Review Article

Mediterranean Marine Science
Volume 8/2, 2007, 65-82

Optimal sampling designs for large-scale fishery sample surveys in Greece

G. BAZIGOS[1] and S. KAVADAS[2]

[1] 20 Tsakalof street, Halandri-Patima 152 35, Athens, Greece
[2] Hellenic Centre for Marine Research, Institute of Marine Biological Resources, Agios Kosmas, Helliniko 16610, Greece

e-mail: stefanos@ath.hcmr.gr

Abstract

The paper presents the quality problem of fishery statistics produced by the conducted land-based and sea-going, large scale sample surveys of the survey programme of the Institute of Marine Biological Resources of the Hellenic Centre for Marine Research (IMBR/HCMR) in Greece, and the optimality strategies developed in their sampling designs for the maximization of precision of the calculated sample estimates for a given cost of sampling.

The optimality problems of the sampling designs of the individual large scale sample surveys are described, and the optimality solutions developed under the sampling variance structure are explained.

The paper deals with the optimization of the following three large scale sample surveys: biological sample survey of commercial landings (BSCL), experimental fishing sample survey (EFSS), and commercial landings and effort sample survey (CLES).

Keywords: Probability sampling; Optimization; Efficient estimators; Stratification; Greece.

Introduction

This paper presents the optimality problems inherent in the sampling designs of various types of land-based and sea-going large-scale sample surveys conducted under the survey programme of IMBR/HCMR, and the optimality strategies developed for their solution, for the maximization of the precision of the calculated sample estimates for a given cost of sampling.

The land-based biological sample survey of commercial landings (BSCL) aims at the calculation of reliable, simple sample estimates of the vital statistics parameters of the survey species, for given types of fishing vessels and gear, and composite estimates of their length frequency distributions and age compositions (relative, absolute). For monthly data collection, a stratified multi-cluster sampling scheme was used. Information on the survey characteristics from the sample fishing vessel-landings was obtained through trained recorders visiting the sample fishing ports.

The sea-going experimental fishing sample survey (EFSS) aims, among other things, at the calculation of reliable sample estimates of the survey species (mean length estimates, mean weight estimates), and estimates of their length frequency distributions and abundance indices. The survey is based on a stratified sampling scheme of combined space/time (seasons) strata and the selection of sampling stations within them. Measurements were obtained through trained recorders on board the sample fishing vessels.

The land-based commercial landings and effort sample survey (CLES) aims at the calculation of reliable, current (= monthly) sample estimates of fishing fleet statistics, fishing effort statistics, landing statistics and statistics on landings per unit effort. For monthly data collection, three independent large-scale sample surveys were conducted concurrently, based on area sampling techniques: Fishing Zeal Sample Survey for the estimation of the number of fishing vessels, Fishing Effort Sample Survey for the estimation of the exerted fishing effort on the survey stocks and Landing Sample Survey for the estimation of the landings by species of the operating commercial fishing vessels. Measurements of the survey characteristics were obtained through trained field staff.

The optimality strategies developed for the BSCL cover the combined design/estimator aspects of the sampling design of the sample survey. On the design aspect, the optimality problems considered concern the level of representativeness of the area sample of the sample survey, the sampling intensity of the selected length samples in space and time, and the degree of homogeneity of the length frequency distributions of the survey species in space and time. On the estimator aspect, the optimality problems considered concern the statistical efficiency and quality of the existing point and precision estimators by domain of interest, and the sufficiency of the selected length samples.

The optimality strategies developed for the EFSS concern the replacement of the existing empirical estimation system with a probabilistic one for multidomain estimates. The sampling scheme of the sample survey was restructured with the use of combined space/time criteria of various levels of stratification, and the development of integrated, efficient point and precision estimators for multidomain estimates and the population as a whole.

The optimality strategies developed for the CLES concern the design of an efficient probability estimation system to replace the existing empirical one, and the use of the principle of integration. A new probability sampling scheme was also designed for trawling and purse seining fisheries, to deal with the bias effect of the mobility pattern of their fishing vessels on the accuracy of calculated sample estimates.

This paper gives an account of the optimality problems inherent in the sampling designs of the conducted large-scale sample surveys, and presents the optimality strategies developed for their solution, for the maximization of the precision of the calculated sample estimates for a given cost of sampling.

Methodology

The sampling designs set the specifications of the individual large-scale fishery sample surveys under consideration concerning the survey populations, sampling frames, measurement procedures, stratification systems, sampling methods, estimators and auxiliary variables. They are the design/estimator combination, which together constitute a defined goal. The design aspect covers the planning components of the survey design and the estimator aspect covers the estimation system components of the survey design (BAZIGOS, 1983a). A clear distinction is made between empirical (subjective) sampling designs and probability sampling designs. In the optimisation process we select the best probability sampling design which comes as close as possible to realizing the defined goal for a given cost of sampling. Empirical sampling designs of various degrees of subjectivity were used in the conducted large-scale sample surveys for data collection, and assumptions were made for the calculation of required multi-fishery sample estimates affecting their validity. There are great differences among the sampling designs of the individual sample surveys, and the optimality problems which are important in one sample survey may be trivial in another (BAZIGOS, 1983b). The optimisation of the individual large-scale sample surveys are presented under separate headings, covering, in summary, the following three main topics of interest for the survey designs of the sample surveys: objectives, sampling design and optimisation (SHOTTON, & BAZIGOS, 1984). A summary table of abbreviations used in this work, is presented in Table 1.

The optimization of the biological sample survey of commercial landings (BSCL)

Objectives

Biological sampling of the most important species of commercial landings, for given types of fishing vessels and gear used, aimed at the estimation of two different kinds of magnitudes:

a. estimation of vital statistics parameters of the survey species (mean length, mean weight, sex ratio, sexual maturity proportion, mean length at age),

b. calculation of composite estimates of the survey species (length distribution estimates and age composition estimates).

