of age-structure indicators for these three stocks is likely to lead to fre-quent false positives/negatives. The outcome of the MSE simulations depend a lot on the assumptions (in particular how recruitment is modelled) and hence using peer reviewed and stock specific MSEs is key to obtaining reliable results.

5.4 Mediterranean stocks

Length-based simulations were implemented for two hake and one sole population from the Mediterranean Sea, using stock parameters derived from peer-reviewed stock assessments. Constant F scenarios were applied for a period > 2x lifespan. Indicators identified by WKTHRESHOLDS (both length-based, and age-based) were calculated on the simulated exploited population testing different F scenarios. Indicators at F_{MSY} were used as threshold and compared to the historical time series of length indicators calculated on survey data. Additionally, an exploratory run testing a harvest control rule driven by age-based indicator target was completed.

5.4.1 Methods

Simulation method: Length based simulations based on the virtualPop function included in the fishdynr R library (Taylor 2015). The method simulates the population dynamics of a fish stock based on stock parameters (natural mortality, growth, length-weight relationship, maturity), a stock-recruitment relationship, data on fishing selectivity (Lc), and values of fishing mortality, generating size structures at equilibrium. The method works at user defined time increment (from monthly to annual increments). For each time interval of the simulation, the population at sea grows according to biological parameters, then natural mortality is applied and ultimately a user defined level of fishing mortality (F) with associated selectivity. The simulation includes a period of development of the non-exploited population (F=0), ensuring equilibrium is reached before the stock starts to be exploited (F>0). A few modifications have been made to the original function: incorporation of natural mortality vector, use of a non-logistic selectivity (double normal selectivity implemented with the dub logistic R function; Duplisea 2019) and biphasic growth parameters to simulate species that exhibit a strong slowdown in the growth rate after a certain age t (such as Common sole, see Carbonara et al., 2023).

Data sources: data used to parametrize the function were derived from the most recent peer reviewed stock assessments (Table 5.4.1). HKE_8_11 was an a4a model not benchmarked. HKE_17_18 was a ss3 model benchmarked. SOL_17 was a ss3 ensemble model benchmarked, with alternative mortality vector, selectivity and steepness.

Scenarios: For each stock, the population was simulated to equilibrium for 2x lifespan and then constant fishing for a period of at least 2x lifespan was applied. Fishing scenario tested were 0.1 intervals between 0 and 1 as well as the F_{MSY} and the $F_{current}$ provided by the peer reviewed stock assessment. In each simulation run, the length and age composition of the population during the last simulated year was recorded.

Validation: simulated population size structure was visually compared to observed trawl survey data to ensure a coherent pattern under similar levels of exploitation.

Indicators estimated: simulated populations size and age structure under fishing scenarios were resampled with trawl survey selectivity to obtain information comparable with the observed trawl survey data. Length based indicators estimated on the simulated data were L90, L90R and

LBI_{MSY} (a length-based version of the age-based indicator ABI_{MSY}). L90 and L90R are reported as $L90_{MSY}$ and $L90R_{MSY}$, obtained dividing the value estimated at F scenarios by the value observed when $F=F_{MSY}$. Age based indicators estimated on the simulated data were ABI_{MSY} and average age. The distribution of the indicators under the F scenario was shown to demonstrate the responsiveness of the indicator to exploitation.

Thresholds at F_{MSY} *vs observed data*: Indicators derived from populations exploited at F_{MSY} were used as a threshold and compared with historical data. All the estimations were done on the population at sea as well as on the population sampled with survey selectivity.

Population structure-based harvest control rule: the age based ABI_{MSY} indicator was used as base to anchor a harvest control rule to a desired population structure. The experiment was done on the simulated population of Hake from the Adriatic Sea. For the calculation of the original ABI_{MSY}, the 90th percentile of the fish numbers-at-age distribution at F_{MSY}, rounding to the nearest integer, is used as a reference age (A_{MSY}). The percentage of the numbers-at-age distribution above A_{MSY}, in a population exploited at F_{MSY} is defined P_{MSY}. For a given numbers-at-age distribution, the indicator ABI_{MSY} is defined as P/P_{REF}, with P representing the observed percentage of the numbers-at-age distribution above A_{MSY}, and $P_{REF} = P_{MSY}$. We were interested in evaluating the responsiveness of the exploitation and of the biomass when testing alternative values for P_{REF}. We defined this configuration of ABI_{MSY} as ABI_{REF}, where the value used as P_{REF} was defined by the user. P_{REF} values tested were 0.01; 0.03; 0.05; 0.09; 0.11. P_{MSY} (0.097) was also tested as P_{REF}. ABI_{REF} was calculated each year, and the harvest control rule imposed was:

$$\begin{cases} F_{y-1} * 1.02 & if\ ABI_{y-1} > 1.1 \\ F_{y-1} & if\ 0.9 < ABI_{y-1} < 1.1 \\ F_{y-1} * 0.95 & if\ ABI_{y-1} < 0.9 \end{cases}$$

Results are shown in terms of SSB/SSB_{B0}, F and L90.

Table 5.4.1 stocks used for length-based simulations and stock assessment used as reference.

Stock identifier	Species	Area	Stock assessment reference
HKE_8_11	European hake (*Merluccius merluccius*)	GSAs 8, 9, 10, 11.	STECF 2023
HKE_17_18	European hake (*Merluccius merluccius*)	GSAs 17, 18	FAO 2019; STECF 2022
SOL_17	Common sole (*Solea solea*)	GSA 17	FAO 2021; STECF 2022

5.4.2 Results

Validation:

- HKE_8_11 (Figure 5.4.10): the simulated population was dominated by individuals under 25 cm and the number of individuals above 75cm was very low. The simulated population sampled with the survey shows a peak at around 10-15 cm in accordance with the observed data. However, the length distribution of the simulated population was less concentrated around the catchability peak. As a result, the simulated population consisted of many individuals between 20 and 40 cm. The maximum length observed was also higher in the simulated population. These results suggest that the simulated population generally resembled the observed survey sample but overestimated the number of large individuals at current F.

- HKE_17_18 (Figure 5.4.11): the simulated population was dominated by individuals under 25 cm and the number of individuals above 75cm was very low. The simulated population sampled with the survey net shows a peak at around 10-15 cm in accordance with the observed data. However, the length distribution of the simulated population was less concentrated around the catchability peak. As a result, the simulated population had a higher proportion of individuals between 20 and 40 cm than the observed survey data. The maximum length observed was also higher in the simulated population. These results suggest that the simulated population generally resembled the observed survey sample but overestimated the number of large individuals at current F.

- SOL_17 (Figure 5.4.12): the simulated population showed a bimodal distribution, with a large peak around 15 cm and a second peak around 25 cm. The simulated population in the survey showed the same shape, except that the two modes have a similar density, in accordance with the observed data. The maximum length observed was similar between the simulated and the observed population. These results suggest that the simulated population was in good agreement with the observed survey sample.

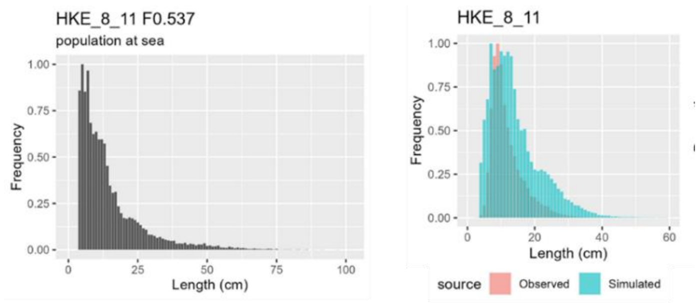


Figure 5.4.10: simulated population of HKE_8_11 under Fcur. Population at sea (left); population sampled with survey selectivity compared with observed survey data; proportion of individuals by length bin in the simulated population versus the observed survey data (right).

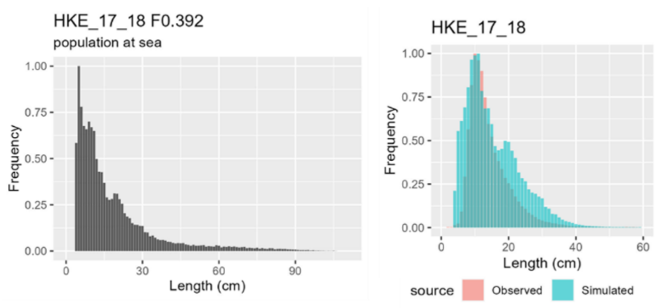


Figure 5.4.11: simulated population of HKE_17_18 under Fcur. P Population at sea (left); population sampled with survey selectivity compared with observed survey data; proportion of individuals by length bin in the simulated population versus the observed survey data (right).

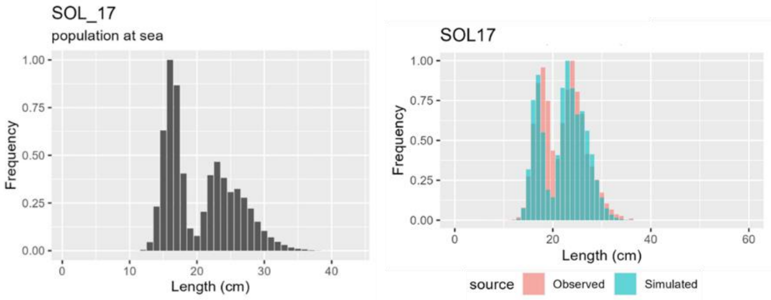


Figure 5.4.12: simulated population of SOL_17 under Fcur. Population at sea (left); population sampled with survey selectivity compared with observed survey data; proportion of individuals by length bin in the simulated population versus the observed survey data (right).

Indicators estimated: for HKE_8_11 (Figure 5.4.13) and HKE_17_18 (Figure 5.4.14) the pattern of length-based indicators was coherent with age-based indicators, showing an exponential decline with increasing fishing pressure. L90R was an exception, showing an almost linear decline. Nevertheless, the range of variation of length-based indicators was lower than age-based indicators, especially when calculated from the simulated population data resampled with survey selectivity. The case of SOL_17 (Figure 5.4.15 and Figure 5.4.16) permitted to evaluate the sensitivity of the indicators to alternative assumptions in the stock assessment, such as alternative fishery selectivity and natural mortality vectors. In SOL_17 length-based indicators, when calculated on the simulated population data resampled with survey selectivity shown low variability. The low variability and a "jagged" pattern in the length-based indicators were likely to be caused by the characteristics of the stock (low growth at large size: see Carbonara et al., 2023) and by the nature of the data (integers). The effect of alternative assumptions in the stock assessment was more evident on length-based indicators than on age-based indicators. High natural mortality vector narrowed down the variability of all the indicators and especially of those length based. Alternative selectivity effect was less evident. These results suggest that length-based indicators can give information that is in general agreement with age-based indicators, however the former are less sensitive to fishing mortality than the latter. Alternative stock assessment configurations can influence the indicator value at F_{MSY} irrespective of the indicator chosen. This evidence highlights how simulations exercises aimed to identify threshold values for population size or age structure at F_{target} will benefit from exploration of stock assessment uncertainty.

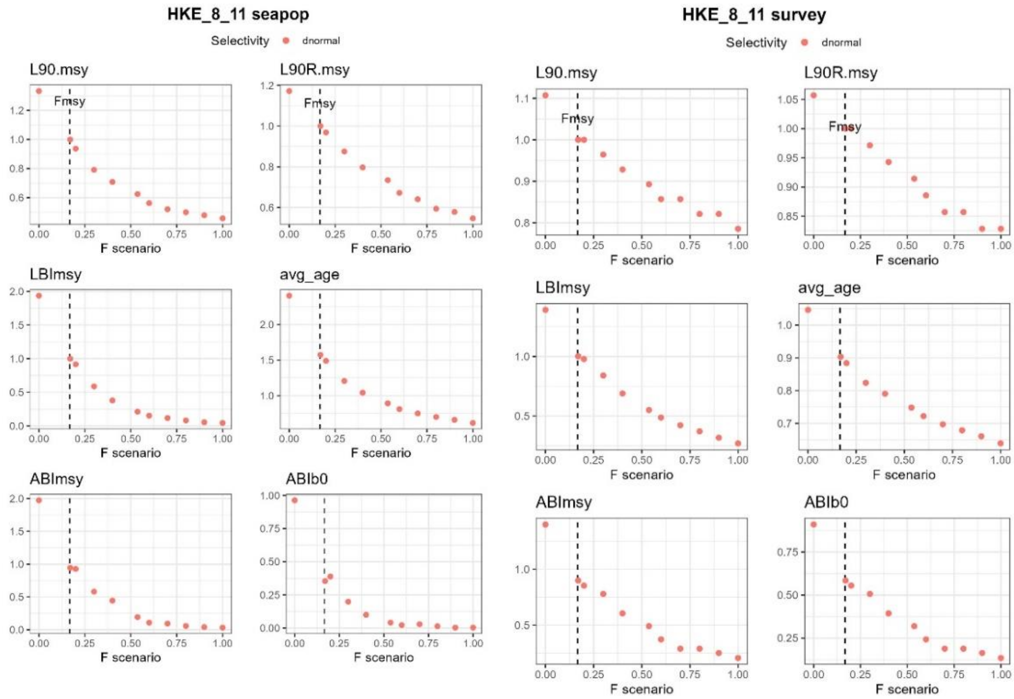


Figure 5.4.13: length based (L90.msy, L90R.msy, LBImsy) and age based (avg_age, ABImsy and ABIb0) indicators calculated on length-based simulations for HKE_8_11 at F scenarios. Left panel show the indicators calculated on the population at sea, right panel show the indicators calculated on the population at sea sampled with survey net selectivity. Dotted vertical line represent the F at msy estimated from the most updated stock assessment.

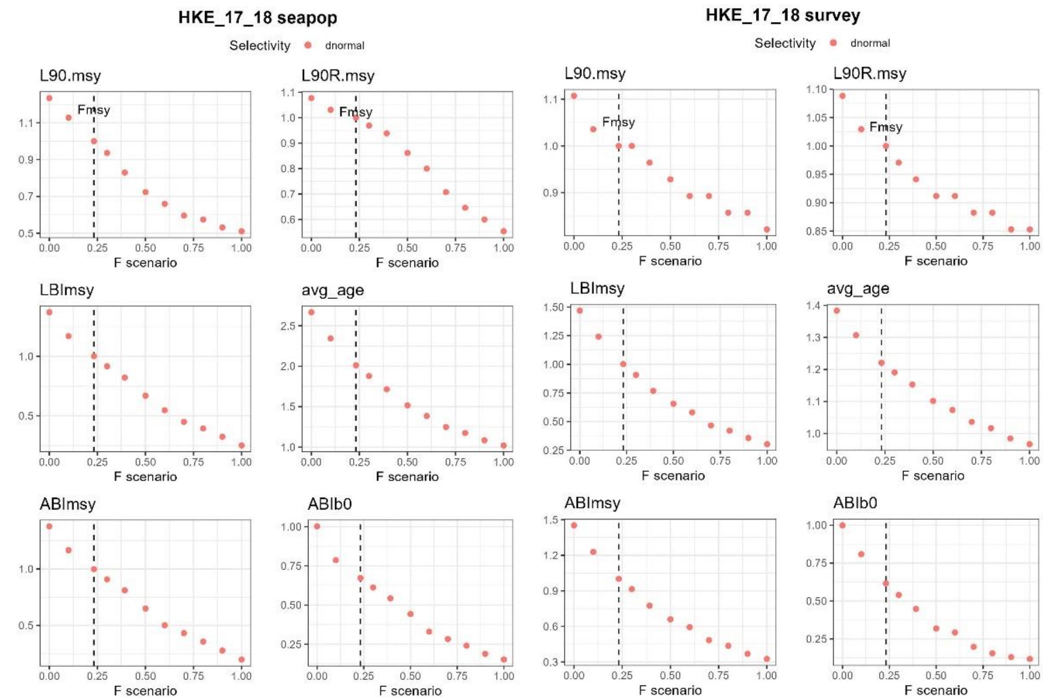


Figure 5.4.14: length based (L90.msy, L90R.msy, LBImsy) and age based (avg_age, ABImsy and ABIb0) indicators calculated on length-based simulations for HKE_17_18 at F scenarios. Left panel show the indicators calculated on the population at sea, right panel show the indicators calculated on the population at sea sampled with survey net selectivity. Dotted vertical line represent the F at msy estimated from the most updated stock assessment.

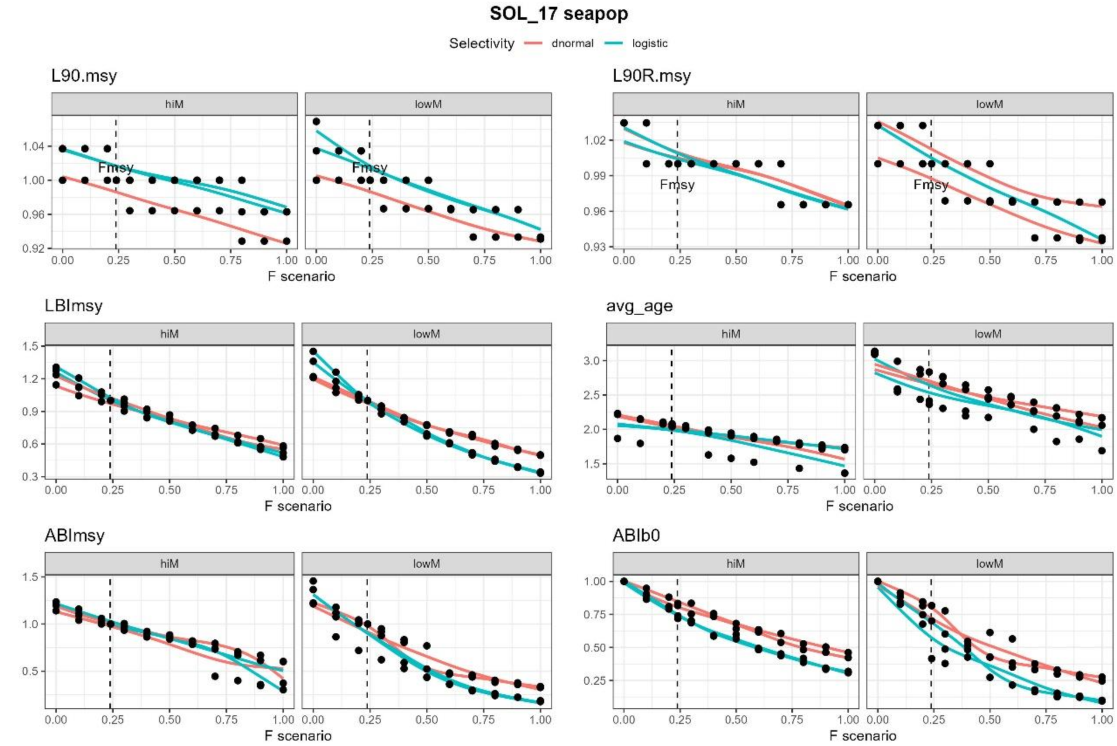


Figure 5.4.15: length based (L90.msy, L90R.msy, LBImsy) and age based (avg_age, ABImsy and ABIb0) indicators calculated on length-based simulations for SOL_17 at F scenarios calculated on the population at sea. Dotted vertical line represent the F at msy estimated from the most updated stock assessment. Panels show the alternative natural mortality vectors used (high mortality hiM and low mortality lowM), solid lines show the trend when using alternative fishery selectivity (doublenormal in pink and logistic in blue).

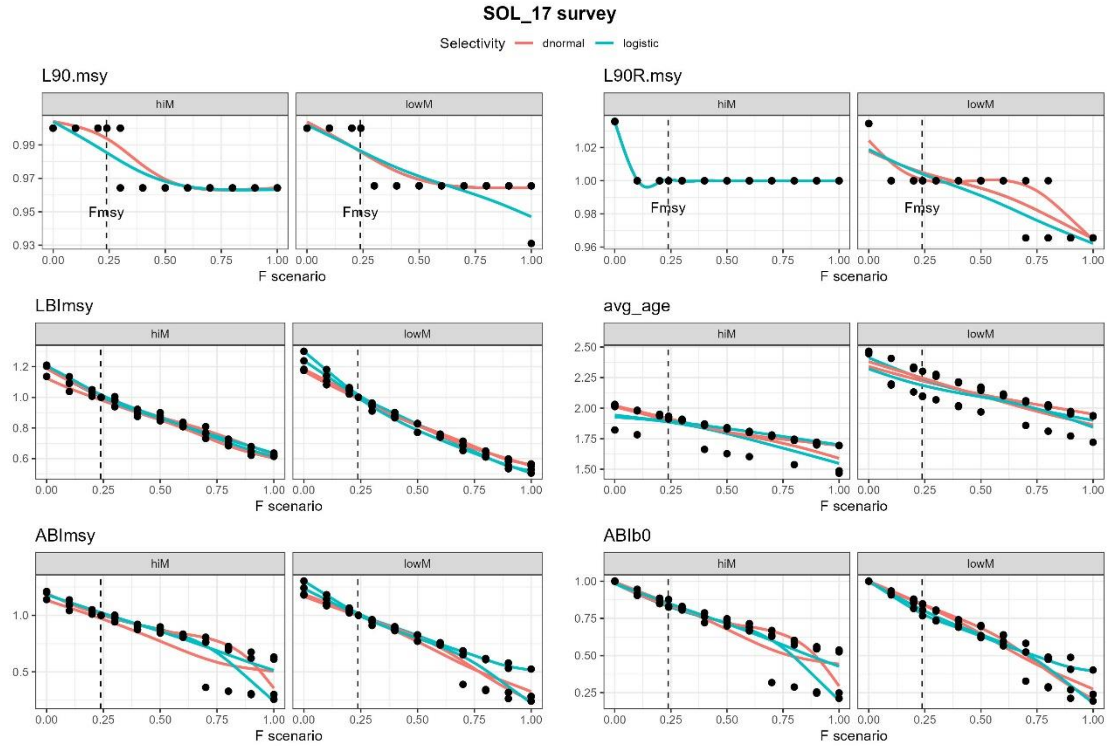


Figure 5.4.16 length based (L90.msy, L90R.msy, LBImsy) and age based (avg_age, ABImsy and ABIb0) indicators calculated on length-based simulations for SOL_17 at F scenarios calculated on the population at sea sampled with survey selectivity. Dotted vertical line represent the F at msy estimated from the most updated stock assessment. Panels show the alternative natural mortality vectors used (high mortality hiM and low mortality lowM), solid lines show the trend when using alternative fishery selectivity (doublenormal in pink and logistic in blue).

Thresholds at F_{MSY} *vs observed data*: Indicator values at F_{MSY} were used as a threshold for the observed indicators calculated on trawl survey data (Figure 5.4.17). The stocks considered were all overexploited for the entire period when trawl survey data was available. However, the magnitude and the pattern of the fishing mortality varied among the stocks.

- HKE_8_11: exploitation has been steady for most of the time series and well above the reference point, with a slight decline observed only in recent years that stopped at values around 5 times the sustainable fishing level. All the tested indicators were well below the threshold. Age based indicators were suggesting a slight recovery, while length-based indicators were almost flat. The results were in line with expectations: the stock has always been in high overexploitation and the slight stock improvement due to the decline in F was so low that it was detected only by the age-based indicators (more sensitive to exploitation level: see section).
- HKE_17_18 the exploitation also has been steady and well above the reference point for most of the time series, however the decline was harsh and in recent years F was around 2 times the reference point. It was not possible to calculate an age-based indicator for this stock. Indicators were generally below the threshold but suggesting a clear positive trend. In this case the decline in F was strong enough to enable length based to detect a recovery in the population structure.
- SOL_17 the exploitation spiked in the first part of the time series and then dropped to values close to the sustainable exploitation in recent years. Length based indicators excluding recruitment (LBI_{MSY} and L90R) were oscillating around the sustainability

threshold. The age-based indicator and L90 both reflected a deterioration in the population structure due to the harsh exploitation peak and suggested a recovery after the exploitation decline.

These results show the behavior of the length based and of the age-based indicators calculated on observed data. Length based indicators are less sensitive than age-based indicators to stock recovery, with the case of HKE_8_11 not showing any trend. Length based indicators can be sensible to the inclusion of the recruitment depending on stock characteristics: L90R was performing better than L90 in HKE_17_18, while the opposite was true in the case of SOL_17. In overall, it is also highlighted how the indicator status can be well below from its value at Fmsy.

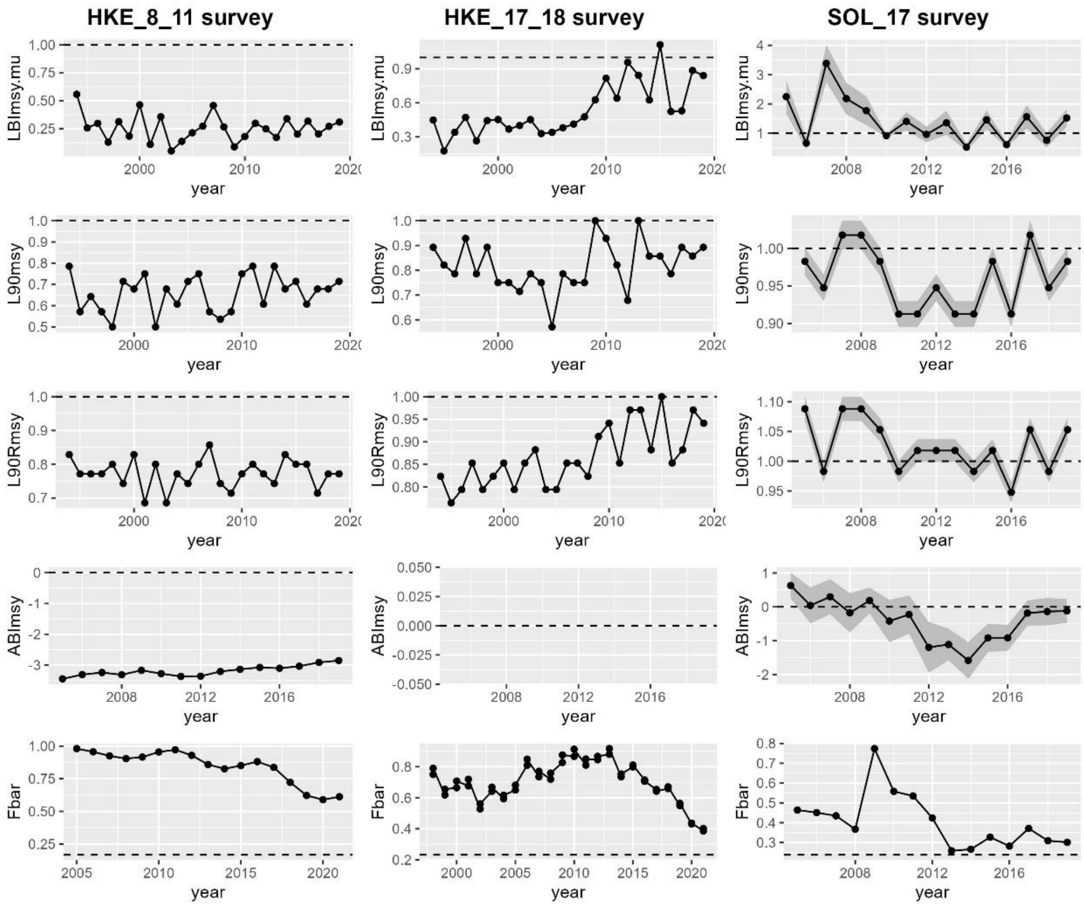


Figure 5.4.17: time series of length based (LBImsy.mu, L90.msy, L90R.msy) and age based (ABI_{MSY}) indicators as well as fishing mortality trend (Fbar) for HKE_8_11, HKE_17_18 and SOL_17. Solid line indicates the reference value at F_{MSY}.

5.4.3 Population structure-based harvest control rule

Population structure-based harvest control rule results of the HCR exercise shown the use of population structure as driver for a harvest control rule. Using a P_{REF} similar (or above) to P_{MSY} allowed the spawning stock biomass to remain well above 0.4 SSB_{B0} (Figure 5.4.18). The improvement in SSB provided by $P_{REF} > P_{MSY}$ was also reflected in the length composition of the population (Figure 5.4.20). The F pattern, according to the characteristics of the HCR, was more oscillatory as the P_{REF} decreased (Figure 5.4.19). It must be acknowledged that these results are exclusively to be intended for exploratory purposes. The simulation model used did not account for adequate process uncertainty (especially on recruitment) neither for model uncertainty (no iterations tested) or structural uncertainty (alternative configurations of the operating model). It is suggested to check the feasibility of implementing this exercise in MSE simulations to verify how different age-based indicators and associated target thresholds perform. The resulting information can be used to better understand how indicator values associate with stock characteristics, and how to choose a threshold that can ensure the stock to be in GES.

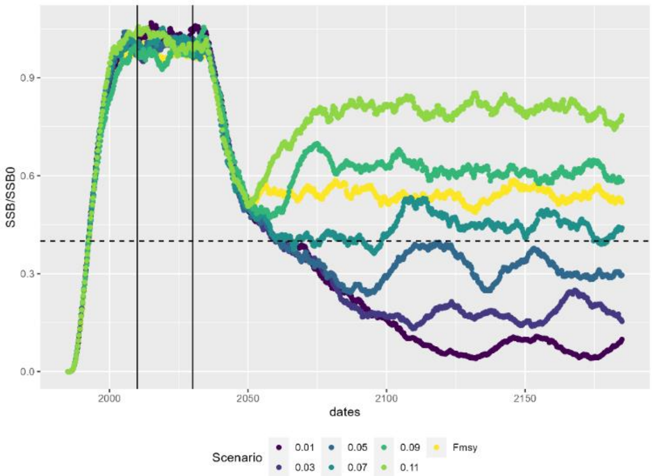


Figure 5.4.18: SSB/SSB B0 trend from simulations on HKE_8_11 with age-based HCR at different PREF. Colors indicate used PREF, dotted line corresponds to 40% SSBB0.

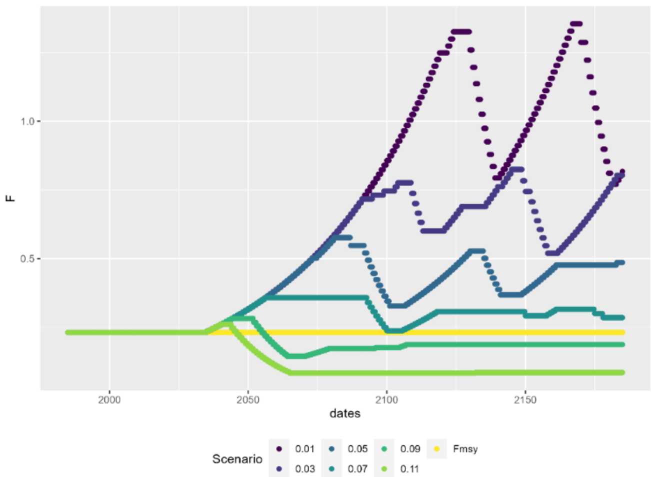


Figure 5.4.19: F trend from simulations on HKE_8_11 with age-based HCR at different $P_{REF.}$ Colors indicate used P_{REF}.

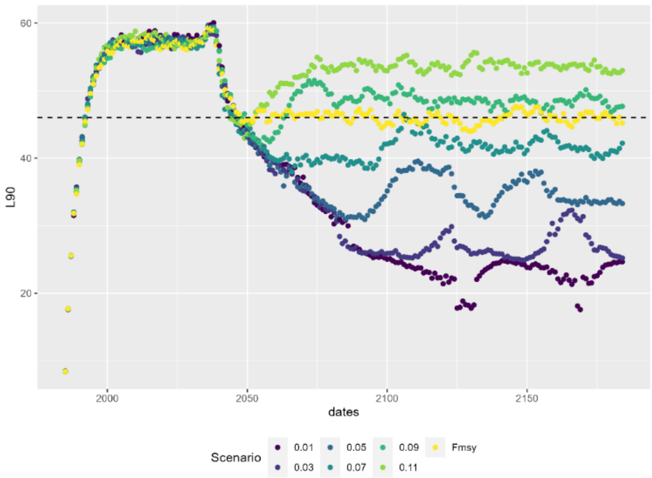


Figure 5.4.20: L90 trend from simulations on HKE_8_11 with age-based HCR at different $P_{REF.}$ Colors indicate used P_{REF}, dotted line corresponds to $L90_{MSY}$.

6 Evaluation of potential thresholds

This section describes the result of the evaluation of the indicator approaches for the 7 demersal and 2 pelagic stocks, where the distribution of the indicators when fishing according to MSY principles were available. Two potential solutions for stocks where an MSE is not available were evaluated by comparing the resulting threshold with that based on 5th percentiles of the indicators at FMSY (section 6.10).

It should be remembered that the observations are based on 7-9 stocks, and further studies should be conducted to investigate the generalisations on thresholds across life history types.

6.1 Estimate the indicator level at which there is a clear decrease productivity at low indicator levels (similar principle to Blim)

Recruitment deteriorated substantially when fishing at levels above F_{P05} for the seven demersal stocks. The age structure indicators all declined rapidly with increasing fishing mortality up to the level at which recruitment was impaired. Unfortunately, for the age structure indicators, there was no evidence of a relationship between the indicators and productivity and there were also no clear breakpoints in the relationship with fishing mortality. Increasing fishing beyond the levels that reduced recruitment led to little further change in the age structure indicators. As a consequence, it was considered that there was insufficient evidence for indicators other than recruitment to suggest a clear decrease in productivity level at a specific value of the indicator. For recruitment, the decline was linked to SSB below Blim, an aspect there is already in focus when setting management strategies and measures for D3C1 and D3C2.

6.2 5th percentiles of the indicator from MSE simulations of fishing according to the MSY approach

This method was the preferred approach among those investigated. It generally (but not always) produced a level that was within the historic range of the indicators. Surprisingly, the threshold generally resulted in the historic values of the indicator fluctuating at or below the threshold rather than above in periods with sustainable (near F_{MSY}) fishing (see saithe and the two plaice stocks). The cause of this could be a lag in the response of age structure indicators to changes in fishing mortality. This issue may need to be addressed in the future. As this threshold method was the preferred method, all other thresholds were compared to this in the evaluation for the stocks where an MSE was available.

For the two pelagic stocks where the distribution of the indicators was available, the variation in age structure indicators when fishing according to the MSY approach encompassed virtually all of the historical time series, possibly due to the larger variability in productivity of these stocks than the other stocks examined. Hence, the age structure indicators are likely to result in very frequent classification of the stock as being above the threshold regardless of fishing pressure. The group felt that information from two stocks was insufficient to deduce if this is a general pattern for pelagic stocks and recommend further studies to be performed to investigate this.

6.3 Estimate 5 percentiles of the indicator using data from a minimum of 3 generations years with near sustainable exploitation

This method produced levels that were substantially lower than those produced by the method using MSY management simulations. The historical timeseries generally exceeded the threshold in the majority of the years, including years with unsustainable fishing. The method was therefore not considered suitable for the estimation of thresholds.

6.4 Estimate the trend of the indicator using historic data

The investigation of historical data revealed that the vast majority of the species have been and often still are fished at levels substantially above F_{MSY}. At these levels of fishing mortality, all simulations confirmed that the interannual variation at a given fishing level greatly exceeded the variation between different (high) levels of fishing. As a result, the trend is not expected to reflect actual change in the population until fishing is reduced to below levels impairing recruitment. Hence, this method is not recommended for stocks fished above the smaller value of $FP_{0.5}$ or $1.5*F_{MSY}$. Its usefulness for other stocks remains to be determined.

6.5 5th percentiles of the indicator from MSE simulations of no fishing

The comparison of indicator values at no fishing and at F_{MSY} showed a clear change in the age structure indicators as a result of fishing, consistent with a decrease in survival of the individuals and a decline in SSB. In contrast, there was little change in recruitment though smaller values did occur at no fishing if the stock recruitment relationship included a declining in survival at high densities (like a Ricker curve, see tur.4). It was discussed that the values are associated with greater uncertainty than those at high levels of fishing mortality as there is no data for commercial stocks for prolonged fishing closures at high stock size. As the potential density dependent effects on growth, mortality and maturity at high stock size are not know, the predictions were considered more uncertain. The age structure indicators declined monotonically with fishing pressure with no clear change in the relationship until fishing reached the level impairing recruitment. The no fishing scenario was considered interesting but not in accordance with the aim of other D3 criteria of exploiting the stocks.

6.6 5th and 95th percentiles of the indicator from MSE simulations of fishing at $F_{P0.5}$

In this analysis, we evaluated the indicators variability at the MSY approach and related that to the indicators' value (and variance) calculated at the fishing level considered as the precautionary limit, FP0.5 (Fig. 6.6.1). The MSE simulations of the FLBEIA framework for 6 stocks from the North Sea and English Channel (cod, saithe, plaice, sole, turbot; see section 5.2) were used. The 5th percentiles of the indicators when fishing according to the MSY approach and the 95th percentiles of those same indicators when fishing at $F_{P0.5}$ were extracted. $F_{P0.5}$ was considered the highest fishing mortality that was consistent with a low risk of impairing recruitment.

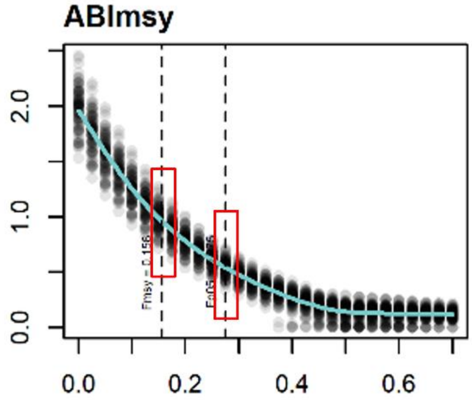


Figure 6.6.1. Example of an indicator (ABImsy) simulations at different fishing levels. Red boxes highlight contain the 100 simulations of the indicator at F_{MSY} and $F_{P0.5}$ selected for the analysis.

Figure 6.6.2 shows the smoothed distribution of the indicators for each species at F_{MSY} and $F_{P0.5}$, with the associated 5th and 95th percentile, respectively. In general, the distributions tend to be quite symmetric, especially at F_{MSY} while at $F_{P0.5}$ there is some tendency for bimodality or the emergence of additional peaks which is likely linked to how multiple S-R have been implemented in the simulations. Distributions of R and SSB/R overlap extensively between the two fishing scenarios examined, which would be problematic for the use of the indicator. The indicators ABImsy and POS are well separated based on the 5th and 95th percentile thresholds in 3 out of 6 stocks (COD-NS, PLE-EC, TUR) and avgAge in 2 out of 6 (PLE-EC, TUR). The difference between $F_{P0.5}$ and F_{MSY} varied substantially between stocks. However, also when the distributions exceed those thresholds the overlap appears in most cases limited to a visual inspection (except for POS in SOL-NS), suggesting that a less stringent but still meaningful threshold for separation could be satisfactory.

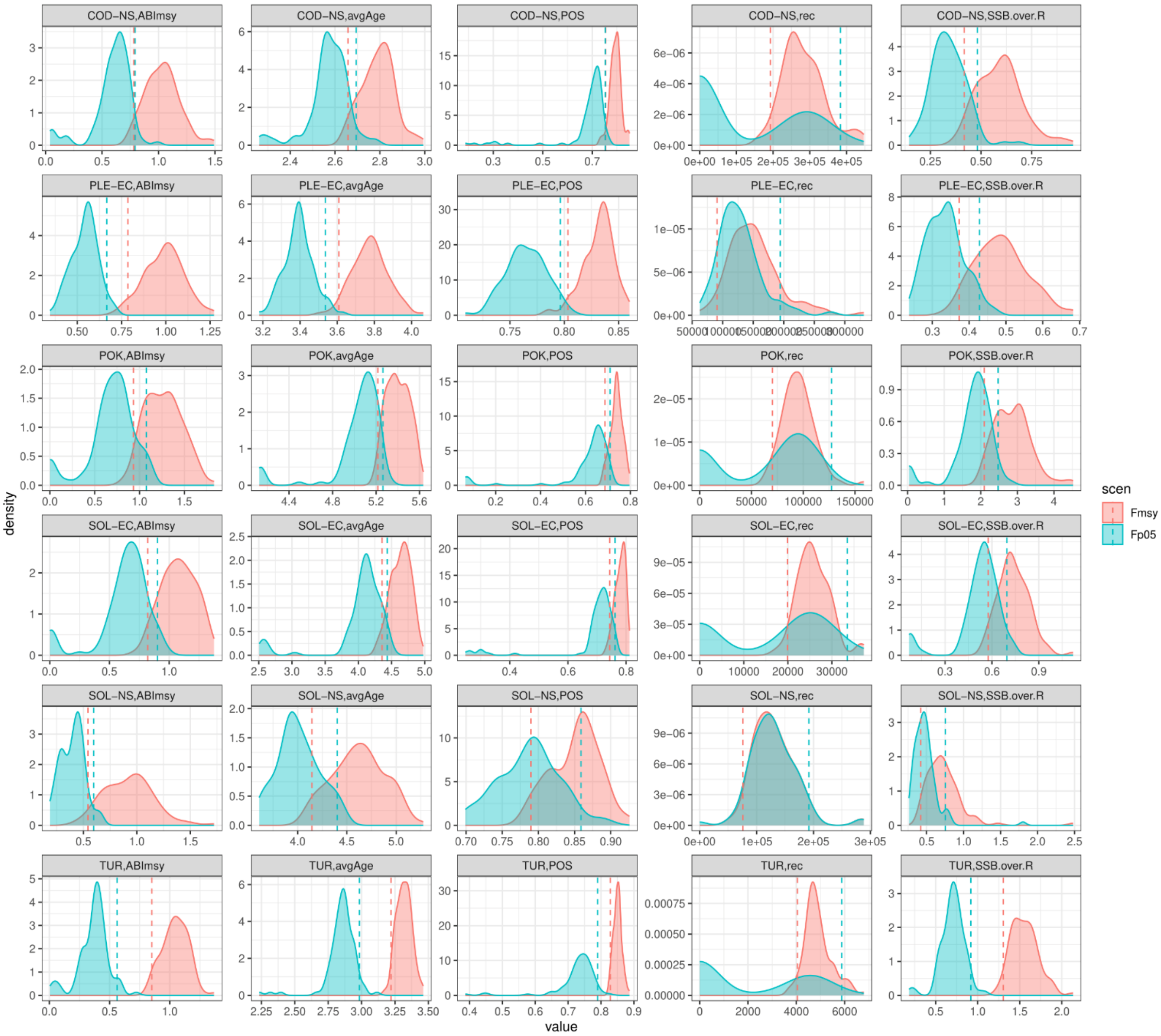


Figure 6.6.2 Distribution of the indicators ABImsy, avgAge, POS, R, SSB/R at F_{MSY} and $F_{P0.5}$, with the associated 5th and 95th percentile (dotted vertical lines), calculated for the stocks of COD-NS, PLE-EC, POK, SOL-EC, SOL-NS, TUR.

The percentage of times the indicator at F_{MSY} would fall within the 95th percentile of the indicator distribution at $F_{P0.5}$ was then estimated (figure 6.6.3). A low percentage was considered good (good distinction between fishing levels) and gives a measure of how many times given the natural variability of the indicator at F target, this would lay below a limit undesirable level. It is noted that the 95th on the $F_{P0.5}$ is a very low risk level (e.g., it would be like to evaluate the stock status in relation to limit reference points by accounting for uncertainty in both SSB and Blim estimates, while in practice Blim is a deterministic threshold). For this reason, these results are tabulated for both the 95th and 50th percentile of the indicator distribution at $F_{P0.5}$ (table 6.6.1).

ABImsy followed by avgAge are the two indicators with the lowest percentage of runs at F_{MSY} falling within the $F_{P0.5}$ distribution. The average values were 7.2% and 9.2%, respectively, although for both the indicators the percentages exceeded 10% for saithe and English Channel sole, and for avgAge also for North Sea sole. SSB/R and R showed the highest percentages with averages of 25.7% and 93.5%, respectively. In 15.5% of the runs, POS had values at F_{MSY} within the 95th percentile of $F_{P0.5}$. When compared to the 50th percentile at $F_{P0.5}$, ABImsy and avgAge fall below the limit in none of the runs, while this still occurs for 6% of the runs of North Sea sole for POS and SSB/R and in 1% of the runs of English Channel Sole for SSB/R. The percentages remain high for R.

Table 6.6.2 Percentage of runs at F_{MSY} where the indicators fall within the 95th percentile (and 50th percentile in brackets) of the indicator distribution at $F_{P0.5}$.

stock	ABImsy	avgAge	POS	rec	SSB/R
COD-NS	5 (0)	9 (0)	4 (0)	93 (4)	17 (0)
PLE-EC	0 (0)	1 (0)	2 (0)	88 (26)	20 (0)
POK	20 (0)	11 (0)	15 (0)	97 (32)	23 (0)
SOL-EC	12 (0)	10 (0)	20 (0)	96 (19)	32 (1)
SOL-NS	6 (0)	24 (0)	52 (6)	95 (49)	62 (6)
TUR	0 (0)	0 (0)	0 (0)	92 (0)	0 (0)

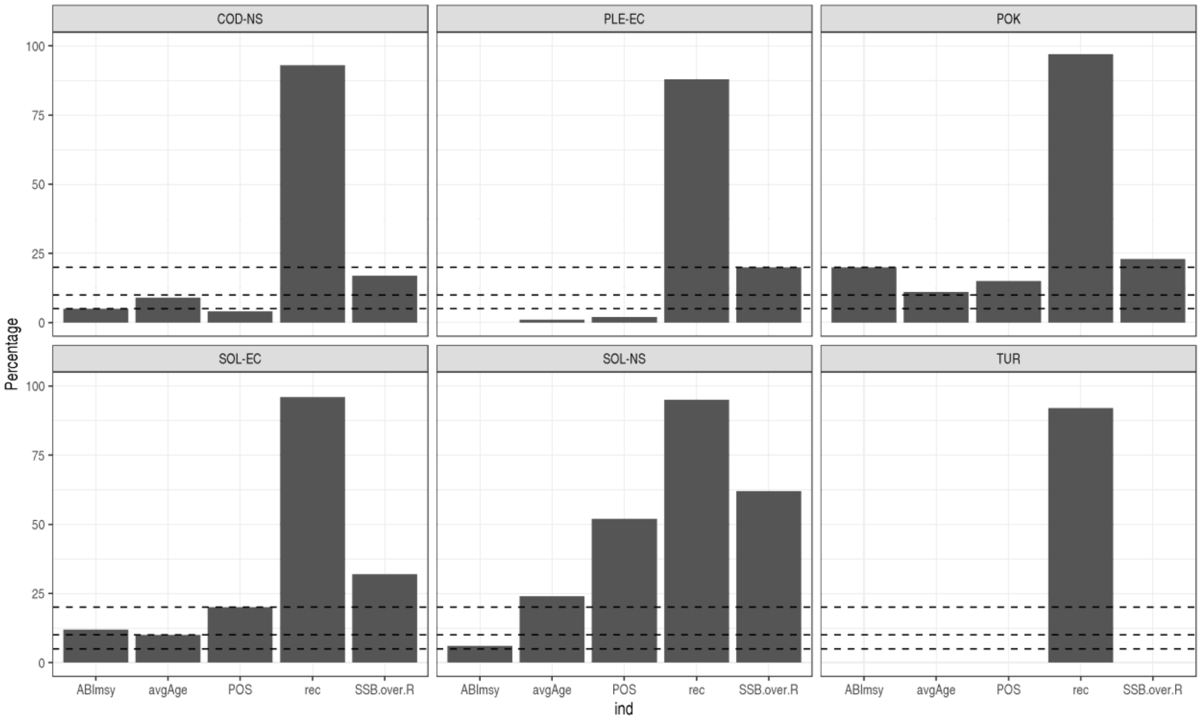


Figure 6.6.3 ***Number of runs*** **(percentage)** ***falling outside the*** **5th percentile of $F_{P0.5}$ threshold calculated for each indicator (ABImsy, avgAge, POS, R, SSB/R) and stock (COD-NS, PLE-EC, POK, SOL-EC, SOL-NS, TUR). Horizontal dotted lines are drawn at 5%, 10%, 20%.**

Indicators were also analyzed by performing a Principal Component Analysis (PCA) considering the distributions obtained from simulations for F_0, $F_{0.5}$, and F_{MSY}. The indicator for recruitment (rec) was excluded from the analysis. The purpose was to investigate and compare these distributions within and between stocks more comprehensively. As expected, all indicators were highly correlated, particularly avgAge and SSB/R, with the first principal component (CS1) carrying more than 70% of the total information. POS was the only indicator with a high coefficient for the second component (CS2). Minimum convex polygons were built considering the three distributions of F better to visualize the overlapping degree between the three distributions. This was done both considering all the stocks together and separately. Considering all the stocks together, the visualization highlighted a high level of overlapping between $F_{0.5}$ and F_{MSY} and, as expected, a very low level with F_0. However, the scenario changed when considering the stocks one by one. Indeed, six up to eight of the considered stocks showed a low level of overlap between the two distributions, with the F_0 distribution entirely separated from the two in all cases.

In particular, the PLE-NS stock was characterized by an almost complete overlap between $F_{0.5}$ and F_{MSY} distributions.

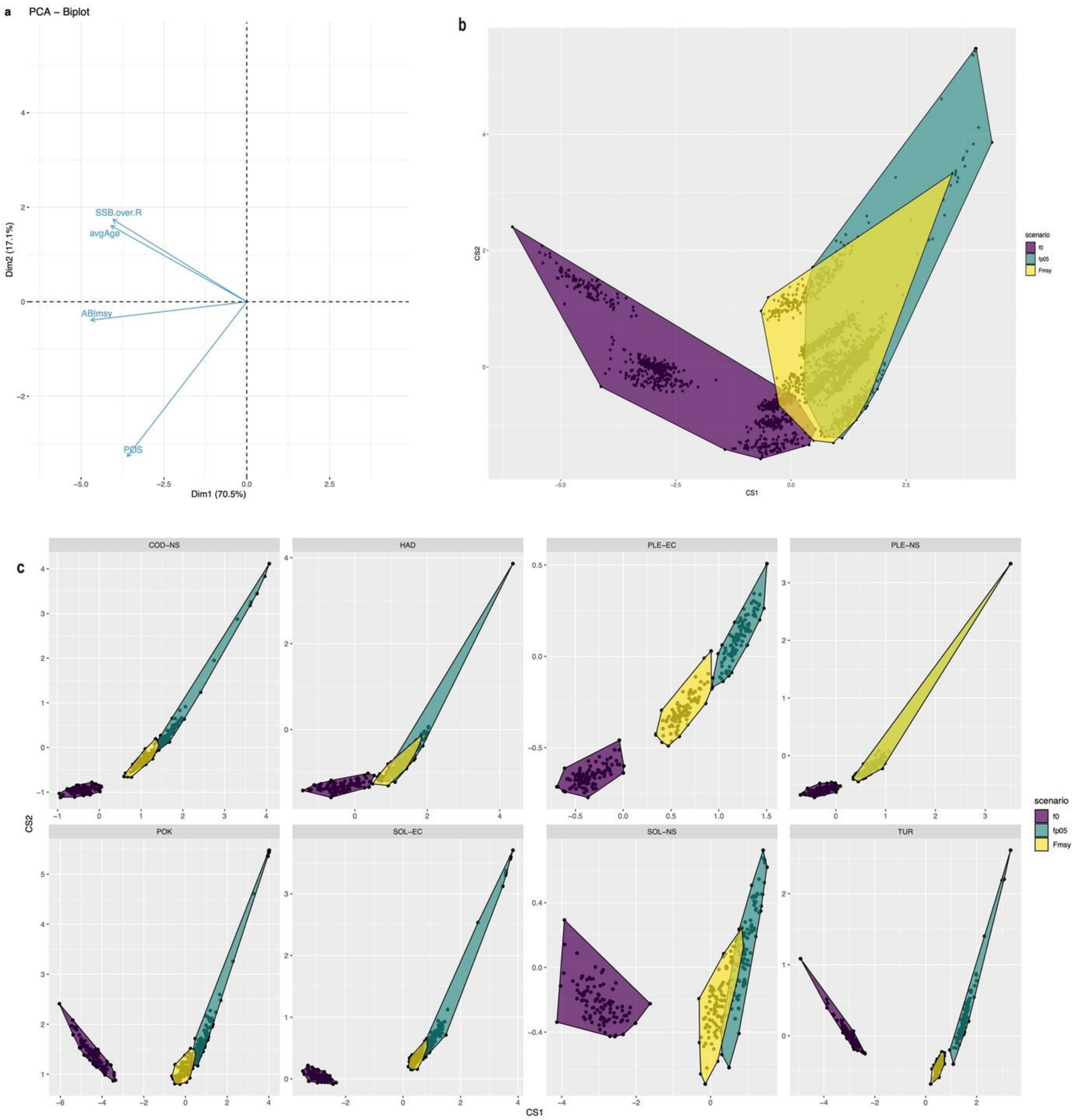


Figure 6.6.4 Bi-plot of the Principal Component Analysis. The original variables are represented with light blue vectors (a). Minimum convex polygons calculated considering all the stocks together (b) and separately (c).

.

6.7 95th percentile of the indicator using data periods of unsustainable exploitation available

This proposed threshold when estimated seemed unconnected to thresholds set from simulations at MSY management, though the values were always lower, indicating that this method would substantially underestimate the threshold. In unsustainably fished stocks, the advice should be to reduce fishing. Once this is accomplished, the potential thresholds can be reevaluated.

6.8 Median or average threshold across MSEs investigated in WKSIMULD3

The group discussed whether a generic threshold was likely to be deductible from all stocks. This would be equivalent to suggesting a common F_{MSY} for all stocks. However, the general perception was that differences in thresholds will occur between stocks of different productivity, mortality, shape of the stock-recruitment relationship and selectivity pattern. These differences reflect biological differences that should not be ignored in monitoring or in management advice.

6.9 95th percentile of the indicator from MSE simulations of fishing at high F

In this analysis, we evaluated the indicators variability at the target fishing level, here corresponding to the MSY approach for all the stocks considered, and then related that to the indicators' value (and variance) calculated at fishing level considered well beyond the limit. This work can be useful to identify a critical threshold for the indicators, indicating a clear non-GES condition for the stocks. For the stocks from the North Sea and English Channel (cod, saithe, plaice, sole, turbot; see section 5.2), the 5th percentiles of the indicators fishing according to the MSY approach and the 95th percentiles of those same indicators when fishing at $F_{0.7}$ were computed. $F_{0.7}$ was the highest fishing mortality scenario tested in the MSE simulations, and its distance to F_{MSY} can vary over stocks considered. To verify if the behavior of the threshold was dependent on how severe is $F_{0.7}$ compared to F_{MSY}, we provide a plot comparing the distance between the thresholds identified and between $F_{0.7}$ and F_{MSY}. We repeated the analysis separately on the MSE runs with a dynamic Harvest Control Rule (Figure 6.9.5 and Figure 6.9.6) and without a dy-namic Harvest Control Rule (Figure 6.9.7 and Figure 6.9.8).

Figure 6.9.5 shows the smoothed distribution of the indicators for each species at F_{MSY} and $F_{0.7}$ in MSE simulations where no dynamic HCR was implemented, with the associated 5th and 95th percentile, respectively. In analogy to chapter 6.6, the distributions at F_{MSY} tended to be quite symmetric, while at $F_{0.7}$ there is some tendency for bimodality or the emergence of additional peaks. Some distributions at $F_{0.7}$ are left-truncated, as the indicator drops to zero, and tends also to have a lower dispersion than distributions at F_{MSY}. Recruitment distribution for $F_{0.7}$ is concentrated around low values, as the absence of a dynamic HCR creates the potential for recruitment failure. In general, there is a low but heterogeneous degree of overlap between the two distributions over the stocks tested. In a few cases (3 out of 7 stocks: PLE-EC; SOL-NS and TUR) the distributions were showing a very low degree of overlap, implying that the threshold identified at $F_{0.7}$ can indicate a stock status that is unequivocally not in line with GES. In other cases, the threshold identified at $F_{0.7}$ were close or overlapping the indicators at F_{MSY}. Figure 6.9.6 indicates that for ABI_{MSY}, average age and POS, a great variability in thresholds behavior is observed when the distance between $F_{0.7}$ and F_{MSY} is large. Indicators including the recruitment shown decreas-ing variability at large distance between $F_{0.7}$ and F_{MSY}, probably due to the largest chance of re-cruitment failure at high F. The heterogeneous behavior suggests that distributions at high fish-ing mortality are dependent on the stock and their ability to indicate a condition for stocks that is not at GES can be limited.

Figure 6.9.7 shows the distribution of the indicators for each species at F_{MSY} and $F_{0.7}$ in MSE simulations where dynamic HCR was implemented, with the associated 5th and 95th percentile, respectively. In respect to figure 6.6.2, the distributions at $F_{0.7}$ tend to have a dispersion that is comparable (or larger) than the distributions at F_{MSY}. The degree of overlap between the two distributions over the stocks tested is also larger of that observed in Figure 6.6.2, on average. In one case (1 out of 7 stocks: PLE-EC) the distributions were showing a very low degree of overlap, implying that the threshold identified at $F_{0.7}$ can indicate a stock status that is unequivocally not in line with GES. In all the other cases, the threshold identified at $F_{0.7}$ were close, overlapping, or even exceeding the indicators at F_{MSY}. Figure 6.9.8 indicates that for average age and POS, a great variability in the difference between thresholds is observed when the distance between $F_{0.7}$ and F_{MSY} is large. In POS, the distance between thresholds is always narrow. The heterogeneous behavior suggests that distributions at high fishing mortality are dependent on the stock and their ability to indicate a condition for stocks that is not at GES can be limited.

Figure 6.9.5 Distribution of the indicators ABImsy, avgAge, POS, R, SSB/R at F_{MSY} and $F_{0.7}$, with the associated 5th and 95th percentile (dotted vertical lines), calculated for the stocks of COD-NS, PLE-EC, POK, SOL-EC, SOL-NS, TUR from MSE without a dynamic Harvest Control Rule.

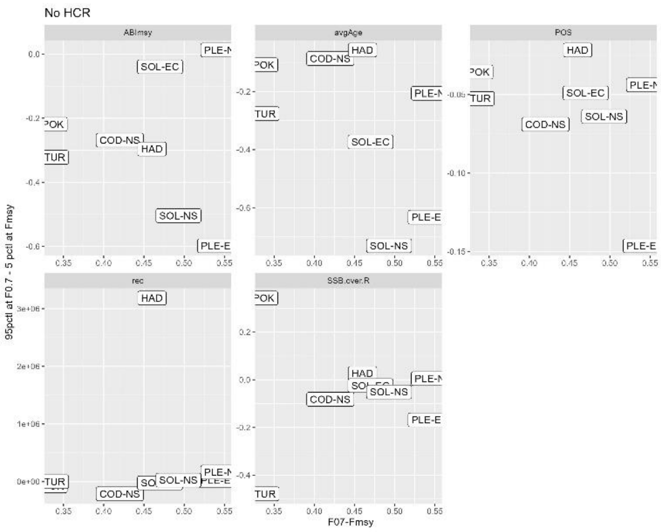


Figure 6.9.6 Distance between 95th percentile of the distribution at $F_{0.7}$ and the 5th percentile of the distribution at F_{MSY} of the indicators ABImsy, avgAge, POS, R, SSB/R compared to the difference between $F_{0.7}$ and F_{MSY} for the stocks of COD-NS, PLE-EC, POK, SOL-EC, SOL-NS, TUR from MSE without a dynamic Harvest Control Rule.

