WORKSHOP TO APPLY THRESHOLDS FOR THE PRESELECTED INDICATORS FOR MSFD D3C3 (WKD3C3THRESHOLDS)

VOLUME 6 | ISSUE 3

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

This document has been produced under the auspices of an ICES Expert Group or Committee. The contents therein do not necessarily represent the view of the Council.

© 2024 International Council for the Exploration of the Sea

This work is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). For citation of datasets or conditions for use of data to be included in other databases, please refer to ICES data policy.

ICES Scientific Reports

Volume 6 | Issue 3

WORKSHOP TO APPLY THRESHOLDS FOR THE PRESELECTED INDICATORS FOR MSFD D3C3 (WKD3C3THRESHOLDS)

Recommended format for purpose of citation:

ICES. 2024. Workshop to apply thresholds for the preselected indicators for MSFD D3C3 (WKD3C3THRESHOLDS).
ICES Scientific Reports. 6:3. 141 pp. https://doi.org/10.17895/ices.pub.25266580

Editors

Anna Rindorf • Giuseppe Scarcella

Authors

Anna Rindorf • Giuseppe Scarcella • Enrico Armelloni • Valerio Bartolino • Jose Maria Bellido • Elisabeth Bolund • Phillip Boulcott • Victoria Campon-Linares • Gema Canal • Helena Caserman • Ilaria Coscia • Tomaso Fortibuoni • Eva Garnacho • Beatriz Giujarro • Patricia Gonçalves • Chris Griffiths • Nis Jacobsen • Bernhard Kuehn • Isabel Maneiro • Ualerson Iran Peixoto da Silva • Nikolaus Probst • Owen Rowe • Sonia Seixas • Amina Tifoura • Paris Vasilakopolous • Ching Maria Villanueva • Håkan Wennhage • Nuria Zaragoza

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

Contents

i Executive summary

The WKD3C3THRESHOLDS 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). WKD3C3THRESHOLDS is the second of a series of three workshops (WKD3C3SCOPE, WKD3C3THRESHOLDS and WKD3SIMUL) to identify operational indicators for MSFD D3C3.

The workshop was organised as a series of presentations of results with intermittent group discussions.

The D3C3 indicators agreed at WKD3C3SCOPE were estimated and documented for a selection of stocks representing different life-histories (Tor a). Plots comparing indicators were investigated for stocks with all estimated indicators. The age structure indicators ABI, ASA, POS and SSB/R generally followed the temporal development of SSB and react similarly to F. A gap of up to 10 years was observed between changes in F and subsequent changes in age structure indicators for long-lived species while SSB responded quickly to changes in F. For medium-lived stocks, the four age structure indicators exhibited similar temporal patterns, with SSB divided by R tending to be more variable. Recruitment and mean weight at age documented shifts in productivity, impacting age structure indicators differently when changes occur. Plots of F, recruitment, weight at age and SSB are considered useful for understanding cases where changes in F do not impact SSB as expected (e.g. rebuilding does not occur or stock remains high in spite of high F). Higher proportions of older fish as measured by ABI/ASA/POS or SSB/R did not appear linked to an immediate increase recruitment. A comparison of length-based and age-based indicators for Mediterranean stocks was also conducted. Recruitment detection from survey time series showed uneven patterns over stocks and time series, and in some cases depended on survey timing. Length-based indicators exhibited weak consistency information from stock assessments, and confounding effects of biological variability and sampling timing on observed recruitment pulses were noted. The indicator L90R, calculated from the length-frequency distribution of fish larger than recruiting length, seemed to perform well among those inspected.

Thresholds for the D3C3 indicators for stocks representing different life-histories, data availability and MSFD (sub)regions (when possible) were discussed (ToR b and c). The suggested thresholds covered all approaches identified by WKD3C3SCOPE. Clear thresholds where the indicator signifies stock productivity declines could not be identified from the data as none of the age structure indicators showed a positive correlation with stock productivity. As a result, threshold levels cannot be determined based on levels at which stock productivity is either impaired or enhanced. In the absence of clear relationships between the indicators and stock health, the workshop used varying percentages (10th percentile, median/50th percentile) of the simulated or observed distributions of indicators to determine good status of the indicator. The analyses presented emphasized the direct influence of recruitment and growth on fisheries yield and precautionary fishing mortality limits. Finally, a decision tree to choose a threshold setting method was proposed for further testing in WKSIMULD3 on the basis of listed pros and cons discussed by WKD3C3THRESHOLD participants.

The SSB/R indicator responded to recruitment in an undesirable manner but there was insufficient evidence to determine which of the three remaining age structure indicators provided a higher signal to noise due to recruitment variability. Selectivity indicators under D3C3 were retained despite unclear guidance in the MSFD guidance document. The retained indicators for medium-lived stocks with age-based assessment data include ABI, POS, ASA, R, ASW, and Fjuv/Fapical. The value of age structure indicators as management indicators was unclear for

short-lived and long-lived species. For short-lived species, no strong link was found between age-structure indicators and F or SSB, and high age at spawning may lead to senescence rather than increased viability of spawning products. For long-lived species, age structure indicators appeared to react substantially later than F and SSB, making their added value for management unclear. The definition of thresholds for these indicators will be further investigated in WKSIMULD3.

The assessment of stock health under D3C3 relies on crucial data such as recruitment, weight at age, and size/age distribution (ToR d). In the absence of this information, D3C3 assessments cannot be conducted, and Member States were encouraged to enhance data collection efforts. For stocks with age-based assessments, these data are considered essential input and/or output for the assessment, and assessments based on age data are preferred over those based solely on length distributions for the estimation of age structure indicators.

Finally, the group drafted a list of actions to be completed for the reparation of WKSIMULD3.

ii Expert group information

Expert group name	WKD3C3THRESHOLD - Workshop to apply thresholds for the preselected indicators for MSFD D3C3
Expert group cycle	Annual
Year cycle started	2023
Reporting year in cycle	1/1
Chair(s)	Anna Rindorf, Denmark, Guiseppe Scarcella, Italy
Meeting venue(s) and dates	18-21 September 2023, Larnaca, Cyprus/online, 31 participants (6/25).

1 Terms of Reference

The workshop to apply thresholds for the preselected indicators for criterion D3C3 (WKD3C3Thresholds), chaired by Anna Rindorf, Denmark and Giuseppe Scarcella, Italy, met in Larnaca, Cyprus, from 18-21 September 2023 to:

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:
    a. Evaluate the consequences of failing to achieve the thresholds
    b. Evaluate consistency and complementarity with D3C1(FMSY) and D3C2 (MSYBtrigger)
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:
    a. The data requirements to assess D3
    b. Recommended indicator(s) for the assessment of D3C3 that are compatible with D3C1 and D3C2
    c. Methods to set thresholds and reference levels.

WKD3C3THRESHOLDS will report by 1st of November 2023 for the attention of ACOM.

2 Workshop approach

The meeting commenced with a round of introduction of all participants. The workshop agenda was briefly described and agreed upon. The workshop was organised as a series of presentations of the analyses' outputs with intermittent group discussions. Presentations focused on the setting of D3C3 thresholds in the MSFD context.

2.1 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 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.

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).
5. 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 WKD3C3THRESHOLDS in relation to WKD3C3SCOPE and WKSIMULD3

The three workshops on D3C3 in this series each cover separate aspects. The first workshop WKD3C3Scope identifies 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 are calculated, validated and evaluated for a selection of stocks representing different life-histories, data availability and MSFD (sub)regions. Subsequently, thresholds for these indicators are 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. 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 will explore the relationships between indicators of population traits/dynamics and healthy population structure through simulation. The workshop will infer 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.

3 ToR 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.

3.1 Indicators investigated at WKD3C3Thresholds

The candidate indicators identified in WKD3C3SCOPE that were also considered in WKD3C3Thresholds are listed in Table 3.1.

Table 3.1. Indicators for D3C3 identified in WKD3C3SCOPE as potentially useful and further investigated at WKD3C3Thresholds.

Indicators estimated in WKD3C3Thresholds	Indicators aspect	Rationale	Refs
ABI_{MSY} Proportion of old spawners by biomass Average age of spawners SSB/R	Proportion of large or old spawners	Ensure stock resilience by maintaining high numbers or proportion of old and large spawners	ICES 2016; 2017; 2021; Probst et al. 2021, Griffiths et al 2023, van Deurs et al 2023 Probst 2023
L90 of all lengths L90 of non-recruit lengths	90th percentile of length of individuals in the population		
Recruitment from stock assessment output Recruitment from survey catch at lengths below recruitment cut off	Recruitment	Adjustment D3C1 and D3C2 thresholds if necessary due to changes in productivity of stock reflecting environmental conditions	ICES 2021, European Commission 2022
Average weight at age anomaly (ASW) from stock assessment input	Growth		
Fjuv/Fbar Fjuv/Fapical	Maintain or decrease F of juveniles relative to older fish	Tracks changes in selectivity pattern	Vasilakopoulos et al. (2020); Probst et al. (2021); STECF (2020; 2021)

For convenience, the description of the indicators from WKD3C3SCOPE is given below.

3.1.1 Relative proportion of fish above A90, ABI_{MSY}

The new age-based indicator for commercial stocks (called ABI_{MSY}) that describes the proportion of older fish currently in a population (in a given year) relative to the proportion of older fish at equilibrium under constant fishing at F_{MSY} (Griffiths et al. 2023; Figure 3.1.1). Older fish are defined as fish above the age closest to the 90th percentile of the numbers at age distribution (often referred to as A90 and shown as A_{MSY} in Figure 4.7.1) at equilibrium when fishing at F_{MSY}. To date, ABI_{MSY} has been applied to 72 Category 1 stocks in the Northeast Atlantic. These data are taken from the FLR (Fisheries Library in R; Kell et al. 2007) stock assessment database that was collated at WKREF1 and is freely available from ICES (ICES, 2022). During the workshop, the indicator was also calculated on additional 11 stocks from the Mediterranean for which FLR stock objects were already available. ABI_{MSY} requires an age-based analytically stock assessment and therefore is only relevant for data-rich stocks.

Work is ongoing to adapt the code for shortlived stocks with quarterly data and for which advice is based on the escapement rule, hence these species were not included in estimations of ABI_{MSY}.

The framework (method used to simulate the age structure of a stock to equilibrium) behind ABI_{MSY} is built to be flexible and can consider any biological and fishing pattern assumptions as well as any chosen F (in terms of target F used). To date we have taken assumptions directly from each stocks assessment model, such that they represent our 'best' current understanding of the stock's status, its biology and the F pattern.

The main benefit of ABI_{MSY} is that it is based on the same methodology used for the calculation of SSB and F (and their respective reference points; B_{MSY} and F_{MSY}) and therefore fits nicely within the stock assessment and advice process. It also has a suggested reference point (and threshold), namely the age structure at equilibrium under F_{MSY} and therefore shares a common currency with assessments of D3C1 and D3C2. On the topic of GES, we have also calculated ABI_0 which is an additional indicator that compares the proportion of older fish in a population to the proportion of older fish at equilibrium under no future fishing.

The rationale behind ABI_{MSY} is based on the value of having older (and larger) fish in the population, as they are expected to provide greater spawning potential and resilience to perturbations (Barnett et al. 2017; Hixon et al. 2014). An added benefit of ABI_{MSY} is that by comparing to a MSY reference level, indicator values around 1 will indicate the number of older fish that are needed to theoretically sustain B_{MSY} (and F_{MSY}) in the long term (Figure 3.1.1).

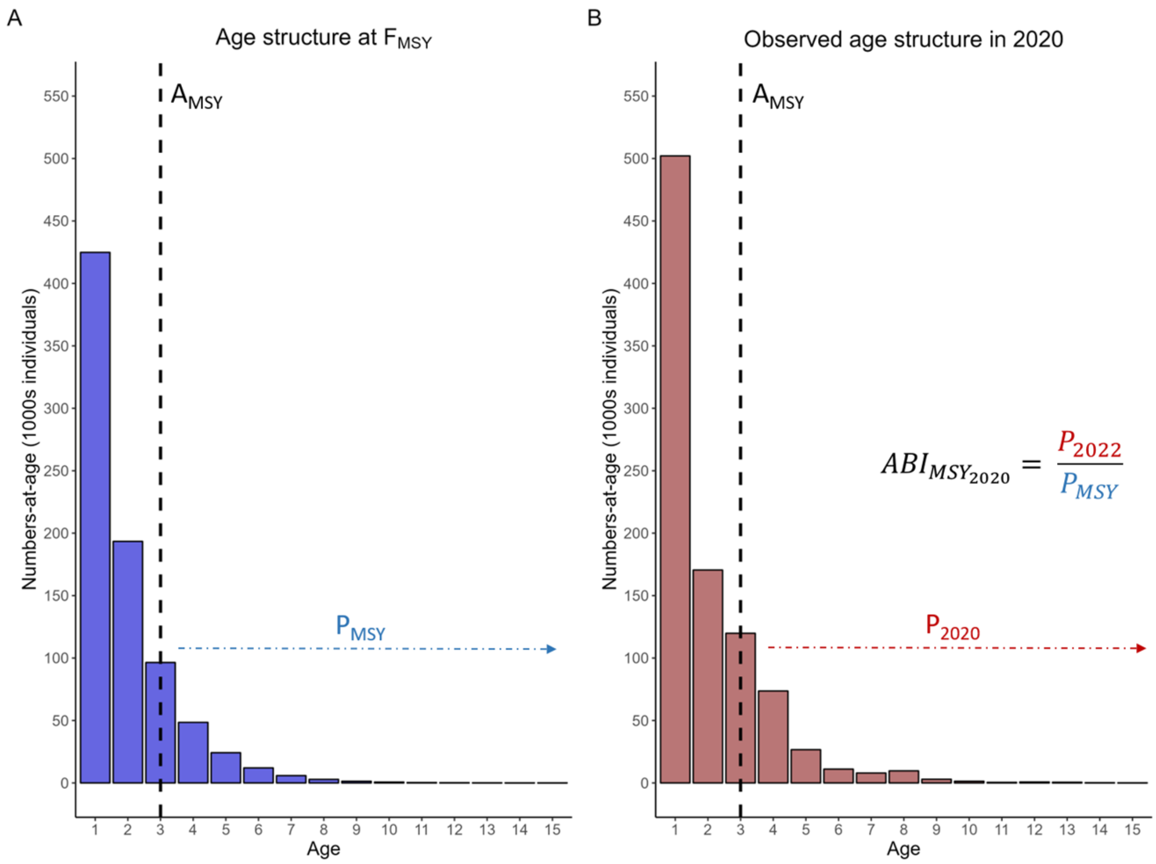

Figure 3.1.1: Graphical example of the methods used to calculate ABI_{MSY}. Shown is the theoretical age structure at equilibrium under F_{MSY} (A) and the estimated age structure in the final year of the assessment (B). Both age structures relate to the hake (*Merluccius merluccius*) stock (hke.27.3a46-8abd) in the Greater North Sea, Celtic Seas, and the northern Bay of Biscay. Dashed vertical lines illustrate the reference age under F_{MSY} (black; A_{MSY}) conditions, whereby the reference age is the closest age to the 90th percentile of the fish numbers-at-age distribution (the A90 age). P_{MSY} and P_{2020} then represent the proportion of individuals above A_{MSY} and are used to calculate ABI_{MSY} in a given year (here 2020).

The indicator is summarised in the table 3.1.1 below.

Table 3.1.1. Indicator summary for ABI_{MSY}

Name of the indicator	ABI_{MSY}
Short description	ABI_{MSY} describes the proportion of older fish in a population in a given year relative to the proportion of older fish in the population at equilibrium under constant fishing at F_{MSY}. To define older, we have used the number of fish above the 90th percentile of the numbers at age distribution. The framework behind ABI_{MSY} is built to be flexible and can consider many biological and fishing pattern assumptions as well as any chosen F (in terms of target F used; e.g. F target associated to any given B target or F_{SPR0}). Indicator values under F_0 are available for comparison. The main benefit of ABI_{MSY} is that is based on the same methodology used for the calculation of SSB and F (and their respective reference points) and therefore fits nicely within the stock assessment and advice process. It also has a reference point, namely the age structure at equilibrium under a given target F.
Reference	Griffiths et al. (2023)

Name of the indicator	ABI_{MSY}
Pressure or state indicator?	State
Aspect of health addressed	Age structure
Benefit to the population of high or low indicator values	Low indicator values = low relative abundance of older fish possibly limiting spawning potential, resilience and recovery High indicator values = high relative abundance of older fish indicating good spawning potential and resilience and recovery.
Benefits supported by empirical evidence	
Benefits supported by simulation/theoretical considerations	Simulation work has shown that ABI_{MSY} has a good classification skill for the biomass threshold of 80% of B_{MSY} and is capable of tracking changes in both B and F in a range of stocks.
CV of indicator	The current version of ABI_{MSY} is deterministic but work (and funding applications) are ongoing to define both the indicator and the reference point in terms of probability via stochastic simulations.

3.1.2 Proportion of old spawners

Proportion of old spawners (POS) were analysed by (van Deurs et al., 2023) is defined as

$$POS = \frac{SSB_{old}}{SSB}$$

Where SSB_{old} if the spawning stock biomass (SSB) over the age a_{old}. This age will vary among stocks. The indicator essentially defined the fraction of the mature stock that it is comprised of old fish. It is measured in biomass rather than in numbers to reflect that older fish are rarer but produce more spawning products. The indicator is summarized in table 3.1.2. below.

Table 3.1.2. Indicator summary for Proportion of old spawners (POS)

Name of the indicator	Proportion of old spawners (POS)
Short description	The indicator measures the relative change in old (age) spawners. $POS = \frac{SSB_{old}}{SSB_y}$
Reference	Van deurs et al. 2023
Pressure or state indicator?	State
Aspect of health addressed	Age structure of the spawning stock to provide good recruitment
Benefit to the population of high or low indicator values	Increased fecundity with high values
Benefits supported by empirical evidence	Several studies in the literature, see e.g., Hixon et al 2014

Name of the indicator	Proportion of old spawners (POS)
Benefits supported by simulation/theoretical considera-tions	Some, however the proposed effect on recruitment is un-clear.
CV of indicator	

3.1.3 Average spawner age in the stock

Average spawner age (ASA) is defined as

$$ASA = \frac{\sum_{a_{min}}^{a_{max}} n_a \, m_a \, a}{\sum_{a_{min}}^{a_{max}} n_a \, m_a}$$

Where a is age, n is numbers at age, and m is maturity. This indicator calculates the average spawner age, weighted by the cohort sizes. The indicator is summarized in table 3.1.3 below.

Table 3.1.3. Indicator summary for Average Spawner Age (ASA)

Name of the indicator	Average spawner age
Short description	The indicator measures the average age of the popula-tion $$ASA = \frac{\sum_{a_{min}}^{a_{max}} n_{a,y} \, mat_y a}{\sum_{a_{min}}^{a_{max}} n_{a,y} \, mat_y}$$
Reference	van Deurs et al. 2023
Pressure or state indicator?	State
Aspect of health addressed	Age structure of the entire stock to provide good recruit-ment
Benefit to the population of high or low indicator values	Increased fecundity with high values
Benefits supported by empirical evidence	Several studies in the literature, see Hixon et al 2014
Benefits supported by simulation/theoretical considerations	Some, however the proposed effect on recruitment is unclear.
CV of indicator	

3.1.4 SSB/R

Probst (2023) identified SSB/R as a very good proxy for the annual mean age (A_{mean}) within a stock (Figure 3.1.2). Contrary to A_{mean}, SSB/R has the advantage of being easily implementable for a wide array of stocks as time series of SSB and R are readily accessible (e.g. through the R-package ‘icesSAG’ or ICES Advice sheets), whereas matrices of number at age from analytical stock assessments - which are necessary to calculate A_{mean} - are often only accessible from stock assessment working group reports (from which data has to be cut and copied manually).

The time series of SSB/R can be assessed in the sense of a surveillance indicator similar to R (see section 4.10). Surveillance indicators are indicators which trigger additional management action once the indicator metric leaves known bounds (Shephard et al., 2015; Rufino et al., 2018). Such additional management actions could be the initiation of a research program or the implementation of more restrictive (and thereby precautionary) management measures.

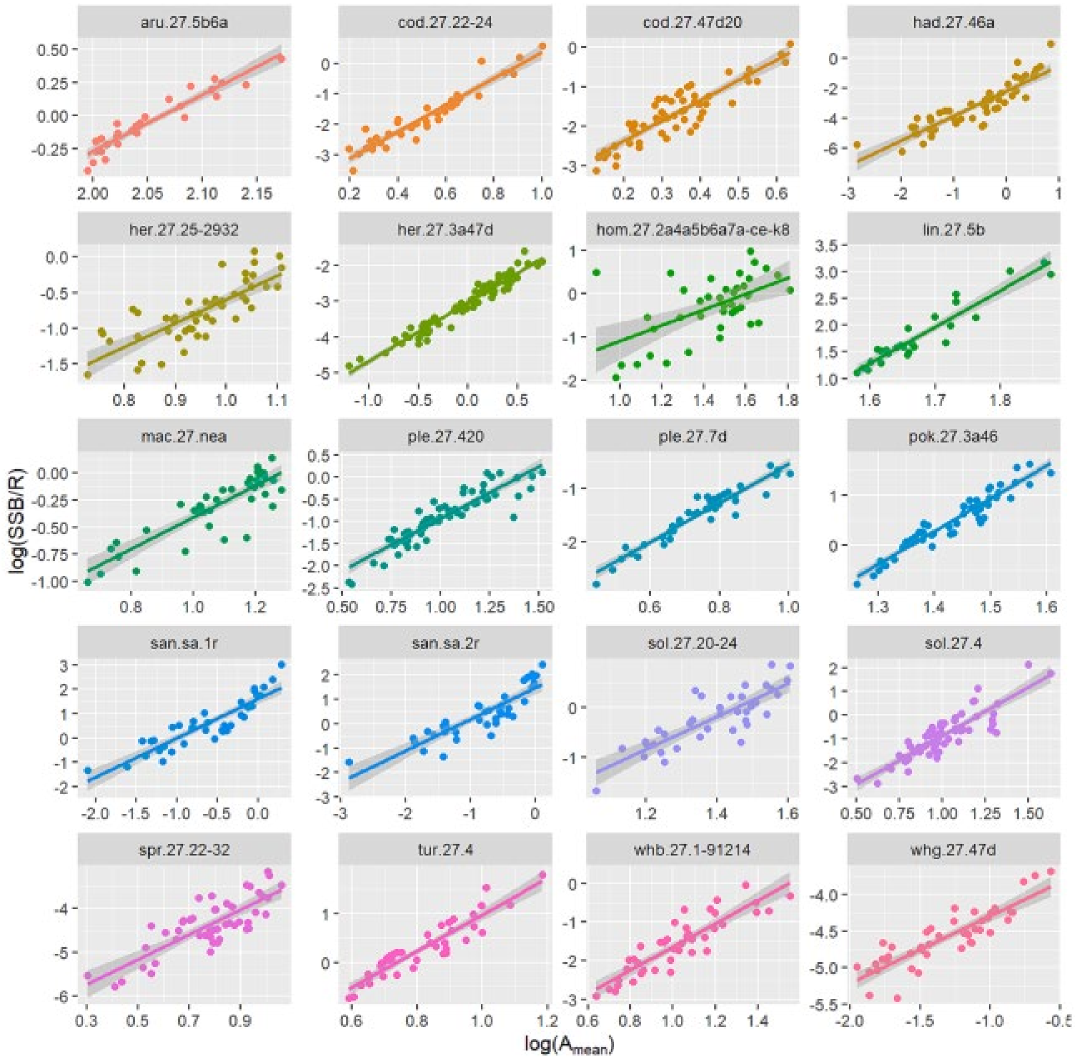

Figure 3.1.2. Relationships between mean age (A_{mean}) and SSB/R for 24 fish stocks from the North Atlantic. Note the logarithmic scale on both axes.

The indicator is summarised in table 3.1.4 below.

Table 3.1.4. Indicator summary for SSB/R

Name of the indicator	SSB/R
Short description	Time series of ratio SSB/R Threshold is minimum of historic time series
Reference	Probst, 2023
Pressure or state indicator?	State indicator

Name of the indicator	SSB/R
Aspect of health addressed	Proportion of old individuals (proxy for mean Age [A_{mean}])
Benefit to the population of high or low indicator values	At high values relative high abundance of old individuals
Benefits supported by empirical evidence	Stocks with large proportion of old individuals might indicate low fishing pressure and hence good conditions for growth and survival
Benefits supported by simulation/theoretical considerations	High abundance of large individuals buffers recruitment against environmental variation
CV of indicator	10s – 100s

3.1.5 L90

A suite of size indicators for coastal fish communities in the Baltic Sea have been suggested. including mean and median length, 10th and 90th-percentile of the length distribution (L10, L90), mean length of the 10% largest fish (Lmax), Large Fish Indices, Size-spectra slope and Size-diversity. Östman et al (2023) compared this suite of size indicators and found good precision and accuracy of most indicators at realistic sample sizes, except for Size-spectra and Size-diversity. Different size indicators were correlated among sites, indicating similar responses to environmental variation. Most size indicators responded positively to lower fishing pressure, especially indicators emphasizing the largest individuals in the population (e.g. L90 and Lmax), whereas eutrophication and physical disturbances had less impact on indicator variation. Östman et al. (2023) concluded that size-based indicators aiming at describing the occurrence of larger fish, like L90 and Lmax, are useful for establishing management targets and evaluate the status of coastal fish. Within HELCOM, the indicator L90 was thus agreed upon and implemented in Helcom HOLAS III (the Helcom core indicator 'Size structure of coastal fish', https://indicators.helcom.fi/indicator/coastal-fish-size/) to assess the status of key coastal (non-commercial) fish species. A threshold for good status has been developed (Bolund et al in prep) and likewise implemented within HELCOM HOLAS III for the key species perch. The indicator is summarized in table 3.1.5 below

Table 3.1.5. Indicator summary for Survey-based L90 for perch in the Baltic Sea

Name of the indicator	Survey-based L90 for perch in the Baltic Sea
Short description	The indicator evaluates the size structure of perch in the Baltic Sea using the size of the fish at the 90th percentile of the length distribution (L90) from fisheries independent surveys (using gill nets and fyke nets). The indicator is also associated with a threshold representing good environmental status.
Reference	Östman et al. in review ICES J Mar Sci; Bolund et al. in prep; HELCOM HOLAS III indicator report: Size structure of coastal fish
Pressure or state indicator?	State
Aspect of health addressed	Size structure of coastal fish populations/stocks in the Baltic Sea

Name of the indicator	Survey-based L90 for perch in the Baltic Sea
Benefit to the population of high or low indicator values	An indicator value above the threshold indicates a more natural size structure where large fish in the population are present. The presence of large fish does in turn have positive effects for reproduction and trophic regulation in the food web.
Benefits supported by empirical evidence	Yes
Benefits supported by simulation/theoretical considerations	To some extent theoretical considerations
CV of indicator	Moderate to low concerning both spatial and temporal variation in the Baltic Sea. The CV of the data used to establish a threshold for the Baltic Sea is 0,1005 in the raw data and 0,0761 in the predicted data from the statistical model used to set the threshold.

3.1.6 Recruitment

Recruits can be considered as an essential component of the stocks age/size structure representing the productivity of a stock (Probst, 2023). Afterall, recruitment will determine the persistence of a stock through time and will shape other fundamental metrics of the stock such as stock biomass (B or SSB) or the abundance of old individuals in subsequent years. Hence recruitment is a vital stock parameter and its assessment as a surveillance indicator (Shephard et al., 2015) can be used to determine if a stock is threatened in its persistence and if more precautionary harvest options might need to be considered.

Time series of recruitment (R) are already a standard product of analytical stock assessments. but it is usually not assessed within the management framework of the CFP.

Recruitment time series may not be available for all stocks that need to be assessed within the MSFD, but for some stocks proxies for R could be obtained from survey data as suggested by Froese et al. (2015). In WKD3C3Thresholds, recruitment was defined as the first mode of the length distribution and different ways to investigate this mode considered.

The indicator is summarized in table 3.1.6 below.

Table 3.1.6. Indicator summary for Recruitment (R)

Name of the indicator	Recruitment (R)
Short description	Time series of recruitment (obtainable from ICES advice sheets) or from analyses of survey data
Reference	ICES Advice, STECF
Pressure or state indicator?	State indicator
Aspect of health addressed	Productivity of the stock
Benefit to the population of high or low indicator values	At high values population is viable
Benefits supported by empirical evidence	Stocks with high recruitment will persist, while stocks with low recruitment become scarce and eventually may disappear

Name of the indicator	Recruitment (R)
Benefits supported by simulation/theoretical considerations	Recruitment is a fundamental process of population dynamics, see Beverton and Holt (1957), (Jennings et al., 2001)
CV of indicator	In the range of recruitment fluctuations

3.1.7 Average weight at age of spawners

Average spawner weight, which takes into account that different years may have changes in weight at age, calculated as

$$ASW = \frac{\sum_{a_{old}}^{a_{max}} n_a \frac{w_a}{\overline{w}_a}}{\sum_{a_{old}}^{a_{max}} n_a}$$

Where w_a is the weight at age, and $\overline{w}_a$is the weight of old spawners. This indicator monitors the relative change in weight across age groups. The indicator is summarized in table 3.1.7 below:

Table 3.1.7. Indicator summary for Average spawner weight

Name of the indicator	Average spawner weight
Short description	The indicator measures the average weight of spawning individuals in a stock $$ASW = \frac{\sum_{a_{old}}^{a_{max}} n_a \frac{w_a}{\overline{w}_a}}{\sum_{a_{old}}^{a_{max}} n_a}$$
Reference	van Deurs et al. 2023
Pressure or state indicator?	State
Aspect of health addressed	Condition and age structure of the stock
Benefit to the population of high or low indicator values	Increased fecundity with high values
Benefits supported by empirical evidence	Several studies in the literature, see Hixon et al 2014
Benefits supported by simulation/theoretical considerations	Some, indication of positive effects on recruitment of some stocks.
CV of indicator	

3.1.8 Fjuv/Fapical

Fisheries science distinguishes between two distinct aspects of exploitation of commercial stocks: exploitation rate, capturing the fishing intensity and typically quantified by Fbar, and population selectivity, capturing the way that fishing is distributed across the different demographic components of a population. Population selectivity is a result of both the gears used (e.g., choice of mesh size) and fish availability (e.g., due to choice of fishing time and location).

Vasilakopoulos et al. 2020 introduced three 'utility criteria' for fisheries selectivity management indicators: i) the ability to track selectivity changes in the fishery; ii) robustness to recruitment variability; iii) robustness to changes in Fbar. Subsequently, a range of different candidate selectivity indicators were tested against these three criteria. First, changes in selectivity, recruitment and Fbar were simulated on a virtual fish stock to study the indicators under controlled conditions. Then, the indicators were applied to six European fish stocks with a known history of technical measures to explore the indicators' response in real-world situations. This process identified indicators estimated as the ratio of F of small/juvenile fish (e.g., Frec – F of the first recruited age-class - or similar) to the F of fully selected fish (e.g., Fbar or similar) as those fulfilling the three utility criteria. By contrast, catch-based and abundance-based indicators were found to be sensitive to recruitment fluctuations.

Further applications of such F-based selectivity indicators were carried out by STECF 20-02 (STECF 2020), STECF 21-07 (STECF 2021), Probst et al. (2021) and Probst (2023). These studies applied F-based selectivity indicators to a great variety of fish stocks, illustrating that alternative configurations of the Frec/Fbar indicator, such as Fjuv/Fbar (Fjuv: the average F of the juvenile age-classes) and Fjuv/Fapical (Fapical: the F of the fully selected age-class), could be also informative.

Currently, the JRC is conducting further research on selectivity indicators to support the implementation of the TMR (Regulation (EU) 2019/1241), with a particular focus on estimating relevant thresholds based on the MSY principle. This ongoing investigation points to Fjuv/Fapical as the most suitable configuration of the indicator.

A key strength of Fjuv/Fapical (or similar) is that it reflects processes related to the fisheries selectivity in a way that is robust to both recruitment variability and overall fishing pressure. Additionally, being a 'pressure' indicator it responds more directly to management decisions than 'state' indicators. Finally, the use of this selectivity indicator would link D3C3 with fisheries regulations (TMR and CFP), similarly to D3C1 and D3C2.

The main limitations of the indicator is that it can be only applied to stocks with age-structured stock assessments and that the quality of the indicator estimates depends on the quality of the stock assessment. The indicator is summarised in the table below.

At present, selectivity indicators such as Fjuv/Fapical are not explicitly addressed within D3, a fact that has already been noted by Probst et al. (2016). These authors suggest that D3 could reflect the impacts of fishing on exploited stocks more comprehensively, if selectivity was included as an additional pressure indicator within D3. The indicator is summarized in table 3.1.8 below

Table 3.1.8. Indicator summary for Fjuv/Fapical

Name of the indicator	Fjuv/Fapical
Short description	Selectivity indicator capturing the differential exploitation of 'small' versus 'big' fish. In this configuration, 'small fish' refer to the juvenile age-class(es) (Fjuv), while 'big fish' refer to the fully selected age-class (Fapical)
Reference	Vasilakopoulos et al. 2020
Pressure or state indicator?	Pressure
Aspect of health addressed	Protection of juveniles to attain higher long-term SSB and yields
Benefit to the population of high or low indicator values	Lower values indicate lower relative exploitation of small fish/juveniles, hence the fishery allows more fish to enter the adult fraction.

Name of the indicator	Fjuv/Fapical
Benefits supported by empirical evidence	Increased mid/long-term SSB (Vasilakopoulos et al. 2011)
Benefits supported by simulation/theoretical considerations	Increased mid/long-term SSB and yields (Vasilakopoulos et al. 2020; STECF 2021)
CV of indicator	Stock-dependent, but generally lower than catch-based or population-based indicators. Typically <1.

3.2 Data used for the 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 meeting, the stock objects for five shortlived stocks of small pelagics (spr.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 3.2.1, 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 described below, 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/FMSY over the whole stock assessment time series, whereby the length of that time series differed across stocks (Figure 3.2.1).

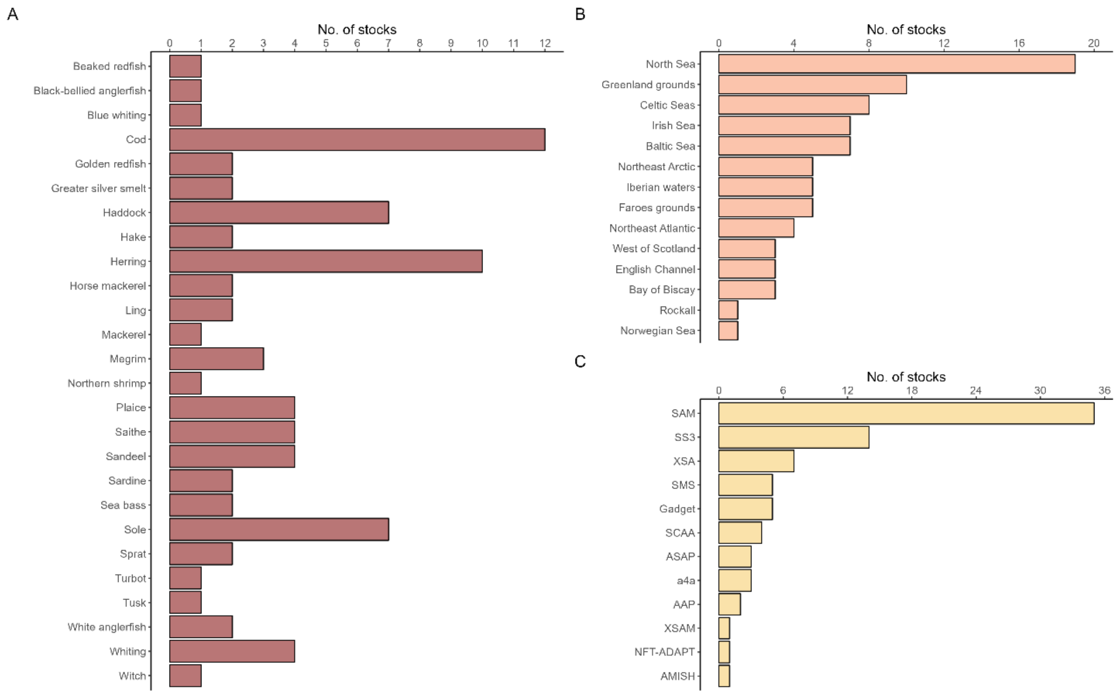

Figure 3.2.1: Graphical of the Northeast Atlantic dataset. The dataset contains 26 species (A) from 14 different areas (B), whereby area is a broad characterisation based on ICES areas and ecoregions. All 81 stocks are assessed using age-structured analytical assessment frameworks (C).

3.3 Data used for the 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 FMSY and BMSY. 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 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 files that landed on the working group SharePoint where the timeseries of the length-based indicators.

Table 3.3.1. 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

3.4 Documentation of calculation procedures

3.4.1 Calculation procedure for ASA and POS

Average stock age (ASA) and Proportion of Old Spawners (POS) are two suggested indicators that depend on the numbers at age, and biomass at age respectively. ASA is calculated as

$$ASA = \frac{\sum_{a_{min}}^{a_{max}} n_{a,y}\, mat_y a}{\sum_{a_{min}}^{a_{max}} n_{a,y}\, mat_y}$$

Where n is the numbers at age, a_{min} is the minimum age of individuals included. In the cases used in this report, recruits were excluded from all calcuations in order for the number of recruits entering the population to not have a negative effect on the indicator. mat is the maturity ogive of the stock. a_{max} is the maximum age used in the calculation, and will often be the plus group, and a is age.

Proportion of old spawners is a biomass based indicator that measures the fraction of old spawners to all spawners in the population as

$$POS = \frac{SSB_{old}}{SSB_y}$$

This requires a choice of which age (or size) that determines when a spawner is old. Following van Deurs et al (2023) we set the aeg at which a spawner is old such that the historical time series has an average POS of as close to 0.5 as possible. For short lived stocks, which may only have 2-3 ages that spawn, there might not be a number that is close to 0.5, and may therefore influence the result. Since this historical choice is fairly arbitrary the raw number of the indicator is not as important as the temporal dynamic of the indicator.

Since fishing often reduces the number of old (large) fish in a stock, both ASA and POS both respond to increasing or decreasing fishing mortality in a population.

3.4.2 Calculation procedure for SSB/R

SSB/R can simply be calculated from time series of SSB and R extracted from stock assessment data. Stock data can either be extracted from the stock assessment summary sheets found in the online ICES advice, from the according working group reports or via the ICES API using the R-package 'icesSAG'.

3.4.3 Calculation procedure for ABI_{MSY}

ABI_{MSY} is calculated in a two-stage approach. Firstly, a stock's age structure is simulated forward to equilibrium based on its current state and a fixed fishing level. We then calculate reference levels from that equilibrium age structure, and compare the realised state of the stock (in a given year) to those reference levels. In general, the calculation of ABI_{MSY} is based on the same procedure used for reference point estimation (e.g., B_{MSY} and F_{MSY}) and the evaluation of stock status in terms of those reference points (e.g., B/B_{MSY} and F/F_{MSY}). A detailed description of the calculation of ABI_{MSY} and its application to 72 stocks can be found in Griffiths et al. (2023). Here, we briefly summarise those calculations.

Age structure at equilibrium

To estimate a stocks age structure at equilibrium, we first run deterministic long-term forecasts (approx. 200 years) under a constant F. For ABI_{MSY}, this constant F is fixed at each stocks ICES F_{MSY} value, although any F could be used. For instance, for the Mediterranean stocks an F_{target} of 0.1 is sometimes used, or one could opt to evaluate the age structure of stock at equilibrium under no future fishing (F = 0; as shown in Figure 3.1.1). Once equilibrium is attained (when SSB achieves steady-state dynamics), we then extract the corresponding age structure at F_{MSY}. It is important to stress that deterministic F_{MSY} is usually higher than stochastic F_{MSY}.

During these forecasts, it is assumed that weight-at-age, maturity-at-age, mortality-at-age and selectivity is the average of the last three years. Thus, our forecasts to equilibrium are based on the current state of a stock. A stock recruitment relationship in the form of a segmented regres-sion with a fixed breakpoint at ICES B_{lim} is used. Again, B_{lim} is currently used but in theory any lower biomass value could be used. For example, in many stocks B_{lim} is assumed to be B_{loss} (the lowest observed biomass value) and this could also be used when estimating a stocks age struc-ture at equilibrium assuming that B_{loss} doesn't cause impaired recruitment.

Calculation of ABI_{MSY}

Once the age structure of a stock at F_{MSY} has been extracted, we first remove the first age class to limit the potential impact of incoming recruitment. ABI_{MSY} is then calculated by first taking the 90th percentile of the equilibrium age structure and rounding to the nearest integer age. We call this age A_{MSY} and calculate P_{MSY}, which is the realised percentage of fish above A_{MSY}. It is note-worthy that P_{MSY} is not necessarily 10% due to the discretisation of numbers-at-age data. Moreo-ver, we have opted to use the 90th percentile as an indicator of 'older' fish (based on tests detailed in Griffiths et al. 2023), however, in principle any cut off could be used.

For any given year *t*, we then identify A_{MSY} in the estimated numbers-at-age distribution and calculate the proportion of fish above it, P_t. ABI_{MSY} is then calculated as:

$$_t = \quad _t /$$

This process is then repeated for all years and all stocks, producing a stock-specific time series of ABI_{MSY}. A graphical example of this calculation is provided in Figure 3.1.1.

3.4.4 Calculation of ASW and R

Average stock weight (ASW) is a stock size independent method of calculation the average change in weight in a stock in a time series, to provide an indicator that describes the trend of somatic growth in the population (van Deurs et al 2023).

$$ASW = \frac{\sum_{a_{min}}^{a_{max}} n_a \frac{w_a}{\overline{w}_a}}{\sum_{a_{min}}^{a_{max}} n_a}$$

Where $\overline{w}$ is the long term average weight in the population, the indicator appears to be somewhat positively associated with fishing (van Deurs et al 2023), where one can speculate that high fishing mortality leads to a low amount of density dependent growth.

3.4.5 Calculation of length based indicators to be compared with age based indicators

MEDITS survey data were analyzed with JRC scripts (Mannini, 2020) to compute standardized annual length frequency distribution (LFD) by stock. LFD output is number/km_2 by length class (cm) and year. The indicators built on LFD were sum of the individuals smaller than length corresponding to separation of recruits and older (R), 90th percentile of the length distribution (L90), 90th percentile of individuals larger than the length corresponding to the separation of recruits and older (L90R). Two methodologies to compute the threshold to separate recruits and older were considered:

- a dynamic threshold based on the application of the Von Bertalanffy growth equation, deriving k,t0 and Linf from available stock assessment (STECF, 2022, 2023) and t equal to the mean period of the survey;
- visual inspection aiming to discriminate the first modal component (when present).

A dynamic threshold results appeared to be unstable, and the fixed threshold approach was then applied to produce the indicators plots. The threshold selected is provided in table 3.3.1. The R indicator was defined as the sum of individuals below the recruitment threshold. L90 was defined as the 90^{th} percentile computed over the LFD of each year and L90R was calculated as L90 after removing the recruits from the data. Code is available on SharePoint/GitHub.

3.4.6 Calculation procedure for selectivity indicators

Three configurations of the F-based selectivity indicator were presented and discussed during the WK, namely Frec/Fbar, Fjuv/Fbar and Fjuv/Fapical. Frec is the fishing mortality on the first age class that appears in the stock assessment, Fjuv is the average fishing mortality over the age-classes with a maturity of less than 0.5, Fbar is the average fishing mortality on the age classes most represented in the catch, and Fapical is the highest fishing mortality observed in any age-class (i.e., that of the fully selected age-class). To calculate the juvenile age-classes for Fjuv, the average maturity by age over the last 20 years was calculated.

Fjuv/Fapical was the configuration retained, as it allowed the most straightforward estimation of threshold values.

3.4.7 Defining longevity

Different approaches to defining longevity groups were investigated. Using maximum age from fishbase resulted in some larger gadoids being identified as short lived whereas some smaller pelagics were identified as long lived. This seemed counterproductive. Longevity groups were instead defined as:

Short: small pelagics (sprat, sardine, sandeel and herring) and pandalus

Medium: all stocks not identified as short- or long-lived

Long: Species identified by WKABSENS as sensitive to fishing (monkfishes, ling, tusk, golden redfish)

This definition does not account for the fact that longevity may also vary within a species, as e.g. sprat grows slower and dies as a lower rate in the Baltic than in the North Sea and the same is true of cod in the Barents Sea compared to other areas. However, it was considered a reasonable approximation for the purpose of the workshop.

3.4.8 R-code available

As part of the WKD3C3Thresholds meeting, the code used to calculate all seven indicators were compiled into easily useable R functions and are available via the meetings GitHub repository (https://github.com/ices-eg/WKD3C3Thresholds). All functions require a single stock object of the 'FLStock' class (Kell et al., 2007) as input and output a vector of indicator values alongside information on years, F level or minimum age (when and if appropriate). The only deviance from this is for ABI_{MSY}, where an F_{target} (default = F_{MSY}) and a threshold value (default = 0.9) is required (as inputs). The threshold refers to the percentile cut off used to represent older fish in a stocks number-at-age distribution at equilibrium.

The GitHub repository also contains code to collate all seven indicators into a single data frame and then plot alongside other indicators of stock status (SSB, R and F). The plotting code is the same used to produce the stock specific plots found in Annex 2. Information on R packages and versions used are provided at the top of each R script.

3.5 Results

The indicators were estimated for all stocks where sufficient data was available. As ABI_{MSY} requires information on F_{MSY} and B_{lim} (or a proxy thereof), ABI for the 81 stocks which had F_{MSY} defined (not the added short lived stocks). However, plots comparing indicators include data only for stocks where all indicators were estimated. The temporal developments in indicators of all stocks are given in Appendix 2.

Inspecting the stock specific plots one by one, the group found that the indicators follows temporal development of SSB and react similarly to F in many stocks. Examining the plots for individual stocks, there appeared to be a time gap of up to 10 years between a change in F and the subsequent occurrence of a change in ABI, ASA and POS in long lived species. In contrast, SSB of long lived stocks responded very quickly to changes in F and reasonably quickly to changes in recruitment. For these stocks, the value of ABI and ASA would therefore not lie in eliciting management action, as this would be more timely if based on other indicators (F, SSB and R). For medium lived stocks, the four age structure indicators show very similar temporal patterns but generally, SSB divided by R was considered by the group to be more variable than the other two indicators when evaluated qualitatively in plenary. The indicators follow temporal development of SSB and react similarly to F in many stocks. The indicators recruitment and mean weight at age documented productivity shifts in many stocks. Trends in recruitment were seen both with and without trends in weight at age and also with opposite trends in weight at age. The changes in recruitment in particular impact the four age structure indicators and under these changes, the age structure indicators will not react to changes in F in the way expected based on simulations without change in recruitment.

The investigated plots of F, recruitment and weight at age were considered useful to understand when changes in F do not impact SSB as expected. There are stocks where high biomass was maintained under decreased recruitment as F was decreased quickly in response, demonstrating that a decreased recruitment does not necessarily lead to a severe decline in SSB. The plots also demonstrated that having a high ABI/ASA/POS or SSB/R does not appear to increase recruitment at lag 0. It was unclear if this was caused by the higher responsiveness of recruitment to other factors such as environmental factors. ABI, ASA and POS generally show a stronger correlation to F than SSB/R.

3.5.1 Comparison of length based and age based indicators

Time series of numbers and length distribution of recruitment and the non-recruits for the Mediterranean stocks were presented to the group. The ability of the survey to intercept recruits and the consistence of recruitment detection over the time series was discussed. It was observed that recruitment detection seemed uneven over the stocks and the time series. In some cases, the recruitment detection was dependent on survey timing. The case of red mullet, strong recruitment events were intercepted in years when the survey was delayed. The group inspected the length-based indicators and visually compared them to the age-based indicators, and to the F and SSB trends for the same stocks. It was noted that there was a weak consistency between the trend observed in the length-based indicators and the other information inspected. However, often the length of the time series did not coincide among the data under consideration, because stock assessments started a few years later than the MEDITS survey. The recruitment indicator was available for hake and red mullet, with the latter oscillating because of the availability to the survey only in the years when the sampling was conducted in late summer/fall (after the expected sampling time). The influence of recruitment detection on the absolute value and the trend of percentile-based length indicators has been discussed elsewhere (ICES, 2017). The confounding effect of the biological variability and of the sampling timing on the observed recruitment pulses has received less attention and requires caution when using length-based indicators from survey data. L90R, which is the L90 calculated on the left truncated LFD, was the indicator performing best among those inspected.

4 ToR b: Derive thresholds for the D3C3 indicators for stocks representing different life-histories, data availability and MSFD (sub)regions

4.1 Synthesis of characteristics of healthy populations from WKD3C3SCOPE

It was broadly considered a central paradigm that healthy stocks allow population persistence under disturbance by exhibiting a sufficiently high productivity through high recruitment, condition and growth, a sufficiently low natural and fisheries related mortality. As a result of these aspects, a healthy stock will have a high biomass with a relatively wide age structure for the given species and will be able to recover quickly from perturbations such as a few poor recruitment or survival years. It was noted that a wide age distribution may provide greater benefits in long lived than short lived stocks, and as a result, short and long lived stocks may require different definitions of a healthy stock. It was also mentioned that environment as well as stock biomass impact productivity and health of a stock, and that the effect of environment is particularly high for recruitment and short lived stocks.

A range of principles which potentially can be used to define healthy stock indicator thresholds were mentioned. These included:

- Level of the unfished stock (stock not fished for at least 3 generations)
- Level at which the indicator aspect is not healthy following principles of D3C2. For example, recruitment may deteriorate below a specific level of average age of spawners
- Level of indicator of stocks considered to be in good health by expert judgement
- Level of indicator of stocks considered to be in good health by quantitative approaches applied to historical data. This includes breakpoint analyses and percentiles in historical data.
- Level of the indicator when the stock is fished according to MSY principles and provides for human needs.
- Level of the indicator when the stock provides for foodweb needs. The exact meaning of this level was unclear to WKD3C3THRESHOLDS and this principle was therefore not considered further.

The characteristics of a fishing mortality and biomass being, consistent with thresholds defined under D3C1 and D3C2 was also mentioned.

4.2 Examples of potential threshold methods

4.2.1 ABI_{MSY}

ABI_{MSY} by definition, relates the current age structure of a stock in terms of older fish to the age structure at equilibrium when constantly fished at F_{MSY}. This means that an indicator value greater than 1 indicates that a stock currently has more older fish than at under F_{MSY}, whereas a value below 1 indicates a lack of older fish. As the threshold of ABImsy is derived from deterministic simulations at equilibrium, we would expect that a stock managed at Fmsy (and with a biomass of Bmsy) would have an age structure that fluctuates around the threshold. This fluctuations suggest that a median is not an ideal threshold and future work is needed to incorporate natural variability and select a more appropriate percentile threshold for ABI.For instance, by many other countries a biomass threshold of 80% of Bmsy is used to define a healthy state due to natural fluctuations around Bmsy. Griffiths et al. (2023) used simulation testing (including differing levels of stochastic recruitment and applied to six stocks) to show that having an ABI_{MSY} value above 0.8 has a good classification skill for the same biomass threshold, albeit this does deteriorate for short-lived species and under high recruitment variability. Therefore, when variability is included this might prove to be a suitable indicator of GES.

4.2.2 Recruitment (suggestion 1), AWS, ASA and POS

If no thresholds are given for a given indicator, one suggestion is to use the historical time series to construct indicator thresholds. Since most fisheries time series are fairly limited in their time span (for instance many stocks in the North Sea begin their assessments around 30 years ago), it is difficult to establish a rank order based on e.g., the 10th and 90th percentile, as the comparison has to be done to a 6 year average. Additionally, indicators like recruitment tend to have several years of very large and very small outliers, and in shorter time frames, these big swings will determine the rank-order percentiles. A potential solution to this issue is to take a smoothed (6 year) mean over the historical period to create a historical trend line, and then take the 10th and 90th percentile of that time series. While it shortens the time series (since the first 5 years cant be used), it also makes the historical dynamic better comparable to the 6 year mean that is required for the MSFD assessment period. The method is visualized in figure 4.2.2.

4.2.3 Recruitment (suggestion 2)

Probst (2023) applies segmented regression to the time series of R within a reference period to determine a historic minimal mean against which average recruitment in the MSFD assessment cycles (2004-2009, 2010-2015 & 2016-2021) are assessed (Figure 4.2.1).

