Annex 5: Benchmarks and prioritisation

This Annex was updated in November 2021

Benchmarks

A.1.1 Executive Summaries of recent benchmarks

Two benchmarks that involved WGNSSK stocks were organised in 2020–2021. The WKNSEA benchmark was convened to evaluate the appropriateness of data and methods to determine stock status for cod in the North Sea, Eastern Channel and Skagerrak (Cod.27.47d20) and sole in the eastern English Channel (Sol.27.7d). Furthermore, two interbenchmark workshops (IBPNSWhiting) were convened in 2021 for North sea whiting (whg.27.47d) to include new natural mortality estimates in the assessment and for witch in the Greater North Sea (wit.27.3a47d) to include new survey indices.

A.1.1.1 Cod in 4, 7.d and 20 (WKNSEA 2021)

The North Sea cod stock was put forward for benchmark in 2021 due to conflicting signals in the underlying data and a developing retrospective bias in the assessment. In addition, the stock ID was put forward as an issue for North Sea cod. To address the latter, a four-day workshop on Stock Identification of North Sea Cod (WKNSCodID) was held in August 2020 to review information on the population structure of cod in the North Sea and adjacent waters. The workshop concluded that North Sea cod includes reproductively isolated Viking and Dogger cod populations, and the Dogger population has some phenotypic structure and extends to 6.a.N. However, the data evaluation workshop found unexplained discrepancies between the spatially-disaggregated data and the data as used in the current assessment, possibly caused by the very short timeframe for data providers to compile the data. Further, the spatially-disaggregated time-series started in 2002 which would truncate the time-series with 40 years. Therefore, the workshop concluded that development of spatial approaches would not be possible in time for a benchmark in 2021, although it was agreed that a spatial-disaggregated cod assessment would be preferable and work to archive this goal should be initiated in the next years. However, after consultation with the ACOM LS it was decided to improve the present combined assessment until a

spatially-disaggregated time-series would be available. At this benchmark;

- recreational catches were considered but not included in the analytic assessment due to data quality issues.
- Updates were made to the base calculations for deriving the subarea-weighted maturity ogive. The first 15 years (1963–1977) were removed and the ogive not smoothed. Further, maturity is now modelled as a process.
- Stock weights have changed to IBTS Q1 survey weights for ages 1–2 and as Q1 catch weights for ages 3+.
- A high-resolution delta-GAM survey indices with a fixed spatial term and yearly independent deviances is now used.
- Introduction of a recruitment index based on the IBTS Q3 at age 0 and shifted to the beginning of the following year has been introduced.
- Smoothed M data from the 2020 SMS key run is included with an addition of adjusted Ms from 2011 for ages 3+ to mimic migration out of the stock area into 6aN.

- Several configuration adjustments were made to the model.
- New reference points were calculated based on a truncated time-series (1998–2019) and a type 6 S–R plot.
- Inclusion of both age 0 and age 1 in the protocol on the reopening of the advice.

A.1.1.2 Sole in 7.d (WKNSEA 2021)

Sole in Division 27.7d had data issues with a commercial tuning series, and an inter-benchmark was set up in August 2019. At the end of the inter-benchmark, it was found that some commercial catch data for 2016 and 2017 were aggregated incorrectly for older ages. During the benchmark in February 2020 (WKFLATNSCS 2020), further data issues were discovered. As a result, the benchmark process was postponed to the WKNSEA 2021 benchmark, and in the data call, the commercial catch data time-series was corrected and re-uploaded. Discard data were available from 2004 onwards. Prior to 2004, discards were reconstructed using the ratio between discards and landings in the period 2004–2008. Stock weight-at-age were set to quarter 1 catch weight-at-age (2004–2019) to improve consistency. They were reconstructed prior to 2004 using the ratio between quarter 1 and yearly catch weight-at-age using data from 2004–2019. Six tuning fleets are currently included in the assessment: three survey indices (UK BTS, FRA YFS and UK YFS) and three commercial indices (BEL CBT, UK CBT, FRA COTB). During the benchmark, the commercial indices were changed to biomass indices in the assessment instead of disaggregating them by age to avoid double counting of commercial data. The French commercial otter trawl fleet (FRA COTB) and Belgian commercial beam trawl fleet (BEL CBT) were revised using the adjusted catch data as input and following a model-based approach to derive an lpue index that is considered to reflect the fishable biomass of the stock.

