



Regional Coordination Group
on Economic Issues

REPORT OF THE WORKSHOP ON AN ALTERNATIVE APPROACH TO THE SEGMENTATION OF FISHING FLEETS

Edited by Erik Sulanke and Jörg Berkenhagen

Workshop organized and hosted by:

Thünen Institute of Sea Fisheries, Bremerhaven, Germany, part of

Johann Heinrich von Thünen Institute

Federal Research Institute for

Rural Areas, Forestry and Fisheries

29 – 31 March 2021

Online Workshop

Table of contents

Introduction	3
Executive summary	3
Terms of Reference	4
Background	4
Statistical Framework	5
ToR 1: Test the user-friendliness, clarity of documentation, and applicability to national	6
ToR 2: Apply the newly developed approach to national fleet data sets and compare the	6
ToR 3: Identify perspectives and possible additions to the approach and the FS-package	7
Annex 1 – List of participants	8
Annex 2 – Workshop agenda	9
Annex 3 – Package manual	10
Annex 4.1 – Case studies - Finland	30
Annex 4.2 – Case studies - Iceland	39
Annex 4.3 – Case studies - Italy	46
Annex 4.4 – Case studies - Netherlands	65
Annex 4.5 – Case studies - Portugal	79
Annex 4.6 – Case studies - Spain	97

Introduction

The workshop on a newly developed, alternative approach to the segmentation of fishing fleets took place from 29th to 31st of March 2021. Due to travel restrictions caused by the global Covid-19 pandemic, the workshop took place online.

The workshop was hosted by the Thünen Institute of Sea Fisheries, part of the Johann Heinrich von Thünen Institute, Germany's Federal Research Institute for rural areas, forestry, and fisheries.

34 experts representing 15 nations, the ICES, and DG MARE, participated in the workshop. 13 national fisheries data sets were analyzed. The List of participants can be found in Annex 1. Erik Sulanke and Jörg Berkenhagen chaired the workshop. The agenda can be found in Annex 2.

Executive summary

A new approach to the segmentation of fishing fleets was developed in a DCF pilot project at the Thünen Institute of Sea Fisheries and transferred to an R package, referred to as "FS-package" in the following. To introduce the package to the scientific community and identify possibilities for further improvement, the developers of the approach organized a workshop following the ToRs:

1. Test the user-friendliness and clarity of documentation for non-experts in the R software and the applicability of the approach and its functions for data preparations to varying national datasets
2. Apply the approach to national fisheries data sets of the participants and test whether they correspond with expert knowledge on the fleet structure.
3. Identify perspectives for extension or amendment of the functions and analyses implemented in the R package to foster its applicability and the quality of its results.

The new approach on fleet segmentation is based on hierarchical agglomerative clustering (HAC) of a catch composition matrix. All functions needed to apply the approach to a fishing fleet data set were included in the FS-package and their usage was illustrated in a manual. The participants rated the user-friendliness of the package documentation and the data preparation process, which was assessed in ToR 1, mostly positive and only minor adjustments had to be made. To account for the overseas pelagic fleets of some participant's nations, stock definitions of the ICCAT were integrated into the stock assigning function of the package.

Regarding ToR 2, participants successfully applied the FS-package to their national fleet datasets and 6 fully documented case studies were produced, which can be found in Annex 4.1-4.6. All authors highlighted the usefulness of the package to identify major fishing patterns, but they all identified possibilities for further improvement. Most importantly, the identification of the appropriate number of clusters to use became very hard in large (> 500 vessels) and heterogenic datasets. Furthermore, the clusters of such datasets often comprised of only a small number of vessels. This would impede data collection and reporting. These findings highlight the necessity of

adequate pre-separation of the data set. In addition, the maintainers of the package aim to improve the applied distance measure of the clustering algorithm and introduce a minimal cluster size.

In the concluding phase of the workshop, the participants identified perspectives for improvement of the new approach in ToR 3. The importance of a reasonable pre-separation of the data set by meaningful fishing gear classification was highlighted. Therefore, a more variable and robust gear classification scheme than the method currently used by the DCF has to be developed and tested. In addition, the participants highlighted the importance of high-value target species for fishing patterns, which are often underrepresented in input matrices based on catch weight. In order to solve this, an adequate approach to consider catch value has to be included in the FS-package. The workshop participants and organizers also agreed on the importance of regional sub-groups to identify and characterize regionally important fisheries to ensure consistency of fisheries-based fleet segmentation approaches. Since the overarching aim of such fleet segmentation approaches is to group vessels with similar cost structures, cost data should be used to validate newly formed fleet segments whenever possible.

Terms of Reference

The overarching aim of the workshop was to test the newly developed approach on the segmentation of fishing fleets created in a DFC-pilot project of the Thünen-Institute between September 2019 and December 2021. After finalizing the statistical framework of the approach and transferring it into a user-friendly package of the R-Software, it was tested in bilateral cooperation with several partner institutions. The results of this preceding testing phase and the remaining knowledge and experience gaps were included in the following ToRs:

- 1. Testing the user-friendliness and clarity of documentation for non-experts in the R software and the applicability of the approach and its functions for data preparations to varying national datasets
- 2. Apply the approach to national fisheries' data sets of the participants and test whether they correspond with expert knowledge on the fleet structure.
- 3. Identify perspectives for extension or amendment of the functions and analyses implemented in the R package to foster its applicability and the quality of its results.

Background

Under the Data Collection Framework (DCF), fleet economic data of European fishing fleets have to be provided by fleet segment. The current fleet segmentation scheme is based on the vessel length class and the main fishing gear, which are technical parameters of the vessels. This segmentation method is well defined and easily applicable, but it does not adequately represent target fisheries. In addition, the length classes used for the classification often fail to group vessels of similar operational scales properly. Vessels with similar technical parameters are often active in varying fisheries that differ in terms of catch composition, fishing activity, and cost structure.

This has also been pointed out at the 2019 PGECON meeting, where a workshop on the issue was recommended.

In order to improve reporting with respect to individual target fisheries, a transferable, systematic approach based on multivariate statistics methods was developed in the pilot project 'Fleet Segmentation' at the Thünen Institute of Sea Fisheries in Germany, between September 2019 and December 2020. The statistical framework was developed and transferred to a user-friendly function package for the R statistics software, which was tested in cooperation with several partners during and after the developmental process.

Statistical Framework

The newly developed fleet segmentation approach is based on the catch composition of the fishing vessels within one reference year. Catches were aggregated on the stock level following ICES, GFCM, and ICCAT stock descriptions based on species and fishing areas. Each vessel's stock-based catch weights are expressed as a proportion of the total catch, scaled to values between 0 and 1, and then transformed into a Euclidean distance matrix. A separate matrix for each fishing gear class is computed. The main fishing gear is identified following the DCF-procedure, i.e., a gear being used for more than 50% of a vessel's fishing time was selected as the vessel's main gear. The matrices provided are the basis of a 'hierarchical agglomerative clustering (HAC) procedure, for which the UPGMA (unweighted pair group method with arithmetic mean) fusion algorithm is used. This fusion algorithm outperformed other standard HAC fusion algorithms in the development process. The performance was tested by comparing the cophonetic correlation of the clustering results of the respective fusion algorithms with the original distance matrix. An array of specific indices, tests, and visual validation methods is applied to determine the optimal number of clusters, of which the Average Silhouette Coefficient, the Mantel test, the dendrogram, and the clustering tree are the most important. The procedure is finalized by a post-hoc validation of the clustering result to identify the actual alternative fleet segments based on similarities in catch composition, vessel size, and expert knowledge. For a detailed description of the application of the R package, see the corresponding manual in Annex 3.

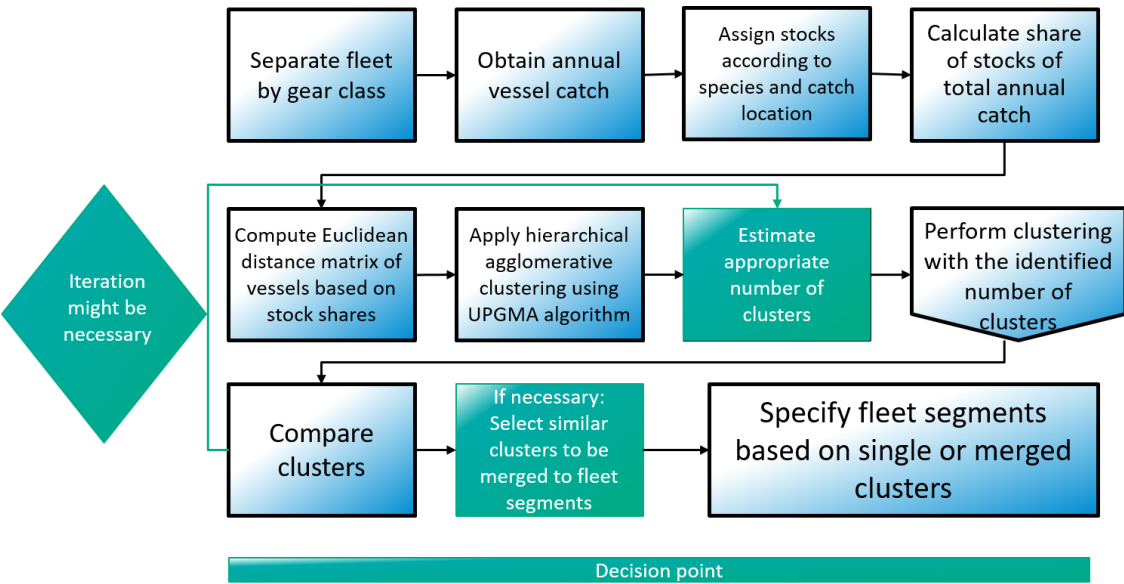


Figure 1: Stepwise flowchart of the newly developed fleet segmentation approach.

ToR 1: Test the user-friendliness, clarity of documentation, and applicability to national datasets regarding data preparation

Following the detailed manual and instructions given in preparation for the workshop, the participants generally encountered none or minor difficulties installing the functions of the package. However, the workshop made clear that installing an R package from a .zip-file, as it is currently necessary for the FS-package, is a time-consuming procedure that can deter non-users of the software. Therefore, an important next step for the near future is to publish the package on a commonly used online software library (CRAN, GitHub, or GitLab), from which it can be installed, maintained, and updated more easily.

In most cases, the participants could successfully prepare national data sets with the documentation given. Nevertheless, the maintainers of the package are considering simplifying the functions used for data preparation in the package to avoid errors. The data preparation with the workshop participants revealed two further weaknesses in the package, which could be resolved during the workshop. Firstly, the package version initially used in the workshop contained only stock information of the ICES and the GFCM. In order to account for vessels operating overseas, particularly those of the Spanish fleet, stock definitions of the ICCAT were integrated into the stock assignment function of the FS-package. Secondly, a catch-weight threshold for bycatch species is integrated into the FS-package. Due to coding errors, the threshold was too restrictive and assigned many target species as "bycatch", especially in small-scale coastal fisheries data sets with overall low catch weights. The coding error was corrected, and the default threshold was lowered to make it less conservative. The user can adjust this threshold when applying the package to ensure the conditions of the analyzed fishing fleet are met.

ToR 2 Apply the newly developed approach to national fleet data sets and compare the results to expert knowledge

The newly developed approach was applied to various national datasets in the workshop, of which 6 case studies were fully documented and are available in the Annex (4.1-4.6). The general feedback of the participants was positive. Using the functions of the FS-package enabled the users to identify and characterize the significant fishing patterns present in their national fleet data sets and to detect newly developed or minor, undescribed fishing patterns. Limitations of the new approach became visible in very large (approximately more than 500 vessels) and heterogenic data sets. The clustering procedure tended to produce numerous clusters containing only a few vessels and a few large clusters containing the majority of the vessels and having a very heterogenic catch composition. Some participants suggested including a minimal cluster size as a systematic controlling feature in the clustering algorithm. In addition, the maintainers of the package presume that a change of the applied Euclidian distance to a metric transformation of the Bray-Curtis index could help resolve this problem. The maintainers will review both possible solutions.

ToR 3 Identify perspectives and possible additions to the approach and the FS-package

As mentioned before, the newly developed approach to fleet segmentation is based on hierarchical agglomerative clustering. Clustering is a data-driven approach; therefore, the quality of the clustering result and the precision of indicators and tests is reduced by large, highly heterogenic datasets. Reasonable pre-separation of the input data has been proven crucial for a valid result and was recommended to be implemented based on the DFC gear classes of the analyzed national fleets. However, the DCF gear classes are assigned based on a relatively conservative procedure, in which a vessel's main gear is classified according to the share it has on a vessel's total operating time. A gear used in more than 50% of the operations is considered the respective vessel's main gear. This method is straightforward and applicable, yet it fails to account for characteristic operational strategies of fisheries such as seasonal gear changes or gears used at nearly equal percentages of a vessel's sea time. Therefore, the participants deemed a revision of the gear classification process to be necessary. The maintainers consider the clustering approach used in the package to be feasible for a gear classification with minor adjustments. Participants of the workshop also highlighted the applicability of the R-package "VMSbase", which was developed for the analysis of métier-based fisheries data. The maintainers will review this option and integrate a comprehensive approach for gear classification in bilateral cooperation with workshop participants.

In most of the case studies carried out at the workshop, fisheries were identified, in which a species that had only a small share on the catch weight but a significant share on the catch value due to its high price. Such high-value species are often drivers of fisher behavior, yet they are under-represented in an approach based on a catch-weight matrix. Therefore, the input data of the FS-package should be replaced or, most ideally, extended by data on the catch value. The maintainers of the package will redraw the data input in a suitable way accounting for this issue.

In all data sets analyzed with the newly developed approach, clusters of only a few vessels with very distinct catch composition were present. Despite representing unique fisheries, such fleet segments contain too few vessels to collect and report data based on those segments. Those fleet segments have to be joined with other segments into larger groups of vessels based on the similarity of the fisheries. In order to do so, a standardized toolset to analyze and define similarity of fishing patterns has to be defined and integrated into the FS-package. Specific fishing patterns present in a particular region should ideally be identified and characterized by sub-groups at a regional level.

The overarching aim of subdividing a fishing fleet into fleet segments is to group vessels with similar cost structures. In fisheries, individual vessel cost data is typically too scarce and too heterogenic to base a fleet segmentation procedure on it. Nonetheless, the results of the alternative fleet segmentation approach should be examined utilizing available cost data sets and derived cost-effort and cost-effectiveness indices. By those means, an actual improvement of the collection framework and advantages for all involved parties can be guaranteed.

Annex 1 – List of participants

Name	Surname	Country/Organisation
Erik	Sulanke	Germany (co-chair)
Jörg	Berkenhagen	Germany (co-chair)
Katrien	Verlé	Belgium
Ivana	Vukov	Croatia
Myrto	Ioannou	Cyprus
Josefine	Egekvist	Denmark
Monika	Sterczewska	EU
Antti	Takolander	Finland
Marie-Dominique	Minne	France
Sébastien	Demaneche	France
Steffi	Meyer	Germany
Elzbieta	Baranowska	Iceland
Sandra	Rybicki	Iceland
Julie	Kellner	ICES
Paolo	Accadia	Italy
Rosaria	Sabatella	Italy
Monica	Gambino	Italy
Jurgen	Mifsud	Malta
Eric	Muscat	Malta
Hans	van Oostenbrugge	Netherlands
Ana Cláudia	Fernandes	Portugal
Suzana	Cano	Portugal
Hugo	Diogo	Portugal
Luís	Costa	Portugal
Francisco Manuel	Fernández Martínez	Spain
Maria	Valiente Viana	Spain
Hanna	Swahnberg	Sweden

Workshop on a novel approach to the segmentation of fishing fleets



Agenda*

* all times are in **CEST** and might be changed in the progression of the workshop

Monday

10:00	Welcome and housekeeping
10:30	Introduction to the current fleet segmentation system and its issues with the German fishing fleet (Dr. Jörg Berkenhagen)
11:00	Discussion on the current fleet segmentation in the participants fishing fleets
11:30	Introduction to the new approach on fleet segmentation (Erik Sulanke)
12:00	Lunch break
13:00	Data preparation, package installation, R workflows. Eventually first package trials.
17:00	End of the first day

Tuesday

10:00	Getting started and last questions before data work
10:30	Full day of data work. Erik Sulanke and Dr. Jörg Berkenhagen will be available for questions and programming aid.
12:00	Lunch break
13:00	Data work, pt.II
16:30	Impressions from data work
17:00	End of the second day

Wednesday

10:00	Summary and accomplishments of data work day (Erik Sulanke)
10:30	Segmentation of the Belgian fleet – Experience presentation (Katrien Verlé)
11:00	Participants start presenting their results
12:00	Lunch break
13:00	Experience presentation, pt. II
15:00	Key results and implications for the future
16:00	End of workshop

Fleet Segmentation - Package Manual *

Erik Sulanke *Thuenen-Institute of Sea Fisheries, Bremerhaven, Germany*

This is the manual for the *FleetSegmentation*-package (FS-package) developed by Erik Sulanke of the Thuenen-Institut of Sea Fisheries in Bremerhaven, Germany. The package contains ready-to-use functions to apply a multivariate approach on the segmentation of fishing fleets. It is based on hierarchical agglomerative clustering (HAC) of vessel catch data. Functions for diagnostic plots, which enable the user to characterize fleet segments, are included. In the manual, the use of the package will be demonstrated with example data, which is also included in the package.

Keywords: Fishing fleet, segmentation, clustering

Introduction

The appropriate segmentation of fishing fleets is controversially discussed in fisheries research and management and a variety of approaches has been introduced. In the current fisheries data management of the STECF and ICES, two major approaches continue to exist side by side. In the economic data collection and the annual economic report, fishing vessels are classified by gear class and vessel length. This is problematic, not only because of the selection of length classes for vessels but also because no information on the fishing strategy is included. The second approach used mostly in biological research uses so-called fisheries metiers, which include information on the main gear, vessel length, mesh width and target assemblage. This is again problematic because the target assemblage contains no information on the stocks used and how mixed or targeted a fishery is. The present approach, developed in a pilot study funded by the DCF, introduces a standardised multivariate approach for characterising fisheries fleet segments by hierarchical agglomerative cluster analysis (HAC) of their catch composition.

1. Getting started

1.1 Loading the package

The FS-package includes all the functions needed to perform the multivariate approach and analyse the results. It loads all packages necessary for its use automatically, but unfortunately, those packages can't be installed automatically when the FS-package is installed from .zip. A code block for the installation of all necessary packages can be found in the appendix (see Appendix - *Package installation*). It is planned to make this package accessible on *GitHub* and automate the process. Make sure to update your R and R Studio to the latest version and install/update all other necessary other packages. Once you are ready, load the FS-package from the library.

*This is a prototype version of the package. Critique and suggestions are highly welcome! **Current version:** July 03, 2021; **Corresponding author:** erik.sulanke@thuenen.de

```
library(FleetSegmentation)
```

The package uses functions from 33 other packages, so loading it might take moment. Some warnings might be shown because of package conflicts about which you do not have to worry.

1.2 Preparing & loading fleet data

Now it is time to prepare and load your fleet data. Please keep in mind, that clustering is a data-driven procedure, so the structure and size of the data frame determines the clustering result and especially the estimates of how many clusters are present in the data. Therefore, it is strongly advised to **pre-separate your data by gear class** into for example demersal trawlers, pelagic trawlers and passive gearers. If a notable part of the fleet is using mixed gears, classify them as a **mixed gear** fleet. Prepare the required data frames for all gear classes and run the segmentation procedure for all of them. This will most likely improve the quality of the clustering. The functions of the FS-package are essentially based on two input data frames: One data frame containing the catch of all vessels of a certain gear class and one data frame containing the length of those vessels. First, we will load the example catch data included in the FS-package to illustrate the required data frame structure. We will store it in a data frame called `catchdata`.

```
data <- example_catchdata
glimpse(data)

## Rows: 287
## Columns: 4
## $ ship_ID <chr> "ship_1", "ship_1", "ship_1", "ship_1", "ship_2", "ship_2", "~
## $ species <chr> "PLE", "SOL", "COD", "TUR", "PLE", "SOL", "COD", "PLE", "SOL"~
## $ fao_area <chr> "27.4.b", "27.4.b", "27.4.b", "27.4.b", "27.4.b", "27.4.b", "~
## $ landings <dbl> 65321.00, 17565.00, 12676.00, 3201.00, 72399.00, 12535.00, 16~
```

As you see, this data frame contains four columns.

1. A character column containing a unique vessel identifier. In the example data frame, it is `ship_[running nr.]`
2. A character column containing a species identifier. Please make sure to use the official ICES three-letter species identifier.
3. A character column containing the FAO area code. Please make sure to use the official and full area codes.
4. A numeric column containing the total catch per species and vessel in one year. The catch should be given in kilograms.

The dataframe is aggregated to the species- and FAO-area level, hence one row of the dataframe resembles the total catch of one vessel for one species in one FAO-area.

Next, we load the vessel length data.

```
shiplength <- example_lengthdata
glimpse(shiplength)

## Rows: 71
## Columns: 2
## $ ship_ID    <chr> "ship_1", "ship_2", "ship_3", "ship_14", "ship_15", "ship_1~
## $ shiplength <int> 16, 19, 13, 22, 15, 18, 21, 19, 22, 21, 18, 23, 17, 20, 14,~
```

This data frame contains two columns.

- 1. A character column containing a unique vessel identifier. In the example data frame it is `ship_[running nr.]`
- 2. A numeric column containing the length of the ship. It can be given in `cm` or `m`, dependent functions will auto-transform values.

1.3 Assign ICES-stocks to data

The clustering procedure’s quality and identification of specific fisheries are considerably better when spatially defined `stocks` instead of species are used as the clustering variable. Therefore, the package contains a build-in function to assign ICES and GFCM stocks based on the target species and the catch area. If the fisheries represented by your data are predominantly operating in the Northeast Atlantic or the Mediterranean, you can use this function to assign stocks to your catch data relatively quickly, as you can see in the example.

```
data_stocks <- assign_stocks(data = data)
glimpse(data_stocks)

## Rows: 287
## Columns: 3
## $ ship_ID    <chr> "ship_1", "ship_1", "ship_1", "ship_1", "ship_10", "ship_11",~
## $ stock      <chr> "cod.27.47d20", "ple.27.420", "sol.27.4", "tur.27.4", "pra.27~
## $ landings   <dbl> 12676.00, 65321.00, 17565.00, 3201.00, 197973.00, 193028.12, ~
```

As you can see, the newly created data frame has only three columns: The vessel identifier, the stock identifier, and the total catch per stock and vessel. Yet, not all species fished in the Northeast Atlantic or the Mediterranean are managed under a specific stock. In such cases, the function will automatically assign stocks to species caught in a reasonable volume (more than 10t in total, adjustable by the argument `threshold.auto.generate`) based on the species and the area identifiers. In case of FAO areas, only the major area (e.g., 27.4 for the North Sea) will be considered. You can switch off this function setting the argument `auto.generate` to `FALSE`, though it is highly recommended to leave it activated. Suppose the fisheries you are analyzing with the package are not operating in the areas covered by the function but are managed by a stock system. In that case, you have to prepare your data structured analogously to the data frame created by the `assign_stocks` function, i.e., with the three columns vessel identifier, stock identifier, and total catch (in kg) per stock and

vessel. If you don’t have access to spatial information or if there is no stock-management system applicable to your data set, you can still use the fleet segmentation approach with the catch per species as an input variable, even though the result of the clustering will most likely be harder to interpret. Just make sure to use the necessary data frame structure: Three columns, first the vessel identifier, then the species identifier, and then the total catch per species and vessel. In our case, the example data resembles a fishery operating in the area covered by ICES, and therefore, ICES-stocks were assigned to the example data by the applied function. Next up, we transform the data to a format suitable for the clustering procedure.

2. Transforme the data

Prior to the clustering, we have to transform the data into a suitable format for clustering and store the transformed data in a new data frame. Therefore, we use the `catchdata_transformation()` function and call the transformed data frame `catchdata`.

```
catchdata <- catchdata_transformation(data = data_stocks)
glimpse(catchdata)

## Rows: 71
## Columns: 18
## $ bwq.27.2425      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ‘cod.27.24-32’   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ cod.27.47d20     <dbl> 0.1283477, 0.0000000, 0.0000000, 0.0000000, 0.000~
## $ ‘ghl.27.1-2’     <dbl> 0.00000000, 0.00000000, 0.00000000, 0.00000000, 0~
## $ ‘her.27.1-24a514a’ <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ‘her.27.25-2932’ <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ‘hke.27.3a46-8abd’ <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ‘lin.27.3a4a6-91214’ <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ‘ple.27.24-32’   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ple.27.420       <dbl> 0.6613914, 0.0000000, 0.0000000, 0.0000000, 0.000~
## $ pok.27.3a46      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ pol.27.3a4       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ‘pra.27.1-2’     <dbl> 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ‘reb.27.1-2’     <dbl> 0.00000000, 0.00000000, 0.00000000, 0.00000000, 0.000~
## $ reb.27.5a14      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ rng.27.1245a8914ab <dbl> 0.00000000, 0.00000000, 0.00000000, 0.00000000, 0~
## $ sol.27.4         <dbl> 0.1778500, 0.0000000, 0.0000000, 0.0000000, 0.000~
## $ tur.27.4         <dbl> 0.03241092, 0.00000000, 0.00000000, 0.00000000, 0~
```

The resulting data frame is a matrix with each stock as a column and each vessel as a row. The catch values are transformed to their relative proportion of a vessels total catch, scaling between 0 and 1.

3. Examine the best number of clusters to use

This is one of the most important, yet most challenging steps of a clustering procedure. There is not one definite index or method to examine the **correct** number of clusters in a dataset. Instead, various indices and tests are used, their results compared and a heuristic decision on the number of clusters made. The multiple clustering procedures using a different number of clusters might have to be repeated to get a final clustering result.

3.1. Tabelize index results

First, five essential indices estimating the best number of clusters are computed. The result can be printed in a basic or an HTML-version or stored in a data frame. The maximum number of clusters to be expected is automatically calculated from the data, up to a maximum of 15 clusters. If you expect your dataset to contain more than 15 clusters, use the `max_clusternumber` argument in the function.

```
# print base format
numberclust_table(catchdata = catchdata)

##           Indices Optimal number of clusters Index value
## 1   Average silhouettes              7         0.852
## 2         Mantel test                7         0.950
## 3   Davis-Bouldin index              8         0.190
## 4         SD index                  7         2.609
## 5 Calinski-Harabasz index           15       2734.682

# store base format as data frame
index_table <- numberclust_table(catchdata = catchdata)

# print html-format
numberclust_table(catchdata = catchdata, style = "html")
```

This pdf-manual was created in R markdown, therefore the HTML-table can't be produced. The HTML-table will be accessible in the R-Studio viewer-tab and can be saved as an image or HTML-file and be included in HTML-formated R-Markdown documents. As you see, most indices suggest a number of 7 to 8 clusters to be present in the data, except for the Calinski-Harabasz-index, which suggests 15 clusters to be present. These indices use different methods for calculating the ideal number of clusters and are likely to produce quite heterogenic results. Average silhouette score and the mantel test have proven to be the most robust for the type of data used for the clustering. They can be compared graphically using the `numberclust_plot` function of the FS-package, along with two other important diagnostic plots: The scree plot and a dendrogram of the clustering.

3.2. Plot clustering diagnostics

```
numberclust_plot(catchdata = catchdata)
```

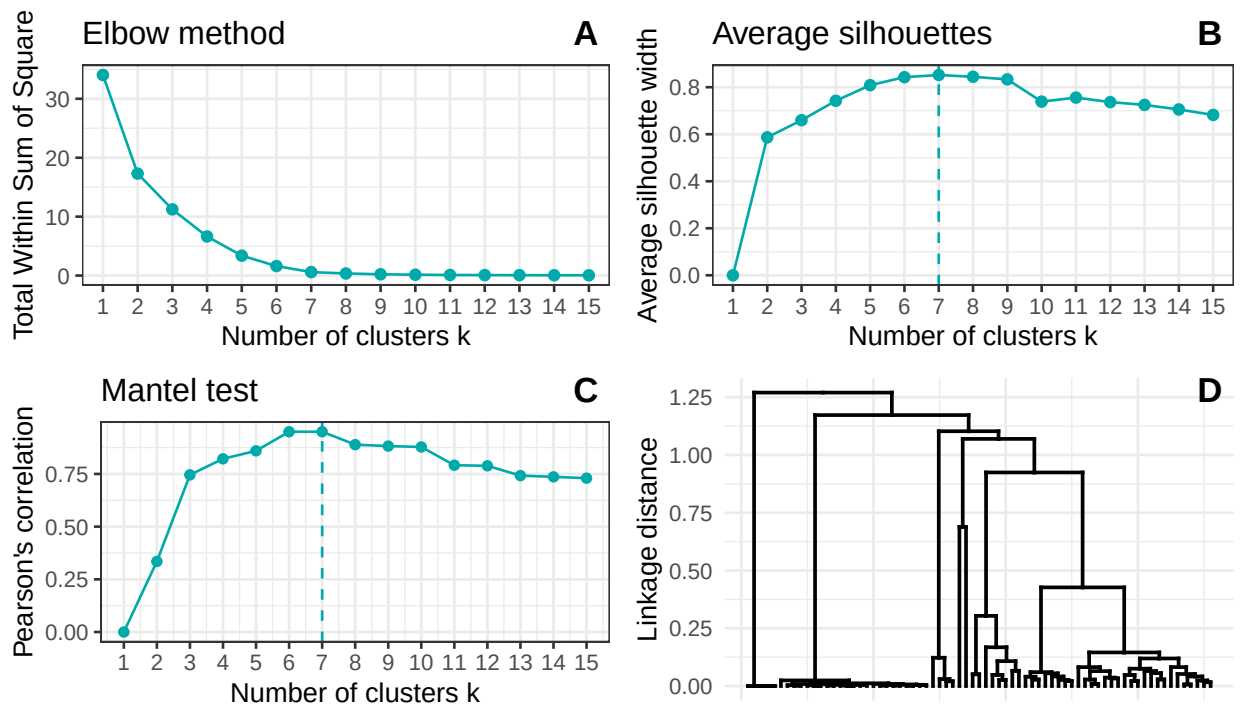


Figure 1: Gridded diagnostic plots of the HAC procedure. A) Scree plot showing the total within sums of squares of the clustering vs. the number of clusters used. B) Average silhouette width of the clusters vs. the number of clusters. C) Mantel test, i.e. the Pearson correlation between the clustering and the original distance matrix vs. the number of clusters. In B) and C), the best score is marked with a vertical line. D) dendrogram of the HAC procedure with the linkage distance of the clusters on the y-axis.

The scree plot on the upper left depicts the total within sums of squares of the clustering data. It is checked for an “elbow” in the curve, i.e. the clustering step with the steepest decline. This position is usually less pronounced in the clustering approach used than in e.g. k-means clustering of environmental data. Yet, the **transition zone** of the curve, in which the slope decreases from steep to nearly zero, gives a good insight into how many clusters are to expect in the data set. In the case of the example data, seven to eight clusters seem to be ideal. This is supported by the average silhouette width, which is high for six to nine clusters and at a maximum at seven clusters and shows a steep decline for ten clusters and forth; and by the mantel test, which reveals a high Pearson correlation between the clustering and the original data for six to seven clusters, with a maximum at seven. The visual examination of the dendrogram is another way to estimate the appropriate number of clusters in the data. The FS-package contains a function to plot only the dendrogram and manipulate the cutting height in different ways, one for cutting at a specific height and one

for cutting at a range of heights. We will plot these options aligned with the `ggarrange()`-function of the `ggpubr`-package. To tidy up the grid, the y-axis label of the second and third dendrogram will be removed. These dendrograms are produced by the `ggplot`-grammar and can be manipulated with `ggplot`-functions (and all functions based on the `ggplot`-grammar).

```
# basic dendrogram
dend1 <- numberclust_dendrogram(catchdata = catchdata)
# dendrogram with specified cutting height at 0.75
dend2 <- numberclust_dendrogram(catchdata = catchdata,
                                dend_method = "cut",
                                dend_cut = .75) +
  theme(axis.title.y = element_blank())
# dendrogram with a range of cutting heights from 0.75 to 1
dend3 <- numberclust_dendrogram(catchdata = catchdata,
                                dend_method = "range",
                                range_min = .5,
                                range_max = 1)+
  theme(axis.title.y = element_blank())

# grid the dendrograms
ggarrange(dend1,dend2,dend3, nrow = 1,align = "h",
          labels = c("A","B","C"), label.x = .9)
```

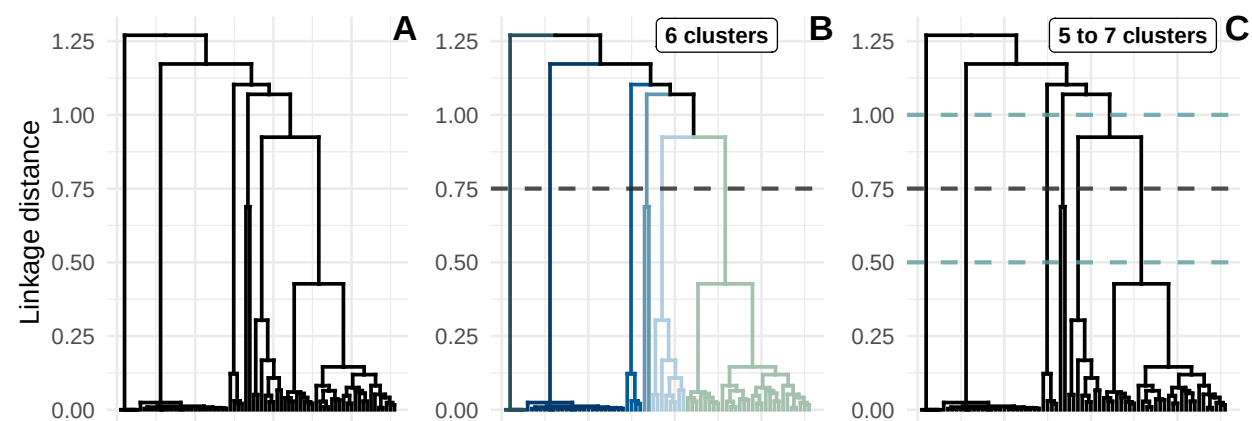


Figure 2: Grided plot of three dendrograms of the HAC procedure. The y-axis shows the linkage distance of clusters. A) is unmodified, B) is cut at a linkage distance of 0.75 and resulting branches are individually coloured and the resulting number of clusters is labelled, C) has cutting lines at linkages distance 0.5, 0.75 and 1 and the range of cluster numbers resulting from those cutting heights is labelled.

As you see, the dendrograms support the conclusion that five to seven distinct clusters are present in the data. Note that 0.5 to 1 is a very wide range of distances and in a more heterogenic dataset can return a very wide range of clusters. In this case, it can be helpful

to narrow down the range of cutting heights using the `range_min` and `range_max` arguments. In general, the more heterogenic and overlapping a dataset is, the harder it is to interpret a dendrogram. In this case, rely on the clustering indices and other diagnostic plots to estimate the appropriate number of clusters.

3.3 Plot clustering tree

Another helpful diagnostic plot for estimating the best number of clusters to use is a clustering tree. This tree plot visualises the proceeding of the clustering splits and the resulting sizes of clusters. It can be used to determine the steps in which significant splits of clusters occur and differentiate between those steps and **over-clustering**, where only small sub-clusters of single ships are produced. The function is a modification of the `clustree()` function from the eponymous package. As the indices and dendrograms already showed, that the data set is unlikely to contain more than 10 clusters, we will adjust this argument when plotting the clustering tree.

```
numberclust_clustree(catchdata = catchdata,max_clusternumber = 10)
```

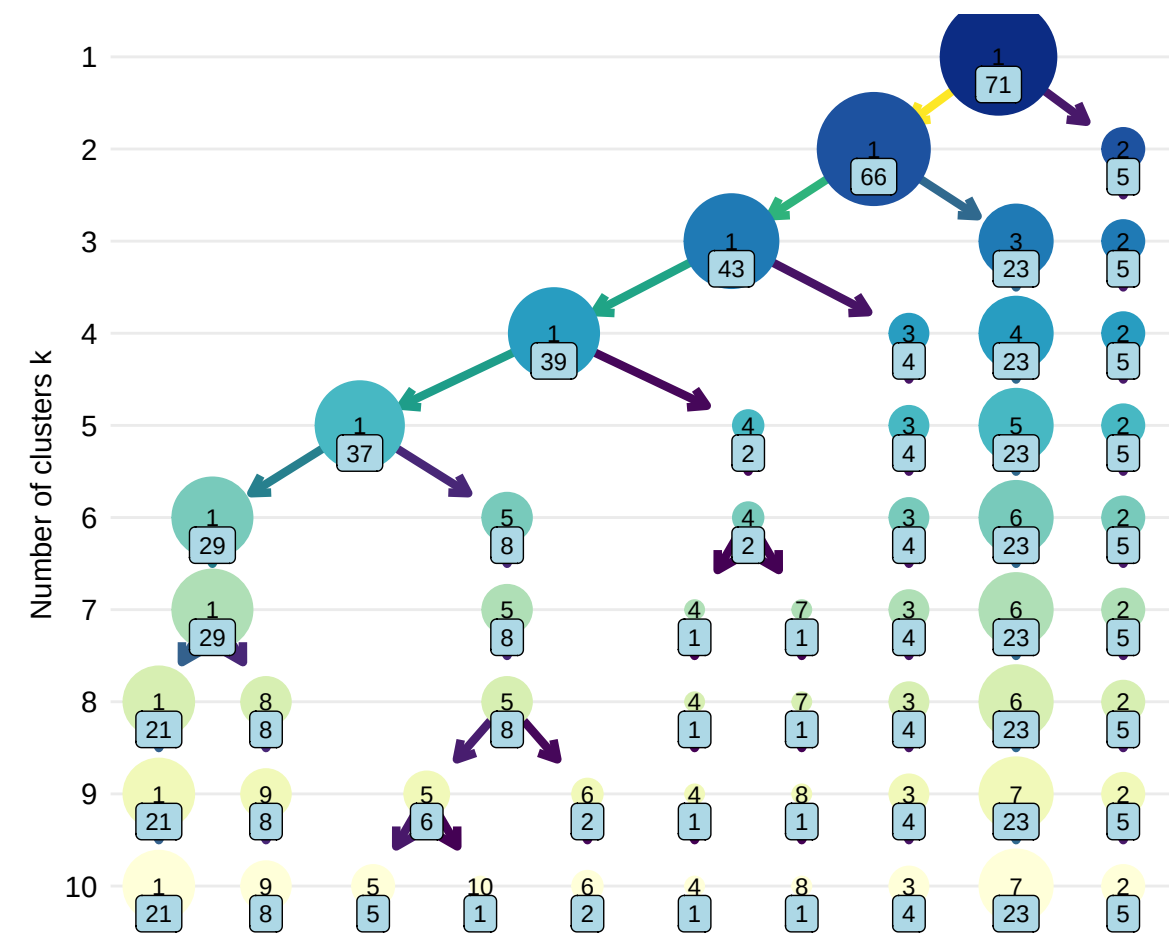


Figure 3: Treeplot of the HAC-procedure. The circle size indicates the cluster size, which is also labelled. The number in the circle shows the number of the cluster. The number of clusters in the respective step is depicted on the y-axis.

The clustering tree makes it easy to retrace the changes happening when new clusters are introduced. When a sixth cluster is introduced, cluster number 1 is split into two new clusters, one containing 29 and the other 8 vessels. These might represent valid fleet segments. When an additional cluster is created, cluster number 4 is split into two clusters, one containing only one ship. This is rather undesirable for one-ship-clusters are rarely a valid fleet segment, yet it is possible that one vessel is engaged in. Considering all the clustering diagnostics we examined, we have to decide to use six to seven clusters. A general strategy on how to proceed at a point like this is to stick with the upper end of the identified range. It is quite simple to re-join clusters in the final step of the segmentation but nearly impossible to separate them. Therefore, we will move forward with our example case and assume that **7 clusters** are present in the dataset.

4. Assign the clusters

After deciding on the number of clusters in the dataset, the clustering can easily be re-computed and assigned in a single line of code. It is advised to store the clustering result in a new data frame, which we will simply call `clustering`. The function to create this clustering data frame is called `segmentation_clustering`, it requires the transformed catch data and the assumed number of clusters as input.

```
clustering <- segmentation_clustering(catchdata = catchdata,n_cluster = 7)
glimpse(clustering)
```

```
## Rows: 71
## Columns: 2
## $ ship_ID <chr> "ship_1", "ship_10", "ship_11", "ship_12", "ship_13", "ship_14~
## $ cluster <ord> cluster 1, cluster 2, cluster 2, cluster 2, cluster 2, cluster~
```

The clustering result is a data frame with two columns.