Table 1
Abbreviations.

Abbreviation	
BSCL	Biological Sample Survey of Commercial Landings
CLES	Commercial Landings And Effort Sample Survey
EFSS	Experimental Fishing Sample Survey
ESU	Elementary Sampling Unit
FESS	Fishing Effort Sample Survey
FZSS	Fishing Zeal Sample Survey
LASS	Landing Sample Survey
MES	Multidomain Estimation System
P-SAU	Primary Survey Area Unit
PSU	Primary Sampling Unit
RFV	Registry of Fishing Vessels

Sampling design

The sampling mode of the sampling design of BSCL is a stratified three-stage cluster sampling in space and time. Its structure is presented in Figure 1.

The BSCL is a cost-oriented, large-scale survey; the sizes of the selected samples were predetermined (EU Regulation 1639/2001). Specifically, the sizes of the following two kinds of samples were predetermined:

1. the total sample size of landings (= length sample)
2. the sample sizes of specimens per sample landing:
   i. length sample: 50 specimens
   ii. age sub-sample: 25 specimens

By using empirical criteria, the procedure used for the allocation of the total sample size of landings in space and time strata was:

Step 1: First allocation by survey area, according to their importance in terms of landings.
Step 2: Second allocation by species/gear, according to their importance in terms of landings.
Step 3: Third allocation by time sub-stratum/ calendar month
Step 4: Fourth allocation by sample port. One sample port was allotted in each month (= subjective).

Three kinds of samples were specified: required (= pre-determined), planned (= nominal), achieved (= actual)

Optimization

The optimality process used for the optimization of the sampling design of the BSCL comprises nine components which are listed below. They are discussed in summary under separate headings.

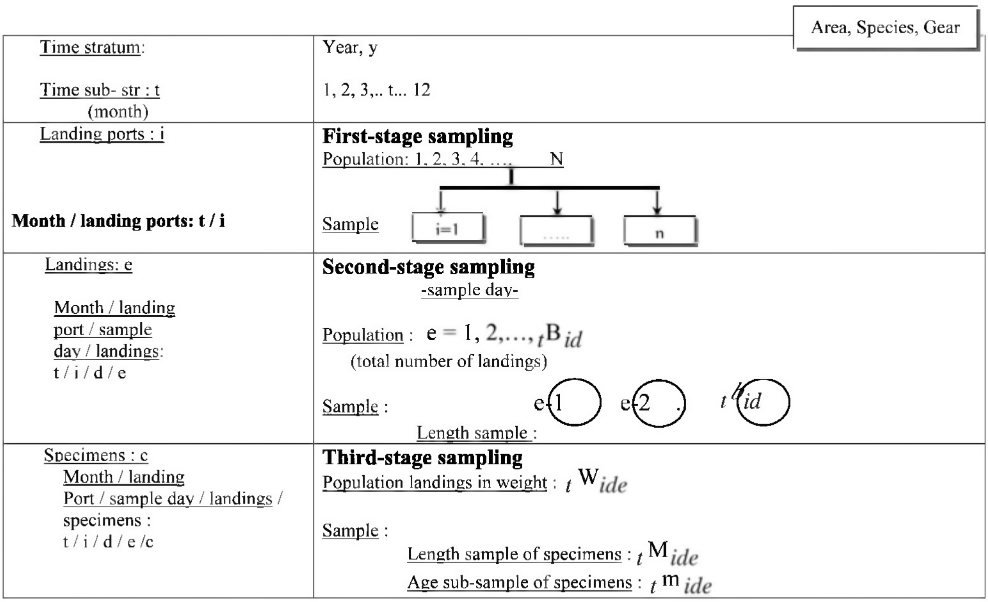

Fig. 1: The statistical structure of the stratified three-stage cluster sampling in space and time.

The structure of the optimality strategies employed is portrayed in the graph below:

The components of the optimization process

1. The level of representativeness of selected fishing ports in space and time 2. The sampling intensity of landings in space and time (= length sample) 3. The level of uniformity of length frequency distributions of the survey species in space and time. 4. The sufficiency of the selected length samples of specimens for the multi-species population 5. The optimization of the sampling design of age composition estimates 6. The optimization of the estimation system of length distribution estimates. 7. The optimization of the estimation system of sex ratio estimates. 8. The efficiency of the estimators of vital statistics parameters used in BSCL 9. The development of efficient estimation procedures by length category of fish estimates A, B, C

The level of representativeness of sample landings ports (= PSUs: Primary Sampling Units)

A series of figures were constructed on a species/gear basis showing the spread of the selected landing ports of the survey in space and time. They indicated the non-random character of the selected area sample, affecting the accuracy of the calculated sample estimates. Further, the inclusion probabilities of selected landing ports were unknown, thus affecting the ability of the selected area sample of the survey to provide population total estimates.

The level of sampling intensity of landings in the sample PSUs

Due to the purposive character of the sampling process used, the property of invariance required in the probability multi-cluster sampling scheme was not held i.e. the use of an identical sub-sampling process within the PSUs. For this reason, sample PSUs with high intensity of sampling carried the most weight in the estimation system, which affected the reliability of the calculated sample estimates.

The uniformity in space and time of length frequency distributions

The Kolmogorov-Smirnov non-parametric test (= two-sample test) was employed, for testing the statistical hypothesis of spatial and temporal uniformity of the length frequency distributions of the selected length samples of the survey species. This information is important in the processes of optimum stratification of the parent population, and the determination and allocation of the number of length samples in space and time (MARE, 1996).