A state–space assessment model (SAM) was chosen for this stock using the three commercial LPUE indices as fishable biomass (FRA COTB, BEL CBT, UK CBT) and three scientific, age-structured survey indices (UK BTS, UK YFS, FRA YFS). Compared to the previous XSA assessment model, the spawning–stock biomass is estimated to be significantly lower, while the fishing mortality is estimated to be higher. Following the changes in the input data and assessment model, the reference points were re-calculated and F_{MSY} is now estimated at 0.193 (similar to previous estimate).

A.1.1.3 Whiting in 4 and 7.d (IBPNSWhiting 2021)

The Inter-Benchmark Protocol of North Sea Whiting (IBPNSWhiting 2021) met to consider the use of updated Natural Mortality estimates from the North Sea multispecies assessment model developed by the Working Group on Multispecies Assessment Methods (WGSAM, 2020) for Whiting in Subarea 4 and Division 7.d (North Sea and eastern English Channel). In this report the estimates of Natural Mortality are compared to previous estimates (WGSAM, 2018), the effects of this change on the assessment model are considered, and reference points recalculated. The estimates of Natural Mortality from the most recent multispecies assessment model were slightly higher than from the previous run, particularly at age 0. Incorporating these revised Natural Mortality estimates into the assessment resulted in only minor changes to the stock size, recruitment and exploitation estimates, and to the quality of the model fit. The updated model showed higher retrospective bias than previously, but was still judged to be acceptable. Following the revision of the assessment model, reference points were re-calculated following the ICES Technical guidance and using the same assumptions as for previous assessments. This resulted in lower biomass reference point (e.g. MSY $B_{trigger}$ decreased from 167 000 t to 144 000 t) and a substantial increase in F_{MSY} (from 0.172 to 0.371).

A.1.1.4 Witch in 3.a, 4 and 7.d (IBPWITCH 2021)

IBPWitch was primarily tasked with establishing a new method to create reproducible survey indices after the method determined in the last full benchmark was no longer available/reproducible. In addition to establishing the new survey indices, the group also considered the assessment model configuration and updated the stock's reference points, to ensure coherence and reproducibility across future assessments. The group selected a Generalised Additive Model modelling approach, which is implemented across many other ICES stocks, to generate indices by age and year across two quarters (Q1 and Q3). The indices' model assumptions and configurations were thoroughly investigated, assessed and documented. The assessment model, a State-space Assessment Model (SAM) that is available on stockassessment.org, was modified so that age1 survey indices were not included in the model. Assumptions of interdependence in fishing pressure between ages that were adopted in the previous benchmark were tested and retained. The short-term forecast methodology was modified, to provide a more consistent estimation of recruitment. The inclusion of different data sources in the calculation of reference points was thoroughly considered and the decision made to utilise the same dataset for the stock-recruitment relationship as in the previous benchmark with just updated recent years. This relationship was modelled according to a "type-two" segmented regression and utilised to estimate B_{lim}. Future work on this stock relies on improved age sampling (spatially and inter-lab calibration) from surveys, investigation of alternate surveys covering deeper waters, and evaluating a shift to a length-based assessment.

A.1.2 Benchmarks for 2022

A.1.2.1 Northern Shelf haddock

Data available/needed

Current assessment issues

Proposed working papers/analyses

Work plan for benchmark

The issue list for Northern shelf haddock (had.27.46a20) is given below.

Issue	Problem/Aim	Work needed / possible direction of solution	Data needed to be able to do this: are these available / where should these come from?	External expertise needed at benchmark type of expertise / proposed names
(New) data to be Considered and/or quantified	SSB is used to indicate both reproductive potential and harvestable biomass, and it may be a poor proxy for both.	Investigate indices of reproductive potential and methods to use them in management advice.	Weight-at-age and fecundity/egg condition data.	Fecundity modelling: Peter Wright (MSS).
	The stock is considered to be homogenous throughout subareas 4 and 6a, but there may be relevant substock structure.	Explore stock ID and structure, using otolith microchemistry, tagging data, and the spatial range of genetic data.	Otolith micro-chemistry, tagging, genetic data.	Stock ID: Peter Wright, Neil Campbell (both MSS)
Tuning series	The survey data used in the assessment cover only the North Sea component.	Explore combining survey indices from the North	Survey data available.	Survey modelling: Andrzej Jaworski (MSS), Casper Berg (DTU-Aqua)

