Stichting Wageningen Research Centre for Fisheries Research (CVO)

HERAS survey indices: automation, TAF and testing

B.J.P. Bergès, S. Sakinan, F. Berg, S. M. Lusseau, M. Schaber, S. O'Connnell

CVO report: 21.007

Commissioned by:

A.P. Doddema MSc
Directie VLG
P.O Box 24001
2500 EK Den Haag

Project number: 4313100145
BAS code: KB-36-002-020

Publication date: 17th of February 2021

Stichting Wageningen Research
Centre for Fisheries Research (CVO)
P.O. Box 68
1970 AB IJmuiden
Phone. +31 (0)317-487418

Visitor address:
Haringkade 1
1976 CP IJmuiden

This research is conducted as part of the statutory task programme "fisheries research" and subsidised by the Dutch Ministry of Agriculture, Nature and Food Quality.

DOI: https://doi.org/10.18174/541760

CVO rapport ENG V11

Table of Contents

Summary

Fisheries acoustic surveys are routinely conducted around the world, and particularly within the ICES community. Following pre-defined transects, these surveys make use of downward active acoustic systems (i.e. so called scientific echosounders) coupled with biological sampling through fishing operations to estimate abundance and distribution of marine species. Of interest here is the HERAS survey that takes place yearly for ~1 month and is a dedicated international survey effort. It is performed across the North Sea, West of Scotland and the Malin Shelf, more specifically in the context of the Western Baltic Spring Spawning (WBSS) and North Sea Autumn Spawning (NSAS) herring stocks. Recently, the calculation of the WBSS and NSAS indices underwent a major update with the introduction of the ICES acoustic trawl database and the use of the StoX software. In the context of these recent changes, the aim of the current study is three-fold:

1. Develop R code for the automatic calculation of the NSAS and WBSS indices with the aim of minimizing manual user input.
2. Run sensitivity tests against various assumption on NSAS and WBSS indices.
3. Generate NSAS and WBSS indices under ICES TAF (Transparent Assessment Framework).

Overall, good agreement was found between previously derived indices and indices calculated using automatic routines. The only large discrepancies found was for 2017 in strata 11 and 141 which is accountable to discrepancies in WBSS/NSAS stock separation in these strata. This should be investigated further but the automatic routines as derived through this project will be used in the future for the derivation of the index alongside rigorous checking of outputs. Yearly, this process will also be running on the ICES TAF framework, providing improved transparency and robustness.

A secondary aim for this study was to run sensitivity tests programmatically. This was done for a range of assumptions around: calibration of acoustic instruments, stock splitting in strata 11 and 141, alternative strata definition and alternative haul allocation strategy. Firstly, it was found that calibration error was very influential on the index. Considering this, it is recommended to run thorough checks of calibration results prior to every survey (e.g. yearly comparing results historically). Secondly, the stock splitting in strata 11 and 141 mostly influences WBSS and the introduction of novel identification methods will be beneficial, rotating out the currently used method based on vertebrae count. Thirdly, the change in strata tested here did not exemplify large differences, suggesting that the current stratification is appropriate. Lastly, the use of automatic haul allocation to transect was tested and was found influential, highlighting the importance of expert input during for the haul allocation process.

1 Introduction

Fisheries acoustic surveys are routinely conducted around the world, and particularly within the ICES community. Following pre-defined transects, these surveys make use of downward active acoustic systems (i.e. so called scientific echosounders) to estimate abundance and distribution of marine species (Mehl et al. 2018; Dalen and Nakken 1983; Simmonds and MacLennan 2005). This type of survey is often used to derive abundance indices for a specific stock which are subsequently used in stock assessments. Of interest here is the HERAS survey that takes place yearly across the North Sea, West of Scotland and the Malin Shelf. This survey consists of an ~1 month international survey effort in the June/July period. In recent years, the following countries have been participating Scotland (GB-SCT), Germany (DE), Denmark (DK), Ireland (IE), Norway (NO) and The Netherlands (NL). A map of the coverage by country for the 2020 survey is shown in Figure 1-1.

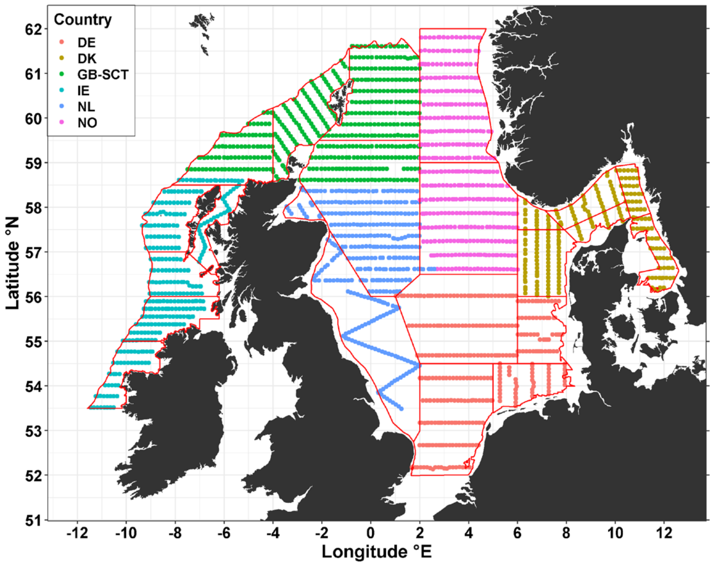


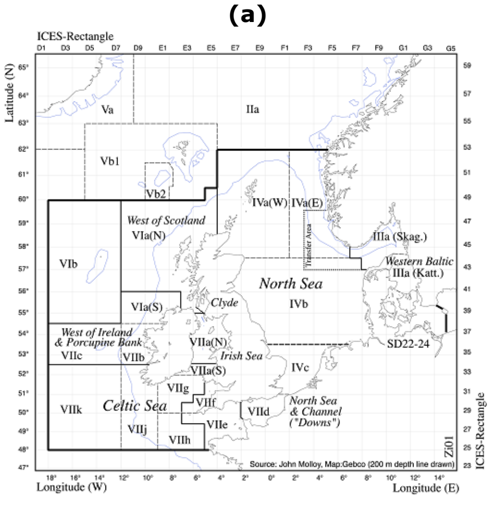
Figure 1-1: *2020 coverage of the HERAS survey with transects covered by country.*

The collation and analysis of the combined survey results is carried out collaboratively annually in November in a post cruise meeting with all survey participants. The HERAS survey delivers indices of abundance as well as key biological parameters (abundance at age, proportion mature at age and weight at age) for use in the assessments of the status of sprat and herring stocks across the North Sea and to the West of Scotland and Ireland (Figure 1 2). More specifically, these indices are delivered for the following stocks for which assessments are carried out in the ICES Herring assessment working group (ICES 2020d):

- North Sea Autumn Spawning herring (NSAS): Autumn spawning herring in subarea 4 and divisions 3.a and 7.d
- Western Baltic Spring Spawning herring (WBSS): Spring spawning herring in subdivisions 20–24 (where subdivision 20-21 is the same as division 3.a)
- West of Scotland herring: Herring in division 6.a (North)
- Malin shelf herring: Herring in Divisions 6.a and 7.b–c
- North Sea Sprat (Sprat in division 3.a and subarea 4)

The geographical extent of the distribution of the different herring and sprat stocks is shown in Figures 1-2(c) and (d) respectively. For NSAS and WBSS, it is essential to note that both stocks are found in division 3.a as well as in the north-eastern part of the North Sea at the time of the HERAS survey. To provide abundance indices and biological parameters specific to each of these two stocks, the biological sampling in these survey areas includes sampling to provide stock ID information for the herring caught during the survey. This permits a splitting of the acoustic herring abundances into the component stocks. This document focuses on the automation of these computations and on sensitivity tests of the resulting indices against a range of assumptions. Only WBSS and NSAS herring stocks are dealt with in this study.

(a)



(b)

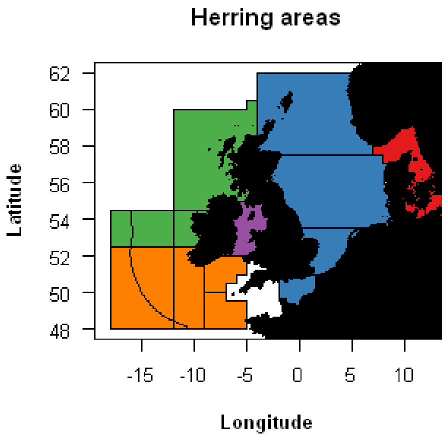


Figure 1-2: *geographical definition of herring and sprat stocks in the North Sea. (a) ICES divisions and subdivisions. (b) ICES areas for the different herring stocks. The HERAS survey provides abundance indices for West of Scotland and Malin shelf herring (green), North Sea Autumn spawning herring (blue) and Western Baltic Spring spawning herring (red). The North Sea Sprat stock is located in in division 3.a and subarea 4.*

Survey indices for NSAS and WBSS derived from the HERAS survey are directly used in their respective stock assessments, ages 1 to 9 for NSAS and ages 3 to 6 for WBSS. Both assessments use the SAM model, a probabilistic, state-space model (Nielsen and Berg 2014). The assessments are the basis for fisheries advice for these two stocks provided by ICES (ICES 2020b, 2020c).

For NSAS, the HERAS index spans the 1989-2020 period and follows closely the trajectory of the stock (Figure 1-3(a)), mostly within the assessment uncertainty, and is an important source of information for the stock assessment model. This is shown by the observation variance per data source as estimated by the model (Figure 1-3(b)). Within the SAM model, the lower the observation variance the more weight the data source has in the model. Here, the HERAS index for ages 2-6 is close to catch at age data in terms of observation variance. Moreover, the survey has a very good correlation between ages (internal consistency), suggesting a good tracking of cohorts. However this cohort tracking is not as close as it has been in recent years (ICES 2020a).

For WBSS, the HERAS index spans the period between 1991-2020. has a long term trend that follows the trajectory of the stock (Figure 1-3(c)) though with yearly variation exceeding assessment uncertainty. Similarly, to NSAS, the HERAS index is a key source of information for the core ages (ages 3-6) of the stock. This is illustrated by low observation variance (Figure 1-3(d)). For WBSS, the level of observation variance for the HERAS survey is close to levels from the GERAS (German Autumn Acoustic Survey, part of the Baltic International Acoustic Survey, BIAS) survey and catch data.

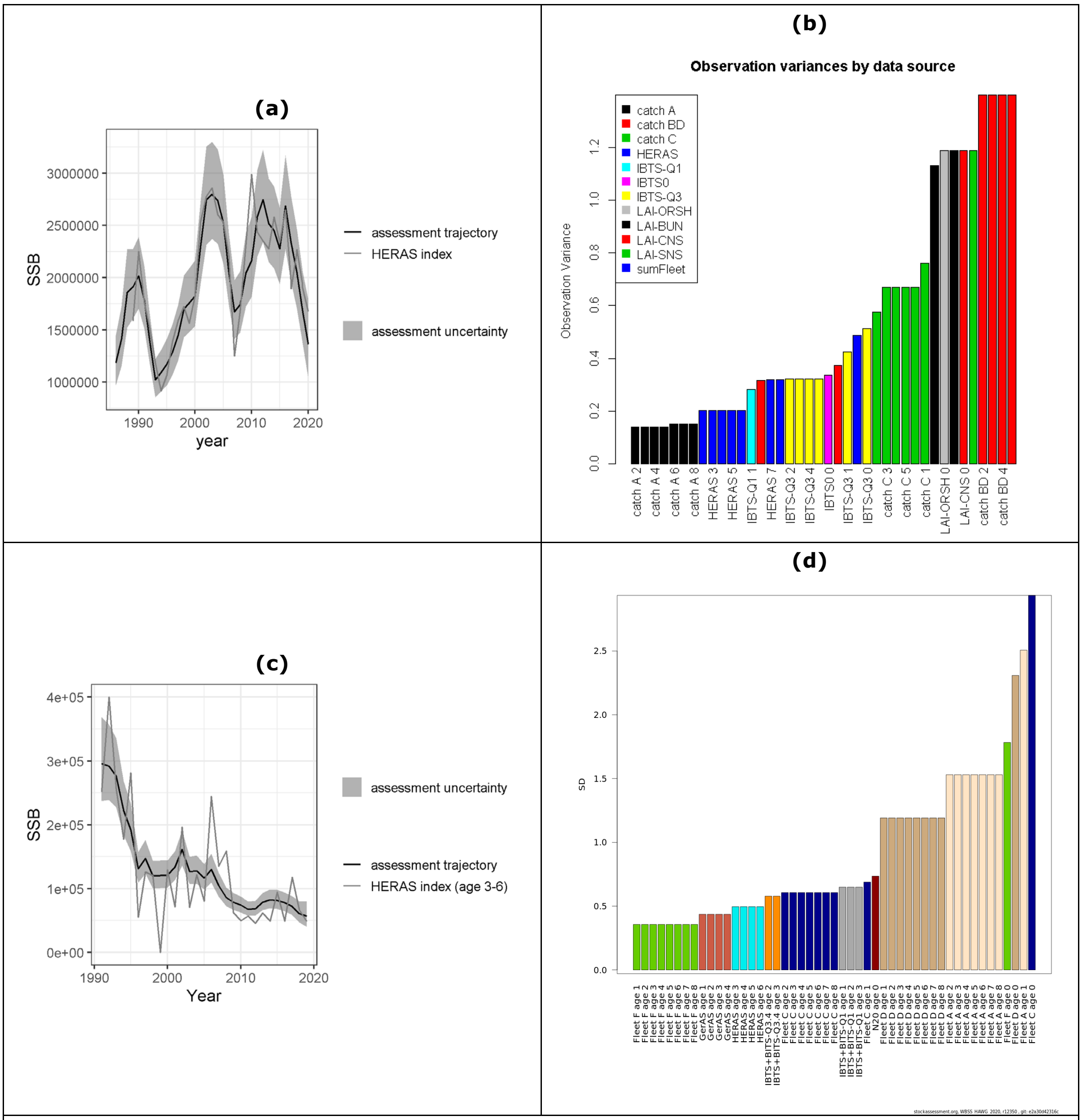


Figure 1-3: *NSAS and WBSS assessments and in the context of the use of the indices derived from the HERAS survey. (a) NSAS SSB trajectory for the assessment and HERAS survey. (b) NSAS observation error of input sources as estimated by the assessment model. (c) WBSS SSB trajectory for the assessment and HERAS survey. (d) WBSS observation error of input sources as estimated by the assessment model.*

Since the start of the time series for NSAS and WBSS in 1989 and 1991 respectively, the method of calculation has changed, especially in recent years (Figure 1-4(a)). Prior to 2015, the calculation of survey indices was performed on a national level per ICES rectangle and collated to provide final overall estimates. This involved processing through national programs and collation using a specific database (FishFrame). This processing involved several manual inputs which was an error-prone process. Since 2015, a transition was undertaken with 1) using the newly developed ICES acoustic trawl survey database[1] for data storage and 2) using a new calculation workflow for the derivation of abundance at age using the StoX software[2] (Johnsen et al. 2019). The use of the ICES acoustic database alongside the StoX software provide several advantages such as: data disaggregation (e.g. to better handle stock mixing), uncertainty estimation, more robust estimation, and traceable process. Moreover, the use of the acoustic database also drove harmonization of the HERAS survey (e.g. biological sampling, acoustic data analysis) across participating countries. A drawback induced using StoX is the loss in spatial heterogeneity of the calculated outputs. With the use of StoX, each stratum used encompasses several ICES rectangles whilst previously, calculations were undertaken on each ICES rectangle, providing a finer spatial scale. Previously, the survey area was stratified in post-processing based on biological composition in the trawls. Since 2016, the survey design is determined based on strata used in post-processing in StoX (e.g. when allocating survey coverage by country). Generally, this provides a more robust and sounder framework for the survey. A comparison of the 2015 indices between previously applied analysis method and calculations using StoX did not detect significant differences overall (Lusseau et al. 2016; ICES 2015c). In 2015, the survey was carried out according to the old design and indices calculations were performed at ICES rectangle levels and using StoX. For the StoX calculations on that year, the stratifications were defined during post-processing (Figure 1-4(b)). Starting from 2016 the newly defined strata definitions were used for the survey design (Figure 1-4(c)) in order to ensure consistency between survey and index calculations (Lusseau et al. 2016; ICES 2015c). In addition, since the first use of the StoX software, the NSAS/WBSS calculation procedure has been optimised between 2015 and 2017 with increased familiarity with the new database and analysis method. Since 2017, the calculations undertaken follow the same streamlined workflow.

[1] http://ices.dk/data/data-portals/Pages/acoustic.aspx
[2] https://www.hi.no/en/hi/forskning/projects/stox

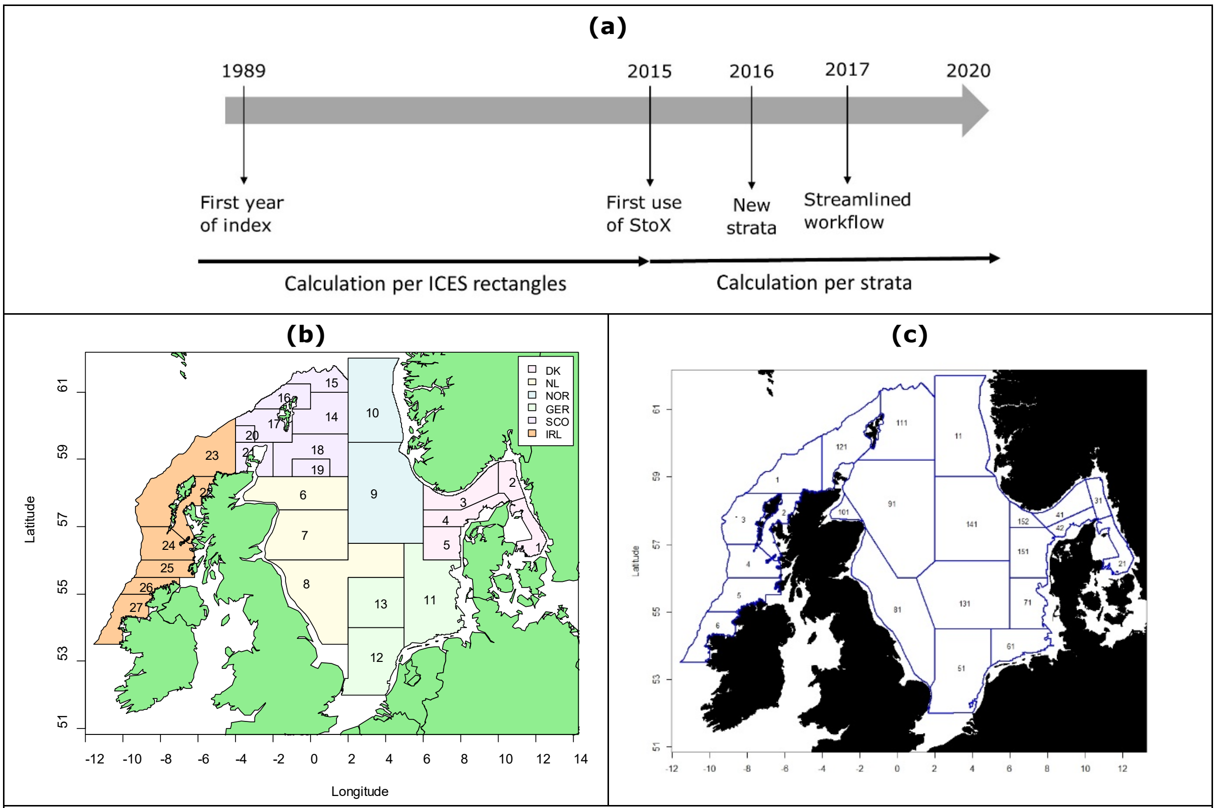


Figure 1-4: *(a) methodological changes in index calculation for the HERAS survey. (b) strata definition of the 2015 survey. (c) strata definition since 2016.*

In an effort to increase transparency, ICES is engaged in building and maintaining a range of databases[3] and has recently developed the Transparent Assessment Framework (TAF)[4] (Figure 1-5). TAF is an online resource used to publish and run R code for annual stock assessments but also for inputting data computations such as survey indices. This framework is open and works toward a transparent and fully traceable stock assessment process, from data all the way to fisheries advice. To date, a range of stock assessments and survey index calculations have been developed on TAF but currently the code used in the calculation of indices from acoustic surveys is not documented.