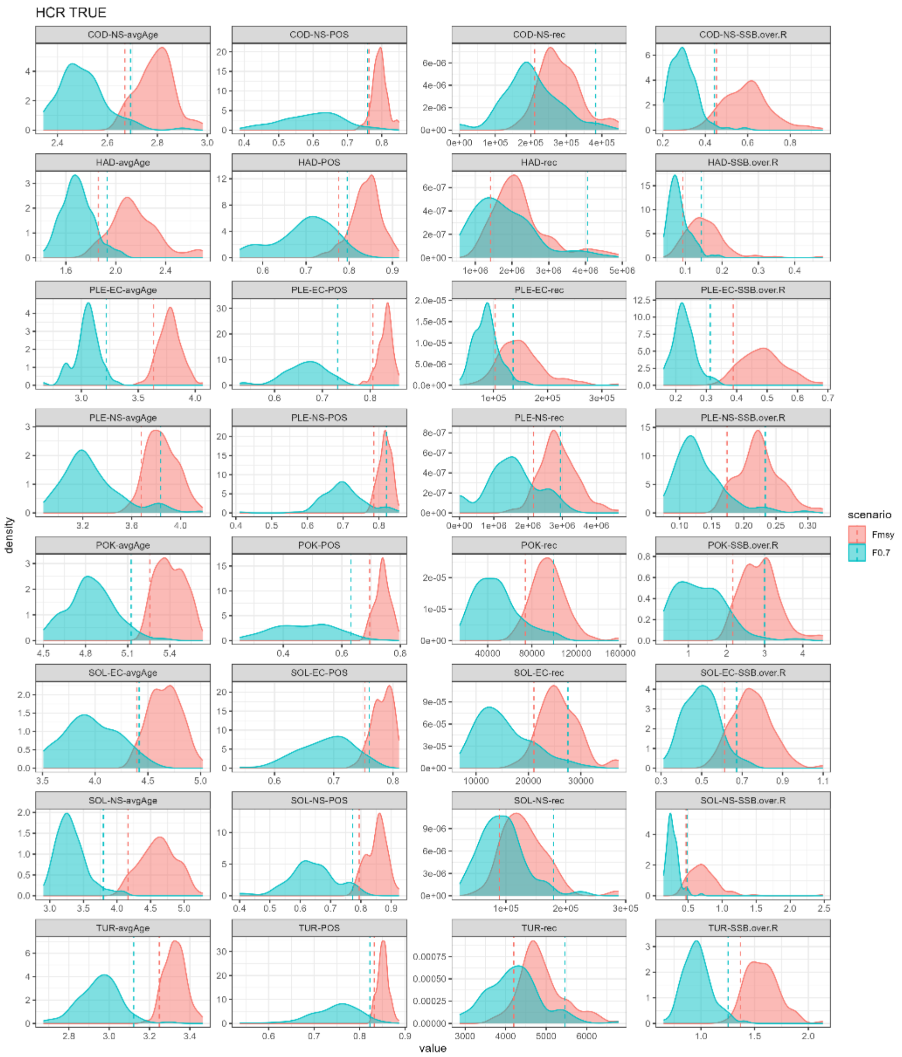


Figure 6.9.7 Distribution of the indicators ABImsy, avgAge, POS, R, SSB/R at F_{MSY} and $F_{0.7}$, with the associated 5th and 95th percentile (dotted vertical lines), calculated for the stocks of COD-NS, PLE-EC, POK, SOL-EC, SOL-NS, TUR from MSE with a dynamic Harvest Control Rule.

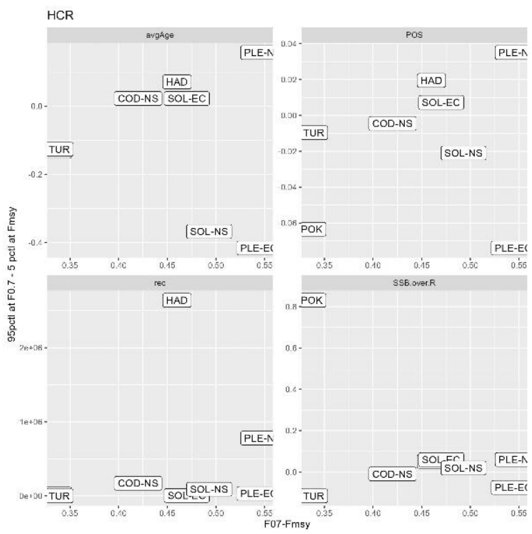


Figure 6.9.8 Distance between 95th percentile of the distribution at $F_{0.7}$ and the 5th percentile of the distribution at F_{MSY} of the indicators ABImsy, avgAge, POS, R, SSB/R compared to the difference between $F_{0.7}$ and F_{MSY} for the stocks of COD-NS, PLE-EC, POK, SOL-EC, SOL-NS, TUR from MSE without a dynamic Harvest Control Rule.

6.10 Comparison of historical development and potential thresholds

The potential thresholds compared to historical data were:

1) The 95th percentile of historic data with F/Fmsy >1.5
2) The 95th percentile of historic data with F/Fmsy >2
3) The 5th percentile of historic data from periods where F/Fmsy < 1 for at least three generations
4) The 5th percentile of simulated data from the MSE

The historical development of the indicators for the 7 demersal stocks are shown in figs. 6.10.1-7. Comparisons for mackerel and sprat are in figs. 5.3.4-5.3.6.

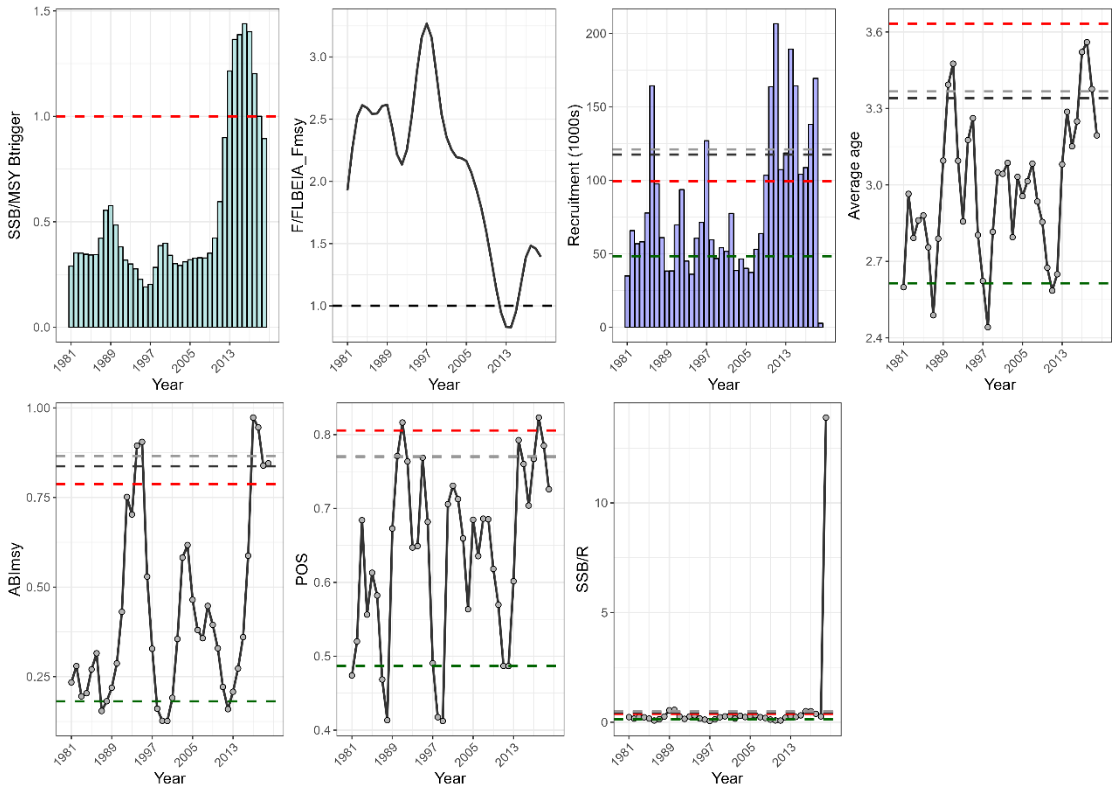


Fig. 6.10.1. Historical development in indicators compared with potential thresholds for Ple7d. The red and black dashed horizontal lines at a value of 1 in the SSB and F plots respectively, signifies equivalence between each metric of stock status and their respective reference points (MSY Btrigger and FMSY). For the other five indicators, dashed horizontal lines are coloured: dark green – 5th percentile of historic data from periods where F/Fmsy < 1, dark grey - 95th percentile of historic data with F/Fmsy >1.5, light grey – 95th percentile of historic data with F/Fmsy >2 and red - 5th percentile of simulated data from the MSE. In some cases, these thresholds/lines either overlap or were not considered by the group. For instance, the 5th percentile of the historic data from periods where F/Fmsy < 1 (dashed dark green line) was only calculated for 4 of the 7 stocks (Ple7d, Ple420, Pok27.3a46 and Tur4)

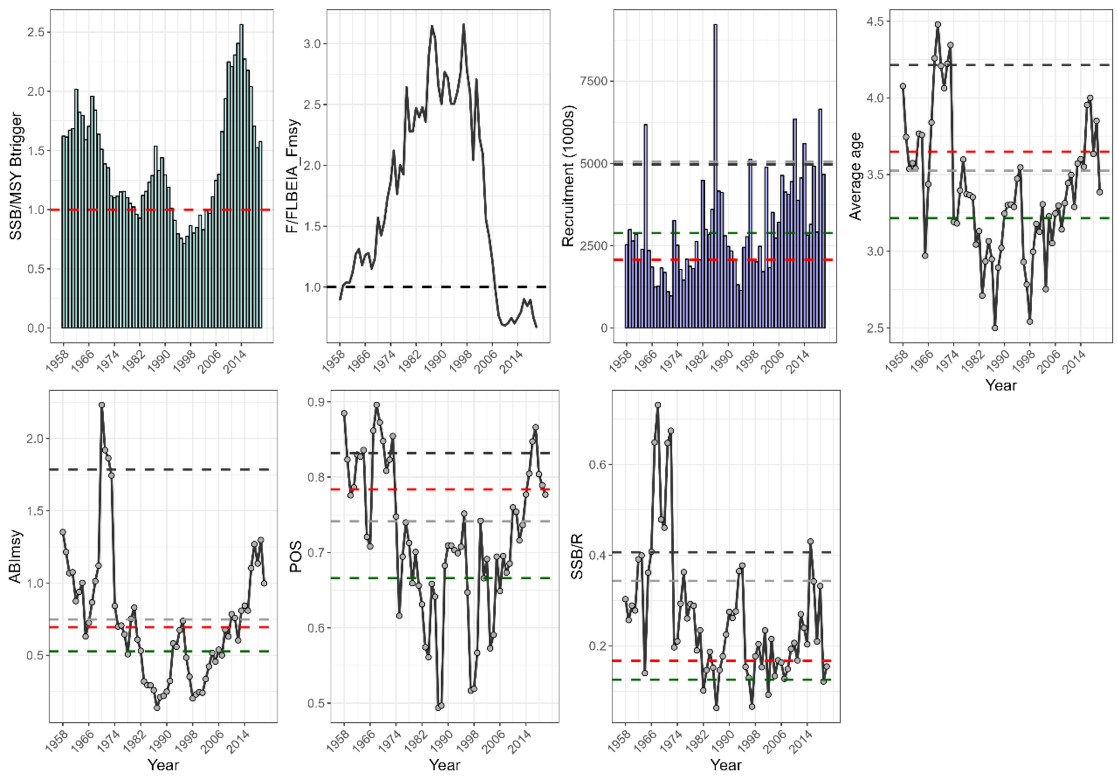


Fig. 6.10.2. Historical development in indicators compared with potential thresholds for Ple420. **The red and black dashed horizontal lines at a value of 1 in the SSB and F plots respectively, signifies equivalence between each metric of stock status and their respective reference points (MSY Btrigger and FMSY). For the other five indicators, dashed horizontal lines are coloured: dark green – 5th percentile of historic data from periods where F/Fmsy < 1, dark grey - 95th percentile of historic data with F/Fmsy >1.5, light grey – 95th percentile of historic data with F/Fmsy >2 and red - 5th percentile of simulated data from the MSE. In some cases, these thresholds/lines either overlap or were not considered by the group. For instance, the 5th percentile of the historic data from periods where F/Fmsy < 1 (dashed dark green line) was only calculated for 4 of the 7 stocks (Ple7d, Ple420, Pok27.3a46 and Tur4).**

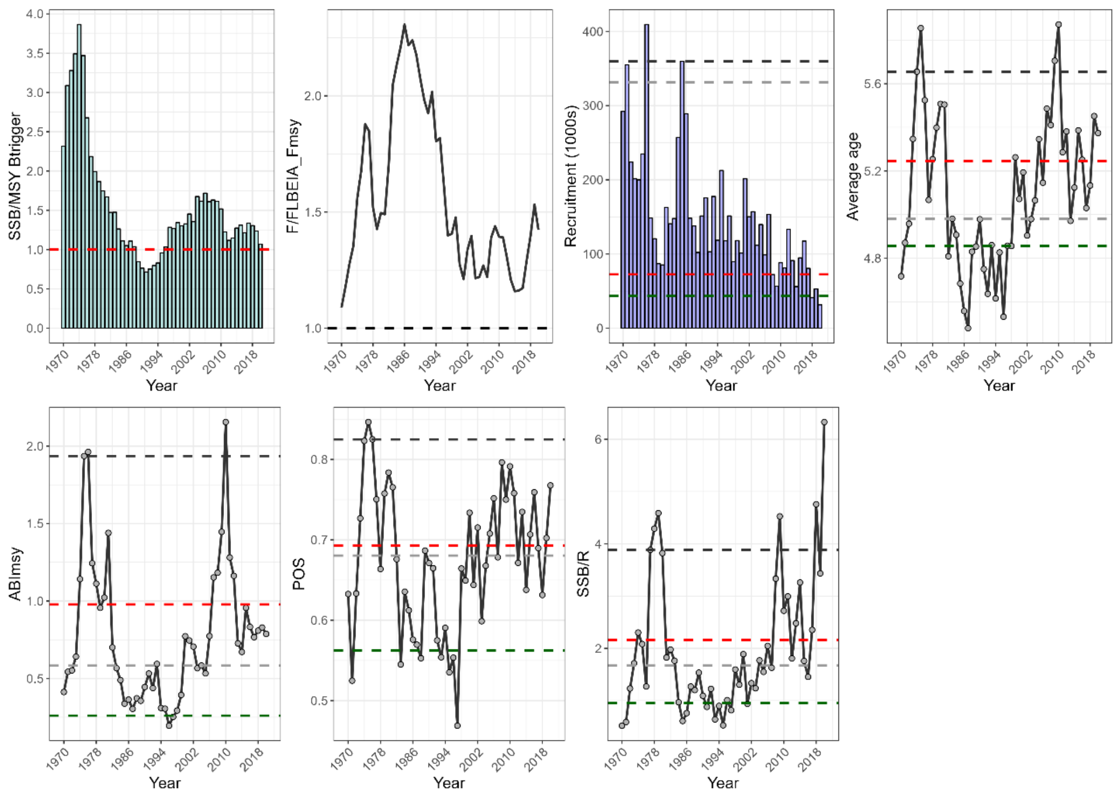


Fig. 6.10.3. Historical development in indicators compared with potential thresholds for Pok27.3a46. **The red and black dashed horizontal lines at a value of 1 in the SSB and F plots respectively, signifies equivalence between each metric of stock status and their respective reference points (MSY Btrigger and FMSY). For the other five indicators, dashed horizontal lines are coloured: dark green – 5th percentile of historic data from periods where F/Fmsy < 1, dark grey - 95th percentile of historic data with F/Fmsy >1.5, light grey – 95th percentile of historic data with F/Fmsy >2 and red - 5th percentile of simulated data from the MSE. In some cases, these thresholds/lines either overlap or were not considered by the group. For instance, the 5th percentile of the historic data from periods where F/Fmsy < 1 (dashed dark green line) was only calculated for 4 of the 7 stocks (Ple7d, Ple420, Pok27.3a46 and Tur4)**

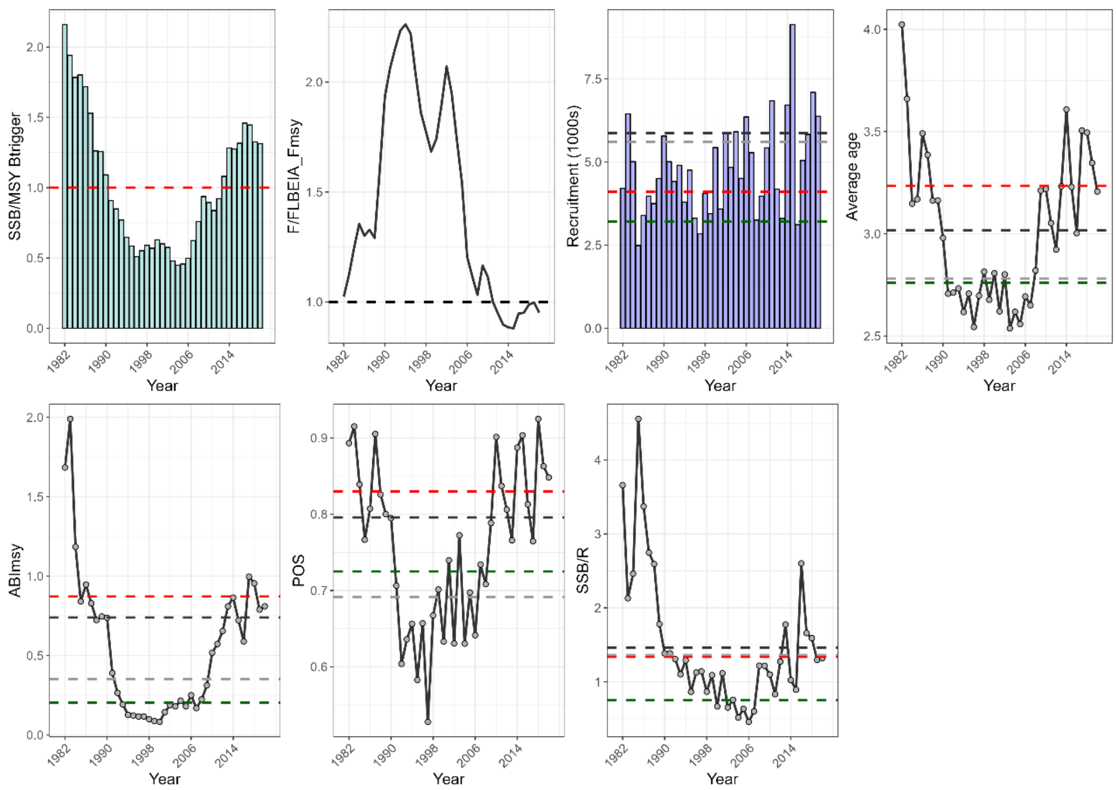


Fig. 6.10.4. Historical development in indicators compared with potential thresholds for Tur4. **The red and black dashed horizontal lines at a value of 1 in the SSB and F plots respectively, signifies equivalence between each metric of stock status and their respective reference points (MSY Btrigger and FMSY). For the other five indicators, dashed horizontal lines are coloured: dark green – 5th percentile of historic data from periods where F/Fmsy < 1, dark grey - 95th percentile of historic data with F/Fmsy >1.5, light grey – 95th percentile of historic data with F/Fmsy >2 and red - 5th percentile of simulated data from the MSE. In some cases, these thresholds/lines either overlap or were not considered by the group. For instance, the 5th percentile of the historic data from periods where F/Fmsy < 1 (dashed dark green line) was only calculated for 4 of the 7 stocks (Ple7d, Ple420, Pok27.3a46 and Tur4)**

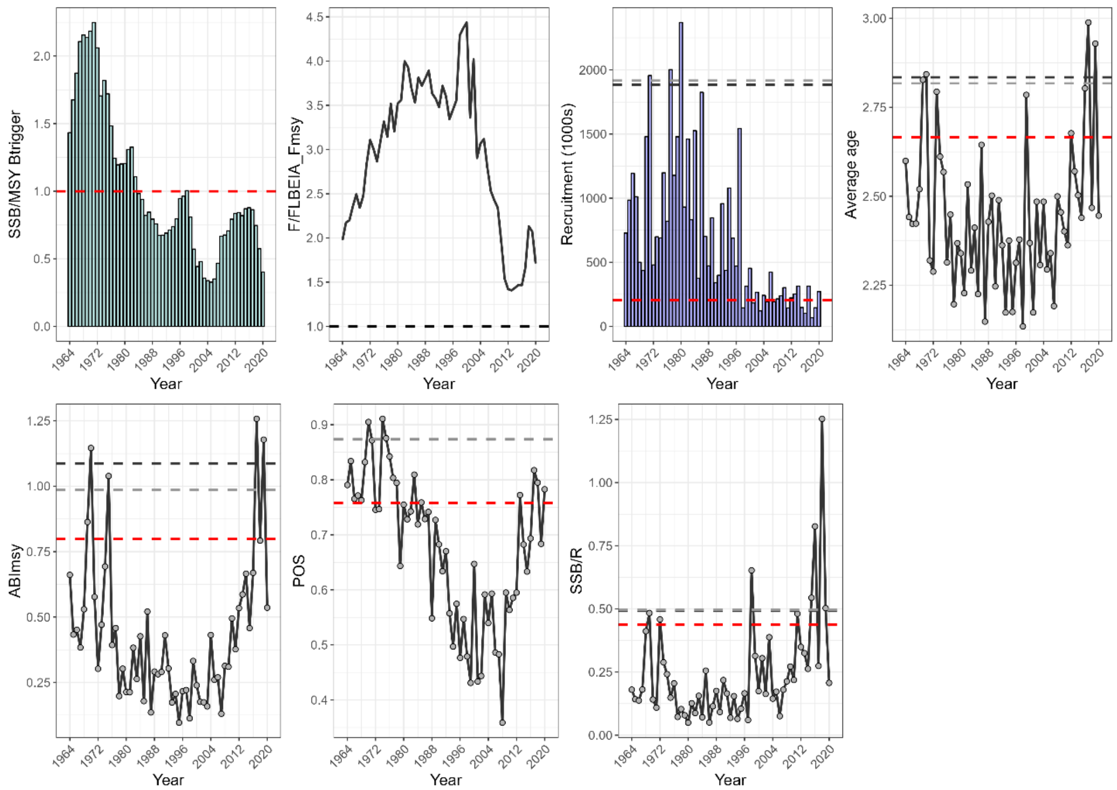


Fig. 6.10.5. Historical development in indicators compared with potential thresholds for cod.27.47d20. **The red and black dashed horizontal lines at a value of 1 in the SSB and F plots respectively, signifies equivalence between each metric of stock status and their respective reference points (MSY Btrigger and FMSY). For the other five indicators, dashed horizontal lines are coloured: dark green – 5th percentile of historic data from periods where F/Fmsy < 1, dark grey - 95th percentile of historic data with F/Fmsy >1.5, light grey – 95th percentile of historic data with F/Fmsy >2 and red - 5th percentile of simulated data from the MSE. In some cases, these thresholds/lines either overlap or were not considered by the group. For instance, the 5th percentile of the historic data from periods where F/Fmsy < 1 (dashed dark green line) was only calculated for 4 of the 7 stocks (Ple7d, Ple420, Pok27.3a46 and Tur4)**

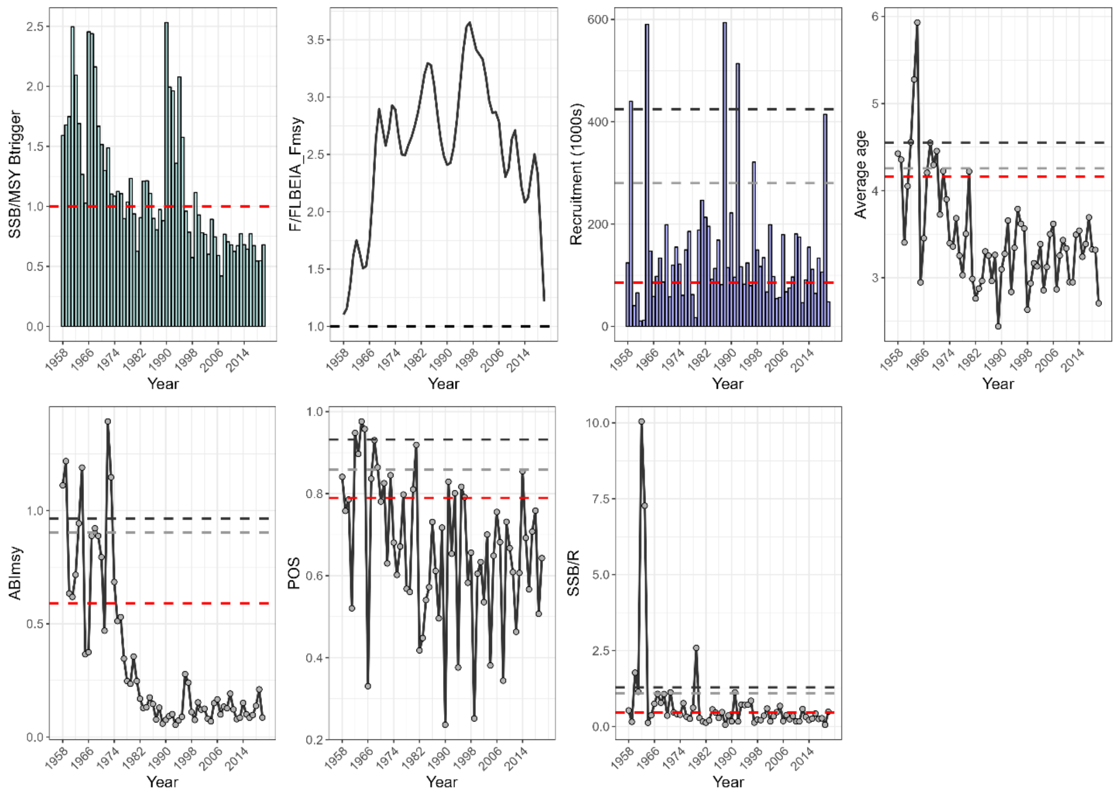


Fig. 6.10.6. Historical development in indicators compared with potential thresholds for Sol27.4. **The red and black dashed horizontal lines at a value of 1 in the SSB and F plots respectively, signifies equivalence between each metric of stock status and their respective reference points (MSY Btrigger and FMSY). For the other five indicators, dashed horizontal lines are coloured: dark green – 5th percentile of historic data from periods where F/Fmsy < 1, dark grey - 95th percentile of historic data with F/Fmsy >1.5, light grey – 95th percentile of historic data with F/Fmsy >2 and red - 5th percentile of simulated data from the MSE. In some cases, these thresholds/lines either overlap or were not considered by the group. For instance, the 5th percentile of the historic data from periods where F/Fmsy < 1 (dashed dark green line) was only calculated for 4 of the 7 stocks (Ple7d, Ple420, Pok27.3a46 and Tur4)**

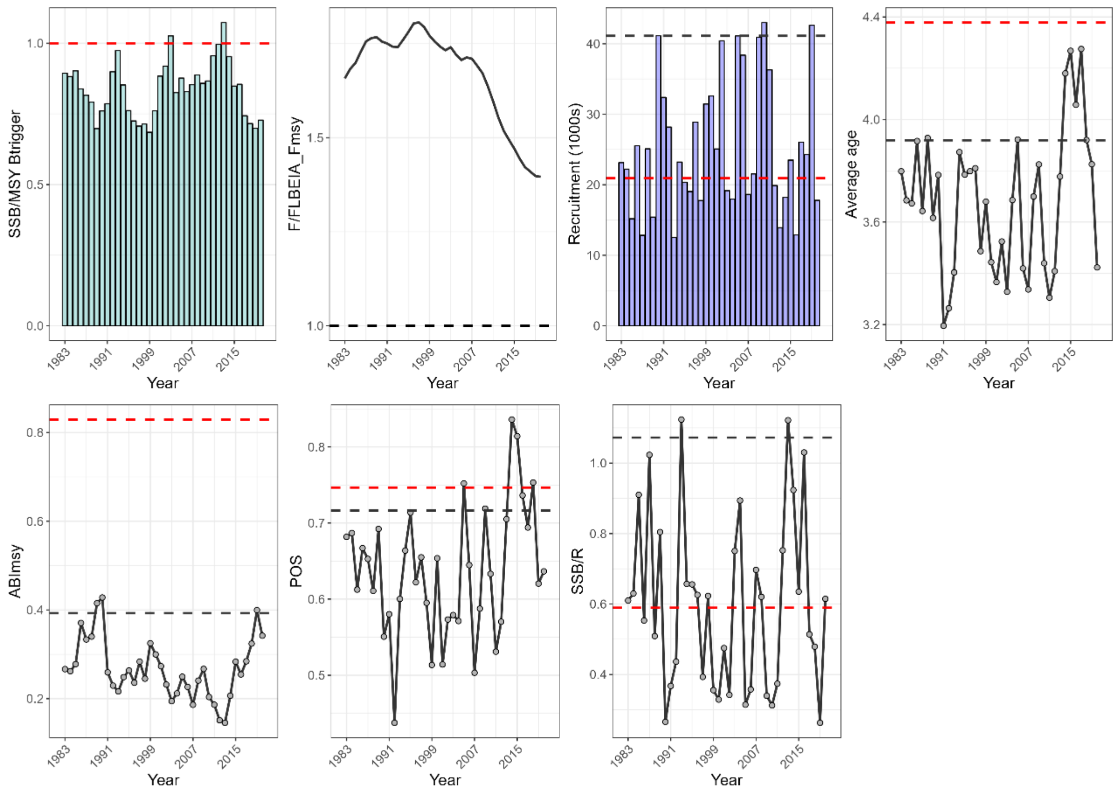


Fig. 6.10.7. Historical development in indicators compared with potential thresholds for Sol7d. **The red and black dashed horizontal lines at a value of 1 in the SSB and F plots respectively, signifies equivalence between each metric of stock status and their respective reference points (MSY Btrigger and FMSY). For the other five indicators, dashed horizontal lines are coloured: dark green – 5th percentile of historic data from periods where F/Fmsy < 1, dark grey - 95th percentile of historic data with F/Fmsy >1.5, light grey – 95th percentile of historic data with F/Fmsy >2 and red - 5th percentile of simulated data from the MSE. In some cases, these thresholds/lines either overlap or were not considered by the group. For instance, the 5th percentile of the historic data from periods where F/Fmsy < 1 (dashed dark green line) was only calculated for 4 of the 7 stocks (Ple7d, Ple420, Pok27.3a46 and Tur4)**

6.11 Length based indicators and thresholds

The length based and of the age-based indicators calculated on observed data were less sensitive than age-based indicators to fishing mortalities above F_{MSY}. In some cases, the length based indicators did not reflect stock changes and though L90R was theoretically less sensitive to recruitment than L90, this could not be demonstrated in the data for the stocks examined. Using the principles of age based indicators, thresholds for stock health indicators based on survey catch at length can be derived from simulations, preferably demonstrating as part of the estimation that the model is capable of reproducing dynamics of the stock. For the stocks examined at WKSIMULD3, it proved very difficult to produce simulated trajectories of the indicators that tracked the observed development, perhaps indicating that there are other factors influencing the indicators such as variable catchabilities between surveys or that the indicators are not very responsive at levels of F greatly above F_{MSY}.

7 Definitions of thresholds for the D3C3 indicators and thresholds

7.1 Considerations for stocks of different life histories

Recruitment is only vaguely related to management action and retaining SSB above the level that impairs recruitment seems the only strong management measure to ensure this. However, there is a potential benefit in tracking recruitment to reveal cases where the effects of a productivity failure on the stock and exploitation can be alleviated through rapid regulation of management targets under D3C1. The same argument can be made for mean weight at age. Hence, these indicators are recommended to be used for surveillance and potential update of thresholds for D3C1 and D3C2.

For species having high spawning mortality, the indicators reflecting the need to have old spawners (POS, ASA, ABI, SSB/R) do not retain the intended positive effect on productivity. Indeed, for some species, delayed spawning reflects deteriorating growth conditions rather than improved stock health. As a result, it is not recommended to use the indicators POS, ASA, ABI and SSB/R for stocks with high spawning or senescence mortality such as e.g. capelin and cephalopods.

For short lived (stocks that mature early and have a natural short life span) or small pelagic stocks, the age structure indicators POS, ASA, ABI and SSB/R were highly variable in simulations even when fished using the MSY approach for generations. This is presumably due to their natural high variability in recruitment. Further, for some species, high age at spawning may lead senescence rather than increased viability of the spawning products (Beverton et al. 2004, Christiansen et al 2008, Benoit et al. 2019). Variability and growth likely also add to this pattern, but these factors were not varied in the MSEs examined at WKSIMULD3 except for mackerel. As a result of this high variability, there was no strong link between the age-structure indicators and F or SSB, leading to a case where the indicators cannot be used for management action. However, given the low number of stocks examined, the results can not firmly be interpreted as applying to other pelagic stocks.

No long lived species where included in the MSEs when defining long lived here as species which are generally longlived (e.g. ling, halibut, spurdog, redfish) as well as species which are longlived under specific conditions (e.g. cod at Iceland and Barents Sea). WKD3C3THRESHOLDS noted that the age structure indicators appeared to react substantially later than F and SSB for long lived stocks, and hence the added value for management action of monitoring them was unclear. Further work is needed to investigate the behaviour of the indicators for these species is needed to determine if they are all relevant in these cases.

8 Framework for estimating D3C3 indicators and thresholds

The indicators recommended by WKD3C3THRESHOLDS were all retained at WKSIMULD3 as appropriate, and the indicator SSB/R added though this can be overly too responsive to recruitment. The indicators encompass three aspects of the stock: production of new individuals, growth of individuals and age structure (Table 8.1). It was unclear to WKSIMULD3 participants whether the indicator of selectivity in the fishery Fjuv/Fapical belonged under D3C3 stock health as it seemed more appropriately placed under D3C1 together with F and F_{MSY}, as the selectivity pattern is a major influencing factor in the estimation of F_{MSY}. These indicators will complement the existing D3 indicators as described in Table 8.2.

Table 8.1. Recommended indicators of D3C3. Their applicability varies between stocks and hence only indicators relevant to the specific stock should be included. As a minimum, recruitment indicators should be included as knowledge of recruitment is key to interpreting age structure indicators.

Aspect	Age based indicator	Length based indicator
Recruitment	R	R
Growth of individuals	ASW	-
Age structure	ABI_{MSY}, POS, ASA (potentially SSB/R)	L90R, L90
Fisheries selectivity	Fjuv/Fapical	-

Table 8.2. Indicators of the three D3 criteria following the guidance in this report.

D3 criterion	Stock component	Indicator	Aspect of stock health	Threshold
D3C1	**All**	F	Total exploitation	F_{MSY}
	Recruits & juveniles	***Fjuv/Fapical***	*Fisheries induced mortality of juveniles*	*To be determined*
D3C2	**All spawners**	**SSB**	Stock size/abundance	MSY $B_{escapement}$ MSY $B_{trigger}$ B_{pa}
D3C3	***Recruits & Juveniles***	***R***	*Stock productivity*	*5%-percentile either from R-values > B_{lim} or from MSE with $F=F_{MSY}$*
	Growth of individuals	***ASW***	*Growth*	*5%-percentile either from ASW-values observed or from MSE with $F=F_{MSY}$*
	Age structure	***ABI_{MSY}, POS, ASA, SSB/R, L90R, L90***	*Chance of survival to older age classes*	*5%-percentile from indicator-values from MSE with $F=F_{MSY}$*

8.1 Assessment based indicators

Where age based assessments are available, the input and output for these provide quality assured data at the level of the stock and hence provide the best possible data foundation to evaluate the health of the stock under D3C3. The indicators investigated here can readily be estimated from input and output of age based assessments. Estimation of thresholds however requires a stock specific MSE appropriately implementing ICES principles for best practice in Management Strategy Evaluations along with historical data. For future estimation of thresholds, using predictions of MSEs recently used to define stock reference points under D3C1 and D3C2 or produced specifically to mimic these, including stock specific decisions on the shape of e.g. the stock recruitment relationship ensures that the resulting thresholds suggested are methodologically consistent with the thresholds for D3C1 and D3C2 and the MSY approach to fisheries management guiding the threshold setting on these criteria. Following the work in WKSIMULD3, the decision tree suggested at WKD3C3THRESHOLDS was updated (fig. 8.1.1 and 8.1.2). The estimation of thresholds should ideally be performed whenever thresholds for D3C1 and D3C2 are updated (for example during benchmarks). This will ensure consistency between the thresholds under the three D3 criteria.

The option to estimate the indicator thresholds independently of methods used to define stock reference points for D3C1 and D3C2 (F_{MSY} and biomass reference points) was also discussed but the group considered that this approach would provide results that were inconsistent as the F_{MSY} and distribution of the indicators at this fishing pressure would differ. The estimation of indicators using existing frameworks for estimation had the advantage that these methods are already peer reviewed and quality assured for advice through independent review (during e.g. ICES benchmarks). Further, as this estimation will be conducted by a group of experts on the specific stock, it is ensured that the indicator is based on the best available knowledge of the stock. This knowledge can influence the setting of MSE submodels, the choice of the indicators to use as well as the setting of thresholds.

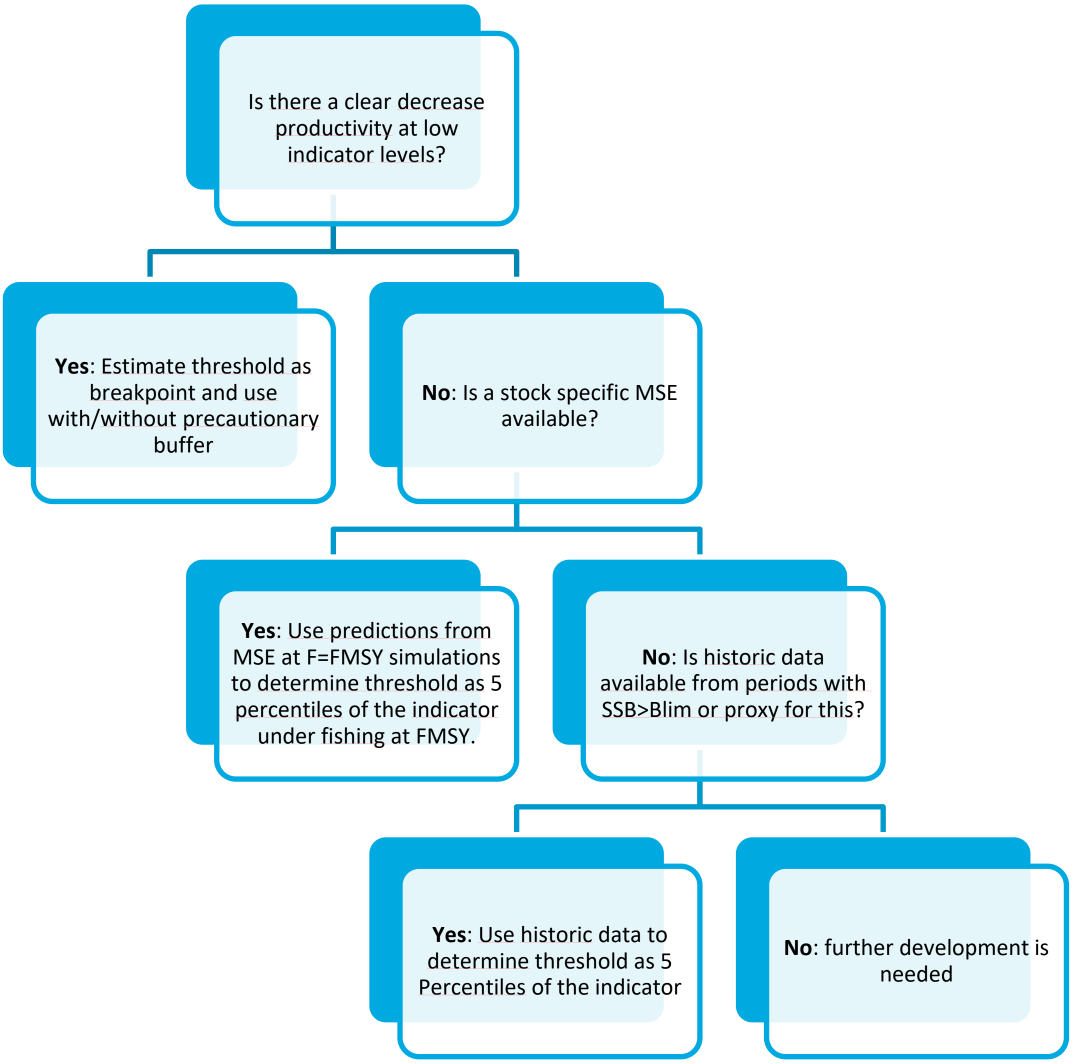


Fig. 8.1.1. Decision tree for assessment-based evaluations of recruitment from WKSIMULD3. ASW follows the same decision tree with the exception that all historic data are used in the bottom box.

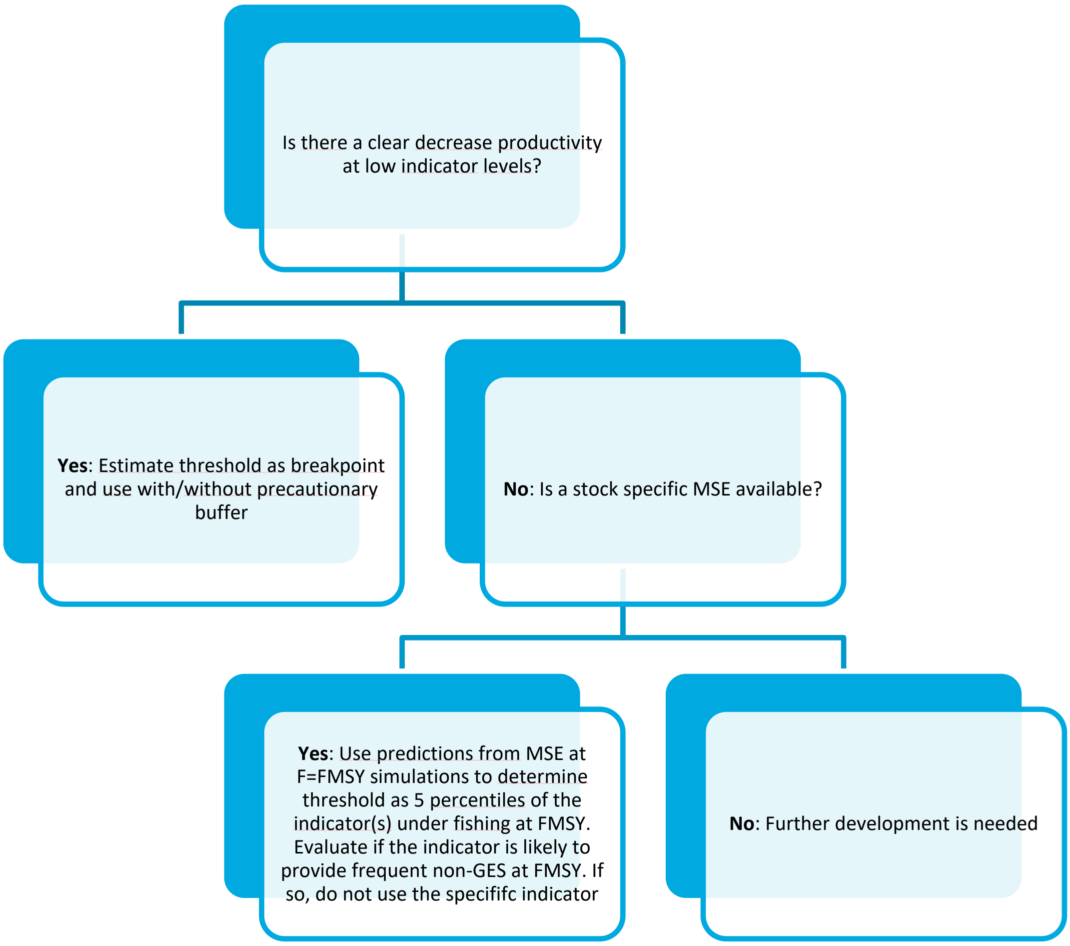


Fig. 8.1.2. Decision tree for assessment-based evaluations of age-structure indicators (ABI, POS, ASA) from WKSIMULD3.

8.2 Survey based indicators

Where age based assessments are not available, the use of length based survey catch rates to evaluate the health of the stock under D3C3 was investigated at WKSIMULD3. For both age and length structure indicators, there was very limited response of the indicator at high fishing mortality but substantial interannual variation. While there is a high theoretical consistency of the length and age based indicators, their consistency was low in the investigated data, most likely due to the high historic fishing mortality. For stocks fished at levels above $1.5F_{MSY}$, it is probably preferable to denote the stock as being below the threshold as these stocks are unlikely to exhibit healthy age structure.

Survey data pose specific challenges as they may only cover part of the geographical extent or size structure of the stock and the timing of the survey may vary somewhat between years leading to different representation of the youngest yearclass. Based on the evaluations at WKSIMULD3 and WKD3C3THRESHOLD, it is recommended to investigate if the survey catches appear to be consistent (e.g. recruits are seen in two consecutive years) before using the data. In the case where data appear inconsistent, the indicator should not be estimated as changes will likely be caused by differences in stock catchability. Following the age based indicators, the decision trees in figs. 8.1.1 and 8.1.2 can also be used for length based indicators.

9 For the stocks identified at c), rank potential management options in decreasing order of expected effectiveness.

There are a wide range of management measures in use and these can broadly be divided into two categories:

1. Measures aiming to reduce fishing pressure through regulating either input (fishing effort) or output (catches) and
2. Measures aiming to reduce fishing pressure on specific stock components such as recruits or spawning fish (closed areas and seasons, realtime closures, technical gear specifications)

The former assumes that the selectivity pattern of the fishery remains unchanged from the present whereas the latter assumes a change. Changes such as closures of areas and times with high juvenile abundance and increased mesh size can reduce F of juveniles whereas spawning closures may reduce F of spawning fish. These changes to F will change the age structure of the stock if all else stays the same. For example, we can imagine that we no longer catch fish of the recruiting age class. This will mean a greater proportion of the fish survive to the subsequent age classes. However, changes to juvenile selectivity will not affect the age structure indicators, as these explicitly exclude recruiting fish to avoid impacts of occasional large recruitments and therefore only investigate the age distribution within ages older than recruitment age. For the age structure indicators to be impacted, it is necessary to increase survival within the non-recruiting ages. This can be attained through decreasing F for all ages or some ages. Which of these measures will be most effective will depend on their operational feasibility for the specific stock and can readily be investigated in MSEs by adding scenarios with realistic changes in selectivity due to e.g. increased mesh size or closure of spawning areas and/or seasons. In accordance with current principles for D3C1 and D3C2, WKSIMULD3 investigated management options relating to rescaling fishing mortality at the existing selection pattern which was a highly effective means of regulating the age-structure indicators at levels of F around F_{MSY} (section 5.4.3). The group considered that the stock specific considerations on most appropriate management measures can be investigated further in designated MSEs.

10 Knowledge gaps and further studies needed

The Management Strategy Evaluation style simulations underlying the conclusions here are, of necessity, based on simplified stock dynamics. It is important to consider the simplifications in the simulations and their potential impact on the performance of the simulations performed here. The following are potentially important processes which are not currently included, together with a short discussion of how they could impact the results of the simulations.

- **Fecundity of big fish**: Including variable fecundity of large fish will matter when the stock drops to near the Blim level. At these low stock levels reduced age structure may give worse impacts on stock recruitment than the current simulations show.
- **Density dependence**: Density dependence will likely matter as you get near carrying capacity, i.e. long term fishing at F=0. Food limitation at these high levels would result in the biomass of old fish being lower than the simulations suggest, since fish will not grow as large as the simulations show. The impact on the numbers of large fish will be less than on the biomass (there will still be some impact due to conversion of F in numbers to catch in weight).
- **Variable food:** Including variable food supply (bottom up control) will act in a similar way to including density dependence, as there will be a larger change in the simulated the old fish indicators in biomass than in numbers.
- **Cannibalism and other reasons for reduced juvenile survival at high stock size (Ricker)**: Including cannibalism (either in the recruitment function or in the simulated stock dynamics) would reduce recruitment at high stock sizes. Ignoring this where it matters means that the simulations will underestimate the "old fish" indicators rise more steeply at low F. Ignoring this may also result in an underestimate of the variability in the indicators (since cannibalism can induce cycles in stock development).
- **Different life histories**: The stocks simulated here represent only stocks with medium life spans (not small pelagics, not long-lived stocks). It is not clear to what extent the simulation results here can be extended to these other life histories. Therefore simulations should be conducted on these other life history stocks before extrapolating any conclusions here to the small pelagics or the long lived stocks.

WKSIMULD3 recommends that further analyses are conducted to investigate particularly the possible common patterns for different life histories once the indicators and thresholds have been estimated in approved MSEs for a larger number of stocks.

Reference list

Barnett, L. A. K., Branch, T. A., Ranasinghe, R. A., and Essington, T. E. (2017). Old-Growth Fishes Become Scarce under Fishing. *Current Biology*, 27: 2843-2848.e2.

Berkeley, S. A., Hixon, M. A., Larson, R. J., & Love, M. S. (2004). Fisheries sustainability via protection of age structure and spatial distribution of fish populations. Fisheries, 29(8), 23-32.

Bertrand, J. A., Gil de Sola, L., Papaconstantinou, C., Relini, G., and Souplet, A. 2002. The general specifications of the MEDITS surveys. Scientia Marina, 66: 9. CSIC Consejo Superior de Investigaciones Cientificas 2.

Benoît, H. P., Benhalima, K., & McDermid, J. L. (2020). Histological evidence of reproductive senescence in atlantic herring (Clupea harengus). Canadian Journal of Zoology, 98(1), 73-78.

Beverton, R. J., Hylen, A., Østvedt, O. J., Alvsvaag, J., & Iles, T. C. (2004). Growth, maturation, and longevity of maturation cohorts of Norwegian spring-spawning herring. ICES Journal of Marine Science, 61(2), 165-175.

Carbonara, P., Masnadi, F., Donato, F., Sabatini, L., Pellini, G., Cardinale, M., and Scarcella, G. 2023. Biphasic versus monophasic growth curve equation, an application to common sole (Solea solea, L.) in the northern and central Adriatic Sea. Fisheries Research, 263: 106694

Christiansen, J. S., Præbel, K., Siikavuopio, S. I., & Carscadden, J. E. (2008). Facultative semelparity in capelin Mallotus villosus (Osmeridae)-an experimental test of a life history phenomenon in a sub-arctic fish. Journal of Experimental Marine Biology and Ecology, 360(1), 47-55.

Delius G, Scott F, Blanchard J, Andersen K (2023). mizer: Dynamic Multi-Species Size Spectrum Modelling. R package version 2.4.1, https://CRAN.R-project.org/package=mizer.

Duplisea (2019). _dublogistic: Explore the double logistic function as a selectivity curve for fisheries modelling_. R package version 0.1.0.

Duskey, E. (2023). Metabolic prioritization of fish in hypoxic waters: an integrative modeling approach. Frontiers in Marine Science, 10. doi:10.3389/fmars.2023.1206506

Eero, M., Cardinale, M., & Storr-Paulsen, M. (2020). Emerging challenges for resource management under ecosystem change: Example of cod in the Baltic Sea. Ocean & Coastal Management, 198, 105314. doi:10.1016/j.ocecoaman.2020.105314

European Commission, 2022. MSFD CIS Guidance Document No. 19, Article 8 MSFD, May 2022.

FAO. 2019. Scientific Advisory Committee On Fisheries (SAC) Working Group on Stock Assessment of Demersal Species (WGSAD) - Benchmark session for the assessment of European hake in GSAs 17-18 , FAO headquarters, Rome, Italy, 15-18 January 2019.

FAO. 2021. Scientific Advisory Committee On Fisheries (SAC) Working Group on Stock Assessment of Demersal Species (WGSAD) - Benchmark session for the assessment of common sole in GSA 17 , Online, 12-16 April 2021.

Froese, R., Demirel, N., and Sampang, A. 2015. An overall indicator for the good environmental status of marine waters based on commercially exploited species. Marine Policy, 51: 230-237.

Froese, R., Winker, H., Gascuel, D., Sumaila, U. R., & Pauly, D. (2016). Minimizing the impact of fishing. Fish and Fisheries, 17(3), 785–802.

Garcia, D., Sánchez, S., Prellezo, R., Urtizberea, A., Andrés, M., 2017. FLBEIA : A simulation model to conduct Bio-Economic evaluation of fisheries management strategies. SoftwareX 6, 141–147. https://doi.org/10.1016/j.softx.2017.06.001

Griffiths, C. A., Winker, H., Bartolino, V., Wennhage, H., Orio, A., & Cardinale, M. (2023). Including older fish in fisheries management: A new age-based indicator and reference point for exploited fish stocks. Fish and Fisheries. Hixon, M. A., Johnson, D. W., and Sogard, S. M. 2014. BOFFFFs: on the importance of conserving old-growth age structure in fishery populations. *ICES Journal of Marine Science*, 71: 2171–2185.

Heino, M., Dieckmann, U., and Godø, O. R. 2002. Measuring probabilistic reaction norms for age and size at maturation. Evolution, 56: 669-678.

ICES.2012. Marine Strategy Framework Directive - Descriptor 3+, ICES CM2012/ACOM:62. 173 pp.

ICES. 2016a. Report of the workshop on guidance on development of operational methods for the evaluation of the MSFD Criterion D3.3 (WKIND3.3i). ICES CM 2016/ACOM:44. 97 pp.

ICES. 2016b. Report of the Workshop on guidance on development of operational methods for the evaluation of the MSFD criterion D3.3 (WKIND3.3ii) ICES CM 2016/ACOM:44. 161 pp.

ICES. 2017. Report of the Workshop on guidance on development of operational methods for the evaluation of the MSFD criterion D3.3 (WKIND3.3ii), 1–4 November 2016, ICES HQ, Copenhagen, Denmark, ICES CM 2016/ACOM:44. 155pp

ICES 2021, sr.2021.14

ICES, 2021. ICES fisheries management reference points for category 1 and 2 stocks. ICES Technical Guidelines. https://doi.org/10.17895/ICES.ADVICE.7891

ICES. 2022. Workshop on ICES reference points (WKREF1). *ICES Scientific Reports*, 4: 70 pp.

ICES, 2022. Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). ICES Scientific Reports 4, 1321. https://doi.org/10.17895/ICES.PUB.19786285

ICES. 2023. Workshop on the Development of Quantitative Assessment Methodologies based on Life-history traits, exploitation characteristics, and other relevant parameters for data-limited stocks (WKLIFEXII). ICES Scientific Reports. 5:103. 111 pp. https://doi.org/10.17895/ices.pub.24581343

Kell, L.T., Mosqueira, I., Grosjean, P., Fromentin, J.-M., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M.A., Poos, J.J., Scott, F., Scott, R.D., 2007. FLR: an open-source framework for the evaluation and development of management strategies. Ices Journal of Marine Science 64, 640–646. https://doi.org/10.1093/icesjms/fsm012

Mannini, A. 2020. The JRC MEDITS R script. JRC119776. Publications Office of the European Union, Luxembourg (Luxembourg).

Marteinsdottir G. & Begg G.A. (2002) Essential relationships incorporating the influence of age, size and condition on variables required for estimation of reproductive potential in Atlantic cod Gadus morhua. Mar. Ecol. Prog. Ser. 235, 235–256

McKendrick, A. (1925). Applications of Mathematics to Medical Problems. Proceedings of the Edinburgh Mathematical Society, 44, 98-130. doi:10.1017/S0013091500034428

Probst, W. N., Stelzenmüller, V., and Kraus, G. 2013. A simulation-approach to assess the size structure of commercially exploited fish populations within the European Marine Strategy Framework Directive. Ecological Indicators, 24: 621-632.

Probst, W. N., Kempf, A., Taylor, M., Martinez, I., and Miller, D. 2021. Six steps to produce stock assessments for the Marine Strategy Framework Directive compliant with Descriptor 3. ICES Journal of Marine Science, 78: 1229-1240.

Probst, W. N. 2023. An approach to assess exploited fish stocks compliant to the requirements of he Marine Strategy Framework Directive (MSFD) including criterion D3C3. Ecological Indicators, 146.

Rufino, M. M., Bez, N., and Brind'Amour, A. 2018. Integrating spatial indicators in the surveillance of exploited marine ecosystems. PLoS ONE, 13: e0207538.

Shephard, S., Greenstreet, S. P. R., Piet, G. J., Rindorf, A., and Dickey-Collas, M. 2015. Surveillance indicators and their use in implementation of the Marine Strategy Framework Directive. ICES Journal of Marine Science, 72: 2269-2277.

Shin, Y., Rochet, M., Jennings, S., Field, J., and Gislason, H. 2005. Using size-based indicators to evaluate the ecosystem effects of fishing. ICES Journal of Marine Science, 62: 384-396.

Simmonds, J., Hjorleifsson, E., Millar, C., 2022. msy: Estimation of Equilibrium Reference Points for Fisheries.

STECF, 2020. Review of technical measures (part 1) (STECF-20-02). EUR 28359 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-27161-1, doi:10.2760/734593, JRC123092.

STECF, 2021. Review of the Technical Measures Regulation (STECF-21-07). Publications Office of the European Union, Luxembourg, 2021, EUR 28359 EN, ISBN 978-92-76-45890-6, doi:10.2760/790781, JRC127718

STECF, 2022a. Stock Assessments: demersal stocks in the western Mediterranean Sea. (STECF-22-09). Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/00380, JRC132120.

STECF, 2022b. Stock Assessments: demersal stocks in Adriatic, Ionian and Aegean Seas and straits of Sicily (STECF-22-16). Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/25344, JRC132157

STECF. 2023. Scientific, Technical and Economic Committee for Fisheries (STECF): stock assessments: demersal stocks in the Western Mediterranean Sea 2022 (STECF-22-09). Publications Office of the European Union, Luxembourg.

Taylor, M.H. (2015) fishdynr: an R package of fisheries science related population dynamics models. R package version 0.4. Available at: https://github.com/marchtaylor/fishdynr.

Taylor, M., & Mildenberger, T. (2015). fishdynr: An R package of fisheries science related population dynamics models.

Tornero Alvarez M.V., Palma M., Boschetti S., Cardoso A.C., Druon J.-N., Kotta M., Louropoulou E., Magliozzi C., Palialexis A., Piroddi C., Ruiz-Orejón L.F., Vasilakopoulos P., Vighi M., Hanke G., Marine Strategy Framework Directive. Review and analysis of EU Member States' 2020 reports on Monitoring Programmes (MSFD Article 11), Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/8457, JRC129363

van Deurs, M., Jacobsen, N., Behrens, J., Henriksen, O., & Rindorf, A. (2023). The interactions between fishing mortality , age , condition and recruitment in exploited fish populations in the North Sea. Fisheries Research, 267, 106822..

Vasilakopoulos P., Konrad Cy., Boschetti S.T., Palialexis A., Marine Strategy Framework Directive, Review and analysis of Member States' 2018 reports. Descriptor 3: Commercial species, EUR 30660 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-79-34175-8, doi:10.2760/40557, JRC124746

Vasilakopoulos, P., Palialexis, A., Boschetti, S.T., Cardoso, A.C., Druon, J.-N., Konrad, C., Kotta, M., Magliozzi, C., Palma, M., Piroddi, C., Ruiz-Orejón, L.F., Salas-Herrero, F., Stips, A., Tornero, V. and Hanke, G., Marine Strategy Framework Directive, Thresholds for MSFD Criteria: state of play and next steps, EUR 31131 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92- 76-53689-5, doi:10.2760/640026, JRC128344.

Vasilakopoulos, P., Jardim, E., Konrad, C., Rihan, D., Mannini, A., Pinto, C., Casey, J., Mosqueira, I., O'Neill, F.G., 2020. Selectivity metrics for fisheries management and advice. Fish and Fisheries 21, 621–638.

Vasilakopoulos, P., O'Neill, F. G., & Marshall, C. T. (2016). The unfulfilled potential of fisheries selectivity to promote sustainability. Fish and Fisheries, 17, 399–416. https://doi.org/10.1111/faf.12117

Vasilakopoulos, P., Jardim, E., Konrad, C., Rihan, D., Mannini, A., Pinto, C., Casey, J., et al. 2020. Selectivity metrics for fisheries management and advice. Fish and Fisheries, 21: 621-638.

