| Alternative Approach to the Segmentation of Fishing Fleets<br>*Daniel Grigoraș, Cătălin Valentin Păun, George Țiganov, Cristian Sorin Danilov, Dragoș Diaconu* | **"Cercetări Marine" Issue no. 53**<br>**Pages 92-105** | **2023** |
|---|---|---|
| DOI: 10.55268/CM.2023.53.92 | | |

# ALTERNATIVE APPROACH TO THE SEGMENTATION OF FISHING FLEETS

**Daniel Grigoraș*, Cătălin Valentin Păun, George Țiganov, Cristian Sorin Danilov, Dragoș Diaconu**

*National Institute for Marine Research and Development "Grigore Antipa", 300 Mamaia Blvd., Constanta 900581, Romania*
**Corresponding author: dgrigoras@alpha.rmri.ro*

## ABSTRACT

The appropriate segmentation of fishing fleets is controversially discussed in fisheries research and management and a variety of approaches has been introduced. The present approach, developed in a pilot study funded by the European Commission – Data Collection Framework (DCF), introduces a standardized multivariate approach for characterizing fisheries fleet segments by hierarchical agglomerative cluster analysis (HAC) of their catch composition. We chose data from 2021 of the Romanian fishing fleet as the basis of our analysis. Statistical analyses were performed using the program RStudio V3.6.1 by running the fleet segmentation package script. The specific indices, tests, and visual validation methods of the package were applied to determine the optimal number of clusters. The procedure was finalized by a post-hoc validation of the clustering result to identify the actual fleet segments. From the basic data, 6 fleet segments for Vessels using active and passive gears (PMP) were highlighted, representing 52 boats, where it was noted that fishing at Rapa whelk (RPW) prevailed with 91.94% of the total catches on the segment, respective 5 fleet segments for Vessels using passive gears (PG) only for vessels <12m, representing 78 vessels where the main catches or recorded at European anchovy (ANE) 25.16%, turbot (TUR) 13.08%, horse mackerel (HMM) 8.75%, thus 130 vessels from two classes of different gears in total. We detected mixed fishing, especially on various assemblages of demersal and pelagic fish, as well as target fishing on demersal and pelagic fish, Rapa whelk, and mussels. For a better understanding of the approach, further research is needed.
**Keywords:** Romanian fleet segmentation, catch composition, Black Sea

## AIMS AND BACKGROUND

As of the end of the year 2021, the fishing fleet comprised 163 vessels with a total capacity of 1575.26 GT and 6198.29 kW. Out of these, 130 vessels were active, and they were categorized into two fishing techniques: 52 vessels using Vessels using active and passive gears (PMP) and 78 vessels using Vessels using passive gears (PG).

For an accurate and realistic analysis of the fishing fleets, it is crucial to adopt a new approach to fleet segmentation, different from the current one. The existing segmentation approach relies only on the dominant gear class and vessel length (LOA).

The new approach to fleet segmentation should involve using statistical programs capable of handling large volumes of data. Specific scripts can be developed to process, analyze, and interpret the data, generating comprehensive and detailed sets of information.

By implementing this approach, member states can ensure the quality of data transmitted on the data collection platforms in the fisheries sector, meeting the requirements of the European Union.

## EXPERIMENTAL

Statistical analyses were conducted using the RStudio V3.6.1 program by executing the fleet segmentation package script (Fig. 1). This package includes functions for a multivariate approach to fleet segmentation in target fisheries. It was developed by Erik Sulanke, a research associate at the Thuenen Institute for Marine Fisheries in Bremerhaven, Germany, economics subunit. The pilot project that led to the development of this package was established in the DCF of Scientific, Technical and Economic Committee for Fisheries (STECF).

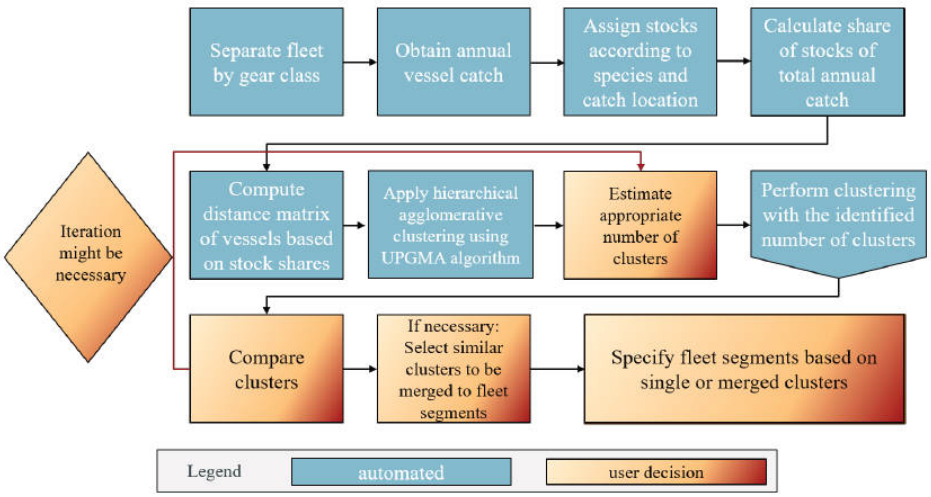


**Fig. 1.** Updated stepwise flowchart of the newly developed fleet segmentation approach (Sulanke E. and Berkenhage J., 2022)

**Preparing and Loading fleet data.** This first step describes uploading data about our fishing fleet. The respective data were structured in a framework of generated data that will be needed later for grouping the data into clusters.

