Food and Agriculture Organization of the United Nations

General Fisheries Commission for the Mediterranean
Commission générale des pêches pour la Méditerranée

Working Group on the Black Sea

Subregional Group on Stock Assessment for the Black Sea (SGSABS) benchmark session for the assessment of Black Sea anchovy in GSA 29

Online, 5–9 July 2021

CONCLUSIONS AND RECOMMENDATIONS

The Subregional Group on Stock Assessment for the Black Sea (SGSABS) benchmark session for the assessment of Black Sea anchovy in geographical subarea (GSA) 29 was held online on 5–9 July 2021. The meeting was attended by experts from within and outside the Black Sea region, including an external reviewer (the list of participants and agenda are included as Appendixes 1 and 2 of this report, respectively).

In line with the SGSABS and benchmark terms of reference as well as on the basis of scientific evidence and of discussions held, the group agreed on the following conclusions and recommendations:

Ecosystem considerations

- The assessments span a period from 1988 to 2020, during which the Black Sea and the anchovy stock in particular have undergone very dramatic changes, which would need to be accounted for by any model capable of adequately determining stock status. The main issues are summarized as follows:
  - The extreme productivity characterizing the northwestern sector of the Black Sea in the past has changed and, due mainly to climate, the high productivity areas have expanded south hence extending the potential spawning area of anchovy but decreasing the strength of the features characterizing the Ocean Triad (Bakun, 1996). Additional issues such as eutrophication and anoxia have negatively affected the northwestern spawning grounds. This pattern is more evident in the last decade with a spawning anchovy population within the Turkish exclusive economic zone and in rather high numbers.
  - The migratory patterns observed in the past have changed and northwestern anchovy, instead of performing an anticlockwise migration along the Ukrainian, Romanian and Bulgarian coasts before reaching Turkey, often head straight for Anatolia and then move east in the overwintering period. In years of warm autumns and cold late winters (e.g. 2018), this pattern is further disrupted with schools lingering in the central part of the Black Sea and then moving straight to Georgia. In very warm years (e.g. 2012), anchovy were found to move south later and migrate offshore.
  - Anchovy landings underwent an important (fivefold) decline in 1989 and 1990. The only country whose catches recovered was Turkey and this was due to change in the migration routes which used to be more coastal before the collapse. Turkey targets anchovy mainly using purse seines operated from large vessels capable of chasing schools, while other countries employ static coastal traps relying on the incidental

passage of the species. *Mnemniopsis* was blamed for the decline and the recovery was thought to be due to the high numbers of its predator, *Beroe ovata*.

  - Azov Sea anchovy spawn in the Azov Sea and migrate along the eastern Black Sea coast; they appear to mix with Black Sea anchovy during the overwintering period. The two subspecies are exactly the same from the morphological point of view but Black Sea anchovy grow faster. In years of more intense mixing this results in peculiar length-frequency distributions (LFDs) that can be difficult to underpin by a model assuming a single species and a single growth curve.

Input data: summary and comments

A summary of the final input data agreed for the anchovy benchmark assessment is provided in Table 1 and the salient issues are summarized below.

Landings

- Missing data from Georgia (1988–1990) and Ukraine (1988–1991) in the past: this may be a problem because ex-Union of Soviet Socialist Republics (USSR) was formerly fishing significant quantities of Black Sea anchovy in Georgian waters.
- Sharp ups and downs in landings occurred and are conducible to:
  - *Mnemniospis* (1989)
  - Bonito outbreak (2005, 2014, 2018)

Landings-at-age

- Turkey: from 1988; Romania: from 2002; Ukraine: from 2014; Bulgaia + Georgia: from 2015
- Strategy used:
  - 1988–2001: all landings added to Turkish landings and used Turkish LFD and age-length key (ALK) assuming age structure was similar. Not necessarily a strong argument, but the bulk of catches until 2002 was dominated by Turkey
  - From 2002–2014:
    - Bulgaria + Romania + Ukraine
    - Georgia + Turkey
  - From 2014
    - Bulgaria + Romania
    - Ukraine
    - Georgia + Turkey
- Georgian catches have a consistently high abundance of age 0 for two reasons: i) Georgia does not apply a minimum landing size and ii) older anchovy and young of the year anchovy behave differently: older ones aggregate earlier than younger ones and it has to do with temperature (15–17°C vs. 12–13°C, respectively). In November–December, adults are aggregated on the coast (Turkey) and in January they move to Georgia and the young aggregate. This pattern is observed through time.
- Turkey:
  - The previously used, "old", catch-at-age matrix was derived from the analysis of LFDs which were used to generate a single ALK applied to all catches
  - Data were collected by Central Fisheries Research Institute (SUMAE) prior 1998 and again from 2005. The gap in the period 1998–2004 was filled with data from the literature (from three different sources according to the years). Coincidentally, the

previous assessment estimates an increase in spawning stock biomass (SSB) during this same period 1998–2004.

- More recently, an ageing workshop gave a revised common ageing protocol. All Turkish otoliths (approximately 4 000) from 2012–2020 (with some gaps) were re-read giving a new ALK that was applied to all catches back to 1988.
- Small fish with high growth rate (k) and low asymptotic length: length distributions overlap soon for older ages making length-based assignment of ages uncertain and sampling of older ages difficult. In this case it is difficult to distinguish age 3 from age 4 and a plus-group at 3+ should be considered.

