Methodology Report for Socio Economic Variables National Fisheries Data Collection Program

Athens, 2018

AGRICULTURE ECONOMICS RESEARCH INSTITUTE (AGR.E.R.I)

Alkmanos str., 115 28, Ilisia, Athens / T.: +30 210 27 55 086 / E-mail: info@agreri.gr

Scientific coordination: Dr Irene Tzouramani

Research Associates: Dr Alexandra Sintori

Dr Angelos Liontakis

Dr Georgios Valakas

AGR.E.R.I partner: Magdalene Bardounioti

Contents

List of Tables

List of Abbreviations

Abbreviation	Term
CNR	Complete non-response errors
CV	Coefficient of variation
DCF	Data Collection Framework
DFN	Drift and/or fixed netters
DRB	Boat dredges
DRH	Hand dredges
DTS	Demersal trawlers and/or demersal seiners
FPO	Vessels using Pots and/or traps
GNC	Encircling gillnets
GTR	Trammel nets
HOK	Vessels using hooks
LHM	Handlines and pole-lines (mechanised)
LHP	Handlines and pole-lines (hand-operated)
LLD	Drifting longlines
LLS	Set longlines
LTL	Troll lines
MAR	Missing at random
MCAR	Missing completely at random
MGO	Vessels using other active gears
MGP	Vessels using Polyvalent 'active' gears only
MNAR	Missing not at random
PGO	Vessels using other Passive gears
PGP	Vessels using Polyvalent 'passive' gears only
PMP	Vessels using active and passive gears
PNR	Partial non-response errors
PS	Purse seiners
SB	Beach seines
TBB	Beam trawlers
TM	Pelagic trawlers
H.C.M.R.	Hellenic Centre for Marine
HAO-DEMETER	Hellenic Agricultural Organization – Demeter
I.M.B.R.I.W	Institute of Marine Biological Resources & Inland Waters
AGR.E.R.I	Agricultural Economics Research Institute
F.R.I	Fisheries Research Institute

Abstract

This report presents the methodological framework developed for the collection and analysis of socio-economic variables, in the context of the National Fisheries Data Collection Programme for the years 2017-2019. It explains the data frame and the need of conducting a probability sampling survey for data collection. In addition, it presents the target population and the frame population as well as the main sources of information, such as the National Fisheries Register, the sampling survey questionnaire and the fishing logbook. The stratified random sampling technique is used for the collection of socio-economic variables. Also, this report explains the ways used to determine the sample size and how the sample is selected. The fishing gears, the vessel length and the importance of fleet segments as well as the geographical location of the vessels are used to determine the sample size and the sample distribution. In cases where the statistical inference about the population cannot be derived by census data, it is based on an inference to the active population of the fishing vessels. Afterwards, the report describes the methods that ensure the quality of both the raw data and the statistical analysis, such as the unbiased indicator (coverage rate) and the indicators of variability (coefficient of variation and confidence intervals). Finally, it presents information on the availability, cohesion and comparability of data, along with the procedures applied to ensure confidentiality of data.

1. Type of Data Collection

1.1 Introduction

The present report describes the methodology adopted to collect socioeconomic data in the fisheries sector, as part of the **National Programme for the Collection of Fisheries Data 2018 framework**, in accordance with Council Regulation (EC) No 199/2008 (European Commission, 2008) and Commission Implementing Decision (EU) 2016/1251 (European Commission, 2016). Also, the report presents the methods used to ensure the data quality according to the European Statistical System (ESS) (Eurostat, 2014). The report includes the description of data accessibility, coherence and comparability as well as the principles of data confidentiality and security.

Hellenic Agricultural Organization – Demeter (**HAO-DEMETER**) and Hellenic Centre for Marine Research (**H.C.M.R.**) are the scientific partners of the Programme. More specifically, the Agricultural Economics Research Institute (AGR.E.R.I) and the Fisheries Research Institute (F.R.I) of the HAO-DEMETER organization participate in the Programme. AGR.E.R.I is responsible for the collection and evaluation of socioeconomic data in the fisheries sector. F.R.I is responsible for collecting scientific data on fisheries in the North and Central Aegean. The Institute of Marine Biological Resources & Inland Waters of Athens (I.M.B.R.I.W-Athens) and the Institute of Marine Biological Resources & Inland Waters of Crete (I.M.B.R.I.W-Crete) from H.C.M.R. contribute also to the implementation of the National Programme and are responsible for collecting scientific data on fisheries in the South Aegean, the Ionian Sea and the Cretan Sea.

1.2 Data Frame

Even though the ideal survey method is the census, the special characteristics of the Greek fishing fleet such as the large number of vessels and the large geographical dispersion of ports, prevent the implementation of such a survey. Particularly, the Greek fleet consists of 14,934 vessels and the 94.34% of them are smaller than 12 meters, while the fishing fleet is distributed across 223 ports. Therefore, the majority of the economic and social variables of the fleet will be collected by **Probability Sample Survey**, using face to face interviews and structured questionnaires. Moreover, other data sources will be used as it is explained on section 3. The sampling frame presented below takes into account the classification of the national fleet according to vessel length and fishing gear, following the Commission Implementing Decision (EU) 2016/1251 (European Commission, 2016).

The chosen method of sampling is **the stratified random sampling** to take into consideration additional features of the Greek fleet, such as individual fishing gear (type) and geographical distribution. Tables A.1, A2, and A.3 in the Annex present the collection method (census or probability sample survey) and the data sources for the activity, economic and social

variables. Since the Greek management system does not involve quotas or other fishing rights, the following economic variables are not collected.

- *Income from leasing out quota or other fishing rights*
- *Value of quotas or other fishing rights*
- *Lease/rental payments for quota or other fishing rights*

The sample unit is the vessel and it is selected from the Greek vessel registry that corresponds to the 31/12 of the reference year. Moreover, the sample unit and, therefore, the sampling frame are common for the economic and effort variables following the Commission Decision 2010/93/EU (European Commission, 2010 - section A.1.1).