		Sea and West of Scotland.		
Biological parameters	The assessment uses a knife-edge maturity at age 3.	Derive time-varying maturity estimate.	Maturity data from IBTS surveys.	Maturity modelling: Peter Wright (MSS), Casper Berg (DTU-Aqua)
	Mean weights-at-age for total catch are used for stock weights.	Derive estimates of mean weights at age for stock.	Weight data from commercial catch and surveys.	Weight modelling: Peter Wright (MSS), Casper Berg (DTU-Aqua).
Assessment method	TSA support likely unavailable after 2021/22.	Consider alternative models which are compatible with high performance computing (for MSE).	Alternatives likely to use same data as TSA, although a spatio-temporal model such as SS3 would require more extensive spatial data.	SAM: Anders Nielsen (DTU-Aqua). Potentially SS3 expert.
	Plus group does not seem to be well fitted.	Investigate poor fit in plus group in view of increasing relative importance of this age class.	No extra data requirements.	SAM: Anders Nielsen (DTU-Aqua). Potentially SS3 expert.
	Exploratory model SURBAR requires further development.	Develop likelihood profiling for ad hoc parameters, and catchability estimation model based on catch curves.	No extra data requirements.	SURBAR: Coby Needle (MSS).
	Haddock is characterised by occasional large year-classes, which do not conform to the usual distributional assumptions for modelling recruitment.	Exploration of modelling techniques for sporadic recruitment is needed (mixed distributions etc.).	No extra data requirements.	SAM: Anders Nielsen (DTU-Aqua)
Biological Reference Points	Reference points will need to be updated following data, assessment and forecast revisions.	Follow the standard processes where appropriate to generate new reference points.	No extra data requirements.	No external expertise required.
Forecast	Growth model used in forecast needs to evaluated.	Investigate extent of cohort effect on growth rate. Ensure consistency between catch components for weight at age cohort modelling. Develop non-spreadsheet approach to forecasting weights.	No extra data requirements.	Growth modelling: Andrzej Jaworski (MSS), Casper Berg (DTU-Aqua).
	Approach for recruitment estimation in the intermediate year	Investigate intermediate year recruitment assumption.	No extra data required.	Statistical modelling.

	needs to be evaluated.	Forecast value for recruitment would benefit from including information on the probability of large year classes occurring.		
Other	There appear to be SOP issues in InterCatch data.	Ensure consistency in catch data used in assessment and advice sheet.	InterCatch database.	InterCatch experts: Henrik Kjems-Nielsen (ICES)

A.1.2.2 Plaice in 4, 20

Data available/needed

Current assessment issues

Proposed working papers/analyses

Work plan for benchmark

The issue list for plaice in 4, 20 is given below.

Issue	Problem/Aim	Work needed / possible direction of solution	Data needed to be able to do this: are these available / where should these come from?	External expertise needed at benchmark type of expertise / proposed names
(New) data to be Considered and/or quantified	Due to a sequence of "low" catch rates in several years, the industry are not in agreement with the ICES estimated stock status.	Dutch commercial LPUE analysis, 1) LPUE by subarea and gear type; 2) LPUE of targeted fisheries	Dutch commercial landing, discards, VMS data	Experts on Dutch fisheries
	Applying smoothed stock/catch/discards at weight, investigate its trend and impact on catch/stock weight	Apply gam		Stock coordinator,
	Update Mortality, maturity, age and length distribution, by subarea (North sea and NW-North sea)	Apply models and evaluating trends Stock ID analysis	BTS, IBTS, commercial catch data	Expert on biological modelling, expert on survey
Tuning series	The delta-gam IBTSQ1 age>=5 indices showed upward revision in last 3 years, this is likely the cause of the	Investigate the data quality and age reading.	IBTS-Q1	Expert in IBTS-Q1 survey, age readers. Stock coordinator, Casper Berg

	upscaling SSB in empirical retro analysis.			
	IBTSQ3 showed strong signals in north west area (around Scotland). Younger ages (>=2) even appear in this area.	Investigate the data quality and age reading, especially around Scotland. Explore gear effect, time-varying spatial random effect. Validate the signals in this area with catches in Scotland	IBTS-Q3 Catches in Scotland	Expert in IBTS-Q3, age readers, Stock coordinator
	Age 0 are moving from coastal area (SNS) to BTS area. This will result in a changed catchability between these 2 surveys.	Explore possibilities of a combined indices; split the time series; include age 0 BTS indices in assessment	BTS, SNS	Casper Berg, Stock coordinator, survey expert in SNS and DFS
	Currently 6 survey indices are included. A clear vision is needed on the contribution of each survey in the assessment.	Sensitivity analysis on survey (LOO), especially when 2 surveys are highly correlated (e.g. BTS and IBTS-Q3) Internal and external correlations within and between surveys (catch as well) SURBAR		Experts on stock assessment, modeling, Coby Needle, Stock coordinator
Assessment method	Residual patterns in both catch survey	Re-define the spline structures in the model; number of knots, age plateau, max (or plus) age		Experts on SAM or AAP
	Confirm the stock status with run on another assessment model	SAM		Experts on SAM
	Large empirical retrospective pattern	Likely caused by the upscaling revision from IBTS-Q1		Stock coordinator
	Currently all discards in assessment are considered dead. A non-zero survival rate could be included in the model.	Include non-zero survival of discards in the model		Expert in survival experiment and stock assessment