Mean length of the catch and mean individual weight-at-age

- Since the mid-2000s, the mean length distribution of the catches has been decreasing.
- During the same period, the mean individual weight-at-age from the catches has also shown a decrease, more pronounced for the older ages.

Turkish catch per unit effort

- Previously a nominal catch per unit effort (CPUE) based on the catch of purse seiners divided by the number of vessels was used.
- A significant amount of effort was devoted to retrieving the daily catches of purse seiners and pelagic trawlers and using them to derive a CPUE standardized by location of catch and time. The standardization had the scope of accounting for changes in the distribution of the stock and the fishery and of possibly coping with issues of hyperstability in catches generated by the schooling behaviour of anchovy.

Acoustic survey

- It was decided to only consider the winter acoustic survey in the assessment, excluding the summer one and possibly 2011 when only the coastal area was sampled.

Maturity

- The previous maturity ogive was revisited also in light of the information available, by country, on the percentage mature at age 0 as well as available information on the sharpness of the spawning peak (e.g. on batch fecundity by month) and on the knowledge available on migrations. The group concluded it unsafe to assume any of the age 0 class was spawning and contributing to SSB, while age 1 were assumed to be fully mature.

Growth and natural mortality

- Growth data from the literature had the common problems of i) having, in many cases, extremely high values of t0 and ii) being of unknown/doubtful quality. Given the impressive amount of work done to re-read otoliths, one single set of von Bertalanffy growth parameters (VBGP) was estimated from this dataset and used in the estimation of natural mortality.
- It was commented that these VBGP, based on Turkish data from 2012, may not fully reflect the situation in the past and may not represent the older age classes adequately, prompting the proposal to try and retrieve otoliths and data from the past in different countries.
- The group decided it would be important to use a vector-at-age, rather than a scalar value. The final vector was chosen based on an average of the values obtained applying different relevant empirical estimators (Hoenig, Then_M_tmax, Then_M_growth, Then_mean and Gislason) and it was decided to only compare it to Gislason, which was used in the last assessment.

Table 1. Summary of the final input data agreed for the anchovy benchmark assessment

Type of datum	Data
Landings	Combined for 1988–2020 Detail by country: Bulgaria: 1970–2020 Georgia: 1991–2020 Romania: 1960–2020 Turkey: 1967–2020 Ukraine: 1992–2020 Russian Federation: NA
Catch-at-age	Combined catch-at-age matrix for 1988–2020 Detail by country: Bulgaria: 2015–2020 (ages 0–4) Georgia: 2015–2020 (ages 0–5) Romania: 2002–2020 (ages 0–4)* Turkey: 1988–2020 (ages 0–5) Ukraine: 2014–2020 (ages 0–5) Russian Federation: NA * Romania provided catch-at-age data from 2002 to 2020 and weight-at-age data from 2008 to 2020. To be able to be consistent with weight-at-age data, the years from 2008 to 2020 for the catch-at-age were used in the model.
Weight-at-age	Combined weight-at-age matrix for 1988–2020 Detail by country: Bulgaria: 2015–2020 (ages 0–4) Georgia: 2015–2020 (ages 0–5) Romania: 2008–2020 (ages 0–4) Turkey: 1988–2020 (ages 0–5) Ukraine: 2014–2020 (ages 0–5) Russian Federation: NA
Catch-at-age and weight-at-age from previous assessment	The old catch-at-age/weight-at-age matrix from previous assessment (prepared by LFDs based ALKs) from 1988 to 2017 was extended to 2020 by adding 2018, 2019 and 2020 catch-at-age data (prepared by same LFDs based ALKs)
Tuning index	Turkish standardized purse seine CPUE 2007–2020; Turkish standardized pelagic trawl CPUE 2007–2020; Turkish winter hydro-acoustic survey 2011–2020
Natural mortality	Age M average* Gislason 0 2.6952 3.098 1 1.3258 1.785 2 1.051 1.165 3 0.967 0.995 4 0.9346 0.932 5 0.921 0.907 (*Average of Hoening, M_tmax, M_growth, Then_mean and Gislason_M methods)
Maturity at length/age	Age Proportion mature 0 0 1 1 2 1 3 1 4 1 5 1
Growth model	Linf = 12.45 cm, k = 0.815, t0 = -0.2793
Length-weight relationship	a = 0.004565408 , b = 3.121758676

The full dataset is available to registered experts participating in the benchmark in a dedicated Sharepoint portal, under Input data and Scripts/Input_data, as Excel files, Lowestoft input data and Rdata files (https://gfcm.sharepoint.com/EG/SitePages/Meetings/SGSABS_TUR29_Yr.aspx?RefYear=2019&Session=TUR_29#).