[3] https://www.ices.dk/data/data-portals/Pages/default.aspx

[4] https://www.ices.dk/data/assessment-tools/Pages/transparent-assessment-framework.aspx

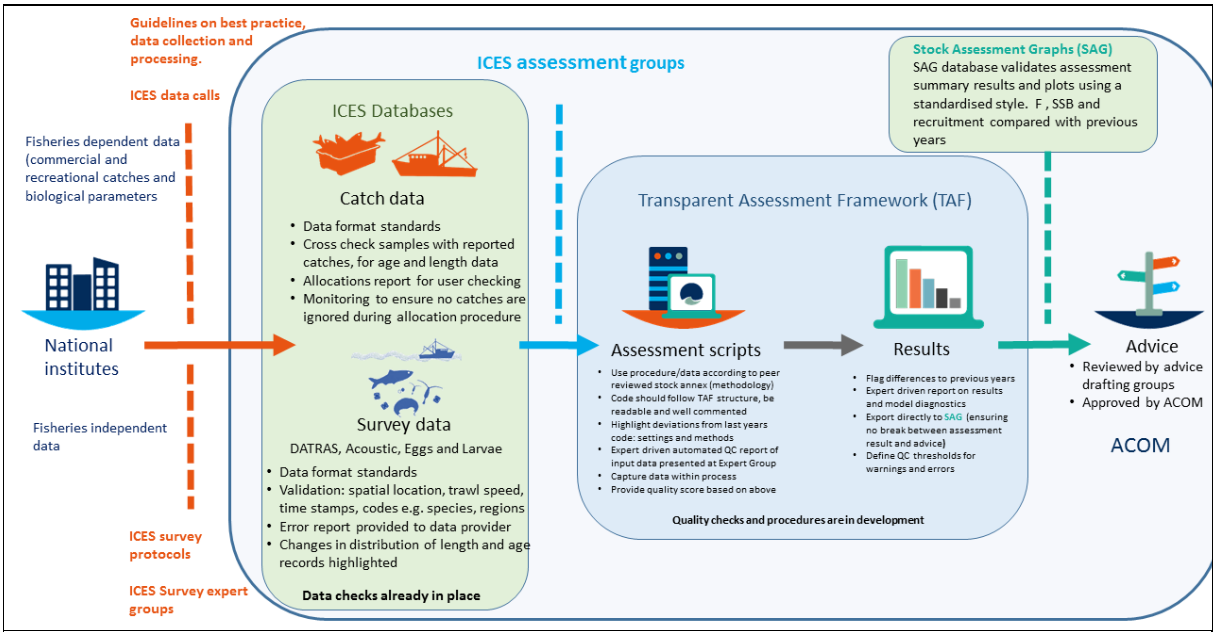


Figure 1-5: *fisheries advice process from national institutes to advice, including role of TAF in process transparency. Extracted from (Millar and Magnusson 2018).*

Using the TAF tool and the StoX software, the aim of this study is three-fold:

1. Develop R code for the automatic calculation of the NSAS and WBSS indices with the aim of minimizing manual user input.
2. Run sensitivity tests against various assumption on NSAS and WBSS indices.
3. Generate NSAS and WBSS indices under the TAF framework.

The code developed through this study has been stored and versioned on git repositories:

- git repository for ongoing development related to aims 1 and 2[5].
- git repositories within the TAF framework (aim 3) for the 2019 survey index calculation[6] and the 2020 survey index calculation[7]. These repositories are private, but access can be easily granted by contacting the first author of this report.

Because the currently used workflow with StoX has been in place since 2017 (Figure 1-4(a)), for consistency, only years from 2017 to 2020 are considered in this study.

In the future, the streamlined workflow should tentatively be applied to years prior to 2017 as far back in time as possible. However, an important caveat in the use of StoX for survey years prior to 2016 (prior to survey design based on StoX strata) is the fact that previous index calculation methods relied on a different survey design. The previous design may be largely compatible with the new analysis method in the area covered by the Scottish and Dutch components, but this might be challenging for the German and Danish components.

5 git@git.wur.nl:berge057/heras_index_kbwot.git
6 git@github.com:ices-taf/2020_her.27.3a47d_acousticIndex.git
7 git@github.com:ices-taf/2021_her.27.3a47d_acousticIndex.git

2 Method

The aim of this study is three-fold: 1) automate the calculation of the NSAS/WBSS indices, 2) run sensitivity tests, 3) run the calculation of the NSAS/WBSS on the TAF framework. Separate runs of index calculation are performed for 1) and 2) with a total of 10 runs with various assumptions and model settings. Aim 3) is an implementation of the base run baseline under TAF. A description of methods for each run are given in the subsequent sections (starting in Section 2.2) but for clarity, the organization and listing of runs is given beforehand:

1. Automation of NSAS/WBSS index calculation
    - i. Base run baseline
    - ii. Base run bootstrap
2. Sensitivity test runs
    - a. Calibration error
        - i. Calibration gain offset
        - ii. Backscattering coefficient (SA) error
    - b. Stock splitting
        - i. Split ratio offset
        - ii. Otolith split
        - iii. Genetics split baseline
        - iv. Genetics split bootstrap
    - c. Strata definition
        - i. New strata definition
    - d. Alternative haul allocation
        - i. Alternative haul allocation

2.1 Underlying principles

A typical acoustic survey from a vessel follows predefined transects, collecting acoustic data continuously and performing directed fishing operations. The latter is essential to 1) ground truth echo traces from the acoustic records and 2) provide information on length and age structure as well as other biological parameters such as maturity, weights at age and mixing of stocks. Outputs from surveys then consist of both acoustic and biotic data.

2.1.1 Acoustic data

The standard agreed acoustic equipment when performing an acoustic survey for fish stock assessment are split-beam active acoustic systems (so called echosounders) operated in narrowband mode. Examples of echosounders that are routinely used are the SIMRAD EK60 and EK80 systems used in the HERAS survey. The definition of quantities associated with acoustic data are summarized in (Maclennan, Fernandes, and Dalen 2002). Acoustic data collected by these systems represents the volumetric level of backscattering in the water column and is expressed as the volume backscattering coefficient. This quantity S_v is calculated as:

$$S_\mathrm{v} = 10\log(p) + 20log(r) + 2\alpha r - 2G - 10\log(U), \qquad (1)$$

with p the digital power amplitude, r the range from the transducer, α the acoustic absorption through the water at the transmit frequency (Francois and Garrison 1982; Ainslie and McColm 1998) and G the calibration gain. For simplicity, effects from beam angle, pulse duration and transducer characteristics have been encapsulated in U. In depth explanation of the calculation of S_v is given in (Demer et al. 2015) or (Lunde et al. 2013; Lunde and Korneliussen 2016). The quantity S_v is expressed in dB re 1 m^{-1} with $S_\mathrm{v} = 10\log(s_\mathrm{v})$ and s_v is expressed in m^{-1}. An important component of Equation (1) is the calibration gain G which accounts for transducer specificities relative to an idealized lossless omnidirectional transducer (G_0) and filter attenuation ($s_{\text{a}\,corr}$): $G = G_0 + s_{\text{a}\,corr}$. With an accurate estimation of G, the echosounder provides absolute level of S_{v}.

In order to determine G, a dedicated calibration trial is necessary. This is most commonly done using the calibration sphere method (Demer et al. 2015; Foote et al. 1987). During the HERAS survey, calibration trials are performed prior to the survey onboard each participating vessel. The built-in calibration tool of the EK80 or EK60 systems are used for that purpose and provide calibration gain G_0 and filter attenuation $s_{\text{a}\,corr}$.

Following calibration, data collected through the survey can be corrected and yield absolute levels of S_{v}. The echograms are then interpreted by allocating fish species to different echotraces. The final output is given in an integrated form across a range (i.e. depth):

$s_{\text{a}} = \int s_{\text{v}}\, dr.$	(2)

The quantity s_{a} is the area backscattering coefficient and is expressed in m^2/m^2. It is most often used relative to nautical miles with $s_{\text{A}} = 4\pi(1852)^2 s_{\text{a}}$ in m^2/nmi^2 (synonym: NASC – Nautical Area Scattering Coefficient). Values of s_{A} can be broken down in different depth bins but are further summarized per distance interval by taking the average over the distance interval for each depth bin. This constitutes the final output from a survey, i.e. s_{A} per depth bin for each distance interval covered. For the HERAS survey, each participating vessel produces a single acoustic output that is stored on the ICES acoustic trawl database.

2.1.2 *Biotic data*

An important component of fisheries acoustic surveys is fishing operations. These are performed when large echo traces are observed on the echosounder. Trawling is attempted on these echo traces as much as possible (similar depth and geographical location) in order to collect fish samples. In terms of catches, there is no absolute requirement for large sample sizes or standardized tows, but rather for a representative sample of the fish schools targeted (enough specimens per species for representative length-frequency and other biological measurements) (ICES 2015a).

The fish samples that are collected through fishing operations are essential for the analysis of the survey. Firstly, they allow ”ground truthing” of the echo traces during scrutinization and post-processing of the acoustic data. Depending on the schools that are targeted, catches do not always consist of “clean” samples of the target species, but often consist of a mix of different species, either mixing in the schools observed and targeted or caught in conjunction with the targeted species. The level of mixing between species varies between different surveys. During HERAS, a high level of mixing with other species (whiting, haddock, Norway pout) can be observed in several regions, as well as a high level of mixing of different clupeid species. In this context, fishing information is essential in the decision making for species allocation of echo traces. Secondly, linking acoustic data to specific species categories provides density to specific fish species categories (herring and sprat for HERAS) but these need to be further disaggregated by length, age, stock id and maturity stage. To that end, biotic data are used to allocate biological data to different transects with the purpose to characterize the biological composition of acoustic abundances detected along the transect. Biological samples typically consist of length, weight, maturity and age readings. For the HERAS survey, these data are stored in the ICES acoustic trawl database.

2.1.3 *Abundance calculation*

Abundance estimation relies on the principle that there is a linear relationship between the number of individuals in the observed portion of the water column and the echosounder output, s_{A} (Simmonds and MacLennan 2005; Maclennan, Fernandes, and Dalen 2002). Therefore, once the expected energy level from individual targets are established, s_{A} can be converted to an abundance estimate. Factors such as species and size composition, behaviour and spatial variability in the distribution of fish have an impact on this estimation. Each of these factors are handled at different levels of processing by the StoX software.

Allocation of acoustic backscatter to species:

The s_a recorded during the HERAS survey is generally representative of multiple fish species in the area (including clupeids such as herring, sprat and occasionally sardine in the southern areas as well as gadoids such as Norway pout, haddock or blue whiting). As described in the earlier section, these data go through an initial interpretation for identification of the echo traces and allocation of s_A to species based on expert judgement and trawl information. Factors considered mainly consist of the catch composition from the targeted trawl hauls, the echo characteristics of the detected fish aggregations and other environmental parameters such as depth, vertical position in the water column and geographical location among others. In certain cases, nearby representative trawl hauls are directly used to calculate the s_A that can be attributed to the target species through a separate process. The final s_A used in the abundance estimation is assumed to be 100% composed of the targeted species (e.g. herring).

Target Strength of fish:

In linear form, the expected energy level from individual fish is termed as backscattering cross-section, σ_{bs}, and in logarithmic (dB) form Target Strength (*TS*) such that $\sigma_{bs} = 10^{TS/10}$. Target strength is a stochastic value, affected by tilting behaviour and morphological variability between individuals. The size dependency is accounted for by known species specific TS/ Length relationship. This requires reliable length frequency data to be acquired by trawl samples and allocated to the s_A values in order to convert s_A to an abundance estimate. For HERAS the TS/Length relationship used is:

$$TS = 20log(L) - m, \qquad (3)$$

with L the fish length and m a variable that is species dependent. For the HERAS survey, $m = 71.2$ dB is used which is standard for North Atlantic clupeid fish and especially herring. Equation (3) does not account for depth but this effect can potentially be included (Ona 2003).

The sampling units:

Echosounder measurements during HERAS are performed continuously typically 1 ping per seconds while the vessel is underway with a speed of 10 knots. These samples are aggregated at different levels in StoX for different purposes. The sampling units are:

- *Pings:* Typically, every 2-5 meters depending on speed and ping rate
- *Elementary Distance Sampling Units (EDSU):* Every 0.1 or 1 nautical mile. EDSUs consists of the aggregation of pings within a standard distance.
- *Primary Sample Unit (PSU):* Transects.
- *Stratum (part of a strata system):* Strata are represented as spatial polygons in the StoX analysis software. Strata are predefined based on known homogeneity in variability within the area with respect to distribution of aggregations and size compositions.

STOX procedures for abundance estimation:

STOX uses the biotic and acoustic inputs that are submitted to the ICES acoustic trawl database and converted to an XML format specific to the STOX. STOX is composed of different modules performing a specific task at each step where different templates are available for different types of surveys. For HERAS, the template "Acoustic abundance by transect and r-model with uncertainty" is used.
The template is constructed in 24 steps in a predefined order. The steps can be visualized, parametrised and executed through the Graphical User Interface (GUI). For the result of each step a text file is generated and stored in the "output/baseline/data" folder of the STOX project. Here a short description of each step is given.

1. ***ReadProcessData*****:** Reading the predefined processing procedures in the project template such as strata definitions, transect definitions and biotic assignment.
2. ***ReadAcousticXML*****:** Reading acoustic inputs from the survey outputs (XML files)
3. ***FilterAcoustic*****:** The input data file may contain different species categories and postprocessing results from different acoustic frequencies. This operator filters the acoustic data for requirements of the analysis.
4. ***NASC*****:** Generation of the acoustic data matrix based on NASC values in the data.
5. ***ReadBioticXML*****:** Reading biotic inputs from the survey outputs (XML files)
6. ***FilterBiotic:*** Filtering biotic data for requirements of the analysis (e.g. species category, minimum number of individuals in the hauls etc.)
7. ***StationLengthDist:*** Station length distribution, definition how the length distribution is handled (e.g. as it is or normalized by tow distance)
8. ***RegroupLengthDist:*** Define the precision of the length classes (e.g. 0.5 cm or 1 cm)
9. ***DefineStrata:*** Read polygons for strata definitions
10. ***StratumArea:*** Calculate the surface area of each stratum
11. **DefineAcousticTransect**: This is a manual procedure performed on the GIS module of the software on the map
    - Transects (PSU) are created by giving individual ID numbers to each of them
    - Each PSU are grouped under the strata they belong to
    - EDSUs that are belong to specific PSU are defined by manually clicking on the points on the map
12. ***MeanNASC – SampleUnitNASC:*** The mean NASC calculated by this function can be a single mean for the entire stratum or per PSU (transect). The inputs are individual EDSUs from the input acoustic file. For HERAS the means are calculated per PSU.
13. **BioticStationAssignment:** Assignment of trawl hauls to stations. This is a manual procedure performed on the GIS module of the software on the map
    - For HERAS the standard way to do this is using manual assignment
    - Expert judgment is used to allocate hauls to individual PSU
    - At this stage one of the main factors of judgment are the length and maturity distributions and reliability of the hauls. This is based on expert judgements taking into account e.g. the number of sampled individuals in the haul, background information for the haul etc.
    - Alternatively a radius distance or the strata can be used to allocate hauls as well.
    - It is necessary to allocate at least one haul per PSU
14. **BioticStationWeighting**; Weighting the biotic assignments performed in the previous step.
    - Standard for HERAS is equal weighting. E.g. length frequencies are aggregated based on proportions at each length class regardless of sample size, haul duration or catch size.
    - Alternatives are
        i. NASC (based on the acoustic density surrounding a station),
        ii. Normalization by catch weight (larger catch more weighting)
        iii. Normalization by catch number (larger number more weighting)
15. ***TotalLengthDist*****:** Here aggregations of length distributions are generated per each PSU based on previously entered parameters. Because of the equal weighting used in HERAS this step executes no changes generated in the visual plots in the STOX GUI.
16. ***AcousticDensity:*** Here the acoustic densities are converted to numbers by using the length depended target strength equation (Equation (3)). Length distribution comes from the previous step. The numbers are calculated for each PSU.
17. ***MeanDensity_*****(Stratum)**: An overall stratum mean is calculated from the mean density (average number of fish at length per square nautical mile) of each PSU.
18. **SumDensity_(Stratum).** Vertical summation. Currently this has no effect for the HERAS data as full water column is used disregarding the vertical bins in the acoustics.

19. **Abundance by length:** The number obtained in step 17 is multiplied with the polygon area to obtain abundance estimate for each stratum
20. **IndividualDataStations**; A list of biotic assignments are created containing information on which stations are assigned to which strata.
21. **IndividualData**; Here a list is created consisting of each individual measurement existing in the input file either only length measurement or full biotic variables such as age, maturity stage, individual weight etc.
22. **SuperIndAbundance** Here, the total abundance by length by strata are allocated to the individuals. For example; in year 2017, in stratum 91, we have 90 individually measured fish at length of 20 cm coming from 10 different stations. As a result of step 19, we calculated that we have 117,668,451 total individuals of 20 cm at Stratum 91. This number is allocated to each of these 90 individuals. Each of them receives 1,307,427. Regardless of whether these individuals have an associated age, maturity or individual weight, or not, each of them will be representative of 1,307,427 fish. The missing biotic variables such as age are assigned in the next stage.

Last two steps are found in the *"**Baseline report"***segment

23. **FillMissingData**, here the missing values are filled using the age variable. For example, if a super individual comes from only length measurement and without age reading, an assignment is done from the same length group and station which has a value for age by a random selection. If that station didn't have the value, then the procedure is repeated on stratum or survey resolution.
    - This is the final table generated from the data processing. From this table the reports for abundance indices are created and uncertainties are calculated.
24. ***EstimateByPopulationCategory:*** Generation of a summary table for final abundance indices

Further details of some of the components of the workflow are given in the next section.

2.2 HERAS index calculation: base run

For the HERAS survey, the level of mixing between the main species (herring and sprat) and other species (e.g. whiting, sardines, haddock) varies throughout. In the Dutch, Irish, Norwegian and Scottish survey (Figure 1-1), scrutiny of the acoustic data is taken to species level. Based on scattering characteristics of echo traces as well as catch composition of corresponding targeted trawl catches, a robust allocation of herring and sprat to echoes originating from detected fish schools and aggregations is feasible. The acoustic categories, herring and sprat, are then allocated to these echo traces and corresponding NASC values are exported from integration results. However, in the German and Danish survey area, this is not possible due to the very high level of mixing and the impossibility to discern mixed clupeid schools from echo traces. For these survey components, there is no direct herring/ sprat categorization. Instead, the species composition of each trawl is used to disaggregate NASC values into species categories by transect. This is achieved using the StoX software with dedicated projects to split NASC values for Danish and German surveys.