- 1. A character column containing the vessel identifier.
- 2. The cluster, to which the vessel was assigned. The cluster is an ordinated factor variable.

This data frame will be the basis of the fleet segmentation. It is likely, that some of the clusters are not valid fleet segments, but rather should be joined in a single segment. To characterize the clusters and decide, whether or not to unite them to fleet segments, the FS-package contains a set of diagnostic methods.

5. Validate the clusters

5.1. Catch of the clusters

5.1.1. Catch composition

The first step to characterize the clusters is to evaluate their catch. The average percentage each stock contributes to a clusters catch can be tabelized and plotted using the following functions. Note that these functions require the basic (untransformed) data as an input argument.

```
# This will print a basic table
clustering_stockshares_table(data = data_stocks,
                             clustering = clustering)
```

```
## # A tibble: 24 x 3
##   cluster stock      share_stock
##   <ord>   <chr>          <dbl>
## 1 cluster 1 ple.27.420      56.2
```

```
## 2 cluster 1 cod.27.47d20      17.7
## 3 cluster 1 sol.27.4         15.9
## 4 cluster 1 tur.27.4         3.70
## 5 cluster 1 her.27.1-24a514a  3.10
## 6 cluster 1 lin.27.3a4a6-91214 2.01
## 7 cluster 1 hke.27.3a46-8abd  1.37
## 8 cluster 2 pra.27.1-2       100
## 9 cluster 3 reb.27.1-2       88.0
## 10 cluster 3 rng.27.1245a8914ab 6.35
## # ... with 14 more rows

# This will store the basic table as a data frame
stockshares_table <- clustering_stockshares_table(data = data_stocks,
                                                  clustering = clustering)

# This will produce an html-table displayed in the viewer
# (disabled in markdown)

# clustering_stockshares_table(data = data,clustering = clustering,
#                               style = "html")

# This produces a ggplot() of the average percentages and a
# ranking of shares colour-indicated
clustering_stockshares_plot(data = data_stocks,clustering = clustering)
```

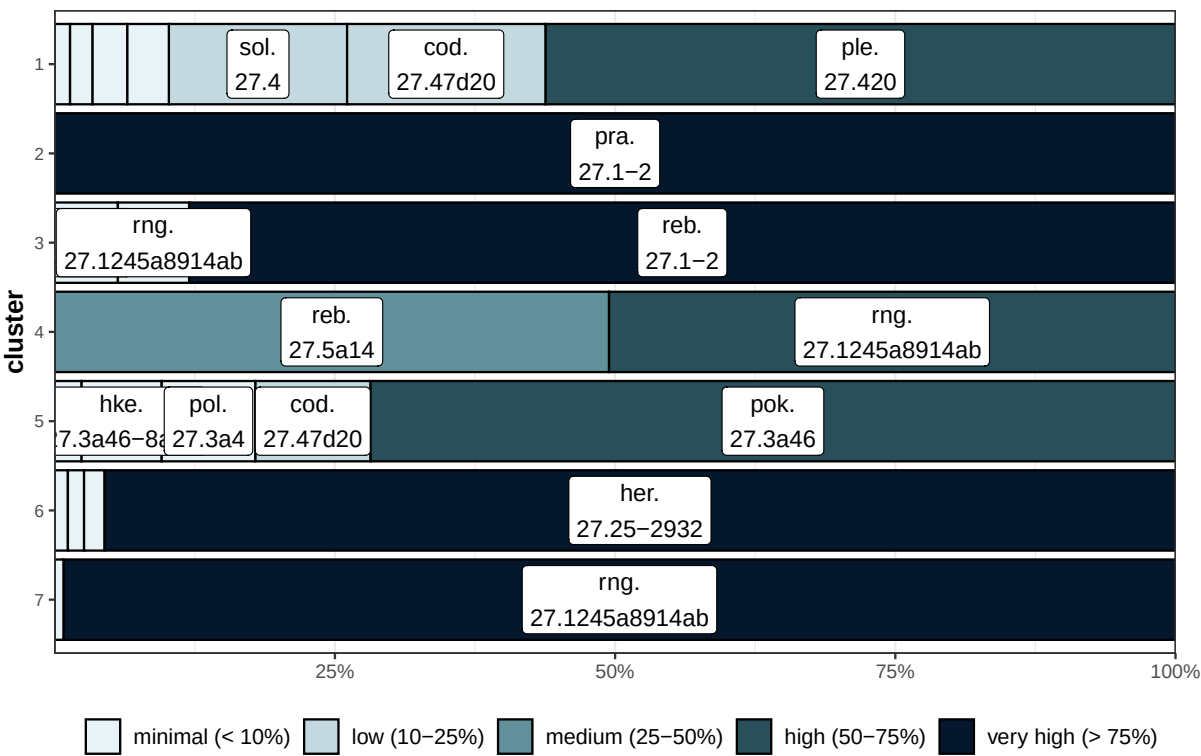


Figure 4: Barplot of the average percentage of each stock on the total catch of each cluster. The average percentage is depicted on the x-axis, the clusters on the y-axis. The colours indicate the magnitude of the stocks contribution to the total average catch. The stock names are indicated with labels, stocks are ordered by their average percentage contribution.

The plot illustrates the differences in catch composition among the clusters. At this point, expert knowledge on stock distribution is very helpful. It can be used to identify clusters representing a target fishery and clusters with similar catch composition which might be joined to fleet segments. In the example data, a few things become obvious with some background knowledge of the fished stocks. First, we can distinguish between clusters using coastal stocks (1,5 and 6) and clusters using subarctic and high seas stocks (2,3,4 and 7). This might be harder in your dataset (due to less distinct stock nomenclature or more widely distributed) but still worth the effort. In our example, the clusters fishing coastal stocks can be further divided into a very targeted herring fishery (cluster 6), a targeted saithe fishery (cluster 5) and a rather mixed coastal fishery (cluster 1). The clusters fishing high sea stocks comprise of a cluster fishing exclusively shrimps (2), a cluster fishing mainly redfish (3), a cluster fishing redfish and grenadier in approximately equal quantities (4) and a cluster fishing mostly grenadier (7). In a next step, we will check how the catch composition of those clusters is inter-related with 2-dimensional and 3-dimensional ordination methods.

5.1.2 Ordination

The FS-package includes functions for 2- and 3-dimensional ordination methods. They are very valuable tools for visualising the compositional distances of clusters, identify overlaps,

and therefore clusters which might be suitable for uniting them to fleet segments. We will start with the 2-dimensional MDS. We can use a metric MDS because of the metric nature of the input data. It is to prefer over an nMDS because it reflects true distances, not ranked distances.

```
clustering_MDS(catchdata = catchdata, clustering = clustering)
```

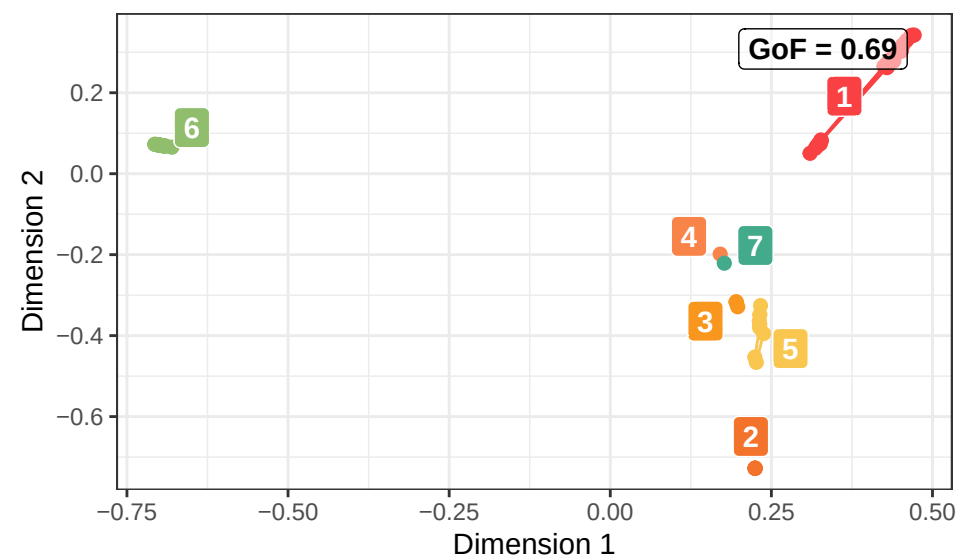


Figure 5: MDS of the transformed vessel catch data. Points represent individual vessels, coloured according to their cluster affiliation. Clusters are labelled, labels are coloured similar to points. The **goodness of fit** (GoF) is labelled.

This MDS reveals some of the relations in the catch composition and shows quite good goodness of fit. We will analyse it along the dimensions, starting with cluster 6. The cluster is distinctly different from the other clusters, which makes sense considering its catch composition. Vessels in cluster 6 caught near to exclusively herring, while no other clusters vessels caught considerable amounts of herring. Moving in the first dimension, cluster 1 opposes cluster 6. Moving down along the 2nd dimension, the clusters 7, 4 and 3 are grouped close to each other by the ordination. This is in accordance with their catch composition, for they all fish on the same stocks (redfish, grenadier and halibut), but in different quantities. Further down in the most negative area of the second dimension, cluster 2 and cluster 5 are ordinated very closely together. This is quite confusing at first, for cluster 2 is formed by vessels performing a targeted high seas shrimp fishery while cluster 5 vessels are mainly fishing coastal saithe and cod. It illustrates the limits of ordinating a catch composition-matrix in a two-dimensional space. The `clustering_MDS`-function therefore also includes the option of plotting a 3-dimensional ordination.

```
clustering_MDS(catchdata = catchdata, clustering = clustering, dim = 3)
```

This will produce an interactive, 3-Dimensional in the RStudio-Viewer looking like this:

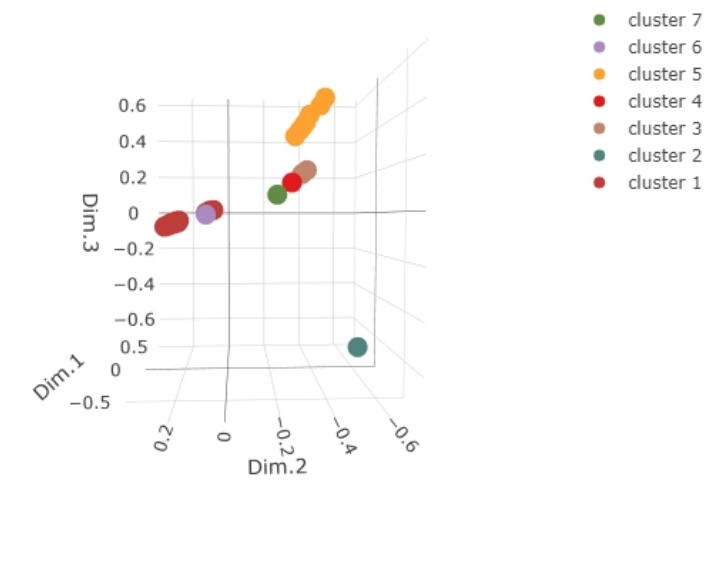


Figure 6: Snapshot of interactive 3D-MDS. Rotated to 2nd dimension being the x-axis, 3rd dimension the y-axis and 1st dimension the z-axis.

As you see in this snapshot, adding a third dimension to the visualization of the MDS shows a large distance between the second and fifth cluster along the third axis. Differences like this can only be revealed in a three-dimensional ordination so I encourage you to always check it in your fleet segmentation progress, even if it might take a moment to fully interpret the ordination. In both ordinations, you might have noted, that two clusters (2 and 6) are only represented by a single point. This is due to the catch composition of the clusters: They heavily fish on one species and there is close to no inner-cluster variation. Because of cases like this (which are also common in real data), it is important to get an impression on the properties of the clusters and the vessels in them besides their catch composition two finally decide on which clusters can be used as fleet segments and which clusters should be joined. The FS-package contains functions to do this which will be covered in the next section.

5.2. Vessel and cluster properties

The main diagnostic features for the characterisation of fleet segments based on the available data besides the clusters catch composition are the number of vessels in the clusters, the length distribution of vessels in the cluster, the total catch of the individual vessels and the total catch of each cluster. The FS-package contains a function to compute a plotgrid visualizing all these features.

```
clustering_plotgrid(data = data_stocks,
                    clustering = clustering,
                    shiplength = shiplength)
```

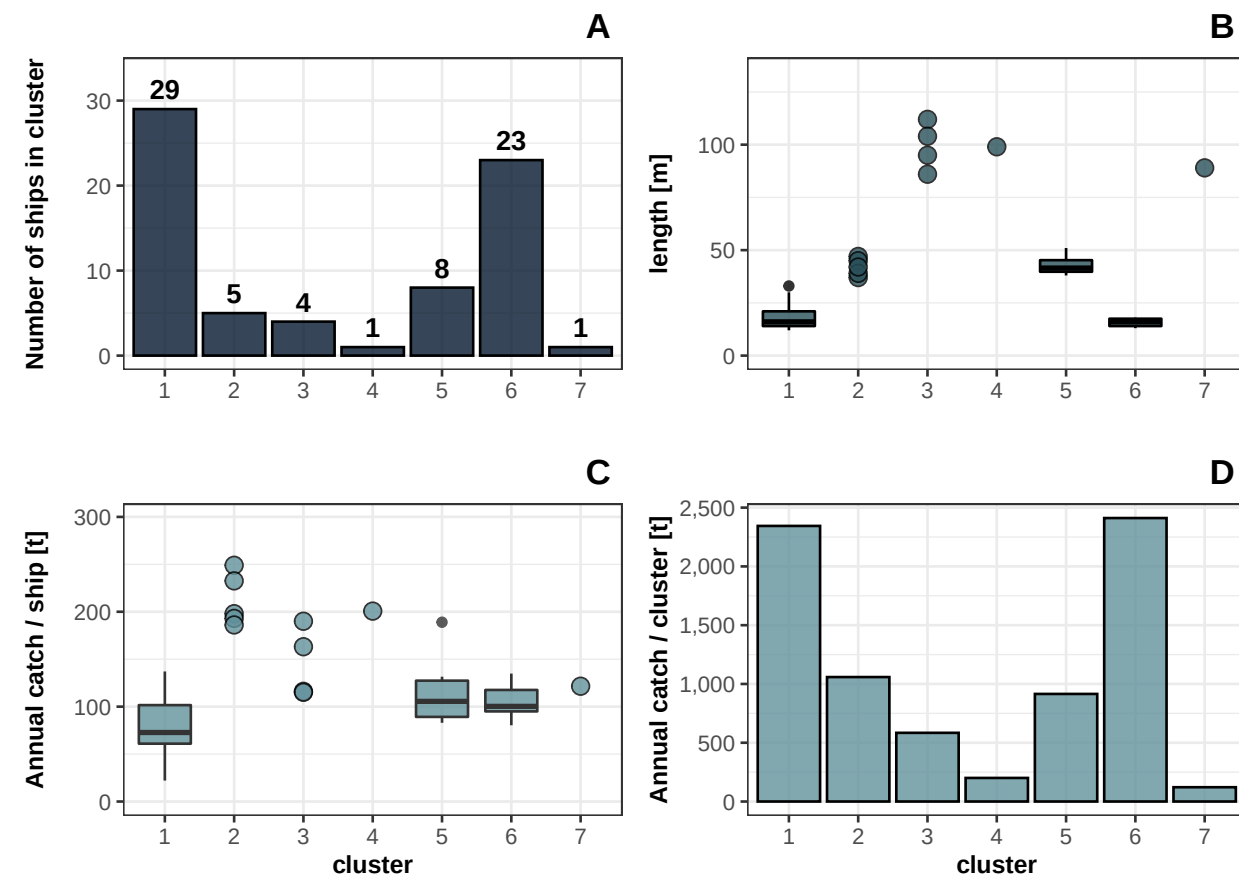


Figure 7: Grid of mixed plot types displaying vessel characteristics. A) Barplot of cluster size. The number of ships in each cluster is plotted on the y-axis vs. cluster number on the x-axis. The number over each bar gives the number of vessels in the cluster. B) Mixed box- and dotplot of the length of the vessels in each cluster. Vessel length in meters is depicted on the y-axis vs. cluster number on the x-axis. Clusters containing more than 5 vessels are depicted with a boxplot, for smaller clusters all vessel lengths are plotted with points. C) Mixed boxplot of the annual catch of individual vessels in tons on the y-axis vs. cluster number on the x-axis. Clusters containing more than 5 vessels are depicted with a boxplot, for smaller clusters each vessel catch is plotted with a point. D) Barplot of the total annual catch of all vessels in the clusters. The total catch in tons is depicted on the y-axis vs. cluster number on the x-axis.

This plot grid contains a lot of information on the clusters and enables us to group them according to their structure. The numerically largest clusters are cluster 1 and cluster 6, both consisting of more than 20 vessels. They contain mostly small vessels of 15 to 20m length. Individual vessels have a medium annual catch but add up to the largest total cluster catches. The high seas clusters 2,3,4 and 7 contain rather few ships; 4 and 7 are actually single-ship clusters. The vessels are around a hundred meters long and have medium to high individual catches. The total cluster catch is consequently low since they are the only members of their clusters. This is just an extract of the information one can draw from this visualization. Even though the data used is artificial, this is a very realistic pattern to encounter in fisheries data. Especially high seas vessels are often operated by larger companies, and fishing quota is distributed among them not equally but region- or stock specifically. They, therefore, tend to form very small subclusters even though they are engaged in similar fishing operations. Cluster 5, which is the coastal fishery focussed on saithe, contains remarkably longer vessels than the other clusters of coastal fishery. They are likely to have different coast structures and other areas of operation than the other coastal clusters. Cluster 1, the highly mixed coastal cluster, contains relatively small vessels with an annual individual catch comparable to vessels from cluster 6. It is upon expert knowledge to decide whether or not these clusters should be joined as a fleet segment. For our example dataset, we assume that due to their differences in fishing strategies and techniques, area of operation, and cost structure, they will be treated as separate fleet segments.

All the plots in this grid can be produced as single plots using

```
cluster_size_plot(clustering = clustering) # For A)
shiplength_plot(clustering = clustering,
                 shiplength = shiplength) # For B)
singleship_catch_plot(data = data_stocks,
                      clustering = clustering) # For C)
clustercatch_plot(data = data_stocks,
                  clustering = clustering) # For D)
```

Note that there are a lot more kinds of fisheries data sources with the potential to characterise clusters and ultimately assign fleet segments, e.g. spatial data (haul positions and VMS-Data), temporal data (total time spent at sea and fishing trip duration, seasonality of fishing activity) as well as data on the cost structure (total costs as well as cost ratios). I encourage you to make use of every data source available and inspect your clusters from all possible angles and invite you to discuss this with me. It will make it easier to assign the fleet segments to your clusters. The actual process is fairly simple.

6. Assign fleet segments

Taking the stock shares on the clusters catch, the result of the ordination and the vessel and cluster characteristics into account, we can make a profound decision on the number of actual fleet segments in our example data.

- Cluster 1 will become the fleet segment **Coastal mixed demersal fishery**

- Cluster **2** will become the fleet segment **High seas shrimp fishery**
- The clusters **3,4** and **7** will be merged to the fleet segment **High seas mixed demersal fishery**
- Cluster **5** will become the fleet segment **Coastal saithe fishery**
- Cluster **6** will become the fleet segment **Coastal herring fishery**

The clustering result is stored as a factor variable in the data frame you created with the `segmentation_clustering`-function and working with factors can be kind of tricky. There is no built-in function in the FS-package, but I will provide you with code to easily call the clusters by number and assign fleet segments to them. I recommend you to create a new data frame in which you assign the fleet segments. We will call it `final_segmentation` in our example and introduce a column called `FS` for the fleet segments.

```
final_segmentation <- clustering

# Create FS-column
final_segmentation$FS <- NA

# Call clusters by factor level
# and assign fleet segments

# 1) Coastal mixed demersal fishery
final_segmentation$FS[final_segmentation$cluster %in%
                      levels(final_segmentation$cluster)[c(1)]] <-
  "Coastal mixed demersal fishery"

# 2) High seas shrimp fishery
final_segmentation$FS[final_segmentation$cluster %in%
                      levels(final_segmentation$cluster)[c(2)]] <-
  "High seas shrimp fishery"

# 3) High seas mixed demersal fishery
final_segmentation$FS[final_segmentation$cluster %in%
                      levels(final_segmentation$cluster)[c(3,4,7)]] <-
  "High seas mixed demersal fishery"

# 4) Coastal saithe fishery
final_segmentation$FS[final_segmentation$cluster %in%
                      levels(final_segmentation$cluster)[c(5)]] <-
  "Coastal saithe fishery"

# 5) Coastal herring fishery
final_segmentation$FS[final_segmentation$cluster %in%
                      levels(final_segmentation$cluster)[c(6)]] <-
  "Coastal herring fishery"
```

```
# check if we forgot to assign any cluster
anyNA(final_segmentation$FS)
```

```
## [1] FALSE
```

```
# if FALSE, make fleet segment a factor variable
# otherwise check assigning code
final_segmentation$FS <- as.factor(final_segmentation$FS)
```

```
# list the segments
levels(final_segmentation$FS)
```

```
## [1] "Coastal herring fishery"      "Coastal mixed demersal fishery"
## [3] "Coastal saithe fishery"       "High seas mixed demersal fishery"
## [5] "High seas shrimp fishery"
```

```
# glimpse the data frame
glimpse(final_segmentation)
```

```
## Rows: 71
## Columns: 3
## $ ship_ID <chr> "ship_1", "ship_10", "ship_11", "ship_12", "ship_13", "ship_14~
## $ cluster <ord> cluster 1, cluster 2, cluster 2, cluster 2, cluster 2, cluster~
## $ FS      <fct> Coastal mixed demersal fishery, High seas shrimp fishery, High~
```

And this is it! We have successfully assigned a fleet segment to each ship in our example data.

A final remark on assigning the fleet segments: The clustering process of the FS-package requires stock-based input data. Fishing vessels are typically designed to perform specific operations and therefore to harvest specific stocks. Consequently, the clustering result should be sufficient to assign fleet segments to your vessels. In case you want to take into account further variables in your segmentation, e.g. the ship length, this is, of course, possible since the clustering result is just a qualitative variable and can easily be combined with other variables. To illustrate such a case, we will assume that there is a *large-scale* (ships above 24m length) and a *small-scale* (ships below 24m length) *coastal mixed demersal fishery*.

```
# join the ship length information to the
# segmentation data
final_segmentation <- final_segmentation %>%
  left_join(example_lengthdata)
```

```
# transform FS back to a character variable
# this is important for manipulation, otherwise
```

```
# an error will occur
final_segmentation$FS <- as.character(final_segmentation$FS)

# assign the small scale mixed demersal coastal fisheries
final_segmentation$FS[final_segmentation$cluster %in%
  levels(final_segmentation$cluster)[c(1,8)]
  & final_segmentation$shiplength < 24] <-
  "Small-scale coastal mixed demersal fishery"

# assign the large scale mixed demersal coastal fisheries
final_segmentation$FS[final_segmentation$cluster %in%
  levels(final_segmentation$cluster)[c(1,8)]
  & final_segmentation$shiplength >= 24] <-
  "Large-scale coastal mixed demersal fishery"

# check if we forgot to assign any cluster
anyNA(final_segmentation$FS)

## [1] FALSE

# if FALSE, make fleet segment a factor variable
# otherwise check assigning code
final_segmentation$FS <- as.factor(final_segmentation$FS)

# list the segments
levels(final_segmentation$FS)

## [1] "Coastal herring fishery"
## [2] "Coastal saithe fishery"
## [3] "High seas mixed demersal fishery"
## [4] "High seas shrimp fishery"
## [5] "Large-scale coastal mixed demersal fishery"
## [6] "Small-scale coastal mixed demersal fishery"
```

As you see, combining the clustering information with further information, such as vessel characteristics, is simple. I hope you found this manual helpful and enjoyable to read. Good luck with your fleet segmentation. If you have any questions, remarks or critique, I am looking forward to hearing from you.

All the best, Erik Sulanke

Final remark: The example dataset included in this package is purely fictional and for illustrative purposes only. Any resemblance of real fisheries is of coincidence.

Appendix

Package installation

This is the package installation code for all the packages you need to make the FS-package work. Simply start a new R-session outside any project and run this code to your console prior to installing the FS-package from the .zip-file.

```
install.packages(c("plotly",
  "knitr",
  "kableExtra",
  "RColorBrewer",
  "dplyr",
  "tidyr",
  "ggplot2",
  "scales",
  "ggpubr",
  "stringr",
  "lme4",
  "rpart",
  "data.table",
  "vegan",
  "factoextra",
  "FactoMineR",
  "dendextend",
  "cluster",
  "NbClust",
  "fpc",
  "ade4",
  "BBmisc",
  "clustree",
  "networkD3",
  "htmlwidgets",
  "labdsv",
  "ggfortify",
  "corrplot",
  "ggforce",
  "concaveman",
  "forcats",
  "ggrepel",
  "ggdendro"))
```

Fleet segmentation pilot report / Finland

Antti Takolander, Natural Resources Institute Finland (antti.takolander@luke.fi)

13.4.2021

The data used

The data used in the analysis consisted of the catch data by Finnish trawlers in 2020 (N = 148). The fleet data also contained ships which used passive gears (N = 5849) and fishing without a ship (N = 373), which was mostly fishing on ice, but these were not analysed in the pilot because of time limitations.

The pelagic trawlers gear class analysed here consist of trawlers which mostly target Baltic herring (*Clupea harengus membras*) in the Northern Baltic Sea, with additional target species being sprat (*Sprattus sprattus*), smelt (*Osmerus eperlanus*) and vendace (*Coregonus albula*). Other species are caught in minor amounts and are mostly sold as fish fodder. All the vessels fish in the Baltic Sea, (27.d), with majority of the fishing occurring in sub-areas 27.d.32,27.d.31, 27.d.29 and 27.d.30. The mean length of trawlers was 25 meters (Figure 1).

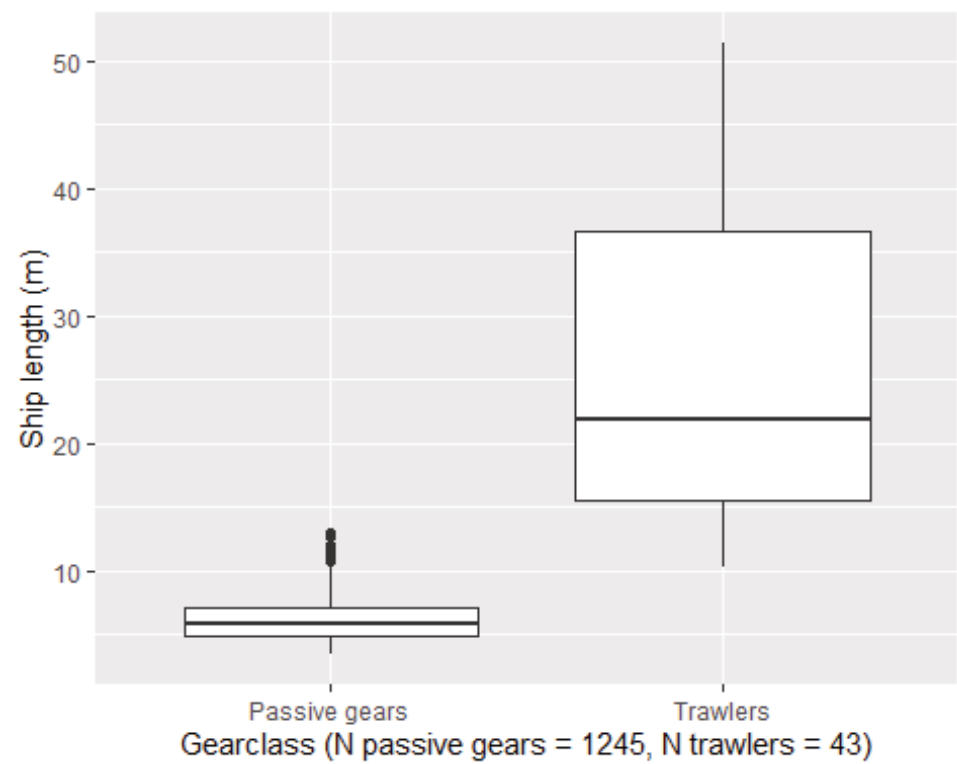


Figure 1. The length of ships in Finnish fleet in 2020 in two gear classes: passive gears and pelagic trawlers.

Results

Pelagic trawlers

To account for rather small amounts of catch of other species aside herring, the assign_stocks function was run with stock autogeneration function on with 0 tonnes threshold. This should generate the stocks for the clustering algorithm for even very small catches which might be present in the data.

With initial run, the numberclust_tabel function indicated that the optimal number of clusters would be between three to six (table 1 & figure 2).

Table 1. Estimated number of clusters for pelagic trawlers gear class.

Indices	Optimal number of clustersIndex value	
Average silhouettes	2	0.699
Mantel test	4	0.862
Davis-Bouldin index	6	0.226
SD index	6	4.216
Calinski-Harabasz index	15	240.187

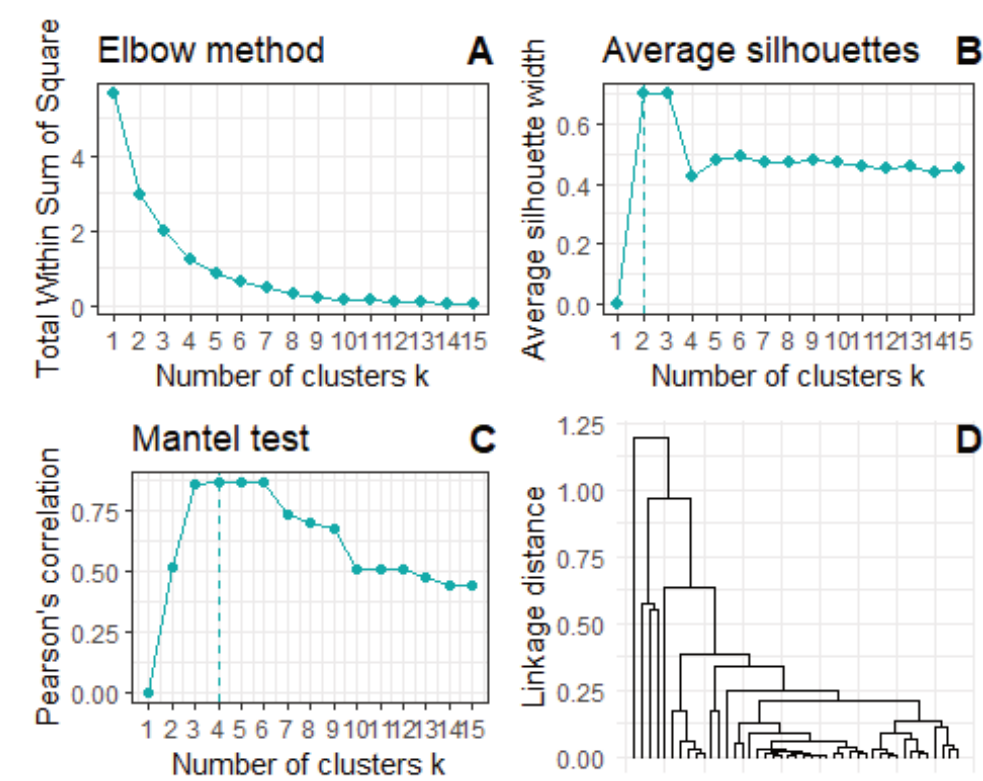


Figure 2. Visual inspection for number of clusters in the pelagic trawlers gear class.

However, the size of the first three clusters were very small. Based on visual inspection (Figure 3), the number of clusters was defined to be 10. This was probably not a realistic number of clusters for the entire pelagic trawlers gear class, as the data set only contained 148 vessels, but the cluster containing majority of ships was only split at this number of clusters formed.

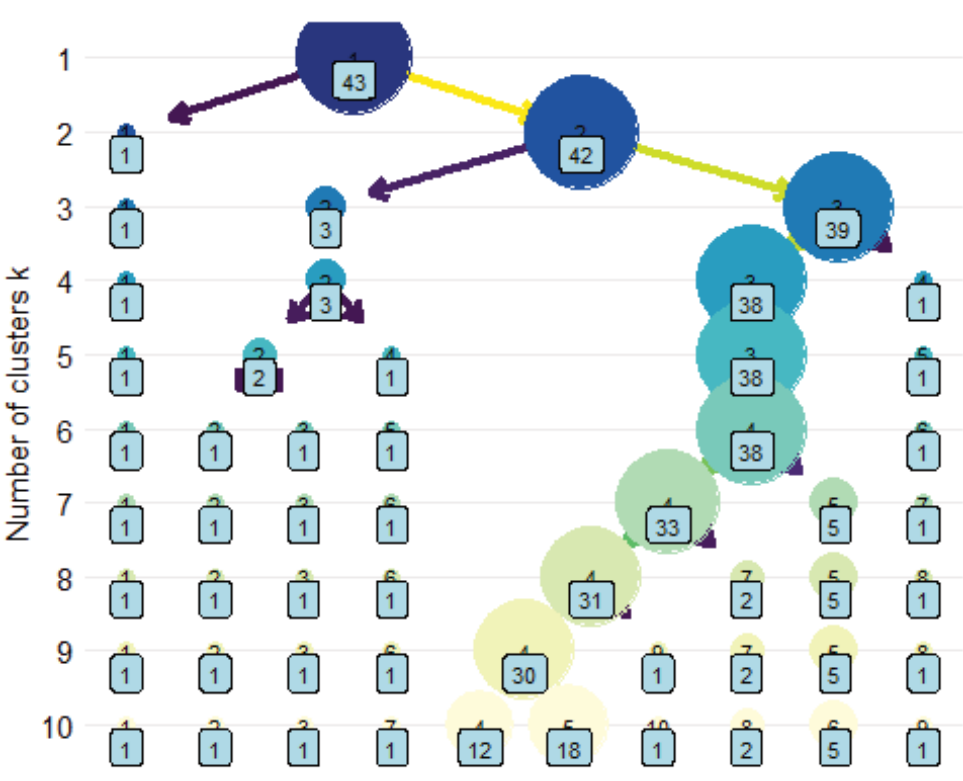


Figure 3. Cluster tree plot for pelagic trawlers.

The catch composition of the clusters differed somewhat. The clusters 4 and 5 targeted mainly herring, and cluster 4 had also some secondary sprat catch. The cluster 10 also targeted herring, but had undefined species (FIN) as secondary catch. Several clusters (2, 3 and 7) targeted vencace (FVE) and only cluster 9 had smelt (SME) as the priority catch (Figure 4). Cluster 6 had herring as the main target species, but sprat formed a substantial amount of the catch.

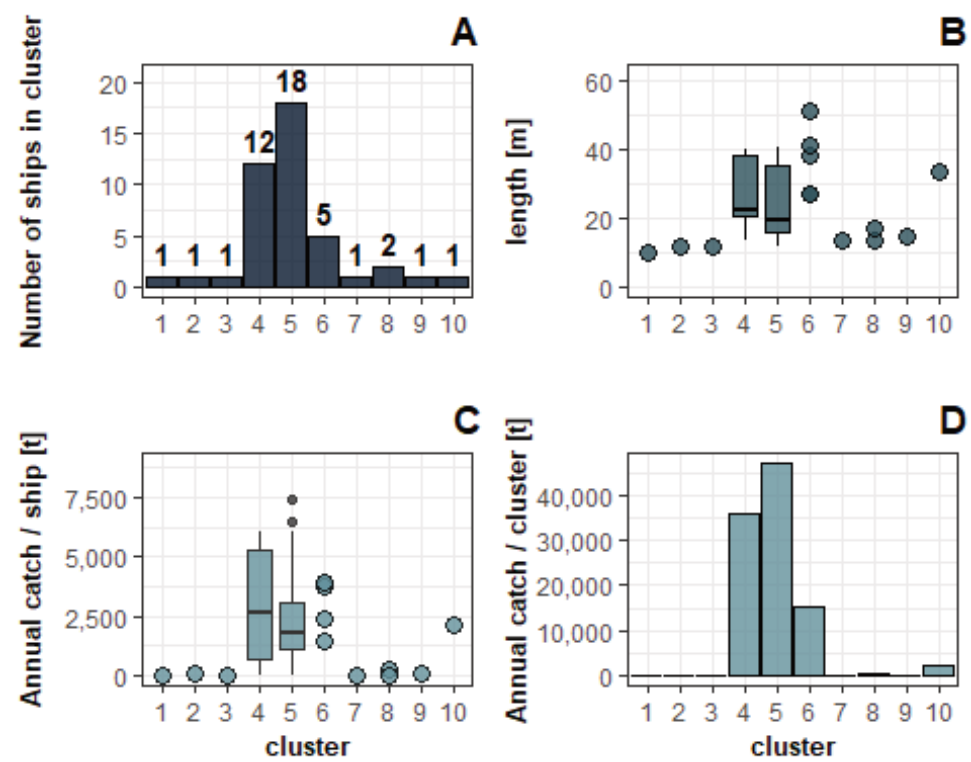
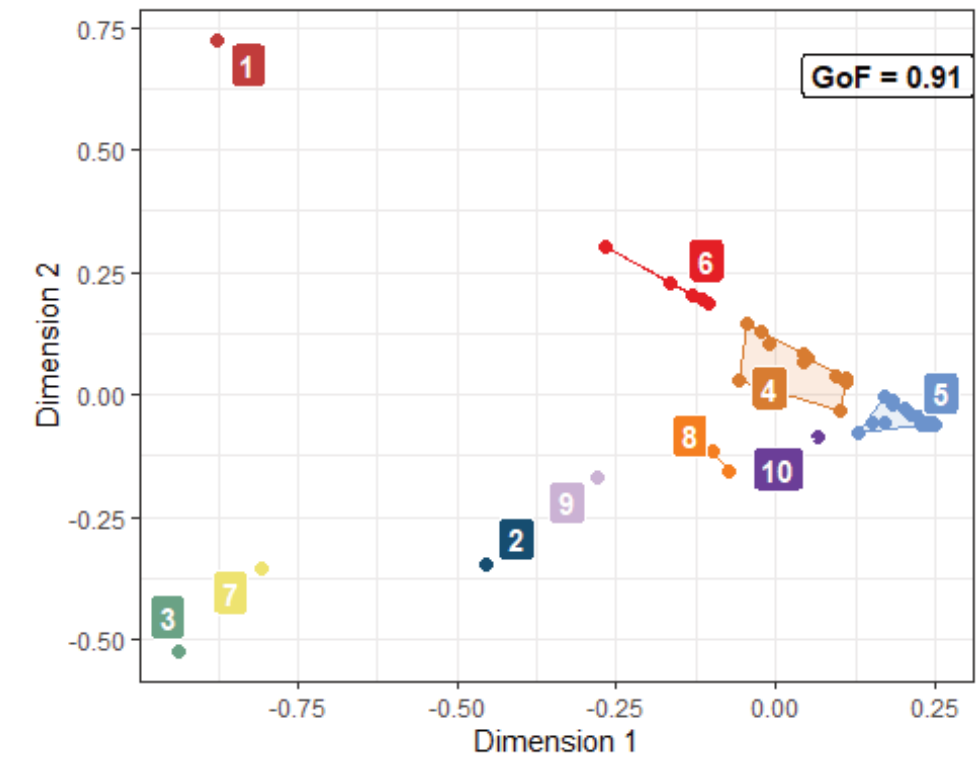
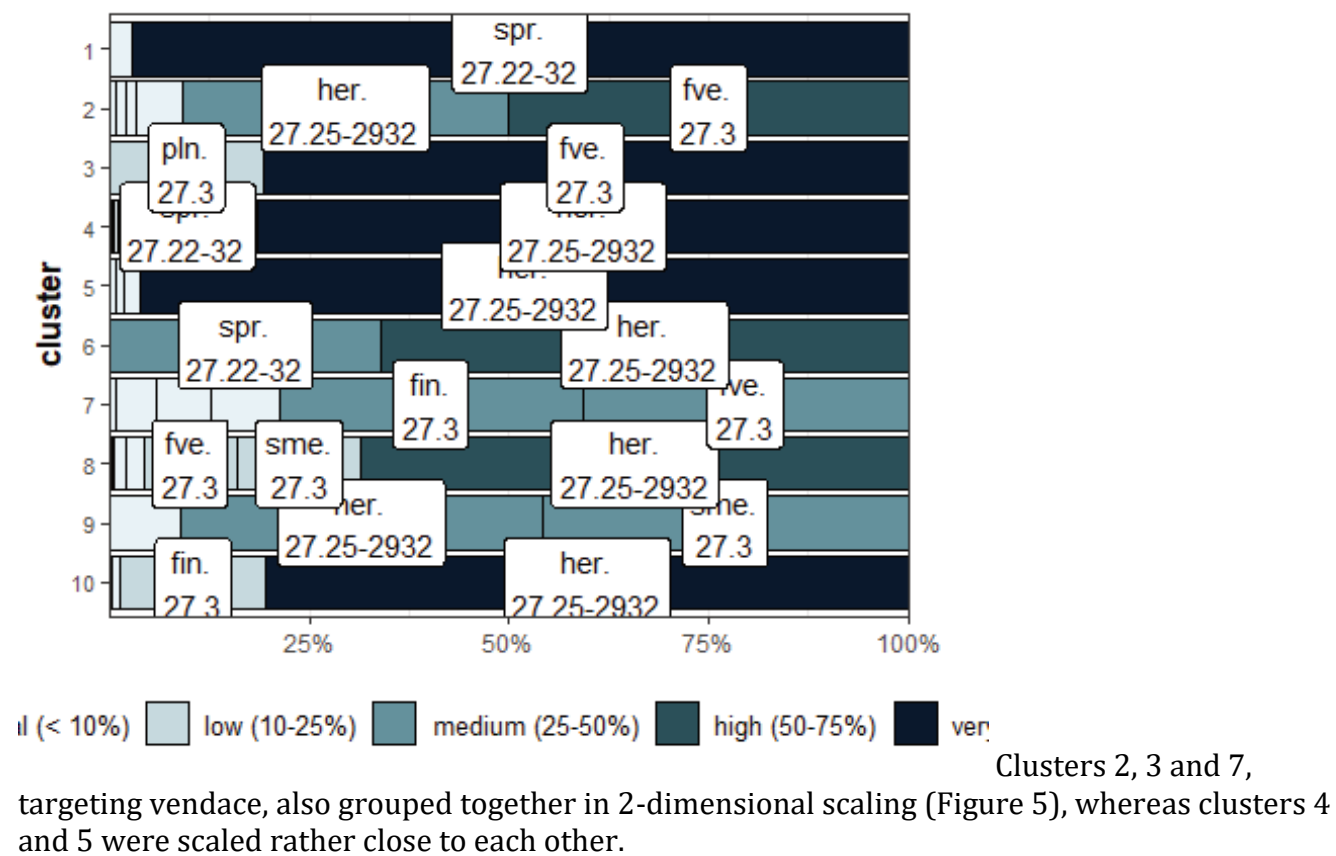


Figure 6. The properties of vessels in each cluster

The number of vessels in all clusters was very low except for clusters 4, 5 and 6. These were also the clusters with larger ships, and caught most of the catch. Cluster no 10 consisted of a single ship, which by vessel length and catch could be grouped with clusters 4,5 or 6. This vessel caught mainly herring but had substantial amount of secondary catch for which the species composition was not known.