2. Target and Frame Population

The target population is the total number of registered vessels in the Greek fleet. The fishing gears are divided into active, passive and polyvalent gears (active and passive), as Table 2.1 presents. According to DCF (European Commission, 2016), the vessel length (in m) in the Mediterranean Sea is categorized into the following groups: 0-6, 6-12, 12-18, 18-24, 24-40 and the respective group names used in this report are VL0006, VL0612, VL1218 and VL2440. Tables 2.1 shows the total number of fishing vessels of the Greek fleet by basic gear and length category, features that determine the segments of fishing fleet. It is worth mentioning that the values of Table 2.1 may diverge from the respective data of the Annual Report. This may happen for the following reasons:

- Table 2.1 includes values of the national fleet register posted on 13/06/2018, while the corresponding table in the annual report uses the latest updated (31/12/2018) national fleet register.
- The population is adjusted according to the sampling results when a different main gear from fleet register is declared.

There is no clear community decision on the classification of the hand dredgers (DRH) in the categories of main gears. Therefore, DRH fishing gear has been classified as boat dredgers (DRB) due to their common features. Adopting the opinion of experts in this area, DRB or DRH fishing gears were allocated to the polyvalent passive gear (PGP) segments in previous Annual Reports, because DRB or DRH are used for shell fishing.

Table 2.2 and Figure 2.1 show the fleet segments that are merged either for the presentation of results or for the sampling design. The clustering refers to the vessel length variable per main gear. In order to categorize the fishing vessels, the clustering is necessary for both sampling and confidentiality purposes[1] following the guidelines of STECF (2009). In most cases, the clustering does not affect the aggregated segment, due to the small number of

[1] See COMMISSION DECISION of 18 December 2009 (adopting a multiannual Community Programme for the collection, management and use of data in the fisheries sector for the period 2011-2013).

clustered vessels per segment. Table 2.3 presents the total number of vessels per fishing gear after the clustering of segments. The number of inactive vessels is not taken into account a-priori as no information is available. Thus, the target population is the same with the frame population, as already mentioned.

		Length Category				
	Gear Classes	**VL0006**	**VL0612**	**VL1218**	**VL1824**	**VL2440**
Active Gears	Beam trawlers (**TBB**)	0	0	0	0	0
	Demersal trawlers and/or demersal seiners (**DTS**)	1	109	28	98	146
	Pelagic trawlers (**TM**)	0	0	0	0	0
	Purse seiners (**PS**)	0	2	84	136	26
	Dredgers (**DRB**/**DRH**)	11	35	0	0	0
	Vessel using other active gears (**MGO**)	0	0	0	0	0
	Vessels using Polyvalent 'active' gears only (**MGP**)	0	0	0	0	0
Passive Gears	Vessels using hooks (**HOK**)	1551	2117	181	8	0
	Drift and/or fixed netters (**DFN**)	3719	6155	130	2	0
	Vessels using Pots and/or traps (**FPO**)	68	321	5	1	0
	Vessels using other Passive gears (**PGO**)	0	0	0	0	0
	Vessels using Polyvalent 'passive' gears only (**PGP**)	0	0	0	0	0
Vessels using active and passive gears (PMP)		0	0	0	0	0

Table 2.1: The segmentation of the Greek fleet according to the European Union Data Collection Framework (DCF) and the national fleet register data (using the main gear class) posted on 13/06/2018.

	Merged Class		**Clustered Class**		
Gear Class	**Length Category**	**Number of Vessels**	**Length Category**	**Number of Vessels**	**Total vessel number after clustering**
DFN	VL1824	2	VL1218	130	132
DRB	VL0006	11	VL0612	35	46
DTS	VL0006	1	VL0612	109	110
FPO	VL1218 VL1824	5 1	VL0612	321	327
HOK	VL1824	8	VL1218	181	189
PS	VL0612	2	VL1218	84	86

Table 2.2: The clustering scheme of Greek fleet segments*

*The first column indicates the main fishing gear for each cluster. The second column indicates the vessel length merged into the segment of fourth column. The third column indicates the number of vessels merged, the fifth column includes the number of vessels per segment before the clustering and the last column presents the total number of vessels after the clustering per segment.

