**Standardization**

1. **Swedish Data**
   1. **Swedish historical data**

After loading the file in R I create an index for each haul.

Then I remove the pelagic trawls from the database.

After that I start creating a different dataset for each species of interest (cod, flounder, herring, sprat and plaice) removing the hauls with wrong information and adding the zero hauls.

* 1. **Fiskdata2**

First I load the files for the years until 2014 and the files for 2015-16 and merge them.

Then I change all the lengths from mm to cm and add the species names.

Then I remove the SD 21 because we are not interested in it and create a new index for each haul.

Then I merge the length frequency distributions with the index file.

Then I select only OTB otter trawl bottom and remove the Invalid and calibration hauls.

Then I remove all the hauls of TVL with sweeps equal to 25 and 50 because Anders says they are calibration hauls.

Then I remove from the index all the hauls without LFD.

I also remove all the hauls for which we know they have fished the species but we don’t have LFD.

Then I change the NAs in the Number per hours into 0 and create a dataset with only 0 catches.

* 1. **Swedish historical data + Fiskdata2**

First I change the names of the columns of my indices data in order to be able to merge them.

After merging them I fix the SDs and ICES rectangles using the shapefiles from ICES.

Then I change the names to my species files to merge them.

After merging each species separately I merge them all in one file and calculate the CPUE standardized by gear and speed.

Then I change the NAs in the standardized CPUE to 0 and fix the SD and ICES rectangles.

I remove the only 3 hauls that are in SD IIIa because we are not interested in that SD.

1. **DATRAS**

First I read the exchange file.

* 1. **HH**

First I put year as numeric.

Then I select only the valid, additional and no oxygen hauls and remove Sweden.

Then I fix SD and ICES rectangles.

There are 4 hauls on land. I remove the 2 from Denmark because I don’t know where they should be and I put the 2 from Poland in SD 26.

There are 7 hauls with no depth registered. I just put an approx. depth for one haul in SD 24, for one in 26 and 4 in SD 22 (the other record is from SD IIIa so I remove it later).

* 1. **HL**

First I put year as numeric, then I select only the hauls in HH.

Then I match SD, ICES rectangle, validity, Standard and bycatch species code.

Then I select the different species and add the 0 catches.

For cod I add the 0s and remove the lines with species validity (SpecVal) equal to 0.

For flounder and plaice I add the 0s, remove them if StdSpecRecCode !=1 and then remove the lines with species validity (SpecVal) equal to 0.

For herring and sprat I add the 0s, remove them if BycSpecRecCode !=1 and then remove the lines with species validity (SpecVal) equal to 0.

Then I create the CPUE unstandardized for SpecVal=1. If DataType = C then CPUEun = HLNoAtLngt, if DataType = R then CPUEun = HLNoAtLngt / (HaulDur / 60), if DataType = S then CPUEun = (HLNoAtLngt \* SubFactor) / (HaulDur / 60). If SpecVal = "zeroCatch" then CPUEun = 0, if SpecVal = 4 we didn’t decide (hauls with no LFD) and so I created a dataset with all those hauls.

Then I sum for the same haul the CPUE of the same length classes if they were sampled with different sub factors or with different sexes.

Then I remove SD IIIa and the sprat > 20 cm because they are errors.

1. **Individual data**
   1. **CA**

First I put year as numeric, then I select only the hauls in HH.

Then I do the matching of SD, rect and haul validity.

Then I select the different species and remove lines with missing individual weights or weight = 0.

I change the flounder of 347 cm to 34.7 and remove sprat >=20.

Then I check for outliers and remove them for cod, flounder and plaice.

Then I create mean condition per year and SD per different length classes to check for trends.

* 1. **Individual data from fiskdata2**

First I load the data for cod and flounder for the years until 2014 and the files for 2015-16 and merge them.

Then I change all the lengths from mm to cm and create a new index.

Then I matched latitude and longitude and removed invalid and calibration hauls.

Then I remove the lines with either length or weight = 0 or with NA in lat and long.

Then I match SD and ices rect.

Then I remove SD IIIa.

Then I check for outliers and remove them.

* 1. **Merge individual data for L/W relationships**

First I remove the year 2017.

Then I select only year, SD, length and weight and I merge all the individual data.

Then I check again for outliers.

Then I create the file for the L/W relationship of plaice **(NOTE!! Only DATRAS data without SWEDEN in the case of plaice)**

1. **Historical Latvian data**
   1. **HH**

First I load the data and create an index.

Then I select only valid, additional and no oxygen hauls.

Then I select the latitude and longitude and fix one haul position that seemed wrong.

Then I fix SD and ICES rect.

Then I remove the pelagic gear LPT.

There are 2 hauls with no depth so I put an approx. depth.

* 1. **HL**

First I load the data and create an index.

Then I select only the hauls in the HH file.

Then I match haul SD, rect, haul validity, StdSpecRecCode and BycSpecRecCode, haul duration and data type.

Then I create a different file for cod and for flounder.

For cod I add the 0s and remove the lines with species validity (SpecVal) equal to 0.

For flounder I add the 0s, remove them if StdSpecRecCode !=1 and then remove the lines with species validity (SpecVal) equal to 0.

Then I create the CPUE unstandardized for SpecVal=1. If DataType = C then CPUEun = HLNoAtLngt, if DataType = R then CPUEun = HLNoAtLngt / (HaulDur / 60), if DataType = S then CPUEun = (HLNoAtLngt \* SubFactor) / (HaulDur / 60). If SpecVal = "zeroCatch" then CPUEun = 0, if SpecVal = 4 we didn’t decide (hauls with no LFD) and so I created a dataset with all those hauls.