Assessment models and preliminary results

Two different assessment methods were explored:

i. SAM
ii. XSA which has been used to provide advice in the past

SAM

- A base model was discussed and decided a priori:

  Base model:

  - Catch-at-age 5: as proposed plus group: 4+
  - Fbar: 0–3
  - M-at-age vector
  - New VBGP
  - Maturity: 0-1-1-1
  - Tuning indices: see if the modelling process can help us prune some of these info also making treasure of the diagnostics made at data level
    - winter surveys: all ages, excluding 2011
    - Purse seine (P)S CPUE: all ages
    - Purse trawler (PT) CPUE: all ages

- The results obtained from the base model and subsequent trials (Table 2) revealed the inability of the SAM model to follow the cohorts owing to an issue related to the input data. This overall outcome was due to an incorrect allocation of variance to the different compartments, resulting for example in very high process error being estimated to compensate for this. The input catch data are such that the model is not able to estimate a strong pattern in fishing mortality (F) to fit the catches but rather estimates high variance in other compartments.
- More work on the input data is required to be able to run these models in a satisfactory manner

Table 2. SAM trials carried out during the benchmark session

Tests for SAM	
With catch-at-age data from data preparation meeting, new M and Mat. Winter acoustic (2011 removed) data and purse seine and trawl CPUE. With pg:4, fbar(0:3), 1988:2020.	SAM did not converge.
Short time series: With catch-at-age data from data preparation meeting, new M and Mat. Winter acoustic (2011 removed) data and purse seine and trawl CPUE. With pg:4, fbar(0:3), 2005:2020.	SAM did not converge.
With catch-at-age data from data preparation meeting, new M and Mat. Winter acoustic (2011 removed) data & purse seine and trawl CPUE. With pg:3, fbar(0:3), 1988:2020.	SAM did not converge.

With catch-at-age data from data preparation meeting, new M and Mat. Winter acoustic (2011 included) data & purse seine and trawl CPUE. With pg:4, fbar(1:3), 1988:2020.	SAM converged, but there are un-estimated parameters, and infinities in upper confidence bound.
With catch-at-age data from data preparation meeting, new M and Mat. Winter acoustic (2011 included) data & purse seine (trawl CPUE removed). With pg:4, fbar(0:3), 1988:2020.	SAM converged, but there are un-estimated parameters, and infinities in upper confidence bound.
With catch-at-age data from data preparation meeting, new M and Mat. Winter acoustic (2011 included) data & purse seine (pelagic trawl CPUE was removed). With pg:4, fbar(0:3), 1988:2020.	SAM converged, but there are un-estimated parameters, and infinities in upper confidence bound. Model result did not advanced considerably.
With catch-at-age data from data preparation meeting, new M and Mat. (Winter acoustic was removed) purse seine and trawl CPUE is used. With pg:4, fbar(0:3), 1988:2020.	SAM converged, model result did not advance at all, which is in line with the low weight associated to the acoustic survey in the other runs.
With catch-at-age data from data preparation meeting, new M and Mat. Winter acoustic (2011 included) data & pelagic trawl CPUE (purse seine CPUE was removed). With pg:4, fbar(0:3), 1988:2020.	Model results changed significantly since the purse seine CPUE had high weight in the other runs. The purpose of this test was to evaluate the importance of purse seine CPUE. The results revealed that purse seine CPUE are highly relevant so a model without them would miss important information, which is the reason why this run could not be used.
Pg3 was tried with new time of the tunings.	SAM did not converge.
Pg3 + CorF=1	SAM did not converge.
Pg4 + CorF=1	SAM did not converge.
Pg3 + CorF=0	SAM did not converge.
Pg4 + CorF=0	SAM did not converge.
Catch-at-age data changed. 1988–2017 (from previous assessment) 2018–2020 (from new data) + pg4 .	SAM converged. Model struggle to estimate age 4 year fish amount since there were small values in the inputs (0.001 or 0.002).
Catch-at-age: 1988–2017 (from previous assessment) 2018–2020 (from new data)+ pg3.	SAM converged. Different settings were tried in the ctrl object. But the model results is not proper for giving advice: smoother and low F value, high SSB & Rec and high process error.
Although these were not considered specific test runs, in addition, the timing of the tuning series (alpha and beta values) were changed to reflect the birthdate of anchovy at 1 June.	SAM converged, but there are un-estimated parameters, and infinities in upper confidence bound.

XSA

- XSA was run using all the combinations of assumptions on data and data sets (Table 3) with no success when using the newly calculated catch-at-age data and the standardized purse seine CPUE. In particular, most runs gave unrealistically high estimates of SSB.
- XSA was also run using the previous year's validated model, assumptions and data updated with the most recent three years (2018, 2019 and 2020) with a view to providing interim advice pending the finalization of the benchmark. Discussions on the goodness, reliability and validity of the outcomes in view of the shortcomings related to the data ensued. The final model exhibited significantly increased residuals of age 0 in all data sets used (purse seine CPUE in particular) compared to the previously validated one. This implied problems related to the input data, also indicating that there were no solid reasons to believe that the assessment could be reliably indicative of trends.

