External Review of the Norwegian Institute of Marine Research Stock Assessment Processes

Final Report

W. Karp (chair), J. Ianelli, A. Rindorf, and G. Stefannson

March 5, 2020

Contents

Executive Summary

Overall the panel found the information for the Stock Assessment Review to be extremely well organized and well documented. The presentations were excellent and thorough with only a few lacking in detail. The presenters were clear and well-prepared to answer our questions. IMR should be commended on the quality of its data collection and the accomplishments and scope of their stock assessment processes. There is, nevertheless, room for innovation and improvement, as might be expected in any enterprise of this kind.

1. Is the **scientific/technical approach** to stock assessment modelling appropriate?

   IMR scientists employ a range of stock assessment models and, in general, models selected are appropriate for the stocks being assessed. For stocks assessed through ICES, the benchmark process provides a mechanism for periodic review of assessment methodology although the frequency of benchmarks varies. Some assessments would benefit from methodological updating, including the data-poor stocks and stocks using methods not accounting for observation error, but in most cases, the presenting scientist reported that a plan for this was in place, indicating knowledge of the important challenges and that work is continuously ongoing to improve assessments. Emphasis on data aspects which drive assessment models is given a high priority at IMR and improvements on this front are proceeding well (see below).

2. Is the assessment process **efficient**, effective and clearly described?

   In most cases the assessment process is efficient, effective and clearly described in the stock descriptions. However, availability of fishery-dependent and fishery-independent data can be problematic in assessments with short time spans between sampling and delivery of advice. For example, catch advice for capelin depends on survey data that is complete only just before the fishery opens, increasing the risk of errors not being discovered in the rushed process to produce advice. In several instances there is a one-year lag in availability of some fishery-dependent data and delays in age determination appear to be problematic (additional staff are being made available to address this need). Stocks that are assessed under the ICES system are generally well documented. For stocks that are assessed within IMR, ICES standards and timing requirements for documentation should be followed whenever possible (see further comments below). The stock descriptions should ideally be expanded in the future with information on how each assessment conforms to the best practice guidelines developed under REDUS.

3. Is there an adequate **peer review** of the stock assessment process?

   As part of the ICES system, the assessment methods receive peer review during benchmark meetings. However, an overview of the timing of the latest benchmarks and the next expected benchmarks was not available. For the review process to be effective and transparent, regular schedules for benchmarks need to be made and published. For example, the capelin assessment seems due for a benchmark review with appropriate planning for upgrading the methods that might be applied.

   An important part of peer review is auditing, including verification of earlier assessments and the data used for these. With the new projects, including Sea2Data, IMR staff can (or will be able to) obtain snapshots of earlier data sets and settings, exactly as used in earlier assessments. This seems to be well facilitated with the version control system built into Sea2Data.

4. Is the **organizational structure**, staffing and funding, sufficient for generating efficient, timely and high-quality stock assessments?

   IMR's organizational structure supports an efficient and effective stock assessment process. Internal communication seems to work well in general. Some specific challenges may be related to geographical separation between small groups of scientists working on similar stocks, making it difficult for these individuals to work together. Similarly, even though IMR is appropriately well-recognized for technological innovation (e.g. acoustics and survey gear/technology), it seems that communication between scientists working in these areas and the stock assessment scientists is sometimes lacking. These concerns can be mitigated to some extent through clearer guidance from leadership. Assigning 'liaison' scientists who are tasked with facilitating communication between localities or groups with different types of expertise may be helpful. These individuals should bring a broad knowledge of relevant activities within IMR.

   Funding appears to be sufficient and allocated appropriately to support assessment needs although some assessment scientists expressed concerns in this area. Availability of funding generated through fees from landed value of the fishery can help focus and prioritize assessment work. There are regular meetings to prioritize the funding used, which allows IMR to address upcoming challenges quickly and efficiently. In addition, internally available funds can be prioritized for use in areas which the fishery stakeholders find less pressing.

   Overall, staffing appears to be sufficient to address assessment needs, especially now that additional resources are being made available for age reading and other support tasks. However, staff highlighted the lack of statistical and assessment modeling expertise within IMR and indicated that retaining these skills within the institute was difficult. This constrains innovation and methodological development. To some extent this deficiency is addressed through contracting (with, for example, the Norwegian Computing Center). We suggest that developing this expertise in-house would be advantageous and having better technical capacity for assessment models, preferably including individuals with appropriate background in biology and fishery data issues would improve the assessment quality. A dedicated effort to ensure collaboration between modelers and methods developers through face-to-face meetings could improve the scientific environment in this research area.

5. Does IMR achieve adequate **assessment accomplishments** relative to mandates and the needs of managers for scientific advice?

   In general, the scientific work that leads to management advice is based on sound research that is published in refereed journals. Whether assessments are conducted under the ICES process or independently, this process results in scientific advice that appears to meet management requirements within the realm of fisheries management based on the precautionary and MSY principles. The stated management goal of optimal economic yield for some stocks is not addressed at present in assessments, as this goal is not shared by ICES and hence does not appear in annual advice tables for shared stocks (see additional comments in Recommendations section).

6. Does IMR have an effective process in place for taking **ecosystem and climate change** factors into consideration in the stock assessment process?

   IMR scientists conduct world class research on ecosystem processes and climate change. There is, however, no formal process for integrating the results of this research in stock assessments. This challenge is common throughout the world and, to some extent, environmental influences are captured through existing stock assessments via rigorous monitoring programs. In several cases, assessment scientists discussed plans for taking climate-change factors into account in providing advice. It is important to note that ecosystem interactions will vary markedly across species groups. Some applications of the model “Atlantis” were shown at the meeting, demonstrating that such a complex model can provide responses to what-if questions (e.g., the effect of snow crab on the biomass of cod, haddock and capelin). The Panel noted such an approach has merit for strategic (i.e., broader long-term) advice to industry and managers.

Ecosystem considerations and climate change are major aspects of IMR’s review of survey priorities. For example, areas where fish distributions have shifted are considered and collecting broader ecosystem data has been given increased priority. As mentioned above, monitoring of individual species as they respond to ecosystem and climate change can be effective. An example of this was provided in the harp seal assessment where the scientists observed changes in blubber content and the potential relationship to pup production. Similarly, changes to mean weight at age, maturity, recruitment success and natural mortality are considered in most assessments, and hence historic changes in productivity are accounted for in advice. However, it is not clear in all cases whether individual assessments have considered such changes.

Communicating ecosystem processes could be improved and included within the assessment and advisory processes. This could be accomplished by assigning coordinators to compile ecosystem considerations that are taken into account in the advisory process. Through this process, diverse scientific expertise can be brought to bear (see this link for an example of how this is accomplished by NOAA for Alaska groundfish fisheries).

The Panel noted that work in this area seems to focus on the impact of ecosystem conditions on fish stocks rather than the effect of fishing on the ecosystem as no reference was made to non-target species and habitats impacted by fisheries.

7. Does IMR adequately engage stakeholders in the stock assessment process and communicate assessment-related results, needs, and research to them effectively?

   Communication to managers and fisheries representatives is achieved through regular meetings organized by the directorate, targeted communication with reference fleet fishers and fisheries representatives. Stakeholders are also invited to ICES benchmarks. In addition to this, scientists responsible for a number of stocks communicate by meeting with fishers in the regions where they conduct their fisheries.

   IMR scientists and leadership also meet with industry organizations on a regular basis. Furthermore, the committee that establishes priorities for use of funds derived from fish landings include industry representatives.

   It was unclear what steps are taken to engage with the broader stakeholder community, including NGOs.

8. Are fishery-dependent and fishery-independent data sufficient in quality, quantity and timeliness to support all stock assessment needs?

   IMR conducts a substantial number of stock and ecosystem surveys using a variety of methods including acoustic/trawl, bottom trawl and pelagic/surface trawl. Active and productive research programs support innovation in these areas. For most stocks, at least one time-series of survey data is available; in some cases, several such data sets are available. Within IMR the process for prioritizing surveys and dealing with survey operational issues is well-established and appears to be effective. Nevertheless, accurate survey data are not available for all stock assessments as all species are not monitored reliably by the existing surveys. For several stocks, this could be addressed by developing new surveys and survey methods.

Fishery-dependent data are obtained from the reference fleet and from catch/delivery sampling. An innovative probabilistic process for sampling catches appears to have increased efficiency and reduced overall sampling needs and this has been coupled with a self-sampling scheme for fishers. Nevertheless, delays in making catch data available for assessment scientists can be problematic, especially from some countries Investment in additional age readers has been effective in improving data timeliness.

It was not clear if data from the reference fleet was representative of commercial catches and specifics regarding the type and extent of data provided by this fleet was not discussed. IMR has documented that the major source of catch variability is between vessels and trips, so it is important to understand the biases introduced by basing catch information on a very limited number of vessels. Also, this fleet seemed appropriate for CPUE analysis but is generally excluded. When CPUE data are included, technological changes were generally ignored.

