



---

# Improvement of greater argentine stock assessment

Author(s): S.Sakinan, E. Beukhof and N. Hintzen

Wageningen University &  
Research report C098/21

---

# Improvement of greater argentine stock assessment

Author(s): S.Sakinan, E. Beukhof and N. Hintzen

Publication date: 1 December 2021

Wageningen Marine Research  
IJmuiden, December 2021r

---

CONFIDENTIAL yes / no

Wageningen Marine Research report C098/21

---

*S.Sakinan, E. Beukhof and N. Hintzen 2020. Improvement of greater argentine stock assessment.*  
Wageningen, Wageningen Marine Research (University & Research centre), Wageningen Marine  
Research report *reportnumber C098/21ber*. 49 pp.

Keywords: Greater argentine, stock assessment, age length key

Client: Martin Pastoors, PFA

This report can be downloaded for free from <https://doi.org/10.18174/558216>  
Wageningen Marine Research provides no printed copies of reports

Wageningen Marine Research is ISO 9001:2008 certified.

© Wageningen Marine Research

Wageningen Marine Research, instituut  
binnen de rechtspersoon Stichting  
Wageningen Research, hierbij  
vertegenwoordigd door  
Drs.ir. M.T. van Manen, directeur  
bedrijfsvoering

KvK nr. 09098104,  
WMR BTW nr. NL 8113.83.696.B16.  
Code BIC/SWIFT address: RABONL2U  
IBAN code: NL 73 RABO 0373599285

Wageningen Marine Research aanvaardt geen aansprakelijkheid voor  
gevolgschade, noch voor schade welke voortvloeit uit toepassingen van de  
resultaten van werkzaamheden of andere gegevens verkregen van Wageningen  
Marine Research. Opdrachtgever vrijwaart Wageningen Marine Research van  
aanspraken van derden in verband met deze toepassing.  
Alle rechten voorbehouden. Niets uit deze uitgave mag weergegeven en/of  
gepubliceerd worden, gefotokopieerd of op enige andere manier gebruikt worden  
zonder schriftelijke toestemming van de uitgever of auteur.

A\_4\_3\_1 V31 (2021)

---

# Contents

<b>Summary</b>	<b>4</b>
<b>1 Introduction</b>	<b>6</b>
1.1 Aim of the project	6
<b>2 Materials and Methods</b>	<b>7</b>
2.1 Storage of PFA data in WMR databases	7
2.2 WMR Market sampling	7
2.2.1 Sample processing	7
2.2.2 Age reading	7
2.3 PFA self-sampling	8
2.3.1 Consistency of the length measurements between WMR market sampling and PFA self-sampling	8
2.3.2 Combined key generation	9
2.3.3 Calculation of catch at age	11
2.4 Assessments using different catch at age matrices	11
2.5 CPUE standardization	11
<b>3 Results</b>	<b>12</b>
3.1 Storage of PFA data in WMR databases	12
3.2 Market sampling vs PFA self-sampling	12
3.2.1 Comparison of the samples collected at a close distance in area and time (one-to-one)	12
3.2.2 Market sampling vs PFA Self-sampling: Group comparison	17
3.3 WMR Market sampling: Overview of age and length and maturity information	19
3.3.1 Age readings and growth patterns in WMR market samples	20
3.3.2 Changes in the age, length and gender structure	21
3.4 PFA self-sampling, an exercise for age estimation	22
3.4.1 Evaluation of the combined key method	23
3.5 Assessments using different catch at age matrices	25
3.5.1 Evaluation of the combined key method	28
3.6 CPUE standardization	33
<b>4 Conclusions</b>	<b>34</b>
<b>5 Quality Assurance</b>	<b>35</b>
<b>Appendix I</b>	<b>36</b>
<b>Appendix II</b>	<b>43</b>
<b>References</b>	<b>44</b>
<b>Justification</b>	<b>45</b>

---

# Summary

An ongoing commercial market sampling for the Dutch greater argentine fishery is carried out by WMR in collaboration with the pelagic fishing fleet of the Netherlands. Start of this sampling programme goes back to year 1990. The aim of this programme is to achieve a haul-level sampling each year, stratified in area and time within the targeted ICES division. Later, market sampling data is pooled with respect to the quarter and the ICES divisions and the inferred age structure to is allocated to the catch, a classical ALK is generated from the collected age-length data which describes the probability of age given length (Fridriksson 1934).

The sample sizes in this long time series can be relatively low in some years and the coverage can be relatively limited in space and time. The additional data sets, collected outside of this program, such as length composition data collected by self-sampling of PFA, maybe useful for filling some of these gaps and improve the level of accuracy of the estimations.

However, this approach requires

- A detailed understanding of the characteristics of the data set and the patterns in the market sampling time series.
- A careful comparison between the datasets while assessing the quality and consistency of the measurements
- Adoption of an unbiased method to incorporate age structure information from one dataset to the other.

In the first section of the report, the characteristics of the two datasets are explained such as the sampling strategy, sample processing, sample size as well as the demographic changes over the years. Secondly the results of the comparison between the two datasets are presented. The initial aim was to focus on the samples from the two different sampling programs overlapping in space and time. However the level of overlap was not as close as expected. Nevertheless the closest samples in space and time selected and compared. It was found that the resemblance between the datasets did not show a pattern that can be explained with the distance or time between the samples. This is most likely due to stochasticity of the length distributions.

A secondary problem that was tackled with was the lack of age information in the PFA data sets. Although length distributions can be an indirect indicator of the age structure a cohort estimation can not be acquired from the length distribution due to slow growth and long life span of the species.

One solution would be inferring such missing ages using the available age/length information. Currently, classical age/length keys are commonly used to extrapolate otolith readings to length composition. Unfortunately the classical ALK obtained with data from one year/area cannot be applied to another year/area, because the underlying structure of the population may change due to variation in recruitment, survival or growth. An alternative is the use of an inverse key, which describes the probability of length given age, and has the benefit that this probability is not affected by variation in recruitment and survival. An inverse key can be applied to other years or areas, but under the assumption that size at age does not vary between years or areas. Another alternative is the combined forward-inverse key (hereafter called 'combined key') which combines the forward and inverse key into one approach (Hoenig et al. 2002). It uses Bayes' rule to implement the forward and inverse key into one maximum likelihood framework through the probability of length at age. In this work, we applied the combined key using an R package implementation ( 'ALKr') by Murta et al. (2016). For this exercise, we used market age/length samples to build a combined key for greater greater argentine and compared it with the results obtained by the classical ALK. We further used the constructed combined key to estimate the age composition for the catch of samples by the PFA based on their length measurements.

The resulting combined key is a promising solution for estimating the age composition in the PFA self-sampling data set, from which the length composition could be used as an input. However before that

---

it is necessary to ensure that the this length composition time series is free from systematic biases with reasonable accuracy. Therefore a comprehensive check was carried out for the PFA length frequency data while using the market sampling dataset as a reference. The results are presented in this report.

Lastly stock assessment tests were performed for the greater argentine by with and without incorporation of the PFA data, and the results are discussed.

---

# 1 Introduction

Greater argentine is one of the commercially important deep semi-pelagic fish species for the Pelagic Dutch fisheries with an annual landing range of approximately 2000 – 4000 tons taking place in ICES divisions 6a and 5b. The assessment of the stocks in this area is however, categorised as a data-poor stock due to lack of readily available time series of age samples from catches and surveys (DLS method 3.2). There is, on the other hand, a wide range of data that can be made available for the improvement of the stock assessment for this species in this region. With such considerations, WGDEEP has been exploring an age-based assessment model for this stock. Examples of potential data are length-frequency measurements from the PFA self-sampling programme of the Pelagic Freezer-trawler Association (PFA), Market samples from Wageningen Marine Research (WMR), Market samples from Germany and survey data collected by countries such as the Faroe Islands and Scotland. Recent studies show that well-assessed stocks are healthier and more sustainable than poorly reviewed stocks and that their management is able to respond to the changing environmental circumstances (Costello et al. 2012, Hilborn et al. 2015). Most advanced methods for assessing stocks are based on information from catch and survey time series, in particular the estimates of the age and time series of commercial catches. The data collected by commercial market sampling programmes are usually the primary source of information for monitoring the performance of a fishery and obtaining necessary inputs for age-based methods. These samples consist of fish collected from the catches during fishing trips from a number of different vessels that are representative for a specific fishery of a specific stock. From these fish samples, age (usually based on the otoliths), length, individual weight, maturity and gender are determined. Such data collected each season contribute to a time series that provides information about the dynamics of the stock, such as mortality and recruitment. When age-length time series are not available, determining a robust management strategy becomes difficult.

This report focusses mainly on examining the historical data collected by the Dutch market sampling programme and generating an overview of the available length, round weight, age and maturity information in order to evaluate the status of the stock and explore the potential use for a data-rich stock assessment. In addition to the market samples, a large set of data is available from the PFA self-sampling programme of the PFA from 2015 onwards. This dataset consists of length frequencies associated with detailed positional and environmental information, and with unique haul numbers. In this work, a comparison study has been carried out between the length distributions of the market samples and PFA samples in order to screen the PFA self-sampling data for any systematic bias to ensure its reliability when incorporating this data into stock assessment. Furthermore, a potential application of the PFA self-sampling data into stock assessment was explored, namely testing an age-composition estimation algorithm that can predict the age composition of the catches from the length distribution data of the PFA and the age-length information of the market samples.

## 1.1 Aim of the project

In order to evaluate the data availability for the greater argentine from the market sampling and PFA self-sampling, the following main objectives and questions were defined at the start of the project:

1. An inventory of how the collected data, in accordance with data standards and confidentiality, can be stored within the current WMR database structure
2. Processing of length, age and condition data from the WMR market sampling and sampling of the pelagic vessels into fleet and stock level

---

## 2 Materials and Methods

### 2.1 Storage of PFA data in WMR databases

A session was organised between WMR and PFA to discuss the potential way forward to store PFA data in WMR databases for use in assessments. A default template (Appendix I) was used to guide the discussion.

### 2.2 WMR Market sampling

WMR, in collaboration with the pelagic fishing fleet of the Netherlands, has been carrying out a market sampling programme on the commercial greater argentine catches. The start of the sampling programme goes back to 1990. Early samples in the dataset come from a government-subsidized fishing experiment to determine the distribution of greater argentine around the Porcupine Bay and British Isles (Heessen and Kuitert 1990). In the following years, data collection is carried on with the same sampling scheme, however, at a smaller scale in terms of sample size. In this on-board sampling programme, the samples are taken directly from the individual hauls before the catch is processed or mixed with previous catches. Sampling and logging on board is carried out by a specialized person from the crew, and samples are labelled with vessel logs, such as position, time, date, etc. This haul-level sampling allows well-stratified sampling within the targeted ICES division during the fishing period.

The fishing vessels involved in the programme are committed to collect one random sample from a haul in each ICES division (sub-area) in each week of a fishing trip. If the same vessel remains in the same division for more than a week, it collects one sample each week in that division. If the vessel is moving between divisions, regardless of the time difference, one additional sample is taken.

The samples are taken from the hauls in which greater argentine dominates the catch. Total sample weight can be between 20 -25 kg. The sample is taken randomly without sorting for species or size and is immediately frozen without any measurement. A detailed logbook is kept with the geographical location (latitude and longitude), time, date, depth and name of the vessel corresponding to each sample. At the end of each fishing trip, the samples are transported to WMR for processing. Commonly 2-3 vessels contribute to these samples while it could rarely be fewer or more up to five. The total weight of samples collected by all ships per year varies from approximately 200-400 kgs.

#### 2.2.1 Sample processing

In the fish laboratories in the WMR, after the frozen samples have been thawed, the length frequency measurements are performed for each sample with a precision of half a cm. Then a subset of 25 specimens is taken for biological parameters based on the obtained distribution of the length classes with an interval of 1 cm. Subsequently, in the second phase of processing, these 25 specimens are measured in detail for a greater precision length (1 mm), individual weight, maturity, gender and age.

#### 2.2.2 Age reading

Throughout the time series, 3 different experts were involved in age determination (from otoliths); first from 1990 to 1994, second, 1995 to 2014, and third 2015 to present. The quality of the age measurements are continually checked according to the WMR routines. In addition, consistency between the age readers are verified before any transition.



---

## 2.3 PFA self-sampling

The PFA self-sampling dataset is another potential source of information to be used for stock assessment (for ICES division 6a). The Dutch pelagic freezer-trawler fleet has traditionally been collecting data, including length, weight, relative fat content and gonad development of individual fish as part of their commercial quality assessment. Their archives include detailed metadata such as time, depth, sea state, wind speed, speed, gear opening etc. Although the registered information goes back more than ten years in the skipper's logs, the data are only available in a standard digital format from 2015 onwards, organised and maintained under the name of self-sampling. Normally, the WMR market samples and the PFA samples come from the same group of pelagic ships. But PFA self-sampling data covers a much wider spatial area and time range, which may have added value to the stock assessment process.

### 2.3.1 Consistency of the length measurements between WMR market sampling and PFA self-sampling

Combining the PFA data with the time series of market samples was one of the objectives of this study to improve the reliability of current knowledge about greater argentine. But, before pooling samples from different sources to construct a new time series, it is necessary to ensure that each set has a common level of accuracy and is free from systematic bias.

In the WMR market sampling programme, data acquisition for a specific species is handled mostly by the same expert with an unchanged method. Each new dataset undergoes a standardized test by a dedicated data analyst before being stored in the database.

In the PFA self-sampling programme of the PFA, the measurements come from multiple vessels and are carried out separately by different crew members. Even though the crew member in charge for the sampling and the measurements on board is a trained and experienced person, a detailed verification of their measurement accuracy has never been performed.

Because of structural differences in the data acquisition and storage procedures, combining the WMR market samples and PFA samples is not a simple task. Therefore, a detailed inspection of both sets has been carried out, including:

- Comparison of samples collected at a close distance in space and time (one-to-one)
- Comparison of the pooled data per vessels per year (taking the within variability into account)

As for the one-to-one comparison, each length frequency distribution from the PFA self-sampling dataset was compared against a market sample that falls within less than 15 days and within 40 nmi in distance. Even though all vessels involved in the market sampling programme are also involved in the PFA self-sampling, not always the same vessels or hauls are sampled by both programmes during a fishing season. Especially in the case of greater argentine, unfortunately, none of the samples in the PFA self-sampling dataset had a corresponding sample in market samples at the haul level.

Furthermore, except for 2018, none of the vessels that provided samples to WMR market sampling programme also provided data to PFA self-sampling dataset (Table 1). Therefore, an ideal comparison and a confident verification was not possible in this way for all years. Nevertheless, the two of vessels involved in market sampling in 2018 were also involved in the PFA self-sampling, thereby giving an opportunity to make illustrative comparisons using the closest-proximity samples. In addition, there are several other samples, despite being from different vessels, that were close to each other in space and time, making them interesting for comparisons.

As part of this project, we explored the potential of using the age data available from the market sampling for estimating age composition for the datasets where only length frequencies are available. For this purpose, we looked at both probability of length at age and probability of age at length. While there are different methods based on such age-length relationships, for this exercise we will focus on 3 methods as described below:

- 
1. Forward key: Probability of age in each length class
  2. Inverse key: Probability of length at each given age, assuming constant growth rate in space and time
  3. Combined key: Combination of the forward and inverse key

The forward or classic key, which describes the probability of age given length, is the simplest method (Fridriksson, 1934) and likely the most commonly used ALK (age-length key) in fisheries science. However it has certain limitations, for example, data from one year cannot be applied to another year, because the underlying structure of the population may change due to variation in recruitment, survival or growth.