Biological Reference Points	Determine MSY reference points	Run EqSim functions	Using the final assessment	Experts in computation of reference points, Stock coordinator
Forecast	Validate the RCT3 method			Experts in RCT3, stock coordinator
	Possibility and quality of including IBTS-Q1 indices in spring forecast.	Validate the prediction performance of IBTS-Q1, ask for possibility of having IBTS-Q1 plaice ready in Spring		Experts in RCT3, IBTS-Q1
	Given large changes in plaice in response to environmental changes (e.g. density dependent growth, differences in age distributions), we need to know the efficacy of ICES advice rule n the long term	MSE		Expert on MSE and stock coordinator

A.1.3 Benchmarks for 2023 and beyond

There remain a few Category 3+ stocks that have not yet been benchmarked, namely bll.27.3a47de (brill), pol.27.3a4 (pollack) and gug.27.3a47d (grey gurnard). The stocks being considered for benchmark in 2023 are mur.27.3a47d (red mullet) and pok.27.3a46 (saithe), the former due to the assessment having been rejected in 2021 and being downgraded to category 5, and the latter to improve data input and account for the recent low productivity regime in the forecast. Full benchmark issue lists for these stocks will be developed in the coming year.

A.2 Benchmark prioritisation

Benchmark prioritisation was conducted according to the scheme described in Table A2.1. Table A2.2 provides a summarised list of benchmark issues for each stock, and applies the scoring scheme to each stock. The finfish stocks listed in Table A2.2 have been ordered from highest to lowest score. *Nephrops* have not been considered in this scheme as the benchmark process for *Nephrops* is handled separately.

Table A2.1. Prioritisation scoring used in Table A2.2.

Category	1. assessment quality	2. Opportunity to improve	3. Management importance	4. Perceived stock status	5. Time since last benchmark
Scoring / weight	**0.4**	**0.3**	**0.1**	**0.1**	**0.1**
5	Assessment judged to be inadequate to provide advice (e.g., bias, stock id, unreliable catches, major change in biological processes/productivity)	New approaches and new data sources will be available for the stock, and these are likely to address issues or change perception of stock dynamics	All 4 attributes: a) Advice on fishing opportunities is requested for the stock. b) Stock is the object of an agreed management plan. c) Stock is the object of a directed fishery. d) Stock is included in a mixed fishery analysis, is a likely choke stock, or the object of a pelagic fishery (meets 1 of the 3)	Most likely below B_{lim}, or stock is in rapid decline, or state of the stock unknown	Stock has never been benchmarked
4	Assessment has high potential & priority to be upgraded to Cat. 1 from Cat. 3 or to Cat. 3 from Cat. 5 and 6	New data sources or corrections in data, or new methods will be available for the stock, and these are likely to address issues or change perception of stock dynamics	3 attributes	Between B_{lim} and MSY $B_{trigger}$	Stock has been benchmarked 10 years or more ago
3	Assessment judged to have substantial deficiencies (models and/or data) but considered acceptable	Some improvement in data /modelling approaches will be available, and unclear whether they will address issues or change perceptions	2 attributes	About MSY $B_{trigger}$	Stock has been benchmarked between 5 and <10 years ago
2	Assessment has no substantial or only minor issues	Minor improvement in data or methods will be available	1 attribute	Above MSY $B_{trigger}$	Stock has been benchmarked between 1 and < 5 years ago
1	Assessment has no obvious issues	No change in data or models will be available	No attributes	Near highest on record	Stock was benchmarked in the last year

Table A2.2. Benchmark prioritisation scoring for WGNSSK finfish stocks along with issues. The weighting for the scoring categories is according to Table A2.1. Stocks have been ordered from highest to lowest total score.