In the application of the tests, the PSUs and months were taken in pairs, and the statistical level of uniformity of the sample length frequency distributions among them was tested with reference to the level of significance of the estimated values of the test-statistics [- : non-significant, * significant (: 5%), ** highly significant (: 1%)]. The results of the tests on the spatial and temporal uniformity of the length sample frequency distributions of the survey species were presented in a series of matrices of uniformity.

The sufficiency of selected length sample sizes of specimens for the multi-species population

Fixed sample sizes of specimens (= 50 specimens), irrespective of species, were pre-determined for the length samples. This approach to sampling introduces bias in the calculated composite sample estimates (= length distribution estimates and age composition estimates), due to the insufficiency of the sample sizes of selected length samples. Insufficient sample sizes affect the level of their representativeness, and therefore the calculated composite sample estimates.

In probability sampling, the sufficiency of the length sample sizes required depends on the skewness of the length distributions of survey species, expressed by their relative Personian measure of kurtosis (=β-coefficient). This in turn means that, in multi-species populations, much bigger length sample sizes of specimens are required for species with a high degree of skewness, than for species with a low degree of skewness. The use of fixed length sample sizes, without taking into account the skewness of length distributions of the survey species, introduces bias into the calculated composite sample estimates (Fig. 2). As is indicated in the figure, insufficient sample sizes produce unrepresentative population length frequency distributions of the survey species, both in shape and location, thus affecting the validity of the calculated sample estimates of means, total and location param-

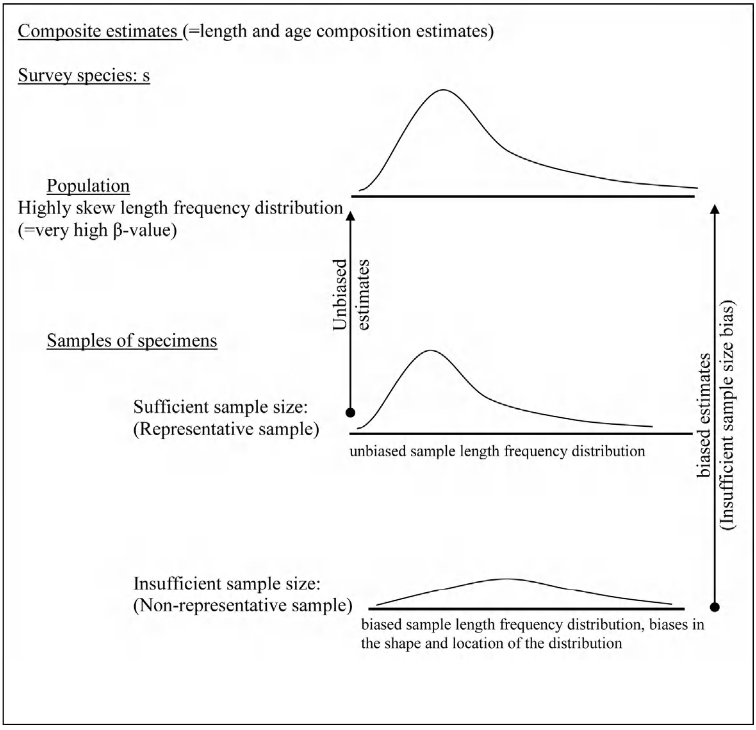

Fig. 2: The biased effect of insufficient sample sizes of specimens in composite estimates.

eters. The upper part of the figure shows the case of representativeness of the sample length frequency distribution based on a sufficient sample size of specimens, and the lower part of the figure shows the case of non-representativeness of the sample length frequency distribution based on an insufficient sample size of specimens.

The optimization of age composition estimates of the survey species

A probability two-phase sampling method (= double-sampling for stratification), was adopted for the selection of age samples. It comprises the following steps:

a. the first-phase length sample determines the relative length distribution of the survey species
b. the second-phase samples for age determination are chosen randomly within the established length intervals (= length strata).
c. the fish to be aged are allotted to the age classes (= age strata) within each length interval.

A systematic presentation of the probability sampling method adopted for the selection of age samples (detailed, combined methods), based on the method of double sampling for stratification is presented in Table 2.

Table 2
The probability sampling method adopted for the selection of age samples.

Length Intervals (j)	**Age classes (k)**						
	1	**2**	**3**	**k**	**....**		
j=1	m_{11}	m_{12}	m_{13}			$m_{1.}$	$M_{1.}$
j=2						$m_{2.}$	$M_{2.}$
j				m_{jk}		$m_{j.}$	$M_{j.}$
. . . .						. . .	. . . .
Total	$m_{.1}$	$m_{.2}$	$m_{.3}$	$m_{.k}$		m	M

Efficient probability estimators based on the method of double sampling for stratification were developed for reliable sample estimates of age composition estimates. Estimates are calculated for the following three levels of estimation:

1. relative age composition estimates
2. absolute age composition estimates
3. absolute age composition estimates by length interval

The probability efficient estimators developed provide reliable point estimates of the parameters of interest, the level of precision of the point estimates and the estimation of the required confidence intervals of the survey parameters, at specified probability levels.

The levels of estimation of the adopted estimation system are:

First level of estimation: relative age composition estimates

Estimation of the population proportion of age k fish in the total commercial landings, and its precision level expressed in terms of sampling variance, sampling error, and relative sampling error (yearly estimates).