Potentially, clusters 4 and 6 could be grouped together, as the stock composition of these two is highly similar (Figures 7 & 8). The real difference is in the relative share of sprat as secondary catch, but this division is not that distinctive.

The lone ship in cluster 1 is catching dominantly sprat, and does not really cluster together with anything else if only catch is considered. However,

Figure 5. Multidimensional scaling of pelagic trawlers gear class.

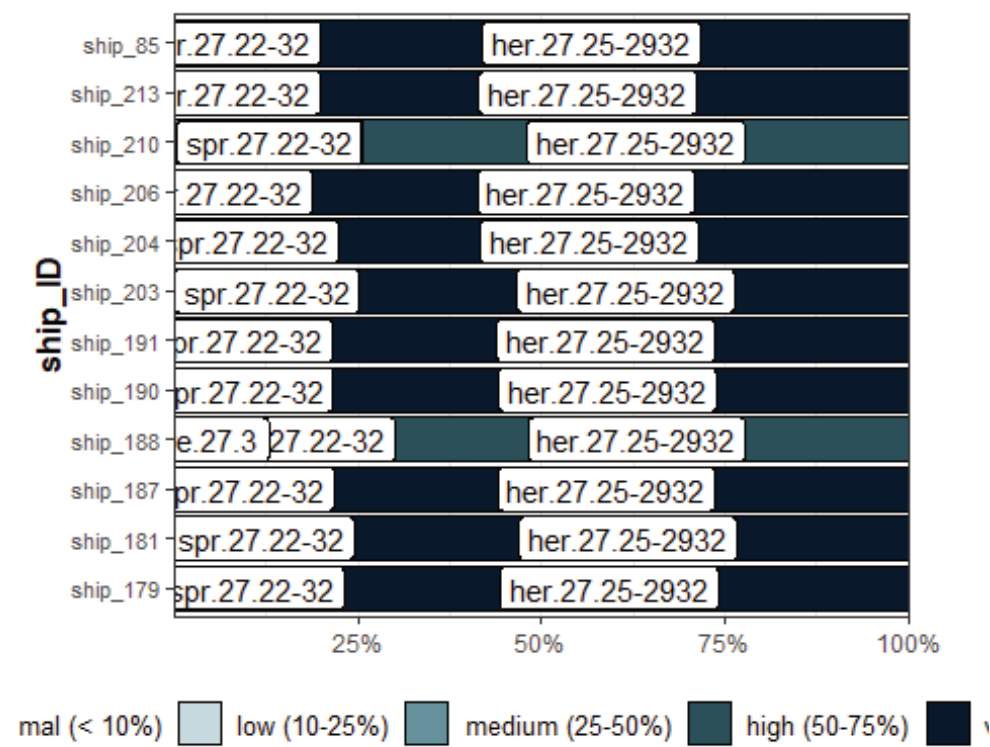


Figure 7. Stock shares of cluster 4.

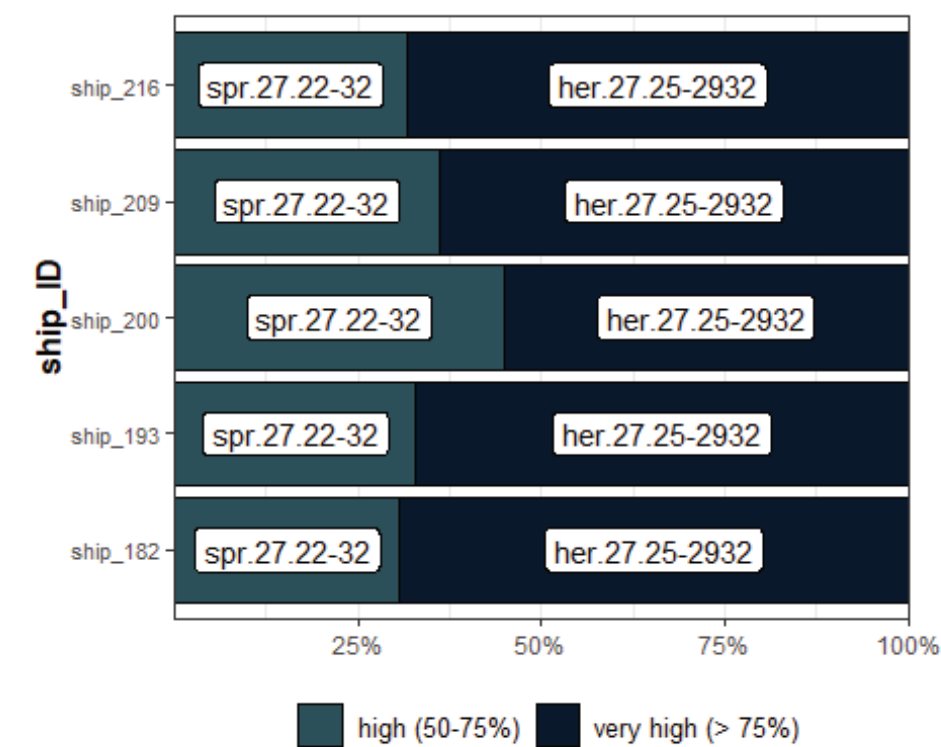


Figure 8. Stock shares of cluster 6.

Final clustering of the pelagic trawlers

Based on the steps of data exploration above, the ten clusters suggested by the clustering algorithm were assigned to the following clusters:

1. “Large” trawlers targeting herring (cluster 5): “Pelagic herring fishery”
2. Trawlers targeting herring but with substantial secondary sprat or other species catch (clusters 4, 6, 10, 8): “Opportunistic herring / pelagic fishery”
3. Small trawlers targeting mixed fishery (clusters 1,2,3, 7, 9): “Mixed fishery”

Vessels in cluster 8 are substantially smaller than those in clusters 4,6 or 10 but they were aggregated into the same final cluster 2 because of similar catch composition. Vessels in clusters 1,2,3, 7 and 9 had mixed main target species (vendace, smelt or sprat), small size and small annual catch. These were grouped into a single final cluster “Mixed fishery”.

Suggestions for future development

1. Raw data return function.

The function `single_cluster_stockshares` is very handy in visualising the stock composition of the suggested clusters. It could be beneficial to have a similar function to extract the raw catch data and vessel ids and length as data frames. If the data frame is printed in “wide” format (i.e. catches per species as columns), it is very easy to see what kind of catch composition has gone into a particular cluster, and compare the clusters. Of course the usability would depend on the number of vessels in a cluster, but it seems that quite often the number of vessels are within reasonable range.

2. Include number of vessels into stockshares plot

The function `clustering_stockshares_plot` could also include the information on number of vessels in each cluster. In Finnish pelagic trawler gear class, the clustering algorithm tended to form rather small clusters, siezes of 1 or 2 ships. To pinpoint these, it would be nice to see these alongside the stock composition of the cluster. This would help to think which clusters could be potentially grouped together.

3. Considering including vessel length or absolute catch as a variable in the clustering?

Should vessel length be included in the clustering as a variable? Or is it already? Now it seems that the vessel length is only used in visual checking of the clustering, and potentially to group different clusters together. To me this is bit puzzling as it seems that vessel length is potentially an important grouping variable, but it is only brought in to the analysis at the cluster validation stage. I understand that since the motivation for the clustering approach has been to get rid of static size classes it might seem like a step back to include vessel length back in. But it seems that the length is included in subjective judgement anyway at the cluster validation phase. So bringing the vessel length in as a variable could be a way toward more formal, non-subjective analysis.

Another way around the problem could be to use absolute catches per vessel in the clustering. These are highly correlated with vessel length, but of course might also reflect some other quantities such as fishing activity. This may be important for fleets which have lots of small ships, and variable fishing activity. For example, the ships of the Finnish fleet which use passive gears

are all about the same size (around or below 10 meters), but the fishing intensity varies widely, because some are used by active fishermen with high investments on gear and fishing, and some are used by part-time fishermen which exercise fishing as a side income activity. These two groups would have about the same catch composition, as both target the same stocks, and since the vessels are of the same size, the clustering approach piloted here would presumably group them into similar clusters. Yet there are large differences in fishing activity, catch, and especially economic performance between these fishermen/ships. Including absolute catches either into the clustering approach or visual inspection of the clusters would potentially allow to tease these groups apart.

4. Function to assign the final clusters?

If the final clustering is done manually after data visualisation, it would be nice to have a ready-made function, which would produce similar clustered object as `segmentation_clustering()`. This would allow the user to re-run the analysis and visualizations on the final clustering, to verify that the clustering is reasonable.

5. A parameter for minimum size of clusters

In the Finnish trawler data analysed here many clusters that were formed were small, usually only a single ship or two ships. This is not desirable, as such “mini-clusters” are not viable to be used in any official analyses due to data privacy restrictions. Also, having a large number of very small clusters is a problem for applicability of the analysis if the number of vessels is large. In these cases, the user would have to visually inspect and verify all the mini-sized clusters. If there are thousands of ships this becomes very laborious.

It would be good to have a parameter setting to set the minimum size of the clusters formed.

Fleet Segmentation package workshop – exploration of Icelandic fleet (ICES division 5.a)

The workshop was very well organized by Erik and Jörg. They gave us good introductions on the purpose of the package, development, and usage. Organizers were helpful in every aspect and tried to assist everyone on their shortcomings during the course of the workshop.

The fleet segmentation package is nicely structured and user friendly with a clear and detailed plot output. There was no need during our trial analysis to convert the Icelandic catch data to the ICES stocks as much of it is not identified as such, therefore the processing of data was quite straightforward. Logbook data from 2019 from three Icelandic fleet segments with passive gear was experimented with during the workshop:

- gillnetters – 248 vessels
- longliners and jiggers combined – 859 boats/small vessels

Gillnets

The output from the gillnet fisheries clustered the majority of the fleet in one cluster where Icelandic cod is very prominent. That is not unusual as the gillnet fisheries are quite seasonal and mostly target large cod. However, in 2019 approximately 50% of the landed catch of Greenland halibut was from gillnets and the package picked up on those “new” fisheries quite nicely.

Optimal number of suggested clusters was between 6-11 clusters and during the workshop 7 clusters were used to begin with for gillnetters. However, after several trials the number of clusters resulted in a total of 12, which enabled the observation of how the largest cluster splits into smaller units. We can still see that most of the gillnet vessels are clustered in one cluster with cod as the main target species.

The gillnet Greenland halibut fisheries are farther from shore and at more depth than the traditional Icelandic gillnet fisheries and the FS package distinguishes them efficiently and quite early in the process.

Characterisation of fleet segments for gillnet fisheries is quite tricky as all of the clusters could be identified as *coastal gillnet cod fisheries* except for the *Greenland halibut gillnet fisheries* (cluster 11), which are very different from traditional gillnet fisheries (Fig 2. and Fig. 5). Cluster 5, with 3 vessels, could also be identified as *gillnet monkfish fisheries* as monkfish can be specifically targeted by this gear type.

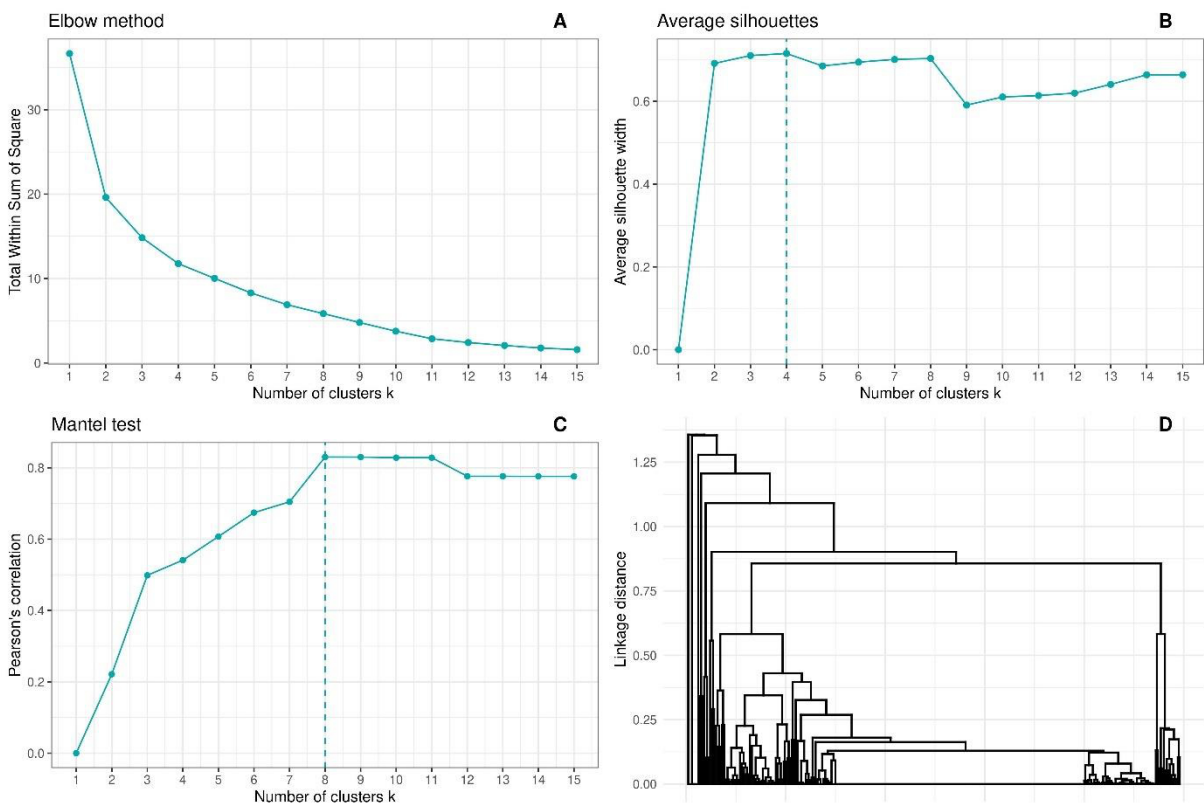


Fig 1. Grided diagnostics of HAC procedure for gillnet fisheries.

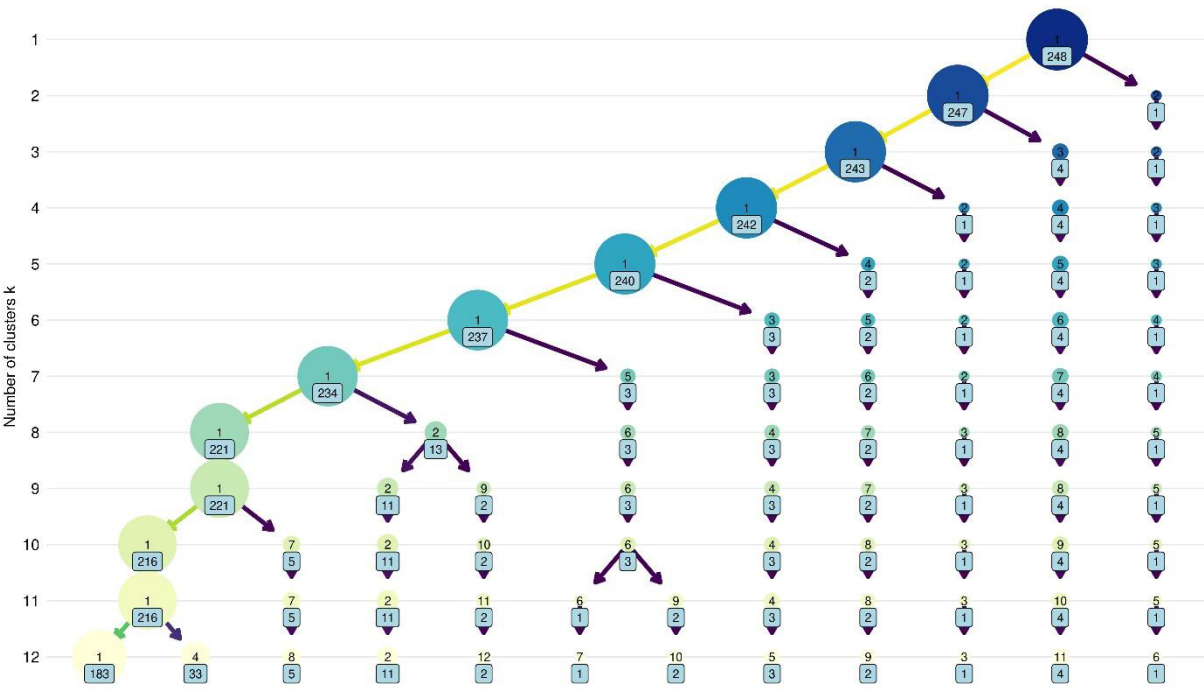


Fig 2. Treeplot of the HAC-procedure for gillnet fisheries.

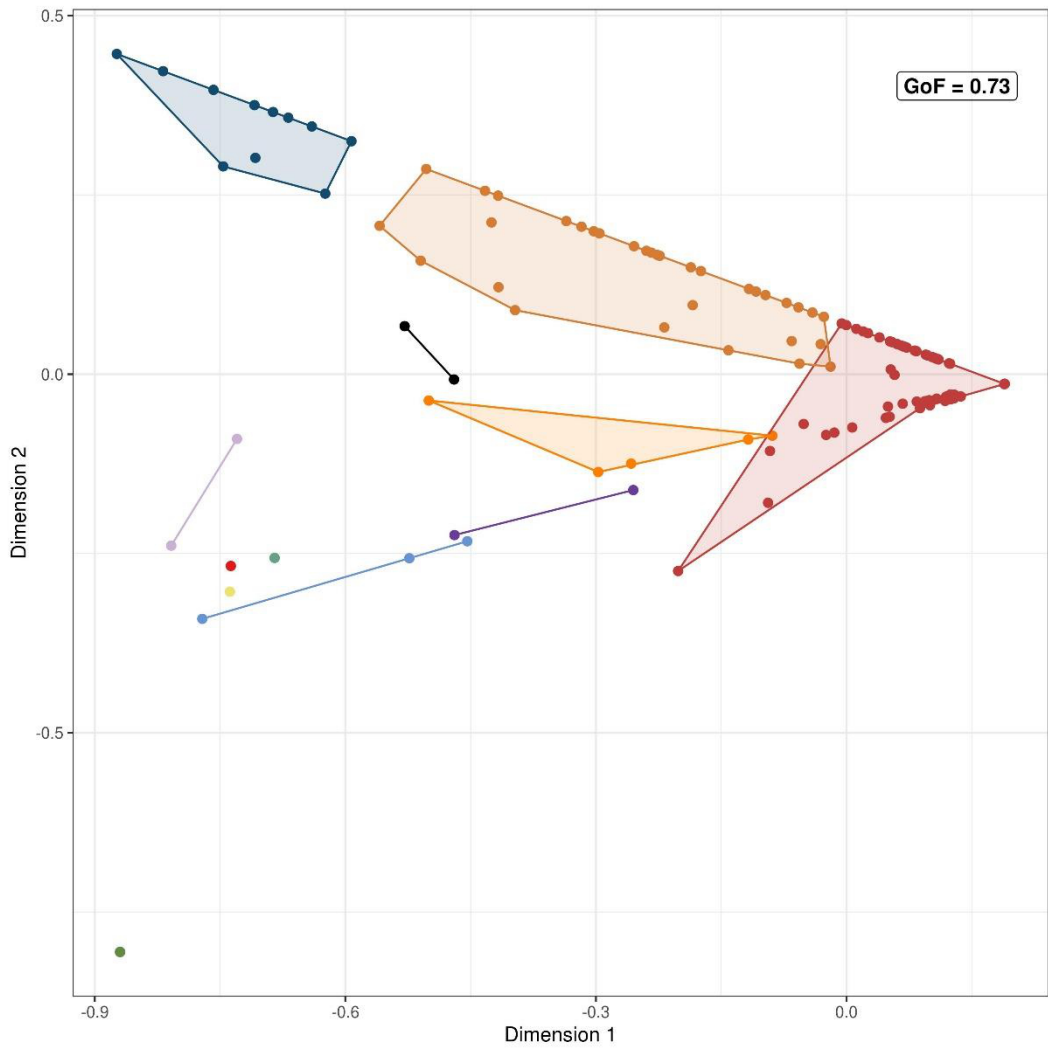


Fig 3. MDS of the transformed vessel catch data from gillnet fisheries.

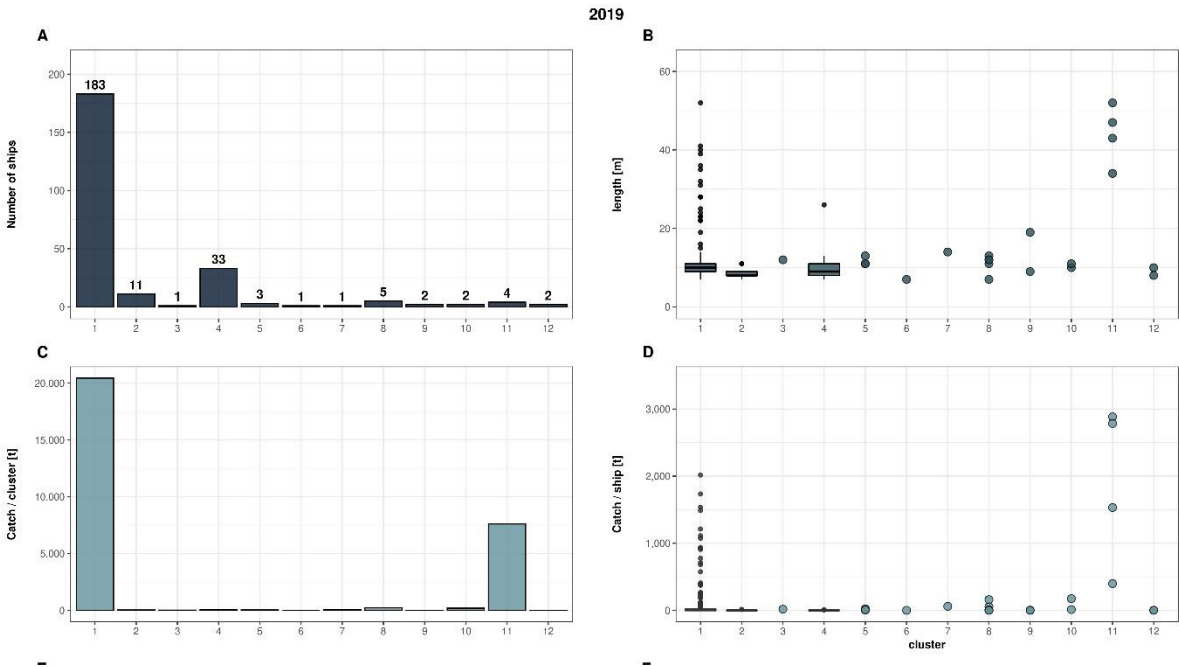


Fig 4. Grid of mixed plot types displaying vessel characteristics from gillnet fisheries.

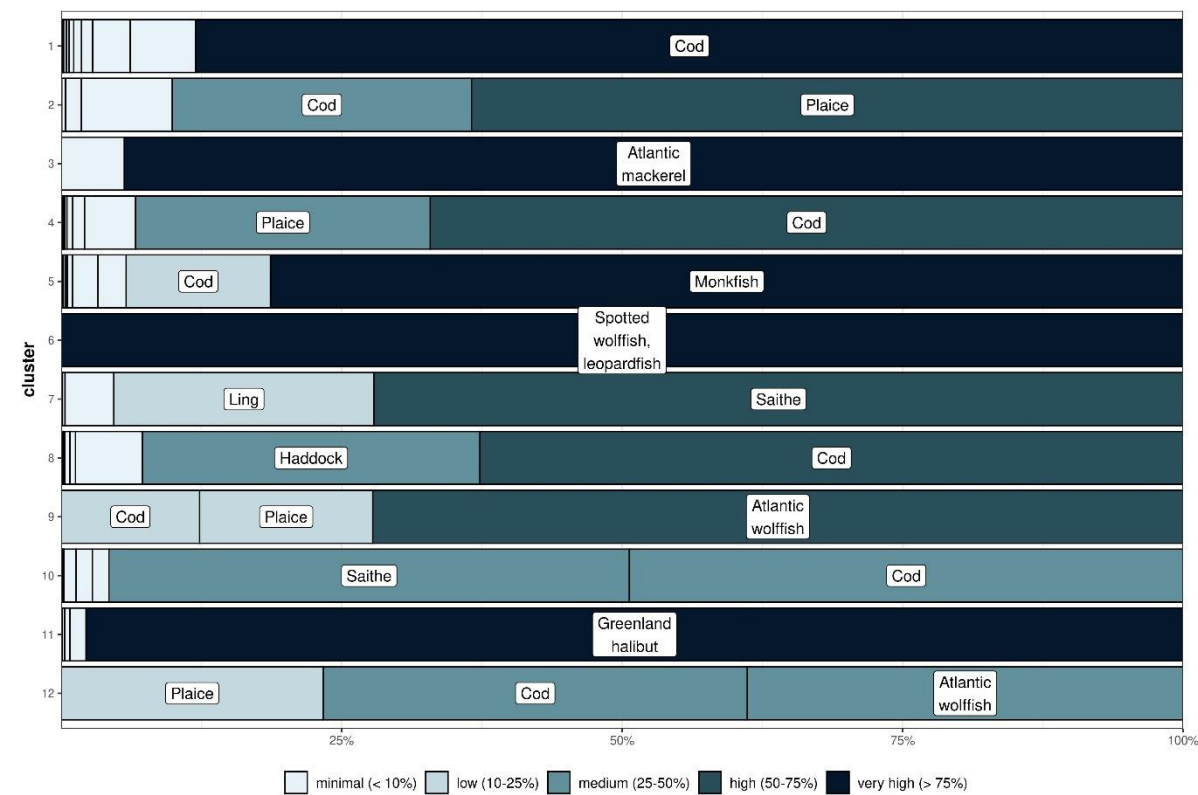


Fig 5. Barplot of the average percentage of each stock on the total catch of each cluster from gillnet fisheries.

Longliners and jiggers

A short view of the longliners and jiggers dataset. Despite the fact that those two fleets are not exactly similar we wanted to see if the package would distinguish between them.

```
Rows: 3,771
Columns: 3
$ ship_ID <chr> "253", "253", "253", "253", "253", "253", "253", "253"...
$ species <chr> "Atlantic wolffish", "Blue ling, European ling", "Cod"...
$ landings <int> 35753, 4213, 960753, 654, 64101, 30152, 7, 119, 7503...
```

We decided to use 9 clusters for this analysis even though the optimal number of suggested clusters was between 2-8. Here again we got one large cluster including majority of the vessels and cod is yet again playing the most important role in that cluster (Fig 7 and fig 10). This clustering result is not unusual as cod is also the main target species of jigger and longliner fleets in Icelandic waters. We can see that the procedure picks up jigger boats that have been targeting mackerel early in the process (cluster 3, fig 7 and fig 10) and the boats that had saithe in their catch composition which is a rather difficult species to catch and/or quite coincidental (cluster 2 and 5). After looking at the output for this dataset, we see that fewer clusters should be used to begin with and that there was no clear partition between the fleets in the clustering analysis. It would be interesting to run the dataset through the package on seasonal basis aimed at observing if the FS package picks up the seasonal jiggers coastal fisheries that start approximately around May, before analysing the data per gear.

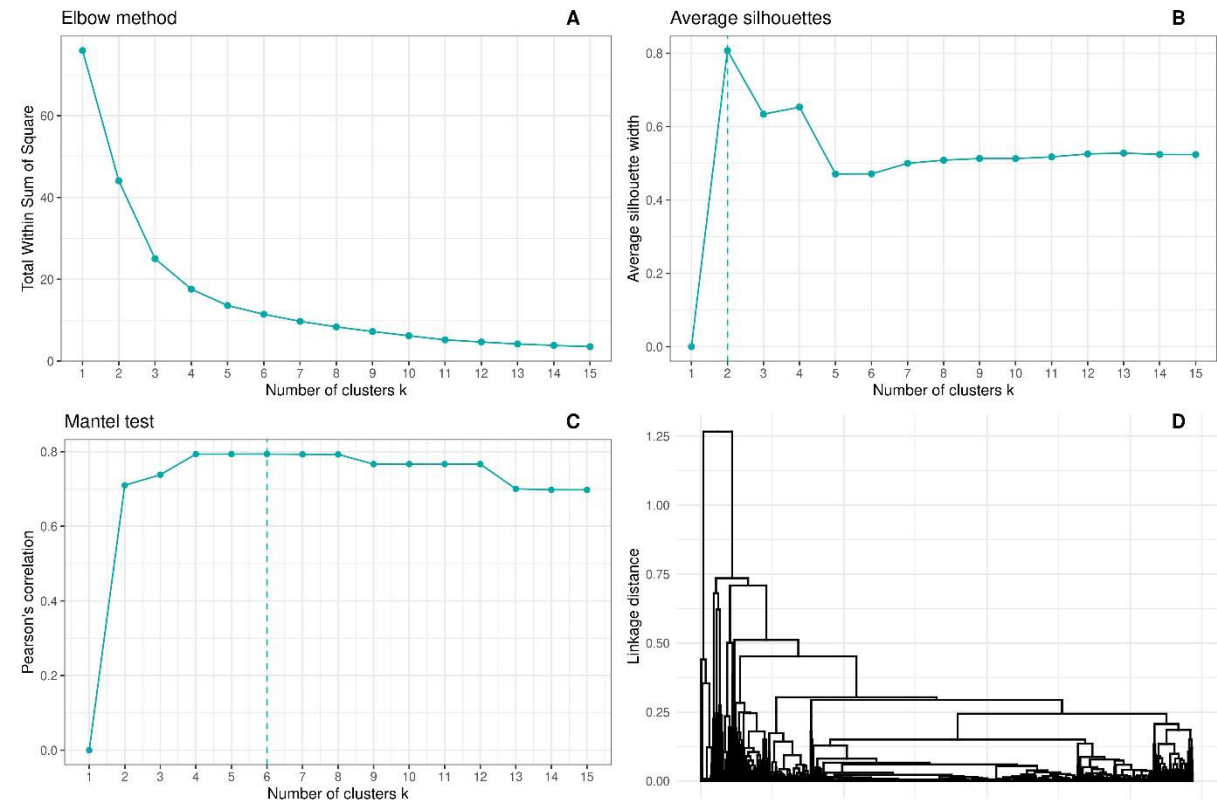


Fig 6. Grided diagnostics of HAC procedure for longline and jigger fisheries.

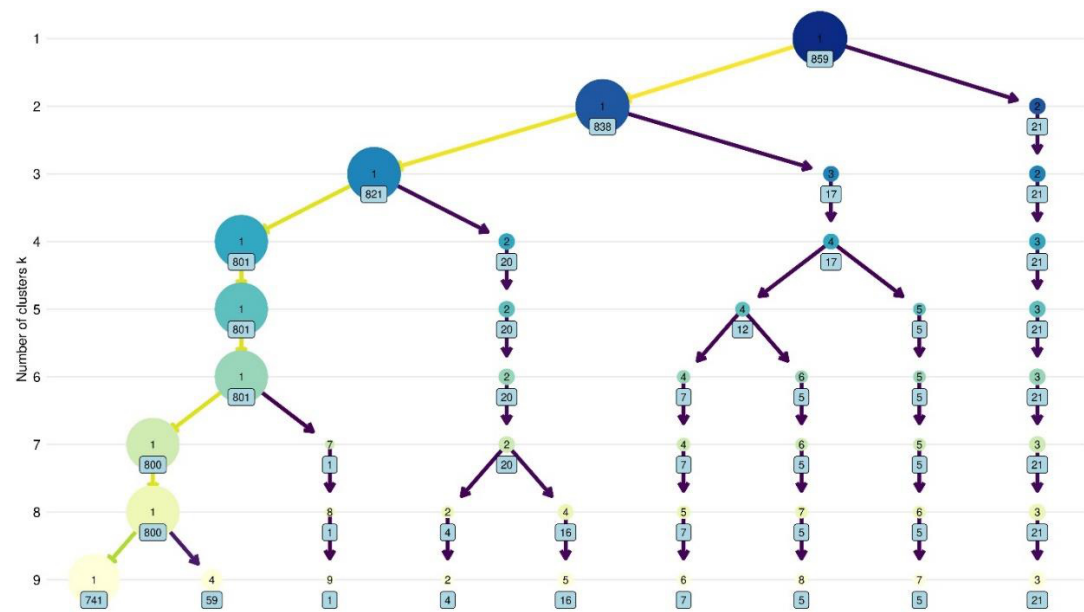


Fig 7. Treeplot of the HAC-procedure for longline and jigger fisheries.

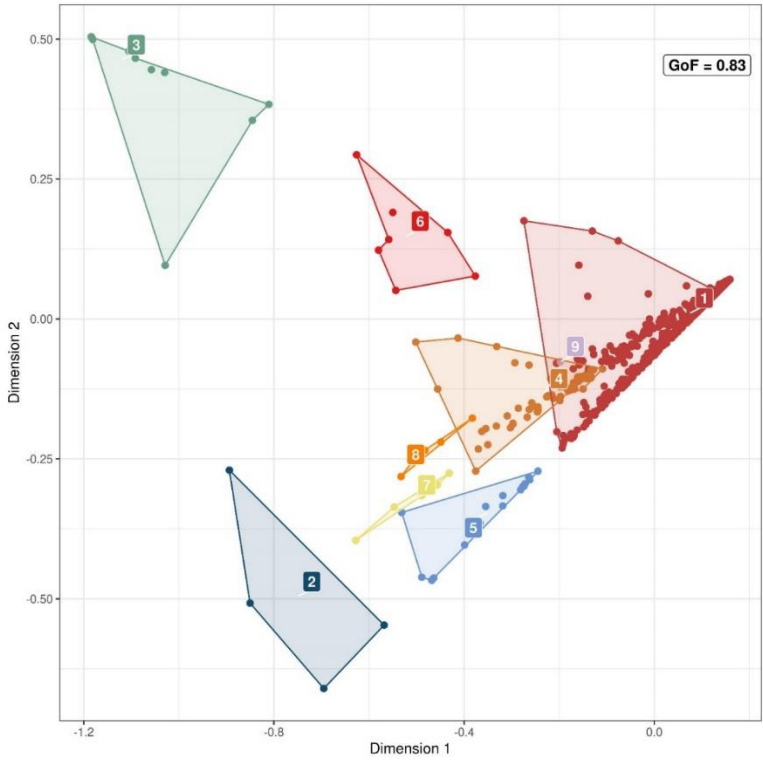


Fig 8. MDS of the transformed vessel catch data from longline and jigger fisheries.

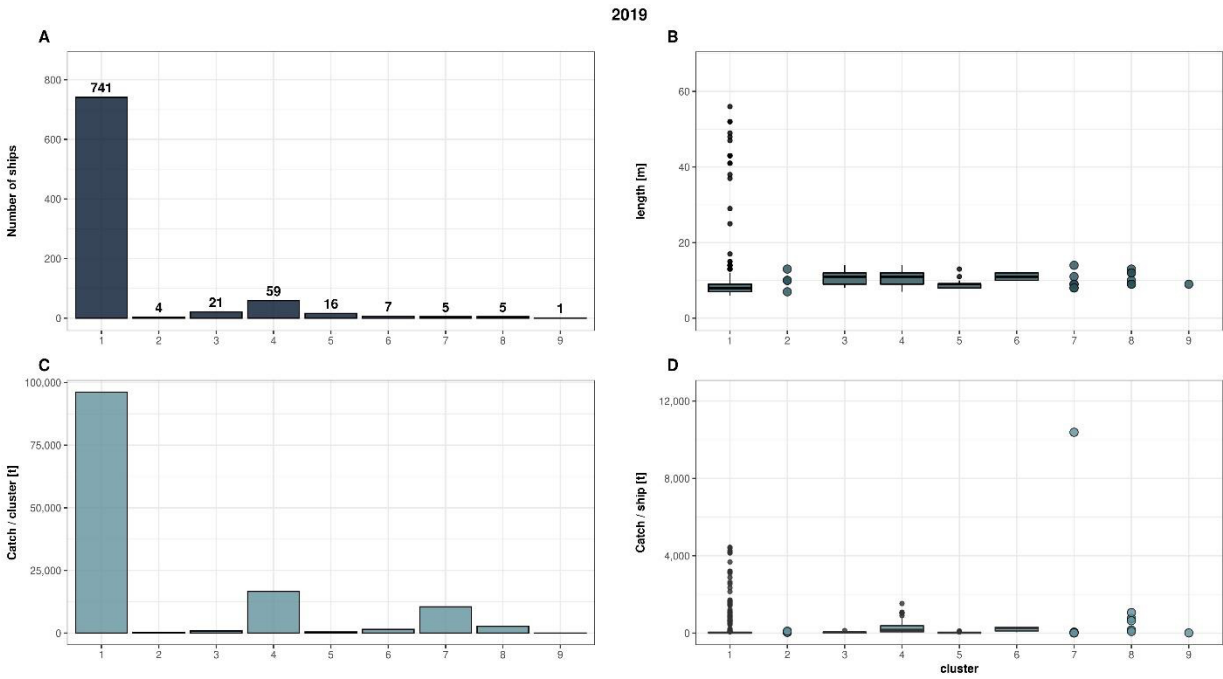


Fig 9. Grid of mixed plot types displaying vessel characteristics from longline and jigger fisheries.

