Italian sampling plan

for fisheries products landed from fishing vessels permitted to weigh on landing in accordance with Article 60(1) of the Control Regulation

Under

COUNCIL REGULATION (EC) No 1224/2009 establishing a Community control system for ensuring compliance with the rules of the common fisheries policy

COMMISSION IMPLEMENTING REGULATION (EU) No 404/2011 laying down detailed rules for the implementation of Council Regulation (EC) No 1224/2009

1. Introduction

The present document represents the national sampling plan which the Italian administration intends to submit to the Commission in accordance with the procedure referred to in Articles 60(1) of the Control Regulation (EC) No. 1224/2009 and in accordance with the risk-based methodology described in Annex XIX of the Commission Implementing Regulation (EU) No 404/2011.

Article 60(1) of the Control Regulation provides that Member States shall ensure that all fishery products are weighed on systems approved by the competent authorities unless it has adopted a sampling plan approved by the Commission and based on the risk-based methodology adopted by the Commission in accordance with the procedure referred to in Article 119.

Annex XIX of the Commission Implementing Regulation lays down the methodology for Member States to establish sampling plans for weighing of landings of fisheries products in accordance with Article 60(1) of the Control Regulation.

2. The population of interest

The population of interest is the Italian fleet operating in the Mediterranean Sea and the list is based on the Fleet Register kept at the Directorate-General of Fisheries and Aquaculture of the Ministry of Agricultural and Forestry Policies.

In 2010, the Italian Mediterranean fleet consisted of about 13 thousands vessels, of which around 9 thousands are classified in the segment of passive gears less than 12 meters (small scale fishery). The Italian Mediterranean fleet is characterised by a strong multi-specificity and multi-gear activity. The fishing sector appears highly fragmented in fourteen regions along the coast and there are many large structural and technical differences in vessels from different geographical areas. Landings from the Adriatic Sea (namely Puglia, Marche, Emilia, Veneto) and the from Sicily Channel account for almost two thirds of national production (Figure 1).

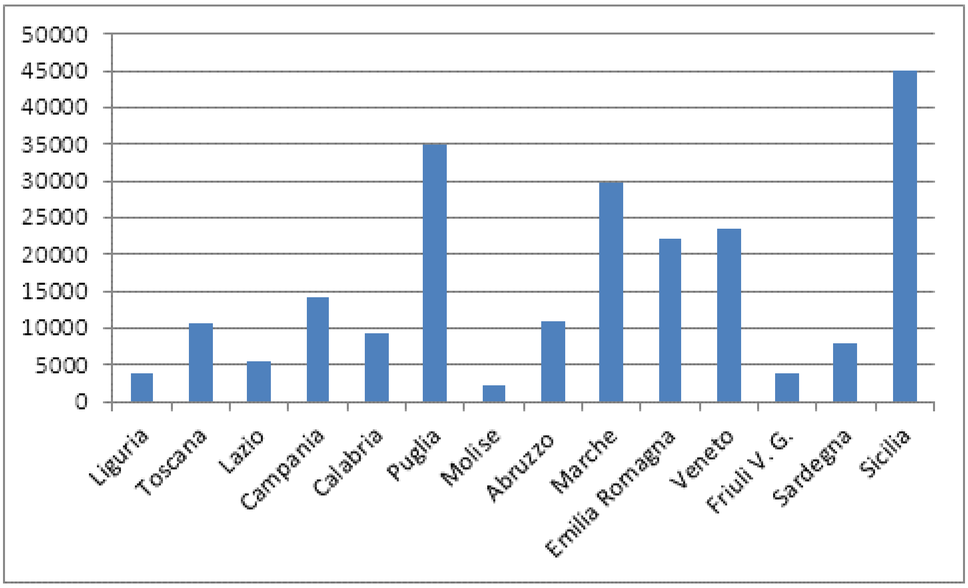


Figure 1. Landings by region, 2010 (Ton.)

The majority of vessels operates in coastal waters around the Italian peninsula. Over 24% of vessels is concentrated in the ports of Sicily, while another 13% of vessels is located in Puglia (Figure 2).

The fleet is classified into the following fleet segments: bottom trawlers, mid water pair trawlers, purse seiners, dredges, multi purposes vessels (using a combination of passive and mobile gears), small scale fishery and long liners.

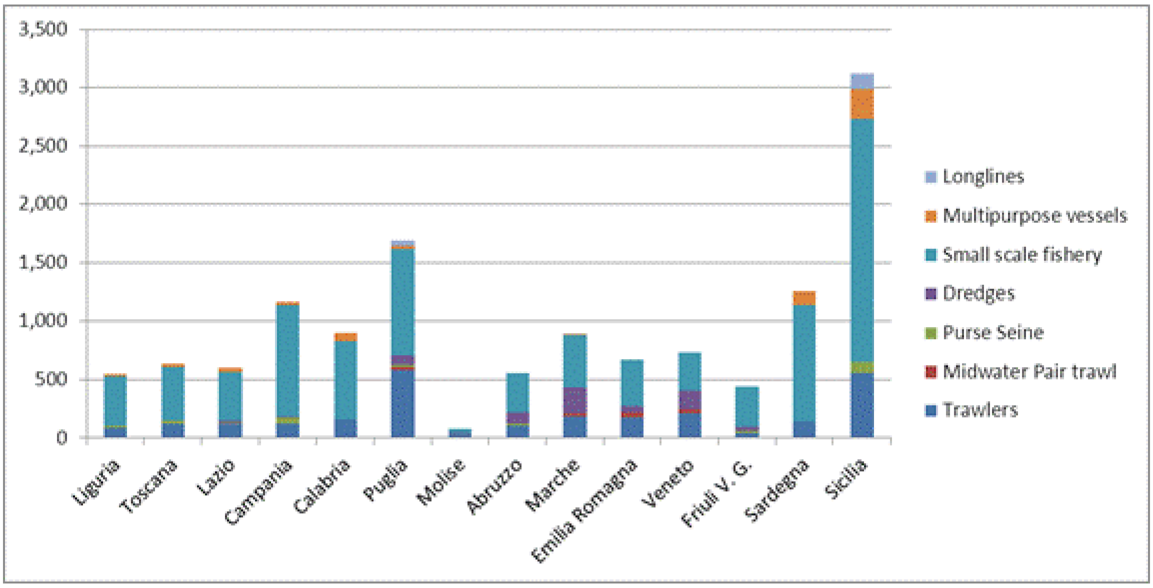


Figure 2. Number of vessels by fleet segment and regions

Table.2 gives an overview of the Italian Mediterranean fisheries covered by the sample plan.

The small-scale fishery (small fishing boats with an overall length of less than 12 meters using passive gears) is the most important fishery in terms of vessels' number, employment and activity. It accounts for a 66% of the total fleet in number and for about a quarter of the national value of landings. Fishermen represent 41% of total employment with an average crew of two men. The bulk of the catch consists of cuttlefish (11%), followed by common octopus (5.4%) and European hake (4.5%). In terms of value, cuttlefish represents 12.5% of total earnings, followed by red mullet (6%), European hake (5.8%) and sole (5.4%).

The trawling segment is the main fleet segment both in volume and value of its landings. In 2010, it produced 35% of total national landings and 51% of the total value of landings, employing around 9,075 fishermen (30% of total employment). This fleet, which mainly operates with otter trawls and beam trawls, is composed of 2636 trawlers and it accounts for a 20% of vessels. The main target species include European hake (11.5% of total volume), striped mullet (5.8%) and Norway lobster (4.1%), followed by musky octopus, cuttlefish and giant rose shrimp.

Pelagic fisheries are exclusively practiced by vessels authorized to mid-water pair trawl and purse seines. While purse seiners are concentrated in Tyrrhenian and Sicilian waters, mid-water pair trawlers fish exclusively in the Adriatic waters. This fleet accounts for a 30% of all landings in volume and 11% in value. The main target species include European anchovy and sardines. In 2010, European anchovy accounted for a 70% of landings (in volume) of mid-water pair trawl and for a 59% of landings of purse seine. Sardines represented around 20% of landings of both segments.

