Methodology Report for Fleet Socio Economic Variables National Fisheries Data Collection Program

Version 4

Athens, 2021

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

Stamatis Mantziaris, MSc

Ass. Prof. Petros E. Maravelakis

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

The present report describes 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 period 2022-2027. It explains the data frame and describes the probability sampling survey for data collection. In addition, it presents the target and frame population as well as the main sources of information utilized. Also, the report explains the ways used to determine the sample size and how the sample is selected. The specific characteristics and the importance of the different fishing gears and of the corresponding métiers, the vessel length, as well as the location of the vessels (place of registration), are used to determine the sample size and the sample distribution. In the majority of 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. In addition, the report describes the methods used to 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 data confidentiality.

1. Type of Data Collection

1.1 Introduction

The present report describes the methodology adopted to collect socioeconomic data in the fisheries sector following Council Regulation 1004/2017 (EC) and Commission Delegated Decision (EU) 2021/1167. Also, it 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 the H.C.M.R. is responsible for the collection of 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 (Data Collection Scheme) is the census, the special characteristics of the Greek fishing fleet such as the large number of vessels and ports along the 16.000 km of the Greek coastline, prevent the implementation of such a survey. Particularly, given the Greek National Fleet Register retrieved from the EU on September 29th 2021 (available online at https://webgate.ec.europa.eu/fleet-europa/search_en) the Greek fleet consists of more than 14,000 vessels, 94% of which are smaller than 12 meters. In addition, the fishing fleet is distributed across approximately 200 ports.

For the above reasons, the majority of the economic and social variables of the fleet is collected by **Probability Sample Survey**, using face to face interviews and structured questionnaires. Moreover, other data sources are utilised as it is explained in section 3. The sampling frame presented below takes into account the classification of the national fleet according to vessel length and fishing segment, following the Commission Delegated Decision (EU) 2021/1167 (European Commission, 2021).

The chosen method of sampling is **the stratified random sampling** using métiers and vessel length to define the strata as well as geographical areas to ensure a realistic geographical allocation of the sample (for a description of stratified random sampling in the EU-MAP, see also section 3.6. of Deliverable 2.1 from MARE/2016/22 SECFISH study). Tables A.1, A.2, and A.3 in

the Annex present the data collection scheme and the data sources for the activity, economic and social variables. Since the Greek management system does not involve individual 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 National Fleet Register that corresponds to December 31st of the reference year. Moreover, the sample units 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 fishing fleet. The fishing gears are divided into active, passive and polyvalent gears (utilization of both active and passive gears), as Table 2.1 presents. Table 2.1 also presents the métiers (fishing activity level 6) that correspond to each fishing gear. According to EU Multi-Annual Programme (EU-MAP), the vessel length in the Mediterranean Sea is categorized into the following classes: 0-6m, 6-12m, 12-18m, 18-24m, 24-40m and the respective group names used in this report are VL0006, VL0612, VL1218 and VL1824 and VL2440. Table 2.1 shows the total number of fishing vessels by main fishing gear (and métier) groups and length class, features that determine the segments of the fishing fleet. It is worth mentioning that the values of Table 2.1 may differ from the respective data of each year Annual Report. This may happen for the following reasons:

- Table 2.1 includes vessels reported in the national fleet register retrieved from the EU on September 29th, 2021, while the corresponding table in each annual report includes all the vessels reported in the national fleet register of December 31st of the reference year.
- Special annual fishing licenses (for beach seines and for the fishing of Thunnus thynnus, Thunnus alalunga and Xiphias gladius) are taken into consideration for the segmentation. In such cases, the main fishing gear of the vessel is the one acquired by these special licenses irrespective of the main fishing gear reported in the Greek National Fleet Register.
- The characteristics of the population may be adjusted according to the sampling results. More specifically, when the field research shows that the dominant gear of a vessel is different from the one reported in the Greek National Fleet Register, the vessel is categorized according to this segment.

Table 2.2 and Figure 2.1 show the fleet segments with less than ten vessels that are merged with other segments to design the sampling plan and to report economic variables, following the guidelines of STECF (2009). In all cases, the segments merged are very similar to the clustered segments, as they include the same métiers and only differ in the length class. Moreover, clustering does not affect the aggregated segment, since the number of vessels of the merged segments is very small compared to the number of vessels of the clustered segments (see the last column of Table 2.2). Finally, Table 2.3 presents the total number of vessels per fishing segment after clustering.

Finally, it should be noted that the number of inactive vessels is not taken into account *a priori* as no information is available. Thus, the target population is equal to the frame population, as already mentioned.

Table 2.1: The segmentation of the Greek fleet according to the European Union Multi-Annual Program (EU-MAP) and the national fleet register retrieved from EU on September 29th 2021.

			Length Class				
	Gears	**métiers (Activ. level 6)**	**VL0006**	**VL0612**	**VL1218**	**VL1824**	**VL2440**
Active Gears	Beam trawlers (**TBB**)	NA in Greece	-	-	-	-	-
	Demersal trawlers and/or demersal seiners (**DTS**)	SB-SV_DEF_0_0_0 OTB_DEF_>=40_0_0*	3	174	38	101	144
	Pelagic trawlers (**TM**)	-	-	-	-	-	-
	Purse seiners (**PS**)	PS_SPF_>=14_0_0	-	2	81	130	29
	Dredgers (**DRB**/**DRH**)	DRB_MOL_0_0_0	11	24	-	-	-
	Vessels using other active gears (**MGO**)	NA in Greece	-	-	-	-	-
	Vessels using Polyvalent 'active' gears only (**MGP**)	NA in Greece	-	-	-	-	-
Passive Gears	Vessels using hooks (**HOK**)	LHP-LHM_FIF_0_0_0 LLD_LPF_0_0_0 LLS_DEF_0_0_0 LTL_LPF_0_0_0	1759	2688	121	8	-
	Drift and/or fixed netters (**DFN**)	GNS_DEF_>=16_0_0 GTR_DEF_>=16_0_0 GTN_DEF_>16_0_0 GNC**	3194	5541	177	2	-
	Vessels using Pots and/or traps (**FPO**)	FPO_DEF_0_0_0	57	269	9	1	-
	Vessels using other Passive gears (**PGO**)	NA in Greece	-	-	-	-	-
	Vessels using Polyvalent 'passive' gears only (**PGP**)	NA in Greece	-	-	-	-	-
Vessels using active and passive gears (PMP)		NA in Greece	-	-	-	-	-

* SB-SV_DEF_0_0_0 corresponds to the length classes: VL0612 and VL1218, while OTB_DEF_>=40_0_0 corresponds to the length classes: VL1824 and VL244

** There are only four vessels in the Greek National Fleet Register that report this gear as their main gear. For this reason, this métier is considered as not significant and is incorporated in the GTR_DEF_>=16_0_0, for reasons of similarity.

Table 2.2: The clustering scheme of Greek fleet segments (based on the Greek National Fleet Register retrieved from EU on September 29th 2021).