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
mur.27.3a47d	Cat 3 No TAC	- Age data from other countries than France need to be provided as everything is actually raised using the French catches in the Eastern Channel and few sample from the North Sea. - Length data is mainly provided by France and a also UK with recently some submission from Netherlands, however the sampling coverage remain to low and further sampling is required. - No survey is available in the North Sea; IBTS/UK BTS should be investigated again. So work was done to assess the representativeness of the Eastern Channel data compared to the stock, but these should be investigated further - CGFS survey data issue in index calculation needs to be fixed. GAM or GLMM methods such as the method developped by Berg et al. or Thorston et al. should be explored to account for missing data UK haul in 2020 and also better account for the change in vessel in 2014.	- Assessment model was rejected by WGNSSK 2021 and a category 5 advice is given for this stock. - With so few age classes exploited the a4a model formulation used needs revision as it is no longer fit to give advice. - Explore other models (SAM, SURBAR, length-based model…) - Explore methods applied to "short lived species" (two stages model)? - SPiCT should be explore again either as basis for advice or to estimate the stock status.	- This stock was downgraded to category 5, so no forecast is done currently. This should be investigated if the assessment method is improved. However, there is no TAC for that stock so Forecast is not a priority, although reference points are still important.	5	4	2	5	3	4.2

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
		- Even if discards are expected to be very low (no minimum landing size, high price), discards data should be re-investigated - Based on the recent WD presented at WGNSSK2020 stock ID should be reinvestigated								
ple.27.420	Cat 1 shared	- The delta-gam IBTSQ1 age>=5 indices showed upward revision in last 3 years, this is likely the cause of the upscaling SSB in empirical retro analysis. Investigate the data quality and ALK. - IBTSQ3 showed strong signals in north west area (around Scotland). Investigate o Quality of samples: gear and age reading. o Why younger ages (age>=2) appear in this area in last 15 years o Indices with/without gear effect, time-invariant and time-varing, including and excluding NWarea o Validate indices with catches in Scotland - WGBEAM indicates an increasing age 0 selectivity in BTS while a decreasing sel in SNS (aim for age0), maybe a combined indices - Investigate the spatial mismatch between survey fishing effort, e.g.	- Solve residual patterns - Investigate the survey leave-one-out results and retro analysis on LOO - "error" in discards due to non-zero survival in assessment (~9%), might lead to overestimate of stock size - explore different assessment models	- RCT3 analysis on recruitment? If not, how to include recuitment survey in assessment, e.g. DYFS - Considering density-dependent growth in reference point calculation?	5	4	5	2	2	4.1

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
		LPUE - Explore stock ID trend and difference between NS and NW-NS: maturity/mortality/sex ratio/growth rate/LF/survey_indices								
tur.27.3a	Cat 3 No TAC	- review of knowledge, including genetic findings, and turbot migrations and spawning grounds - Stock definition - dealing with the missing Swedish catches - overview of recreational catches - dealing with a reduction in sampling for length - survey data to be investigated and mapped in more detail (including options for a combined Delta-GAM index for the entire stock area) - update of Cardinale et al (2009) survey time series	-advance assessment (SPiCT)	-develop reference points	5	5	2	2	1	4
pol.27.3a4	Cat 5 No TAC	- Examine if data exist that allows the determination of age and size of maturity ; - Explore the potential availability of data that would allow the determination of size/age in catches and the possibility to determine reference points	-develop an assessment if possible	-develop reference points if possible	5	2	1	5	5	3.7

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
whg.27.3a	Cat 3 PA	-explore stock ID	-develop assessment (SPiCT)	-develop reference points -develop advice based on short term forecast	4	4	3	5	1	3.7
pol.27.3a4		- Examine if data exist that allows the determination of age and size of maturity ; - Explore the potential availability of data that would allow the determination of size/age in catches and the possibility to determine reference points most likely through the use of data limited approaches	-develop an assessment if possible	-develop reference points if possible	4	3	1	5	5	3.6
cod.27.47d20	Cat 1 shared	-develop spatial approaches to better account for stock structure -investigate the significance of spawner age on reproductive potential -investigate perceived catchability problems in IBTS surveys (age reading issues as well as emmigration?) -investigate the possibility of including recreational catches	-develop spatial assessment approaches to better account for stock structure	-explore potential biases in the forecast and how to deal with these	3	4	5	5	1	3.5
pok.27.3a46	Cat 1 shared	Stock definition – The North Sea saithe stock is influenced by migrations to and from the North Sea. This can potentially lead to the observed year effects in survey indices. It needs to be analysed if the inclusion of spawning	Variance by age – The last inter-benchmark for saithe in 2019 revealed that uncoupling of the variance parameters for the observations by age (i.e. age 3 receiving a separate parameter) could improve the model fit statistics (e.g.	The effect of the current low productivity regime of the stock (i.e. lower recruitment) on reference pooints should be investigated.	3	4	5	4	3	3.6