Inverse keys describe the probability of length given age, and have the benefit that this probability is not affected by variation in recruitment and survival. An inverse key can be applied to other years or areas, but under the assumption that size at age does not vary between years or areas. Changes in size at age may be caused by changes in growth rate or in the selectivity of the fishery. However, as long as this assumption holds, the inverse key can thus be applied to other years or areas with missing or incomplete age data (Murta *et al.*, 2016; Ailloud and Hoenig, 2019).

The combined key uses Bayes' rule to implement the forward and inverse key into one maximum likelihood framework through the probability of length at age (see (Ailloud and Hoenig, 2019) and (Murta *et al.*, 2016) for more details and mathematical background). In years without age data and only length measurements, it uses the length at age information from other years to estimate the probability of length at age. In years with age data it uses the principles of the forward key (i.e. age at length), but the estimates are penalized by the algorithm when they deviate too much from the length at age information known from other years. Note that the combined key also relies on the assumption of no-spatio-temporal variation in size at age

The method applies an iterative process by first giving initial values to the parameters, then completing the dataset by calculating the expected values for the missing parameters using the model with the initial values for the parameters. Finally, using the complete dataset, the algorithm calculates the maximum-likelihood estimates of the parameters. With the new parameters' estimates, a new iteration starts by recalculating the expected values of the missing parameters until the model convergences.

### 2.3.2 Combined key generation

Once a combined key is constructed, it can be applied to a given length frequency set regardless of its contribution to the initial set from which the key is generated. Based on the age-length data from the market samples we generated a combined key, followed by estimating the proportional number of fish at age in the market samples. These results were tested against the conventional forward key that was applied to the same market samples. Finally, we estimated the age composition of PFA samples based on length measurements using the constructed combined key. The construction of ALKs were performed with the ALKr package in R (Murta *et al.*, 2016).

Prior to the analysis, only market sampling stations were selected that contained as much or more length classes than age classes, as this is required for the optimization algorithm behind the combined key to work (Ailloud and Hoenig, 2019). Sampling stations outside the stock area were also removed to minimize the risk of not meeting the assumption of no varying size at age.

Generation of a combined key required inclusion of wide range of samples from different years, point locations and age readers. In order to account for the bias potentially coming from combining these different sets, a number of scenarios were tested. The tests were carried out under two broad assumptions; 1) There is no spatial variability in stock structure within the area for a given year 2)

Each station is an independent sample for the given year capturing different spatial components. Each scenario was tested twice: once for using the dataset where all stations were aggregated per year (based on assumption 1) and once for when samples were not aggregated by year (based on assumption 2). The scenarios are:

- Control scenario: all data are used as they are (either aggregated by year or not).
- Due to the few years with data that include old ages, age 20 is considered as the maximum age and used as plus group in age-length data (i.e. ages greater than 20 were treated as if they were 20).
- From the data it appeared that between 2004-2007 age 15 was treated as a plus group by the otolith expert. In this scenario, this plus group is removed from the age-length data used to construct the combined key in the years 2004-2007.
- Similar to scenario c, but the years 2004-2007 are entirely removed from the dataset used to construct the combined key (due to presence of plus group of 15).
- A comparison between the theoretical Von Bertalanffy growth curve (table 4) and the actual age reading was made and some years were found to strongly deviating from the expected growth (years: 1996, 1998, 2000, 2001-2003, 2016). In this scenario, these years are removed from the dataset used to construct the combined key.

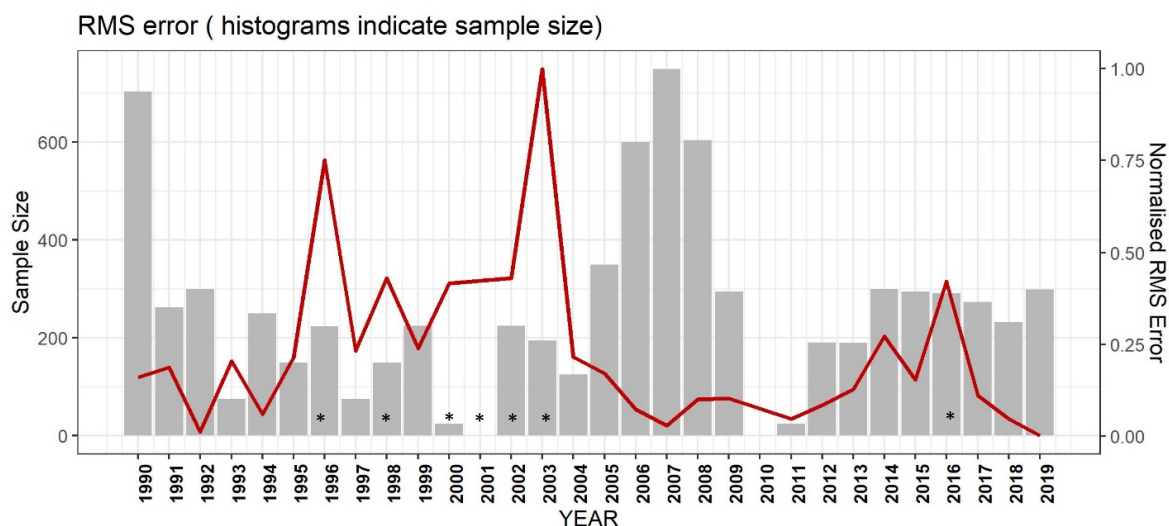


Figure 1 Red curve shows the deviation of the length at age relative to the theoretical growth curve ( Von Bertalanffy). Higher error rates were considered due to combination of low sample size and erratic age reading. Grey bars show the sample size (N). The years that are excluded are indicated with asterix.

For each of these scenarios the length frequency data from the year (or station) of interest were then used by the combined key (that was constructed based on the age-length and length frequency data of all years and stations) to estimate the age composition for that year (or station). These results were then compared against the results of the forward key. The comparisons of the predicted age compositions by the two keys were made by calculating the mean squared error per year and the sum of these errors for each scenario. The sum of mean squared errors were used for scoring the scenarios, and the scenario with the lowest score was then selected as the best "combined key" (and later on applied to the length frequency data of the PFA).

---

### 2.3.3 Calculation of catch at age

The ultimate output needed as input for the stock assessment is the number at age for the catch in a given season. The standard routines used to calculate these numbers before submission to the ICES database (InterCatch Kjems-Nielsen et al., 2006) use the forward key and apply a weighting to the samples based on the sample sizes (number of fish in the sample). In order to compare the effects of using different age-length key methods and, furthermore, inclusion of PFA self-sampling data to the process, these numbers were recalculated. In addition, one potential concern was that weighting the samples by number of individuals may have been misleading when there are samples containing too many small-sized individuals in few stations, even though the weight (i.e. biomass) of the samples are similar. To see the effect of this, an additional raising was carried out by applying a weighting based on the sample weight rather than the number of individuals in the samples. Eventually, raising (i.e. extrapolating the sub-samples to the corresponding catch) was carried out in five different ways as;

- 1) Identical to the InterCatch routines
- 2) Applying weighting based on the sample weight
- 3) Applying forward key by equal weighting
- 4) Applying the combined key
- 5) Applying the combined key, incorporating the PFA dataset

## 2.4 Assessments using different catch at age matrices

The stock coordinator for greater argentine combined all catch at age data, including data submitted by WMR, and raised it accordingly to one catch at age matrix. WMR used FLSAM to assess the stock using the new catch at age matrix and standardized indices. An exercise of data validation as well as assessment model configuration was undertaken using a standardized approach in which internal consistency of catch and survey data was investigated, plus-group of the stock was considered and optimal model parameter bindings were evaluated.

Furthermore, stock assessments were run under three different conditions: applying the standard forward age-length key methodology, the combined key and combined key + PFA samples methodology.

## 2.5 CPUE standardization

CPUE standardisation was undertaken by Floor Quirijns with statistical support from WMR. No analyses were undertaken by WMR but extensive comments and suggestions were provided that led to the successful completion of CPUE standardization.

---

## 3 Results

### 3.1 Storage of PFA data in WMR databases

The discussion focused on ownership of the data, use of the data for other purposes, storage location, quality control of the data and availability of data after the project (see Appendix I). In this project, no data was stored in WMR databases but the discussion on data storage did trigger a number of questions that PFA had to resolve internally. No results on this aspect can be described here.

### 3.2 Market sampling vs PFA self-sampling

In this section the results of the comparisons between the length distributions of the market sampling programme and PFA self-sampling programme are presented. The comparisons include one-to-one tests at sample level and group comparisons at vessel and week level.

#### 3.2.1 Comparison of the samples collected at a close distance in area and time (one-to-one)

The pair comparison results showed considerable variation in the differences between the length distributions of the market and PFA samples that fell within 2 weeks/40 nmi spatio-temporal range. In several cases, the length distributions between the PFA and market samples were comparable, whereas in other cases differences were considerably large. Figure 2 shows an example of the one-to-one comparison from the year 2018, for week 16 where the market sample has a size of 70 and the PFA self-sampling a size of 118. In this example, the median length of the samples differs 1 cm from each other and the distribution is relatively similar. The samples were taken relatively close to each other in space and time (22.8 nmi and 1 day respectively) and come from the same vessel. Figure 3 is an example for which the compared vessels are not the same, and the differences between samples are 13 days and 33.8 nmi, with only 0.4 cm difference in median length. The length distribution in the market sample ( $n = 84$ ) is characterized by a bimodal distribution, while the PFA sample ( $n = 64$ ) is characterized by a relatively normal distribution with a single peak. Figure 4 is another example where the vessels and sampling date are the same, while there is 15 nmi distance between the stations. The median length of the self-sample (33 cm) is smaller than that of the market sample (35 cm).

Summary of the one-to-one comparisons are presented on Figure 6 and Figure 7. On these figures, if the differences in the median lengths (the y axis) were controlled by the distance between the hauls (x axis), a change would have been expected with changing distance. Otherwise, if the controlling parameter was the time difference between the hauls, then a color gradient would have been expected between dark blue and yellow, along the direction of the y axis. If the sample size would account for the differences, then symbol sizes would indicate a pattern. However, all of these factors, regardless of being from the same vessel or not, seems to be randomly distributed without a clear pattern in neither direction.

Although the differences in length distributions and median length in the examples above could still be partially attributed to the small differences in distance and time of sampling, they could also can be due to random variability, low sample size or a measurement bias in one of the data sets.

Furthermore, even if the compared samples are close to each other, they were not from the same hauls. If similar sized fish are aggregating together in schools and hauls consist of catches from schools with differently sized fish, the resulting differences in composition could be a result of this as well. Consequently, neither geographical nor time difference can explain the differences (see Figure 6 for 2018 and Figure 7 for 2019).

Table 1 Summary of the pairwise comparison results. For this comparison market sampling and PFA self-sampling data are compared against each other based on distance and date. Only the pairs that have less than 15 days and 40 nmi distance were used. Second column shows how many of the pairs met these criteria. Third column shows number of pairs that have the same median length and fourth column shows the number of pairs for which the median length difference is equal or less than 2 cm. Last column shows how many of these pairs are from the same vessel.

YEAR	Number of pairs within range	Identical medians	Median length difference $\leq 2$ cm	Same Vessel paired
2015	185	27	134	0
2016	38	5	27	0
2017	52	15	49	0
2018	49	10	27	28
2019	51	17	47	0

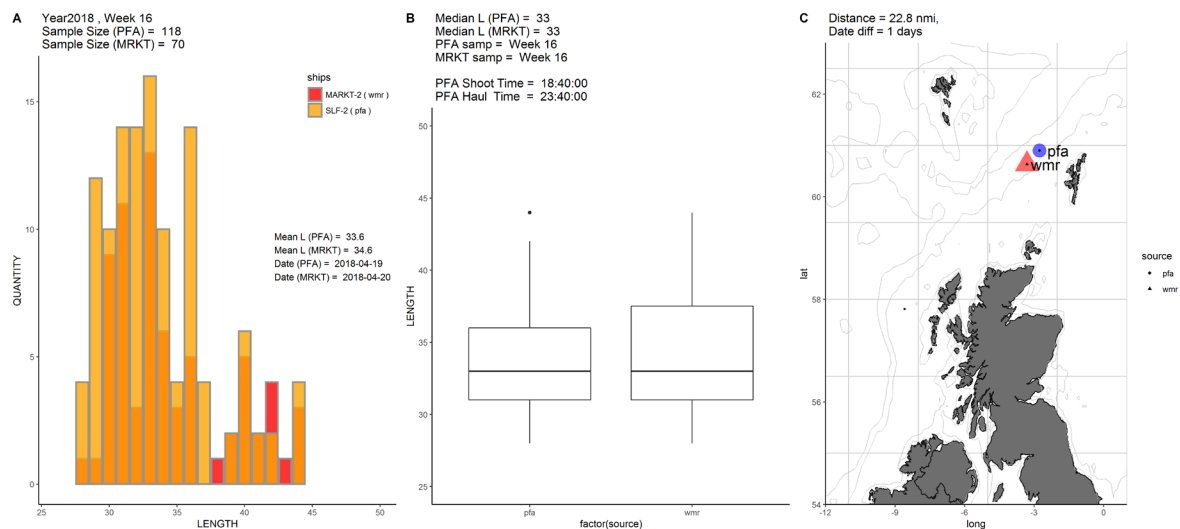
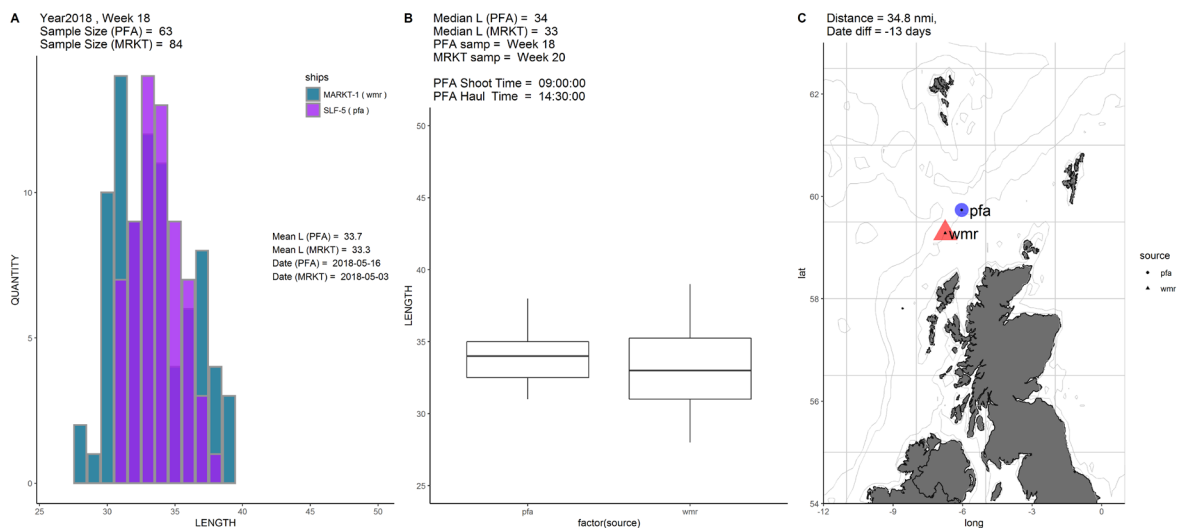


Figure 2 An example of a pairwise comparison between a market sample and a self-sample. The histogram on the left panel (A) show the market sample distribution as red, PFA samples as yellow and orange is where they overlap. The middle panel (B) shows a box plot comparison where spread of the length data is shown as quantiles. The median length of these samples are identical and there is no statistical differences between these distributions. The right panel (C) shows the location of these samples. The distance between these samples are 22.8 nmi and date difference is only one day.