Vasilakopoulos, P., O'Neill, F. G., & Marshall, C. T. (2011). Misspent youth: Does catching immature fish affect fisheries sustainability? *ICES Journal of Marine Science* 68, 1525–1534.

Vasilakopoulos, P., Jardim, E., Konrad, C., Rihan, D., Mannini, A., Pinto, C., Casey, J., Mosqueira, I. & O'Neill, F.G. et al. (2020). Selectivity metrics for fisheries management and advice. *Fish and Fisheries* 21, 621-638.

von Foerster, H. (1959). The Kinetics of Cellular Proliferation, F. Stohmann Jr (Ed.), New York: Grune and Stratton, pp. 382–407.

Östman, Ö., Hommik, K., Bolund, E., Heikinheimo, O., Olin, M., Lejk, A. M., Svirgsden, R., Smoliński, S., Olsson, J. (2023). Size-based indicators for assessments of ecological status of coastal fish communities. ICES Journal of Marine Science, in press.

Annex 1: List of participants

Name	Institute	Email
Anna Rindorf (Chair)	DTU	ar@aqua.dtu
Giuseppe Scarcella (Chair)	CNR	giuseppe.scarcella@cnr.it
Enrico Armelloni	National Research Council	enrico.armelloni@irbim.cnr.it
Valerio Bartolino	SLU	valerio.bartolino@slu.se
Jose Maria Bellido	IEO	josem.bellido@ieo.csic.es
Elisabeth Bolund	SLU	elisabeth.bolund@slu.se
Victoria Campon-Linares	CEFAS	victoria.linares@cefas.gov.uk
Gema Canal	IEO	gema.canal@ieo.csic.es
Ilaria Coscia	Marine Institute Ireland	Ilaria.Coscia@marine.ie
Elizabeth Duskey		eduskey@okstate.edu
Tomaso Fortibuoni	ISPRA	tomaso.foribuoni@isprambiente.it
Eva Garnacho	CEFAS	eva.garnacho@cefas.gov.uk
Patracia Gonçalves	IPMA	patricia@ipma.pt
Chris Griffiths	SLU	christopher.griffiths@slu.se
Daniel Howell	Institute of Marine Research Norway	Daniel.howell@hi.no
Nis Jacobsen	DTU	nsja@aqua.dtu.dk
Peter Kidd	CEFAS	peter.Kidd@cefas.gov.uk
Bernhard Kuehn	Thuenen Institute	bernhard.kuehn@thuenen.de
Susana Junquera	IEO	susana.junquera@ieo.csic.es
Stefano Moro	SZN	stefano.moro@szn.it
Natasha Murphy	Marine Institute Ireland	natasha.murphy@marine.ie
Nikolaus Probst	Thuenen Institute	nikolaus.probst@thuenen.de
Owen Rowe	HELCOM	Owen.Rowe@helcom.fi
Sonia Seixas	MARE	Sonia.seixas@uab.pt
Jessica Tengvall	University of Bergen	Jessica.Tengvall@uib.no

Name	Institute	Email
Amina Tifoura	IEO	Amina.tifoura@ieo.csic.es
Ching Maria Villanueva	IFREMER	Ching.Villanueva@ifremer.fr
Håkan Wennhage	SLU	hakan.wennhage@slu.se
Josephine Woronoff	PEW	jworonoff@pewtrusts.org
Nuria Zaragoza	IEO	nuria.zaragoza@ieo.csic.es
Tomas Zolubas	Marine Research Institute	tomas.zolubas@ku.lt
Barbara Zorica	IZOR	zorica@izor.hr

Annex 2: Stock specific plots of proposed D3C3 indicators alongside metrics of stock status (SSB, R and F)

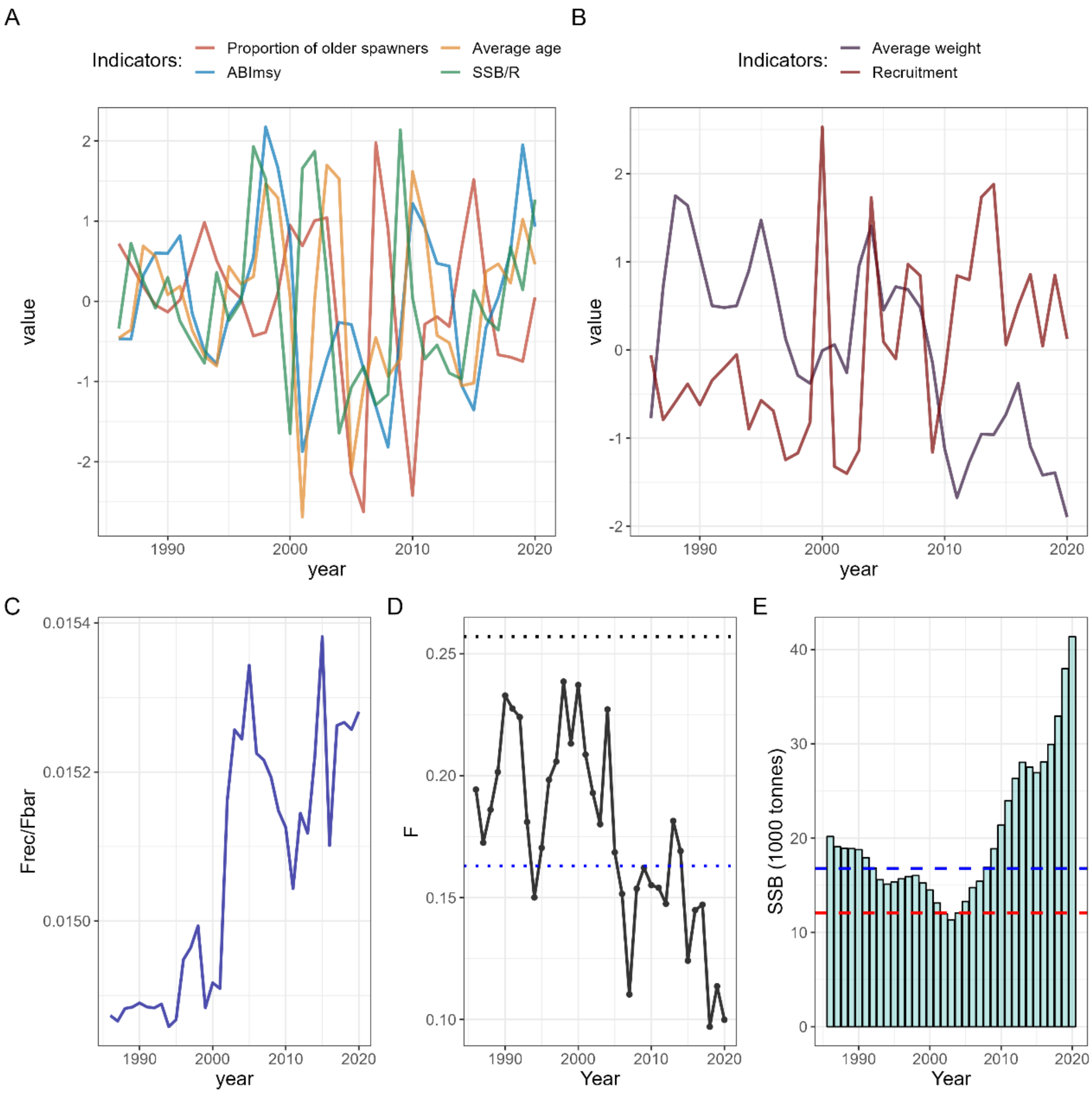


Figure A2.1. Estimated D3C3 indicators and stock status for Black-bellied anglerfish in Subarea 7 and Divisions 8.a-b and 8.d (ank.27.78.abd). Time series of the indicators are shown in panels A-C, as well as F (panel D) and SSB (panel E). Indicators in panels A and B have been normalised by subtracting the mean and dividing by the standard deviation, thus allowing them to be plotted and compared on the same scale. All F and SSB values are taken directly from each stocks stock assessment. ICES reference points in terms of F and SSB are also plotted when available, specifically B_{lim} (red dashed) and MSY $B_{trigger}$ (blue dashed), F_{lim} (red dotted), F_{pa} (black dotted) and F_{MSY} (blue dotted).

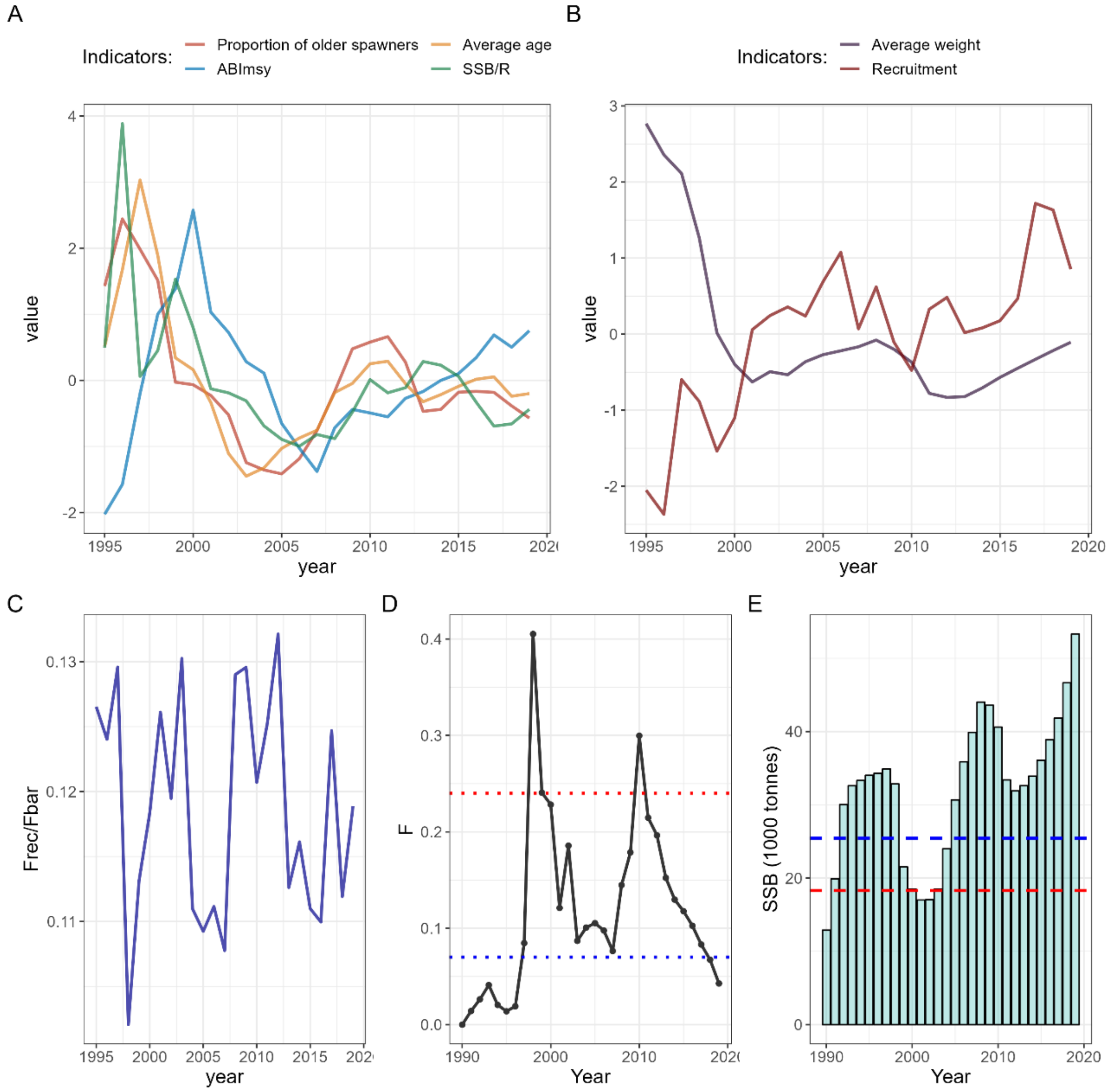


Figure A2.2. Estimated D3C3 indicators and stock status for Greater silver smelt (aru.27.5a14). See Figure legend A2.1 for further details.

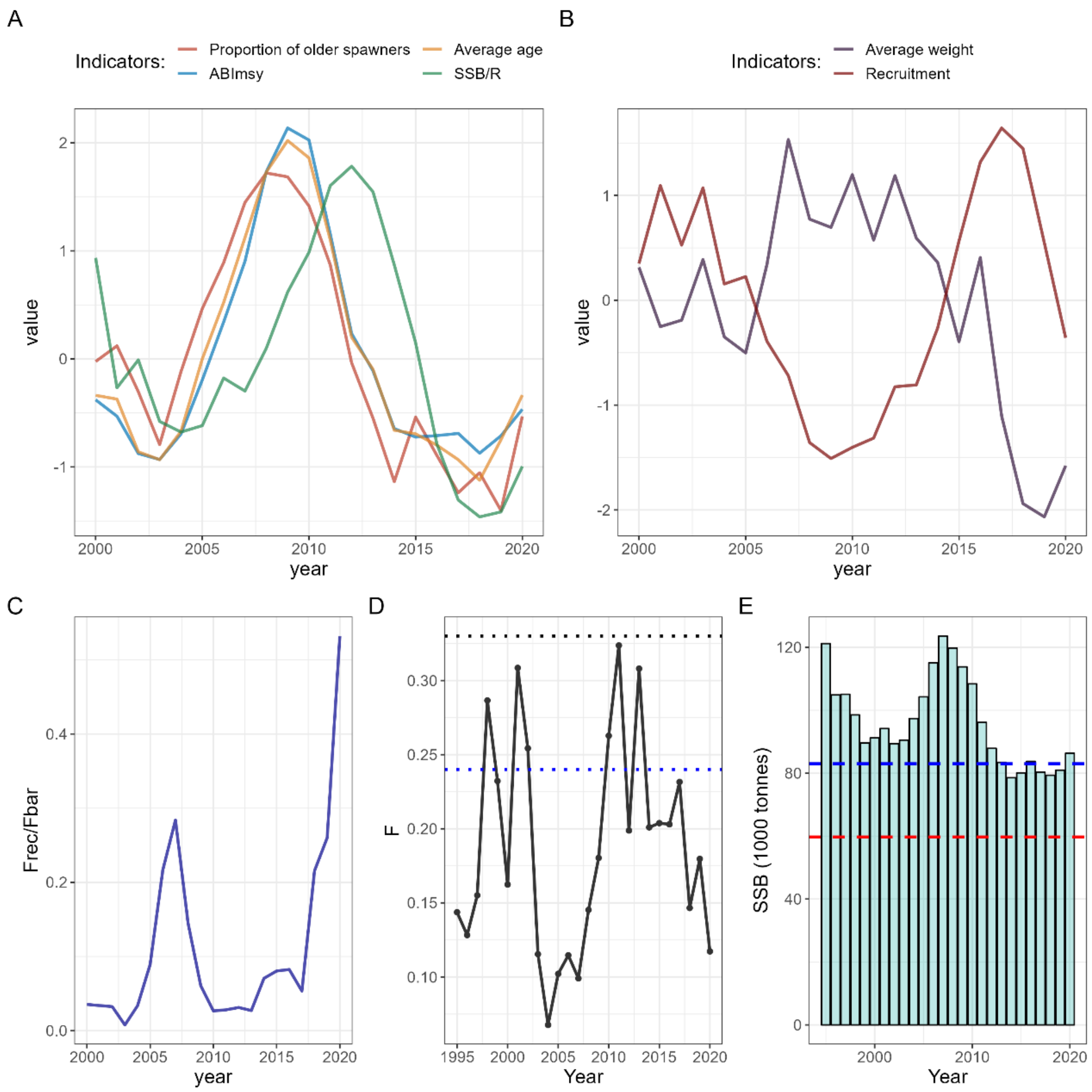
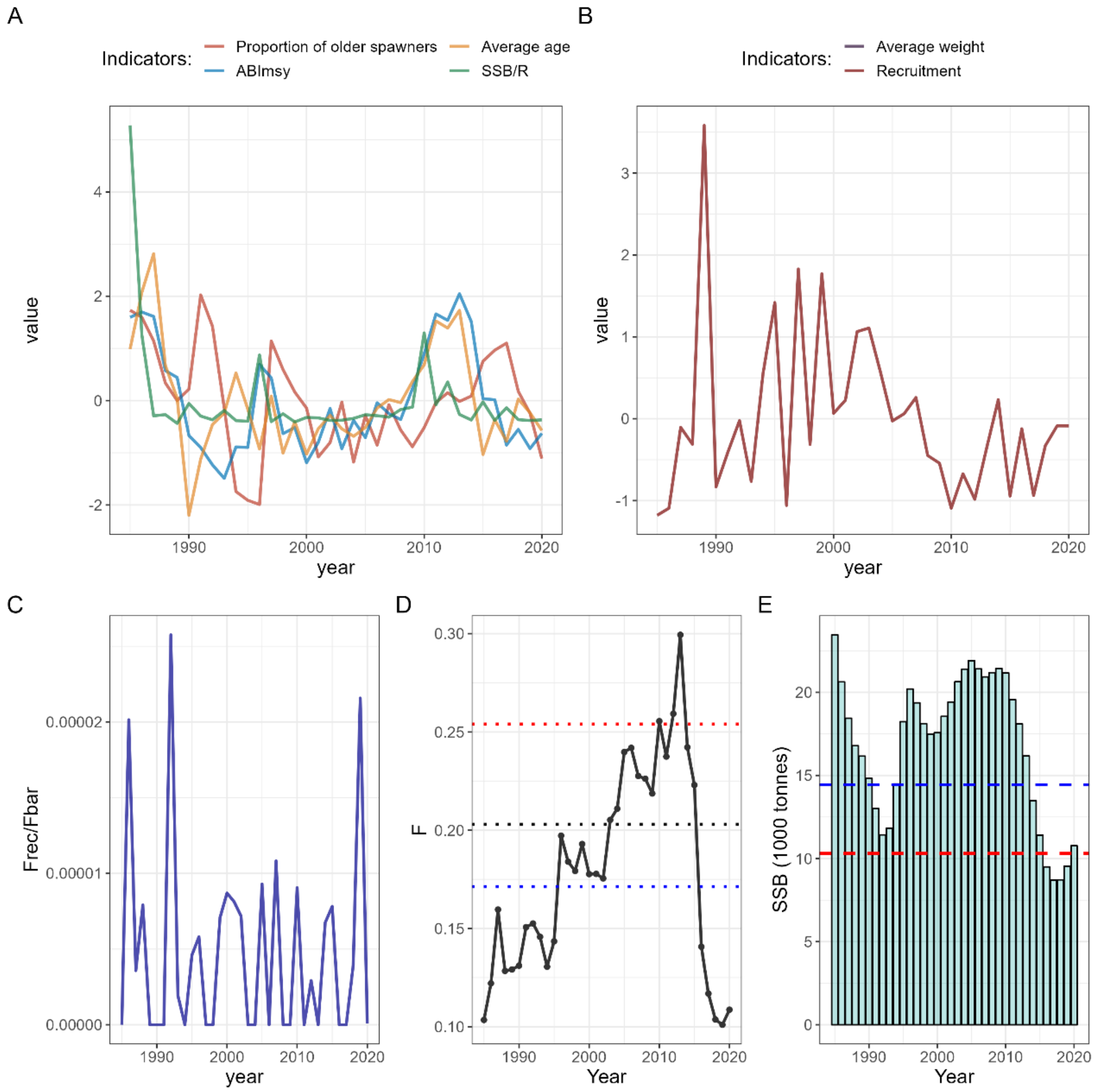


Figure A2.3. Estimated D3C3 indicators and stock status for Greater silver smelt (aru.27.5b6a). See Figure legend A2.1 for further details.



,

Figure A2.4. Estimated D3C3 indicators and stock status for Sea bass (bss.27.4bc7ad-h). See Figure legend A2.1 for further details.

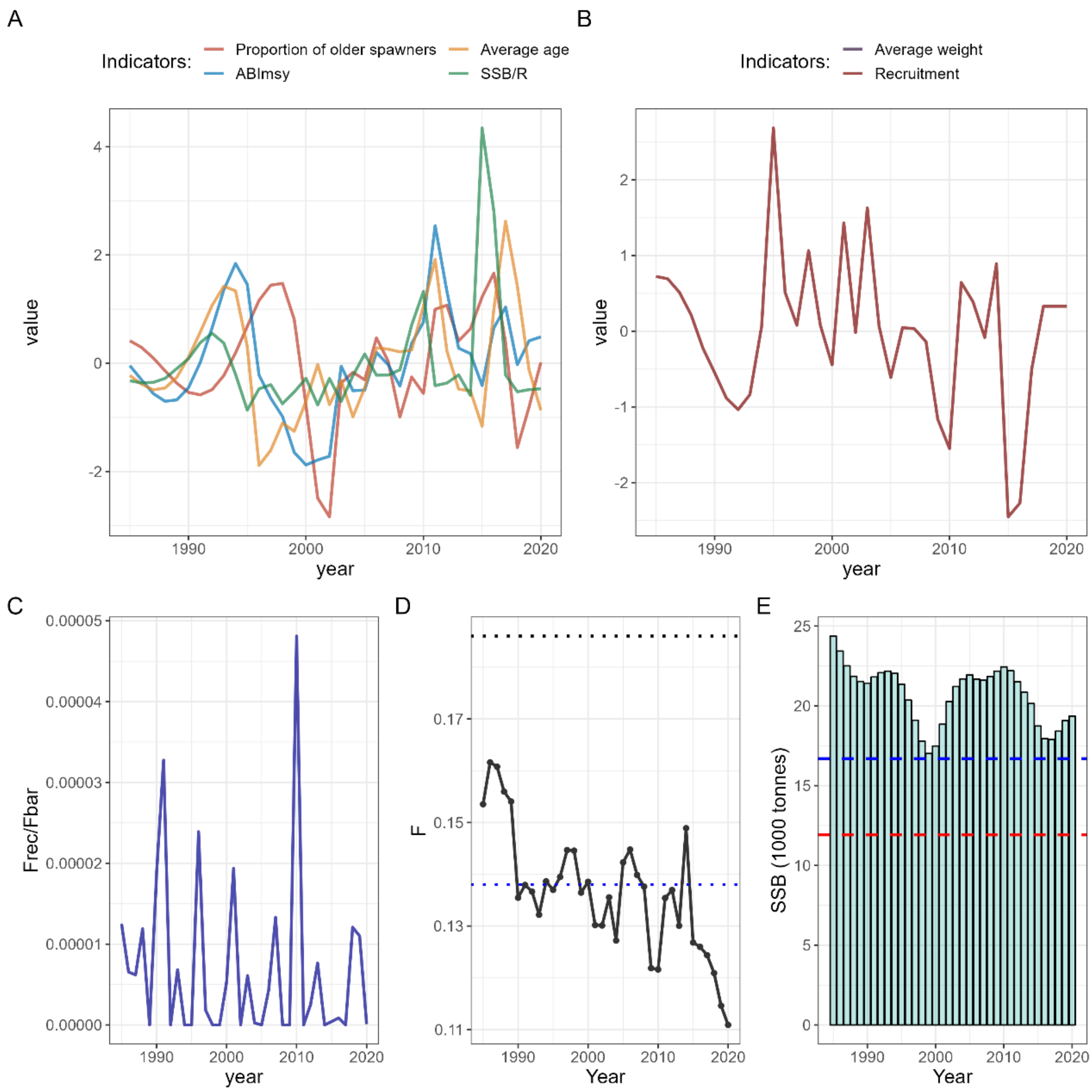


Figure A2.5. Estimated D3C3 indicators and stock status for Sea bass (bss.27.8ab). See Figure legend A2.1 for further details.

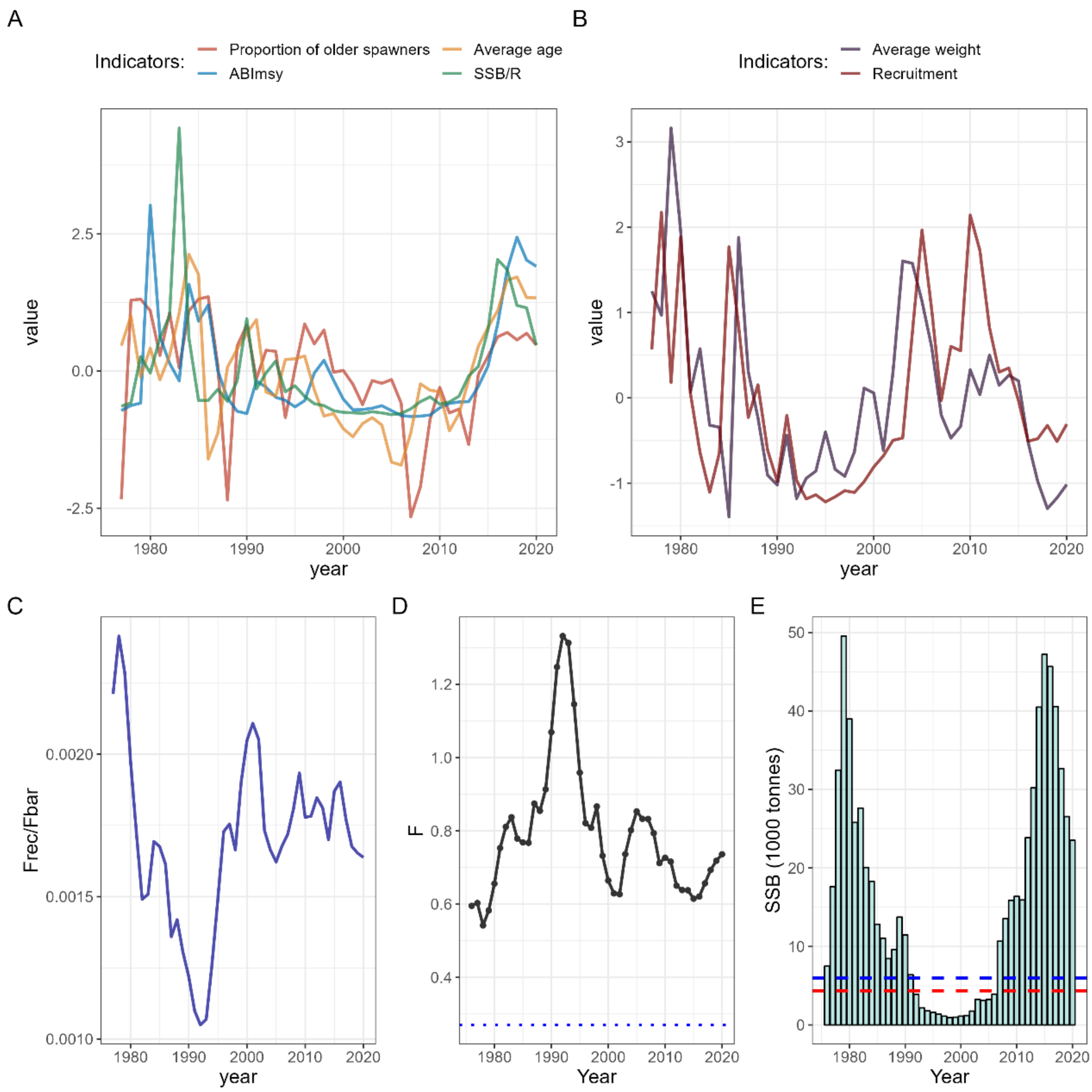


Figure A2.6. Estimated D3C3 indicators and stock status for Atlantic cod (cod.21.1). See Figure legend A2.1 for further details.

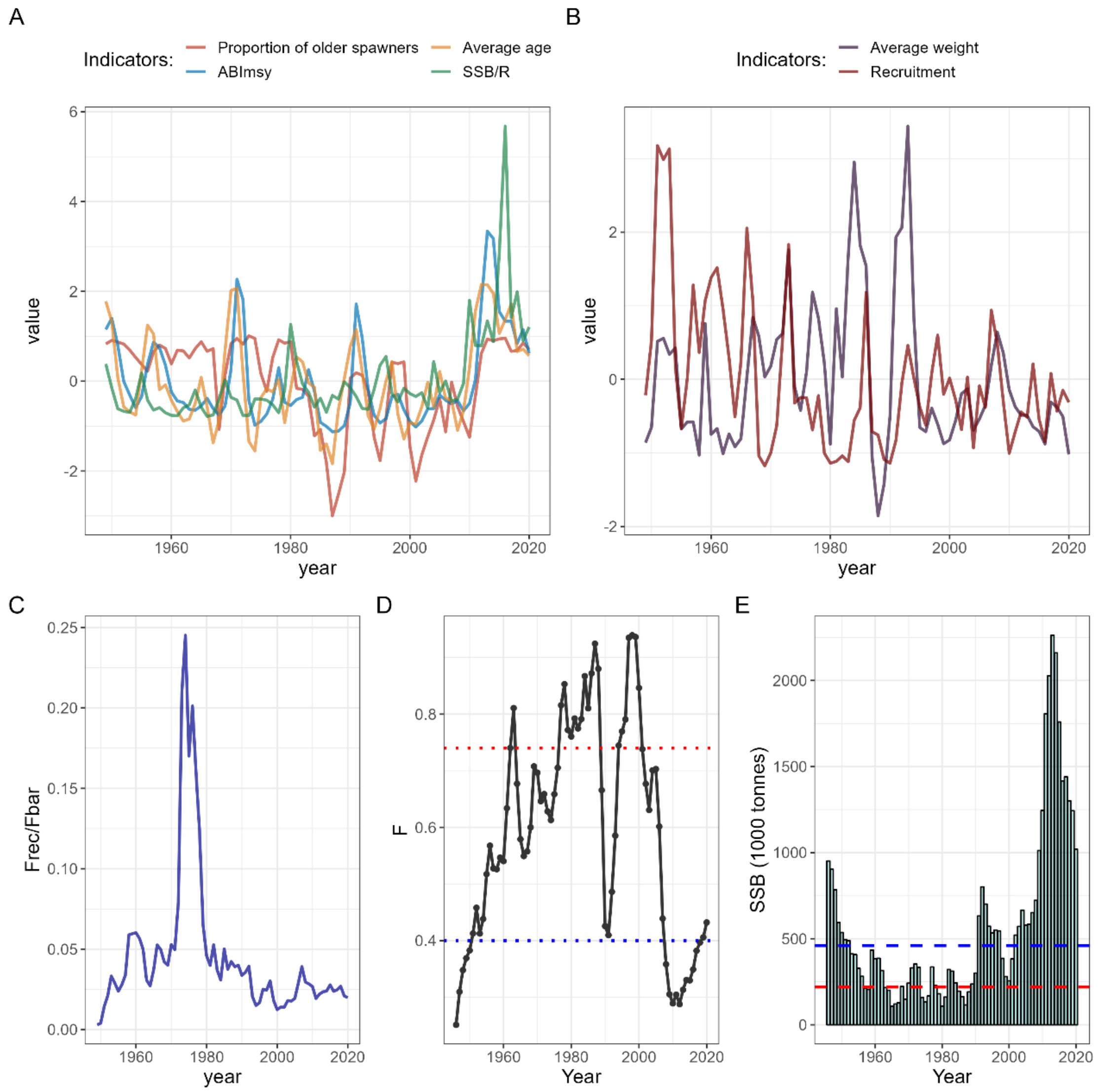


Figure A2.7. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.1-2). See Figure legend A2.1 for further details.

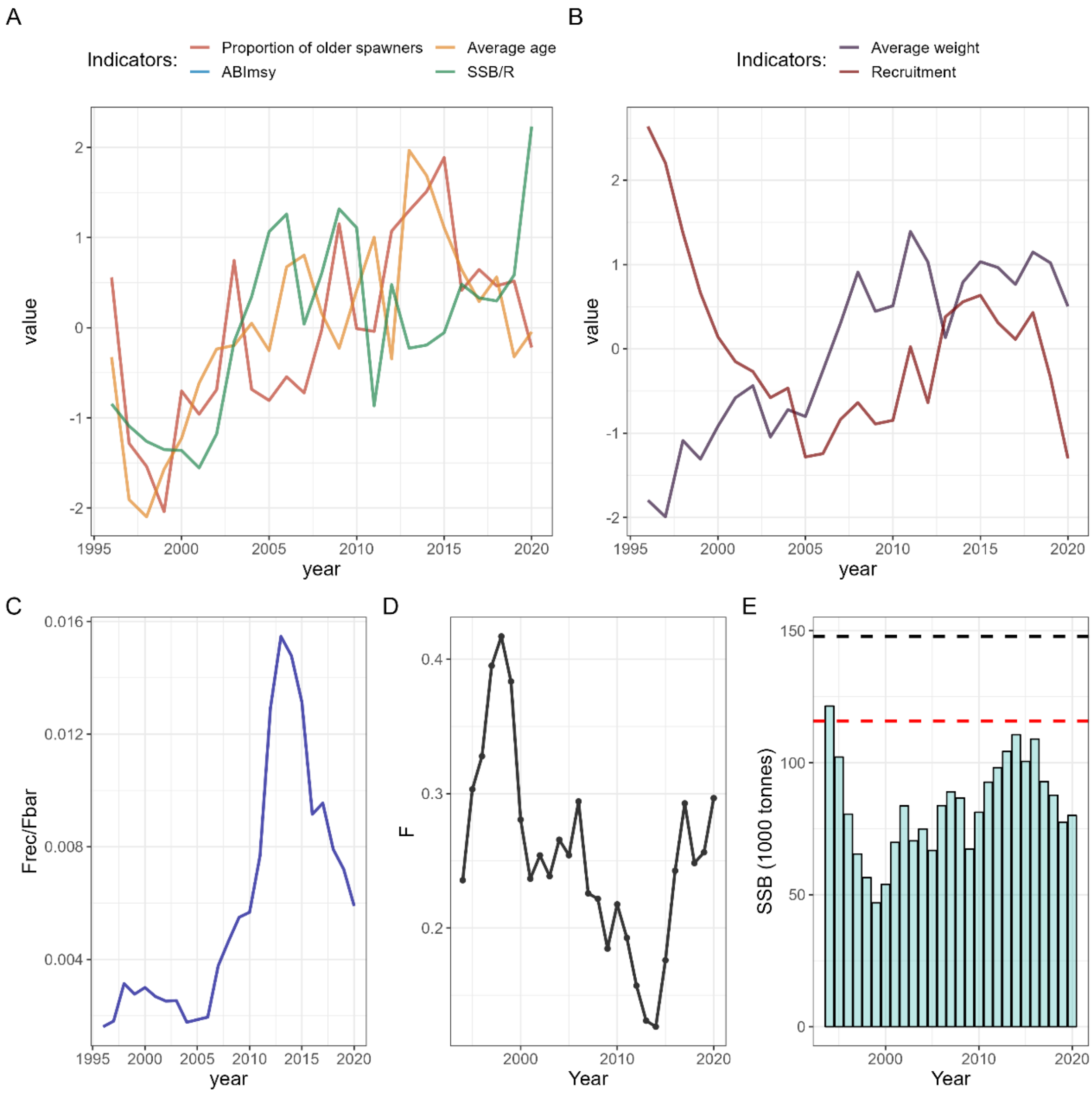


Figure A2.8. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.1-2coastN). See Figure legend A2.1 for further details.

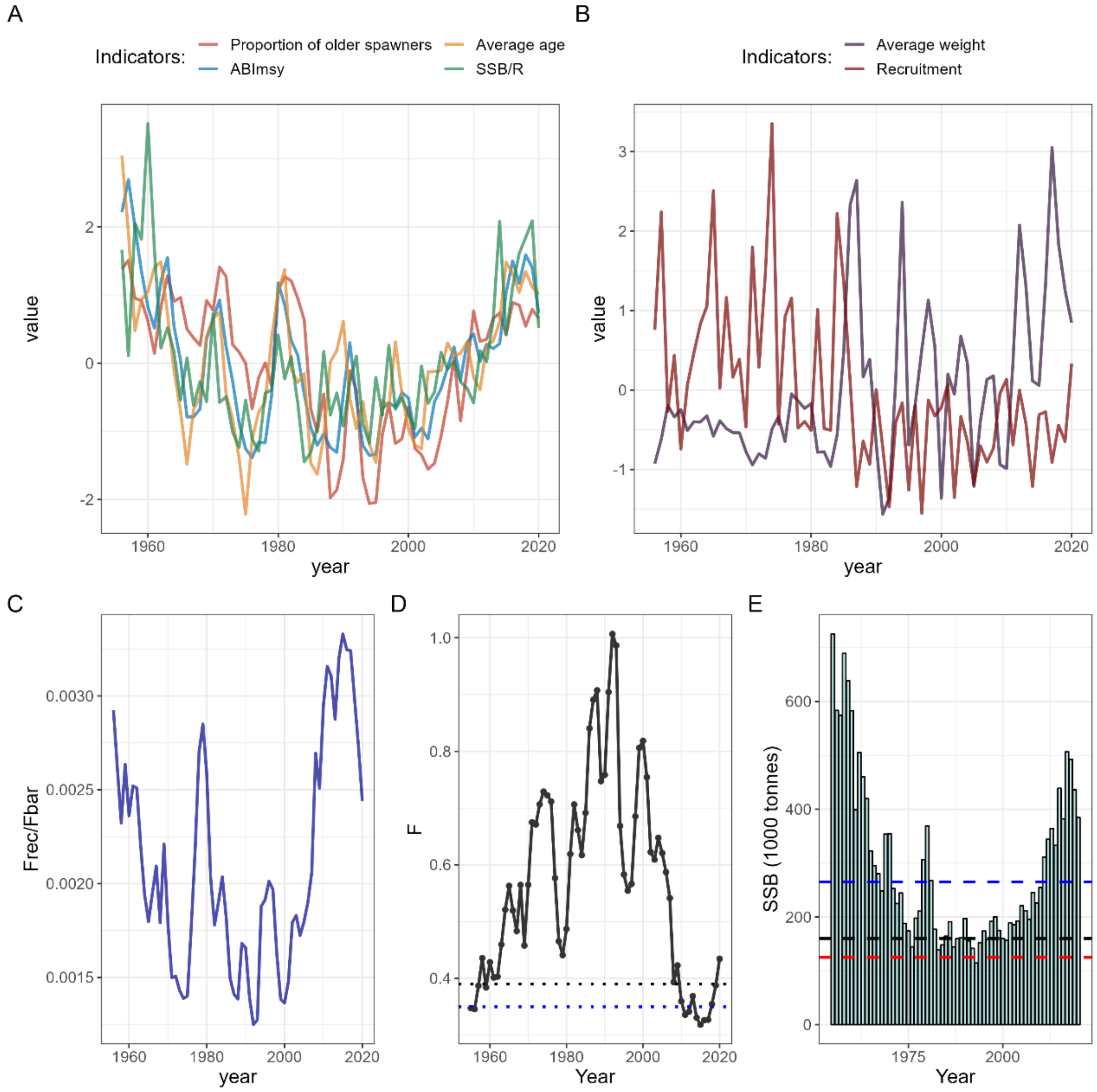


Figure A2.9. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.5a). See Figure legend A2.1 for further details.

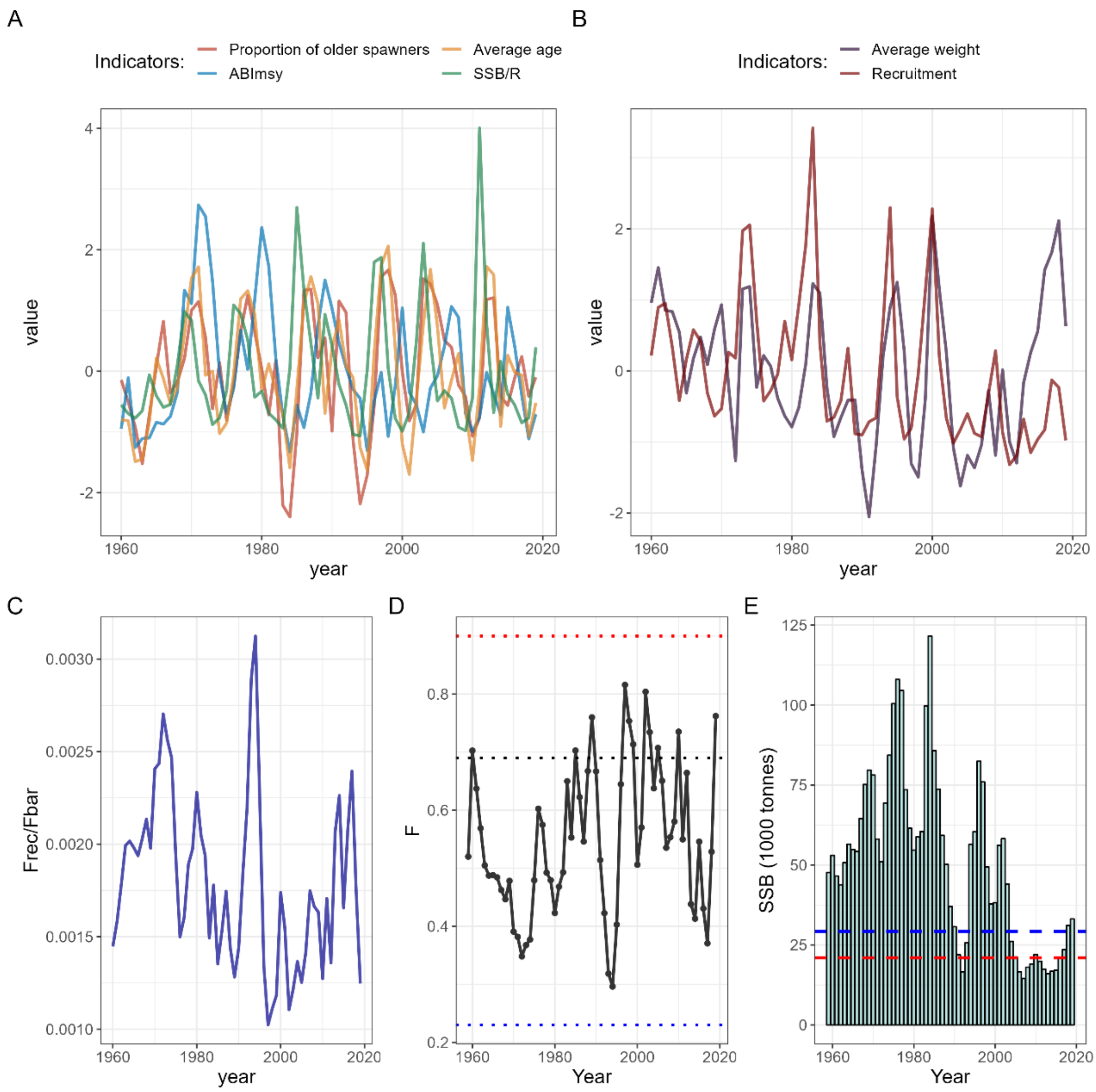


Figure A2.10. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.5b1). See Figure legend A2.1 for further details.

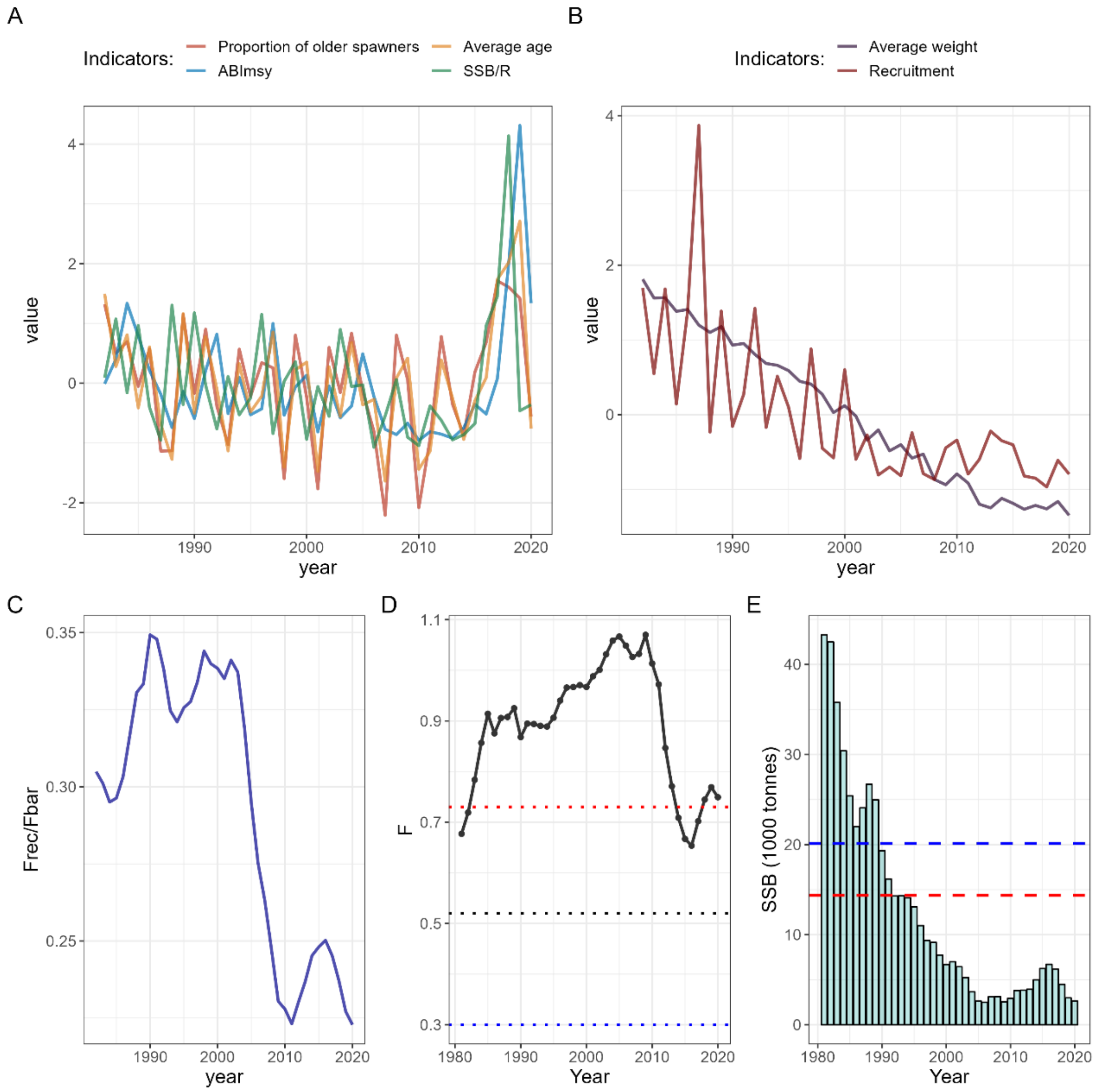


Figure A2.11. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.6a). See Figure legend A2.1 for further details.

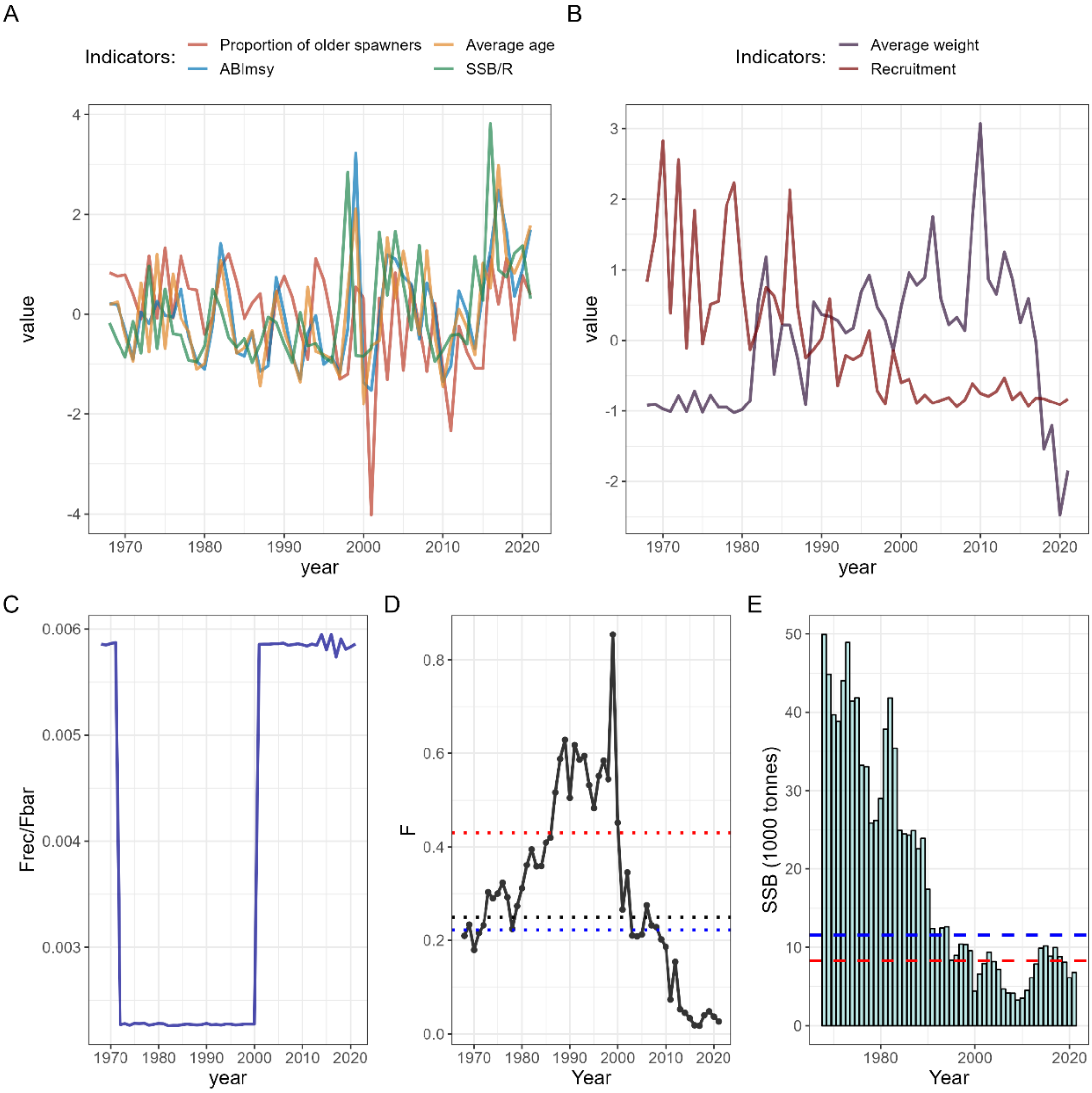


Figure A2.12. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.7a). See Figure legend A2.1 for further details.

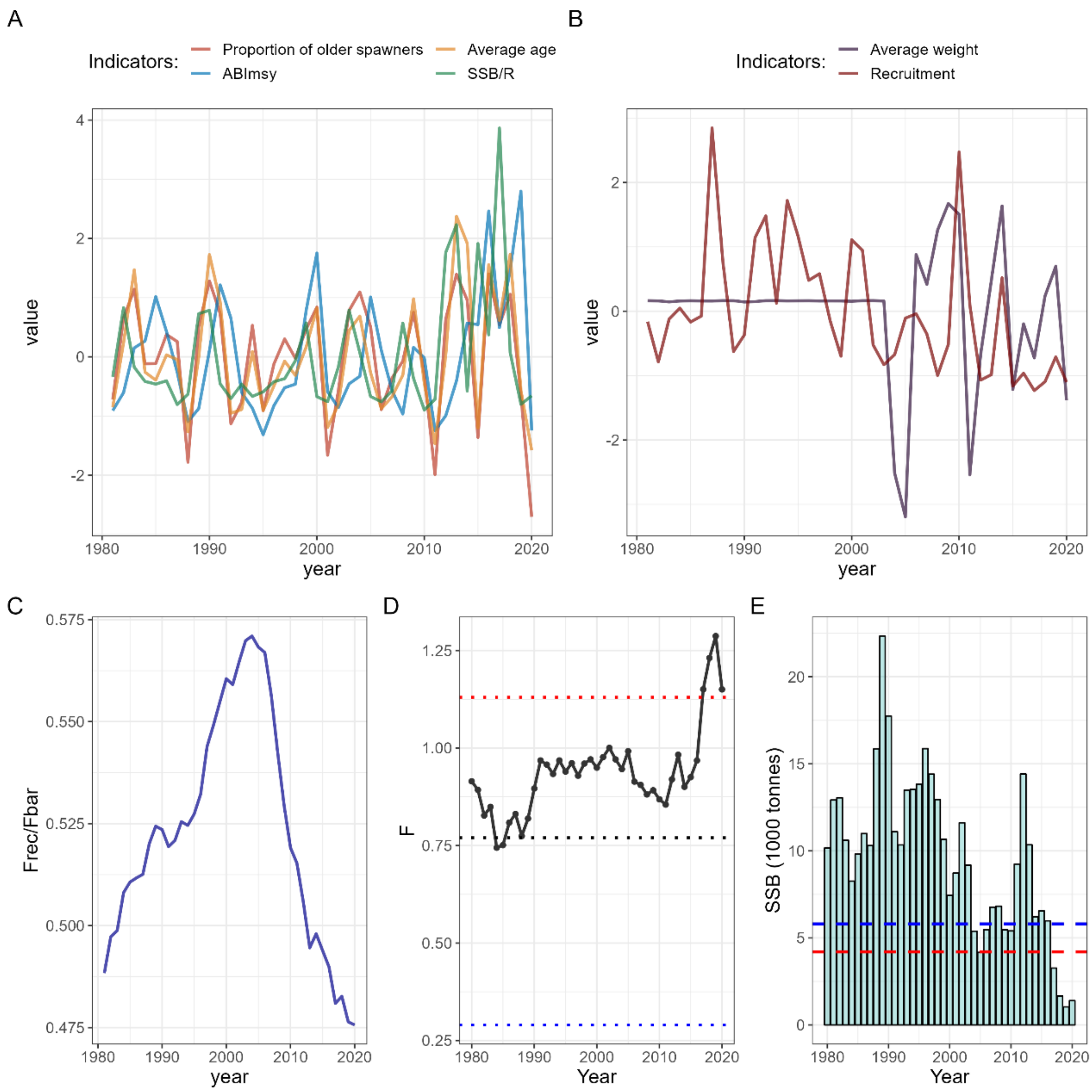


Figure A2.13. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.7e-k). See Figure legend A2.1 for further details.

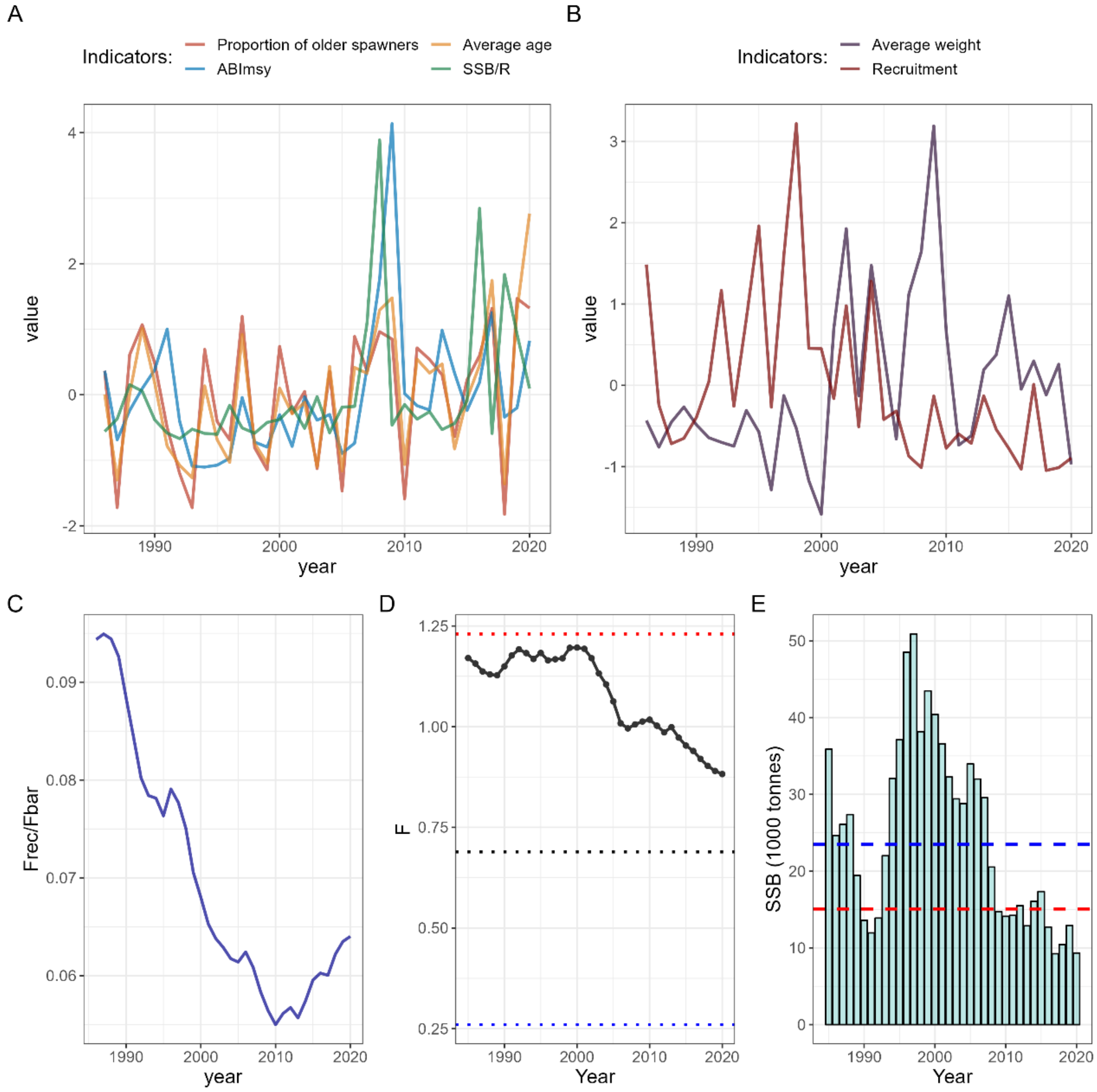


Figure A2.14. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.22-24). See Figure legend A2.1 for further details.