Development of greatly improved systems for managing data and analyzing survey data (Sea2Data and StoX) are very promising and already beginning to bring improvements in consistency and transparency to some stock assessments. IMR should strive to employ these tools to support all assessments and ensure full compatibility with ICES data systems.

Here we also note the impressive list of courses offered by IMR’s “Academy”. While primarily focused on data collection and survey procedures (which is essential relative to assuring data quality), they also offer training opportunities in a broad range of related areas of expertise. Encouragement for IMR scientists to ‘teach each other’ should have beneficial effects not only by increasing the knowledge level but also by increasing the awareness of expertise within the institute, hence potentially improving internal collaboration.

9. Is **uncertainty** associated with data and analytical methods properly characterized and included in the assessment results?

The extent to which uncertainty is considered and incorporated into assessments varies by stock and, in some cases, modeling approaches preclude full consideration of uncertainty. The lack of in-house statistical and assessment-modeling expertise may be an impediment to advancing in this area. Furthermore, data quality (including catch data) influences uncertainty and this may not be recognized. Consideration of structural uncertainty may be lacking and, in some cases, the observation error may be specified as known and correct. A specific challenge (in IMR and internationally) is the correct estimation of uncertainty for stocks where the abundance is increasing at a constant rate unrelated to catches.

The notion of benefits in terms of greater yield and accuracy by reducing uncertainty in stock assessments as a general principle and as specified by the REDUS project may be misleading. Generally, there are cases where “improved” assessments explicitly account for all types of uncertainty. As such, overall estimated uncertainty may increase as the specification of this becomes more correct. A more focused approach would involve comprehensively accounting for process and observation errors such that uncertainty estimates are more reliable (as opposed to being reduced). The more realistic uncertainty estimates of e.g. survey indices can be used as realistic lower boundaries for uncertainty in stock assessment models. This is, perhaps, a more appropriate goal and differs from approaches directed towards improved survey precision, e.g. by increasing the number of stations.

A method of estimating uncertainty, using a bootstrap approach on blocks of data aggregated within areas was presented. Such new methods of estimating uncertainty and carrying that uncertainty all the way through the assessments into the eventual quota need to be tested. For example, it is not obvious how the (a) choice of spatial units for aggregating the echo abundance will affect the perceived or real accuracy of estimates and it is not clear what the net effect will be of (b) increasing the number of transects, nor the interaction between (a) and (b). These issues can potentially be investigated using simulated transects across a smoothed population.

10. The panel recognizes the importance of the REDUS project in developing an integrated, thorough and well-documented approach to the overall assessment process. We see benefits in extending this approach to all assessments. Linked to this is the Assessment Best Practice Guide, another tool and work-in-progress which, when refined, should be used as part of the assessment process. Are there opportunities for improving stock assessments and the stock assessment process?

    Almost all the presenting scientists recognized that there are opportunities for improving stock assessments and the overall process and demonstrated comprehensive knowledge of their stocks in their suggestions for further work. The new tools and processes such as Sea2Data, StoX, REDUS (and the Best Practice Guide) will lead to overall improvements in consistency and quality and we encourage further support and implementation of these projects. Documentation of analytical and research needs should be completed for each of the stock overviews. This will be helpful in cataloging overall needs, setting priorities, and improving assessments.

    Analytical innovation is essential to improved stock assessments and many staff identified this challenge. As noted above, developing in-house expertise would encourage exploration and development of new modeling approaches. We also note that evaluating sampling priorities for both survey and catch data may be helpful in considering research redirections and in testing models.

Preface

In the evaluation, the panel assumed that the variety of stocks shown in the oral presentations were representative of the methods and quality of the other stocks.

Direct responses to the terms of reference are provided in the executive summary above. The sections below follow the meeting agenda and provide general observations and comments from the panel, followed by detailed recommendations.

Assessment Infrastructure

Here we consider a number of projects and activities designed to improve data quality and data flow, integrate assessment processes and address challenges related to properly estimating uncertainty and propagating uncertainty through the assessment process. These include the related Sea2Data and StoX projects, the REDUS project, and the Best Practice Guide. These projects are designed to improve various processes and facilitate internal collaboration. They also lead to better planning and organization of key functions within IMR. To some extent, however, they duplicate similar processes being developed and implemented by the international community (through ICES) even though IMR is implementing some enhancements and innovations that are not currently available elsewhere. The extent of international collaboration in development and integration of these capabilities is unclear. Early inclusion of external research scientists and full integration with similar ICES systems will increase the likelihood of attaining results that are widely applicable and acceptable.

Sea2Data and StoX

Sea2Data is an initiative designed to automate, organize and integrate functions related to assimilation and management of survey data and catch statistics. It provides for a high degree of quality control, greatly improved data management, and consistent interfaces for reviewing, combining and extracting data. Sea2Data interfaces directly with StoX, standardized survey estimation software, which calculates indices and associated variances for acoustic/trawl surveys, trawl surveys and fishery-dependent surveys (this terminology is unclear). Versioning built into these systems seems powerful and allows tracking of changes\corrections in the data. This also serves to track underlying decisions in preparing survey estimates for stock assessments and will help understand changes and data corrections if needed over time. The products derived from this type of project are essential to effective and efficient stock assessment and should be considered a “necessary” activity whether it is undertaken by an individual institution or through broader international collaboration.

IMR researchers expressed concern that Sea2Data/StoX establishes something of a “straightjacket”. This is because the dataflow and estimation methodology is, to some extent, hardwired into the system. Furthermore, providing data to researchers through APIs which access data in XML format is relatively uncommon. Many modern researchers access raw databases using packages (such as dplyr) in R, typically reading directly from the databases for efficiency. In other cases, institutes provide the entire databases in the form of R data structures. This capability is appears to be provided in Sea2Data and Stox but should be developedand documented further; R interfaces could be written to allow researchers to easily access the types of files normally accessed from the databases. These R-interfaces and/or of R data structure versions could then be stored to ensure history storing and transparency.

The importance of open access in these systems (or transparency) should be emphasized and should include documented code as well as data. We recognize that transparency for fishery data may not always be possible, but it should be possible for surveys and will likely increase interest from external modelers who wish to collaborate.

StoX can currently provide estimates from acoustic and bottom-trawl area-swept data using design-based approaches (e.g., stratified random-sampling). Alternative estimates can be compared. However, the facility to test these for optimality (within and over different species) is unclear. Therefore, the Panel **recommends that a simulation framework be developed to evaluate survey data.** For example, Kotwicki and Ono 2019 developed an approach for generating population data (conditioned on survey observations) to provide realistic simulated samples for testing alternative survey designs.

REDUS (Reduced Uncertainty in Stock Assessment) Project

This is an ambitious project which is nearing completion. It draws heavily on the tools described above (Sea2Data and StoX) and addresses multiple sources of error and issues associated with data quality. Management Strategy Evaluation (MSE) approaches are included in this work to a limited extent (see below). This project is also important as a process for improving coordination, integration, and overall assessment quality within IMR.

The panel noted that the concept of reducing assessment uncertainty may be misleading. Experience suggests that improving assessments requires a more complete consideration and accurate estimate of uncertainty (which often can result in increased uncertainty). We think that improving assessments means providing advice that clearly estimates the uncertainty and relative risks.

The panel discussed how the REDUS work products (WP1-WP4) differ from necessary tasks of any agency or group providing science advice for fisheries management. Also, it was noted that once established, generic implementation of these may be challenging. However, the panel noted that the project as designed has intrinsic value by highlighting where improvements can be made.

The panel noted that developing the capacity for comprehensive MSEs is beyond the scope of this project even though this should be considered a priority area for investment by IMR.

The panel noted that since “Optimum Economic Yield” is a priority for many Norwegian fisheries, it would be helpful to indicate how this requirement is recognized within the REDUS project. Similarly, some recognition and discussion of the linkages between uncertainty, risk evaluation, and precaution should be provided.

Uncertainty in survey indices is complicated by the likelihood that some underlying assumptions are violated. For example, the population fraction accessible to survey gear may vary by year. Additionally, the temporal and spatial coverage of surveys is generally more limited than for the commercial fishery. Increasing survey efficiency is stock-specific so the multi-species nature of most surveys is constraining.

While communication with stakeholders is incorporated in this project to a limited extent, the panel noted that ongoing and meaningful involvement of stakeholders throughout the process is essential and should be emphasized.

Implementation of the REDUS framework for the two demonstration stocks appears to be near completion. Following thorough review, this framework should be adapted to include all assessed stocks.

Best Practice Guide

This is a useful guide and should be implemented to the extent practicable for all assessments. The guidelines themselves and individual checklists should be date stamped. The panel noted that in Table 4, growth information, e.g. weight in the stock or weight in the catch should be included. While these are observed historically, they also need to be forecast, and this can be a major source of uncertainty; a best practice should be to evaluate the relative impact. Data processing and assessment model code should be checked minimally at benchmarks but ideally also through technical consultations with colleagues tasked to scrutinize these aspects.