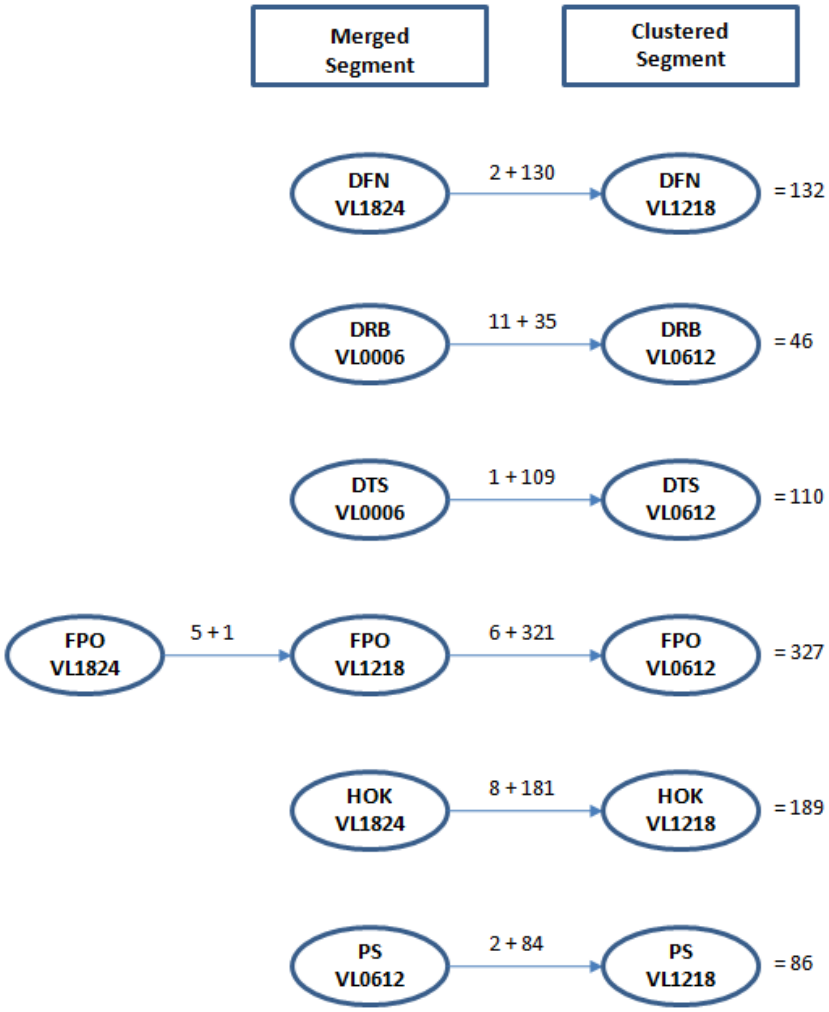

Figure 2.1: The process of clustering and the corresponding number of vessels per cluster.

Gear Class	VL0006	VL0612	VL1218	VL1824	VL2440	Total
DFN	3719	6115	132	0	0	**10006**
DRB	0	46	0	0	0	**46**
DTS	0	110	28	98	146	**382**
FPO	68	327	0	0	0	**395**
HOK	1551	2117	181	8	0	**3857**
PS	0	0	86	136	26	**248**

Table 2.3: The total number of vessels of Greek fleet after the clustering of segments.

3. Data Sources

3.1 National fleet register

The Greek National Fleet Register (NFR) is held by the Hellenic Ministry of Mercantile Marine and Island Policy and it is the primary source of information for the categorization of fleet

segments per gear class and vessel length. The NFR is frequently updated at regular intervals and is available at http://ec.europa.eu/fisheries/fleet/index.cfm. The NFR includes information on the characteristics and the activity information of each vessel, such as the port of registry, the vessel length, the year of construction, the capacity, the horsepower as well as the main and secondary fishing gear.

3.2 Questionnaires

For the sample probability survey of the non-available variables, a structured questionnaire drawn up by AGR.E.R.I has been used from the beginning of the programme and is continually updating and improving to better serve the purposes of the survey. The questionnaire is complemented by face-to-face interviews with fishermen, while correspondents are fish scientists, who receive the appropriate training on a yearly basis for the correct completion of the questionnaires. Moreover, correspondents receive written instructions and they are in contact with AGR.E.R.I research team for questions and clarifications. The completion of the questionnaires for data collection in 2018 will take place in the first half of 2019. H.C.M.R and F.R.I are the responsible institutes for the collection of questionnaires.

3.3 Other Sources

Data of electronic report system (ERS) from vessels and satellite-based Vessel Monitoring System (VMS) are additional sources of information. The Ministry of Rural Development and Food has an internal agreement with HAO-Demeter and H.C.M.R. to provide OSPA data exclusively for the proper implementation of the Multi-annual Programme. This data refer to the logbook of the fishing vessels with length more than ten meters, the beach seiners (SB) with length less than 10 meters and the vessels which are larger than six meters and utilize the following fishing gears: handlines and pole-lines (mechanised) (LHM) and drifting longlines (LLD). Thus, it is possible to obtain activity variables, such as: *days at sea*, *landing weight* and *value of landings*. The transmission of the above information does not include the vessel identity for confidentiality reasons.

Finally, it should be noticed that the Ministry of Rural Development and Food provide HAO-Demeter with the list of vessels that that utilize fishing gears that need a special fishing permit, i.e. SB, LHM LLD. This information is also used to categorize the Greek fleet by basic tool.

4. Sampling Frame

4.1 Sampling Strategy

A sampling scheme of stratified random sampling without replacement is chosen for the probability sample survey. The sample unit is the vessel and it is selected randomly from each segment as it is described below. The stratified random sampling of Greek fleet is considered the most proper sampling technique due to the heterogeneous feature of segments. It is noted that the stratified random sampling was also applied in the previous years of the National Fisheries Data Collection Program implementation.

4.2 Stratification

The stratification of Greek fleet is done by the variables: gear class and vessel length. Additional stratums (gear type level 4), are considered for vessels using drift and/or fixed netters (DFN) : GNS, GTN and GTR). Also, two stratums are considered for the gear class Hook: Hook A includes the gear types (level 4): LHP, LLS and LHM. Hook B includes the gear types (level 4): LHM and LLD. Table 4.1 shows the gear types (1st column) that have been merged to the strata GTR, HOK_A και HOK _B (4th column). The gear type GNC has been categorized into GTR strata due to statistical and confidential reasons.

Gear Type	Length Category	Number of Vessels	Sampling Stratum	Length Category
GNC	VL0006	3	**GTR**	VL0006
GNC	VL0612	1	**GTR**	VL0612
GTR	VL1824	2	**GTR**	VL1218
LHP	VL0006	117	**HOK_A**	VL0006
LHP	VL0612	71	**HOK_A**	VL0612
LHP	VL1218	3	**HOK_A**	VL1218
LLS	VL0006	1418	**HOK_A**	VL0006
LLS	VL0612	1890	**HOK_A**	VL0612
LLS	VL1218	45	**HOK_A**	VL1218
LTL	VL0006	15	**HOK_A**	VL0006
LTL	VL0612	19	**HOK_A**	VL0612
LHM	VL0006	1	**HOK_B**	VL0612
LHM	VL0612	47	**HOK_B**	VL0612
LHM	VL1218	43	**HOK_B**	VL1218
LLD	VL0612	90	**HOK_B**	VL0612
LLD	VL1218	90	**HOK_B**	VL1218
LLD	VL1824	8	**HOK_B**	VL1218

Table 4.1: List of gear types that have been merged to the strata GTR, HOK_A and HOK _B.