The other important fishery is represented by dredges (707 vessels in 2010), almost exclusively located in the central-north Adriatic coast. This fishery is highly specialised targeting mainly clams (Venus gallina), which account for 92% of their catch. In 2010, the production amounted to around 22,000 tonnes for an economic value of € 63 million.

The segment of multi-purpose vessels is composed of polyvalent vessels using passive gears (mainly nets) in combination with mobile gears (mainly trawls) according to season, demand and fishing grounds. In 2010, they accounted for a 4% of total fleet and represented around 3% of national landings in volume and value. In 2010, production by vessels with mixed passive gears totalled 8,426 tonnes for an economic value of € 66 million. The most significant species in terms of volumes landed were swordfish, albacore, hake and Atlantic bonito. Taken together, these products accounted for almost 40% of total catches and revenues.

The segment of longlines comprises many types of set and longlines used to catch different species, such as swordfishes, Bluefin tuna, albacore tuna and hakes. The production is concentrated in the Tyrrhenian littoral and particularly in Sicily, where there is the largest fleet. In 2010, the total volume of landings amounted to 5148 tonnes, around 3% of national landings for a total economic value of € 44 million. The most important species are swordfish (2,345 tonnes, 46% of the total), albacore (734 tonnes, 14% of the total) and hake (473 tonnes, 9% of the total). In 2010, 390 tonnes (amounting to € 4 million) of Bluefin tuna were caught by authorised longliners.

Table 1. Capacity and economic indicators by fleet segments, 2010

	Total fleet	Trawl	Mid. Pair trawl	Purse Seine	Dredges	Small scale	Multi purpose	Longlines
Volume of landings ('000ton)	223,007	78,182	44,393	31,506	21,794	33,559	8,426	5,148
Value of landings (EUR million)	1,103	555	47	53	63	276	66	44
Number of vessels	13,223	2,636	131	292	707	8,776	493	188
Total GT ('000)	176	110	10	18	9	17	7	6
Total kW ('000)	1,076	524	47	74	76	248	70	38
Employment	28,982	9,075	691	1,731	1,440	14,047	1,292	707

Source: MIPAAF - IREPA

3. The planning of the survey sampling design

According to point 1 of Annex XIX of the Commission Implementing Regulation (EU) No 404/2011, the aim of the sampling plans shall be to ensure accurate weighing of fisheries products on landing.

The sample unit is each Italian vessel selected from the Fleet Register operating in coastal waters around the Italian peninsula.

The variables of interest are the landings of fisheries products by metier.

The sampling is of a stratified nature in that the fishing vessels of the fleet are divided into homogenous groups based on four stratification variables (landings, LOA classes, geographical sub-regions, fleet segments) and independent samples are taken from each of these clusters.

The estimate of the sample size shall be determined on the basis of two approaches:

- the risk of non-compliance with the rules of the Common Fisheries Policy for the port/location/region in the Member State where landings take place in accordance with point 2 of Annex XIX
- the precision level required for the relevant estimates, which is expressed in terms of the coefficient of variation (CV= ratio between the standard error of the estimate and the estimate itself).

The coefficient of variation takes into account the level of landings by metier and region as it is also indicated in point 4 of Annex XIX. It will be included in the analysis through the Bethel's procedure (1989), which is a generalization of the method of Neyman (known as a method of univariate optimal allocation) and allows to determine total sample size and allocation of units in strata, under the constraints of defined precision levels of estimates (coefficient of variation in the multivariate case). Input to this algorithm is given by the information on distribution characteristics (total and variance) of target variable (landings) and other auxiliary information (prices, fishing days) in the population strata. For this reason it allows to include in the analysis also other relevant variables such as the price levels for the landed fisheries products and the metier activity.

The approach used by this method is to transform the analysis into a linear programming model that allows the identification of the sample size and the allocation across strata, minimising the variances of all variables simultaneously. In simple terms, the aim of Bethel's procedure is to ascertain the «minimum cost» of the sample, given the precision limits required for each stratum. The cost C is defined as:

$$C = c_0 + \sum_{h=1}^{H} c_h n_h \qquad \text{(eq.1)}$$

where c_0 represents a fixed cost correlated with the organisation of the collection of data, c_h represents the costs of the sampling of a unit within the stratum h-th (h= 1...H), while n_h represents the number of units selected from within the h-th stratum.

Given that the sampling is stratified, the precision limits on the estimate can be expressed as follows:

$$\text{var}\left(\hat{Y}_j\right) = \sum_{h=1}^{H} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{hj}^2}{n_h} \leq \tilde{v}_j^2 \qquad \text{(eq.2)}$$

j=1 … J

where $\left(\hat{Y}_j\right)$ represents the total for the j-th variable (j=1....J), S_{hj}^2 represents an estimate (or a hypothetical value) of the variance of the j-th variable within the h-th stratum and $\tilde{v}_j^2$ represents the threshold level (the limit), in absolute terms, for the value of the variance of the total estimator for the j-th variable.

As input data to start the procedure, the variance estimates for each stratum of landings by species and the estimates of the totals are needed. These estimates will be obtained from the data available, at the time of the analysis, for the most recent year. In particular, on the basis of the past experience developed in the data collection framework, the average value of coefficient of variation will not exceed 3.5% for the main species landed by the Italian fleet.

The methodological aspects of the Bethel's procedure are described in detail in Annex 1.

Table 2 below synthetizes how all relevant criteria laid down in Annex XIX of the Commission Implementing Regulation (EU) No 404/2011 will be implemented in the sample strategy. In accordance with point 5 in Annex XIX, the sampling plan shall be representative and at least as efficient as simple random sampling.

Table 2: Relevant criteria according to Article 4 of Annex XIX Reg 404/2011 and Sampling plan

RELEVANT CRITERIA	APPROACH
1. levels of landings	Bethel
2. level of previously detected infringements	Risk analysis
3. levels of inspection activity	Risk analysis
4. availability of quota	Not relevant
5. fluctuation of market price levels	Bethel
6. risk of fraud	Risk analysis

Total landings per stratum will be obtained by multiplying the total number of active vessels with the average daily catch per given stock:

$$\hat{Y}_{HT} = \sum_{i}^{n} w_i y_i \qquad \text{(eq.3)}$$

where $w_i = N/n$, and y_i are the observed values from the sample units.

More specifically, in order to estimate total landings per stratum, the Horvitz-Thomson estimator (1952) will be applied, which is a generalization of the equation 3 in the case of a stratified random sampling and in the hypothesis of extracting the sampling units with equal probability and without re-pooling:

$$\hat{Y}_{HT} = \sum_{i=1}^{n_h} w_{hi} y_{hi} \quad \text{(eq.4)}$$

with:

n_h sample size in the stratum h;

$w_{hi} = N_h/n_h$:

y_{hi}, sample data of the unit i in the stratum h.

4. Risk analysis

The risk analysis conducted includes consideration of previous history of infringement linked to landings and risk of fraud.

In accordance with point 4 in Annex XIX of Commission Implementing Regulation (EU) No 404/2011 and on the basis of available and most relevant data, three variables have been used to estimate the level of risk of non-compliance with the rules of the Common Fisheries Policy:

1. Number of infringements
2. Number of inspections
3. Level of priority of infringements linked to landings.

The level of risk was calculated as the product of the likelihood of a violation of the CFP rules and the potential impact of violations:

Level of risk = (*No. infringements /No inspections) * level of priority of each infringements*

The likelihood of a violation of the CFP rules was expressed as the ratio between the number of infringements detected by vessel to the number of physical inspections carried out at sea in 2010.

The potential impact of violations was assessed according to the estimated seriousness of the infringement according to article 90 of Control Regulation (EC) No. 1224/2009 and to article 3 of Council Regulation (EC) No 1005/2008.