The vessels were grouped according to the type of fishing gear they use in two fishing techniques, combining mobile and passive gears (PMP) and vessels with passive gears (PG), this separation was necessary because most of the vessels in the fishing fleet in Romania use a combination of types of gear, both the active ones and the most active ones.

This segmentation procedure uses two input data frames, one frame containing catch data for all vessels of a certain gear class and the second frame containing information about the length of the vessels (Sulanke, 2022).

**Examine the best number of clusters to use.** In the second step, technical knowledge related to the segmentation of fishing fleets is necessary to correctly interpret the data and determine the appropriate number of clusters.

The script developed by Sulanke 2022, utilizes several indices and tests in this step to make an accurate choice of the number of clusters. To obtain the most reliable grouping results, it is necessary to run several scripts with different numbers of clusters. This allows for result comparison and facilitates the final decision-making process.

The indices presented in (Table 1 and 2) act as measures for evaluating different cluster configurations and help identify the most suitable number of clusters for the given dataset.

**Table 1.** Indices optional of clusters for PMP vessels

| **Indices optional** | **No. of clusters** | **Index value** |
|---|---|---|
| Average silhouettes* | 8 | 0.788 |
| Mantel test* | 9 | 0.949 |
| Davis_Bouldin index* | 14 | 0.088 |
| SD index* | 5 | 2.884 |
| Calinski-Harabasz indexv* | 15 | 1129.278 |

*Average silhouettes – The approach measures the quality of a clustering. That is, it determines how well each object lies within its cluster. A high average silhouette width indicates good clustering.

*Mantel test – The matrices must be of the same dimension; in most applications, they are matrices of interrelations between the same vectors of objects.

*Davi_Bouldin index – This is an internal evaluation scheme, where the validation of how well the clustering has been done is made using quantities and features inherent to the dataset.

*SD index – The standard deviation index is a measurement of bias (how close your value is to the target value)

*Calinski-Harabasz index – Clustering validation has been recognized as one of the important factors essential to the success of clustering algorithms. How to effectively and efficiently assess the clustering results of clustering algorithms is the key to the problem

**Table 2.** Indices optional of clusters for PG vessels

| Indices optional | No. of clusters | Index value |
|---|---|---|
| Average silhouettes* | 9 | 0.559 |
| Mantel test* | 15 | 0.861 |
| Davis_Bouldin index* | 11 | 0.531 |
| SD index* | 6 | 2.11 |
| Calinski-Harabasz index* | 14 | 44.184 |

For our segmentation, we selected the average silhouette index as the most representative measure for both PMP and PG fishing techniques, resulting in 8 and 9 clusters, respectively.

**Test plot clustering diagnostics.** By applying this hierarchical agglomerative clustering (HAC) procedure (Fig. 2 and 3), with various graphical tests, indices, and methods, the correct number of clusters in a dataset can be determined.