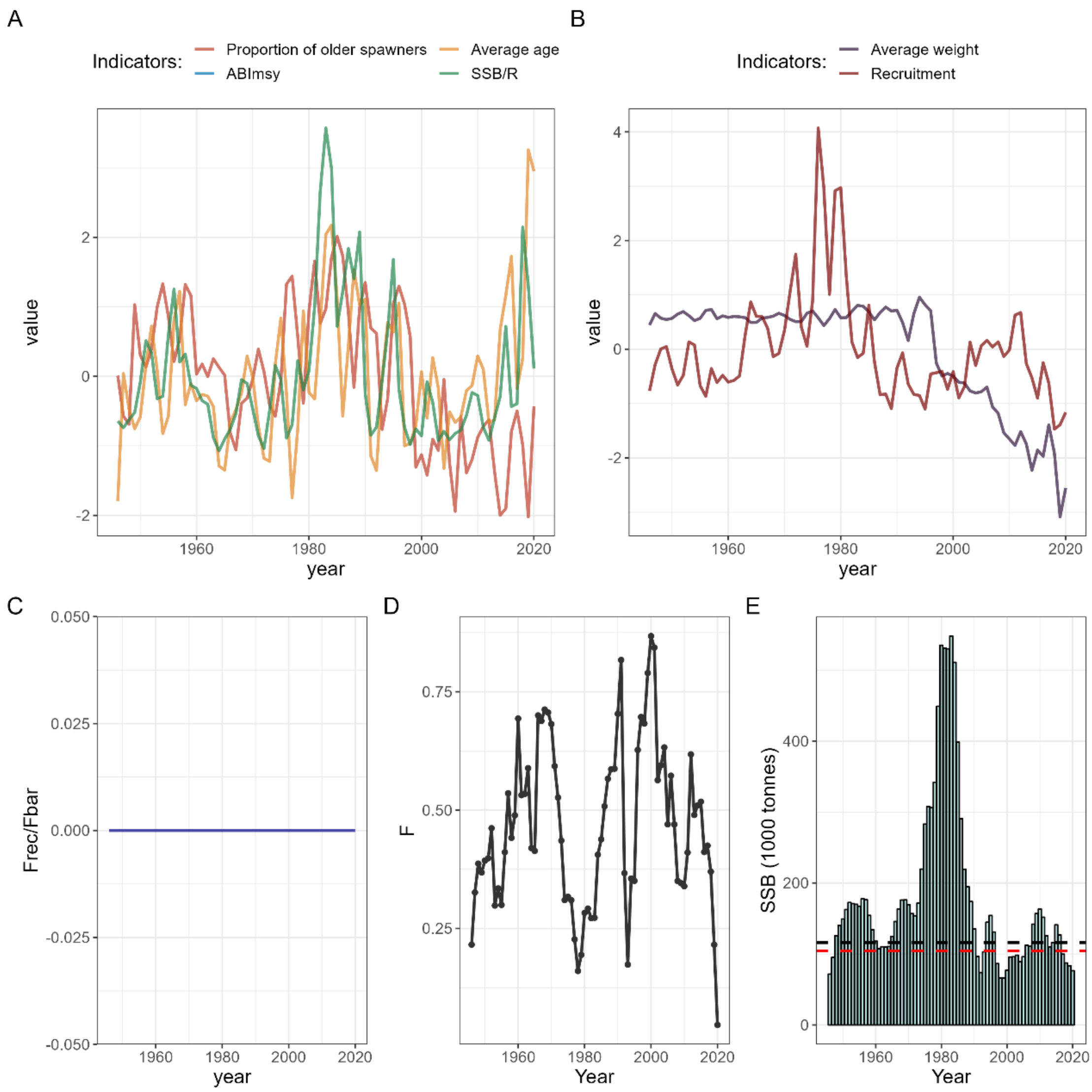


Figure A2.15. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.24-32). See Figure legend A2.1 for further details.

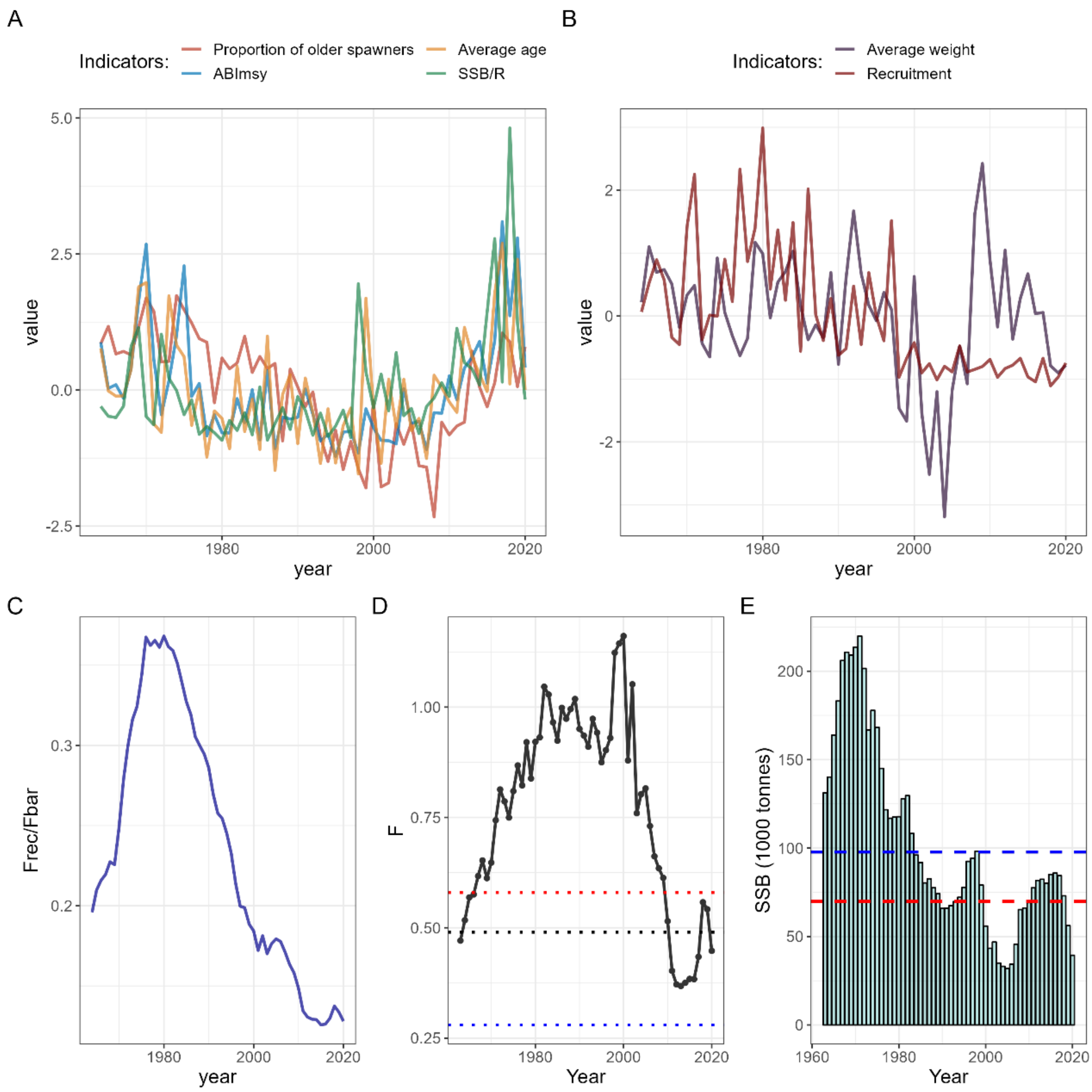


Figure A2.16. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.47d20). See Figure legend A2.1 for further details.

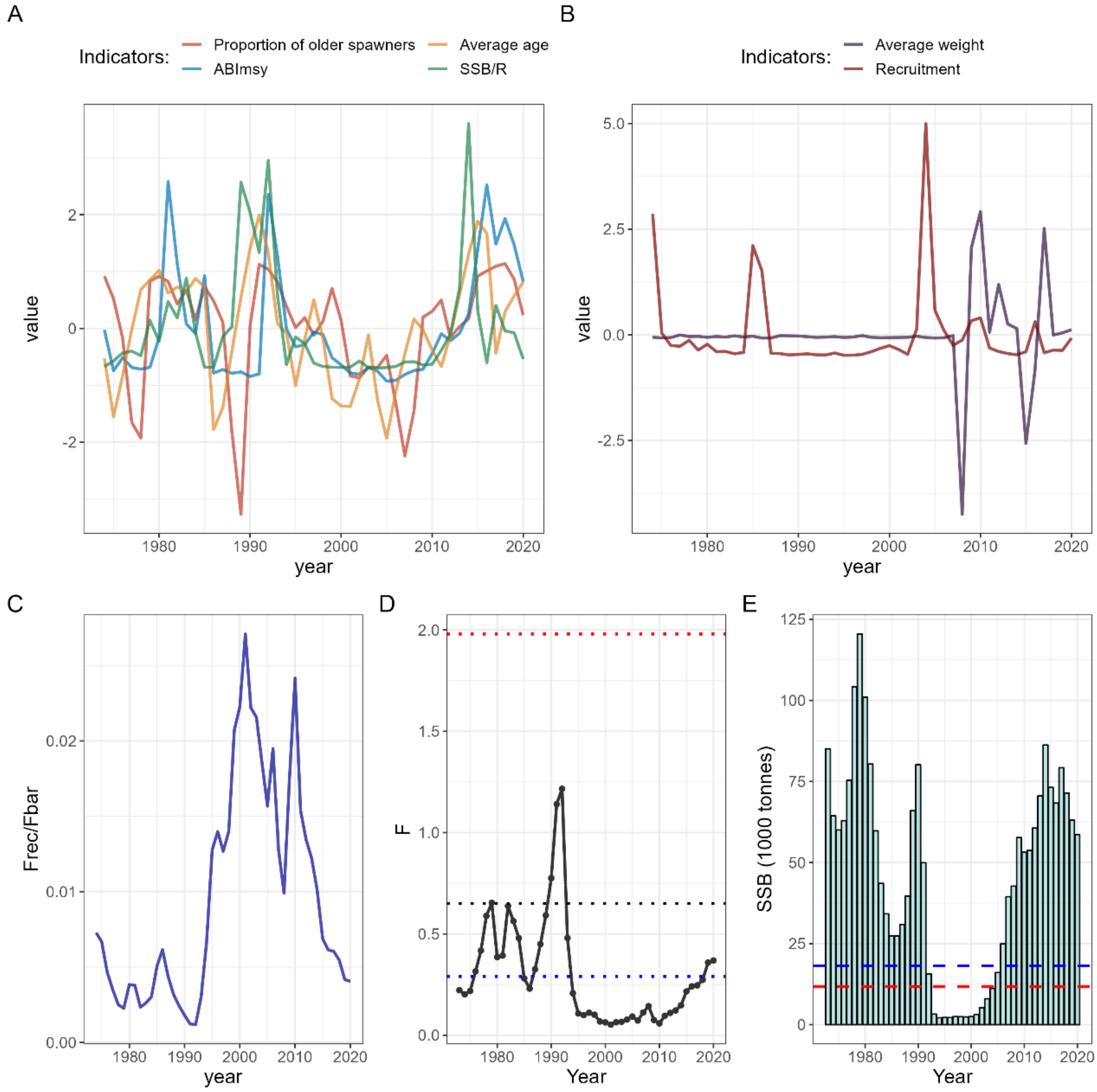


Figure A2.17. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.2127.1f14). See Figure legend A2.1 for further details.

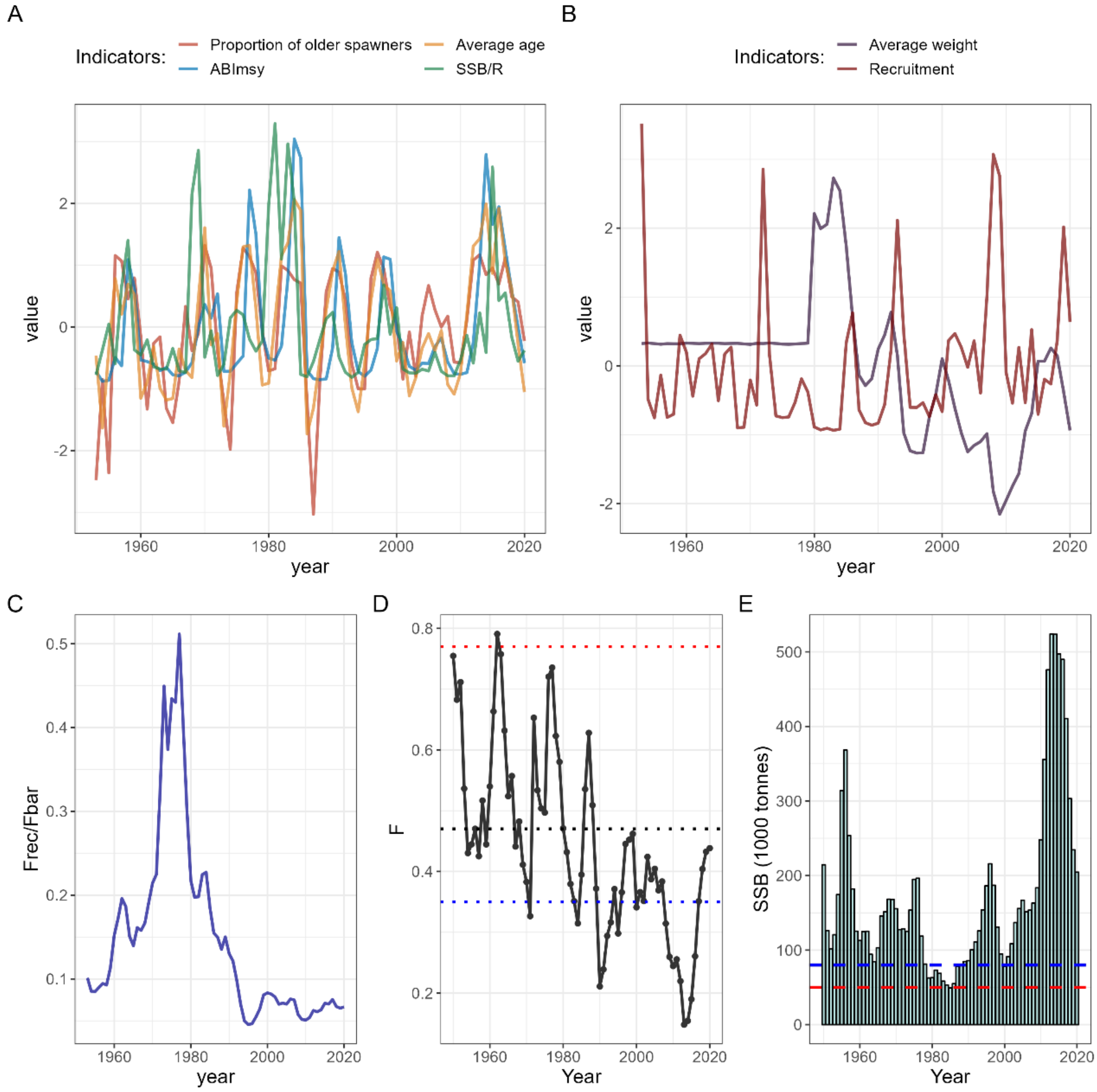


Figure A2.18. Estimated D3C3 indicators and stock status for Haddock (had.27.1-2). See Figure legend A2.1 for further details.

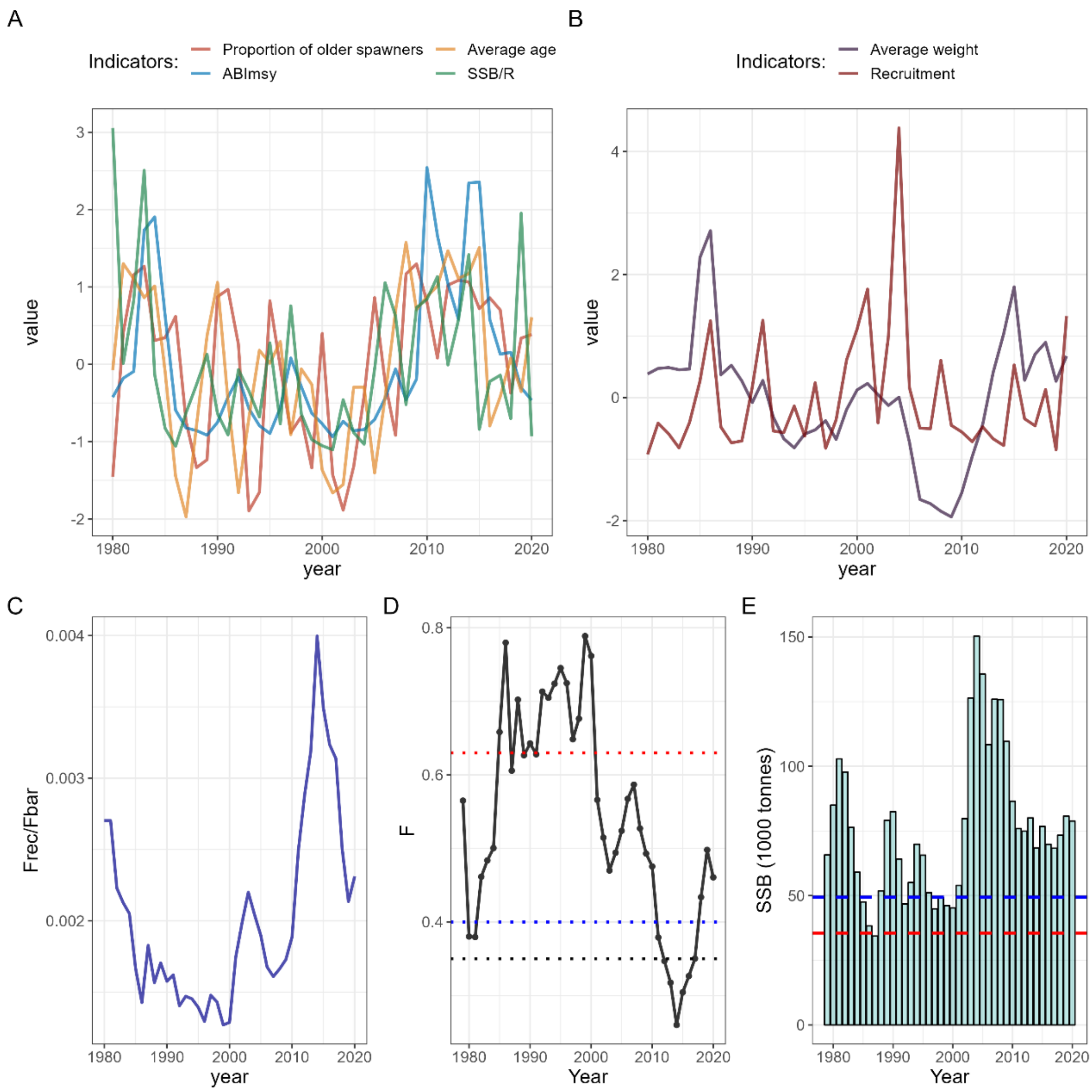


Figure A2.19. Estimated D3C3 indicators and stock status for Haddock (had.27.5a). See Figure legend A2.1 for further details.

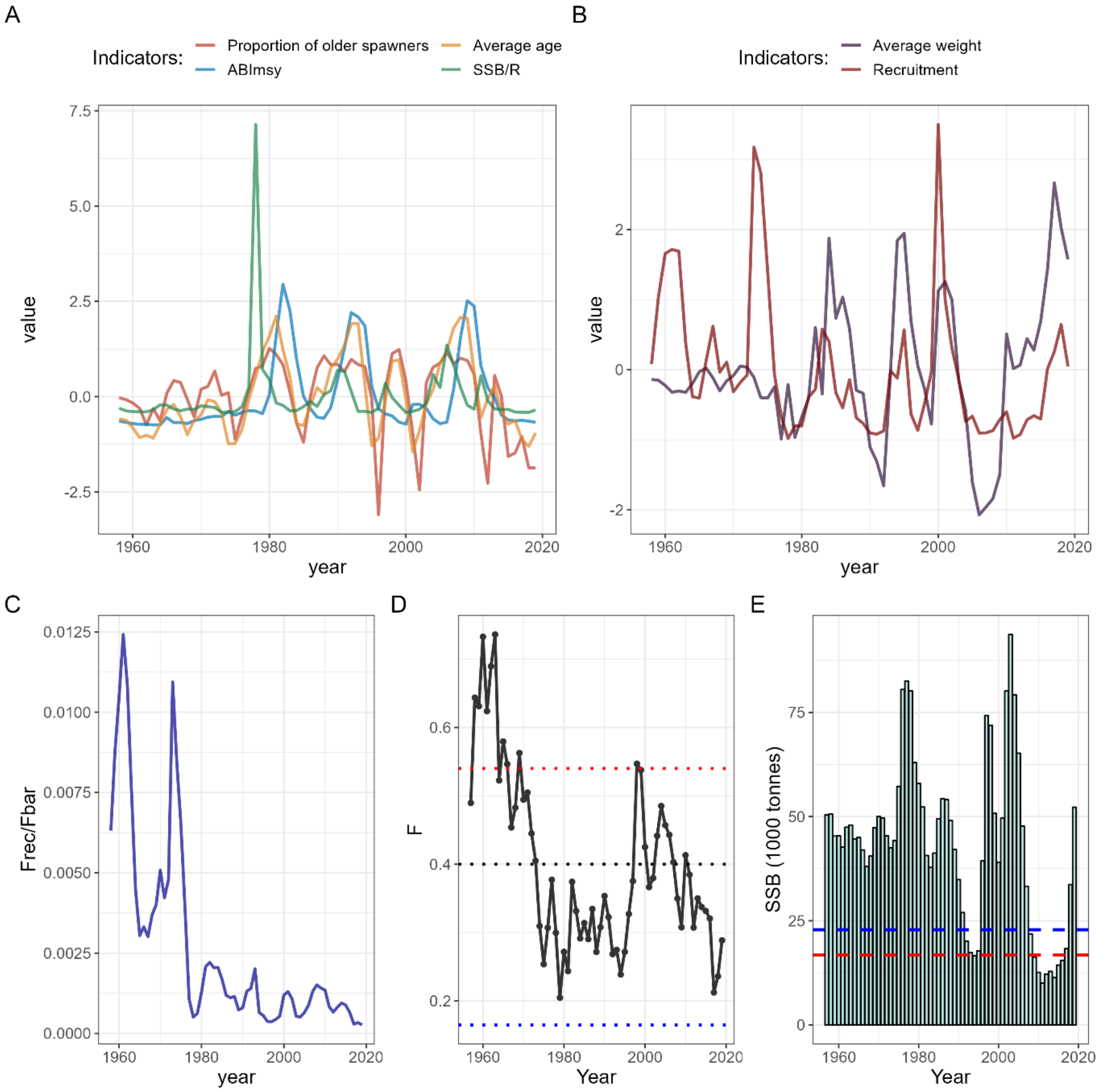


Figure A2.20. Estimated D3C3 indicators and stock status for Haddock (had.27.5b). See Figure legend A2.1 for further details.

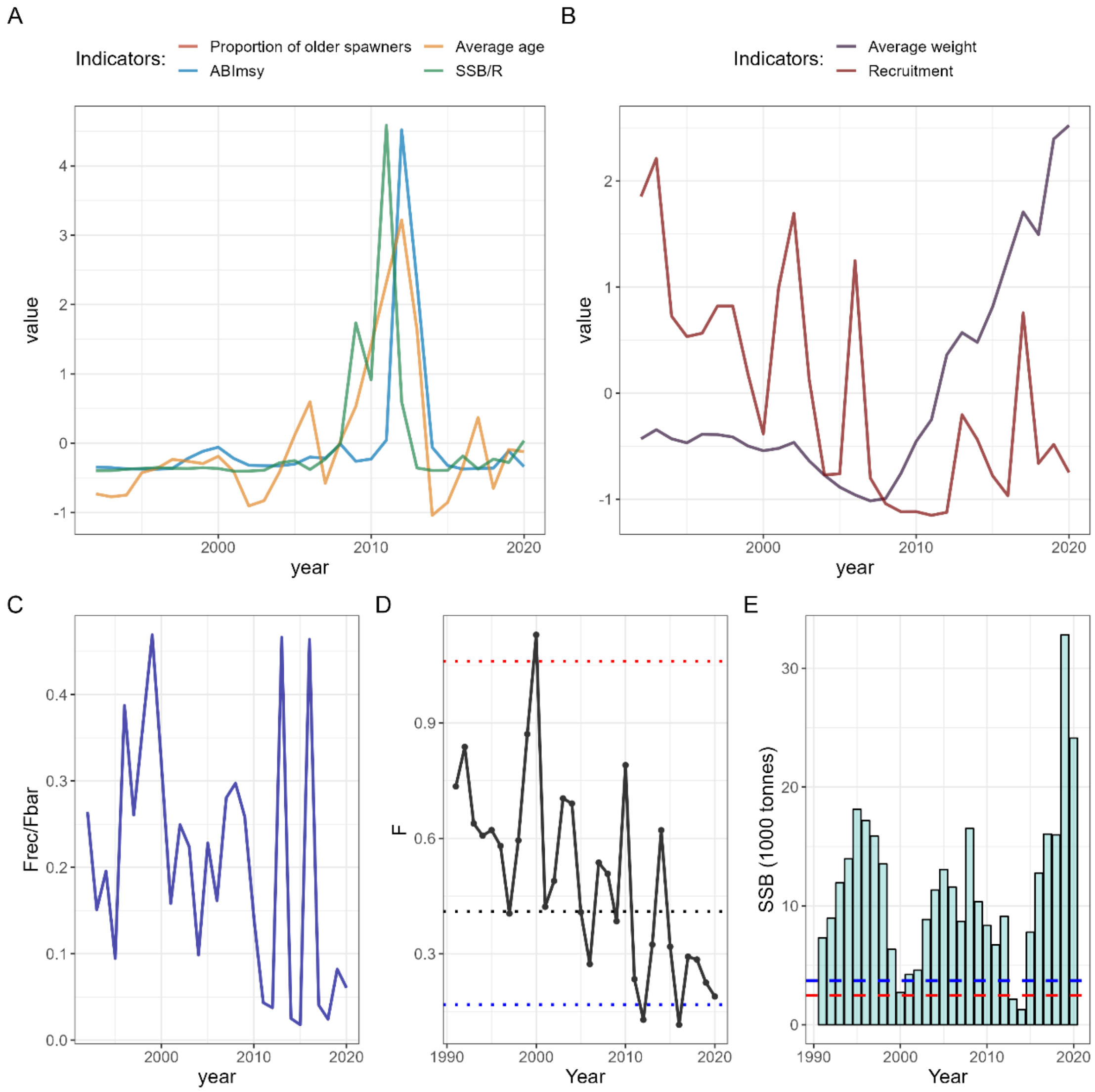


Figure A2.21. Estimated D3C3 indicators and stock status for Haddock (had.27.6b). See Figure legend A2.1 for further details.

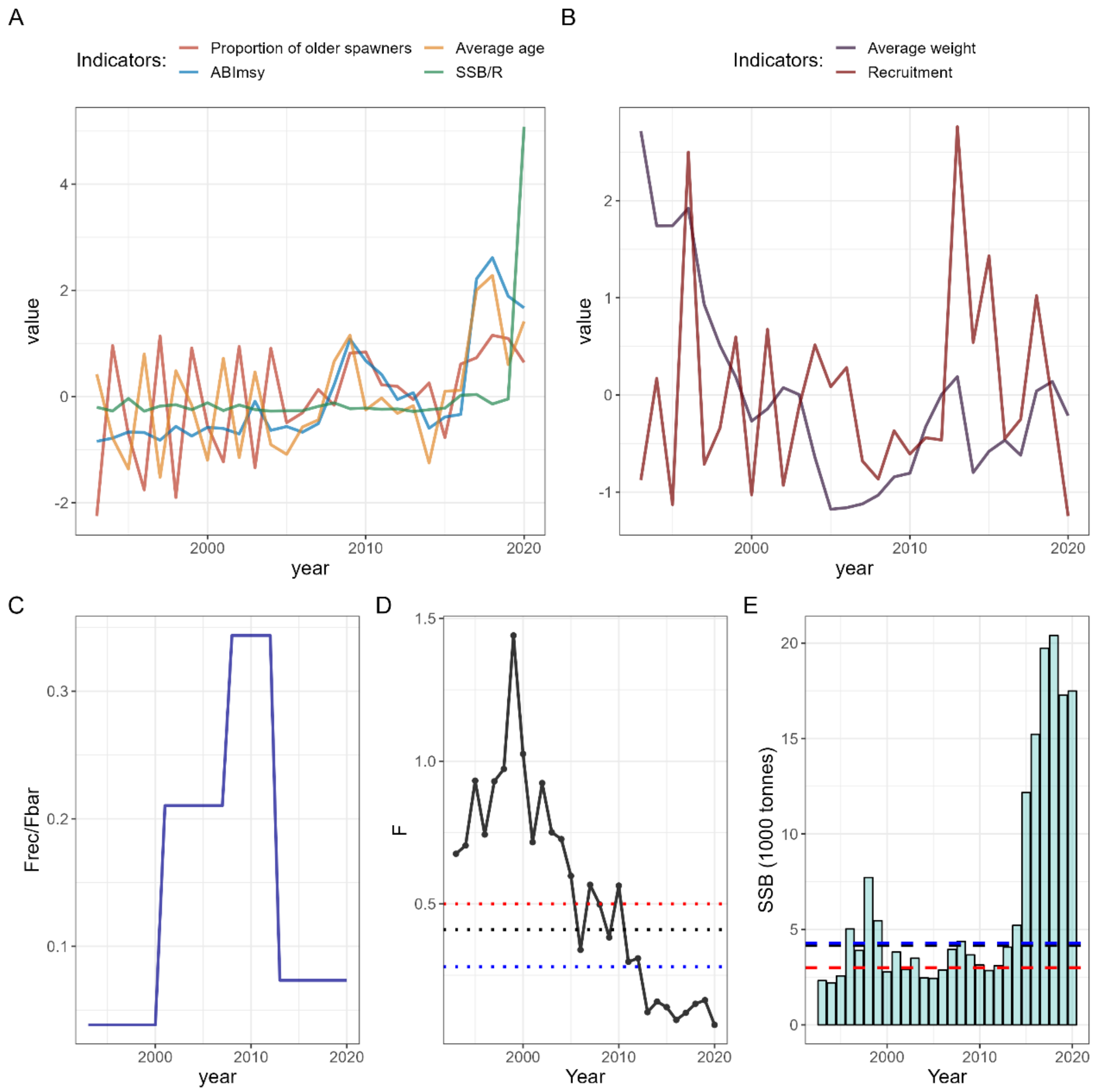


Figure A2.22. Estimated D3C3 indicators and stock status for Haddock (had.27.7a). See Figure legend A2.1 for further details.

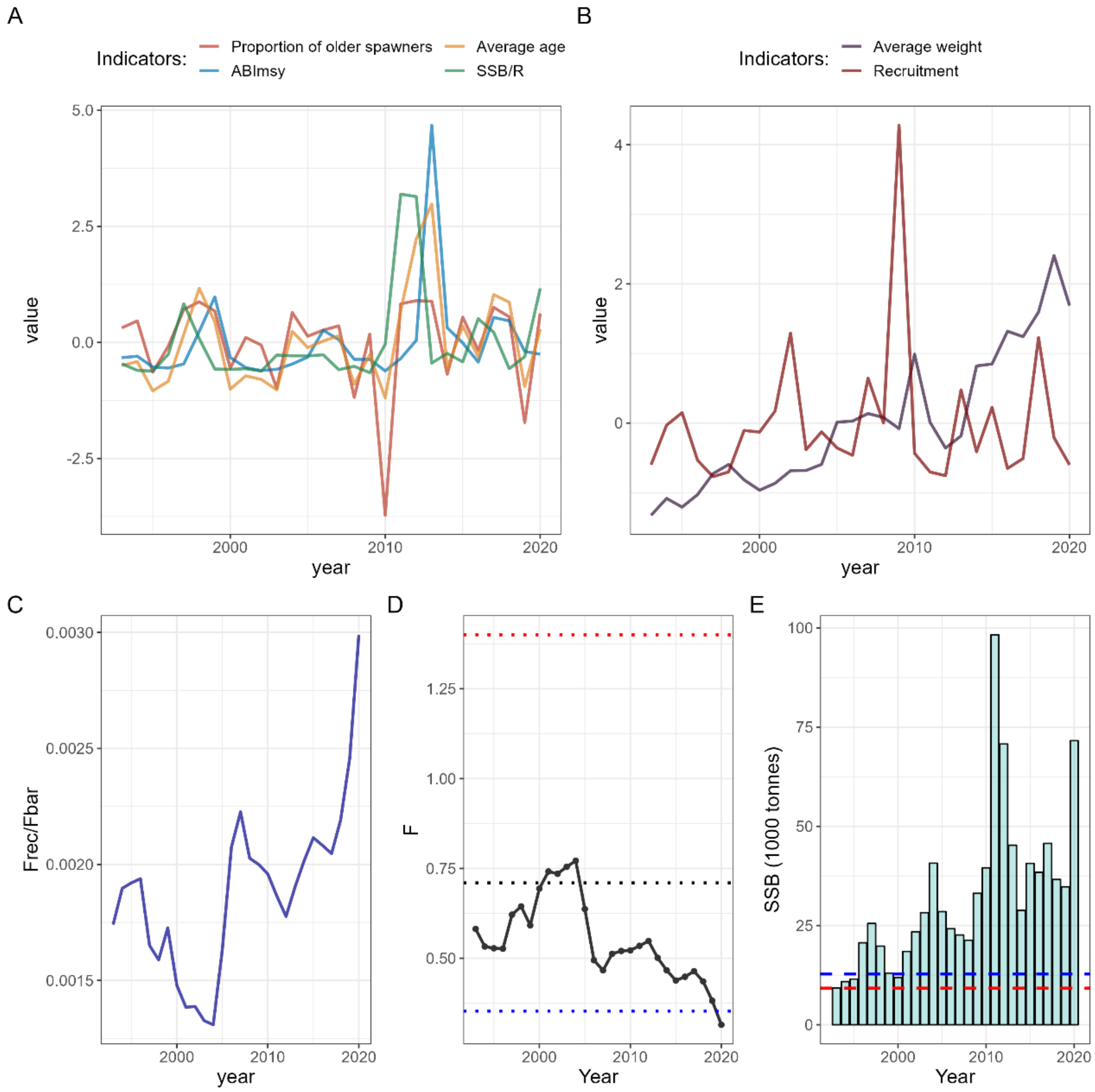


Figure A2.23. Estimated D3C3 indicators and stock status for Haddock (had.27.7b-k). See Figure legend A2.1 for further details.

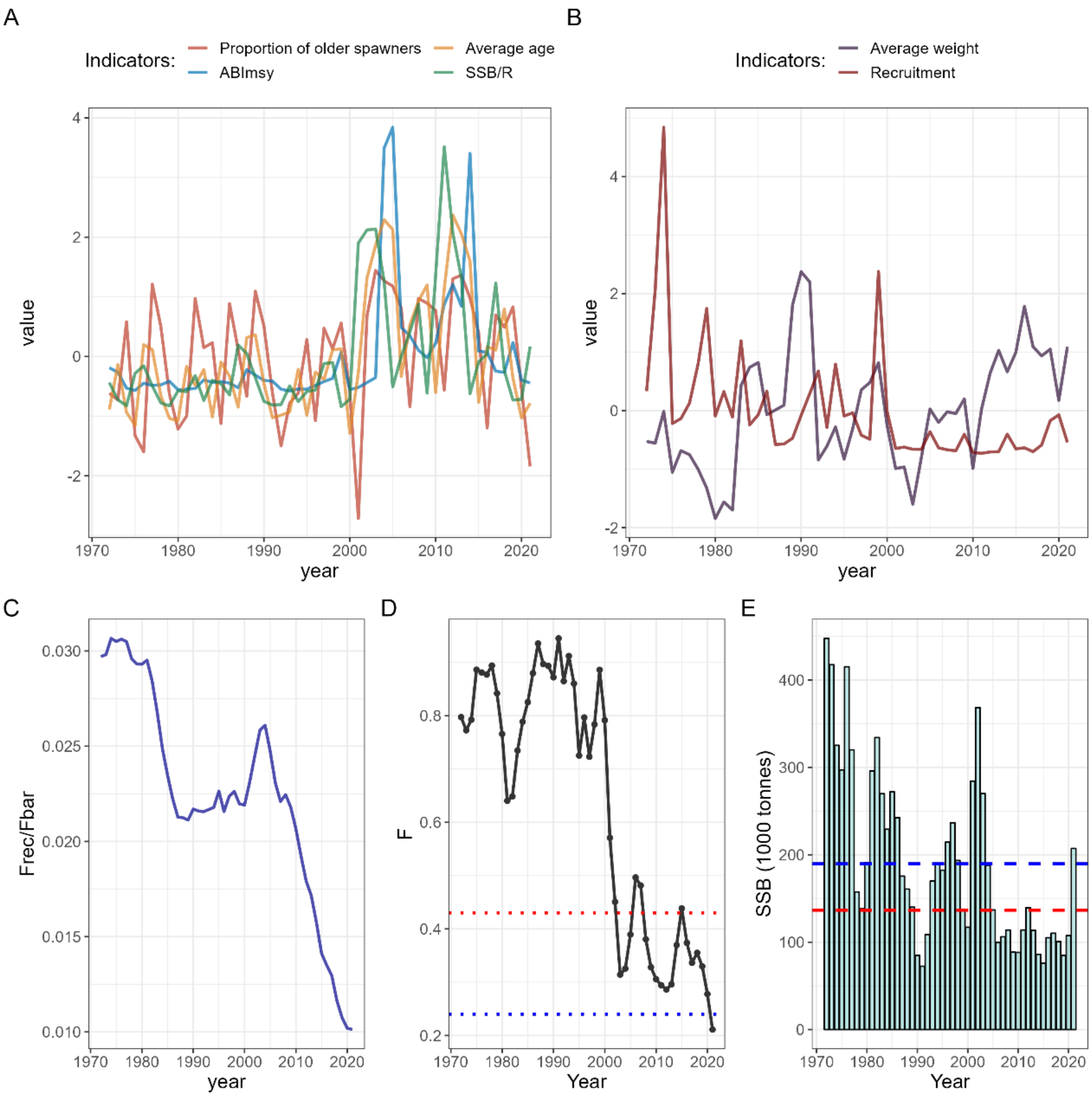


Figure A2.24. Estimated D3C3 indicators and stock status for Haddock (had.27.46a20). See Figure legend A2.1 for further details.

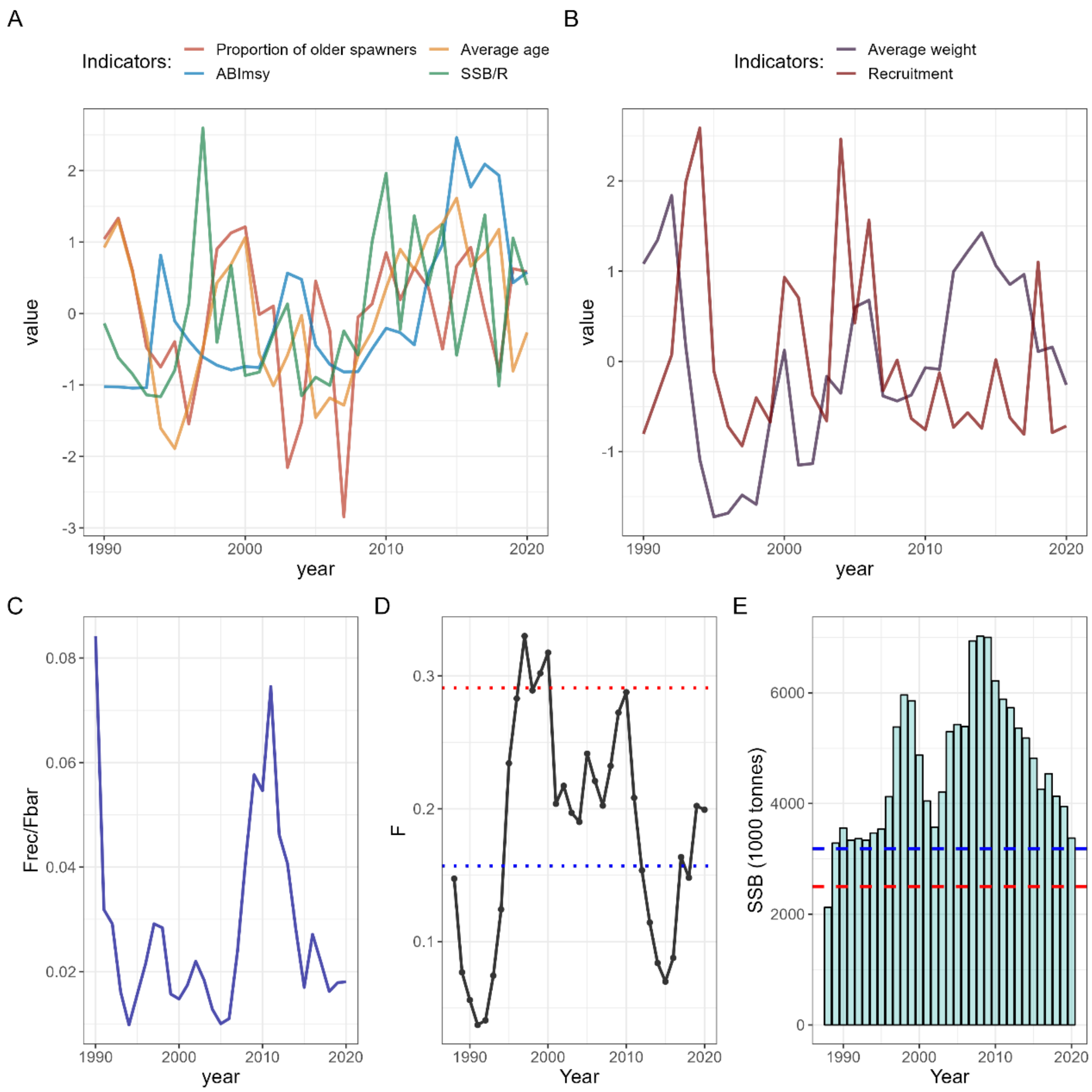


Figure A2.25. Estimated D3C3 indicators and stock status for Herring (her.27.1-24a514a). See Figure legend A2.1 for further details.

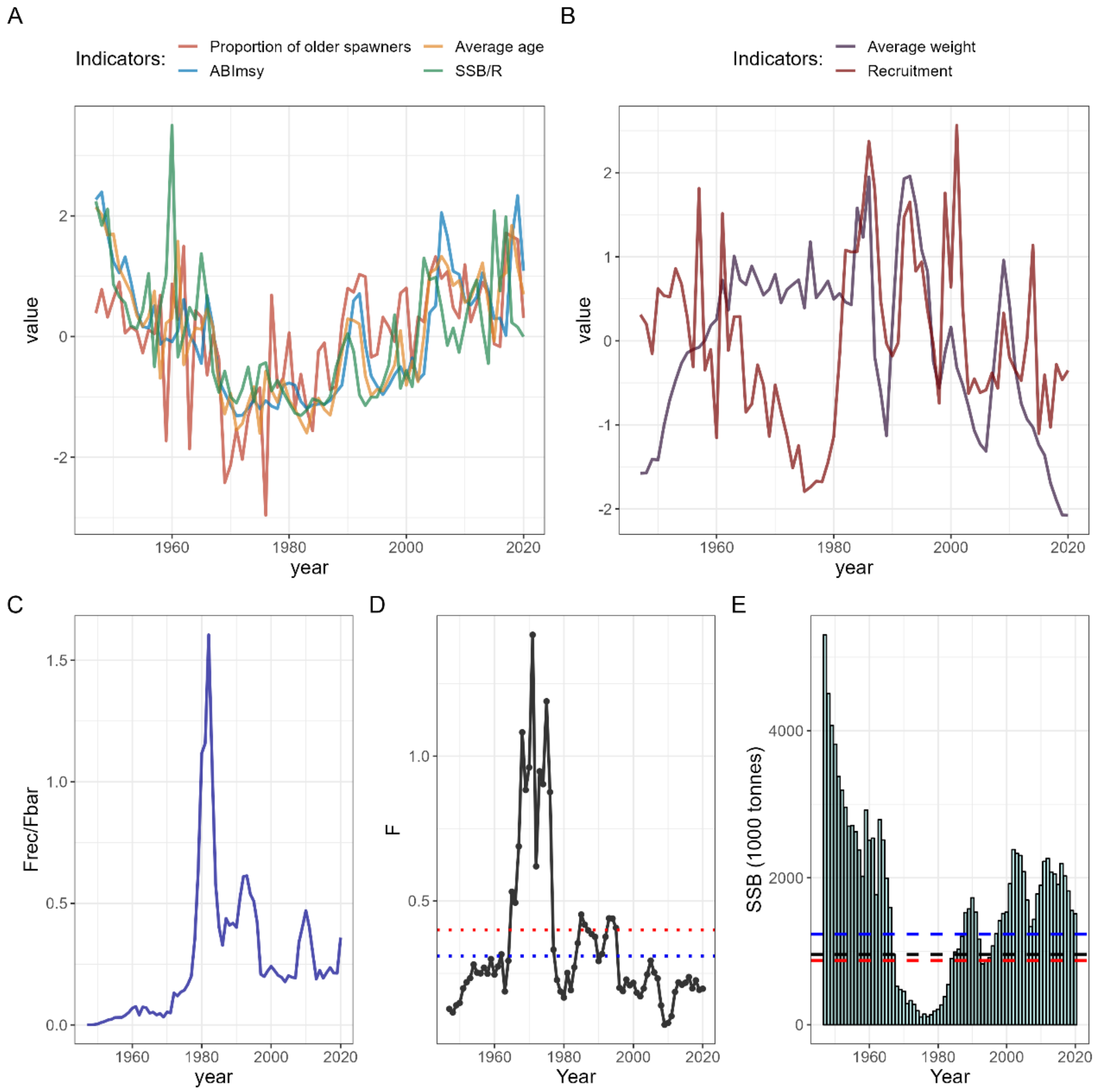


Figure A2.26. Estimated D3C3 indicators and stock status for Herring (her.27.3a47d). See Figure legend A2.1 for further details.

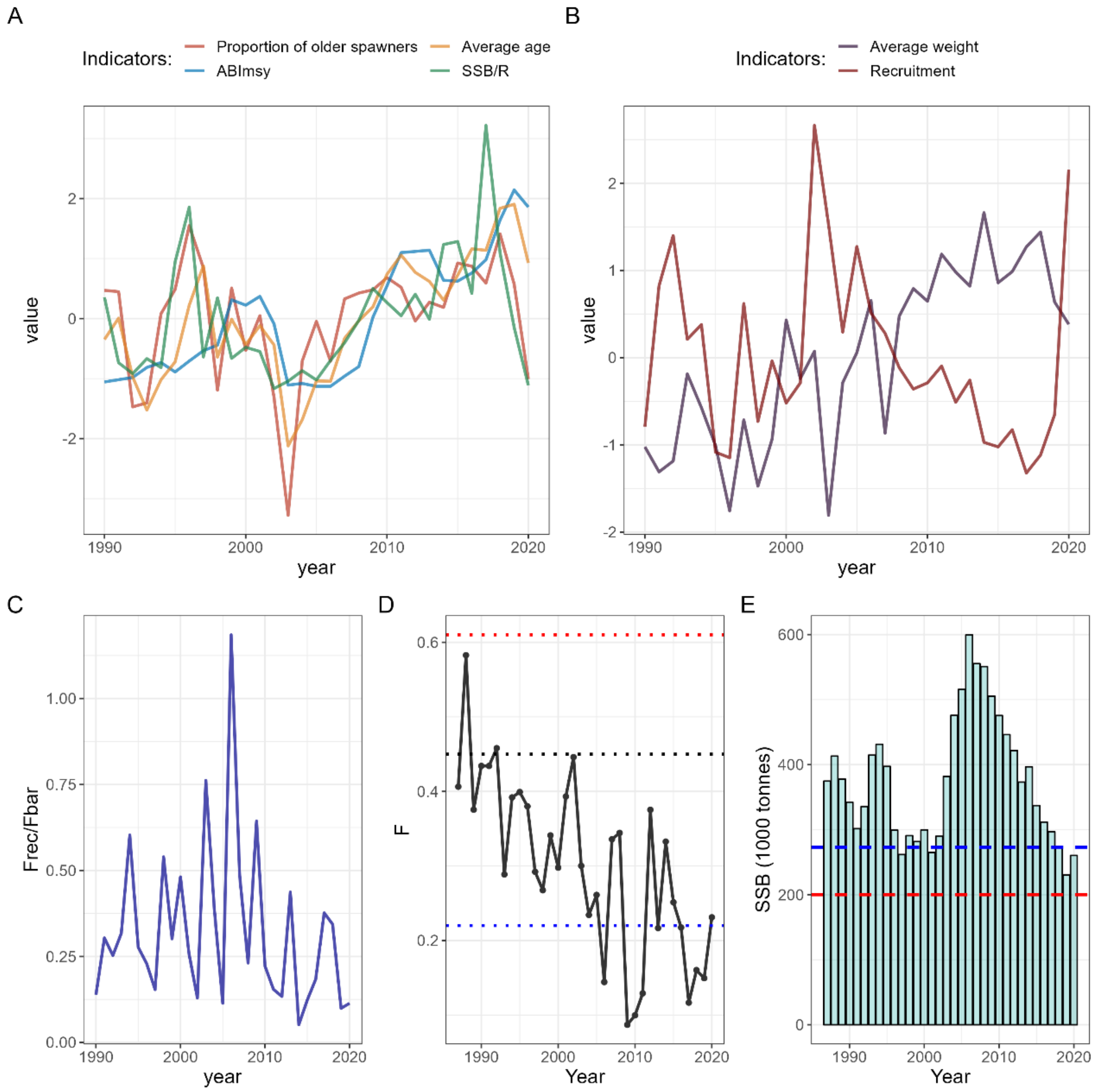


Figure A2.27. Estimated D3C3 indicators and stock status for Herring (her.27.5a). See Figure legend A2.1 for further details.

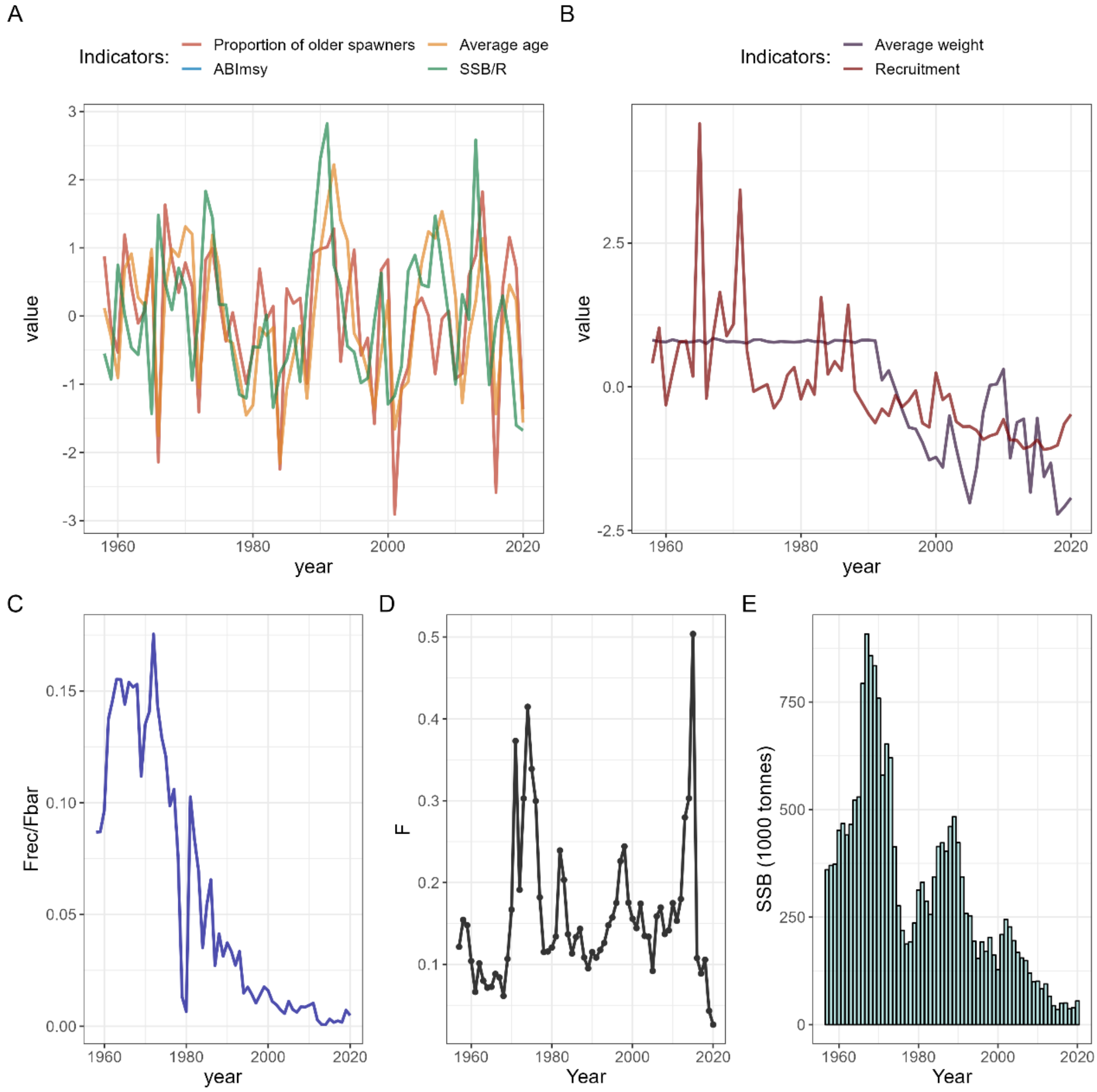


Figure A2.28. Estimated D3C3 indicators and stock status for Herring (her.27.6a7bc). See Figure legend A2.1 for further details.

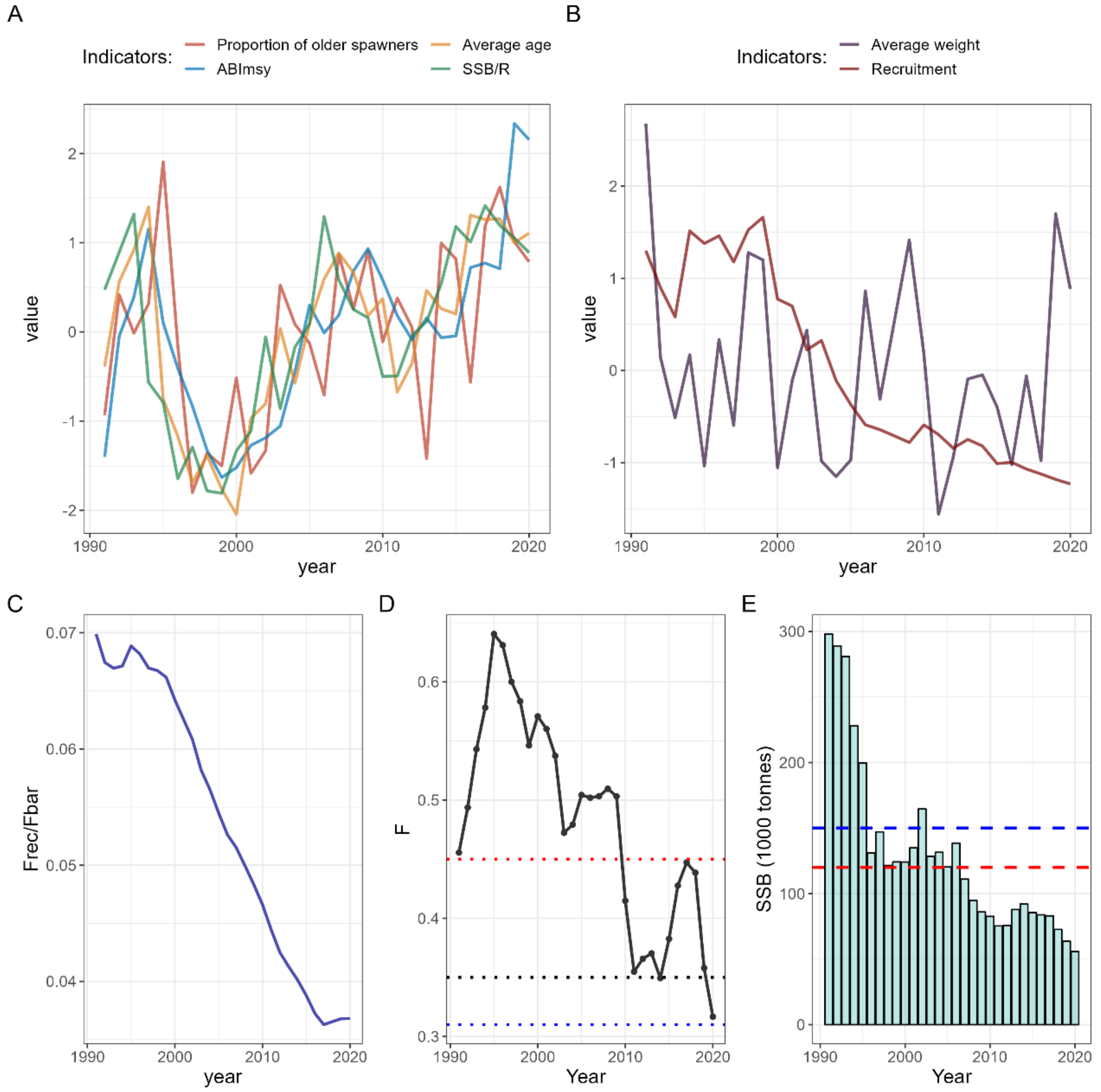


Figure A2.29. Estimated D3C3 indicators and stock status for Herring (her.27.20-24). See Figure legend A2.1 for further details.

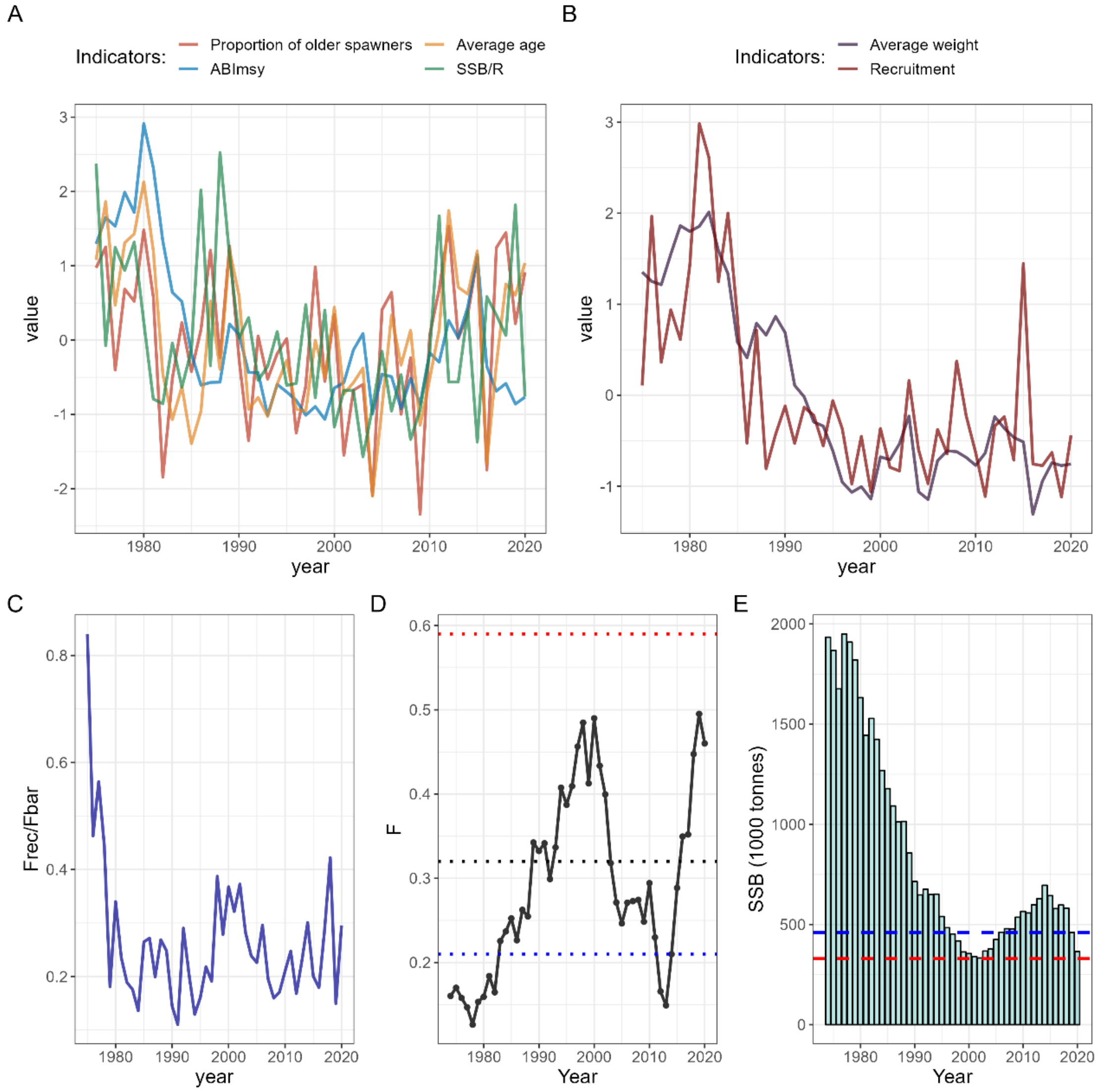


Figure A2.30. Estimated D3C3 indicators and stock status for Herring (her.27.25-2932). See Figure legend A2.1 for further details.