The level of previously detected infringements, total number of inspections carried out by metier and background and/or potential risk of fraud can be considered consistent and exhaustive criteria to determine the level of risk of non-compliance of the fleet of interest. As outlined above, the other criteria set out in point 4 of Annex XIX for the identification of risk are less relevant in this case because:

- Fisheries products are not always landed in regulated standardized boxes
- No fishing quotas are allocated to Italian vessels: Bluefin tuna is the only stock managed by IQ in Mediterranean and all landings are already recorded on a separate base following the EU Recovery plan
- Fluctuation of market price levels for the landed fisheries products and levels of metier activity are already included in the Bethel's procedure and, hence, in the sampling plan.

Procedure for the Assessment of the level of risk of non-compliance with the rules of the Common Fisheries Policy

Data source

Physical inspections and infringements records in 2010 provided by the Italian Coast Guard have been used. This database was not properly designed for the purpose of the present analysis and therefore there is room for improvements. However, the Italian administration is committed to improve the database in next years, through the full implementation of the Control Regulations.

During 2010, Italian Coast Guard reported 24,657 inspections and 1,916 penal and administrative infringements. Data on inspections were available only at regional level (namely for 15 Maritime Units). Infringements records included different types of infringements by legal (entities) person, classified in naval units, individuals, corporate entities and unknown[1]. However, for only 563 infractions (almost 30% of total) it was indicated a valid registration number. Most part of infractions collected, in fact, did not report any registration number. While another 2% of infractions referred to vessels not present in the Fleet Register.

The number of inspections and infringements detected at sea by region are shown in Table 3. In accordance with point 8 in Annex XIX ("*Any risk analysis, data assessment, validation procedure, audit procedure, or other documents supporting the establishment, and further amendments, of the sampling plan shall be documented and made available for audits and inspection*"), the available data set is reported in Appendix I of this document. Appendix II reports records about 563 infractions used for the calculation of the likelihood of a violation by region and fleet segment.

Table 3: Number of inspections and infringements detected at sea by region in 2010

Maritime unit	REGION	No Inspections	No administrative infringements	No penal infringements	Total infringements
Genova	LIGURIA	1155	27	4	31
Livorno	TOSCANA	1463	65	7	72
Roma	LAZIO	1984	100	10	110
Napoli	CAMPANIA	1188	128	63	191
Reggio Calabria	CALABRIA	1869	265	97	362
Bari	PUGLIA	2237	160	46	206
Ancona	MARCHE	1991	157	0	157
Ravenna	EMILIA ROMAGNA	962	132	2	134
Venezia	VENETO	901	46	0	46
Trieste	FRIULI VENEZIA GIULIA	994	39	0	39
Catania	SICILIA	3684	166	33	199
Palermo	SICILIA	2069	182	15	197
Cagliari	SARDEGNA	1788	21	4	25
Pescara	ABRUZZO	1534	99	7	106
Olbia	Sardegna	838	18	23	41
Total		**24657**	**1605**	**311**	**1916**

Source: Italian Coast Guard

Likelihood of occurrence = No. infringements / No inspections

As previously stated, the first variable used to estimate the level of risk of non-compliance is the likelihood of occurrence, which was estimated as the ratio between the 563 infringements detected at sea (concerning vessels with a registration number) to the number of physical inspections carried out at sea in 2010 by administrative region. Figure 1 shows the likelihood of occurrence by region and types of infringements. According to this ratio, Lazio, Tyrrhenian Calabria, Ionic Calabria, Campania and Marche present the highest likelihoods of violations.

[1] Unknown means gears abandoned at sea.

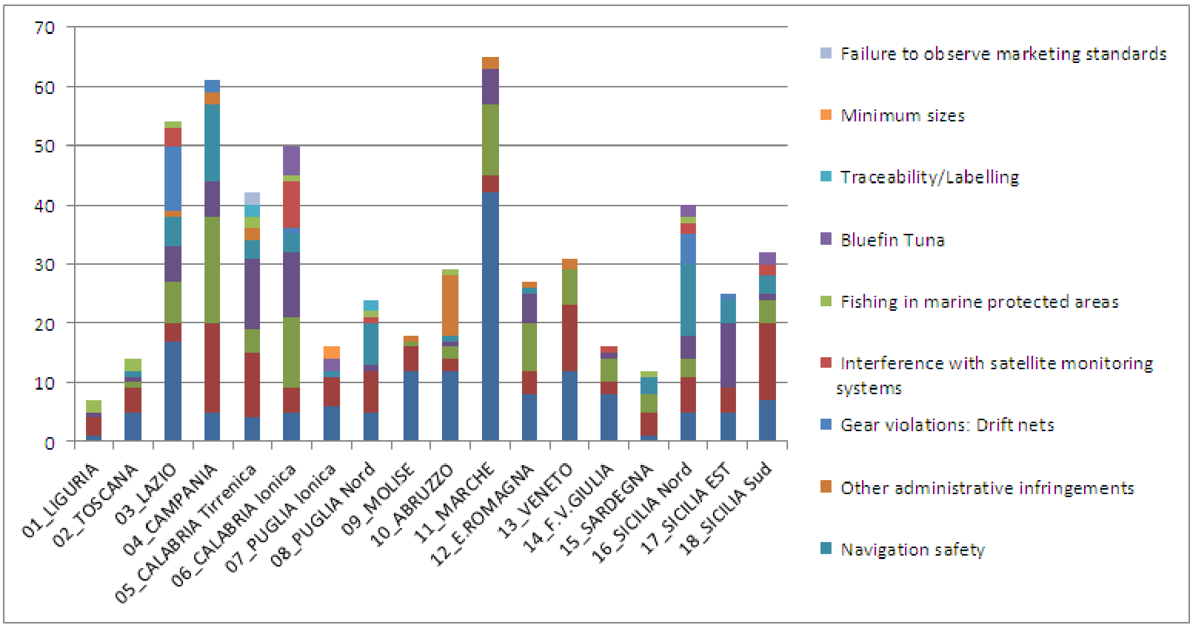


Figure 1: Likelihood of a violation of the CFP rules by region and type of infringements.

Similarly, Figure 2 shows the likelihood of occurrence by fleet segment and type of infringements. Bottom trawl presents the highest ratio, followed by small scale fishery and dredge. Detailed data and calculations by each sample stratum are available in Appendix III.

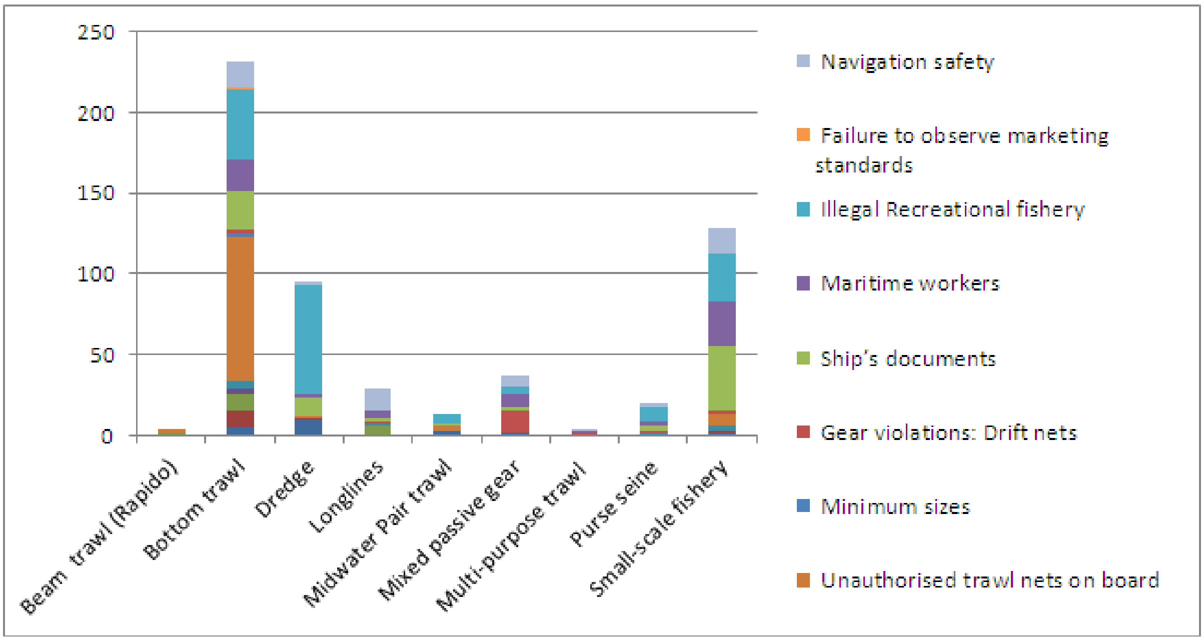


Figure 2: Likelihood of a violation of the CFP rules by fleet segment and type of infringements.