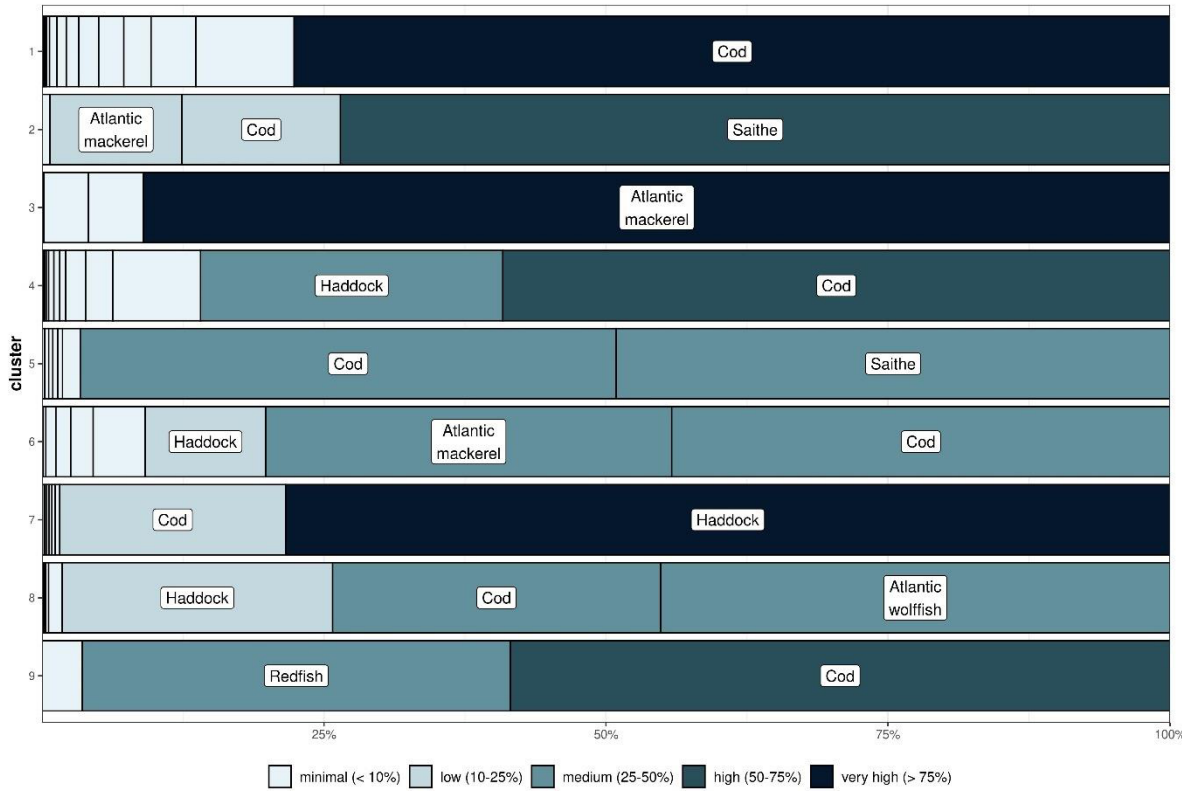


Fig 10. Barplot of the average percentage of each stock on the total catch of each cluster from longline and jigger fisheries.

Overall, the procedure identifies main fisheries for fleets investigated above. Gillnets and longline/jigger fisheries are quite seasonal and small scale in Icelandic waters with cod as the target species, therefore one large cluster with cod as main species is expected for both datasets.

Case .. :Italian fleet segmentation.

Description of analysis population

For the Italian fleet, the segmentation analysis was applied in the segmentation of:

- Hydraulic dredges (at national level)
- Purse seiners (at national level)
- Demersal trawlers (at national level)
- Demersal trawlers (at GSA level)

Basic data available

Landings per species: official logbooks

Length data: fleet register

Coverage: logbooks are available for all the vessels using active gears and for the vessels using passive gears with LOA>10 m.

Hydraulic dredges (DRB)

Data available: 554 vessels

Description of the fishery

Bivalve mollusc fishing with hydraulic dredges is mainly concentrated on the Adriatic sea. The main target species is the streped venus (*Chamelea gallina*), which is harvested with hydraulic dredges. Clams are short-lived, sedentary species and stock status can vary from year to year with environmental conditions as well as due to exploitation patterns. Clams make up 90% of the landings in this segment, 84% in revenues.

Hydraulic dredges are standardised boats from a technical and size perspective (average size 12 GT, average engine power 105 kW). In economic terms, the contribution of the hydraulic dredge segment to the gross saleable production of the entire Italian fisheries sector is 12%.

DCMAP fleet segmentation for hydraulic dredges

GSA	LOA	No. vessels
9	VL1218	3
10	VL1218	14
17	VL1218	552
18	VL1218	39
Total		608

Alternative fleet segmentation

The initial analysis on the hydraulic dredges fleet segment resulted in the grided diagnostics shown in Fig 1a. As shown in in 1c and in dendrograms in Fig 2a, the optimal number of clusters is reached at 7.

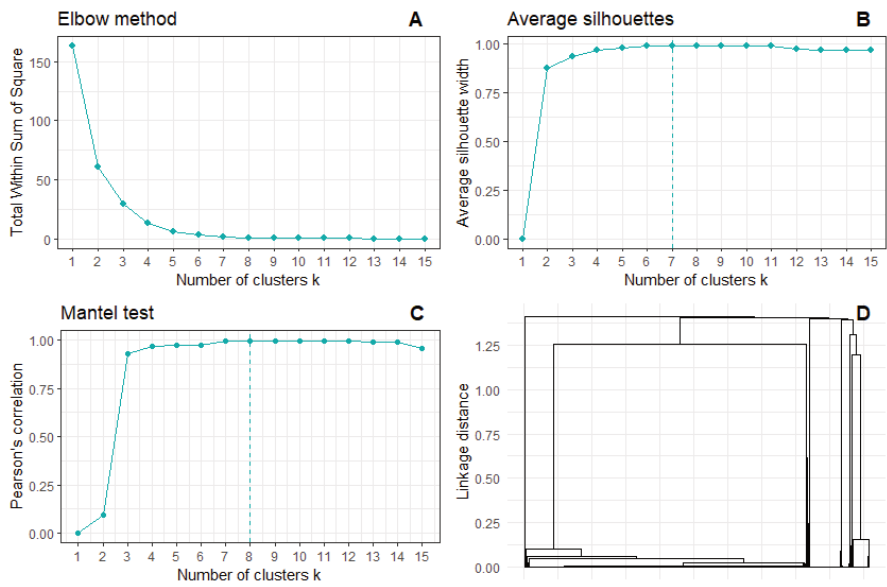


Fig 1a: Grided diagnostic plots of the HAC procedure for **hydraulic dredges**.

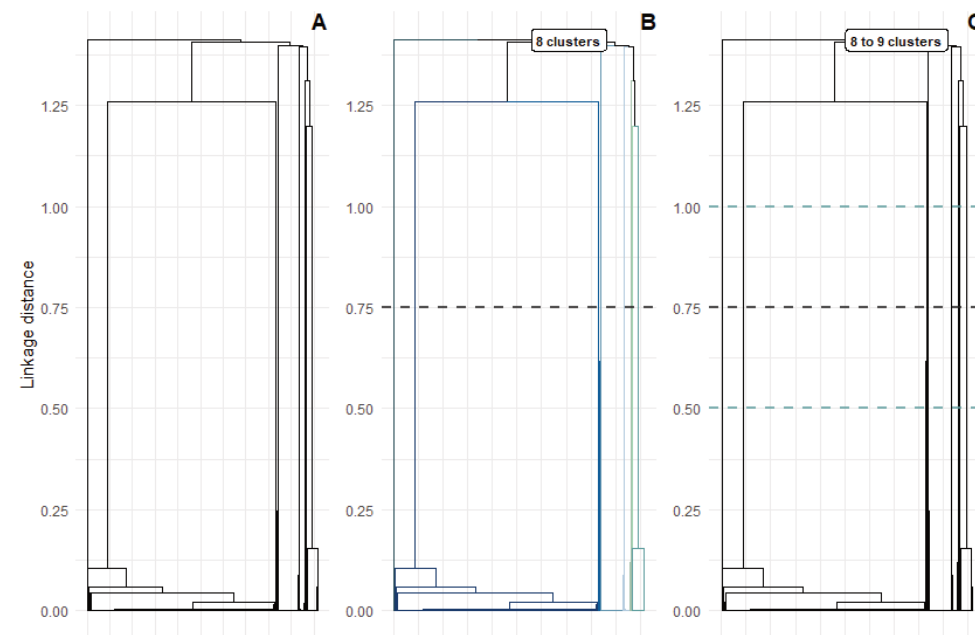


Figure 2a: Grided plot of three dendrograms of the HAC procedure for **hydraulic dredges**.

The clustering tree (fig 3a) shows one cluster with the majority of the vessels (413 on a total of 554).

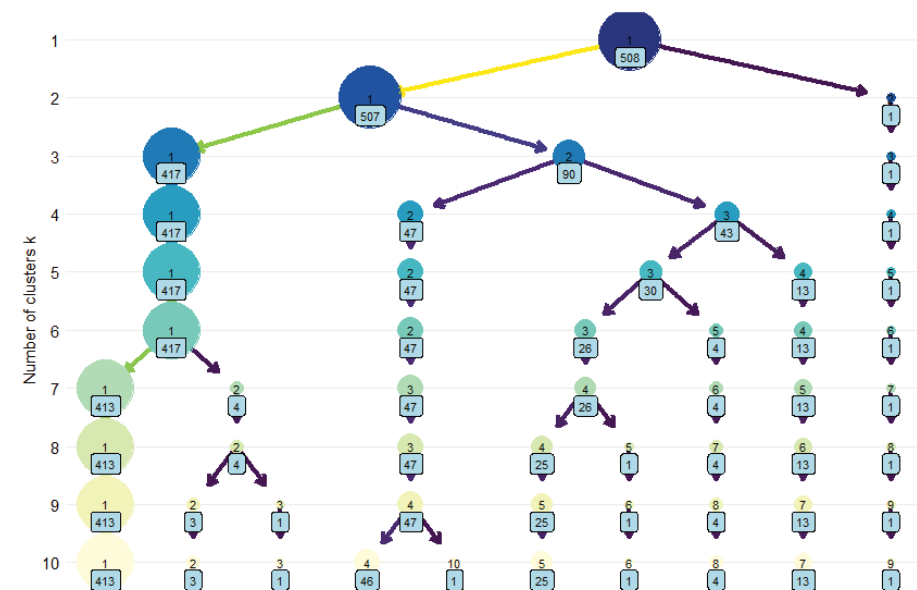
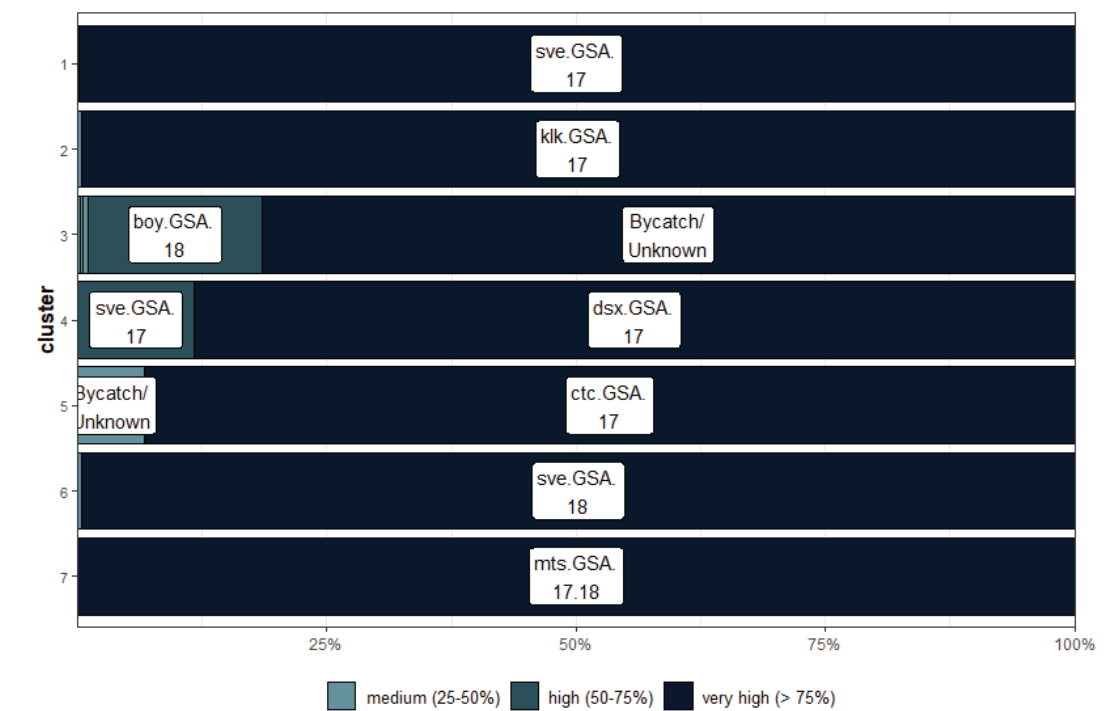


Figure 3a : Treeplot of the HAC-procedure for **hydraulic dredges**.

The species composition of the various clusters shows a low variability in consideration of the concentration of the landings in bivalve molluscs (fig. 4a).



4a: Barplot of the average percentage of each stock on the total catch of each cluster of **hydraulic dredges**

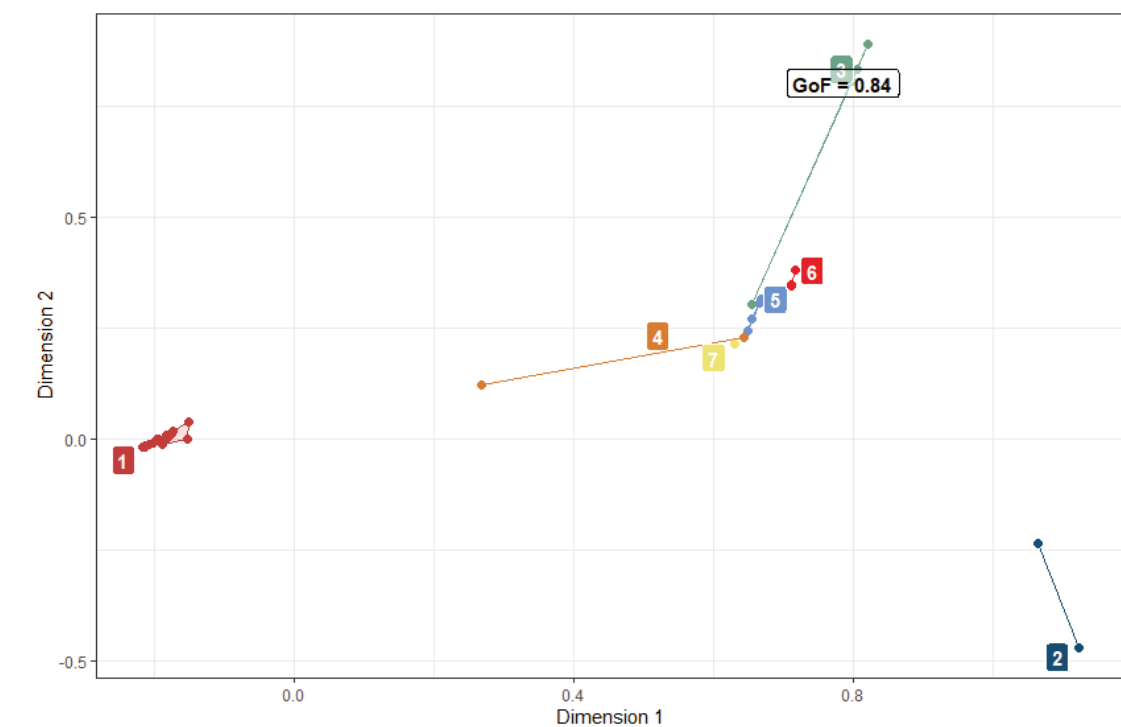


Figure 5a: MDS of the transformed vessel catch data for **hydraulic dredges**.

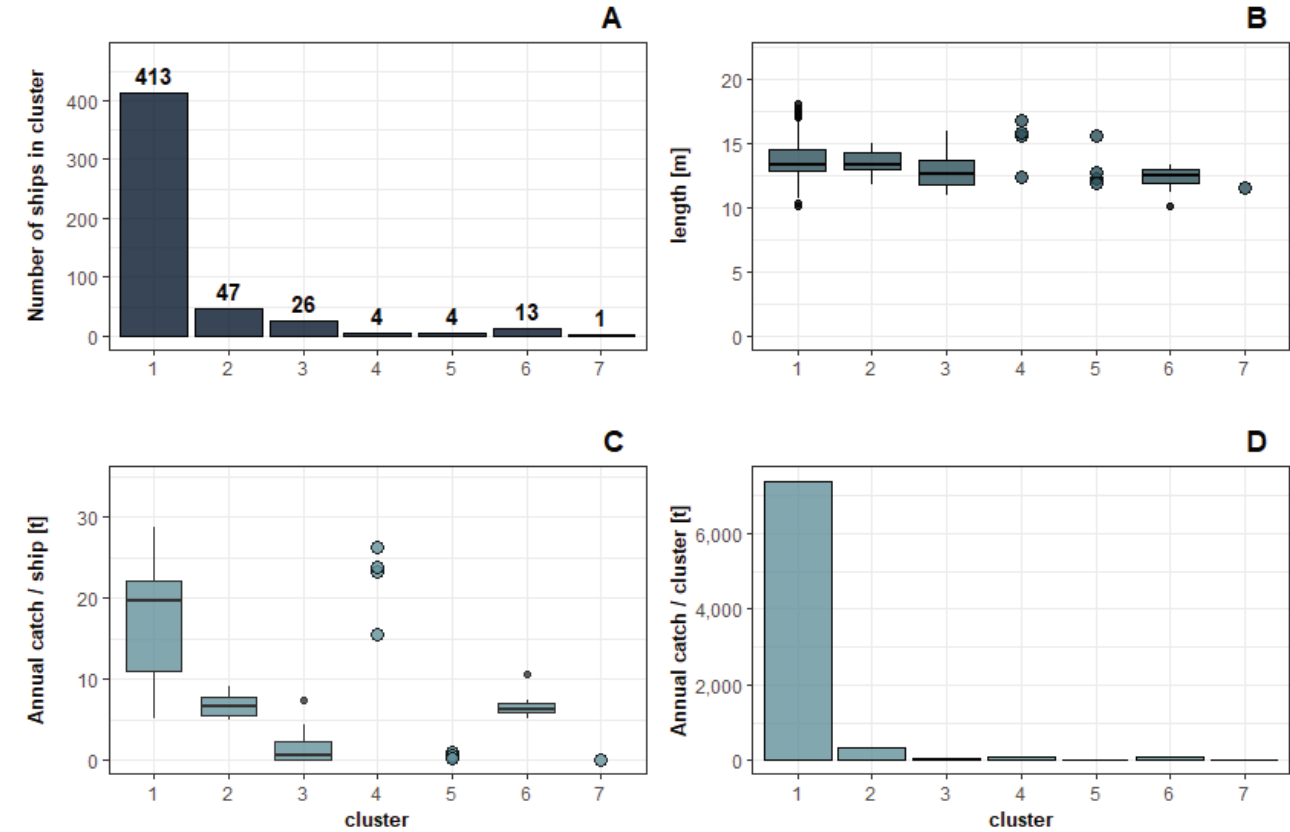


Figure 6a: Grid of mixed plot types displaying vessel characteristics for **hydraulic dredges**.

Characterization of the fleet segments

Cluster 1: hydraulic dredges in Northern Adriatic Sea, target specie Striped venus

Cluster 2: hydraulic dredges in Northern Adriatic Sea, target specie Smooth callista

Cluster 3: undefined catch profile (unknown); in the cluster we find hydraulic dredges in Tyrrhenian sea targeting different bivalve molluscs (Cockles nei, Donax clams, European razor clam) but also hydraulic dredges in Southern Adriatic Sea targeting other species. The annual catch of the cluster is very low (fig.6a.D)

Cluster 6: hydraulic dredges in Southern Adriatic Sea, target specie Striped venus

Cluster 4, 5 and 7: different gears used in the year (the target specie is not the same of the other clusters)

Purse seiners (PS)

Data available: 213 vessels

Description of the fishery

The purse seiners are widespread all along the Italian coastline, with a major concentration on Southern Tyrrhenian sea. The target species were small-pelagics, in particular anchovy and pilchard, with the exception of the biggest purse seiners authorized to exploit bluefin tuna. The segment includes very different typology of vessels with an average LOA that can range from 10 to 40 metres; the average crew size ranges from 4 to 15 fishermen. The cost structure differs accordingly.

DCMAP fleet segmentation for purse seiners

GSA	LOA	No. vessels
	VL0612	1
	VL1218	24
	VL1824	11
	VL2440	10
9		46
	VL0612	124
	VL1218	74
	VL1824	18
	VL2440	10
	VL40XX	6
10		232
	VL1218	6
	VL1824	4
11		10
	VL0612	2
	VL1218	9
	VL1824	9
	VL2440	7
	VL40XX	0
16		27
	VL0612	3
	VL1218	7
	VL1824	0
	VL2440	14
17		24
	VL1218	1
	VL2440	6
	VL40XX	1
18		8
	VL0612	9
	VL1218	19
	VL1824	6

	VL2440	3
	VL40XX	1
19		38
Total		385

Alternative fleet segmentation

The initial analysis on the purse seiners fleet segment resulted in the grided diagnostics shown in Fig 1b. As shown in 1c the optimal number of clusters is not reached before 15.

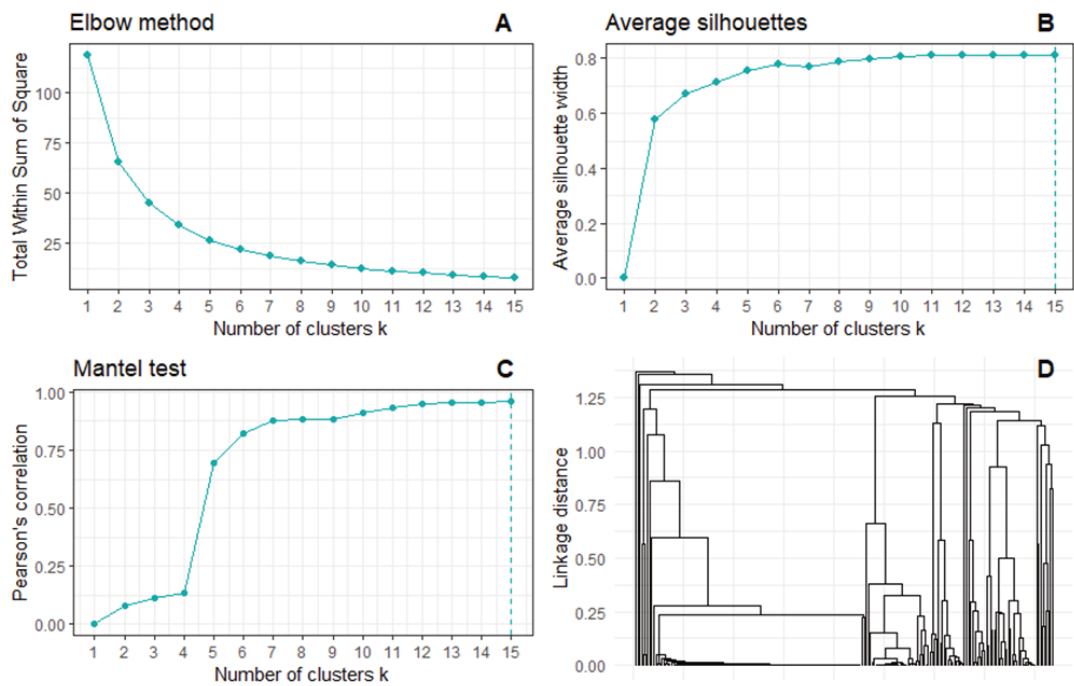


Fig 1b: Grided diagnostic plots of the HAC procedure for **Purse seiners**.

The fact that the optimal clusters is large for the PS fleet segment is also illustrated by the grided plots of the dendrograms in Fig 2b. For this fleet segment the optimal clusters is between 16-27 clusters.

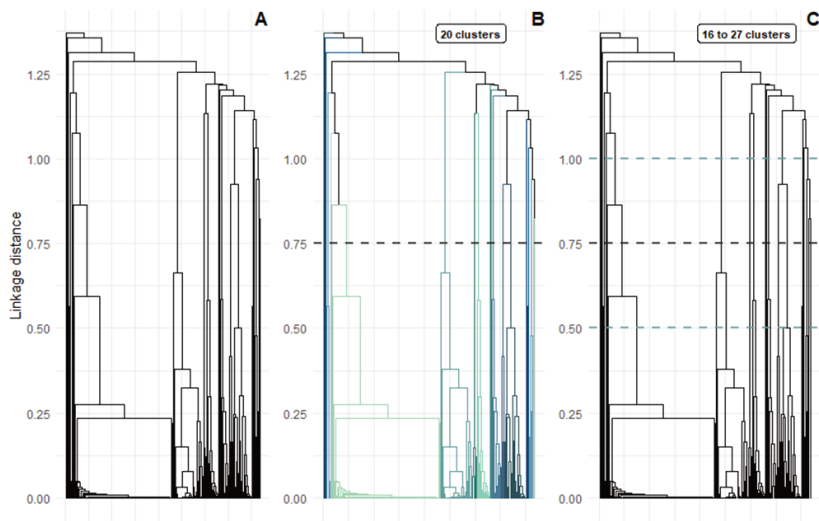


Figure 2b: Grided plot of three dendrograms of the HAC procedure for **Purse seiners**.

The clustering tree (fig 3b) shows a large number of very small clusters consisting of one or two vessels.

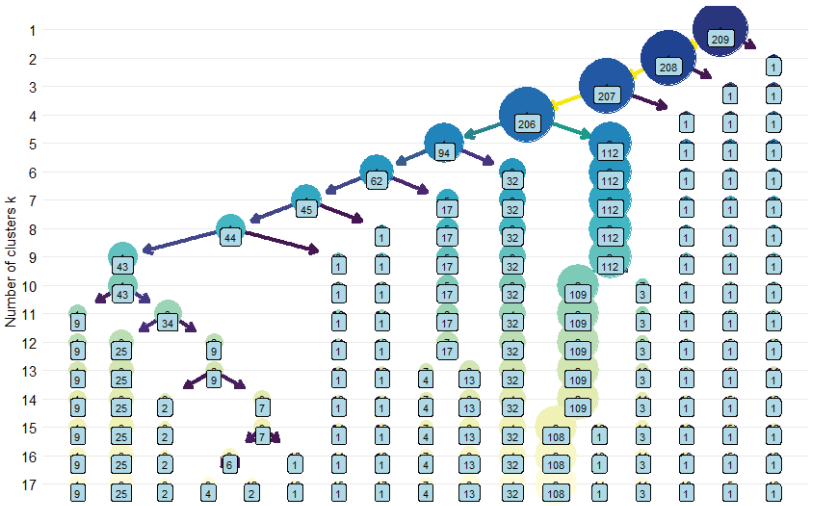


Figure 3b : Treeplot of the HAC-procedure for **Purse seiners**.

The species composition of the various clusters shows a low variability in fish stocks targeted by vessels of the purse seiners (fig. 4b). In order to interpret these graphs it is good to consider the number of vessels in each of the clusters as shown in fig.7b.

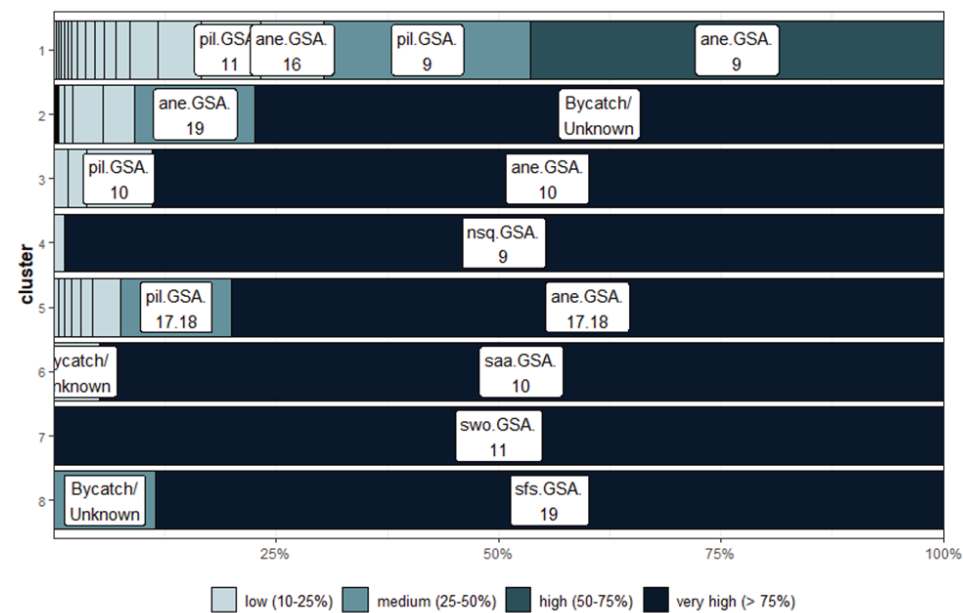


Figure 4b: Barplot of the average percentage of each stock on the total catch of each cluster of the purse seiners.

Figures 5b shows the results of the clustering in MDS figures.

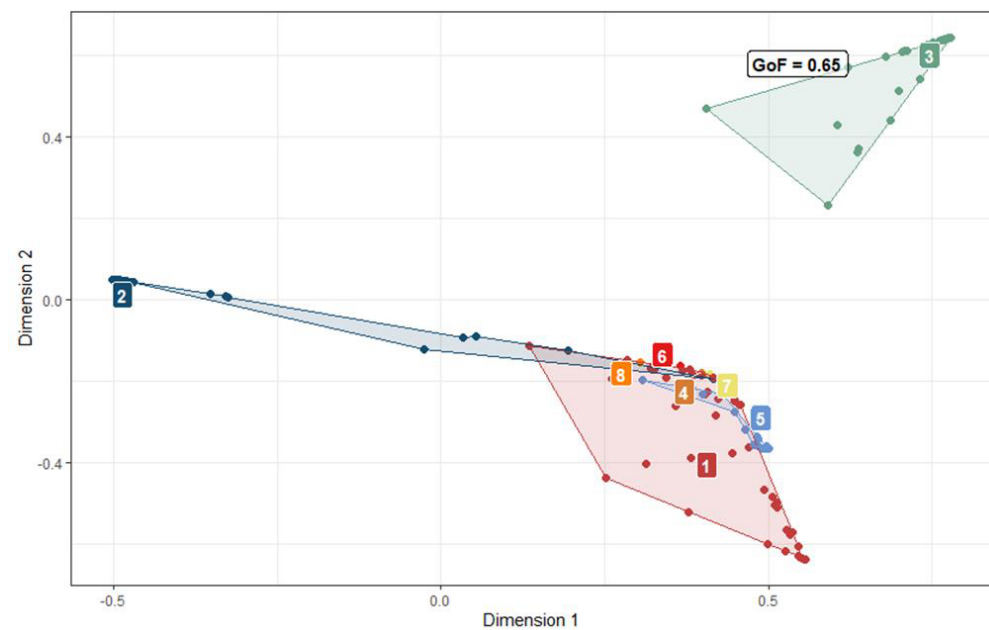


Figure 5b: MDS of the transformed vessel catch data for **purse seiners**.

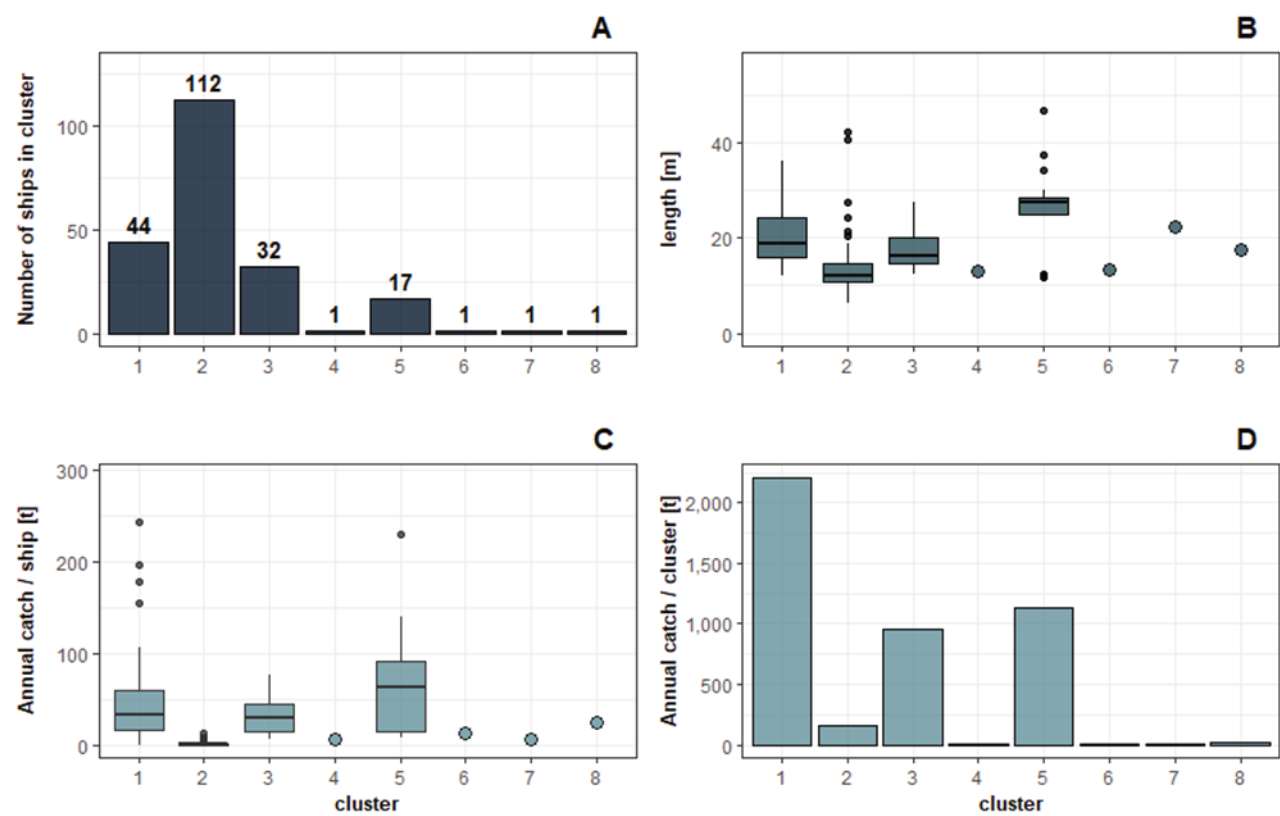


Figure 6b: Grid of mixed plot types displaying vessel characteristics for **purse seiners**.

Characterization of the fleet segments

Cluster 1: two species, three different GSAs: we lost information at GSA level

Cluster 2: high number of vessels with low annual catch volume; we have not a defined catch profile (unknown); cost structure is not the same for all the vessels in the cluster as a consequence of the different size (graf. 7a.B)

Cluster 5: small pelagic fishery in Adriatic Sea

Cluster 4, 6, 7 and 8: only 1 vessel; different gear used in the year (the target specie is not the same of the other clusters)

Demersal trawlers (DTS)

Data available: 1,715 vessels

Description of the fishery

Demersal trawler is the most productive fleet segment in terms of landings value; the fleet targets high-value demersal stocks, such as deep-water rose shrimp, European hake, red mullet, giant red shrimp, spottail mantis squillid, deep-water rose shrimp, common cuttlefish and Norway lobster. Combined, these seven species make up about 55% percent of the total value of landings

DCMAP fleet segmentation for demersal trawlers

GSA	LOA	No. vessels
9	VL0612	30
	VL1218	135
	VL1824	114
	VL2440	12
		291
10	VL0612	16
	VL1218	132
	VL1824	56
	VL2440	13
		217
11	VL0612	14
	VL1218	65
	VL1824	30
	VL2440	20
		129
16	VL0612	19
	VL1218	136
	VL1824	122
	VL2440	87
		364
17	VL0612	61
	VL1218	248
	VL1824	169
	VL2440	51
		529
18	VL0612	30
	VL1218	275
	VL1824	66
	VL2440	11
		382
	VL0612	14

VL1218	189
VL1824	28
VL2440	6
19	237
Total	2149

Alternative fleet segmentation

The initial analysis on the hydraulic dredges fleet segment resulted in the gridded diagnostics shown in Fig 1c. As shown in fig. 1c, the optimal number of clusters is not reached before 15

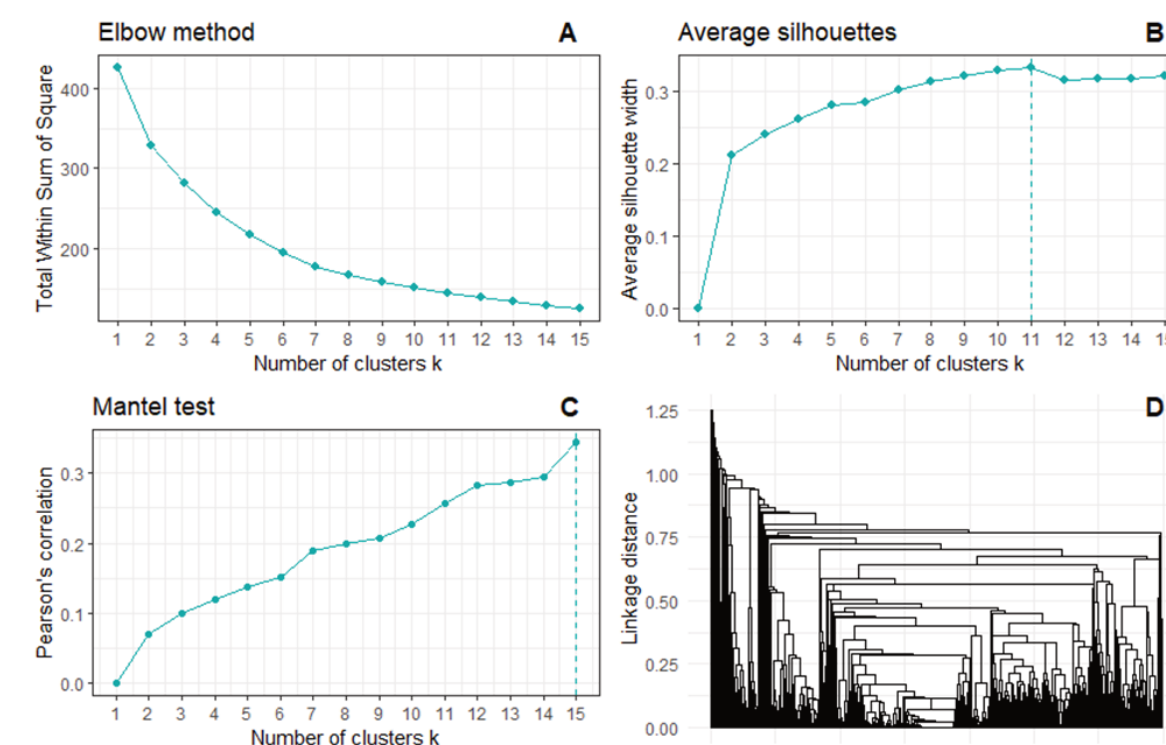


Fig 1c: Gridded diagnostic plots of the HAC procedure for **Demersal trawlers**.

The fact that the optimal clusters is large for the DTS is also illustrated by the gridded plots of the dendrograms in Fig 2c. For the complete fleet the optimal clusters are between 16-68 clusters

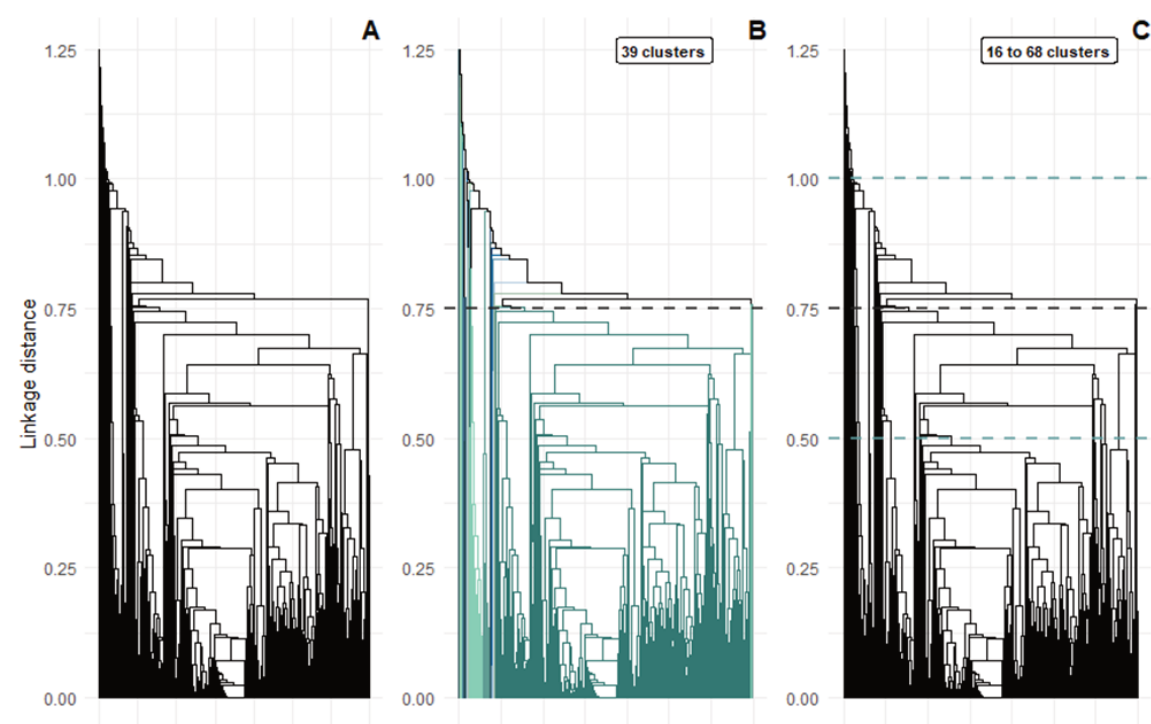


Figure 2c: Grided plot of three dendrograms of the HAC procedure for **Demersal trawlers**.