Gear	**Merged segment**		**Clustered segment**		**No of vessels after clustering**
	Length Class	**No. of Vessels**	**Length Class**	**No. of Vessels**	
DFN	VL1824	2	VL1218	177	**179** (98.9% DFNVL1218)
DTS	VL0006	3	VL0612	174	**177** (98.3% DTS0612)
FPO	VL1218 VL1824	9 1	VL0612	269	**279** (96.4% FPOVL0612)
HOK	VL1824	8	VL1218	121	**129** (93.8% HOKVL1218)
PS	VL0612	2	VL1218	81	**83** (97.6% PSVL1218)

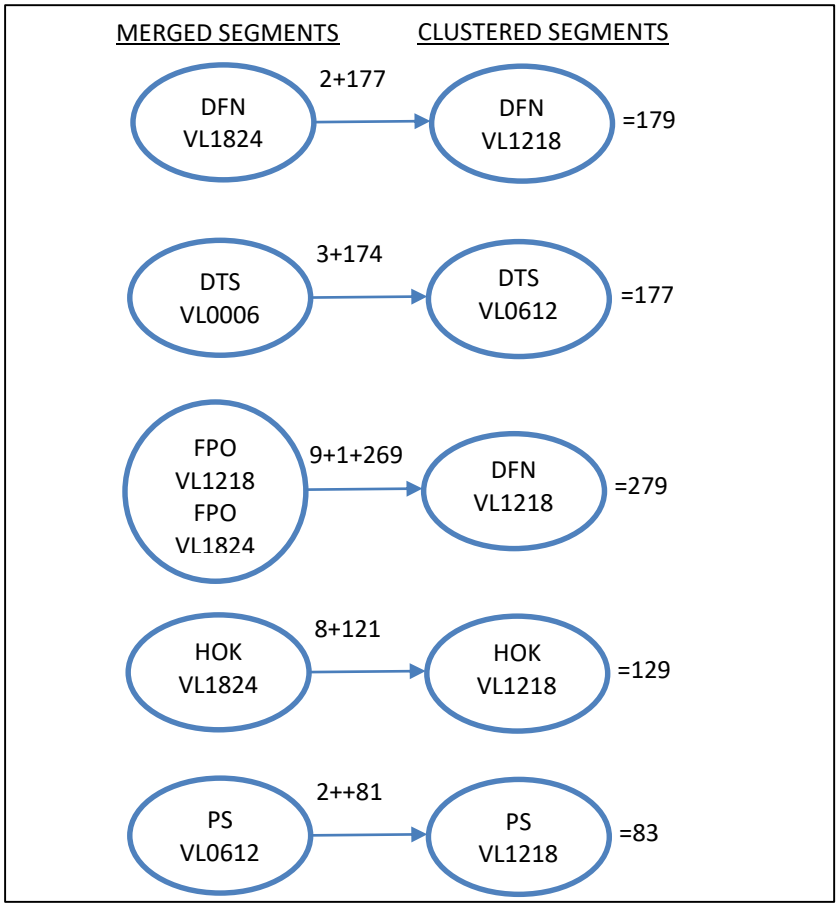

Figure 2.1: The process of clustering and the corresponding number of vessels per cluster.

Table 2.3: The total number of vessels of the Greek fleet after the clustering of segments (based on the Greek National Fleet Register retrieved from the EU on September 29th 2021).

Gear	VL0006	VL0612	VL1218	VL1824	VL2440	Total
DFN	3194	5541	179	-	-	**8914**
DRB	11	24	-	-	-	**35**
DTS		177	33	106	144	**460**
FPO	57	279	-	-	-	**336**
HOK	1759	2688	129	-	-	**4576**
PS	-	-	83	130	29	**242**

3. Data Sources

3.1 National fleet register

The Greek National Fleet Register (NFR) is held by the Hellenic Ministry of Maritime Affairs and Insular Policy and it is the primary source of information for the categorization of fleet segments per gear class and vessel length. The NFR is updated at regular intervals and is available at http://ec.europa.eu/fisheries/fleet/index.cfm. The NFR includes information on the capacity and some other characteristics of each vessel, such as the port of registration, the vessel length, the year of construction as well as the main and secondary fishing gear.

3.2 Questionnaires

For the probability sample survey of the no-census variables, a structured questionnaire drawn up by AGR.E.R.I has been used since 2013 and is continually updated to better serve the purposes of the survey. The questionnaire is filled out by face-to-face interviews with fishermen, while the interviewers are fisheries scientists who receive the appropriate yearly training for the correct completion of the questionnaires. Moreover, the interviewers receive written instructions and they are in regular contact with the AGR.E.R.I research team for questions and clarifications. H.C.M.R and F.R.I are the responsible institutes for the collection of the 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 ERS data exclusively for the proper implementation of the Multi-annual Programme. These data refer to the logbook of all fishing vessels with length more than ten meters as well as those that acquire a special fishing license (beach seiners-SB, handlines and pole lines-LHM and drifting longlines-LLD. Thus, these data can be utilized to estimate activity variables for some of the fleet segments, such as *days at sea* and *landing weight*. The transmission of the above information excludes vessel identity for confidentiality reasons.

Moreover, since 2020 the Ministry of Rural Development and Food provides HAO-Demeter with the list of vessels 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 the basic gear. Finally, data from the Managing Authority of Fisheries and Maritime Operational Programme is utilized for the estimation of the variable “subsidies on investments”, along with questionnaires derived by the probability sample survey.

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 stratum as is described below. The stratified random sampling of the Greek fishing fleet is considered the most proper sampling technique due to the high heterogeneity of the population. 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 the Greek fishing fleet is implemented using métiers (fishing activity level 6) and vessel length classes. In this way, the strata developed and their relationship with métier and length classes is provided in Table 4.2.

Table 4.1. Strata developed and relation with métier s and fleet segments.

Strata	Métier	Fleet Segments
SB_0612	SB-SV_DEF_0_0_0	Demersal trawlers and/or demersal seiners (**DTS**) VL0612
SB_1218		Demersal trawlers and/or demersal seiners (**DTS**) VL1218
OTB_1824	OTB_DEF_>=40_0_0	Demersal trawlers and/or demersal seiners (**DTS**) VL1824
OTB_2440		Demersal trawlers and/or demersal seiners (**DTS**) VL2440
PS_1218	PS_SPF_>=14_0_0	Purse seiners (**PS**) VL1218
PS_1824		Purse seiners (**PS**) VL1824
PS_2440		Purse seiners (**PS**) VL2440
DRB_0006	DRB_MOL_0_0_0	Dredgers (**DRB**/**DRH**) VL006
DRB_0612		Dredgers (**DRB**/**DRH**) VL0612
LHP-LHM_0006	LHP-LHM_FIF_0_0_0	Vessels using hooks (**HOK**) **VL0006**
LLD_0006	LLD_LPF_0_0_0	
LLS_0006	LLS_DEF_0_0_0	
LTL_0006	LTL_LPF_60_0_0	
LHP-LHM_0612	LHP-LHM_FIF_0_0_0	Vessels using hooks (**HOK**) **VL0612**
LLD_0612	LLD_LPF_0_0_0	
LLS_0612	LLS_DEF_0_0_0	
LTL_0612	LTL_LPF_0_0_0	
LHP-LHM_1218	LHP-LHM_FIF_0_0_0	Vessels using hooks (**HOK**) **VL1218**
LLD_1218	LLD_LPF_0_0_0	
LLS_1218	LLS_DEF_0_0_0	
LTL_1218	LTL_LPF_0_0_0	
GNS_0006	GNS_DEF_>=16_0_0	Drift and/or fixed netters (**DFN**) VL0006
GTR_0006	GTR_DEF_>=16_0_0	
GTN_0006	GTN_DEF_>16_0_0	

GNS_0612	GNS_DEF_>=16_0_0	Drift and/or fixed netters (**DFN**) VL0612
GTR_0612	GTR_DEF_>=16_0_0	
GTN_0612	GTN_DEF_>16_0_0	
GNS_1218	GNS_DEF_>=16_0_0	Drift and/or fixed netters (**DFN**) VL1218
GTR_1218	GTR_DEF_>=16_0_0	
GTN_1218	GTN_DEF_>16_0_0	
FPO_0006	FPO_DEF_0_0_0	Vessels using Pots and/or traps (**FPO**) VL0006
FPO_0612		Vessels using Pots and/or traps (**FPO**) VL0612

Furthermore (and following the sampling scheme of the biological data collection), 12 major (geographical) areas are used for a proportional (geographical) allocation of the sample. These areas are the following (see also Figure 4.1): 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). The 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. The H.C.M.R. is responsible for the remaining areas. The total population per stratum (based on métier and length class) and per major area are presented in Table 4.2.