Potential impact of violations = level of priority of each infringements

The second variable used to estimate the level of risk of non-compliance is the potential impact of violations. It was calculated by assigning risk weights to each type of infringements.

Table 4 shows the scoring system used to assign a level of priority to each type of infringement. It was based on the elements of article 90 of Control Regulation (EC) No. 1224/2009 and article 3 of Council Regulation (EC) No 1005/2008 and basically reflects the nature of the damage, its value, the economic situation of the offender and the extent of the infringement.

The score varies between 2 and 7 on a scale of increasing severity according to the point system for serious infringements as referred to in Legislative Decree 11 November 2011 and in Article 42(1)(a) of Regulation (EC) No 1005/2008 (Table 4). In particular, infringements linked to landings, as those related to the minimum size or to marketing standards are ranked with the highest point.

Table 4: Type of infringement and level of priority

Infringement	letters art. 3 Reg 1005/2008	Point
Maritime workers		2
Navigation safety		2
Interference with satellite monitoring systems	b) not fulfilled obligations to transmit data by satellite vessel monitoring system)	3
Gear violations: Drift nets	e) used prohibited or non-compliant fishing gear	4
Unauthorised trawl nets on board	e)	4
Ship's documents	g) concealed, tampered with or disposed of evidence relating to an investigation	5
Traceability/Labelling	d) i)	7
Other administrative infringements	c) fished in a closed area, during a closed season, without or after attainment of a quota or beyond a closed depth	6
Fishing in marine protected areas	c)	6
Illegal Recreational fishery	a) fished without a valid licence, authorisation or permit issued by the flag State or the relevant coastal State	7
Failure to observe marketing standards	i)	7
Bluefin Tuna	d) engaged in directed fishing for a stock which is subject to a moratorium or for which fishing is prohibited	7
Minimum sizes	i) taken on board, transhipped or landed undersized fish in contravention of the legislation in force	7

*Quantitative assessment of level of risk = (No. infringements /No inspections) * level of priority of each infringements*

The potential impact of violations by administrative region was calculated as the product of likelihood of occurrence and the corresponding level of priority. Appendix IV reports the level of risk and its calculation by individual stratum (region, segment, LOA class). Appendix V shows the total assessment of the level of risk by strata. On average, the Italian Mediterranean fleet presents a level of risk equal to 1.05 and it varies between 0.03 and 11.05.

The level of risk was measured in terms of percentage deviation from the mean (1.05) as shown in table 5 below. It varies between 1 (very low) and 5 (very high) on a scale of increasing gravity as provided in accordance with point 4 in Annex XIX.

Table 5: Level of risk and deviation from the mean

Deviation from the mean (Xi)	Level of risk	Point
Xi=<-75%	Very low	1
-25% =<Xi<-74%	Low	2
-25% <Xi<=+25%	Medium	3
25% <Xi=<75%	high	4
Xi>75%	Very high	5

Around 25% of sample strata presents a very low risk level, while another 39% a low level. A 14% of strata has a level of risk around the total average. A 6% presents a high level of risk and the remaining 17% a very high level. At regional level (Figure 3), highest level of risk are concentrated in Lazio, Campania, Calabria, Marche and Veneto. Bottom trawl followed by small scale fishery and dredge present large number of vessels with a very high risk of non compliance with the CFP rules. (Figure 4)

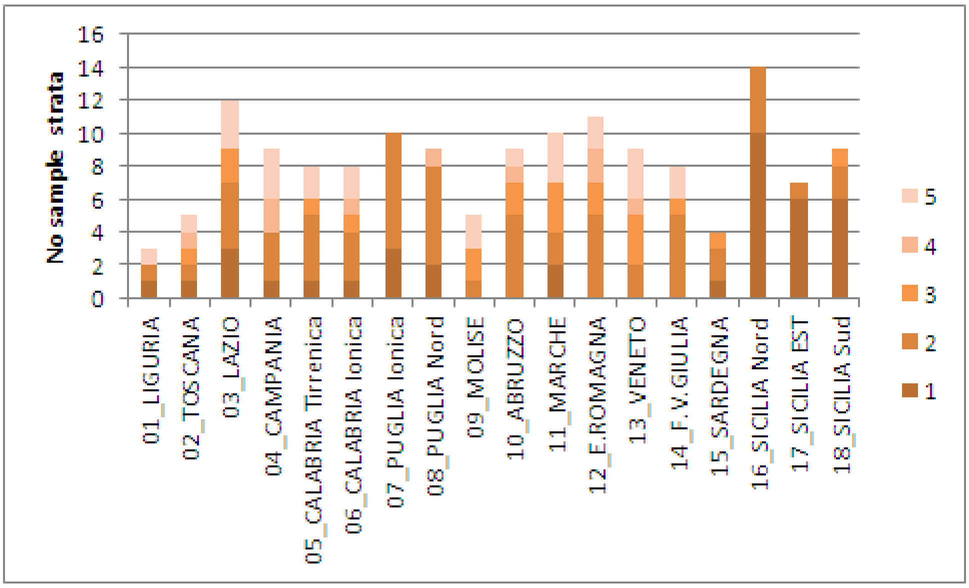


Figure 3: Level of risk by region

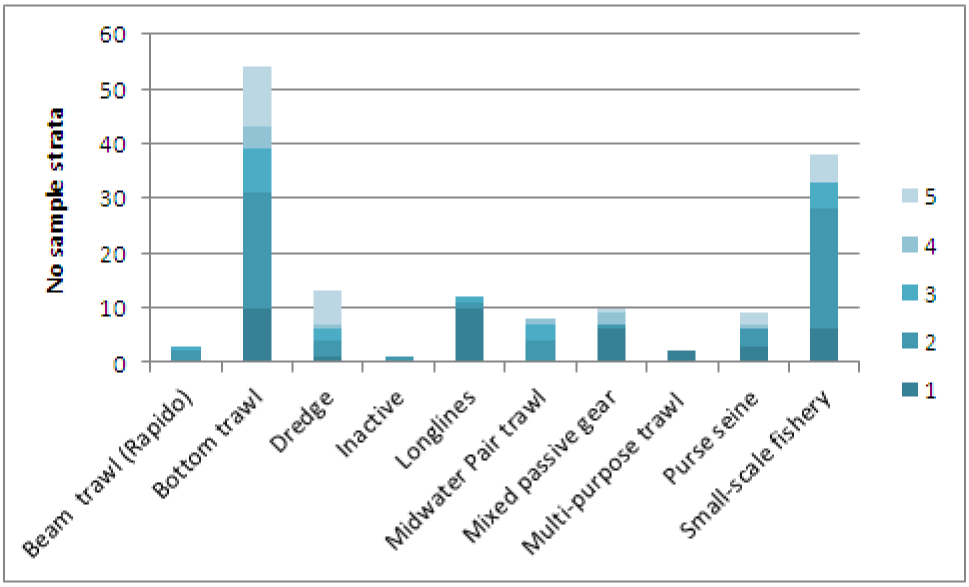


Figure 4: Level of risk by fleet segment

5. Integration between the risk analysis and the Bethel's procedure for the estimation of the sample size

As outlined in paragraph 3, in order to take into account other qualitative and quantitative information related to the social and economic importance of the fish production at regional level, the first quantitative assessment of the level of risk is further compared with sample size (Bethel's allocation) as estimated in the Bethel's procedure.