Furthermore, in the Norwegian and Danish survey components, strong mixing of herring stocks occurs between NSAS, WBSS and Norwegian Spring Spawning (NSS) herring. Historically the mixing with the NSS stock is not considered and only separation between WBSS and NSAS is performed. This is accounted for in two different ways:

- Danish survey (in strata 21, 31, 41, 42, 151, 152 in Figure 1-4c): use of otolith microstructure and shape (Clausen et al. 2007; ICES 2018b) to determine stock ID for the sampled fish. Stock information is embedded in the biotic data for each sampled fish and submitted to the ICES acoustic trawl database.
- Norwegian survey (strata 11, 141 in Figure 1-4c): use of vertebrae counts to determine fraction of WBSS at age for each year of survey. The proportion of each stock is determined using ordered statistics on vertebrae counts from the biological data. This split proportion cannot be used to

allocated stock id to individual fish biological samples. Biological information as submitted to the ICES acoustic trawl database therefore does not contain direct information on stock id. Instead, stock splitting is applied after the abundance at age estimation process.

Because of the multi-country survey components, species mixing and herring stock mixing, the procedure for the derivation of NSAS and WBSS indices involve the use of several inputs and StoX projects. Figure 2-1 gives a workflow summary of this process.

The processing of the biotic and acoustic data using StoX is first divided into two separate components: the main project with data from all strata except 11 and 141 and a side project that include data from strata 11 and 141 only (mostly consisting of data from the Norwegian survey but not exclusively). For the main project, a further two specific StoX projects are needed for the splitting of acoustic densities into fish species categories (herring and sprat). This is done through separate projects for the Danish and German survey components. The resulting split acoustic data alongside biotic data is further fed into the main project with data from other surveys (The Netherlands and Scotland). This specific process is depicted in Figure 2-2(a). The main and side StoX projects are then run, yielding abundance of herring for each stratum disaggregated by:

- Age
- Length per 0.5 cm bins
- Maturity stage (mature/immature)
- Stock id (NSAS, WBSS)

As a consequence of the survey biological sampling strategy, combinations of length, age, stock and maturity are missing sporadically. These are propagated through by StoX and the resulting final estimates will also include cases with missing allocation. In order to account for these missing fields in the final abundance estimates, these are filled using the following protocol:

1. Maturity of age 0 allocated to immature.
2. Maturity of length < 8.5 cm allocated to immature
3. Missing lengths: filled in using inverse of weight/length growth relationship inferred from data.
4. Missing weights: filled in using the weight/length growth relationship inferred from data.
5. Missing ages: iterative process comparing abundance proportions in associated length group.
6. Missing maturities: iterative process comparing abundance proportions in associated age group and/or length group.
7. Missing stock id: iterative process comparing abundance proportions in associated age group and/or length group.

In the filling in process, steps 6 and 7 are the most complex. Historically, this has been done manually in spreadsheets but is now automated, and more traceable and reproduceable. It is important to note that the filling in of stock id is only performed for the main project but not the side project. For the side project, the abundance in each category is broken down into NSAS/WBSS using split proportions calculated directly from the raw biotic data. NSAS/WBSS estimates from the side and main projects are then combined to produce final abundance at age. The procedure is visualised in Figure 2-1(b). With the newly developed automatic routines, the outputs from the processing are stored and managed in multiple dimensions (maturity, age, area, year) using the FLR R package (Kell et al. 2007)[8].

[8] https://flr-project.org/

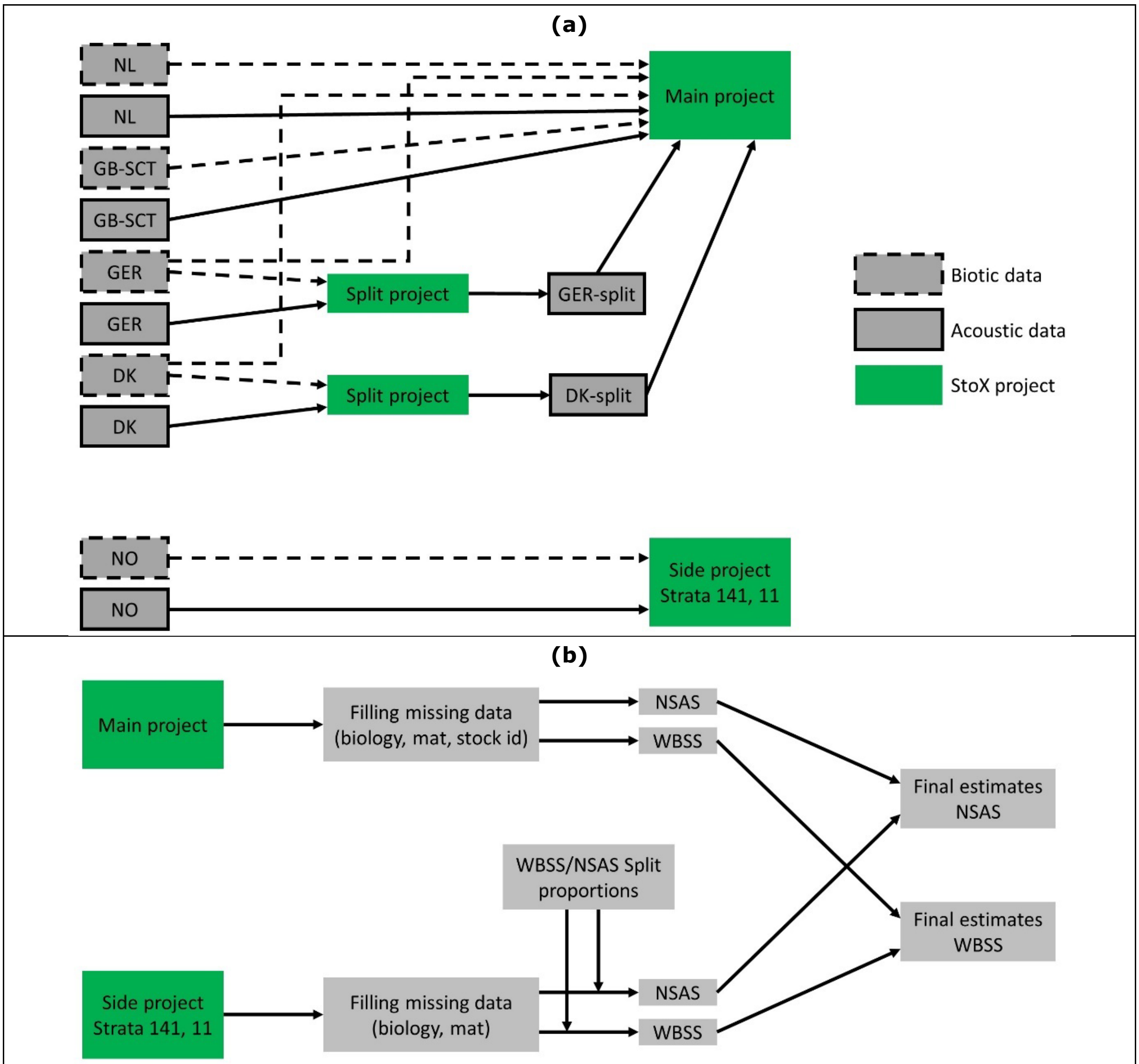


Figure 2-1: *workflow of NSAS and WBSS index calculations. (a) flow diagrams of input data and StoX projects. (b) Use of StoX projects to derive abundance at age.*

In strata 11 and 141, the abundance derived from the side project is divided into NSAS/WBSS using split proportions. This quantity is computed every year using vertebrae count measured through biological sampling. The fraction of WBSS in those strata is given by:

$$F_{\mathrm{WBSS}} = \left(56.5 - \bar{v}_{\mathrm{c}}(a, y, s)\right)/(56.5 - 55.8). \quad (4)$$

With $\bar{v}_{\mathrm{c}}(a, y, s)$ the mean vertebrae count in strata s, for year y at age a. It is important to note that this is only calculated for ages 1 to 4. For ages 4+, values for age 4 are used. In the rarer case of the presence of age 0, the value for age 1 is used. Equation (4) transforms the vertebrae count $\bar{v}_{\mathrm{c}}$ into a [0,1] interval. This is exemplified in Figure 2-2(a). Categories with $\bar{v}_{\mathrm{c}}$ greater than 56.5 corresponds to $F_{\mathrm{WBSS}} = 0$ (only NSAS), categories with $\bar{v}_{\mathrm{c}}$ smaller than 55.8 are allocated $F_{\mathrm{WBSS}} = 1$ (only WBSS) and any category with $\bar{v}_{\mathrm{c}}$ comprised between 56.5 and 55.8 is allocated F_{WBSS} with a linear decrease with increasing vertebrae count. The resulting WBSS split proportion for each age, strata and year is shown in Figure 2-2(b).

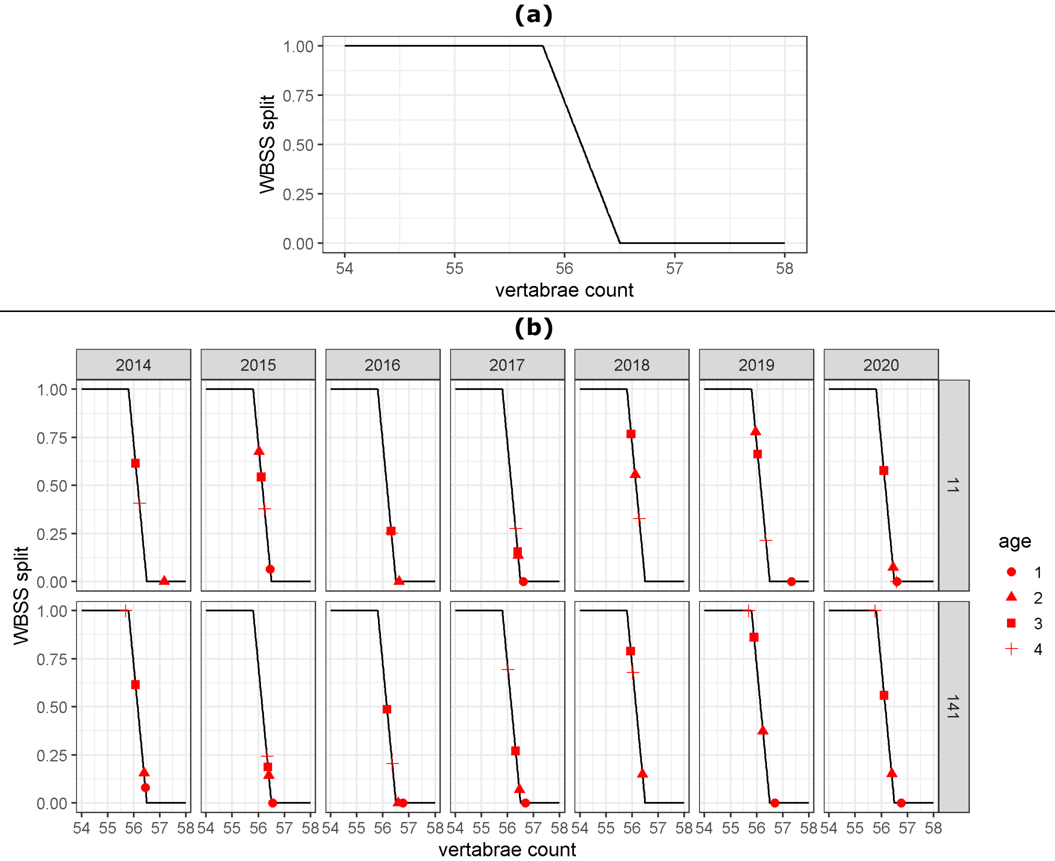


Figure 2-2: *split proportion of WBSS as applied in strata 11 and 141. (a) function to determine the split proportion of WBSS given mean vertebrae counts at a given age, strata and year. (b) WBSS split proportion function and associated values per age, strata and year.*

Moreover, the StoX software can compute the uncertainty for the abundance at age using a bootstrapping method on transects. Because of the filling in process and the use of external split proportions (in strata 11 and 141), manual inputs were required in the past. As these two steps are now automated, it is now possible to compute a bootstrap for both the side and main projects and combine projects for each iteration of the bootstrapping, in turn yielding uncertainties for the final abundance estimates. Two separate runs are then computed for the base run, one using baseline (all transects included) and another one using the Acoustic Trawl bootstrap option from StoX using 500 iterations.

2.3 Transparent Assessment Framework (TAF)

TAF is a framework developed by ICES to organize data, methods, and results used in fish stock assessments, so they are easy to reference and re-run with new data or methods. The framework uses the R language and is computed on an ICES server and dedicated website to display results. The link between the R code developed and the app is performed through individual repositories on GitHub[9]. Through this platform, all data input and output are fully traceable and versioned.

TAF is used to run various stock assessment models and compute survey indices (e.g. from trawl surveys) but there is currently no example of the computation of indices from acoustic surveys. With tools such as the ICES acoustic trawl database and the StoX software, it is now possible to automate the calculation of

[9] https://github.com/ices-taf/doc

abundance estimates for complex cases such as NSAS/WBSS and run this on TAF. With this addition, the NSAS assessment now consists of the following TAF repositories:

- Acoustic index: calculation of the HERAS indices for NSAS and WBSS (the focus of the current study).
- Single fleet stock assessment
- Multi-fleet stock assessment
- Forecast as used to provide catch advice

The working principle of a TAF repository is shown in Figure 2-3. To produce the NSAS and WBSS indices the processing is divided into four main parts:

1. Data: fetching acoustic and biotic data from the acoustic database and data stored locally on the repository. The latter includes:
    - Individual *.xml StoX project files.
    - DK and GER split acoustic data
    - Strata files
    - File with split proportion for strata 11 and 141
2. Input: organize folder tree on the repository for swift running of StoX projects.
3. Model: run side and main StoX projects (Figure 2-1)
4. Output: create outputs and plots from final estimates.

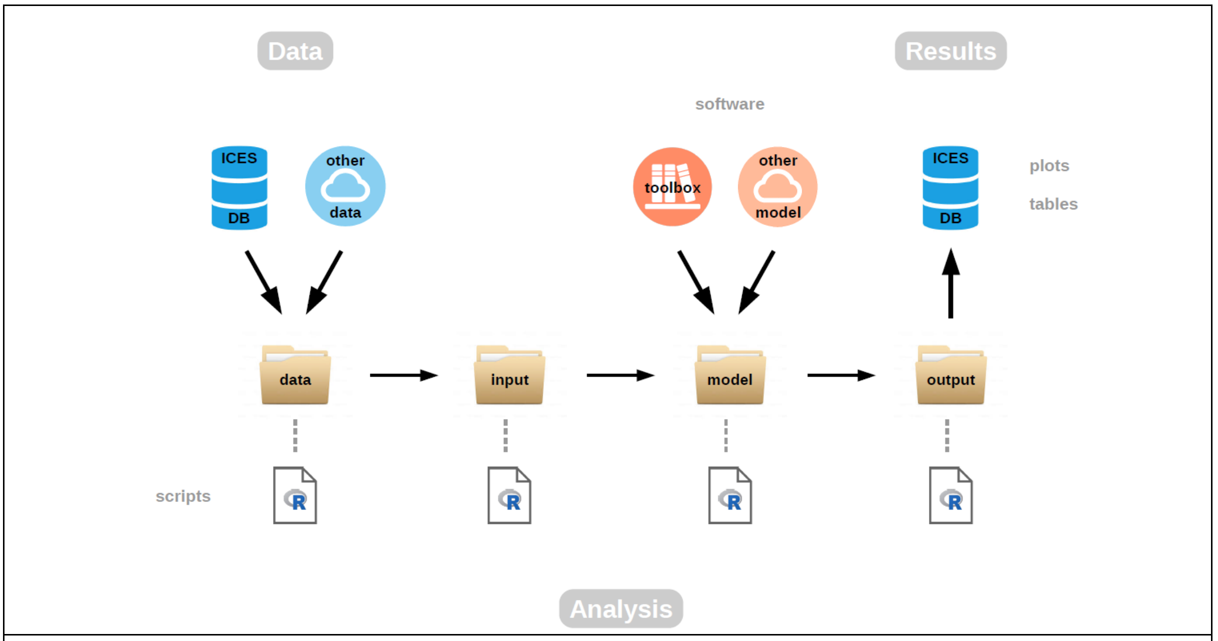


Figure 2 3: *TAF working principles from data to results. Extracted from (Millar and Magnusson 2018).*

2.4 Sensitivity test runs

2.4.1 Calibration error

During an acoustic survey such as HERAS, the calibration of the echosounder from each participating vessel is paramount to provide absolute estimates of S_v and in turn yields consistent measurements between platforms. Any offset from genuine calibration gain G directly influences abundance estimates as it impacts the level of acoustic density that is measured. Small variabilities in calibration gain can occur because of factors such as variation in sphere parameters, weather conditions during calibration trials or inherent variability in calibration measurement. Moreover, larger discrepancies can sometime arise from echosounder malfunctioning, miss handling during calibration trials or bugs in the software. In that context, it is important to investigate the impact of various levels and types of calibration offsets can have on final abundance estimates. In this study, two scenarios are considered:

- Constant offset of calibration gain.
- Nonlinear offset of S_A values based on ping to ping evaluation of malfunctioning echosounder from RV Tridens II (Dutch component of the HERAS survey).

Calibration gain offset:
Here, the sensitivity of NSAS/WBSS abundance estimates is tested by adding an error to the calibration gain as:

$$G_{\text{erroneous}} = G + e, \quad (5)$$

with e the calibration error in dB. Here, a range of errors between 0.2 and 2 dB in 0.2 dB increments are tested. Each calibration error is introduced in the raw acoustic files (input to StoX) and tested for each survey component (NL, GB-SCT, DK, NO, GER) and survey year separately (2017 to 2020).

SA error from RV Tridens II echosounder malfunctioning:
In 2020, (Sakinan and Berges 2020) used data from the Dutch component of the HERAS survey to test the comparability of the EK60 and EK80. This was done by running a multiplexer, producing ping to ping data between EK60 and EK80 for effective comparison. Similar to findings in other studies (Macaulay et al. 2018; Sakinan et al. 2018), both systems yielded similar results. However, this study revealed a malfunctioning of the EK80 38 kHz echosounder onboard RV Tridens II (vessel participating in the HERAS survey). Onboard RV Tridens II, the EK60 was used for survey purposes until 2018. Therefore, the malfunctioning of the EK80 in 2017 had no implications for the HERAS survey. The ping to ping data of opportunity produced with the malfunctioning EK80 provides insight on a type of malfunctioning that can easily be overlooked but has implications for the survey results.

The origin of the EK80 issue was tracked down to 13$^{\text{th}}$ July 2017 and is due to malfunctioning of the EK80 transceiver (not the transducer). The transition in the echogram is shown in Figure 2-3(a) and is very subtle and hard to identify in survey conditions. An analysis of the impedance level of each echosounder quadrant shows the malfunctioning more clearly Figure 2-3(b). The inspection of calibration results in perspective to older results also revealed the malfunctioning with a drop of ~2 dB in calibration gain whilst expected variations usually do not exceed 0.1-0.2 dB Figure 2-3(c). The ping to ping data set was further used to build the relationship between the EK60 data and the erroneous EK80 CW data Figure 2-3(d).