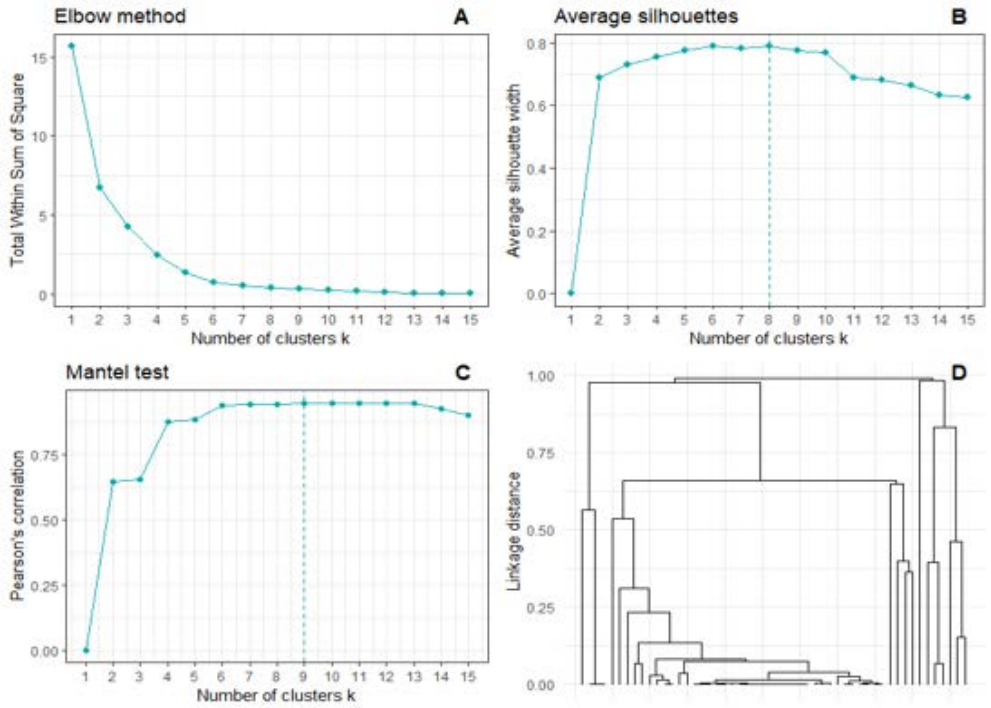


**Fig. 2**. Gridded diagnostic plots of the HAC procedure – PMP vessels

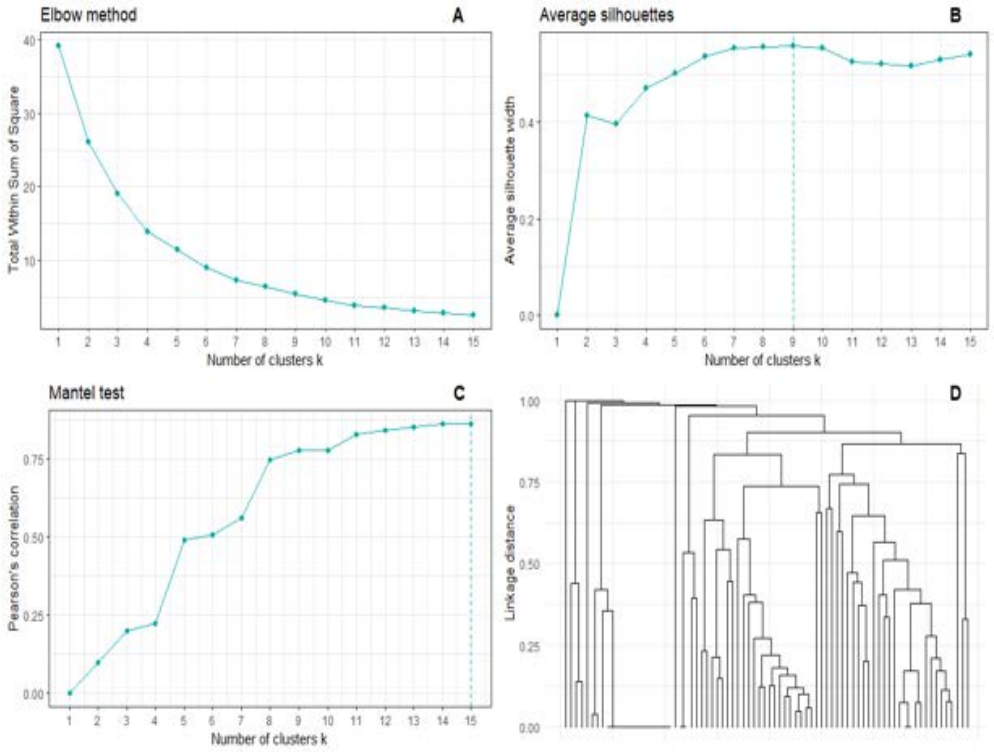


**Fig. 3.** Gridded diagnostic plots of the HAC procedure – PG vessels

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.
D) dendrogram of the HAC procedure with the linkage distance of the clusters on the y-axis. The y-axis shows the linkage distance of clusters (Sulanke, 2022).

In the plots generated by the script for both fishing techniques (Fig. 2 and 3) at the average silhouette index, it can be observed that the number of clusters remains 8 for PMP and 9 for PG, which is the same number of clusters as in the previous step.

**Test dendrogram clustering diagnostics.** In the continuation of the Hierarchical Agglomerative Clustering (HAC) procedure, dendrograms are used (Fig. 4 and 5). Dendrograms are graphical representations that display the hierarchical clustering results obtained during the agglomeration process. As the HAC algorithm iteratively merges similar data points or clusters, the dendrogram visually illustrates these merging steps.