Furthermore, the geographical distribution of vessels is taken into consideration for the construction of strata. As it is explained below, the features *gear class* and *vessel length* are accounted for the determination of sample size, while the registry port for the distribution of sample over geographical areas. Following the sampling scheme of the biological data collection, the Hellenic coastline and marine area is divided in 12 major areas and stratified sampling will be carried out using these areas as -geographic- strata. These areas are the following: Argosaronikos (ARGSAR), Chios - Mytilene (CH-MIT), Central Ionion (C-ION), Crete (CRETE), Cyclades (CYCL), Dodecanese (DODEC), Evia (EVIA), North Ionion (N-ION), South Ionion (S-ION), Thermaikos Gulf (THERM), Thracian Sea-Limnos (THR-LIM) and Volos-Sporades (VOL-SPOR). F.R.I is responsible for the coordination and collection of questionnaires in the regions: Chios – Mytilene, Thermaikos Gulf, Thracian Sea-Limnos and Volos-Sporades. H.C.M.R. is responsible for the remaining areas. The specific fishing gear, the population size for each area and the total population for each stratum are presented in Table 4.2.

4.3 Sample Size

The sample size is determined according to the specific gears and the length category as it is presented in the first column of Table 4.2. The registry port of vessel is accounted for the distribution of sample over geographical areas. The variable *days at sea* was selected among the activity variables in order to determine the sample size in each segment and the margin of error affecting the sample size was determined by the importance of each individual fishing gear (see Table 4.3). The level of statistical significance for all segments is set at 10% (z = 1.64). The sample size of each segment is estimated using the following equation (Dattalo, 2008; Moura, 2016):

$$\mathrm{n} = \frac{\mathrm{n}_0 \cdot \mathrm{N}}{\mathrm{n}_0 + (\mathrm{N} - 1)} \tag{4.1}$$

where N is the population size of each segment and

$$\mathrm{n}_0 = \frac{z^2 \cdot s^2}{\mathrm{e}^2 \cdot \bar{\mathrm{x}}} \tag{4.2}$$

where s is the standard deviation και $\bar{\mathrm{x}}$ is the average of auxiliary variable (*days at sea*), which were estimated by the data of year 2017 (see Annex Tables A.4 and A.5).

The sample size of each segment, as it was calculated by equation 4.2, is adjusted according to equation 4.3 when the population size of segment is very small and the sample size is relatively large (n/N>0.05) (see e.g. Thomson, 2002).

$$\mathrm{n}_{\mathrm{adj}} = \frac{\mathrm{n}}{1 + \mathrm{n}/\mathrm{N}} \tag{4.3}$$

	ARGSAR	CHI-MIT	C-ION	CRETE	CYCL	DODEC	EVIA	N-ION	S-ION	THERM	THR-LIM	VOL-SPOR	Total
DRB_0612	27	2	1	0	1	0	0	1	0	4	9	1	**46**
DTS_0612	20	9	9	1	13	19	7	21	7	3	1	0	**110**
DTS_1218	5	0	0	3	9	3	1	0	0	0	3	0	**24**
DTS_1824	16	7	9	1	6	6	8	6	2	11	26	4	**102**
DTS_2440	29	12	5	7	4	4	13	2	2	39	25	4	**146**
FPO_0006	5	8	3	0	0	4	4	5	0	17	21	1	**68**
FPO_0612	11	12	2	0	0	16	13	7	0	166	93	7	**327**
GNS_0006	85	260	83	41	21	49	60	50	28	74	112	70	**933**
GNS_0612	178	241	143	46	23	40	52	123	31	149	133	78	**1237**
GNS_1218	10	0	3	2	3	1	0	1	1	2	2	0	**25**
GTN_0006	18	20	28	8	54	27	8	12	12	10	7	5	**209**
GTN_0612	72	23	50	8	100	30	25	55	33	20	38	20	**474**
GTN_1218	0	0	0	0	7	0	1	0	0	0	2	0	**10**
GTR_0006	231	258	318	118	85	155	244	371	150	262	206	179	**2577**
GTR_0612	583	303	471	276	255	322	388	626	247	402	403	168	**4444**
GTR_1218	13	9	5	7	27	15	5	2	1	1	10	2	**97**
HOK_A_0006	230	175	138	152	63	129	153	114	98	121	90	87	**1550**
HOK_A_0612	359	154	185	185	120	240	206	150	128	117	58	78	**1980**
HOK_A_1218	9	2	1	3	8	7	6	0	5	0	5	2	**48**
HOK_B_0612	31	8	17	15	4	20	3	13	0	6	9	12	**138**
HOK_B_1218	12	5	11	18	13	37	4	2	4	6	17	12	**141**
PS_1218	21	5	12	3	5	4	19	4	0	4	7	2	**86**
PS_1824	27	4	12	4	4	1	24	8	1	13	24	14	**136**
PS_2440	1	0	0	1	1	0	6	0	0	12	5	0	**26**
Total	**1993**	**1517**	**1506**	**899**	**826**	**1129**	**1250**	**1573**	**750**	**1439**	**1306**	**746**	**14934**

Table 4.2: The population of fishing vessels per stratum.

Sampling Strata	Margin of Error
DTS_1824, DTS_2440, GNS_0612, GTR_0612, HOK_A_1218 PS_1218, PS_1824, PS_2440	5%
FPO_0006, FPO_0612, GNS_0006, GNS_1218, GTR_0006, GTR_1218, HOK_A_0006, HOK_A_0612, LLS_0006, LLS_0612	15%
DTS_0612, DTS_1218, GTN_0006, GTN_0612, GTN_1218, DRB_0612	20%

Table 4.3: The fleet strata with the respective margin of errors used to determine the sample size.

After the determination of the sample size of fleet segments, the sample size per geographical stratum is determined using the method of proportional allocation (Eurostat, 2008):

$$n_g = \frac{n \cdot N_g}{N} \tag{4.4}$$

where n is the sample size per fleet segment, as it was calculated by equation 4.3, N_g is the sum of vessels of geographical area per fleet segment and N is the population size for each segment. Using the proportional allocation, the proportions are respected and, consequently, the share of a geographic stratum population to the total population will be similar to the share of this stratum to the sample size. Decimal values of sample size were rounded up to the nearest integer. The sample size per stratum is presented in Table 4.4.