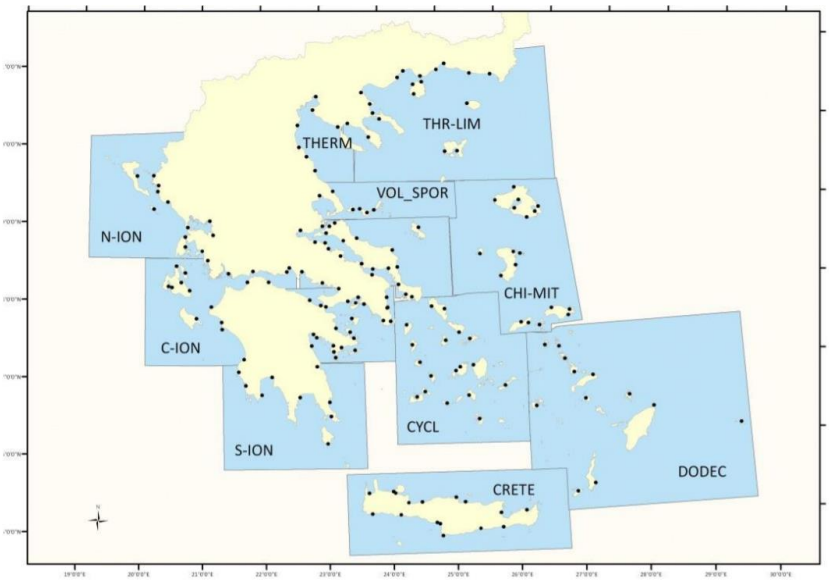

Figure 4.1. the 12 Greek Major Areas

Strata	ARGSAR	CHI-MIT	C-ION	CRETE	CYCL	DODEC	EVIA	N-ION	S-ION	THERM	THR-LIM	VOL-SPOR	Total
DRB_0006	6	1	0	0	0	0	1	0	0	0	3	0	**11**
DRB_0612	11	2	1	0	1	0	0	0	0	3	5	1	**24**
FPO_0006	5	6	2	0	0	4	4	5	0	14	16	1	**57**
FPO_0612	7	12	2	1	0	11	12	8	0	136	83	7	**279**
GNS_0006	88	247	91	40	24	61	63	70	31	68	97	77	**957**
GNS_0612	179	230	156	52	34	56	45	151	43	148	139	93	**1326**
GNS_1218	11	0	4	4	1	1	1	4	1	1	4	3	**35**
GTN_0006	10	16	26	7	45	23	8	12	11	7	5	4	**174**
GTN_0612	66	17	45	5	78	25	25	46	26	16	33	18	**400**
GTN_1218	2	0	0	0	11	1	0	0	0	0	3	0	**17**
GTR_0006	170	207	264	91	66	124	203	305	120	215	158	140	**2063**
GTR_0612	520	276	416	232	211	279	330	538	220	320	346	127	**3815**
GTR_1218	13	10	6	12	33	22	8	1	4	3	13	2	**127**
LHP-LHM_0006	25	25	28	2	7	16	3	7	1	10	11	12	**147**
LHP-LHM_0612	16	16	15	5	6	9	10	7	4	12	14	5	**119**
LHP-LHM_1218	1	0	1	0	2	0	1	0	0	1	0	0	**6**
LLD_0006	6	8	3	3	0	0	2	4	1	1	3	0	**31**
LLD_0612	17	8	12	11	0	6	10	8	1	4	7	1	**85**
LLD_1218	1	0	4	5	1	23	0	0	0	1	2	1	**38**
LLS_0006	207	171	126	155	61	111	175	134	97	134	111	88	**1570**
LLS_0612	425	181	216	216	164	260	237	197	144	201	121	108	**2470**
LLS_1218	12	6	3	7	11	13	5	1	4	3	11	9	**85**
LTL_0006	3	1	0	0	0	2	0	1	0	3	0	1	**11**
LTL_0612	0	0	0	0	0	11	0	0	1	0	0	2	**14**
OTB_1824	18	7	10	2	5	6	9	6	2	11	26	4	**106**
OTB_2440	24	13	5	7	4	4	13	2	2	41	25	4	**144**
PS_1218	20	4	11	3	4	4	19	4	0	5	7	2	**83**
PS_1824	27	4	11	3	4	1	22	8	1	13	22	14	**130**
PS_2440	1	2	0	0	1	0	7	0	0	11	7	0	**29**
SB_0612	28	13	26	2	19	24	12	35	9	5	4	0	**177**
SB_1218	6	0	2	3	11	3	2	0	2	0	4	0	**33**
Total	**1919**	**1482**	**1486**	**868**	**804**	**1100**	**1226**	**1554**	**725**	**1387**	**1277**	**724**	**14552**

Table 4.2: The population of fishing vessels per stratum (based on the Greek National Fleet Register retrieved from the EU on September 29th 2021)

4.3 Sample Size

The sample size (n) per stratum is estimated as (Dattalo, 2008; Moura, 2016):

$$n = \frac{n_0 \cdot N}{n_0 + (N - 1)} \tag{4.1}$$

where N is the population size of each métier and

$$n_0 = \frac{z^2 \cdot s^2}{e^2 \cdot \bar{x}} \tag{4.2}$$

where s is the standard deviation και $\bar{x}$ is the average of the auxiliary variable selected as the one that can better represent the stratum variation. According to Deliverable 2.1 from MARE/2016/22 SECFISH study, auxiliary variables can be utilised for efficient sampling design. In this study, the variable "*days at sea*" is selected as an auxiliary variable, as, according to experts' knowledge, effort is one of the most important sources of variation among variables The level of significance (z) and the margin of error (e) are selected by the importance of each stratum, taking into consideration effort and landings data, were available, as well as the number of vessels and expert knowledge. Four different groups of strata have been considered using a different combination of values for e and z (see Table 4.3).

Table 4.3: The fleet strata with the respective level of significance (z) and margin of error (e) used to determine the sample size.

Very Important strata (z=1.64, e=10%)	Important strata (z=1.64, e=15%)	Less Important strata (z=1.64, e=20%)	Not Important strata (z=1.28, e=20%)
GNS_0612	FPO_0612	FPO_0006	DRB_0006
GTR_0612	GTR_0006	GNS_0006	DRB_0612
LLS_0612	GTN_0612	GNS_1218	GTN_1218
OTB_1824	GTR_1218	GTN_0006	LHP-LHM_1218
OTB_2440	LLS_1218	GTR_1218	LLD_0006
PS_1824	PS_1218	LHP-LHM_0006	LTL_0612
PS_2440		LHP-LHM_0612	LTL_0006
		LLD_0612	SB_0612
		LLD_1218	SB_1218
		LLS_0006	
		LLS_1218	

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

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

As mentioned earlier, after the determination of the sample size of the strata, the number of vessels in the sample per major area is determined using the method of proportional allocation (Eurostat, 2008):

$$n_g = \frac{n \cdot N_g}{N} \quad (4.4)$$

where n is the sample size per stratum, as it was calculated by equation 4.3, N_g is the sum of vessels of the geographical area per stratum and N is the population size for each stratum. Using the proportional allocation, the proportions are respected and, consequently, the share of a geographic area 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 total sample size for the reference year 2021 is set to 761 vessels and is presented in Table 4.4.

In conclusion, the determination of sample size and the sample geographical distribution is based on the specific métiers and their importance, the vessel length class and the variability of the strata, by taking into account a representative auxiliary variable (days at sea) and the 12 major geographic areas. Combining the information providing in Tables 4.2 and 4.4, Tables 4.5 and 4.6 present the planned sample size and sample rate per segment, respectively.