*Figure 3 Another example of a pairwise comparison between a market sample and a self-sample. In this case data from two different vessels are compared. The histogram on the left panel (A) show the market sample distribution as greenish blue, and PFA samples as light-purple and dark-purple is where they overlap. The median of these samples are very similar (panel B), the distance between the vessels is 34.8 nmi in space and 13 days in date of sampling (panel C). There is no statistically meaningful difference between these length distributions.*

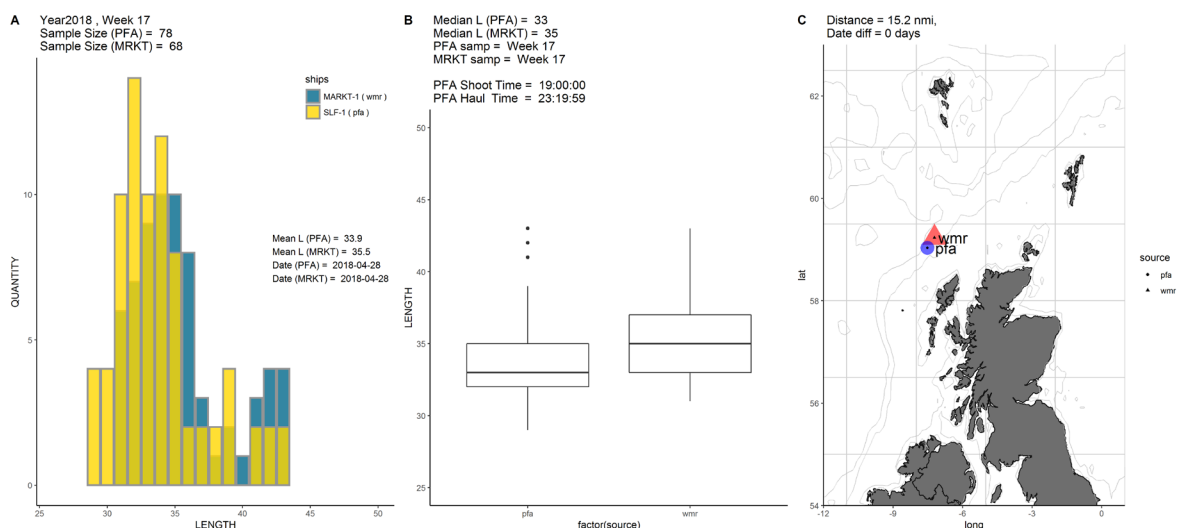


Figure 4 Another example of a pairwise comparison between a market sample and self-sample. The samples of both sampling programmes are from the same vessel. The histogram on the left panel (A) show the market sample distribution as dark-blue, PFA samples as yellow and dark-yellow where they overlap. The median of these samples are different by 2 cm (market sample is greater; panel B), the distance between the vessels is 15 nmi in space, while they are from the exact same date (panel C). There is no statistically meaningful difference between these length distributions.



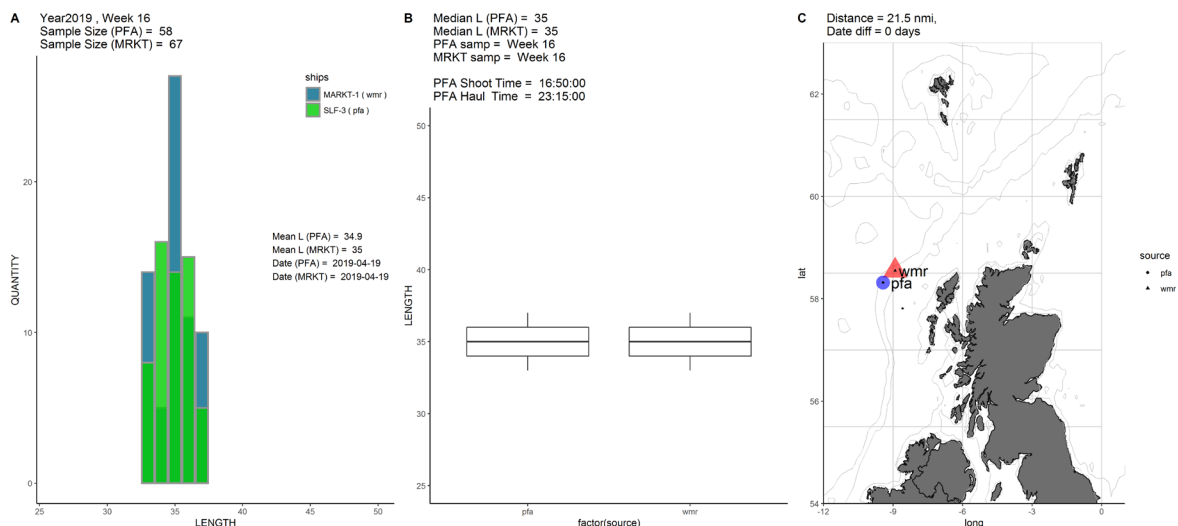


Figure 5 An example for the 2019 pairwise comparisons. In this year none of the pair samples are coming from the exact same vessels. In this example, the sampling location is close (21.5 nmi) and both samples are from the same date. The median lengths are identical and distributions are very similar.

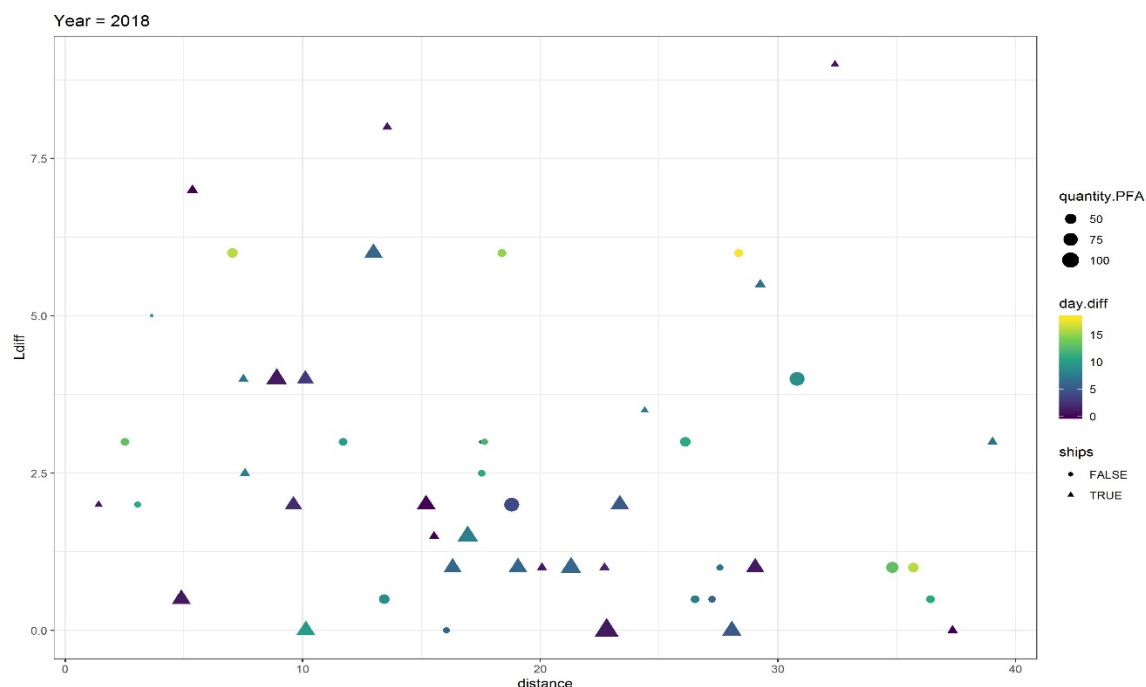


Figure 6 Summary of the pairwise vessel comparison for 2018. X axis show the geographic distance in nautical miles and Y axis show the differences in the median lengths (in cm). The triangular shapes indicate that the compared samples are from the same vessel while circular shapes are from different vessels. The blue-green-yellow scale indicates differences in days of sampling between market and PFA self-sampling, with blue no/little difference and yellow a large difference. The size of the shapes indicates the sample size that is coming from the PFA self-sampling data.

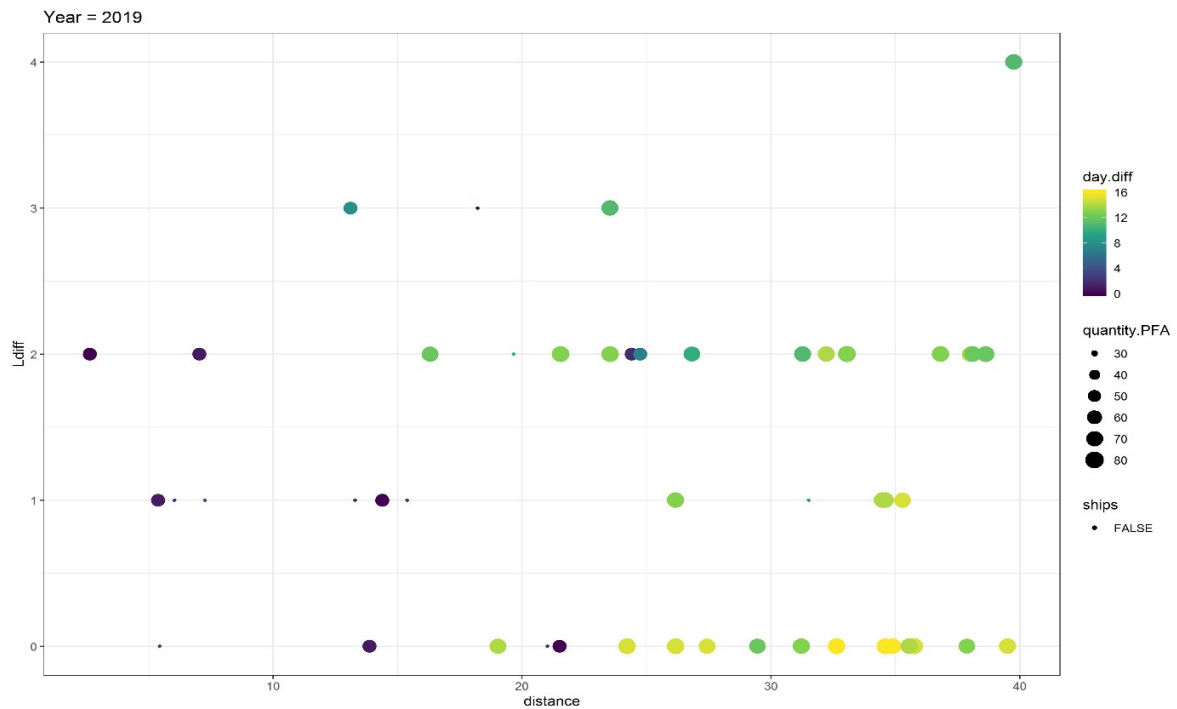


Figure 7 Summary of the pairwise vessel comparison for the 2019. X axis show the geographic distance in nautical miles and Y axis show the differences in the median lengths (in cm). All comparisons were made with the same vessels. The blue-green-yellow scale indicates differences in days of sampling between market and PFA self-sampling, with blue no/little difference and yellow a large difference. The size of the shapes indicates the sample size that is coming from the PFA self-sampling data.

### 3.2.2 Market sampling vs PFA Self-sampling: Group comparison

In order to understand to what extent the random variabilities underlie the differences, a one-way ANOVA test was performed among the groups (e.g. samples pooled by vessels). This test takes into account both within and between variability of the groups and the results can be used to tell whether the differences between groups are significant or not. Example of the results for the year 2018 are presented here. 2018 is the only year where corresponding vessels are included both in WMR market sampling and PFA self-sampling, where, the market samples come from 2 different vessels, which are also involved in the PFA self-sampling (Figure 8). On Figure 8, the distribution of the length measurements are demonstrated for each vessel together with box plots. The vessels with corresponding numbers indicate the same vessel (i.e. MARKT-1 vs SLF-1 and MARKT-2 vs SLF-2 are same vessels sampled by both programmes). The results are presented in Table 2 where the pink background of the table indicates significant difference between the pairs judged by the p-values (post-hoc Tukey HSD test). Nine out of 21 comparisons showed no significant difference in the median length, for which the difference in median length was smaller than 1 cm. The remaining compared vessels had differences in median length of at least 1 cm and were significantly different from each other. These results indicate that, since more than half of the comparisons showed significant differences, it is necessary to understand the reason behind having such differences before any attempt of integrating these datasets.

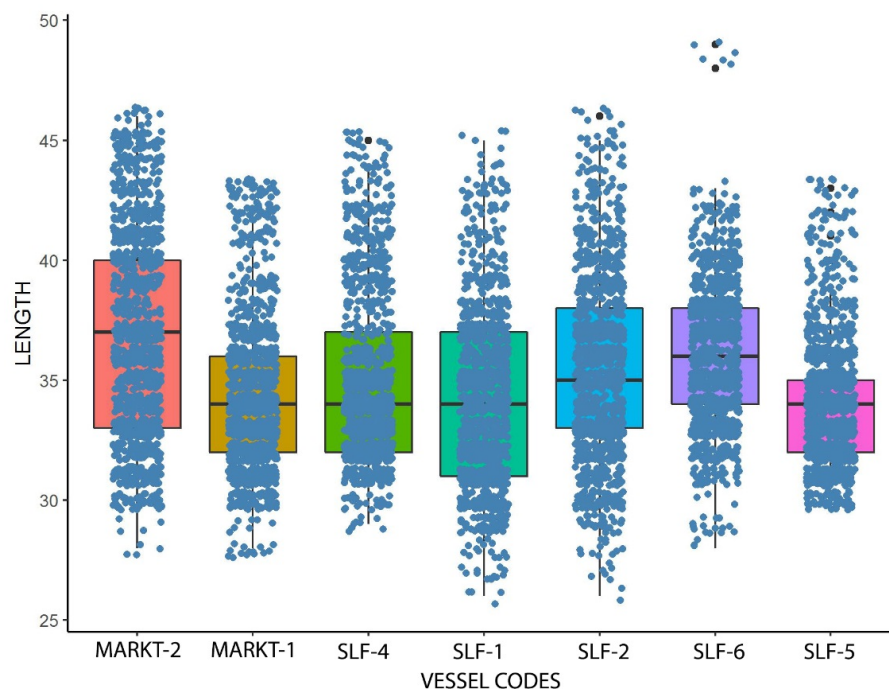


Figure 8 Distribution of the length measurements from different vessels in 2018. First two sets are from the WMR market samples. Rest of the groups are the vessels that contributed to the PFA self-sampling dataset of the PFA. MARKT-1 the same vessel as SLF-1 (and same for MARKT-2 and SLF-2). The center of the box plot shows the median length.

Table 2 Pairwise summary of the one-way ANOVA test. Columns indicate the compared pair of vessels, the differences in the overall length means and the associated p-values. Pink background indicate significant difference between the groups.