Table 3. Summary of XSA runs

	Catch-at-age	Weight-at-age	Maturity-at-age ogive	Natural mortality	TUNING
RUN01	New	New	New	New	New (PS+PT)
RUN02	New	New	New	New	New (PS)
RUN03	New	New	New	New	New (PT)
RUN04	New	New	New	Scalar M (0.847)	New (PS)
RUN05	New	New	New	Very low scalar M	New (PS)
RUN06	1998–2004 averaged	New	New	New	New (PS)
RUN07	2005–2020 only	New	New	New	New (PS)
RUN08	New (PG 3)	New	New	New	New (PS)
RUN09	Old (2018–20 updated)	New	New	New	New (PS)
RUN10	Old (2018–20 updated)	New	New	New	New (PS+PT)
RUN11	Old (2018–20 updated)	Old (updated)	New	Old (Gislasson)	New (PS)
RUN12	Old (2018–20 updated)	Old (updated)	New	New	New
RUN13	Old (2018–20 updated)	Old (updated)	Old	Old	New (PS)
RUN14	Old (2018–20 updated)	Old (updated)	Old	Old	New (PS)
RUN15	Old (2018–20 updated)	Old (updated)	Old	Old	New (PS+PT)
RUN16	Old (2018–20 updated)	Old (updated)	Old	Old	Old (updated)

Table 4. Advice

GSA	Species	Methods	Time series of catches	$F_{current}$ *$E_{current}$	F_{unique} *E=0.4	F/F_{unique} *E/E=0.4	$B_{current}$	B_{MSY} *B_{pa} **B_{lim}	B/B_{MSY} *B/B_{pa} **B/B_{lim}	Stock status	Scientific advice	Comments
29	Black Sea anchovy (*Engraulis encrasicolus ponticus*)	SAM XSA	1988–2020	-	-	-	-	-	-	Uncertain	Do not increase fishing mortality	Benchmark not concluded. Contrasting signals in the data did not allow the status of the stock to be ascertained. An appraisal of all information is provided in Table 5, including a summary of the conclusions that could be drawn from each data set. A roadmap for the finalization of the benchmark was suggested.

Table 5. Appraisal of all information available on the anchovy stock and the conclusions that can be drawn from each to assist in determining stock status

Information available	Signal	Cause of uncertainty
Landings	Decreasing with sharp declines	Vessel buy-back program applied in Turkey. Environment-driven recruitment success. Climate-driven variability in the overwintering/fishing grounds influencing the availability.
CPUE	Increasing	Hyperstability issue due to overwintering aggregations.
Individual size	Decreasing	Mix of different anchovy forms displaying different growth rate/pattern.
Abundance index	Increasing	Abrupt fluctuations and short time series.

General conclusions and future work

- Data were provided by all countries, with the exception of the Russian Federation, up to 2020; addressing the request from the GFCM to be able to provide advice to the Working Group on the Black Sea (WGBS) and the annual session of the GFCM based on year n-1 data.
- Coordinated by the BlackSea4Fish project, data preparation efforts were carried out prior to the meeting. These significantly improved the input data (especially in terms of catch-at-age information and tuning indices – both commercial CPUE and scientific surveys) and advanced the discussion on basic criteria and assumptions for the assessment of anchovy.
- The first session of the benchmark assessment on Black Sea anchovy identified a number of issues related to the data that prevented the benchmark from being finalized. These issues were related mainly to:
    i. Age, growth and migration at Black Sea level: the input data used in the assessment relied mostly on Turkish catches and surveys. In particular, ALKs relied on the fact that the processing of otoliths and LFDs derived from Turkish catches covered a limited range of sizes not always representative of the entire basin. This could be ascribable, among other things, to the migratory pattern of anchovy at the basin level – which has also been found to be changing over time – such that the larger spawner mostly reside in the western sector with Turkish catches comprising mostly age 1 individuals and Georgian catches mostly age 0 individuals. In addition to this, the possibility of more or less variable amounts of Azov Sea anchovy – whose growth rate is slower – has been found to mix with Black Sea anchovy causing with the application of a single growth curve and ALK.
    ii. The use of commercial CPUE to tune the model and issues related to the possibility of hyperstability.
    iii. Acoustic survey.
- No quantitative or qualitative advice could be provided by the assessment models scrutinized during this session of the benchmark. In the absence of such conclusive results, other sources of information were sought as a basis for advice. Attention was turned to the new data in an attempt to find solid indications on the status of the stock (standardised CPUE and catch time series, and acoustic survey) but conflicting signals suggested the stock could only be diagnosed as uncertain pending finalization of the benchmark (Tables 4 and 5). An appraisal of all information is summarized in Table 5, including a summary of the conclusions that can be drawn from each data set.
- The benchmark was not concluded and a roadmap of future work was devised in order to address these issues to improve the assessment of stock status and conclude the benchmark, as outlined below.
- The report of the external reviewer, Dr Valerio Bartolino, is provided as Appendix 3.

Draft roadmap for future work on Black Sea anchovy

1. Work on data

1.1 Age, growth and migration at Black Sea level

The need to investigate the following three main issues was identified: a) the length frequency distributions (LFD), b) ageing and otoliths and c) migratory patterns. Together information on these aspects will shed light on variability in growth at a basin level and allow for the consideration of the peculiarities of the life cycle of the species as it varies across the region.

The first action proposed under this theme was a **workshop to set a process to review the issue at a basin scale.** This should take place as soon as possible in order to identify and collate all available information and start the practical work towards collecting new data, which by theme, according to future actions are as follows:

a) LFD:
   - Retrieve LFDs from the past
     - Georgia in 2009, 2010 and 2012 (possible)
     - Romania
     - Bulgaria
     - Others

b) Otoliths
   - Retrieve otoliths:
     - Georgia: from (2015) 2017
     - Romania: from 2008 to today
     - Bulgaria: back to the 1960s maybe
     - Others

c) Review migration patterns over time and information available (discuss juvenile and/or egg and larvae surveys).

d) Discrimination of Azov and Black Sea anchovy from their otoliths.

e) Phenology: evidence of changes in the timing of spawning? In relation to climate change?

f) Make a plan for future work and establish a process forward.