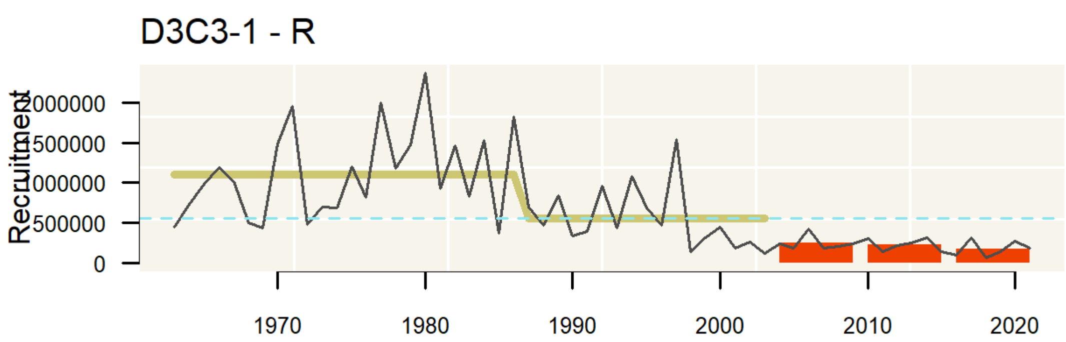

Figure 4.2.1. Tine series-based assessment of recruitment of North Sea cod Gadus morhua (cod.27.47d20). The brown line represents the segmented regression of the reference time series to obtain the minimal mean (blue dashed line), which is used as assessment threshold for the six-year means of R in the according MSFD assessment cycles (here red bars).

Besides the break point analysis suggested by Probst (2023) (see Figure 4.2.1), alternative methods to define thresholds for recruitment are possible. For example, a threshold definition by the

10th-percentile of the recruitment-time series in the reference period was suggested in WKD3C3THRESHOLD (Figure 4.2.2). Similar to the break point analysis, the 10th-percentile indicates situations in which recruitment falls.

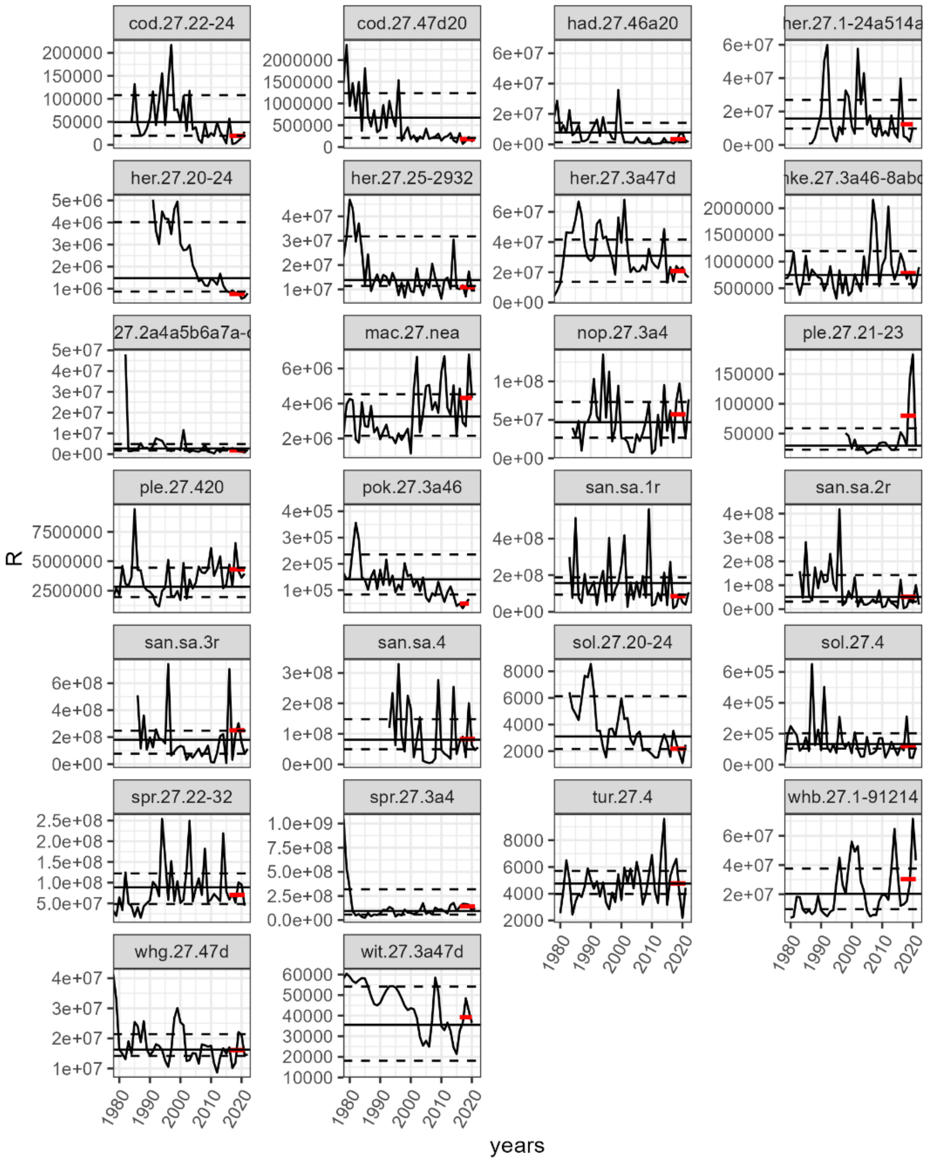

Figure 4.2.2: Recruitment time series of 26 fish stocks from the North Sea and Baltic Sea with assessment thresholds. Time series were smoothed with a six-year mean to obtain the 10th- and the 90th-percentile (lower and upper dashed lines) and the median (solid black line) of the recruitment time series in the reference period. The mean of the last six years defines the assessment period (2016-2021, red bar).

Another alternative to define a threshold value for recruitment could be based on the SSB-R relationship (Figure 4.2.3). Under this method, recruitment fails the threshold if it is below certain percentile of the recruitment observed when the SSB was above Blim. The likelihood that the indicator for recruitment fails the threshold when fished at FMSY is direct function of the percentile selected. At present there are no indications on a meaningful percentile to be selected. The advantage of calculating a threshold for R based on a SSB-R-relationship lies in the fact that it was consistent with reference values for SSB and thereby would be rooted in the framework for the assessment of D3C1 and D3C2.

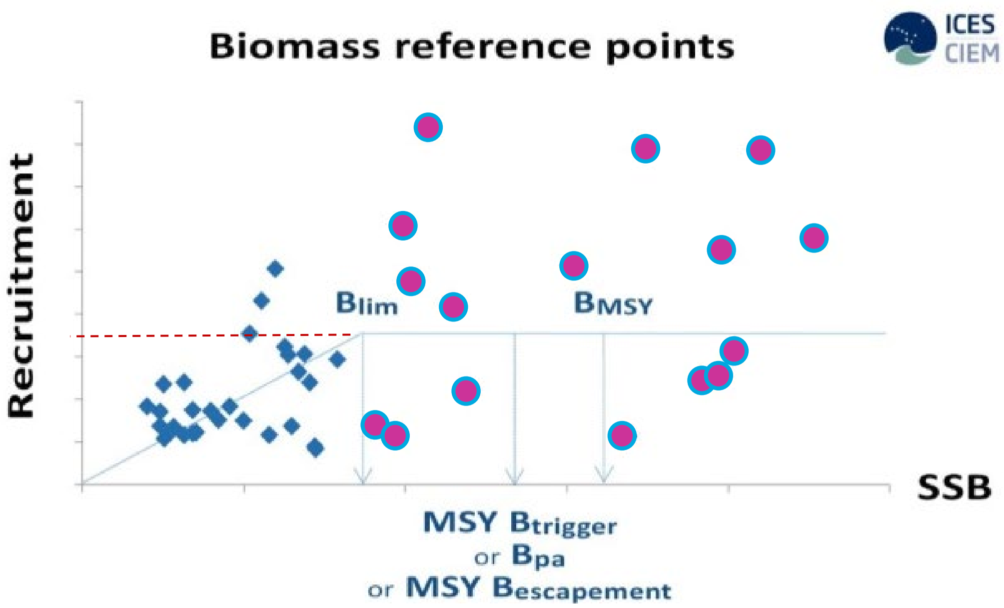

Figure 4.2.3. The classical spawner-recruit-relationship can be fitted with a hockey stick curve to obtain reference values for spawning stock biomass (SSB), e.g. MSY $B_{trigger}$, B_{pa} or MSY $B_{escapement}$. Vice a versa, the SSB reference values could be used to define thresholds for the assessment of recruitment (R). For example, the vertical line of the hockey-stick could be extended (red dashed line) or quantile-values such as the 10th-perctile of all R-values above a given SSB-reference point (e.g. B_{lim}, blue-pink dots) could be calculated. Modified from ICES Advice on fishing opportunities 2023 (https://doi.org/10.17895/ices.advice.22240624).

4.3 Conclusions

The suggested thresholds covered all approaches identified by WKD3C3SCOPE and used varying percentage (10th percentile, median/50th percentile) of the simulated or observed distributions of indicators to determine poor status of the indicator. They are further evaluated in section 5.

5 ToR 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.

5.1 Suitable indicators for D3C3 assessment for stocks without age based assessment data but length data available

For stocks without age based assessment data but length data available from surveys (this includes stocks assessed with production models and data limited approaches) the retained indicators were:

1. L90 of non-recruit lengths
2. Recruitment from survey catch at lengths below recruitment cut off
3. Where recruitment cut off cannot be estimated or recruits are not present in data, L90 of all lengths

No indicators reflecting growth and selectivity were evaluated for stocks without age based assessment data.

The temporal development of these indicators was compared for stocks with both age based assessment data and length based data available.

The length based indicators appeared sensitive to changes in survey design that may impact catchability of juveniles. In general, there was poor correlation between the indicators derived based on length distributions and indicators derived based on age based data. This led to concern that the length based indicators may be prone to having a substantially lower signal to noise ratio than age based indicators, in particular where the length distribution is based on low number of sampled individuals. L90 of all lengths (No. 3 above) reaches high precision and accuracy (within one centimetre) at sample sizes of 200-300 individuals (Östman et al. 2023), and these sample sizes can be challenging to achieve in survey-based data.

5.2 Suitable indicators for D3C3 assessment for stocks with age based assessment data

5.2.1 Criteria to select among the identified D3C3 indicators for further testing and setting of thresholds from WKD3C3SCOPE

WKD3C3SCOPE produced a list of evaluation criteria for use in WKD3C3THRESHOLD. These criteria are listed in Table 5.2.1. Issues 1-6, 8-9 and 12 were considered fulfilled for the indicators tested here, and focus was on:

7. Threshold clearly linked to poorer health

10. Responsive to management decisions - describe which and whether reactive (e.g. climate impact where we need to adapt other criteria targets) or regulatory (e.g. length distribution which can be changed through adapting F)

11. Threshold estimable and CV of estimated threshold acceptable

13. Indicator correlation & ambiguity – investigated through correlation between indicators F, SSB, R and age structure indicators.

Table 5.2.1. Criteria agreed by WKD3C3SCOPE for WKD3C3THRESHOLD

Topic	Issue
Availability of underlying data (Measurable)	1.Data accessible and easy to use for many stocks
	2. Relevant spatial coverage
	3. Relevant temporal coverage
Quality of underlying data (Sensitivity) (Responsive)	4. Indicators should be technically rigorous (well described) and peer reviewed (tangible)
	5. Reflects changes in health of stocks/key process (identify which)
	6. Magnitude, direction and variance of indicator estimable
Conceptual (Theoretical Basis)	7. Threshold clearly linked to poorer health
	8. Unambiguous
Communication (Con'crete) (PubleAware)	9. Comprehensible – was considered fulfilled for most if not all
Management	10. Responsive to management decisions - describe which and whether reactive (e.g. climate impact where we need to adapt other criteria targets) or regulatory (e.g. length distribution which can be changed through adapting F)
	11. Threshold estimable and CV of estimated threshold acceptable
	12. Cost-effectiveness – not considered relevant for indicators where data are already collected but possibly for new data
Indicator suites (Redundancy)--post criteria evaluation	13. Indicator correlation & ambiguity – information from indicators that are correlated in time can be useful if a change in correlation will tell you something useful
Type of stock	Fish or shellfish, short-lived or long-lived, sensitive or other
	data needed (age based, length based, catch only)

5.3 Responsive to management decisions and unambiguity

The indicators for recruitment and growth were considered to be appropriate for signalling a need to adapt D3C1 and D3C2 thresholds whereas the indicators of age structure and selection pattern were considered regulatory as they could be at least partly changed by adapting technical rules or F. Below, the responsiveness and unambiguity of the age structure indicators are evaluated.

The age/size-structure of fish stocks is expected to impact to stock recruitment (or recruitment success) and be impacted by F. In the case of recruitment, a positive effect should be visible in the recruitment year. In the case of F, the relationship should be with the fishing pressure induced in the previous or earlier years. Two simple analyses (test1-2) were performed at the workshop to evaluate the impact of time delays on the correlation of the D3C3 indicators ABImsy, ASA, POS, SSB/R (Table 3.2.1) with recruitment (R) and fishing mortality (F). The analyses were based on data from 72 ICES stocks with analytical stock assessment described in section 3.2.

Test1: D3C3 indicators aim to capture changes which consolidate in the age/size structure of stocks and strong negative correlation with recruitment was considered an undesirable feature. For this purpose, correlation of the indicators with R at lag 0 and -1 was used to evaluate the instantaneous and short-term sensitivity of the indicators to recruitment pulse. The lag -1 was included as the recruits enter the ABI_{MSY} estimation the year after recruitment, whereas recruits enter the estimation of ASA and POS only once the fish mature.

The median correlation of all indicators are negative with recruitment. The spread of correlations with R at all the lags evaluated were comparable among indicators. At both lag 0 and -1, the correlation was closest to zero for ABImsy and POS closely followed by ASA. Median correlation was markedly more negative for SSB/R (table 5.3.1, figure 5.3.1).

Table 5.3.1. Median correlations at lag -1, 0, +1 of 72 ICES stocks with analytical assessment calculated between R and the D3C3 indicators ABImsy, ASA, POS, SSB/R.

Lag	ABImsy	ASA	POS	SSB/R
-1	-0.043	-0.054	-0.033	-0.099
0	-0.026	-0.051	-0.041	-0.274
1	-0.132	-0.231	-0.08	-0.313

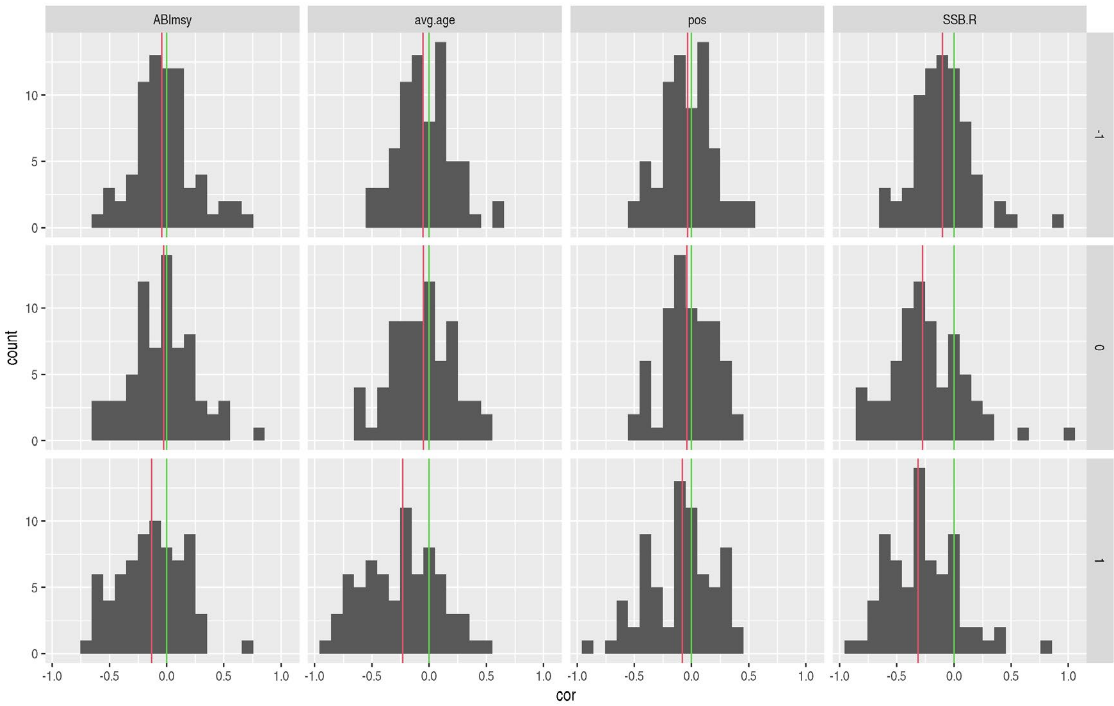

Figure 5.3.1. Histogram of correlations at lag -1, 0, +1 for 72 ICES stocks with analytical assessment calculated between R and the D3C3 indicators ABImsy, ASA (termed avg_age in the figure), POS, SSB/R. The red vertical line is the median correlation, the green vertical line is centred on zero as reference (no correlation).

Contrary to expectations, none of the indicators showed a positive median correlation with R (Fig. 5.3.1). Several reasons could explain this lack of response, e.g. mechanistic link although theoretically sound is not proved for many species and it is not supported by the data, the indicator formulations are not suitable to detect the expected response, drivers other than stock structure have stronger influence on recruitment and a simple univariate correlation is not able to disentangle likely interacting drivers.

Test 2: Age structure indicators theoretically respond to fishing exploitation. However, the effects of changes in fishing pressure are unlikely to affect the age/size structure of fish stocks instantaneously. In populations with complex demography and many age classes, changes in the intensity of fishing may take years to integrate in the stock and be detected by an indicator. For this reason, the correlation of the indicators with F (in most cases F_{bar}) from the stock assessment was evaluated over a broad range of lag from 0 to -10 years. At a stock level, we expect that long-living species would have higher correlation at longer time lags than short living species. Independently from the time lag, a strong (negative) correlation with F is considered a desirable feature of the indicators.

All the indicators show a negative median correlation with fishing mortality (Figure 5.3.2). The correlation is stronger at lag 0 for ABImsy and ASA compared to POS and SSB/R. The correlation decreases progressively at longer time lags but more rapidly for POS and SSB/R compared to ASA and especially ABImsy where it is still visible until lag -8 (Figure 5.3.2, although it has not been statistically tested).

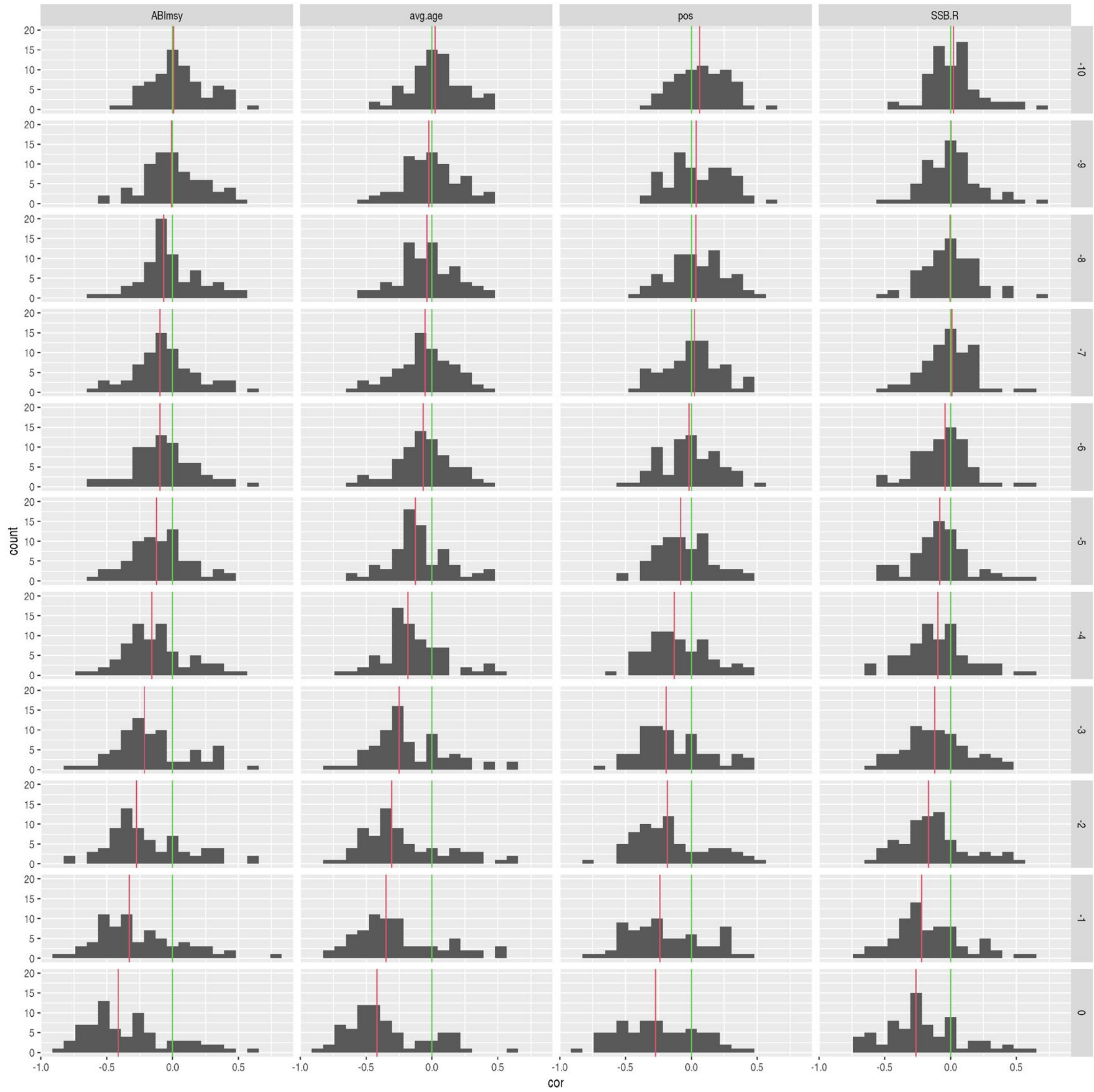

Figure 5.3.2. Histogram of correlations at lag from -10 to 0 for 72 ICES stocks with analytical assessment calculated between R and the D3C3 indicators ABImsy, ASA, POS, SSB/R. The red vertical line is the median correlation, the green vertical line is centred on zero as reference (no correlation).

5.4 Cross stock comparisons

The relationship of the different indicators was evaluated for all stocks for longevity groups as well as for groups differing in historical fishing pressure. The historic fishing pressure was investigated by estimating the median F over the entire timeseries and classified as

Medium F: median F/FMSY less than 1

High F: 1<median F/FMSY<2

Very high F: 2<median F/FMSY

The correlations observed in each of these groups are shown in ‘violin’ plots in Fig. 5.4.1 and 5.4.2. Values on the y-axis show correlations and where the ‘violin’ is wide at a specific value, this value is frequently observed whereas a narrow band on the ‘violin’ shows a value rarely observed. The correlation between SSB/R and recruitment was large for all longevity groups and levels of historical fishing pressure. Among the other three indicators, the distribution of correlations was quite similar, and they all exhibited correlations closer to zero for short lived species for F, R and SSB whereas correlations with SSB were higher in long lived species than in medium species (Figure 5.4.1).

Stocks fished historically at medium or high fishing mortality showed a positive correlation between the indicators and SSB (Figure 5.4.2). However, this correlation became close to zero for stocks with very high fishing mortality. This could indicate that the indicator became less accurate at high fishing pressure or that the two are decopuled at very high fishing pressure.

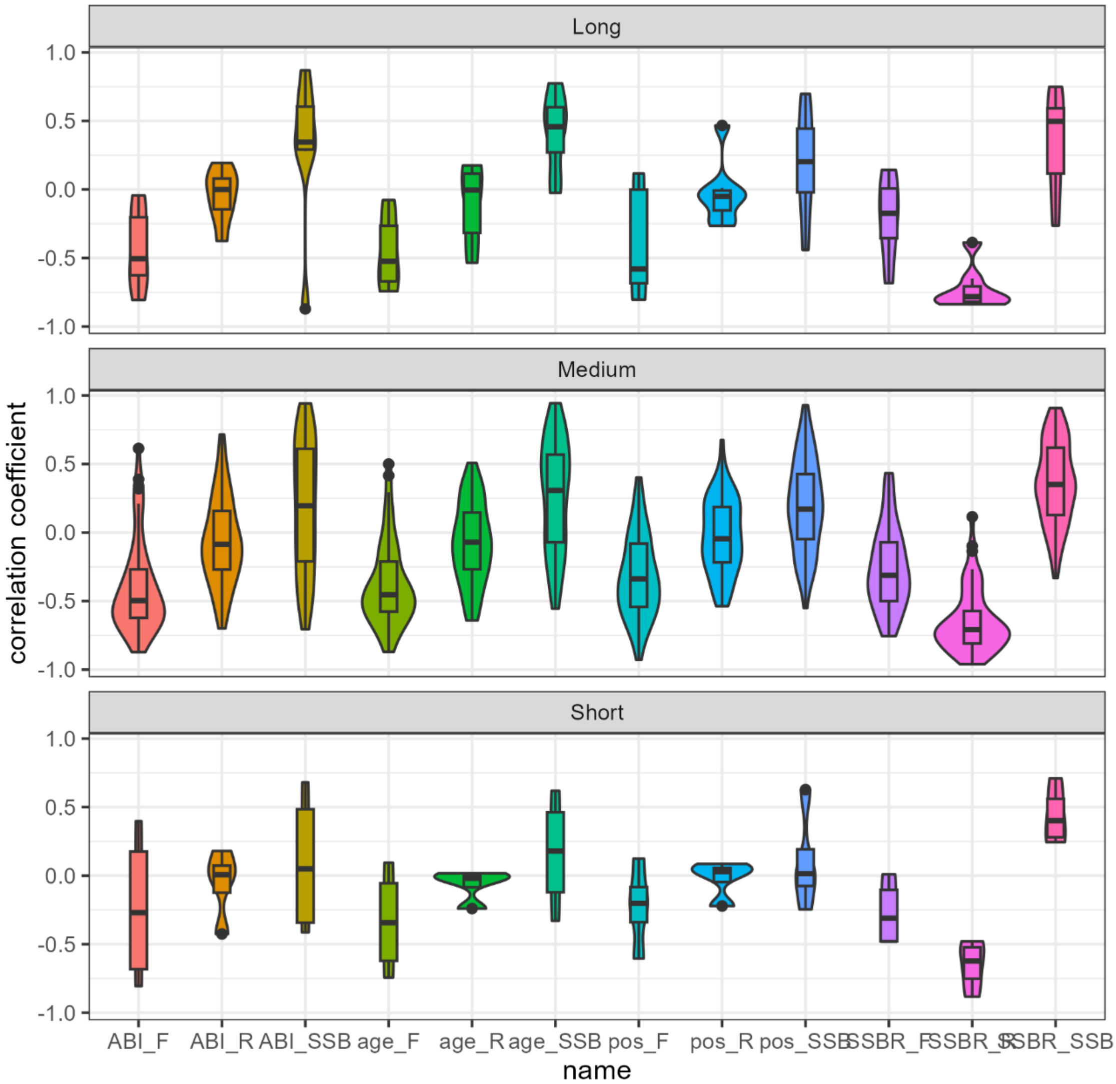

Figure 5.4.1. Difference in the correlation between the indicators ABI, ASA (termed age), POS and SSB/R (termed SSBR). The name of the indicator is given before the _ and the term after the _ denotes the factor to which the indicator was correlated. Grouped by longevity groups.

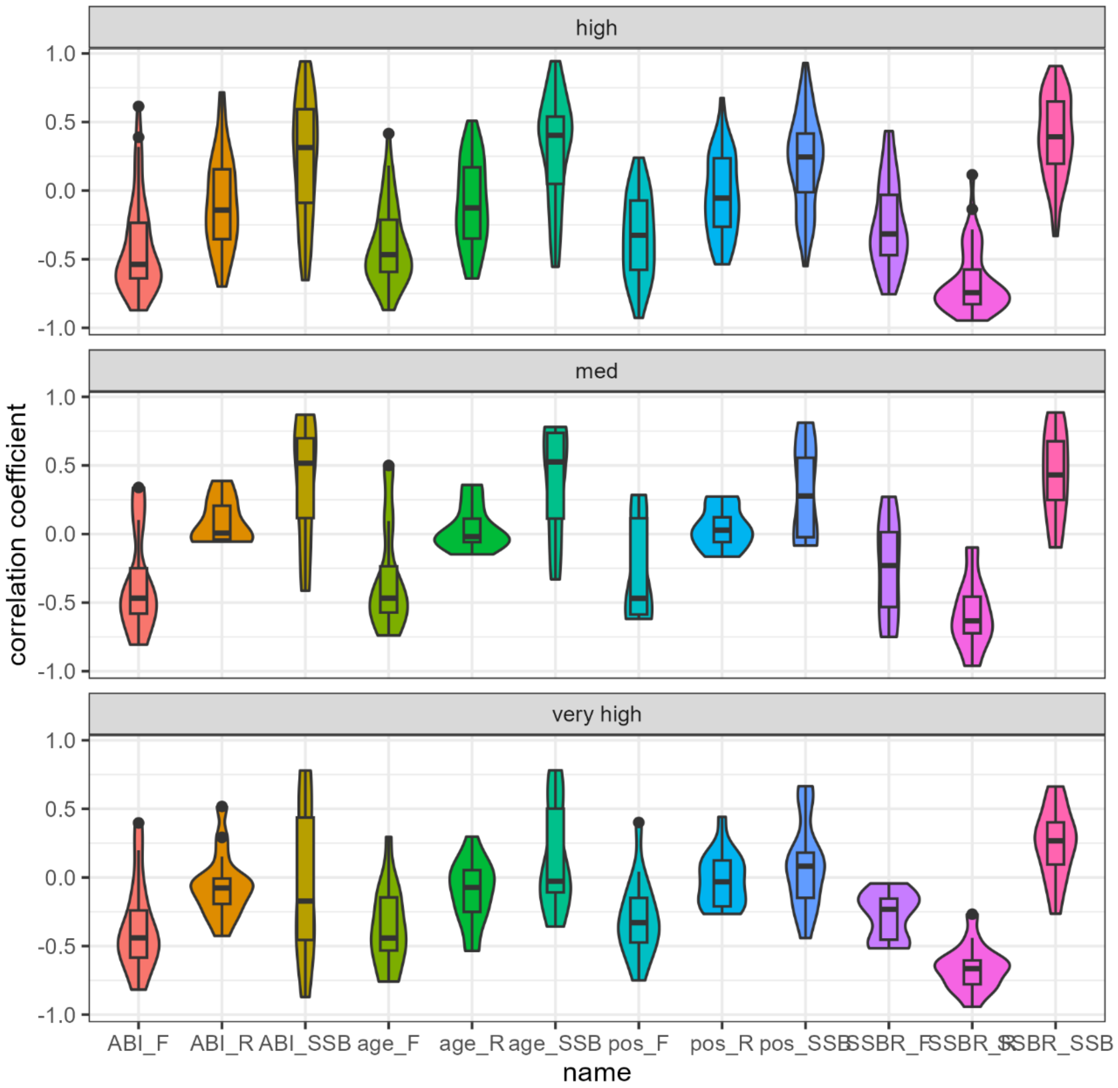

Figure 5.4.2. Difference in the correlation between the indicators ABI, ASA (termed age), POS and SSB/R (termed SSBR). The name of the indicator is given before the _ and the term after the _ denotes the factor to which the indicator was correlated. Grouped by historical median fishing pressure.

The MSFD reporting for D3C1 and D3C2 is based on 6-year averages (assessment period averages) and to determine if 6-year averages exhibited a different correlation than annual values, the correlations are compared in figures 5.4.3 and 5.4.4. Generally, annual values provided correlations closer to zero than 6-year averages, indicating that annual values had a lower signal to noise ratio than 6-year averages.

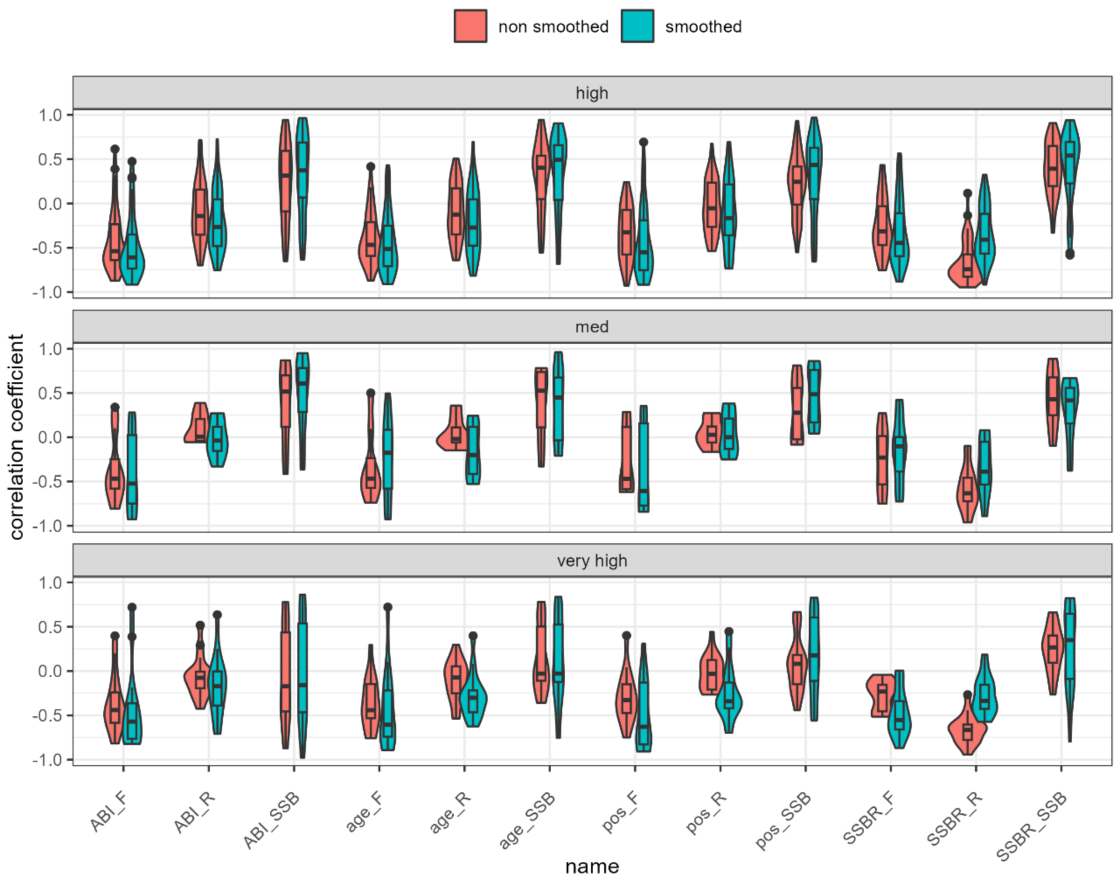

Figure 5.4.3. Comparison of correlations estimated based on annual (non-smoothed) and 6-year averages (smoothed) values of the indicators ABI, ASA (termed age), POS and SSB/R (termed SSBR). The name of the indicator is given before the _ and the term after the _ denotes the factor to which the indicator was correlated. Grouped by longevity.

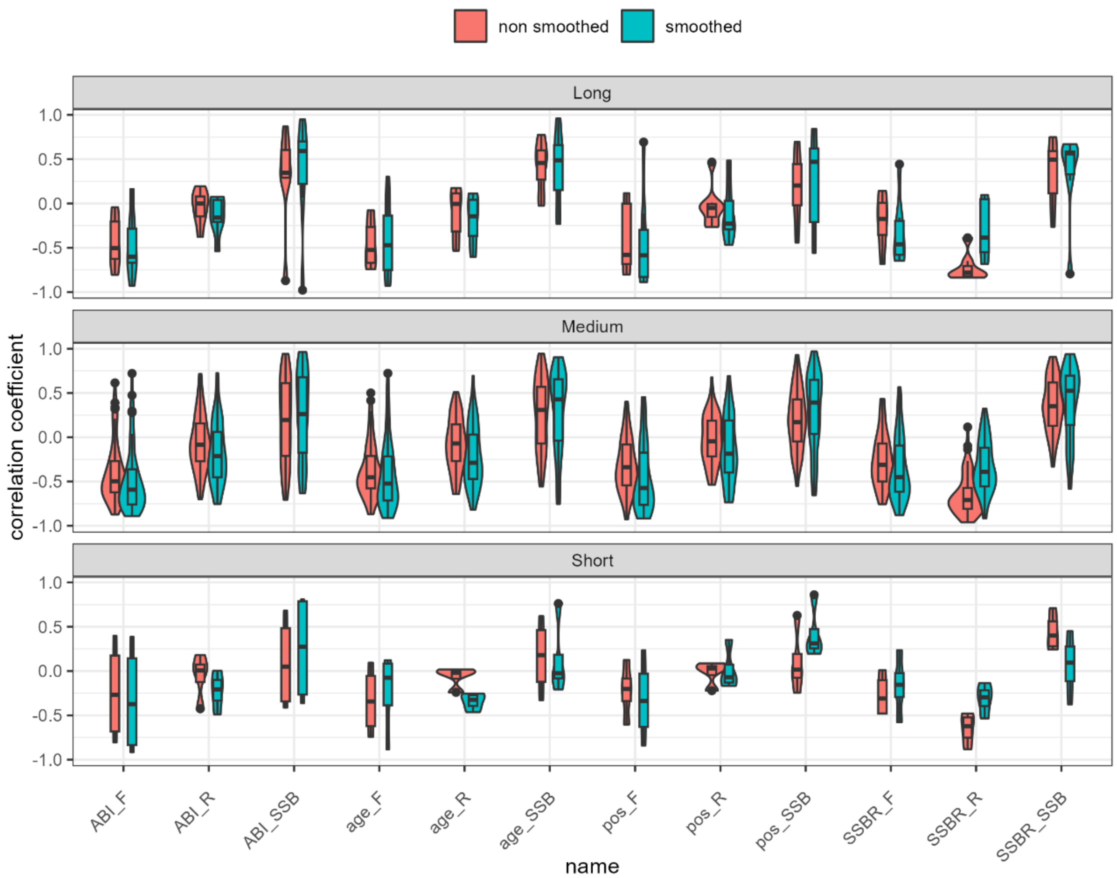

Figure 5.4.4. Comparison of correlations estimated based on annual (non-smoothed) and 6-year averages (smoothed) values of the indicators ABI, ASA (termed age), POS and SSB/R (termed SSBR). The name of the indicator is given before the _ and the term after the _ denotes the factor to which the indicator was correlated. Grouped by historical median fishing pressure.

5.5 Methods to define thresholds for D3C3 assessment for stocks with different data availability

5.5.1 Consequences of failing to achieve different suggested thresholds

Recruitment and growth directly impact fisheries yield and precautionary fishing mortality limits. However, there was not a clear threshold where this impact becomes so strong action must be taken and that can be identified without simulations in an MSE framework. As a result, the consequences of failing different candidate thresholds will have to be investigated in WKD3C3SIMUL.

None of the age structure indicators demonstrated a positive link with stock productivity, and hence threshold levels could not be based on levels at which stock productivity is impaired or enhanced. Rather, it could be concluded that a substantial positive impact of a specific age structure on productivity cannot be expected.

5.5.2 Evaluate consistency and complementarity with D3C1(FMSY) and D3C2 (MSYBtrigger)

The threshold approaches were each evaluated for their pros and cons, including their consistency with the MSY approach outlined under D3C1 and D3C2.

Pros	Cons	consistency and complementarity with D3C1(F_{MSY}) and D3C2 ($MSYB_{trigger}$) objectives for the stock
Level of the unfished stock		
Can possibly be used if there is agreement on accepted distance from unfished as in e.g. D6	Very difficult to estimate, simulation models and assumptions of biology of the unfished stock is required as such a stock has not observed in catches.	Inconsistent with objective to fish stocks according to MSY principles under D3.
Level at which the indicator aspect is not healthy following principles of D3C2.		
Easy to explain and in accordance with MSFD guidance wording. Based on historic data analysis and hence exhibits no sensitivity to simulation model formulation	Difficult to estimate in practice. There is no evidence in the data examined at the workshop that such a level exists, possibly because the effect is small.	Consistent with D3C2 approach to threshold setting.
Level of indicator of stocks considered to be in good health by expert judgement		
Expert knowledge is highly expert dependent and has low reproducibility	Difficult to ensure comparable methods across experts and member states even for the same stocks. Dependent on length of time series.	Can be consistent with D3C1 and D3C2 thresholds if using FMSY and MSYBtrigger to identify healthy periods.
Level of indicator of stocks considered to be in good health by quantitative approaches applied to historical data		
Consistent across researchers and member states and produces reproducible thresholds independent of the expert. Produces ok thresholds for stocks which have not been constantly overfished and for indicators not directly affected by fishing pressure (e.g. recruitment when SSB is greater than B_{lim}, weight at age, selectivity).	Breakpoints and percentiles are uncertain when based on few data. If there is a change in productivity or selectivity, the threshold value will depend on how this is addressed. For percentile approaches, it is necessary to agree on a percentile level. Indicators directly impacted by fishing (e.g. age structure indicators) will not get reliable thresholds for stocks that have been continuously overfished.	Consistent with D3C1 and D3C2 thresholds for stocks with reasonably long historical data, except for age structure indicators for continuously overfished stocks.
Level of the indicator when the stock is fished according to MSY principles		
Consistent with D3C1 threshold setting methods (population simulation models). Can account for natural variability if simulation model includes this (e.g. recruitment ag growth variability).	Only possible to estimate if F_{MSY} and B_{lim} are available. MSY level can be difficult to estimate if stock has been continuously and heavily overfished. In addition to stock productivity, the threshold will depend on the selectivity pattern. Processes not linked to F (selectivity pattern, growth, recruitment when the stock is above B_{lim}) will not provide thresholds that differ from those	Consistent with D3C1 and D3C2 thresholds.

Pros	Cons	consistency and complementarity with D3C1(F_{MSY}) and D3C2 ($MSYB_{trigger}$) objectives for the stock
	derived from using quantitative methods on historical data. If simulation model is not including natural variability, statements about the likelihood of false alarms and false green lights cannot be made.	

The listed pros and cons were used to define a decision tree to find the best threshold estimation approach (Fig. 5.5.1)

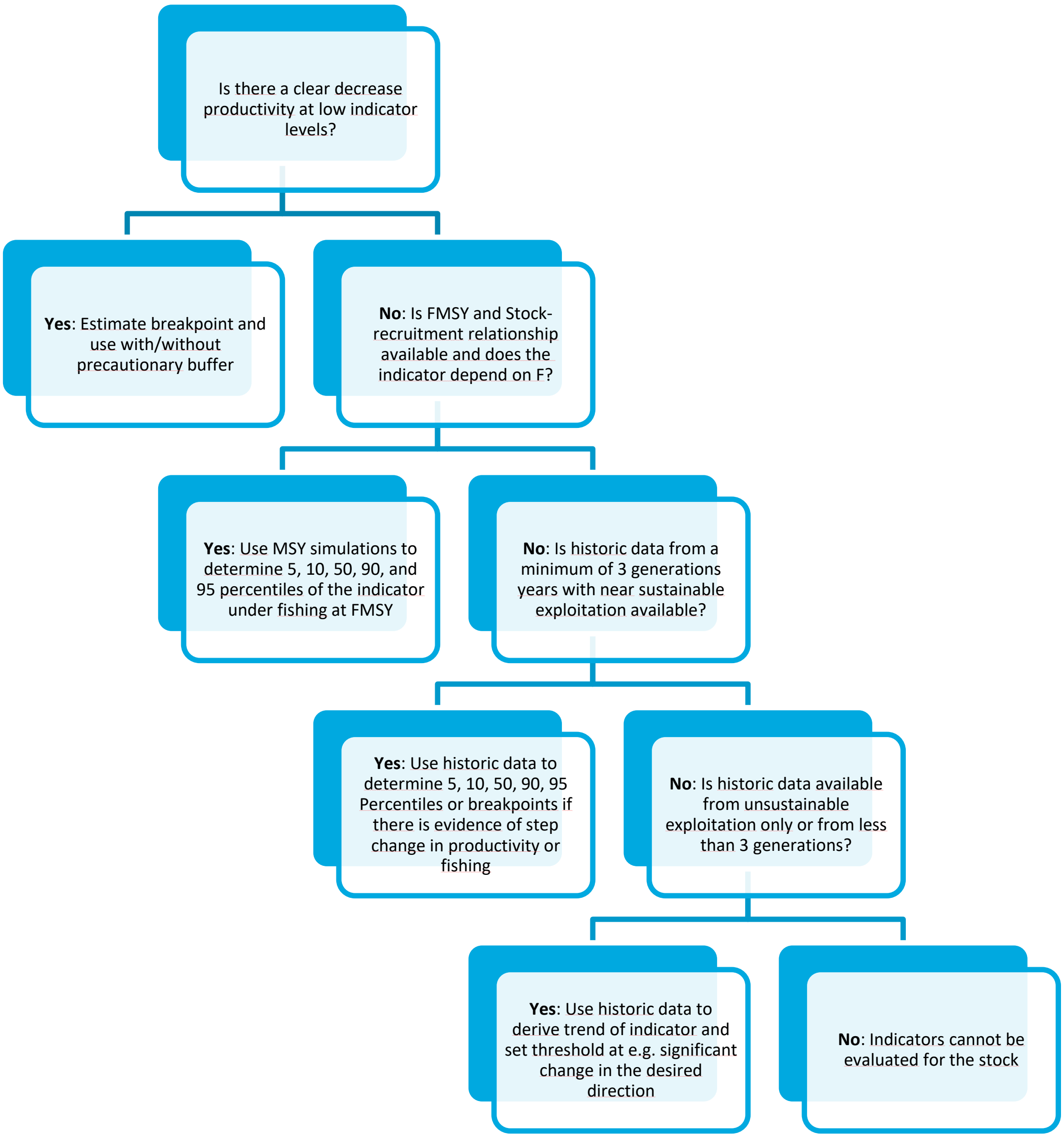

Figure 5.5.1. Decision tree to choose a threshold setting method.

5.5.3 Decisions on most appropriate percentages to use for threshold definition

As no link was clear between age structure indicators and productivity, historical analyses or simulations are going to be the most likely methods to derive thresholds. In both historical observations and simulations, the indicators are expected to vary between as a result of variation in e.g. productivity of the stock. This means that a decision is required on which proportion of these observed or simulated values are considered to be in poor health. The participants discussed using percentages such as 5, 10, 50, 90 and 95 but a decision on which was most appropriate was postponed until the results from WKD3C3SIMUL on the distribution of the indicators under fishing at F_{MSY} become available.

6 ToR d: Draft a framework for the comprehensive assessment of D3 stocks

6.1 Data requirements to assess D3

The participants 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.

6.2 Recommended indicator(s) for the assessment of D3C3 that are compatible with D3C1 and D3C2

The participants concluded that the indicator SSB/R was responsive to recruitment in an undesirable way. They considered that there was not sufficient evidence to determine whether one of the three remaining indicators provided a higher signal to noise ratio than the others and therefore proceeded with all three age structure indicators. The role of selectivity indicators under D3C3 was unclear in the guidance but these were also retained. For recruitment and growth, only one indicator was evaluated (though estimation differed depending on available data), and both were retained. This resulted in the following indicators being retained for medium lived stocks with age based assessment data:

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

For short lived and long lived species, the value of age-structure indicators (no. 1-3 in the list above) as management indicators was unclear. For short lived species, the group as a whole could not see a no strong link between the age-structure indicators and F or SSB, and 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). For long lived species, the group found that age structure indicators appeared to react later than F and SSB, and hence the added value for management of monitoring them was unclear.

Details on definition of thresholds will be further investigated in WKSIMULD3.

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.

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.

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.

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.

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.

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

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.

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.

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, 00, 1–20. https://doi.org/10.1111/faf.12789.Heino, M., Dieckmann, U., and Godø, O. R. 2002. Measuring probabilistic reaction norms for age and size at maturation. Evolution, 56: 669-678.

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.

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

ICES. 2016. Report of the Workshop on Guidance on Development of Operational Methods for the Evaluation of the MSFD Criterion D3.3 (WKIND3.3i), 14–17 March 2016, Copenhagen, Denmark. ICES CM 2016/ACOM:44. 99 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) EU request for a Technical Service on MSFD Article 8 guidance on undertaking assessments for Descriptor 3 (commercially exploited fish and shellfish) and Descriptor 4 (marine foodwebs). ICES Advice 2021 – sr.2021.14 – https://doi.org/10.17895/ices.advice.8817 1

ICES. (2022). Workshop on ICES reference points (WKREF1). ICES Scientific Reports, 4(2), 70. http://doi.org/10.17895/ices.pub.9822.

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(4), 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

Probst, W. N., Kempf, A., Taylor, M., Martinez, I., & 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(4), 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.

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

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

Scientific, Technical and Economic Committee for Fisheries (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.

Scientific, Technical and Economic Committee for Fisheries (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

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.

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., 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., 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., Mosqueira, I. & O'Neill, F.G. et al. (2020). Selectivity metrics for fisheries management and advice. Fish and Fisheries 21, 621-638.

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.

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
Phillip Boulcott	Marine Science Scot-land	Philip.Boulcott@gov.scot
Victoria Campon-Li-nares	CEFAS	victoria.linares@cefas.gov.uk
Gema Canal	IEO	gema.canal@ieo.csic.es
Helena Caserman	IZVRS	helena.caserman@izvrs.si
Ilaria Coscia	Marine Institute Ire-land	Ilaria.Coscia@marine.ie
Tomaso Fortibuoni	ISPRA	tomaso.foribuoni@isprambiente.it
Eva Garnacho	CEFAS	eva.garnacho@cefas.gov.uk
Beatriz Giujarro	IEO	beatriz.guijarro@ieo.csic.es
Patricia Gonçalves	IPMA	patricia@ipma.pt
Chris Griffiths	SLU	christopher.griffiths@slu.se
Nis Jacobsen	DTU	nsja@aqua.dtu.dk
Bernhard Kuehn	Thuenen	bernhard.kuehn@thuenen.de
Isabel Maneiro	IEO	isabel.maneiro@ieo.csic.es
Danai Mantopoulou Palouka	European Commis-sion	Danai.MANTOPOULOU-PALOUKA@ec.europa.eu
Ualerson Iran Peixoto da Silva	IMAR	ualerson.ip.silva@uac.pt
Nikolaus Probst	Thuenen Institute	nikolaus.probst@thuenen.de
Owen Rowe	HELCOM	Owen.Rowe@helcom.fi
Sonia Seixas	MARE	Sonia.seixas@uab.pt

Name	Institute	Email
Amina Tifoura	IEO	Amina.tifoura@ieo.csic.es
Paris Vasilakopolous	European Commission	paris.vasilakopoulos@ec.europa.eu
Ching Maria Villanueva	IFREMER	Ching.Villanueva@ifremer.fr
Håkan Wennhage	SLU	hakan.wennhage@slu.se
Nuria Zaragoza	IEO	nuria.zaragoza@ieo.csic.es

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

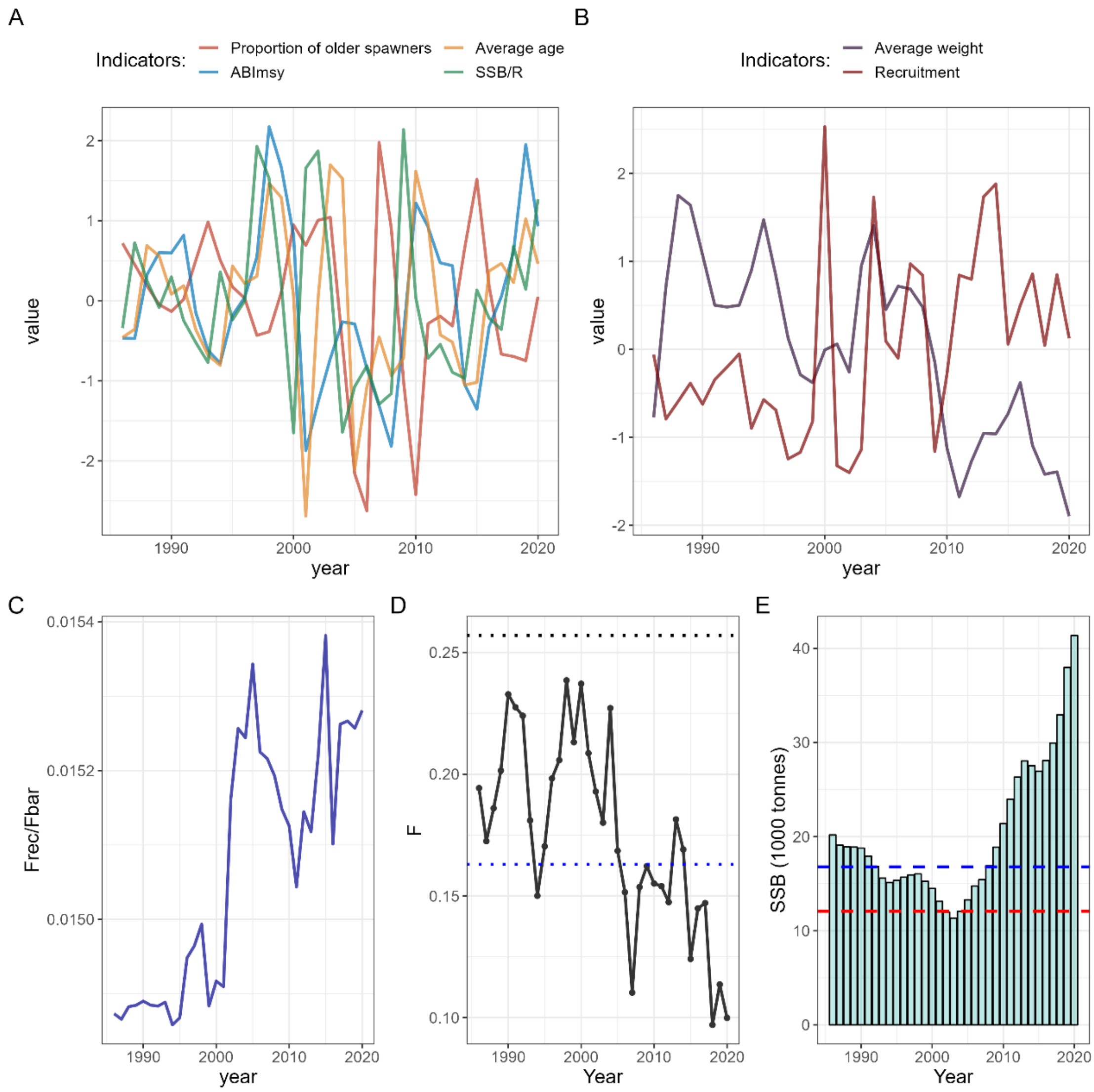

Figure A2.1. Estimated D3C3 indicators and stock status for Black-bellied anglerfish in Subarea 7 and Divisions 8.a-b and 8.d (ank.27.78.abd). Time series of the indicators are shown in panels A-C, as well as F (panel D) and SSB (panel E). Indicators in panels A and B have been normalised by subtracting 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).