The result is a second estimate of the sample size, where the coverage of the population of strata with high and very high levels of risk have been increased respectively to 50% and 75% in order to take into account such risk. Appendix VI presents the proposed sample dimension by sample stratum. The total sample size (1521 vessels) corresponds to the 11% of the total fleet. The 8 Clustering of segments and the proposed sample dimension are also reported in the Annex 2 of this document.

6. Conclusions

The probability sample survey carried out to estimate fisheries products landed from Italian fishing vessels is a multivariate sample survey. The sample unit is the single vessel and this unit is selected from the Vessel Register. The sampling is of a stratified nature in that the fishing vessels of the fleet are divided into homogenous groups based on suitable variables and independent samples are taken from each of these clusters.

The optimum sample number per stratum is defined according to Bethel's procedure (1989), the vessels are selected using PPS methodology (Probability Proportional to Size) and, to be more exact, using the algorithm of Hanurav-Vijayan. In each of these phases the data is elaborated using the R software language.

In order to obtain the optimum sample number per stratum, and to meet the requirements of EC Regulation 404/2011, the following stratification criteria were used:

1. Geographical area of registration of vessels

2. Segmentation technique based on the most used gear

3. Size of vessels

4. Level of risk of non-compliance with the rules of the Common Fisheries Policy

In accordance with point 7, second paragraph, of the Annex XIX the figure resulting from the weighing shall be used for the completion of landing declarations, transport document, sales notes and take-over declarations.

Furthermore, in order to ensure that operators comply with the established sampling levels (point 7, first paragraph in Annex XIX of EC Regulation 404/2011), the weighing of vessels of less than 10 metres' length overall (N. 8075) not subject to fishing logbook requirements and landing declaration requirements will be cross checked with data resulting from the sample system that the Italian administration has previously requested in accordance with Articles 16 , 25 and 65 of Regulation (EC) No. 1224/2009.

The weighing of vessels over 10 metres' length overall (N. 4949) will be cross checked with data from log books, sales notes and from the sample survey.

Finally, in accordance with point 7 third paragraph of the Annex XIX, the Italian Administration establishes that at least a minimum of 10% (n. 151) of the sample landings described in Appendix IV for each month will be weighed in the presence of officials of the local Coast Guard. The Italian Administration is committed to improve the quality and the quantity of controls in next years, once all requirements related to the Control Regulation have been complied.

REFERENCES

Bethel, J. (1989). Sample allocation in multivariate surveys. *Survey Methodology* 15, 47-57.

Cochran W.G. (1977), "*Sampling Techniques*", 3rd ed., Wiley, New York

EUROSTAT (2008), Survey sampling reference guidelines. Introduction to sample design and estimation techniques. Methodologies and Working papers

Hanurav, T.V. (1967). Optimum utilization of auxiliary information: π ps sampling of two units from a stratum. *Journal of the Royal Statistical Society* B 29,374-391.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47,* 663-685.

Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics 5*, 119-127.

Vijayan, K. (1968). An exact π ps sampling scheme: generalization of a method of Hanurav. *Journal of the Royal Statistical Society B 30,* 556-566.

Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society B 15,* 253-261.

Annex 1. Bethel's procedure

Bethel's procedure (1989) is a mathematical algorithm to achieve the optimum sample allocation in a multivariate sample survey, that is to say the study of several subject variables which are also stratified.

The aim of Bethel's procedure is to ascertain the «minimum cost» of the sample, given the precision limits required for each stratum. The cost C is defined as:

$$C = c_0 + \sum_{h=1}^{H} c_h n_h \qquad (*)$$

where c_0 represents a fixed cost correlated with the organisation of the collection of data, c_h represents the costs of the sampling of a unit within the stratum h-th (h= 1...H), while n_h represents the number of units selected from within the h-th stratum.

Given that the sampling is stratified, the precision limits on the estimate can be expressed as follows[2]:

$$\mathrm{var}\left(\hat{Y}_j\right) = \sum_{h=1}^{H} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{hj}^2}{n_h} \le \tilde{v}_j^2 \qquad j=1 \ldots J \qquad (**)$$

where $\hat{Y}_j$ represents the total for the j-th variable (j=1....J), S_{hj}^2 represents an estimate (or a hypothetical value) of the variance of the j-th variable within the h-th stratum and $\tilde{v}_j^2$ represents the threshold level (the limit), in absolute terms, for the value of the variance of the total estimator for the j-th variable.

This set of limited J can be equivalently expressed in an alternative form:

$$\sum_{h=1}^{H} N_h^2 \frac{S_{hj}^2}{n_h} \le \tilde{v}_j^2 + \sum_{h=1}^{H} N_h S_{hj}^2 \iff \frac{\sum_{h=1}^{H} N_h^2 \frac{S_{hj}^2}{n_h}}{\tilde{v}_j^2 + \sum_{h=1}^{H} N_h S_{hj}^2} \le 1 \iff \sum_{h=1}^{H} \left(\frac{N_h^2 S_{hj}^2}{\varepsilon_j^2 Y_j^2 + \sum_{h'=1}^{H} N_{h'} S_{h'j}^2} \cdot \frac{1}{n_h} \right) \le 1$$

where $\hat{Y}_j$ represents the total estimated (or hypothesized) for the variable j-th, and ε_j represents the relative error (absolute error of the estimate divided by the value of the estimate) acceptable for the j-th variable.

Thus using a_{hj} to indicate the term on the left of the product in brackets of the last inequality, with x_h at a value of $1 / n_h$, all the last inequality can be expressed in the form:

$a_j' \underline{x} \le 1$ j=1 … J

or, equivalently,

---

[2] In Bethel's original article, the correction for finite populations was not considered, and therefore (considering the differences due to the fact that in the article the quantities to be estimated were averages and not totals) the formula (*) in in fact presented as: $\mathrm{var}\left(\hat{Y}_j\right) = \sum_{h=1}^{H} N_h^2 \frac{S_{hj}^2}{n_h} \le \tilde{v}_j^2$

A'$\underline{x} \leq \underline{1}$

dove $\underset{(H \times J)}{A} = \lfloor a_{hj} \rfloor$ e $\underset{(H \times 1)}{\underline{x}}$ the vector of values 1/n_h.

The whole problem of the minimum limit can be expressed in the following way:

$$\begin{cases} \min_{\underline{x}} g(\underline{x}) = \sum_{h=1}^{H} \frac{c_h}{x_h} \\ s.v.: A'\underline{x} \leq \underline{1} \end{cases}$$

Bethel demonstrated that this problem always has a solution, and that this corresponds to the following formula:

$$x_h{}^* = \frac{\sqrt{c_h}}{\sqrt{\sum_{j=1}^{J} \alpha_j{}^* \, a_{hj}} \cdot \sum_{k=1}^{H} \sqrt{c_k \sum_{j=1}^{J} \alpha_j{}^* \cdot a_{kj}}}$$

Where the α_j* are suitable normalised constants (Lagrange multipliers), that is to say those for which $\sum_{j=1}^{J} \alpha_j{}^* = 1$.

The solution of the problem of the minimum: Chromy's algorithm

To solve the problem of the minimum limit, Bethel proposes the use of an algorithm which is neither particularly efficient nor easy to apply. At that time, in fact, another algorithm was already available, formulated by Chromy (1987) and also put forward in the same publication of Bethel, which made it easier to find a solution to the problem from the point of view of the development of the code and quicker in terms of elaboration time.

Once the initial values of α_j ,equivalent to1/J, are in place, this algorithm develops fundamentally in two steps, which are repeated continually until reaching an acceptable criteria of convergence.