Estimators: Double sampling for stratification

$$\begin{pmatrix}\text{Absolute age}\\ \text{composition}\\ \text{estimates}\end{pmatrix} = \begin{pmatrix}\text{Relative age}\\ \text{composition}\\ \text{estimates}\end{pmatrix} * \begin{pmatrix}\text{Total number of speciment in the}\\ \text{commercial landings (estimate)}\end{pmatrix}$$

Second level of estimation: absolute age composition estimates

Estimation of the number of age k fish in the total commercial landings and its precision level (yearly estimates)

Estimators: Simple compound estimators

$$\begin{pmatrix}\text{Absolute age composition}\\ \text{estimates, by length}\\ \text{interval}\end{pmatrix} = \begin{pmatrix}\text{Relative age}\\ \text{composition}\\ \text{estimates}\end{pmatrix} * \begin{pmatrix}\text{Relative length distribution}\\ \text{estimates (lenght sample)}\end{pmatrix}$$

Third level of estimation: absolute age composition estimates by length interval

Estimation of the number of age k fish in the total commercial landings by length interval, and their precision level (yearly estimates)

Estimators: Extended compound estimators

In the developed probability estimation system, the first level estimates (= relative age composition estimates) are the control estimates on which the whole estimation system is built up. Therefore, the maximization of precision of the first level estimates for a given cost of sampling is very desirable.

In the optimization process, the question which arises is how the total sample of fish to be aged can be distributed over the length intervals, so that the sampling variance of the estimates should be minimal. Out of the existing three methods of allocation (= fixed, proportional, optimum), the proportional allocation chosen is the best practical method of allocation. Its Relative Efficiency in precision (=REp) is 20% higher than that of the fixed method of allocation, which was used in the existing estimation system.

The optimization of length distribution estimates

In the optimization process, efficient probability estimators were developed for reliable length distribution estimates of the survey species. For the formulation of the estimators, the following factors were taken into account in combination:

(i) the levels of estimation (lower, upper levels of estimation)
(ii) the hierarchy of the estimation process (landing estimates, fishing port estimates, survey area estimates, population as a whole estimates)
(iii) the winds of estimates (proportions, means, totals)
(iv) the sampling methods of estimation (simple random sampling, stratified sampling, Ratio estimation method, compound method of estimation).

The optimization of sex ratio estimates

The sex ratio of females (F) to males (M) of the individual species is statistically a first type ratio. It measures the average number of females per male. The estimated ratios signify the sex composition of the survey species by length interval. They do not depend on the particular size of either member (F, M), but on the relation in size between them.

In the optimization process, efficient probability estimators were developed for sex ratio R estimates and their level of precision, based on the method of the ratio of two independently estimated parameters (estimated sex ratio- estimated total number of specimens-F/ estimated total number of specimens-M). The efficient estimators developed replaced the biased Ratio estimators used in the existing estimation system.

The optimization of vital statistics parameters estimates

In the existing estimation system, estimators based on the method of simple random sampling were used for the calculation of sample estimates of the:

Mean length,
Mean weight
Mean length at age
Maturity proportion estimates

In the optimization process, efficient probability estimators were developed by taking into account the sampling scheme of the sample survey, in order to improve the precision of the calculated sample estimates.

The efficient estimators by length category of fish A, B, C

In the optimization process, efficient probability estimators were developed for reliable sample estimates of the survey species by length category of fish A, B, C, for the following domains of interest:

1. Length distribution estimates
2. Age composition estimates
3. Mean length at age estimates
4. Sex ratio estimates
5. Sexual maturity estimates

For the formulation of the efficient probability estimators, the length categories of fish were taken as subdivisions of the parent populations, and cutting points of division, corresponding to the range of values of the length categories of fish, were introduced into the established length class distribution estimates of the survey species.

The optimisation of the experimental sampling fishing sample survey (EFSS)

Objectives

The objectives of the conducted cost-oriented EFSS were the collection of the required fishery sample data for accessing the status of the survey stocks in the study areas, and the estimation of abundance and their equilibrium in the ecosystem.

Sampling design

Combined strata of study areas/seasons were established and sample stations were purposively selected within them for data collection. Four time strata (seasons) were defined: spring, summer, autumn and winter.

In the measurement process (deck sampling), the total weight of each sample haul was recorded and the fished species were identified. For the individual commercially important species and species with high abundance in the catch, measurements were obtained on their total weight, number of specimens and their biological data (length, sex, maturity, age).

Estimators based on the method of simple random sampling and simple Ratio estimation methods were employed for mean estimates and abundance indices estimates, respectively.

Optimization of the experimental fishing sample survey (EFSS)

In the optimization process of EFSS, an efficient probability estimation system called the multidomain estimation system (MES) was developed for sample estimates, replacing the existing empirical one,. Its multidomain sampling scheme involved geographical and time divisions, followed by the supporting space and time stratification of various levels, and the selection of independent probability samples within the established combined space/time strata (= cells = basic domains) (UNIVERSITY OF CRETE, DEPARTMENT OF BIOLOGY, 1999).

The objectives of MES were:

1. calculation of reliable sample estimates by domain and the populations as a whole
2. avoiding bias in comparison among domain–estimates
3. minimization of the sampling variance of sample estimates for a given cost of sampling.

Statistically, the creation of multidomains in EFSS from the combination of space-time stratification of various levels designates subpopulations of the parent population, for which separate sample estimates can be calculated by using the principle of integration. The probability structure of the multidomain probability sampling scheme developed for two criteria of stratification (space-survey area, time-season) is portrayed in Figure 3. In the figure, the levels of estimation are also presented. More complex sampling schemes are also used in EFSS, based on three criteria of stratification (space-survey area, depth domain, time-season). They are presented in Figure 4.

The multidomain character of EFSS, envisaged the construction of a multidomain integrated estimation system (= MES), for the presentation of the output of EFSS. For the formulation of MES, the following three factors were taken into account in combination:

1. the modes of sampling used
2. the lines of estimation
3. the levels of estimation

Figure 5 portrays the developed, complete MES of EFSS. It consists of 35 domains of interest, for which probability sample estimates can be calculated.