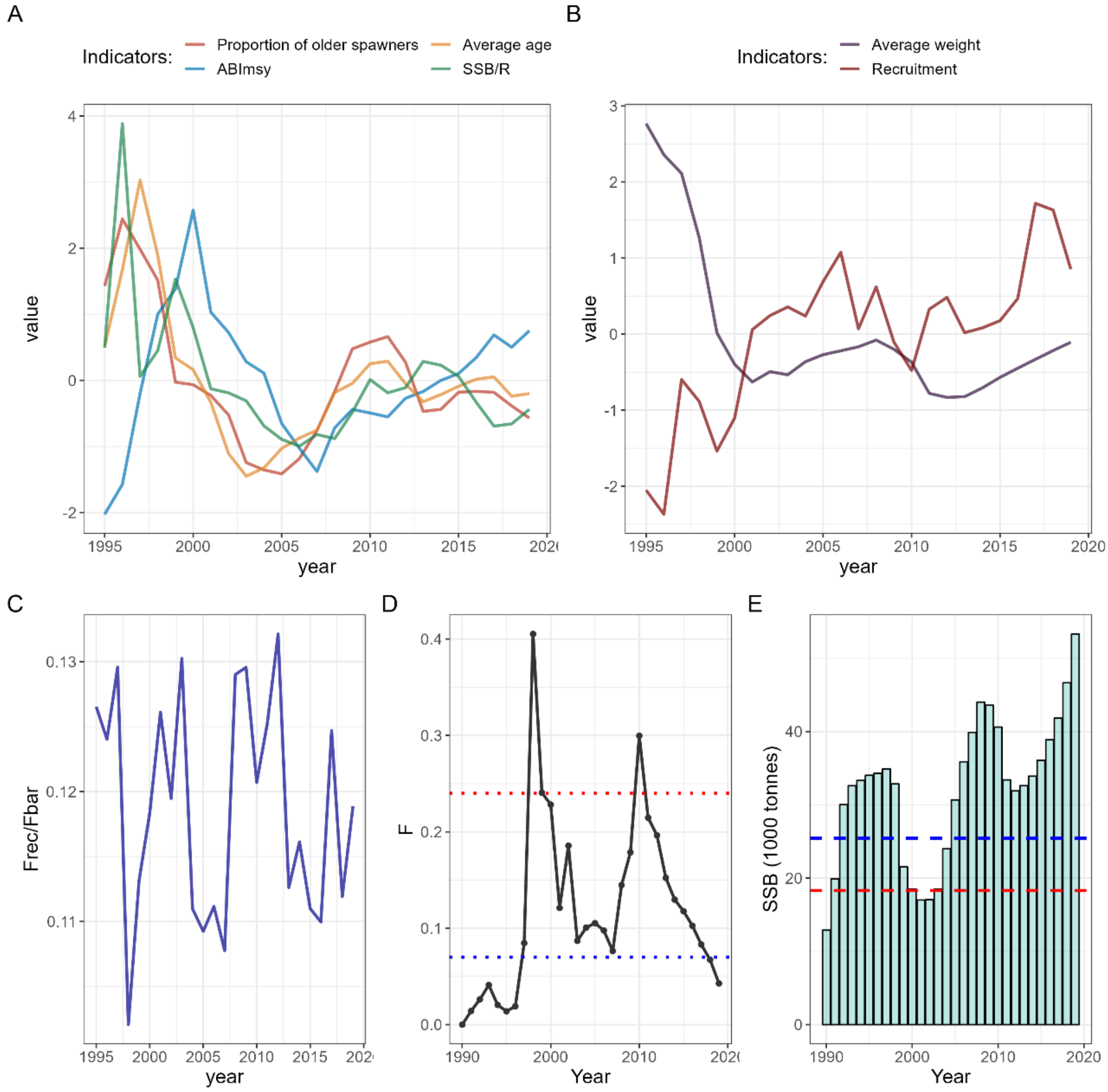

Figure A2.2. Estimated D3C3 indicators and stock status for Greater silver smelt (aru.27.5a14). See Figure legend A2.1 for further details.

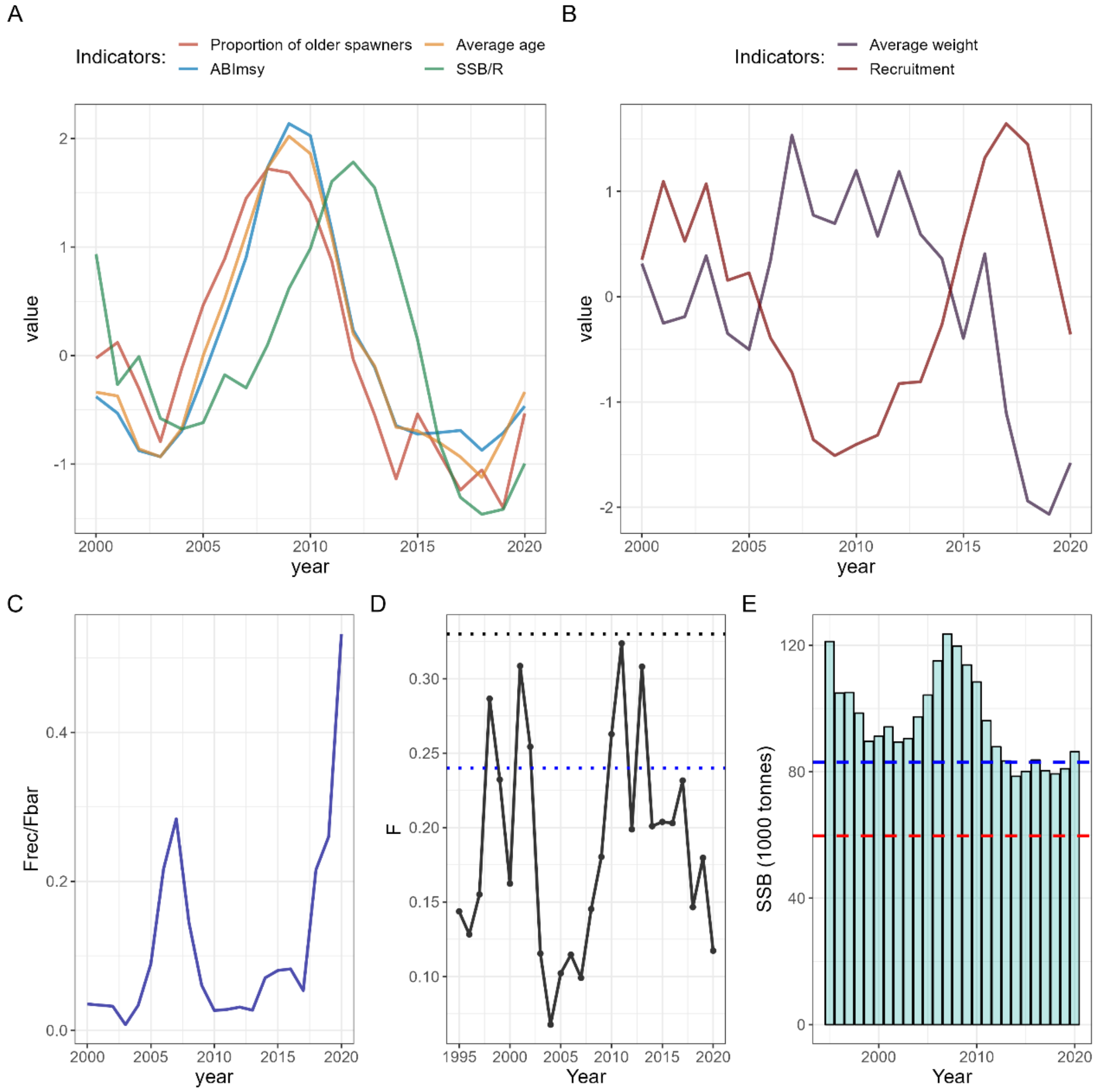

Figure A2.3. Estimated D3C3 indicators and stock status for Greater silver smelt (aru.27.5b6a). See Figure legend A2.1 for further details.

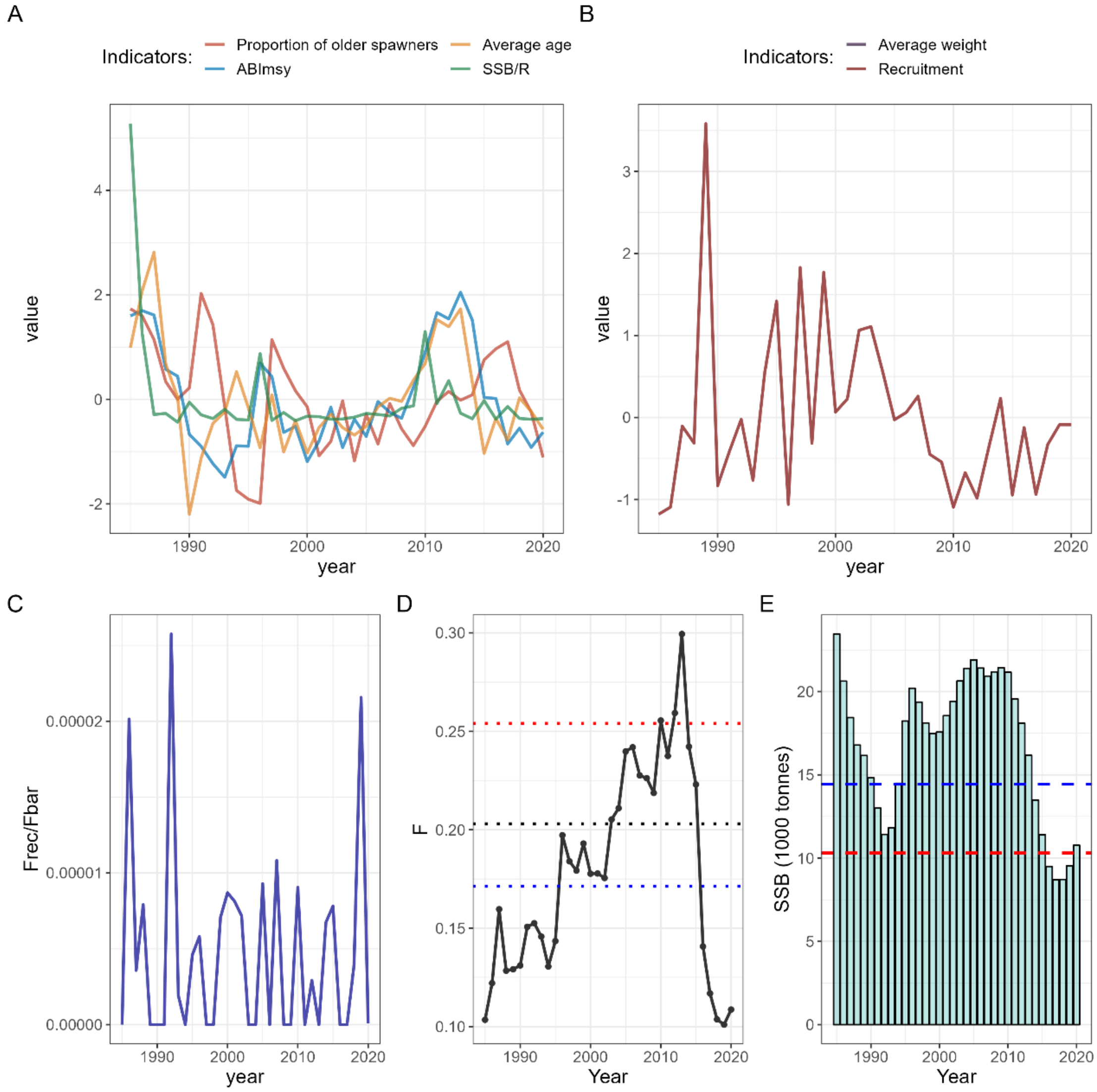

Figure A2.4. Estimated D3C3 indicators and stock status for Sea bass (bss.27.4bc7ad-h). See Figure legend A2.1 for further details.

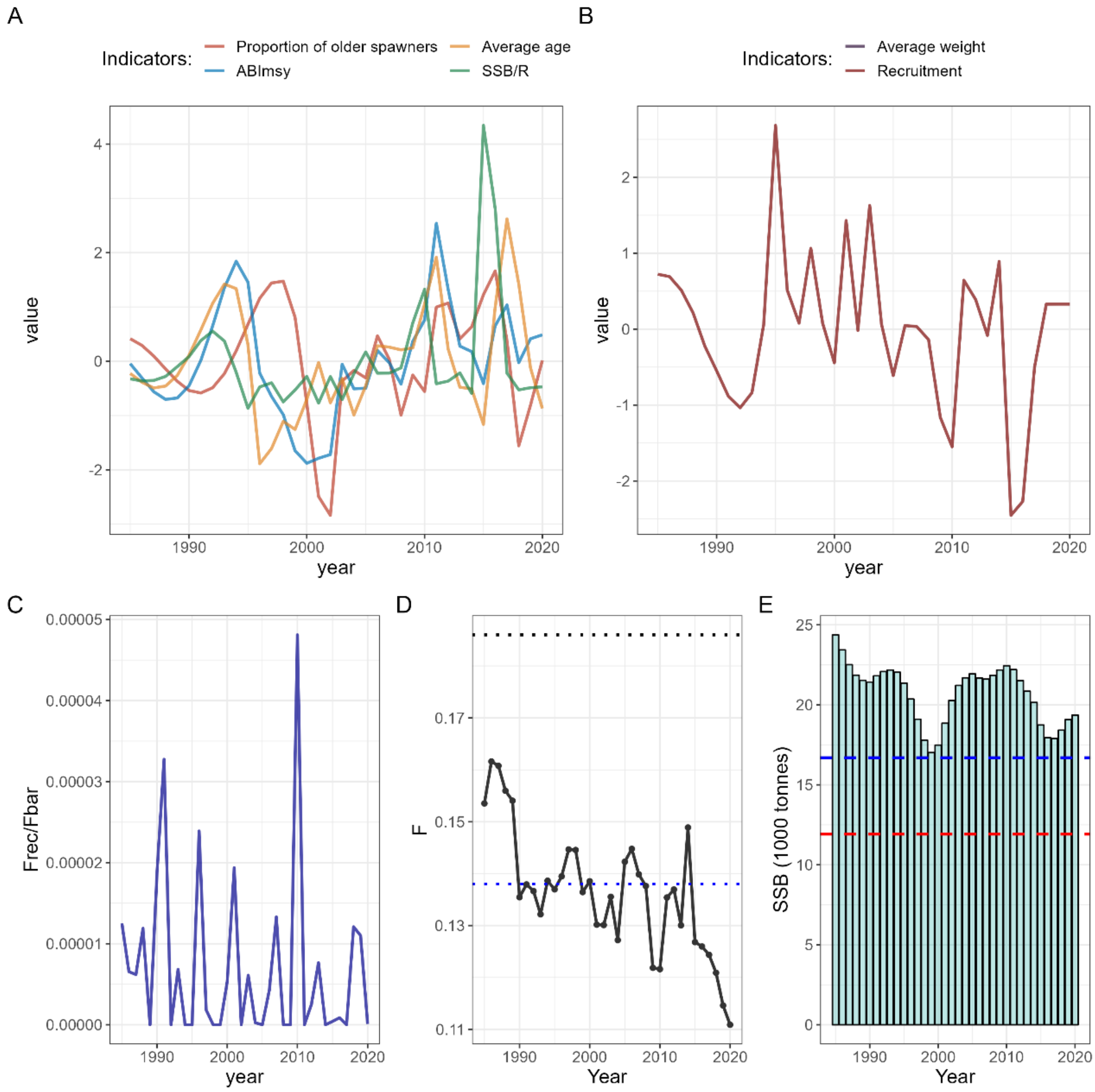

Figure A2.5. Estimated D3C3 indicators and stock status for Sea bass (bss.27.8ab). See Figure legend A2.1 for further details.

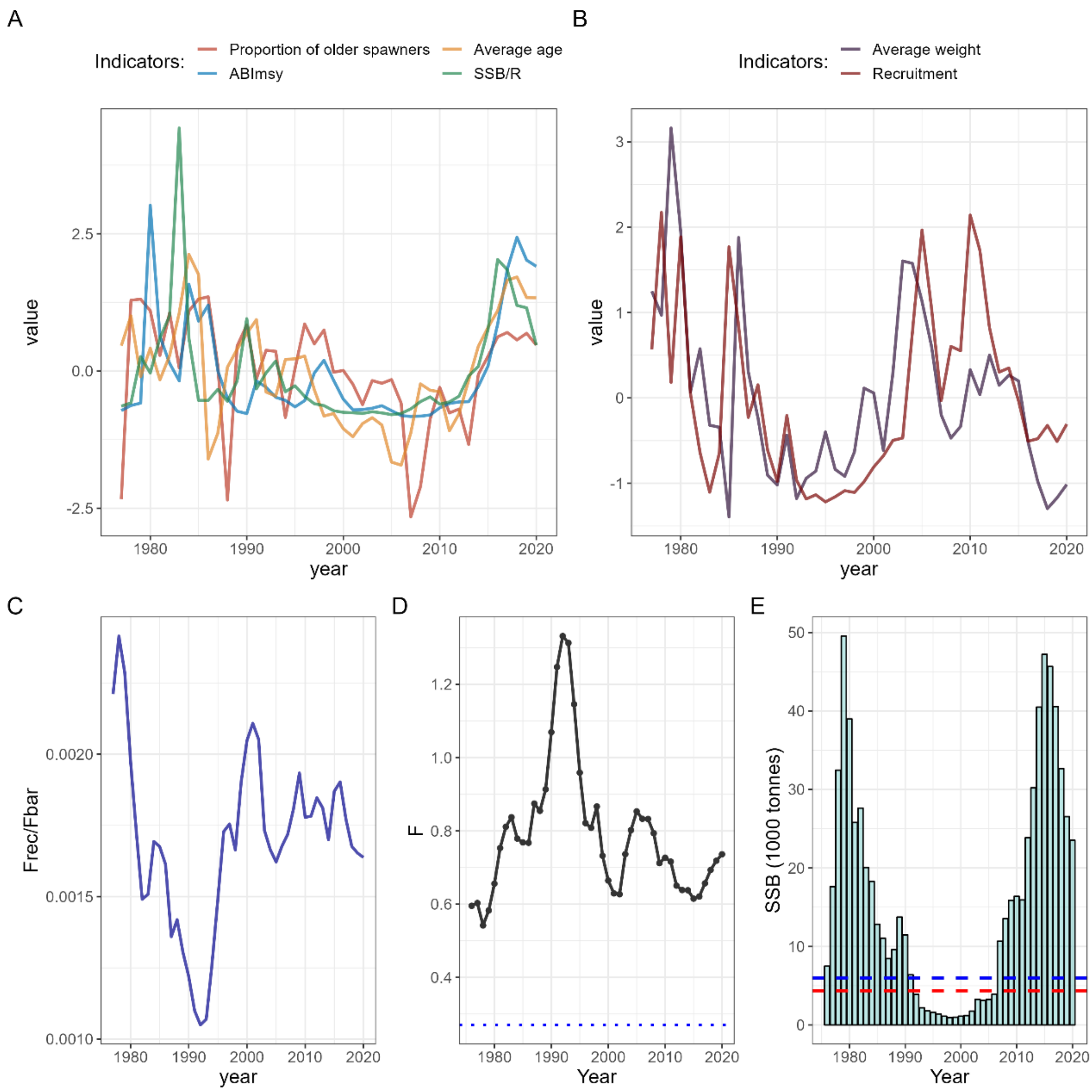

Figure A2.6. Estimated D3C3 indicators and stock status for Atlantic cod (cod.21.1). See Figure legend A2.1 for further details.

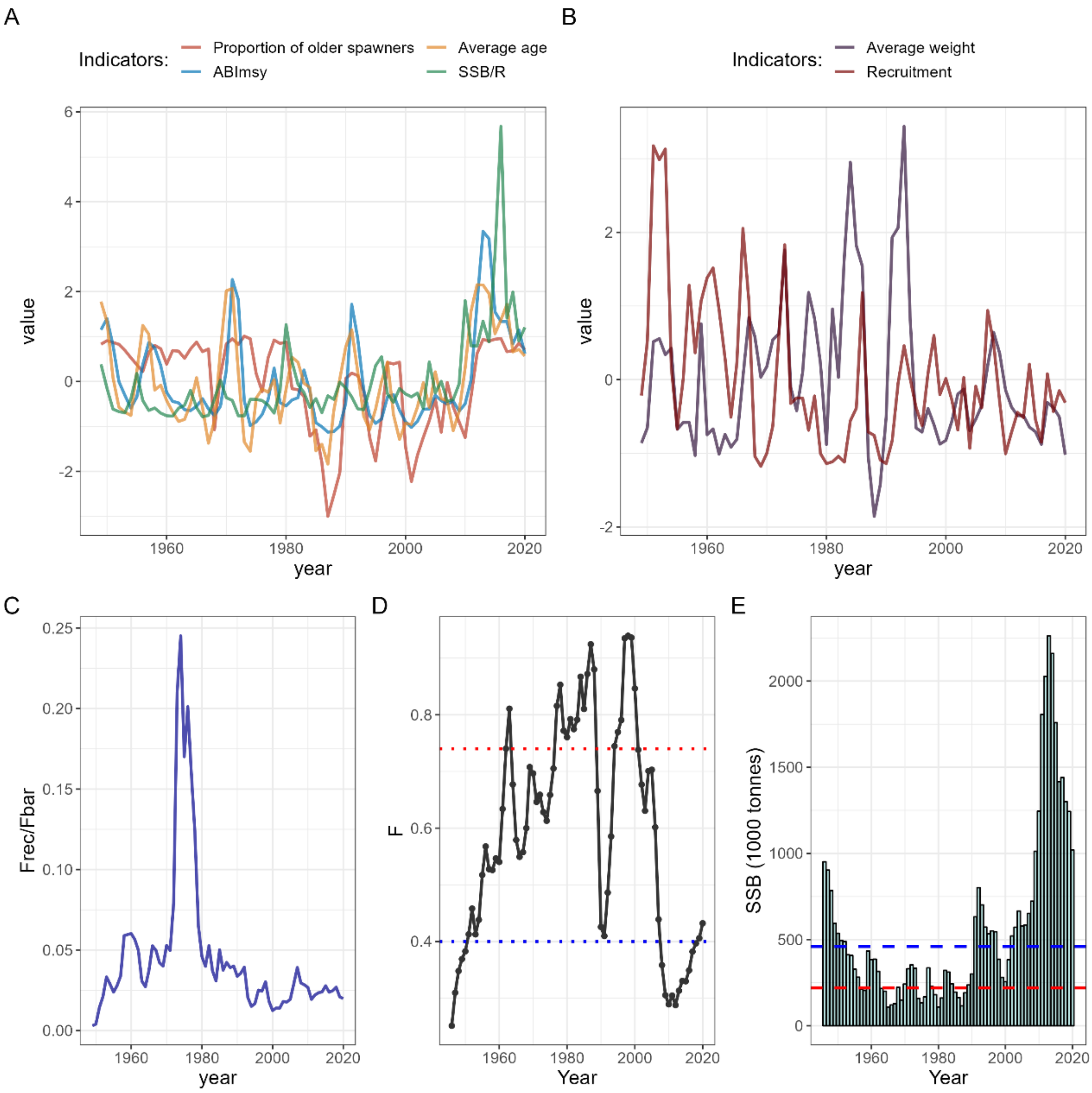

Figure A2.7. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.1-2). See Figure legend A2.1 for further details.

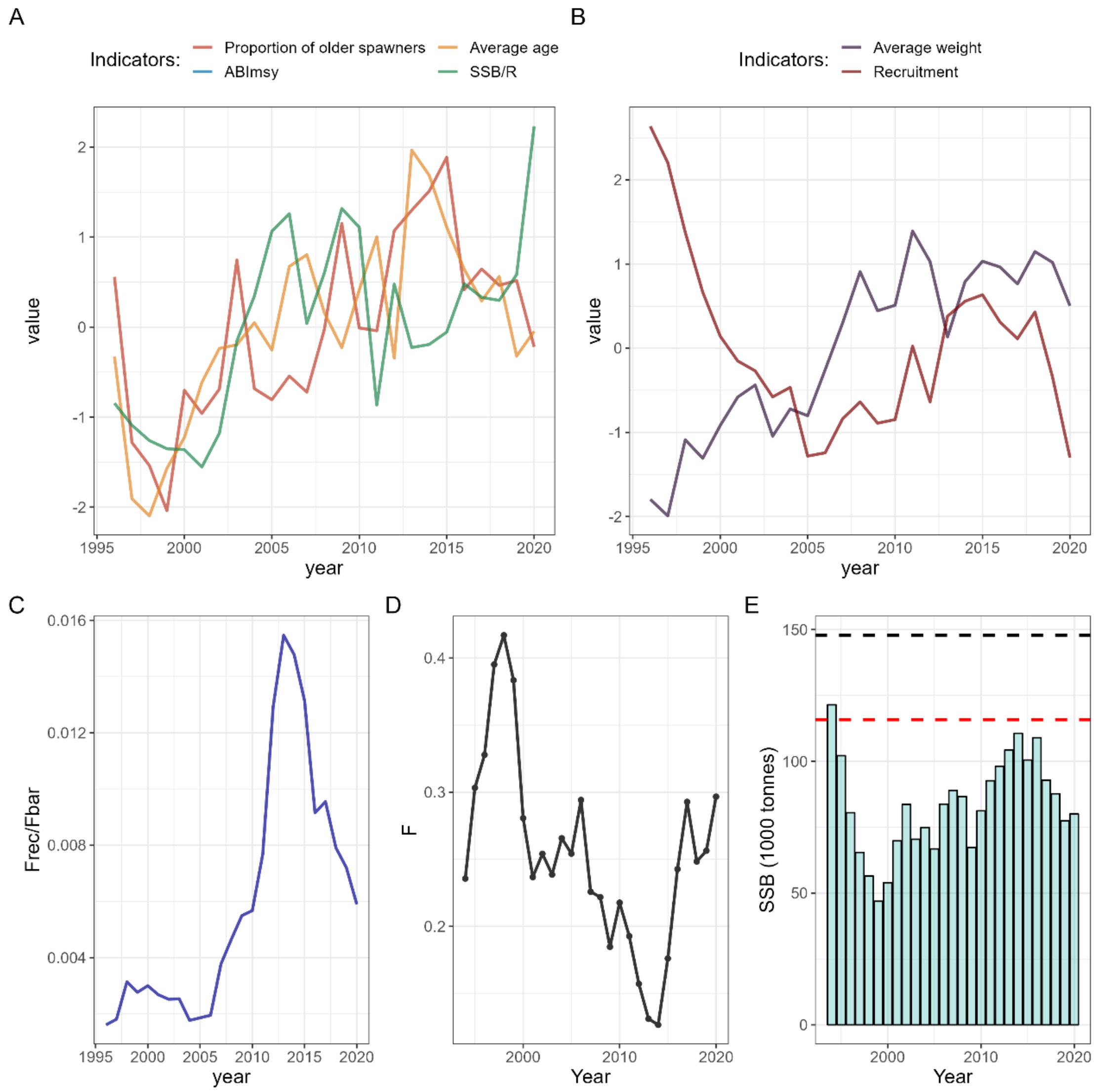

Figure A2.8. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.1-2coastN). See Figure legend A2.1 for further details.

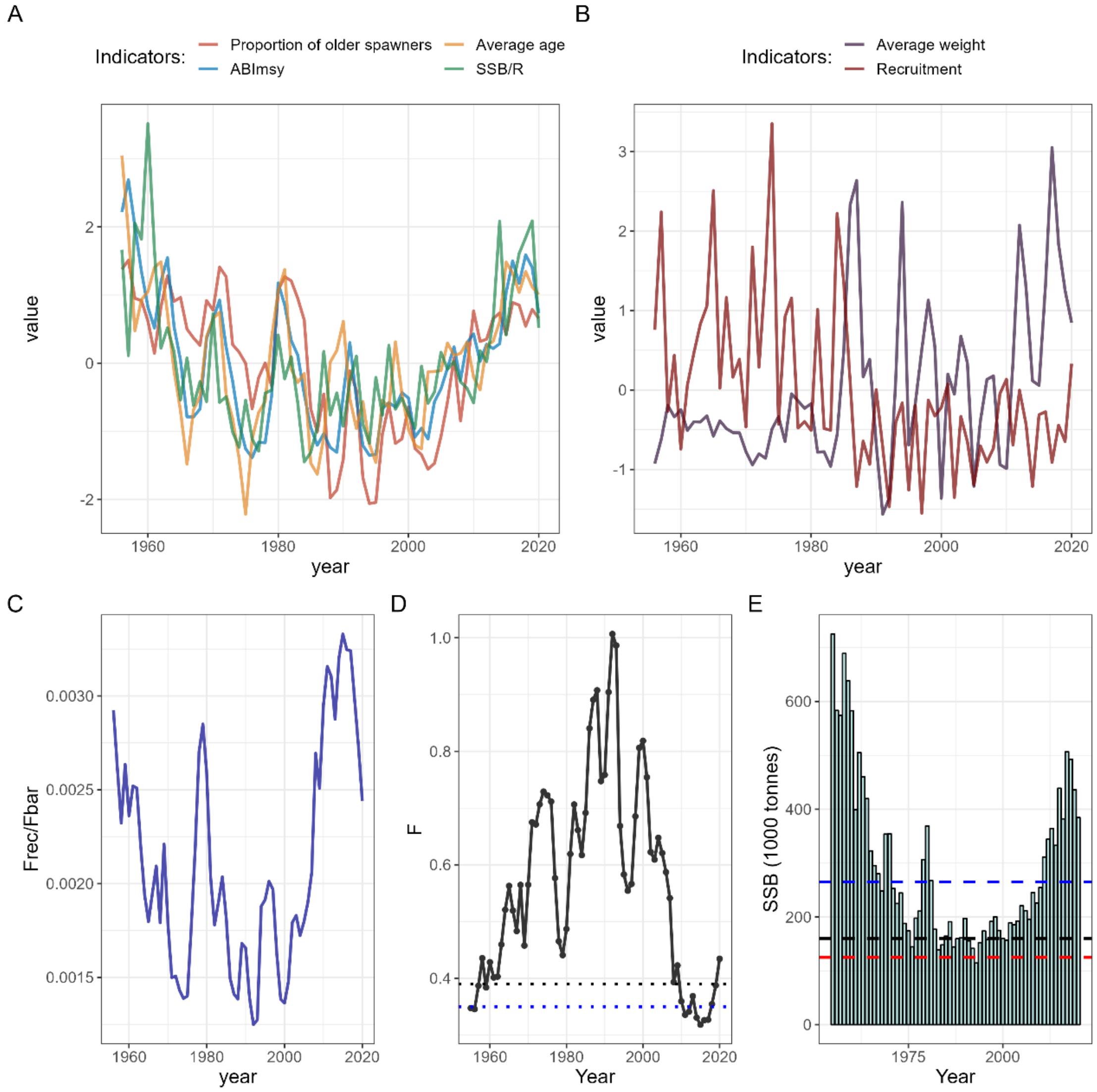

Figure A2.9. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.5a). See Figure legend A2.1 for further details.

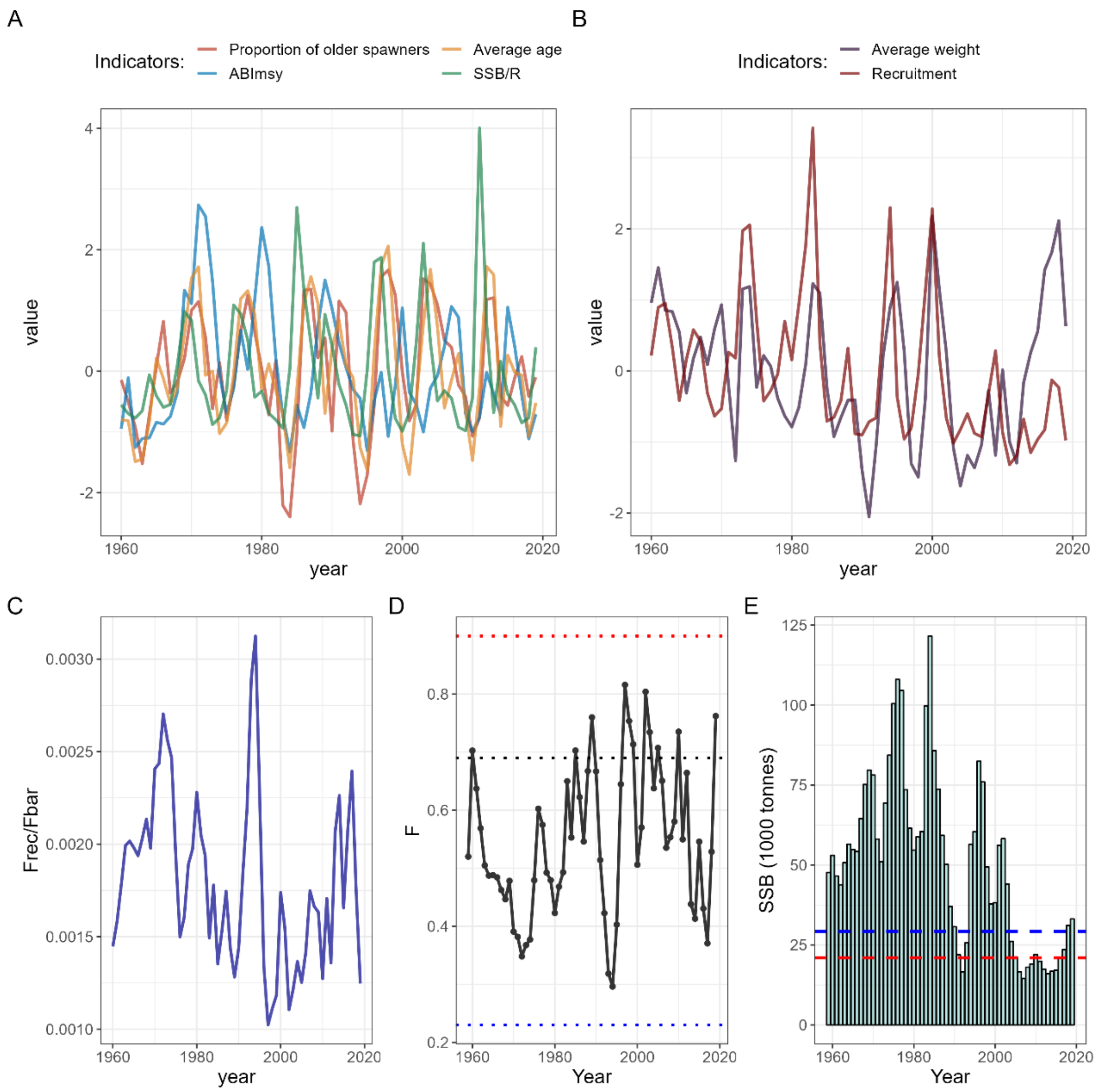

Figure A2.10. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.5b1). See Figure legend A2.1 for further details.

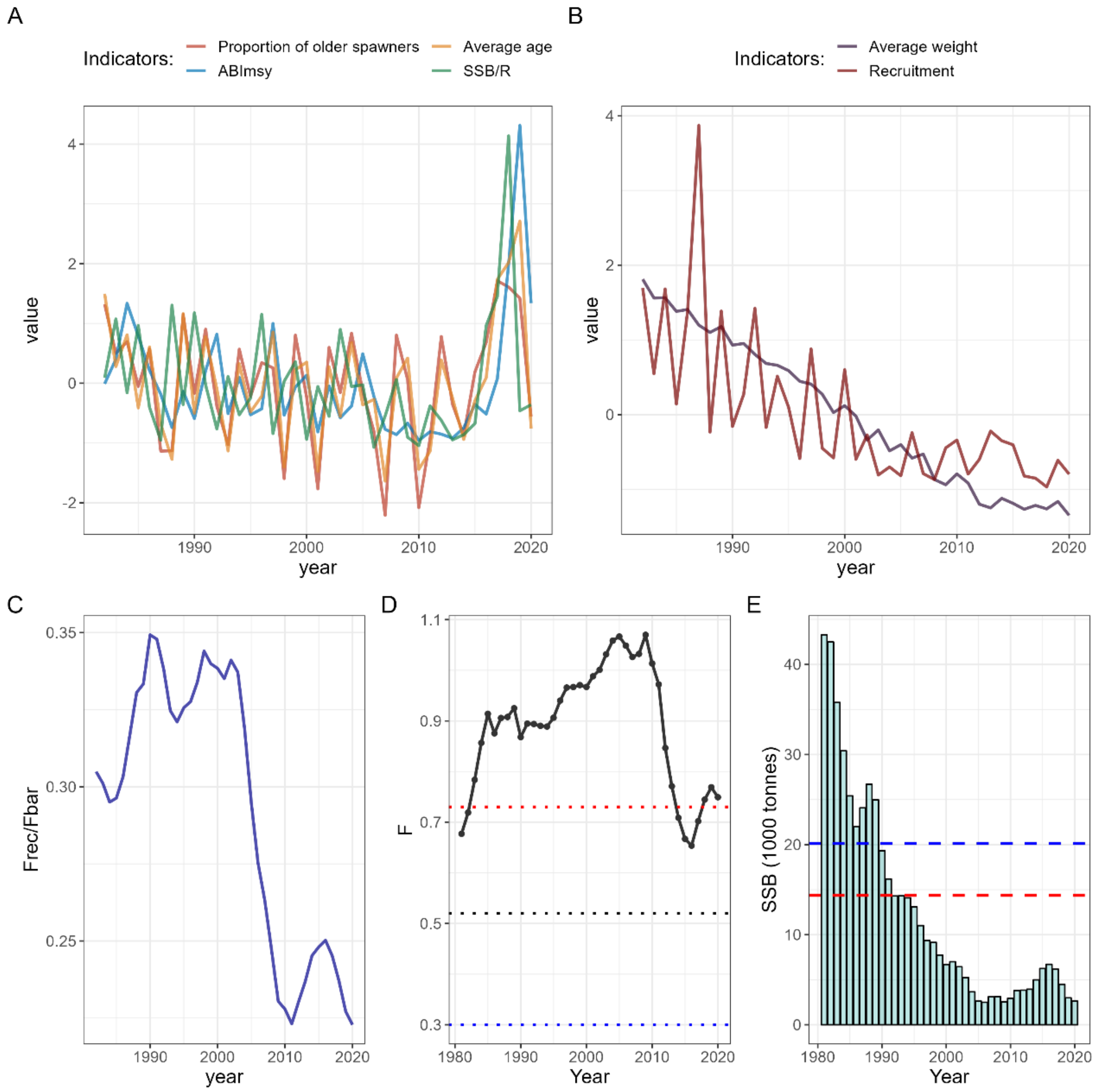

Figure A2.11. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.6a). See Figure legend A2.1 for further details.

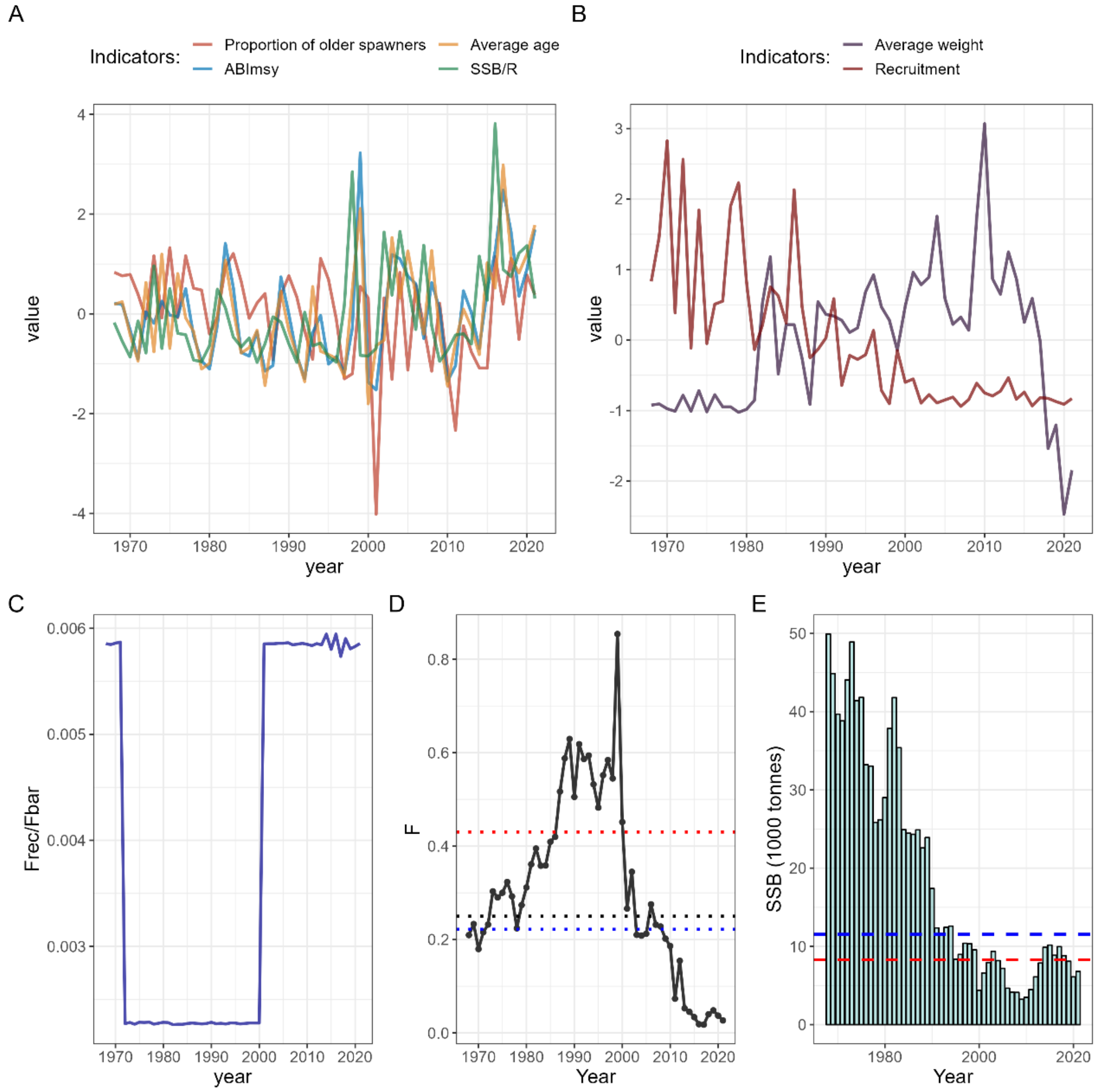

Figure A2.12. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.7a). See Figure legend A2.1 for further details.

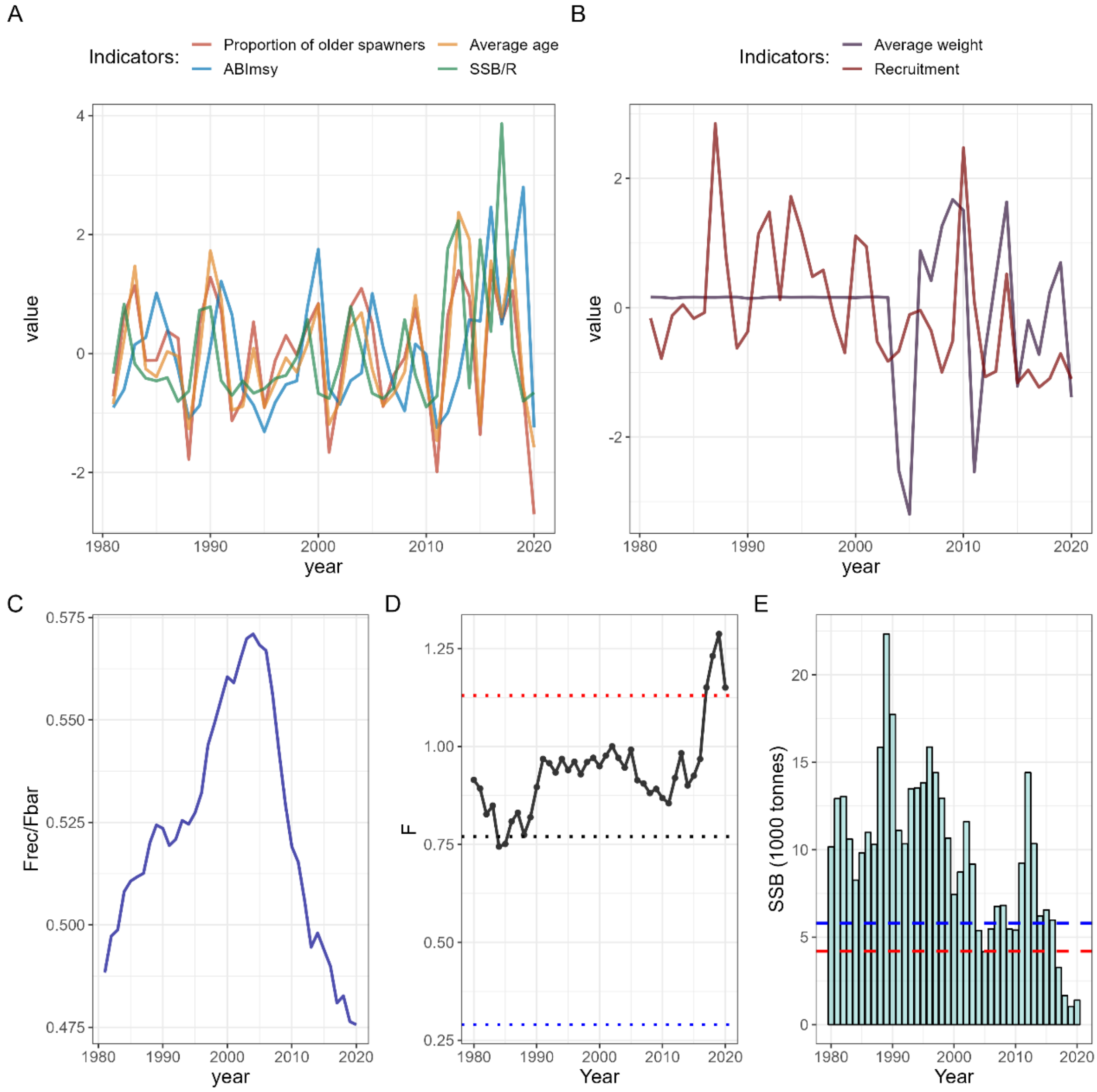

Figure A2.13. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.7e-k). See Figure legend A2.1 for further details.

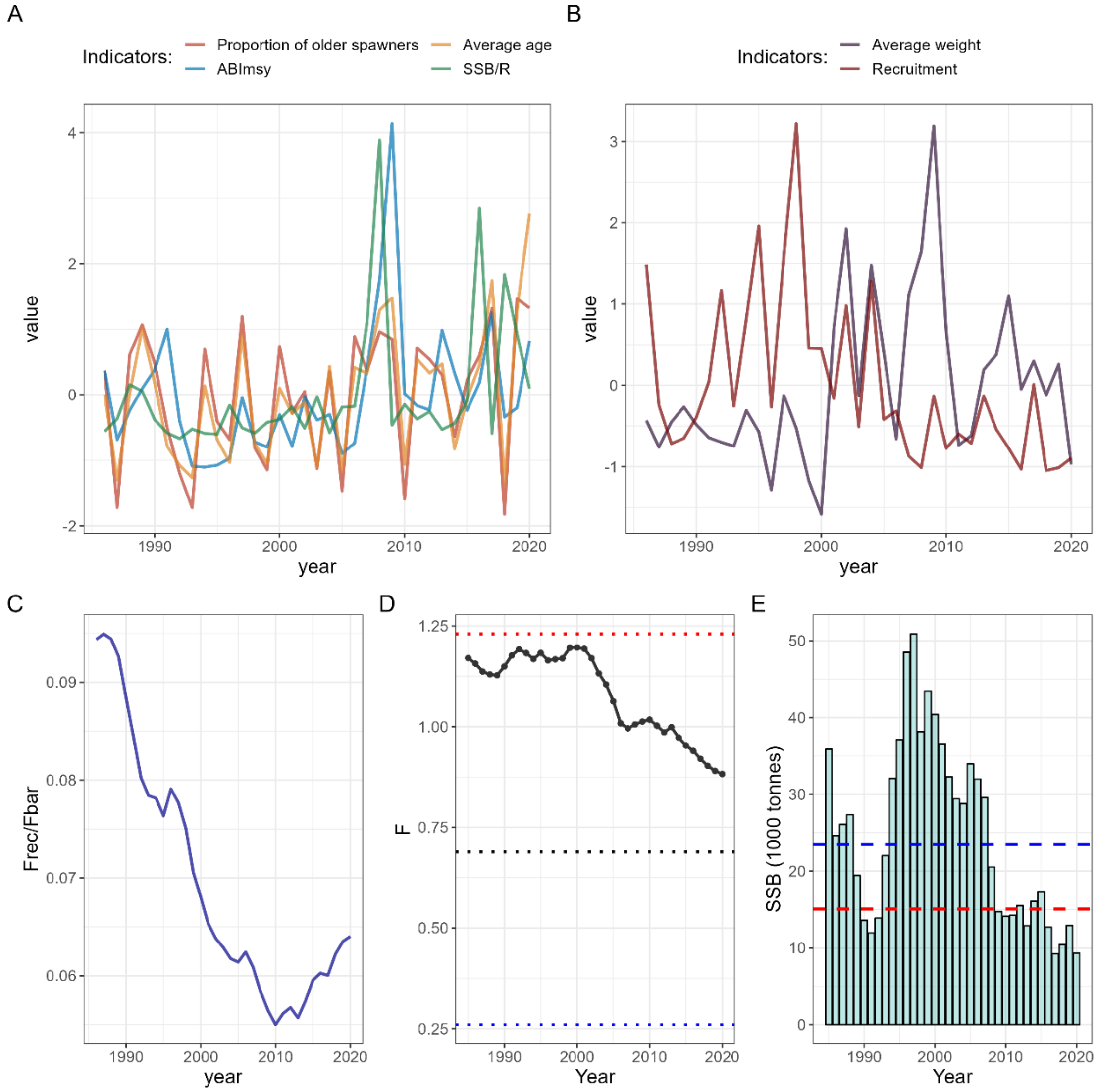

Figure A2.14. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.22-24). See Figure legend A2.1 for further details.