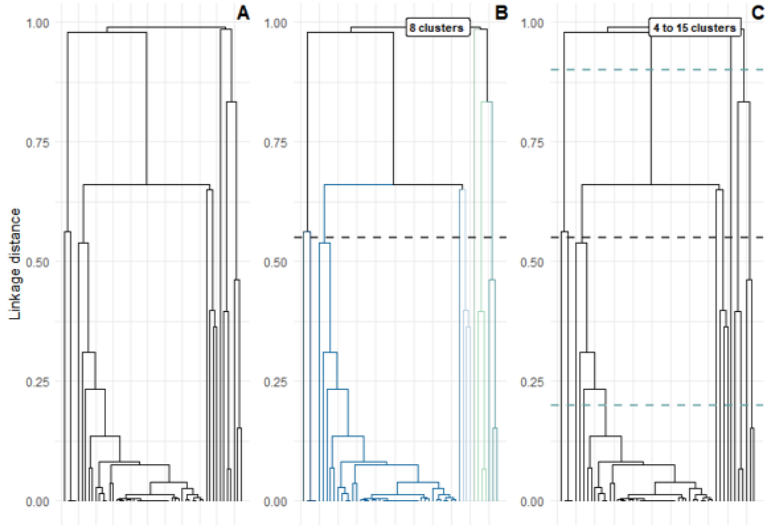


**Fig. 4.** Grided plot of three dendrograms of the HAC procedure – PMP vessels

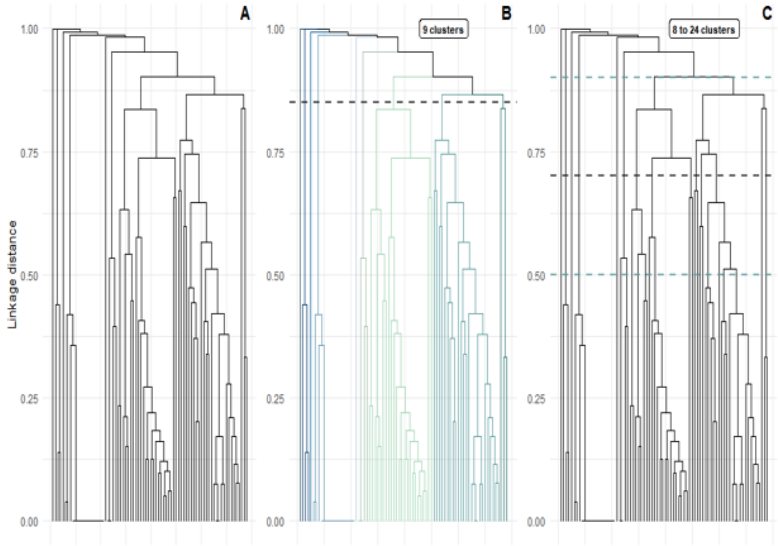


**Fig. 5**. Grided plot of three dendrograms of the HAC procedure – PG vessels

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 labeled.
C) has cutting lines at linkages distance 0.5 and 0.9 and the range of cluster numbers resulting from those cutting heights are labeled (Sulanke, 2022).

Based on the information provided in the dendrograms, it becomes apparent that they both support the conclusion that there are 8 to 9 distinct clusters in the dataset. A dendrogram visually represents the hierarchical

clustering process and illustrates how data points or clusters are grouped together based on their similarity or distance from each other.

**Plot clustering tree**. During the various tests for analyzing the number of clusters, we experimented with different combinations of cluster configurations and evaluation measures. This was done to assess the quality and suitability of the clustering solutions at each step of the analysis. Toward the end of the test, the clustering tree procedure was introduced. A clustering tree is another method of clustering analysis, and it serves as a visual representation of the hierarchical clustering process.


**Fig. 6.** Tree plot of the HAC-procedure – PMP vessels

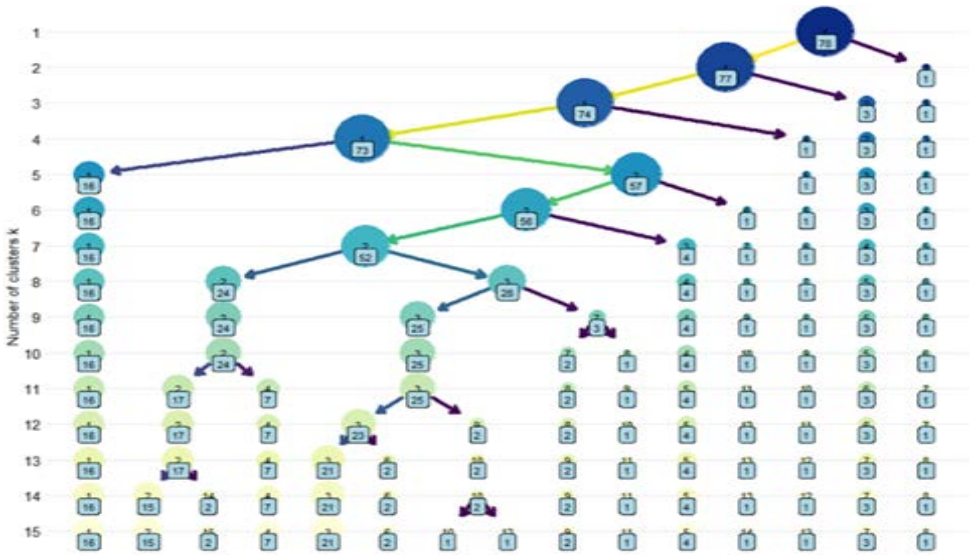


**Fig. 7.** Tree plot of the HAC-procedure – PG vessels

The clustering tree makes it easy to trace back changes that occur when new clusters are introduced. When a sixth group is added, cluster number 1 is divided into two new groups: one contains 37 vessels, and the other contains 4 vessels for PMP, while for PG, one contains 16 vessels and the other contains 57 vessels. These newly formed groups may represent valid fleet segments.

## RESULTS AND DISCUSSION

Following the HAC procedure, several variants of cluster selection resulted from the entered data set. After the correlation and analysis of each variant separately, it was established that the most correct variant is 8 clusters for the PMP vessels and 9 clusters for the PG vessels.

**Catch composition of clusters**. According to the script, the first step for cluster characterization is to evaluate their catch (Fig. 8 and 9). The average percentage that each stock contributes to the catches of a cluster can be tabulated and graphed (Sulanke, 2022).