1.2 Standardization of Turkish CPUE indices

- Work towards better addressing the issue of hyperstability in CPUE by using vessel monitoring system (VMS) data to quantify search time and the issue of vessels coming together to search.
  - Investigate composition of the vessels contributing to the CPUE indices and possibly establish reference fleet(s).

1.3 Acoustic survey

- Investigate the effect of sampling location (offshore vs coastal) on the perception of the stock provided by the Turkish acoustic survey.
- Work towards a standardized acoustic survey covering the main area of the fishery (at least Turkey and Georgia) and beyond – facilitated by BlackSea4Fish project.
- Ensure there are no gaps in the future time series.

This work should be followed by the **launch of a regional reflection on ageing and modelling of these fast-growing short-lived species (second stage).**

The work should **encompass all coastal countries** and as such is foreseeable to be **coordinated by the BlackSea4Fish project**.

2. Finalization of the benchmark

- Organize an intermediate session of the benchmark to investigate alternative modelling approaches.
- Finalize the benchmark once all issues are resolved.

3. Other activities

Transfer of knowledge

Two main areas were identified as in need of further transfer of knowledge:

i. Ageing, including:
   - greater detail in the existing protocol with the aim distributing it for wider use; and
   - further hands-on transfer of knowledge on otolith reading.

ii. Assessment and modelling

Appendix 1

List of participants

Gizem AKKUŞ
Research Assistant
Institute of Marine Sciences, Middle East Technical University

Mohab Mahmoud BADRELDIN
Teaching Assistant in the College of Fisheries and Aquaculture Technology
Arab Academy for Science, Technology and Maritime Transport

Pinelopi BELEKOU
Policy Officer
European Commission –DGMARE

Thomas BRUNEL
Senior Scientist
Wageningen Marine Research

Georgi DASKALOV
Professor
IBER

Adham EL EBRASHI
Teaching Assistant
AASTMT

Ercan ERDEM
Fisheries Officer
Ministry of Agriculture and Forestry

Madalina GALATCHI
Researcher
NIMRD Grigore Antipa

Yaşar GENÇ
Fisheries Management Division
Central Fisheries Research Institute-Trabzon-Turkey

Ali Cemal GÜCÜ
Professor
Middle East Technical University

Erdinç GÜNEŞ
Senior Fisheries Officer
DG Fisheries and Aquaculture

Bohdan HULAK
Researcher
Odessa center YugNIRO

Salih İLHAN
Researcher
Central Fisheries Research Institute, Turkey

IRINE LOMASHVILI
Chief Specialist
Ministry of Environmental Protection and Agriculture of Georgia

Guranda MAKHARADZE
Ichtyologist/Invited Specialist
National Environmental Agency/Department of Fishery, Aquaculture and Water Biodiversity

Valodia MAXIMOV
Senior Fisheries Scientist
National Institute for Marine Research and Development "Grigore Antipa"

Mihaela MIREA
Project Officer
Black Sea Advisory Council

Serhii SNIHIROV
Leading Researcher
Odessa national I.I. Mechnikov university

Erdal USTUNDAG
Fisheries Data Officer
Republic of Turkey, Ministry of Agriculture and Forestry, General Directorate of Fisheries and Aquaculture

External reviewer

Valerio BARTOLINO
Researcher
Swedish University of Agricultural Sciences

GFCM Secretariat

Elisabetta MORELLO
Fishery Resources Officer

Huseyin ÖZBILGIN
BlackSea4Fish Project Coordinator

Konstantin PETROV
Coordinator of the Black Sea subregional technical unit

Yoana GEORGIEVA
BlackSea4Fish Project Consultant

Mihail VATSOV
BlackSea4Fish Project Consultant

Rusi ATANASOV
BlackSea4Fish project Junior Consultant

Matteo STARNONI
Liaison Consultant

Appendix 2

Agenda

1. Opening session
2. Review of previous advice and identified issues
3. Review of available fisheries dependent and fisheries independent information
4. Analysis of potential assessment models, including detected issues with previous assessment models, and identification of candidate models and assumptions
5. Practical session; assessment runs and compilation of tentative results
6. General discussion on assessment outcomes
7. Simulations and reference points
8. Wrap up and finalization of model runs and simulations
9. Conclusions and preparation of draft advice

Appendix 3

External reviewer's report

Valerio Bartolino

This section has been prepared by myself (Valerio Bartolino) as external reviewer invited to participate to the Black Sea anchovy benchmark. It should be noted that this review is limited to the material presented during the benchmark workshop (5-9 July 2021) and the discussion carried on during the workshop, at which point no section of the benchmark report was written.

The benchmark was focused on the Black Sea stock of anchovy (GSA 29) and excluded the Azov Sea stock which was considered only in relation to the former.

Migration and distribution

The Black Sea stock performs seasonal migrations from the NW spawning grounds to the E and SE. Such migration may progress along the Turkish coast or occur more offshore shortcutting the route directly from the NW to the SE of the basin. Interannual variability in the distribution may be high and it has been shown that in some years overwintering can occur much more north outside the Crimea peninsula (Chashchin et al. 2015). Primary productivity and temperature are suggested as important drivers of anchovy abundance and distribution, and it is suggested that the stock promptly responds to temperature changes which trigger at approx. 15-16 Celsius degree its migrations. The working group also reported that in recent years spawning activity extended to the south-western waters within the Turkish EEZ possibly as a response to temperature driven changes in productivity.