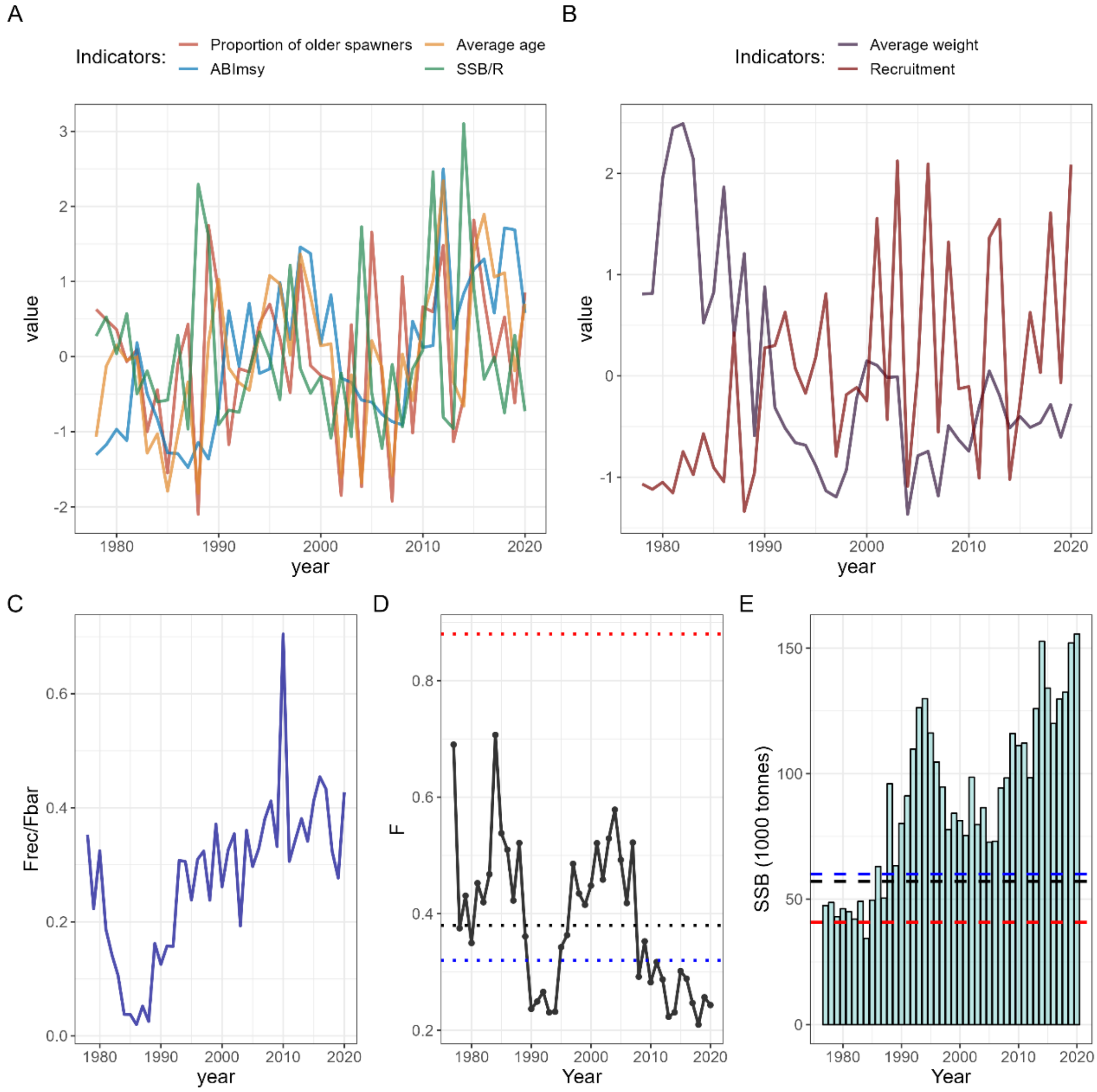


Figure A2.31. Estimated D3C3 indicators and stock status for Herring (her.27.28). See Figure legend A2.1 for further details.

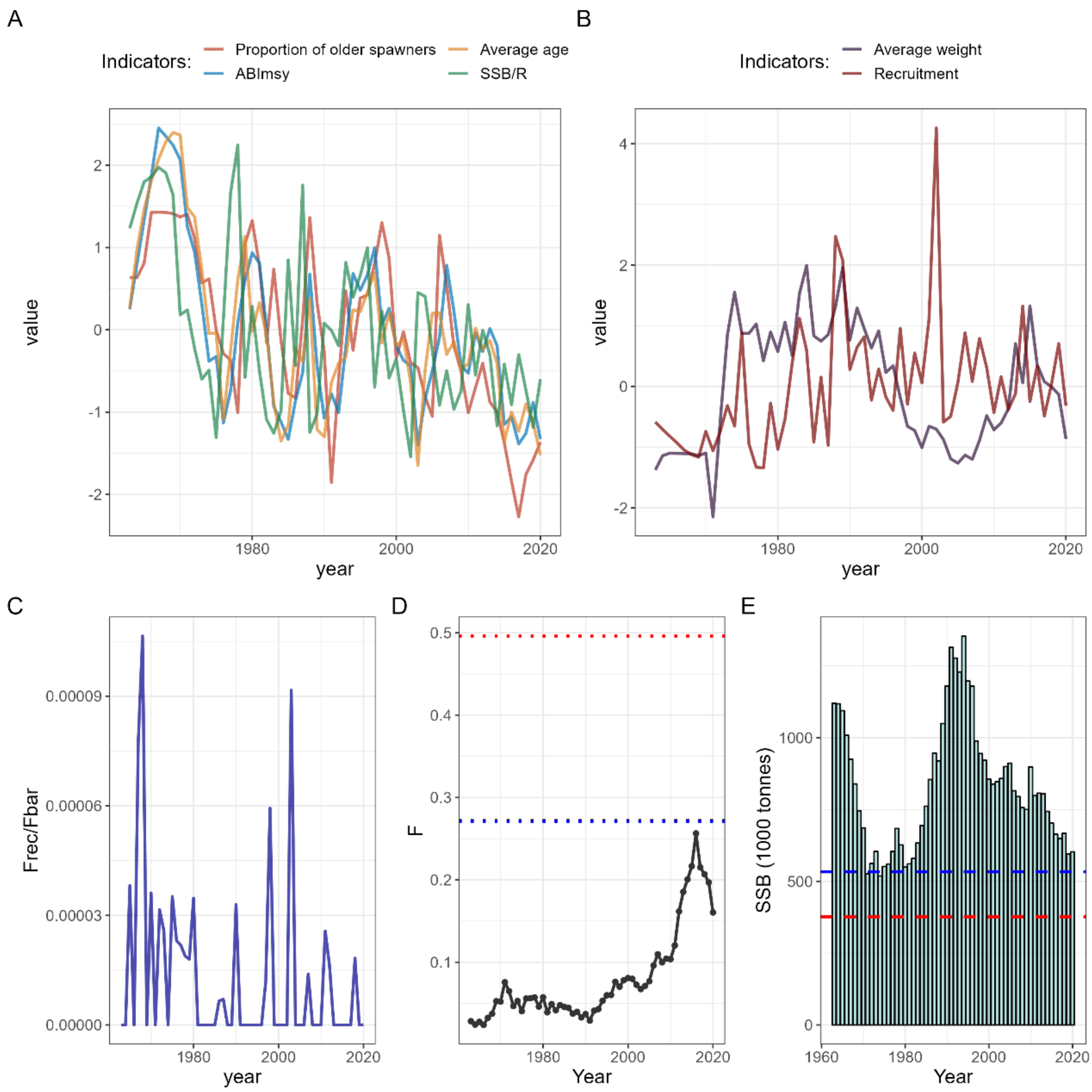


Figure A2.32. Estimated D3C3 indicators and stock status for Herring (her.27.3031). See Figure legend A2.1 for further details.

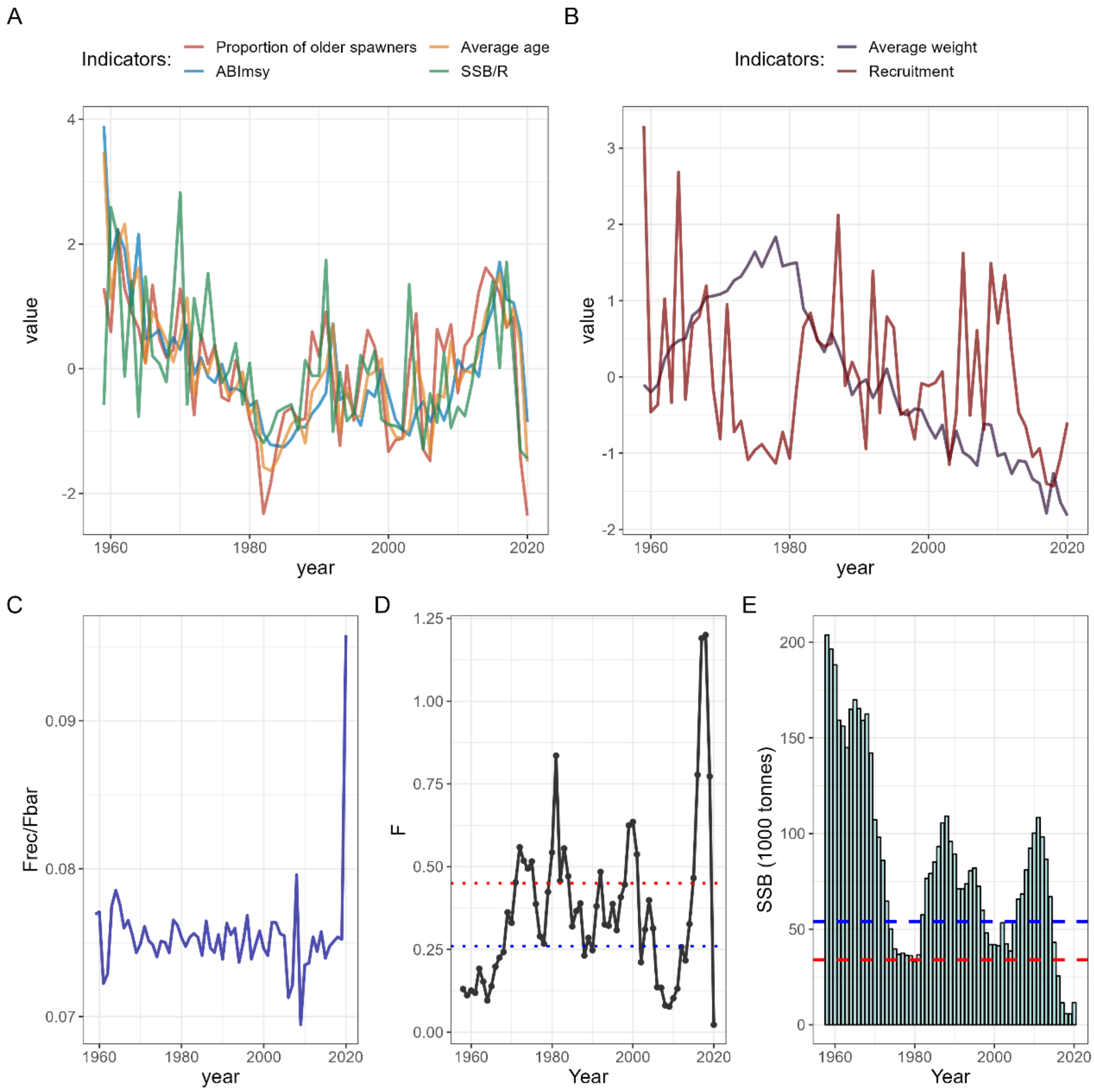


Figure A2.33. Estimated D3C3 indicators and stock status for Herring (her.27.irls). See Figure legend A2.1 for further details.

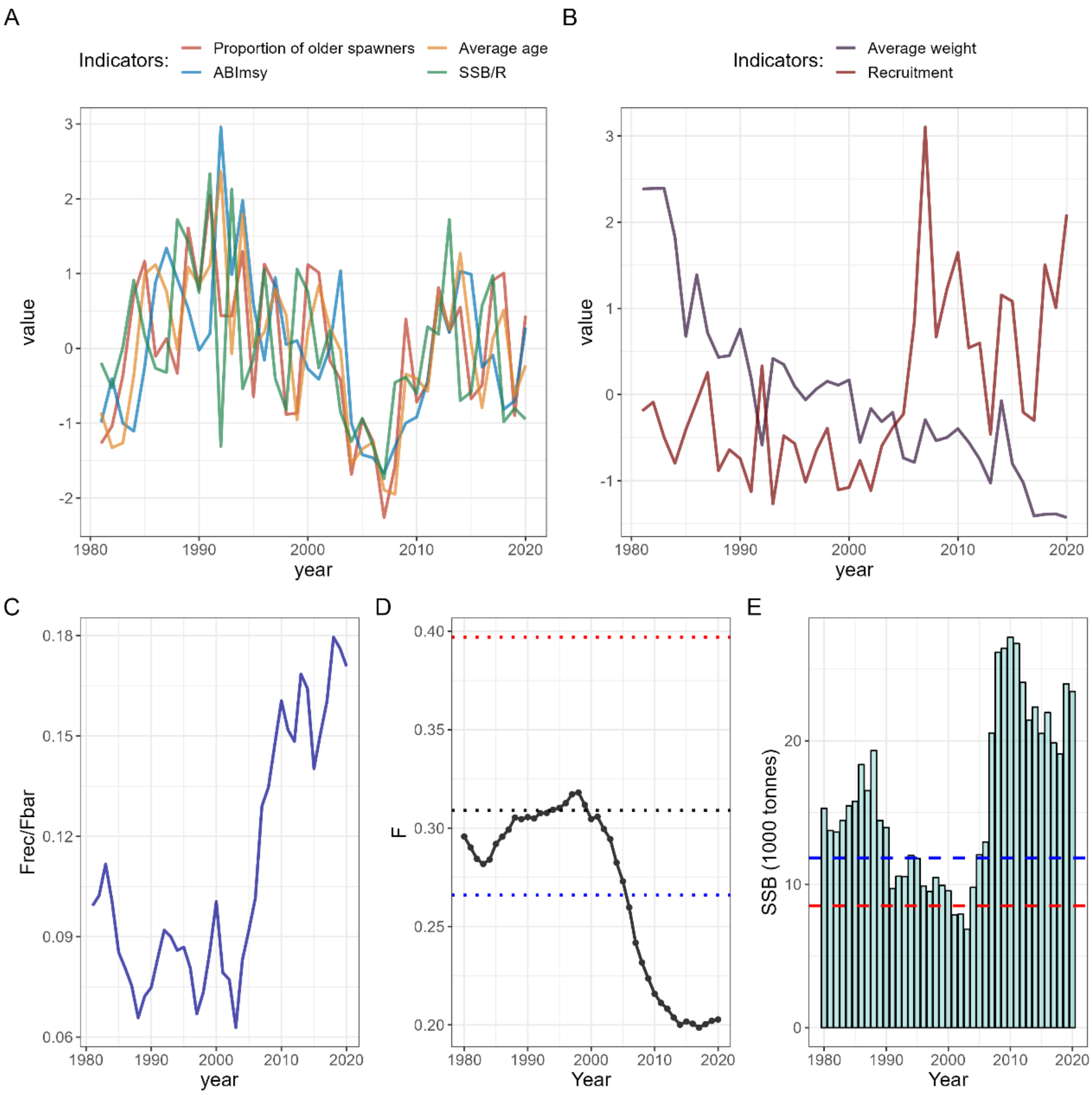


Figure A2.34. Estimated D3C3 indicators and stock status for Herring (her.27.nirs). See Figure legend A2.1 for further details.

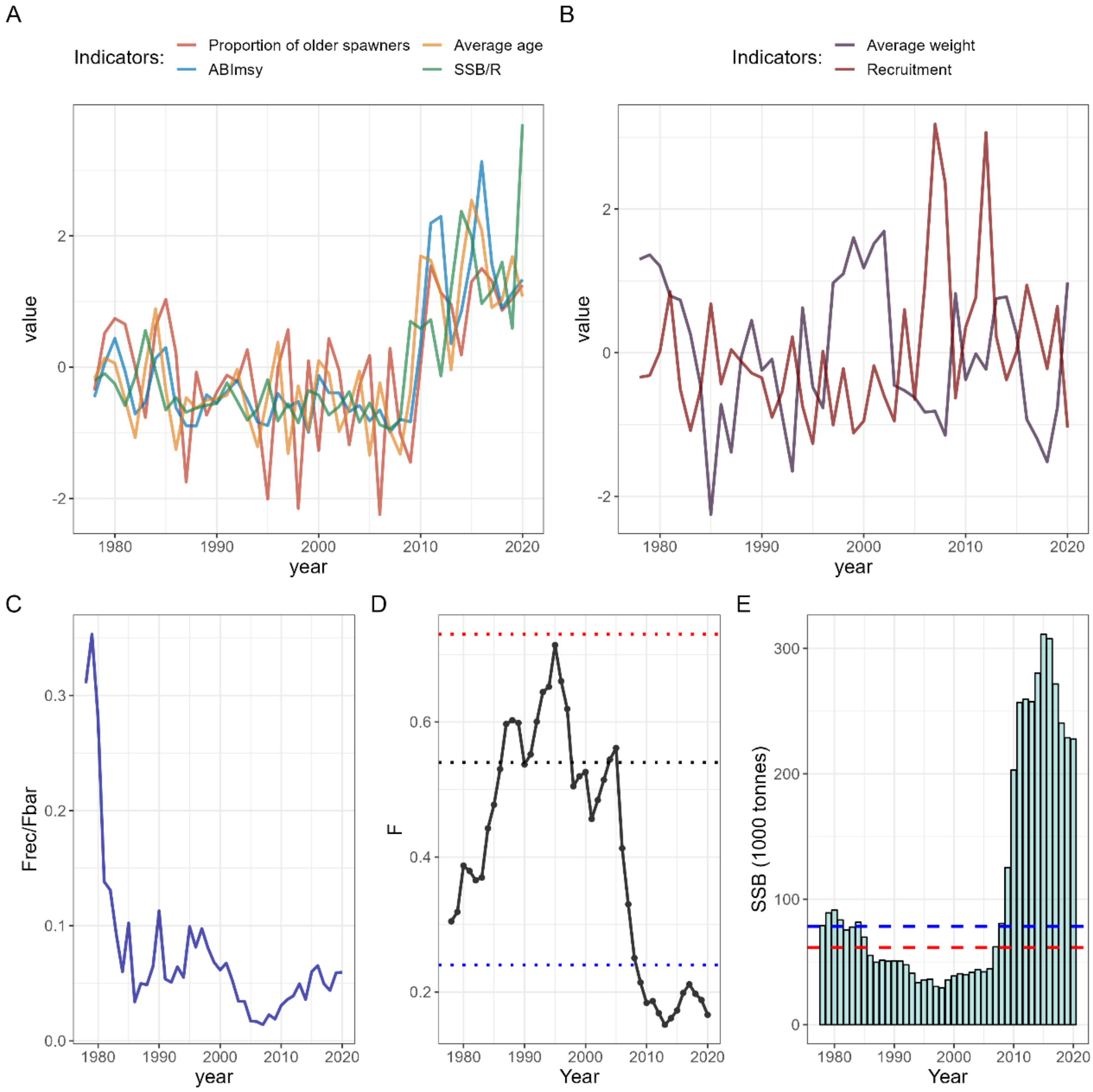


Figure A2.35. Estimated D3C3 indicators and stock status for Hake (her.27.3a46-8abd). See Figure legend A2.1 for further details.

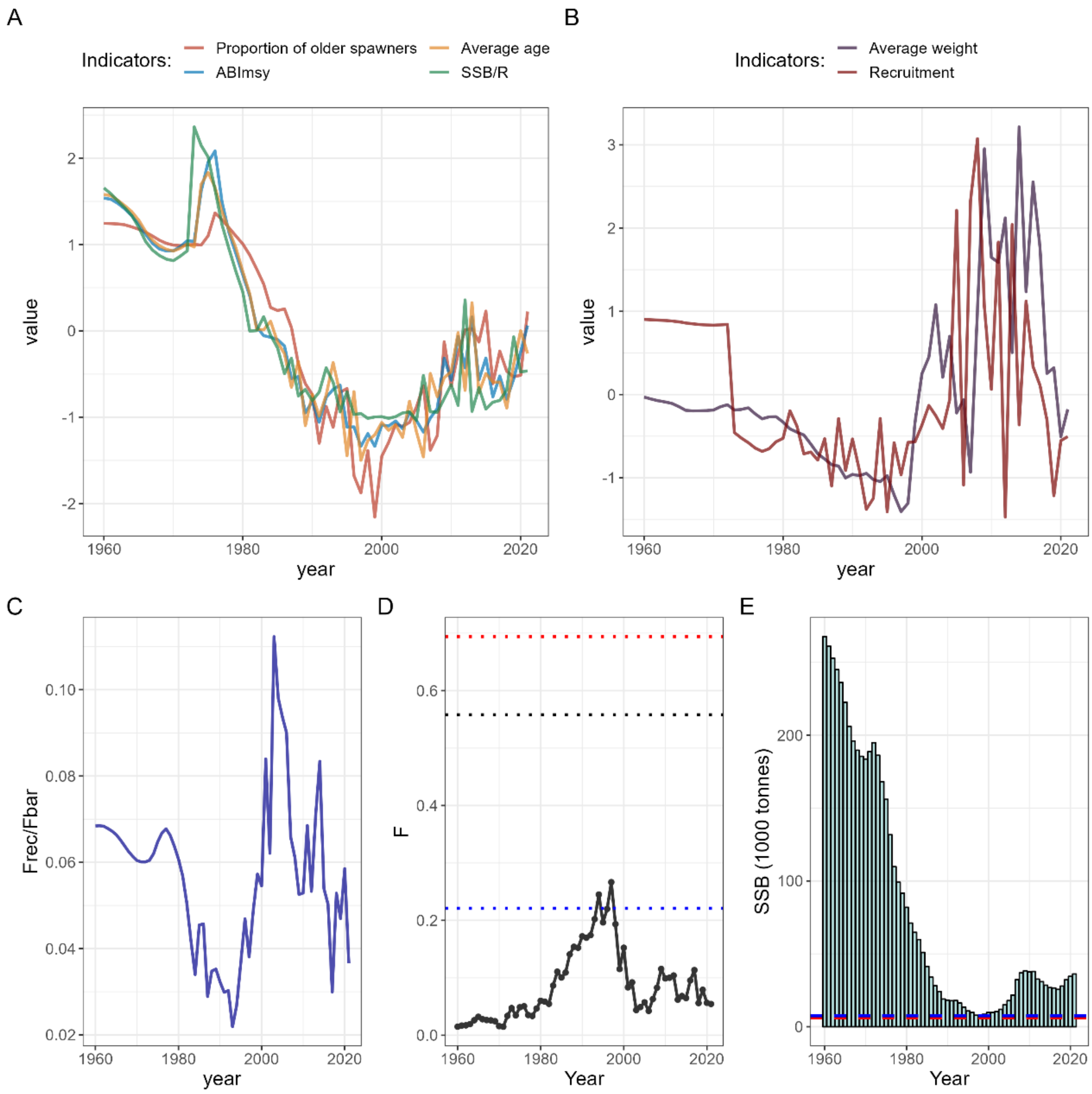


Figure A2.36. Estimated D3C3 indicators and stock status for Hake (her.27.8c9a). See Figure legend A2.1 for further details.

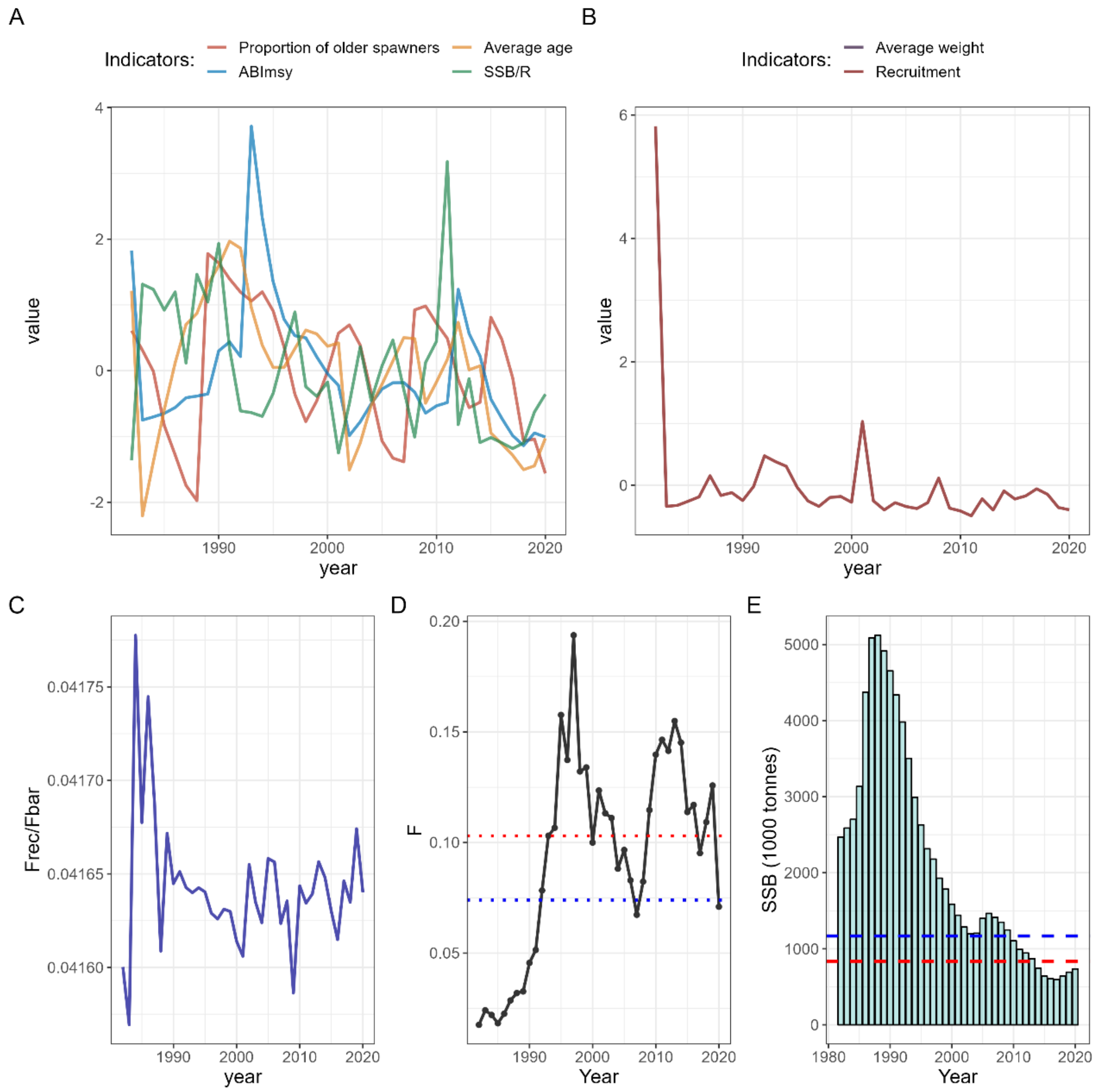


Figure A2.37. Estimated D3C3 indicators and stock status for Horse mackerel (hom.27.2a4a5b6a7a-ce-k8). See Figure legend A2.1 for further details.

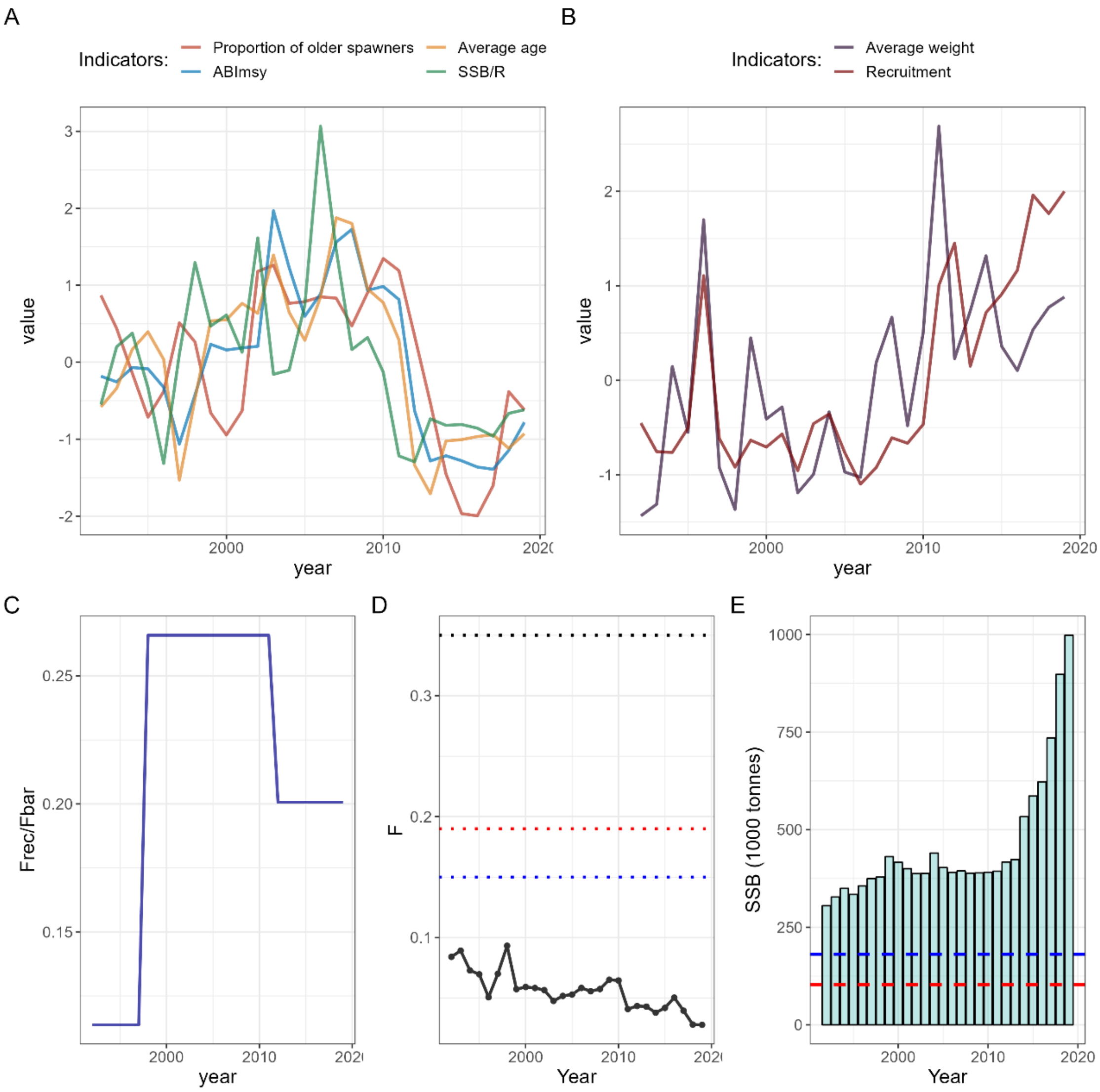


Figure A2.38. Estimated D3C3 indicators and stock status for Horse mackerel (hom.27.9a). See Figure legend A2.1 for further details.

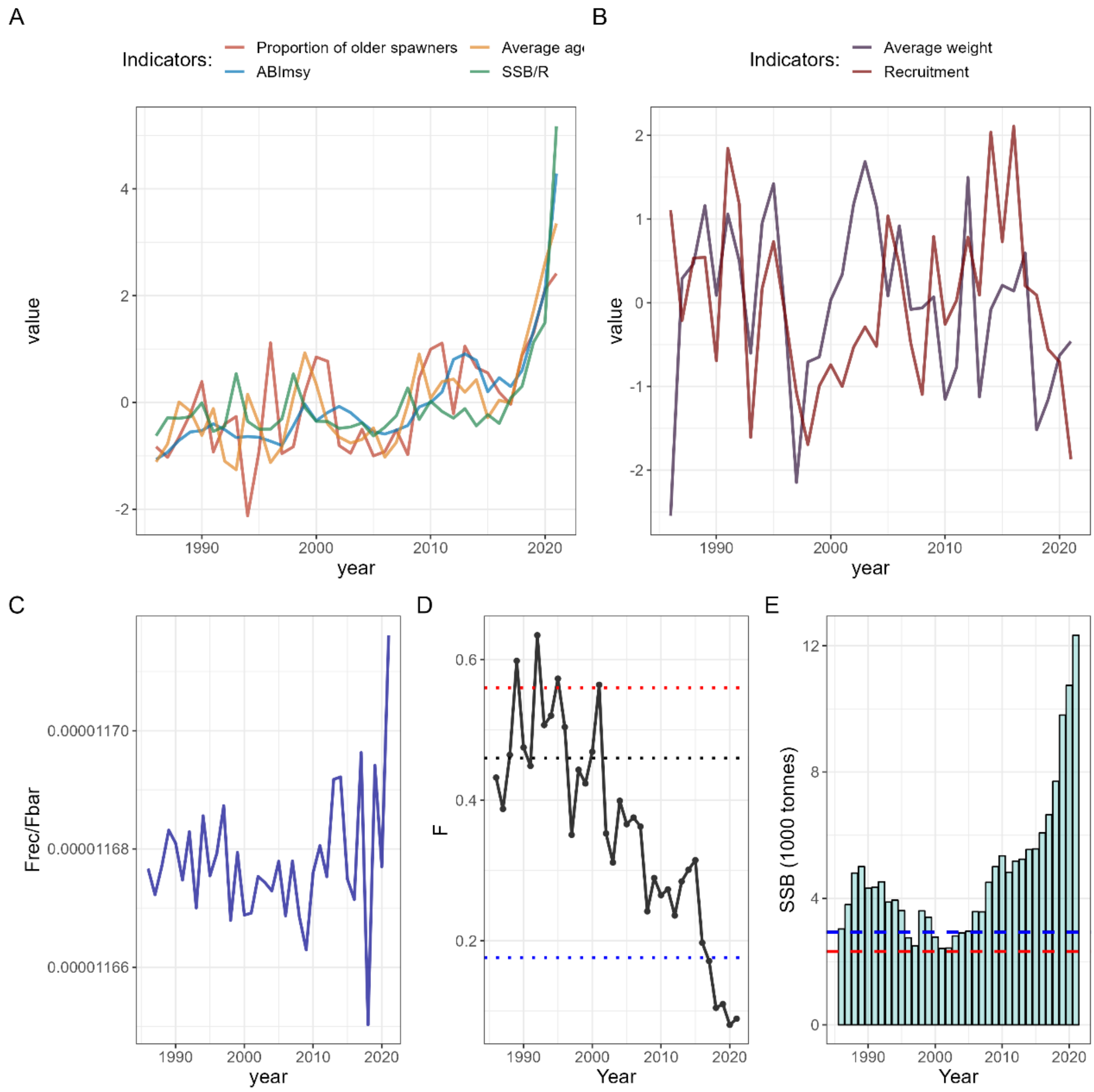


Figure A2.39. Estimated D3C3 indicators and stock status for Megrim (ldb.27.8c9a). See Figure legend A2.1 for further details.

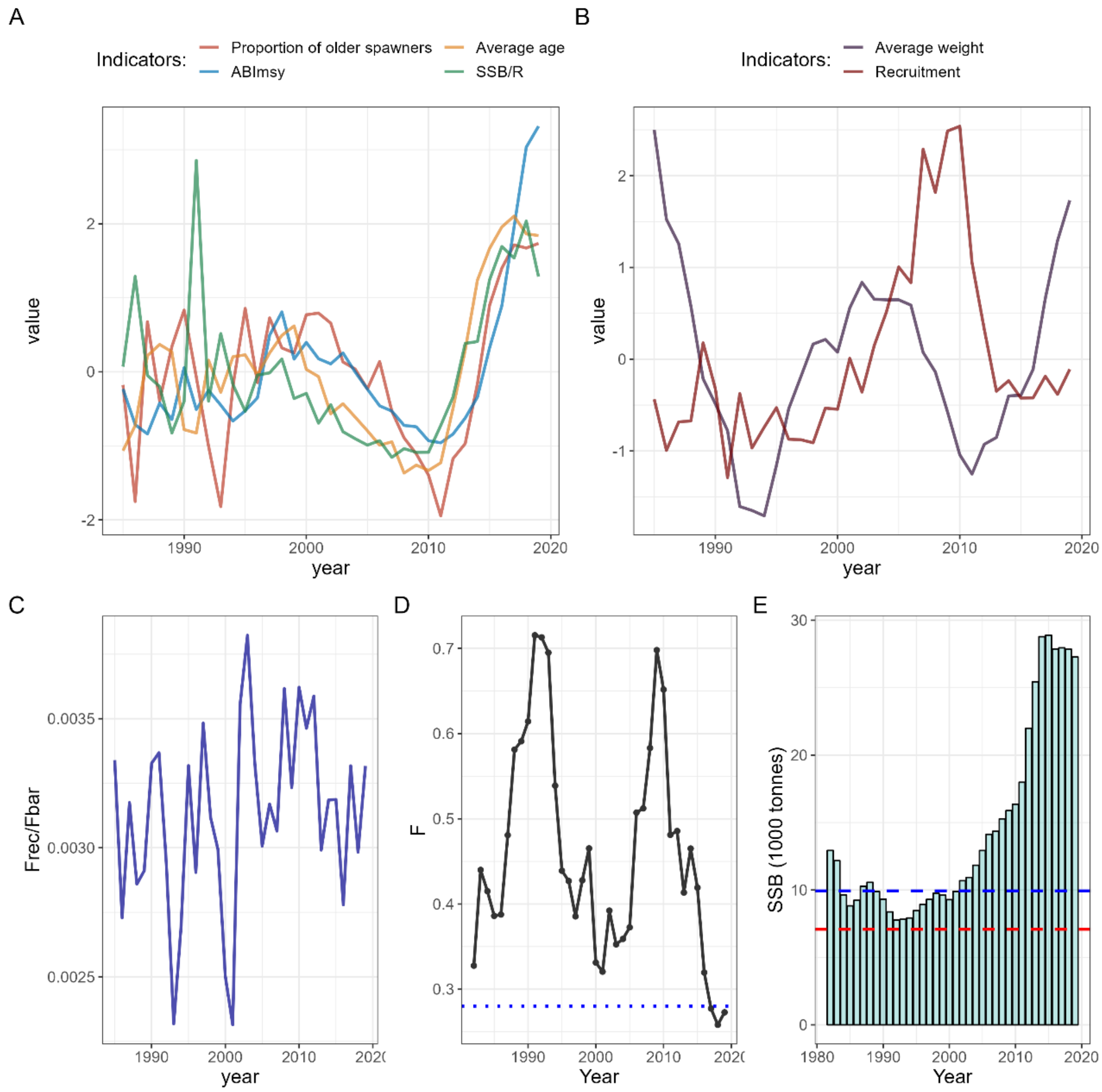


Figure A2.40. Estimated D3C3 indicators and stock status for Ling (lin.27.5a). See Figure legend A2.1 for further details.

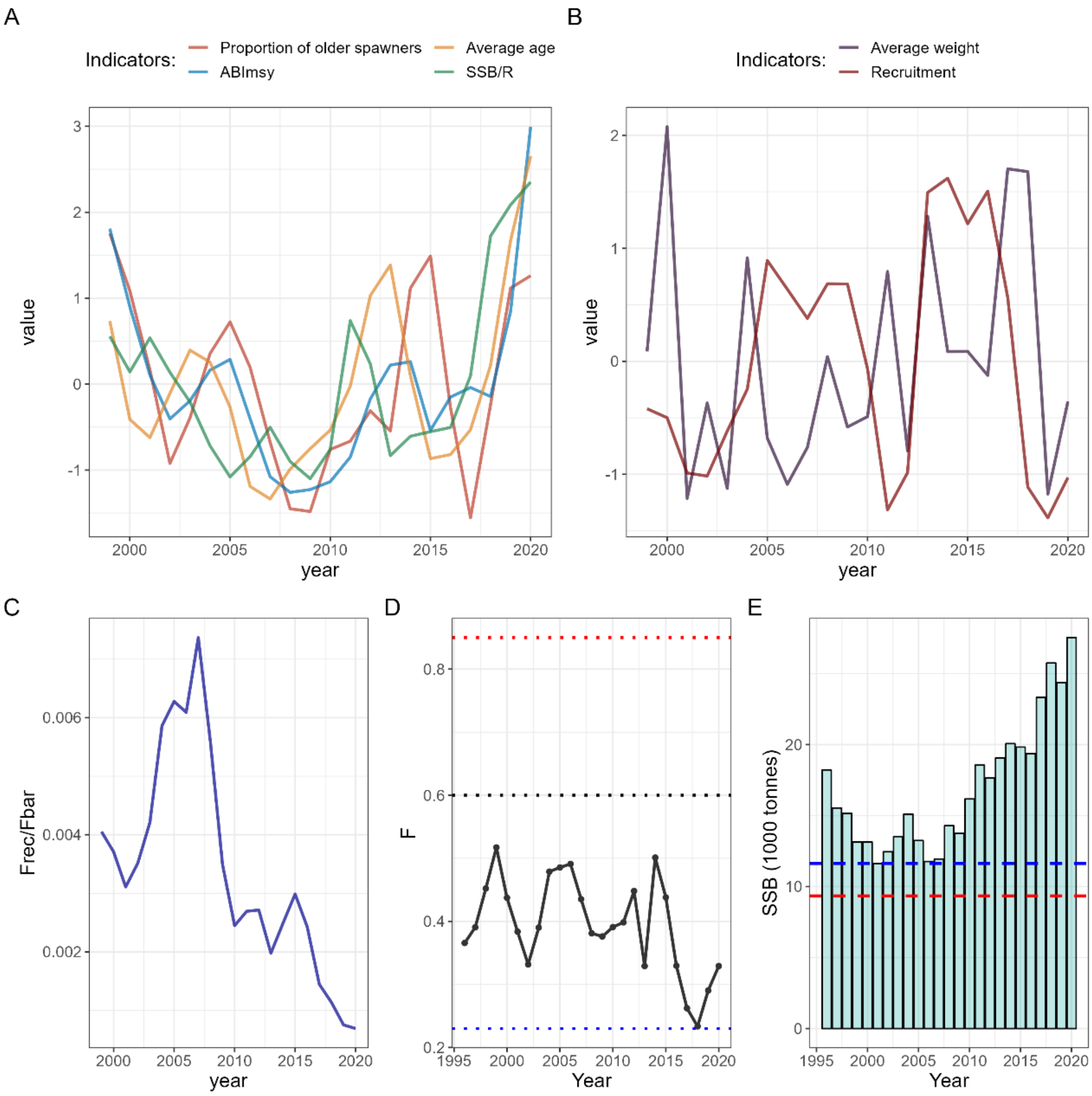


Figure A2.41. Estimated D3C3 indicators and stock status for Ling (lin.27.5b). See Figure legend A2.1 for further details.

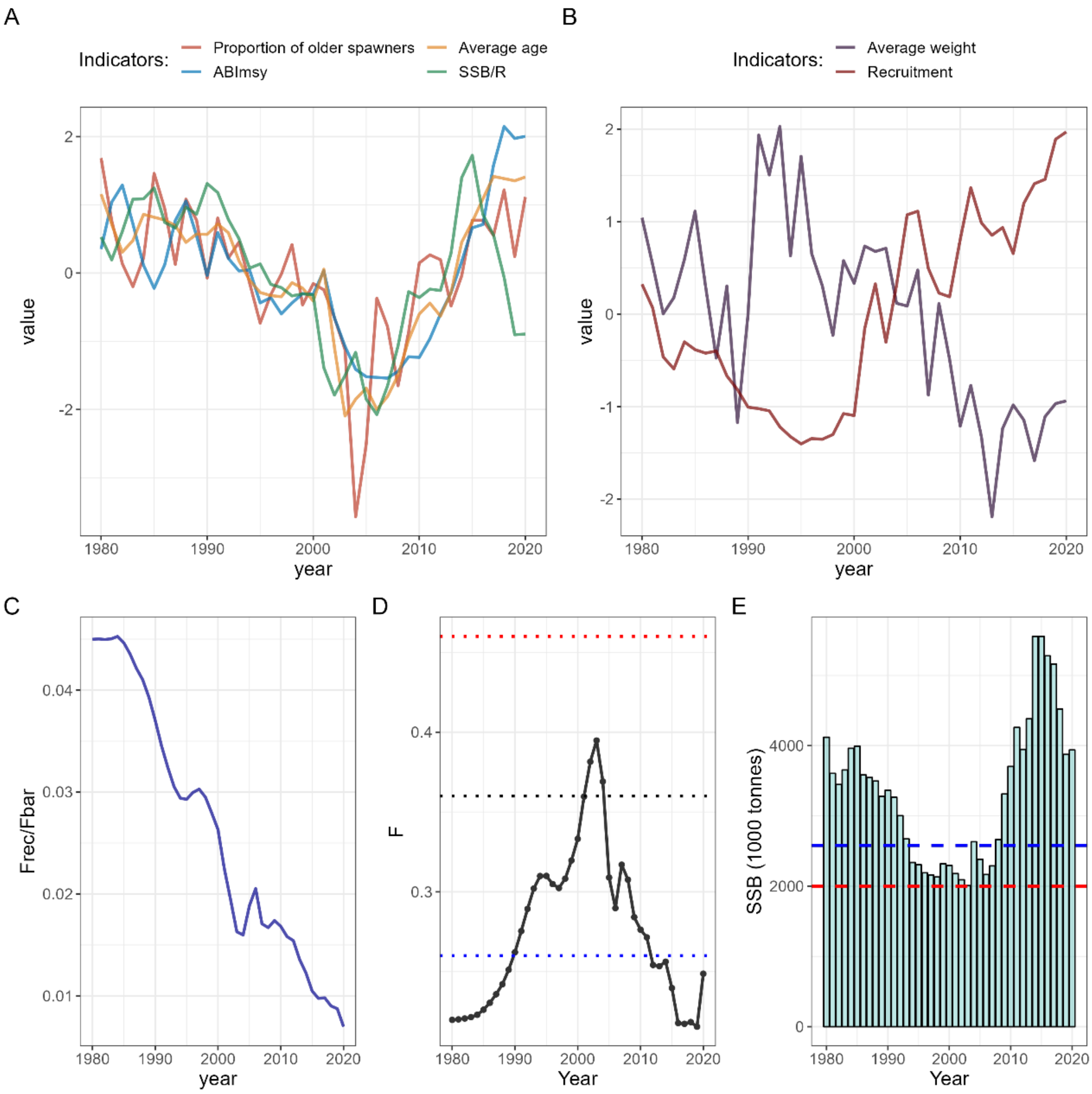


Figure A2.42. Estimated D3C3 indicators and stock status for Mackerel (mac.27.nea). See Figure legend A2.1 for further details.

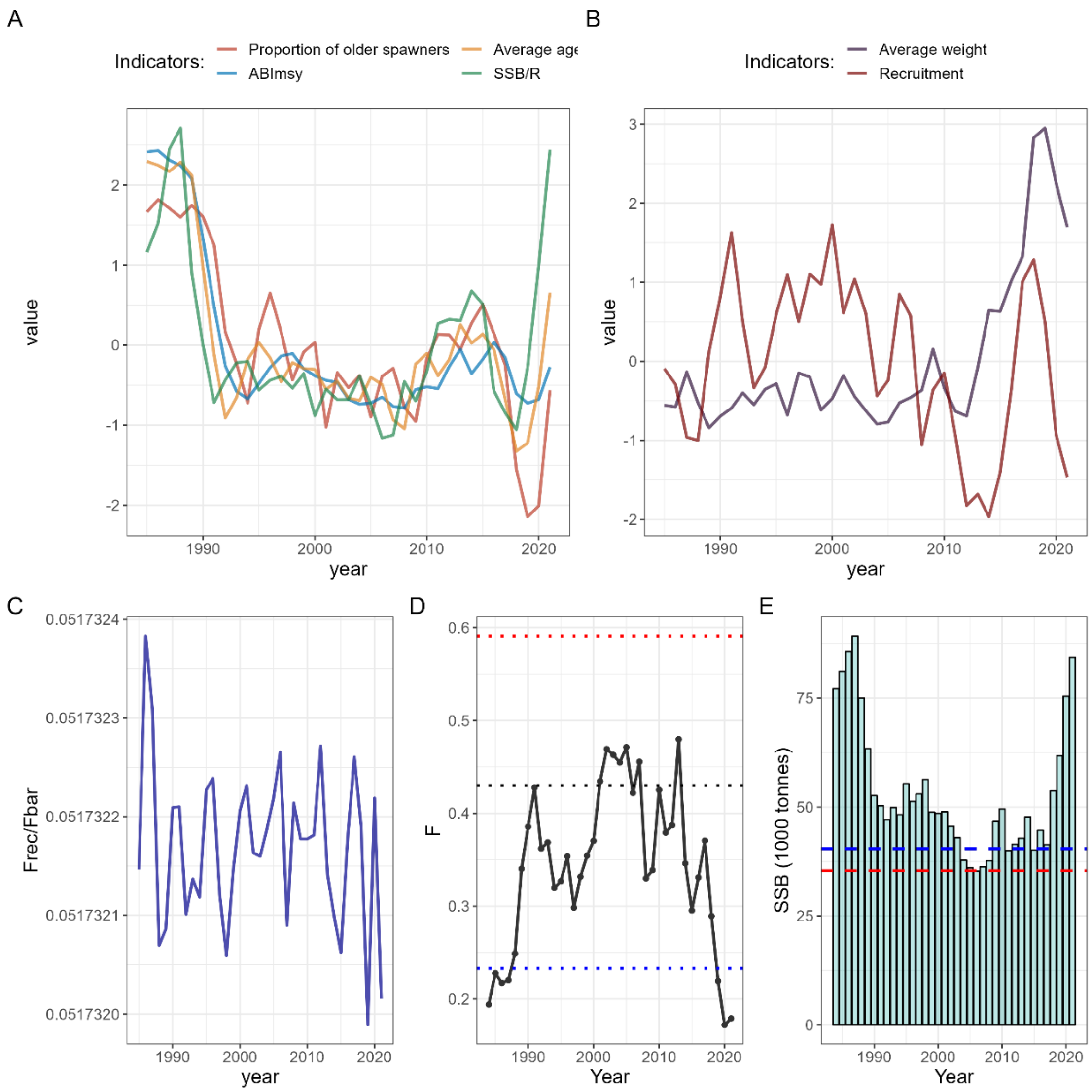


Figure A2.43. Estimated D3C3 indicators and stock status for Megrim (meg.27.7b-k8abd). See Figure legend A2.1 for further details.

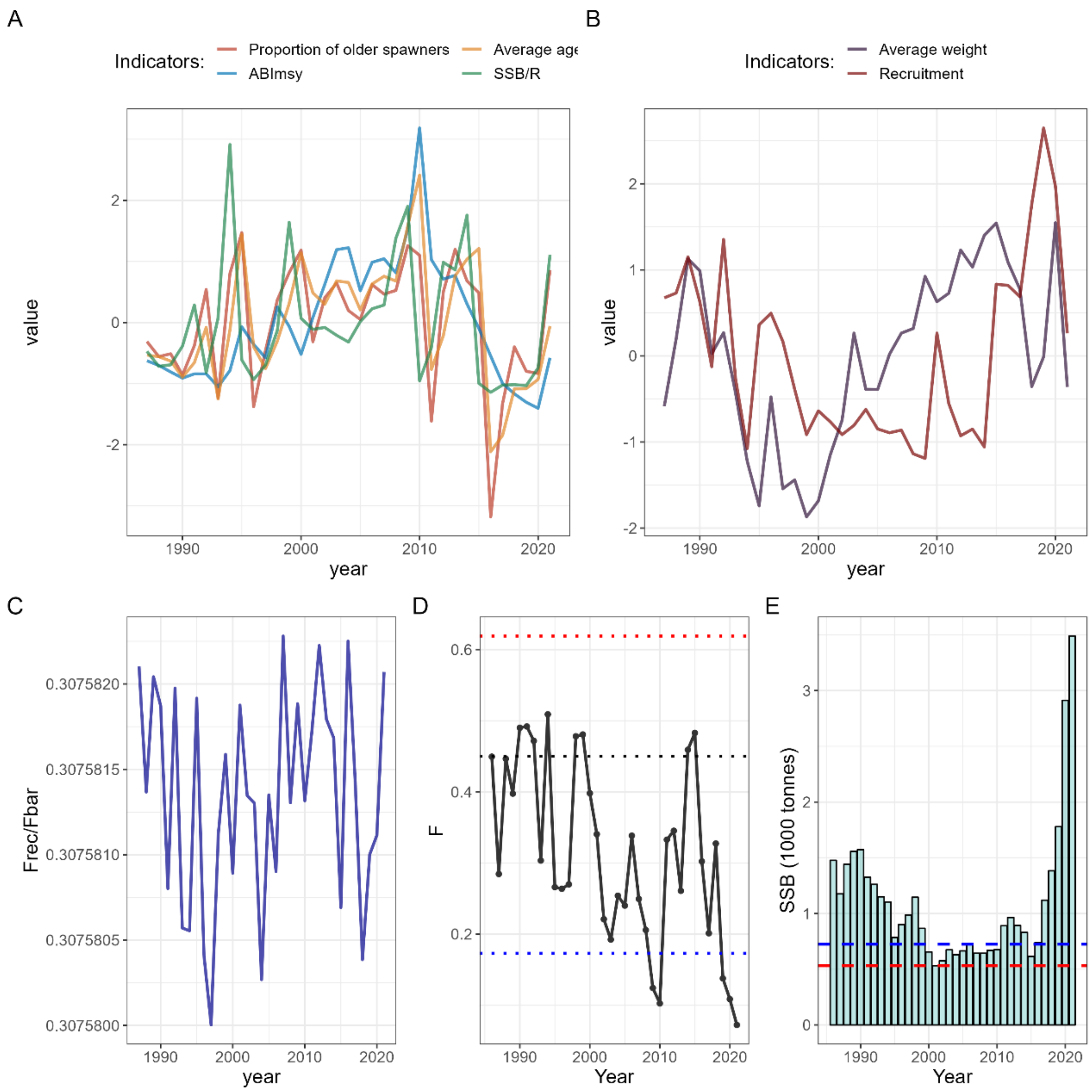


Figure A2.44. Estimated D3C3 indicators and stock status for Megrim (meg.27.8c9a). See Figure legend A2.1 for further details.

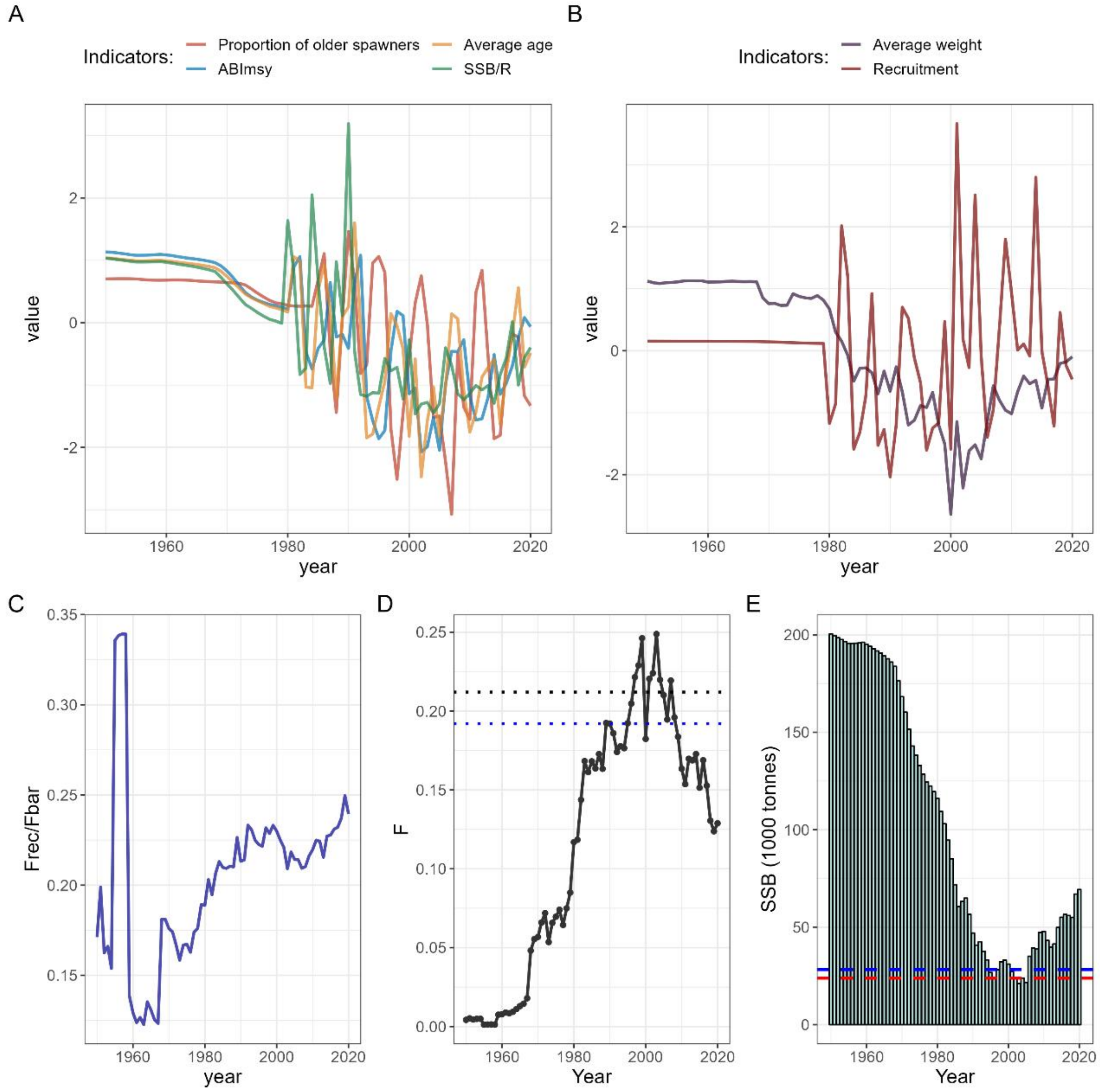


Figure A2.45. Estimated D3C3 indicators and stock status for White anglerfish (mon.27.78abd). See Figure legend A2.1 for further details.