Strata	ARGSAR	CHI-MIT	C-ION	CRETE	CYCL	DODEC	EVIA	N-ION	S-ION	THERM	THR-LIM	VOL-SPOR	Total
DRB_0006	2	-	-	-	-	-	-	-	-	-	1	-	**3**
DRB_0612	3	-	-	-	-	-	-	-	-	1	2	-	**6**
FPO_0006	1	2	-	-	-	1	1	1	-	3	3	-	**12**
FPO_0612	1	1	-	-	-	1	1	1	-	10	7	1	**23**
GNS_0006	3	6	3	1	1	2	2	2	1	2	3	2	**28**
GNS_0612	5	6	4	2	1	2	2	4	2	4	4	3	**39**
GNS_1218	4	-	2	2	-	-	-	2	-	-	2	1	**13**
GTN_0006	1	2	2	-	3	2	1	1	1	-	-	-	**13**
GTN_0612	6	2	4	-	7	2	2	4	3	2	3	2	**37**
GTN_1218	1	-	-	-	4	-	-	-	-	-	1	-	**6**
GTR_0006	4	5	6	2	2	3	5	7	3	5	4	3	**49**
GTR_0612	12	7	10	6	5	7	8	13	5	8	8	3	**92**
GTR_1218	3	2	2	3	6	4	2	-	1	1	3	-	**27**
LHP-LHM_0006	4	4	5	-	2	3	-	2	-	2	2	2	**26**
LHP-LHM_0612	3	3	3	1	2	2	2	2	1	3	3	1	**26**
LHP-LHM_1218	-	-	-	-	1	-	-	-	-	-	-	-	**1**
LLD_0006	2	3	1	1	-	-	1	2	-	-	1	-	**11**
LLD_0612	4	2	3	2	-	2	2	2	-	1	2	-	**20**
LLD_1218	-	-	1	2	-	5	-	-	-	-	-	-	**8**
LLS_0006	5	4	3	4	2	3	4	3	2	3	3	2	**38**
LLS_0612	20	9	10	10	8	12	11	10	7	10	6	5	**118**
LLS_1218	3	2	1	2	3	3	2	-	1	1	3	3	**24**
LTL_0006	2	-	-	-	-	1	-	-	-	2	-	-	**5**
LTL_0612	-	-	-	-	-	4	-	-	-	-	-	1	**5**
OTB_1824	5	2	3	-	2	2	3	2	-	3	7	1	**30**
OTB_2440	6	3	2	2	1	1	3	-	-	9	6	1	**34**
PS_1218	4	1	3	1	1	1	4	1	-	1	2	-	**19**
PS_1824	7	1	3	1	1	-	6	2	-	3	6	4	**34**
PS_2440	-	1	-	-	-	-	3	-	-	5	3	-	**12**
SB_0612	2	1	2	-	2	2	1	3	1	-	-	-	**14**
SB_1218	2	-	-	1	3	1	-	-	-	-	1	-	**8**
Total	**115**	**69**	**73**	**43**	**57**	**66**	**66**	**64**	**28**	**79**	**86**	**35**	**781**

Table 4.4: Sample size of fishing vessel per stratum (based on the Greek National Fleet Register retrieved from the EU on September 29th 2021).

Table 4.5: The planned sample size per fleet segment

	Length category				
	VL0006	VL0612	VL1218	VL1824	VL2440
DFN (GNS+GTN+GTR)	90	168	46	0*	0
DRB	3	6	0	0	0
DTS (OTB+SB)	0	14	8	30	34
FPO	12	23	0	0	0
HOK (LHP/LHM+LLD+LLS+LTL)	80	169	33	0	0
PS	0	0	19	34	12

*The sample size is zero when there are no vessels on a population segment after the clustering

Table 4.6: The planned sample rate per segment

	Length category				
	VL0006	VL0612	VL1218	VL1824	VL2440
DFN	2.82%	3.03%	25.70%	-	-
DRB	27.27%	25.00%	-	-	-
DTS	-	7.91%	24.24%	28.30%	23.61%
FPO	21.05%	8.24%	-	-	-
HOK	4.55%	6.29%	25.58%	-	-
PS	-	-	22.89%	26.15%	41.38%

4.4 Sample Selection

The sample selection, as already mentioned, is random. A random number between 0 and 1 is attached to each vessel of the fleet and the vessels of a specific major area that belong to a specific stratum are sorted from the lowest to the highest random number. Then, the first n^{th} vessels are included in the sample of size n. In this way, the selection of the sample is achieved without replacement and each vessel has an equal probability to be selected.

In the same manner, a complementary sample is also selected to address the non-response of sampling units. If the complementary sample is exhausted in the defined geographical areas, sampling shall be carried out in adjacent areas and vessels with similar characteristics in terms of vessel length and main fishing gear.

4.5 Sample Evolution

The main differences in the proposed sample size during the years occur due to the changes in the segments' significance. Table 4.7 shows the sample rate per segment from 2018 onwards.

Table 4.7: Evolution of proposed sample coverage rate during the period 2018-2021.

Gear Segment	2018	2019	2020	2021
DFN_VL0006	1.64%	1.64%	1.64%	2.82%
DFN_VL0612	2.86%	2.86%	2.86%	3.03%
DFN_VL1218	40.15%	40.15%	40.15%	25.70%
DRB_VL0006				27.27%
DRB_VL0612	17.39%	17.39%	17.39%	25.00%
DTS_VL0612	18.18%	18.18%	18.18%	7.91%
DTS_VL1218	37.50%	37.50%	37.50%	24.24%
DTS_VL1824	45.10%	45.10%	45.10%	28.30%
DTS_VL2440	28.08%	28.08%	28.08%	23.61%
FPO_VL0006	27.94%	27.94%	27.94%	21.05%
FPO_VL0612	6.12%	6.12%	6.12%	8.24%
HOK_VL0006	2.06%	2.06%	2.06%	4.55%
HOK_VL0612	8.31%	8.31%	8.31%	6.29%
HOK_VL1218	46.56%	46.56%	46.56%	25.58%
PS_VL1218	48.84%	48.84%	48.84%	22.89%
PS_VL1824	40.44%	40.44%	40.44%	26.15%
PS_VL2440	57.69%	57.69%	57.69%	41.38%

5. Estimation Procedure and Modelling

5.1 Estimation

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 (raising), where applicable. Following, among others, the Deliverable 2.1 from MARE/2016/22 SECFISH study, the Horvitz-Thompson (HT) estimator (Horvitz and Thompson, 1952) is used to allow correct generalization of the sample statistics to the population parameters per statum:

$$\widehat{\mathrm{Y}} = \sum_{i=1}^{\mathrm{n}} \mathrm{y}_i \, \pi_i{}^{-1} \qquad (5.1)$$

where $\pi_i = \mathrm{n}_{\mathrm{act}}/\mathrm{N}_{\mathrm{act}}$ is the inclusion probability for each stratum, $\mathrm{n}_{\mathrm{act}}$ the number of active vessels in sample per stratum and $\mathrm{N}_{\mathrm{act}}$ is the number of active vessels in population per stratum. Then, the estimation of a variable per fleet segment is done by combining the values of the corresponding strata (e.g., a variable of the segment DFN_VL006 is calculated by adding up the values of the strata: GNS_006+GTN_0006+GTR_0006).

Using the above formula, the number of inactive vessels of the population in each stratum is also estimated, before the estimation of other variables[1]. This estimation is done for the strata with vessels less than 12 meters in length, except for the strata utilising the gears SB, LLD and LHM, where inactivity can be based on ERS information. Then, the rest of the capacity variables are also estimated for the inactive vessels.