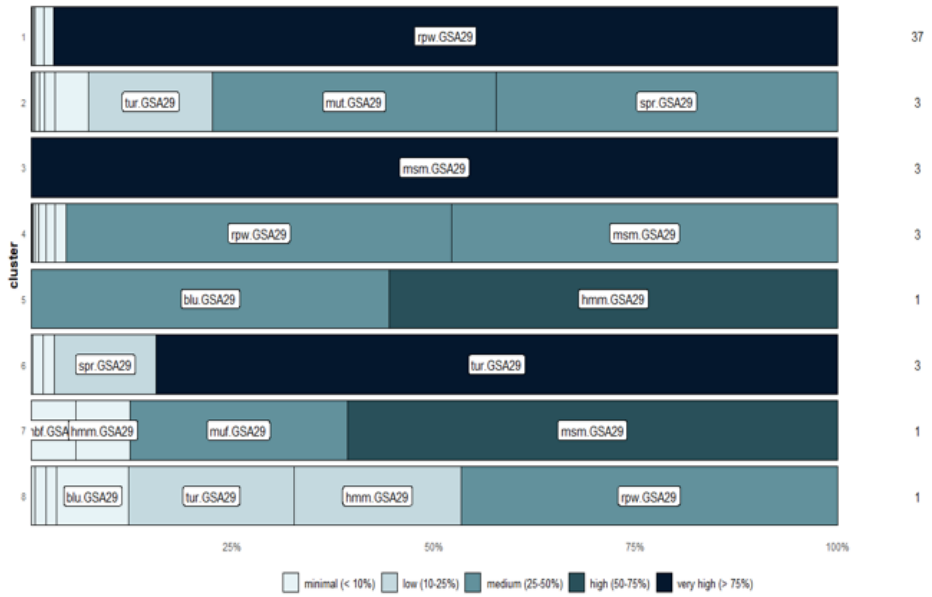


**Fig. 8**. Bar plot of the average percentage of each stock on the total catch of each cluster, PMP vessels

The average percentage is depicted on the x-axis, the clusters on the y-axis. The colors indicate the magnitude of the stock's 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 (Sulanke, 2022).

From the basic data, 6 fleet segments for Vessels using active and passive gears (PMP) were highlighted, representing 52 boats, where it was

noted that fishing at Rapa whelk (RPW) prevailed with 91.94% of the total catches on the segment, respective 5 fleet segments for Vessels using passive gears (PG) only for vessels <12m, representing 78 vessels where the main catches or recorded at European anchovy (ANE) 25.16%, turbot (TUR) 13.08%, horse mackerel (HMM) 8.75%, thus totaling 130 vessels from two classes of different gears. We detected mixed fishing, especially on various assemblages of demersal and pelagic fish, as well as target fishing on demersal and pelagic fish, Rapa whelk and mussels.

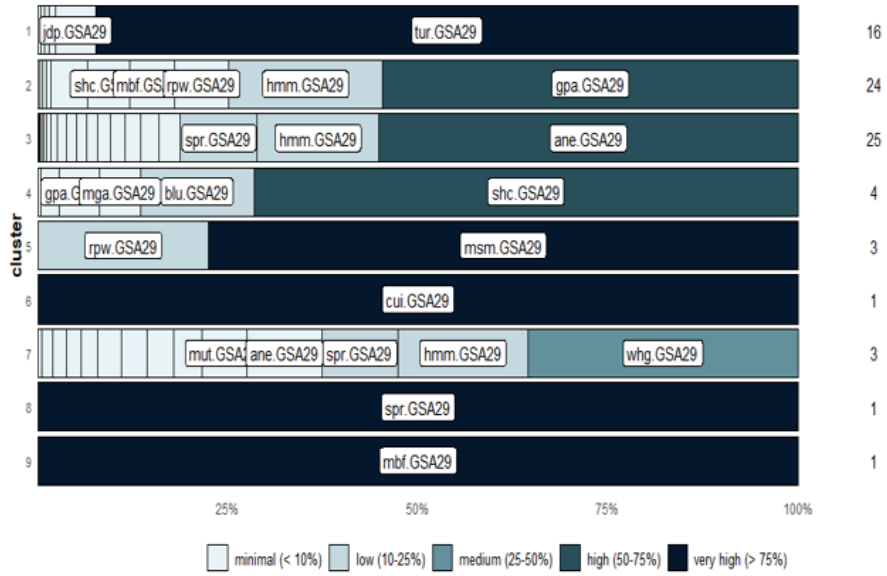


**Fig. 9.** Bar plot of the average percentage of each stock on the total catch of each cluster, PG vessels

**Plotting a 2-dimensional.** This method of ordering in 2 dimensions helps with a better visual identification of the overlaps and therefore, the clusters can be united by matching them in the specific segment of the fleet (Fig. 10-11). In this step of the script, a metric MDS was used because it reflects real distances, not ranked distances (Sulanke, 2022).

Points represent individual vessels, colored according to their cluster affiliation. Clusters are labeled, and labels are colored like points. The goodness of fit (GoF) is labeled. This MDS reveals some of the relationships in the catch composition and shows a good fit (Sulanke, 2022).