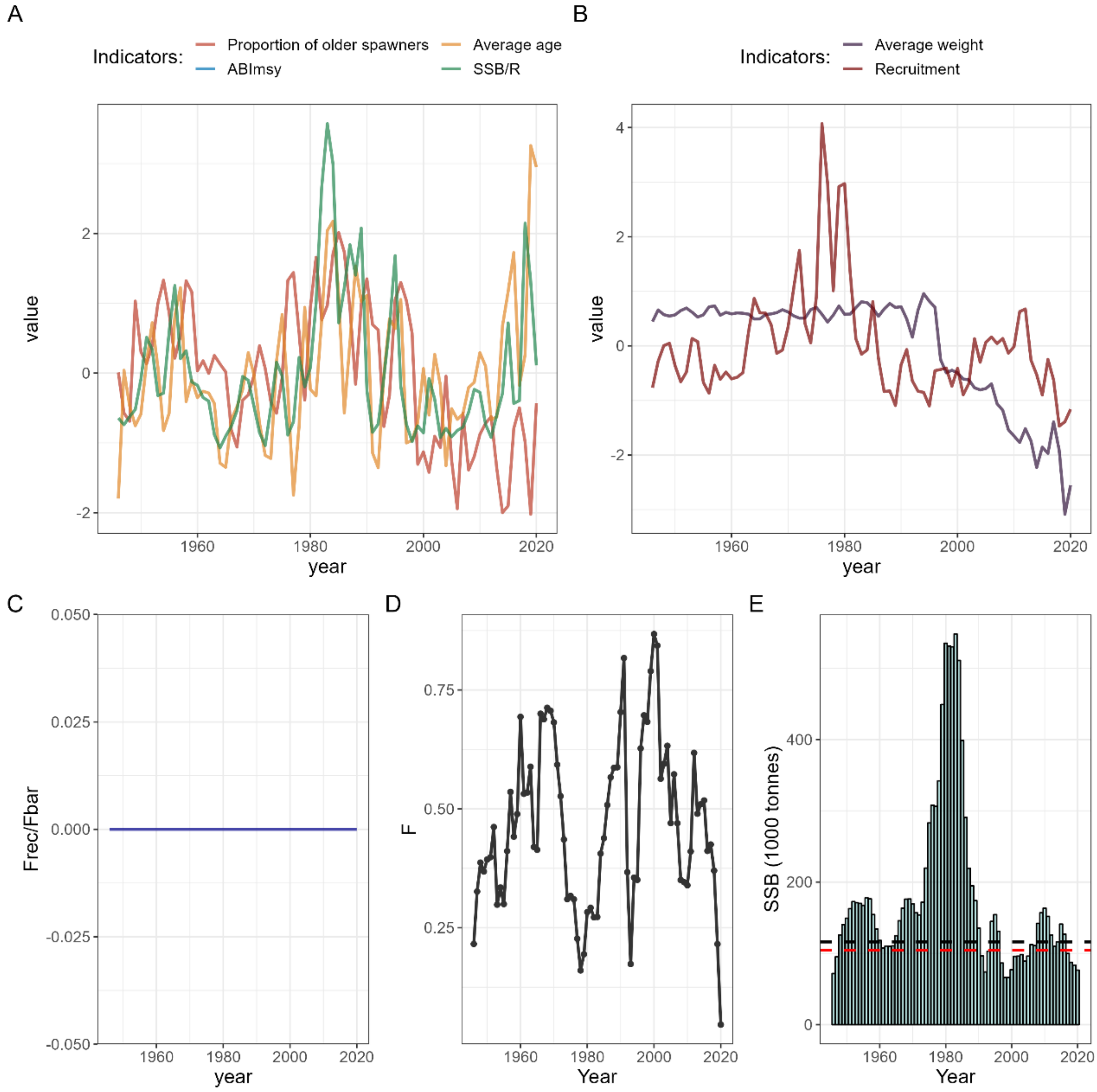

Figure A2.15. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.24-32). See Figure legend A2.1 for further details.

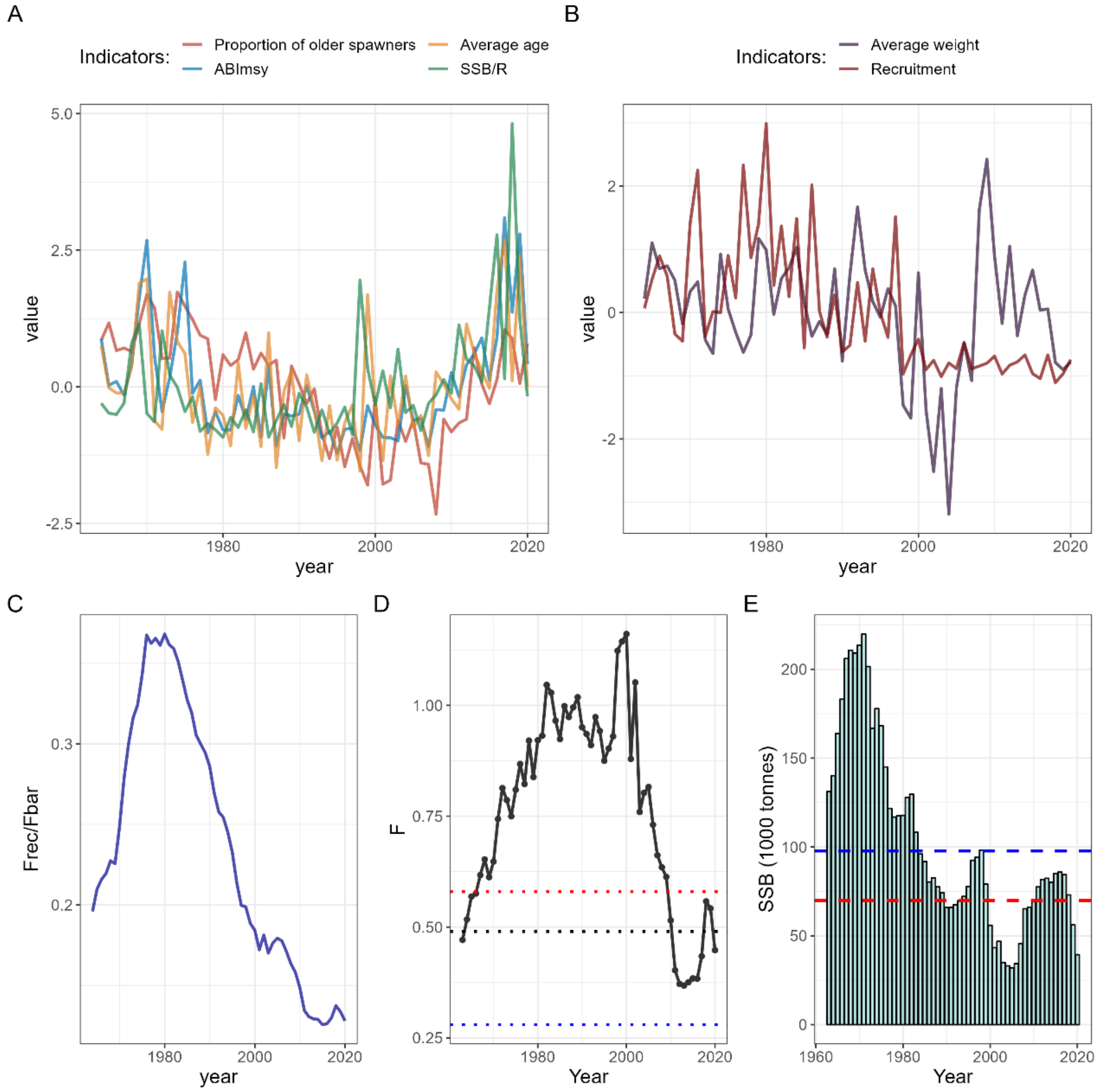

Figure A2.16. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.47d20). See Figure legend A2.1 for further details.

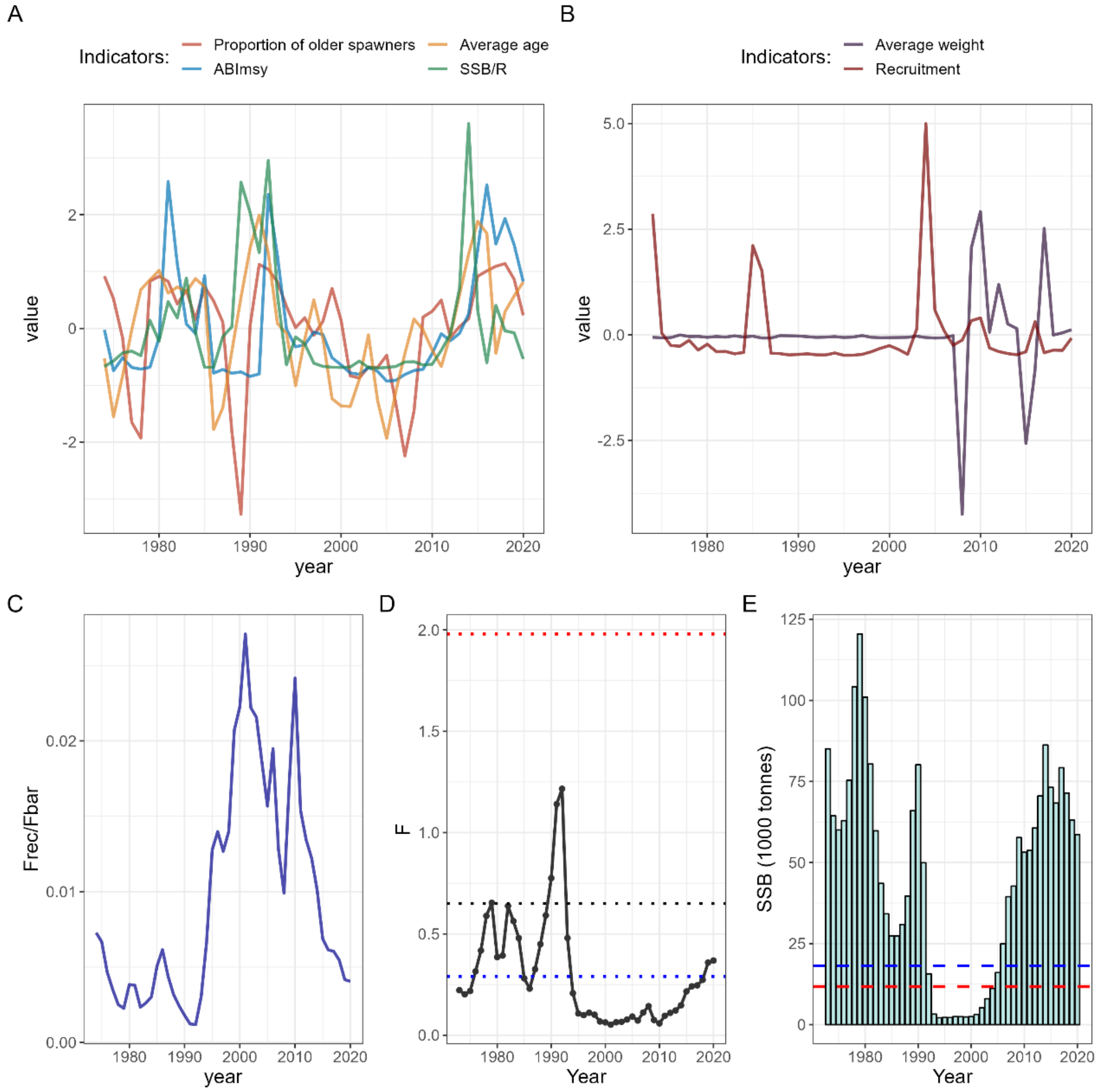

Figure A2.17. Estimated D3C3 indicators and stock status for Atlantic cod (cod.27.2127.1f14). See Figure legend A2.1 for further details.

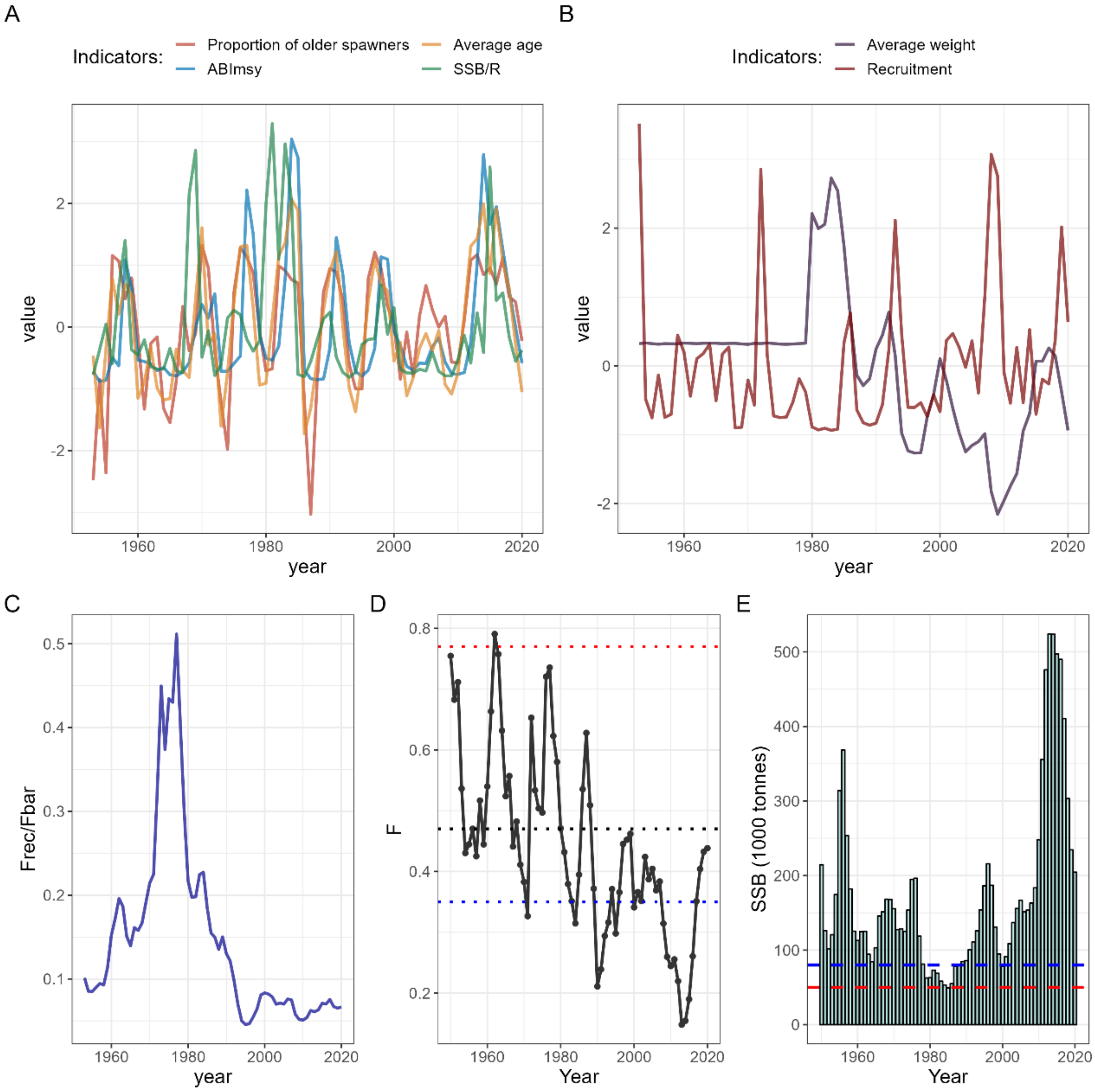

Figure A2.18. Estimated D3C3 indicators and stock status for Haddock (had.27.1-2). See Figure legend A2.1 for further details.

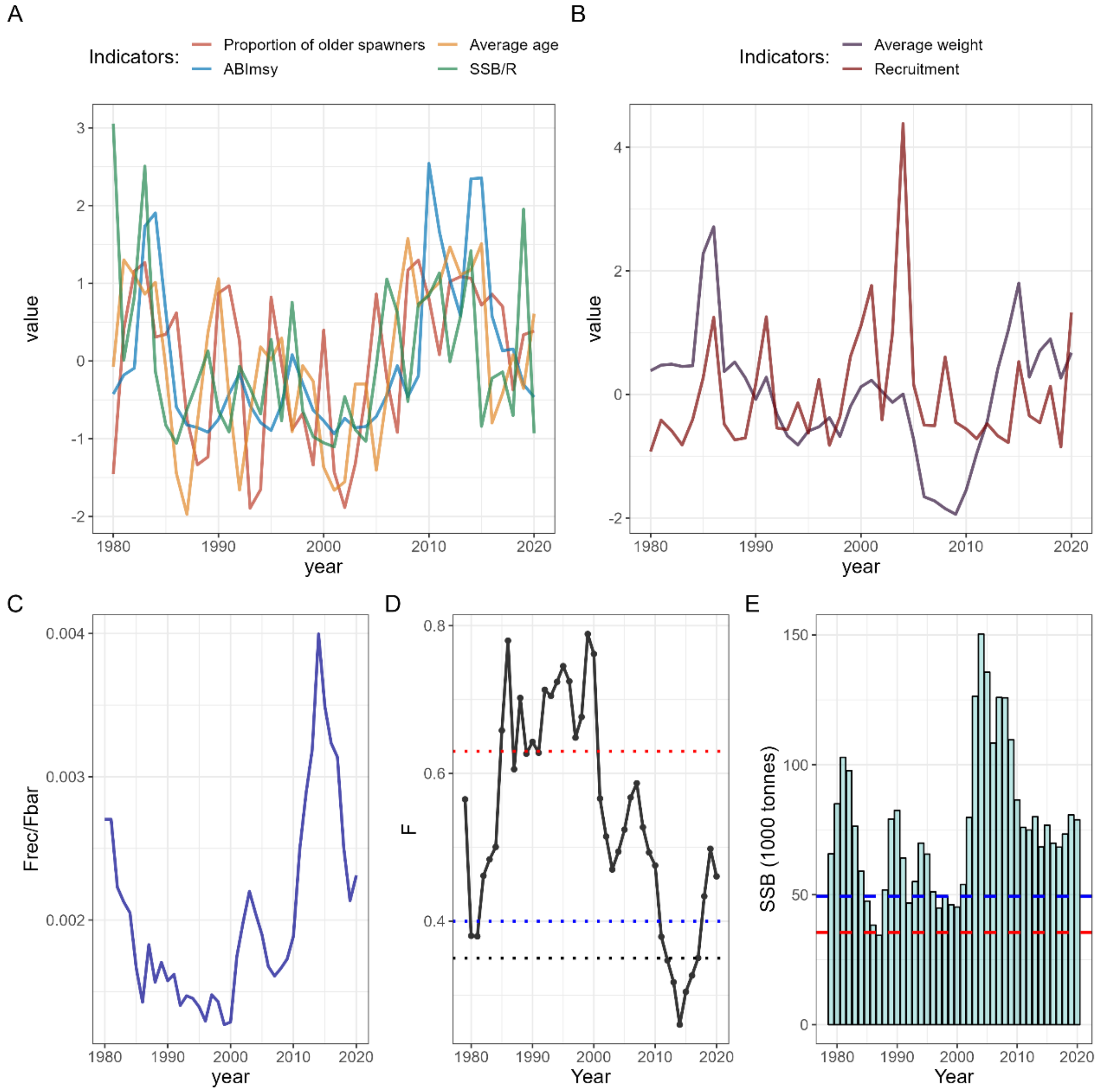

Figure A2.19. Estimated D3C3 indicators and stock status for Haddock (had.27.5a). See Figure legend A2.1 for further details.

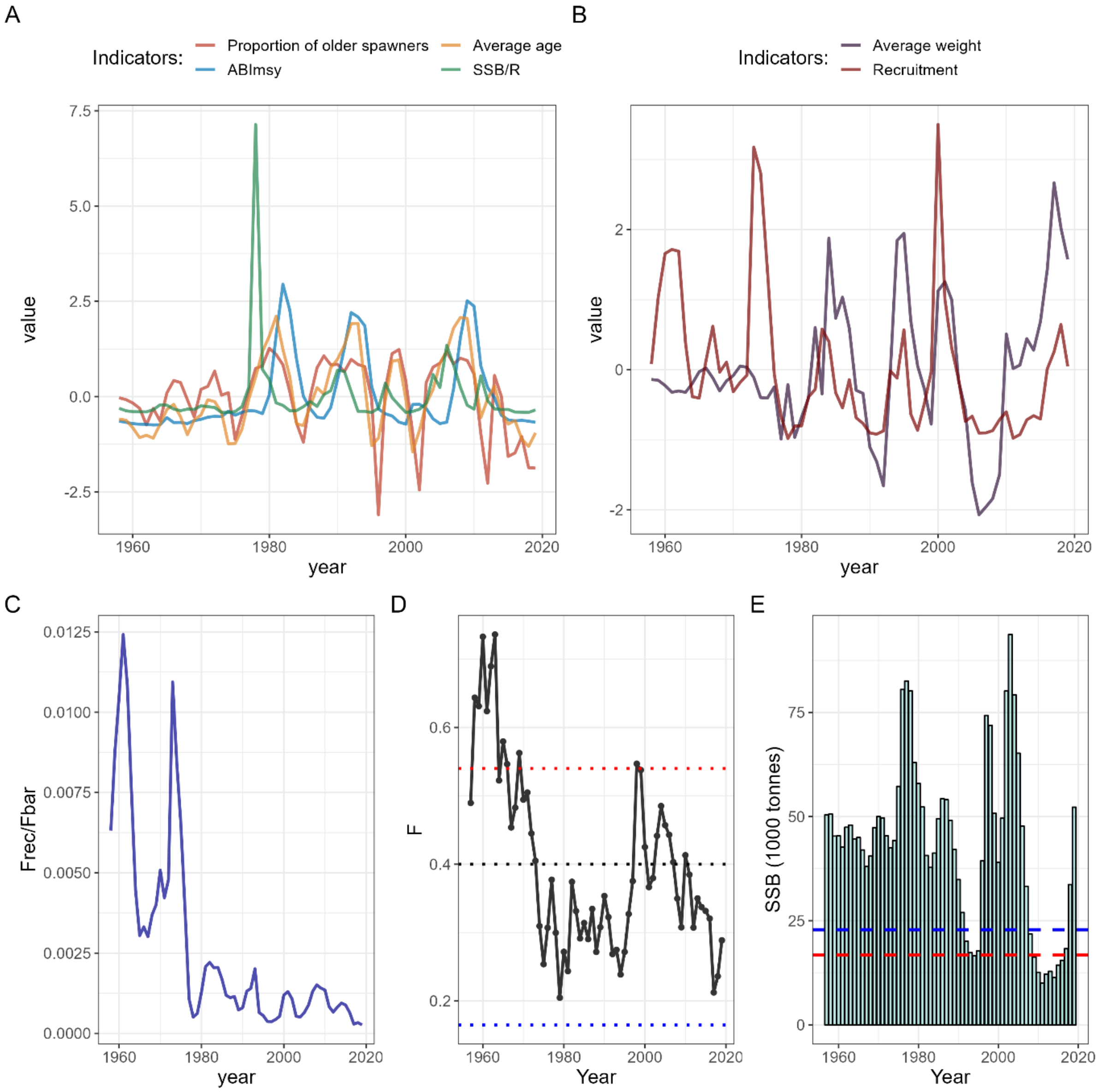

Figure A2.20. Estimated D3C3 indicators and stock status for Haddock (had.27.5b). See Figure legend A2.1 for further details.

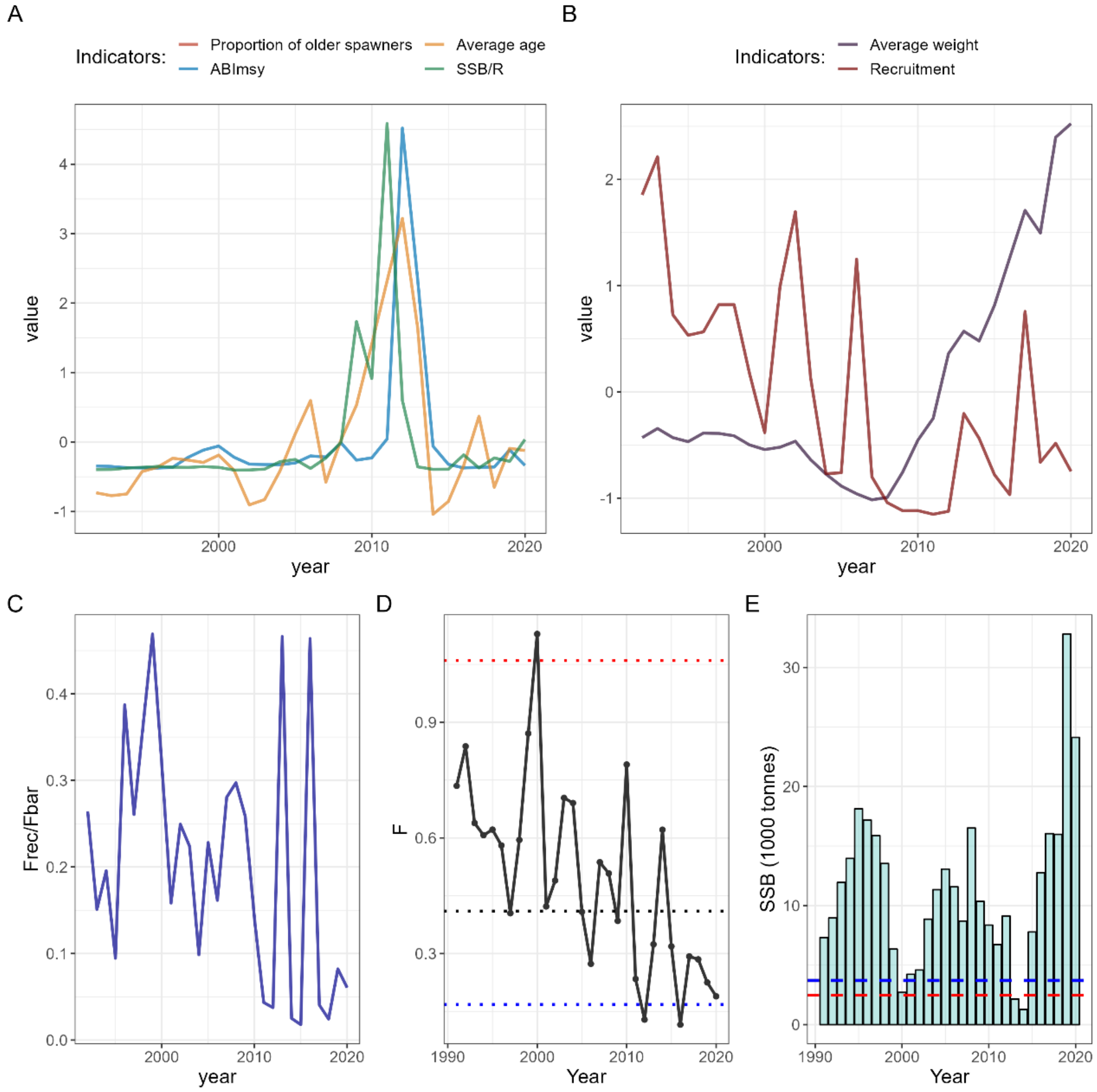

Figure A2.21. Estimated D3C3 indicators and stock status for Haddock (had.27.6b). See Figure legend A2.1 for further details.

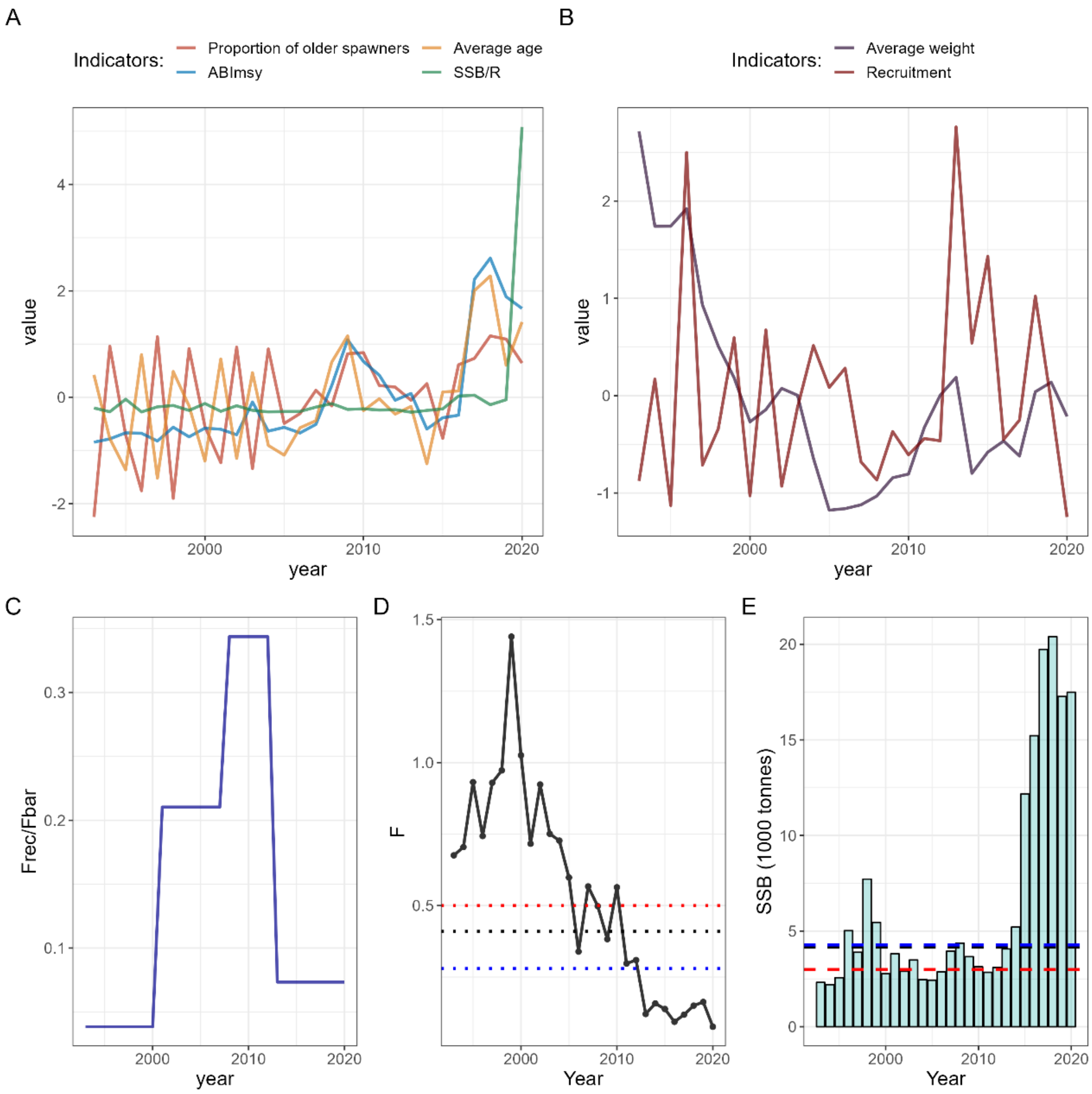

Figure A2.22. Estimated D3C3 indicators and stock status for Haddock (had.27.7a). See Figure legend A2.1 for further details.

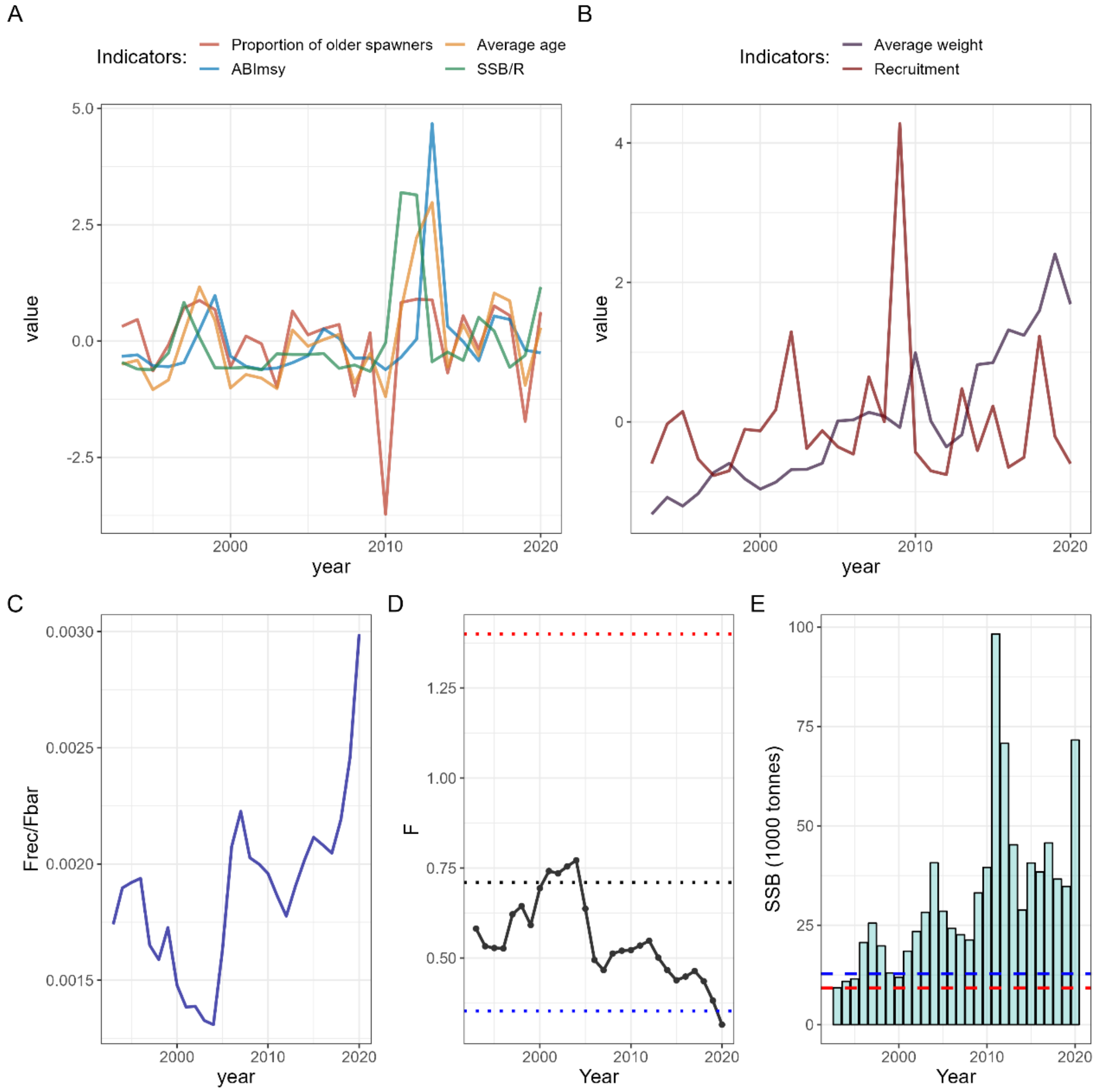

Figure A2.23. Estimated D3C3 indicators and stock status for Haddock (had.27.7b-k). See Figure legend A2.1 for further details.

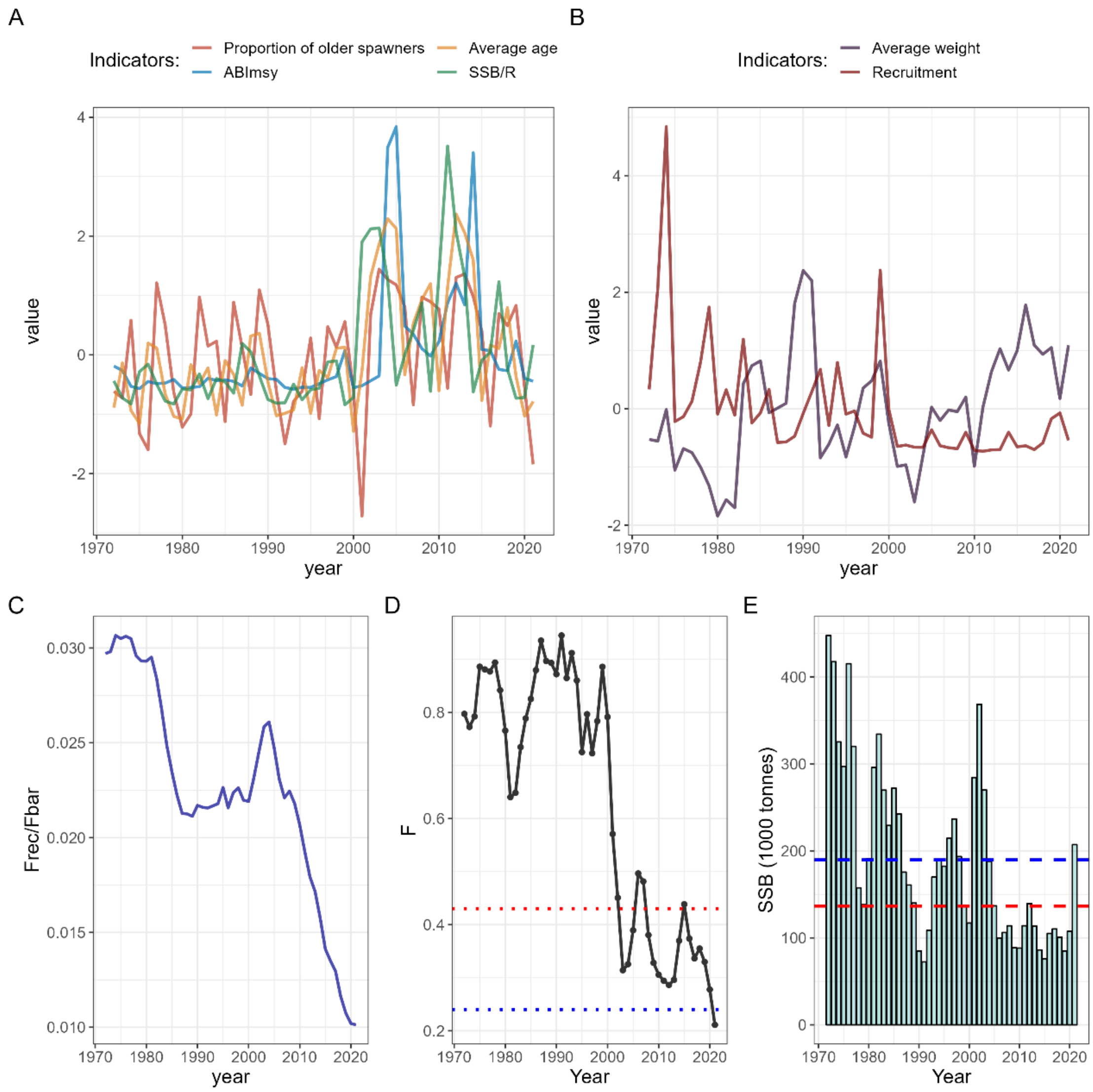

Figure A2.24. Estimated D3C3 indicators and stock status for Haddock (had.27.46a20). See Figure legend A2.1 for further details.

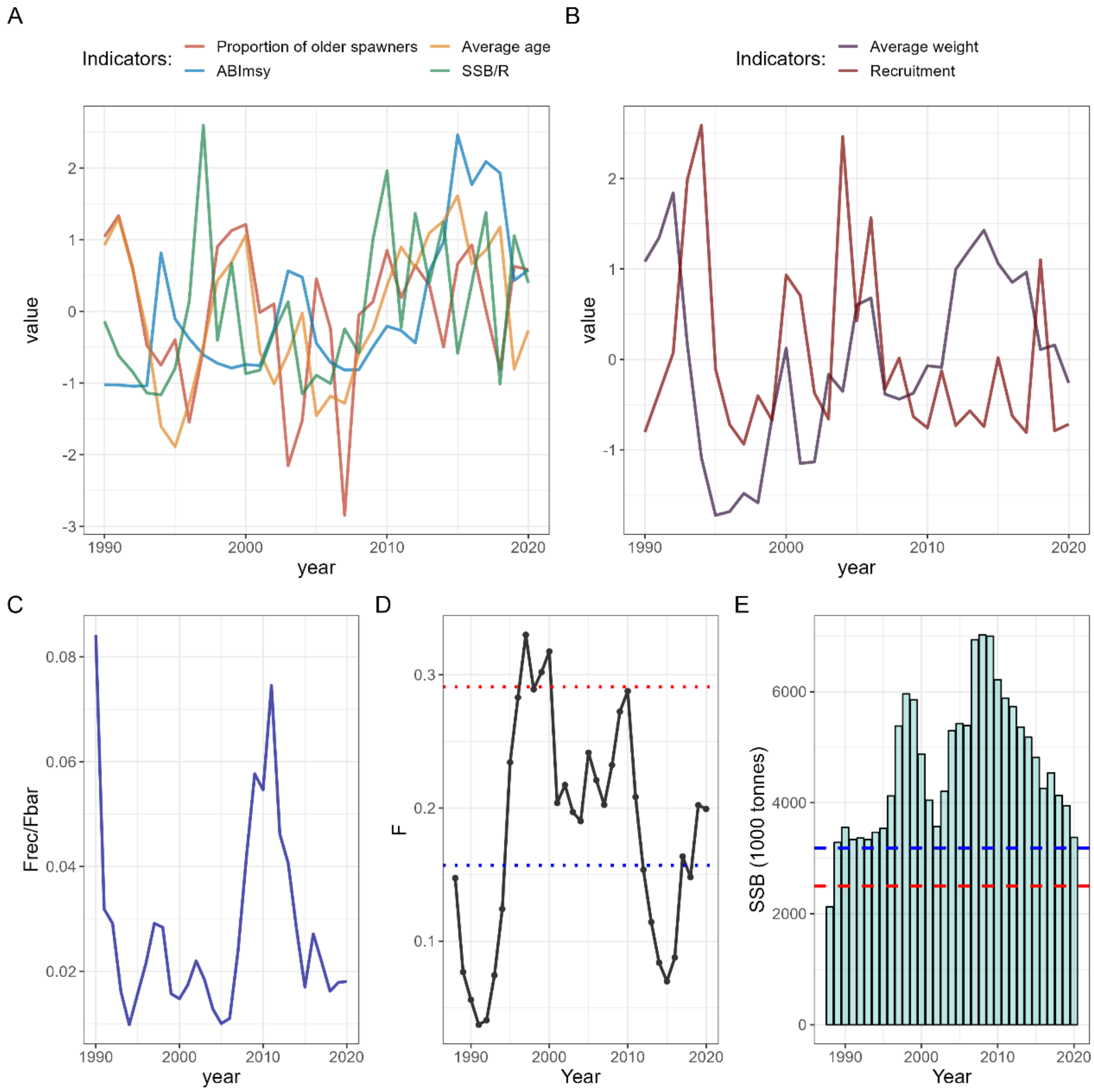

Figure A2.25. Estimated D3C3 indicators and stock status for Herring (her.27.1-24a514a). See Figure legend A2.1 for further details.

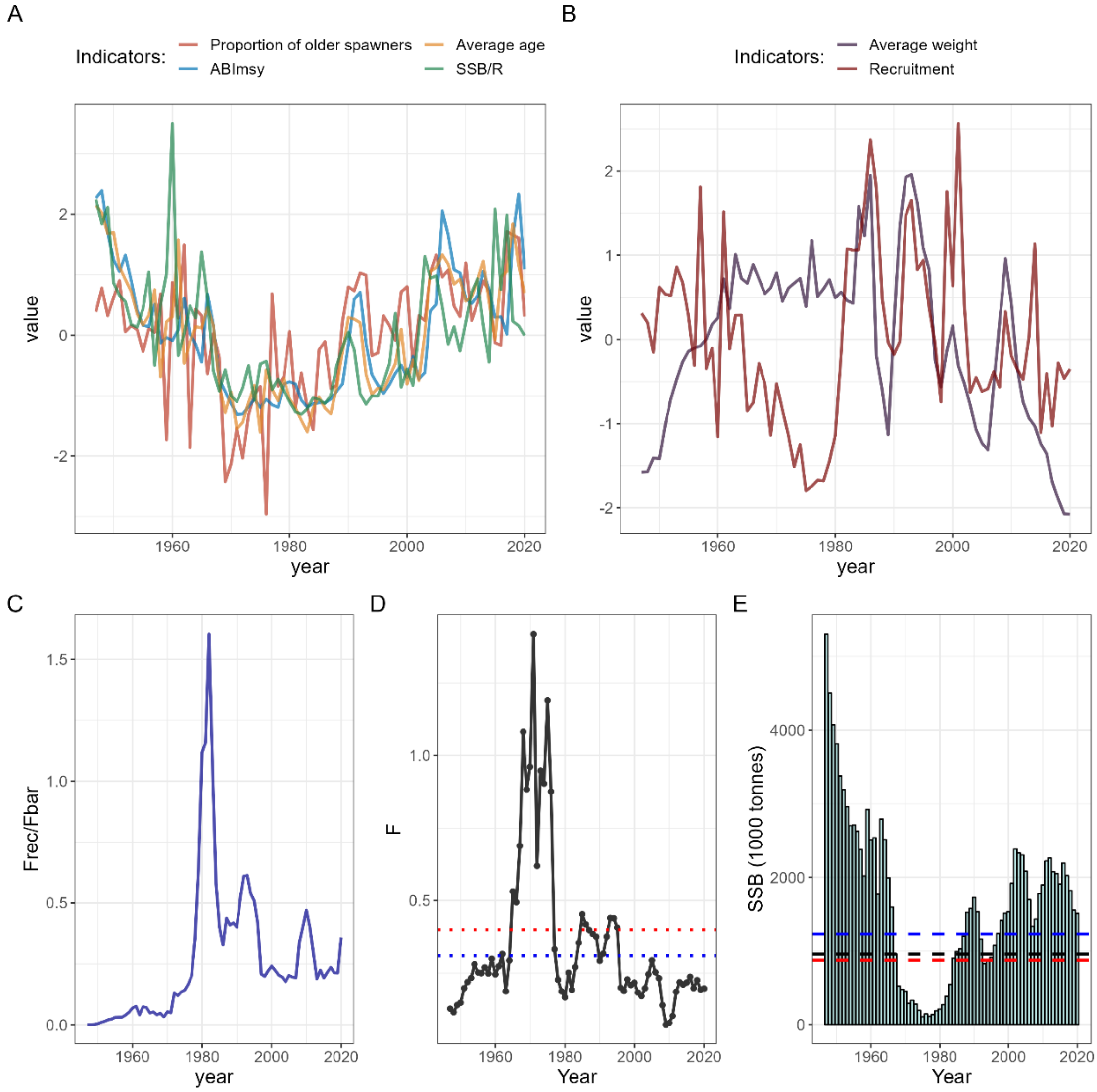

Figure A2.26. Estimated D3C3 indicators and stock status for Herring (her.27.3a47d). See Figure legend A2.1 for further details.

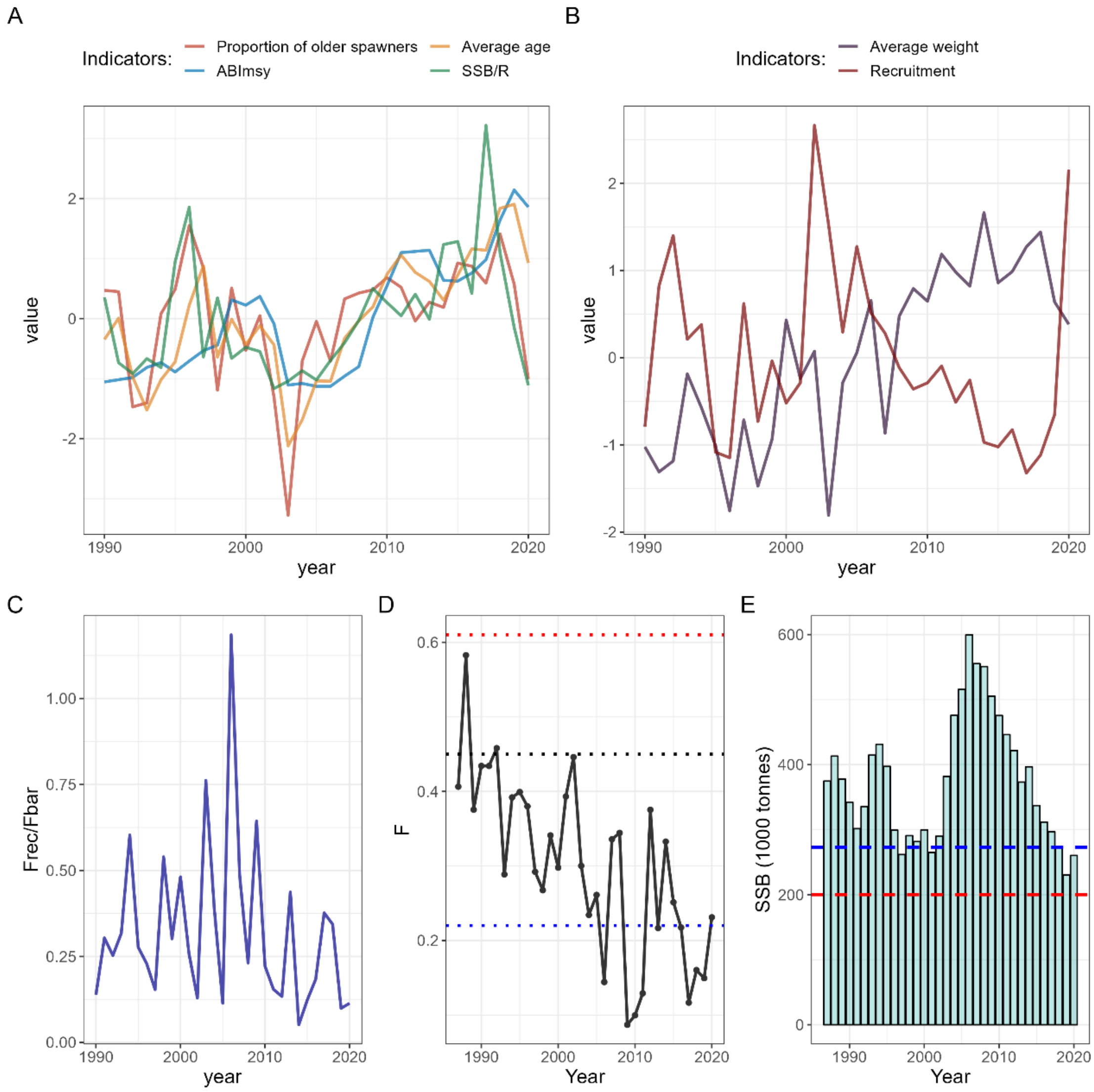

Figure A2.27. Estimated D3C3 indicators and stock status for Herring (her.27.5a). See Figure legend A2.1 for further details.