The economic variables: *Consumption of fixed capital* and *Value of physical capital* are estimated using data from both the probability sample survey (replacement value) and the NFR (mean LOA and number of vessels per fleet segment), as 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 equipment – 10%

[1] 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.

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 considered to evaluate the results of the survey.

5.2 Modelling

Statistical modelling is able to explain the variability of crucial socio-economic variables. Standard statistical techniques like linear regression provide a flexible tool to address such problems. Moreover, linear regression can be used as an alternative method to handle outliers and non-response errors. The use of a powerful model can give the researcher the ability to predict values of the dependent variable based on a specific value(s) of the independent(s) variable(s). A model that can fit the data neatly offers the researcher the ability to conclude different policies and future change scenarios (e.g., environmental risks). Under this perspective, we present in the following subparagraphs linear regression models to aid the researcher in explaining the inner relationships between variables using IBM SPSS Statistics 26. It is important to clarify that this kind of modelling is still in a pilot phase and under consideration, so it is not yet utilized for variables estimation. The results provided in the following sections are based on the data regarding the reference year 2018.

5.2.1. Full-Time Equivalents (FTEs)

The level of Full-time equivalents (FTE) is modelled using segment, length, days at sea and fishing time. The results are given in Table 5.1.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,912[a]	,831	,829	1,10416

a. Predictors: (Constant), DAYS_AT_SEA, OTB, DFN, LLD, FPO, LLS, VL0006, VL0612, VL1218, VL1824, FISHING_TIME

ANOVA[a,b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4288,201	11	389,836	319,754	,000[b]
	Residual	870,493	714	1,219		
	Total	5158,694	725			

a. Dependent Variable: FTE

b. Predictors: (Constant), DAYS_AT_SEA, OTB, DFN, LLD, FPO, , LLS, VL0006, VL0612, VL1218, VL1824, , FISHING_TIME

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	8,410	,296		28,449	,000
	DAYS_AT_SEA	,004	,001	,094	4,250	,000
	FISHING_TIME	,0003494	,000	,131	5,561	,000
	VL0006	-3,783	,379	-,554	-9,975	,000
	VL0612	-3,749	,365	-,700	-10,283	,000
	VL1218	-2,898	,325	-,371	-8,918	,000
	VL1824	-1,398	,247	-,140	-5,657	,000
	DFN	-4,921	,260	-,923	-18,899	,000
	FPO	-5,139	,298	-,514	-17,249	,000
	LLD	-3,924	,297	-,283	-13,196	,000
	LLS	-4,901	,277	-,743	-17,688	,000
	OTB	-4,198	,242	-,395	-17,339	,000

a. Dependent Variable: FTE

Table 5.1: Linear Regression between Full-time equivalents and segment, length, days at sea, fishing time.

The equation is:

$$\widehat{FTE} = 8.41 + 0.004\ days\ at\ sea + 0.00035\ Fishing\ Time - 3.783\ VL0006 - 3.749\ VL0612 - 2.898\ VL1218 - 1.398\ VL1824 - 4.921\ DFN - 5.139\ FPO - 3.921\ LLD - 4.901\ LLS - 4.198\ OTB.$$

The interpretation of the coefficient of Days at Sea (0.004) is that for every unit increase in days at sea, FTE increases on average by 0.004 given that the remaining variables remain constant. Using the same way of thinking we can interpret the remaining coefficients of the model. This model is statistically significant since the p-value is approaching zero (smaller than α=0.05, and

also smaller than all the usually used significance levels). This model can explain R^2=83.1% of the total variability of FTE.

5.2.2 Fuel Consumption

Fuel Consumption is modelled using segment, capacity unit(s), days at sea and trip duration. The results are given in Table 5.2.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,934[a]	,873	,871	17512,22892

a. Predictors: (Constant), Trip_duration, Days_at_Sea, FPO, LLD, LLS, OTB, Capacity_Units, DFN

ANOVA[a,b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1490825788297	8	1863532235370	607,651	,000[b]
	Residual	2174348166670	709	3066781601		
	Total	1708260604968	717			

a. Dependent Variable: Fuel_Consumption

b. Predictors: (Constant), Trip_duration, Days_at_sea, FPO, LLD, LLS, OTB, DFN, Capacity_Units

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3968,542	3512,849		1,130	,259
	DFN	-16817,018	2753,702	-,172	-6,107	,000

	FPO	-15151,766	3560,297	-,082	-4,256	,000
	LLD	-5743,486	4041,399	-,022	-1,421	,156
	LLS	-14583,416	3003,954	-,121	-4,855	,000
	OTB	72119,082	3841,424	,373	18,774	,000
	Capacity_Units	855,851	42,000	,495	20,377	,000
	Days_at_sea	55,248	10,578	,072	5,223	,000
	Trip_duration	370,098	130,249	,042	2,841	,005

a. Dependent Variable: Fuel_Consumption

Table 5.2: Linear Regression between Fuel Consumption and segment, capacity unit(s), days at sea, trip duration.

The equation is equal to:

$$\widehat{Fuel\ Consumption} = 3968.54 + 55.248\ days\ at\ sea + 855.851\ Capacity\ Units + 370.098\ Trip\ duration - 16817.018\ DFN - 15151.77\ FPO - 5743.49\ LLD - 14583.42\ LLS + 72119.08\ OTB.$$

The interpretation of the coefficient of Days at Sea (55.248) is that for every unit increase in days at sea, Fuel Consumption increases on average by 55.248 given that the remaining variables remain constant. Using the same way of thinking we can interpret the remaining coefficients of the model. This model is statistically significant since the p-value is approaching zero (smaller than 0.05=α, and also smaller than all the usually used significance levels). This model can explain R^2=87.3% of the total variability of Fuel Consumption.

5.2.3 Fuel Cost

Fuel Cost is modelled using the product Fuel consumption by segment and segment. The results are given in Table 5.3.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate

1	,968[a]	,938	,937	6220,72069

a. Predictors: (Constant), OTB, LLD, FPO, LLS, FuelConsSEG, DFN

ANOVA[a,b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	417355602300,077	6	69559267050,013	1797,519	,000[b]
	Residual	27823406075,844	719	38697365,891		
	Total	445179008375,921	725			

a. Dependent Variable: fuel_cost

b. Predictors: (Constant), OTB, LLD, FPO, LLS, FuelConsSEG, DFN

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	115,653	999,055		,116	,908
	FuelConsSEG	,061	,001	,876	43,184	,000
	DFN	3384,408	1046,431	,068	3,234	,001
	FPO	3655,978	1289,744	,039	2,835	,005
	LLD	12370,474	1471,908	,096	8,404	,000
	LLS	2277,799	1106,983	,037	2,058	,040
	OTB	15309,868	1510,385	,155	10,136	,000

a. Dependent Variable: fuel_cost

Table 5.3: Linear Regression between Fuel Cost and the product of Fuel consumption by segment and segment

The equation is:

$$\widehat{Fuel\ Cost} = 115.65 + 0.061\ Fuel\ Cons * Segm + 3384.41\ DFN + 3655.98\ FPO + 12370.47\ LLD + 2277.8\ LLS + 15309.87\ OTB$$

The interpretation of the coefficient of Fuel Cons*Segm 0.061 is that for every unit increase in the product Fuel Consumption by Segment, Fuel Cost increases on average by 0.061 given that the remaining variables remain constant. Using the same way of thinking we can interpret the

remaining coefficients of the model. This model is statistically significant since the p-value is approaching zero (smaller than 0.05=α, and also smaller than all the usually used significance levels). The model can explain R^2=93.8% of the total variability of Fuel Cost.