As can be seen in the PMP vessels (Fig. 10), a matching of the vessels that form cluster 1 (which includes most of the vessels) and cluster 4 can be distinguished, because they are consistent with their catch composition that fish the same stocks, but in different amounts. Regarding the PG vessels (Fig. 11), it can be clearly observed that most of the clusters are in the right area of the MDS forming 2 representative clusters that have most of the vessels in their competence, are cluster 2 and 3, it can also be observed that

cluster 1 is at a great distance from the other clusters, having turbot and common stingray in its composition.

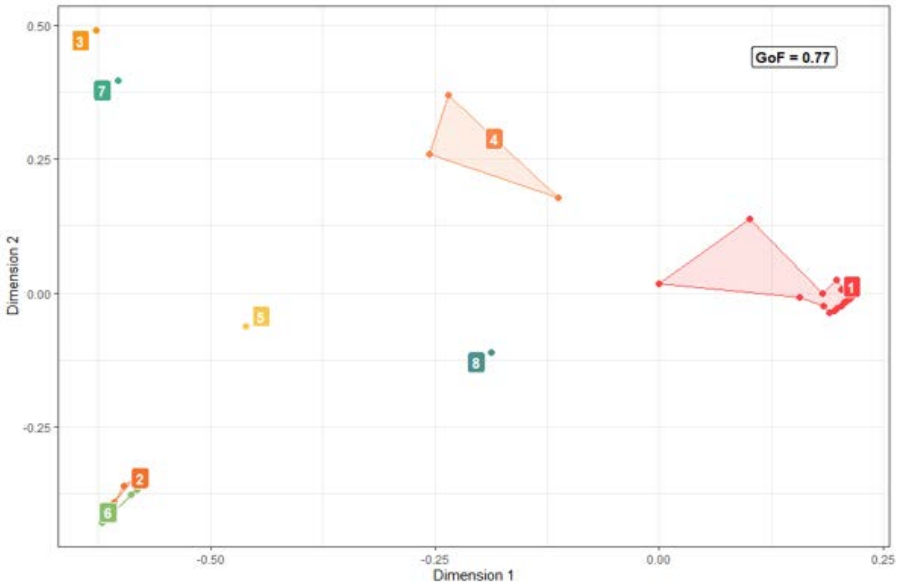


**Fig. 10.** MDS of the transformed vessel catch data – PMP vessels

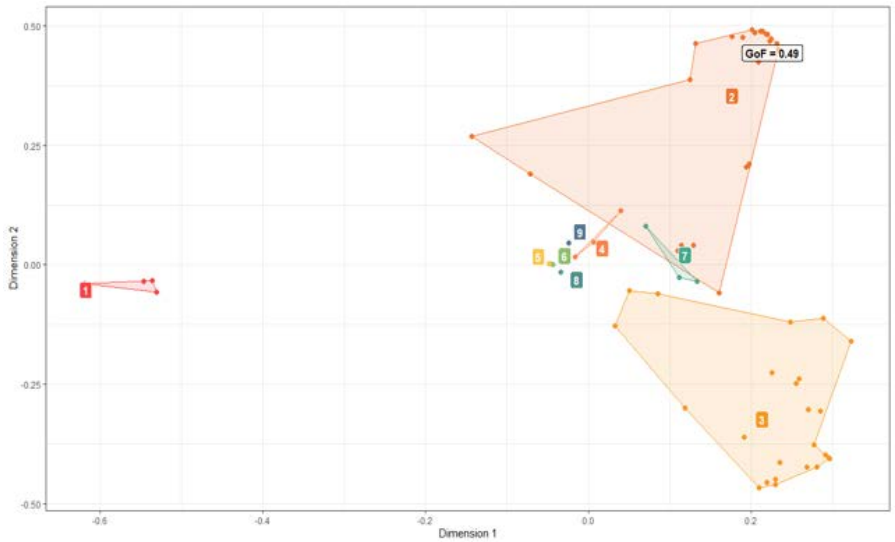


**Fig. 11**. MDS of the transformed vessel catch data – PG vessels

**Vessel and cluster properties.** This last step characterizes the clusters obtained based on the available data: the number of vessels in the clusters, the length distribution of the vessels in the cluster, the total catch of individual vessels, and the total catch of each cluster (Fig. 12-13), (Sulanke, 2022).