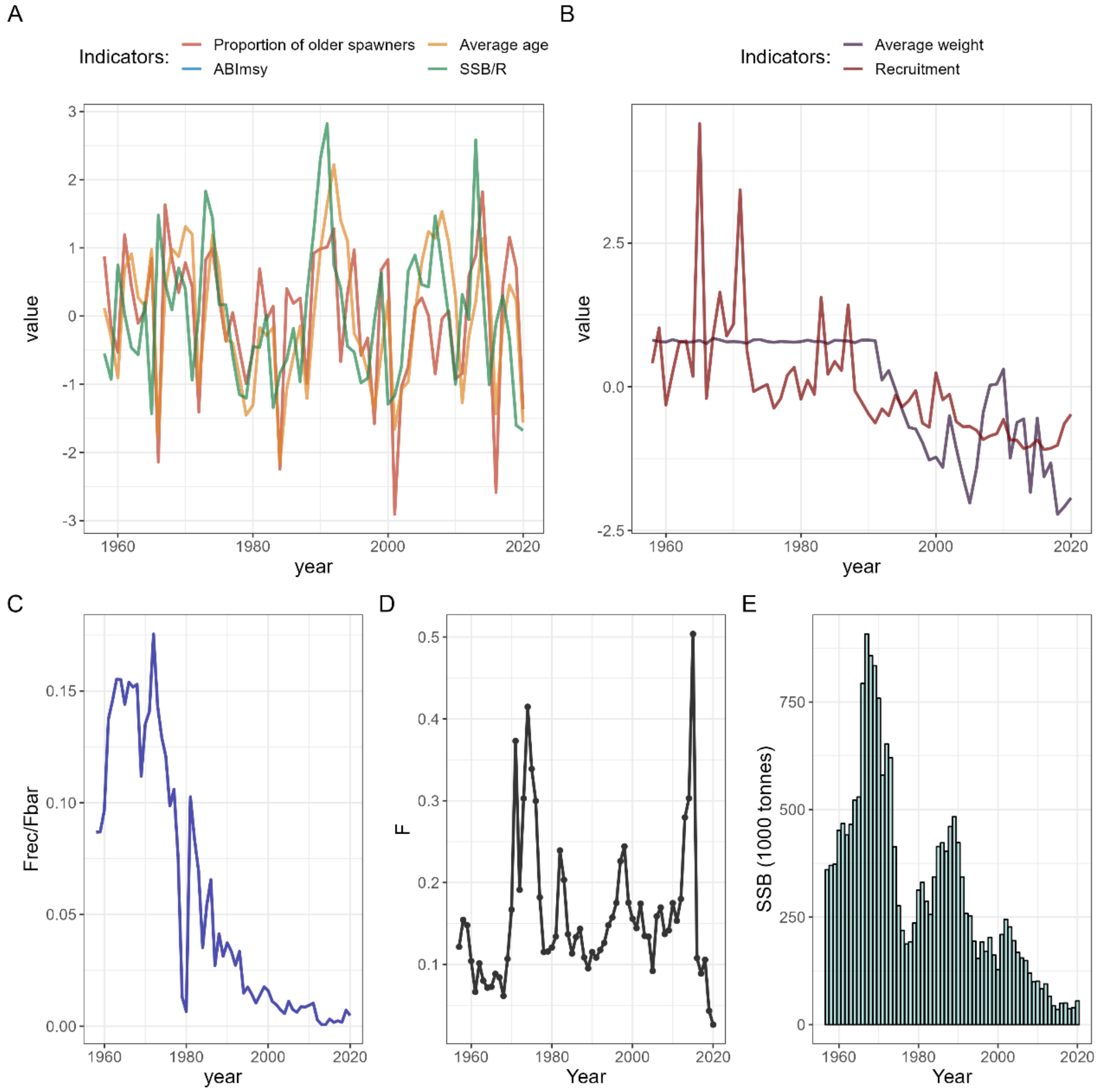

Figure A2.28. Estimated D3C3 indicators and stock status for Herring (her.27.6a7bc). See Figure legend A2.1 for further details.

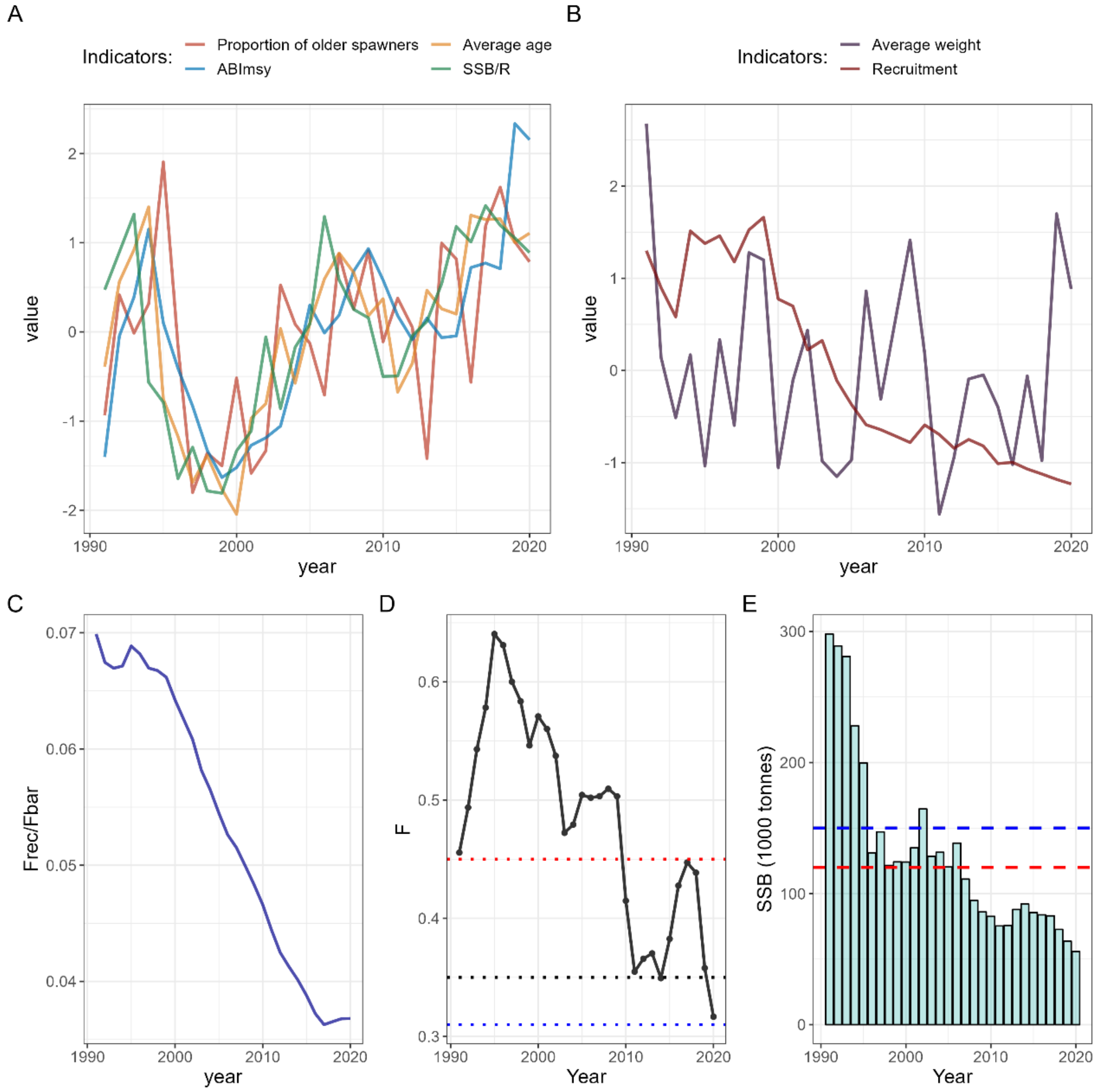

Figure A2.29. Estimated D3C3 indicators and stock status for Herring (her.27.20-24). See Figure legend A2.1 for further details.

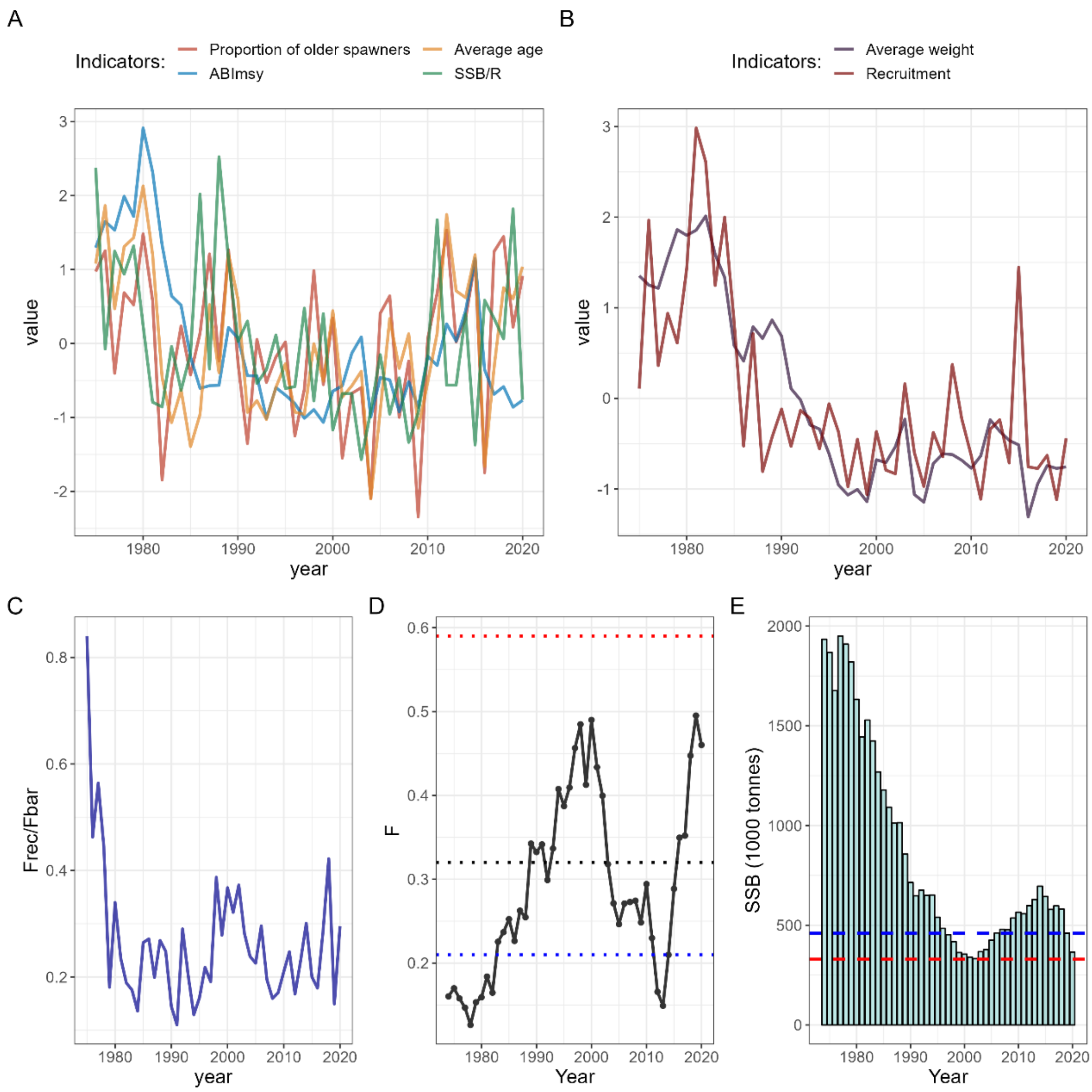

Figure A2.30. Estimated D3C3 indicators and stock status for Herring (her.27.25-2932). See Figure legend A2.1 for further details.

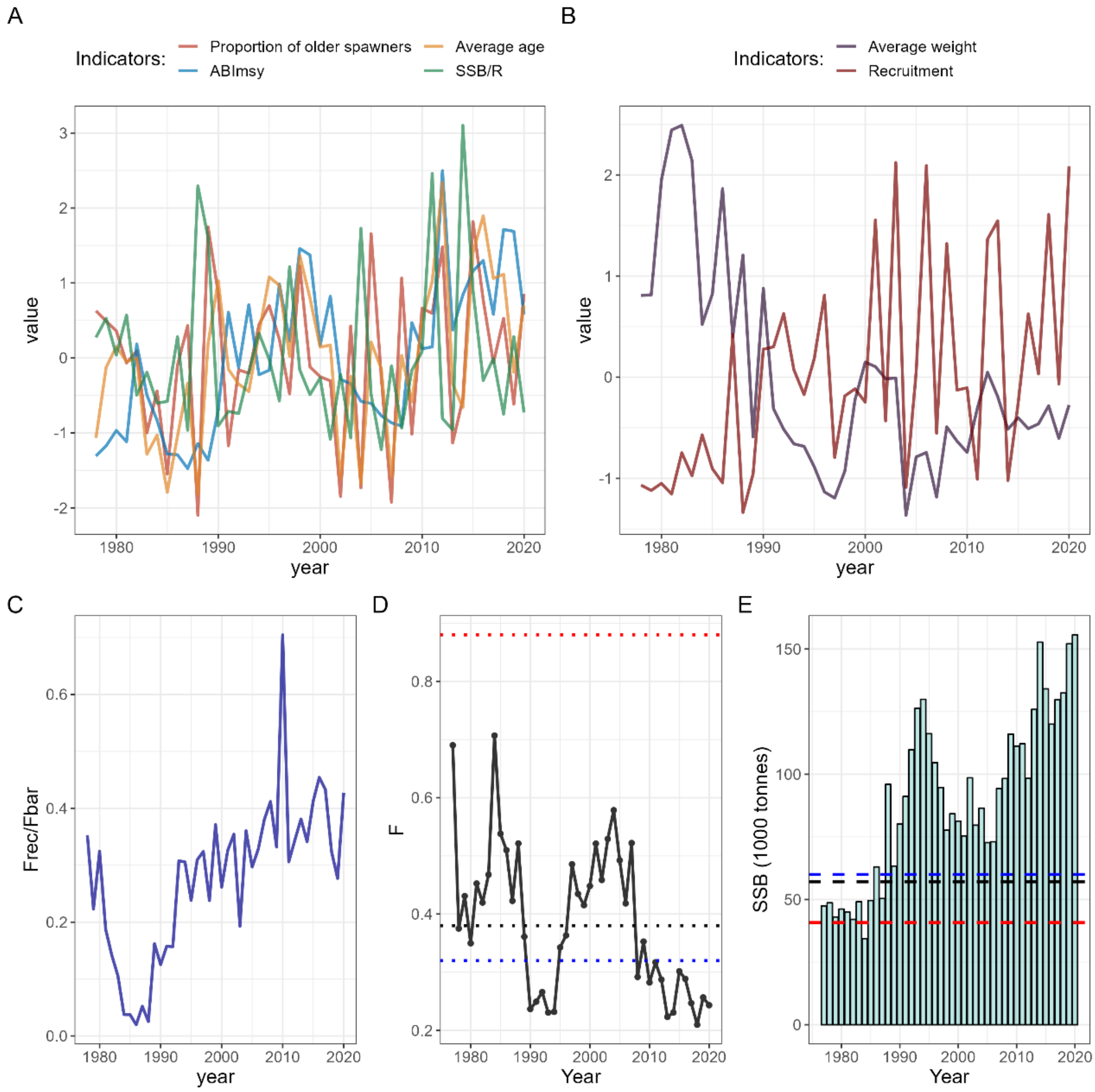

Figure A2.31. Estimated D3C3 indicators and stock status for Herring (her.27.28). See Figure legend A2.1 for further details.

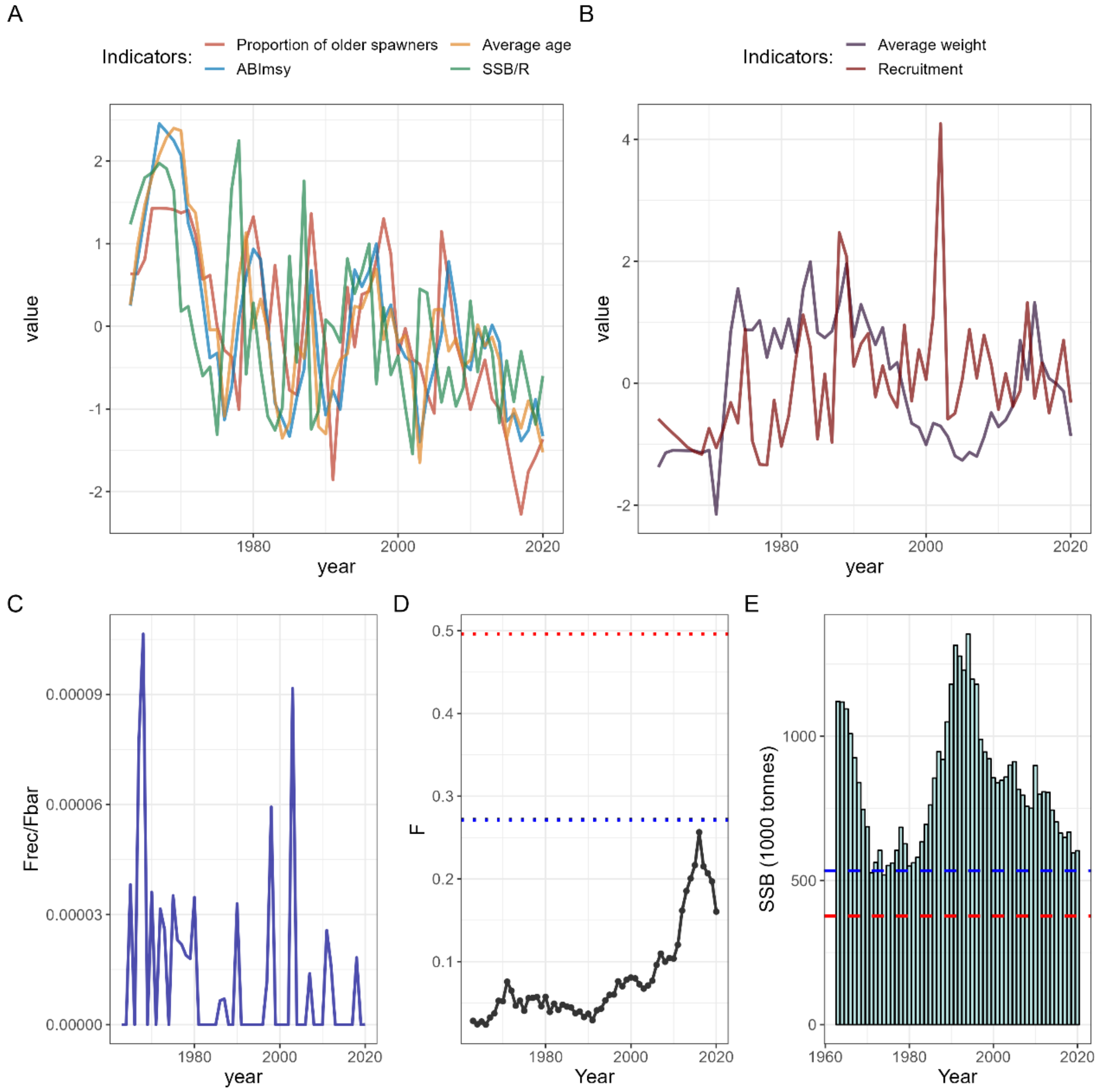

Figure A2.32. Estimated D3C3 indicators and stock status for Herring (her.27.3031). See Figure legend A2.1 for further details.

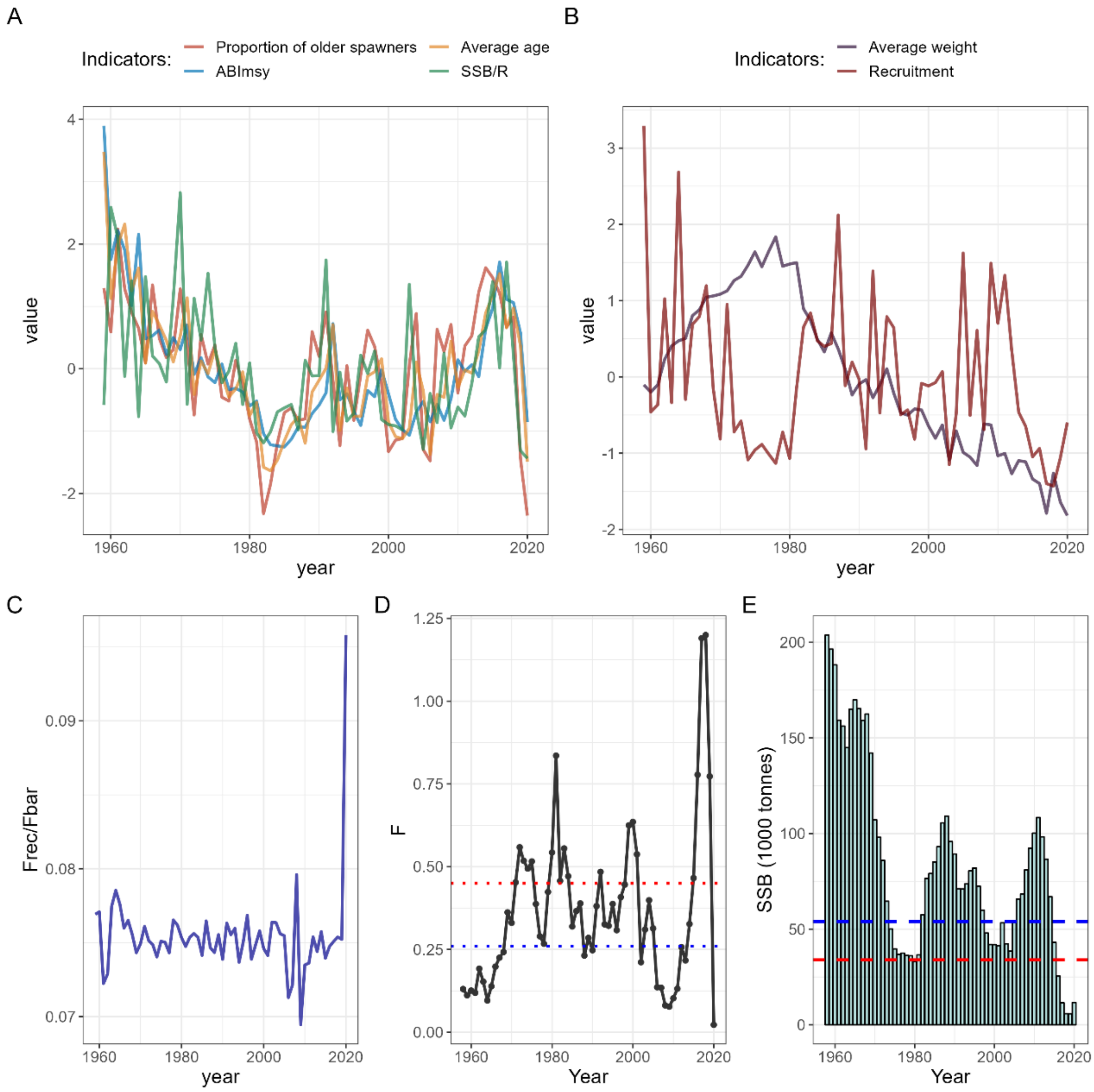

Figure A2.33. Estimated D3C3 indicators and stock status for Herring (her.27.irls). See Figure legend A2.1 for further details.

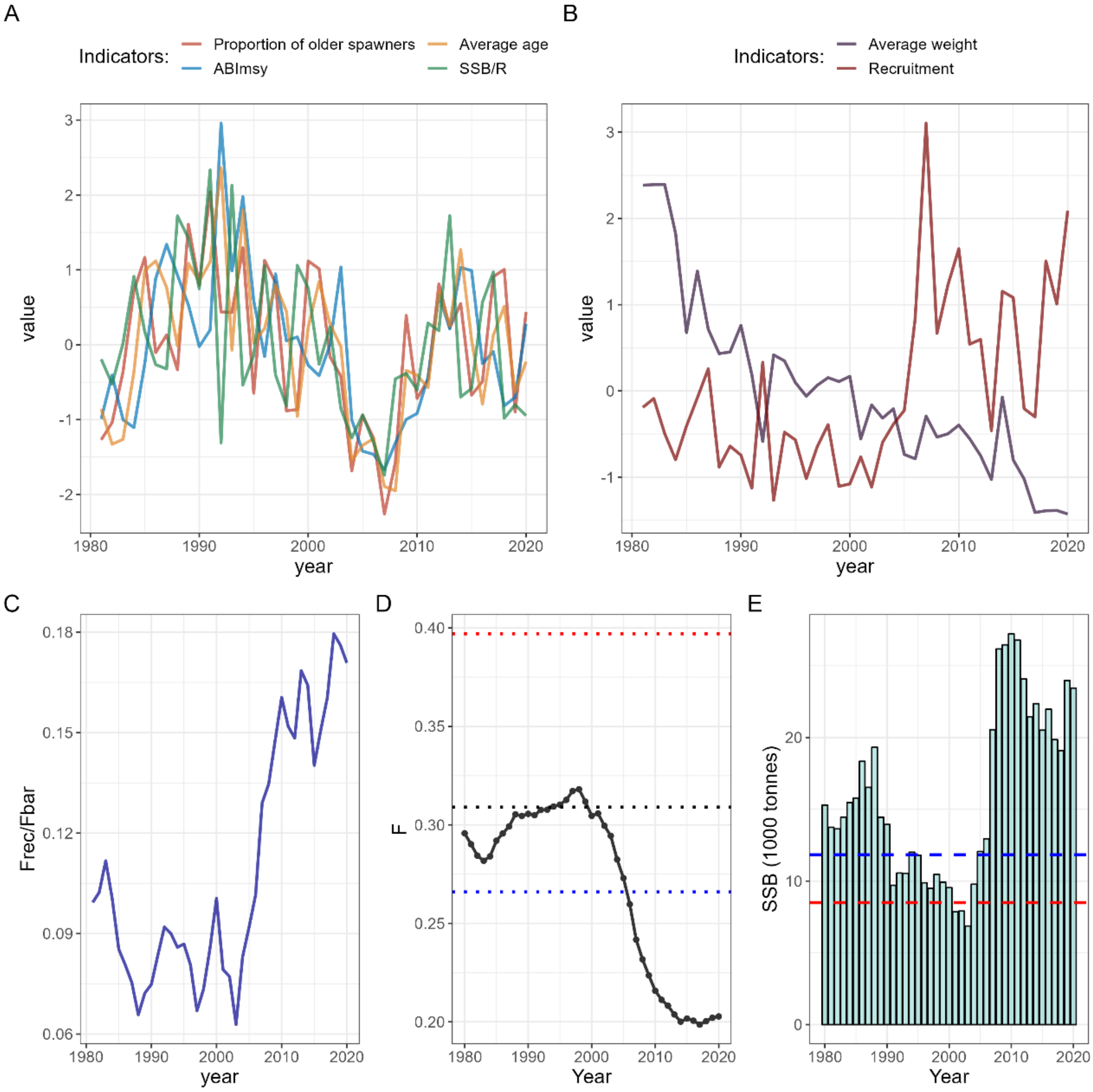

Figure A2.34. Estimated D3C3 indicators and stock status for Herring (her.27.nirs). See Figure legend A2.1 for further details.

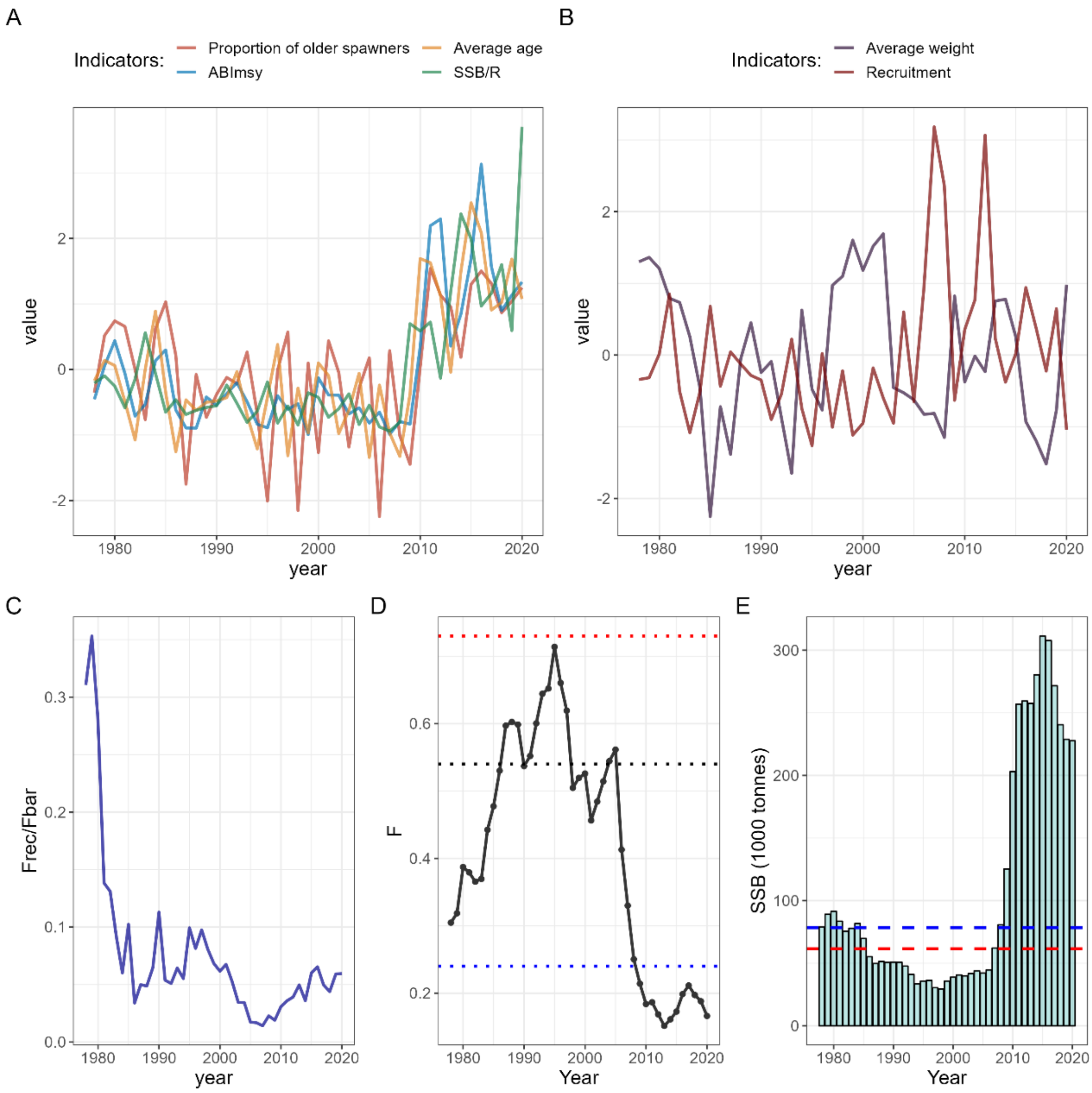

Figure A2.35. Estimated D3C3 indicators and stock status for Hake (her.27.3a46-8abd). See Figure legend A2.1 for further details.

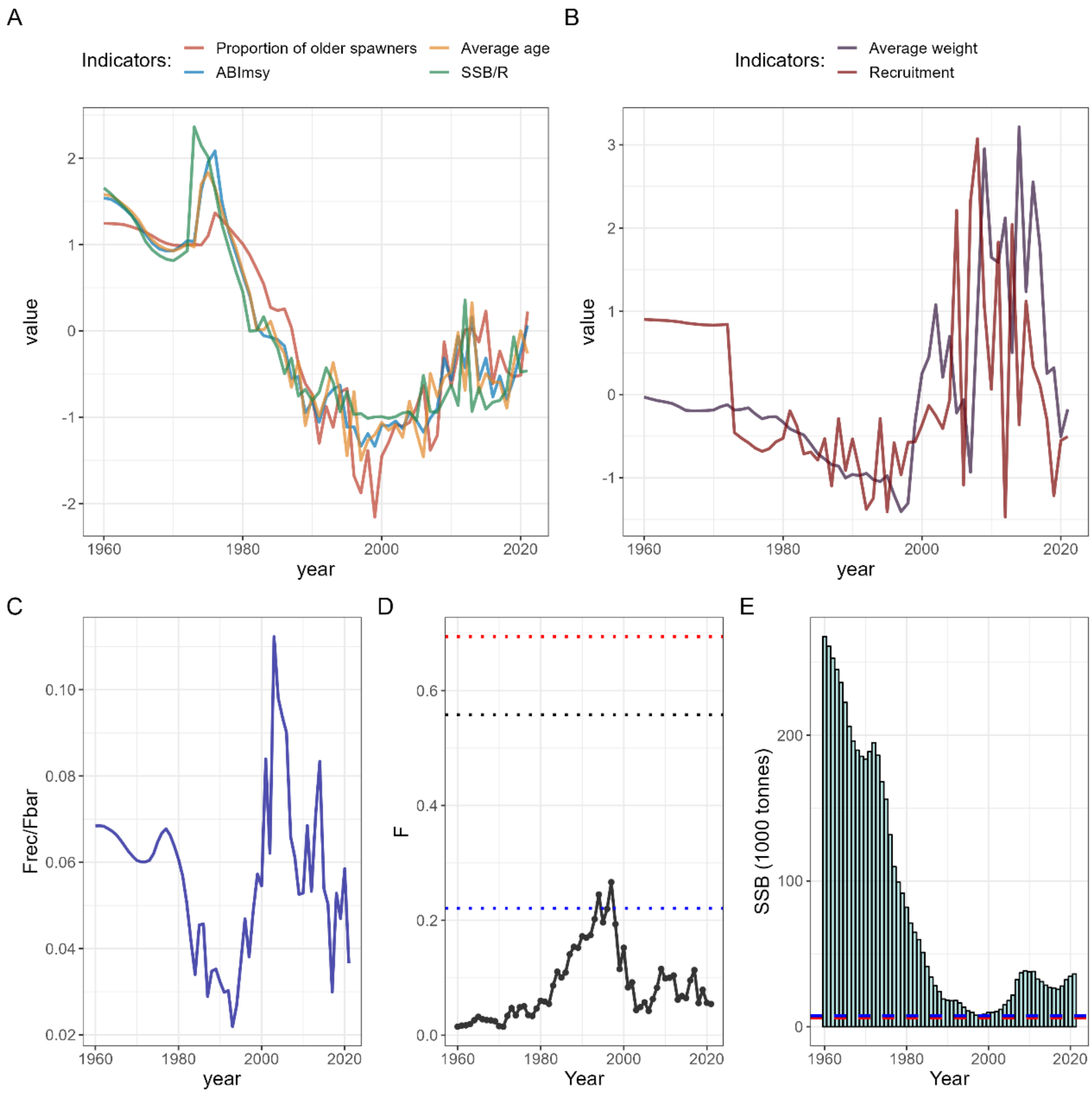

Figure A2.36. Estimated D3C3 indicators and stock status for Hake (her.27.8c9a). See Figure legend A2.1 for further details.

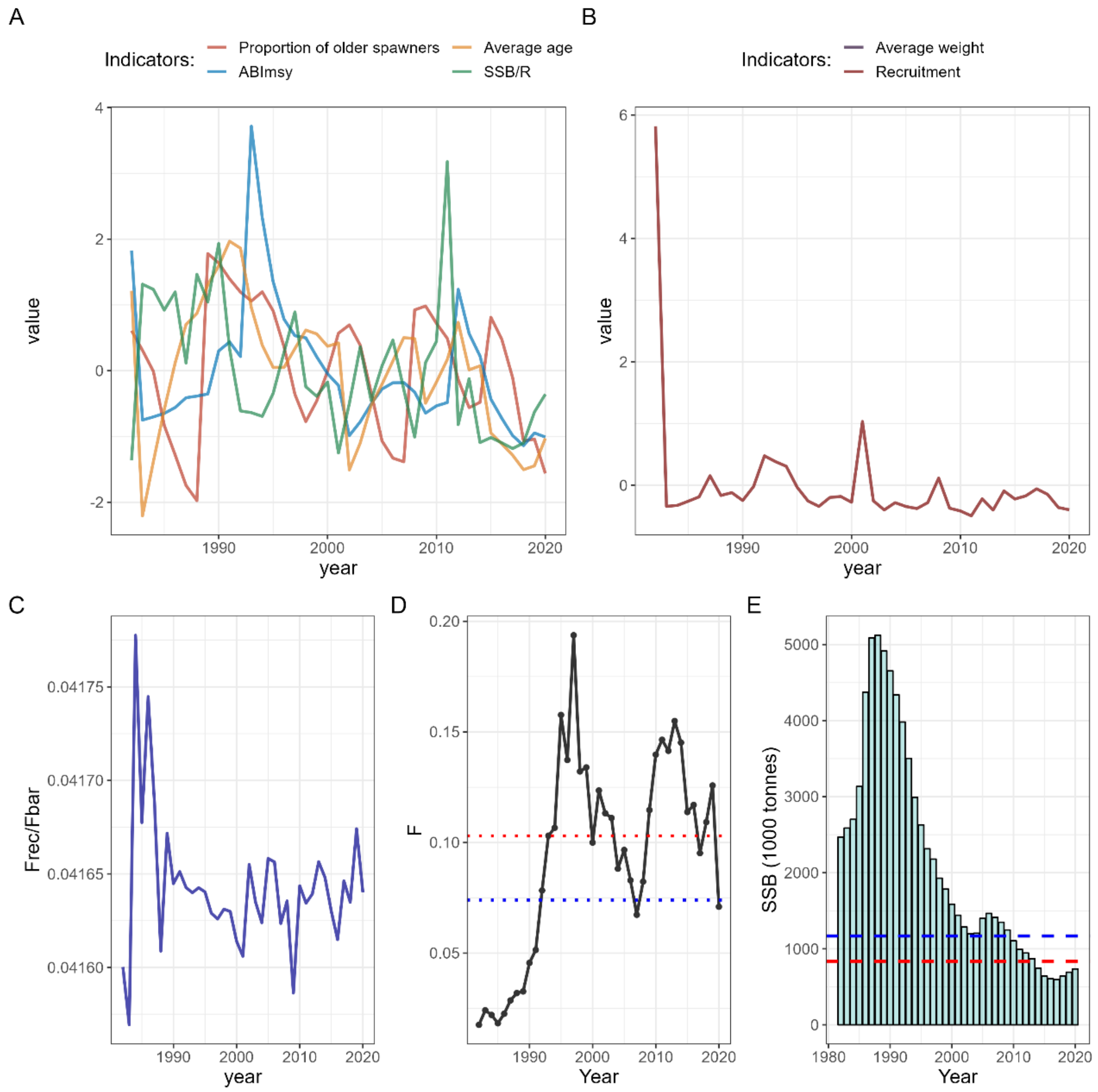

Figure A2.37. Estimated D3C3 indicators and stock status for Horse mackerel (hom.27.2a4a5b6a7a-ce-k8). See Figure legend A2.1 for further details.

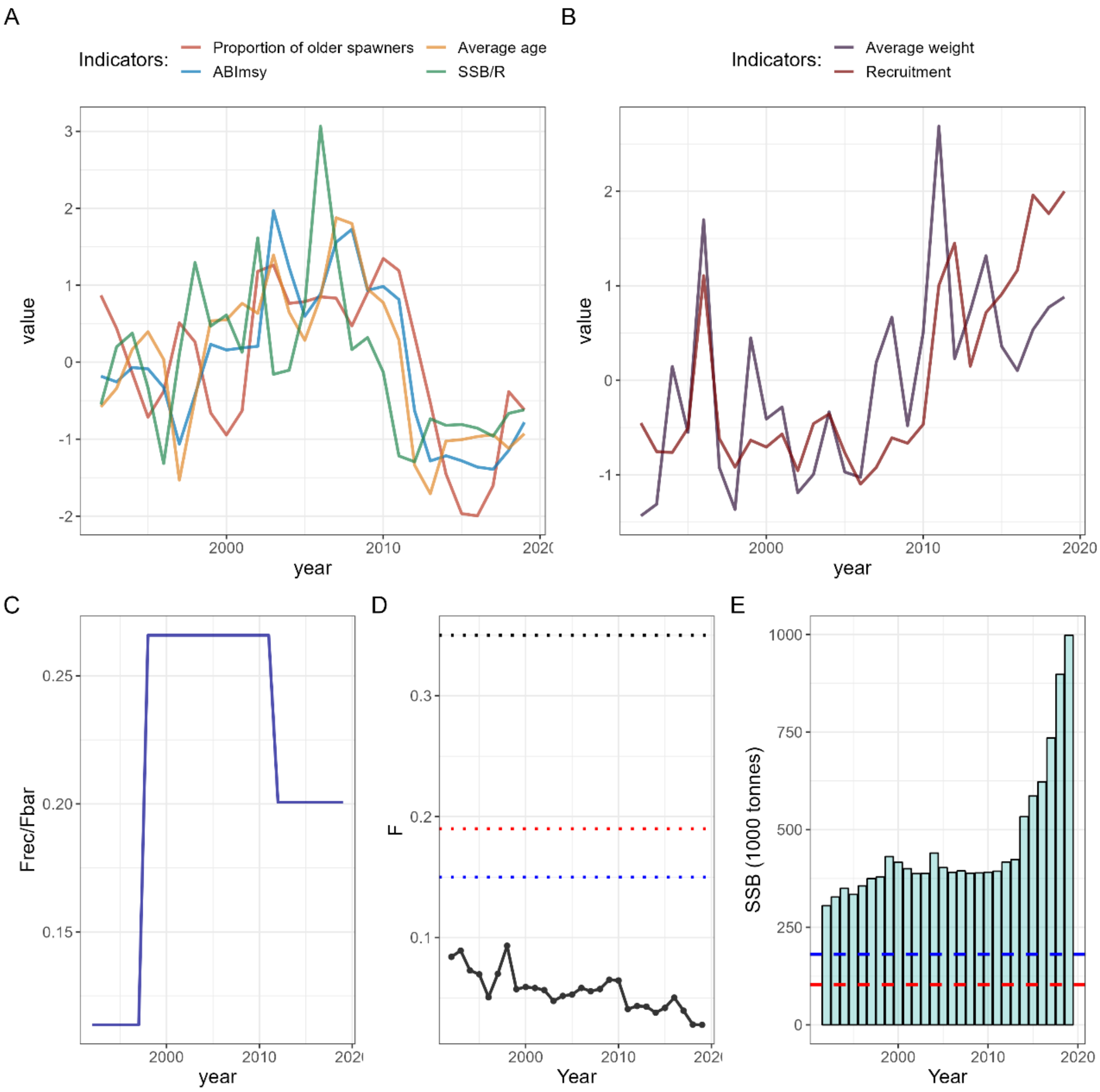

Figure A2.38. Estimated D3C3 indicators and stock status for Horse mackerel (hom.27.9a). See Figure legend A2.1 for further details.

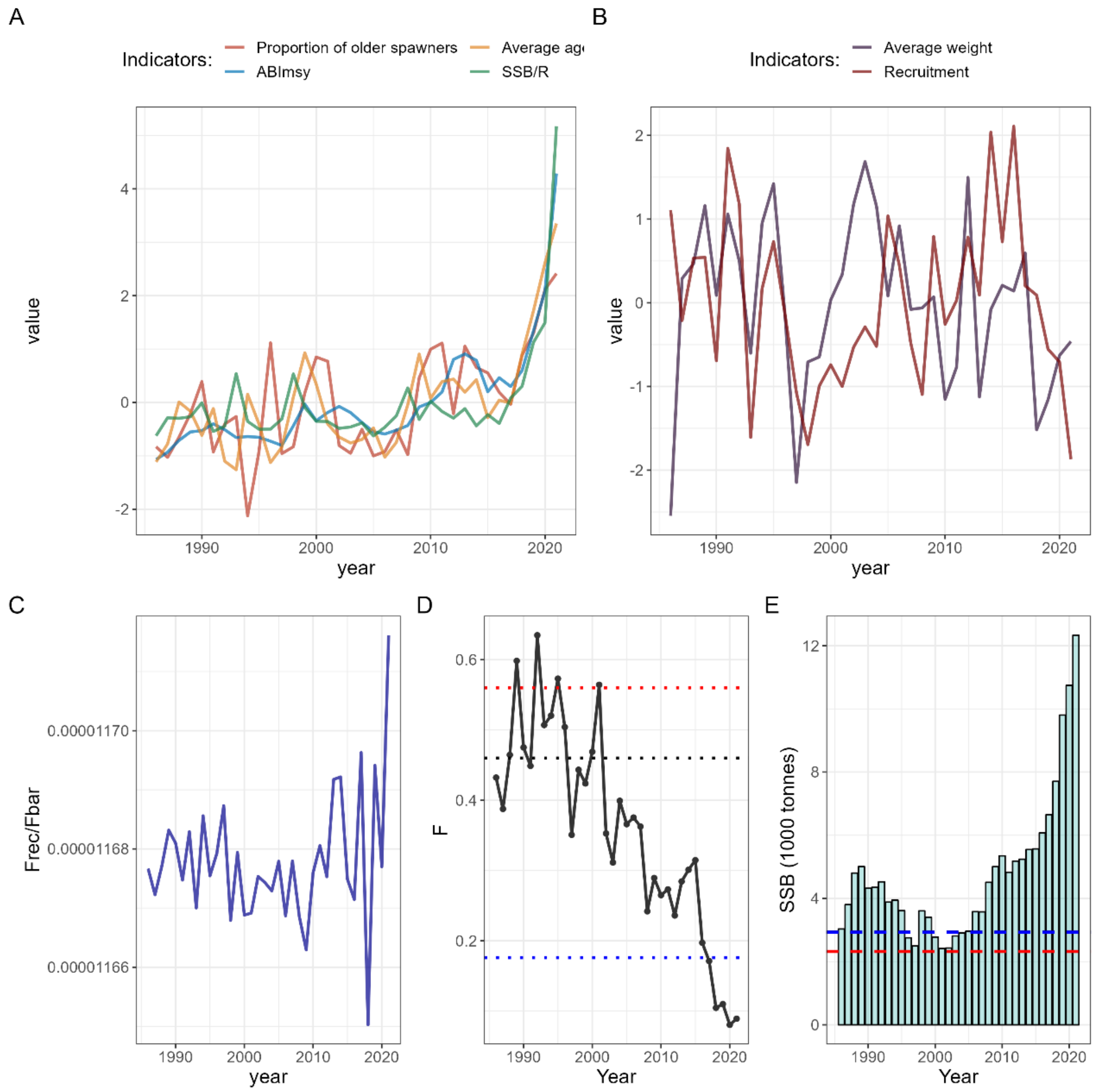

Figure A2.39. Estimated D3C3 indicators and stock status for Megrim (ldb.27.8c9a). See Figure legend A2.1 for further details.

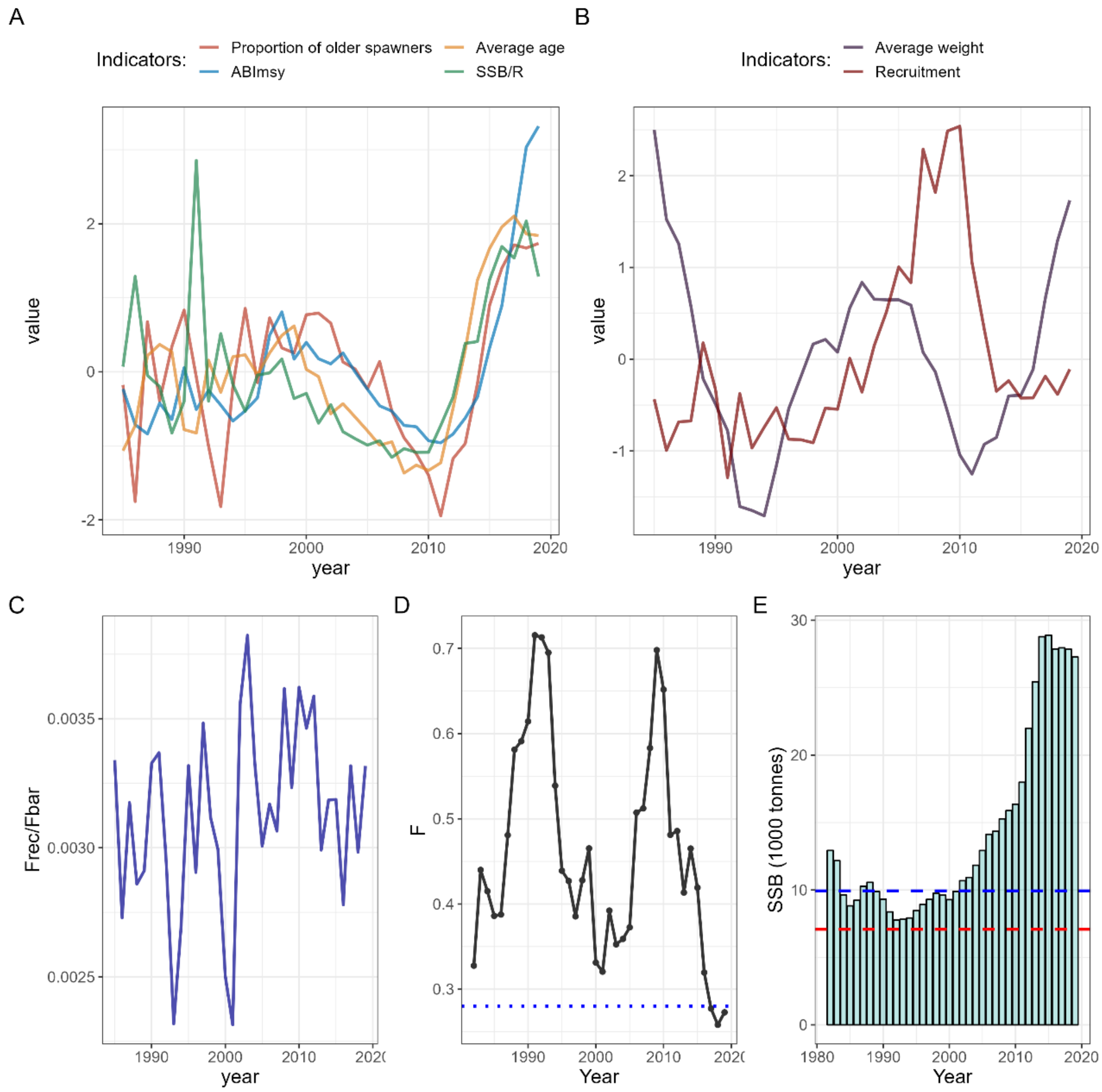

Figure A2.40. Estimated D3C3 indicators and stock status for Ling (lin.27.5a). See Figure legend A2.1 for further details.

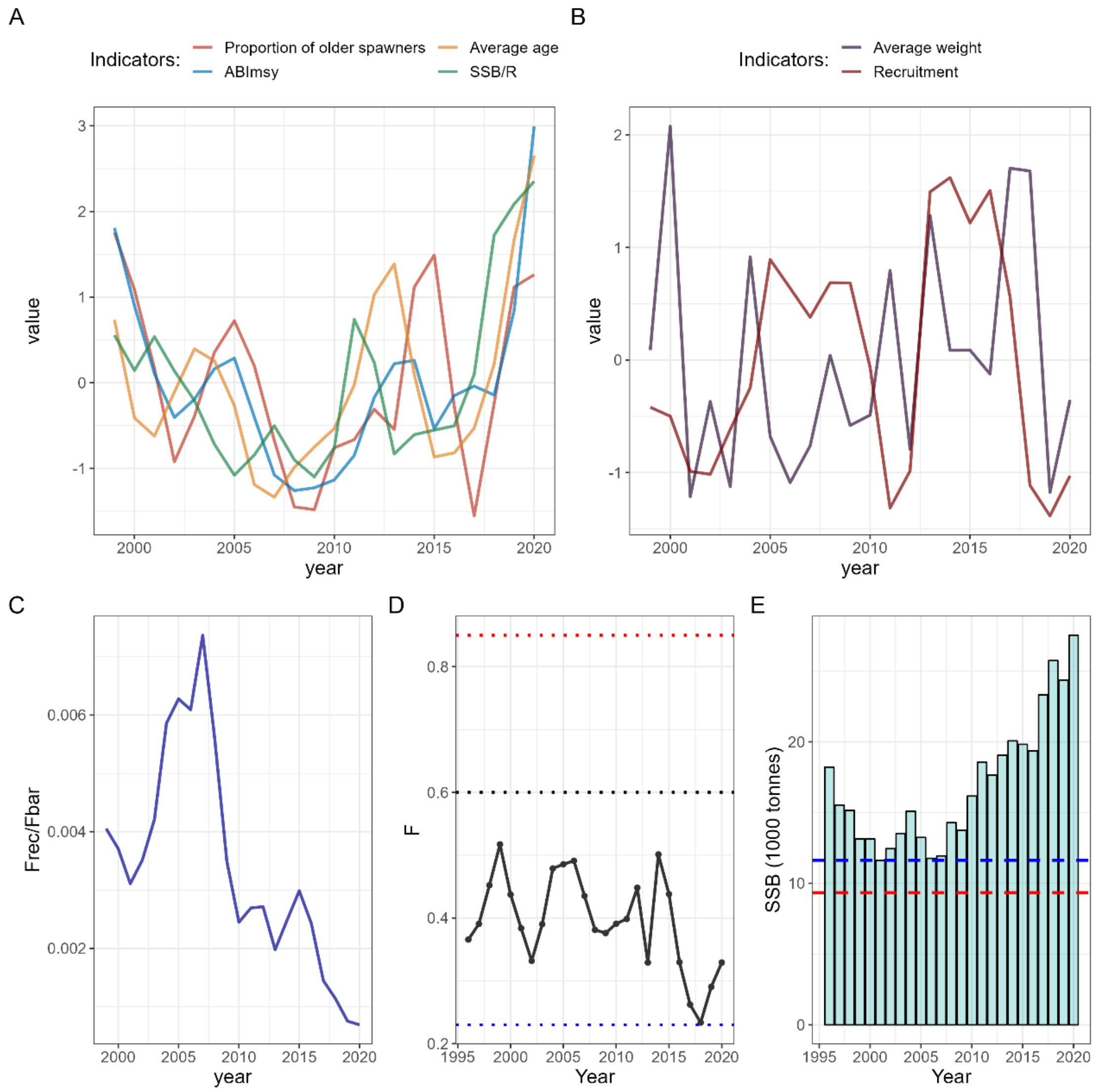

Figure A2.41. Estimated D3C3 indicators and stock status for Ling (lin.27.5b). See Figure legend A2.1 for further details.