5.2.4 Other Variable Costs

The "Other variable costs" variable is modelled by Commercial costs, segment and length. The results are given in Table 5.4.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,949[a]	,901	,899	11097,32276

a. Predictors: (Constant), OTB, DFN, LLD, FPO, , LLS, Commercial_costs, VL0006, VL0612, VL1218, VL1824

ANOVA[a,b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	768304531317,844	10	76830453131,784	623,874	,000[b]
	Residual	84850744338,419	689	123150572,334		
	Total	853155275656,263	699			

a. Dependent Variable: Other_variable_cost

b. Predictors: (Constant), OTB, DFN, LLD, FPO, , LLS, Commercial_costs, VL0006, VL0612, VL1218, VL1824,

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	56353,557	3751,938		15,020	,000
	Commercial_costs	,680	,021	,656	31,948	,000

VL0006	-45863,390	4290,879	-,518	-10,689	,000
VL0612	-45192,375	4201,071	-,642	-10,757	,000
VL1218	-38880,479	3824,352	-,367	-10,167	,000
VL1824	-13778,127	2758,732	-,101	-4,994	,000
DFN	-10595,346	3196,804	-,152	-3,314	,001
FPO	-10272,541	3512,559	-,080	-2,925	,004
LLD	10732,045	3518,427	,058	3,050	,002
LLS	-8387,966	3328,816	-,098	-2,520	,012
OTB	-30231,076	2820,707	-,206	-10,718	,000

a. Dependent Variable: Other_variable_cost

Table 5.4: Linear Regression between Other variable costs and Commercial costs, segment, length.

The equation is:

$$\widehat{Other\_variable\_cost} = 56353.56 + 0.68\ Commercial\_Costs - 45863.37VL0006 - 45192.38VL0612 - 38880.48VL1218 - 13778.13VL1824 - 10595.35\ DFN - 10272.54\ FPO + 10732.05\ LLD - 8387.97\ LLS - 30231.08\ OTB$$

The interpretation of the Commercial Costs' coefficient (0.68) is that for every unit increase in Commercial Costs, Other Variable Cost increases on average by 0.68 given that the remaining variables remain constant. Using the same way of thinking we can interpret the remaining coefficients of the model. This model is statistically significant since the p-value is approaching zero (smaller than 0.05=α, and also smaller than all the usually used significance levels). The model can explain R^2=90.1% of the total variability of the variable *other variable costs*.

6. Data Quality Evaluation

6.1 Methodology Relevance

The methodology adopted and described in the previous chapters is controlled for its proper implementation at all stages. In particular, during the sampling period, AGR.E.R.I research team communicates regularly with the interviewers to ascertain the proper process of data collection. When the field research is finalised, the questionnaires are evaluated; for example, the number of questionnaires collected per interviewer and the completeness of the data are checked.

The planned sample rate of the proposed sample per fleet segment (Table 4.6) is compared with the actual sample rate. Small discrepancies are allowed in the assessment for reasons such as misclassification of the fishing fleet (see section 6.2) or non-updating of the NFR data during the sampling design.

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 to a main and secondary fishing gear. 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 1004/2017 and the recent Commission Delegated Decision (EU) 2021/1167. More specifically, the annual report includes all socio-economic variables and activity variables for all segments 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 following the principles of Impartiality, Reliability and Objectivity using only official sources of information. 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 is described in the next sections. The presentation of the results for all socio-economic and activity variables is detailed for all segments of the fishing fleet.

6.3 Accuracy sampling Errors

6.3.1 Sampling Errors

The desired accuracy for each fleet segment is determined according to the importance of the strata included in the fleet segments (see 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 whose variance better reflects the variability of all variables has to be chosen. This variable is "*days at sea*".

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

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

where SD is the standard deviation and $\overline{\mathrm{X}}$ 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):

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

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 degrees of freedom and S^* the unbiased estimator of the standard deviation of the sample, which is determined by the following equation:

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

As the assumption that the data $X_1, X_2, \ldots X_n$ follows normal distribution does not usually hold, we also use the Wilcoxon-signed rank confidence interval. Let us consider the $N = n(n+1)/2$ possible means of the form $(X_i + X_j)/2$ for all $i \leq j = 1, \ldots, n$. We have to find those means that are larger than a value m (equal to zero unless differently stated). Let $w_{a/2}$ denote the critical values of the statistical function T^+ of the Wilcoxon signed-rank test, where T^+ is equal to the sum of the positive signed ranks. We find the k=$w_{a/2}$ larger and the k=$w_{a/2}$ smaller such means. The confidence interval with at least 100(1-α)% confidence is given by $\left[X^{(k)}, X^{(N-k+1)}\right]$ where $X^{(k)}$ is the *k*-th ordered observation (Hollander et al., 2014).

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, to avoid convenience sampling. 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 4.5 shows the coverage rate which is the percentage of sample-to-population for each fleet segment of the fishing fleet.

6.3.3 Measurement Errors

Before the data analysis, data are checked for completeness, cohesiveness and comparability over time. More specifically, exploratory data analysis is used, to locate measurement and processing errors, through unreasonable and 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 non-negative by definition variable, e.g. landings weight). An extreme value is a considerably remote value, compared to the majority of the rest values.

To address the problems related to the unreasonable or extreme values that appear on some vessels in basic technical and economic parameters that are used for the estimation of the data call variables, we attempt to construct benchmark tables with control ranges per segment for each parameter. Then, to cope with the unreasonable or extreme value problem, the values that are outside the ranges of the benchmark tables are replaced by the median value of the parameter. For example, “Gross value of landings per kilogram of live weight”, “Landings per day” and “Days at sea” are auxiliary parameters for the calculation of the variable “Gross value of landings” at vessel level. The parameters "Days at sea" and "Value of landings" are used to "homogenize" the data and subtract that portion of the variance caused by the diversification of the fishing effort. The auxiliary parameters are presented in Table 6.1.

The data was used to determine the ranges derive from observations per vessel during the period 2012-2018. The steps to determine the control ranges and to fix the parameter values are summarized as follows:

1. Construct a cumulative frequency distribution for the parameter under consideration per fleet segment.
2. Calculate descriptive statistics.
3. Set the limit equal to the 95% percentage point of the observations.
4. Incorporate an additional threshold up to the level of the two standard deviations (if applicable).
5. Use of rational/natural boundaries (for example values that have to be larger than zero) and experts’ knowledge.
6. Check the values per vessel (if many observations lie beyond the specified range then it is expanded by ± 20%).
7. Replace by the parameter median value of the fleet segment if required.
8. Transform each parameter properly, to have annually based values.

Table 6.1: Parameters utilized to build control ranges for the benchmark tables.

Parameters (at the vessel level)
Days at sea
Crew members
Energy consumption per day (Liters/day)
Landings or Production per day (kg/day)
Price (euros/kg)
Repair and maintenance costs per day (euros/day)
Other Variable costs per day (euros/day)
Energy cost per day (euros/day)
Personnel cost per day (euros/day)
Energy cost / Value of landings (%)
Other Variable costs / Value of landings (%)
Repair and maintenance costs / Value of landings (%)
Fixed costs/ Value of landings (%)
Personnel cost / Value of landings (%)
Total costs / Value of landings (%)

An alternative process can be also used to locate extreme values. Specifically, a value may be characterized as extreme if it is located outside the following interval for each variable (Tukey, 1977):

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

where Q1 and Q3 are the first and third quartile of the variable, respectively and IQR is the interquartile range.

When unreasonable or extreme values refer to data that has been collected by the census, the source of information is checked and data are corrected in case of mismatch. In case of an accord, the unreasonable or extreme values are replaced with missing values. The same applies to data that has come from the probability survey sampling, after communicating with the interviewers 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):

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

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. To find possible omissions in the questionnaire's design or the interviewer'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 interviewer and
- incorrect answers of the interviewee

The incorrect values due to the questionnaire's design include the values that are repeated in different segments of the fishing fleet and are located in data coming from different interviewers. The incorrect answers caused by the interviewer consist of the values that are repeated in the data coming from the same interviewer 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 questions in free text format.