The clustering tree (fig 3c) also confirms that the clustering exercise for the complete fleet results in a large number of very small clusters consisting of one or two vessels and only 3 clusters consist of more than 3 vessels (minimal for EUmapp).

This outcome could be due to the high number of combinations species_GSA; for this reason, it was decided to apply the methodology to a subset of data, namely the demersal trawlers in GSA16.

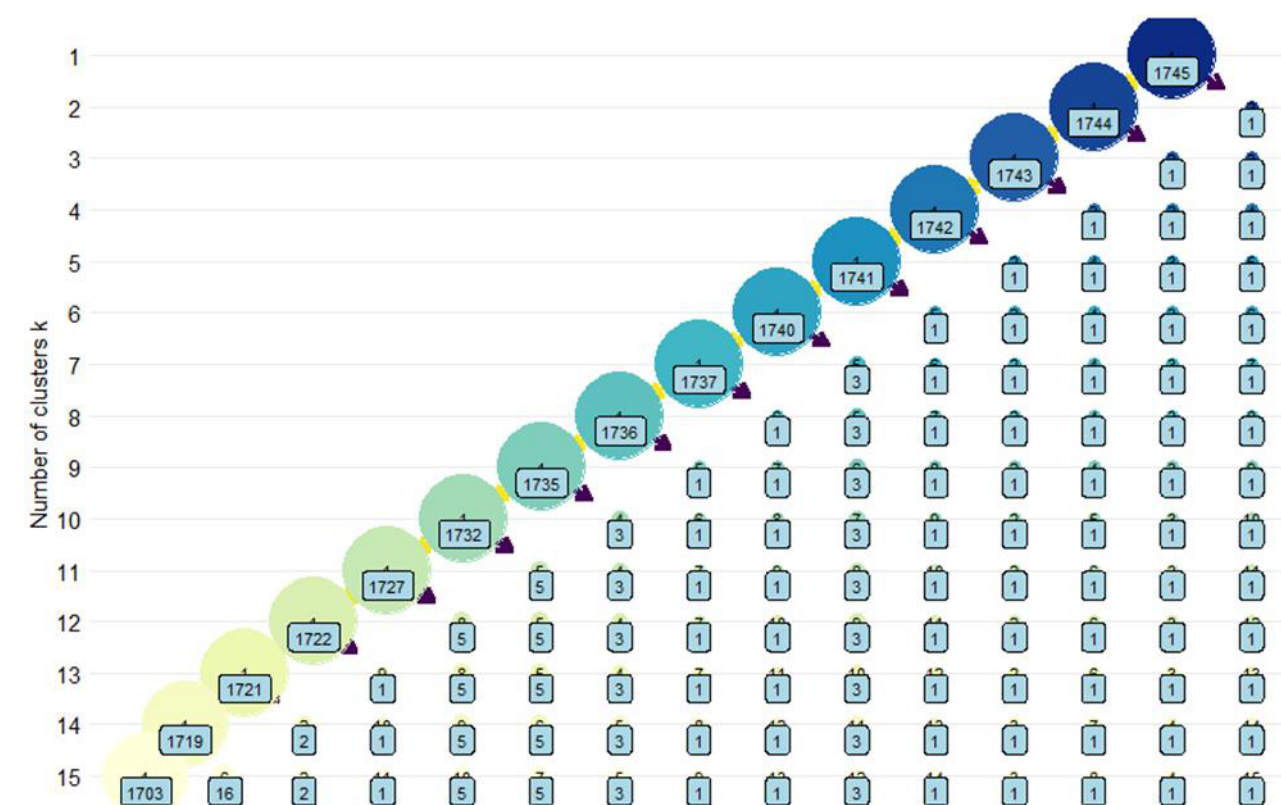


Figure 3c : Treeplot of the HAC-procedure for **demersal trawlers**.

Demersal trawlers (DTS) in GSA 16

Data available: 314 vessels

Alternative fleet segmentation

On the basis of the analysis of fig.1d and fig.2d, we decided that the optimal number of clusters is reached at 10.

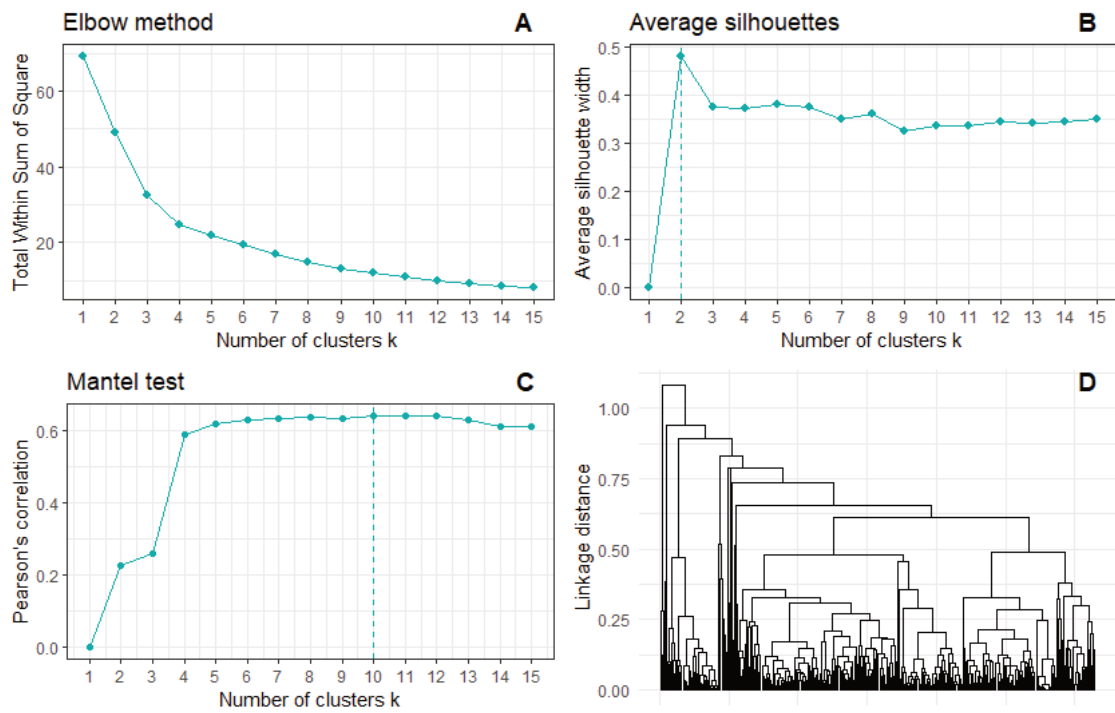


Fig 1d: Gridded diagnostic plots of the HAC procedure for **demersal trawlers in GSA16**.

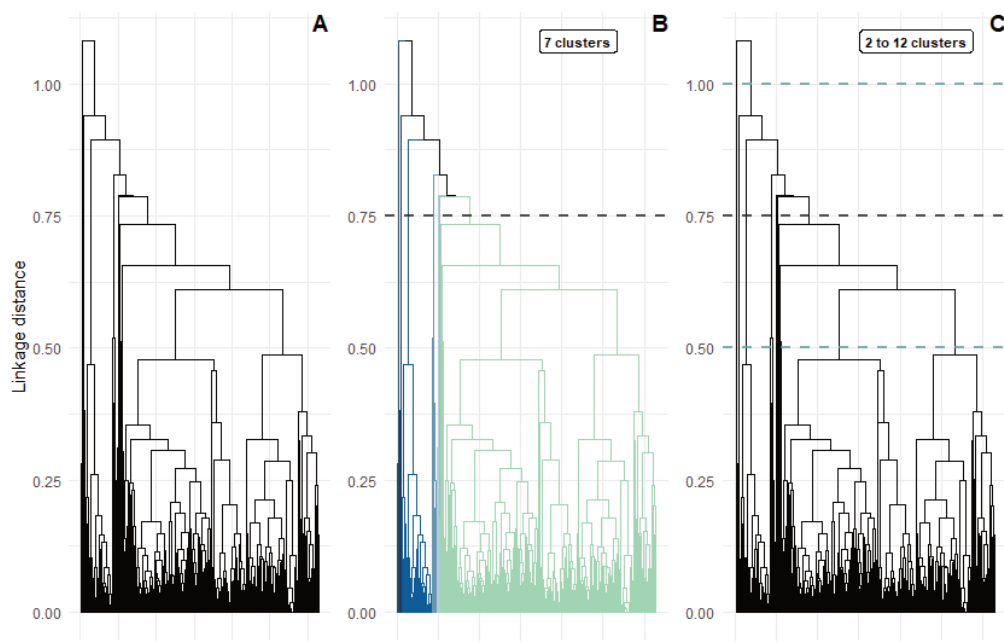


Figure 2d: Gridded plot of three dendrograms of the HAC procedure for **demersal trawlers in GSA16**

On 10 clusters, 3 consisted of an high number of vessels (respectively 96, 162 and 37), while the others are very small (less than 3 vessels).

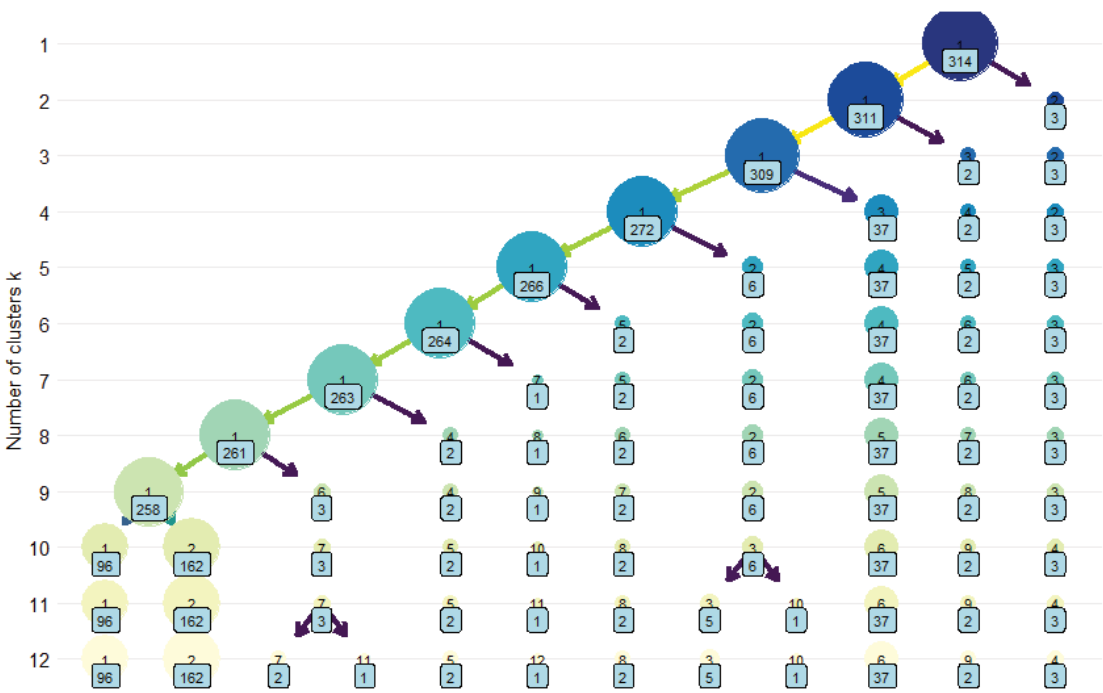


Figure 3d : Treeplot of the HAC-procedure for **demersal trawlers in GSA16**.

The species composition of the various clusters shows the high variability in fish stocks targeted by vessels of the demersal trawlers in GSA16 (fig. 4d). In the total demersal trawlers in GSA16 fleet (fig. 4d) the main clusters are 1,3 and 6. The catch profile of cluster 1 is not clear, while cluster 2 includes vessels targeting both demersal species (hake) and crustaceans (shrimps). Cluster 6 is better defined and it includes vessel targeting crustaceans.

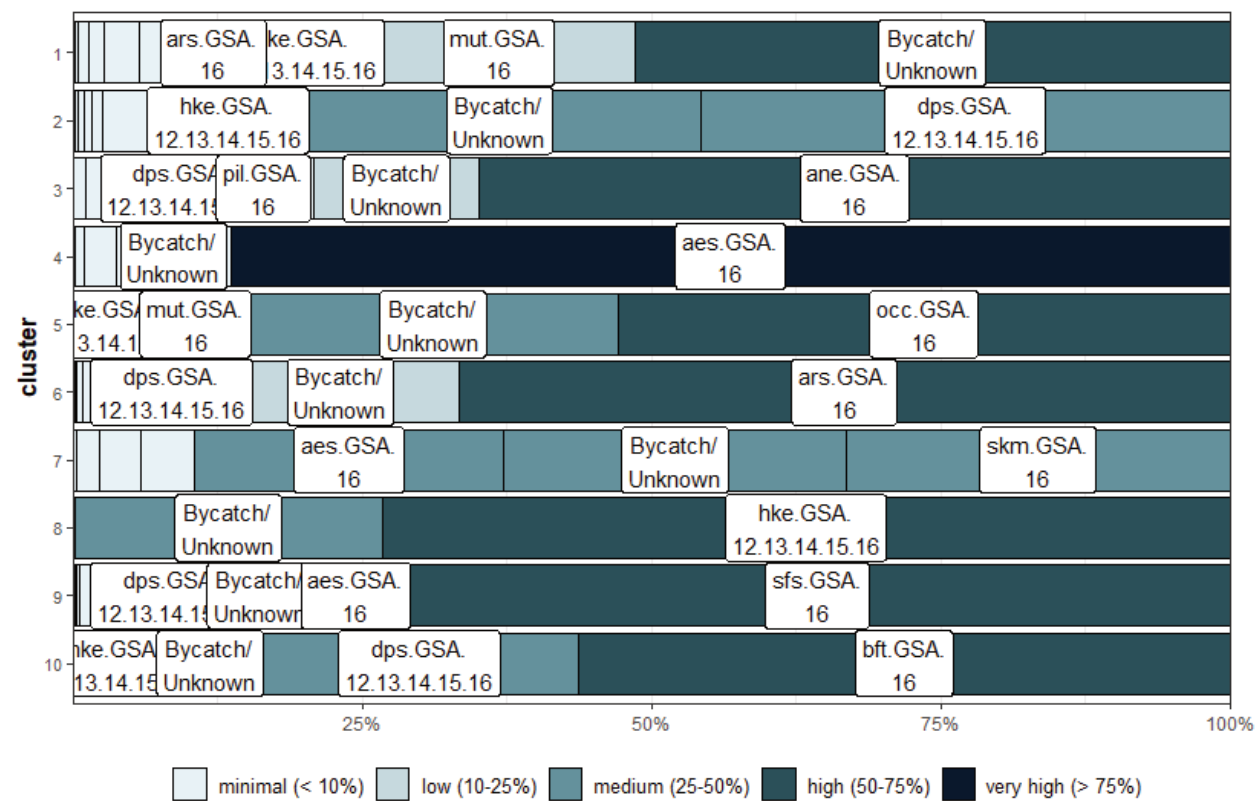


Figure 4d: Barplot of the average percentage of each stock on the total catch of each cluster of the demersal trawlers in GSA16.

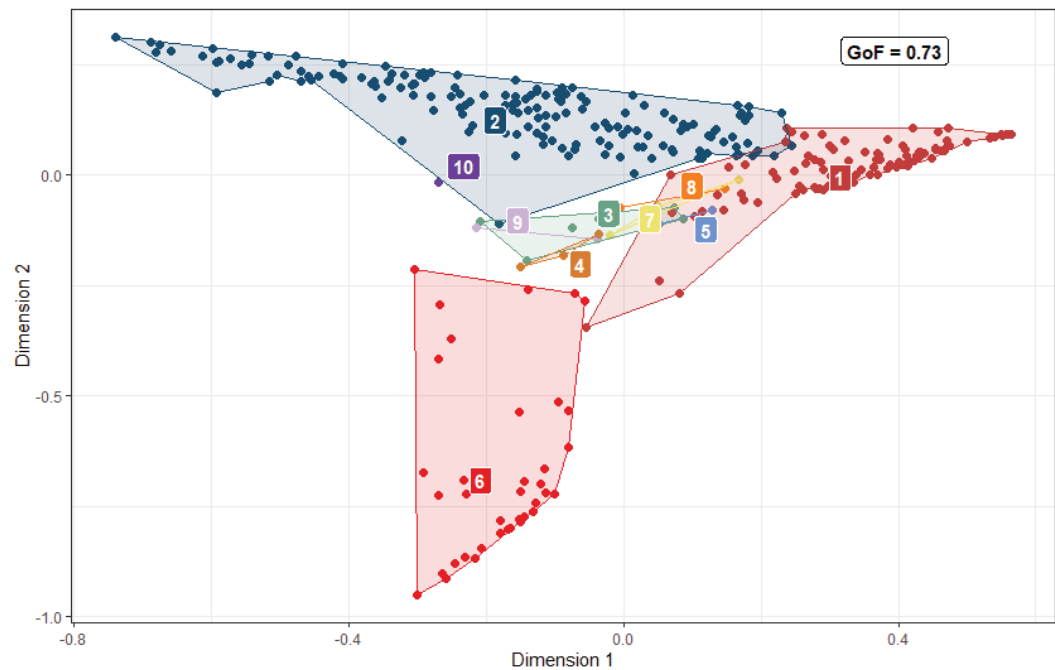


Figure 5d: MDS of the transformed vessel catch data for demersal trawlers in GSA16

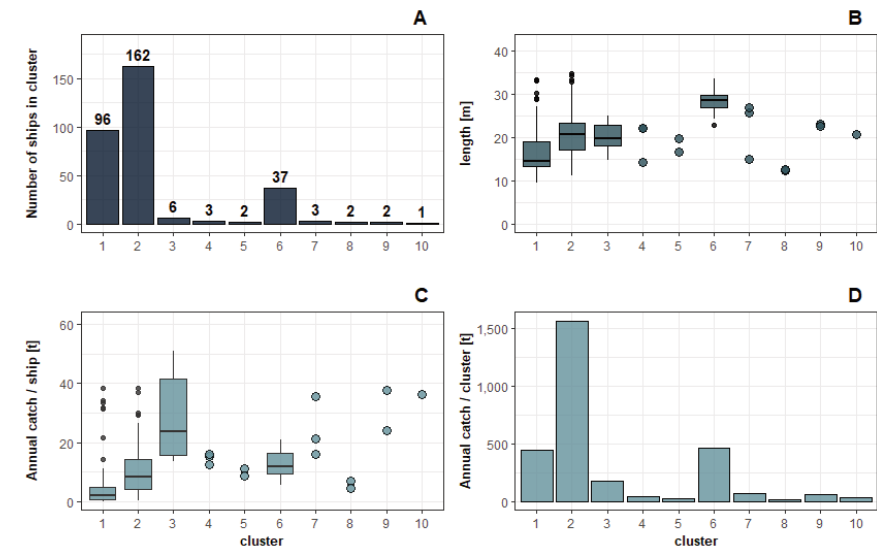


Figure 6d: Grid of mixed plot types displaying vessel characteristics for demersal trawlers in GSA16.

For the trawler sector in GSA16, the species composition of the various clusters shows some of the major fisheries (cluster 6) but other clusters are not clearly identified. In the cluster 2 as reported in fig. 6d.B, the size of the vessels differs greatly (from 10 metres to 35 metres) and this can have a direct impact on the cost structure.

Characterization of the fleet segments

Cluster 1: undefined catch profile

Cluster 2: high number of vessels with high annual catch volume; we have not a defined catch profile (unknown); cost structure is not the same for all the vessels in the cluster as a consequence of the different size (graf. 7d.B)

Cluster 6: shrimp fishery in Strait of Sicily

Other clusters: few vessels per cluster; different gears used in the year (the target specie is not the same of the other clusters)

Conclusions

The novel approach to the segmentation of the fishing fleet at Italian fishery sectors could improve the economic analysis in terms of:

- highlight some important fisheries in terms of revenue (this is the case for example of the cluster 6 in the DTS_GSA16)
- split the fleet segments on the basis of the target species (this is the case of cluster 1 and 2 in DRB).

In general, the new approach is in line with the new management orientation in Mediterranean sea, based on the stocks and non on the technical aspects of the vessels (i.e WestMedPlan, multi-annual plan for small pelagic stocks in the Adriatic sea, limitation of fishing capacity for giant red shrimp and blue and red shrimp in the Ionian Sea in Gsa 19 and Gsa 16).

The application of the procedure to Italian data leads to some issues; among these:

- data availability: official logbooks are the source of information; they are available for all the vessels using active gears and for the vessels using passive gears with LOA>10 m (3,999 active vessels on a total of 10,874 vessels); the vessels using passive gears (6,875 active vessels) are therefore excluded from the analysis.
- representativity of the data: mis-reporting of catches and imputation errors of the species landed may occur in some cases; this last issue is due to the large range of species and combinations between species and gears.

From the four analyses reported, we can conclude:

- for the fleet segments with highly variable catch compositions (i.e. DTS), the procedure fails to identify consistent clusters; the output is a lot of clusters with one of these with the majority of the vessels and an huge number of clusters very small (less than 3 vessels)
- the catch profile is not always well defined inside the cluster with the category “unknown” that occurs frequently.
- in some cases, we have clusters with the same catch composition but different effort and cost structures (different size of the vessels)

In order to improve the analysis, the procedure should be applied at GSA level rather than national level. In this case, the number of combinations species_fishing area decreases (sea DTS in GSA 16).

Next steps:

logbook data should be deeply explored to detect and correct any inconsistencies (i.e. check on the species and combination species_gear)

Apply the procedure at GSA level and to other fleet segments (TBB, TM)

Case .. :Dutch fleet segmentation.

Description of analysis population

For the Dutch fleet, the segmentation analysis was applied in two ways:

- Segmentation of the overall population of active vessels
- Segmentation of the so called Cutters (vessels using active gears targeting fish) and the pelagic trawlers.

Basic data available

For all vessels landings per species are available from official logbooks, together with the length data of the vessels

Best number of clusters

The initial analysis on the complete fleet resulted in the grided diagnostics shown in Fig 1a. As shown in 1a the Dutch fleet consists of a wide variety of vessels targeting multiple species. As a result the clustering results in a large number of very small clusters of vessels that target specific species (e.g. razor clams, surf clams, crab, mullets). As the species composition from these vessels differs a lot from the other vessels, they are situation in their own cluster, where as most of the other vessels using active gears are positioned in a mixed cluster. Because of this the optimal number of clusters is not reached before 15.

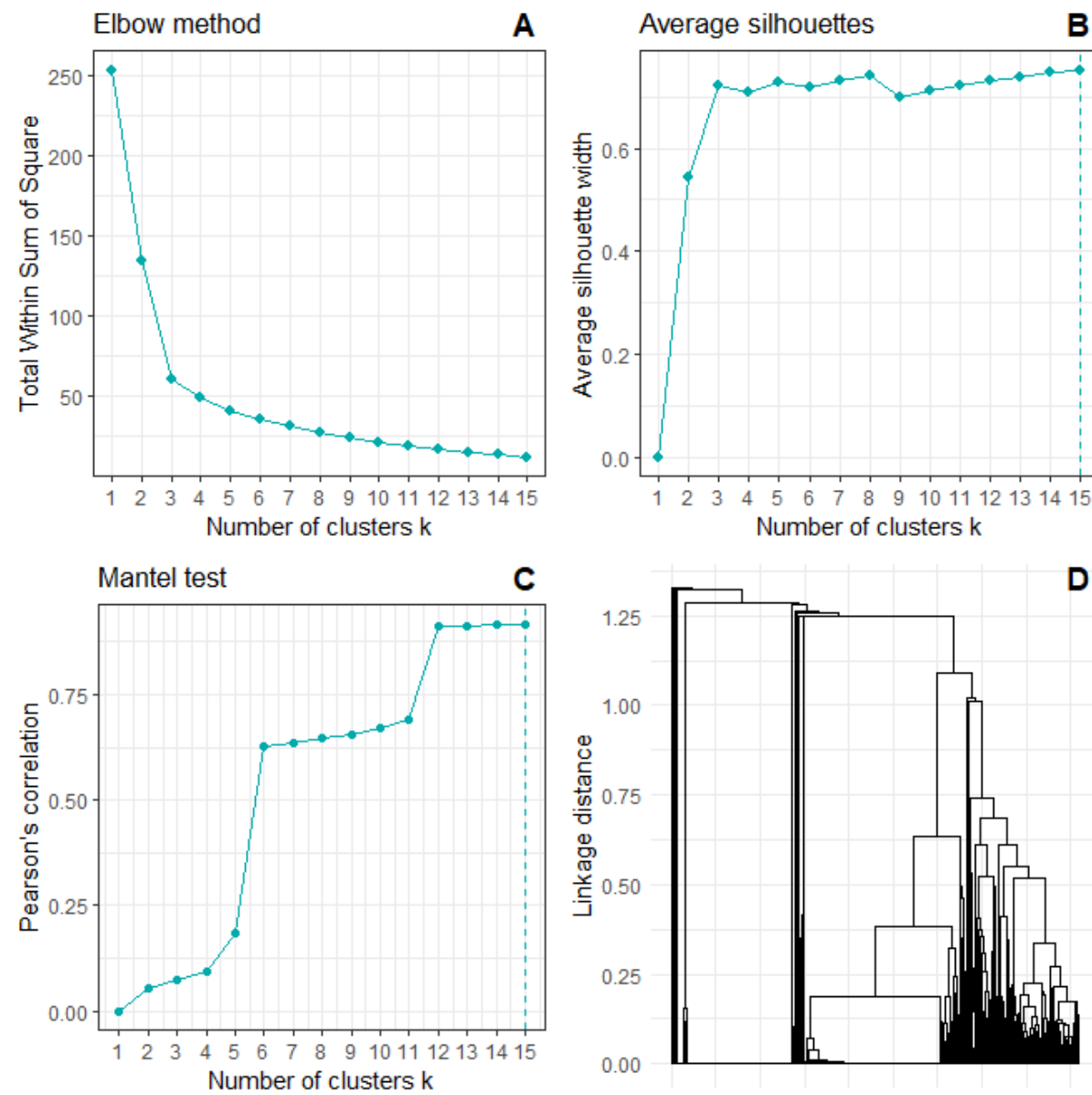


Fig 1a: Gridded diagnostic plots of the HAC procedure for **the complete Dutch fleet**. A) Scree plot showing the total within sums of squares of the clustering vs. the number of clusters used. B) Average silhouette width of the clusters vs. the number of clusters. C) Mantel test, i.e. the Pearson correlation between the clustering and the original distance matrix vs. the number of clusters. In B) and C), the best score is marked with a vertical line. D) dendrogram of the HAC procedure with the linkage distance of the clusters on the y-axis.

As for the cutter and trawler fleet the groups are more homogeneous and the optimal number of clusters is somewhere between 2 and 8 (fig 1.b).

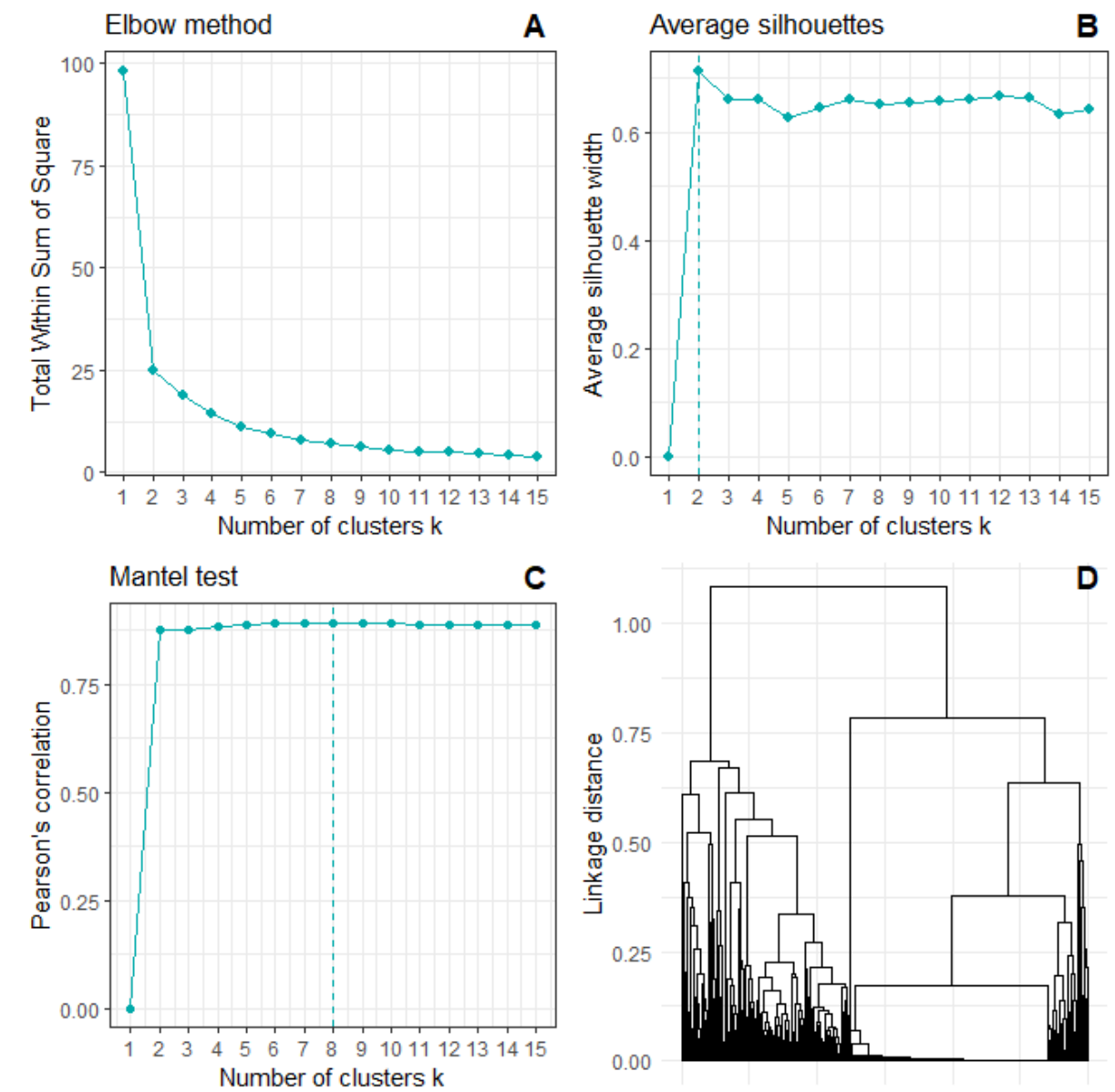


Fig 1b: Gridded diagnostic plots of the HAC procedure for **the Dutch cutter and trawler fleet**. A) Scree plot showing the total within sums of squares of the clustering vs. the number of clusters used. B) Average silhouette width of the clusters vs. the number of clusters. C) Mantel test, i.e. the Pearson correlation between the clustering and the original distance matrix vs. the number of clusters. In B) and C), the best score is marked with a vertical line. D) dendrogram of the HAC procedure with the linkage distance of the clusters on the y-axis.

The fact that the optimal clusters is large for the overall fleet is also illustrated by the gridded plots of the dendrograms in Fig 2a. For the complete fleet the optimal clusters is between 14-24 clusters.

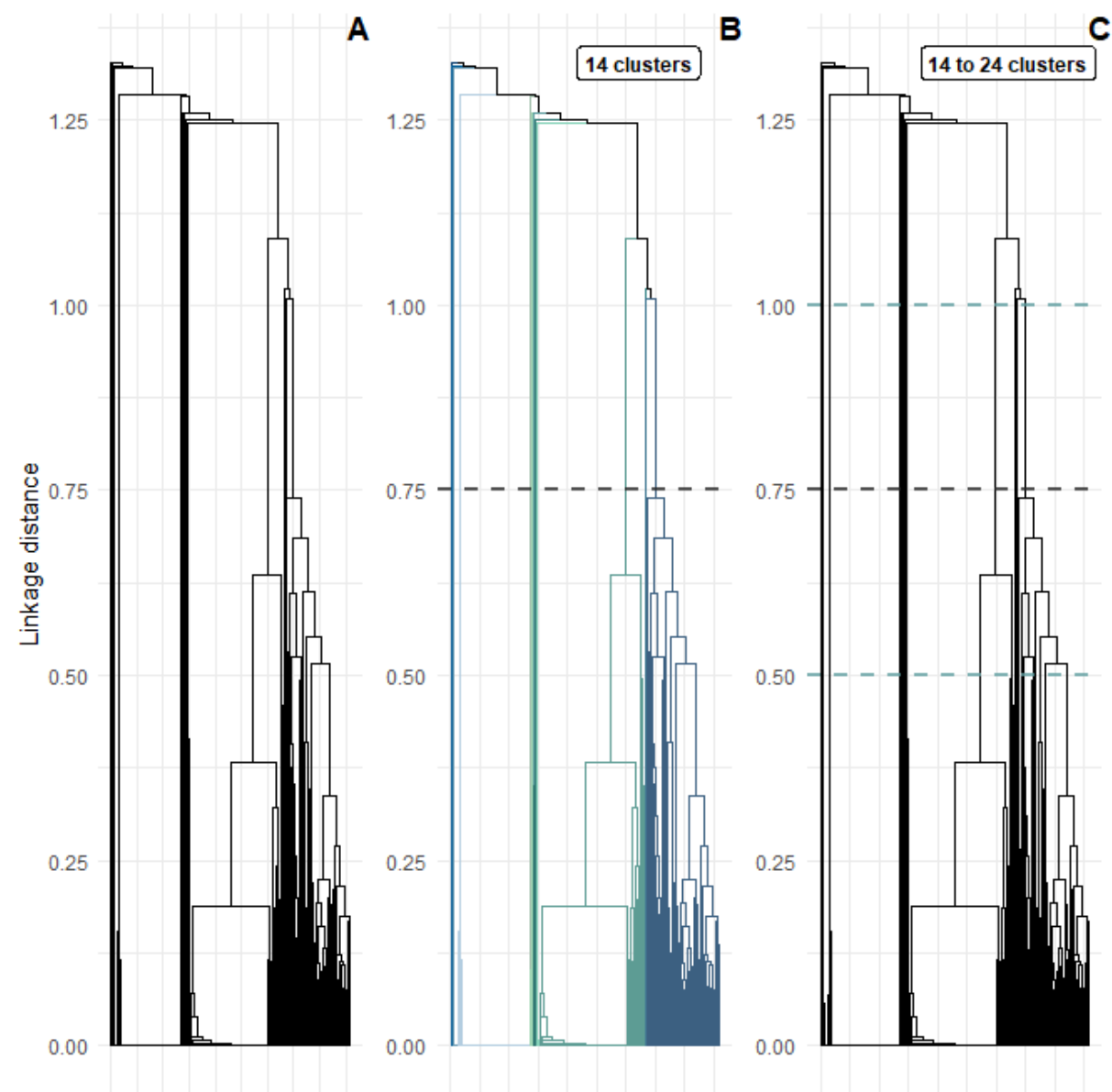


Figure 2a: Grided plot of three dendrograms of the HAC procedure for **the complete Dutch fleet**. The y-axis shows the linkage distance of clusters. A) is unmodified, B) is cut at a linkage distance of 0.75 and resulting branches are individually coloured and the resulting number of clusters is labelled, C) has cutting lines at linkages distance 0.5, 0.75 and 1 and the range of cluster numbers resulting from those cutting heights is labelled.

For the cutter and trawler fleet the optimal number of clusters is between 2-11 clusters (fig 2b).

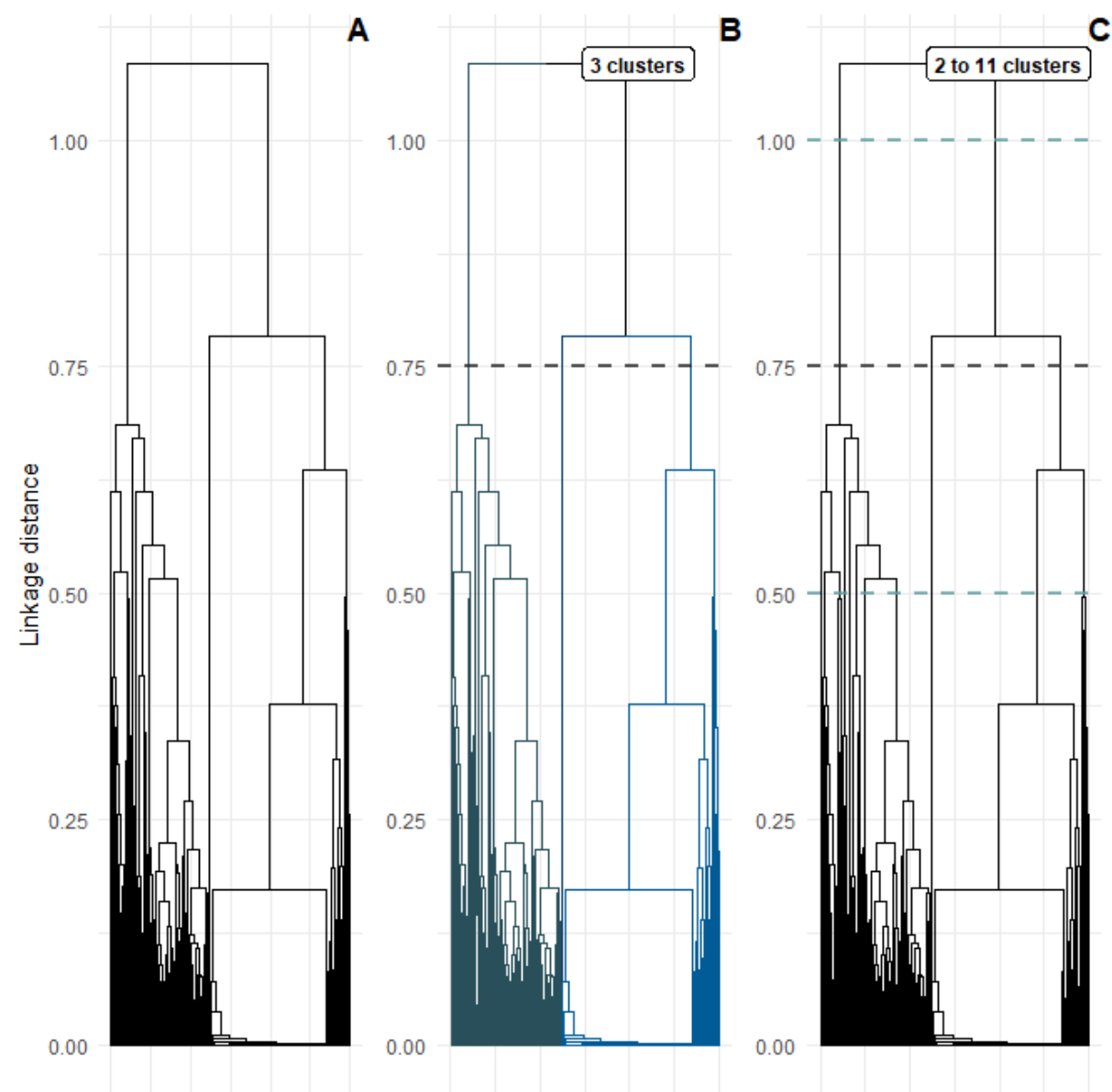


Figure 2b: Grided plot of three dendrograms of the HAC procedure **for the Dutch cutter and trawler fleet**. The y-axis shows the linkage distance of clusters. A) is unmodified, B) is cut at a linkage distance of 0.75 and resulting branches are individually coloured and the resulting number of clusters is labelled, C) has cutting lines at linkages distance 0.5, 0.75 and 1 and the range of cluster numbers resulting from those cutting heights is labelled.

The clustering tree (fig 3a) also confirms that the clustering exercise for the complete fleet results in a large number of very small clusters consisting of one or two vessels and only 3 clusters consist of more than 10 vessels (minimal for EUMap).