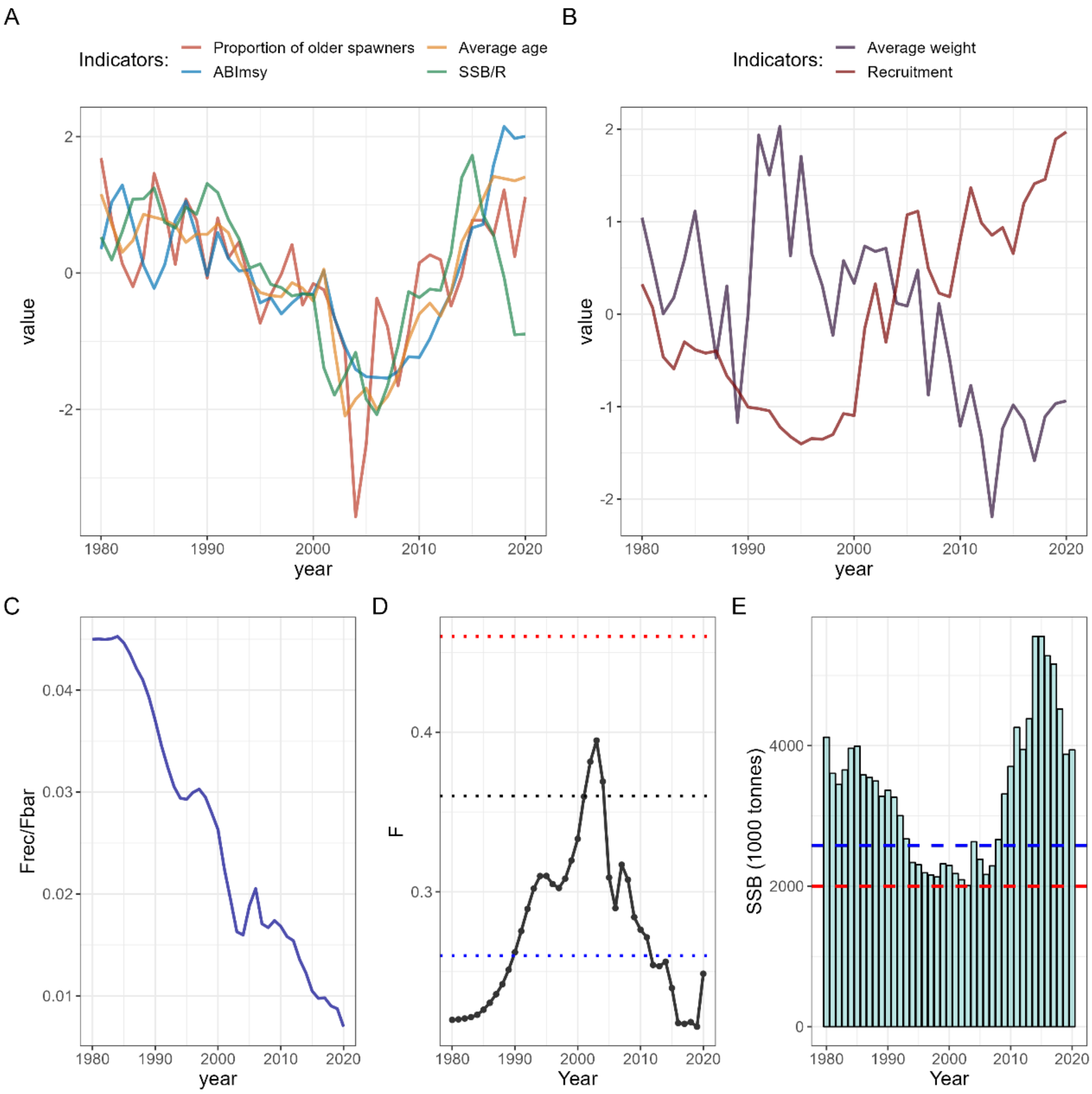

Figure A2.42. Estimated D3C3 indicators and stock status for Mackerel (mac.27.nea). See Figure legend A2.1 for further details.

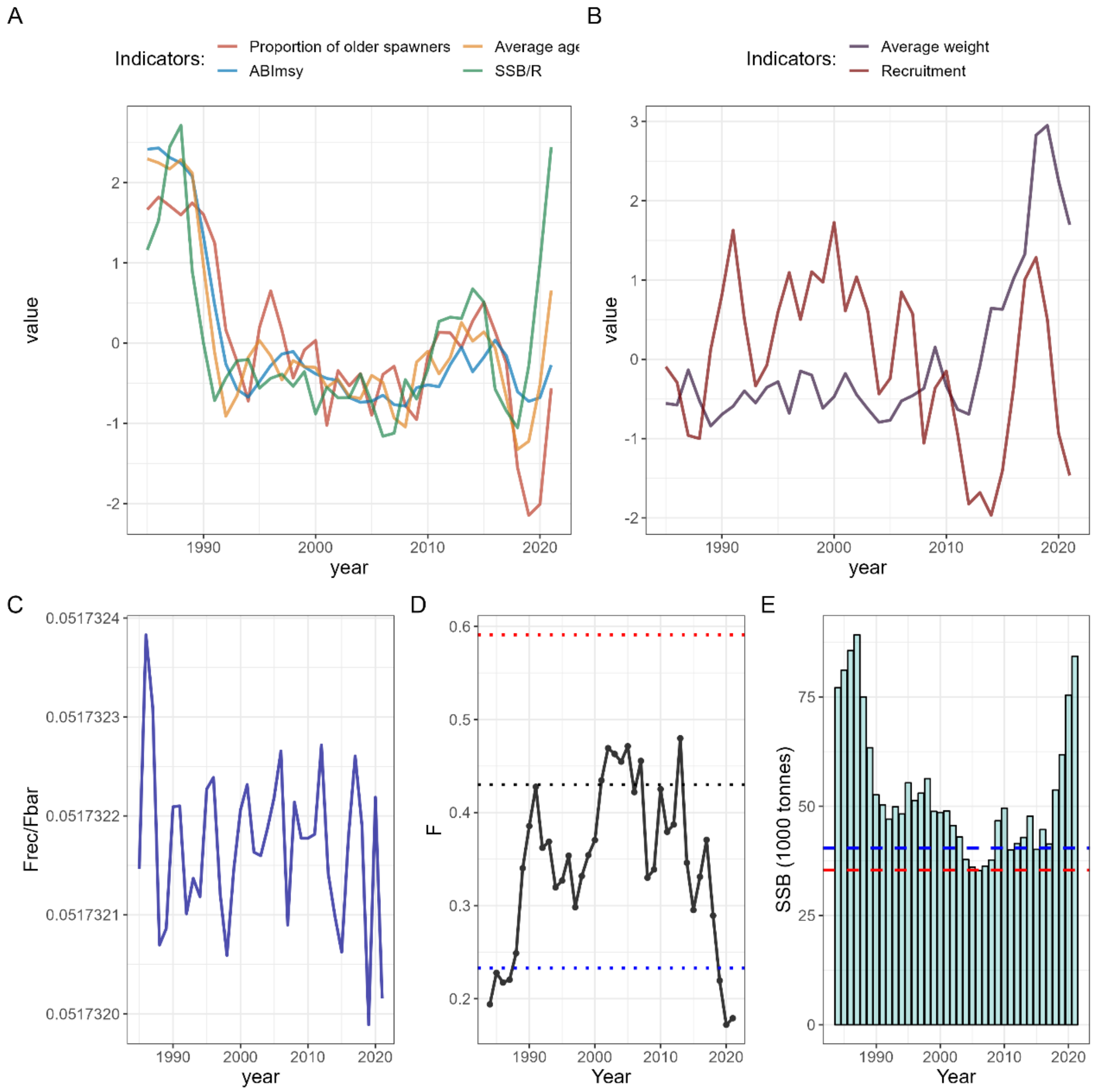

Figure A2.43. Estimated D3C3 indicators and stock status for Megrim (meg.27.7b-k8abd). See Figure legend A2.1 for further details.

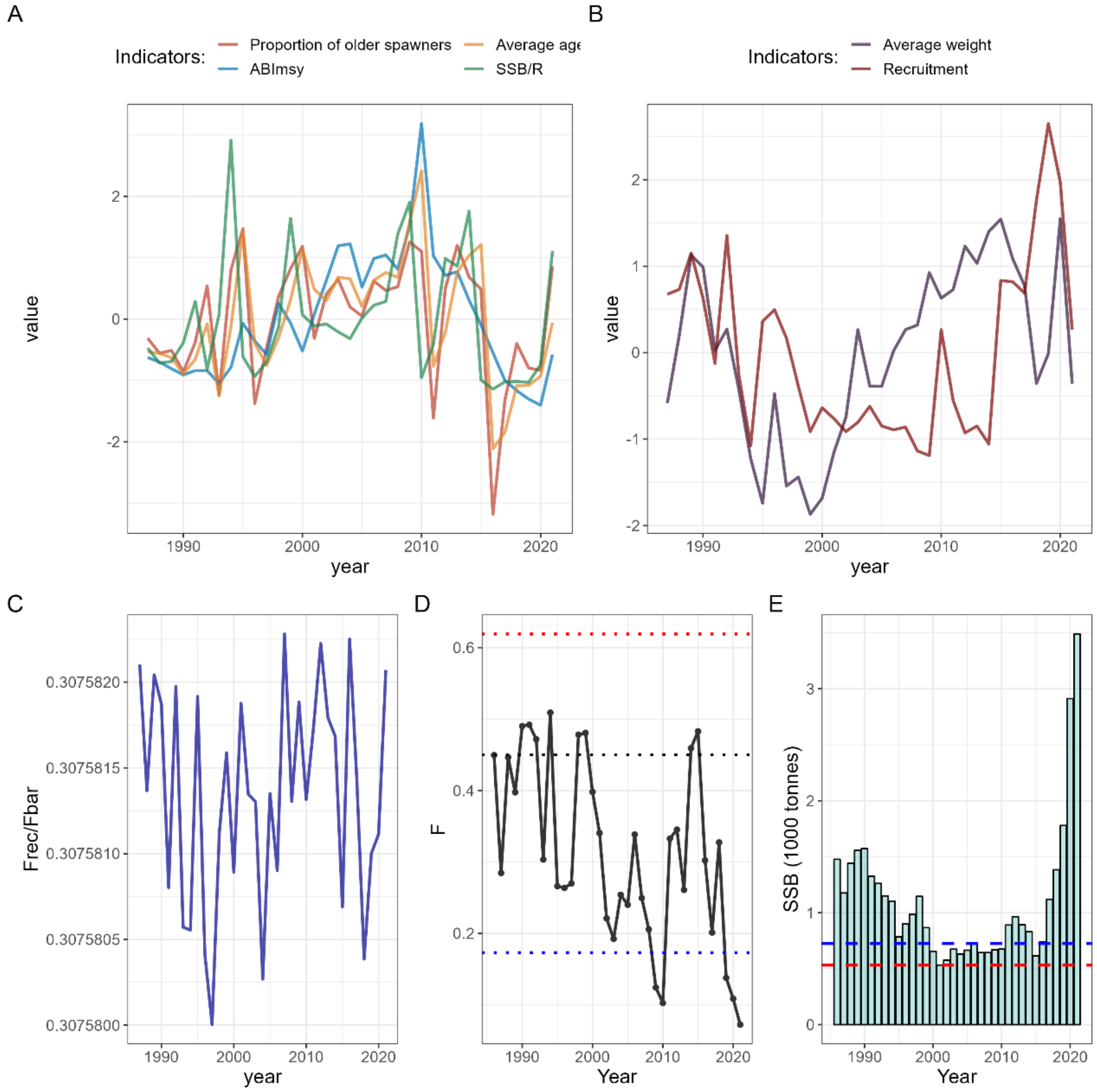

Figure A2.44. Estimated D3C3 indicators and stock status for Megrim (meg.27.8c9a). See Figure legend A2.1 for further details.

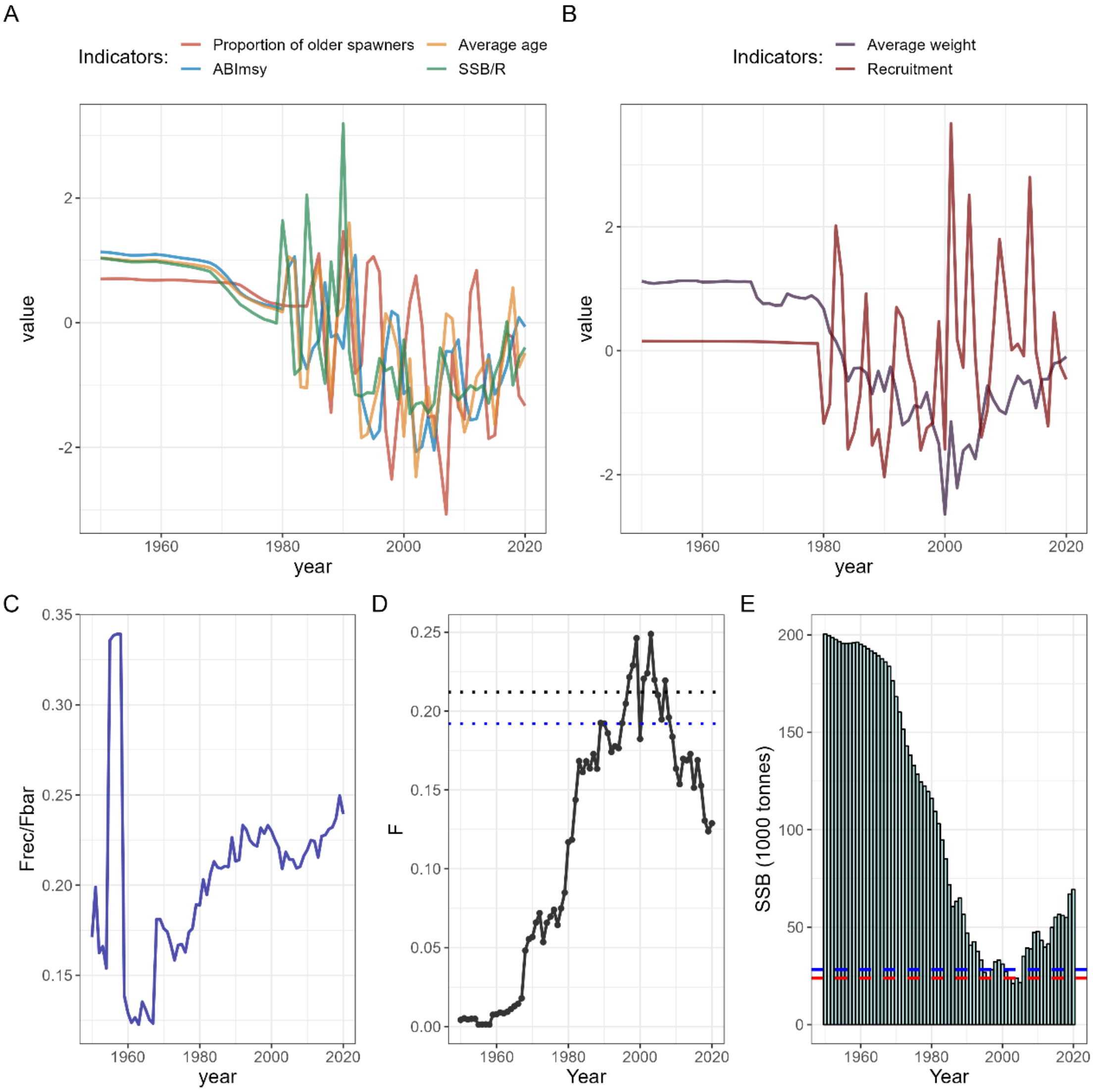

Figure A2.45. Estimated D3C3 indicators and stock status for White anglerfish (mon.27.78abd). See Figure legend A2.1 for further details.

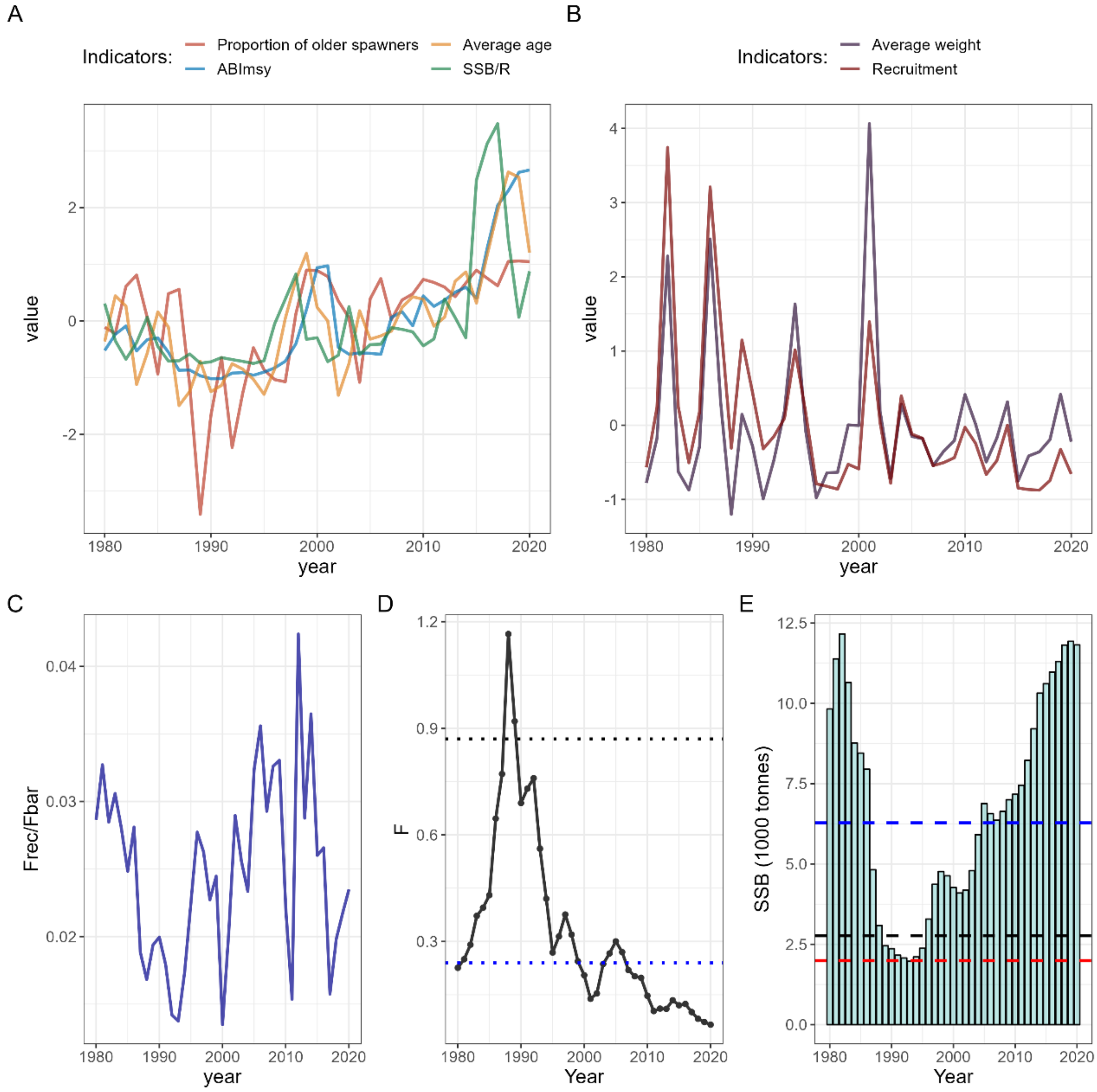

Figure A2.46. Estimated D3C3 indicators and stock status for White anglerfish (mon.27.8c9a). See Figure legend A2.1 for further details.

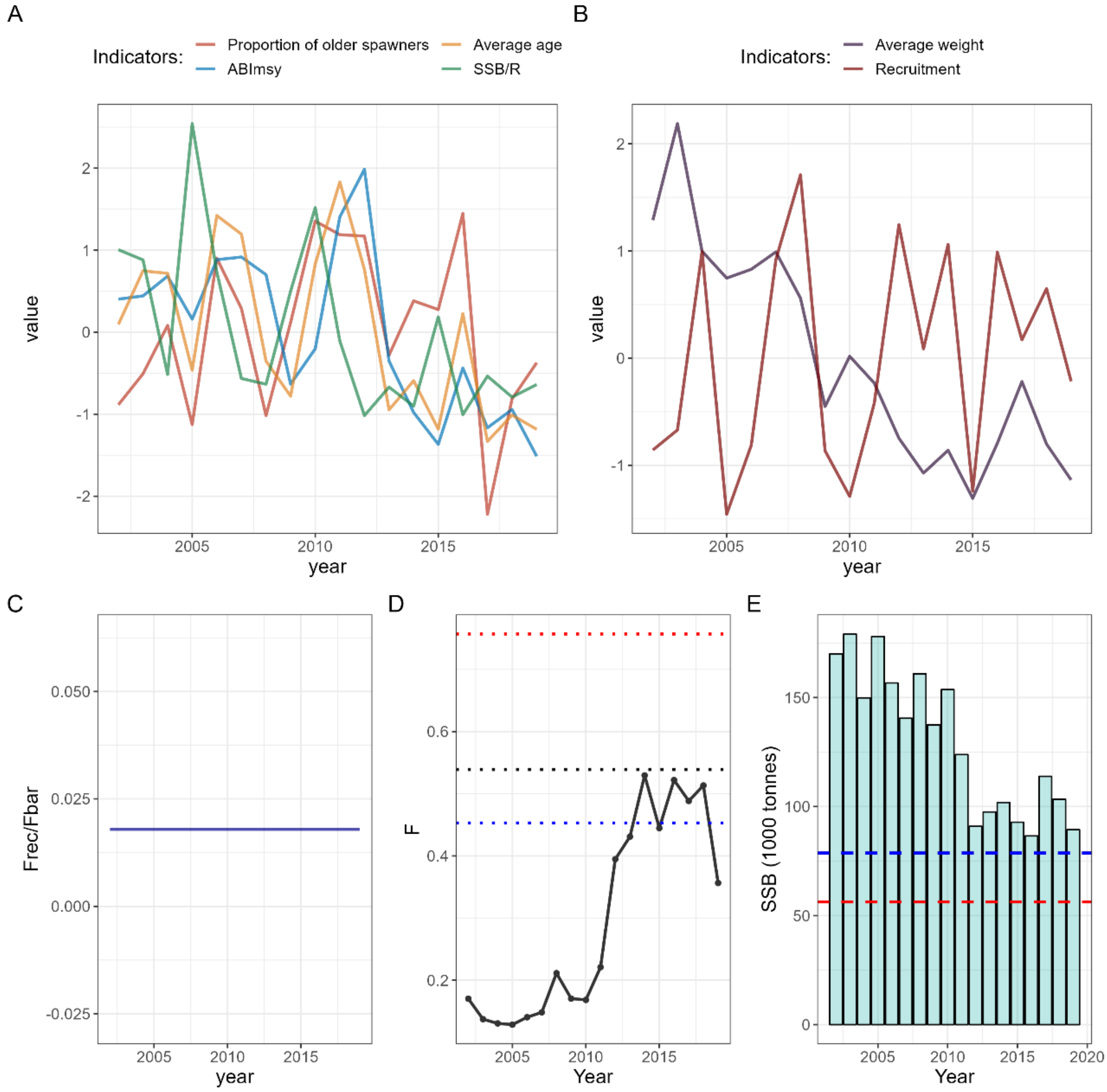

Figure A2.47. Estimated D3C3 indicators and stock status for Sardine (pil.27.8abd). See Figure legend A2.1 for further details.

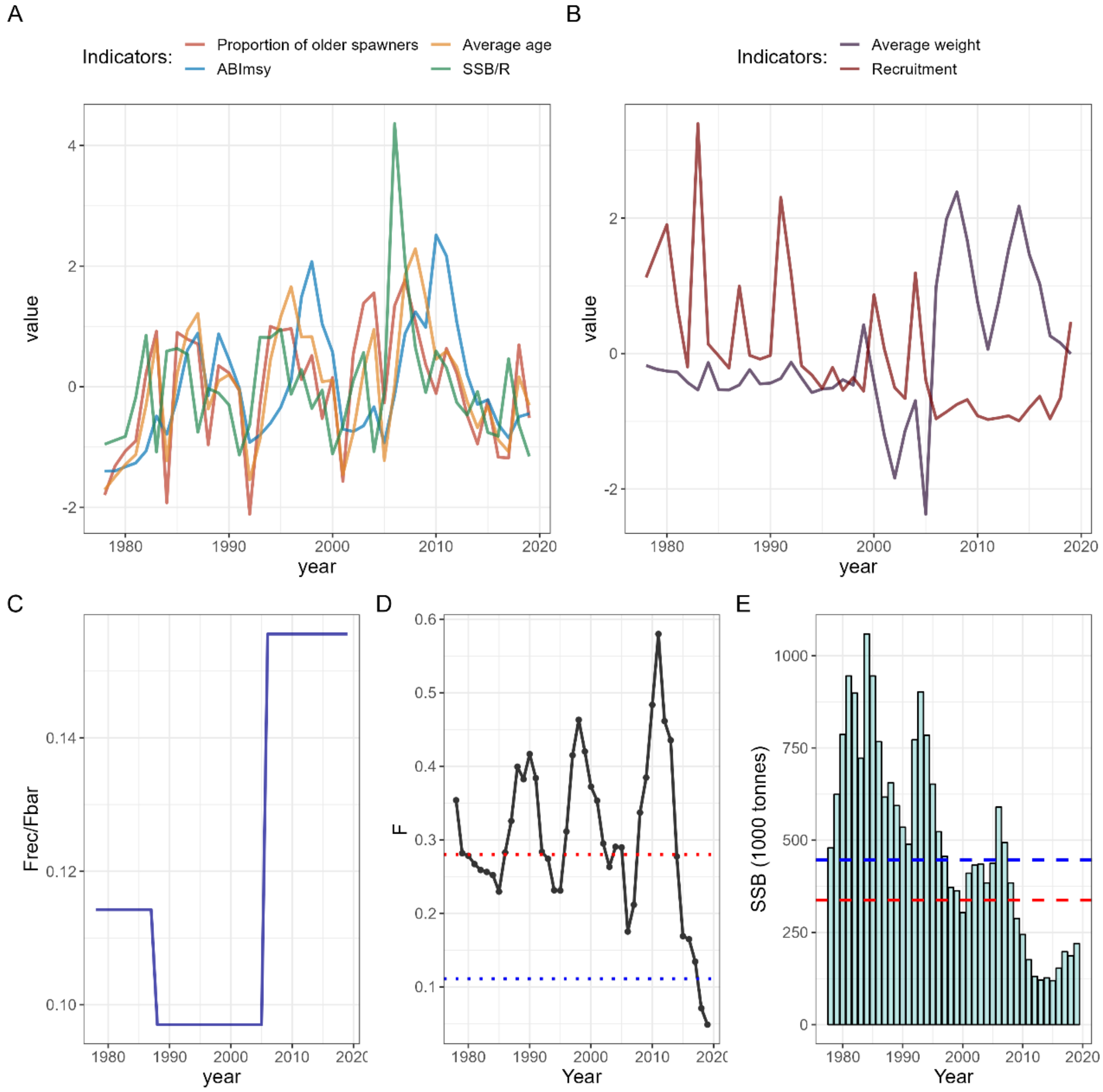

Figure A2.48. Estimated D3C3 indicators and stock status for Sardine (pil.27.8c9a). See Figure legend A2.1 for further details.

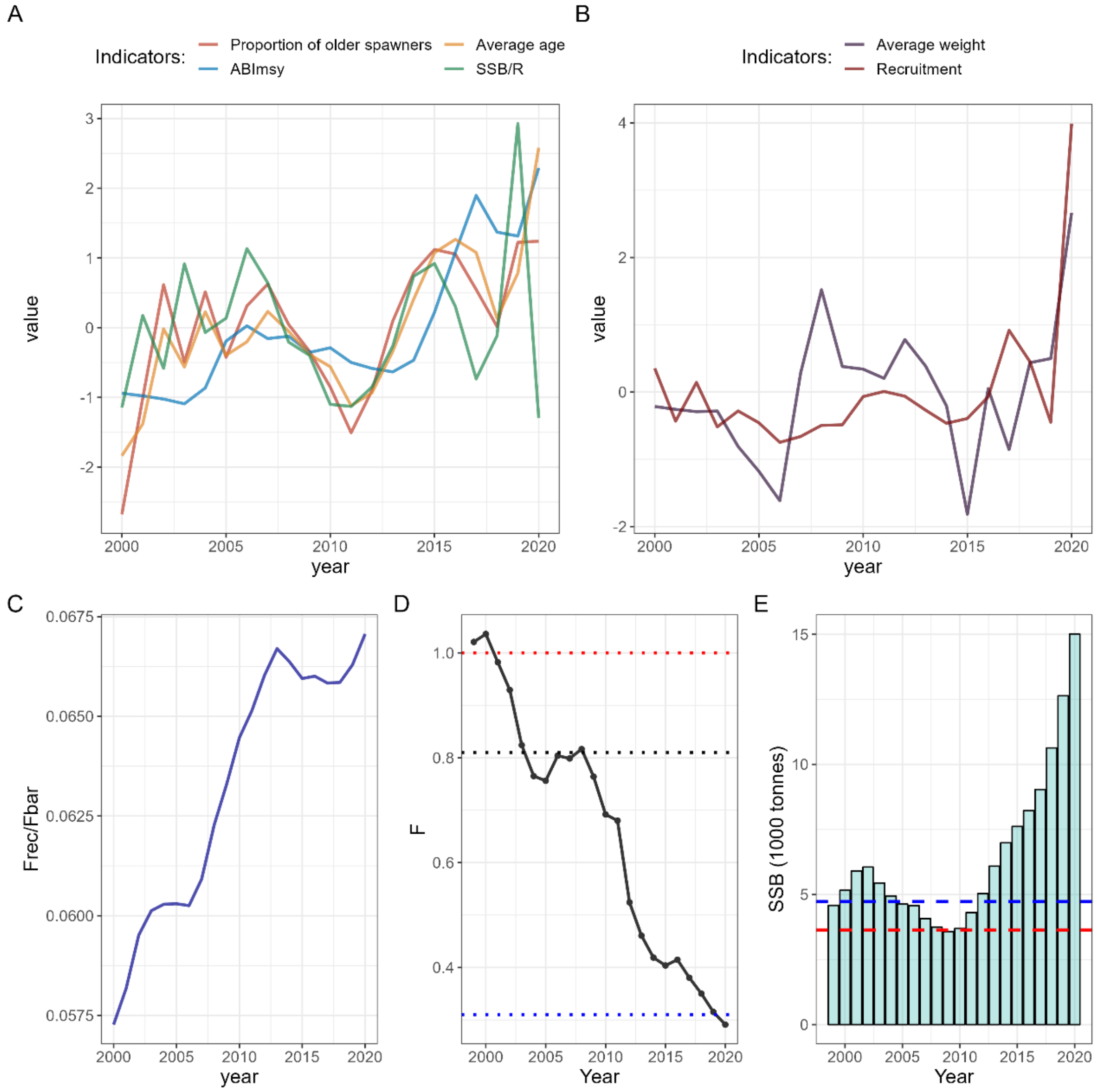

Figure A2.49. Estimated D3C3 indicators and stock status for European plaice (ple.27.21-23). See Figure legend A2.1 for further details.

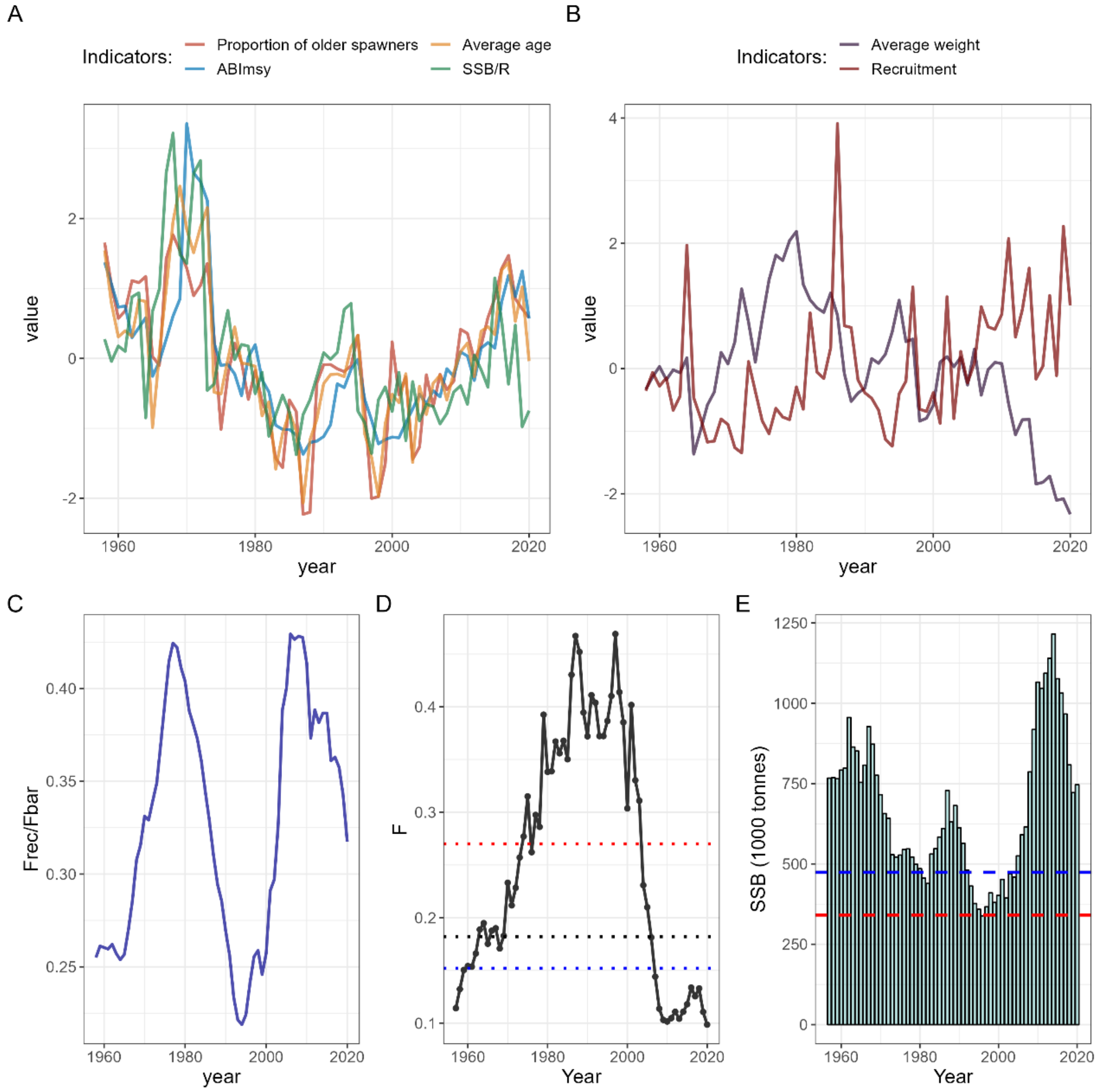

Figure A2.50. Estimated D3C3 indicators and stock status for European plaice (ple.27.420). See Figure legend A2.1 for further details.

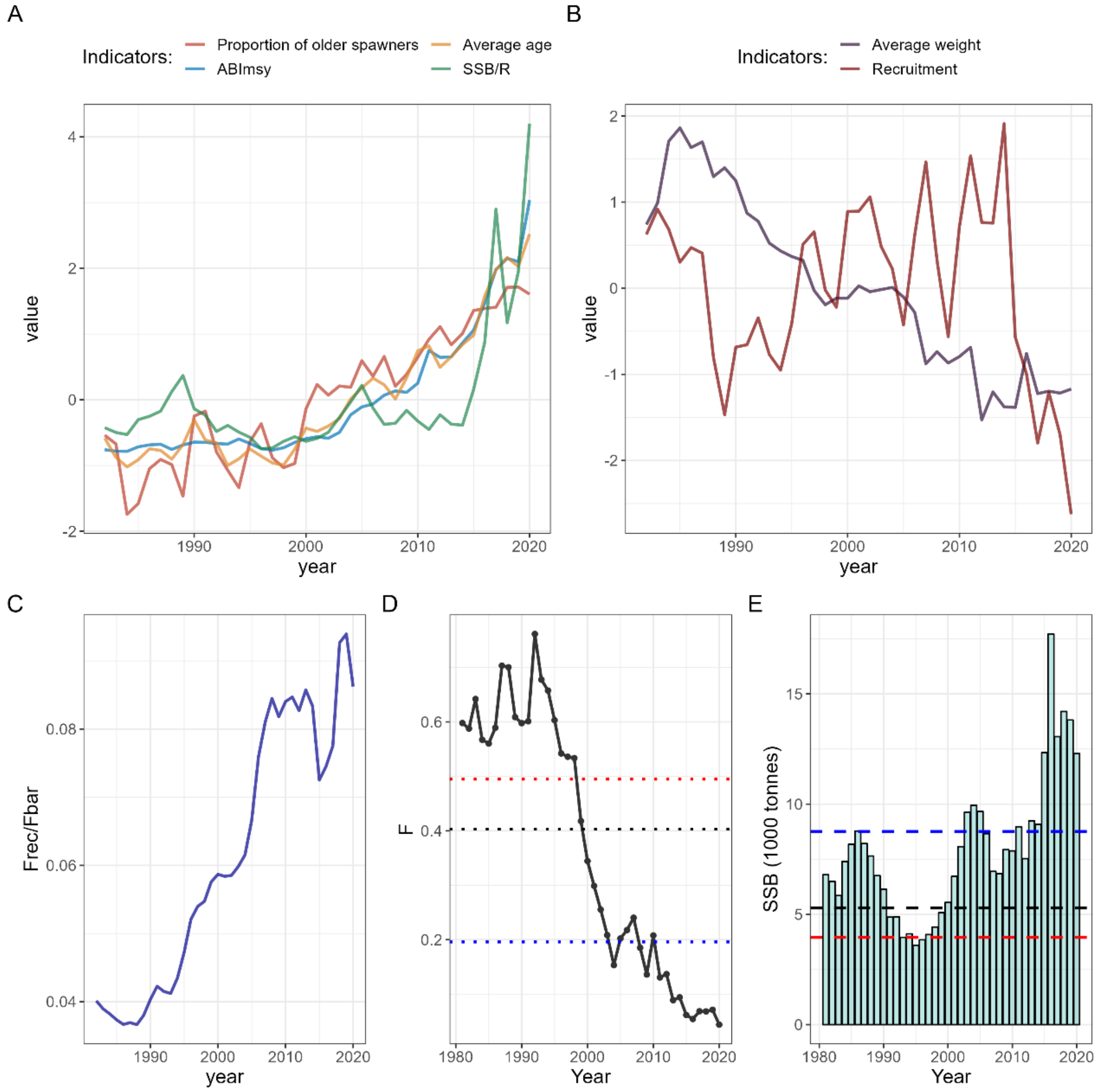

Figure A2.51. Estimated D3C3 indicators and stock status for European plaice (ple.27.7a). See Figure legend A2.1 for further details.

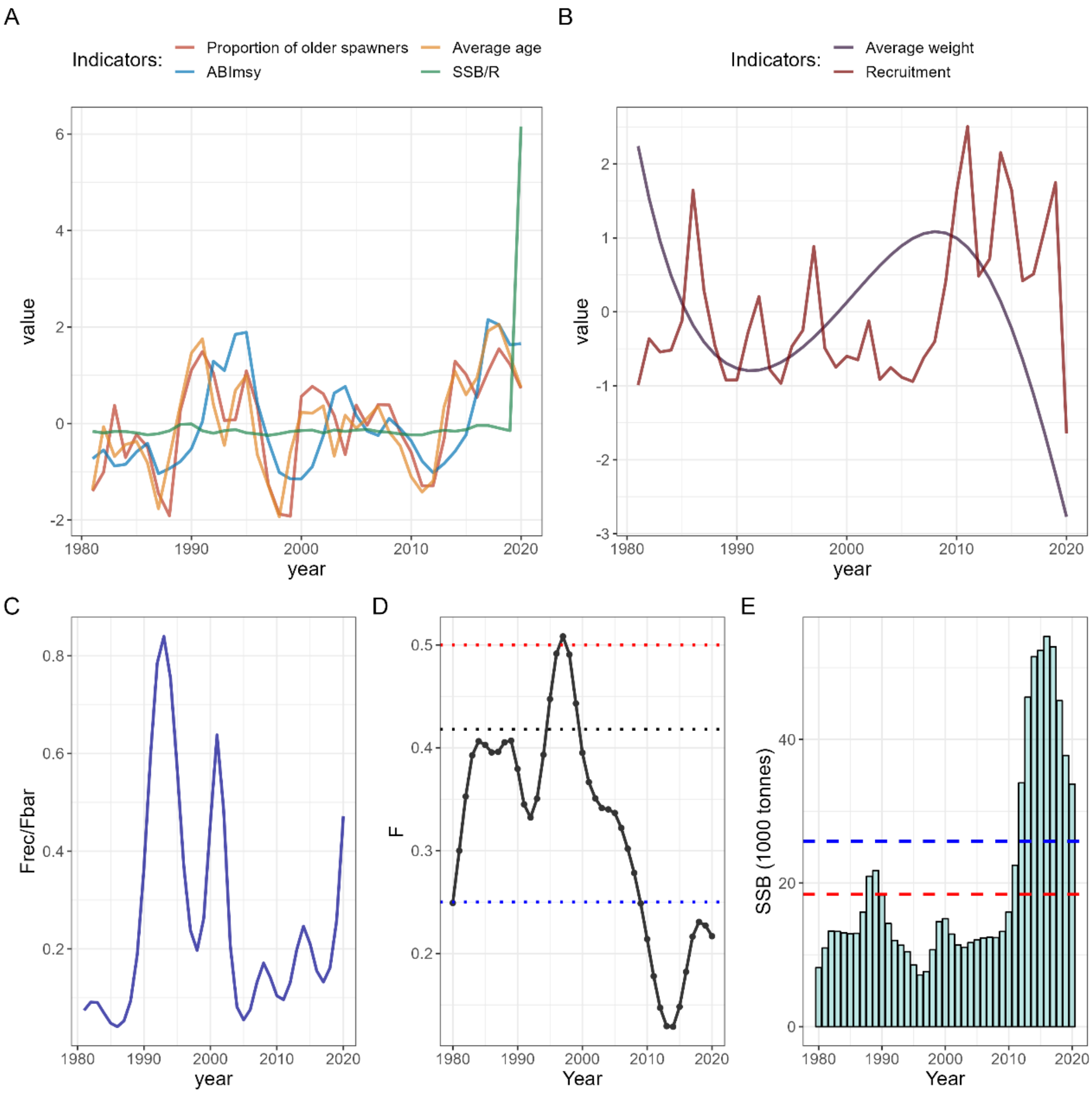

Figure A2.52. Estimated D3C3 indicators and stock status for European plaice (ple.27.7d). See Figure legend A2.1 for further details.

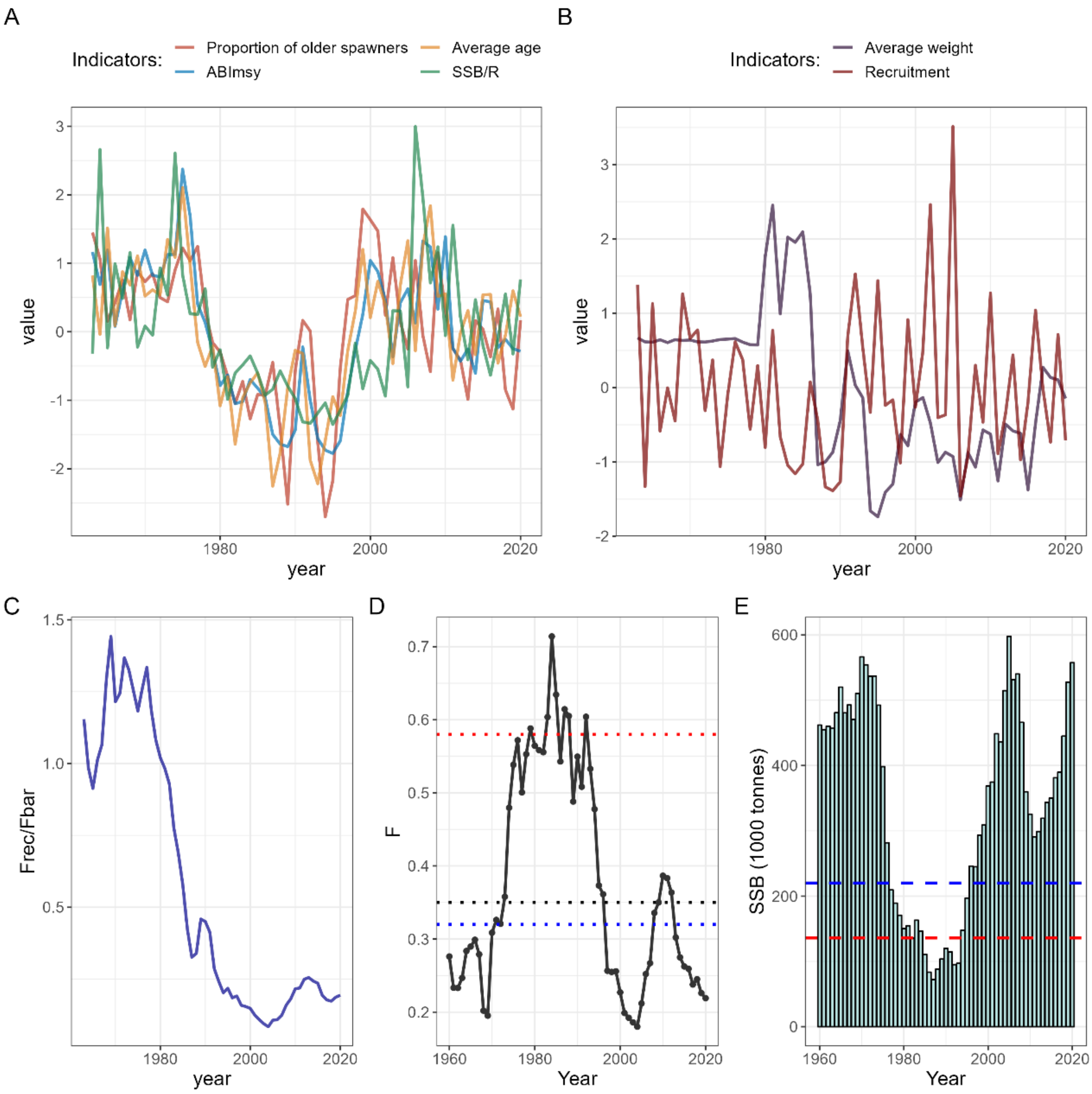

Figure A2.53. Estimated D3C3 indicators and stock status for Saithe (pok.27.1-2). See Figure legend A2.1 for further details.

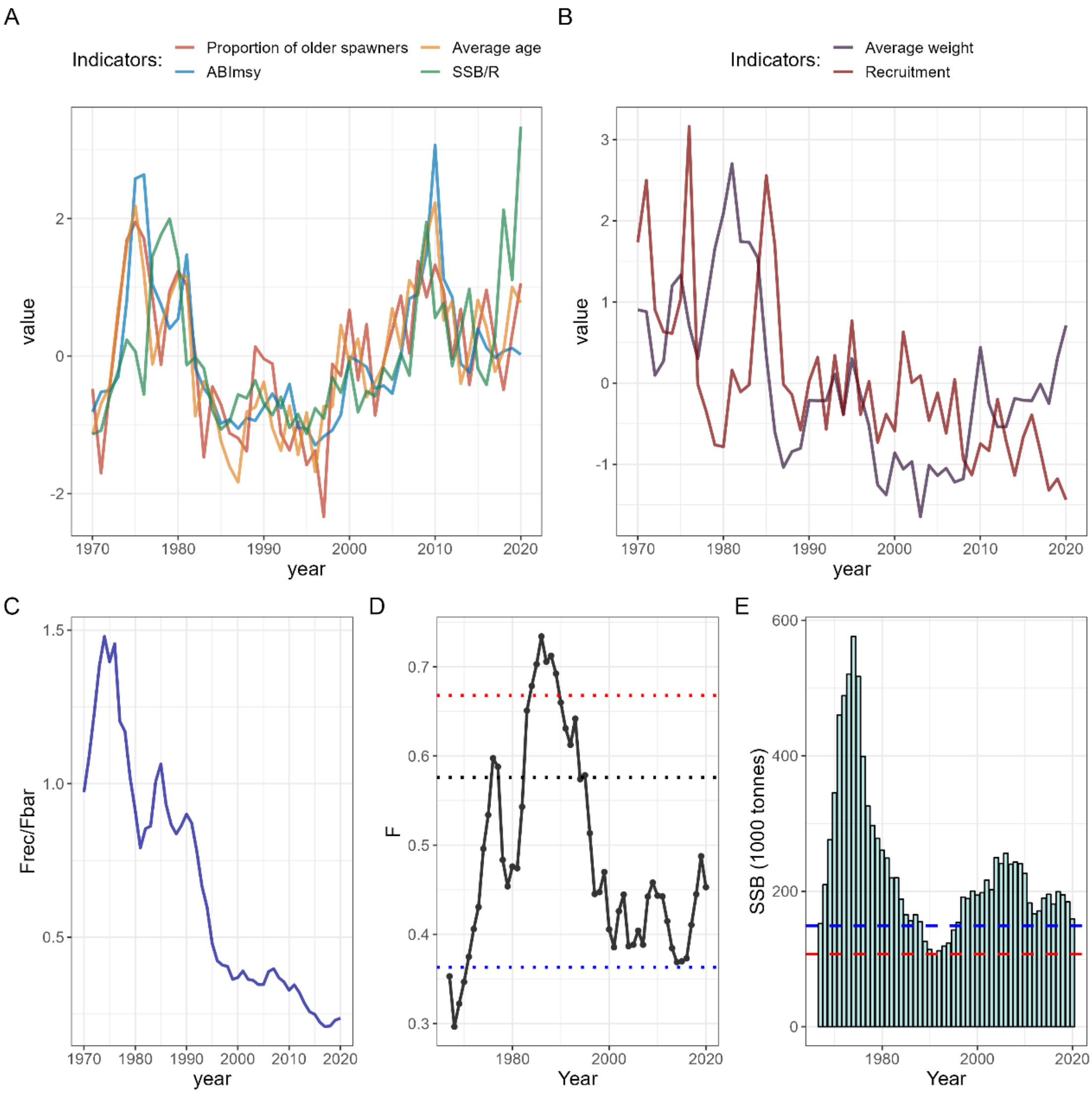

Figure A2.54. Estimated D3C3 indicators and stock status for Saithe (pok.27.3a46). See Figure legend A2.1 for further details.