The general picture from the benchmark discussion and review is that of a highly dynamic system with pronounced interannual variability in the patterns of anchovy migrations and spatial distribution that affect the availability of the stock to the different national fisheries, especially the targeted fisheries by Turkey and Georgia which are carried on outside the spawning period and contribute to most of the catches.

A spatial model was briefly shown for predicting migration patterns based on simple temperature rules. The model is not published, was not specifically reviewed at the benchmark and in depth discussion was outside the scope of the workshop, but it was claimed that the model has good consistency with information from both the acoustic survey and fishery concerning the distribution of the stock.

A predictive model of the seasonal distribution of anchovy, on the line of that briefly presented at the workshop, is judge as a highly valuable tool to interpret the complex relationship between the catches and the availability of the stock to the different national fisheries.

Mixing

Black Sea and Azov Sea anchovy stocks are known to mix primarily in the eastern part of the Black Sea especially during winter time. Based on available knowledge on stock migration and distribution of the different fisheries, the two stocks are expected to mix in the catches of Turkey and especially Georgia. The mixture in the catches is expected to be seasonally and annually variable depending on the time and extent of migrations of the two stocks. Length frequency distributions are occasionally characterized by additional intermediate peaks which are interpreted as cohorts of the Azov Sea stock. Such interpretation seems reasonable given the known different growth patterns of the two stocks, but corroborating evidence based on stock identification methods would be highly valuable and recommended.

The mixing between the two stocks in both the catches and survey remains an important but unaccounted confounding effect which should be addressed with dedicated work on stock discrimination. Despite differences in the growth patterns, it is unlikely that the two stocks could be distinguished based on growth, for instance at the workshop it was not mentioned that the two stocks are characterized by largely different Linf, and other techniques should be considered including genetics. The spawning grounds for the two stocks are well known and individuals in spawning activity are available (at least for the Black Sea stock) which make a good opportunity to start work on the development of baselines for stock discrimination. Recommendation is that the initial work should at least consider a variety of discrimination techniques including morphometrics, growth, otolith shape, microchemistry and most importantly genetics to be later tailored towards a subset of cost effective and efficient methods.

Anchovy and the Black Sea ecosystem

Temporal variability in the BS anchovy has been related to multiple causes. The ctenophores *Mnemiopsis* could have contributed to the 1989 drop (although *Mnemiopsis* abundance was quite stable in that period along the Turkish waters), as well as it is suggested that the lows recorded in 2005, 2014 and 2018 could have been associated to the occurrence of large abundance of bonito in the basin. While it is impression that a comprehensive picture of the relative role of major drivers is still lacking, there are good evidences for strong ecosystem links that should be further investigated. In this respect, I would suggest to invest in parallel work to the classic single species stock assessment based on multispecies modelling, with the objective to better understand and quantify temporal variability in the natural mortality (M) of the BS anchovy. This would make sense not only in the context of an ecosystem approach to fisheries management, but also it could directly serve stock assessment.

Maturity

The previous assessment considered that 25% of age0 fish contributes to the spawning stock biomass. While it was not possible to reconstruct this exact figure, it was mentioned to be related to early spawned fish which had the time to mature and contribute to SSB in the same year.

This assumption and the data behind it were discussed at the benchmark and it was agreed to change to 0% the age0 contribution to the SSB. The revision is fully supported. The spawning peak of anchovy is relatively well defined in June-August with the peak in July (Lisovenko and Andrianov, 1996) and even if some age0 fish originating from early spawning event during the year are found mature in late Summer-Autumn their contribution to spawning in the same year is expected to be of minor relevance and it cannot be related in any way to the initial 25% estimate. The new assumption on maturity of age0 is not only better supported by data and knowledge of anchovy's phenology, but it has also the benefit to free the assessment from an unfortunate dependency between SSB and recruitment in the terminal year which added uncertainty to the evaluation of the status of the stock.

It is important to note that the Turkish data show a recent reduction in the length at 50% maturity. While this was not sufficient to work on a time variable maturity ogive, such development should be followed closely in the future especially if accompanied with changes in the time of spawning on which no specific analysis or conclusive evidences were presented.

Natural mortality

At the benchmark it was not possible to reconstruct the old estimates of natural mortality at age (M) based on the available sources. Estimates of M derived from a literature review are quite different from each other and difficult to reconstruct. For this reason, the working group explored new estimates of M using empirical relationships based on life history parameters and derived the final estimates as their ensemble. The approach is reasonable and also supported by the parallel pattern shared by most of the estimators, with some exception for the method by Gislason.

M is calculated as a fixed vector at age and given the available information this is appropriate. Methods based on von Bertalanffy growth parameters used parameters estimated from the new agreed aging methodology which also seems appropriate with the only limit that the otolith sample is based on otoliths from 2012. The need to extend the new aging method also to otoliths prior to 2012 (see comment on aging) could have implications also for M which should be consequently re-evaluated.

The assumption of a constant M is strong for a stock and ecosystem that went through important documented changes as anchovy in the Black Sea but at the moment there were no sufficient elements to provide alternative estimates of time variable M (see also comments to natural mortality).