In conclusion, the determination of sample size and the sample distribution was based on the specific fishing gears and their importance by taking into account the most representative auxiliary variable and the 12 major geographic areas. The total sample size for the reference year 2018 is set to 861 vessels, divided into the main fishing gear and the length category, as shown in Table 4.5. Table 4.6 presents the sample ratio to the population per main gear and vessel length.

	ARGSAR	CHI-MIT	C-ION	CRETE	CYCL	DODEC	EVIA	N-ION	S-ION	THERM	THR-LIM	VOL-SPOR	Total
DRB_0612	1	1	1	0	1	0	0	1	0	1	1	1	**8**
DTS_0612	3	2	2	1	2	3	1	3	1	1	1	0	**20**
DTS_1218	2	0	0	1	3	1	1	0	0	0	1	0	**9**
DTS_1824	7	3	4	1	3	3	4	3	1	5	10	2	**46**
DTS_2440	8	3	2	2	1	1	4	1	1	10	7	1	**41**
FPO_0006	2	2	1	0	0	1	1	2	0	4	5	1	**19**
FPO_0612	1	1	1	0	0	1	1	1	0	8	5	1	**20**
GNS_0006	2	5	2	1	1	1	2	1	1	2	2	2	**22**
GNS_0612	11	14	9	3	2	3	4	8	2	9	8	5	**78**
GNS_1218	5	0	2	1	2	1	0	1	1	1	1	0	**15**
GTN_0006	1	1	2	1	3	2	1	1	1	1	1	1	**16**
GTN_0612	1	1	1	1	2	1	1	1	1	1	1	1	**13**
GTN_1218	0	0	0	0	3	0	1	0	0	0	1	0	**5**
GTR_0006	2	2	3	1	1	2	2	3	1	2	2	2	**23**
GTR_0612	11	6	9	5	5	6	7	12	5	8	8	3	**85**
GTR_1218	4	3	2	2	8	5	2	1	1	1	3	1	**33**
HOK_A_0006	4	3	3	3	2	3	3	2	2	3	2	2	**32**
HOK_A_0612	23	10	12	12	8	15	13	10	8	8	4	5	**128**
HOK_A_1218	5	1	1	2	4	4	3	0	3	0	3	1	**27**
HOK_B_0612	10	3	6	5	2	7	1	5	0	2	3	4	**48**
HOK_B_1218	5	2	5	8	6	15	2	1	2	3	7	5	**61**
PS_1218	9	3	6	2	3	2	9	2	0	2	3	1	**42**
PS_1824	10	2	5	2	2	1	9	3	1	5	9	6	**55**
PS_2440	1	0	0	1	1	0	3	0	0	6	3	0	**15**
Total	**128**	**68**	**79**	**55**	**65**	**78**	**75**	**62**	**32**	**83**	**91**	**45**	**861**

Table 4.4: Sample size of fishing vessel per stratum.

	Length category				
Fishing gear	**VL0006**	**VL0612**	**VL1218**	**VL1824**	**VL2440**
DFN	61	176	53	0	0
DRB	0	8	0	0	0
DTS	0	20	9	45	41
FPO	19	20	0	0	0
HOK	32	176	88	0	0
PS	0	0	42	55	15

Table 4.5: The sample size of fleet segments. The sample size is zero when there are not vessels on the segments of population after the clustering.

	Length category				
Fishing gear	**VL0006**	**VL0612**	**VL1218**	**VL1824**	**VL2440**
DFN	1,64%	2,86%	40,15%	-	-
DRB	-	17,39%	-	-	-
DTS	-	18,18%	37,50%	45,10%	28,08%
FPO	27,94%	6,12%	-	-	-
HOK	2,06%	8,31%	46,56%	-	-
PS	-	-	48,84%	40,44%	57,69%

Table 4.6: The coverage rate per main gear and vessel length.

4.4 Sample Selection

The sample selection, as already mentioned, is random, producing one random number for each fishing vessel of the EAM list, i.e. 14934 random numbers are produced. The random numbers are produced by computerized routine and get values from 0 to 1. For each stratum of sampling, vessels are classified from the lowest to the highest random number. Then, the number of vessels is selected according to the sample size in each stratum. The vessels of each stratum are selected according to the smallest random value. In this way the selection of sample is achieved without replacement and each vessel has an equal probability to be selected.

A complementary sample is selected according to the above classification, in order to address the non-response of the sampling unit. If the complementary sample is exhausted in the defined geographical areas, sampling shall be carried out in adjacent areas and in vessels belonging to the same fishing gear.

4.5 Sample Evolution

The main differences of proposal sample size during the years occur due to the changes of the segments' significance. Table 4.7 shows the coverage rate of proposed sample in relation to the

population from 2012 to 2018 for each fleet segment. Over the years there are small differences in coverage rates between fleet segments. These small differences are due to the population size and not due to the sample size. Also, in some cases, as in the DTS segment, even greater significance is attributed to larger vessels. For example, DTS segment with length 18-24 m has 45.10% coverage rate compared to 28.57% of the year 2017. Moreover, the previous schemes classified some vessels to PGP segment, while the present does not. The main reason of this decision is that vessels classified to PGP segment have high variability of the second gear type, while using a main gear type systematically.