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
		grounds north of 62°N could improve the assessment. Planned tagging studies may also aid in this. New survey indices – IMR-Norway has set-up a new hydro-acoustic survey targeting spawning aggregations in Quarter 1. Germany has also participated in this survey in recent years. The inclusion of this survey in the assessment should be evaluated once a sufficiently long time series has been developed. The use of the Norwegian summer accoustic survey (NORACU) - formerly dissmissed during the 2016 benchmark on the ground of (now corrected) inconsistencies - should also be re-evaluated. Catch-per-effort index – The current commercial CPUE index is standardized for area and engine power effects and is not able to account for spatial and temporal effects interactions. The inclusion of alternative explanatory variables (e.g. vessel effect) and spatio-temporal effects should be evaluated.	log-likelihood, AIC). This should be investigated further. CPUE index - issues exist on the calculation method / model. Improved methods exist for deriving yearly indices in the CPUE model. The fix maturity ogive assumption should be re-evaluated, especially in the light of improved sampling during the spawning season (Q1 acoustic survey). Survey Index - time series has been updated using new ALK-matching methodology								

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
had.27.46a20	Cat 1 shared	Explore combining survey indices. Derive time-varying maturity estimate. Derive estimates of mean weights at age for stock. Investigate indices of reproductive potential and methods to use them in management advice. Explore stock id and structure, using otolith micro-chemistry, tagging data, and the spatial range of genetic data. Ensure consistency in catch data used in assessment and advice sheet (SOP issues in InterCatch data).	Investigate poor fit in plus group in view of increasing relative importance of this age class. Investigate alternative models which are compatible with high performance computing (simulation runs). TSA shows some bias in prediction errors for Age 0 IBTS Q3 survey. TSA support likely unavailable after 2021/22 so need to consider alternative models. Exploratory assessment model SURBAR – develop likelihood profiling for ad hoc parameters, and catchability estimation model based on catch curves. If TSA is retained, an objective criteria are needed to decide if a year class is significantly large to warrant special treatment in TSA. Alternatively, some exploration of modelling techniques for sporadic recruitment is needed (mixed distributions etc).	Investigate extent of cohort effect on growth rate. Ensure consistency between catch components for weight at age cohort modelling. Investigate intermediate year recruitment assumption. Forecast value for recruitment would benefit from including information on the probability of large year classes occurring.	3	4	5	2	3	3.4
nep.27.4outFU	Cat 4 PA	Data from the Dutch landings and discards length sampling programme from 2015 onwards contain errors due to issues with processing codes and need to be resubmitted to InterCatch. On the basis of the revised sampling data,	No changes to the assessment are anticipated	No reference points have been determined	3	3	3	5	5	3.4

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
		raised discards will then be recalculated.								
nep.fu.32	Cat 4 PA	Sampling of trawl catches by the Norwegian coast guard should be improved by sexing individuals and sampling discards and landings components separately to enable discards estimations. An UWTV survey should be carried out to explore and map distribution and density	Assessment methods for data poor stocks should be explored		3	4	3	5	2	3.4
nep.fu.5	Cat 4 PA	Data from the Dutch landings and discards length sampling programme from 2015 onwards contain errors due to issues with processing codes and need to be resubmitted to InterCatch. On the basis of the revised sampling data, raised discards will then be recalculated. Also, the individual mean weights in landings and discards will be recalculated.	The assessment is based on the harvest rate estimate in relation to the MSY proxy of 7.5%. With the revised discard rates and mean weights, the harvest rates from 2015 will be revised, with potential impacts on the next advice due in 2022.	No change to the reference point is anticipated	3	3	3	5	5	3.4
nop.27.3a4	Cat 1 shared	Investigate size-at-age and derived weight-at-age in how it affects model estimation in terms of sampling accuracy and precision achieved under the current design and the most statistically rigorous way to impute values for years where these data are missing or in question.	Investigate retrospective patterns in the assessment among other in relation to the Mohn´s Rho values for recruitment, SSB and F. Introduce procedure in SESAM to make one-out-standard analyses of tuning time series.	The consumption amount of Norway pout by its main predators should be evaluated in relation to production amount in the Norway pout stock under consideration of consumption and production of other prey species for those predators in the ecosystem.	3	4	3	2	3	3.2