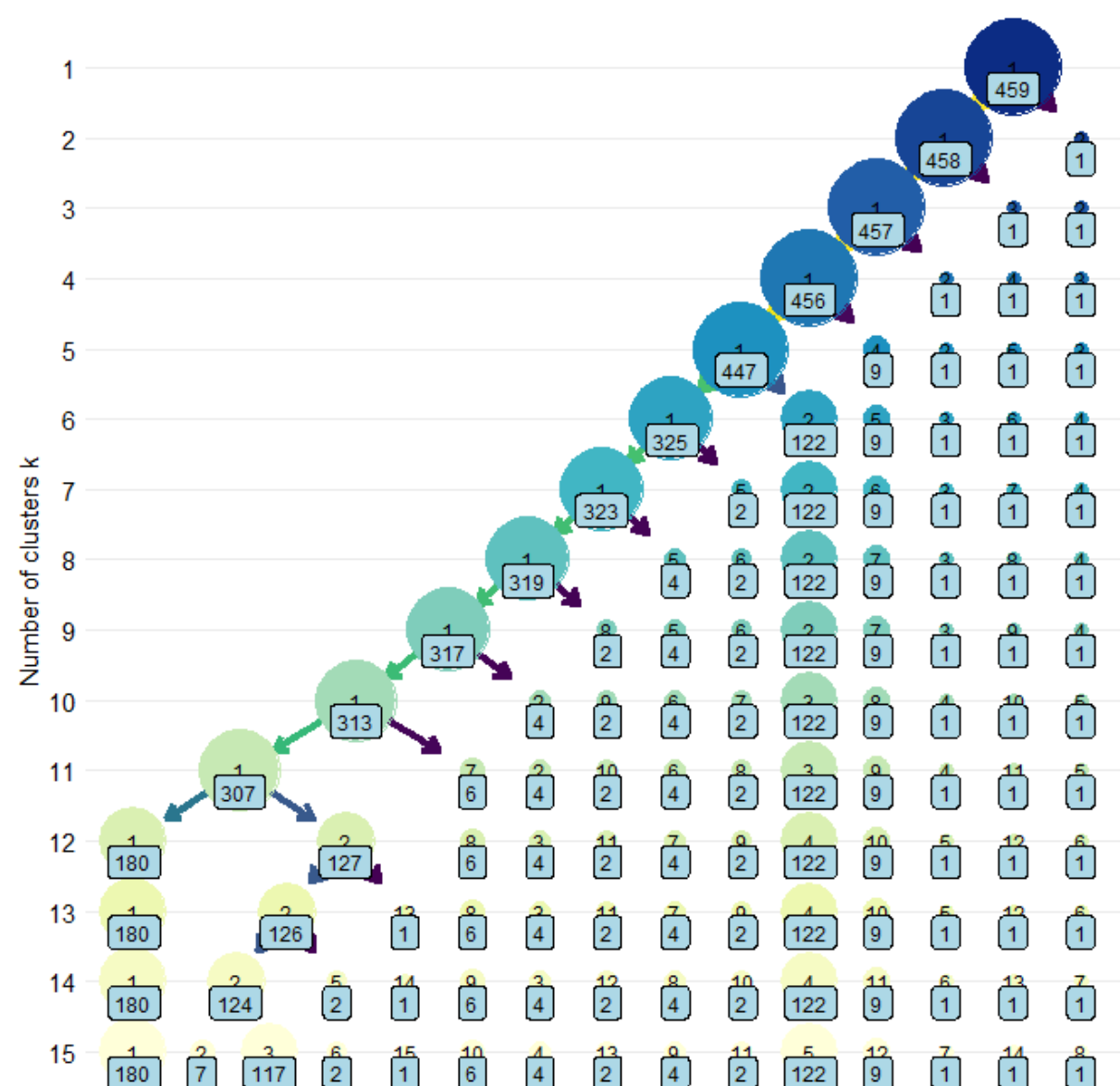


Figure 3a : Treeplot of the HAC-procedure for **the complete Dutch fleet**. The circle size indicates the cluster size, which is also labelled. The number in the circle shows the number of the cluster. The number of clusters in the respective step is depicted on the y-axis.

For the Dutch cutter and trawler fishery the clustering results in a more meaningful separation of groups (Fig 3b). Here also some specific vessels are classified as separate clusters, but among the 7 clusters also 3 clusters have more than 10 vessels. These clusters do not coincide with the clusters from the clustering exercise for the whole fleet.

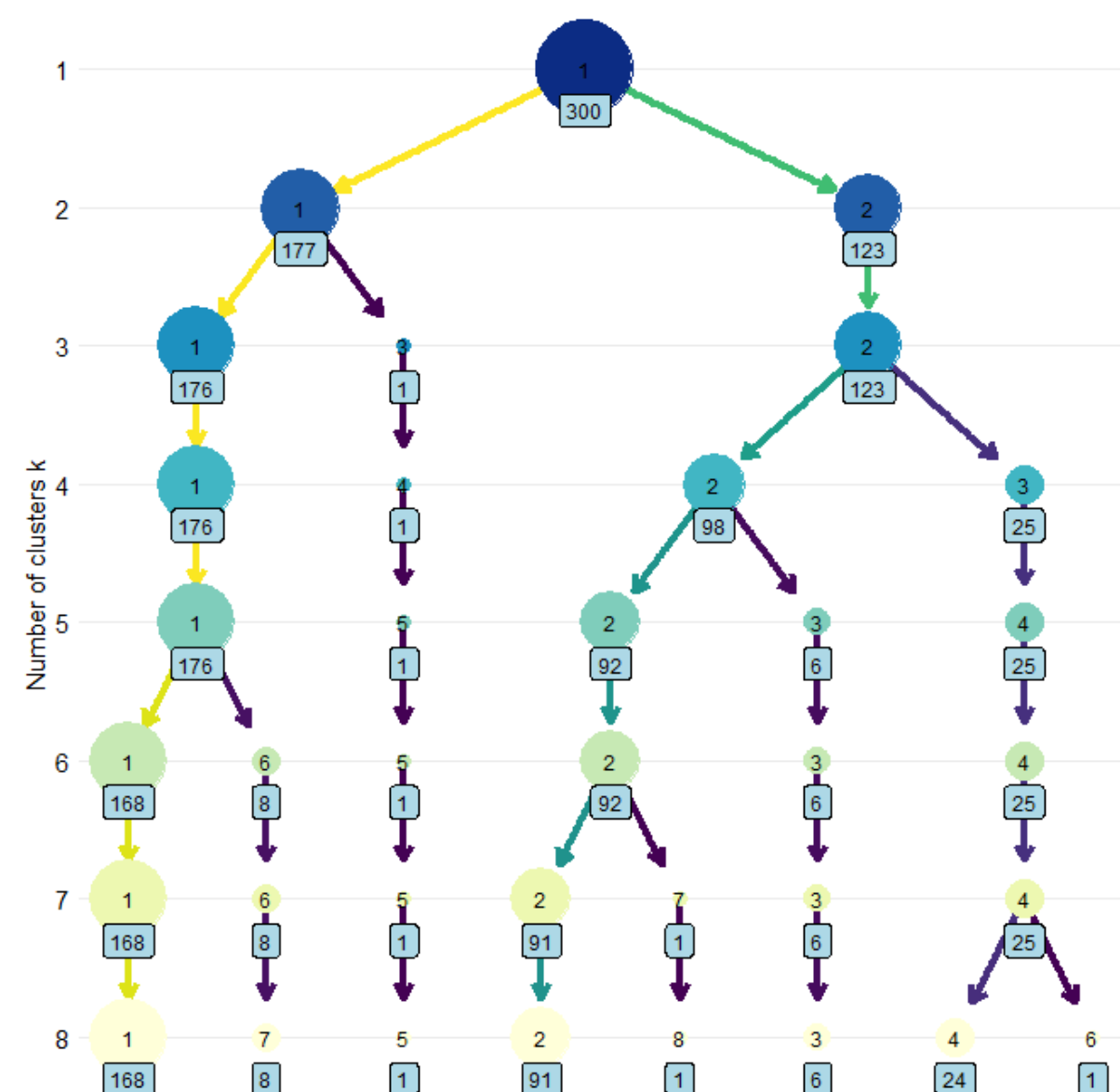


Figure 3b : Treeplot of the HAC-procedure for **the Dutch cutter and trawler fleet**. The circle size indicates the cluster size, which is also labelled. The number in the circle shows the number of the cluster. The number of clusters in the respective step is depicted on the y-axis.

The species composition of the various clusters shows the high variability in fish stocks targeted by vessels of the Dutch fleet (fig. 4a and 4b). In order to interpret these graphs it is good to consider the number of vessels in each of the clusters as shown in fig.7a and 7b.

In the total Dutch fleet (fig. 4a) the main clusters are 1,3 and 5, targeting resp. shrimp, pelagic species and bycatch/unknown. The latter is due to the fact that most small scale vessels had landings per species which were below the threshold from the analysis (5000 kg). Most other clusters of vessels target shellfish/crabs (oysters, whelks, razor clams), mullets and seabass using lines, or smelt. The other small clusters are two vessels that catch shrimp and FLE and which are clustered as a separate group

besides the shrimp cluster . It seems strange that in this clustering sole does not come up as an important species.

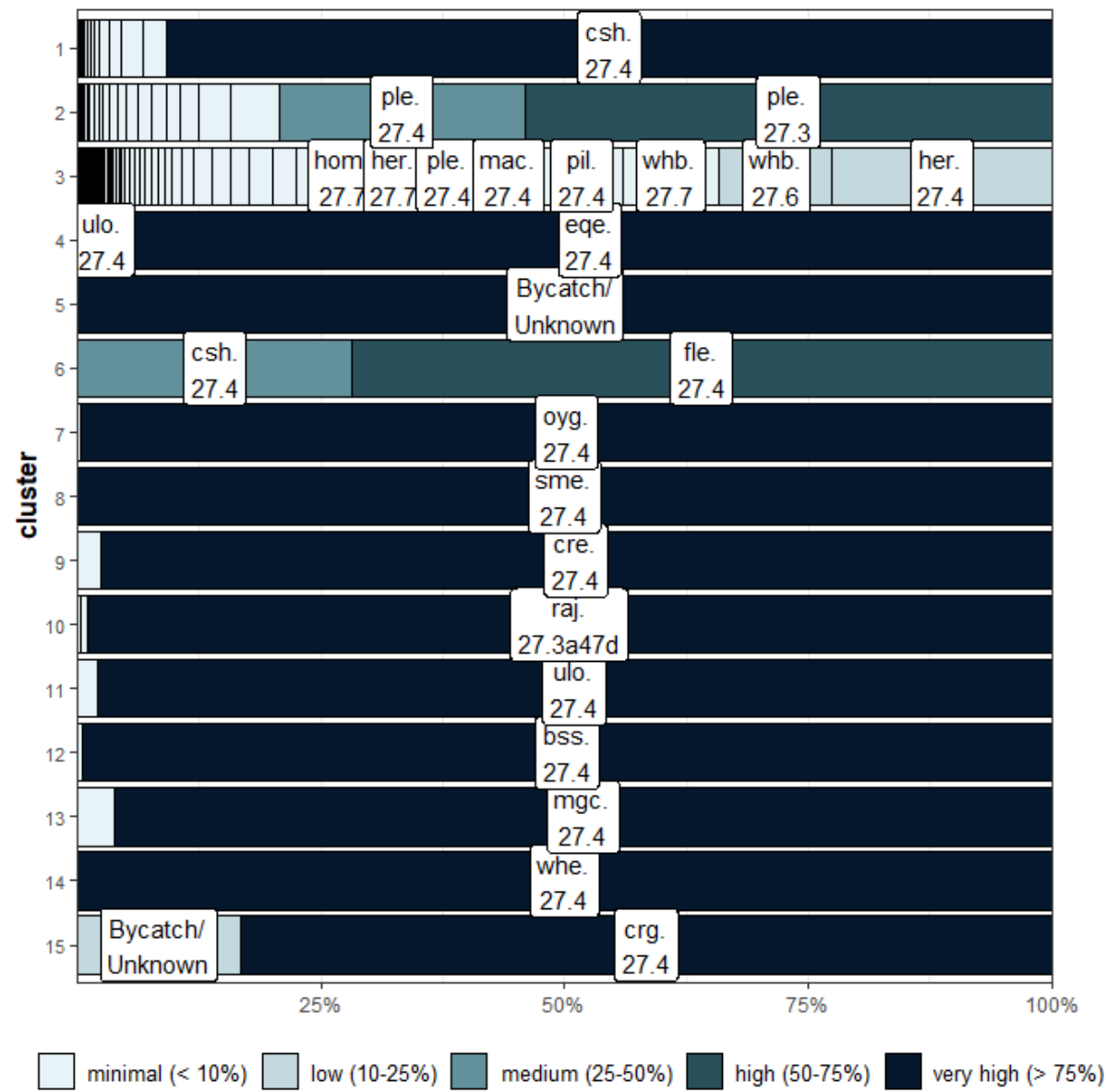


Figure 4a: Barplot of the average percentage of each stock on the total catch of each cluster of the complete Dutch Fleet. The average percentage is depicted on the x-axis, the clusters on the y-axis. The colours indicate the magnitude of the stocks contribution to the total average catch. The stock names are indicated with labels, stocks are ordered by their average percentage contribution.

For the Dutch cutter and trawler sector the species composition of the various clusters shows some of the major fisheries, whereas some others remain hidden in larger clusters. The first cluster includes the fishery targeting shrimps, and combined with the vessel of cluster 5 this cluster seems to represent the “real” schrimp fishery. The second cluster consists of the vessels targeting mainly plaice, and the third cluster includes the vessels targeting sole and plaice combined. The fourth cluster consists of the vessels using danish and scottish seines, targeting a wide variety of species and also includes the pelagic trawlers. Based on the amount of pelagic species in the landings (mainly

mackerel) these two very distinct fisheries are grouped together. Also the vessel from cluster 7 have a comparable catch profile.

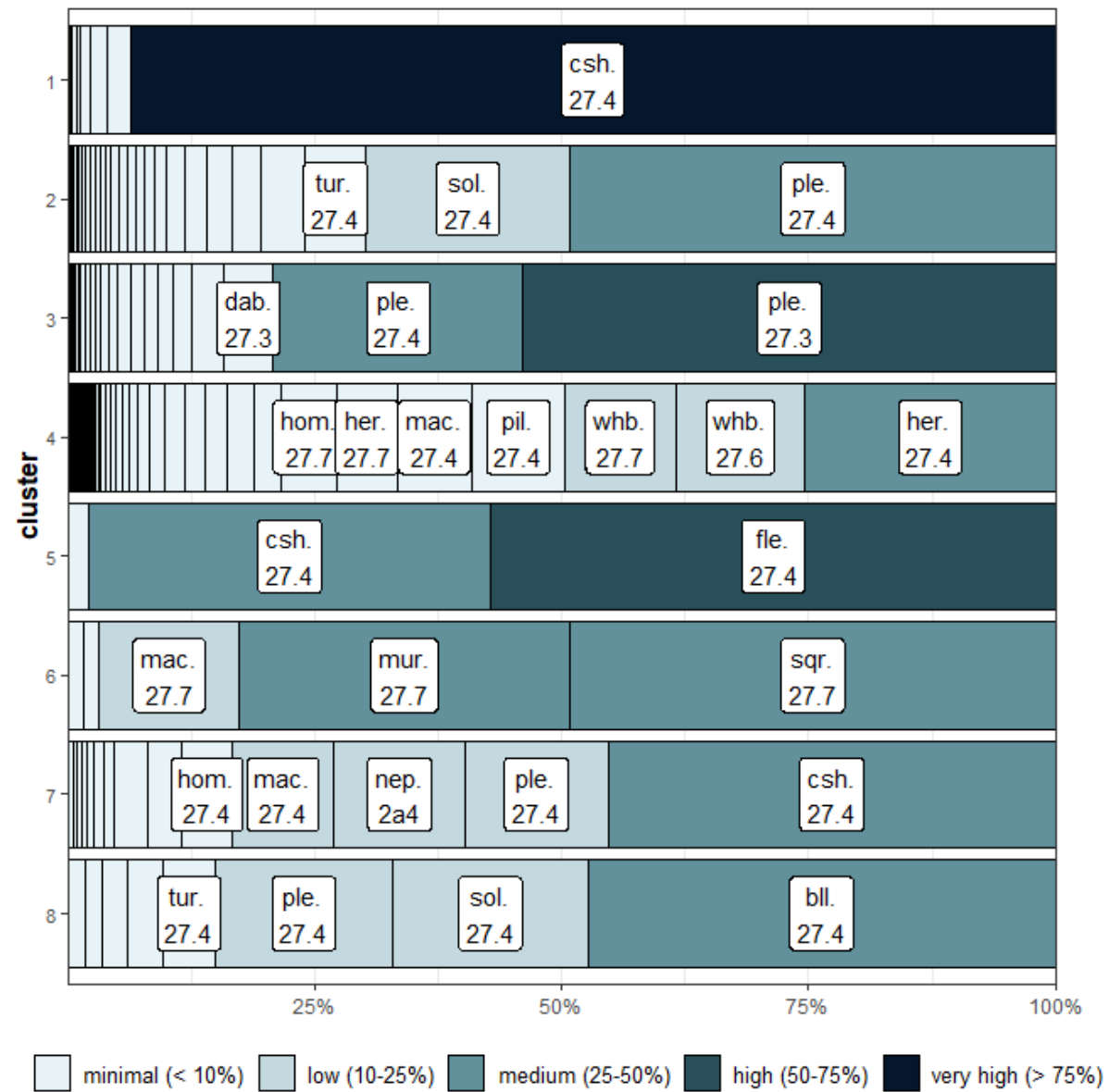


Figure 4b: Barplot of the average percentage of each stock on the total catch of each cluster of the Dutch cutter and trawler fleet. The average percentage is depicted on the x-axis, the clusters on the y-axis. The colours indicate the magnitude of the stocks contribution to the total average catch. The stock names are indicated with labels, stocks are ordered by their average percentage contribution.

Figures 5a and 5b shows the results of the clustering in MDS figures. Here one can also see the differences between the clustering for the complete fleet and the cutter and trawler sector on the other hand.

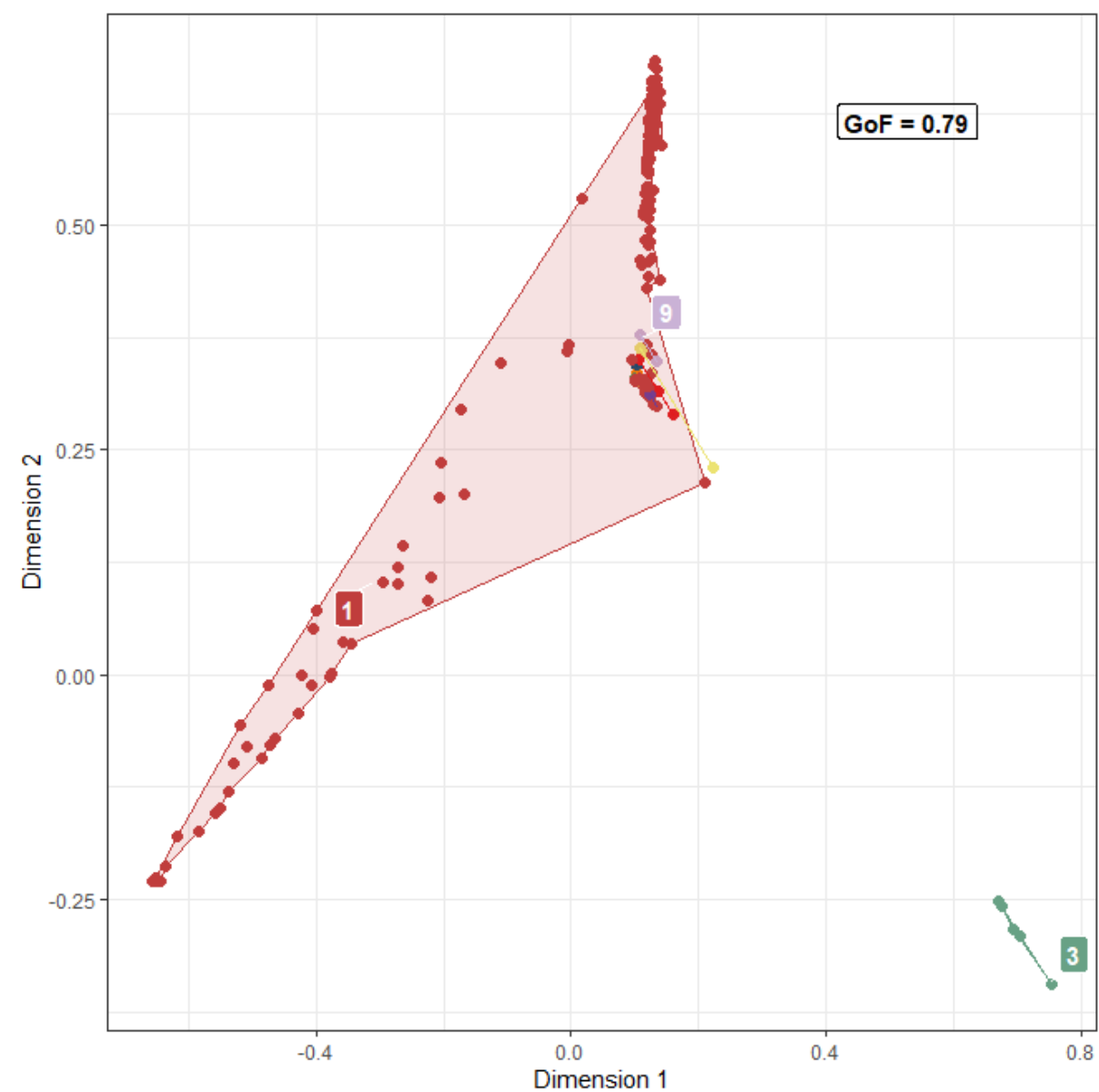


Figure 5a: MDS of the transformed vessel catch data for **the complete Dutch fleet**. Points represent individual vessels, coloured according to their cluster affiliation. Clusters are labelled, labels are coloured similar to points. The goodness of fit (GoF) is labelled.

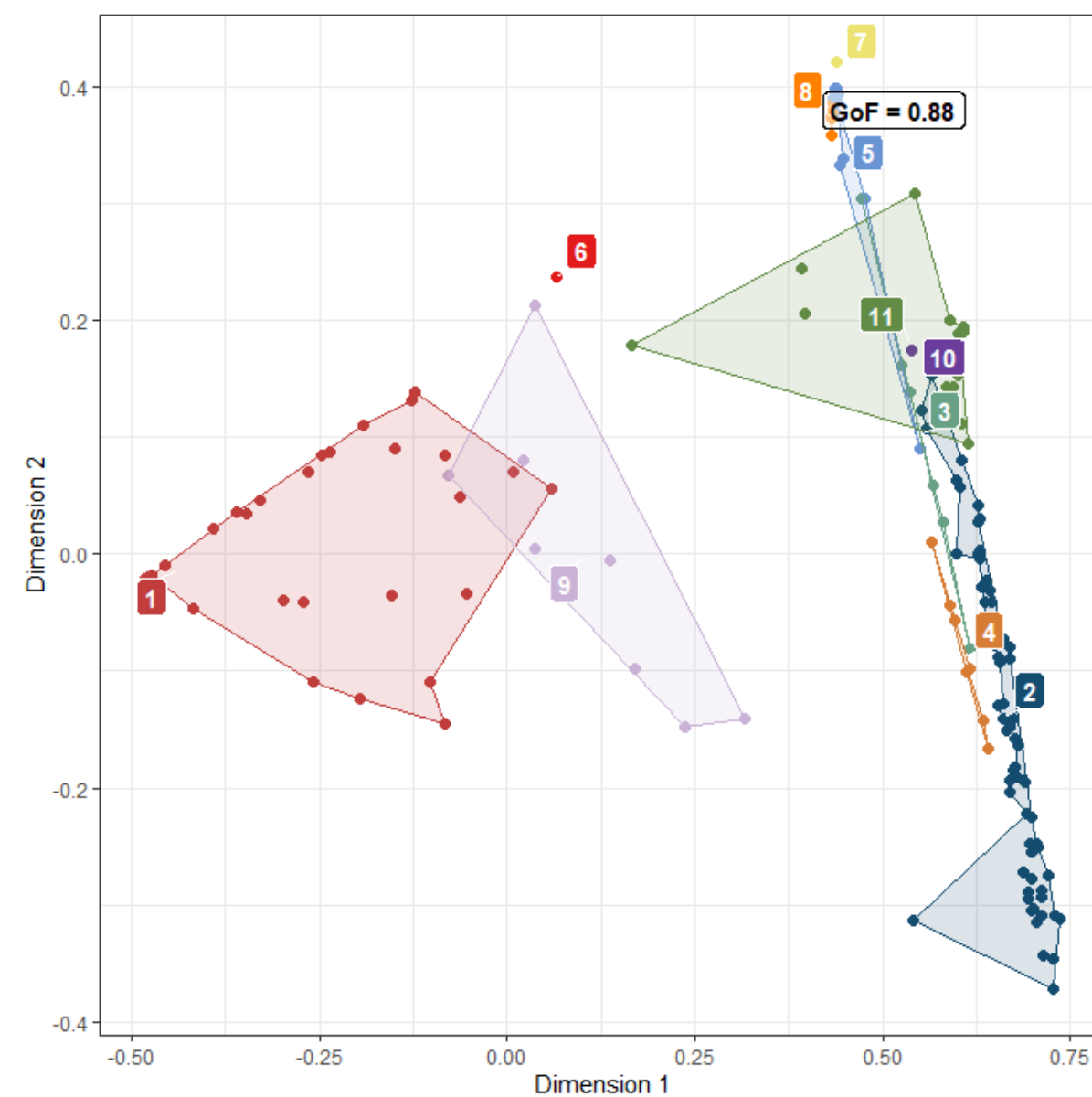


Figure 5b: MDS of the transformed vessel catch data for **the Dutch cutter and trawler fleet**. Points represent individual vessels, coloured according to their cluster affiliation. Clusters are labelled, labels are coloured similar to points. The goodness of fit (GoF) is labelled.

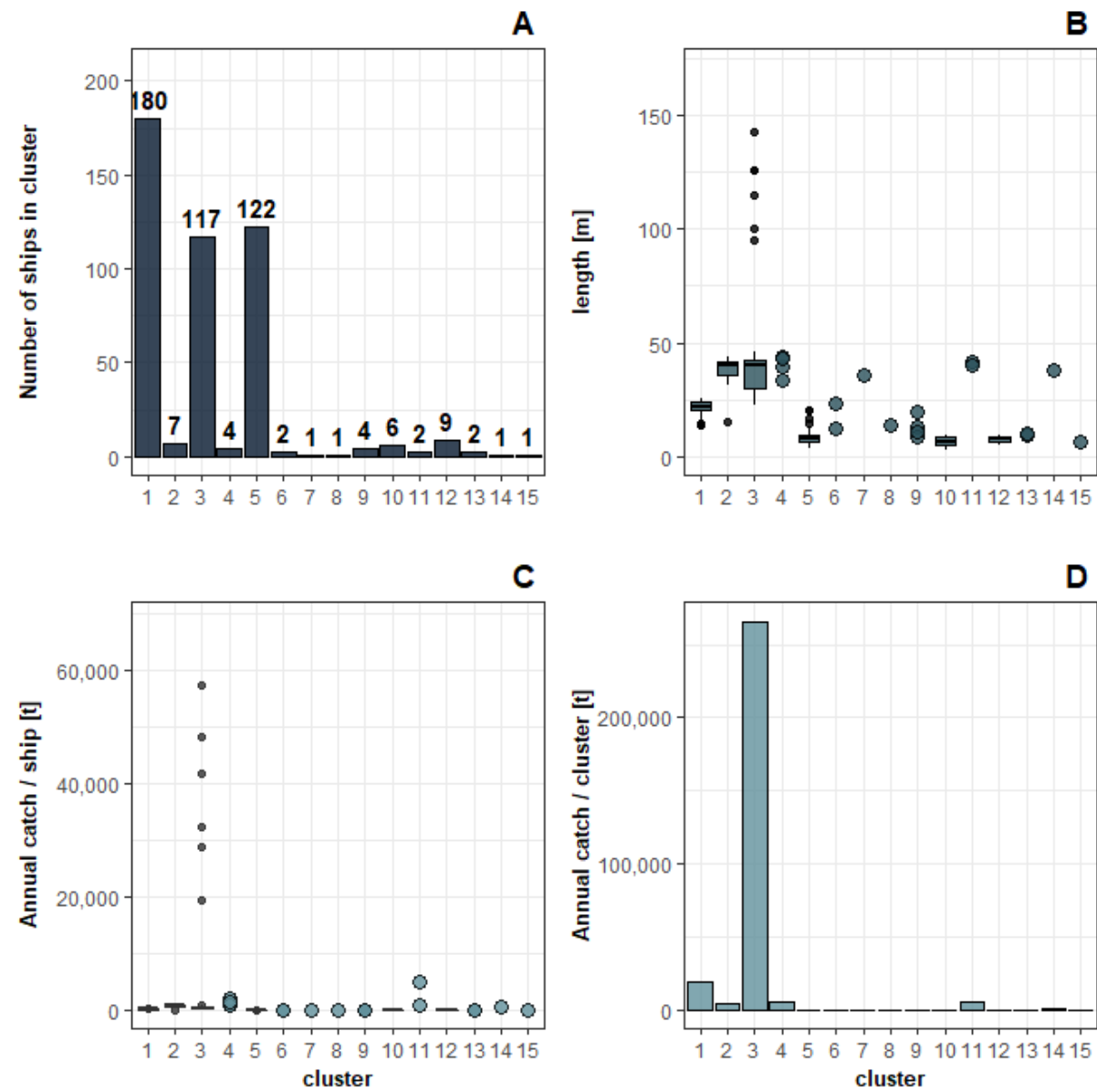


Figure 7a: Grid of mixed plot types displaying vessel characteristics for **the complete Dutch fleet**. A) Barplot of cluster size. The number of ships in each cluster is plotted on the y-axis vs. cluster number on the x-axis. The number over each bar gives the number of vessels in the cluster. B) Mixed box- and dotplot of the length of the vessels in each cluster. Vessel length in meters is depicted on the y-axis vs. cluster number on the x-axis. Clusters containing more than 5 vessels are depicted with a boxplot, for smaller clusters all vessel lengths are plotted with points. C) Mixed boxplot of the annual catch of individual vessels in tons on the y-axis vs. cluster number on the x-axis. Clusters containing more than 5 vessels are depicted with a boxplot, for smaller clusters each vessel catch is plotted with a point. D) Barplot of the total annual catch of all vessels in the clusters. The total catch in tons is depicted on the y-axis vs. cluster number on the x-axis.

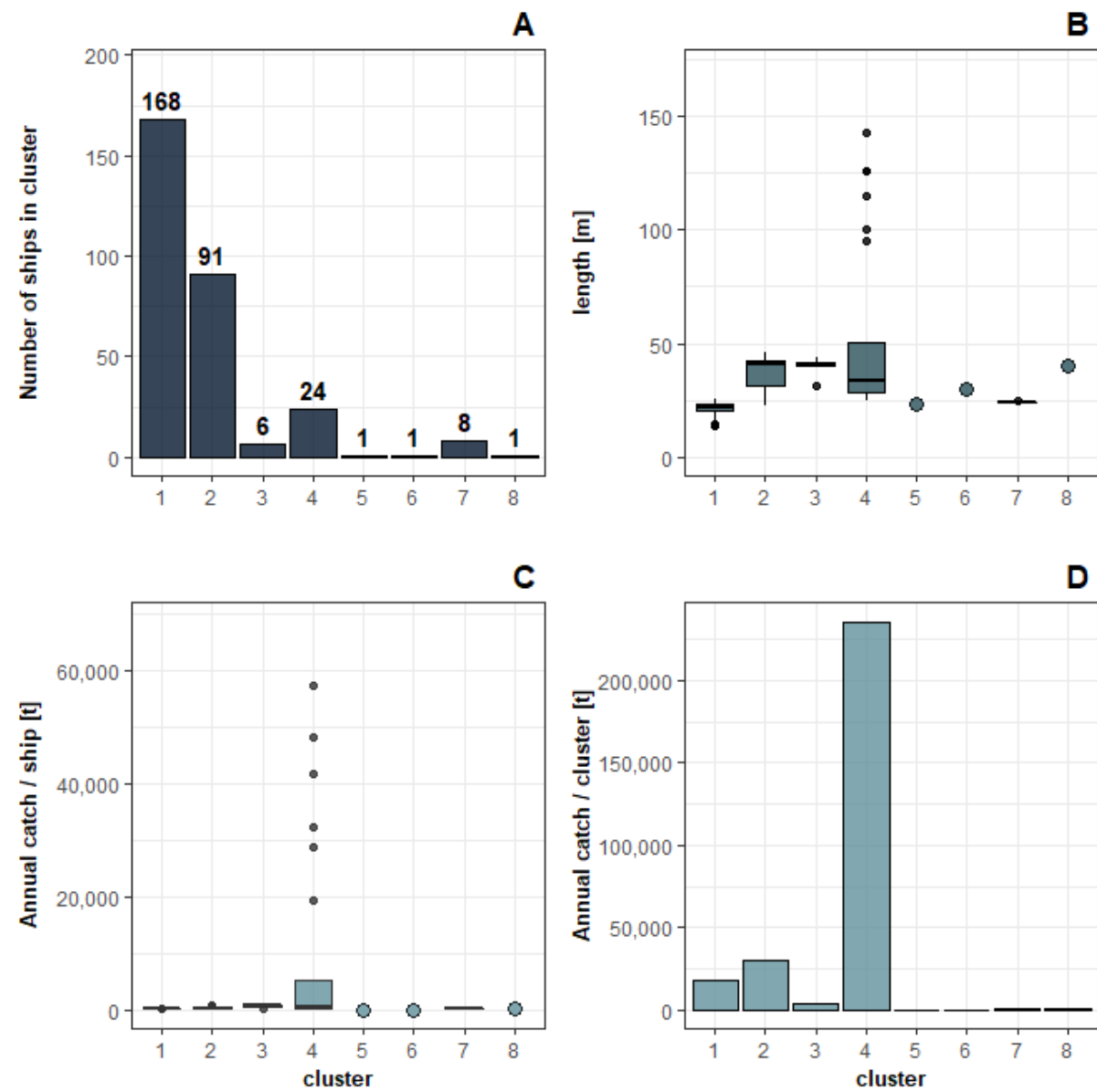


Figure 7a: Grid of mixed plot types displaying vessel characteristics for **the Dutch trawler and cutter fleet**. A) Barplot of cluster size. The number of ships in each cluster is plotted on the y-axis vs. cluster number on the x-axis. The number over each bar gives the number of vessels in the cluster. B) Mixed box- and dotplot of the length of the vessels in each cluster. Vessel length in meters is depicted on the y-axis vs. cluster number on the x-axis. Clusters containing more than 5 vessels are depicted with a boxplot, for smaller clusters all vessel lengths are plotted with points. C) Mixed boxplot of the annual catch of individual vessels in tons on the y-axis vs. cluster number on the x-axis. Clusters containing more than 5 vessels are depicted with a boxplot, for smaller clusters each vessel catch is plotted with a point. D) Barplot of the total annual catch of all vessels in the clusters. The total catch in tons is depicted on the y-axis vs. cluster number on the x-axis.

Conclusions

From the two analyses the following can be concluded:

- For complete sector with highly variable catch compositions, the procedure does not identify the main fisheries. As discussed in the group, the analysis of the total fleet are biased because of individual vessels with highly specific landings composition.
- For the cutter and trawler sector, the cluster analysis identifies some of the main fisheries (shrimp fisheries, plaice fisheries, plaice and sole fisheries) but combines some other distinct fisheries (pelagic trawlers and flyshoot fisheries).
- These issues makes simple application of this clustering analysis to develop a alternative fleet segmentation impossible and make a stepwise analysis necessary. As proposed an analysis per gear will result in better outcomes than an overall analysis of the complete fleet.
- Incorporating other information on effort and geartypes used could definitely be beneficial for the application of the method. E.g. a cluster analysis of the effort per geartyp/metier would potentially also have added value.
- Ultimately the method used should be in line with the objectives of the data collection and data publication (which might be different). E.g. for data collection purposes the clustering of vessels with similar cost structures will be favorable, whereas for publication the link with the biological stocks will be important, but also the ecological effects of the gear use will be more and more important. Moreover, historic consistency of the timeseries and simple rules for classification might have the advantage that changes over time might be easier to understand/explain.

Case .. :Portuguese fleet segmentation

Description of analysis population

The Portuguese fishing fleet is characterized to present multi-specific and multi-gear fishing patterns. Major part of the vessels has multiple gear licenses (polyvalent fleet) and can use more than one gear/métier in each single trip, besides changing it along the year. For this reason, it is very difficult to classify the annual fishing patterns of these vessels.

For this workshop, the classification of the vessels to include in each gear group was performed based on vessels licenses linked to the logbook information (vessels length ≥ 12 m) and sales notes (vessels length < 12 m).

In Portugal there are fisheries targeting species that are not assessed by ICES but are very important at national level (e.g. Atlantic chub mackerel). In this work we removed the ICES stock assignment limitation to enable the classification of all 'species *versus* area' as a new 'stock'.

For the Portuguese fleet, the segmentation analysis was applied to three different gear groups:

- i) Segmentation of the Demersal Trawlers (DTS) operating in Iberian Waters (ICES Divisions 27.9.a and 27.8)
- ii) Segmentation of the Polyvalent Passive Gears (PGP), in ICES Division 27.9.a.
- iii) Segmentation of the Beam Trawlers (TBB), in ICES Division 27.9.a

The selection of these gears resulted from the fact that they present different types of concerns in the vessels classification procedure.

Basic data available

For all vessels in each gear group, landings per species are available from official logbooks and/or sales notes, together with the length data of the vessels.

i) Segmentation of the Demersal Trawlers (DTS) operating in Iberian Waters (ICES Divisions 27.9.a and 27.8)

This group of gear is composed by 87 vessels with a mean length of 23.8 m (range 9m - 34m). Vessels usually have annual fixed licenses to operate with bottom otter trawl and, considering the mesh sizes used, they can target demersal species (65-67mm or ≥ 70 mm) or crustacean species (55-59mm and ≥ 70 mm). The areas considered for this group reflect their fishing behavior: it is known that some vessels can operate in ICES Div 27.8 (Spanish waters) part of the year. The data used in the analysis includes also the landings of 292 different stocks (species*area).

Best number of clusters

The initial analysis of the fleet resulted in the diagnostics shown in **Figure i.1**. The DTS fleet consists of homogeneous groups of vessels targeting multiple species. The graphics presented in **Figure i.1** indicate that 5 clusters can produce adequate results.

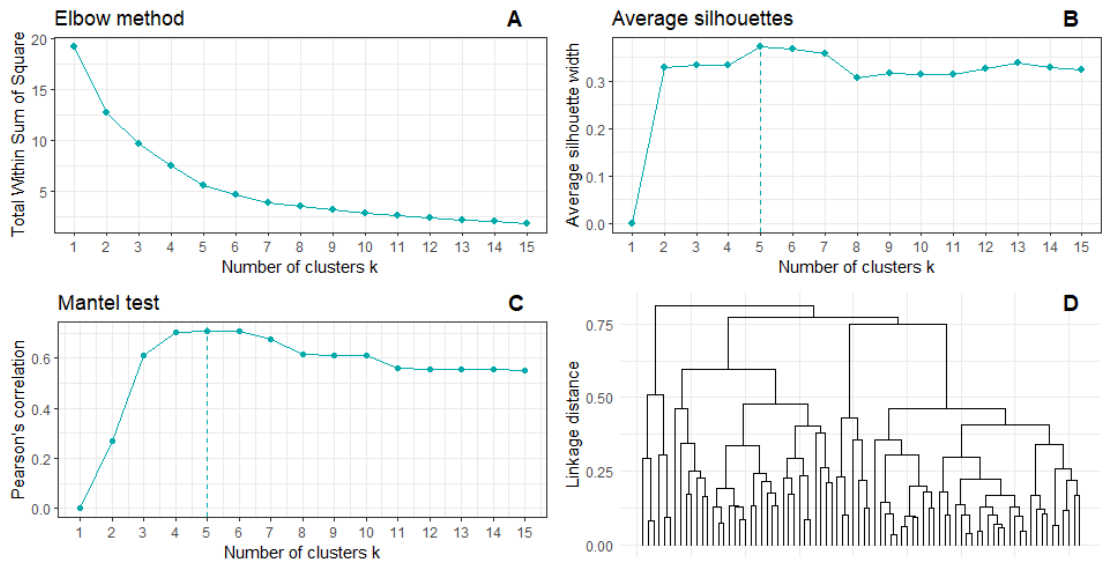


Figure i.1: Diagnostic plots of the HAC procedure for the **Portuguese DTS fleet**. A) Screen plot showing the total within sums of squares of the clustering vs. the number of clusters used. B) Average silhouette width of the clusters vs. the number of clusters. C) Mantel test, i.e. the Pearson correlation between the clustering and the original distance matrix vs. the number of clusters. In B) and C), the best score is marked with a vertical line. D) dendrogram of the HAC procedure with the linkage distance of the clusters on the y-axis.