6.3.4 Non-response Errors

To address the non-response of the sampling unit (vessel), a complementary sample is selected. The complementary sample, as well as the main sample, is drawn randomly. The selection order of the complementary sample is determined in advance to avoid convenient sampling. 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 interviewers and the ability to resolve any questions in close cooperation with the AGR.E.R.I research team. During the interview, the interviewers pay special attention to the variables which had high non-response rates in previous years. Variables with frequent non-response are highlighted before the design of the questionnaire. Finally, the percentage of PNR is calculated for each variable.

The PNR cases as well as 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 of 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 a high number of employers. Finally, as MNAR are classified the missing values that are correlated with the variable. For example, the values of annual revenues are missing for vessels with high revenues. 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 either by unconditional or conditional imputation. In the first case, the missing values of continuous variables are replaced directly by the median of the non-missing values whereas the missing values of nominal variables are replaced by the respective mode value of the variable in a stratum. If two or more values are missing for a specific variable, all the missing values will be substituted by the same number, a fact that leads to underestimation of the true variance due to the imputation of missing values at the centre of the distribution.

For the above reasons, conditional median imputation, which is also known as regression imputation (RG), is used in this analysis (see also section 5.3.4. of Deliverable 2.1 from

MARE/2016/22 SECFISH study). To illustrate the idea of the RG method, let us assume that we have (n × k) data set, where the columns are $x_1, \ldots, x_k$. Consider a univariate missing pattern with $x_1, \ldots, x_{k-1}$ fully observed and x_k observed for the first r values and missing for the last $n - r$ values. The RG method computes the regression of x_k on $x_1, \ldots, x_{k-1}$ based on the r complete cases, and then fills in the missing values by the $n - r$ predicted values using the regression model. The models given in section 5.2 can be used to predict the missing values for each observation by substituting the values of the present variables in the regression equations.

Another method that can be used for handling missing values is the complete case analysis (CC). The CC is the most widely used method of handling missing values. This method considers only samples where all the variables are present and discards those whose information is incomplete. It is also known as case deletion and listwise deletion in the literature (Little and Rubin, 2002). The main advantages of this approach are simplicity and comparability; simplicity since standard statistical analysis (for data without missing values) can be applied without modifications, and comparability of univariate statistics since they are calculated using the same pool of cases. Moreover, if the mechanism of missing values is MCAR, the remaining cases (reduced sample) will be a random subsample of the original sample. This implies that, for any parameter of interest, the estimates are unbiased for the full data set and that they are also unbiased for the listwise deleted data set. Disadvantages of the CC analysis are loss of information, which leads to loss of precision, and bias when the missing data mechanism is not MCAR. A straightforward result of the bias is that complete cases are not a random subsample of the original sample. It should be emphasized that complete case analysis is only used in extreme cases of questionnaires with large missing or problematic sections.

7. Accessibility and Clarity

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 a short version presented in the Work Plan. They are also informed about the quality of the statistical results, following the standards of the European Statistical System. 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.

The survey files including case information are created using a specific layout. The survey data can be unequivocally mapped to the internationally required fields, codes and formats if asked. Deviations from the prescribed record format must be reconciled before the data can be integrated for further processing.

All data collected are imported into a database, following specifications in the corresponding operational manuals and national record layouts. All data are verified for structural consistency within and across sources and for agreement with the defined formats and record layouts. It has to be assured that sample design and disposition data are recorded for every case. All adaptations are thoroughly tested prior to the production use of the data integration software.

Data are imported regularly as the survey progresses. Adaptations to the context are reflected in the record layout before data are imported, based on the corresponding documentation. The verification of the database includes the following: 1) a unique ID check, 2) a valid value check for nominal/ordinal variables, 3) a valid range check for continuous variables and 4) cross-table consistency checks.

The delivery of data follows the adapted record layout, prior to any necessary recoding or mapping to re-establish the international record layout, if such a layout exists. Necessary structural adjustments following data submission will be done unless agreed differently. Data delivery is made through secure channels (i.e. the project's SharePoint site or a secured FTP connection), in a folder to which only specific users have access. All data available on web services are accessible only through encrypted connections (HTTPS/SSL) and access control mechanisms. Each user has only a limited set of user rights.

8. Coherence and comparability

The data collection follows the European Regulations and therefore are comparable to those of the other European countries, as long 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 section 4.5.

The interviewers could greatly affect the data collection process which in turn influences the comparability of the results over time. The scientific coordinator reviews the planning reports to determine whether the survey requirements have been satisfied. To ensure comparability of the results across countries, the survey design plan will be consistent with the objectives and standards set, methodologically sound and operationally practical. The review of the scientific coordinator also gives an advanced opportunity to detect potential problems with the survey design plan and to provide advice on dealing with these issues.

To facilitate comparability in data analysis, the dataset is mapped into a highly structured, standardized record layout. In addition to specifying the position, format and length of each field, the record layout includes a description of each variable and indicates the categories and codes to be provided for that variable. Upon receiving the file, the data manager performs a series of range checks to ensure compliance with the prescribed format and run flow and consistency edits on the file. When anomalies are detected, proper personnel is notified of the problem and asked to submit cleaned files.

9. Confidentiality, Transparency and Security

9.1 Introduction

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 takes all appropriate precautions to insure 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 Maritime Affairs and Insular 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.

9.2 Respondents

Participants in the survey are provided with the following information on the data collection delivered in person by the interviewer:

- funding, purpose and duration of research
- explanation of how the respondent was selected for the study
- interview procedures to be followed by the respondent
- voluntary nature of participation
- expected risks and benefits
- maintenance of confidentiality
- right to withdraw from the study at any time without penalty

The institute will maintain the confidentiality of respondent data. All the involved staff sign a pledge of confidentiality and a non-disclosure form.

9.3 Researchers-AGR.E.R.I

Researchers maintain a copy of the following documents:

- scripts, letters, fact sheets and any other materials provided to give respondents information they need to make an informed decision about participation;

- consent protocols;
- confidentiality procedures and protocols;
- confidentiality agreements completed by staff;

AGR.E.R.I possesses substantial practical experience in the following areas:

- Survey management;
- Probability sample design and sample selection;
- In-person data collection;
- Computer-assisted personal interviewing (CAPI);
- Instrument and materials translation/adaptation;
- Interviewer training;
- Achievement of high response rates on surveys;
- Maintenance of data confidentiality;
- Data processing, including data capture, coding and editing;
- Sample weighting and estimation.

The staff is encouraged to report any confidentiality concerns that limit data sharing. A document describes in detail the encryption software on computers, the group policies which limit user access on computers, who has administrative control on the computers and how is security on folder level is realized in the host operating system of computers.

A confidentiality agreement is signed by the involved staff including Field supervisors; Interviewers; Coders; and Data processing personnel. These agreements are kept in a physical file for future use if appropriate. There is special education of training interviewers on the importance of data security. The main goal is to maintain the continued confidentiality and security of the survey materials and respondent data during data capture, coding, scoring, and processing. All of the work for coding, scoring, and processing will be carried out within the premises and reach of the survey organization.

Training of the interviewers includes the following components:

• An introduction to survey research, providing examples of types of survey questions and interviewing terminology;

• The conventions for asking survey questions and recording answers;

• Written and oral exercises on asking questions, recording responses and applying probing techniques to obtain accurate data;

• Gaining respondent co-operation, in which the following concepts are discussed:

  - o Interviewer behaviour and style when making contact with the respondent;
  - o The importance of making effective and fast connections with the respondent;

- Methods to overcome resistance and address respondent concerns;
- Written and oral exercises on refusal aversion techniques and how to answer respondent questions.