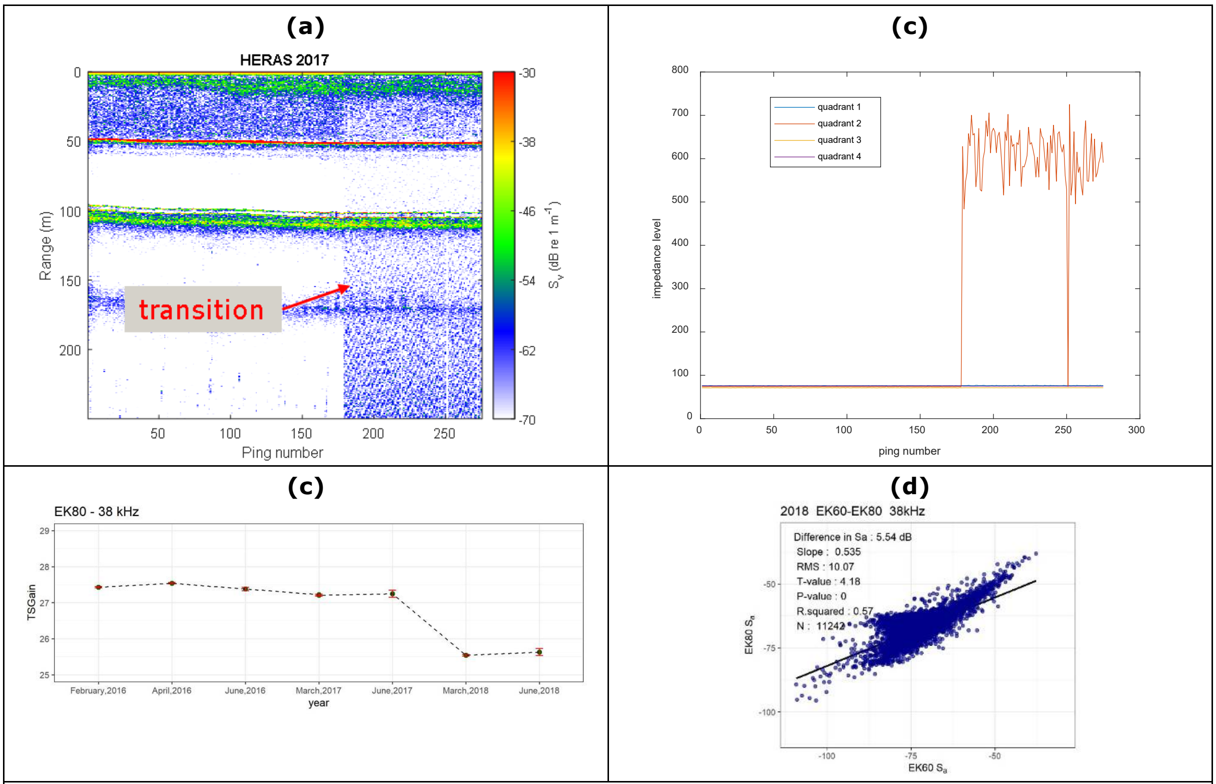


Figure 2-3: *ping to ping analysis of EK80 CW malfunctioning onboard Tridens II RFV during HERAS 2017 and 2018. (a) transition to erroneous data in echogram. (b) time series of impedance levels around the occurrence of echosounder malfunctioning. (c) comparison of results from EK80 CW calibration trials onboard Tridens II. Erroneous calibration starts on March 2018. (d) EK60 vs EK80 CW ping to ping with malfunctioning EK80 CW echosounder.*

For the current study, the relationship shown in Figure 2-3(d) is modelled using a GAM model with a smoothing function, with EK80 SA as the response variable and EK60 SA as the sole explanatory variable. Here, EK80 SA is the erroneous SA measurement whilst EK60 SA are the calibrated measurements. The model can be used to generate erroneous SA values given a random draw of initial SA values. An example distribution is shown in Figure 2-4. This process was applied to acoustic input files from the Dutch component of the survey for each of the survey years 2017-2020. For each specific year, error is introduced to each input file prior to running the StoX projects. Results for abundance at age can then be compared with the results from the base run.

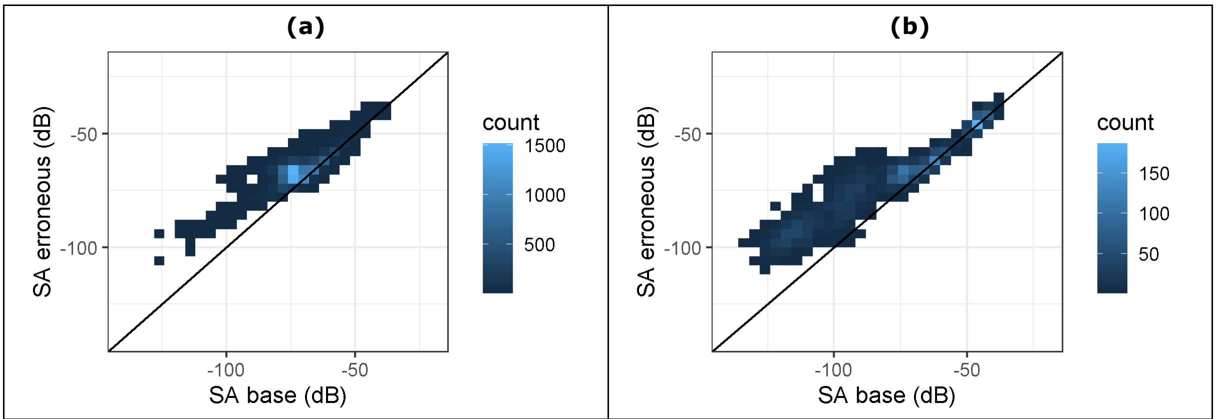


Figure 2-4: *Simulation or erroneous SA values using the study by (Sakinan and Berges 2020). (a) distribution of ping to ping data. (b) modelled distribution with randomly drawn SA values between -130 dB and 25 dB.*

2.4.2 Stock splitting

An important aspect of the calculation of NSAS/WBSS indices is the mixing between herring stocks at the border between the Baltic Sea and the North Sea. The area of interest corresponds to the following strata (Figure 1-4(c)): 11, 141, 152, 151, 42, 41, 31, 21. In these regions, mixing is taking place between NSAS and WBSS but also NSS along the Norwegian coast. Historically mixing with NSS has been assumed small and not considered for calculations of NSAS/WBSS indices. Currently, the way mixing is handled is different between strata 11/141 and other strata with mixing. In strata 11 and 141, the ratio of WBSS and NSAS is computed using Equation (4) which relies on biological measurements of vertebrae counts. The data set used for that purpose is shown in Figure 2-5 and resulting WBSS split ratios are shown in Figure 2-2(b) and Figure 2-6. Overall, it is observed that it can vary significantly between years. Of importance when calculating WBSS split ratios using Equation (4) are the 56.5 and 55.8 threshold values first introduced by (ICES 1994). Beside that these values might be somewhat outdated, (Gröger and Gröhsler 2001) noted that no stochastic assumption or theoretical background is available for Equation (4). Overall, the use of Equation (4) combined with potentially outdated vertebrae count thresholds and uncertainty associated with the use of vertebrae count for stock splitting could introduce bias to the final NSAS/WBSS estimates. Moreover, new developments in research and of stock identification methods (Berg et al. 2017) should enable a more accurate splitting of the stock in future years. It is therefore of interest to test the following:

- Sensitivity of final WBSS/NSAS estimates against variation in vertebrae count
- Use of new methods for stock splitting, namely otolith shape recognition (Berg et al. 2019) and genetics (Berg et al. 2021).

It is important to note that the testing undertaken here is only for strata 11 and 141. For other strata (152, 151, 42, 41, 31, 21), stock splitting is performed on individual herring samples using otolith microstructure and shape (Berg et al. 2021; ICES 2018b). It would be of interest to also perform sensitivity tests on stock mixing in these strata but due to the complication in implementation of such scenarios (which would require artificially altering biotic data), this is not considered in this study.

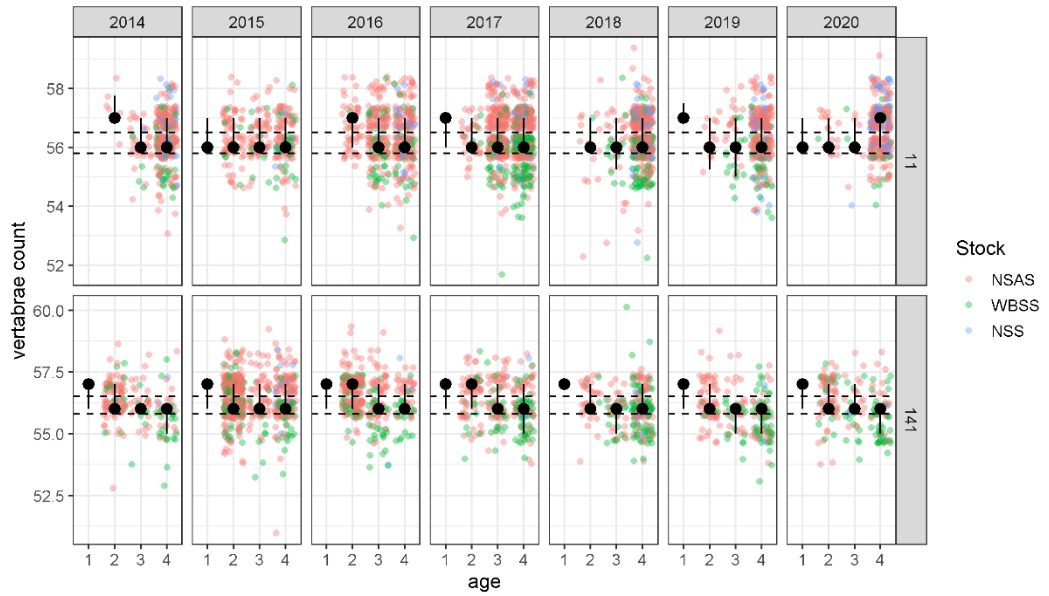


Figure 2-5: *statistics of vertebrae count in each year strata and age groups. Black circle markers are the median of vertebrae count in each category and the extend of vertical black lines are the 25th and 75th quantiles. Coloured circle markers are the vertebrae counts for each fish sample with stock identification determined using otolith shape (Berg et al. 2019): NSAS (red), WBSS (green), NSS (blue). The dashed horizontal lines are the 55.8 and 56.5 thresholds as used in Equation (4).*

Split ratio with offset in vertebrae count:

Here, the sensitivity of index calculation is tested against changes in vertebrae count that in turn changes split proportions used in strata 11 and 141. The metric used in Equation (4) is the mean vertebrae count in each category (age, year, strata) $\overline{v}_c$. For testing, this is altered using the standard deviation in vertebrae count in each category and computing a new split ratio using the following for $\overline{v}_c$:

$\overline{v_{c\ \mathrm{error}}}(a, y, s) = \overline{v}_c(a, y, s) + \beta \times \sigma(a, y, s).$	(6)

The term β is a scalor that is used to introduce a deviation in $\overline{v}_c$. The range of values tested for β are -0.2 to 0.2 in 0.1 increments. The change in split ratio is exemplified in Figure 2-6 for a limited range of β values.

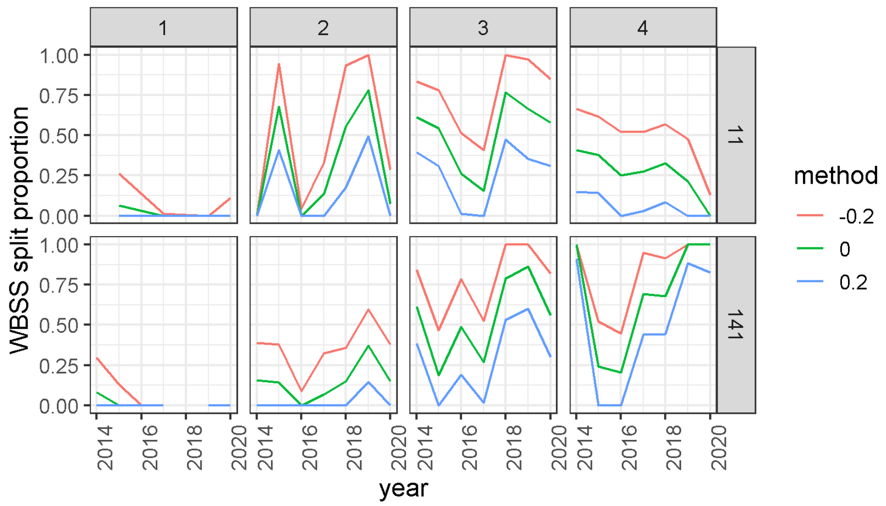


Figure 2-6: *WBSS split proportions in each age group, strata and year. Also shown is the effect of deviation from mean vertebrae count by -0.2/+0.2 sd.*

Split based on otolith shape:

The method by (Berg et al. 2019) uses otolith shape recognition using machine learning to identify the stock id of a given otolith. The discrimination is made between NSAS, WBSS and NSS. This method has been applied to the Norwegian biological samples for the period 2014-2020, yielding alternative split proportion estimates to those calculated using vertebrae count (Equation (4)). Using otolith photographs, probabilistic estimates are derived for each individual otolith. Here, a threshold of 0.75 on the probability is used to filter out individual samples that are not discriminated with a high enough confidence. The proportion of individuals in each category (age, year and strata) is then calculated. The comparison between split proportions inferred from vertebrae counts and otolith shape recognition (with a probability threshold of 0.75) is shown in Figure 2-7. There are significant differences between the results from the vertebrae count and otolith shape recognition methods. Notably, the otolith shape recognition exemplifies fewer variable results (e.g. age 2 in strata 11). Using these new split proportions, alternative NSAS/WBSS abundance estimates can be generated and compared to the base run.

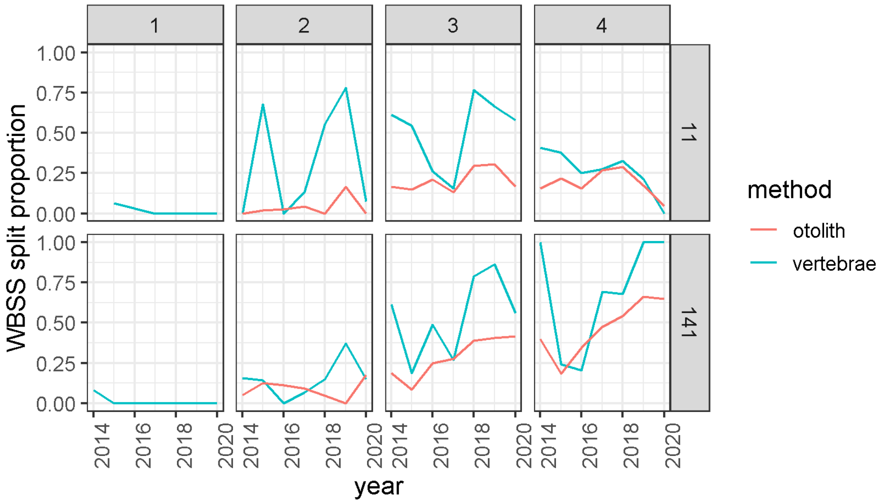


Figure 2-7: *time series of split proportion at age as derived using vertebrae count and otolith shape (Berg et al. 2019) from 2014 to 2020 in strata 11 and 141. The split proportion from the otolith method is computed using a probability threshold of 0.75.*

Split based on Genetic data:

In 2020, genetic data was available for biological samples taken in strata 11 and 141. Biological samples could thus be allocated a stock id at an individual level in the biotic files from the Norwegian survey. This allows for combining all data from individual surveys and run a single StoX project as opposed to having to break down the processing into a main project and a side project (Figure 2-1). This new project can be run as a baseline but also using transect bootstrapping for uncertainty estimation. These results can then be compared with results from the base run.

2.4.3 Strata definition

The HERAS survey extends over a wide geographical area covering the habitat of herring and sprat at their different life stages. Their distribution is heterogenous due the impact of different biological and physical drivers, such as age, reproductive state, temperature, depth, food availability etc. This heterogeneity reflects on their aggregation densities and size distribution and results in geographical patchiness and gradients. One of the objectives of the survey design is to allocate the survey effort in an optimal way with the aim to keep the survey precision as high as possible. This is achieved by stratification (Figure 1-4(c)). Normally, the main goal of the stratification is to minimize the variability within each stratum. Ideally the sampling variability between the strata should be independent from each other. Although the stratification has a direct effect on the variability in the results, the data of an individual survey cannot be used to determine the strata since the survey design is one of the factors that determines the results. For example, as a survey design strategy, in a stratum where the variability is known to be higher than the others, the effort would be increased to improve the precision.

Prior to 2016, the HERAS survey strata were either based on ICES statistical rectangles or based on historic catch data. Beginning from the year 2016 this was changed based on internal work carried out in 2015 (Lusseau et al. 2016; ICES 2015b). The new survey strata was formulated which is independent from the ICES rectangles. This new design had two main goals:

- Standardization of survey design and effort within co-surveyed strata
- Improvement in precision in the abundance estimate

This new stratum design has been in use for the surveys from 2016 to 2020. While this design was effective in maintaining homogeneity (e.g. variability in the aggregation patterns, maturity level and size distributions), some minor concerns arose during post cruise meetings for certain exceptional cases with a potential change of strata to take into account historically known north-south gradients and coastal-offshore gradients,.

A clear example is from the 2018 HERAS survey, where aggregations of both small-immature and large-mature herring were observed along the same transect in strata 91. Despite being found along the same transect, there was a large distance between these aggregations in the inshore-offshore direction. Normally, the stratum 91 is historically characterized by large – mature herring with distinct aggregations in high densities. The inshore strata 81 and 101 are more characterized by the earlier life stages and small sizes of herring. Although not as clear as in the 2018 example, this issue also occurred in strata 81 and 101 repeatedly. Having two distinct age/size groups with different aggregation behaviour and spatial structure in the same strata have the potential to violate the survey design principles.

Here an alternative strata design was created and compared against the current design (Figure 1-4(c)). For this purpose, the spatial variability in the mean length distributions, mean age and maturity levels, the biological data from individual survey hauls from 2016 to 2019 were combined for the entire survey area and distribution maps were created using geostatistics. Although all parameters show specific patterns, it was decided to use maturity proportions as it was found the most informative parameter. Figure 2-8 shows the underlying spatial structure in maturity levels. Figure 2-8(a) shows the original strata design overlaid and Figure 2-8(b) shows a slightly modified version of the current stratification with modifications to the following strata: 111, 101, 91 and 81. In this design, the inshore strata 81 and 101 are extended to follow the maturity pattern.

Based on these new polygons, the survey indices are recalculated. For the sake of this exercise, the transects are also modified such that; when a section exceeds the limit of one stratum, it is split into another section and either added to an existing transect or identified as a new transect. This trial allowed for separation of young and adult herring aggregations that were initially observed on the same transect within the same stratum. Note that, this is only a hypothetical test, and in principle, the transect design should not be changed after the survey has been conducted.