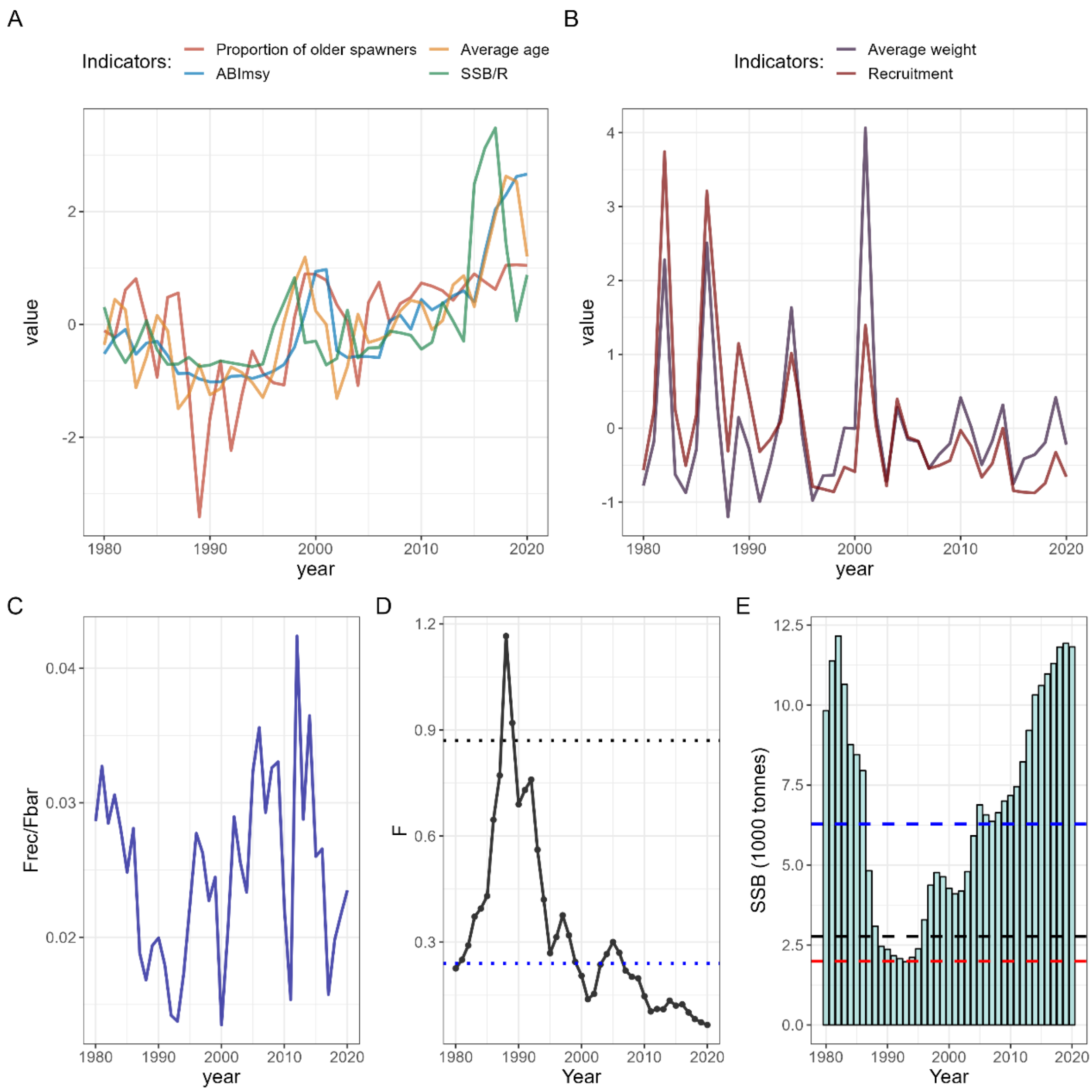


Figure A2.46. Estimated D3C3 indicators and stock status for White anglerfish (mon.27.8c9a). See Figure legend A2.1 for further details.

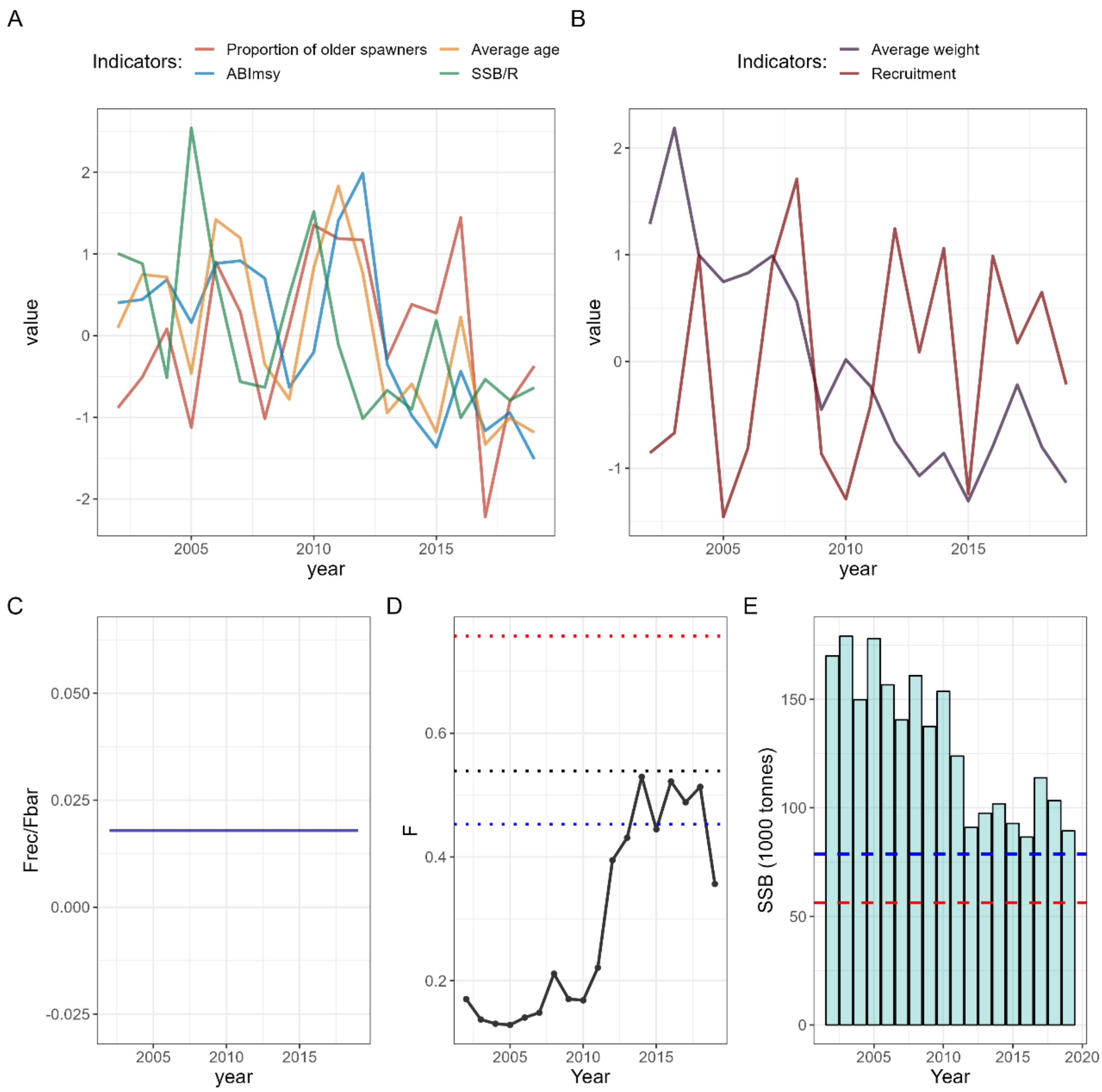


Figure A2.47. Estimated D3C3 indicators and stock status for Sardine (pil.27.8abd). See Figure legend A2.1 for further details.

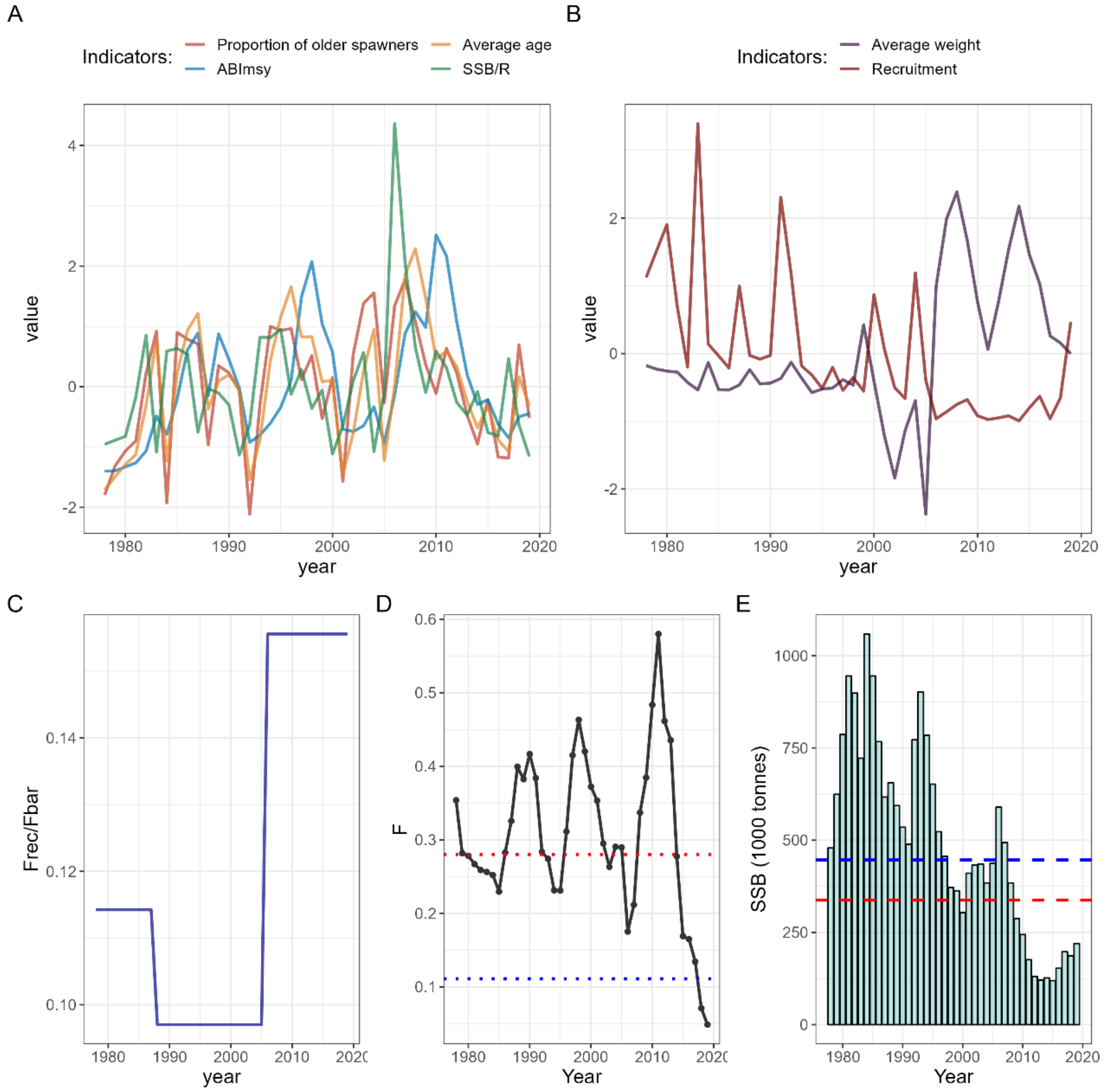


Figure A2.48. Estimated D3C3 indicators and stock status for Sardine (pil.27.8c9a). See Figure legend A2.1 for further details.

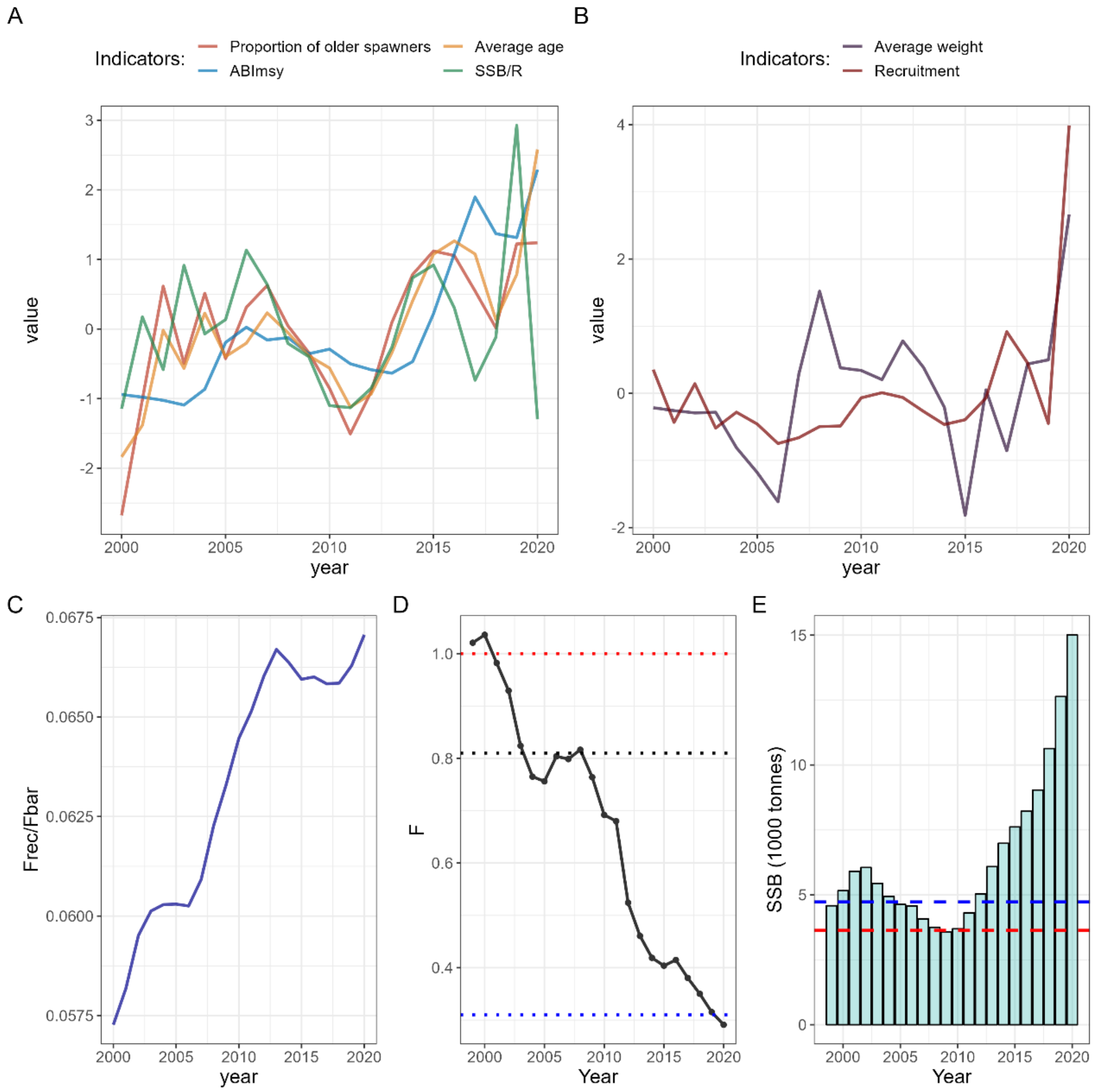


Figure A2.49. Estimated D3C3 indicators and stock status for European plaice (ple.27.21-23). See Figure legend A2.1 for further details.

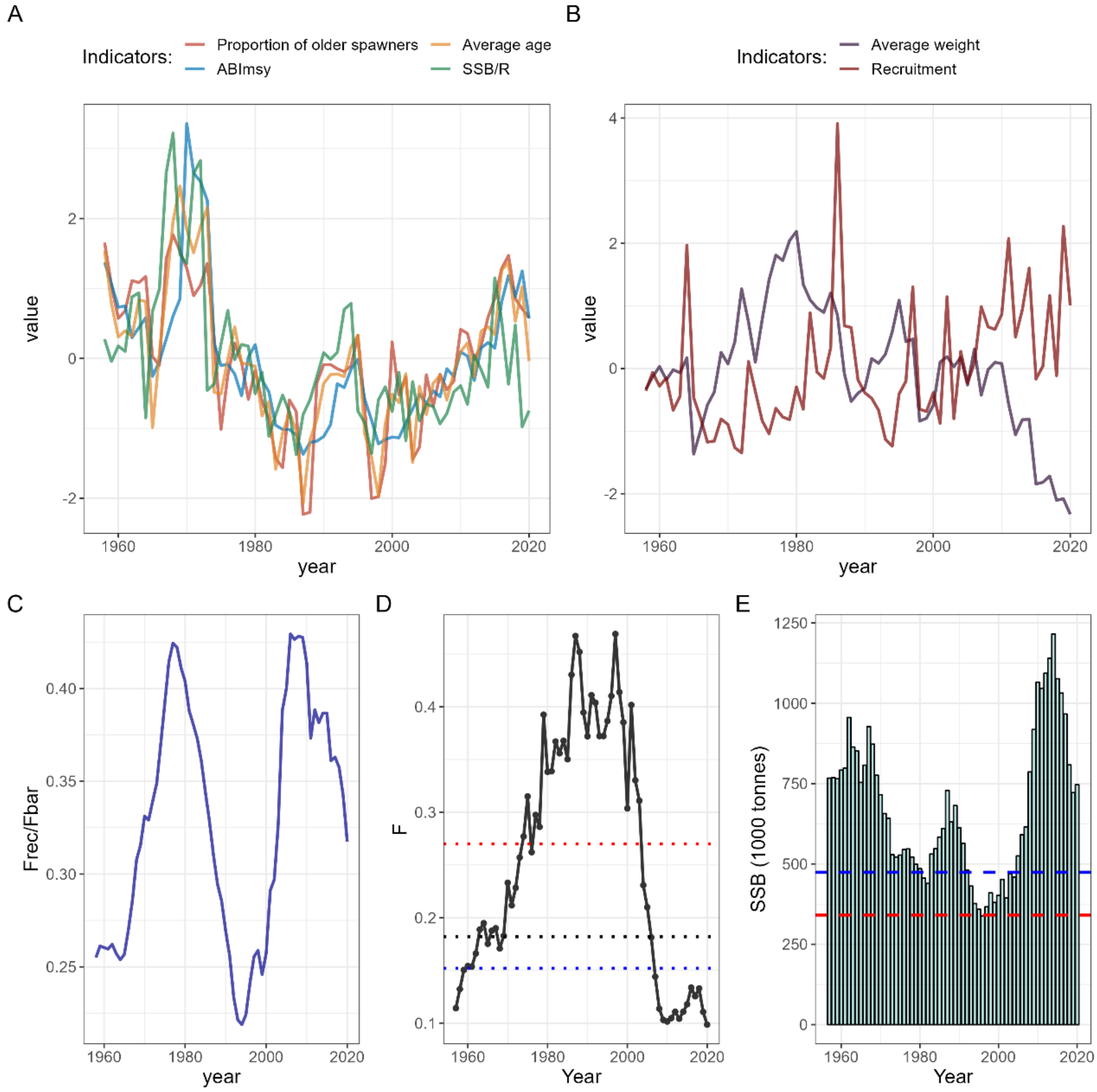


Figure A2.50. Estimated D3C3 indicators and stock status for European plaice (ple.27.420). See Figure legend A2.1 for further details.

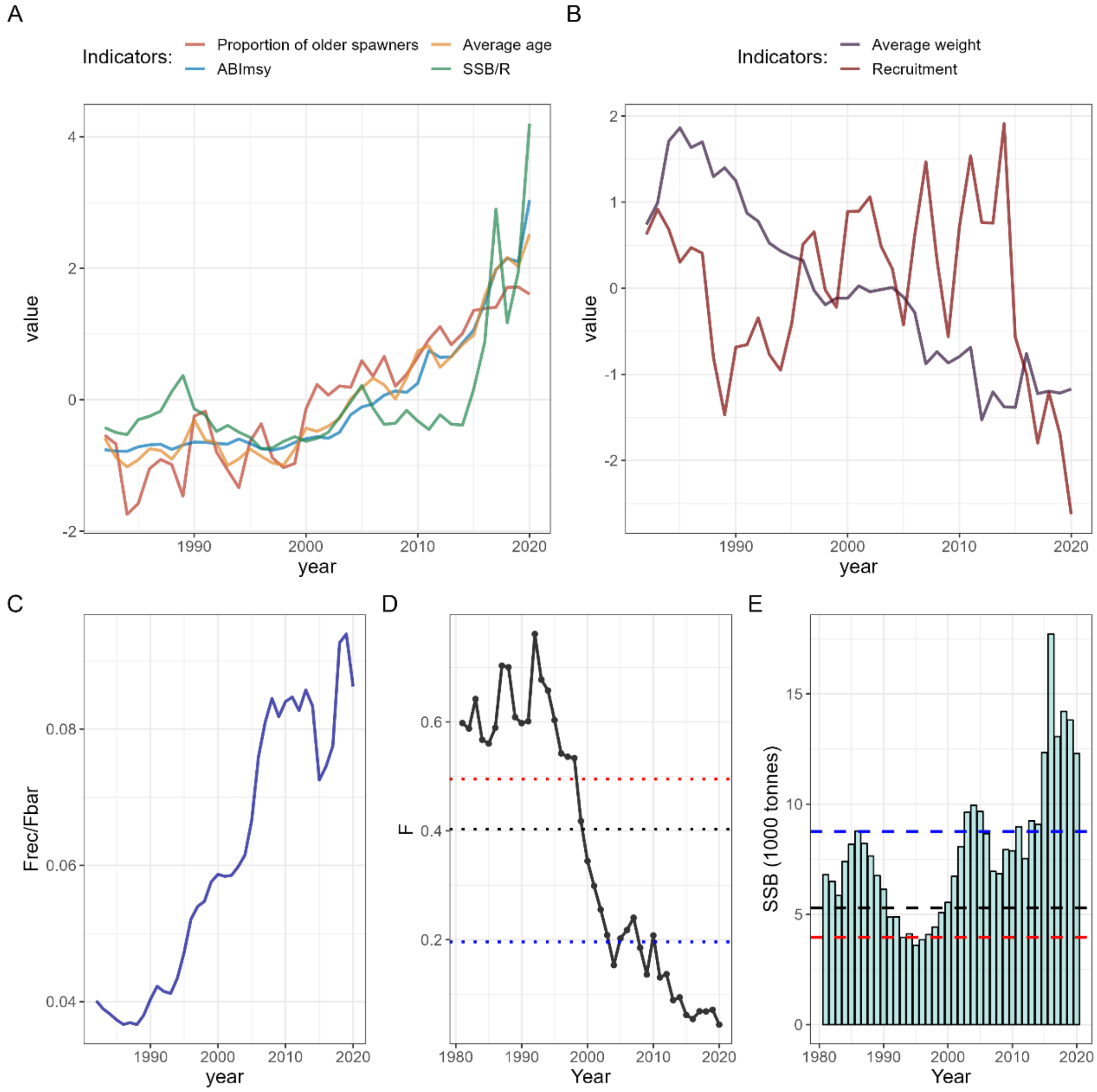


Figure A2.51. Estimated D3C3 indicators and stock status for European plaice (ple.27.7a). See Figure legend A2.1 for further details.

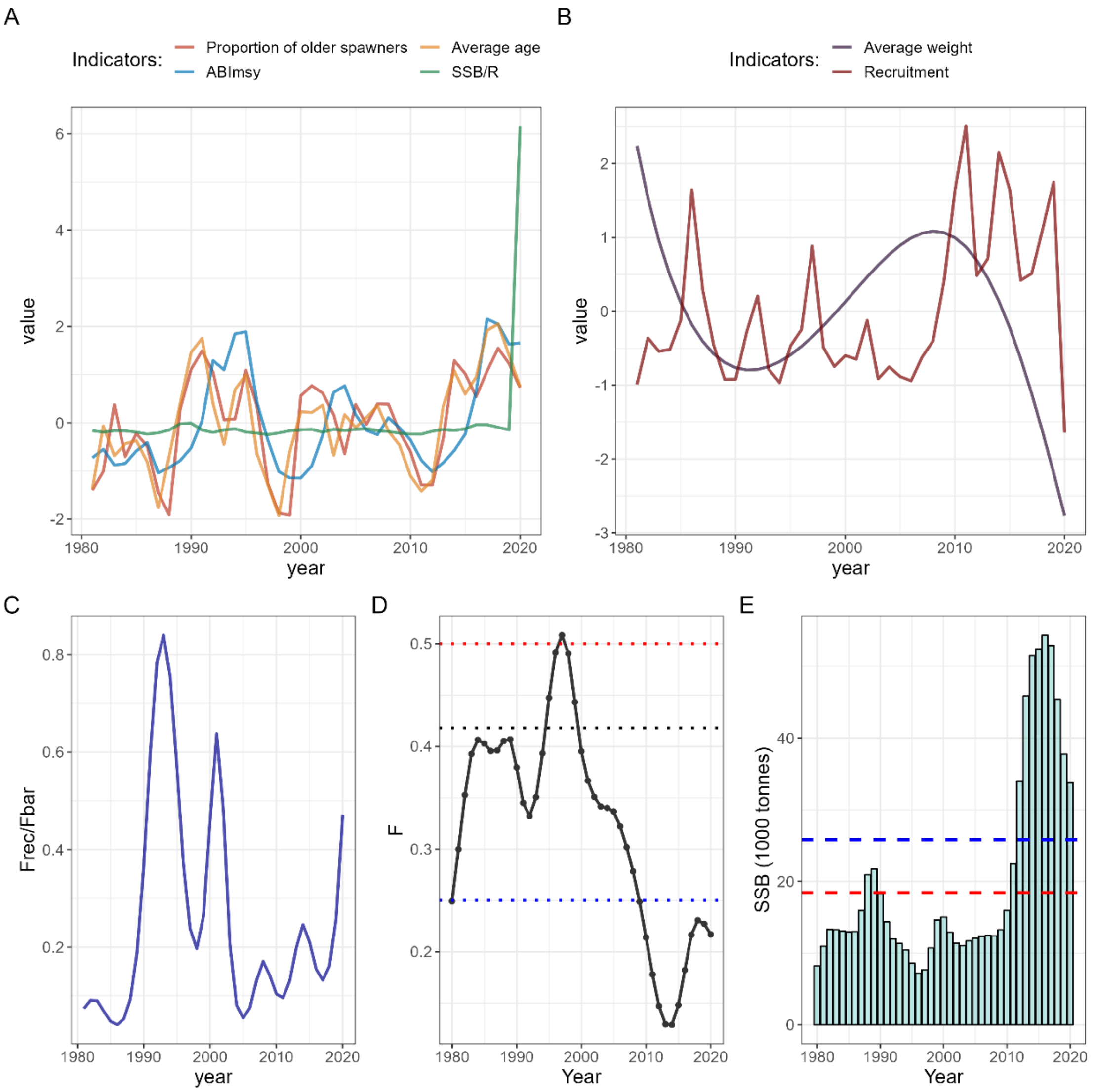


Figure A2.52. Estimated D3C3 indicators and stock status for European plaice (ple.27.7d). See Figure legend A2.1 for further details.

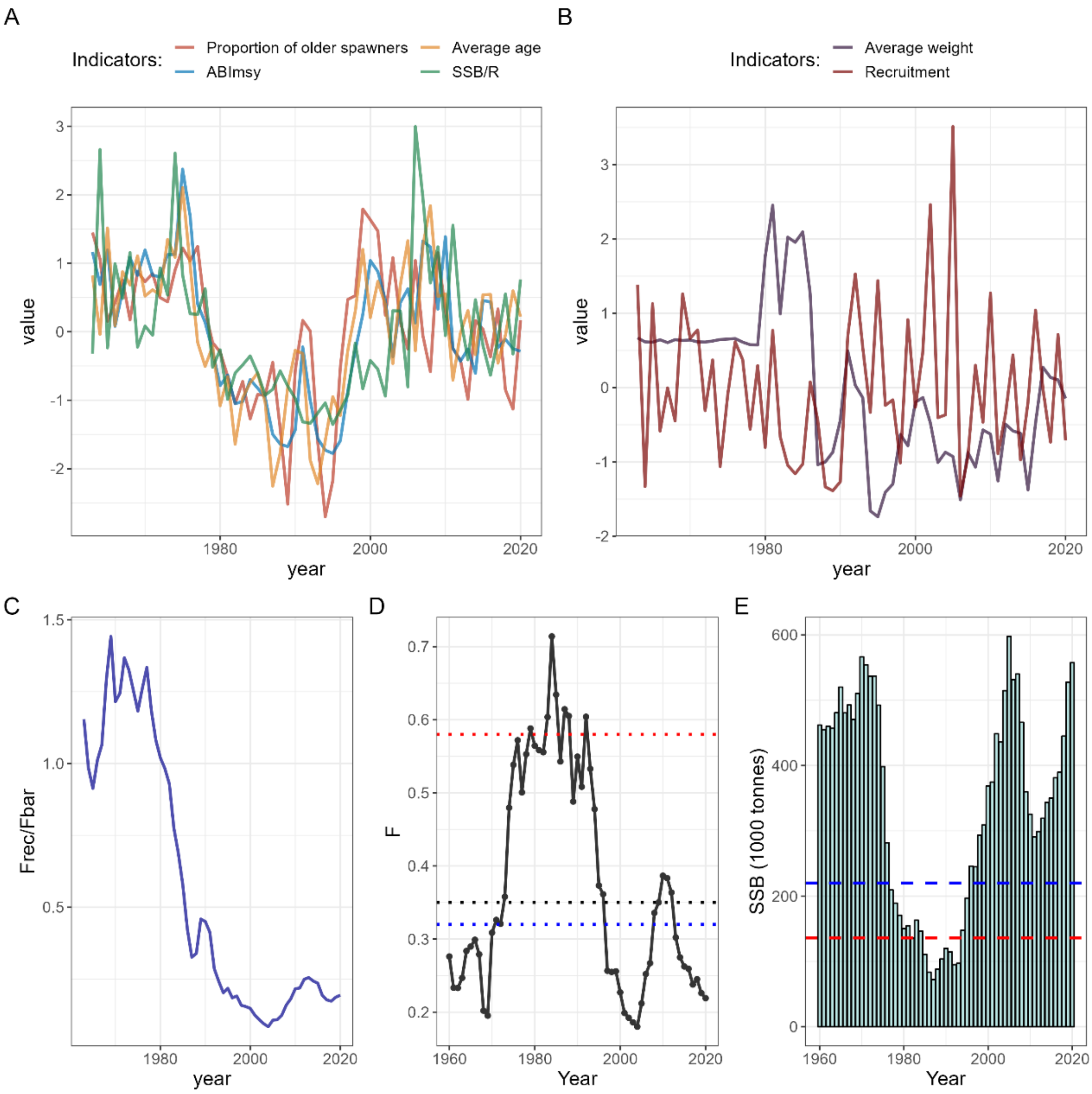


Figure A2.53. Estimated D3C3 indicators and stock status for Saithe (pok.27.1-2). See Figure legend A2.1 for further details.

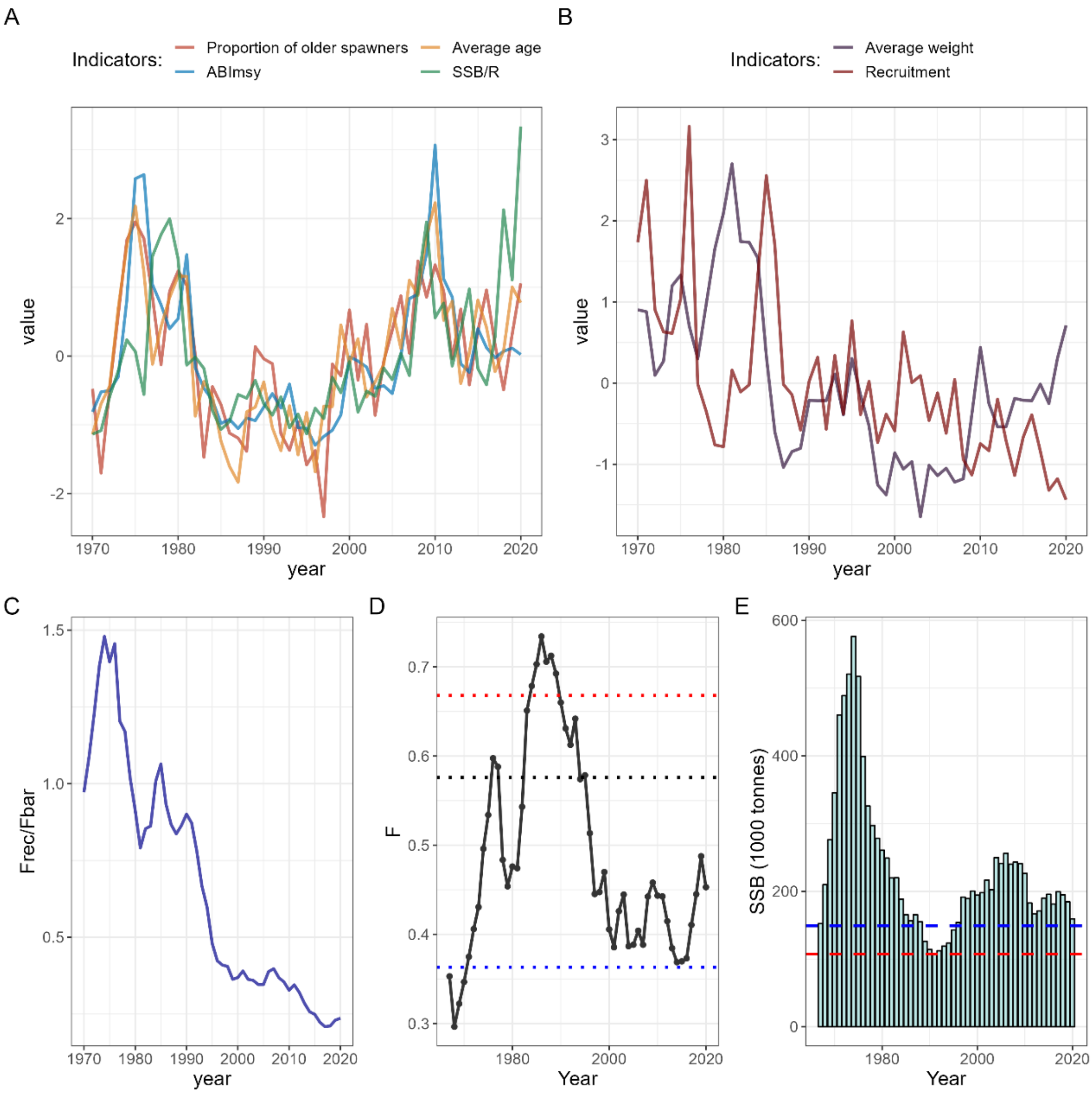


Figure A2.54. Estimated D3C3 indicators and stock status for Saithe (pok.27.3a46). See Figure legend A2.1 for further details.

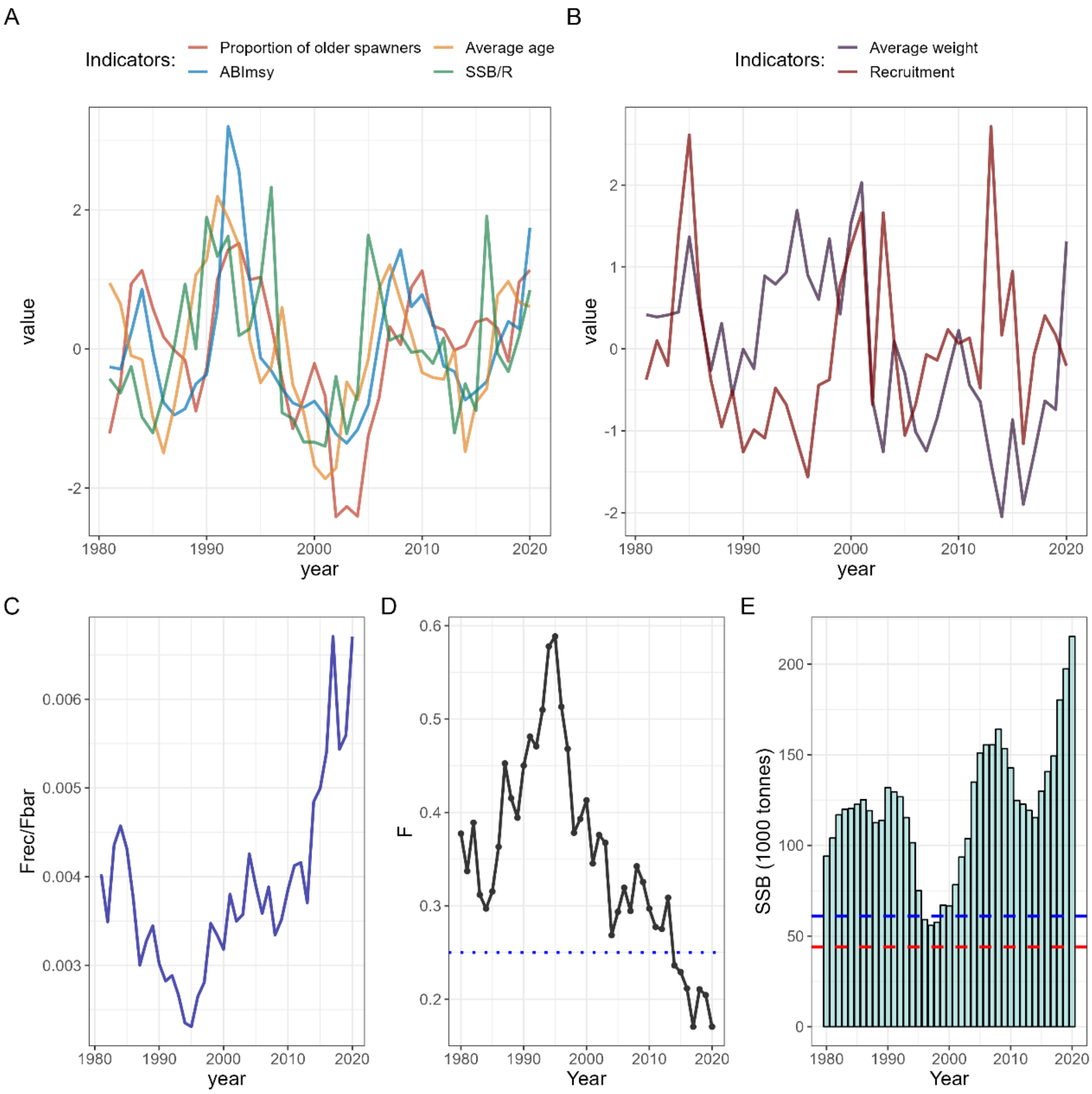


Figure A2.55. Estimated D3C3 indicators and stock status for Saithe (pok.27.5a). See Figure legend A2.1 for further details.

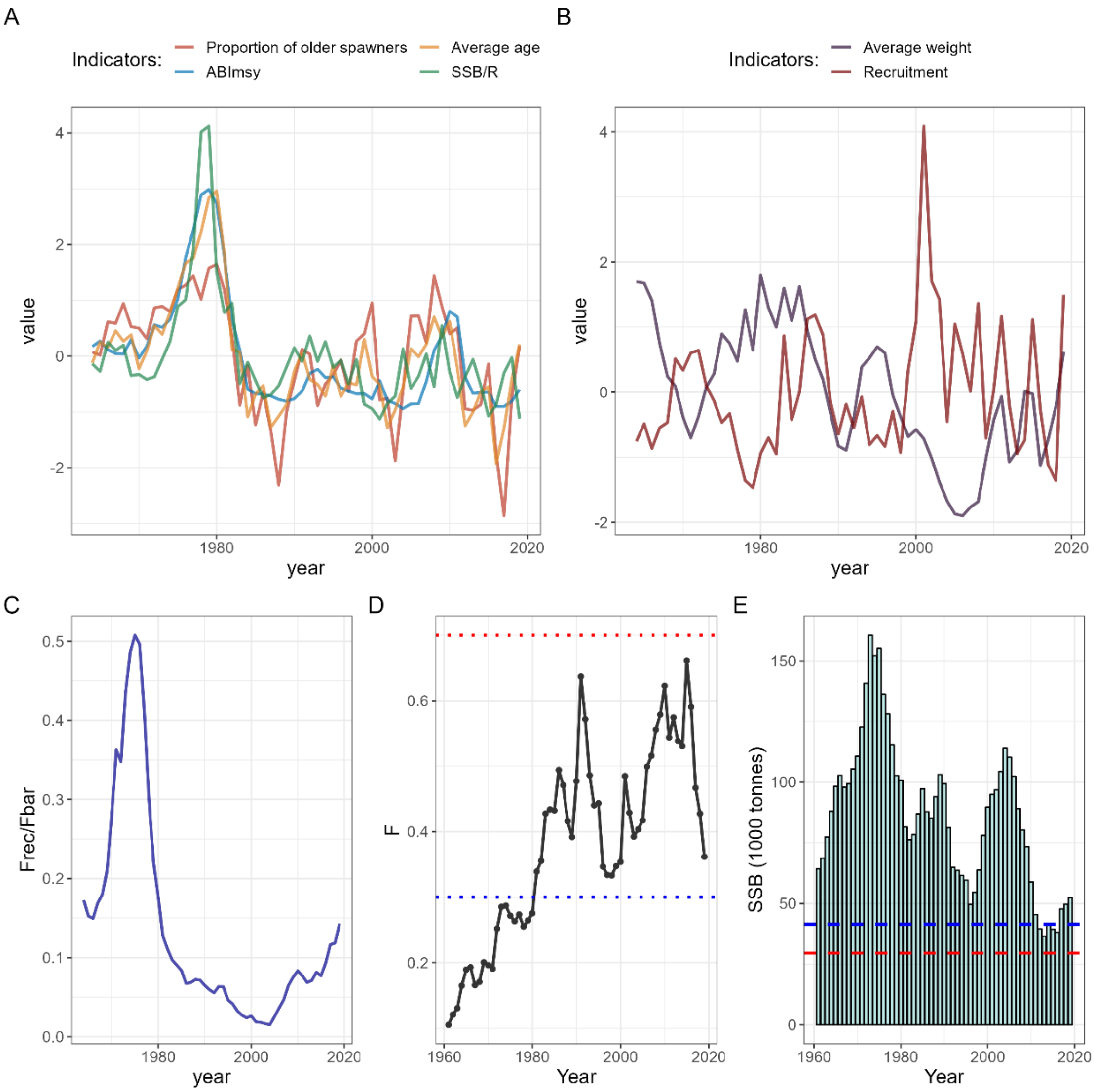


Figure A2.56. Estimated D3C3 indicators and stock status for Saithe (pok.27.5b). See Figure legend A2.1 for further details.

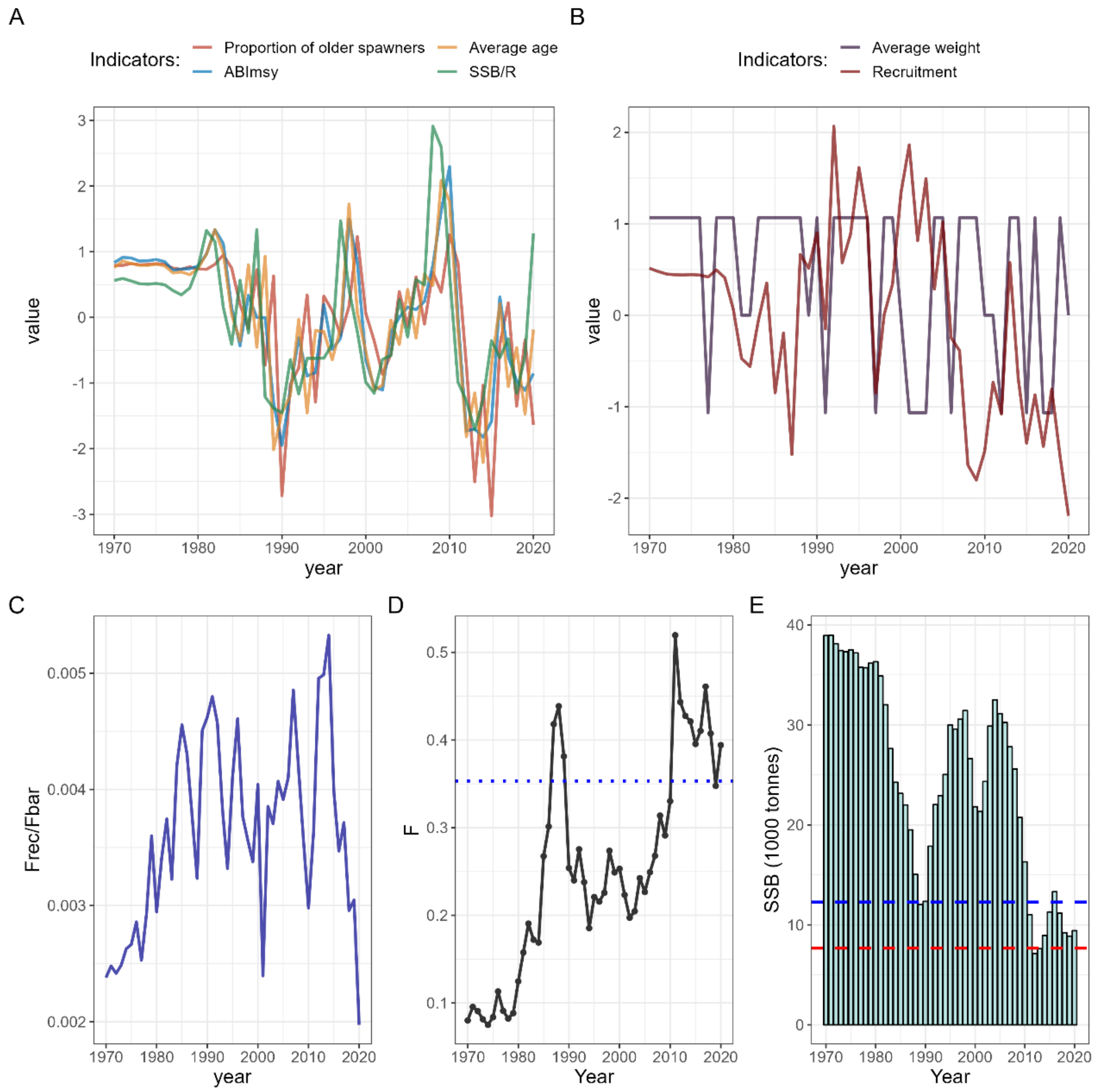


Figure A2.57. Estimated D3C3 indicators and stock status for Northern shrimp (pra.27.3a4a). See Figure legend A2.1 for further details.

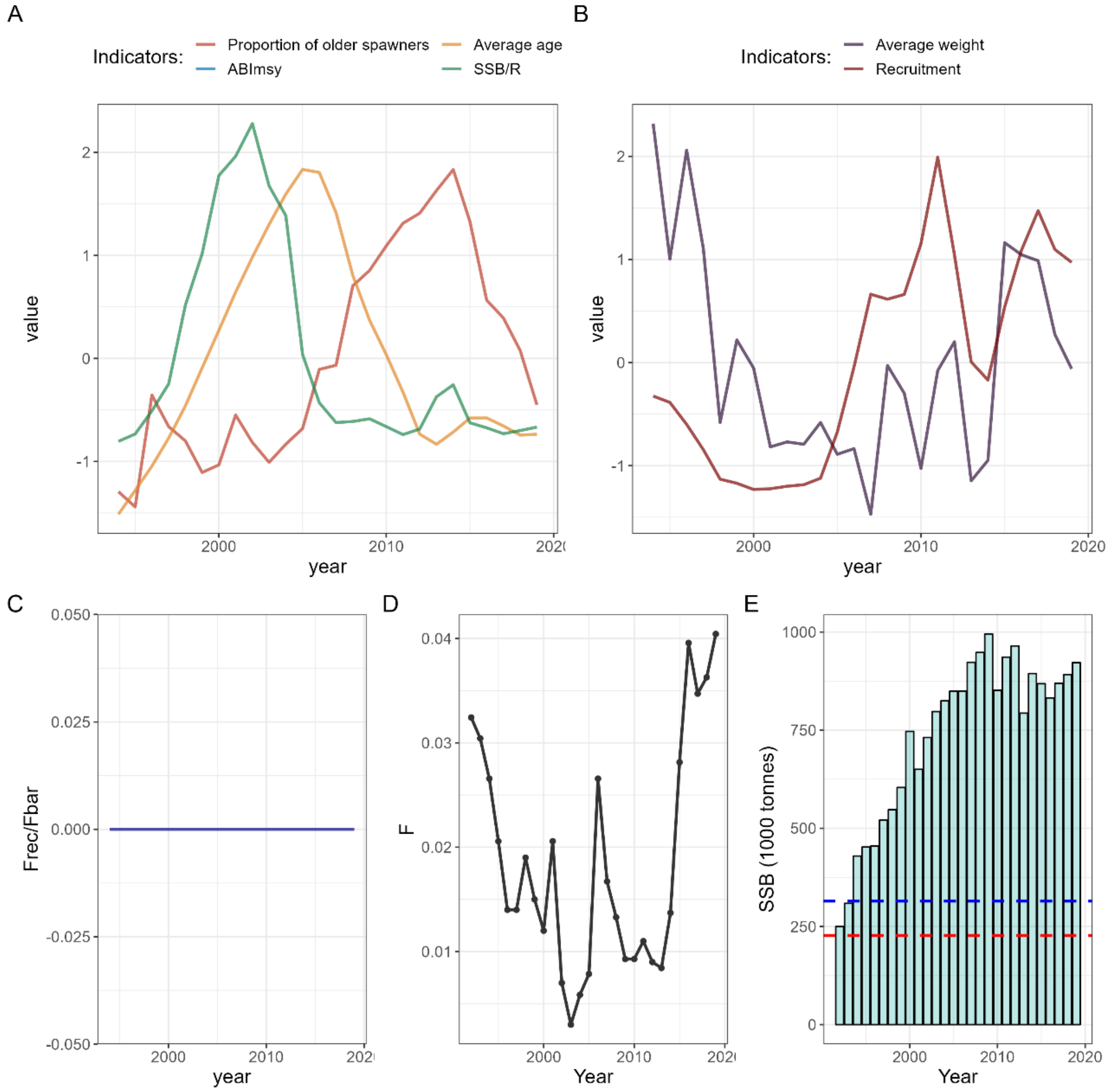


Figure A2.58. Estimated D3C3 indicators and stock status for Beaked redfish (reb.27.1-2). See Figure legend A2.1 for further details.

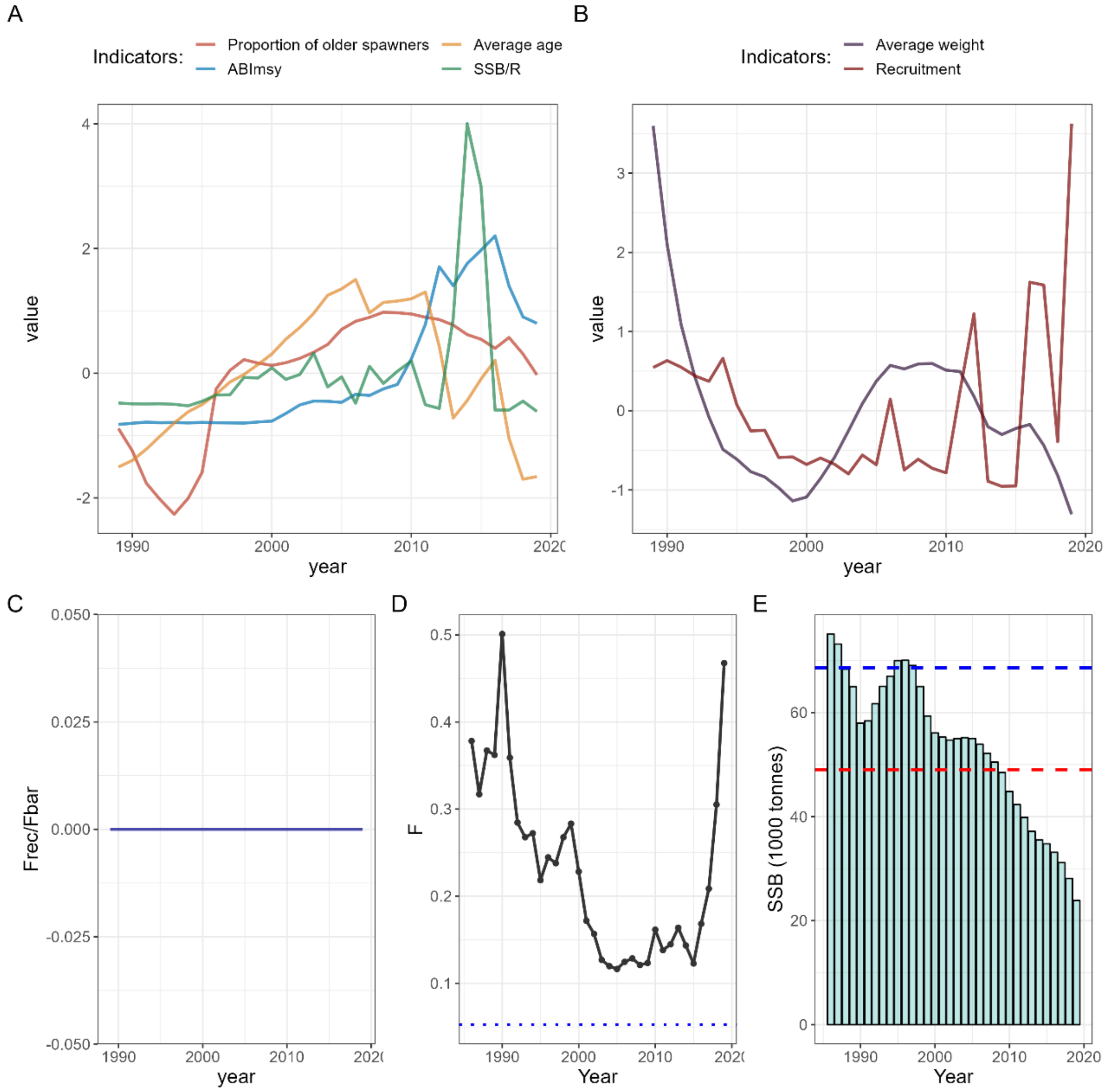


Figure A2.59. Estimated D3C3 indicators and stock status for Golden redfish (reg.27.1-2). See Figure legend A2.1 for further details.

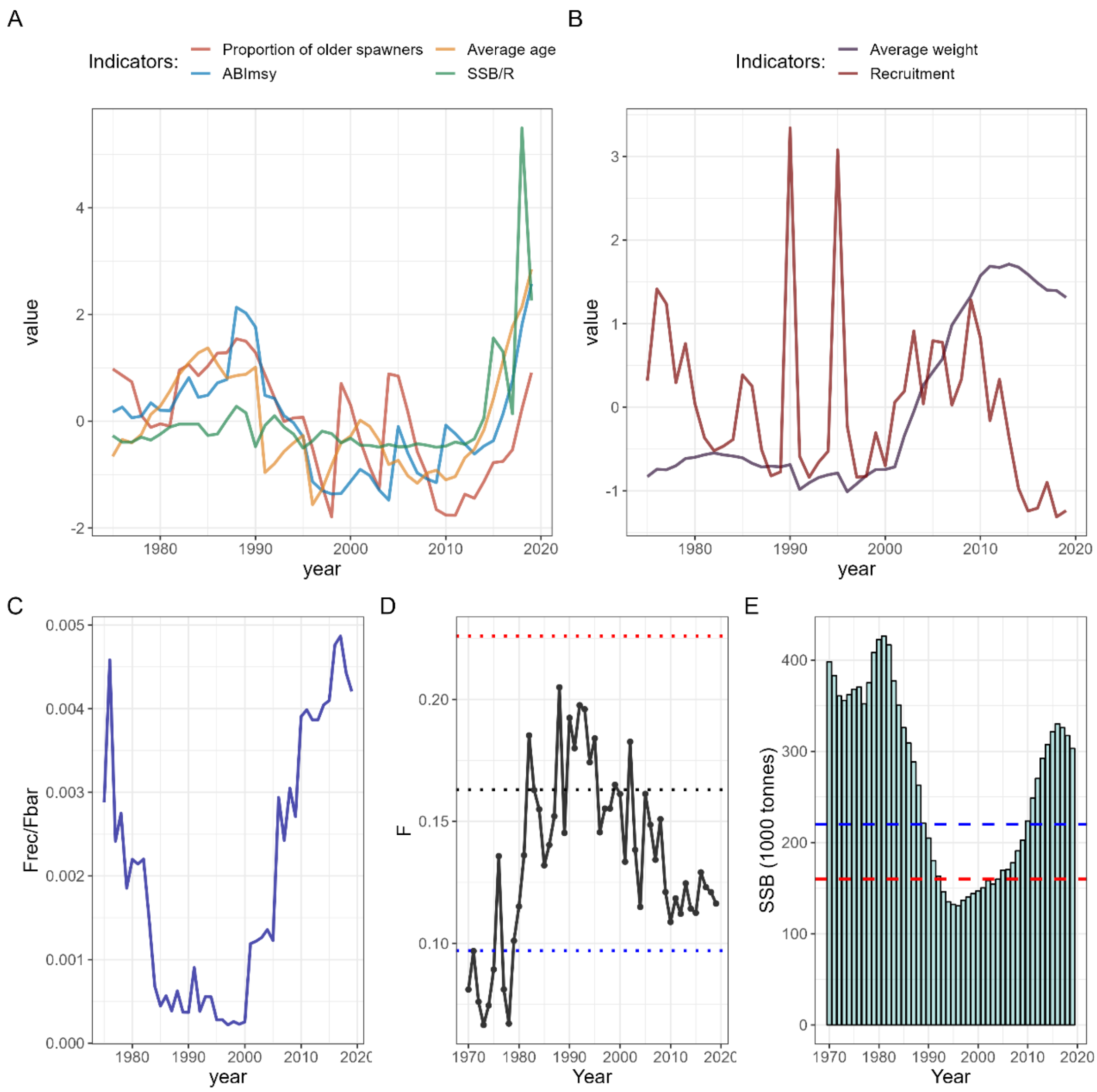


Figure A2.60. Estimated D3C3 indicators and stock status for Golden redfish (reg.27.561214). See Figure legend A2.1 for further details.

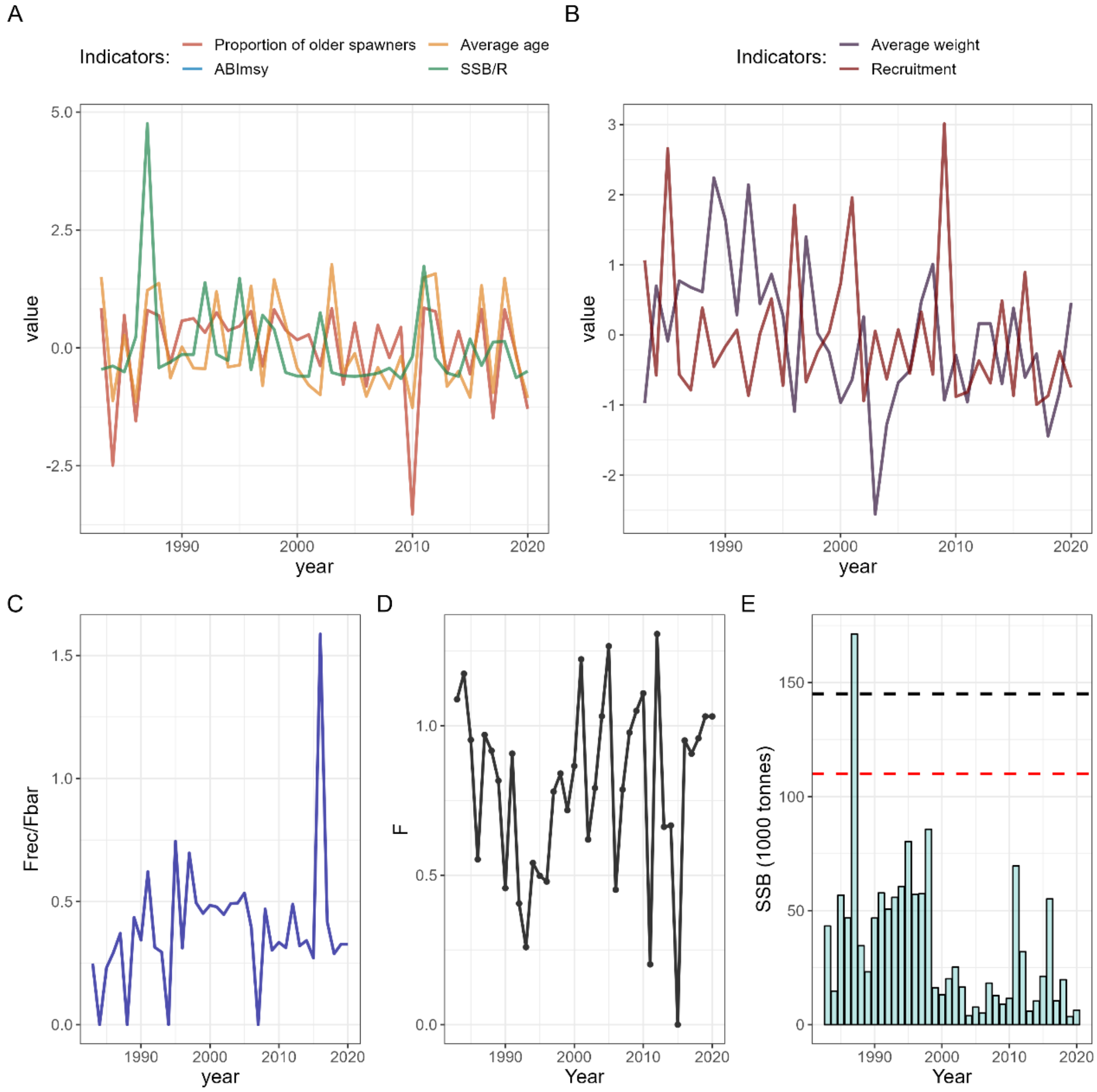


Figure A2.61. Estimated D3C3 indicators and stock status for Sandeel (san.sa.1r). See Figure legend A2.1 for further details.