Aging and age-length key (ALK)

A recent workshop provided the basis for an agreed methodology among countries and a re-evaluation of >4000 otoliths from 2012-2020. The methodology and outcomes from the workshop were not part of this review, but in principle are seen as an improvement. In addition, age-length data were modelled to reconstruct a single ALK by assuming normal length distributions by age. Several caveats are found in the approach and work is recommended in relation to each:

- The new ALK is based only on re-analysed Turkish otolith and should be expanded to include samples from other countries and fisheries
- the use of a single ALK for the whole assessment period is a strong assumption, year and gear specific ALKs should be calculated. Assessment methods to handle directly ALKs in the model would enhance this application
- a multinomial model should be considered for the modelling of ALKs. Such models could improve inference on age-length data, including testing (i.e. among fleets/countries) and standardization.

Landing data

The working group presented a time series of total landings of BS anchovy from 1970. The time period considered for the assessment is limited to 1988-present, because of the availability of compositional data, which is considered largely sufficient for the purposes of assessment given that anchovy has a rather short life span.

Landings are not available for Georgia and Ukraine during the first few years of the assessment period. These unaccounted catches are considered quite small for Ukraine but were possibly not so small and negligible for Georgia (1988-1990). Some additional work aimed to evaluate at least their relative size in those missing years is recommended and eventually it should be considered to start the assessment from 1992.

It is acknowledged that the working group has done fundamental work to retrieve national landings over several decades which is pre-requisite for stock assessment. It is my understanding that Russian catches of BS anchovy are not available. While assessment methods can cope with some level of uncertainty in the total catches, systematic underestimation could be more problematic depending on the extent of missing catches. It is the working group's perception that those Russian catches are rather small over the total, but from a stock assessment perspective it is recommended to continue working to obtain the most comprehensive and best possible picture of landings.

Discards are reported only by Romania in 2008-2014 but it seems reasonable assumption to consider them negligible for anchovy.

Turkey has a minimum landing size that to my understanding is achieved with a combination of fishing behaviour and mesh size. Several studies have shown that survival of fish which escape from fishing gears is rarely 100%, and in the case of towed gears and small pelagic fish such as herring it has been reported as potentially substantial (Suuronen et al. 1995, 1996). I am not aware of specific survival studies on BS anchovy, but because purse seiners have the largest share of catches in the Turkish fleet compared to pelagic trawlers potential unaccounted mortality of escaping fish should not be expected to be a major issue in this specific case.

Composition of commercial catches

Length frequency distributions by month are now weighted by the corresponding monthly catch before pulling which is considered a good improvement as part of the data preparation.

Age composition of the commercial catches is reconstructed for the entire assessment period (1988-present) based on length distributions and age-length keys. These compositional datasets are available only for the Turkish fleet from 1988, while start in 2002 for Romania (but according to the working group data quality is dubious until 2007), and are available only from 2014-2015 for the other countries. The gaps have been handled by assuming same age compositions among different countries with missing information in certain periods. Turkey which has a complete time series of ALKs had the largest contribution to the total catches, and the borrowing of information for filling the gaps for the other countries has been based on reasonable assumptions of similarities among the fleets (i.e., Bulgaria and Romania both use passive gears and common age distribution of the catches has been assumed for these two countries from 2014). However, this was not always possible and biases have likely been introduced when catches of countries with known distinct fishing pattern have been pooled under the same age composition forced by lacked of data (i.e., prior 2002 catches from all countries are assumed to share the same age composition as from Turkey). Inspection of the age compositions over the period 2015-2020 shows some systematic differences among the countries (i.e., Georgia has generally larger proportion of age0 compared to Turkey, or Bulgaria and Ukraine have larger proportion of older fish compare to the others) which confirm the worries for biases in the reconstruction of a full catch-at-age for the whole assessment period.

Catch data are also affected by a gap in the length frequency distributions in the period 1998-2004 which is filled in with various sources. It was not possible to verify the quality of the different sources and there are concerns also shared by the group on how to deal with this gap. Some sensitivity test was run which is meaningful approach under the current availability of data and models.

The use of integrated assessment models is recommended to reduce dependency on some of the major assumptions on the catch composition thanks to the fact that such models do not require a full catch-at-age matrix as input. This type of models allows to fit diverse datasets such as length frequency distributions and age-length data with their gaps directly into the assessment model without too many a priori assumptions. In this case, the approach could be particularly effective if different national fleets could be treated separately in the model.

Commercial CPUE

The basic rational in the use of commercial CPUE is that as a fish population decreases catches will decrease accordingly. But is it always true? What if catch rates remain high even as fish are rapidly depleted? Hyperstability is a combination of both fish and human behaviour. Fish may shoal maintaining same densities or even more densely as their number decrease, and fishermen also had centuries of knowledge about when and where to find the remaining schools.

It is generally accepted that purse seine CPUE metrics are likely to be relatively insensitive to changes in stock abundance over a relatively wide range of stock sizes, primarily due to the concentration of fishing effort on fish aggregations. However, with appropriate dedicated standardisations commercial CPUE from pelagic fisheries have been used when other sources of information on abundance are missing. In this sense the attempt to standardise the CPUE is supported and considered an improvement compared to the previous use of nominal CPUE but some reserves remain.