* 1. **CA**

First I load the data and create the index.

Then I select only the hauls in HH and match SD, rect and haul validity.

Then I select the different species and remove lines with missing individual weights or weight = 0.

Then I check for outliers and remove them.

Then I create mean condition per year and SD per different length classes to check for trends.

1. **Merge all the individual data for L/W relationships**

First I remove the year 2017.

Then I select only year, SD, length and weight and I merge all the individual data.

Then I check again for outliers.

Then I create mean condition per year and SD per different length classes to check for trends.

1. **L/W relationships**

First I transform length and weight in a Log scale.

Then I create an index for year and SD. In the case of cod I create also one for SD 22-23 only.

Then I run the L/W relationship using two different approach: the first is using one relationship for all the years and the other is using one relationship for each year.

For the cod in SDs 22-23 before 1992 I use the L/W relationship obtained by using the data of SDs 22-23 in 1992-1994.

For the cod in SDs 24-32 before 1976 I use the L/W relationship obtained by using the data of SDs 24-32 in 1976 and 1978-1979.

For the cod in SDs 24-32 in 1977 I use the L/W relationship obtained by using the data of SDs 24-32 in 1976 and 1978.

For the flounder in 1977-1978 I use the L/W relationship obtained by using the data in 1976 and 1979.

For the flounder in 1987 I use the L/W relationship obtained by using the data in 1986 and 1988.

For the flounder in 1991 I use the L/W relationship obtained by using the data in 1990 and 1992.

For the plaice before 1994 I use the L/W relationship obtained by using the data in 1994-1996.

Then I transform all the coefficients to a linear scale and create a data frame with all the coefficients.

Then I calculate the estimated weight for each length class and convert the CPUE in biomass.

I do this for all the different datasets (DATRAS, Swedish data and Latvian data).

1. **Standardization of gear and speed**

First I create a column for date in the HH files.

Then I read the files with gear and speed info.

* 1. **Gear assumptions**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Country** | **Gear** | **Fishing line** | **Sweep length** |  |  |  |  |
| DEN | CAM |  |  |  |  | Not standardized | |
| DEN | EGY |  | 60 |  |  | Data from DATRAS | |
| DEN | EXP |  | 110? |  |  | Data from expert | |
| DEN | GRT |  | 60/110? |  |  | Data in DATRAS not for all the years | |
| DEN | TVL | 63.46 | 75 |  |  | Pelagic gears removed | |
| DEN | TVS | 33.22 | 60? |  |  |  |  |
| EST | ESB |  |  |  |  |  |  |
| EST | TVS | 33.22 | 75 |  |  |  |  |
| GFR | CHP | 35.2 | 53.5 |  |  |  |  |
| GFR | H20 | 36 | 60 |  |  |  |  |
| GFR | SON | 16.54 | 53.5 |  |  |  |  |
| GFR | TVS | 33.22 | 75 |  |  |  |  |
| LAT | DT |  |  |  |  |  |  |
| LAT | LBT | 28 | 82/100 |  |  |  |  |
| LAT | LPT |  |  |  |  |  |  |
| LAT | TVL | 63.46 | 75 |  |  |  |  |
| LAT | TVS | 33.22 | 75 |  |  |  |  |
| LTU | TVS | 33.22 | 75 |  |  |  |  |
| POL | P20 | 39.8 | 104.5 |  |  |  |  |
| POL | PEL | 53.1 | 89.5 |  |  |  |  |
| POL | TVL | 63.46 | 75 |  |  |  |  |
| RUS | DT |  |  |  |  |  |  |
| RUS | HAK |  |  |  |  |  |  |
| RUS | TVL | 63.46 | 75 |  |  |  |  |

* 1. **Speed assumptions**

For Denmark: I used the mean speed of 1996 (2.8) for the years 1991-1996.

For 1999 TVS I used the mean TVS speed in 1999 (2.7).

For 1999 TVL I used the mean TVL speed in general (3).

For 2000-2002, 2008-2009 I used for each gear the mean speed for the same gear in the same year.

For 2010 I have changed a speed of 0.1 to the mean speed of the same gear and year.

For Estonia: I changed all the NAs with the mean values of the speed present in the data independently by the year.

For Germany: For 1991 I have used the mean speed of all the gears used in 1991.

For 1992, 1995-1996, 1999 and 2014 I used for each gear the mean speed for the same gear in the same year.

For Latvia: We have the speed data apart from two hauls (I used for each gear the mean speed for the same gear in the same year).

From 2004 they report always a speed of 3.

For Lithuania: I used the mean speed... they are always reporting=3

For Poland: I have put 3 the speed for tvl since they are always reporting it as 3 but we don't have any info for the other gears (I have put 3 for everything).

In 2015 Q1 I used the mean speed of 2015 Q4 and for 2016 I used the mean speed of each quarter.

For Russia: I have put in 1994 the mean speed of 1993 and 1995 for the same gear and boat.

* 1. **Standardization**

I match all the missing information and calculate the standardized CPUE.

Then I merge all the datasets for each species.

Then I remove all the hauls with no gear standardization.

For herring there are a few records in the Swedish data that have length but no numbers so I remove them.

Then I remove all the pelagic gears.

Then I create a data frame with the haul positions.

And then I merge all the species.

1. **Flounder with no LFD**

First I match missing information to the flounder data without LFD.

Then I merge the flounder datasets with no LFD.

Then I take only SDs from 24 to 28-2 because are the one that I use in my analyses.

Then I create a file with the CPUE for flounder <20 cm, >20 cm and the total CPUE.

After fixing the names of the variables I merge the flounder LFD data with the one without LFD.