Survey Planning and Logistics

IMR has a substantial fleet of seven research vessels and is able to charter commercial vessels under multiyear contracts. While it is not possible to provide dedicated surveys for all stocks, surveys conducted by IMR or through collaboration with other countries supports a noteworthy and comprehensive effort. IMR's processes for setting priorities, managing the fleet, and dealing with unforeseen breakdowns and needs appears to be thorough and effective. Like most similar institutions, IMR is faced with "mission creep" associated with increasing demands for data and sea time but resources appear to be sufficient to address priority needs. Comments regarding survey design, processing of survey data, etc. are provided elsewhere.

Fishery Dependent Data

The panel noted that the ECA program used for estimating catch at- age seems quite useful, the documentation is quite complete and also has been brought into an R package. This application is not implemented for all data sets and some ad-hoc processes for filling sampling gaps and otherwise dealing with data limitations were reported.

Fishery dependent data from multiple sources must sometimes compiled and delays in data delivery from some fishery participants were also noted. Data compilation and interaction is carried out both by IMR and ICES; approaches seem to be generally consistent, but this should be monitored carefully.

Representative sampling of catches is often problematic and IMR has ongoing research to address this concern. The lottery-(self) sampling approach appears to have merit in this regard and sampling by the reference fleet is an important source of fishery-dependent data. Nevertheless, the lack of on-board observers in Norwegian fisheries constrains catch sampling. This is especially problematic because accurate documentation of discard is lacking, because a major contributing factor is that regulations prohibit discarding in many fisheries. Port sampling of retained catch may also be inadequate when landings occur at multiple isolated locations and this could introduce further bias in catch estimation. Spatial and temporal limitations in catch sampling should be monitored carefully and addressed fully to ensure representativeness and for picking up on spatial differences in growth or age composition (both of interest when considering spatial management).

Sampling by the reference fleet is an essential source of high-quality catch data. During our site visit, very little information about the functioning of the research fleet was presented. For example, the concept of using the research fleet to develop CPUE indices as inputs for assessment was raised by panelists on several occasions but did not stimulate a serious discussion. We suggest this be pursued further.

Age determination can be time consuming and requires special skills. Several presenters noted delays in availability of age data as well as inconsistencies in age data from different (national) fleets. Ongoing work in intercalibration and technique development should continue. It seems that new age readers have recently been hired by IMR and this should be beneficial.

General Observations

The need for developing in-house assessment modeling and statistical expertise is mentioned elsewhere in this report. It is evident that some IMR assessment scientists view assessment and MSE software packages as 'black boxes' requiring experts in modeling to avoid improper interpretation of settings and/or results. Consequently, some scientists are unable or unwilling to consider methodological improvements. This concern was especially apparent regarding unclear expectations for data-limited stocks (i.e., that success is only measured if a full age-based assessment is conducted).

IMR is encouraged to become a leading player in data analysis (as they already are in acoustic analysis) and statistical modeling of data and stock assessments. This would accommodate development of models tailored to long-lived species (including marine mammals), invasive and steadily increasing species (where traditional models tend to produce poor results), forage species (informed by ecosystem information), as well as deep sea and data limited stocks, in particular where these are caught in single species fisheries.

The panel (and IMR staff) noted that better understanding of factors that influence productivity is a necessary aspect of the ecosystem approach. Such factors could affect short term predictions as well as the expected long-term changes. IMR often takes changes in productivity into account when characterizing the historical development of stocks, perhaps not always recognizing that this is an important aspect of EAFM. Work that considers how fishing has affected the ecosystem should be developed further.

Comments on individual stock assessments: pelagic stocks

Pelagic stocks presented and discussed during this session included Norwegian Spring Spawning (NSS) Herring, Blue Whiting, North Atlantic Mackerel, Capelin and Sandeel

NSS Herring

This is an important stock for Norway and is assessed within the ICES framework. Even though other nations participate, the assessment, conducted by the ICES Working Group on Widely Distributed Stocks, WGWIDE, is led and coordinated by Norwegian scientists.

Survey coverage is extensive and includes the Norwegian acoustic survey on spawning grounds (Ages 3-12+; February/March) and the International Ecosystem Survey in the Nordic Seas (Age 2 from Barents Sea; ages 3-12+ from Norwegian Sea). Concerns regarding temporal and spatial changes in survey availability should be followed carefully.

Catch sampling intensity varies markedly among the countries that harvest this stock. Catch-at-age estimation (and estimation of error structure) is carried out using the ECA method but this is applied only to Norwegian data (Norwegian catch exceeded 55% of the total for the example presented). Norwegian catch sampling uses the catch sampling lottery approach which claims to achieve close-to optimal sampling through a probabilistic method that distributes sampling effort by time, area, and gear type. Questions remain regarding the consequences of other countries using different sampling schemes. Further, similar schemes in other species are often ground-truthed by samples taken by control or observers, ensuring that the samples taken by the fishers are representative of the catch.

Catch varies substantially by season. Largest catches are generally taken in Q4, followed by Q1. Temporal shifts in harvest patterns may substantially affect estimates of F, M and mean weight at age as the assessment model uses an annual time step. Such changes may also impact realized fishing mortality, causing this to differ from the recommended level.

Extensive RFID tagging has been carried out in recent years, but detection systems are only partially deployed. This approach holds promise for improved estimation of migration patterns and important life-history parameters and should be advanced to the extent possible in the near future.

Survey age-reading methodology varies among the countries involved (otoliths vs. scales). This should be resolved to reduce error and uncertainty (note the following extract from the assessment summary document: "*There have been concerns regarding dissimilarities between age distribution from the different nations participating in the IESNS survey. For NSSH some nations use otoliths while others use scales for age reading. A workshop in 2015 concluded that the different ages obtained from scale and otolith readings of the same fish could be due to several issues relating to identification of the first winter ring and age interpretation of older fish, additionally confounded by stock mixing issues. Final conclusions could not be reached based on the samples from this workshop. Since the problem could not be resolved it was suggested to explore of effects of ageing-errors on the assessment of the stock size by using the variance-covariance matrix from the latest age reading workshop to introduce uncertainty in age reading in the assessment model (XSAM). This will be done before the next WGWIDE in 2020.*"

Assessments have been carried out using the XSAM model since the 2016 benchmark. Model development has been supported through a substantial contract with the Norwegian Computing Center. This implementation of catch-at-age modeling takes into account multiple fleets which cover different groups of ages and different fishing areas. Retrospective analyses indicate low values for Mho's rho.

If age-reading error structure is included in the assessment model, it should be estimated using both age-at-length keys and length distributions. Otherwise high uncertainty in length groups with few individuals will lead to overestimation of age reading uncertainty. It should also be recognized that age reading errors is only one of the ways in which correlation between age groups in catches may occur. Other possibilities include correlations in size dependent mortality and autocorrelation in recruitment.

Overall this assessment appears to be appropriate, well-designed and effectively implemented. The work (in Norway) is well-funded although additional staff resources would be helpful. Peer review (under the ICES system) is good and engagement with stakeholders seems to be sufficient. The assessment model includes some aspects of uncertainty making it possible to explore impacts of e.g ageing error. Exploration of climate and ecosystem considerations is lacking and should be implemented to link the large body of work focusing on understanding the biological processes in this stock to stock advice.

Blue Whiting

Norwegian harvests of blue whiting are substantial (438 kt in 2018 out of a total 1711 kt). The stock is assessed by ICES through the Working Group on Widely Distributed Stocks (WGWIDE) and the assessment is led by Denmark. The primary source of fishery-independent data is the (acoustic) International Blue Whiting Spawning Stock Survey (IBWSS). Norway participates in this survey together with Ireland, the Netherlands, and the Faroe Islands. Abundance indices derived from the survey are provided to the assessment as a tuning index using StoX. A second tuning index estimated from data collected during the International Ecosystem Surveys in the Nordic Seas (IESSNS) will be considered at the next assessment benchmark. Since IBWSS is conducted in March/April and IESSNS is conducted in July, this can potentially provide useful information on mortality and growth.

Fishery-dependent data from all fishing countries are provided for the assessment although samples sizes from some countries are very small (i.e. The relative sampling per ton of catch varies considerably by country). Norwegian sampling uses the lottery/ERS process as described for NSS herring. Also, ECA is used to provide catch-at-age inputs for the assessment. The same caveats as mentioned above apply and it is noted that this scheme is only followed by IMR. Issues regarding intercalibration of age reading among countries should be addressed.

Since ICES procedures are followed, peer-review and communication with stakeholders is good. Funding for this work with IMR appears to be adequate but climate change and ecosystem factors are not taken into account directly in the assessment. Unresolved issues regarding stock structure are challenging and should continue to receive attention.