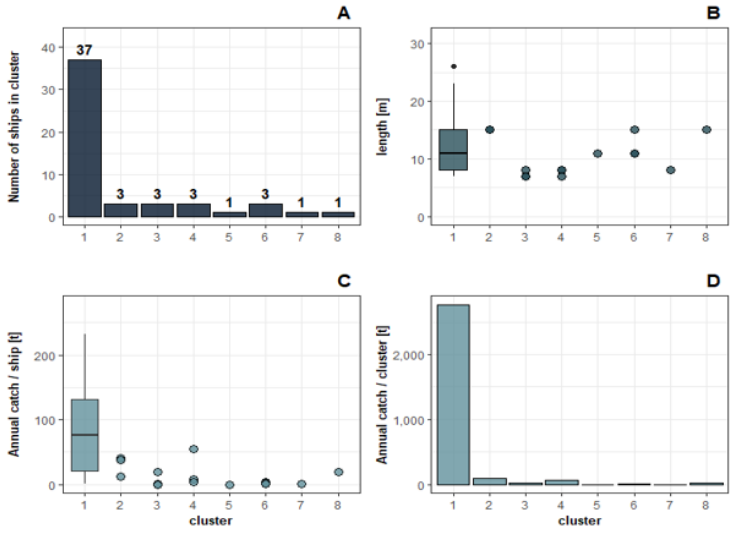


**Fig. 12.** Grid of mixed plot types displaying vessel characteristics – PMP vessels

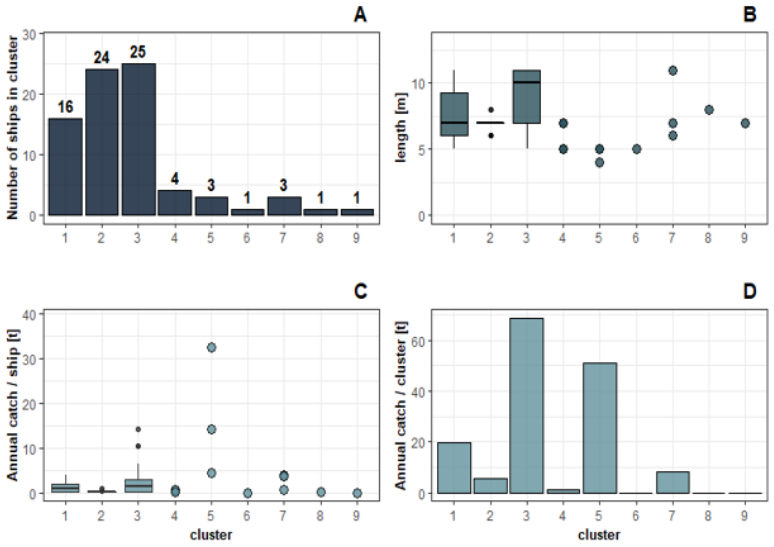


**Fig. 13.** Grid of mixed plot types displaying vessel characteristics – PG vessels

A) number of ships in each group is plotted on the y-axis versus the group number on the x-axis. The number above each bar gives the number of vessels in the group.
B) vessel length in meters is plotted on the y-axis versus the group number on the x-axis.
C) annual catches by vessels in tones on the y-axis versus group number on the x-axis.
D) total catch in tons is represented on the y-axis versus the group number on the x-axis (Sulanke, 2022).

This graphical grid contains a lot of information about the clusters and allows us to group them according to their structure.

Table 3 provides information on fishing vessel activity based on two fishing techniques, PG and PMP. Table 4 highlights the activities of the fishing vessels belonging to the most representative clusters within the PMP fishing technique. More precisely, it shows the distribution of vessels in three clusters, with cluster 1 having 37 vessels, and clusters 2 and 4 each having 3 vessels. Similarly, (Table 5) presents the activities of fishing vessels within the PG fishing technique, focusing on the most representative clusters. The table indicates that cluster 1 consists of 16 vessels, cluster 3 has 25 vessels, while clusters 5 and 7 comprise 3 vessels each.

Only the relevant information was extracted based on which an analysis of the fishing fleet can be made in 2021, Cluster 1 from PMP vessels and cluster 3 from PG vessels add up to the largest total catches in the fishing fleet 2 837.219 tons representing 90.72%.

**Table 3.** Activity of fishing vessels by fleet segment in 2021

| | PMP | | | | PG | | |
|---|---|---|---|---|---|---|---|
| | **Vessels** | **Catches (tons)** | **Amount (euro)** | | **Vessels** | **Catches (tons)** | **Amount (euro)** |
| no. | 52 | 2973.056 | 1880725 | no. | 78 | 154.0905 | 355759 |
| % | 40.00 | 95.07 | 84.10 | % | 60.00 | 4.93 | 15.90 |

**Table 4.** Activity of fishing vessels on combining mobile and passive gears (PMP) clusters in 2021

| **CLUSTER 1 - 93.12%** | | | **CLUSTER 2 - 3.08%** | | | **CLUSTER 4 - 2.21%** | | |
|---|---|---|---|---|---|---|---|---|
| **Stock*** | **Catch (t)** | **Vessels** | **Stock*** | **Catch (t)** | **Vessels** | **Stock*** | **Catch (t)** | **Vessels** |
| RPW | 2693.279 | 37 | SPR | 38.865 | 3 | RPW | 31.515 | 3 |
| MSM | 33.352 | 4 | MUT | 32.215 | 3 | MSM | 31.551 | 3 |
| TUR | 29.825 | 21 | TUR | 14.101 | 3 | | | |
| HMM | 5.374 | 9 | HMM | 3.860 | 3 | | | |
| BLU | 5.141 | 9 | BLU | 1.179 | 3 | | | |
| **Total** | | **37** | **TOTAL** | | **3** | **TOTAL** | | **3** |
| DOMI NANT | MIN | MAX | DOMI NANT | MIN | MAX | DOMI NANT | MIN | MAX |
| 15 m | 7 m | 26 m | 15 m | 15 m | 20 m | 8 m | 7 m | 8 m |