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
		There are currently two recruit indices (age 0 from SGFS and EGFS) being used in model parameter estimation. To avoid duplicative information being introduced into the assessment, a method should be developed that combines the Scottish and the English indices into a single robust index. In general GAMM analyses should be conducted to explore further integration of survey time series. Investigate error variances of the data that concerns sampling mechanics, sampling theoretics and sampling designs for both fishery-independent data, and for those obtained from the fleets.	Develop additional standard diagnostic tools for performance for the new SESAM model: (i) a better format for displaying and interpreting standardized model residuals over time (the bubble plots are horizontally compressed and very difficult to read and interpret); (ii) performance statistics based on prediction skill (e.g., how well does the model predict when a data point is removed?); (iii) likelihood profiles (if there is tension in the model, where does it occur?); (iv) some depictions of any gradient problems that may exist; (v) summary tables with AIC/BIC values for models using the same data (i.e., documentation of all intermediate models tested before arriving at the final choice of parameter coupling); (vi) statistics for model goodness-of-fit.	This has implications for setting of Blim levels. Sensitivity runs on the assumptions of time invariant growth, maturity and natural mortality may need to be considered. For the short term, projections that include different ways to handle mean weight-at-age, including projecting forward with specified uncertainty, should be more fully explored (smoothed historic time series, average over some recent time period, etc.).					
sol.27.4	Cat 1 EU	- Explore data giving rise to larger discards estimates for fish aged 6+	- Investigate retrospective patterns appearing in 2019	-validate RCT3 method	3	3	5	4	1	3.1
gug.27.3a47d	Cat 3 No TAC	- investigate ways to raise discards for métiers with zero landings but no discards reported - investigate potentially better ways to deal with the "generic gurnard grouping" problem for	- exploratory SPiCT model	- investigate the use of rfb, chr HCR	3	2	1	5	5	2.9

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
		some nations (e.g. Germany and the UK)								
lem.27.3a47d	Cat 3 PA	The erroneous length data submitted to InterCatch for 2013 also needs to be corrected. Further work may indicate an alternative method of collating the survey data that could be more appropriate for lemon sole.The current survey indices used for North Sea lemon sole are not able to track cohort strength on a consistent basis, and they exhibit generally poor catchability characteristics which limit the reliability of the advice based thereon. It would be very beneficial to be able to include commercial catch data in the assessment in order to improve reliability and reduce variability.	A new method of estimating age-based survey catchability coefficients is needed to help to address the problem of negative Z estimates. Age data are lacking from commercial catch data, so a (spatial) length-based assessment using both catch and survey data should be explored (for example, Stock Synthesis 3).	Reference points are currently based on length-based indicators, and further work could help derive more robust estimates. If a length-based assessment can be developed using commercial and survey data, a full stochastic forecast method should be explored.	3	3	1	5	2	2.9
bll.27.3a47de	Cat 3 PA	- Investigate the availability of more data on this stock (including discards and BMS landings or historical catches); - Explore the availability of more appropriate tuning fleets (both commercial and survey) or revise the current biomass index series (cfr. Tur4 assessment); - investigate biological parameters	- Explore whether other assessment methods can be used (Spict/SAM).	-calculate reference points based on any new assessment for the stock	3	3	2	2	3	2.8

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
		- Investigate how the biomass index should be corrected for technological creep (Dutch fleet has an increasing amount of pulse trawlers compared to the beginning of the series, who switch back to beam trawl in the most recent year);								
tur.27.4	Cat 1 EU	-The available scientific surveys (SNS and BTS-ISIS Q3) have a low internal consistency especially for older ages leading to a low ability to track cohorts over time. - Estimates of discards are available (e.g. Dutch discards are available for 1999-present), however, age-length information is very limited. - More work needed on obtaining LPUE data from other Member States, given the heavy reliance of the assessment on the Dutch LPUE data. - A detailed analysis of delta GAM indices with various settings should be carried out once more age information becomes available. -alternatives to smoothing of mean weights-at-age from the fishery to be investigated	- The over-reliance of the assessment on a single LPUE time series is potentially a problem that may need further investigation, for example by using CVs associated with the estimated index directly in the assessment. - Investigate the use of a more appropriate selectivity in the assessment to construct a model-equivalent index for LPUE	- uncertainty in recruitment and forecast is based on landings instead of catches.	3	3	2	2	2	2.7