Gear Segment	2012	2013	2014	2015	2016	2017	2018
DFN_VL0006	-	-	5.04%	5.02%	2.15%	2.11%	1.64%
DFN_VL0612	-	-	5.43%	5.01%	1.86%	1.86%	2.86%
DFN_VL1218	-	-	30.40%	30.30%	28.65%	28.78	40.15%
DRB_VL0612	-	-	-	-	-	-	17.39%
DTS_VL0612	31.75%	30.14%	-	-	-	28.43%	18.18%
DTS_VL1218	-	52.29%	-	-	-	51.43%	37.50%
DTS_VL1824	33.78%	34.17%	30.78%	30.77%	30.00%	28.57%	45.10%
DTS_VL2440	10.87%	21.56%	21.69%	23.03%	20.00%	20.00%	28.08%
FPO_VL0006	-	-	31.82%	30.07%	40.00%	38.36%	27.94%
FPO_VL0612	-	-	30.14%	30.23%	16.72%	16.47%	6.12%
FPO_VL1218	-	-	50.00%	53.84%	-	-	-
HOK_VL0006	-	-	15.35%	15.04%	6.28%	6.15%	2.06%
HOK_VL0612	-	-	5.16%	5.01%	9.14%	8.90%	8.31%
HOK_VL1218	35.07%	32.08%	33.05%	30.39%	43.56%	43.44%	46.56%
PGP_ V0006	8.43%	8.30%	-	-	37.50%	33.33%	-
PGP_ V0612	6.28%	6.00%	-	-	38.46%	35.71%	-
PGP_ V1218	34.93%	32.54%	-	-	-	-	-
PS_VL1218	36.60%	35.63%	30.69%	31.03%	36.14%	36.14%	48.84%
PS_VL1824	33.93%	33.58%	30.36%	30.60%	22.22%	22.22%	40.44%
PS_VL2440	50.32%	51.61%	51.61%	51.61%	50.00%	50.00%	57.69%

Table 4.7: Evolution of proposed sample coverage rate during the years.

5. Estimation Procedure

Statistical inference of population for the economic and social variables is based on the inference of the sample to the active population of fishing vessels, where applicable. No estimation is made for the variables that result from census. The Horvitz-Thompson (HT) estimator (Horvitz and Thompson, 1952) is used to allow correct generalization of the sample statistics to the population parameters per fleet segment:

$$\hat{Y} = \sum_{i=1}^{n} y_i \, \pi_i^{-1} \qquad (5.1)$$

where $\pi_i = \mathrm{n_{act}}/\mathrm{N_{act}}$ is the inclusion probability for each stratum, $\mathrm{n_{act}}$ the number of active vessels in sample per stratum and $\mathrm{N_{act}}$ is the active vessel in population per stratum. At a pilot stage, the population features will be estimated using linear regression analysis, to explore possible estimation models that can be used in the future.

Using the above formula the number of inactive vessels of the population per stratum is also estimated, before the estimation of other variables. The estimation is done for the stratums less than 12 meters in length, except for the stratums using the gears SB, LLD and LHM. The number of inactive vessels of the rest of stratum is determined by census data using information ofthe fishing logbook. According to Commission Decision 2010/93/EU (European Commission, 2010), inactive are the vessels that have not been engaged in fishing operations during the reference year.

The economic variables: *Consumption of fixed capital* and *Value of physical capital* will be estimated from both the data of the probability sample survey (replacement value) and the NFR (mean LOA and number of vessels per fleet segment), as it is suggested by Perpetual Inventory Methodology (PIM) (European Commission, 2006). More specifically, the "degressive" depreciation function is used for the variable *Consumption of fixed capital* and the capital values are determined assuming that the engine is renovated every 10 years, electronics and other equipment every 5 and 7 years respectively, while the hull is never renovated. The share of each asset item in the total vessel price is:

- Hull – 60%
- Engine – 20%
- Electronics – 10%
- Other equipments – 10%

For the variable *Value of the physical capital*, the unit price is determined by direct survey. The selling prices of second-hand vessels and their insurance costs are also used to evaluate the results of survey.

6. Data Quality Evaluation

6.1 Methodology Relevance

The methodology adopted and described above is controlled for its proper implementation at all stages. In particular, during the sampling period, AGR.E.R.I research team communicates with the correspondents at regular intervals in order to ascertain the proper process of collecting the questionnaires. When the questionnaires are collected, the material is evaluated, for example the number of questionnaires collected per correspondent and the completeness of the data is checked.

The coverage rate of the proposed sample (Table 4.6) is compared with the response rate of sample units per fleet segment. Small discrepancies are allowed in the assessment of rates for reasons such as misclassification of the fishing fleet (see section 6.2) or non-updating of the NFR data during the sampling design. Also, the percentage of inactive vessels in the sample is compared with the percentage of inactive vessels resulting from the sources of census in order to assess the relevance of the sampling methodology.

Moreover, there is an evaluation of the sources of information that contributed to the categorization of the Greek fleet, such as NFR, in which each vessel is assigned a main fishing gear and a secondary fishing and the fishing licenses of 2018. The evaluation is carried out by calculating the percentage of fishing gears declared in the sampling process and differentiated from the initial classification of the fleet.

6.2 Results/Output Completeness

The annual report meets the requirements of Council Regulation (EC) No 199/2008 (European Commission, 2008) and Commission Implementing Decision (EU) 2016/1251 (European Commission, 2016). More specifically, the annual report includes all socio-economic variables and activity variables for all sections of the Greek fleet. As already mentioned, the following financial variables are not collected since the Greek management system does not include commercial quotas or other fishing rights:

- *Income from leasing out quota or other fishing rights*
- *Value of quotas or other fishing rights*
- *Lease/rental payments for quota or other fishing rights*

Data collection is compiled in accordance with the principles of impartiality, reliability and objectivity using only official sources of information. The probability sampling survey that was carried out as well as the statistical inference of the population units are consistent at a level of Member States, regional and EU. The adopted methodology for collecting fisheries data follows international standards and best practices.

Before the statistical inference, the data collected either by census or by probability sampling survey are evaluated using a process of error detection as it is described in the next sections. The presentation of the results for all socio-economic variables and activity variables is detailed for all segments of the fishing fleet and there are no cases of clustering segments for confidentiality reasons.

6.3 Accuracy sampling Errors

6.3.1 Sampling Errors

The desired accuracy for each segment is determined according to the importance of the segments (Table 4.3). As already described in Section 4, this process requires, in general, prior knowledge of the population variance for each variable. Therefore, an auxiliary variable was chosen from the activity variables of the previous Annual Report, whose variance better reflects the variability of all variables. The variable "*days at sea*" is selected as auxiliary variable and the level of statistical significance for all segments is set at 10% (z = 1.64).