NA Mackerel

Mackerel is harvested by several countries in the North Atlantic. Overall annual harvest has fluctuated around 1 million t or higher in recent years, of which the Norwegian harvest constitutes about 20%.

NA Mackerel is assessed under the ICES framework by WGWIDE and Norway plays a substantial role in this process. The assessment utilizes SAM but model configurations have changed frequently following a series of benchmarks during the last 5 years. Three survey indices are used based on:

- Triannual egg survey (SSB) (1992-2019)
- International bottom trawl survey IBTS (recruitment index) (1998-2018)
- International Ecosystem Survey (IESSNS) (ages 3 to 11) (2010, 2012:2019)

Swept area pelagic survey methods are innovative but may be sensitive to catchability and availability assumptions. Furthermore, trawl and egg survey results are conflicting, and this is unresolved.

Mark-recapture data from 1980-2006 and 2014 and 2018 and catch data are also utilized in the assessment. Steel tag data has been collected by IMR since the 1960's. Concerns with consistency and quality of these data have been substantial and some years of data have been eliminated from the assessment. For groups of years where data quality is considered acceptable, SAM uses a table of data showing numbers of steel tagged fish per year class in each release year, and the corresponding numbers scanned and recaptured of the same year classes in all years after release. Recent, innovative work with RFID tags is promising but further work to refine the approach and improve the recovery process appear to be necessary. Overall, the tagging data appear to be quite influential in determining recent trends as omitting them reduced the recent stock estimates substantially. As such, there is concern about which data were selectively omitted (i.e., fish at large for longer were excluded due to concerns about tag shedding). Also, the organization of the data by cohort and ages may be problematic since individual fish lengths may be a poor indicator of true age for this species. A thorough review of tagging data is recommended, and this should document analytical approaches and criteria for including or excluding subsets of the data.

Catch data quality and reliability issues are also of concern. Catch data from before 2000 is down weighted because of unresolved issues related to underreporting. An alternative and likely more appropriate way to do this would be to estimate an 'unallocated mortality factor' for uncertain years. Problems relating to misreporting, underreporting and discarding persist. The process currently employed for applying length and weight information from samples to unsampled catches appears to be somewhat arbitrary; well-documented and more rigorous methods should be adopted (the Norwegian ECA approach may be more suitable).

The assessment process has been challenging in recent years. The issues discussed above regarding conflicting survey results, catch data uncertainties and ongoing refinement of mark-recapture data contribute to these challenges. Marked changes in distribution and productivity have been observed and appear to be ongoing. It should also be noted that stock structure questions are unresolved even though NEA mackerel is assessed and managed as a single stock.

As a result, the assessment has been benchmarked three times since 2014. To some extent this reflects the reality of changing reference points (a moving target) that are likely related to broader ecosystem changes. These are not taken into account explicitly in the assessment, but research should consider whether available ecosystem information could be used to improve e.g. growth, mortality, maturity or recruitment predictions. Such work may also provide quantitative indicators that can be brought to bear in deciding which ages and time periods to include within SAM. It would be interesting to see whether available data can be used to better-define the thermal habitat of mackerel and to use this understanding in survey design or analysis. The use of tagging data for a species with consistent age readings is unusual. While these tagging data present a wealth of information about e.g. migration and growth, it is unclear how they add to the assessment quality and if this addition is in correspondence with the costs of the programme.

While this assessment is difficult and complex, IMR has invested heavily and effectively in improving survey data and investigating alternative analytical approaches. Earlier comments regarding lack of analytical stock assessment expertise arose during the discussion of this stock. It was commented that the current model does not incorporate the bounding of survey error to that estimated from surveys. It should be noted that this should be done with extreme caution as the survey uncertainty as estimated from samples can at best only be a minimum uncertainty given that other factors such as time of survey, weather, catchability on the particular day, also add uncertainty. Technically, correlation between e.g. age group catches as a result of ageing uncertainty has been included in previous SAM assessments in other areas.

Without the presence of on-board observers, reliable information on bycatch of e.g. mammals and seabirds is lacking. It should be remembered that in such a large fishery, a low bycatch percentage of other species may still amount to a large biomass of bycatch. With the wealth of studies on NEA Mackerel, potential ecosystem information that could be used to improve e.g. growth, natural mortality, maturity or recruitment predictions may exist and should be examined. For example, an M of 0.15 for younger ages may be too low and should be evaluated.

Capelin

In the Barents Sea, capelin supports important Norwegian and Russian fisheries as well as being a key prey for a variety of natural predators. Because of the ecosystem importance of capelin, and it's unique and simple (short-lived) life history, the stock is managed according to a target escapement strategy as agreed under the Joint Norwegian-Russian Fisheries Commission (JNRFC): "The TAC for the following year should be set so that, with 95% probability, at least 200 000 t of capelin (Blim) will be allowed to spawn." This HCR has been in place since 2002 and will be re-evaluated in 2021.

The assessment relies primarily on the annual joint Russian-Norwegian Barents Sea Ecosystem Survey (BESS), a trawl-acoustic survey that takes place in September. The abundance estimate derived from this survey is considered an absolute estimate of stock size. Survey data analysis is straightforward but is subject to the normal assumptions of the methodology. Sampling uncertainty could be estimated but, instead, a fixed CV is used in the projection model.

The capelin assessment is formally part of the work of the ICES Arctic Fisheries Working Group (AFWG). However, the working group leaves the assessment to the parties responsible for the Barents Sea Ecosystem Survey (BESS), i.e. IMR and PINRO to accommodate the need for rapid processing of survey data and provision of management advice.

While the general approach seems to be efficient and effective, some improvements and advances should be considered. The rapid turnaround of survey data to provide management advice limits opportunities for thorough QA/QC and peer review. This would be exacerbated if a winter survey were to be implemented. However, using the ICES process for retrospective (post assessment/advice) peer review would be beneficial and this would allow improvements/corrections to take place in future years. The pilot project for monitoring the spawning stock in the winter should be adopted if the quality of advice is improved and the time between the survey and advice still allows proper quality control of input data and assessment. We also recommend changes in the projection methodology which would eliminate use of Excel, provide full documentation and transparency, and allow updating of parameters and, potentially, incorporation of ecosystem considerations such as additional predators. The current escapement biomass is derived in what appears to be an ad-hoc manner and should be revisited in an ICES benchmark setting. While funding in Norway to support this work is sufficient, concerns regarding adequacy of stock assessment expertise and redundancy should also be recognized.

Sandeel

Sandeel in the Norwegian EEZ is assessed first by ICES in January and updated by IMR in May. The stock is managed independently by Norway. The assessment is carried out through an explicit management plan that includes spatial components designed to allow areal management to limit or prevent local depletion. An area is closed for fishing unless acoustic abundance estimates for that area are relatively high; if an area is open for fishing, some adjacent areas will be closed.

The May assessment update is based primarily on acoustic/trawl surveys that are carried out in late April/early May. These result in absolute abundance estimates that are used to update the overall assessment and provide management advice by May 15, one month after the fishery opens. Preliminary advice is based on the previous year's acoustic estimate after factoring in estimates of mortality, growth and other life history parameters as well as the ICES stock advice in January. Low/medium quality, high uncertainty dredge surveys conducted in November/December also inform the assessment. Regardless of the final advice, TACs cannot be set lower than the preliminary advice.

Data for 2010-2019 indicate that final advice has been >= preliminary advice since this system was initiated. Final advice sometimes exceeds preliminary advice by a substantial amount and in one out of the previous three years has exceeded the ICES advice based on the January assessment. This suggests that assumptions associated with survey methodology and the projection model should be carefully evaluated. For example, survival estimates rely on an estimate of M=0.7 for all lengths and ages; this should be revisited and the sensitivity of the projection to changes in M should be evaluated. Growth estimation is also challenging and should be researched to the extent possible.

While the overall approach is innovative and pragmatic and appears to be efficient and effective, QA/QC and peer review are lacking as is a formal international review of the management strategy using peer reviewed management strategy evaluation methods. Even if the May advice is not provided through ICES, subjecting the assessment to ICES peer review (as a post assessment exercise) should be considered.

Comments on individual stock assessments: demersal and deep-sea stocks

Stocks presented and discussed during this session included NEA Cod, NEA Saithe, Wrasses, Ling, Greenland and Beaked Redfish.

NEA Cod

Northeast Arctic Cod is harvested primarily by Norway and Russia. The fishery is managed jointly by these two countries and the stock is assessed by ICES through the Arctic Fisheries Working Group (AFWG). The current stock coordinator is Russian. During the last decade, annual harvests have fallen within the range of approximately 0.8-1.0 million t and SSB appears to be at a relatively high level.

Survey indices are derived from the Joint Russian-Norwegian winter (January-March) survey, Norwegian Lofoten survey and the Joint Russian-Norwegian Ecosystem survey (August-September). These provide acoustic/trawl and bottom trawl indices. Survey time series (for the joint surveys) are extensive but have been adjusted from time-to-time and, in some cases, there are gaps in the time series resulting from technical issues. Even though the joint winter survey area was increased in 2014, indices from the increased area are not yet included in the assessment. Different software systems are used to manage data and calculate survey indices. We recommend that these processes be thoroughly documented and integrated within Sea2Data/StoX or a similar system ensuring traceability and quality control.