A modelling approach based on regression models (i.e., GLM) was applied to the Turkish commercial CPUE from the purse seiners and pelagic trawlers with the intent to standardise the catches. However, there are concerns that such standardisation could not be effective against the potential risk of hyperstability. In general, changes in CPUE through time may be due either to changes in the stock density (D) or to changes in the catchability coefficient (q). The need to standardize arises because q and D are not constant. This type of standardisation methods can deal with changes in catchability between vessels/skippers (so called "vessel effect") that anyway previous analyses of the BS anchovy data by the working group (not included in this review) found to be not significant on the CPUE, but I do not think it could discriminate if changes in local density are due to changes in stock abundance or changes in catchability due to other reasons such as hyperstability. Although it may be difficult to completely rule out the risk of hyperstability (which would require also some understanding of how stock size could affect school characteristics and possibly catch rates), closer inspection of vessels search time and behaviour is recommended.

Given the problem with the use of commercial CPUE, the working group should also consider the possibility to use a reference fleet for the calculation. Criteria for selection of such a fleet should be considered carefully and among those criteria temporal persistence of vessels in the fishery need to be included to overcome potential biases due to the recent decommissioned of part of the fleet.

Ability to follow cohorts is a fundamental pre-requisite of abundance data by age. Analysis of internal consistency in the commercial CPUE (and also in the acoustic survey) specifically looks at cohorts and it shows that such signal is rather poor in the present commercial CPUE. Several biological reasons such as variability in natural mortality can contribute to reduce internal consistency in abundance data, but complete lack of internal consistency between age1-2 and age2-3 in the standardised purse seine CPUE (labelled "NEW" in the documentation available) is problematic for the use in the assessment.

Acoustic survey

I support the decision to retain only the Winter survey which is more consistent, excluding the year 2011 which includes only the coastal transects and may be under-representative of anchovy abundances off-shore. On the contrary, in those few years when only the off-shore transects are performed, they seem to get sufficiently well into the coastal zone and be still representative of the whole stock. Some validation of the representativeness of the off-shore transects against the off-shore + coastal is recommended.

Fbar

Given the large proportion of age0 fish in the catch, the working group suggested to extend the Fbar from age1-3 to age0-3 which is supported. Only concern is on potential detrimental effects that this could have for the predictability of F for the advice and some test is recommended.

Modelling

During the assessment two models were tested, XSA which has been used in the previous assessment, and SAM.

Several exploratory runs were tried with SAM but major issues were encountered which prevented from achieving a satisfactory fitting of the data during the benchmark workshop. Among these issues:

- most of the tested configurations had problem of convergence which is revealing of difficulties of the model in “interpreting” the present dataset;
- a too weak signal on cohort development in the data resulted in an unrealistic allocation of variance in the assessment model. As a consequence, the process error on the fish numbers was estimated very large, often considerably larger than the amount of fish removed by the fishery, and with a pronounced trend. Fixing the process error is not seen as a valid solution with this type of model, but it was a useful attempt to better understand the model behaviour. By fixing the process error SAM had problems with the estimation of other parameters, suggesting that the model is not able to interpret the current data by allocating variance to other compartments except the process error which confirms the severity of the issue;
- the estimated F is unrealistically flat and with large confidence intervals which is likely the result of a weak signal in the catch data, as also suggested by the large observation variance of

the catches and their poor fitting; On the contrary, some overfitting was encountered with the commercial CPUE;

- none of the tests, including those on the correlation and variance of the F random walk, did fix sufficiently the major issues above.

Fitting of XSA with the revised data (i.e., new catch-at-age matrix, commercial CPUE and M) provided unrealistic estimates of stock size and development, in most cases with unsupported too high SSB and a drop in recruitment in the early 2000s. Several tests suggested that the new catch-at-age matrix was the main driver of such unrealistic patterns estimated by the model. The use of the old assessment extended with the last three additional years (2018-2020) was also attempted but found impractical given a deterioration of the age0 fitting in the commercial CPUE. Deterioration of the residuals is important but per se could be a weak rational to drop the previous assessment. However, it is noted that after the work done by the working group for this benchmark, the numerous caveats and limitations of the old dataset were now interpreted by the group as more severe issues than in the past leading to a lack of confidence in an updated assessment based on the old time series and assumptions.

Commercial CPUE increase for age0-2 but decrease for age3+, while the acoustic survey has only few years and with a record low in the last year. These contrasting signals made interpretation of trends based on indices difficult and the working group was reluctant to use them for the purposes of advice which seems a difficult but possibly best solution under these conditions.

A wide range of information and data are available on Black Sea anchovy which is only to a limited extent used for the assessment. Diversity in the type of data and differences in their temporal and spatial coverage may represent a challenge but there are modelling frameworks that could help integration of these different source. For instance, the Turkish egg and larval survey which could provide useful indications of emerging spawning activity along the southern coast appears currently underutilized. I understand that this is not only a matter of the assessment model used but a more flexible framework would contribute in including more of the available knowledge on the ecology of the stock.

Comments on the GFCM anchovy benchmark process and conclusions

Overall, preparation for this benchmark workshop stimulated a considerable amount of work on the data. This work has not only certainly improved the quality of the data but it has also shown where and what type of more work is required for an analytical assessment of the stock. A roadmap was prepared by the group as a conclusive act of the benchmark workshop. It includes also elements from this review and it is seen as an effective step forward.