Pair of vessels compared (Year 2018)	Difference in Mean Length	P
MARKT-1-MARKT-2	-2.3	0
SLF_5-MARKT-2	-2.2	0
SLF_1-MARKT-2	-2.1	0
SLF_4-MARKT-2	-1.8	0
SLF_5-SLF_6	-1.5	0
SLF_5-SLF_2	-1.1	0
SLF_2-MARKT-2	-1	0
SLF_2-SLF_1	1.18	0
SLF_6-SLF_4	1.19	0.01
SLF_2-MARKT-1	1.21	0
SLF_6-SLF_1	1.63	0
SLF_6-MARKT-1	1.65	0
SLF_2-SLF_4	0.74	0.16
SLF_6-MARKT-2	-0.6	0.39
SLF_6-SLF_2	-0.5	0.53
SLF_5-SLF_4	-0.4	0.83
SLF_4-MARKT-1	0.46	0.9
SLF_1-SLF_4	-0.3	0.97
SLF_1-MARKT-1	0.03	1
SLF_5-MARKT-1	0.18	1
SLF_5-SLF_1	0.15	1

### 3.3 WMR Market sampling: Overview of age and length and maturity information

The first set of data in the market sampling database comes from an experimental fishing campaign conducted in the Porcupine Bay and British Isles in 1990 (Heessen and Kuitert 1990). From the collected samples during this initial year, 780 fish were aged, 706 of them were from the 6A division. In the following years, commercial fishery by the Dutch fleet carried on as well as the onboard market sampling. Data from early years (until 1995) are characterized by a relatively larger mean length (Figure 9 (A)) and greater number of old individuals reaching up to 40 years. Figure 10 shows the age at length from the WMR market sampling per year with Von Bertalanffy growth curves fitted to pooled data over entire time series based on gender. The older age groups start to disappear in 1995 (Figure 11). There is also a change in the length composition over the years (Figure 10 and 11). The mean length shows a decline from the 1990s until the year 2000, then increases slightly until 2010 and then shows another sharp decline until present (Figure 9 (A)). Furthermore, the proportion of females, similar to the decline in mean length, also shows a negative trend from 2010 onwards (Figure 9 (D)). These negative trends correspond to a similar decline in the CPUE of the commercial fishery in the ICES division 5b (Figure 9 (C)). A complete overview of the market sampling data set for the greater argentine can be seen in Appendix II.

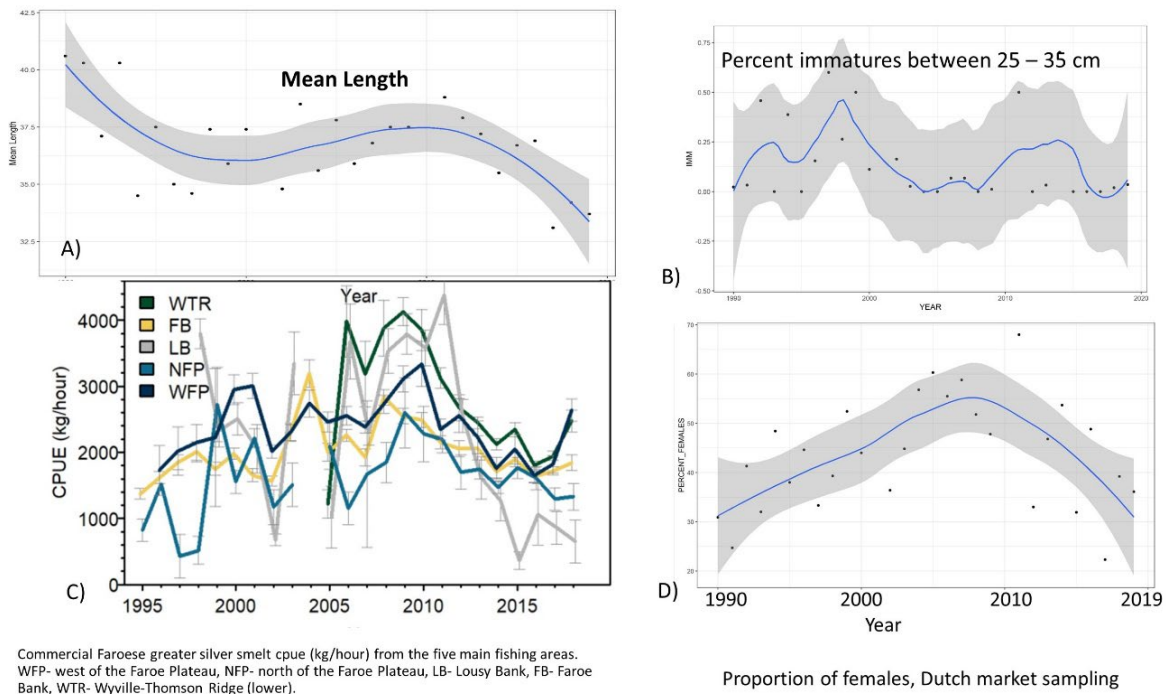
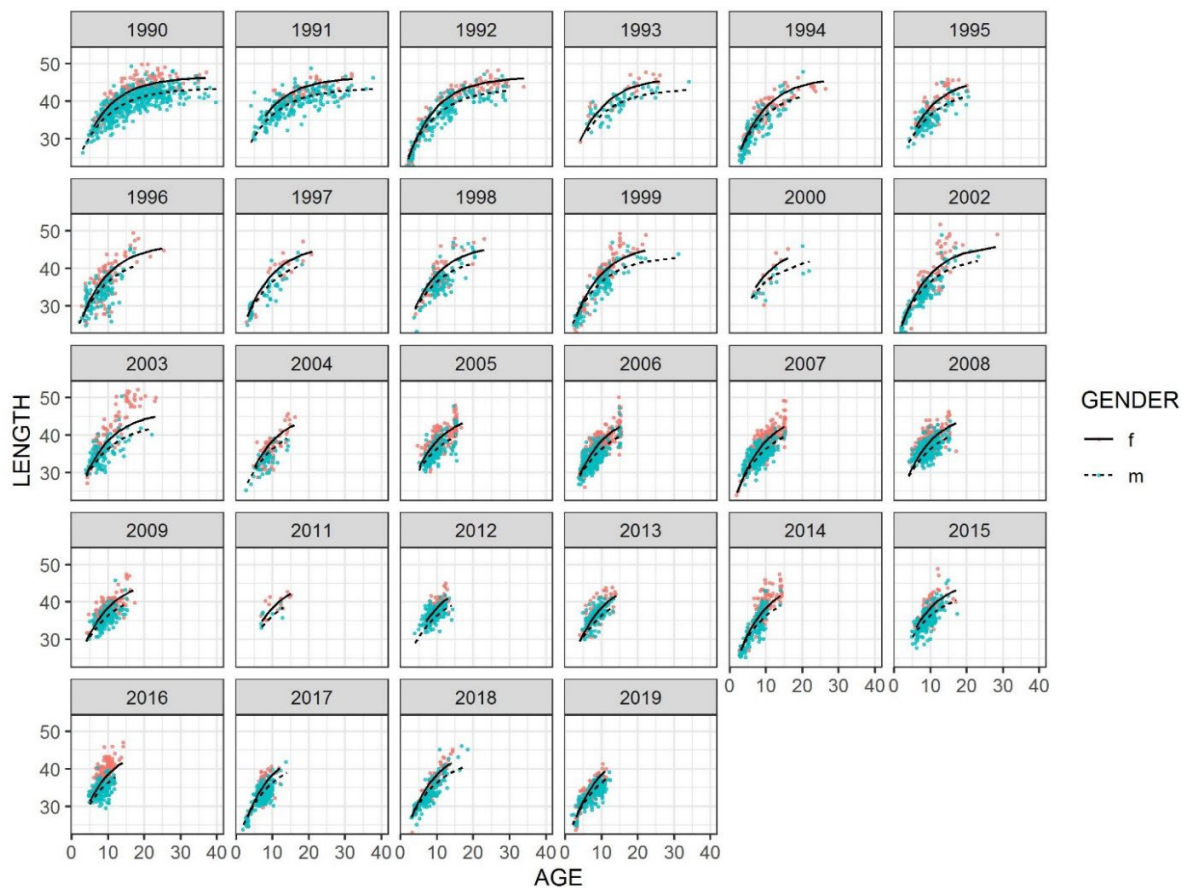


Figure 9 A) Mean length per year in the market sampling time series. B) Percentage of the immatures within the 25-53 cm length classes C) CPUE estimations for the commercial Faroes fishing fleet for different sections of the division 5b d) percentage of females. Note: the size of the panel C is adjusted in a way that the time scale in the x-axis visually matches each other.

### 3.3.1 Age readings and growth patterns in WMR market samples

A clear asymptotic growth pattern over the first 5 years is visible on the age-length distributions which starts to change in 1995 and onwards. This pattern changes as the age classes older than 20 disappear (Figure 10). On this figure, there is a visible accumulation around the 14-20 years age classes in some years (e.g. 1996, 1999, 2002 and 2003) and starts to become truncated from 2004 onwards. Based on personal communication with the second and third age reading experts of greater argentine, it was revealed that recognition of the otolith rings beyond 15 years age group has been difficult. Therefore implied change in growth rate over the years is potentially not genuine as the age data is noisy at oldest ages and sample sizes are low in several years. There is an improvement from 2012 onwards regarding such deviations as the distribution of the age vs length are consistent without an apparent truncation. This improvement, ruling out the age reading related issues, could be an indication that older age classes have disappeared as a result of exploitation. This change also correlates with the decline in the average lengths (Figure 9 (A) and Figure 11) and in the proportions of the females in the samples (Figure 9 (B)). It is known for this species that the female growth and habitat preference and size range is different from the males. This can be seen on the age-length scatter plots discriminated by gender on Figure 12. The different growth rate of the females and males becomes apparent after 7-8 years old where the females growth rate is higher.



*Figure 10 Scatterplots for the historical age length distribution. The colors indicate the gender, the pink indicate the female and the blue indicate male. The earlier years from 1990 to 1995 show a wide range of length classes including many old age classes up to 40 years old. In the later years these age classes disappear. It is also important to note that after 1995 the otolith reading started to be carried out by a new person.*



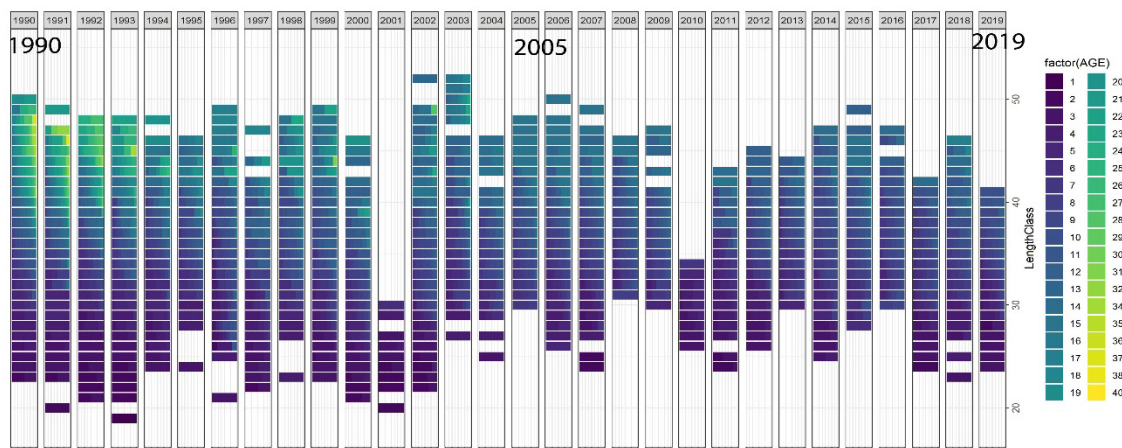


Figure 11 Greater argentine length distributions decomposed by the proportions of the ages in each length classes in each year as indicated by the colours. The yellowish colours indicate older fish

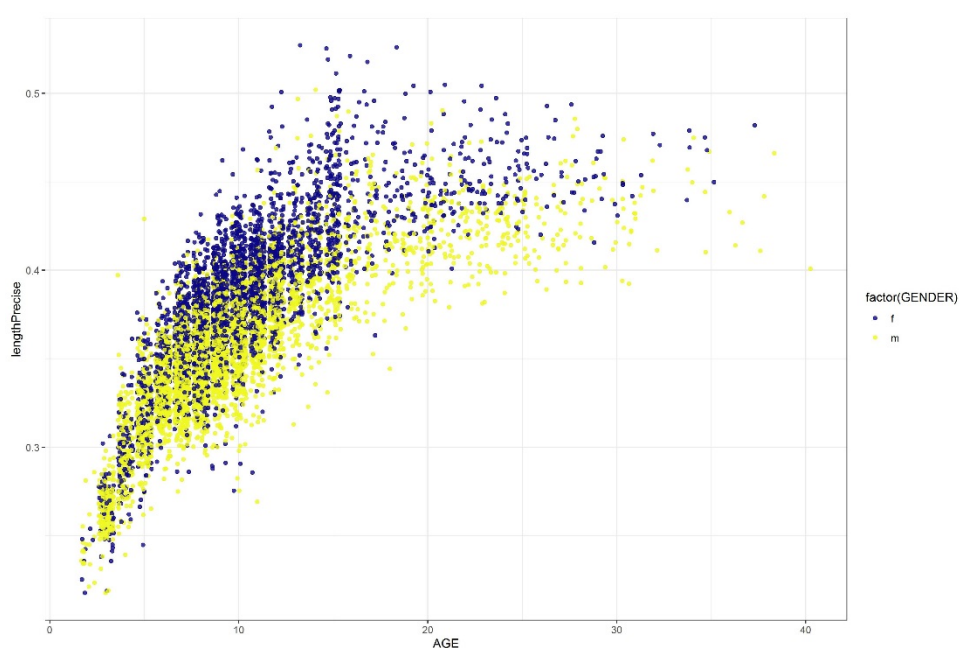


Figure 12 Market sampling age readings vs length, all years combined. Colours indicate the gender. Dark-blue indicate the females and yellow indicate males.

### 3.3.2 Changes in the age, length and gender structure

Individual growth rate of greater argentine differs between the females and males (also observed in this data, see Table 4) particularly after the age of 8-9 years old. Normally, it may be assumed that, as long as the male vs female proportions remain the same, the overarching growth rate remains consistent even though its standard deviation would be higher relative to a case where there is no difference between male and female growth rate. However, because the fishery may go through adaptations based on the environmental changes and gained knowledge and experience over time, the catch composition may change. For example, the male/female proportions are known to be dependent, and changed fishing pattern may lead to changes in the gender ratios in the catches over years. Such changes, especially in the older ages, may result in changes in the observed average growth between the years if gender discrimination is not applied. This might lead to violating the assumption of no variation in size at age over time. In addition, throughout the time series, the sample size, age reading precision and accuracy is not uniform. This is potentially leading to a low signal-to-noise ratio in some years which could be causing a discrepancy between the combined and forward key predictions. Furthermore, a pattern in the data suggests that in certain years (e.g. 2005 to 2009), 15 year age class was treated as the plus group, which resulted in an accumulation of

different size groups at this age class (Figure 10 and 12). It is assumed that the accuracy in the age readings from the otoliths is generally not higher than 65% (pers. comm. with staff within WMR and internationally in ICES). This noisiness also potentially adds to the discrepancies.

*Table 3 Von Bertalanffy growth parameters calculated from the pooled data separately for males and females*

	<b>L<sub>INF</sub></b>	<b>K</b>	<b>T<sub>0</sub></b>
<b>FEMALE</b>	46.49	0.13	-3.99
<b>MALE</b>	43.54	0.12	-5.34

In the years such as 1994 and 2019 where the age variability in the length groups are relatively small, the combined method predictions seems to be more similar to the forward method. This promising similarity may suggest that, the combined method may help in dealing with the noisiness in the otolith readings. It would be interesting to perform similar tests on the data of the Faroese fishery on greater argentine, as well as on data of a different species where the sample size is larger and the age-reading accuracies are higher.