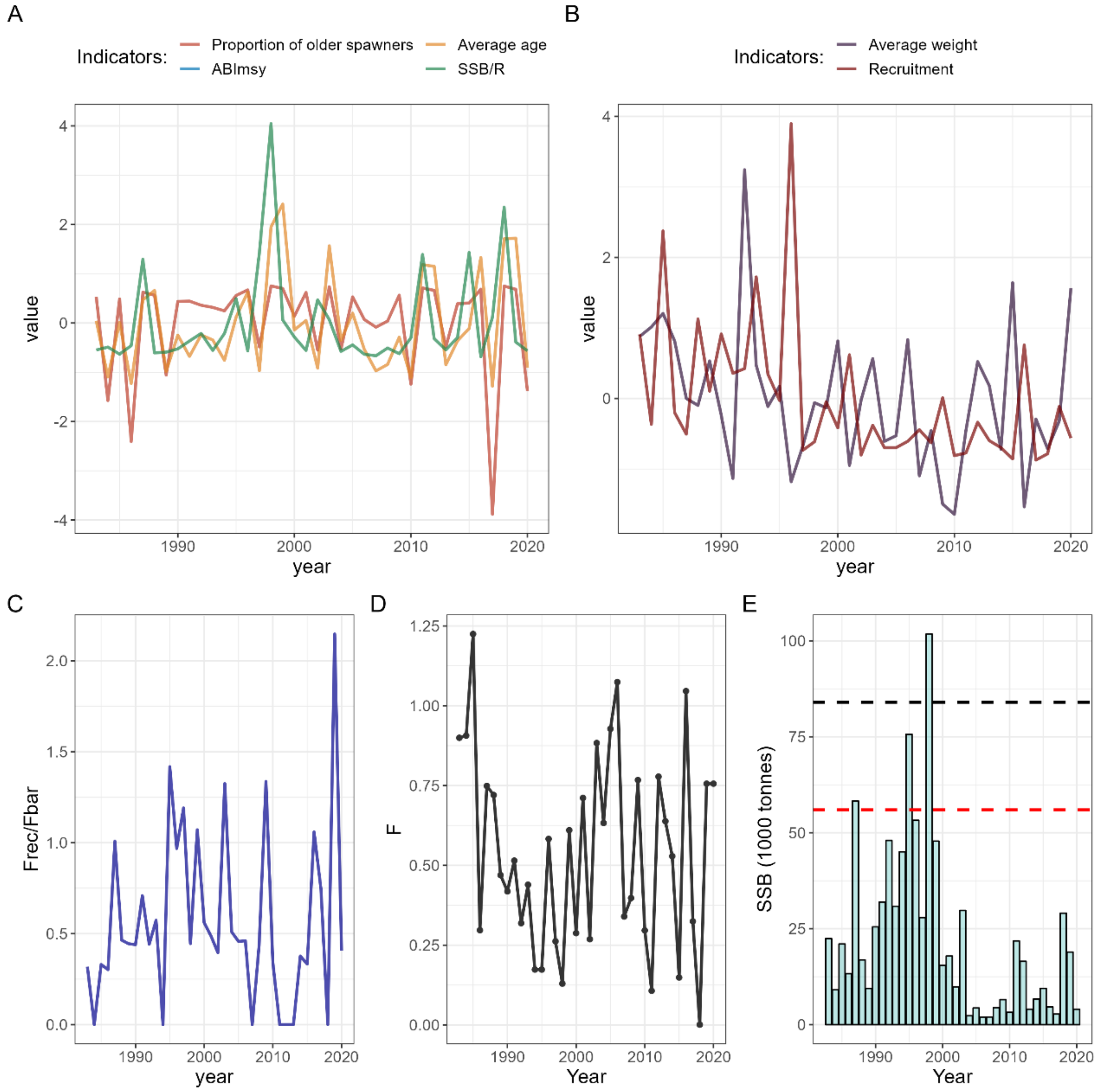


Figure A2.62. Estimated D3C3 indicators and stock status for Sandeel (san.sa.2r). See Figure legend A2.1 for further details.

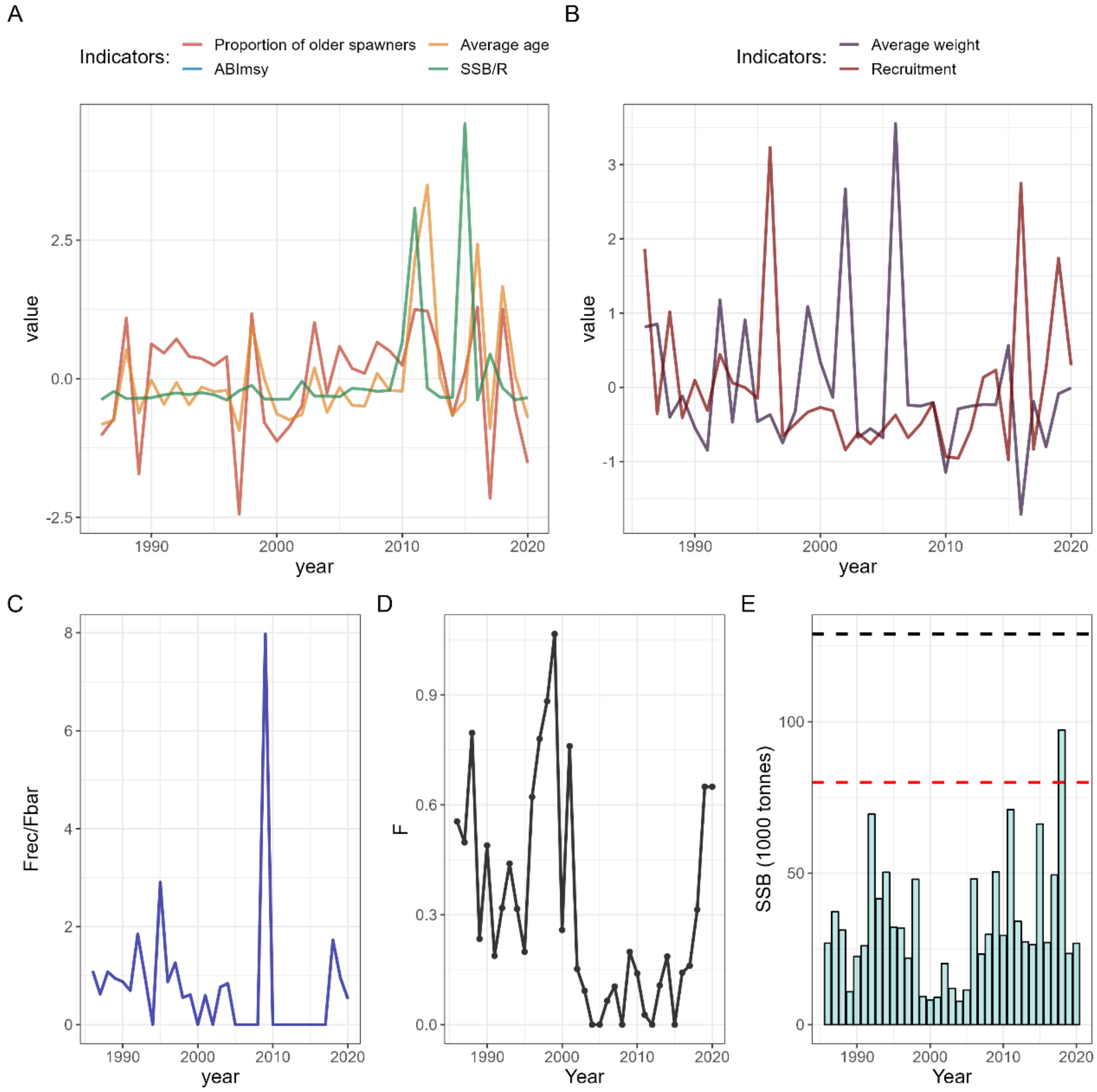


Figure A2.63. Estimated D3C3 indicators and stock status for Sandeel (san.sa.3r). See Figure legend A2.1 for further details.

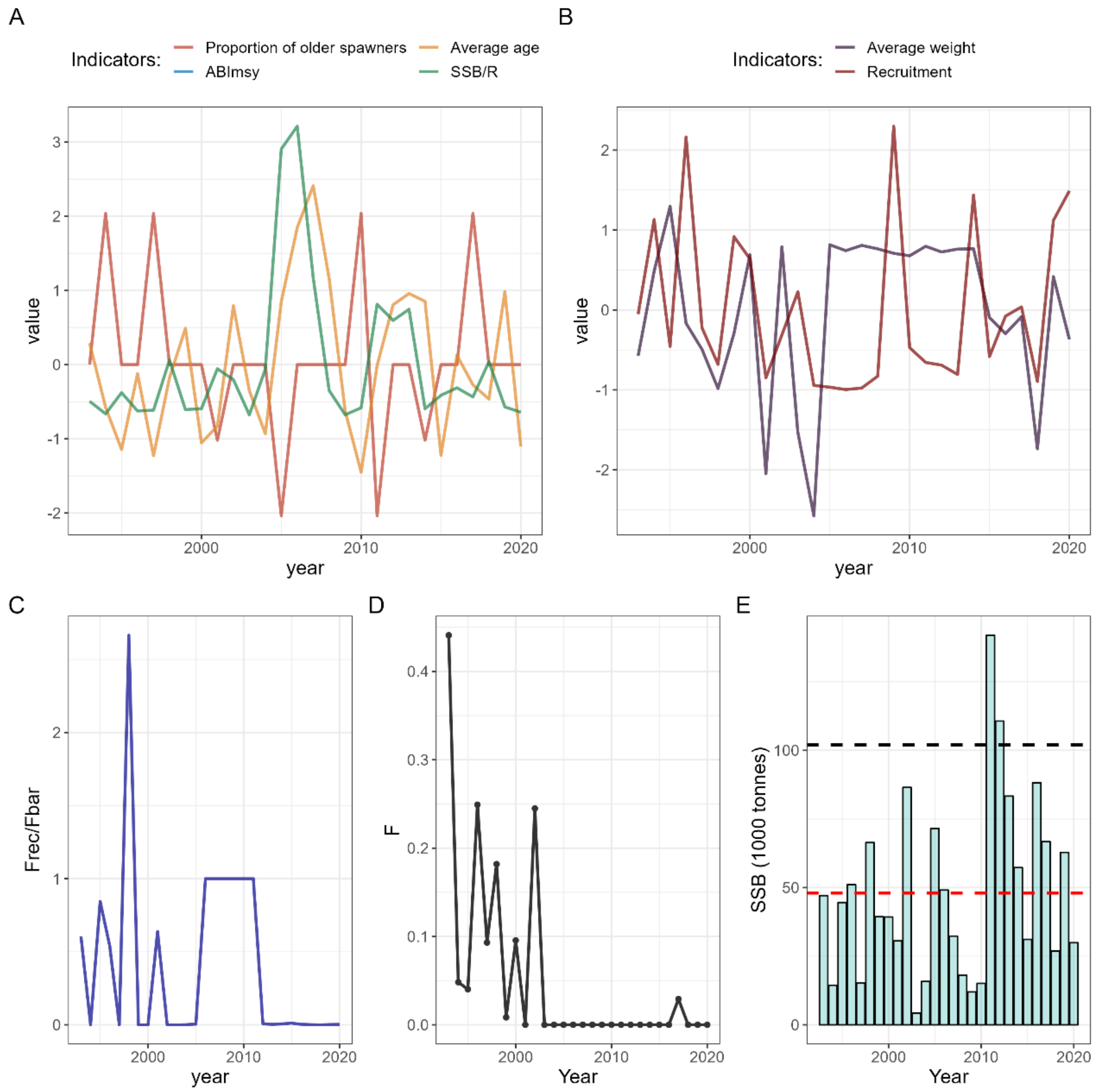


Figure A2.64. Estimated D3C3 indicators and stock status for Sandeel (san.sa.4r). See Figure legend A2.1 for further details.

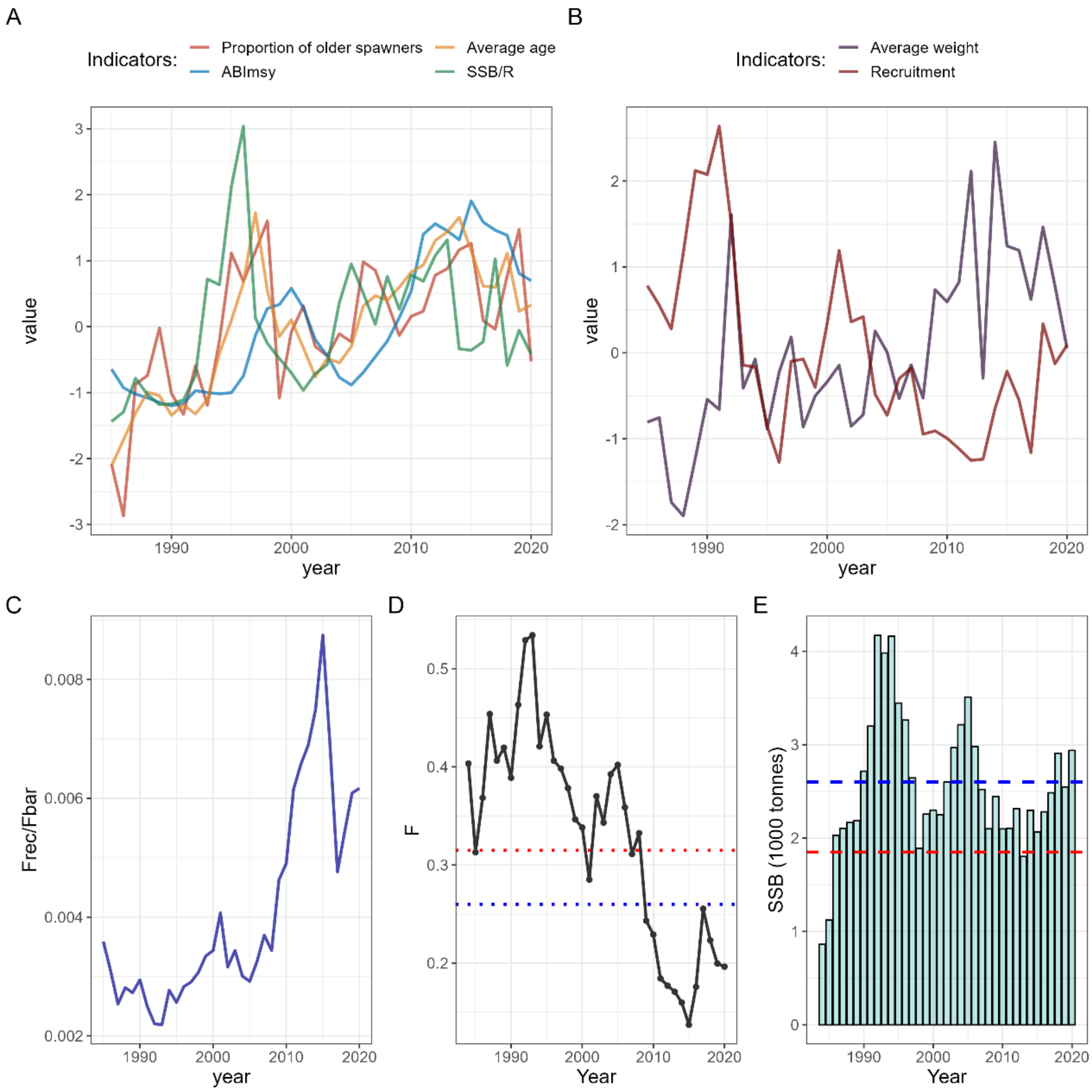


Figure A2.65. Estimated D3C3 indicators and stock status for Sole (sol.27.20-24). See Figure legend A2.1 for further details.

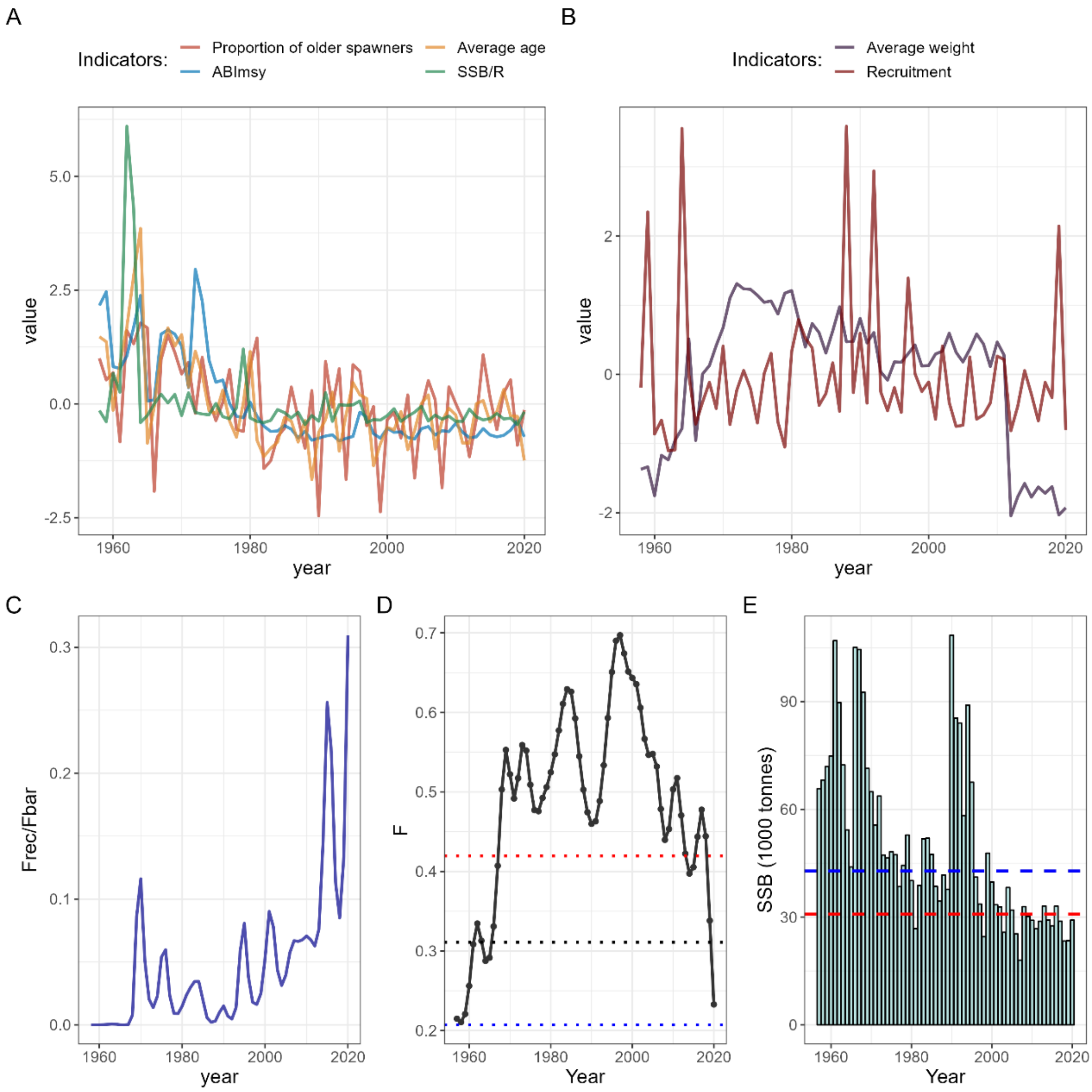


Figure A2.66. Estimated D3C3 indicators and stock status for Sole (sol.27.4). See Figure legend A2.1 for further details.

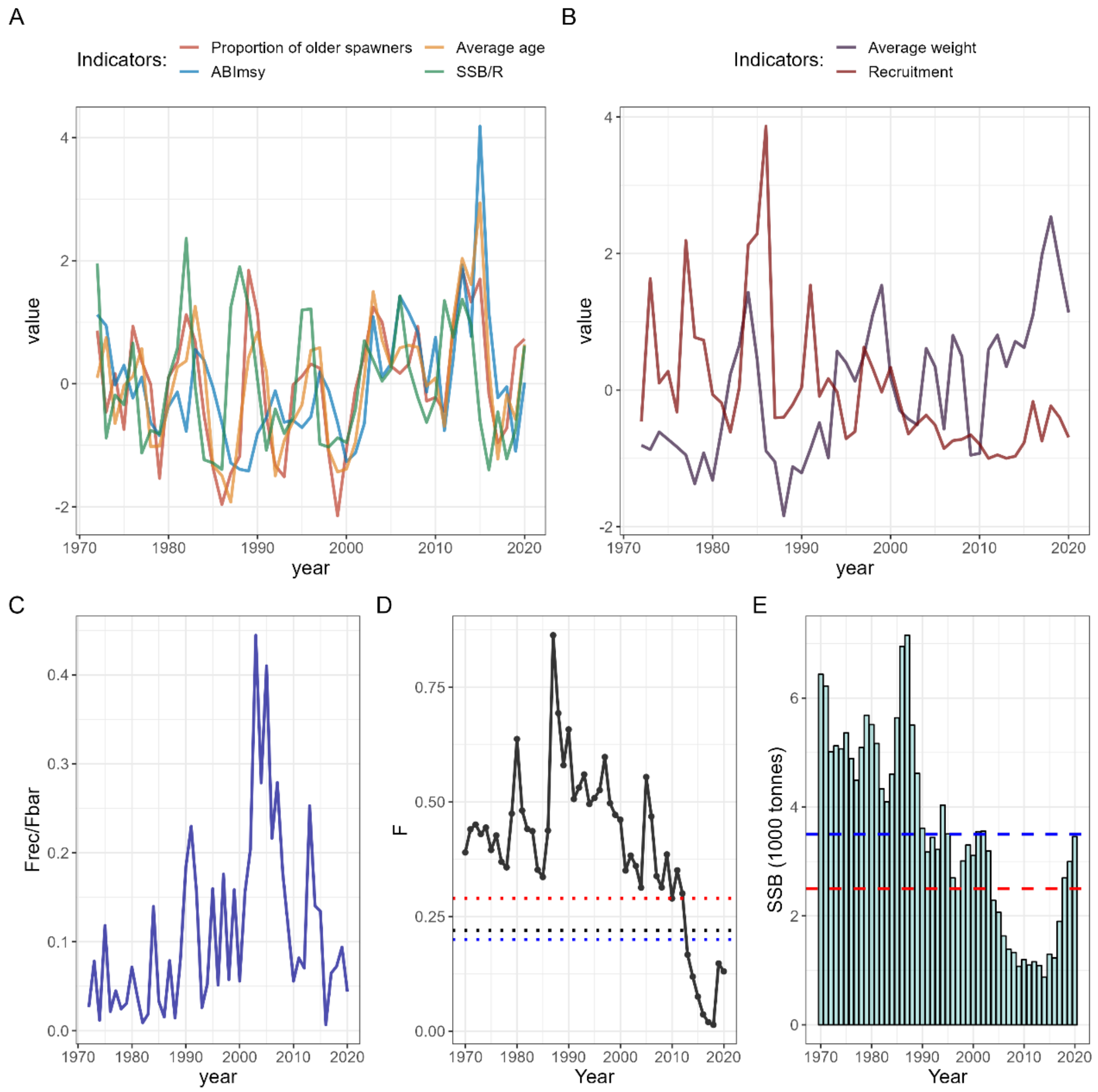


Figure A2.67. Estimated D3C3 indicators and stock status for Sole (sol.27.7a). See Figure legend A2.1 for further details.

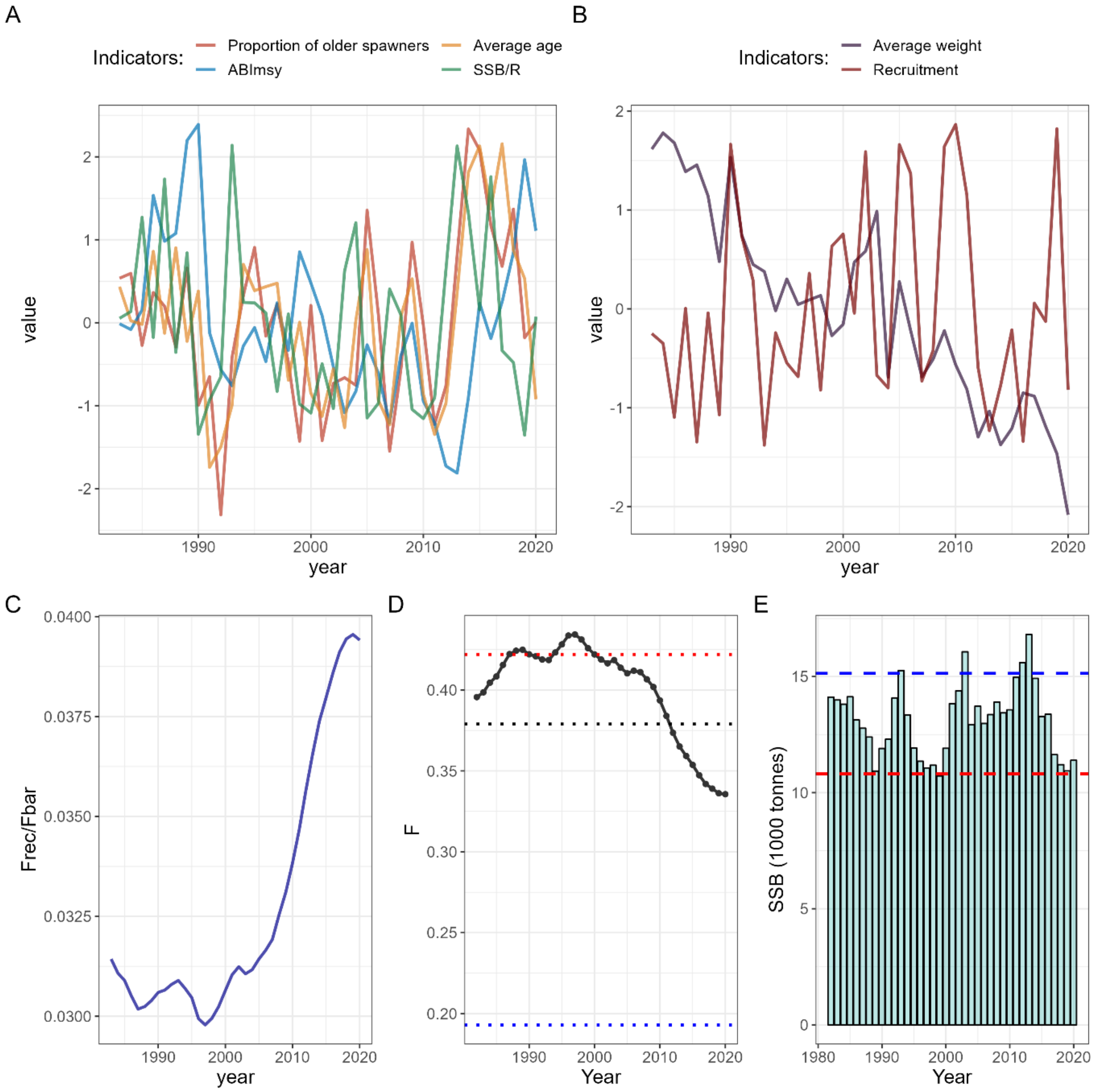


Figure A2.68. Estimated D3C3 indicators and stock status for Sole (sol.27.7d). See Figure legend A2.1 for further details.

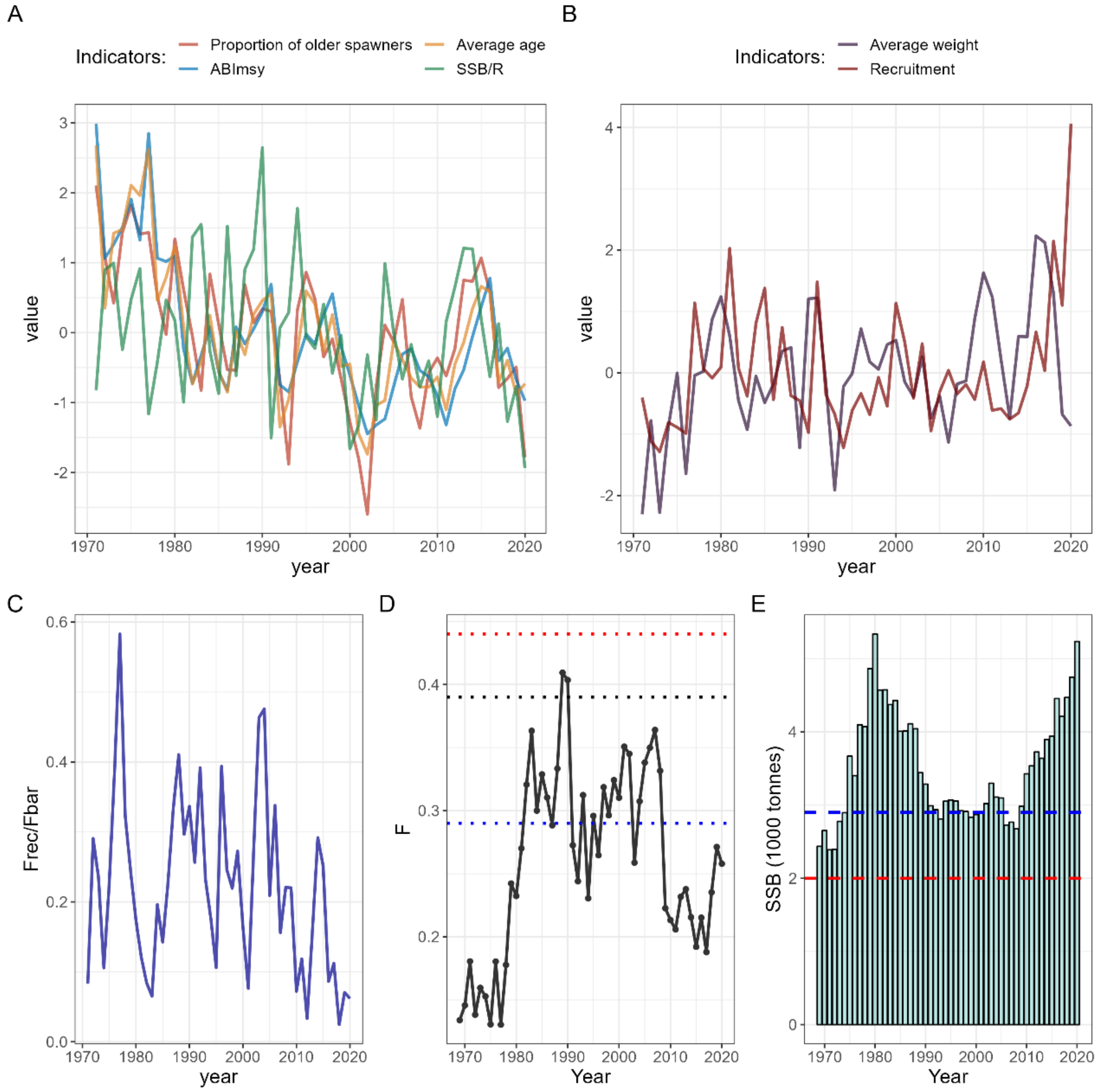


Figure A2.69. Estimated D3C3 indicators and stock status for Sole (sol.27.7e). See Figure legend A2.1 for further details.

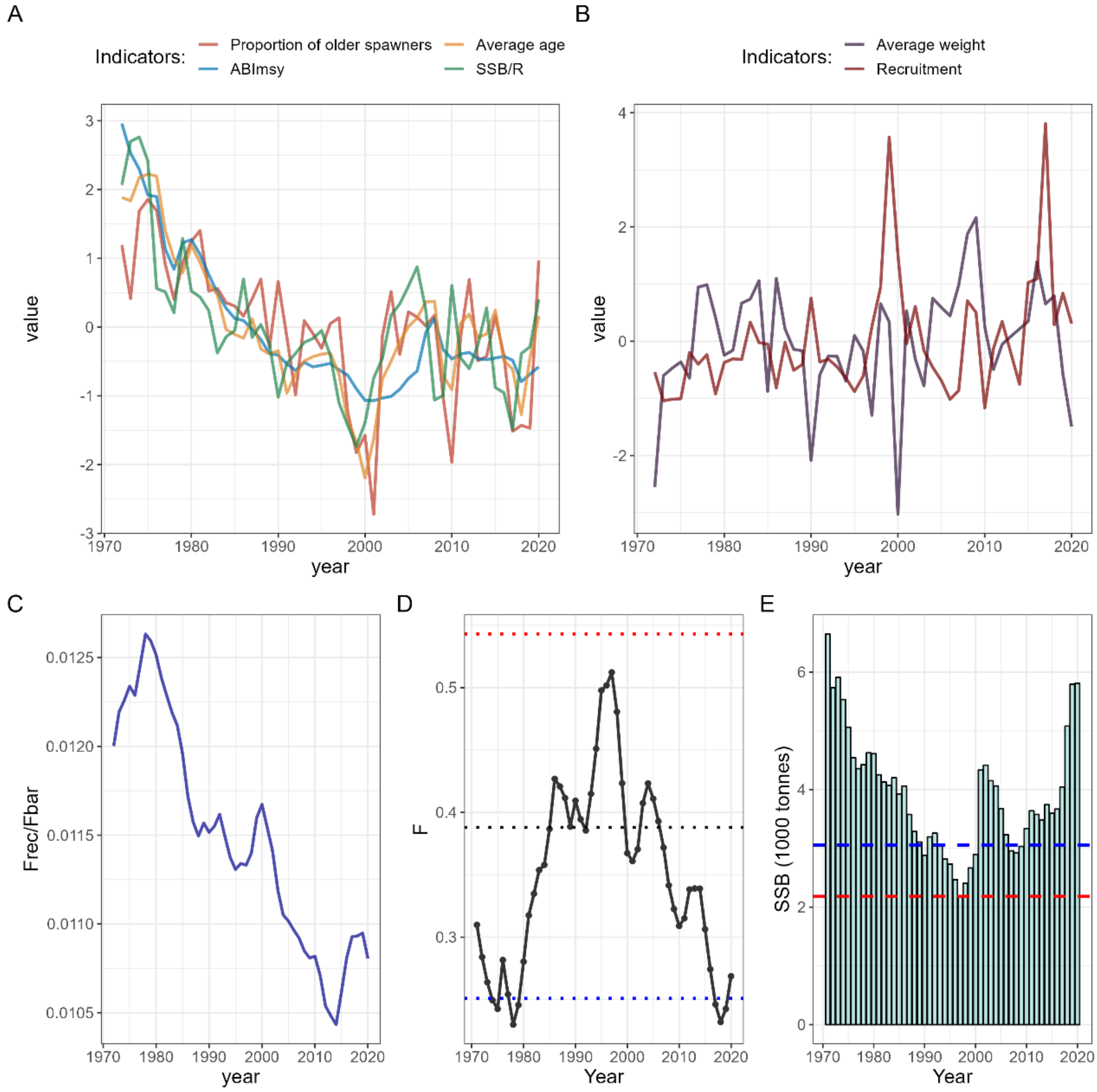


Figure A2.70. Estimated D3C3 indicators and stock status for Sole (sol.27.7fg). See Figure legend A2.1 for further details.

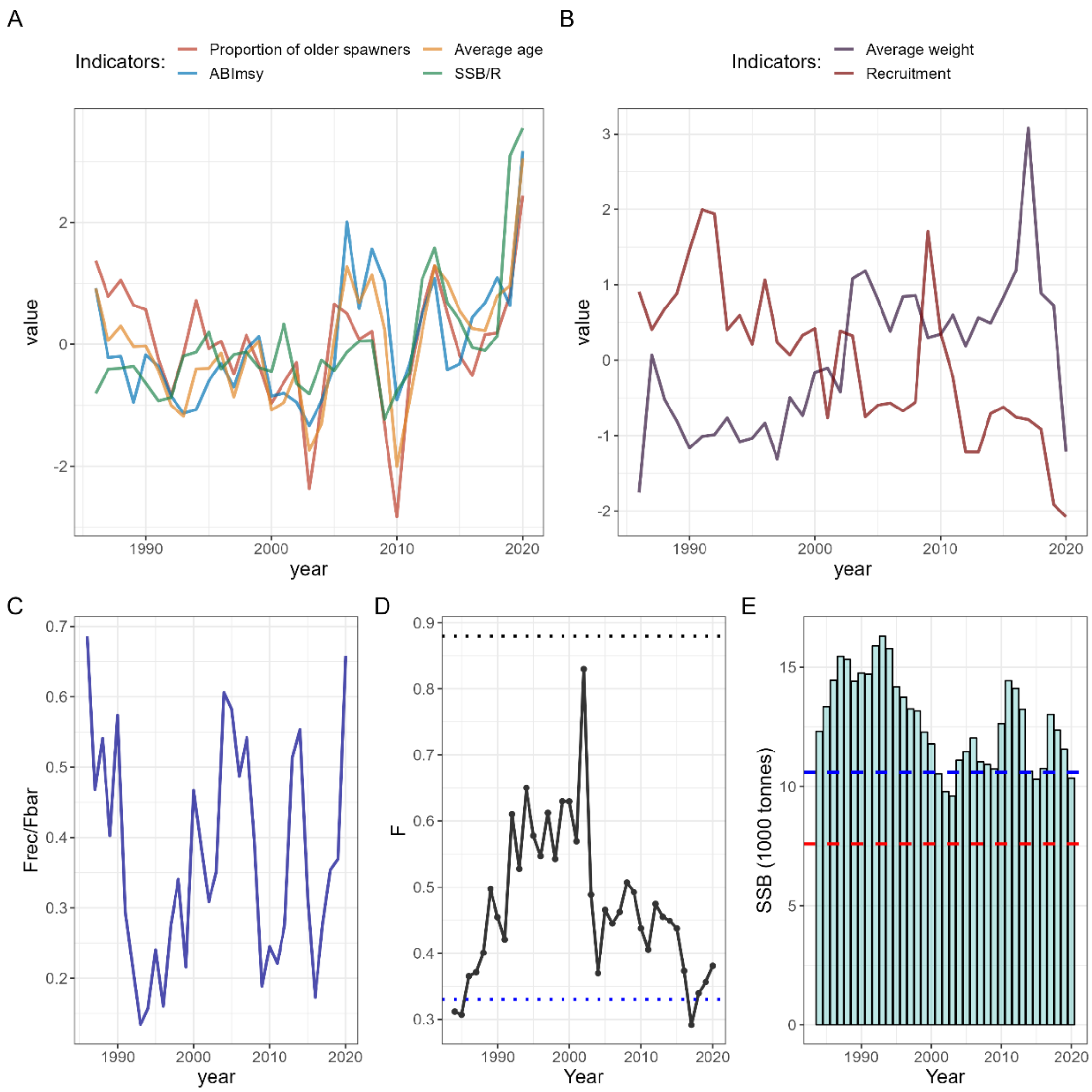


Figure A2.71. Estimated D3C3 indicators and stock status for Sole (sol.27.8ab). See Figure legend A2.1 for further details.

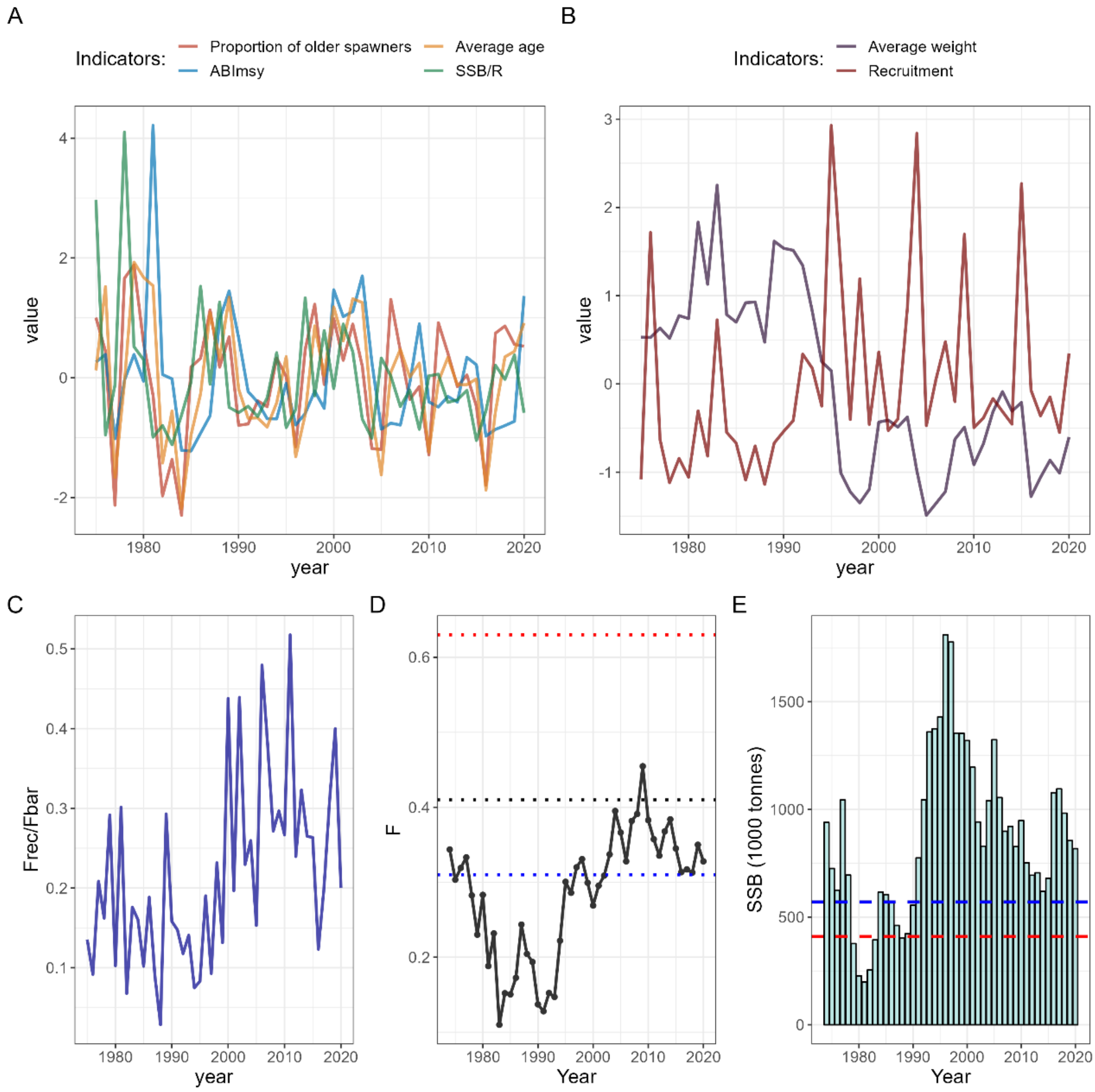


Figure A2.72. Estimated D3C3 indicators and stock status for Sprat (sol.27.22-32). See Figure legend A2.1 for further details.

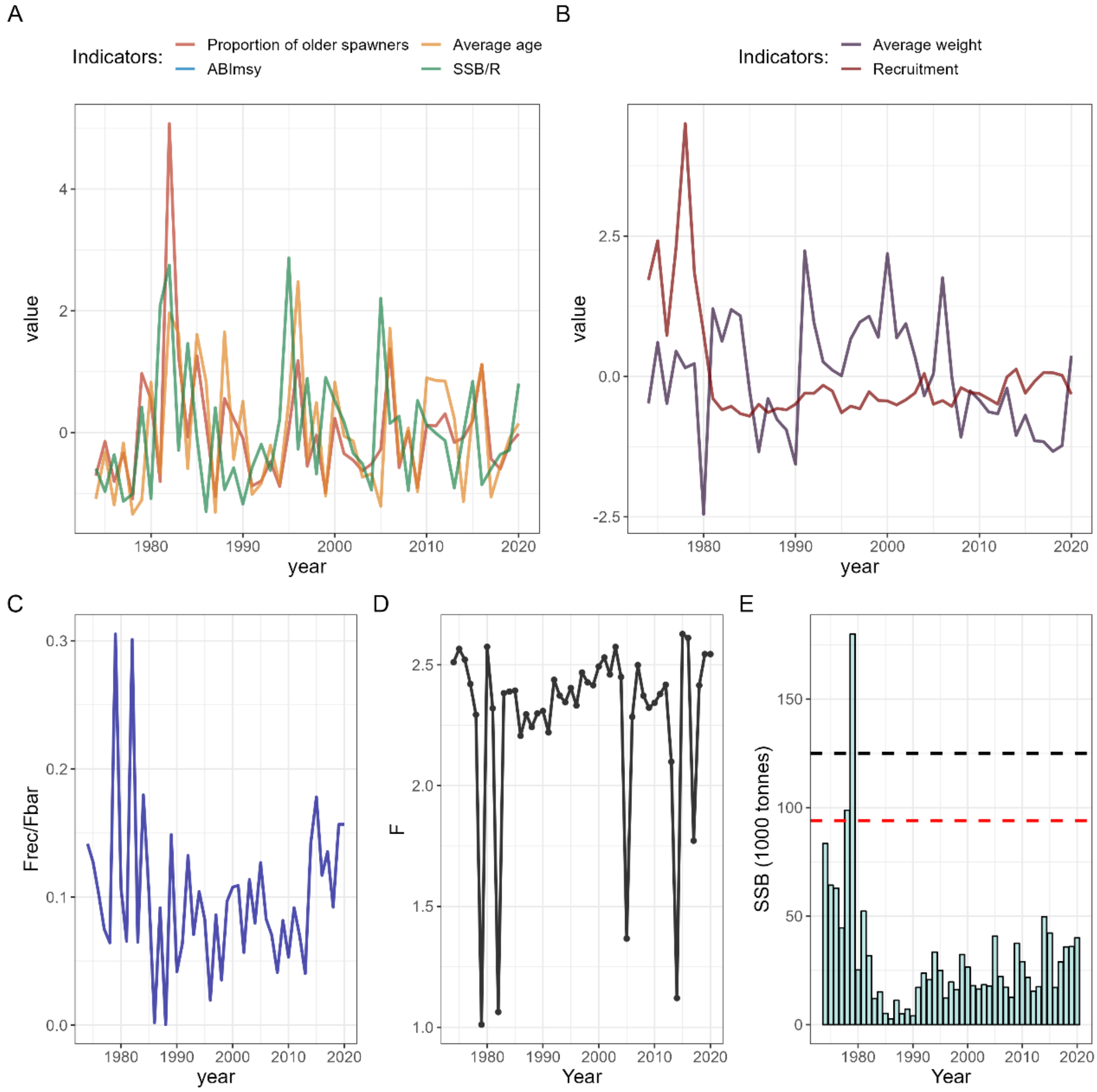


Figure A2.73. Estimated D3C3 indicators and stock status for Sprat (sol.27.3a4). See Figure legend A2.1 for further details.

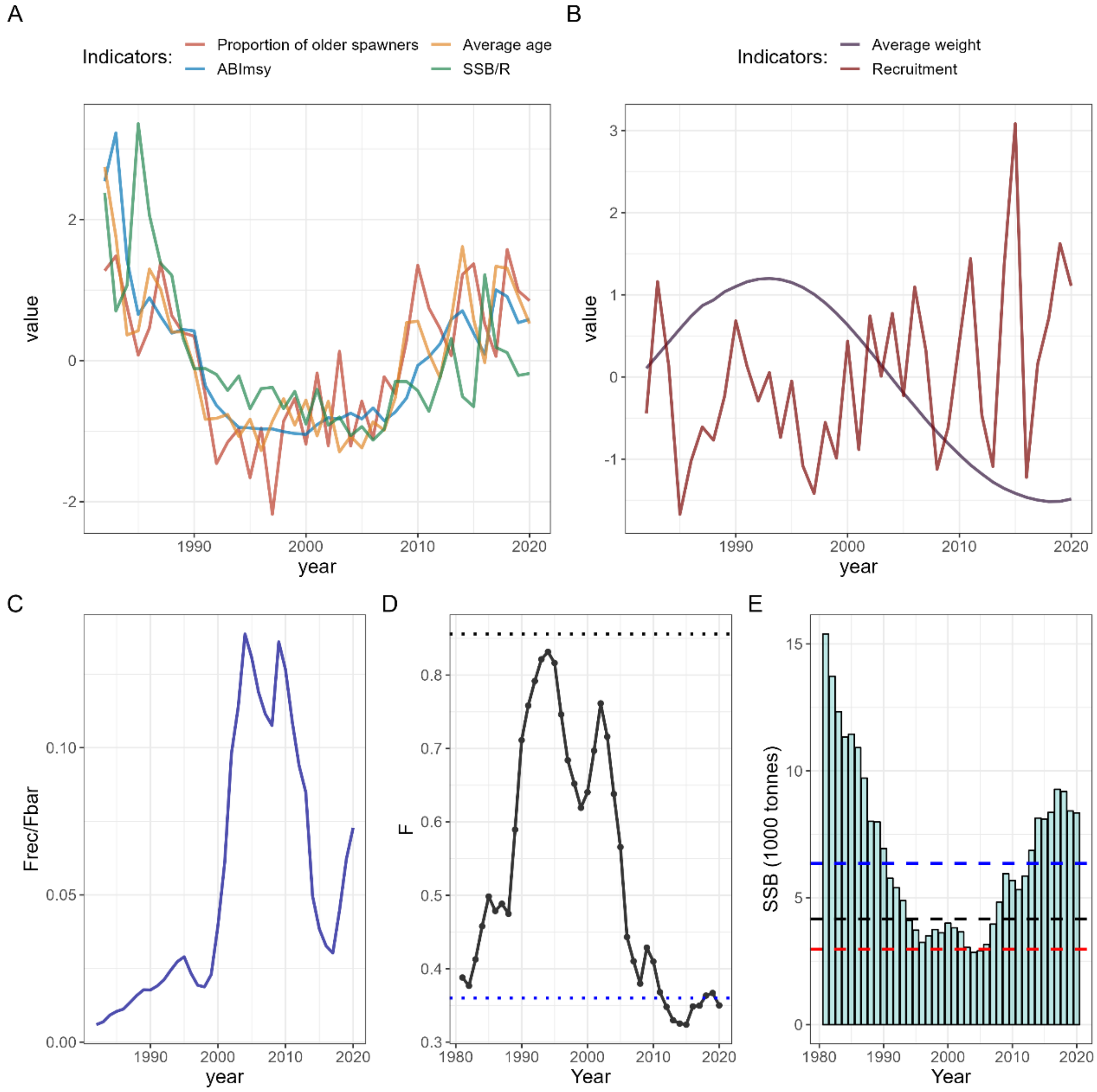


Figure A2.74. Estimated D3C3 indicators and stock status for Turbot (tur.27.4). See Figure legend A2.1 for further details.

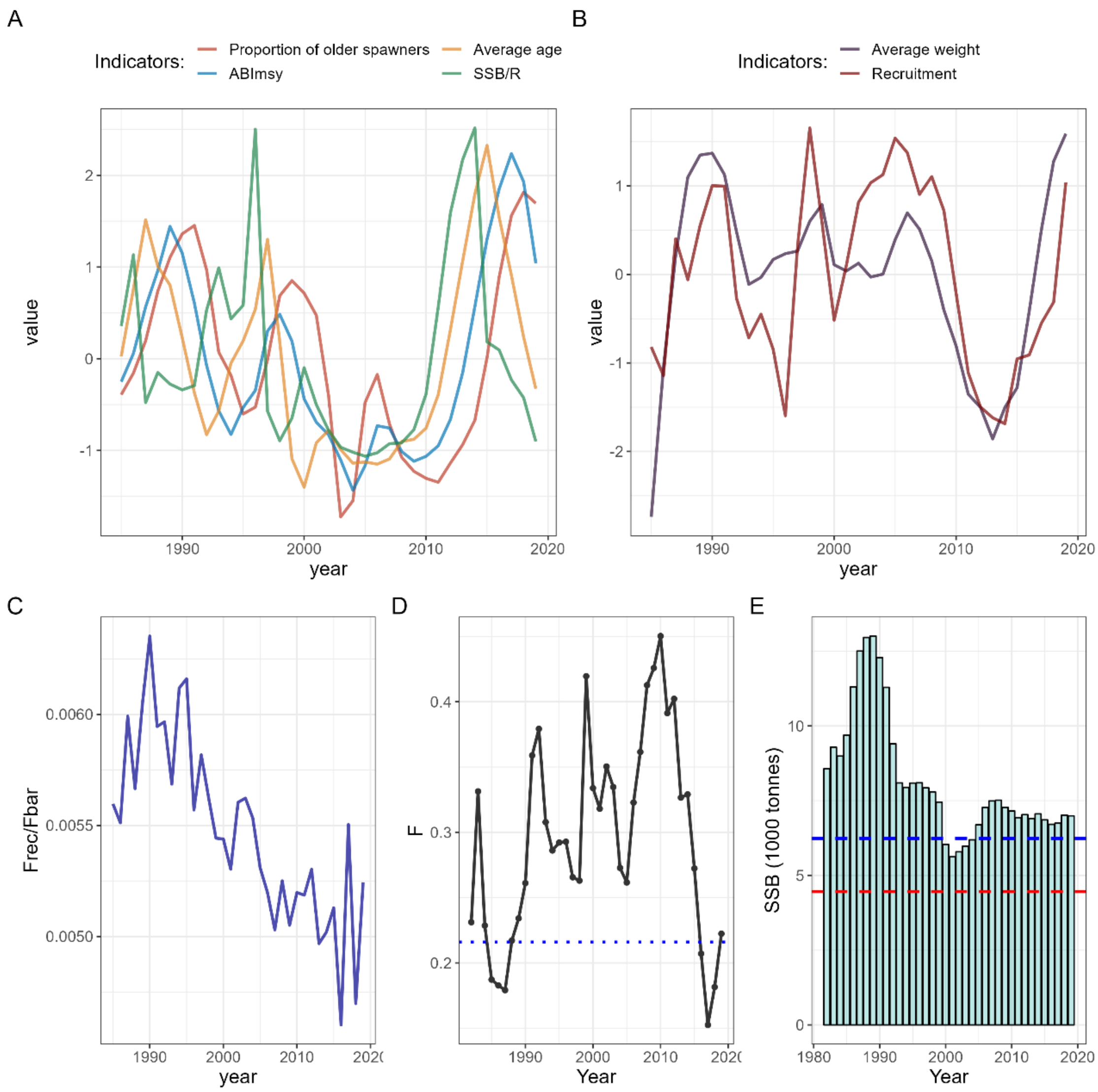


Figure A2.75. Estimated D3C3 indicators and stock status for Tusk (usk.27.5a14). See Figure legend A2.1 for further details.

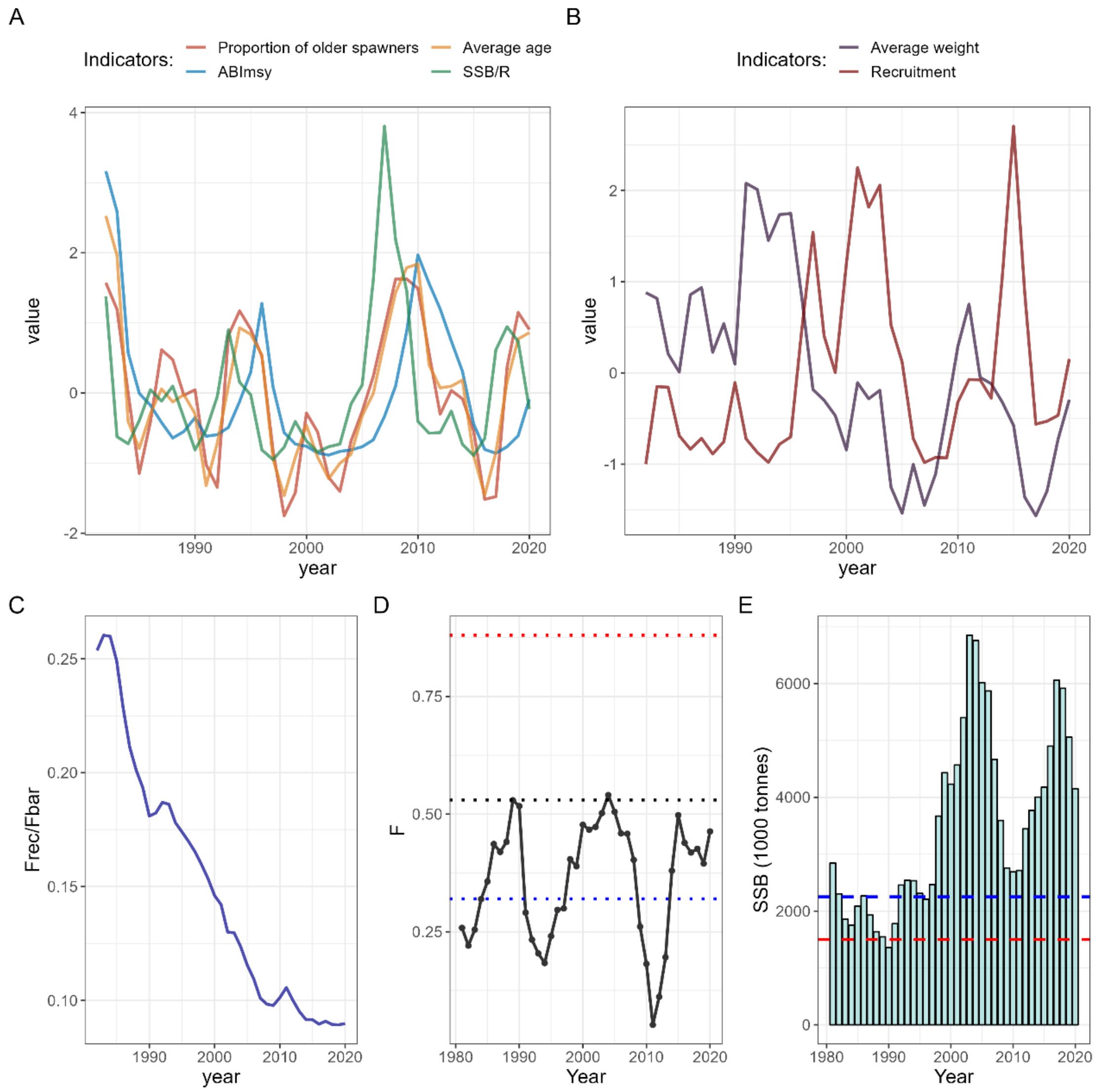


Figure A2.76. Estimated D3C3 indicators and stock status for Blue whiting (whb.27.1-91214). See Figure legend A2.1 for further details.