Integrated efficient probability estimators were developed in MES. They provide reliable sample estimates (= point, precision) by domain, combination of domains and the population as a whole, and for all levels of estimation. Separate estimators were developed for the two types of mul-

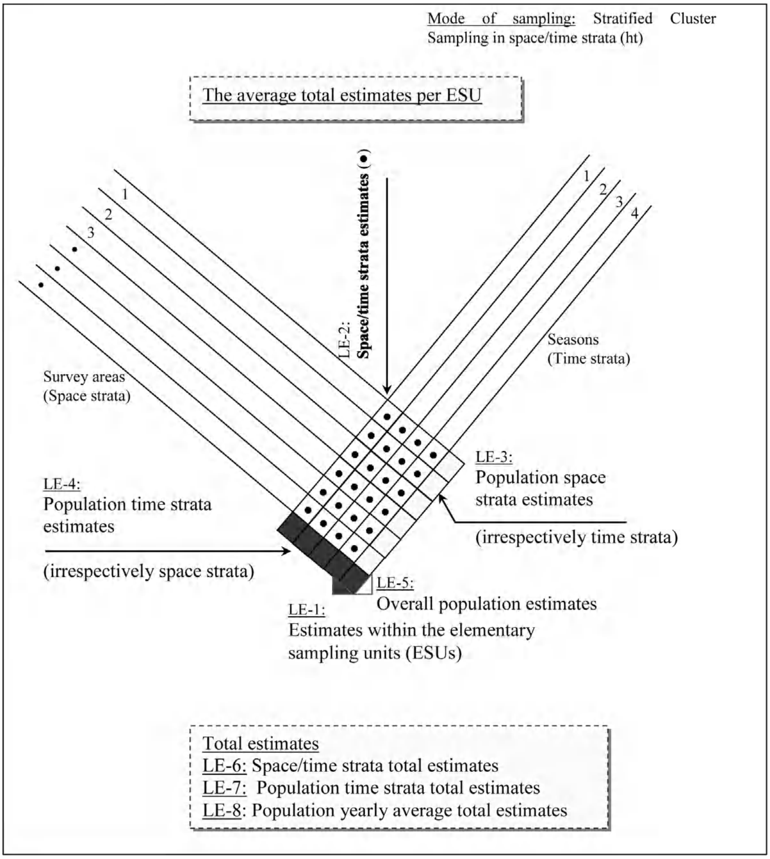

Fig. 3: The probability structure of the multidomain sampling scheme based on the combination of survey/area season stratification (ESU=Elementary Sampling Unit, LE= Level of Estimation).

tidomain sampling scheme i.e. a sampling scheme based on domains with two criteria of stratification, and a sampling scheme based on domains with three criteria of stratification.

The optimization of the commercial landings and effort sample survey (CLES)

Objectives

The CLES is statistically classified within the category of cost-oriented current (monthly) large-scale sample surveys, aiming at the calculation of multiple point estimates of different finite population parameters. The main domains covered by CLES are:

1. Fishing fleet statistics
2. Fishing effort statistics
3. Landing statistics
4. Landings per unit effort statistics

The current results of CLES are used, among other things, to assess the status of operating fisheries and decision-making. Therefore the quality profile of the sample estimates of CLES is of a great interest to

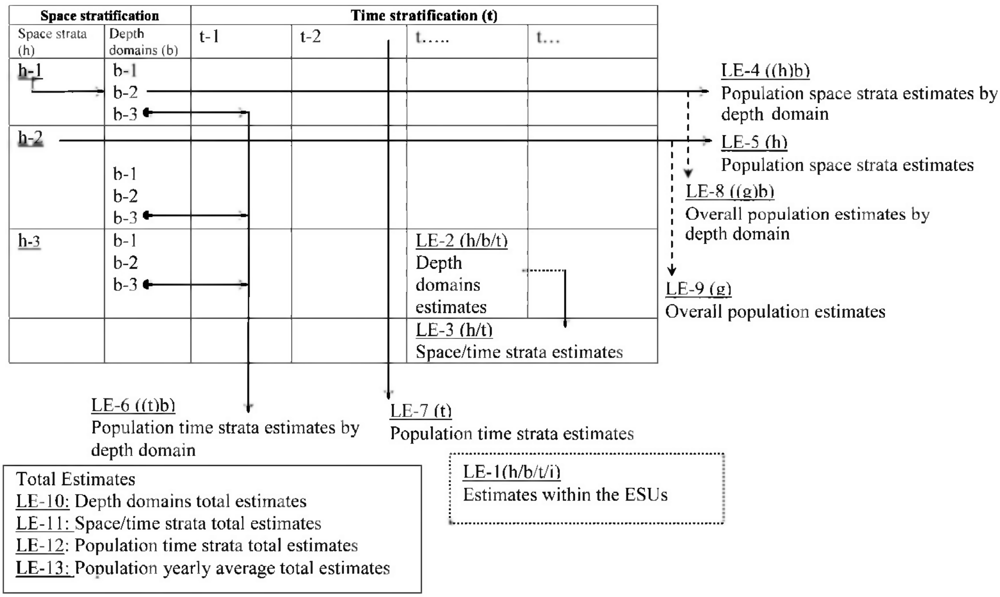

Fig. 4: The probability structure of the multidomain sampling scheme based on the combination of survey area / depth domain / season stratification (ESU=Elementary Sampling Unit, LE= Level of Estimation).

the various users (EU, other users), including its strengths and weaknesses. The level of quality of CLES is the end result of the quality of a series of survey operations, including the sampling frames used, procedures in data collection, measurement procedures, sampling methods and estimation systems.