Figure i.2 presents the dendrograms with different ways of manipulating the cutting height. In this case, the dendrograms support the conclusion that 1 to 6 distinct clusters are present in the data. We decided to rely on the previous diagnostics (**Figure i.1**), selecting 5 clusters for the vessels grouping.

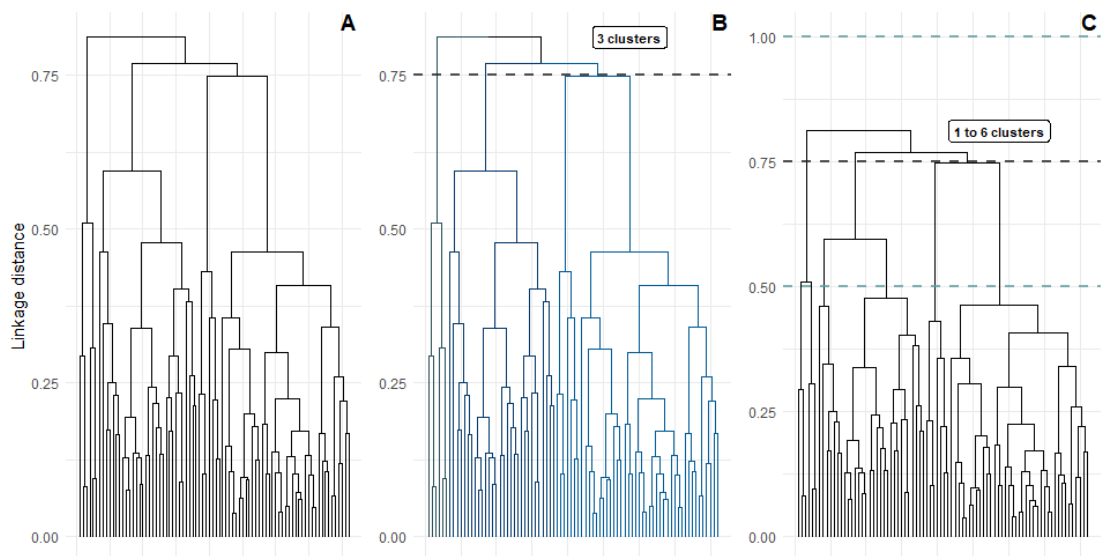


Figure i.2: Gridded plot of three dendrograms of the HAC procedure for the **Portuguese DTS fleet**. The y-axis shows the linkage distance of clusters. A) is unmodified, B) is cut at a linkage distance of 0.75 and

resulting branches are individually colored and the resulting number of clusters is labelled, C) has cutting lines at linkage distance 0.5, 0.75 and 1 and the range of cluster numbers resulting from those cutting heights is labelled.

The clustering tree was another type of diagnostic used to help investigating the best number of clusters to consider in the analysis (**Figure i.3**). It shows that with 5 clusters there is a reasonable separation of the groups, with each cluster including more than 5 vessels. It's also noticeable the separation of a cluster with a very distinct landing profile (Cluster 3), since the beginning of the tree, that would only be separated if 6 clusters were considered.

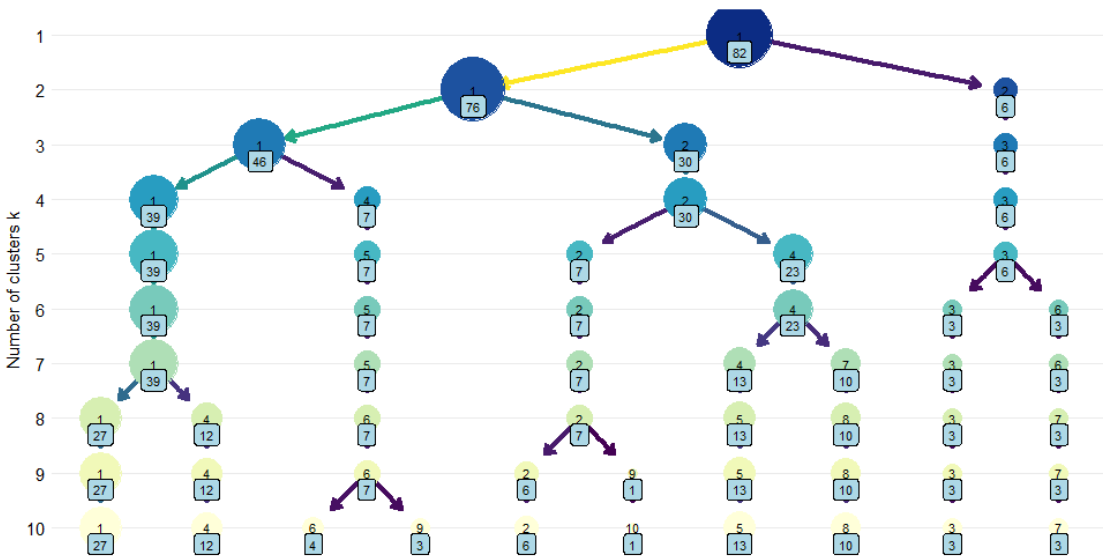


Figure i.3: Treeplot of the HAC-procedure for the **Portuguese DTS fleet**. The circle size indicates the cluster size, which is also labelled. The number in the circle shows the number of the cluster and in the squares is the number of vessels. The number of clusters in the respective step is depicted on the y-axis.

Cluster characterization and validation

The clustering computation, using the five clusters, resulted in the stockshares presented in **Figure i.4**. According to species composition, it is possible to identify the two main métiers for trawlers from the Portuguese fleet (cluster 4 to OTB_CRU and clusters 1, 3 and 5 to OTB_DEF), and another cluster that presents a particular species composition (cluster 2). The analysis of the Portuguese landings data shows that mackerel (mac.27.nea) are mainly caught (91% of total landings) by the Portuguese fleet in ICES Div 27.8, while horse mackerel (hom.27.9.a) and blue whiting (whb.27.1-91214) are caught in ICES Div. 27.9.a (100% and 91% respectively). Regarding the OTB_DEF clusters, and since vessels in the same cluster present different pattern of targeted species, it may indicate that they present a seasonal fishing pattern regarding the fishing area.

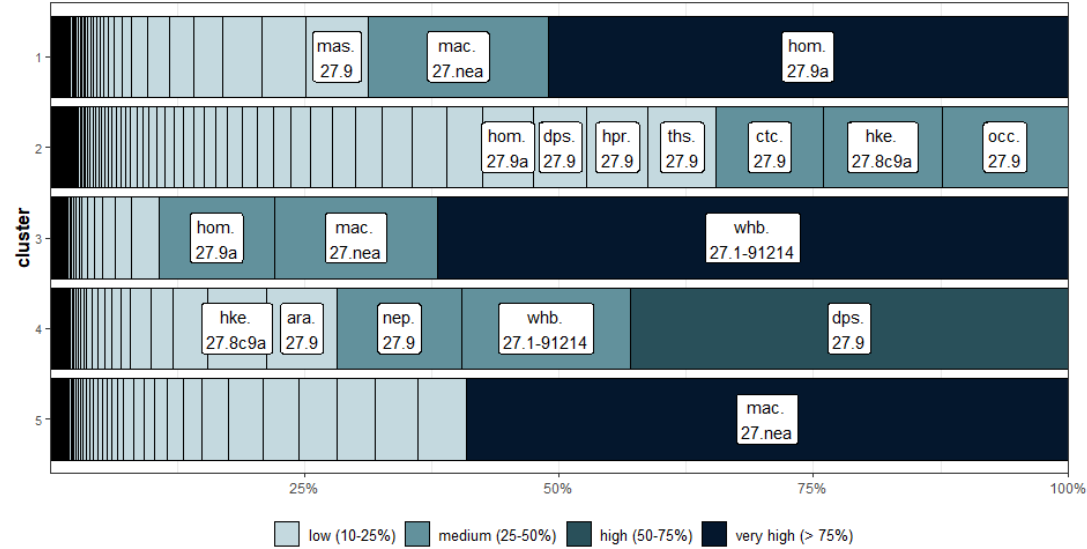


Figure i.4: Barplot of the average percentage of each stock on the total landings of each cluster of the **Portuguese DTS fleet**. The average percentage is depicted on the x-axis, the clusters on the y-axis. The colors indicate the magnitude of the stocks contribution to the total average landings. The stock names are indicated with labels; stocks are ordered by their average percentage contribution.

The cluster 2 presents a landing profile mainly composed by mixed cephalopods and fish species, constituting a different pattern from what was “expected”. The plot with the stockshares by vessel was performed to better investigate this group of vessels (Figure i.5). It shows some similar target species among vessels although differing in their proportions.

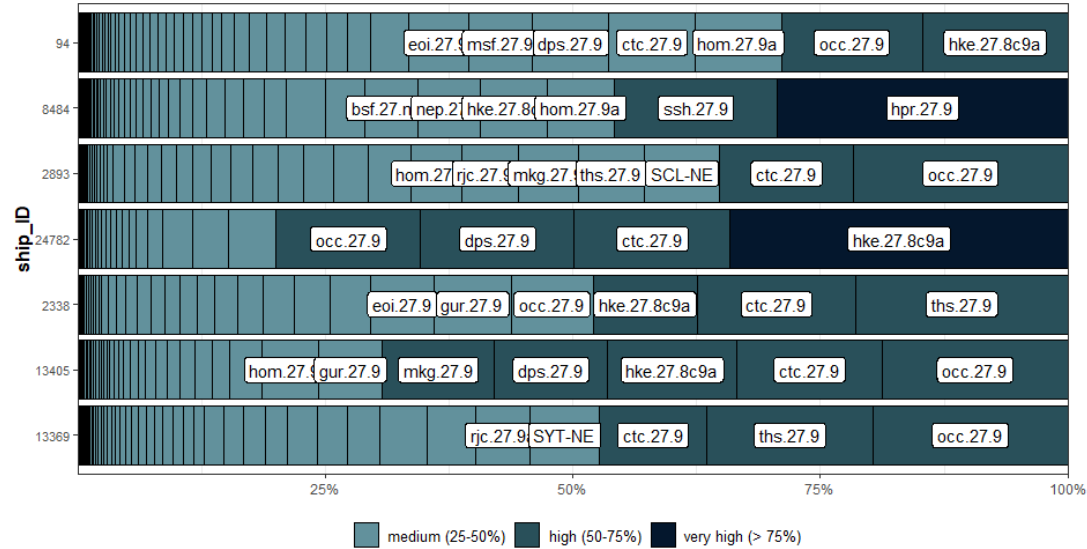


Figure i.5: Barplot of the average percentage of each stock on the total landings of each vessel from cluster 2 for the **Portuguese DTS fleet**. The average percentage is depicted on the x-axis, the vessels' id on the y-axis. The colors indicate the magnitude of the stocks contribution to the total average landings. The stock names are indicated with labels; stocks are ordered by their average percentage contribution.

The 2-dimensional MDS, used for visualising the compositional distances of clusters and identify which clusters could/should be united into a single fleet segment, is presented in Figure i.6. For the DTS fishery, there's no cluster overlapping probably indicating that they

are distinct from each other, and it also shows a goodness of fit quite good (GoF=68%). Moving in the first dimension, clusters 1 and 4 are in opposite sides, reflecting the differences in the landings composition: cluster 1 for trawlers targeting demersal species of fish and cluster 4 targeting crustaceans. This is also noticeable in the 3-dimensional plot presented in Figure i.7.

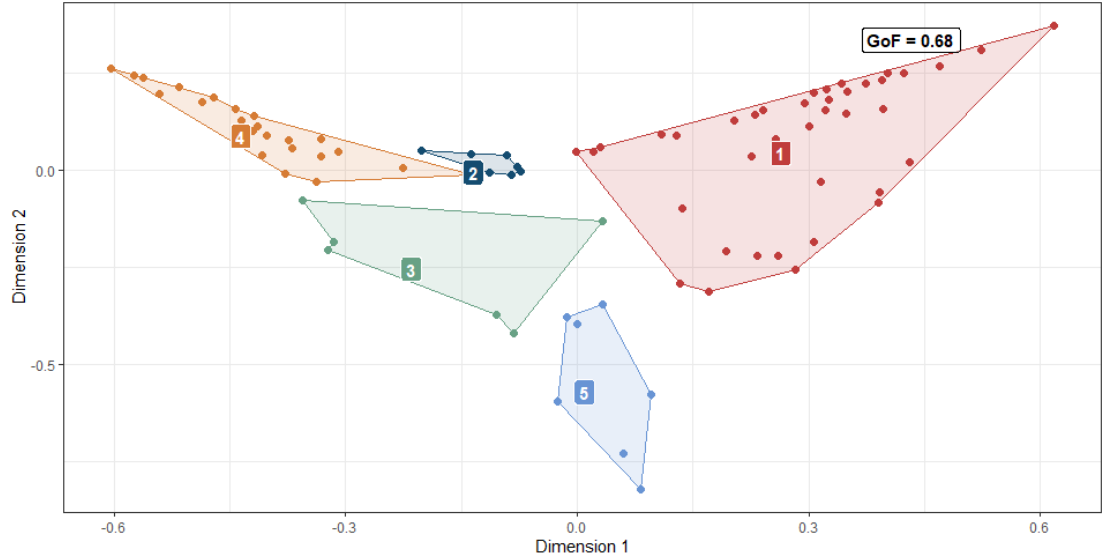


Figure i.6: MDS of the transformed vessel catch data for the **Portuguese DTS fleet**. Points represent individual vessels, colored according to their cluster affiliation. Clusters are labelled, labels are colored similar to points. The goodness of fit (GoF) is labelled.

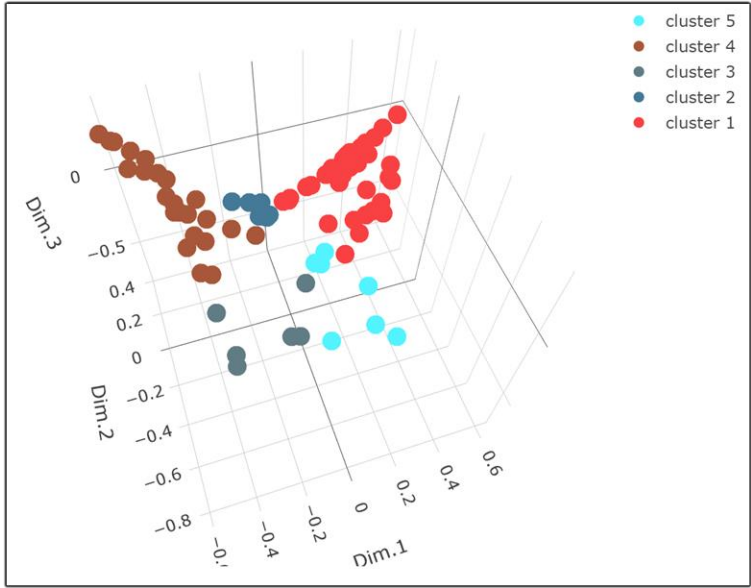


Figure i.7: Snapshot of interactive 3D-MDS for the **Portuguese DTS fleet**. Rotated to 2nd dimension being the x-axis, 3rd dimension the y-axis and 1st dimension the z-axis.

Vessel and cluster properties

The clusters including higher number of vessels are the clusters 1 and 4 (Figure i.8.A). These correspond to the main métiers of this fleet (OTB_DEF and OTB_CRU respectively). The Figure i.8 also shows that cluster 1 contributes to the highest landings volume of the fleet, both in annual landings per vessel (Figure i.8.C) and per cluster (Figure i.8.D). On

the other hand, cluster 4 landings are very low when related to the fishing effort in terms of number of vessels. The main reason for that is because in this fishery (targeting crustaceans) the species value per kilo is much higher than in the other fisheries. The clusters 3 and 7 are composed by a lower number of vessels (6 and 7) but can also be included in OTB_DEF métier, although presenting some differences in the species proportions. The cluster 2 seems to have a different fishing pattern: is composed by the smaller vessels of the fleet (Figure i.8.B), and presents a different species composition (Figure i.4). It also represents a small part of the fleet (7 vessels), having the lower annual landings per vessel and per cluster.

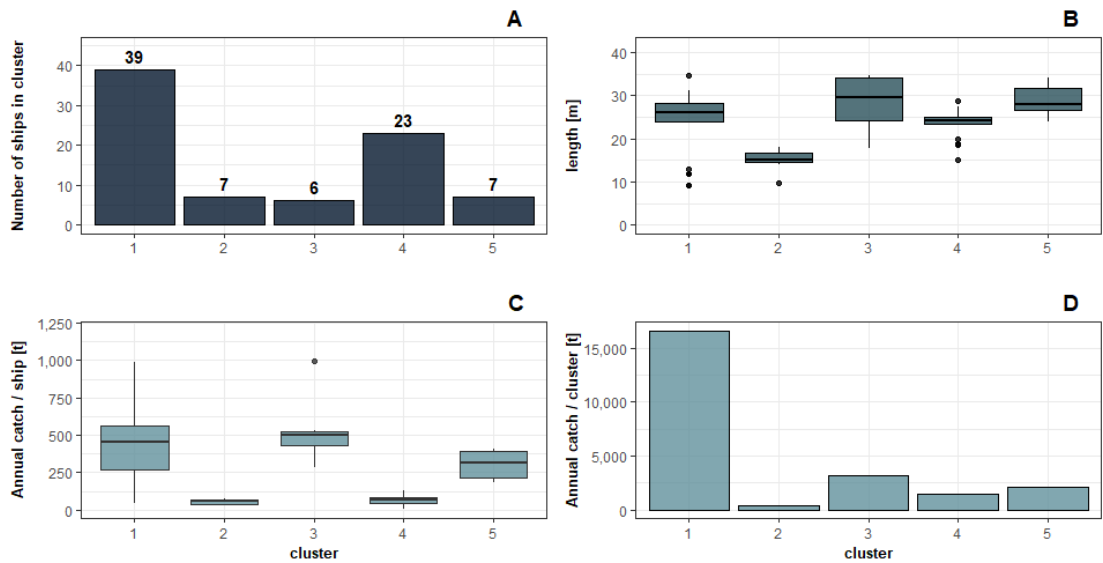


Figure i.8: Grid of mixed plot types displaying vessel characteristics for the Portuguese DTS fleet. **A)** Barplot of cluster size with the number of ships in the cluster over each bar. **B)** Mixed box- and dotplot of the length of the vessels in meters, each cluster. **C)** Mixed boxplot of the annual catch of individual vessels in tons. **D)** Barplot of the total annual catch of all vessels in the clusters. (For B) and C), clusters containing more than 5 vessels are depicted with a boxplot, for smaller clusters each vessel catch is plotted with a point.)

Fleet segment assignment

Besides the characterization of the landings composition of the clusters, it was also considered the vessel length category in the classification of the fleet segments. Two groups of vessels were considered: <12 meters for Small-scale fishery and ≥12 meters for Large-scale fishery. The characterization and fleet segment assigned per cluster are listed below:

Cluster 1 - OTB_DEF fishery: composed by vessels that present seasonal fishing patterns: one targeting MAC in 27.8 and other targeting HOM in 27.9:

- "Coastal fishery targeting HOM (27.9.a) and MAC (27.8) (Large-scale fishery)"
- "Coastal fishery targeting HOM (27.9.a) and MAC (27.8) (Small-scale fishery)"

Cluster 2 - OTB_MCF (Mixed Cephalopod and Fish species) fishery - includes a group of smaller vessels with distinct target assemblage; small number of vessels with low annual landing volumes per vessel and total:

- "Coastal fishery targeting Mixed Cephalopod and Fish species (MCF) (Large-scale fishery)"
- "Coastal fishery targeting Mixed Cephalopod and Fish species (MCF) (Small-scale fishery)"

Cluster 3 - OTB_DEF fishery - includes 'bigger' vessels targeting mainly WHB (and HOM) in 27.9, and MAC in 27.8 (dependent of fishing area/season)

- "Coastal fishery targeting WHB (27.9.a) and MAC (27.8) (Large-scale fishery)"

Cluster 4 - OTB_CRU fishery - present a common pattern of this fishery: lower landing volumes (but higher income per kilo, when compared to OTB_DEF fishery) - importance of this fishery is evaluated in terms of species value and not volume. However, it's still traceable with this procedure.

- "Coastal crustacean fishery (Large-scale fishery)"

Cluster 5 - OTB_DEF fishery - group of vessels operating mainly in 27.8, presenting reasonable annual landings per vessel but with a lower contribution for the total annual landings of the fleet.

- "Coastal fishery targeting MAC (27.8) (Large-scale fishery)".

Conclusions

From this analysis we may conclude that:

- For this gear group the main métiers of the fleet were identified: OTB_DEF and OTB_CRU;
- It also identified a new group of vessels, with smaller total lengths, that present a different target assemblage (mixed cephalopods and fish): OTB_MCD. This segment can be further explored, to understand better its fishing dynamic;
- It highlighted the seasonality of some vessels from the OTB_DEF clusters regarding the fishing areas: 27.9.a and 27.8. This can be further investigated by performing this analysis in a quarterly basis instead of the annual.

ii) Segmentation of the Polyvalent Passive Gears (PGP), in ICES Division 27.9.a

This group of gears is composed by 1579 vessels with a mean length of 6.8 m (range 3m - 20m), meaning that the major part of these vessels is from small scale fisheries (vessel length < 12m). They are characterized to have licenses for multiple types of gears and don't have logbooks for registering their catches. So, in these cases, the sales notes along with the licenses information are used for classifying the gears used in the trips, which is not a straightforward procedure because a vessel can use more than one gear/métier in the same trip. The data used in this analysis includes the landings of 235 different 'stocks' (species*area).

Best number of clusters

The initial analysis of the fleet resulted in the diagnostics shown in Figure ii.1. The PGP fleet consists of very heterogeneous groups with a wide variety of vessels targeting multiple groups of species. It is also expected to find small groups of vessels with specific target assemblages. For these reasons, and also because the number of vessels to analyze is extremely high (1579), the graphics presented in Figure ii.1 show that the optimal number of clusters cannot be reached before 14 to 15 clusters.

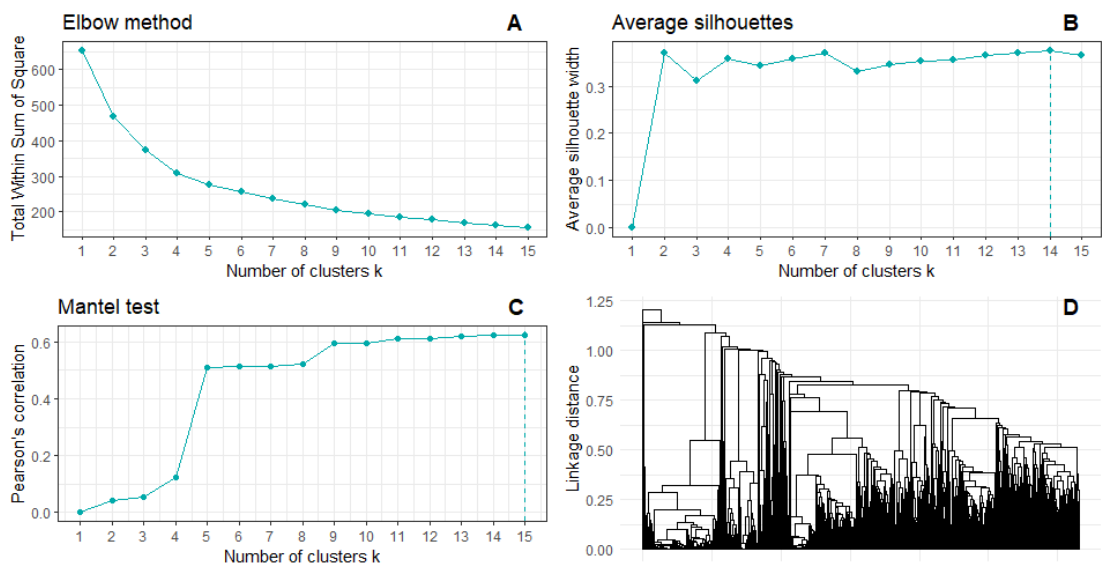


Figure ii.1: Diagnostic plots of the HAC procedure for the **Portuguese PGP fleet**. A) Screen plot showing the total within sums of squares of the clustering vs. the number of clusters used. B) Average silhouette width of the clusters vs. the number of clusters. C) Mantel test, i.e. the Pearson correlation between the clustering and the original distance matrix vs. the number of clusters. In B) and C), the best score is marked with a vertical line. D) dendrogram of the HAC procedure with the linkage distance of the clusters on the y-axis.

Figure ii.2 presents the dendrograms with different ways of manipulating the cutting height. In this case, the dendrograms support the extremely high number of possible vessels groupings (8 to 74 clusters). We decided to use the outcomes of the previous diagnostics (Figure ii.1), selecting 14 clusters because increasing this number would result in a higher number of clusters with only one vessel.

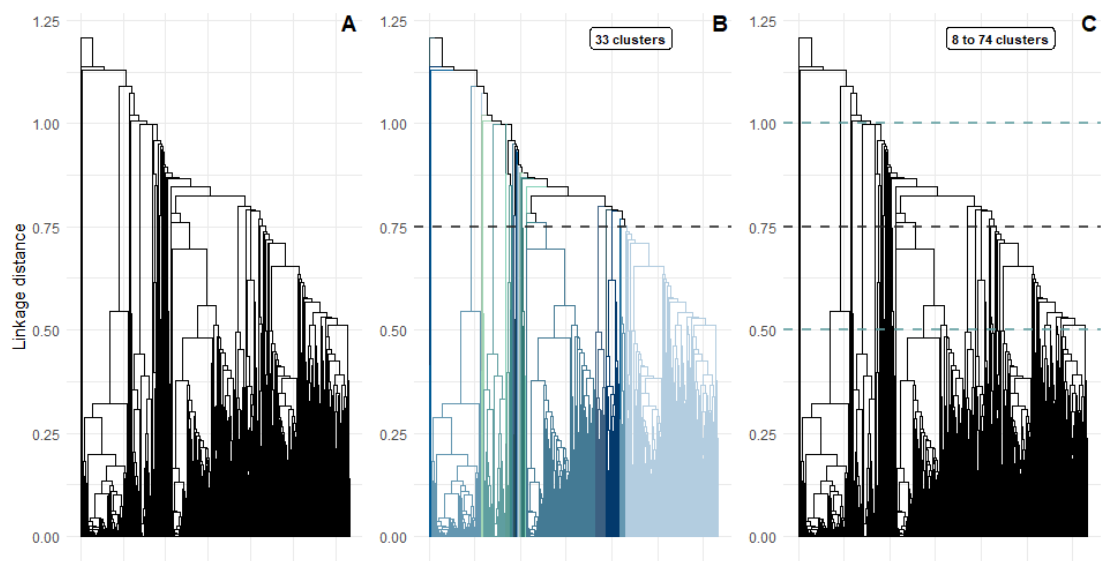


Figure ii.2: Gridded plot of three dendrograms of the HAC procedure for the **Portuguese PGP fleet**. The y-axis shows the linkage distance of clusters. A) is unmodified, B) is cut at a linkage distance of 0.75 and resulting branches are individually colored and the resulting number of clusters is labelled, C) has cutting lines at linkages distance 0.5, 0.75 and 1 and the range of cluster numbers resulting from those cutting heights is labelled.

The clustering tree presented in Figure ii.3, also confirms the expected 'problems' when dealing with such a large number of vessels, with a wide variety of different gear types and species targeted. There is always a main group that includes a large number of vessels which is being separated into groups sometimes with very small number of vessels: 7 from the 14 clusters have less than 5 vessels. This highlights the big difficulty in classifying these gear group vessels.

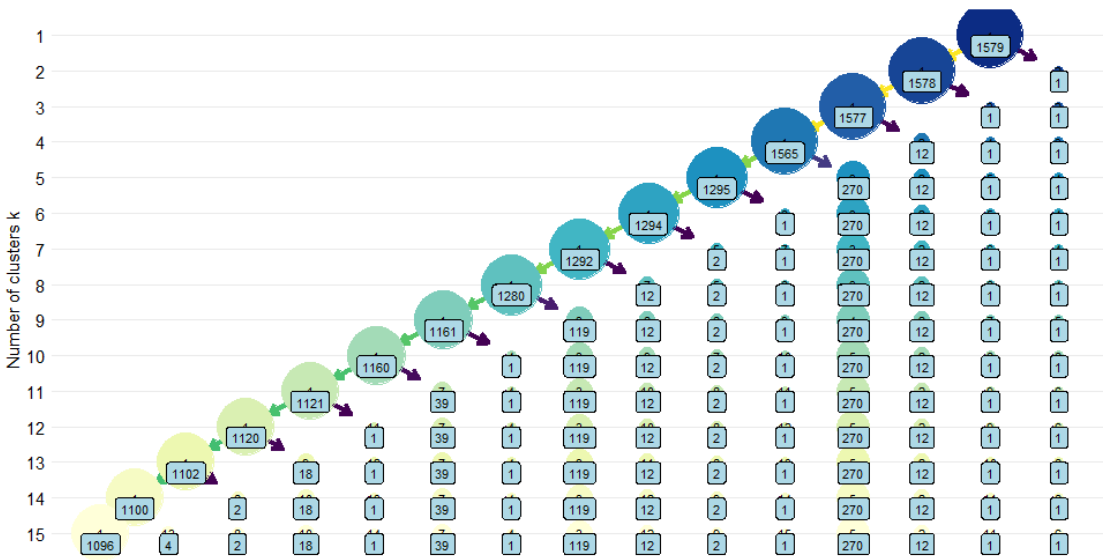


Figure ii.3: Treeplot of the HAC-procedure for the **Portuguese PGP fleet**. The circle size indicates the cluster size, which is also labelled. The number in the circle shows the number of the cluster and in the squares is the number of vessels. The number of clusters in the respective step is depicted on the y-axis.

Cluster characterization and validation

The clustering computation, using the fourteen clusters, resulted in the stockshares presented in Figure ii.4. According to the species composition, it is possible to identify very specific fisheries already known (e.g. sea lamprey and eel fisheries) but also some other groups composed by one vessel, presenting particular target assemblages which were not of common knowledge. The landing profiles per cluster also highlight the fact that for the polyvalent fleet (in this case specially the cluster 1), the species present are caught by different gears/métiers. This can occur along the year but also within the same trip. These vessels constitute the major part of the polyvalent fleet.

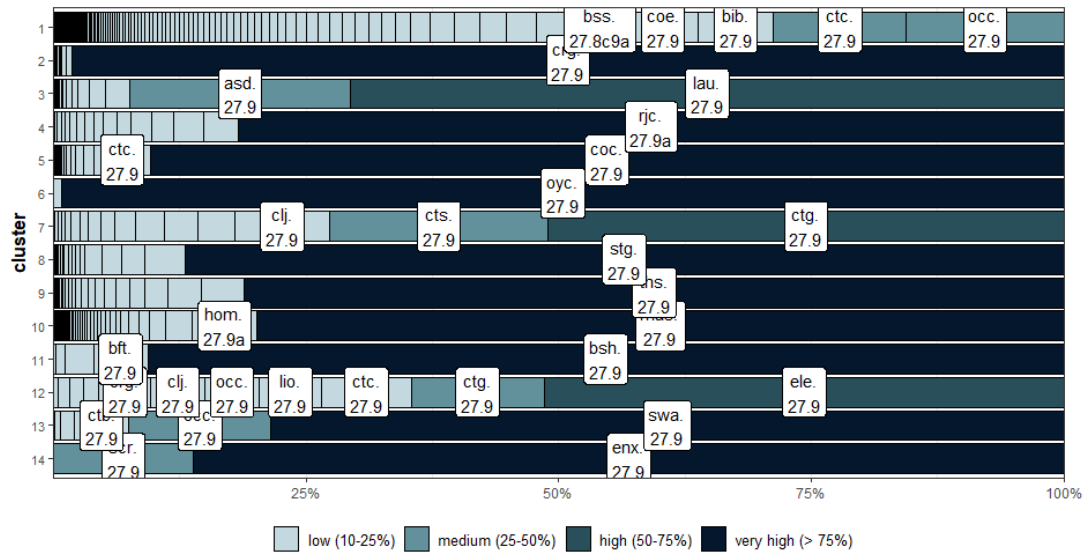


Figure ii.4: Barplot of the average percentage of each stock on the total landings of each cluster of the Portuguese PGP fleet. The average percentage is depicted on the x-axis, the clusters on the y-axis. The colors indicate the magnitude of the stocks contribution to the total average landings. The stock names are indicated with labels; stocks are ordered by their average percentage contribution.

The 2-dimensional MDS, used for visualising the compositional distances of clusters and identify which clusters could/should be united into a single fleet segment, is presented in Figure ii.5. For the PGP fishery, there are overlapping clusters (Figure ii.5 and Figure ii.6) and, because some clusters have only one vessel, they are difficult to find specially in this 2-dimensional plot (Figure ii.5). The cluster that stands more apart for the others is the cluster 5 that is composed by the second higher number of vessels of this fleet (Figure ii.7.A) and targets cockle species. Also the vessels from cluster 3 are very closed together probably because it's a very species specific fishery targeting sea lamprey. The high variability of the fisheries included in this gear group is reflected in the value obtained for the goodness of fit (GoF=0.49).

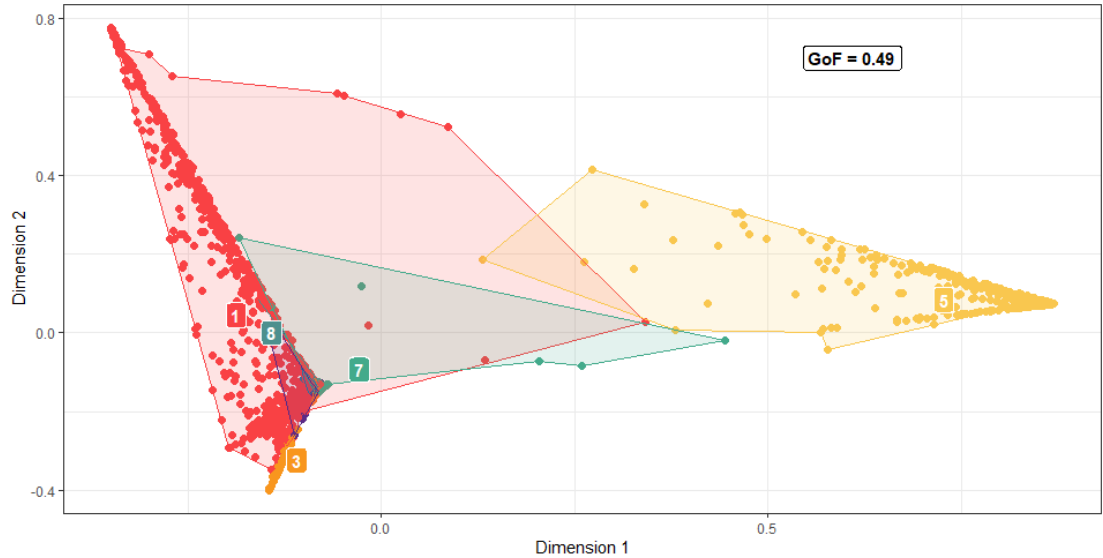


Figure ii.5: MDS of the transformed vessel catch data for the Portuguese PGP fleet. Points represent individual vessels, colored according to their cluster affiliation. Clusters are labelled, labels are colored similar to points. The goodness of fit (GoF) is labelled.

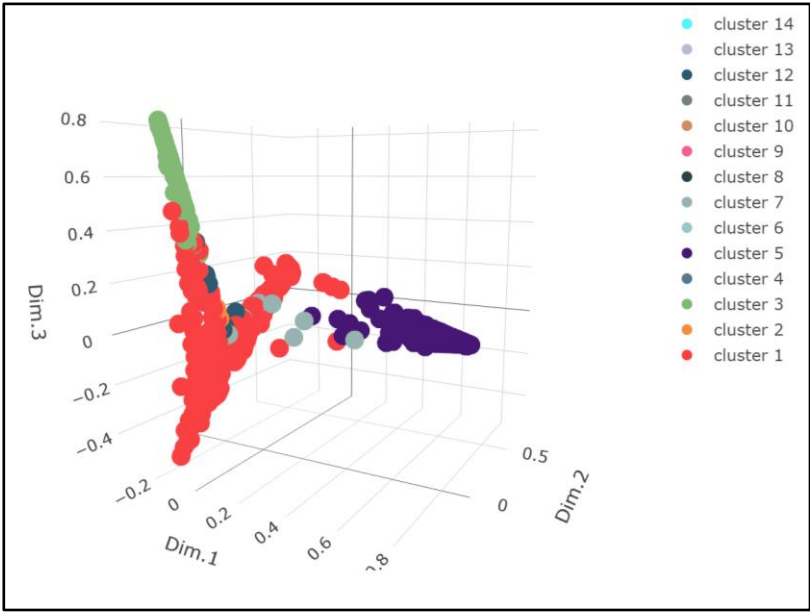


Figure ii.6: Snapshot of interactive 3D-MDS using Portuguese PGP data. Rotated to 2nd dimension being the x-axis, 3rd dimension the y-axis and 1st dimension the z-axis.

Vessel and cluster properties

The main part of this fleet is represented in cluster 1: higher number of vessels (Figure ii.7.A) and higher annual landings (Figure ii.7.C). This cluster also shows a very high variability in the vessel length ranging from 3 to 20m (Figure ii.7.B) and includes vessels both from the small-scale (<12m) and large-scale fisheries (≥12m). This fact may have some influence on the fishing pattern of these vessels because, p.e. smaller vessels may have lower capacity to operate with more than one gear/métier in the same trip, although they can change between trips. Vessels from this cluster present very high annual cluster landings and the landings per vessel are variable but rather low. The plots in Figure ii.7

also show other important species specific fisheries that can be detected when combining the number of vessels involved in the fishery and the annual volumes. In the case of the cockle fishery, represented by cluster 5, its importance is reflected by the number of vessels and by both the annual landings per vessel and per cluster. However, in the case of cluster 3, targeting sea lamprey, there is a significant number of vessels involved but the annual landed volumes are low. In this case, if species value was also taken into account the result could be more illustrative of the economic importance of this fishery. In the case of cluster 10, although having only 18 vessels included, it presents the higher landings per vessel and represents the 3rd cluster with higher annual landings. Given its target assemblage (mainly small pelagic fish) and the volumes landed, it may be questionable if they are fishing with passive gears. This fishing pattern can give some indications that those vessels may be operating with beach seines what can be very reasonable given that vessels from this fleet have multigear licenses.

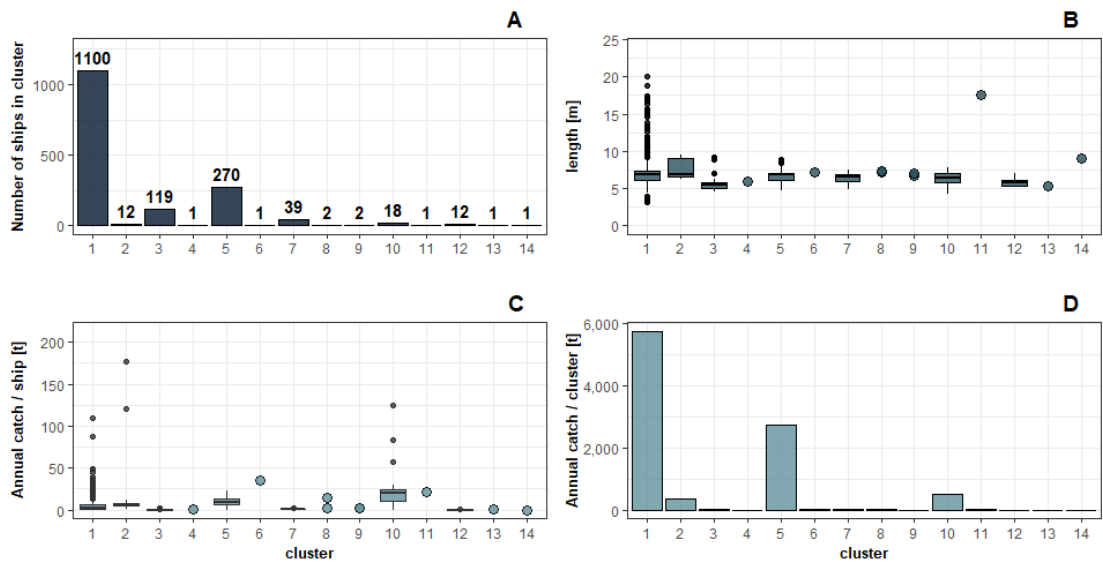


Figure ii.7: Grid of mixed plot types displaying vessel characteristics for the **Portuguese PGP fleet**. **A)** Barplot of cluster size with the number of ships in the cluster over each bar. **B)** Mixed box- and dotplot of the length of the vessels in meters, each cluster. **C)** Mixed boxplot of the annual catch of individual vessels in tons. **D)** Barplot of the total annual catch of all vessels in the clusters. (For B) and C), clusters containing more than 5 vessels are depicted with a boxplot, for smaller clusters each vessel catch is plotted with a point.)