The early years (first half of the 90s) in the data are characterized by relatively larger mean length and greater presence of old individuals reaching up to 40 years. The older age groups start to disappear in 1995. There is also a change in the length composition over the years. The mean length shows a decline from the 1990s until the year 2000, then increases slightly until 2010 and then shows another decline until the most recent year. Correlated with the decline from 2010 onwards, there is also a decrease in the proportion of females in the gender composition. There is also a discernible pattern in the proportion of the immatures in the group negatively correlated with the changes in mean length and gender ratios. In order to calculate the maturity ratio, the focus was given to the 25-35 cm fish (1st and 50<sup>th</sup> percentiles). This period of decline also corresponds to a similar trend in the CPUE predictions from the commercial greater argentine fishery for the different sections of the ICES division 5b ( see panel C on Figure 9).

Trends in the market sampling data indicate changes in the population structure with respect to age, length, gender and maturity. The decline of the proportions of the females in the samples maybe a signal of a risk. The fact that size of females are larger than the males and that females are found in deeper part of the water column separated from males, may have led to a disproportional depletion of the females. This may be a concern for the reproductive success of the stock. It is important to further investigate how the depth range of the fishing effort has changed over the years and whether there is a link between the changed gender ratio and the age composition with changed fishing depth. Results from preliminary assessments on greater argentine indicate the stock to be declining over the past 10 years while being at or just below Bpa, being well above Blim. Fishing mortality is preliminary estimated to be below Fmsy.

### 3.4 PFA self-sampling, an exercise for age estimation

A combined key is generated from the market sampling age data to be used for estimating the age composition of the greater argentine catches based on the length distributions in by PFA self-sampling dataset. For constructing the combined key, samples were pooled separately for each year and each year were treated as an independent sample. Because market sampling is composed of a long set of time series and there are some concerns regarding the accuracy of the ages determination in the older age classes in different years, different trials were carried out to see the effects of filtering/manipulating data in different ways. As explained in the section 2.1.1, 5 different scenarios were constructed by applying an upper limit by either treating the older ones as plus group or removing them altogether (15 and 20 years) and also removing some of the erratic data points those with small sample size or those having large deviations from the theoretical growth curve. In addition, although the pooled data comes from the same ICES region, the locations of the samples geographically vary. Therefore in order to take such variability into account, a set of trials were also performed by using each sample separately. As a result, the scenario (d) in section 2.1.1, produced

the best score (lowest RMSE) and has been accepted as the final combined key. In this scenario the years 2004-2007 are entirely removed from the dataset and the rest of the data points were pooled by the year. It is important to the years 2009-2011 are particularly important for determining the best fit for the scenario d. Otherwise, except from the scenario b, similar results have been produced.

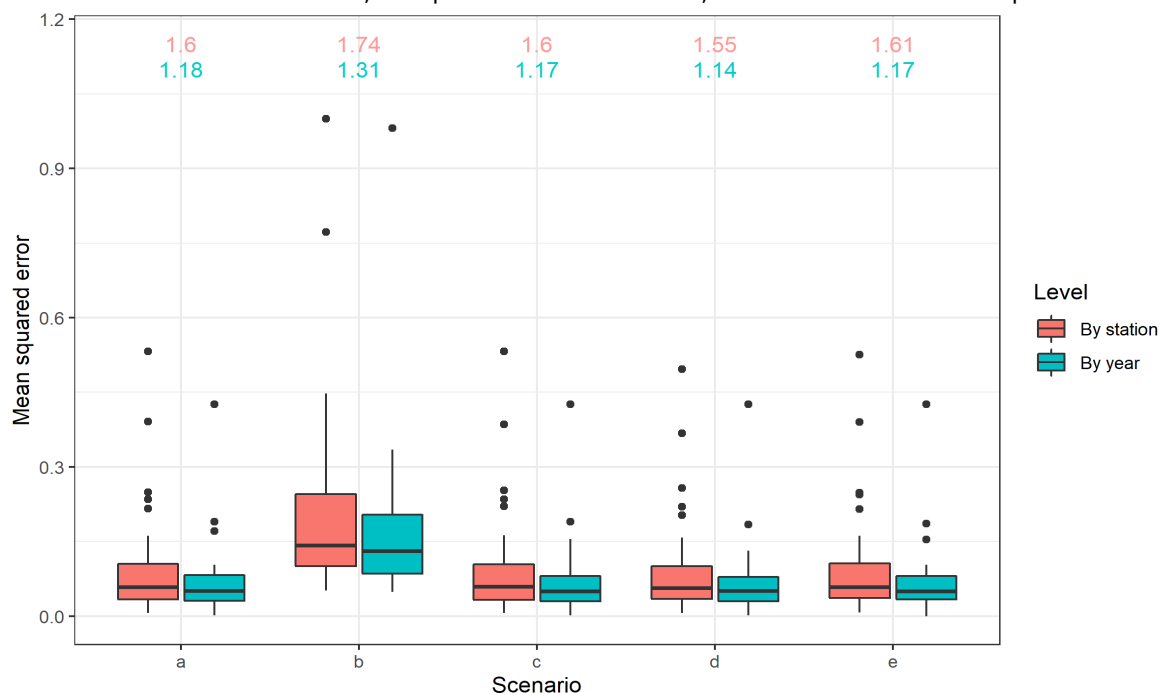


Figure 13 The tested scenarios for combined key generation and their normalised root mean squared errors used as a score for picking the best scenario (the scenario d).

### 3.4.1 Evaluation of the combined key method

Figure 14 show the age composition predicted by the combined key method and pooled age compositions per each year in separated panels. In each panel the normalised root mean square errors is also indicated. In most years there is a reasonable fit where the best was in the first year (1990). Figure 15 shows normalised RMSE scores for different years based on different scenarios. In this first year the sample size is large and older ages are well represented up to 40 years old. This is also the year where there is a reasonable fit between the fitted Von Bertalanffy curve and the age/length relationship. The combined key tends to generate a smooth curve suggesting loss of some of the information. On the other hand, the accuracy of the otolith readings are not expected to be very high (i.e. < 65%), it may also be possible that the combined key is correcting some of the errors in the age readings.



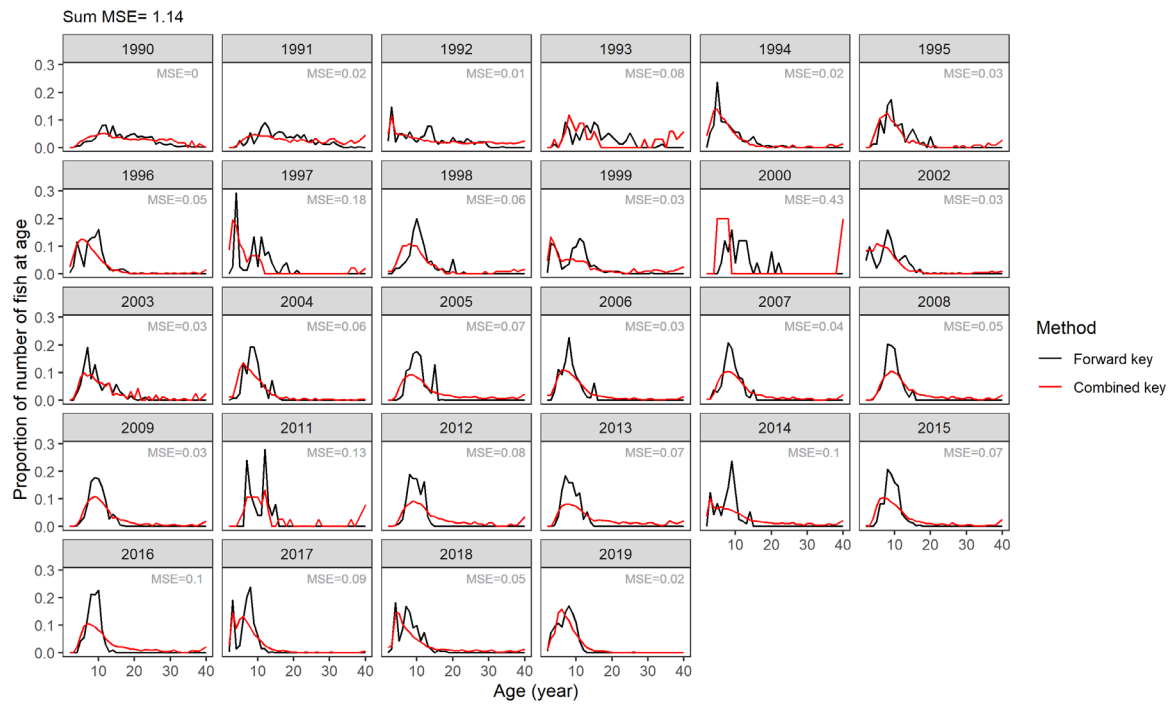


Figure 14 Comparison of the combined vs forward age/ length key methods applied on the market samples. The black line show the age composition based on the classical ALK ( forward) method and orange line shows the combined method in which a pooled set of age/length information is used to predict the size at age.

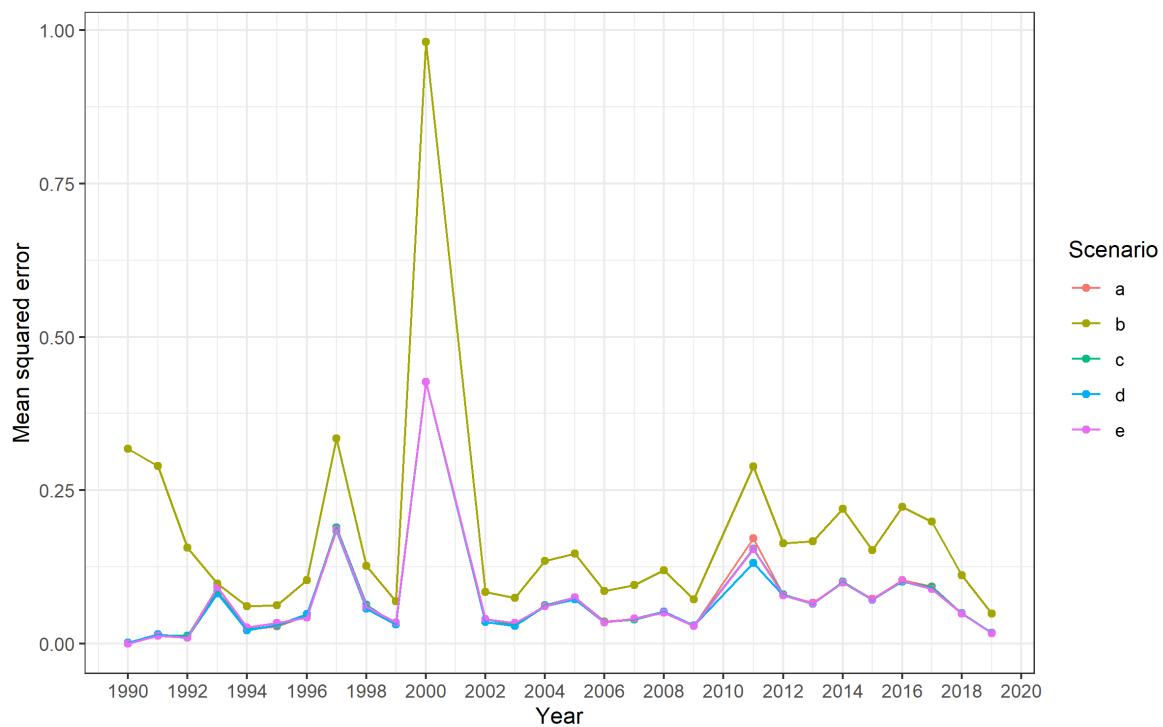


Figure 15 Normalised root mean squared errors for different years for different scenarios

---

### 3.5 Assessments using different catch at age matrices

An assessment making use of the ICES WGDEEP benchmark agreed data were used in an FLSAM assessment exercise. The exercise consisted of testing for plus group (Figure 16, table 5), highest true age with independent selection estimation (Figure 17) and general configuration of the SAM model (Table 5).

The assessment results were sensitive to the plus-group setting showing markedly different trends in especially SSB with plus-groups 20 and 21. From 19 and lower, there are hardly any differences in stock trends, except for plus-group age 15 where Fbar shows a different trend around the 2000s. From this, it is concluded that a plus-group of 18 or 19 would suite the model best also given the low Mohns-rho statistics and low CV on process-error which is one of the parameters that is more difficult to estimate given the poorer quality of the data.

*Table 5. Mohns-rho statistic (3-year peal) for the different plus-group runs*

age	21	20	19	18	17	16	15
SSB	-0.70	0.25	-2.45	-1.2	-0.81	-0.87	-7.36
Fbar	1.35	-0.11	4.80	2.50	1.39	0.53	4.79
Rec	5.24	5.70	5.69	5.87	6.15	6.06	3.60

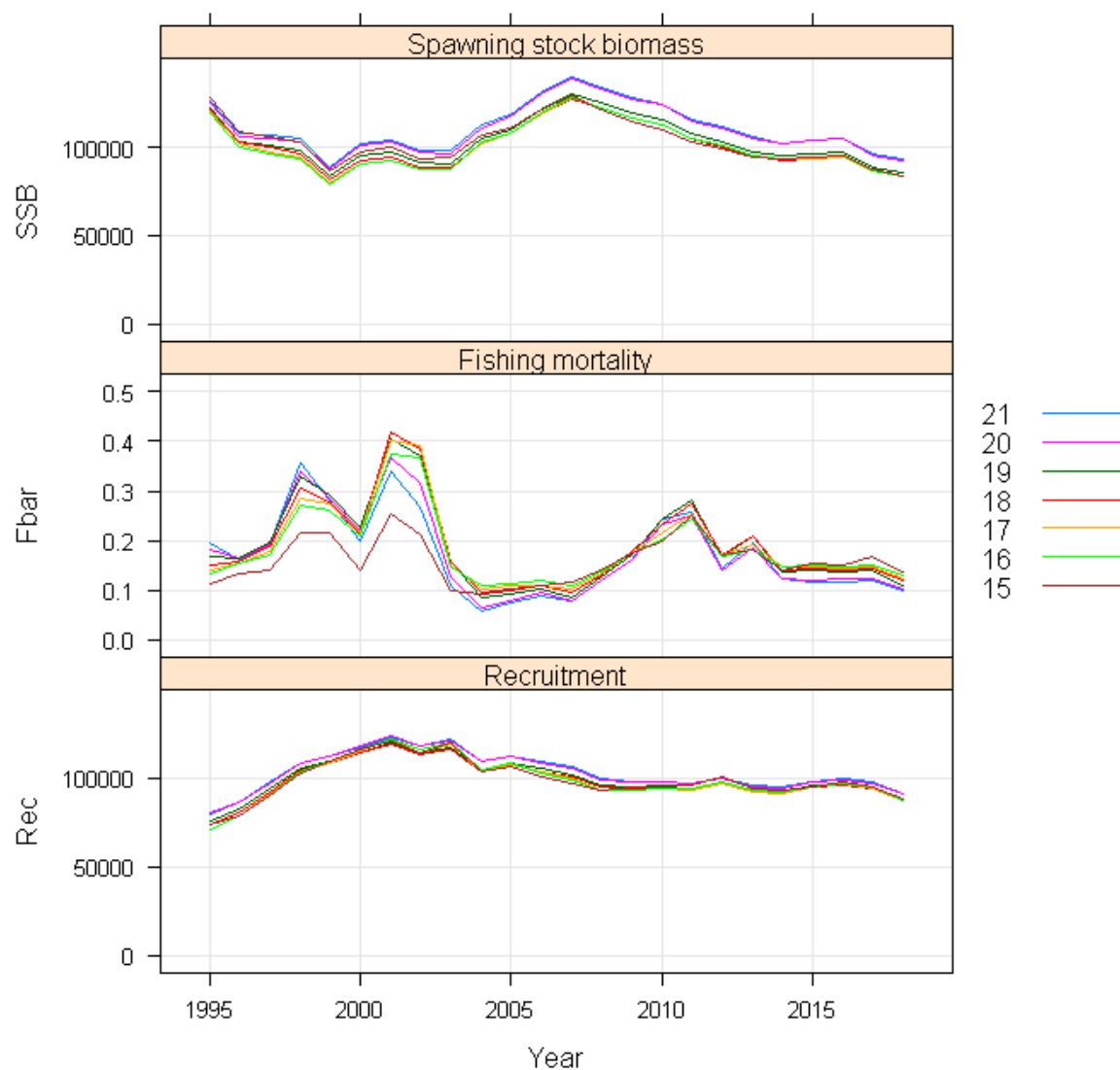


Figure 16. Trends in SSB,  $F_{bar}$  and Recruitment varying the plus-group age from 21 down to 15.