Sampling design

An in-depth stratification/classification system was used in the sampling design of CLES to improve the level of survey population homogeneity. The space stratification involved area and sub-area space strata, and the time stratification involved yearly and monthly time strata. The space primary survey area unit (P-SAU) was the sampling station (= cluster of fishing ports and landing places), and the time survey unit was the day. The classification variables of the fishing vessels included the type of fishing vessels, gear used, length, activity level (active, inactive) and occupational status (FT-full time, PT-part time, OC-occasional).

Three independent, non-probability, large-scale sample surveys were conducted concurrently for sample data collection: Fishing Zeal Sample Survey (FZSS), Fishing Effort Sample Survey (FESS) and Landings Sample Survey (LASS).

The FZSS was based on stratified cluster sampling, aimed at the collection of monthly sample data from the number of harboured fishing vessels at the sample sampling stations by type, activity level and occupational status, using the arithmetical approach. The counted fishing vessels were subsequently converted into regression units (R) by using the assumption that a multi-regression relationship exists between the regression units and fishing vessels of the form R=1FT+0.5PT+0.25OC). Monthly estimates were also calculated on the total number of regression units in the population by using a subjective estimation procedure. It involves the sample estimation of the proportion of regression units to the counted fishing vessels, and its application to the total monthly population of fishing vessels provided by the existing current (monthly) Registry of Fishing Vessels (RFVs).

The FESS was based on a stratified two-stage sampling method, aimed at monthly estimates of the exerted fishing effort on the survey stocks by the operating active fishing vessels, expressed in the total number of days at sea. Monthly, simple random samples of active full-time fishing vessels were selected within the sample sampling stations for sample data collection. Estimates were first calculated of the average fishing effort per active full-time fishing vessel, and subsequently multiplied by the estimated population total number of regression units (FZSS) for total monthly population fishing effort estimates.

The LASS was based on a stratified two-stage sampling method, aimed at the calculation of the monthly population landing estimates by species of the active fishing vessels. Monthly, simple random samples of active, full-time fishing vessel-

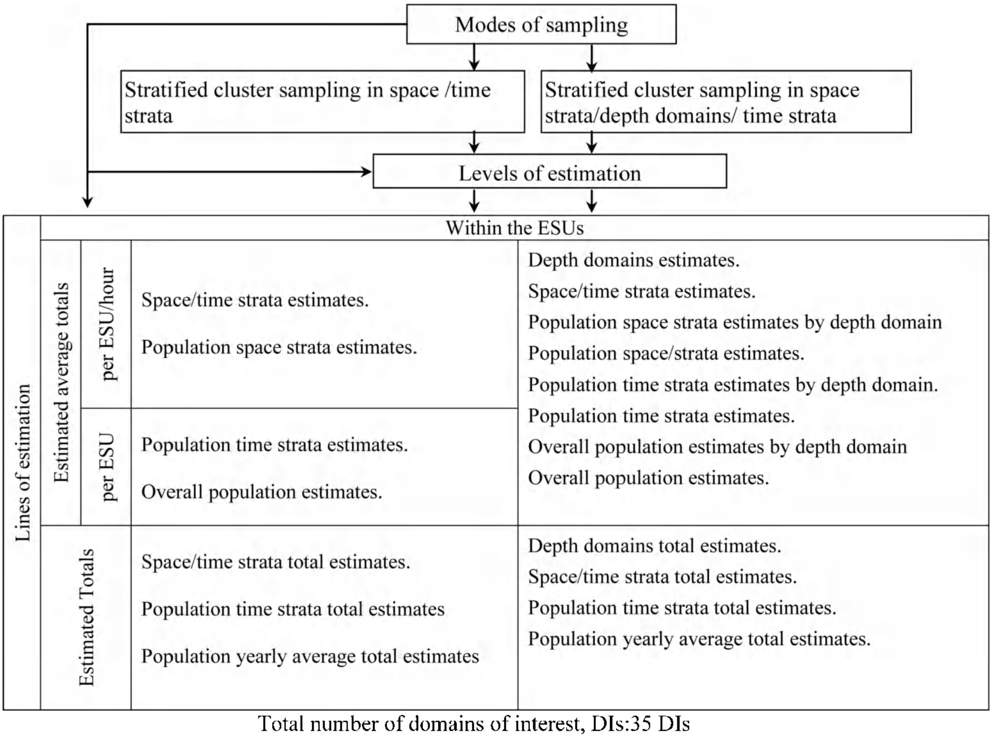

Fig. 5: The multidomain estimation system (MES) (ESU=Elementary Sampling Unit).

landings were selected in the sample sampling stations for sample data collection. Estimates were first calculated of the average landings/day per active full-time fishing vessel, and subsequently multiplied by the estimated population total number of days at sea (FESS) for total monthly population landing estimates.

Optimization

The optimization process developed for the optimization of the sampling design of CLES is presented, in summary, below:

The need for using probability sampling systems

A statistical evaluation of the efficiency and quality of the existing survey design of CLES, and recent demands by the EU for reliable probability sample estimates with a high level of precision expressed in terms of relative sampling error (=RSE =5%), necessitated the switch over from the existing arithmetical survey system of CLES to a probabilistic one, using optimality strategies.

The probability raising factors

(i) The required current (monthly) sampling frames of fishing vessels (= Fs), based on the existing current Registries of Fishing Vessels were constructed, providing reliable information on the size, composition and area distribution of the clustered (= unequal cluster sizes) population of fishing vessels

(ii) The inclusion probabilities of the selected sampling stations were calculated (unequal probabilities), based both on the number of fishing vessels by type and their GT-gross tonnage (= control variables)

(iii) The required probability raising factors of the sample sampling stations were calculated, expressed by the inverse of their inclusion probabilities.