Fleet segment assignment

Besides the characterization of the landings composition of the clusters, it was also considered the vessel length category in the classification of the fleet segments. Two groups of vessels were considered: <12 meters for Small-scale fishery and ≥12 meters for Large-scale fishery. Since some clusters only included 1 vessel assigned to a specific fishery (clusters 4, 6, 11, 13 and 14), they were grouped into a single fleet segment. The fleet segments assigned per cluster are listed below:

Cluster 1 - "Coastal mix fishery - traps and nets (Large-scale)"
- "Coastal mix fishery - traps and nets (Small-scale)"

Cluster 2 - "Crab fishery (Small-scale)"

Cluster 3 - "Sea lamprey fishery (Small-scale)"

Clusters 4, 6, 11, 13 and 14 grouped:

- "Other single vessel fisheries (Large-scale)"
- "Other single vessel fisheries (Small-scale)"

Cluster 5 - "Cockle fishery (Small-scale)"

Cluster 7 - "Carpet shells fishery (Small-scale)"

Cluster 8 - "Corvinata fishery (Small-scale)"

Cluster 9 - "Thisckback soles fishery (Small-scale)"

Cluster 10 - "Small pelagic fishery (Small-scale)"

Cluster 12 - "Eel fishery (Small-scale)"

Conclusions

From this analysis we may conclude that:

- A cluster with the majority of vessels and annual landings is detected (cluster 1 for mix fishery) and reflects the difficulty in classifying these trips/vessels where landing profiles indicate that different gears are used by vessel. In these cases, the analysis performed is not sufficient to classify the fisheries included in the cluster also because there are very well known target assemblages that cannot be differentiated within the high number of species detected;

- This approach identifies well-known fisheries (e.g. sea lamprey, eel and cockle fisheries), but for others maybe some finer scale criteria (e.g. subsetting the data) should be performed before the clustering procedure;

- The grouping of clusters/fisheries with only one vessel was performed but probably some other criteria should be used taking into account the minimum number required for reporting data from EUMAP.

iii) Segmentation of the Beam Trawlers (TBB), in ICES Division 27.9.a

This group of gear is composed of 48 vessels with a mean length of 7.5 m (range 4m - 11m), meaning that all vessels are included in the small-scale fishery. Vessels can have different licenses, but the ones for TBB are closely linked to the fishing area. The data used in the analysis includes also the landings of 98 different stocks (species*area).

Best number of clusters

The initial analysis of the fleet resulted in the diagnostics shown in **Figure iii.1**. The TBB fleet consists of homogeneous groups of vessels targeting multiple species. The graphics presented in **Figure iii.1** indicate 4 and 7 clusters as the best number according to the

average silhouettes and Mantel test respectively. However, these vessels catch and land a wide variety of species but they are very area related so considering 4 clusters seems appropriated, as a first approach.

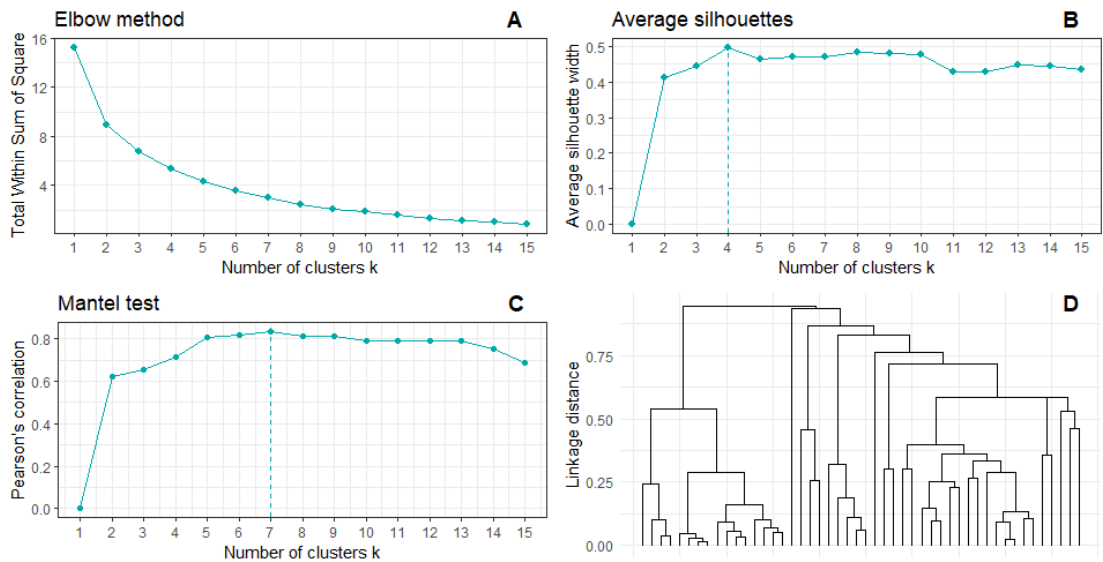


Figure iii.1: Diagnostic plots of the HAC procedure for the **Portuguese TBB fleet**. A) Screen plot showing the total within sums of squares of the clustering vs. the number of clusters used. B) Average silhouette width of the clusters vs. the number of clusters. C) Mantel test, i.e. the Pearson correlation between the clustering and the original distance matrix vs. the number of clusters. In B) and C), the best score is marked with a vertical line. D) dendrogram of the HAC procedure with the linkage distance of the clusters on the y-axis.

Figure iii.2 presents the dendrograms with different ways of manipulating the cutting height. In this case, the dendrograms support the conclusion that 1 to 11 distinct clusters can be identified in the data. We decided to rely on the number obtained from the average silhouette (Figure iii.1), selecting 4 clusters for the vessels grouping.

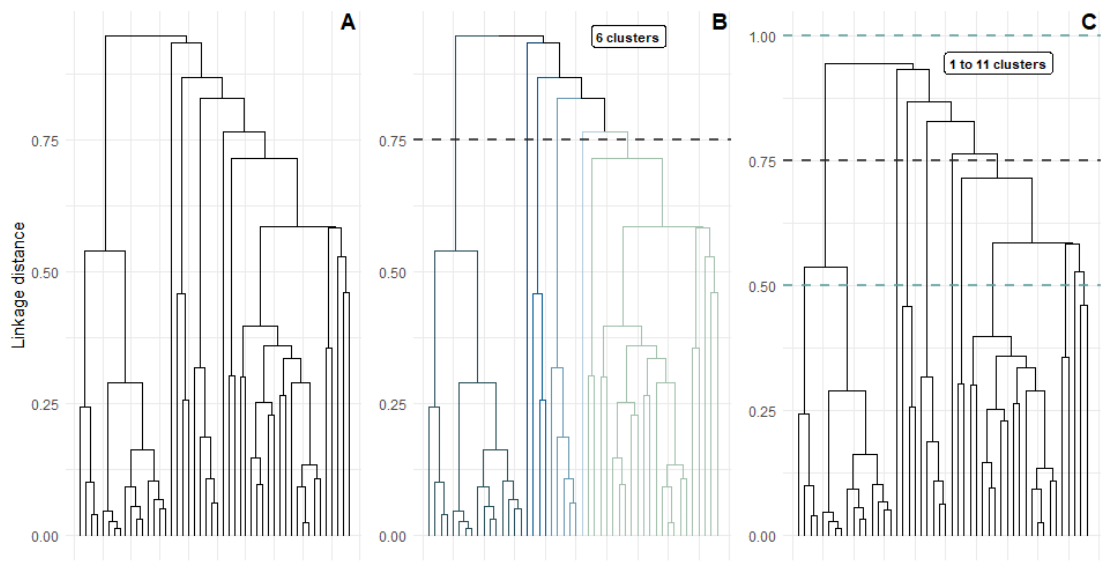


Figure iii.2: Gridded plot of three dendrograms of the HAC procedure for the **Portuguese TBB fleet**. The y-axis shows the linkage distance of clusters. A) is unmodified, B) is cut at a linkage distance of 0.75 and resulting branches are individually colored and the resulting number of clusters is labelled, C) has cutting lines at linkages distance 0.5, 0.75 and 1 and the range of cluster numbers resulting from those cutting heights is labelled.

The clustering tree (Figure iii.3) shows that considering more than 4 clusters would result in several clusters with low number of vessels so that may also indicate that number of clusters to be appropriate in this preliminary analysis. With this separation, two main clusters are identified including almost all vessels of the fleet (44 of 48 vessels in total).

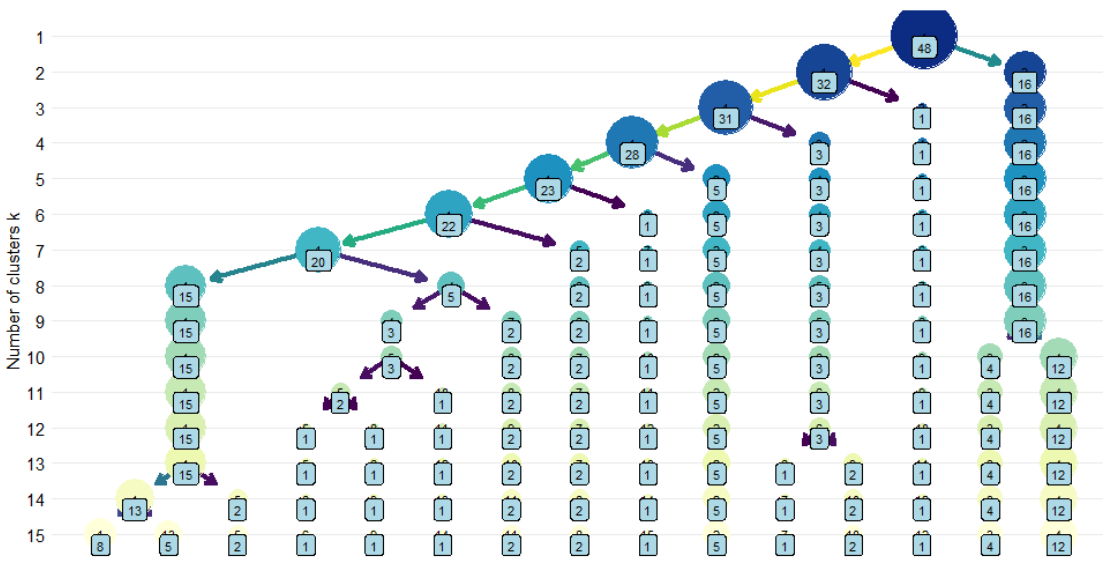


Figure iii.3: Treeplot of the HAC-procedure for the **Portuguese TBB fleet**. The circle size indicates the cluster size, which is also labelled. The number in the circle shows the number of the cluster and in the squares is the number of vessels. The number of clusters in the respective step is depicted on the y-axis.

Cluster characterization and validation

The clustering computation, using the four clusters, resulted in the stockshares presented in Figure iii.4. According to species composition, it is possible to identify the two main métiers for the beam trawlers from the Portuguese fleet (cluster 1 to TTB_MCD and cluster 2 to TTB_CRU). The target species of the other two clusters indicate that probably these vessels are misclassified. The species in cluster 3 give some indications that traps are eventually used, while for cluster 4 the species present in the landing profile (Figure iii.4) may be considered caught by hooks or nets. However these two last clusters should be deeper investigated to get the right gear/métier used by those vessels.

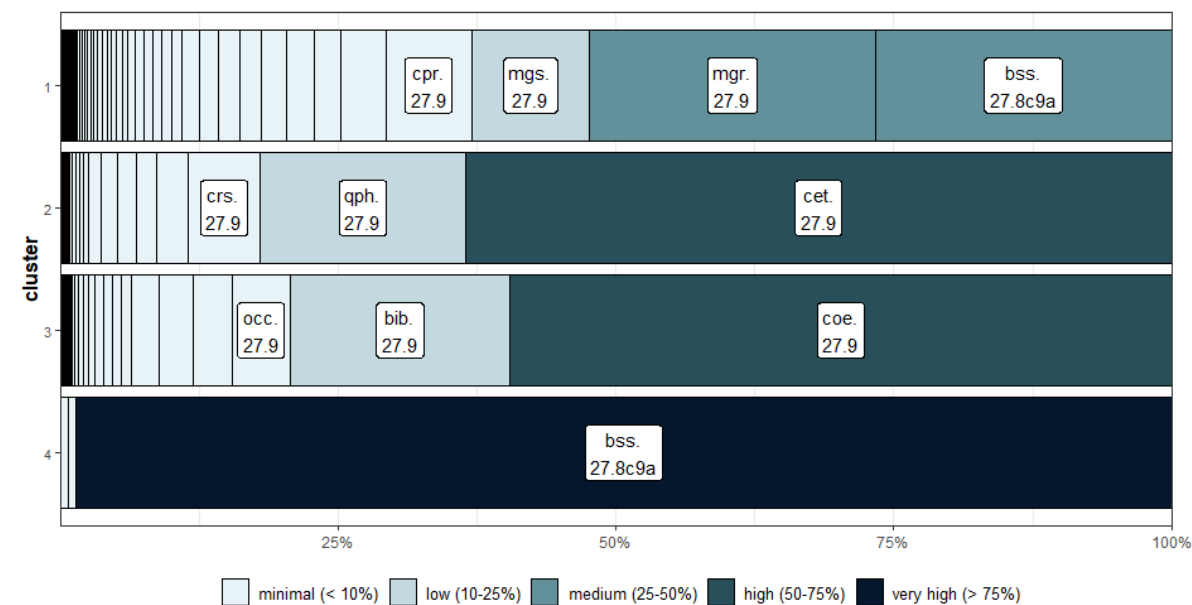


Figure iii.4: Barplot of the average percentage of each stock on the total landings of each cluster of the **Portuguese TBB fleet**. The average percentage is depicted on the x-axis, the clusters on the y-axis. The colors indicate the magnitude of the stocks contribution to the total average landings. The stock names are indicated with labels; stocks are ordered by their average percentage contribution.

The 2-dimensional MDS (Figure iii.5) and 3-dimensional MDS (Figure iii.6) confirm the presence of two main clusters (cluster 1 and cluster 2), very distant from each other. Regarding the other two clusters, and combining the information from the two figures, the vessel in cluster 4 seems to be very close to cluster 1 probably because the main species landed is also landed by vessels in cluster 4. On the other side, Figure iii.6 shows that cluster 3 is well separated from the others. The goodness of fit obtained in this analysis is quite good (GoF=0.62).

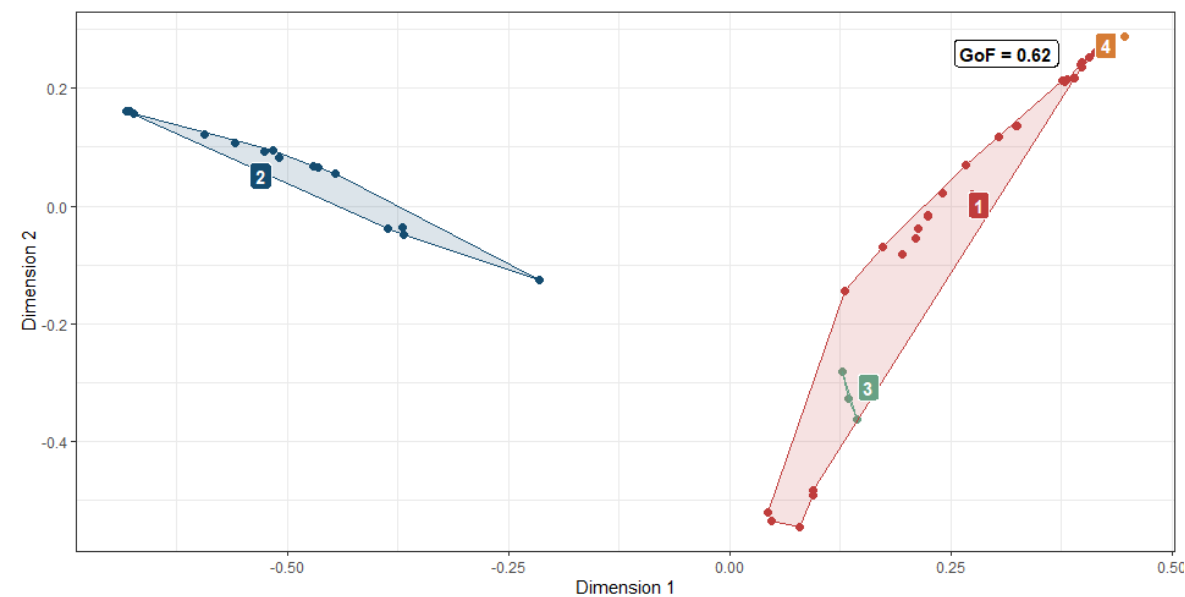


Figure iii.5: MDS of the transformed vessel catch data for the **Portuguese TBB fleet**. Points represent individual vessels, colored according to their cluster affiliation. Clusters are labelled, labels are colored similar to points. The goodness of fit (GoF) is labelled.

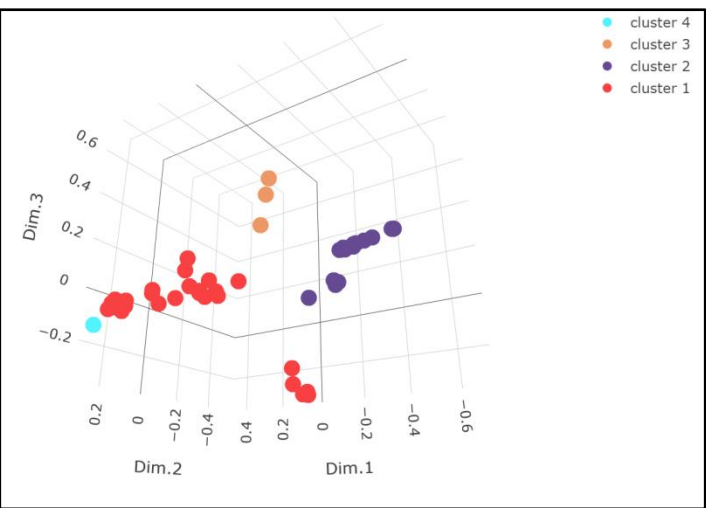


Figure iii.6: Snapshot of interactive 3D-MDS for the **Portuguese TBB fleet**. Rotated to 2nd dimension being the x-axis, 3rd dimension the y-axis and 1st dimension the z-axis.

Vessel and cluster properties

The clusters including higher number of vessels are the clusters 1 and 2 (Figure iii.7.A). These correspond to the main métiers for this fleet (TTB_CRU and TTB_MCD respectively). Vessels in cluster 1 are smaller and the lower landings per vessels and cluster is the expected for a crustacean fishery. However, if considering value instead of weight probably this pattern would be inverted. The cluster 2 has bigger vessels (Figure iii.7.B), able to 'accomodate' higher volumes of landed species, which can be confirmed by the higher landings per vessel and per cluster observed (Figures iii.7.C and D). The other two clusters are composed by vessels that were probably misclassified and should not be included in this fleet.

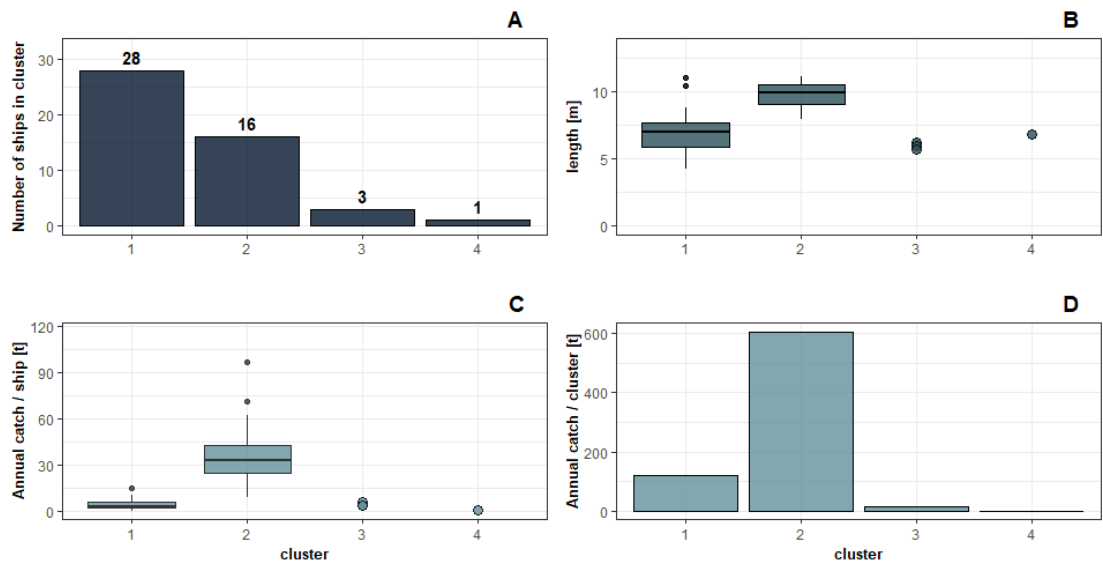


Figure iii.7: Grid of mixed plot types displaying vessel characteristics for the **Portuguese TBB fleet**. **A)** Barplot of cluster size with the number of vessels in the cluster over each bar. **B)** Mixed box- and dotplot of the length of the vessels in meters, each cluster. **C)** Mixed boxplot of the annual catch of individual vessels in tons. **D)** Barplot of the total annual catch of all vessels in the clusters. (For B and C, clusters containing more than 5 vessels are depicted with a boxplot, for smaller clusters each vessel catch is plotted with a point.)

The characterization and fleet segment assigned per cluster are listed below:

- Cluster 1** - includes main part of the fleet; target mixed crustaceans (prawns) and fish (seabass and mullets) according to the landing proportions; but low catches per vessel may indicate that they target specially prawns -> TBB_CRU
- Cluster 2** - includes also a good part of the fleet, with bigger vessels; target mixed fish (wedge soles and others) and crustaceans (crabs); the high annual catches may indicate that this is a TBB_MCD fishery
- Cluster 3** - according to landings profile this are vessels probably working with traps/pots
- Cluster 4** - vessel that probably uses hooks or nets

Conclusions

From this analysis we may conclude that:

- This approach identified the main fisheries of this fleet: TBB_MCD and TBB_CRU;
- When crustacean species are present in the data, one may also consider to perform the analysis accounting for the value of the landings and not only the weight;
- This analysis allowed the identification of vessels that are misclassified.

CONCLUSIONS FROM THE WORKSHOP ON A NOVEL APPROACH TO THE SEGMENTATION OF FISHING FLEETS

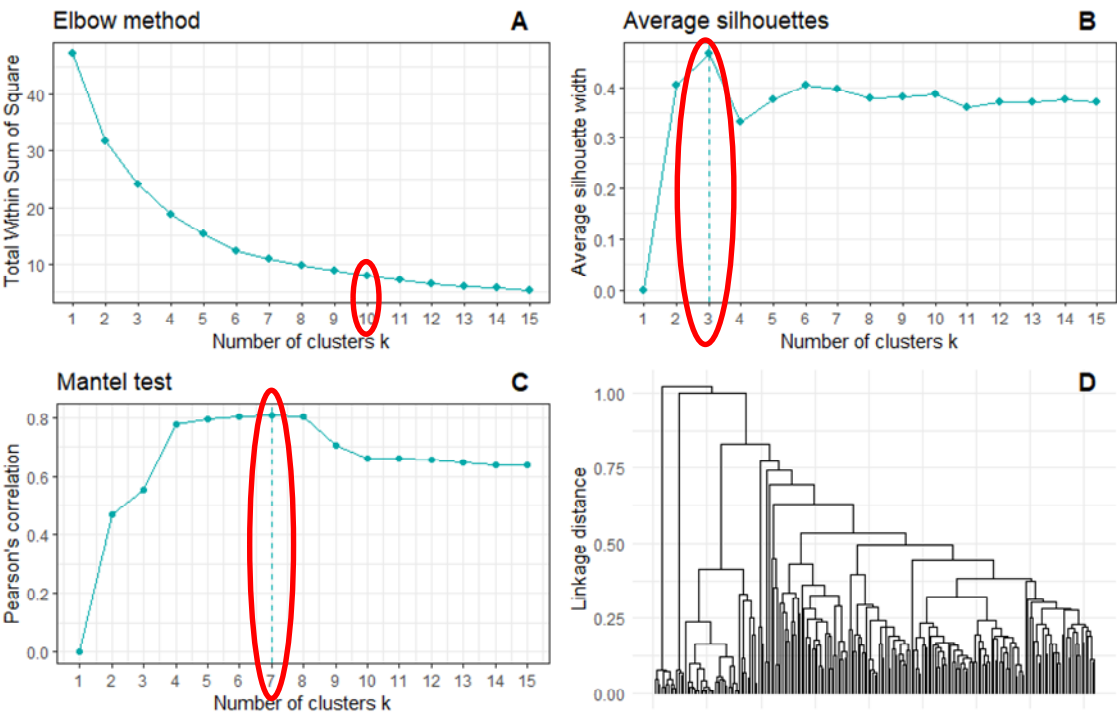
INTRODUCTION

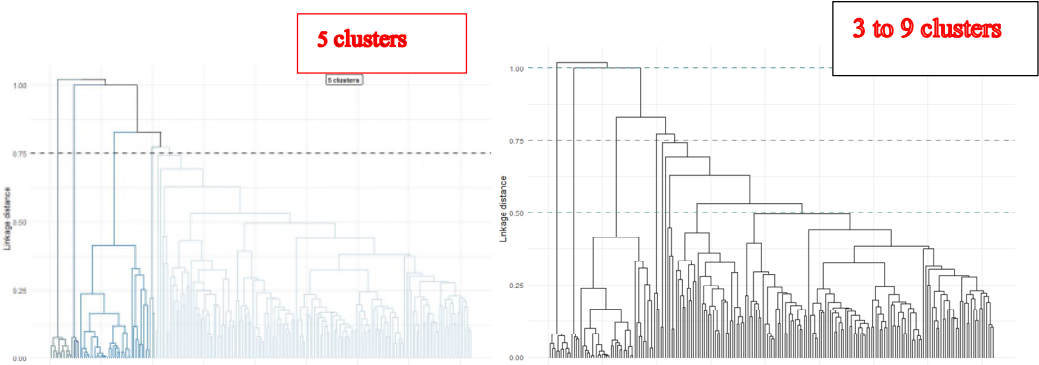
Spain has applied the “Fleet Segmentation-Package” which contains ready-to-use functions to apply a multivariate approach on the segmentation of fishing fleets. The present approach introduces a standardized multivariate approach for characterizing fisheries fleet segments by hierarchical agglomerative cluster analysis (HAC) of their **catch composition**.

Spain has applied this tool to our trawlers of 2019 which have fished on 27.VIII.c and 27.IX.a areas. This float is composed by 205 vessels.

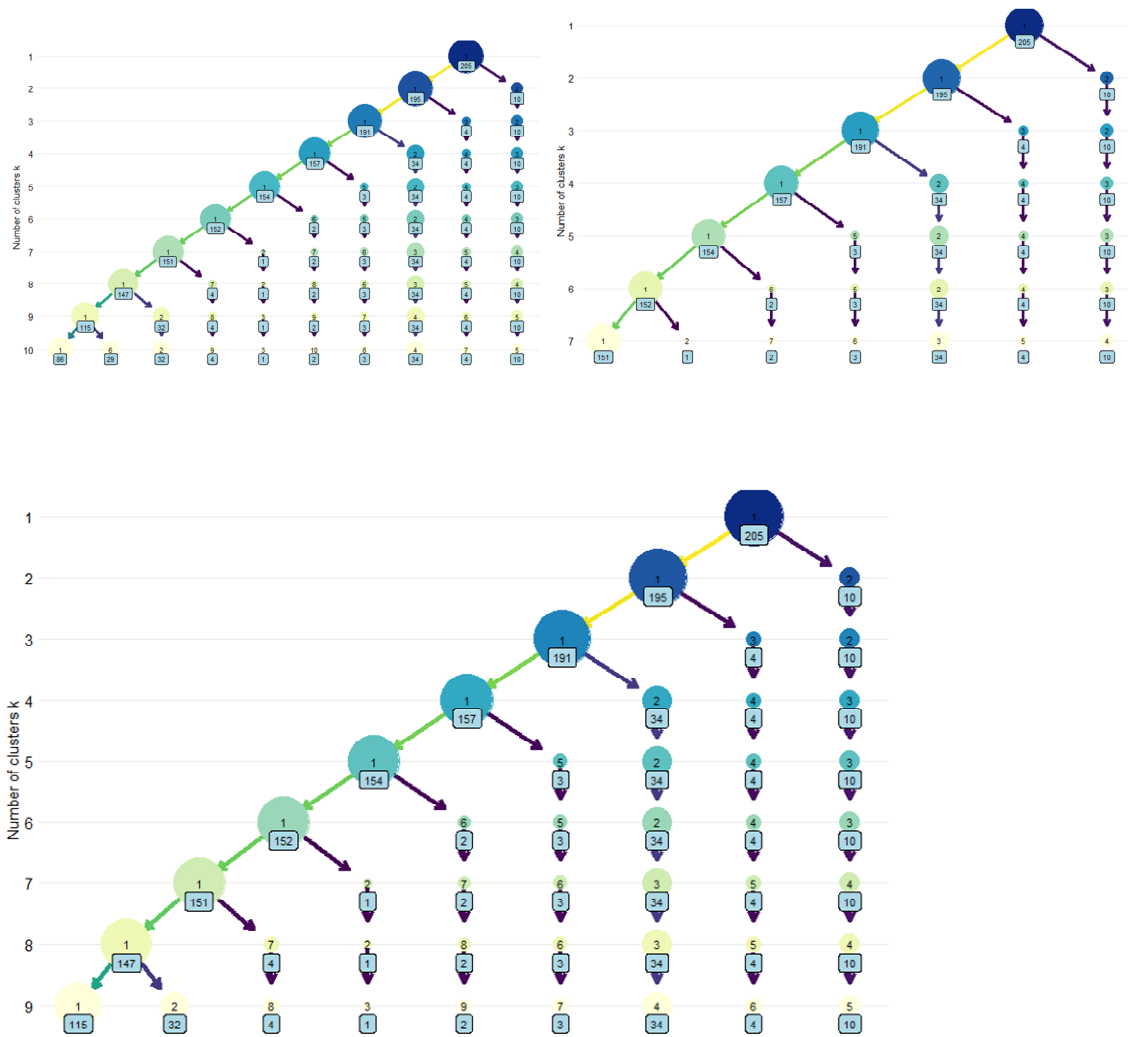
IMPLEMENTATION

The first step for doing the segmentation is to decide how many clusters we want to have. In order to do it, the Package use five essential indices which estimate the best number of clusters. Moreover, the program uses other tests as we can see in the following graphics. Agree with this, the best number of clusters for our trawlers vary between 3 to 10.

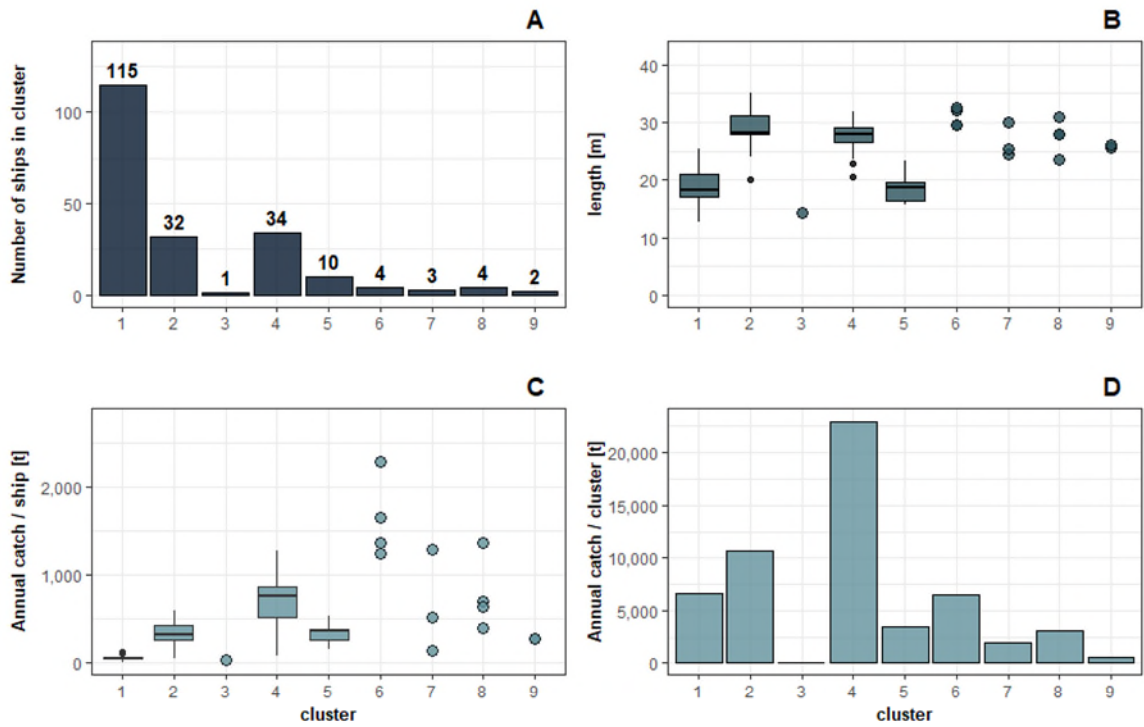
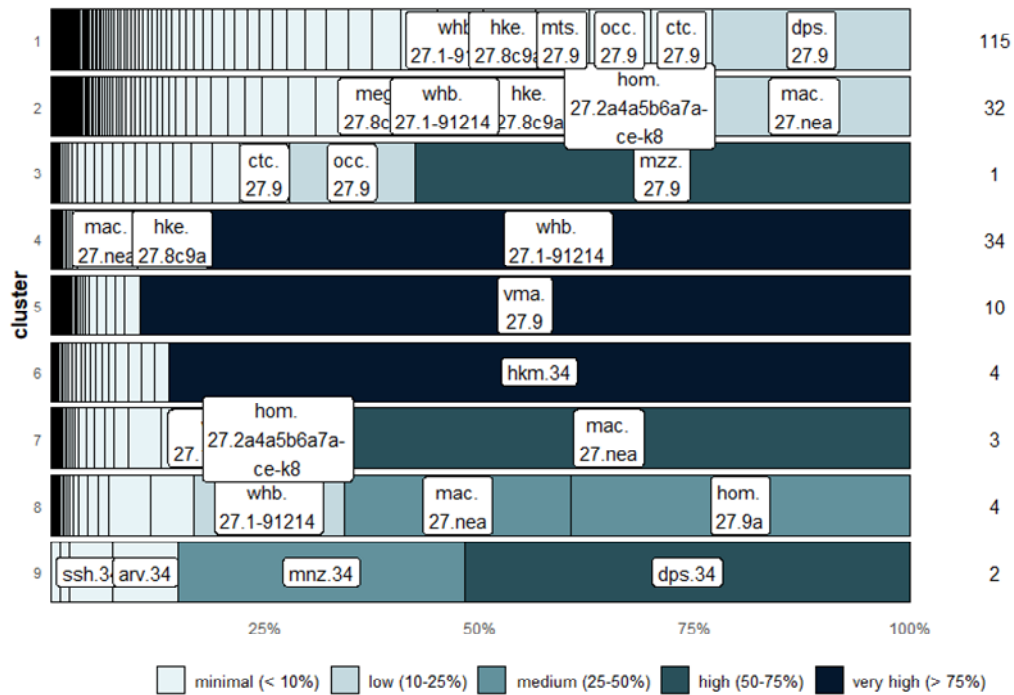




In order to take a final decision, we have used the “clustering tree”. We have worked with 10, 9 and 7 clusters. After studying all the assumptions, we have decided that the best number of clusters is 9.



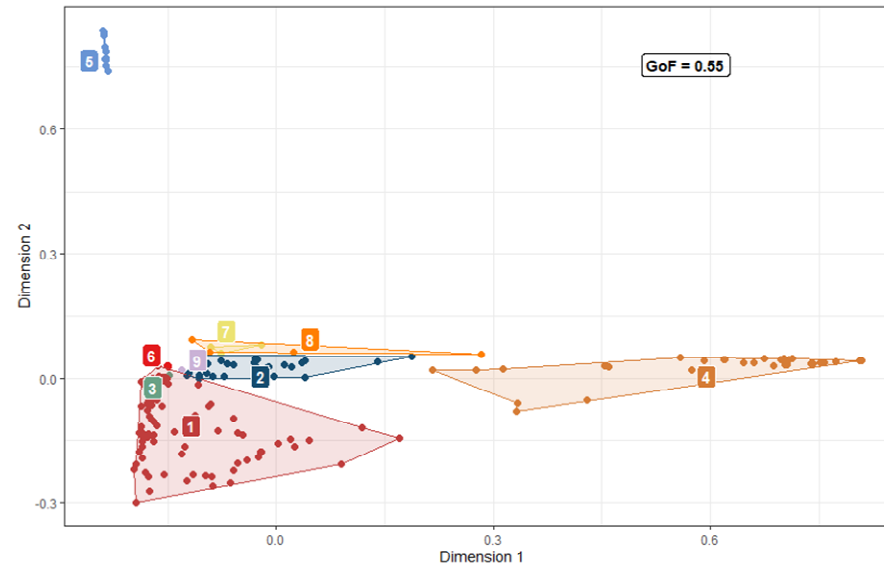
Step 2: According to the differences in catch composition, the Package divides the fleet in the number of clusters desired.
Considering 9 clusters, the vessels distribution was the following:





Step 3. It is possible to unit clusters if we consider they are similar. For that, the Package includes functions for 2 and 3-dimensional ordination methods in order to show the compositional distances of cluster and identify overlaps.

We have only used the 2-dimensional ordination method because our program was not able to open the 3-dimensional ordination.



According to this, we have decided to join:

- Cluster number 1 and 3: cluster 3 is composed just by one vessel and it is better to join it with another cluster. Moreover, in cluster 3 the majority of their catches were “marine fishes” and other species such as octopus and cuttlefish which are also included in cluster 1.
- Cluster number 7 and 8: both clusters have few vessels (3 and 4 vessels respectively) and both catch mackerel.

Step 4. Final segmentation

- Cluster 1 + 3: 116 vessels. They caught crustacean and mollusks. Individually, their catches are not really representative, but together represent 5.000 Tn.
- Cluster 2: 32 vessels. These vessels caught mackerel, blue whiting, hake and so on. They caught 10.000 Tn.
- Cluster 4: 34 vessels. The main specie caught was blue whiting (over 75%). They caught more than 20.000 Tn
- Cluster 5: 10 vessels. The main specie caught was starling (over 75%), their catches represent less than 5.000 Tn.
- Cluster 6: 4 vessels. The main specie caught was hake (over 75%). They caught more than 5.000 Tn.
- Cluster 7 + 8: 7 vessels. The main specie caught was mackerel (between 50-75%). They caught around 5.000 Tn.
- Cluster 9: 2 vessels. The main specie caught was prawn (between 50-75%).



CONCLUSIONS

- The Fleet Segmentation-Package is really useful for dividing the vessels that fish in different regions and different species. It has been important for us, in order to separate the trawlers that fish in the Gulf of Cadiz from the trawlers that fish in the CNW (27.VIII.c), whose economic, biologic and technical characteristics are really different. In this sense:
 - o Cluster 1+3 and 5 fish in the Gulf of Cádiz
 - o Cluster 2, 4 and 7+8 fish in CNW
 - o Cluster 6 and 9 fish in Portugal.
- The program not use the variable “days at sea” or “the value” of their catches that could be interesting for the segmentation. For this reason, some vessels that we have considered that spend most of the time in Portugal has been classified in Gulf of Cádiz or CNW because of the weight of their catches.
- A lot of clusters show few vessels. It makes the economic indicator’s calculation unviable.
- Basing the segmentation on catches, It seems that, could generate a high variability between one year and the following, so it would be really difficult to apply measures a long time in an action’s plan.

To sum up, the Fleet Segmentation-Package is a good tool as a complement to the actual segmentation based on gear class and vessel length, but it seems difficult to apply as a main tool for the segmentation.