The coefficient of variation is selected as measure of sampling error for each variable and it is given by equation:

$$\mathrm{CV} = \frac{\mathrm{SD}}{\bar{\mathrm{X}}} \tag{6.1}$$

where SD is the standard deviation and $\bar{\mathrm{X}}$ is the average of the variable. Also, the 100(1-α) % confidence interval of mean is estimated for each variable according to the equation (e.g. Särndal et al. 1992):

$$\bar{\mathrm{X}} \pm \mathrm{t}_{\mathrm{n}-1,1-\alpha/2} \frac{\mathrm{S}^*}{\sqrt{\mathrm{n}}} \tag{6.2}$$

where n is the sample size of a fleet segment $\mathrm{t}^2_{\mathrm{n}-1,1-\alpha/2}$ the percentile of t-student distribution with n degree of freedom and S^* the unbiased estimator of standard deviation of sample, which is determined by the following equation:

$$\mathrm{S}^* = \sqrt{\frac{\sum_{\mathrm{i}}^{\mathrm{n}} (\mathrm{X}_{\mathrm{i}} - \bar{\mathrm{X}})^2}{\mathrm{n}-1}} \tag{6.3}$$

6.3.2 Coverage Errors

The selection of sample units (vessels) is random and it is determined before the conduction of the probability sample survey, in order to avoid convenience sample. Also, there are no units of the target population that are excluded from the sample selection procedure. Therefore, the sample is considered to be representative and unbiased to the target population. As a consequence, the coverage error is zero. It is noted that Table 4.6 shows the coverage rate which is the percentage of sample-to-population for each segment of the fishing fleet rather than the percentage of units that cannot be included in the sample.

6.3.3 Measurements Errors

Before the data analysis, data are controlled for their completeness, cohesiveness and comparability over time. The data that are collected, either by census or by the probability sampling survey, are controlled by using exploratory data analysis, in order to locate measurements errors and processing errors. Measurements and processing errors are located by controlling the unreasonable values and the extreme values of the data. An unreasonable value is a value that has no natural meaning of interpretation of the variable (e.g. a negative value of a variable that can only take positive values). An extreme value is a value that is considerably remote, compared to the majority of the rest of the variable values. In particular, extreme values are located outside the following limit for each variable (Tukey, 1977):

$$(Q_1 - 1.5\ IQR\ , Q_3 + 1.5\ IQR\ ) \qquad (6.3)$$

where Q1 and Q3 are the first and third quartile of the variable respectively and IQR is the interquartile range. When unreasonable values or extreme values refer to data that have been collected by census, the source of the information is controlled and data are corrected in case of mismatch. In case of an accord, the unreasonable values or the extreme values are replaced with missing values. The same applies for data that have come from the probability sampling survey, communicating with the correspondents in order to clarify the validity of these values. It should be noted that very extreme values are defined as the values that are outside the following limit (Tukey, 1977):

$$(Q_1 - 3\ IQR\ , Q_3 + 3\ IQR\ ) \qquad (6.4)$$

Since equation 6.4 may yield minimum or maximum values that do not have a physical interpretation for the variable (e.g. the variable days at sea cannot be higher than the value of 365 days), the boundaries are adjusted using values from historical data, literature review (e.g. Pinello et al., 2017) and experts' opinion. In order to find possible omissions in the questionnaire's design or in the correspondent's training, the unreasonable values and the extreme values are segmented into:

- incorrect values due to the questionnaire's design,
- incorrect answers caused by the correspondent/interviewer and
- incorrect answers of the interviewee

The incorrect values due to questionnaire's design include the values that are repeated in different segments of the fishing fleet and are located in data coming from different correspondents. The incorrect answers caused by the correspondent consist of the values that are repeated in the data coming from the same correspondent and are related either to guided responses during the interview, incorrect coding or data entry. The incorrect answers of the interviewee contain the values that cannot be classified into the other two categories. It should be noted that the questionnaire does not include question in free text format.

6.3.4 Non-response Errors

In order to address the non-response of the sampling unit (vessel), a complementary sample is selected as it was done in previous years. The complementary sample, as well as the main sample, is taken randomly. The selection order of the complementary sample is determined in advance to avoid the collection of convenience sample. This action results in the complete non-response rate (CNR) being zero. Also, the partial non-response (PNR) rate of the collected variables is small due to the appropriate training of correspondents and the ability to resolve any questions in close cooperation with the AGR.E.R.I research team. During the interview, the correspondents pay special attention to the variables which had high non-response rates in previous years. Variables with high non-response rates have been taken into account before the designing of the questionnaire. Finally, the percentage of PNR is calculated for each variable.

The PNR cases and the missing values which replace the unreasonable/extreme values resulting from the exploratory data analysis, are classified into the following categories: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). A missing value is classified in MCAR if the probability of missing is independent from the variable or from other variables. A missing value is classified in MAR if the missing is random in the variable but it correlates with other variables, for example the education level is often missing in vessels with high number of employers. Finally, as MNAR are classified the missing values that are correlated with the variable. For example, the values of annual revenue are missing for vessels with high revenue. The percentage of the three missing values categories is calculated for each variable per fleet segment.

Only the missing values classified in MCAR and MAR are replaced. The missing values of continuous variables are replaced by the median and the missing values of nominal variables by the respective mode value of the variable in stratum.

7 Accessibility and Clarity

The present methodological framework applies to the data collection of 2019 (reference year: 2018) and the results will be presented in the Annual Report of 2020. In the Annual report, the results of merged segments (Table 2.2) are not presented for confidentiality and statistical reasons.

Data and metadata are presented and archived in a form that facilitates correct interpretation and comparisons. The end-users are informed about the methodology of statistical processes through the present methodological framework and through the Work Plan. They are also informed about the quality of the statistical results, following the standards of European Statistical System. Customized data analysis beyond the analysis presented in the Annual report is provided upon request. The methodological framework and the annual report

are available on the European Union website. The data are stored in databases and are not available for editing online.

8 Coherence and comparability

The data collection is follows the European Regulations and therefore are comparable to those of the other European countries, as the same concepts, definitions and classifications are used. All values of the variables either collected by census or by the probability sampling survey have the same target population and the same reference year. Long-term comparability is possible for most segments of the fishing fleet according to the section 4.5.