The probability sampling methods

The probability sampling method of stratified two-stage sampling in space and time with unequal probabilities (=PPS-method: probability proportional size), was introduced into the sampling designs of FESS and LASS, by using the principle of integration. For fishing fleet estimates, the stratified cluster method with unequal probabilities was employed. For probability sample estimates of landings per unit effort, the method of stratified two-stage sampling with unequal probabilities, with ratio estimation was used (PAPACONSTANTINOU, KAVADAS, TSIMENIDIS, ECONOMOU, & BAZIGOS, 2002).

In the mode of two-stage sampling, the sampling station was taken as P-SAU (= primary survey area unit, first-stage sampling). Unequal probabilities were used for their selection, in order to control their variability. Sub-sampling (second stage sampling) of fishing vessels (secondary sampling units, SSUs) was taken within the sample P-SAUs (= equal probabilities), by using the principles of invariance and independence. The optimality strategies employed for the maximization of the precision of the calculated sample estimates for a given cost of sampling involved the determination of optimum inclusion probabilities of P-SAUs, and the development of efficient point and precision estimators (= unbiased, precise).

The probability estimators

Integrated efficient point and precision estimators were developed covering all levels of estimation, providing reliable sample estimates of the survey characteristics and their level of precision, expressed in terms of sampling variance, sampling error and relative sampling error.

The new sampling design for trawling and purse seining fisheries (=TPSS)

A new sampling scheme for trawling and purse seining fisheries was designed to deal with the biased effect of the mobility pattern of their fishing vessels.

During the fishing period and survey months, the bulk of trawlers and purse seiners, move away from their home ports to the fishing grounds of the Aegean and Ionian seas. Statistically, they become 'absent' fishing vessels from their home ports and 'simple visitors' to the fishing ports located in their operating areas. In the existing survey system, monthly sample data collected from the 'remaining' present trawlers and purse seiners were considered as representative samples of the absent fishing vessels and were used for the calculation of the required total estimates, introducing substantial 'estimation bias' and 'species composition bias' into the calculated sample estimates.

The designed new probability sampling scheme comprises the following two steps:

Step 1: A Certainty Stratum of P-SAUs was created consisting of the sampling stations where the bulk of the trawlers and purse seiners are located (= probability of selection of P-SAUs equal to 1).

Step 2: In the survey months, within the certainty-selected P-SAUs, the survey trawlers and purse seines were grouped into two categories:
1. present
2. absent.

Separate total estimates are calculated for each of the above two categories of fishing vessels. Total estimates by certainty P-SAU are obtained by adding the estimates of the above two categories of fishing vessels. Certainty stratum total estimates are calculated by adding the estimated totals of the individual P-SAUs. Total population estimates are calculated by adding the certainty stratum estimates to the ones of the remaining population estimates (= total population minus certainty stratum). The probability sampling design of trawlers and purse seiners is fully integrated with the developed integrated probability sampling scheme of CLES. Figures 6 and 7 present the integrated probability sampling design developed for reliable trawling and purse seining sample estimates.

Conclusions

The study carried out within the framework of the survey programme of IMBR/HCMR aimed at the assessment of the quality and efficiency of the design aspect and estimator aspect of the sampling designs of the conducted land-based and sea-going, large-scale sample surveys, and the maximization of the precision of the calculated sample estimates for a given cost of sampling.

A variety of optimality strategies were developed to deal with the optimality problems inherent in the design and estimator aspects of the independently designed and conducted large-scale sam-

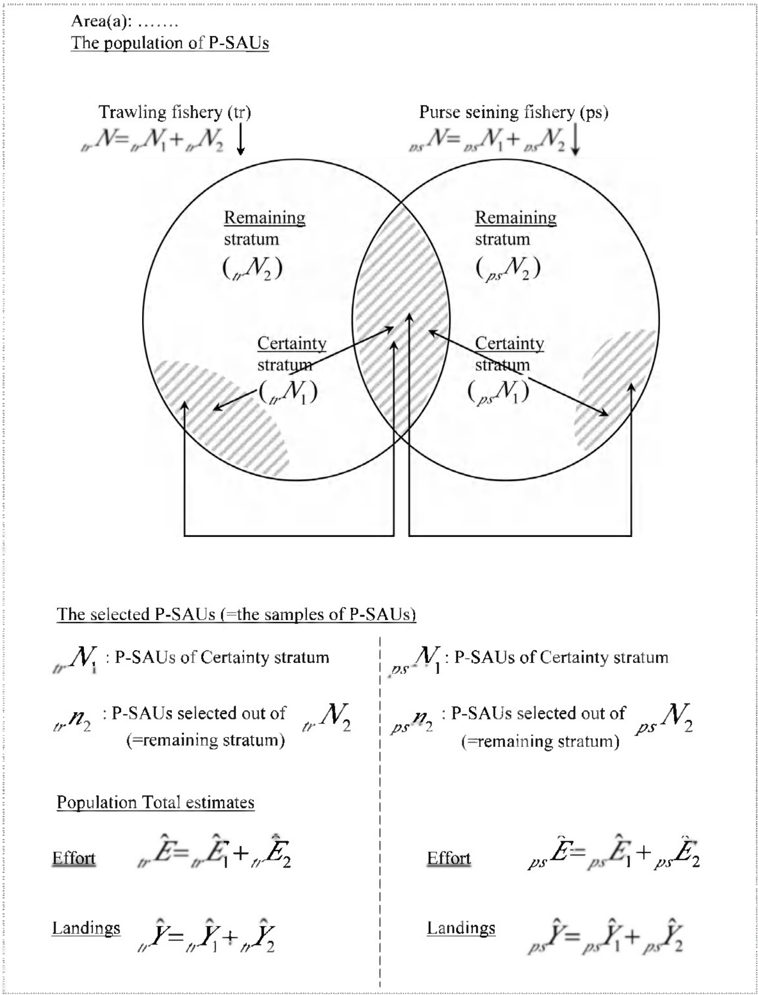

Fig. 6: The integrated probability sampling design for reliable trawling and purse seining sample estimates (P-SAU= Primary Survey Area Unit).

ple surveys, and with the optimization of their sampling designs.