1. Calculate: $x_h\left(\underline{\alpha}^{(r-1)}\right) = \frac{\sqrt{c_h}}{\sqrt{\sum_{j=1}^{J} \alpha_j{}^{(r-1)} a_{hj}} \cdot \sum_{k=1}^{H} \sqrt{c_k \sum_{j=1}^{J} \alpha_j{}^{(r-1)} \cdot a_{kj}}}$

2. Calculate: $\alpha_j{}^{(r)} = \frac{\alpha_j{}^{(r-1)} \left[a_j' \underline{x}\left(\underline{\alpha}^{(r-1)}\right)\right]^2}{\sum_{k=1}^{J} \alpha_k{}^{(r-1)} \left[a_k' \underline{x}\left(\underline{\alpha}^{(r-1)}\right)\right]^2}$

The estimate of the average , in analogy with the estimate of the totals, will be given by $\hat{\bar{Y}}_h = \frac{1}{N_h} \sum_{i=1}^{n_h} \frac{y_{hi}}{\pi_{(h)i}}$, for the single stratum h, by $\hat{\bar{Y}} = \frac{1}{N} \sum_{h=1}^{H} N_h \hat{\bar{Y}}_h = \frac{1}{N} \sum_{h=1}^{H} \sum_{i=1}^{n_h} \frac{y_{hi}}{\pi_{(h)i}} = \frac{\hat{Y}}{N}$ for the total of the variable Y.

For the estimate of the variance of the total the Sen-Yates-Grundy formula (1953) is used:

$\hat{\sigma}^2(\hat{Y}_h) = \sum_{i=1}^{n_h}\sum_{j>i}^{n_h} \left(\frac{\pi_{(h)i}\pi_{(h)j}}{\pi_{(h)ij}} - 1\right)\left(\frac{y_{hi}}{\pi_{(h)i}} - \frac{y_{hj}}{\pi_{(h)j}}\right)^2$,for the single stratum h, while having obtained the sample of H independent selection in each stratum, the total variance is obtained from the sum of the variances of (from within) each single stratum:

$$\hat{\sigma}^2(\hat{Y}) = \sum_{h=1}^{H}\hat{\sigma}^2(\hat{Y}_h) = \sum_{h=1}^{H}\ \sum_{i=1}^{n_h}\sum_{j>i}^{n_h} \left(\frac{\pi_{(h)i}\pi_{(h)j}}{\pi_{(h)ij}} - 1\right)\left(\frac{y_{hi}}{\pi_{(h)i}} - \frac{y_{hj}}{\pi_{(h)j}}\right)^2 .$$

The relationship between the estimate of the standard deviation of the total and the estimate of the total itself, provides the estimate of the sampling error committed ($\hat{\sigma}(\hat{Y}_h)/\hat{Y}_h$ or $\hat{\sigma}(\hat{Y})/\hat{Y}$ depending on whether or not reference is made to the single stratum).

For the estimate of the variance of the population relative to each stratum the formula of Chaudhuri is used:

$$\hat{\sigma}_h^2 = \frac{1}{2N_h^{\ 2}}\sum_{i\neq j}^{n_h}\frac{\left(y_{hi} - y_{hj}\right)^2}{\pi_{(h)ij}} = \frac{1}{N^2}\sum_{i=1}^{n_h}\sum_{j>i}^{n_h}\frac{\left(y_{hi} - y_{hj}\right)^2}{\pi_{(h)ij}} .$$

This last value can be used as an input parameter for the procedure of Bethel

The sample selection procedure: the algorithm of Hanurav(1967)-Vijayan (1968)

The sampling design adopted requires the extraction, without repetition, of the sampling units based on the PPS (probability proportional to size) method; in simpler terms, this sampling plan involves the extraction of various units with a first-order inclusion probability which is not constant, but is proportional to a suitably selected auxiliary variable (Cochran, 1977). The use of such a sampling plan, and thus its use in place of simple random sampling, is justified by the intention of wanting to exploit the information given by the auxiliary variable. This auxiliary variable obviously must be noted for all units in the reference population, and must be «linked» to the unknown variable, the estimate of which is being attempted. This link, in statistical terms, is translated in «proportional relation» between the variable to be estimated and the noted auxiliary variable. The use of information supplied by the auxiliary variable aims to improve the estimate; put in other words, the «stronger» this proportional relation is, the smaller the variability of the estimator (or variance), and so the estimate is much more precise. In the theoretical situation limit of exact proportionality, the estimator would have zero variance and would assume ,in any sample, the exact total to estimate. In the case considered, the noted auxiliary variable is the LOA, the use of which as an accessory variable was preceded by an exploratory analysis, which confirmed the hypothesis of proportionality between the LOA on the one hand, and the quantity fished and revenue on the other (this obviously does not refer to an «exact» relationship between the variables).

The algorithm of Hanurav-Vijayan defines a series of steps to be taken to select a sample of a pre-defined number (n), without replacement, and with a non-uniform probability of each individual unit being included in the sample. By following this algorithm, a sample is obtained which has a series of properties, some of which are worthy of note:

i. $\pi_i = n X_i / X$, where π_i represents the inclusion probability (also called probability of inclusion of the first order) of the i-th unit, n indicates the pre-determined size of the sample, Xi represents the size of the noted variable (or «accessory» measure) from which the inclusion probability is calculated and X is the sum of the values X_i for i=1…N, where N is used to denote the size of the universe being sampled. This identity is «required for construction» and necessitates some special treatment in specific circumstances (considered further on).

ii. $\pi_{ij}>0$, where π_{ij} represents the probability (called of the second order) of the simultaneous presence of units i and j. The very fact of being able to determine these probabilities exactly and relatively simply, a consequence of the sampling procedure, is already a notable result which assures the existence of an unbiased estimate of the variance.

iii. $\pi_{ij} \leq \pi_i \pi_j$. This characteristic is notable because it guarantees a positive Sean-Yates-Grundy estimator of the variance of the total

iv. $\pi_{ij} - \pi_i \pi_j > \beta$, for β nor too close to 0. This property guarantees the stability of the Sean-Yates-Grundy total variance estimator.

The values π_i and π_{ij} (for i,j=1…N) satisfy the following two properties:

$$\sum_{i=1}^{N} \pi_i = n$$

$$\sum_{i=1}^{N} \sum_{j>i}^{N} \pi_{ij} = \frac{n\,(n-1)}{2}$$

It is interesting to note that the sum of the probabilities of the first order never equals 1 (unless the sample is composed of only a single unit).The same can be said for probabilities of the second order (unless the sample is composed of only 2 units).It is also to be noted how the application of the formula (i) can sometimes cause the inclusion probability of the first order to be more than 1. In this case corrections in the procedure of sample selection and the probabilities of inclusion must be applied. Specifically, the inclusion probability of the first order is assigned equal to 1, to the k units of which the probability results more than 1, and the n-k units within the entire population are selected, once the unit with the probability of 1 is excluded. It is clear that, once the units with a probability greater than 1 are «set aside» (or rather, selected with a probability of 1), should others with a inclusion probability greater than 1 appear within the remaining N-k, a gradual «setting aside» of these must be provided for, as for all other units, until a population of units with all the probabilities of being selected randomly in the first order inferior to 1, is obtained. Finally a sample is selected of (n-h) units among the (N-h) units of the entire population (where h (≤n) represents the number of units «set aside» or «pre-sampled»).

The sample selection procedure

For the description of the procedure, focus will be placed on the simple random sample. For a stratified sampling it is sufficient to apply the following procedure to every population sub-set.

It is thus presumed, without loss of generality, to have a population composed of N units, pre-ordered with respect to an accessory measure X_j (j=1….N) (in our case this measure was given by the value of the LOA). Thus $X_1 \leq X_2 \leq \dots \leq X_N$. is obtained.