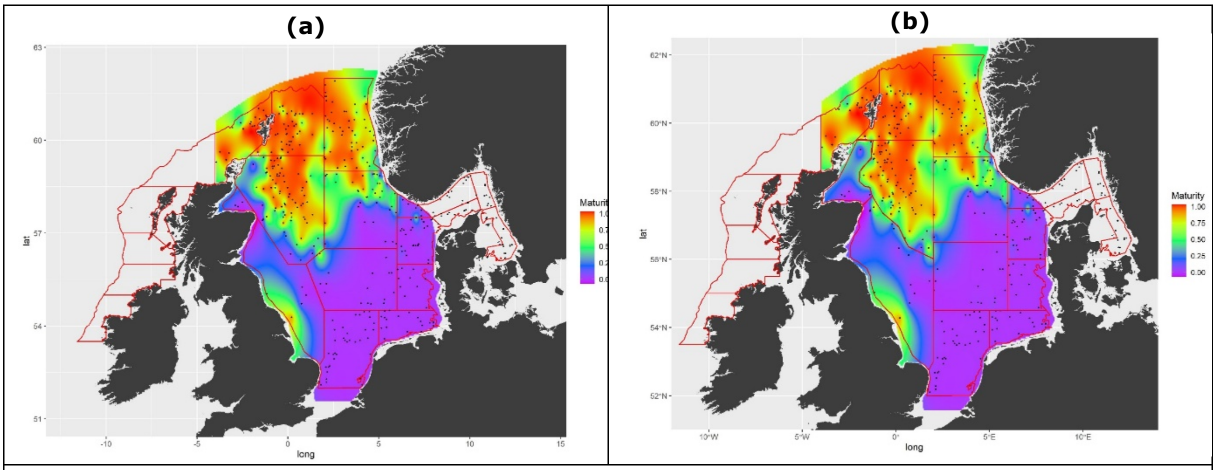


Figure 2-8: *Defining new strata based on spatial distribution of maturity levels. (a) base strata definition. (b) newly developed strata definition. The underlying colormap is the spatially interpolated distribution of maturity levels using biotic data from 2016 to 2019*

2.4.4 Haul allocation

In StoX, it is necessary to assign at least one set of biological measurement to each acoustic transect. This allows for the calculation of length dependent acoustic target strength (TS) and estimation of disaggregated abundance (age, length, maturity) for each transect taking the most representative length distribution into account. In theory, all the hauls within a stratum should be representative for all transects. However, this is not considered to be true in practice as there are local variabilities in addition to global variability and anisotropy within each stratum. Especially in strata 91 and 111 there is a well-known gradient of size with larger herring typically being further north. Therefore, the standard way of performing this allocation in the HERAS survey is to use neighboring hauls and survey knowledge with a manual assignment based on distance and expert judgment. The downside of this approach is that it introduces subjectivity and operator dependency to the analysis.

In this exercise, the difference between the manual assignment and the stratum assignment was tested in StoX. The stratum assignment basically aggregates all the samples from all the hauls within a stratum by equal weighting and assigning them to all transects in that stratum in the same way. Ideally it is expected that during the survey there are representative numbers of hauls within each stratum. However, in practice, there is limited time that can be dedicated to trawling during the survey. The location selection is generally a tradeoff between resolving local variability in high density areas and effective coverage of all strata. During the HERAS surveys, on few occasions, there were no hauls corresponding to a stratum. To be able to perform this exercise, in such cases, most representative hauls (based on expert knowledge and background information) from the neighboring strata were used. Furthermore, in all strata, the samples with too few measurements (<30 herring samples in haul) to produce a clear statistical distribution were identified as outliers and removed.

3 Results and discussion

3.1 Base run and TAF

3.1.1 Automatic calculation of the index

The method for the calculation of the WBSS and NSAS abundance indices described in Section 2 is applied to the period 2017-2020. This is the period where a well-defined workflow for processing is in place (Figure 1-4(a)). Prior to adoption of the newly automated procedure implemented in R, it is important to check the routines against potential bugs and investigate discrepancies with existing indices as used in the assessment. This is shown in Figure 3-1 for the abundance at age and SSB (in % relative to values used in assessments). For both NSAS and WBSS, whilst the discrepancy is marginal for 2018 and 2019, it is significant for 2017. In order to investigate those more in depth, a thorough check of the results for each year was performed, comparing manually derived numbers from spreadsheets and results from automatic routines. As in Figure 3-1 very good agreement was found for the years 2018-2019 (and 2020). The remaining minor discrepancies are due to small differences in filling of missing fields (Figure 2-1(b)), which are arguably correctly attributed by the automated routine, and rounding. Closer investigation of the higher differences seen in 2017 comparison, the discrepancies are mainly due to differences in WBSS split ratio for that year used in the automated process and those used in the original analysis in 2017. This needs to be followed up and an update of the index for the assessment should be performed in the future for both NSAS and WBSS.

In addition to abundance at age, the results from the HERAS survey also provide biological parameters for NSAS: weight at age in the stock and maturity at age in the stock. These are used in the calculation of SSB. The effect of the new calculation routines on these parameters is shown in Figure 3-2. Only small differences were observed, most notably in 2017. For WBSS, weight at age and maturity at age calculated from the data of the HERAS survey is not directly used in the assessment because of the timing of the survey. Therefore, no sensible comparison can be made.

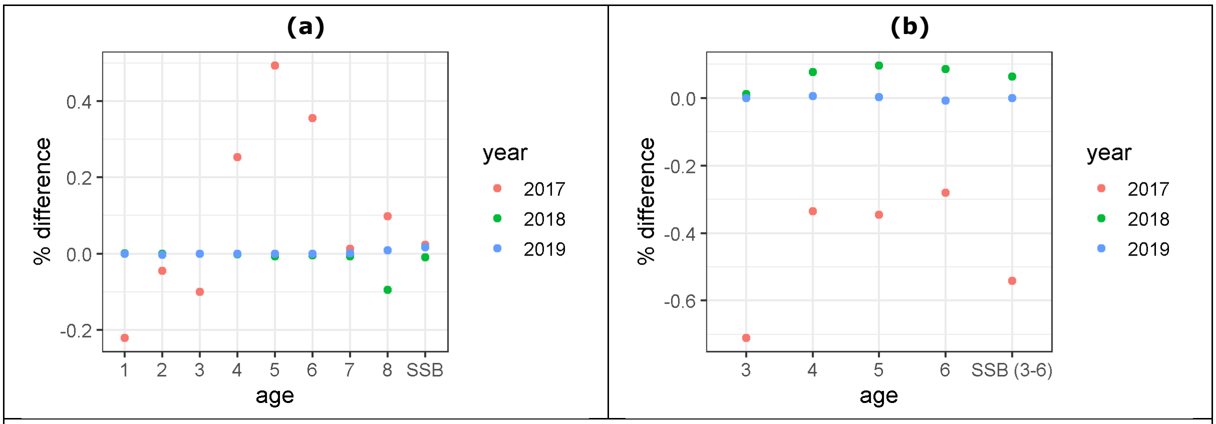


Figure 3-1: *comparison of abundance at age and SSB as computed using the automated routines against those used in the assessment. (a) NSAS. (b) WBSS. For WBSS, only ages 3-6 are used in the assessment.*

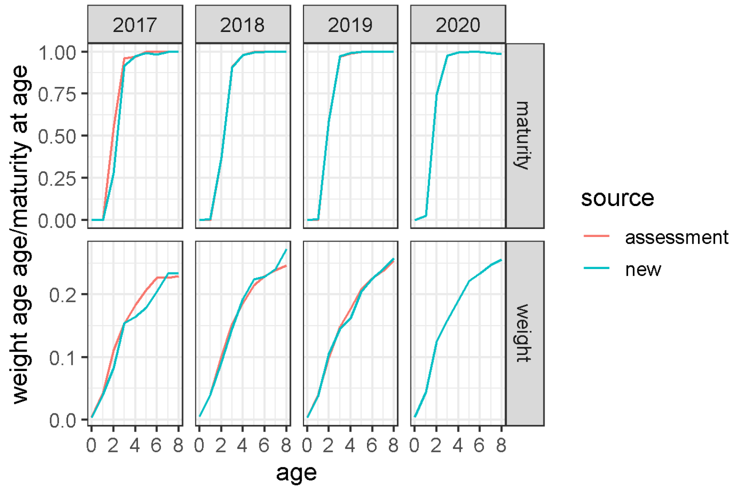


Figure 3-2: *comparison of maturity at age and weight at age computed using the automated routines against those used in the assessment for NSAS. For WBSS, these quantities can be derived from the HERAS survey but are not used in the assessment due to survey timing and are therefore not shown.*

3.1.2 2017-2020 NSAS and WBSS results

With the newly developed framework for the automated calculation of the NSAS/WBSS indices and the use of FLR, results can be plotted efficiently in a generic way. This is exemplified in Figure 3-3 and Figure 3-4 for the years 2017-2020 with abundance at age and abundance for mature/immature categories in each stratum (Figure 1-4(b)).

Results for NSAS abundance at age are shown in Figure 3-3(a) per strata and for all strata combined. It shows the bulk of the stock for ages above 2 is concentrated in the Northern strata (91, 111, 101, 121, 11). Since 2017, there is a decreasing trend in stock trajectory for NSAS (Figure 1-3(a)). This is notable from the abundance at age for all strata combined in Figure 3-3(a) with a decrease since 2018 which is driven by decrease in stock abundance in strata such as 111, 91 or 11. Strata with predominantly younger ages show large variations in abundance in recent years (e.g. strata 81, 131 or 141). The age composition of the stock is similar since 2017. The abundance for mature and immature fish as shown is overall consistent between years for the different strata (Figure 3-3(b)).

WBSS herring stock has followed a downward trajectory for the last ~15 years (Figure 1-3(c)). The abundance estimates from the HERAS survey is also an important source of information for the assessment of WBSS. The survey has with some inter-annual variations given an overall downward trend for the last 15 years, and this is reflected in the stock assessment. Though, since 2017, there is a slight increase in abundance. This is shown in Figure 3-4(a). Only strata 11 (the northern most one) shows a strong consistent decrease in abundance which is partly caused by a decrease in stock proportions in that strata (Figure 2-6). Strata 141 with the highest WBSS abundance shows steady abundance levels and even an increase in 2020. Also, to note is the large proportion of age 1 individuals observed in 2020. The proportions of mature/immature abundances remain at similar levels Figure 3-4(b).

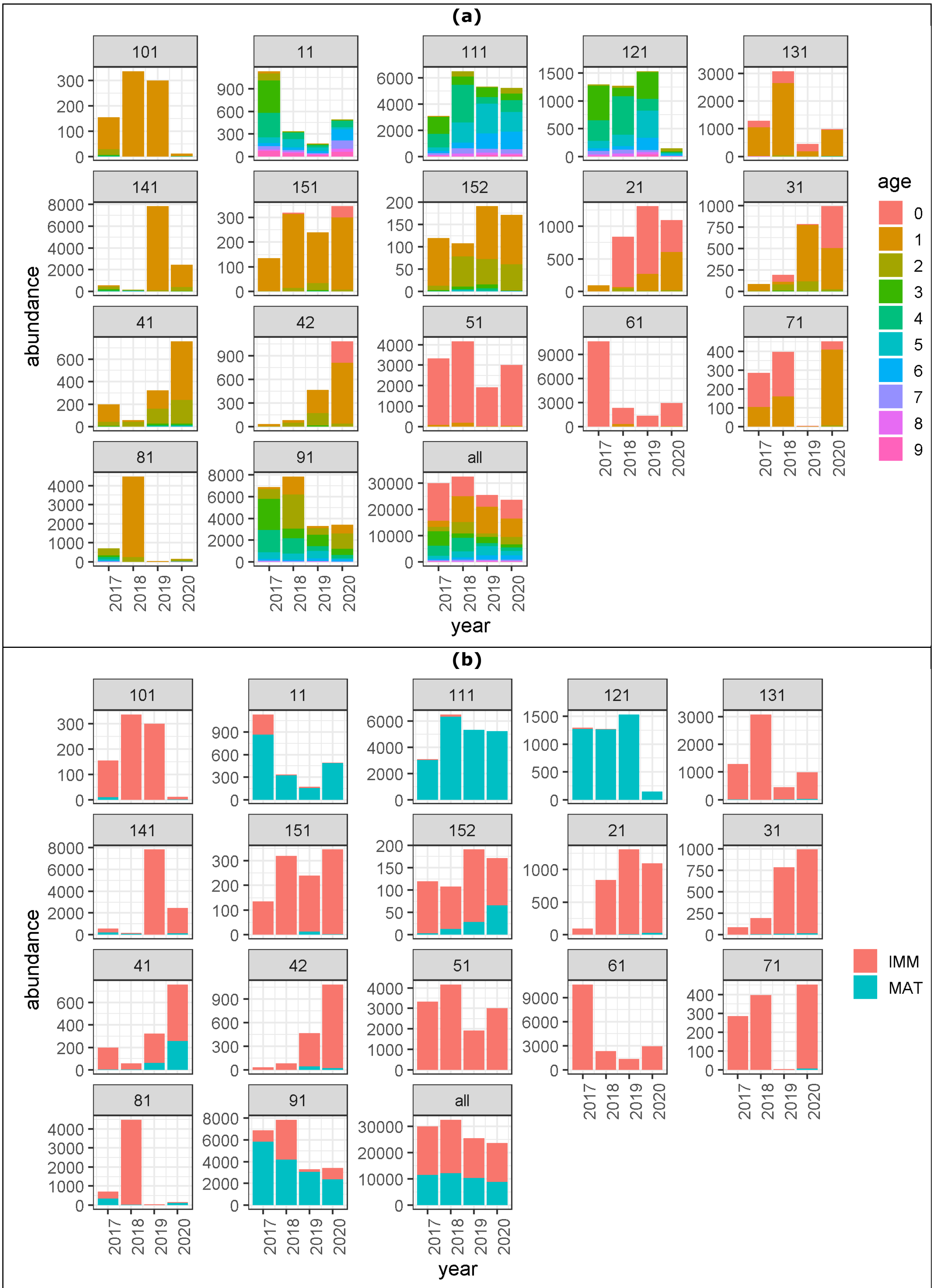


Figure 3-3: *NSAS base run abundance per strata between 2017 and 2020. (a) abundance by age group. (b) abundance for mature/immature stages*

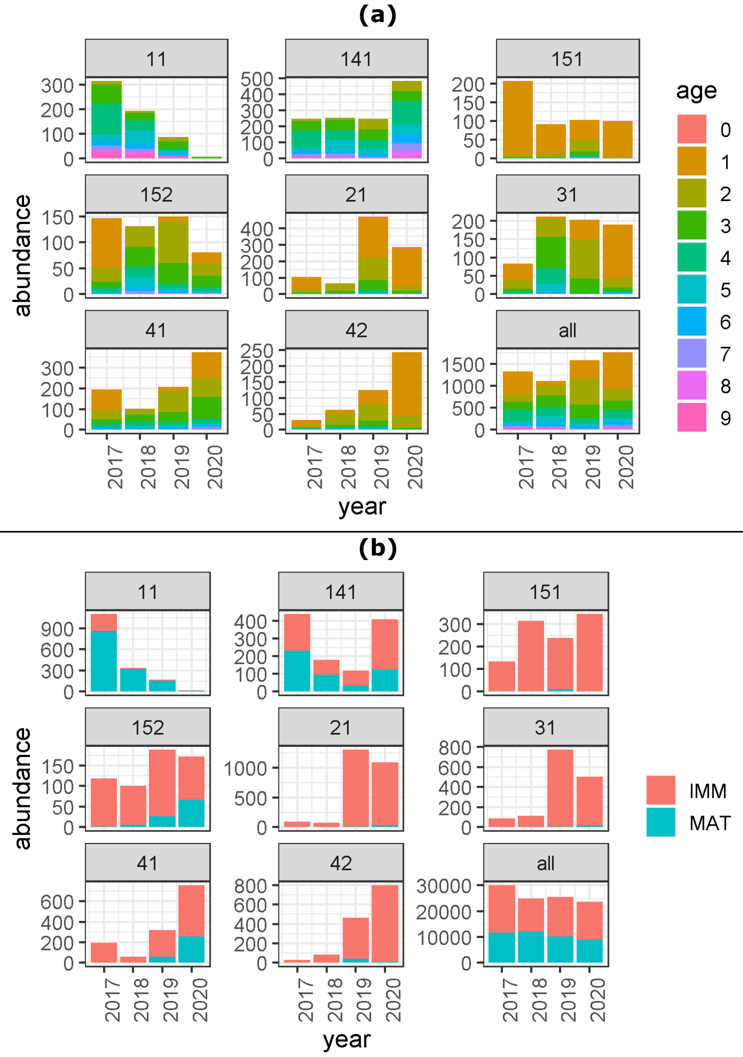


Figure 3-4: *WBSS base run abundance per strata between 2017 and 2020. (a) abundance by age group. (b) abundance for mature/immature stages.*

3.1.3 *Index uncertainty through Bootstrapping*

Because WBSS and NSAS indices are currently derived using two separate StoX projects (Figure 2-1), it has not been possible to use the acoustic trawl bootstrap feature in StoX which enables uncertainty estimation. However, in order to derive uncertainty from bootstrapping for final NSAS/WBSS abundance estimates, one needs to combine projects for each iteration with filling of missing fields (stock id, maturity, age, length) and applying split proportion ratio for strata 11 and 141. With 500 iterations, this was not possible with manual processing but is now enabled with automatic routines. For both NSAS and WBSS, uncertainties associated with the index calculation process can therefore now be estimated with ease and delivered with the annual survey results.

Uncertainties for SSB is shown in Figure 3-5. For NSAS (Figure 3-5(a)), both the 25th and 75th quantiles are within the assessment uncertainty. For WBSS, assessment uncertainty is larger than the 25th/75th quantiles interval. Moreover, a large difference exists for 2017 between the new index and the index currently used in the assessment which is attributable to data discrepancies (e.g. in split proportions) for that year (Figure 3-1). Of importance is the difference (in SSB and abundance at age) for both NSAS and WBSS between baseline estimates and median values of bootstrap. The use of median values from bootstrap is possibly a more robust estimation.

The comparison between baseline and bootstrap runs for both NSAS and WBSS is presented in Figure 3-6(a) for NSAS and Figure 3-6 (b) for WBSS. Whilst there is consistency between the results, baseline results can be somewhat higher (e.g. NSAS age 1 in 2019).