Some Russian and Norwegian commercial CPUE data are available but CPUE indices are not currently included. This should be revisited to determine if a useful time series can be generated, with appropriate adjustments for technology differences. Research on catchability changes over time is likely to be required.

Extensive catch data is available. Norway, Russia, Germany and Spain provide catch-at-age and weight-at-age data, while other countries only provide catch weights. The Norwegian data is based on reference fleet samples and no systematic port or on-board sampling is conducted. Historically, much of this data was provided in Excel spreadsheets but this is now compiled by ICES using Intercatch. Sea2Data is used to some extent as well but whether this complements or supersedes Intercatch is unclear. Potentially, the data can also be integrated using the ICES RDBES framework. Age reading consistency among nations/laboratories appears to be good. Weight-at-age estimation has evolved over time and some challenges remain. Again, integration under a single analytical system, such as ECA is recommended. The stock annex reports that ECA has been developed to utilize all sampling information to estimate catch at age in some areas. As mentioned in the assessment document, the time series for weight and maturity at age should be revised following the revision of the time series for the acoustic estimates in the Norwegian winter survey Since discarding is prohibited in the Norwegian fishery, discards are not reported or accounted for even though reports of discarding continue to appear. Furthermore, a constant round weight/gutted weight conversion factor is used even though there is evidence of seasonal variability (inter-annual variability is not mentioned).

NEA cod play an important ecosystem role both as predators and prey. A considerable amount of work has been done on consumption and cannibalism, and cannibalism is considered to some extent in determining natural mortality. The positive relationship between cod recruitment and temperature is also taken into account in predicting recruitment. Multispecies interactions beyond cannibalism are not considered.

The assessment is carried out using an age-structured model calibrated with survey data (SAM). This approach has been endorsed through ICES benchmarks.

The model does not bound survey error to that estimated from the surveys themselves. Although this should be remedied, this should be done cautiously as the survey uncertainty as estimated from samples can only be considered to be a minimum estimate because other factors such as time of survey, weather, catchability on the particular day, add uncertainty.. More research is necessary to better characterize uncertainty associated with several aspects of life history, catch and survey results.

Peer review of the methodology by ICES is sound but process peer-review is lacking. This can be addressed through better standardization of methodology as mentioned above as well as provision of additional skilled staff within IMR. This would enable better QA/QC as well as innovation in assessment modeling/statistics. Full implementation of REDUS and the best practice check list is also recommended to address this concern. Improved interaction between the survey/technology and assessment personnel is also encouraged.

Overall this assessment appears to be efficient and effective and IMR invests heavily in data collection and analysis and analytical support. Nevertheless, continued research to better characterize assessment inputs and related uncertainties is strongly recommended.

Exploration of climate and ecosystem considerations (both on stocks and of fisheries) is lacking and should be implemented to link the large body of work focusing on understanding the biological processes in this stock to stock advice.

NEA Saithe

Northeast Arctic saithe is harvested primarily by Norway which has taken 85-91% of the catch since 2008. Russia, Germany, France, the UK and Greenland also harvest this stock although only the Russian harvest is of consequence. The predominant gear type is bottom trawl although purse seine, gillnet and other gear types (longline, Danish and handline) are utilized. Recent annual catches have been close to 150,000 t.

The model is tuned with a survey index derived from the annual coastal survey which is an acoustic survey consisting of parallel transects with bottom trawls at fixed locations. Information regarding matching of the depth range covered by acoustics with the depth range covered by the bottom trawl (as well as other aspects of availability and selectivity) were not presented. These should be evaluated carefully if this has not been done. Survey indices computed using StoX are now available and are used in the assessment. However, even though StoX can provide survey index variance and this can be input to SAM, this is not currently being done. Plans are, apparently, in place to do this in the near future and this is to be encouraged, but it is important to recognize that an external estimate of sampling uncertainty is commonly an underestimate (hence “bounding by” the external estimate may be a better term). The lack of a recruitment index was noted and further work to develop this would likely improve the assessment.

Catch-at-age matrices are computed using ECA. This seems to work well for Norwegian data although the process for filling gear-type and/or areal data gaps related to insufficient sampling should be carefully reviewed.

Stock assessment is carried out by the ICES Arctic Fisheries Working Group (AFWG). The stock coordinator is Norwegian, and the assessment is conducted by Norwegian scientists.

Norway is also responsible for surveys that contribute to the assessment and for collection of most fishery-dependent data and integration of catch data from other (primarily Russian) sources.

Russian length data is converted to catch-at-age using Norwegian age-length keys; collecting otoliths from Russian catches for age reading by IMR should be considered to ensure consistency. We note Norwegian concerns regarding the practice of storing and processing Russian catch data in Excel files.

We also strongly suggest that comparability of samples taken by the reference fleet be validated relative to fully characterizing the commercial fishery. As with some other stocks considered during the review, the possibility for deriving useful CPUE indices from the reference fleet should also be examined.

Overall, the modeling approach appears to be appropriate and the process is efficient. Peer review and quality control of the assessment process is ensured by the ICES process although improvements and more comprehensive quality control in the provision of fishery dependent data is recommended.

Even though it is evident that ecosystem factors influence the status of the stock, this is not taken into account in the assessment. IMR has embarked on a Climate Vulnerability Assessment as the first step in recognizing and addressing climate change in characterizing and assessing the dynamics of major stocks.

Staff and financial resources to support data collection, analysis and the stock assessment are sufficient but internal communications among fishery dynamics staff and the demersal fish staff could be improved.

Wrasses

Five species of wrasse are harvested to supply a lucrative market for live fish which feed on lice when introduced into salmon rearing pens. Overall, approximately 20 million individuals are harvested each year and the vast majority of these are corkwing and goldsinny wrasse. The fishery takes place in shallow water (< 5m depth) and utilizes pots and fyke nets. Habitat preferences, gear vulnerability and sensitivity to fishing associated with life history characteristics vary by species.

Even though this fishery is relatively new, considerable research has been conducted to better understand life histories (age, growth, maturity, natural mortality) of these species as well as catchability and gear selectivity. This has resulted in minimum size limit regulations and seasonal (spawning) closures. This work includes mark-recapture studies and evaluation of potential survey technologies.

Fishery independent data are lacking although a recruitment time series is available which is derived from a beach seine time series. These data indicate high recruitment variability but no indication of decreasing recruitment as the fishery has intensified.

Fishery dependent data are collected from a small reference fleet of 16 vessels. Detailed data from two traps per set are documented and data loggers (depth, temperature, soak time) are attached to these traps. Wave exposure is determined from time and location data. Catch data have been collected from the reference fleet since 2011 but collection of this more detailed data was not initiated until 2019. Since buyers will only purchase live fish in good condition, mortality resulting from discard of dead or damaged fish is likely to be important.

Analysis of CPUE data has been carried out but the time series is considered to be unreliable for evaluating population trends for a number of reasons, including the relationships between catchability and environmental conditions. Improvements in data collection initiated in 2019 hold promise for developing a usable CPUE index but further work is necessary before this can be determined.

We encourage further development of the CPUE index and enlargement of the reference fleet. The mark-recapture work should be continued, and evaluation of depletion methods should be considered. Research on survey methodologies should also be continued. Further, assessment methods appropriate for sedentary fish should be investigated in recognition of the fact that large spatial scale age-based assessments are unlikely to be appropriate for these species.

Ling

Ling support important commercial fisheries in Norwegian coastal waters (Subareas 1 and 2) as well as more broadly in waters adjacent to Iceland, the Faroe Islands, along the continental shelf west and north of the British Isles and to the south. Work conducted by ICES (WGDEEP) provides no evidence of genetically distinct populations. Nevertheless, the fisheries off Iceland and the Faroes are managed separately as is the Norwegian Coastal Fishery (ICES areas 1 and 2) and the fishery that takes place in waters off the British Isles and to the west and south (ICES Areas 3.a, 4, 6, 7, 8, 9, 10, 12 and 14). Here we are asked to consider two assessments, the first, which is in ICES areas 1 and 2 is harvested almost exclusively by the Norwegian fleet has supported catches of approximately 10,000-12,000 t in recent years (an increase over earlier harvests) with an increasing CPUE biomass index. The second, in areas 3.a, 4, 6, 7, 8, 9, 10, 12 and 14, has harvested 10,000-20,000 t in recent years (a marked decrease over earlier catches) but also showing an increasing CPUE biomass index. Curiously, zero discards are assumed for the first fishery although moderate levels of discard are included in catch data for the second. While this may seem warranted by the Norwegian discard ban, generally, investigations in areas with discard bans show a continuation of discards unless all trips are monitored using observers or remote observation technology. Multiple gear types are employed including longline, gillnet, trawl and handline. In waters adjacent to Norway, longlines and gillnets predominate.