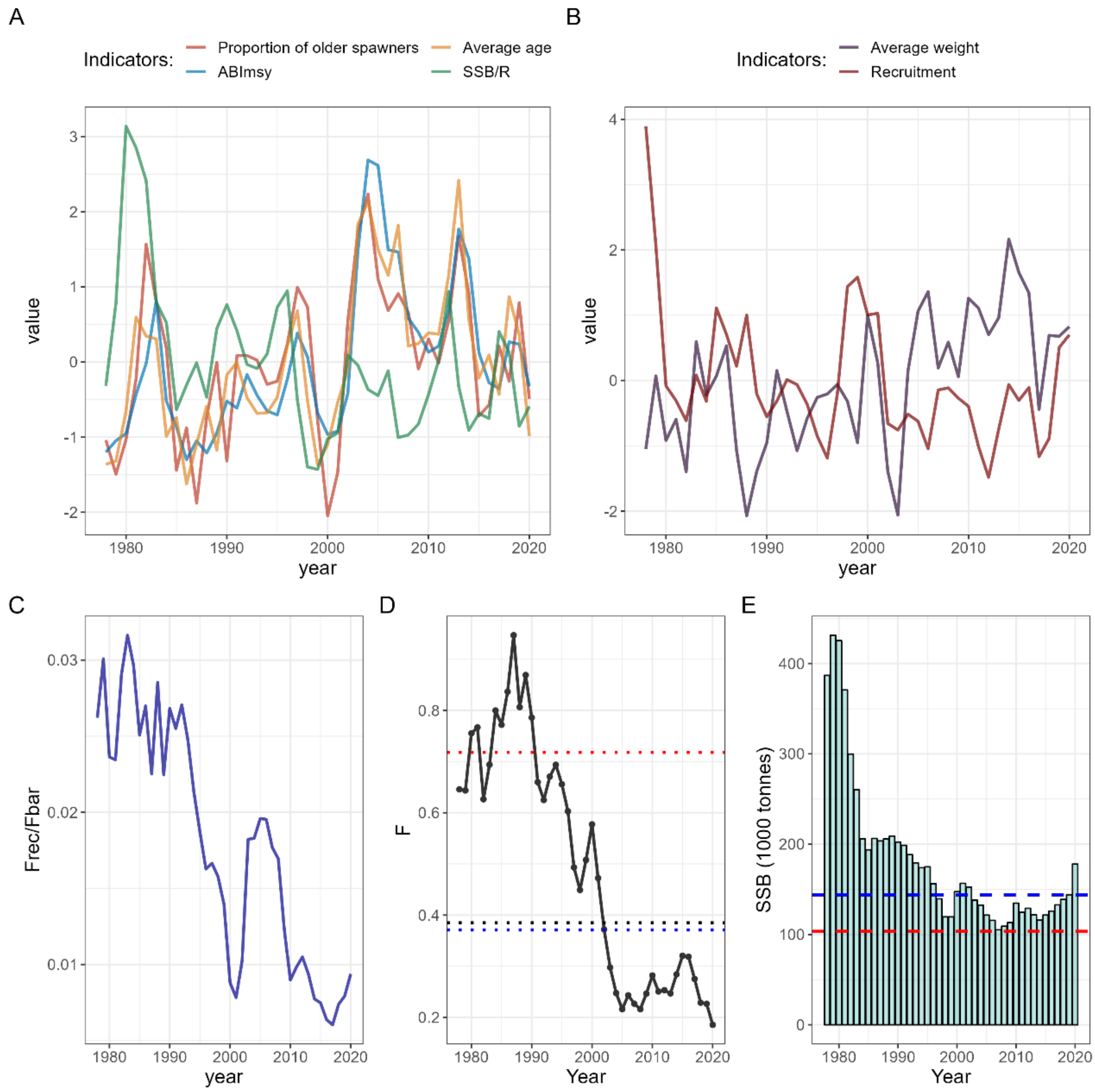


Figure A2.77. Estimated D3C3 indicators and stock status for Whiting (whg.27.47d). See Figure legend A2.1 for further details.

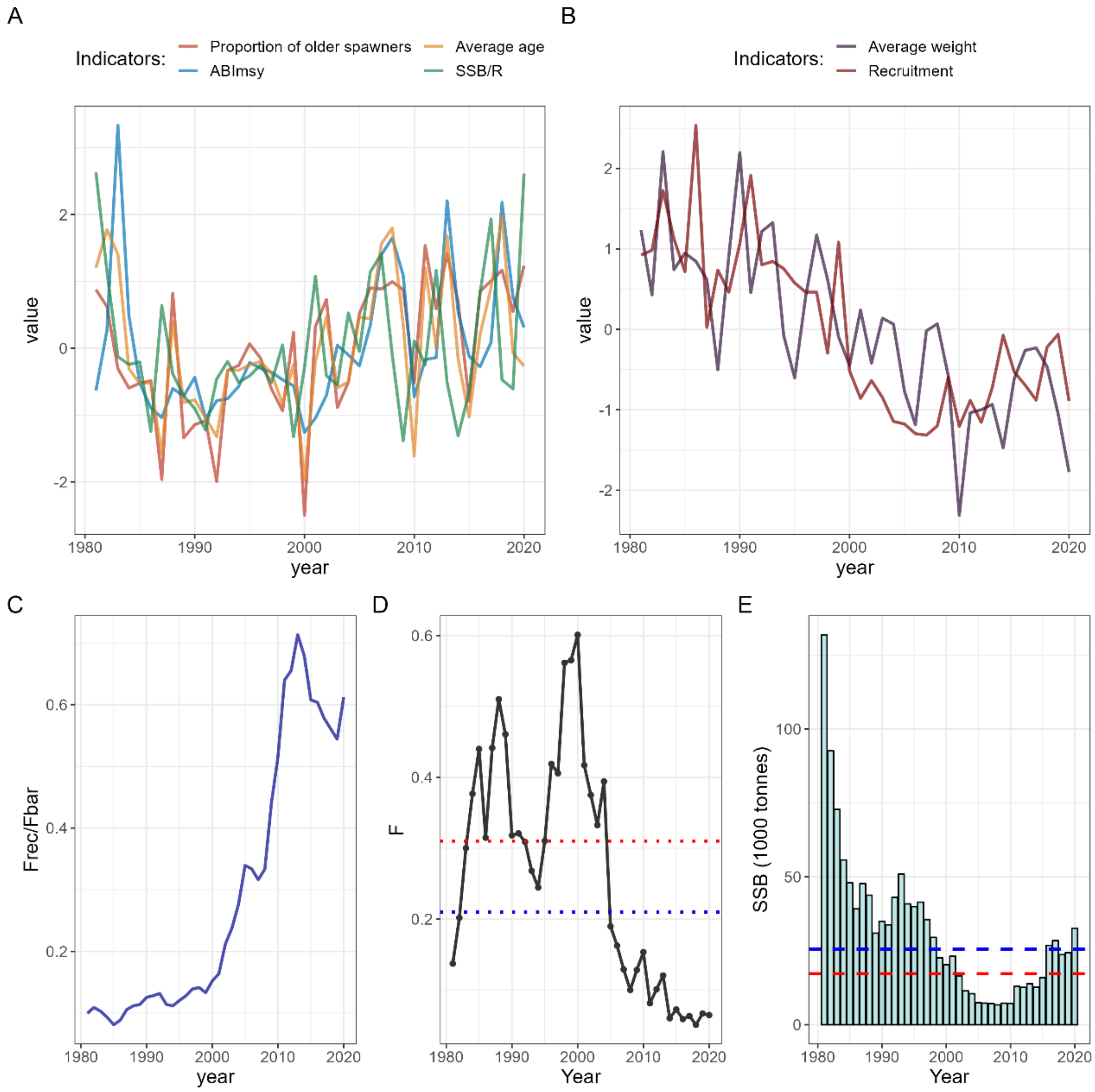


Figure A2.78. Estimated D3C3 indicators and stock status for Whiting (whg.27.6a). See Figure legend A2.1 for further details.

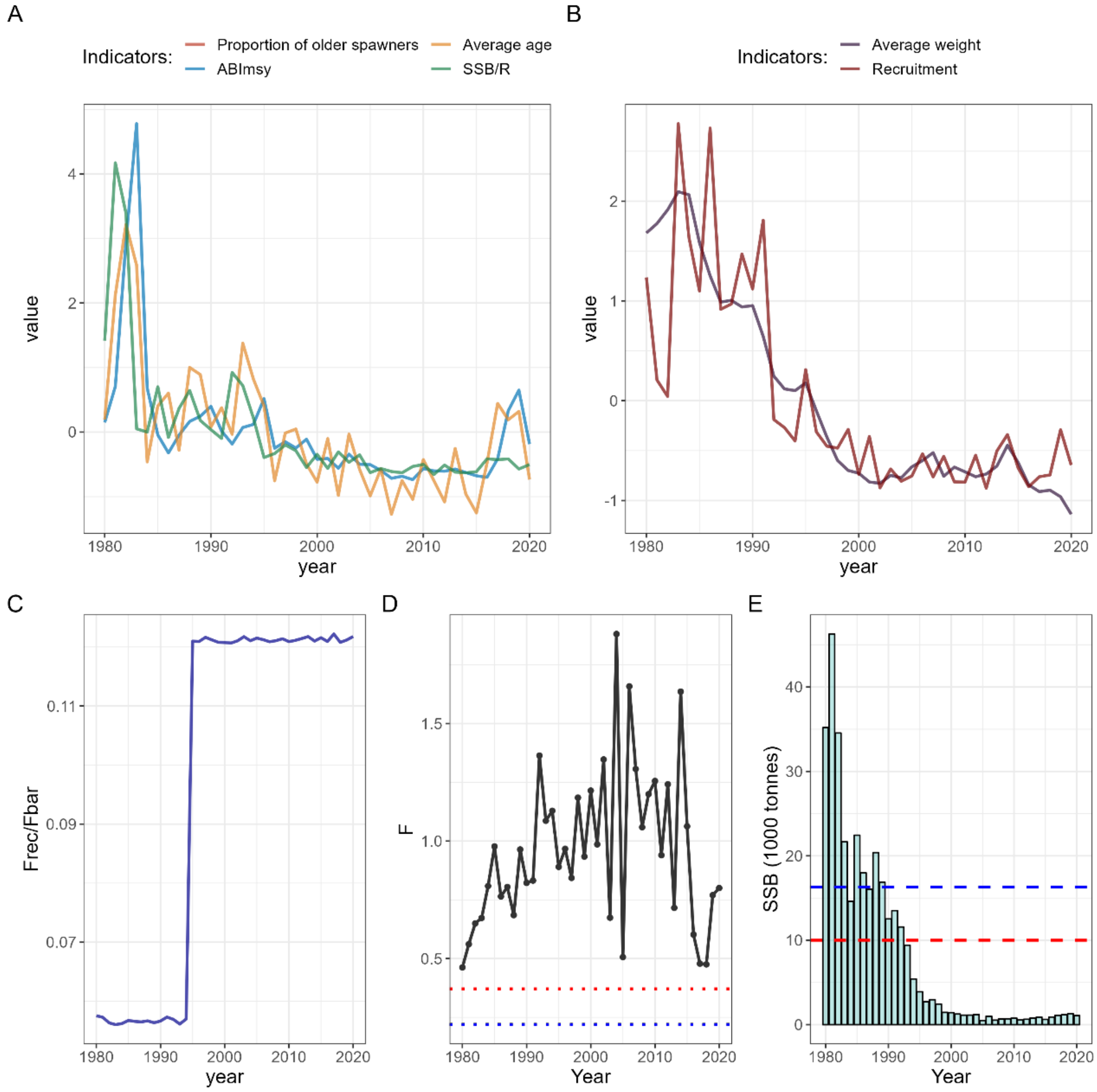


Figure A2.79. Estimated D3C3 indicators and stock status for Whiting (whg.27.7a). See Figure legend A2.1 for further details.

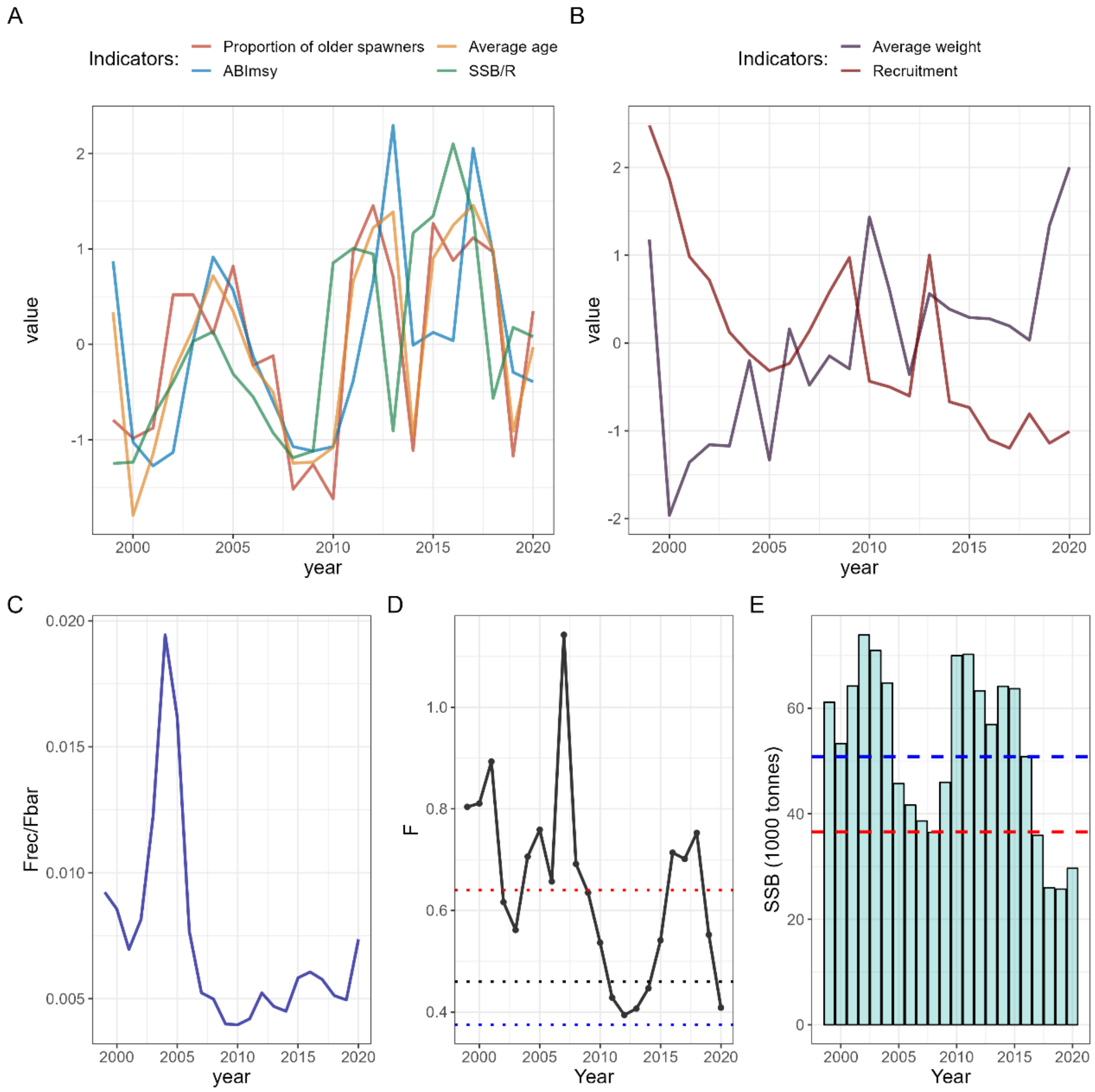


Figure A2.80. Estimated D3C3 indicators and stock status for Whiting (whg.27.7b-ce-k). See Figure legend A2.1 for further details.

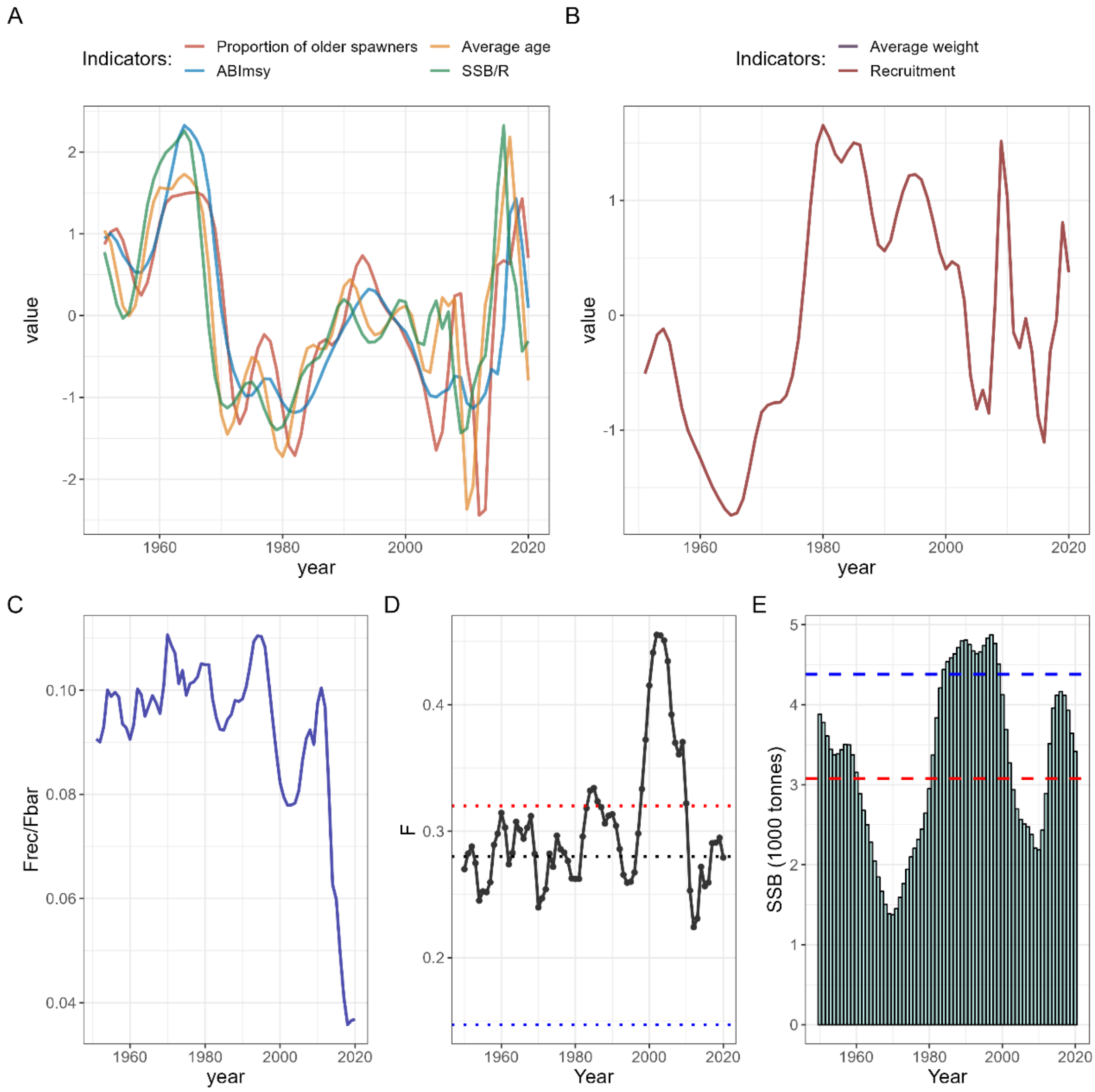


Figure A2.81. Estimated D3C3 indicators and stock status for Witch (wit.27.3a47d). See Figure legend A2.1 for further details.

Annex 3: Agenda

<table>
<tr><td>Time</td><td>Issue</td><td>Responsible</td></tr>
<tr><td>Terms of Reference</td><td colspan="2">a) Calculate, validate and evaluate the D3C3 indicators agreed at WKD3C3SCOPE for a selection of stocks representing different life-histories, data availability and MSFD (sub)regions. Calculation procedures will be documented and provided as technical guidance.
b) Derive thresholds for the D3C3 indicators for stocks representing different life-histories, data availability and MSFD (sub)regions and:
1. Evaluate the consequences of failing to achieve the thresholds
2. Evaluate consistency and complementarity with D3C1(F_{MSY}) and D3C2 ($MSYB_{trigger}$)
c) Discuss and agree on suitable indicators and threshold definitions for D3C3 assessment for stocks with different life-histories, data availability and MSFD (sub)regions.
d) Draft a framework for the comprehensive assessment of D3 stocks that includes:
1. The data requirements to assess D3
2. Recommended indicator(s) for the assessment of D3C3 that are compatible with D3C1 and D3C2
3. Methods to set thresholds and reference levels.</td></tr>
<tr><td>Monday September 18th</td><td colspan="2"></td></tr>
<tr><td>14.30-15.30</td><td>Welcome, practicalities round of presentation, discussion of scope and terms of reference and adoption of agenda</td><td>Anna and Guiseppe</td></tr>
<tr><td>15.30-16.30</td><td>Presentations of historical timeseries of indicators for different stocks (tor a)</td><td>Please send an email to Anna and Guiseppe if you wish to provide figures to document historical development
Nis
Chris
Paris</td></tr>
<tr><td>16.30-16.45</td><td>Coffee break</td><td></td></tr>
</table>

16.45-18.30	Presentations of historical timeseries of indicators for different stocks (tor a)	Please send an email to Anna and Guiseppe if you wish to provide figures to document historical development
Tuesday September 19th		
9.00-10.45	Discuss patterns in historical timeseries of indicators for different stocks and agree on suitable indicators for D3C3 assessment for stocks with different life-histories, data availability and MSFD (sub)regions (tor c)	All
10.45-11.00	Coffee break	
11.00-13.00	Documentation of calculation procedures provided as technical guidance for stocks with different life-histories, data availability and MSFD (sub)regions (tor a).	Please send an email to Anna and Guiseppe if you wish to submit a calculation procedure for use by others
13.00-14.00	Lunch break	
14.00-15.30	Describe calculation methods to derive thresholds and the associated uncertainty of the threshold for D3C3 indicators using the approached agreed in WKD3C3SCOPE: • Level of the unfished stock • Level at which the indicator aspect is not healthy (e.g. recruitment declines) following principles of D3C2 • Level of indicator of stocks considered to be in good health (stocks identified as good by expert judgement) • Level of the indicator when the stock is fished according to MSY principles • Level of the indicator when the stock provides for foodweb needs	Please send an email to Anna and Guiseppe if you wish to submit a calculation procedure for use by others
15.30-15.45	Coffee	
15.45-17.00	Discuss pros and cons of different threshold approaches including their consistency and complementarity with D3C1(F_{MSY}) and D3C2 ($MSYB_{trigger}$) objectives for the stock	
17.00-18.00	Wrap up of the day in plennary	

Wednesday September 20th		
9.00-10.45	Derive agreed thresholds for the D3C3 indicators for stocks representing different life-histories, data availability and MSFD (sub)regions	
10.45-11.00	Coffee	
11.00-13.00	Present and agree on thresholds for different stocks	
13.00-14.00	Lunch	
14.00-15.30	Describe the consequences of failing to achieve the thresholds	
15.30-16.00		
16.00-17.00	Discuss methods to evaluate consistency and complementarity with D3C1(F_{MSY}) and D3C2 ($MSYB_{trigger}$) quantitatively	
17.00-18.00	Wrap up of the meeting in plennary	
Thursday September 21st		
9.00-10.45	Agree on indicators and thresholds to be further investigated in simulations in WKD3C3SIMUL	
10.45-11.00	Coffee	
11.00-12.30	Assigning writing tasks and write report	
12.30-13.00	Wrap up of the meeting in plenary	

Annex 4: Reviewers Reports

Synthesis focusing on key aspects in relation to the advice request

See also the individual reports for more details and additional comments

The three workshops met to identify operational indicators and define useable threshold values for MSFD D3C3: "*The age and size distribution of individuals in the populations of commercially-exploited species is indicative of a healthy population. This shall include a high proportion of old/large individuals and limited adverse effects of exploitation on genetic diversity.*"

The intended use of the request output is to

1. Address a legal requirement for D3C3 under Decision (EU) 2017/848;
2. Contribute to the MSFD CIS process;
3. Allow using D3C3 in the integration rules for D3 (request to ICES pending);
4. Contribute to standardised EU Member States' reporting for MSFD descriptor 3 (reporting provided for under MSFD Article 17);
5. Contribute to MSFD article 8 assessment guidance.

Each of the three workshops was designed to cover the more specific parts of the request, but as the conclusions from preceding workshops were partially further developed under later ones, their outcomes are assessed together in the synthesis.

Part 1. "Define characteristics of a 'healthy population structure' for species with different life history traits and identify relevant indicators for these characteristics"

- The workshops identified healthy fish stocks as "characterized by high productivity, wide age and size structuring in the population, and the ability to quickly recover from disturbances".
- They, further, identified and retained age-based and length-based indicators applicable to various parts of this definition. The indicators were related to recruitment (R, assessment and survey-based), growth of individuals (ASW), age structure (ABI_{MSY}, POS, ASA and possibly SSB/R), and length structure (L90R and L90). The workshops found it unclear whether Fjuv/Fapical belongs under D3C3 but it was nonetheless also retained within D3. Indicators for genetic diversity and parasite load were not identified due to lack of relevant data.

Technical aspects, including appropriateness of methods, application of the best available science, technical assessment of the credibility of scientific findings

- The workshop conclusions rely on the combined expertise and contributions of the workshop participants. The selected indicators have been peer-reviewed and / or reviewed and applied for management, and analyses were conducted using established models and quality-assured data.

Scope and depth of the science in relation to the request?

- The workshop did not explicitly or structurally relate to all D3C3 aspects lifted by EU (2022; table 5.5.1-2; also in ICES 2022). For example, it is not clear if the workshop intentionally omitted genetic diversity from their definition, compared to the existing MSFD definition, or if it was just omitted from the further discussions at the workshop.

- The assessment-based indicators (ABI_{MSY}, POS, ASA, SSB/R, R, ASW and Fjuv/Fapical) use data that are input to or output from age-based stock assessments and are therefore readily applicable to stocks with Category 1 assessments. Survey-based indicators (R, L90R and L90) could provide an alternative for some data-limited stocks but were evaluated only for a handful of species and the requirement of reliable length distributions potentially restricts application to Category 3–4 stocks.
- While consideration was given to a range of life history characteristics, the results from historical evaluations combined with availability of MSEs limit the findings to stocks with medium life spans.

Does the analysis contain the knowledge to answer the request for advice?

- Consequences for individual stocks or marine foodwebs of applying the identified definition of "healthy population structure" (transferred to "healthy fish stocks"?) operationally are not comprehensively addressed
- Nonetheless, several of the identified indicators are likely to become operationally useful for D3C3 assessment in the near future, at least for some stocks. However, the work still needs to be continued before this can happen (relating to further comments within Parts 2-5 below).
- Future work could explore simpler options (than relying on MSEs) to increase applicability to more stocks.

Part 2. "Identify thresholds of 'healthy population structure' indicators and for species with different life history characteristics"

- The work suggests that generic approaches for setting threshold values can be agreed upon, although more work will be needed to ensure that these are widely acceptable and adequate. A decision tree on setting thresholds was defined and consolidated. Although it may appear not comprehensive enough for all stocks, it a solid basis to be expanded in the future.
- The work also suggests that any generic approach should be flexible enough to ascertain that it will not override any specific circumstances that need to be taken into account for individual stocks or sea regions. As one potential way forward, the setting of specific threshold values could be done by regional/local stock experts based on commonly agreed guidance. The advantage would be that this estimation will be conducted by a group of experts on the specific stock, ensuring that the indicator is based on the best available knowledge of the stock.

Technical aspects, including appropriateness of methods, application of the best available science, technical assessment of the credibility of scientific findings

- Various approaches for identifying threshold values are identified and systematically tested for the focal stocks and methods, conclusions are provided on when different approaches are likely applicable (or not applicable).

Scope and depth of the science in relation to the request?

- The absence of conclusive results could be attributed to the wide and diverse task.
- The scope of the workshops was not sufficient to obtain decisions on threshold values or approaches for determining limits between desired and not desired states.
- A decision tree was defined in order to work on common guidelines to encompass all the stocks as much as possible, according to data availability.

Does the analysis contain the knowledge to answer the request for advice?

- The potential use of indicators for recruitment and growth for signalling a need to adapt D3C1 and D3C2 thresholds is mentioned in WKD3C3Thresholds but consequences of different options for D3C3 thresholds in relation to D3C1-2 are not extensively explored there or in the subsequent workshop. (Current EC guidance for developing D3C3 states that "in developing D3C3 indicators, compatibility between any threshold values of D3C3 and the threshold values of criteria D3C1 and D3C2 should be ensured to ascertain that all criteria can in fact be attained simultaneously. The current guidance does not address which criterion to adapt, i.e. which criterion should take precedence if two criteria are in conflict" (EC 2022)).
- A recommendation to define indicator thresholds using stock specific MSEs would undoubtedly limit the number of species to which D3C3 indicators can be applied.

Part 3. "Explore the relationship between population traits/dynamics and healthy population structure' indicators and thresholds through simulations and infer cases where management in the context of CFP objectives -and equally of MSFD D3C1 and D3C2- alone may be insufficient and additional management measures should be envisaged. In such cases, and depending for example on the characteristics and exploitation patterns of the populations concerned, suggest a set of management options, ranked in decreasing order of expected effectiveness".

- The workshops give useful background information on a wide range of aspects, exploring links between the assessed indicators and stock biomass/fishing pressure for the focal stocks, and responses in stocks of different lifespan and levels of fishing pressure. So far, historical evaluations of the age structure indicators show good correlation with F and SSB for medium-lived stocks while MSE simulations show decreasing values of all D3C3 indicators with increasing long-term F. It also shows some examples of how environmental factors may influence on the D3C3 assessment. In all, the work shows how evaluating indicators for CFP and MSFD D3 management objectives is complex due to complex responses of indicators to both fishing pressure and environmental variation.
- Management measures are identified in relation to fisheries management, identifying and comparing 1) Measures aiming to reduce fishing pressure through regulating fishing effort or catches and 2) Measures aiming to reduce fishing pressure on specific stock components such as recruits or spawning fish. Rescaling fishing pressure is identified as a highly effective means of regulating age-structured indicators at levels around FMSY.

Technical aspects, including appropriateness of methods, application of the best available science, technical assessment of the credibility of scientific findings

- The applied methods and technical approaches appear suitable and appropriate for data-rich stocks and within the specified workshop aims.

Scope and depth of the science in relation to the request?

- Results are for a limited number of stocks, and further studies will be needed to investigate the validity of these initial results across life history types, data availabilities and environmental settings.
- The evaluations of functional relationships focus on effects on productivity but do not address other properties included in the definition of healthy fish stocks (see part 1).

Does the analysis contain the knowledge to answer the request for advice?

- The aspect "infer cases where management in the context of CFP objectives -and equally of MSFD D3C1 and D3C2- alone may be insufficient and additional management measures should be envisaged" appears not fully addressed.
- The work reflects awareness of the operational connections between D3C3 and D3C1-2 in the definition of stock status, from both technical and biological sides. It does not similarly reflect on potential impacts on the fulfilment of the overall MSFD objectives. For example, the work so far does not explore effects of D3C3 thresholds on the range of biodiversity ecosystem components in the MSFD (D1), foodwebs (D4), or genetic effects on stocks. Any final conclusions on how to operationalize D3 (C3) indicators may need to clarify how to related to these aspects as well, to ensure that approaches taken are widely adequate and acceptable within the MSFD framework, although it could be understandable that they could not be covered under the current request.

Part 4. "Advise indicators and thresholds most suitable for D3C3 assessment for species with different life history characteristics, giving preference to indicators that are derived from easily collected data (e.g. data routinely collected under the DCF)"

- Tables 8.1-2 in the report of WKSIMULD3 gives an overview of advised indicators and thresholds and how they could relate to D3C1-3.
- The participants at WKD3C3THRESHOLDS agreed that health of the stock as assessed under D3C3 can only be evaluated if information on recruitment, weight at age, size/age distribution and/or fisheries selection pattern of the stock is available. The participants at , WKD3C3THRESHOLDS suggested that if these data are not available, D3C3 cannot be assessed and Member States should be encouraged to remedy through increased data collection.

Technical aspects, including appropriateness of methods, application of the best available science, technical assessment of the credibility of scientific findings

- Adequate within the identified scope and in relation to agreed results

Scope and depth of the science in relation to the request?

- It is not clear how to deal with indicators suggested in the existing MSFD guidance but (for different reasons) were not covered by the workshops.

Does the analysis contain the knowledge to answer the request for advice?

- Not fully, as the analysis is dependent on results from the preceding parts.

Part 5: "Prepare a framework for comprehensively assessing D3 criteria for commercially-exploited fish and shellfish populations (= stocks), including data-limited stocks"

Technical aspects, including appropriateness of methods, application of the best available science, technical assessment of the credibility of scientific findings

- The focus on age-based indicators, with lesser emphasis on length-based indicators makes the results less relevant for many stocks and management regions;
- The age-based indicators incorporate the conversion process from length to age classes; this includes different aspects as, for example, harmonization of 1) age reading protocols along the years and among the institutes working on the areas where the stock is distributed, 2) of methods to fit von Bertalanffy growth curve and age slicing procedures. For this reason, the exploration and application of length-based indicators in

parallel with the age-based is necessary to disentangle possible effects of age conversion.

- Also, the fish condition indicator can be another indicator to be considered in the future, linking strongly to the health status of the stock; this indicator is based on data easily available under DCF.
- Typical length and mean maximum length indicators are two other length-based indicators that could be considered in the future and that can be estimated on survey data for many stocks and regions.

Scope and depth of the science in relation to the request?

- The examples that were explored, and which form the basis for the framework, do not claim to reflect the wide range of commercial fish stocks present in European Seas, as evident from previous advice (ICES 2022).
- For example, mediterranean stocks are not sufficiently covered; it is important to increase the number of applications of the selected and other indicators in Med basin, where survey and commercial data collected under DCF are available since long time (MEDITS from 1994, DCF from 2002). More Mediterranean experts need to be encouraged to participate to the next workshops.

Does the analysis contain the knowledge to answer the request for advice?

- How deal with the evaluation of D3C3 for the full existing range of stocks and data availabilities is not yet fully elaborated on. Different life-histories, species-specific factors and regional or sub-regional variation need to be further considered, including diversities within the assessment of stocks based on survey data.
- The workshops emphasised that a single indicator is unlikely to suffice for all stocks. How to potentially aggregate indicators within D3C3 was not addressed, but is probably more relevant to address at the point when finally operational indicators are defined.

Reviewer 1

Review of WKD3C3SCOPE, WKD3C3THRESHOLDS and WKSIMULD3

Summary findings in relation to the request:

The three workshops met to identify operational indicators and define useable threshold values for MSFD D3C3: "*The age and size distribution of individuals in the populations of commercially-exploited species is indicative of a healthy population. This shall include a high proportion of old/large individuals and limited adverse effects of exploitation on genetic diversity.*"

The workshops identified and retained seven age-based and three length-based indicators of D3C3 relating to recruitment (R, assessment and survey-based), growth of individuals (ASW), age structure (ABI_{MSY}, POS, ASA and possibly SSB/R), length structure (L90R and L90) and fisheries selectivity (Fjuv/Fapical), that reflect the range of '*healthy population structure*' characteristics identified during WKD3C3SCOPE, but barring genetic diversity and parasite load through lack of relevant data. The selected indicators have been peer-reviewed and / or reviewed and applied for management (e.g., within HELCOM).

Historical evaluations of the age structure indicators show good correlation with F and SSB for medium-lived stocks while MSE simulations show decreasing values of all D3C3 indicators with increasing long-term F, justifying their selection. It is unclear whether Fjuv/Fapical belongs under D3C3 but was retained, nonetheless.

The assessment-based indicators (ABI_{MSY}, POS, ASA, SSB/R, R, ASW and Fjuv/Fapical) use data that are input to or output from age-based stock assessments and are therefore readily applicable to stocks with Category 1 assessments. Survey-based indicators (R, L90R and L90) could provide an alternative for some data-limited stocks but were evaluated only for a handful of species and the requirement of reliable length distributions potentially restricts application to Category 3–4 stocks.

While consideration was given to a range of life history characteristics, the results from historical evaluations combined with availability of MSEs limit the findings to stocks with medium life spans.

In the absence of a clear decrease in productivity at low indicator levels, it is recommended to define indicator thresholds using stock specific MSEs. This will undoubtedly limit the number of species to which D3C3 indicators can be applied.

Comments on the science and analyses:

From the reports it appears much of the code is available on the group SharePoint and GitHub sites. However, as reviewer, I did not have access to this, and my review is based on the reports alone.

- While the issue of survey catchability with regards to survey timing, geographic extent and size structure is acknowledged in the WKD3C3THRESHOLDS and WKSIMULD3 reports, it should be emphasized that any state indicator derived from survey data should be representative of the underlying population. The groups approach of excluding recruits from indicator calculations is sensible provided any threshold is adjusted accordingly (although very large recruitment events may track through to older ages in future years), but catchability of older ages also needs to be considered when evaluating the appropriateness of survey data and defining thresholds, e.g., dome-shaped selectivity may invalidate the underlying rationale.
- The Northeast Atlantic data set covers a wide range of species, areas, life history groups and exploitation levels. Although the extremes of the long-lived group may be unrepresented because there are no elasmobranchs included (and only one Category 1 elasmobranch stock within ICES that could have been considered). Furthermore, no long-lived species were included in the MSEs.
- Only four species were included in the Mediterranean dataset used to assess the length structure indicators and it is unclear into which life history category they fall. Furthermore, only two species were included in the MSE simulations. Future work could consider combining the Northeast Atlantic data set with length-based survey data from DATRAS, to evaluate these indicators for more stocks.
- The group state that estimation of thresholds requires a stock specific MSE. MSE development is a complex and time-consuming task, and it is therefore unlikely that MSEs will be developed for all stocks for which D3C3 criteria are to be assessed. Future work could explore simpler options to increase applicability to more stocks. These could include (1) building the estimation of D3C3 thresholds into the EQSIM framework (although this may be what is meant by 'using predictions of MSEs recently used to define stock reference points under D3C1 and D3C2' in Section 8.1); (2) use of generic MSEs to define thresholds by life history characteristics; and (3) use of ROC curves to optimise threshold values (based on stock and exploitation status) and explore potential generalisations across life history types.

Specific comments from each workshop report:

WKD3C3SCOPE

- The acronym PMRN is undefined (pages 5 and 28).
- It would be helpful to include species common names on page 16.
- It is not clear how a_{old} is defined in Section 3.8 and whether there is a general rule for setting it even though it varies among stocks. Furthermore, the use of old spawners and explanation of $\overline{w}_a$ is not sufficient in this report to understand how the equation for ASW works. It wasn't until reading the report from WKD3C3THRESHOLDS that it became clear $\overline{w}_a$ is used to mean-standardise the time-series of w_a.
- It is unclear in section 3.9 whether time-series of SSB and R were lagged in the cases where the recruitment age is not zero.

WKD3C3THRESHOLDS

- It would be useful to include species common names in Table 3.3.1 to give a better sense of the species covered and what life history group they fall into.
- The calculations of ABI_{MSY} assumed three-year averages for biological and selectivity parameters and a segmented regression with breakpoint at B_{lim}, which is appropriate here for application to many stocks. In practise, these could match the stock specific EQSIM settings used to define reference points.
- The report recommends retaining Fjuv/Fapical. I assume this is following the results of the existing studies referred to in section 3.1.8 as this indicator was not specifically evaluated and results in the appendix are for Frec/Fbar.

WKSIMULD3

- As for THRESHOLDS, it would be useful to include species common names in Table 2.6.1 to give a better sense of the species covered and what life history group they fall into.
- The reasons for dismissing the used of ROC curves is unclear, given the historical evaluations conducted in THRESHOLDS, and the MSEs considered here, are based/conditioned on stock assessment outputs.
- The use of 'avgAge' for average spawner age throughout the report is inconsistent with the use of ASA in the previous reports.
- For some short-lived species, R and ASW are compared to historic F. Why was this comparison not done also with biomass reference points? Particularly for sprat which is managed based on escapement rather than an F-based strategy. Furthermore, it is not clear whether MSE or stock assessment outputs are used for these comparisons (I assume the latter).
- For the two hake MSEs (currently labelled Figures 5.4.10–5.4.11), could the overestimation of large individuals be a consequence of simulating a selection pattern that is less domed than in reality?
- It is not clear which results are being referred to in the first paragraph of Section 6.2, reference to a figure would be helpful.
- In Section 6.10, potential thresholds are presented against historical developments, but no synthesis of the outcomes given.

Reviewer 2

Individual comments on the report of WKD3C3SCOPE

- WKD3C3SCOPE defined overall, healthy fish stocks as characterized by high productivity, wide age and size structuring in the population, and the ability to quickly recover from disturbances. The MSFD already defines healthy fish stocks ("healthy population structure") that indications of a healthy age and size distribution of in populations of commercially exploited species include a high proportion of old/large individuals and limited adverse effects of exploitation on genetic diversity. Hence, WKD3C3SCOPE appear to have extended this definition to also covering sufficient prevalence of younger individuals and functional consequences in the sense of characteristics leading to high productivity and the ability to quickly recover from disturbances. All of these aspects are relevant to explore in the definition of D3 indicators, and most of the aspects are probably applicable to criterion 3 (although this could be subject to further discussion).
- It is not clear if the workshop intentionally omitted genetic diversity from their definition, or if it was just omitted from the further discussions at the workshop.
- The workshop did not explicitly relate to all D3C3 aspects lifted by EU (2022; table 5.5.1-2; also ICES 2022).
- WKD3C3SCOPE suggested a list of indicators for further evaluation in the next workshop(s). The participants did not find it suitable to rank the suggested indicators from a general perspective, as they found that a ranking would depend on additional factors, e.g. what data is used to estimate them. Instead, they prepared a joint new set of evaluation criteria for use in WKD3C3THRESHOLDS. These are listed in Table 4.2 and appear motivated and useful.
- (in the further work, the two subsequent workshops continued evaluating potential indicators for some of the aspects identified by WKD3C3SCOPE. For example, "ability to recover from disturbances" appear not have been in focus in the further evaluations. Proportion of fish with parasite infestation was lifted omitted due to lack of data, and it could also be discussed if this is relevant to be assessed as part of D3).
- The workshop emphasised that a single indicator is unlikely to suffice for all stocks. How to potentially aggregate indicators within D3C3 was not addressed but is probably more relevant to address at the point when finally operational indicators are defined.

Individual comments on the report of WKD3C3THRESHOLDS

The report was reviewed in draft format and the key conclusions are not always very clearly expressed, so it is possible that part of the work is not fully or correctly understood at this point.

- The age-based indicators were analysed by groups of stocks structured by life-length. This was seen as a useful way forward at this stage of analysis, when the aim was to screen for overall patterns, but caution should be taken not to interpret the results too far for individual stocks. This caution is also emphasised in the report.
- The most informative results were obtained for stocks classified as medium-lived. Several of the addressed indicators show promising properties that are worth exploring further. Especially ABI, ASA and POS were seen to follow the temporal development of SSB and react to F similarly and in the expected direction in many of the evaluated stocks.

- For long-lived species, age-structure indicators appeared to react slowly. A gap of up to 10 years was observed between changes in F and subsequent changes in age structure indicators for long-lived species. These results seem to contradict expectations expressed in EC (2022) that "…criteria D3C3 can function as an early warning indicator for D3C2 and thereby D3C1", for age-structure indicators, and the report states that the value of ABI and ASA would in those cases not lie in eliciting management action, which would be more timely based on F, SSB and R. The report discusses that this could lift questions about the value of age structure indicators as management indicators, especially for long-lived species. Similar concerns are lifted also for short-lived species, due to observed weak links between stock structure and productivity. However, the use of age structure indicators as status indicators still appears motivated in relation to the general objectives of D3C3, although this is not discussed in the report. (This distinction could be expressed for example in section 6.2, if wanted)
- The workshop discussed and explored potential thresholds for stocks representing different life histories, data availability and MSFD regions when possible. Clear thresholds where the indicators would signify stock productivity declines could not be identified, based on that none of the age-structure indicators showed a positive correlation with stock productivity. The group hence excluded the option to determine threshold levels based on levels at which stock productivity is either impaired or enhanced. As the next option, historical analyses or simulations were identified as the most promising approaches to derive thresholds values, which appears motivated.
- Extensive test evaluations were carried out for selected stocks and cases, and these give valuable information to support the further development. However, decisions on threshold values and approaches for determining limits between desired and not desired states were not possible within the workshop.
- The results are to large extent limited to stocks with which age-based assessment data, which represented only part of the stocks for which D3 assessments are required.
- For length-based indicators, the indicator L90R, calculated from the length-frequency distribution of fish larger than recruiting length, seemed to perform well when applied on survey data. The group noted the importance of quality-assuring input data prior to calculating indicators so that the indicator trends are not influenced by inconsistent sampling routines, annual environmental variation, seasonal effects, etc. This is in alignment with established practise for how to analyse long-time series for environmental data also generally.
- The selectivity indicators were retained although they may rather be supplementary indicators under D3C1.

Specific comments to the report of WKD3C3THRESHOLDS

Executive summary

- "The SSB/R indicator responded to recruitment in an undesirable manner": Clarify what is meant by "undesirable" to help the reader. This also appears in the beginning of section 6.2.
- "high age at spawning may lead to senescence rather than increased viability of spawning products.": While there are references to support this statement, it could be argued that they in reality don't reflect very likely circumstances for most stocks. The same statement appears in several places of the three workshop reports. I would recommend the authors to tone it down or provide more information about how likely

this actually is a management concern, as the text may currently suggest this. See also eg section 6.2 in the current report.

- "assessments based on age data are preferred over those based solely on length distributions for the estimation of age structure indicators": This appears quite evident, look over wording
- For structuring the conclusions about retained indicators, it would be helpful if the evaluated indicators were grouped according to the aspects shown in Table 5.5-1 in EC (2022), to give an overview of which aspects were covered by the workshop and which were not attempted (if the latter, also possibly clarifying for what reasons).

Main text

- Consider using the same generic short name for potential D3C3-indicators throughout the report, e.g. stock structure indicators (SSI). Currently, the term "health indicator" is sometimes used in this sense, but another term might be better suited as "health indicator" could be misinterpreted both in a narrower (signifying explicitly "condition" indicators) or wider (signifying general ecosystem health) sense than intended here.
- On section 3.2: "Stocks were split into three life span groups of short, medium and long-lived" and similarly in section 3.4.7. As these are very coarse divisions given the high diversity in life histories present among different fish species and among stocks in different sea regions, care should be taken so that the workshop results are not overinterpreted. A note could be made to reflect this in section 3.2. In section 3.4.6 it would be good to provide information about the ranges of life spans that were represented within each of the three groups. The first sentence in section 3.4.7 appears to dismiss the information available in Fishbase but doesn't motivate why the currently applied division would be less arbitrary.
- Section 5.5.3 table "Level at which the indicator aspect is not healthy following principles of D3C2" This is unclear.

Individual comments on the report of WKSIMULD3

The report was reviewed in draft format and the key conclusions are not always very clearly expressed, so it is possible that part of the work is not fully or correctly understood at this point.

- The report summarises the main workshop conclusions as: "Overall, the workshop findings highlighted the complexity of evaluating indicators for CFP and MSFD D3 management objectives, particularly in relation to the responsiveness of the indicators to fishing pressure and environmental variation. … Additionally, it underscores the challenges associated with using age- and length-based indicators for different species and the importance of considering environmental and recruitment variability in simulations." In this, would it be possible to also conclude on (or add a comment about) likely responses of indicators to fishing pressure and environmental variation, respectively, under different fishing pressure regimes. Can any patterns be identified at this stage? (For example, to what extent can the intensity or variability of fishing pressure, or other aspects, be expected to affect likely observed indicator-pressure relationships). It appears based on the various examples in the report that it could be interesting to take this aspect a bit further.
- Section 8.1 mentions: "The option to estimate the indicator thresholds independently of methods used to define stock reference points for D3C1 and D3C2 (FMSY and biomass reference points) was also discussed but the group considered that this approach

would provide results that were inconsistent as the FMSY and distribution of the indicators at this fishing pressure would differ". - This interpretation would be worth elaborating on to clarify the last part ("…as the FMSY and distribution of the indicators at this fishing pressure would differ"). This appears to be a too hasty conclusion from the workshop but could also reflect a need to add more explanations to the section explaining why this would be an issue.
- It would be worthwhile to discuss under which circumstances a D3C3 assessment is most urgently needed to improve management (and similarly guide priorities for further development work). Some hints are included in the report, but no clear conclusions. For example, the report at one-point states: "For stocks fished at levels above 1.5FMSY, it is probably preferable to denote the stock as being below the threshold as these stocks are unlikely to exhibit healthy age structure." As another example, if results for all criteria are to be combined in the end to obtain one status assessment for each stock, and assuming that the one-out-all-out principle is retained, evaluating C3 status would only be needed in cases when C1-2 do not meet their threshold values?

Specific comments to the report of WKSIMULD3

General

- I assume the report will still be edited for spelling and grammar errors (there are cases when words seem to be missing in the sentences), table and figure numberings need to be checked, etc.
- Some technical parts of the report will be difficult to fully understand for non-experts on the particular models and stocks.

Executive summary

- "Overall, it is highlighted how the indicator status can be well below from its value at Fmsy." This is not clear, consider rewording (also at page 39)
- Section 3 (about short-lived stocks) is not very easy to understand if it is about age-based, length-based indicators, or if results for the first are inferred from the latter (and in what order).

Main text

- Check wording in section 2.2 so that the section reflects what was aimed for in the workshop and what was achieved, respectively (perhaps use of past tense would be more suitable)
- Section 2.6.1 "During some of the analyses in [W]KD3C3THRESHOLDS, stocks were aggregated by exploitation level and life span." See same comment for that report above: As this is a coarse grouping given the high diversity in life histories present among different fish species, a note could be made to reflect, and it would be good to provide information about the ranges of life spans that were represented within each of the three groups.
- Section 2.6.2. Spell out abbreviation GSA.
- Table 2.6.2 and other similar: it would be very helpful to spell out the full names of the species, not only their letter codes.
- Section 3.1 "In the absence of evidence to suggest a threshold linked to observed harm, a percentage of 5 percent was agreed to be the most relevant for options 2 and 3 as this would mean that there was no inherent conflict between D3C1, D3C2 and D3C3 thresholds." It is not clear what is meant by "no inherent conflict" please specify.

- Section 3.1 "Even when fishing according to an MSY management approach, the indicators are expected to vary between [x?] as a result of variation in e.g. productivity of the stock. As a result, the inspections also considered whether the natural variability when fishing according to the MSY approach was so large that a change due to increased fishing would not be detected, in which case the indicator was thought not to be useable for management decisions (approximately equal number of false positives, true positive, false negatives and true negatives." I might be misunderstanding the intent here, but it is not clear why the MSY approach was used, could there no be a case for expecting a different relationship between fishing level and the indicator values under other strategies than MSY?
- Figure 3.1.1 It could be good to clarify that this figure is there for background content and that a reworked version is provided further down in the report (Figs in Chapter 8).
- Section 5.4.1. Please add clarification (page 33) about why indicator values corresponding to those at FMSY is preferred as a threshold
- Section 6.5. The conclusions for this option appear to underrate its potential as an approach for identifying operational threshold values for individual stocks, as flexibility could be added to allow the setting of target levels for individual stocks in line with local/regional management aims and trade-offs (not necessarily using 5th percentile for all) including allowing for exploitation (in line with what is suggested at the top of page 61).
- Figures 8.1.1-2. It is not clear why no further development is needed in the last box, how to solve/develop cases when the preceding boxes are not applicable?

Reviewer 3

Review of the draft report by WKD3C3SCOPE, WKD3C3THRESHOLDS and WKSIMULD3

Summary findings in relation to the request:

The three workshops met to identify operational indicators and define useable threshold values for MSFD D3C3: "*The age and size distribution of individuals in the populations of commercially-exploited species is indicative of a healthy population. This shall include a high proportion of old/large individuals and limited adverse effects of exploitation on genetic diversity.*"

However, I found more focus on age indicators rather than length based. I suggest in the future, if there would be the opportunity, to explore other length based indicators, already discussed during ICES WGECO 2012 and 2014) as the mean maximum length and the typical length. Following their definition, it is possible that these indicators are more responsive to the recruitment than the length-based indicators L90, LR90. Moreover, the fish condition indicator can be another indicator to be considered in the future, because quite linked to the health status of the stock. All this indicators could also be potentially estimated by survey data and not necessarily from the stock assessment results.

For Mediterranean the work presented, including 5 stocks in the first 2 workshops and in the last actually only 2, in the report in my opinion represents a starting point, but more work is needed to cover at least the main stocks for which a quantitative assessment is available.

Moreover, the link between the length-based and age-based indicators should be better explored, also in the light of the conversion methods applied in the different cases from length to age and specifically taking into consideration if harmonised age reading protocols, slicing procedures, von Bertalanffy fitting methodologies were applied. I recognise that this task is time consuming but in my opinion is crucial to identify spurious correlation or reason of not evident correlations (that instead, are expected).

a) Is the analysis technically correct?

I found all the analysis technically correct according to my experience and knowledge. I think that the three workshops were very well organised and that a huge amount of work was made in few time, contributing valuable insights into the complexities of evaluating and comparing D3C3 indicators for MSFD D3 management objectives.

b) Are the scope and depth of the science appropriate for the request?

I found the scope and depth of science appropriate for the request; probably, considering that more work was needed to be done in Mediterranean Sea, in the next workshops I encourage the participation of more Mediterranean experts to cover more stocks in that area.

c) Does the analysis contain the knowledge to answer the request for advice?

Yes. The three workshops met to identify operational indicators and define useable threshold values for MSFD D3C3. Although additional work is needed especially to better cover Med areas, a huge amount of work has been done in few time with a very high level of scientific reliability, including review of relevant indicators, exploration of correlations with F, SSB and R and MSE simulations, comparing different methods. The decision tree produced in the WKD3C3THRESHOLD and consolidated in WKD3C3SIMUL is a crucial product of the work carried out by the three workshops and very useful for the applications of the key indicators to a wider number of stocks.

Specific comments to the report of WKD3C3SCOPE

The workshop had the objective to identify a list of indicators informing of the health status of commercially exploited fish and shellfish stocks in line with descriptor D3C3 of the MSFD and to provide a list of criteria for the selection of a shorter list for the evaluation of the corresponding thresholds in WKDeC3THRESHOLD.

The workshop discussed about how to define a healthy population and the key characteristics that a healthy stock should have. I found efficient the followed approach based on sub-groups, allowing to highlight in short time the main aspects to be taken into consideration for this definition. Among the different factors the monitoring of the natural mortality was cited. This is an important factor but quite difficult to monitor especially in wild populations.

Moreover, from the report I found the separation between short and long-living species as related to the different impact of environment on their health status, but no mention was done on the difference between pelagic and demersal stocks. In this sense, pelagic, especially small pelagic, are more affected by environment respect to demersal stocks. This should be also considered.

In relation with the climate change, also a separation between thermopile and not thermophile species could be considered in the future (as attempted in several works on mean catch temperature).

Within the sub-groups a high individual condition was also mentioned as a characteristic of a healthy stock, but this indicator was not considered anymore in the final list of indicators.

Also the review of the available indicators through a common structure based on operative aspects is in my opinion very efficient. Although I found the list quite extensive, I think that the typical length (Geometric mean length of fish community, weighted by body mass, based on ICES 2014 and OSPAR 2017) needed to be considered. In Bitetto et al. (2019), this indicator was explored at community level as complementary with the mean maximum length (as already highlighted by Lynam and Rossberg, 2017), because the joint application of both indicators can help disentangle changes in community composition from changes in size structure. An exploration also considering 1, 5 and 10-year time lag was also carried out. This indicator, if applied at stock level could be considered as tracker of a change in the size composition of the stock along the years.

When the list of indicators is provided (e.g. chapter 3.3), it would be better specified if the indicator is calculated on fish (including elasmobranchs) or also shellfish or if it is considered by stock.

Table 1.3: in the caption should be indicated D3C3(not only D3).

chapter 2.6: it is not clear what means "healthy maturity ogive". It should be clarified.

chapter 2.6: examples of food web needs should be indicated for completeness (also highlighted in WKD3C3THRESHOLDS draft report, page 27).

Specific comments to the report of WKD3C3THRESHOLD

I acknowledge the huge amount of work carried out during the workshop to address all the ToRs and to apply a many as possible indicators to as many as possible stocks in the Atlantic and in Mediterranean Sea.

My feeling is that the focus was mainly on the age structure indicators (ASA, ABImsy, POS, ASW). It should be noticed that these indicators are strongly based on age data (Age- length key, age slicing procedures, depending on the stock assessment and data available) and are prone to bias on age reading protocol agreement among the readers and/or harmonisation on the von Bertalanffy growth curve estimation method. These could be also the reasons of poor correlation between age structured and length-based correlation (p. 31 WKD3C3THRESHOLDS draft report). Although the length-based indicators have not this issue and can be estimated also without and age structured stock assessment, they were explored only marginally by the WKD3C3THRESHOLDS. The typical length indicator could represent an additional indicator to be included in the list. Moreover, other length-based indicators indicated in the WKD3C3SCOPE were not fully explored during the workshop: proportion of individuals with length > L95 (p. 15 WKD3C3SCOPE), mean maximum length, proportion of individuals longer than Lm50%, average somatic condition (p. 11 WKD3C3SCOPE).

It is also not clear the sentence at page 31 of the draft report indicating that length-based indicators are more prone to lower signal to noise ratio than age based indicators when the length distributions are based on low number of sampled individuals. This is unclear to me because it is generally expected that for the same species more length data are available respect to age data, due to the amount of work needed for otolith reading.

I acknowledge the effort to explore ASW indicator as a representation of the somatic growth in the population; in my opinion also the fish condition indicator (e.g. Le Kreen) could be considered depending on the data availability.

Table 5.2.1: Another column with WKTHRESHOLD follow-up findings for each criterions would be useful.

In the text the time lag 1 is missing (not very clear what it means a time lag of +1) when referring to Table 5.3.1.

I would suggest to carry out the analysis described at page 33 of the draft report with time lags (also wider, like 5 and 10 years) by longevity group to verify if in all the cases there is still this undesirable negative correlation with recruitment.

In any case, considering the formulation of ASA and POS is not surprising that no correlation with R was found, being based completely on older individuals; moreover, the ratio SSB/R, by definition, is expected to be negatively correlated with R. It would be interested to test the same correlation, including time-lags set according to the longevity group, with the length-based indicators.

I found very informative the decision tree defined and discussed by the group to establish a clear roadmap for the selection of best threshold estimation approach.

Page 43: not clear the sentence related to the noisy results of the length-based assessment; where it is demonstrated in the report?

Specific comments to the report of WKSIMULD3

The amount of work carried out to run MSE simulations during the WKSIMULD3 is very impressive.

I found very important the use of different methods to estimate thresholds and reference point before MSE (Eqsim, FLBEIA MSE).

The indicator A90prop/A90fmsy was never mentioned in the previous workshops (see page 18 of the draft WKD3C3SIMUL report). More details should be provided, because can represent another indicator to be included in the age-based indicators in the future.

I found the MIZER approach very useful to explore and identify the impact of food web and of the environment on the indicators. This is a challenge to be taken into consideration also in the future following the possibly increasing data availability also in other areas.

Page 22, chapter 5.2.1.2: please include the number of runs, for completeness, even if it is reported in chapter 5.2.2.1.

I suggest to update the MEDITS reference, including also Spedicato et al, 2019 in addition to Bertrand 2002.

Chapter 5.4.1: it is not clear <u>why it is stated that HKE8_11 a4a model is not benchmarked</u>. As far as I know, the STECF assessment should be an update of the benchmark assessment.

Concerning the Mediterranean Sea, 3 stocks were analysed, using the VirtualPop function in the R package fishdynr, allowing to carry out length based simulations at different levels of F. Although the application of this tool was done including all the available information from the stock assessment reports, it cannot be considered to incorporate the stock assessment results, especially the current status of the stock; this it is quite far from an MSE. More work need to be done for the stocks in Mediterranean Sea, starting from the ones having the biological reference points already defined.