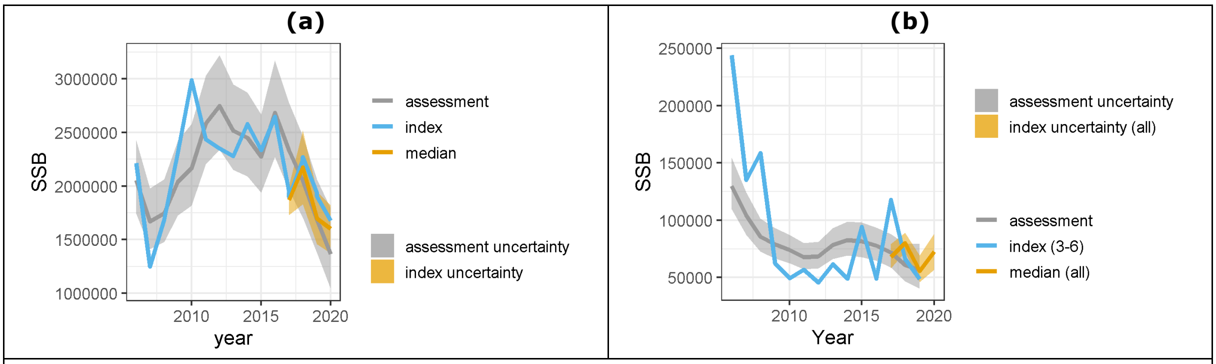


Figure 3-5: *comparison of SSB from the assessment with index sampling uncertainty derived from bootstrap computation. (a) NSAS. (b) WBSS. The index uncertainty bounds are the 25th/75th quantiles from each year's pool of independently derived SSB for each iteration.*

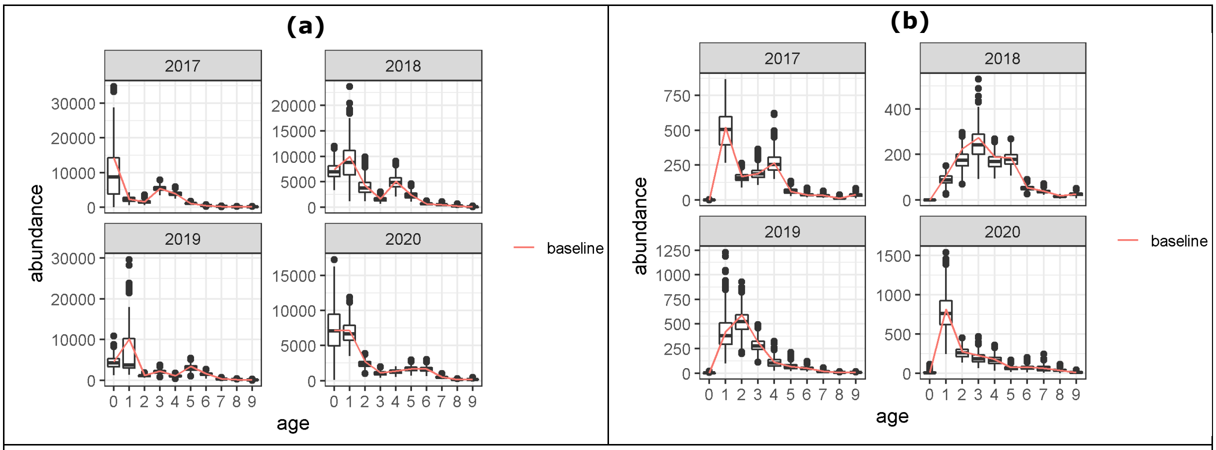


Figure 3-6: *base run abundance at age comparison between baseline and bootstrap computations. (a) NSAS. (b) WBSS.*

3.2 Sensitivity tests

3.2.1 Sensitivity to calibration error

Error in calibration gain can be introduced by either echosounder malfunctioning or measurement variability. The latter is caused by factors such as weather conditions (i.e. sea state), accuracy of sphere parameters or quality of calibration trial (e.g. number of target hits acquired). Unpredictable and hardly identifiable errors can notably come from slight malfunctioning of equipment or software/firmware bugs. Calibration errors from bugs can be difficult to spot and are often present at the release of new equipment. This has been a point of attention with the introduction of the EK80 system from SIMRAD which superseded the EK60 system, used for the last nearly 2 decades in fisheries acoustics. Even though it has been shown that the two systems yield comparable results (Macaulay et al. 2018; Demer et al. 2017; ICES 2018a, 2017; Sakinan and Berges 2020), an underlying bug in the EK80 calibration tool was identified by (Sakinan et al. 2018)[10]. As a result of this bug, (Sakinan et al. 2018) observed discrepancies of up to 1.76 dB in $s_{a\,corr}$. Considering the potential of such discrepancies or echosounder malfunctioning, it is important to screen for any calibration error prior to the start of survey, ideally during calibration trials. This can be achieved by:

- Keeping track of abnormal RMS (Root Mean Square) errors in calibration results.
- Keeping records of calibration gain and $\boldsymbol{s}_{\mathbf{a}\,corr}$ for the specific vessel, and contrasting new calibration results with historical time series, therefore tracking any suspicious changes in time.
- Monitoring impedance level of each transducer quadrant regularly using the BITE tool from the EK80 software.

Here, the sensitivity of WBSS/NSAS index calculations are tested against two scenarios: an offset in gain, like an error in calibration and an echosounder malfunctioning as observed in 2018 during the Dutch component of the HERAS survey.

<u>*Calibration gain offset:*</u>
In order to test the effect of bugs or calibration error, a fixed offset in calibration was introduced in each acoustic input file separately. Results are shown in Figure 3-7 (NSAS) and Figure 3-8 (WBSS) for each participating country from 2017 to 2020.

For NSAS, the effect of calibration error on SSB is greatest for the Scottish and Dutch surveys as they cover the bulk of the adult stock. For 2020, the effect is limited for the Dutch survey as the distribution of herring was observed to have shifted further north. Depending on the extend of calibration error, a change in SSB of up to 50% can be observed (2dB calibration error, 2020 Scottish survey). In that context, it is interesting to note that the large spike in 2010 (Figure 1-3(a)) is apparently attributed to the handling of calibration results in the postprocessing software and would need to be corrected when running the calculation method back in time. Errors in calibration on other participants would have a smaller impact on absolute SSB levels. However, the impact can be substantial for specific ages, e.g. 2019 NO survey having a 50% decrease with a 2 dB error. Besides a bias in absolute level, such errors in abundance at age can also lead to inconsistencies in the age structure because of the relative bias of the erroneous survey with the other surveys. Such an inconsistency in age structure can lead to issues when running the stock assessment model, especially if the calibration error is both variable in scale and persistent in time.

For WBSS, error in both the Norwegian and Danish surveys are very influential in term of age structure and SSB level as both surveys observe a large part of the stock. Because WBSS relies on two individual surveys (as opposed to five for NSAS), error in calibration is potentially more influential.

---

[10] https://www.echoview.com/products-services/news/important-information-for-simrad-ek80-software-users

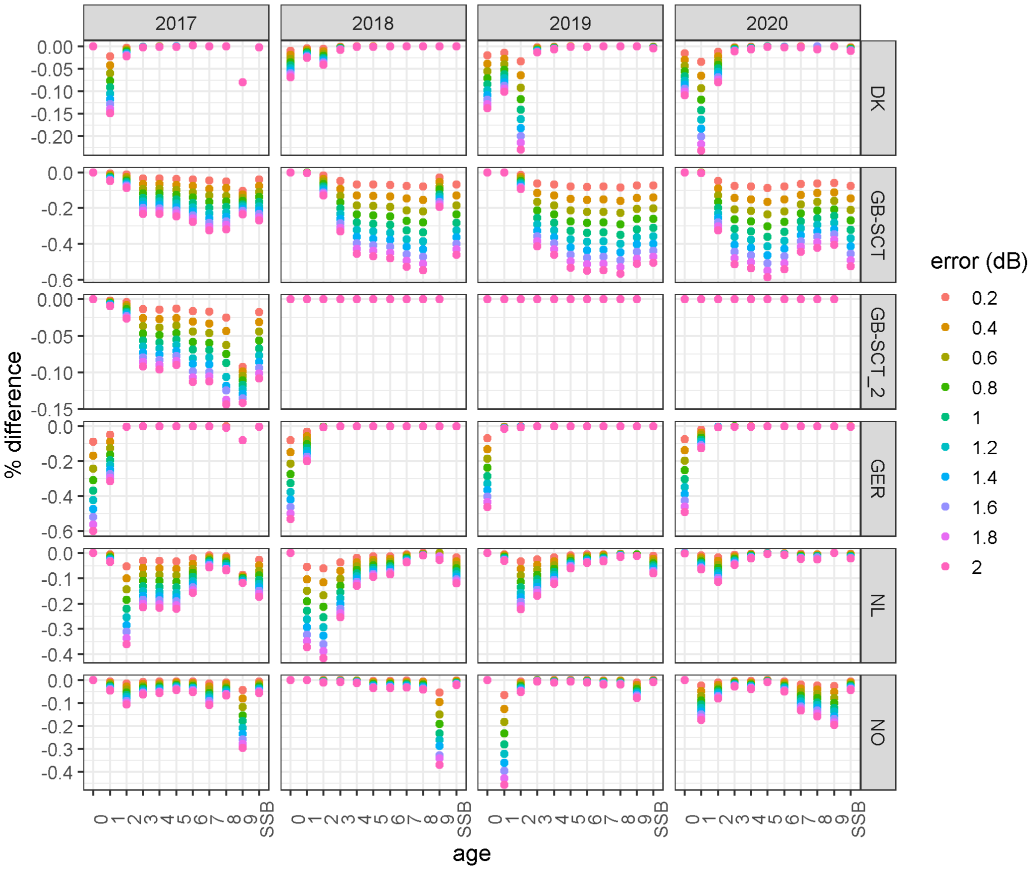


Figure 3-7: NSAS *% difference relative to base run abundance at age and SSB for a range of calibration errors by country and year of survey. For 2017, the Scottish part of the survey was covered through two distinct surveys, here identified as GB-SCT and GB-SCT_2. The % difference is relative to the baseline of the base run.*

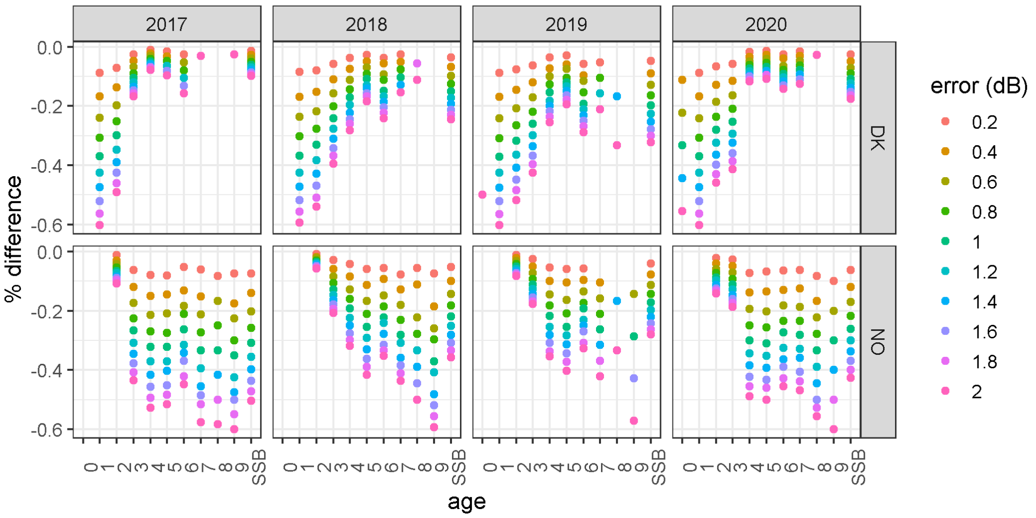


Figure 3-8: *WBSS % difference relative to base run abundance at age and SSB for a range of calibration errors by country and year of survey.*

SA error from RV Tridens II echosounder malfunctioning:

Further to calibration gain offset, the echosounder malfunctioning revealed by the ping to ping analysis by (Sakinan and Berges 2020) can provide realistic insights on the impact of such an issue. Because the ping to ping data used here were collected onboard RV Tridens II, the vessel conducting the Dutch component of the HERAS survey, the test carried out here only considers alteration of NL data. Using the mapping of genuine SA values to erroneous SA values as presented in Figure 2-4, the SA values in the NL acoustic input files are altered. The NSAS/WBSS automatic routines are then applied, yielding alternative NSAS index (the Dutch component of the survey does not cover a stratum with WBSS/NSAS mixing). The comparison of the results with the base run is presented in Figure 3-9 for the strata of interest (101, 81 and 91) and for the overall NSAS index from 2017 to 2020. Whilst the effect for 2020 and 2019 are limited in terms of SSB deviation (~1%), it is higher for 2018 (5%) and 2017 (6%). For 2018, the discrepancy is driven by differences in strata 91 whilst for 2017, it originates from strata 81. The direct effect of these discrepancies on the assessment is shown in Figure 3-10(a) for the model estimates of SSB and in Figure 3-10(b) for the catchability of the survey. Results are given in % deviation relative to the 2020 assessment (mostly using data up to 2019) for the cases of erroneous SA in 2017, 2018, 2019 and over the 2017-2019 period. From these plots, it is clear that the most influential year is 2017 but erroneous SA over the 2017-2019 period yield the largest differences in both SSB and catchability. It is also important to note that to be able to run the assessment with erroneous SA over the 2017-2019 period, it was necessary to provide initial values to assessment model for convergence, suggesting greater instability with the introduction of these erroneous data.

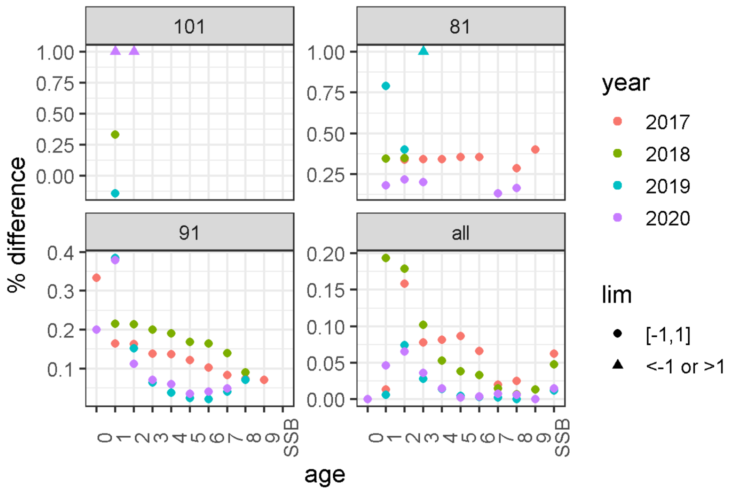


Figure 3-9: *error in NSAS abundance at age and SSB induced by malfunctioning of echosounder (SA error). Results are given as % difference relative to base run for each affected strata and combined strata. Triangle markers show values that are higher than 1 for display purposes.*

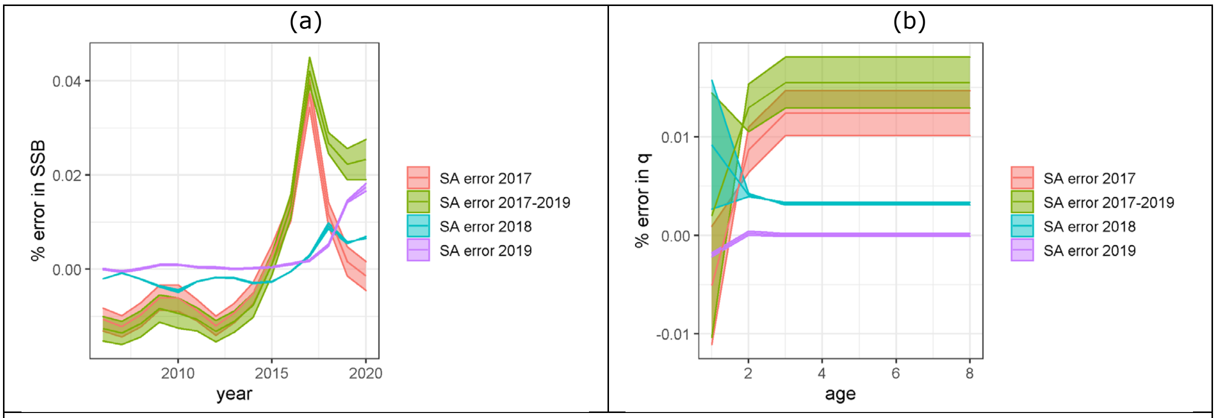


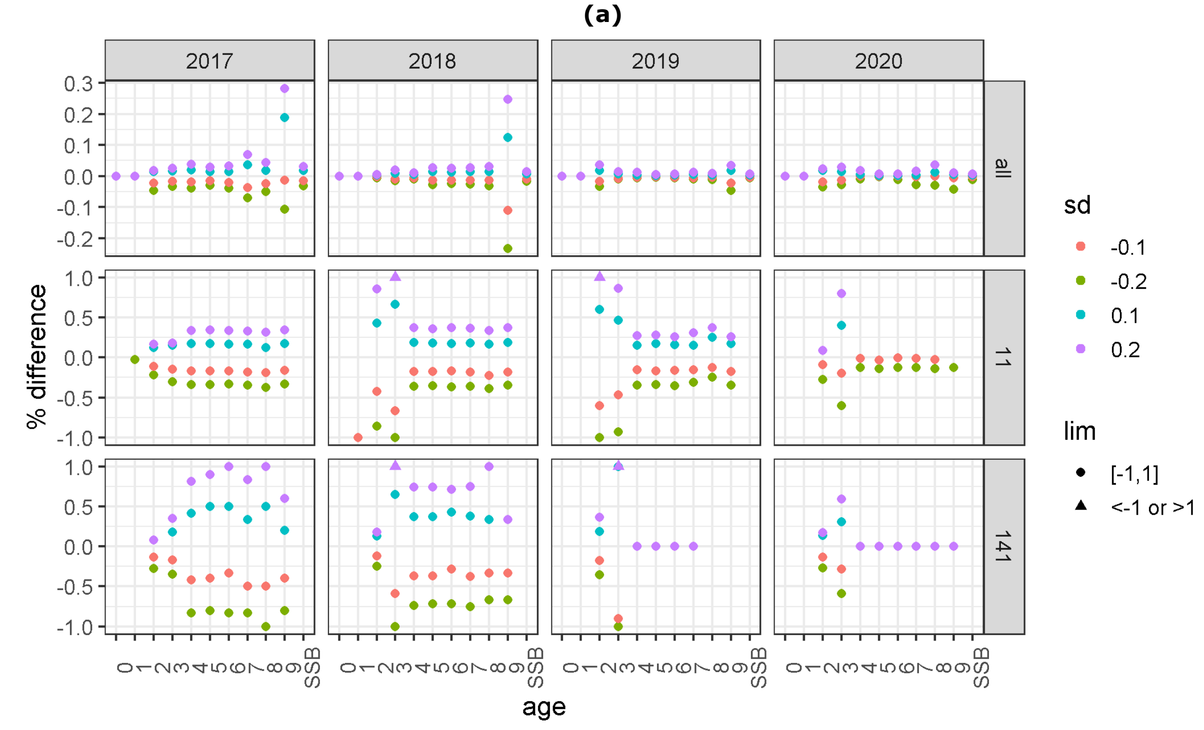
Figure 3-10: *error in NSAS assessment by malfunctioning of echosounder (SA error). (a) difference in SSB trajectory. (b) difference in catchability for the HERAS survey.*

3.2.2 Stock splitting

Stock splitting is an important aspect of the WBSS/NSAS calculation. As described in Section 2, this is handled differently between the Norwegian survey and the Danish survey. Using otolith microstructure and shape, the data issued by the Danish survey provide stock identification at individual fish level. However, this is currently not possible for the Norwegian survey and a split of the abundance is applied instead. This split is currently undertaken using a simple formula (Equation (4)) based on vertebrae count (ICES 1994; Gröger and Gröhsler 2001). The use of such a method is surrounded with uncertainties as the split proportion is highly variable in time (Figure 2-6). Moreover, newer more accurate methods are now available. In this section, three scenarios are tested:

- WBSS/NSAS split ratio with deviation in vertebrae counts**.**
- WBSS/NSAS split ratio inferred from otolith shape.
- Stock identification using genetics**.**

<u>*Split ratio with offset in vertebrae count:*</u>
The mean vertebrae count in each category (age, year, strata) is altered using a scalar β on the standard deviation, as described in Section 2.4.2. The range of values used is -0.2 to 0.2 in 0.1 increments. This introduces an error in split ratio between WBSS and NSAS (Figure 2-6). Results are shown in Figure 3-11(a) for NSAS and Figure 3-11(b) for WBSS. Expectedly, the impact is larger for WBSS with a change of ~20% with $\beta = \pm 0.2$. For WBSS, any change in split ratio has the potential to impact the age structure of the index significantly whilst discrepancies induced for NSAS are more marginal relative to the total abundance for this stock.