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
fle.27.3a4	Cat 3 No TAC	- investigate ways to raise discards for métiers with zero landings but no discards reported - investigate ways to raise discards for shrimper fleets operating in coastal waters for which no suitable data are available	- Investigate what could be done/changed to improve the SPiCT model (e.g. include effort data) - Investigate the use of alternative stock indices (DYFS, DFS) which are able to better reflect the stock status -other stock indicators available? e.g. WFD monitoring from coastal areas	- Investigate available growth darta and use of rfb and chr HCR	3	2	1	5	2	2.6
wit.27.3a47d	Cat 1 MSY	- no issues currently	-The choice of proportion of fishing mortality and natural mortality before spawning (Fprop and Mprop) to be equal to 0.5 should be evaluated for its biological reasoning.	- The calculation of reference points is based on the whole time series (1940 - 2016), which includes the period before the data start (1940 – 1949) and the period where catch is the only available information (1950 – 1982). The adequacy of the assessment to estimate SSB and recruitment during that period should be evaluated, especially concerning their use in estimating reference points.	2	3	3	2	3	2.5
ple.27.7d	Cat 1 EU	- evaluate FR GFS index, remove potential vessel affect from the data (possibility of splitting the time serie of the index) - there was a lack sampling during CGFS 2020 (stations in UK waters were not sampled)	- test new index produced and evaluate its impact on survey residuals and the assessment - test new maturity ogive and Q1 removal investigate the important decrease of recuitment in 2020	- no issues currently	3	3	5	2	3	3.1

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
		- investigate if new maturity data are available and useable - data required to update Q1 migration - lack of information of weight at age for ages 1 and 2 for Q2 - Fix the coding issue of CGFS index (update the calculation of the index)								
whg.27.47d	Cat 1 shared	-stock identity (SURBAR runs by component, not an issue yet) -historical stock weights at age re-estimated every year (reconsider if significant changes in historical time series, not issue yet) -include natural mortality estimates (WGSAM) when available (not an issue yet) -DATRAS indices (new French data upload for historical series), exploration of delta GAM method for index calculation	- impact of new 2020 SMS keyrun (WGSAM, 2021) estimates of natural mortality on assessment model: SSB retros just within acceptable limits defined by WKFORBIAS -use of unsmoothed maturity and natural mortality estimates as input (using the new SAM method to estimate missing historical values and forecast)	-further investigate alternative SAM forecast (recruitment assumption, split of catches) -Reference points estimated with EqSim with the new survey indices are slightly different from the ones estimated during the 2018 benchmark.	2	2	4	2	2	2.2
dab.27.3a4	Cat 3 No TAC	- investigate ways to raise discards for métiers with zero landings but no discards reported - investigate ways to raise discards for shrimper fleets operating in coastal waters for which no suitable data are available - Investigate extending the delta-	- Investigate the use of DYFS, DFS inshore surveys to estimate a recruitment index - Investigate which effort data are available and if these could be used as further input for the SPiCT model	- investigate HCR from WKLIFE X	3	2	1	1	2	2.2

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
		GAM index with Belgian and German BTS data (prior to 2002).								
sol.27.7d	Cat 1 shared	- further investigation of the perceived subpopulations; - investigate the mechanism of the declining trend in weights-at-age; - investigate natural mortality and maturity	none	none	1	2	5	4	1	2
nep.fu.10	Cat 4 PA									
nep.fu.33	Cat 4 PA	Data from the Dutch landings and discards length sampling programme from 2015 onwards contain errors due to issues with processing codes and need to be resubmitted to InterCatch. On the basis of the revised sampling data, raised discards will then be recalculated. Also, the individual mean weights in landings and discards will be recalculated.	The assessment is based on the harvest rate estimate in relation to the MSY proxy of 7.5%. With the revised discard rates and mean weights, the harvest rates from 2015 will be revised, with potential impacts on the next advice due in 2022.	No change to the reference point is anticipated						0
nep.fu.34	Cat 4 PA									0
nep.fu.3-4	Cat 1 EU									0
nep.fu.6	Cat 1 EU									0
nep.fu.7	Cat 1 EU									0
nep.fu.8	Cat 1 EU									0

stock	Type	Benchmark Issues			Scoring Categories					Total
		data and stock ID	assessment	forecast and reference points	1	2	3	4	5	(weighted)
nep.fu.9	Cat 1 EU									0