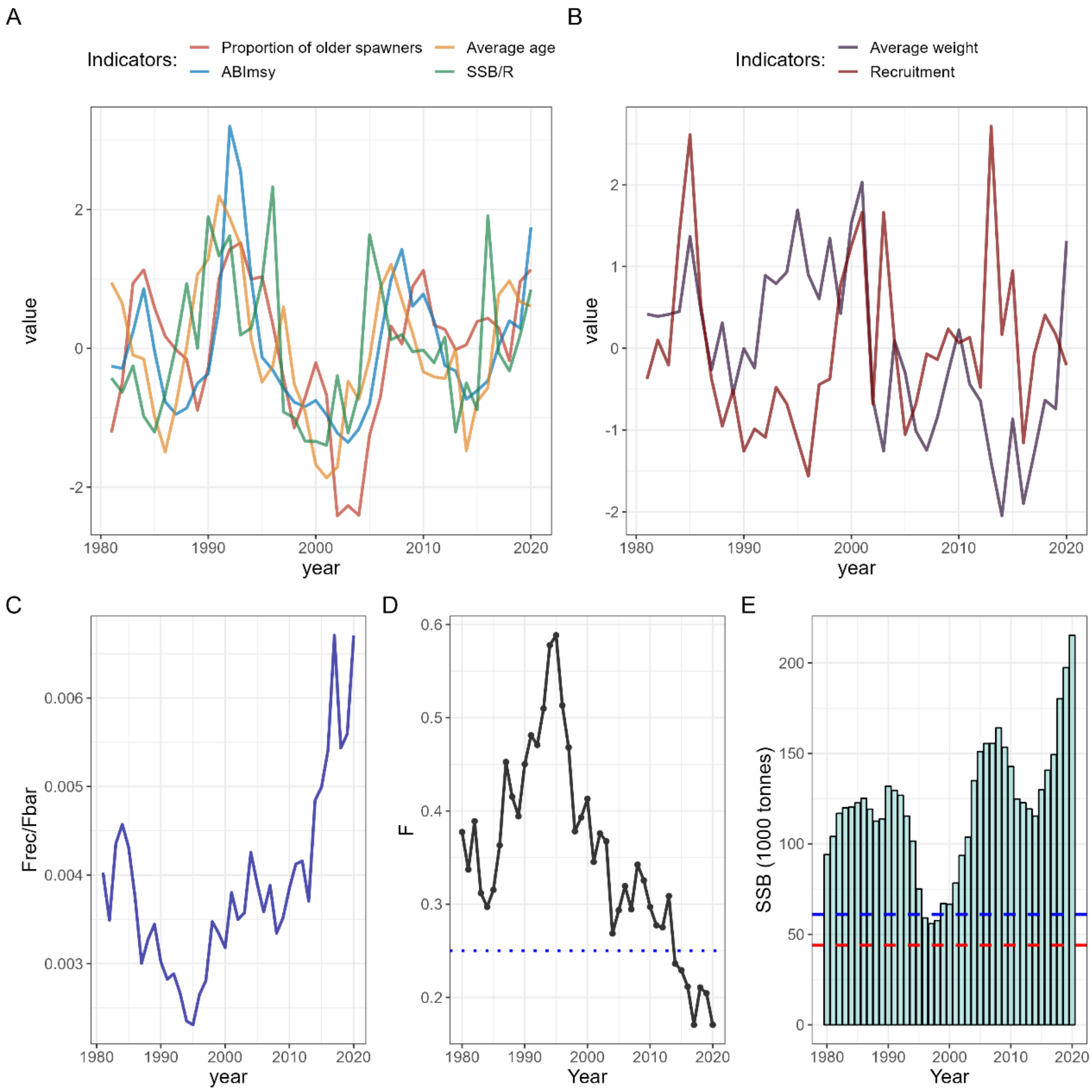

Figure A2.55. Estimated D3C3 indicators and stock status for Saithe (pok.27.5a). See Figure legend A2.1 for further details.

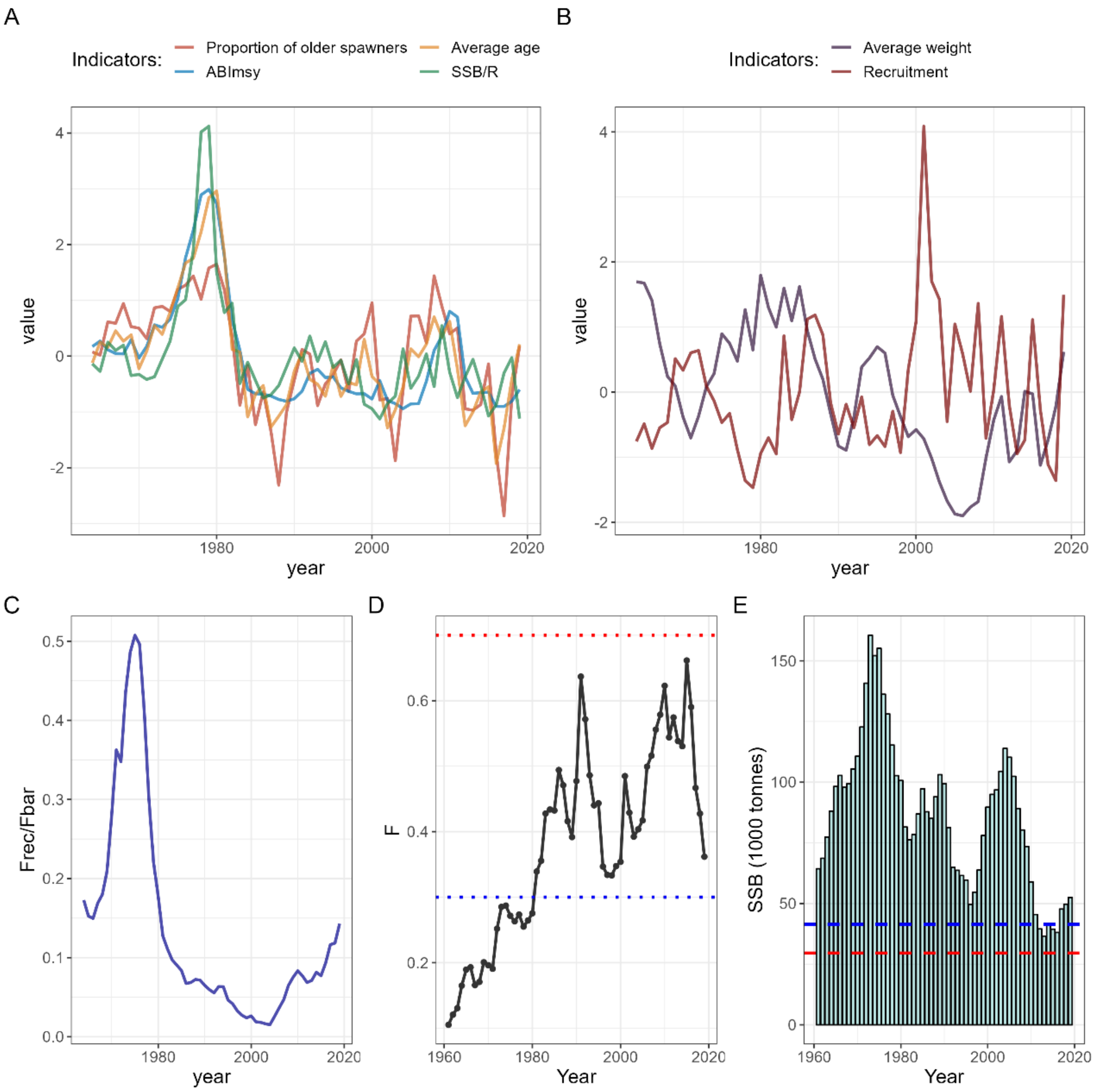

Figure A2.56. Estimated D3C3 indicators and stock status for Saithe (pok.27.5b). See Figure legend A2.1 for further details.

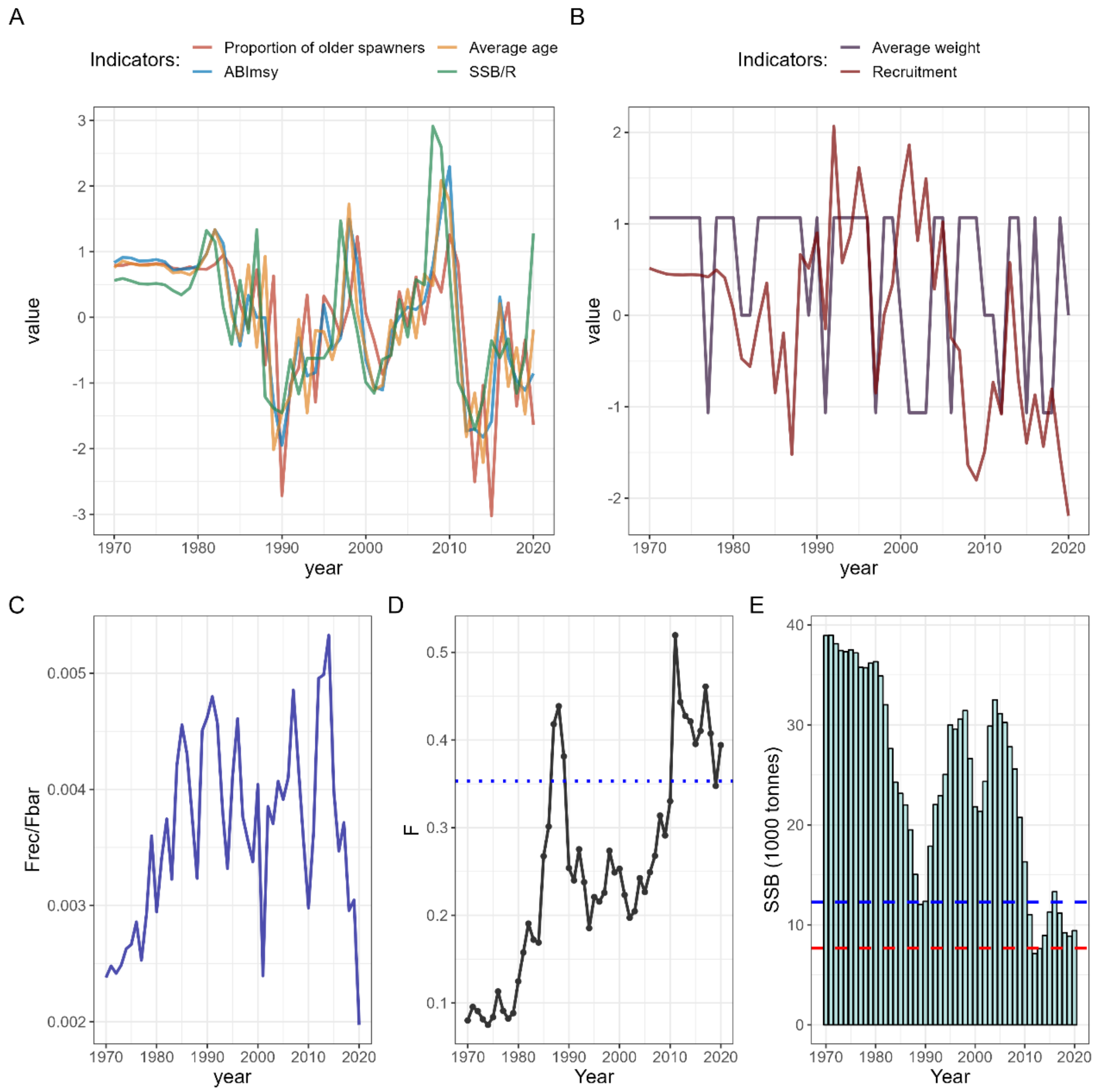

Figure A2.57. Estimated D3C3 indicators and stock status for Northern shrimp (pra.27.3a4a). See Figure legend A2.1 for further details.

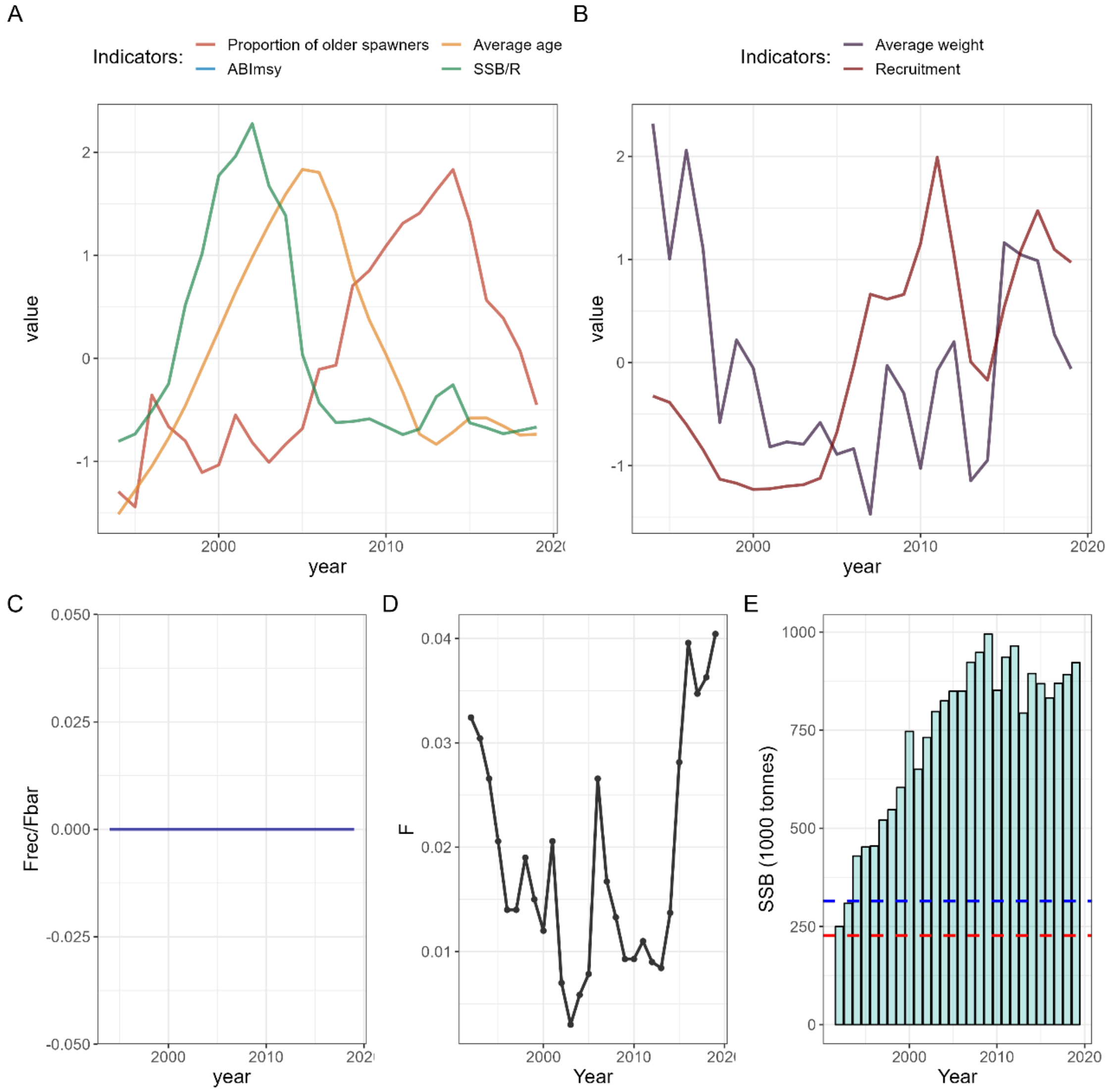

Figure A2.58. Estimated D3C3 indicators and stock status for Beaked redfish (reb.27.1-2). See Figure legend A2.1 for further details.

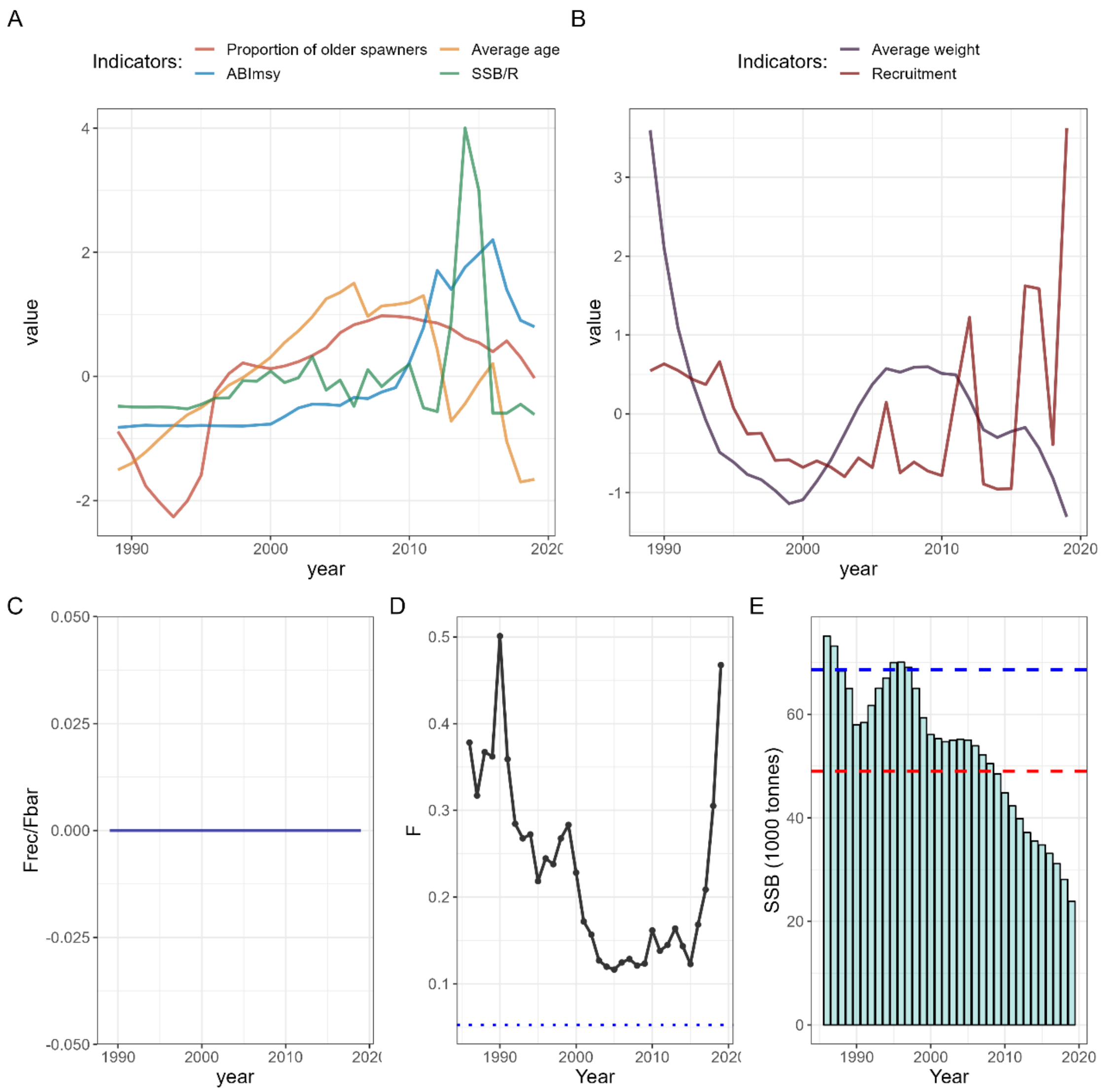

Figure A2.59. Estimated D3C3 indicators and stock status for Golden redfish (reg.27.1-2). See Figure legend A2.1 for further details.

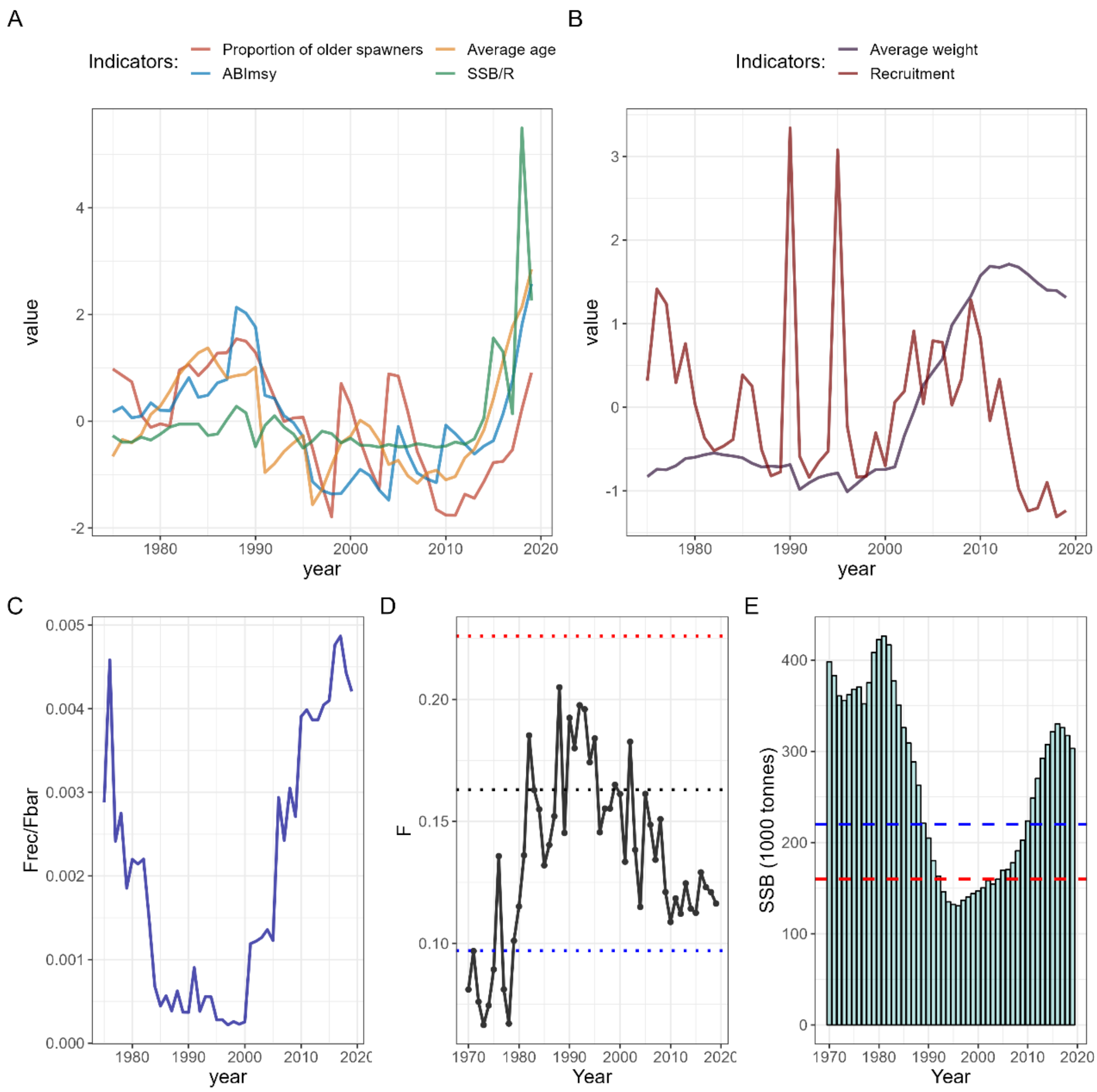

Figure A2.60. Estimated D3C3 indicators and stock status for Golden redfish (reg.27.561214). See Figure legend A2.1 for further details.

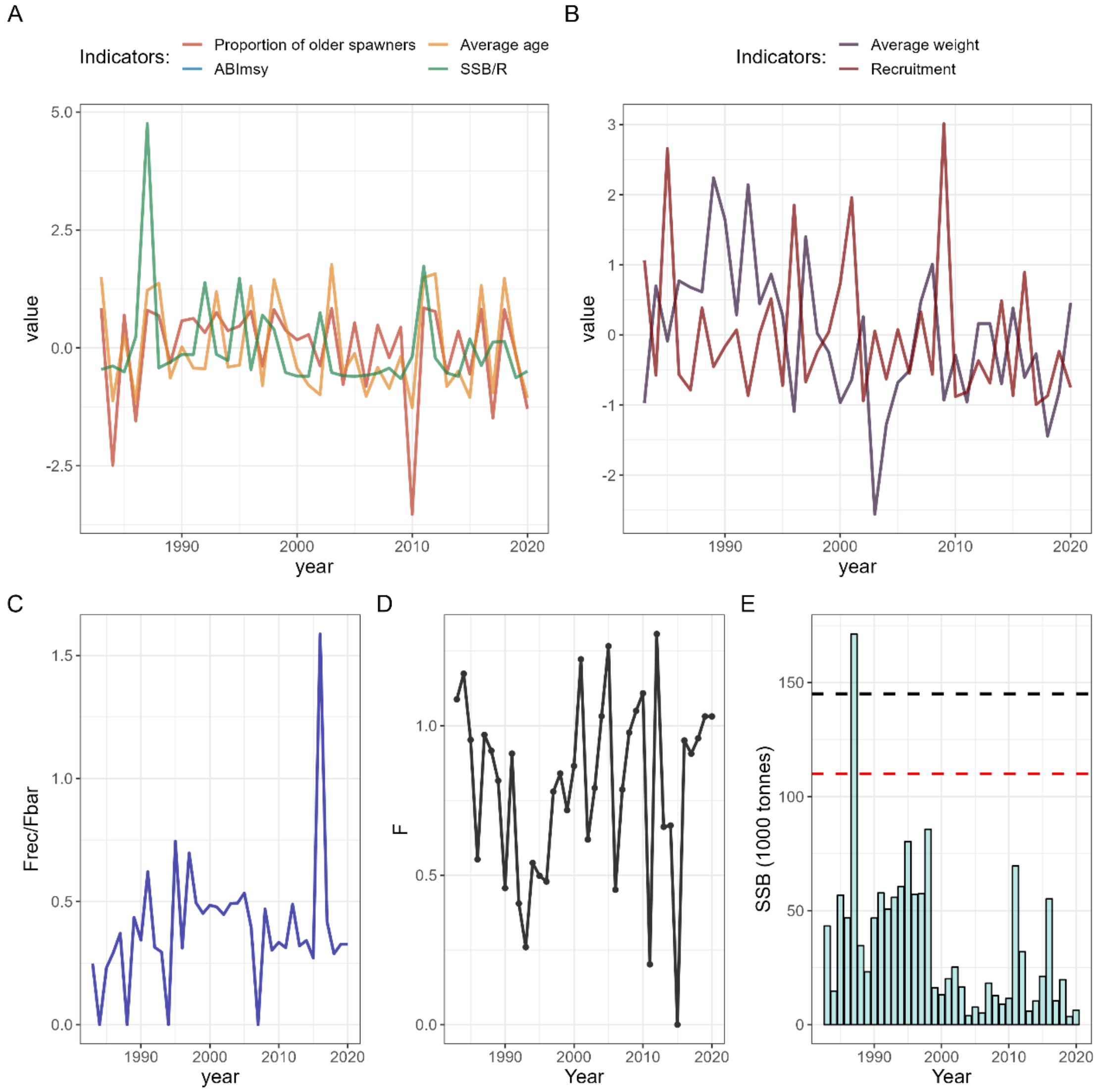

Figure A2.61. Estimated D3C3 indicators and stock status for Sandeel (san.sa.1r). See Figure legend A2.1 for further details.

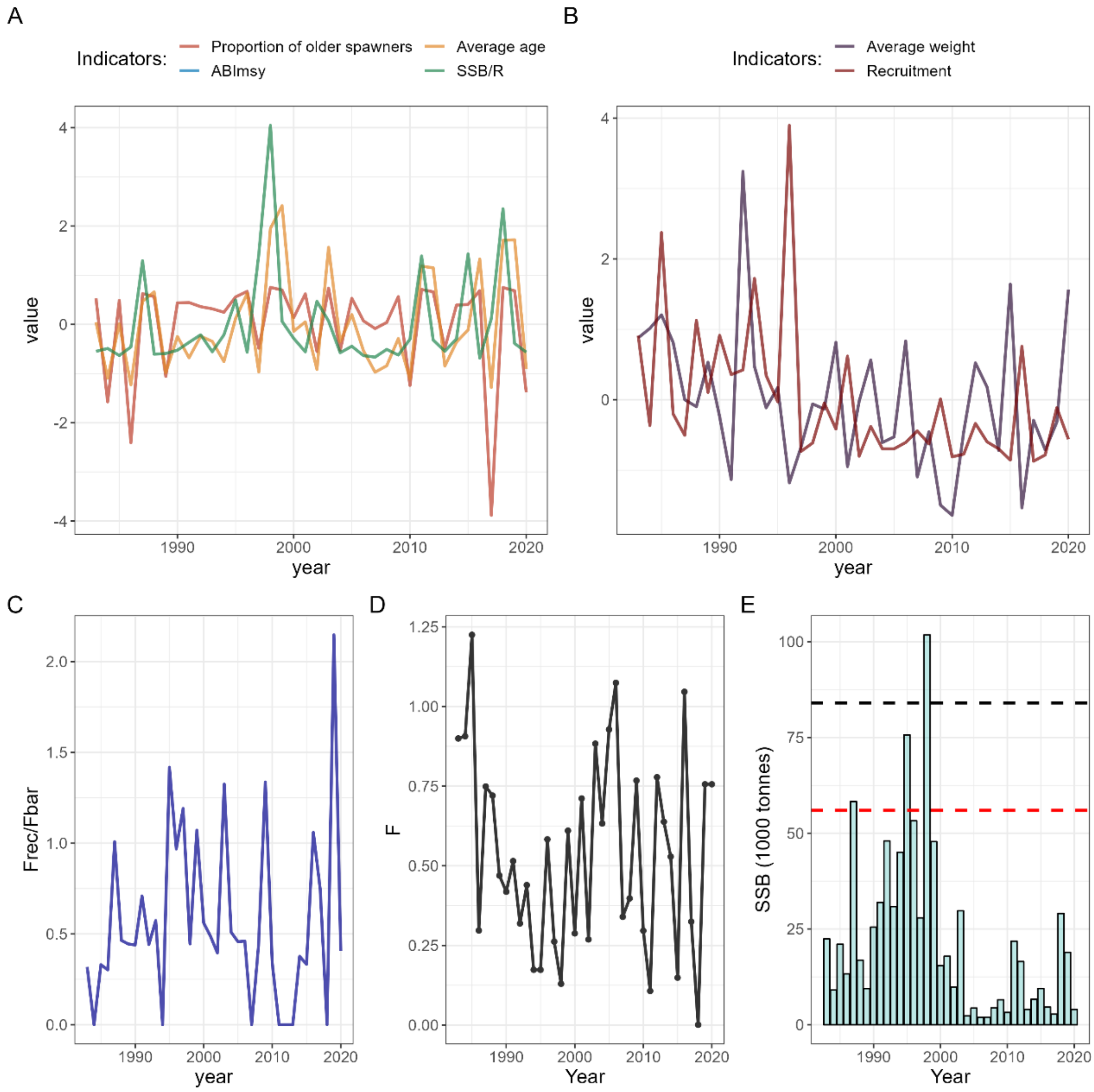

Figure A2.62. Estimated D3C3 indicators and stock status for Sandeel (san.sa.2r). See Figure legend A2.1 for further details.

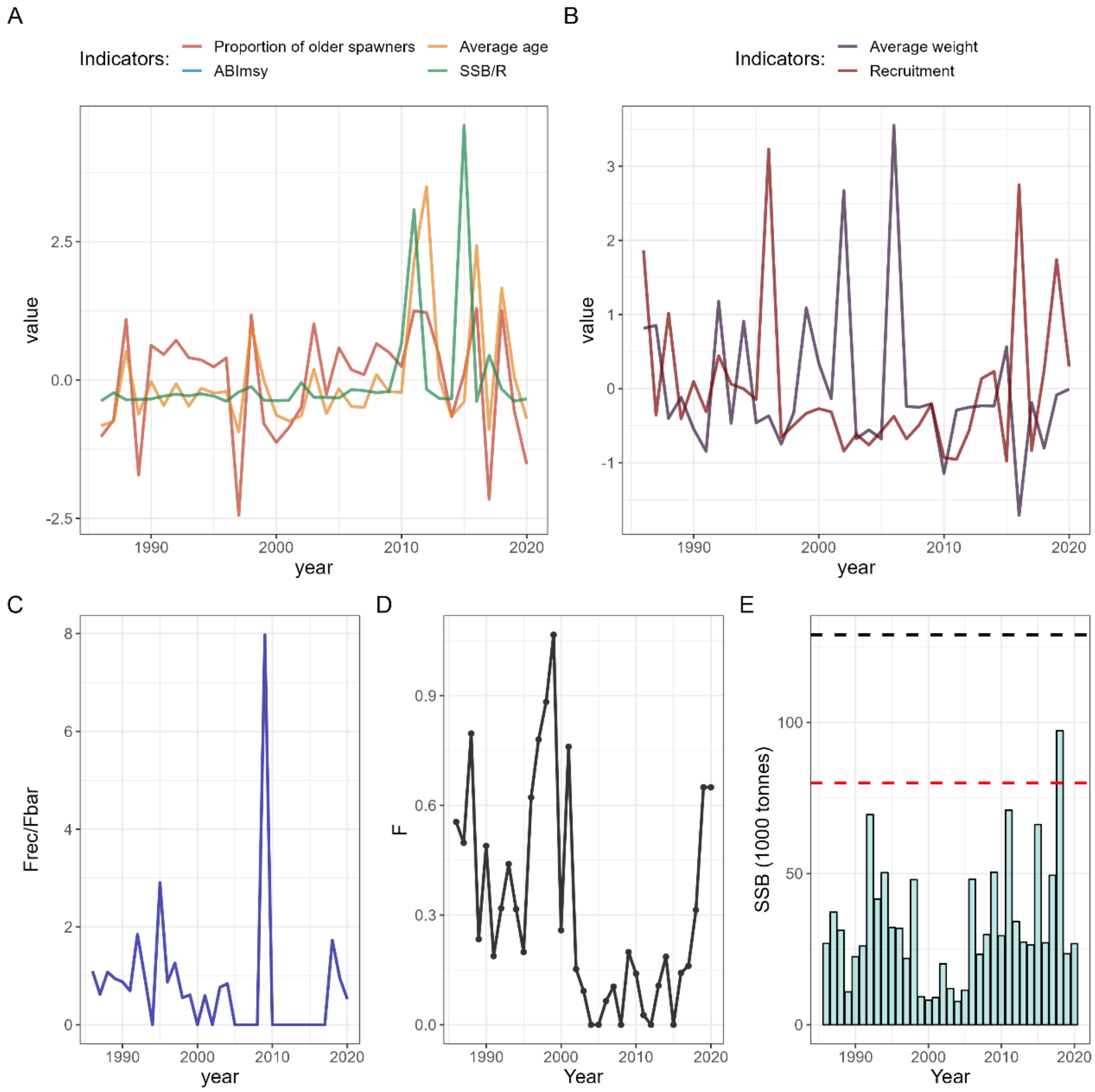

Figure A2.63. Estimated D3C3 indicators and stock status for Sandeel (san.sa.3r). See Figure legend A2.1 for further details.

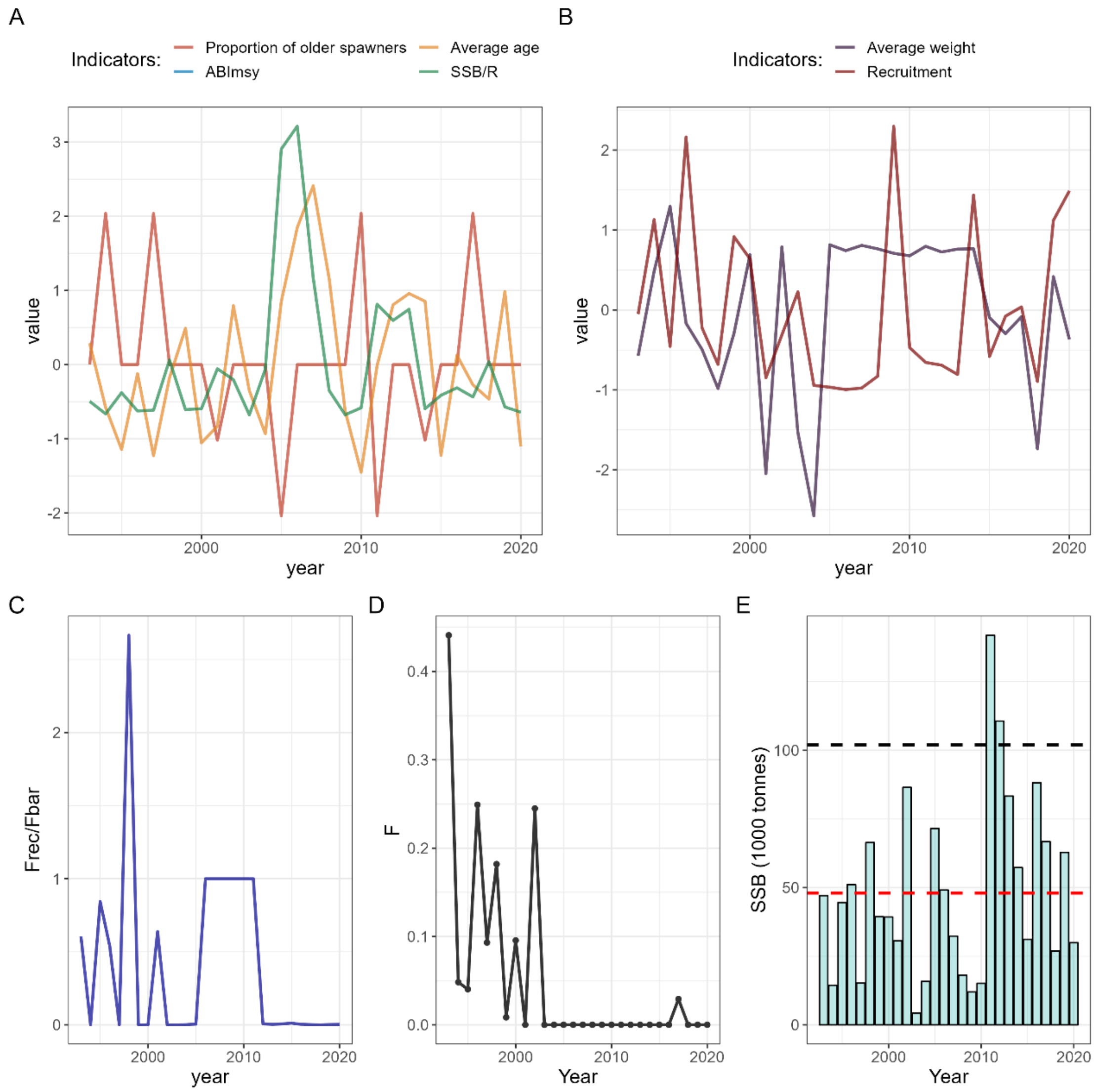

Figure A2.64. Estimated D3C3 indicators and stock status for Sandeel (san.sa.4r). See Figure legend A2.1 for further details.

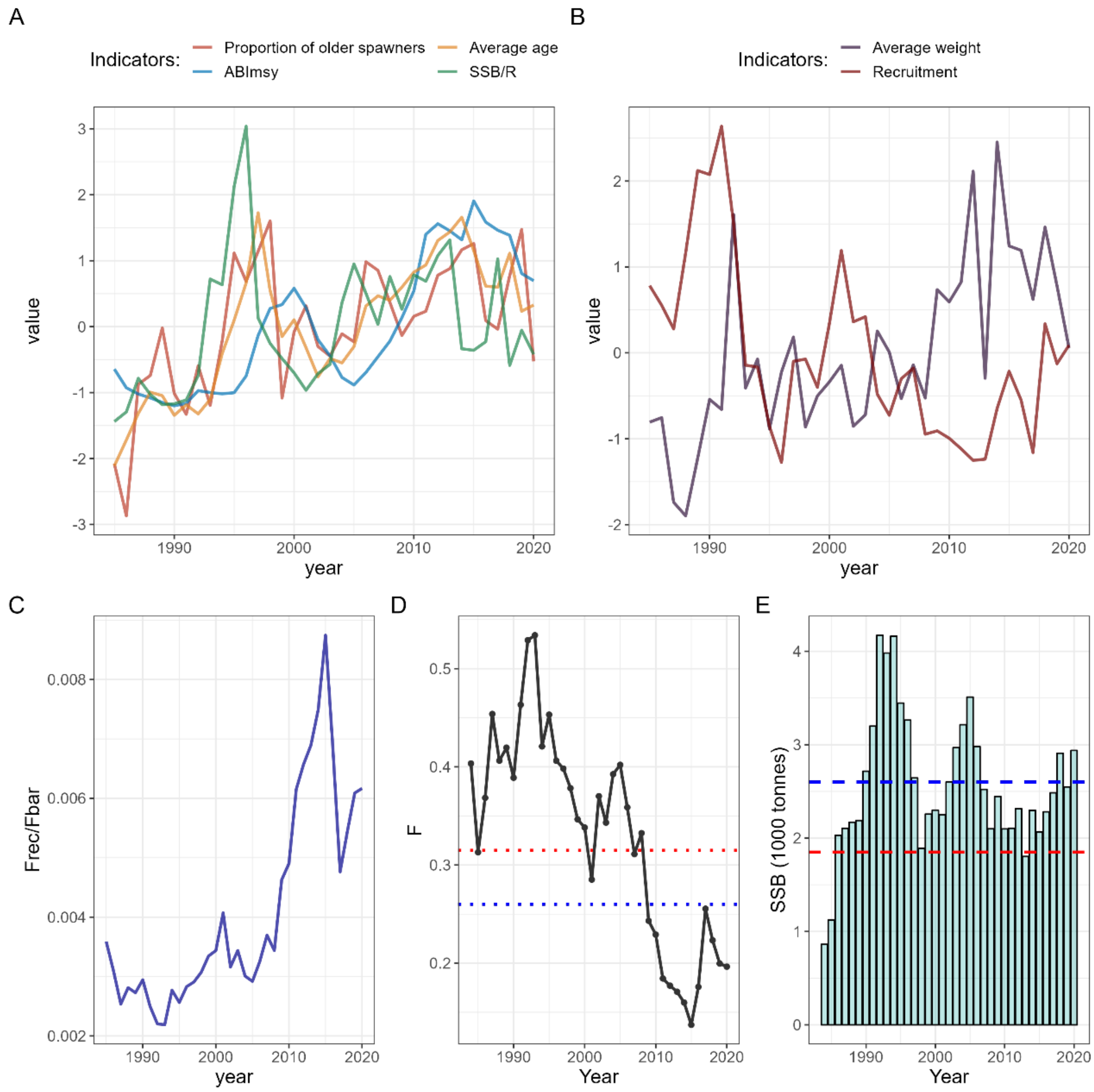

Figure A2.65. Estimated D3C3 indicators and stock status for Sole (sol.27.20-24). See Figure legend A2.1 for further details.

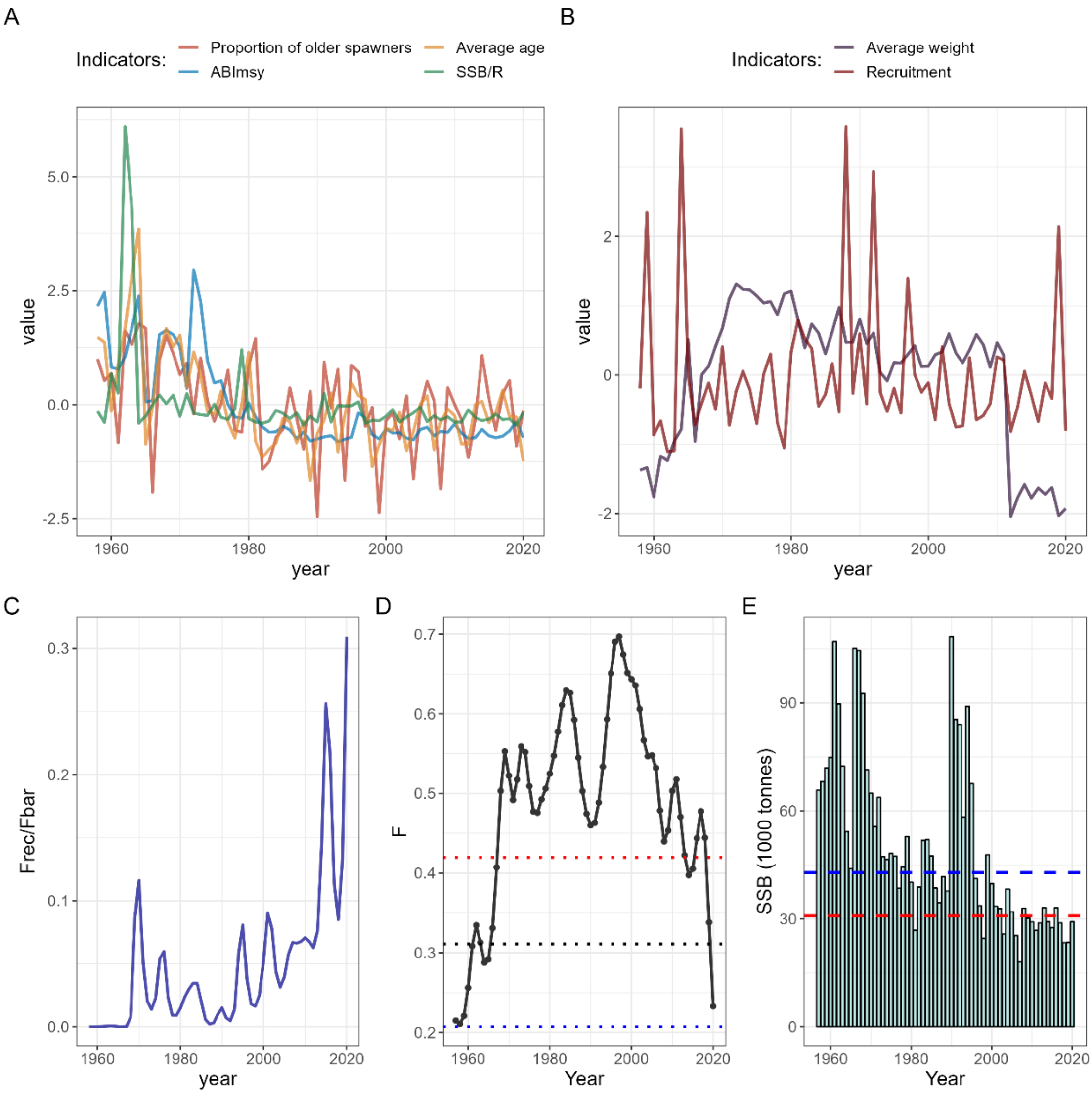

Figure A2.66. Estimated D3C3 indicators and stock status for Sole (sol.27.4). See Figure legend A2.1 for further details.

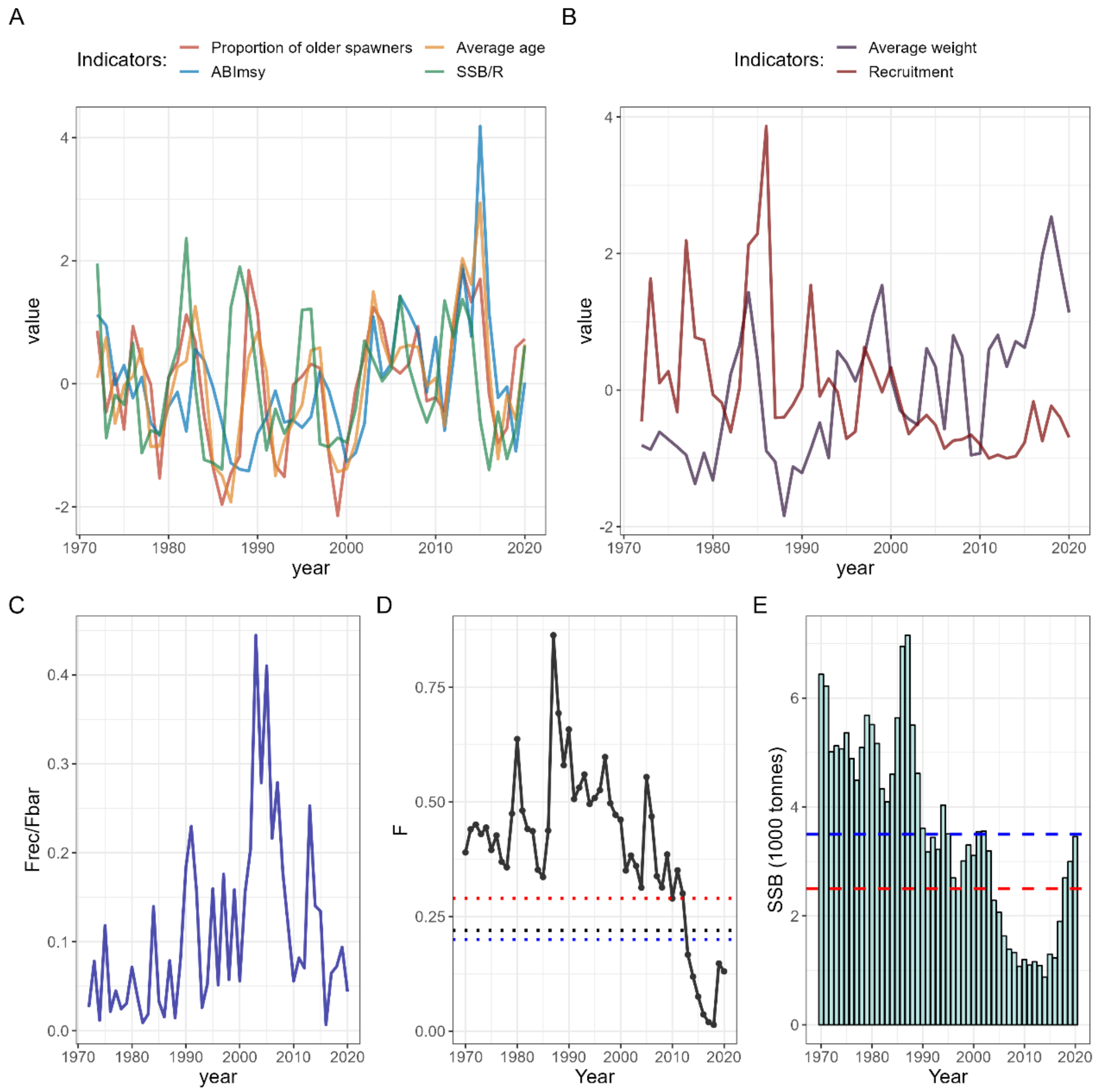

Figure A2.67. Estimated D3C3 indicators and stock status for Sole (sol.27.7a). See Figure legend A2.1 for further details.

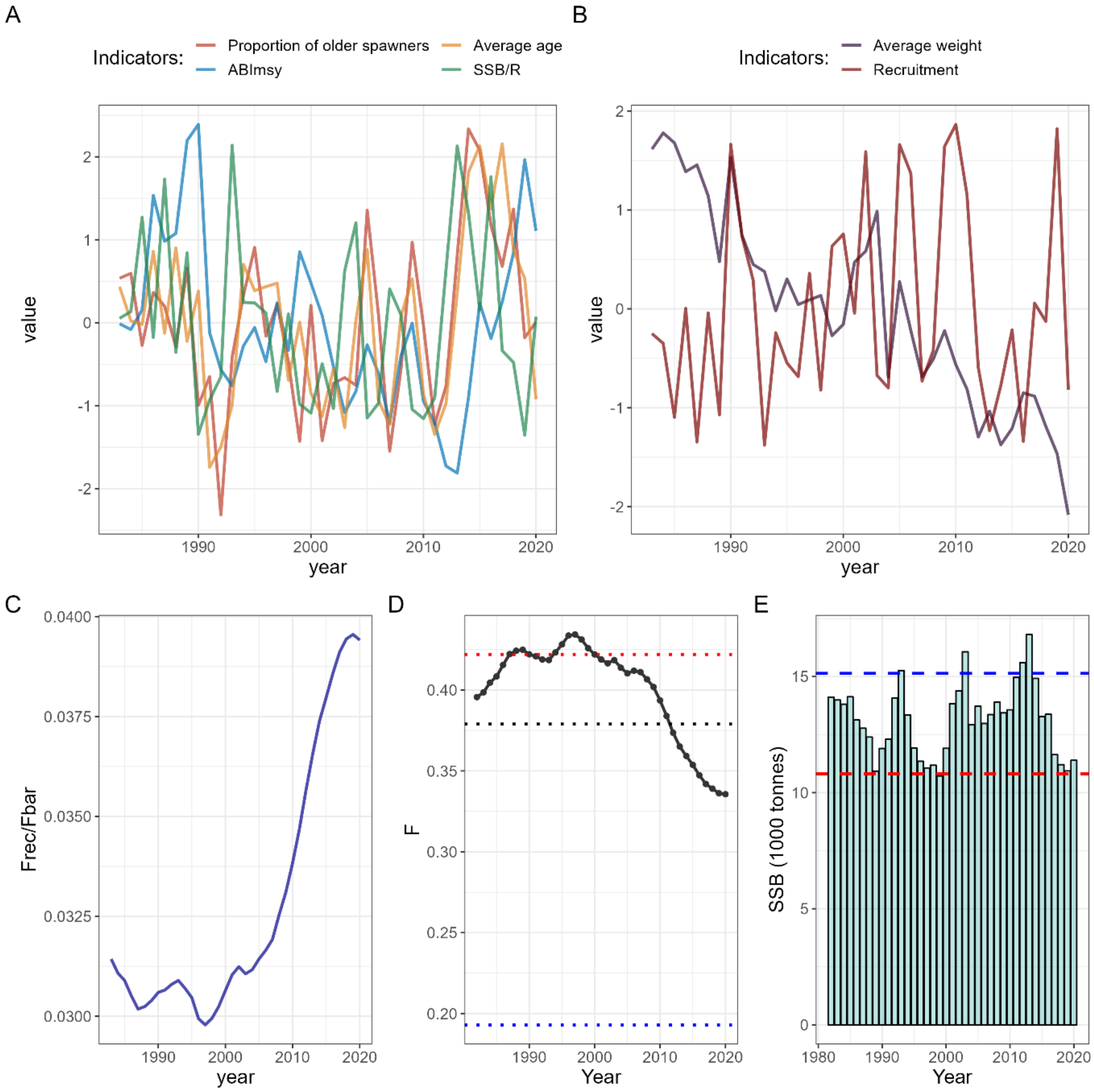

Figure A2.68. Estimated D3C3 indicators and stock status for Sole (sol.27.7d). See Figure legend A2.1 for further details.