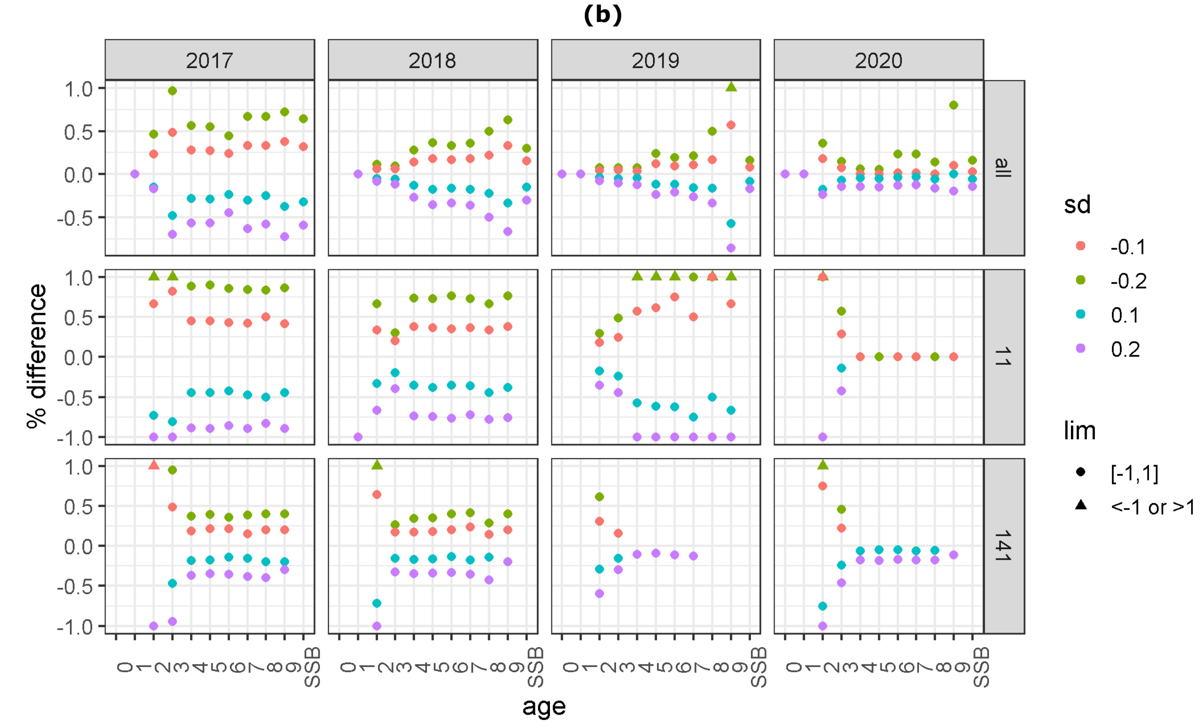
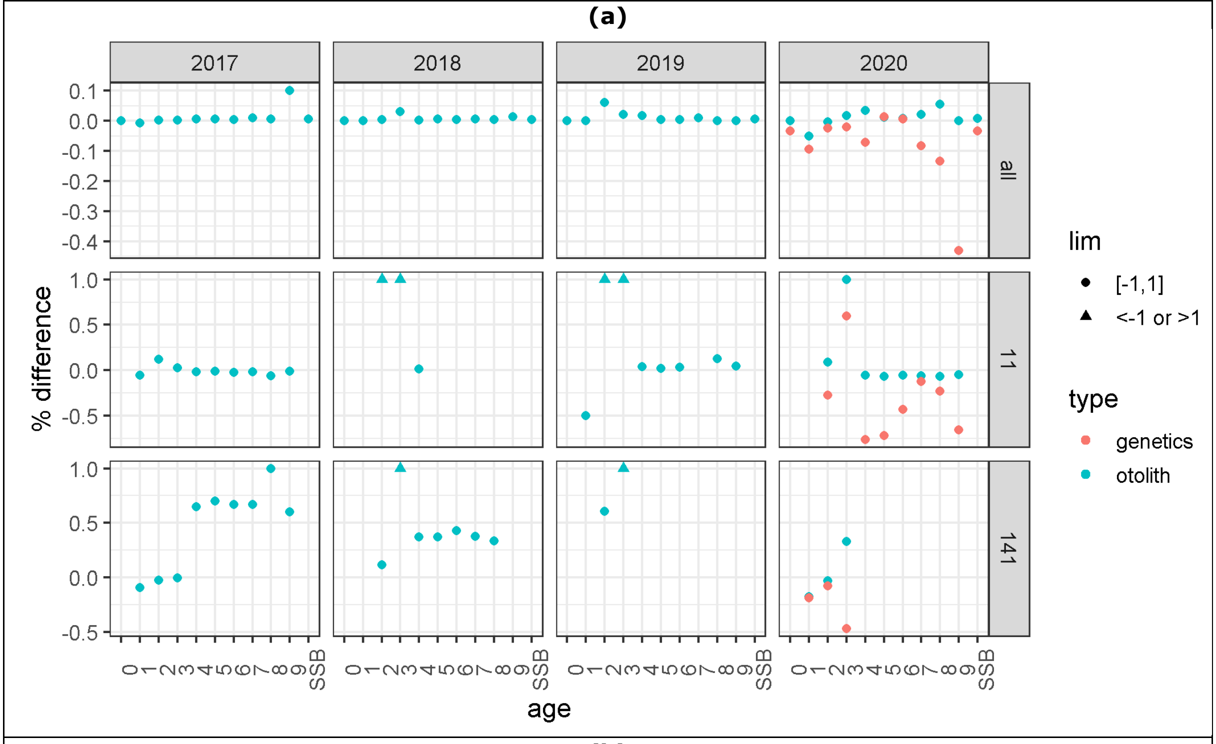


Figure 3-11: *% difference relative to base run for abundance at age and SSB with varying alternative ratios (offset in vertebrae count). (a) NSAS. (b) WBSS. Triangle markers show values that are higher than 1 for display purposes.*

Split based on otolith shape and genetics:

The use of vertebrae counts and derived split proportions has two main flaws: 1) it forces the use of two separate StoX projects (Figure 2-1), therefore complicating the calculations, 2) it is highly variable and sensitive. Whilst strata 11 and 141 only account for a small proportion of NSAS, a large proportion of the WBSS stock is observed in these strata each year. As shown in Figure 3-11(b), a small deviation in vertebrae counts can lead to discrepancies of ~20%. In that context it will be advantageous to change the NSAS/WBSS mixing calculation procedure in strata 11 and 141 using newer methods. In recent years, methods such as otolith shape recognition (Berg et al. 2019) or genetics (Berg et al. 2021) have been introduced. For the HERAS survey in strata 11 and 141, data from otolith shape recognition is available for the period 2014-2020 but genetics were only applied in 2020. Using these stock splitting methods, alternative computation of NSAS/WBSS indices are computed and compared to the base run. This is shown in Figure 3-12 for both stocks. It can be observed that the change induced by the use of otolith recognition for NSAS (Figure 3-12(a)) is marginal (<1% in SSB for all years,) whilst the use of genetics is more influential though limited (~5%). For WBSS (Figure 3-12(b)), the use of genetics data leads to large discrepancies in abundance at age (e.g. age 0, 7, 8 and 9+) but the overall change in SSB is limited (~10%). Using data from otolith shape recognition gives larger discrepancies with a decrease of SSB of ~25% in 2020.



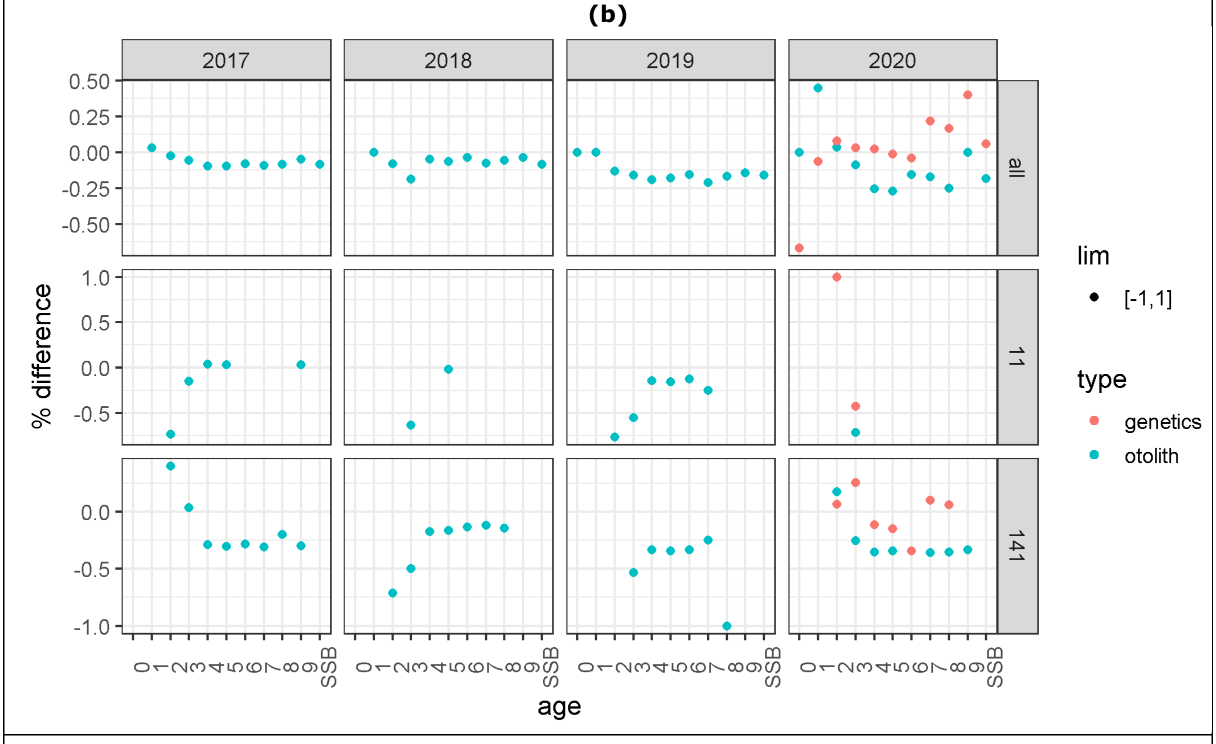


Figure 3-12: *% differences from baseline with genetics and otolith structure.*

One advantage in the use of genetics data is the allocation of stock ID in the biotic files. This lifts the constraint of having two separate StoX projects (Figure 2-1) and allows to compute the WBSS/NSAS indices using the same procedure in all strata. It also simplifies the use of the bootstrapping in StoX (used to compute uncertainty), alleviating the need to combine iterations from two separate bootstrapped projects. The comparison between baseline and bootstrap runs for both the base run and the run using genetic data is presented in Figure 3-13. Though trends are similar, some differences can be observed, e.g. at age 1 for both stocks.

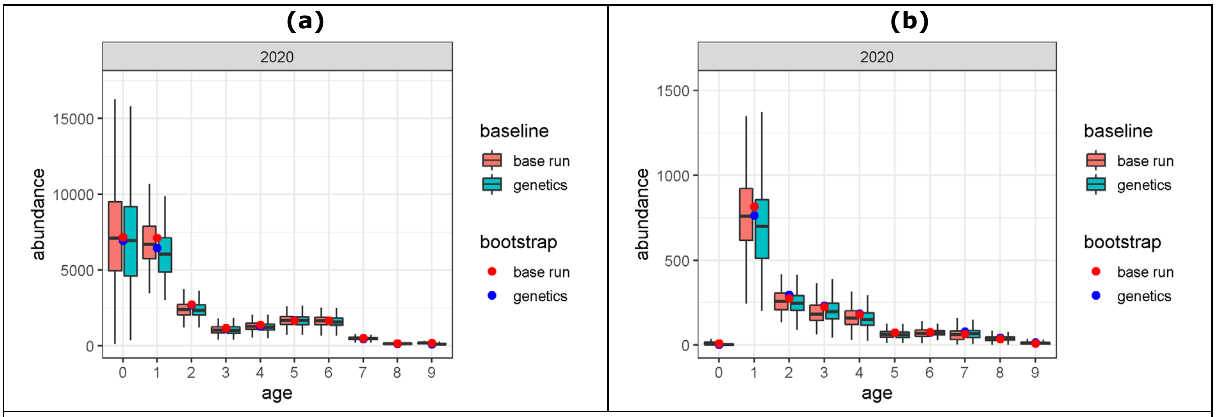


Figure 3-13: *comparison of baseline and bootstrap runs. (a) NSAS. (b) WBSS.*

3.2.3 *Change in strata definition*

A set of tests were performed to compare the effect of the design of the original strata and the modified strata. Figure 3-14 shows both the original and modified strata together with their ID numbers. In terms of areal coverage, the largest change was in stratum 101, corresponding to the Moray Firth region. Originally this is the smallest stratum in the survey area and considered as an important nursery spot for herring, therefore characterized mainly by small sized early stages herring as well as sprat and other small fish. On the other hand, the neighboring stratum in the east, stratum 91 captures the main adult herring aggregation areas in the survey region, generally containing nearly 50% of the total adult abundance. Together with stratum 111, they make up more than 90% of adult herring.

The extrapolation based on aggregated data set of maturities showed that the characteristics of stratum 101 extent a bit further to the East and North (Figure 2-8). As a result, in the hypothetical design, this stratum increased in size by 3-fold. Similarly, the inshore stratum 81, also increased in size by 32 percent while stratum 91 shrunk by 27 percent (Table 1).

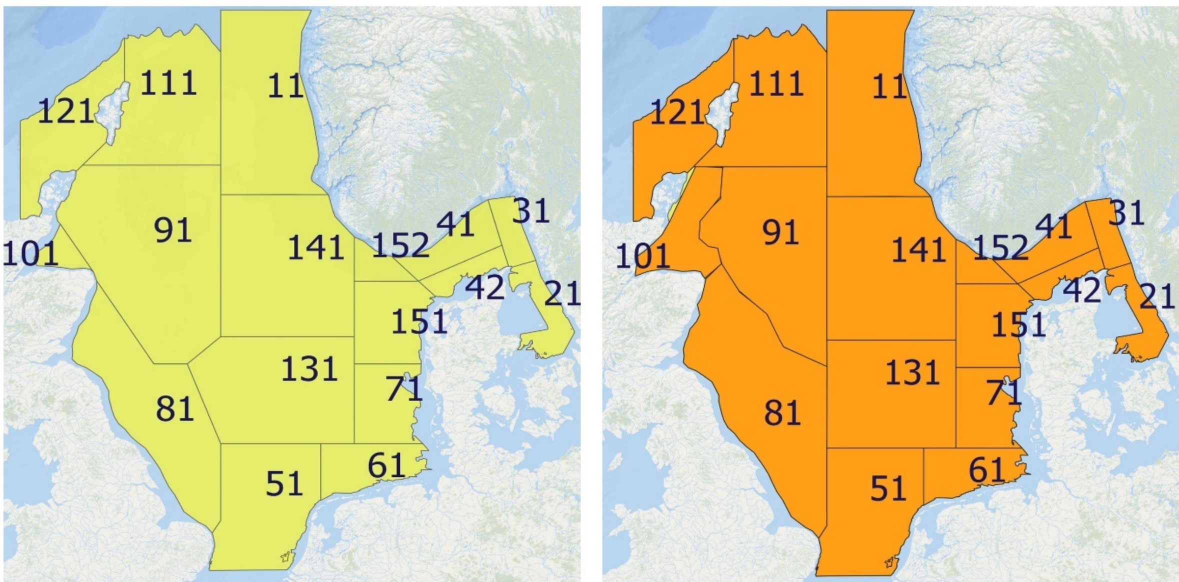


Figure 3-14: *Map showing the HERAS Strata Id numbers. The left panel shows the original strata and the right panel shows a hypothetical stratum where stratum 81,91,101 and 131 are modified based on the maturity percentages in the samples from 2016 to 2019.*

Table 1: *The strata area in nautical square miles for both original and modified strata.*

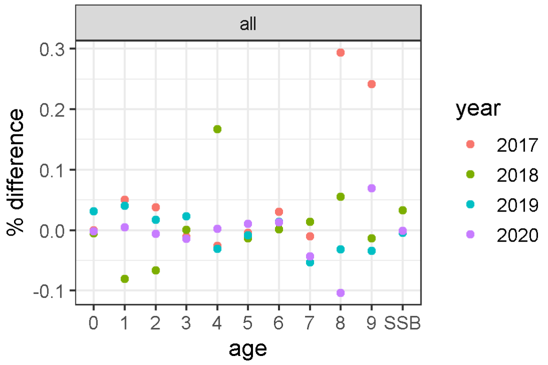
Stratum ID	Original Area	Modified Area	Percentage changed
11	14174	14174	0.00%
21	2981	2950	0.00%
31	1809	1809	0.00%
41	3377	3377	0.00%
42	1742	1742	0.00%
51	14919	14835	0.00%
61	5321	5307	0.00%
71	5521	5521	0.00%
81	17593	23213	31.94%
91	24558	17872	-27.23%
101	1272	4829	279.55%
111	11550	11550	0.00%
121	6968	6968	0.00%
131	18375	16328	-11.14%
141	18569	18569	0.00%
151	6103	6103	0.00%
152	1596	1596	0.00%

The effect of the strata modification is twofold: 1) it changes the total area of strata and 2) it changes allocation of acoustic transects and trawl samples. As an example, the results are shown in Table 2 for the most important adult herring strata, the 91 and 111 for the year 2018. A more complete table showing the rate of changes based on test scenarios are given on Table 3. In the test scenario where the haul allocation remained nearly the same while transects were split and reallocated, the total abundance of mature herring decreased only by 6 % in stratum 91 in 2018 despite a 27% reduction in the total area of this stratum. While the change in adult abundance being minor for this stratum, the total contribution of immature herring is largely decreased (by 38 %). This result indicates that overall, the new design achieves a better separation in the geographical distribution between mature and immature herring. For the stratum 111, the strata definition had not changed therefore no difference in the outputs were observed based on the stratum definition. Figure 3-15(a) shows the changes in abundance at age and SSB for the entire survey area for different years based on the changed strata definition. One can observe that the overall impact is marginal. As for the proportion of mature/immature individuals, this is shown in Figure 3-15(b). A small change is observable in 2018 but proportions overall remain like those from the base run.

Table 2: *Results of the tests of 4 different assumptions for the two most important stratum, example.*

Results for Stratum 91				
Design	**Haul Allocation**	**IMM**	**MAT**	**TOTAL**
Original	Original	3666263	4165122	7849550
Modified	Original	2294023	3909307	6216679
	Change:	**-37%**	**-6%**	**-21%**
Modified	By Strata	2588629	3746531	6344355
	Change:	**-29%**	**-10%**	**-19%**
original	By Strata	3107573	4376850	7499487
	Change:	**15%**	**-5%**	**-4%**
Results for Stratum 111				
Design	**Haul Allocation**	**IMM**	**MAT**	**TOTAL**
Original	Original	178606	6311074	6546024
Modified	Original	176347	6313109	6546024
	Change:	**-0.1%**	**0%**	**0%**
Modified	By Strata	76917	6602956	6738959
	Change:	**-57%**	**5%**	**3%**
Original	By Strata	88172	6491553	6657163
	Change:	**-51%**	**-3%**	**-2%**

(a)



(b)

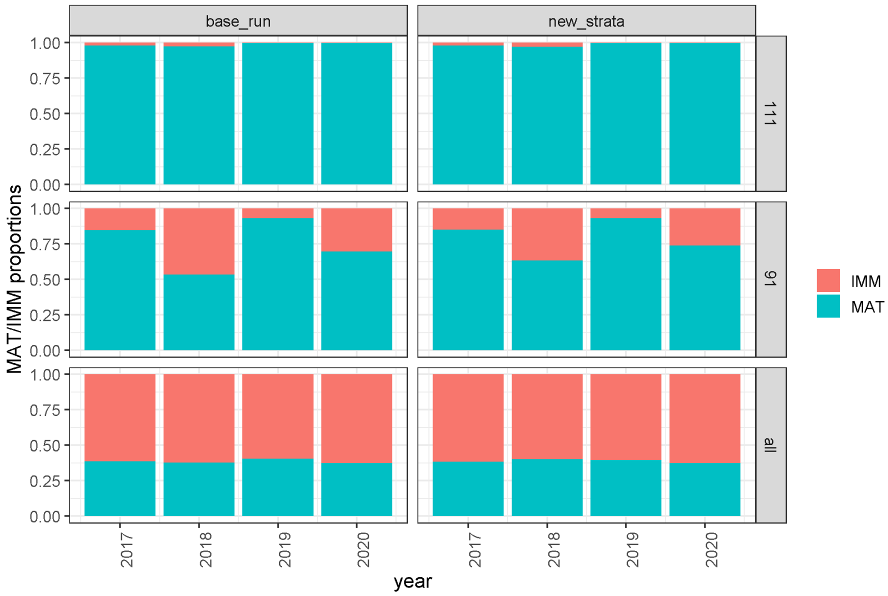


Figure 3-15: *Impact of change in strata definition with a stratum following maturity contour. (a) NSAS differences in abundance. (b) NSAS differences in mature/immature abundance at age proportions for strata 91 and 111 and overall results.*

3.2.4 Haul allocation

The changed method in haul allocation from manual assignment to stratum assignment did not have a large effect in the estimated mature herring numbers for stratum 91 and 111 where the changes were 5% and 3% respectively for 2018 (Table 3). However, the effect was large for the estimated immature numbers. In stratum 91, the immature numbers increased by 15% and decreased by 51% in stratum 111. In 2018, immature numbers increased by 28.9% for this scenario. In 2019, this method of haul allocation resulted strikingly 316% and 140% increase in the immature numbers for the stratum 91 and 111. This discrepancy shows the importance of the expert interpretation of the assignments during the post-cruise meetings. Although it can be considered as subjective, this is mainly incorporation of the observations during the survey in terms of aggregation characteristics and their most likely demographic characteristics. Currently there is no straightforward method to systematically incorporate this knowledge into an automated procedure. The Figure 3-16 show the changes based on each age group and SSB for the entire survey area for different years based on the changed strata definition. The differences are very significant with for example a change in age 1 abundance in 2017 for NSAS of ~30% (Figure 3-16(a)). For WBSS (Figure 3-16(b)), the change is substantial, especially for 2019. Overall, these results illustrate the sensitivity of index calculation to haul allocation and the need for expert knowledge and human input.