Maximum age with flexible selection (free parameter for states) was tested fitting a linear model through as many older ages as possible without intersecting with the confidence bounds of a model (plus-group of 21) with all free states parameters. This showed that assuming the same selection from age 14 onwards could be considered appropriate.

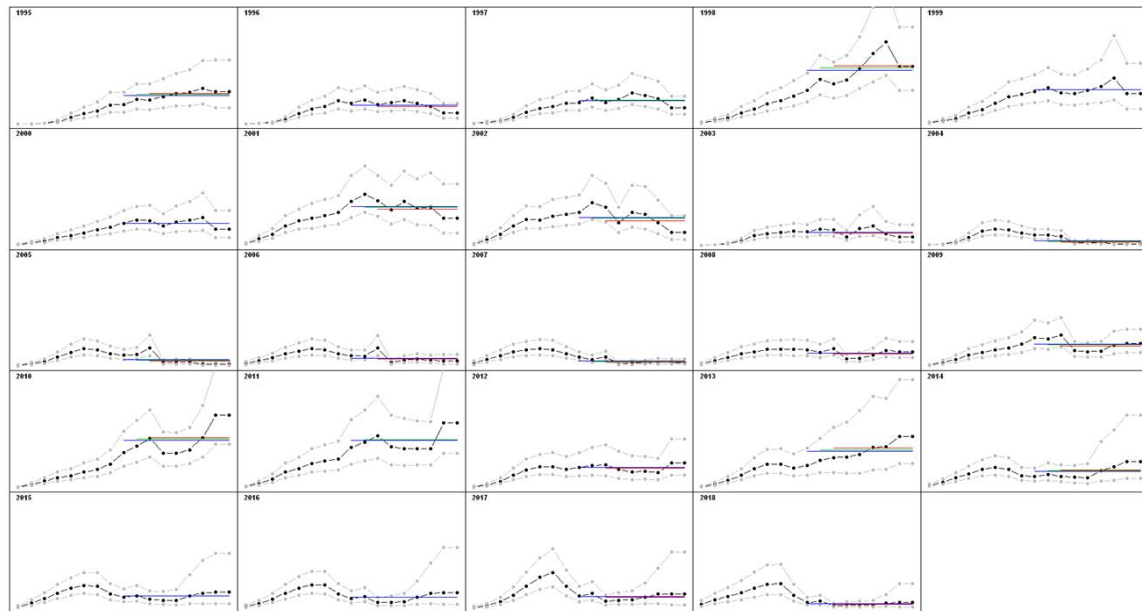


Figure 17. Fitted mean selection with averages over a range of ages plotted on-top (straight lines)

Table 6. FLSAM settings of the WMR proposed model

\$minAge	5	
\$maxAge	18	
\$maxAgePlusGroup	1 0 0 0 0	
\$keyLogFsta		# Coupling of the fishing mortality states (nomally only first row is used).
catch	0 1 2 3 4 5 6 7 8 9 9 9 9 9	
Faroese summer surv	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
Faroese deepw surv	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
Scottish deepw surv	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
Faroese/EU CPUE	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
\$corFlag	2	# Correlation of fishing mortality across ages (0 independent, 1 compound symmetry, 2 AR(1), 3 separable AR(1).
\$keyLogFpar		# Coupling of the survey catchability parameters (nomally first row is not used, as that is covered by fishing mortality).
catch	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
Faroese summer surv	0 1 2 3 4 5 6 6 -1 -1 -1 -1	
Faroese deepw surv	9 10 11 12 13 13 13 13 13 -1 -1 -1	
Scottish deepw surv	7 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
Faroese/EU CPUE	8 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
\$keyQpow	All -1	# Density dependent catchability power parameters (if any).
\$keyVarF		# Coupling of process variance parameters for log(F)-process (nomally only first row is used)
catch	0 1 2 3 3 3 3 3 3 3 3 3 3	
Faroese summer surv	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
Faroese deepw surv	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
Scottish deepw surv	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
Faroese/EU CPUE	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
\$keyVarLogN	0 1 1 1 1 1 1 1 1 1 1 1	# Coupling of process variance parameters for log(N)-process
\$keyVarObs		# Coupling of the variance parameters for the observations.
catch	2 3 4 4 4 4 4 4 4 5 5 5 5	
Faroese summer surv	6 6 6 6 7 8 8 8 -1 -1 -1 -1	
Faroese deepw surv	9 10 10 10 10 10 10 10 10 -1 -1 -1	
Scottish deepw surv	0 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
Faroese/EU CPUE	1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
\$obsCorStruct	"ID" "AR" "AR" "ID" "ID"	# Covariance structure for each fleet ("ID" independent, "AR" AR(1), or "US" for unstructured).   Possible values are: "ID" "AR" "US"

<i>\$keyCorObs</i>		# Coupling of correlation parameters can only be specified if the AR(1) structure is chosen above.
<i>catch</i>	NA NA NA NA NA NA NA NA NA NA NA	# NA's indicate where correlation parameters can be specified (-1 where they cannot).
<i>Faroese summer surv</i>	0 0 0 0 0 0 0 -1 -1 -1 -1 -1	#5-6 6-7 7-8 8-9 9-10 10-11 11-12 12-13 13-14 14-15 15-16 16-17 17-18 18-19 19-20 20-21
<i>Faroese deepw surv</i>	-1 1 1 1 1 1 1 1 1 -1 -1 -	
<i>Scottish deepw surv</i>	1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
<i>Faroese/EU CPUE</i>	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
<i>\$stockRecruitmentModelCode</i>	0	# Stock recruitment code (0 for plain random walk, 1 for Ricker, 2 for Beverton-Holt, and 3 piece-wise constant).
<i>\$noScaledYears</i>	0	# Number of years where catch scaling is applied.
<i>\$keyScaledYears</i>		# A vector of the years where catch scaling is applied.
<i>\$keyParScaledYA</i>		# A matrix specifying the couplings of scale parameters (nrow = no scaled years, ncols = no ages).
<i>\$fbarRange</i>	9 18	# lowest and highest age included in Fbar
<i>\$keyBiomassTreat</i>	-1 -1 -1 0 0	# To be defined only if a biomass survey is used (0 SSB index, 1 catch index, 2 FSB index, 3 total catch, 4 total landings and 5 TSB index).
<i>\$obsLikelihoodFlag</i>	"LN" "LN" "LN" "LN" "LN"	# Option for observational likelihood   Possible values are: "LN" "ALN"
<i>\$fixVarToWeight</i>	0	# If weight attribute is supplied for observations this option sets the treatment (0 relative weight, 1 fix variance to weight).
<i>\$fracMixF</i>	0	# The fraction of t(3) distribution used in logF increment distribution
<i>\$fracMixN</i>	0	# The fraction of t(3) distribution used in logN increment distribution
<i>\$fracMixObs</i>	0 0 0 0 0	# A vector with same length as number of fleets, where each element is the fraction of t(3) distribution used in the distribution of that fleet
<i>\$constRecBreaks</i>		# Vector of break years between which recruitment is at constant level. The break year is included in the left interval. (This option is only used in combination with stock-recruitment code 3)

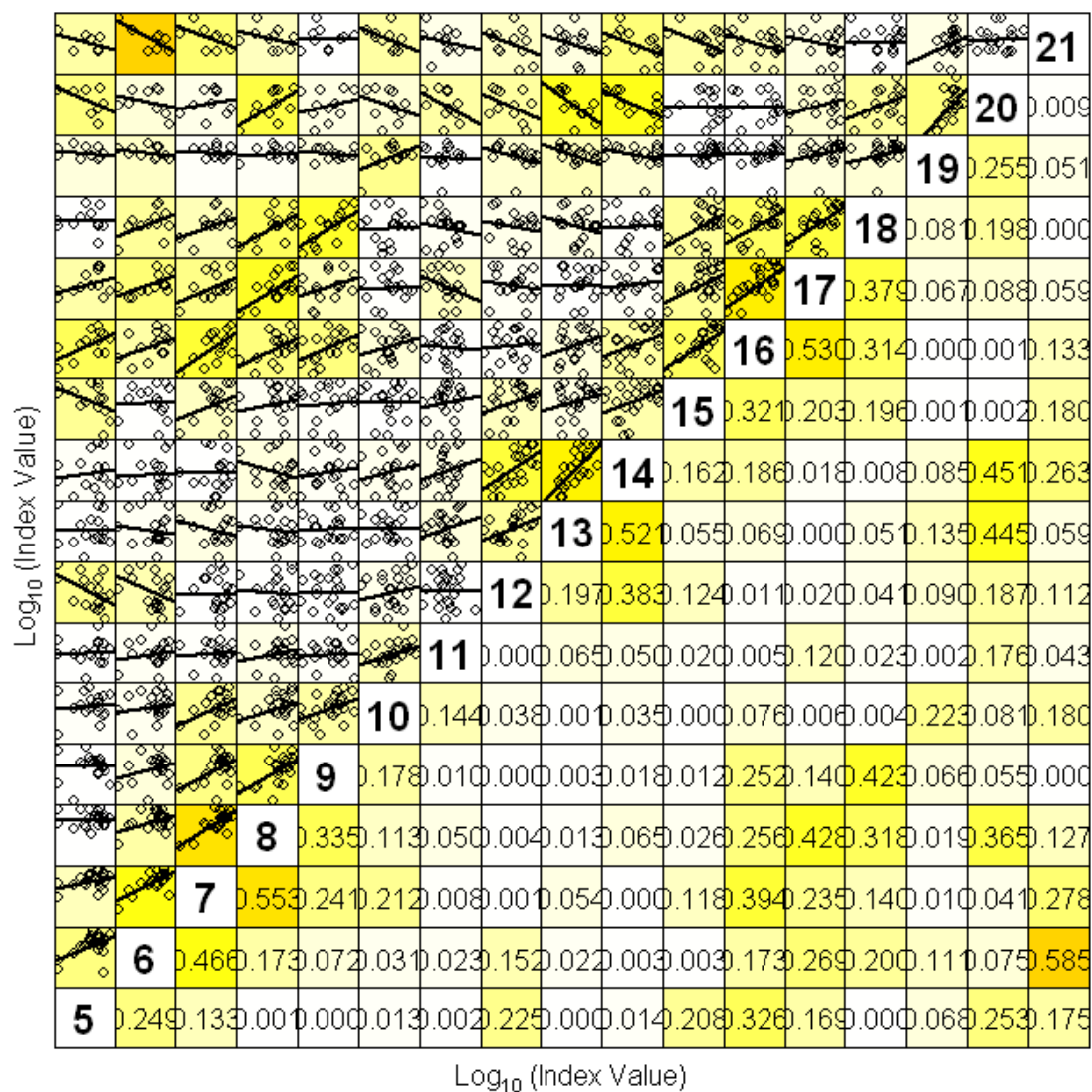
### 3.5.1 Evaluation of the combined key method

We evaluate the internal consistency of the catch@age data when substituting the Dutch data (submitted to InterCatch using the forward-key) with the combined key and the combined key with PFA data included. The internal consistency plots are shown in the three panels below (Figure 18). The results show clearly that the internal consistency goes down when using the combined key and go down every further (although minimal) when using the combined key and the PFA self-sampling data.

One imminent question would be whether the deterioration of the internal inconsistency is simply due to the application of the combined key instead of the forward key. The key differences between the forward and combined key are given in the table below. A further evaluation in a simpler case with a different species would be useful to evaluate level of confidence that can be expected from this method.

Forward key	Combined key
<b>Change in growth rate as observed in the data</b>	Assumes no change in growth rate
<b>→ Variable length-at-age relationship</b>	<b>→ Constant length-at-age relationship</b>
<b>Representative for one year/subarea combination</b>	Representative for multiple years/subarea combinations
<b>Not fit for purpose when using external length measurements</b>	Least amount of bias when using external length measurements
<b>Variable results with potential noise</b>	Smoothed results

Although the quality of the catch-at-age data deteriorates (Figure 18) and total catch estimates are almost similar (Figure 19) the difference in perception is very large (Figure 20). It should be noted however that this results in a change in level, not in trend.



Lower right panels show the Coefficient of Determination ( $r^2$ )

Figure 18a. Internal consistency of the cohort signal in the catch-at-age data for the forward key. Each panel represents the log-log linear relationship between estimated numbers at age  $x$  vs age  $y$  (given on the diagonal). A positive slope is expected with observations (dots) close to the fitted line (solid lines in upper part of the panels) indicating strong internal consistency. Colour scale ranges from white (no consistent pattern, value  $\sim 0$ ) to dark red (highly consistent pattern, value  $\sim 1$ ).