For Norwegian landings, comprehensive catch data are available, including reference fleet data on length, weight, sex and maturity. Otolith samples for aging and tissue samples for genetic research are also collected. Some CPUE (LPUE) data are provided for French fisheries.

Age composition data are available but do not appear to be used in the assessment (the language in the two assessment documents is somewhat ambiguous).

Fishery independent data are not available because of mismatches in temporal and spatial coverage (as well as gear type) of current surveys. However, the Spanish bottom trawl survey in ICES area 7c and 7k does provide biomass and abundance indices. These were not presented or discussed.

The presentation and summary documents were rather brief, but it appears that the same assessment approach is used in both cases:

- Two CPUE series based on data from the Norwegian reference fleet for ling, one using all data available and the other using only data when ling were targeted (>30% of the total catch).

- The assessments are considered to be data limited and are carried out by ICES/WGDEEP (ICES category 3; CPUE trend-based assessment). IMR is the stock coordinator and provides much (most) of the necessary data.

The following caveat is provided for both areas:

As always, it should be emphasized that commercial catch data are typically observational data when used to estimate trends in abundance; that is, there were no scientific controls on how or from where the data were collected from the actual fish population. Therefore, it is not known with certainty if a cpue series tracks the population and how accurate the measures of uncertainty associated with the series are.

Other approaches have, apparently, been attempted, including a surplus production model (ASPIC). Nevertheless, fishery-dependent data appears to be substantive and reliable so further investigation of more sophisticated modeling approaches should take place. The lack of good fishery-independent data is limiting to some extent so re-analysis of existing data to account for survey gear differences, expansion of existing surveys or initiation of a new survey time series would also likely to improve the assessment.

Changes in the productivity of both "stocks" is apparent from the analyses provided and this suggests the need for more research on life history. Ecosystem and climate change factors are not considered in these assessments. This reflects the lack of information as well as the low level of staff support for this assessment. Communication among researchers working on similar assessment problems at different IMR locations appears to be limited and this should be rectified to the extent possible. Furthermore, comments made earlier about the lack of strong stock assessment expertise at IMR may constrain development of improvements in these assessments.

Greenland Halibut

Greenland halibut is assessed and managed as a unit stock in ICES areas 1 and 2 even though questions regarding stock structure are unresolved. Catches have ranged between 20,000 and 28,000 t since 2000 and have averaged 17,000 t since 1960. Trawl and gillnet are the primary gear types and almost all the catch is taken by Russia and Norway under the JNFRC.

Norwegian catch data is provided by quarter, area and gear type. Discards are not reported and are assumed to be zero. Information on length and biological characteristics is obtained from the reference fleet and port sampling. Even though the goal is to have adequate biological samples in each of these cells, this is often not accomplished, and data are aggregated across cells. Only catch data is provided from the Russian fishery, aggregated by area, quarter and gear type.

Four fishery data streams are utilized in the current assessment, each with sex-specific selectivities. These include the Norwegian and Russian bottom trawl fisheries, the Norwegian "gillfleet" (gillnet and longline) and the Russian "gillfleet" (also gillnet and longline). Catches by other nations (minor part) are included in "Norwegian Trawl."

Four survey-derived indices are also input:

- EggaNor – based on the Norwegian Greenland halibut slope survey
- EcoJuv – a juvenile index based on data from the northern/eastern areas of the Joint Ecosystem survey
- EcoSouth – an index for the Barents Sea south of 76.5°N, based on data from the Joint Ecosystem survey in the Barents Sea
- Russian – Russian bottom-trawl survey in the Barents Sea

These surveys differ in temporal and spatial coverage and none cover the entire area of distribution. Neither age nor CPUE data are used in the assessment.

Improvements in QA/QC of fishery dependent and fishery independent data and handling efficiency and calculation of survey indices would likely be achieved through full implementation of Sea2Data and StoX.

The assessment is conducted by the ICES AFWG. The stock assessment model GADGET was adopted following the 2015 benchmark. While this model appears to be effective and appropriate it is inefficient in the sense that it is difficult to configure. Even though the model is designed to work with both catch-at-age and catch-at-length data, only catch-at-length data are currently available. Unresolved issues with age determination should be addressed as a high priority.

A new benchmark will likely occur in 2022. We recommend evaluation of alternative, age-based modeling approaches (such as stock-synthesis) as well as updating life history parameters (especially M) to the extent possible. Ecosystem and climate factors are not currently considered in this assessment and a term of reference to consider these aspects in the next benchmark is suggested.

Management objectives for this stock should be clearly defined to allow evaluation of appropriate management strategies. The current management approach allows marked in catches following recruitment spikes, but as these only occur around twice per decade, this strategy is likely to result in high interannual catch variability with associated issues of overcapacity in years of low catch advice. An MSY strategy with the objective of reducing catch fluctuations while maintaining reproductive capacity should be evaluated. Recruitment seems to have declined in the last 15 years, and the implications of such long-term changes in productivity should also be considered.

Beaked Redfish

The directed fishery for this species in ICES areas 1 and 2 is exclusively a trawl (demersal and pelagic) fishery although limited bycatch is taken by other gears. Most catches are taken by Norwegian and Russian vessels in the Norwegian EEZ. This stock was heavily fished in the past and has been subject to closures and restrictions (primarily bycatch only). It is now managed through a TAC. Overall catches have been low since the early 1990s, fluctuating between approximately 10,000 and 30,000 t. Redfish catches are predominantly of beaked redfish (*Sebastes mentella*) but golden redfish (*Sebastes norvegicus*) are also found in the catch and it is difficult to distinguish between these two species. Both species are slow growing and long-lived, thereby potentially making them sensitive to small increases in mortality. Mixed species fisheries are always difficult to manage as stock development and sensitivity to fishing may differ substantially among species in the catch.

Norwegian commercial fishery data are provided by the Directorate of Fisheries and sampling of catches is carried out to provide size, age structures and biological information (details are lacking in the presentation and the assessment document; the ICES Stock

Annex is only accessible to authorized users). Details regarding availability and quality of Russian data were not provided. It is evident, however, that preparation and processing of fishery dependent data is neither timely nor transparent. Evaluation and improvement of QA/QC is recommended and integration and processing of data through well documented tools such as ECA is essential. Norwegian catch data does not include estimates of discard.

Fishery independent information is derived from four surveys, three in the Barents Sea and one in the Norwegian Sea (one of the Barents Sea surveys, the Russian bottom trawl survey in the Svarlbard and Barents Sea areas was initiated in 1978 and discontinued after 2017). Spatial and temporal coverage of these surveys is linked to ecosystem survey priorities rather than known patterns of redfish distribution and this is thought to introduce bias in the redfish survey indices and, consequently, in the assessment. The efficiency and transparency of management of survey data and calculation of survey indices would be greatly improved through documentation of estimation methods in e.g. ICES data products, Sea2Data or StoX so this should be prioritized. Inclusion of indices from two slope surveys (Egga South and Egga North) should be carefully evaluated.

In addition to the data handling and management concerns for fishery dependent and fishery independent data discussed above, concerns regarding quality, consistency and timeliness of age data were described. Ongoing efforts to address these concerns should also be prioritized. Similarly, the assessment would be improved by updating the methods used to apportion catches between the two Sebastes species and adoption of a common assessment platform.

A statistical catch-at-age model is used for this assessment. SCAA was adopted in 2012; it was first implemented in ADMB and was later migrated to TMB. While the software will run on a laptop and can deal with some data limitations, it requires considerable, specialized data preparation; furthermore, software QA/QC and transparency are problematic. While details were lacking, it appears that the assessment is carried out by ICES AFWG so methodological peer review is likely sufficient. Nevertheless, the concerns expressed above regarding data reliability, quality and timeliness require attention.

Even though shifts in distribution have been observed, these are not readily accounted for in the surveys or the overall assessment; ecosystem and climate change factors are not considered. Assessment uncertainty is estimated by the model and serious concerns regarding bias due to inadequate survey coverage were raised. This may introduce assessment bias of up to 50%. Due to high uncertainty and bias, advice provided is precautionary only.

Poor communication between staff working on this assessment but working in different locations should be addressed as well as the need for additional expertise to support the assessment itself and provide some degree of redundancy.

Comments on individual stock assessments: shellfish

Stocks presented and discussed during this session included Red King Crab, Snow Crab and Barents Sea Shrimp

Red King Crab

Red king crab is an introduced species in the Barents Sea. It is found in Russian and Norwegian waters, but the Norwegian fishery is managed independently. The Norwegian management strategy is to 1) Maintain a long-term fishery in the Quota Regulated Area (QRA) and 2) Limit expansion and reduce abundance in the Open Access Area (OAA). The QRA is a relatively small area which encompasses several fiords in the Eastern portion of the Norwegian Economic Zone, adjacent to the border with Russian waters and the OAA covers a larger area which extends further offshore. Recent landings have fluctuated between c 1,000 and 2,000 t Within the QRA, the fishery is carried out by small (<= 21m) vessels employing pot gear. Landings correspond closely with quotas but some degree of IUU is expected. In the OAA, harvesting is not limited but catch and effort reporting is mandatory. Recent annual catches have fluctuated between 180 and 400 t.