The study provided answers to a series of crucial questions concerning the quality and efficiency of the sampling designs of the various types of large scale sample surveys under consideration:

1. What sampling designs are optimal in a series of sampling designs of the conducted land-based and sea-going, large-scale sample surveys, if the objective is the maximization of the precision of the calculated sample estimates for a given cost of sampling?
2. What are the precise mathematical formulations which can be used for the best designs of the large scale sample surveys?
3. What are the optimality problems and optimality solutions in the estimation

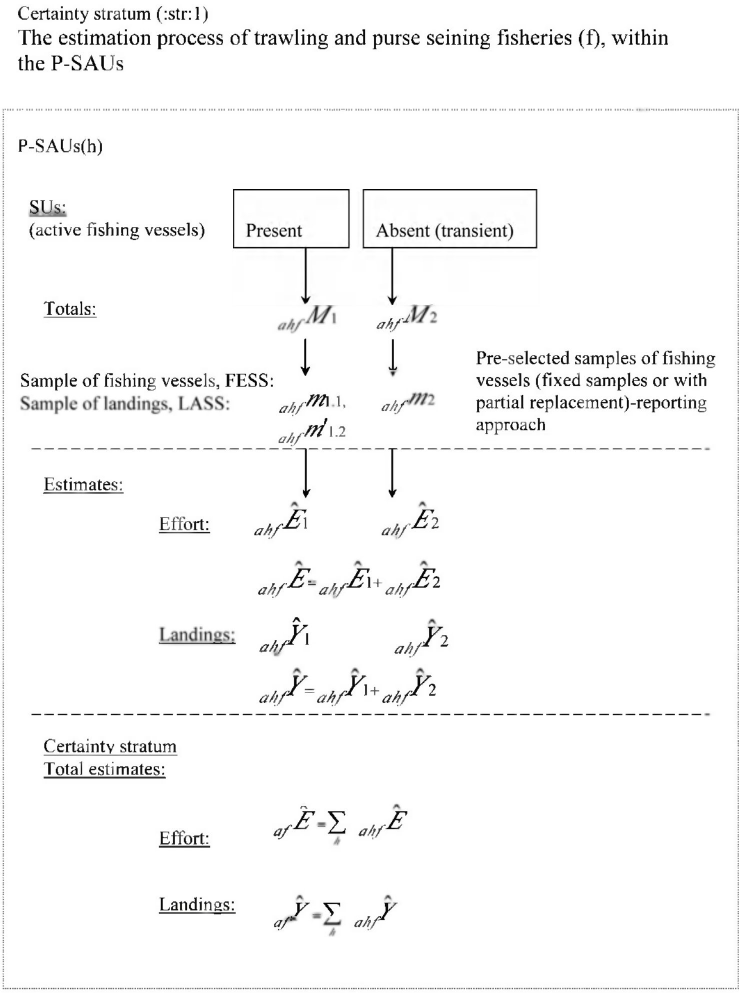

Fig. 7: The estimation process of trawling and purse seining fisheries.

systems of the sampling designs for the optimisation of large scale sample surveys?

In the study, many optimality problems inherent in the sampling designs of the conducted large scale sample surveys were detected and solved, leading to the optimization of their sampling designs and the level of precision of the calculated sample estimates for a given cost of sampling.

The results of the study can be used, among other things, as a source of guidance to other workers of the same field of fishery statistics, on the adequate use of optimality strategies for improving the quality and efficiency of the sampling designs of their respective large scale sample surveys, under the sampling variance criterion and given cost of sampling.

Acknowledgements

To the scientists of IMBR/HCMR responsible for the technical and operational aspects of the conducted large scale sample surveys, C. Mitilineau and J. Charalabus (BSCL, EFSS) and A. Kapantagakis and A. Machias (CLES)) for their cooperation in providing the supporting documentation and information required for the completion of the paper.

References

SHOTTON, R. & BAZIGOS, G., 1984. Techniques and considerations in the design of acoustic surveys, Rapports et Proces-Verbaux des Reunions du Conseil International pour l'Exploration de la Mer, 184: 34-57.

BAZIGOS, G., 1983a. Design of fishery statistical surveys in inland waters, FAO Fisheries Technical Paper 133 (third printing): 1-119.

BAZIGOS, G., 1983b. Applied fishery statistics, FAO Fisheries Technical Paper 135 (third printing): 1-164.

MARE, 1996. Design of a system for the collection and compilation of Basic Fishery Statistics in Greece. (Contract No. 94/067). EU Final report, 103 p. (+Appendices, +Addenda).

UNIVERSITY OF CRETE, DEPARTMENT OF BIOLOGY, 1999. The optimization of the sampling strategies of biological sample surveys and surveys of discards of the commercial catches in Greece and definition of their efficiency expressed in terms of cost: accuracy ratio. (Contract No 97/0034), Final Report.

PAPACONSTANTINOU, C., KAVADAS, S., TSIMENIDIS, N., ECONOMOU, A., & BAZIGOS, G., 2002. Design of a system for the collection and compilation of basic fishery statistics in the Mediterranean: A case study for Greece. *Fisheries Research,* 57 : 155-167.

Accepted in 2007