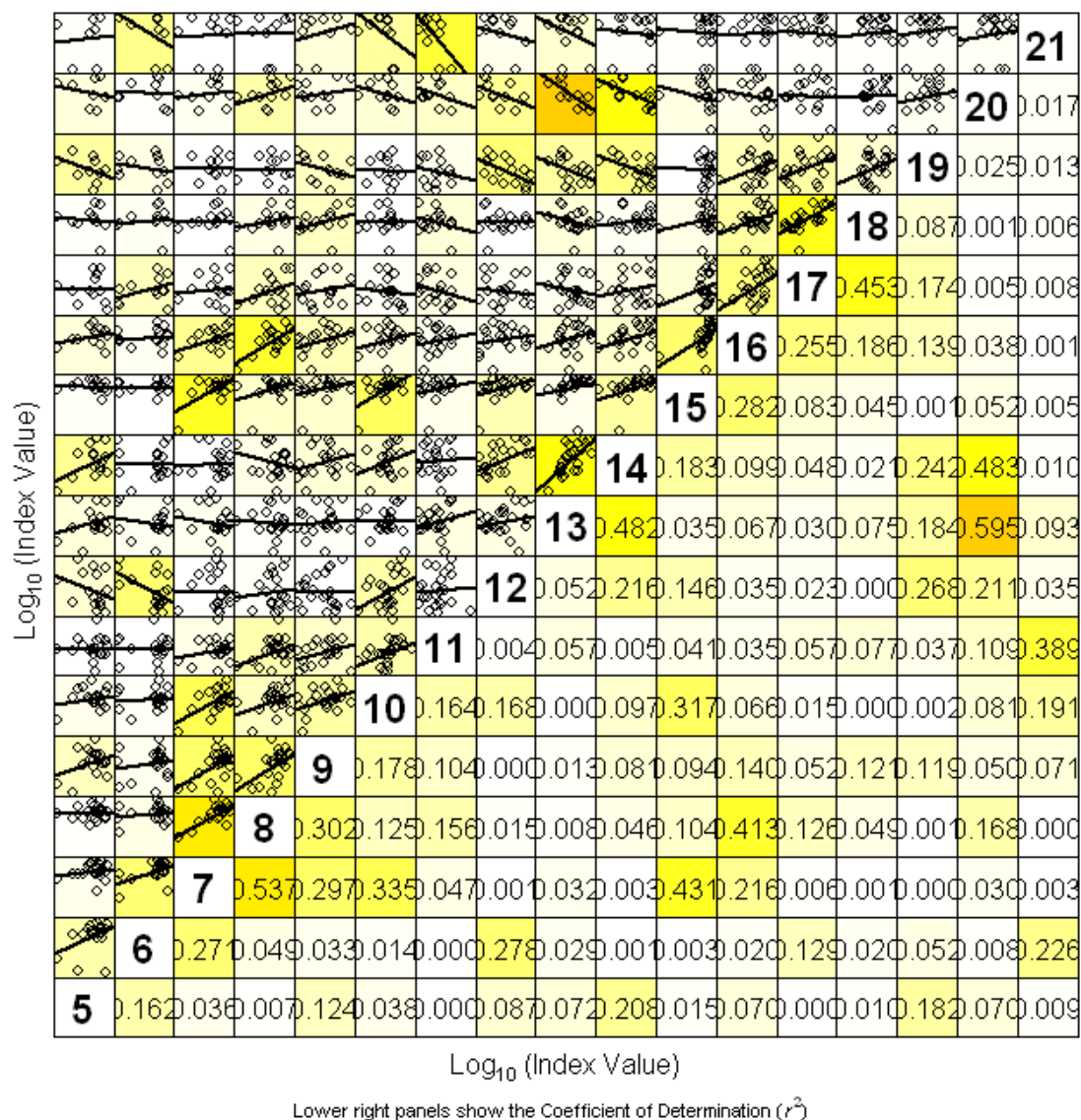
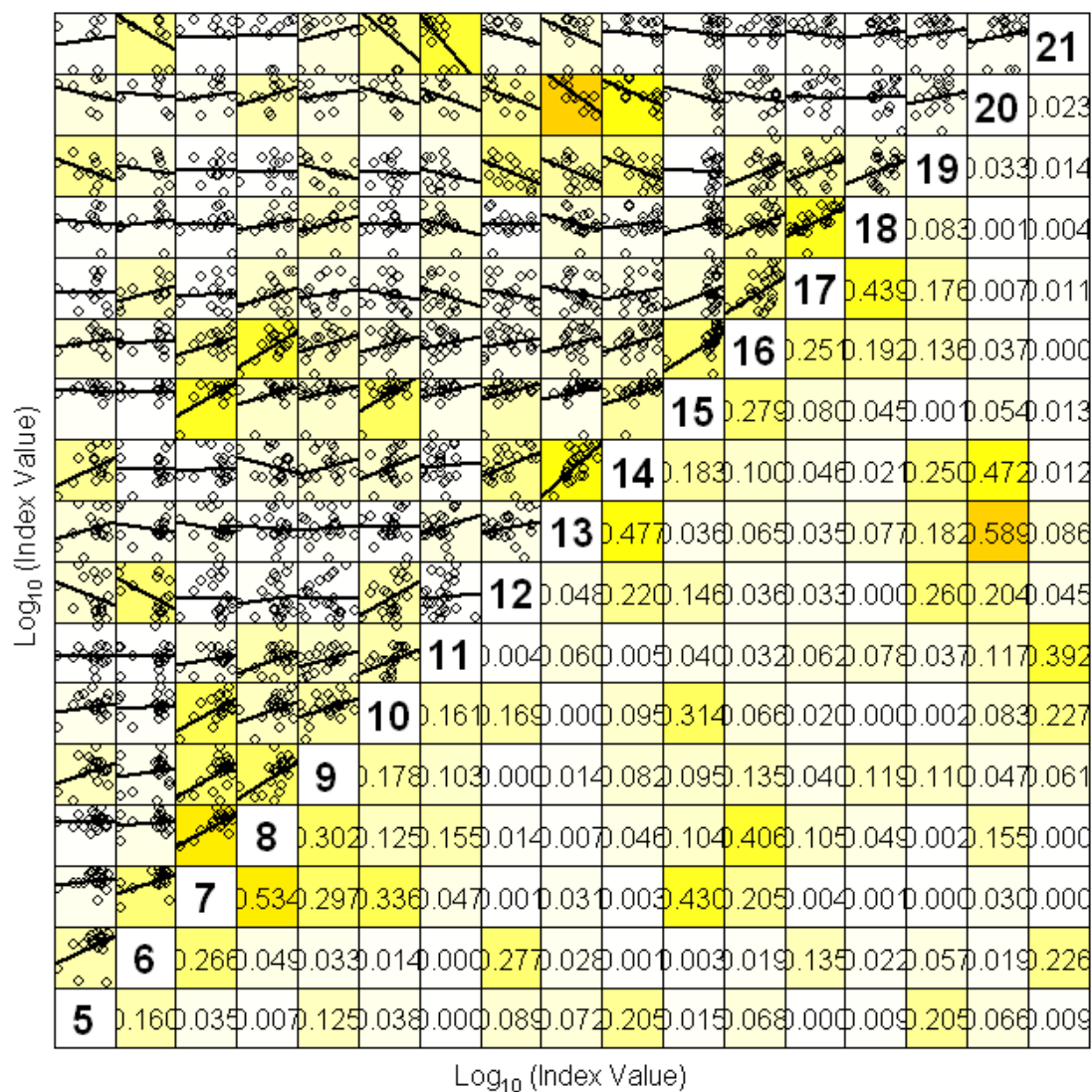


Figure 18b. Internal consistency of the cohort signal in the catch-at-age data for the combined key. Each panel represents the log-log linear relationship between estimated numbers at age x vs age y (given on the diagonal). A positive slope is expected with observations (dots) close to the fitted line (solid lines in upper part of the panels) indicating strong internal consistency. Colour scale ranges from white (no consistent pattern, value ~0) to dark red (highly consistent pattern, value ~1).



Lower right panels show the Coefficient of Determination ( $r^2$ )

Figure 18. Internal consistency of the cohort signal in the catch-at-age data for the combined key with PFA data included. Each panel represents the log-log linear relationship between estimated numbers at age  $x$  vs age  $y$  (given on the diagonal). A positive slope is expected with observations (dots) close to the fitted line (solid lines in upper part of the panels) indicating strong internal consistency. Colour scale ranges from white (no consistent pattern, value ~0) to dark red (highly consistent pattern, value ~1).



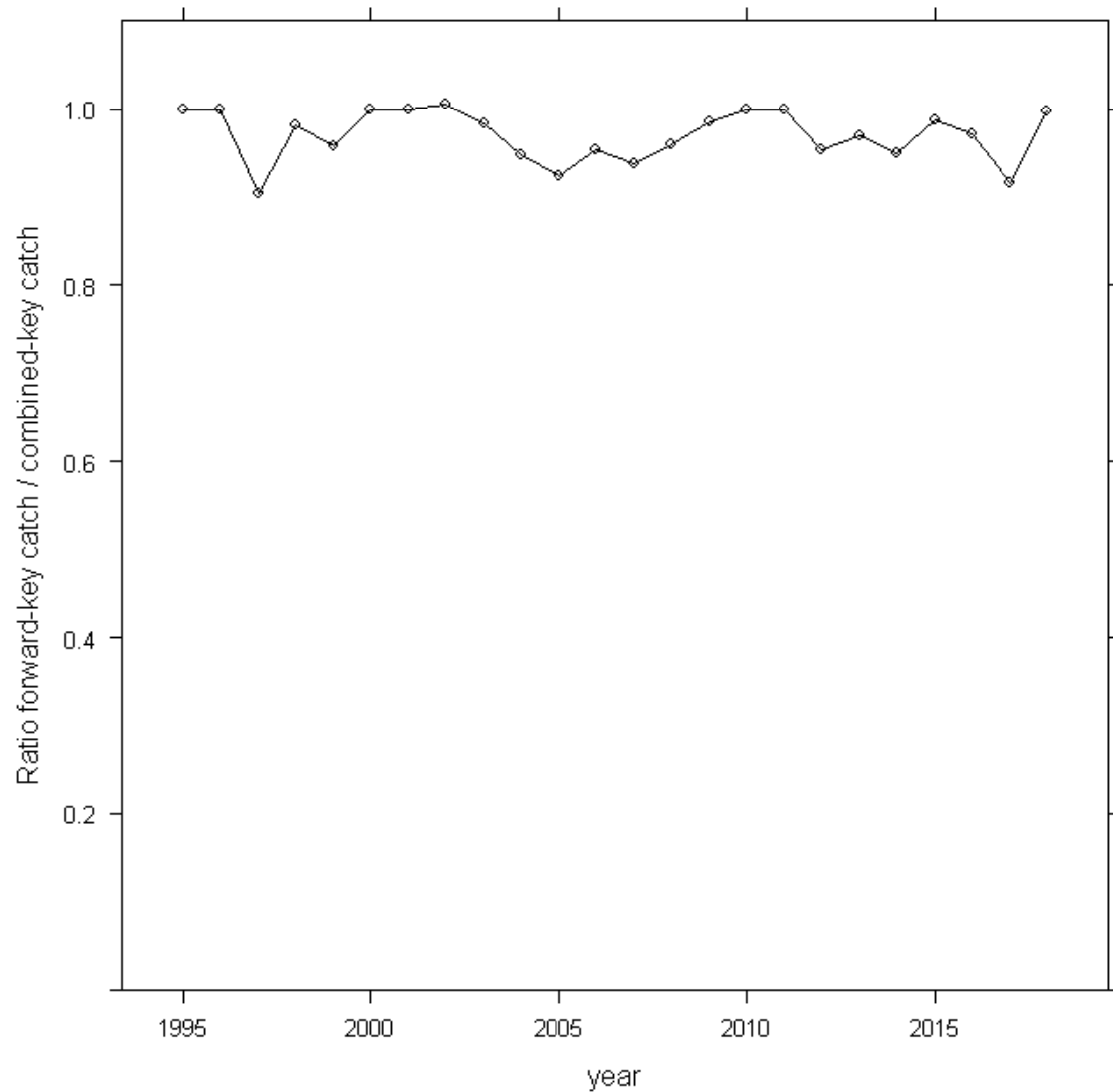


Figure 19. The assessment making use of the forward-key and combined-key estimates total catch. The ratio in estimated total catch between these two models is given. Both assessment models estimate similar levels of total catch.

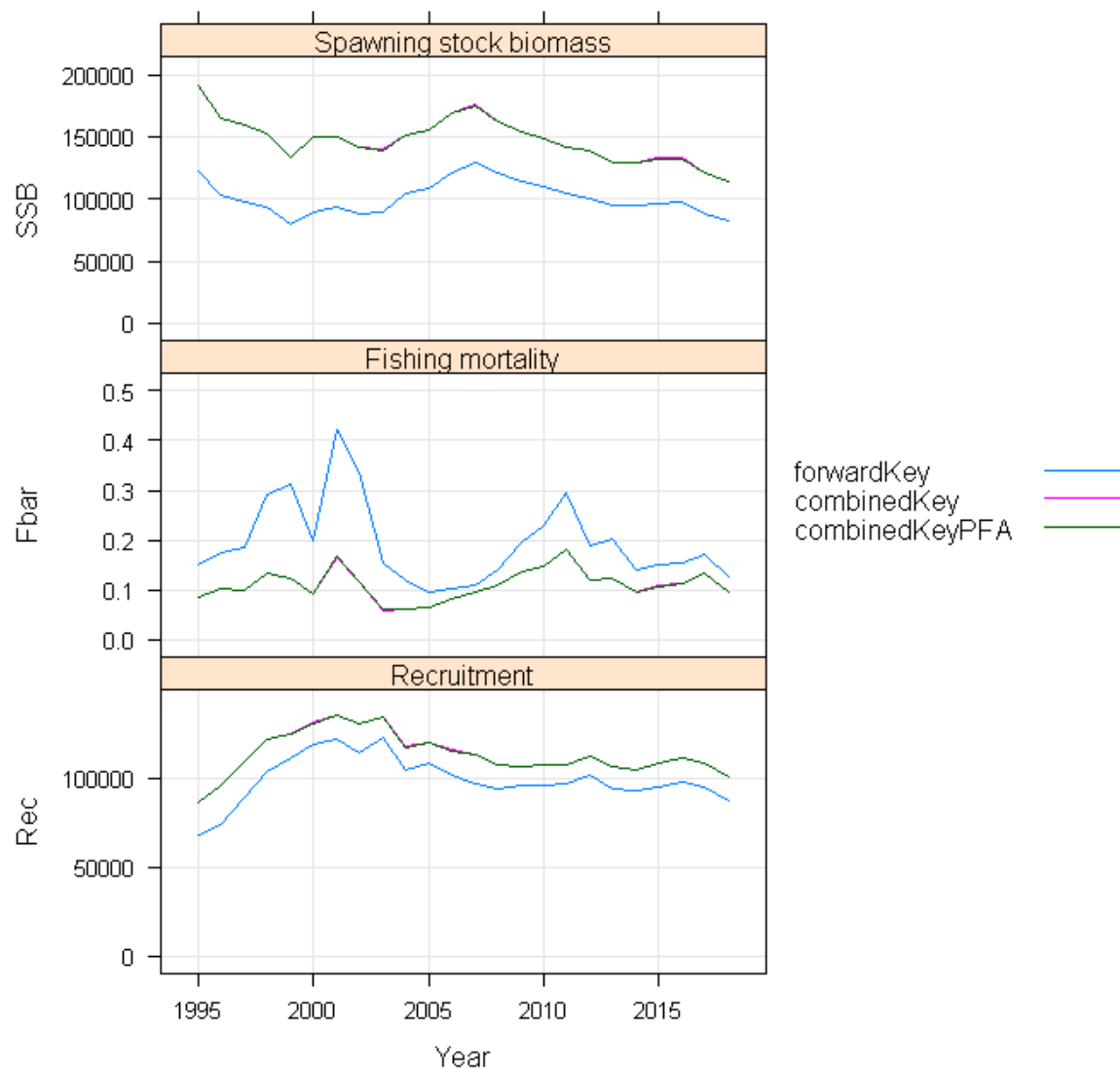


Figure 20. Perception in SSB, Fbar and recruitment under three different uses of age-length keys. Note that the green line is plotted almost entirely on top of the pink line. The forward key and combined keys estimate markedly different levels of SSB, Fishing mortality and recruitment, while estimated total catch is similar (see Figure 19).

### 3.6 CPUE standardization

No results of the CPUE standardisation are presented here as they were part of a WD by Floor Quirijns. Only expertise was delivered to support the standardisation process.

---

## 4 Conclusions

Relatively few market samples for each year come from a small number of fishing trips of few vessels, whereas from 2015 onwards self-sampling gives a much wider coverage and resolution in area and time, making its incorporation interesting for stock assessment. A detailed comparison was carried out between the self-sampling and market sampling length composition to assess its potential to be incorporated into the stock assessment process. Similarities as well as differences were found mostly in a random fashion. The natural stochasticity together with the lack of close spatio-temporal overlap between samples from different vessels makes it difficult to draw clear conclusions on the quality of the self-sampling data. While the similarities in length frequencies of the WOT market sampling and PFA self-sampling prevail when data is pooled over ICES subdivision, over the fishing season, comparisons between the individual samples taken show highly random discrepancies in length frequencies. In conclusion, although the data collection method by the PFA is without bias and conforming the needed level of quality in terms of measurement accuracy and precision, some small issues exist such as difficulty in distinguishing between the standard length and total length records in the data set. This problem can be tackled by comparing the sample weight versus the estimated weight of the measured specimens using a length – weight relationship. It is important to pinpoint the exact source of the issues highlighted above, i.e. which vessel or sampler (crew member) introduces these discrepancies, to mitigate future errors and maintain a more robust and trustworthy data delivery.