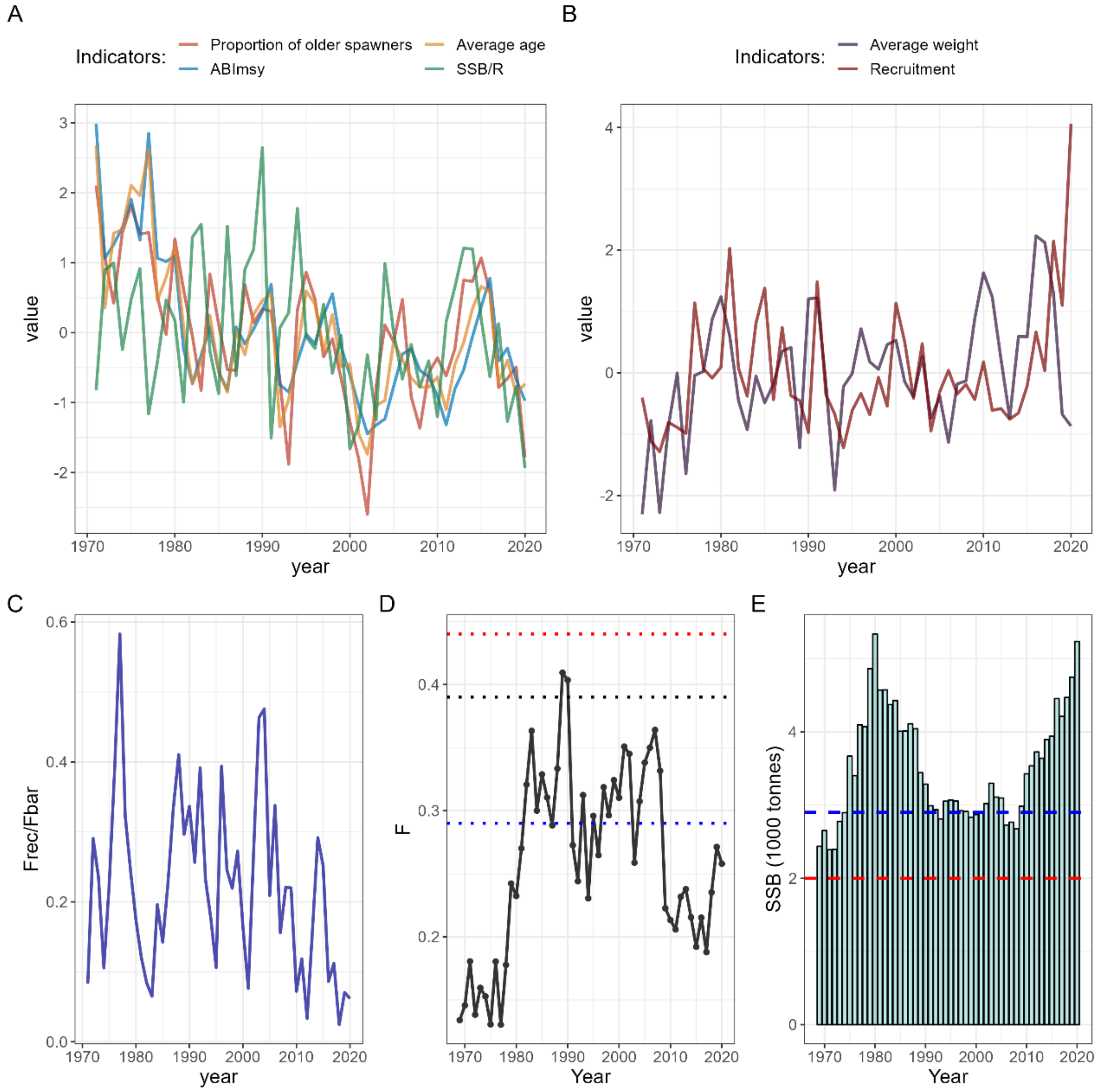

Figure A2.69. Estimated D3C3 indicators and stock status for Sole (sol.27.7e). See Figure legend A2.1 for further details.

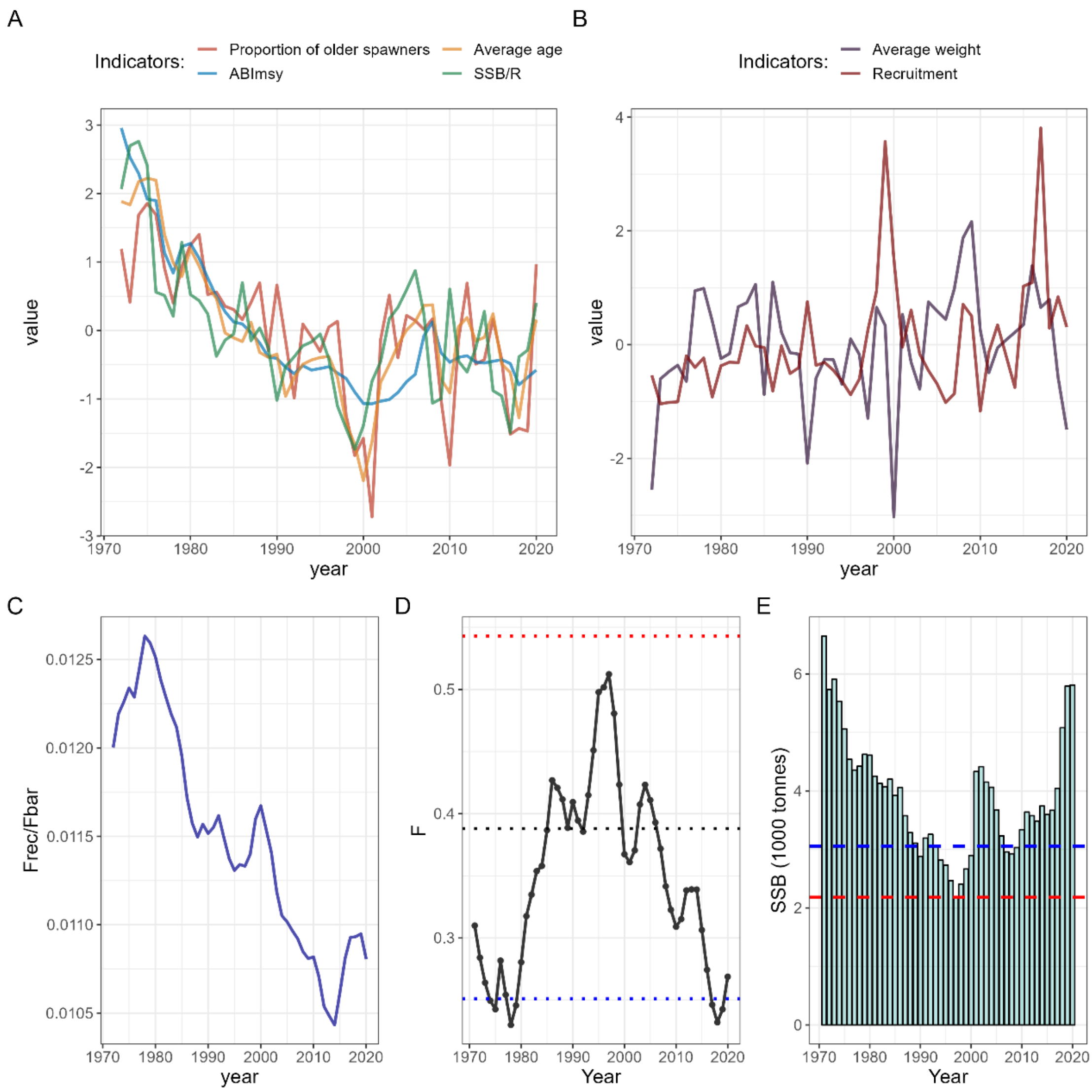

Figure A2.70. Estimated D3C3 indicators and stock status for Sole (sol.27.7fg). See Figure legend A2.1 for further details.

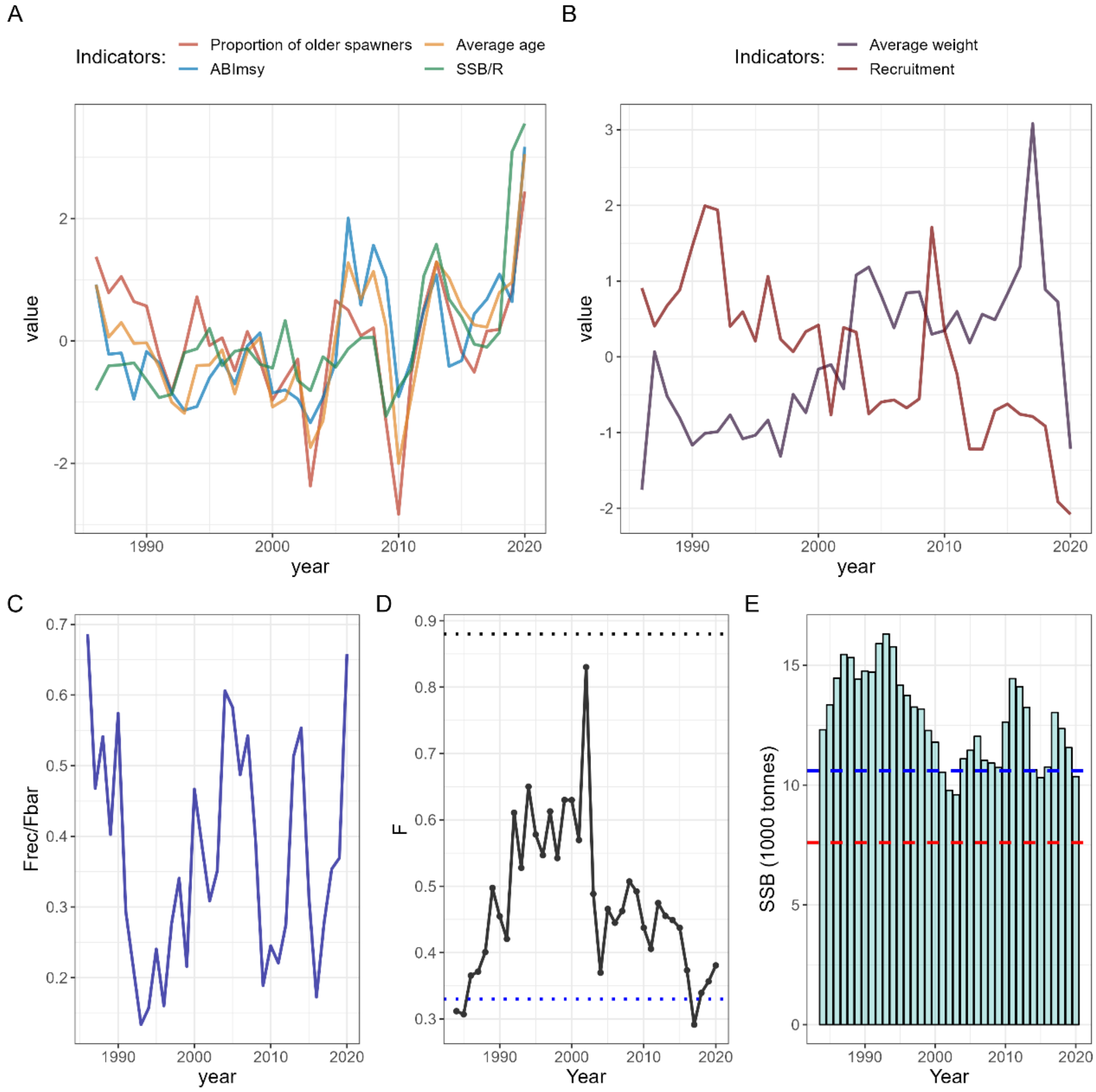

Figure A2.71. Estimated D3C3 indicators and stock status for Sole (sol.27.8ab). See Figure legend A2.1 for further details.

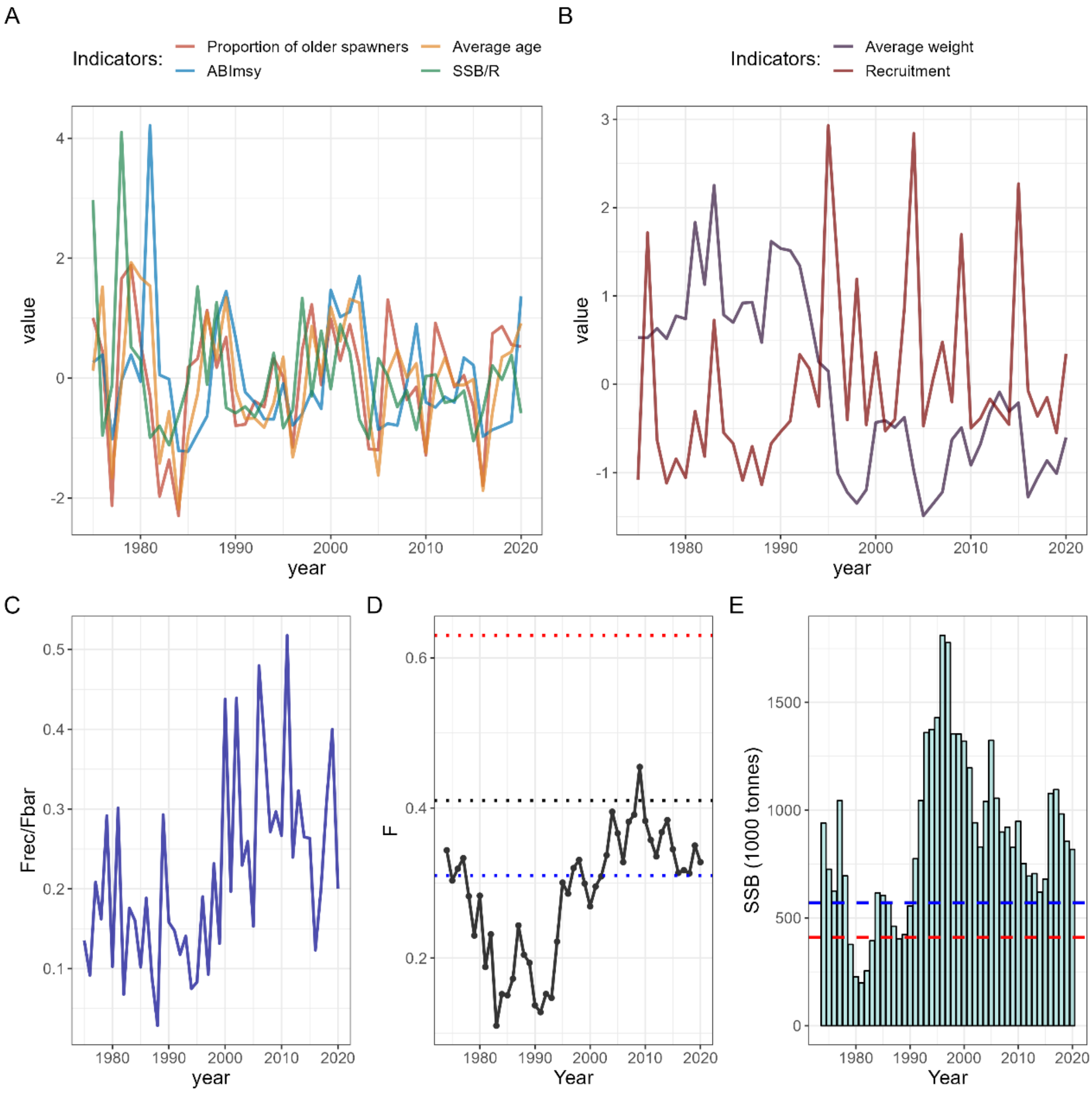

Figure A2.72. Estimated D3C3 indicators and stock status for Sprat (spr.27.22-32). See Figure legend A2.1 for further details.

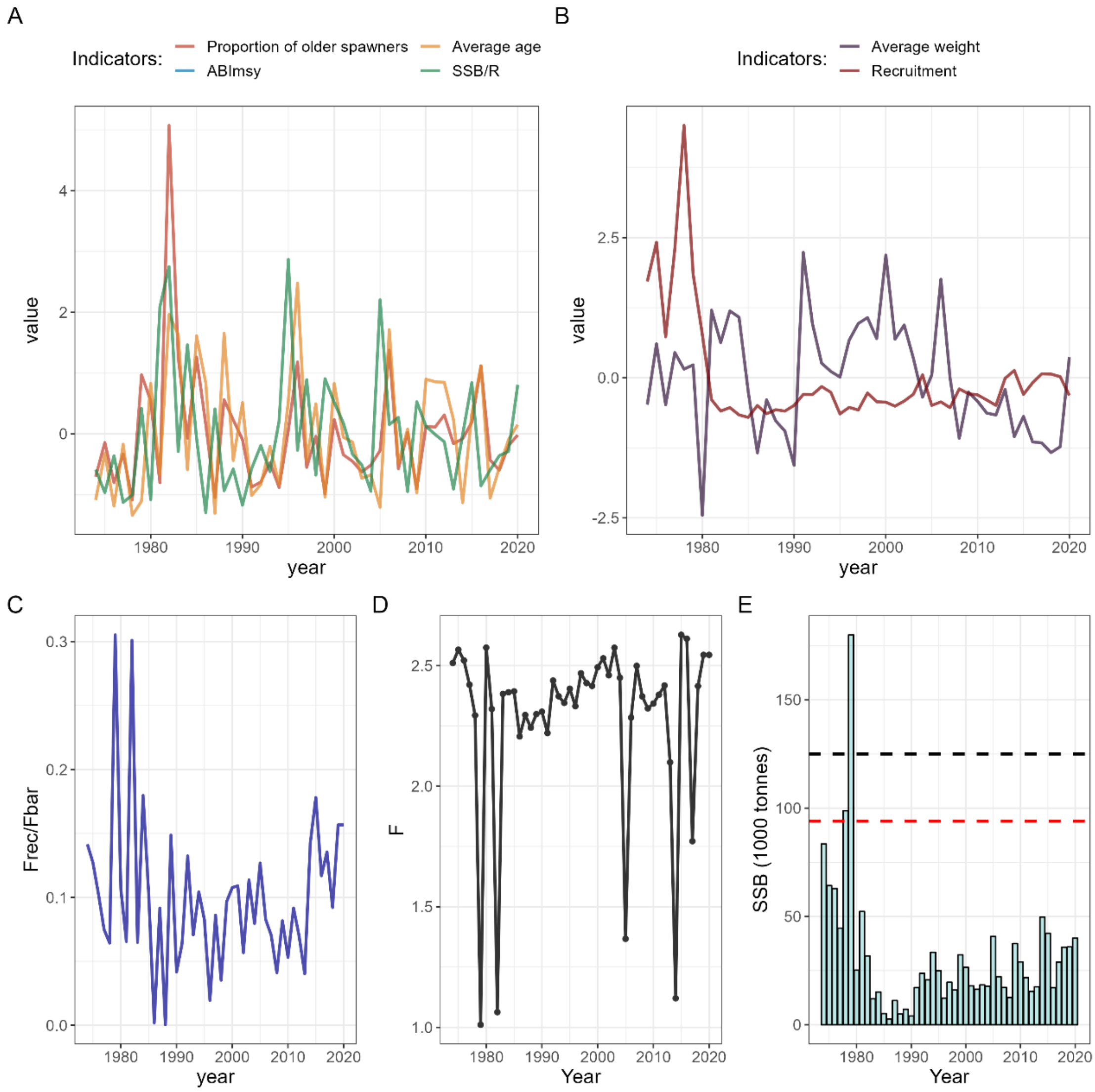

Figure A2.73. Estimated D3C3 indicators and stock status for Sprat (spr.27.3a4). See Figure legend A2.1 for further details.

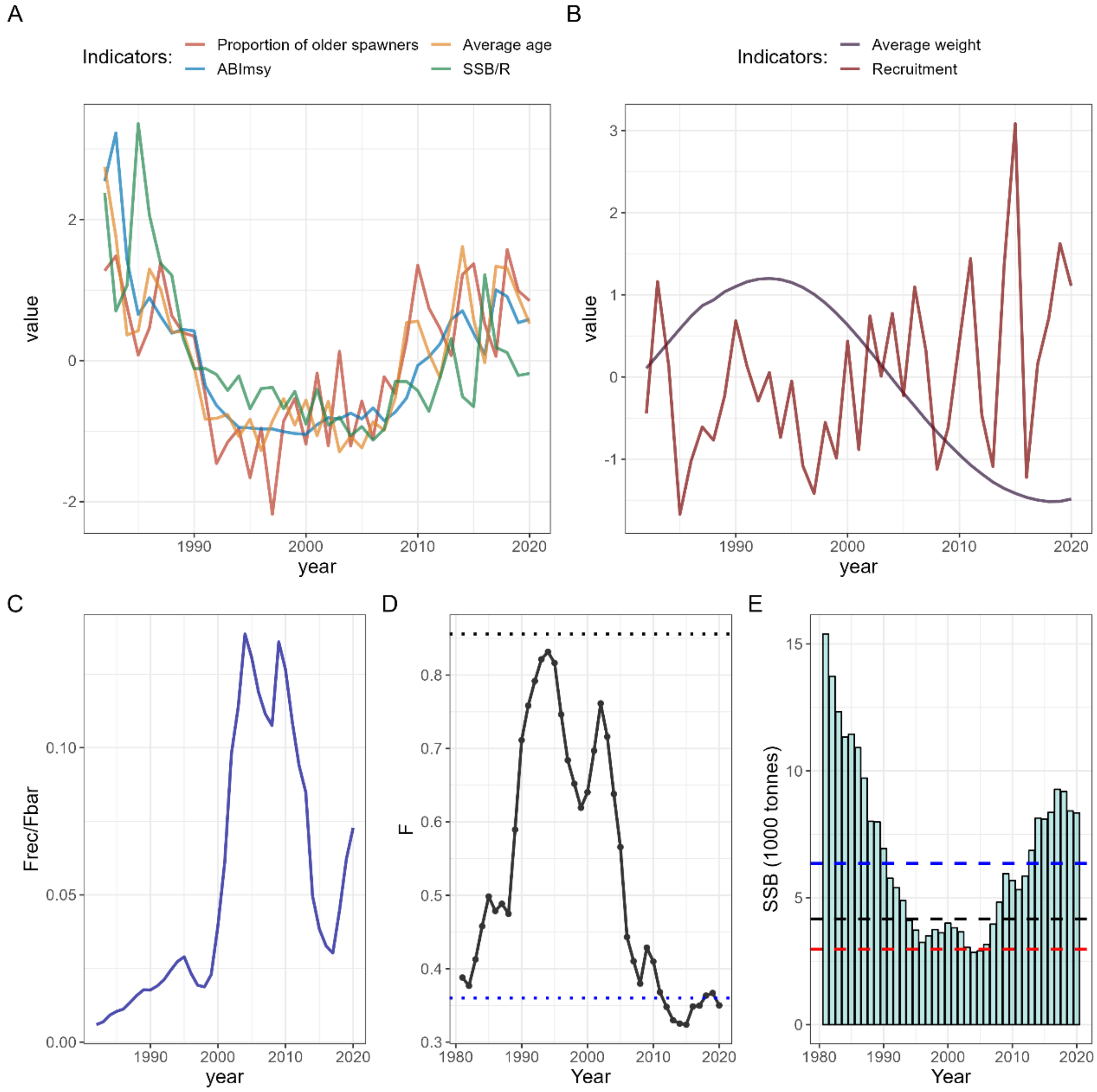

Figure A2.74. Estimated D3C3 indicators and stock status for Turbot (tur.27.4). See Figure legend A2.1 for further details.

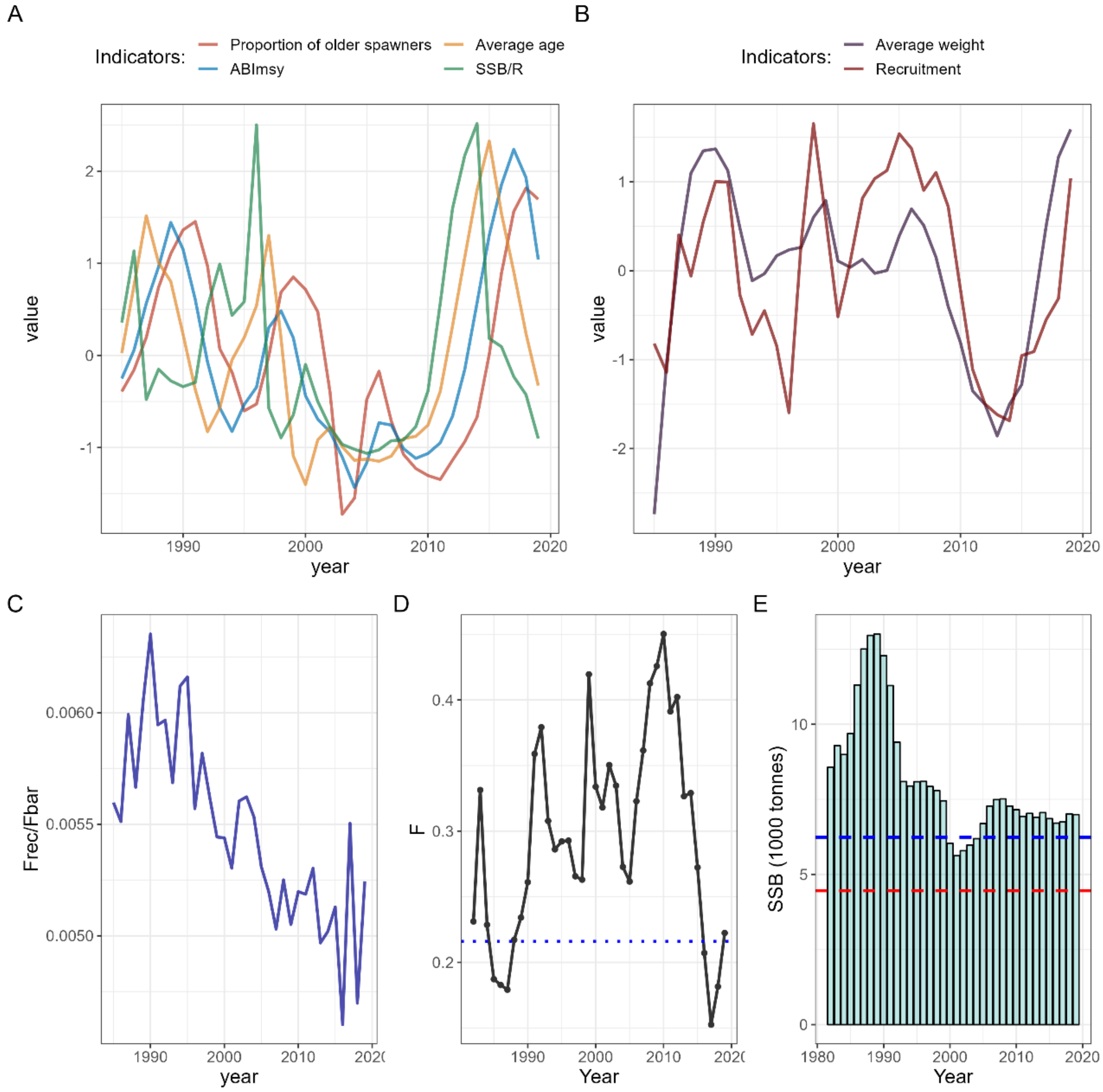

Figure A2.75. Estimated D3C3 indicators and stock status for Tusk (usk.27.5a14). See Figure legend A2.1 for further details.

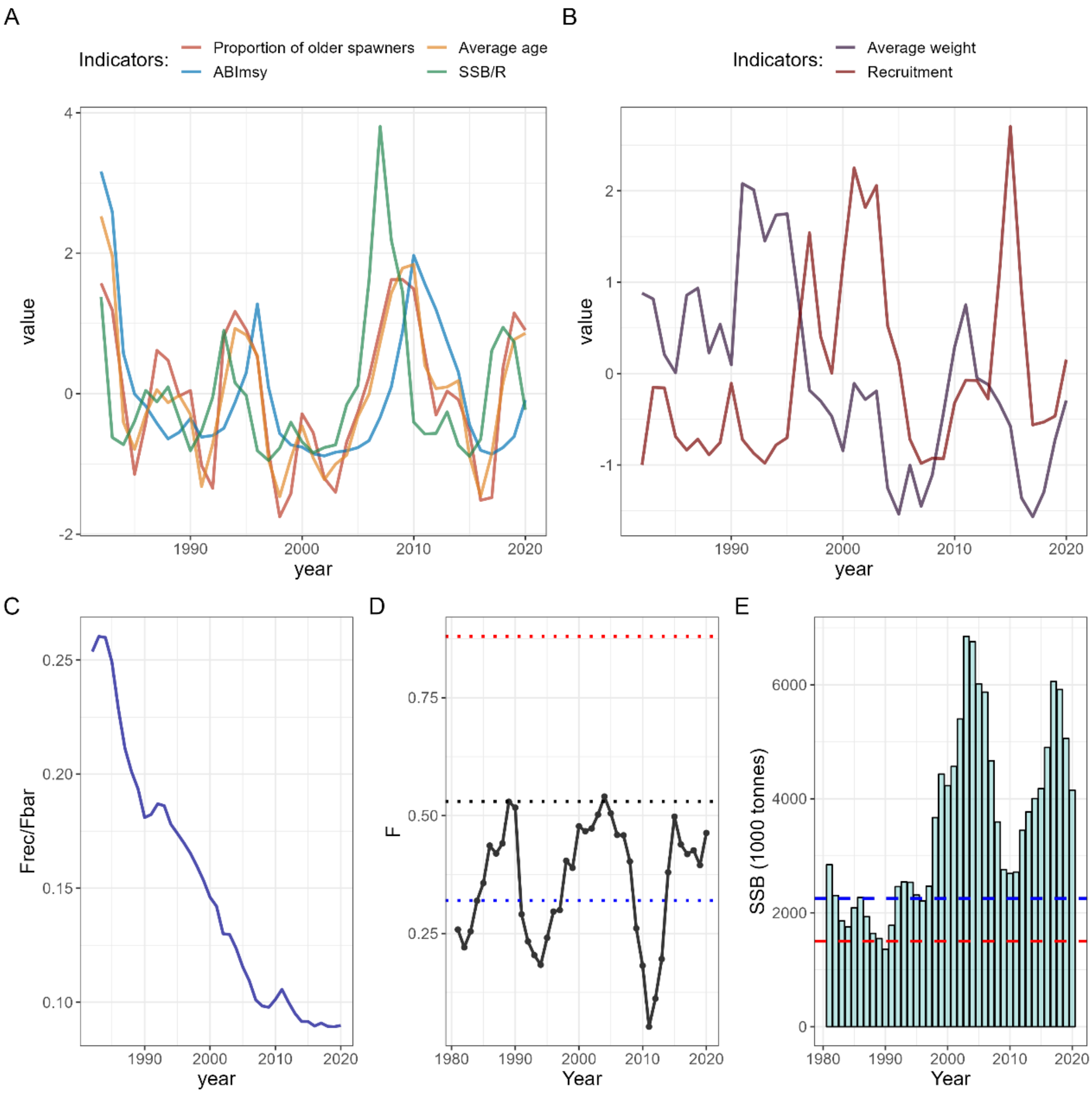

Figure A2.76. Estimated D3C3 indicators and stock status for Blue whiting (whb.27.1-91214). See Figure legend A2.1 for further details.

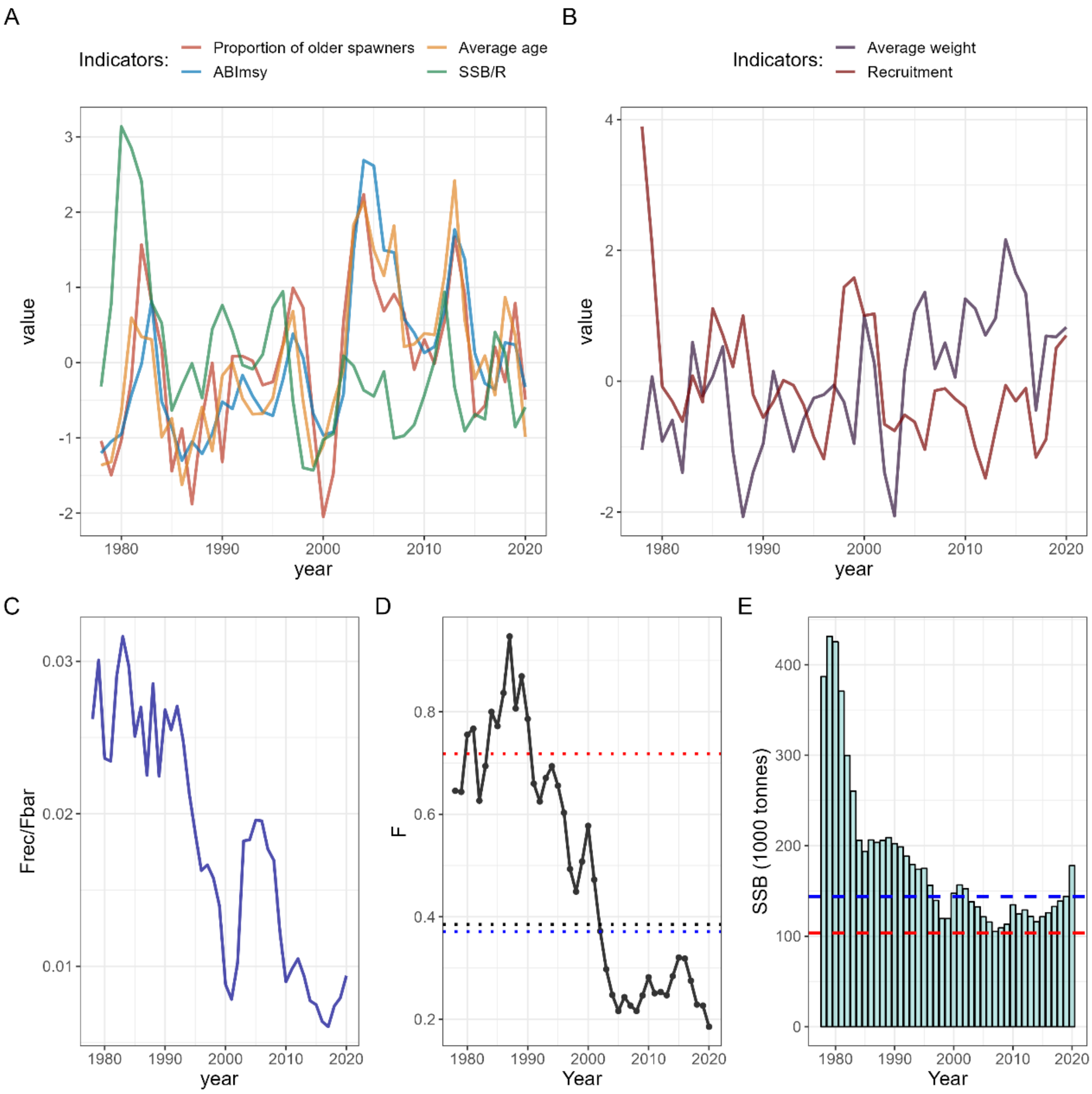

Figure A2.77. Estimated D3C3 indicators and stock status for Whiting (whg.27.47d). See Figure legend A2.1 for further details.

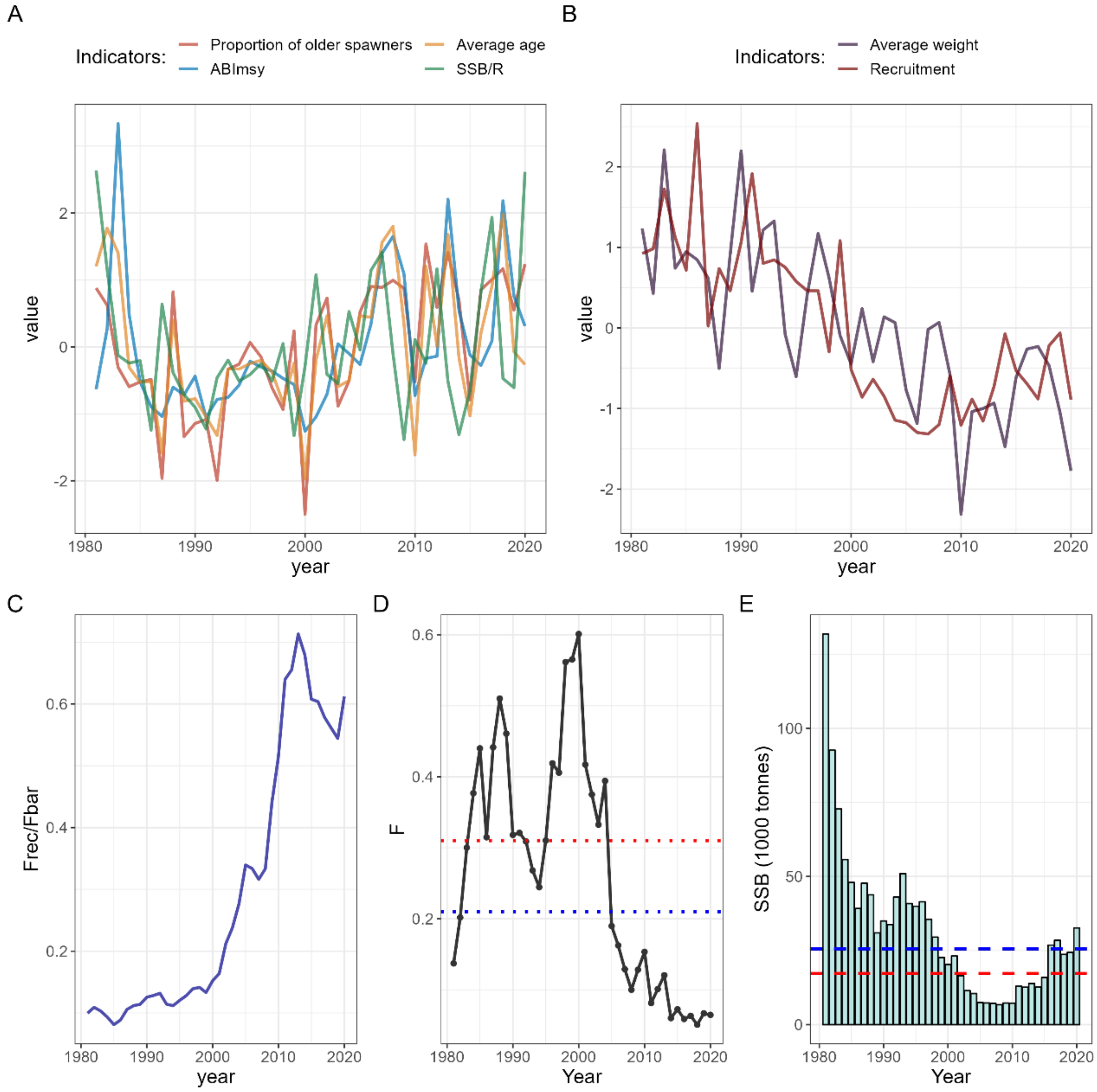

Figure A2.78. Estimated D3C3 indicators and stock status for Whiting (whg.27.6a). See Figure legend A2.1 for further details.

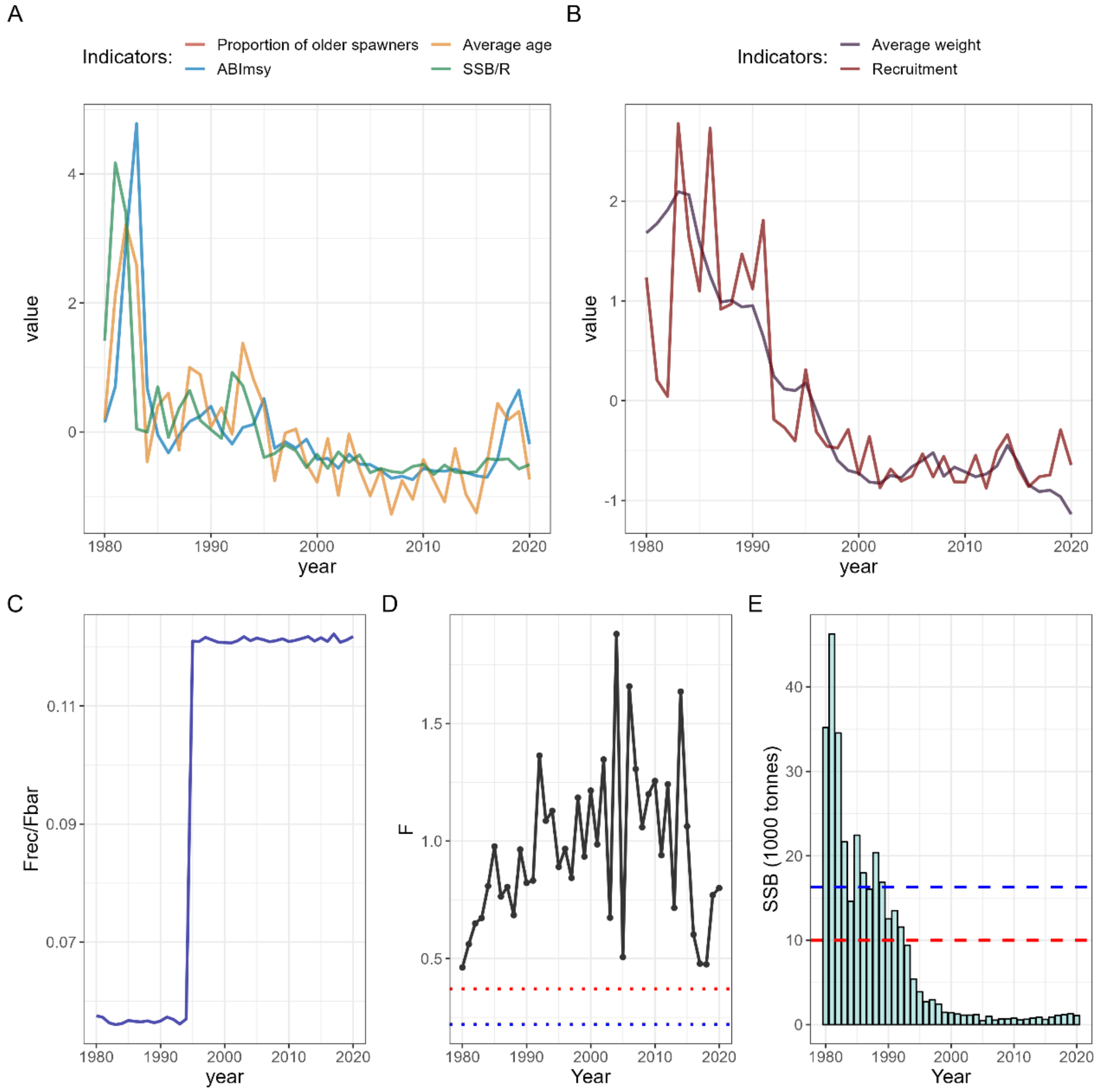

Figure A2.79. Estimated D3C3 indicators and stock status for Whiting (whg.27.7a). See Figure legend A2.1 for further details.

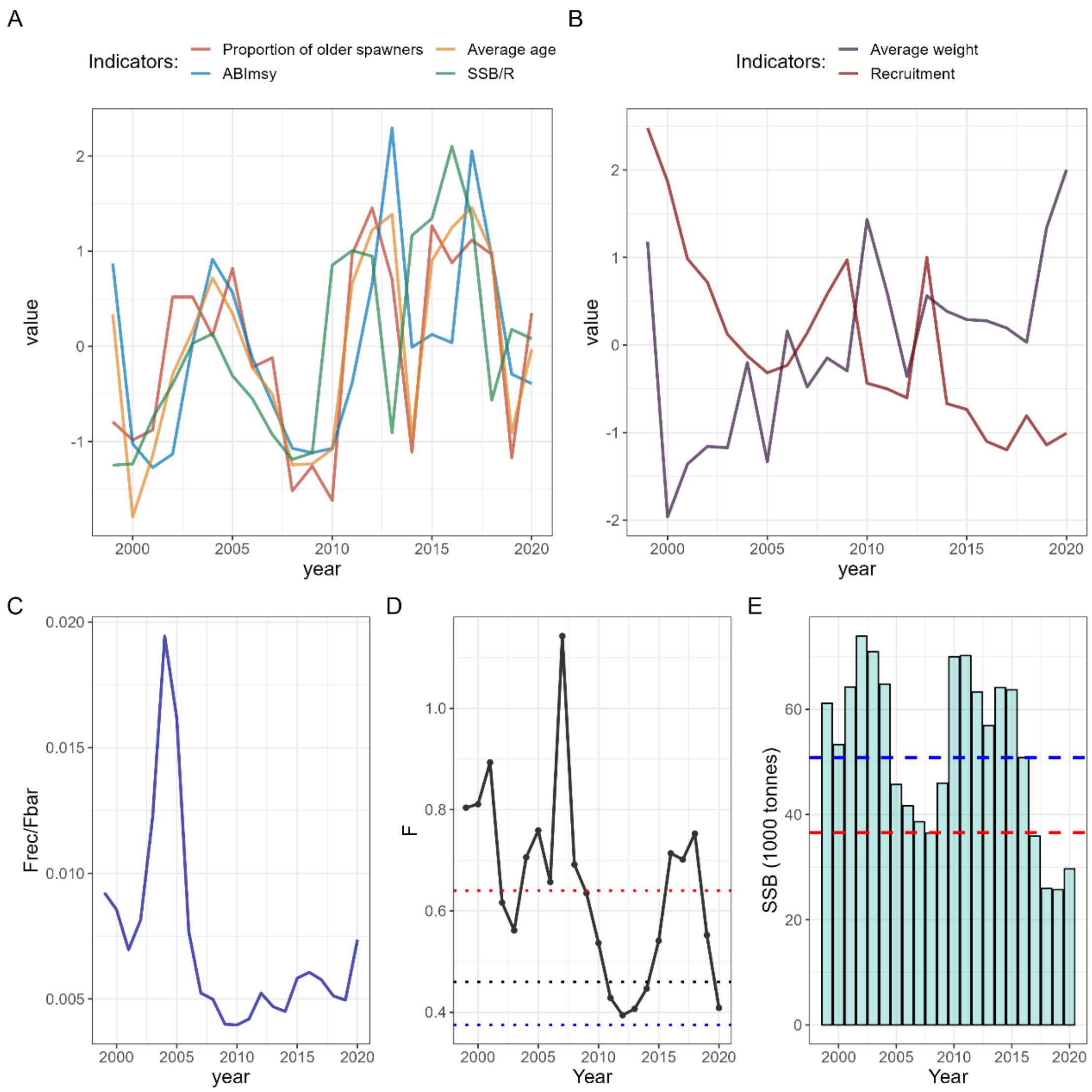

Figure A2.80. Estimated D3C3 indicators and stock status for Whiting (whg.27.7b-ce-k). See Figure legend A2.1 for further details.

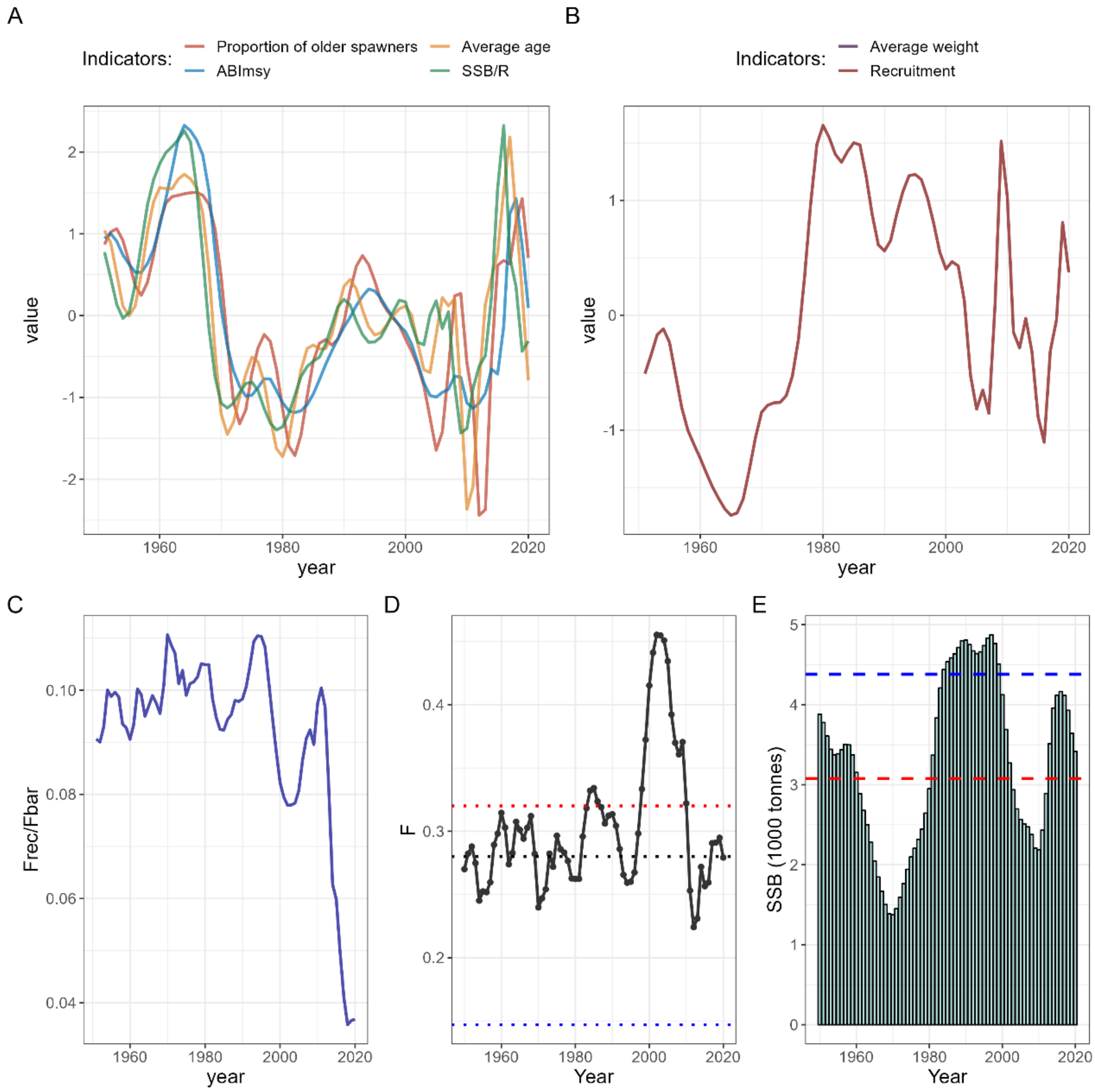

Figure A2.81. Estimated D3C3 indicators and stock status for Witch (wit.27.3a47d). See Figure legend A2.1 for further details.

Annex 3: Agenda

Time	Issue	Responsible
Terms of Reference	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.	
Monday September 18th		
14.30-15.30	Welcome, practicalities round of presentation, discussion of scope and terms of reference and adoption of agenda	Anna and Guiseppe
15.30-16.30	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 Nis Chris Paris
16.30-16.45	Coffee break	
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 plenary	
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 plennary	

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 why it is stated that HKE8_11 a4a model is not benchmarked. 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.