- Standards and ethics in survey research, including information on informed consent, data confidentiality, and data security and written exercises that include consent and confidentiality scenarios that interviewers may confront during data collection;

9.4 Data Collection Process

Initially, the respondent is located in the port and a letter printed on official stationery is provided. An attractive brochure is developed and it is given in conjunction with the official letter to the respondent during the contact. This brochure further serves to legitimize the study, stress the study's importance and motivate respondent co-operation. The following information is included in the brochure:

- Why the study matters and why the respondent should participate
- Topics included in the survey
- How the respondent's name and the port was obtained
- Why another person cannot substitute the respondent
- Confidentiality of the data
- Users of the data
- Length of the interview
- URL of the survey website

Upon agreement of the respondent to participate in the study and the signing of the Confidentiality declaration form (see Section 9.5) a face-to-face interview is contacted and the collected data are transferred to the database by the interviewer.

9.5 Confidentiality Declaration Form

SAMPLE CONFIDENTIALITY FORM

Confidentiality Agreement

Name: _____

Phone number: _____

E-mail: _____

Address: _____

_____

This is to certify that I, _____, have agreed to provide _____ services within the framework of the:

National Fisheries data Collection Program

I understand that the signing of this form is an acknowledgement of my professional responsibility to maintain complete integrity of security for this project. I declare that I will not divulge any project information, trial materials, test materials, processes, contents or results, or any other materials, documents, or information pertaining to the project, or its clients or suppliers, to any person or organization, as directed under the terms of the project.

I understand that the above does not apply to information that is in the public domain.

I have read and accepted the conditions as outlined above.

Signature _____

Date _____

10 General Data Protection Regulation

10.1 Data Collection

AGR.E.R.I collects personal information from respondents using alternative methodologies. The collection is done with the consent of the respondents. The data collected relate to demographic, social and/or economic data relevant to the purpose of each survey.

10.2 Time of Keeping the Data in Identifiable Form

When conducting quantitative surveys, which are carried out by the researcher using the method of personal interviews, the Institute shall keep the data of the respondents in a personalized form. Such data shall be kept in this format for a limited time, only to check the quality and authenticity of the data and the reliability of the research.

The institute checks at least 15% of the completed questionnaires during or immediately after the survey is completed. Respondents' personal data, at the latest after three months, are anonymized to make it impossible to identify the personal data with the survey data.

10.3 Rights

Respondents, as long as the institute holds their personal data in identifiable form, have the following rights:

- The right to withdraw their consent. The revocation shall take effect upon its submission.
- The right to disclose their data. Respondents will be notified of their data within 15 days of receiving the request at the latest.
- The right to correct their data. The correction will apply to any processing that may occur after the initial disclosure of the data to the Institute.
- The right to delete their data. Deletion will take place no later than 30 days after the request is received.
- The right to report to the supervisory authority in case of breach of data security or illegal processing.

The above rights are exercised for all types of rights upon request to the Institute. The institute will contact interested parties to confirm the request while the email address with the request to the Institute's Data Protection Officer (DPO) is: Vasileios Paliouras (dpo@elgo.gr).

10.4 Transmission of Data to Third Parties

The Institute does not transmit personal data to third parties. Personal data are transmitted only with the consent of the respondent after being informed of the time the data is to be kept in an identifiable form and the full contact details of the recipient. Once the data has been transmitted, the respondent may exercise his rights to the recipient for as long as his data is retained. The institute shall ensure the confidentiality of the interviewee's data before their dispatch, by agreement with the recipient. When the Institute uses subcontractors for the processing of personal data, it shall enter into a contract with them containing the standard contractual terms drawn up by the EU for this purpose to fully guarantee the rights of the respondents.

10.5 Confidentiality Agreements regarding the collection and processing of research data

The staff (Employees, Partners, and Technicians) of the Institute sign Confidentiality Agreement, under which they are bound by the data protection rules when collecting, processing, and analyzing the data of the respondents.

10.6 Data Protection Officer and Contact Details

Data Protection Officer (DPO) is Vasileios Paliouras, dpo@elgo.gr. The institute's offices are in Athens, Terma Alkmanos str., 11528 Ilisia. The contact number is 210 2756596

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, 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.

European Commission, 2021. Commission Implementing Decision (EU) 2021/1167: establishing the multiannual Union programme for the collection and management of biological, environmental, technical and socioeconomic data in the fisheries and aquaculture sectors from 2022. *Official Journal of the European Union*, C/2021/2797, 27 April 2021.

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.

Hollander, M., Wolfe, D.A. and Chicken, E. (2014). Nonparametric Statistical Methods, 3rd Edition, Wiley.

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 study No. FISH/2005/03

Little, R.J.A. and Rubin, D.B. (2002). Statistical analysis with missing data. (Second Edition.) New York: Wiley.

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

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

Fishing activity variables		Data Source	Data Collection
Capacity	Number of vessels	Fleet register	A
	GT	Fleet register	A
	kW	Fleet register	A
	Vessel Age	Fleet register	A
Effort	Days at sea	ERS/VMS/Questionnaire	A/B
	Fishing days	ERS/VMS/Questionnaire	A/B
	kW * Days at sea	Fleet register / Questionnaire	A/B
	GT * Days at sea	Fleet register / Questionnaire	A/B
	kW * Fishing Days	Fleet register / Questionnaire	A/B
	GT * Fishing days	Fleet register / Questionnaire	A/B
	Number of trips	ERS/VMS/Questionnaire	A/B
	Number of fishing operations	ERS/VMS/Questionnaire	A/B
	Length of nets (m) * soak time (days)	ERS/VMS/Questionnaire	A/B
	Number of nets/Length	ERS/VMS/Questionnaire	A/B
	Number of hooks, Number of lines	ERS/VMS/Questionnaire	A/B
	Numbers of pots, traps	ERS/VMS/Questionnaire	A/B
	Number of FADs/buoys	ERS/VMS/Questionnaire	A/B
	Number of support vessels	ERS/VMS/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
	Average price per species	Fleet register / Questionnaire	A/B

A: Census, B: Probability Sample Survey

Table A.2: The economic 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	NA	NA
	Operating Subsidies	Questionnaire	A
	Subsidies on Investment	Managing Authority of Fisheries and Maritime Operational Programme/ Questionnaire	A/B
	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	Other Variable costs	Questionnaire	B
	Other Non-variable costs	Questionnaire	B
	Lease/rental payments for quota or other fishing rights	NA	NA
Capital costs	Consumption of fixed capital	Fleet Register/ Questionnaire	C
Investments	Investments in tangible assets, net	Questionnaire	B
Financial position (assets and liabilities)	Gross Debt	Questionnaire	B
	Total assets	Questionnaire	B
	Value of physical capital	Fleet Register/ Questionnaire	C
	Value of quota and other fishing rights	NA	NA
Employment	Paid Labour	Questionnaire	B
	Unpaid labour	Questionnaire	B
	Full Time Equivalent	Questionnaire	B
	Total hours worked per year (optional)	Questionnaire	B
Fleet	Number of vessels	Fleet Register	A
	Mean LOA of vessels	Fleet Register	A
	Total vessel tonnage	Fleet Register	A
	Total vessel power	Fleet Register	A
	Mean age of vessels	Fleet Register	A
Effort	Days at sea	ERS/VMS/Questionnaire	A/B
	Energy consumption	ERS/VMS/Questionnaire	A/B
Number of fishing enterprises/units	Number of fishing enterprises/units	Fleet Register	A

A: Census, B: Probability Sample Survey, C: Indirect Survey, NA: Not Applicable.

Table A.3: The social 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

B: Probability Sample Survey