In order to make use of self-sampling data in stock assessment, number at age needs to be derived. A combined key was successfully generated and tested against the conventional age-length key results. This combined key can be used for predicting the age composition for greater argentines when only the length composition available. Combined key tend to produce relatively smooth age distribution.

The effect of different approaches to age-length-key modelling are substantial. Applying the combined key resulted in a completely different perception of stock level compared to the conventional forward key. The results likely change due to a change in how important each of the datasets is regarded by the stock assessment model (more weight on surveys when using the combined key as cohort signal is lower in that data). Although these results are premature, they do provide insights into the need for data quality control and validation when using industry self-sampling data. Once the reason behind the potential differences has been found and the issues solved, the Industry self-sampling data can be used for stock assessments, but if there is limited age-length information available, converting self-sampling data to ages using a modelled age-length key can result in unexpected trends.

---

## 5 Quality Assurance

Wageningen Marine Research utilises an ISO 9001:2015 certified quality management system. This certificate is valid until 15 December 2021. The organisation has been certified since 27 February 2001. The certification was issued by DNV GL.

## Data management plan for projects

Version WMR 1.0 , 9 november 2018

This datamanagement plan may be added as an Annex to the project proposal (Offerte), or be used as an internal document. If any of the categories is not relevant to the project, please mention. Filled in data management plans can be sent to WMR Kernteam Data: [kernteamdata.marine-research@wur.nl](mailto:kernteamdata.marine-research@wur.nl)

### General information data management plan

Roles	
Project (name and number)	OSW pelagic
WMR Project manager	Niels Hintzen, Serdar Sakinan
WMR Contact person data management plan (may be identical to project manager)	Serdar Sakinan
WMR Data manager (should be confirmed by TTT Data, <a href="mailto:ingeborg.deboois@wur.nl">ingeborg.deboois@wur.nl</a> Cc: <a href="mailto:kernteamdata.marine-research@wur.nl">kernteamdata.marine-research@wur.nl</a> )	Ingeborg de Boois

### Define data management roles within the *project*

If multiple types of data are collected and by different partners, be clear about the roles. Please list the project partner, even if individuals may collect the data on behalf of the project partner.

Roles	Project partner(s) and/or names
Who is <b>collecting</b> the data?	PFA and WMR/LNV (VISSTAT and Frisbe)
Who is <b>analysing</b> the data?	WMR together with PFA
Who is the <b>data owner</b> ? <sup>1</sup>	PFA
Who is responsible for <b>quality assurance and quality control</b> of the collected data?	PFA
Who is responsible for <b>data storage</b> ?	PFA, WMR

<sup>1</sup> when major software developments take place, also take licensing into account here

## Give an overview of expected types of research data and storage choices for WMR

NB: only use storage solutions in line with WUR Data Policy:

<https://www.wur.nl/en/Value-Creation-Cooperation/WDCC/Data-Management-WDCC/Data-policy/Storage.htm>

Data storage consists of two components:

- The storage of the data itself, sometimes in different levels of aggregation (raw/unaggregated, processed, maps, etc.). **Check** with the **datamanager** if standard import software and/or templates are available for the data that will be collected in the project.
- The storage of the scripts/code/queries/models used for data processing. The scripts should at least be stored in a sustainable manner, and re-usable by others without too much effort. This requires clear description of the purpose, the different steps within the script/model, and reference to external data used –if relevant.

### *a. Data*

Data stage	Description of type of data	Estimated size (MB/GB/TB)	Storage (type & location*)
Raw/unaggregated data	PFA: Catch per haul by species, If per haul, total weight per haul. No individual fish information. Other available information: depth, position, temperature at fishing depth, date, time, haul duration, shipname, net characteristics. LNV: total catch weight FRISBE: Market samples	Less than 1 GB	PFA: LNV: VISSTAT WMR: FRISBE
Processed data	N/A		
External data	N/A		
Model output	N/A		
Privacy data (names, shipcodes, etc.)	Ship codes are in the data, will be pseudonomised before publication.		
Other?			

\*this may be a link to a network drive, a reference to one of the WMR databases (preferred for unaggregated data), or external databases.

*b. Data processing*

Processing type	Specification of type of software	Storage (type & location)
Data processing code	N/A	
Models	N/A	
Other? <sup>2</sup>		

## Documentation of dataset/software

Fill in the table below to describe the dataset(s) on a general level. This facilitates reporting, re-use, findability, etc. as it provides metadata.

Theme	Element	Information
Content	What: Short description on the information collected, and reference to the project	Length frequency data of the greater argentine from PFA vessels collected within the scope of self-sampling together with the meta data. Aim is to incorporate this data into stock assessment procedures. Further details are in the report above.
	When: startdate, enddate (if not continuous) If the data are only collected in a specific time of year, please mention	2015-2019
	Where: general area (e.g. Wadden Sea, IJsselmeer, North Sea) and if possible minimum and maximum longitude and minimum and maximum latitude	ICES divisions 6a and 5b
Context	Contact person	Niels Hintzen
	Data access restrictions	Connected to frisbe and VISSTAT
	Project abstract	See project proposal

**If WMR is data owner and responsible for the data storage**, the information from the table above (metadata) can directly be made publicly available by means of a metadata record (contact [kernteamdata.marine-research@wur.nl](mailto:kernteamdata.marine-research@wur.nl) for help). The information on the data is then available for the public, and therefore Findable. The data itself won't be publicly available by doing this.

A metadata record should be created for the dataset(s) collected in this project	Yes/No**
--	----------

\*\*select relevant option

<sup>2</sup> when major software developments take place, take into account here

## Sharing, ownership and privacy<sup>3</sup>

WMR Data Policy article 5 (in prep.) applies to all data exchanges from WMR to other parties. **Preferably, the topic is discussed with all parties involved in the project prior to the start of the project, in order to manage expectations.**

	Topic	Appointment
Data <b>sharing from WMR to others</b>	Dataset description	
	Aggregation level	Unaggregated. Lf by haul.
	Pseudonimisation for data that can lead to physical persons	Yes, for vessel ids, Martin asks if it leads to problems. Registration of quality manager is possible, Martin will check to which extent this is possible. At least to the level of the vessel.
	Exchange format <sup>4</sup> and/or file type	Rdata
	Timing of data delivery	
	Channel for data delivery (e.g. email, sharepoint)	e-mail
	Storage at the receiver's end: access rights, duration of storage, etc.	Only delivery (combined WMR and PFA data in aggregated manner) to ICES, SPRFMO, STECF without further discussion. Data delivery to other organisations should be checked first with PFA (Martin).
	Allowed data use (e.g. only for the specific project purpose, is merging with other datasets without informing WMR allowed, etc.)	Data hasn't reached to a reliable level of accuracy throughout the project. Therefore use of this data set is not possible for the tasks other than the experimental trials within the scope of the current project.
	Dataset ownership	PFA
Data <b>sharing from others to WMR</b>	Dataset description	Length Frequency, metadata
	Aggregation level	Unaggregated. Lf by haul.
	Pseudonimisation for data that can lead to physical persons	Yes, for vessel ids, Martin asks if it leads to problems. Registration of quality manager is possible, Martin will check to which extent this is

<sup>3</sup> when major software developments take place, take into account here

<sup>4</sup> If a standard exchange format is available, name it here and use it in the project. Provide documentation on the format to the other parties. Check the datamanager if you're not sure if any standard format exists.



		possible. At least to the level of the vessel.
	Exchange format <sup>5</sup> and/or file type	To be discussed
	Timing of data delivery	For assessment purposes: ICES data calls should be followed. VISSTAT update is given end January/beginning February. If relevant to HAWG (end January), WGDEEP (end February), WGWIDE (end May). Use of within year data is not priority.  Data quality checking is done, data availability is realtime. Include coarse description of data quality checks.
	Channel for data delivery (e.g. email, sharepoint)	
	Storage at the receiver's end (WMR): location, access rights, duration of storage, etc.	FRISBE (proposal)
	Allowed data use (e.g. only for the specific project purpose, is merging with other datasets without informing WMR allowed, etc.)	Points of attention: --measurement type: standard length, fork length or total length. Is registered. --what do we do if other signals come from the WMR data and the industry data? What is the risk? There is always a possibility when WMR hasn't received samples for a certain area. If spatial-temporal coverage is the same, no differences are to be expected.  Length sample data: for general use. Information of hauling locations and the catch size is under discussion.
	Dataset ownership	PFA is owner. If ownership is not with WMR, then it is hard to guarantee consistency for stock assessments. Martin to discuss in PFA.

<sup>5</sup> If a standard exchange format is available, name it here and use it in the project. Provide documentation on the format to the other parties. Check the datamanager if you're not sure if any standard format exists. Standard(ised) exchange formats are crucial when data from others has to be stored at WMR. Consistent exchange formats help re-use of existing processing code (R/SAS/...)

---

		<p>WKindustry will discuss similar topics about submitting to RDB directly.</p> <p>Can WMR use it for other research purposes than within this project?</p> <p>Pay back, information on use of results? If we arrange it in a contract</p>
--	--	--

## After the project, when others are storing WMR data or when WMR stores project data

To be discussed at the end of the project

	Question	Answer
	Will the data be <b>stored</b> <i>after</i> the end of the project?	The Netherlands is obliged to be able to resubmit the data if data is used for stock assessment.
If no:	Where will the data be <b>archived</b> ? <sup>6</sup>	Data remains stored in PFA achieves.
If yes:	Who is responsible for <b>safe and durable</b> storage? (project partner, institute, if relevant: contact person)	The data hasn't reached to the final level of accuracy throughout the project. Therefore this is not yet applicable
	Till <b>when</b> will the data be stored?	not applicable
	Who pays for the <b>safe and durable</b> storage? (NB mainly direct storage costs)	Not applicable
	Who should be <b>informed</b> when the storage end date is due?	Not applicable
	Where will the data be stored or archived?	PFA achieves.
	Will the data be <b>maintained</b> (e.g. correction of outliers etc., checking if data still exist, apply format shifts if relevant) <i>after</i> the end of the project?	Yes/No
If yes:	Who is responsible for <b>maintenance</b> ? (project partner, institute, if relevant: contact person)	Data needs further maintenance. This remains under responsibility of PFA. Further involvement of WMR yet to be negotiated.
	Till <b>when</b> will the data be maintained?	Not applicable
	Who pays for the <b>maintenance</b> ? (NB mainly personnel costs) (project partner, institute, if relevant: contact person)	Not applicable
	Does maintenance also include resubmission to external data portal(s)? If yes: which portal(s)? If yes: till when?	Not applicable

<sup>6</sup> Data have to be archived for a number of years, even when they don't have to be re-used.

## Appendix II

**Summary of the data available from the WMR market samples. Until 2004, the only available data from the WMR FRISBE database are the aged subsets.**

YEAR	N VESSELS	N STATIONS	N LF SAMP	N AGED SAMP	PERCENT IMMATURE	PERCENT MATURE	PERCENT FEMALES	PERCENT MALE
1990	5	28	706	706	0.3	99.7	30.9	68.8
1991	2	11	275	275	0.4	99.6	24.7	75.3
1992	2	12	300	300	20.3	79.7	41.3	58.7
1993	2	3	75	75	0	100	32	68
1994	2	10	250	250	25.2	74.8	48.4	51.6
1995	2	6	150	150	0	100	38	62
1996	2	9	225	224	10.3	89.7	44.6	55.4
1997	2	3	75	75	36	64	33.3	66.7
1998	2	6	150	150	16.7	83.3	39.3	60.7
1999	2	9	225	225	21.8	78.2	52.4	47.6
2000	1	1	25	25	4	96	44	56
2002	3	9	225	225	15.1	84.9	36.4	63.6
2003	2	8	194	194	1	99	44.8	55.2
2004	2	5	278	125	0	100	56.8	43.2
2005	2	14	747	350	0	100	60.3	39.7
2006	2	24	1805	600	3.5	96.5	55.5	44.5
2007	3	30	2223	750	2.8	97.2	58.8	41.2
2008	2	25	1716	604	0	100	51.8	48.2
2009	1	12	659	295	0.3	99.7	47.8	52.2
2011	1	1	61	25	8	92	68	32
2012	1	8	453	191	0	100	33	67
2013	2	8	560	190	1.1	99	46.8	53.2
2014	2	12	823	300	35.3	64.7	53.7	46.3
2015	2	12	755	295	0	100	31.9	68.1
2016	2	12	818	291	0	100	48.8	51.2
2017	2	11	1532	273	0	100	22.3	77.7
2018	2	10	808	232	1.3	98.7	39.2	60.8
2019	2	12	1079	299	3.3	96.7	36.1	63.9

---

# References

- Ailloud LE, Hoenig JM (2019) A general theory of age-length keys: combining the forward and inverse keys to estimate age composition from incomplete data. *ICES J Mar Sci*.
- Ailloud LE, Lauretta M V, Walter JF, Hoenig JM (2019) Estimating age composition for multiple years when there are gaps in the ageing data: the case of western Atlantic bluefin tuna. *ICES J Mar Sci*.
- Coggins Jr, L. G., Gwinn, D. C., & Allen, M. S. (2013). Evaluation of age-length key sample sizes required to estimate fish total mortality and growth. *Transactions of the American Fisheries Society*, 142(3), 832-840.
- Costello, C., D. Ovando, R. Hilborn, S. D. Gaines, O. Deschenes, and S. E. Lester. 2012. Status and solutions for the world's unassessed fisheries. *Science* **338**: 517-520.
- Fridriksson A (1934) On the calculation of age-distribution within a stock of cod by means of relatively few age-determinations as a key to measurements on a large scale. *Rapp Proces-Verbaux Des Reun Cons Int Pour l'Exploration La Mer* 86:1-5.
- Hilborn, R., E. A. Fulton, B. S. Green, K. Hartmann, S. R. Tracey, and R. A. Watson. 2015. When is a fishery sustainable? *Canadian Journal of Fisheries and Aquatic Sciences* **72**: 1433-1441.
- Hoenig JM, Hanumara RC, Heisey DM (2002) Generalizing double and triple sampling for repeated surveys and partial verification. *Biometrical J*.
- Murta AG, Loff JF, Neves M, Wise L (2016) ALKr: A R package of methods based on age-length keys to estimate the age structure of fish populations.

---

# Justification

ReportC098/21

Project Number: 4311100070

The scientific quality of this report has been peer reviewed by a colleague scientist and a member of the Management Team of Wageningen Marine Research

Approved: Gerben IJntema  
Researcher

Signature:



Date: 3 December 2021

Approved: Jakob Asjes  
MT Member Integration

Signature:



Date: 3 december 2021

---

Wageningen Marine Research  
T +31 (0)317 48 09 00  
E: [marine-research@wur.nl](mailto:marine-research@wur.nl)  
[www.wur.eu/marine-research](http://www.wur.eu/marine-research)

Visitors' address

- Ankerpark 27, 1781 AG Den Helder
- Korringaweg 7, 4401 NT Yerseke
- Haringkade 1, 1976 CP IJmuiden

---

With knowledge, independent scientific research and advice, **Wageningen Marine Research** substantially contributes to more sustainable and more careful management, use and protection of natural riches in marine, coastal and freshwater areas.

Wageningen Marine Research is part of Wageningen University & Research. Wageningen University & Research is the collaboration between Wageningen University and the Wageningen Research Foundation and its mission is: 'To explore the potential for improving the quality of life'