The following steps are then followed:

a whole number between 1 and n is chosen randomly with a probability

1. $\theta_i = n\,(p_{N-n+i+1} - p_{N-n+i})\,(S + i\,p_{N-n+1}) / S$ (i=1 … n)

   where $p_j = X_j / X$, $S = \sum_{j=1}^{N-n} p_j$ and $p_{N+1} = 1/n$ to ensure that $\sum_{i=1}^{n} \theta_i = 1$

2. If at step (1.) the value i is selected, the last (n-i) elements of the population are selected and the next step is used to obtain the remaining i.
3. New normalized measures are defined in place of p_j , which are then indicated as p_j^*:

$$p_j^*\,(i) = \begin{cases} \dfrac{p_j}{S + i\,p_{N-n+1}} & se\ j \leq N-n+1 \\ \dfrac{p_{N-n+1}}{S + i\,p_{N-n+1}} & se\ N-n+1 < j \leq N-n+i \end{cases}$$

The missing units are selected, in order, using for each selection (indicated by 1 (1=0…i-1), probability values proportional to $a_j(l, j_{l-1})$, where:

$$a_j(l, j_{l-1}) = \begin{cases} (i-l)\ p_j^*\ (i) & \text{se } j = j_{l-1}+1 \\ (i-l)\ p_j^*\ (i) \prod_{k=j_{l-1}+1}^{j-1} \left[1-(i-l-1)\ P_k^*\ (i)\right] & \text{se } j = j_{l-1}+2 \ldots N-n+l+1 \end{cases}$$

where j_l represents the position of the unit selected in the 1-th selection (1=0…i-1) and P_k*(i) is calculated as:

$$P_k{}^*(i) = p_k{}^*\ (i) \Big/ \sum_{h=k+1}^{N-n+i} p_k{}^*\ (i)$$

The inclusion probabilities

For construction, the first-order inclusion probability π_j for the randomly sampled units (thus excluding the «pre-selected» units with a inclusion probability greater than 1) are equal to:

$\pi_j = np_j$ (taking the number of «pre-selected» units to be equal to 0).

The probability of inclusion in the second order is instead equal to:

$$\pi_{ij} = \sum_{r=1}^{n} \theta_r K_{ij}{}^{(r)}$$

Where

$$K_{ij}{}^{(r)} = \begin{cases} 1 & \text{se } N-n+r < i \le N-1 \\ \dfrac{r\ p_{N-n+1}}{S+r\ p_{N-n+1}} & \text{se } N-n < i \le N-n+r \text{ e } j > N-n+r \\ \dfrac{r\ p_i}{S+r\ p_{N-n+1}} & \text{se } 0 < i \le N-n \text{ e } j > N-n+r \\ \pi_{ij}{}^{(r)} & se \ \ j \le N-n+r \end{cases}$$

and where

$$\pi_{ij}{}^{(r)} = n\,(n-1) P_i{}^*(r) \prod_{k=1}^{r-1} \left[1-P_k{}^*(r)\right] .$$

Annex 2. Clustering of segments

The following table reports the segments that have been clustered. Clusters are named after the biggest segment in terms of number of vessels.

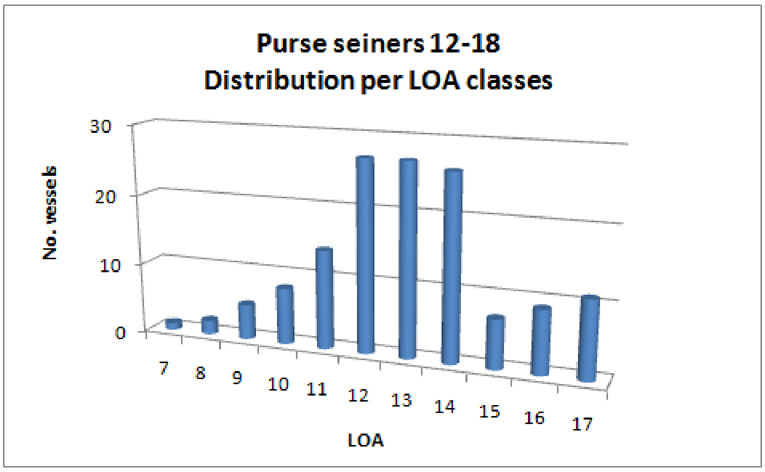
Name of the clustered fleet segments	LOA class	Population	Sample dimension
Beam trawl (Rapido)	2.4_LFT =>12 <18	12	4
	3.1_LFT => 18 <24	27	6
	4.1_LFT => 24 < 40	31	12
Bottom trawl	1.2_LFT=>6 <10	120	47
	2.1_LFT =>10 <12	137	29
	2.4_LFT =>12 <18	1418	198
	3.1_LFT => 18 <24	724	174
	4.1_LFT => 24 < 40	227	94
Dredge	1.2_LFT=>6 <10	4	4
	2.4_LFT =>12 <18	704	77
Longlines	2.4_LFT =>12 <18	138	28
	3.1_LFT => 18 <24	48	12
Mid water pair trawl	2.4_LFT =>12 <18	26	8
	3.1_LFT => 18 <24	44	12
	4.1_LFT => 24 < 40	80	18
Mixed gears	2.1_LFT =>10 <12	15	4
	2.4_LFT =>12 <18	37	4
Mixed passive gear	2.4_LFT =>12 <18	447	71
Purse seine	1.2_LFT=>6 <10	3	3
	2.4_LFT =>12 <18	125	37
	3.1_LFT => 18 <24	41	16
	4.1_LFT => 24 < 40	51	20
Small-scale fishery	1.1_LFT < 6	2852	191
	1.2_LFT=>6 <10	5190	357
	2.1_LFT =>10 <12	801	95
		13302	1521

Clustering is necessary in order to design the sampling plan and to report economic variables. The economic sample is stratified by segments according to Appendix III and by geographical sub areas (GSA).
This double level of stratification of the population (technical and geographical) may generate very small strata that have to be grouped in order to get a statistical sample. When a strata is too small (less than 10 vessels) it is very difficult to randomly select a sample. At the same time, the sampling plan is subject to budget constraints and clustering of small segments is also necessary to reach cost efficiency.
The proposed clustering also guarantees continuity in the time series.
In the following section, the scientific evidence justifying the clustering is explained for each clustering reported in the table below.

Purse seiners 12-18 m*

Name of the clustered fleet segments	*Total number of vessels in the cluster*	*Fleet segments which have been clustered*	*No. Of vessels*
purse seiners 12-18 m	125	purse seiners >-12<18 m	113
		purse seiners >-6<12 m	12

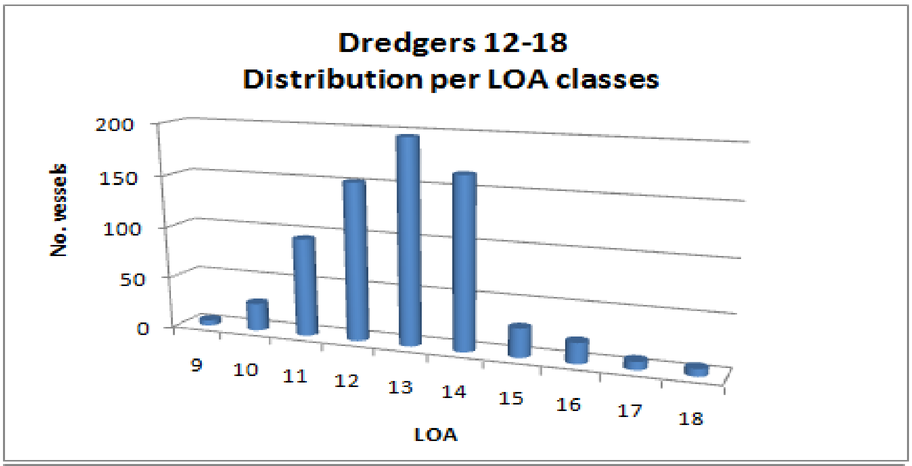
The clustered segment (purse seiners 12-18 m) is composed by 125 vessels with an average LOA of 13.5 m. Vessels are concentrated near the average value, as shown by the graph. Therefore the clustered segment is homogenous from a statistical point of view.