Fishery independent data consist of two annual surveys in the QRA using trawl and pot gear. Both surveys provide indices of crab density and distribution. Annual surveys are also carried out to monitor the spread of this species to the west

Landings are reported to the Directorate of Fisheries from both areas and provide information on total catches per landing broken down by sex and weight category. Information provided in the documents and presentation is insufficient to address questions regarding QA/QC but we recommend that all fishery dependent and fishery independent data be managed and processed efficiently and transparently using the systems developed by IMR for this purpose (e.g. Sea2Data/StoX).

Assessment is carried out using a hierarchical Bayesian production model which is informed by catch data and a stock index derived by modeling survey data. The model is designed to account for sequential westward migration (fiord by fiord) by including local and global parameters. Due to limited knowledge of stock dynamics and life history, informative priors for parameters such as MSY and q are drawn from expert knowledge on similar stocks.

The assessment is carried out by IMR without involvement of ICES. Formal peer review is strongly recommended together with the above-mentioned improvements in data handling and processing to improve transparency and QA/QC. Collaboration with Russia is essential to allow a comprehensive assessment of this stock and its dynamics. Further work on thermal habitat is encouraged since this will allow better evaluation of the potential for range expansion. Assessment and process transparency and QA/QC would be more rigorous if carried out under a formal ICES framework. We also suggest evaluation of an ICES Data Limited Stock method (e.g. increase with % seen in surveys capped by +-20%).

Snow Crab

Snow crab has recently become established in the Barents Sea. It is thought that this is due to natural colonization and range extension although deliberate introduction could have occurred. The fishery commenced in 2012 and a total of 2.5 t was landed during that year. Catches were initially unregulated and increased rapidly, peaking at 16,000 t in 2016. From 2017 regulations were implemented which allowed Norwegian vessels to fish only on the Norwegian Continental Shelf and Russian vessels in the Russian portion of the Barents Sea. A TAC was established and total (Norwegian + Russian) annual catches of ~ 12,000 t were taken in 2017 and 2018. The fishery is prosecuted with baited traps, but design, size and operations vary among regions.

Fishery dependent data are provided from logbooks and include information on time and location of fishing event, depth, effort (number of traps) and catch weight. Details regarding catch sampling as well as data handling, management and disposition were not provided.

Survey abundance estimates are derived from the annual Norway/Russia joint ecosystem survey. Monitoring of snow crab occurrence in cod stomachs also provides a useful index of abundance which tracks well with the survey abundance trend. Little information on survey methodology, data handling or index estimation was provided.

The modeling approach is similar to that implemented for red king crab but without the sequential colonization feature. This employs a logistical model informed by survey indices, catch data and priors for essential life history parameters.

Also similar to the red king crab assessment, the assessment is carried out by IMR without involvement of ICES. Formal peer review is strongly recommended together with any necessary improvements in data handling and processing to improve transparency and QA/QC. Collaboration with Russia is essential to allow a comprehensive assessment of this stock and its dynamics. As with red king crab, assessment and process transparency and QA/QC would be more rigorous if carried out under a formal ICES framework. We also suggest evaluation of an ICES Data Limited Stock method (e.g. increase with % seen in surveys capped by +-20%).

Limited information on life history and ecosystem effects are also apparent and additional research would be appropriate. Directed research should also be implemented to better understand changes in distribution and improve survey methodology (for both snow crab and red king crab).

The quality and scope of the assessment work for both crab species is strong. Hence, the benefit of adopting the ICES process is not related to the immediate quality of advice but more to the benefits of exchanging experience on crab fisheries and fisheries on emerging species and to ensure ongoing transparency, peer review and quality control.

Barents Sea Shrimp

The fishery for Northern Shrimp in ICES areas 1 and 2 has taken place since 1970. Currently, Norwegian and Russian vessels exploit the stock throughout its entire range while vessels from other countries operate only in the Svalbard fishery zone and the “Loop Hole”. During the last 10 years, annual catches have fluctuated between 20,000 and 50,000 t. Large factory trawlers (6000-7000 HP; utilizing double and triple trawls) participate in this fishery. Norway, Russia and the EU each harvest about 33% of the total catch. There is no overall TAC. Instead, the fishery is regulated by a specific TAC in the Russian zone and by effort control elsewhere. Additional gear and effort restrictions are placed on non-Russian or -Norwegian participants.

Fishery dependent data are limited. Harvest information is available from all participants, but detailed logbook information is available only from the Norwegian fleet. A CPUE index is developed from these data which accounts for vessel, season, area and gear type. It is not clear how or if technology creep is taken into account and this should be elucidated.

Data are provided from three surveys. The Norwegian shrimp survey took place 1982-2004 and the Russian shrimp survey took place 1984-2005. These have been replaced by the joint Norwegian-Russian ecosystem survey which has been operational since 2004.

Details regarding handling and processing of fishery and survey data are not provided so questions regarding transparency and QA/QC cannot be evaluated. However, since the assessment is carried out by a joint ICES-NAFO working group (Joint NAFO/ICES Pandalus Assessment Working Group; NIPAG) these concerns are likely to have been addressed. The ICES advice sheet notes that input data are of good quality but are sometimes subject to challenging delays in availability.

The assessment involves Bayesian fitting of a surplus production model. It appears that this assessment methodology has not been benchmarked so this should be considered a priority. Alternative production-model software might be considered (e.g., JABBA).

Some climate and ecosystem linkages have been investigated but further work is necessary. Given our understanding of the thermal tolerance of Northern shrimp, research on temperature-related survey availability may be important.

Comments on individual stock assessments: marine mammals

Harp Seals

Two of the three distinct harp seal stocks are found in the Northeast Atlantic. These two stocks, the Greenland Sea (GS) stock and the Barents Sea / White Sea (BS/WS) stock, are managed by Norwegian authorities, and assessments are carried out regularly by IMR, often in collaboration with international partners. Both stocks have a long history of exploitation. Quotas were imposed for the GS stock in 1971 and for the BS/WS stock in 1955. The BS/WS stock has been subject to very low harvest levels in recent years while moderate harvests of pups and 1+ adults (<10,00 individuals per year) continue to be taken from the GS stock. These animals are considered to be generalist predators. Animals from the GS stock often undertake annual migrations from breeding sites into the Barents Sea and there appears to be a high degree of overlap in the autumn and winter.

The ecosystems inhabited by these stocks are undergoing rapid change. Changes in migratory patterns are likely and these require further investigation. Changes in body condition may be correlated and could be driven by changes in prey availability – this also merits further investigation.

Basic fishery dependent data in the form of annual hunt catch statistics are available from Russia and Norway since 1946 but information on reproductive parameters is scarce since it is collected only (approximately) every 5 years. The timing and low frequency of these collections creates data gaps and incomplete characterization of reproductive parameters. This would be problematic if the interannual variation exceeds the range of long term changes; however if population characteristics such as productivity and, abundance change slowly and variation is not great, the 5 year gap between sampling points may not be a major limitation. Fishery independent data consist of pup counts taken during surveys (~ every 5 years) of whelping areas in the GS and WS.

Assessment of total population size and population trajectories is carried out using a deterministic, age-structured model. The model has known shortcomings related to sparse data, lack of accounting for uncertainty and the fact that only three parameters are estimated. ICES WGHARP recognizes these limitations and is encouraging further model development. We encourage this ongoing work, including evaluation and development of the SAM framework. A more rigorous statistical approach could incorporate probabilities that are linked to management guidelines (e.g., PBRs and N70 catch level probabilities). The upcoming (2020) benchmark will be important.

Concerns regarding assessment methodology, paucity of survey/biological data and impacts of climate change on these stocks remain. Guidance of ICES WGHARP regarding assessment development will be important but availability of resources to improve fishery dependent data collection may also be necessary to improve these assessments (see above) however, If population characteristics such as productivity and, abundance change slowly, the 5 year gap between sampling points may not be a major limitation .

Minke Whales

Minke whale assessments are carried out following the rigorous Revised Management Plan (RMP) process defined by the IWC. This process requires accurate catch data and abundance estimates with uncertainty. The abundance estimates are derived from surveys conducted by IMR n areas “E Medium” and CM: Jan Mayen) using a mosaic design which covers the entire area every 6 years. Information on survey QA/QC was not provided but oversight by the IWC Scientific Committee should address this concern. Similarly, the process for establishing catch limits, the Catch Limit Algorithm (CLA) is clearly defined and subject to review by the SC. It should also be noted that harvest levels are very low relative to CLA advice.