*FAO fish species code: RPW - *Rapana venosa;* MSM - *Mytilus galloprovincialis;* TUR - *Scophthalmus maximus;* HMM - *Trachurus mediterraneus;* BLU - *Pomatomus saltatrix;* SPR - *Sprattus sprattus*; MUT - *Mullus barbatus*; JDP - *Dasyatis pastinaca*; SHC - *Alosa pontica*; RJC - *Raja clavata*; MGA - *Liza aurata*; MBF - *Mesogobius batrachocephalus*; GPA – *Gobiidae*; ATB - *Atherina boyeri*; ANE - *Engraulis encrasicolus*; WHG - *Merlangius merlangus.*

**Table 5.** Activity of fishing vessels on passive gears (PG) clusters in 2021

| **CLUSTER 1 - 12.81%** | | | **CLUSTER 3 – 44.00%** | | | **CLUSTER 5 – 5.36%** | | | **CLUSTER 7 – 5.36%** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Stock*** | **Catch (t)** | **Vessels** | **Stock*** | **Catch (t)** | **Vessels** | **Stock*** | **Catch (t)** | **Vessels** | **Stock*** | **Catch (t)** | **Vessels** |
| TUR | 18.256 | 16 | TUR | 1.635 | 6 | RPW | 11.402 | 2 | WHG | 2.940 | 2 |
| JDP | 1.045 | 2 | SPR | 6.957 | 6 | MSM | 39.527 | 3 | HMM | 1.413 | 3 |
| | | | SHC | 1.419 | 12 | | | | | | |
| | | | RPW | 0.848 | 2 | | | | | | |
| | | | RJC | 0.185 | 2 | | | | | | |
| | | | MUT | 1.229 | 9 | | | | | | |
| | | | MGA | 0.880 | 4 | | | | | | |
| | | | MBF | 0.594 | 8 | | | | | | |
| | | | JDP | 0.963 | 6 | | | | | | |
| | | | HMM | 10.946 | 21 | | | | | | |
| | | | GPA | 1.961 | 11 | | | | | | |
| | | | BLU | 0.898 | 2 | | | | | | |
| | | | ATB | 1.256 | 2 | | | | | | |
| | | | ANE | 37.842 | 25 | | | | | | |
| **TOTAL** | | **16** | **TOTAL** | | **25** | **TOTAL** | | **3** | **TOTAL** | | **3** |
| DOMINANT | MIN | MAX | DOMINANT | MIN | MAX | DOMINANT | MIN | MAX | DOMINANT | MIN | MAX |
| 7 m | 5 m | 11 m | 11 m | 5 m | 11 m | 5 m | 4 m | 5 m | 7 m | 6 m | 7 m |

## CONCLUSIONS

In 2021, there was a decrease in both the total quantity landed and the value in euro compared to 2020. The total quantity landed in 2021 amounted to 3.127.146 tons, while the corresponding value in euro was 2.236 484. This represented a decline of 29.93% in the share of landings and a decrease of 19.31% in the value when compared to the previous year.

Rapa whelk is represented in 2021 by (87.82 %) of the total catches in Romania, followed by mussels with (4.00%) and turbot with (2.38 %).

The fishing vessels grouped by fishing techniques combining mobile and passive gears (PMP) used during the activities of both active and passive fishing tools beam trawl (TBB), midwater otter trawl (OTM), gillnet set (GNS), beach seine (SB) and manual to collect Rapa whelk and mussel. Such activities were performed by almost all ship categories 24-40 PMP, 18-24 PMP, 12-18 PMP, and 06-12 PMP.

The fishing vessels with passive gears (PG) fishing techniques were those of the 00-06 PG and 06-12 PG class segments, the main tools used were: set gillnet (GNS), stationary uncovered pound nets (FPN), hand and pole lines (LHP) and set longlines (LLS).

**Acknowledgment.** This study case was part of the workshop organized by Erik Sulanke and Jorg Berkenhagen from the Thünen Institute for Sea Fisheries in Bremerhaven, Germany under the coordination of data collection framework (DCF).

## REFERENCES

Sulanke E. and Berkenhage J. (Eds) (2022*).* Report of the workshop on an alternative approach to the segmentation of fishing fleets, 29-30 March 2021, doi: 10.13140/RG.2.2.15792.23043

Sulanke E. (2022), Fleet segmentation R-package script. Available online at https://github.com/ESulanke/FleetSegmentation/blob/main/vignettes/FleetSegmentation_vignette.Rmd

MADR (Ministerul Agriculturii și Dezvoltării Rurale), 2022. Romania's annual report on efforts to achieve a sustainable balance between fishing capacity and fishing opportunities for 2021. Available online at https://oceans-and-fisheries.ec.europa.eu/system/files/2022-09/2021-fleet-capacity-report-romania_en.pdf

RStudio V3.6.1. Avaible online at https://cran.r-project.org/bin/windows/base/old/3.6.1/