9 Confidentiality, Transparency and Security

The dissemination of statistics deriving from the socio-economic variables is carried out by AGR.E.R.I following the statistical principles of the European Statistics Code of Practice, in particular the principle of statistical confidentiality. AGR.E.R.I shall take all appropriate precautions to ensure that individual statistical units (fishing vessels) cannot be identified by technical or other means reasonably practicable by third parties.

The confidential information transmitted by the departments of the Ministry of Rural Development and Food and the Ministry of Shipping and Island Policy to AGR.E.R.I are used exclusively for statistical purposes and only the authorized members of the AGR.E.R.I research team has the exclusive right of access to that information. In addition, the members of the AGR.E.R.I research team , in any employment relationship are bounded by confidentiality and have the obligation to use the data accessed exclusively for statistical purposes. Any other use of such data is prohibited beyond the end of their duties.

References

European Commission. 2006. Evaluation of the capital value, investments and capital costs in the fisheries sector. Study No. FISH/2005/03 (available at www.framian.nl/includes/download.asp?file_id=173).

European Commission, 2008. Council Regulation (EC) No. 199/2008 concerning the establishment of a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the Common Fisheries Policy. *Official Journal of the European Union*, L60, 5 March 2008.

European Commission, 2010. Commission Decision (2010/93/EC): Adopting a multiannual Community programme for the collection, management and use of data in the fisheries sector for the period 2011−2013. *Official Journal of the European Union*, L41/8, 18 December 2009.

European Commission, 2016. Commission Implementing Decision (EU) 2016/1251: Adopting a multiannual Union programme for the collection, management and use of data in the fisheries and aquaculture sectors for the period 2017-2019. *Official Journal of the European Union*, L 207/113, 12 July 2016.

Eurostat, 2008. Survey sampling reference guidelines 2008 edition: Introduction to sample design and estimation techniques. Luxembourg: Office for Official Publications of EC.

Eurostat, 2014**.** ESS handbook for quality reports. General and regional statistics Collection: Manuals and guidelines. Office of the European Union.

Horvitz, D. & Thompson, D., 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association,* 47: 663‑685. EC studyNo. FISH/2005/03

Moura, C., 2016. *Quality guidelines for the DCF. Guidelines for socio-economic data. Ad hoc contract on quality.*

Pinello, D., Gee, J. & Dimech, M., 2017. Handbook for fisheries socio-economic sample survey: principles and practice. FAO Fisheries and Aquaculture Technical paper No 163, Rome, FAO.Dattalo, P. , 2008. Determining sample size. Oxford: Oxford University Press.

Särndal, C. E., Swensson, B. and Wretman, J., 1992. Model assisted survey sampling. New York: Springer-Verlag

STECF (Scientific, Technical and Economic Committee for Fisheries), 2009. Report of the Sub-group on Research Needs (SGECA/SGRN 09-02). Edited by Paul Connolly & Tiit Raid. European Commission, 22-27 JUNE 2009, GALWAY.

Thompson, S., 2002. Sampling. Wiley New York.

Tukey, J.W., 1977. Exploratory Data Analysis. Addison Wesley, Reading, MA

Annex

Fishing activity variables		Data Source	Data Collection
Capacity	GT	Fleet register	A
	kW	Fleet register	A
	Vessel Age	Fleet register	A
Effort	Days at sea	Fleet register / Questionnaire	A/B
	Fishing days	Fleet register / Questionnaire	A/B
	kW * Fishing Days	Fleet register / Questionnaire	A/B
	GT * Fishing days	Fleet register / Questionnaire	A/B
	Number of trips	Fleet register / Questionnaire	A/B
	Number of fishing operations	Fleet register / Questionnaire	A/B
	Number of nets/Length	Fleet register / Questionnaire	A/B
	Number of hooks, Number of lines	Fleet register / Questionnaire	A/B
	Numbers of pots, traps	Fleet register / Questionnaire	A/B
Landings	Value of landings total and per commercial species	Fleet register / Questionnaire	A/B
	Live Weight of landings total and per species	Fleet register / Questionnaire	A/B
	Prices by commercial species	Fleet register / Questionnaire	A/B

A: Census, B: Probability Sample Survey

Table A.1: The activity variables, the data sources and the data collection method.

	Economic variables	Data Source	Data Collection
Income	Gross value of landings	Questionnaire	B
	Income from leasing out quota or other fishing rights	N/A	N/A
	Other income	Questionnaire	B
Labour costs	Personnel costs	Questionnaire	B
	Value of unpaid labour	Questionnaire	B
Energy costs	Energy costs	Questionnaire	B
Repair & maintenance costs	Repair and maintenance costs	Questionnaire	B
Other operating costs	Variable costs	Questionnaire	B
	Non-variable costs	Questionnaire	B
	Lease/rental payments for quota or other fishing rights	N/A	N/A
Subsidies	Operating subsidies	Questionnaire	B
	Subsidies on investments	Questionnaire	B
Capital costs	Consumption of fixed capital	Fleet Register/ Questionnaire	A/B
Capital value	Value of physical capital	Fleet Register/ Questionnaire	A/B
	Value of quota and other fishing rights	N/A	N/A
Investments	Investments in tangible assets, net	Questionnaire	B
Financial position	Long/short Debt	Questionnaire	B
	Total assets	Questionnaire	B
Employment	Engaged crew	Questionnaire	B
	Unpaid labour	Questionnaire	B
	Total hours worked per year	Questionnaire	B

A: Census, B: Probability Sample Survey, N/A: Not Applicable.

Table A.2: The economic variables, the data sources and the data collection method.

Social variables	Data Source	Data Collection
Employment by gender	Questionnaire	B
FTE by gender	Questionnaire	B
Unpaid labour by gender	Questionnaire	B
Employment by age	Questionnaire	B
Employment by education level	Questionnaire	B
Employment by nationality	Questionnaire	B
Employment by employment status	Questionnaire	B
FTE National	Questionnaire	B

B: Probability Sample Survey

Table A.3: The social variables, the data sources and the data collection method