The presentation provided details of the assessment and the deliberative process for setting up the catch limit algorithm and the management procedure by the IWC.

This assessment approach would benefit from more collaboration with colleagues working on population models (e.g., those working on fish population dynamics) since methodological advances are no longer occurring within the IWC itself. This could be accomplished by establishing a strong international collaboration outside the IWC. The existing information seems to be somewhat underutilized (e.g. length from catches could be included and related to e.g. quota or stock development). Assessment methodology could also be advanced to consider migration in the abundance estimate and to take into account information on habitat and stock structure and/or establish related research priorities.

Recommendations

We provide generic and overarching recommendations in this section. More specific recommendations regarding individual assessments can be found above.

Staffing

in general staff resources to support data collection, data processing and assessment are adequate. However, several constraints were identified during the review and these should be addressed:

- Lack of statistical and stock assessment expertise hampers the stock assessment process. IMR should establish an internal stock assessment development team and build capacity for carrying out comprehensive Management Strategy Evaluations
- For many assessments, key responsibilities are assigned to individual staff without backup. This lack of redundancy increases risk and should be alleviated
- In some cases, individuals working on similar assessments are located in different facilities which constrains communication and collaboration. Changes in organizational structure and/or relocation of key staff should be considered to address this concern
- Poor communication between scientists working on different aspects of assessments (fisheries dynamics, assessments per se, surveys/technology, ecosystem processes, IEAs) should be improved. Designation of stock coordinators should be considered
- Expertise in some other key areas is lacking. IMR is training additional age readers and should continue to review and address technical staffing needs.

Fishery Independent Data

- A simulation framework should be developed to evaluate survey design. For example, Kotwicki and Ono 2019 developed an approach for generating population data (conditioned on survey observations) to provide realistic simulated samples for testing alternative survey designs.
- Temporal and spatial mismatches may cause bias in survey results, especially when survey objectives involve multiple species and/or ecosystem monitoring requirements. This should be monitored closely. Where changes in survey design are warranted, IMR should pay careful attention to calibration requirements and work closely with ICES.
- Proper characterization of uncertainty in survey biomass indices is challenging. While sampling variance can be estimated by new software (including StoX) and this should become a universal practice, other sources of uncertainty are more difficult to assess. Ongoing work in this area is essential and will require expertise in statistics.
- Development of innovative survey technologies is essential to address increasing needs for fishery-dependent information. IMR leadership in this area should be supported by the necessary staff and financial resources.
- Ongoing research on catchability is essential to address fishery-independent and fishery-dependent information needs. IMR has significant capacity in this area and work should be aligned with outstanding questions related to stock assessment
- All processes and steps associated with production of survey indices should be fully documented and integrated within Sea2Data/StoX or a similar system to ensure traceability and quality control.
- For some species availability and catchability may be related to temperature; this should be investigated with the goal of improving or better-characterizing survey efficiency.

Fishery Dependent Data

- Catch Sampling and Estimation - Biases related to illegal/unreported catches, dissimilar sampling methods among different countries, and perverse consequences of the Norwegian discard ban are problematic, especially in some fisheries. Since Norway lacks an at-sea observer program and does not monitor fishing operations by CCTV, accurate characterization of discard is challenging and assessments for many stocks are unable to properly account for this source of mortality. Improved catch data should be considered a high priority and will require close collaboration with industry and with other countries involved in fisheries on shared stocks.
- CPUE indices are maintained for some fisheries but rarely used to tune assessment models. This may be because CPUE trends conflict with other tuning indices and/or because changes in fishing practice and technology creep are not properly accounted for. We recommend that IMR examine existing CPUE data to determine if corrections to existing data sets or improvements in data collection are merited. In particular, the potential for use of data collected by the reference fleet as CPUE indices should be evaluated.
- Mark-Recapture – use of historic tagging data and results from recent RFID tagging for some important fisheries was discussed (e.g. herring and mackerel). These data sets hold promise for improved estimation of migration patterns, life history parameters and fishery selectivity. This work should be continued following a thorough review of tagging, including documentation of analytical approaches and criteria for excluding subsets of the data.

- Age Determination
    - Age-reading methodology varies among countries involved in shared-stock fisheries. In some cases, otoliths and scales are used while in others, inconsistent results are obtained from reading the same structures. Consistency and inter-calibration is essential in this regard. IMR should champion emphasize this need even though it may be unable to effect necessary changes by all parties.
    - Age-reading methodology varies among countries involved in shared-stock fisheries. In some cases, otoliths and scales are used while in others, inconsistent results are obtained from reading the same structures. Consistency and inter-calibration are essential in this regard. IMR should champion emphasize this need even though it may be unable to effect necessary changes by all parties. Accurate determination of catch-at-age is essential for most assessment models utilized by IMR. Integration of this process under a single analytical system, such as ECA is recommended (note that the panel did not evaluate ECA in detail). Concerns regarding sampling bias should be documented and addressed to the extent possible. This includes documentation of steps taken to fill un- or under-sampled strata
- Data Timeliness - There are challenges with the analyses and incorporation of Russian data for several stocks. These seem to be linked to the lack of common databases and agreed extraction and analyses methods. Common databases may not be an option where countries consider input data to be sensitive. However, in other areas, such challenges have been solved by agreeing on common database formats which are then used in locally hosted databases. This ensures that joint extraction and analyses code can be agreed, quality assured and applied by each country (e.g. the efforts with RDBES in ICES). While achieving this requires a long-term effort, it would be beneficial to initiate the process. Meanwhile, benchmarks can explore different data assumptions, including greater-than estimated ageing bias, bias in reported catches (both landings and discards) by making exploratory assessment runs based on different assumptions (e.g. estimating unreported landings in the assessment model).

Climate, Ecosystem and Stock Structure

- Characterizing ecosystem processes and change could be improved and included within the assessment and advisory processes. This could be accomplished by assigning coordinators to compile ecosystem considerations that are taken into account in the advisory process. Through this process, diverse scientific expertise can be brought to bear (see this link for an example of how this is accomplished by NOAA for Alaska groundfish fisheries).
- Unresolved stock structure issues were reported for many fisheries. Ongoing work in this area is essential, especially relative to climate-change related distributional changes

Assessment and Projection Methodology

- In general, stock assessment and projection methods were deemed appropriate by the panel although investigation of alternative approaches for some stocks are detailed in the stock-specific sections.
- Lack of comprehensive peer-review for certain assessments (e.g. capelin, sandeel, crab stocks) was noted. See suggestions below for addressing this limitation.
- Use of Excel for assessment modeling and projections should be discontinued as soon as possible to improve quality control, transparency and documentation (this recommendation is more broadly applicable).

Stakeholder Communication

- Stakeholder communication is generally good but could be improved; engagement with NGOs is unclear. In particular, explicit involvement of fishers and other stakeholders through formal MSE would allow comprehensive evaluation of tradeoffs.

Peer Review

- Where assessments are carried out under the ICES framework, peer review is generally of a high standard. In some cases, however, it was evident that review of individual processes (e.g. data preparation, calculation of survey indices) carried out with IMR were lacking in QA/QC. Full implementation of standardized, peer-reviewed procedures is essential to address this concern.
- Assessments that are conducted independently by IMR should adhere to the ICES process or, preferably, be carried out within the ICES framework When management advice is required soon after surveys have been conducted (or fishery data have been compiled) using the ICES process for retrospective (post assessment/advice) peer review would be beneficial and this would allow improvements/corrections to take place in future years.
- In the short term, peer review of assessments not carried out by ICES should be conducted.

Provision of Management Advice

- While IMR's role is to provide advice to managers, the panel suggests that proactive engagement may be appropriate in some instances. For example, if precautionary reference points are lacking or an HCR has not been established, IMR could initiate work to address these deficiencies, perhaps by proposing precautionary limits to F and SSB where these do not exist (through an MSE process). Providing feedback to managers when catches exceed advice or when stocks are severely depleted should be considered as a priority.

General

- The extent of international collaboration in development of assessment support systems (software) developmental work (e.g. REDUS, Sea2Data/StoX) is unclear. Early inclusion of international research scientists and full integration with similar ICES systems will increase the likelihood of attaining results that are widely applicable and acceptable.
- We recommend continued investment in, and full implementation of tools and procedures currently under development within IMR to improve processes for data handling, data management, survey index calculation, estimation of uncertainty, documenting procedures, etc. (REDUS, Best Practice Guide, ECA, Sea2Data, StoX, etc.) It is essential that these systems be fully compatible with similar systems developed and utilized by ICES and that they be subjected to comprehensive peer review.
- IMR policy allows the institute to express disagreement with ICES advice to the press. This circumvents the quality assurance process of peer review integral to ICES advice and thereby increases the risk of basing advice on information that has not been sufficiently peer reviewed. It may also increase the risk of bias in response to political pressure (or the perception thereof). Even though this type of intervention does not occur frequently, we consider its continuation to be unwise.