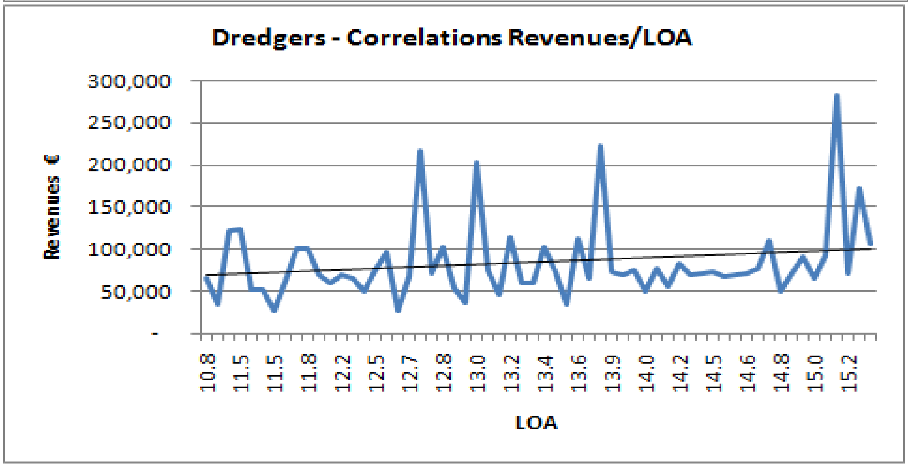


Dredgers >-12<18 m*

Name of the clustered fleet segments	*Total number of vessels in the cluster*	*Fleet segments which have been clustered*	*No. Of vessels*
dredgers 12<18 m*	704	dredgers >-6<12 m	127
		dredgers >-12<18 m	570
		dredgers >-18<24 m	7

Dredgers are based almost exclusively in central-north Adriatic cost. Vessels are very specialised targeting only clams and smooth-callista (Venus gallina and Callistachione) and they are homogenous in terms of size, gears and fishing practises. Vessels have an average LOA of 13.4 meters and 81% of them belong to the class 12-18 meters. Therefore, the split into the class <12 m and > 12 meters is not statistical reliable for this segment. Moreover, it is demonstrated that revenues are not correlated with the LOA of the vessels (see graph) and this fact proves the high level of homogeneity of the vessels.

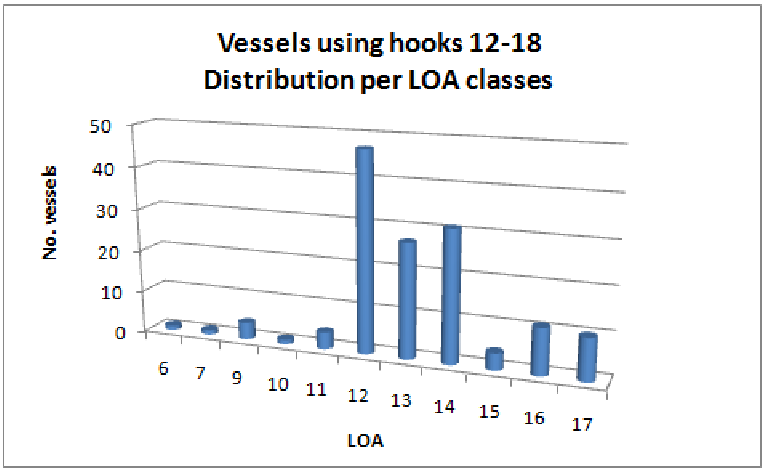




Vessels using hooks 12-18 m *

Name of the clustered fleet segments	*Total number of vessels in the cluster*	*Fleet segments which have been clustered*	*No. Of vessels*
vessels using hooks 12-18 m *	138	vessels using hooks >-6<12 m	11
		vessels using hooks >-12<18 m	127

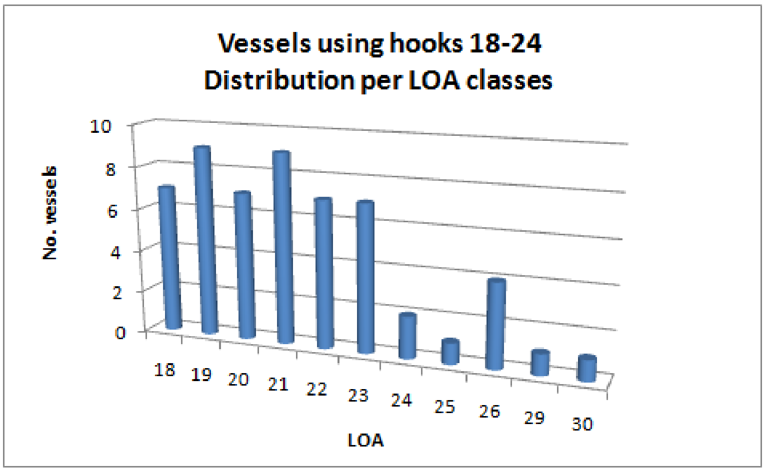
This is composed by 11 vessels and they operate in different areas (GSA 10, GSA 16, GSA 18, GSA 19). The sampling plan is stratified by area, therefore the rule that allows the clustering of segments with less than 10 vessels, is applied in each GSA. The distribution of the vessels per LOA classes shows a concentration around the average value of 13.8.



Vessels using hooks >-18<24 m*

Name of the clustered fleet segments	*Total number of vessels in the cluster*	*Fleet segments which have been clustered*	*No. Of vessels*
vessels using hooks 18-24 m*	48	vessels using hooks 18-24 m	39
		vessels using hooks 24-40 m	9

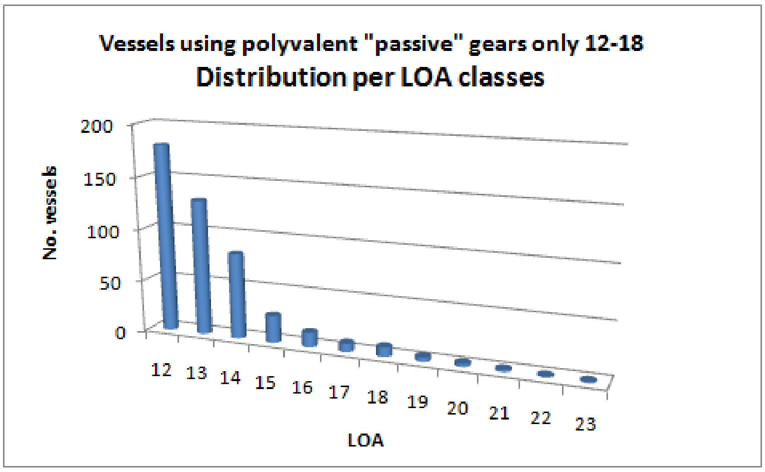
The class 24-40 m is composed by 9 vessels, 5 of which operate in GSA 19, 3 in GSA 16 and 1 in GSA 10. The average LOA of these vessels is 26.7 metres, therefore they are quite homogenous in terms of size with the vessels in the LOA class 18-24 m (whose average length is 21 meters). All these vessels operate in the same way, exploiting the same fishing grounds and targeting the same species.



Vessels using polyvalent "passive" gears only >-12<18 m*

Name of the clustered fleet segments	*Total number of vessels in the cluster*	*Fleet segments which have been clustered*	*No. Of vessels*
vessels using polyvalent "passive" gears only >-12<18 m*	447	vessels using polyvalent "passive" gears only >-12<18 m	427
		vessels using polyvalent "passive" gears only >-18<24 m	20

The clustering of the 20 vessels >18 m into the lower LOA class is necessary in order to design the sampling plan. In fact these vessels are scattered along the Italian coast (GSA 9, GSA 10, GSA 11, GSA 17, and GSA 19). The average length of these 20 vessels is 19.6 meters, therefore very close to the upper limit of the 12-18 m class.
The graph shows the high concentration of the vessels on the left side of the distribution.



<u>Vessels using active and passive gears >-12<18 m*</u>

Name of the clustered fleet segments	*Total number of vessels in the cluster*	*Fleet segments which have been clustered*	*No. Of vessels*
vessels using active and passive gears >-12<18 m*	37	vessels using active and passive gears >-12<18 m	36
		vessels using active and passive gears >-18<24 m	1

There is only one vessel in the class 18-24 meters. It is obvious that this vessel is grouped in the neighbouring class because it is impossible to get a statistical random sample from a stratum of only one vessel.