Table 3: *Percent changes based on the original project in the results from strata 91 and 111.*

Year	Stratum	Design	Allocation	IMM	MAT	TOTAL
2017	91	Modified	Manual	-12%	-8%	-8%
2017	91	Modified	Stratum based	-28%	-8%	-11%
2017	91	Original	Stratum based	8%	-2%	-1%
2017	111	Modified	Manual	-1%	0%	0%
2017	111	Modified	Stratum based	-57%	5%	3%
2017	111	Original	Stratum based	-51%	3%	2%
2018	91	Modified	Manual	-37%	-6%	-21%
2018	91	Modified	Stratum based	-29%	-10%	-19%
2018	91	Original	Stratum based	-15%	5%	-4%
2018	111	Modified	Manual	-1%	0%	0%
2018	111	Modified	Stratum based	-57%	5%	3%
2018	111	Original	Stratum based	-51%	3%	2%
2019	91	Modified	Manual	-12%	-10%	-11%
2019	91	Modified	Stratum based	29%	-12%	-9%
2019	91	Original	Stratum based	316%	-11%	11%
2019	111	Modified	Manual	-7%	0%	0%
2019	111	Modified	Stratum based	140%	3%	3%
2019	111	Original	Stratum based	122%	3%	3%
2020	91	Modified	Manual	-20%	-1%	-7%
2020	91	Modified	Stratum based	53%	-12%	7%
2020	91	Original	Stratum based	9%	-33%	-21%

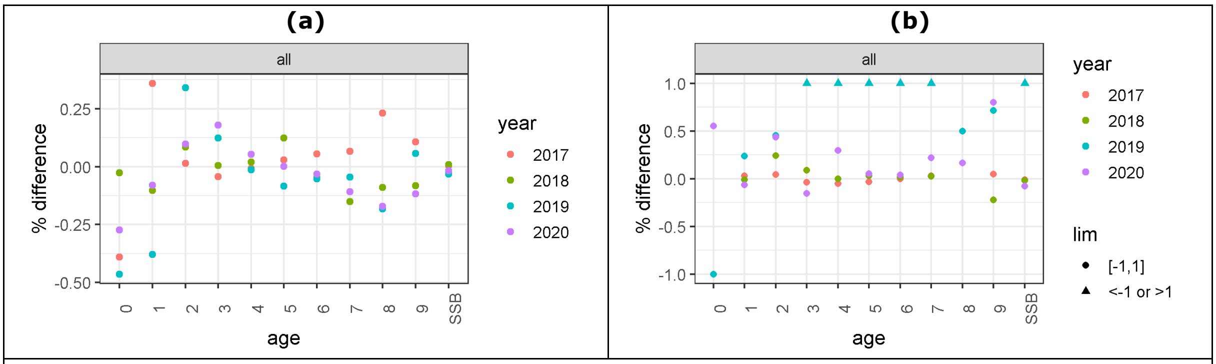


Figure 3-16: *difference in abundance with alternative haul allocation per strata. (a) NSAS. (b) WBSS.*

4 Conclusion and leads for future work

Automatic routines for the calculation of WBSS and NSAS indices were developed through this project. In the future, this will allow the process to be more traceable and will expand possibilities of index calculation (e.g. deriving uncertainties). Prior to the adoption of the automatic processing, it was important to perform thorough checks of results to ensure no discrepancies with the process currently in place. For the period 2018-2020, very good agreement was found between manually and automatically derived results with small discrepancies only accountable to bugs in the manual procedure or to difficult decision making. Large discrepancies were found for 2017, mostly due to differences in split proportions used to split the WBSS/NSAS stocks in strata 11 and 141. This should be investigated further. Overall, it is recommended to use the automatic procedure described here to compute the WBSS/NSAS indices as it offers advantages in terms of processing time, traceability and consistency. However, in the forthcoming years, the manual procedure should still be run alongside the automated routines to cross check results to prevent potential overlooked bugs and inconsistencies. Besides, it should be investigated if the processing of survey data prior to 2017 can be performed according to the latest processing workflow, alongside the processing of forthcoming HERAS survey data. With the automated procedures it will also likely be possible to implement better decision-making tools for the allocation of unallocated "super individuals" to category. The automated procedure will be able to handle more complicated decision making than the manual procedure could, for example performing a split out on proportions rather than the present procedure where all are assigned to where the majority lies for example. Once a large portion of the time series has been handled using the newly developed automatic routines, and agreements to the most correct allocation algorithms have been made, a revision of the HERAS index should be performed for both NSAS and WBSS.

The code developed through this project was stored on git repositories and is fully accessible. This includes a repository used for development[11] and repositories for the calculation of the 2019[12] and 2020[13] indices under the ICES TAF framework. The integration of calculation of acoustic survey indices for the HERAS index brings more transparency to the fisheries advice process of both NSAS and WBSS herring stocks. In addition, the R code developed for TAF can now easily be applied to other acoustic surveys.

A secondary aim for the current study was to run sensitivity tests programmatically. This was done for the following range of assumptions:

- Introduction of calibration errors
- Introduction of error in split in strata 11 and 141
- Testing of alternative WBSS/NSAS split in strata 11 and 141
- Testing of alternative strata definition
- Testing of haul allocation using objective assignment by stratum

Expectedly, error in calibration of echosounder was found to be influential for both NSAS and WBSS, though of varying importance depending on which survey component is affected. In that context, consistency in calibration gain for each participating vessel should undergo specific attention to identify and troubleshoot echosounder malfunctioning prior to surveying. Experience from an echosounder malfunctioning onboard RV Tridens II (Sakinan and Berges 2020) showed that effective ways to identify issues can consist of keeping a historical record of calibration gain and $s_{a\,corr}$ and regularly monitoring the impedance levels of each transducer quadrants.

---

[11] git@git.wur.nl:berge057/heras_index_kbwot.git
[12] git@github.com:ices-taf/2020_her.27.3a47d_acousticIndex.git
[13] git@github.com:ices-taf/2021_her.27.3a47d_acousticIndex.git

Currently, the stock splitting in strata 11 and 141 is handled externally to StoX using derived split proportions between NSAS and WBSS based on vertebrae counts. A sensitivity analysis showed the final WBSS indices to be sensitive to this quantity which has significant fluctuations since 2014. The introduction of genetics and otolith shape recognition data was tested and found to be influential, especially for WBSS, though this influence is fluctuating depending on the year of the survey. Overall, the introduction of genetic data in forthcoming HERAS surveys in strata 11 and 141 will enable one to improve the accuracy and consistency of stock splitting.

The use of the new strata was effective in separating the immature and mature aggregations and thereby improving the homogeneity in the strata for improved precision. However, the results did not demonstrate significant differences in terms of abundances and proportion mature for NSAS. This suggests that the stratification as currently used is suitable with a manual haul assignment based on expert judgment.

The use of allocation of all hauls to stratum is clearly influential, emphasizing the need for accurate haul allocation to transects. Manual assignment enables incorporation of the observations during the survey in terms of aggregation characteristics and their most likely demographic characteristics and therefore preferable to stratum-based assignment.

With automatic routines as developed in this project, there is potential for further testing of the WBSS/NSAS indices. Such work includes

- Testing the use of depth dependent TS
- Testing the use of error in biological samples
- Testing the effect of changes to biological sampling strategies
- Investigating the impact on NSAS/WBSS assessments and advice

References

Ainslie, Michael A., and James G. McColm. 1998. "A Simplified Formula for Viscous and Chemical Absorption in Sea Water." *The Journal of the Acoustical Society of America* 103 (3): 1671–72. https://doi.org/10.1121/1.421258.

Berg, Florian, S. Lusseau, Valerio Bartolino, T. Gröhsler, Cecilie Kvamme, R. D.M. Nash, and Aril Slotte. 2019. "Using Otolith Shape Analysis and Machine Learning Techniques to Discriminate between Stocks of Atlantic Herring." *ICES Annual Science Conference 2019, 9-12 September 2019, Gothenburg, Sweden*.

Berg, Florian, G. Seljestad, D. Bekkevold, C. Kvamme, G. Dahle, and K. Glover. 2021. "Stock Splitting to Estimate Individual Survey Estimates for Herring Stocks Occurring during the HERAS Survey – Genetics as New Method." In *Report of the Working Group of International Pelagic Surveys (WGIPS), 11-15 January 2021*.

Berg, Florian, Aril Slotte, Arne Johannessen, Cecilie Kvamme, Lotte Worsøe Clausen, and Richard D. M. Nash. 2017. "Comparative Biology and Population Mixing among Local, Coastal and Offshore Atlantic Herring (Clupea Harengus) in the North Sea, Skagerrak, Kattegat and Western Baltic." Edited by Daniel E. Duplisea. *PLOS ONE* 12 (10): e0187374. https://doi.org/10.1371/journal.pone.0187374.

Clausen, L. A. W., D. Bekkevold, E. M. C. Hatfield, and H. Mosegaard. 2007. "Application and Validation of Otolith Microstructure as a Stock Identification Method in Mixed Atlantic Herring (Clupea Harengus) Stocks in the North Sea and Western Baltic." *ICES Journal of Marine Science* 64 (2): 377–85. https://doi.org/10.1093/icesjms/fsl036.

Dalen, John, and Odd Nakken. 1983. "On the Application of the Echo Integration Method." ICES CM 1983/B: 19.

Demer, D. A., L. N. Andersen, C. Bassett, L. Berger, D. Chu, J. Condiotty, G. R. Cutter, et al. 2017. "2016 USA–Norway EK80 Workshop Report: Evaluation of a Wideband Echosounder for Fisheries and Marine Ecosystem Science." In *ICES Cooperative Research Report*, 336:69. http://doi.org/10.17895/ices.pub.2318.

Demer, D. A., L. Berger, M. Bernasconi, E. Bethke, K. Boswell, D. Chu, R. Domokos, et al. 2015. "Calibration of Acoustic Instruments." *ICES Cooperative Research Report* 326: 133.

Foote, K. G., H. P. Knudsen, G. Vestnes, D. N. MacLennan, and E. J. Simmonds. 1987. *Calibration of Acoustic Instruments for Fish Density Estimation : A Practical Guide*. Copenhagen Denmark: International Council for the Exploration of the Sea.

Francois, R. E., and G. R. Garrison. 1982. "Sound Absorption Based on Ocean Measurements. Part II: Boric Acid Contribution and Equation for Total Absorption." *The Journal of the Acoustical Society of America* 72 (6): 1879. https://doi.org/10.1121/1.388673.

Gröger, J., and T. Gröhsler. 2001. "Comparative Analysis of Alternative Statistical Models for Differentation of Herring Stocks Based on Meristic Characters." *Journal of Applied Ichthyology* 17 (5): 207–19. https://doi.org/10.1046/j.1439-0426.2001.00254.x.

ICES. 1994. "Report of the Herring Assessment Working Group for the Area South of 62°N." *ICES CM 1994/Assess: 13.*

———. 2015a. "Manual for International Pelagic Surveys (IPS)."

———. 2015b. "Report of the Benchmark Workshop on West of Scotland Herring (WKWEST)." *ICES CM 2015\ACOM:34*.

———. 2015c. "Report of the Workshop on Evaluating Current National Acoustic Abundance Estimation Methods for HERAS Surveys (WKEVAL)." *24-28 August 2015, ICES Headquarters, Copenhagen, Denmark. ICES CM 2015/SSGIEOM:16*.

———. 2017. "Interim Report of the Working Group on Fisheries, Acoustics, Science and Technology." In *WGFAST 2017*, 40. 4-7 April 2017. Nelson, New Zealand, ICES CM 2017/SSGIEOM:12.

———. 2018a. "Interim Report of the Working Group on Fisheries Acoustics, Science and Technology (WGFAST)."

———. 2018b. "Report of the Benchmark Workshop on Pelagic Stocks (WKPELA 2018)." In *ICES HQ, Copenhagen, Denmark. ICES CM 2018/ACOM:32*, 313.

———. 2020a. "Herring (Clupea Harengus) in Subarea 4 and Divisions 3.a and 7.d, Autumn Spawners." In *Herring Assessment Working Groupfor the Area South of 62°N (HAWG)*.

———. 2020b. "Herring (Clupea Harengus) in Subarea 4 and Divisions 3.a and 7.d, Autumn Spawners (North Sea, Skagerrak and Kattegat, Eastern English Channel)." *ICES Advice on Fishing Opportunities, Catch, and EffortBaltic Sea and Greater North Sea Ecoregions*.

———. 2020c. "Herring (Clupea Harengus) in Subdivisions 20–24, Spring Spawners (Skagerrak, Kattegat, and Western Baltic)." *ICES Advice on Fishing Opportunities, Catch, and EffortBaltic Sea and Greater North Sea Ecoregions*.

———. 2020d. "Herring Assessment Working Groupfor the Area South of 62°N (HAWG)."

Johnsen, Espen, Atle Totland, Åsmund Skålevik, Arne Johannes Holmin, Gjert Endre Dingsør, Edvin Fuglebakk, and Nils Olav Handegard. 2019. "StoX: An Open Source Software for Marine Survey Analyses." *Methods in Ecology and Evolution* 10 (9): 1523–28. https://doi.org/10.1111/2041-210x.13250.

Kell, L. T., I. Mosqueira, P. Grosjean, J-M. Fromentin, D. Garcia, R. Hillary, E. Jardim, et al. 2007. "FLR: An Open-Source Framework for the Evaluation and Development of Management Strategies." *ICES Journal of Marine Science* 64 (4): 640–46. https://doi.org/10.1093/icesjms/fsm012.

Lunde, Per, and Rolf Korneliussen. 2016. "Power-Budget Equations and Calibration Factors for Fish Abundance Estimation Using Scientific Echo Sounder and Sonar Systems." *Journal of Marine Science and Engineering* 4 (3): 43. https://doi.org/10.3390/jmse4030043.

Lunde, Per, O. Audun Pedersen, Rolf J. Korneliussen, Frank E. Tichy, and Håvard Nes. 2013. *Power-Budget and Echo-Integrator Equations for Fish Abundance Estimation*. *Fisken Og Havet No. 10/2013*. Institute of Marine Research, Bergen, Norway. https://brage.bibsys.no/xmlui/handle/11250/113867.

Lusseau, S., Sacha Fassler, Cecilie Kvamme, N. Cormac, Matthias Schaber, K. Staehr, E. Amstrong, C. O'Donnell, and E. Johnsen. 2016. "Sensitivity Analysis HERAS 2015. Effect of Changing the Estimation Tool to StoX for the 2015 HERAS." In *Report of the Working Group of International Pelagic Surveys (WGIPS), 18-22 January 2016, Dublin, Ireland. ICES CM 2016/SSGIEOM:05.*

Macaulay, Gavin J, Ben Scoulding, Egil Ona, and Sascha M M Fässler. 2018. "Comparisons of Echo-Integration Performance from Two Multiplexed Echosounders." *ICES Journal of Marine Science* 75 (6): 2276–85. https://doi.org/10.1093/icesjms/fsy111.

Maclennan, D, Paul G. Fernandes, and John Dalen. 2002. "A Consistent Approach to Definitions and Symbols in Fisheries Acoustics." *ICES Journal of Marine Science* 59 (2): 365–69. https://doi.org/10.1006/jmsc.2001.1158.

Mehl, Sigbjørn, Espen Johnsen, Åsmund Skålevik, and Asgeir Aglen. 2018. "Estimation of Acoustic Indices with CVs for Northeast Arctic Saithe in the Norwegian Coastal Survey 2003-2017 Applying the Sea2Data StoX Software." *19*. Fisken og Havet;1-2018: Institute of Marine research, Bergen, Norway. https://brage.bibsys.no/xmlui/handle/11250/2562514.

Millar, C., and A. Magnusson. 2018. "Stock Assessment Data Flow: Status Update." *WGCHAIRS, 24 January 2018*. https://taf.ices.dk/pdf/2018-01-24-chairs.pdf.

Nielsen, Anders, and Casper W. Berg. 2014. "Estimation of Time-Varying Selectivity in Stock Assessments Using State-Space Models." *Fisheries Research* 158 (October): 96–101. https://doi.org/10.1016/j.fishres.2014.01.014.

Ona, E. 2003. "An Expanded Target-Strength Relationship for Herring." *ICES Journal of Marine Science* 60 (3): 493–99. https://doi.org/10.1016/S1054-3139(03)00031-6.

Sakinan, Serdar, and B.J.P. Berges. 2020. "Investigation of the Use of the EK80 CW during Acoustic Surveys on Board Tridens II." *IJmuiden: Wageningen Marine Research (Wageningen Marine Research Rapport)*, 41.

Sakinan, Serdar, Dick de Haan, Dirk Burggraaf, and Sacha Fassler. 2018. "Investigation of Echosounder Finger Prints of Dutch Pelagic Freezer Trawlers (SEAT II) Evaluation of the SEAT II Joint-Industry Project : Evaluation of the SEAT II Joint-Industry Project." https://library.wur.nl/WebQuery/wurpubs/541398.

Simmonds, E John, and D N MacLennan. 2005. *Fisheries Acoustics : Theory and Practice*. *Fish and Aquatic Resources Series ;* 2nd ed. Oxford ; Ames, Iowa: Blackwell Science.

Quality assurance

CVO is certified to ISO 9001:2015 (certificate number: 268632-2018-AQ-NLD-RvA). This certificate is valid until December 15th, 2021. The certification was issued by DNV GL Business Assurance B.V

Justification

CVO Report: 21.007
Project number: 4313100145

The quality of this report has been peer reviewed by a colleague scientist and the head of CVO.

Approved by: C. van Damme
Colleague scientist

Signature: 

Date: 17th of February 2021

Approved by: Ing. S.W. Verver
Head Centre for Fisheries Research

Signature: 

Date: 17th of February 2021